A Slide-Save Based Framework for Multi-Source DOA Extraction With Closely Spaced Sources

Jianhua Geng, Sifan Wang and Xin Lou*

ShanghaiTech University

{gengjh, wangsf, louxin}@shanghaitech.edu.cn

May 1, 2022

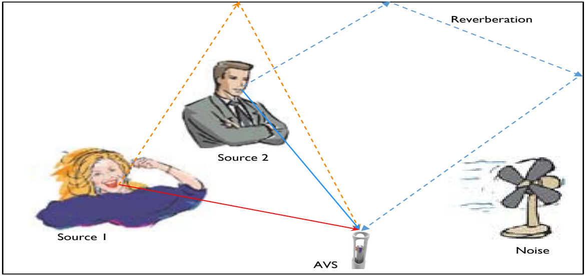
Outline

Direction of arrival (DOA) estimation

Task: estimate DOAs of multiple speech sources in an enclosed environment.

Challenges: reverberation, noise and interference between adjacent active speakers.

Sensor: a single acoustic vector sensor (AVS).



Acoustic Vector Sensor (AVS)

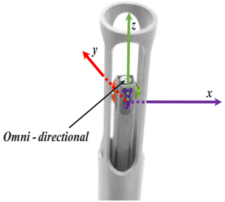
- Measuring both the sound pressure and particle velocity.
- Having frequency independent manifold vector.

Manifold vector:

$$\mathbf{a} = \begin{bmatrix} 1 \\ \cos\psi\cos\phi \\ \cos\psi\sin\phi \\ \sin\psi \end{bmatrix}$$

Measurements:

$$\mathbf{x}[n] = \begin{bmatrix} x_p[n] \\ x_{v_x}[n] \\ x_{v_y}[n] \\ x_{v_z}[n] \end{bmatrix}$$

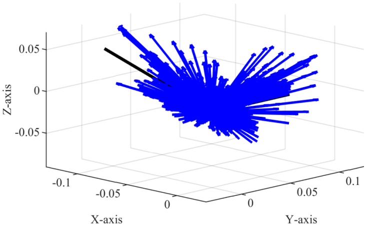


The structure of AVS

Intensity vector

The instantaneous active intensity vector in the time-frequency (TF) domain is defined as

$$\mathbf{I}(k,l) = \boldsymbol{\mathcal{R}}\left\{ x_p^{\star}(k,l)\begin{bmatrix} x_{v_x}(k,l) \\ x_{v_y}(k,l) \\ x_{v_z}(k,l) \end{bmatrix} \right\}$$

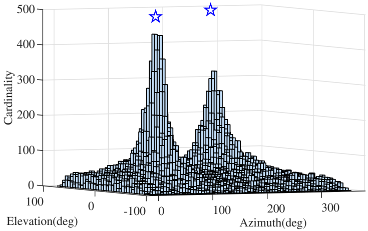


The distribution of intensity vectors

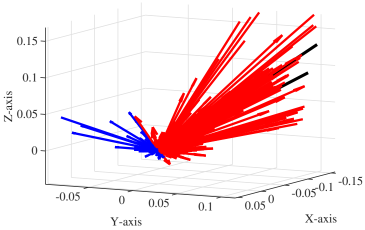
The direction of $\mathbf{I}(k,l)$ denotes an estimated DOA at the TF bin (k,l).

Multi-source DOAs extraction

- Intensity-based approaches can estimate a rough DOA at each TF bin.
- Multi-source DOAs can be extracted by histogram or clustering.



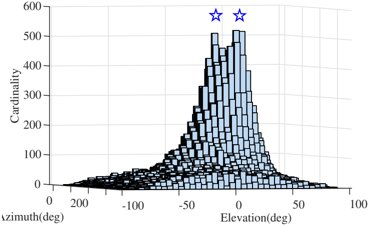
Histogram



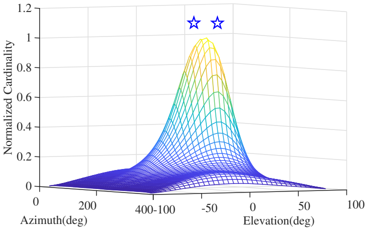
Clustering

Histogram-based DOA extraction

- Azimuth and elevation corresponding to the peaks are extracted as DOAs.
- Spatial smoothing is applied to emphasize the peaks.



