ZE ADVENTUROUS FRENCH AUDIO RESEARCHER (ZAFAR)

Zafar Rafii

PhD in Electrical Engineering & Computer Science

ABSTRACT

We present Zafar, Ze Adventurous French Audio Researcher. The proposed researcher has a PhD in electrical engineering and computer science from Northwestern University, with a focus on audio signal analysis. He has more than 40 publications, including conference papers, journal articles, and patents, with more than 2,400 citations in total. He is actively involved within the research community, as a reviewer, chair, and editor for several conferences and journals, an organizer of a networking meetup group in the San Francisco Bay Area, and an organizer of an international mentoring program for underrepresented groups in music information retrieval. He is currently a research scientist at Audible Magic.

Index Terms— Research, audio signal analysis, separation, recognition, classification.

1. INTRODUCTION

The proposed researcher is named Zafar Rafii. He received a PhD in electrical engineering and computer science from Northwestern University in 2014. He was with the Interactive Audio Lab under the supervision of professor Bryan Pardo. Prior to that, he was a research engineer at Audionamix, in France. He was then a research engineer manager at Gracenote. He is now a research scientist at Audible Magic.

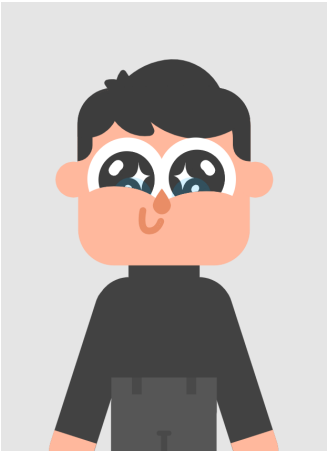

Fig. 1. Overview of the proposed researcher.

The proposed researcher has interest and expertise in audio signal analysis. He has worked on a number of projects, including:

- Blind source separation
- Spatial source separation
- Digital audio effects
- Audio fingerprinting
- Cover song identification
- Audio encoding analysis
- Audio beamforming
- Audio watermarking
- Audio/video segmentation
- Audio classification

For more information on the proposed researcher, the reader is referred to the following materials:

- CV
- GitHub
- LinkedIn
- Google Scholar

For other relevant information related to the proposed researcher, such as the meetups he organizes, the mentoring program he is involved in, or the audio dataset he created, the reader is referred to the following links:

- SF-BISH Bash meetup
- Widening Inclusion in Music Information Retrieval
- MUSDB18 dataset

The rest of the website is organized as follows. In Section 2, we present a selection of projects in which the proposed researcher has worked. In Section 3, we introduce his PhD thesis work on the REpeating Pattern Extraction Technique (REPET) for blind source separation. In Section 4, we share links to his GitHub repositories where some of his source codes reside. In Section 5, we provide references to all of his publications, presentations, and other materials.

2. RESEARCH

2.1. Adaptive Reverberation Tool (2008)

People often think about sound in terms of subjective concepts which do not necessarily have known mappings onto the controls of existing audio tools. For example, a bass player may wish to use a reverberation effect to make her/his bass sound more ”boomy”, but unfortunately there is no ”boomy”

knob to be found. We developed a system that can quickly learn an audio concept from a user (e.g., a "boomy" effect) and generate a simple controller than can manipulate sounds in terms of that audio concept (e.g., make a sound more "boomy"), bypassing the bottleneck of technical knowledge of complex interfaces and individual differences in subjective terms.

For this study, we focused on reverberation effects. We developed a digital reverberator, mapping the parameters of the digital filters to measures of the reverberation effect, so that the reverberator can be controlled through meaningful descriptors such as "reverberation time" or "spectral centroid." In the learning process, a sound is first modified by a series of reverberation settings using the reverberator. The user then listens and rates each modified sound as to how well it fits the audio concept she/he has in mind. The ratings are finally mapped onto the controls of the reverberator and a simple controller is built with which the user is able to manipulate the degree of her/his audio concept on a sound. Several experiments conducted on human subjects showed that the system learns quickly (under 3 minutes), predicts user responses well (mean correlation of 0.75), and meets users' expectations (average human rating of 7.4 out of 10).

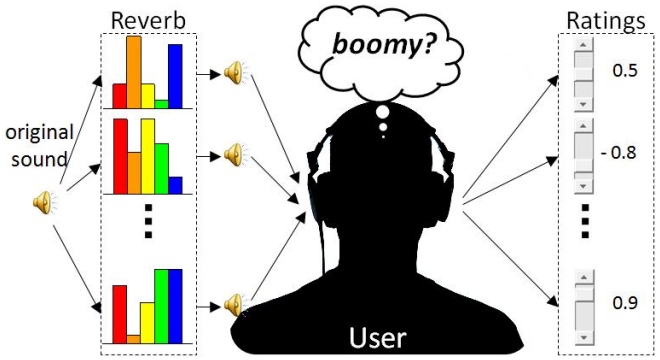

Fig. 2. A listener rating a sound modified by a series of reverberation settings as to how well it fits the audio concept of "boomy" they have in mind.

For more information on this project, the reader is referred to [25], [45], and [48].

2.2. DUET using the CQT (2011)

The Degenerate Unmixing Estimation Technique (DUET) is a blind source separation method that can separate an arbitrary number of unknown sources using a single stereo mixture. DUET builds a two-dimensional histogram from the amplitude ratio and phase difference between channels, where each peak indicates a source, with peak location corresponding to the mixing parameters associated with that source. Provided that the time-frequency bins of the sources do not overlap too much - an assumption generally validated by speech mixtures, DUET partitions the time-frequency representation of the mixture by assigning each bin to the source with the closest mixing parameters. However, when time-frequency bins of the sources start to overlap more - as generally seen in music mixtures when using the common short-time Fourier transform (STFT), peaks start to fuse in the 2d histogram and DUET cannot perform separation effectively.

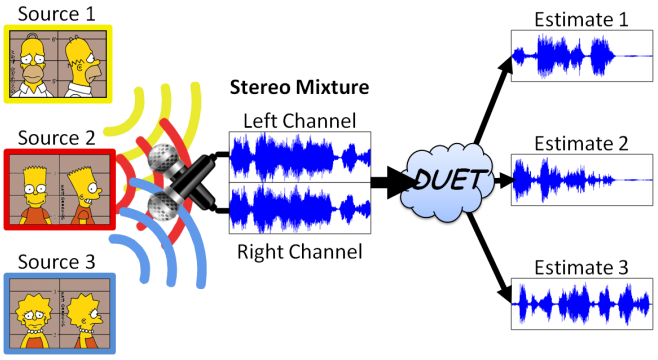

Fig. 3. Blind source separation of a stereo recording of Homer, Bart, and Lisa using DUET.

We proposed to improve peak/source separation in DUET by building the 2d histogram from an alternative time-frequency representation based on the constant-Q transform (CQT). Unlike the Fourier transform, the CQT has a logarithmic frequency resolution, mirroring the human auditory system and matching the geometrically spaced frequencies of the Western music scale, therefore better adapted to music mixtures. We also proposed other contributions to enhance DUET, such as adaptive boundaries for the 2d histogram to improve peak resolving when sources are spatially too close to each other, and Wiener filtering to improve source reconstruction. Experiments on mixtures of piano notes and harmonic sources showed that peak/source separation is overall improved, especially at low octaves (under 200 Hz) and for small mixing angles (under $\pi/6$ rad).

