Sequential Estimation Methods for Acoustic Source Localization

Ph.D. mid-term review seminar

Eric A. Lehmann
Department of Telecommunications Engineering
Research School of Information Sciences and Engineering
The Australian National University, Canberra, Australia
Eric.Lehmann@anu.edu.au

April 16, 2003
Presentation Outline

- **Introduction**: existing concepts and theory
  - Acoustic source localization and tracking
  - Bayesian filtering
  - Sequential estimation and particle filtering

- **Contribution of current Ph.D. research**: 
  - Particle filters applied to acoustic source localization
  - Results obtained so far

- **Future Ph.D. research**: 
  - Further tasks
  - Possible approaches
**Source Localization with Sensor Array**

**Acoustic Source Localization (ASL): 2D problem formulation**

**Aim:** determine the physical location of an acoustic source using signals $s_m(t), m = 1, \ldots, M$, received at an array of $M$ sensors
Traditional array-based methods:

- Steered Beamforming (SBF)
- Methods based on Time Delay Estimates (TDE) including:
  - Cross-Correlation Function (CCF)
  - Adaptive Eigenvalue Decomposition Algorithm (AEDA)
    

Working principle of traditional methods:

- Transformation of the signals $s_m(t), m = 1, \ldots, M$, into a localization function that exhibits a peak at the source position
ASL: Traditional TDE Approaches

① TDE-based methods

For example, Cross-Correlation Function (CCF):

- **Signal model:**
  \[
  s_i(t) = \alpha_i s(t - \tau_i) + v_i(t)  \\
  s_j(t) = \alpha_j s(t - \tau_j) + v_j(t)
  \]

- **CCF:**
  \[
  R_{ij}(\tau) = \int_{-\infty}^{\infty} s_i(t)s_j(t + \tau) \, dt
  \]

⇒ Determine TDE\(_{ij}\) as \(\arg \max_{\tau} \{R_{ij}(\tau)\}\)

- Given a vector of TDE’s obtained from \(P\) sensor pairs, estimate source location as that minimizing some LS criterion
ASL: Traditional TDE Approaches

**Properties**: traditional TDE methods

- TDE-based approaches are **indirect** (two step) methods:
  1. **Compute localization function** (e.g. CCF) and determine TDE for \( P \) sensor pairs
  2. **Combine TDE’s** to determine source location estimate

- Usually require \( P \) one-dimensional searches over scalar space of possible time delays

Examples of GCC localization functions:

- **Ideal conditions**: low level of reverberation, high SNR
Examples of GCC localization functions:

- **Ideal conditions**: low level of reverberation, high SNR

- **Practical conditions**: reverberation, low SNR
Steered Beamforming (SBF) methods:

- Beamformer measures the average power $P(\ell)$ as a function of location $\ell = [x, y, z]$.

- Basic delay-and-sum beamformer:
  
  $$
  P(\ell) = \int_{-\infty}^{\infty} \left| \sum_{m=1}^{M} s_m(t - \tau_m) \right|^2 \, dt
  $$

  with steering delays $\tau_m \triangleq \tau_m(\ell), m = 1, \ldots, M$.

- Determine source location estimate $\hat{\ell}_S$ as $\arg\max_{\ell}\{P(\ell)\}$.
ASL: Traditional SBF Approach

Properties: traditional SBF method

- SBF approach is a direct (single step) method:
  - Searching the SBF output power function delivers the source location estimate directly

- Requires one multi-dimensional search over potential source locations (potentially computationally very demanding)

- More sophisticated versions of the basic delay-and-sum beamformer: filter-and-sum, frequency weights, etc.
Example of SBF localization function: practical conditions
Problem definitions:

• **State variable** $\mathbf{x}_k$: source position and velocity in state-space at time $k$

  \[
  \mathbf{x}_k = \begin{bmatrix} x, y, z, \dot{x}, \dot{y}, \dot{z} \end{bmatrix}^T
  \]

• **Observation (measurement)** $\mathbf{y}_k$: localization function computed from microphone array data (e.g. GCCF, SBF, etc.)

