Master's Thesis On

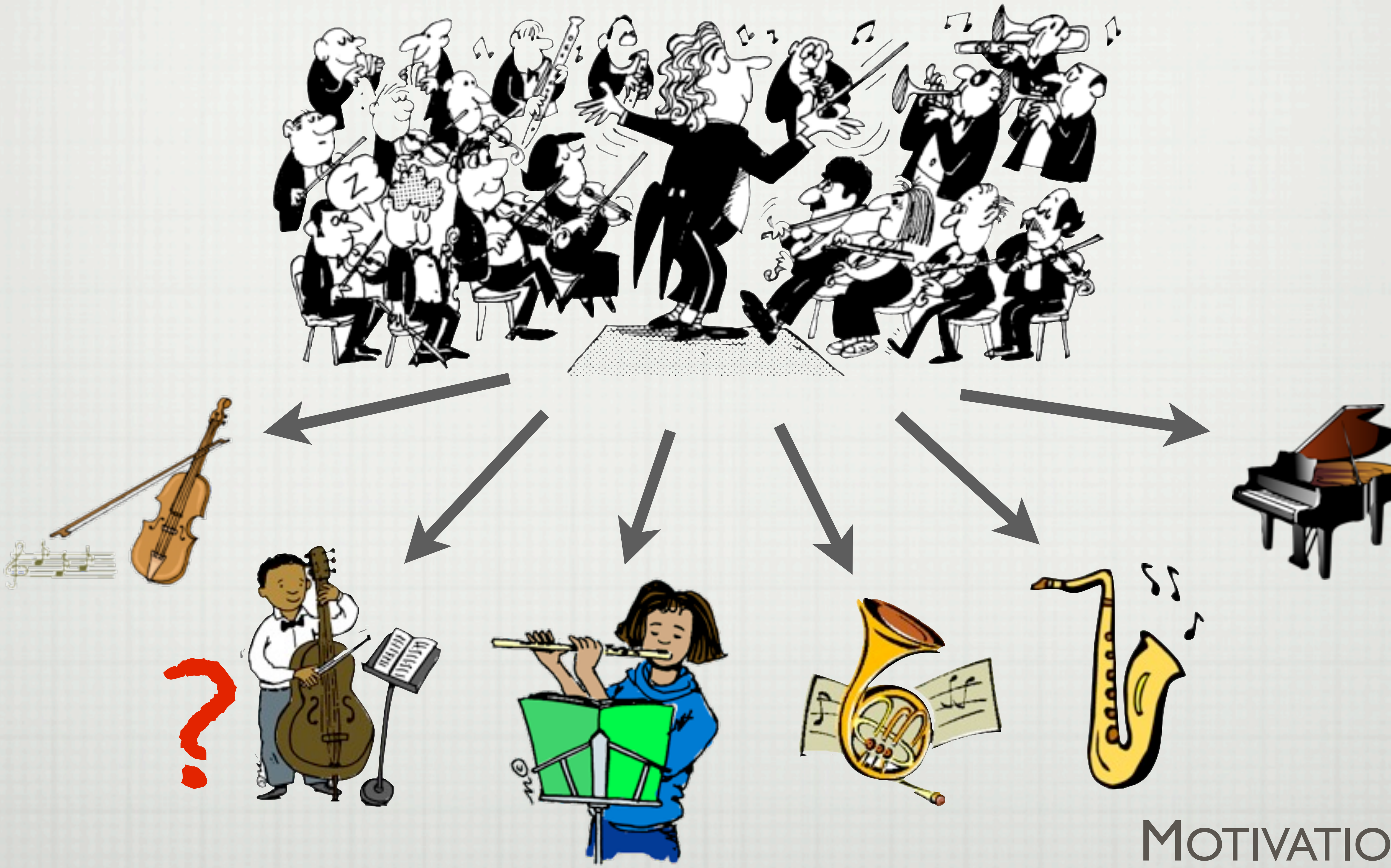
Semi-blind Source Separation via Sparse Representations and Online Dictionary Learning

By

Sirisha Rambhatla

Advisor: Prof. Jarvis Haupt

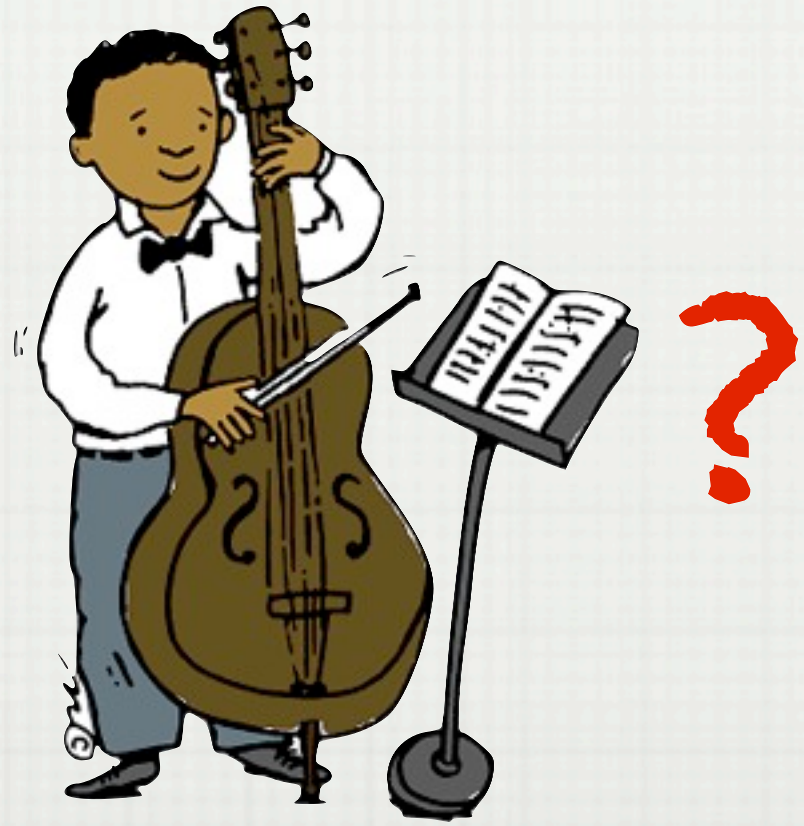
The Last Time I was at The Minnesota Orchestra..



Motivation

Sunday, December 9, 12

How Do I Identify What Cello Sounds Like?

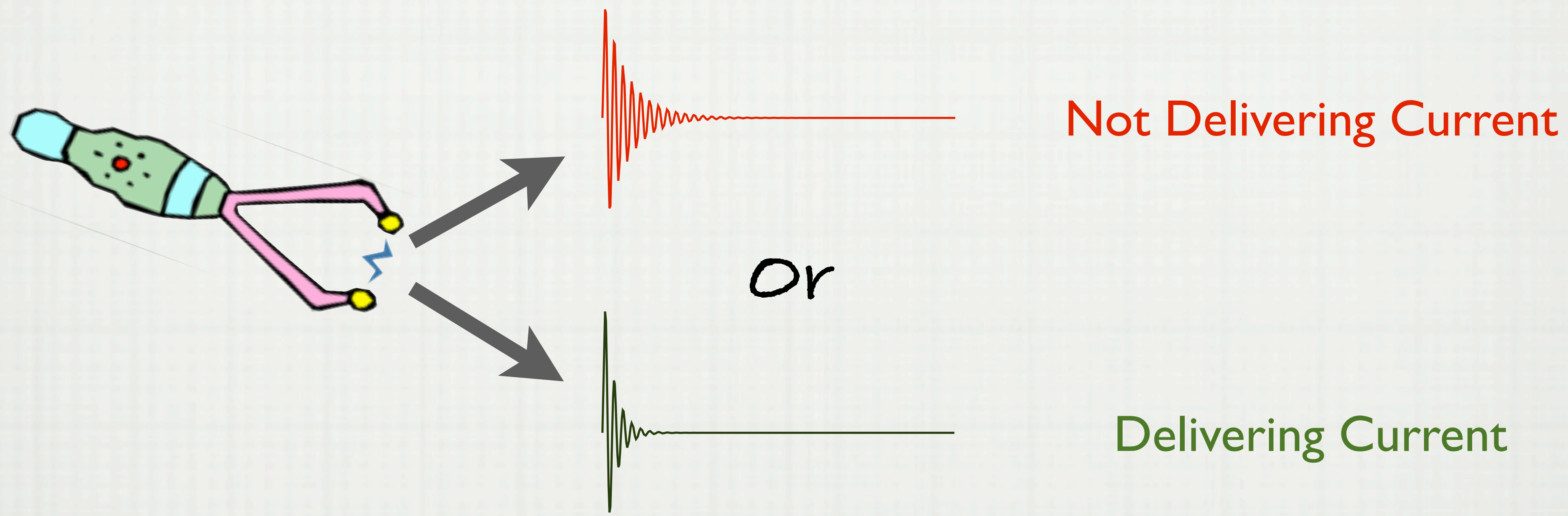


Motivation

THE *COCKTAIL PARTY* PROBLEM : BLIND SOURCE SEPARATION

- Multiple speakers are simultaneously speaking
- The aim is to separately comprehend each speaker.
- None of the sources are known a-priori - ***Blind Source Separation Problem***

Consider an Audio Forensics Application



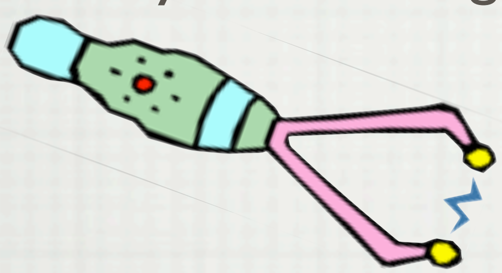
- Electro-shock Law enforcement devices generate characteristic **Nominally Periodic** signals, indicating discharge of current.
- These signals are often corrupted by background noise like speech, etc., not known *a-priori*.

