User Assisted Separation of Repeating Patterns in Time and Frequency using Magnitude Projections

Derry FitzGerald[1] Zafar Rafii[2] Antoine Liutkus[3]

[1]Cork School of Music, Cork Institute of Technology, Cork, Ireland derry.fitzgerald@cit.ie
[2]Gracenote, Applied Research, Emeryville, CA 94608 USA zrafii@gracenote.com
[3]Inria, Villiers-Lès-Nancy, 54600 France antoine.liutkus@inria.fr

Introduction

– Audio editors let users manipulate recordings but still lack tools to allow for the separation of sounds.
– Proposed user interface systems for audio source separation depend on heavy manual annotations.
+ Researchers have demonstrated the importance of repetition for efficient audio source separation.
= We propose to leverage repetition to perform audio source separation for a more intuitive system.
= This is done by performing PROJET-MAG between similar regions in a time-frequency spectrogram

System

Constant Q Transform is first used to transform a recording into a time-frequency representation with a logarithmic frequency resolution.
Normalized 2D Cross-Correlation is then used to identify the most similar region to a selected region from which a user wishes to recover a repeating pattern.
PROJET-MAG is used to project the two similar regions against each other in order to recover the underlying repeating pattern.

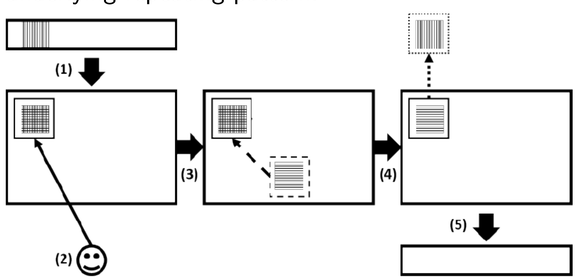

Figure 1: Overview of the system. (1) A signal with an undesired element is transformed into a log-frequency spectrogram. (2) The user selects the region of the undesired element in the spectrogram. (3) The selected region is cross-correlated with the spectrogram to identify the most similar region where the underlying pattern repeats. (4) The identified regions are projected against each other to recover the common repeating pattern. (5) The filtered spectrogram is inverted back to the time-domain with the undesired element removed.

Applications

Recovering a Melody obscured by a Cough

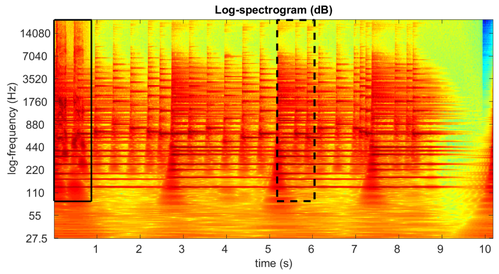

Figure 2: Log-spectrogram of a melody with a cough masking the first notes. The user selected the region of the cough (solid line) and the system identified the most similar region where the underlying notes repeats (dashed line).

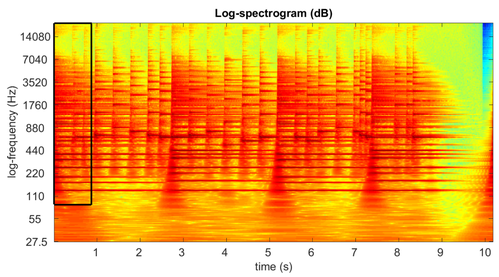

Figure 3: Log-spectrogram of the melody with the first note recovered. The system recovered the repeating elements between the two regions in 2 and filtered the cough from the selected region.

	SDR	SIR	SAR
recovered notes (REP)	8.73	13.63	13.60
extracted cough (REP)	5.91	6.55	11.90
recovered notes (PM)	11.85	16.94	14.99
extracted cough (PM)	8.90	12.76	11.40

Table 1: Separation performance for the recovered note, and the extracted cough for baseline (REP) and PROJET-MAG (PM) (in dB).

