DISTRIBUTED COMPRESSION IN ACOUSTIC SENSOR NETWORKS USING OVERSAMPLED A/D CONVERSION

Olivier Roy [†] *and Martin Vetterli* [†§]

[†] School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[§] Department of Electrical Engineering and Computer Sciences
University of California at Berkeley (UCB), Berkeley, CA 94720, USA

email : {olivier.roy, martin.vetterli}@epfl.ch

ABSTRACT

We address the problem of distributed compression in acoustic sensor networks. A typical scenario consists of a set of microphones that record a sound source located at some unknown position. The goal then is to convey the corresponding audio signals to a central base station for processing. We thus aim at designing efficient distributed communication schemes that minimize the overall bit-rate needed in order to achieve a given reconstruction accuracy. In this paper, we show how the a priori knowledge of the system's geometry can be beneficially used in order to lower the transmission rates. We propose a distributed coding scheme for simple synthetic sources. These results are then applied to more general signals by means of oversampled analog-to-digital conversion. Simulation results confirm the effectiveness of our approach.

1. INTRODUCTION

Sensor networks have recently emerged as a promising technology for a variety of applications ranging from surveillance to environmental monitoring. The scenario generally envisioned consists of a large number of distributed nodes that collaborate in order to process data available at multiple distant locations. Typical such devices are self-powered low-cost processing units with wireless communication capabilities. Beyond the appealing perspectives offered by such distributed infrastructures, this new paradigm has given rise to numerous challenging problems from both a theoretical and practical viewpoint.

In this work, we focus on acoustic sensor networks, where multiple audio capture devices are deployed in order to record the spatio-temporal sound field created by an acoustic source. The properties and sampling of this field were recently investigated by Ajdler *et al.* under the name of *plenacoustic function* [1]. These sensors then wirelessly transmit the acquired data to a central base station which aims at extracting relevant features of the original sound field by means of judicious signal processing. The tasks that can be performed at the base station in this scenario are manifold. Possible examples are time-delay estimation for source localization and tracking or beamforming for enhanced audibility. In the centralized framework, these problems have received considerable attention over the

last decades and are relatively well understood. In the distributed scenario, however, the very limited computation and communication capabilities of current sensor technology motivate the quest for efficient protocols that specifically target these new design criteria. In this context, the results obtained in the mid 70's by Slepian-Wolf [2] and Wyner-Ziv [3] on distributed compression of correlated sources, have laid the theoretical grounds for promising advances. Applications of these results to the plenacoustic function can be found in [4]. From a practical standpoint, however, such distributed code designs have only recently emerged from the research community.

In this paper, we wish to design an efficient distributed coding scheme in the scenario where multiple microphones receive the signal emitted by a sound source whose location is unknown. An equivalent problem has been addressed by Gehrig and Dragotti in [5] for the distributed compression of the plenoptic function. In this sense, we follow their approach in the context of the plenacoustic function. In particular, we show how the different parameters of our setup can be efficiently used in order to decrease the overall bit-rate needed to transmit the acquired signals to a central base station while ensuring a given reconstruction fidelity. We first illustrate the main ideas of our distributed coding scheme by means of simple synthetic sources made of diracs. These results are then extended to periodic bandlimited signals with appropriate use of *oversampled analog-to-digital (A/D) conversion*. Apart from its economical justification, the use of oversampling to increase conversion accuracy is usually considered to be dramatically inferior, in rate-distortion sense, to quantization refinement. We however show that, in our setup, this type of conversion is much more amenable to low-complexity distributed coding and thus provide significant gains over simple A/D conversion. Note that the benefits of oversampled A/D conversion in sensor networks have already been demonstrated by Ishwar et al. in [6]. We finally present some simulation results in order to confirm the effectiveness of our scheme.

The outline of the paper is as follows: in Section 2, we present the setup under consideration along with our problem statement. Section 3 provides a precise description of the distributed compression scheme for both streams of diracs and periodic bandlimited signals. Simulation results are given in Section 4. We finally offer some conclusions and future directions of research in Section 5.

2. ACOUSTIC SENSOR NETWORKS

The problem setup that we consider in this paper consists of M perfectly synchronized acoustic sensors (microphones) evenly spaced
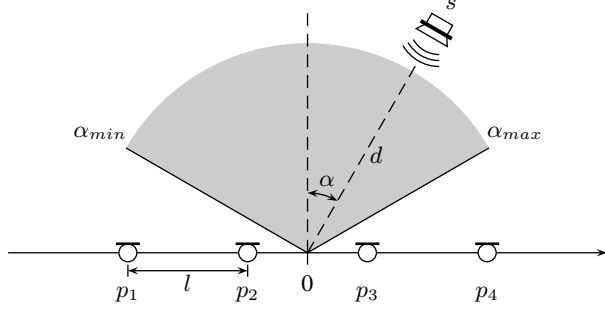
Fig. 1. The acoustic sensor network under consideration with $M = 4$ microphones.

along a line. The distance between two consecutive microphones is l. The origin of the axis is set in the middle of the array such that the position of microphone m, denoted p_m, is given by

$$p_m = \left(m - \frac{M+1}{2} \right) l \tag{1}$$

for $m = 1, 2, \ldots, M$. The source $s(t)$ is an acoustic wave propagating at speed c with a direction $\alpha \in [\alpha_{min}, \alpha_{max}] \subset [-90, 90]$ [deg] from a distance $d \geq 0$ of the origin. We assume that the above parameters are known to both the sensors and the central base station. Our setup is illustrated in Figure 1. For the scope of this paper, we will work under the far-field assumptions, i.e. that the signal received at the microphone array can be accurately modelled as a plane wave with a constant attenuation factor. Without lost of generality, this factor will be set to 1. The signal $x_m(t)$ recorded at microphone m can thus be generally expressed as

$$x_m(t) = s(t - \tau_m) + n_m(t) \tag{2}$$

where τ_m is the propagation delay and $n_m(t)$ is an additive white Gaussian noise independent of both $s(t)$ and $n_{m'}(t)$ for $m' \neq m$. In this work, we will however not take noise into account and consider the received signals to be simply delayed versions of the source signal. Extension of the material presented here to the noisy case is currently under investigation. Note that τ_m can be easily computed as

$$\tau_m = \frac{d - p_m \sin \alpha}{c}. \tag{3}$$

In this context, we address the problem of designing an efficient distributed compression scheme that allows the acoustic sensors to transmit their signal to a central base station in a power-efficient manner, i.e. using low-complexity operations and limited bit-rates.

3. DISTRIBUTED COMPRESSION

3.1. Background

In their celebrated paper [2], Slepian and Wolf showed that the lossless encoding of two correlated discrete random sources can be performed separately without any loss in the overall rate, provided that the correlation is known at both the encoders and the decoder. For this result to be applicable from a signal processing viewpoint, it is thus necessary to represent the signals encountered in practice in such a way that they fall into the framework considered by the aforementioned authors. To this end, we show in the next two subsections how this goal can be achieved using oversampled A/D conversion and some a priori knowledge about our problem setup.

3.2. Streams of diracs

Consider first the source $s(t)$ modelled as a sequence of K diracs of unit weight, located at time positions t_k for $k = 1, 2, \ldots, K$. The purpose of this model is not to provide an accurate representation of real audio signals, but to introduce the main ideas behind our distributed coding strategy. Note that the concepts developed in this section can be applied to the dual problem where diracs in the frequency domain translate into complex exponentials in the time domain. Let us denote by $t_{k,m}$ the time of arrival (TOA) of the k-th dirac at microphone m. It can be easily computed as

$$t_{k,m} = t_k + \tau_m = t_k + \frac{d - p_m \sin \alpha}{c}. \tag{4}$$

Assume that the K diracs arrive at all the microphones within some time interval $[0, T)$ for $T > 0$.

Let us first consider a setup with only $M = 2$ microphones. Their goal is to provide the central base station with an accurate description of their observation. In this scenario, this amounts to transmit the TOA of the diracs. Since we want a limited transmission rate, the observation interval $[0, T)$ is divided into n bins of duration $T_s = T/n$. Each $t_{k,m}$ is quantized into the bin index $b_{k,m} = \lfloor t_{k,m}/T_s \rfloor \in \{0, 1, \ldots, n-1\}$ and transmitted using $B = \lceil \log_2(n) \rceil$ bits. The reconstructed TOAs $\hat{t}_{k,m}$ are computed as

$$\hat{t}_{k,m} = T_s b_{k,m} + T_s/2. \tag{5}$$

At this point, the overall transmission rate is $R = 2KB/T$ bits per second. However, by taking into account the geometrical properties of the model introduced in Section 2, we can see that the delay between the signals received at microphones 1 and 2 is given by $\tau_{2,1} = t_{k,2} - t_{k,1} = -(l \sin \alpha)/c$ for all k. Since $\alpha \in [\alpha_{min}, \alpha_{max}]$, it holds that

$$t_{k,2} - t_{k,1} \in \left[-\frac{l \sin \alpha_{max}}{c}, -\frac{l \sin \alpha_{min}}{c} \right]. \tag{6}$$

This can be equivalently expressed in terms of quantization indexes as

$$b_{k,2} - b_{k,1} \in \left[-\left\lceil \frac{l \sin \alpha_{max}}{T_s c} \right\rceil, -\left\lfloor \frac{l \sin \alpha_{min}}{T_s c} \right\rfloor \right]. \tag{7}$$

The difference between $b_{k,1}$ and $b_{k,2}$ can thus be computed using only $B_1 = \lceil \log_2(\delta + 1) \rceil$ bits where $\delta = \lceil (l \sin \alpha_{max})/(T_s c) \rceil - \lfloor (l \sin \alpha_{min})/(T_s c) \rfloor$. Note that when $\alpha_{min} = \alpha_{max} \neq 0$, this difference is non-zero since quantization still introduces a one-bin uncertainty. A similar analysis is provided by Gehrig and Dragotti in the context of the distributed compression of the plenoptic function [5]. In that paper, they present a simple and efficient distributed coding scheme that can be directly applied to our problem. The idea is that microphones 1 and 2 respectively send the last B_1 bits of $b_{k,1}$ and $\tilde{b}_{k,2} = b_{k,2} + \lceil (l \sin \alpha_{max})/(T_s c) \rceil$. They also transmit complementary subsets of the first $B_2 = B - B_1$ bits. This allows to perfectly decode both bin indexes at the central base station and is proved to be optimal (in the sense that it achieves the Slepian-Wolf bounds) when $b_{k,1}$ and $\tilde{b}_{k,2} - b_{k,1}$ are uniformly distributed in $\{0, 1, \ldots, 2^B - 1\}$ and $\{0, 1, \ldots, \delta\}$, respectively, with $\delta = 2^{B_1} - 1$. The interested reader is referred to [5] for a complete description of the scheme.

The use of the aforementioned coding strategy is easily seen to require an overall transmission rate of $R = K(B + B_1)/T$ bits per second. It thus provides a gain of $G = KB_2/T$ bits per second with respect to the case where the geometrical information is not taken
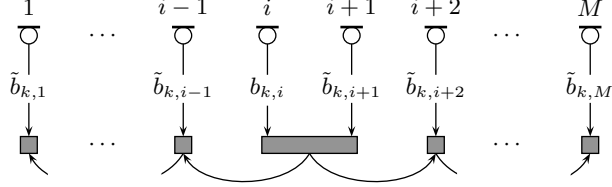
Fig. 2. Sequential decoding with multiple microphones.

Fig. 3. Quantization threshold crossings of two periodic bandlimited signals: (a) $s_1(t) = s(t - \tau_1)$ and (b) $s_2(t) = s(t - \tau_2)$ for $\tau_1 = 0.1$ [s], $\tau_2 = 0.15$ [s], $T = 1$ [s] and $W = 9$. The quantization thresholds are marked with dashed lines and the reconstruction levels with dotted lines.

into account. As the sampling frequency $1/T_s$ is increased, we have that

$$G = \frac{K}{T}(B - B1) \tag{8}$$

$$\leq \frac{K}{T}\log_2\left[\frac{cT}{l(\sin\alpha_{max} - \sin\alpha_{min}) + T_s c}\right] \tag{9}$$

$$\xrightarrow{T_s \to 0} \frac{K}{T}\log_2\left[\frac{cT}{l(\sin\alpha_{max} - \sin\alpha_{min})}\right]. \tag{10}$$

The gain is therefore bounded and depends on the parameters of our system. In particular, decreasing the distance l between the two microphones allows to increase the gain by reducing the possible disparity between the TOAs. Also, having a more precise knowledge of the source location allows to tighten the bounds α_{min} and α_{max} in order to further decrease the overall bit-rate. This suggest an iterative algorithm, where the central base station would feedback to the sensors an estimate of the direction of arrival (DOA). This estimate would in turn allow the sensors to reduce their transmission rates while maintaining a constant reconstruction accuracy or conversely, to decrease the distortion while keeping constant bit-rates.

Let us now turn our attention to the case with $M > 2$ microphones. When quantization is not considered, the geometry of our system allows to compute the TOAs of the K diracs at any microphones based solely on what is recorded at two different positions. The quantization process, however, introduces uncertainty (noise) in the reconstructed TOAs. In this context, the use of a larger number of microphones becomes beneficial since it exploits spatial redundancy. It allows, for example, to increase the robustness of time-delay estimation schemes that may be performed at the base station. Note that a similar argument motivates the encoding of every dirac's position at each microphone. When considering more than two sensors, our distributed coding scheme can easily be extended as follows: for each dirac k, two arbitrary consecutive sensors, say i and $i + 1$, encode their bin index as described previously. Microphone $j \notin \{i, i + 1\}$ then only send the last B_1 bits of $\tilde{b}_{k,j}$ defined as

$$\tilde{b}_{k,j} = b_{k,j} + \lceil (l\sin\alpha_{max})/(T_s c)\rceil. \tag{11}$$

The decoder retrieves the bin index of microphones i and $i + 1$ and then *sequentially* decodes the bin index of their direct neighbor. This process is illustrated in Figure 2.

In this case, the overall coding rate can be computed as $R = K[B + (M - 1)B_1]/T$ bits per second. This strategy provides thus a gain of $G = K(M - 1)B_2/T$ bits per second with respect to the case where the geometrical information is not taken into account. We note a linear increase with the number of sensors.

3.3. Periodic Bandlimited Signals

We will now extend the distributed coding scheme explained in the previous subsection to periodic bandlimited signals by means of oversampled A/D conversion. Improving the reconstruction accuracy by
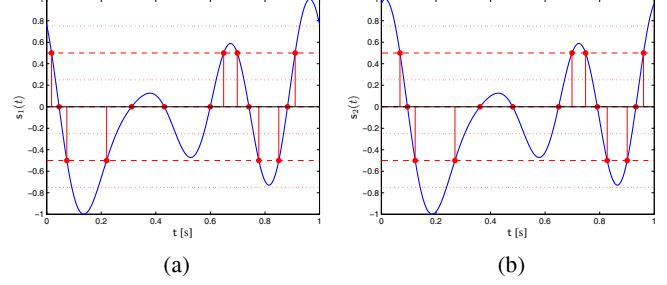
oversampling is often considered as significantly inferior to refining quantization (simple A/D conversion). This is mainly motivated by the error behavior given by the white noise model under linear reconstruction, namely, for a signal $s(t)$ and its reconstruction $\hat{s}(t)$,

$$E[|s(t) - \hat{s}(t)|^2] = \frac{\Delta^2}{12r} \tag{12}$$

where Δ denotes the quantization step size, r the oversampling ratio and $E[\cdot]$ the expectation. However, it has been recently proved that the error-rate characteristics of oversampled A/D conversion can be drastically improved using non-linear reconstruction algorithms, referred to as *consistent*. More precisely, it can be shown [7] that for real periodic bandlimited signals of period T and having $W = 2M + 1$ non-zero Fourier coefficients, the reconstruction mean-squared error (MSE) behaves as $O(1/r^2)$ as opposed to the $O(1/r)$ given by Equation (12). This is provided that the number Q of quantization threshold crossings (QTC) in the interval $[0, T]$ is greater than W. Moreover, for a small enough sampling interval T_s, the sequence of QTCs completely determines the quantized samples of $s(t)$ and vice-versa [8]. Note that the first QTC level in the interval $[0, T]$ is described with q_1 bits and the subsequent ones can be specified using only 1 bit [8]. In this case, the quantized signal can be encoded with a rate of

$$R = \frac{Q}{T}B + \frac{Q - 1 + q_1}{T} \tag{13}$$

bits per second, where B is defined as in the previous subsection. Combined with the MSE behavior of consistent reconstruction algorithms, we directly obtain an exponentially decaying error-rate characteristic.

We can now make use of the above considerations for the problem at hand. The key to our approach is to describe the signals to be transmitted by the different microphones as a sequence of QTCs. More precisely, for each time interval $[nT, (n + 1)T)$ with $n \in \mathbb{N}$, the microphones sample their observed signal with a sampling frequency $1/T_s$, high enough such as to ensure a one-to-one mapping between the quantized samples and the QTCs. The number of reconstruction levels is chosen such as to have at least W QTCs. This process is illustrated in Figure 3. For each QTC k, the time position can be encoded using the method described in the previous subsection. Note that each QTC level only needs to be encoded by one microphone. At the central base station, the time positions of the QTC sequences can be decoded within an accuracy of T_s seconds.
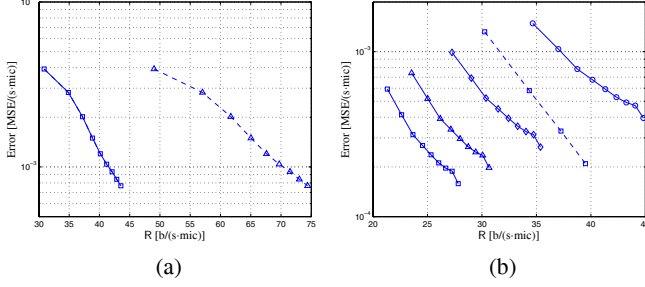
Fig. 4. Error-rate curves obtained for $\alpha_0 = 90$ [deg]. (a) Oversampled A/D conversion with (plain) and without (dashed) geometrical considerations for $M = 2$ microphones. (b) Simple A/D conversion (dashed) and our distributed coding scheme (plain) for $M \in \{2, 3, 4, 5\}$ (from right to left).
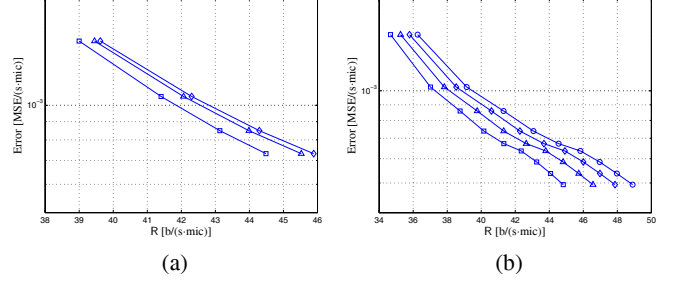
Fig. 5. Error-rate curves obtained for $M = 2$ microphones. Our distributed scheme for (a) $\alpha_0 \in \{10, 50, 90\}$ [deg] (left to right) and (b) $l \in \{0.5, 1, 1.5, 2\}$ [m] (left to right).

The quantized samples of the transmitted signals are then recovered and a consistent reconstruction algorithm is applied to reconstruct the signals. Note that we could also base our reconstruction directly on the QTCs using stable interpolation techniques such as the ones presented in [9]. For M microphones, the overall rate can thus be computed as

$$R = \frac{Q}{T}\left(B + (M-1)B_1\right) + \frac{Q - 1 + q_1}{T} \qquad (14)$$

bits per second, where B and B_1 are defined as in the previous subsection. Notice that q_1 can be made arbitrarily small if we encode the QTC levels over multiple periods [8]. It will be thus neglected in the simulation results presented in the next section.

4. SIMULATION RESULTS

We now present the error-rate characteristics obtained numerically for our distributed coding strategy. The simulations are averaged over 500 randomly generated periodic bandlimited signals $s(t)$ with period $T = 1$ [s] and $W = 7$ non-zero Fourier coefficients. The parameters of our system are $c = 330.7$ [m/s], $l = 0.5$ [m] and $\alpha_{max} = -\alpha_{min} = \alpha_0$. All the curves are plotted on a per microphone basis.

We first set $\alpha_0 = 90$ [deg]. In Figure 4(a), we show the results obtained for $M = 2$ microphones using oversampled A/D conversion, with and without taking into account the geometrical properties of our setup. We clearly see the gain achieved by our distributed coding scheme. Figure 4(b) compares the error-rate curves between our coding strategy and simple A/D conversion applied independently at each sensor for various array configurations. We observe that for $M > 2$ sensors, our technique outperforms simple A/D conversion for which distributed coding techniques would involve significantly higher complexity and latency. Moreover, we are confident that more efficient oversampled A/D conversion schemes can be used to further improve these results, in particular for the two-microphone setup. We now consider $M = 2$ sensors. Figure 5(a) depicts the curves corresponding to different values of α_0 to illustrate the gain provided by a more accurate knowledge of the DOA. Finally, Figure 5(b) investigates how the distance l affects the error-rate characteristics.

5. CONCLUSIONS

In this paper, we have presented a distributed compression scheme for acoustic sensor networks. We have shown how oversampled A/D conversion can be efficiently used in this context to significantly reduce the overall transmission bit-rate. Ongoing research is focusing on extending our approach to noisy signals and oversampled A/D conversion schemes with better accuracy. Tradeoffs between time and frequency domain coding are also under investigation.

6. REFERENCES

[1] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *to appear in IEEE Trans. Signal Processing*, 2005.

[2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, 1973.

[3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[4] R. L. Konsbruck, E. Telatar, and M. Vetterli, "On the multiterminal rate-distortion function for acoustic sensing," *to appear in IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2006.

[5] N. Gehrig and P. L. Dragotti, "Distributed compression of the plenoptic function," in *IEEE International Conference on Image Processing*, October 2004, vol. 1, pp. 529–532.

[6] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: a bit-conservation principle," in *IEEE International Symposium on Information Processing in Sensor Networks*, 2003, pp. 17–31.

[7] N. T. Thao and M. Vetterli, "Reduction of the MSE in R-times oversampled A/D conversion from $O(1/R)$ to $O(1/R^2)$," *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 200–203, January 1994.

[8] Z. Cvetković and M. Vetterli, "Error-rate characteristics of oversampled analog-to-digital conversion," *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1961–1964, September 1998.

[9] Z. Cvetković and I. Daubechies, "Single-bit oversampled A/D conversion with exponential accuracy in the bit-rate," in *IEEE Data Compression Conference*, March 2000, pp. 343–352.