REpeating Pattern Extraction Technique (REPET)

EECS 352: Machine Perception of Music & Audio
Repetition is a fundamental element in generating and perceiving structure.
Observation

• Musical works are often characterized by an underlying repeating structure over which varying elements are superimposed.
Observation

• Musical works are often characterized by an underlying repeating structure over which varying elements are superimposed.
Assumption

- There should be patterns that are more or less repeating in time and frequency

Mixture Spectrogram
Assumption

• The repeating patterns could be identified and extracted using a **time-frequency mask**
Idea

• REpeating Pattern Extraction Technique!
  1. Identify the repeating elements
  2. Derive a repeating model
  3. Extract the repeating structure
Idea

• Simple **music/voice separation** method!
  - Repeating structure = background music
  - Non-repeating structure = foreground voice
Practical Advantages

- Does not depend on special parametrizations
- Does not rely on complex frameworks
- Does not require external information
Practical Interests

• Karaoke gaming (need the music)
• Query-by-hhumming (need the voice)
• Audio remixing (need both components)
Intellectual Interests

• Music understanding
• Music perception
• Simply based on repetition!
Parallels

• **Background subtraction** in computer vision
Parallels

• **Background subtraction** in computer vision
Parallels

- **Background subtraction** in computer vision
  - In audio, we also need to identify the repetitions!
Parallels

- **Background subtraction** in computer vision
  - In audio, we also need to identify the repetitions!

![Waveform graphs](image_url)
Parallels

- **Auditory segregation** in human listeners

  - Unknown audio mixtures with the same target and different distractors
  - Target identified as the repeating object
Parallels

• **Auditory segregation** in human listeners

As the number of mixtures increases, the target becomes more apparent... [courtesy of Josh McDermott]
REPET

Mixture Signal $x$

Step 1

Step 2

V

1p

2p

Median

Step 3

S

Repeating Spectrogram $W$

Time-Frequency Mask $M$

Zafar Rafii, Winter 2014
1. Repeating Period

- We compute the **autocorrelations** of the frequency rows of the mixture spectrogram.
1. Repeating Period

- We take the mean of the autocorrelation rows and obtain the **beat spectrum**
1. Repeating Period

- The beat spectrum reveals the **repeating period** $p$ of the underlying repeating structure.
2. Repeating Segment

- We then use the repeating period to segment the mixture spectrogram at period rate.
2. Repeating Segment

- We derive a **repeating segment model** by taking the element-wise median of segments.
2. Repeating Segment

- The median helps to derive a clean repeating segment, removing the non-repeating outliers

Zafar Rafii, Winter 2014
REPET

Step 1
Mixture Signal $x$

Step 2
Mixture Spectrogram $V$

Step 3
Beat Spectrum $b$

Repeating Segment $S$

Repeating Spectrogram $W$

Time-Frequency Mask $M$

Zafar Rafii, Winter 2014
3. Repeating Structure

- We take the element-wise \textbf{min} between the repeating segment model and the segments
3. Repeating Structure

- We obtain a **repeating spectrogram model** for the repeating background
3. Repeating Structure

• The repeating spectrogram should not have values higher than the mixture spectrogram
3. Repeating Structure

- We then **divide**, element-wise, the repeating spectrogram by the mixture spectrogram.
3. Repeating Structure

- We obtain a **soft time-frequency mask** (with values between 0 and 1)
3. Repeating Structure

• In the soft t-f mask, the more/less a t-f bin is repeating, the more it is weighted toward 1/0
3. Repeating Structure

- We could further derive a **binary t-f mask** by fixing a threshold between 0 and 1
3. Repeating Structure

- We multiply, element-wise, the t-f mask with the mixture STFT to get the background STFT.
3. Repeating Structure

- We obtain the **repeating background** signal by inverting its STFT into the time domain.
3. Repeating Structure

- We obtain the **non-repeating foreground signal** by subtracting background from mixture.
Summary

- Repeating background ≈ *music component*
- Non-repeating foreground ≈ *voice component*

**REPET**

1. Repeating period
2. Repeating segment
3. Repeating structure
Music/Voice Separation

- A variety of techniques has been proposed to separate **music** and **voice** from a mixture
  - Accompaniment modeling, Pitch-based inference, Non-negative Matrix Factorization (NMF), etc.
Music/Voice Separation

