Source-Filter Based Clustering for Monaural Blind Source Separation

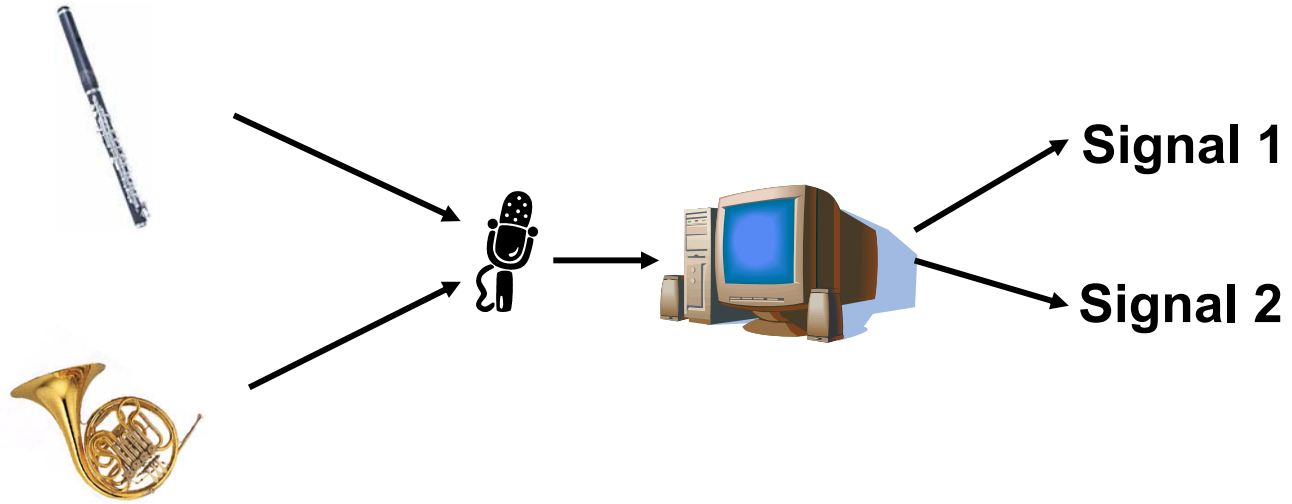
Institut für Nachrichtentechnik
RWTH Aachen University

Outline

- Motivation
- Separation Framework
- Clustering
  - MFCC + k-Means
  - NMF Clustering
- Examples

Motivation

- Scenario

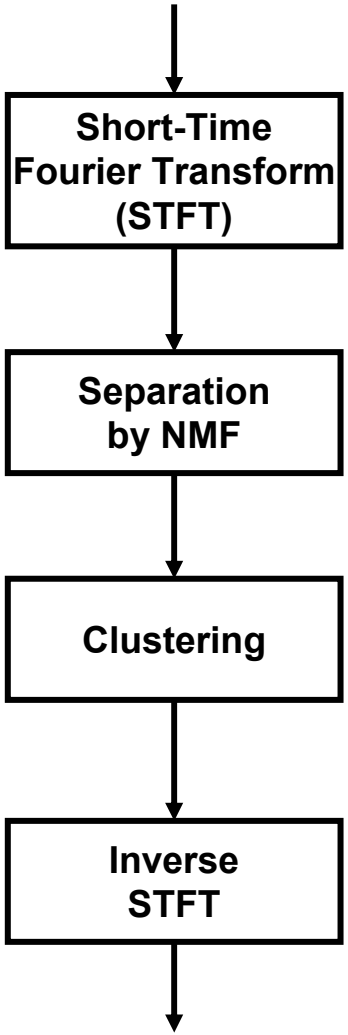


- Applications
  - Remixing
  - Noise Cancellation
  - Karaoke
  - …

Outline

- Motivation
- Separation Framework
- Clustering
  - MFCC + k-Means
  - NMF Clustering
- Examples

