

RESEARCH

Open Access

N -dimensional N -microphone sound source localization

Ali Parsayan* and Seyed Mohammad Ahadi

Abstract

This paper investigates real-time N -dimensional wideband sound source localization in outdoor (far-field) and low-degree reverberation cases, using a simple N -microphone arrangement. Outdoor sound source localization in different climates needs highly sensitive and high-performance microphones, which are very expensive. Reduction of the microphone count is our goal. Time delay estimation (TDE)-based methods are common for N -dimensional wideband sound source localization in outdoor cases using at least $N + 1$ microphones. These methods need numerical analysis to solve closed-form non-linear equations leading to large computational overheads and a good initial guess to avoid local minima. Combined TDE and intensity level difference or interaural level difference (ILD) methods can reduce microphone counts to two for indoor two-dimensional cases. However, ILD-based methods need only one dominant source for accurate localization. Also, using a linear array, two mirror points are produced simultaneously (half-plane localization). We apply this method to outdoor cases and propose a novel approach for N -dimensional entire-space outdoor far-field and low reverberation localization of a dominant wideband sound source using TDE, ILD, and head-related transfer function (HRTF) simultaneously and only N microphones. Our proposed TDE-ILD-HRTF method tries to solve the mentioned problems using source counting, noise reduction using spectral subtraction, and HRTF. A special reflector is designed to avoid mirror points and source counting used to make sure that only one dominant source is active in the localization area. The simple microphone arrangement used leads to linearization of the non-linear closed-form equations as well as no need for initial guess. Experimental results indicate that our implemented method features less than 0.2 degree error for angle of arrival and less than 10% error for three-dimensional location finding as well as less than 150-ms processing time for localization of a typical wideband sound source such as a flying object (helicopter).

Keywords: Sound source localization; ITD; TDE; TDOA; PHAT; ILD; HRTF

1. Introduction

Source localization has been one of the fundamental problems in sonar [1], radar [2], teleconferencing or videoconferencing [3], mobile phone location [4], navigation and global positioning systems (GPS) [5], localization of earthquake epicenters and underground explosions [6], microphone arrays [7], robots [8], microseismic events in mines [9], sensor networks [10,11], tactile interaction in novel tangible human-computer interfaces [12], speaker tracking [13], surveillance [14], and sound source tracking [15]. Our goal is real-time sound source localization in outdoor environments, which necessitates a few points to be considered. For localizing such sound sources, a far-field assumption is usual. Furthermore, our experiments

confirm that placing localization system in suitable higher heights often reduces the reverberation degree, especially for flying objects. Also, many such sound source signals are wideband signals. Moreover, outdoor high-accuracy sound source localization in different climates needs highly sensitive and high-performance microphones which are very expensive. Therefore, reduction of the number of microphones is very important, which in turn leads to reduced localization accuracies using conventional methods. Here, we intend to introduce a real-time accurate wideband sound source localization system in low degree reverberation far-field outdoor cases using fewer microphones.

The structure of this paper is as follows. After a literature review, we explain HRTF-, ILD-, and TDE-based methods and discuss TDE-based phase transform (PHAT). In Section 4, we explain sound source angle of arrival and location calculations using ILD and PHAT. Section 5

* Correspondence: pourmohammad@aut.ac.ir
Electrical Engineering Department, Amirkabir University of Technology,
424 Hafez Ave., Tehran 15914, Iran

covers the introduction of TDE-ILD-based method to two-dimensional (2D) half-plane sound source localization using only two microphones. Section 6 includes simulation of this method for 2D cases, where according to simulation results and due to the use of ILD, we introduce source counting. In Section 7, we propose, and in Section 8, we implement our TDE-ILD-HRTF-based method for 2D whole-plane and three-dimensional (3D) entire-space sound source localization. Section 9 includes the implementations. Finally conclusions will be made in Section 10.

2. Literature review

Passive sound source localization methods, in general, can be divided into direction of arrival (DOA) [16], time difference of arrival (TDOA) or TDE or interaural time difference (ITD)- [17-20], ILD- [21-24], and HRTF-based methods [25-30]. DOA-based beamforming and sub-space methods typically need a large number of microphones for estimation of narrowband source locations in far-field cases and wideband source locations in near-field cases. Also, they have higher processing needs in comparison to other methods. Many localization methods for near-field cases have been proposed in the literature, such as maximum likelihood (ML), covariance approximation, multiple signal estimation (MUSIC), and estimation of signal parameters via rotational invariance techniques (ESPRIT) [16]. However, these methods are not applicable to the localization of wideband signal sources in far-field cases with small number of microphones. On the other hand, ILD-based methods are mostly applicable to the case of a single dominant sound source (high signal-to-noise ratio (SNR)) [21-24]. TDE-based methods with high sampling rates are commonly used for 2D and 3D high-accuracy wideband near-field and far-field sound source localization. In the case of ILD or TDE-based methods, minimum number of microphones required is three for 2D positioning and four for the 3D case [17-20,31-39]. Finally, HRTF-based methods are applicable only to the case of calculating arrival angle in azimuth or elevation [30].

TDE- and ILD-based outdoor far-field accurate wideband sound source localization in different climates needs highly sensitive and high-performance microphones which are very expensive. In the last decade, some papers were published which introduce 2D sound source localization methods using just two microphones in indoor cases using TDE- and ILD-based methods simultaneously. However, it is noticeable that using ILD-based methods requires only one dominant source to be active, and it is known that, by using a linear array in the proposed TDE-ILD-based method, two mirror points will be produced simultaneously (half-plane localization) [40]. In this paper, we apply this method in outdoor (low-degree reverberation) cases for a dominant sound source.

We also propose a novel method to have 2D whole-plane (without producing two mirror points) and 3D entire-space dominant sound source localization using TDE-, ILD-, and HRTF-based methods simultaneously (TDE-ILD-HRTF method). Based on the proposed method, a special reflector for the implemented simple microphone arrangement is designed, and source counting method is used to find that only one dominant sound source is active in the localization area.

In TDE and ILD localization approaches, calculations are carried out in two stages: estimation of time delay or intensity level differences and location calculation. Correlation is most widely used for time delay estimation [17-20,31-39]. The most important issue in this approach is high-accuracy time delay estimation between microphone pairs. Meanwhile, the most important issue in ILD-based approach is high accuracy level difference measurement between microphone pairs [21-24]. Also, numerous results were published in the last decades for the second stage, i.e., location calculation. Equation complexities and large processing times are the important obstacles faced at this stage. In this paper, we propose a simple microphone arrangement that solves both these problems simultaneously.

In the above-mentioned first stage, the classic methods of source localization from time delay estimates by detecting radio waves were Loran and Decca [31]. However, generalized cross-correlation (GCC) using a ML estimator, proposed by Knapp and Carter [32], is the most widely used method for TDE. Later, a number of techniques were proposed to improve GCC in the presence of noise [3,33-36]. As GCC is based on an ideal signal propagation model, it is believed to have a fundamental weakness of inability to cope well with reverberant environments. Some improvements were obtained by cepstral prefiltering by Stephenne and Champagne [37]. Even though more sophisticated techniques exist, they tend to be computationally expensive and are thus not well suited for real-time applications. Later, PHAT was proposed by Omologo and Svaizer [38]. More recently, a new PHAT-based method was proposed for high-accuracy robust speaker localization, known as steered response pattern-phase or power-phase transform (SRP-PHAT) [39]. Its disadvantage is higher processing time in comparison to PHAT, as it requires a search of a large number of candidate locations. According to the fact that in the application of this research, the number of candidate locations is much higher due to direction and distance estimation; this disadvantage does not allow us to use it in real-time applications.

In the last decade, according to the fact that the received signal energy is inversely proportional to the squared distance between the source and the receiving sensor, there has been some interest in using the

received signal level at different sensors for source localization. Due to the spatial separation of the sensors, the source signal will arrive at the sensors with different levels so that the level differences can be utilized for source localization. Sheng and Hu [41] followed by Blatt and Hero [42] have proposed different algorithms for locating sources using a sensor network based on energy measurements. Birchfield and Gangishetty [22] applied ILD to sound source localization. While these works used only ILD-based methods to locate a source, Cui et al. [40] tried 2D sound source localization by a pair of microphones using TDE- and ILD-based methods simultaneously. When the source signal is captured at the sensors, both time delay and signal level information are used for source localization. This technique is applicable for 2D localization with two sensors only. Also, due to the use of a linear array, it generates two mirror points simultaneously (half-plane localization). Ho and Sun [24] addressed a more common scenario of 3D localization using more than four sensors to improve the source location accuracy.

Human hearing system allows finding sound sources direction of arrival in 3D with just two ears. Pinnae, shoulders, and head diffract the incoming sound waves [43]. These propagation effects collectively are termed the HRTF. Batteau reported that the external ears play an important role in estimating the elevation angle of arrival [44]. Roffler and Butler [45], Oldfield and Parker [46], and Hofman et al. [47] have tried to find experimental evidence for this claim by using a Plexiglas headband to flatten the pinna against the head [45]. Based on HRTF measuring, Brown and Duda have made an extensive experimental study and provided empirical formulas for the multipath delays produced by pinna [48]. Although more sophisticated HRTF models have been proposed [49], the Brown-Duda model has the advantage that it provides an analytical relationship between the multipath delays and the azimuth and elevation angles. Recently, Sen and Nehorai considered the Brown-Duda HRTF model as an example to model the frequency-dependent head-shadow effects and the multipath delays close to the sensors for analyzing a 3D direction finding system with only two sensors inspired by the human auditory system [43]. However, they did not consider white noise gain error or spatial aliasing error in their model. They computed the asymptotic frequency domain Cramer-Rao bound (CRB) on the error of the 3D direction estimate for zero-mean wide-sense stationary Gaussian source signals. It should be noted that HRTF-based works are just able to estimate the azimuth and elevation angles [50]. In the last decades, some papers were published which tried to apply HRTF along with TDE for azimuth and elevation angle of arrival estimation [30]. According to this ability, we apply HRTF in

our TDE-ILD-based localization system for solving the ambiguity in the generation of two mirror location points. We named it TDE-ILD-HRTF method.

Given a set of TDEs and ILDs from a small set of microphones using PHAT- and ILD-based methods, respectively, the second stage of a two-stage algorithm determines the best point for the source location. The measurement equations are non-linear. The most straightforward way is to perform an exhaustive search in the solution space. However, this is computationally expensive and inefficient. If the sensor array is known to be linear, the position measurement equations are simplified. Carter focused on a simple beamforming technique [1]. However, it requires a search in the range and bearing space. Also, beamforming methods need many more microphones for high-accuracy source localization. The linearization solution based on Taylor-series expansion by Foy [51] involves iterative processing, typically incurs high computational complexity, and for convergence, requires a tolerable initial estimate of the position. Hahn proposed an approach [20] that assumes a distant source. Abel and Smith proposed an explicit solution that can achieve the Cramer-Rao Lower Bound (CRLB) in the small error region [52]. The situation is more complex when sensors are distributed arbitrarily. In this case, emitter position is determined from the intersection of a set of hyperbolic curves defined by the TDOA estimates using non-Euclidean geometry [53,54]. Finding the solution is not easy as the equations are non-linear. Schmidt has proposed a formulation [18] in which the source location is found as the focus of a conic passing through three sensors. This method can be extended to an optimal closed-form localization technique [55]. Delosme [56] proposed a gradient method for search in a localization procedure leading to computation of optimal source locations from noisy TDOA's. Fang [57] gave an exact solution when the number of TDOA measurements is equal to the number of unknowns (coordinates of transmitter). This solution, however, cannot make use of extra measurements, available when there are extra sensors, to improve position accuracy. The more general situation with extra measurements was considered by Friedlander [58], Schau and Robinson [59], and Smith and Abel [55]. These methods are not optimum in the least-squares sense and perform worse in comparison to the Taylor-series method.

