Musical Source Separation: Principles and State of the Art

Juan José Burred
Équipe Analyse/Synthèse, IRCAM
burred@ircam.fr

ircam
Centre Pompidou

2nd International Workshop on Learning Semantics of Audio Signals (LSAS),
Paris, 21st June 2008

Presentation overview

1. Introduction
   - Paradigms, tasks, applications
   - Mixing models
2. Solving the linear mixing model
   - Joint and staged separation
3. Estimation of the mixing matrix
   - The need for sparsity
   - Independent Component Analysis
   - Clustering methods, other methods
4. Estimation of the sources
   - Norm minimization
   - Time-frequency masking
5. Methods using advanced source models
   - Adaptive basis decomposition methods
   - Sinusoidal methods
   - Supervised methods
6. Conclusions

Presentation overview

1. Introduction
    - Paradigms, tasks, applications
    - Mixing models

Sound Source Separation

- "Cocktail party effect"
  - E. C. Cherry, 1953.
  - Ability to concentrate attention on a specific sound source from within a mixture.
  - Even when interfering energy is close to energy of desired source.

- "Prince Shotoku Challenge"
  - Legendary Japanese prince Shotoku (6th Century AD) could listen and understand simultaneously the petitions by ten people.
  - Concentrate attention on several sources at the same time!
  - "Prince Shotoku Computer" (Okuno *et al.,* 1997)

- Both allegories imply an extra step of semantic understanding of the sources, beyond mere acoustical isolation.

[Cherry53] E. C. Cherry. Some Experiments on the Recognition of Speech, With One and Two Ears. *Journal of the Acoustical Society of America,* Vol. 25, 1953.

[Okuno97] H. G. Okuno, T. Nakatani and T. Kawabata. Understanging Three Simultaneous Speeches. *Proc. Int. Joint Conference on Artificial Intelligence (IJCAI),* Nagoya, Japan, 1997.

The paradigms of Musical Source Separation

- (based on [Scheirer00])
  - *Understanding without separation*
    - Multipitch estimation, music genre classification
    - "Glass ceiling" of traditional methods (MFCC, GMM) [Aucouturier&Pachet04]
  - *Separation for understanding*
    - First (partially) separate, then feature extraction
    - Source separation as a way to break the glass ceiling?
  - *Separation without understanding*
    - BSS: Blind Source Separation (ICA, ISA, NMF)
    - Blind means: only very general statistical assumptions taken.
  - *Understanding for separation*
    - Supervised source separation (based on a training database)

[Scheirer00] E. D. Scheirer. *Music-Listening Systems*. PhD thesis, Massachusetts Institute of Technology, 2000.

[Aucouturier&Pachet04] J.-J. Aucouturier and F. Pachet. Improving Timbre Similarity: How High is the Sky? *Journal of Negative Results in Speech and Audio Sciences*, 1 (1), 2004.

Required sound quality

- Regarding the quality of the separated sounds, source separation tasks can be divided into:
- Audio Quality Oriented (AQO)
  - Aimed at full unmixing at the highest possible quality.
  - Applications:
    - Unmixing, remixing, upmixing
    - Hearing aids
    - Post-production
- Significance Oriented (SO)
  - Separation quality just enough for facilitating semantic analysis of complex signals.
  - Less demanding, more realistic.
  - Applications:
    - Music Information Retrieval
    - Polyphonic Transcription
    - Object-based audio coding

Musical Source Separation Tasks

- Classification according to the nature of the mixtures:

\+ Difficulty -

Source position	Mixing process	Source/mixture ratio	Noise	Musical texture	Harmony
• changing • static	• echoic (changing impulse response) • echoic (static impulse response) • delayed • instantaneous	• underdetermined • overdetermined • even-determined	• noisy • noiseless	• monodic (multiple voices) • heterophonic • homophonic / homorhythmic • polyphonic / contrapuntal • monodic (single voice)	• tonal • atonal

- Classification according to available a priori information:

\+ A priori knowledge -

\+ Difficulty -

Source position	Source model	Number of sources	Type of sources	Onset times	Pitch knowledge
• unknown • statistical model • known mixing matrix	• none • statistical independence • sparsity • advanced/trained source models	• unknown • known	• unknown • known	• unknown • known (score/MIDI available)	• none • pitch ranges • score/MIDI available

Linear mixing model

- Only amplitude scaling before mixing (summing)

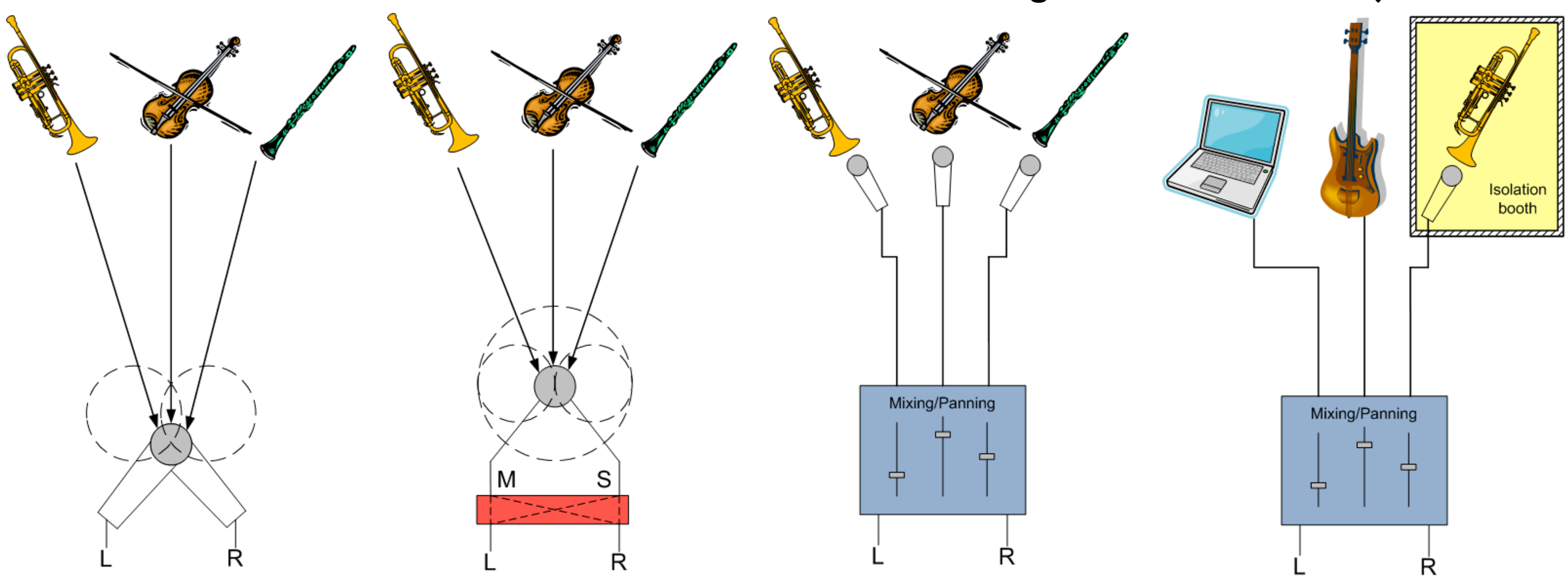
$$x_m(t) = \sum_{n=1}^{N} a_{mn} s_n(t), \qquad m = 1, \dots, M.$$

