NEAR-FIELD SPHERICAL MICROPHONE ARRAY FOR SPEECH

Etan Fisher[1] *and Boaz Rafaely*[2]

Department of Electrical and Computer Engineering
Ben-Gurion University of the Negev , Beer-Sheva 84105, Israel.
[1]fisher@ee.bgu.ac.il, [2]br@ee.bgu.ac.il

ABSTRACT

When close enough to a microphone array, the spherical nature of radiating sources allows for sound field processing in terms of distance as well as direction. As part of an on-going study on beamforming given sources close to a spherical microphone array, radial filters have been developed for attenuating near-field and far-field interference in the source direction. This paper discusses the use of the radial filters for the case of speech. A simulation is presented to demonstrate the desired results. Estimation of near-field source direction and distance is presented for speech signals. Finally, experiments are presented demonstrating radial filtering capabilities given speech signals.

1. INTRODUCTION

Following a growing interest in spatial sound processing, the spherical microphone array has become the subject of substantial study. The majority of microphone array studies relate to sources distant enough from the array to be considered plane-waves. However, when sources are close to the array, such as for the case of close-talk speech [1], or higher order near-field ambisonics [2][3], the far-field assumption may not hold, and could lead to design or analysis errors. Furthermore, in the near-field, the spherical wave-front of near-field sources includes important spatial information which may be utilized for better spatial separation and design [4]. Therefore, in the near-field, it is beneficial to assume point sources, or an integration of point sources.

A recent study discussed the possibility of using a low order spherical microphone array as a robust close-talk microphone [1]. Methods for estimating the distance and direction of a near-field source were proposed. In [4], the behavior of the sound field due to point sources close to the array was analyzed. It was shown that near-field processing allows radial, as well as directional separation. Later, a method was presented for radial separation of sound sources propagating from the same direction using radial filters [5]. Several radial filtering techniques have been developed for processing sources within the near-field of the spherical array [6]. One radial filtering technique bases the polynomial design on the Dolph-Chebyshev beampattern. This beampattern, applied along the radius of the array, attenuates far-field interference given a desired source close to the array surface. Another potentially useful filter is the radial notch filter. This is a simple first order notch. The notch can be applied for attenuating near-field point-sources with little attenuation of other near-field or far-field sources propagating from the same direction. The radial filter has been shown to be robust, even when the desired source is in the far-field.

This research was supported by the Israel Science Foundation (ISF grant no. 155/06).

This study demonstrates the performance of the radial filters. Experimentation was carried out using an mhacoustics em32 (large version), a 32-microphone spherical array with a radius of 8.4cm. This array is capable of processing the sound field up to an order of 3 or 4, depending on the signal frequency.

Results demonstrate spherical microphone array radial filtering capabilities given sources in the array near-field, theoretically and for real audio signals. This further indicates the potential of the spherical microphone array as a means for true 3D recording.

2. NEAR-FIELD SPHERICAL MICROPHONE ARRAY PROCESSING

Consider a sound field with pressure denoted by $p(k, r, \theta, \phi)$, where k is the wave number, r is the radial distance and (θ, ϕ) is the direction in standard spherical coordinates [7]. The spherical Fourier transform of the pressure is given by [8]:

$$p_{nm}(k, r) = \int_{\Omega \in S^2} p(k, r, (\theta, \phi)) Y_n^{m^*}((\theta, \phi)) d(\theta, \phi) \ , \quad (1)$$

with the inverse transform relation:

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

where $Y_n^m(\theta, \phi)$ is the spherical harmonic [8] of order n and degree m. The sound field due to a point source located at $\mathbf{r}_s = \{r_s, \theta_s, \phi_s\}$ and emitting a signal $s(k)$ is [8] :

$$p(r_s, \theta_s, \phi_s) = s(k) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} b_n^s(kr, kr_s) Y_n^{m*}(\theta_s, \phi_s) Y_n^m(\theta, \phi) \ , \quad (3)$$

where $b_n^s(kr, kr_s)$ is [1] :

$$b_n^s(kr, kr_s) = i^{-(n-1)} k b_n(kr) h_n(kr_s) \ , \quad (4)$$

$b_n(kr)$ depends on the sphere boundary and is, for open and rigid sphere configurations:

$$b_n(kr) = 4\pi i^n \begin{cases} j_n(kr), & \text{Open sphere} \\ j_n(kr) - \frac{j_n'(ka)}{h_n'(ka)} h_n(kr), & \text{Rigid sphere} \end{cases} \quad (5)$$

$a < r$ is the sphere radius, $j_n(kr)$ is the spherical Bessel function, and $h_n(kr_s)$ is the spherical Hankel function which expresses the radial decay of the spherical wavefront. The spherical Fourier transform of (3) is :

$$p_{nm}(r_s, \theta_s, \phi_s) = s(k) b_n^s(kr, kr_s) Y_n^{m^*}(\theta_s, \phi_s) \ . \quad (6)$$

The array order, N, is limited by the number of microphones, typically $(N+1)^2 \leq M$ [9]. Therefore, the array output may be written as :

$$y(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} w_{nm}^*(k) p_{nm}(k) \ , \tag{7}$$

with w_{nm}^*, the complex conjugate of the spherical Fourier transform of the array spatial weighting function, $w(\theta, \phi)$, controlling the beam-pattern. Given an array of radius a and assuming r_s is known, the sound field can be separated from the array parameters by choosing [1] :

$$w_{nm}^* = d_n \frac{1}{b_n^s(ka, kr_s)} Y_n^m(\theta_l, \phi_l) \ , \tag{8}$$

where (θ_l, ϕ_l) is the look direction and d_n are the beamforming coefficients . Applying the spherical harmonics addition theorem [8], yields,

$$y = s(k) \sum_{n=0}^{N} d_n \frac{2n+1}{4\pi} P_n(\cos\Theta) \ , \tag{9}$$

where Θ is the angle between (θ, ϕ_s) and (θ_l, ϕ_l). When $d_n = 1$ and $N \to \infty$, this yields a directional impulse which signifies ideal separation.
Assuming r_s is not known, a more general weighting function is:

$$w_{nm}^* = d_n(k) \frac{1}{i^{-(n-1)} k b_n(ka)} Y_n^m(\theta_l, \phi_l) \ , \tag{10}$$

which yields :

$$y_N(k, r_s, \Theta) = s(k) \sum_{n=0}^{N} d_n(k) h_n(kr_s) P_n(\cos\Theta) \ . \tag{11}$$

The frequency-dependent coefficients, $d_n(k)$, can be chosen for design of the directional beampattern. This beampattern has been studied in [4] and is radially dependent with a decaying nature. Alternatively, the coefficients may be used to control the radial behavior of the array output. Focusing on radial design in the look direction, $\Theta = 0$, the array output, (11), becomes :

$$y_N(k, r_s) = s(k) \sum_{n=0}^{N} d_n(k) h_n(kr_s) \ . \tag{12}$$

3. RADIAL FILTERING

The radial behavior of the sound field close to the spherical array is determined by the spherical Hankel function. Assuming $s(k) = 1$, (12) can be written explicitly as a polynomial using the the spherical Hankel function [7]:

$$y(x) = h_0(x) \sum_{n=0}^{N} d_n(k) \sum_{m=0}^{n} g_m^n x^{-m} \ , \tag{13}$$

where

$$g_m^n = \frac{i^{(m-n)}}{m! 2^m} \frac{(n+m)!}{(n-m)!} \ , \tag{14}$$

$x = kr_s$ and $h_0(x) = \frac{e^{ix}}{ix}$. Note the amplitude of $h_0(x)$ expresses the natural $1/r_s$ decay of a point source at $\mathbf{r}_s$. By expanding the spherical Hankel function into its polynomial form, a polynomial-based filter can be designed along the radius of the array. Given the array order, any polynomial of this order can be applied to the radial filter. The coefficient vector is calculated directly:

$$\mathbf{d} = (\mathbf{C}^{-1})^T \mathbf{q} \ . \tag{15}$$

where

$$\mathbf{C} = \begin{bmatrix} c_0^0 & 0 & \cdots & 0 \\ c_1^0 & c_1^1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ c_N^0 & c_N^1 & \cdots & c_N^N \end{bmatrix} \ , \tag{16}$$

$c_m^n(k) = g_m^n \cdot (ka)^{-m}$ [6]. $\mathbf{q} = [q_0, \ q_1, \ \cdots, \ q_N]^T$ are the coefficients of the polynomial $\sum_{n=0}^{N} q_n z^n$ and

$$z = \frac{ka}{x} \ . \tag{17}$$

Transformation from x to z was performed to overcome the negative orders of x in the Hankel function expansion (13).

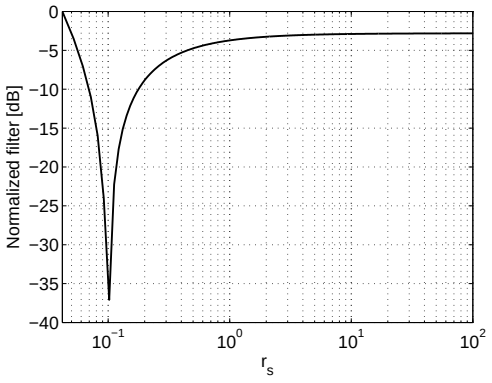


Figure 1: Radial notch filter for suppressing source r_0.

3.1. Low order radial notch filter

A simple first order notch filter can be applied for attenuating near-field point-sources with little attenuation of other near-field or far-field sources propagating from the same direction. Consider a single notch at $r_s = r_0$. The corresponding value of z is $z_0 = \frac{a}{r_0}$ yielding the first order polynomial

$$y_N(z) = z - z_0 \ . \tag{18}$$

Figure 1 shows the radial notch filter given a notch at $r_0 = 0.1 \ m$.

3.2. Dolph-Chebyshev radial filter

The Dolph-Chebyshev beampattern enables control over the trade-off between side lobe level and main lobe width. In [4] the directional beampattern was transformed into a radial design. Applied along the radius of the array, the Dolph-Chebyshev beampattern attenuates far-field interference given a desired source close to the array surface. The radial filter coefficient vector is:

$$\mathbf{d} = \frac{1}{R} (\mathbf{C}^{-1})^T \mathbf{X}_0 \mathbf{t} \ . \tag{19}$$

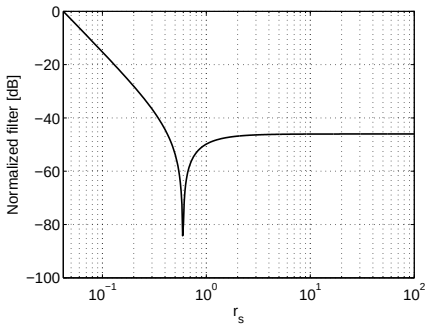


Figure 2: Radial Dolph-Chebyshev filter for suppressing far-field interference.

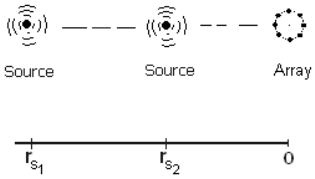


Figure 3: Speech sources located in the same direction.

where $\mathbf{d} = [d_0, \ d_1, \ \cdots, \ d_N]^T$. R, $\mathbf{X}_0$ and $\mathbf{t}$ are predetermined Dolph-Chebyshev design parameters[6]. Figure 2 shows the radial pattern of a second-order radial Dolph-Chebyshev filter for $R = 200$.

4. RADIAL FILTERING SIMULATION

A simulation was performed to demonstrate near-field source suppression using a radial notch filter and far-field source suppression using a radial Dolph-Chebyshev filter. Two speech sources located in the same direction relative to the array were simulated (see Figure 3). The first speech source, uttering the word 'one', was located at $r_{s1} = 0.1 \ m$. The second speech source, uttering the word 'two', was located at $r_{s2} = 1 \ m$. Given these source distances, source 'two' is naturally attenuated by a factor of 10 (20 dB). In order to emphasize radial filter performance, source 'two' was given a ten-times greater amplitude. Source direction was chosen to be $(\theta, \phi) = (\pi/2, 0)$.
Radial separation can be achieved for both signals spoken simultaneously. However, for display purposes, the signals are displayed one after the other. Figure 4 shows the speech signals as they would be picked up by a single microphone located at the array center. The source distance is estimated using the method of [1]. It was shown that, at low frequencies, the source distance is approximately :

$$r_L = \frac{a}{2} \frac{b_0^s}{b_1^s} \ . \tag{20}$$

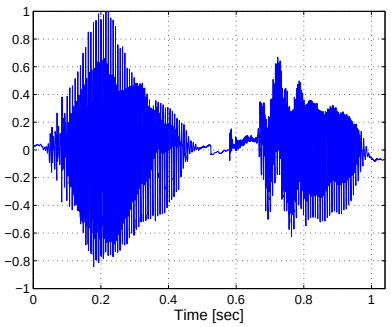


Figure 4: The signals as they would be picked up by a single microphone at the array center.

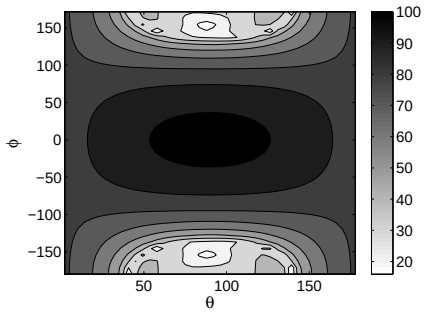


Figure 5: Directional contour plot for source at $r_{s1} = 0.1 \ m$.

In terms of the spherical Fourier transform of the pressure, this is equal to :

$$r_L = \sqrt{\frac{3}{4} a^2 \frac{|p_{00}|^2}{\sum_{m=-1}^{1} |p_{1m}|^2}} \ . \tag{21}$$

Given the theoretical point sources of the simulation, both r_{s1} and r_{s2} were estimated exactly. Source direction was estimated by applying a directional radial compensation filter [6]. Figure 5 shows a contour plot of the array output for source 'one'. The equivalent plot for source 'two' yields the same direction, but is slightly wider.
In order to attenuate the distant source, 'two', a Dolph-Chebyshev filter can be applied. The radial Dolph-Chebyshev filter appearing above in Figure 2 was used. The array output appears in Figure 6. Source 'two' has been significantly attenuated.
In order to attenuate the near-field source 'one' without compromising source 'two', a radial notch filter is applied. The radial notch filter of Figure 1 is used. Figure 7 shows the array output. This time, source 'one' has been significantly attenuated.

5. NEAR-FIELD AUDIO ANALYSIS

In order to examine the possibility of applying the radial filters to real speech, an experiment was carried out in an anechoic chamber. Initially a wide-band chirp signal was played through an omni-directional loudspeaker at various distances close to the spherical microphone array (at the array surface, at 20cm, and at 50cm). In order to separate between the actual signal and the

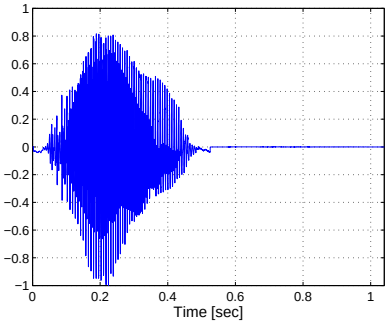


Figure 6: Array output signal after radial Dolph-Chebyshev filter.

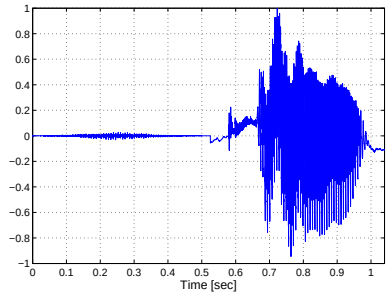


Figure 7: Array output signal after radial notch filter.

spatial source model, the signal was estimated from the zero-order spherical harmonic coefficient, using (6) and assuming a point source :

$$\hat{s}(k) = \frac{p_{00}}{b_0^s Y^{*0}_0(\theta_s, \phi_s)} \ . \tag{22}$$

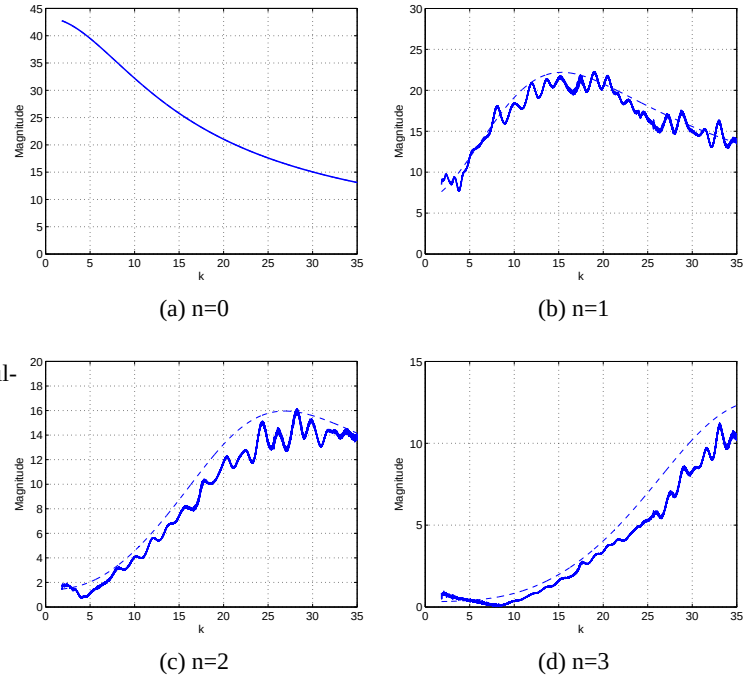
The radial sound field components, b_n^s, were estimated using

$$\hat{b}_n^s(k, a, r_s) = \frac{p_{nm}}{\hat{s}(k) Y^{*m}_n(\theta_s, \phi_s)} \ . \tag{23}$$

In order to realize equation (12) given the source sound field, the $h_n(kr_s)$ component was replacd by dividing $\hat{b}_n^s$ by $i^{-(n-1)}kb_n(ka)$.The estimated sound field components were compared to the theoretical sound field components of point sources. Figure 8 shows the amplitude comparison for a source $20\ cm$ from the array surface ($r_s = 0.284\ m$). The amplitude is similar, but a phase shift exists which varies with the order, n, and with kr_s.

6. RADIAL FILTERING OF SPEECH

In the second experiment, a speaker was sampled uttering a series of digits between 3 and 90 centimeters from the array, in the same direction. Due to the non-stationary nature of the speech signals, Welch-based spectral averaging was used for estimating the radial sound field components [10]. Assuming the source distance to be unknown, the following process was performed:
(a) An initial guess of the source signal was taken using one of the array microphones.
(b) $\hat{b}_n^s$ was estimated given this source signal and equation (23), and assuming $r_s = a$.



(a) n=0 (b) n=1

(c) n=2 (d) n=3

Figure 8: Amplitude of $\hat{h}_n(kr_s)$ (solid) vs. $h_n(kr_s)$ (dashed) for a source at $r_s = 0.284\ m$.

(c) $\hat{r}_s$ was estimated from equation (20). Since a low frequency is required, the fundamental frequency of the estimated signal is used (in this experiment $\sim 180\ Hz$). This limits the distance of sources that can be estimated correctly, as can be seen from Figure 9. This figure shows the estimate of $\Delta\hat{r}_s$ versus Δr_s, where the Δ signifies the distance from the array surface rather than the distance from the array center.
(d) Estimation of $\hat{b}_n^s$ is performed given $\hat{r}_s$.
Figure 10 shows the mode strength amplitude estimate $\hat{h}_n(k\hat{r}_s)$ compared to $h_n(k\hat{r}_s)$. The estimate was similar for all sources.
In order to demonstrate radial filtering of speech, a radial notch filter at the array surface was applied to the speech signals. Phase was synthesized from $h_n(k\hat{r}_s)$. The array output appears in

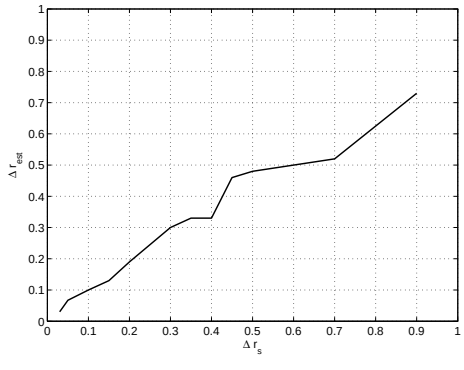
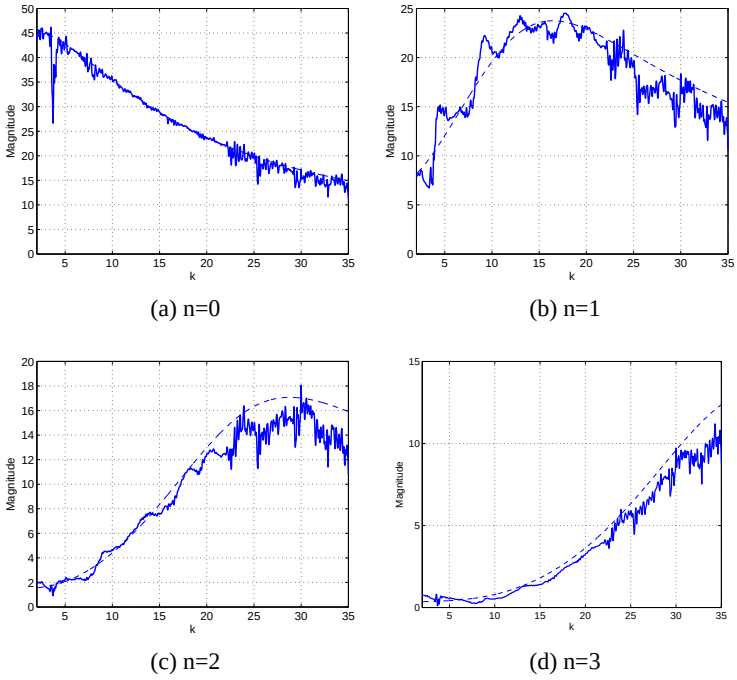


Figure 9: Source distance estimation for speech signals between 3 and 90 centimeters from the array surface.



(a) n=0 (b) n=1

(c) n=2 (d) n=3

Figure 10: Amplitude of $\hat{h}_n(k\hat{r}_s)$ (solid) vs. $h_n(k\hat{r}_s)$ (dashed) for $r_s = 0.284\ m$.

Figure 11. The notch significantly attenuates near-field sources relative to the far-field. Although this experiment did not fully re-create the simulation scenario (simultaneous sources), it indicates there is enough information which may be used for radial separation of real speech signals.

7. CONCLUSION

A simulation and experiments were carried out to examine the possibilities of using radial filters on real speech. Direction estimation was performed using a radial compensation filter. Then, existing methods for point source distance estimation were adapted for real speech. Finally, radial sound field components were estimated and used for demonstrating radial filtering. This paper showed the point source assumption is acceptable for near-field speech and demonstrated radial filtering on real speech signals.

8. REFERENCES

[1] J. Meyer and G. W. Elko, "Position independent close-talk microphone," *Signal Processing*, 2005.

[2] J. Daniel, "Spatial sound encoding including near field effect," in *Proc. AES 23nd International Conference.*, 2003.

[3] S. Favrot and J. Buchholtz, "Distance perception in loudspeaker-based room auralization," in *Proc. 127th AES Convention.*, 2009.

[4] E. Fisher and B. Rafaely, "The near-field spherical microphone array," in *Proceedings of the IEEE International Conference on Acoustics and Speech Signal Processing (ICASSP'08)*, 2008.

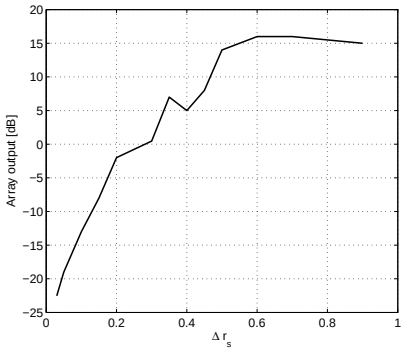


Figure 11: Radial notch array output for speech signals at Δr_s.

[5] ——, "Dolph-chebyshev radial filter for the near-field spherical microphone array," in *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA09)*, 2009.

[6] ——, "Near-field spherical microphone array with radial filtering," *IEEE Transactions on Speech, Audio and Language Processing*, 2010, in press.

[7] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 6th ed. Oxford: Elsevier Academic Press, 2005.

[8] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. New York, NY: Academic Press, 1999.

[9] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.

[10] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio Electroacoustics*, vol. 15, pp. 70–73, 1967.