Although closed-form solutions have been developed, their estimators are not optimum. The divide and conquer (DAC) method [60] from Abel can achieve optimum performance, but it requires that the Fisher information is sufficiently large. To obtain a precise position estimate at reasonable noise levels, the Taylor-series method [51] is commonly employed. It is an iterative method that starts with an initial guess and improves the estimate at each step by determining the local linear least-squares (LS)

solution. An initial guess close to the true solution is needed to avoid local minima. Selection of such a starting point is not simple in practice. Moreover, convergence of the iterative process is not assured. It also suffers from convergence problem and large LS computational burden as the method is iterative. Within the last few years, some papers were published on improving LS and closed-form methods [16,23,61-65]. Based on closed-form hyperbolic-intersection method, we will explain our proposed method, which using a simple arrangement of two microphones for 2D cases and three microphones for 3D cases, can simplify non-linear equations of this method to have a linear equation. Although there have been attempts to linearize closed-form non-linear equations through algebraic means, such as [7,16,56,63], our proposed method with simple pure geometrical linearization needs less microphones and features accurate localization and less processing time.

3. Basic methods

3.1. HRTF

Human beings are able to locate sound sources in three dimensions in range and direction, using only two ears. The location of a source is estimated by comparing difference cues received at both ears, among which are time differences of arrival and intensity differences. The sound source specifications are modified before entering the ear canal. The head-related impulse response (HRIR) is the name given to the impulse response relating the source location and the ear location. Therefore, convolution of an arbitrary sound source with the HRIR leads to what would have been received at the ear canal. The HRTF is the Fourier transform of HRIR and thus represents the filtering properties due to diffractions and reflections at the head, pinna, and torso. Hence, HRTF can be computed through comparing the original and modified signals [25-29,43-50]. Consider $X_c(k)$ being the Fourier transform of sound source signal at one ear (modified signal at either left or right ear) and $Y_c(k)$ to be that of the original signal. HRTF of that source can be found as [30]:

$$H_c(k) = \frac{Y_c(k)}{X_c(k)} \quad (1)$$

In detail:

$$|H_c(k)| = \frac{|Y_c(k)|}{|X_c(k)|} \quad (2)$$

$$\arg H_c(k) = \arg Y_c(k) - \arg X_c(k) \quad (3)$$

$$H_c(k) = |H_c(k)| e^{j \arg H_c(k)}. \quad (4)$$

In fact, $H_c(k)$ contains all the direction-dependent and direction-independent components. Therefore, in order to have pure HRTF, the direction-independent elements have to be removed from $H_c(k)$. If $C_c(k)$ is the known CTF (common transfer function), then DTF (directional transfer function), $D_c(k)$, can be found as:

$$D_c(k) = \frac{Y_c(k)}{C_c(k)X_c(k)}. \quad (5)$$

3.2. ILD-based localization

We consider two microphones for localizing a sound source. Signal $s(t)$ propagates through a generic free space with noise and no (or low degree of) reverberation. According to the so-called inverse-square law, the signal received by the two microphones can be modeled as [21-24,40-42]:

$$s_1(t) = \frac{s(t-T_1)}{d_1} + n_1(t) \quad (6)$$

and

$$s_2(t) = \frac{s(t-T_2)}{d_2} + n_2(t), \quad (7)$$

where d_1 and d_2 are the distances and T_1 and T_2 are time delays from source to the first and second microphones, respectively. Also $n_1(t)$ and $n_2(t)$ are additive white Gaussian noises. The relative time shift between the signals is important for TDE but can be ignored in ILD. Therefore, if we find the delay between the two signals and shift the delayed signal in respect to the other one, the signal received by the two microphones can be modeled as:

$$s_1(t) = \frac{s(t)}{d_1} + n_1(t) \quad (8)$$

and

$$s_2(t) = \frac{s(t)}{d_2} + n_2(t). \quad (9)$$

Now, we assume that the sound source is audible and in a fixed location. Also, it is available during the time interval $[0, W]$ where W is the window size. The energy received by the two microphones can be obtained by integrating the square of the signal over this time interval [21-24,40-42]:

$$E_1 = \int_0^w s_1^2(t) dt = \frac{1}{d_1^2} \int_0^w s^2(t) dt + \int_0^w n_1^2(t) dt \quad (10)$$

$$E_2 = \int_0^w s_2^2(t) dt = \frac{1}{d_2^2} \int_0^w s^2(t) dt + \int_0^w n_2^2(t) dt. \quad (11)$$

According to (10) and (11), the received energy decreases in relation to the inverse of the square of the distance to the source. These equations lead us to a simple relationship between the energies and distances:

$$E_1 \cdot d_1^2 = E_2 \cdot d_2^2 + \eta, \quad (12)$$

where $\eta = \int_0^w [d_1^2 n_1^2(t) + d_2^2 n_2^2(t)] dt$ is the error term. If (x_1, y_1) is the coordinates of the first microphone, (x_2, y_2) is the coordinates of the second microphone and (x_s, y_s) is the coordinates of the sound source with respect to the origin located at array center, then:

$$d_1 = \sqrt{(x_1 - x_s)^2 + (y_1 - y_s)^2} \quad (13)$$

$$d_2 = \sqrt{(x_2 - x_s)^2 + (y_2 - y_s)^2}. \quad (14)$$

Now using (12), (13), and (14), we can localize the sound source (Section 4.1.).

3.3. TDE-based localization

Correlation-based methods are the most widely used time delay estimation approaches. These methods use the following simple reasoning for the estimation of time delay. The autocorrelation function of $s_1(t)$ can be written in time domain as [17-20,31-39]:

$$R_{s_1 s_1}(\tau) = \int_{-\infty}^{+\infty} s_1(t) \cdot s_1(t - \tau) dt. \quad (15)$$

Dualities between time and frequency domains for autocorrelation function of $s_1(t)$ with the Fourier transform $S_1(f)$, results in frequency domain presentation as:

$$R_{s_1 s_1}(\tau) = \int_{-\infty}^{+\infty} S_1(f) \cdot S_1^*(f) e^{j2\pi f \tau} df. \quad (16)$$

According to (15) and (16), if the time delay τ is zero, this function's value will be maximized and will be equal to the energy of $s_1(t)$. The cross-correlation of two signals $s_1(t)$ and $s_2(t)$ is defined as:

$$R_{s_1 s_2}(\tau) = \int_{-\infty}^{+\infty} S_1(f) \cdot S_2^*(f) e^{j2\pi f \tau} df. \quad (17)$$

If $s_2(t)$ is considered to be the delayed version of $s_1(t)$, this function features a peak at the point equal to the time delay. This delay can be expressed as:

$$\tau_{12} = \operatorname{argmax}_{\tau} R_{s_1 s_2}(\tau). \quad (18)$$

In an overall view, the time delay estimation methods are as follows [17-20,31-39]:

- Correlation-Based Methods: Cross-Correlation (CC), ML, PHAT, Average Square Difference Function (ASDF)
- Adaptive Filter-Based Methods: Sync Filter, LMS.

Advantages of PHAT are accurate delay estimation in the case of wideband and quasi-periodic/periodic signals, good performance in noisy and reflective environments, sharper spectrum due to the use of better weighting function and higher recognition rate. Therefore, PHAT is used in cases where signals are detected using arrays of microphones and additive environmental and reflective noises are observed. In such cases, the signal delays cannot be accurately found using typical correlation-based methods as the correlation peaks cannot be precisely extracted. PHAT is a cross-correlation-based method used for finding the time delay between the signals. In PHAT, similar to ML, weighting functions are used along with the correlation function, i.e.,

$$\phi_{\text{PHAT}}(f) = \frac{1}{|G_{s_1 s_2}(f)|}, \quad (19)$$

where $G_{s_1 s_2}(f)$ is the cross-correlation-based power spectrum. The overall function used in PHAT for the estimation of delay between two signals is defined as:

$$R_{s_1 s_2}(\tau) = \int_{-\infty}^{+\infty} \phi_{\text{PHAT}}(f) \cdot G_{s_1 s_2}(f) e^{j2\pi f \tau} df \quad (20)$$

$$D_{\text{PHAT}}(f) = \operatorname{argmax}_{\tau} R_{s_1 s_2}(\tau), \quad (21)$$

where D_{PHAT} is the delay calculated using PHAT. $G_{s_1 s_2}(f)$ is found as:

$$G_{s_1 s_2}(f) = \int_{-\infty}^{+\infty} r_{s_1 s_2}(\tau) e^{j2\pi f \tau} d\tau \quad (22)$$

where

$$r_{s_1 s_2}(\tau) = E[s_1(t) s_2(t + \tau)]. \quad (23)$$

4. ILD- and PHAT-based angle of arrival and location calculation methods

4.1. Using ILD method

Assuming two microphones are on x -axis and have a distance of R ($R = D/2$) from origin (Figure 1), we can rewrite (13) and (14) as:

$$d_1 = \sqrt{(R - x_s)^2 + y_s^2} \quad (24)$$

$$d_2 = \sqrt{(-R - x_s)^2 + y_s^2}. \quad (25)$$

Therefore, we can rewrite (12) as:

$$\left(\frac{E_1}{E_2}\right) \cdot d_1^2 = d_2^2 + (\eta/E_2). \quad (26)$$

Assuming $m = \frac{E_1}{E_2}$ and $n = \eta/E_2$, (26) is written as:

$$m[(R - x_s)^2 + y_s^2] = [(-R - x_s) + y_s^2] + n. \quad (27)$$

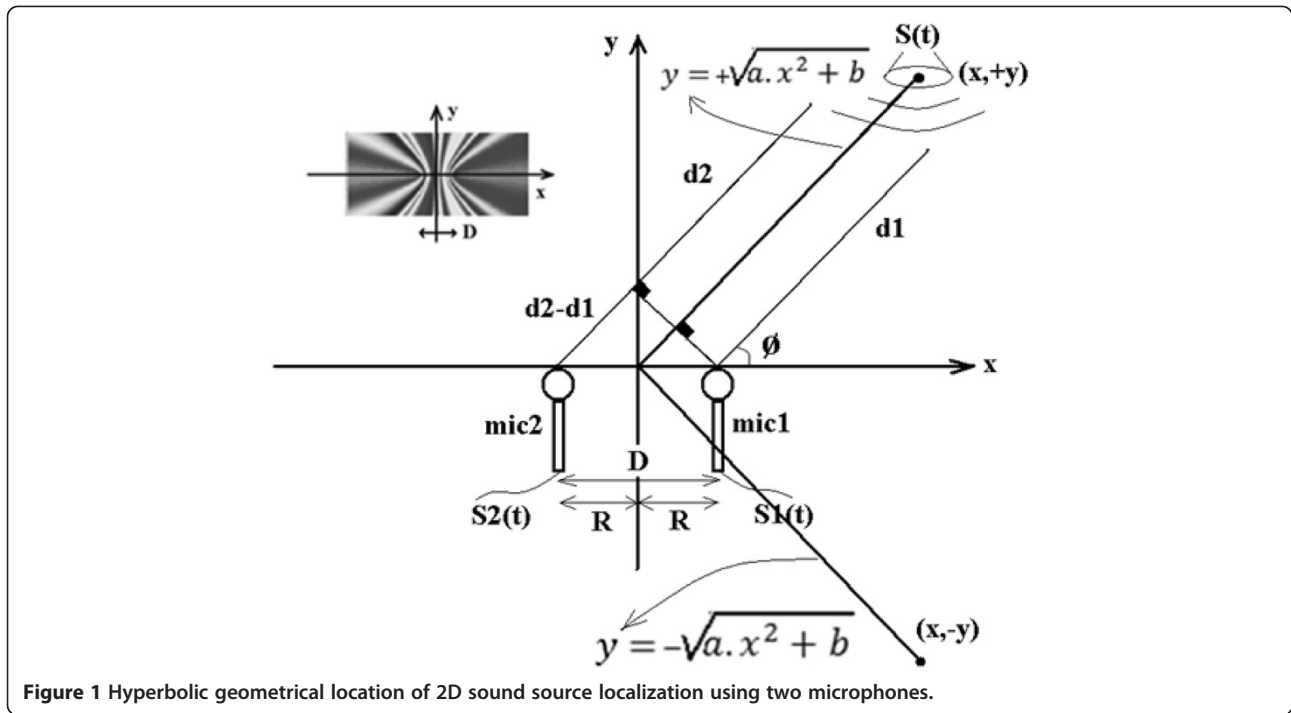


Figure 1 Hyperbolic geometrical location of 2D sound source localization using two microphones.

