

ON THE MULTITERMINAL RATE-DISTORTION FUNCTION FOR ACOUSTIC SENSING

Robert L. Konsbruck, Emre Telatar, and Martin Vetterli

École Polytechnique Fédérale de Lausanne (EPFL)
School of Computer and Communication Sciences (I&C)
CH-1015 Lausanne, Switzerland
E-mail: {robert.konsbruck, emre.telatar, martin.vetterli}@epfl.ch

ABSTRACT

We consider the rate-distortion problem for recording a Gaussian acoustic field with an array of sensors equipped with microphones. Our analysis is based on the field's spatio-temporal correlation structure, which is induced by the physics of sound propagation. Under certain assumptions about the spectrum of the wave kernel, we determine the rate-distortion functions for various distributed source coding schemes and sampling lattices. Our results also include an answer to the multiterminal rate-distortion problem for the setup considered in the paper.

1. INTRODUCTION

Wireless sensor networks consist of a large number of small, low-power devices, equipped with limited processing and communication capabilities, which measure a physical phenomenon in some region of interest. The goal is typically to reconstruct the measured physical field (e.g. temperature, pressure, light) at an intelligent central unit to within some prescribed accuracy, and this at the lowest possible cost on the communication link and power consumption by the sensors. Wireless sensor networks have been intensely studied over the past years by both the information theory and the signal processing communities [1], and some fundamental results on the performance of such networks under various operation constraints have been obtained.

However, in the previous work on sensor networks, the physics that describes the underlying phenomenon has often not been taken into account. The data sources, although allowed to have a generic spatial correlation, have usually been assumed to be memoryless. For real physical processes however, there is a spatio-temporal correlation structure, which is determined by physical laws. We investigate the role of physical models in source coding problems appearing in typical sensor network applications with the aim of quantifying the potential gains in terms of rate-distortion performance that exploiting these models can provide and also of devising practical strategies to approach the theoretical bounds [2].

In this paper, we consider the rate-distortion problem for recording a spatio-temporal acoustic field in some region, which is generated by sound sources at a distant location. Sensors equipped with microphones sample the sound field, quantize the samples, and transmit the compressed samples to a central base station, which produces an estimate of the original sound field at any point in the recording region.

The recorded sound field is the solution of the acoustic wave equation with appropriate boundary conditions and a driving function determined by the sound sources. In other words, it is the result of the spatio-temporal convolution of the source field with the Green's function of the wave equation. Referring to Ajdler *et al.* [3], we call that Green's function the *plenacoustic function* (PAF). The interesting observation is that for temporally bandlimited sources, the recorded field is essentially bandlimited in space, and thus, it can be sampled with a microphone array. For this setup, we determine rate-distortion functions under various communication and complexity constraints. We also analyze and compare the performance of sampling and coding strategies that are adapted to the particular spectral shape of the PAF.

2. THE PLENACOUSTIC FUNCTION

The scenario that we consider in this paper, consists in a sensor network recording a spatio-temporal acoustic field on an infinite line \mathcal{V} , which is generated by sound sources located on a parallel line \mathcal{U} at a distance d from the recording line \mathcal{V} . The setup is shown in Figure 1. The sound sources emit an acoustic field $U(x, t)$, which induces a sound field $V(x, t)$ on the recording line. Sensors equipped with microphones are equally spaced on the recording line and sample the induced field $V(x, t)$.

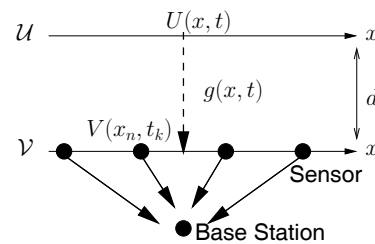


Fig. 1. Sound sources located on a line \mathcal{U} emit an acoustic field $U(x, t)$, which induces another sound field $V(x, t)$ on a parallel line \mathcal{V} at a distance d .

