SOUND LOCALIZATION IN A SOUND FIELD REPRESENTED BY SPHERICAL HARMONICS

Amir Avni, Boaz Rafaely

Department of Electrical and Computer Engineering
Ben-Gurion University of the Negev
Beer-Sheva, 84105,Israel
{amirav, br}@ee.bgu.ac.il

ABSTRACT

Sound localization is an important aspect in human hearing, with interaural time difference (ITD) and interaural level difference (ILD) as two binaural cues that complement each other in this goal. Head related transfer functions (HRTFs) can be used to estimate and analyze these cues, where an HRTF representation in the spherical harmonics domain has recently been suggested for analysis. The aim of this paper is the study of the effect of incomplete representation of the sound field in the spherical harmonics domain on the cues of sound localization. An objective study is presented on the relation between the sound field order and the errors in sound localization cues of ITD and ILD.

1. INTRODUCTION

According to Lord Rayleigh's 'duplex theory' [1], interaural time difference (ITD) and interaural level difference (ILD) are two binaural cues that complement each other for sound localization. ITD is the time difference between the sound measured at the left ear and the sound measured at the right ear; it is usually effective at low frequencies where the wavelength is larger than the head. ILD, on the other hand, is the difference in dB between the magnitude of the sound pressure measured at the left ear and the sound pressure measured at the right ear, which is mostly affected by head shading; hence ILD is mostly noticeable at high frequencies. ITD and ILD can be estimated objectively using head-related transfer functions (HRTFs), or subjectively by listening experiments. Several methods have been suggested to estimate ITD using HRTF and head related impulse responses (HRIRs). Kistler and Wightman [2] suggested computing the cross correlation between the left HRIR and the right HRIR and determining ITD when the cross correlation is maximized. Jot et al. [3] and Huopaniemi and Smith [4] suggested linear fitting of the excess phase to estimate ITD by the time-of-arrival for both HRIRs. Nam et al. [5] suggested computing the delay of the HRIR for each ear using a maximization of the cross correlation between the HRIR and the HRIR minimum phase. ILD is easier to compute using the ratio of the HRTFs magnitude; however it is frequency dependent and generally increases as frequency increases; therefore it is common to estimate ILD for a certain frequency [6] or in 1/3 octave bands [7].

Spherical microphone arrays have been recently studied both theoretically and experimentally. One application of spherical microphone arrays is for spatial sound recording and sound reproduction, where the sound field in a given location is represented by spherical harmonics. The sound field, typically captured by a finite number of microphones, has incomplete representation in the spherical harmonic domain, due to the finite number of spherical harmonics coefficients [8].

The effect of incomplete sound field representation, denoted by the maximum order in the spherical harmonics domain, on the spatial attributes of the sound field, has been studied by Rafaely and Avni [9] [10]. In these studies, the Interaural Cross Correlation (IACC) was used to quantify the spatial behavior of the sound field. Good accuracy of IACC computation has been shown for a sufficient order of spherical harmonics representation, where the order was directly related to the maximum frequency of the selected sound field.

The following paper examines the effect of a reconstructed HRTF database on ITD and ILD, where the reconstruction is done using a finite order N in the spherical harmonics domain. The error of the binaural cues estimation is analyzed where for each cue a finite order N is suggested in order to diminish the error below the cue's just noticeable differences (JNDs).

2. PLANE WAVES AND SPHERICAL FOURIER TRANSFORM

Consider a sound pressure function $p(k, r, \theta, \phi)$, with (r, θ, ϕ) the standard spherical coordinate system, which is square integrable over (θ, ϕ), with $k = \frac{2\pi f}{c}$ the wavenumber, where f is the frequency and c is the speed of sound. Its spherical Fourier transform (SFT), $p_{nm}(k, r)$, and the inverse spherical Fourier transform (ISFT) are defined [11] by:

$$p_{nm}(k,r) = \int_0^{2\pi} \int_0^{\pi} p(k,r,\theta,\phi) Y_n^{m^*}(\theta,\phi) \sin\theta d\theta d\phi \quad (1)$$

$$p(k,r,\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} p_{nm}(k,r) Y_n^m(\theta,\phi) \quad (2)$$

where the spherical harmonics are defined by:

$$Y_n^m(\theta,\phi) \equiv \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{im\phi} \quad (3)$$

