# The Ray Space Transform: A New Framework for Wave Field Processing

Lucio Bianchi, *Member, IEEE*, Fabio Antonacci, *Member, IEEE*, Augusto Sarti, *Senior Member, IEEE*, and Stefano Tubaro, *Senior Member, IEEE*

The authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria at Politecnico di Milano, 20133 Milano, Italy (e-mail: lucio.bianchi@polimi.it; fabio.antonacci@polimi.it; augusto.sarti@polimi.it; stefano.tubaro@polimi.it).

***Abstract*—Soundfield imaging is a special analysis methodology aimed at capturing the directional components of the acoustic field and mapping them onto a domain called "ray space", where rele- vant acoustic objects become linear patterns, i.e., sets of collinear points. This allows us to overcome resolution issues while easing far-field assumptions. In this paper, we generalize this concept by introducing the ray space transform for acoustic field repre- sentation. The transform is based on a short space-time Fourier transform of the signals captured by a microphone array, using discrete Gabor frames. The resulting transform coefficients are parameterized in the same ray space used for soundfield imaging. The resulting transform enables the definition of analysis and syn- thesis operators, which exhibit perfect reconstruction capabilities. We show examples of applications of the ray space transform to source localization and spot spatial filtering.**

***Index Terms*—Array signal processing, wave field processing, sound field analysis, near-field spatial filtering, plenacoustic func- tion, space-time analysis, source localization.**

## I. Introduction

THE need of developing new and effective methodologies for sound-field analysis and processing has been steadily growing for the past few decades and has become crucial in application domains such as multimedia communications [1, Chap. 3], noise control [2] and the study of vibrating structures [3]. The initial focus of research was on localizing, tracking, extracting and identifying sound sources in an acoustic scene.

Spatial information of the acoustic field is usually captured with microphone arrays [4]. State-of-the-art microphone array processing techniques can be classified into two broad categories. The first category is based on an *omnidirectional* analysis of array data, which means that no directional components of the wave field are considered. These techniques capture *local* information of the acoustic field by analyzing the impulse responses at multiple microphone locations [5]. Prominent examples include techniques for noise reduction [6]–[14], multichannel acoustic echo cancellation [15], channel identification [16], source localization [17].

The second category of array processing methodologies, on the other hand, relies on the analysis of directional components of the acoustic wave field, extracted through a joint analysis of microphone signals, with reference to a specific wave-field decomposition. For example, if the source is located at a sufficiently large number of array lengths from the microphones (far field assumption), wavefronts can be safely considered as planar and the underlying data model can explicitly rely on plane-wave decomposition. This enables the assumption of a space-invariant impulse response over the spatial extension of the array. Prominent examples of techniques relying on this assumption are [18]–[24]. However, performance degradations occur when the far-field assumption is not fully satisfied, as it can happen in many real-world scenarios.

Manuscript received March 3, 2016; revised June 3, 2016; accepted June 27, 2016. Date of publication July 14, 2016

A more general setting than the two above was proposed in [25], with the introduction of the *plenacoustic function* $f(\mathbf{r}, \mathbf{k})$, defined (in a time-harmonic scenario) as a function of the observation position $\mathbf{r}$ and the spatial wavenumber vector $\mathbf{k}$. As the spatial wavenumber vector for propagating waves can be interpreted as a direction of propagation, this function can be interpreted as the complex amplitude of an acoustic directional component traveling through point $\mathbf{r}$. This mathematical setting, already exploited in the optical domain [26], [27], was introduced in [25], [28], [29] with the goal of deriving a sampling condition. Mignot *et al.*, [30] later revisited the sampling problem of the plenacoustic function in the light of compressive sampling.

Later, Pinto and Vetterli [31], [32] proposed a time varying representation of acoustic signals. These works introduce a representation that is approximated for sources that do not fully satisfy the far-field condition, i.e. sources that are closer than $r = 2\Delta^2/\lambda, r$ being the source distance, $\Delta$ the length of the array and $\lambda$ the wavelength [33]. The representation proposed in [31], [32] enabled the authors to approximate the acoustic field produced by sources in the near field by far field components, which facilitates the estimation of the Direction-of-Arrival of acoustic sources and enables spatial filtering [34].

