Localization of Simultaneous Moving Sound Sources for Mobile Robot Using a Frequency-Domain Steered Beamformer Approach

Jean-Marc Valin, François Michaud, Brahim Hadjou, Jean Rouat

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

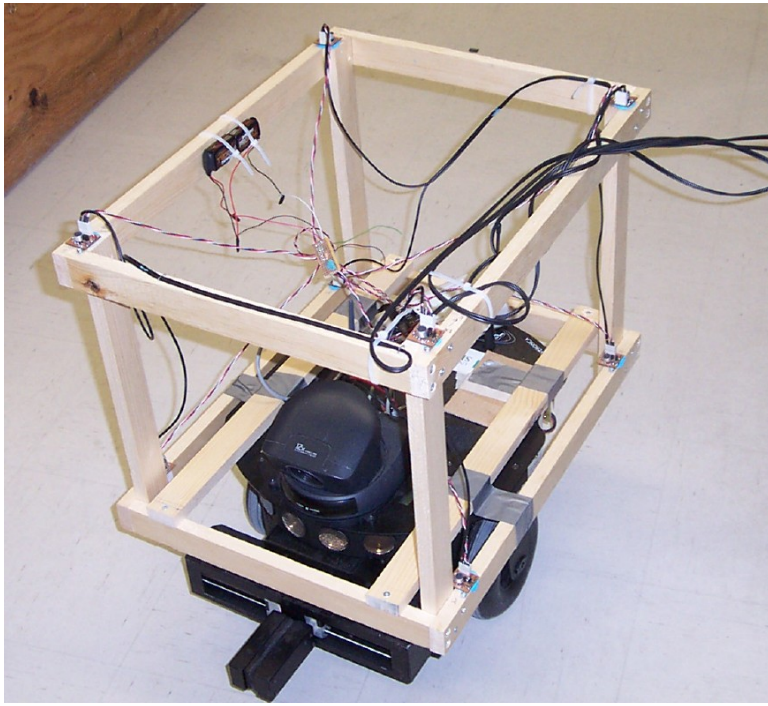
Approaches to Sound Source Localization

Binaural audition

- Two microphones
- Interaural phase difference
- Interaural intensity difference
- Imitate human auditory system

Microphone array audition

- Larger number of microphones
- Phase difference only
- Increased redundancy compensating for high complexity of human audition



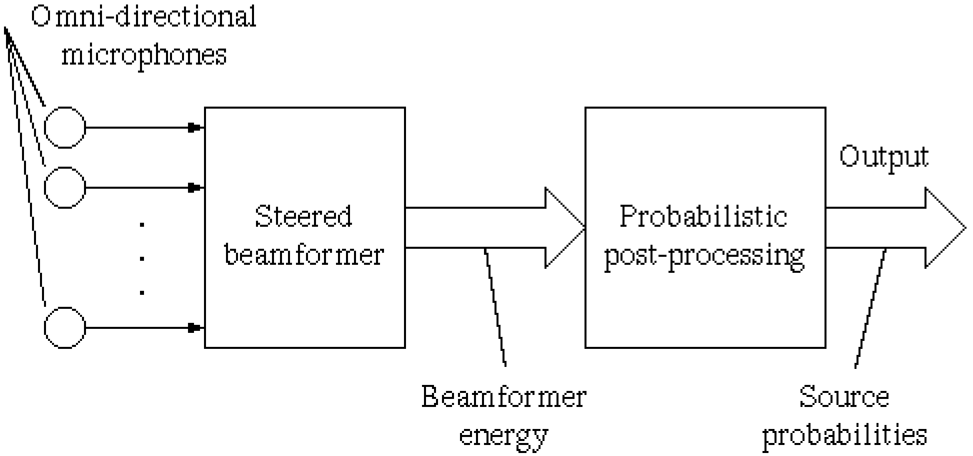
Approach Overview

Sounds arrive at microphones with different delays (depending on distance)

