EPSRC
Engineering and Physical Sciences Research Council

Imperial College London

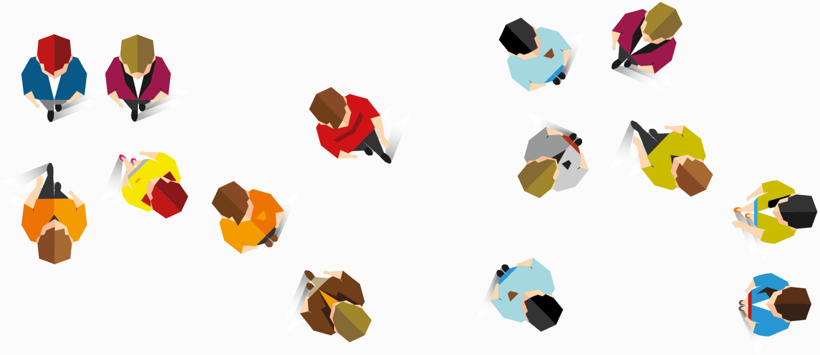
Bayesian Learning for Robot Audition

Keynote, HSCMA 2017, San Francisco

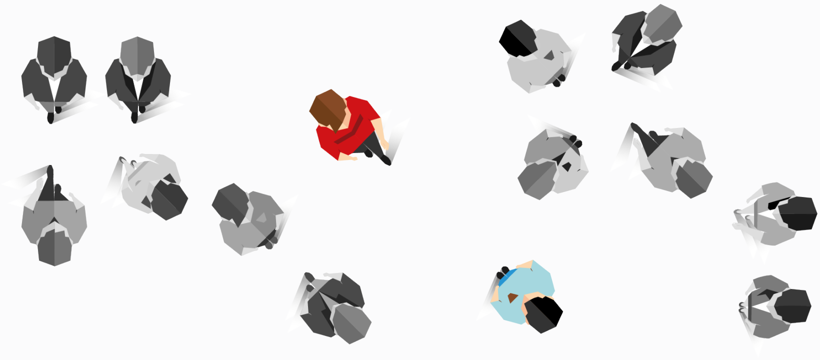
Christine Evers
Imperial College London
Dept. Electrical and Electronic Engineering

Introduction

Listening in Realistic Environments



Listening in Realistic Environments

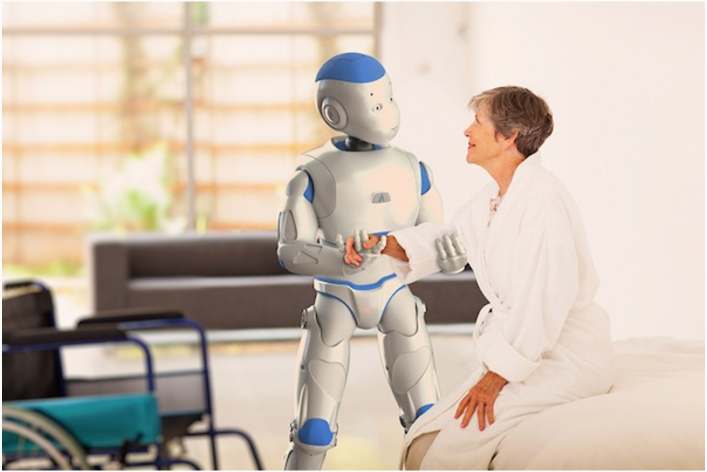


Acoustic Scene Mapping

- Estimate the positions of surrounding sources
- Focus and interact with users
- Situational awareness to react to stimuli in the environment

Impact

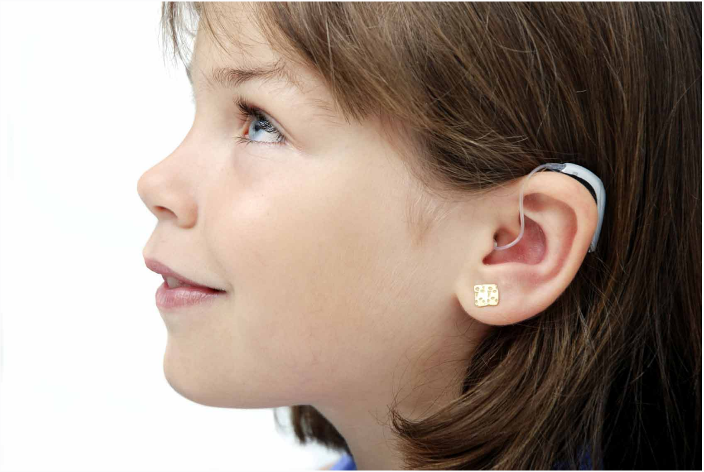
Imperial College London



Socially assistive robots. Photo credit: SoftBank Robotics

Search-and-rescue robots. Photo credit: U.S. Army Space and Missile Defense Command

Smart homes. Photo credit: theappsolutions.com



Hearing aids. Photo credit: Oregon Lions Sight & Hearing Foundation

Impact

Virtual reality. Photo credit: Samsung

Sound Source Localization

- **Aim:** Estimate source direction(s) at one time instance
- Generalized cross correlation (GCC), Multiple Signal Classification (MUSIC), Intensity vectors, ...

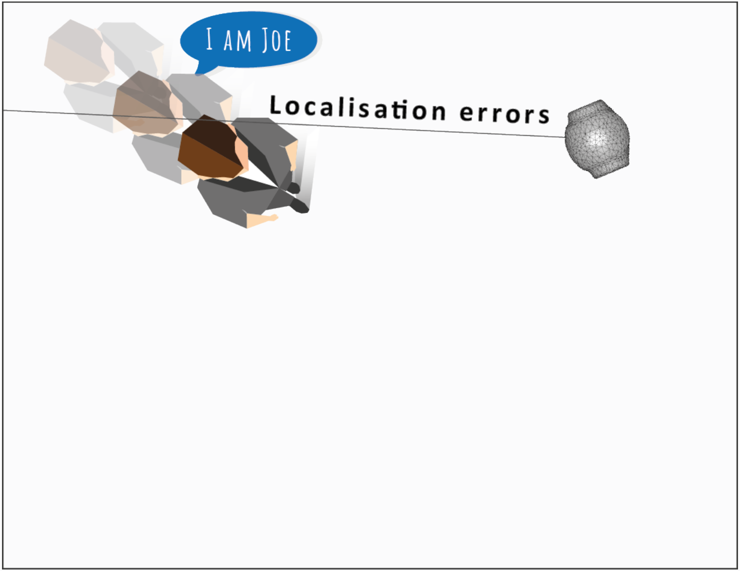
[1] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust Localization in Reverberant Rooms," in *Microphone Arrays*, Springer, 2001.

[2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer 2008.

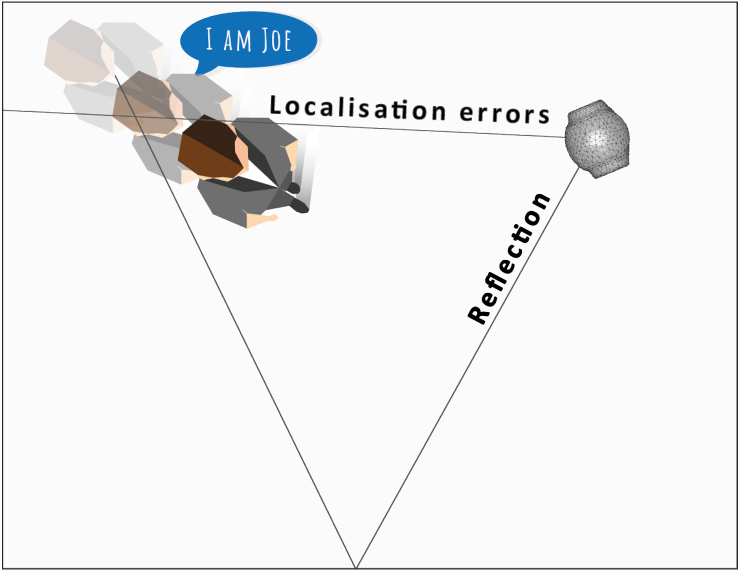
Sound Source Localization

- **Aim:** Estimate source direction(s) at one time instance
- Generalized cross correlation (GCC), Multiple Signal Classification (MUSIC), Intensity vectors, ...

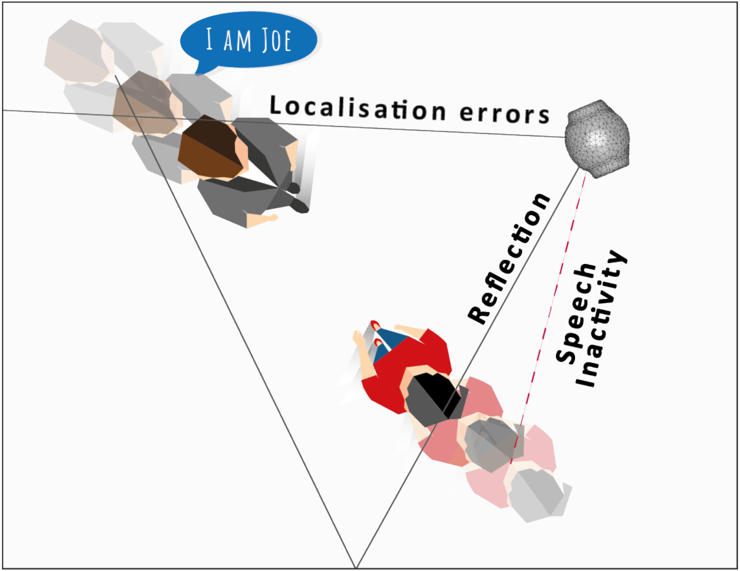
Localization Challenges



Localization Challenges



Localization Challenges



Sound Source Localization

- **Aim:** Estimate source direction(s) at one time instance

Localization & Tracking

Imperial College London

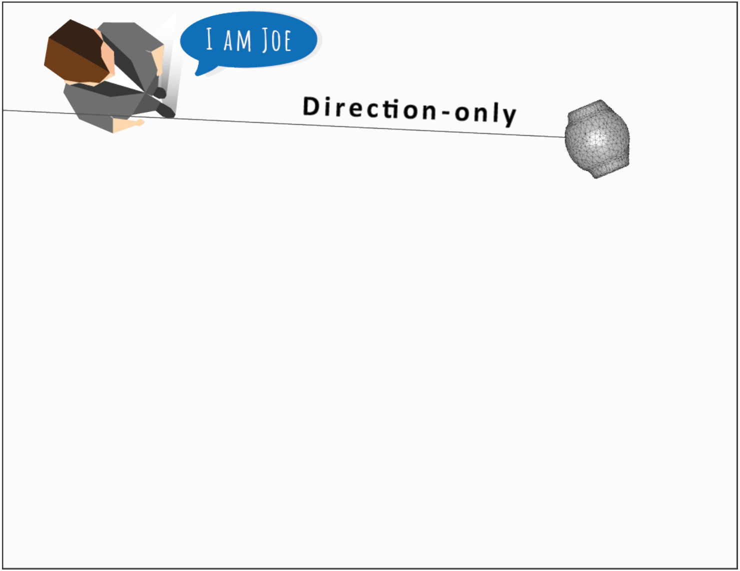
Sound Source Localization

- **Aim:** Estimate source direction(s) at one time instance

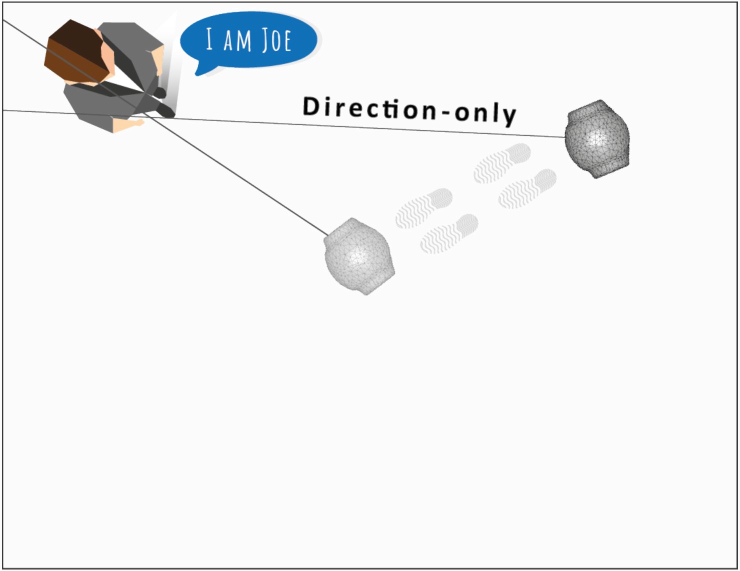
Sound Source Tracking

- Utilize localization estimates as "measurements"
- Exploit temporal models of the source dynamics
- Estimate smoothed source trajectories over time

