Heterogeneous Convolutive Non-negative Sparse Coding

Dong Wang

2012-08-27

Non-negative matrix factorization

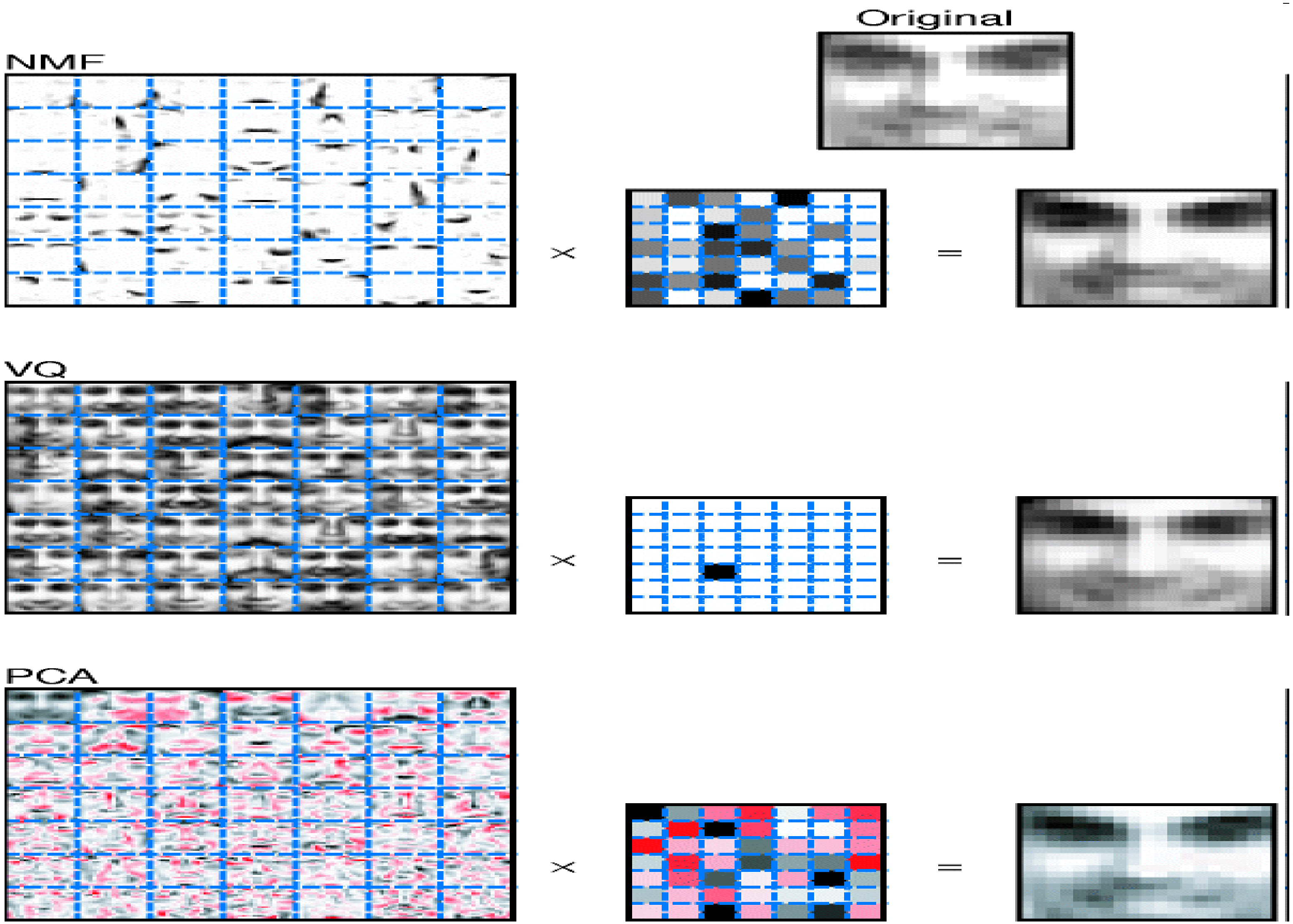
- X: signal, W:bases, H: coefficients

$$L(W, H) = | X - WH |_2 \qquad X_{i,j}, W_{i,j}, H_{i,j} \geq 0,$$

$$\widehat{W}, \widehat{H} = \arg\min_{W,H} L(W, H)$$

1. http://en.wikipedia.org/wiki/Non-negative_matrix_factorization
2. Daniel D. Lee and H. Sebastian Seung (1999). "Learning the parts of objects by non-negative matrix factorization". *Nature* **401** (6755): 788–791

Non-negative matrix factorization



Sparse NMF

$$L(W,H) = | X - WH |_2 + \lambda | H | + \tau | W |$$

$$X_{i,j}, W_{i,j}, H_{i,j}, \lambda, \gamma \geq 0$$

$$\widehat{W}, \widehat{H} = \arg\min_{W,H} L(W,H)$$

- P. Hoyer, “Non-negative matrix factorization with sparseness constraints,” Machine Learning Research, vol. 5, pp. 1457–1469, 2004.

