UNIVERSITÉ DE SHERBROOKE

IMSI — INSTITUT DES MATÉRIAUX ET SYSTÈMES INTELLIGENTS — INTELLIGENT MATERIALS AND SYSTEMS INSTITUTE

LABORIUS

Microphone Array Post-Filter for Separation of Simultaneous Non-Stationary Sources

Jean-Marc Valin, Jean Rouat, François Michaud

Department of Electrical Engineering and Computer Engineering
Université de Sherbrooke, Québec, Canada
Jean-Marc.Valin@USherbrooke.ca

CANADA RESEARCH CHAIRS
CHAIRES DE RECHERCHE DU CANADA

Fondation canadienne pour l'innovation

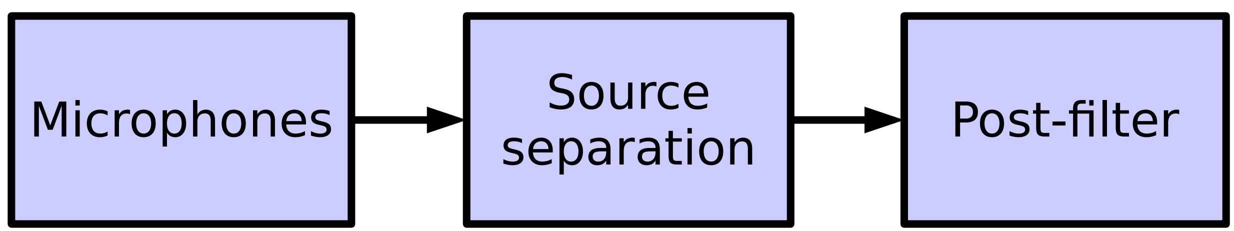
NSERC CRSNG

Motivations

The context: sound source separation

The problem: beamforming and similar techniques provide limited noise reduction

The solution: use a post-filter to further reduce noise and interference



Approach

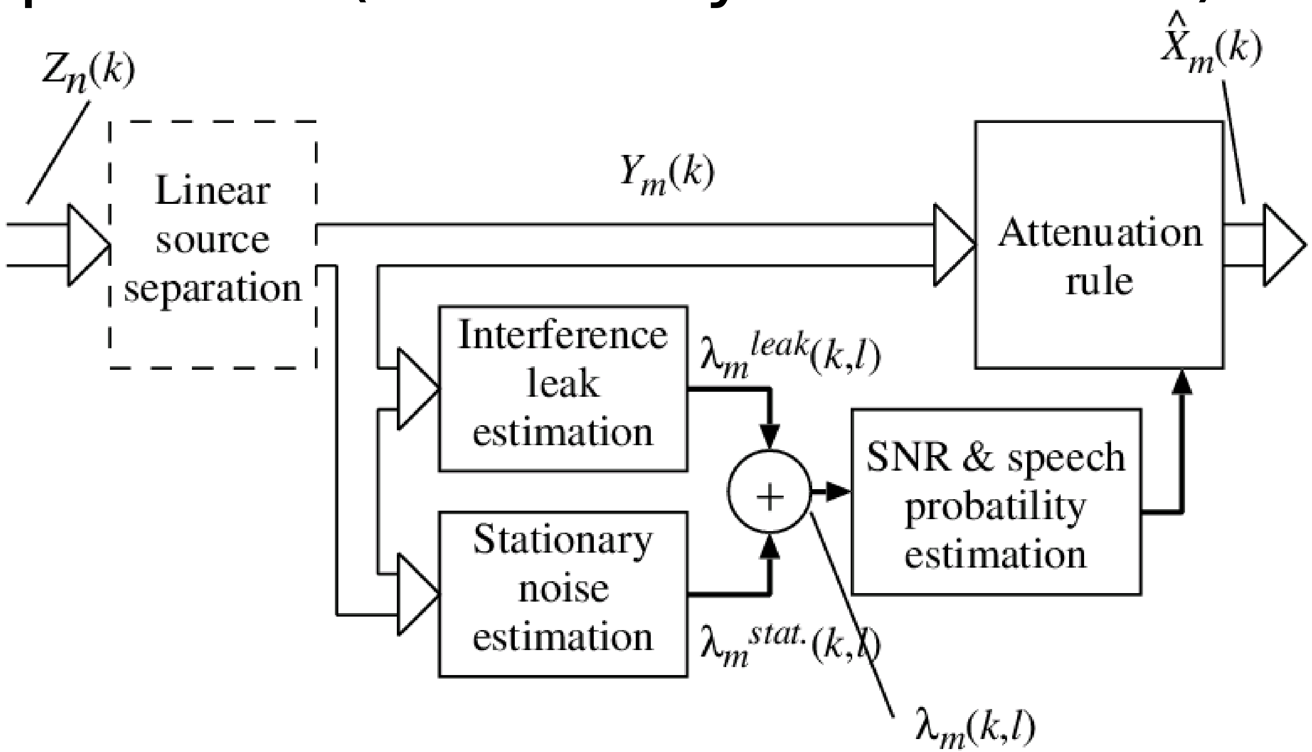
- Linear source separation
  - Geometric Source Separation (Parra) is used
  - Works for any linear separation algorithm
- Post-filter
  - Frequency-domain processing
  - Based on the optimal Ephraim and Malah estimator
  - Gain modification according to probability of speech presence (Cohen)

