Expectation-Maximization (EM) Framework for Multiple Speaker Localization and Tracking

Sharon Gannot
Joint work with Ofer Schwartz, Yuval Dorfan & Gershon Hazan
Faculty of Engineering, Bar-Ilan University, Israel

The 3rd Annual Underwater Acoustics Symposium
Tel-Aviv University, June 19th, 2014
Multiple Speaker Localization using a Network of Microphone Pairs

1. Tracking algorithm for moving sources (centralized processing).
2. Localization algorithm for static sources (distributed processing):
   - Constrained communication bandwidth.
   - Limited Computation capabilities at the nodes.

Outline

- Problem formulation & Maximum Likelihood (ML).
- Expectation-Maximization (EM).
- Recursive EM (REM).
- Distributed EM (DEM).
- Simulation results.
Received Data @ microphone pair $m$

- $z^1_m$ & $z^2_m$ - Signals @ microphone 1 & 2 of node $m$.
- $z^i_m(t, k) = \sum_{s=1}^{S} a^i_{sm}(t, k) \cdot b_s(t, k) + n^i_m(t, k)$.
- Pair-wise relative complex phase ratio (PRP): $\phi_m(t, k) \triangleq \frac{z^1_m(t, k)}{z^2_m(t, k)} \cdot \frac{|z^2_m(t, k)|}{|z^1_m(t, k)|}$.
Problem Formulation

Statistical Model

Probabilistic Model @node \(m\)

**Assumptions**

- Define a grid of positions in the region of interest: \(\mathbf{p} \in \mathcal{P}\).
- TDOA from any grid point to the microphone pair:
  \[
  \tau_m(\mathbf{p}) \triangleq ||\mathbf{p} - \mathbf{p}_m^2|| - ||\mathbf{p} - \mathbf{p}_m^1||.
  \]
- Each T-F bin is solely dominated by one speaker (\textit{W-disjoint}).

**Phase @node \(m\) as Mixture of Gaussian (MoG)**

\[
  f(\phi_m) = \prod_{t,k} \sum_{\tau_m} \psi_{\tau_m} \cdot \mathcal{N}^c(\phi_m(t,k); \tilde{\phi}_m^k(\tau_m), \sigma^2)
\]

- \(\tilde{\phi}_m^k(\mathbf{p})\) - Mean of phase differences \textbf{pre-calculated} for all grid positions \(\mathbf{p}\).
- \(\sigma^2\) - Known and constant variance of the Gaussians.
- \(\psi_{\tau_m}\) - Probability that \(\phi_m \triangleq \text{vec}_{t,k}(\{\phi_m(t,k)\})\) originates from TDOA \(\tau_m\).
Problem Formulation

Statistical Model

Probabilistic Model from Array Perspective

Definitions & Relations

- $\phi = \text{vec}_m(\phi_m)$.
- Multiple source positions give rise to the same TDOA.
- $\psi_p$ - Probability that $\phi$ originates from position $p$.

$$\psi_{\tau_m} = \int_{p' \rightarrow \tau_m} \psi_{p'} p' \approx \sum_{p' \rightarrow \tau_m} \psi_{p'}$$

Augmented Phase as Mixture of Gaussian (MoG)

$$f(\phi) = \prod_{t,k,m} \sum_{p} \psi_p \cdot \mathcal{N}^c(\phi_m(t,k); \tilde{\phi}_m^k(\tau_m(p)), \sigma^2)$$
Maximum Likelihood

Straightforward ML

Let $\psi = \text{vec}_p \{\psi_p\}$:

$$f(\phi) = \prod_{t,k,m} \sum_p \psi_p \cdot \mathcal{N}^c(\phi_m(t,k); \tilde{\phi}_m(p), \sigma^2)$$

$$\hat{\psi} = \arg\max_{\psi} \log f(\phi; \psi)$$

Goal

Estimate the most probable grid points that “explains” the received phases.
Iterative Solution using EM [Dempster et al., 1977]

Estimate-Maximize Procedure

- Solving the ML is a cumbersome task.
- Selecting a hidden data $x$ that can simplify the solution.
- E-step: $Q(\psi|\hat{\psi}^{(\ell-1)}) \triangleq E\{ \log(f(\phi, x; \psi)) | \phi; \hat{\psi}^{(\ell-1)} \}$.
- M-step: $\hat{\psi}^{(\ell)} = \operatorname{argmax}_\psi Q(\psi|\hat{\psi}^{(\ell-1)})$.

Hidden Data [Mandel et al., 2007, Schwartz and Gannot, 2014]

- $x(t, k, p) \sim I_{t,k}(p)$ (Speech sparsity assumption)
- $I_{t,k}(p)$ - Indicator that bin $(t, k)$ belongs to a (single) speaker @position $p$. 
Batch EM

**E-step**

\[
\mu^{(\ell-1)}(t, k, p) \triangleq E \left\{ x(t, k, p) | \phi(t, k); \hat{\psi}^{(\ell-1)} \right\} \\
\hat{\psi}_p^{(\ell-1)} \prod_m \mathcal{N}^c \left( \phi_m(t, k); \tilde{\phi}_m^k(p), \sigma^2 \right) \\
= \frac{\sum_p \hat{\psi}_p^{(\ell-1)} \prod_m \mathcal{N}^c \left( \phi_m(t, k); \tilde{\phi}_m^k(p), \sigma^2 \right)}{\sum_p \hat{\psi}_p^{(\ell-1)} \prod_m \mathcal{N}^c \left( \phi_m(t, k); \tilde{\phi}_m^k(p), \sigma^2 \right)}
\]

**M-step**

\[
\hat{\psi}_p^{(\ell)} = \frac{\sum_{t,k} \mu^{(\ell-1)}(t, k, p)}{T \cdot K}
\]

\(T:\ #\ of\ frames\ and\ K:\ #\ of\ frequencies.\)
Recuesive EM [Schwartz and Gannot, 2014]

Procedures

- Replace iteration index with time index.
- Execute one iteration per time index.
- Recursively estimate $Q$ [Cappé and Moulines, 2009]:
  
  $Q_R(\psi|\psi_R^{(t)}) = Q_R(\psi|\psi_R^{(t-1)}) + \gamma_t \left[ Q(\psi|\psi_R^{(t)}) - Q_R(\psi|\psi_R^{(t-1)}) \right]$.
  
