Applied Acoustics 110 (2016) 194–206

Contents lists available at ScienceDirect

Applied Acoustics

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

Design of robust differential microphone arrays with the Jacobi–Anger expansion ☆

Liheng Zhao [a], Jacob Benesty [a], Jingdong Chen [b,*]

[a] INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, QC H5A 1K6, Canada

[b] Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, 127 Youyi West Rd, Xi'an, Shaanxi 710072, China

ARTICLE INFO

Article history:
Received 2 August 2015
Received in revised form 14 January 2016
Accepted 20 March 2016

Keywords:
Differential microphone arrays (DMAs)
Robust DMAs
Beampattern
Beamforming
White noise gain
Directivity factor
First-order DMA
Second-order DMA
Third-order DMA
Jacobi–Anger expansion

ABSTRACT

Due to their small size, differential microphone arrays (DMAs) are very attractive. Moreover, they have been effective in combating noise and reverberation. Recently, a new class of DMAs of different orders have been developed with the MacLaurin's series and the frequency-independent patterns. However, the MacLaurin's series does not approximate well the exponential function, which appears in the general definition of the beampattern, when the intersensor spacing is not small enough. To circumvent this problem, we propose in this paper to approximate the exponential function with the Jacobi–Anger expansion. Based on this approximation and the frequency-independent Chebyshev patterns, we derive first-, second-, and third-order DMAs. Furthermore, in order to improve the robustness of DMAs against white noise amplification, we propose to use more microphones combined with minimum-norm filters. It is also shown that the Jacobi–Anger expansion is optimal from a mean-squared error perspective. Simulations are carried out to evaluate the performance of the proposed DMAs.

1. Introduction

It is well known that noise and reverberation are detrimental to the speech quality and intelligibility. As a consequence, the performance of many applications, such as hands-free telecommunication and hearing aids, can be severely degraded. Over the past decades, approaches based on microphone arrays and beamforming techniques have been widely studied in the difficult context of noisy and reverberant environments [1–4]. Recently, methods based on differential microphone arrays (DMAs) have received a great deal of attention due to their small size and potential of high directivity factors [5–8]. As early as in the 1940s, DMAs of different orders were constructed and their anti-noise characteristics were analyzed [9,10]. Since then, a good amount of progress has been made. In [11,12], adaptive DMAs were developed to suppress spatially non-stationary noise. In [13], an approach based on sensor calibration was designed to increase DMAs' robustness against sensor mismatch, which may seriously damage their performance. In [14], DMAs were used to estimate the noise power spectral density (PSD), and the spectral subtraction algorithm was then applied to suppress noise. In [15,16], approaches for the design of higher-order DMAs were developed. In [6,7], DMAs were systematically studied from a signal processing perspective. Specifically, the design, implementation, and performance analysis of DMAs were presented.

In [6,8], the exponential function, which appears in the general definition of the beampattern, was approximated with the MacLaurin's series; this lead to the design of DMAs of different orders. It has been reported that DMAs based on the MacLaurin's series are capable of achieving high directivity factors. However, it has been observed that when the intersensor spacing is not very small, the MacLaurin's series is no longer a good approximation of the exponential function. As a result, the performance of DMAs is affected. To avoid this problem, we propose in this paper to use the Jacobi–Anger expansion to approximate the exponential function. We first derive the traditional[1] first-, second-, and third-order DMAs. Many simulation results show that the traditional DMAs with the Jacobi–Anger expansion significantly improve the directivity

☆ This work was supported in part by the NSFC "Distinguished Young Scientists Fund" under Grant No. 61425005.

* Corresponding author.
E-mail address: jingdongchen@ieee.org (J. Chen).

