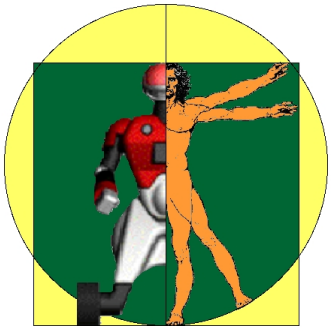
UNIVERSITÉ DE SHERBROOKE

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

LABORIUS



Enhanced Robot Audition Based on Microphone Array Source Separation with Post-Filter

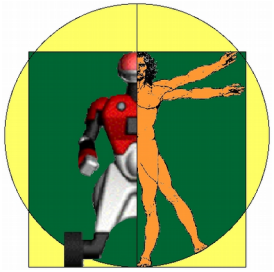
Jean-Marc Valin, Jean Rouat, François Michaud

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

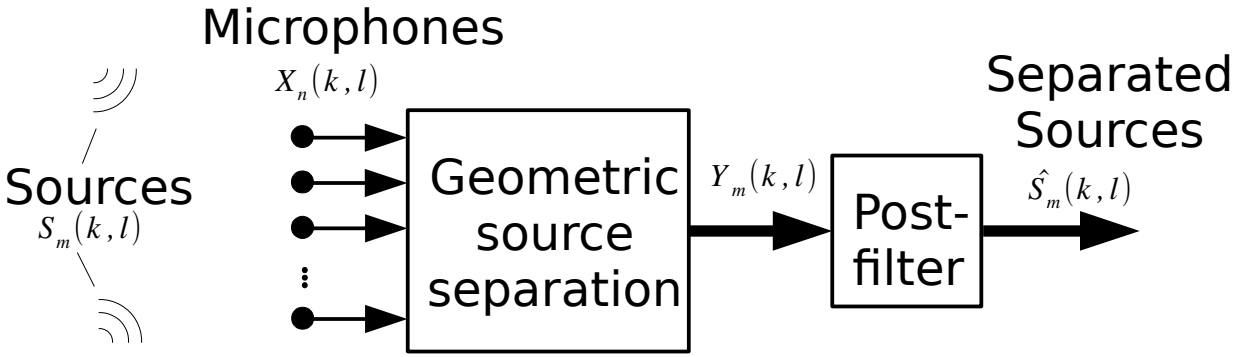


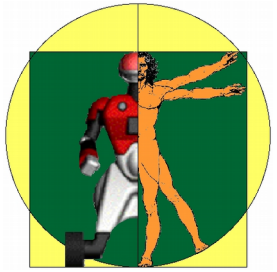
Motivations

The context: mobile robot and cocktail party effect

The problem: separating sound sources

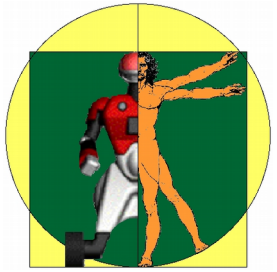
The solution: microphone array with both linear and non-linear processing





Approach

- Frequency-domain processing
- Geometric Source Separation (GSS)
  - Minimize leakage under constraints
  - Adapted for real-time processing
- Post-filter
  - Cancels remaining interferences
  - Based on Ephraim and Malah estimator
  - Handles both stationary and non-stationary noise/interference



Geometric Source Separation

Frequency domain:

$$\mathbf{x}(k) = \mathbf{A}(k)\mathbf{s}(k) + \mathbf{n}(k)$$

Constrained optimization $\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k)$

- Minimize correlation of the outputs:

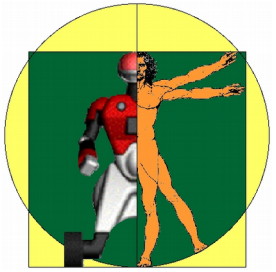
$$J_1(\mathbf{W}(k)) = \|\mathbf{R}_{\mathbf{yy}}(k) - \mathrm{diag}\,[\mathbf{R}_{\mathbf{yy}}(k)]\|^2$$