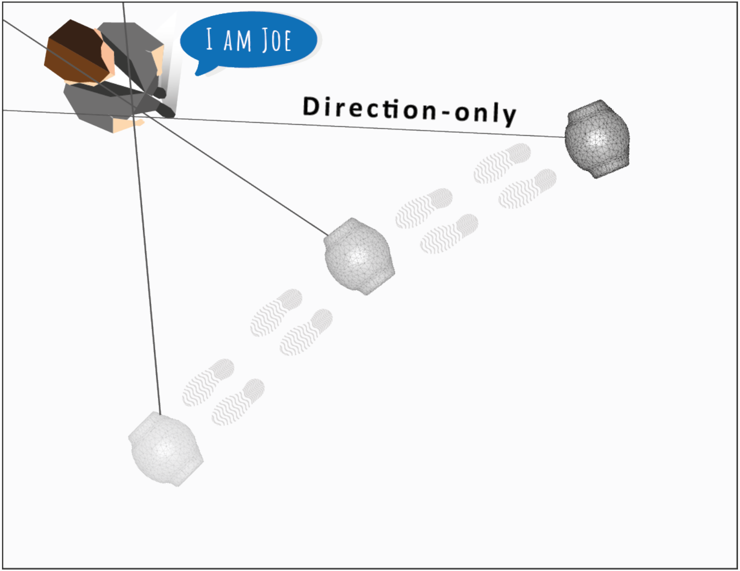
Source Position Tracking



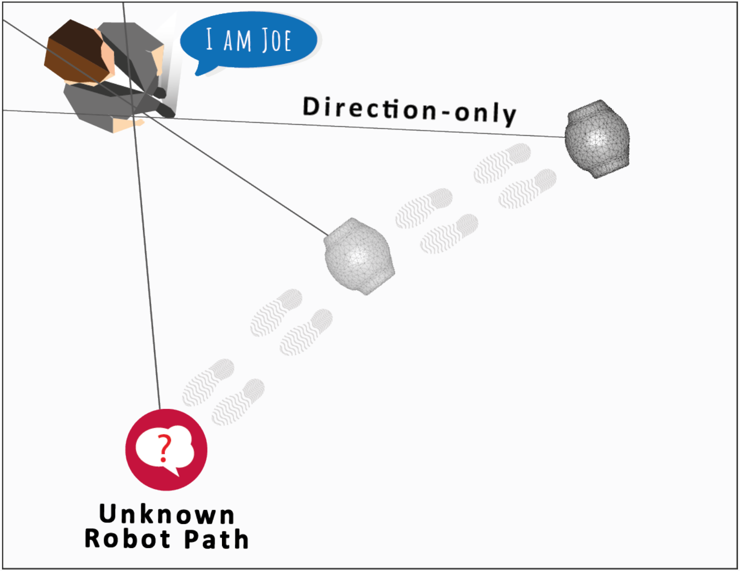
Source Position Tracking



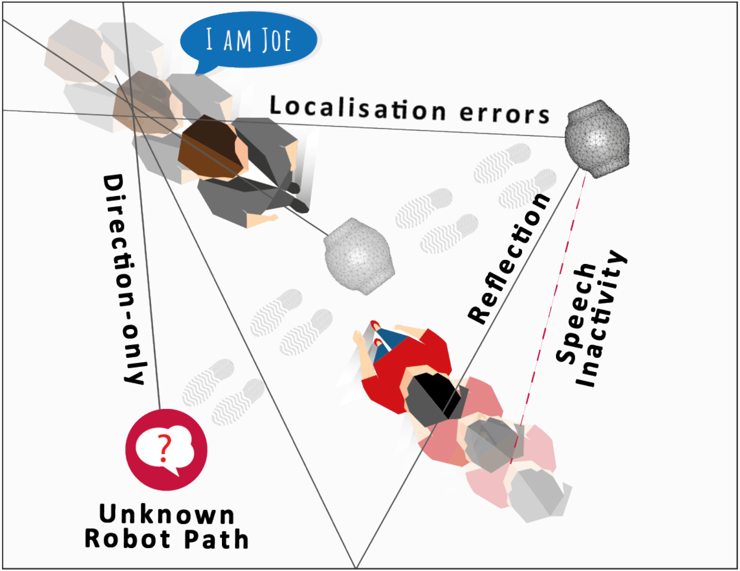
Source Position Tracking



Source Position Tracking



Acoustic Scene Mapping Challenges



Bayesian Inference

Real-time Scene Mapping

- Sequential estimation required
- Tracking domain:
  - Each window of speech data considered one time step, t
  - Localization estimates at t: Measurements, $\mathbf{z}_t$
  - **Aim:** Estimate source position, $\mathbf{s}_t$ from trajectory $\mathbf{z}_{1:t}$

- **Probabilistic perspective:** Propagation of posterior density function (pdf)

Real-time Scene Mapping

- Sequential estimation required
- Tracking domain:
  - Each window of speech data considered one time step, t
  - Localization estimates at t: Measurements, $\mathbf{z}_t$
  - **Aim:** Estimate source position, $\mathbf{s}_t$ from trajectory $\mathbf{z}_{1:t}$

- **Probabilistic perspective:** Propagation of posterior density function (pdf)

$$p\left(\mathbf{s}_{t-1} \mid \mathbf{z}_{1:t-1}\right) \xrightarrow{f(\cdot)} p\left(\mathbf{s}_t \mid \mathbf{z}_{1:t}\right) \xleftarrow{} \mathbf{z}_t$$

- Sequential estimation required
- Tracking domain:
  - Each window of speech data considered one time step, t
  - Localization estimates at t: Measurements, $\mathbf{z}_t$
  - **Aim:** Estimate source position, $\mathbf{s}_t$ from trajectory $\mathbf{z}_{1:t}$

- **Probabilistic perspective:** Propagation of posterior density function (pdf)

$$p\left(\mathbf{s}_{t-1} \mid \mathbf{z}_{1:t-1}\right) \xrightarrow{f(\cdot)} p\left(\mathbf{s}_t \mid \mathbf{z}_{1:t}\right) \xrightarrow{f(\cdot)} p\left(\mathbf{s}_{t+1} \mid \mathbf{z}_{1:t+1}\right)$$

$$\mathbf{z}_t \downarrow \qquad \mathbf{z}_{t+1} \downarrow$$

Step 1 - Probability transformation

$$\text{Dynamical model:} \quad \mathbf{s}_t = f(\mathbf{s}_{t-1}, \mathbf{v}_t) \quad \overset{\text{Prior}}{\Rightarrow} \quad p\left(\mathbf{s}_t \mid \mathbf{s}_{t-1}\right)$$

$$\text{Measurement model:} \quad \mathbf{z}_t = h(\mathbf{s}_t, \mathbf{w}_t) \quad \overset{\text{Likelihood}}{\Rightarrow} \quad p\left(\mathbf{z}_t \mid \mathbf{s}_t\right)$$

Bayesian Recipe

Step 1 - Probability transformation

Dynamical model: $\mathbf{s}_t = f(\mathbf{s}_{t-1}, \mathbf{v}_t) \quad \overset{\text{Prior}}{\Rightarrow} \quad p\left(\mathbf{s}_t \mid \mathbf{s}_{t-1}\right)$

Measurement model: $\mathbf{z}_t = h(\mathbf{s}_t, \mathbf{w}_t) \quad \overset{\text{Likelihood}}{\Rightarrow} \quad p\left(\mathbf{z}_t \mid \mathbf{s}_t\right)$

Step 2 - Prediction using the prior

$$p\left(\mathbf{s}_t \mid \mathbf{z}_{1:t-1}\right) = \int p\left(\mathbf{s}_{t-1} \mid \mathbf{z}_{1:t-1}\right) p\left(\mathbf{s}_t \mid \mathbf{s}_{t-1}\right) d\mathbf{s}_{t-1}$$

Bayesian Recipe

Step 1 - Probability transformation

Dynamical model: $\mathbf{s}_t = f(\mathbf{s}_{t-1}, \mathbf{v}_t) \quad \overset{\text{Prior}}{\Rightarrow} \quad p\left(\mathbf{s}_t \mid \mathbf{s}_{t-1}\right)$

Measurement model: $\mathbf{z}_t = h(\mathbf{s}_t, \mathbf{w}_t) \quad \overset{\text{Likelihood}}{\Rightarrow} \quad p\left(\mathbf{z}_t \mid \mathbf{s}_t\right)$

Step 2 - Prediction using the prior

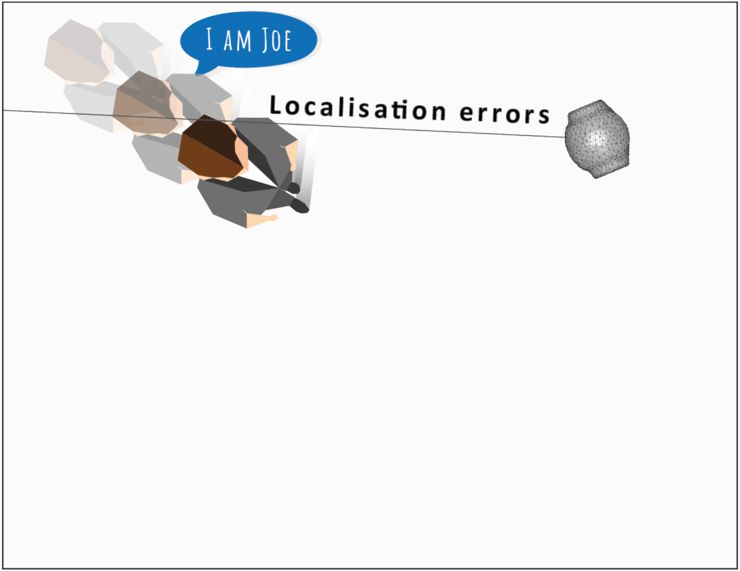
$$p\left(\mathbf{s}_t \mid \mathbf{z}_{1:t-1}\right) = \int p\left(\mathbf{s}_{t-1} \mid \mathbf{z}_{1:t-1}\right) p\left(\mathbf{s}_t \mid \mathbf{s}_{t-1}\right) d\mathbf{s}_{t-1}$$

Step 3 - Bayes's theorem

$$p\left(\mathbf{s}_t \mid \mathbf{z}_{1:t}\right) = \frac{p\left(\mathbf{z}_t \mid \mathbf{s}_t, \mathbf{z}_{1:t-1}\right) p\left(\mathbf{s}_t \mid \mathbf{z}_{1:t-1}\right)}{p\left(\mathbf{z}_t \mid \mathbf{z}_{1:t-1}\right)}$$

Source DoA Tracking

Tracking and Localization Errors



Classical DoA Tracking - Models

Dynamical Model of DoA $\boldsymbol{\omega}_t \triangleq \left[\phi_t, \theta_t, \dot{\phi}_t, \dot{\theta}_t\right]^T$

$$\textbf{Source state:} \qquad \boldsymbol{\omega}_t = \vartheta\left(\mathbf{F}_t \boldsymbol{\omega}_{t-1} + \mathbf{v}_t\right) \tag{1}$$

$$\textbf{Constant-velocity dynamics:} \qquad \mathbf{F}_t \triangleq \begin{bmatrix} \mathbf{I}_2 & \Delta_t \mathbf{I}_2 \\ \mathbf{0}_{2\times 2} & \mathbf{I}_2 \end{bmatrix} \tag{2}$$

where $\vartheta\left(\boldsymbol{\omega}_t\right) = \mod\left(\boldsymbol{\omega}_t, [2\pi, \pi, 2\pi, \pi]^T\right)$ and $\mathbf{v}_t$: process noise with covariance $\mathbf{Q}$.

Classical DoA Tracking - Models

Dynamical Model of DoA $\boldsymbol{\omega}_t \triangleq \left[\phi_t, \theta_t, \dot{\phi}_t, \dot{\theta}_t\right]^T$

Source state:
$$\boldsymbol{\omega}_t = \vartheta\left(\mathbf{F}_t \boldsymbol{\omega}_{t-1} + \mathbf{v}_t\right) \tag{1}$$

Constant-velocity dynamics:
$$\mathbf{F}_t \triangleq \begin{bmatrix} \mathbf{I}_2 & \Delta_t \mathbf{I}_2 \\ \mathbf{0}_{2\times 2} & \mathbf{I}_2 \end{bmatrix} \tag{2}$$

where $\vartheta\left(\boldsymbol{\omega}_t\right) = \mod\left(\boldsymbol{\omega}_t, [2\pi, \pi, 2\pi, \pi]^T\right)$ and $\mathbf{v}_t$: process noise with covariance $\mathbf{Q}$.