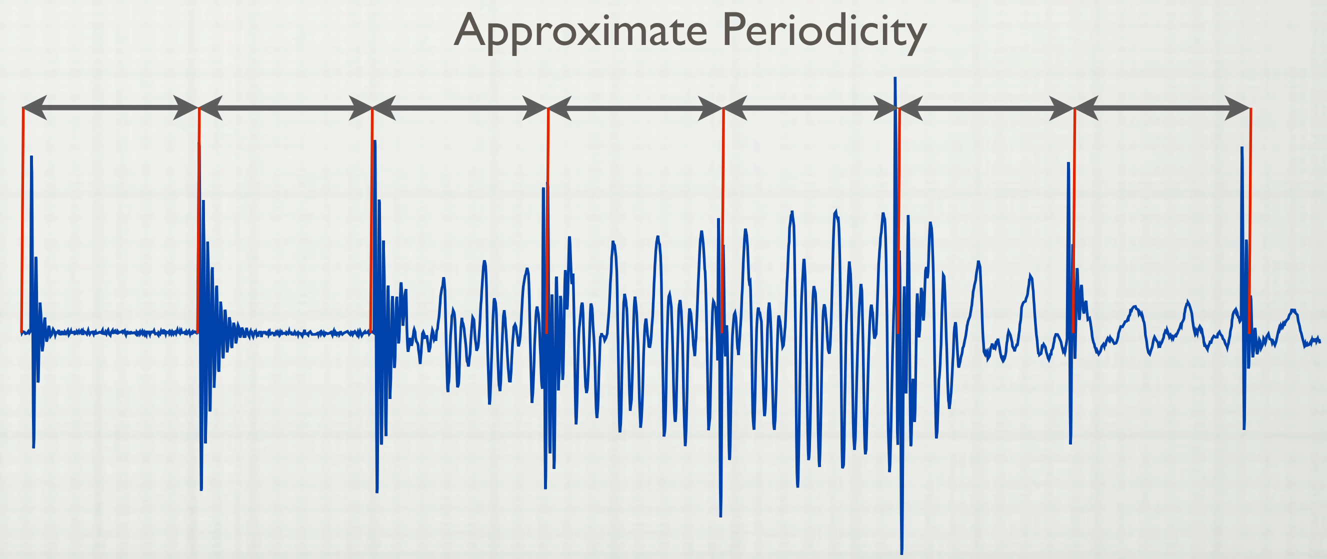
Motivation

Audio Forensics Application: Single Channel Source Separation

- It is of interest to detect if the device is delivering current or not from a single mixture of the sources, referred to as **Single-Channel Source Separation**
- On the whole, ***A Single Channel Semi-Blind Source Separation problem.***

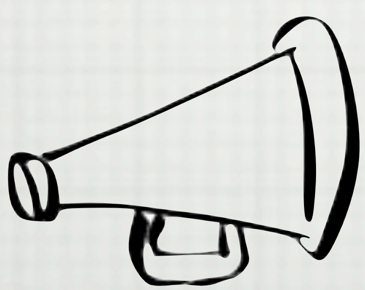
Nominally Periodic signal $x_p \in \mathbb{R}^n$





Single Linear Mixture

$x \in \mathbb{R}^n$



Background Noise $x_u \in \mathbb{R}^n$

Motivation

Outline : Bird's Eye View of The Presentation

- ☐ Background: Setting the stage
- ☐ Semi-Blind Morphological Component Analysis (SBMCA)
- ☐ Evaluation of SBMCA : Simulation Specifics
- ☐ Conclusions and Future Work

BACKGROUND

Model

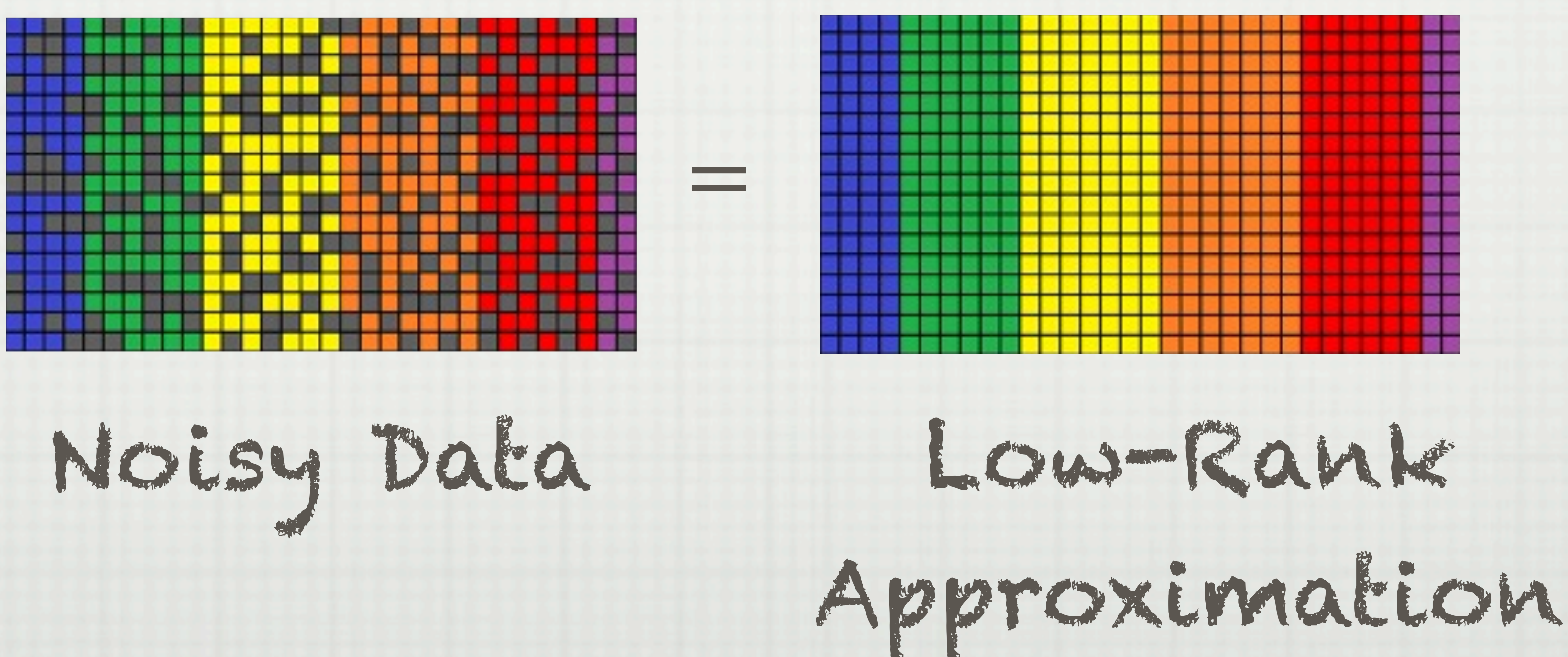
- We suppose that, m is an integer which divides n into q equal parts
  - x is represented by matrix $X \in \mathbb{R}^{m \times q}$, and

$$X = X_p + X_u$$

- The aim is to separate X into its constituent matrices X_p and X_u

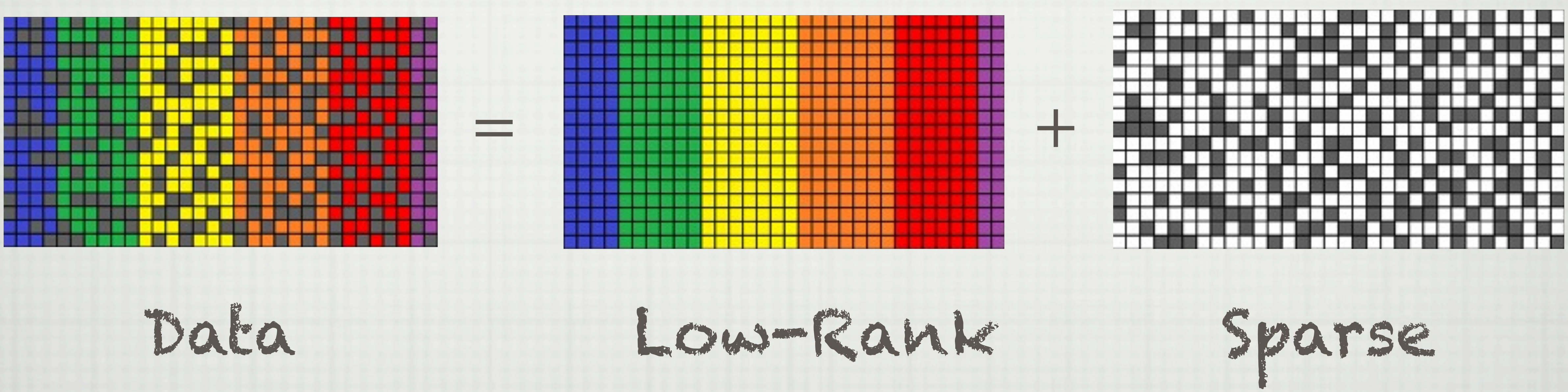
Truncated-Singular Value Decomposition(SVD)

- Let X_p be a rank-r matrix and X_u be random Gaussian noise.
- Estimating X_p is equivalent to finding a rank-r approximation of X



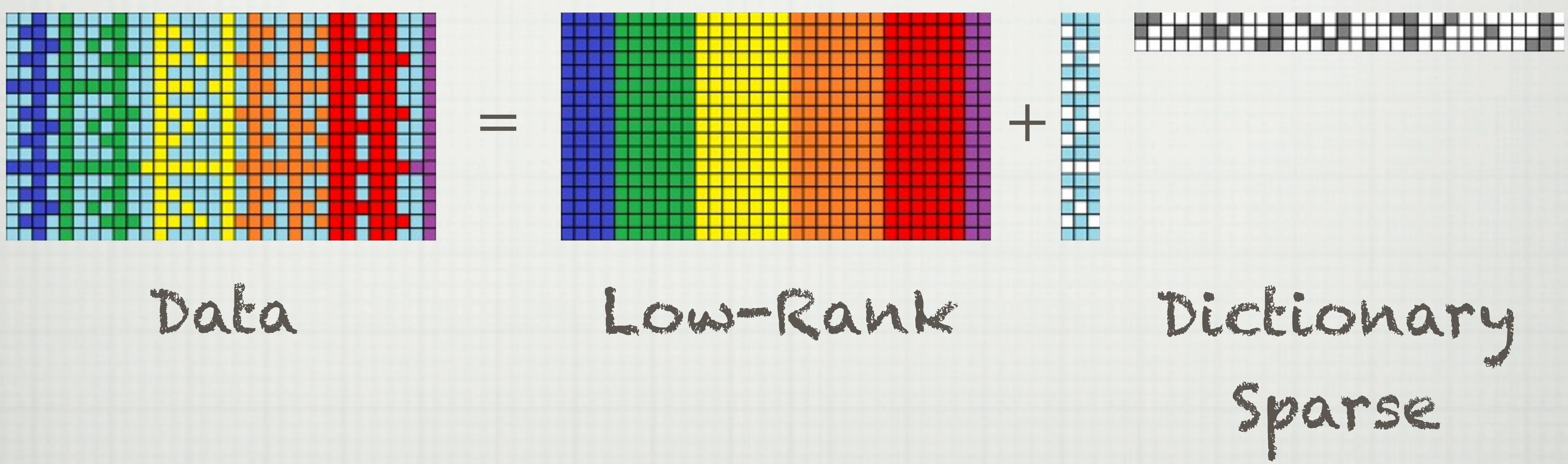
ROBUST PCA[1,2]

- Let X_u be comprised of impulsive noise, in such a case we adopt Robust PCA, leading to following decomposition :



LOW-RANK PLUS SPARSE IN A KNOWN DICTIONARY[3]

- In case X_u is sparse in some known dictionary,



BACKGROUND

Morphological Component Analysis[4,5,6]

- ☐ Another extension: X_p and X_u are both sparse in some known dictionaries, represented as :

Background

Revisiting The Cellist:
So, How Do I Identify What Cello Sounds Like?

- ☐ I can listen to a sample of Cello before the next act.
- ☐ I **Train** my ears to Cello.
- ☐ Training data required.

Or

- ☐ I know how other instruments sound.
- ☐ I **Learn** the features of Cello by employing my prior **Experience** with other instruments.
- ☐ An **Online** methodology.

Revisiting The Cellist: How Do I Identify What Cello Sounds Like?