- Subject to geometric constraint:

$$J_2(\mathbf{W}(k)) = \|\mathbf{W}(k)\mathbf{A}(k) - \mathbf{I}\|^2$$

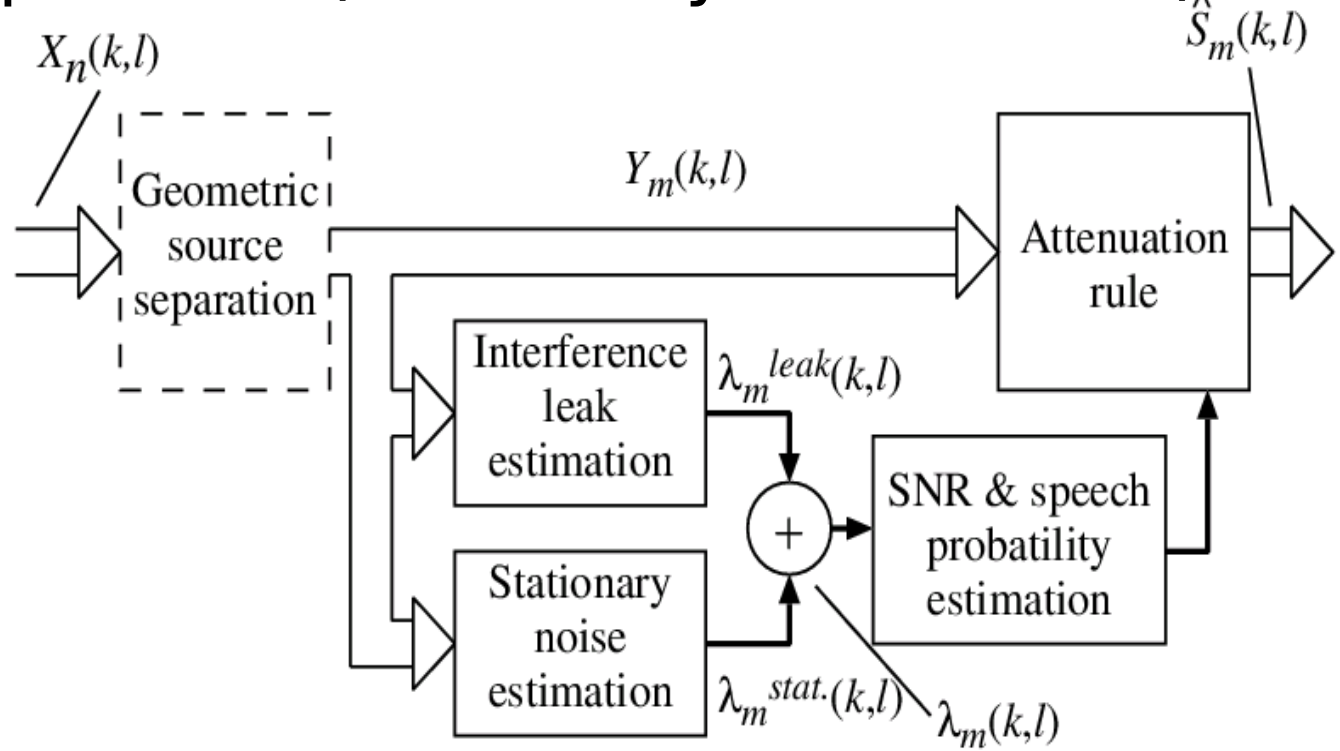
Modifications to original GSS algorithm

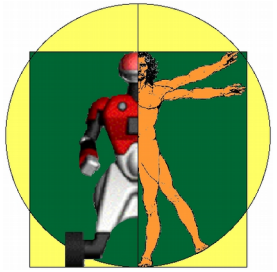
- Instantaneous computation of correlations
- Stochastic-gradient descent



Post-Filter Overview

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

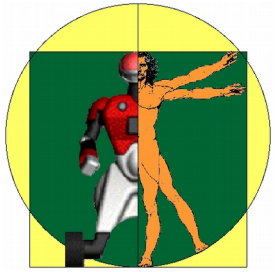




Background Noise Estimation

- Minima-Controlled Recursive Average (Cohen)
  - Noise estimate is adapted during quiet periods
  - Applied for each source of interest
- Initial estimate provided directly from the microphones