Unlike the classic DUET based on the Fourier transform, DUET combined with the CQT can resolve adjacent pitches in low octaves as well as in high octaves thanks to the log frequency resolution of the CQT:

- Mixture of 3 piano notes
- Estimates: A2 - Bb2 - B2
- Originals: A2 - Bb2 - B2

DUET combined with the CQT and adaptive boundaries helps to improve separation when sources have low pitches (for example, here, between the two cellos) and/or are spatially too close to each other:

- Mixture of 4 instruments
- Estimates: cello 1 - cello 2 - flute - strings
- Originals: cello 1 - cello 2 - flute - strings

For more information on this project, the reader is referred to [44].

2.3. Live Music Fingerprinting (2014)

Suppose that you are at a music festival checking on an artist, and you would like to quickly know about the song that is being played (e.g., title, lyrics, album, etc.). If you have a smartphone, you could record a sample of the live performance and compare it against a database of existing recordings from the artist. Services such as Shazam or SoundHound will not work here, as this is not the typical framework for audio fingerprinting or query-by-humming systems, as a live performance is neither identical to its studio version (e.g., variations in instrumentation, key, tempo, etc.) nor it is a hummed or sung melody. We propose an audio fingerprinting system that can deal with live version identification by using image processing techniques. Compact fingerprints are derived using a log-frequency spectrogram and an adaptive thresholding method, and template matching is performed using the Hamming similarity and the Hough Transform.

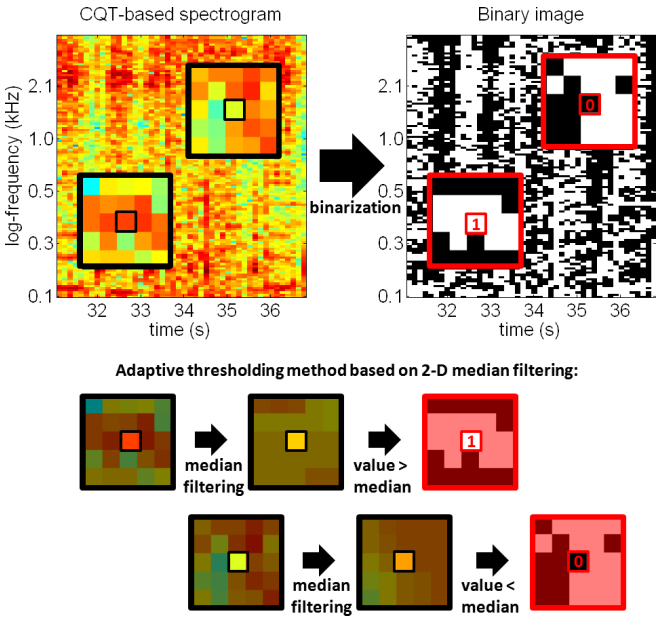

Fig. 4. Overview of the fingerprinting stage. The audio signal is first transformed into a log-frequency spectrogram by using the CQT. The CQT-based spectrogram is then transformed into a binary image by using an adaptive thresholding method.

For more information on this project, the reader is referred to [37].

2.4. Lossy Audio Compression Identification (2018)

We propose a system which can estimate from an audio recording that has previously undergone lossy compression the parameters used for the encoding, and therefore identify the corresponding lossy coding format. The system analyzes the audio signal and searches for the compression parameters and framing conditions which match those used for the encoding. In particular, we propose a new metric for measuring traces of compression which is robust to variations in the audio content and a new method for combining the estimates from multiple audio blocks which can refine the results. We evaluated this system with audio excerpts from songs and movies, compressed into various coding formats, using different bit rates, and captured digitally as well as through analog transfer. Results showed that our system can identify the correct format in almost all cases, even at high bitrates and with distorted audio, with an overall accuracy of 0.96.

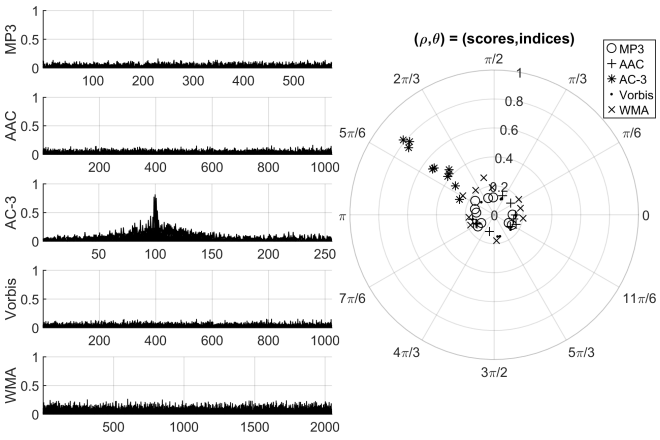

Fig. 5. Results for an audio example encoded with AC-3. The system identified traces of compression corresponding to AC-3, but not to other lossy coding formats such as MP3, AAC, Vorbis, or WMA.

For more information on this project, the reader is referred to [28].

2.5. Sliding DFT with Kernel Windowing (2018)

The sliding discrete Fourier transform (SDFT) is an efficient method for computing the N-point DFT of a given signal starting at a given sample from the N-point DFT of the same signal starting at the previous sample. However, the SDFT does not allow the use of a window function, generally incorporated in the computation of the DFT to reduce spectral leakage, as it would break its sliding property. We show how windowing can be included in the SDFT by using a kernel derived from the window function, while keeping the process computationally efficient. In addition, this approach allows for turning other transforms, such as the modified discrete cosine transform (MDCT), into efficient sliding versions of themselves.

For more information on this project, the reader is referred to [20].

2.6. Constant-Q Harmonic Coefficients (2022)

Timbre is the attribute of sound that makes, for example, two musical instruments playing the same note sound different. It

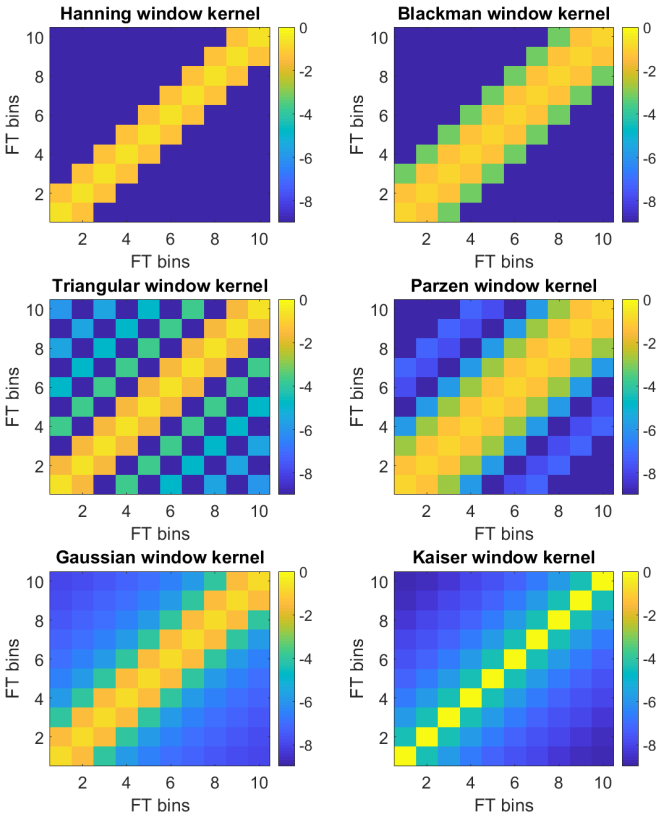

Fig. 6. Kernels derived from the Hanning, Blackman, triangular, Parzen, Gaussian (with $\alpha = 2.5$), and Kaiser (with $\beta = 0.5$) windows. The kernels were derived for an N-point DFT where N = 2,048 samples. Only the first 100 coefficients at the bottom-left corner of the N-by-N kernels are shown. The values are displayed in log of amplitude.