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{pmatrix} \cdot \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{pmatrix}$$

$$\boxed{\mathbf{X} = \mathbf{A}\mathbf{S}}$$

- Linear stereo recording setups:

XY Stereo | MS Stereo | Close miking | Direct injection



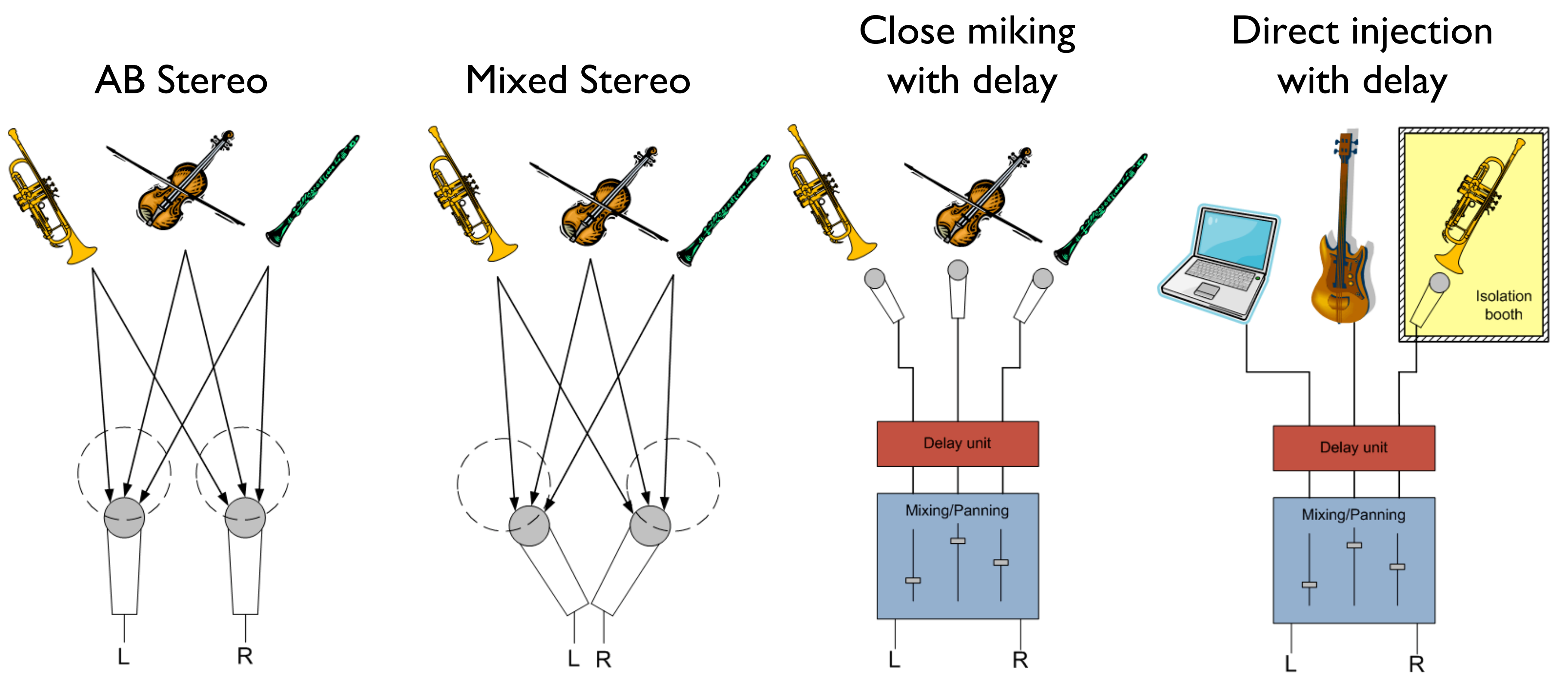
Delayed mixing model

- Amplitude scaling and delay before mixing

$$x_m(t) = \sum_{n=1}^{N} a_{mn} s_n(t - \delta_{mn}), \qquad m = 1, \ldots, M. \qquad \mathbf{A} = \begin{pmatrix} a_{11}\delta(t-\delta_{11}) & \ldots & a_{1N}\delta(t-\delta_{11}) \\ \vdots & \ddots & \vdots \\ a_{M1}\delta(t-\delta_{M1}) & \ldots & a_{MN}\delta(t-\delta_{MN}) \end{pmatrix}$$

$$\boxed{\mathbf{x} = \mathbf{A} * \mathbf{s}}$$

- Delayed stereo recording setups:



Convolutive mixing model

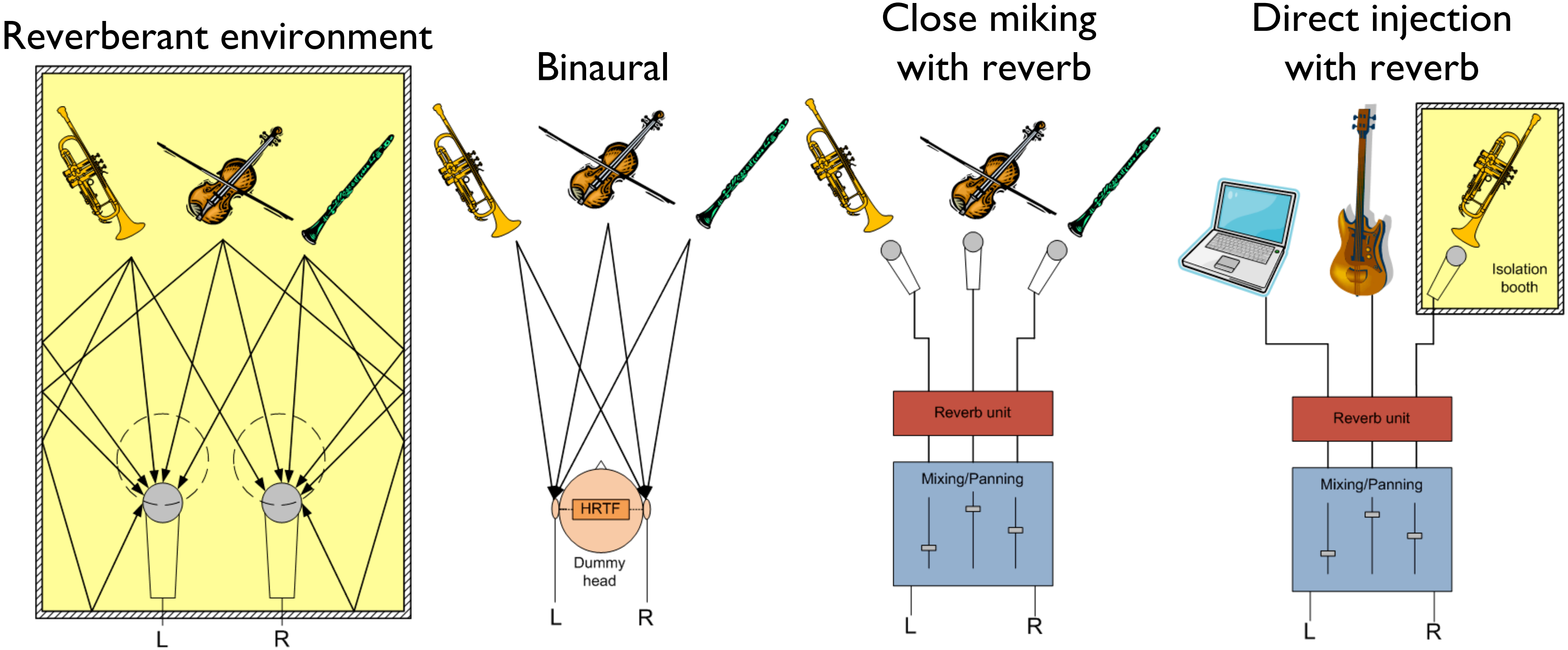
- Filtering between sources and sensors

$$x_m(t) = \sum_{n=1}^{N} h_{mn}(t) * s_n(t) = \sum_{n=1}^{N} \sum_{k=1}^{K_{mn}} a_{mnk} s_n(t - \delta_{mnk}), \qquad m = 1, \ldots, M.$$

$$\mathbf{A} = \begin{pmatrix} h_{11}(t) & \ldots & h_{1N}(t) \\ \vdots & \ddots & \vdots \\ h_{M1}(t) & \ldots & h_{MN}(t) \end{pmatrix}$$

$$\boxed{\mathbf{x} = \mathbf{A} * \mathbf{s}}$$

- Convolutive stereo recording setups:



Some terminology

- System of linear equations: $\boxed{\mathbf{X} = \mathbf{AS}}$
  - Usual algebraic methods from high school: $\mathbf{X}$ known, $\mathbf{A}$ known, $\mathbf{S}$ unknown
  - But in source separation: unknown variables ($\mathbf{S}$, sources) AND unknown coefficients ($\mathbf{A}$, mixing matrix)
- Algebra terminology is retained for source separation:
  - More equations (mixtures) than unknowns (sources): overdetermined
  - Same number of equations (mixtures) than unknowns (sources): determined (square $\mathbf{A}$)
  - Less equations (mixtures) than unknowns (sources): underdetermined
- The underdetermined case is the most demanding, but also the most important for music!
  - Music is (still) mostly in stereo, with usually more than 2 instruments
  - Overdetermined and determined situtations are only of interest for arrays of sensors or arrays of microphones (localization, tracking)
- Alternative interpretation of the linear model as a linear transform from signal space to mixture space, with $\mathbf{A}$ the transformation matrix and the columns of $\mathbf{A}$ the transformation bases.

Presentation overview

1. Introduction
    - Paradigms, tasks, applications
    - Mixing models
2. Solving the linear mixing model
    - Joint and staged separation

Solving the linear model

- Direct way to tackle the problem:
  - Mean Square Error (MSE) minimization: $\min_{\mathbf{A},\mathbf{S}} \|\mathbf{X} - \mathbf{AS}\|_F^2$
  - F is the Frobenius norm ("matrix energy")
  - BUT: this has infinitely many solutions
- One must assume probability distributions for the involved variables
  - Maximum A Posteriori (MAP) approach: maximize $P(\mathbf{A}, \mathbf{S}|\mathbf{X})$
  - Applying Bayes' theorem $P(\mathbf{A}, \mathbf{S}|\mathbf{X}) = \dfrac{P(\mathbf{X}|\mathbf{A}, \mathbf{S})P(\mathbf{A})P(\mathbf{S})}{P(\mathbf{X})}$ and
  - Assuming $\mathbf{A}$ has a uniform distribution (all source positions are equally equal) and
  - Assuming the sources are statistically independent this finally yields

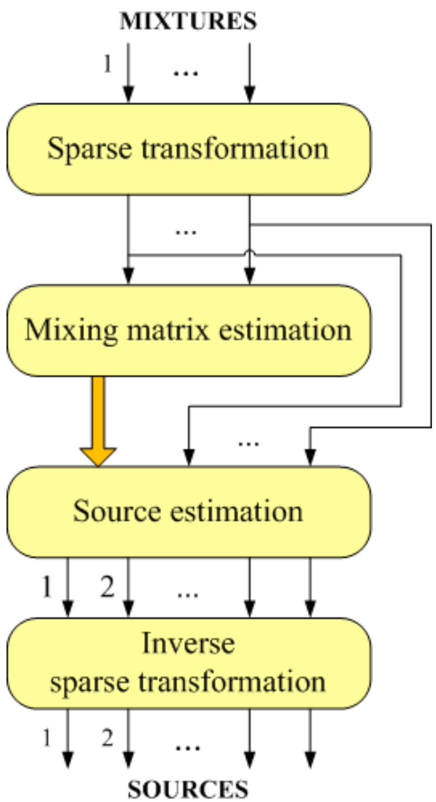
$$\min_{\mathbf{A},\mathbf{S}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{AS}\|_F^2 - \sum_{n,t} l_n(s_n(t)) \right\}$$

  - σ^2 is the noise variance (if any) and l_n is the assumed log-density of the sources

Staged separation

- However, such a joint estimation of $\mathbf{A}$ and $\mathbf{S}$ is:
  - Extremely computationally demanding
  - Unstable with respect to convergence

- Most methods follow thus a staged approach: first estimate the mixing matrix, then estimate the sources.



- Note that, if $\mathbf{A}$ is square (determined source separation) and invertible (virtually always for usual mixtures), then the sources can be readily obtained by $\boxed{\hat{\mathbf{S}} = \hat{\mathbf{A}}^{-1}\mathbf{X}}$

  (^ denotes estimation)

- In that case, source separation amounts to mixing matrix estimation!

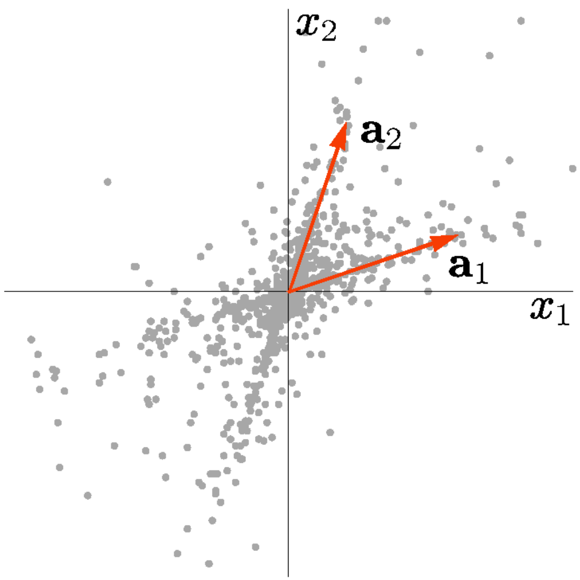
- In the underdetermined case, $\mathbf{A}$ is rectangular and thus non-invertible. Thus, a second source estimation stage is needed!

