SPRINGER BRIEFS IN
ELECTRICAL AND COMPUTER ENGINEERING

Jacob Benesty
Jingdong Chen
Chao Pan

Fundamentals of Differential Beamforming

Springer

SpringerBriefs in Electrical and Computer Engineering

More information about this series at http://www.springer.com/series/10059

Jacob Benesty · Jingdong Chen
Chao Pan

Fundamentals of Differential Beamforming

Springer

Jacob Benesty
INRS-EMT, University of Quebec
Montreal, QC
Canada

Chao Pan
Northwestern Polytechnical University
Xi'an
China

Jingdong Chen
Northwestern Polytechnical University
Xi'an
China

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISBN 978-981-10-1045-3 ISBN 978-981-10-1046-0 (eBook)
DOI 10.1007/978-981-10-1046-0

Library of Congress Control Number: 2016937935

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Contents

Abstract

Microphone arrays can be used in a broad range of applications from telecommunications, teleconferencing, and smart home systems to intelligent human–machine interfaces to deal with many important acoustic problems such as noise reduction, source separation, dereverberation, source localization/tracking, robust hands-free speech recognition, to name a few. Significant efforts have been devoted to the design of such arrays and the associated processing algorithms since the 1970s. In the literature, microphone arrays are generally classified into two major categories: additive and differential. The former refers to arrays with large sensor spacing whose outputs are responsive to the acoustic pressure field; whereas the latter refers to arrays with small sensor spacing whose outputs are responsive to the differential acoustic pressure field of different orders. While both types of arrays have their own pros and cons, when applied to solving a real problem, differential microphone arrays (DMAs) are more appropriate for high-fidelity signal enhancement applications as they have the potential to form frequency-invariant directivity patterns and attain large directional gains with small and compact apertures.

This book is intended to provide a systematic study of the fundamental theory and the methods of beamforming with DMAs, or differential beamforming in short. From a physical perspective, a DMA of some order is defined as an array that measures the differential acoustic pressure field of that order; such an array has a beampattern in the form of a polynomial whose degree is equal to the DMA order. Therefore, the fundamental and core problem of differential beamforming boils down to the design of beampatterns with orthogonal polynomials. But constraints have to be considered so that the resulting beamformer does not seriously amplify sensors' self-noise and mismatches among sensors. In this work, we first present a brief overview of differential beamforming and some popularly used DMA beampatterns such as dipole, cardioid, hypercardioid, and supercardioid. Then, some background knowledge on orthogonal functions and orthogonal polynomials is provided, which forms the basis of differential beamforming. Several performance criteria are subsequently revisited, which can be used to evaluate the performance of the derived differential beamformers. Next, differential beamforming is cast into a framework of optimization and linear system solving and it is shown how different beampatterns can be designed with this optimization framework. After that, several approaches are presented to the design of differential beamformers with the maximum DMA order, with the control of the white noise gain, and with the control of both the frequency invariance of the beampattern and the white noise gain. Finally, a joint optimization method is explained, which can be used to derive differential beamformers that have almost frequency-invariant beampatterns and meanwhile are robust to sensors' self-noise.

Chapter 1
Introduction

In this chapter, we briefly discuss microphone array beamforming in general and differential beamforming in particular. We also explain why this latter approach is fundamental and should be considered in most speech and audio acquisition systems.

1.1 Introduction

Sound signal acquisition has been an essential part of speech processing since the invention of telephone systems in the late 19th century. Most early sound acquisition systems use only a single microphone; but such systems were not found to be very good, to say the least, in challenging acoustic environments where there are noise, echo, reverberation, and interferences. For a better control of the mentioned problems and preservation of the spatial sound realism, multiple-microphone systems were then invented which make them much more powerful than single-microphone systems in terms of sound acquisition quality, system functionality, and flexibility in developing the associated processing algorithms.

Depending on how sensors are arranged in space, multiple-microphone systems have two basic forms, i.e., organized and ad hoc arrays. In an organized array, which is generally referred to as a microphone array, the sensors are arranged to form a particular geometry such as a line, circle, or sphere in which all the sensors' positions with respect to a reference point are known and such information can be used in subsequent processors. These sensors are generally required to have the same characteristics (e.g., sensitivity, dynamic range, gain, noise floor, etc.) and their outputs are converted to digital signals with a synchronized multichannel A/D system. By processing the outputs of the sensors and combining the results together, many functionalities can be implemented such as direction-of-arrival (DOA) estimation, source localization, noise reduction, signal enhancement, source separation, just to name a few. In comparison, in an ad hoc array, sensors are arbitrarily placed in different positions to form a sensor network without a fixed geometry. The sensors are generally not required to have the same characteristics and their outputs may be sampled

J. Benesty et al., *Fundamentals of Differential Beamforming*, SpringerBriefs
in Electrical and Computer Engineering, DOI 10.1007/978-981-10-1046-0_1

with separate A/D systems that may use different clocks. Both organized and ad hoc arrays have their own pros and cons. Generally, ad hoc arrays are logistically easy to install since they do not have much constraint on the sensors' quality and sensors' positions. They are preferable to be used in applications where a large acquisition area needs to be covered as the number of sensors and the size of such arrays can be large. But the associated signal processing and fusion algorithms can be complicated and usually are *ad hoc* as the reliability and the amount of information from different sensors may vary significantly. Clock skew is another important issue that adds difficulty in processing and fusing the sensors' outputs. At last but not least, it is still not clear whether beamforming makes sense in such scenario. In comparison, the design of a microphone (organized) array usually takes much professional experience as the sensors are required to be uniform in response (if not but difference is not dramatic, compensation can help). The selection of a proper geometry is also very important, which depends not only on the performance expectation but also on the application constraints. The advantage with microphone arrays is that the associated theory and methods of signal processing are more rigorous to develop. The performance is also more consistent over different environments. As a result, significant efforts have been devoted to microphone arrays in the literature than to the ad hoc ones. This book also focuses on microphone arrays.

A microphone array system consists of two important components: hardware and associated processing algorithms. The design of the former involves the selection of sensors, the array geometry, as well as the design of pre-amplifiers and multichannel A/D convertors. While it is important, this part can be done by experienced audio engineers; so we choose not to devote much effort to it in this book. For the latter, a large variety of processing algorithms have been studied in the literature either to estimate some important parameters or enhance certain signals or signal components from the microphones' outputs, e.g., beamforming, channel identification, channel equalization, multichannel noise reduction, and blind source separation. Many more methods are still emerging, which is more than one book can cover in detail. In this work, we focus on one major topic that we deem to be very important and useful: beamforming. Briefly, beamforming consists of designing a good, in some sense, spatial filter that can take advantage of the spatiotemporal information embedded in the microphone array outputs to form a response with different sensitivities to sounds arriving from different directions. It can be used in many applications to deal with sound signal acquisition and enhancement, including but not limited to

- teleconferencing,
- multi-party telecollaboration,
- hands-free speech communication,
- distance speech recognition,
- robotics,
- gaming,
- virtual reality,
- high-fidelity audio recording,
- acoustic surveillance (security and monitoring),

- acoustic scene analysis,
- smart television and smart home system,
- hearing aids.

The number of applications is still growing. While the potential of beamforming is huge, solutions are still far from being satisfactory and much more efforts in this area of research are indispensable.

1.2 Microphone Array Beamforming: A Brief Overview

Research in microphone array beamforming started in the late 1960s although some of the fundamental principles can be traced back to the 1930s when directional microphones were invented [1, 2]. Early works in this area were strongly influenced by the sensor array theory developed in the field of radar and sonar. The most popular microphone array structure investigated in the literature is the uniform linear array (ULA) combined with the delay-and-sum (DS) beamformer. The basic idea underlying this algorithm is to delay each microphone output in the array by a proper amount of time so that the signal components from the desired look direction are synchronized across all sensors. These delayed signals are then weighted and summed together. This beamformer has been intensively studied for enhancing broadband signals of interest from their noisy observations. However, the biggest issue with this beamformer is that its beampattern, which is defined and discussed in Chaps. 2 and 4, varies greatly with frequency. When applied to processing speech and audio signals where frequencies can range from 60 Hz to 20 kHz, the DS beamformer suffers from a number of well-known problems and drawbacks. They are as follows.

- Its beampattern varies with frequency. As a result, noise is not uniformly attenuated over its entire spectrum, resulting in some disturbing artifacts in the array output [3].
- Its directivity factor, which will be discussed in Chap. 4, is small at low frequencies and, as a consequence, it is not effective in suppressing noise and interference at low frequencies [4].
- Its beamwidth is inversely proportional to the frequency. If the incident angle of the acoustic source is different from the array's look direction even though it is still within the range of the mainlobe, the signal can be significantly distorted.

To overcome the drawbacks of the DS beamformer, the so-called broadband beamforming techniques have been investigated. One way to obtain a broadband beamformer is to use harmonically nested subarrays where every subarray is designed for operating at a small frequency range [5–7]. An example of a nested array is given in [5], which is illustrated in Fig. 1.1 where four subarrays are used in the frequency range between 0.5 and 8 kHz. The DS beamformer is applied to each subarray. By controlling the spacing and the number of sensors in every subarray, the overall nested array can obtain a similar beamwidth across the frequency range of interest, for example, from 0.5 to 8 kHz in Fig. 1.1. But such a solution requires a large array

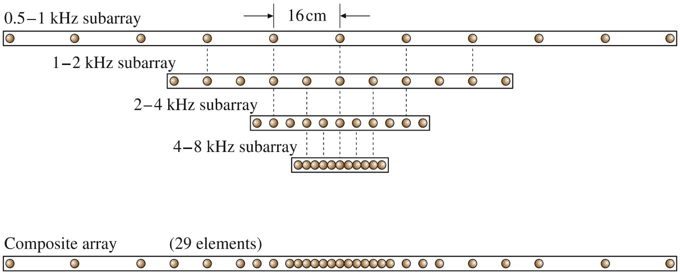

Fig. 1.1 Illustration of broadband beamforming using a nested array

size (aperture) with many microphones, even though different subarrays may share sensors in the array. This large size makes this kind of arrays impractical in real-world applications.

Another way to deal with broadband beamforming is through narrowband decomposition of a broadband signal and then design a narrowband beamformer at each subband as shown in Fig. 1.2 [3, 8, 9]. The beamwidth at each subband is controlled in such a way that all the beamformers at different frequencies have a similar beamwidth. Though it can make similar beamwidth across a wide range of frequencies, this way of broadband beamforming sacrifices the array performance at high frequencies.

The subband structure of broadband beamforming can be equivalently transformed into a time-domain framework as shown in Fig. 1.3, where a finite-impulse-response (FIR) filter is applied to each sensor output, and the filtered sensor signals are then added up together to form a single output. This is widely known as the filter-and-sum beamformer originally developed by Frost [10] and, hence, it is also called the Frost beamformer. The core problem in this structure is then to determine

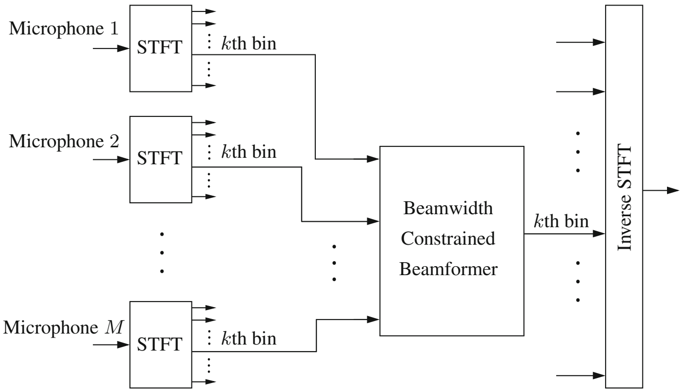

Fig. 1.2 Broadband beamforming using narrowband decomposition, where STFT stands for short-time Fourier transform