Measurement Model of localized DoA $\mathbf{z}_t \triangleq \left[\hat{\phi}_t, \hat{\theta}_t\right]^T$

Source direction:
$$\mathbf{z}_t = \vartheta\left(\mathbf{H}\, \boldsymbol{\omega}_t + \mathbf{w}_t\right) \tag{3}$$

where $\mathbf{H} \triangleq \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_{2\times 2} \\ \multicolumn{2}{c}{\mathbf{0}_{2\times 4}} \end{bmatrix}$ and $\mathbf{w}_t$: measurement noise with covariance $\mathbf{R}$.

Single Source Tracking - Kalman Filter

For source constrained to area where 2π crossing is avoided:

- Approximate by linear state space
- Prior and likelihood: Gaussian

For source constrained to area where 2π crossing is avoided:

- Approximate by linear state space
- Prior and likelihood: Gaussian

Prediction - Marginalization with Gaussian prior:

$$p\left(\boldsymbol{\omega}_t \mid \mathbf{z}_{1:t-1}\right) = \mathcal{N}\left(\boldsymbol{\omega}_t \,\middle|\, \mathbf{m}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$$

where $\mathbf{m}_{t|t-1} = \mathbf{F}_t\,\mathbf{m}_{t-1}$ and $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{F}_t\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}_t^T + \mathbf{Q}_t$

Single Source Tracking - Kalman Filter

For source constrained to area where 2π crossing is avoided:

- Approximate by linear state space
- Prior and likelihood: Gaussian

Prediction - Marginalization with Gaussian prior:

$$p\left(\boldsymbol{\omega}_t \mid \mathbf{z}_{1:t-1}\right) = \mathcal{N}\left(\boldsymbol{\omega}_t \,\middle|\, \mathbf{m}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$$

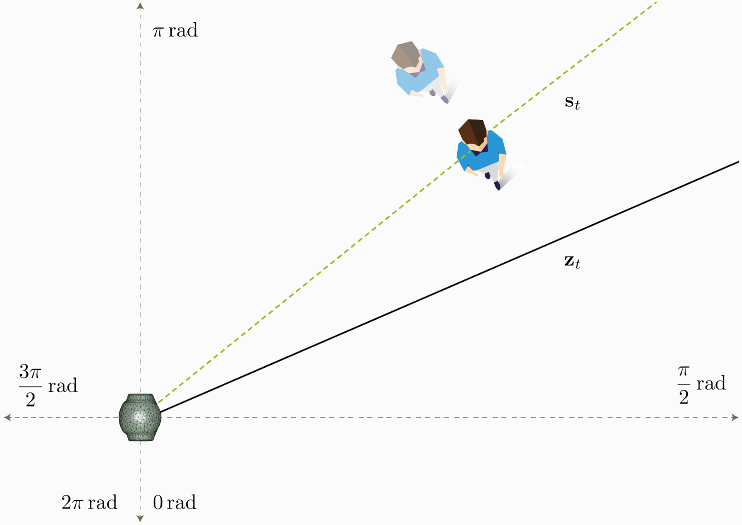
where $\mathbf{m}_{t|t-1} = \mathbf{F}_t \, \mathbf{m}_{t-1}$ and $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{F}_t \, \boldsymbol{\Sigma}_{t-1} \, \mathbf{F}_t^T + \mathbf{Q}_t$

Update - Bayes's theorem for Gaussian likelihood:

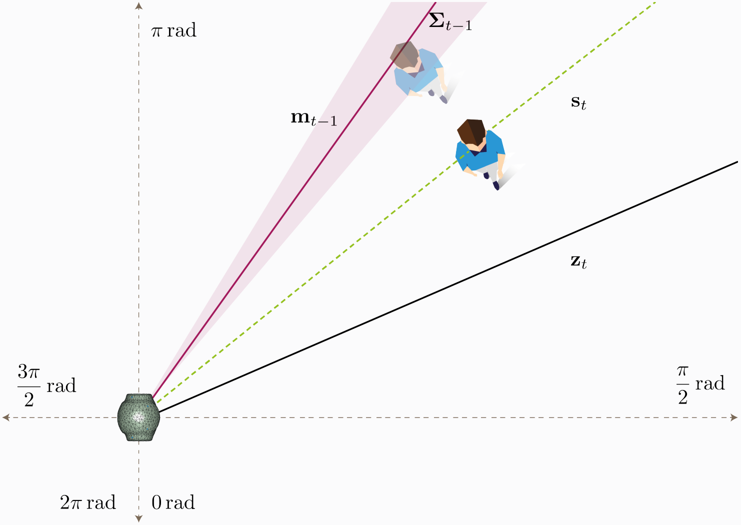
$$p\left(\boldsymbol{\omega}_t \mid \mathbf{z}_{1:t}\right) = \mathcal{N}\left(\boldsymbol{\omega}_t \,\middle|\, \mathbf{m}_t, \boldsymbol{\Sigma}_t\right)$$

where $\mathbf{m}_t = \mathbf{m}_{t|t-1} + \mathbf{K}_t \left(\mathbf{z}_t - \mathbf{H} \, \mathbf{m}_{t|t-1}\right)$ and $\boldsymbol{\Sigma}_t = \left(\mathbf{I}_4 - \mathbf{K}_t \mathbf{H}\right) \boldsymbol{\Sigma}_{t|t-1}$.

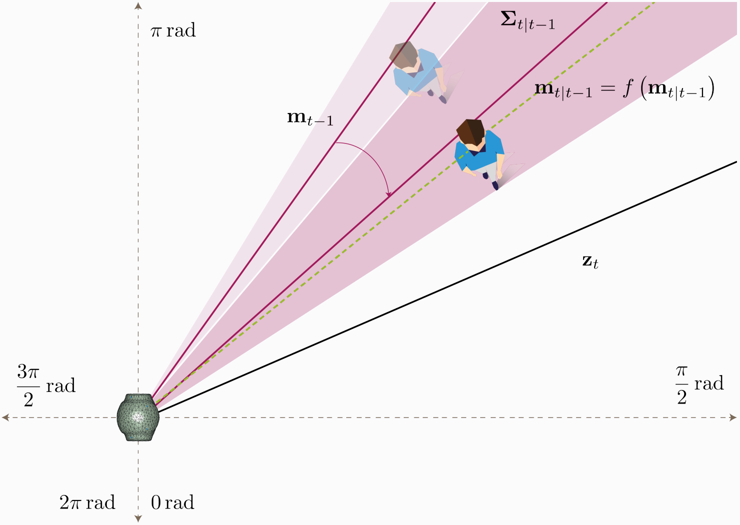
Single Source Tracking - Kalman Filter



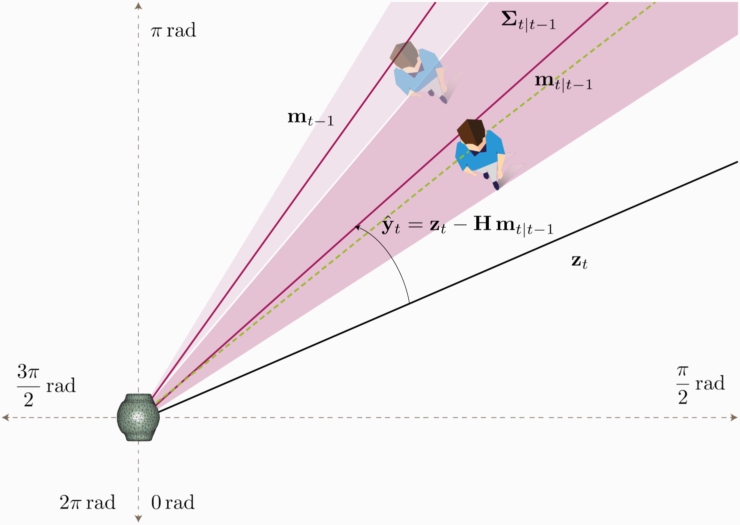
Single Source Tracking - Kalman Filter



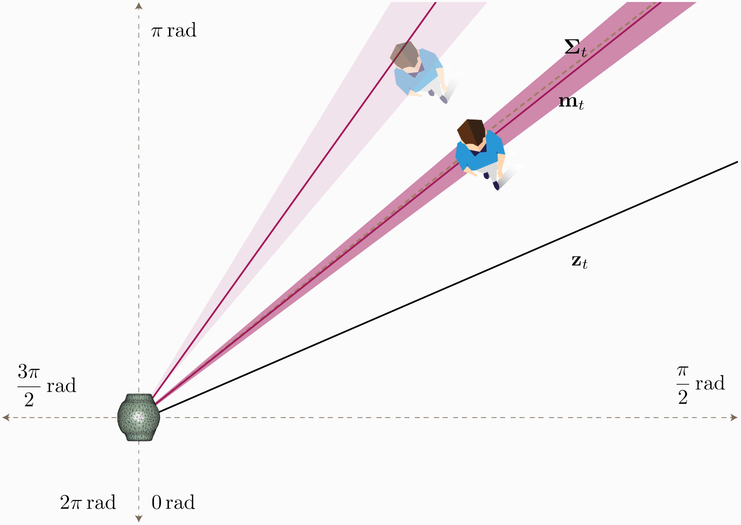
Single Source Tracking - Kalman Filter



Single Source Tracking - Kalman Filter

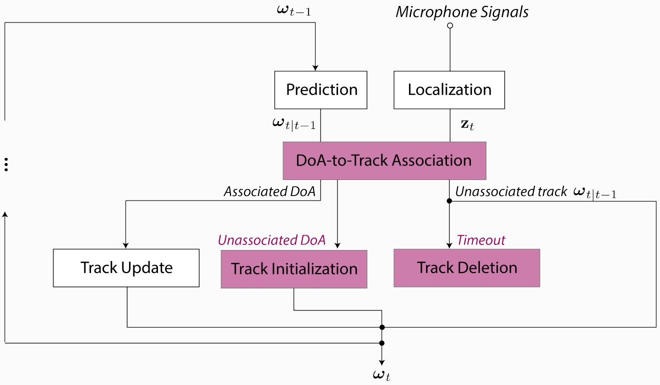


Single Source Tracking - Kalman Filter



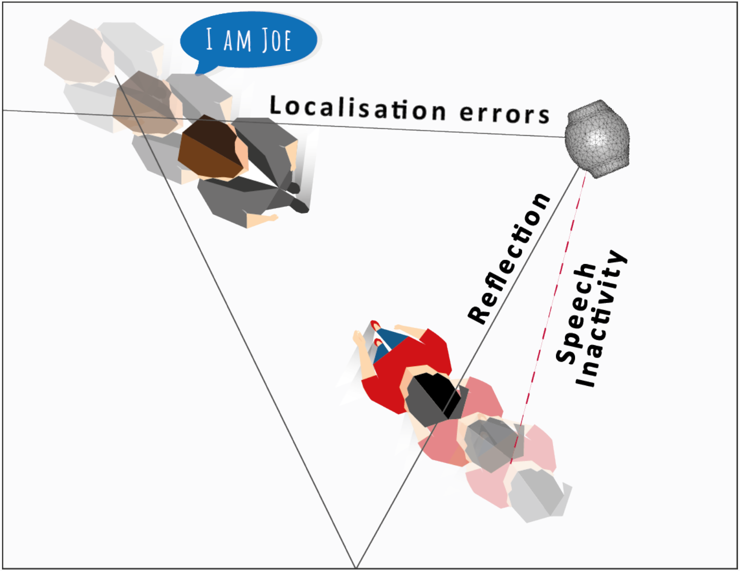
Classical DoA Tracking - Kalman Filter++

- Missing detections due to speech inactivity
- Sources may enter / leave the scene
- Initial source position unknown *a priori*

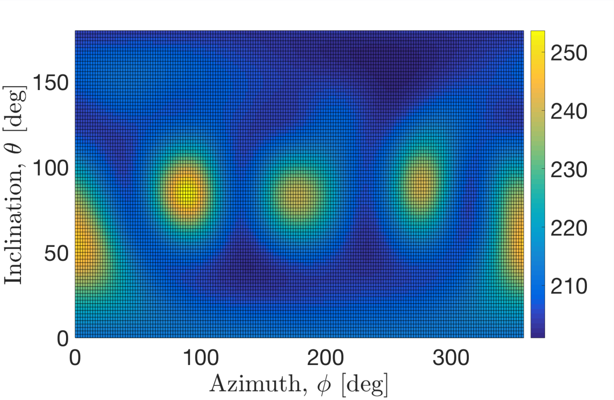


Multi-Source Tracking

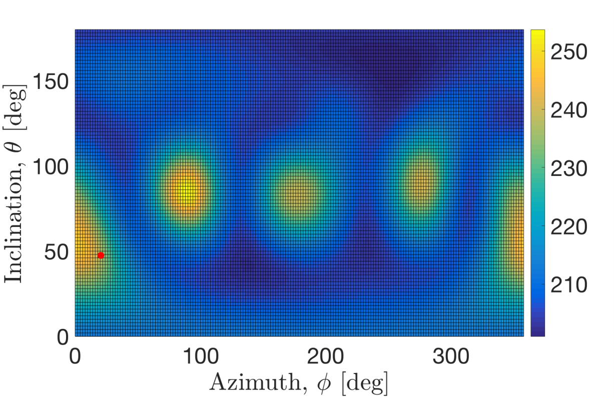
Tracking, spurious & missing detections



Multi-Source Localization



Multi-Source Localization



- Time-varying number of unlabelled measurements
- Unknown, time-varying number of unknown source states

Multi-Hypothesis Tracking (MHT)

- **Hypothesis tree:** Enumerate exhaustively all association hypotheses across time
- Final time step: Backward-propagation to identify most likely path
- Computationally prohibitive for real-time processing

[1] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," IEEE Aerosp. Electr. Sys. Mag., 2004.