Contribution

- Multiple sources of interest
  - Generalize post-filters to separation of multiple sources
- Non-stationary noise
  - Decouple background noise (stationary) and directional interference (transient)
  - Fast estimation of interference
    - Direct estimation from initial separation

Post-Filter Overview

Noise estimate as the sum of two components (stationary + transient)



Background Noise Estimation

- Minima-Controlled Recursive Average (Cohen)
  - Applied for each source of interest
- Initial estimate provided directly from the microphones

$$\lambda_m^{stat.}(k,\ell_0) = \frac{1}{N^2}\sum_{n=0}^{N-1}\sigma_{x_n}^2(k)$$

Interference Estimation

- Source separation leaks
  - Incomplete adaptation
  - Inaccuracy in localization
  - Reverberation
  - Imperfect microphones
- Estimation from other separated sources

$$\lambda_m^{leak}(k,l) = \eta \sum_{i=0, i \neq m}^{M-1} S_i(k,l)$$

$$S_m(k,l) = \alpha_s S_m(k,l-1) + (1-\alpha_s) Y_m(k,l)$$

Suppression Rule

Loudness-domain optimal estimator

$$\hat{X}_m(k,l) = G_m(k,l) Y_m(k,l)$$

Assuming speech is present:

$$G_{H_1}(k) = \frac{\sqrt{\upsilon(k)}}{\gamma(k)} \left[\Gamma\left(1 + \frac{\alpha}{2}\right) M\left(-\frac{\alpha}{2}; 1; -\upsilon(k)\right)\right]^{\frac{1}{\alpha}}$$

$$\gamma(k) \triangleq |Y(k)|^2 / \lambda(k) \qquad \xi(k) \triangleq E\left[|X(k)|^2\right] / \lambda(k)$$

$$\upsilon(k) \triangleq \gamma(k)\xi(k) / \left(\xi(k) + 1\right)$$

Speech Presence Uncertainty

Optimal gain modification for loudness-domain

$$G(k) = \left[p(k)G_{H_1}^{\alpha}(k) + (1-p(k))G_{min}^{\alpha}\right]^{\frac{1}{\alpha}}$$