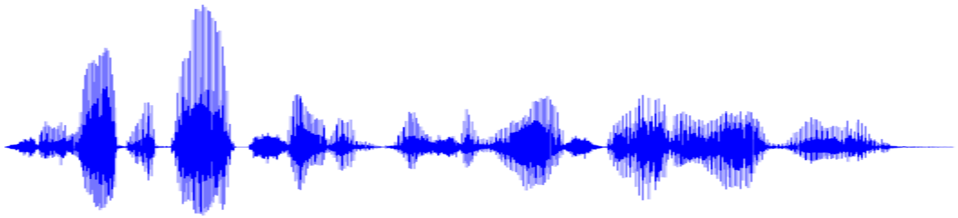
Convolutive NMF

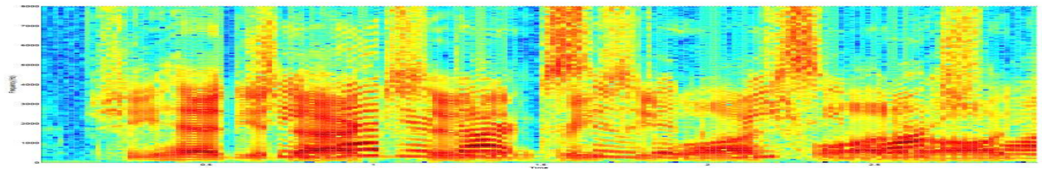
$$L(W,H) = | X - \tilde{X} |_2 \qquad X_{i,j} \geq 0,$$

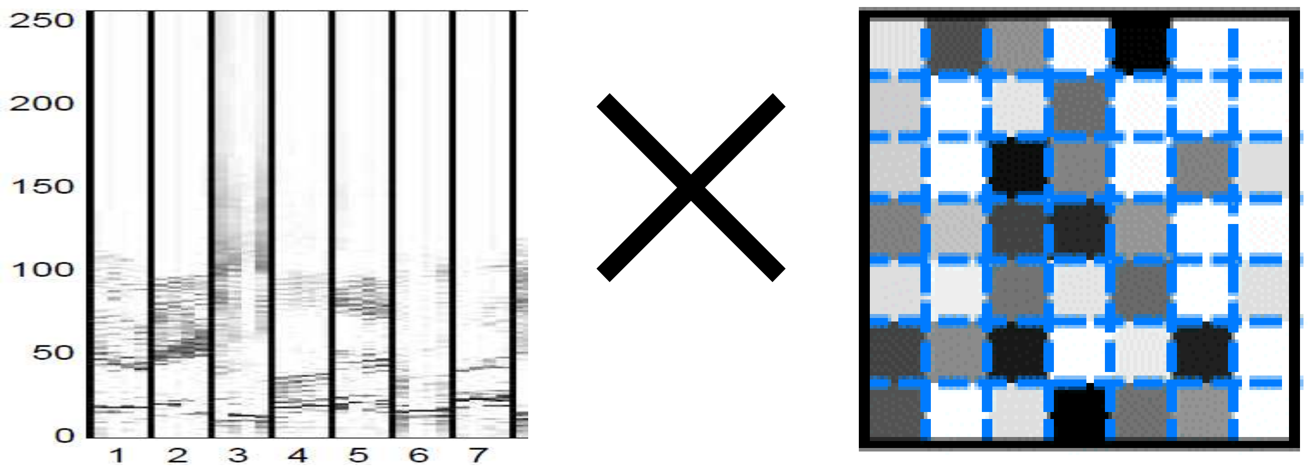
$$\tilde{X} = \sum_{p=0}^{P-1} W_p \overset{p\rightarrow}{H} \qquad W_{p;i,j}, H_{i,j} \geq 0$$

W. Wang, A. Cichocki, and J. A. Chamber, “A multiplicative algorithm for convolutive non-negative matrix factorization based on squared Euclidean distance,” IEEE Trans. on Signal Process., vol. 57, no. 5, pp. 2858–2864, 2009.

