Real-time sound source localization using Microphone Array based on Kalman Filter

Name: Yuanzhi Lou

Supervisor: Dr Shun Bai

Abstract

With the development of signal processing technology, sound source localization technology is maturing and plays an important role in more and more areas, such as fault diagnosis, robotics perception and voice pickup. However, this technology also has limitations in real-time, accuracy and anti-interference ability. In this study, an algorithm based on microphone array is introduced and applied to the sound source localization system, and different songs are used as sound sources. This algorithm mainly consists of two techniques, Direction of Arrival (DOA) and Kalman Filtering (KF). The DOA is achieved by Time Difference of Arrival (TDOA) estimation. There are three TDOA methods mentioned in this study, which are Cross Correlation Estimation, Cross Power Spectrum Estimation and Generalized Cross Correlation Estimation. Their performances are analyzed and compared. Besides, the impact of different coordinate system selection is discussed. In terms of KF, due to the nonlinearity of the DOA model, Extended Kalman Filtering (EKF) is adopted in the algorithm.

1. Introduction

1.1 Motivation and Objectives

An accurate and reliable real-time sound source localization system (SSLS) is currently a hot topic of research because of its wide application. For example, in a conference, SSLS can be used to filter out noise and obtain the location of speaker. Thus, the talker can be focused by some external equipment such as cameras or microphones. The positioning of sound mainly relies on a localization technique called direction of arrival (DOA) estimation. As there are measurement noise due to environmental factors, estimation techniques can be applied to improve the robustness of system and Kalman Filter (KF) is one of the most common methods. This study aims to build a robust and real-time SSLS for tracking the single sound source.

However, there are challenges including noise and divergence of Kalman filter, which will be detailly discussed in 1.3 Challenges.

1.2 Sound Source Localization

This section addresses the basic conception of DOA, different types of Microphone Array, Near-field and Far-field model and KF.

1.2.1 Direction of Arrival (DOA)

Direction Of Arrival (DOA) estimation, also known as Angle of Arrival (AOA) estimation, which has widely applications in communications, Electronics, radar, and other research fields. Its function is to estimate the distance and position information of the transmitting source by processing the incoming wave signal received by the microphone array. Its physical principle is based on the linear propagation principle of electromagnetic wave.

Microphone Array (MA) is a device acting as the receiver of the system, which plays an important role in DOA estimation. Each Microphone in the MA is an array element and each of them has a relatively fixed position in MA. Through the geometric relationship of different microphones and the principle of linear propagation of electromagnetic waves, equations containing the position relationship of single sound source can be established, which is the DOA model.

Different number of array elements and different shapes of MA will significantly affect the DOA model and the performance of the DOA estimation. Although in theory, this estimation requires only two array elements to determine the position of the sound source, but in general, more than two array elements are used, as the more Microphones the system has, the more accurate the results are. For different shapes of microphone array, as the geometric relationship of each microphone changes, the DOA model established through it will also change. In this study, all practical experiments were carried out based on the 6-microphone uniform circular microphone array (6-UCMA) (Figure 1-1).

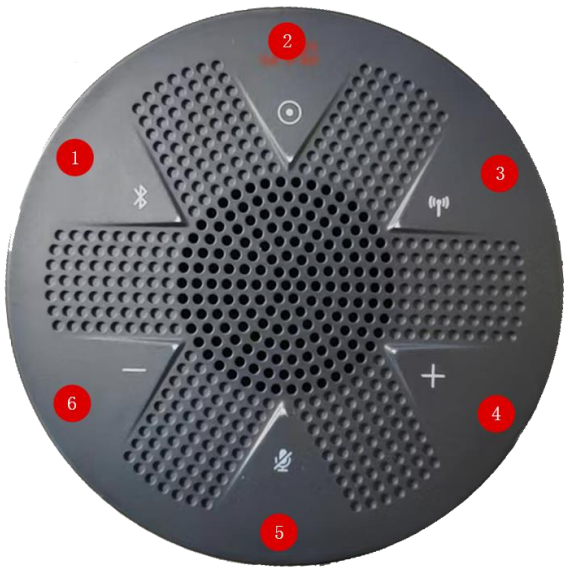

Figure 1-1 6-microphone uniform circular microphone array used in this study

1.2.2 Dimension of Microphone Array

The dimension of MA limits the position information of the sound source that can be obtained. For a one-dimension MA like the uniform linear Microphone array (ULMA), due to its spatial symmetry, it cannot distinguish whether the sound source is above or below it, left or right. If we imaginarily divide the space into four parts by two orthogonal planes that intersect at the line where ULMA is located (Figure 1-2). The motion range of the sound source can only be limited to one of the four parts, otherwise wrong results will be obtained.

