DESIGN AND ANALYSIS OF MINIATURE AND THREE TIERED B-FORMAT MICROPHONES MANUFACTURED USING 3D PRINTING

Matthew Dabin, Christian Ritz and Muawiyath Shujau

School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, Wollongong, NSW, Australia

ABSTRACT

This paper describes miniature and three-tiered B-format microphone array designs for accurate sound source localisation that are manufactured using 3D printing and MEMs devices. The implications of pressure gradient resolution reduction in spatial-temporal sampling on the accuracy of Direction of Arrival (DOA) estimation is analysed through simulated room impulse response measurements and characterized by the directional signal to noise ratio. It is shown how the capsule spacing can be optimally chosen based on microphone capsule sensitivity and the required DOA accuracy. Through this method a new three tiered B-format microphone array is proposed, where each tier optimally records 3D sound for a given frequency sub-band to achieve highly accurate DOA estimation for the full audible frequency range of 50 Hz to 20 kHz.

Index Terms— *B-format, 3D localization, microphone array, direction of arrival, 3D printing*

1. INTRODUCTION

The need for highly accurate sound source Direction of Arrival (DOA) estimation persists for many types of speech and audio processing applications. This is typically achieved using a microphone array [1]. The B-format microphone array is very practical due to its compact design, four channel processing, configurable polar response and standardized B-format decoding to surround sound. Similar to other differential microphones, the B-format microphone estimates the pressure gradient of a sound wave based on processing of pairs of closely spaced microphones. The physical separation between any two microphones limits the highest attainable error free frequency, f_{err}, that can be achieved before spatial aliasing occurs [2] and can be described by:

$$f_{err} = \frac{c}{2d} \tag{1}$$

where c is the speed of sound and d is capsule separation.

Typically the inter-capsule separation used for B-format microphones is 1.47 cm as suggested by Michael Gerzon in 1978 [3]. This achieves error free pressure gradient measurement of up to 11.6 kHz covering the majority of the speech frequency range. Estimating the pressure gradient via differential processing results in a high pass frequency effect, where the maximum pressure difference reduces relative to wave length, λ, and characterized by:

$$\Delta = \frac{2d}{\lambda} \tag{2}$$

Hence, at low frequencies, the pressure gradient has a small magnitude which can be compensated using gain equalization. However, this results in a Signal to Noise Ratio (SNR) that depends on the wavelength and leads to poor DOA estimation accuracy [3] at low frequencies and therefore limits the frequency range for which reliable DOA estimates can be obtained.

This paper proposes an optimal design of two types of B-format microphones using 3D printing and MEMS microphones: a miniature design occupying a volume of only 1 cm³ and a three-tiered tetrahedral design where each tier is designed to optimally record sound for a given frequency sub-band. Section 2 describes B-Format microphones and derives the corresponding Directional Signal to Noise Ratio (DSNR). Section 3 presents simulation results for the designs and analyses the impact of microphone separation on DSNR and DOA estimation accuracy. Section 4 presents the manufactured designs while conclusions are presented in Section 5.

1.1. Relation to Prior Work

One approach to improving the 3D sound processing performance across a broad range of frequencies is the use of a dual concentric spherical microphone array (SMA) [4]. This achieved higher pressure gradient readings for a large frequency range than other microphone arrays using an approach that optimally chose selections of microphones based on the sub-band of interest. This resulted in highly accurate spatial recording but requires a significant number of microphones (64 in total). In [5] the authors proposed an alternative approach based on four microphones arranged on a rigid object that exploited acoustic shadowing to achieve improved DOA accuracy over a wider frequency range. Their approach required a cylinder of approximately 8 cm diameter. In contrast, this paper first examines and compares a much smaller microphone array (1 cm³) as well as an array of similar size (8 cm diameter) but using microphones arranged in three concentric tetrahedrons. This second design has some relation to [6] but utilizes an additional tetrahedron and is designed and manufactured using 3D printing technology to enable flexible and highly accurate placement of microphones on their structural support. A more theoretical study examining the noise statistics of acoustic gradient sensors formed from differential microphones is described in [14].

2. B-FORMAT MICROPHONE SIGNAL ANALYSIS

This section derives the directional signal to noise ratio (DSNR) for a tetrahedron microphone array. In this analysis the additive noise for each sensor is considered uncorrelated and proportionate to the original SNR for each sensor. Consider the tetrahedral microphone array of Figure 1, with the origin located at the center of the prism such that a vector from the center to any sensor has equal magnitude.