Histogram

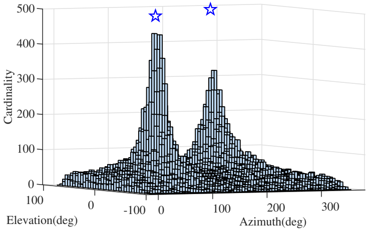


Smoothing

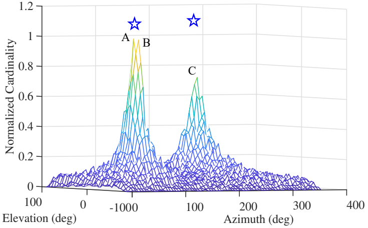
A strong smoothing may merge close peaks in adjacent sources scenarios.

Histogram-based DOA extraction

- Azimuth and elevation corresponding to the peaks are extracted as DOAs.
- Spatial smoothing is applied to emphasize the peaks.



Histogram

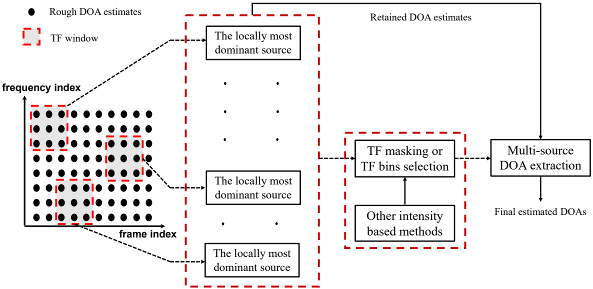


Smoothing

A weak smoothing may result in irregular peaks corresponding to one active source being identified as multiple sources.

Outline

Overview of the proposed framework



- Saving the DOA estimates corresponding to the locally most dominant source within a sliding TF window.
- Extracting final DOAs from the set of retained DOA estimates.
- Other intensity-based algorithms can also be incorporated into the proposed framework.

Overview of the proposed framework

Let $d(t,f)$ be the DOA estimate at TF bin (t,f) and d_{ci} be the core direction corresponding to the most dominant source within TF window $\mathcal{W}_i$. The set of DOA estimates close to d_{ci} is defined as

$$\mathcal{D}(d_{ci}, \mathcal{W}_i) = \{d(t,f) | \measuredangle\{d(t,f), d_{ci}\} \leq \theta, (t,f) \in \mathcal{W}_i\}, \tag{1}$$

After determining the core direction d_{ci}, the core set of critical DOA estimates corresponding to all the locally most dominant sources can be obtained by

$$\Lambda = \mathcal{D}(d_{c1}, \mathcal{W}_1) \cup \mathcal{D}(d_{c2}, \mathcal{W}_2) \cup ... \cup \mathcal{D}(d_{cM}, \mathcal{W}_M). \tag{2}$$

Here M is the total number of sliding windows.

How to determine the core direction d_{ci} ?

Implementing the proposed framework based on histogram

- **Determining the core direction** :
  For histogram-based scheme, the core direction d_{ci} is determined by

$$d_{ci} = \arg\max_{(\psi,\phi)} \mathbf{C}_{si}(\psi, \phi), \tag{3}$$

  where the smoothed histogram $\mathbf{C}_{si}(\psi, \phi)$ is constructed by the rough DOAs estimated from TF bins within TF window $\mathcal{W}_i$.

- **Obtaining the core set** :
  The core set of critical DOA estimates Λ can be obtained by substituting the core direction d_{ci} into (1) and (2).