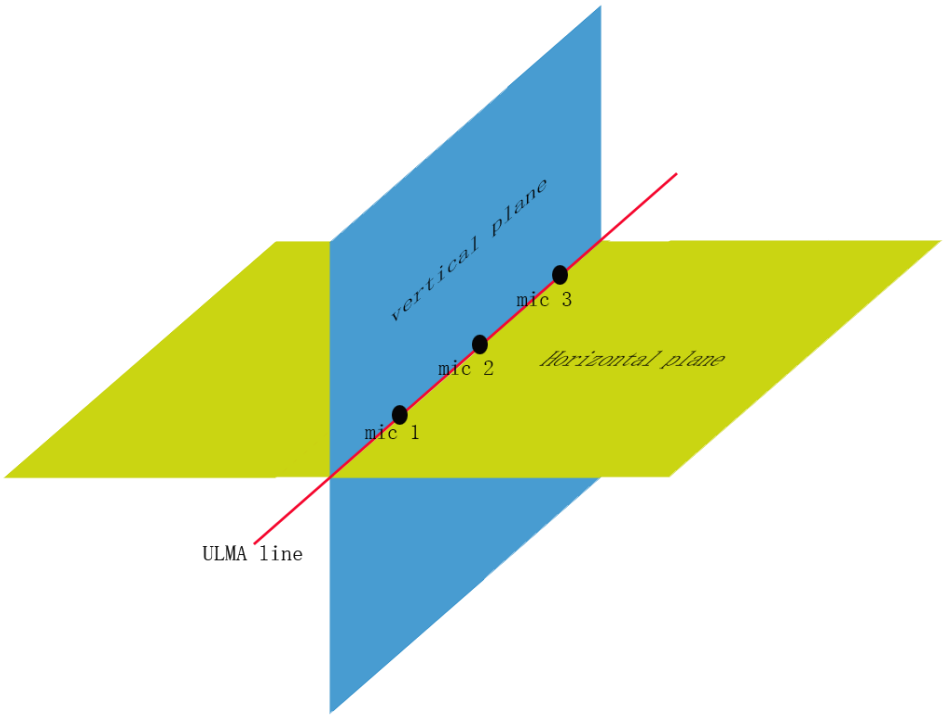

Figure 1-2 A space divided into four parts by two orthogonal planes

For a two-dimension MA like 6-UCMA used in this study, considering using the spherical Coordinate and the MA is horizontally placed, the horizontal angle of the sound source can be determined, while for the pitch angle, only the absolute value of it can be determined. In other words, the two-dimension MA cannot distinguish whether the sound source is above or below it when horizontally placed. Therefore, in this study, it is artificially specified that all sound sources are located above the 6-UCMA.

For a three-dimension MA such as the spherical microphone array, all position

information of sound source can be determined. However, while acquiring more location information, its DOA model also becomes more complex. Therefore, the moderate one, 6-UCMA, is chosen, which has a relatively considerable location information and low complexity.

1.2.3 Near-field and Far-field model

The DOA model can be divided into two types including Near-field model and Far-field model according to the distance between the microphone array and the sound source. A Far-field and Near-field model for ULMA is shown in Figure 1-3. The Near-field model is the actual model, and the Far-field model is the approximation and simplification of the Near-field model. In the Near-field model, the sound wave is regarded as spherical wave from point source and the amplitude difference between the received signals of the microphone array is considered. The Far-field model, however, treats the sound wave as plane wave and ignores the amplitude difference between the received signals. In terms of the condition for using Far-field model, there is an empirical equation (1.1) below:

$$r > \frac{2d^2}{\lambda_{min}} \tag{1.1}$$

Where r is the distance between the microphone array and the sound source; d is the spacing between microphones; λ_{min} is the minimum wavelength of sound waves.

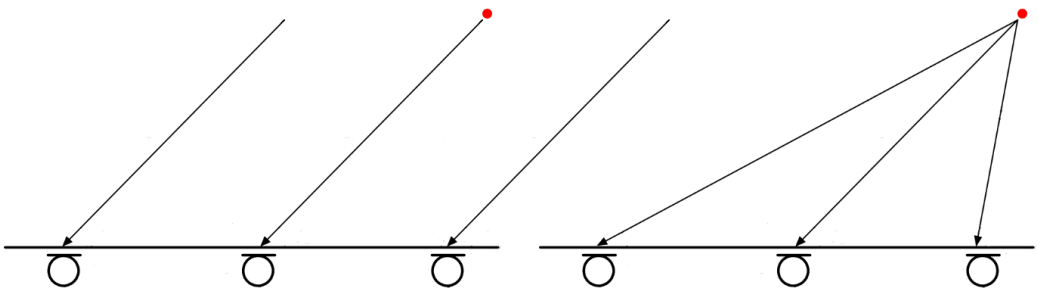

Figure 1-3 Far-field model (on the left) and Near-field model (on the right)

In this study, the Near-field model is used, but the amplitude difference between the received signals of the microphone array is ignored.