- Hypothesis: point sound sources

Steered beamformer: scans all directions for energy peaks

Probabilistic post-processing: applies Bayesian inference



Steered Beamformer

Delay-and-sum beamformer

$$y(n) = \sum_{m=0}^{M-1} x_m\left(n - \tau_m\right)$$

Beamformer energy

$$\begin{aligned} E &= \sum_{n=0}^{L-1} \left[y(n)\right]^2 \\ &= \sum_{n=0}^{L-1} \left[x_0\left(n-\tau_0\right) + \ldots + x_{M-1}\left(n-\tau_{M-1}\right)\right]^2 \end{aligned}$$

Frequency Domain Computation

$$
\begin{aligned}
E &= \sum_{m=0}^{M-1} \sum_{n=0}^{L-1} x_m^2 (n - \tau_m) \\
&+ 2 \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{m_1-1} \sum_{n=0}^{L-1} x_{m_1}(n - \tau_{m_1}) x_{m_2}(n - \tau_{m_2})
\end{aligned}
$$

$$
E = K + 2 \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{m_1-1} R_{x_{m_1}, x_{m_2}} (\tau_{m_1} - \tau_{m_2})
$$

$$
R_{ij}(\tau) \approx \sum_{k=0}^{L-1} X_i(k) X_j(k)^* e^{\jmath 2\pi k \tau / L}
$$

Spectral Weighting

- Cross-correlation peaks are very wide
  - Poor angular accuracy
  - Overlap between close sources
- Solution: spectral weighting
  - Whiten spectrum
  - Give less weight to noisy regions of spectrum

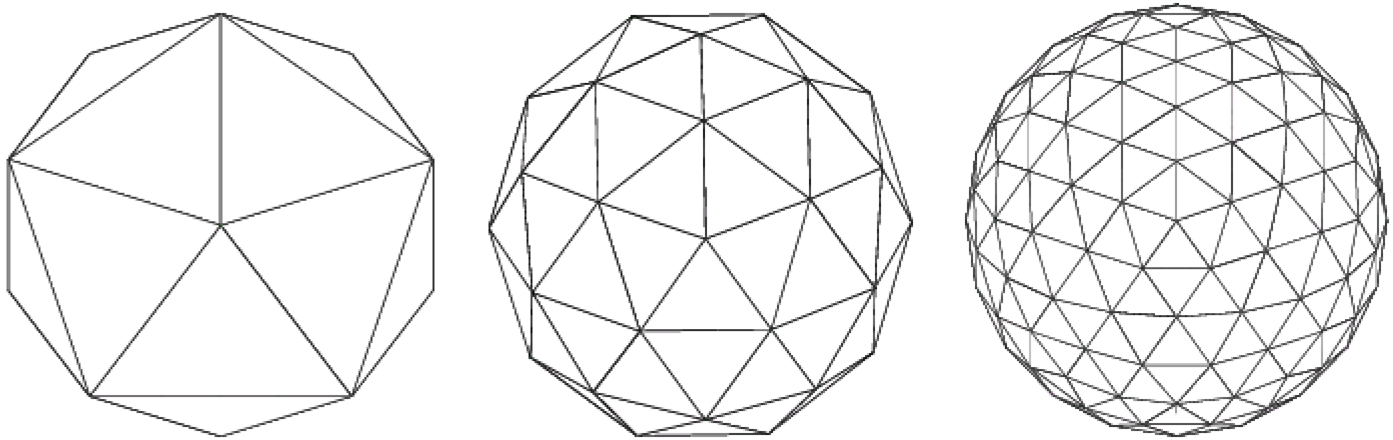
$$R_{ij}^{(e)}(\tau) = \sum_{k=0}^{L-1} \frac{w^2(k) X_i(k) X_j(k)^*}{|X_i(k)|\,|X_j(k)|} e^{\jmath 2\pi k\tau/L}$$