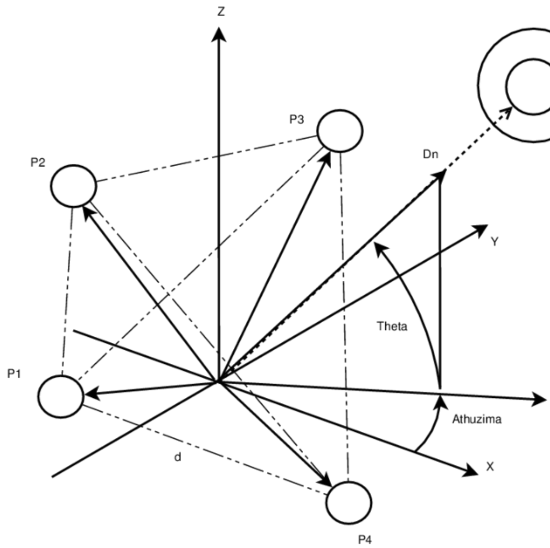

Figure 1: Tetrahedron microphone capsule configuration

If the distance d is the separation between any two capsules, then the sensor's 3D vectors as (used in the Section 3) are defined by:

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix}^T = d \times \begin{bmatrix} -1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \\ -\sqrt{3}/2 & \sqrt{3}/2 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix} \quad (3)$$

The n^{th} source vector as a function of azimuth θ_n and elevation ϕ_n is defined as:

$$d_n = [\cos(\theta_n)\cos(\phi_n) \quad \sin(\theta_n)\cos(\phi_n) \quad \sin(\phi_n)]^T \quad (4)$$

Considering a single source and disjoint frequencies using a short time window, the source signal model can be simplified. Where ω frequency and t is time sample. The four sensors can then be represented using the pressure $P_0(\omega, t)$ of the source, the impulse response of the acoustic path from source to the sensor n, $h_n(\omega, t)$ and the self-noise of the sensor $w_n(\omega, t)$.

$$P_n(\omega, t) = P_0(\omega, t) * h_n(\omega, t) + w_n(\omega, t) \quad (5)$$

Using the B-format equations [7] the directional components can then be calculated:

$$X(\omega, t) = P_0(\omega, t)\big(h_1(\omega, t) - h_2(\omega, t) + h_3(\omega, t) - h_4(\omega, t)\big) + w_n(\omega, t) \quad (6)$$

$$Y(\omega, t) = P_0(\omega, t)\big(h_1(\omega, t) + h_2(\omega, t) - h_3(\omega, t) - h_4(\omega, t)\big) + w_n(\omega, t) \quad (7)$$

$$Z(\omega, t) = P_0(\omega, t)\big(h_1(\omega, t) - h_2(\omega, t) + h_3(\omega, t) - h_4(\omega, t)\big) + w_n(\omega, t) \quad (8)$$

$$W(\omega, t) = \frac{1}{\sqrt{2}}(P_0(\omega, t)\big(h_1(\omega, t) - h_2(\omega, t) + h_3(\omega, t) - h_4(\omega, t)\big) + w_n(\omega, t)) \quad (9)$$

The microphone is assumed uncorrelated between capsules such that the total spectral energy density of the noise remains constant.

As shown in [8] the resulting phase difference $h_i(\omega, t) - h_j(\omega, t)$ can be derived using 3D vector coordinates, between any pair of microphones $p_i(d)$ and $p_j(d)$ for source direction $dn(d)$ as:

$$\delta(\omega, d) = \frac{2\pi f}{s}\left(p_i(d) - p_j(d)\right)dn(d) \simeq \angle\frac{p_i(d)}{p_j(d)} \quad (10)$$