Implementation of the framework based on histogram

- **Extracting final DOAs** :
  Let δ_m be the contribution of the mth detected peak (ψ_m, ϕ_m). It can be calculated by

$$\delta_m = \mathbf{C}_s^m(\psi, \phi) \odot \mathbf{h}_r(\psi - \psi_m, \phi - \phi_m), \tag{4}$$

  where the smoothed histogram $\mathbf{C}_s(\psi, \phi)$ is constructed based on $d \in \Lambda$, $\odot$ denotes element-wise multiplication and $\mathbf{h}_r$ is a 2D Gaussian filter. Then the contribution from $\mathbf{C}_s^m(\psi, \phi)$ is removed by

$$\mathbf{C}_s^{m+1}(\psi, \phi) \leftarrow \mathbf{C}_s^m(\psi, \phi) - \delta_m. \tag{5}$$

  The iterative procedures proceed to detect (ψ_{m+1}, ϕ_{m+1}) and calculate δ_{m+1} from $\mathbf{C}_s^{m+1}(\psi, \phi)$ until reach the number of source J. Afterwards, $\{(\psi_m, \phi_m)\}_{m=1}^J$ are saved as the final DOAs.

Implementing the proposed framework based on clustering

- **Determining the core direction** :
  The core direction is set to $d_{ci} = (\psi_k, \phi_k)$, where (ψ_k, ϕ_k) represents the direction of the center of the kth cluster $\mathcal{C}_k$. Here the index k is determined by

$$k = \arg\max_{k} \{Card(\mathcal{C}_1), ..., Card(\mathcal{C}_k), ..., Card(\mathcal{C}_J)\}, \tag{6}$$

  where $Card(\cdot)$ counts the number of elements in each cluster.
- **Obtaining the core set** :
  The core set of critical DOA estimates Λ can be obtained by substituting the core direction d_{ci} into (1) and (2).
- **Extracting final DOAs** :
  Partitioning $d \in \Lambda$ into J clusters and output the corresponding J centers as the final DOAs.

Implementing the proposed framework based on GMM

- **Determining the core direction** :
  For GMM-based scheme, the core direction d_{ci} is determined by the mean direction μ_k of the kth Gaussian component, i.e., $d_{ci} = \mu_k$. Here the index k is given by

$$k = \arg\max_{j} \left\{ \frac{w_1}{\sqrt{2\pi}\sigma_1}, ..., \frac{w_j}{\sqrt{2\pi}\sigma_j}, ..., \frac{w_J}{\sqrt{2\pi}\sigma_J} \right\}, \tag{7}$$

  where σ_j and w_j are the standard deviation and weight corresponding to the kth Gaussian component, respectively.

- **Obtaining the core set** :
  The core set of critical DOA estimates Λ can be obtained by substituting the core direction d_{ci} into (1) and (2).

Implementation of the framework based on GMM

- **Extracting final DOAs** :
  An extra Gaussian distribution is introduced to describe the outlier component. Let $\{\mu_j, \sigma_j, w_j\}_{j=1}^{J+1}$ denote the parameter set of $J+1$ components derived via an EM algorithm. The index corresponding to the outlier component can be determined by

$$o = \arg\min_j \left\{ \frac{w_1}{\sqrt{2\pi}\sigma_1}, ..., \frac{w_j}{\sqrt{2\pi}\sigma_j}, ..., \frac{w_{J+1}}{\sqrt{2\pi}\sigma_{J+1}} \right\}. \quad (8)$$

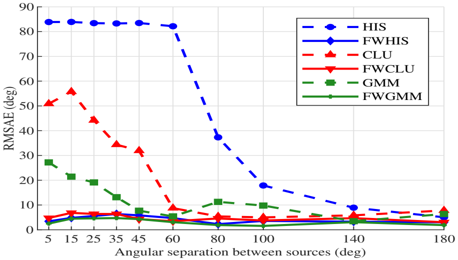
  After removing the outlier component, other mean directions $\{\mu_j\}_{j=1}^{J}$ are saved as the final estimated DOAs.

