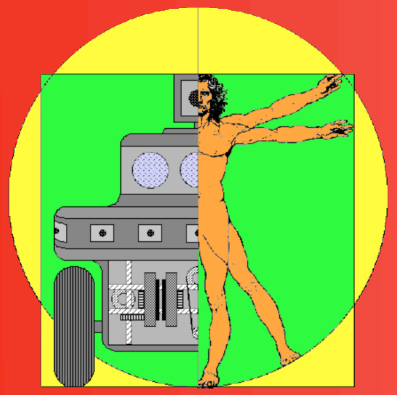


Laboratory on Mobile Robotics and Intelligent Systems
LABORIUS

FCAR

CRSNG
NSERC

Fondation canadienne pour l'innovation

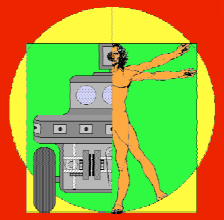
Robust Sound Source Localization Using a Microphone Array on a Mobile Robot

Jean-Marc Valin, François Michaud, Jean Rouat, Dominic Létourneau
Department of Electrical Engineering and Computer Engineering
Université de Sherbrooke, Québec CANADA
jean-marc.valin@usherbrooke.ca
http://www.gel.usherb.ca/laborius

CANADA RESEARCH CHAIRS
CHAIRES DE RECHERCHE DU CANADA

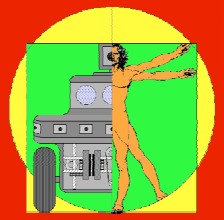
UNIVERSITÉ DE SHERBROOKE

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



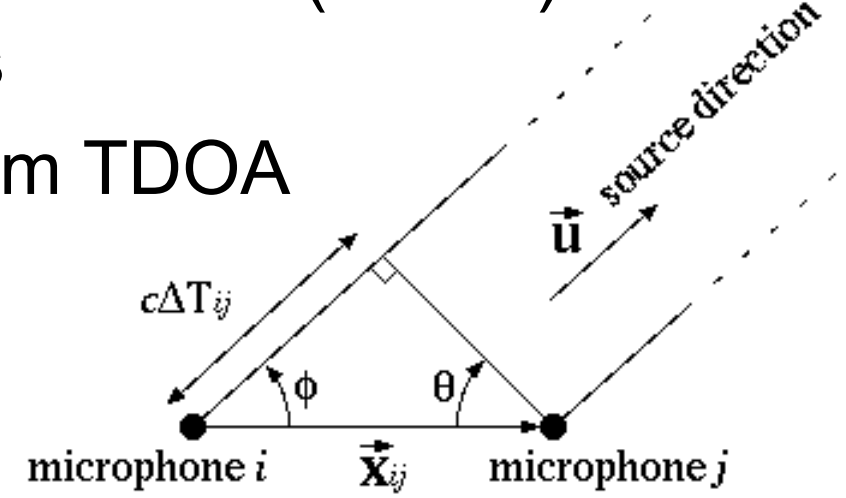
Sound Source Localization

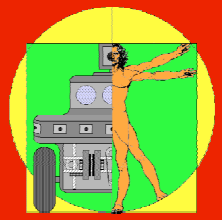
- Determining where the sources of sounds are
  - Humans
    - Two ears
    - Head transfer function (acoustic shadow, reflections of sound by the ridges of the ear)
  - Robots
    - Two microphones (phase difference only)
      - Locate sounds over a planar area, without distinguishing the front from the back or high precision if the sound source is in the same axis
    - Eight microphones
      - Compensate for high level of complexity of the hearing sense
      - Filter out noise by discriminating multiple sound sources



Approach Overview

- Sounds arrive at microphones with different delays (depending on distance)
  - Hypothesis: Punctual sound source, far field
- Extract **Time Delay of Arrival** (TDOA) between different microphones
- Compute direction from TDOA





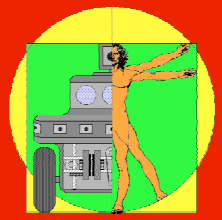
TDOA by Cross-Correlation

- Delay found as peak in cross-correlation

$$R_{ij}(\tau) = \sum_{n=0}^{N-1} x_i[n] x_j[n-\tau]$$

- Performed in frequency domain (faster)