In general, an acoustic field $\tilde{V}(x, y, z, t)$ depends on a sound source $\tilde{U}(x, y, z, t)$ through the scalar wave equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \tilde{V}(x, y, z, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \tilde{V}(x, y, z, t) = -\tilde{U}(x, y, z, t), \quad (1)$$

where c is the wave velocity. In the setup of Figure 1, we study the evolution of the sound fields along the x -axis, i.e., we consider the restriction $V(x, t) = \tilde{V}(x, 0, 0, t)$ of the solution to (1) with a driving function $\tilde{U}(x, d, 0, t) = U(x, t)$. The sound field $V(x, t)$ can then be expressed as

$$V(x, t) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x - \xi, t - \tau) U(\xi, \tau) d\xi d\tau,$$

where $g(x, t)$ is the PAF for this configuration and is given by

$$g(x, t) = \frac{\delta\left(t - \frac{1}{c}\sqrt{x^2 + d^2}\right)}{4\pi\sqrt{x^2 + d^2}}.$$

Its Fourier transform can be shown to be [3]

$$G(\Phi, \Omega) = -\frac{j}{4} H_0^* \left(d \sqrt{\left(\frac{\Omega}{c}\right)^2 - \Phi^2} \right), \quad (2)$$

where Φ and Ω represent the spatial and temporal frequencies, and $H_0^*(\cdot)$ denotes the complex conjugate of the Hankel function of order zero. Figure 2(a) shows the squared norm of $G(\Phi, \Omega)$ in dB. We observe that for fixed Ω , the decay along the spatial frequency axis (i.e. as a function of Φ) is exponential, and most of the energy is contained in the part of the spectrum satisfying $|\Phi| < |\Omega|/c$. The region where $|\Phi| > |\Omega|/c$ corresponds to the evanescent mode of the wave, where the wave loses its propagating character. In the following, we consider an approximation of the spectrum, and we assume that it vanishes when $|\Phi| > |\Omega|/c$.

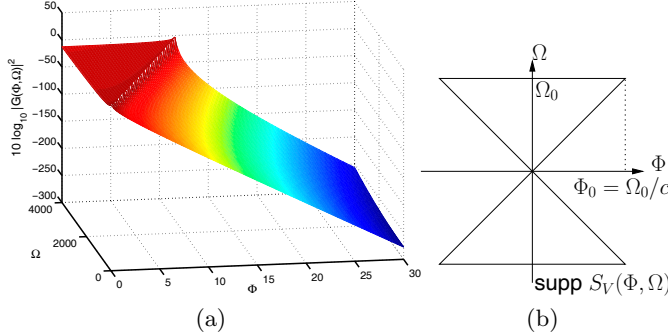


Fig. 2. (a) Squared norm of the Fourier transform $G(\Phi, \Omega)$ of the PAF in dB. (b) Butterfly-like spectral support of the PSD $S_V(\Phi, \Omega)$ of the induced acoustic field $V(x, t)$.

3. PROBLEM STATEMENT

We model the sound source $U(x, t)$ as a continuous stationary Gaussian stochastic process with a flat and bandlimited PSD $S_U(\Phi, \Omega)$, i.e.,

$$S_U(\Phi, \Omega) = \sigma^2 1_{[-\tilde{\Phi}_0, \tilde{\Phi}_0]}(\Phi) 1_{[-\Omega_0, \Omega_0]}(\Omega),$$

where σ is some real parameter, and $\tilde{\Phi}_0$ and Ω_0 are the spatial and temporal bandwidths. The acoustic field $V(x, t)$ that is induced in the recording region, is then also a stationary Gaussian stochastic process, whose PSD is given by

$$S_V(\Phi, \Omega) = S_U(\Phi, \Omega) |G(\Phi, \Omega)|^2,$$

where $G(\Phi, \Omega)$ is given in (2). If we assume, as we have mentioned above, that the spectrum of the Green's function

vanishes in the evanescent mode, then the support of the PSD $S_V(\Phi, \Omega)$ has a butterfly-like shape that is bandlimited on both frequency axes, as it is shown in Figure 2(b). We observe that the spatial bandwidth is at most equal to $\Phi_0 = \Omega_0/c$, independently of the actual spatial bandwidth $\tilde{\Phi}_0$ of the source. According to Shannon's sampling theorem, the field $V(x, t)$ is thus completely described by its samples taken on a sufficiently dense sampling grid in the spatio-temporal plane.