with n the order of the spherical harmonics and P_n^m is the associated Legendre function. The main advantage of the analysis of the sound fields in the spherical harmonics domain is the ability to decompose the pressure function into plane waves [12], thereby estimating the number and amplitudes of the plane waves composing the sound field. The analysis usually requires a spherical microphone array, configured around an open or a rigid sphere. When using a finite number of microphones, a finite number of spherical harmonics order will be analyzed leading to measurement errors that depend on the spherical harmonics order, the number and locations of the microphones (samples over the sphere), and the maximum frequency of the sound field [8]. In addition, a rigid sphere may provide an approximation for a human head, and therefore can be useful when analyzing the expected sound field with the presence of a listener.

3. HRTF REPRESENTATION IN THE SPHERICAL HARMONICS DOMAIN

Head related transfer function (HRTF) analysis using spherical harmonics has been previously presented [13] [9]. HRTF modeling as a function of spherical harmonics order N and frequency will be briefly described in this section. Suppose we use spherical harmonics order N to represent the HRTF, sampled in L positions, for a certain frequency; the representation can be displayed by using (2):

$$\mathbf{H}\left(k\right) = \mathbf{Y}\mathbf{H}_{\mathbf{nm}}\left(k\right) \tag{4}$$

where $H\left(k\right)$ is the right or left HRTF function, represented by an $L \times 1$ vector, $H_{nm}\left(k\right)$ is the spherical Fourier transform of $H\left(k\right)$, an $(N+1)^2 \times 1$ vector, and Y is an $L \times (N+1)^2$ transformation matrix, as defined in (3), sampled at L sample points and presented up to order N. The spherical Fourier transform $H_{nm}\left(k\right)$ can be derived using the pseudo-inverse of Y, giving a numerical solution for (4) in the least squares sense:

$$\mathbf{H}_{\mathbf{nm}}\left(k\right) = \mathbf{Y}^{\dagger}\mathbf{H}\left(k\right) \tag{5}$$

Now, reconstructing the HRTF using a finite order coefficient will produce error that diminishes as the order is increased. Rafaely et al. [9] have shown that use of $N \approx kr$, where r can be estimated as an average radius of a human head, can reduce the normalized error to less than 3dB. In the next sections this reconstruction will be used to analyze the effect of finite order spherical harmonics coefficients on the two binaural cues discussed above.

4. THE EFFECT OF SPHERICAL HARMONICS ORDER ON ITD AND ILD

4.1. HRTF database

The CIPIC HRTF database [14] has been used for the analysis in this paper. The database consists of measured HRTFs for a dense grid of directions and a large number of human subjects including a KEMAR manikin. The CIPIC HRTF directions are presented by inter-aural polar coordinates where for a chosen azimuth and varying elevation, the cone of confusion is defined where ITD and ILD have similar values along this cone. Two sets of databases have been used in the analysis, the original CIPIC database and a reconstructed CIPIC database. The reconstructed CIPIC database was constructed considering the theory in Section 3, where the CIPIC database was transformed to the spherical harmonics domain and then reconstructed by a varying number of spherical harmonics coefficients with finite order N. The CIPIC database has missing samples around the south pole, which results in a large condition number of the transformation matrix Y, hence its pseudo-inverse is non-robust and can cause numerical errors. Nevertheless, by using numerical programming software such as MATLAB, as used in the current work, the results are satisfactory.

4.2. The effect of the spherical harmonics order on ITD

The following simulation study examines the ITD errors for a spherical harmonics finite order HRTF with respect to the azimuth in the inter-aural polar coordinates. Around the left and right ears the azimuth would be $\pm\pi$ and in front of the head the azimuth would be 0. In each simulation a single direction of arrival was chosen where the ITD was estimated for both HRTF sets and the error was computed. For a chosen azimuth, several elevations were selected such that all the directions of arrival would be around the cone of confusion of the selected azimuth. The root-mean-square of all errors around the cone of confusion was then computed, and averaged for 7 different subjects' HRTF sets. In order to achieve better time resolution for estimating ITD, the HRIRs were first interpolated [15] so the sample rate was 441KHz. In addition, the HRIRs were filtered by a 1.5KHz low-pass filter in order to examine ITD in the low frequencies. Two of the suggested ITD estimations were examined. The first estimation computed a cross-correlation of the selected HRIRs where the τ that results in the maximum correlation is the estimated ITD. Suppose $h_l(\theta,\phi)$ and $h_r(\theta,\phi)$ are two HRIRs for sound arriving from direction (θ,ϕ), the ITD will be estimated as:

$$\hat{\tau}(\theta,\phi) = \arg\max_{\tau}\left[\int_t h_l\left(t,\theta,\phi\right) h_r\left(t+\tau,\theta,\phi\right) dt\right] \tag{6}$$

