Signal Processing 118 (2016) 177–187

Contents lists available at ScienceDirect

Signal Processing

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

Synchronization based on mixture alignment for sound source separation in wireless acoustic sensor networks

Cosme Llerena-Aguilar *, Roberto Gil-Pita, Manuel Rosa-Zurera, David Ayllón, Manuel Utrilla-Manso, Francisco Llerena

Department of Signal Theory and Communications, Polytechnic School of the University of Alcalá, Campus Universitario, Ctra. Madrid-Barcelona Km. 33.600, Alcalá de Henares, 28871 Madrid, Spain

ARTICLE INFO

Article history:
Received 6 March 2015
Received in revised form
1 June 2015
Accepted 26 June 2015
Available online 4 July 2015

Keywords:
Mixture alignment
Speech separation
Short-term log-energy
Wireless acoustic sensor network

ABSTRACT

Desynchronization degrades the performance of many signal processing algorithms in Wireless Acoustic Sensor Networks. It is mainly caused by the different distances between the source and each node and by the clock phase offset and frequency skew. Classical solutions use clock synchronization protocols and algorithms in the communication layer, but these alternatives do not tackle the lack of synchronization caused by the distances between sources and nodes.

In this paper, we present a novel study of the synchronization problem in acoustic sensor networks from a signal processing point of view. First, we propose a theoretical framework that allows us to study the effects of misalignment over any short-time based algorithm, focusing on the requirements of the effective length of the analysis time frame. From this framework, a theoretical synchronization delay is established aimed at reducing the required length of the time frame. Second, two novel alignment methods are developed and are tuned up to reduce the amount of synchronization information required for transmission. The results obtained demonstrate that our proposed methods represent a good solution in terms of performance over the quality of a standard Blind Source Separation algorithm, allowing us to reduce the transmission bandwidth required for synchronization data.

1. Introduction

Wireless Acoustic Sensor Networks (WASNs) are composed of spatially distributed autonomous nodes with acoustic sensors (microphones) to monitor sound sources and with wireless communication links to cooperatively pass data through the network. They represent the next-generation technology for audio acquisition and processing. In a WASN, each node wirelessly communicates with other nodes transmitting information and provides in-network signal processing, for instance, to either separate or estimate a desired signal (e.g., Blind Source Separation (BSS) of speech signals) or to extract certain parameters (e.g., the location or identity of speakers) [1,2].

One of the main problems in a WASN is the desynchronization of mixtures, which may degrade the performance of the separation and localization methods [3]. The problem of synchronization is mainly caused by two factors. First, because each node of a WASN has its own clock, there is an inevitable clock phase offset and clock frequency skew from a standard reference clock frequency. Second, the different distances between nodes and sources give rise to different acoustic propagation delays, which in

* Corresponding author.
E-mail addresses: cosme.llerena@uah.es (C. Llerena-Aguilar), roberto.gil@uah.es (R. Gil-Pita), manuel.rosa@uah.es (M. Rosa-Zurera), david.ayllon@uah.es (D. Ayllón), manuel.utrilla@uah.es (M. Utrilla-Manso), francisco.llerena@edu.uah.es (F. Llerena).

http://dx.doi.org/10.1016/j.sigpro.2015.06.023
0165-1684/

some configurations can affect the performance of the algorithms.

In the case of non-stationary sources such as speech, signal processing algorithms are usually implemented using short-time analysis tools. These tools are based on segmenting the analyzed signals into time frames because it is assumed that the signal is considered nearly stationary in short-time frames. Desynchronization of mixtures entails that there would not be the same source contributions in simultaneous time frames at the different microphones, which can decrease the performance of the algorithms. One possible solution consists of increasing the length of the time frames [4], but in many cases, the required length is not suitable for analyzing non-stationary sources.

Classical solutions to this problem are based on the use of clock synchronization protocols and algorithms in the communication layer [5–7]. In a WASN with dedicated and uniform hardware, synchronization of the sampling rates for analog-to-digital converters (ADCs) is usually manageable [6] and is sometimes even unnecessary if the oscillators are of sufficient quality. One example is presented in [8], where a blind method is proposed to estimate the sampling frequency mismatch using a two-stage procedure. On the other hand, in non-uniform ad hoc WASNs with different devices from different manufacturers or without good management of the communication layer, synchronization of the ADCs may be hard (or impossible), and the resulting signal drift must then be taken into account by the signal processing algorithms [1]. Moreover, wireless communication-based solutions only consider the clock phase offset and clock frequency skew (the clock problem), but they omit the effects of unknown propagation delays on the performance of separation and localization algorithms.

