

How to Design a Delay-and-Sum Beamformer for Rigid Rotationally Symmetric Arrays?

Karim Helwani¹, Sascha Spors², and Herbert Buchner³

¹Quality and Usability Lab, Telekom Innovation Laboratories, ³Machine Learning Group,

^{1,3}Technische Universität Berlin, 10587 Berlin, Germany

² Institute of Communications Engineering, Universität Rostock, 18119 Rostock, Germany

Email: karim.helwani@telekom.de, sascha.spors@uni-rostock.de, hb@buchner-net.com

Abstract—In this paper we present two novel techniques with low computational complexity and high robustness for beamforming using rigid microphone arrays. For the first one we show which approximations have to be made in terms of applying techniques that are equivalent to delay-and-sum beamforming on rigid microphone arrays. In the second approach we exploit the technique of focused sinks for creating virtual linear microphone arrays with flexible dimensions and spacing. As a result, simple techniques for beamforming using linear arrays can be applied. We derive our approaches analytically, state their theoretical limits and provide simulation results and discuss the results.

I. INTRODUCTION

Using cylindrical or spherical microphone arrays for analyzing soundfields in two or three dimensions respectively is obviously a good choice due to the rotational symmetry. Intuitively, with a reasonable arrangement of microphones on the surface, sound capture from all directions should work evenly well. Soundfield analysis bases on the so-called (Kirchhoff-)Helmholtz integral equation and depending on the application scenario the mentioned integral equation can be formulated for an exterior or interior domain [1].

A common model for acoustic scene analysis is to assume the soundfield to be a superposition of an infinite number of plane waves, arriving from all directions at the sensor surface. Hence, acoustic scene analysis can be regarded as equivalent to a plane wave decomposition: The goal is to determine the frequency dependent amplitudes of the plane waves from the captured microphone signals. This can also be understood as spatial beamforming, by applying steerable directional filters to the array output with an (ideally infinitely) narrow main lobe in the desired look direction and maximal suppression elsewhere.

It is convenient for cylindrical / spherical boundaries to express the pressure and its gradient in terms of the solutions of the homogeneous wave equation in cylindrical / spherical coordinates. This results in an expansion of the angular components into exponential functions / spherical harmonics and the radial components in terms of (spherical) Hankel functions. These principles are termed modal beamforming techniques and have been used for a number of practical realizations using spatial distributions of microphones on a cylinder / sphere [2], [3], [4], [5], [6].

For a practical realization of the above discussed principle it is desirable to measure only the acoustic pressure at a finite

number of positions on the cylindrical / spherical boundary to derive the expansion coefficients of the observed sound field in terms of circular / spherical harmonics. However, this typically results in nonuniqueness of the expansion coefficients. A number of techniques have been proposed to overcome this problem, e.g., scattering at a rigid sphere [3], [5] or measuring at multiple radii [7]. Modal beamforming, assuming maximum directivity design maintains the good high-frequency directivity at frequencies up to the aliasing frequency of the array, but at the expense of poor robustness at low frequencies.

The requirements for high quality audio are on the one hand a high cutoff frequency which is often related to a high number of sampling points and hence, high complexity and on the other hand a high white-noise-gain as a measure for the attenuation of diffuse white noise from the sensors to the array output. It is therefore desired to find efficient approximations for fullband beamforming techniques. Delay-and-sum beamforming techniques guarantee maximum white-noise-gain, or robustness, and therefore the directivity is best at high frequencies and reduces towards the low frequencies [8].

Since a rigid surface adds a scattered field to the original soundfield, techniques based on delay-and-sum beamforming cannot be transferred straightforwardly. In this paper we propose a microphone selection criterion based on the sound intensity vector [9] that allows us to perform beamforming with low computational complexity. We derive analytically the beamforming filters and state the limits of our approach. The idea is to create a virtual soundfield which is conjugate in the wave-number space (k -space) to the desired look direction. Due to the matched filter principle the desired component will be extracted. In the limiting case of a reference line being coincident with the array surface we get a delay-and-sum beamformer. To create this conjugate virtual soundfield we use elements of the wave field synthesis theory and the time reversal principle.

