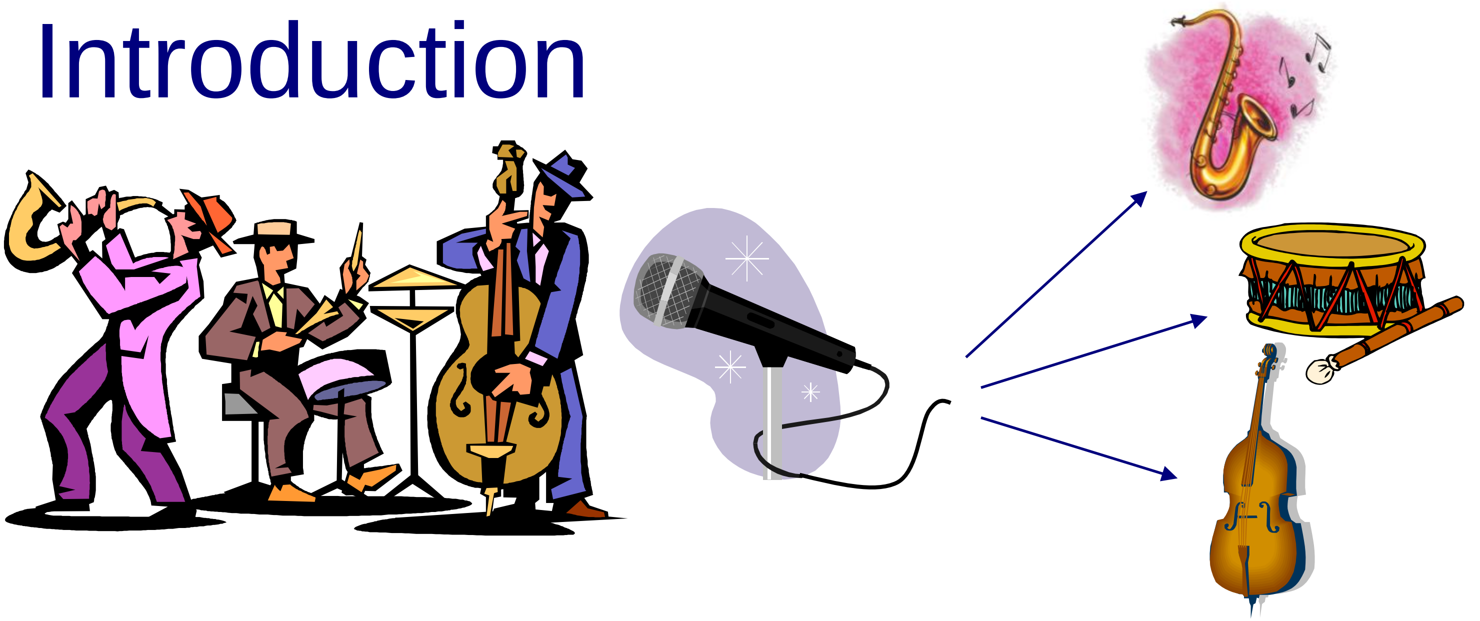
SIPL SIGNAL & IMAGE PROCESSING LAB

Audio Source Separation With a Single Sensor

Performed by: Kfir Gedalyahu, Michal Levy

Supervisor: Guy Rapaport

Introduction

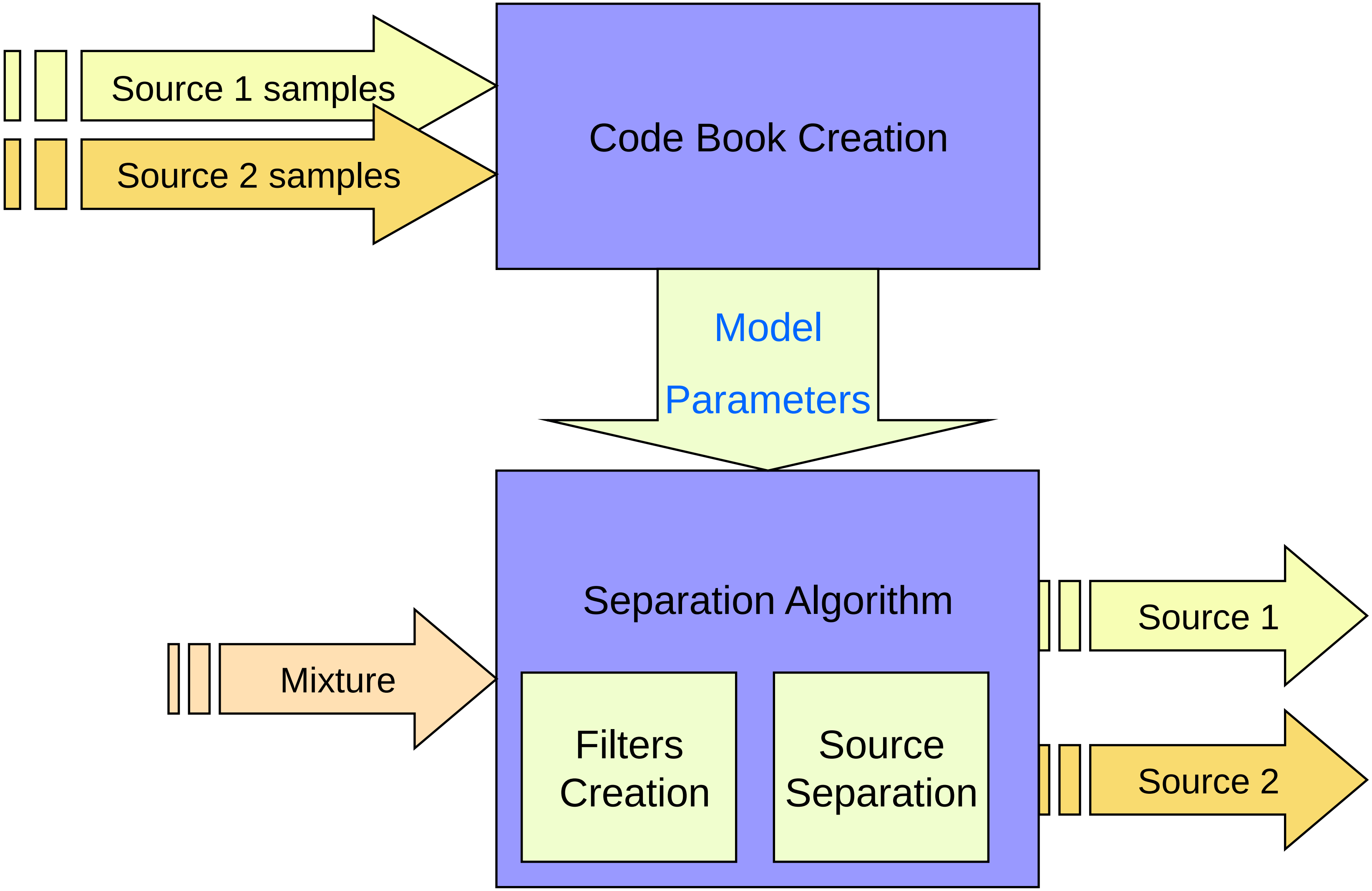


- The basic idea behind single sensor source separation is trying to extract sources from their mixture.
- In this project, we will focus on a codebook based method for source separation.