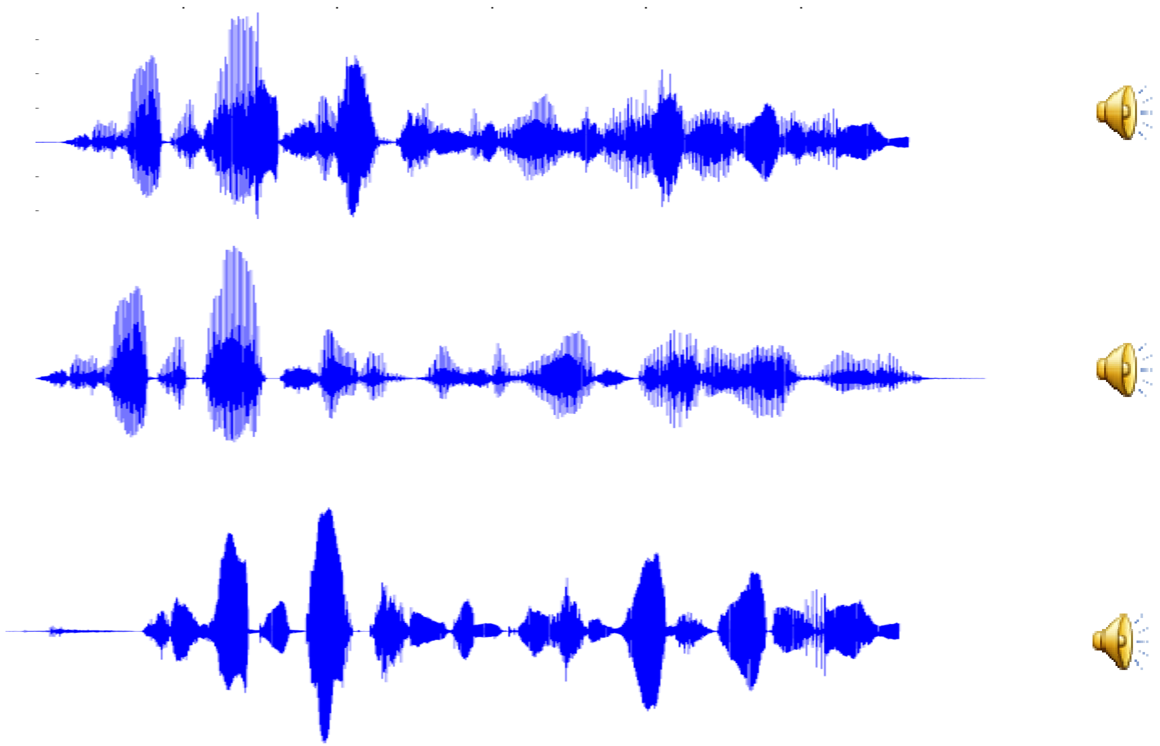
Convolutive NMF for speech signals





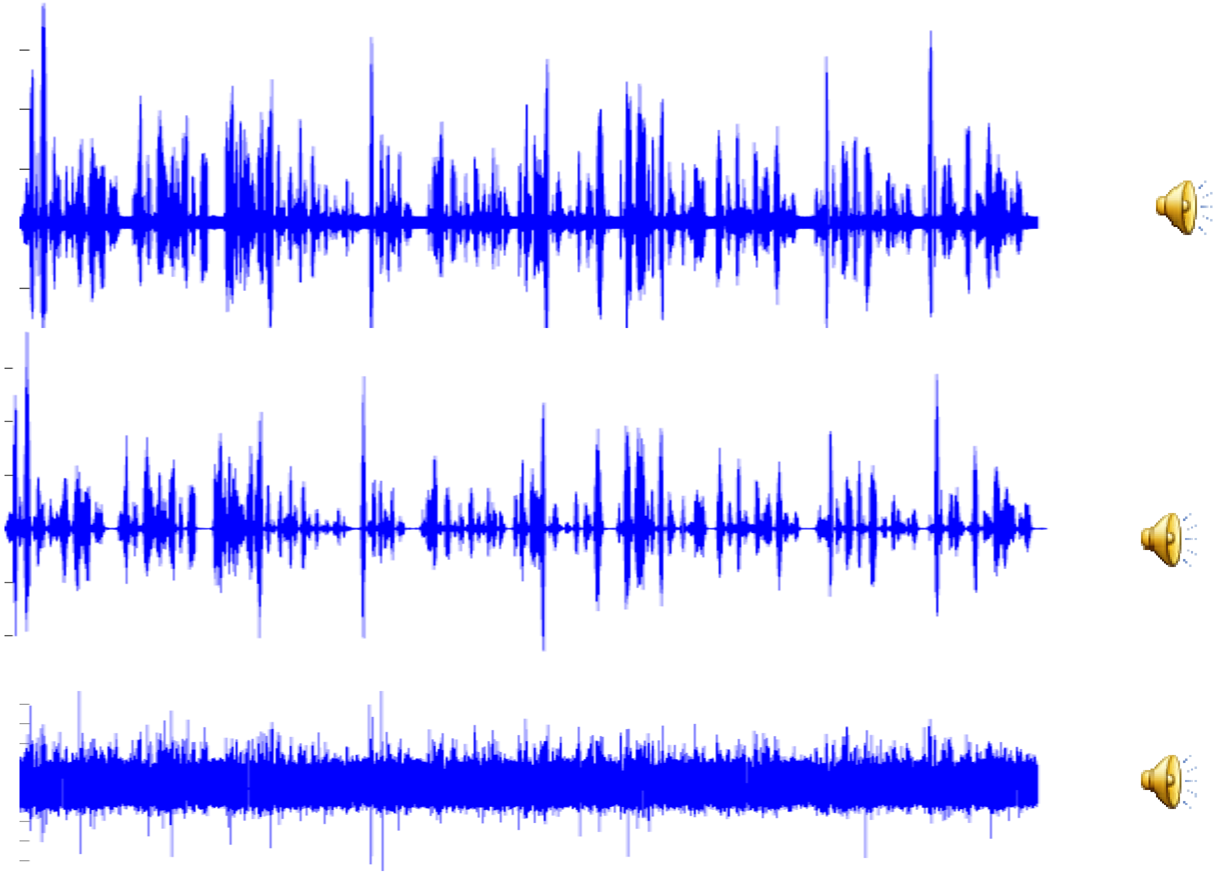


Convolutive NMF for speech separation



P. Smaragdis, “Convolutive speech bases and their application to