The sampling is performed by sensors located on a regular lattice on the recording line. The sensors quantize their observations at a given rate and transmit the quantized samples to the base station over parallel rate-constrained channels. The latter produces an estimate $\hat{V}(x, t)$ of the original field $V(x, t)$ at any point on the recording line. We use a rate-constrained communication model in keeping with current digital communication architectures. In other words, the sensors encode their observations into bit streams, and the base station reconstructs an estimate from these bits. The goal is to minimize the distortion D , measured in MSE per meter and per second, for a given total rate R , measured in nats per meter and per second, spent by the sensors for communicating with the base station. The MSE distortion is defined as

$$D = \lim_{\substack{L \rightarrow \infty \\ T \rightarrow \infty}} \frac{1}{2L} \frac{1}{2T} \int_{-L}^L \int_{-T}^T \mathbb{E} \left[\left(V(x, t) - \hat{V}(x, t) \right)^2 \right] dt dx.$$

For this setup, we determine rate-distortion (RD) functions under various constraints on the inter-sensor communications and the extent to which the spatio-temporal correlation can be taken into account for the quantization. An RD function is defined to be the optimal trade-off between the rate and the distortion under the given constraints. In particular, we compute the *centralized* RD function, where the sensors are allowed to collaborate through free inter-sensor communications to jointly encode the spatio-temporal samples of the sound field $V(x, t)$. Next, we determine the *spatially independent* RD function, where inter-sensor communications are precluded, and each sensor quantizes the locally observed temporal stochastic process ignoring the spatial dimension. Having the sensors consider and exploit the spatial correlation without communicating with each other leads to the general *multiterminal* RD problem, which is the true problem of interest in sensor network applications, but which remains an unsolved question to date. However, the corresponding RD function is lower bounded by the centralized RD function and upper bounded by the spatially independent RD function, so that, depending on the size of the gap between these two functions, the precise determination of the multiterminal RD function may be less relevant for practical applications. Below, we construct a case for which we devise a sampling and coding scheme achieving the lower bound, so that we determine the multiterminal RD function exactly for this case.

4. SAMPLING OF THE ACOUSTIC FIELD

Concerning the sampling of the acoustic field $V(\mathbf{z}) = V(x, t)$, with $\mathbf{z} = (x, t)$, we consider regular lattices in the spatio-temporal domain. Let \mathbf{M} denote the generating matrix of the lattice. Sampling the continuous process $V(\mathbf{z})$ produces the discrete stationary Gaussian stochastic process $V[\mathbf{l}] = V(\mathbf{M} \cdot$

\mathbf{l}), where \mathbf{l} is an integer vector representing the indices of the samples along the columns of \mathbf{M} . The density of samples per unit area is given by $1/|\det \mathbf{M}|$, and this number should be sufficiently large so as to avoid aliasing. The stationarity of $V(\mathbf{z})$ is inherited by $V[\mathbf{l}]$, and the PSD $S_V[\psi]$ of the process $V[\mathbf{l}]$ is given by

$$S_V[\psi] = \frac{1}{|\det \mathbf{M}|} \sum_{\mathbf{k}} S_V \left(\mathbf{M}^{-T} \cdot (\psi - 2\pi \mathbf{k}) \right),$$

where the sum is over all integer vectors \mathbf{k} , and \mathbf{M}^{-T} denotes the transposed inverse of the matrix \mathbf{M} . For the sampling of the acoustic field, we consider the rectangular and the quincunx lattices.