In (10) $\angle$ represents the phase difference. The Directional Signal to Noise Ratio (DSNR) of each of the gradient component (X, Y and Z) of (6)-(8) can be then characterized by the new directional output and the self-noise, for a source position θ, ϕ:

$$DSNR(\omega, t, d, \theta, \phi) = \log_{10}\left(\frac{P_0(\omega, t) * \left(\delta_i(\omega, d) + \delta_j(\omega, d)\right)}{w(\omega, t)}\right)^2 \quad (11)$$

where $w(\omega, t)$ is the uncorrelated self-noise of the microphone, $P_0(\omega, t)$ is the pressure of sound source and δ_i is the resulting pressure difference of the first capsule pair and δ_j is the resulting pressure difference of the second capsule pair as used in (6) to (9). Low frequencies, corresponding to longer wavelengths, result in lower pressure differences than higher frequencies. Since the self-noise is uncorrelated, its magnitude remains constant following pressure difference calculation and hence from (12) this results in poorer DSNR values at low frequencies. Hence, there will always be a trade-off between high frequency aliasing and low frequency directivity. This is also related to the sensitivity (SNR) rating of the microphone capsules, where at longer wavelengths the magnitude of the pressure difference becomes close to the magnitude of the noise and hence cannot result in reliable pressure differences. This affects DOA estimation accuracy based on the ratio of *X*, *Y* and *Z* signals of (9) – (11) as used in Section 3 and existing research [9].

3. SIMULATION

This section describes the simulation methodology and results for the proposed B-microphone array designs.

3.1. Setup and Measurement

The array designs are evaluated here by examining the microphone polar patterns and sound source DOA estimation accuracy from directivity calculations. The Matlab based MCRoomSim [10] was used to simulate the acoustic environment as it supports both non conformal array configurations and accurately processes phase information of the acoustic paths [10]. A single sinusoidal tone is generated using the defined frequencies and filtered with the impulse response. Using the sensors noise rating in dB(a), white Gaussian distributed noise is added to the pressure signals. The B-format equations (6-9) are used to calculate the gradient signals. Source azimuths ranged from 0° to 360° in 5° intervals to obtain a polar response and directivity of the directional signals. The measured angle is calculated using directional intensity vectors from the pressure gradient signals [9].

$$\theta_{n,m} = atan^{360°}\left(\frac{X_i}{Y_i}\right) \quad (12)$$

For DOA estimation performance, the Average Angular Error (AAE) was estimated as:

$$AAE = \frac{1}{N}|\theta_{n,m} - \theta_{n,a}| \quad (13)$$

The AAE was calculated for source frequencies ranging from 62.5 kHz to 20 kHz, where $\theta_{n,m}$ is the measured angle, $\theta_{n,a}$ is the actual angle and N is the number of samples [11].

3.2. Results: Gradient Reponses

In the first set of simulations the directional polar response is a function of frequency and inter-capsule distances. The sensors SNR_{dB} was configured to 59 dB(a), as this is the rating value of