Using x and y instead of x_s and y_s , (27) will become:

$$x^2 - \left[\frac{2R(m+1)}{m-1} \right] x + y^2 = \frac{n}{m-1} - R^2 \quad (28)$$

$$\left(x - \frac{R(m+1)}{m-1} \right)^2 + y^2 = \frac{1}{m-1} \left(n + \frac{4m}{m-1} R^2 \right) \quad (29)$$

$$\begin{cases} (x-k)^2 + y^2 = l \\ k = \frac{R(m+1)}{m-1} \\ l = \frac{1}{m-1} \left(n + \frac{4m}{m-1} R^2 \right) \end{cases} \quad (30)$$

Therefore, source location is on a circle with center coordinates $(k, 0)$ and radius (\sqrt{l}) . Now, using a new microphone to find a new equation, in combination with one of the first or second microphones, helps us to have another circle which leads to source location with different center coordinates and different radii relative to the first circle. Intersection of the first and second circles gives us source location x and y [22,40].

4.2. Using PHAT method

Assuming a single frequency sound source with a wavelength equal to λ to have a distance from the center of two microphones equal to r , this source will be in far-field if [51-65]:

$$r > \frac{2D^2}{\lambda}, \quad (31)$$

where D is the distance between two microphones. In the far-field case, the sound can be considered as having the same angle of arrival to all microphones, as shown in Figure 1. If $s_1(t)$ is the output signal of the first microphone and $s_2(t)$ is that of the second microphone (Figure 1), taking into account the environmental noise, and according to the so-called inverse-square law, the signal received by the two microphones can be modeled as (6) and (7). The relative time shift between the signals is important for TDOA but can be ignored in ILD. Also, the attenuation coefficients $(1/d_1$ and $1/d_2)$ are important for ILD but can be ignored in TDOA. Therefore, assuming T_D is the time delay between the two received signals, the cross-correlation between $s_1(t)$ and $s_2(t)$ is [51-65]:

$$R_{s_1 s_2}(\tau) = \int_{-\infty}^{+\infty} s_1(t) s_2(t + \tau) dt. \quad (32)$$

Since $n_1(t)$ and $n_2(t)$ are independent, we can write

$$R_{s_1 s_2}(\tau) = \int_{-\infty}^{+\infty} s(t) s(t - T_D + \tau) dt. \quad (33)$$

Now, the time delay between these two signals can be measured as:

$$\tau = \operatorname{argmax}_{TD} R_{s_1 s_2}(\tau). \quad (34)$$

Correct measurement of the time delay needs the distance between the two microphones to be:

$$D \leq \frac{\lambda}{2}, \quad (35)$$

since when D is greater than $\frac{\lambda}{2}$, T_D is greater than π , and

therefore, time delay is measured as $\tau = -(T_D - \pi)$. According to Figure 1, the cosine of the angle of arrival is

$$\cos(\phi) = \frac{d_2 - d_1}{D} = \frac{(t_2 - t_1) v_{\text{sound}}}{D} = \frac{T_D \cdot v_{\text{sound}}}{D} = \frac{\tau_{21} \cdot v_{\text{sound}}}{D}. \quad (36)$$

Here, v_{sound} is sound velocity in air. The delay time τ_{21} is measurable using the cross-correlation function between the two signals. However, the location of source cannot be measured this way. We can measure the distance between source and each of the microphones as (13) and (14). The difference between these two distances will be

$$d_2 - d_1 = \tau_{21} \cdot v_{\text{sound}}. \quad (37)$$

Using x and y instead of x_s and y_s , τ_{21} will be

$$\tau_{21} = \frac{\sqrt{(x-x_2)^2 + (y-y_2)^2} - \sqrt{(x-x_1)^2 + (y-y_1)^2}}{v_{\text{sound}}}. \quad (38)$$

This is an equation with two unknowns, x and y . Assuming the distances of both microphones from the origin to be R ($D = 2R$) and both located on x axis,

$$\tau_{21} = \frac{\sqrt{(x+R)^2 + y^2} - \sqrt{(x-R)^2 + y^2}}{v_{\text{sound}}}. \quad (39)$$

Simplifying the above equation will result in:

$$\begin{cases} y^2 = a \cdot x^2 + b \\ a = \frac{4R^2}{v_{\text{sound}} \cdot \tau_{21}^2} - 1 \\ b = \frac{v_{\text{sound}}^2 \cdot \tau_{21}^2}{4} - R^2 \end{cases}, \quad (40)$$

where y has hyperbolic geometrical location relative to x , as shown in Figure 1. In order to find x and y , we need to add another equation to (38) for the first and a new (third) microphone so that:

$$\begin{cases} \tau_{21} = \frac{\sqrt{(x-x_2)^2 + (y-y_2)^2} - \sqrt{(x-x_1)^2 + (y-y_1)^2}}{v_{\text{sound}}} \\ \tau_{31} = \frac{\sqrt{(x-x_2)^2 + (y-y_3)^2} - \sqrt{(x-x_1)^2 + (y-y_1)^2}}{v_{\text{sound}}} \end{cases}. \quad (41)$$

It is noticeable that these are non-linear equations (Hyperbolic-intersection Closed-Form method) and numerical analysis should be used to calculate x and y , which will increase localization processing times. Also in this case, the solution may not converge.

5. TDE-ILD-based 2D sound source localization

Using either TDE or ILD to calculate source location (x and y) in 2D cases needs at least three microphones. Using TDE and ILD simultaneously helps us calculate source location using only two microphones. According to (26) and (37) and this fact that in a high SNR environment, the noise term $\eta/E2$ can be neglected, after some algebraic manipulations, we derive [40]

$$(x_s - x_1)^2 + (y_s - y_1)^2 = \left(\frac{\tau_{21} \cdot v_{\text{sound}}}{1 - \sqrt{m}} \right)^2 = r \frac{2}{1} \quad (42)$$

and

$$(x_s - x_2)^2 + (y_s - y_2)^2 = \left(\frac{\tau_{21} \cdot v_{\text{sound}} \cdot \sqrt{m}}{1 - \sqrt{m}} \right)^2 = r_2^2 \quad (43)$$

Intersection of two circles determined by (42) and (43), with center (x_1, y_1) and (x_2, y_2) , and radius r_1 and r_2 , respectively, gives the exact source position. In $E_1 = E_2$ ($m = 1$) case, both the hyperbola and the circle determined by (26) and (37) degenerate a line perpendicular bisector of microphone pair. Consequently, there will be no intersection to determine source position. Trying to obtain a closed-form solution to this problem, we transform the expression by [40]:

$$x_1 x_s + y_1 y_s = \frac{1}{2} (k_1^2 - r_1^2 + R_s^2) \quad (44)$$

and

$$x_2 x_s + y_2 y_s = \frac{1}{2} (k_2^2 - r_2^2 + R_s^2), \quad (45)$$

where

$$k_1^2 = x_1^2 + y_1^2, k_2^2 = x_2^2 + y_2^2 \text{ and } R_s^2 = x_s^2 + y_s^2. \quad (46)$$

Rewriting (44) and (45) into matrix form results in:

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{1}{2} \left\{ \begin{bmatrix} k_1^2 - r_1^2 \\ k_2^2 - r_2^2 \end{bmatrix} + R_s^2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \quad (47)$$

and

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1} \left(\frac{1}{2} \left\{ \begin{bmatrix} k_1^2 - r_1^2 \\ k_2^2 - r_2^2 \end{bmatrix} + R_s^2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \right). \quad (48)$$

If we define

$$p = \begin{bmatrix} p1 \\ p2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x1 & y1 \\ x2 & y2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (49)$$

and

$$q = \begin{bmatrix} q1 \\ q2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x1 & y1 \\ x2 & y2 \end{bmatrix}^{-1} \begin{bmatrix} k_1^2 - r_1^2 \\ k_2^2 - r_2^2 \end{bmatrix}, \quad (50)$$

then the source coordinates can be expressed with respect to R_s :

$$X = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} q_1 + p_1 R_s^2 \\ q_2 + p_2 R_s^2 \end{bmatrix}. \quad (51)$$

Inserting (46) into (51), the solution to R_s is obtained as:

$$R_s^2 = \frac{0_1 \pm 0_2}{0_3}, \quad (52)$$

where

$$\begin{aligned} 0_1 &= 1 - p_1 q_1 + p_2 q_2, \\ 0_2 &= \sqrt{(1 - p_1 q_1 + p_2 q_2)^2 + (p_1^2 + p_2^2)(q_1^2 + q_2^2)}, \\ 0_3 &= p_1^2 + p_2^2. \end{aligned}$$

The positive root gives the square of distance from source to origin. Substituting R_s into (51), the final source coordinate will be obtained [40].

However, a rational solution requires prior information of evaluation regions. It is known to us that, by using a linear array, two mirror points will be produced simultaneously. Assuming two microphones are on x -axis ($y_1 = y_2 = 0$) and have distance R from origin (Figure 1), According to (49) and (50), we cannot find p and q . Therefore, we cannot consider such a microphone arrangement. However, using this microphone arrangement simplifies equations. According to (26) and (37), we can intersect circle and hyperbola to find source location, x and y . For intersection of circle and hyperbola, firstly we rewrite (42) and (43), respectively, as

$$(x_s - x_1)^2 + (y_s - y_1)^2 = r_1^2 \quad (53)$$

and

$$(x_s - x_2)^2 + (y_s - y_2)^2 = r_2^2. \quad (54)$$

Using microphones coordinate values x and y instead of x_s and y_s , we will have

$$(x - R)^2 + (y - 0)^2 = r_1^2 \quad (55)$$

and

$$(x + R)^2 + (y - 0)^2 = r_2^2; \quad (56)$$

therefore,

$$r_1^2 - (x - R)^2 = r_2^2 - (x + R)^2 \quad (57)$$

which results in

$$r_2^2 - r_1^2 = 4Rx. \quad (58)$$

Hence, the sound source location can be calculated as:

$$x = (r_2^2 - r_1^2) / 4R \quad (59)$$

and

$$y = \pm \sqrt{r_1^2 - (x - R)^2} \quad (60)$$

We remember again that by using a linear array, two mirror points will be produced simultaneously. This means that we can localize 2D sound source only in half-plane.

