



HAL
open science

DJ Mix Transcription with Multi-Pass Non-Negative Matrix Factorization

Étienne Paul André, Dominique Fourer, Diemo Schwarz

► **To cite this version:**

Étienne Paul André, Dominique Fourer, Diemo Schwarz. DJ Mix Transcription with Multi-Pass Non-Negative Matrix Factorization. 2024. <hal-04995821>

HAL Id: hal-04995821

<https://hal.science/hal-04995821v1>

Preprint submitted on 18 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

DJ Mix Transcription with Multi-Pass Non-Negative Matrix Factorization

Étienne Paul André

UMR STMS – Ircam, Sorbonne Universités,
CNRS, Ministère de la Culture
Paris, France
andre@ircam.fr

Dominique Fourer

IBISC (EA4526)
Univ. Evry Paris-Saclay
Évry-Courcouronnes, France
dominique.fourer@univ-evry.fr

Diemo Schwarz

UMR STMS – Ircam, Sorbonne Universités,
CNRS, Ministère de la Culture
Paris, France
schwarz@ircam.fr

Abstract—DJ mix transcription is a crucial step towards DJ mix reverse engineering, which estimates the set of parameters and audio effects applied to a set of existing tracks to produce a performative DJ mix. We introduce a new approach based on a multi-pass NMF algorithm where the dictionary matrix corresponds to a set of spectrogram slices of the source tracks present in the mix. The multi-pass strategy is motivated by the high computational cost resulting from the use of a large NMF dictionary. The proposed method uses inter-pass filtering to favor temporal continuity and sparseness and is evaluated on a publicly available dataset. Our comparative results considering a baseline method based on dynamic time warping (DTW) are promising and pave the way of future NMF-based applications.

Index Terms—DJ mix transcription, mix re-engineering, Non-negative Matrix Factorization (NMF), cue point estimation, fade curve estimation.

I. INTRODUCTION

The transcription of DJ mixes is a challenging problem which aims at estimating how music tracks are combined into a DJ set in time and predicts the digital audio processing parameters (eg. fades, equalization, effects). It allows further Music Information Retrieval (MIR) research into DJ practices for musicology and computer support or automation of DJing. Mix transcription often requires the following sub-tasks [1]:

Identification of the tracks contained in the mix to obtain the playlist, **Alignment** of each track to determine the start and end time in the mix and the speed changes applied by the DJ to achieve beat-synchronicity, **Unmixing** to estimate the cue regions where cross-fading is applied and the curves for gain and equalization, and also the parameters of other additional effects (dynamic range compression, echo, etc.), **Content and context analysis** to derive the genre, social tags and other metadata of the music that can inform us about the choices a DJ makes when creating a mix.

While the identification sub-task is well handled by fingerprinting [2]–[5], our work focuses on the alignment and unmixing tasks, and paves the way for future work based on DJ-related MIR content analysis tasks [6]. Thus, we introduce a new method based on multi-pass non-negative matrix factorization (NMF) that is able to extract arbitrary time-warping transformations, including skips or loops, and the mixing gains, while being robust to noise.

II. RELATED WORK

Several existing works focus on *studio mixes*, where a stereo track is produced from multi-track recordings and software instruments by means of a mixing desk or DAW [7]–[12]. The term *mix reverse engineering* (in the context of multi-track studio mixing) was first used for a method to invert linear processing (gains and delays, including short FIR filters typical for EQ) and some dynamic processing parameters (compression) [13], of interest for our aim of DJ unmixing. The unmixing problem was already tackled for radio broadcast mixes [14], i.e. retrieving the fader positions of the mixing desk for several known input signals (music tracks, jingles, reports), and one unknown source (the host and guests’ microphones in the broadcast studio). These two latter references both assume having sample-aligned source signals at their disposal, with no time-scaling applied, unlike our use-case where each source track only covers part of the mix, can appear partially, repeatedly, and out of order, and can be time-scaled for beat-matched mixing. More recent work treats alignment [1], [15]–[20] and mix parameter estimation for unmixing [17], [21], [22]. There is rare work related to analysis [23] and inversion of non-linear processing applied to the signal such as dynamic-range compression [24] which remains challenging and full of interest for unmixing and source separation. Other works on information retrieval from DJ mixes treats content-based analysis of playlist choices [25], and track boundaries estimation in mixes [26], [27].