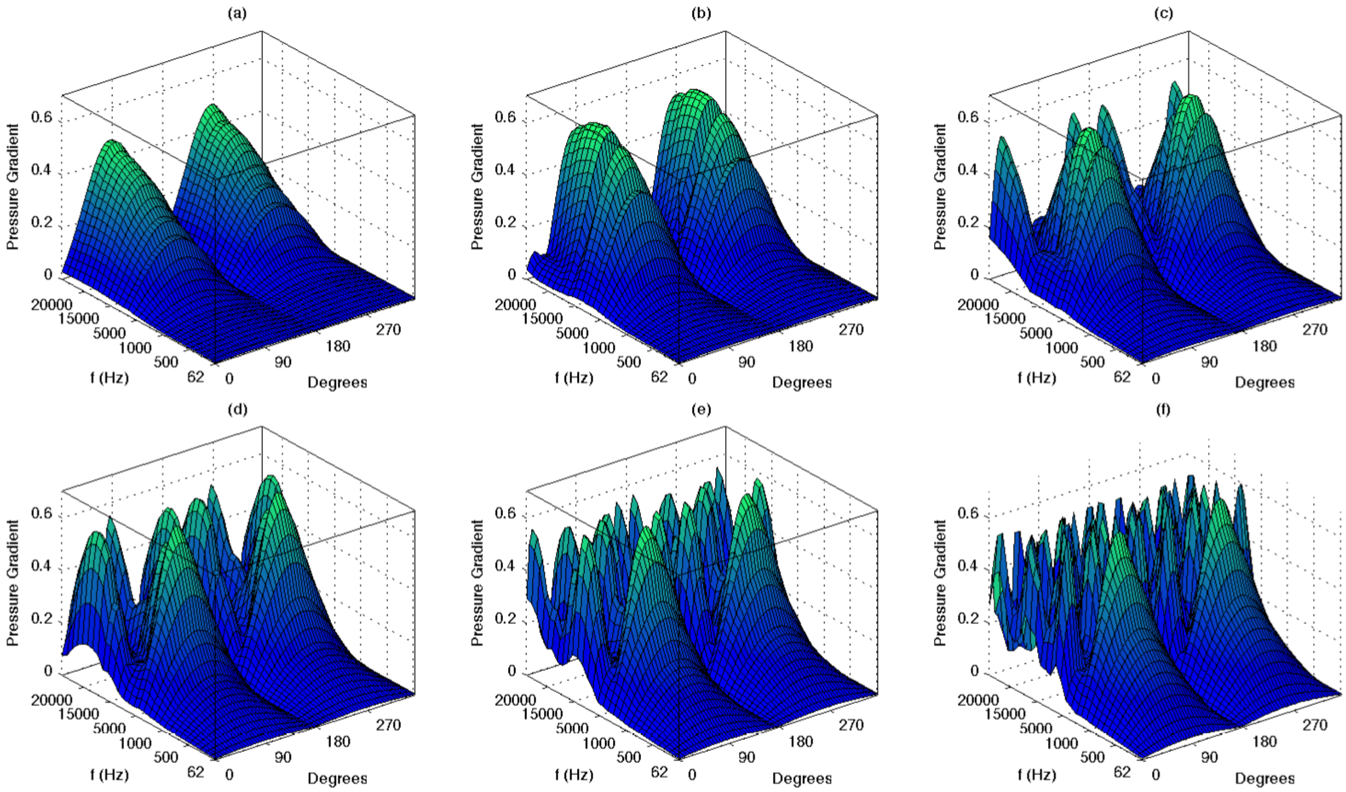

Figure 2: 3D surface plot of simulated X-axis pressure gradient signals at a range of capsule separations. (a) $8\ mm$, (b) $15\ mm$, (c) $25\ mm$, (d) $35\ mm$, (e) $50\ mm$, (f) $80\ mm$. This shows the X directional sensitivity from $62.5\ Hz$ to $20\ kHz$.

the MEMS sensors used in physical arrays (Knowles analogue omnidirectional microphones)[12].

Multiple inter-capsule distances were simulated to test the effect of spatial reduction as a function of directional sensitivity and frequency. Figure 2 shows that greater spatial distances result in better lower frequency directional sensitivity as theory states. Microphone arrays inter-capsule separations often vary from 1.2 cm to 8 cm [1, 6]. Literature [13] suggests a separation of 1.47 cm as it captures error free frequencies up to 11.6 kHz which is the important range of human perceptual hearing [8]. This is shown in Figure 2 (b), due to the high error free range it also results in very low sensitivity at longer wave lengths (low frequency). In Figure 2 (f) for a capsule separation distance of 80 mm the on-axis peaks are well formed and have a high sensitivity at low frequencies but as expected for frequencies over 2.1 kHz aliasing starts to occur as indicated by noise like peaks in this frequency region.

3.3. Results: AAE Dimension Reduction

Figure 4 shows the AAE results for multiple simulated tetrahedron arrays with different capsule separation distances. The points on each line indicate when aliasing errors have started to occur (values beyond this become very large and are omitted for clarity). It can be seen that for each inter capsule distance there is an optimal frequency range for source localization. Larger capsule separation distances result in better lower directivity pattern and as a result better localization accuracy at lower frequencies. It can be noted that the point at which the AAE is minimum can be approximated by:

$$\mathrm{AAE_{min}} \cong \frac{\mathrm{s}}{6\mathrm{d}} \cong \frac{\mathrm{F_{null}}}{3} \tag{14}$$

where $\mathrm{f_{err}}$ is given in (1). For frequencies below this point the low DSNR starts to affect the localization accuracy where off-axis DOA estimation errors become larger. As an observation directivity patterns start to begin to rotate and elongate above $\mathrm{AAE_{min}}$ which causes the localization errors.

3.4. Results: AAE SNR Reduction

The following simulation evaluates the effect of microphone self-noise on the accuracy of source localization. Figure 5 shows the results as a function of frequency and AAE for a range of SNR's using a fixed capsule separation size of 8 mm, which are the limitations of the 3D printing technology used to manufacture the array. The microphone array exhibits highly accurate results around frequencies given by $\mathrm{f_{err}}/2$ to $\mathrm{f_{err}}/3$ for this array.