The second method estimated the delay of the selected HRIRs using linear fitting of their excess phase and computing the ITD. Suppose the left or right HRTFs $H(\omega)$ is defined by magnitude $|H(\omega)|$ and a phase that consists of the minimum phase and the excess phase, the angle of the excess phase $\gamma(\omega)$ can be derived by:

$$\gamma(\omega) = \angle\left(H(\omega)H^{*}_{mp}(\omega)\right) \tag{7}$$

where the delay is the slope of the linear fitting of $\gamma(\omega)$, and ITD can be estimated by the difference between the delays. Fig. 1 presents an example of the decaying curve for both ITD estimations as a functions of N for the cone of confusion around -80°. Assuming an average head with radius of 0.09 m and a maximum frequency of 1.5KHz, the value of kr would be around 2.47. Examination of $N \approx kr$ order in Fig. 1 matches earlier work [9] where it was defined as the cut-off order.

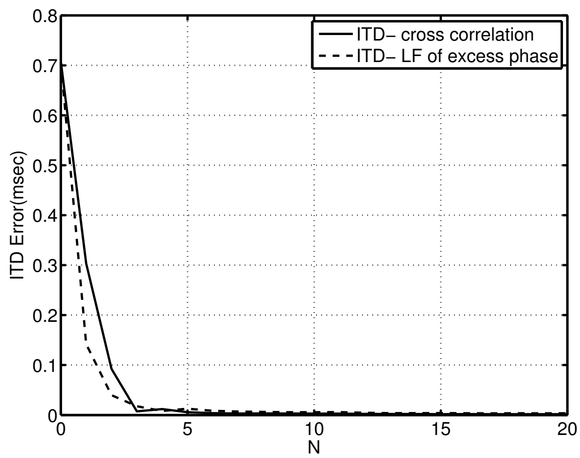

Figure 1: Two ITD error estimations for Az=-80° as a function of N

The ITD just noticeable differences (JND) have been shown to be around 10–20 μsec [16], [17], which are 5–9 sample points using the selected sample rate. Therefore, an estimation leads to an error that is smaller than the ITD JND might be considered as sufficient. Table 1 presents the ITD error estimations where ITD was estimated using (6). For different azimuth values a finite order N is presented where from this order a reconstructed HRTF set will result in ITD error estimation of less than a certain threshold. The thresholds were defined with respect to the ITD JNDs. It seems that a spherical harmonics representation with finite order of around 3–4 is sufficient to reconstruct the ITD values with error less than than 20 μsec. Notice that for lower ITD values (i.e., in front of the listener), the order is smaller, which implies a similar response to the two ears. Table 2 presents a similar analysis; however, the ITD estimations were computed using (7). As before, the error seems to decay when using a higher number of spherical harmonics coefficients; however, it seems a higher spherical harmonics order is needed in order to receive similar error values as before. The results might apply to the lack of robustness of this ITD estimation method.

Az. (deg)	$N_{50\mu sec}$	$N_{20\mu sec}$	$N_{10\mu sec}$	Avg. ITD (μsec)
-80	3	3	5	717
-45	2	4	5	472
-15	2	3	4	170
0	0	0	3	22
10	1	3	4	70
30	1	3	5	266
40	2	4	5	362
65	3	4	6	614
80	3	3	5	673

Table 1: ITD estimation using cross-correlation: The first rows are the spherical harmonics order needed to produce an error smaller than the threshold mentioned. The last row is the average ITD estimated value.

Az. (deg)	$N_{50\mu sec}$	$N_{20\mu sec}$	$N_{10\mu sec}$	Avg. ITD (μsec)
-80	2	3	6	729
-45	1	4	14	463
-15	1	3	15	157
0	0	2	15	26
10	1	4	18	139
30	1	4	13	377
40	1	5	20	468
65	3	7	10	695
80	3	4	5	800

Table 2: ITD estimation using linear fitting of the excess phase: The first rows are the spherical harmonics order needed to produce an error smaller than the threshold mentioned. The last row is the average ITD estimated value.