II. DERIVATION OF BEAMFORMING WEIGHTS FOR DELAY-AND-SUM BEAMFORMER

A. Fundamentals

In the following we briefly review some fundamentals of wave field synthesis which will form the basis of our approaches. Wave field synthesis aims at reproducing a sound

field of a particular source, e.g., point or line source, using a distribution of secondary sources surrounding a listening area. Secondary sources are realized by loudspeakers in practical implementations. The Kirchhoff-Helmholtz integral occupies a central position in the theory of the wave field synthesis. It states that the pressure at any point inside a bounded source free volume is uniquely determined by the pressure and its normal gradient at the boundary of this volume Ω_S . This reads

$$P(\mathbf{x}, \omega) = \oint_{\Omega} \left(-\frac{\partial}{\partial \mathbf{n}} P(\mathbf{x}_0, \omega) G(\mathbf{x}|\mathbf{x}_0, \omega) + \frac{\partial}{\partial \mathbf{n}} G(\mathbf{x}|\mathbf{x}_0, \omega) P(\mathbf{x}_0, \omega) \right) \Big|_{\mathbf{x}_0 \in \Omega_S} d\Omega. \quad (1)$$

$G(\mathbf{x}|\mathbf{x}_0, \omega)$ corresponds to the Green's function which is the solution of the inhomogeneous wave equation for a spatio-temporal Dirac pulse. Under a free-field assumption, the Green's function describes a monopole source. The directional gradient of the Green's function under a free field assumption can be interpreted as a dipole source. Arbitrary convex secondary source distributions are usually treated as locally planar (linear). This approximation originates from the scattering theory and is known as Kirchhoff or physical optics approximation [10] and holds for small wave lengths compared to the dimensions of the secondary source distribution. A step towards specializing the Kirchhoff-Helmholtz integral to linear secondary source distributions is done using the Rayleigh integral and assuming the secondary source distributions to be infinitely long. Employing this approximation requires a rule for secondary source selection. A secondary source is selected if the normal vector \mathbf{n} of the secondary source and the averaged acoustic intensity vector of the virtual source $\mathbf{I}(\mathbf{x}_0, \omega)$ form an acute-angle. In this study we are interested in rigid arrays (arrays with sound hard surfaces). The gradient of the pressure on the surface of a rigid array vanishes and we have to approximate the secondary source distribution as a dipole distribution. Hence, the synthesis equation using a continuous distribution of dipoles and using the above mentioned approximations and assumptions reads [9]

$$P(\mathbf{x}, \omega) \approx \oint_{\Omega} \left(a(\mathbf{x}_0) \frac{\partial}{\partial \mathbf{n}} G(\mathbf{x}|\mathbf{x}_0, \omega) P(\mathbf{x}_0, \omega) \right) \Big|_{\mathbf{x}_0 \in \Omega_S} d\Omega, \quad (2)$$

with

$$a(\mathbf{x}_0) = \begin{cases} 1, & \text{if } \langle \mathbf{I}(\mathbf{x}_0, \omega), \mathbf{n}(\mathbf{x}_0) \rangle > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Hence, the driving function $D(\mathbf{x}_0)$ for synthesis of a specific sound field within Ω_S is simply the pressure at the boundary due to the sound field of the desired source multiplied by a window function:

$$D(\mathbf{x}_0) = a(\mathbf{x}_0) \cdot P(\mathbf{x}_0). \quad (4)$$

B. The Time-Reversal Cavity

So far, we have reviewed the fundamentals of the wave field synthesis technique. The basic task for far-field beamforming

techniques is rather the extraction of plane waves from the captured sound field using a spatial filter for the sound field. It can easily be verified that the optimal spatial filter for extracting a plane wave is a plane sink which is coincident to the front of the desired plane wave. A sound source can be converted into a sink using the time-reversal principle. Since only second order derivatives with respect to space and time occur in the lossless wave equation, its solutions are invariant to a time-reversal. The physical background of the time-reversal cavity and its properties have been discussed in [11]. It can be shown that converting a source into a sink in the frequency domain can be done by conjugation. Hence, the well studied theory of wave field synthesis can be applied for the synthesis of sinks as we will show later on.

C. Approach I

A direct approach for deriving the coefficients of the desired beamformer is to use the array to directly synthesize a plane sink (a virtual plane wave annihilating the plane wave in look direction using time-reversal). The desired sink can be obtained using the conjugate of the driving functions for the synthesis of a plane wave from θ_0 using a linear distribution of dipoles. Furthermore, we introduce a normalization to ensure obtaining the unit impulse if the captured sound field is a plane wave in the look direction with an amplitude equal to 1

$$\check{\mathbf{D}}_{\theta_0}(\mathbf{x}_0) := \frac{\mathbf{D}_{\theta_0}^H(\mathbf{x}_0)}{\|\mathbf{D}_{\theta_0}(\mathbf{x}_0)\|_2}, \quad (5)$$

