openBliSSART: Design and Evaluation of a Research Toolkit for Blind Source Separation in Audio Recognition Tasks

Felix Weninger[1], Alexander Lehmann[2], Björn Schuller[1]

[1]Institute for Human-Machine Communication, [2]Group of Applied Informatics / Cooperative Systems, Technische Universität München, Germany

ICASSP 2011 International Conference on Acoustics, Speech and Signal Processing, May 22–27, 2011, Prague, Czech Republic

Background and Motivation

- Manifold applications of Blind Audio Source Separation (BASS):
  - Music Information Retrieval / Transcription
  - Automatic Speech Recognition / Speech Enhancement
- openBliSSART: Focus on **monaural** separation by Non-Negative Matrix Factorization (NMF)
- Increasing amount of computational power available
  ⇝ NMF-based algorithms ready to be used in real-life applications
- openBliSSART: **First comprehensive open-source framework** for NMF
  - Useful for music and speech processing
  - Seamless integration with other audio toolkits
  - Highly efficient C++ implementation using BLAS and FFTW libraries

Implemented Algorithms

- Extract arbitrary number sources (*components*) from monophonic audio files
  - Spectrogram matrix $\mathbf{V} \in \mathbb{R}_+^{M \times N}$ from STFT decomposed as

$$\mathbf{V} = \mathbf{WH}$$

  by minimizing cost function $d(\mathbf{V}|\mathbf{WH})$ (Lee-Seung algorithms)
  - Reconstruction of R component spectrograms $\mathbf{V}^{(j)}$ e. g. by 'Wiener filter':

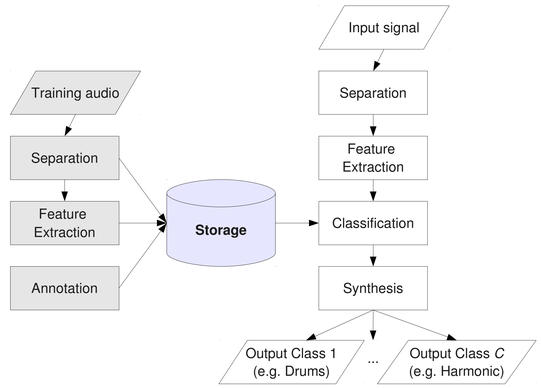
$$\mathbf{V}^{(j)} = \mathbf{V} \otimes \frac{\mathbf{w}^{(j)}\mathbf{h}^{(j)}}{\mathbf{WH}}$$

- Component classification by SVM
  - Useful, e. g., for instrument separation (1 instrument = multiple components)
  - Features: MFCC, noise-likeness, auto-correlation of activations, ...
- Supervised NMF (given $\mathbf{W}$) supported by component databases
- NMF extensions: Deconvolution (NMD), sparse priors in cost function, ...
- Optimized Euclidean NMF
  - Complexity determined by products of the form $\mathbf{W}^T\mathbf{WH}$
  - $\mathbf{W}^T(\mathbf{WH})$: $O(MNR)$ (NMF-ED-ov); $(\mathbf{W}^T\mathbf{W})\mathbf{H}$: $O(R^2(M+N))$ (NMF-ED-in)
  - openBliSSART: use $(\mathbf{W}^T\mathbf{W})\mathbf{H}$ iff $MN > R(M+N)$

preprocessing	algorithms	$d(\mathbf{V}\|\mathbf{WH})$	reconstruction
Mel filter	NMF	IS div.	default
Power spec.	NMD	KL div.	Wiener filtering
Sliding window		Eucl. dist.	comp. classif.
		+sparsity	

Use Cases

Unsupervised Separation with Classification

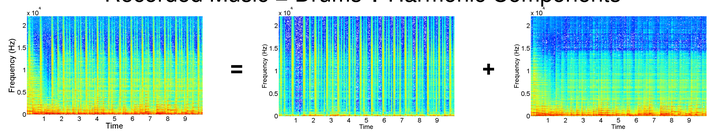


Source signals:

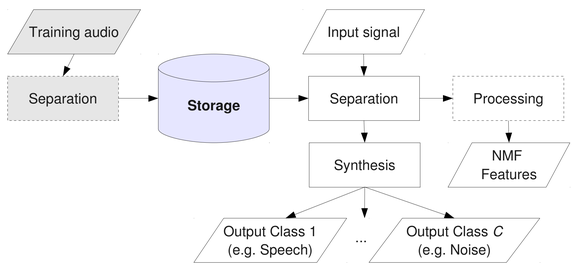
$$J_c = \{j : (\mathbf{w}^{(j)}, \mathbf{h}^{(j)}) \text{ classified as class } c\}, \mathbf{V}_c = \sum_{j \in J_c} \mathbf{V}^{(j)}.$$

Example: Drum Beat Separation

Recorded Music = Drums + Harmonic Components



(Semi-)Supervised Separation and Acoustic Feature Extraction

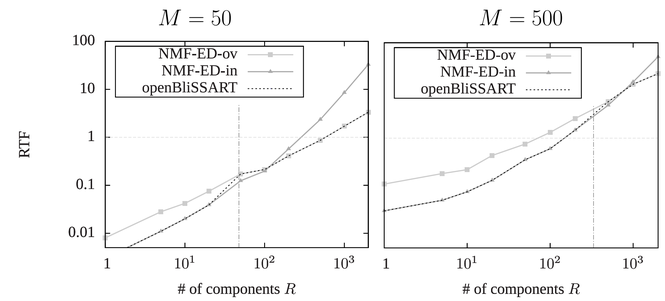


- Compute NMF basis $\mathbf{W}$ for signal class
- Use $\mathbf{W}$ in supervised NMF for filtering
- Feature extraction: $\mathbf{H}$ = activations of given spectra $\mathbf{W}$

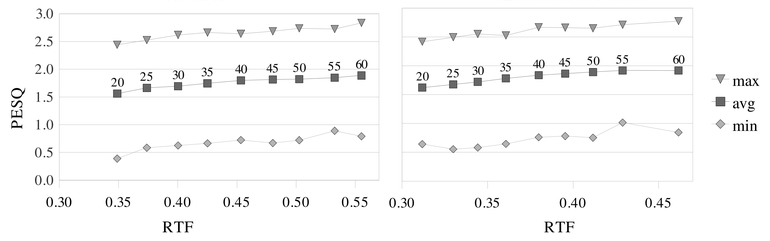
Evaluation

- RTFs on single 2.4 GHz Intel Q6600 core
- Supervised speaker separation: 12 pairs of male and female speakers from the TIMIT database

Optimized Euclidean NMF



Performance of Supervised Speaker Separation



Conclusions

- openBliSSART has led us to great success in various research:
  - Instrument separation
  - Noise-robust speech recognition
  - Detection of non-linguistic vocalizations in speech
- Effect of NMF optimizations could be demonstrated through benchmarks
- Source separation by NMF is real-time capable on desktop PC
- Future work: On-line / incremental audio processing

Software and audio examples available at `http://openblissart.github.com/openBliSSART`

This work was supported by the German Research Foundation (DFG) under the grant no. SCHU 2508/2-1 ("Non-Negative Matrix Factorization for Robust Feature Extraction in Speech Processing").