Abstract

This paper presents a new sparse representation for acoustic signals based on a mixing model defined in the complex-spectrum domain (where additivity holds):

$$F_{k,t} = \sum_{i=1}^{I} H_k^{(i)} U_t^{(i)} e^{\mathrm{j}\phi_{k,t}^{(i)}}.$$

This allows us to extract recurring patterns of magnitude spectra that underlie observed complex spectra and the phase estimates of constituent signals. An efficient iterative algorithm is derived, which reduces under a particular condition to the multiplicative update algorithm for non-negative matrix factorization (NMF) developed by Lee et al.

Even though the terms “complex” and “non-negative” may seem to conflict with each other and the decomposition is no longer a matrix factorization, we named this framework “Complex NMF”, because it shares with NMF the ability to generate non-negative matrices $\boldsymbol{H}$ and $\boldsymbol{U}$, while the input matrix $\boldsymbol{Y}$ is assumed to be a complex matrix and the algorithm also generates a third-rank complex-valued tensor $Z_{k,t}^{i} = e^{\mathrm{j}\phi_{k,t}^{(i)}}$.

1. Introduction

◆Sparse representation

Given a set of observed vectors $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_T$, the goal is to find a set of basis vectors $\boldsymbol{h}_1,\cdots,\boldsymbol{h}_I$ such that any observations can be succinctly represented as a linear combination of a small number of “active” bases.

$$\boldsymbol{y}_t \simeq \sum_i u_{i,t}\boldsymbol{h}_i \quad (t=1,\cdots,T) \Longrightarrow \boldsymbol{Y} \simeq \boldsymbol{H}\boldsymbol{U} \quad \left\{ \begin{array}{l} \boldsymbol{Y} \equiv (\boldsymbol{y}_1,\cdots,\boldsymbol{y}_T) \\ \boldsymbol{H} \equiv (\boldsymbol{h}_1,\cdots,\boldsymbol{h}_I) = (h_{k,i})_{K\times I} \\ \boldsymbol{U} \equiv (u_{i,t})_{I\times T} \end{array} \right.$$

sparse

▪Sparse coding (SC) (Olshausen et al., 1996)[1]

$$\text{minimize} \quad \|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{U}\|_F^2 + \lambda\,\mathrm{sparse}(\boldsymbol{U}) \quad \text{w.r.t.} \quad \boldsymbol{H},\boldsymbol{U}$$

