DMRN+14: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2019

QUEEN MARY UNIVERSITY OF LONDON
TUE 17 DECEMBER 2019

Unmixer: An Interface for Extracting and Remixing Loops

Jordan B. L. Smith[1], Yuta Kawasaki[1] and Masataka Goto[1]*

[1]National Institute of Advanced Industrial Science and Technology (AIST), Japan,
unmixer-ml@aist.go.jp

***Abstract*— Someone planning to remix a song would like to have segmented stem tracks at their disposal; that is, isolated instances of the loops and sounds that were used to compose the original song. We present Unmixer, a web service that will analyze and extract loops from any audio uploaded by a user (see Fig. 1). The loops are presented in an interface that allows them to be immediately remixed, and if users upload multiple tracks, they can create mash-ups with the loops, which are automatically matched in tempo.**

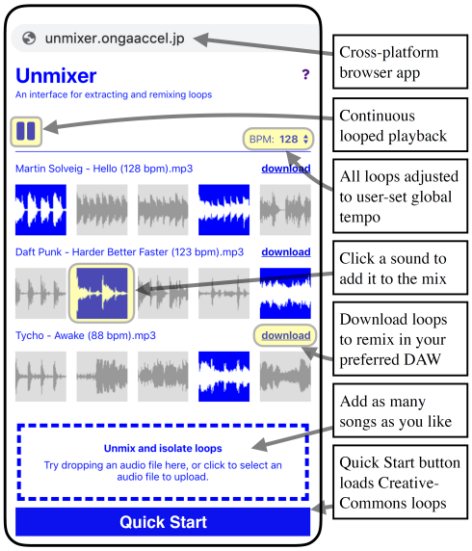

Figure 1: Screenshot of Unmixer website annotated with main features.

I. PROJECT DETAILS

To analyze the audio, we adapt a method of source separation that we recently proposed [1], in which a 2D spectrogram is split at each downbeat and stacked into a 3D "spectral cube", allowing us to model periodic repetitions. We estimate the nonnegative Tucker decomposition, which describes the signal very naturally as the product of a set of sounds, rhythms, and loop activations, the latter directly estimating the compositional layout of the estimated loops (see Fig. 2).

*This work was supported in part by JST ACCEL Grant Number JPMJAC1602, Japan.

J. B. L. Smith was a Visiting Researcher at AIST Japan when this project was initiated.

Y. Kawasaki is a Creative Engineer in the Information Technology Research Institute at AIST Japan.

M. Goto is Prime Senior Researcher at AIST Japan.

All authors may be reached regarding this project at unmixer-ml@aist.go.jp.

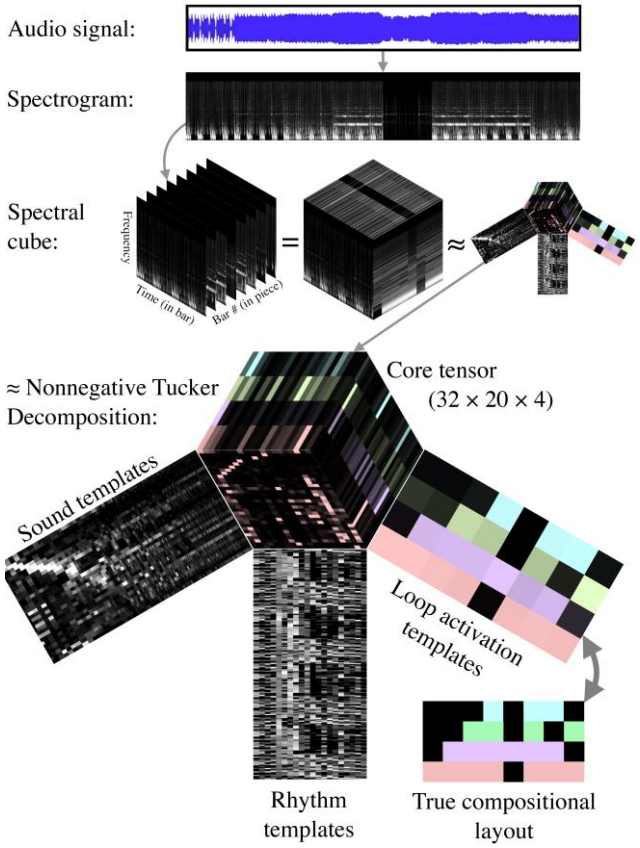

Figure 2: Illustration of non-negative Tucker decomposition being applied to a song of length 8 bars, decomposed as a product of 32 sounds, 20 rhythms, and 4 loop activation templates. The loop activation templates reproduce the true composition of this synthetic example.

To reduce the redundancy of some loops, we propose an extra factorization step with a sparseness constraint and demonstrate (in a test using synthesized pieces) that it improves the source separation result. We also propose a method for selecting the best instances of the extracted loops (maximizing their loudness and minimizing cross-talk from other loops) and demonstrate its effectiveness in an evaluation. Both of these improvements are incorporated into the back end of the interface.

REFERENCES

[1] J. B. L. Smith and M. Goto, "Nonnegative tensor factorization for source separation of loops in audio" in *Proceedings of ICASSP*, Calgary, Canada, pp. 171–175.