Outline

Evaluate the accuracy

- Average angular error: $e = \frac{1}{J}\sum_{i=1}^{J} \measuredangle\{d_i, (\psi_i, \phi_i)\}$.
- Root-mean-square angular error (RMSAE): $\sqrt{\mathbb{E}\{e^2\}}$.

RMSAE versus angular difference at $T_{60} = 0.35s$ and $SNR = 20dB$.

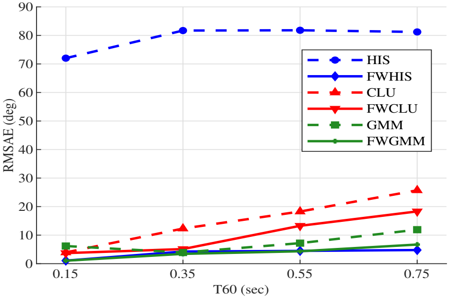


"HIS"and "CLU" denote histogram and cluster, "FWHIS" and "FWCLU" denote the corresponding frameworks.

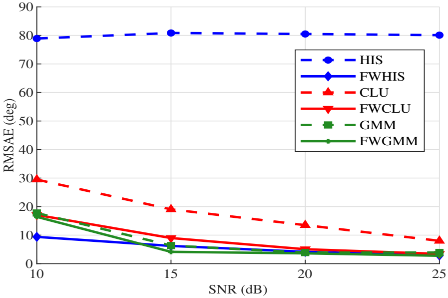
The proposed method is accuracy, even in adjacent sources scenario.

Evaluate the robustness

Two active sources are separated at an angle of $60°$.



RMSAE versus T_{60} at $SNR = 20dB$



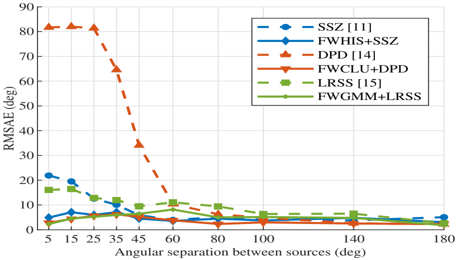
RMSAE versus SNR at $T_{60} = 0.35s$

The proposed method is robust to T_{60} and SNR.

Evaluate the accuracy

- Average angular error: $e = \frac{1}{J}\sum_{i=1}^{J} \measuredangle\{d_i, (\psi_i, \phi_i)\}$.
- Root-mean-square angular error (RMSAE): $\sqrt{\mathbb{E}\{e^2\}}$.

RMSAE versus angular difference at $T_{60} = 0.35s$ and $SNR = 20dB$.

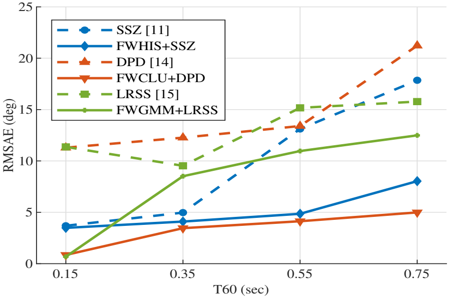


"HIS"and "CLU" denote histogram and cluster, "FWHIS" and "FWCLU" denote the corresponding frameworks.

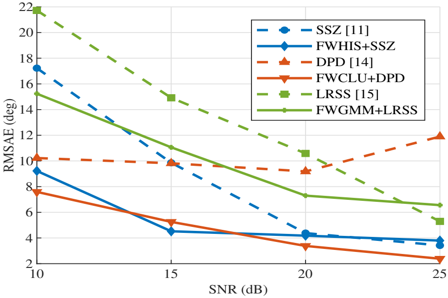
The proposed method is accuracy, even in adjacent sources scenario.

Evaluate the robustness

Two active sources are separated at an angle of $60°$.



RMSAE versus T_{60} at $SNR = 20dB$



RMSAE versus SNR at $T_{60} = 0.35s$

The proposed method is robust to T_{60} and SNR.

Thanks