Motivation for our approach

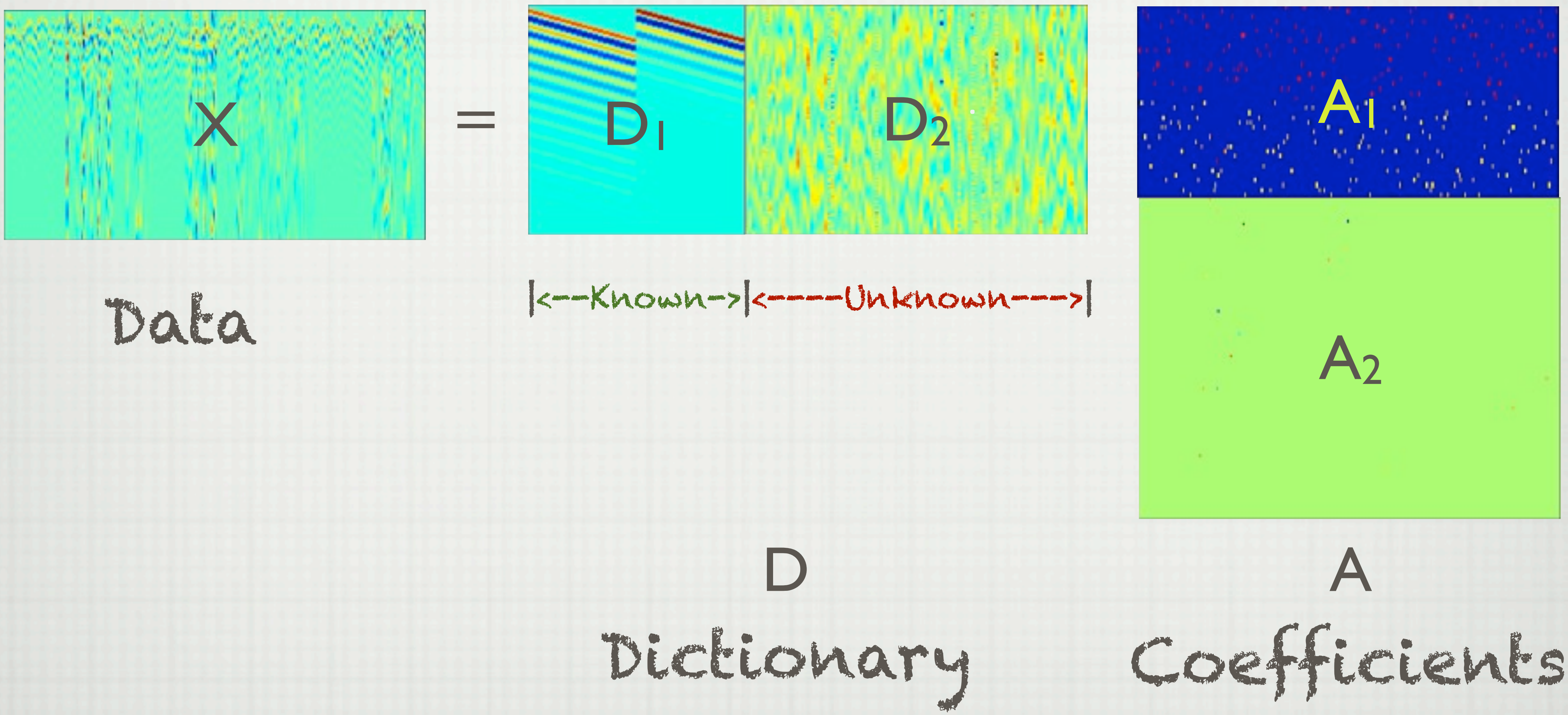
- I can listen to a sample of Cello before the next act.
- I **Train** my ears to Cello.
- Training data required.

Or

- I know how other instruments sound.
- I **Learn** the features of Cello by employing my prior **Experience** with other instruments.
- An **Online** methodology.

SEMI-BLIND MORPHOLOGICAL COMPONENT ANALYSIS (SBMCA)

Semi-Blind Morphological Component Analysis[7]



SBMCA

SEMI-BLIND MORPHOLOGICAL COMPONENT ANALYSIS: ALGORITHMIC CONSIDERATIONS

- Formally we set out to solve

$$\{\widehat{A}_1, \widehat{A}_2, \widehat{D}_2\} = \underset{A_1, A_2, D_2}{\arg\min} \|X - D_1 A_1 - D_2 A_2\|_F^2 + \widetilde{\lambda}_1 \|A_1\|_1 + \widetilde{\lambda}_2 \|A_2\|_1$$

$$for\ \widetilde{\lambda}_1,\ \widetilde{\lambda}_2 > 0$$

- This optimization problem is
  - Not jointly convex
  - Sensitive to initialization
- We adopt
  - Alternating Minimization based approach for Online Dictionary Learning[8,9,10]

SBMCA

Semi-Blind Morphological Component Analysis
Algorithmic details

Algorithm 1: Semi-Blind MCA Algorithm

Input: Original Data $X \in \mathbb{R}^{m \times q}$, Known Dictionary $D_1 \in \mathbb{R}^{m \times d}$,
Regularization parameters $\lambda_1, \lambda_2, \lambda_3 > 0$,
Number of elements in unknown dictionary ℓ.

Initialize: $\widetilde{A}_1 \leftarrow \arg\min_{A_1} \|X - D_1 A_1\|_F^2 + \lambda_1 \|A_1\|_1$
(or other suitable initialization depending on the problem.)

Iterate (repeat until convergence):

repeat

Dictionary Learning:

$$\{\widetilde{D}_2, \widetilde{A}_2\} \leftarrow \arg\min_{D_2, A_2} \|X - D_1 \widetilde{A}_1 - D_2 A_2\|_F^2 + \lambda_2 \|A_2\|_1$$

Coefficient Update:

$$\widetilde{D} = [D_1 \ \widetilde{D}_2]$$

$$[\widetilde{A}_1^T \ \widetilde{A}_2^T]^T \triangleq \widetilde{A} \leftarrow \arg\min_{A} \|X - \widetilde{D} A\|_F^2 + \lambda_3 \|A\|_1$$

until convergence

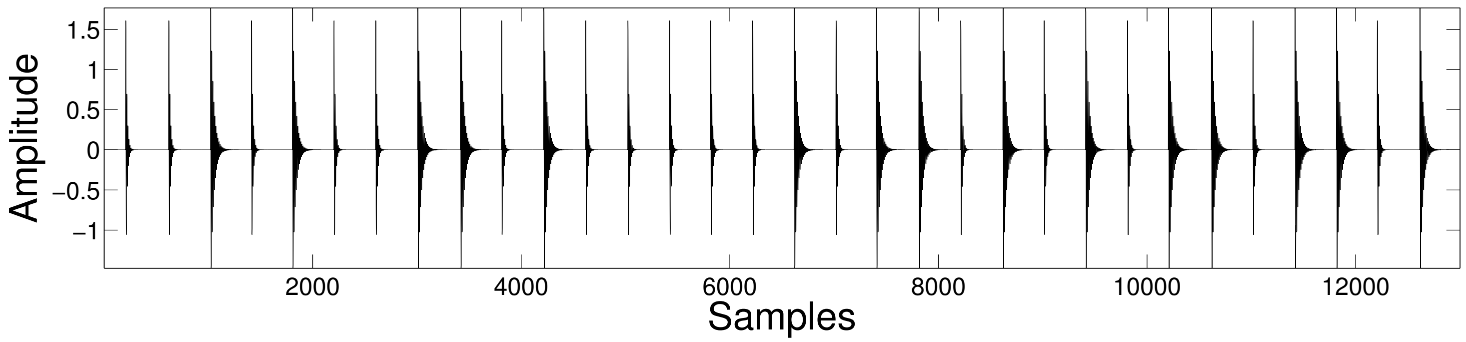
Output: Learned dictionary $\widehat{D}_2 \leftarrow \widetilde{D}_2$,
Coefficient estimates $\widehat{A}_1 = \widetilde{A}_1, \ \widehat{A}_2 = \widetilde{A}_2$.

