Estimation of Trajectory and Location for Mobile Sound Source

Mehmet Cem Catalbas, Merve Yildirim, Arif Gulten, Hasan Kurum, and Simon Dobrišek
1 University of Firat Faculty of Electrical and Electronics Engineering, Elazig, Turkey
2 University of Ljubljana Faculty of Electrical Engineering Ljubljana, Slovenia

Abstract—In this paper, we present an approach to estimate mobile sound source trajectory. An artificially created sound source signal is used in this work. The main aim of this paper is to estimate the mobile object trajectory via sound processing methods. The performance of generalized cross correlation techniques is compared with that of noise reduction filters for the success of trajectory estimation. The azimuth angle between the sound source and receiver is calculated during the whole movement. The parameter of Interaural Time Difference (ITD) is utilized for determining azimuth angle. The success of estimated delay is compared with different types of Generalized Cross Correlation (GCC) algorithms. In this study, an approach for sound localization and trajectory estimation on 2D space is proposed. Besides, different types of pre-filter method are tried for removing the noise and signal smoothing of recorded sound signals. Some basic parameters of sound localization process are also explained. Moreover, the calculation error of average azimuth angle is compared with different GCC and pre-filtered methods. To conclude, it is observed that estimation of location and trajectory information of a mobile object from a stereo sound recording is realized successfully.

Keywords—Sound processing; sound source localization; azimuth angle estimation; generalized cross-correlation; interaural time difference; interaural level difference

I. INTRODUCTION

The general definition of sound is pressure change versus time which is examined by microphone or microphone arrays. Also, these differences are used for sound recognition, localization, and classification, etc. The sound localization is implemented by using this information. Generally, Sound localization process is inspired from the human hearing system [1-2]. There are several areas of research for sound processing [3]. One of the most important areas is sound localization [4]. In parallel to technological improvement, sound localization is used in very different areas [5-6] such as security systems, cell phone, voice command application, and conference system, etc. The term of sound localization is determining the coordinates of sound source in 2D or 3D Space [7-8]. The 2D coordinate information is adequate for some of the basic applications [9-10]. Besides, the information of 3D coordinates is necessary for complex sound processing applications. In this study, we obtain 2D coordinates information for moving person via stereo sound recording [11]. A sound dataset which is created artificially is used for this work [12-14].

In this paper, Section II explains sound localization. In Section III, 2D sound localization via microphone pair is described. Estimations of sound delay and sound source trajectory are carried out in Section IV and V, respectively. Finally, conclusions are given in the last section.

II. SOUND LOCALIZATION

The term of sound localization is determining of the coordinates of sound sources by using some of the signal processing methods. Generally, the users must have least three microphones or sensors for localization.

We propose a different approach for sound localization in this paper. In contrast to the other studies, the locations of sound sources are taken by only two microphones. Furthermore, two common terms for sound localization process which are ITD and Interaural Level Differences (ILD) are used [15-16]. The parameters of ITD and ILD are time and amplitude difference between signals received by microphones or sensors [17]. The parameter of ITD is used generally for determining azimuth angle [18-19]. ITD and ILD are illustrated in Fig. 1 and Fig. 2, respectively.

![Interaural Time Difference (ITD)](image1)

![Interaural Level Difference (ILD)](image2)

Fig. 1. Interaural Time Difference (ITD)

Fig. 2. Interaural Level Difference (ILD)

The parameter of ITD is much more reliable compared to that of ILD. The ITD is easily calculated by cross-correlation methods. The result of ITD is sample difference between received signals. The azimuth angle is also calculated by this parameter and sampling frequency is taken as $f_s = 44100 \, Hz$ in this study. The calculation of azimuth angle is shown in Fig.
3. The speed of sound is very sensitive to environmental conditions such as especially ambient humidity and temperature [20].

The calculation of the speed of sound depended on ambient temperature and azimuth angle is shown in (1) and (2), respectively [21].

\[
V_{\text{sound}} = 20.05\sqrt{273.15 + C_t} \quad (1)
\]

\[
\alpha = \arcsin\left(\frac{V_{\text{sound}}T_d}{d}\right) \quad (2)
\]

where \(C_t\), \(d\), \(V_{\text{sound}}\), and \(T_d\) are the ambient temperature (°C), the distance between sound receivers (meter), speed of sound (m/sec), and time differences between received signals (sec), respectively.