[1] By traditional, we mean that the order of the DMA is equal to $M-1$, where $M \geqslant 2$ is the number of microphones.

http://dx.doi.org/10.1016/j.apacoust.2016.03.015
0003-682X/

factor, but have the problem of white noise amplification, like any other approaches. To deal with this serious side effect, we derive robust DMAs by using more microphones combined with minimum-norm filters. It is shown that the robust DMAs with the Jacobi–Anger expansion improve the white noise gain considerably and, therefore, are more robust against any imperfections in the system. In comparison with DMAs based on the MacLaurin's series, DMAs based on the Jacobi–Anger expansion perform better by giving higher directivity factors and white noise gains, confirming that the latter approximation is preferable in the derivation of DMAs. It is also shown that the Jacobi–Anger expansion is optimal from an MSE perspective.

The rest of this paper is organized as follows. In Section 2, some basic concepts of DMAs are introduced. In Section 3, frequency-independent patterns and the approximation based on the Jacobi–Anger expansion are presented. The traditional and robust first-, second-, and third-order DMAs are derived in Sections 4–6, respectively. Simulations are carried out to evaluate the performance of DMAs in Section 7, followed by our conclusions in Section 8.

2. Signal model, problem formulation, and definitions

We consider a source signal (plane wave), in the farfield, that propagates in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s, and impinges on a uniform linear sensor array consisting of M omnidirectional microphones, where the distance between two successive sensors is equal to δ (see Fig. 1). The direction of the source signal to the array is parameterized by the azimuth angle θ. In this scenario, the steering vector (of length M) is given by

$$\mathbf{d}(\omega, \theta) = \begin{bmatrix} 1 & e^{-\jmath\omega\tau_0\cos\theta} & \cdots & e^{-\jmath(M-1)\omega\tau_0\cos\theta} \end{bmatrix}^T, \tag{1}$$

where the superscript ^{T} is the transpose operator, $\jmath = \sqrt{-1}$ is the imaginary unit, $\omega = 2\pi f$ is the angular frequency, $f > 0$ is the temporal frequency, and $\tau_0 = \delta/c$ is the delay between two successive sensors at the angle $\theta = 0$. The acoustic wavelength is $\lambda = c/f$.

In order to avoid spatial aliasing [3], which has the negative effect of creating grating lobes (i.e., copies of the main lobe, which usually points toward the desired signal), it is necessary that the inter-element spacing is less than $\lambda/2$, i.e.,

$$\omega\tau_0 < \pi. \tag{2}$$

The condition (2) easily holds for small values of δ and at low frequencies but not at high frequencies.

We consider fixed beamformers, such as DMAs [6,7,13,15,16], where the main lobe is at the angle $\theta = 0$ (endfire direction) and the desired signal propagates from the same angle. Our focus is on the design of different orders DMAs that are robust to white noise amplification. For that, a complex weight, $H_m^*(\omega), m = 1, 2, \ldots, M$, is applied at the output of each microphone, where the superscript * denotes complex conjugation. The weighted outputs are then summed together to form the beamformer output as shown in Fig. 1. Putting all the gains together in a vector of length M, we get

$$\mathbf{h}(\omega) = \begin{bmatrix} H_1(\omega) & H_2(\omega) & \cdots & H_M(\omega) \end{bmatrix}^T. \tag{3}$$

Then, the objective is to design such a filter so that the array obeys a given DMA pattern.

The vector containing the microphone signals can be expressed as

$$\mathbf{y}(\omega) = \begin{bmatrix} Y_1(\omega) & Y_2(\omega) & \cdots & Y_M(\omega) \end{bmatrix}^T = \mathbf{d}(\omega, 0)X(\omega) + \mathbf{v}(\omega), \tag{4}$$

where $\mathbf{d}(\omega, 0)$ is the steering vector at $\theta = 0$ (direction of the source), $X(\omega)$ is the desired signal, and