For the rectangular lattice, the sensors are separated by a distance $\Delta x = \pi c / \Omega_0$, and they record their samples at integer multiples of $\Delta t = \pi / \Omega_0$. These sampling rates correspond to the individual Nyquist rates on the spatial and temporal frequency axes. The corresponding sampling matrix is given by

$$\mathbf{M}_r = \frac{\pi}{\Omega_0} \begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix}.$$

Figure 3(a) shows the support of the PSD of the sampled process $V(\mathbf{M}_r \cdot \mathbf{l})$, and in particular, it shows the imperfect tiling of the frequency plane with spectral replicas. The gaps that remain between the replicas, translate into an efficiency loss for the rate-distortion behavior.

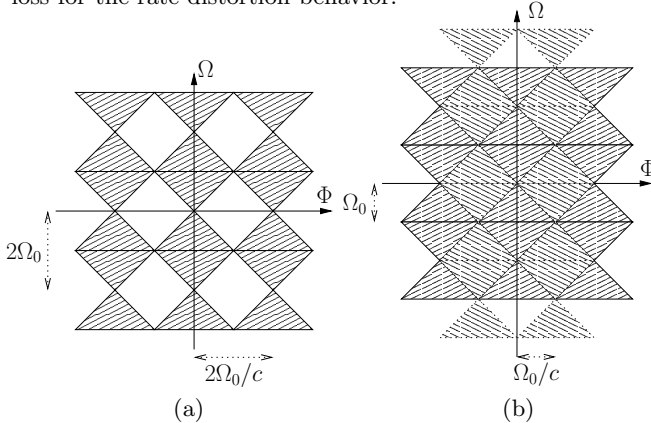


Fig. 3. Spectral supports of the PSDs of the sampled fields for the rectangular lattice (a) and the quincunx lattice (b).

The use of the quincunx sampling lattice to achieve a tighter packing of the spectrum has been suggested in [3]. The sensors are still separated by a distance $\Delta x = \pi c / \Omega_0$, but each sensor records its samples at integer multiples of $2\Delta t = 2\pi / \Omega_0$ and with an offset of $\Delta t = \pi / \Omega_0$ with respect to its neighboring sensors. The corresponding sampling matrix is given by

$$\mathbf{M}_q = \frac{\pi}{\Omega_0} \begin{pmatrix} c & c \\ -1 & 1 \end{pmatrix}.$$

With quincunx sampling, the sampling density is half as large as with the rectangular grid, yet there is no aliasing. There is even a gap-free tiling of the frequency plane with spectral replicas, as it is shown in Figure 3(b). Although each sensor samples the locally observed temporal process below the Nyquist rate and therefore generates aliasing, the overall sampling density is high enough for the base station to be able to produce an aliasing-free estimate.

5. RATE-DISTORTION FUNCTIONS UNDER THE FLAT SPECTRUM ASSUMPTION

In this paper, we determine the RD functions for the coding schemes described in Section 3 under the additional assumption that the PSD $S_V(\Phi, \Omega)$ is constant on its support, i.e.,

$$S_V(\Phi, \Omega) = \sigma^2 \Pi(\Phi, \Omega), \quad (3)$$

where $\Pi(\Phi, \Omega)$ is the indicator function of the butterfly-like region shown in Figure 2(b). The same PSD would also result from the far-field assumption used in the acoustics literature. To compute the RD functions, we use the so-called reverse waterfilling technique as well as the results for stationary Gaussian random processes in [4]. The results are summarized in the following proposition.

Proposition 1. *Under the flat spectrum assumption (3), the RD functions corresponding to the coding schemes defined in Section 3 and the sampling lattices described in Section 4 are given by the following expressions:*

- for centralized coding:

$$R(D) = \frac{\Omega_0^2}{4\pi^2 c} \log \left(\frac{\sigma^2 \Omega_0^2}{2\pi^2 c D} \right); \quad (4)$$

- for rectangular sampling and independent coding:

$$R(D) = \frac{\Omega_0^2}{2\pi^2 c} \log \left(\frac{\sigma^2 \Omega_0^2}{e\pi^2 c D} \frac{1}{2} \left(1 + \sqrt{1 - 2 \frac{\pi^2 c D}{\sigma^2 \Omega_0^2}} \right) \right) + \frac{\Omega_0^2}{2\pi^2 c} \left(1 - \sqrt{1 - 2 \frac{\pi^2 c D}{\sigma^2 \Omega_0^2}} \right); \quad (5)$$

- for quincunx sampling and independent coding:

$$R(D) = \frac{\Omega_0^2}{4\pi^2 c} \log \left(\frac{\sigma^2 \Omega_0^2}{2\pi^2 c D} \right); \quad (6)$$

where $D \in (0, (\sigma^2 \Omega_0^2) / (2\pi^2 c))$.