is typically associated with the spectral (but also the temporal) envelope and assumed to be independent from the pitch (but also the loudness) of the sound. We show how to design a simple but effective pitch-independent timbre feature, well adapted to musical data, by deriving it from the constant-Q transform (CQT), a log-frequency transform that matches the typical Western musical scale. The decomposition of the CQT spectrum into an energy-normalized pitch component and a pitch-normalized spectral component is demonstrated, the latter from which a number of harmonic coefficients are extracted. The discriminative powers of these constant-Q harmonic coefficients (CQHCs) are then evaluated on the NSynth data set, a publicly available, large-scale data set of musical notes, where they are compared with the mel-frequency cepstral coefficients (MFCCs), a feature originally designed for speech recognition but commonly used to characterize timbre in music.

An online Python implementation of CQHCs is provided with some examples here. For more information on this project, the reader is referred to [19].

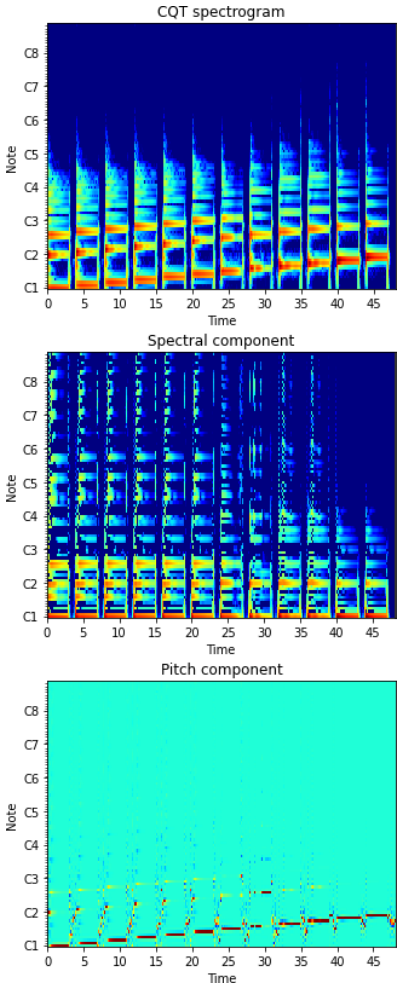

Fig. 7. The deconvolution of the CQT spectrogram (shown in dB) of 12 acoustic bass notes playing from C1 to B1, into a pitch-normalized spectral component (shown in dB), and an energy-normalized pitch component (shown in [0, 1]).

2.7. Cheap Music Detection using a Simple Multilayer Perceptron with Temporal Integration (2024)

We will show how to design a cheap system for detecting when music is present in audio recordings. We will make use of a small neural network consisting of a simple multilayer perceptron (MLP), along with compact features derived from the mel spectrogram by means of temporal integration. Temporal integration is the process of combining information over time, for example, by computing statistics over a sequence of spectra in audio. We will experiment with common statistics and compare the performances of various small MLPs, with each other and with more complex models, on two large music detection datasets. We note that this article does not claim to propose a state-of-the-art music detection model but rather to demonstrate how anybody can design their own efficient system from scratch using simple ideas and with limited re-

sources.

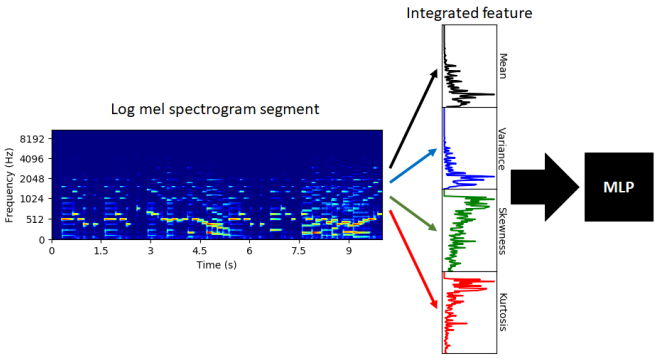

Fig. 8. The temporal integration over a 10-s segment of a log mel spectrogram by computing the first four moments (mean, variance, skewness, and kurtosis) and concatenating them into a 1D feature that can then be fed into an MLP..

For more information on this project, the reader is referred to [18].

3. REPET

Repetition is a fundamental element in generating and perceiving structure. In audio, mixtures are often composed of structures where a repeating background signal is superimposed with a varying foreground signal. On this basis, we present the REpeating Pattern Extraction Technique (REPET), a simple approach for separating the repeating background from the non-repeating foreground in an audio mixture. The basic idea is to find the repeating elements in the mixture, derive the underlying repeating models, and extract the repeating background by comparing the models to the mixture. Unlike other separation approaches, REPET does not depend on special parameterizations, does not rely on complex frameworks, and does not require external information. Because it is only based on repetition, it has the advantage of being simple, fast, blind, and therefore completely and easily automatable.

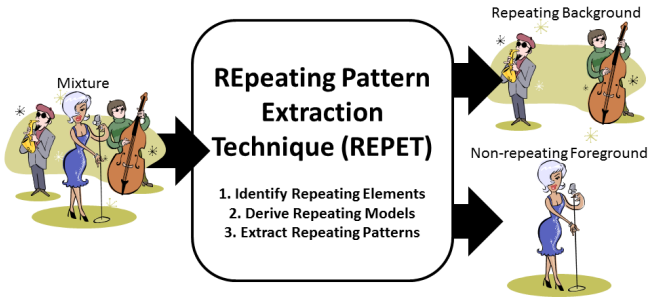

Fig. 9. Overview of REPET.

3.1. Original REPET (2011)

The original REPET aims at identifying and extracting the repeating patterns in an audio mixture, by estimating a period of the underlying repeating structure and modeling a segment of the periodically repeating background.

