GENERALIZED FRAMEWORK FOR SPHERICAL MICROPHONE ARRAYS: SPATIAL AND FREQUENCY DECOMPOSITION

Thushara D. Abhayapala

Department of Information Engineering
Research School of Information Sciences & Engineering
Australian National University
e-mail: Thushara.Abhayapala@anu.edu.au

ABSTRACT

This paper provides a generalized framework to decompose a sourcefield into its spatial and frequency components using a spherical microphone array. Spherical arrays are used to decompose a soundfield into spherical harmonics which are the natural basis functions for space over directions. This paper extends this theory by including basis functions in frequency to decompose a soundfield into spatial-frequency components. Specifically, we use spherical Bessel functions as a set of basis in frequency. The paper also shows how to avoid numerical ill-conditioning inherited in open spherical arrays at frequencies corresponding to Bessel zeros.

Index Terms— Spherical microphone array, spatial decomposition, sound field, beamforming, Fourier spherical Bessel series

1. INTRODUCTION

Spherical microphone arrays have been introduced [1, 2] to use in spatial sound recording and beamforming applications. Over the last 5 years, there has been a strong interest [3–6] in theory and design of spherical microphone arrays due to their possible use in various applications such as in recording directional sounds for surround sound creation, speech enhancement and surveillance. The theory of such arrays is based on decomposition of soundfields to spherical harmonics which are the natural basis functions for valid soundfields over three dimensional space.

A soundfield in an *observation region*[1] is due to sources outside of the observation region. Source signals (such as speech) are generally frequency dependent as well as directional when observed from an observation region. Spherical microphone arrays are used to sample the soundfield over space and then to decompose the soundfield to the spherical harmonic components which are frequency dependent. These harmonic coefficients are manipulated to beamform or to record directional sources.

---

[1] Spatial region where we use microphones to sample the soundfield

The observed soundfield is a spatial and frequency filtered version of the original source signals. Specifically, individual spherical harmonic component of a soundfield obtained on a surface of a sphere is equal to filtering of the corresponding component of the source signal by a spherical Bessel function. Due to the unavoidable zeros of Bessel functions which lies within the desired band, it is difficult to recover the harmonic component of the sources around these frequencies. Rigid [2] and dual sphere microphone arrays [5, 7, 8] have used to overcome this limitations in the literature.

In this paper, we extend the general theory of spherical microphones to include frequency decomposition. Specifically, we use an orthogonal property of Bessel functions to form frequency basis functions. These basis functions are useful in expressing any source signal as a Fourier-spherical Bessel series. Then we show how to exploit Bessel zeros to decompose soundfield into these basis functions. In this approach, we not only avoid ill-conditioning associated with spherical arrays but also decompose the source signal into an efficient frequency basis set.

2. SOURCEFIELD AND SOUNDFIELD

2.1. Sourcefield

We define the *sourcefield* as the equivalent field created at an origin of an *observation region* as a function of angular (directional) position of the source and frequency. A sourcefield could consists of farfield/nearfield discrete/continuous (distributed) sources in space. We write an arbitrary sourcefield as

$$A(\hat{\phi}; k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm}(k) Y_{nm}(\hat{\phi}) \qquad (1)$$

where $\hat{\phi}$ is a unit vector pointing to the direction of the source (in 3D), $Y_{nm}(\cdot)$ are the spherical harmonics, $\alpha_{nm}(k)$ are the spherical harmonic coefficients of the source field at wavenumber k. Note that $k = 2\pi f/c$ where f is the frequency and c is the speed of sound propagation. In this paper, we use the

term 'frequency' for wavenumber since f is proportional to k assuming constant speed of sound.

2.2. Soundfield

Consider a spherical observation region Ω with radius R. Let $\boldsymbol{x} \in \Omega$ be a position vector connecting the origin of the sphere to an arbitrary point within the spherical region. We use $\hat{\boldsymbol{x}} = \boldsymbol{x}/\|\boldsymbol{x}\|$ to denote a unit vector pointing to the direction of $\boldsymbol{x}$, which we use to show directions in 3D.

We can write the soundfield at a point $\boldsymbol{x}$ in terms of source-field as

$$f(\boldsymbol{x}; k) = \begin{cases} \int A(\hat{\boldsymbol{\phi}}; k) \, e^{ik\boldsymbol{x}\cdot\hat{\boldsymbol{\phi}}} d\hat{\boldsymbol{\phi}} & \text{for farfield sources,} \\ \int A(\hat{\boldsymbol{\phi}}; k) \, \frac{e^{ik\|S\hat{\boldsymbol{\phi}}-\boldsymbol{x}\|}}{\|S\hat{\boldsymbol{\phi}}-\boldsymbol{x}\|} d\hat{\boldsymbol{\phi}} & \text{for nearfield sources,} \end{cases}$$
(2)

where S is a radius of a sphere with *equivalent* nearfield sources, and the integration is over the unit sphere. Using (1) and the spherical harmonic expansion of $e^{ik\boldsymbol{x}\cdot\hat{\boldsymbol{\phi}}}$ and $\frac{e^{ik\|S\hat{\boldsymbol{\phi}}-\boldsymbol{x}\|}}{\|S\hat{\boldsymbol{\phi}}-\boldsymbol{x}\|}$ (see [9, p. 30–32] equations (2.42) and (2.45)), we write

$$f(\boldsymbol{x}; k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \beta_{nm}(k) j_n(k\|\boldsymbol{x}\|) Y_{nm}(\hat{\boldsymbol{x}}),$$
(3)

where

$$\beta_{nm}(k) = \begin{cases} 4\pi i^n \alpha_{nm}(k) & \text{for farfield,} \\ -4\pi i k \alpha_{nm}(k) h_n^{(2)}(kS) & \text{for nearfield,} \end{cases}$$
(4)

where $j_n(\cdot)$ is the spherical Bessel function of order n, and $h_n^{(2)}(\cdot)$ is the spherical Hankel function of the second kind of order n. For the rest of the paper, we only consider the case of farfield sources, however extension to nearfield sources is possible.

Equation (3) represents the soundfield within the observation region as a weighted sum of spherical harmonic components. The coefficients $\beta_{nm}(k)$ in (3) are called soundfield coefficients. The relationship between the sourcefield coefficients and the soundfield coefficients are given by (4).

In the case of rigid spherical microphone array, the sound-field is disturbed by the rigid sphere. In this case, the factor $j_n(k\|\boldsymbol{x}\|)$ in (3) is replaced by

$$j_n(k\|\boldsymbol{x}\|) - (j_n'(kR_1)/h_n^{(1)'}(kR_1)) h_n^{(1)'}(k\|\boldsymbol{x}\|)$$

where $(\cdot)'$ represents the derivative of a function with respect to its argument. In this paper, we only consider the *open* sphere configuration.

2.3. Spherical Harmonic Decomposition

Generally, spherical microphone arrays can be used to estimate spherical harmonic coefficients $\beta_{nm}(k)$ of the sound-field (3), and then the sourcefield coefficients $\alpha_{nm}(k)$. These

coefficients could be further processed to form beams in certain directions, record them for surround sound applications, and use them for localization/ direction of arrival estimation (DOA).

Suppose we know the soundfield over a sphere of radius $r = \|\boldsymbol{x}\|$. Using the orthonormal property of spherical harmonics, we write

$$\gamma_{nm}(r, k) \triangleq \int f(r\hat{\boldsymbol{x}}, k) Y_{nm}^*(\hat{\boldsymbol{x}}) d\hat{\boldsymbol{x}} = \beta_{nm}(k) j_n(kr) \quad (5)$$

where the integration is over the unit sphere. In the spherical microphone array literature [1, 4, 5], finite number of microphones on a fixed radius are used to measure the soundfield and then estimate the soundfield as

$$\beta_{nm}(k) = \frac{\gamma_{nm}(r, k)}{j_n(kr)},$$
(6)

for frequencies where $j_n(kr) \neq 0$.

2.4. Filtering in Frequency Domain

We can clearly see the filtering of source signal by observing (5), which is a result of spatial decomposition on a given radius. Using (4) and (5), for a farfield sources, we write

$$\gamma_{nm}(r, k) = 4\pi i^n \alpha_{nm}(k) j_n(rk).$$
(7)

Note that in (7), sourcefield coefficients $\alpha_{nm}(k)$ is filtered by $j_n(kr)$. Let the frequency band of the source signal be $k \in (0, k_u)$, and let $z_{n\ell}$ be the ℓth zero of $j_n(z_{n\ell})$, where $z_{n1} < z_{n2} < \dots$. If $z_{n\ell_1} < k_u r < z_{n\ell_2}$ then there are ℓ_1 zero-crossings of $j_n(kr)$ within the desired bandwidth. Hence it is difficult to recover $\alpha_{nm}(k)$ around these frequencies.

3. FREQUENCY DECOMPOSITION

3.1. Fourier Bessel Series

Bessel functions are quasi periodic with successive zero crossing intervals. In the past Bessel function series have been used to represent speech and sound signals and considered as more efficient than Fourier domain description [10]. A finite duration signal $f(t)$ in the interval of $0 < t < t_u$ can be represented by an infinite Fourier Bessel series expansion [11] as

$$f(t) = \sum_{\ell=1}^{\infty} C_\ell J_v(\frac{z_\ell}{t_u} t), \text{ for } 0 < t < t_u$$
(8)

where z_ℓ, $\ell = 1, 2, \dots$ are the roots of $J_v(z_\ell) = 0$ and v is a real number. Using the orthogonality of $J_v(\frac{z_\ell}{t_u} t)$ for different ℓ,

$$C_\ell = \frac{2}{t_u^2 J_{v+1}^2(z_\ell)} \int_0^{t_u} t f(t) J_v(\frac{z_\ell}{t_u} t) \, dt.$$
(9)

In this paper, we make use of Fourier-(spherical) Bessel (FSB) series which has spherical Bessel functions instead of cylindrical Bessel functions. The synthesis and analysis of FSB series can be derived using (8) and (9) as

$$f(t) = \sum_{\ell=1}^{\infty} C_\ell j_n(\frac{z_\ell}{t_u}t), \text{ for } 0 < t < t_u \quad (10)$$

and

$$C_\ell = \frac{2}{t_u^3 j_{n+1}^2(z_\ell)} \int_0^{t_u} t^2 f(t) j_n(\frac{z_\ell}{t_u}t) \, dt. \quad (11)$$

Also note that the relevant orthogonality property

$$\int_0^{t_u} j_n(\frac{z_{\ell_1}}{t_u}t) j_n(\frac{z_{\ell_2}}{t_u}t) \, t^2 dt \quad (12)$$

$$= \begin{cases} \frac{t_u^3}{2} j_{n+1}^2(z_\ell) & \text{if } z_{\ell_1} = z_{\ell_2}, \\ 0 & \text{otherwise.} \end{cases}$$

3.2. FSB Expansion

We choose FSB series (10) to expand the sourcefield coefficients $\alpha_{nm}(k)$ within the desired frequency band $0 < k < k_u$,

$$\alpha_{nm}(k) = \sum_{\ell=1}^{\infty} a_{nm\ell} \, j_n(\frac{z_{n\ell}}{k_u}k). \quad (13)$$

For reasons to be understood in the next section, we use the nth order spherical Bessel function corresponding to the order of the sourcefield coefficient. One advantage of using (13) is that once we estimate $a_{nm\ell}$ for $n = 0, \ldots, m = -n : n$, and $\ell = 1, \ldots$, we could recalculate sourcefield coefficients $\alpha_{nm}(k)$ for any frequency k.

Thus, a general farfield source distribution (sourcefield) can be written as

$$A(\hat{\phi}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \sum_{\ell=1}^{\infty} a_{nml} \, j_n(\frac{z_{n\ell}}{k_u}k) Y_{nm}(\hat{\phi}), \quad (14)$$

and

$$a_{nm\ell} = \frac{2 \int_0^{k_u} \int A(\hat{\phi}, k) Y_{nm}^*(\hat{x}) j_n(\frac{z_{n\ell}}{k_u}k) d\hat{\phi}k^2 dk}{k_u^3 j_{n+1}^2(z_{n\ell})} \quad (15)$$

where the second integration is over the unit sphere. Equation (15) is derived by applying the orthonormal property of spherical harmonics and orthogonal property of spherical Bessel functions (12).

3.3. Exploitation of Bessel Zeros

We start with the following theorem:

Theorem 1 (Spatial-Frequency Decomposition) *Let $f(\boldsymbol{x}, k)$, $x \in \Omega$, $k \in (0, k_u)$ be a valid soundfield over a spatial region* Ω. *Then the sourcefield corresponding to $f(\boldsymbol{x}, k)$ is given by (14) with spatial-frequency harmonic coefficients*

$$a_{nml} = \frac{1}{2\pi i^n k_u^3 j_{n+1}^2(z_{n\ell})} \int_0^{k_u} \int f(r_{n\ell}\hat{\boldsymbol{x}}, k) Y_{nm}^*(\hat{\boldsymbol{x}}) k^2 d\hat{\boldsymbol{x}} dk \quad (16)$$

where $f(r_{n\ell}\hat{\boldsymbol{x}}, k)$ are the soundfield over discrete radii $r_{n\ell} = z_{n\ell}/k_u$ for $n = 0, 1, \ldots, \ell = 1, 2, \ldots,$ and $z_{n\ell}$ is the ℓth zero of the nth spherical Bessel function.

Proof:
Since $f(\boldsymbol{x}, k)$, $x \in \Omega$, $k \in (0, k_u)$ is induced by the sourcefield $A(\hat{\boldsymbol{\phi}}, k)$, from (4) and (5), we have

$$4\pi i^n \alpha_{nm}(k) j_n(kr) = \int f(r\hat{\boldsymbol{x}}, k) Y_{nm}^*(\hat{\boldsymbol{x}}) d\hat{\boldsymbol{x}}. \quad (17)$$

By substituting (13) in (17) for $\alpha_{nm}(k)$ we obtain

$$4\pi i^n \big[\sum_{\ell=1}^{\infty} a_{nm\ell} j_n(\frac{z_{n\ell}}{k_u}k) \big] j_n(kr) = \int f(r\hat{\boldsymbol{x}}, k) Y_{nm}^*(\hat{\boldsymbol{x}}) d\hat{\boldsymbol{x}}. \quad (18)$$

Let $r = z_{n\ell}/k_u$. We multiply both sides of (18) by k^2 and integrate over $(0, k_u)$, and then apply orthogonal property (12) to get (16). □

Theorem 1 states that by sampling the soundfield over a number of discrete spheres, we can decompose the sourcefield to its basis coefficients $a_{nm\ell}$. However, we have not answered the following questions: (i) What is the maximum number of ℓ that sufficiently characterizes a typical sourcefield; (ii) Does $\ell_{\max}$ depend on n?; (iii) How many spheres of microphones do we need?. These questions need to be answered in a future publication. However, we think that the number of co-centered spheres needed is too many for a practical implementation. In the next section, we show two alternative methods to use (i) dual sphere, (ii) single sphere measurements to estimate $a_{nm\ell}$.

4. ALTERNATIVE IMPLEMENTATIONS

4.1. Dual Sphere Method

Suppose we have two co-centered open spherical microphone arrays of radius r_1 and r_2 as in [5,7]. Then it is can be shown

$$a_{nml} = \frac{1}{2\pi i^n k_u^3 j_{n+1}^2(z_{n\ell})} \times \quad (19)$$

$$\int_0^{k_u} \big[\frac{\int [f(r_1\hat{\boldsymbol{x}}, k) + f(r_2\hat{\boldsymbol{x}}, k)] Y_{nm}^*(\hat{\boldsymbol{x}}) d\hat{\boldsymbol{x}}}{j_n(kr_1) + j_n(kr_2)} \big] j_n(\frac{z_{n\ell}}{k_u}k) k^2 dk.$$

We omit the proof due to space limitations. Having measurements from two radii will help to have improved SNR of the measured signal for frequencies where one of the radius corresponds to a Bessel zero. One may also use a method given in [5] to find the optimum ratio of two radii or combine the two measurements in some optimum way.

4.2. Omitting Low SNR measurements

In this method, we only use measurements from a single sphere of radius r_0. It is clear that we can not recover the source field components $\alpha_{nm}(k)$ in the vicinity of $k = z_{n\ell}/r_0$.

Suppose that there is an acceptable SNR outside of the frequency range $(z_{n\ell}/r_0) - \delta < k < (z_{n\ell}/r_0) + \delta$, where δ is a suitable small constant. Then we estimate $\alpha_{nm}(k_q)$, $k_q \notin [(z_{n\ell}/r_0) - \delta < k < (z_{n\ell}/r_0) + \delta]$ as

$$\alpha_{nm}(k_q) = \frac{1}{4\pi i^n j_n(k_q r_0)} \int f(r_0 \hat{\boldsymbol{x}}, k_q) Y_{nm}^*(\hat{\boldsymbol{x}}) d\hat{\boldsymbol{x}}. \quad (20)$$

Provided there are sufficient frequency sampling $q = 1, \ldots, Q$ which avoid low SNR frequency interval, we can show

$$a_{nm\ell} \approx \frac{2}{k_u^3 j_{n+1}^2(z_\ell)} \sum_{q=1}^{Q} \alpha_{nm}(k_q) j_n\left(\frac{z_{n\ell}}{k_u} k_q\right) k_q^2 \Delta k_q \quad (21)$$

where Δk_q is the difference between qth and $q + 1$th sample.

We have following comments:

1. We believe that this is a promising method to use with existing spherical microphone array methods to estimate spatial-frequency coefficients of the sourcefield.

2. The method avoids the difficulty associated with Bessel zeros.

3. One can argue that this method is an extrapolation as using (13) together with estimated $a_{nm\ell}$ from (21) can calculate $\alpha_{nm}(k)$ over the frequency bands $(z_{n\ell}/r_0) - \delta < k < (z_{n\ell}/r_0) + \delta$ which corresponds to Bessel zeros.

4. We still need to find practical value(s) for δ, as well as few example illustrations to support these methods. Unanswered details will be addressed in a future publication.

5. SUMMARY

This paper combines two sets of orthogonal functions (i) spherical harmonics, (ii) spherical Bessel functions, to form basis sets in angle and in frequency respectively, to decompose a soundfield using spherical microphone arrays. We show that by sampling the soundfield over a natural set of spheres, we can estimate the corresponding sourcefield coefficients in the above basis set. Since it is not practical to have a large number of sampling spheres, we show two alternative methods (i) dual sphere measurement, and (ii) single sphere with omitting low SNR measurements, to estimate the sourcefield coefficients.

However, there are number of unanswered questions in this paper, which we hope to provide in a future publication together with illustrations to support the claims in this paper.

6. REFERENCES

[1] T.D. Abhayapala and D.B. Ward, "Theory and design of higher order sound field microphones using spherical microphone array," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2002, vol. 2, pp. 1949–1952.

[2] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2002, vol. 2, pp. 1781–1784.

[3] Z. Li, R. Duraiswami, and L.S. Davis, "Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays," in *IEEE International Conference Acoustics, Speech and Signal Processing,*, May 2004, vol. 4, pp. 41–44.

[4] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Acoust. Speech Sig. Proc.*, vol. 13, pp. 135–143, Jan. 2005.

[5] I. Balmages and B. Rafaely, "Open sphere designs for spherical microphone arrays," *IEEE Trans. Audio, Speech, and Language Proc.*, vol. 15, pp. 727–732, Feb. 2007.

[6] Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Trans. Acoust. Speech Sig. Proc.*, vol. 15, pp. 702–714, Feb. 2007.

[7] C. Jin A. Parthy and A. Avan Schaik, "Optimisation of co-centred rigid and open spherical microphone arrays," in *AES 120th Convention*, Paris, France, May 2006, pp. 388–391.

[8] T.D. Abhayapala and M.C.T. Chan, "Limitation and eriror analysis of spherical microphone arrays," in *14th International Congress on Sound and Vibration (ICSV14)*, Cairns, Australia, July 2007.

[9] David Colton and Rainer Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer, New York, 2nd edition, 1998.

[10] K. Gopalan, "Speech modification by selective fourier-bessel series expansion of speech signals," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, Orlando, FL, USA, 1999, vol. 1, pp. 388–391.

[11] F. Bowman, *Introduction to Bessel Functions*, Dover, New York, 1958.