supervised speech separation,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 1, pp. 1–12, January 2007.

Convolutive NMF for denoising

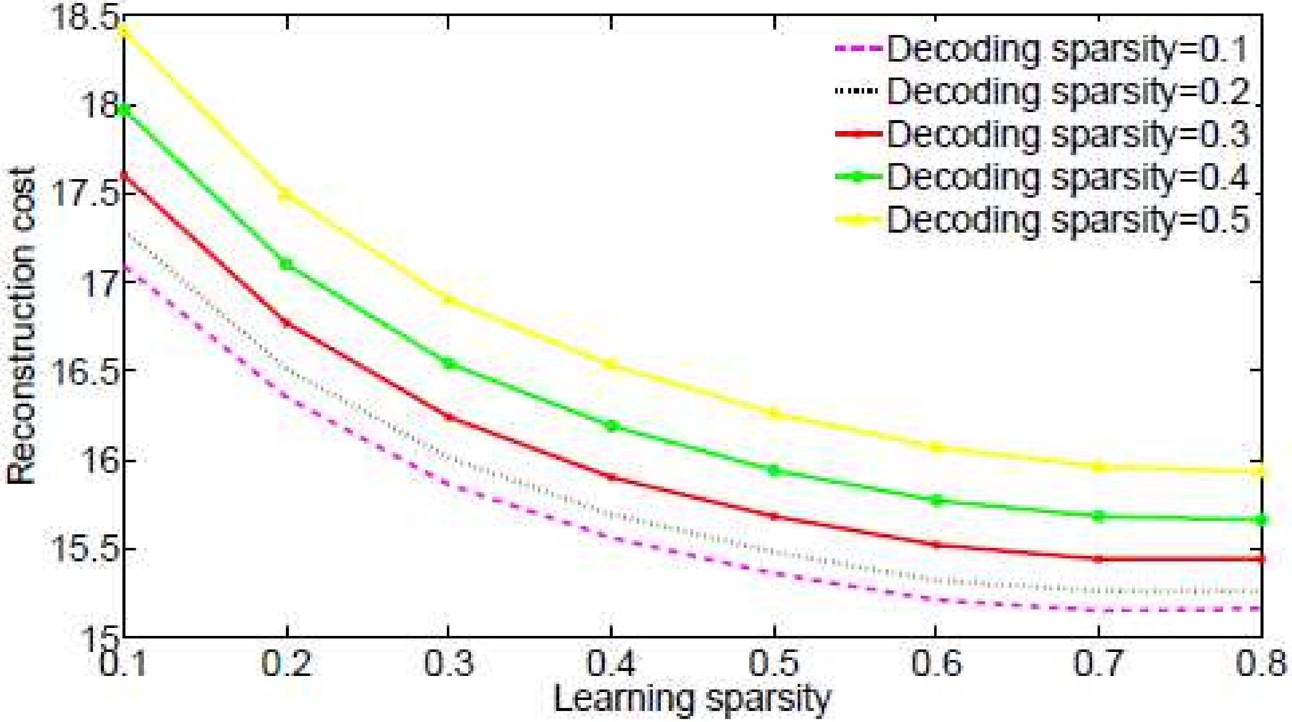


Convolutive non-negative sparse coding (CNSC)

