Audio Engineering Society

Convention Paper 7091

Presented at the 122nd Convention
2007 May 5–8 Vienna, Austria

Sound source localization and B-format enhancement using soundfield microphone sets

Dimoulas C. A.[1], Avdelidis K. A.[2], Kalliris G. M. [3], and Papanikolaou G. V. [4]

[1] Aristotle University of Thessaloniki, Thessaloniki, 54644, Greece
babis@eng.auth.gr

[2] Aristotle University of Thessaloniki, Thessaloniki, 54644, Greece
avdel@auth.gr

[3] Aristotle University of Thessaloniki, Thessaloniki, 54644, Greece
gkal@jour.auth.gr

[4] Aristotle University of Thessaloniki, Thessaloniki, 54644, Greece
pap@eng.auth.gr

ABSTRACT

The current work focuses on the implementation of sound-field microphone arrays for sound source localization purposes and B-format enhancement. There are many applications where spatial audio information is very important, while reverberant sound-field and ambient noise deteriorate the recording conditions. As examples we may refer to sound recordings during movie production, virtual reality environments, teleconference and distance learning applications using 3D audio capabilities. B-format components, provided from a single soundfield microphone, are adequate to estimate sound source direction of arrival, while the combination of two soundfield microphones allows estimating the exact source location. In addition, the eight (or more) available signal components can be used to apply delay and sum techniques, enabling SNR improvements and virtual positioning of a signal B-format microphone to any desired place. Simplicity, reduced computational load and effectiveness are some of the advantages of the proposed methodology, which is evaluated via software simulations.

1. INTRODUCTION

The several different approaches in the field of spatial audio analysis have as common ground the usage of multiple signals originating from equal number of sensors. Those needs are distinguished in the localization of sound sources, the algorithms of signal enhancement and multi-channel recording and reproduction techniques according to stereophonic and peripheral sound models [1]-[3]. The methods used to accomplish the above tasks utilize more or less common algorithms and their combination is often needed [4]-[5]. In general, we may distinguish three basic approaches: direction of arrival / beam forming methods

[3], [5]-[7], time delay of arrival estimation [3]-[5], [8] and energy based localization [3], [6]-[11].

The sound source localisation is often accomplished by the use of microphone arrays, having as major application the improvement of the speech intelligibility (i.e. conference systems). Spatial sound processing mainly focuses in beam forming techniques, improving the total signal-to-noise ratio via the application of time delay compensation (TDC) techniques [2]-[6]. Energy based localization techniques offer the advantage of easy implementation and reduced resolution demands [3], [6], [9], [10].

In the above context, special interest presents the soundfield microphone and the ambisonics (also known as periphony), developed by the Gerzon [12]-[13] in the effort of complete spherical sound field recording and reproduction. The resulting acoustic pressure signal and the acoustic gradient pressure vector information for each direction (X, Y, Z) of the sound excitation are used to record a spatial sound image. This representation offers a flexible and fast way of tackling with spatial sound information, accentuating the surround sound recording as the main soundfield microphone application [12]-[18]. In addition, second order ambiosonics microphones have been already implemented by incorporating additional signal components to increase spatial audio processing capabilities [15]-[16].

The current paper tries to extend the potentials of the soundfield microphone usage in the areas of sound source localization and signal enhancement. With respect to these objectives, several topologies and setups are going to be discussed and evaluated in the following paragraphs.

2. B-FORMAT MODEL

The B-format model, known as the system of four components W, X, Y, Z, provides possibilities of spatial analysis mainly for creation of suitable "sonic image". In this case, the acquisition system is placed in the centre of supposed unitary sphere, while from the four components is possible the calculation of address of sound field (without is possible the finding of precise place of sound excitation). Thus, if ϕ is the horizontal angle in the XY level and θ is the angle of the Z-axis (elevation angle), then the following equations apply:

