TDE-ILD-HRTF-Based 2D Whole-Plane Sound Source Localization Using Only Two Microphones and Source Counting

Ali Pourmohammad, *Member, IACSIT* and Seyed Mohammad Ahadi

***Abstract*—In outdoor cases, TDOA-based methods with lower accuracy than DOA based methods but fewer microphones and less computation time, are used for 2D wideband sound source localization using only three microphones. Using these methods, outdoor (far-field) high accuracy 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 to reduce the microphones count in indoor 2D sound source localization using TDE and ILD based methods simultaneously. However, these papers do not mention that using ILD-based methods need only one dominant source to be active for accurate localization. Also it is known that using a linear array, two mirror points will be produced simultaneously. This issue means we can localize 2D sound source only in half-plane. In this paper we propose a novel method to have 2D whole-Plane dominant sound source localization using TDE, ILD and HRTF-based methods simultaneously. Based on the proposed method, a special reflector (instead of dummy head) for microphones arrangement is designed and source counting method is used to find that only one dominant sound source is active in the localization area. Simulation results indicate that this method is useful in outdoor and low degree reverberation cases when we try to raise SNR using spectral subtraction and source counting methods.**

***Index Terms*—Sound source localization, TDOA, TDE, PHAT, ITD, ILD, HRTF.**

I. Introduction

Passive sound source localization methods, in general, can be divided into direction of arrival (DOA), time delay of arrival (TDOA) or time difference estimation (TDE) or interaural time difference (ITD), intensity level difference or interaural level difference (ILD) and head related transfer function (HRTF) based methods. DOA-based beamforming and sub-space methods need many more microphones for high accuracy narrowband source localization which they are not applicable to localization of wideband signals sources in far-field cases. ILD-based methods need high accuracy level measurement hardware and need one source to be dominant enough for high accuracy sound source localization. These methods are applicable to the case of only a dominant sound source (high SNR). TDE-based methods with high sampling rate are used for 2D high accuracy wideband near-field and far-field sound source localization. The minimum number of microphones required for 2D positioning is 3 [1][2]. Recently some papers were published which introduce 2D sound source localization method using only two microphones in indoor cases using TDE and ILD based methods simultaneously [3][4]. In this paper we apply this method in outdoor and low reverberation cases for a dominant sound source and evaluate it in different noise powers. Also we propose a novel method to have 2D whole-plane dominant sound source localization using HRTF, TDE and ILD-based methods simultaneously. Based on the proposed method, a special reflector for microphones arrangement is designed and source counting method is used to find that only one dominant sound source is active in the localization area.

The structure of this paper is as follows. Firstly we explain HRTF, ILD and TDE-based methods and remember TDE-based PHAT method. Then we explain sound source angle of arrival and location calculations using ILD and PHAT methods. Then we introduce TDE-ILD-based method to 2D half-plane sound source localization using only two microphones. After introducing source counting method, we propose and simulate our TDE-ILD-HRTF-based method for 2D whole-plane sound source localization. Finally conclusions will be made.

Manuscript received February 20, 2012; revised April 26, 2012.

The authors are with the Electrical Engineering Department, Amirkabir University of Technology, Hafez Ave., Tehran 15914, Iran (emails: pourmohammad@aut.ac.ir; sma@aut.ac.ir)

II. HRTF, ILD and TDE (ITD) Based Methods

A. HRTF-Based Method [5]

Humans have just two ears, but can locate sounds in three dimensions in range and direction. This is possible because the brain, inner ear and the extErnal ears (pinnas) work together to make inferences about location. Humans estimate the location of a source by taking cues derived from one ear, and by comparing cues received at both ears (binaural or difference cues). Among the difference cues are time differences of arrival and intensity differences. The monaural cues come from the interaction between the sound source and the human anatomy, in which the original source sound is modified before it enters the ear canal for processing by the auditory system (Fig.1). These modifications encode the source location, and may be captured via an impulse response which relates the source location and the ear location. This impulse response is termed the head-related impulse response (HRIR). Convolution of an arbitrary source sound with the HRIR converts the sound to that which would have been heard by the listener if it had been played at the source location, with the listener's ear at the receiver location. The HRTF is the Fourier transform of HRIR. Therefore, HRTF describes how a given sound wave input is filtered by the diffraction and reflection properties of the head, pinna, and torso, before the sound reaches the transduction machinery of

the eardrum and inner ear. The pinna is an important object for localizing a sound source in the elevation direction. This is because the pinna spectral features like peaks and notches in HRTFs are caused by the direction dependent acoustic filtering due to the pinna. Notches are created in the frequency spectra when incident wave is cancelled by the reflected wave from the artificial concha (Fig.1).