Existing datasets with ground truth DJ mix annotations can be automatically generated [28], hand-crafted [17], or derived from collaborative annotations [18].

This paper extends our previous study [1] replacing multi-scale alignment using dynamic time warping and subsequent estimation of fade curves by a novel method of combined estimation via multi-pass NMF. It is organized as follows: Section III presents our DJ mix model and our multi-pass NMF method is described in Section IV. Comparative results are presented in Section V before a conclusion in Section VI.

III. MODEL OF THE DJ MIXING PROCESS

A. DJ Mixing Model

Despite the variety of DJ hardware and software options, which come with different features, the DJ’s core workflow

remains the same across all setups. This consistency allows us to outline a general signal path from the recorded tracks to the final mixed output, as illustrated in Fig. 1(a).

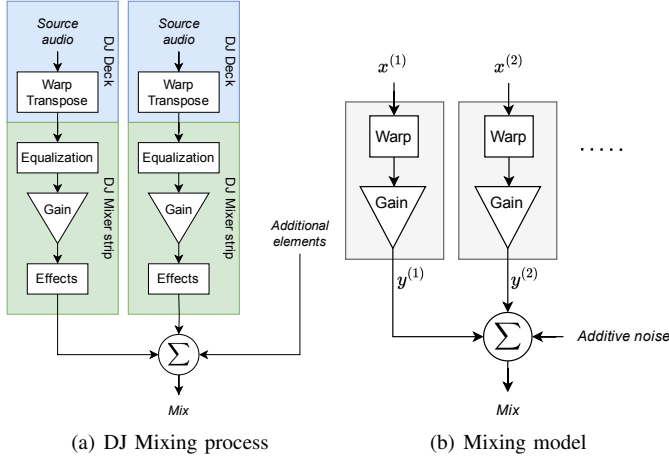


Fig. 1. Generalized DJ mixing process and model.

The process can be described as follows:

- Two or more DJ decks or turntables are used as signal sources. They play pre-recorded tracks and apply pitch-shifting and time-warping.
- The signal from the decks is routed to the DJ mixer, which performs a weighted sum of the input signals after an equalization stage. The mixer can also apply various effects (e.g. reverberation, distortion, bit-crushing) and additional elements (e.g. drum machines, speech, sirens).

We disregard pitch-shifting and equalization to simplify this model to a simple weighted sum of time-warped tracks, illustrated in Fig. 1(b). Effects such as reverberation, distortion, and positive equalization are assumed to be strictly additive and are modeled by additive noise.

B. Capturing Warping and Gain Modulation

Assuming the I source tracks of the mix are known, let $x^{(i)}$ be the signal of length $T^{(i)}$ corresponding to the input track $i \in \{1, 2, \dots, I\}$. The spectrogram matrix $\mathbf{X}^{(i)} \in \mathbb{R}_{M \times T^{(i)}}^+$ of $x^{(i)}$ is defined as the square modulus of the Short Time Fourier Transform (STFT) computed using an analysis window w . We model the time-warping and gain-modulation operations by:

- $f^{(i)} : \tau \mapsto t$ a time-warping injective function that maps a mix time step τ to a track time step t .
- $g^{(i)}$ the positive gain function.

