14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

SPATIO-TEMPORAL SAMPLING AND DISTRIBUTED COMPRESSION OF THE SOUND FIELD

Thibaut Ajdler[*], *Robert L. Konsbruck*[*†], *Olivier Roy*[*], *Luciano Sbaiz*[*], *Emre Telatar*[†] *and Martin Vetterli*[*◇]

[*]Laboratory for Audio-Visual Communications (LCAV)
Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

[†]Laboratory of Information Theory (LTHI)
Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

[◇]Department of Electrical Engineering and Computer Science (EECS)
University of California at Berkeley, Berkeley CA 94720, USA

Email: {thibaut.ajdler, robert.konsbruck, olivier.roy, luciano.sbaiz, emre.telatar, martin.vetterli}@epfl.ch

ABSTRACT

We investigate how the sound field induced by an acoustic event evolves over space and time. The characteristics of its bidimensional Fourier spectrum are analyzed and spatio-temporal sampling results using an array of microphones are provided for different scenarios of interest. We then address the distributed compression problem using an information-theoretic point of view. In this context, optimal rate-distortion tradeoffs are derived for two scenarios of interest. A linear network setup is first considered, where a central base station aims at recovering with minimum distortion the signals recorded by an infinite line of microphones. A hearing aid problem is then studied, where two hearing devices exchange data over a rate-constrained wireless link in order to provide spatial noise reduction.

1. INTRODUCTION

Sensor networks have emerged as a powerful tool to acquire data distributed over a large area by means of self-powered and low-cost sensing units. They allow to observe physical fields at different time instants and space locations, thus acting as spatio-temporal sampling devices. Most envisioned deployments of such distributed infrastructures are tailored to a particular sensing task. Examples are environmental monitoring (temperature, humidity or pressure measurements) [1], target tracking [2] or acoustic beamforming [3]. The design of these networked architectures usually involves a complex interplay of source and channel coding principles as a mean to reproduce the sensed data within prescribed accuracy. In this context, a thorough understanding of the physical phenomenon under observation is crucial. It allows to accommodate the design of sensing devices, sampling schemes and transmission protocols to the targeted application, hence providing significant gains over blind communication strategies.

In this work, we examine the spatio-temporal sampling and distributed compression of the sound field acquired with an array of microphones. We first review the simple setup which consists of one emitting source and one recording device. We then look at more complex scenarios by means of the plenacoustic function [4] which describes the evolution of the sound field over space and time. More precisely, given an acoustic event, the plenacoustic function corresponds to the sound that would be recorded at any given position and time. Its Fourier spectrum is shown to exhibit exponential spatial decay rates beyond an essential spectral support. This almost-bandlimited character allows to derive spatio-temporal sampling results using different sampling lattices. Experimental results are provided to confirm the theory.

We then turn our attention to the distributed compression problem, where the samples acquired by the microphones must be efficiently coded and transmitted to a central base station for reconstruction. Note that in the scope of this paper, the problem is solely addressed from a source coding standpoint, the channel coding perspective being matter of current research. Under these assumptions, we focus on two limiting scenarios of interest. In the first setup [5], the source is modelled as a continuous stationary Gaussian space-time process on a line and is recorded using an infinite linear array of microphones (linear network). Closed-form rate-distortion (RD) formulas are provided for various sampling strategies. In particular, we show that under restricting hypotheses, the best achievable RD tradeoff can be obtained by judicious signal processing at the sensors. We thus provide, for this particular example, the solution to the multi-terminal source coding problem whose general solution remains unknown to date [6]. In the second setup [7], we consider the problem where two digital hearing aids, each equipped with an omnidirectional microphone, exchange their sensed acoustic data using a wireless communication link in order to provide collaborative beamforming. In other words, we study the beamforming gain provided by a rate-constrained two-sensor array. Our hearing aids setup is first identified as a remote source coding problem with side information at the decoder. Under assumptions similar to the first setup, we compute a closed-form RD formula in a simple scenario. We then define the gain-rate function, which describes the best achievable gain-rate tradeoff, and provide its corresponding closed-form description. Extension to more complex acoustic environments is also discussed.

The paper is organized as follows: in Section 2, we present the spatio-temporal characteristics of the sound field along with sampling results. The distributed compression task is then addressed for the linear network setup in Section 3 and in the context of the hearing aids problem in Section 4. We present the conclusions in Section 5.

