SOUND SOURCE LOCALIZATION USING SVD-PHAT

Girish Chandar G

Indian Institute of Technology Gandhinagar
Electrical Engineering

Bedmutha Manas

Indian Institute of Technology Gandhinagar
Electrical Engineering

ABSTRACT

Sound Source Localization is a very important processing step for multiple sound driven applications. The given problem requires us to estimate the direction of arrival of a sound source based on a given array of 8 microphones arranged on the cube. We propose using an SVD based method by creating a spherical mesh of artificial sources to estimate the most probable direction.

Index Terms— 3D localization, DOA

1. METHOD

We create a mesh that represents a network of angles spanning the sphere. This is used to make multiple sound sources from which we estimate the most probable direction based on the SVD based approach of [1] which depends closely on [2].The dataset used for validation is taken from [3] as provided by the IEEE SP Cup team. The block diagram shown in Figure 1 describes our method. The approach followed for estimating the direction of sound is described below.

1.1. TDOA calculation and SRP-PHAT weights

- Spherical mesh of artificial sources is generated by varying azimuth θ from -180 to 180 degrees and elevation ϕ from -90 to 90 degrees with a difference of 5 degrees.
- The unit vectors from the spherical mesh is used to calculate the Time Direction of Arrival (TDOA) for large distance approximation.
- The matrix W containing the SRP-PHAT weights is generated from TDOA calculated in the previous step and its singular value decomposition is found by the formula, $W = USV^H$[4].

1.2. Processing of signal

- STFT of each microphone for N-frames is found by applying a sine window after noise reduction.
- Normalized cross-spectrum for each pair of microphones is found and stored in matrix X.

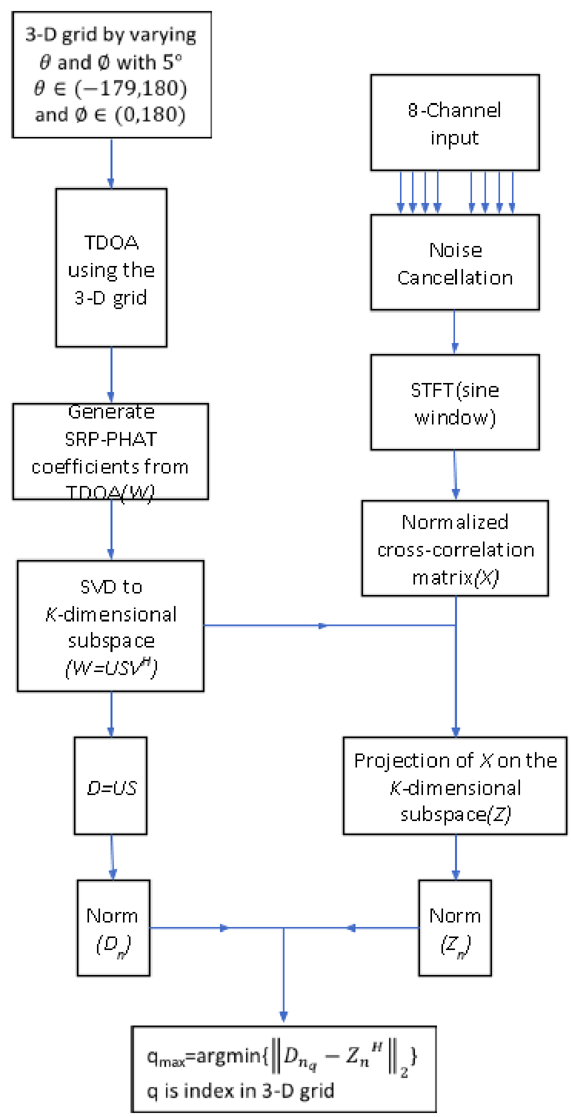

Fig. 1. Pipeline

1.3. Direction Estimation

- The vector Z represents the projection of the observations Z in the K-dimensional space. It is given by $Z = V^H X$.
- The matrix D is defined such that $D = US = [D_1 T D_2 T \dots D_Q T]$ which will give us the maximum value of the output across Z, i.e. $\Re(D_{q0} * Z^H)$ for optimum value $q0$. Here we use the interpretation from [1] to define the maximum of $\Re(D_{q0} * Z^H)$ to be the same as $1 - (\frac{1}{2})||D_q - Z^H||_2^2$ where we the new vectors are $D_q = D_q/||D_q||_2$ and the normalized Vector is $\hat{Z} = \frac{Z}{||Z||_2}$. The value $q \in Q$ for which it is maximum is the required source.
- This q is found using a simple iteration to get minimum of $||D_q - Z^H||_2^2$ in the space of $q \in [1, 2, \dots Q]$.

2. REFERENCES

[1] François Grondin and James Glass, "Svd-phat: A fast sound source localization method," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 4140–4144.

[2] Hoang Do, Harvey F Silverman, and Ying Yu, "A real-time srp-phat source location implementation using stochastic region contraction (src) on a large-aperture microphone array," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. IEEE, 2007, vol. 1, pp. I–121.

[3] Martin Strauss, Pol Mordel, Victor Miguet, and Antoine Deleforge, "Dregon: Dataset and methods for uav-embedded sound source localization," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1–8.

[4] Charles Blandin, Alexey Ozerov, and Emmanuel Vincent, "Multi-source tdoa estimation in reverberant audio using angular spectra and clustering," *Signal Processing*, vol. 92, no. 8, pp. 1950–1960, 2012.