4.3. The effect of the spherical harmonics order on ILD

The following simulations examine the ILD errors regarding the reconstruction of the HRTFs with a spherical harmonics finite order. The ILD is estimated in dB for single frequencies and around 1.5KHz–3.5KHz it is mostly affected by the head shading effect; in general, as opposed to the ITD, it is more noticeable at higher frequencies. Consider $H_l(f,\theta,\phi)$ and $H_r(f,\theta,\phi)$ to be an HRTF set from direction (θ,ϕ), the ILD for a chosen frequency f will be:

$$a(f,\theta,\phi) = 20\log\left(\frac{|H_l(f,\theta,\phi)|}{|H_r(f,\theta,\phi)|}\right) \quad (8)$$

The ILD, as in the previous simulations, was estimated using HRTFs in different azimuths and elevations, where for each azimuth the elevations around the cone of confusion were selected and the root-mean-square error was computed and averaged for 7 different subjects. Fig. 2 presents the ILD error at azimuth -80° as a function of the spherical harmonics order N for frequencies $1KHz$, $2KHz$, and $3KHz$. It seems that the error diminished as N increased, where for higher frequencies, a larger number of coefficients is needed.

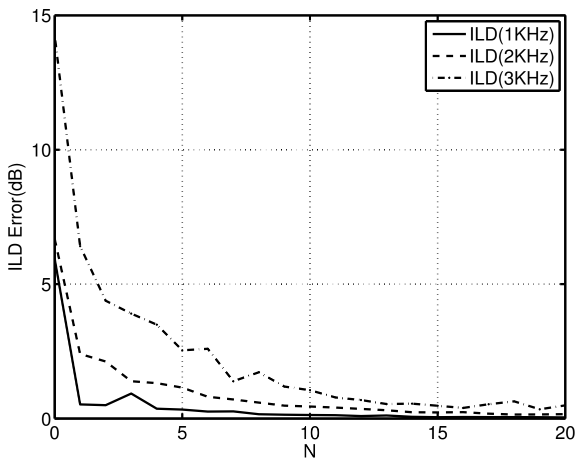

Figure 2: ILD error for Az=-80° as a function of N and different frequencies

Tables 3, 4, and 5 present the ILD error analysis for frequencies $1KHz$, $2KHz$, and $3KHz$, respectively. The thresholds were chosen with respect to the ILD JND, which is around 1dB [18]. For $1KHz$, $kr = 1.64$ and orders 1-2 were sufficient for an error smaller than 2dB; for $2KHz$, orders 3-4 and $kr = 3.29$; and for $3KHz$, orders 4-7 and $kr = 4.94$. For errors smaller than 1dB, higher orders were needed. It seems as before that an order of $N \approx kr$ is sufficient for a threshold that is near the ILD JND. Another aspect that can be seen in the tables is the difference between the orders for different azimuths. Similar to the ITD results, HRTFs that are in front of the listener need a smaller number of spherical harmonics coefficients to receive more accurate ILD values.

The results can be compared with a past work [9], where a similar reconstruction of the CIPIC HRTF database has been done; however, a complex sound field was simulated using the database and the inter-aural cross correlation (IACC) in different octave bands was examined. In the paper the aspect of direction of arrival was not examined since the pressure was a result of a sound field in a reverberant hall; however, the error of the IACC in different octave bands was examined as a function of the spherical harmonics order. In the same way as this work, it has been shown that the order of $N \approx kr$ would be a sufficient reconstruction for the sound field so the IACC could be computed with minor errors. Notice that the error of the IACC time bias τ was mentioned to be very precise using any spherical harmonics orders as opposed to the ITD estimations, which might occur as a result of the diffuseness of complex sound field which result in low τ values, similar to the ITD estimations in front of the listener.