1.2.4 Kalman Filter

Kalman Filter is a widely used algorithm for tracking moving objects. It optimizes system state estimation by comparing system estimation data with observation data. However, it only applies to linear, discrete, and finite dimensional systems, but DOA model is not a linear system. Therefore, its variants, Extended Kalman Filtering (EKF), is adopted in this study. EKF approximatively changes the non-linear function into linear function using Taylor Series.

1.3 Challenges

1.3.1 Noise

Generally, localization performance is disturbed by a lot of environmental noise, which is called measurement noise. In addition, because the near-field model used in this study ignores the influence of the received signal amplitude difference between the array elements, there is system noise in the SSLS. By assuming these noises are uncorrelated white noise, the measurement noise can be dealt with by Generalized Cross Correlation (GCC) TDOA Estimation and both measurement and system noise can be filtered out by the EKF, which will be explained in 2.2.3 Generalized Cross Correlation (GCC) Estimation and 2.3 EKF procedure. However, in practice, the noise is usually correlated, which makes noise removal a challenge.

1.3.2 Divergence of Kalman filter

Ideally, for KF, with the increase of observation times, the more observation data, the more accurate the state estimation will be, and the posterior estimation error variance P

will gradually decrease and tend to a certain value. However, in practice, P sometimes increases with the increase of observation times, and finally reaches an unacceptable level. This phenomenon is called divergence. The result is meaningless when divergence occurs. The reason for divergence may be that the system model has errors. With the progress of filtering, the accumulated model errors become larger and larger, resulting in divergence.

2. System Model

2.1 DOA model with different coordinate systems

In the sound source localization process, DOA model obtained by the topological structures of microphone array can be fed into the Extended Kalman Filtering (EKF) as the Measurement Equation (ME). In this study, the case of Near-field without considering amplitude difference is applied and the 6-microphone uniform circular microphone array (6-UCMA) is used. In DOA model, the center of the microphone array is regarded as the reference point. The time delays τ between the time when the acoustic signal reaches each microphone on the circle and the center of the circle are the outputs of this model. In this section, the different DOA models will be established with the spherical coordinate system and cylindrical coordinate system.

2.1.1 Spherical coordinate system

A spherical coordinate system with the reference point of the 6-UCMA as the origin is represented in the Figure 2-1. The coordinate of the sound source is represented as (r, θ, φ), where r is the distance between the reference point and sound source, θ is the pitch angle of the sound source and φ is its horizontal angle. Note that θ should be between 0 and $\frac{\pi}{2}$, which means the sound source must be located above the 6-UCMA. This is mentioned in 1.2.2 Dimension of Microphone Array.