The sound localization process is given in Fig. 4. The angles and distances between points are taken as (\(\beta=\gamma=\delta=60^\circ\)) and (\(d_x = d_y = d_z\)), respectively. \(V_m\) is movement speed of the sound source.

III. 2D SOUND LOCALIZATION VIA MICROPHONE PAIR

In this section, 2D localization via microphone pair is explained in detail. The sound localization problem is solved by three or more microphones and the researchers use Time Difference of Arrival (TDOA) between microphone pairs [22-23]. The advantage of this approach is that the user can determine \(x\) and \(y\) coordinates via only one microphone pair. However, the users have to obtain not only the parameter of ITD but also ILD parameter of sound signals [24]. In this approach, the inverse square law of energy propagation is used for sound localization. The received sound signals are defined as \(s_1\) and \(s_2\) [21]. \(n_i(t)\) represents the white noise on environment.

The energy received from microphone is

\[
E_i = \int_0^w x_i^2(t) dt \quad i=1,2,...,n
\]

\[
E_i = \frac{1}{d^2} \int_0^w s^2(t) dt + \int_0^w n_i^2(t) dt
\]

where \(w\), \(E_i\), and \(d\) are the screen size of the observation, the energy of sound signal \(x_i\), and the distance between sound source and receiver, respectively.

The relationship between the energies and distances can be obtained by using these equations

\[
E_1d_1^2 = E_2d_2^2 + \eta \quad (4)
\]

where \(\eta = \int_0^w [d_1^2n_1^2(t) - d_2^2n_2^2(t)] dt\) is error term and \((x_n, y_n)\) and \((x_n, y_n)\) are the coordinates of an \(i\)-th microphone and the sound source, respectively.

\[
d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2} \quad (5)
\]

Since the signals acquired by each microphone can be assumed as delayed replicas of the source signal, localizing a source is to estimate time delay estimation (TDE) between the signals of two microphones. \(\tau_1\) is time delay and \(d_i\) is ignored in TDE model. When time delay is measured, hyperbolic equation can be satisfied by Cartesian coordinate.

\[
\sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2} - \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2} = c\tau_{12} \quad (6)
\]

where \(c\) and \(\tau_{12}\) are sound speed and TDOA of mic1 and mic2, respectively. By using (4) and (5), (7) can be obtained by

\[
(x_1 - x_s)^2 + (y_1 - y_s)^2 = \frac{1}{\gamma^2}(x_2 - x_s)^2 + (y_2 - y_s)^2 + \frac{\eta}{n_i} \quad (7)
\]

where \(\gamma = \sqrt{E_1/E_2}\). The noise term \(\eta/E_1\) can be ignored in a high Signal to Noise Ratio (SNR) environment. According to this, (7) can be inserted into (6) and following equations are obtained as

\[
(x_s - x_1)^2 + (y_s - y_1)^2 = \left(\frac{\tau_{12}}{1 - \gamma}\right)^2 \quad (8)
\]

\[
(x_s - x_2)^2 + (y_s - y_2)^2 = \left(\frac{\tau_{12}\gamma}{1 - \gamma}\right)^2 \quad (9)
\]

The exact source position can be found by composing (8) and (9). The existence of solution is defined by

\[
\frac{\tau_{12}}{1 - \gamma} |1 - \gamma| \leq d \leq \left(\frac{\tau_{12}}{1 - \gamma}(1 + \gamma)\right) \quad (10)
\]

where \(d\) is the distance between two circle centers. When \(\tau_{12}\) is greater than zero, it means that source reaches mic1 later than mic2 and \(\gamma\) is also less than one. Thus, \(\tau_{12}\) and \((1 - \gamma)\) are positive or negative. In case of \(E_i \neq E_2\), it can be determined by

\[
c|\tau_{12}| \leq d \leq c|\tau_{12}|(1 + \gamma) \quad (11)
\]