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



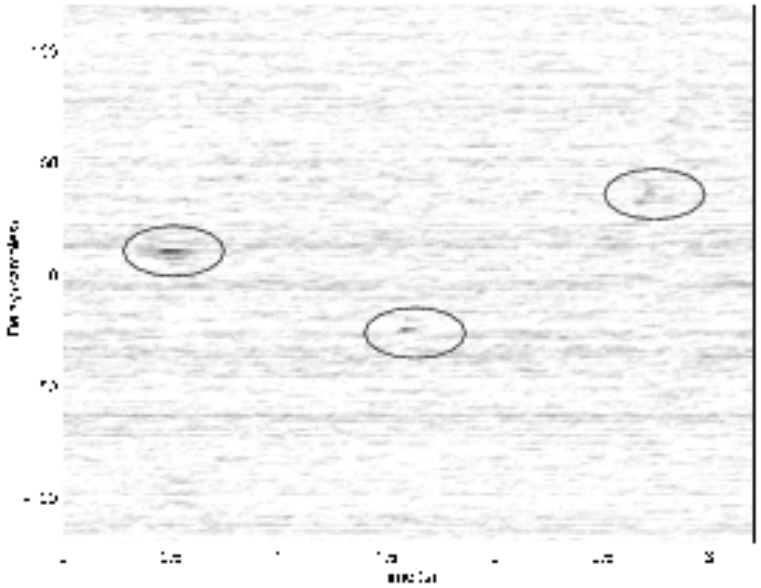
Enhanced Cross-Correlation

- Whitened cross-correlation
  - Cross-correlation on low-pass signal generates wide peaks in frequency: must narrow the wide maxima caused by the correlations within the received signals
  - Normalize spectrum (only phase information is preserved)
- Spectral weighting
  - Whitening gives less weight for frequencies dominated by noise: must give more weight to frequencies with high power

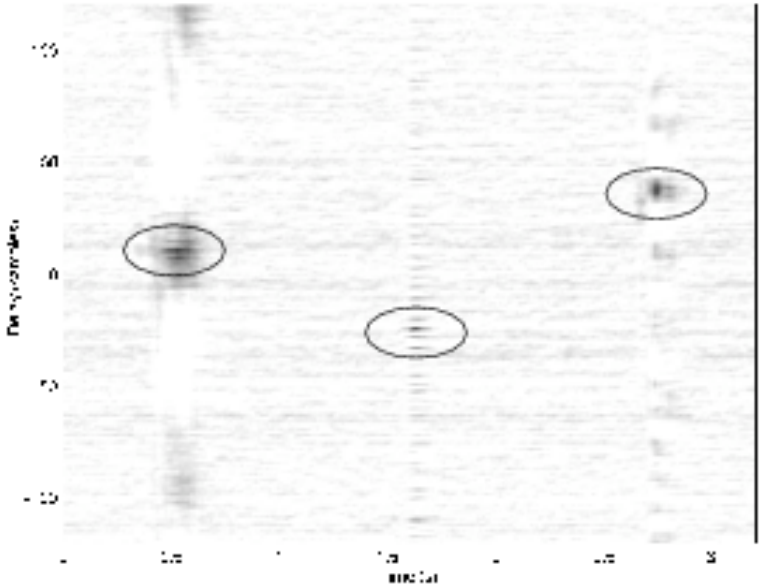
Spectral Weighting

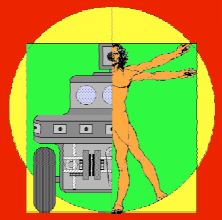
- Effect of weighting on cross-correlation
  - Example

No weighting (whiten only)



With weighting



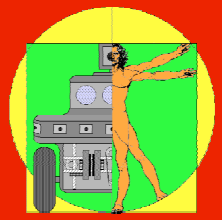


Peak Extraction

- For each microphone pair:

$$\Delta T_{ij} = \underset{\tau}{\operatorname{argmax}} \; R_{ij}(\tau)$$

- Extract *M* peaks (*M*=8) for each pair
  - To make sure the source is detected



Peak Coherence Search

- *N*(*N*-1)/2 microphone pairs, *N*-1 deg. of freedom
- Dependent TDOAs satisfy:

$$\Delta T_{ij} = \Delta T_{1j} - \Delta T_{1i}$$

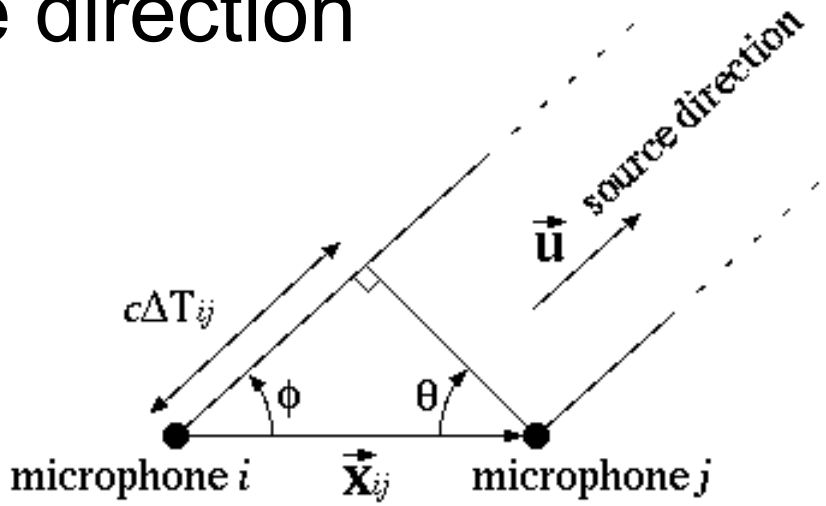
$$\Delta T_{23} = \Delta T_{13} - \Delta T_{12}$$

$$T_3 - T_2 = (T_3 - T_1) - (T_2 - T_1)$$

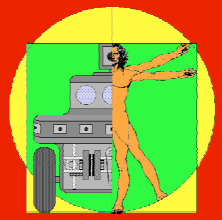
- Source detected if *most* constraints are met
- Depth-first search with pruning
- If more than one solution, only keep best

