Sound Source Localization using HRTF database

Sungmok Hwang*, Youngjin Park and Younsik Park

* Center for Noise and Vibration Control, Dept. of Mech. Eng., KAIST, Deajeon, Korea
(Tel: +82-42-869-3060, Email: tjdahr78@kaist.ac.kr)

Abstract: We propose a sound source localization method using the Head-Related-Transfer-Function (HRTF) to be implemented in a robot platform. In conventional localization methods, the location of a sound source is estimated from the time delays of wave fronts arriving in each microphone standing in an array formation in free-field. In case of a human head this corresponds to Interaural-Time-Delay (ITD) which is simply the time delay of incoming sound waves between the two ears. Although ITD is an excellent sound cue in stimulating a lateral perception on the horizontal plane, confusion is often raised when tracking the sound location from ITD alone because each sound source and its mirror image about the interaural axis share the same ITD. On the other hand, HRTFs associated with a dummy head microphone system or a robot platform with several microphones contain not only the information regarding proper time delays but also phase and magnitude distortions due to diffraction and scattering by the shading object such as the head and body of the platform. As a result, a set of HRTFs for any given platform provides a substantial amount of information as to the whereabouts of the source once proper analysis can be performed. In this study, we introduce new phase and magnitude criteria to be satisfied by a set of output signals from the microphones in order to find the sound source location in accordance with the HRTF database empirically obtained in an anechoic chamber with the given platform. The suggested method is verified through an experiment in a household environment and compared against the conventional method in performance.

Keywords: Sound source localization, Head-Related-Transfer-Function (HRTF), Phase criterion, Magnitude criterion

1. Introduction

The sound source localization is about finding the whereabouts of a sound source using measurements from a number of microphones. The studies for developing localization model have a long history and many researchers have studied different methods for sound source localization. These days, mobile robot technology is gaining much attention in many application fields. The sound localizing ability of a robot is essential for human-robot communication and interaction. A robot operating in a household environment should detect diverse sound events and take notice of them to achieve robust recognition and interaction with user. So, sound source localization can be said to be one of the cores of the robot technology.

ITD (Interaural Time Delay) plays an important role in most conventional methods for localization. Although many different sound source localization methods such as beamforming[1], spatial spectrum[2], biological cues are developed, the ITD method[3] is one of the most popular methods in practical applications. ITD indicates the time delay between two microphones when acoustic waves emitted from a sound source reach each microphone. The ITD method estimates time delay and localizes the sound source with free-field assumption. Although ITD is an excellent sound cue in stimulating a lateral perception on the horizontal plane, confusion is raised when estimating the sound source location from ITD alone because many positions sharing the same ITD in 3-dimensional space can exist[4]. The ITD method also assumes that microphones are placed in free-filed, but this assumption is not valid for an actual platform used in real environments. For example, microphones are embedded in the robot head. Therefore the phase and magnitude of signals are distorted due to diffraction and scattering by the shading object such as the head and body of the platform.

HRTF (Head-Related-Transfer-Function) summarizes the direction dependent acoustic filtering which a free-filed sound undergoes due to the head, torso, shoulder and pinna[5]. HRTF associated with a dummy head microphone system or a robot platform with several microphones contains not only the information regarding proper time delays but also phase and magnitude distortions. So, we propose a new localization method using HRTF database empirically obtained in an anechoic chamber with a given platform. Performance of the proposed method is shown through experiments carried out in an anechoic chamber and a household environment. In addition, appropriate filtering method of noise existing in daily environment is proposed and the result is shown.

2. HRTFs for Dummy head

In this paper, we apply the proposed method to the B&K HATS. First, we took measurements and constructed the HRTF database of the dummy head with azimuth varying from 0° to 180° and elevation from -30° to 90° in an anechoic chamber. The sampling frequency was $44.1kHz$. The HRTFs were calculated by dividing the pressure at each ear by the free-field pressure at the center of the head. Figure 1 shows the HRTFs for horizontal plane sources from $2m$. When the source is placed at the very front and back of the head, the left and right ear HRTFs are the same due to symmetry of the head. As the source moves in a counterclockwise direction, the magnitude of the left ear HRTF increases and that of the right ear decreases due to the shadowing effect. However, when the source is located right in front of the left ear, the so-called “bright spot” occurs at the right ear: All the waves propagating around the head arrive at the right ear in phase resulting in a slight magnitude boost. As HRTFs vary according to change of azimuth, HRTFs also vary with change of elevation. Figure 2 shows magnitude of the left ear HRTF for median plane (azimuth=0°) sources from $1m$. The main causes of variety are diffraction, reflection, and scattering by the torso, shoulder, and pinna.