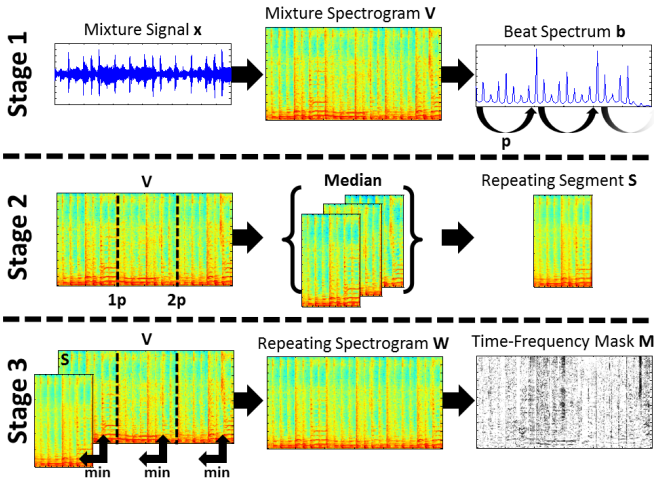

Fig. 10. Overview of the original REPET. Stage 1: calculation of the beat spectrum b and estimation of a repeating period p. Stage 2: segmentation of the mixture spectrogram V and calculation of the repeating segment S. Stage 3: calculation of the repeating spectrogram W and derivation of the time-frequency mask M.

Experiments on a data set of song clips showed that REPET can be effectively applied for music/voice separation. Experiments also showed that REPET can be combined with other methods to improve background/foreground separation; for example, it can be used as a preprocessor to pitch detection algorithms to improve melody extraction, or as a postprocessor to a singing voice separation algorithm to improve music/voice separation.

- Mixture
- Estimates: background - foreground
- Originals: accompaniment - vocals

REPET can be easily extended to handle varying repeating structures, by simply applying the method along time, on individual segments or via a sliding window. Experiments on a data set of full-track real-world songs showed that this method can be effectively applied for music/voice separation.

For more information on this project, the reader is referred to [43], [24], and [47].

3.2. Adaptive REPET (2012)

The original REPET works well when the repeating background is relatively stable (e.g., a verse or the chorus in a song); however, the repeating background can also vary over

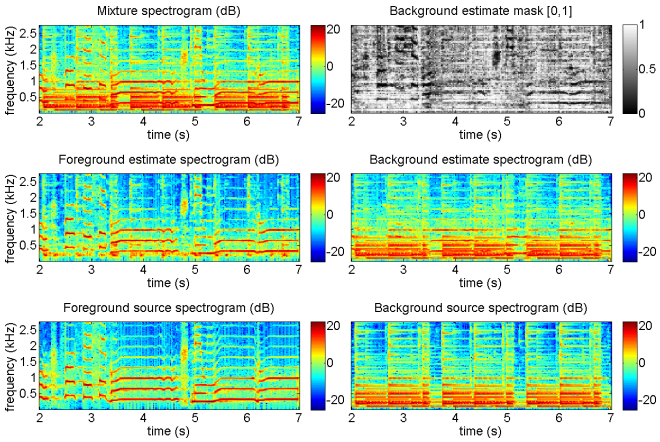

Fig. 11. Music/voice separation using REPET. The mixture is a female singer (foreground) singing over a guitar accompaniment (background). The guitar has a repeating chord progression that is stable along the song. The spectrograms and the mask are shown for 5 seconds and up to 2.5 kHz.

time (e.g., a verse followed by the chorus in the song). The adaptive REPET is an extension of the original REPET that can handle varying repeating structures, by estimating the time-varying repeating periods and extracting the repeating background locally, without the need for segmentation or windowing.

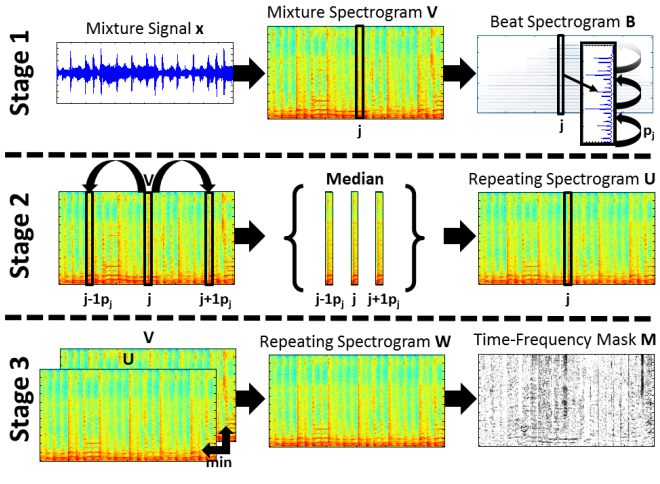

Fig. 12. Overview of the adaptive REPET. Stage 1: calculation of the beat spectrogram B and estimation of the repeating periods pj's. Stage 2: filtering of the mixture spectrogram V and calculation of an initial repeating spectrogram U. Stage 3: calculation of the refined repeating spectrogram W and derivation of the time-frequency mask M.

Experiments on a data set of full-track real-world songs showed that the adaptive REPET can be effectively applied for music/voice separation.

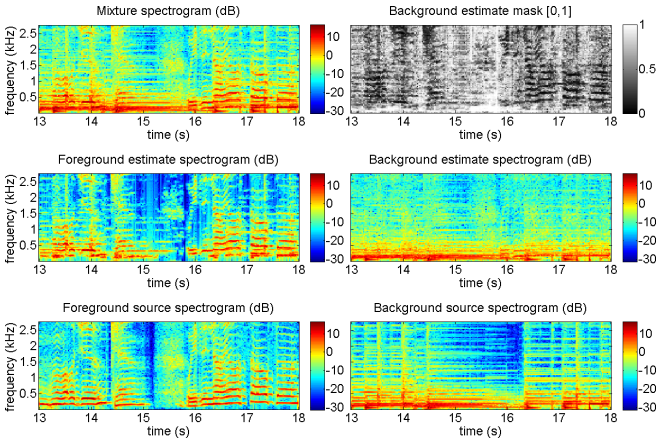

Fig. 13. Music/voice separation using the adaptive REPET. The mixture is a male singer (foreground) singing over a guitar and drums accompaniment (background). The guitar has a repeating chord progression that changes around 15 seconds. The spectrograms and the mask are shown for 5 seconds and up to 2.5 kHz.

- Mixture
- Estimates: background - foreground
- Originals: accompaniment - vocals

For more information on this project, the reader is referred to [41] and [47].

3.3. REPET-SIM (2012)

The REPET methods work well when the repeating background has periodically repeating patterns (e.g., jackhammer noise); however, the repeating patterns can also happen intermittently or without a global or local periodicity (e.g., frogs by a pond). REPET-SIM is a generalization of REPET that can also handle non-periodically repeating structures, by using a similarity matrix to identify the repeating elements.

Experiments on a data set of full-track real-world songs showed that REPET-SIM can be effectively applied for music/voice separation.

REPET-SIM can be easily implemented online to handle real-time computing, particularly for real-time speech enhancement. The online REPET-SIM simply processes the time frames of the mixture one after the other given a buffer that temporally stores past frames. Experiments on a data set of two-channel mixtures of one speech source and real-world background noise showed that the online REPET-SIM can be effectively applied for real-time speech enhancement.

- Mixture
- Estimates: foreground - background
- Originals: speech - noise

For more information on this project, the reader is referred to [40], [39], and [47].