Separation Framework



$x(n)$ **Mixture**

$X(k,t)$ **Spectrogram of mixture**

$C_i(k,t)$, $1 \le i \le I$ **Spectrogram of separated channels**

$\tilde{S}_m(k,t)$, $1 \le m \le 2$ **Spectrogram of estimated sources**

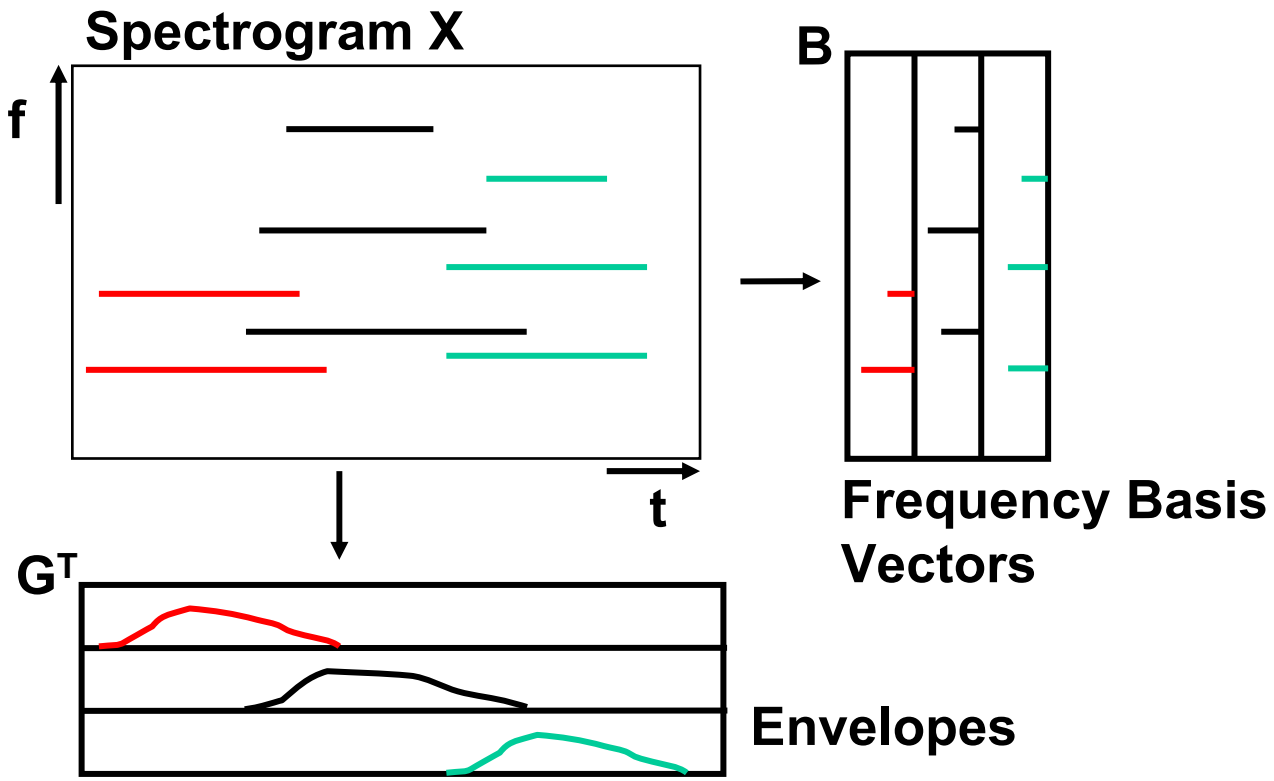
$\tilde{s}_m(n)$ **Estimated sources**

Separation by Non-Negative Matrix Factorization (NMF)

- Audio: approximates magnitude spectrogram by frequency basis vectors and corresponding envelopes

