Speech Denoising via Nonnegative Matrix Factorization

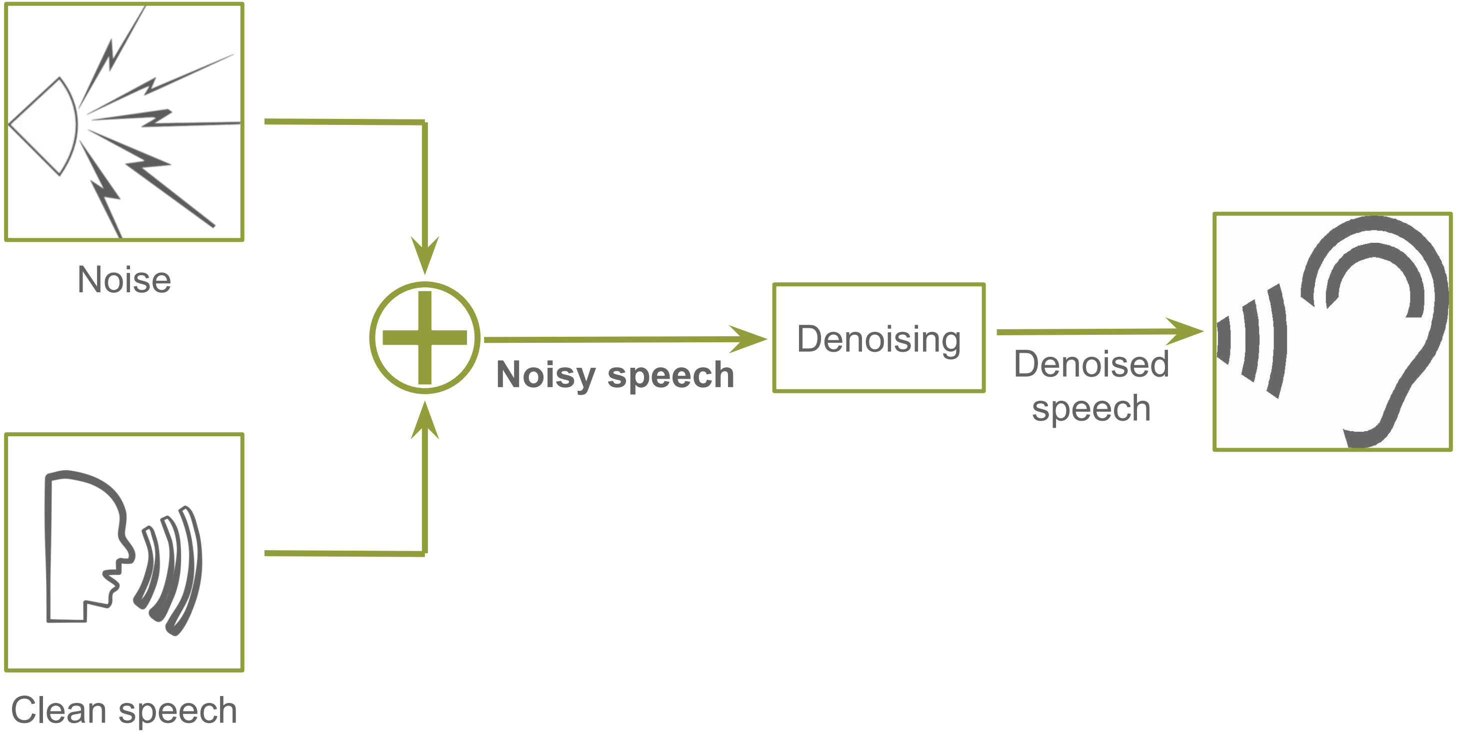
Сколково

Your Answer

Oleg Alenkin
Artyom Chashchin
Vladimir Chernykh
German Novikov

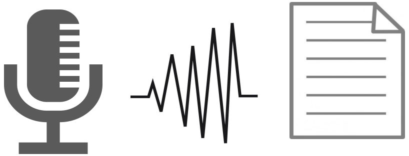
December 2016

Problem


Noise
Clean speech
Noisy speech
Denoising
Denoised speech

Applications

Automatic Speech Recognition



Telephone conversations

Hearing aids

Data

- CHiME Speech Separation and Recognition Challenge
  http://spandh.dcs.shef.ac.uk/chime_challenge/chime_download.html
  Recordings of WSJ utterances + 8 hours of noise