Presentation overview

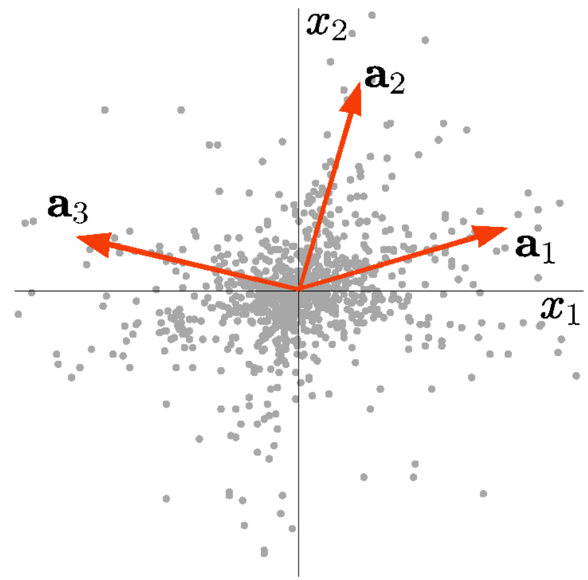
1. Introduction
    - Paradigms, tasks, applications
    - Mixing models
2. Solving the linear mixing model
    - Joint and staged separation
3. Estimation of the mixing matrix
    - The need for sparsity
    - Independent Component Analysis
    - Clustering methods, other methods

Mixing matrix estimation

- Simple examples can be visualized by means of scatter plots



Determined mixture
(2 channels, 2 sources)



Underdetermined mixture
(2 channels, 3 sources)

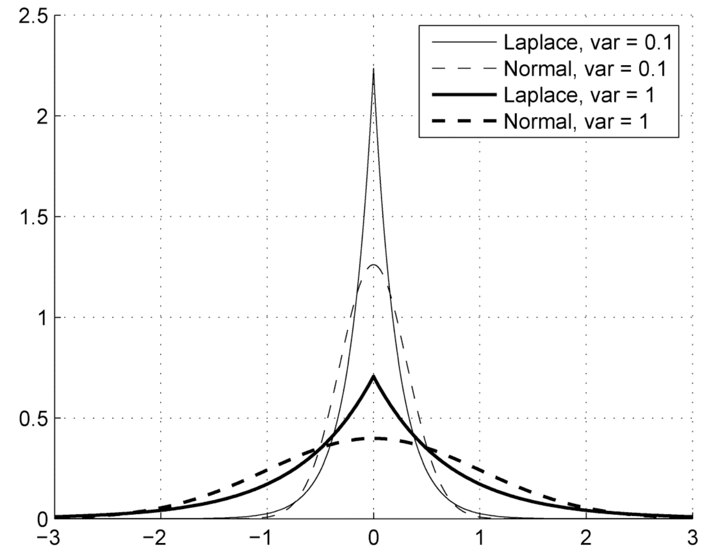
- The coordinates of each data point are the values of a certain signal coefficient (time sample, time-frequency bin) in each of the mixtures.

- Data points tend to concentrate around the vectors defined by the columns of the mixing matrix: the mixing directions.

- The goal of mixing matrix estimation is thus to find such vectors.

The need for sparsity

- A signal is said to be sparse if most of its coefficients (in some domain) are zero or close to zero.
- Sparse signals will have a peaked probability distribution.
  - Example: Laplacian signals are sparser than Gaussian signals



Laplace distribution: $p(c) = \frac{\lambda}{2} e^{-\lambda|c-\mu|}$

- Geometrical perspective:
  - The sparser the signals, the more their coefficients will be concentrated around the mixing directions, and the easier will be the detection of the directions.
- Analytical perspective:
  - Remember the MAP problem: $\min_{\mathbf{A},\mathbf{S}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \sum_{n,t} l_n(s_n(t)) \right\}$

Penalty for sparsity

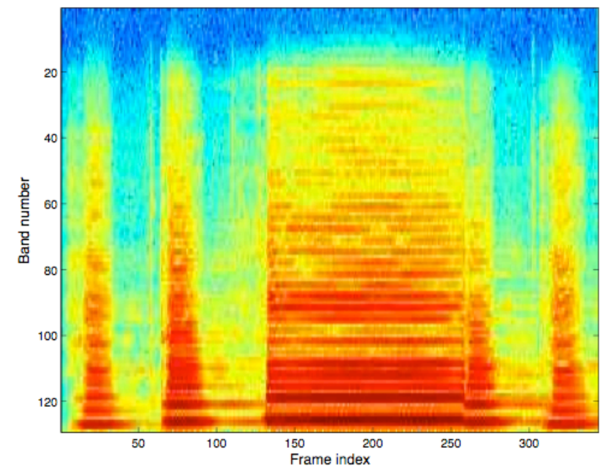
- Measures of sparsity
  - L1-norm
  - Kurtosis
  - Negentropy

L1-norm: $\|\mathbf{c}\|_1 = \sum_{i=1}^{C} |c_i|$

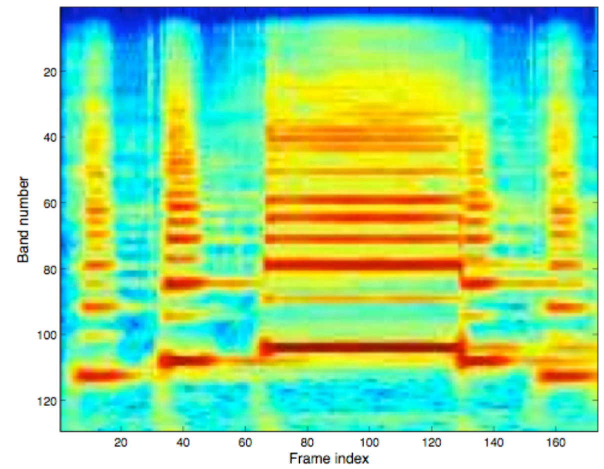
How to increase sparsity

- Time-frequency domain much sparser than time domain
  - Short Time Fourier Transform (STFT)
- Logarithmic resolution front-ends
  - Constant-Q Transform (CQT)
  - Discrete Wavelet Transform (DWT)
- Auditory resolution front-ends
  - Bark
  - ERB (Equal Rectangular Bandwidth)
  - Mel
- Adaptive signal decompositions
  - Basis Pursuit
  - Matching Pursuit

Spectrogram (|STFT|)