Let $y^{(i)}$ be a signal of length K , time-warped by $f^{(i)}$, and gain-modulated by $g^{(i)}$, transformation of $x^{(i)}$. The spectrogram matrix $\mathbf{Y}^{(i)} \in \mathbb{R}_{M \times K}^+$ of $y^{(i)}$ is defined as follows, and can be expressed in terms of $\mathbf{X}^{(i)}$:

$$\begin{aligned} \mathbf{Y}_{m\tau}^{(i)} &= \left| g^{(i)}[\tau] \sum_{n=0}^{M-1} x^{(i)}[n + f^{(i)}[\tau]] w[n] e^{-j2\pi n \frac{m}{M}} \right|^2 \\ &= g^{(i)}[\tau]^2 \mathbf{X}_{m, f^{(i)}[\tau]}^{(i)} \end{aligned} \quad (1)$$

It follows that the warping and gain-modulation operations can be expressed as a matrix product of $\mathbf{X}^{(i)}$ with an *activation matrix* $\mathbf{H}^{(i)} \in \mathbb{R}_{T^{(i)} \times K}^+$:

$$\mathbf{Y}^{(i)} = \mathbf{X}^{(i)} \mathbf{H}^{(i)}. \quad (2)$$

The *ideal activation matrix* $\tilde{\mathbf{H}}^{(i)}$, can be expressed in terms of the Dirac delta function δ such as:

$$\tilde{\mathbf{H}}_{t\tau}^{(i)} = g^{(i)}[\tau]^2 \delta_{t, f^{(i)}[\tau]} \quad (3)$$

where the warping and gain-modulation operations can be retrieved from the following estimators:

$$\hat{g}^{(i)}[\tau] = \sqrt{\sum_{t=1}^{T^{(i)}} \tilde{\mathbf{H}}_{t\tau}^{(i)}} \quad (4)$$

$$\hat{f}^{(i)}[\tau] = \underset{t \in \{1, 2, \dots, T^{(i)}\}}{\operatorname{argmax}} \tilde{\mathbf{H}}_{t\tau}^{(i)}. \quad (5)$$

C. Matrix Representation of the DJ Mixing Process

Let $\mathbf{Y} \in \mathbb{R}_{M \times K}^+$ be the spectrogram of the mix which can be expressed as the sum of its transformed source tracks' spectrograms $\mathbf{Y}^{(i)}$, and of the residual noise matrix \mathbf{N} as:

$$\mathbf{Y} = \mathbf{N} + \sum_{i=1}^I \mathbf{Y}^{(i)}. \quad (6)$$

Now, we define two additional matrices $\bar{\mathbf{X}}$ and $\bar{\mathbf{H}}$ of compatible dimensions so that $\mathbf{N} = \bar{\mathbf{X}}\bar{\mathbf{H}}$. We can rewrite Eq. (6) as a product of two large matrices $\bar{\mathbf{X}} \in \mathbb{R}_{M \times T}^+$ and $\bar{\mathbf{H}} \in \mathbb{R}_{T \times K}^+$, where $T = \sum_{i=1}^I T^{(i)}$, by concatenation as:

$$\begin{aligned} \mathbf{Y} &= \bar{\mathbf{X}}\bar{\mathbf{H}} + \sum_{i=1}^I \mathbf{X}^{(i)} \mathbf{H}^{(i)} \\ &= \underbrace{(\bar{\mathbf{X}} \ \mathbf{X}^{(1)} \ \mathbf{X}^{(2)} \ \dots \ \mathbf{X}^{(M)})}_{\mathbf{X}} \underbrace{\begin{pmatrix} \bar{\mathbf{H}} \\ \mathbf{H}^{(1)} \\ \mathbf{H}^{(2)} \\ \vdots \\ \mathbf{H}^{(I)} \end{pmatrix}}_{\mathbf{H}}. \end{aligned} \quad (7)$$

Hence, the DJ mix transcription aims to estimate the coefficients of the submatrices $\mathbf{H}^{(1)}$ to $\mathbf{H}^{(I)}$. Moreover, an additive noise profile can be learned with the coefficients of $\bar{\mathbf{X}}$ and $\bar{\mathbf{H}}$. This can be understood as a Non-negative Matrix Factorization (NMF) problem which was intensively used for audio source separation tasks [29].