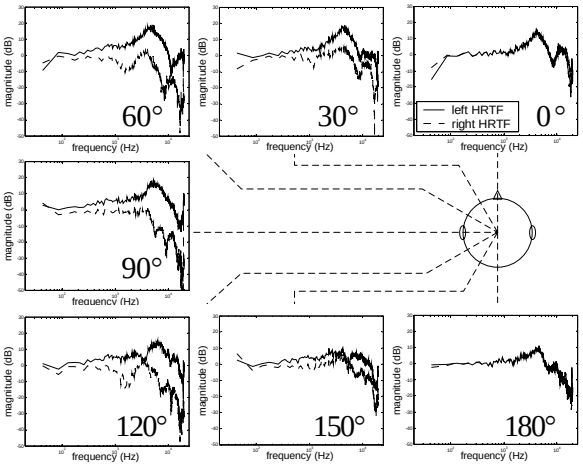

Fig. 1 Magnitude of HRTF for horizontal plane

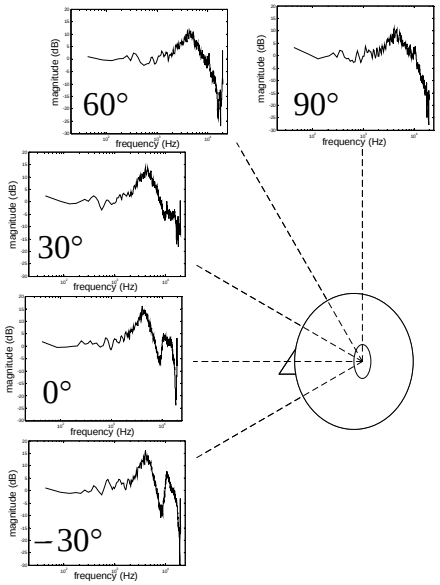

Fig. 2 Magnitude of left ear HRTF for median plane

3. Localization Cues

A set of HRTFs for any given platform provides a substantial amount of information about whereabouts of the sound source. In this study, we introduce new phase and magnitude criteria to be satisfied by a set of output signals from a dummy head microphone system in order to find the sound source location in accordance with the HRTF database empirically obtained in the anechoic chamber. The phase and magnitude criteria are defined as follows.

$$e_{phase} = \int_{\omega} \gamma_{RL}^2 \left\{ \theta_{RL}(\omega) - \theta_{HRTF}(\omega) \right\}^2 d\omega \qquad (1)$$

$$e_{mag} = \int_{\omega} \gamma_{RL}^2 \left\{ M_{RL}(\omega) - M_{HRTF}(\omega) \right\}^2 d\omega \qquad (2)$$

γ_{RL}^2 : *coherence between right and left ear outputs*

θ_{RL} : *phase differnece between right ear and left ear outputs*

M_{RL} : *magnitude ratio between right ear and left ear outputs*

θ_{HRTF} : *phase differnece between right ear HRTF and left ear HRTF*

M_{HRTF}: *magnitude ratio between right ear HRTF and left ear HRTF*

If the noise can be ignored, the phase difference and magnitude ratio between the two ear outputs can be directly obtained from the HRTFs, corresponding to the actual location of sound source. So, we can detect the sound direction by finding the θ_{HRTF}, M_{HRTF} set minimizing the phase and magnitude criteria and this set directly corresponds to the actual location of the source. Coherence function can be used a weighting function. It is a measure of evaluating linear relationship between the two signals and represents how much uncorrelated noise contaminates the signals. As a result, we can reduce the uncorrelated noise effects by using the coherence function as a weighting function.