Elongated polar patterns (omitted for brevity) were observed of 2 kHz to 16 kHz. For an accurate DOA estimation below 2° of the important speech spectrum below 12 kHz the miniature array would require the microphone capsules to have a self-noise rating above 80 dB(a).

3.5. Three Tiered Concentric Tetrahedral Array

Results of the previous simulations show that there are limitations when using a single fixed array. From the AAE results over a range of capsule separations, the dimensions of a 3 tier multi-dimensional concentric tetrahedron microphone array was chosen. Tier 1 has a capsule separation of 80 mm and targets frequencies from 100 Hz to 1 kHz. Tier 2 has a capsule separation of 35 mm

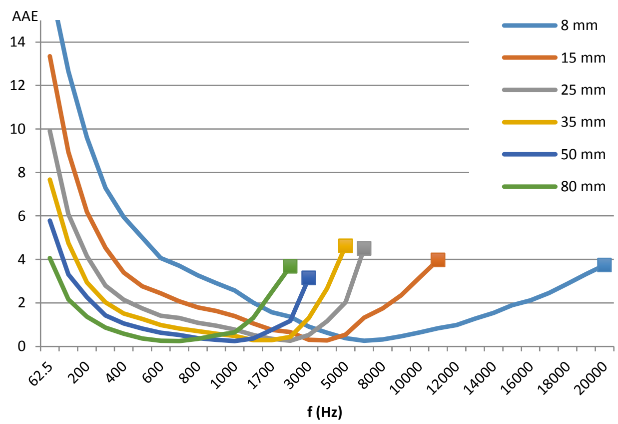

Figure 3: AAE Plot of the Simulated Tetrahedron Array with SNR of 59 $\mathrm{dB(a)}$

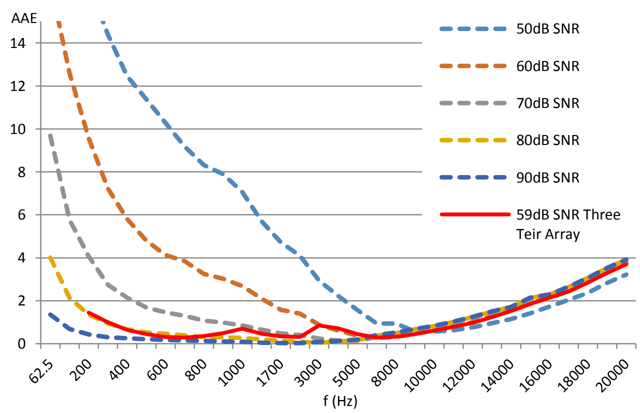

Figure 5: AAE Plot of Simulated Tetrahedron Array with a range of SNR values for 8 mm and three tiered array

and targets frequencies from 1 kHz to 3 kHz. Tier 3 has a capsule separation of 8 mm and targets frequencies from 3 kHz to 21 kHz.

Figure 5 shows the results of combining DOA estimates using these three tiers, with errors less than 1° for 250 Hz to 11 kHz. This result is something that cannot be achieved with a single array without an unrealistic SNR rating of microphones. The 3D surface map of the multi-dimensional array is shown in Figure 6, with each peak along the frequency axis signifying the boundary point of the arrays. There is a much higher linear sensitivity at lower frequencies compared to the fixed array.

4. DESIGNS

Here, 3D printing was used to manufacture the two microphone arrays. The printer used for development was a multi jet VU cured plastic printer with layer accuracy of 0.127 mm, which results in a reduced phase error at high frequencies. The miniature array prototype shown in Figure 3 aims to test the feasibility of 3D recording and source localization for small portable devices. As outlined previously the multi-dimensional array is designed so that different sets of microphones optimally record a particular frequency sub-band. The inner array design is of similar construction to the miniature array. The spatial distance was increased to 12 mm due to 3D printer manufacturing limitations of the second array. The second array uses a Type 2 tetrahedron array with inter-capsule distance of 35 mm. This