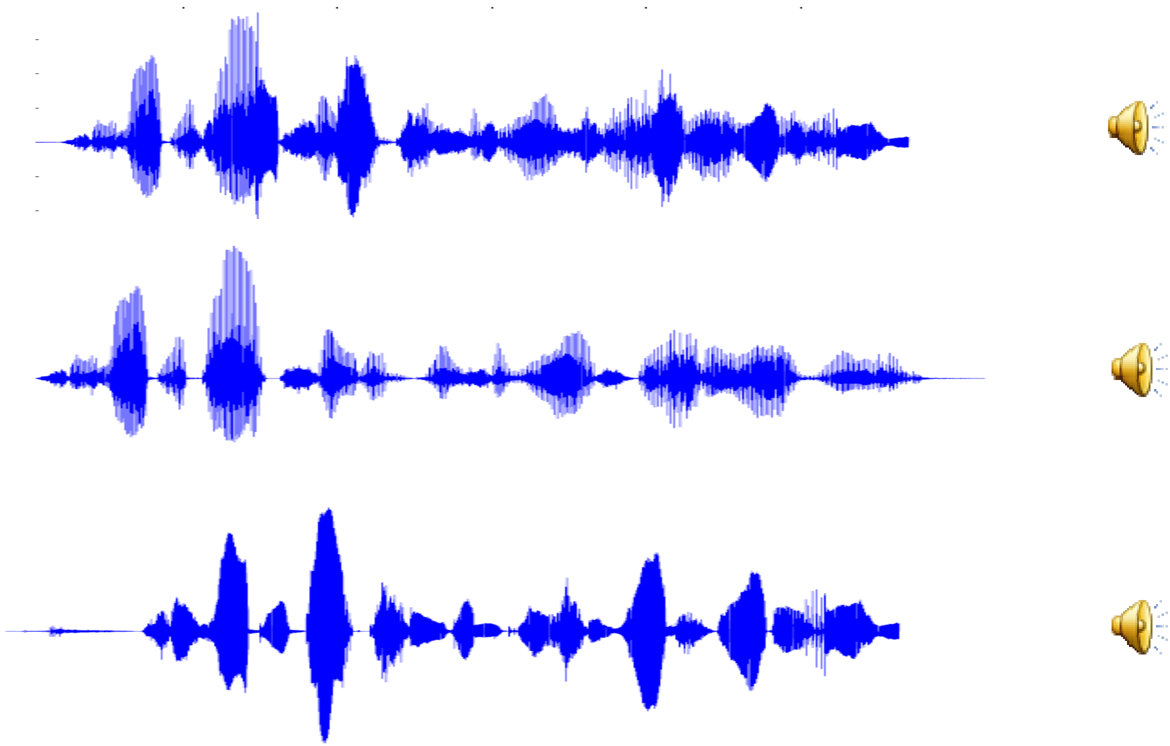
$$L(W,H) = | X - \tilde{X} |_2 + \lambda | H | \qquad X_{i,j}, \lambda, H_{i,j} \geq 0,$$

$$\tilde{X} = \sum_{p=0}^{P-1} W_p \overset{p\rightarrow}{H} \qquad W_{p;i,j}, H_{i,j} \geq 0$$