Search

Set of possible directions of arrival represented as sphere

Defining a homogeneous grid

- Recursive subdivision of icosahedron
- Resulting grid with 2562 points



UNIVERSITÉ DE SHERBROOKE

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

LABORIUS

Search

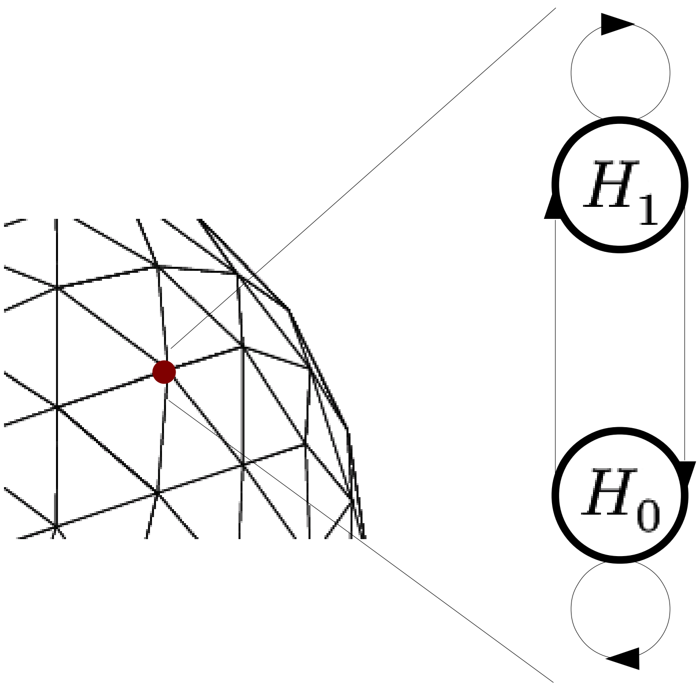
Find directions with highest energy

for $k = 1$ to desired number of sources **do**
 for all grid index d **do**
  $E_d \leftarrow 0$
  for all microphone pair ij **do**
   $\tau \leftarrow lookup(d, ij)$
   $E_d \leftarrow E_d + R_{ij}^{(e)}(\tau)$
 $D_k \leftarrow \text{argmax}_d \left(E_d\right)$
 for all microphone pair ij **do**
  $\tau \leftarrow lookup(D_k, ij)$
  $R_{ij}^{(e)}(\tau) \leftarrow 0$

Bayesian Post-filter

- Data from beamformer is noisy
- Express localization in terms of source probability of presence
- Probability computed for each grid point
- Use Bayes' rule to compute probability using past and present observations

Bayesian Post-filter



$P\left(H_1^n \middle| o_n\right)$ beamformer probability

$P\left(H_1^n \middle| \mathbf{O}_{n-1}\right)$ *a priori* probability

$P\left(H_1^n \middle| \mathbf{O}_n\right)$ combined probability

Estimator Combination

- All previous steps computed twice
  - Short frames (~40 ms)
  - Medium frames (~200 ms)
- Need to combine both estimators
  - Estimators are not independent
- Weighted geometric average of the dependent case and the independent case:

$$P\left(H_1 \middle| \mathbf{O}^s, \mathbf{O}^m\right) \approx \left[P_d\left(H_1 \middle| \mathbf{O}^s, \mathbf{O}^m\right)\right]^{\beta} \cdot \left[P_i\left(H_1 \middle| \mathbf{O}^s, \mathbf{O}^m\right)\right]^{1-\beta}$$

