Interactive Dynamic Soundfield Rendering with Visual Feedback

F. Antonacci, L. Bianchi, D. Marković,
A. Canclini, A. Sarti, S. Tubaro
Politecnico di Milano
Image and Sound Processing Group
p.zza Leonardo da Vinci 32
20133 Milano - IT

R. Magalotti
B&C Speakers
via Poggiomoro, 1
Bagno a Ripoli
50012 Firenze - IT

Abstract

In this demonstration we show a system for the rendering of directional sources based on a loudspeaker array. The rendering methodology is based on the idea of the decomposition of the sound field in plane waves. The speaker array is divided into overlapping sub-arrays, each generating a plane-wave component. Individual plane waves are weighed by the desired directivity pattern. We demonstrate the capability of the rendering system in an acoustic scene consisting of two sources (a male and a female speaker) partially occluded by walls. The listener can freely move in the scene to appreciate the spatial audio effect. A tracking and visualizazion system based on a Kinect camera and a 2D display is used to provide a visual feedback of the acoustic scene from the viewpoint of the listener.

1. Introduction and motivations

The problem of sound field rendering is of growing interest in multimedia for potential applications in immersive gaming, telepresence and all the scenarios in which the delivery of a realistic sound impression is an issue. In particular, the rendering of the directivity is of crucial importance for improving the immersivity. There are evidences in the literature [4] that the disparities in the sound field at different listening positions created by the directional properties of sources greatly contribute in making realistic and immersive the sound scene. Some works in the literature focus on the problem of rendering directional cues of the source, in both nearfield [1] (for Ambisonics) and farfield [7] (for WFS).

In this demonstration we show an alternate route, partially inspired by the novel idea of plenacoustic rendering [2]. In particular, the loudspeaker array acts as an Emission Window (EW) and it is subdivided into sub-arrays, each in charge of rendering the planar portion of the acoustic wavefront that, departing from the virtual acoustic source, passes through the considered sub-array. The overall acoustic wavefront is obtained through the superposition effect, after a phase alignment stage, which guarantees that the planar wavefronts produced by all the sub-arrays are aligned in phase. The beamforming filter of each sub-array is obtained by means of the Superdirective Beamforming (SD), described in [6].

The directivity is implemented in the plenacoustic rendering framework by weighing the planar wavefronts of each sub-array through the directivity pattern of the virtual source to be rendered.

The acoustic scene rendered consists of two sources partially occluded by two walls. The listener can freely move in the acoustic scene to appreciate the different attenuation of the sources caused by the occlusion from the walls. A Kinect camera and a 2D display are also used to provide to the listener a visual feedback of the scene from its location.

2. Scientific and technical description

Let us consider an array of M loudspeakers. Transducers locations in Cartesian coordinates are $\mathbf{x}_m = [x_m y_m]^T,\ m = 0, ;M-1$. We adopt the linear array configuration with uniform spacing q along the x axis and we fix the reference frame so that $y_m = y_0$ for all the loudspeakers. The line on which the array lies determines two half-spaces. We are interested in rendering the sound field in the half-space $y > y_0$. The virtual source emits the signal $s(t)$ and it is also characterized by the polar pattern $\tilde{D}(\alpha, \omega)$, where α is the angle with respect to the x axis and ω is the radial frequency. Our goal is to render the sound field generated by the virtual source, including the radiation pattern. As depicted in Fig. 1, the array is decomposed into maximally overlapped sub-arrays, each consisting of W adjacent loudspeakers. The number of sub-arrays is $M - W + 1$. Our goal is to operate a plane-wave decomposition. In a 2-dimensional sound field the dispersion

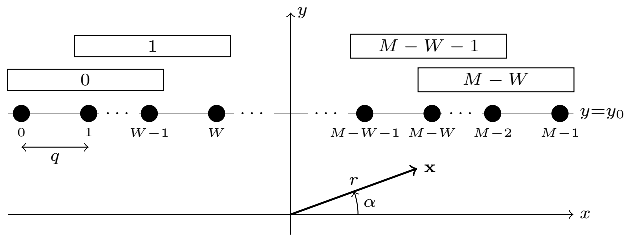

Figure 1. Reference frame and decomposition of the loudspeaker array into sub-arrays.