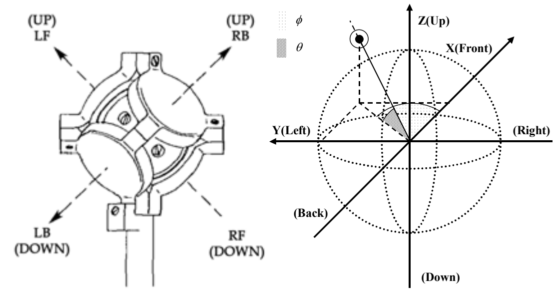


Figure 1 Soundfield microphone

$$\begin{aligned} p_X &= p_{LF} - p_{RB} + p_{RF} - p_{LB} = \cos\phi \cdot \sin\theta \cdot p_S \\ p_Y &= p_{LF} - p_{RB} - p_{RF} + p_{LB} = \sin\phi \cdot \sin\theta \cdot p_S \\ p_Z &= p_{LF} - p_{RB} - p_{RF} - p_{LB} = \cos\theta \cdot p_S \\ p_W &= p_{LF} + p_{RB} + p_{RF} + p_{LB} = 0.707 \cdot p_S \end{aligned} \quad (1)$$

Therefore, considering an isotropic model, the direction upon which the sound source is located can be calculated from the equations (1) according to the following:

$$\begin{aligned} \phi &= \tan^{-1}\left(\frac{p_Y}{p_X}\right) \qquad 0 \le \phi < 2\pi \\ \theta &= \tan^{-1}\left(\frac{p_Z}{\sqrt{p_X^2 + p_Y^2}}\right), -\frac{\pi}{2} \le \theta < \frac{\pi}{2} \end{aligned} \quad (2)$$

In addition, phase comparisons between each one of the axis components and the W reference signal may be used to determine the exact direction of arrival (the exact position of the ϕ, θ angles in the 0-2π interval) [1], [2], [4], [12]-[18].

It is important to mention that B-Format works as an entity rather than many individual signal components. In this context simple algebraic operations are adequate to produce various virtual auralization effects such as source / audience rotation and displacement [1], [2], [4]. These characteristics are useful in various spatial audio applications movie / film production, surround sound, virtual reality etc.

3. INVERSE SQUARE LAW MODEL

The simplest model of sound source localization concerns in the case far field, where the source is considered very small, emitting omni-directionally (isotropic source), while the space is free of obstacles (free sound field). Thus, the distribution is considered homogenous and linear, while the Inverse Square Low applies (*ISL*: *the sound intensity attenuates with the square of distance*) [9], [10].

$$I = k_p \frac{P_s}{4\pi d^2} \tag{3}$$

where the propagation constant k_p is equal to 1 in most sound propagation cases. Furthermore, the intensity is proportional with the square of sound pressure for the far field, obstacle free case:

$$I = \frac{p_{rms}^2}{\rho_o c} \tag{4}$$

In the cases where obstacles or confined areas with reflective surfaces apply, the components of the reverberant field can be modelled as undesirable additive noise. Even if in practice deviations from the above ideal model exist, however the results of localisation are satisfactory for the needs of the majority of applications [9]-[11].

Let's consider the usage of two omni-directional microphones in the 2D plane. In favour of simplicity, the coordinates system is fitted so as the first microphone is placed in (0,0) and the second in (1,0). The final setup is presented in Figure 2.

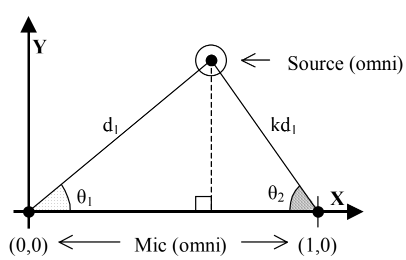


Figure 2 Two omni-directional microphones in 2D plane setup

If the conditions for (3) and (4) apply then:

$$\left.\begin{array}{c} \dfrac{kP_s}{4\pi d^2} = \dfrac{p_s^2}{\rho_o c} \Rightarrow P_s = \dfrac{4\pi}{k\rho_o c} p_s^2 d^2 \\ P_{S1} = P_{S2} \end{array}\right\} \Rightarrow p_1^2 d_1^2 = p_2^2 d_2^2 \Rightarrow$$

$$\Rightarrow d_2 = \frac{p_1}{p_2} d_1 \Rightarrow \begin{cases} d_2 = k \cdot d_1 \\ k = \dfrac{p_1}{p_2} \end{cases} \tag{5}$$