[2] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," IEEE Control Syst. Mag., 2010.

Classical Multi-Source Tracking

Imperial College London

Multi-Hypothesis Tracking (MHT)

- **Hypothesis tree:** Enumerate exhaustively all association hypotheses across time
- Final time step: Backward-propagation to identify most likely path
- Computationally prohibitive for real-time processing

Joint Probabilistic Data Association (JPDA)

- At each time step consider all association hypotheses, including spurious detections
- Use pruning after each time step to reduce computational complexity

[1] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," IEEE Aerosp. Electr. Sys. Mag., 2004.

[2] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," IEEE Control Syst. Mag., 2010.

Classical Multi-Source Tracking

Imperial College London

Multi-Hypothesis Tracking (MHT)

- **Hypothesis tree:** Enumerate exhaustively all association hypotheses across time
- Final time step: Backward-propagation to identify most likely path
- Computationally prohibitive for real-time processing

Joint Probabilistic Data Association (JPDA)

- At each time step consider all association hypotheses, including spurious detections
- Use pruning after each time step to reduce computational complexity
- Estimation of the number of sources typically selected heuristically

---

[1] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," IEEE Aerosp. Electr. Sys. Mag., 2004.

[2] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," IEEE Control Syst. Mag., 2010.

Multi-Source Tracking - Set Model

Random Finite Set (RFS) of sources

$$\boldsymbol{\Omega}_t = \left[\bigcup_{n=1}^{N_{t-1}} P(\boldsymbol{\omega}_{t-1,n}) \right] \cup B_t. \tag{4}$$

$$P(\boldsymbol{\omega}_{t-1,n}) = \begin{cases} \boldsymbol{\omega}_{t,n}, & \text{if } \boldsymbol{\omega}_{t-1,n} \text{ persists} \\ \emptyset, & \text{otherwise.} \end{cases} \tag{5}$$

where $|\boldsymbol{\Omega}_t| = N_t$ and B_t models the birth of previously inactive sources.

Multi-Source Tracking - Set Model

Imperial College London

Random Finite Set (RFS) of sources

$$\mathbf{\Omega}_t = \left[\bigcup_{n=1}^{N_{t-1}} P(\boldsymbol{\omega}_{t-1,n})\right] \cup B_t. \tag{4}$$

$$P(\boldsymbol{\omega}_{t-1,n}) = \begin{cases} \boldsymbol{\omega}_{t,n}, & \text{if } \boldsymbol{\omega}_{t-1,n} \text{ persists} \\ \emptyset, & \text{otherwise.} \end{cases} \tag{5}$$

where $|\mathbf{\Omega}_t| = N_t$ and B_t models the birth of previously inactive sources.

Detection RFS of sources and early reflections

$$\mathbf{Z}_t = \left[\bigcup_{n=1}^{N_t} D(\boldsymbol{\omega}_{t,n})\right] \cup C_t, \tag{6}$$

$$D(\boldsymbol{\omega}_{t,n}) = \begin{cases} \mathbf{z}_{t,m}, & \text{if } \boldsymbol{\omega}_{t,n} \text{ detected} \\ \emptyset, & \text{otherwise} \end{cases} \tag{7}$$

where $|\mathbf{Z}_t| = M_t$ and C_t models spurious detections.

Multi-Source Tracking - Generalization

Imperial College London

- Number of sources, N_t, in RFS $\mathbf{\Omega}_t$ is time-varying and unknown

Integrals for random finite sets

$$\int p(\mathbf{\Omega}_t)\,\delta\mathbf{\Omega} \stackrel{?}{=}$$

[1] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," IEEE Tran. Aerosp. Electr. Sys, 2003.

Multi-Source Tracking - Generalization

- Number of sources, N_t, in RFS $\boldsymbol{\Omega}_t$ is time-varying and unknown

Integrals for random finite sets

$$\int p(\boldsymbol{\Omega}_t)\,\delta\boldsymbol{\Omega} \stackrel{?}{=} p(\emptyset).$$

[1] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," IEEE Tran. Aerosp. Electr. Sys, 2003.

- Number of sources, N_t, in RFS $\mathbf{\Omega}_t$ is time-varying and unknown

Integrals for random finite sets

$$\int p(\mathbf{\Omega}_t)\,\delta\mathbf{\Omega} \stackrel{?}{=} p(\emptyset) + \int p(\{\boldsymbol{\omega}_{t,1}\})d\boldsymbol{\omega}_{t,1}.$$

[1] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," IEEE Tran. Aerosp. Electr. Sys, 2003.

Multi-Source Tracking - Generalization

- Number of sources, N_t, in RFS $\boldsymbol{\Omega}_t$ is time-varying and unknown

Integrals for random finite sets

$$\int p(\boldsymbol{\Omega}_t)\,\delta\boldsymbol{\Omega} = p(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int \dots \int p(\{\boldsymbol{\omega}_{t,1}, \dots, \boldsymbol{\omega}_{t,n}\}) d\boldsymbol{\omega}_{t,1} \dots d\boldsymbol{\omega}_{t,n}.$$

[1] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," IEEE Tran. Aerosp. Electr. Sys, 2003.

Multi-Source Tracking - Generalization

- Number of sources, N_t, in RFS $\mathbf{\Omega}_t$ is time-varying and unknown

Integrals for random finite sets

$$\int p(\mathbf{\Omega}_t)\,\delta\mathbf{\Omega} = p(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int \dots \int p(\{\boldsymbol{\omega}_{t,1}, \dots, \boldsymbol{\omega}_{t,n}\}) d\boldsymbol{\omega}_{t,1} \dots d\boldsymbol{\omega}_{t,n}.$$

- Multi-source pdf is combinatorially intractable
- Approximate multi-source pdf by its first order moment:
  - **Probability hypothesis density (PHD)**, $\lambda\left(\boldsymbol{\omega}_t \,|\, \mathbf{Z}_{1:t}\right)$
  - Probability that one of the sources has state $\boldsymbol{\omega}_t$

[1] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," IEEE Tran. Aerosp. Electr. Sys, 2003.

Multi-Source Tracking - Generalization

- Number of sources, N_t, in RFS $\mathbf{\Omega}_t$ is time-varying and unknown

Integrals for random finite sets

$$\int p(\mathbf{\Omega}_t)\,\delta\mathbf{\Omega} = p(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int \dots \int p(\{\boldsymbol{\omega}_{t,1}, \dots, \boldsymbol{\omega}_{t,n}\}) d\boldsymbol{\omega}_{t,1} \dots d\boldsymbol{\omega}_{t,n}.$$

- Multi-source pdf is combinatorially intractable
- Approximate multi-source pdf by its first order moment:
  - **Probability hypothesis density (PHD)**, $\lambda\left(\boldsymbol{\omega}_t \mid \mathbf{Z}_{1:t}\right)$
  - Probability that one of the sources has state $\boldsymbol{\omega}_t$

$$\int \lambda\left(\boldsymbol{\omega}_t \mid \mathbf{Z}_{1:t}\right) d\boldsymbol{\omega}_t = \mathbb{E}\left[N_t\right]$$

[1] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," IEEE Tran. Aerosp. Electr. Sys, 2003.

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

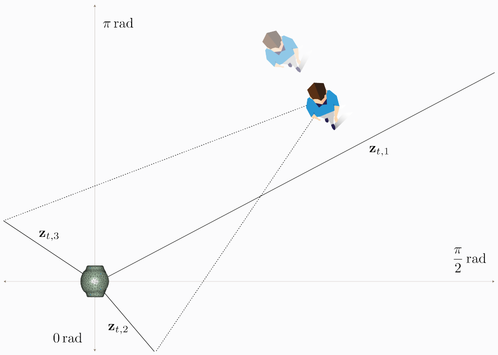
$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$

[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$



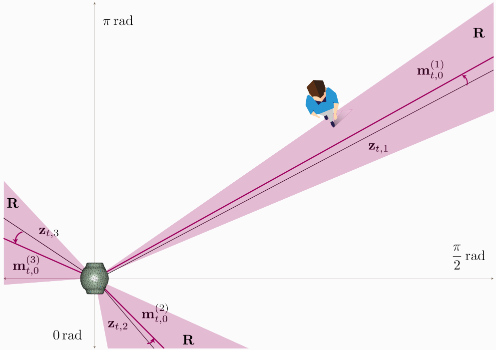
[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Imperial College London

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}(\mathbf{s}_t \mid \mathbf{Z}_t) + \lambda^{\text{missed}}(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}) + \lambda^{\text{detect}}(\mathbf{s}_t \mid \mathbf{Z}_{1:t})$$



[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$

Newborn sources: Each DoA may originate from a new source

$$\lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) = \sum_{m=1}^{M_t} p_b \, \mathcal{N}\left(\mathbf{s}_t \middle| \mathbf{z}_{t,m}, \mathbf{R}\right)$$

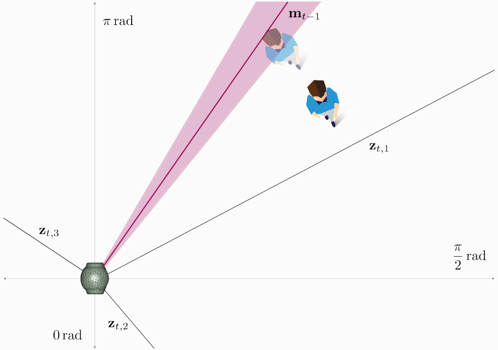
where p_b is the probability of source birth.