equation can be written as $k^2 = (\omega/\nu)^2 = k_x^2 + k_y^2$, where k_x and k_y are the components of the wavenumber vector $\mathbf{k}$ and ν is the speed of sound. Each sub-array is in charge of reproducing a plane-wave component with travel direction α_i, where α_i is the angle formed by the x axis and the position vector of the central loudspeaker of the ith sub-array. The space-frequency pressure sound field $\tilde{p}(\mathbf{x}, \omega)$ is approximated through Plane Wave Decomposition (PWD) by

$$\tilde{p}(\mathbf{x}, \omega) \approx \sum_{i=0}^{M-W} \tilde{d}_i(\omega) e^{j\frac{\omega}{\nu}(\hat{\mathbf{k}}_i^T \mathbf{x})} = \tilde{p}_{\mathrm{P}}(\mathbf{x}, \omega), \qquad (1)$$

where $\tilde{d}_i(\omega)$ is related to the directivity $\tilde{D}(\alpha_i, \omega_l)$ through

$$\tilde{d}_i(\omega) = \tilde{D}(\alpha_i, \omega). \qquad (2)$$

3. Implementation and use

The overall scheme of the rendering system is shown in Fig. 3.1. In this Section we show how each component is implemented.

3.1. Beamforming

In order to render the individual plane wave components we have adopted the data-independent reformulation of the MVDR beamformer, at our knowledge introduced in [6]. This choice maximizes the directivity with the limited numbers of transducers available in each sub-array. In order to include in the framework wideband signals, we operate a subdivision of the frequency axis into L sub-bands. Let ω_l be the central frequency of the lth sub-band and

$$\omega_{\mathrm{s},i} = \frac{\omega_l}{\nu} q \cos \alpha_i \qquad (3)$$

be the spatial frequency of a plane wave with time radial frequency ω_l and travel direction α_i, $i = 0, \ldots, M - W$.

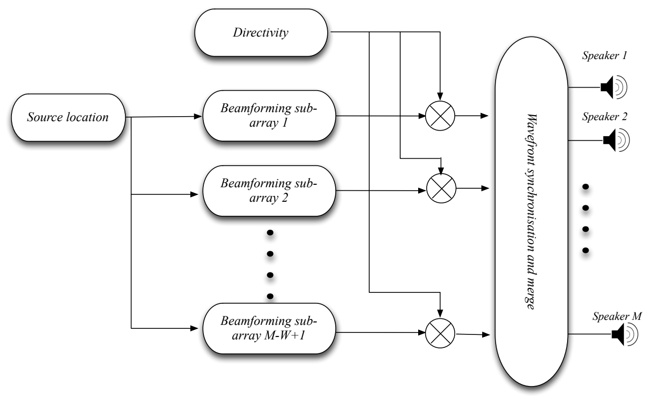

Figure 2. The compuational scheme to derive the filters applied to each loudspeaker.

Denoting the noise covariance matrix for the lth sub-band with $\tilde{\mathbf{R}}(l)$, the MVDR spatial filter $\tilde{\mathbf{f}}_i(l)$ is [3]

$$\tilde{\mathbf{f}}_i(l) = \frac{\tilde{\mathbf{R}}^{-1}(l)\tilde{\mathbf{a}}_i(l)}{\tilde{\mathbf{a}}_i^H(l)\tilde{\mathbf{R}}^{-1}(l)\tilde{\mathbf{a}}_i(l)}, \qquad (4)$$

The element at row r and column c of the coherence matrix is [6]

$$\tilde{\mathbf{\Gamma}}(l)|_r^c = \mathrm{sinc}\left(\frac{2\pi\|\mathbf{x}_r - \mathbf{x}_c\|}{c}\frac{F_s l}{L}\right), \qquad (5)$$

where $\mathbf{x}_r$ and $\mathbf{x}_c$ are the coordinates of the r and cth elements of the loudspeaker array, respectively; and F_s is the sampling frequency.

3.2. Directivity, Wavefront synchronization and merge

The rendering coefficient relative to the mth loudspeaker can be expressed as

$$h_m(l) = \sum_{i=i_0}^{i_1} \tilde{f}_{i,m-i}(l) \cdot e^{-j\omega_l \boldsymbol{\tau}_{i,m-i}} \cdot d_i(l) \cdot z(m-i), \quad (6)$$

where $i_0 = \max(0, m-W)$, $i_1 = \min(M-W, m+W-2)$ and $d_i(l) = d_i(\omega_l)$. The index $i = i_0, \ldots, i_1$ spans the sub-arrays that include the mth loudspeaker and $\boldsymbol{\tau}_{i,m-1}$ performs the alignment of the beamforming filters from different sub-arrays. The windowing function z is a squared Hanning window [5] of length W. Further details can be found in [2]. Coefficients $h_m(l)$ are organized in a matrix $\mathbf{H}$ of dimensions $M \times L$. Finally, the M time-domain rendering filters $\mathbf{h}$ are obtained with the L-point Inverse Fourier transform over the rows of the matrix $\mathbf{H}$.