UNIVERSITÉ DE SHERBROOKE

INSTITUT DES MATÉRIAUX ET SYSTÈMES INTELLIGENTS
INTELLIGENT MATERIALS AND SYSTEMS INSTITUTE

LABORIUS

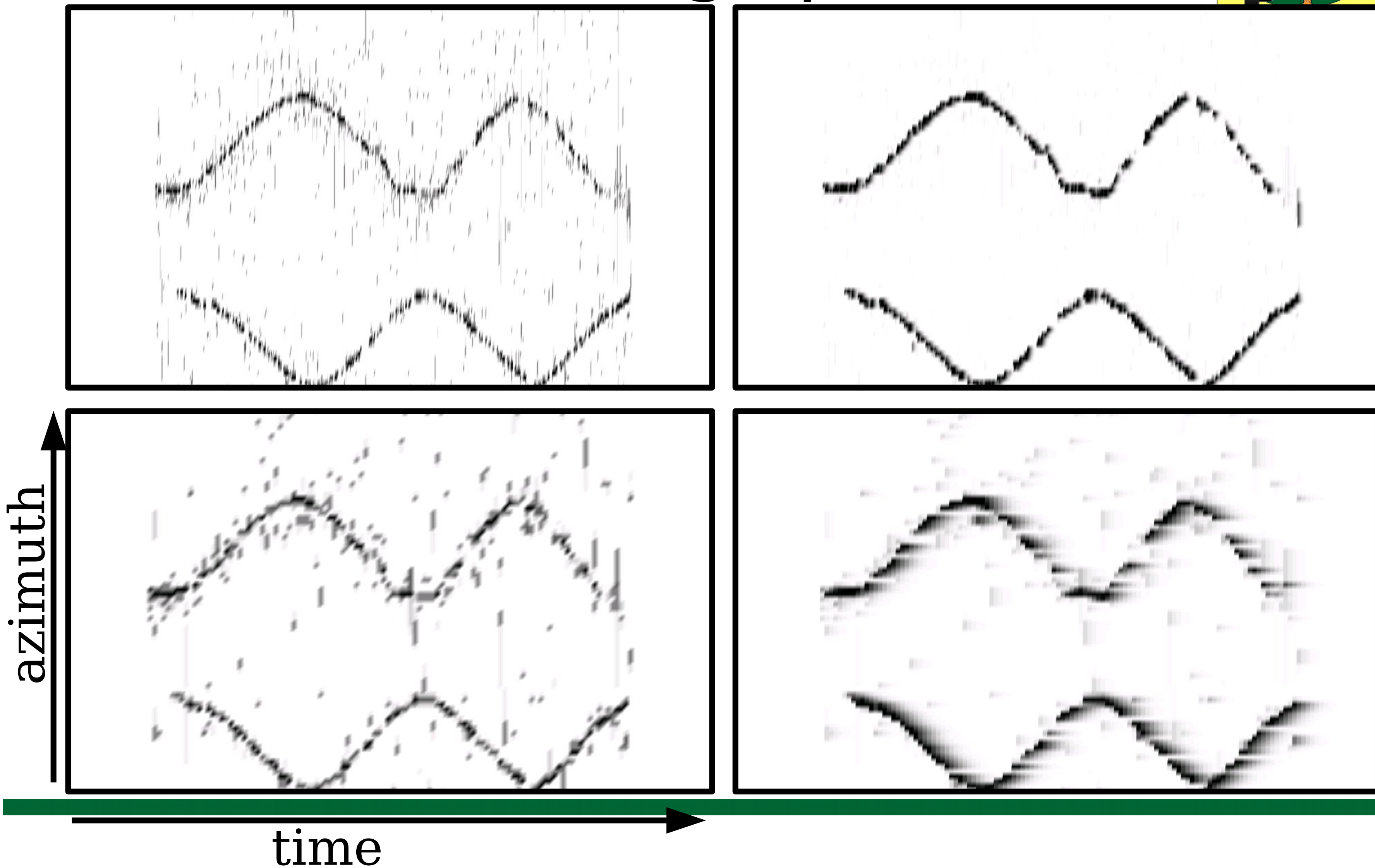
Results

Detection accuracy over distance

- Different sounds
- Rate of detection(#detections / #occurences)

Sound source	3 m	5 m	7 m
Hands clapping	92%	94%	84%
Speech ("test")	100%	90%	42%
Noise burst (250 ms)	100%	100%	100%

Results (2 moving speakers)