• **Set of all observations**: $\mathbf{y}_{1:k} = [\mathbf{y}_1, \ldots, \mathbf{y}_k]$

• **System dynamics (transition) equation**: $\mathbf{x}_k = g(\mathbf{x}_{k-1}, \mathbf{u}_{k-1})$
ASL: State-Space Approach

Problem definitions:

- **State variable** $\mathbf{x}_k$: source position and velocity in state-space at time $k$
  \[ \mathbf{x}_k = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T \]

- **Observation (measurement)** $\mathbf{y}_k$: localization function computed from microphone array data (e.g. GCCF, SBF, etc.)

- **Set of all observations**: $\mathbf{y}_{1:k} = [\mathbf{y}_1, \ldots, \mathbf{y}_k]$

- **System dynamics (transition) equation**: $\mathbf{x}_k = g(\mathbf{x}_{k-1}, \mathbf{v}_{k-1})$

**Aim**: given data $\mathbf{y}_{1:k}$, compute posterior PDF $p(\mathbf{x}_k | \mathbf{y}_{1:k})$

$\Rightarrow$ Bayesian filtering problem
ASL: State-Space Approach

- **Bayesian filtering solution:** if posterior PDF \( p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \) known at time \( k-1 \), compute current posterior PDF as follows:

  **Predict:** \( p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \, d\mathbf{x}_{k-1} \)

  **Update:** \( p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) \)

where \( p(\mathbf{y}_k|\mathbf{x}_k) \) is the **likelihood function** (measurement PDF)
ASL: State-Space Approach

- **Bayesian filtering solution**: if posterior PDF $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ known at time $k-1$, compute current posterior PDF as follows:

  **Predict:**
  
  $p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \, d\mathbf{x}_{k-1}$

  **Update:**
  
  $p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$

  where $p(\mathbf{y}_k|\mathbf{x}_k)$ is the likelihood function (measurement PDF)

- **Problem**: usually no closed-form solutions available for practical cases
ASL: State-Space Approach

- **Bayesian filtering solution**: if posterior PDF $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ known at time $k-1$, compute current posterior PDF as follows:

  **Predict**: $p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \, d\mathbf{x}_{k-1}$

  **Update**: $p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$

  where $p(\mathbf{y}_k|\mathbf{x}_k)$ is the **likelihood function** (measurement PDF)

- **Problem**: usually no closed-form solutions available for practical cases

- **Current approximations**: Kalman filter, extended Kalman filter, Gaussian sum methods, grid-based methods, etc.
ASL: State-Space Approach

• **Bayesian filtering solution:** if posterior PDF $p(x_{k-1}|y_{1:k-1})$ known at time $k - 1$, compute current posterior PDF as follows:

  Predict: $p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1}) p(x_{k-1}|y_{1:k-1}) \, dx_{k-1}$

  Update: $p(x_k|y_{1:k}) \propto p(y_k|x_k) p(x_k|y_{1:k-1})$

  where $p(y_k|x_k)$ is the likelihood function (measurement PDF)

• **Problem:** usually no closed-form solutions available for practical cases

• **Current approximations:** Kalman filter, extended Kalman filter, Gaussian sum methods, grid-based methods, etc.

  ⇝ **Sequential Monte Carlo** methods, i.e. **Particle Filters** (PF)
Basic PF Algorithm

Assumption: a set of $N$ state samples and corresponding weights $\{x_{k-1}^{(i)}, w_{k-1}^{(i)}, i = 1, \ldots, N\}$ represents the posterior density $p(x_{k-1}|y_{1:k-1})$ at time $k - 1$.