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

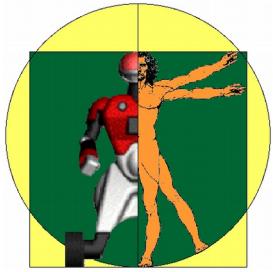


Interference Estimation

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

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

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



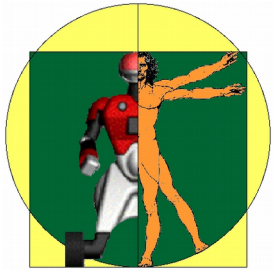
Suppression Rule

Ephraim & Malah spectral estimator

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

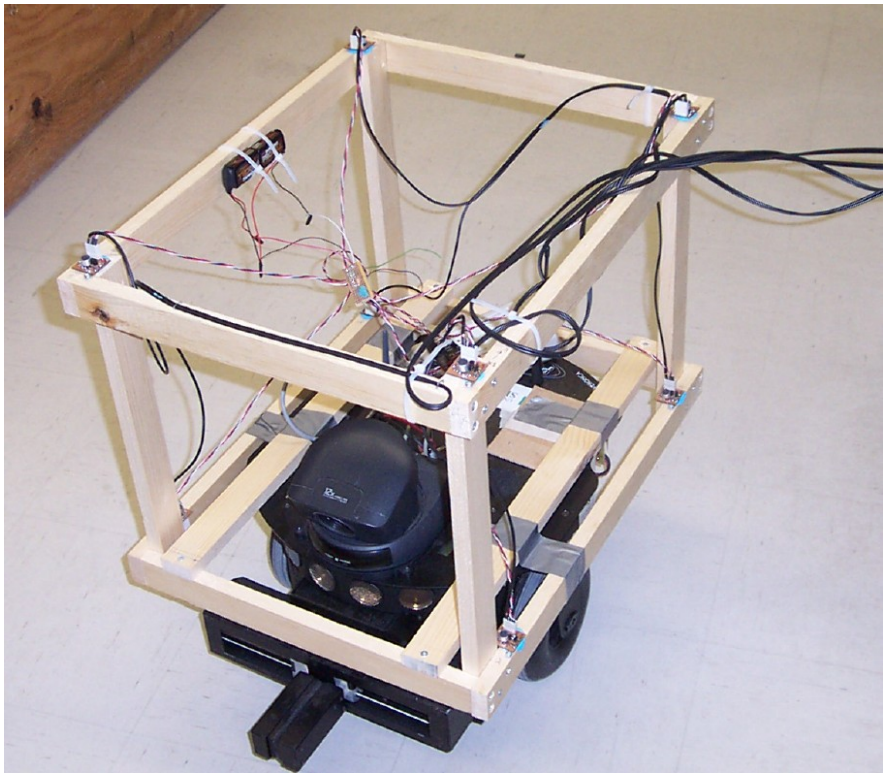
Gain is modified to take into account probability of source being present (Cohen)

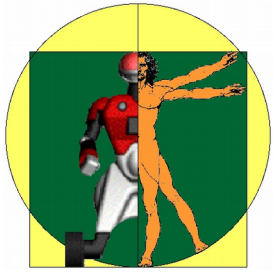
$$G(k) = p^2(k)G_{H_1}(k)$$



Experimental Setup

- Array of 8 inexpensive microphones on a Pioneer2 robot
- Automatic localization
- Noisy conditions
- 350 ms reverberation time

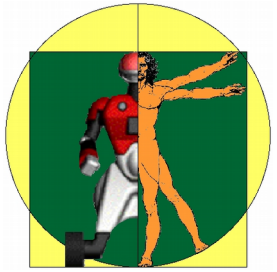




Results (Signal-to-Noise Ratio)