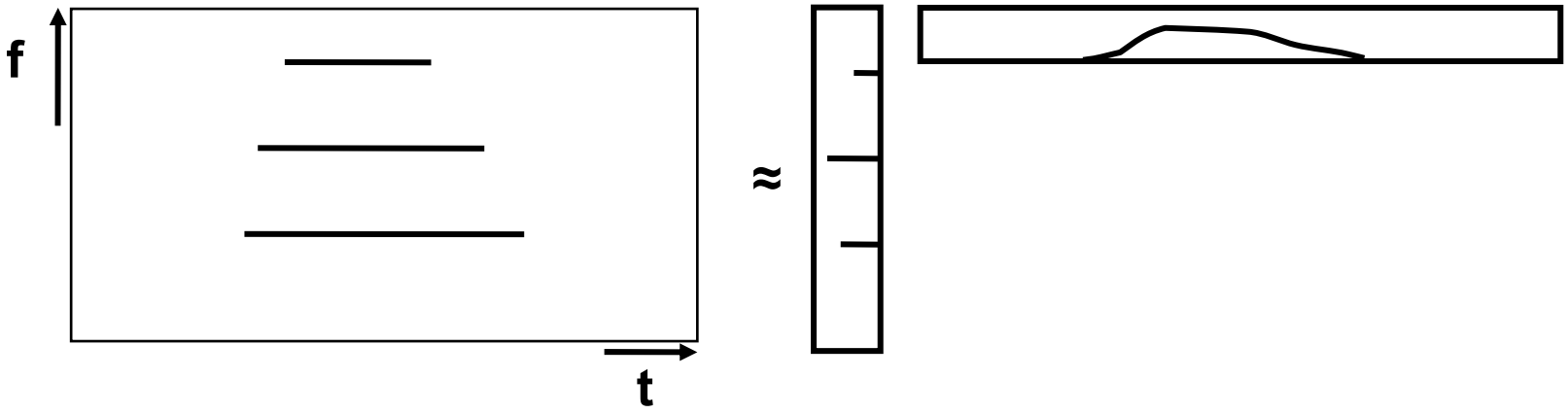
$$X \approx BG^T$$

- Toy example:

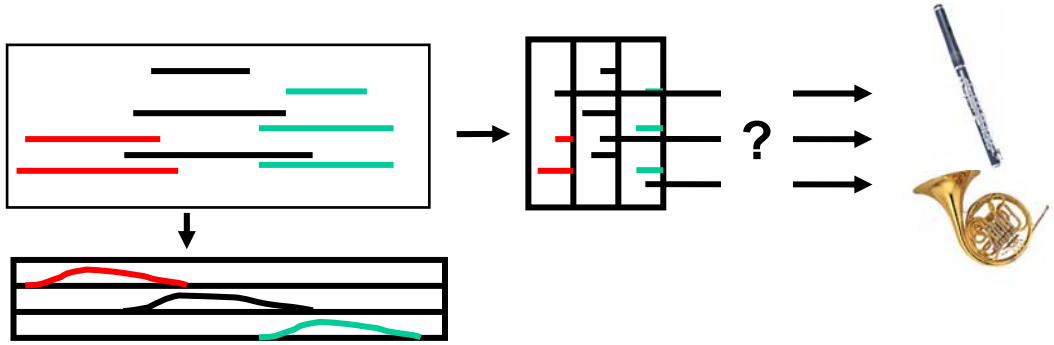


Separation by Non-Negative Matrix Factorization (NMF)

- Reconstruction of single channel spectrogram:



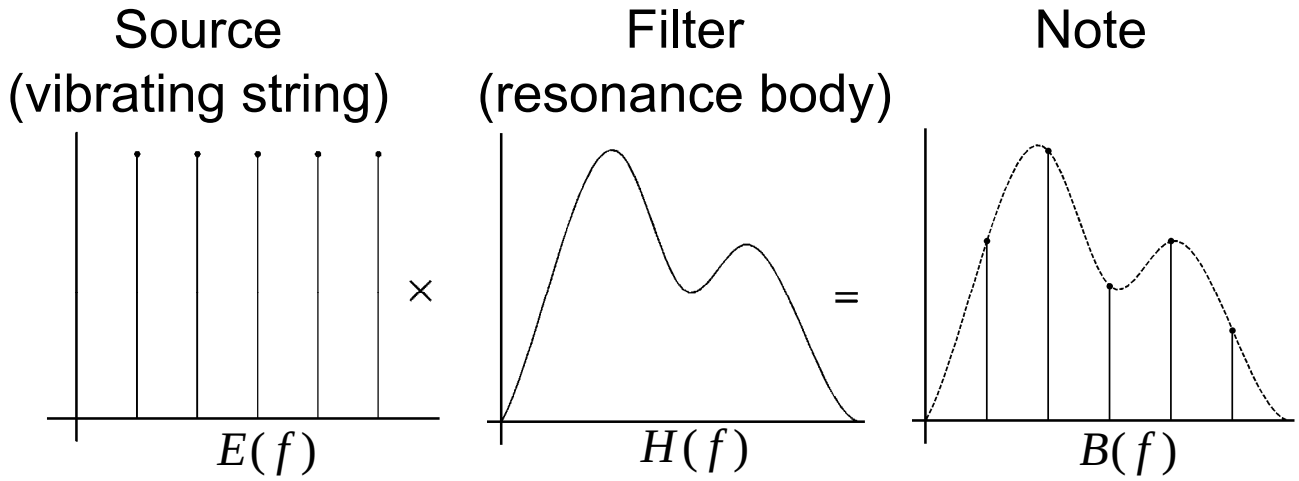
- Result: rank 1 approximation for single channel
- Factorizes notes not instruments → clustering