Biologically, the source-location-specific prefiltering effects of these external structures aid in the neural determination of source location, particularly the determination of the source's elevation. Comparing the original and modified signal, we can compute HRTF which tell us in frequency domain how a sound changes on the way from its source to the listener's ear. In a simple style, if $X_c(k)$ is a Fourier transform of the left or right ear sound source, $Y_c(k)$ is a DFT of the recorded one, HRTF of that signal can be formally found as:

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

In detail:

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

$$\arg H_c(k) = \arg Y_c(k) - \arg X_c(k) \tag{3}$$

$$H_c(k) = |H_c(k)| e^{jarg H_c(k)} \tag{4}$$

Generally, $H_c(k)$ contains all the direction-dependent and direction-independent components (directional transfer function or DTF and common transfer function or CTF, respectively). For the pure HRTF, we have to remove direction-independent elements from $H_c(k)$. Mathematically, if $C_c(k)$ is the known CTF, then DTF $D_c(k)$ can be computed as:

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

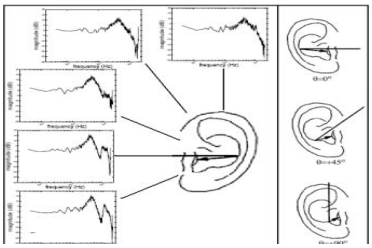


Fig.1. Pinna reflections of sound for different elevations.

B. ILD-Based Method [3][4][6]

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:

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

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

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 method 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:

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

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

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:

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

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

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:

$$E_1 . d_1^2 = E_2 . d_2^2 + \eta, \tag{12}$$

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:

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

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

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

C. TDE-Based Methods [1][2][6]

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:

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

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:

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

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:

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

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:

$$\tau_{12} = argmax_\tau \; R_{s_1 s_2}(\tau). \tag{18}$$

In an overall view, the time delay estimation methods are as follows:

Correlation-based methods:

(Cross-correlation (CC), ML, PHAT, AMDF)

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.

D. TDE-Based PHAT Method [2]

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 as:

$$\Phi_{PHAT}(f) = \frac{1}{|G_{s_1 s_2}(f)|}, \tag{19}$$

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:

$$R_{s_1 s_2}(\tau) = \int_{-\infty}^{+\infty} \Phi_{PHAT}(f).G_{s_1 s_2}(f)e^{j2\pi f\tau}df \tag{20}$$

$$D_{PHAT}(f) = argmax_\tau \ R_{s_1 s_2}(\tau), \tag{21}$$

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

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

where

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

The main reason for using ϕ_{PHAT} is to sharpen the correlation function peak leading to more accurate measurement of the delay. Another advantage of PHAT is its simplicity in implementation. In this method, in fact, the weighting function is the same as a normalizing function that relates the spectrum information to the phase of the spectrum. This method features a much higher accuracy in comparison with other methods when dealing with periodic signals. The reason is that, in this case, the spectrum includes local maxima resulting from periodicity of the signal. As the spectrum is whitened in this method, the delay is estimated without any problem. Another advantage of this method is its high performance in noisy and reflective environments. This is due to the fact that all the frequency components are of equal importance due to the normalization. Obviously, this is only true when the signal-to-reverberation ratio (SRR) remains constant, which is almost correct in real conditions, since the amount of reverberation in any frequency is related to the signal energy in that frequency.

III. ILD and PHAT Based Angle of Arrival and Local Calculations Methods

A. Using ILD-based Method [3][4][6]

Assuming two microphones are on x axis and have a distance of R (R=D/2) from origin (Fig.3), we can rewrite (13) and (14) as:

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

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

Therefore we can rewrite (12) as:

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

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

$$m[(R - x_s)^2 + {y_s}^2] = [(-R - x_s)^2 + {y_s}^2] + n \tag{27}$$

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

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

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

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

Therefore source location is on a circle (Fig.2) with centre coordinate (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 centre coordinate and different radii relative to the first circle. Intersection of the first and second circles gives us source location x and y.