IV. MULTI-PASS NMF

A. IS-NMF

The NMF algorithm aims to minimise a similarity measure \mathcal{D} between the *target matrix* \mathbf{Y} and the *estimated matrix* $\mathbf{X}\mathbf{H}$. To this end, we use the Itakura-Saito divergence defined as:

$$\begin{aligned} \mathcal{D}_{IS}(\mathbf{Y} \mid \mathbf{X}\mathbf{H}) &= \sum_{f=1}^F \sum_{t=1}^T d_{IS} \left(\mathbf{Y}_{ft} \mid (\mathbf{X}\mathbf{H})_{ft} \right) \\ &\text{with } d_{IS}(x \mid y) = \frac{x}{y} - \ln(xy) - 1 \end{aligned} \quad (8)$$

for which the estimation of $\bar{\mathbf{X}}$ and \mathbf{H} can be computed through the multiplicative gradient descent algorithm presented in Algorithm 1 [30], where \odot and \oslash respectively stand for the Hadamard (element-wise) product and division.

Algorithm 1 IS-NMF function based on partial multiplicative gradient descent.

Input: $\mathbf{Y}, \mathbf{X}, \mathbf{H}$

Initialize \mathbf{H} randomly

repeat

$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{X}^T((\mathbf{X}\mathbf{H})^{-2} \odot \mathbf{Y})}{\mathbf{X}^T(\mathbf{X}\mathbf{H})^{-1}}$$

$$\bar{\mathbf{X}} \leftarrow \bar{\mathbf{X}} \odot \frac{((\mathbf{X}\mathbf{H})^{-2} \odot \mathbf{Y})\mathbf{H}^T}{(\mathbf{X}\mathbf{H})^{-1}\mathbf{H}^T}$$

until convergence criterion is reached

return $\bar{\mathbf{X}}, \mathbf{H}$

B. Multi-pass NMF

Experiments suggest that applying the NMF to music spectrograms produces similarity sub-matrices between \mathbf{Y} and \mathbf{X} instead of the ideal activation matrices described above, on which our estimators yield poor results. Using a large overlap ($> 80\%$) for the analysis windows is crucial to emphasize temporal coherence, but not sufficient.

Algorithm 2 Multi-pass NMF

Input: $hlenInit, hlenTarget, overlap$

$hlen \leftarrow hlenInit$

while $hlen > hlenTarget$ **do**

$wlen \leftarrow hlen \times overlap$

for $i = 1$ **to** I **do**

$\mathbf{X}^{(i)} \leftarrow \text{spectrogram}(x^{(i)}, hlen, wlen)$

if $hlen = hlenInit$ **then**

$\mathbf{H}^{(i)} \leftarrow \text{random noise}$

else

$\mathbf{H}^{(i)} \leftarrow \text{interpassFilter}(\mathbf{H}^{(i)})$

$\mathbf{H}^{(i)} \leftarrow \text{scale}(\mathbf{H}^{(i)}, hlen)$

end if

end for

$\bar{\mathbf{X}}, \bar{\mathbf{H}} \leftarrow \text{random noise}$

$\mathbf{Y} \leftarrow \text{spectrogram}(y, hlen, wlen)$

$\mathbf{X}, \mathbf{H} \leftarrow \text{IS-NMF}(\mathbf{Y}, \mathbf{X}, \mathbf{H})$

$hlen \leftarrow hlen/2$

end while

return \mathbf{H}

Thus, we propose a multi-pass NMF algorithm, described in Algorithm 2 and illustrated in Fig. 2. The method involves performing the NMF on spectrograms with increasing temporal resolution. The activation submatrices obtained at each pass are filtered to remove noise, blurred, thresholded then scaled to serve as initialization for the subsequent pass. The iterative process continues until the desired temporal resolution is reached. Note that, (1) with each increase in resolution, the activation matrices become less noisy and more sparse, (2) regions of the matrix set to zero by the thresholding step

remain zero during the optimization process, thereby avoiding spurious activations, (3) when coupled with an appropriate block-sparse matrix representation, this approach significantly enhances processing efficiency and memory usage, which is efficient for processing a large number of tracks in the DJ mix.