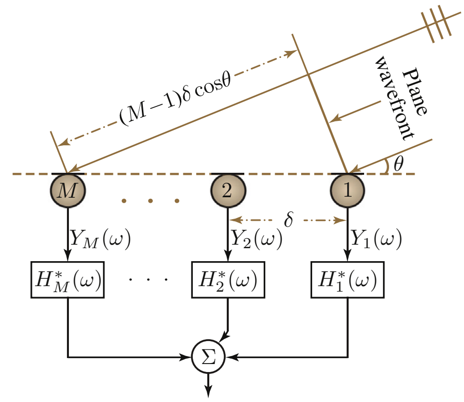

Fig. 1. A uniform linear microphone array with processing.

$$\mathbf{v}(\omega) = \begin{bmatrix} V_1(\omega) & V_2(\omega) & \cdots & V_M(\omega) \end{bmatrix}^T \tag{5}$$

is the additive noise signal vector.

The beamformer output is simply [4]

$$Z(\omega) = \mathbf{h}^H(\omega)\mathbf{y}(\omega) = \mathbf{h}^H(\omega)\mathbf{d}(\omega, 0)X(\omega) + \mathbf{h}^H(\omega)\mathbf{v}(\omega), \tag{6}$$

where $Z(\omega)$ is the estimate of the desired signal, $X(\omega)$, and the superscript ^{H} is the conjugate-transpose operator.

If we take microphone 1 as the reference, we can define the input signal-to-noise ratio (SNR) with respect to this reference as

$$\text{iSNR}(\omega) = \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)}, \tag{7}$$

where $\phi_X(\omega) = E\left[|X(\omega)|^2\right]$ and $\phi_{V_1}(\omega) = E\left[|V_1(\omega)|^2\right]$ are the variances of $X(\omega)$ and $V_1(\omega)$, respectively, with $E[\cdot]$ denoting mathematical expectation.

The output SNR is obtained from the variance of $Z(\omega)$:

$$\begin{aligned} \text{oSNR}[\mathbf{h}(\omega)] &= \phi_X(\omega)\frac{\left|\mathbf{h}^H(\omega)\mathbf{d}(\omega, 0)\right|^2}{\mathbf{h}^H(\omega)\mathbf{\Phi}_\mathbf{v}(\omega)\mathbf{h}(\omega)} \\ &= \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)} \times \frac{\left|\mathbf{h}^H(\omega)\mathbf{d}(\omega, 0)\right|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_\mathbf{v}(\omega)\mathbf{h}(\omega)}, \end{aligned} \tag{8}$$

where $\mathbf{\Phi}_\mathbf{v}(\omega) = E\left[\mathbf{v}(\omega)\mathbf{v}^H(\omega)\right]$ and $\mathbf{\Gamma}_\mathbf{v}(\omega) = \frac{\mathbf{\Phi}_\mathbf{v}(\omega)}{\phi_{V_1}(\omega)}$ are the correlation and pseudo-coherence matrices of $\mathbf{v}(\omega)$, respectively.

The definition of the gain in SNR is easily derived from the previous definitions, i.e.,

$$\mathcal{G}[\mathbf{h}(\omega)] = \frac{\text{oSNR}[\mathbf{h}(\omega)]}{\text{iSNR}(\omega)} = \frac{\left|\mathbf{h}^H(\omega)\mathbf{d}(\omega, 0)\right|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_\mathbf{v}(\omega)\mathbf{h}(\omega)}. \tag{9}$$

We are interested in two types of noise.

- The temporally and spatially white noise with the same variance at all microphones.[2] In this case, $\mathbf{\Gamma}_\mathbf{v}(\omega) = \mathbf{I}_M$, where $\mathbf{I}_M$ is the $M \times M$ identity matrix. Therefore, the white noise gain (WNG) is

$$\mathcal{G}_{\text{wn}}[\mathbf{h}(\omega)] = \frac{\left|\mathbf{h}^H(\omega)\mathbf{d}(\omega, 0)\right|^2}{\mathbf{h}^H(\omega)\mathbf{h}(\omega)}. \tag{10}$$

The delay-and-sum (DS) beamformer:

[2] This noise models the sensor noise.