SBMCA

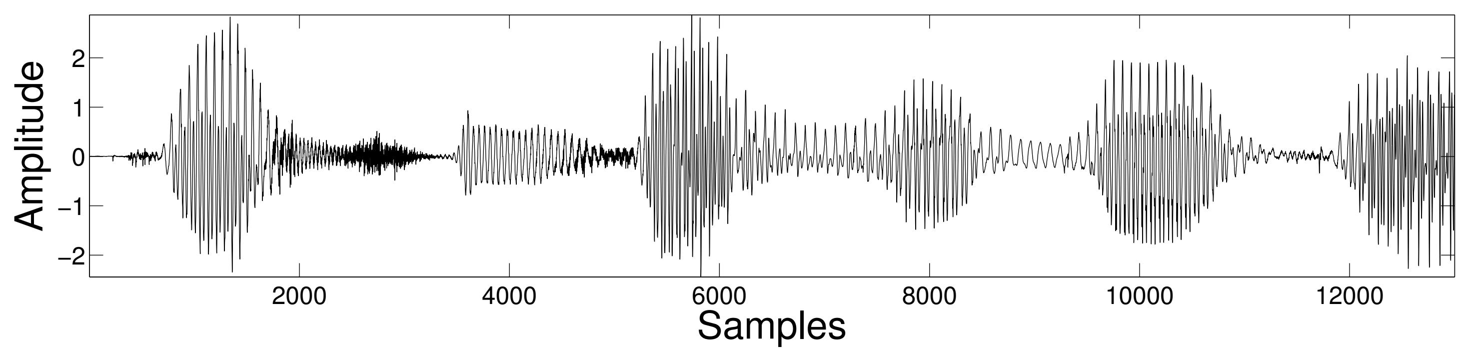
EVALUATION OF PERFORMANCE

Signal Configuration

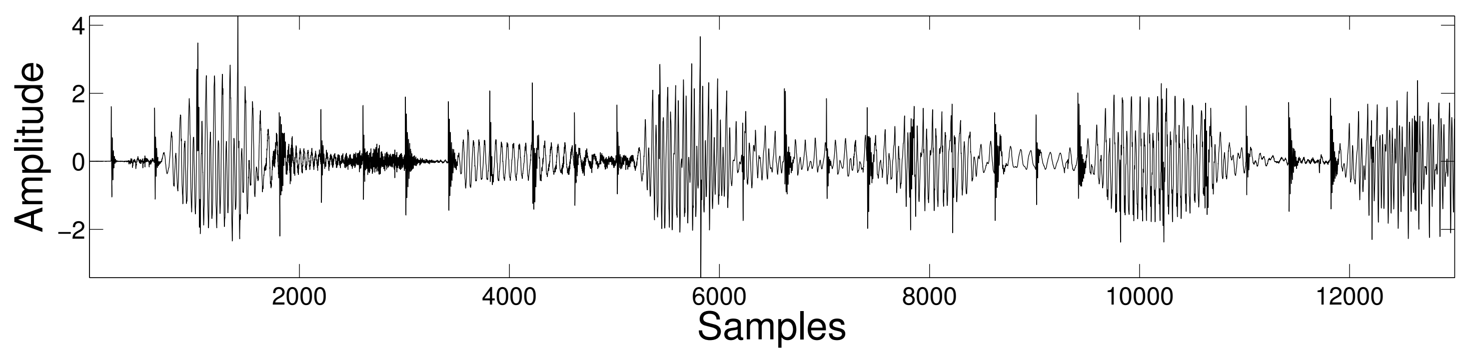
Nominally Periodic Signal



Unknown Background Signal



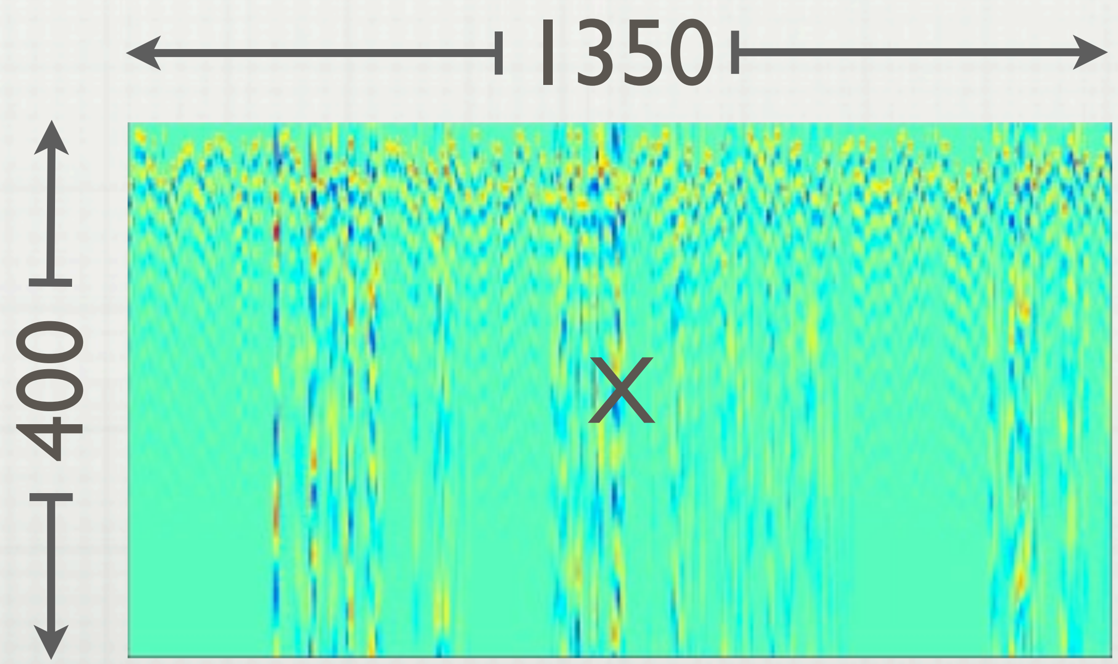
Linear Mixture



Evaluation

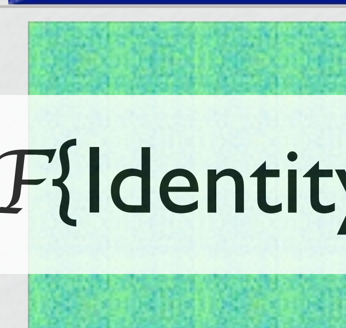
Data Generation

- Data formed from a mixture of Speech[1] (unknown) and nominally periodic signal (one per period)
- Data matrix looks like,



[1] Speech Samples obtained from VoxForge Speech Corpus: www.voxforge.org/home

Frequency Domain Source Separation

Method	D_1	D_2
SBMCA		 Unknown
MCA-DCT-Fourier		 $\|\mathcal{F}\{\text{DCT}\}\|$
MCA-Identity-Fourier		 $\|\mathcal{F}\{\text{Identity}\}\|$

Source Separation in Frequency Domain

Table 1: Analysis of reconstruction SNR(in dB): Frequency Domain Separation

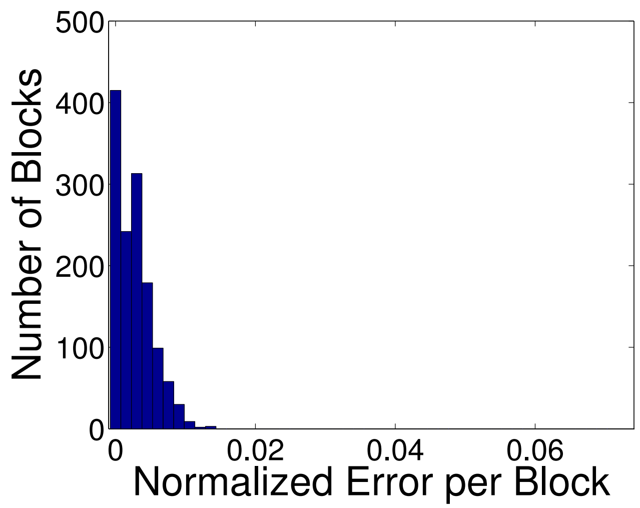
Noise $\mathcal{N}(0,\sigma^2)$	$\sigma = 0$		$\sigma = 0.001$		$\sigma = 0.01$		$\sigma = 0.1$	
Method \ Signal	x_p	x_u	x_p	x_u	x_p	x_u	x_p	x_u
SBMCA	**8.95**	**15.21**	**8.91**	**15.17**	8.83	15.09	6.80	11.56
MCA-DCT-Fourier	8.81	15.16	8.81	15.16	**8.88**	**15.19**	**6.82**	**12.50**
MCA-Identity-Fourier	1.19	-19.07	1.19	-19.07	1.19	-18.90	1.34	-9.57

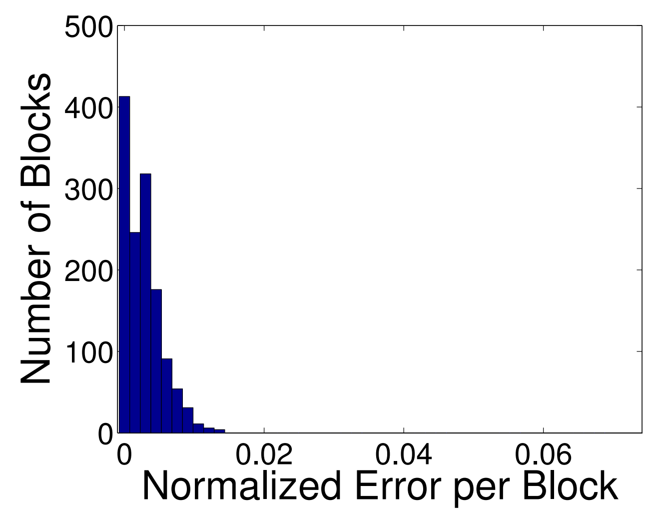
Evaluation

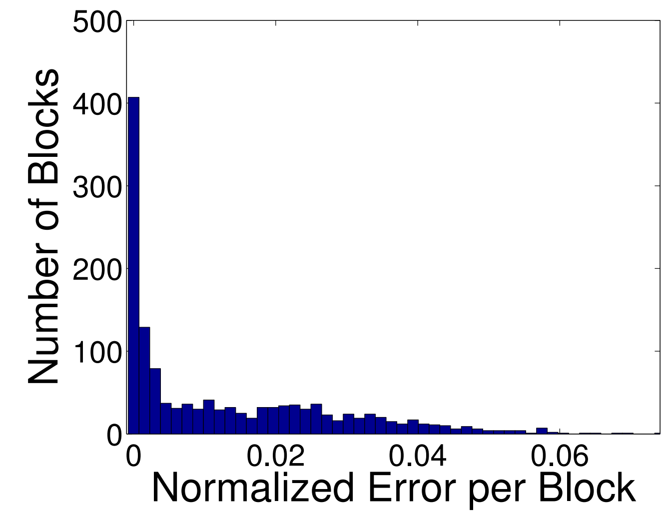
Source Separation in Frequency Domain: Nominally Periodic Signal