The idea of approximating fields produced by nearby sources with far field components was later revised and put into work with the introduction of the ray space as a domain of definition, both in Euclidean [35] and in projective form [36]. This led to soundfield imaging methods that enable the effective encoding and the decoding of "local" information in the acoustic field. Operatively, the soundfield image is estimated in [35] by subdividing the array into smaller portions and by estimating the amplitude of directional components for each portion of the array through beamforming. Through a nonlinear mapping of the beamforming power output onto the *ray space* [37], all acoustic primitives become linear features, thus enabling super-resolution methods for acoustic scene analysis, such as:

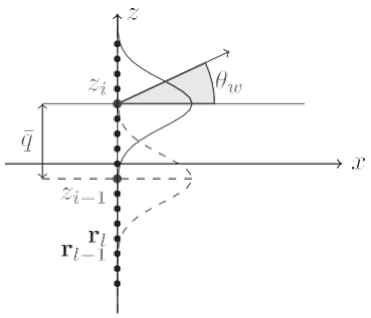


Fig. 1. Microphone array, local analysis points and directions. Array data are partitioned into subgroups, weighed by a spatial window, and modulated in order to phase-align acoustic contributions from directions $\{\theta_w\}$.

localization of sources or reflectors [35]; source separation [38]; "sharpening" of plenacoustic images [39]; and extraction of other spatial information [40]; all accomplished with pattern analysis techniques. Also worth noticing is the fact that the definition of ray space is Euclidean when working with a single microphone array [35] but becomes projective when multiple arrays are being used [41].

The methods in [35], [41] ease the far field hypothesis by basically scaling it down. Sources no longer need to be in far field with respect to the length $\Delta$ of the global array, but only with respect to the subarray size $\Delta_1 < \Delta$. Considering each subarray, the methods keep relying on the plane-wave decomposition. This limits the performance of spatial filtering, which degrades at low-frequencies and when the source distance becomes comparable to the size of the subarrays. In this article we focus on overcoming this limitation by generalizing and formalizing the soundfield analysis introduced in [35], [41] through the adoption of discrete Gabor frames [42]–[48] in the space-time domain. More specifically, the wave field decomposition that we define relies on a new (overcomplete) basis of wave field functions of both temporal and spatial local validity. We will discuss this decomposition and how to guarantee its invertibility using a countable number of components.

We perform local Fourier analysis through a computation of the similarity between the array data and shifted and modulated copies of a prototype spatial window applied to the microphone data, as shown in Fig. 1. This turns out to be equivalent to a beamforming that extracts directional components on a predefined set of directions, but without the far field and frequency limitations discussed above. Our goal, however, is not just to generalize soundfield imaging: we want to define a linear operator $\mathbf{\Psi}$ that maps array data $\mathbf{u}(\omega)$ onto the ray space and, using Gabor signal expansions, define its inverse $\tilde{\mathbf{\Psi}}$ for a perfect reconstruction of array data. This way, as shown in Fig. 2, we can do the processing directly in the transformed domain (ray space), with no loss of information.

This newly-defined *ray space transform* exhibits several important properties: 1) it leads to an inherently *local* representation, as only the coefficients in the proximity of an observation point actually contribute to the acoustic field in that location; 2) it is invertible, which means that working in the transformed domain (ray space) can be done in a lossless fashion; 3) the


Fig. 2. Analysis, synthesis and processing operations in the ray space.

ray space parameterization bears strong geometric insight on the spatial properties of the acoustic field (i.e., spatial position of the acoustic sources, etc.) and preserves all the properties of soundfield imaging.

The ray space transform, however, goes beyond simple compatibility with soundfield imaging. We will show this through a more general application of processing in the transformed domain (in the ray space). We will introduce a "spot" spatial filtering as a form of windowing in the ray space, which leaves selected spatial contributions as undistorted (i.e., sound coming from specific locations in space), while attenuating others. In particular, we will show an example of ray-space filtering based on a minimum mean-squared error design criterion. The design of the spot spatial filter is intuitive in the domain of the ray space transform, whereas other techniques require optimized approaches [49], [50]. We remark that the examples of application proposed in Section V are intended with the mere purpose of proving the generality and the flexibility of the proposed framework in various scenarios.