In the case of \(E_i = E_2\), it means that there will not be an intersection to determine source position. To solve this problem, (8) and (9) are used again as follows.
\[(x_s - x_1)^2 + (y_s - y_1)^2 = \left(\frac{ct_{12}}{1 - \gamma}\right)^2 = r_1^2 \tag{12}\]
\[(x_s - x_2)^2 + (y_s - y_2)^2 = \left(\frac{ct_{12}}{1 - \gamma}\right)^2 = r_2^2 \tag{13}\]

According to this, the equations are obtained as follows
\[x_i x_2 + y_i y_2 = \frac{1}{2} (K_i - r_i^2 + R_s^2); \quad i = 1, 2 \tag{14}\]

where
\[K_i = x_i^2 + y_i^2; \quad (i = 1, 2)\]
\[R_s = \sqrt{x_s^2 + y_s^2} \tag{15}\]

Matrix form is defined by following equations
\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix} = \frac{1}{2} \begin{bmatrix}
  R_s^2 & 1 \\
  R_s^2 & 1
\end{bmatrix} \begin{bmatrix}
  K_1 - r_1^2 \\
  K_2 - r_2^2
\end{bmatrix} \tag{16}
\]
\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix} = \frac{1}{2} \begin{bmatrix}
  R_s^2 & 1 \\
  R_s^2 & 1
\end{bmatrix} \begin{bmatrix}
  K_1 - r_1^2 \\
  K_2 - r_2^2
\end{bmatrix} \tag{17}
\]
\[
a = \frac{[a_2] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}}{[a_1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}} \tag{18}
\]
\[
b = \frac{[b_1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}}{[b_2] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}} \tag{19}
\]
\[
x = \frac{[x_1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}}{[y_1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}} \tag{20}
\]

where \(R_s\) is the source coordinate.

Equation (21) is obtained by inserting (20) into (15)
\[(a_i^2 + a_s^2) R_s^4 + 2(a_i b_1 + a_s b_2) R_s^2 + b_1^2 + b_2^2 = 0 \quad (21)\]

The solution to \(R_s^2\) is
\[R_s^2 = \frac{c_a c_b c_c}{c_a} \tag{22}\]

where
\[c_a = a_1^2 + a_s^2 \]
\[c_b = 1 - a_s b_1 + a_s b_2 \]
\[c_c = \sqrt{(1 - a_s b_1 + a_s b_2)^2 - (a_1^2 + a_s^2)(b_1^2 + b_2^2)} \quad (23)\]

Positive root means the square of the distance from the source to an origin. The final source coordinate can be found by using \(R_s\) calculation in (20).

IV. SOUND DELAY ESTIMATION

In this section, trajectory estimation of sound sources and determination of the coordinates are explained. The dual-microphone source location method is used in 2-D space. This method is realized by database created artificially. The ambient temperature is selected as 20°C for this work. Stereo sound recording is shown in Fig. 5.

The sliding window techniques are used for energy and delay estimation of sound channels. The width of sliding window and step size are selected as 1024 sample and 10, respectively. Delays between sound channels are calculated by GCC algorithms [25]. Two different types of GCC algorithms which are basic GCC and Generalized Cross Correlation with Phase Transform (GCC-PHAT) are used in this paper [26-27]. The GCC of the signals recorded by two microphones is given by
\[R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_{ij}(w) X_1(w) X_2(w) e^{j\omega\tau} dw \tag{24}\]

where \(X_1(w)\) and \(X_2(w)\) are the signals recorded by one and two microphones in Fourier domain and \(\omega\) is angular frequency. The weighting function \(\psi_{ij}(w)\) is designed to optimize the given performance criteria. Many different types of weighting function are used in the literature. The most common one is Phase Transform (PHAT) which is defined in (25).
\[\psi_{ij}(w) = \frac{1}{|X_1(w) X_2^*(w)|} \tag{25}\]

