

Estimation of Trajectory and Location for Mobile Sound Source

¹Mehmet Cem Catalbas, ¹Merve Yildirim, ¹Arif Gulden, ¹Hasan Kurum, and ²Simon Dobrišek

¹University of Firat Faculty of Electrical and Electronics Engineering, Elazig, Turkey

²University 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.

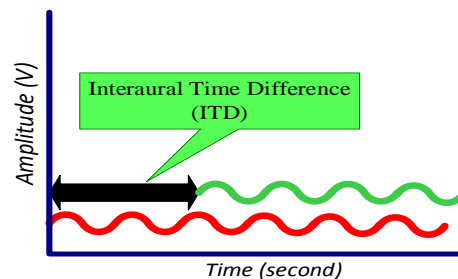


Fig. 1. Interaural Time Difference (ITD)

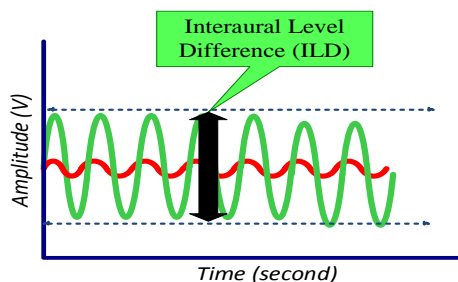


Fig. 2. Interaural Level Difference (ILD)

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].

$$V_{sound} = 20.05\sqrt{273.15 + C_t} \quad (1)$$

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

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

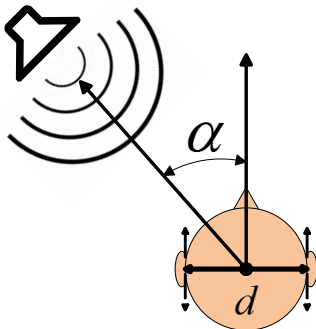


Fig. 3. Azimuth angle determination

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.

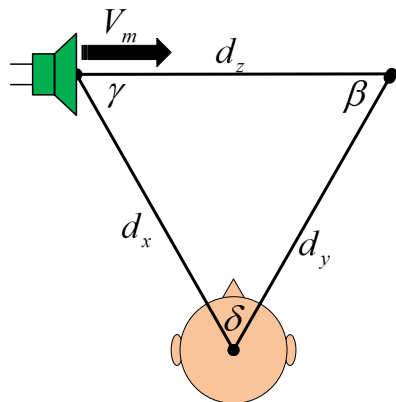


Fig. 4. Illustration of test environment

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 to the white noise on environment.

The energy received from microphone is

$$E_i = \int_0^w x_i^2(t)dt \quad i=1,2,\dots,n \quad (3)$$

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

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

$$E_1 d_1^2 = E_2 d_2^2 + \eta \quad (4)$$

where $\eta = \int_0^w [d_1^2 n_1^2(t) - d_2^2 n_2^2(t)]dt$ is error term and (x_i, y_i) and (x_s, y_s) 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. τ_i 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 τ_{12} are sound speed and TDOA of mic1 and mic2, respectively. By using (4) and (5), (7) can be obtained by

$$(x_1 - x_s)^2 + (y_1 - y_s)^2 = \frac{1}{\gamma^2} [(x_2 - x_s)^2 + (y_2 - y_s)^2] + \frac{\eta}{E_1} \quad (7)$$

where $\gamma = \sqrt{E_1/E_2}$. The noise term η/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

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

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

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

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

where d is the distance between two circle centers. When τ_{12} is greater than zero, it means that source reaches mic1 later than mic2 and γ is also less than one. Thus, τ_{12} and $(1 - \gamma)$ are positive or negative. In case of $E_1 \neq E_2$, it can be determined by

$$c|\tau_{12}| \leq d \leq c|\tau_{12}| \frac{1+\gamma}{|1-\gamma|} \quad (11)$$

In the case of $E_1 = E_2$, 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

$$(x_s - x_1)^2 + (y_s - y_1)^2 = \left(\frac{c\tau_{12}}{1-\gamma}\right)^2 = \tau_1^2 \quad (12)$$

$$(x_s - x_2)^2 + (y_s - y_2)^2 = \left(\frac{c\tau_{12}\gamma}{1-\gamma}\right)^2 = \tau_2^2 \quad (13)$$

According to this, the equations are obtained as follows

$$x_i x_s + y_i y_s = \frac{1}{2}(K_i - r_i^2 + R_s^2); \quad i = 1, 2 \quad (14)$$

where

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

Matrix form is defined by following equations

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{1}{2} \left\{ R_s^2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} K_1 - r_1^2 \\ K_2 - r_2^2 \end{bmatrix} \right\} \quad (16)$$

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1} \frac{1}{2} \left\{ R_s^2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} K_1 - r_1^2 \\ K_2 - r_2^2 \end{bmatrix} \right\} \quad (17)$$

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (18)$$

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1} \begin{bmatrix} K_1 - r_1^2 \\ K_2 - r_2^2 \end{bmatrix} \quad (19)$$

$$x = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} b_1 & a_1 R_s^2 \\ b_2 & a_2 R_s^2 \end{bmatrix} \quad (20)$$

where R_s is the source coordinate.