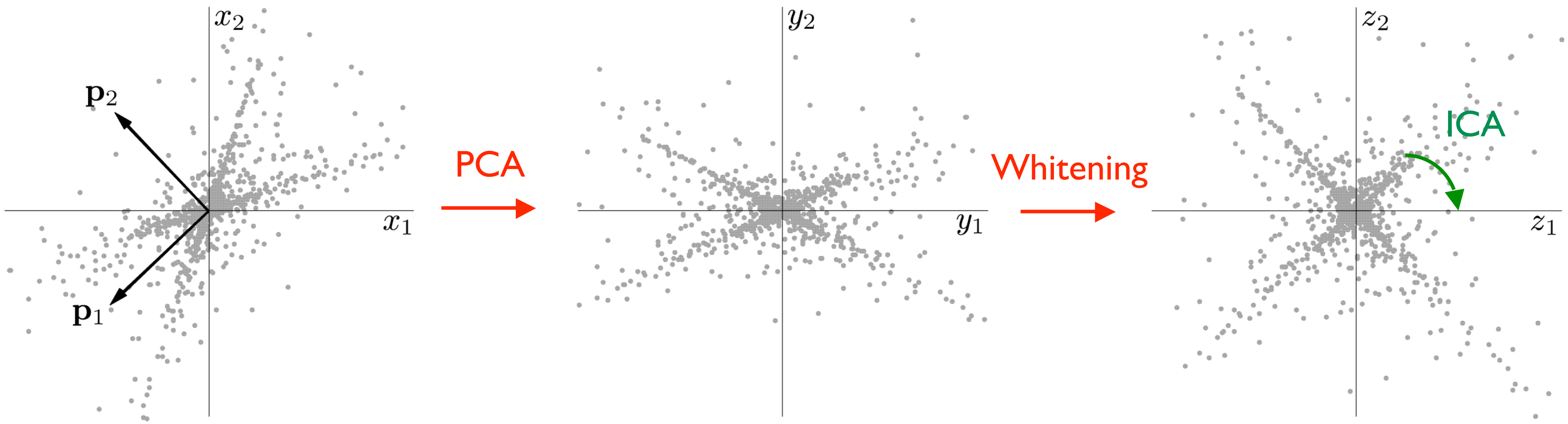


ERB



Independent Component Analysis (1)

- ICA tries to find the mixing directions by aligning the coefficient clusters to the (orthogonal) scatter axes.



  - Note that Principal Component Analysis (PCA), which finds the directions of greatest variance, is not enough for the alignment.
  - However, PCA is used as a first step for ICA because, when followed by whitening (variance normalization), it makes the mixing directions orthogonal, and thus ICA reduces to finding the remaining rotation.
  - Also, note that this is only possible for determined mixtures $\rightarrow$ not very useful for music!
- Axis alignment corresponds to the sources being statistically independent.

Independent Component Analysis (II)

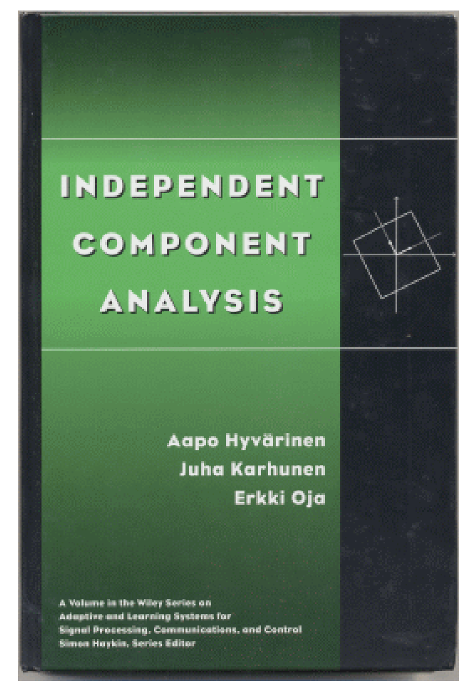
- ICA works by maximizing some objective measure of statistical independence between candidate sources.
- Methods based on maximizing nongaussianity of the sources
  - FastICA based on kurtosis or negentropy
- Methods based on minimizing mutual information between sources
- Methods based on Maximum Likelihood (ML) estimation
  - Bell-Sejnowski (BS) algorithm
  - Natural gradient algorithm
  - FastICA based on ML
- Tensorial methods (“decorrelate” higher order statistics)
  - FOBI (Fourth-Order Blind Identification)
  - JADE (Joint Approximate Diagonalization of Eigenmatrices)
- Sound examples (Hyvärinen *et al.*)

Original sources

Mixtures

Separated sources

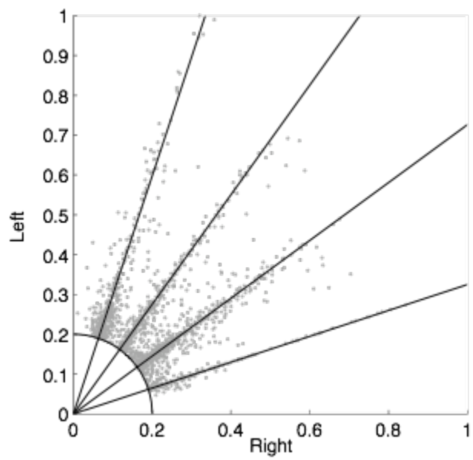
Hyvärinen + Karhunen + Oja



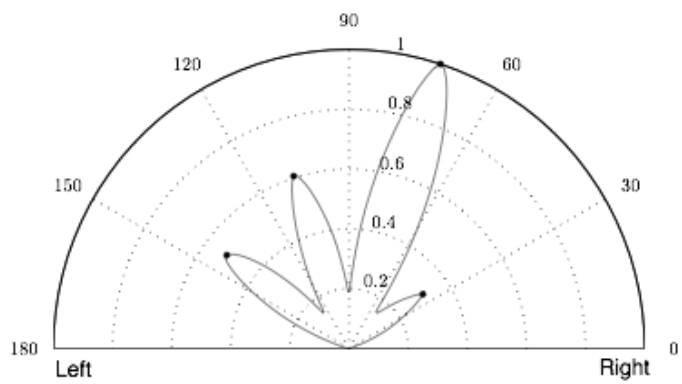
Clustering methods

- Explore the mixture space to find the clusters.
- Allow underdetermined separation!
- Direct inspection of the scatter plot: sparsity is crucial!
- Example: kernel-based angular clustering
  - [Bofill&Zibulevsky01]
  - Kind of smoothed histogram

Mixture scatter and found directions



Estimated density (polar)

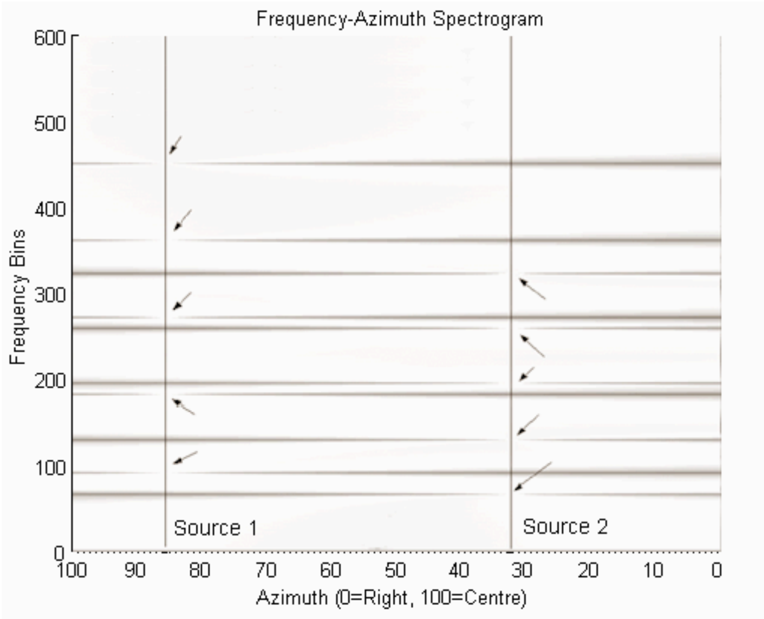


- Also: methods based on k-Means, fuzzy C-means clustering...

