Reduction of non-stationary noise using a non-negative latent variable decomposition

Mikkel M. Schmidt and Jan Larsen
DTU Informatics
Technical University of Denmark
www.imm.dtu.dk

Noise Reduction

- Single channel recording
- Unknown speaker / signal of interest
- Focus on modeling noise

Single channel source separation is hard

- The is no spatial information hence
  - beamforming
  - independent component analysis

are not feasible

- Maybe higher-level cognitive capabilties such as *context detection* could help

- We will use a data-driven approach to learn a good noise representation

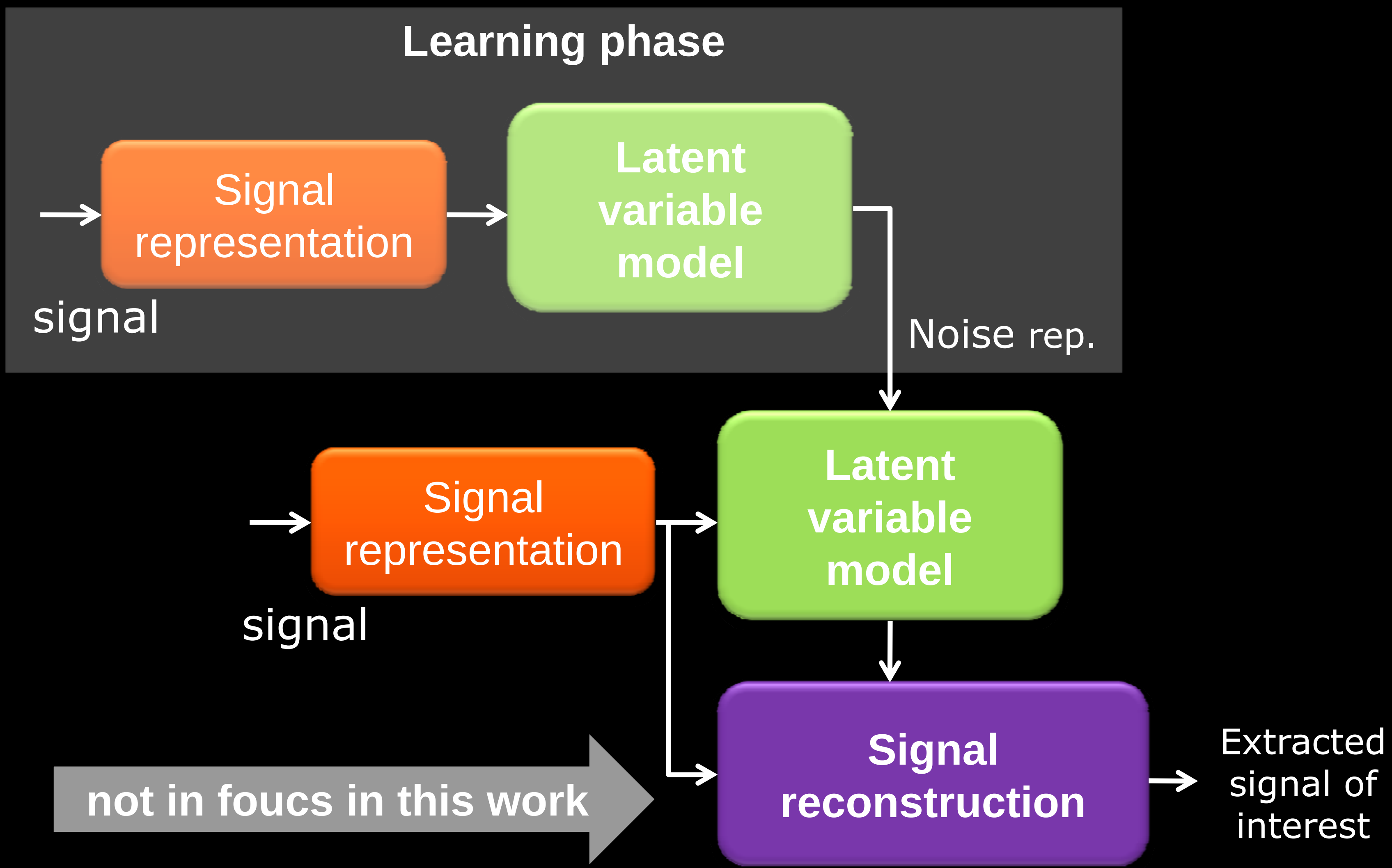
The spectrum of alternative methods

- Wiener filter (Wiener, 1949)
- Spectral subtraction (Boll 1979; Berouti et al. 1979)
- AR codebook-based spectral subtraction (Kuropatwinski & Kleijn 2001)
- Minimum statistics (Martin et al. 1994,2001, 2005)
- Masking techniques (Wang; Weiss & Ellis 2006)
- Factorial models (Roweis 2000,2003)
- Non-negative sparse coding (Casey&Westner 2000, Wang&Plumbley2006, Schmidt&Olsson 2006, Schmidt,Larsen&Hsiao 2007)

Several method requires a VAD

Largely fail for fast changing non-stationary noise

Our approach in brief



Signal Representation

- Exponentiated magnitude spectrogram

$$X = |\mathrm{STFT}\{x(t)\}|^{\gamma}$$

