REPET: REpeating Pattern Extraction Technique

A Simple Music/Voice Separation Method based on the Extraction of the Underlying Repeating Structure

Zafar RAFII
Plan

I. Introduction

II. REPET Algorithm
   1. Identify a repeating period
   2. Model a repeating segment
   3. Extract the repeating structure

III. Music/Voice Separation
   1. Experimental results
   2. Audio examples
   3. Future work

IV. Conclusion
I. Introduction

• **Repetition** is a core principle in music: a musical piece has generally a distinguishable underlying repeating structure.
I. Introduction

- Assuming sparsity in time and frequency, there should be (roughly) periodically repeating & non-repeating t-f bins
I. Introduction

- Assuming sparsity in time and frequency, there should be (roughly) periodically repeating & non-repeating t-f bins.
I. Introduction

• **Idea:** identify time-frequency bins periodically repeating in the spectrogram and extract them via binary masking

• **Method:** REpeating Patterns Extraction Technique (REPET)
  1. Identify a repeating period
  2. Model a repeating segment
  3. Extract the repeating structure

• **Result:** a simple music/voice separation system!
  – The underlying repeating structure $\approx$ *musical "background"
  – The overlying non-repeating structure $\approx$ *vocal "foreground"
1. Introduction

- Parallel with *background subtraction* in computer vision:
  1. Compare frames to estimate a "background model"
I. Introduction

- Parallel with **background subtraction** in computer vision:
  2. Extract the "background" from the "foreground"
I. Introduction

- Parallel with **background subtraction** in computer vision:
  
  → With audio, we also need to identify a repeating period!
I. Introduction

- **Practical Interest:**
  - Instruments/vocalist identification
  - Music/voice transcription
  - Post production
  - Karaoke

- **Practical Advantages:**
  - Not feature-dependent
  - No complex framework
  - No prior training

- **Intellectual Interest:**
  - Simply based on repetition
  - Musical structure extraction
I. Introduction

II. REPET Algorithm
   1. Identify a repeating period
   2. Model a repeating segment
   3. Extract the repeating structure

III. Music/Voice Separation
   1. Experimental results
   2. Audio examples
   3. Future work

IV. Conclusion
II. REPET – 1. Repeating Period

- A correlogram is calculated from the **autocorrelation** of the rows of the power spectrogram to detect periodicities.
II. REPET – 1. Repeating Period

• By taking the mean of the rows of the correlogram, we obtain the "beat spectrum"
II. REPET – 1. Repeating Period

- The beat spectrum reveals the repeating period $P$ of the underlying repeating musical structure.
II. REPET – 1. Repeating Period

- The beat spectrum reveals the repeating period $P$ of the underlying repeating musical structure.
Plan

I. Introduction

II. REPET Algorithm
   1. Repeating period
   2. Repeating segment
   3. Repeating structure

III. Music/Voice Separation
   1. Evaluation
   2. Examples
   3. Future

IV. Conclusion
II. REPET – 2. Repeating Segment

- The **repeating period** is used to segment the magnitude spectrogram at period rate
II. REPET – 2. Repeating Segment

The **repeating segment** is modeled as the median of the segments for every t-f bins.
II. REPET – 2. Repeating Segment

- As the middle value of a distribution, the **median** helps to model a smooth repeating segment, eliminating outliers.
II. REPET – 2. Repeating Segment

- As the middle value of a distribution, the **median** helps to model a smooth repeating segment, eliminating outliers.
Plan

I. Introduction

II. REPET Algorithm
   1. Identify a repeating period
   2. Model a repeating segment
   3. Extract the repeating structure

III. Music/Voice Separation
   1. Experimental results
   2. Audio examples
   3. Future work

IV. Conclusion
II. REPET – 3. Repeating Structure

- The **repeating segment** is used to divide bin-wise each segment in order to obtain a median-scaled spectrogram.
II. REPET – 3. Repeating Structure

- The **repeating segment** is used to divide bin-wise each segment in order to obtain a median-scaled spectrogram.
II. REPET – 3. Repeating Structure

- The **repeating segment** is used to divide bin-wise each segment in order to obtain a median-scaled spectrogram.
II. REPET – 3. Repeating Structure

- Repeating t-f bins are similar to the repeating segment, so have values around 1 in the median-scaled spectrogram.
II. REPET – 3. Repeating Structure

- Repeating t-f bins are similar to the repeating segment, so have values around 1 in the median-scaled spectrogram.