Using the right triangles and the cosine law in Figure 2, the distance d_1 can be expressed as $f_+\left(k,\theta_1\right)$.

$$\left.\begin{array}{r} d_1 \cos\theta_1 + kd_1 \cos\theta_2 = 1 \\ d_1^2 = 1 + \left(kd_1\right)^2 - 2kd_1 \cos\theta_2 \end{array}\right\} \Rightarrow$$

$$\Rightarrow \left(1-k^2\right) d_1^2 - 2\cos\theta_1 d_1 + 1 = 0 \Rightarrow$$

$$d_1 = \begin{cases} \dfrac{\cos\theta_1 \pm \sqrt{\cos^2\theta_1 - \left(1-k^2\right)}}{\left(1-k^2\right)}, & k<1 \\ 1/\left(2\cos\theta_1\right), & k=1 \\ \dfrac{\cos\theta_1 - \sqrt{\cos^2\theta_1 - \left(1-k^2\right)}}{\left(1-k^2\right)}, & k>1 \end{cases} \tag{6}$$

with $\sin\theta_1 < k$. The above function gives two valid solutions for the case of $k<1$.

Finally, the above function is transferred into 3D space according to Figure 3.

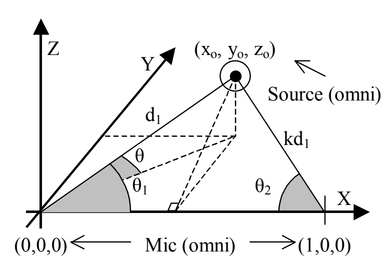


Figure 3 Two omni-directional microphones in 3D space setup

$$\left.\begin{array}{l} x_o = d_1 \cos\phi\cos\theta \\ \cos\theta_1 = \dfrac{x_o}{d_1} \end{array}\right\} \Rightarrow \cos\theta_1 = \cos\phi\cos\theta \qquad (7)$$

From (6) and (7), the $f_+(k,\theta,\phi)$ is expressed as follows

$$d_1 = \begin{cases} \dfrac{\cos\phi\cos\theta \pm \sqrt{\cos^2\phi\cdot\cos^2\theta-\left(1-k^2\right)}}{\left(1-k^2\right)}, & k<1 \\ 1/\left(2\cos\phi\cdot\cos\theta\right), & k=1 \\ \dfrac{\cos\phi\cos\theta - \sqrt{\cos^2\phi\cdot\cos^2\theta-\left(1-k^2\right)}}{\left(1-k^2\right)}, & k>1 \end{cases} \qquad (8)$$

with $\cos\phi\cos\theta < k$.

If the real distance of the microphones is L then the equation that provides the real distance of the source from the beginning of the axes d_{acc} results from (8) into the following:

$$d_{acc} = L\cdot d_1 = L\cdot f_+\left(k,\theta,\phi\right) = F_+\left(k,\theta,\phi\right) =$$

$$= \begin{cases} L\cdot\dfrac{\cos\phi\cos\theta \pm \sqrt{\cos^2\phi\cdot\cos^2\theta-\left(1-k^2\right)}}{\left(1-k^2\right)}, & k<1 \\ L/\left(2\cos\phi\cdot\cos\theta\right), & k=1 \\ L\cdot\dfrac{\cos\phi\cos\theta - \sqrt{\cos^2\phi\cdot\cos^2\theta-\left(1-k^2\right)}}{\left(1-k^2\right)}, & k>1 \end{cases} \qquad (9)$$