where $\mathbf{D}_{\theta_0}(\mathbf{x}_0)$ is the discretized version $D_{\theta_0}(\mathbf{x}_0)$, the driving function for synthesizing a plane wave with the angle of incidence θ_0 , and $\check{\mathbf{D}}_{\theta_0}(\mathbf{x}_0)$ contains the beamformer coefficients. Note that the obtained spatial filter in Eq. (5) coincides with the definition of the pseudoinverse of $\mathbf{D}_{\theta_0}^H(\mathbf{x}_0)$. Hence, we can write

$$\check{\mathbf{D}}_{\theta_0}(\mathbf{x}_0) = \mathbf{D}_{\theta_0}^\dagger(\mathbf{x}_0). \quad (6)$$

Extraction of a particular plane wave from the sound field $\mathbf{P} = [P_1, \dots, P_N]^T$ captured by an N -elements array is performed by the spatial filtering $\check{\mathbf{D}}_{\theta_0}(\mathbf{x}_0) \cdot \mathbf{P}$.

D. Approach II: Synthesis of linear arrays of focused sinks

As expected, the simulations given in Sect. III of the presented approach I show a limited directivity performance especially at low frequencies using arrays with small dimensions relative to the wavelength. In the following we outline a second approach to cope with the resulting limited directivity at low frequencies. For clarity, we illustrate our approach for capturing 2-dimensional sound fields. However, the extension to the 3-dimensional case is straightforward. The synthesis of a sink for extracting a desired plane wave can be done based on the synthesis of multiple point sinks on a line. The synthesis of point sink is equivalent to a near-field beamformer [12]. Here, we use the outlined wave field synthesis and the time reversal principle to synthesize a virtual sink at \mathbf{x}_s on a line which is perpendicular to the desired look direction, see Fig. 1. The distance between the virtual array and the physically real array

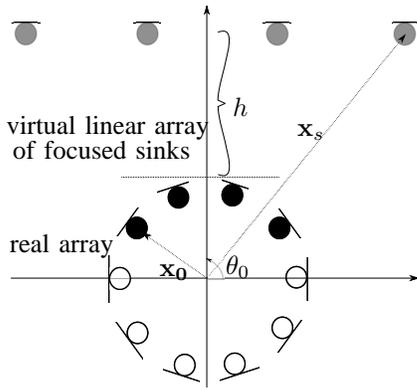


Fig. 1. Illustration for the presented approach II.

h has to be chosen in a way such that $\frac{\omega}{c} \|\mathbf{x}_s - \mathbf{x}_0\|_2 \gg 1$. To obtain the weights for synthesizing a point sink we again use the pseudoinverse of the driving function for the synthesis of a point source by a dipole distribution. Note that the number of the independent virtual sinks that can be synthesized depends on the array dimensions and order.

For an improved sidelobe cancellation, window functions $w(\mathbf{x}_s)$ can be applied on the virtual array, such as Chebychev or Hann window functions. Finally, the coefficients of the desired beamformer are given as

$$\check{\mathbf{D}}_{\theta_0}(\mathbf{x}_0) = \left\{ \sum_{s=1}^{s=S} w(\mathbf{x}_s) \mathbf{D}_s(\mathbf{x}_0) \right\}^\dagger. \quad (7)$$

In general, flexible placement of the virtual sinks, e.g., logarithmic arrays can be simply emulated, although this is not in the scope of this paper.

III. SIMULATION RESULTS

To prove our concept we show the performance of the presented approaches based on simulating a compact rigid cylindrical microphone array with 25 omnidirectional pressure microphones mounted on the surface of the cylinder at $z = 0$. The array has a diameter of 15 cm. In order to simulate approach II we created a virtual line array of 10 m total length and a spacing of 0.5 m, the distance between the line array and the real array h was set to 1.5 m. The two most important measures to evaluate the performance of a beamformer are the directivity and the white noise gain.