6. Simulations of TDE-ILD-based method and discussion

In order to use the introduced method for sound source localization in low reverberant outdoor cases, firstly we performed simulations. We tried to evaluate the accuracy of this method in a variety of SNRs for some environmental noises. We considered two microphones on x -axis ($y_1 = y_2 = 0$) with 1-m distance from the origin ($x_1 = 1 \text{ m}$, $x_2 = -1 \text{ m}$ ($D = 2R = 2$)) (Figure 1). In order to use PHAT for the calculation of time delay between the signals of the two microphones, we used a helicopter sound (wideband and quasi-periodic signal) with a length of approximately 4 s, downloaded from the internet (www.freesound.com) as our sound source. For different source locations and for an ambient temperature of 15°C, first we calculated sound speed in air using:

$$v_{\text{sound}} = 20.05 \sqrt{273.15 + \text{Temperature}(\text{Centigrade})}. \quad (61)$$

Then, we calculated d_1 and d_2 using (24) and (25), and using (37), we calculated time delay between the received signals of the two microphones. For time delay positive values, i.e., sound source nearer to the first microphone (mic1 in Figure 1), we delayed second microphone signal with respect to the first microphone signal, and for negative values, i.e., sound source nearer to the second microphone (mic2 in Figure 1), did the opposite. Then using (6) and (7), we divided the first microphone signal by d_1 and the second microphone signal by d_2 to have correct attenuation in signals according to the source distances from microphones. Finally, we tried to calculate source location using the proposed TDE-ILD method (Section 5) in a variety of SNRs for some environmental noises.

For a number of signal-to-noise ratios (SNRs) for white Gaussian, pink and babble noises from 'NATO RSG-10 Noise Data,' 16-bit Quantization, and 96,000-Hz sampling frequency (hence, we upsampled NATO RSG-10 from its original 190,00 to 96,000 Hz), simulation results are shown in Figure 2 for source location of ($x = 10$, $y = 10$).

Assuming a high audio sampling rate, fractional delays are negligible and delays are rounded to nearest sampling point. Simulation results show larger localization error for SNRs lower than 10 dB. This issue occurs due to the use of ILD. Hence, we have to use this method for the case of only a dominant source to have accurate localization. Using spectral subtraction and source counting are useful for reducing localization error in SNRs lower than zero.

7. Our proposed TDE-ILD-HRTF method

Using TDE-ILD-based method, dual microphone 2D sound source localization is applicable. However, it is known that, by using a linear array in TDE-ILD-based method, two mirror points will be produced simultaneously (half-plane localization) [40]. Also, according to TDE-ILD-based simulation results (Section 6), it is noticeable that using ILD-based method needs only one dominant

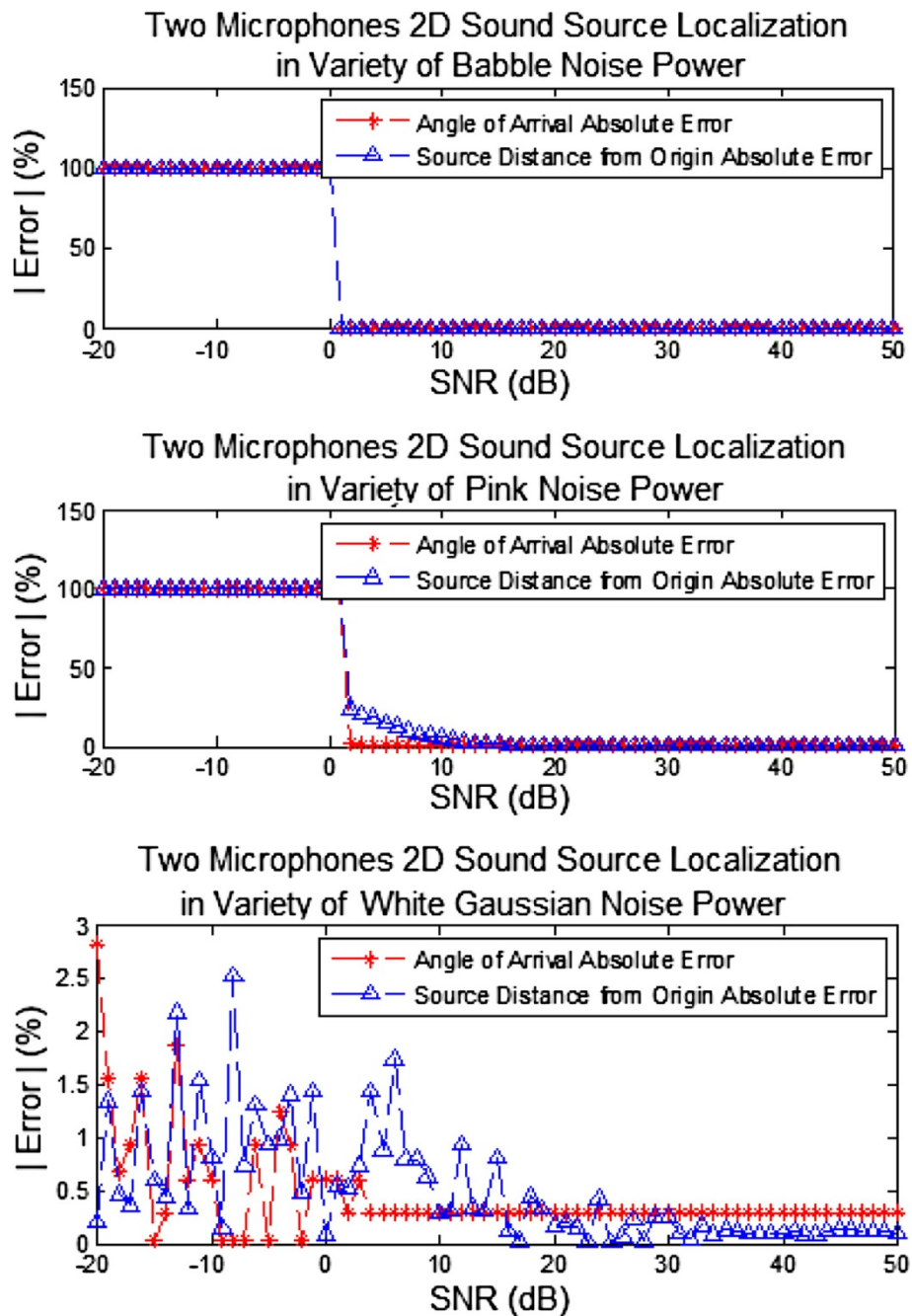


Figure 2 Simulation results for a variety of SNRs.

high SNR source to be active in localization area. Our proposed TDE-ILD-HRTF method tries to solve these problems using source counting, noise reduction using spectral subtraction, and HRTF.

7.1. Source counting method

According to previous discussions, ILD-based method needs to use source counting to find that one dominant source is active for high-resolution localization. If more than one source is active in localization area, it cannot calculate $m = E_1/E_2$ correctly. Therefore, we would need to count active and dominant sound sources and decide on localization of one sound source if only one source is dominant enough. PHAT gives us the cross-correlation vector of two microphone output signals. The number of dominant peaks of the cross-correlation vector gives us the number of dominant sound sources [66]. We consider only one source signal to be a periodic signal as:

$$s(t) = s(t + T). \quad (62)$$

If the signals window is greater than T, calculating cross-correlation between the output signals of the two microphones gives us one dominant peak and some weak peaks with multiples of T distances. However, using signals window of approximately equal to T or using non-periodic source signal would lead to only one dominant peak when calculating cross-correlation between the output signals of the two microphones. This peak value is delayed equal to the number of samples between the two microphones' output signals. Therefore, if one sound source is dominant in the localization area, only one dominant peak value will be in cross-correlation vector. Now, we consider having two sound sources $s(t)$ and $s'(t)$ in high SNR localization area. According to (6) and (7), we have

$$s_1(t) = s(t-T_1) + s'(t-T'_1) \quad (63)$$

$$s_2(t) = s(t-T_2) + s'(t-T'_2). \quad (64)$$

According to (32), we have

$$R_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} (s(t-T_1) + s'(t-T'_1))(s(t-T_2 + \tau) + s'(t-T'_2 + \tau))dt \quad (65)$$

$$R_{s_1s_2}(\tau) = R1_{s_1s_2}(\tau) + R2_{s_1s_2}(\tau) + R3_{s_1s_2}(\tau) + R4_{s_1s_2}(\tau) \quad (66)$$

where

$$R1_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} (t-T_1) \cdot s(t-T_2 + \tau)dt \quad (67)$$

$$R2_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} (t-T_1) \cdot s'(t-T'_2 + \tau)dt \quad (68)$$

$$R3_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} s'(t-T'_1) \cdot s(t-T_2 + \tau)dt \quad (69)$$

$$R4_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} s'(t-T'_1) \cdot s'(t-T'_2 + \tau)dt. \quad (70)$$

Using (34), $\tau_1 = T_2 - T_1$ gives us a maximum value for $R1_{s_1s_2}(\tau)$, $\tau_2 = T'_2 - T_1$, gives us a maximum value for $R2_{s_1s_2}(\tau)$, $\tau_3 = T_2 - T'_1$, gives us a maximum value for $R3_{s_1s_2}(\tau)$ and $\tau_4 = T'_2 - T'_1$, and gives us a maximum value for $R4_{s_1s_2}(\tau)$. Therefore, we will have four peak values in cross-correlation vector. However, according to this fact that (67) and (70) are cross-correlation functions of a signal with delayed version of itself, and (68) and (69) are cross-correlation functions of two different signals, τ_1 and τ_4 maximum values are dominant with respect to τ_2 and τ_3 values. Now, we conclude in two dominant sound sources area, cross-correlation vector will have two dominant values and therefore equal count dominant values for more than two dominant sound sources signals as multiple power spectrum peaks in DOA-based multiple sound source beamforming methods [16]. Therefore, counting dominant cross-correlation vector values, we can find the number of active and dominant sound sources in localization area.

7.2. Noise reduction using spectral subtraction

In order to apply ILD in TDE-ILD-based dual microphone 2D sound source localization, source counting is used to find that one dominant high SNR source is active in localization area. Source counting was proposed to calculate the number of active sources in localization area. Furthermore, spectral subtraction can be used for noise reduction and therefore increasing active source's SNR. Also, according to the background noise, such as wind, rain, and babble, we can consider a background spectrum estimator. In the most practical cases, we can assume that the noisy signal can be modeled as the sum of the clean signal and the noise [67]:

$$s_n(t) = s(t) + n(t). \quad (71)$$

Also according to this fact that the signal and noise are generated by independent sources, they are considered uncorrelated. Therefore, the noisy spectrum can be written as:

$$S_n(w) = S(w) + N(w). \quad (72)$$

During the silent periods, i.e., periods without target sound, it can be estimated background noise spectrum, considering the noise to be stationary. Then, the noise magnitude spectrum can be subtracted from the noisy input magnitude spectrum. In non-stationary noise cases,

there can be used an adaptive background noise spectrum estimation procedure [67].