![Diagram showing median and division steps in spectrograms](image_url)
II. REPET – 3. Repeating Structure

- By assigning bins around 1 to the repeating structure (white) and the rest to 0 (black), we get a binary t-f mask.
II. REPET – 3. Repeating Structure

- By assigning bins around 1 to the repeating structure \((white)\) and the rest to 0 \((black)\), we get a \textbf{binary t-f mask}
II. REPET – 3. Repeating Structure

- By assigning bins around 1 to the repeating structure (white) and the rest to 0 (black), we get a **binary t-f mask**
II. REPET – 3. Repeating Structure

- In practice, bins can overlap and repetitions can involve variations, therefore we introduced a **tolerance factor** $T$.
II. REPET – 3. Repeating Structure

- In practice, bins can overlap and repetitions can involve variations, therefore we introduced a **tolerance factor** $T$.
II. REPET – 3. Repeating Structure

• In practice, bins can overlap and repetitions can involve variations, therefore we introduced a **tolerance factor** $T$.
II. REPET – 3. Repeating Structure

- The **binary t-f mask** is used to extract the spectrogram of the underlying repeating structure.
II. REPET – 3. Repeating Structure

- The **binary t-f mask** is used to extract the spectrogram of the underlying repeating structure.
II. REPET – 3. Repeating Structure

- Finally, the audio signals can be reconstructed from their respective spectrograms.
II. REPET – 3. Repeating Structure

- Finally, the audio signals can be reconstructed from their respective spectrograms.

REPET algorithm

1. Repeating period
2. Repeating segment
3. Repeating structure
II. REPET – 3. Repeating Structure

- Overlying non-repeating structure \( \approx \text{vocal foreground} \)
- Underlying repeating structure \( \approx \text{musical background} \)

REPET algorithm

1. Repeating period
2. Repeating segment
3. Repeating structure
Plan

I. Introduction

II. REPET Algorithm
   1. Identify a repeating period
   2. Model a repeating segment
   3. Extract the repeating structure

III. Music/Voice Separation
   1. Experimental results
   2. Audio examples
   3. Future work

IV. Conclusion
III. M/V separation – 1. Experiment

There are only few papers dealing with m/v separation; the different approaches can be summarized as follows:

1. Non-negative Matrix Factorization (NMF)
   - Need to know the number of components
   - Need a proper initialization

2. Train spectra for accompaniment from non-vocal segments
   - Need vocal/non-vocal segmentation
   - Need sufficient amount of non-vocal frames

3. Pitch-based inference
   - Cannot extract unvoiced vocal frames
   - Harmonic structure of the instruments can interfere
III. M/V separation – 1. Experiment

- REPET vs. Hsu et al. [2010]:
  - Last state-of-the-art m/v separation approach:
    - Vocals separation using pitch-based inference
    - Identification of unvoiced vocal frames
    - Spectral subtraction to eliminate interferences
  - Best automatic version of Hsu et al.:
    - Estimated pitch (not human-labeled)
    - Computer-detected unvoiced vocal frames (not human-labeled)
    - Voiced vocal enhancement (spectral subtraction)
  - Dataset:
    - 1,000 Chinese pop song clips
    - 3 sets of mixtures for 3 different "voice-to-music ratio" (-5, 0, 5 dB)
III. M/V separation – 1. Experiment

- REPET vs. Hsu et al. [2010]:

Global music/voice separation performance between Hsu, REPET (Rafii) and using the ideal binary mask.
III. M/V separation – 1. Experiment

- REPET vs. Hsu et al. [2010]:

Music/voice separation performance at voice-to-music ratio of 0 dB using REPET and successive enhancements.

Rafii & Pardo, ICASSP 2011
Plan

I. Introduction

II. REPET Algorithm
   1. Identify a repeating period
   2. Model a repeating segment
   3. Extract the repeating structure

III. Music/Voice Separation
   1. Experimental results
   2. Audio examples
   3. Future work

IV. Conclusion
III. Music/Voice – 2. Examples

- REPET vs. Ozerov et al. [2008]:
  - Adapt accompaniment model from non-vocal frames

1. The Prodigy – Breathe
III. Music/Voice – 2. Examples

• REPET vs. Ozerov et al. [2008]:
  - Adapt accompaniment model from non-vocal frames

2. *The Doors* – *People are strange*
III. Music/Voice – 2. Examples

• More audio examples:

3. **RJD2 – Ghostwriter (no vocals)**

4. **Rebecca Black – Friday (because tomorrow is Saturday...)**
Plan

I. Introduction

II. REPET Algorithm
   1. Identify a repeating period
   2. Model a repeating segment
   3. Extract the repeating structure

III. Music/Voice Separation
   1. Experimental evaluation
   2. Audio examples
   3. Future work

IV. Conclusion
III. Music/Voice – 3. Future

• Extraction of a hierarchical repeating structure:
  – Extract multiple repeating layers at different period rates
III. Music/Voice – 3. Future

• Use of a smart segmentation:
  – Model different repeating segments from/for different regions

![Diagram showing different model segments for verse 1, verse 2, chorus, non-repeating foreground, and repeating background over time.](image-url)
III. Music/Voice – 3. Future

- Use of a similarity matrix:
  - Identify and extract individual repeating elements
Plan

I. Introduction

II. REPET Algorithm
   1. Identify a repeating period
   2. Model a repeating segment
   3. Extract the repeating structure

III. Music/Voice Separation
   1. Experimental results
   2. Audio examples
   3. Future work

IV. Conclusion
IV. Conclusion

• **REPET**: REpeating Pattern Extraction Technique

• **Music/voice separation** by extraction of the underlying repeating musical structure

• **Strengths:**
  – Simple
  – Fast
  – Blind
  – Automatable
  – Promising...
Questions?