In the case that the second microphone in Figure 3 is situated in the position (-1,0), the eq. (9) is transformed into the following:

$$d_{acc} = L\cdot d_1 = L\cdot f_-\left(k,\theta,\phi\right) = F_-\left(k,\theta,\phi\right) =$$

$$= \begin{cases} L\cdot\dfrac{-\cos\phi\cos\theta \pm \sqrt{\cos^2\phi\cdot\cos^2\theta-\left(1-k^2\right)}}{\left(1-k^2\right)}, & k<1 \\ -L/\left(2\cos\phi\cdot\cos\theta\right), & k=1 \\ L\cdot\dfrac{-\cos\phi\cos\theta - \sqrt{\cos^2\phi\cdot\cos^2\theta-\left(1-k^2\right)}}{\left(1-k^2\right)}, & k>1 \end{cases} \qquad (10)$$

4. SOURCE LOCALIZATION

The eq. (9) and (10) expresses a surface of the possible sound source positions, given the sound pressure values measured by two omni-directional microphones in specific positions. Based on this and the B-format model, we propose several microphone setups in order to achieve source localization.

4.1. 1 Soundfield + 1 Omni Setup

The configuration where the source localization is accomplished using a soundfield and omni-directional microphone is derived from the one presented in Figure 3, using a soundfield type microphone in the (0,0,0) position. Eq. (2) specifies a B-format direction via a pair ϕ and θ angles (ϕ, θ and $\phi+\pi$, $-\theta$) and the k coefficient of eq. (5) is calculated from the omni acoustic pressure component of the soundfield microphone p_W (eq. (1)) and the reading of the omni-directional microphone.

$$k = \frac{p_W}{p_{omni}} \qquad (11)$$

Subsequently, the distance of the sound source from the axis centre is exported by directly applying eq. (9) for each of the ϕ, θ pairs and by selecting the positive d_{acc}.

The fact that for $k<1$, eq. (9) gives two possible source locations, push us to examine further configurations or to look at the constraints posed by the practical conditions. For example, the Z-coordinate of sound sources is usually known in most real world recording applications. In addition, in the case of slowly moving sources, where windowed signal processing localization is used, position history may facilitate the process.

4.2. 2 Soundfield Setup

The two soundfield microphone technique is identical with the one presented in 4.1 based on the omni acoustic pressure components of both soundfield microphones p_{W1}, p_{W2}. Moreover, source location may be found by converging the estimated directions. An indicative setup of this setup is presented in Figure 4.

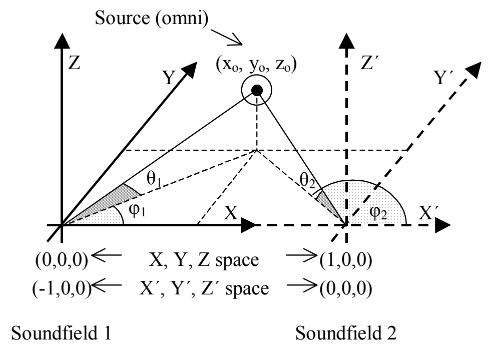


Figure 4 Source localization using two soundfield microphones

In the optimal case of non-reverberant field, the lack of noise and ideal microphone frequency responses, the calculated sound source positions from eq. (9), (10) must coincide. However, the presence of these deteriorative factors may lead to crossing direction lines exported the angular information and to different source positions calculations from each soundfield microphone.

Thus, as the angular information is provided for both microphone positions, the above redundant configuration can be exploited in order to increase the accuracy of the localization method. The direction calculation is relied on the microphone closer to the sound source as it is less affected by the fault factors. According to the ISL, the higher omni acoustic pressure component indicates the nearest sensor, which by either eq. (9) or (10) specifies the exact source location as the case $k<1$ no longer apply.

$$d_{acc} = \begin{cases} F_-\left(k^{-1},\theta_2,\phi_2\right) & k<1 \\ F_+\left(k,\theta_1,\phi_1\right) & k>1 \end{cases} \tag{12}$$