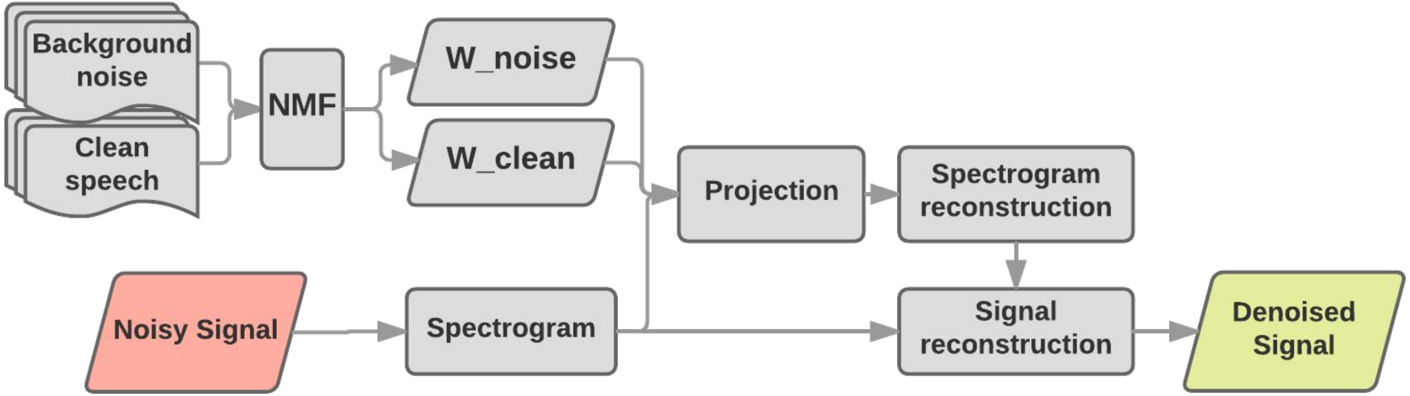
- Berlin Database of Emotional Speech
  http://www.emodb.bilderbar.info/download/
  Clean utterances

- Aurora noising
  http://aurora.hsnr.de/download.html

Skoltech
Skolkovo Institute of Science and Technology

Solution

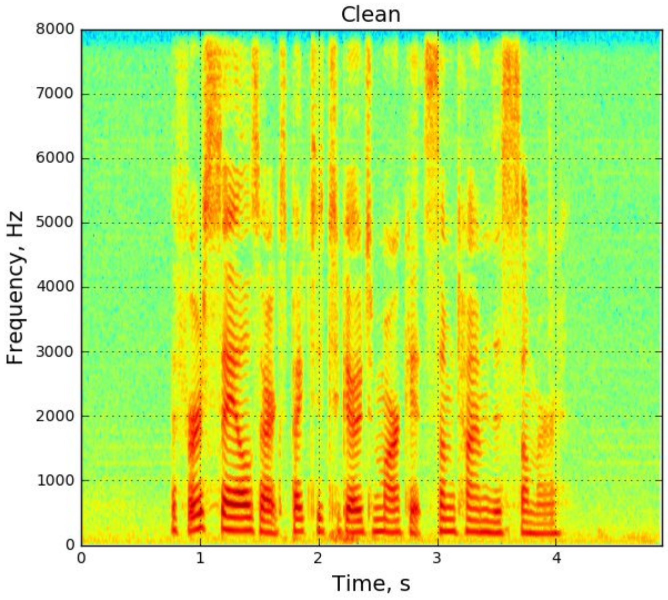
- Learn frequency patterns from speech and noise via NMF
- Decompose new signal on joint “dictionary” of patterns
- Take only projection corresponding to “clean speech”