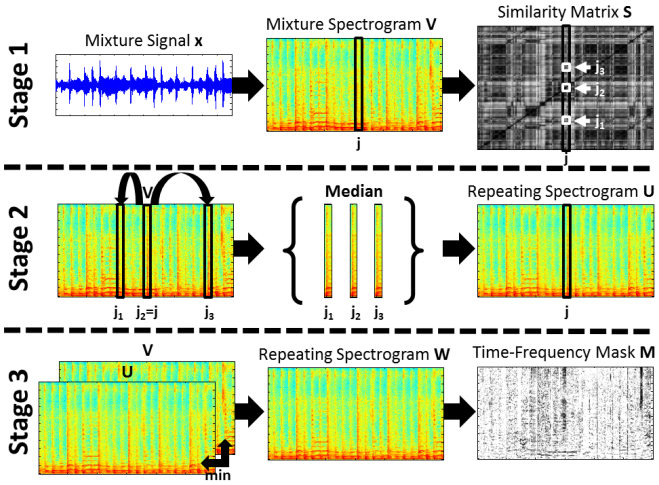

Fig. 14. Overview of REPET-SIM. Stage 1: calculation of the similarity matrix S and estimation of the repeating indices jk's. Stage 2: filtering of the mixture spectrogram V and calculation of an initial repeating spectrogram U. Stage 3: calculation of the refined repeating spectrogram W and derivation of the time-frequency mask M.

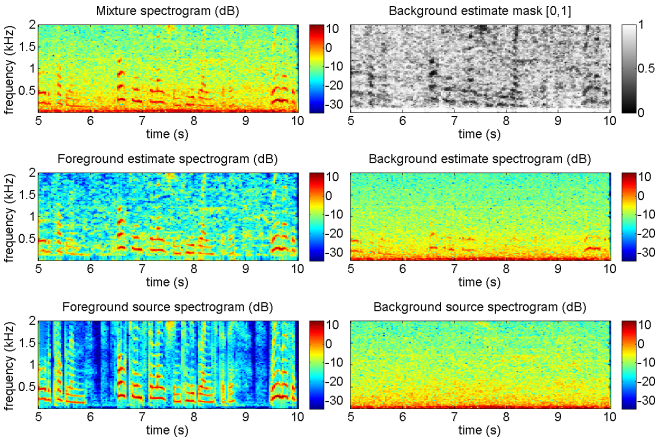

Fig. 15. Noise/speech separation using REPET-SIM. The mixture is a female speaker (foreground) speaking in a town square (background). The square has repeating noisy elements (passers-by and cars) that happen intermittently. The spectrograms and the mask are shown for 5 seconds and up to 2 kHz.

3.4. uREPET (2015)

Repetition is a fundamental element in generating and perceiving structure in audio. Especially in music, structures tend to be composed of patterns that repeat through time (e.g., rhythmic elements in a musical accompaniment), and also frequency (e.g., different notes of the same instrument). The auditory system has the remarkable ability to parse such patterns by identifying repetitions within the audio mixture. On this basis, we propose a simple user interface system for recovering patterns repeating in time and frequency in mixtures of sounds. A user selects a region in the log-frequency spectrogram of an audio recording from which she/he wishes to recover a repeating pattern covered by an undesired element (e.g., a note covered by a cough). The selected region is then cross-correlated with the spectrogram to identify similar regions where the underlying pattern repeats. The identified regions are finally averaged over their repetitions and the repeating pattern is recovered.

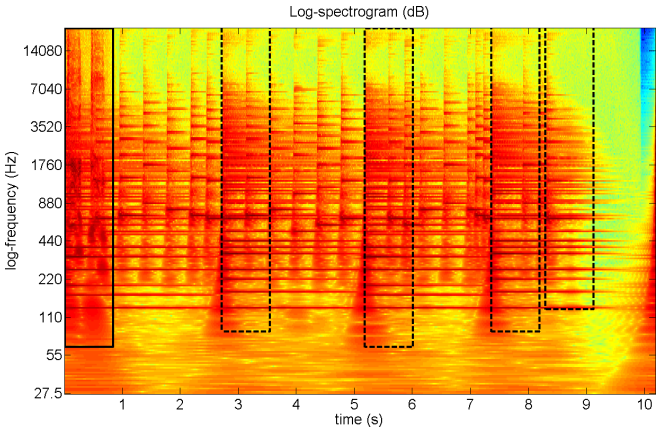

Fig. 16. Log-spectrogram of a melody with a cough covering the first note. The user selected the region of the cough (solid line) and the system identified similar regions where the underlying note repeats (dashed lines).

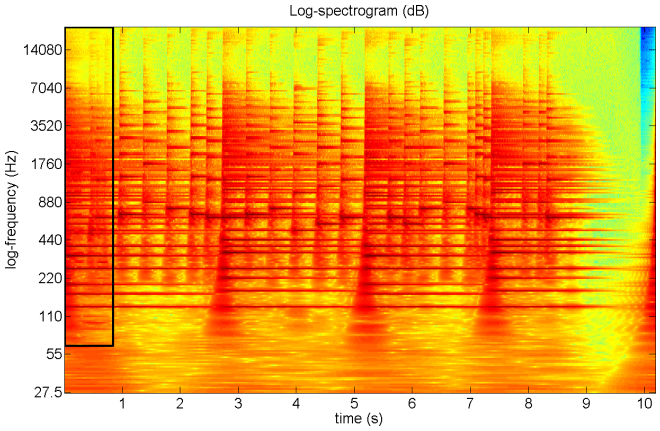

Fig. 17. Log-spectrogram of the melody with the first note recovered. The system averaged the identified regions over their repetitions and filtered out the cough from the selected region.

- Melody covered by a cough
- Recovered melody
- Original melody - original cough
- Accompaniment covered by vocals

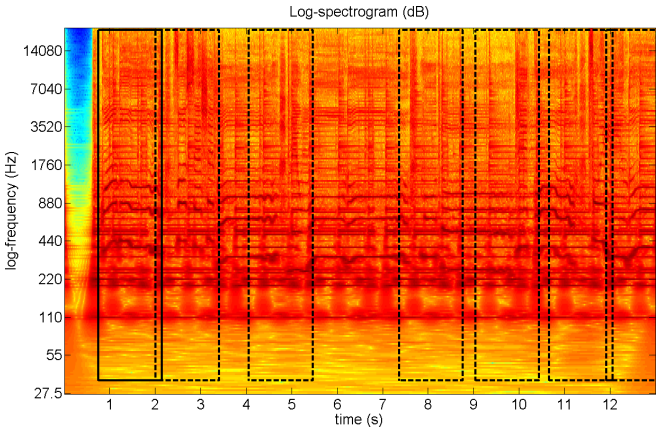

Fig. 18. Log-spectrogram of a song with vocals covering an accompaniment. The user selected the region of the first measure (solid line) and the system identified similar regions where the underlying accompaniment repeats (dashed lines).