[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}(\mathbf{s}_t \mid \mathbf{Z}_t) + \lambda^{\text{missed}}(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}) + \lambda^{\text{detect}}(\mathbf{s}_t \mid \mathbf{Z}_{1:t})$$

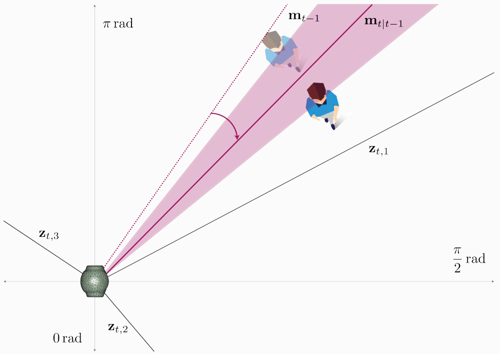


[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$



[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Multi-Source Tracking - Gaussian Mixture PHD

Imperial College London

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}(\mathbf{s}_t \mid \mathbf{Z}_t) + \lambda^{\text{missed}}(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}) + \lambda^{\text{detect}}(\mathbf{s}_t \mid \mathbf{Z}_{1:t})$$

Missed detections: Any previously detected source may be missed:

$$\lambda^{\text{missed}}(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}) = \sum_{j=1}^{J_{t|t-1}} w_{t|t-1}^{(j)} \mathcal{N}\left(\mathbf{s}_t \,\middle|\, \mathbf{m}_{t|t-1}^{(j)}, \boldsymbol{\Sigma}_{t|t-1}^{(j)}\right)$$

where $\mathbf{m}_{t|t-1}$ and $\boldsymbol{\Sigma}_{t|t-1}$: Predicted Kalman filter mean / covariance

[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$

Missed detections: Any previously detected source may be missed:

$$\lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) = \sum_{j=1}^{J_{t|t-1}} w_{t|t-1}^{(j)}\, \mathcal{N}\left(\mathbf{s}_t \,\middle|\, \mathbf{m}_{t|t-1}^{(j)}, \boldsymbol{\Sigma}_{t|t-1}^{(j)}\right)$$

$$w_{t|t-1}^{(j)} = p_s\left(1 - p_d\right) w_{t-1}^{(j)}$$

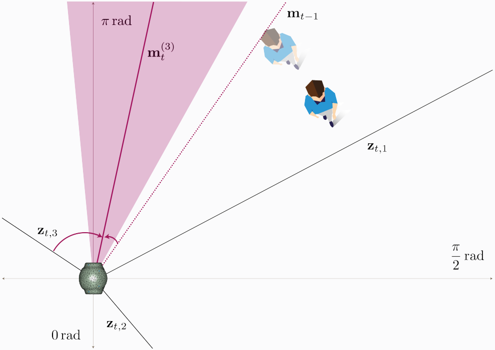
where $\mathbf{m}_{t|t-1}$ and $\boldsymbol{\Sigma}_{t|t-1}$: Predicted Kalman filter mean / covariance

[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$

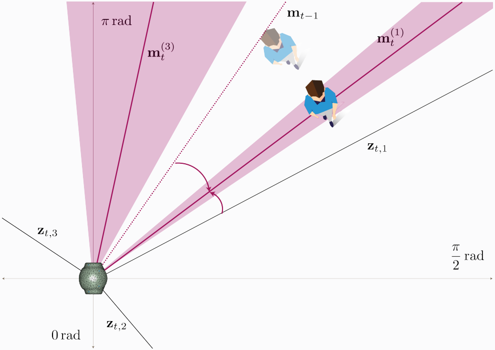


[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$



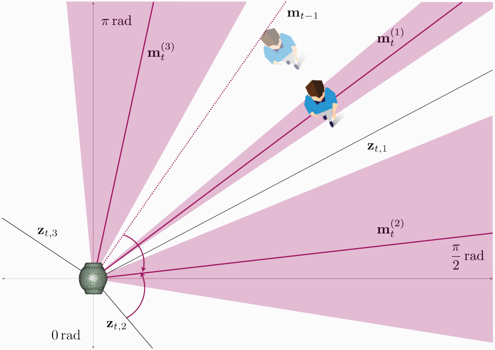
[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Imperial College London

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$



[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$

Detected sources: Any DoA may originate from the source

$$\lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right) = \sum_{m=1}^{M_t} p_d \sum_{j=1}^{J_{t|t-1}} w_t^{(j)} \mathcal{N}\left(\mathbf{s}_t \,\middle|\, \mathbf{m}_t^{(j,m)}, \mathbf{\Sigma}_t^{(j,m)}\right)$$

[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

Multi-Source Tracking - Gaussian Mixture PHD

Imperial College London

Posterior Source PHD

Explicitly model sources, missing detections, and spurious detections:

$$\lambda(\mathbf{s}_t|\mathbf{Z}_{1:t}) = \lambda^{\text{birth}}\left(\mathbf{s}_t \mid \mathbf{Z}_t\right) + \lambda^{\text{missed}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t-1}\right) + \lambda^{\text{detect}}\left(\mathbf{s}_t \mid \mathbf{Z}_{1:t}\right)$$

Detected sources: Any DoA may originate from the source

$$w_t^{(j,m)} = p_d \frac{\mathcal{N}\left(\mathbf{z}_{t,m} \,\middle|\, \mathbf{H}\,\mathbf{m}_{t|t-1}^{(j)}, \mathbf{S}_t^{(j)}\right)}{\kappa\left(\mathbf{z}_{t,m}\right) + \sum\limits_{i=1}^{J_{t|t-1}} \mathcal{N}\left(\mathbf{z}_{t,m} \,\middle|\, \mathbf{H}\,\mathbf{m}_{t|t-1}^{(i)}, \mathbf{S}_t^{(j)}\right)}$$

where $\mathbf{S}_t^{(j)} \triangleq \mathbf{H}\,\mathbf{\Sigma}_{t|t-1}\,\mathbf{H}^T + \mathbf{R}$ and $\kappa\left(\mathbf{z}_{t,m}\right) \triangleq \text{FAR} \times \mathcal{U}\left[\text{FoV}\right]$

[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015

GMMs for Audio-Visual Fusion

- **Audio:** Speech inactivity, reverberation / noise, interference
- **Vision:** Limited field-of-view, visual obstructions, lighting
- Complementary modalities

Multi-Modal Measurements

Visual face detector: $z_t^{(v)} = g_v\left(\mathbf{x}_t, \mathbf{w}_t^{(v)}\right)$

Sound source localization: $z_t^{(a)} = g_a\left(\mathbf{x}_t, \mathbf{w}_t^{(a)}\right)$

Find out the details...

Friday, March 3, 13:00 - 14:40 (BATGIRL):
ROBOT.5: Audio-Visual Tracking by Density Approximation in a Sequential Bayesian Filtering Framework

GMMs for Broadband Speech

- Broadband nature of speech signals
- Peak picking difficult
- Each combination of M_t localized DoAs may contain information about the source

Multi-detection likelihood

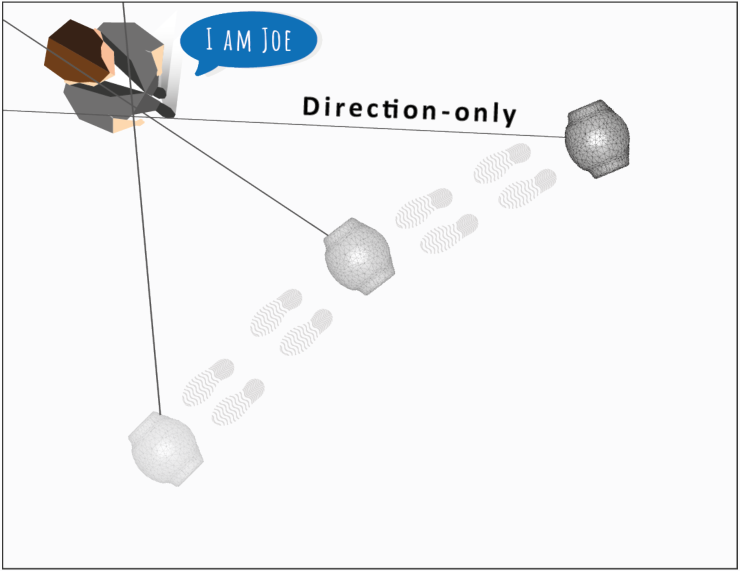
$$p\left(\mathbf{Z}_t \mid \boldsymbol{\omega}_t\right) = \sum_{\mathbf{P} \boxminus \mathbf{Z}_t} \prod_{\ell \in (\mathbf{Z}_t - \mathbf{P})} \kappa\left(\mathbf{z}(t, \ell)\right) \prod_{p \in \mathbf{P}} p\left(\mathbf{z}(t, p) \mid \boldsymbol{\omega}_t\right),$$

Find out the details...

Friday, March 3, 15:10 - 16:50 (BATGIRL):
LOC.4: Speaker Tracking in Reverberant Environments using Multiple Directions of Arrival

Source Position Tracking

Tracking Source Positions



Source Position Tracking - Model

Dynamical Model of $\mathbf{s}_t \triangleq [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$

Source state:

$$\mathbf{s}_t = \mathbf{D}_t \mathbf{s}_{t-1} + \mathbf{v}_t \tag{8}$$

where $\mathbf{D}_t$: Langevin model[1], and Δ_t is the time step

Measurement Model of $\mathbf{z}_t \triangleq \left[\hat{\phi}_t, \hat{\theta}_t\right]^T$

Source direction:

$$\mathbf{z}_t = h\left(\mathbf{s}_t\right) + \mathbf{w}_t \tag{9}$$

where $h(\cdot)$ is the Cartesian-to-spherical transformation.

[1] E. A. Lehmann and R. C. Williamson, "Particle Filter Design Using Importance Sampling for Acoustic Source Localisation and Tracking in Reverberant Environments," EURASIP J. Adv. Signal Process., 2006.

Estimate 6D states from 2D measurements

Estimate 6D states from 2D measurements

Kalman filter (KF) update equation

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(\mathbf{z}_t - h\left(\boldsymbol{\mu}_{t|t-1}\right)\right)$$

where $\boldsymbol{\mu}_{t|t-1}$ and $\boldsymbol{\mu}_t$ are the predicted / updated KF mean of $\mathbf{s}_t$.

Estimate 6D states from 2D measurements

Kalman filter (KF) update equation

$$\underbrace{\boldsymbol{\mu}_t}_{6\times1} = \underbrace{\boldsymbol{\mu}_{t|t-1}}_{6\times1} + \underbrace{\mathbf{K}_t}_{6\times2} (\underbrace{\mathbf{z}_t}_{2\times1} - \underbrace{h(\boldsymbol{\mu}_{t|t-1})}_{2\times1})$$

where $\boldsymbol{\mu}_{t|t-1}$ and $\boldsymbol{\mu}_t$ are the predicted / updated KF mean of $\mathbf{s}_t$.

Unmeasured Range Inference

Estimate 6D states from 2D measurements

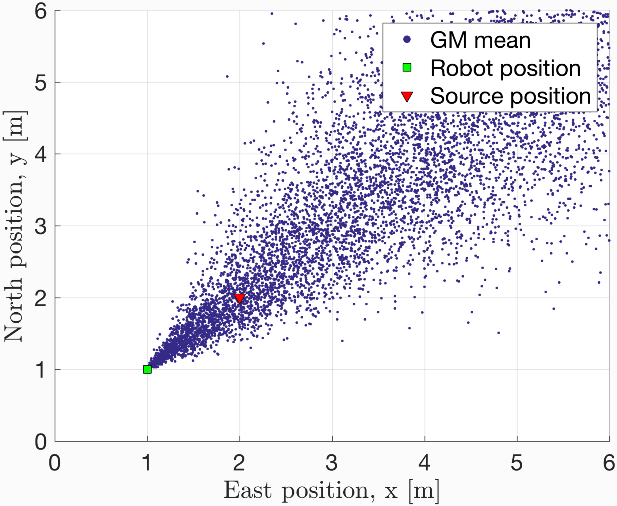
Kalman filter (KF) update equation

$$\underbrace{\boldsymbol{\mu}_t}_{6\times 1} = \underbrace{\boldsymbol{\mu}_{t|t-1}}_{6\times 1} + \underbrace{\mathbf{K}_t}_{6\times 2} (\underbrace{\mathbf{z}_t}_{2\times 1} - \underbrace{h(\boldsymbol{\mu}_{t|t-1})}_{2\times 1})$$

where $\boldsymbol{\mu}_{t|t-1}$ and $\boldsymbol{\mu}_t$ are the predicted / updated KF mean of $\mathbf{s}_t$.

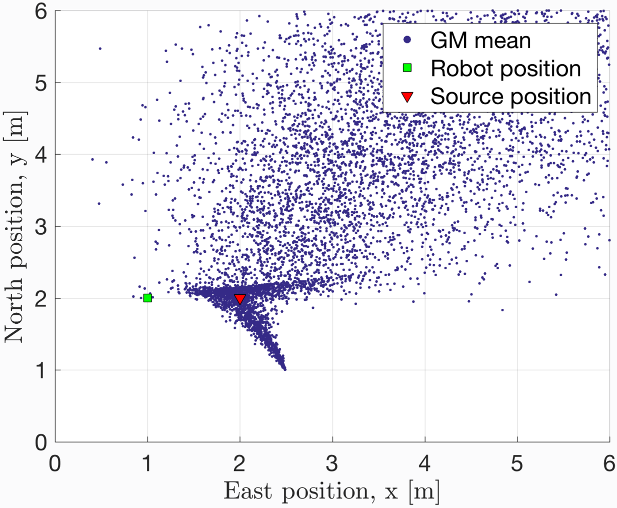
- Kalman gain: Map between state and measurement space
- Pre-populate KF mean with range hypothesis
- New directional information updated through measurements
- Range component is extrapolated through time

Range Inference - Probabilistic Triangulation



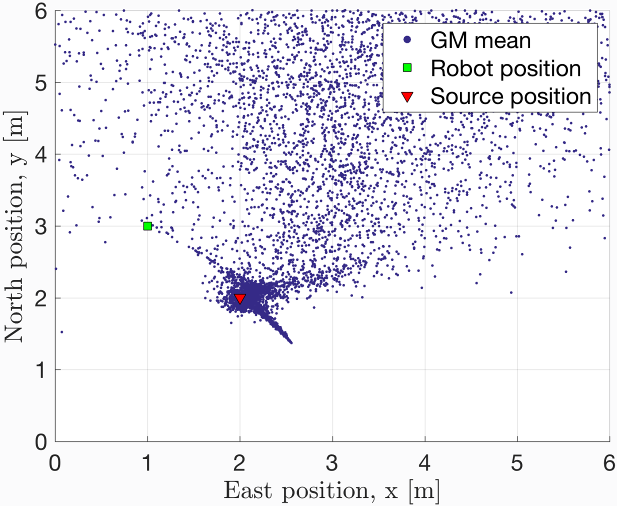
[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015.