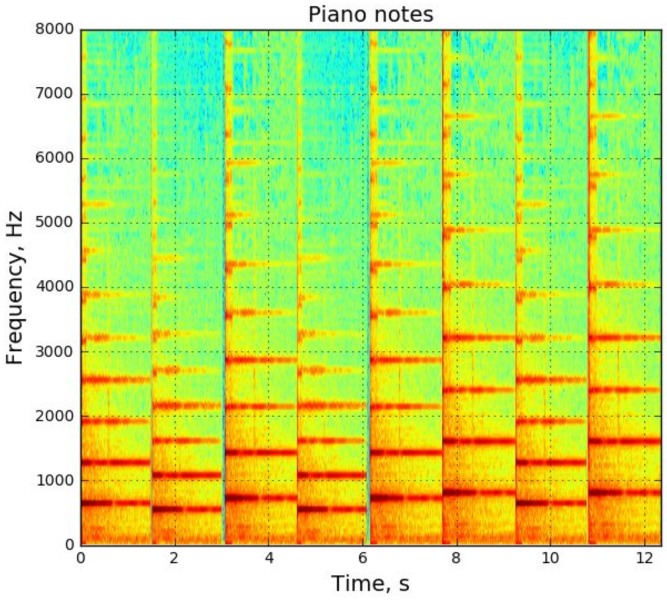


Skoltech
Skolkovo Institute of Science and Technology

Spectrogram

- Represent signal in Time-Frequency domain
- Built via Short-Term Fourier Transform - FFT with sliding overlapped window
- STFT - Complex spectrogram S ⇒ Amplitudes V

