Applied Acoustics 74 (2013) 987–1000

Contents lists available at SciVerse ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

On the use of microphone arrays to visualize spatial sound field information

Francesco Martellotta *

Dipartimento di Scienze dell'Ingegneria Civile e dell'Architettura, Politecnico di Bari, via Orabona 4, I-70125 Bari, Italy

ARTICLE INFO

Article history:
Received 3 September 2012
Received in revised form 12 November 2012
Accepted 15 February 2013

Keywords:
Microphone arrays
Sound field visualization
Room acoustics

ABSTRACT

Microphone arrays represent today a state of the art solution to many acoustic problems. In architectural acoustics, for example, one of the most interesting applications is the possibility to analyse the directional information associated to a given reflection. Ambisonics microphones could provide similar information based on zeroth and first order spherical harmonic decomposition, but larger microphone arrays allow the determination of higher order components providing even better accuracy. In this case, directional information may be obtained through beamforming techniques that, although potentially more accurate and capable of resolving simultaneous reflections, are computationally heavier and provide a "discrete" sampling of the sound field. The paper compares the localization accuracy of a 32 channel microphone array by processing its output using a simple Ambisonics decomposition and a spatial sampling carried out using 32 "virtual" third-order hyper cardioid microphones. In addition, a comparison with conventional Ambisonics microphones is provided in order to point out possible differences. Results show that, when single reflections are involved and the sound field is highly polarized, the Ambsionics decomposition given by the microphone array gives good accuracy over the whole spectrum, while conventional Ambisonic microphones shows less stable results and greater variations as a function of frequency. Spatial sampling is intrinsically less accurate but allows a clearer resolution of simultaneous reflections.

1. Introduction

Sound source identification and localization has been one of the most interesting outcomes of early sound intensity measurements [1,2]. The most widespread applications are based on 1D transducers (requiring two spaced pressure microphones) so that spatial information about sound sources can be grasped only by a proper scanning of the space. However, there have been applications involving 3D transducers (using six pressure microphones). In this way the three Cartesian components of the particle velocity could be determined simultaneously and, hence, the corresponding direction of arrival of a given sound. The most recent developments in this field are represented by the pressure–velocity sensors, Ref. [3] showing quite interesting performance.

A viable alternative, although originally not well understood in its potential, was proposed by Gerzon [4,5], who first introduced the idea of Ambisonics decomposition of the sound field. As it will be described in detail in the next section, Ambisonics is based on the measurement of the sound field by means of four nearly coinciding microphones arranged on a tetrahedral volume. Such set of signals (named "A-format"), after a proper processing provides an omni-directional (pressure information) and three figure-of-eight signals oriented along the Cartesian axes (particle velocity information), known together as "B-format". This technique ideally corresponds to decomposing the sound field according to the zeroth and first order spherical harmonics, from which directional information can be grasped, and reproduction over several spatial loudspeaker arrays is made possible. This method, originally patented in 1975 by Gerzon and Craven [5], is now implemented on many commercially available Ambisonics microphones, offering a wide spectrum of choices. One of the most interesting aspects of this technique is the possibility to determine the directional characteristics of the sound field by means of just four channels. However, the intrinsic limits of the hardware (microphones and sound processing devices) may sometimes affect the accuracy of the localization [6]. In addition, when used to playback the recorded sound field this system has several disadvantages [7] due to the low order of the spherical decomposition. In fact, the reproduced signals tend to be highly coherent, resulting in colouring and distortion of the spatial image, so preventing a satisfactory subjective listening experience. In order to get improved realism in sound reproduction, higher-order Ambisonics components need to be determined [8], but this requires the use of more microphones than just the four used for first-order Ambisonics.

* Tel.: +39 080 5963631.
E-mail address: f.martellotta@poliba.it

0003-682X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2013.02.004

However, after many years during which large microphone arrays were only developed by isolated research groups [9–13], mostly focussing on the more appropriate combination of hardware setup and mathematical processing to optimize a reliable retrieval of spatial information [14–18], a few commercial solutions are now available. This is opening new perspectives also for end-users not interested in (or not able to understand) the complex mathematical details that lay behind the microphone array theory. In fact, most commercial solutions include a basic set of tools to process the signals and interpret the results according to the desired purpose of the measurement. Different fields of the acoustics, from industrial and machinery noise identification, to forensic applications, architectural acoustics, and virtual sound field reproduction are now interested in exploring the extremely wide potential offered by such microphones.

The present paper aims at discussing the performance of such type of microphones in the specific field of sound field analysis for architectural acoustic purposes, comparing different "low level" ways to use the microphone outputs and, at the same time, comparing the accuracy of the spatial information resulting from microphone arrays with that resulting from traditional 4-channel microphones.