  $\psi_R^{(t+1)} = \arg\max_{\psi} Q_R(\psi|\psi_R^{(t)})$.

- Maximize using Newton’s method [Titterington, 1984] (with constraints [Schwartz and Gannot, 2014]).

Solution (for both recursive procedures!))

$\psi_R^{(t+1)} = \psi_R^{(t)} + \gamma_t (\psi^{(t+1)} - \psi_R^{(t)})$
Distributed EM [Dorfan et al., 2014]

Centralized Computation

- Estimating the global hidden data depends on the availability of all PRPs in one point.
- Requires: powerful fusion center, communication bandwidth, ...

Local Hidden Data ⇔ Global Hidden Data

\[ y(t, k, \tau_m(p)) \triangleq I_{t,k,m}(\tau_m(p)) \]
\[ x(t, k, p) \equiv \prod_{m} y(t, k, \tau_m(p)) \]

Multiple positions \( p \) can induce the same \( \tau_m \).
Incremental EM \cite{Neal and Hinton, 1998} - Ring Topology

\begin{align*}
\mu^{(i)}(t, k, p) &= E\{x(t, k, p)|\phi(t, k); \psi_p^{(i-1)}\} \\
i &= (\ell - 1)M + m; \\
m &= 0, \ldots, M - 1.
\end{align*}

Becomes sparse after few iterations.

\textbf{M-Step: Global Parameter Estimation}

\[ \psi_p^{(i)} = \frac{\sum_{t,k} \mu^{(i)}(t, k, p)}{T \cdot K} \]
Increment @Node $m$

M-Step: Global Parameter Estimation (Reminder)

$$\psi^{(i)}_{\tau_m(p)} \triangleq \int_{p'\rightarrow\tau_m(p)} \psi^{(i)}_{p'} dp'$$
**Increment @Node** \( m \)

\[ \mu^{(i)}(t, k, p) \]

**M-Step**

\[ \phi_m(t, k) \]

**E-Step**

\[ \psi^{(i)}_p \]

\[ \psi^{(i)} \]

\[ \psi^{(i)}_m(t, k, \tau_m(p)) \]

\[ \frac{1}{\psi^{(i-M)}_m(t, k, \tau_m(p))} \]

\[ \mu^{(i+1)}(t, k, p) \]

**E-step: Local Hidden**

\[ \psi^{(i)}_m(t, k, \tau_m(p)) \triangleq E \left\{ y(t, k, \tau_m(p)) | \phi_m(t, k); \psi^{(i)}_p \right\} \]

\[ \psi^{(i)}_m(t, k, \tau_m(p)) \mathcal{N}^c \left( \phi_m(t, k); \tilde{\phi}^k_m(\tau_m(p)), \sigma^2 \right) \]

\[ \frac{1}{\sum_{\tau_m(p)} \psi^{(i)}_m(t, k, \tau_m(p)) \mathcal{N}^c \left( \phi_m(t, k); \tilde{\phi}^k_m(\tau_m(p)), \sigma^2 \right)} \]
Simulation results

Simulation Setup

- **Tracking**
  - 2D setup: $10 \times 10$ cm grid.
  - Trajectory: line, arc.
  - 12 nodes.
  - Inter-microphone pair: 20 cm.
  - $T_{60} = 0.7$ Sec.
  - Performance criterion: curve fit.

**Distributed Localization**

- 2D setup: $10 \times 10$ cm grid.
- Randomly located sources.
- 12 nodes.
- Inter-microphone pair: 50 cm.
- $T_{60} = 0.3$ Sec.
- Performance criteria:
  - Detection rate.
  - False Alarm (FA) rate.
  - Mean Square Error (MSE).
Simulation Results: Distributed EM

(a) Delay & Sum BF

(b) Distributed EM

<table>
<thead>
<tr>
<th># Sources</th>
<th>Detection [%]</th>
<th>FA [%]</th>
<th>MSE [cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>22</td>
<td>3.9</td>
</tr>
<tr>
<td>2</td>
<td>98</td>
<td>6.5</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Table: Results for 100 Monte-Carlo simulations
Simulation Results: Recursive EM

(c) \( \gamma = 0.1 \)

(d) \( \gamma = 0.5 \)

(e) \( \gamma = 1 \)

(f) \( \gamma = 0.1 \)

(g) \( \gamma = 0.5 \)

(h) \( \gamma = 1 \)
Summary

Recursive EM Algorithm for Tracking

1. Speech sparsity utilized to derive EM-based Localization.
2. Two versions of tracking algorithms were proposed based on [Cappé and Moulines, 2009],[Titterington, 1984].
3. A Constrained version of [Titterington, 1984] was derived.

Distributed EM Algorithm for Localization

1. No central processing unit required.
2. Decomposing the global hidden data to local hidden data is the key step in distributed algorithm derivation.
3. Detection and localization of multiple concurrent sources with minimal a priori information.
4. Only two global iterations required in our simulations.
5. No significant dependency on initial conditions observed.