[Bofill&Zibulevsky01] P. Bofill and M. Zibulevsky. Underdetermined Blind Source Separation Using Sparse Representations. *Signal Processing*, Vol. 81, 2001.

Other methods for mixing matrix estimation

- Phase cancellation methods
  - ADRess (Azimuth Discrimination and Resynthesis) [Barry04]
  - Artificial stereo panning retains phase and only changes amplitude between channels → phase cancellation in the inter-channel difference spectrogram



(Fig. from [Barry04])

- Methods from image processing applied to the scatter plots
  - Example: application of the Hough transform to detect straight lines created by the direction clusters [Lin97]

[Barry04] D. Barry, B. Lawlor and E. Coyle. Sound Source Separation: Azimuth Discrimination and Resynthesis. *Proc. Int. Conf. on Digital Audio Effects (DAFX)*, Naples, Italy, 2004.

[Lin97] J. K. Lin, D. G. Grier and J. D. Cowan. Feature Extraction Approach to Blind Source Separation. *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP)*, 1997.

Presentation overview

1. Introduction
   - Paradigms, tasks, applications
   - Mixing models
2. Solving the linear mixing model
   - Joint and staged separation
3. Estimation of the mixing matrix
   - The need for sparsity
   - Independent Component Analysis
   - Clustering methods, other methods
4. Estimation of the sources
   - Norm minimization
   - Time-frequency masking

Source estimation by norm minimization

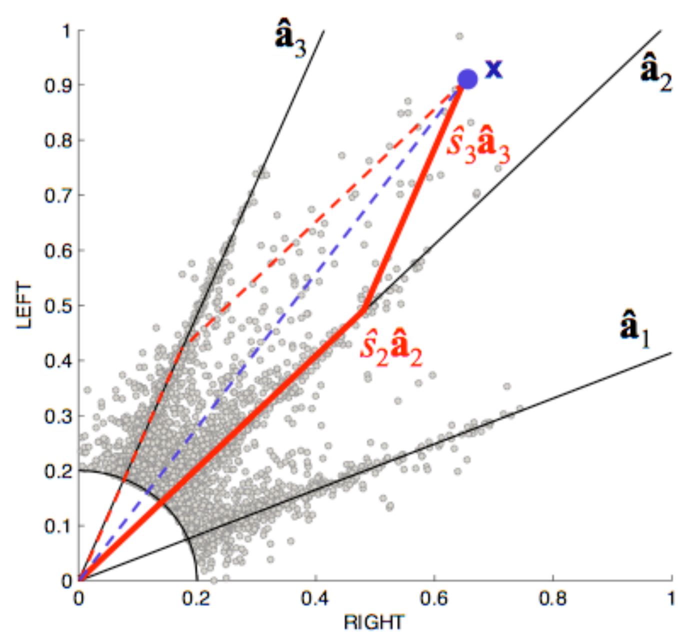
- In the underdetermined case, $\mathbf{A}$ is rectangular and thus non-invertible. Thus, a second source estimation stage is needed!

- Norm minimization methods
  - Recall (again) the minimization problem $\min_{\mathbf{A},\mathbf{S}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{AS}\|_F^2 - \sum_{n,t} l_n(s_n(t)) \right\}$
  - Assuming no noise, known $\mathbf{A}$ and Laplacian (sparse) sources, this simplifies to an L1-norm minimization problem:

$$\hat{\mathbf{S}} = \underset{\mathbf{X}=\hat{\mathbf{A}}\mathbf{S}}{\operatorname{argmin}} \left\{ \sum_{n,t} |s_n(t)| \right\}$$

  - A realization thereof is the shortest-path algorithm
  - Sound examples for angular kernel clustering plus shortest-path estimation:

	Independent melodies				Musical performance	
			Original sources			
			Mixtures			
			Separated sources			

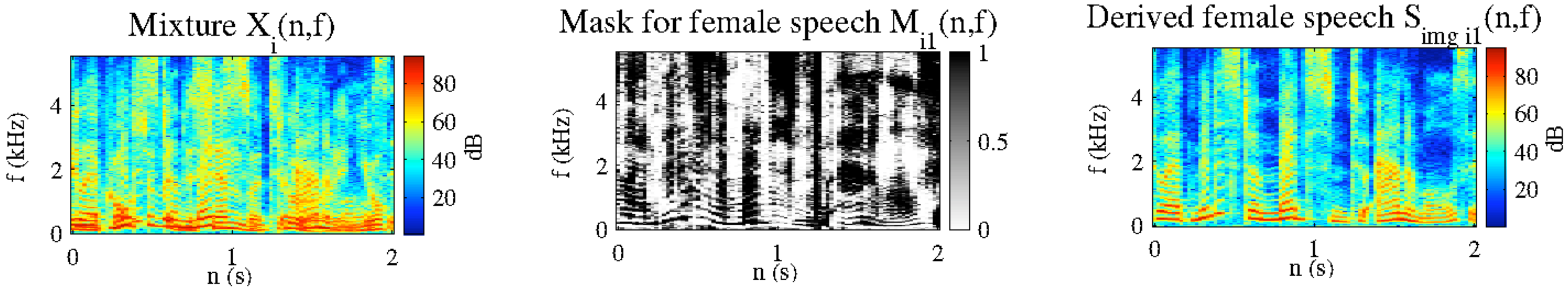


Time-frequency masking (1)

- Goal: find a mask M that retrieves one source when used to filter a given time-frequency representation.

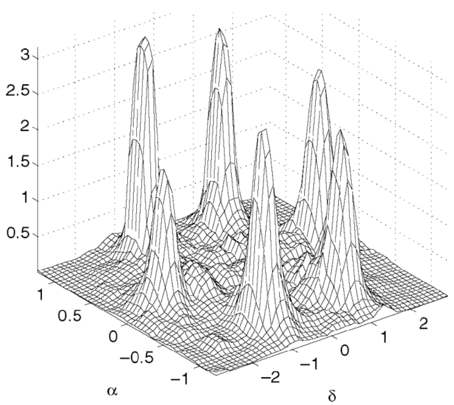
$$\boxed{\hat{\mathbf{S}}_n(r,k) = \mathbf{M}_{mn}(r,k) \circ \mathbf{X}_m(r,k)}$$

$\circ$ is the Hadamard (element-wise) product



(Fig. from [Vincent06])

- Adaptive Wiener filtering
- Binary time-frequency masking
  - DUET (Degenerate Unmixing Estimation Technique) [Yilmaz&Rickard04]
    - Histogram of Interchannel Intensity (IID) and Phase (IPD) Differences
    - Binary Mask created by selecting bins around histogram peaks.