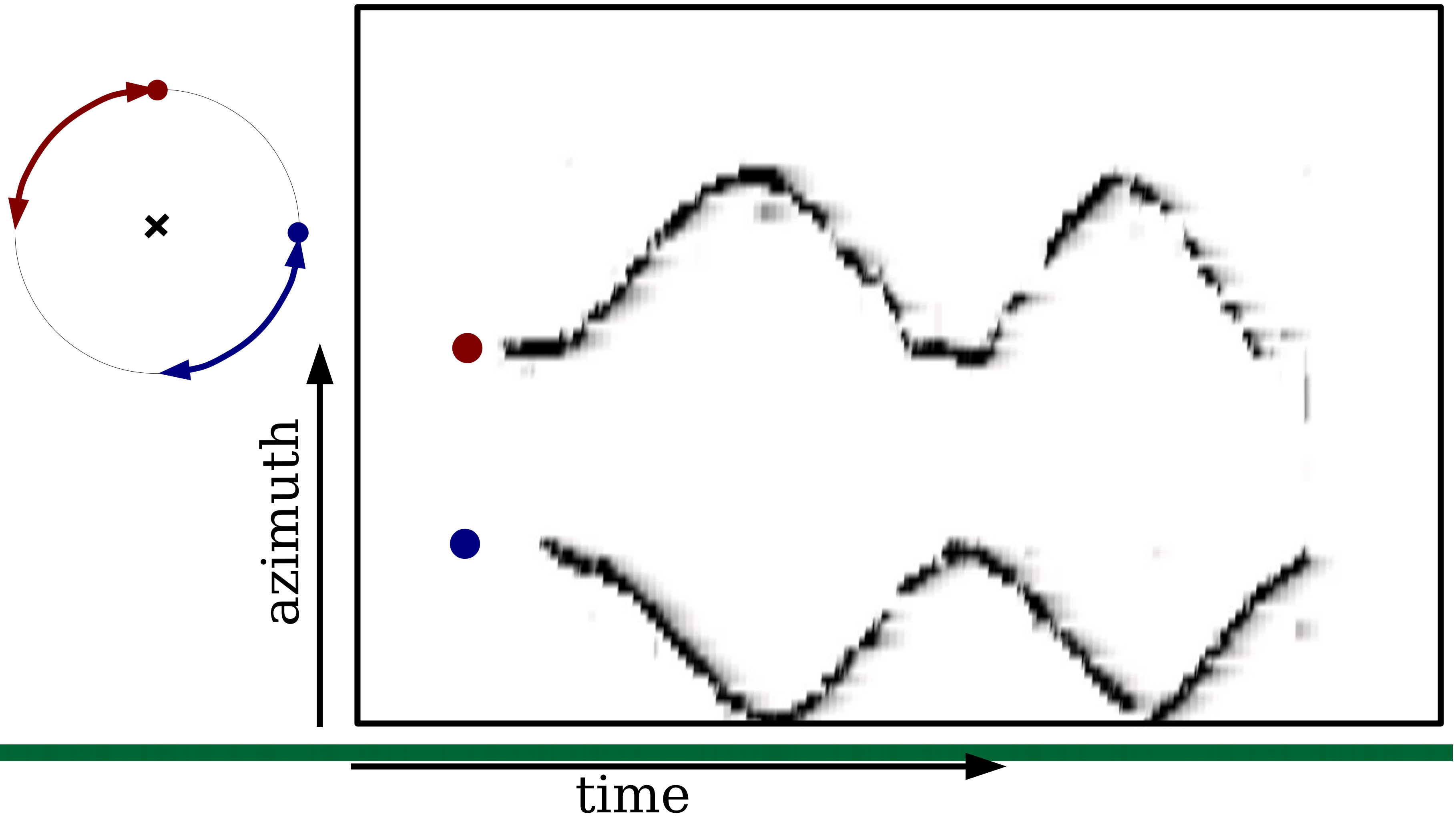


UNIVERSITÉ DE SHERBROOKE

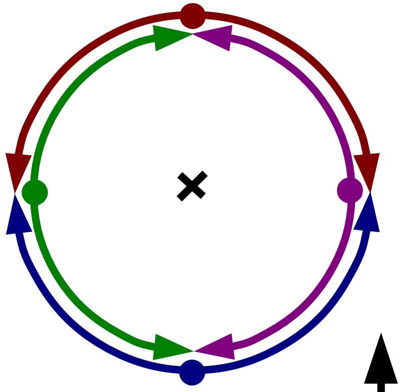
IMSI — INSTITUT DES MATÉRIAUX ET SYSTÈMES INTELLIGENTS — INTELLIGENT MATERIALS AND SYSTEMS INSTITUTE