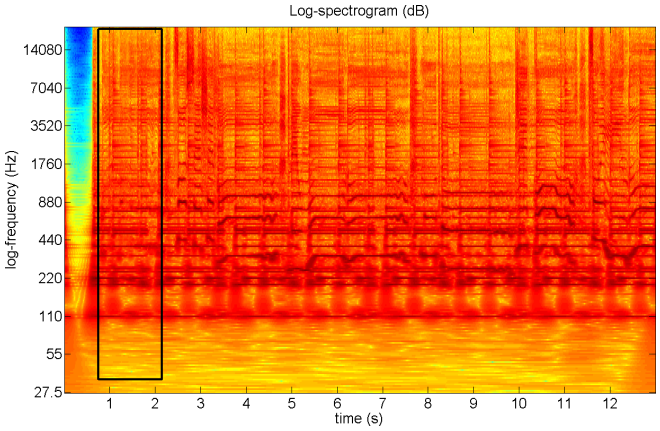

Fig. 19. Log-spectrogram of the song with the first measure of the accompaniment recovered. The system averaged the identified regions over their repetitions and filtered out the vocals from the selected region.

- Recovered accompaniment
- Original accompaniment - original vocals
- Speech covering a noise
- Recovered speech
- Original speech - original noise

For more information on this project, the reader is referred to [33].

3.5. PROJET-MAG (2017)

We propose a simple user-assisted method for the recovery of repeating patterns in time and frequency which can occur in mixtures of sounds. Here, the user selects a region in a logfrequency spectrogram from which they seek to recover the underlying pattern which is obscured by another

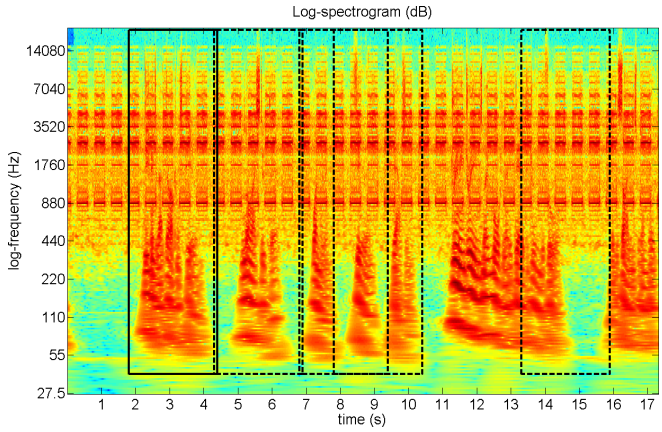

Fig. 20. Log-spectrogram of a speech covering a noise. The user selected the region of the first sentence (solid line) and the system identified similar regions where the underlying noise repeats (dashed lines).

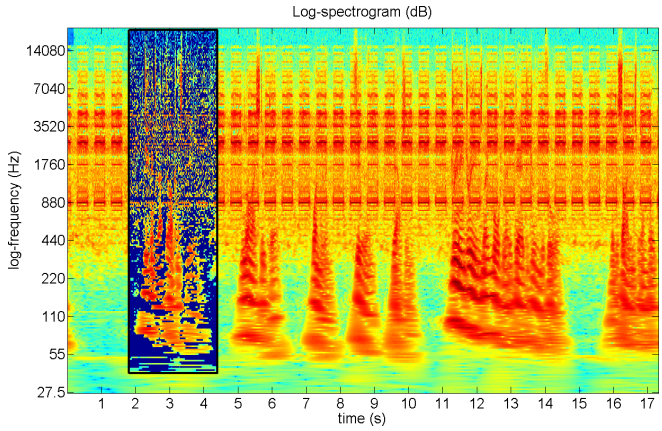

Fig. 21. Log-spectrogram of the first sentence of the speech extracted. The system averaged the identified regions over their repetitions and extracted the speech from the selected region.

interfering source, such as a chord masked by a cough. A cross-correlation is then performed between the selected region and the spectrogram, revealing similar regions. The most similar region is then selected and a variant on the PROJET algorithm, termed PROJET-MAG, is then used to extract the common time-frequency components from the two regions, as well as extracting the components which are not common to both. The results obtained are compared to another user-assisted method based on REPET, and the PROJET-MAG method is demonstrated to give improved results over this baseline.

- Melody covered by a cough
- Recovered melody uREPET - PROJET-MAG
- Original melody - original cough
- Accompaniment covered by vocals

- Recovered accompaniment uREPET - PROJET-MAG
- Original accompaniment - original vocals
- Speech covering a noise
- Recovered speech uREPET - PROJET-MAG
- Original speech - original noise

For more information on this project, the reader is referred to [30].

4. CODES

4.1. Zaf-Python

This GitHub repository contains Zafar's Audio Functions in **Python** for audio signal analysis: STFT, inverse STFT, mel filterbank, mel spectrogram, MFCC, CQT kernel, CQT spectrogram, CQT chromagram, DCT, DST, MDCT, inverse MDCT.

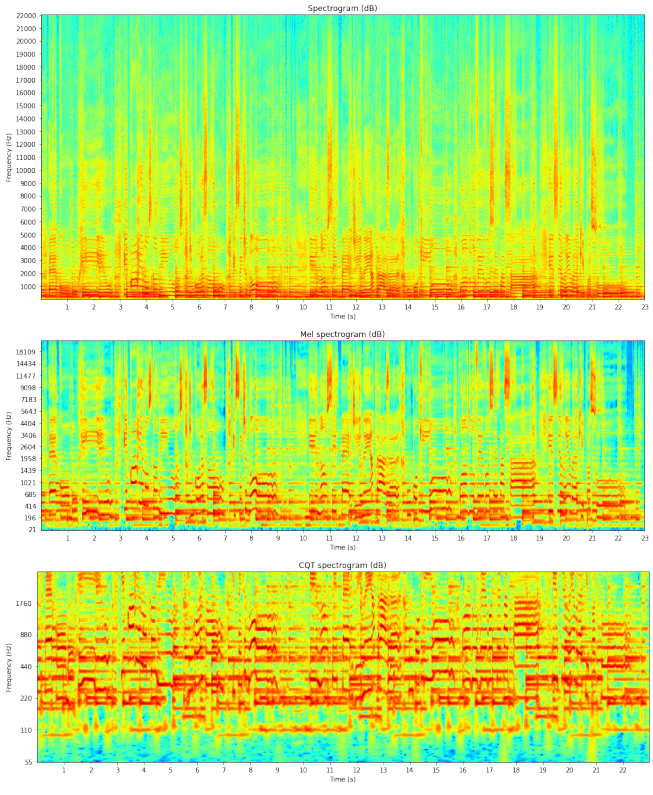

Fig. 22. STFT spectrogram, mel spectrogram, and CQT spectrogram (from top to bottom) computed using the Zaf-Python module.

4.2. Zaf-Matlab

This GitHub repository contains Zafar's Audio Functions in **Matlab** for audio signal analysis: STFT, inverse STFT, mel filterbank, mel spectrogram, MFCC, CQT kernel, CQT spectrogram, CQT chromagram, DCT, DST, MDCT, inverse MDCT. See also MathWorks' File Exchange.

4.3. Zaf-Julia

This GitHub repository contains Zafar's Audio Functions in **Julia** for audio signal analysis: STFT, inverse STFT, mel filterbank, mel spectrogram, MFCC, CQT kernel, CQT spectrogram, CQT chromagram, DCT, DST, MDCT, inverse MDCT.

4.4. REPET-Python

This GitHub repository contains REPET in **Python** for audio source separation: original REPET, REPET extended, adaptive REPET, REPET-SIM, and online REPET-SIM.