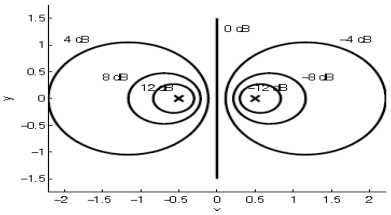


Fig. 2. Isocontours of (30) for R=0.5m and different values of 10 log (m). The sound source lies on a circle unless the two energies are equal, in which case it lies on a line [3].

B. Using PHAT Method [1][2][6]

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

$$r > \frac{2D^2}{\lambda} \tag{31}$$

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 Fig.3. If $s_1(t)$ is the output signal of the first microphone and $s_2(t)$ is that of the second microphone (Fig.3), 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 method 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:

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

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

$$R_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} s(t)s(t - T_D + \tau)\, dt. \quad (33)$$

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

$$\tau = argmax_{T_D} R_{s_1s_2}(\tau). \quad (34)$$

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 Fig.3, the angle of arrival is:

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

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:

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

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

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

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:

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

Simplifying the above equation will result in:

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

where y has hyperbolic geometrical location relative to x, as shown in Fig.3. 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:

$$\begin{cases} \tau_{21} = \frac{\sqrt{(x-x_2)^2+(y-y_2)^2} - \sqrt{(x-x_1)^2+(y-y_1)^2}}{v_{sound}} \\ \tau_{31} = \frac{\sqrt{(x-x_3)^2+(y-y_3)^2} - \sqrt{(x-x_1)^2+(y-y_1)^2}}{v_{sound}} \end{cases} \quad (41)$$

It is noticeable that these are nonlinear 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.

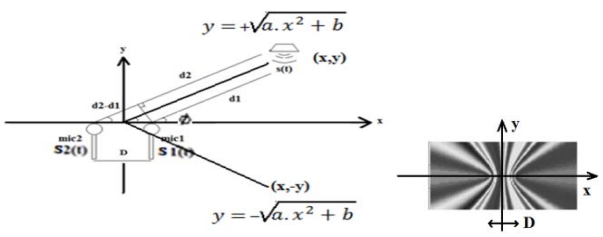


Fig.3. Hyperbolic geometrical location of 2D sound source localization using two microphones

IV. TDE-ILD-Based 2D Sound Source Localization Method [4]

Using only TDE or ILD method to calculate source location (x and y) in 2D cases needs at least three microphones. Using simultaneously both TDE and ILD methods 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:

$$(x_s - x_1)^2 + (y_s - y_1)^2 = (\frac{\tau_{21}.v_{sound}}{1-\sqrt{m}})^2 = r_1^2 \quad (42)$$

and

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

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. Try to obtain a closed form solution to this problem, transforming the expression by:

$$x_1x_s + y_1y_s = \frac{1}{2}(k_1^2 - r_1^2 + R_s^2) \quad (44)$$

and

$$x_2x_s + y_2y_s = \frac{1}{2}(k_2^2 - r_2^2 + R_s^2), \quad (45)$$

where

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

rewrite (44) and (45) into matrix form:

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

results:

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

If we define:

$$p = \begin{bmatrix} p1 \\ p2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (49)$$

and

$$q = \begin{bmatrix} q1 \\ q2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}^{-1} \begin{bmatrix} k_1^2 - r_1^2 \\ k_2^2 - r_2^2 \end{bmatrix} \quad (50)$$

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

$$X = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} q_1 + p_1R_s^2 \\ q_2 + p_2R_s^2 \end{bmatrix} \quad (51)$$

Insert (46) into (51), the solution to Rs is obtained as:

$$R_s^2 = \frac{O_1 \pm O_2}{O_3} \quad (52)$$

where:

$$O_1 = 1 - p_1q_1 + p_2q_2,$$
$$O_2 = \sqrt{(1 - p_1q_1 + p_2q_2)^2 + (p_1^2 + p_2^2)(q_1^2 + q_2^2)},$$
$$O_3 = p_1^2 + p_2^2.$$

The positive root gives the square of distance from source to origin. Substituting Rs into (51), the final source

coordinate will be obtained. 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 (Fig.3), According to (49) and (50), we cannot find p and q. Therefore we cannot consider such a microphones arrangement. However, using this microphones arrangement simplifies equations. According to (26) and (37) we can intersect circle and hyperbola (Fig.4) 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:

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

Therefore 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 issue means we can localize 2D sound source only in half plane.