2. Overview of theoretical background

2.1. Ambisonics decomposition

Microphone arrays based on Ambisonic decomposition of the sound field may provide two different outputs. The raw signal, corresponding to the four capsules is the so called A-format, while B-format corresponds to zeroth and first order components of a spherical decomposition of the sound field. Traditionally such components are identified as *W*, *X*, *Y*, and *Z*. The first one (0th order component) represents the omni-directional response of the microphone at the centre of the array and its value is traditionally normalized by dividing the amplitude by 2 in order to have four signals with comparable amplitude. In the following discussion only the "ideal" *W* signal will be considered, assuming therefore that any "normalization" is immediately compensated before any subsequent analysis. The latter three (1st order components) correspond to figure-of-eight (or better 1st order dipoles) responses of microphones oriented along the three Cartesian axes, so that they provide the sound pressure multiplied by the vector of the sound direction along that axis. In other words, considering that particle velocity (u) is a vector quantity oriented along the direction of sound propagation, and that (at least for plane waves) u is proportional to sound pressure (p) through characteristic impedence Z_0 ($u = p/Z_0$), *X*, *Y*, and *Z* may also be assumed as the Cartesian components of the particle velocity [19] and used to determine sound intensity properties [20]. All this stated, considering the relationship between sound intensity, sound pressure and particle velocity, the following equations may be written for the instantaneous intensity components (time and position dependencies are omitted for brevity):

$$\begin{aligned} I_x &= p \cdot u_x = w \cdot x / Z_0 \\ I_y &= p \cdot u_y = w \cdot y / Z_0 \\ I_z &= p \cdot u_z = w \cdot z / Z_0 \end{aligned} \tag{1}$$

where w, x, y, and z are the output signals of the B-format microphone. And hence the norm of the vector is:

$$|I| = \sqrt{I_x^2 + I_y^2 + I_z^2} = \frac{1}{Z_0}\sqrt{(w \cdot x)^2 + (w \cdot y)^2 + (w \cdot z)^2} \tag{2}$$

In addition, considering that energy density in a reverberant field [2] is given by:

$$E = \frac{1}{2}\rho_0\left[\frac{p^2}{Z_0^2} + u^2\right] = \frac{1}{2}\frac{\rho_0}{Z_0^2}\left(w^2 + x^2 + y^2 + z^2\right) \tag{3}$$

it is now possible to calculate the "degree of diffusion" (ψ) of the sound field as the ratio between the norm of the time-averaged intensity vector (given by the vector sum of the instantaneous values), divided by the speed of sound c, and the time-averaged density, so that:

$$\psi = 1 - \frac{|\langle I/c \rangle|}{\langle E \rangle} \tag{4}$$

In this way ψ may vary between 0 and 1, where 0 corresponds to a fully polarized sound field (resulting from direct sound or strong reflections), and 1 corresponds to a fully diffuse sound field (resulting from an intensity vector that, being the sum of contributions from uniformly distributed directions, is very small compared to fluctuating pressure). Other approaches have been proposed to calculate this parameter [21], but a discussion on this topic is beyond the scope of this paper.

With reference to the B-format components and replacing the average over a given time window with a discrete summation, Eq. (4) yields:

$$\psi = 1 - \frac{2\sqrt{\left(\sum w \cdot x\right)^2 + \left(\sum w \cdot y\right)^2 + \left(\sum w \cdot z\right)^2}}{\sum (w^2 + x^2 + y^2 + z^2)} \tag{5}$$

A further development of this way of interpreting the data is that, given the I_x, I_y, and I_z components of the sound intensity, the azimuth (θ) and elevation (ϕ) of the direction of arrival of the sound at a given time may be easily calculated from the following equations:

$$\theta = \arctan\left(\frac{I_y}{I_x}\right); \quad \phi = \arctan\left(\frac{I_z}{\sqrt{I_x^2 + I_y^2}}\right) \tag{6}$$

So, combining together the direction of sound with its intensity and degree of diffusion, sound field properties may be reproduced using innovative multichannel techniques [22] or, as in the present case, used to obtain a spatial map of the sound intensity as a function of time. Different approaches may be followed in order to graphically render this information, depending on the purpose to be obtained. The most simple and, at the same time, detailed approach may be that of using an equirectangular projection which transposes angular coordinates over a Cartesian plane, using colours (or grayscale) to represent the intensity of sound in that point. This way of mapping the information may also be particularly useful in combination with a panoramic view of the room (such as those used in virtual tours) to allow easy identification of actual origin of the sound reflections.

2.2. Beamforming

The basic output of any microphone array is a set of raw signals recorded by each single microphone capsule (conceptually equivalent to Ambisonic A-format). However, in order to make such information of any practical use, a complex mathematical processing (named "beamforming") is required. Beamforming takes into account the delays and phase changes of each of the signals arriving to each of the microphones in order to accurately determine the direction of the impinging sound wave (which, to keep calculations simpler, is assumed to be a plane wave).

The general idea behind beamforming is that, given a single monopole source at an assumed position, and a set of M microphones in an array, a proper processing of the output signals of the microphones via a set of linear filters may allow an estimate