In a different approach, it is possible to synchronize the mixtures using signal processing techniques by analyzing and comparing the signals received at the microphones. This idea has been explored for long recordings to correct the clock skew and offset in the Time–Frequency (T–F) domain [9]. However, the design of systems working in wide areas with moving sources and the analysis of WASN particularities for these algorithms are still open issues [10].

This paper addresses the synchronization problem of WASN nodes from a signal processing point of view, which can be divided into two important parts. First, a theoretical analysis is developed and presented, which allows a study of the requirements of the analysis frame length in a BSS system as a function of the delays caused by clock desynchronization and acoustic propagation. Specifically, a lower bound of the time frame length is obtained, which helps us determine how alignment of the mixtures must be performed, minimizing the length constraint. Second, once a lower bound of the frame length is established, two novel synchronization methods are proposed based on this theoretical analysis, which allows the use of shorter analysis time frames for BSS without decreasing the performance.

One method is based on the cross correlation of the mixtures, and the second method uses the Short-Term Log-Energy (STLE). Our solutions are specifically designed for WASNs, analyzing the distributed and collaborative processing in terms of both computational cost and transmission bandwidth. In some works such as [11], compressing algorithms are proposed to reduce bandwidth usage in WASNs. The proposed solutions can avoid the problem of clock desynchronization, and they also improve the performance of separation algorithms over clock-synchronized systems for WASNs with larger distances between nodes.

This paper is organized as follows. In Section 2, the synchronization problem is formulated and studied from a signal processing point of view. Section 3 includes our theoretical study to determine the bounds for the frame length to improve the performance of BSS algorithms. In Sections 4 and 5, two novel mixture alignment methods are proposed. In both cases, the methods are tuned up to reduce the amount of synchronization information required for transmission. Section 6 describes the database and the protocol used for the experiments and shows the obtained results. Finally, conclusions are presented in Section 7.

2. Problem formulation: signal processing framework

BSS algorithms usually assume that the mixing process is modeled as a linear and time-invariant system. The impulse responses from the sources to the sensors are used to model the different phenomena, for instance, delays of the signals [12], path loss or multipath effects [13]. We are interested in BSS in WASNs. Fig. 1 shows a block diagram of a WASN with two nodes, detailing the synchronization block, source separation block, and communication links.

As we can see in Fig. 1, the first block is the synchronization block, which is the block this paper addresses. In a traditional approach, this block is mainly focused on clock and sampling synchronization, which can be solved using wireless network timing [6]. In those applications in which source localization is required, clock timing synchronization is important because it controls the precision of the results. In the case of BSS algorithms, precise positioning of sources and sensors is not needed, but a good characterization of the mixing process is required. Thus, in the BSS case, clock synchronization can be replaced by any other type of synchronization that guarantees a good separation performance. To obtain an alternative approach, in this paper, we present a novel theoretical study of this block

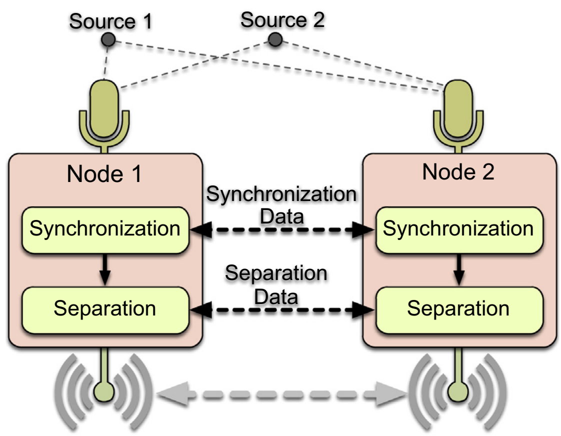

Fig. 1. Scheme of a WASN for sound separation, including the mixture alignment stage.