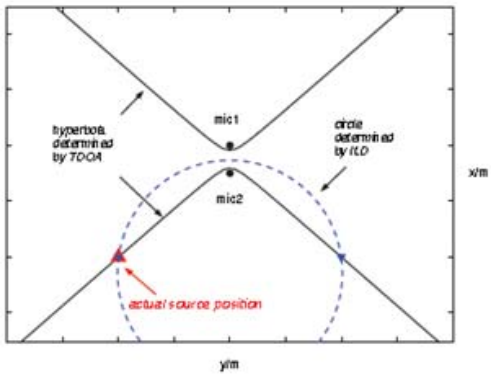


Fig.4. Intersection of circle and hyperbola which both of them conclude source location [4]

V. Proposed TDE-ILD-HRTF Method

Using TDE-ILD-based method, dual microphone 2D sound source localization is applicable. But it is known that, by using a linear array in TDE-ILD-based method, two mirror points will be produced simultaneously (half-plane localization in Fig.4) [4].

Also, it is noticeable that using ILD-based method needs only one dominant 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 spectrum subtraction, and HRTF methods. According to previous discussions, ILD-based method needs to use source counting to find that one dominant source is active for high resolution localizing. If more than one source is active in localization area, it cannot calculate $m = \frac{E_1}{E_2}$ 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. Using PHAT method 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. We consider only one source signal to be a periodic signal as:

$$s(t) = s(t + T) \quad (61)$$

If the signals window is greater than T, calculating cross correlation between the output signals of the two microphones give 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:

$$s_1(t) = s(t - T_1) + s`(t - T`_1) \quad (62)$$

$$s_2(t) = s(t - T_2) + s`(t - T`_2) \quad (63)$$

According to (32) we have:

$$R_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty}\big(s(t - T_1) + s`(t - T`_1)\big)\big(s(t - T_2 + \tau) + s`(t - T`_2 + \tau)\big)\,dt \quad (64)$$

$$R_{s_1s_2}(\tau) = R1_{s_1s_2}(\tau) + R2_{s_1s_2}(\tau) + R3_{s_1s_2}(\tau) + R4_{s_1s_2}(\tau) \quad (65)$$

where:

$$R1_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} s(t - T_1).s(t - T_2 + \tau)\,dt \quad (66)$$

$$R2_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} s(t - T_1).s`(t - T`_2 + \tau)\,dt \quad (67)$$

$$R3_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} s`(t - T`_1).s(t - T_2 + \tau)\,dt \quad (68)$$

$$R4_{s_1s_2}(\tau) = \int_{-\infty}^{+\infty} s`(t - T`_1).s`(t - T`_2 + \tau)\,dt \quad (69)$$

Using (34), $\tau_1 = T_2 - T_1$ gives us a maximum value of $R1_{s_1s_2}(\tau)$, $\tau_2 = T`_2 - T_1$ gives us a maximum value of $R2_{s_1s_2}(\tau)$, $\tau_3 = T_2 - T`_1$ gives us a maximum value of $R3_{s_1s_2}(\tau)$ and $\tau_4 = T`_2 - T`_1$ gives us a maximum value of $R4_{s_1s_2}(\tau)$. Therefore we will have four peak values in cross correlation vector. But according to this fact that (66) and (69) are cross correlation functions of a signal with delayed version of itself, and (67) and (68) are cross

correlation functions of two different signals, τ_1 and τ_4 respect maximum values are dominant respect to τ_2 and τ_3 respect values. Now we conclude in two dominant sound sources area, cross correlation vector will have two dominant values and therefore for more than two dominant sound sources signals. Therefore counting dominant cross correlation vector values, we can find the number of active and dominant sound sources in localization area.

Using ILD-based method in TDE-ILD-based dual microphone 2D sound source localization method constraints to use source counting to find that one dominant high SNR source is active in localization area. Source counting method was proposed to calculate active source numbers in localization area. Also spectrum subtraction method is usable for noise reduction and raising alone active source's SNR. If the input signal sampling rate is 96 kHz, the bandwidth of the input signal should be limited to 48 kHz. Since no anti-aliasing analog low-pass filter with 48 kHz cut-off frequency was available, aliasing will cause higher frequency (more than 48 kHz) components to cause some distortion in the lower frequencies. Also, according to the background noise, such as wind, rain and babble sound signals, we can consider a background spectrum estimator.