- The main assumption of this method is that each source can be represented by a dictionary.

  This assumption simplifies the separation process .

Solution

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



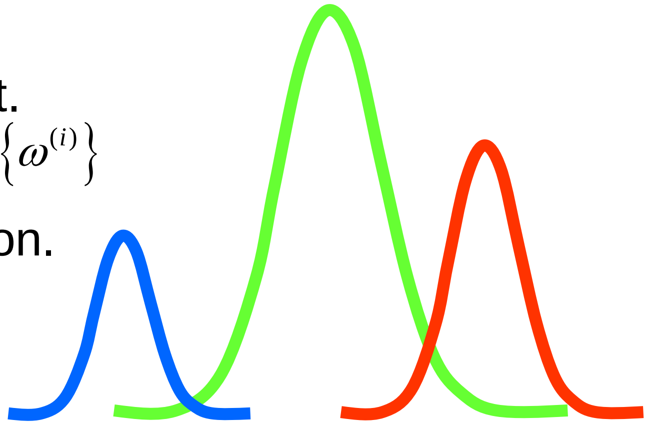
Gaussian Mixture Models (GMM)

Gaussian mixture prior density:

$$G\left(y,\left\{\omega^{(i)}\right\},\left\{\Sigma^{(i)}\right\}\right)=\sum_{i=1}^{K}\omega^{(i)}g\left(y,\Sigma^{(i)}\right),\qquad \sum_{i=1}^{K}\omega^{(i)}=1$$

Observation is obtained by:

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



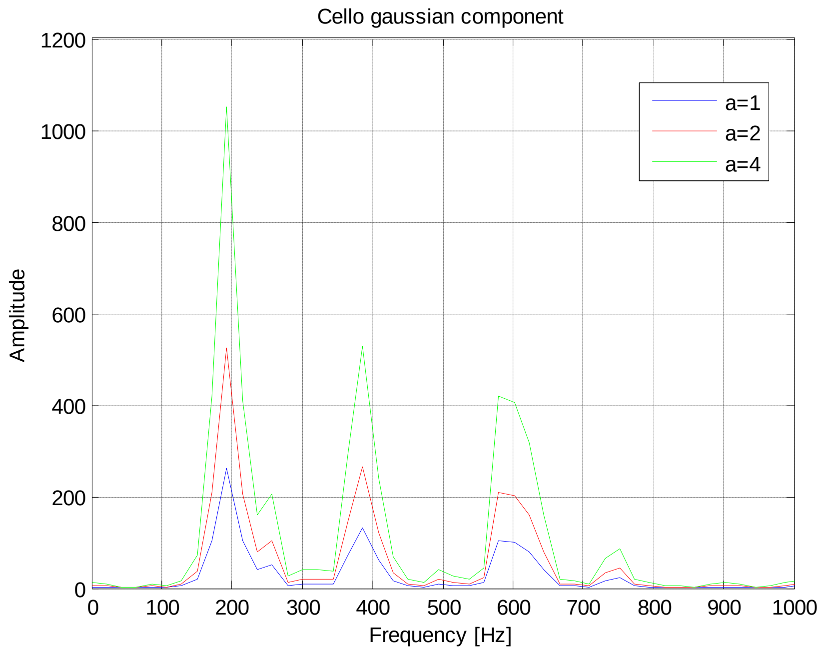
This model permits dealing with multiple covariance matrices corresponding to multiple PSD shapes.

Gaussian Scaled Mixture Models (GSMM)

Gaussian scaled mixture prior density:

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

- $a^{(i)}$ 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.

$$\left(\hat{a}_i^1, \hat{a}_j^2\right) = \underset{a_1 \geq 0, a_2 \geq 0}{\arg\max}\, \gamma_{i,j,a_i^1,a_j^2}(x)$$

posterior probabilities of components (i, j):

$$\gamma_{i,j,a_i^1,a_j^2}(x) \propto \varpi_1^{(i)} \varpi_2^{(j)} g(x, a_i^1 \Sigma_1^{(i)} + a_j^2 \Sigma_2^{(j)} + \sigma^2 I)$$

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

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

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:

$$\boxed{\mathcal{S}x(t,f) = \mathcal{S}s_1(t,f) + \mathcal{S}s_2(t,f) + \mathcal{S}b(t,f)}$$

- The covariance matrices $\Sigma_1^{(i)}, \Sigma_2^{(j)}$ assumed to be diagonal (in the STFT domain), with running elements $\sigma_1^{(i)}(f)^2,\ \sigma_2^{(j)}(f)^2$

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:

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

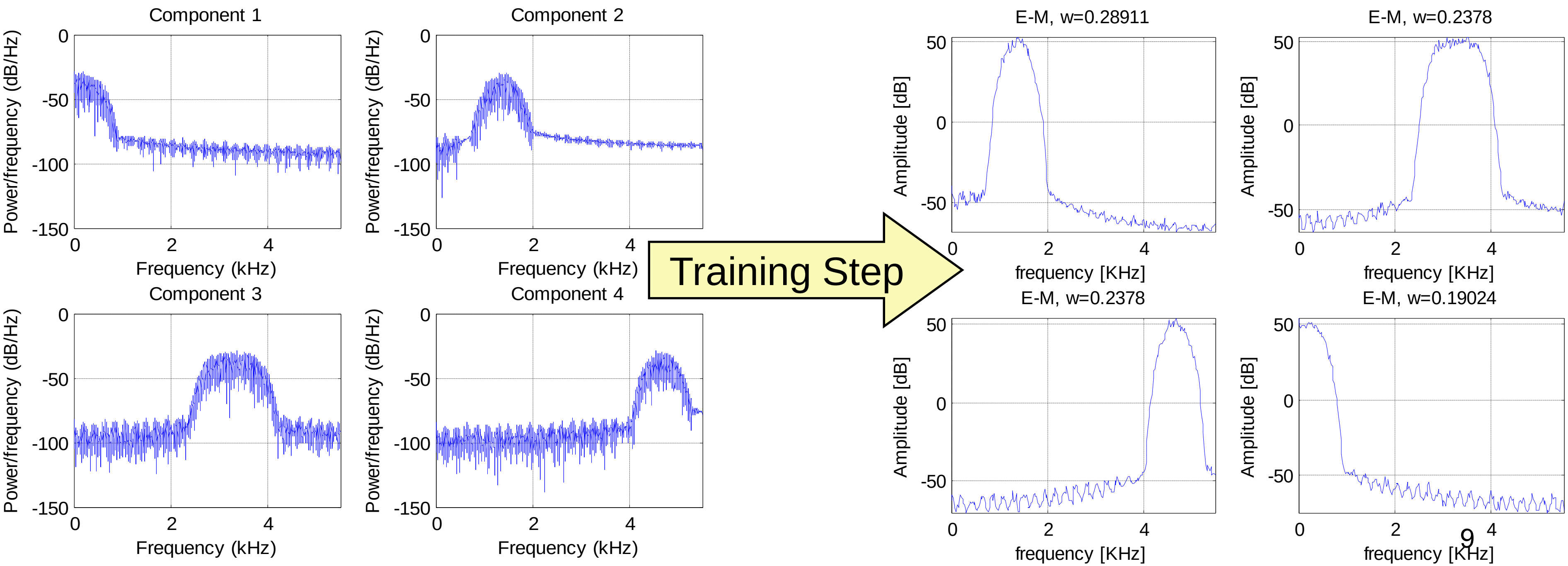
$$\hat{S}s_1(t,f)=\frac{\sigma_{1,}^{(\hat{i})}(f)^2}{\sigma_1^{(\hat{i})}(f)^2+\sigma_2^{(\hat{j})}(f)^2+\sigma^2}Sx(t,f)$$

$$\hat{S}s_2(t,f)=\frac{\sigma_2^{(\hat{j})}(f)^2}{\sigma_1^{\hat{i}}(f)^2+\sigma_2^{\hat{j}}(f)^2+\sigma^2}Sx(t,f)$$

Training step

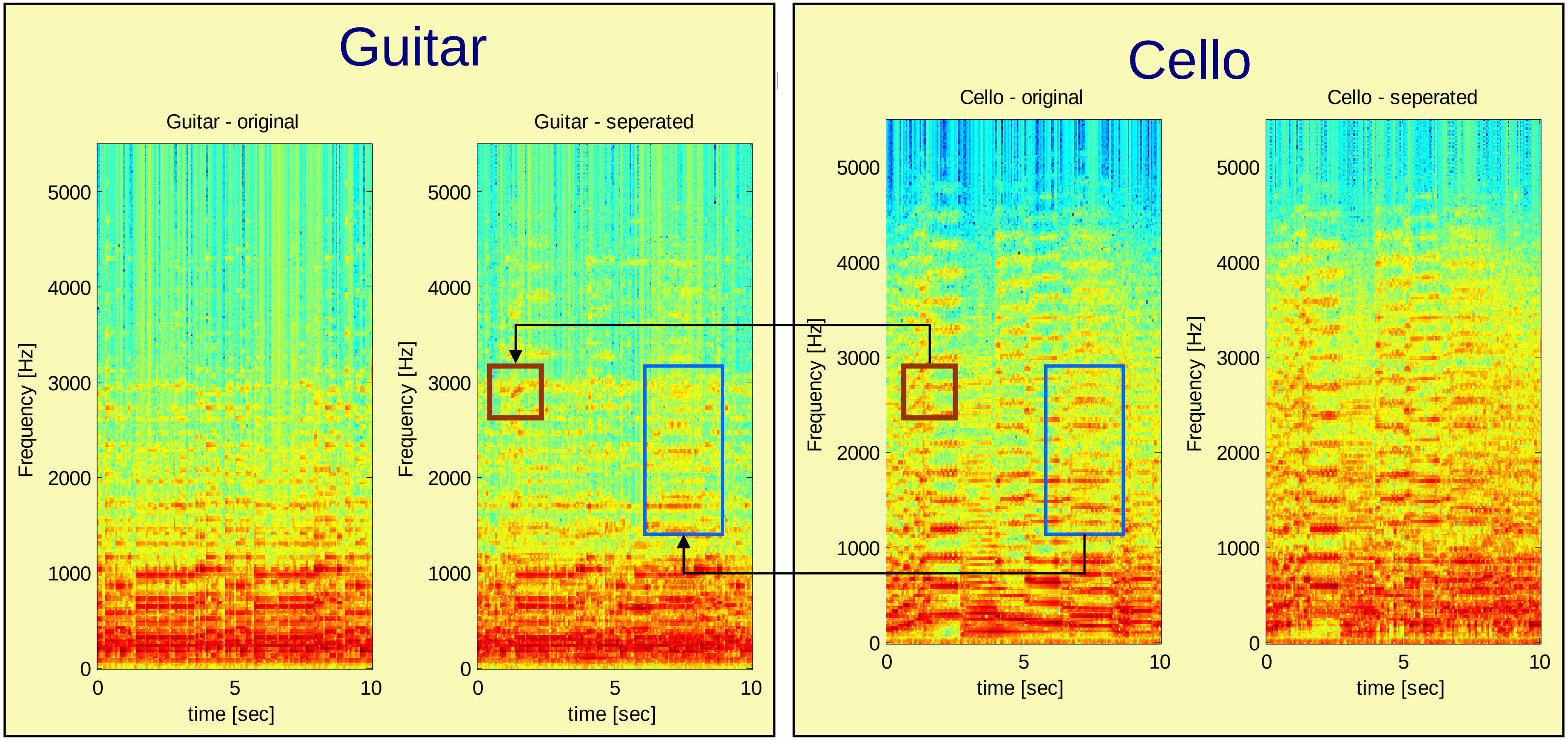
- Using E-M algorithm, model parameters of each source are estimated separately:
  1. PSD of each Gaussian component.
  2. Priori probability of each component.