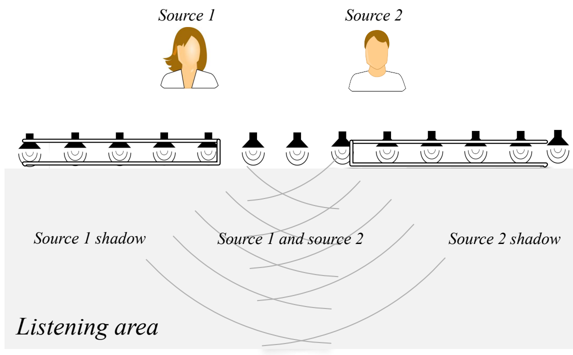

Figure 3. The acoustic scene rendered by the rendering system.

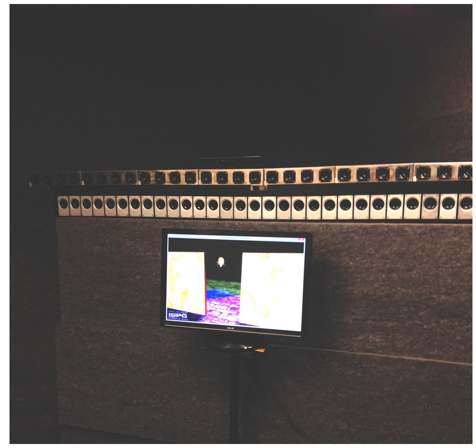

Figure 4. Example of the installation of the system

3.3. Example of Use

The scene rendered in the demonstration is shown in Fig. 3. Two sources (a female and a male speakers) are placed behind the corner of two parallel walls that form an aperture in front of the listening area. Due to the reciprocal configuration of walls and speakers, we can identify three distinct zone in the listening area: a region where only the source 2 is audible, a region where both sources are audible , and a region where only source 1 is audible. Notice also that the region where both sources are audible narrows as the listener moves away from the aperture, until both sources become inaudible. The loudspeakers form a linear array two meters long on the lying line of the virtual walls. Both sources are characterized by a proper cardioid directivity $c(l)$ and are oriented towards the listening area. The directivity functions $d_i^{(j)}(l), j = 1, 2$ defined in 1 are related to$c(l)$ through

$$d_i^{(j)}(l) = \begin{cases} c(l) & \text{in the audible area} \\ \frac{c(l)}{1000} & \text{in the shadow area} \end{cases} \tag{7}$$

The actual average attenuation of the sources in the shadow areas has been measured to be of approximately 25 dB, quite close to the desired effect. In this demonstration we did not include diffractive effects on the part of the corners of the walls. Figure 3.3 shows an example of installation of the system. In order to provide the listener a visual feedback, a tracking system is included, based on a Kinect camera and a display. The camera localizes in space the listener, and the display visualizes the acoustic scene from her/his viewpoint. The 2D display in front of the array only helps understanding the border between the different regions in the listening area and appreciating the correspondence between audio and visual cues. The tracking and audio systems are independent and no transmission between them is in place.

4. Conclusions and future developments

In this demonstration we proposed a system for rendering directional sources using a loudspeaker array, based on the Plane Wave Decomposition of the desired sound field. We envision this framework to be the engine of a more sophisticated rendering system which adopts an object-based description of the acoustic scene.

References

[1] J. Ahrens and S. Spors. Rendering of virtual sound sources with arbitrary directivity in higher order ambisonics. In *Audio Engineering Society Convention 123*, 2007.

[2] L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro. Rendering of directional sources through loudspeaker arrays based on plane wave decomposition. In *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pages 13–18, 2013.

[3] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969.

[4] E. Corteel. Synthesis of directional sources using wave field synthesis, possibilities, and limitations. *EURASIP Journal on Advances in Signal Processing*, 2007, Feb. 2007.

[5] F. J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

[6] K. Kumatani, L. Lu, J. McDonough, A. Ghoshal, and D. Klakow. Maximum negentropy beamforming with superdirectivity. In *European Signal Processing Conference (EUSIPCO)*, Aalborg, DK, Aug. 2010.

[7] E. Verheijen. *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1998.