By using a linear array in TDE-ILD-based dual microphone 2D sound source localization method, two mirror points will be produced simultaneously (Fig.4). Adding HRTF method, whole plane dual microphone 2D sound source localization is applicable. The scattering of incident sound wave by the pinna cues spectral notches. Notch is created in the frequency spectra when incident wave is cancelled by the reflected wave from the concha wall. The position of the notch varies linearly with elevation. Researchers use an artificial ear that has a spiral shape. This is because a spiral shaped artificial ear is a special type of pinna that can vary the distance from a microphone placed in the centre of the spiral to a concha wall linearly according to the sound direction. But we consider a half-cylinder instead of artificial ear. Due to using such reflector, constant notch position is created for all variation of sound source angle of arrival in front of reflector. Of course the reflector scatters the sound source waves which are in back of it. Therefore we consider some circular slits in half-cylinder's surface (Fig.5).

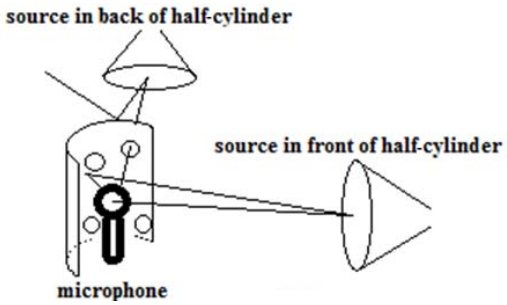


Fig.5. Used half-cylinder instead of artificial ear for 2D cases

If d is the distance between the reflector (half-cylinder) and the microphone (is placed in centre), a notch will create when it is equal to quarter of the wavelength of the sound λ plus any multiple of λ/2. For these wavelengths, the incident waves are cancelled (reduced) by reflected waves:

$$n.\left(\frac{\lambda}{2}\right)+\left(\frac{\lambda}{4}\right)=d \qquad n=0,1,2,3\ldots \tag{70}$$

These notches will appear at the following frequencies:

$$f=\frac{c}{\lambda}=\frac{(2n+1).c}{4d} \tag{71}$$

Covering only microphone 2 in Fig.3 by reflector, calculating the Interaural spectral difference gives:

$$|\Delta H(f)|=|10\log_{10}H_1(f)-10\log_{10}H_2(f)|=\left|10\log_{10}\frac{H_1(f)}{H_2(f)}\right| \tag{72}$$

Valuable $|\Delta H(f)|$ indicates that sound source is in front. Also negligible value indicates that sound source is in back. Of course need to well design of circular slits to have same frequency spectrum in both microphones when sound source is in back. Using half-cylinder or others shape of reflectors decreases accuracy of time delay and intensity level deference estimation between microphones 1 and 2 due to changing spectrum of second microphone's signals. Multiplying inverse function of notch-filter in second microphone's spectrum increases accuracy.

VI. The Proposed Method's Algorithm

According to the discussion in previous sections, we can consider the following steps for our proposed method:

- Setup of the microphones, reflector and hardware
- Calculating the sound recording hardwares set (microphone, preamplifier and sound card) amplification normalizing factor
- Obtain $s_1(t)$ and $s_2(t) \rightarrow m=\frac{E_1}{E_2}$
- Remove DC from the signals and Normalize the signals regarding the sound intensity
- Window signals regarding their periods or their stationary parts (for example at least about 100ms for wideband quasi-periodic helicopter sound or twice that)
- Hamming windowing
- Noise cancelation in real world applications (e.g. using spectral subtraction and band-pass filtering)
- Apply PHAT to the signals in order to calculate τ_{21} in frequency domain (index of first maximum value of cross correlation vector in time domain)
- Finding second maximum value of cross correlation vector.
- If the first maximum value is not enough dominant with respect to the second maximum value, go to the next widows of signals and do not calculate sound source location. else:

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

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

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

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

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

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

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

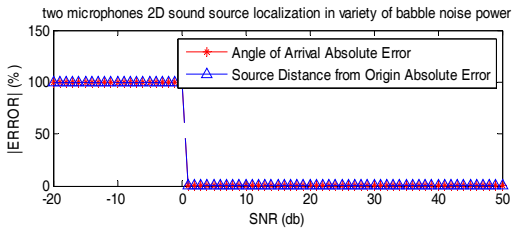
VII. Simulations and Discussion

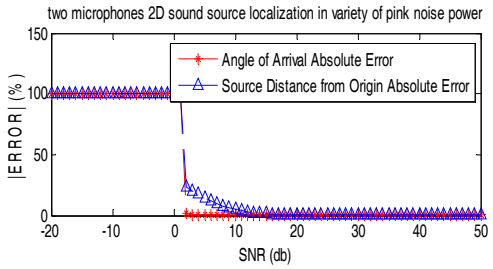
In order to use the introduced method for sound source localization in low reverberant outdoor cases, we simulated. We tried to evaluate the accuracy of this method in noise-free environment, and a variety of SNRs for some environmental noises. We considered two microphones on x axis ($y_1=y_2=0$) with one meter distance from origin ($x_1=1, x_2=$