Procedure: update the particle set to represent the posterior density $p(x_k|y_{1:k})$ for current time $k$ according to following iterations.
\[ \{ \chi_{k-1}^{(i)}, w_{k-1}^{(i)} \} \sim p(\chi_{k-1} | \mathbf{Y}_{1:k-1}) \]
Basic PF: Symbolic Representation

\[
\{ \mathbf{x}^{(i)}_{k-1}, w^{(i)}_{k-1} \} \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \leftarrow \text{resampling}
\]

\[
\{ \tilde{\mathbf{x}}^{(i)}_{k-1}, 1/N \} \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})
\]
Basic PF: Symbolic Representation

\[ \{ \mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)} \} \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \leftarrow \text{resampling} \]

\[ \{ \tilde{\mathbf{x}}_{k-1}^{(i)}, 1/N \} \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \leftarrow \text{prediction} \]

\[ \{ \mathbf{x}_{k}^{(i)}, 1/N \} \sim p(\mathbf{x}_{k} | \mathbf{y}_{1:k-1}) \]
Basic PF: Symbolic Representation

\[
\{ \mathbf{x}^{(i)}_{k-1}, w^{(i)}_{k-1} \} \sim p(\mathbf{x}^{(i)}_{k-1}|\mathbf{y}_{1:k-1})
\]
\[\Leftarrow \text{resampling}\]
\[
\{ \tilde{\mathbf{x}}^{(i)}_{k-1}, 1/N \} \sim p(\mathbf{x}^{(i)}_{k-1}|\mathbf{y}_{1:k-1})
\]
\[\Leftarrow \text{prediction}\]
\[
\{ \mathbf{x}^{(i)}_{k}, 1/N \} \sim p(\mathbf{x}^{(i)}_{k}|\mathbf{y}_{1:k-1})
\]
Basic PF: Symbolic Representation

\[ \{ \mathbf{x}^{(i)}_{k-1}, w^{(i)}_{k-1} \} \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \]
\[ \leftarrow \text{resampling} \]
\[ \{ \tilde{\mathbf{x}}_{k-1}^{(i)}, 1/N \} \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \]
\[ \leftarrow \text{prediction} \]
\[ \{ \mathbf{x}^{(i)}_k, 1/N \} \sim p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \]
\[ \leftarrow \text{measurement & update} \]

\[ \{ \mathbf{x}^{(i)}_k, w^{(i)}_k \} \sim p(\mathbf{x}_k | \mathbf{y}_{1:k}) \]
PF Methods for ASL

Summary: particle filtering methods

- **Numerical** method to solve **nonlinear** and/or **non-Gaussian** Bayesian filtering problems

- **Algorithm design choices:**
  - **Source dynamics model:** various models available
  - **Localization function:** GCCF, AEDA, SBF, etc.
  - **Likelihood function:** topic of coming slides . . .

Current Ph.D. research:

- Transformation of **observations** from traditional ASL methods into **likelihood function** for PF algorithm
- Research work in collaboration with Darren Ward, Imperial College, London
PF Likelihood for ASL

- Likelihood function $p(\mathbf{y}_k | \mathbf{x}_k)$ measures probability of receiving measurement $\mathbf{y}_k$
  - Peaks in measurement indicate likely source location
  - Occasionally no peak corresponding to true source location
  - Peak positions may have errors

- Built of the basis of a localization function, for example: GCCF
PF Likelihood for ASL

Building likelihood function from observations: two possibilities proposed

1. **Gaussian likelihood:**

![Gaussian likelihood graph](image-url)
PF Likelihood for ASL

Building likelihood function from observations: two possibilities proposed

1. **Gaussian likelihood:**

![Gaussian likelihood graph](image)

- ✓ No prior knowledge of source location: all peaks have *same* weight
PF Likelihood for ASL

Building likelihood function from observations: two possibilities proposed

1. **Gaussian likelihood:**

   ![Gaussian likelihood graph]

- ✓ No prior knowledge of source location: all peaks have **same** weight
- ✗ Entire localization function must be evaluated to find peaks
Building likelihood function from observations:

2 Pseudo-likelihood:
PF Likelihood for ASL

Building likelihood function from observations:

② **Pseudo-likelihood:**

✓ Likelihood value obtained from **pointwise evaluation** of localization function only (no need to compute entire function)
PF Likelihood for ASL

Building likelihood function from observations:

2. **Pseudo-likelihood:**

- Likelihood value obtained from pointwise evaluation of localization function only (no need to compute entire function)
- Variable weighting imposed on potential source positions
Building likelihood function from observations:

2. **Pseudo-likelihood:**

- ✓ Likelihood value obtained from **pointwise evaluation** of localization function only (no need to compute entire function)
- ✗ Variable weighting imposed on potential source positions
- ✓ Minor peaks also included in likelihood function
Proposed PF Algorithms