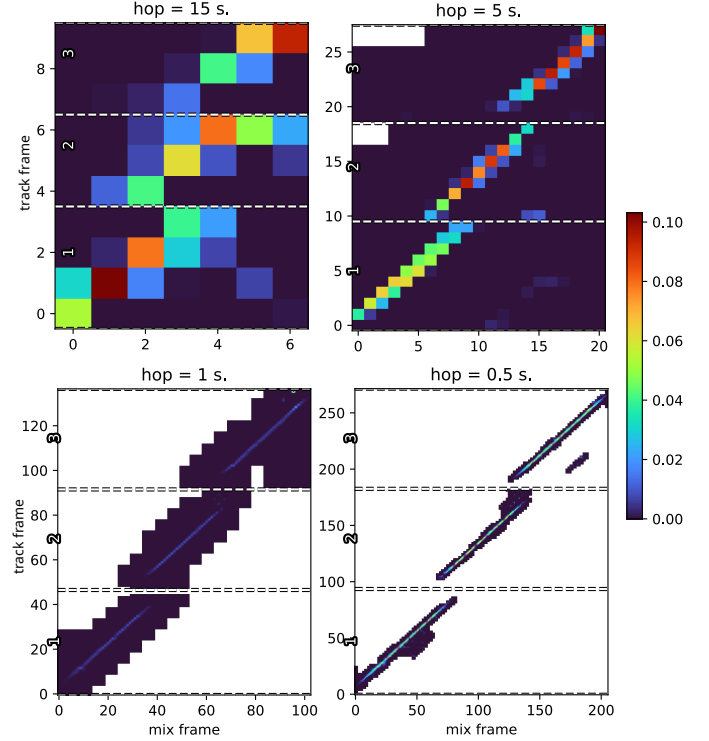


Fig. 2. Example of intermediate activation matrices of the multi-pass NMF algorithm for decreasing hop sizes, on a 3-track mix. Zero-valued cells are colored in white. The submatrices' boundaries are indicated by white dashed lines.

C. Inter-pass Filtering Procedure

The *interpassFilter* procedure of Algorithm 2 performs these steps:

Morphological filtering A line-enhancing filter is applied on each submatrix $\mathbf{H}^{(i)}$, inspired by [31]. The filter kernels $\mathbf{K}(s, d)$ are one-pixel-wide line kernels of slope s and length d . A morphological opening (\circ) is performed and accumulated given slope limits s_{min} and s_{max} , removing activations shorter than d : $\forall i : \mathbf{H}^{(i)} \leftarrow \max_{s \in [s_{min}, s_{max}]} (\mathbf{H}^{(i)} \circ \mathbf{K}(s, d))$

Blurring A Gaussian blur is applied on each submatrix with a small kernel.

Thresholding Activations smaller than a threshold value, that can be considered negligible, are set to zero.

V. RESULTS

The method was implemented in Python and runs on CPU and/or GPU. The code for reproducing all results is available online.¹ To illustrate the method's capacity to extract complex