$\gamma = 2$ Power spectrogram

$\gamma = 1$ Magnitude spectrogram

$\gamma = 0.67$ Cube root compression

(Steven's power law - perceived intensity)

- Any other representation could be used – wavelets, perceptually weighted etc.
- Ignores phase information. Reconstruct by re-filtering

Non-negative latent variable model

- Speech

Binary activation

$$\boldsymbol{s}^{(i)} = a^{(i)} \sum_{k=1}^{K_s} \boldsymbol{\sigma}_k b_k^{(i)}$$

Non-negative basis

Weights

- Noise

$$\boldsymbol{n}^{(i)} = \sum_{k=1}^{K_n} \boldsymbol{\nu}_k c_k^{(i)}$$

- Noisy speech

Residual

$$\boldsymbol{x}^{(i)} = \boldsymbol{s}^{(i)} + \boldsymbol{n}^{(i)} + \boldsymbol{r}^{(i)}$$

Non-negative latent variable model

$$X = SBA + NC + R,$$

- Use a probabilistic Bayesian setting
- Exponential priors (sparsity) on ***B*** and ***C***
- Gaussian residual ***R***

Goals:

Posterior of all parmeters ***A***, ***B***, ***C***, ***S***, ***N*** and noise variance

Marginalized mean estimate of signal component

A three-step simple approximate learning procedure for speech

1. Compute speech activation using state-of-the-art voice activity detector (Qualcomm-ICSI-OGI)
2. Compute noise basis representation using non-speech signal frames
3. Jointly compute noise weights, speech basis, and speech weights and reconstruct speech signal