4. Experiment in an anechoic chamber

4.1 Azimuth estimation

Figure 3 and Figure 4 show the calculated phase and magnitude criteria on horizontal plane with varying azimuth of an actual sound source in an anechoic chamber and the voice frequency band (VFB, i.e. $300\,Hz \sim 4000\,Hz$) is used for calculation. For comparison with the ITD method, the phase criterion using free-field data, calculated from eq. (1) with replacing θ_{HRTF} by θ_{ff}, is also shown. θ_{ff} is the phase difference between the two ears under the free-field assumption and it can be analytically calculated as follows.

$$\theta_{ff}(f) = 2\pi f \tau, \quad \tau : ITD \qquad (3)$$

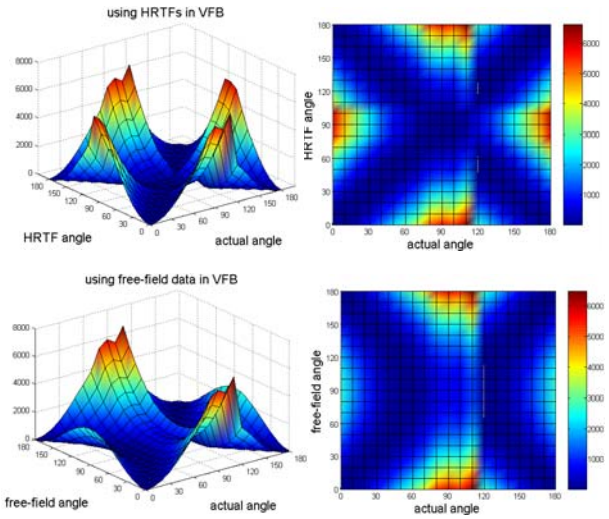

Fig. 3 Phase criterion for azimuth estimation

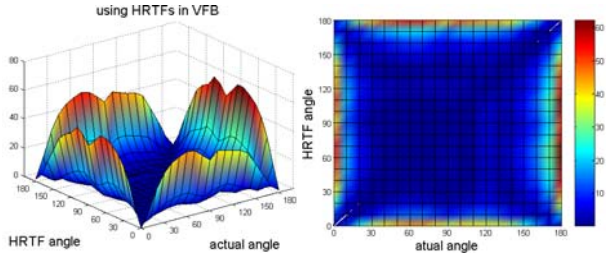

Fig. 4 Magnitude criterion for azimuth estimation

From Figure 3, it can be seen that the phase criterion calculated from HRTFs corresponding to the actual source location has a minimum value at the true angle. However, there is another HRTF angle, which is almost symmetric

position about the interaural axis, making the phase criterion low. This means that the front-back confusion results from the phase assessment alone and this confusion also appears in the free-field result. In general, the estimation performance using the HRTF is better than that using the free-field data. Specially, as the source leans toward one ear, the scattering and diffraction effects for the hidden ear due to the head are most dominant. As a result, the phase criterion is faint in the free-field result. On the other hand, it is clear in the result using the HRTF data because HRTFs contain the information about the scattering and diffraction due to the shading object. In Figure 4, the magnitude criterion doesn't give sufficient information about the source location. Although low values of criterion are faintly shown, we can not determine the azimuth of the actual source.

From the results, it can be said that the estimation performance for azimuth localization based on the HRTFs is better than the performance under the free-field assumption. And the phase criterion is more useful for azimuth localization than the magnitude criterion.

4.2 Elevation estimation

Figure 5 shows the phase and magnitude criteria calculated in the voice frequency band with varying elevation of a sound source from -30° to 30° at some selected azimuth. According to the result of the phase criterion, the criterion has low value not only at the HRTF angle corresponding to the true angle but also at the symmetric position to that. It can be said that up-down confusion is generated as the front-back confusion occurred in the azimuth localization case. However, the shape is somewhat different from that of azimuth due to the vertical asymmetry of the dummy head about the interaural axis. The both sides about the median plane are almost symmetric whereas, the upper and lower halves of the dummy head are asymmetric about the horizontal plane. The magnitude criterion contains this asymmetry, therefore the magnitude criterion has the minimum value at the HRTF angle corresponding to the true angle without confusion in estimation.

As a result, for elevation estimation, it can be said that the magnitude criterion is more useful than the phase criterion.