Direction Estimations

- Once peaks are located, use them to compute direction



$$\vec{\mathbf{u}} \cdot \vec{\mathbf{x}}_{ij} = c\Delta T_{ij}$$

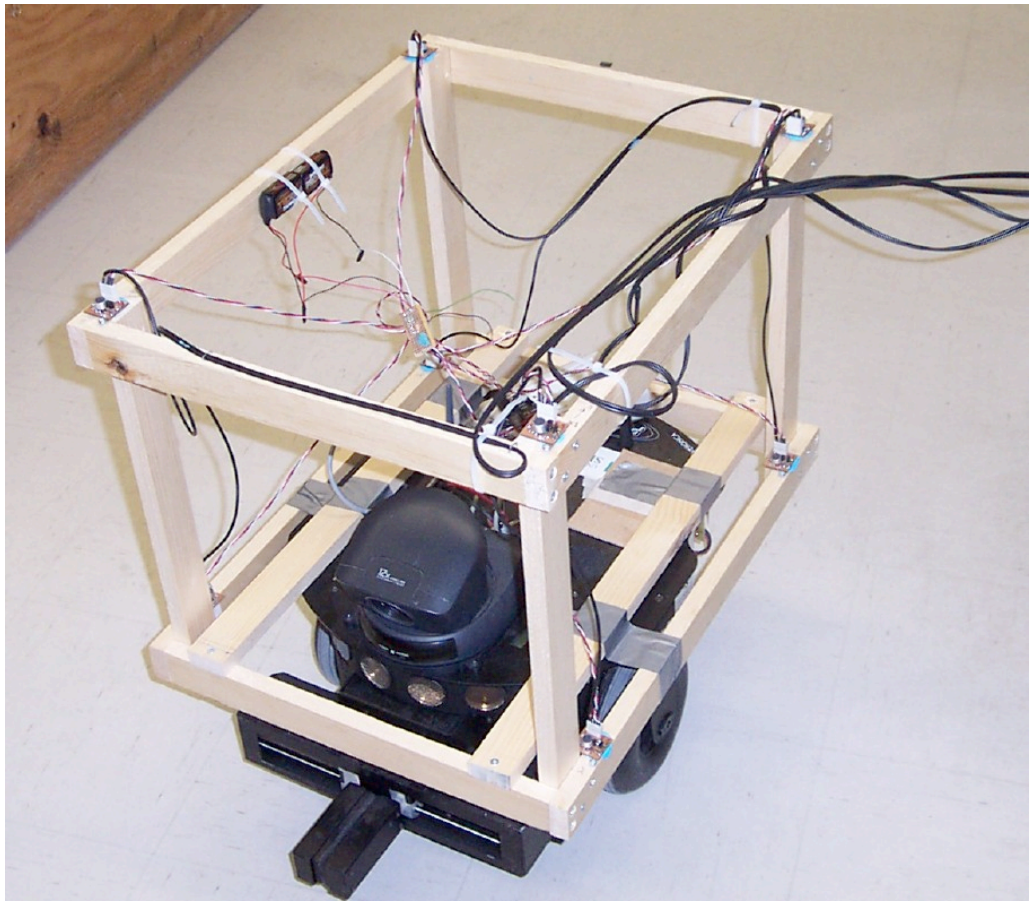


Direction Estimation

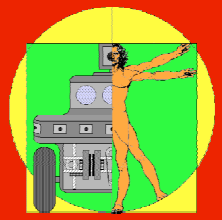
- $$\begin{bmatrix} (x_2 - x_1) & (y_2 - y_1) & (z_2 - z_1) \\ (x_3 - x_1) & (y_2 - y_1) & (z_3 - z_1) \\ \vdots & \vdots & \vdots \\ (x_N - x_1) & (y_N - y_1) & (z_N - z_1) \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} c\Delta T_{12} \\ c\Delta T_{13} \\ \vdots \\ c\Delta T_{1N} \end{bmatrix}$$

- Over-constrained (least square solution)
- Pseudo-inverse of matrix is constant and pre-computed

Experimental Setup



Experiments



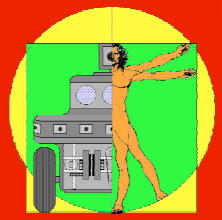
Results

Distance, Elevation	Mean Ang. Error
3 m, -7°	1.7°
3 m, 8°	3°
1.5 m, -13°	3.1°
0.9 m, 24°	3.3°

- Error caused by reverberation, near-field effects, measurement precision, source size
- Accuracy shows no dependencies on angle (unlike binaural localization)

Results

- Pictures taken of detected sources



Conclusion

- Sound source localization based on TDOA
  - Frequency-domain cross-correlation
  - Peak finding, coherence search
- Accuracy of ±3 degrees
- Works in noisy environments