4.3. 1 Soundfield + 2 Omni Setup and beyond

The same principal governs the case of two omni-directional and one soundfield microphone, depicted in Figure 5. The direction is provided from the soundfield microphone. In order to find the sound source position the closest omni is selected, according to ISL. This actually creates a partitioning of the 3D space in two volumes. The calculation of the source position for each volume is made with the use of eq. (9) and (10) from the closer omni increasing the localization accuracy. The corresponding equations are shown below.

$$d_{acc} = \begin{cases} F_-\left(k_2,\theta,\phi\right) & k_2<k_1 \\ F_+\left(k_1,\theta,\phi\right) & k_1<k_2 \end{cases} \tag{13}$$

where

$$k_1 = \frac{p_W}{p_{omni1}} \text{ and } k_2 = \frac{p_W}{p_{omni2}} \tag{14}$$

In case that d_{acc} is calculated with i.e. $k_1<1$, the sound source position is also calculated for the other microphone, which gives only one solution as $k_2>1$. The correct location is selected to be the closest to the one calculated by the far microphone.

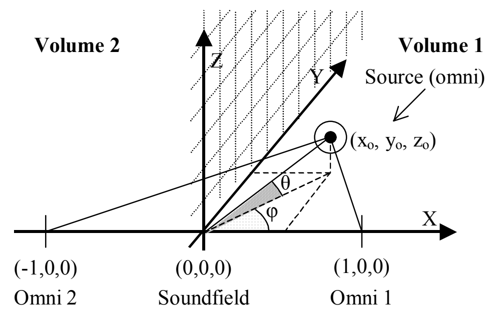


Figure 5 Source localization using two omni-directional and one soundfield microphone

The above configurations can be extended to multiple microphone combinations setups of the soundfield and omni-directional type. Similar eq. (9) and (10) can be derived for various microphone positions in the 3D space as well as allowing the creation of more complex volume that increase the redundancy as well as the accuracy of the method. The closest point of approach rule may be used to initiate the process, as well as to constrain the localization area taking advantage of the signal components of the highest energy [2], [3].

5. B-FORMAT ENHANCEMENT

After the localization of the sound source is applied, the B-format enhancement procedure is the next step. As already stated, the fault factors mainly involve reverberation and noise. Let's initially assume that the location of the sound source is known and that the acoustic pressure $p_1(t)$ as well as the location of an omni-directional microphone is available. This setup is depicted in Figure 6.

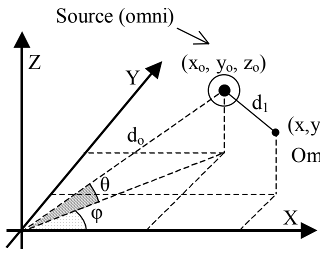


Figure 6 Omni to B-Format transferring

Given the above and the sound propagation velocity using the air as medium (c_{sound}≈*344 m/sec*), the delay of signal arrival on positions (0,0,0) and (x,y,z) as well as their difference is given by the following equations

$$t_{(0,0,0)} = \frac{d_o}{c_{sound}} \text{ and } t_{(x,y,z)} = \frac{d_1}{c_{sound}}$$

$$\Delta T = t_{(x,y,z)} - t_{(0,0,0)} = \frac{d_1 - d_o}{c_{sound}} \quad (15)$$

The resulting sound pressure signal due to the ISL difference in position (0,0,0) is provided by eq. (5) and eq. (15).

$$p_W(t) = \frac{d_1}{d_o} p_1(t - \Delta T) = \frac{d_1}{d_o} p_1\left(t - \frac{d_1 - d_o}{c_{sound}}\right) \quad (16)$$

The application of eq. (16) on multiple omni-directional readings (either normal or soundfield microphones) followed by their weighted addition is given by the equation below.

$$\left.\begin{aligned} p_W(t) &= \frac{1}{N}\sum_{i=1}^{N} w_i \cdot \frac{d_i}{d_o} \cdot p_i\left(t - \frac{d_i - d_o}{c_{sound}}\right) \\ w_i &= \frac{\min\{d_i \mid i = 1,2,...,N\}}{d_i},\ i = 1,2,...,N \end{aligned}\right\} \Rightarrow$$

$$\begin{aligned} p_W(t) &= \frac{w_o}{N} \cdot \sum_{i=1}^{N} p_i\left(t - \frac{d_i - d_o}{c_{sound}}\right) \\ w_o &= \frac{\min\{d_i \mid i = 1,2,...,N\}}{d_o} \end{aligned} \quad (17)$$