Equation (21) is obtained by inserting (20) into (15)

$$(a_1^2 + a_2^2)R_s^4 + 2(a_1 b_1 + a_2 b_2)R_s^2 + b_1^2 + b_2^2 = 0 \quad (21)$$

The solution to R_s^2 is

$$R_s^2 = \frac{c_b \pm c_c}{c_a} \quad (22)$$

where

$$c_a = a_1^2 + a_2^2 \\ c_b = 1 - a_1 b_1 + a_2 b_2 \\ c_c = \sqrt{(1 - a_1 b_1 + a_2 b_2)^2 - (a_1^2 + a_2^2)(b_1^2 + b_2^2)} \quad (23)$$

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

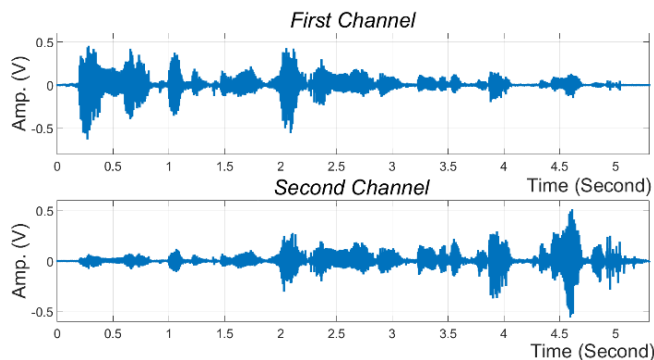


Fig. 5. Sound signals

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 $20\text{ }^\circ\text{C}$ 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

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \psi_{ij}(w) X_1(w) X_2(w) e^{jw\tau} dw \quad (24)$$

where $X_1(w)$ and $X_2(w)$ are the signals recorded by one and two microphones in Fourier domain and w 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).

$$\psi_{ij}(w) = \frac{1}{|X_1(w)X_2^*(w)|} = \frac{1}{|X_1(w)||X_2(w)|} \quad (25)$$

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).

$$y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i+j] \quad (26)$$

$$y[i] = \frac{1}{M} \sum_{j=0}^{M-1} w_j x[i+j] \quad \sum_{i=1}^M w_i = 1 \quad (27)$$

V. ESTIMATION OF SOUND SOURCE TRAJECTORY

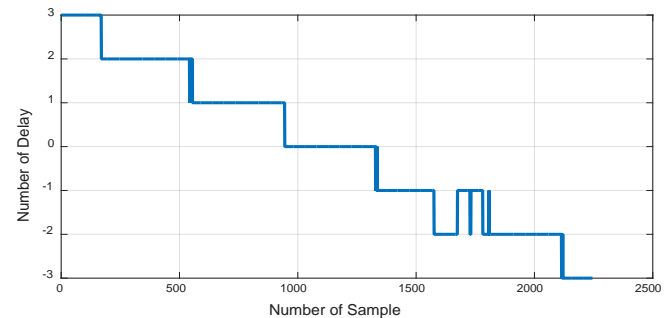


Fig. 6. Calculated delays

The estimation of sound trajectory is mentioned in this section. The implementation of our approach consists of several steps. Firstly, noise removal filters are applied to raw sound recordings. After that, the estimation of delays between

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.