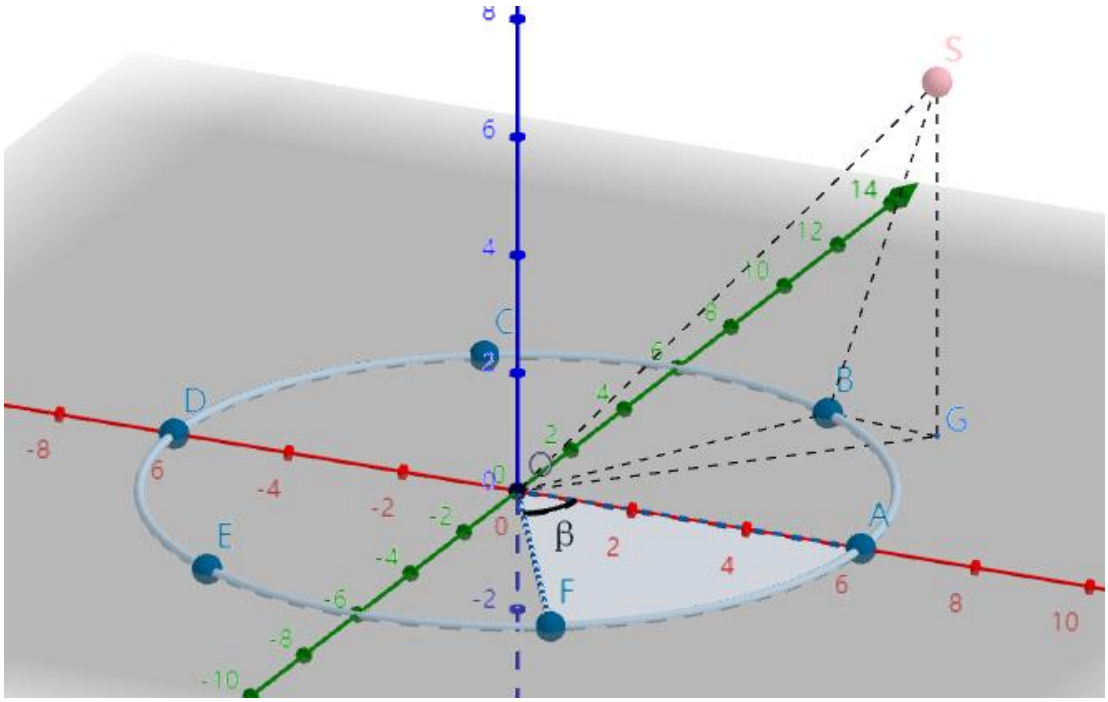

Figure 2-1 Position relationship between 6-UCMA and sound source

In Figure 2-1, S is the sound source, O is the reference point of 6-UCMA and A, B, C, D, E, F are microphones in 6-UCMA. β is the phase between two adjacent microphones, and OA has a phase of 0 degree. The time delay τ between the time when the acoustic signal reaches B and O is calculated as an example. Firstly, BS is calculated by (2.1) below:

$$\boldsymbol{BS = \sqrt{SG^2 + BG^2} = \sqrt{(r\cos\theta)^2 + [R^2 + (r\sin\theta)^2 - 2Rr\sin\theta\cos(\beta - \varphi)]}}$$
$$\boldsymbol{= \sqrt{r^2 + R^2 - 2Rr\sin\theta\cos(\beta - \varphi)}} \qquad \textbf{(2.1)}$$

Where R is the radius of 6-UCMA, which is $0.06\,m$ in this study; $\beta = \frac{\pi}{3}$.

Then, the time delay τ is calculated (2.2).

$$\boldsymbol{\tau = \frac{OS - BS}{c} = \frac{r - \sqrt{r^2 + R^2 - 2Rr\sin\theta\cos(\beta - \varphi)}}{c}} \qquad \textbf{(2.2)}$$

Where c is the speed of sound, which is $340\,m/s$.

Almost same as (2.2), the generalized time delay equation for each microphone is obtained by (2.3).

$$\boldsymbol{\tau = \frac{OS-BS}{c} = \frac{r-\sqrt{r^2+R^2-2Rr\,sin\,\theta\,cos(m\beta-\varphi)}}{c}, m = 0,1,2,3,4,5} \qquad \textbf{(2.3)}$$

2.1.2 Cylindrical coordinate system

Same as spherical coordinate system, the reference point of the 6-UCMA is set to be the origin in the cylindrical coordinate system. The coordinate of the sound source is represented as (r_c, φ_c, z_c) and z_c should be larger than zero. The r, θ, φ of spherical coordinate can be obtained through the r_c, φ_c, z_c of cylindrical coordinate (2.4).

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{\sqrt{r_c{}^2 + z_c{}^2}} \\ \boldsymbol{\theta} &= \boldsymbol{\sin^{-1}\left(\frac{r_c}{\sqrt{r_c{}^2+z_c{}^2}}\right)} \\ \boldsymbol{\varphi} &= \boldsymbol{\varphi_c} \end{aligned} \qquad \textbf{(2.4)}$$

Therefore, the generalized time delay equation for cylindrical coordinate system can be obtained from (2.3) and (2.4).

$$\boldsymbol{\tau = \frac{\sqrt{r_c{}^2+z_c{}^2}-\sqrt{r_c{}^2+z_c{}^2+R^2-2Rr_c\,cos(m\beta-\varphi_c)}}{c}, m = 0,1,2,3,4,5} \qquad \textbf{(2.5)}$$

The localization performance by using spherical coordinate system and cylindrical coordinate system will be compared and discussed in 3.2 Impact of Different Coordinate System Selection.

2.2 TDOA methods

In this study, the time delay of arrival (TDOA) Estimation is used to obtain the time

delays τ mentioned in 2.1 DOA model with different coordinate systems by analyzing the sound signals received from the reference point microphone and each microphone in the microphone array. However, it is shown in Figure 1-1 that there are only 6 microphones evenly distributed on the circle, and the center reference point microphone is not existed. In fact, a virtual center reference point microphone is simulated, and its received sound signal is simulated to be the average of six microphones in 6-UCMA.

In this section, three TDOA methods including Cross Correlation (CC) Estimation, Cross Power Spectrum (CPS) Estimation and Generalized Cross Correlation (GCC) Estimation will be explained in detail.