1. **GCC-GL**: [Vermaak and Blake, *ICASSP*, 2001]
   - Compute $P$ separate GCCF’s over entire range of delays
   - Requires $P$ 1D searches to find peaks in localization function

2. **GCC-PL**:
   - Evaluate $P$ separate GCCF’s only at delays corresponding to particles’ positions

3. **SBF-GL**:
   - Compute SBF response over entire range of source locations
   - Requires 2D search to find peaks in localization function

   - Evaluate SBF response only at locations corresponding to particles’ positions
Tracking Assessment Parameters

- **Mean square error (MSE):** average square error $\varepsilon_k$ over all signal frames, with:

  $$\varepsilon_k = \| \ell_S - \hat{\ell}_S \|^2$$

- **Frame convergence ratio (FCR):** percentage of frames for which PF correctly converges, i.e. true source location $\ell_S$ within one standard deviation $\sigma_k$ of estimate $\hat{\ell}_S$:

  $$\| \ell_S - \hat{\ell}_S \| \leq \sigma_k$$
Simulation and Experimental Setup

- Room dimensions: 2.9m × 3.8m × 2.7m (2D formulation)
- 8 sensor array: omnidirectional, constant height
Image Method Simulations

- Synthetic impulse responses
- Speech signals
- Reverberation time $RT_{60}$: 0.13s
- Additive noise: 20dB SNR

Classical GCC

Classical SBF

GCC-GL

SBF-PL
Image Method Simulations

- one reference audio sample and trajectory with variable \(RT_{60}\)
- 100 simulation runs for PF methods: SBF-PL, GCC-GL
- one simulation run for traditional methods: SBF, GCC, AEDA

![Graph showing MSE vs. RT60 for different methods: SBF-PL, GCC-GL, SBF, GCC, AEDA.](Image)
Experimental Results

- Room B129 at RSISE
- Loudspeaker as moving source
- Reverberation time $RT_{60}: 0.39s$
- Average SNR: $\sim 10dB$
- Tracking results with SBF-PL:
Experimental Comparison

Practical experiments:

- 6 trial recordings of real audio data: various speech signals and source trajectories

- 100 simulation runs for PF methods: SBF-PL, SBF-GL, GCC-PL, GCC-GL

- one simulation run for traditional methods: SBF, GCC, AEDA
Experimental Comparison: MSE

![MSE Graph]

- SBF-PL
- SBF-GL
- GCC-PL
- GCC-GL
- SBF
- GCC
- AEDA
Experimental Comparison: FCR

![Bar chart showing FCR percentages for different trials with groups SBF-PL, SBF-GL, GCC-PL, and GCC-GL.]

E. Lehmann, Ph.D. Mid-Term Review
Summary and Conclusions

- **Acoustic source localization:**
  - Traditional methods: GCCF, SBF, AEDA, etc.
  - State-space approach: Bayesian filtering
  - Sequential estimation methods (PF)
    * Localization function
    * Likelihood function

- **Sequential estimation** improves performance of traditional methods in reverberant environments:
  - Uses complete history of measurement data
  - Exploits fact that true source peak follows a dynamical model from frame to frame

- Compared performance of 4 PF methods vs. traditional approaches
Outlook – Future Research

- **Enhanced PF variants**: APF, UPF, ICondensation, hybrid bootstrap, fast weighted bootstrap, etc.
  - In particular: sequential importance sampling method, design of optimal importance function

- **Statistical room acoustics**: reverberation process and influence on PF likelihood, design of optimal likelihood function
  - Further investigation of pseudo vs. Gaussian likelihood
  - Further investigation of SBF vs. GCC methods

- **Multiple source tracking**: data association

- **Real-time ASL PF** demonstration under C/Linux
  - Design of practical tracking system: voice activity detector, particle set initialization, etc.
The End

Thanks for your attention...

ANY QUESTIONS?

Thanks to: Darren Ward
Kris Modrak

Typeset by FoilTEX using hyperref, pdfLaTEX and PPower4