Signal Processing 107 (2015) 4–20

Contents lists available at ScienceDirect

Signal Processing

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

Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks ☆

Shmulik Markovich-Golan [a], Alexander Bertrand [b], Marc Moonen [b], Sharon Gannot [a]

[a] Bar-Ilan University, Faculty of Engineering, 5290002 Ramat-Gan, Israel
[b] KU Leuven, Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form
17 July 2014
Accepted 21 July 2014
Available online 1 August 2014

Keywords:
Wireless acoustic sensor networks
Distributed speech enhancement
Minimum variance beamforming

ABSTRACT

In multiple speaker scenarios, the linearly constrained minimum variance (LCMV) beamformer is a popular microphone array-based speech enhancement technique, as it allows minimizing the noise power while maintaining a set of desired responses towards different speakers. Here, we address the algorithmic challenges arising when applying the LCMV beamformer in wireless acoustic sensor networks (WASNs), which are a next-generation technology for audio acquisition and processing. We review three optimal distributed LCMV-based algorithms, which compute a network-wide LCMV beamformer output at each node without centralizing the microphone signals. Optimality here refers to equivalence to a centralized realization where a single processor has access to all signals. We derive and motivate the algorithms in an accessible top-down framework that reveals their underlying relations. We explain how their differences result from their different design criterion (node-specific versus common constraints sets), and their different priorities for communication bandwidth, computational power, and adaptivity. Furthermore, although originally proposed for a fully connected WASN, we also explain how to extend the reviewed algorithms to the case of a partially connected WASN, which is assumed to be pruned to a tree topology. Finally, we discuss the advantages and disadvantages of the various algorithms

☆ The work of A. Bertrand was supported by a Postdoctoral Fellowship of the Research Foundation – Flanders (FWO). This work was carried out in the frame of KU Leuven Research Council CoE PFV/10/002 (OPTEC), Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P7/23 (BESTCOM, 2012-2017), Research Project FWO nr. G.0763.12 'Wireless acoustic sensor networks for extended auditory communication', Research Project FWO nr. G.0931.14 'Design of distributed signal processing algorithms and scalable hardware platforms for energy-vs-performance adaptive wireless acoustic sensor networks', and project HANDiCAMS. The project HANDiCAMS acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open Grant number: 323944. The scientific responsibility is assumed by its authors.

E-mail addresses: shmulik.markovich@gmail.com (S. Markovich-Golan), alexander.bertrand@esat.kuleuven.be (A. Bertrand), marc.moonen@esat.kuleuven.be (M. Moonen), sharon.gannot@biu.ac.il (S. Gannot).

http://dx.doi.org/10.1016/j.sigpro.2014.07.014
0165-1684/

1. Introduction

A general problem of interest in the field of speech processing is to extract a set of desired speech signals from microphone recordings that are contaminated by interfering speakers or other noise sources in a reverberant enclosure. By exploiting the spatial properties of the speech and noise signals, array-processing techniques can significantly outperform single-channel techniques in terms of improved interference suppression and reduced speech distortion, especially in scenarios with non-stationary noise sources (such as interfering speakers).

A family of array-processing techniques, known as *beamforming*, typically performs a linear filter-and-sum operation on the microphone signals, where the filters are optimized according to certain design criteria [1–3]. In classical speech beamformer (BF) setups, a microphone array is placed somewhere within the enclosure, preferably close to the desired speakers (as in mobile phone or personal computer applications [4]). In this case, the received signal-to-noise ratio (SNR) and direct-to-reverberant ratio (DRR) are often sufficiently large, enabling the BF to obtain adequate performance. However, in applications where the desired sources are far away from the array, or if the array contains too few microphones to obtain the required speech enhancement performance, it may be useful to add additional microphone arrays at other places within the enclosure to collect more data over a wider area.