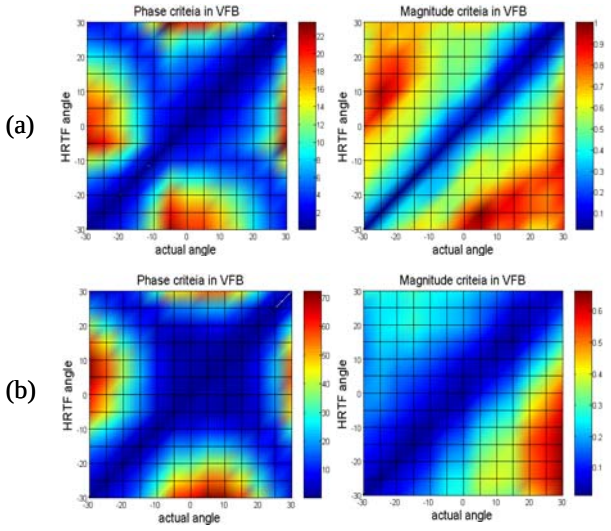

Fig. 5 Phase and magnitude criteria
(a) azimuth = 30° , (b) azimuth = 60°

5. Experiment in a household environment

Earlier works are conducted in an anechoic chamber and this means that we have no regard for noise. However, in daily environment, many noise sources such as the ambient noise, reflection, reverberation exist. Here, the proposed method is verified in a household environment and the method to reduce the noise effects is introduced in the following.

5.1 Azimuth estimation

Figure 6 shows the experimental results. For verification of the proposed method, the result based on the conventional method which uses ITD calculated by GCC (Generalized Cross-Correlation method) with the free-field assumption is shown. In the range of azimuth from about 60° to 90° , the ITD in the free-field exceeds the maximum value, $\mathrm{ITD}_{\max}$. When the distance between the two ears is $2a$, $\mathrm{ITD}_{\max}$ is determined by

$$\mathrm{ITD}_{\max} = \frac{2a}{c}, \quad c : \text{speed of sound} \tag{4}$$

As a result, the ITD method fails to estimate an accurate sound source location and this arises due to the noise contaminating the cross-correlation between the two ear outputs. On the other hand, the proposed method based on the phase criterion can detect the azimuth in general without noise filtering. However, there is an error in estimation.

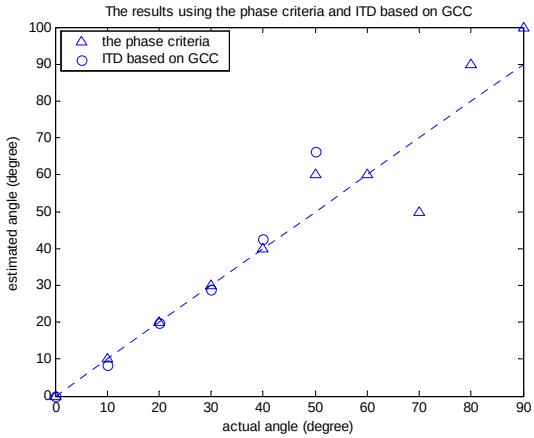

Fig. 6 Azimuth estimation results

5.2 Elevation estimation

Table 1 shows the experimental result for elevation localization at some selected azimuth. The estimation performance is poor because the magnitude ratio between the two ear outputs is contaminated by background noise, reflection from household goods and secondary sound sources. An example of this contaminating effect is shown in Figure 7. Ideally, or in the anechoic chamber, the magnitude ratio between the two ear outputs should match one of the magnitude ratios between the two ear HRTFs and this HRTF set corresponds to the sound source location. However, if noise exists, the information about the magnitude of pure output signals becomes inaccurate.

Table 1 Elevation estimation results

Elevation (degree)	Azimuth (degree)		
	30	60	90
-10	20	-10	10
0	-10	-10	10
10	-10	0	10
20	-10	20	10

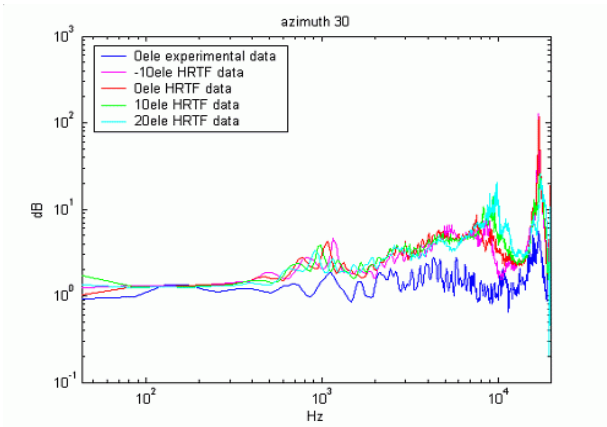

Fig. 7 An example of contaminated magnitude ratio

Up to the present, the experimental results of the proposed method without noise filtering are shown. Although the estimation performance in an anechoic chamber is good, the performance is poor in a household environment due to noise effects. As a result, for precise localization in the real world, noise reduction is necessary and this has direct relation to the estimation performance.