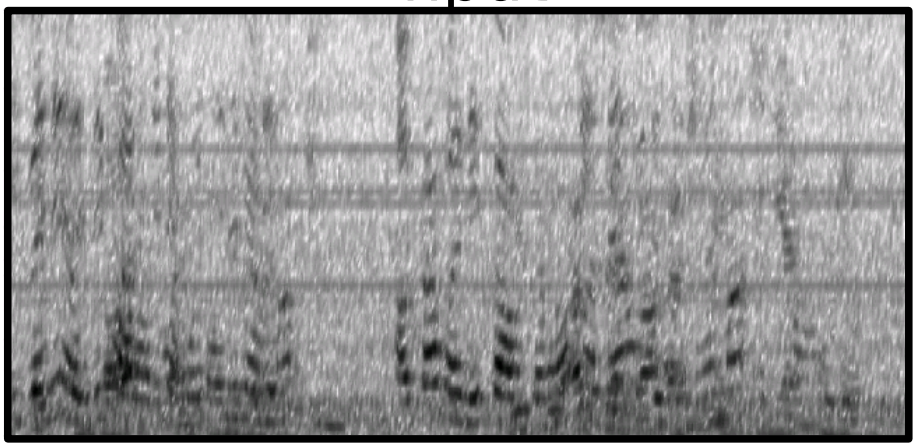
Three voices recorded separately so clean signal is available

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

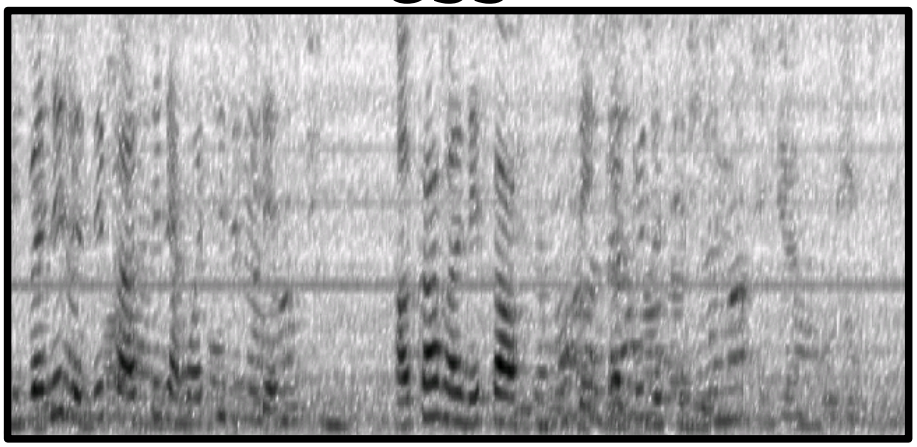


Results (spectrograms)