• Accompaniment modeling
  - Modeling of the musical accompaniment from the non-vocal segments in the mixture

→ Need an accurate vocal/non-vocal segmentation!
→ Need a sufficient amount of non-vocal segments!
Music/Voice Separation

• **Pitch-based inference**
  - Separation of the vocals using the predominant pitch contour extracted from the vocal segments

→ Need an accurate predominant pitch detection!
→ Cannot extract unvoiced vocals!
Music/Voice Separation

• **Non-negative Matrix Factorization (NMF)**
  - Iterative factorization of the mixture spectrogram into non-negative additive basic components

→ Need to know the number of components!
→ Need a proper initialization!
Evaluation

• **REPET** [Rafii et al., 2013]
  - Automatic period finder
  - Soft time-frequency masking

• **Competitive method** [Durrieu et al., 2011]
  - Source-filter modeling with NMF framework
  - Unvoiced vocals estimation

• **Data set** [Hsu et al., 2010]
  - 1,000 song clips (from karaoke Chinese pop songs)
  - 3 voice-to-music mixing ratios (-5, 0, and 5 dB)
Evaluation

SDR (dB)

Music

Voice

D = Durrieu
D+H = Durrieu + High-pass
R = REPET
R+H = REPET + High-pass

Zafar Rafii, Winter 2014
Evaluation

• Conclusions
  - REPET can compete with state-of-the-art (and more complex) music/voice separation methods
  - There is room for improvement (+ high-pass, + optimal period, + vocal frames)
  - Average computation time: 0.016 second for 1 second of mixture! (vs. 3.863 seconds for Durrieu)
Examples

• REPET vs. Durrieu (source-filter + NMF)
Examples

• REPET vs. Ozerov (accompaniment modeling)

Zafar Rafii, Winter 2014
Examples

• REPET vs. Virtanen (NMF + pitch-based)
Examples

- REPET (more examples...)

RJD2 - Ghostwriter

Rebecca Black - Friday

Background estimate

Foreground estimate

Zafar Rafii, Winter 2014
Demo
Thank you!
References

Extensions

- REPET works well on excerpts with a relatively stable repeating background (e.g., 10 s verse)
Extensions

• For full-track songs, the repeating background is likely to vary over time (e.g., verse/chorus)
Prior Segmentation

• We could do a **prior segmentation** of the song and apply REPET to the individual sections

![Diagram showing a prior segmentation of a song into verses and chorus sections, with REPET applied to each section and a full repeating background.]
Sliding Window

- We could apply REPET to local sections of the song over time via a fixed *sliding window*
Adaptive REPET

• We could directly adapt REPET along time by locally modeling the repeating background.
Adaptive REPET

Step 1

Mixture Signal $x$

Mixture Spectrogram $V$

Beat Spectrogram $B$

Step 2

$V$

Median

Repeating Spectrogram $W$

Step 3

$U$

Repeating Spectrogram $W$

Time-Frequency Mask $M$

Zafar Rafii, Winter 2014
Generalization

• REPET (and its extension) assumes periodically repeating patterns
Generalization

• Repetitions can also happen *intermittently* or without a global (or local) period
Generalization

• Instead of looking for periodicities, we can look for similarities, using a similarity matrix
Generalization

- The **similarity matrix** is a matrix where each bin measures the (dis)similarity between any two elements of a sequence given a metric.
REPET-SIM

Step 1

Mixture Signal $x$

Step 2

Mixture Spectrogram $V$

Step 3

Median

Repeating Spectrogram $U$

Similarity Matrix $S$

Repeating Spectrogram $W$

Time-Frequency Mask $M$

Zafar Rafii, Winter 2014
Adaptive REPET

Step 1
Mixture Signal $x$

Step 2
Mixture Spectrogram $V$
Beat Spectrogram $B$

Step 3
Repeating Spectrogram $U$

Median

Time-Frequency Mask $M$

Zafar Rafii, Winter 2014
REPET + Pitch

• REPET models the **background rhythm**

Mixture spectrogram → “Period” → Rhythmic mask → Background rhythm

- **repetitions (in time)**

• Pitch-based methods model the **lead melody**

Mixture spectrogram → “Pitch” → Melodic mask → Lead melody

- **harmonics (in frequency)**
REPET + Pitch

• Auditory processing in human listeners