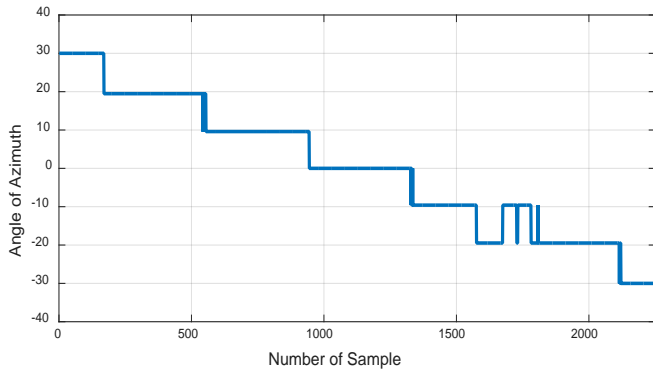


Fig. 7. Azimuth angle of movement

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

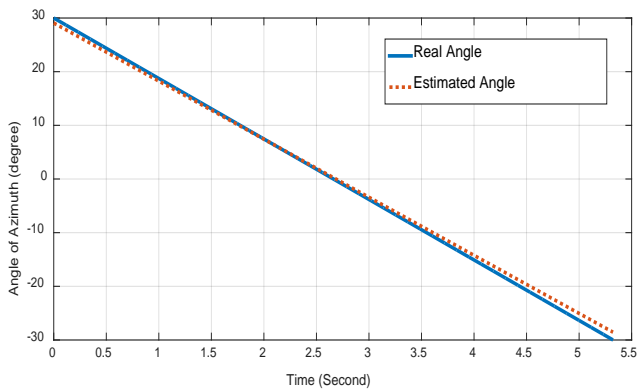


Fig. 8. Comparison of real and estimated azimuth angle

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.

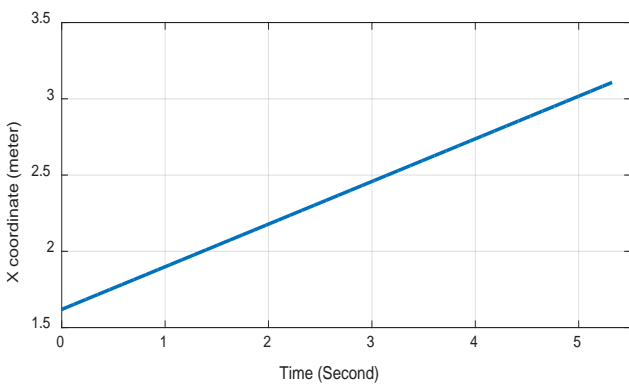


Fig. 9. X coordinate change versus time

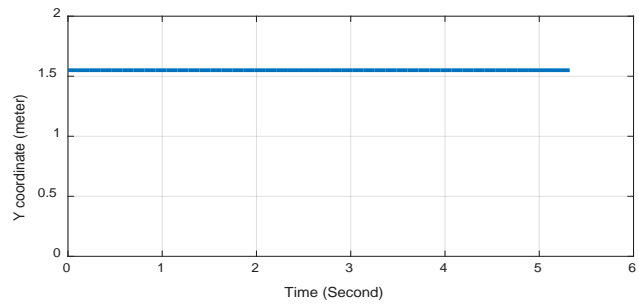


Fig. 10. Y coordinate change versus time

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.

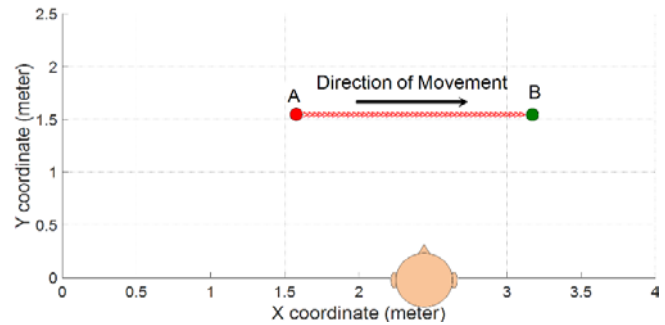


Fig. 11. Direction of motion on coordinate map

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.

TABLE I. CALCULATION OF MEAN ERROR OF THE AZIMUTH ANGLE

GCC Method/Filter	Moving Average	Moving Weighted Window	Median Filter
Unfiltered Correlation	0.053	0.0372	0.0257
PHAT Filter	0.1763	0.7568	0.0641