Outline

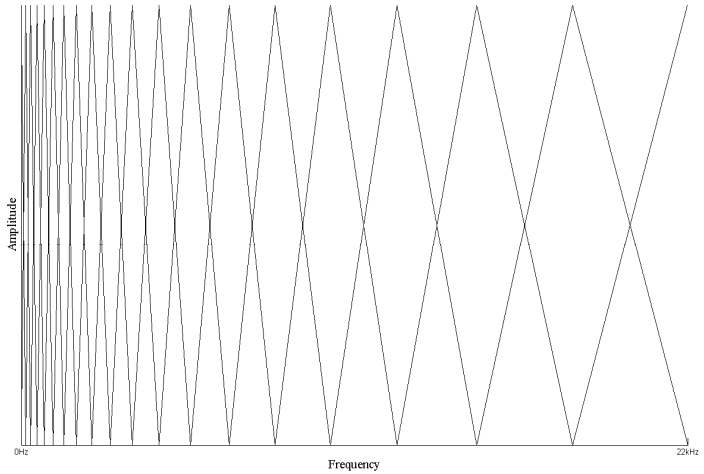
- Motivation
- Separation Framework
- Clustering
  - MFCC + k-Means
  - NMF Clustering
- Examples

Source-Filter Model



- Assumption:
  - Product of source and filter approximates spectrum
  - B(f) represents spectrum
- Idea: clustering by instrument specific filter

- MFCC
  - Mel-filtering (reduction from 2049 to 20 samples)