5.3 Filtering of HRIR

For noise reduction, we propose a filtering of HRIR (Head Related Impulse response). HRIR is a time domain version of HRTF and it can be obtained by the inverse Fourier transform of HRTF. Figure 8 shows an example of measured HRIR in a household environment. In this figure, the small ripples representing the background noise in the room and reflections from household furniture can be observed by the irregular shape of the second peak, third peak and so on. However, the part related with the first peak is almost uncontaminated to background noise or reflections, so it can be said that this part directly reflects the pure effect by the actual sound source. Thus we can get rid of the noise effect and obtain uncontaminated HRIRs by applying a filtering as shown in Figure 8 and the length, having the unity value as its magnitude, corresponds to the meaningful length of the first peak part in the measured HRIR in an anechoic chamber. By taking the fast Fourier transform to this filtered HRIR again, we can get a filtered HRTF which is almost noise-free.

In Figure 9, some filtered HRTFs are shown with HRTFs in a household environment and an anechoic chamber. Through the filtering, small ripples in the magnitude of HRTF are smoothed out, thus the filtered HRTF almost agrees with that of the anechoic chamber. In addition, the phase of filtered HRTF is almost the same with that from the anechoic chamber and the filtering overcomes the problem that experimental HRTF phase is different from that of the anechoic chamber's HRTF due to incompleteness of unwrap. It can be said that this incompleteness arises from the noise, thus the absolute value of phase in a household environment is not in accordance with that in the anechoic chamber although the group delays, which mean the gradient of phase, are almost the same except at several frequencies that noise seriously distorts. However, completely unwrapped phase can be obtained by the filtering and the phase of filtered HRTF closely follows the anechoic chamber data.

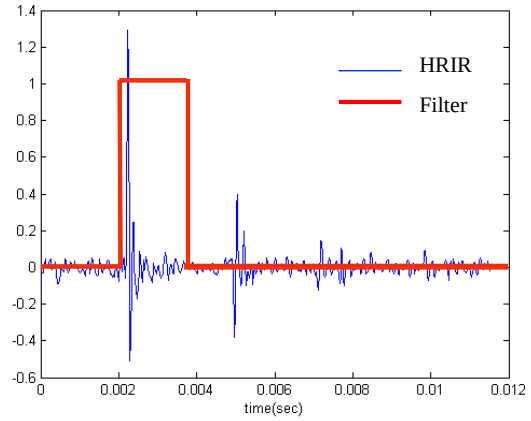

Fig. 8 HRIR filtering

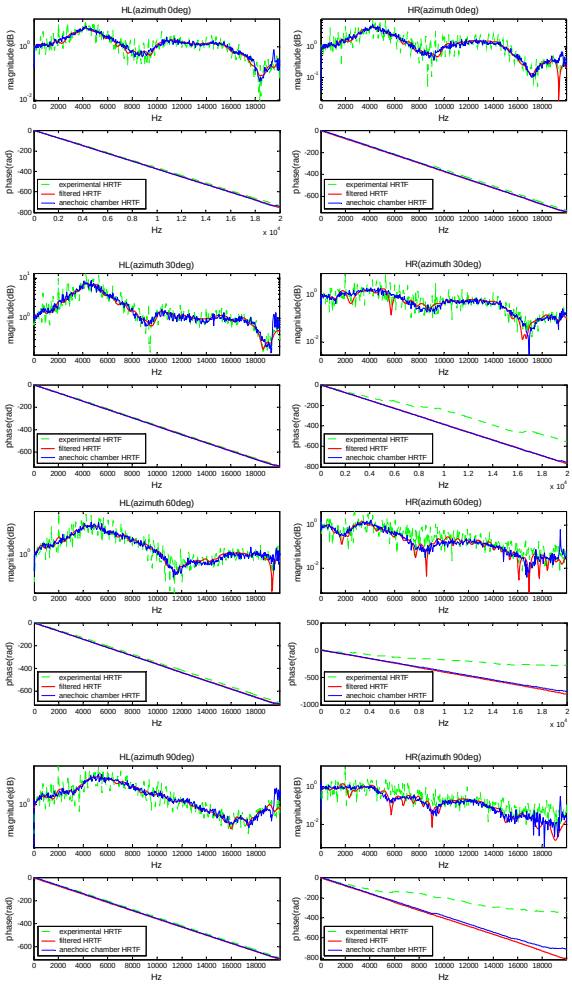

Fig. 9 Magnitude and phase of HRTFs

5.4 Localization using the filtered HRTFs

By filtering the HRIR, we can obtain the filtered HRTF almost uncontaminated to noise and apply the proposed localization algorithm using the phase and magnitude criteria based on the filtered HRTF. Figure 10 shows the experimental result for azimuth localization on the horizontal plane. As mentioned before, azimuth is estimated based on the phase criterion.