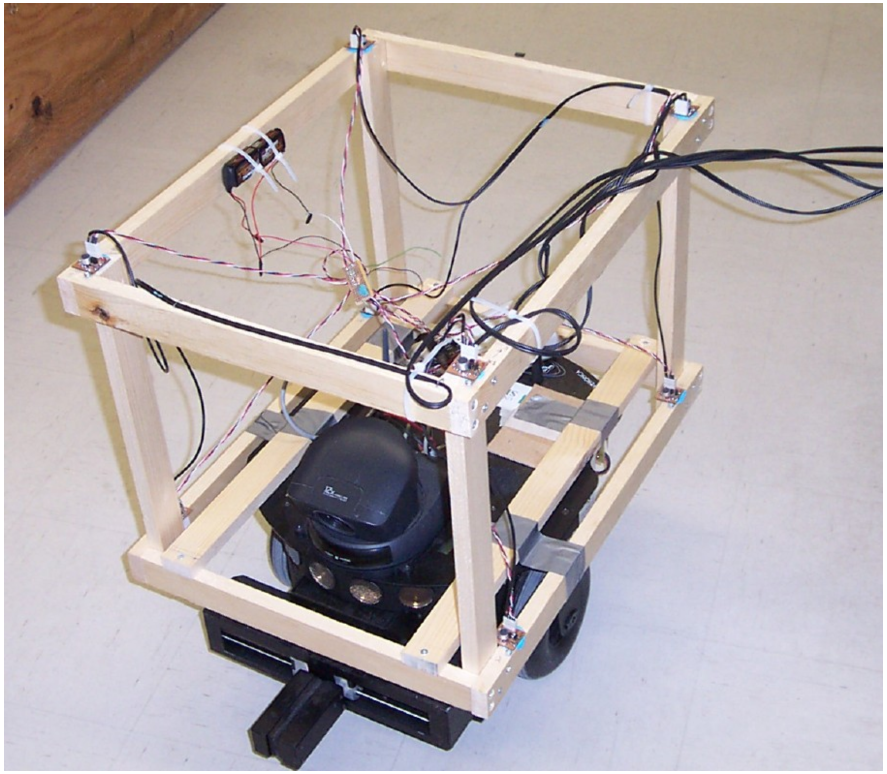
Setting $G_{min} = 0$ leads to

$$G(k) = p^2(k)G_{H_1}(k)$$

Unlike log-domain estimator, no arbitrary limit on attenuation

Experimental Setup

- Array of 8 inexpensive microphones on a mobile robot
- Automatic localization
- Noisy conditions
- Moderate reverberation



Results (Signal-to-Noise Ratio)

Three voices recorded separately so clean signal is available

SNR (dB)	female 1	female 2	male 1
Microphone input	-1.8	-3.7	-5.2
GSS only	9.0	6.0	3.7
GSS+single channel	9.9	6.9	4.5
GSS+proposed	12.1	9.5	9.4

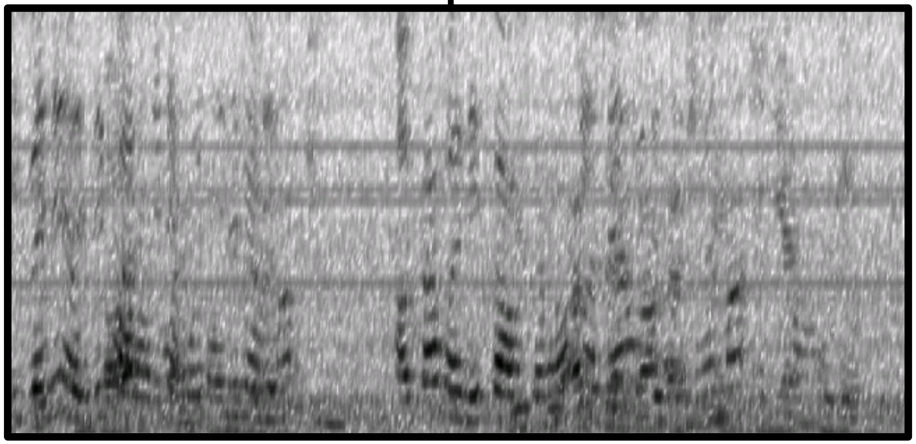
Results (Log-Spectral Distortion)

$$LSD = \frac{1}{L}\sum_{l=0}^{L-1}\left[\frac{1}{K}\sum_{k=0}^{K-1}\left(20\log_{10}\frac{|X(k,l)|+\epsilon}{\left|\hat{X}(k,l)\right|+\epsilon}\right)^2\right]^{\frac{1}{2}}$$

