Audio Source Separation With a Single Sensor

Performed by: Kfir Gedalyahu, Michal Levy
Supervisor: Guy Rapaport
Introduction

- The basic idea behind single sensor source separation is trying to extract sources from their mixture.
- In this project, we will focus on a codebook based method for source separation.
- The main assumption of this method is that each source can be represented by a dictionary.

This assumption simplifies the separation process.
Solution

The solution relies on building a statistical model of the audio sources:

1. Source 1 samples
2. Source 2 samples

- Code Book Creation
- Model Parameters
- Separation Algorithm
  - Filters Creation
  - Source Separation

- Mixture
- Source 1
- Source 2
Gaussian Mixture Models (GMM)

Gaussian mixture prior density:

\[
G\left( y, \{ \omega^{(i)} \}, \{ \Sigma^{(i)} \} \right) = \sum_{i=1}^{K} \omega^{(i)} g\left( y, \Sigma^{(i)} \right), \quad \sum_{i=1}^{K} \omega^{(i)} = 1
\]

Observation is obtained by:

1. Selecting one active component. According to priori probabilities \( \{ \omega^{(i)} \} \)
2. Generating Gaussian observation.

This model permits dealing with multiple covariance matrices corresponding to multiple PSD shapes.
Gaussian Scaled Mixture Models (GSMM)

Gaussian scaled mixture prior density:

\[
G\left( y,\{\omega^{(i)}\},\{\Sigma^{(i)}\} \right) = \sum_{i=1}^{K} \omega^{(i)} g(y, a^{(i)} \Sigma^{(i)}), \quad \sum_{i=1}^{K} \omega^{(i)} = 1
\]

- \(a^{(i)}\) is a positive gain factor.
- Separates the PSD shape from the amplitude information.
Source Separation in GSMM case

Estimating the most probable gain factors for each pair of active components.

\[
\left( \hat{a}_i^1, \hat{a}_j^2 \right) = \arg \max_{a_i \geq 0, a_j \geq 0} \gamma_{i,j,a_i^1,a_j^2}(x)
\]

Calculating the probability of each pair of active components, given the observation \( x \).

\[
\gamma_{i,j,a_i^1,a_j^2}(x)
\]

posterior probabilities of components \((i, j)\):

\[
\gamma_{i,j,a_i^1,a_j^2}(x) \propto \omega_1^{(i)} \omega_2^{(j)} g(x, a_i^1 \Sigma_1^{(i)} + a_j^2 \Sigma_2^{(j)} + \sigma^2 I)
\]

Building the filters.

PM

Or

MAP
Separation Algorithm implementation

- Audio Sources are **locally stationary** in general.
- It is natural to work with the **short-time Fourier transform** (STFT).
- STFT is linear so the mixing equation can be expressed as:

\[
S_x(t, f) = Ss_1(t, f) + Ss_2(t, f) + Sb(t, f)
\]

- The covariance matrices \( \Sigma^{(i)}, \Sigma^{(j)} \) assumed to be **diagonal** (in the STFT domain), with running elements \( \sigma^{(i)}(f)^2, \sigma^{(j)}(f)^2 \)
PM Estimator:

\[ \hat{S}_s_1(t, f) = \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \gamma_{i,j}(t) \frac{a_1^{(i)} \sigma_1^{(i)}(f)^2}{a_1^{(i)} \sigma_1^{(i)}(f)^2 + a_2^{(j)} \sigma_2^{(j)}(f)^2 + \sigma^2} Sx(t, f) \]

\[ \hat{S}_s_2(t, f) = \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \gamma_{i,j}(t) \frac{a_2^{(j)} \sigma_2^{(j)}(f)^2}{a_1^{(i)} \sigma_1^{(i)}(f)^2 + a_2^{(j)} \sigma_2^{(j)}(f)^2 + \sigma^2} Sx(t, f) \]

MAP Estimator:

\[ \hat{i}, \hat{j} = \arg \max_{i,j} \gamma_{i,j}(x) \]

\[ \hat{S}_s_1(t, f) = \frac{\sigma_1^{(i)}(f)^2}{\sigma_1^{(i)}(f)^2 + \sigma_2^{(j)}(f)^2 + \sigma^2} Sx(t, f) \]

\[ \hat{S}_s_2(t, f) = \frac{\sigma_2^{(j)}(f)^2}{\sigma_1^{(i)}(f)^2 + \sigma_2^{(j)}(f)^2 + \sigma^2} Sx(t, f) \]
Training step

- Using E-M algorithm, model parameters of each source are estimated separately:
  1. **PSD** of each Gaussian component.
  2. **Priori probability** of each component.

Example:
Frequency components from the cello can be found in the separated guitar spectrogram.
Improvement: Separation in several frequency bands

- Splitting the frequency domain into several frequency bands.
- Performing separation in each band separately.
- Advantages:
  - Better local resemblance (in frequency domain) between the mixture and the codebook representatives.
  - Working with lower dimension Gaussian vectors.
  - Effectively larger codebook.

![Graph showing separation in several frequency bands](image)

**Band 1**

**Band 2**
Conclusions

- We have presented a codebook based algorithms for single source separation.
- The main assumption of this method is that each source can be represented by a dictionary.
- GMM have been used:
  - each source is represented by “typical” PSD and their priori probabilities.
- We have shown that this model is too simplistic for music instruments:
  - There is no “close enough” representative in the codebook.
  - Music instruments PSDs are too diverse for this model.
- There is a need to use a more adequate model for music instrument, that takes into account their properties.