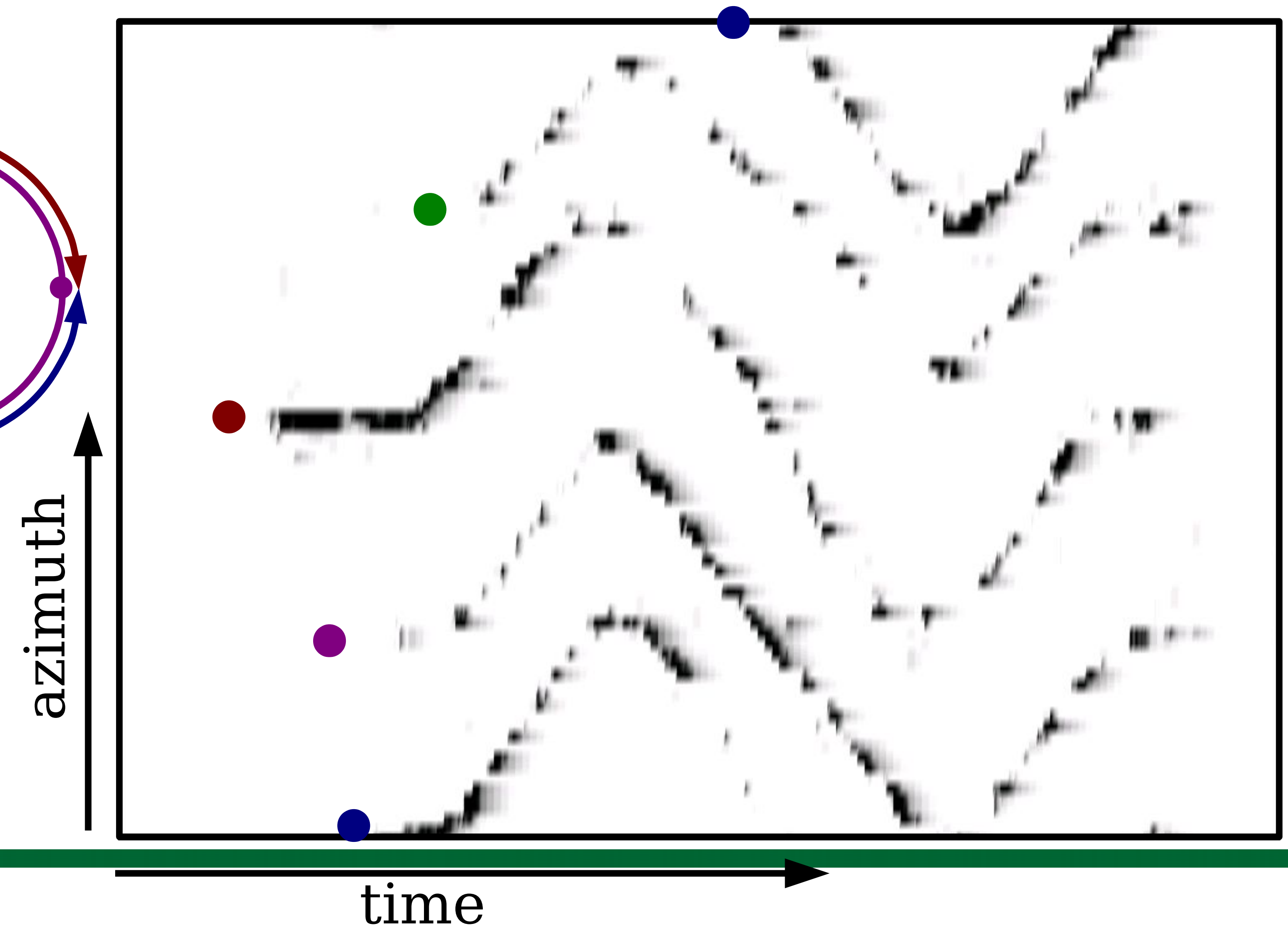
LABORIUS

Results (2 moving speakers)

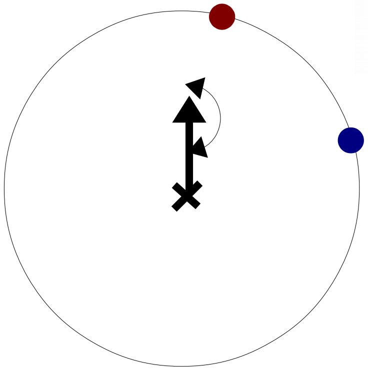


Results (4 moving speakers)