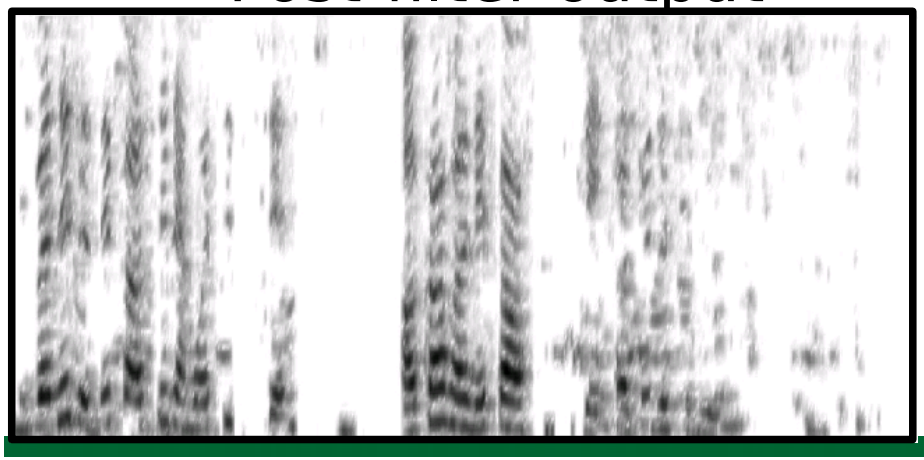
Input



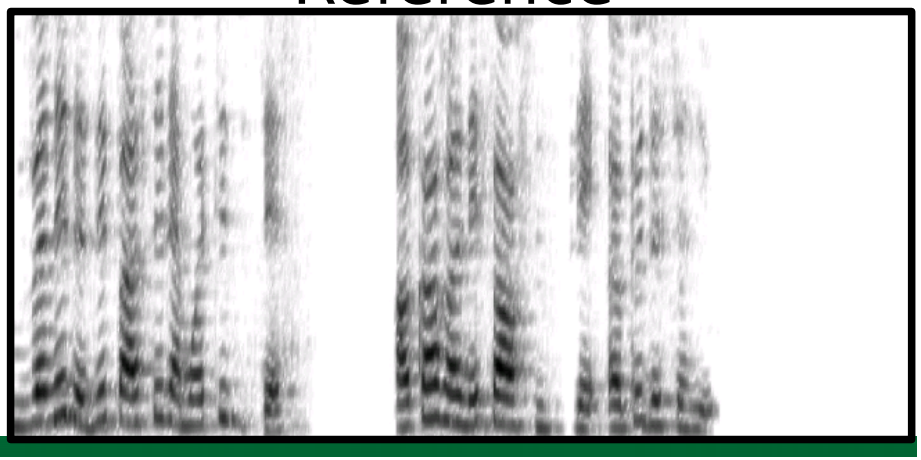
GSS

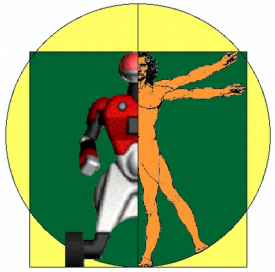


Post-filter output



Reference





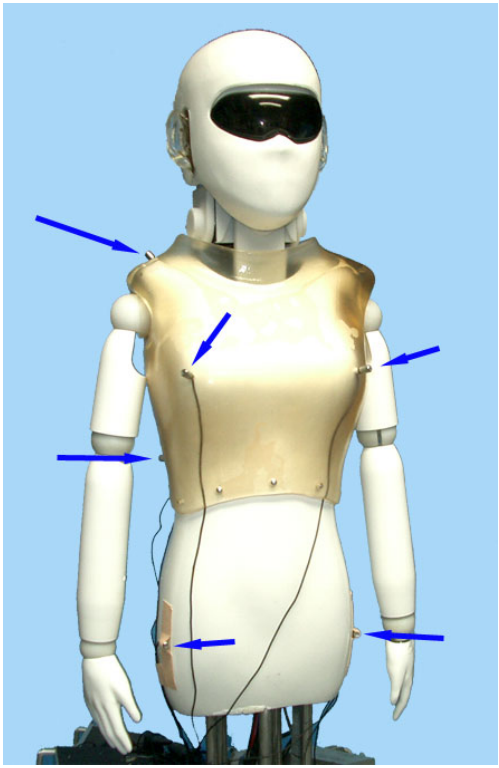
Results (recognition with post-filter)

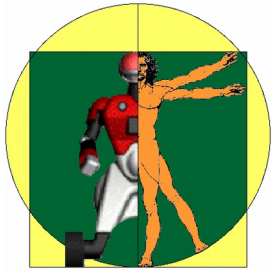
Japanese isolated word recognition (SIG2 robot)

- 3 simultaneous sources
- 200 word vocabulary
- 90 degrees separation

	GSS only	GSS+pf
mixed		
right	66%	71%
left	15%	21%
center	41%	53%

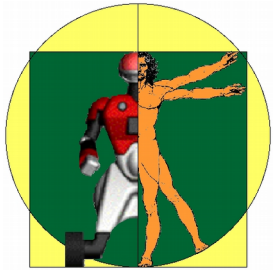
14% reduction in error rate





Conclusion

- Geometric Source Separation
  - Real-time minimization of leakage
- Source separation post-filter
  - Interference estimated using other sources
- Future work
  - Robustness to reverberation
    - original
    - processed
  - Better integration with speech recognition
    - Using the post-filter to estimate ASR feature reliability



Questions?