Audio Engineering Society
Convention Paper 5501
Presented at the 112th Convention
2002 May 10–13 Munich, Germany

Applications of a 3-D Microphone Array

Juha Merimaa

Helsinki University of Technology, Lab. of Acoustics and Audio Signal Processing, P.O.Box 3000, FIN-02015, HUT, Finland

Correspondence should be addressed to Juha Merimaa (juha.merimaa@hut.fi)

ABSTRACT

Soundfield inside an enclosed space depends in a complex way upon interactions between emitted sound waves and different reflecting, diffracting, and scattering surfaces. 3-D microphone arrays provide tools for investigating and recording these interactions. This paper discusses several existing array techniques introducing a variety of application targets for the HUT microphone probe. Applications include directional measurement, analysis, and visualization of room responses, estimation of room parameters and analysis of source and surface positions. In a dynamic case the probe can be utilized in source tracking and beam steering, as well as in tracking its own position. Furthermore, the probe can be used to simulate some microphone arrays commonly used in surround sound recording. In each application case both general theory and its relation to the HUT probe is discussed.

INTRODUCTION

In all practical situation soundfield inside an enclosed space consists of acoustical waves propagating in several different directions. Sound waves emitted by a source are reflected, diffracted, and scattered by different obstacles including the walls of the enclosure. This results in a complex field, the properties of which cannot be comprehensively captured in one-dimensional signals or parameters.

Different practitioners have different viewpoints into spatial sound. Researchers and acoustical engineers often want to measure a response to a given stimulus in a room, in order to gain information about the reasons why the room sounds like it does. The motivation for this may be an attempt to change the acoustics, or a pure scientific interest in the underlying phenomena, including perception of spatial sound. A recording engineer, on the other hand, may want to capture a performance in a room so that an illusion of the room can be later reproduced somewhere else. Another related problem is selective recording of certain sound sources. Third distinct area of interest is acoustical orientation, i.e., localization of sound sources, reflecting obstacles and surfaces, or the receiver itself based on received sound signals. Of course, there

Fig. 1: The HUT 3-D microphone probe.

are also intersections between these viewpoints.

This paper describes applications of a 3-D microphone array in all the previously mentioned tasks. The HUT microphone probe consists of 12 miniature electret microphone capsules arranged as two concentric pairs in each of x-, y-, and z-coordinate axes. The inner pairs are set with a spacing of 10 mm and the outer pairs with a spacing of 100 mm between the capsules. The probe was originally designed for measurement purposes but has proven useful in other applications as well. A picture of the probe is shown in Fig. 1. Related hardware and software are described in [1] and [2].

The paper is divided into three parts. The first part discusses directional measurement and analysis of room responses. Second part presents some possibilities of using the probe in sound recording. Finally, the third part introduces source localization techniques applicable both in measurement purposes and in acoustical orientation.

ROOM RESPONSE MEASUREMENTS

Measurement of room responses and analysis of related attributes is a common task in audio and acoustics. Most often an omnidirectional response to a preferably omnidirectional stimulus is acquired. This is sufficient for calculation of several room-acoustical parameters. However, single omnidirectional responses and standard parameters provide only limited information about the actual acoustics of the room and its perceptual properties.

Microphone array techniques utlilized in room response measurements can be roughly divided into two categories. In the first category large arrays spanning a significant distance, area, or volume in the room are utilized. This gives a representation of the evolving soundfield as a function of spatial position. Application of a long line array in this purpose has been described in [3]. In the second category small arrays such as the HUT microphone probe are used to give a listener centered view of the directional soundfield. The following discussion concentrates on methods in the latter category.

Directional sound pressure components

Ideal omnidirectional microphones are sensitive to sound pressure, which as a scalar quantity does not include any directional information. Systems with varying directional sensitivity can be formed by appropriately combining the signals of two or more closely spaced omnidirectional microphones. An attractive feature of using an array of omnidirectional microphones in place of directional microphones is the possibility to vary and steer the directivity patterns later in the postprocessing phase.

First-order differential directivity patterns can be easily created using the signals of a closely spaced pair of microphones. This kind of beamforming methods are analogous to the construction of microphones with built-in directionality [4, 5]. Basically, all that is needed is some equalization and delay, and a weighted summation of the resulting signals. An ideal dipole has a directivity pattern of the form

$$E(\theta) = \cos(\theta) \tag{1}$$

where θ is the angle of arrival of a plane wave related to the line connecting the pair of microphones. This kind of a pattern is constructed from the equalized difference of the signals of the microphones. A cardioid pattern, on the other hand, is based on a difference where one of the signals is delayed by the time corresponding to the propagation of sound over the distance between the microphones. Another way of forming a cardioid is to sum two signals with omnidirectional and dipole directivity patterns. More generally, any first-order differential pattern can be formed as a weighted sum of an omnidirectional pattern and a dipole pattern resulting in the directivity

$$E(\theta) = \alpha + (1-\alpha)\cos(\theta) \tag{2}$$

Differential directivity patterns with some common α are shown in Fig. 2.

A simple steering method for first-order differential patterns has been patented by Elko [6]. His method is based on steering a dipole pattern, after which any first-order pattern can be created as a combination with an omnidirectional signal as described. In the case of the HUT microphone probe dipoles in x-, y-, and z-directions can be readily formed from corresponding microphone pairs. A signal of a steered dipole can then be written in the form

$$E(\theta,\phi) = \cos(\theta)\sin(\phi)E_x + \sin(\theta)\sin(\phi)E_y + \cos(\phi)E_z \tag{3}$$

where θ is the azimuth angle and ϕ is the elevation angle of the resulting dipole pattern, and E_x, E_y, and E_z are signals with dipole patterns facing at the directions of x-, y-, and z-axis, respectively.