2. SPATIO-TEMPORAL CHARACTERISTICS OF THE SOUND FIELD

This section is devoted to the analysis of the sound field along both the spatial and temporal dimension. For the sake of simplicity, we will work under the free-field assumption [8], i.e. that the reverberation effect due to surrounding objects can be neglected. Extension

The work presented in this paper was supported in part by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

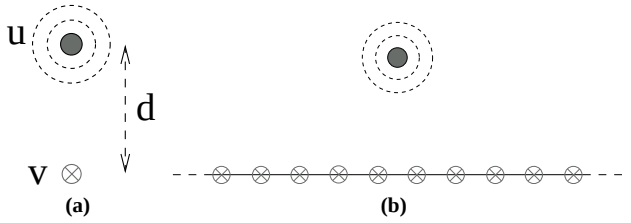

Figure 1: Setups for the study of the sound field in free field. (a) One source and one microphone. (b) One source and an infinite line of microphones.

of the results presented here to the case with room reflections are presented in [4]. Among different spatial setups, the case of an infinite line of microphones is discussed in details. In particular, the 2-dimensional Fourier transform (2D-FT) of the sound field along a line is computed. It is shown that its essential support has a bow-tie-like shape. Different spatio-temporal sampling schemes are then investigated.

Let us first consider the setup depicted in Figure 1(a). A sound $u(t)$ is emitted in free field by an omnidirectional source located at position (x_u, y_u, z_u). The sound recorded at a receiver (microphone) located at position (x_v, y_v, z_v) is written $v(t)$. Considering the channel between the source and the receiver as a linear and time invariant system, we can define an impulse response between these two points. This impulse response, denoted as $h(t)$, is a function of the time and depends on the distance d between the source and the receiver. It is given by [8]

$$h(t) = \frac{\delta\left(t - \frac{d}{c}\right)}{4\pi d}, \tag{1}$$

where c is the speed of sound propagation. Under these considerations, the sound heard at the microphone is simply obtained as the convolution of the source with the impulse response.

We now consider more general setups where we do not only measure the sound field at one point but over larger areas. To this end, we define the *plenacoustic function* (PAF) as the sound field recorded at any possible location for a source located at one particular position. The PAF has been studied for different microphone arrays, such as lines or planes [4], and for sources moving along random trajectories [9]. In the sequel, we will focus on the PAF along a line (the x-axis) as shown in Figure 1(b). For the case of a source emitting a Dirac at time instant t_u, the PAF simply corresponds to the Green's function [8]. In this case, the PAF is described by

$$g(x,t) = \frac{\delta\left(t - t_u - \frac{\sqrt{(x-x_u)^2+(y_v-y_u)^2+(z_v-z_u)^2}}{c}\right)}{4\pi\sqrt{(x-x_u)^2+(y_v-y_u)^2+(z_v-z_u)^2}}. \tag{2}$$

Taking the 2D-FT of (2) leads to a spectrum $G(\Phi,\Omega)$ where Φ stands for the spatial frequency, measured in radians per meter, and Ω for the temporal frequency, measured in radians per second. A closed-form expression of $G(\Phi,\Omega)$ can be obtained and is plotted in Figure 2(a). It can be seen that most of the energy present in the signal is contained in the region of space satisfying

$$|\Phi| \leq \frac{|\Omega|}{c}. \tag{3}$$

Outside of this region, the energy of the spectrum can be shown to decay exponentially fast. As the spectrum is almost bandlimited, the interpolation of the sound field can be achieved using sinc interpolation. A quantitative sampling theorem, trading off sampling

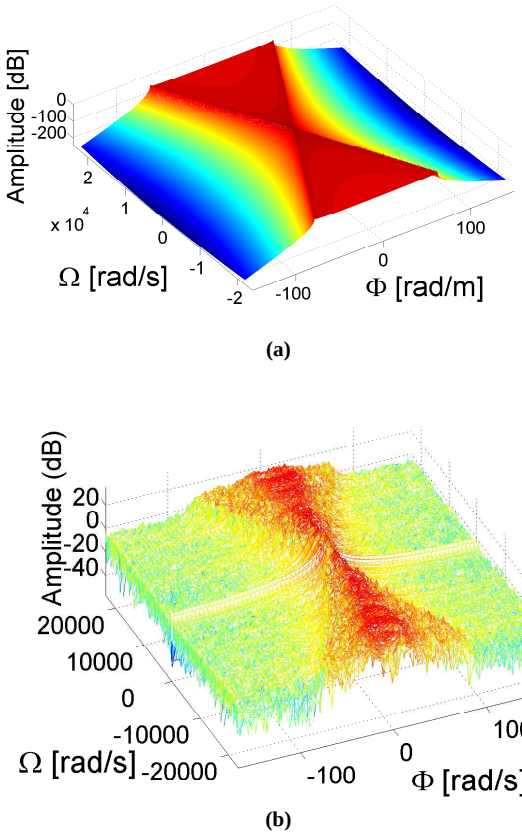

Figure 2: Two-dimensional Fourier transform of the PAF. (a) Obtained from the theory. (b) Obtained from experimental measurements.