Recent technological advances in the design of miniature and low-power electronic devices enable the deployment of so-called *wireless sensor networks* (*WSNs*) [5–7]. A WSN consists of autonomous self-powered devices or nodes, which are equipped with sensing, processing, and communicating facilities. The WSN concept is quite versatile and has applications in environmental monitoring, biomedicine, security and surveillance. In this paper we consider WSNs designed for *acoustic* signal processing tasks, often referred to as wireless acoustic sensor networks (WASNs) [8], where each node is equipped with one or more microphones. A WASN allows to deploy a large number of microphone arrays at various positions, and can be exploited in hearing aids [9–11], (hands-free) speech communication systems [12–14], acoustic monitoring [15–20], ambient intelligence [21], etc.

Alongside their numerous advantages, AWASNs introduce several challenges, in particular related to the limited per-node energy resources, since the finite battery life constrains the communication and computational energy usage at each node. These energy limitations, combined with the fact that each node has access only to partial data, require special attention when developing WASN algorithms. These algorithms can be either *distributed*, to reduce the wireless data transfer and to share the processing burden between multiple nodes, or *centralized*, where all the data is transferred to a so-called fusion center (FC) for further processing. A distributed approach is typically preferred in terms of energy consumption and scalability (or in absence of a powerful FC), although the algorithm design is much more challenging, especially when pursuing a similar performance as in a centralized procedure.

Distributed BF or speech enhancement algorithms typically rely on compression techniques to minimize the data that is exchanged between the nodes. However, applying straightforward signal compression methods on the microphone signals (at each node independently) usually results in a suboptimal BF performance. Moreover, common speech or audio compression methods introduce distortion that may destroy important spatial information, and render the beamforming process useless.

Several distributed BFs or speech enhancement algorithms have been proposed in the literature, ranging from heuristic or suboptimal methods [12,22–24] to algorithms for which optimality can be proven [9–11,25–28]. In this context, 'optimality' refers to the fact that the algorithm obtains the same BF outputs as its centralized counterpart algorithm, i.e., as if each node would have access to the full set of microphone signals. In this paper, we confine ourselves to the review of *optimal* distributed minimum-variance BF algorithms where nodes share (compressed) signals and parameters, and where the general aim is to achieve the same speech enhancement performance as obtained with a centralized minimum-variance BF. We mainly focus on the BF algorithm design challenges, and we disregard several other (but equally important) challenges, such as synchronization [29–32], node subset selection [33,34], topology selection, distortion due to audio compression [22,35,36], packet loss, input-output delay management [37], etc.

We review three state-of-the-art distributed minimum-variance BF algorithms, namely the distributed LCMV (D-LCMV) BF [26], the linearly constrained distributed adaptive node-specific signal estimation (LC-DANSE) algorithm [38], and the distributed generalized sidelobe canceler (DGSC). Although these algorithms were originally proposed independently from each other, they are implicitly related as they are based on a similar LCMV optimization criterion. However, despite this common underlying BF design criterion, the actual relation between the algorithms is not immediately apparent from the original publications [26,27,38], as they start from different problem statements and algorithm design principles. For example, while the generalized sidelobe canceler (GSC) can be derived from the linearly constrained minimum variance (LCMV) BF in a centralized context, there is currently no analogy in which the DGSC in [27] is derived from the D-LCMV BF in [26]. In fact, the two algorithms even have a slightly different communication cost (while theoretically achieving the same BF solution), and it is unclear where and why this discrepancy originates.

Therefore, a first goal of this review paper is to provide a top-down description of these algorithms, in a way such that they can be described within the same generic framework. This generic framework allows to introduce the three algorithms in an accessible way, while also revealing the important similarities between them. The common framework in which the three algorithms are described then also explains how they are fundamentally different at certain crucial points, and we compare the advantages and disadvantages that result from these differences. Furthermore, we will explain why the DGSC cannot be straightforwardly inferred from the D-LCMV BF (as opposed to the centralized case), and why there is a discrepancy between them in terms of communication cost.