The resolution provided by first-order differential directivity patterns may not always be optimal for measurement purposes. Okubo et al. [7] have proposed a product of dipole and cardioid signals in order to construct a unidirectional directivity pattern with more selectivity than the first-order cardioid. The resulting pattern is shown in Fig. 3. For broadband signals this method is, of course, applicable in the frequency domain only, since a time domain multiplication would result in harmonics at sum and difference frequencies of original signal components. Okubo et al were using an array with single

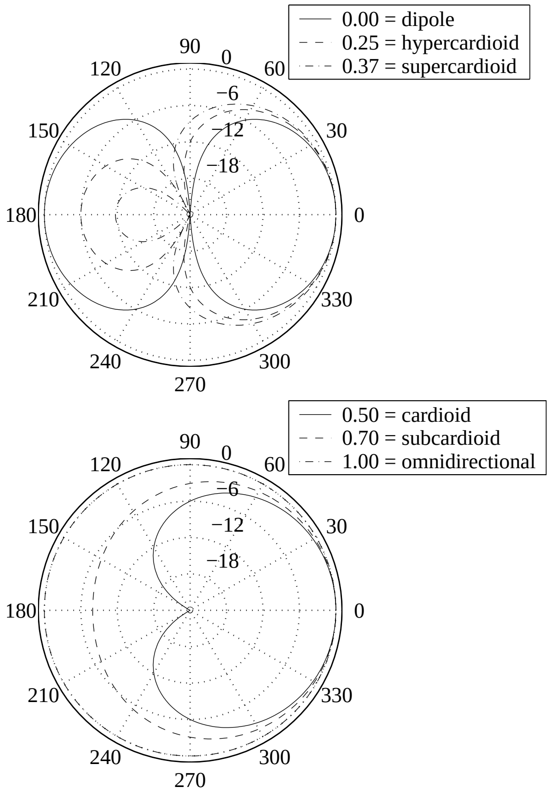

Fig. 2: Polar plots of the logarithmic magnitude responses of first-order differential directivity patterns with some typical α in Eq (2).

concentric pairs of omnidirectional microphones on horizontal (x- and y-) coordinate axis, complemented with a microphone in the middle of the array.

The difference of the signals of a pair of microphones approximates the gradient of the soundfield at low frequencies. The upper frequency limit is constrained by the distance between the microphones. At frequencies where the wavelength of sound is comparable to the microphone spacing, the approximation of gradient is no longer valid and the directivity pattern is distorted. The lower limit, on the other hand, is defined by the phase errors and signal-to-noise ratio of the measurement system. Furthermore, the so called proximity effect causes the magnitude response of a differential system to depend on the distance of a sound source located close to the microphones [4]. In the HUT microphone probe these frequency range problems have been alleviated by positioning two sets of microphones at different distances from the center, allowing thus combined operation on two frequency scales.

Apart from differential techniques, a common method for beamforming is summation of several microphone signals delayed so that they are in phase for sound waves arriving from the desired direction. A problem related to this kind of methods is changing directivity as a function of frequency. The narrowing of the beam with increasing frequency can be prevented by frequency dependent shading of the outer elements in an array [8, 9]. Arrays with non-uniform spacing between microphones have also been used [10]. However, delay-and-sum beamforming at low frequencies requires the size of an array to be comparable to the wavelength of sound, while the spacing between the microphones used at high frequencies needs be comparable to the wavelength at those frequencies in order to avoid spatial aliasing. For a small array, such as the HUT microphone probe, differential methods provide thus better results for wideband beamforming. If the HUT probe were to be used as a delay-and-sum beamformer, its geometry could be utilized either as line or plane arrays, or as a sparsely sampled spherical shell or volume array [11, 12].

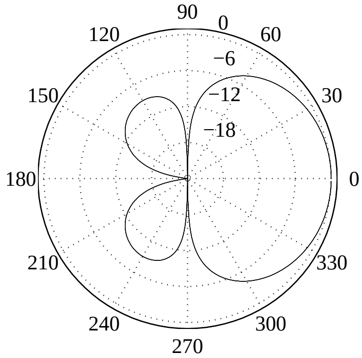

Fig. 3: A polar plot of the logarithmic magnitude response of the directivity pattern proposed by Okubo et al [7].

Separation of measured impulse responses into directional components allows isolated investigation of reflections and diffraction caused by objects in a chosen direction as seen from the listening position. This may sometimes be useful, but in general separate treatment of different directions results in a vast amount of data that is hard to interpret. That is why patterns with higher directivity have not been found very practical in measurement applications. On the contrary, methods for more compact representation of data have been explored. However, Broadhurst [13] has described the design of a sparse volumetric array having a beamwidth of 32° for acoustical measurements. Unfortunately he does not discuss the actual measurements and their interpretation in his paper. Some higher order beamforming methods are also applicable with the HUT microphone and will be discussed in this paper in connection with acoustic communications applications and source localization.

Sound intensity

Sound intensity describes the propagation of energy in a soundfield. Being a vector quantity, it naturally includes a directional aspect. Measured sound intensity data complements nicely the pressure-related impulse responses in characterization of a directional soundfield. Especially the differences between sound pressure and intensity fields may be of interest. For example, in a completely diffuse soundfield, intensity should be zero irrespective of the sound pressure level.

Intensity is defined as the product of sound pressure and particle velocity. In a standard intensity measurement technique the pressure signals $p_1(t)$ and $p_2(t)$ of a pair of closely placed omnidirectional microphones are used to approximate the pressure and velocity components in a point halfway between the microphones. With these signals the instantaneous intensity in the direction of a line connecting the two microphones can be approximated with

$$I(t) \approx \frac{1}{2\rho_0 d} \left[p_1(t) + p_2(t)\right] \int_{-\infty}^{t} \left[p_1(\tau) - p_2(\tau)\right] \mathrm{d}\tau \qquad (4)$$

where ρ_0 is the mean density of air and d is the distance between the microphones [14]. A 3-D intensity vector can be determined using three concentric pairs of microphones in the directions of x-, y-, and z-coordinate axis, such as in the HUT microphone probe.

Sound intensity can be divided into active and reactive components. The active component describes the net transport of sound energy and can be calculated as the average of instantaneous intensity. Frequency distribution of active intensity can also be determined with an FFT-based method from the imaginary part of the cross-spectral density G_{p1p2} of the microphone signals [14, 15]

$$I(\omega) \approx -\frac{j}{\rho_0 \omega d} \mathrm{Im}\left\{G_{p1p2}(\omega)\right\} \qquad (5)$$

where ω is the angular frequency, j is the imaginary unit, and G_{p1p2} is given by

$$G_{p1p2}(\omega) = 2P_1^*(\omega)P_2(\omega) \qquad (6)$$

where * denotes the complex conjugate and $P_1(\omega)$ and $P_2(\omega)$ are the Fourier transforms of the microphone signals $p_1(t)$ and $p_2(t)$, respectively.

The finite difference approximations of sound pressure and particle velocity in the point halfway between the microphones cause frequency dependent systematic error to the approximation. The actual error depends on the acoustical field being measured [14]. Chung [16] has proposed a practical upper limit of

$$kd = 1 \Leftrightarrow f_L = \frac{c}{2\pi d} \qquad (7)$$

where k is the wave number, f_L is the limiting frequency and c is the speed of sound. The practical lower frequency limit, on the other hand, depends on the phase errors of the measurement system. In the HUT microphone probe concentric microphone pairs with two different spacings can once again be used to extend the frequency range. With a larger array having microphones outside the coordinate axis, a method proposed by Kuttruff [17] could also be utilized in order to further reduce the approximation errors.