SBMCA | MCA-DCT-FOURIER | MCA-Identity-Fourier

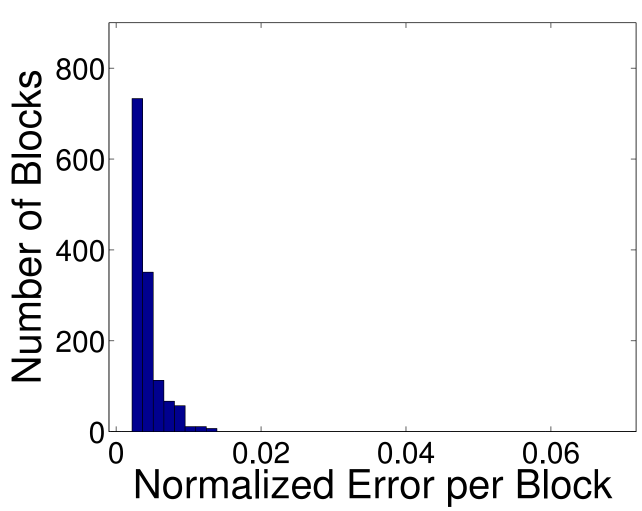
$\sigma = 0$

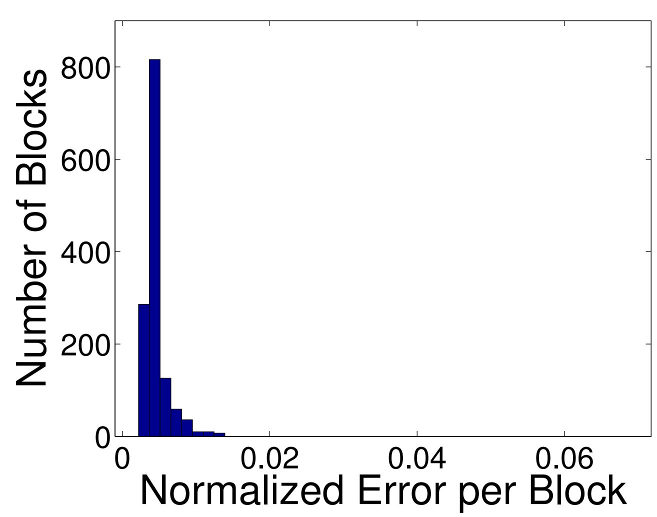


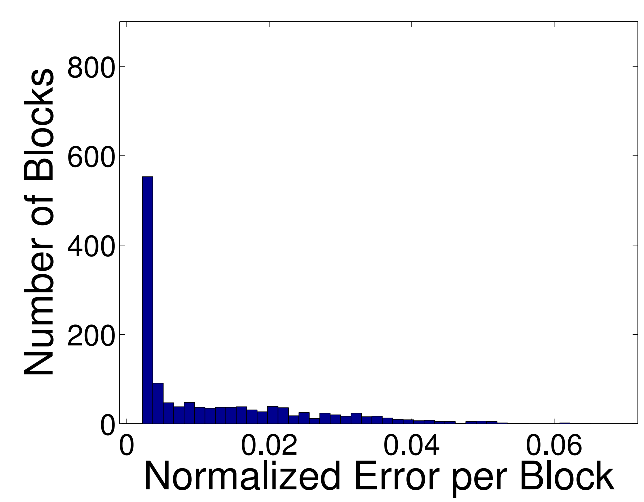




$\sigma = 0.1$





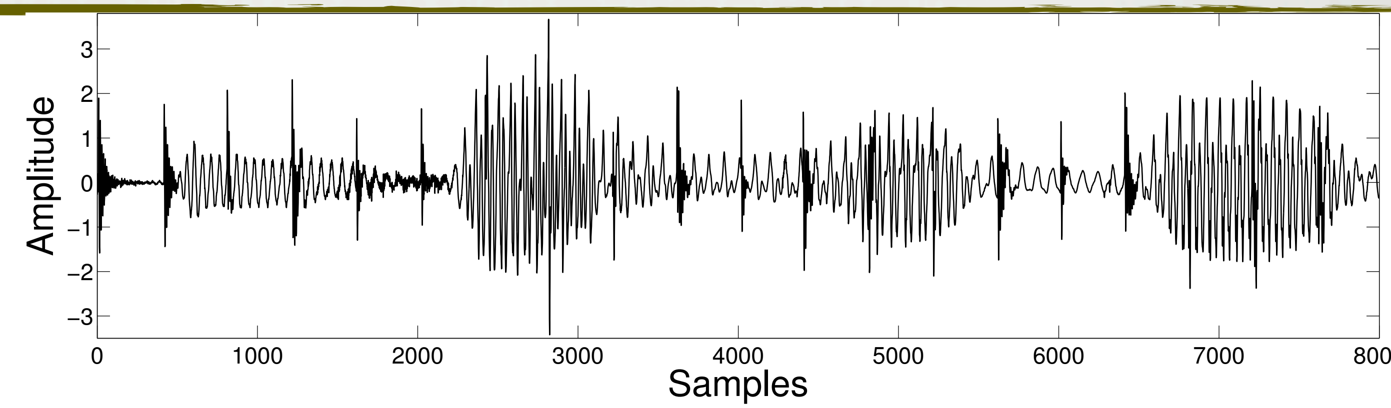


Evaluation

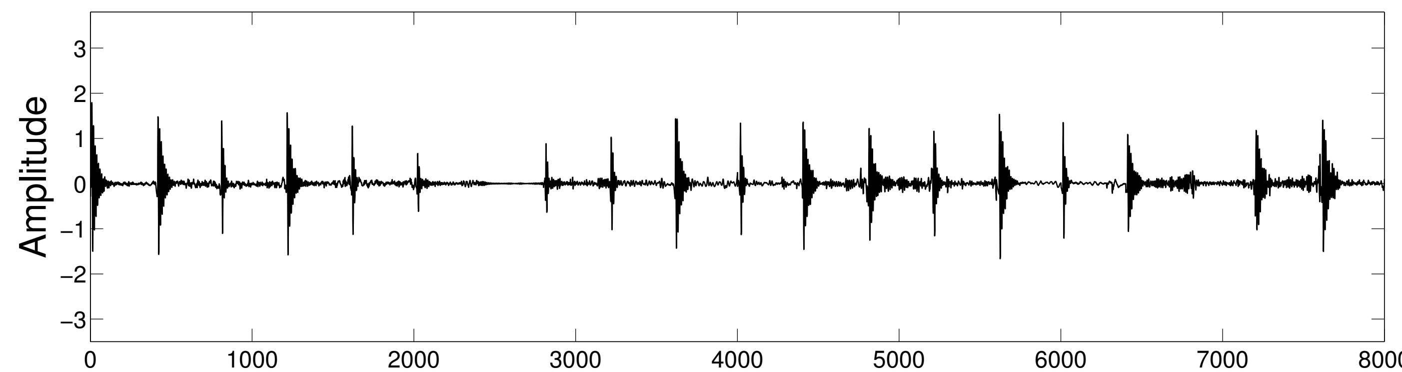
Source Separation in Frequency Domain: Nominally Periodic Signal