Range Inference - Probabilistic Triangulation



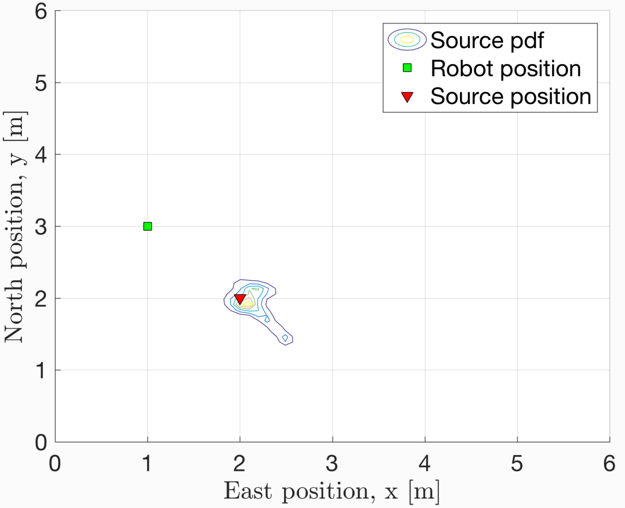
[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015.

Range Inference - Probabilistic Triangulation



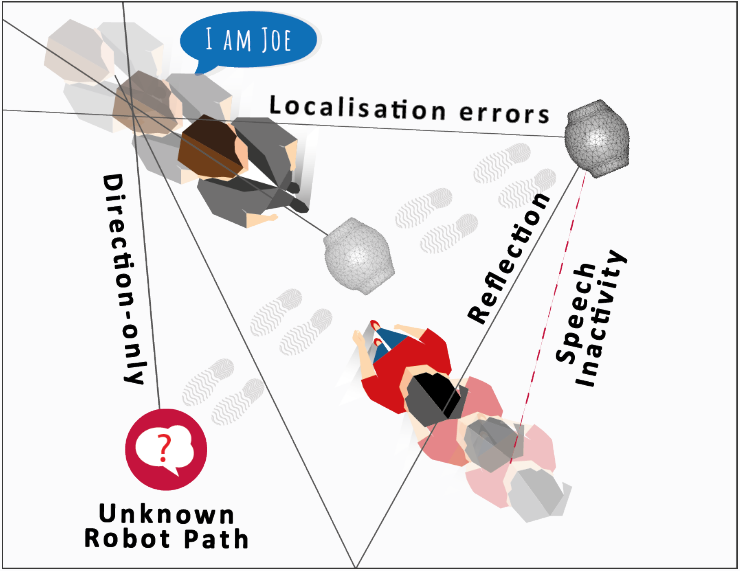
[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015.

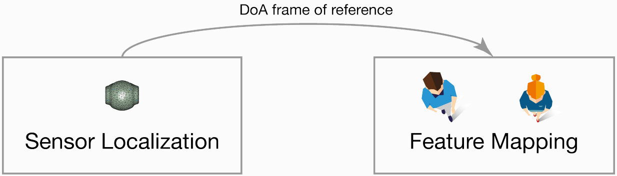
Range Inference - Probabilistic Triangulation



[1] C. Evers *et al.*, "Bearing-only Acoustic Tracking of Moving Speakers for Robot Audition", Proc. IEEE DSP, Singapore, 2015.

Acoustic SLAM


I AM JOE
Localisation errors
Direction-only
Reflection
Speech Inactivity
Unknown Robot Path

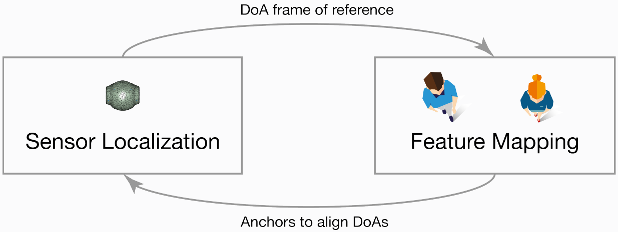


[1] C. Evers *et al.*, "Acoustic Simultaneous Localization and Mapping (a-SLAM) of a Moving Microphone Array and its Surrounding Speakers", Proc. ICASSP, Shanghai, 2016.

[2] C. Evers *et al.*, "Localization of Moving Microphone Arrays from Moving Sound Sources for Robot Audition", Proc. EUSIPCO, Budapest, 2016.

[3] C. Evers, A. H. Moore, and P. A. Naylor, "Towards Informative Path Planning for Acoustic SLAM", Proc. DAGA, Aachen, 2015.

Simultaneous Localization and Mapping

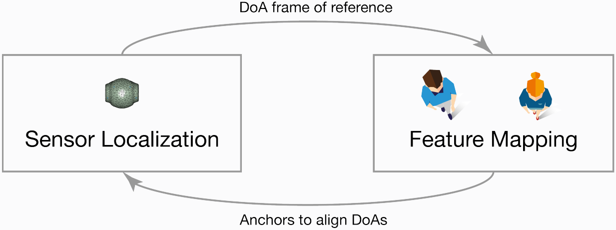


[1] C. Evers *et al.*, "Acoustic Simultaneous Localization and Mapping (a-SLAM) of a Moving Microphone Array and its Surrounding Speakers", Proc. ICASSP, Shanghai, 2016.

[2] C. Evers *et al.*, "Localization of Moving Microphone Arrays from Moving Sound Sources for Robot Audition", Proc. EUSIPCO, Budapest, 2016.

[3] C. Evers, A. H. Moore, and P. A. Naylor, "Towards Informative Path Planning for Acoustic SLAM", Proc. DAGA, Aachen, 2015.

Simultaneous Localization and Mapping



- Acoustic Simultaneous Localization & Mapping (aSLAM):
  - Require sensor position to map moving sources
  - Use the scene map to estimate the sensor position / orientation

[1] C. Evers *et al.*, "Acoustic Simultaneous Localization and Mapping (a-SLAM) of a Moving Microphone Array and its Surrounding Speakers", Proc. ICASSP, Shanghai, 2016.

[2] C. Evers *et al.*, "Localization of Moving Microphone Arrays from Moving Sound Sources for Robot Audition", Proc. EUSIPCO, Budapest, 2016.

[3] C. Evers, A. H. Moore, and P. A. Naylor, "Towards Informative Path Planning for Acoustic SLAM", Proc. DAGA, Aachen, 2015.

SLAM for Audio - Literature

Approach 1: Active SLAM

- Active sensing for room geometry inference and sensor localization
- Cannot estimate source trajectories
- Intrusive / in current form not suitable for human-robot interaction

SLAM for Audio - Literature

Approach 1: Active SLAM

- Active sensing for room geometry inference and sensor localization
- Cannot estimate source trajectories
- Intrusive / in current form not suitable for human-robot interaction

Approach 2: Classic visual-based SLAM

- **Fundamental assumption:** Static, continuously active sources
- Problematic for speech inactivity
- Prone to divergence for high false alarm rates

Acoustic SLAM - Objectives

Imperial College London

Novel, Generalized Framework

- Passive sensing for human-robot interaction
- Exploit directional information gleaned from surrounding sources
- Crucial robustness against reverb, speech inactivity, source motion
- Novel, generalized SLAM framework that jointly estimate:
  a) A PHD filter for multi-source tracking
  b) Estimate robot positions, by exploiting source map, robot reports, prior information about the robot motion

[1] C. Evers and P. A. Naylor, "Generalized Dynamic Scene Mapping", submitted, IEEE Tran. Sig. Proc.

Acoustic SLAM - Objectives

Novel, Generalized Framework

- Passive sensing for human-robot interaction
- Exploit directional information gleaned from surrounding sources
- Crucial robustness against reverb, speech inactivity, source motion
- Novel, generalized SLAM framework that jointly estimate:
  a) A PHD filter for multi-source tracking
  b) Estimate robot positions, by exploiting source map, robot reports, prior information about the robot motion

GEneralized Motion SLAM (GEM-SLAM)

$$\lambda\left(\mathbf{S}_t, \mathbf{r}_t \,|\, \mathbf{Z}_{1:t}, \mathbf{y}_{1:t}\right) = \lambda\left(\mathbf{r}_t \,|\, \mathbf{y}_{1:t}\right) \lambda\left(\mathbf{S}_t \,|\, \mathbf{r}_t, \mathbf{Z}_{1:t}\right)$$

[1] C. Evers and P. A. Naylor, "Generalized Dynamic Scene Mapping", submitted, IEEE Tran. Sig. Proc.

Novel, Generalized Framework

- Passive sensing for human-robot interaction
- Exploit directional information gleaned from surrounding sources
- Crucial robustness against reverb, speech inactivity, source motion
- Novel, generalized SLAM framework that jointly estimate:
  a) A PHD filter for multi-source tracking
  b) Estimate robot positions, by exploiting source map, robot reports, prior information about the robot motion

GEneralized Motion SLAM (GEM-SLAM)

$$\lambda\left(\mathbf{\Omega}_t, \mathbf{r}_t \mid \mathbf{Z}_{1:t}, \mathbf{y}_{1:t}\right) = \lambda\left(\mathbf{r}_t \mid \mathbf{y}_{1:t}\right) \overbrace{\lambda\left(\mathbf{S}_t \mid \mathbf{r}_t, \mathbf{Z}_{1:t}\right)}^{\text{Source PHD}}$$

[1] C. Evers and P. A. Naylor, "Generalized Dynamic Scene Mapping", submitted, IEEE Tran. Sig. Proc.

Acoustic SLAM - Objectives

Novel, Generalized Framework

- Passive sensing for human-robot interaction
- Exploit directional information gleaned from surrounding sources
- Crucial robustness against reverb, speech inactivity, source motion
- Novel, generalized SLAM framework that jointly estimate:
  a) A PHD filter for multi-source tracking
  b) Estimate robot positions, by exploiting source map, robot reports, prior information about the robot motion

GEneralized Motion SLAM (GEM-SLAM)

$$\lambda\left(\mathbf{\Omega}_t, \mathbf{r}_t \mid \mathbf{Z}_{1:t}, \mathbf{y}_{1:t}\right) = \overbrace{\lambda\left(\mathbf{r}_t \mid \mathbf{y}_{1:t}\right)}^{\text{Robot PHD}} \lambda\left(\mathbf{S}_t \mid \mathbf{r}_t, \mathbf{Z}_{1:t}\right)$$

[1] C. Evers and P. A. Naylor, "Generalized Dynamic Scene Mapping", submitted, IEEE Tran. Sig. Proc.

Executed robot state, $\mathbf{r}_t \triangleq \left[x_t, y_t, z_t, v_t, \gamma_t\right]^T = \left[\mathbf{p}_t^T, \gamma_t\right]^T$:

$$\mathbf{p}_t = \mathbf{G}\left(\gamma_t\right)\,\mathbf{p}_{t-1} + \mathbf{v}_{t,\mathbf{p}} \qquad \mathbf{G}\left(\gamma_t\right) \triangleq \begin{bmatrix} 1 & 0 & 0 & \Delta_t \sin\gamma_t \\ 0 & 1 & 0 & \Delta_t \cos\gamma_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\gamma_t = \vartheta(\gamma_{t-1} + v_{t,\gamma})$$

where $\mathbf{p}_t \triangleq \left[x_t, y_t, z_t, v_t\right]^T$ with speed, v_t, and orientation, γ_t.

Executed robot state, $\mathbf{r}_t \triangleq \left[x_t, y_t, z_t, v_t, \gamma_t\right]^T = \left[\mathbf{p}_t^T, \gamma_t\right]^T$:

$$\mathbf{p}_t = \mathbf{G}\left(\gamma_t\right) \mathbf{p}_{t-1} + \mathbf{v}_{t,\mathbf{p}}$$
$$\gamma_t = \vartheta(\gamma_{t-1} + v_{t,\gamma})$$

$$\mathbf{G}\left(\gamma_t\right) \triangleq \begin{bmatrix} 1 & 0 & 0 & \Delta_t \sin\gamma_t \\ 0 & 1 & 0 & \Delta_t \cos\gamma_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where $\mathbf{p}_t \triangleq \left[x_t, y_t, z_t, v_t\right]^T$ with speed, v_t, and orientation, γ_t.

