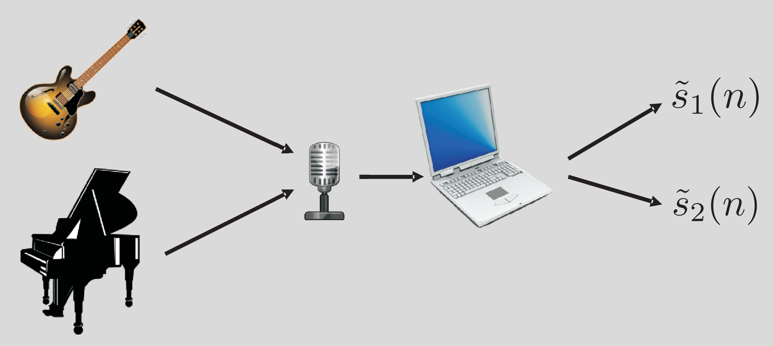
Beta Divergence for Clustering in Monaural Blind Source Separation

Martin Spiertz, Volker Gnann
Institute of Communications Engineering
RWTH Aachen University

8130

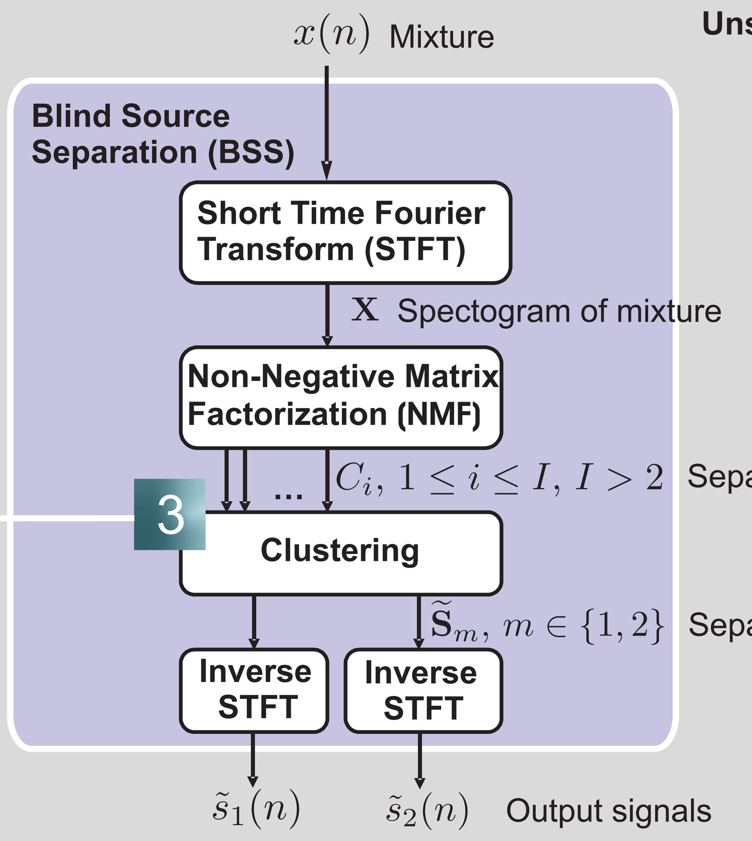
1 Blind Source Separation Scenario



Advantages of monaural scenario:
- Most flexible in number of sensors
- Multichannel case: useful as preprocessing step

Possible applications:
- Remixing
- Denoising
- Automatic transcription
- Instrument classification

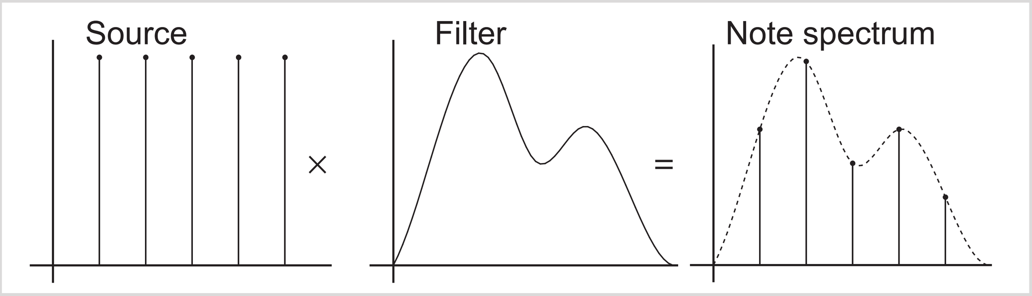
2 Signal Flow of NMF-Based Separation Algorithm