rate versus reconstruction signal-to-noise ratio (SNR), has been derived in [4]. Experimental measurements have also been carried out to confirm the developed theory and a similar bow-tie-like spectrum has been obtained when calculating the spectrum corresponding to the measured room impulse responses. This spectrum is shown in Figure 2(b). Note that any spatio-temporal field governed by the wave equation would result in a similar Fourier spectrum. Typically, in the electromagnetic case, c would then correspond to the speed of light.

When sampling the PAF in the spatio-temporal domain, spectral repetitions appear in the 2D-FT domain as observed in Figures 3(a) and 3(b). Owing to the particular shape of the PAF's spectrum, a better packing can be obtained using quincunx sampling as shown in Figures 3(c) and 3(d). This leads to a gain of factor 2 in the processing and will be proved crucial in the RD analysis provided in the next section.

3. RATE-DISTORTION FUNCTIONS FOR THE LINEAR NETWORK SETUP

The scenario that we consider in this section, consists of a sensor network recording a spatio-temporal acoustic field on an infinite line $\mathscr{V}$, which is generated by sound sources located on a parallel line $\mathscr{U}$ at a distance d from the recording line $\mathscr{V}$. The setup is shown in Figure 4. The sound sources emit an acoustic field $U(x,t)$, which induces a sound field $V(x,t)$ on the recording line through the convolution with a space and time invariant filter $g(x,t)$, which we obtain from equation (2) by setting $t_u = 0$, $x_u = 0$ and

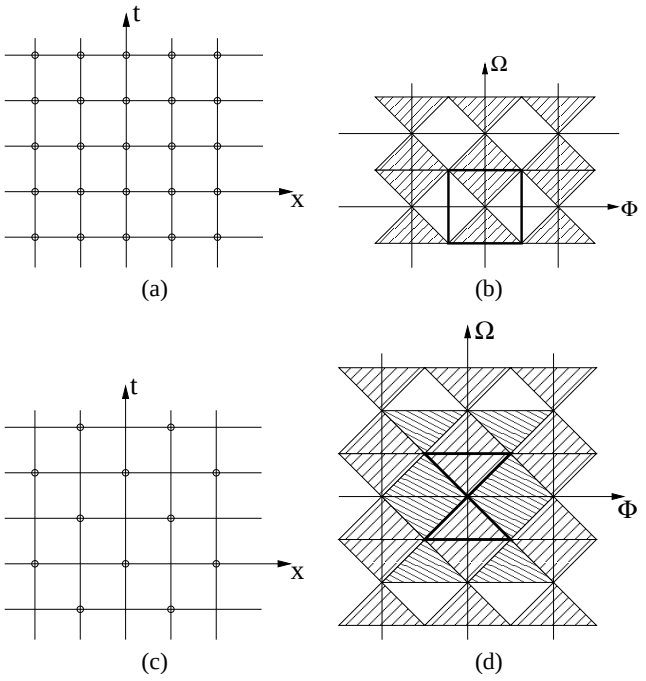

Figure 3: Sampling and interpolation of the PAF. (a) Rectangular sampling grid. (b) Plenacoustic spectrum with its repetitions for a rectangular sampling grid. The interpolation filter is in bold. (c) Quincunx sampling grid. (d) Plenacoustic spectrum with its repetitions for a quincunx sampling grid. The interpolation filter is in bold.