Reported robot state: $\mathbf{y}_t \triangleq \left[y_{t,v} \quad y_{t,\gamma}\right]^T$

$$y_{t,v} = v_t + w_{t,v}, \qquad w_{t,v} \sim \mathcal{N}\left(0,\, \sigma^2_{w_{t,v}}\right)$$
$$y_{t,\gamma} = \vartheta\left(\gamma_t + w_{t,\gamma}\right), \qquad w_{t,\gamma} \sim \mathcal{N}\left(0,\, \sigma^2_{w_{t,\gamma}}\right)$$

GEM-SLAM - Model

Executed robot state, $\mathbf{r}_t \triangleq [x_t, y_t, z_t, v_t, \gamma_t]^T = [\mathbf{p}_t^T, \gamma_t]^T$:

$$\mathbf{p}_t = \mathbf{G}(\gamma_t)\, \mathbf{p}_{t-1} + \mathbf{v}_{t,\mathbf{p}}$$
$$\gamma_t = \vartheta(\gamma_{t-1} + v_{t,\gamma})$$

$$\mathbf{G}(\gamma_t) \triangleq \begin{bmatrix} 1 & 0 & 0 & \Delta_t \sin\gamma_t \\ 0 & 1 & 0 & \Delta_t \cos\gamma_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where $\mathbf{p}_t \triangleq [x_t, y_t, z_t, v_t]^T$ with speed, v_t, and orientation, γ_t.

Reported robot state: $\mathbf{y}_t \triangleq [y_{t,v} \quad y_{t,\gamma}]^T$

$$y_{t,v} = v_t + w_{t,v}, \qquad w_{t,v} \sim \mathcal{N}\left(0,\, \sigma^2_{w_{t,v}}\right)$$
$$y_{t,\gamma} = \vartheta\left(\gamma_t + w_{t,\gamma}\right), \qquad w_{t,\gamma} \sim \mathcal{N}\left(0,\, \sigma^2_{w_{t,\gamma}}\right)$$

Non-linear orientation and dependency of position on orientation

- Particle filter to approximate robot density
- Classic SLAM: Prior importance sampling

GEM-SLAM - Optimized Particle Filter

GEM-SLAM - Optimized Particle Filter

Optimized particle filter of $\hat{\mathbf{r}}_t^{(i)} = \left[\left[\hat{\mathbf{p}}_t^{(i)}\right]^T, \hat{\gamma}_t^{(i)}\right]^T$

- Sampling of orientation, $\hat{\gamma}_t^{(i)}$, from wrapped Kalman filter:
- For each $\hat{\gamma}_t^{(i)}$, optimal sampling of pose, $\hat{\mathbf{p}}_t^{(i)}$, from Kalman filter

$$\lambda\left(\mathbf{r}_t \,|\, \mathbf{y}_{1:t}, \mathbf{Z}_t\right) \approx \sum_{i=1}^{L} \alpha_t^{(i)} \, \delta_{\hat{\mathbf{r}}_t^{(i)}}\left(\mathbf{r}_t\right).$$

GEM-SLAM - Optimized Particle Filter

Optimized particle filter of $\widehat{\mathbf{r}}_t^{(i)} = \left[\left[\widehat{\mathbf{p}}_t^{(i)} \right]^T, \widehat{\gamma}_t^{(i)} \right]^T$

- Sampling of orientation, $\widehat{\gamma}_t^{(i)}$, from wrapped Kalman filter:
- For each $\widehat{\gamma}_t^{(i)}$, optimal sampling of pose, $\widehat{\mathbf{p}}_t^{(i)}$, from Kalman filter

$$\lambda \left(\mathbf{r}_t \,|\, \mathbf{y}_{1:t}, \mathbf{Z}_t \right) \approx \sum_{i=1}^{L} \alpha_t^{(i)} \, \delta_{\widehat{\mathbf{r}}_t^{(i)}} \left(\mathbf{r}_t \right).$$

Particle weights

$$\alpha_t^{(i)} = \alpha_{t-1}^{(i)} \, p(\mathbf{z}_t | \widehat{\gamma}_t^{(i)}, \widehat{\mathbf{p}}_t^{(i)}) \, \mathcal{L}(\mathbf{Z}_t | \mathbf{r}_t^{(i)})$$

Optimized particle filter of $\hat{\mathbf{r}}_t^{(i)} = \left[\left[\hat{\mathbf{p}}_t^{(i)} \right]^T, \hat{\gamma}_t^{(i)} \right]^T$

- Sampling of orientation, $\hat{\gamma}_t^{(i)}$, from wrapped Kalman filter:
- For each $\hat{\gamma}_t^{(i)}$, optimal sampling of pose, $\hat{\mathbf{p}}_t^{(i)}$, from Kalman filter

$$\lambda\left(\mathbf{r}_t \mid \mathbf{y}_{1:t}, \mathbf{Z}_t\right) \approx \sum_{i=1}^{L} \alpha_t^{(i)} \, \delta_{\hat{\mathbf{r}}_t^{(i)}}\left(\mathbf{r}_t\right).$$

Particle weights

$$\alpha_t^{(i)} = \underbrace{\alpha_{t-1}^{(i)}}_{\text{Previous particle weight}} \quad p(\mathbf{y}_t | \hat{\gamma}_t^{(i,p)}, \hat{\mathbf{p}}_t^{(i,p)}) \, \mathcal{L}(\mathbf{Z}_t | \mathbf{r}_t^{(i,p)})$$

GEM-SLAM - Optimized Particle Filter

Imperial College London

Optimized particle filter of $\hat{\mathbf{r}}_t^{(i)} = \left[\left[\hat{\mathbf{p}}_t^{(i)}\right]^T, \hat{\gamma}_t^{(i)}\right]^T$

- Sampling of orientation, $\hat{\gamma}_t^{(i)}$, from wrapped Kalman filter:
- For each $\hat{\gamma}_t^{(i)}$, optimal sampling of pose, $\hat{\mathbf{p}}_t^{(i)}$, from Kalman filter

$$\lambda\left(\mathbf{r}_t \mid \mathbf{y}_{1:t}, \mathbf{Z}_t\right) \approx \sum_{i=1}^{L} \alpha_t^{(i)} \delta_{\hat{\mathbf{r}}_t^{(i)}}\left(\mathbf{r}_t\right).$$

Particle weights

$$\alpha_t^{(i)} = \alpha_{t-1}^{(i)} \underbrace{p(\mathbf{y}_t | \hat{\gamma}_t^{(i)}, \hat{\mathbf{p}}_t^{(i)})}_{\text{Likelihood of robot reports}} \mathcal{L}(\mathbf{Z}_t | \mathbf{r}_t^{(i)})$$

GEM-SLAM - Optimized Particle Filter

Optimized particle filter of $\hat{\mathbf{r}}_t^{(i)} = \left[\left[\hat{\mathbf{p}}_t^{(i)} \right]^T, \hat{\gamma}_t^{(i)} \right]^T$

- Sampling of orientation, $\hat{\gamma}_t^{(i)}$, from wrapped Kalman filter:
- For each $\hat{\gamma}_t^{(i)}$, optimal sampling of pose, $\hat{\mathbf{p}}_t^{(i)}$, from Kalman filter

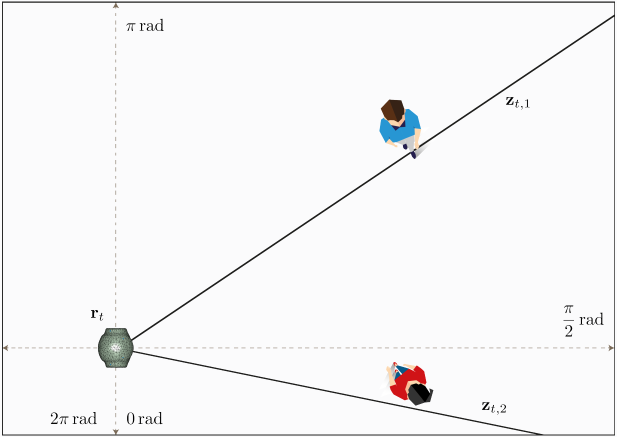
$$\lambda \left(\mathbf{r}_t \mid \mathbf{y}_{1:t}, \mathbf{Z}_t \right) \approx \sum_{i=1}^{L} \alpha_t^{(i)} \, \delta_{\hat{\mathbf{r}}_t^{(i)}} \left(\mathbf{r}_t \right).$$

Particle weights

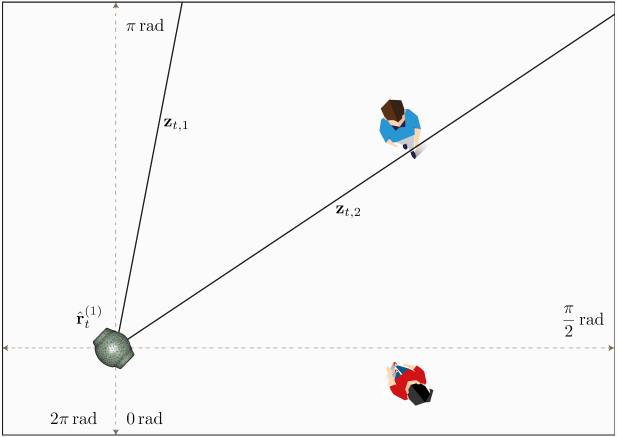
$$\alpha_t^{(i)} = \alpha_{t-1}^{(i)} \, p(\mathbf{y}_t | \hat{\gamma}_t^{(i)}, \hat{\mathbf{p}}_t^{(i)}) \qquad \underbrace{\mathcal{L}(\mathbf{Z}_t | \mathbf{r}_t^{(i)})}_{\text{Multi-measurement evidence}}$$

where $\mathcal{L} \left(\mathbf{Z}_t \mid \mathbf{r}_t \right) = e^{-N_{t|t-1} - N_{t,c}} \prod_{m=1}^{M_t} \left[\kappa \left(\mathbf{z}_{t,m} \mid \mathbf{r}_t \right) + p \left(\mathbf{z}_{t,m} \mid \mathbf{r}_t, \mathbf{Z}_{1:t-1} \right) \right]$

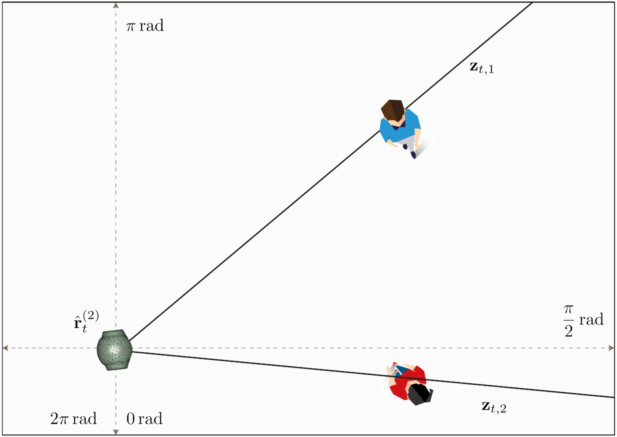
Probabilistic anchoring by weighting with multi-measurement evidence



Probabilistic anchoring by weighting with multi-measurement evidence



Probabilistic anchoring by weighting with multi-measurement evidence



LOCATA Challenge

IEEE AASP Challenge: LOCalization and TrAcking (LOCATA)

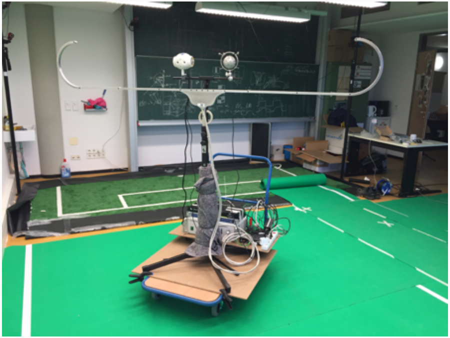
H. Löllmann, C. Evers, H. Barfuß, P. A. Naylor, W. Kellermann

IEEE AASP Challenge: LOCalization and TrAcking (LOCATA)

H. Löllmann, C. Evers, H. Barfuß, P. A. Naylor, W. Kellermann

Static & Moving Arrays:

- 15-channel linear array
- 12-channel robot head
- 32-channel spherical mh-acoustic eigenmike



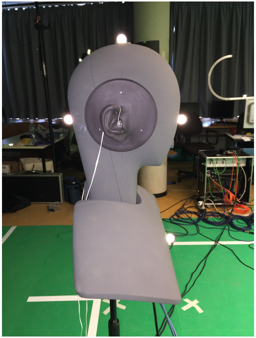
LOCATA Challenge & Corpus

IEEE AASP Challenge: LOCalization and TrAcking (LOCATA)