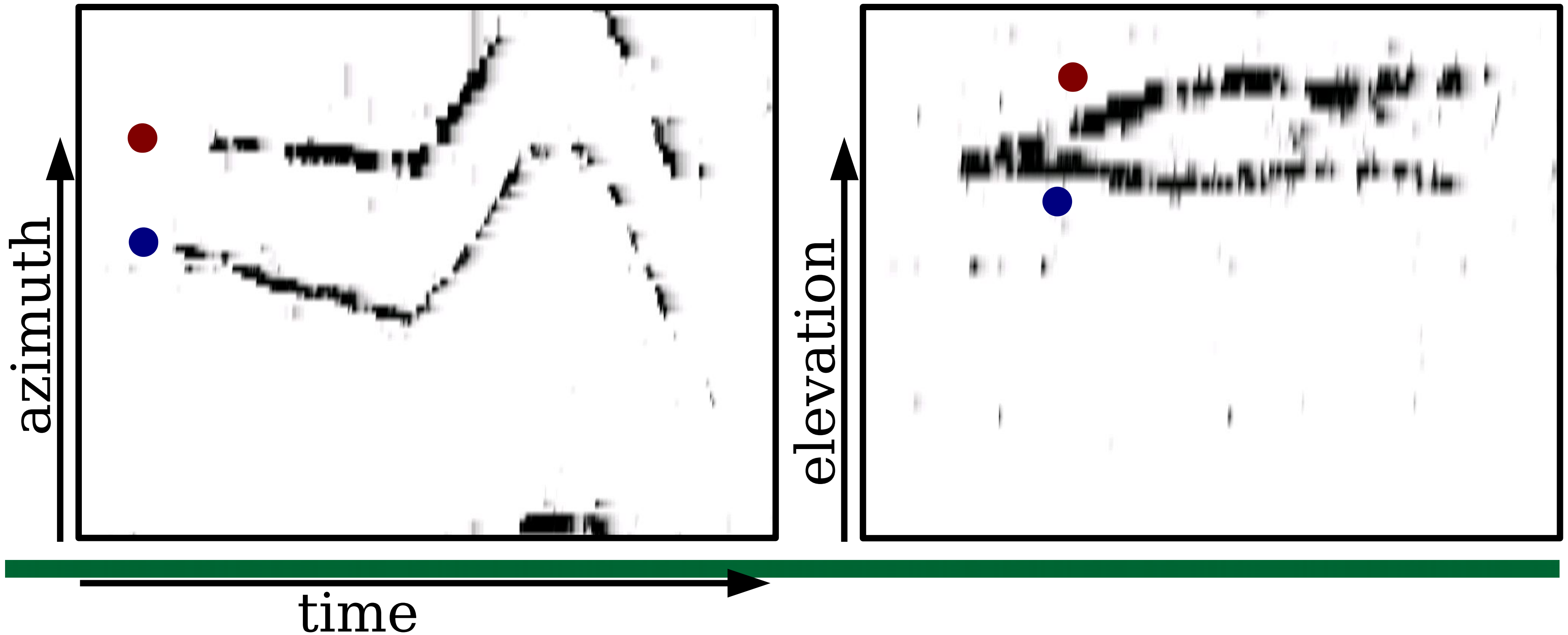




Results (moving robot)



Localization in 3D



Conclusion

- Robust localization of sound sources
  - Moving sources or robot
  - Up to 4 simultaneous sources reliably
  - Reliable detection up to 5 meters
- Two-step method
  - Steered beamformer
  - Bayesian post-filter
- Related work
  - Tracking sources over time
  - Separating sound one mic separated

Questions?

Search (cont.)