2.2.1 Cross Correlation (CC) Estimation

CC Estimation is a method to estimate the degree of correlation between two sequences. The essence of CC is convolution. Assuming that signal $x_2(n)$ is obtained from the signal $x_1(n)$ by a time delay of τ, the cross-correlation function $R_{x_1x_2}(m)$ of the two is (2.6).

$$\boldsymbol{R_{x_1x_2}(m) = \frac{1}{N}\sum_{n=0}^{N} x_1(n)x_2(n+m)} \qquad \textbf{(2.6)}$$

Where N is the number of sampling points.

When $R_{x_1x_2}(m)$ reaches the maximum, the time delay τ can be obtained as below.

$$\boldsymbol{\tau = \Delta T \times argmax(R_{x_1x_2}(m))} \qquad \textbf{(2.7)}$$

Where ΔT is the sampling period of microphones.

2.2.2 Cross Power Spectrum (CPS) Estimation

The principle of CPS Estimation is like CC Estimation, and the only difference is that the CC obtains cross-correlation function $R_{x_1x_2}(m)$ by convolution in the time domain while the CPS gets $R_{x_1x_2}(m)$ through Fourier transform (FT) and Inverse Fourier transform (IFT) in the frequency domain. For the CPS Estimation, the $R_{x_1x_2}(m)$ can be expressed as below,

$$\boldsymbol{R_{x_1x_2}(m) = IFT[FT[x_1(n)] \times FT^*[x_2(n)]]} \quad \textbf{(2.8)}$$

Then the time delay τ can be obtained by (2.7).

2.2.3 Generalized Cross Correlation (GCC) Estimation

GCC Estimation is an upgraded version of CPS Estimation. A weighting function $H(\omega)$ in frequency domain is added to the GCC Estimation, which is used to deal with the noise problem. Therefore, the cross-correlation function for GCC can be expressed below,

$$\boldsymbol{R_{x_1x_2}(m) = IFT[H(\omega) \times FT[x_1(n)] \times FT^*[x_2(n)]]} \quad \textbf{(2.9)}$$

The weighting function $H(\omega)$ works as a frequency domain filter to reduce the impact of noise and increase the robustness of the GCC Estimation. There are several commonly used weighting functions, which are listed in Table 2-1 below.

Table 2-1 Weighting functions of GCC

Name of Weighting Functions	$H(\omega)$
ROTH	$\frac{1}{S_{x_1x_1}(\omega)}$ or $\frac{1}{S_{x_2x_2}(\omega)}$
SCOT	$\frac{1}{\sqrt{S_{x_1x_1}(\omega)S_{x_2x_2}(\omega)}}$
PHAT	$\frac{1}{\lvert S_{x_1x_2}(\omega)\rvert}$

Where $S_{x_1x_1}(\omega)$, $S_{x_2x_2}(\omega)$ and $S_{x_1x_2}(\omega)$ are self-power spectrum of $x_1(n)$, self-power spectrum of $x_2(n)$ and cross-power spectrum of $x_1(n)$ and $x_2(n)$ respectively.

In this study, the PHAT weighting function is used in GCC Estimation. The performances of CC, CPS and GCC-PHAT will be analyzed and compared in 3.3 Comparison of CC, CPS and GCC-PHAT.

2.3 EKF procedure

In this section, the EKF procedure including State Model, Observation Model, Initialization, Prediction and Update will be explained.

Different State Model and Observation Model will be obtained for different coordinate systems. However, the principle is the same. To avoid ambiguity, the following algorithm explained is based on the spherical coordinate system.

2.3.1 State Model

The State Model for EKF in this study is linear, which can be expressed as (2.10).

$$\boldsymbol{x_k = Fx_{k-1} + w} \qquad \textbf{(2.10)}$$

Where x_k is the state in time k; F is the state transition matrix of $\boldsymbol{x_k}$; w is the system noise which satisfies the normal distribution $N(0, Q)$; Q is the covariance matrix of w.

The state x is expressed as below.

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{\theta} \\ \boldsymbol{\varphi} \\ \boldsymbol{v_r} \\ \boldsymbol{v_\theta} \\ \boldsymbol{v_\varphi} \end{bmatrix} \tag{2.11}$$

Where r, θ, φ are the coordinate of the sound source mentioned in 2.1.1 Spherical coordinate system, and v_r, v_θ, v_φ are the changing velocities of r, θ, φ respectively.

The state transition matrix F is expressed as:

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Delta T} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Delta T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Delta T} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \tag{2.12}$$

Where ΔT is the time interval between each EKF procedure. ΔT is related to the sample rate B and the number of samples taken by performing one EKF procedure N. In this study, N is $512\, samples$ and B is $16000\, samples/s$, so ΔT can be calculated by using the equation (2.13).

$$\boldsymbol{\Delta T} = \frac{\boldsymbol{N}}{\boldsymbol{B}} = \frac{\boldsymbol{512}}{\boldsymbol{16000}} = \boldsymbol{0.032\ s} \tag{2.13}$$

Assuming the system noise w affecting all six states similarly, Q can be expressed as below.