Figure 4(a) shows the graphs of the RD functions given in Proposition 1. We observe that the expressions (4) and (6) are identical. Thus, under the flat spectrum assumption, the strategy of using the quincunx sampling lattice and independent coding is optimal. The explanation for this fact is that the PSD of the sampled sound field $V(\mathbf{M}_q \cdot \mathbf{l})$ is a constant function as a consequence of the flatness of the spectrum and the perfect tiling of the frequency plane shown in Figure 3(b), and that the processes sampled by different sensors are thus independent. Therefore, encoding these processes independently does not result in any loss in terms of rate-distortion. This also implies that for this scenario, the RD function for multiterminal source coding is known and coincides with the one for centralized coding.

6. SUBBAND SAMPLING

Quincunx sampling requires a small amount of cooperation between the sensors to set up the sampling time offsets. Another way of getting a tighter packing of the spectrum without cooperation consists in decomposing the sound field $V(x, t)$ into M spectral subbands through filtering in the temporal domain, as shown in Figure 5. Each of the resulting processes $V^{(m)}(x, t)$ for $m = 1, \dots, M$ is sampled on a rectangular lattice at the respective Nyquist rates by an appropriate subset

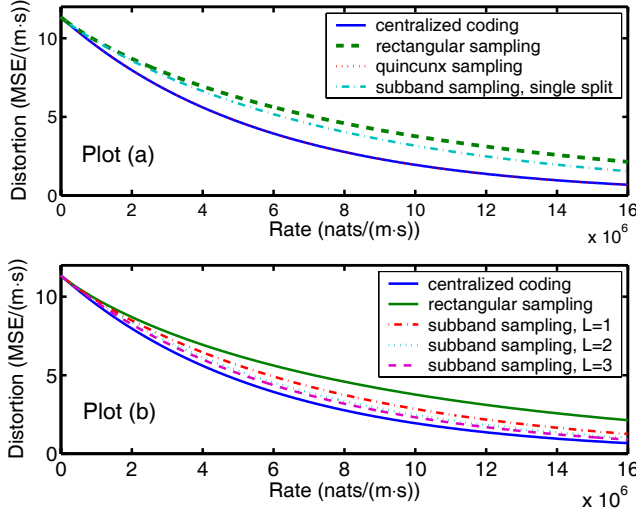


Fig. 4. (a) RD functions for different sampling and coding schemes. The curve for quincunx sampling coincides with the one for centralized coding. (b) RD functions for the subband decomposition method with $L = 1, 2, 3$ and $M = 9L + 1$. The outermost curve corresponds to rectangular sampling and the innermost curve to centralized coding.

of the sensors. Since this method does not require any spatial filtering, which would be impossible in practice, it can be executed by the sensors. Moreover, since the subbands have non-overlapping spectral supports, the processes $V^{(m)}(x, t)$ are independent. As a consequence, they may be encoded independently if the rate allocation among the subbands is done according to the reverse waterfilling principle. We assume again that the sensors perform the encoding independently from each other.

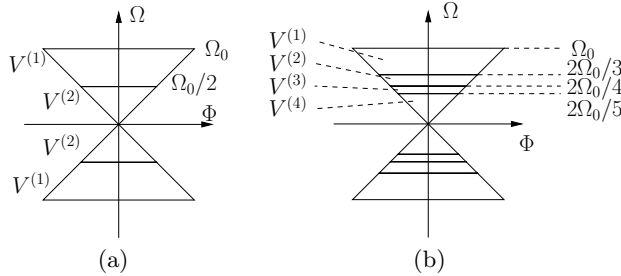


Fig. 5. Spectral decomposition of the sound field $V(x, t)$ into two subbands of equal width (a) and into multiple subbands with $L = 2$ and $M = 4$ (b).