$(y_v - y_u)^2 + (z_v - z_u)^2 = d^2$. Sensors equipped with microphones are equally spaced on the recording line and sample the induced field $V(x,t)$. We model the sound source $U(x,t)$ as a continuous stationary Gaussian stochastic process with a flat and bandlimited power spectral density (PSD) $S_U(\Phi,\Omega)$, i.e.,

$$S_U(\Phi,\Omega) = \sigma_U^2 \, 1_{[-\Phi_0,\Phi_0]}(\Phi) \, 1_{[-\Omega_0,\Omega_0]}(\Omega) \,,$$

where σ_U is some real parameter, and Φ_0 and Ω_0 are the maximal spatial and temporal frequencies. 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

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

Because of the PAF's fast decay in the region where $|\Phi| > |\Omega|/c$, we assume that $G(\Phi,\Omega)$ vanishes in that region. The support of the

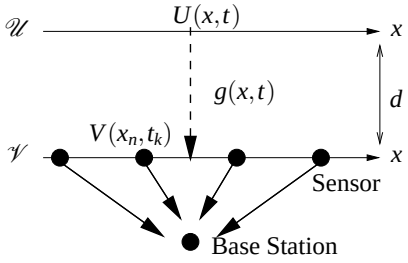

Figure 4: Sound sources located on a line $\mathscr{U}$ emit an acoustic field $U(x,t)$, which induces another sound field $V(x,t)$ on a parallel line $\mathscr{V}$ at a distance d.

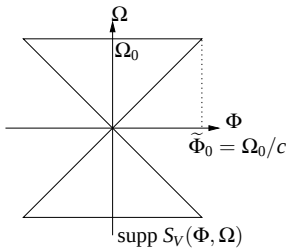

Figure 5: Bow-tie-like spectral support of the PSD $S_V(\Phi,\Omega)$ of the induced acoustic field $V(x,t)$.

PSD $S_V(\Phi,\Omega)$ then has a bow-tie-like shape that is bandlimited on both frequency axes, as it is shown in Figure 5. We observe that the spatial frequency is at most equal to $\widetilde{\Phi}_0 = \Omega_0/c$, independently of the actual maximal spatial frequency Φ_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. In the sequel, we consider the rectangular and the quincunx sampling lattices described in Section 2.

The sensors located on the recording line sample the sound field $V(x,t)$, 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 $\widehat{V}(x,t)$ 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 mean squared error (MSE) per meter and per second, for a given total rate R, measured in bits per meter and per second, spent by the sensors for communicating with the base station. The MSE distortion is defined as

$$D = \lim_{\substack{L\to\infty \\ T\to\infty}} \frac{1}{2L}\frac{1}{2T} \int_{-L}^{L}\int_{-T}^{T} \mathrm{E}\left[\left(V(x,t) - \widehat{V}(x,t)\right)^2\right] \mathrm{d}t\mathrm{d}x \,.$$

For this setup, we determine 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.

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

$$S_V(\Phi,\Omega) = \sigma_V^2 \, \Pi(\Phi,\Omega) \,, \tag{4}$$

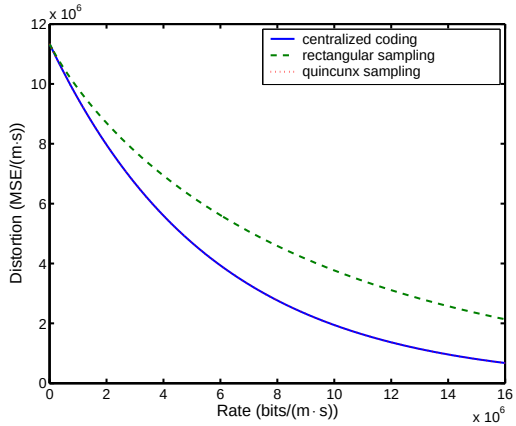

Figure 6: RD functions for different sampling and coding schemes. The curve for quincunx sampling coincides with the one for centralized coding.

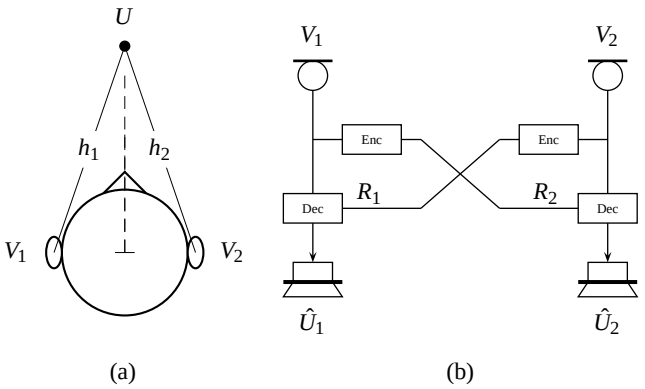

Figure 7: Our hearing aids setup. (a) Typical head-related configuration. (b) Collaboration using a wireless communication link.

where $\Pi(\Phi,\Omega)$ is the indicator function of the bow-tie-like region shown in Figure 5. 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 "water-filling" technique as well as the results for stationary Gaussian random processes in [10]. The results are summarized in the following proposition [5].