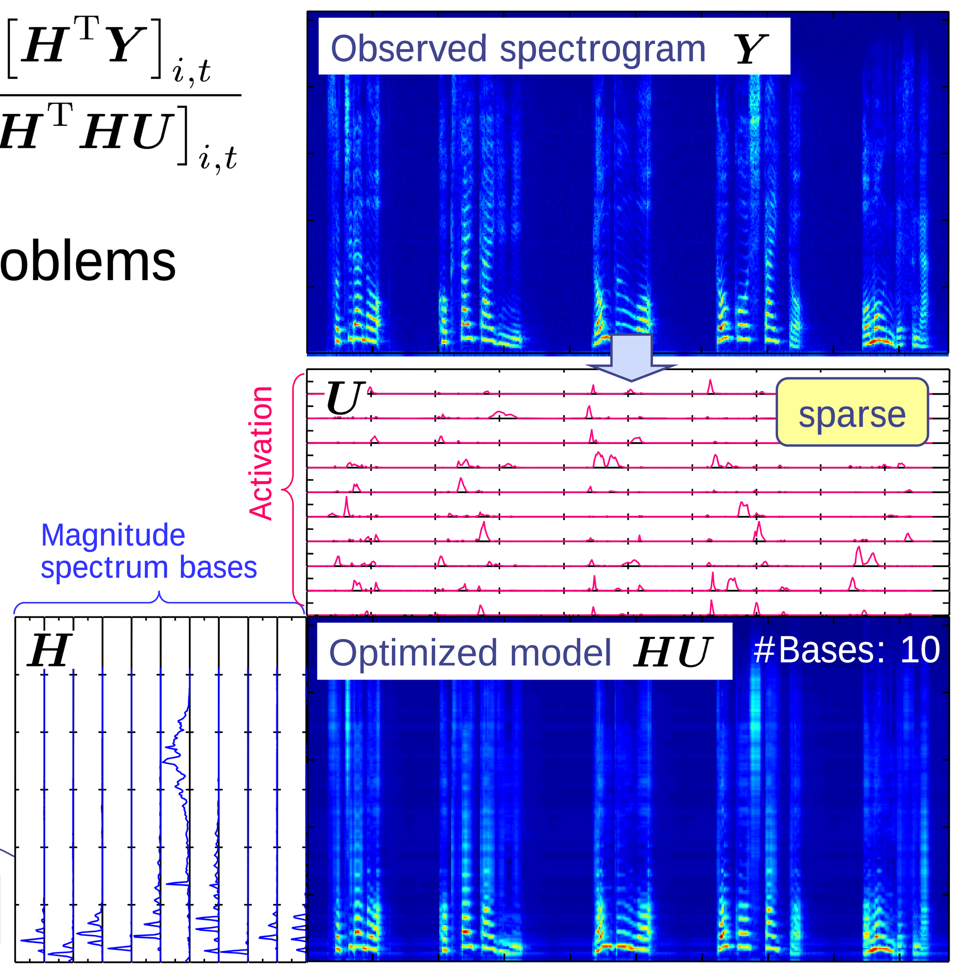
▪Non-negative matrix factorization (NMF) (Lee et al., 1999)[2]

$$\text{minimize} \quad \|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{U}\|_F^2 \quad \text{(or other measure)}$$
$$\text{subject to} \quad \forall_{k,i}\; h_{k,i}\ge 0,\; \forall_{i,t}\; u_{i,t}\ge 0$$

$\boldsymbol{U}$ becomes sparse as a result of a side effect caused by the non-negativity constraints

[multiplicative update algorithm] → guarantees convergence to a stationary point

$$H_{k,i} \leftarrow H_{k,i}\frac{[\boldsymbol{Y}\boldsymbol{U}^{\mathrm{T}}]_{k,i}}{[\boldsymbol{H}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}]_{k,i}} \quad U_{i,t} \leftarrow U_{i,t}\frac{[\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Y}]_{i,t}}{[\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{U}]_{i,t}}$$



◆NMF applied to audio processing problems

→ parts-based decomposition of magnitude spectrogram

- Music transcription [3,4]
- 1ch source separation [5]
- Speech denoising
- Bandwidth expansion
- Audio classification
- Dereverberation [7]

etc.

Recurring spectral patterns

2. Research objective

◆Develop new framework that concurrently offers advantages of NMF-based and Sparse coding/Semi-NMF*-based acoustic modeling

* Semi-NMF (Ding et al., 2006) (Le Roux et al., 2008)[11]

Allows to project all signals that have the same spectral shape onto a single basis → variety of acoustic phenomena with compact set of spectral bases

▪ In NMF framework,

- ☺ Assume mixing model defined in magnitude domain.
- ☺ Model can be optimized by efficient algorithm.
- ☹ Magnitude spectra are assumed additive. →inaccurate modeling
- ☹ Phase information are discarded. →limits the range of application

NMF model

Observed magnitude spectrogram

$$|\boldsymbol{Y}| \simeq \boldsymbol{H}\boldsymbol{U} \Longrightarrow |Y_{k,t}| \simeq \sum_i H_k^{(i)} U_t^{(i)} \quad (\forall_{k,i} H_k^{(i)}\ge 0,\; \forall_{i,t} U_t^{(i)}\ge 0)$$

frequency, time; i th magnitude spectrum basis; activation of i th basis

magnitude spectra are non-additive ☹

Static magnitude spectrum templates
Time-varying activation coefficients

SC/Semi-NMF model

Observed complex spectrogram

$$\boldsymbol{Y} \simeq \boldsymbol{V}\boldsymbol{U} \Longrightarrow Y_{k,t} \simeq \sum_i V_k^{(i)} U_t^{(i)} \quad (\forall_{k,i} V_k^{(i)}\in\mathbb{C},\; \forall_{i,t} U_t^{(i)}\in\mathbb{R} \text{ or } \forall_{i,t} U_t^{(i)}\ge 0)$$