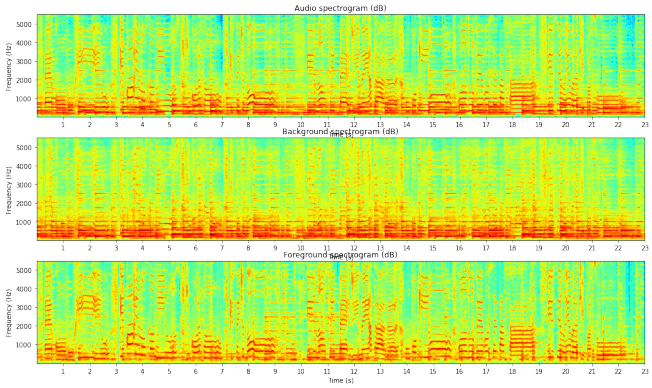

Fig. 23. Separation of a musical excerpt into its accompaniment background and vocal foreground estimates (from top to bottom) using the original REPET from the REPET-Python module.

4.5. REPET-Matlab

This GitHub repository contains REPET in **Matlab** for audio source separation: original REPET, REPET extended, adaptive REPET, REPET-SIM, and online REPET-SIM. See also MathWorks' File Exchange.

4.6. REPET-GUI-Matlab

This GitHub repository contains Matlab GUIs to demo the original REPET and REPET-SIM. See also MathWorks' File Exchange.

4.7. uREPET-Matlab

This GitHub repository contains a Matlab GUI for uREPET, a simple user interface system for recovering patterns repeating in time and frequency in mixtures of sounds. See also MathWorks' File Exchange.

4.8. Zap-Matlab

This GitHub repository contains a Matlab GUI which implements Zafar's audio player (Zap), featuring some practical functionalities such as a playback line, a select/drag tool, and a synchronized spectrogram. See also MathWorks' File Exchange.

4.9. CQHC-Python

This GitHub repository contains a Python implementation for the CQHCs, including a module with the MFCCs, the CQT spectrogram, the CQT deconvolution, and the CQHCs themselves, and notebooks with examples, tests, and notes.

4.10. Problems-Python

This GitHub repository contains Jupyter notebooks with Python coding problems (and solutions). These can be good exercises for beginners and more experienced users to improve and review their programming skills in Python.

5. REFERENCES

5.1. Patents

[1] Z. Rafii, M. Cremer, and B. Kim, "Methods and Apparatus to Identify Sources of Network Streaming Services," 19/029,710, 2025.

[2] Z. Rafii, "Methods and Apparatus for Harmonic Source Enhancement," 19/021,910, 2025.

[3] M. K. Cremer, R. Coover, Z. Rafii, A. Vartakavi, A. Schmidt, and T. Hodges, "Methods and Apparatus to Control Light Pulsing Effects based on Media Content," 19/005,774, 2025.

[4] Z. Rafii, E. Wold, T. P. McGee, and R. Boulderstone, "Detecting and Removing Media Modifications for Identification Services and Copyright Compliance," 18/482,655, 2025.

[5] Z. Rafii, D. G. Robert Coover, A. Topchy, *et al.*, "Use of Audio Classification as Basis to Control Audio Identification," 18/348,202, 2025.

[6] Z. Rafii and P. Seetharaman, "Audio Identification based on Data Structure," 18/824,127, 2024.

[7] M. Cremer, Z. Rafii, R. Coover, and P. Seetharaman, "Automated Cover Song Identification," 18/790,730, 2024.

[8] A. Berrian, T. Hodges, R. Coover, M. Wilkinson, and Z. Rafii, "Audio Content Recognition Method and System," 18/790,749, 2024.

[9] X. Liu, J. Renner, J. Morris, T. Hodges, R. Coover, and Z. Rafii, "Cover Song Identification Method and System," 18/768,497, 2024.

[10] Z. Rafii, "Methods and Apparatus to Extract a Pitch-independent Timbre Attribute from a Media Signal," 18/743,215, 2024.

[11] Z. Rafii, "Methods and Apparatus to Improve Detection of Audio Signatures," 18/740,270, 2024.

[12] M. K. Cremer, R. Coover, Z. Rafii, A. Vartakavi, A. Schmidt, and T. Hodges, "Methods and Apparatus to Control Lighting Effects," 12,035,431, 2024.

[13] Z. Rafii, M. Cremer, and B. Kim, "Methods and Apparatus to Identify Sources of Network Streaming Services using Windowed Sliding Transforms," 11,430,454, 2022.

[14] Z. Rafii, "Audio Matching based on Harmonogram," 11,366,850, 2022.

[15] Z. Rafii, "Methods and Apparatus to Perform Windowed Sliding Transforms," 10,726,852, 2020.

[16] R. Coover and Z. Rafii, "Methods and Apparatus to Fingerprint an Audio Signal via Normalization," 16/453,654, 2020.

[17] B. Pardo and Z. Rafii, "Audio Separation System and Method," 9,093,056, 2015.

5.2. Journal Articles

[18] Z. Rafii, E. Wold, and R. Boulderstone., "How to Design a Cheap Music Detection System Using a Simple Multilayer Perceptron With Temporal Integration," *IEEE Signal Processing Magazine*, vol. 41, no. 4, Nov. 2024.

[19] Z. Rafii, "The Constant-Q Harmonic Coefficients: A timbre feature designed for music signals," *IEEE Signal Processing Magazine*, vol. 39, no. 3, May 2022.

[20] Z. Rafii, "Sliding Discrete Fourier Transform with Kernel Windowing," *IEEE Signal Processing Magazine*, vol. 35, no. 6, Nov. 2018.

[21] Z. Rafii, A. Liutkus, F.-R. Stöter, D. F. Stylianos Ioannis Mimilakis, and B. Pardo, "An Overview of Lead and Accompaniment Separation in Music," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 26, Aug. 2018.

[22] Z. Rafii, Z. Duan, and B. Pardo, "Combining Rhythm-based and Pitch-based Methods for Background and Melody Separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, Dec. 2014.

[23] A. Liutkus, D. FitzGerald, Z. Rafii, B. Pardo, and L. Daudet, "Kernel Additive Models for Source Separation," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, Aug. 2014.

[24] Z. Rafii and B. Pardo, "REpeating Pattern Extraction Technique (REPET): A Simple Method for Music/Voice Separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 1, Jan. 2013.

[25] A. T. Sabin, Z. Rafii, and B. Pardo, "Weighting-Function-Based Rapid Mapping of Descriptors to Audio Processing Parameters," *Journal of the Audio Engineering Society*, vol. 59, no. 6, Jun. 2011.

5.3. Conference Proceedings

[26] G. Peeters, Z. Rafii, M. Fuentes, *et al.*, "Twenty-Five Years of MIR Research: Achievements, Practices, Evaluations, and Future Challenges," in *50th IEEE International Conference on Acoustics, Speech and Signal Processing*, Hyderabad, India, Apr. 2025.

[27] A. Vartakavi, A. Garg, and Z. Rafii, "Audio Summarization for Podcasts," in *29th European Signal Processing Conference*, Dublin, Ireland, Aug. 2021, (poster).