1) Steered beamformer direction search

Finding the direction with highest energy

for all grid index d **do**

$\quad E_d \leftarrow 0$

$\quad$ **for all** microphone pair ij **do**

$\quad\quad \tau \leftarrow lookup(d, ij)$

$\quad\quad E_d \leftarrow E_d + R_{ij}^{(e)}(\tau)$

$\quad$ **end for**

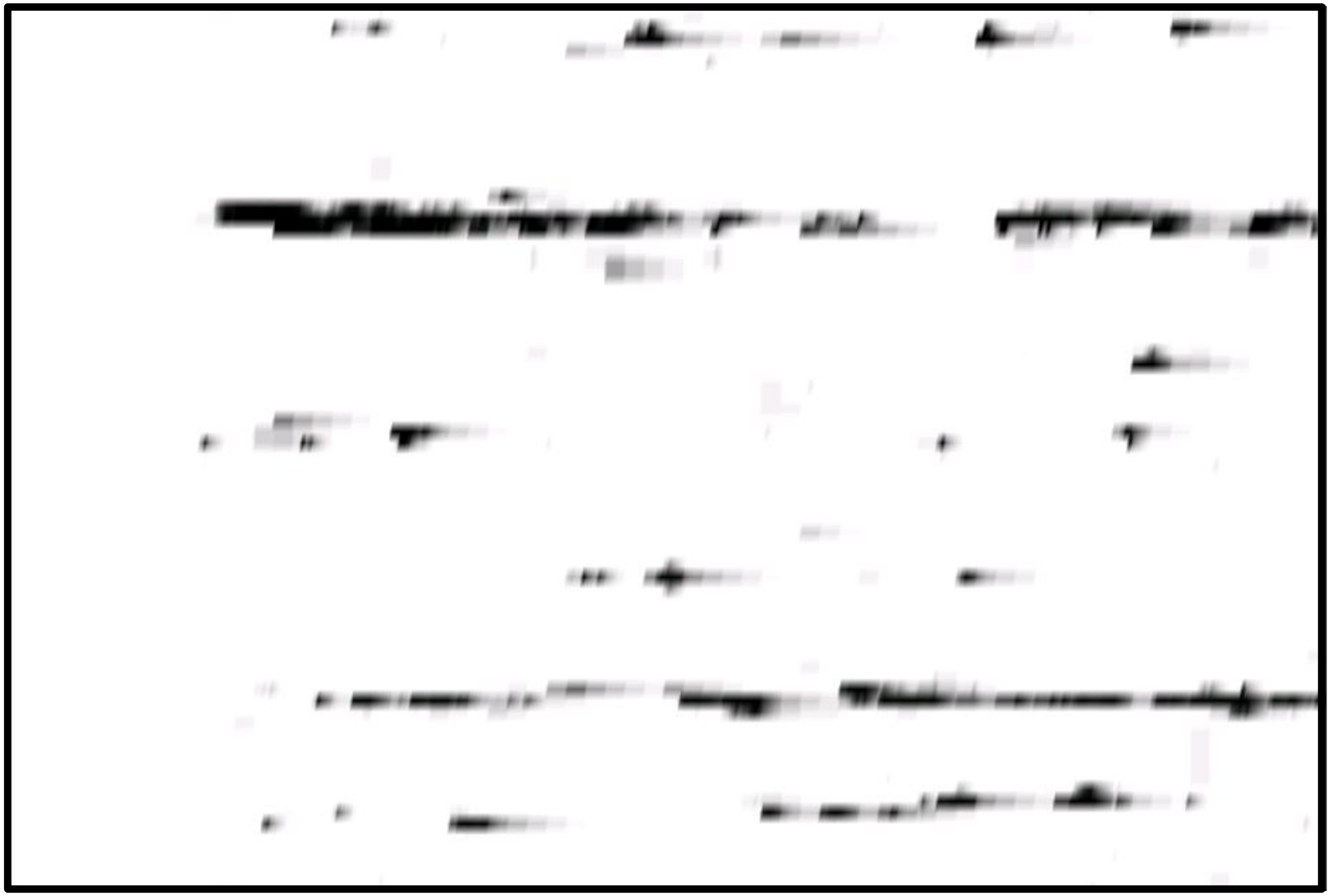
end for

$direction\ of\ source \leftarrow \mathrm{argmax}_d\ (E_d)$

Bayesian Post-filter (cont.)

- Beamformer assigns instantaneous probability $P(H_1^n | o_n)$ for each grid point
- *A priori* probability $P(H_1^n | \mathbf{O}_{n-1})$ assuming a Markov process
- Current probability $P(H_1^n | \mathbf{O}_n)$

Results (7 sources)



UNIVERSITÉ DE SHERBROOKE

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

LABORIUS

Search (cont.)

2) Complete search

Finding all sources

for $k = 1$ to desired number of sources **do**

  $D_k \leftarrow$ Steered beamformer direction search

  for all microphone pair ij **do**

    $\tau \leftarrow lookup(D_k, ij)$

    $R_{ij}^{(e)}(\tau) = 0$

  end for

end for