Visualization of directional room responses

As mentioned earlier, directional processing of room responses results in a large amount of data that is difficult to interpret. Examining several figures of directional sound pressure and intensity components is not very informative. A simple visualization method for directional room responses was proposed in [18]. The method is a combination of two overlaid plots. Active intensity is represented as vectors (quiver plot) laid on top of a sound pressure related spectrogram. Both quantities are analyzed with same time-frequency resolution. One combination plot illustrates horizontal information and another one elevation information in the median plane. In the original paper both uniform and auditory frequency resolutions were applied.

An example of the visualization method is shown in Fig. 4. The plot clearly shows the direct sound followed by discrete reflections from the floor, the ceiling, and the front wall of the room. After these the density of reflections gets higher and the chosen time-frequency resolution cannot distinguish single wideband reflections. In a larger room the first reflections are sparser in time and interesting phenomena can still be seen much later in the responses.

SOUND RECORDING

In conventional studio recording applications microphone arrays may not provide much added value. Typically, a recording engineer can choose the microphones he likes and set them the way he likes. Arrays could, of course, be used to simulate directional microphones, but there is no point in using a number of expensive studio quality microphones to replace single devices that are readily available. However, in natural stereo and surround sound recording, as well as in teleconferencing applications, microphone arrays have proven useful. The following discussion concentrates on surround sound and teleconferencing arrays. For a review of the fairly well established stereo recording techniques see [19].

Surround sound

Several current and past surround sound reproduction systems are discussed in a review by Steinke [20]. Corresponding multi-channel recording techniques can be roughly divided into two categories similar to the measurement arrays. In the first category several microphones are placed in different locations in a room or a concert hall. Typically some of them are located close to sound sources to get the direct sound, while others are used to capture the reverberation and spatiousness from a distance. Often more ambience is added with artificial reverberators. The resulting signals are then processed and mixed down into a set of channels for reproduction with a specific loudspeaker setup. This kind of a recording techniques do not necessarily require any microphone arrays. If the acoustics of a hall is tried to reproduce as naturally as possible, a small number of microphones with a definite placement related to each other is, however, often used as a main array. Principles of this kind of recording techniques and some common geometries for main arrays for 5.1 reproduction, are discussed in [21].

Recording arrays in the second category try to capture the authentic soundfield in a single position in a room. These arrays can be further divided into coincident and spaced setups, where coincident arrays aim for recording the soundfield in a single point of space while in spaced arrays microphones are set with small distances between them. Lipschitz [19] still categorizes spaced arrays into quasi-coincident and spaced setups based on the distances between microphones. In this text, however, both quasi-coincident and spaced arrays are referred to as spaced.

Coincident recording arrays

The earliest surround sound reproduction systems were so called quadraphonic setups where four loudspeakers were placed in corners of a rectangle with two speakers in front of and two speakers behind the listener. (The term quadraphonics also refers to matrixing of four discrete mono reproduction channels into a stereo signal [20, 22, 23].) Design of a single quadraphonic microphone consisting of four coincident cardioids aligned in the directions of the loudspeakers is described in [24]. A quadraphonic loudspeaker setup cannot, however, create an illusion of sound sources on the sides of the listener because of the known incapability to create stable phantom sources between speakers positioned in front

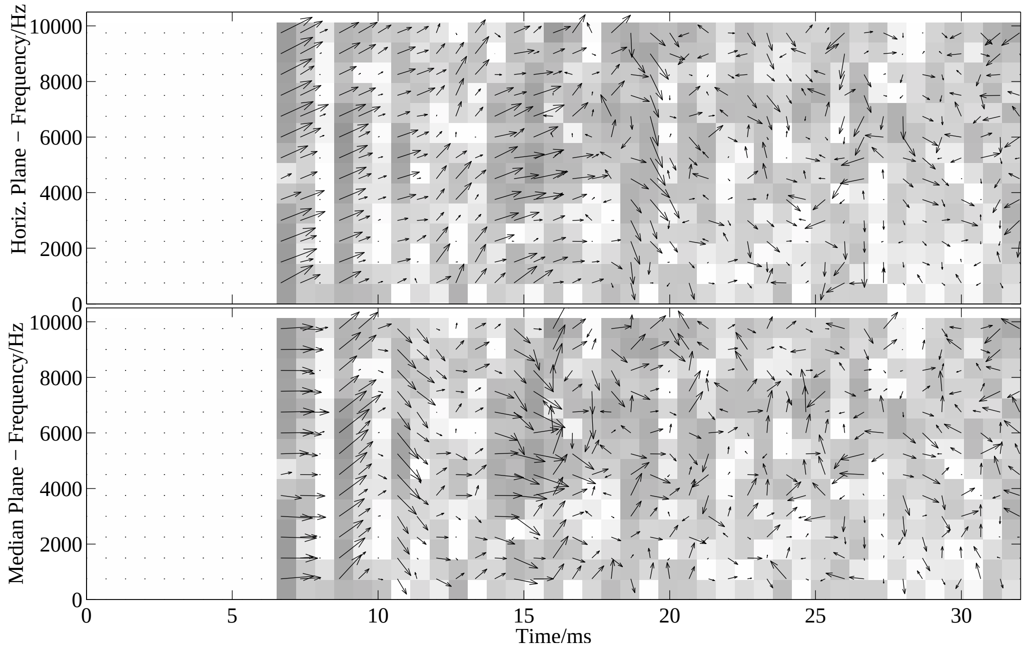

Fig. 4: Visualization of the directional room responses measured from the listening room of HUT Acoustics laboratory, analyzed with a uniform time-frequency resolution. The direct sound arrives in about 7 ms from a front left direction in the horizontol plane. The next events are reflections from the floor and from the ceiling and can be seen to propagate to the same direction in the horizontal plane. In about 12 ms these there can be seen some more diffuse reflections from the diffracting constructions on the left side wall followed by a clear discrete reflection from the front wall of the room in about 15 ms. After 20 ms no clear wideband reflections can be seen anymore.

of and behind the listener [25]. Also the insufficient directional resolution provided by cardioid microphones causes a recorded sound source to spread into several or all loudspeakers in reproduction.

Ambisonics [26, 27, 28] is a generalized 3-D system approach to surround sound. The method is based on reconstruction of soundfield with spherical harmonic functions. In theory, functions up to any order could be used [28]. However, only first- and second order systems have been found to be practical. In the so called B-format of first-order ambisonics the soundfield is divided into an omnidirectional component W (zeroth-order spherical harmonic) and three dipole components X, Y, and Z in the directions of corresponding cartesian coordinate axis (three linearly independent first-order spherical harmonics). The signal $P(\theta, \phi)$ to be reproduced by a loudspeaker in any direction defined by the azimuth angle θ and the elevation angle ϕ, can then be written in the form

$$P(\theta, \phi) = W + 3[\cos(\theta)\sin(\phi)X + \sin(\theta)\sin(\phi)Y + \cos(\phi)Z] \quad (8)$$

where it is assumed that the level of W equals the level of the dipole signals in the direction of their maximum gain [28].