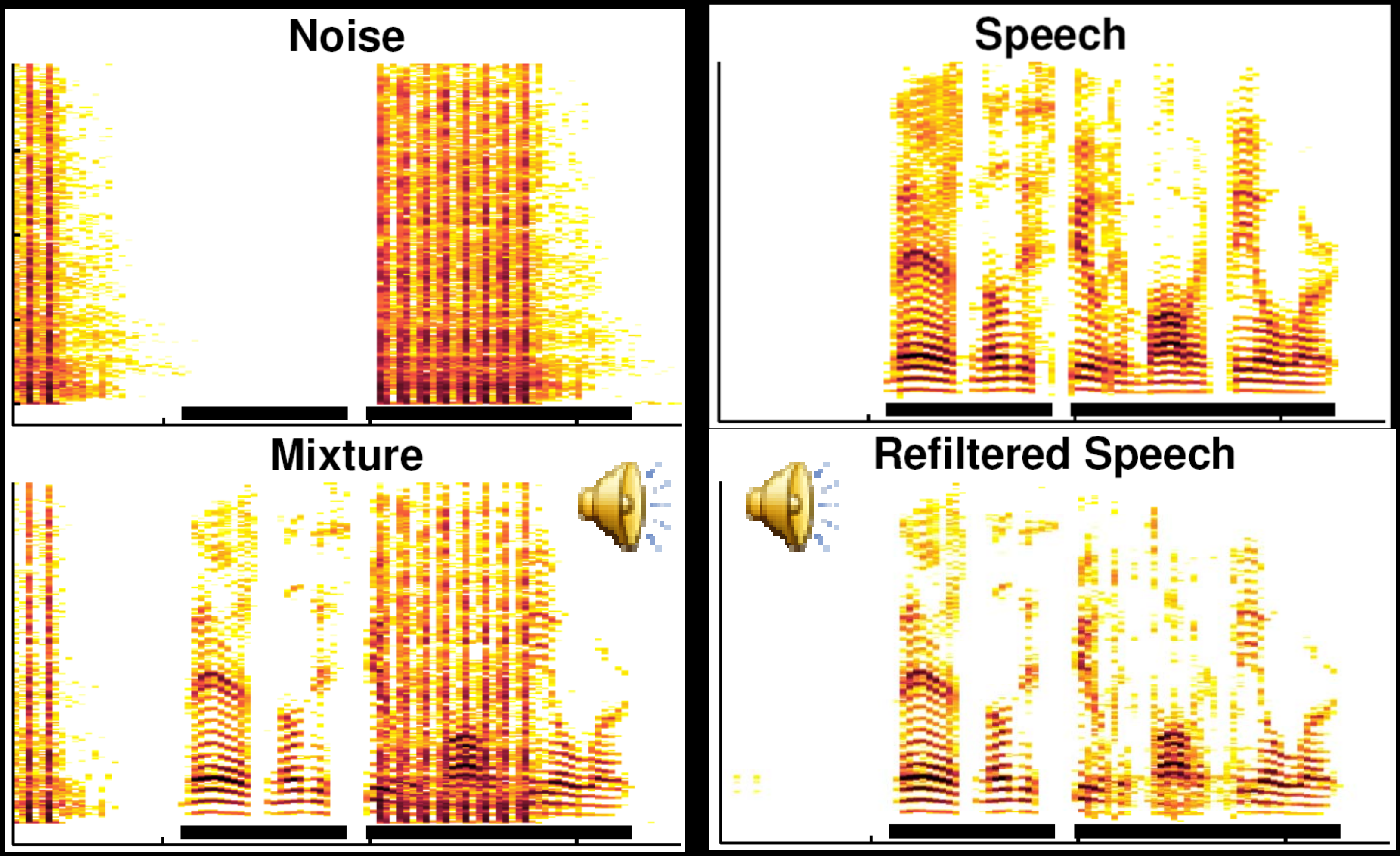
Experimental setup

- Four different noise types: machine gun, string quartet, restaurant noise, traffic noise
- Mixed by 100 sentences from the TIMIT database with SNR in the range -9dB to 6dB
- Signal represented by SFTF using 64ms 50% overlapping Hann windowed frames and mapped onto 32 MEL frequency bins [20Hz; 4kHz]
- 256 bases for signal and noise
- Optimal sparsity (hyper-parameters): $\lambda_B=0.1$, $\lambda_C=0$
- Qualcomn-ICSI-OGI voice activity detector (VAD)

Quality Measure

- Signal to noise ratio
  - Simple measure, has only indirect relation to perceived quality
- Representation-based metrics
  - In systems based on time-frequency masking, evaluate the masks
- Perceptual models
  - Promising to use PEAQ or PESQ
- High-level Attributes
  - For example word error rate in a speech recognition setup
- Listening-tests
  - Expensive, time-consuming, aspects (comfort, intelligibility)