[28] B. Kim and Z. Rafii, "Lossy Audio Compression Identification," in *26th European Signal Processing Conference*, Rome, Italy, Sep. 2018, (poster).

[29] P. Seetharaman and Z. Rafii, "Cover Song Identification with 2d Fourier Transform Sequences," in *42nd IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, LA, USA, Mar. 2017, (poster).

[30] D. FitzGerald, Z. Rafii, and A. Liutkus, "User Assisted Separation of Repeating Patterns in Time and Frequency using Magnitude Projections," in *42nd IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, LA, USA, Mar. 2017, (poster).

[31] A. Liutkus, F.-R. Stöter, Z. Rafii, *et al.*, "The 2016 Signal Separation Evaluation Campaign," in *13th International Conference on Latent Variable Analysis and Signal Separation*, Grenoble, France, Feb. 2017.

[32] N. Ono, Z. Rafii, D. Kitamura, N. Ito, and A. Liutkus, "The 2015 Signal Separation Evaluation Campaign," in *12th International Conference on Latent Variable Analysis and Signal Separation*, Liberec, Czech Republic, Aug. 2015.

[33] Z. Rafii, A. Liutkus, and B. Pardo, "A Simple User Interface System for Recovering Patterns Repeating in Time and Frequency in Mixtures of Sounds," in *40th IEEE International Conference on Acoustics, Speech and Signal Processing*, Brisbane, QLD, Australia, Apr. 2015, (poster).

[34] A. Liutkus, D. FitzGerald, and Z. Rafii, "Scalable Audio Separation with Light Kernel Additive Modelling," in *40th IEEE International Conference on Acoustics, Speech and Signal Processing*, Brisbane, QLD, Australia, Apr. 2015, (slides).

[35] D. FitzGerald, A. Liutkus, Z. Rafii, B. Pardo, and L. Daudet, "Harmonic/Percussive Separation using Kernel Additive Modelling," in *25th IET Irish Signals and Systems Conference*, Limerick, Ireland, Jun. 2014.

[36] A. Liutkus, Z. Rafii, B. Pardo, D. FitzGerald, and L. Daudet, "Kernel Spectrogram Models for Source Separation," in *4th Joint Workshop on Hands-free Speech Communication Microphone Arrays*, Nancy, France, May 2014, (slides).

[37] Z. Rafii, B. Coover, and J. Han, "An Audio Fingerprinting System for Live Version Identification using Image Processing Techniques," in *39th IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, May 2014, (poster).

[38] Z. Rafii, F. G. Germain, D. L. Sun, and G. J. Mysore, "Combining Modeling of Singing Voice and Background Music for Automatic Separation of Musical Mixtures," in *14th International Society for Music Information Retrieval*, Curutiba, PR, Brazil, Nov. 2013, (poster).

[39] Z. Rafii and B. Pardo, "Online REPET-SIM for Real-time Speech Enhancement," in *38th IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, May 2013, (poster).

[40] Z. Rafii and B. Pardo, "Music/Voice Separation using the Similarity Matrix," in *13th International Society for Music Information Retrieval*, Porto, Portugal, Oct. 2012, (slides).

[41] A. Liutkus, Z. Rafii, R. Badeau, B. Pardo, and G. Richard, "Adaptive Filtering for Music/Voice Separation Exploiting the Repeating Musical Structure," in *37th IEEE International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, Mar. 2012, (slides).

[42] M. Cartwright, Z. Rafii, J. Han, and B. Pardo, "Making Searchable Melodies: Human vs. Machine," in *3rd Human Computation Workshop*, San Francisco, CA, Aug. 2011, (poster).

[43] Z. Rafii and B. Pardo, "A Simple Music/Voice Separation Method based on the Extraction of the Repeating Musical Structure," in *36th IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011, (poster).

[44] Z. Rafii and B. Pardo, "Degenerate Unmixing Estimation Technique using the Constant Q Transform," in *36th IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011, (poster).

[45] Z. Rafii and B. Pardo, "Learning to control a Reverberator using Subjective Perceptual Descriptors," in *10th International Society for Music Information Retrieval*, Kobe, Japan, Oct. 2009, (poster).

5.4. Book Chapters

[46] B. Pardo, Z. Rafii, and Z. Duan, "Audio Source Separation in a Musical Context," in *Handbook of Systematic Musicology*, H. Springer Berlin, Ed. 2018.

[47] A. L. Zafar Rafii and B. Pardo, "REPET for Background/Foreground Separation in Audio," in *Blind Source Separation*, H. Springer Berlin, Ed. 2014.

5.5. Technical Reports

[48] Z. Rafii and B. Pardo, "A Digital Reverberator controlled through Measures of the Reverberation," Northwestern University, 2009.

5.6. Tutorials

[49] J. McDermott, B. Pardo, and Z. Rafii, *Leveraging Repetition to Parse the Auditory Scene*, 13th International Society for Music Information Retrieval, Porto, Portugal, Oct. 2012.

5.7. Workshops

[50] Z. Rafii, *An Audio Fingerprinting System for Live Version Identification using Image Processing Techniques*, Midwest Music Information Retrieval Gathering, Northwestern University, Evanston, IL, USA, Jun. 2014.

[51] Z. Rafii, *REPET*, Midwest Music Information Retrieval Gathering, Northwestern University, Evanston, IL, USA, Jun. 2011.

[52] Z. Rafii, R. Blouet, and A. Liutkus, *Discriminant Approach within Non-negative Matrix Factorization for Musical Components Recognition*, DMRN+2: Digital Music Research Network One-day Workshop, Queen Mary University of London, London, UK, Dec. 2007.

5.8. Talks

[53] Z. Rafii, *Sliding DFT with Kernel Windowing*, TokyoBISH Bash, Tokyo, Japan, Mar. 2021.

[54] Z. Rafii, *Identifying Video Sources by Identifying Audio Compression*, Télécom ParisTech, Paris, France, Apr. 2018.

[55] Z. Rafii, *Source separation by repetition*, Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, USA, Jul. 2015.

[56] Z. Rafii, *Source separation by repetition*, Center for New Music and Audio Technologies, University of California, Berkeley, Berkeley, CA, USA, May 2015.

[57] Z. Rafii, *A simple music/voice separation method based on the extraction of the underlying repeating structure*, Télécom ParisTech, Paris, France, Jul. 2011.

5.9. Lectures

[58] Z. Rafii, *Audio Fingerprinting*, EECS 352: Machine Perception of Music & Audio, Northwestern University, Evanston, IL, USA, Dec. 2014.

[59] Z. Rafii, *REPET*, EECS 352: Machine Perception of Music & Audio, Northwestern University, Evanston, IL, USA, Dec. 2014.

[60] Z. Rafii, *Rhythm Analysis in Music*, EECS 352: Machine Perception of Music & Audio, Northwestern University, Evanston, IL, USA, Dec. 2014.

[61] Z. Rafii, *Time-frequency Masking*, EECS 352: Machine Perception of Music & Audio, Northwestern University, Evanston, IL, USA, Dec. 2014.

5.10. Data Sets

[62] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, and R. Bittner. "MUSDB18-HQ – an uncompressed version of MUSDB18." (2019).

[63] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, and R. Bittner. "The MUSDB18 corpus for music separation." (2019).