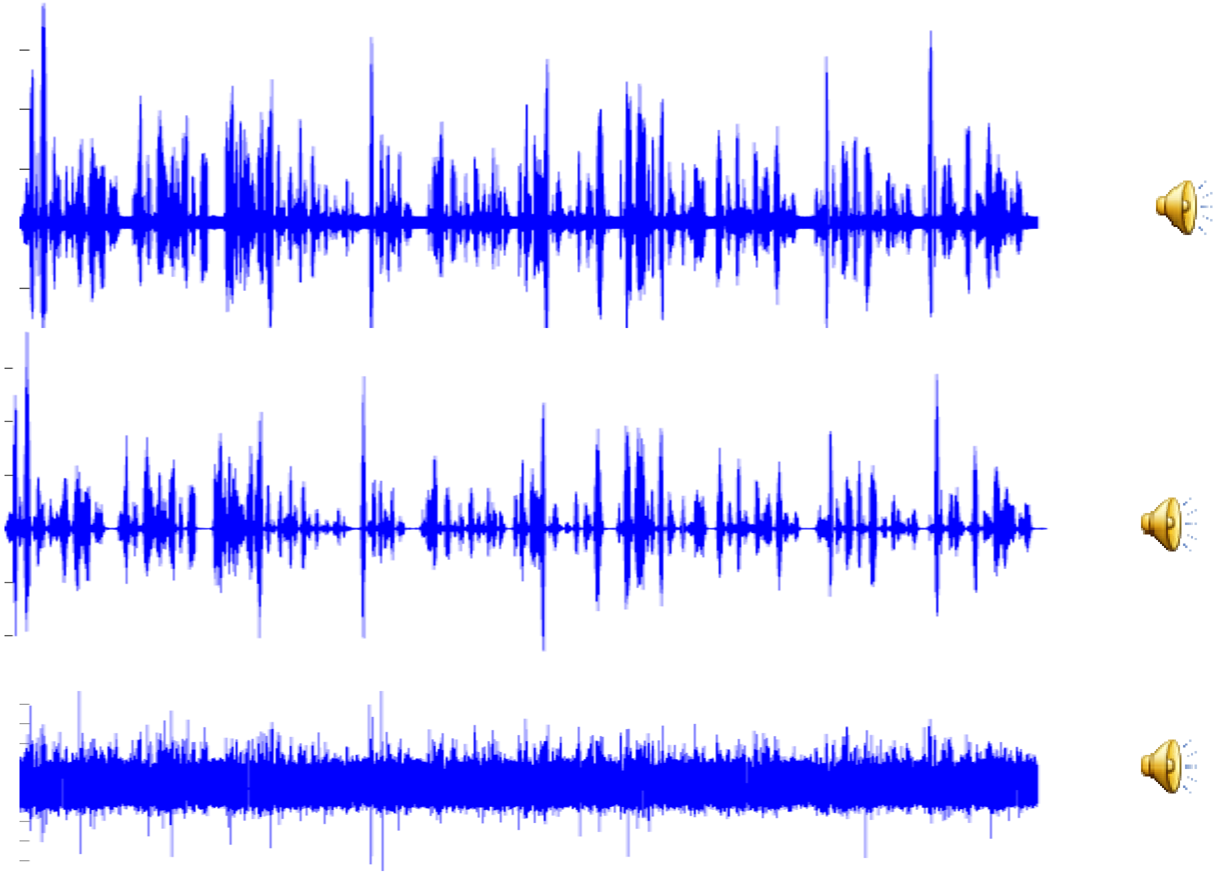
Convolutive non-negative sparse coding (CNSC)