  - Logarithm

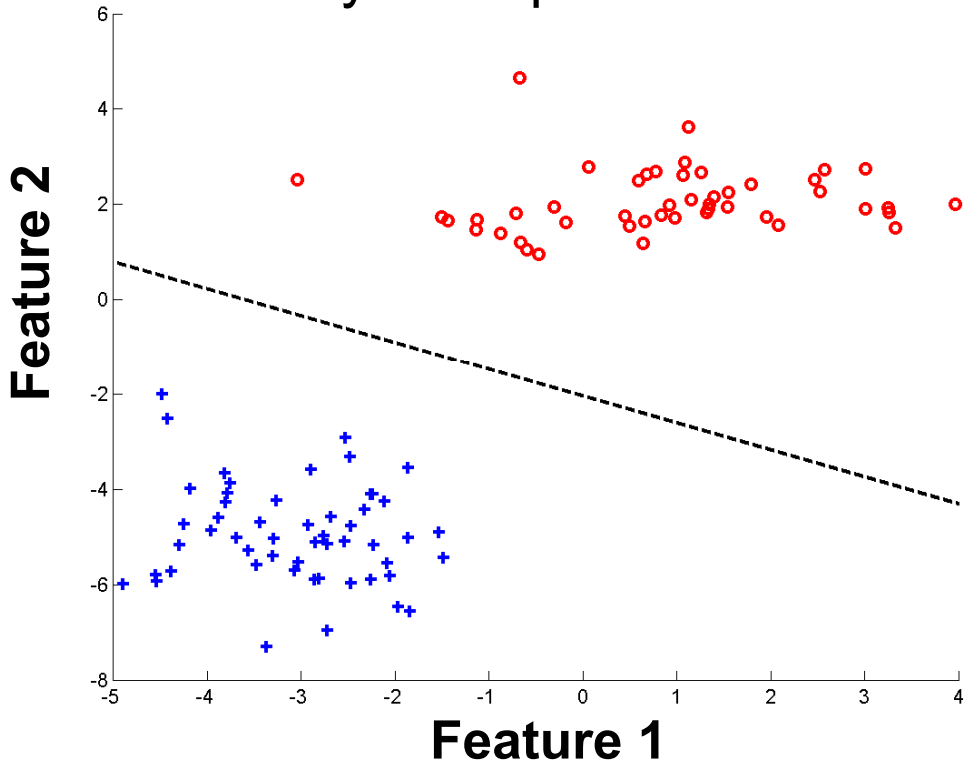
$$B \approx EH$$

$$\log(B) \approx \log(E) + \log(H)$$

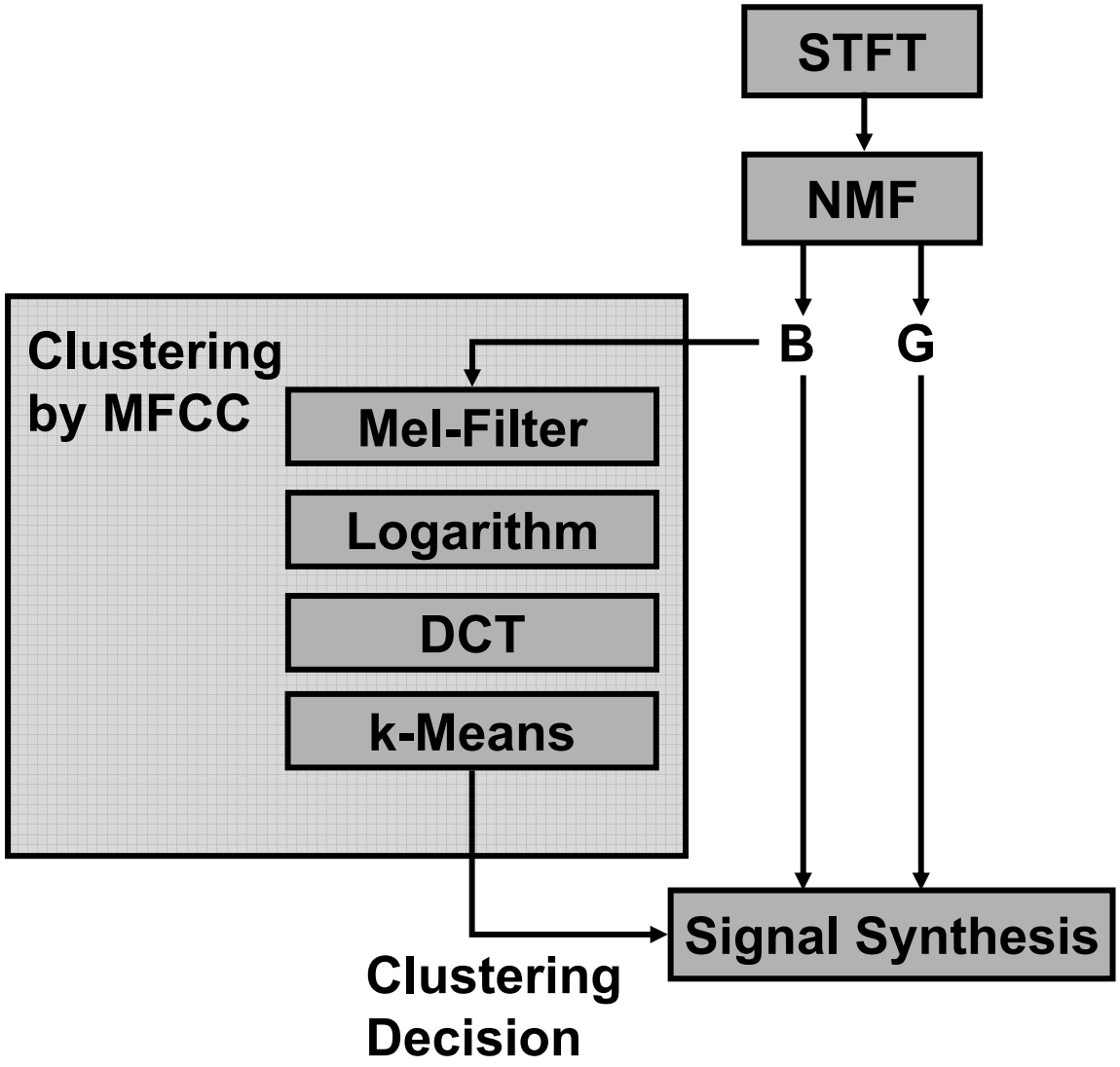
  - Discrete Cosine Transformation for decorrelation