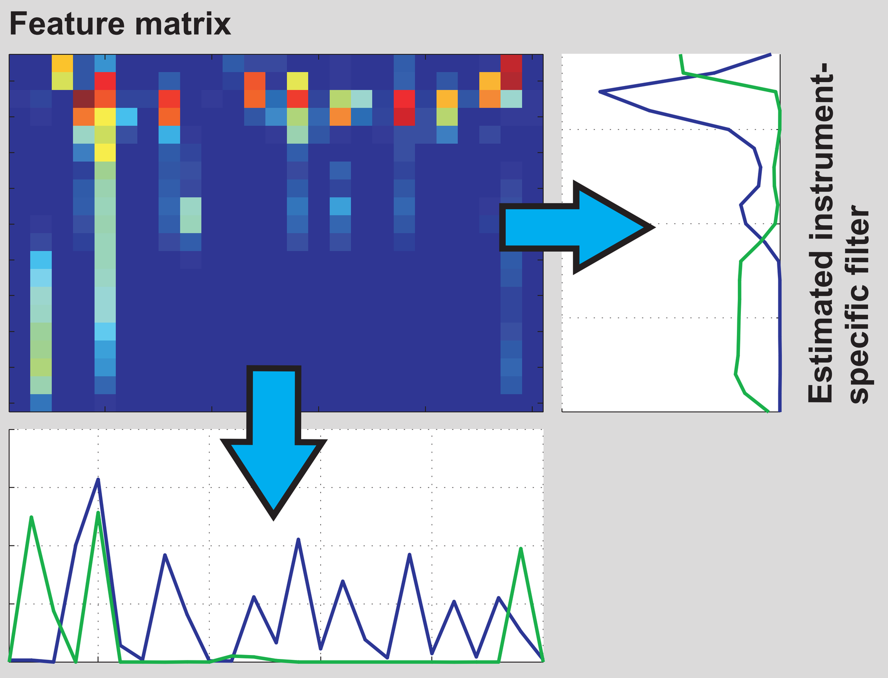
Unsupervised separation
- Input:
  - Monaural mixture
  - Number of sources
- No training step necessary
- No instrument classification

3 Clustering of Separated Sound Events

Feature according to source-filter model
- Evaluation motivated by MFCC
  - Mel filter bank
  - Logarithm
  - But: no decorrelation

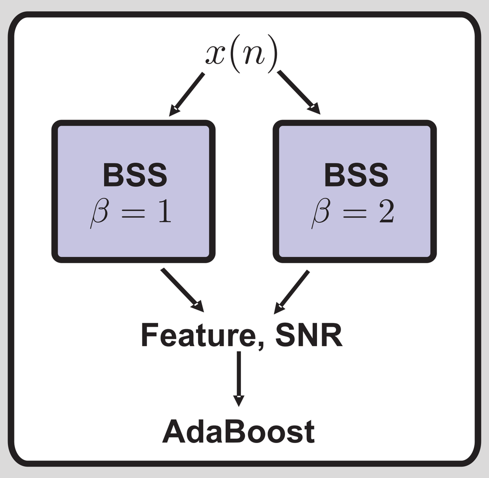


Clustering by second NMF
- Estimates two instrument-specific Filters
- Parameter β specifies factorization by target cost function: β-divergence
  - $\beta = 1$ divergence
  - $\beta = 2$ Euclidean distance
- Activity matrix corresponds to clustering decision



4 Adaptive β Decision

Classify optimal β
- Try both ($\beta = 1$ and $\beta = 2$)
- Evaluate features for both cases
- Features based on
  - Estimation of dynamic differences
  - Common assumptions for BSS e.g. statistical independence

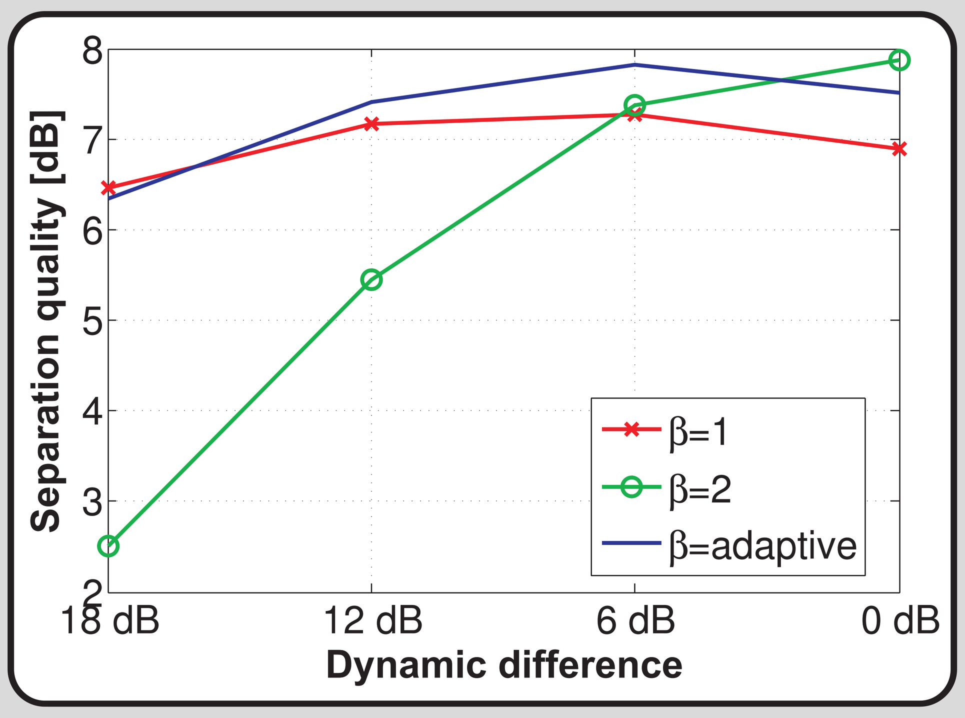


AdaBoost
- Classifier: which β leads to higher SNR (decision between 2 classes)
- Combines weak classifiers to single strong classifier
- Weak classifier
  - 1 dimensional: one feature + treshold