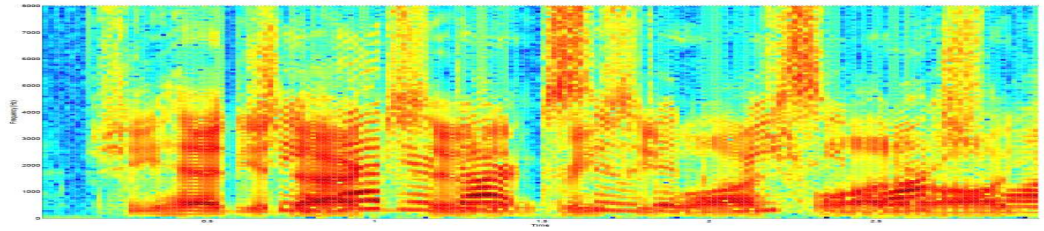
CNSC for speech separation

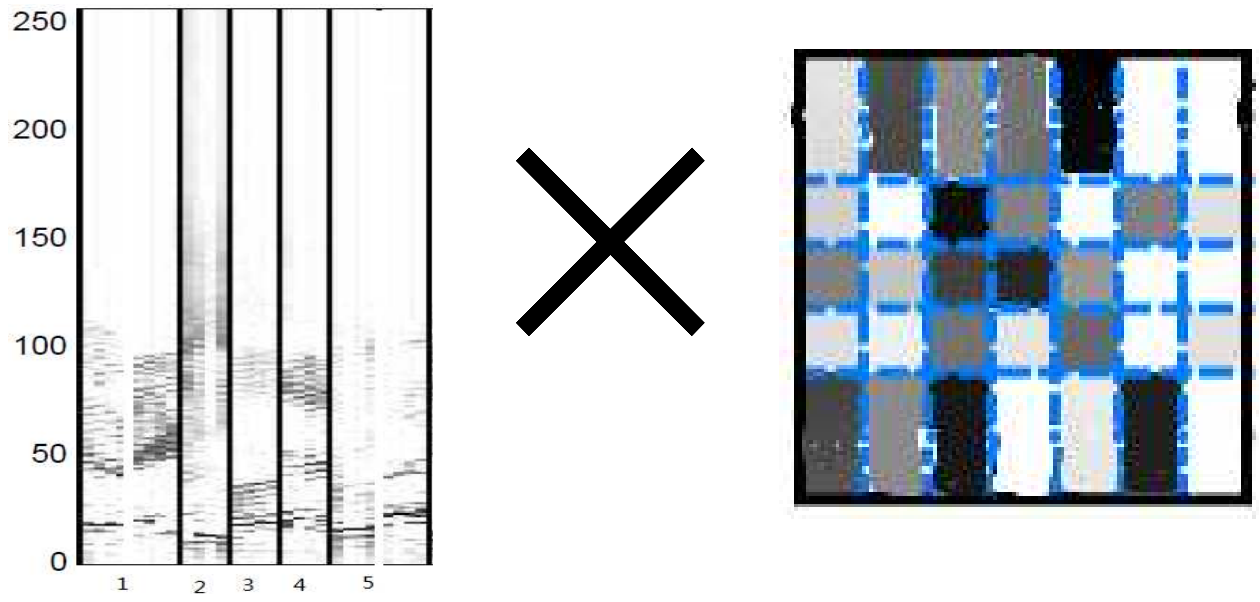


CNSC for denoising



Heterogeneous CNSC





Heterogeneous CNSC

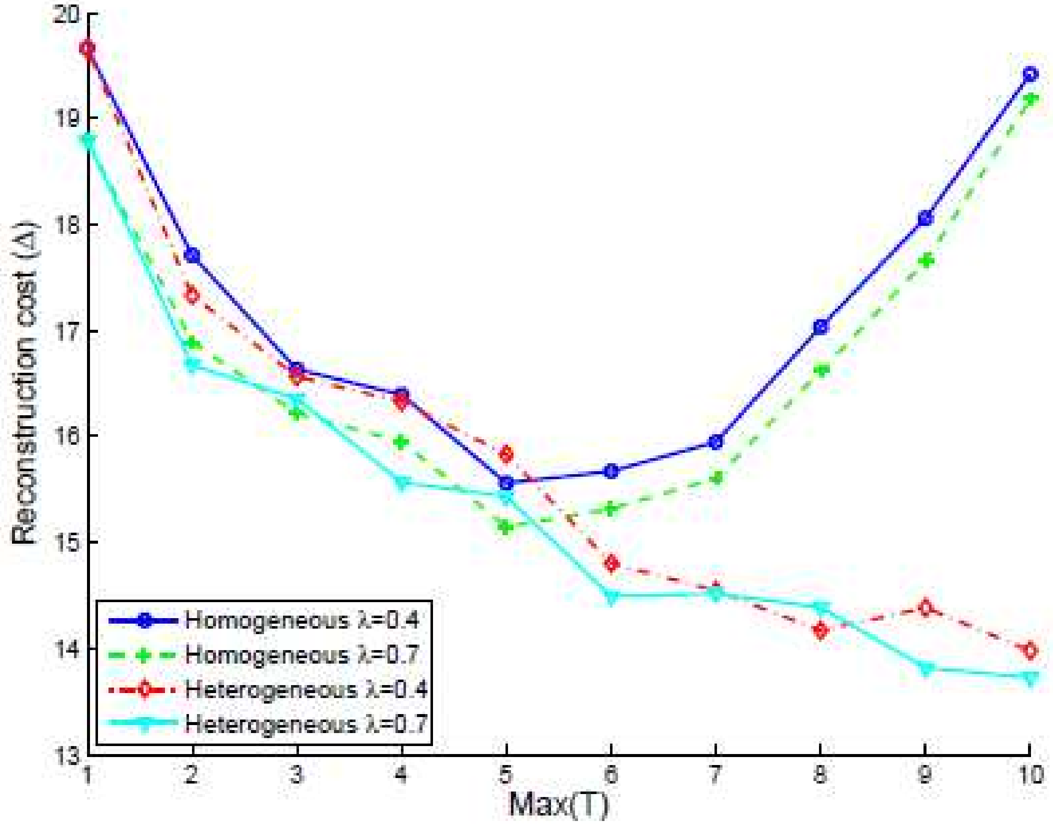
$$L(W,H) = | X - \tilde{X} |_2 + \lambda | H | \qquad X_{i,j}, \lambda, H_{i,j} \geq 0,$$

$$\tilde{X} = \sum_{r=0}^{R-1} \sum_{p=0}^{P(r)-1} W_p \overset{p\rightarrow}{H} \qquad W_{p;i,j}, H_{i,j} \geq 0$$