A. Directivity Gain

The directivity gain is defined as

$$\text{DIR}(\theta_0, \omega) = 10 \log_{10} \left(\frac{|\check{\mathbf{D}}_{\theta_0} \mathbf{P}|^2}{\check{\mathbf{D}}_{\theta_0} \check{\mathbf{D}}_{\theta_0}^H} \right), \quad (8)$$

where \mathbf{P} is a diffuse sound field simulated as an ensemble of dense sampled and equally weighted plane waves. The scattering of the plane waves on the surface of the sound hard array surface was implemented according the results in [5]. As the simulation shows, see Fig. 2, an overall satisfying directivity gain can be achieved using the direct approach I. Unfortunately, at low frequencies only limited directivity can be achieved as shown in Fig.3. Figure 5 depicts the achieved

directivity gain by the approaches I and II using the same simulation scenario as in Fig. 2 but at the single temporal frequency $f = 300\text{Hz}$. The blue curve shows the delay-and-sum directivity gain, the red and green curves show the gain achieved by applying the approach II. The

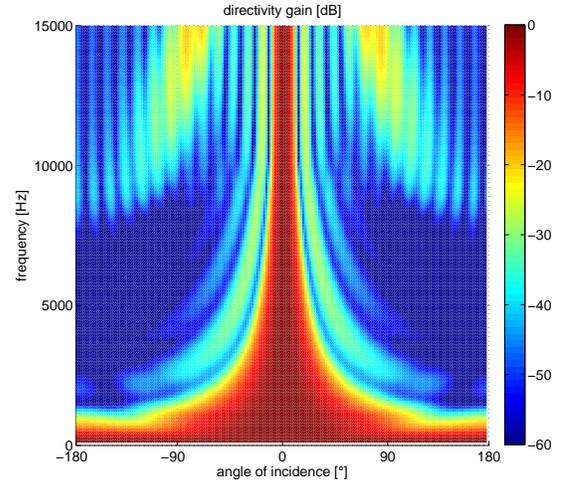


Fig. 2. Directivity of 25 elements rigid cylindrical microphone array using the delay-and-sum approach and the proposed selection criterion.

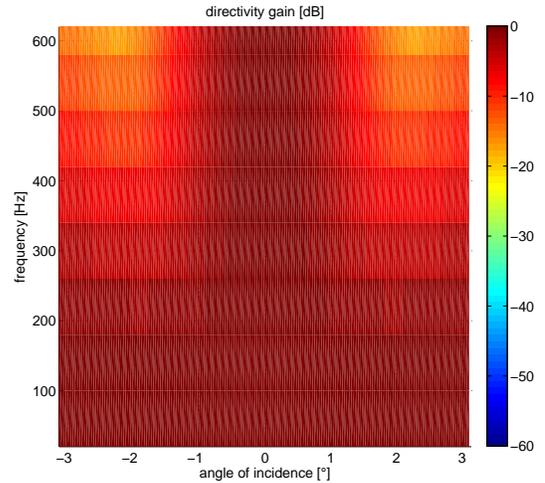


Fig. 3. Achieved directivity gain over the frequency band [20-600] Hz with a 25 elements array of a radius of .075 m using the proposed approach I.

directivity in the frequency range [20-600] Hz using approach II and a Hann windowing function is given in Fig. 4.

B. White Noise Gain

A common performance measure for the robustness of a beamformer is the white noise gain. It quantifies the attenuation of diffuse white noise from the sensors to the array output. It is defined as

$$\text{WNG}(\theta_0, \omega) = 10 \log_{10} \left(\frac{|\check{\mathbf{D}}_{\theta_0} \mathbf{P}|^2}{\check{\mathbf{D}}_{\theta_0} \check{\mathbf{D}}_{\theta_0}^H} \right), \quad (9)$$

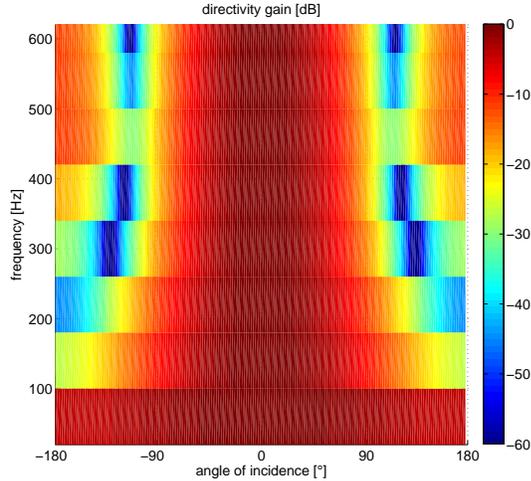


Fig. 4. Achieved directivity gain over the frequency band [20-600] Hz with a 25 elements array of a radius of .075 m using the proposed approach of focused sinks.