$\sigma = 0$

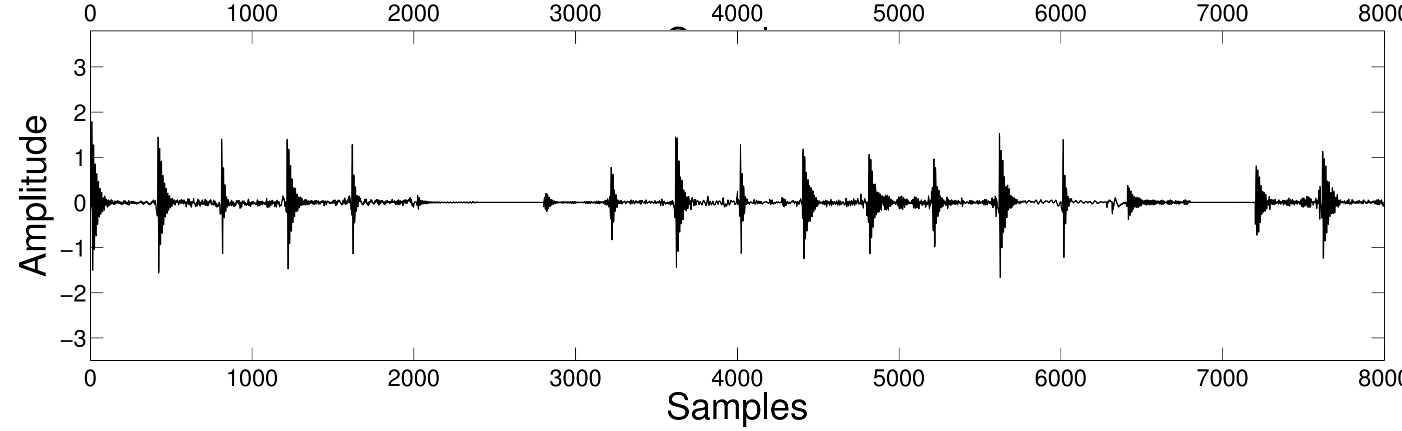
Original Mixture



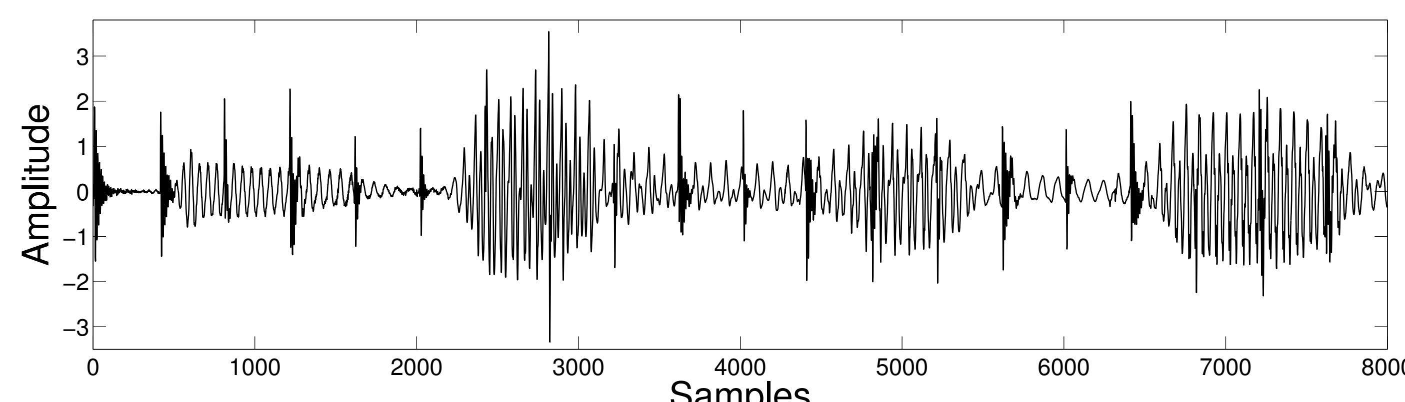
SBMCA



MCA-DCT-Fourier

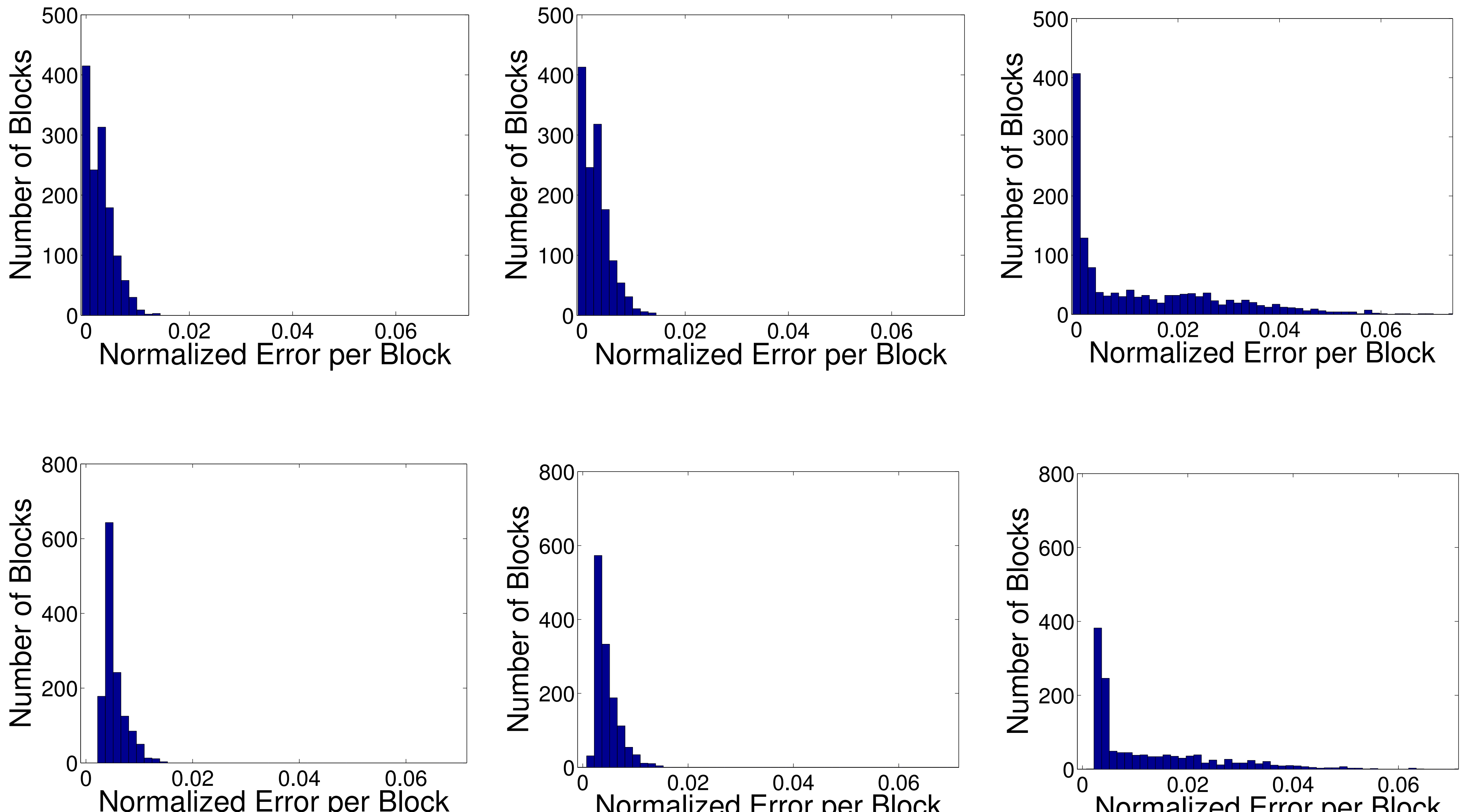


MCA-Identity-Fourier



Source Separation in Frequency Domain: Background Signal

	SBMCA	MCA-DCT-FOURIER	MCA-Identity-Fourier
$\sigma = 0$			
$\sigma = 0.1$			

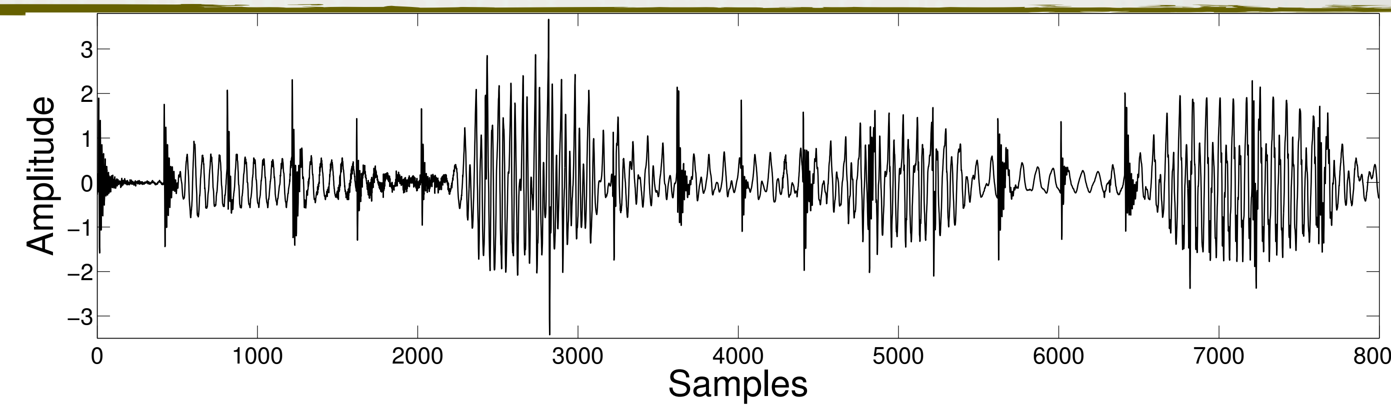


Evaluation

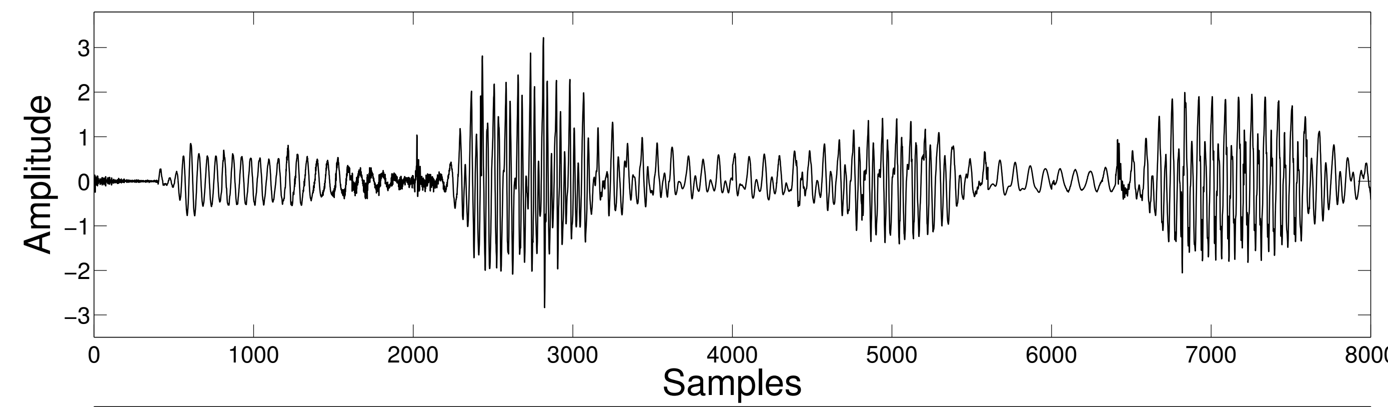
Source Separation in Frequency Domain: Background Signal

$\sigma = 0$

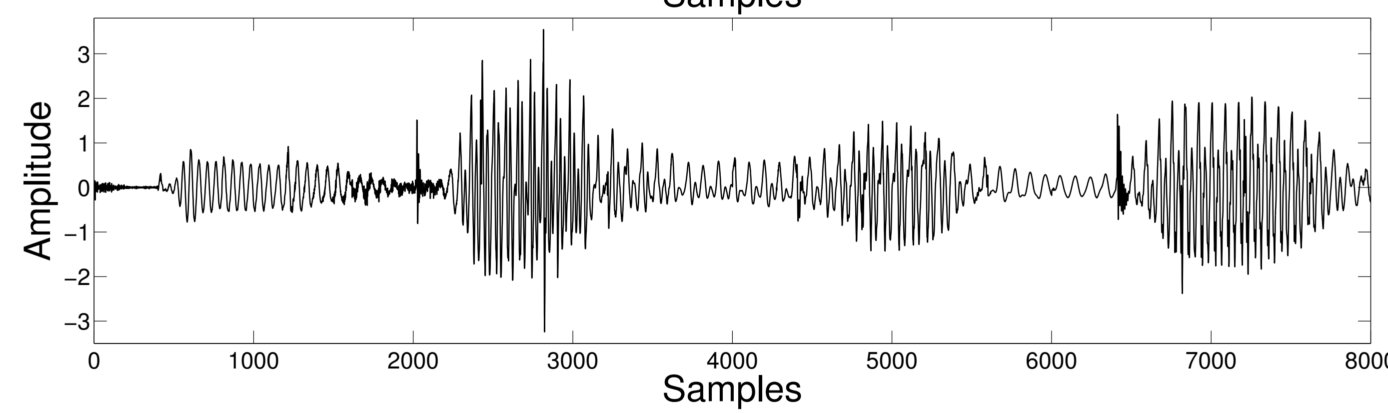
Original Mixture



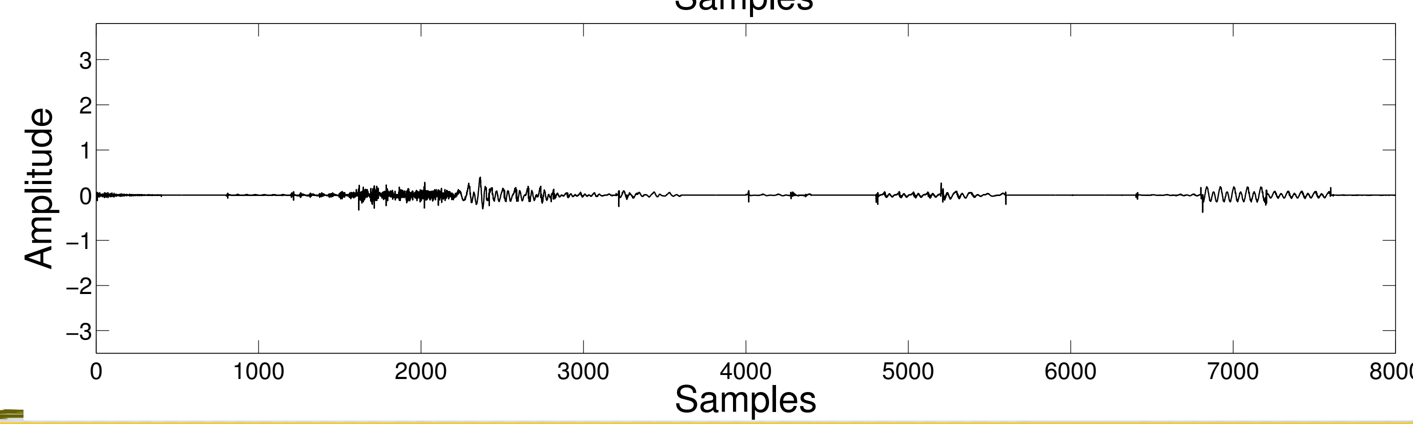
SBMCA



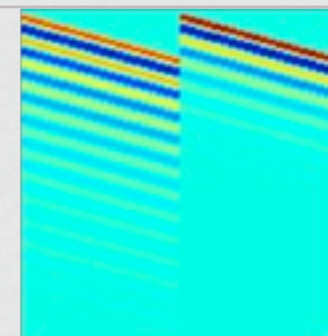
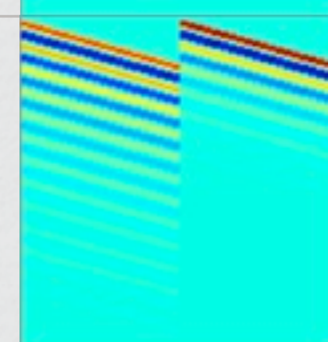
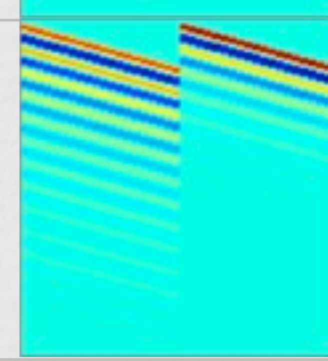
MCA-DCT-Fourier



MCA-Identity-Fourier



Time Domain Source Separation

Method	D_1	D_2
SBMCA		
MCA-DCT		
MCA-Identity		

Source Separation in Time Domain

Table 2: Analysis of reconstruction SNR(in dB): Time Domain Separation

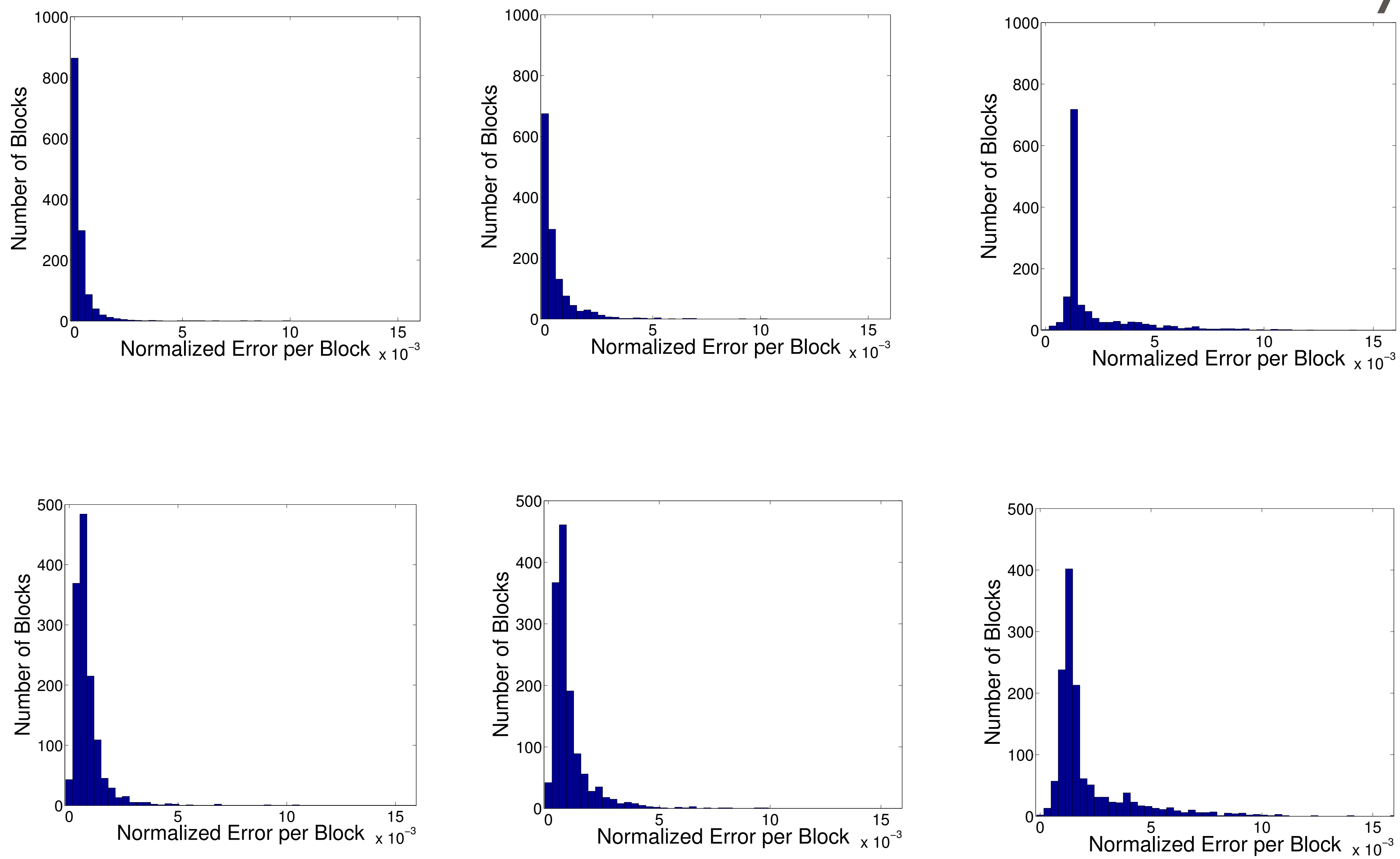
Noise $\mathcal{N}(0, \sigma^2)$	$\sigma = 0$		$\sigma = 0.001$		$\sigma = 0.01$		$\sigma = 0.1$	
Method \ Signal	x_p	x_u	x_p	x_u	x_p	x_u	x_p	x_u
SBMCA	**23.72**	**29.32**	**23.73**	**29.32**	**23.08**	**27.40**	**19.72**	**16.84**
MCA-DCT	20.44	26.02	20.46	26.02	20.19	24.96	18.09	16.72
MCA-Identity	10.90	16.06	10.90	16.44	10.90	16.33	10.78	11.44