-1 (R = 1)) (Fig.3) and half-cylinder reflector for second microphone. In order to use PHAT for the calculation of time delay between the signals of the two microphones, we downloaded a wave file with a length of approximately four seconds of helicopter sound (wideband and quasi-periodic signal) from internet as our sound source. For different source locations and for an ambient temperature of 15 degrees Celsius, first we calculated sound speed in air using (73).

$$v_{sound} = 20.05\sqrt{273.15 + Temperature(Centigrade)} \quad (73)$$

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. According to time delay positive values (sound source nearer to the first microphone (mic1 in Fig.3)), we delayed second microphone signal with respect to the first microphone signal, and for the time delay negative values (sound source nearer to the second microphone (mic2 in Fig.3)) delayed the first microphone signal with respect to the second microphone signal. 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, using the introduced method, we tried to calculate source location in noise free environment and a variety of SNRs for some environmental noises as follows: For a variety of sound signal to noise ratios (SNRs) for white Gaussian, pink and babble noises from "NATO RSG-10 Noise Data", 16 bit Quantization and 96000 Hz sampling frequency, simulation results are shown in Fig.6 for source location (x = 10, y = 10). Simulation Results show more localization error under SNR 10dB. This issue occurs due to using ILD.





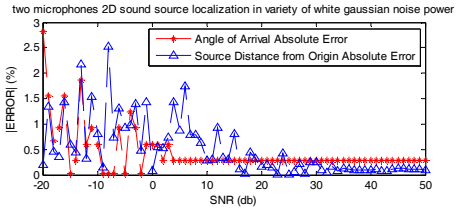


Fig. 6. Simulation results for a variety of SNRs in presence of babble, pink and white Gaussian noises.

VIII. Conclusions

In this paper, we simulated spectral subtraction and source counting methods for proposed TDE-ILD-HRTF-based 2D sound source localization using only two microphones method for low degree reverberation outdoor cases. Simulation results show accuracy in source location measurement in comparison with similar researches [4] and [6] which did not use spectral subtraction and source counting methods. Also indicate that covering one of the microphones by a half-cylinder reflector leads us to have whole-plane 2D sound source localization.

References

[1] M. S Brandstein, J. E. Adcock, H. F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Transactions on Speech and Audio Processing*. vol. 5, no. 1, pp. 45-50. Jan. 1997.

[2] P. Svnizer, M. Matnssoni, and M. Omologo, "Acoustic Source Location in a THREE-Dimensional Space Using Crosspower Spectrum Phase," *IEEE, ICASSP-97*, pp. 231-234, 1997.

[3] T. S. Birchfield and R. Gangishetty, "Acoustic Localization by Interaural Level Difference," in *Proc. ICASSP2005*, pp. iv/1109–iv/1112, Mar. 2005.

[4] W. Cui, Z. Cao, and J. Wei, "DUAL-Microphone Source Location Method in 2-D *Space,"* in *Proc* .pp. IV845-848, 2006.

[5] C. I. Cheng and G. H. Wakefield, "Introduction to Head-Related transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space," *Journal of the Audio Engineering Society,* vol. 49, no. 4, pp.231-248, 2001.

[6] N. Ikoma, O. Tokunaga, H. Kawano, and H. Maeda," Tracking of 3D Sound Source Location by Particle Filter with TDOA and Signal Power Ratio," *ICROS-SICE International Joint Conference*, pp. 18-21, 2009.

Ali Pourmohammad 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 also teaches several courses (C++ programming, multimedia systems, Microprocessor Systems, digital audio processing, digital image processing and digital signal processing I & II). His research interests include digital signal processing and applications (audio and speech processing and applications, digital image processing and applications, sound source localization, sound source separation (determined and under-determined blind source separation (BSS), Audio, Image and Video Coding, Scene Matching and ...) and multimedia and applications.