MFCC + k-Means

- k-Means
  - Known number of clusters
  - Normalized variances for features → Euclidean distance
  - 2-dimensional toy example



Separation by MFCC-Clustering



MFCC vs. NMF

- Example:

MFCC

3.63 dB

3.89 dB

- Test set of 780 mixtures: 6.02 dB
- Perfect clustering: 12.01 dB

Clustering by NMF

- Assumption:

$$B \approx EH$$

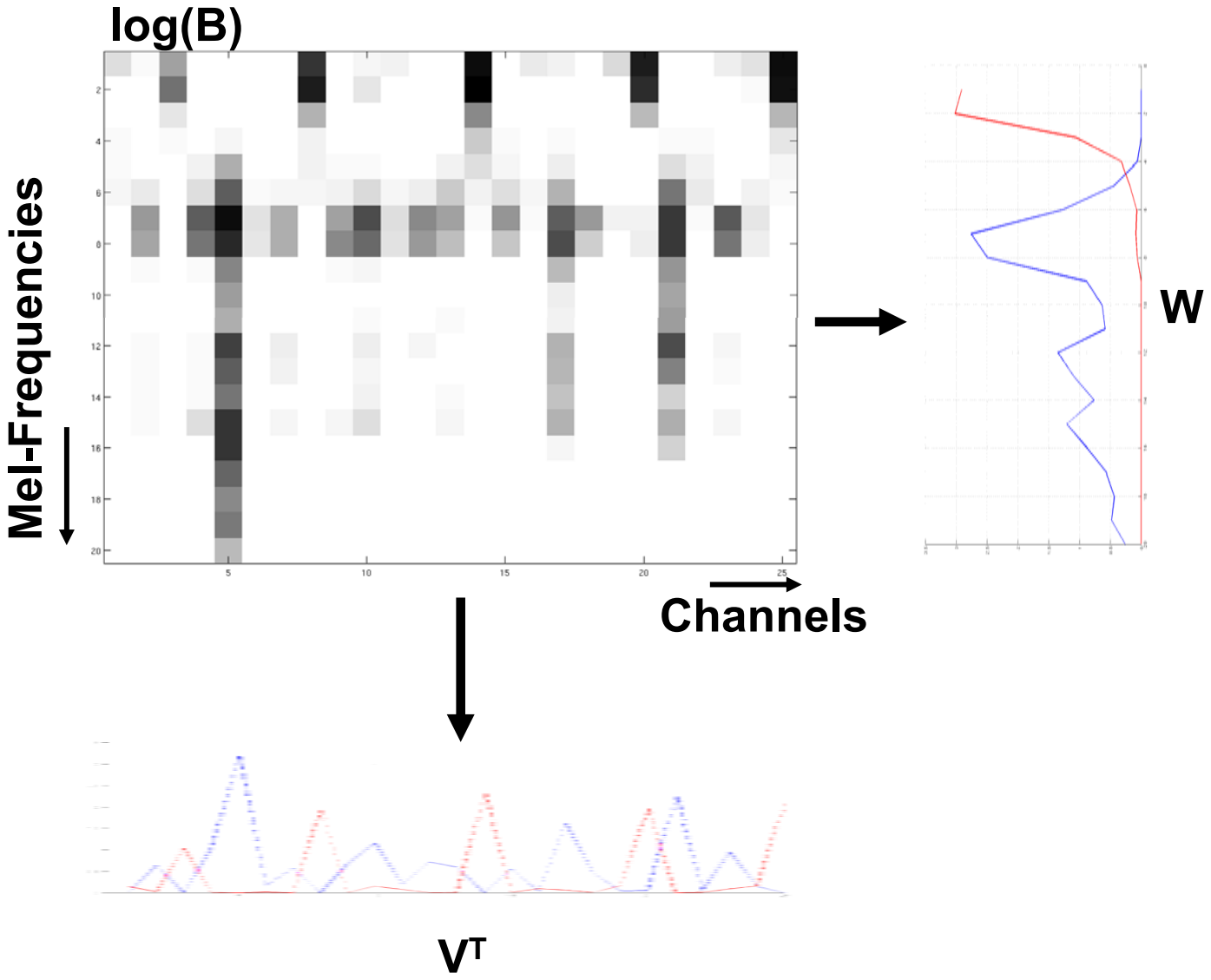
$$\log(B) \approx \log(E) + \log(H)$$

- MFCC: decorrelation of single basis vector
- Scenario: 2 instruments
  - 2 significant resonance filters in all basis vectors
- Apply second NMF on **ALL** basis vectors:

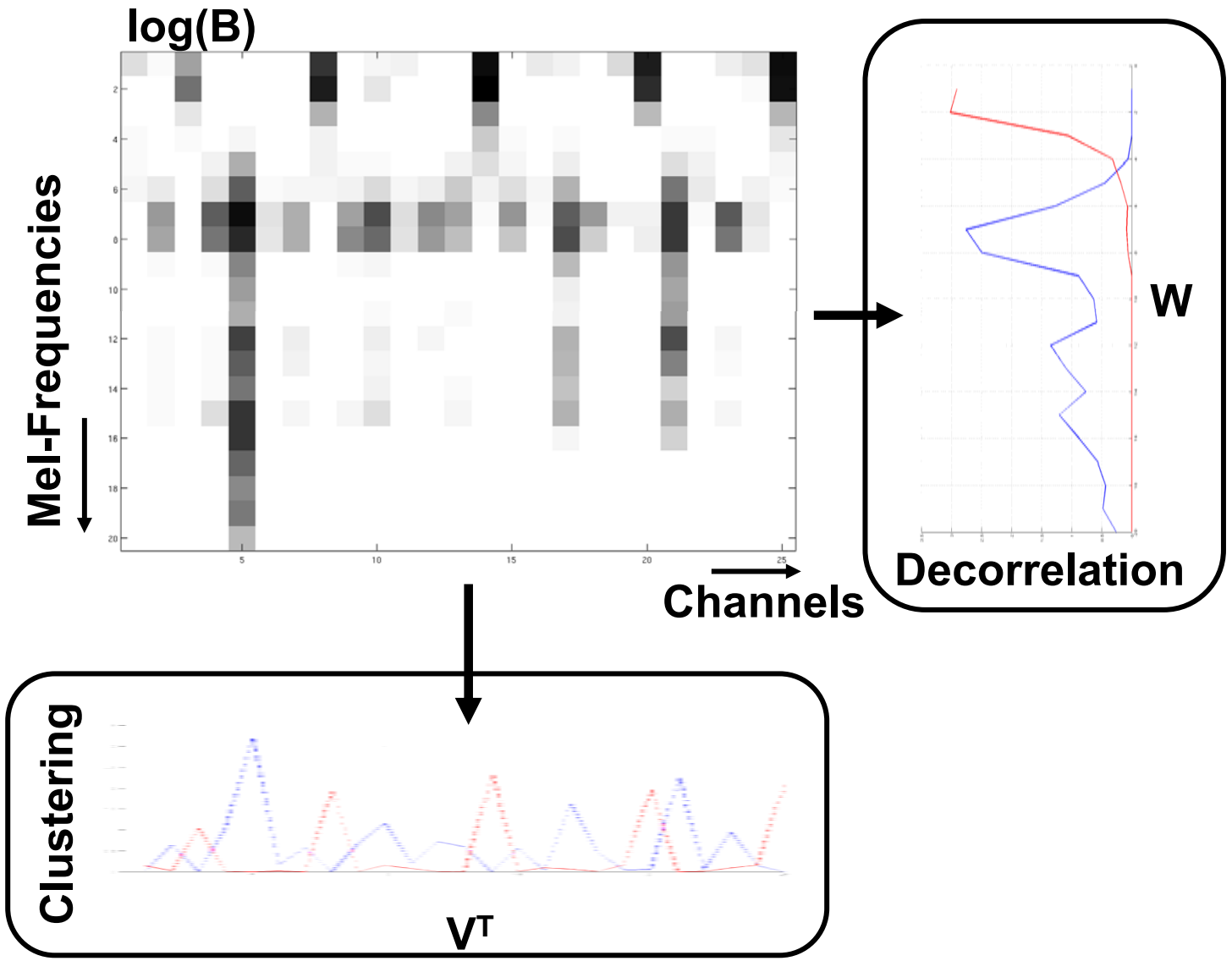
$$\log(B) \approx WV^T$$

  - Envelopes V correspond to activation matrix
  - Combined decorrelation/clustering for all basis vectors