The gain coefficient 3 in Eq (8) compensates for the lower pickup energy of dipoles in a diffuse field. The resulting sum corresponds to a signal recorded from a soundfield with a hypercardioid microphone pointing in the direction of the loudspeaker used for reproduction. This gives maximal directional discrimination that is available with first-order directivity patterns. However, the large rear-lobe of a hypercardioid causes a recorded signal to be reproduced in opposite phase from loudspeakers in opposite directions. Depending on loudspeaker configuration less resolution may sometimes give better sounding results, and can be achieved by reducing the gain factor [28]. In [26, 29, 30] a coefficient of 2 has been utilized resulting in a pattern close to supercardioid.

Typically first-order ambisonics is recorded with a soundfield microphone constructed of four coincident cardioid capsules mounted in the form of a tetrahedron [31, 32]. A cardioid pattern consists of zeroth- and first-order spherical harmonics, and the B-format signals can thus be constructed from the cardioids with simple linear algebra. The easiest way to use the HUT microphone probe for ambisonics recordings is, however, to directly extract the omnidirectional and dipole signals needed in the B-format.

A second-order ambisonics system can be formed by adding five microphone channels with linearly independent second-order spherical harmonic (clover-leaf) pickup patterns to a first-order system. The second-order patterns are illustrated in Fig. 5. All the signals needed for construction of these patterns, as well as the first-order patterns, can be obtained using an array of twelwe small cardioid capsules mounted to form a regular dodecahedron. In this case, the unavoidable non-coincidence of the microphones is taken advantage of by creating the second-order patterns by applying differential tech-

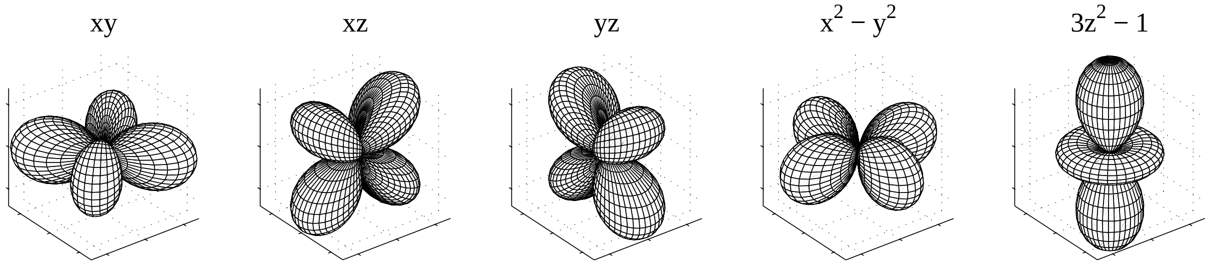

Fig. 5: Second order spherical harmonic functions.

niques to the cardioid signals [28]. The required second-order patterns can also be formed for a limited frequency range with the HUT microphone probe. However, steering them to correct directions related to each other is not possible and thus second-order abisonics cannot be recorded. Construction of second-order differential patterns with the probe is discussed in more detail in the section of acoustic communications applications in this paper.

The absolute minimum number of loudspeakers for ambisonic reproduction is defined by the number of encoded channels, i.e., for first-order ambisonics the minimum is 4 and for second-order 9. More than that are needed to fully reproduce all the encoded information about the spatial distribution of sound energy around a listener. However, in ambisonics the directional resolution is naturally limited by the order of the spherical harmonic reconstruction and thus adding more loudspeakers will not sharpen the images after a certain limit [28].

The focus of phantom images resulting from several coincident multi-channel recording and panning techniques in a 2-D reproduction system has been compared by Martin et al [29]. In their tests with an 8-speaker setup, second-order ambisonics was able to produce very focused images in all directions. First-order ambisonics was less focused and the effect of reproduction of signals with opposite phase on the sides of a listener was apparent. The most blurred images were produced by a panning method simulating recordings with 8 cardioid microphones. An interesting comparison of simulated directional localization cues produced by different recording and reproduction methods has also been presented by Pulkki [30].

Spaced recording arrays

Recordings with an array of coincident microphones capture only localization cues that are recreated in reproduction with some form of amplitude panning between loudspeakers. Theile [21] has argued that microphones with a wider spacing are able to produce more natural spatial impression due to their inherent inter-channel temporal differences between recorded channels. However, the perceived images of discrete sound sources will usually be more ambiguous [19, 30]. This scheme may also lead to severe colorization due to comb filtering, if signals are mixed down with a simple summation for reproduction with a system with less channels.

Implementation of several spaced surround sound recording arrays has been described in the literature. Johnston and Lam [33] have reported experiments with a 3-D array of seven directional microphones. In their array, five hypercardioid microphones were placed on a circle in the horizontal plane and two microphones with more directivity were facing up and down. Special arrays incorporating a diffracting sphere and a dummy head have been described in [34] and [35], respectively. This kind of constructions are not possible with the HUT microphone probe, although coincident patterns similar to the ones used by Johnston and Lam could be formed.

Williams and Le Dû [36] have developed an interesting technique for construction of 2-D recording arrays from cardioid microphones. To take full advantage of their method the positions and orientations of the recording microphones need to be adjustable, but some principles can also be applied to an array of coincident directional microphones. The method is based on manipulation of so called coverage angles of pairs of directional microphones. A coverage angle determines the limits for directional angles of sound sources, as seen by the microphones, that can be naturally reproduced with a pair of loudspeakers. If a source gets outside the coverage angle it will be localized into one loudspeaker only. The angle depends on the distance and the angle between microphones forming a pair and it can be offset (rotated) by non-symmetrical movement of the microphones, or by introducing delay or gain to one of the microphone signals. The rotation can be seen as a form of time-intensity trading used to adjust the localization of a phantom source created from a real source in a defined direction. The trading, of course, results in conflicting ILD and ITD cues in reproduction, but within reasonable limits the coverage angles can be adjusted.

The idea of the method of Williams and Le Dû is to manipulate coverage angles of pairs of adjacent microphones in an array such that they finally cover 360° with no overlap. The recorded and manipulated signals of microphones are then played back with corresponding loudspeakers. The number of microphones and loudspeakers must be the same but their directions can differ. If the angle between a pair of loudspeakers is different from the coverage angle of the corresponding pair of microphones, the soundscape inside the coverage angle will be stretched or compressed in playback, accordingly.

Since in the case of the HUT microphone probe, adjustment of coverage angles by changing distances between pairs of directional microphones is not possible, the usability of the algorithm is limited. A new degree of freedom can, however, be introduced by allowing small changes in the directivity of a simulated microphone. This will once again lead to different distortion of the reproduced surrounding sound field. A more formal analysis of the different warpings created by this method, with or without a possibility to adjust the directivity

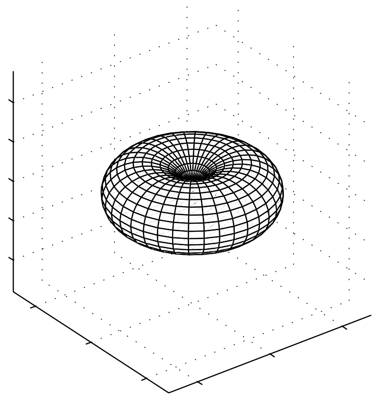

Fig. 6: Second order toroidal directivity pattern.

patterns, would require further testing.

Acoustic communications applications
For teleconferencing applications several special microphones and microphone arrays have been developed. Automatic microphone steering is one possibility for enhancing the pickup of a single speaker. Steering of first-order differential directivity patterns was already discussed in measurement applications and source localization methods will be introduced in the following section. A different approach is taken in systems with special directivity patterns. One common construction is a toroidal pattern (Fig. 6). The idea is that a toroidal microphone is placed on a table where it effectively picks up the speech of people sitting around the table while rejecting the sound from an overhead speaker.

A first-order toroid having a dipole pickup pattern in any vertical plane can be constructed by summing the signals of two orthogonal coincident horizontal dipoles in quadrature phase [4]. This can be easily realized with the HUT microphone probe by forming two dipole signals and using a Hilbert transformer [38] to rotate the phase of one of them by 90°. Second-order toroids, on the other hand, can be directly created by summing the signals of two orthogonal coincident second-order dipoles.

An implementation of a second-order toroidal microphone system with four closely spaced first-order dipoles is described by Sessler et al [39, 40]. The same method can be applied to the HUT microphone probe as follows. The placement of the omnidirectional microphones in the horizontal plane of the probe is illustrated in Fig. 7. Four closely spaced first-order dipole patterns can be formed as the difference signals of pairs $(1, 5)$, $(2, 6)$, $(3, 7)$, and $(4, 8)$. Second-order dipoles now result from the summation of dipoles $(1, 5)$ with $(3, 7)$, and $(2, 6)$ with $(4, 8)$. A toroidal pattern is, thus, created as the sum of all the four dipole signals. Unofortunately this kind of use of second-order techniques prevents the extension of frequency range by combined use of two pairs of microphes with different spacings. Moreover, problems with signal-to-noise ratio and proximity effect get more severe with higher directivity. Thus, either frequency range must be limited or directivity needs to be compromised when using the probe as a toroidal microphone.

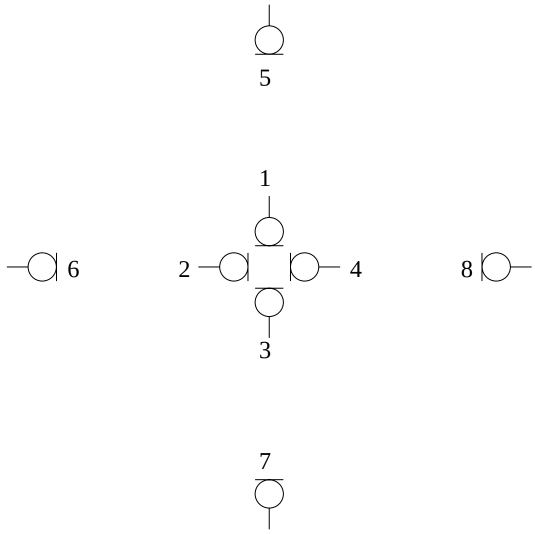

Fig. 7: Positions of the omnidirectional microphones in the horizontal plane of the HUT microphone probe.

Another useful feature in microphone systems designed for communications applications is higher order unidirectionality. More directivity is especially needed in noisy environments in order to pick up as little background noise as possible. Several different second-order unidirectional patterns can be constructed by differential techniques based on first-order differential signals. Design of a unidirectional pattern as a cardioid formed from two dipole signals is described in [41] and similar patterns can also be created with the HUT microphone probe. Elko et al have also proposed an image-derived method for constructing second-order unidirectional or toroidal directivity patterns with an omnidirectional microphone, a first-order dipole microphone and a reflecting table top or a wall [37].

An application with a varying directivity pattern has been introduced by Ishigaki et al. In [42] they describe an electronically zoomable microphone system to be used with a video camera based on mixing and filtering three cardioid signals. Another system with varying directivity as a function of frequency is described by Woszczyk [43]. He describes a system where directivity is increased for lower frequencies in order to pick up less reverberation when recording far from a sound source. The changing directivity is implemented with simple frequency dependent filtering of signals before using them to form a differential directivity pattern. For example, a dipole can be turned into an omnidirectional pattern at a chosen frequency by filtering another one of the original omnidirectional signals to zero.

SOURCE LOCALIZATION

Source localization has been studied extensively in connection with sonar applications, underwater acoustics, and automatic tracking of speakers in conference rooms. Antenna literature is also a good source for information on this topic. A comprehensive review of existing localization techniques is far beyond the scope of this paper. Instead, a brief overview of the methodology is given and the techniques based on time delay estimation are introduced in more detail. Finally, use of source localization in measurement and recording applications is discussed.

Localization problems arise in active and passive applications. In active applications the same system controls both a sound source and a receiver. This gives some distinct advantages. First, the signal emitted by the source can be chosen and is always precisely known. Furthermore, the whole system can often be designed to operate at certain limited bandwidth, which alleviates problems related to beamforming and design of the receiving array. On the other hand, in a general case of passive localization there is no a priori knowledge of the sound source. The goal of passive localization is usually to find the position or the direction of a sound source or several sources, while active systems aim at finding objects reflecting sound waves. The following discussion concentrates on methods useful in both active and passive applications operating at audio frequencies.

Array-based localization procedures can be loosely divided into three categories [44]. In the first category source is localized by maximizing the output of a steerable beamformer. One system for automatic search and recording of a single speaker in a larger room has been introduced by Flanagan et al [45]. Their system utilized a larger 2-D array of microphones with delay-and-sum beamforming, but a corresponding scheme can be implemented with the HUT microphone probe using steerable first-order differential directivity patterns. In the implementation of Flanagan et al simultaneous beams were utilized, one of which was used for recording and another one for scanning the room for a new speaker.

Statistical performance of beam steering based source localization has been discussed in [46] and is compared to that of the cross-correlation based time delay estimation localization scheme (see the next section in this paper) in [47]. Presence of several concurrent sound sources is one of the major problems in this kind of localization. Flanagan et al ended up implementing a speech detection algorithm to distinguish between speakers and acoustical background noise. Wax and Kailath [48] have extended the methodology to simultaneous localization of multiple sound sources, but their technique requires prior knowledge or the power spectra of all sound sources.

In the second category, methods of high-resolution spectral analysis are utilized in beamforming. Most methods in this group are originally designed for narrowband signals [44]. Extensions to wideband localization have been introduced based on analysis of signal components at several subbands both independently and interdependently. A coherent signal-subspace method being able to separate several sound sources has been introduced by Wang and Kaveh [49], and a related method exploiting higher order statistics has been recently proposed by Bourennane and Bendjama [50].

Time delay estimation based methods

The localization methods in the first two categories require a search over the potential directions of sound sources. With time delay estimate (TDE) based methods searching can be avoided. In these methods the relative delays between microphones in an array are first calculated. This is typically done by finding the maximum of a measure of similarity between the signals of each pair of microphones as a function of their relative time lag. Delays are then used to determine the location or direction of a sound source.

A common measure of similarity is the generalized cross-correlation (GCC) function [51]. GCC is defined as the inverse Fourier transform of the weighted cross-spectrum of two microphone signals

$$R_{GCC}(\tau) = \int_{-\infty}^{\infty} \psi(f) G_{x1x2}(f) e^{j2\pi f\tau} df \qquad (9)$$

where τ is the time lag, $\psi(f)$ is the weight defined as a function of frequency, and $G_{x1x2}(f)$ is the cross spectral density function

$$G_{x1x2}(f) = X_1(f) X_2^*(f) \qquad (10)$$

where $X_1(f)$ and $X_2(f)$ are the Fourier transforms of microphone signals $x_1(t)$ and $x_2(t)$, respectively, and * denotes the complex conjugate. An estimate for the time delay now corresponds to the lag value maximizing $R_{GCC}(\tau)$. An optimal weighting function depends on source signal, the acoustical environment, and possible interfering noise. The standard cross-correlation function is obtained with $\psi_{XC}(f) = 1$. Two other typical choices are

$$\psi_{SCOT}(f) = \frac{1}{\sqrt{G_{x1x1}(f) G_{x2x2}(f)}} \qquad (11)$$

resulting in the so called smoothed coherence transform, and

$$\psi_{PHAT}(f) = \frac{1}{|G_{x1x2}(f)|} \qquad (12)$$

which is called the phase transform.

In discrete time implementations some kind of interpolation is needed in order to get better estimates for the time delays. Instead of finding the maximum of the phase transform, Brandstein and Silverman [44] fitted a line to the phase response of the cross-spectrum of two microphone signals and used its slope to estimate the delay. Jacovitti and Scarano [52], on the other hand, propose parabolic interpolation of discrete cross-correlation function around its maximum value. They also suggest replacing the computationally expensive cross-correlation with square difference function (ASDF) or average magnitude difference function (AMDF) defined as

$$R_{ASDF}(\tau) = \frac{1}{N} \sum_{k=1}^{N} [x_1(kT) - x_2(kT + \tau)]^2 \qquad (13)$$

and

$$R_{AMDF}(\tau) = \frac{1}{N} \sum_{k=1}^{N} |x_1(kT) - x_2(kT + \tau)| \qquad (14)$$

where N is the number of samples and T is the sampling period. With these signal distance measures the estimate for time delay corresponds to the lag value minimizing the functions.

The second step of TDE localization is the determination of source location based on the estimated delays. The TDE of each pair of microphones corresponds to the other half of a hyperboloid with two sheets. Ideally the location estimate for a source is found as the intersection of all hyperboloids defined by all microphone pairs in an array. In practice, however, there is always error in the TDEs and source location needs thus to be estimated as the best fit to available data. Several different error criteria for this fit have been discussed in literature. In general the location estimation involves minimization of a set of non-linear functions and thus requires numerical search methods [44]. However, a closed form solution using exactly three TDEs derived from four different microphone pairs exists [53]. This does not use any averaging to reduce errors caused by noisy TDEs, but it is computationally light and can be readily used with the HUT microphone probe. A number of closed form approximations to be used with a larger number of TDEs have also been described in literature. For a comparison see [54].

Due to inaccuracy in the determination of TDEs, estimation of distance is not reliable for sound sources far from a microphone array compared to the dimensions of the array. An estimation method for direction only based on TDEs is discussed in [55]. However, for remote sources the estimation of angle of arrival can be alleviated by approximating the hyperboloid surfaces by cones having a vertex halfway between corresponding pair of microphones. With this kind of approximation the angle of arrival related to the line combining the microphones can be written in the form

$$\theta = \cos^{-1}\left(\frac{c\tau}{d}\right) \tag{15}$$

where c is the speed of sound, τ is the TDE, and d is the distance between the microphones. In [56] two such angles calculated from orthogonal and concentric pairs of microphones were used to define two lines of potential locations in 3-D space, one for each side of the plane formed by the microphones. The actual location of a source was then approximated with a linear intersection of the lines defined by two or more microphone quadruples.

With the HUT microphone probe the angle of arrival of a sound source can be unambiguously defined with three existing orthogonal and concentric pairs of microphones located on x-, y-, and z-coordinate. For better resolution it is advisable to use the outer pairs. The procedure is outlined as follows. Let τ_x, τ_y, and τ_z be the TDEs of the microphone pairs on x-, y-, and z-axis, respectively. With Eq (15) and some simple trigonometric manipulations the elevation angle can be written in the form

$$\phi = \tan^{-1}\left(\frac{\tau_z}{\sqrt{\tau_x^2 + \tau_y^2}}\right) \tag{16}$$

and the azimuth is given by

$$\theta = \tan^{-1}\left(\frac{\tau_y}{\tau_x}\right) \tag{17}$$

where the correct quadrant has to be determined from the signs of τ_x and τ_y.

Reliable time delay estimation is an essential prerequisite for determining the location with the previous methods. Several modifications to the basic estimation scheme have been developed in order to improve performance in noisy, multi-source, or reverberant environments. Zhang and Er [57] introduced a hybrid of TDE and coherent signal-subspace [49] methods for source localization. Benesty [58] used eigenvalue decomposition to derive better TDEs in reverberant environments. Griebel and Brandstein [59] argued that reduction of cross-correlation functions to single TDE parameters causes problems in reverberant conditions. Instead, they estimated a delay vector directly based on cross-correlations between different microphone pairs. Finally, Liu et al [60] introduced an interesting method for localization of multiple independent sound sources, incorporating features from models of human binaural processing.

Applications of source localization

Passive source localization methods have been typically used for automatic beam steering in systems where isolated pickup of selected sources is desired. Application areas include for example teleconferencing, handsfree telecommunication in a car environment, hearing aids, and speech recognition [44]. In video conferencing and surveillance systems, steering of a camera based on acoustical source localization has also been discussed.

In many active systems only directional localization needs to be defined using the previous methods. Locations of objects producing low order reflections can then be easily determined based on angles of arrival and times of propagation from the active sound source to the receiving array. Sonars are typical examples of active systems localizing surrounding objects this way.

Room response measurements are actually not that different from sonar applications, although their emphasis is not on finding out the geometry of a room but on the effect of the geometry and materials on the soundfield. TDE based methods can, for instance, be used to give more accuracy to directional analysis of measured responses. Passive methods can, on the other hand, be utilized in directional noise measurements and localization of noise sources. Passive estimation of room responses and geometry is, however, very difficult since it is not easy to separate several coherent reflected signals from each other without exact knowledge of the original signal.

CONCLUSIONS AND FUTURE WORK

Several methods related to application of microphone arrays were discussed. The measurement section introduced methods for directional measurement, analysis, and visualization of room responses. In the context of recording applications, several directional microphone arrays and techniques for surround sound recording were discussed. Some special directional systems related to acoustic communications applications were also described. Finally, the problem of sound source localization was addressed and the principles of time delay estimation based localization were introduced. In each application case special emphasis was given to methods that can be used with the HUT 3-D microphone probe.

Future work consists of incorporating source localization techniques into measurement applications. Automatic extraction and analysis of low-order reflections and propagation paths in a room from measured directional responses is one such task. Effective separation of single reflections from response measurements could provide new means for analysis of reflection properties of the walls of the room. New directional descriptors for room and concert hall acoustics are also being developed.

In passive applications, better distance localization based on known or estimated properties of a room, instead of properties of the emitted signal, is one of the research goals. Future work also includes examination of microphone array localization properties in relation to the human hearing, thus connecting microphone array techniques with auditory modelling and binaural processing.

ACKNOWLEDGEMENTS

This study is part of the VÄRE technology program, project TAKU (Control of Closed Space Acoustics), funded by Tekes (National Technology Agency). J. Merimaa is supported by the Graduate School in Electronics, Telecommunication and Automation.

REFERENCES

[1] T. Peltonen, T. Lokki, B. Goutarbès, J. Merimaa, and M. Karjalainen, "A System for Multi-Channel and Binaural Room Response Measurements," *AES 110th Convention*, Amsterdam, The Netherlands, 2001.

[2] T. Peltonen, "*A Multichannel Measurement System for Room Acoustics Analysis*," M. Sc. Thesis, Helsinki University of Technology, Department of Electrical and Communications Engineering, Laboratory of Acoustics and Audio Signal Processing, 2000.

[3] A. J. Berkhout, D. de Vries, and J. J. Sonke, "Array technology for acoustic wave field analysis in enclosures," *J. Acoust. Soc. Am.*, Vol. 102, No. 5, Pp. 2757–2770, 1997.

[4] J. Eargle, "*The Microphone Book*," Focal Press, Boston, MA, USA, 2001.

[5] H. F. Olson, "Directional Microphones," *J. Audio Eng. Soc.* 1967, Reprinted in Microphones, An Anthology. Vol. 1 – Vol. 27. 1979.

[6] G. W. Elko, "Steerable and variable first-order differential microphone array," *U.S. Patent*, No. 6,041,127.

[7] H. Okubo, M. Otani, R. Ikezawa, S. Komiyama, and K. Nakabayashi, "A system for measuring the directional room acoustical parameters," *Applied Acoustics*, Vol. 62, No. 2, Pp. 203–215, 2001.

[8] J. L. Flanagan, "Use of acoustic filtering to control the beamwidth of steered microphone arrays," *J. Acoust Soc. Am.*, Vol. 78, No. 2, Pp. 423–428, 1985.

[9] J. Jarzynski and W. J. Trott, "Array shading for a broadband constant directivity transducer," *J. Acoust. Soc. Am.*, Vol. 64, No. 5, Pp. 1266–1269, 1978.

[10] D. B. Ward, R. A. Kennedy, and R. C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *J. Acoust. Soc. Am.*, Vol. 97, No. 2, Pp. 1023–1034, 1995.

[11] W. H. Kummer, "Basic Array Theory," *Proceedings of the IEEE*, Vol. 80, No.1, Pp. 127–140, 1992.

[12] R. Voles, "Spherical shell and volume arrays," *IEE Proc. Microw. Antennas Propag.* Vol. 142, No. 6, Pp. 498–500, 1995.

[13] A. D. Broadhurst, "Sparce Volume Array for Architectural Acoustic Measurements," *Acustica*, Vol. 50, No. 1, Pp. 33–38, 1982.

[14] F. J. Fahy, "*Sound Intensity*," Elsevier Science Publishers Ltd, Essex, England, 1989.

[15] F. J. Fahy, "Measurement of acoustic intensity using cross-spectral density of two microphone signals," *J. Acoust. Soc. Am.*, Vol. 62, No. 4, Pp. 1057–1059, 1977.

[16] J. Y. Chung, "Cross-spectral method of measuring acoustic intensity without error caused by instrument phase mismatch," *J. Acoust. Soc. Am.*, Vol. 64, No. 6, Pp. 1613–1616, 1978.

[17] H. Kuttruff and A. Schmitz, "Measurement of Sound Intensity by Means of Multi-Microphone Probes," *Acustica*, Vol. 80, Pp. 388–396, 1994.

[18] J. Merimaa, T. Lokki, T. Peltonen, and M. Karjalainen, "Measurement, Analysis, and Visualization of Directional Room Responses," *AES 111th Convention*, Preprint 5449, New York, NY, USA, 2001.

[19] S. P. Lipshitz, "Stereo Microphone Techniques... Are the Purists Wrong?" *J. Audio Eng. Soc.*, Vol. 34, No. 9, Pp. 716–744, 1986.

[20] G. Steinke, "Surround Sound — The New Phase. An Overview," *AES 100th Convention*, Preprint 4286, Copenhagen, Denmark, 1996.

[21] G. Theile, "Multichannel Natural Recording Based on Psychoacoustic Principles," *AES 108th Convention*, Preprint 5156, Paris, France, 2000.

[22] B. B. Bauer, R. G. Allen, G. A. Budelman, and D. W. Graveraux, "Quadraphonic Matrix Perspective — Advances in SQ Encoding and Decoding Technology," *J. Audio Eng. Soc.*, Vol. 21, No. 5, Pp. 342–350, 1973.

[23] D. Cooper and T. Shiga, "Discrete-Matrix Multichannel Sound," *J. Audio Eng. Soc.*, Vol. 20, No. 5, 1972.

[24] T. Yamamoto, "Quadraphonic One Point Pickup Microphone," *J. Audio Eng. Soc.*, Vol. 21, No. 4, Pp. 256–261, 1973.

[25] G. Theile and G. Plenge, "Localization of Lateral Phantom Sources," *J. Audio Eng. Soc.*, Vol. 25, No. 4, Pp. 196–200, 1977.

[26] J. S. Bamford and J. Vanderkooy, "Ambisonic Sound for Us," *AES 99th Convention*, Preprint 4138, New York, NY, USA, 1995.

[27] R. K. Furness, "Ambisonics — An Overview," in *Proc. AES 8th International Conference*, Pp. 181-189. 1990.

[28] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, Vol. 21, No. 1, Pp. 2–10, 1973.

[29] G. Martin, W. Woszczyk, J. Corey, and R. Quesnel, "Controlling Phantom Image Focus in a Multichannel Reproduction System," *AES 107th Convention*, Preprint 4996, New York, NY, USA, 1999.

[30] V. Pulkki, "Microphone techniques and directional quality of sound reproduction," *AES 112th Convention*, Munich, Germany, 2002.

[31] K. Farrar, "Soundfield microphone: Design and development of microphone and control unit," *Wireless World*, October 1979.

[32] K. Farrar, "Soundfield microphone — 2: Detailed functioning of control unit," *Wireless World*, November 1979.

[33] J. D. Johnston and Y. H. Lam, "Perceptual Soundfield Reconstruction," *AES 109th Convention*, Preprint 5202, Los Angeles, CA, USA, 2000.

[34] J. Bruck, "The KFM 360 Surround — A Purist Approach," *AES 103rd Convention*, Preprint 4637, New York, NY, USA, 1997.

[35] J. Klepko, "5-Channel Microphone Array with Binaural Head for Multichannel Reproduction," *AES 103rd Convention*, Preprint 4541, New York, NY, USA, 1997.

[36] M. Williams and G. Le Dû, "Microphone Array Analysis for Multichannel Sound Recording," *AES 107th Convention*, Preprint 4997, New York, NY, USA, 1999.

[37] G. W. Elko, J. E. West, R. A. Kubli, and J. P. McAteer, "Image-derived second-order differential microphones," J. Acoust. Soc. Am., Vol. 95, No. 4., Pp. 1991–1997, 1991.

[38] A. V. Oppenheim and R. W. Schafer, "*Digital Signal Processing*," Prentice-Hall, New Jersey, USA, 1975.

[39] G. M. Sessler, J. E. West, and M. R. J. Schroeder, "Toroidal Microphones," *J. Acoust. Soc. Am.*, Vol. 46, No. 1, Pp. 28–36, 1969.

[40] G. M. Sessler and J. E. West, "A Simple Second-Order Toroid Microphone," *Acustica*, Vol. 57, No. 4/5, Pp. 193–199. 1985.

[41] G. M. Sessler, J. E. West, and R. A. Kubli, "Unidirectional, second-order gradient microphone," *J. Acoust. Soc. Am.*, Vol. 86, No. 6, Pp. 2063–2066, 1989.

[42] Y. Ishigaki, M. Yamamoto, K. Totsuka, and N. Miyaji, "Zoom Microphone," *AES 67th Convention*, Preprint 1713, New York, NY, USA, 1980.

[43] W. R. Woszczyk, "A Microphone Technique Applying the Principle of Second-Order-Gradient Unidirectionality," *J. Audio Eng. Soc.*, Vol. 32, No. 7/8, Pp. 507–530, 1984.

[44] M. S. Brandstein and H. F. Silverman, "A Practical methodology for speech source localization with microphone arrays," *Computer Speech and Language*, Vol. 11, No. 2, Pp. 91–126, 1997.

[45] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Am.*, Vol. 78, No. 5, Pp. 1508–1518, 1985.

[46] G. Carter "Variance bounds for passively locating an acoustic source with a symmetric line array," *J. Acoust. Soc. Am.*, Vol. 62, Pp. 922-926, 1977.

[47] W. Hahn, "Optimum signal processing for passive sonar range and bearing estimation," *J. Acoust. Soc. Am.*, Vol. 58, Pp. 201–207, 1975.

[48] M. Wax and T. Kailath, "Optimum localization of multiple sources by passive arrays," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 31, Pp. 1210–1217, 1983.

[49] H. Wang, and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 33, Pp. 823–831, 1985.

[50] S. Bourennane and A. Bendjama, "Locating wide band acoustic sources using higher order statistics," *Applied Acoustics*, Vol. 63, Pp. 235–251, 2002.

[51] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 24, No. 4, Pp. 320–327, 1976.

[52] G. Jacovitti and G. Scarano, "Discrete Time Techniques for Time Delay Estimation," *IEEE Transactions on Signal Processing*, Vol. 41, No. 2, Pp. 525–533, 1993.

[53] B. T. Fang, "Simple Solutions for Hyperbolic and Related Position Fixes," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 26, No. 5, Pp. 748–753, 1990.

[54] Y. T. Chan and K. C. Ho, "A Simple and Efficient Estimator for Hyperbolic Location," *IEEE Transactions on Signal Processing*, Vol. 42, No. 8, Pp. 1905–1915, 1994.

[55] B. Berdugo, M. A. Doron, J. Rosenhouse, and H. Azhari, "On direction finding of an emitting source from time delays," *J. Acoust. Soc. Am.*, Vol. 105, No. 6, Pp. 3355–3363, 1999.

[56] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A Closed-Form Location Estimator for Use with Room Environment Microphone Arrays," *IEEE Transactions on Speech and Audio Processing*, Vol. 5, No. 1, Pp. 45–50, 1997.

[57] M. Zhang and M. H. Er, "An Alternative Algorithm for Estimating and Tracking Talker Location by Microphone Arrays," *J. Audio Eng. Soc.*, Vol. 44, No. 9, 1996.

[58] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Am.*, Vol. 107, No. 1, 2000.

[59] S. M. Griebel and M. S. Brandstein, "Microphone Array Source Localization Using Realizable Delay Vectors," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.

[60] C. Liu, B. C. Wheeler, W. D. O'Brien Jr., R. C. Bilger, C. R. Lansing, and A. S. Feng, "Localization of multiple sound sources with two microphones," *J. Acoust. Soc. Am.*, Vol. 108, No. 4, Pp. 1888–1905, 2000.