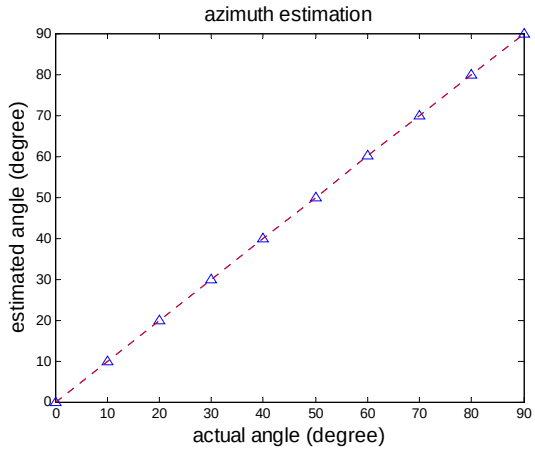

Fig. 10 Azimuth estimation results using the filtered HRTF

From the result, it is clear that the proposed method using the filtered HRTF has the ability of precise azimuth estimation. However, since the HRTF database is obtained on the horizontal plane from 0° to 180° with 10° increment, the above result does not indicate that we can find the source location within an accuracy of several degree. That is to say, we can localize the sound source within an accuracy of about 10° and if we construct the HRTF database with less increment, the resolution of estimation can be increased.

The result of elevation experiment is shown in Table 2. Elevation testing was performed for -10, 0, 10, 20° with the azimuth sets of 30, 60, 90° , respectively. For localization, the magnitude criterion is used.

Table 2 Elevation estimation results using the filtered HRTF

Elevation (degree)	Azimuth (degree)		
	30	60	90
-10	-10	-10	-10
0	10	0	10
10	20	10	10
20	20	20	20

When comparing with the Table 1, which represents the results of elevation estimation without the noise filtering, the estimation performance is improved. Although about 10° estimation error exists, we can estimate the elevation of the sound source approximately and distinguish the ups and downs of the source position. For your information, conventional methods such as the ITD method cannot find the azimuth and elevation of a sound source simultaneously by using two microphones. Above result, however, shows that the proposed method can find both azimuth and elevation by using only two microphones.

6. Conclusion

In this paper, we describe a sound source localization method using HRTF database. The phase difference and magnitude ratio between the two microphones are good localization cues and the HRTF contains information about that. Based on this, we propose two localization cues which are the phase and magnitude criteria and show experimental results using these cues in an anechoic chamber. Experimental results in a household environment are also shown. Although the estimation performance in the anechoic chamber is good, the performance is poor in the household environment due to the noise effects such as reflection, background noise, and additional sources.

For reducing the noise effects, we propose a filtering of HRIR and this yields the filtered HRTF. Although the filter structure is simple, by using this filtering we can get appropriate filtered HRTF. Based on these filtered HRTF, we apply the proposed localization algorithm in the household environment and the estimation performance is improved. When using only two microphones, the conventional method cannot find the azimuth and elevation simultaneously, however, the proposed method which uses the HRTF database can overcome this problem.

In the proposed method, we should know information about the free-field pressure since HRTF means the ratio of the surface pressure to the free-field pressure. However, in practical application, measuring the free-field pressure is not easy, so we will deal with the method which can localize the sound source without the information about the free-field pressure. This is left for future work.

REFERENCES

[1] M. Wax & T. Kailath, “Optimum localization of multiple sources by passive array,” *IEEE Tran. On Acoustics, Speech and Signal Processing*, vol. 31, no.5, pp. 1210-1217, Oct, 1983.

[2] R. Schmitdt, “A signal subspace approach to multiple emitter location and spectral estimation,” Ph. D thesis, Stanford University, Stanford, CA, USA, MUSIC, 1981.

[3] M. S. Brandstein & H. F. Silverman, “A robust method for speech signal time-delay estimation in reverberation rooms,” *Proc. ICASSP-97*, vol. 1, pp. 375-378, April, 1997.

[4] C. I. Cheng & 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.

[5] R. O. Duda & W. L. Martens, “Range dependence of the response of a spherical head model,”, *Journal of Acoustic Society of America*, 104 (5), November, 1998.