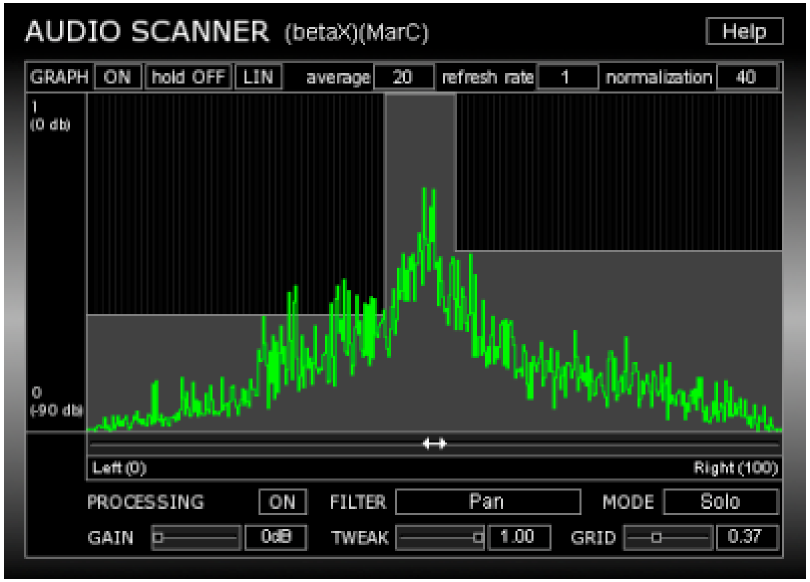
(Fig. from [Yilmaz&Rickard04])

- Drawback of t-f masking: “musical noise” or “burbling” artifacts

[Yilmaz&Rickard04] Ö. Yilmaz and S. Rickard. Blind Separation of Speech Mixtures via Time-Frequency Masking. *IEEE Trans. on Signal Processing*. Vol. 52(7), July 2004

Time-frequency masking (2)

- Human-assisted time-frequency masking [Vinyes06]
  - Human-assisted selection of the time-frequency bins out of the DUET-like histogram for creating the unmixing mask
  - Implementation as a VST plugin ("Audio Scanner")



[Vinyes06] M. Vinyes, J. Bonada and A. Loscos. Demixing Commercial Music Productions via Human-Assisted Time-Frequency Masking. *120th AES convention,* Paris, France, 2006.

Presentation overview

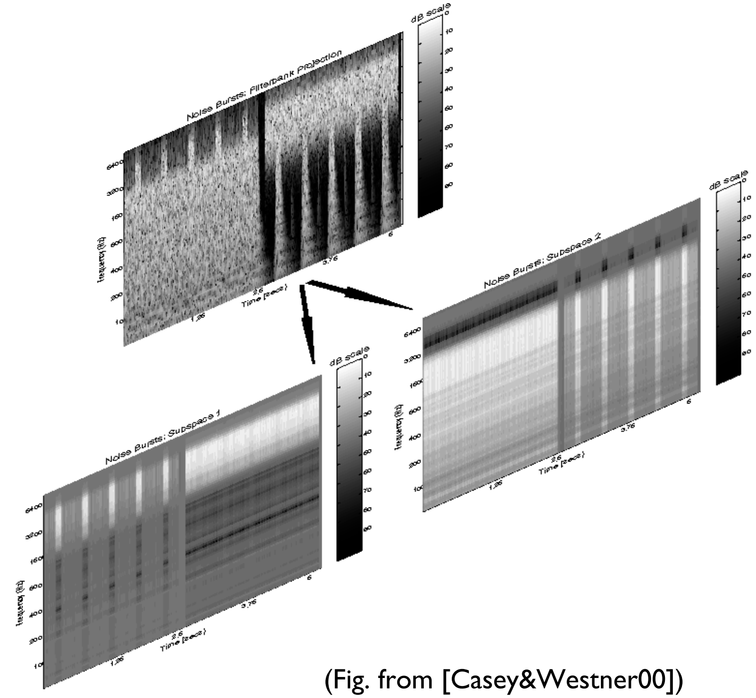
1. Introduction
   - Paradigms, tasks, applications
   - Mixing models
2. Solving the linear mixing model
   - Joint and staged separation
3. Estimation of the mixing matrix
   - The need for sparsity
   - Independent Component Analysis
   - Clustering methods, other methods
4. Estimation of the sources
   - Norm minimization
   - Time-frequency masking
5. Methods using advanced source models
   - Adaptive basis decomposition methods
   - Sinusoidal methods
   - Supervised methods

Advanced source models methods

- Until now: blind approaches (only general, statistical assumptions)
- The use of (sometimes music-specific) advanced source models allow to improve separation quality and to handle highly underdetermined situations (e.g. separation from mono mixtures)
- Classification according to a priori knowledge
  - Supervised
    - Based on training the model with a sound example database
    - Better quality and more demanding situations at the cost of less generality
  - Unsupervised
- Classification according to model type
  - Adaptive basis decompositions (ISA, NMF, NSC)
  - Sinusoidal Modeling
- Classification according to mixture type
  - Monaural systems
  - Hybrid systems combining advanced source models with spatial diversity

Independent Subspace Analysis

- Application of ISA to audio: Casey and Westner, 2000.
- Application of ICA to the spectogram of a mono mixture.
- Each independent component corresponds to an independent subspace of the spectrogram.



(Fig. from [Casey&Westner00])

- Component-to-source clustering
  - The extracted components usually do not directly correspond to the sources.
  - They must be clustered together according to some similarity criterion.
  - Casey&Westner use a matrix of Kullback-Leibler divergences called the ixegram.

[Casey&Westner00] M. Casey and A. Westner. Separation of Mixed Audio Sources by Independent Subspace Analysis. *Proc, Int. Computer Music Conference (ICMC),* Berlin, Germany, 2000.

Nonnegative Matrix Factorization

- Matrix factorization ($\mathbf{X} = \mathbf{AS}$) imposing non-negativity.
- Needed when using magnitude or power spectrograms.
- NMF does not aim at statistical independence, but:
  - It has been proven that, under some conditions, non-negativity is sufficient for separation.
  - NMF yields components that very closely correspond to the sources.
  - To date, there is no exact theoretical explanation why is that so!
- Use for transcription:
  - P. Smaragdis and J.C. Brown. Non-Negative Matrix Factorization for Polyphonic Music Transcription. *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA),* New Paltz, USA, 2003.
- Use for separation:
  - B. Wang and M. D. Plumbley. Musical Audio Stream Separation by Non-Negative Matrix Factorization. *Proc. UK Digital Music Research Network (DMRN) Summer Conf.,* 2005.

Nonnegative Sparse Coding

- Combination of non-negativity and sparsity constraints in the factorization.
- [Virtanen03]: NSC is optimized with an additional criterion of temporal continuity.
  - Measured by the absolute value of the overall amplitude difference between consecutive frames.

$$c(A) = \frac{1}{2}\sum_{t=1}^{T}\sum_{n=1}^{N}\left|a_{t-1,n} - a_{t,n}\right|$$

Mixture Component 1 Component 2

- [Virtanen04]: Convolutive Sparse Coding
  - Improved temporal accuracy by modeling the sources as the convolution of spectrograms with a vector of onsets.

$$(\boldsymbol{M}_n)_{t,f} = \left(\sum_{n=1}^{N}[\boldsymbol{a}_n \otimes \boldsymbol{s}_{n,f}]\right)_t$$

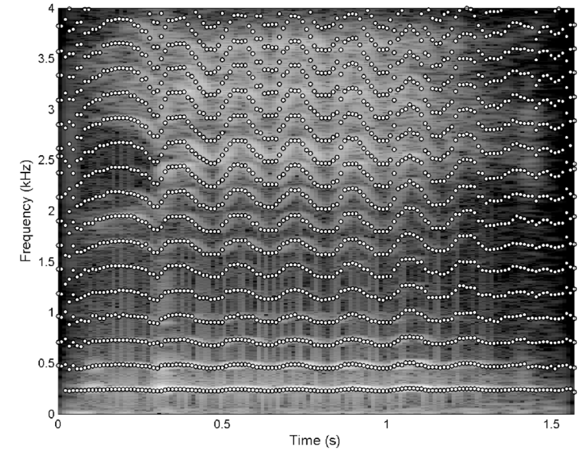
Mixture Component 1 Component 2

[Virtanen03] T. Virtanen. Sound Source Separation Using Sparse Coding with Temporal Continuity Objective. *Proc. Int. Computer Music Conference (ICMC),* Singapore, 2003.