Source Separation in Time Domain: Nominally Periodic Signal

SBMCA | MCA-DCT | MCA-Identity

$\sigma = 0$

$\sigma = 0.1$

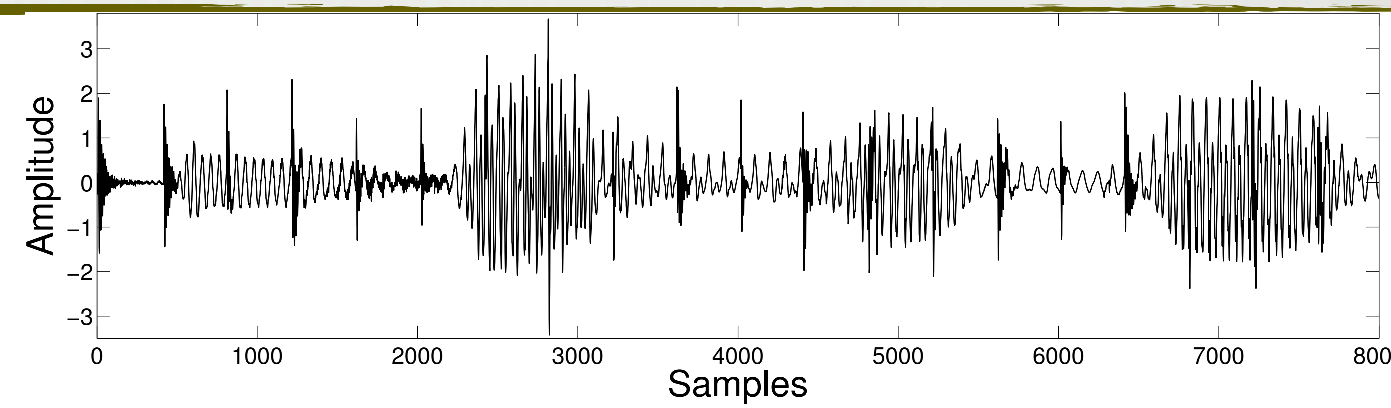


Evaluation

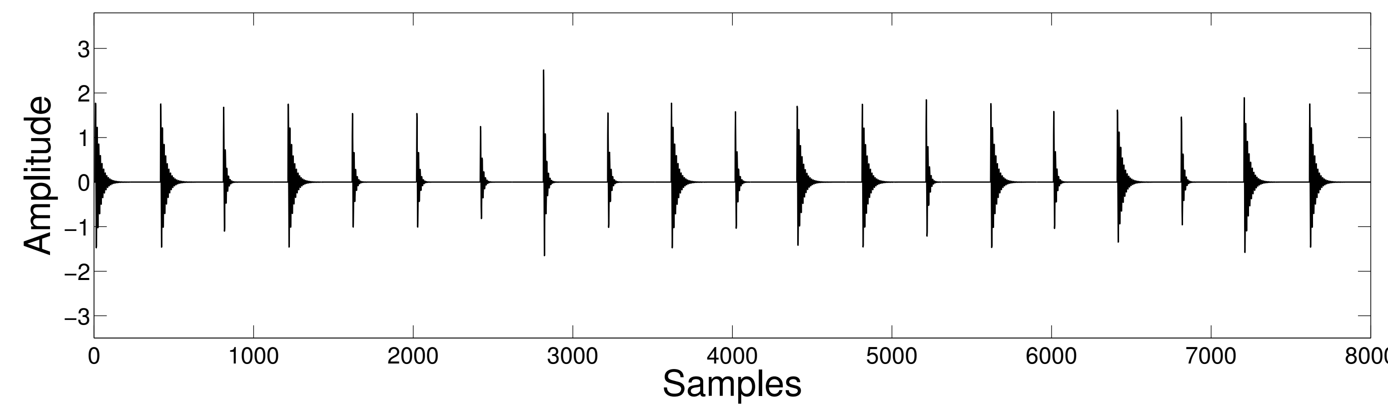
Source Separation in Time Domain: Nominally Periodic Signal