EA-based Heterogeneous CNSC

- Searching for the best distribution with evolutionary algorithms (EA).
  - Set a base distribution r
  - Generate a set of distributions M based on a Dirichlet distribution whose base distribution is r
  - For each m in M, offspring a set of distributions based on Dirichlet(m)
  - Two offsprings can marry, producing new distributions
  - After the offspring and marriage, leave only G distributions whose reconstruction cost is the least.
  - Repeat until the min cost does not change within 2 generations.

Behavior of homogeneous and heterogeneous CNSC learning

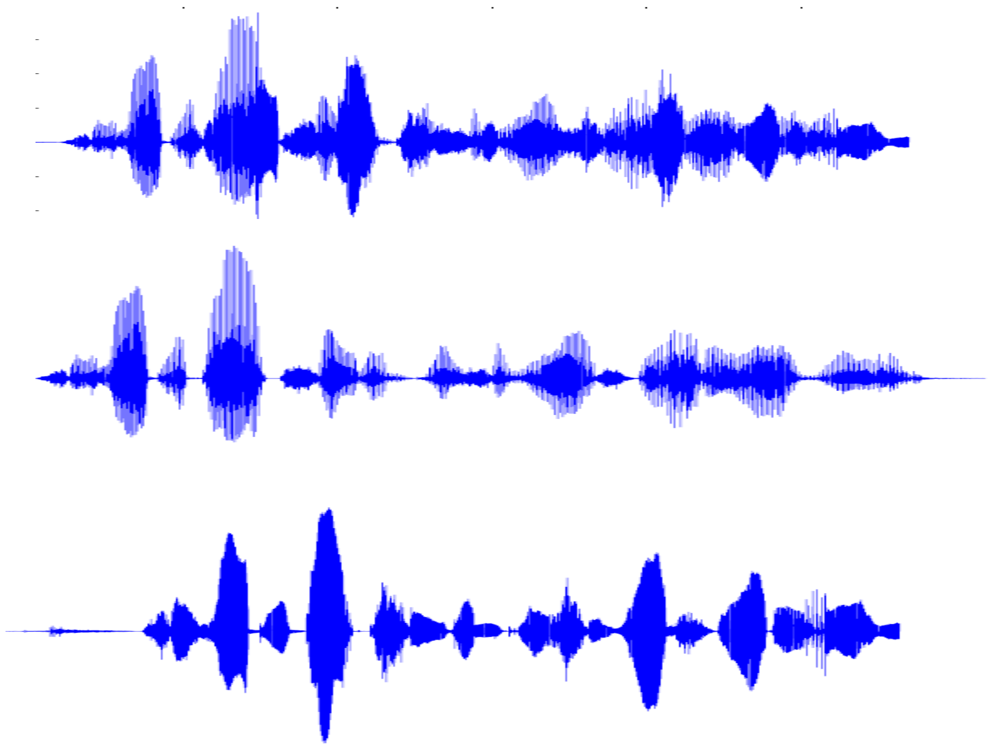


Patterns obtained with EA-based heterogeneous CNSC

The table means how many bases are distributed to each specific T.
It shows that the learning prefers patterns of moderate length.

$\max(T)$	$\lambda = 0.4$	$\lambda = 0.7$
1	[40]	[40]
2	[1 39]	[1 39]
3	[8 3 29]	[6 5 29]
4	[0 10 14 16]	[0 13 10 17]
5	[3 4 9 6 18]	[4 22 2 4 8]
6	[5 1 5 0 0 29]	[9 3 0 1 0 27]
7	[1 7 0 0 0 13 19]	[0 5 3 0 0 8 24]
8	[1 8 0 0 0 2 4 25]	[14 1 0 2 0 4 2 2 17]
9	[0 0 0 5 0 8 6 21 0]	[2 9 1 1 1 0 17 9 0]
10	[3 2 1 0 0 8 2 9 15 0]	[0 1 1 1 1 6 5 17 8 0]

Heterogeneous CNSC for speech separation



Conclusions

- CNSC is a vital approach for speech signal analysis, and can be applied to wide applications such as signal separation and de-noising.
- Heterogeneous CNSC can significantly improve the quality of patterns in representative power and discriminative capability.