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{\sigma_Q^2} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\sigma_Q^2} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\sigma_Q^2} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\sigma_Q^2} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\sigma_Q^2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\sigma_Q^2} \end{bmatrix} \tag{2.14}$$

Where σ_Q^2 is the variance of w. Q is the hyper-parameter which affects the sensitivity of EKF.

2.3.2 Observation Model

The observation z is a 6×1 matrix which contains time delays of six microphones in 6-UCMA. It is shown below.

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{\tau_1} \\ \boldsymbol{\tau_2} \\ \boldsymbol{\tau_3} \\ \boldsymbol{\tau_4} \\ \boldsymbol{\tau_5} \\ \boldsymbol{\tau_6} \end{bmatrix} \tag{2.15}$$

In this study, the Observation Model of EKF is shown in (2.3). Obviously, it is non-linear. Therefore, the generalized Observation Model of KF (2.16) cannot be applied directly.

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x_k} + \boldsymbol{v} \tag{2.16}$$

Where H is the state transition matrix of z; $v \sim N(0, R)$ is the measurement noise; R is the covariance matrix of v.

H is the Jacobian of observation z with respect to state x, which is shown as below.

$$\boldsymbol{H} = \boldsymbol{J}(\boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \boldsymbol{v_r}, \boldsymbol{v_\theta}, \boldsymbol{v_\varphi}) = \begin{bmatrix} \frac{\partial \tau_1}{\partial r} & \frac{\partial \tau_1}{\partial \theta} & \frac{\partial \tau_1}{\partial \varphi} & \frac{\partial \tau_1}{\partial v_r} & \frac{\partial \tau_1}{\partial v_\theta} & \frac{\partial \tau_1}{\partial v_\varphi} \\ \frac{\partial \tau_2}{\partial r} & \frac{\partial \tau_2}{\partial \theta} & \frac{\partial \tau_2}{\partial \varphi} & \frac{\partial \tau_2}{\partial v_r} & \frac{\partial \tau_2}{\partial v_\theta} & \frac{\partial \tau_2}{\partial v_\varphi} \\ \frac{\partial \tau_3}{\partial r} & \frac{\partial \tau_3}{\partial \theta} & \frac{\partial \tau_3}{\partial \varphi} & \frac{\partial \tau_3}{\partial v_r} & \frac{\partial \tau_3}{\partial v_\theta} & \frac{\partial \tau_3}{\partial v_\varphi} \\ \frac{\partial \tau_4}{\partial r} & \frac{\partial \tau_4}{\partial \theta} & \frac{\partial \tau_4}{\partial \varphi} & \frac{\partial \tau_4}{\partial v_r} & \frac{\partial \tau_4}{\partial v_\theta} & \frac{\partial \tau_4}{\partial v_\varphi} \\ \frac{\partial \tau_5}{\partial r} & \frac{\partial \tau_5}{\partial \theta} & \frac{\partial \tau_5}{\partial \varphi} & \frac{\partial \tau_5}{\partial v_r} & \frac{\partial \tau_5}{\partial v_\theta} & \frac{\partial \tau_5}{\partial v_\varphi} \\ \frac{\partial \tau_6}{\partial r} & \frac{\partial \tau_6}{\partial \theta} & \frac{\partial \tau_6}{\partial \varphi} & \frac{\partial \tau_6}{\partial v_r} & \frac{\partial \tau_6}{\partial v_\theta} & \frac{\partial \tau_6}{\partial v_\varphi} \end{bmatrix} \tag{2.17}$$

Due to the non-linearity, H is no longer fixed during the EKF procedure, as the $\frac{\partial z}{\partial x}$ is no longer a constant. Therefore, H needs to be updated every EKF procedure.

Because all the microphones bear similar measurement noise v, R is defined as below,

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{\sigma_R^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma_R^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma_R^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma_R^2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma_R^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma_R^2} \end{bmatrix} \quad \textbf{(2.18)}$$

Where σ_R^2 is the variance of v. Similar with Q, R is also a hyper-parameter which affects the sensitivity of EKF. Intuitively, R is related to the GCC Estimation which reduces the measurement noise.