$\sigma = 0$

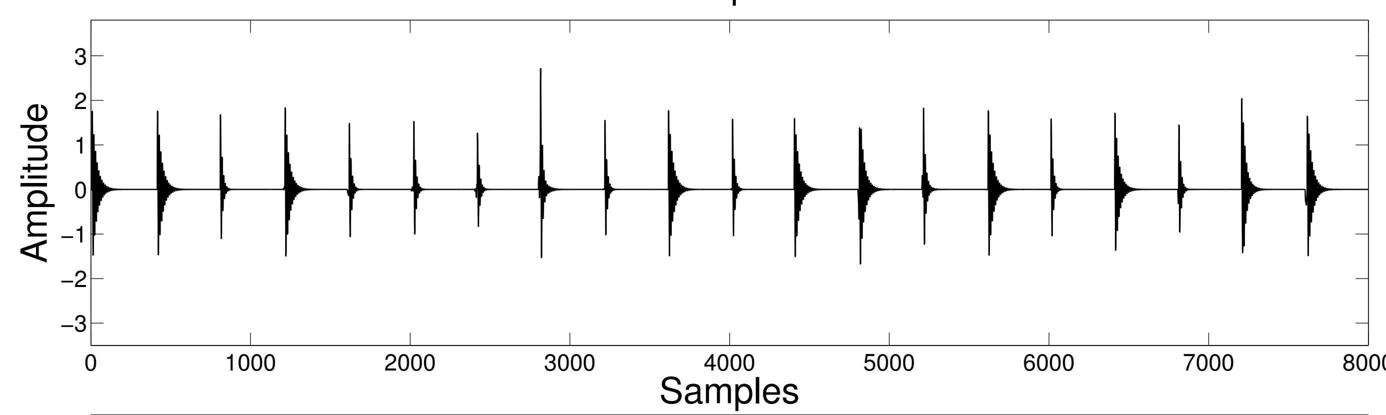
Original Mixture



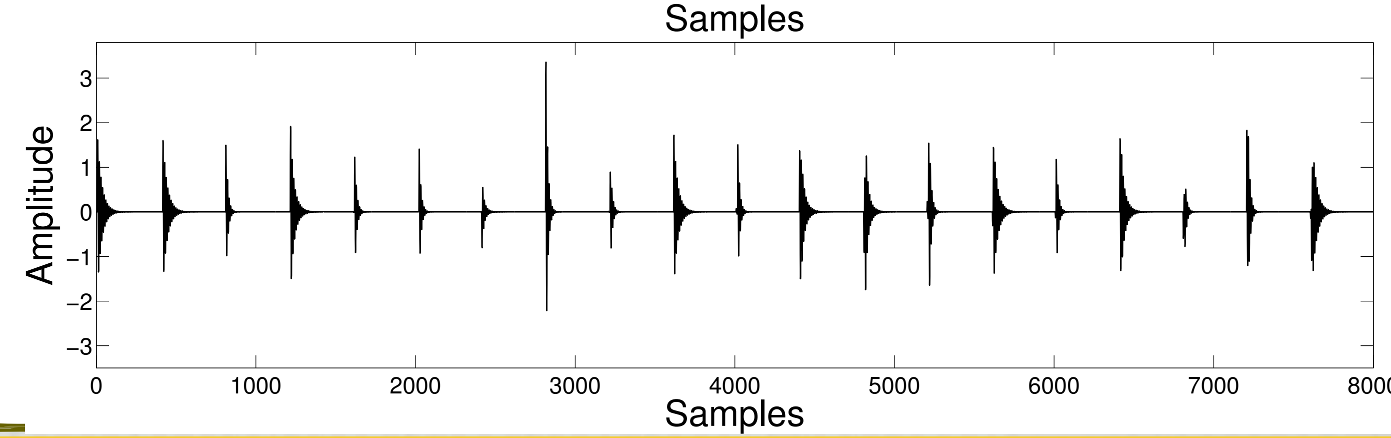
SBMCA



MCA-DCT



MCA-Identity

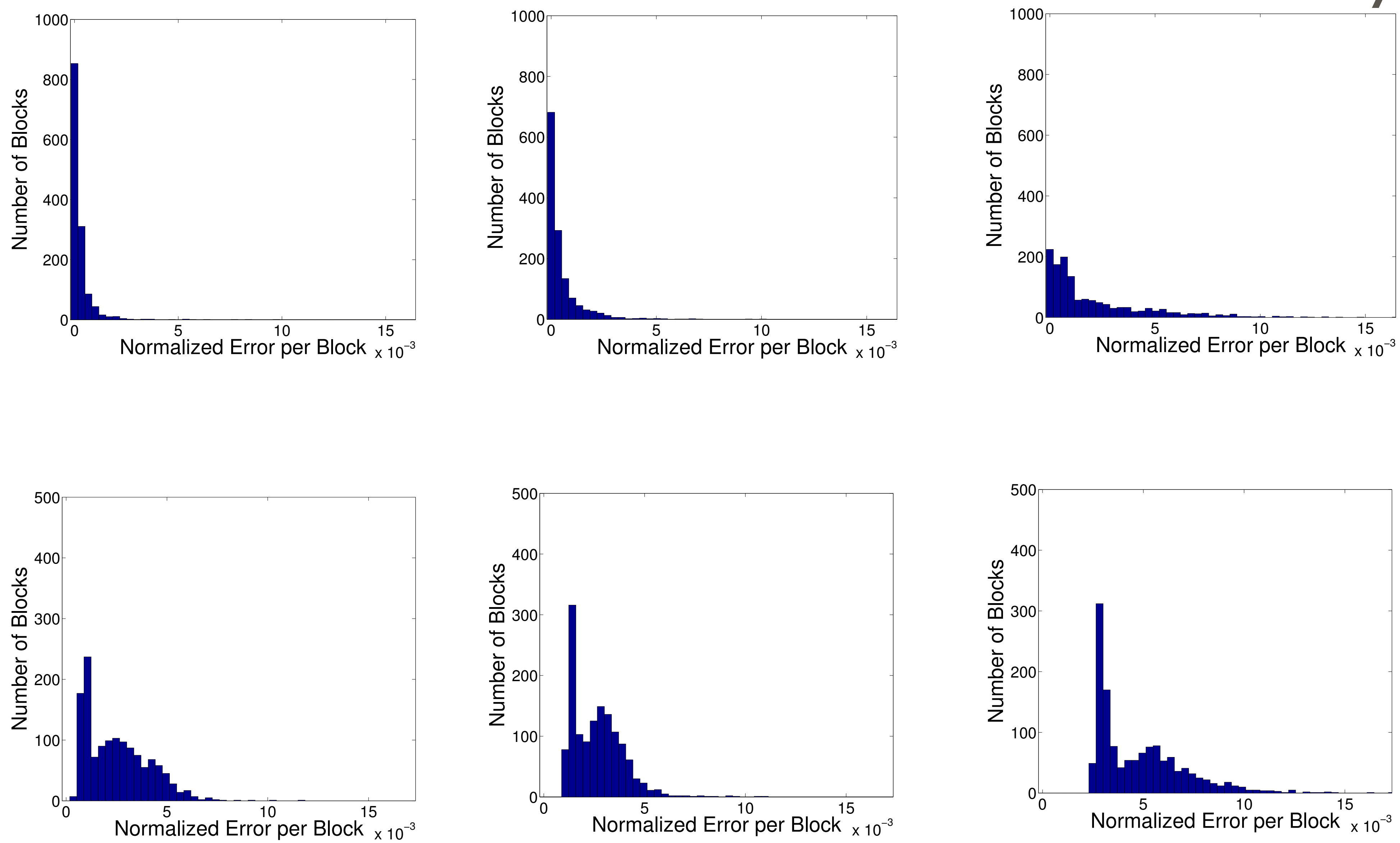


Source Separation in Time Domain: Background Signal

SBMCA | MCA-DCT | MCA-Identity

$\sigma = 0$

$\sigma = 0.1$

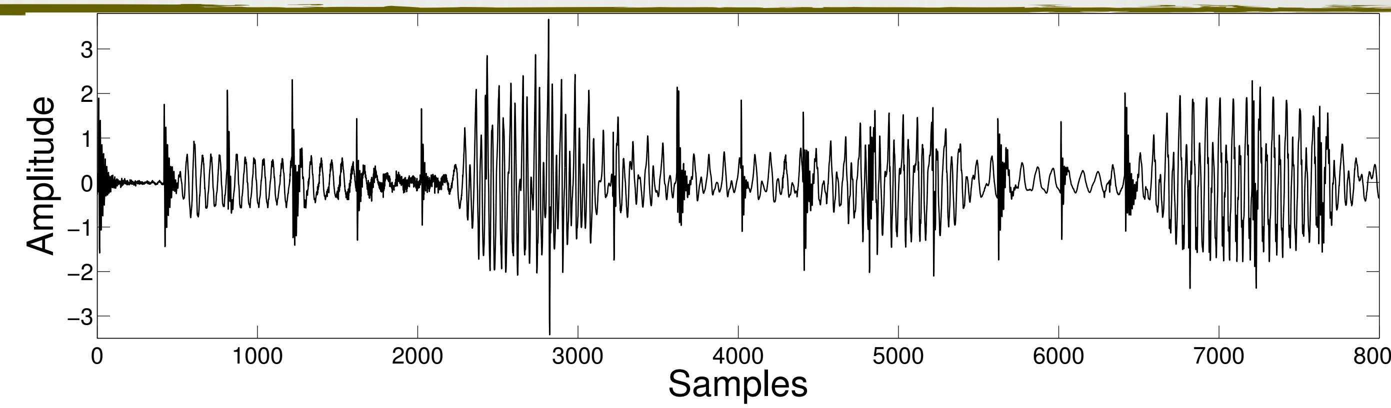


Evaluation

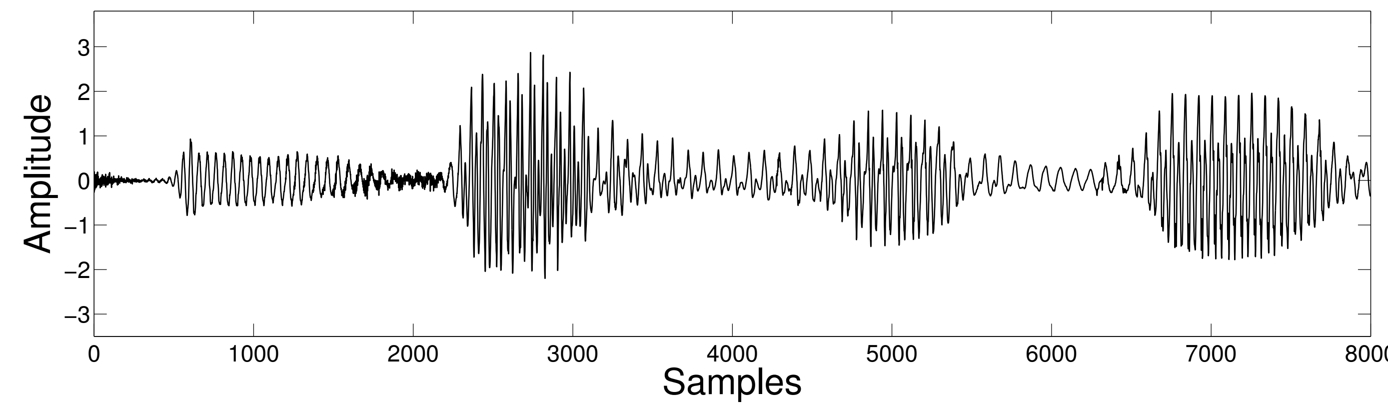
Source Separation in Time Domain: Background Signal

$\sigma = 0$

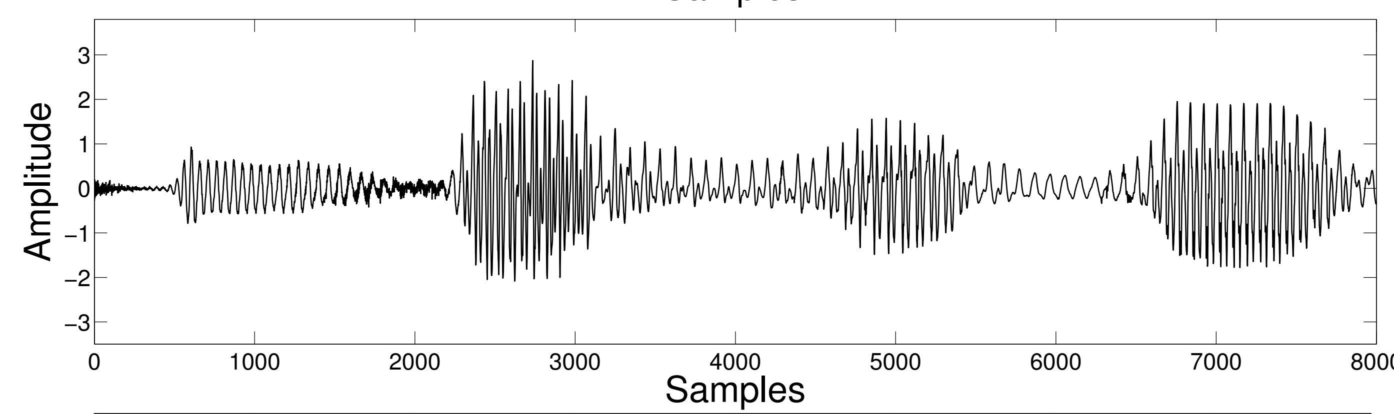
Original Mixture



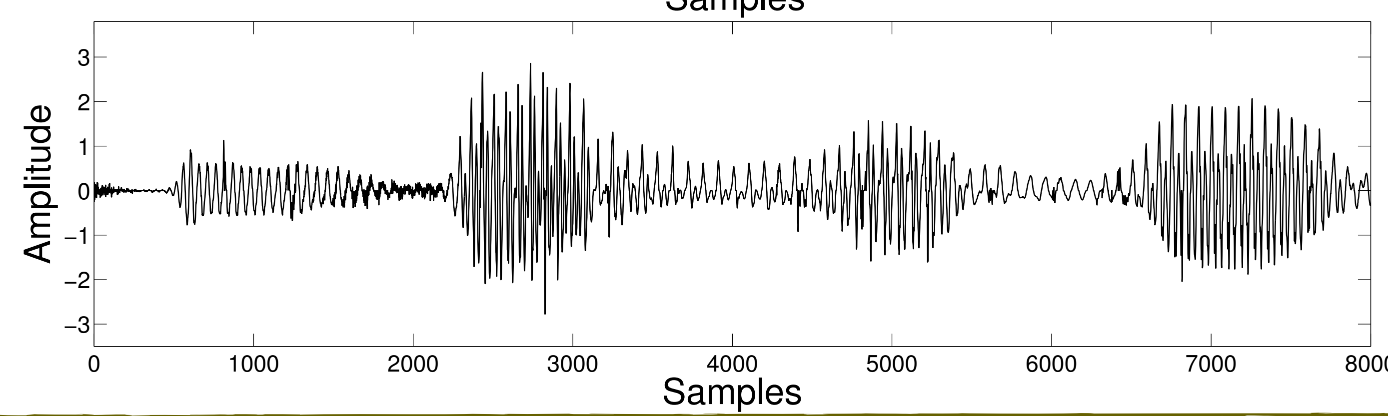
SBMCA



MCA-DCT



MCA-Identity



CONCLUSIONS

- Our approach exploits partial prior knowledge of one of the sources, in the form of a dictionary which sparsely represents local segments of one of the sources. A key feature being online learning of a dictionary (from the mixed source data itself) for representing the unknown background source.

- The timing uncertainty inherent in our application suggests that our approach may be combined with other existing alignment techniques [11, 12, 13].

- More recently [14] proposed a robust alignment procedure that can be viewed as an extension of robust PCA.

- We defer these extensions, as well as the investigation of our approach to other applications (e.g., in image or video processing)to future efforts.

References

1. E. J. Cand`es, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis? Journal of the ACM, 58(3):11, 2011.
2. V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-Sparsity Incoherence for Matrix Decomposition. Society for Industrial and Applied Mathe- matics (SIAM) Journal on Optimization, 21(2):572–596, 2011.
3. M. Mardani, G. Mateos, and G. B. Giannakis. Recovery of Low-Rank Plus Com- pressed Sparse Matrices with Application to Unveiling Traffic Anomalies. IEEE Transactions on Information Theory, 2012. Online: arXiv:1204.6537v1 [cs.IT].
4. J. L. Starck, M. Elad, and D. L. Donoho. Image Decomposition via the Combination of Sparse Representations and a Variational Approach. IEEE Transactions on Image Processing, 14(10):1570–1582, 2005.
5. D. L. Donoho and G. Kutyniok. Microlocal Analysis of the Geometric Separation Problem. CoRR, abs/1004.3006, 2010.
6. J. Bobin, J. L. Starck, J. Fadili, Y. Moudden, and D. L. Donoho. Morphological Component Analysis: An Adaptive Thresholding Strategy. IEEE Transactions on Image Processing, 16(11):2675–2681, 2007.
7. S. Rambhatla and J. Haupt. Semi-Blind Source Separation via Sparse Represen- tations and Online Dictionary Learning. submitted to IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013. Online: arXiv:1212.0451 [cs.SD].
8. B. A. Olshausen and D. J. Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vision Research, 37(23):3311–3325, 1997.
9. M. Aharon, M. Elad, and A. Bruckstein. K-SVD: Design of Dictionaries for Sparse Representation. In Proceedings of SPARS05, pages 9–12, 2005.
10. J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online Learning for Matrix Fac- torization and Sparse Coding. Journal of Machine Learning Research, 11:19–60, 2010.
11. B. J. Frey and N. Jojic. Transformed Component Analysis: Joint Estimation of Spatial Transformations and Image Components. In Proceedings of the IEEE International Conference on Computer Vision, pages 1190–1196, 1999.
12. B. J. Frey and N. Jojic. Transformation-Invariant Clustering using the EM Algo- rithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25:1–17, 2003.
13. A. Vedaldi, G. Guidi, and S. Soatto. Joint Data Alignment Up to (Lossy) Trans- formations. In CVPR. IEEE Computer Society, 2008.
14. Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. RASL: Robust Alignment by Sparse and Low-Rank Decomposition for Linearly Correlated Images. In CVPR, pages 763–770. IEEE, 2010.

THANK YOU!