The proposed system is tested against a REPET-based algorithm which peforms similar tasks. In practice, the whole process only takes a couple of seconds, as the system involves efficient algorithms.

Recovering an Accompaniment Masked by Vocals

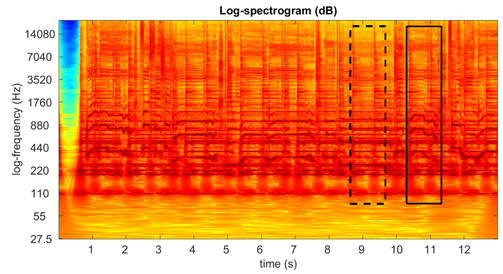

Figure 4: Log-spectrogram of a song with vocals obscuring an accompaniment. The user selected the region (solid line) and the system identified the most similar region where the underlying accompaniment repeats (dashed lines).

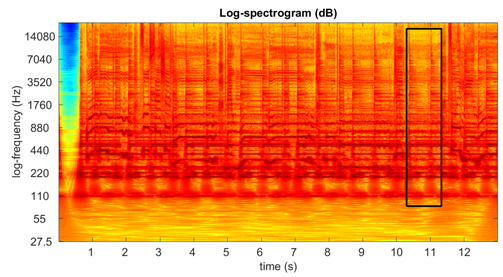

Figure 5: Log-spectrogram of the song with the first measure of the accompaniment recovered. The system recovered the repeating elements between the two regions in 4 and removed the vocals

	SDR	SIR	SAR
recovered accompaniment (REP)	5.11	5.79	13.76
extracted vocal (REP)	6.66	19.83	10.98
recovered accompaniment (PM)	7.52	9.44	11.78
extracted vocal (PM)	9.08	15.03	12.60

Table 2: Separation performance for the recovered accompaniment, and the extracted vocals (in dB).

SDR measures the overall separation performance, with SIR measuring the degree of separation between the sources and SAR measuring the quality of the separation of the estimates. It can be seen the proposed PROJET-MAG method outperforms the baseline REPET-based method on these examples.

Extracting a Speech Masking a Noise

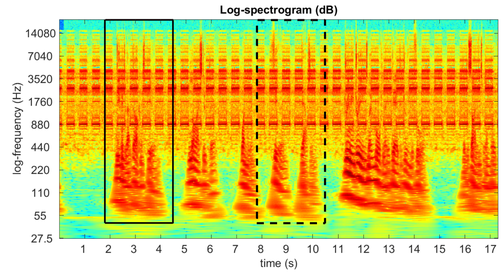

Figure 6: Log-spectrogram of speech masking a noise. The user selected a region of speech (solid line) and the system identified the most similar region where the alarm noise repeats (dashed lines).

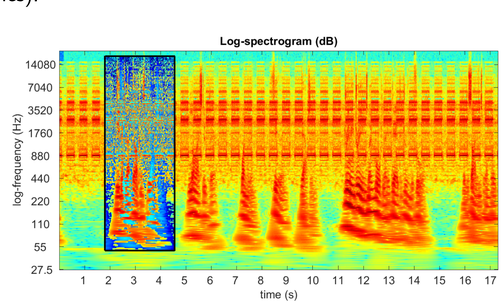

Figure 7: Log-spectrogram of the first sentence of the speech extracted. The system filtered out the repeating pattern once it had been extracted from the two similar regions

	SDR	SIR	SAR
recovered accompaniment (REP)	5.98	15.17	7.77
extracted vocal (REP)	9.71	10.77	15.78
recovered accompaniment (PM)	6.43	12.4	7.38
extracted vocal (PM)	10.17	12.95	13.36

Table 3: Separation performance measures for the extracted speech, and the filtered noise (in dB).

Note that, here, we recovered the non-repeating pattern instead of the repeating pattern.
The reader will find the audio examples online (`http://www.zafarrafii.com/repet.html`).
This work was partly supported by the research programme KAMoulox (ANR-15-CE38-0003-01), funded by ANR, the French State agency for research.