2.3.3 Initialization

In this study, the initialization of the state x and its covariance matrix P is necessary but inconsequential, as they will be updated during the EKF procedure. The initialization of x is as follow,

$$\boldsymbol{x_0} = \begin{bmatrix} \mathbf{1} \\ \frac{\boldsymbol{\pi}}{\mathbf{2}} \\ \frac{\boldsymbol{\pi}}{\mathbf{4}} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad \textbf{(2.19)}$$

P_0 is shown as below,

$$\boldsymbol{P_0} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \quad \textbf{(2.20)}$$

2.3.4 Prediction

The prior state estimate in time k, x_k^-, can be predicted through the State Model in (2.10).

$$\boldsymbol{x_k^- = F x_{k-1}} \quad \textbf{(2.21)}$$

Where x_{k-1} is the posteriori state estimate in time $k-1$.

Meanwhile, the prior state estimate covariance in time k, P_k^- can be predicted by the equation below.

$$\boldsymbol{P_k^- = F P_{k-1} F^T + Q} \quad \textbf{(2.22)}$$

Where P_{k-1} is the posteriori state estimate covariance in time $k-1$.

2.3.5 Update

For EKF, the first thing to do in this step is to plug in x_k^- into the Observation Model (2.3) in order to get the predicted observation z_{pred} and update H which is the Jacobian of z_{pred} with respect to x_k^- (2.17).

Then the Kalman Gain K_k in time k can be updated by P_k^- and updated H,

$$\boldsymbol{K_k = P_k^- H^T (H P_k^- H^T + R)^{-1}} \quad \textbf{(2.23)}$$

After K_k is updated, the posteriori state estimate x_k and posteriori state estimate covariance P_k are updated, which is shown in (2.24) and (2.25) respectively.

$$\boldsymbol{x_k = x_k^- + K_k(z_k - z_{pred})} \quad \textbf{(2.24)}$$

Where z_k is the observation calculated from the TDOA method which is mentioned in 2.2 TDOA methods.

$$\boldsymbol{P_k = (1 - K_k H)P_k^-} \quad \textbf{(2.25)}$$

The posteriori state estimate x_k is the output of EKF, which is continuously corrected by the EKF procedure.

3. Tests, Results and Discussions

3.1 Background Noise for Experiments

In this study, all the tests are taken placed in a conference room which has a size of $3m \times 3m$. The microphones are placed at the center of the room on the floor, and above them are a conference table. To measure the background noise, two seconds of sound is recorded at $16{,}000\ Hz$ in the absence of a sound source. The background noise in both time domain and frequency domain is shown in Figure 3-1 and Figure 3-2 respectively.

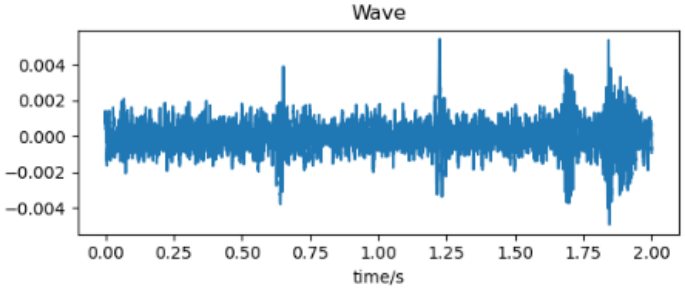

Figure 3-1 Background noise of the tested room in time domain

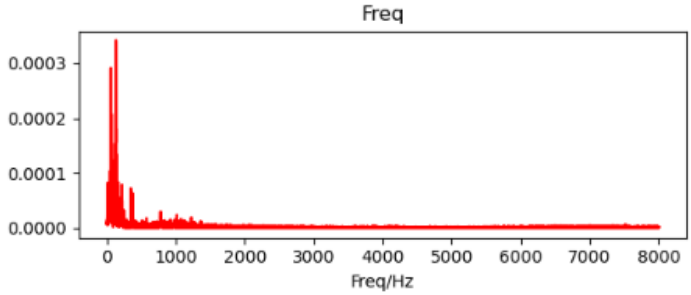

Figure 3-2 Background noise of the tested room in frequency domain

From Figure 3_1, it is observed that the background noise in the tested room has an average amplitude of about 0.002 and there are some peaks which can reach an amplitude of about 0.004. Figure 3_2 shows that the background noise in the tested room are basically consists of low frequency components in the range of $0\ Hz$ to $1400\ Hz$.

3.2 Impact of Different Coordinate System Selection

In this section, the impact of spherical and cylindrical coordinate system to the accuracy of DOA estimation is analyzed and compared.