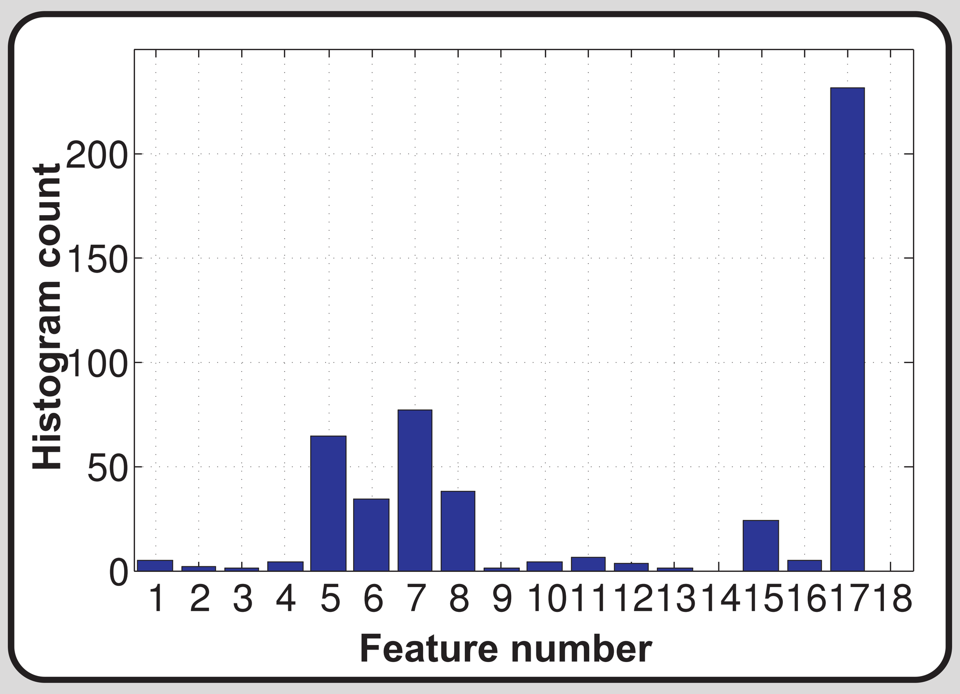
List of features
- Features evaluated for mixture $x(n)$ and output signals $\tilde{s}_m(n)$

Signal Features for Signal x, $\mathbf{X}$	
1	Estimated dynamic differences: $10\,\log_{10}\left(\lVert\mathbf{X}\rVert_2^2\right)$
2	Estimated dynamic differences: $10\,\log_{10}\left(\lVert x\rVert_2^2\right)$
3	Mean of temporal dynamics $\mathbf{d}_t$
4	Variance of temporal dynamics $\mathbf{d}_t$
(Dis-)Similarities of separated signal features	
5-8	Mean of features (1-4) for both active sources $\tilde{s}_m(n)$, $\widetilde{\mathbf{S}}_m$
9-12	Difference of features (1-4) between both active sources $\tilde{s}_m(n)$, $\widetilde{\mathbf{S}}_m$
Correlation Features	
13	Cross-correlation between both $\widetilde{\mathbf{S}}_m$
14	Pearsons rank correlation between both $\widetilde{\mathbf{S}}_m$
15	Cross-correlation between both $\tilde{s}_m(n)$
16	Pearsons rank correlation between both $\tilde{s}_m(n)$
Statistical Independence	
17	Histogram for $\tilde{s}_m(n)$
18	Histogram for $\widetilde{\mathbf{S}}_m$

5 Experimental Results



- SNR values for different dynamic differences of input signals
- Evaluated for a large test set (1770 mixtures)
- $\beta = 2$ (Euclidean distance)
  - Better for equal loudness
- $\beta = 1$ (divergence)
  - Better for large dynamic differences
  - Best overall results for constant β
- Adaptive β, $\beta \in \{1, 2\}$
  - Best results for unknown dynamic differences
  - Only slightly better results with $1 \leq \beta \leq 2$



- Histogram of chosen features
  - 100 training cycles of AdaBoost with different partitions of training/test set
  - Five features per training cycle
- Roughly six features sufficient
  - Robust and meaningful training

6 Conclusions

- Few features (~6) sufficient for adaption of β
- Large gains for unknown dynamic differences

Future Work

- Extend concept to adapt other parameters of algorithm, e.g. number of mel filters for estimation of instrument-specific filter

{Spiertz,Gnann}@ient.rwth-aachen.de
www.ient.rwth-aachen.de
Institut für Nachrichtentechnik, Melatener Str. 23, 52072 Aachen, Germany

RWTH RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN