



Audio coding in wireless acoustic sensor networks [☆]



Adel Zahedi ^{a,*}, Jan Østergaard ^a, Søren Holdt Jensen ^a,
Søren Bech ^{a,b}, Patrick Naylor ^c

^a Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark

^b Bang & Olufsen, 7600 Struer, Denmark

^c Electrical and Electronic Engineering Department, London Imperial College, London SW7 2AZ, UK

ARTICLE INFO

Article history:

Received 18 February 2014

Received in revised form

22 July 2014

Accepted 25 July 2014

Available online 7 August 2014

Keywords:

Wireless acoustic sensor network

Distributed source coding

Rate-distortion function

Covariance distortion constraint

Zero-error coding

ABSTRACT

In this paper, we consider the problem of source coding for a wireless acoustic sensor network where each node in the network makes its own noisy measurement of the sound field, and communicates with other nodes in the network by sending and receiving encoded versions of the measurements. To make use of the correlation between the sources available at the nodes, we consider the possibility of combining the measurement and the received messages into one single message at each node instead of forwarding the received messages and separate encoding of the measurement. Moreover, to exploit the correlation between the messages received by a node and the node's measurement of the source, we propose to use the measurement as side information and thereby form a distributed source coding (DSC) problem. Assuming that the sources are Gaussian, we then derive the rate-distortion function (RDF) for the resulting remote DSC problem under covariance matrix distortion constraints. We further show that for this problem, the Gaussian source is the worst to code. Thus, the Gaussian RDF provides an upper bound to other sources such as audio signals. We then turn our attention to audio signals. We consider an acoustical model based on the room impulse response (RIR) and provide simulation results for the rate-distortion performance in a practical setup where a set of microphones record the sound in a standard listening room. Since our reconstruction scheme and distortion measure are defined over the direct sound source, coding and dereverberation are performed in a joint manner.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The fact that sensor nodes in a Wireless Acoustic Sensor Network (WASN) are wireless, necessitates parsimony in spending the resources, namely power and bandwidth [1].

Two possible remedies are to allow the sensor nodes to communicate with neighboring nodes in order to save communication power, and to use source coding to compress the data acquired by the microphones before transmission in order to reduce the required transmission rate, leading to a better power-bandwidth trade-off.

There are three main consumers of power in a typical sensor node, i.e. sensing, signal processing, and communication units, where the last one is the dominant [2]. Electromagnetic wave power is a highly superlinear function of the inverse of distance, and thus fades rapidly as the distance increases. This means that a multihop scenario where instead of directly being sent to a far-end

[☆] The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant agreement n° ITN-GA-2012-316969.

* Corresponding author.

E-mail addresses: adz@es.aau.dk (A. Zahedi),

jo@es.aau.dk (J. Østergaard), shj@es.aau.dk (S.H. Jensen),

sbe@es.aau.dk (S. Bech), p.naylor@imperial.ac.uk (P. Naylor).

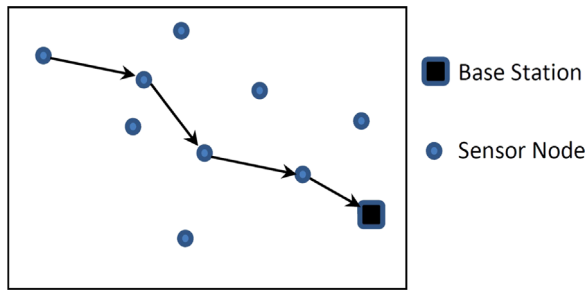


Fig. 1. Message forwarding by multihop communication in a WASN.

destination, the message goes through a sequence of transmissions in short distances, can be more power-efficient. To implement such a scenario, one should assume that neighboring nodes are able to communicate, and thus it is possible to let the nodes which are far from the base station deliver their messages by sequential forwarding through neighboring nodes. This is illustrated in Fig. 1.

To lower the required transmission rate, one may consider source coding for data compression. The audio signals acquired by the microphones are typically highly redundant and can therefore be compressed before transmission. The redundancy is due to correlation in time between the samples of the data acquired by a microphone, as well as spatial correlation between the measurements of the sound field performed by microphones placed at different locations in the environment.

Applying a data compression technique separately to the sequences (the data acquired by the microphones and also those received from neighboring microphones) is equivalent to making use of the time correlation between the samples of a sequence, but ignoring the spatial correlation between different sequences of data.

There are several possibilities to exploit the spatial correlation to further reduce the rate. One possibility is to jointly encode the sequences received by a microphone and the microphone's measurement into one single message instead of separate encoding [3]. Another possibility is to consider the spatial correlation between the message sent from a node and the measurement of the sound field available at the destination node. This latter can be utilized as side information at the decoder to reduce the rate at the encoder. Finally, assuming that nodes transmitting to a neighboring receiving node have the knowledge that which other nodes are sending messages to the same receiving node, it is possible to reduce in the rates of the transmitting nodes due to correlation between their messages by applying e.g. asymmetric or non-symmetric Slepian–Wolf coding [4]. However, in this work, we do not make this assumption, and thus consider only the two first cases as done in [5].

To exploit the availability of the side information at the encoder for reduction in the rate, we rely on results from distributed source coding (DSC), which was started by Slepian and Wolf in [6]. In particular, it is possible to separately encode two correlated discrete sources at a sum-rate equal to the joint entropy of the two sources and reconstruct the sources at a joint decoder. This result was

then extended in [7,8] to the case of continuous sources, assuming that one source is available at the decoder and the other one is to be discretized and encoded for a given distortion. It was shown that for Gaussian sources under an MSE distortion constraint, the rate can be as low as the case where the side information is also available at the encoder.

Measurements made by wireless microphones are generally digitized to discrete-time and discrete-amplitude sequences. This incurs some distortion depending on the quantization step-size. For a given distortion, there is a minimum achievable rate given by the so-called rate-distortion function (RDF). The RDF can be used as a lower bound to assess the performance of any source coding scheme. The interested reader is referred to [9,10] for more information on source coding theory.

In this paper, we derive the local RDF for an arbitrary node in a WASN, assuming that multiple sources at a transmitting node are jointly encoded taking into account (in a distributed sense) the side information available at the receiving node. We will then show in the simulations that local RDFs can be used for rate allocation in the network to achieve the optimal sum-rate. We will also design a coding scheme, which in theory and under certain asymptotical conditions, achieves the theoretical RDF. Vector sources will be considered to allow for modeling of memory in the sources, and distortion constraints are defined in form of covariance matrices for generality. This paper extends our previous work [5] by first proving that Gaussian RDF is an upper bound to other distributions including audio, and then applying our results to Gaussian as well as audio signals. We also provide a complete proof for the RDF, which was only sketched in [5]. In Section 2, the problem is formulated, notations and assumptions are mentioned, and the acoustic channel model used in this paper will be discussed. In Section 3, sufficient statistics are used for joint encoding of correlated sources into a single source in a DSC scenario. Then the RDF will be derived for vector Gaussian sources with noisy measurements under covariance matrix distortion constraints. We further prove that for this setup, Gaussian distribution is the worst case for coding in terms of rate-distortion. In Section 4, we provide simulation results for Gaussian sources and real audio measurements. Section 5 concludes the paper.

2. Preliminaries

2.1. Notation, assumptions, and problem formulation

We denote random vectors by lowercase boldface, matrices by uppercase boldface, and scalars by italic letters. The operations $\text{tr}(\cdot)$, $I(\cdot; \cdot)$, $h(\cdot)$, and $E[\cdot]$ stand for trace of a matrix, mutual information, differential entropy and expectation, respectively. Markov chains are denoted by two-headed arrows; e.g. $\mathbf{x} \leftrightarrow \mathbf{y} \leftrightarrow \mathbf{z}$. Probability density functions are denoted by $f(\cdot)$ and covariance and cross-covariance matrices are denoted by symbol Σ followed by a subscript indicating the random vectors involved and possibly conditioning. We assume that all covariance matrices are of full rank. For two positive semidefinite

Download English Version:

<https://daneshyari.com/en/article/6959599>

Download Persian Version:

<https://daneshyari.com/article/6959599>

[Daneshyari.com](https://daneshyari.com)