Example: Bursts of machine gun shots



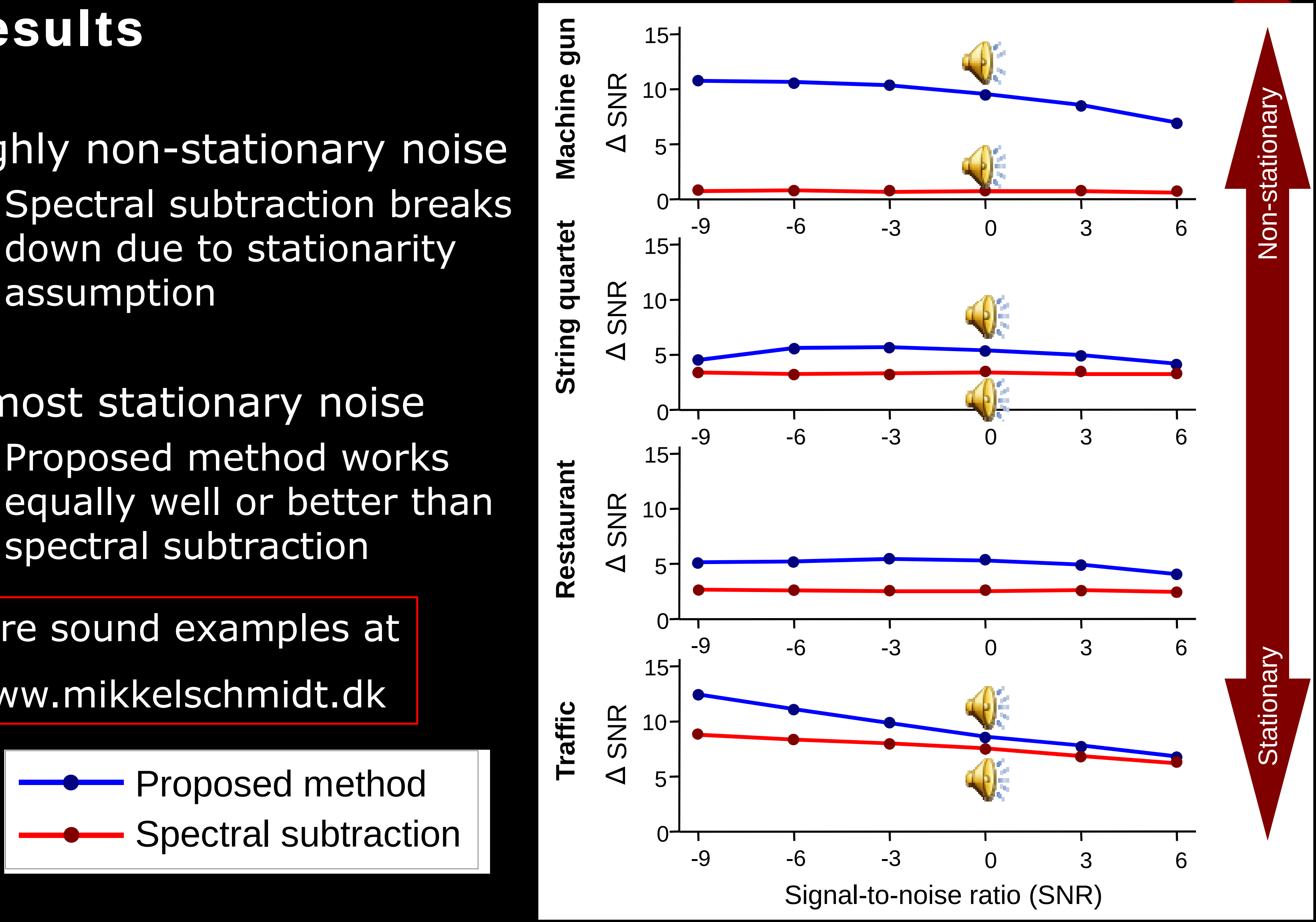
Results

- Highly non-stationary noise
  - Spectral subtraction breaks down due to stationarity assumption

- Almost stationary noise
  - Proposed method works equally well or better than spectral subtraction

More sound examples at

www.mikkelschmidt.dk



Potential ways ahead

- Prior model speech activity pattern, e.g. using HMM
- Harmonic prior for speech basis
- Full Bayesian inference in the model (working on Gibbs sampling approach)
- Better residual models be including phase uncertainty (Rayleigh distribution) (Parry and Essa 2007)
- Advanced post-processing (weighted, thresholded spectral subtraction, and smoothing) can help

Conclusion

- A probabilistic non-negative latent variable decomposition method was presented
- A full Bayesian inference is possible, we resorted to simple a step-wise procedure
- The method has potential over classical methods for handling very non-stationary strong noise conditions
- As an essential output of the method is a good noise estimate, it can also be used as an integral part of other methods we are based on noise estimation.

Thank you for your attention!