H. Löllmann, C. Evers, H. Barfuß, P. A. Naylor, W. Kellermann

Static & Moving arrays:

- 15-channel Linear array
- 12-channel robot head
- 32-channel spherical mh-acoustic eigenmike
- 4-channel hearing aids



Imperial College London

EPSRC Engineering and Physical Sciences Research Council

IEEE AASP Challenge: LOCalization and TrAcking (LOCATA)

H. Löllmann, C. Evers, H. Barfuß, P. A. Naylor, W. Kellermann

Static & Moving arrays:

- 15-channel Linear array
- 12-channel robot head
- 32-channel spherical mh-acoustic eigenmike
- 4-channel hearing aids

Static loudspeakers

LOCATA Challenge & Corpus

IEEE AASP Challenge: LOCalization and TrAcking (LOCATA)
H. Löllmann, C. Evers, H. Barfuß, P. A. Naylor, W. Kellermann

Static & Moving arrays:

- 15-channel Linear array
- 12-channel robot head
- 32-channel spherical mh-acoustic eigenmike
- 2-channel binaural hearing aids

Static loudspeakers
Moving talkers

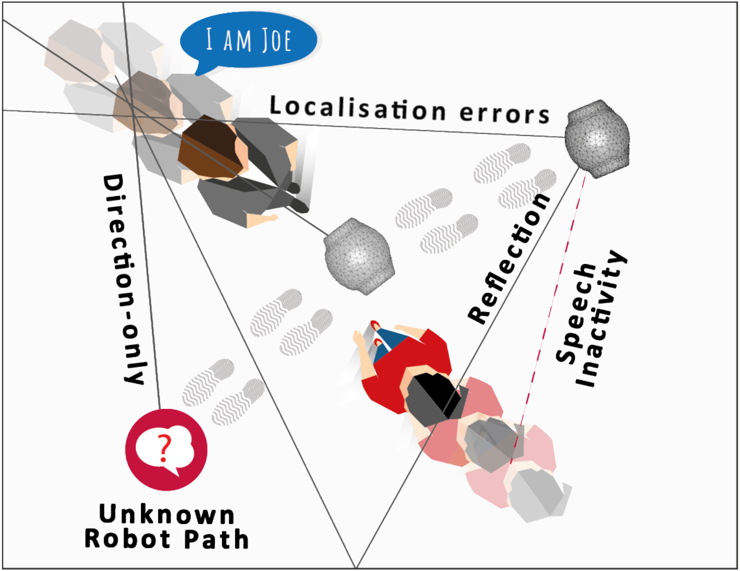
LOCATA Challenge & Corpus

IEEE AASP Challenge: LOCalization and TrAcking (LOCATA)

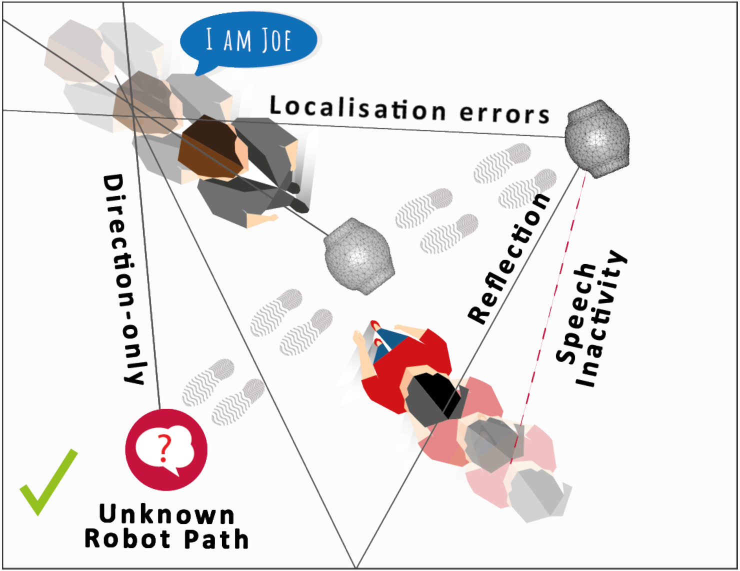
H. Löllmann, C. Evers, H. Barfuß, P. A. Naylor, W. Kellermann

@ LOCATAChallenge

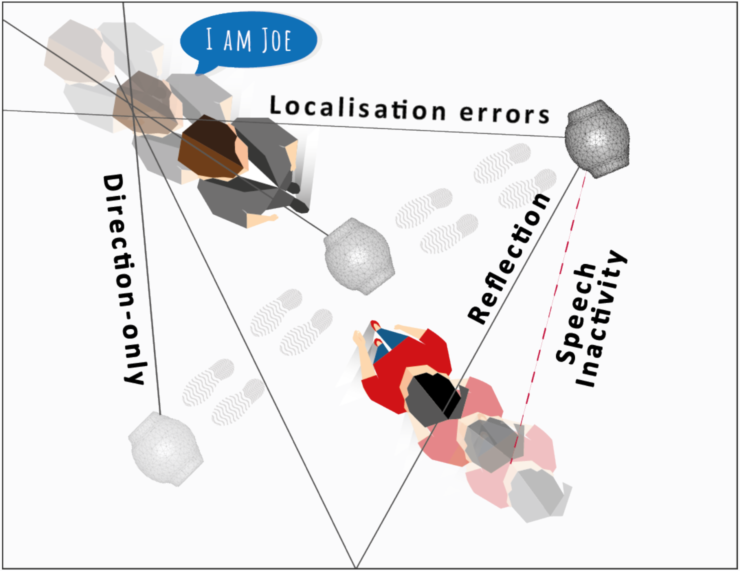
Conclusion

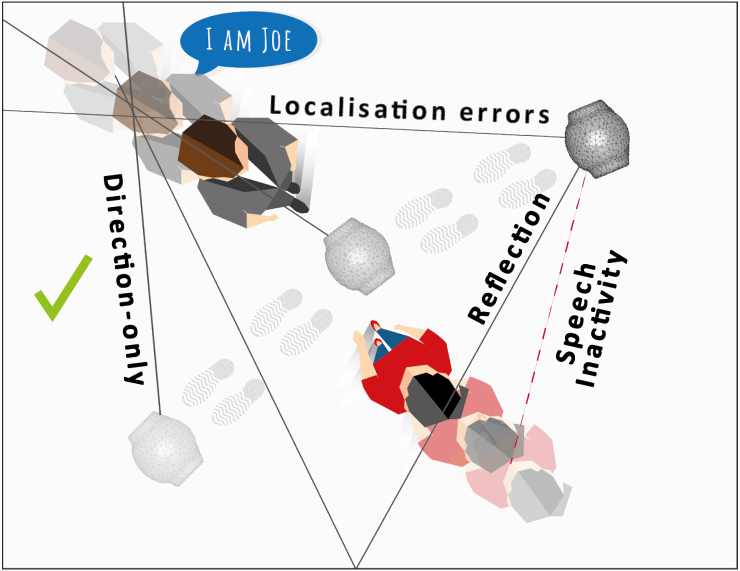

I AM JOE
Localisation errors
Direction-only
Reflection
Speech Inactivity
Unknown Robot Path

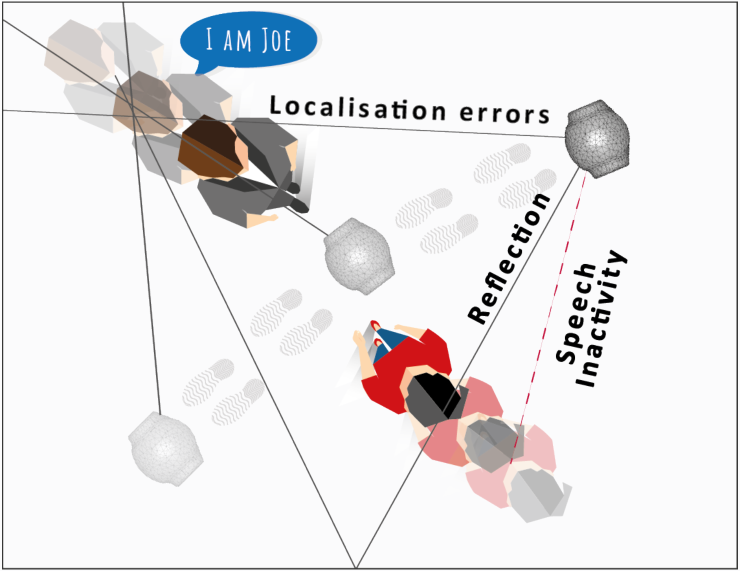
Imperial College London

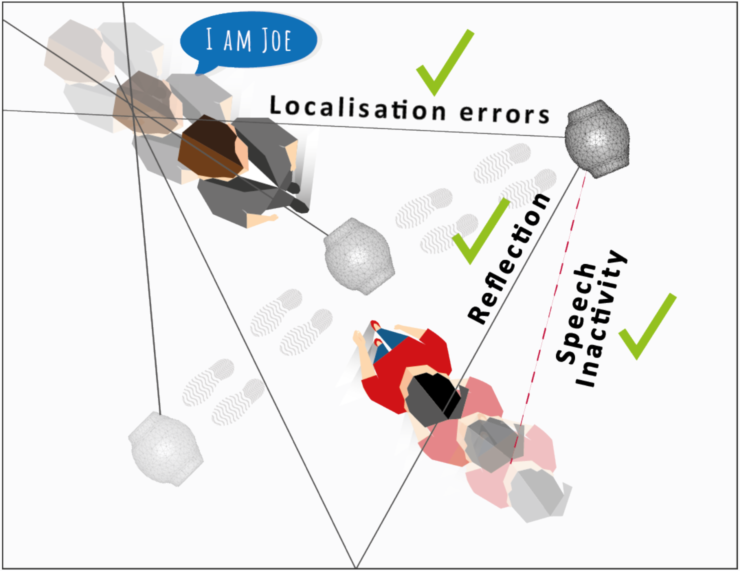


Conclusion


I AM JOE
Localisation errors
Direction-only
Reflection
Speech Inactivity

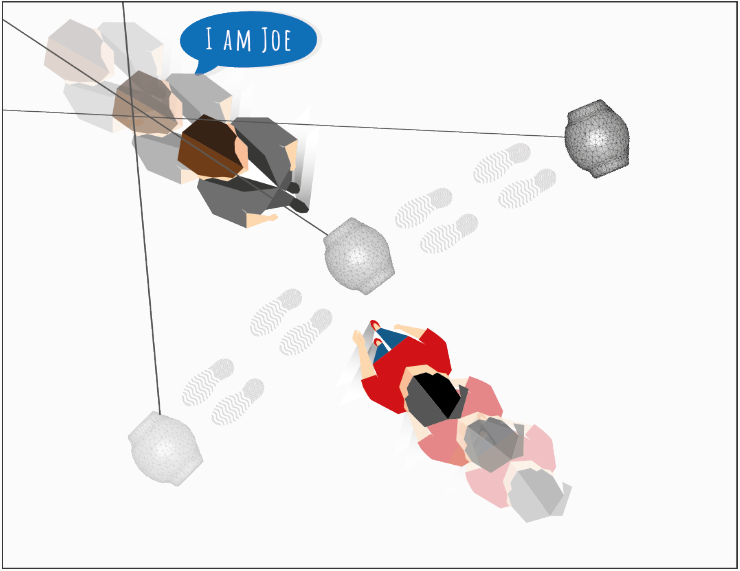

I AM JOE
Localisation errors
Direction-only
Reflection
Speech Inactivity


I AM JOE
Localisation errors
Reflection
Speech Inactivity


I AM JOE
Localisation errors
Reflection
Speech
Inactivity

Conclusion

Imperial College London



Acknowledgement

- **The HSCMA Organizing Committee**
- **Collaborators and Co-Authors:**
  - Patrick A. Naylor (Imperial College London)
  - Sharon Gannot, Yuval Dorfan (Bar-Ilan University)
  - Boaz Rafaely (Ben-Gurion University)
  - Radu Horaud, Israel D. Gebru (INRIA)
  - Walter Kellermann, Heinrich Löllmann, Hendrik Barfuss, Alexander Schmidt (Friedrich Alexander University)
  - Verena Hafner, Heinrich Mellmann, Claas-Norman Ritter (Humboldt University Berlin)
  - James R. Hopgood (University of Edinburgh)
  - Daniel Clark (Heriot Watt University)
- **Funding**
  - EPSRC Fellowship "Acoustic Signal Processing and Scene Analysis for Socially Assistive Robots"
  - FP7 "Embodied Audition for RobotS" (EARS)