

SOUND FIELD DECOMPOSITION USING SPHERICAL MICROPHONE ARRAYS

Dmitry N. Zotkin, Ramani Duraiswami, and Nail A. Gumerov

Perceptual Interfaces and Reality Lab, UMIACS, University of Maryland, College Park

ABSTRACT

Spherical microphone arrays offer a number of attractive properties such as direction-independent acoustic behavior and ability to reconstruct the sound field in the vicinity of the array. Such ability is necessary in applications such as ambisonics and recreating auditory environment over headphones. We compare the performance of two scene reconstruction algorithms – one based on least-squares fitting the observed potentials and another based on computing the far-field signature function directly from the microphone measurements. A number of features important for the design and operation of spherical microphone arrays in real applications are revealed. Results indicate that it is possible to reconstruct the sound scene up to order p with p^2 microphones.

Index Terms— Acoustic fields, spherical microphone arrays, array signal processing, acoustic position measurement.

1. INTRODUCTION

Spherical microphone arrays offer a number of properties attractive for the development of the acoustic and audio systems with 3-D listening capability. Due to 3-D symmetry of the array, the array beamforming pattern is independent of the steering direction and the spatial structure of the acoustic field can be captured without distortion. The spherical configuration leads naturally to an elegant mathematical framework based on elementary solutions of Helmholtz equation in spherical coordinates [1] [2] [3]. Accordingly, spherical arrays are being used in real-time beamforming [4], in capturing of the spatial acoustic field [5], and in other applications.

Recently there has been an interest in representing the acoustic field in the array neighborhood over the plane-wave and spherical wavefunction bases. Plane-wave decompositions are useful in many application scenarios, including source localization and beamforming [4], capture and individualized HRTF-based reproduction of acoustic scenes [5], etc. Ref. [6] presented a framework for performing decomposition using spherical convolution under the assumption of a continuous pressure-sensitive microphone array surface. In case of discrete microphones positioned on the sphere surface this assumption is invalid, and a quadrature formulae that preserves the orthonormality of spherical harmonics should be used as in [2]. Quadrature based on Fliege points [8] was presented and evaluated and two plane-wave decomposition algorithms were developed in [7]. The current work analyzes the performance of those algorithms under realistic operating conditions – finite number of microphones, environmental noise, and aliasing effects – using both synthetic and experimental data.

2. BACKGROUND

In a space with no acoustic sources, acoustic wave propagation at a wavenumber k is governed by the Helmholtz equation [7]

$$\nabla^2 \psi(k, \mathbf{r}) + k^2 \psi(k, \mathbf{r}) = 0, \quad (1)$$

Thanks to the U.S. Department of Veterans Affairs for funding this work.

where $\psi(k, \mathbf{r})$ is the Fourier transform of the pressure. Solutions of the Helmholtz equation can be expanded as a series of regular $R_n^m(k, \mathbf{r})$ and singular $S_n^m(k, \mathbf{r})$ spherical basis functions (see [7]):

$$R_n^m(k, \mathbf{r}) = j_n(kr)Y_n^m(\theta, \varphi); \quad S_n^m(k, \mathbf{r}) = h_n(kr)Y_n^m(\theta, \varphi), \quad (2)$$

where (r, θ, φ) are spherical coordinates of the radius vector \mathbf{r} , $j_n(kr)$ and $h_n(kr)$ are the spherical Bessel and Hankel functions, and $Y_n^m(\theta, \varphi)$ are the orthonormal spherical harmonics.

Any regular acoustic field $\Psi(k, \mathbf{r})$ near a point \mathbf{r}^* in a region that does not contain sources can be represented as a sum of regular functions with some complex coefficients $C_n^m(k)$ as

$$\Psi(k, \mathbf{r}) = \sum_{n=0}^{p-1} \sum_{m=-n}^n C_n^m(k) R_n^m(k, \mathbf{r} - \mathbf{r}^*) + \varepsilon(p, k|\mathbf{r} - \mathbf{r}^*|). \quad (3)$$

To achieve negligible truncation error it is sufficient to set [9]

$$p = (\varepsilon k D - 1)/2. \quad (4)$$

3. SOLVING THE ACOUSTIC SCENE

The potential $\psi(\mathbf{s}', \mathbf{s})$ created at point \mathbf{s}' on the surface of the sound-hard sphere of radius a by plane wave $e^{i\mathbf{k}\mathbf{s}\cdot\mathbf{r}}$ propagating in the direction \mathbf{s} is given by

$$\psi(k, \mathbf{s}', \mathbf{s}) = \frac{i}{(ka)^2} \sum_{n=0}^{\infty} \frac{i^n (2n+1) P_n(\mathbf{s}' \cdot \mathbf{s})}{h_n'(ka)}, \quad (5)$$

where $P_n(\mathbf{s}' \cdot \mathbf{s})$ is the Legendre polynomial of degree n and $h_n'(ka)$ is the derivative of the spherical Hankel function. Assume that L_i microphones are placed at directions \mathbf{s}'_i comprising an ‘‘M-grid’’. The goal is to measure the potentials $\Psi(k, \mathbf{s}'_i)$ at those microphones and to recover the signature function $\mu(k, \mathbf{s}_j)$ over L_j directions \mathbf{s}_j comprising an ‘‘S-grid’’. Two algorithms were introduced in [7].

K-method: This method is essentially based on least-squares fitting of the observed potentials with the plane-wave magnitudes $\mu(\mathbf{s}_j)$. Using equation (5), one can construct a linear system for $\mu(\mathbf{s}_j)$ as follows. Define kernel $K(\mathbf{s}'_i, \mathbf{s}_j)$:

$$K(\mathbf{s}'_i, \mathbf{s}_j) = \frac{i}{(ka)^2} \sum_{n=0}^{p-1} \frac{i^n (2n+1) P_n(\mathbf{s}'_i \cdot \mathbf{s}_j)}{h_n'(ka)} \quad (6)$$

Then, due to linearity of the system, the following equation holds for each direction \mathbf{s}'_i :

$$\Psi(\mathbf{s}'_i) = \sum_{j=1}^{L_j} K(\mathbf{s}'_i, \mathbf{s}_j) \mu(\mathbf{s}_j). \quad (7)$$

Denote by Ψ the $L_i \times 1$ vector of measured $\Psi(\mathbf{s}'_i)$ values, by K the $L_i \times L_j$ matrix of computed $K(\mathbf{s}'_i, \mathbf{s}_j)$ values, and by Λ the $L_j \times 1$ vector of unknown $\mu(\mathbf{s}_j)$. The system then becomes

$$\Psi = K\Lambda. \quad (8)$$

This system is underdetermined when $L_j > L_i$; overdetermined and solved for Λ in least squares sense when $L_j < L_i$, obtaining best-fitting set of $\mu(\mathbf{s}_j)$; and has a unique solution when $L_j = L_i$.

M-method: It is also possible to explicitly compute Λ from Ψ without requiring matrix inversion and for any number of directions in S-grid. The coefficients C_n^m can be computed as

$$C_n^m = -i(ka)^2 h_n'(ka) \int_{S_u} \psi(\mathbf{s}') Y_n^{-m}(\mathbf{s}') dS(\mathbf{s}') \quad (9)$$

(see [7] for derivation). The expression for $\mu(\mathbf{s})$ then can be written:

$$\mu(\mathbf{s}) = -i \frac{(ka)^2}{4\pi} \sum_{n=0}^{p-1} (2n+1) i^{-n} h_n'(ka) \int_{S_u} \psi(\mathbf{s}') P_n(\mathbf{s}' \cdot \mathbf{s}) dS(\mathbf{s}') \quad (10)$$

using addition theorem for the spherical harmonics. In the discrete case the integral is replaced by quadrature over points \mathbf{s}'_i with weights w'_i to obtain kernel $M(\mathbf{s}_j, \mathbf{s}'_i)$:

$$M(\mathbf{s}_j, \mathbf{s}'_i) = -i \frac{(ka)^2}{4\pi} \sum_{n=0}^{p-1} (2n+1) i^{-n} h_n'(ka) P_n(\mathbf{s}_j \cdot \mathbf{s}'_i). \quad (11)$$

Then, one can form matrix M of size $L_j \times L_i$ and find Λ directly as $\Lambda = M\tilde{\Psi}$, where $\tilde{\Psi}$ is an $L_i \times 1$ vector with values $w'_j \Psi(\mathbf{s}'_i)$. No linear system solution is needed.

Hemispherical Array: The above two equations can be adapted to the case of a hemispherical array [4]. Assume that an acoustic source is placed on one side of an infinite sound-hard plane. By the acoustic image principle [4], the potential $\psi_h(\mathbf{s}', \mathbf{s})$ created at point \mathbf{s}' is given by summing up potentials due to two (original and reflected) plane waves given by equation (5):

$$\psi_h(k, \mathbf{s}', \mathbf{s}) = \psi(k, \mathbf{s}', \mathbf{s}) + \psi(k, \mathbf{s}', \bar{\mathbf{s}}), \quad (12)$$

where $\bar{\mathbf{s}}$ is \mathbf{s} reflected in the array base plane. Therefore, the kernel $K(\mathbf{s}'_i, \mathbf{s}_j)$ should be replaced with the kernel $K_h(\mathbf{s}'_i, \mathbf{s}_j)$:

$$K_h(\mathbf{s}'_i, \mathbf{s}_j) = K(\mathbf{s}'_i, \mathbf{s}_j) + K(\mathbf{s}'_i, \bar{\mathbf{s}}_j), \quad (13)$$

In addition, the S-grid should cover only the upper hemisphere (the other one contains image sources). The M-method requires no modifications for use with the hemispherical array.

Observations: 1) The K-kernel has $h_n'(ka)$ in the denominator, whereas the M-kernel has it in the numerator (recall that the function $h_n'(ka)$ grows exponentially when $n > ka$). It follows that the M-method is expected to be sensitive to the correct choice of p and at high p to the measurement noise. 2) Low-frequency acoustic waves create very small differences in potentials at different microphones, which could lead to poor conditioning of the K matrix.

4. SIMULATION SETUP

Simulation of plane-wave decomposition with a spherical microphone array ($a = 0.106m$) was performed. The M-grid was set to be 64-point Fliege grid (“64F”) [7] [8]. For each frequency (0.5 through 6 kHz), each source direction from a list of 1024 random directions, and each noise variance (0, 0.5, and 1.0), the potentials at microphone locations were computed using equation (5) with $p = p_m$. Then, each potential was corrupted by Gaussian noise with zero mean and given variance, and either K-method or M-method was applied to solve for $\mu(\mathbf{s})$ over the S-grid using $p = p_s$. In the plots, p_m and p_s are given in terms of p^* computed as prescribed by equation (4). The direction in S-grid with the largest magnitude of $\mu(\mathbf{s})$ was taken to be the detected direction of the source. The error



Fig. 1. Spherical (left) and hemispherical (right) microphone arrays.

measure was the angular difference, in radians, between the “true” and detected source directions, averaged over 1024 source directions for each frequency, each noise variance, and each method. S-grids used were “64F” and “49F”.

5. EXPERIMENTAL SETUP

The experiments were performed in a large office room. Two test signals played via typical computer speaker were used: a continuous sine wave and a 2.46 ms long up-sweep (chirp) signal. Microphone signals were recorded via two 32-channel NI PCI-6071E data acquisition boards. The analog outputs and inputs of both boards were synchronized to run off a common clock to allow for time averaging.

Spherical array experiment: The 60-microphone spherical array (shown in Figure 1), made out of a hollow plastic lamp shade of 0.101 m radius, was placed in the center of the room on a tripod. The microphone arrangement followed 64-point Fliege grid [8] with nodes 12, 24, 29, and 37 removed, forming a “60F” grid. The spatial aliasing frequency for the array is about 3.85 kHz. Four random positions in the room, all at a distance of about 1.5 m from the array, were chosen to place the source. A 2 s recording was done for each of the sine wave signals. The chirp signal was time-averaged over 10 repetitions with one-second pauses to minimize reverberant noise. The recorded signal was then windowed with 5.0 ms rectangular or Hann window and the potentials at each microphone were computed, normalized, and used as inputs to the two solution methods. The direction with the largest $\mu(\mathbf{s})$ magnitude in the S-grid was taken to be the detected source direction. Two S-grids were used: “49F” and “60F”.

Hemispherical array experiment: The 64-microphone hemispherical array (shown in Figure 1), made out of half of a bowling ball of 0.109 m radius, was mounted in the center of a circular table of 0.457 m radius. The M-grid for the array, referred to as “64H”, was obtained as described in [4]. The spatial aliasing frequency for the array is about 4.66 kHz. The rest of the experimental setup is the same as for the spherical array experiment. Two S-grids covering the upper hemisphere were used. The first S-grid was the “64H” grid and the second S-grid was the 121-point Fliege grid modified to remove all points having negative z , resulting in a 62-point “62H” grid.

6. RESULTS

6.1. Simulation results

In all plots throughout the paper, the M-grid and S-grid used are annotated in the plot, and the K-method and M-method errors are shown in dashed and dotted lines, respectively. For each method, three lines shown correspond to noise variances of 0, 0.5, and 1.0. In all plots, higher noise variance causes higher error; therefore, these lines are not annotated separately.

Figure 2 shows the error plots obtained with $p_m = p_s = p^*$ (i.e., using the p prescribed by equation (4) both for computing the

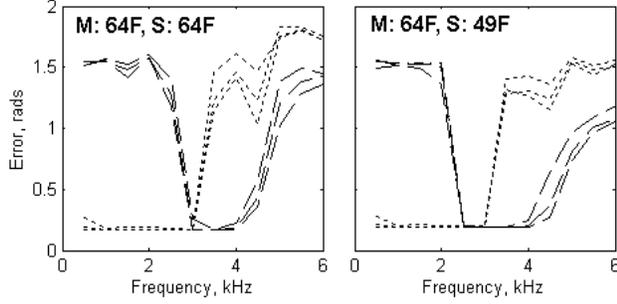


Fig. 2. Simulation, $p_m = p_s = p^*$. In this and all other plots, K-method (M-method) error is shown in dashed (dotted, respectively) line, and M-grid and S-grid used are annotated in the plot.

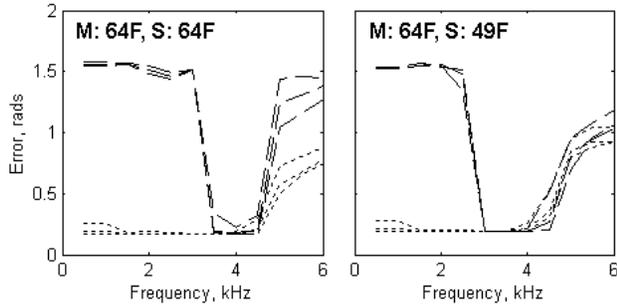


Fig. 3. Simulation, $p_m = p^*$, $p_s = \frac{7}{8}p^*$.

field at the array and for solving this field). The error is always positive due to the discrete grid nature. Interestingly, for the 64F S-grid M-method shows good behavior for up to 3 kHz, whereas K-method does exactly the opposite. When the frequency is increased beyond the spatial aliasing limit (above approximately 4 kHz), the K-method performance gracefully degrades. For the 49F S-grid, the M-method shows the same behavior (because it computes $\mu(\mathbf{s})$ directly from $\Psi(\mathbf{s}'_i)$) and the K-method operating range starts at slightly lower frequency. No significant differences are observed for different noise values. That means that noise with magnitude comparable to the signal does not significantly interfere with localization.

In light of Observation 1 in Section 3, the next experiment was performed with p_s decreased compared to p^* . Figure 3 shows the error plots obtained for $p_m = p^*$ and $p_s = \frac{7}{8}p^*$. The plots show that the M-kernel instability problem is successfully solved by decreasing p_s – the M-method is now operating up to the spatial frequency limit and the error increases very gradually above this limit. However, the K-method operation range is decreased.

Analysis of matrix K (equation (8)) shows that it is indeed poorly conditioned in the low frequency region where the K-method shows high localization error. Therefore, a regularization term was added to the LSE solution in an attempt to improve algorithm's performance:

$$K^T \Psi = (K^T K + \varepsilon I) \Lambda, \quad (14)$$

where I is the identity matrix and ε is the regularization constant. Figure 4 shows obtained error plots with $p_m = p_s = p^*$ and $\varepsilon = 10^{-4}$; inclusion of the regularization term totally removed poor localization performance at low frequencies, and the K-method range

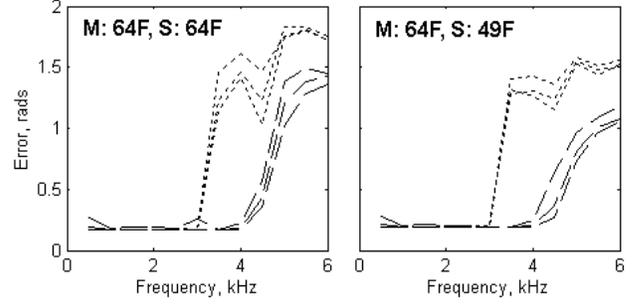


Fig. 4. Simulation, $p_m = p_s = p^*$, regularization $\varepsilon = 10^{-4}$.

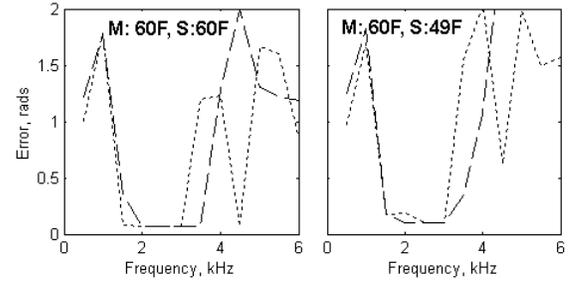


Fig. 5. Experiment, spherical array, sine wave signal.

of validity is now improved, p to be the whole operating range of array (in fact, error plots of K-method and M-method are identical up to about 2.5 kHz). Also, the particular value of ε ranging from 10^{-7} to 1.0 only marginally influences the results.

In summary, from all the simulations presented, it is apparent that with proper choice of parameters and with regularization K-method and M-method are behaving substantially equivalently over the useful array frequency range and either method can successfully extract directional information from the acoustic field presented to the array. Some differences do occur at the frequencies above the spatial aliasing limit, where error exhibited by M-method appears to be lower, and random noise does not seem to affect either algorithm.

6.2. Experimental results

The experiments with the real spherical and hemispherical arrays were designed and carried out as described earlier in this document. No substantial differences were observed in the results when the signal windowing function (rectangular or Hann) was changed and when ε was varied (within the same range as in simulations). Therefore, in all the plots below $p_s = p^*$, $\varepsilon = 10^{-2}$, and rectangular windowing is used. For the spherical (hemispherical) array, 60F and 49F (64H and 62H, respectively) S-grids were used.

Spherical array, sine wave signal: Figure 5 demonstrates the localization performance obtained when the data frame is randomly selected in the middle of the sine wave signal playback so that the measurements are corrupted with reverberation. The operating frequency of the array appears to be from about 1.5 to about 3.5 kHz. The M-method exhibits earlier performance degradation similar to that observed in simulations. The general trends observed when varying p_s and ε were consistent with those seen in simulations.

Spherical array, chirp signal: Figure 6 shows localization performance for the chirp signal repeated 10 times with 1 s pauses and

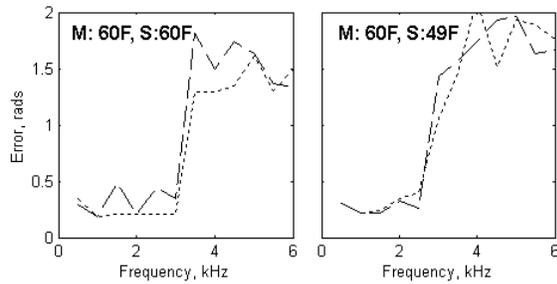


Fig. 6. Experiment, spherical array, chirp signal.

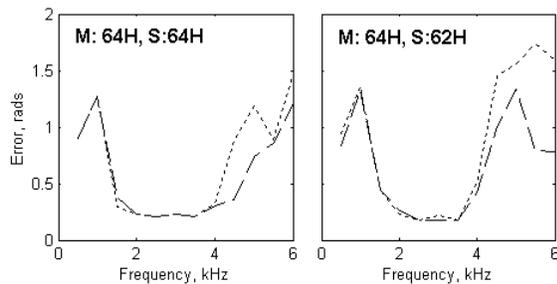


Fig. 7. Experiment, hemispherical array, sine wave signal.

time-averaged. In this plot, there no significant difference between K- and M-methods. Note an increase in localization accuracy in the low-frequency range due to absence of reverberation.

Hemispherical array, sine wave signal: Figure 7 illustrates the localization error for the sine wave signal. The general structure of the plot is similar to the one seen for the spherical array. The range of good localization is extended up to about 4.5 kHz.

Hemispherical array, chirp signal: This plot is substantially similar to Figure 7 and is not included for lack of space. Unlike the spherical array case, the error is still high in the low frequency range (i.e., absence of reverberation does not improve localization).

Summary: The experimental plots for the spherical array show slightly narrower range of good localization performance (up to 3 – 3.5 kHz) compared to the simulation (up to 4 kHz), possibly due to four bottom microphones missing in 60F grid compared to 64F grid used for simulations. Also, it is seen that the reverberation harms the localization significantly at low frequencies.

The experimental plots for the hemispherical array show higher upper limit due to denser microphone spacing and to the sound-hard array support. However, the localization performance at low frequencies is always poor due to the array being mounted on a relatively small table causing deviation from the acoustic image principle (remember that the theoretical foundation assumes that the array is placed on an infinite, sound-hard plane). For reference, the frequency of the sound for which the wavelength is equivalent to the table radius is 750 Hz. It can be expected that if the array were mounted directly on a floor or on a wall the localization performance at lower frequencies would be significantly improved.

An important conclusion can be derived regarding the number of microphones necessary for successful recovery of the scene spatial structure up to order p . Earlier work [3] suggested that $2p^2$ is the minimum number necessary; however, the experimental results pre-

sented here confirm that p^2 microphones arranged over Fliege grid provide adequate quadrature approximation. When less than p^2 microphones are used, spatial aliasing occurs and the scene structure can not be recovered correctly [7].

Also, in the present work auditory scene decomposition over plane-wave basis (i.e., in terms of $\mu(\mathbf{s})$ coefficients) was considered. However, once captured, it can be easily converted to the spherical wavefunction basis (i.e., in terms of C_n^m coefficients) using Gegenbauer expansion [7]. Thus, the developed algorithms are useful not only for localization but also for capturing spatio-temporal acoustic field representation for headphone playback as in [10].

7. CONCLUSIONS AND FUTURE WORK

This experimental study has showed that with the Fliege grid based microphone arrangement the spherical and hemispherical arrays operate successfully up to the spatial aliasing frequency and that no significant performance differences are found between K-method and M-method within the array effective frequency band, assuming that regularization is applied for K-method and the truncation number is adjusted appropriately for M-method. In fact, the effective frequency band is even wider than can be expected from the theoretical requirement of having $2p^2$ microphones for exact integration. Also, somewhat surprisingly, it was observed in simulations that both algorithms are very robust against random noise. Future work will use the developed algorithms in virtual audio applications.

8. REFERENCES

- [1] T. Abhayapala and D. Ward (2002). “Theory and design of high order sound field microphones using spherical microphone array”, Proc. IEEE ICASSP 2002, Orlando, FL, vol. 2, pp. 1949-1952.
- [2] J. Meyer and G. Elko (2002). “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield”, Proc. IEEE ICASSP 2002, Orlando, FL, vol. 2, pp. 1781-1784.
- [3] B. Rafaely (2005). “Analysis and design of spherical microphone arrays”, IEEE Trans. Speech Audio Proc., vol. 13(1), pp. 135-143.
- [4] Z. Li and R. Duraiswami (2005). “Hemispherical microphone array for sound capture and beamforming”, Proc. IEEE WAS-PAA 2005, New Paltz, NY, pp. 106-109.
- [5] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov, and L. S. Davis (2005). “High-order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues”, Proc. AES 119th Conv., New York, NY, preprint #6540.
- [6] B. Rafaely (2004). “Plane-wave decomposition of the sound field on a sphere by spherical convolution”, J. Acoust. Soc. Am., vol. 116(4), pp. 2149-2157.
- [7] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov (2005). “Plane-wave decomposition analysis for the spherical microphone arrays”, Proc. IEEE WASPAA 2005, New Paltz, NY, pp. 150-153.
- [8] J. Fliege and U. Maier (1999). “The distribution of points on the sphere and corresponding cubature formulae”, IMA Journal on Numerical Analysis, vol. 19(2), pp. 317-334.
- [9] N. A. Gumerov and R. Duraiswami (2005). “Fast multipole methods for the Helmholtz equation in three dimensions”, Elsevier, The Netherlands.
- [10] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov (2007). “Efficient conversion of X.Y surround sound content to binaural head-tracked form for HRTF-enabled playback”, Proc. IEEE ICASSP 2007, Honolulu, HI, vol. 1, pp. 21-24.