Three different types of signal smoothing filter such as the moving average filter are utilized for increasing the success of delay estimation in this study. The moving average filter is shown in (26). The parameters including \(x[i]\), \(y[i]\), and \(M\) refer to the input signal, the output signal, and the number of points used in the moving average filter, respectively [28]. The weighted moving average filter is also given in (27).
\[y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i + j] \tag{26}\]
\[y[i] = \frac{1}{M} \sum_{j=0}^{M-1} w_j x[i + j] \quad \sum_{i=1}^{M} w_i = 1 \tag{27}\]

V. ESTIMATION OF SOUND SOURCE TRAJECTORY

![Fig. 5. Sound signals](image)

In this section, trajectory estimation of sound sources and determination of the coordinates are explained. The dual-microphone source location method is used in 2-D space. This method is realized by database created artificially. The ambient temperature is selected as 20°C for this work. Stereo sound recording is shown in Fig. 5.

The sliding window techniques are used for energy and delay estimation of sound channels. The width of sliding window and step size are selected as 1024 sample and 10, respectively. Delays between sound channels are calculated by GCC algorithms [25]. Two different types of GCC algorithms which are basic GCC and Generalized Cross Correlation with Phase Transform (GCC-PHAT) are used in this paper [26-27]. The GCC of the signals recorded by two microphones is given by
\[R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_{ij}(w) X_1(w) X_2(w) e^{j\omega\tau} dw \tag{24}\]

where \(X_1(w)\) and \(X_2(w)\) are the signals recorded by one and two microphones in Fourier domain and \(\omega\) is angular frequency. The weighting function \(\psi_{ij}(w)\) is designed to optimize the given performance criteria. Many different types of weighting function are used in the literature. The most common one is Phase Transform (PHAT) which is defined in (25).
\[\psi_{ij}(w) = \frac{1}{|X_1(w) X_2^*(w)|} \tag{25}\]

Three different types of signal smoothing filter such as the moving average filter are utilized for increasing the success of delay estimation in this study. The moving average filter is shown in (26). The parameters including \(x[i]\), \(y[i]\), and \(M\) refer to the input signal, the output signal, and the number of points used in the moving average filter, respectively [28]. The weighted moving average filter is also given in (27).
\[y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i + j] \tag{26}\]
\[y[i] = \frac{1}{M} \sum_{j=0}^{M-1} w_j x[i + j] \quad \sum_{i=1}^{M} w_i = 1 \tag{27}\]
sound channels is realized by sliding windows. The estimated delays are converted to the angle by using (2). First order polynomial is fitted to determine the angle of linear movement of sound source versus time. The calculated delays and azimuth angle of the sound signal are shown in Fig. 6 and 7, respectively.

The first order polynomial interpolation result of azimuth angle is illustrated in Fig. 8.

By using dual-microphone source location method, the estimation and curve fitting change of X and Y coordinate values versus time are shown in Fig. 9 and 10, respectively.

The 2D coordinate map of sound source motion is shown in Fig. 11. As seen in the figure, the estimated motion of sound source is from A to B by using 2D sound source localization approach.

The mean calculation error of azimuth angle versus different algorithms is shown in Table I. The order of filters is selected as N=5 for the whole process.

As shown in Table I, the best combinations of GCC method and pre-filter are generic GCC and Median filter. The mean difference between the estimated and values of the real coordinates is 8.544 cm.

VI. CONCLUSION

In this paper, we propose a different approach for sound source localization which is 2D sound localization microphone pair in contrast to common approaches. Some basic parameters of sound signal processing are explained and they give an idea about sound signal localization, azimuth angle determination and exact localization of sound sources. This method is related not only delay between sound channels but also the energy ratio between recording sound signals. The 2D sound localization using microphone pair approach is implemented for sound signal created artificially. The result of calculated
azimuth angle is compared with different GCC and noise removing filters methods. Mobile sound source trajectory depended on the time is also calculated. Furthermore, curve fitting method is applied for estimation of the sound source motion. The energy and delay parameter are also used for calculating the coordinates of the sound source. The optimal combination of methods is examined for sound source localization. The coordinates of mobile sound source are calculated very precisely and successfully depending on the time. It is seen that the azimuth angle for linear motion of sound sources is calculated in a minor error. In future work, real time sound localization application with similar approach will be performed by embedded systems such as Raspberry Pi.

REFERENCES