3.2.1 Experiment Configuration

The conference room mentioned in 3.1 Background Noise for Experiments is used as the environment for this experiment. The background noise has an average amplitude of 0.002 and it consists of low frequency components. The microphone array shown in (Figure 1-1) is placed at the center of the room on the floor. A static sound source which plays the music *Without You* is placed on a floor $0.5\ m$ away from the center of 6-UCMA, and its horizontal angle is 60 degrees according to the DOA model of 6-UCMA. By applying the spherical and cylindrical coordinate system in 2.1 DOA model with different coordinate systems, the spherical and cylindrical coordinate of the sound source can be expressed as $(r, \theta, \varphi) = (0.5, 90°, 60°)$ and $(r_c, \varphi_c, z_c) = (0.5, 60°, 0)$ respectively. The DOA

estimation process is continued for 10s with the sampling frequency of $16{,}000\ Hz$ and 512 samples are taken by performing one EKF procedure. What's more, the σ_Q^2 and σ_R^2 for the EKF hyper-parameters Q and R are determined to be 0.001 and $1e-9$ respectively.

The spherical and cylindrical coordinate system are applied to estimate the horizontal angle of the sound source with respect to 6-UCMA, which can be also called the DOA angle. For both coordinate system, the simplest TDOA method, CC, and the EKF procedure are implemented.

3.2.2 Results

The estimated DOA angle for these two coordinate systems over the number of EKF procedure N is shown as below.

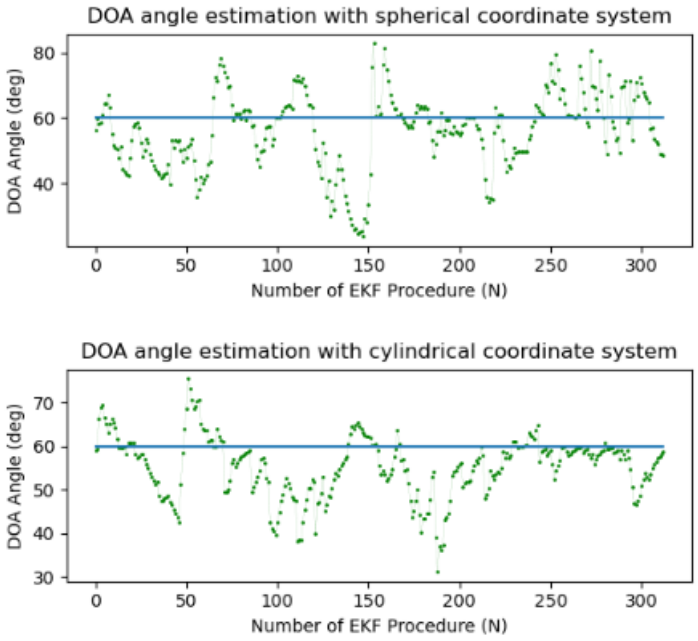

Figure 3-3 DOA angle estimation at 60° using Spherical Coordinate System (top) and Cylindrical Coordinate System (bottom)

From Figure 3-3, it is shown that for both Spherical and Cylindrical Coordinate System, the estimated DOA angle fluctuates up and down at 60°, and the fluctuation ranges for both are similar, which is about 40° - 80°. This fluctuation range may be caused by the background noise. The relationship between accuracy rate of DOA angle estimation for these two Coordinate System and error tolerance is shown in Figure 3-4.

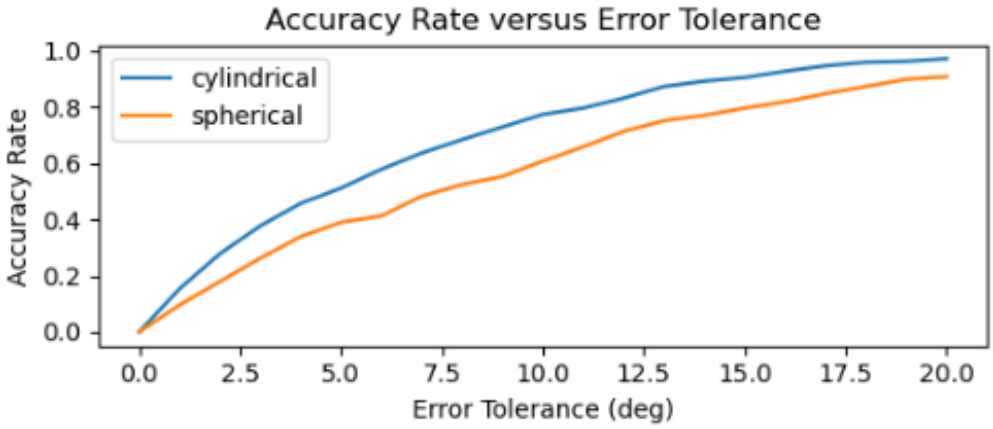

Figure 3-4 Accuracy Rate versus Error Tolerance by using Spherical Coordinate System (orange) and Cylindrical Coordinate System (blue)

From Figure 3-4, it is obvious that for all degree of error tolerance, the one using Cylindrical Coordinate System has a higher accuracy rate than the Spherical Coordinate System.