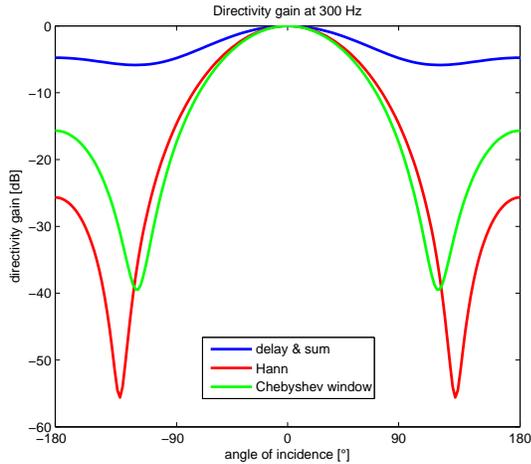


Fig. 5. Achieved directivity gain at 300 Hz with a 25 elements array of a radius of .075 m using the proposed approach of focused sinks.

where \vec{D}_{θ_0} is the row vector of complex weights for each microphone, and \mathbf{P} is the column vector of the complex pressure caused by a plane wave from the incidence direction θ_0 . Note that a positive WNG value indicates an improvement of the signal-to-noise ratio, while a negative one means degradation. Figure 6 depicts the achievable WNG using the presented approaches and a phase-mode beamformer of order 4 as presented in [5].

IV. CONCLUSION

In this paper we presented two approaches for deriving a delay-and-sum beamformer. Both approaches offer high flexibility and low computational complexity. As we innervated with simulations, both presented approaches achieve a high white-noise-gain especially at low frequencies compared to traditional phase-mode beamformers. The first approach which

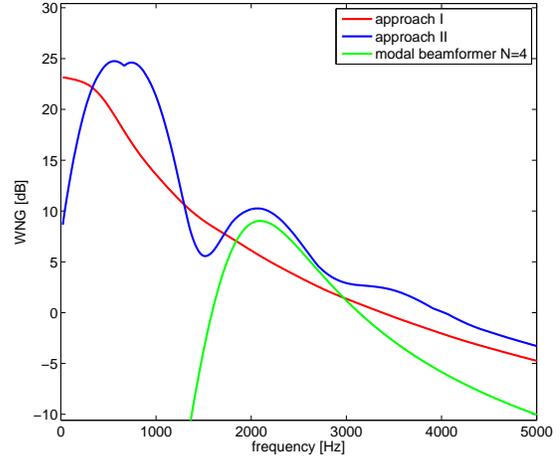


Fig. 6. Comparison of the achieved white-noise-gain over the frequency using the presented approach of virtual sinks with Hann windowing function, the presented Delay-and-Sum approach, and the traditional modal beamformer of order 4 and a 25-elements rigid cylindrical array with a radius of 7.5 cm.

can be seen as a direct delay-and sum beamformer suffers from limited achievable directivity at low frequencies using arrays with small dimensions. To cope with this limitation we presented an alternative approach which stands to benefit from creating large virtual linear arrays such that high directivity can be obtained.

REFERENCES

- [1] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*. Academic Press, 1999.
- [2] B. Rafaely, "Analysis and design of spherical microphone arrays," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 1, pp. 135–143, 2005.
- [3] J. Meyer and G. Elko, "Spherical microphone arrays for 3D sound recording," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, 2004, pp. 67–89.
- [4] Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 702–714, 2007.
- [5] H. Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*. Springer Verlag, 2007.
- [6] F. Zotter, "Analysis and synthesis of Sound-Radiation with spherical arrays," Doctoral Thesis, University of Music and Performing Arts Graz, Graz, Austria, 2009.
- [7] T. Abhayapala and A. Gupta, "Spherical harmonic analysis of wavefields using multiple circular sensor arrays," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1655–1666, Aug. 2010.
- [8] B. Rafaely, "Phase-mode versus delay-and-sum spherical microphone array processing," *Signal Processing Letters, IEEE*, vol. 12, no. 10, pp. 713–716, 2005.
- [9] S. Spors, "Extension of an analytic secondary source selection criterion for wave field synthesis," in *123th AES Convention, New York, USA*, oct. 2007.
- [10] A. Berkhout and C. Wapenaar, "One-way versions of the Kirchhoff integral," *Geophysics*, vol. 54, no. 4, p. 460467, 1989.
- [11] D. Cassereau and F. M., "Time-reversal of ultrasonic fields part iii: Theory of the closed time-reversal cavity," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 39, no. 5, pp. 579–592, Sept.
- [12] T. Abhayapala, "Modal analysis and synthesis of broadband nearfield beamforming arrays," Ph.D. dissertation, The Australian National University, 1999.