7.3. Using HRTF method

Using a linear array in TDE-ILD-based dual microphone 2D sound source localization method leads to two mirror points produced simultaneously. Adding an HRTF-inspired idea, whole-plane dual microphone 2D sound source localization would be possible. This idea was published in [68], and it is reviewed here again. Researchers have used an artificial ear with a spiral shape. This is a special type of pinna with a varying distance from a microphone placed in the center of its spiral [30]. However, we consider a half-cylinder instead of artificial ear [68]. Due to the use of such a reflector, a constant notch position is created for all variations (0 to 180°) of sound signal angle of arrival (Figure 3). However, clearly, and as shown in the figure, a complete half-cylinder scatters the sound waves from sources behind it. In order to overcome this problem we consider slits on the surface of the half-cylinder (Figure 3). If d is the distance between the reflector (half-cylinder) and the microphone (placed at the center), a notch will be created when it is equal to quarter of the wavelength of sound, λ , plus any multiples of $\lambda/2$. For such wavelengths, the incident waves are reduced by reflected waves [30]:

$$n \cdot \left(\frac{\lambda}{2}\right) + \left(\frac{\lambda}{4}\right) = d \quad n = 0, 1, 2, 3 \dots \quad (73)$$

These notches will appear at the following frequencies:

$$f = \frac{c}{\lambda} = \frac{(2n + 1) \cdot c}{4d} \quad (74)$$

Covering only microphone 2 in Figure 1 by reflector, calculating the interaural spectral difference results in

$$\begin{aligned} |\Delta H(f)| &= \left| 10 \log_{10} H_1(f) - 10 \log_{10} H_2(f) \right| \\ &= \left| 10 \log_{10} \frac{H_1(f)}{H_2(f)} \right|. \end{aligned} \quad (75)$$

High $|\Delta H(f)|$ values indicate that the sound source is in front, while negligible values indicate that sound source is at the back. One important point is that in order to have the same spectrum in both microphones when the sound source is at the back, careful design of the slits is necessary.

7.4. Extension of dimensions to three

According to the use of half-cylinder reflector in the proposed localization system, this approach is only applicable in 2D cases. Using a half-sphere reflector instead of the half-cylinder makes it usable in 3D sound source localization (Figure 4). Adding the half-sphere reflector only to microphone 2 in Figure 5 allows the localization system to localize sound sources in 3D cases.

For 3D case, we can consider a plane that passes through source position and x -axis and which makes an angle of θ with the x - y plane, and we name it source-plane (Figure 5). This plane consists of the microphones 1 and 2. According to these assumptions, using (36), we can calculate ϕ which is the angle of arrival in source-plane and is equal to angle of arrival in x - y plane. Then, using (59) and (60), we can also calculate x and y in 2D source-plane (not in x - y plane in 3D case) and therefore calculate sound source distance $r\sqrt{x^2 + y^2}$. Introducing a new microphone (mic3 in Figure 5) with $y = 0$ and $x =$

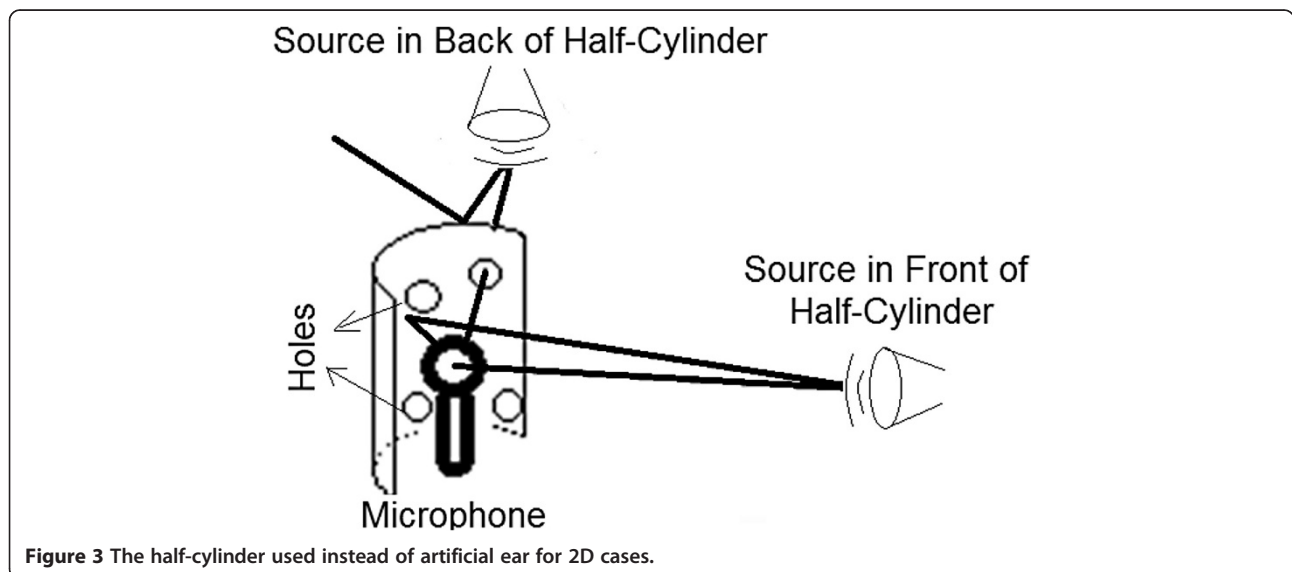


Figure 3 The half-cylinder used instead of artificial ear for 2D cases.

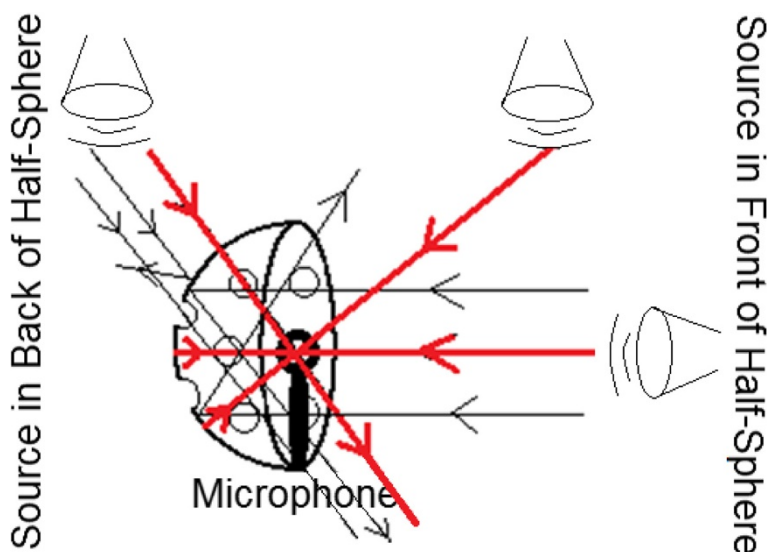


Figure 4 Half-sphere used instead of artificial ear for 3D cases.

$z=R$ helps us calculate the angle θ . Therefore, using (36), we can calculate θ as:

$$\theta = \cos^{-1}\left(\frac{V_{\text{sound}} \cdot T_{31}}{R}\right). \quad (76)$$

Using half-sphere reflector decreases accuracy of time delay and intensity level deference estimation between microphones 1 and 2 due to the change in the spectrum of second microphone's signal. However, covering the third microphone instead of the second microphone by half-sphere reflector only decreases the accuracy of time

delay estimation between the first and third microphones (Figure 5). Multiplying the inverse function of notch-filter in third microphone's spectrum leads to increase in accuracy. Now using r , ϕ , and θ , we can calculate source location x , y and z in 3D case:

$$\begin{cases} x = r \cdot \cos(\phi) \cdot \sin(\theta) \\ y = r \cdot \sin(\phi) \cdot \sin(\theta) \\ z = r \cdot \cos(\theta) \end{cases} \quad (77)$$

The reasons for choosing the shape of sphere for the reflector are as follows [69]. The simplest type of

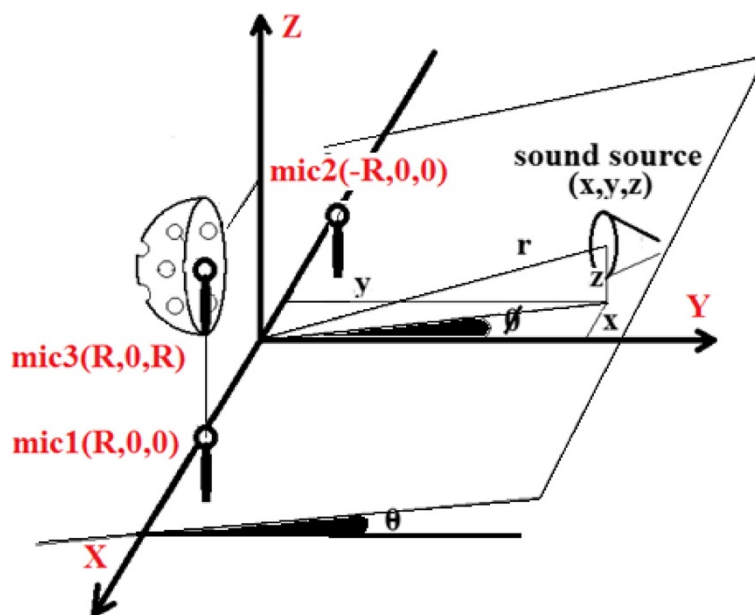


Figure 5 3D sound source localization using TDE-ILD-HRTF method.

reflector is a plane reflector introduced to direct signal in a desired direction. Clearly, using this type of reflector, the distance between reflector and microphone, d in (73), varies with respect to source position in 3D cases leading to a change in notch position within spectrum. The change in notch position may not be suitable as it might occur out of the spectral band of interest. To better adjust the energy in the forward direction, the geometrical shape of the plane reflector must be changed so as not to allow radiation in the back and side directions.

A possible arrangement for this purpose consists of two plane reflectors joined on one side to form a corner. This type of reflector returns the signal exactly in the same direction as it is received. Because of this unique feature, reflected wave acquired by the microphone is unique, which is our aim. Whereas having more than one reflected wave causes a higher energy value with respect to a single reflected wave at the microphone, which causes a deep notch. But using this type of reflector, d is varied with respect to the source position in 3D cases which is related to notch position in spectrum. It has been shown that if a beam of parallel waves is incident upon a parabola reflector, the radiation will focus at a spot which is known as the focal point. This point in spherical reflector does not match to the center in accordance with this fact that the center has equal distance (d) to all points of the reflector surface. Because of this feature, reflected wave which is received by the microphone is unique, whatever the position of the sound source, in 3D cases and belongs to the wave passing the center. The center of the spherical reflector with radius R is located at O (Figure 6). The sound wave axis strikes the reflector at B . From the Law of Reflection and angle geometry of parallel lines, the marked angles are equal. Hence, BFO is an equilateral triangle. Dropping the median from F which is stapled to BO and using trigonometry, the focal point is calculated as:

$$\frac{R}{2b} = \cos(\theta) \rightarrow b = \frac{R}{2 \cos \theta} \quad (78)$$

$$\rightarrow F = R - \frac{R}{2 \cos(\theta)}. \quad (79)$$

F is not equal to R for all θ values. The half-spherical reflector scatters the sound source waves hitting it from back. Therefore, we consider some slits on its surface (Figure 4). Obviously, the slits need to be designed in a fashion that leads to the same spectrum in both microphones when sound source is in back. When a plane wave hits a barrier with a single circular slit narrower than signal wavelength (λ), the wave bends and emerges from the slit as a circular wave [70]. If L is the distance