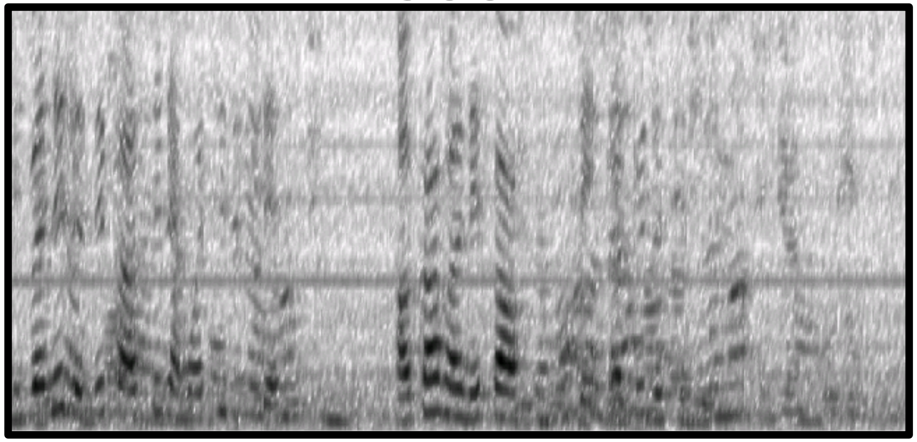
LSD (dB)	female 1	female 2	male 1
Microphone input	17.5	15.9	14.8
GSS only	15.0	14.2	14.2
GSS+single channel	9.7	9.5	10.4
GSS+proposed	6.5	6.8	7.4

Results (spectrograms)

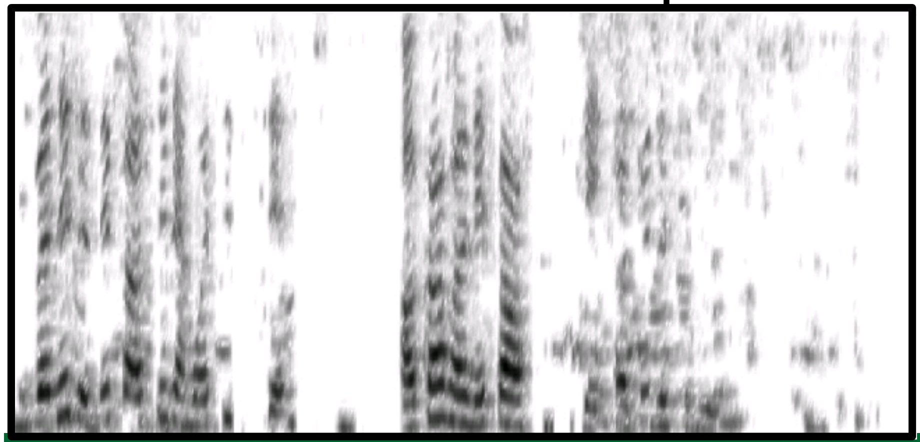
Input



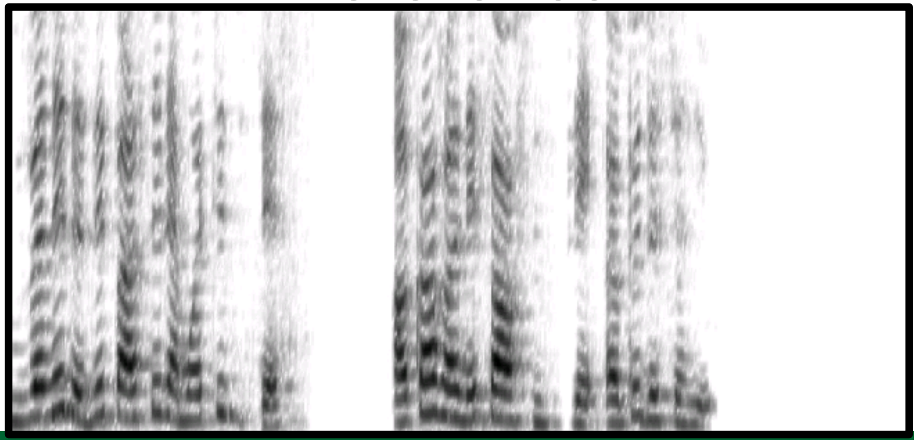
GSS



Post-filter output



Reference



Conclusion

- Source separation post-filter
  - Based on optimal loudness-domain estimator
  - Interference estimated using other sources
- Future work
  - Robustness to reverberation
    - original processed
  - Integration with speech recognition

Questions?