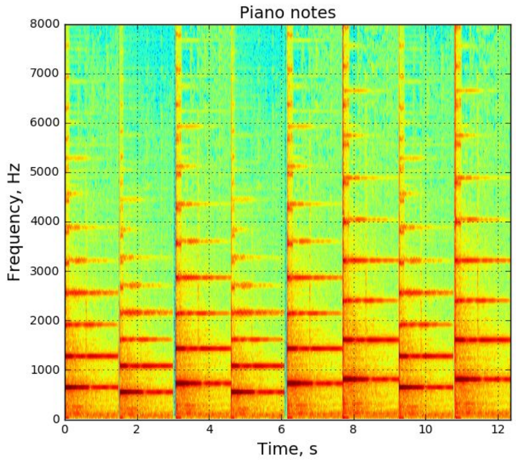


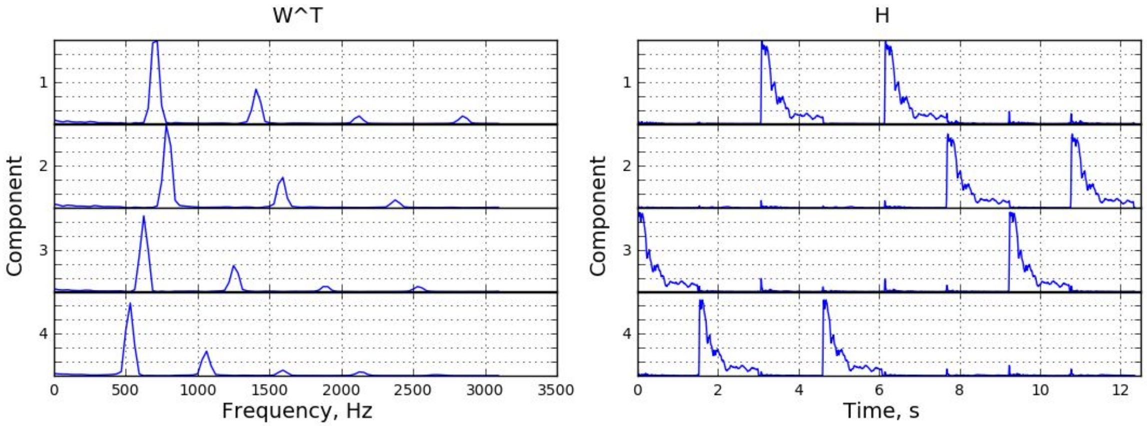


Skoltech
Skolkovo Institute of Science and Technology

NMF

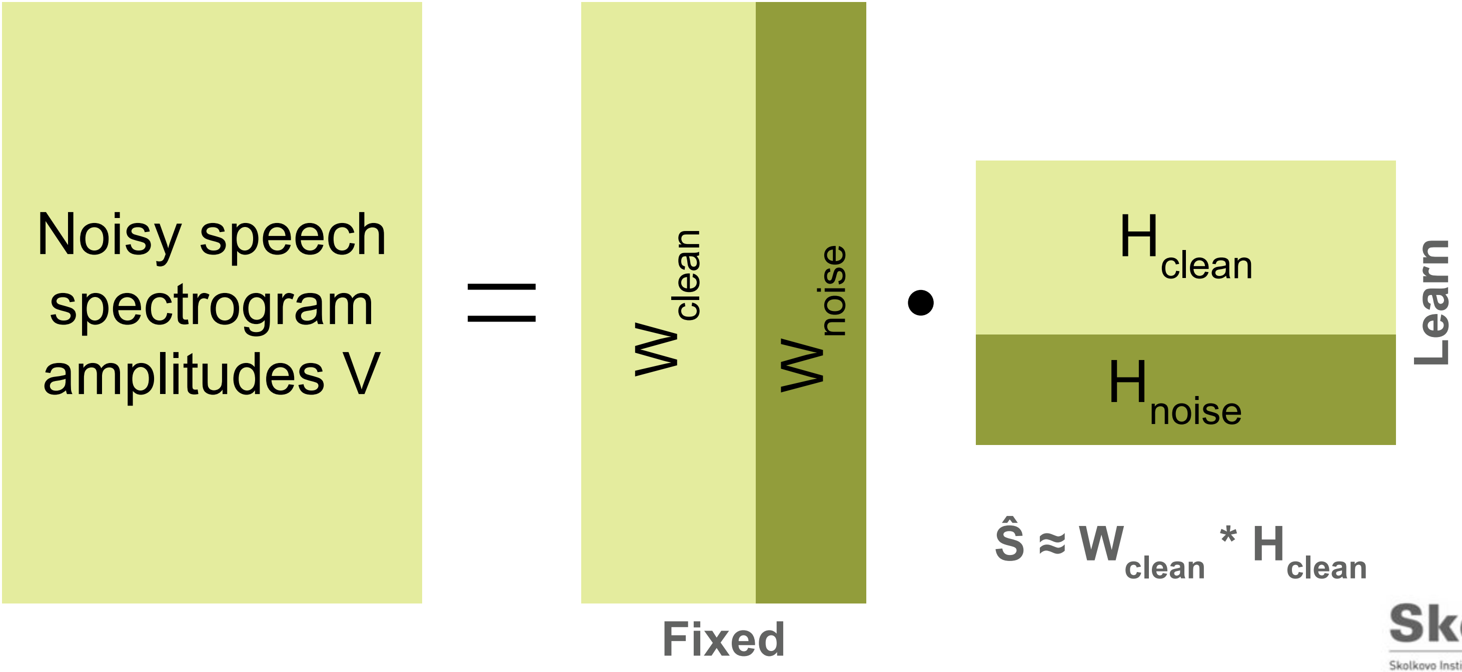
- Factorization **V ≈ W * H**, where V, W, H - real nonnegative
- Interpretation: W - frequency patterns, H - time-activation matrix
- Hidden dimension ≈ number of phonemes ≈ 40
- Learn speech and noise “building blocks”





Projection

- Join “dictionaries” - concatenate matrices W_{noise} and W_{clean}
- Project signal onto them



$$\hat{S} \approx W_{clean} * H_{clean}$$

NMF: how to compute

Optimization problem:

$$(W^*, H^*) = \text{argmin}_{W \geq 0, H \geq 0} D(V, WH)$$

$$D(P, Q) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(p_{ij}, q_{ij})$$

The most popular metrics:

$$d(p, q) = (p - q)^2$$ Frobenius norm

$$d(p, q) = p \ln(\frac{p}{q}) - p + q$$ KL divergence

Multiplicative Update Method:

$$[\nabla_H]_{kj} = \frac{\partial D(V, WH)}{\partial h_{kj}} = [\nabla_H^+]_{kj} - [\nabla_H^-]_{kj}$$

$$h_{kj} \leftarrow h_{kj} - \frac{h_{kj}}{[\nabla_H^+]_{kj}} \left([\nabla_H^+]_{kj} - [\nabla_H^-]_{kj}\right)$$

Skoltech

Skolkovo Institute of Science and Technology

NMF: how to compute

Alternating Nonnegative Least Squares:

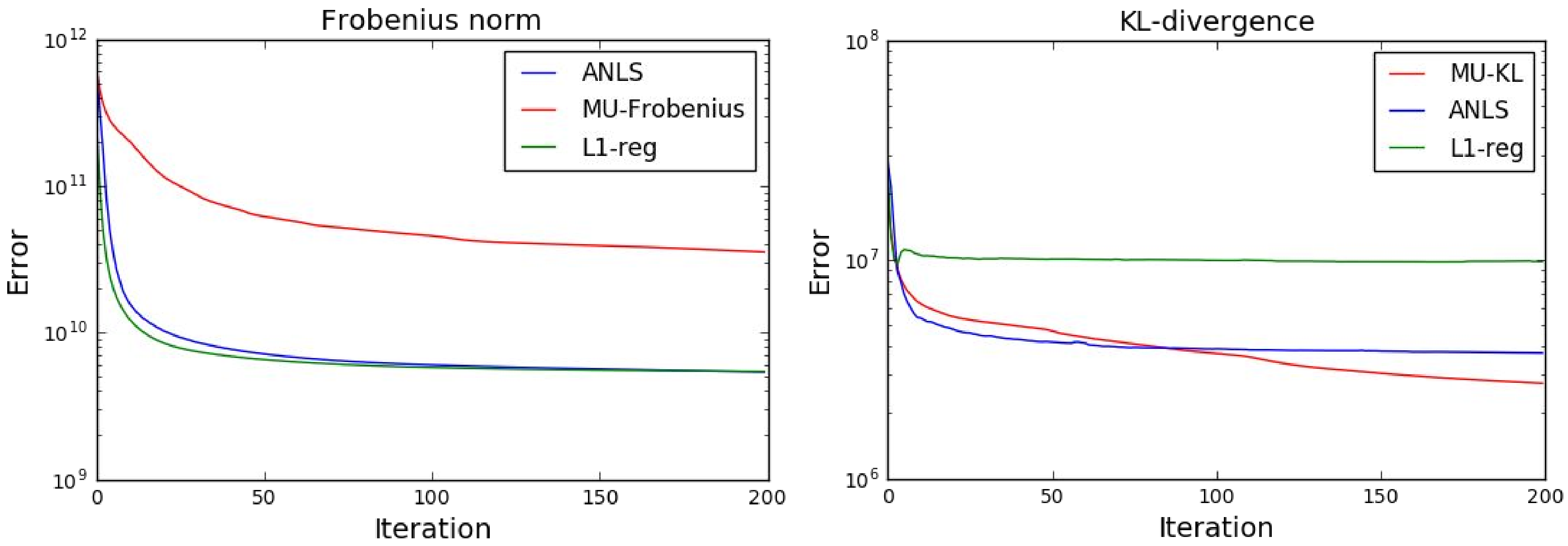
1) Initialize $W^1_{ia} \geq 0, H^1_{bj} \geq 0, \forall a, i, b, j.$

2) For k=1,2,...

$$W^{k+1} = \mathrm{argmin}_{W \geq 0} D\left(V, W H^k\right),$$
$$H^{k+1} = \mathrm{argmin}_{H \geq 0} D\left(V, W^{k+1} H\right).$$

Skoltech
Skolkovo Institute of Science and Technology

Methods’ Convergence



Skoltech
Skolkovo Institute of Science and Technology

Quasi-Newton method

$$(W^*, H^*) = \mathrm{argmin}_{W>0,H>0} D_{KL}(V, WH)$$

$$W \leftarrow \max(\varepsilon, W - H_W^{-1} \nabla_W D_{KL})$$

$$\nabla_W D_{KL} = H^T J_{M\times K} - H^T (V \oslash (WH))$$

$$H_W = \mathrm{diag}\{h_{W,m},\ m = 1, \dots, M\}$$

$$h_{W,m} = H \mathrm{diag}\{[V \oslash (Q \otimes Q)]_{m,:}\} H^T$$

$$H \leftarrow \max(\varepsilon, H - H_H^{-1} \nabla_H D_{KL})$$

$$\nabla_H D_{KL} = J_{M\times K} W^T - (V \oslash (WH)) W^T$$

$$H_H = \mathrm{diag}\{h_{H,k},\ k = 1, \dots, K\}$$

$$h_{H,k} = W^T \mathrm{diag}\{[V \oslash (Q \otimes Q)]_{:,k}\} W$$

Problem: hard to compute inverse Hessian

Signal reconstruction

- Naive method with zero phase:

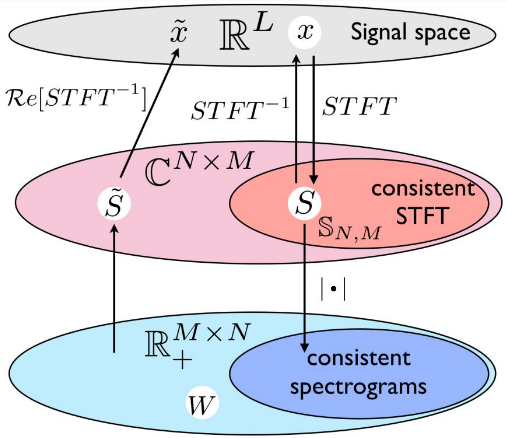
$$\hat{X} = STFT^{-1}(\hat{S})$$

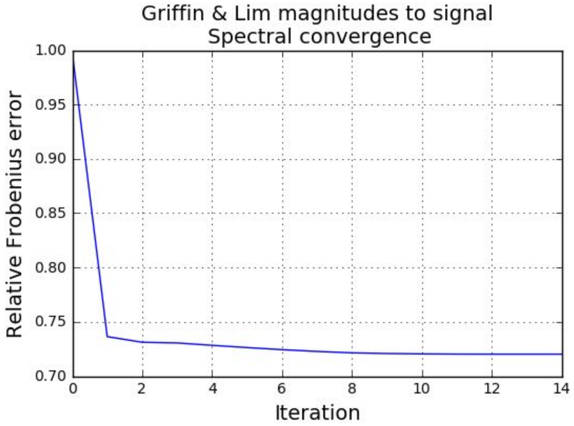
- Noisy signal phases:

$$\hat{X} = STFT^{-1}(\hat{S} \times \exp\left(i\angle STFT(X_{noisy})\right))$$

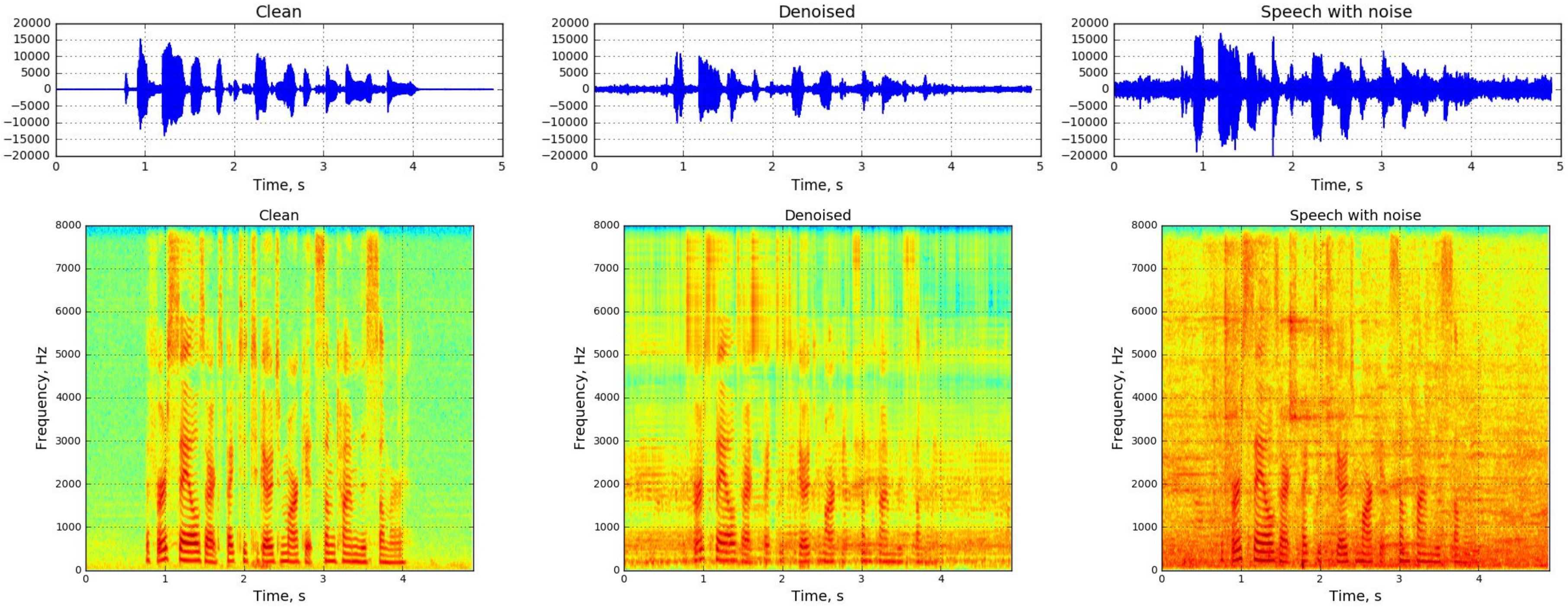
- Griffin & Lim iterative method:

$$\hat{X}_n = STFT^{-1}(\hat{S} \times \exp(i\angle STFT(\hat{X}_{n-1})))$$





Demo



Skoltech
Skolkovo Institute of Science and Technology