First, we consider the spectral decomposition of the acoustic field $V(x, t)$ into two subbands of equal width, as shown in Figure 5(a). Figure 4(a) shows the graph of the resulting RD function in comparison with the ones given in Proposition 1. We observe that the performance of the rectangular sampling lattice is improved when it is combined with the subband decomposition method.

Next, we consider the spectral subdivision into multiple subbands. Given the spatial bandwidth $\Phi_0 = \Omega_0/c$ of the acoustic field $V(x, t)$ and an integer parameter L , we choose an inter-sensor spacing of $\tilde{\Delta x} = \Delta x/L = (\pi c)/(\Omega_0)$, which corresponds to L times the spatial Nyquist rate of the process

$V(x, t)$. We define $V^{(m)}(x, t)$, for $m = 1, \dots, M$, to be the subband of $V(x, t)$ with

$$|\Omega| \in \begin{cases} \left[\frac{L\Omega_0}{L+m}, \frac{L\Omega_0}{L+m-1} \right] & \text{for } m = 1, \dots, M-1, \\ \left[0, \frac{L\Omega_0}{L+M-1} \right] & \text{for } m = M. \end{cases}$$

Figure 5(b) shows the resulting decomposition for $L = 2$ and $M = 4$. The subband process $V^{(m)}(x, t)$ has the spatial bandwidth $\Phi_0^{(m)} = L\Omega_0/c(L+m-1)$, and can thus be sampled with an inter-sensor spacing of $\Delta x^{(m)} = \pi c(L+m-1)/L\Omega_0$. This sampling density can be realized by having only one sensor out of $L+m-1$ sample the process $V^{(m)}(x, t)$. The rate allocation among the subbands is again done according to the reverse waterfilling principle. Note that increasing the parameter L requires a proportional increase in the number of sensors deployed, whereas an increase of the number of subbands M only adds to the local processing complexity. Given L , we set $M = 9L + 1$, such that the temporal bandwidth of $V^{(M)}(x, t)$ is equal to $\Omega_0/10$. Figure 4(b) shows the resulting RD functions for $L = 1, 2, 3$. With a moderate increase in the number of sensors deployed, the performance of the rectangular sampling strategy combined with the subband decomposition method approaches the one of centralized coding.

7. CONCLUSIONS AND FUTURE WORK

The physics of sound propagation provides the recorded acoustic field with a particular spatio-temporal correlation. We have included the physical model into the analysis of the rate-distortion performance of several distributed source coding schemes. Under certain assumptions about the spectrum of the wave kernel, we have determined the rate-distortion functions for these schemes. In particular, we have shown the optimality of the spatially independent coding scheme with the quincunx sampling lattice, and at the same time, we have given an answer to the multiterminal rate-distortion problem for the setup considered in this paper.

Our current research includes the extension of our results to multidimensional sensor arrays. We also want to determine the rate-distortion functions for the problem of estimating the sound source $U(x, t)$ instead of the induced field $V(x, t)$. Finally, since most physical processes are not bandlimited, we intend to study the sampling of non-bandlimited fields and to determine bounds on the resulting reconstruction error.

8. REFERENCES

- [1] J. P. Hubaux, T. Gross, J. Y. Le Boudec, and M. Vetterli, "Towards self-organized mobile ad hoc networks: the terminodes project," *IEEE Communications Magazine*, vol. 39, no. 1, January 2001.
- [2] B. Beferull-Lozano, R. L. Konsbruck, and M. Vetterli, "Rate-distortion problem for physics based distributed sensing," in *Proc. 3rd International Symposium on Information Processing in Sensor Networks (IPSN)*, 2004.
- [3] Th. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," to appear in the *IEEE Transactions on Signal Processing*.
- [4] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, 1971.