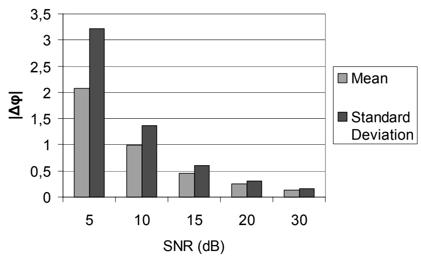
In the addition above, the weight factor w_i favours the reading that are closer to the sound source, as they include a "clearer" representation of the direct field signal. This estimation of the acoustic pressure signal in position (0,0,0) is a weighted Time Delay Compensation (wTDC) technique [2]-[10]. The resulting B-Format signals in position (0,0,0) are finally calculated from eq. (1) according to the following:

$$\begin{aligned} p_S(t) &= p_W(t)/0.707 \\ p_X(t) &= p_S(t)\cos\phi\cos\theta \\ p_Y(t) &= p_S(t)\sin\phi\cos\theta \\ p_Z(t) &= p_S(t)\sin\theta \end{aligned} \quad (18)$$

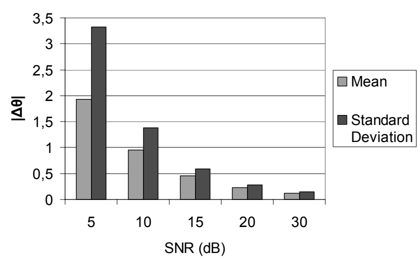
In the position (0,0,0), the resulting B-Format signals are temporally compensated with regard to the direct field of the sound source. As a consequence, the final signal contains an amplified version of the direct field versus the reverberant field signal, leading to improved SNR ratios and de-reverberation effects. Spatial audio information is retained via the enhanced B-Format components, fact that introduces novelty over traditional spatial audio processing techniques. Furthermore, virtual positioning of one or more B-Format microphones in the recording area is possible, issue that can extend the functionality of the current virtual reality approaches.

6. SIMULATION RESULTS

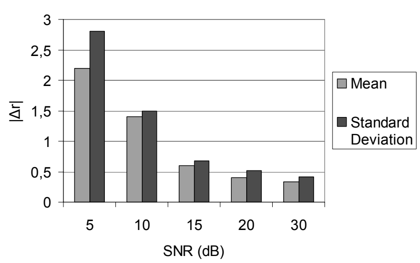
In order to evaluate the basis of the above theoretic approaches, a simulation process was setup and carried out. Specifically, random source positions were simulated in a computer environment, where the simplest model of inverse square law was used for sound propagation. Gaussian white noise was added to the microphone components to simulate ambient noise and reverberant field factors. This approach also gave us the ability to examine the efficiency of the method in various SNR conditions. An additional model for indoor conditions was used on the basis of mirroring sources in simple geometry rectangular rooms. Figure 7, describes the error estimation results with regard to the worst SNR among all the involved components. Figures 7a) and 7b) depict the direction estimation error (mean, standard deviation) that is provided by a single soundfield microphone. Figure 7c) describes the corresponding distance error for the "two B-Format" source localization case. It is important to mention that these simulations were conducted to validate our theoretical approach rather than fully evaluate the proposed methodology. Besides, experimental approaches in real recording conditions are currently under scheduling.



a)



b)



c)

Figure 7 Source localization estimation error results: a) angular deviation $|\Delta\phi|$, b) angular deviation $|\Delta\theta|$ and c) radial deviation $|\Delta r|$

7. CONCLUSION

The current work provides a theoretical approach for sound source localization and spatial audio enhancement using soundfield and omni-directional microphone sets. The single point omni-directional source model was considered in combination with additive broadband noise and mirroring techniques that simulated reverberant field and ambient noise. Based on the first preliminary results, we believe that the proposed methodology is very promising with the additional advantage of reduced complexity mathematics and fast implementation. The extension, to the cases of multiple and moving sources is under investigation. Future work includes, the conduction of various tests, both simulative and practical using actual laboratory equipment, for full evaluation of the algorithms under discussion. The implementation of interactive software environments for experimental and educational purposes is an additional target.