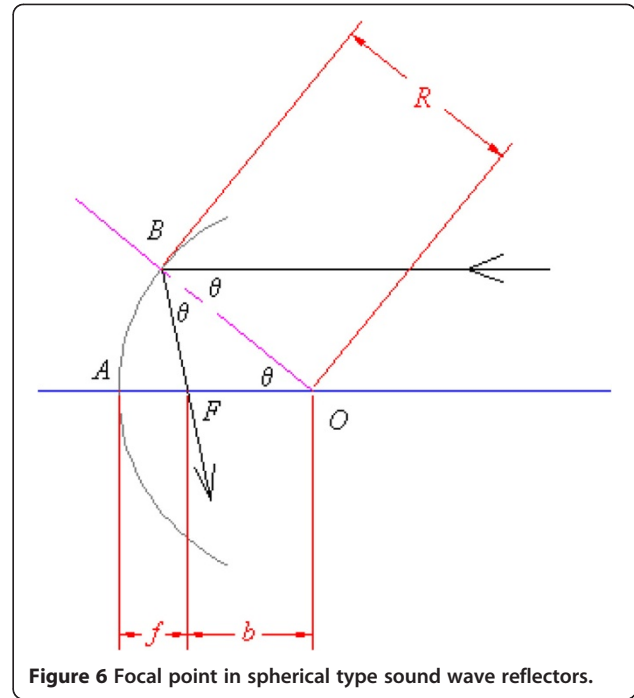


Figure 6 Focal point in spherical type sound wave reflectors.

between the slit and the viewing screen and d is the slit width, based on the assumption that $L \gg d$ (Fraunhofer scattering), the wave intensity distribution observed (on a screen) at an angle θ with respect to the incident direction is given by:

$$if k = \frac{\pi \cdot d}{\lambda} \sin(\theta) \rightarrow \frac{I(\theta)}{I(0)} = \frac{\sin^2(k)}{k^2}, \quad (80)$$

where $I(\theta)$ is wave intensity in direction of observation and $I(0)$ is maximum wave intensity of diffraction pattern (central fringe). This relationship will have a value of zero each time $\sin^2(k) = 0$. This occurs when

$$k = \pm m\pi \text{ or } \frac{\pi \cdot d}{\lambda} \sin(\theta) = \pm m\pi, \quad (81)$$

yielding the following condition for observing minimum wave intensity from a single circular slit:

$$\sin(\theta) = \frac{m\lambda}{d} \quad m = 0, \pm 2, \dots \quad (82)$$

This relationship is satisfied for integer values of m . Increasing values of m gives minima at correspondingly larger angles. The first minimum will be found for $m = 1$, the second for $m = 2$, and so forth. If $\frac{d}{\lambda} \sin(\theta)$ is less than one for all values of θ , i.e., when the size of the aperture is smaller than a wavelength ($d < \lambda$), there will be no minima. As we need more than one circular slit on half-sphere reflector's surface, we consider a parallel wave incident on a barrier that consists of two closely spaced narrow

circular slits S1 and S2. The narrow circular slits split the incident wave into two coherent waves. After passing through the slits, the two waves spread out due to diffraction and interfere with one another. If this transmitted wave is made to fall on a screen some distance away, an interference pattern of bright and dark fringes are observed on the screen, with the bright ones corresponding to regions of maximum wave intensity while the dark ones corresponding to those of minimum wave intensity. As discussed before, for all points on the screen where the path difference is some integer multiple of the wavelength, the two waves from the slits S1 and S2 arrive in phase and bright fringes are observed. Thus, the condition for producing bright fringes is as in (82). Similarly, dark fringes are produced on the screen if the two waves arriving on the screen from slits S1 and S2 are exactly out of phase. This happens if the path difference between the two waves is an odd integer multiple of half-wavelengths:

$$\sin(\theta) = \frac{(m + \frac{1}{2}) \cdot \lambda}{d} \quad m = 0, \pm 1, \pm 2, \dots \quad (83)$$

Of course, this condition is changed using a half-sphere instead of a plane (Figure 7), where according to the same distance R to the center, the two waves from the slits S1 and S2 arrive in phase at the center and bright fringes are observed there. This signal intensity magnification is not suitable for the case of estimating intensity level difference between two microphones where one of them is covered by this reflector. However, covering the third microphone by half-sphere reflector is suitable where there is no need to estimate the intensity level difference between two microphones 1 and 3.

8. TDE-ILD-HRTF approach algorithm

According to previous discussions, the following steps can be considered for our proposed method:

1. Setup microphones and hardware.
2. Calculate the sound recording hardware set (microphone, preamplifier, and sound card) amplification normalizing factor.
3. Apply voice activity detection. Is there valuable signal? yes \rightarrow go to 4 no \rightarrow go to 3.
4. Obtain $s_1(t)$, $s_2(t)$, and $s_3(t) \rightarrow m = E_1/E_2$.
5. Remove DC from the signals. Then normalize them regarding the sound intensity.
6. Hamming window signals regarding their stationary parts (at least about 100 ms for wideband quasi-periodic helicopter sound or twice that).
7. Apply FFT to signals.
8. Cancel noise using spectral subtraction.
9. Apply PHAT to the signals in order to calculate τ_{21} and τ_{31} in frequency domain (index of first maximum value). Then find the second maximum value of cross-correlation vector.
10. If the first maximum value is not dominant enough with respect to the second maximum value, go to the next windows of signals and do not calculate sound source location, otherwise:

a.

$$\Phi = \cos^{-1}\left(\frac{v_{\text{sound}} \cdot \tau_{21}}{2R}\right) \text{ and } \theta = \cos^{-1}\left(\frac{v_{\text{sound}} \cdot \tau_{31}}{R}\right)$$

b.

$$v_{\text{sound}} = 20.05 \sqrt{273.15 + \text{Temperature}(\text{Centigrade})}$$

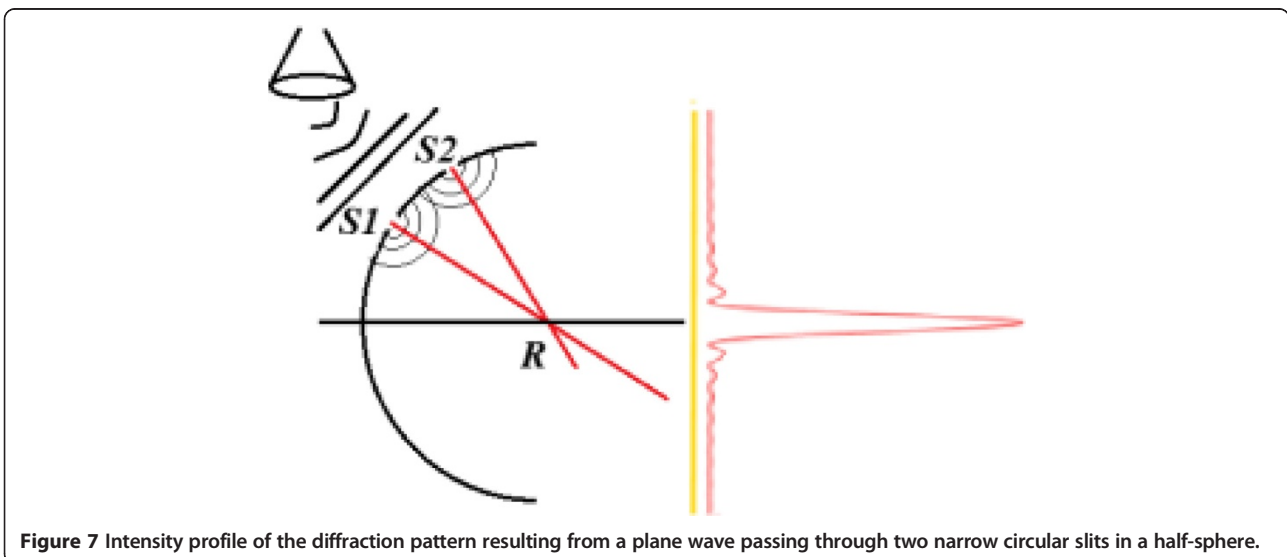


Figure 7 Intensity profile of the diffraction pattern resulting from a plane wave passing through two narrow circular slits in a half-sphere.

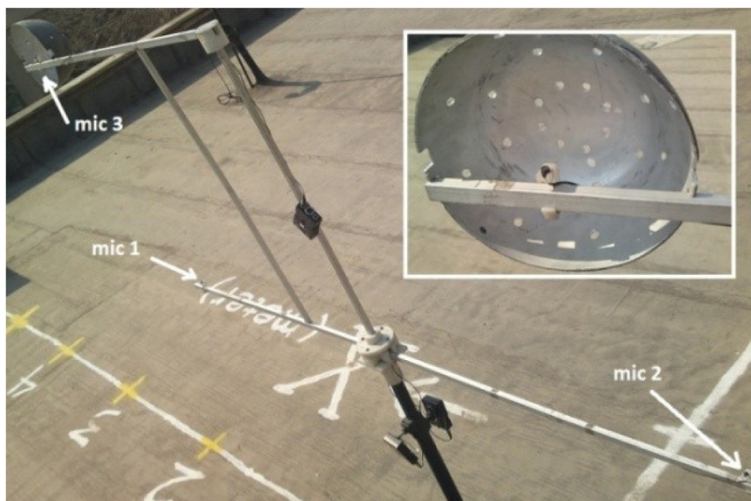


Figure 8 Setup of the proposed 3D sound source localization method using four microphones.

c.

$$r_1 = \frac{r_{21} \cdot v_{\text{sound}}}{1 - \sqrt{m}} \text{ and } r_2 = \frac{r_{21} \cdot v_{\text{sound}} \cdot \sqrt{m}}{1 - \sqrt{m}}$$

d.

$$x = (r_2^2 - r_1^2) / 4R \text{ and } y = \pm \sqrt{r_1^2 - (x - R)^2}$$

e.

$$|\Delta H(f)| = \left| 10 \log_{10} \frac{H_1(f)}{H_3(f)} \right|$$

f.

$$\text{if } (|\Delta H(f)| \approx 0) \text{ } y = -\sqrt{r_1^2 - (x - R)^2}$$

$$\text{else } y = \sqrt{r_1^2 - (x - R)^2}$$

$$\rightarrow \begin{cases} x_s = r \cdot \cos(\Phi) \cdot \sin(\theta) \\ y_s = r \cdot \sin(\Phi) \cdot \sin(\theta) \\ z_s = r \cdot \cos(\theta) \end{cases}$$

g. Go to 3.