¹<https://github.com/etiandre/icassp2025-dj-transcription>

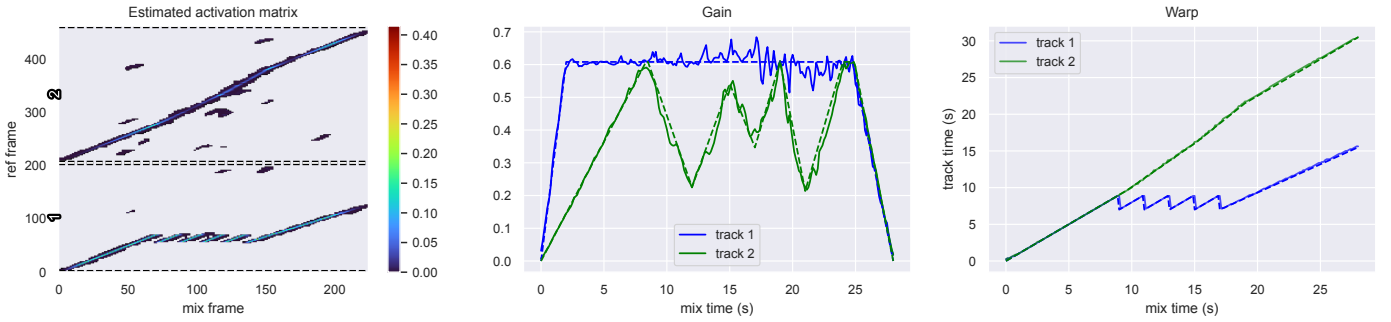


Fig. 3. Estimated activation matrix, gain, and warp, on a mash-up of two tracks with complex gain and warp transformations. Ground truth is represented by a dashed line.

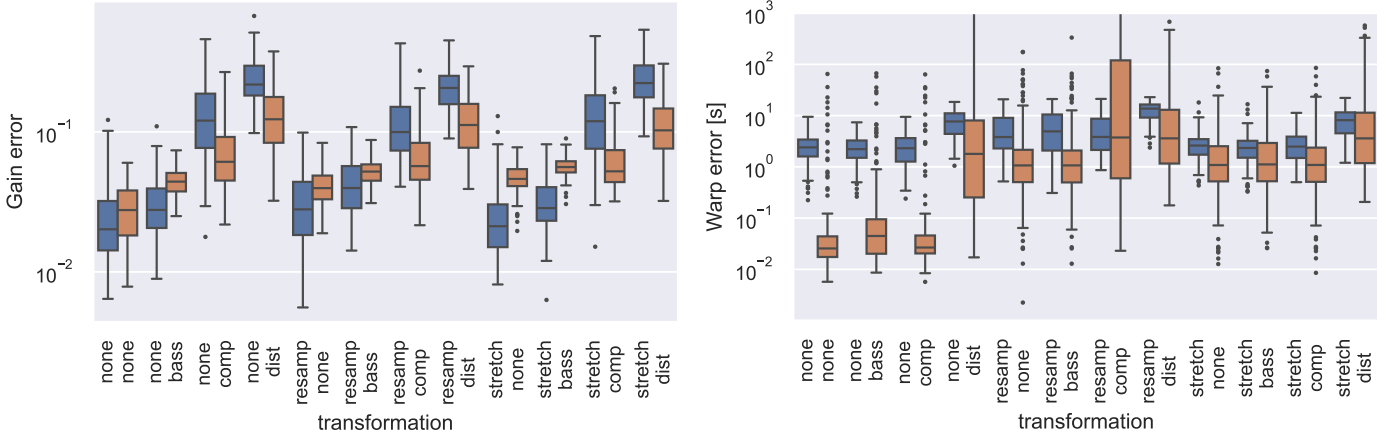


Fig. 4. Box plot of mean absolute estimation error of our method (blue) and DTW-based unmixer (orange), grouped by transformation pair.

warp and gain transformations, Fig. 3 presents the results on a mash-up of two tracks on which fade-in and fade-outs were applied. Track 1 was looped, while track 2 was time-stretched and its gain modulated. We can see that the activation matrix contains very little errors and thus both the time warping and gain curves $f^{(i)}$ and $g^{(i)}$ are estimated with only small deviations.

The algorithm was evaluated on the *UnmixDB* [28] public dataset version 1.1, containing 1931 artificially generated mixes of 3 track excerpts with 3 types of time-warping (*none*, *resampling* changing pitch, *time-stretch* preserving pitch) and 4 effects (*none*, *bass boost*, *compression*, *distortion*). We compare our method with the existing DTW-based unmixer [1] in terms of the mean absolute error metric defined by $\text{MAE}(\hat{x}, x) = \frac{1}{TK} \sum_{i=1}^I \sum_{\tau=1}^K |\hat{x}^{(i)}[\tau] - x^{(i)}[\tau]|$, computed for the gain as $g_{\text{error}} = \text{MAE}(\hat{g}, g)$ and the warp as $f_{\text{error}} = \text{MAE}(\hat{f}, f)$.