Proposition 1 *Under the flat spectrum assumption* (4)*, the RD functions corresponding to the coding schemes defined above and the sampling lattices described in Section 2 are given by the following expressions:*

- *for centralized coding:*

$$R(D) = \frac{\Omega_0^2}{4\pi^2 c}\log\left(\frac{\sigma_V^2\Omega_0^2}{2\pi^2 cD}\right); \qquad (5)$$

- *for rectangular sampling and independent coding:*

$$R(D) = \frac{\Omega_0^2}{2\pi^2 c}\log\left(\frac{\sigma_V^2\Omega_0^2}{e\pi^2 cD}\,\frac{1}{2}\left(1+\sqrt{1-2\frac{\pi^2 cD}{\sigma_V^2\Omega_0^2}}\right)\right) + \frac{\Omega_0^2}{2\pi^2 c}\left(1-\sqrt{1-2\frac{\pi^2 cD}{\sigma_V^2\Omega_0^2}}\right); \qquad (6)$$

- *for quincunx sampling and independent coding:*

$$R(D) = \frac{\Omega_0^2}{4\pi^2 c}\log\left(\frac{\sigma_V^2\Omega_0^2}{2\pi^2 cD}\right); \qquad (7)$$

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

Figure 6 shows the graphs of the RD functions given in Proposition 1. We observe that equations (5) and (7) 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 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(d), 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.

4. GAIN-RATE FUNCTION FOR THE HEARING AIDS SETUP

In this section, the setup consists of two hearing aids, each equipped with an omnidirectional microphone, a processing unit and wireless communication capabilities. For simplicity of exposition, we will concentrate on the simple scenario illustrated in Figure 7(a). The signal observed at microphone k ($k = 1,2$) can be expressed as

$$V_k(t) = U_k(t) + N_k(t) = h_k(t) * U(t) + N_k(t),$$

where U is the point source of interest and N_k some ambient noise. The processes U and N_k are modelled as independent continuous stationary Gaussian stochastic processes with mean zero and PSD S_U and S_{N_k}, respectively. The quantity h_k corresponds to the impulse response of the filter that models the path between the source's position and microphone k. Under the near-field assumption, it is given by the PAF of the source U evaluated at the position of microphone k. It can also account for the shadowing effect induced by the head by means of the corresponding head-related impulse response [11]. As in Section 3, we will work under the far-field assumption. This allows us to provide a closed-form solution to our problem. In this case, $h_k(t)$ is simply given by

$$h_k(t) = \delta(t - \tau_k),$$

where τ_k is the propagation delay from the source U to microphone k. We will further assume that U and N_k have flat PSDs over the frequency band $[-\Omega_0, \Omega_0]$, i.e.

$$\begin{aligned} S_U(\Omega) &= \sigma_S^2\, 1_{[-\Omega_0,\Omega_0]}(\Omega), \\ S_{N_k}(\Omega) &= \sigma_N^2\, 1_{[-\Omega_0,\Omega_0]}(\Omega). \end{aligned}$$

The goal of hearing aid k is to beamform in the direction of U in order to mitigate the effect of surrounding noise. More precisely, it aims at recovering, with minimum MSE, the signal U_k that would have been observed in a noise-free environment. To this end, each device receives a compressed version of its neighbor's acquired signal as depicted in Figure 7(b). In this context, we wish to characterize the best achievable gain, at each hearing aid, that can be provided by the availability of a wireless link as a function of the communication bit-rate. In the sequel, we look at this problem from the perspective of hearing aid 1. Under these assumptions, our setup corresponds to a remote source coding problem with side information at the decoder. For a given rate $R_1 = R$, measured in bits per second, we wish to encode V_2 such as to minimize the MSE between

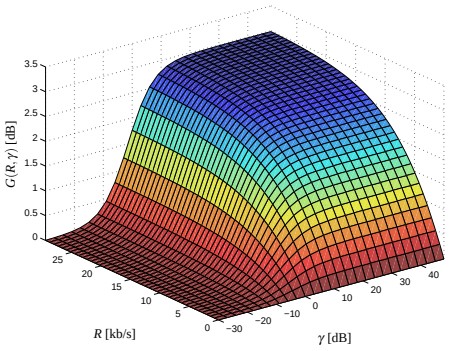

Figure 8: Gain provided by the wireless communication link as a function of the communication rate R and the input SNR γ.

the desired source U_1 and its reconstruction $\hat{U}_1$, assuming the presence of some side information V_1 at the decoder. A general characterization of the corresponding RD function can be found in [7]. For the problem at hand, this RD function evaluates as follows.