$V_k^{(i)} = H_k^{(i)} e^{\mathrm{j}\varphi_k^{(i)}}$

frequency, time; i th complex spectrum basis; activation of i th basis

complex spectra are additive ☺

phase coherence between frequency components can be easily destroyed ☹

Static complex spectrum templates
Time-varying activation coefficients

3. Complex NMF (model)

◆Model

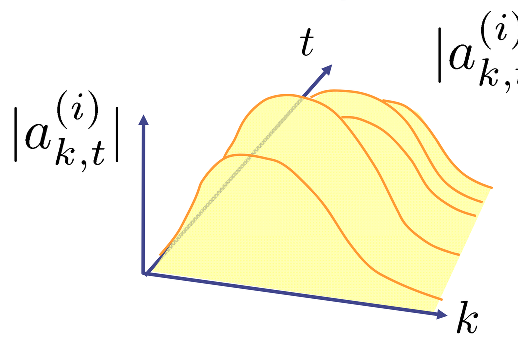
Assume that the short-term Fourier transform (STFT) of an acoustic signal is composed of I complex-valued elements:

$$F_{k,t} = \sum_{i=1}^{I} a_{k,t}^{(i)}.$$

frequency, time

Each constituent, $a_{k,t}^{(i)}$, is assumed to have a magnitude spectrum which is constant up to the gain over time:

$$|a_{k,t}^{(i)}| = H_k^{(i)} U_t^{(i)} \quad (\forall_{k,i} H_k^{(i)} \ge 0, \forall_{i,t} U_t^{(i)} \ge 0),$$



Gain at frame t

Normalized magnitude spectrum template: $\forall_i \sum_k H_k^{(i)} = 1$

and a time-varying phase spectrum:

$$\arg(a_{k,t}^{(i)}) = \phi_{k,t}^{(i)}.$$

Proposed model — Complex NMF model

Observed complex spectrogram

$$Y_{k,t} \simeq F_{k,t} = \sum_i H_k^{(i)} U_t^{(i)} e^{\mathrm{j}\phi_{k,t}^{(i)}}$$
$$(\forall_{k,i} H_k^{(i)} \ge 0,\; \forall_{i,t} U_t^{(i)} \ge 0)$$

frequency, time; i th magnitude spectrum basis; activation of i th basis

Cannot be expressed in matrix notation!
→ new class of sparse representation model

Static magnitude spectrum templates
Time-varying gains & phase spectra

◆Objective function

Given a set of observations $Y=\{Y_{k,t}|1\le k\le K, 1\le t\le T\}$, the goal is to find the optimal estimate of $\theta=\{H_k^{(i)},U_t^{(i)},\phi_{k,t}^{(i)}|1\le k\le K, 1\le t\le T, 1\le i\le I\}$.

(1) Model reconstruction accuracy

Generative model: $Y_{k,t} = F_{k,t} + \epsilon_{k,t}$ where $\epsilon_{k,t} \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2)$

($\epsilon_{k,t}$ and $\epsilon_{k',t'}$ are independent if $(k,t)\ne(k',t')$)

$$P(Y|\theta) = \prod_{k,t}\frac{1}{\pi\sigma^2}\exp\left(-\frac{|Y_{k,t}-F_{k,t}|^2}{\sigma^2}\right) \quad \text{(likelihood)}$$

(2) Sparsity of basis activations

As we want to achieve as parsimonious a representation as possible, we would like the $U_t^{(i)}$'s to be sparse.

This basically means that any observed complex spectrum should be well represented using only a few active magnitude spectrum bases each of which is paired with an arbitrary phase spectrum.

$$P(\theta) \propto \prod_{i,t}\frac{1}{2\Gamma(1+\frac{1}{p})b}\exp\left(-\frac{|U_t^{(i)}|^p}{b^p}\right) \quad \text{(prior)}$$

Generalized Gaussian distribution (GGD)

→ promotes sparsity when $0<p<2$ and the norm of H is bounded

MAP (Maximum A Posteriori) estimation problem leads to...

$$\text{minimize} \quad f(\theta) \equiv \sum_{k,t}|Y_{k,t}-F_{k,t}|^2 + 2\lambda\sum_{i,t}|U_{i,t}|^p$$
$$\text{subject to} \quad \sum_k H_k^{(i)} = 1 \quad (i=1,\cdots,I)$$

4. Complex NMF (algorithm)

◆Utilizing auxiliary function concept

Definition.

$G^+(\theta,\bar{\theta})$ is an *auxiliary function* of the objective function $G(\theta)$ if it satisfies $G(\theta) = \min_{\bar{\theta}} G^+(\theta,\bar{\theta})$.

Theorem 1 (Auxiliary function method).

$G(\theta)$ is non-increasing under the updates $\theta \leftarrow \operatorname{argmin}_{\theta} G^+(\theta,\bar{\theta})$ and $\bar{\theta} \leftarrow \operatorname{argmin}_{\bar{\theta}} G^+(\theta,\bar{\theta})$.

Proof

Assume that we set $\boldsymbol{\theta}$ at an arbitrary value θ_ℓ. Let $\bar{\theta}_{\ell+1} = \operatorname{argmin}_{\bar{\theta}} G^+(\theta_\ell,\bar{\theta})$ and $\theta_{\ell+1} = \operatorname{argmin}_{\theta} G^+(\theta,\bar{\theta}_{\ell+1})$. It is obvious that $G(\theta_\ell) = G^+(\theta_\ell,\bar{\theta}_{\ell+1})$. We deduce from $\theta_{\ell+1} = \operatorname{argmin}_{\theta} G^+(\theta,\bar{\theta}_{\ell+1})$ that $G^+(\theta_\ell,\bar{\theta}_{\ell+1}) \geq G^+(\theta_{\ell+1},\bar{\theta}_{\ell+1})$. By definition, from $G^+(\theta_{\ell+1},\bar{\theta}_{\ell+1}) \geq G(\theta_{\ell+1})$ we verify that $G(\theta_\ell) \geq G(\theta_{\ell+1})$.

Theorem 2 (Auxiliary function for Complex NMF).

$$f^+(\theta,\bar{\theta}) \equiv \sum_{k,t}\sum_{i} \frac{1}{\beta_{k,t}^{(i)}} \left| \bar{Y}_{k,t}^{(i)} - H_k^{(i)} U_t^{(i)} e^{j\phi_{k,t}^{(i)}} \right|^2$$
$$+ \lambda \sum_{t}\sum_{i} \left\{ p|\bar{U}_t^{(i)}|^{p-2} U_t^{(i)2} + (2-p)|\bar{U}_t^{(i)}|^p \right\},$$

with $\bar{\theta} = \{\bar{Y}_{k,t}^{(i)}, \bar{U}_t^{(i)} | 1 \leq k \leq K, 1 \leq t \leq T, 1 \leq i \leq I\}$, is an auxiliary function for $f(\theta)$, if

$$\forall_{k,t} \sum_i \bar{Y}_{k,t}^{(i)} = Y_{k,t}, \forall_{k,t} \sum_i \beta_{k,t}^{(i)} = 1, \forall_{i,k,t} \beta_{k,t}^{(i)} \in (0,1), p \in (0,2].$$

Proof

(1) By differentiating the Lagrangian

$$\sum_i \frac{1}{\beta_{k,t}^{(i)}} \left| \bar{Y}_{k,t}^{(i)} - H_k^{(i)} U_t^{(i)} e^{j\phi_{k,t}^{(i)}} \right|^2 + \gamma \left(\sum_i \bar{Y}_{k,t}^{(i)} - Y_{k,t} \right)$$

partially w.r.t. $\bar{Y}_{k,t}^{(i)}$ and setting it at zero, we determine the minimum value $|Y_{k,t} - F_{k,t}|^2$.

(2) $p|\bar{U}|^{p-2}U^2 + (2-p)|\bar{U}|^p$ amounts to a convex quadratic function that is tangent to $|U|^p$ at argument $U = \pm\bar{U}$.

◆Update formulae

Non-negativities are preserved if we start with non-negative initial conditions.

Step 1) $\theta \leftarrow \operatorname{argmin}_{\theta} f^+(\theta,\bar{\theta})$

$$H_k^{(i)} \leftarrow \frac{\sum_t \frac{U_t^{(i)}}{\beta_{k,t}^{(i)}} \mathrm{Re}\left[\bar{Y}_{k,t}^{(i)*} e^{j\phi_{k,t}^{(i)}} \right]}{\sum_t \frac{U_t^{(i)2}}{\beta_{k,t}^{(i)}}}, \quad U_t^{(i)} \leftarrow \frac{\sum_k \frac{H_k^{(i)}}{\beta_{k,t}^{(i)}} \mathrm{Re}\left[\bar{Y}_{k,t}^{(i)*} e^{j\phi_{k,t}^{(i)}} \right]}{\sum_k \frac{H_k^{(i)2}}{\beta_{k,t}^{(i)}} + \lambda p \left| \bar{U}_t^{(i)} \right|^{p-2}}, \quad e^{j\phi_{k,t}^{(i)}} \leftarrow \frac{\bar{Y}_{k,t}^{(i)}}{\left| \bar{Y}_{k,t}^{(i)} \right|}$$

$$H_k^{(i)} \leftarrow H_k^{(i)} \Big/ \sum_n H_k^{(n)}$$ (Projection onto the constraint space)

Step 2) $\bar{\theta} \leftarrow \operatorname{argmin}_{\bar{\theta}} f^+(\theta,\bar{\theta})$

$$\bar{Y}_{k,t}^{(i)} \leftarrow H_k^{(i)} U_t^{(i)} e^{j\phi_{k,t}^{(i)}} + \beta_{k,t}^{(i)} (Y_{k,t} - F_{k,t}), \quad \bar{U}_t^{(i)} \leftarrow U_t^{(i)}$$

◆Condition for equivalence to NMF with Frobenius norm criterion

We assume a particular situation: $\forall_i \; e^{j\phi_{k,t}^{(i)}} = \frac{Y_{k,t}}{|Y_{k,t}|}$.

Update formula for $\bar{Y}_{k,t}^{(i)}$: $\bar{Y}_{k,t}^{(i)} \leftarrow Y_{k,t} \left[\frac{H_k^{(i)} U_t^{(i)}}{|Y_{k,t}|} + \beta_{k,t}^{(i)} \left(1 - \frac{\sum_n H_k^{(n)} U_t^{(n)}}{|Y_{k,t}|} \right) \right]$.

Update formula for $H_k^{(i)}$: $H_k^{(i)} \leftarrow \frac{\sum_t \left[H_k^{(i)} U_t^{(i)2} \Big/ \beta_{k,t}^{(i)} + U_t^{(i)} \left(|Y_{k,t}| - \sum_n H_k^{(n)} U_t^{(n)} \right) \right]}{\sum_t U_t^{(i)2} \Big/ \beta_{k,t}^{(i)}}$.

$\beta_{k,t}^{(i)}$ is a parameter that can be chosen arbitrarily subject to $\forall_{k,t} \sum_i \beta_{k,t}^{(i)} = 1, \forall_{i,k,t} \beta_{k,t}^{(i)} \in (0,1)$.

What if we update it at each iteration by $\beta_{k,t}^{(i)} \leftarrow \frac{H_k^{(i)} U_t^{(i)}}{\sum_n H_k^{(n)} U_t^{(n)}}$?

$$H_k^{(i)} \leftarrow H_k^{(i)} \frac{\sum_t U_t^{(i)} |Y_{k,t}|}{\sum_t U_t^{(i)} \sum_n H_k^{(n)} U_t^{(n)}}$$

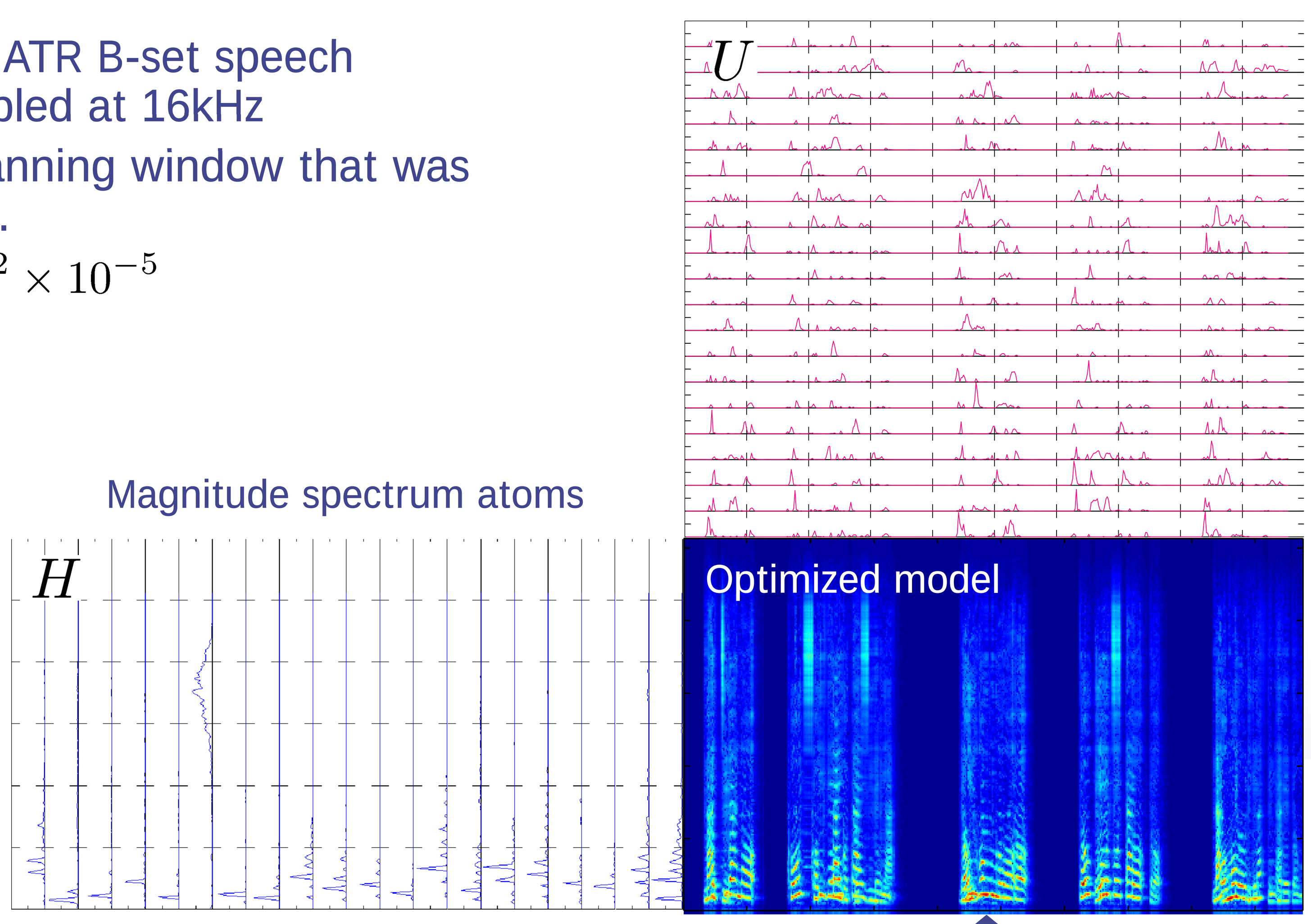
Reduces to Lee's multiplicative update formula for Frobenius norm criterion!

5. Experiments

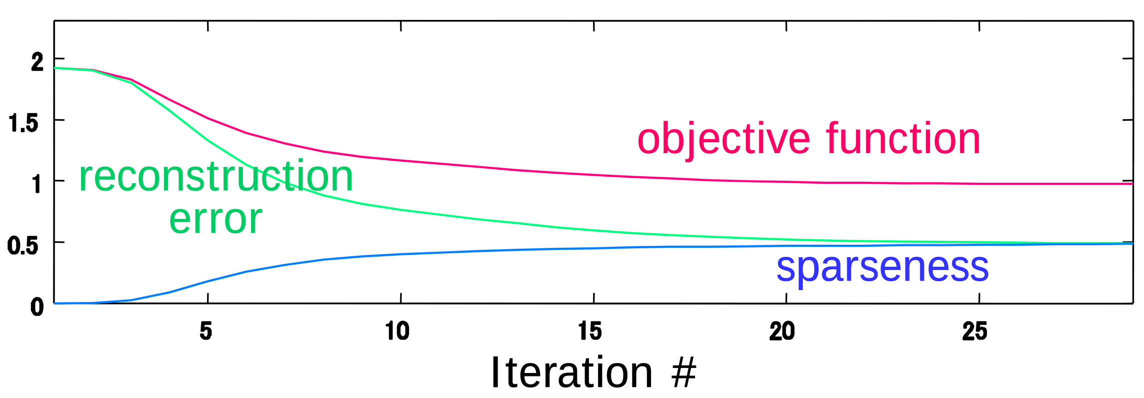
◆Conditions

- Speech data: excerpted from ATR B-set speech database, monaural and sampled at 16kHz
- STFT was computed using Hanning window that was 32ms long with 16ms overlap.
- $p = 1.2, \;\; \lambda = \sum_{k,t} |Y_{k,t}|^2 / I^{1-p/2} \times 10^{-5}$
- Iteration #: 30
- Initialization: random
- $I = 30$

The aim of the experiments was to determine whether Complex NMF has, similarly to NMF, the effect of extracting recurring spectral patterns underlying the observed audio data.

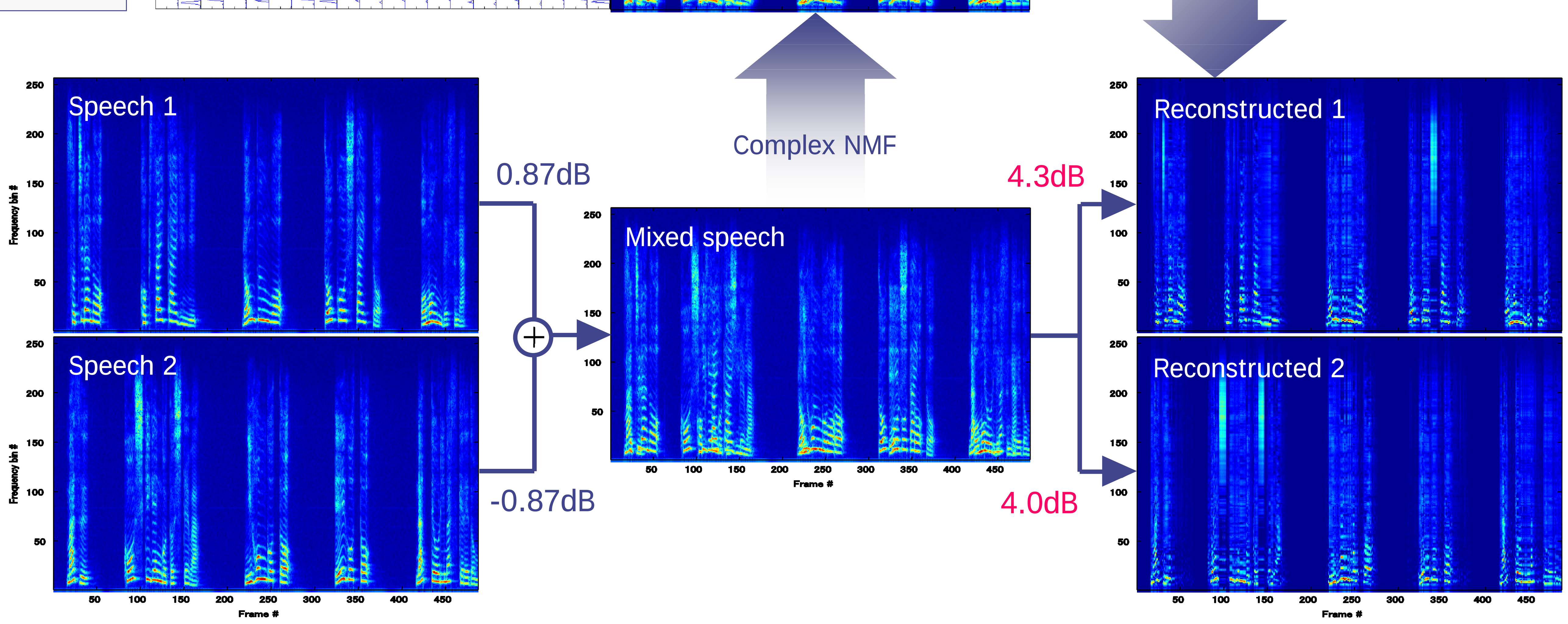


◆Convergence of Complex NMF algorithm



How did we obtain the separate signals?

The magnitude spectrum atom closest to the true spectrum was selected for each frame, and the framewise signals, each of which we constructed using the selected atom and the corresponding activation coefficient and phase spectrum, were concatenated to synthesize the whole signal stream.



References cited in the paper

[1] B.A. Olshausen and D.J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature, **381**, 607–609, 1996.
[2] D.D. Lee and H.S. Seung, "Algorithms for nonnegative matrix factorization," Proc. NIPS'00, 556–562, 2000.
[3] M.D. Plumbley, S.A. Abdallah, J.P. Bello, M.E. Davies, G. Monti and M.B. Sandler, "Automatic music transcription and audio source separation," Cybernetics and Systems, **33**(6), 603–627, 2002.
[4] P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for music transcription," Proc. WASPAA'03, 177–180, 2003.
[5] T. Virtanen, "Sound source separation using sparse coding with temporal continuity objective," Proc. ICMC'03, 231–234, 2003.
[6] S.A. Abdallah, M.D. Plumbley, "Unsupervised analysis of polyphonic music using sparse coding," IEEE Trans. NN, **17**(1), 179–196, 2006.
[7] H. Kameoka T. Nakatani and T. Yoshioka, "Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms," Proc. ICASSP'09, to appear.
[8] R.M. Parry and I. Essa, "Phase-Aware Non-negative Spectrogram Factorization," Proc. ICA'07, 536–543, 2007.
[9] C. Fevotte, N. Bertin and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence with application to music analysis," Technical Report TELECOM ParisTech 2008D006, 2008.
[10] T. Blumensath and M.E. Davies, "Sparse and shift-invariant representations of music," IEEE Trans. ASLP, **14**(1), 50–57, 2006.
[11] J. Le Roux, A. de Cheveigne and L.C. Parra, "Adaptive template matching with shift-invariant semi-NMF," Proc. NIPS'08, 2008.
[12] A.T. Cemgil, "Bayesian inference for nonnegative matrix factorisation models," Technical Report CUED/F-INFENG/TR.609, University of Cambridge, 2008.