Example:



Separation example

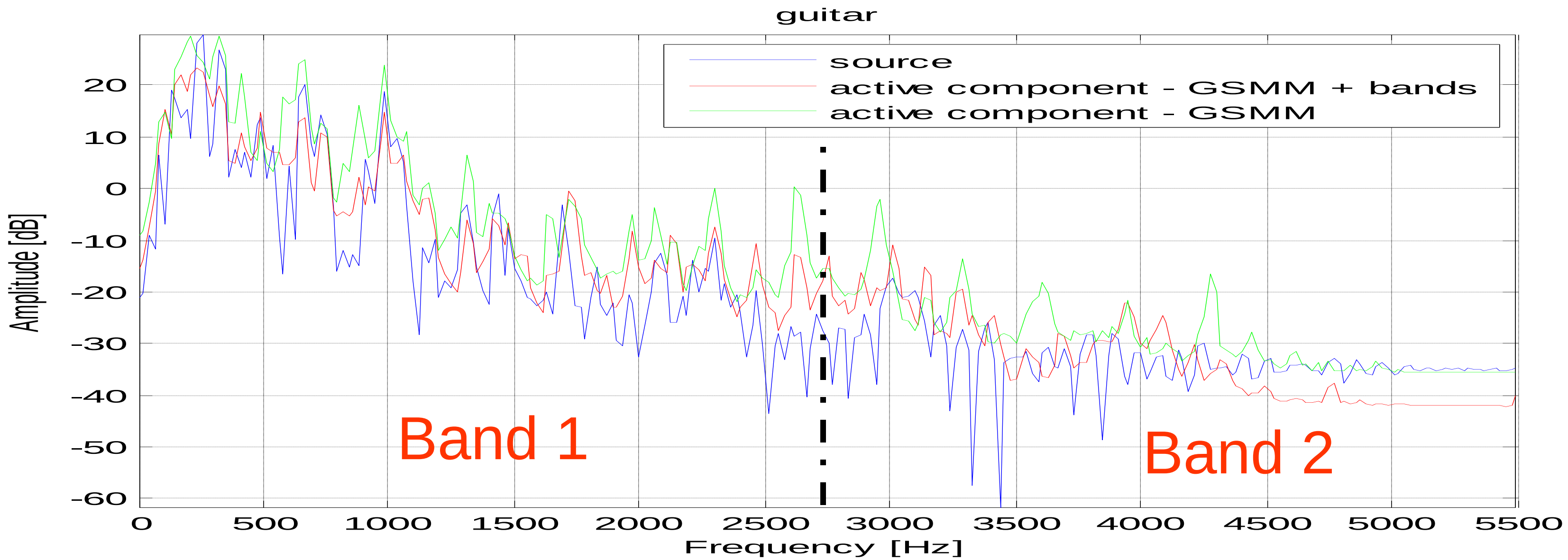
$$x = cello + guitar$$



- 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.



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.