Proposition 2 *The RD function under the above assumptions is given by*

$$R(D) = \frac{\Omega_0}{2\pi}\log_2\left(\frac{\sigma_S^2}{\sigma_S^2+\sigma_N^2}\right) - \frac{\Omega_0}{2\pi}\log_2\left(\frac{2\sigma_S^2+\sigma_N^2}{\sigma_S^2\,\sigma_N^2}\frac{\pi D}{\Omega_0} - 1\right),$$

where $D \in (0, (\Omega_0\,\sigma_S^2\,\sigma_N^2)/(\pi(\sigma_S^2+\sigma_N^2))]$.

The gain achieved by this collaborative beamforming is now expressed as a function of R as

$$G(R) = \frac{D(0)}{D(R)}.$$

The function $G(R)$ is referred to as the *gain-rate function*. It characterizes the optimal tradeoff between the communication bit-rate R and the resulting beamforming gain. From Proposition 2, we can straightforwardly provide the following result.

Proposition 3 *The gain-rate function under the above assumptions is given by*

$$G(R) = \frac{2\gamma+1}{\gamma+1}\left(\frac{\gamma}{\gamma+1}\,2^{-2\pi R/\Omega_0}+1\right)^{-1},$$

where $\gamma = \sigma_S^2/\sigma_N^2$ *is the input SNR.*

We plot in Figure 8 the beamforming gain obtained as a function of the communication bit-rate R and the input SNR γ. As $R \to \infty$, the gain remains bounded and corresponds to that of a two-sensor array with no rate constraint. At high SNR, this gain approaches $10\log_{10}(K)$ [dB] where $K = 2$ is the number of sensing devices. We also observe that, in this scenario, the result depends neither on the actual position of the source nor on the geometrical properties of the hearing aids setup. This is due to the far-field assumption and the fact that the noise is uncorrelated across sensors. A similar analysis can be carried in the presence of interfering point sources. In that case, the spatial extent provided by the head becomes crucial. The interested reader is referred to [7] for a more detailed analysis.

5. CONCLUSIONS

In this paper, the spatio-temporal characteristics of the sound field have been studied. We have introduced the plenacoustic function as a means to describe, at any position and time, the sound induced by a given acoustic source. In particular, its bidimensional Fourier spectrum has been computed in the case of an infinite line of microphones and shown to exhibit an almost-bandlimited character. Based on this insight, we have presented sampling results for different sampling lattices. The intuition provided by this analysis has then been applied to two distributed compression scenarios. The first setup has considered the reconstruction of the measured sound field at a central base station. Under restricting assumptions, efficient distributed processing at the sensors has been proved optimal, hence providing an answer to the general multi-terminal source coding problem for this particular scenario. The second setup has looked at a hearing aids problem, where two hearing devices perform collaborative beamforming through a rate-constrained wireless link. The optimal tradeoff between beamforming gain and communication bit-rate has been derived.

REFERENCES

[1] S. Simić and S. Sastry, "Distributed environmental monitoring using random sensor networks," *International Workshop on Information Processing in Sensor Networks*, pp. 582–592, April 2003.

[2] D. Li, K. Wong, Y. Hu, and A. Sayeed, "Detection, classification, tracking of targets in micro-sensor networks," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 17–29, March 2002.

[3] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 30–39, March 2002.

[4] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *to appear in IEEE Transactions on Signal Processing*, 2006.

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

[6] T. Berger, "Multiterminal source coding," *Lectures presented at CISM Summer School on the Information Theory Approach to Communications*, July 1977.

[7] O. Roy and M. Vetterli, "Rate-constrained beamforming for collaborating hearing aids," *IEEE International Symposium on Information Theory*, July 2006.

[8] P. Morse and K. Ingard, *Theoretical Acoustics*. McGraw-Hill, 1968.

[9] T. Ajdler, L. Sbaiz, A. Ridolfi, and M. Vetterli, "On a stochastic version of the plenacoustic function," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2006.

[10] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ:Prentice-Hall, 1971.

[11] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *Journal of the Acoustical Society of America*, vol. 5, no. 104, pp. 3048–3058, November 1998.