[Virtanen04] T. Virtanen. Separation of Sound Sources by Convolutive Sparse Coding. *Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing (SAPA),* Jeju, Korea, 2004.

Sinusoidal Methods

- Sinusoidal Modeling: detection and tracking of the sinusoidal partial peaks on the spectrogram.



(Fig. from [Every06])

- Based on Auditory Scene Analysis (ASA) cues of good-continuation, common fate and smoothness of sinusoidal tracks.
- Overall, very good reduction of interfering sources, but moderate timbral quality.
    - Appropriate for Significance-Oriented applications
- [Virtanen&Klapuri02]: model of spectral smoothness of harmonic sounds
    - Based on basis decomposition of harmonic structures
    - Additive resynthesis of partial parameters

Mixture

Separated sources

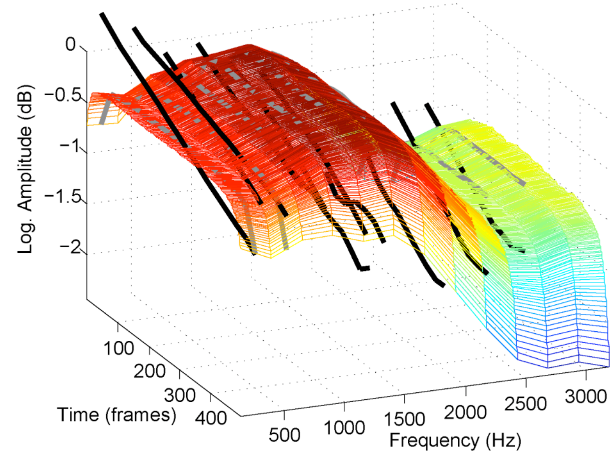
- [Every&Szymanski06]
    - Spectral subtraction instead of additive resynthesis

[Virtanen&Klapuri02] T. Virtanen and A. Klapuri. Separation of Harmonic Sounds Using Linear Models for the Overtone Series. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Orlando, USA, 2002.

[Every&Szymanski06] M. R. Every and J. E. Szymanski. Separation of Synchronous Pitched Notes by Spectral Filtering of Harmonics. *IEEE Trans. on Audio, Speech and Signal Processing.* Vol. 14(5), 2006.

Supervised Methods (1)

- Use of a training database to create a set of source models, each one modeling a specific instrument.
  - Better separation as a trade-off for generality.
- Supervised sinusoidal methods
  - [Burred&Sikora07]
    - The source models are compact descriptions of the spectral envelope and its temporal evolution.
    - The detailed temporal evolution allows to ignore harmonicity constraints, and thus separation of chords and inharmonic sounds is possible.



Separation of chords

Mixture

Separated sources

Inharmonic separation

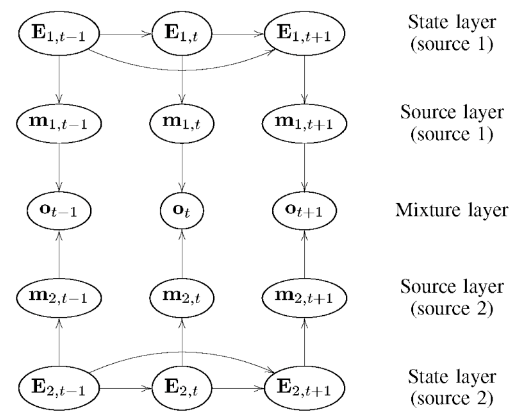
Mixture

Separated sources

[Burred&Sikora07] J.J. Burred and T. Sikora. Monaural Source Separation from Musical Mixtures Based on Time-Frequency Timbre Models. *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, Vienna, Austria, September 2007.

Supervised Methods (2)

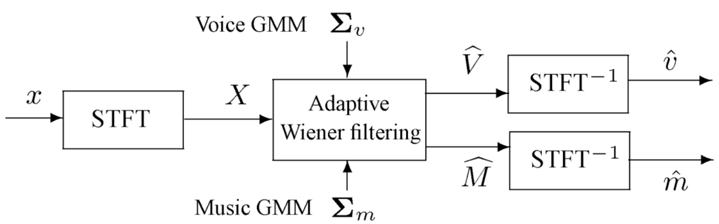
- Bayesian Networks
  - [Vincent06]
  - Multilayered model describing note probabilities (state layer), spectral decomposition (source layer) and spatial information (mixture layer).
  - Trained on a database of isolated notes.
  - Allows separation of sounds with reverb.



Mixture

Separated sources

- Learnt priors for Wiener-based separation
  - [Ozerov05]
  - Single-channel
  - HMM models of singing voice and accompaniment.



Mixture

Separated sources

[Vincent06] E. Vincent. Musical Source Separation Using Time-Frequency Source Priors. *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 14 (1), 2006.

[Ozerov05] A. Ozerov, O. Philippe, R. Gribonval and F. Bimbot. One Microphone Singing Voice Separation Using Source-Adapted Models. *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, USA, 2005.

Conclusions

- Still far from fully-general, audio-quality-oriented system.
- More realistic: significance oriented
  - Separation good enough to facilitate content analysis
- Methods based on adaptive models, time-frequency masking:
  - More realistic mixtures, but more artifacts and interferences
- Methods based on sinusoidal modeling:
  - More artificial timbre, but less interferences.
- Current polyphony limitations:
  - Mono signals: up to 3, 4 instruments
  - Stereo signals: up to 5, 6 instruments

Literature

- Very few overview materials on Musical Source Separation

- P. D. O´Grady, B. A. Pearlmutter and S. T. Rickard. **Survey of sparse and non-sparse methods in source separation**. International Journal of Imaging Systems and Technology, 15(1). 2005.

- E. Vincent, M. G. Jafari, S. A. Abdallah, M. D. Plumbley and M. E. Davies. **Model-based audio source separation**. Technical Report C4DM-TR-05-01, Queen Mary University, London, UK, 2006.

- T. Virtanen. **Unsupervised Learning Methods for Source Separation in Monaural Music Signals**. Chapter in A. Klapuri, M. Davy (Eds.), *Signal Processing Methods for Music Transcription*, Springer 2006.

- Stereo Audio Source Separation Evaluation Campaign:

  http://sassec.gforge.inria.fr