8. REFERANCES

[1] Streicher R., and Everest A.F., "The new stereo soundbook", second edition, ISBN 0 9665162 0 6, Audio Engineering Associates, Pasadena, California, 1998.

[2] Dimoulas C. A., "Audio-visual processing and content management techniques, for the study of (human) bioacoustics' phenomena", PhD Dissertation (in Greek), Dept. of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Nov. 2006.

[3] Chen J.C., Yao K., and Hudson R.E., "Source localization and beamforming", IEEE Signal Processing Magazine, vol. 19, no. 2, pp. 30-39, March 2002.

[4] Malham, D. G., "TUTORIAL ARTICLE: Approaches to spatialisation", Organised Sound, vol. 3, no. 2, pp. 167–177, 1998.

[5] Ginn B., Christensen J.J., Hald J., Mørkholt J., Schuhmacher A. and Blaabjerg C., "A review of array techniques for noise source location", 10th International Congress on Sound and Vibration, pp. 1889-1896, 2003.

[6] Chen J. C., Yao K., Hudson R. E., "Acoustic Source Localization and Beamforming: Theory and Practice", EURASIP Journal on Applied Signal Processing, vol. 4, pp. 359–370, 2003.

[7] Christensen J.J., and Hald J., "Beamforming", Brüel & Kjær Technical Review, No. 1, 2004, (Ed.) Zaveri H.K., available online: www.bk.dk, pp. 1-48, 2004.

[8] Weng J., and Guentchev K.Y., "Three-dimensional sound localization from a compact non-coplanar array of microphones using tree-based learning", J. Acoust. Soc. Am., vol. 110, no. 1, pp. 310-323, July 2001.

[9] Li D. and Hu Y-H, "Energy-based collaborative source localization using acoustic microsensor array", EURASIP Journal on Applied Signal Processing, vol. 2003, no. 4, pp. 321–337, 2003.

[10] Sheng X. and Hu Y-H, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks", IEEE Transactions on Signal Processing, vol. 53, no. 1, pp.44-53, Jan. 2005.

[11] G. Kalliris, K. Avdelidis, N. Vagiokas, C. Gousios, C. Dimoulas, G. Papanikolaou, K. Pastiadis, C. Sevastiadis, "Noise and Vibration in the working environment: Evaluation – estimation, protection", 2nd Pan-Hellenic Conference of Acoustics ELINA: "AKOUSTIKI 2004", Thessaloniki, 27/09-28/09, 2004.

[12] Gerzon M., "Practical Periphony: The Reproduction of Full-Sphere Sound", 65th Convention of Audio Eng. Soc., preprint number 1571, London, Feb. 1980.

[13] Gerzon M., "Ambisonics in Multichannel Broadcasting and Video", J. Audio Eng. Soc., vol. 33, no.11, pp. 859-871, Nov. 1985.

[14] Jagger D. S., "New Developments and Improvements in Soundfield Microphone Technology", 75th AES Convention, Preprint Number 2064, Feb. 1984.

[15] SoundField, "SoundField Technology", available online: www.soundfield.com, 2006.

[16] Rafaely Boaz, "Design of a Second-Order Soundfield Microphone", 118th Convention of Audio Eng. Soc., Preprint Number: 6405, May 2005.

[17] Benjamin Eric and Chen Thomas, "The Native B-Format Microphone", 119th Convention of Audio Eng. Soc., Preprint Number: 6621, Oct. 2005.

[18] Benjamin Eric and Chen Thomas, "The Native B-Format Microphone: Part II", 120th Convention of Audio Eng. Soc., Preprint Number: 6640, May 2006.