9. Hardware and software implementations and results

We implemented our proposed hardware using 'MB800H' motherboard and DELTA IOIO LT sound card from M-Audio featuring eight analogue sound inputs, 96-kHz sampling frequency, and up to 16-bit resolution. Three MicroMic microphones of type 'C 417' with B29L pre-amplifier were used for the implementation (Figure 8). The reason for using this type of microphone is its very

flat spectral response up to 5 kHz. Visual C++ used for software implementation. According to the very low sensitivity of the used microphones (10 mV/Pa), a loudspeaker was used for the sound source. Initially, we used microphones one and two to evaluate the accuracy of the angle of arrival ϕ as well as microphones one and three to evaluate the accuracy of the angle of arrival θ . We considered 1 m ($R = D/2 = 1$) distance between every microphone and the origin. Sound localization results showed that sound velocity changed in different temperatures. Since the sound velocity was used in most of the steps of the proposed methods, sound source location calculations were carried out inaccurately. Therefore, (61) was used based on a thermometer reading in order to more

Table 1 Results of the azimuth angle of arrival based on hardware implementation

ϕ Real angle (degrees)	ϕ Proposed method results for angle of arrival (degrees)	Absolute value of error (degrees)
0	0.18	0.18
15	15.13	0.13
30	29.85	0.15
45	45.18	0.18
60	60.14	0.14
75	74.91	0.09
90	89.83	0.17
105	104.82	0.18
120	120.14	0.14
135	135.11	0.11
150	150.14	0.14
165	165.09	0.09
180	179.93	0.07

Table 2 Results of the elevation angle of arrival based on hardware implementation

θ Real angle (degrees)	θ Proposed method results for angle of arrival (degrees)	Absolute value of error (degrees)
-10	-10.18	0.18
-5	-5.16	0.16
0	0.05	0.05
15	14.15	0.15
30	29.91	0.09
45	45.19	0.19
60	60.13	0.13
75	75.11	0.11
90	89.82	0.18

accurately find the sound velocity in different temperatures. Angle of arrival calculations using downloaded wideband quasi-periodic helicopter sound resulted in Tables 1 and 2. Acquisition hardware was initialized to 96-kHz sampling frequency and 16-bit resolution. Also signal acquisition window length was set to 100 ms. A custom 2-cm-diameter reflector (Figure 8) was placed behind the third microphone. Using the aforementioned sound source, the sound spectra shown in Figure 9 were measured using the third microphone. Figure 9a shows the spectrum when

the sound source has been placed behind the reflector and Figure 9b when placed in front of it. Notches can be spotted in the second spectrum according to the reflector diameter and (74). Finally, using our proposed TDE-ILD-HRTF approach, first we tried to find that a dominant source is active in localization area. Then, using (72), we tried to reduce background noise and localize sound source in 3D space. Table 3 shows the improved results after the application of noise reduction method. Although PHAT features high performance in noisy and reflective environments, due to the rather constant and high valued signal-to-reverberation ratio (SRR) in outdoor cases, this feature could not be evaluated. We tried to find the effect of probable reverberation on our localization system results. In order to alleviate reverberation, our experiments indicated that at least a distance of 1.2 m with surrounding buildings would be necessary. Although the loudspeaker's power, the hardware set (microphone, preamplifier, and sound card) amplification factor and the distance of the microphones with the surrounding buildings indicate the SNR value for reverberate signals in a real environment, this distance can be calculated, but it can also be indicated experimentally, which is easier. Tables 1 and 2 indicate that our implemented 3D sound source localization method features less than 0.2 degree error for angle of arrival. Table 3 indicates less than 10% error for 3D

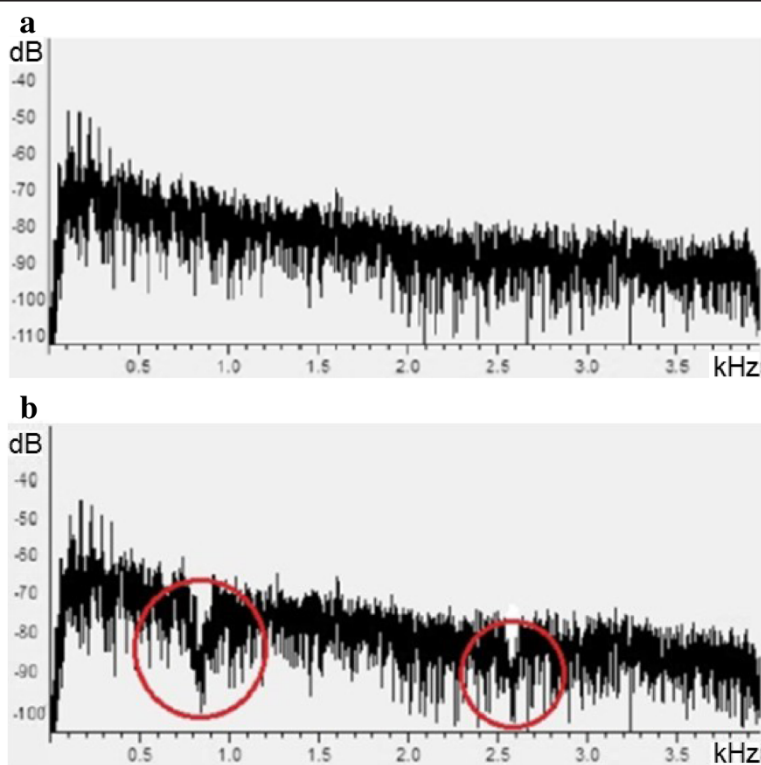


Figure 9 Recorded sound spectrum using the third microphone. (a) When the source has been placed behind the reflector; (b) when the source has been placed in front of the reflector. The logarithmic (dB) graphs indicate spectral level in comparison to 16-bit full scale.

Table 3 Results for proposed 3D sound source localization method (Section 7) using noise reduction procedure

Real location (m)			Proposed method results (m)			Absolute error (m)		
x	y	z	x	y	z	x	y	z
1	10	5	1.9	9.2	4.1	0.9	0.8	0.9
2	10	5	2.9	9.2	4.3	0.9	0.8	0.7
3	10	5	2.2	10.7	5.6	0.8	0.7	0.6
4	10	5	3.2	10.7	5.8	0.8	0.7	0.8
5	10	5	5.5	9.5	5.8	0.5	0.5	0.8
6	10	5	6.5	10.5	4.5	0.5	0.5	0.5
7	10	5	7.7	10.8	4.4	0.7	0.8	0.6
8	10	5	8.6	10.8	4.4	0.6	0.8	0.6
9	10	5	9.8	9.1	4.2	0.8	0.9	0.8
10	10	5	9.5	9.4	5.9	0.5	0.6	0.9
10	15	5	10.6	15.7	5.5	0.6	0.7	0.5
10	20	5	10.7	19.2	5.6	0.7	0.8	0.6
10	25	5	9.2	23.9	5.7	0.8	1.1	0.7
10	30	5	10.5	31.7	4.3	0.5	1.7	0.7
10	9	5	10.8	9.7	5.8	0.8	0.7	0.8
10	8	5	9.3	8.7	5.8	0.7	0.7	0.8
10	1	5	10.6	6.5	5.9	0.6	0.5	0.9
10	6	5	9.5	6.6	4.1	0.5	0.6	0.9
10	5	5	10.5	4.5	4.2	0.5	0.5	0.8
10	4	5	10.6	3.4	5.7	0.6	0.6	0.7
10	3	5	9.3	3.6	5.5	0.7	0.6	0.5
10	2	5	10.8	±2.8	5.8	0.8	0.8	0.8
10	1	5	10.1	±0.2	4.3	0.1	0.8	0.7
10	-1	5	10.6	±0.4	4.1	0.6	0.6	0.9
10	-2	5	10.5	±2.7	5.6	0.5	0.7	0.6
10	-3	5	10.4	-2.8	5.5	0.4	0.2	0.5
10	-4	5	9.7	-3.7	5.5	0.3	0.3	0.5
10	-5	5	9.4	-5.2	5.9	0.6	0.2	0.9

location finding. It also indicates ambiguity in sign finding between 0.5° and -0.5° as well as 175° and 185° due to the implemented reflector edges. We measured processing times of less than 150 ms for every location calculation. Comparing angle of arrival results with the results reported in various cited papers, the best reported results (on approximately similar tasks) were 1 degree limit error in [71], 2 degree limit error in [72], and within 8 degree error in [73]. Note that the fundamental limit of bearing estimates determined by Cramer-Rao Lower Bound (CRLB) for $D = 1$ cm spacing of the microphone pair, sampling frequency of 25 kHz and estimation duration of 100 ms, is computed to be 1° [71]. This is found equal to 1.4° for our system. However, only few respectable references have calculated 2D or 3D source locations accurately. Most of such research reports, besides calculating

the angle of arrival, used either least-square error estimation or Kalman filtering techniques to predict motion path of sound sources, in order to overcome their local calculation errors. Examples are [2], [13], [15], and [19]. Meanwhile, comparison of our results with these and other works (results on approximately similar tasks), such as [72] and [73], shows the approximately same accuracy nature of our proposed source location measurement approach.

10. Conclusion

In this paper, we reported on the simulation of TDE-ILD-based 2D half-plane sound source localization using only two microphones. Reduction of the microphone count was our goal. Therefore, we also proposed and implemented TDE-ILD-HRTF-based 3D entire-space sound

source localization using only three microphones. Also, we used spectral subtraction and source counting methods in low-degree reverberation outdoor cases to increase localization accuracy. According to Table 3, implementation results show that the proposed method has led to less than 10% error for 3D location finding. This is a higher accuracy in source location measurement in comparison with similar researches which did not use spectral subtraction and source counting. Also, we indicated that partly covering one of the microphones by a half-sphere reflector leads to entire-space N -dimensional sound source localization using only N -microphones.

Competing interests

The authors declare that they have no competing interests.

Authors' information

AP was born in Azerbaijan. He has a Ph.D. in Electrical Engineering (Signal Processing) from the Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran. He is now an invited lecturer at the Electrical Engineering Department, Amirkabir University of Technology and has been teaching several courses (C++ programming, multimedia systems, digital signal processing, digital audio processing, and digital image processing). His research interests include statistical signal processing and applications, digital signal processing and applications, digital audio processing and applications (acoustic modeling, speech coding, text to speech, audio watermarking and steganography, sound source localization, determined and underdetermined blind source separation and scene analyzing), digital image processing and applications (image and video coding, image watermarking and steganography, video watermarking and steganography, object tracking and scene matching) as well as BCI.

SMA received his B.S. and M.S. degrees in Electronics from the Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran, in 1984 and 1987, respectively, and his Ph.D. in Engineering from the University of Cambridge, Cambridge, U.K., in 1996, in the field of speech processing. Since 1988, he has been a Faculty Member at the Electrical Engineering Department, Amirkabir University of Technology, where he is currently an associate professor and teaches several courses and conducts research in electronics and communications. His research interests include speech processing, acoustic modeling, robust speech recognition, speaker adaptation, speech enhancement as well as audio and speech watermarking.