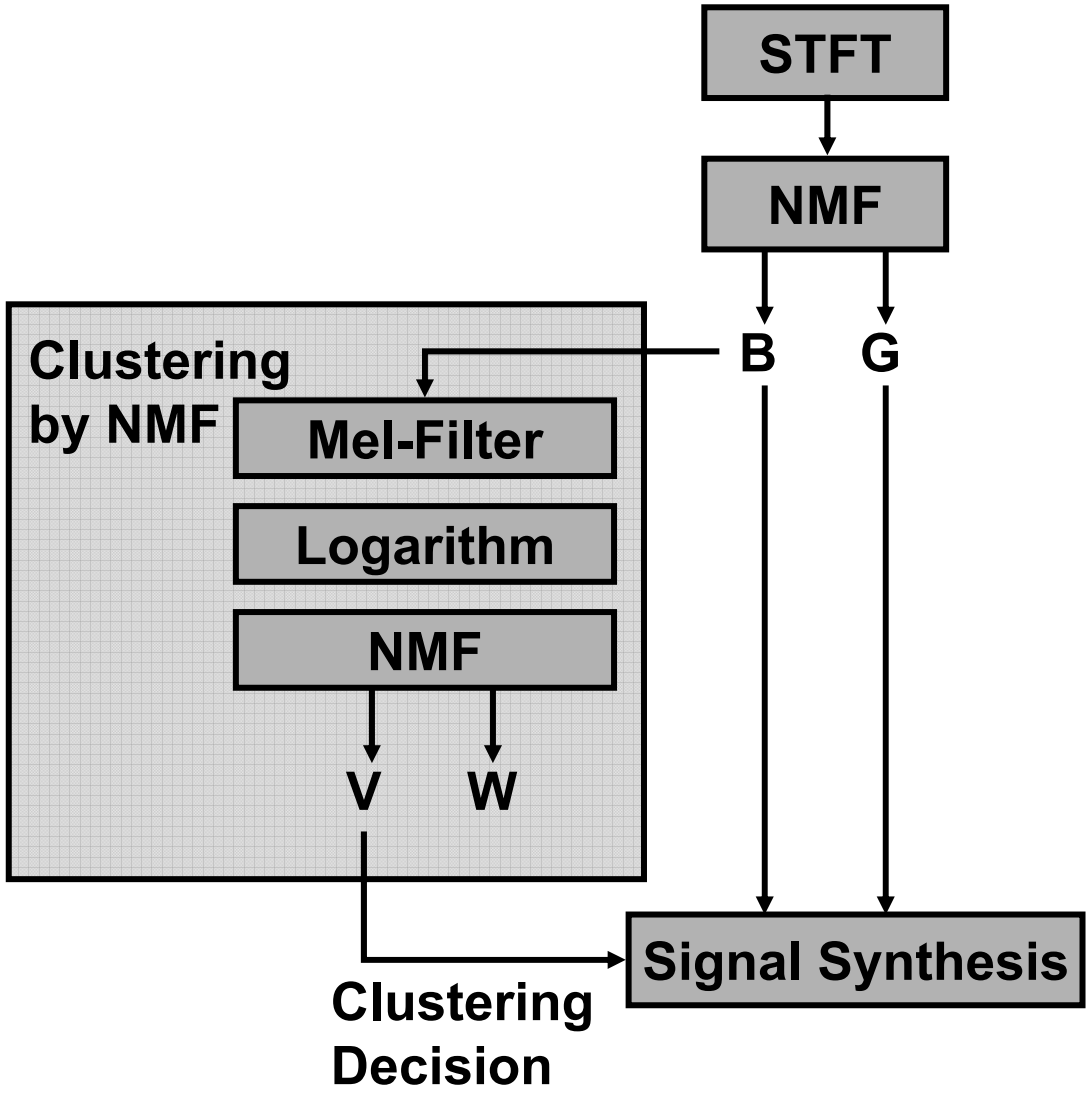
Clustering by NMF



Clustering by NMF



Separation by NMF-Clustering



MFCC vs. NMF

- Example:

	MFCC	NMF
	**3.63 dB**	**9.93 dB**
	**3.89 dB**	**10.02 dB**

- Test set of 780 mixtures: 6.02 dB → 7.77 dB
- Perfect clustering: 12.01 dB

- Thank you for your attention