We observe in Fig. 4, left, that (1) the gain error is slightly better than the baseline in the no effects and bass boost conditions but slightly worse under the more extreme and unrealistic conditions of compression and distortion, and (2) the model performs well with time-scaling by resampling, even though it is not explicitly pitch-invariant.

However, as shown in Fig. 4, right, our method performs worse in terms of warp error, because it does not impose

any specific assumptions on the form of the time-warping function f . In contrast, the baseline method achieves higher accuracy by assuming an affine time-warping function, which aligns well with the dataset’s fixed time-stretching factors and avoids complexities such as jumps or irregularities in time-scaling.

VI. CONCLUSION

We presented a novel application of NMF for the transcription of a DJ mix paving the way for estimating and undoing the articulations and effects a DJ applies to the source tracks. Our innovative multi-pass algorithm greatly reduces the memory requirements of single-pass NMF by exploiting the inherent sparseness of the activation matrices, making it applicable to realistic mixes of one hour length, and is able to capture loops and jumps. The results on complex test mixes are promising, although the quantitative evaluation on the *UnmixDB* dataset does not always surpass the state of the art, due to the dataset’s unrealistically extreme transformations, and its limitation to constant time-scaling. Future more realistic datasets [32] will be able to better show the strengths and limitations of our algorithm; and research into pitch-invariance or estimation, equalization estimation, and regularization of the NMF algorithm could allow to further improve the method.