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

- [1] L. A. Jeffress, "A place theory of sound localization," *Journal of comparative and physiological psychology*, vol. 41, no. 1, 1948.
- [2] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," *Annual Review of Psychology*, vol. 42, no. 1, pp. 135-159, 1991.
- [3] S. K. Mitra and K. Yonghong, "Digital signal processing: a computer-based approach," *New York: McGraw-Hill*, vol. 2, 2006.
- [4] S. L. Gay and J. Benesty, "Acoustic signal processing for telecommunication," *Springer Science & Business Media*, vol. 551, 2012.
- [5] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, and N. Sugie, "A model-based sound localization system and its application to robot navigation," *Robotics and Autonomous Systems*, vol. 27, no. 4, pp. 199-209, 1999.
- [6] F. Asano, M. Goto, K. Itou, and H. Asoh, "Real-time sound source localization and separation system and its application to automatic speech recognition," *INTERSPEECH*, pp. 1013-1016, 2001.
- [7] Y. Liang, Z. Cui, S. Zhao, K. Rupnow, Y. Zhang, D. L. Jones, and D. Chen, "Real-time implementation and performance optimization of 3D sound localization on GPUs," *IEEE Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 832-835, 2012.
- [8] S. Zhao, S. Ahmed, Y. Liang, K. Rupnow, D. Chen, and D. L. Jones, "A real-time 3D sound localization system with miniature microphone array for virtual reality," *7th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2012.
- [9] L. Battista, E. Schena, G. Schiavone, S. A. Sciuto, and S. Silvestri, "Calibration and uncertainty evaluation using monte carlo method of a simple 2D sound localization system," *IEEE Sensors Journal*, vol. 13, no. 9, pp. 3312-3318, 2013.
- [10] A. Deleforge and H. Radu "2D sound-source localization on the binaural manifold," *IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1-6, 2012.
- [11] S. Aoki, T. Masayoshi, and T. Norihisa, "Sound localization of stereo reproduction with parametric loudspeakers," *Applied Acoustics*, vol. 73, no. 12, pp. 1289-1295, 2012.
- [12] B. C. J. Moore, "An introduction to the psychology of hearing," *Academic Press*, 5th edition, 2003.
- [13] N. Birbaumer and R. F. Schmidt, "Biologische psychologie," *Springer*, 4th edition, 1999.
- [14] Frithjof Hummes, Hendrik Buschmeier. 2005. Stereo Perception and Sound Localisation. [ONLINE] Available at: <http://buschmeier.org/bh/study/soundperception/>. [Accessed 22 June 2016].
- [15] R. H. Gifford, D. W. Grantham, S. W. Sheffield, T. J. Davis, R. Dwyer, and M. F. Dorman, "Localization and interaural time difference (ITD) thresholds for cochlear implant recipients with preserved acoustic hearing in the implanted ear," *Hearing research* 312, pp. 28-37, 2014.
- [16] M. Kyweriga, W. Stewart, C. Cahill, and M. Wehr, "Synaptic mechanisms underlying interaural level difference selectivity in rat auditory cortex," *Journal of Neurophysiology*, vol. 112, no. 10, pp. 2561-2571, 2014.
- [17] X. R. Xiong, F. Liang, H. Li, L. Mesik, K. Zhang, D. B. Polley, and L. I. Zhang, "Interaural level difference-dependent gain control and synaptic scaling underlying binaural computation," *Neuron* vol. 79, no. 4, pp. 738-753, 2013.
- [18] V. Benichoux, M. Stimberg, B. Fontaine, and R. Brette, "A unifying theory of ITD-based sound azimuth localization at the behavioral and neural levels," *BMC Neuroscience*, 2013.
- [19] C. P. Brown and O. R. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476-488, 1998.
- [20] N. Bilaniuk and SK W. George, "Speed of sound in pure water as a function of temperature," *The Journal of the Acoustical Society of America*, vol. 93, no. 3, pp. 1609-1612, 1993.
- [21] A. Pourmohammad and M. A. Seyed, "N-dimensional N-microphone sound source localization," *EURASIP Journal on Audio, Speech, and Music Processing*, pp. 1-19, 2013.
- [22] F. Gustafsson and G. Fredrik, "Positioning using time-difference of arrival measurements," *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'03*, vol. 6, 2003.
- [23] G. Mellen, M. Pachter, and J. Raquet, "Closed-form solution for determining emitter location using time difference of arrival measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 3 pp. 1056-1058, 2003.
- [24] W. Cui, Z. Cao, J. Wei, "DUAL-Microphone source location method in 2-D space," *Toulouse in Proceedings of the ICASSP*, pp. 845-848, 2006.
- [25] GC Carter, "Time delay estimation for passive sonar signal processing," *IEEE T Acoust. S.*, pp. 462-470, 1981.
- [26] J. Velasco, M. J. Taghizadeh, A. Asaei, H. Bourlard, C. J. Martín-Arguedas, J. Macias-Guarasa, and D. Pizarro, "Novel GCC-PHAT model in diffuse sound field for microphone array pairwise distance based calibration," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2669- 2673, 2015.
- [27] B. Kwon, P. Youngjin, and P. Youn-sik. "Analysis of the GCC-PHAT technique for multiple sources," *IEEE International Conference on Control Automation and Systems (ICCAS)*, 2010.
- [28] S. W. Smith, "The scientist and engineer's guide to digital signal processing," 1997.