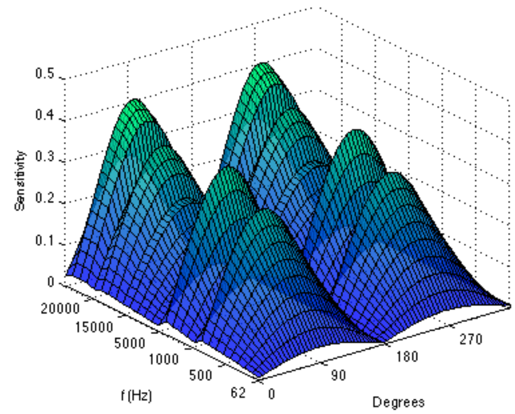

Figure 4: Multi-Tier Array X-Gradient Plot

Figure 7: 8 mm B-format Microphone (Scale in cm)

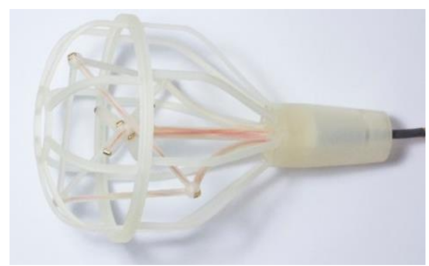

Figure 6: Three Tier B-format Microphone

microphone array is configured to avoid acoustic shadowing on the first array via the second array. This was designed in a single casing to keep element positions relative to each other. The third array consists of a hollow Type 1 tetrahedron array where the inter-capsule distance is 80 mm. The third array fits around the inner arrays and is secured in at the base. A casing is used to protect the sensor array. A benefit of this design is that the 12 channels can be filtered down to the standard four B-format signals.

5. CONCLUSION

This paper described the design of B-format microphones to ensure high directional sensitivity across a broad range of frequencies and implementation using 3D printing technology. The proposed miniature B-format microphone provided highly accurate localisation at high frequencies, but suffers from poorer directivity at low frequencies due to the relationship between inter-microphone spacing, signal wavelengths and self-noise. To overcome these limitations, a second design based on microphones located on three concentric type 1 and type 2 tetrahedrons was proposed. Simulation results showed that this microphone achieves highly accurate DOA estimation across a broad range of frequencies and is viable to manufacture. Future work will investigate the testing and use of this microphone for other applications such as sound source separation.

6. REFERENCES

[1] J. Merimaa, "Applications of a 3-D microphone array," in *Audio Engineering Society Convention 112*, 2002.

[2] J. Eargle, *The microphone book*: Taylor & Francis, 2004.

[3] B. Friedlanger, "On the Cramer-Rao bound for time delay and doppler estimation," *IEEE Transactions on Information Theory,* vol. 30, pp. 575-580, 1984.

[4] A. Parthy, C. Jin, and A. van Schaik, "Acoustic holography with a concentric rigid and open spherical microphone array," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 2173-2176.

[5] J. Ahonen, "Microphone front-ends for spatial sound analysis and synthesis with Directional Audio Coding," 2013.

[6] J. Ahonen and V. Pulkki, "Broadband direction estimation method utilizing combined pressure and energy gradients from optimized microphone array," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 97-100.

[7] K. Farrar, "Soundfield microphone," *Wireless World,* vol. 85, pp. 48-50, 1979.

[8] M. Cobos, J. J. Lopez, and S. Spors, "A sparsity-based approach to 3D binaural sound synthesis using time-frequency array processing," *EURASIP Journal on Advances in Signal Processing,* vol. 2010, p. 2, 2010.

[9] B. Gunel, H. Hacihabiboglu, and A. M. Kondoz, "Acoustic source separation of convolutive mixtures based on intensity vector statistics," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 16, pp. 748-756, 2008.

[10] A. Wabnitz, N. Epain, C. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *Proceedings of the International Symposium on Room Acoustics*, 2010.

[11] M. Shujau, C. Ritz, and I. Burnett, "Designing Acoustic Vector Sensors for localization of sound sources in air," *EUSIPCO 2009,* 2009.

[12] K. ACOUSTIC, ""Slim UltraMini" SiSonicTM Microphone Specification," in *SPQ0410HR5H-B datasheet*, ed, 2013.

[13] M. A. Gerzon, "The design of precisely coincident microphone arrays for stereo and surround sound," in *Audio Engineering Society Convention 50*, 1975.

[14] A. Y. Olenko & K. T. Wong, "Noise Statistics of a Higher Order Directional Sensor, Realized by Computing Finite Differences Spatially Across Multiple Isotropic Sensors," IEEE Transactions on Aerospace and Electronic Systems, vol. 49, no. 4, 2792-2798, October 2013.