REFERENCES

- [1] D. Schwarz and D. Fourer, "Methods and Datasets for DJ-Mix Reverse Engineering," in *Perception, Representations, Image, Sound, Music* (K.-M. R., Y. S., and A. M., eds.), vol. 12631 of *Lecture Notes in Computer Science (LNCS)*, pp. 31–47, Springer, Cham, Mar. 2021. extended version of CMMR 2019 conference article hal-02172427.
- [2] P. Cano, E. Battle, T. Kalker, and J. Haitsma, "A review of audio fingerprinting," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 41, no. 3, pp. 271–284, 2005.
- [3] J. Six, "Olaf : A lightweight, portable audio search system," *Journal of Open Source Software*, vol. 8, no. 87, 2023.
- [4] R. Sonnleitner, A. Arzt, and G. Widmer, "Landmark-based audio fingerprinting for DJ mix monitoring," in *International Symposium on Music Information Retrieval (ISMIR)*, (New York, NY), 2016.
- [5] A. L.-C. Wang, "An industrial strength audio search algorithm.," in *International Symposium on Music Information Retrieval (ISMIR)*, vol. 2003, pp. 7–13, Washington, DC, 2003.
- [6] A. Lerch, *An introduction to audio content analysis: Applications in signal processing and music informatics*. Wiley-IEEE Press, 2012.
- [7] E. Perez-Gonzalez and J. Reiss, "Automatic gain and fader control for live mixing," in *Proc. IEEE WASPAA*, pp. 1–4, Oct. 2009.
- [8] J. A. Maddams, S. Finn, and J. D. Reiss, "An autonomous method for multi-track dynamic range compression," in *Proc. Digital Audio Effects (DAFx)*, 2012.
- [9] S. Mansbridge, S. Finn, and J. D. Reiss, "An autonomous system for multitrack stereo pan positioning," in *Audio Engineering Society Convention*, 2012.
- [10] B. Brecht De Man, R. King, and J. D. Reiss, "An analysis and evaluation of audio features for multitrack music mixtures," in *International Symposium on Music Information Retrieval (ISMIR)*, 2014.
- [11] B. De Man, M. Mora-Mcginity, G. Fazekas, and J. D. Reiss, "The open multitrack testbed," in *Audio Engineering Society Convention 137*, Oct. 2014.
- [12] B. De Man and J. D. Reiss, "Crowd-sourced learning of music production practices through large-scale perceptual evaluation of mixes," *Innovation in Music II*, p. 144, 2016.
- [13] D. Barchiesi and J. Reiss, "Reverse engineering of a mix," *Journal of the Audio Engineering Society*, vol. 58, no. 7/8, pp. 563–576, 2010.
- [14] M. Ramona and G. Richard, "A simple and efficient fader estimator for broadcast radio unmixing," in *Proc. Digital Audio Effects (DAFx)*, pp. 265–268, Sept. 2011.
- [15] S. Ewert, M. Muller, and P. Grosche, "High resolution audio synchronization using chroma onset features," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, (Taipei, Taiwan), pp. 1869–1872, IEEE, Apr. 2009.
- [16] M. Ramona and G. Peeters, "Automatic alignment of audio occurrences: application to the verification and synchronization of audio fingerprinting annotation," in *Proc. Digital Audio Effects (DAFx)*, pp. 429–436, Jan. 2011.
- [17] L. Werthen-Brabants, "Ground truth extraction & transition analysis of DJ mixes," Master's thesis, Ghent University, Belgium, 2018.
- [18] T. Kim, M. Choi, E. Sacks, Y.-H. Yang, and J. Nam, "A Computational Analysis of Real-World DJ Mixes Using Mix-to-Track Subsequence Alignment," *International Symposium on Music Information Retrieval (ISMIR)*, 2020.
- [19] D. Yang, K. Ji, and T. Tsai, "Aligning Unsynchronized Part Recordings to a Full Mix Using Iterative Subtractive Alignment," *International Symposium on Music Information Retrieval (ISMIR)*, 2021.
- [20] J. Six, "DiscStitch : Towards audio-to-audio alignment with robustness to playback speed variabilities," in *Extended Abstracts for the Late-Breaking Demo Session of the 23rd International Society for Music Information Retrieval Conference*, International Symposium on Music Information Retrieval (ISMIR), 2022.
- [21] D. Schwarz and D. Fourer, "Towards Extraction of Ground Truth Data from DJ Mixes," in *Late-breaking Session of ISMIR*, (Suzhou, China), Oct. 2017.
- [22] T. Kim, Y.-H. Yang, and J. Nam, "Joint Estimation of Fader and Equalizer Gains of DJ Mixers Using Convex Optimization," *Proc. Digital Audio Effects (DAFx)*, 2022.
- [23] D. Fourer and G. Peeters, "Objective characterization of audio signal quality: Application to music collection description," in *Proc. IEEE ICASSP*, pp. 711–715, Mar. 2017.
- [24] S. Gorlow and J. D. Reiss, "Model-based inversion of dynamic range compression," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1434–1444, 2013.
- [25] T. Kell and G. Tzanetakis, "Empirical analysis of track selection and ordering in electronic dance music using audio feature extraction.," in *International Symposium on Music Information Retrieval (ISMIR)*, pp. 505–510, 2013.
- [26] T. Scarfe, W. Koolen, and Y. Kalnishkan, "Segmentation of electronic dance music," *International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications*, vol. 22, no. 3, p. 4, 2014.
- [27] N. Glazyrin, "Towards automatic content-based separation of DJ mixes into single tracks," in *International Symposium on Music Information Retrieval (ISMIR)*, pp. 149–154, Oct. 2014.
- [28] D. Schwarz and D. Fourer, "Unmixdb: A dataset for DJ-mix information retrieval," in *Late-breaking Session of ISMIR*, (Paris, France), Sept. 2018.
- [29] P. Smaragdis, "Non-negative Matrix Factor Deconvolution; Extraction of Multiple Sound Sources from Monophonic Inputs," in *Independent Component Analysis and Blind Signal Separation* (C. G. Puntonet and A. Prieto, eds.), vol. 3195, pp. 494–499, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.
- [30] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis," *Neural Computation*, vol. 21, pp. 793–830, Mar. 2009.
- [31] M. Muller and F. Kurth, "Enhancing similarity matrices for music audio analysis," in *Proc. IEEE ICASSP*, vol. 5, pp. V–V, IEEE, June 2006.
- [32] Étienne Paul André, "Dj mix reverse engineering using multi-pass non-negative matrix factorization," Master's thesis, ATIAM, Sorbonne Universités, 2024.