Received: 24 June 2013 Accepted: 21 November 2013

Published: 6 December 2013

References

1. GC Carter, Time delay estimation for passive sonar signal processing. *IEEE T Acoust S.* **ASSP-29**, 462–470 (1981)
2. E Weinstein, Optimal source localization and tracking from passive array measurements. *IEEE T. Acoust. S.* **ASSP-30**, 69–76 (1982)
3. H Wang, P Chu, *Voice source localization for automatic camera pointing system in videoconferencing*, in *Proceedings of the ICASSP, New Paltz, NY*, 19–22, 1997
4. J Caffery Jr, G Stüber, Subscriber location in CDMA cellular networks. *IEEE Trans. Veh. Technol.* **47**, 406–416 (1998)
5. JB Tsui, *Fundamentals of Global Positioning System Receivers* (Wiley, New York, 2000)
6. F Klein, *Finding an earthquake's location with modern seismic networks* (Earthquake Hazards Program, Northern California, 2000)
7. Y Huang, J Benesty, GW Elko, RM Mersereau, Real-time passive source localization: A practical linear-correction least-squares approach. *IEEE T. Audio P.* **9**, 943–956 (2001)
8. VJM Michaud, F Rouat, J Letourneau, *Robust Sound Source Localization Using a Microphone Array on a Mobile Robot* (in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, Las Vegas), pp. 1228–1233
9. BLF Daku, JE Salt, L Sha, AF Prugger, An algorithm for locating microseismic events, in *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering*, Niagara Falls. **2–5**, 2311–2314 (May 2004)
10. N Patwari, JN Ash, S Kyperountas, AO Hero III, RL Moses, NS Correal, Locating the nodes. *IEEE Signal Process. Mag.* **22**, 54–69 (2005)
11. S Gezici, T Zhi, G Giannakis, H Kobayashi, A Molisch, H Poor, Z Sahinoglu, Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks. *IEEE Signal Process. Mag.* **22**, 70–84 (2005)
12. JY Lee, SY Ji, M Hahn, C Young-Jo, *Real-Time Sound Localization Using Time Difference for Human-Robot Interaction* (in *Proceedings of the 16th IFAC World Congress*, Prague, 2005)
13. WK Ma, BN Vo, SS Singh, A Baddeley, Tracking an unknown time-varying number of speakers using TDOA measurements: a random finite set approach. *IEEE Trans. Signal. Process.* **54**(9), 3291–3304 (2006)
14. KC Ho, X Lu, L Kovavisaruch, Source localization using TDOA and FDOA measurements in the presence of receiver location errors: analysis and solution. *IEEE Trans Signal. Process.* **55**(2), 684–696 (2007)
15. V Cevher, AC Sankaranarayanan, JH McClellan, R Chellappa, Target tracking using a joint acoustic video system. *IEEE Trans. Multimedia.* **9**(4), 715–727 (2007)
16. X Zhengyuan, N Liu, BM Sadler, *A Simple Closed-Form Linear Source Localization Algorithm* (in *IEEE MILCOM*, Orlando, FL, 2007), pp. 1–7
17. JPV Etten, Navigation systems: fundamentals of low and very-low frequency hyperbolic techniques. *Elec. Comm.* **45**(3), 192–212 (1970)
18. RO Schmidt, A new approach to geometry of range difference location. *IEEE Trans. Aerosp. Electron Syst.* **AES-8**(6), 821–835 (1972)
19. HB Lee, A novel procedure for assessing the accuracy of hyperbolic multilateration systems. *IEEE Trans. Aerosp. Electron Syst.* **110**(1), 2–15 (1975)
20. WR Hahn, Optimum signal processing for passive sonar range and bearing estimation. *J. Acoust. Soc. Am.* **58**, 201–207 (1975)
21. P Julian, A Comparative, Study of Sound Localization Algorithms for Energy Aware Sensor Network Nodes. *IEEE Trans. Circuits and Systems - I.* **51**(4) (2004)
22. ST Birchfield, R Gangishetty, *Acoustic Localization by Interaural Level Difference* (in *Proceedings of the ICASSP*, Philadelphia, PA, 2005), pp. 1109–1112
23. KC Ho, M Sun, An accurate algebraic Closed-Form solution for energy-based source localization. *IEEE Trans. Audio, Speech, Lang. Process.* **15**, 2542–2550 (2007)
24. KC Ho, M Sun, Passive Source Localization Using Time Differences of Arrival and Gain Ratios of Arrival. *IEEE Trans. Signal. Process.* **56**, 2 (2008)
25. JH Hebrank, D Wright, Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.* **56**, 1829–1834 (1974)
26. JC Middlebrooks, JC Makous, DM Green, Directional sensitivity of sound-pressure level in the human ear canal. *J Acoust Soc Am.* **1**, 86 (1989)
27. F Asano, Y Suzuki, T Sone, Role of spectral cues in median plane localization. *J. Acoust. Soc. Am.* **88**, 159–168 (1990)
28. RO Duda, WL Martens, Range dependence of the response of a spherical head model. *J. Acoust. Soc. Am.* **104**(5), 3048–3058 (1998)
29. CI Cheng, GH Wakefield, Introduction to head-related transfer functions (HRTFs): representations of HRTFs in time, frequency, and space. *J. of the Audio Engin. Soc.* **49**(4), 231–248 (2001)
30. A Kulaib, M Al-Mualla, D Vernon, *2D Binaural Sound Localization for Urban Search and Rescue Robot*. in *Proceedings of the 12th Int* (Conference on Climbing and Walking Robots, Istanbul, 2009), pp. 9–11
31. S Razin, Explicit (non-iterative) Loran solution. *J. Inst. Navigation* **14**(3) (1967)
32. C Knapp, G Carter, The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-24**(4), 320–327 (1976)
33. MS Brandstein, H Silverman, A practical methodology for speech localization with microphone arrays. *Comput Speech Lang.* **11**(2), 91–126 (1997)
34. MS Brandstein, JE Adcock, HF Silverman, A Closed-Form location estimator for use with room environment microphone arrays. *IEEE Trans. Speech Audio Process.* **5**(1), 45–50 (1997)
35. P Svnizer, M Matnssoni, M Omologo, *Acoustic Source Location in a THREE-Dimensional Space Using Crosspower Spectrum Phase* (in *Proceedings of the ICASSP*, Munich, 1997), p. 231
36. E Lleida, J Fernandez, E Masgrau, *Robust continuous speech recog. sys. based on a microphone array* (in *Proceedings of the ICASSP*, Seattle, WA, 1998), pp. 241–244
37. A Stephenne, B Champagne, *Cepstral prefiltering for time delay estimation in reverberant environments* (in *Proceedings of the ICASSP*, Detroit, MI, 1995), pp. 3055–3058

38. M Omologo, P Svaizer, *Acoustic Event Localization using a Crosspower-Spectrum Phase based Technique* (in Proceedings of the ICASSP, vol. 2, Adelaide, 1994), pp. 273–276
39. C Zhang, D Florencio, Z Zhang, *Why Does PHAT Work Well in Low Noise, reverberative Environments* (in Proceedings of the ICASSP, Las Vegas, NV, 2008), pp. 2565–2568
40. W Cui, Z Cao, J Wei, *DUAL-Microphone Source Location Method in 2-D Space* (in Proceedings of the ICASSP, Toulouse, 2006), pp. 845–848
41. X Sheng, YH Hu, Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Trans Signal Process* **53**(1), 44–53 (2005)
42. D Blatt, AD Hero III, Energy-based sensor network source localization via projection onto convex sets. *IEEE Trans Signal Process* **54**(9), 3614–3619 (2006)
43. S Sen, A Nehorai, Performance analysis of 3-D direction estimation based on head-related transfer function. *IEEE Trans. Audio Speech Lang. Process* **17**(4) (2009)
44. DW Batteau, The role of the pinna in human localization. *Series B, Biol. Sci* **168**, 158–180 (1967)
45. SK Roffler, RA Butler, Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.* **43**(6), 1255–1259 (1968)
46. SR Oldfield, SPA Parker, Acuity of sound localization: a topography of auditory space. *Perception* **13**(5), 601–617 (1984)
47. PM Hofman, JGAV Riswick, AJV Opstal, Rearing sound localization with new ears. *Nature Neurosci.* **1**(5), 417–421 (1998)
48. CP Brown, *Modeling the Elevation Characteristics of the Head-Related Impulse Response* (San Jose State University, San Jose, CA, M.S. Thesis, 1996)
49. P Satarzadeh, VR Algazi, RO Duda, *Physical and filter pinna models based on anthropometry* (in Proceedings of the 122nd Conv, Audio Eng. Soc, Vienna, 2007), p. 7098
50. S Hwang, Y Park, Y Park, *Sound Source Localization using HRTF database* (in ICCAS2005, Gyeonggi-Do, 2005)
51. WH Foy, Position-location solution by Taylor-series estimation. *IEEE Trans. Aerosp. Electron Syst.* **AES-12**, 187–194 (1976)
52. JS Abel, JO Smith, Source range and depth estimation from multipath range difference measurements. *IEEE Trans. Acoust. Speech. Signal. Process.* **37**, 1157–1165 (1989)
53. DMY Sommerville, *Elements of Non-Euclidean Geometry* (Dover, New York, 1958). pp. 260–1958
54. LP Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (New York, Dover, 1960), pp. 270–271. originally Ginn, 1909
55. JO Smith, JS Abel, Closed-Form least-squares source location estimation from range-difference measurements. *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-35**, 1661–1669 (1987)
56. JM Delosme, M Morf, B Friedlander, *A linear equation approach to locating sources from time-difference-of-arrival measurements* (in Proceedings of the ICASSP, Denver, CO, 1980), pp. 09–11
57. BT Fang, Simple solutions for hyperbolic and related position fixes. *IEEE Trans. Aerosp. Electron Syst.* **26**, 748–753 (1990)
58. B Friedlander, A passive localization algorithm and its accuracy analysis. *IEEE J. Oceanic Eng.* **OE-12**, 234–245 (1987)
59. HC Schau, AZ Robinson, Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-35**, 1223–1225 (1987)
60. JS Abel, A divide and conquer approach to least-squares estimation. *IEEE Trans. Aerosp. Electron. Syst.* **26**, 423–427 (1990)
61. A Beck, P Stoica, J Li, Exact and approximate solutions of source localization problems. *IEEE Trans. Signal. Process.* **56**(5), 1770–1778 (2008)
62. HC So, YT Chan, FW Chan, Closed-Form Formulae for Time-Difference-of-Arrival Estimation. *IEEE Trans Signal Process* **56**, 6 (2008)
63. MD Gillette, HF Silverman, A linear closed-form algorithm for source localization from time-differences of arrival. *IEEE Signal Process. Lett.* **15**, 1–4 (2008)
64. EG Larsson, D Danev, Accuracy comparison of LS and squared-range LS for source localization. *IEEE Trans. Signal. Process* **58**, 2 (2010)
65. N Ono, S Sagayama, R-means localization: a simple iterative algorithm for range-difference-based source localization. in Proceedings of the ICASSP. **14–19**, 2718–2721 (March 2010)
66. ZE Chami, A Guerin, A Pham, C Servière, *A PHASE-based Dual Microphone Method to Count and Locate Audio Sources in Reverberant Rooms* (in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, 2000), pp. 18–21
67. SV Vaseghi, *Advanced Digital Signal Processing and Noise Reduction* (3rd edn, John Wiley & Sons, Ltd, Hoboken, NJ, 2006)
68. A Pourmohammad, SM Ahadi, TDE-ILD-HRTF-based 2D whole-plane sound source localization using only two microphones and source counting. *Int. J. Info. Eng.* **2**(3), 307–313 (2012)
69. CA Balanis, *Antenna Theory: Analysis and Design* (John Wiley & Sons Inc (NJ, Hoboken, 2005)
70. DC Giancoli, *Physics for Scientists and Engineers*, vol. I (Upper Saddle River, Prentice Hall, 2000)
71. B Friedlander, On the Cramer-Rao bound for time delay and Doppler estimation. *IEEE Trans. Inf. Theory.* **30**(3), 575–580 (1984)
72. A Gore, A Fazel, S Chakraborty, Far-Field Acoustic Source Localization and Bearing Estimation Using $\Sigma\Delta$ Learners. *IEEE Trans. circuits and systems-l: regular papers* **57**(4) (2010)
73. SH Chen, JF Wang, MH Chen, ZW Sun, MJ Liao, SC Lin, SJ Chang, *A Design of Far-Field Speaker Localization System using Independent Component Analysis with Subspace Speech Enhancement* (in Proc. 11th IEEE Int. Symp. Mult., San Diego, CA, 2009), pp. 14–16

doi:10.1186/1687-4722-2013-27

Cite this article as: Parsayan and Ahadi: N-dimensional N-microphone sound source localization. *EURASIP Journal on Audio, Speech, and Music Processing* 2013 **2013**:27.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com