Az.(deg)	N_{2dB}	N_{1dB}	Avg. ILD value (dB)
-80	1	1	3.17
-45	1	4	6.20
-15	1	3	1.25
0	0	0	1.80
10	1	2	3.99
30	1	4	8.63
40	1	4	9.26
65	1	1	6.40
80	1	1	6.43

Table 3: ILD error analysis for 1KHz, $kr = 1.64$

Az.(deg)	N_{2dB}	N_{1dB}	Avg. ILD value (dB)
-80	3	6	6.76
-45	4	8	7.40
-15	3	5	2.79
0	0	2	0.92
10	4	6	3.52
30	3	6	6.60
40	4	6	7.34
65	8	12	15.18
80	3	6	8.26

Table 4: ILD error analysis for 2KHz, $kr = 3.29$

Az.(deg)	N_{2dB}	N_{1dB}	Avg. ILD value (dB)
-80	7	11	10.53
-45	5	15	7.65
-15	5	7	3.16
0	4	6	1.13
10	5	14	3.38
30	6	11	8.21
40	5	10	8.83
65	7	12	11.02
80	6	10	12.05

Table 5: ILD error analysis for 3KHz, $kr = 4.94$

5. CONCLUSION

This work has analyzed the effect of a reconstructed sound field with a finite order of spherical harmonics coefficients on the binaural cues (i.e., interaural time differences (ITD) and interaural level differences (ILD)). The results show that a relatively small number of coefficients can represent an HRTF set and still preserve some of its spatial attributes. Furthermore, it seems that some HRTF directions of arrival, such as around the ears, compared with in front or in back of the head need a larger number of spherical harmonics coefficients.

6. ACKNOWLEDGMENT

This research was supported by The Israel Science Foundation (grant No. 155/06).

7. REFERENCES

[1] J. Strutt, “On our perception of sound direction,” *Philosophical Magazine*, vol. 13, pp. 214–232, 1907.

[2] D. Kistler and F. Wightman, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction.” *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1637–1647, 1992.

[3] J. Jot, V. Larcher, and O. Warusfel, “Digital signal processing issues in the context of binaural and transaural stereophony,” *98th Audio Engineering Society Convention*, 1995.

[4] J. Huopaniemi and J. Smith, “Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters,” *16th Audio Engineering Society International Conference on Spatial Sound Reproduction*, 1999.

[5] J. Nam, J. Abel, and J. Smith, “A Method for Estimating Interaural Time Difference for Binaural Synthesis,” *125th Audio Engineering Society Convention*, 2008.

[6] D. Phillips and S. Hall, “Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level,” *Hearing Research*, vol. 202.

[7] K. Watanabe, K. Ozawa, Y. Iwaya, Y. Suzuki, and K. Aso, “Estimation of interaural level difference based on anthropometry and its effect on sound localization,” *J. Acoust. Soc. Am.*, vol. 122, pp. 2832–2841, 2007.

[8] B. Rafaely, “Analysis and design of spherical microphone arrays,” *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 1, pp. 135–143, January 2005.

[9] B. Rafaely and A. Avni, “Interaural cross correlation in a sound field represented by spherical harmonics,” *J. Acoust. Soc. Am.*, vol. 127, no. 2, pp. 823–828, February 2010.

[10] A. Avni and B. Rafaely, “Interaural cross correlation and spatial correlation in a sound field represented by spherical harmonics,” *Ambisonics Symposium*, June 2009.

[11] J. R. Driscoll and D. M. Healy, “Computing fourier transforms and convolutions on the 2-sphere,” *Adv. App. Math.*, vol. 15, pp. 202–250, 1994.

[12] B. Rafaely, “Plane-wave decomposition of the pressure on a sphere by spherical convolution,” *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157, October 2004.

[13] M. Evans, J. Angus, and A. Tew, “Analyzing head-related transfer function measurements using surface spherical harmonics,” *J. Acoust. Soc. Am.*, vol. 104, no. 4, pp. 2400–2411, October 1998.

[14] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The CIPIC HRTF database,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2001)*, New Paltz, NY, USA, October 2001, pp. 99–102.

[15] S. Hwang and Y. Park, "Time delay estimation from HRTFs and HRIRs," *8th International Conference of Motion and Vibration Control*, 2006.

[16] A. Mills, "On the minimum audible angle." *J. Acoust. Soc. Am.*, vol. 30, pp. 237–246, 1958.

[17] R. Domnitz and H. Colburn, "Lateral position and interaural discrimination," *J. Acoust. Soc. Am.*, vol. 61, p. 1586, 1977.

[18] E. C. Carterette, *Hearing (Handbook of Perception, Vol 4) (v. 4)*. Academic Pr, 1978.