It is worth noticing that the Gabor expansion, already quite novel for acoustic signal analysis, has been widely used for the representation of wave fields in other contexts, ranging from optical imaging [51], [52], through antenna arrays [53], [54] and seismic signal processing [55]–[57]. Thanks to its generality, we believe that the Ray Space Transform proposed in this paper can find application also in these fields.

The rest of the manuscript is organized as follows. Section II describes the signal model and goes over the necessary background information. Section III introduces the ray-space transform, first in the case of a continuous aperture; then for a discrete array of sensors. In this second case, the transform is defined as a matrix multiplication. The inverse transform is then derived, based on the discrete Gabor dual frame [58], [59], both for the continuous and the discrete setting. In Section IV we show the relation between the proposed ray space transform and the soundfield mapping introduced in [35]. Section V features two case studies, where the properties of the ray space representation are exploited for localization and spot spatial filtering. Finally, Section VI draws conclusions.

## II. Signal Model and Background

### A. Notation

In this manuscript we adopt the following notation:

- We use lowercase letters for scalars, boldface lowercase letters for column vectors; boldface uppercase letters for matrices.
- Superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote transposition, conjugation, and conjugate transposition, respectively.
- $\langle \cdot, \cdot \rangle$ denotes inner product, while $\|\cdot\|$ denotes $L_2$ norm.
- $j$ is the imaginary unit ($j = \sqrt{-1}$).

- $[\mathbf{A}]_{i,w}$ is the element in row $i$ and column $w$ of the matrix $\mathbf{A}$. $[\mathbf{A}]_{i,:}$ is the $i$th row of matrix $\mathbf{A}$.
- $E\{\cdot\}$ denotes statistical expectation.

### B. Signal Model

We consider an acoustic source, in spatial coordinates $\mathbf{r}'$, and a microphone on the $z$ axis in position $\mathbf{r} = [0, 0, z]^T$. A system of this sort is generally characterized by a space-time varying acoustic impulse response. A common way to deal with time-varying acoustic impulse responses is to adopt a short-time processing approach [60]–[62], where signals are temporally spliced and processed in a frame-by-frame fashion, and pick a length of our temporal segments that is short enough that the acoustic impulse response will be time-invariant within the frame itself. This way the output of the system for a single frame can be written in the temporal frequency domain as

$$p(z, \omega) = h(z|\mathbf{r}'; \omega)s(\omega), \tag{1}$$

where $s(\omega)$ denotes the temporal Fourier transform of the signal emitted by the source in $\mathbf{r}'$ and $h(z|\mathbf{r}'; \omega)$ is the acoustic transfer function between the source and a microphone in $z$. In what follows we assume a free-field propagation model, in which

$$h(z|\mathbf{r}', \omega) = \frac{e^{-j\omega\|\mathbf{r}-\mathbf{r}'\|/c}}{4\pi\|\mathbf{r}-\mathbf{r}'\|}, \tag{2}$$

$c$ being the speed of sound. The extension of this model to accommodate multiple acoustic sources is quite straightforward.

The output of the acoustic system for an array with $L$ microphones is given as the multidimensional generalization of (1)

$$\mathbf{p}(\omega) = [p(z_0, \omega), \ldots, p(z_{L-1}, \omega)]^T \in \mathbb{C}^{L\times 1}, \tag{3}$$

where $z_l, l = 0, \ldots, L-1$, is the location of the $l$th microphone. Eq. (3) can be rewritten as $\mathbf{p}(\omega) = \mathbf{h}(\omega)s(\omega)$, $\mathbf{h}(\omega)$ being the propagation vector

$$\mathbf{h}(\omega) = [h(z_0|\mathbf{r}'; \omega), \ldots, h(z_{L-1}|\mathbf{r}'; \omega)]^T. \tag{4}$$

In general, signals received by a microphone array are affected by additive sensor noise. For this reason, we introduce the following model for microphone signals

$$\mathbf{u}(\omega) = \mathbf{p}(\omega) + \mathbf{v}(\omega) = \mathbf{h}(\omega)s(\omega) + \mathbf{v}(\omega), \tag{5}$$

where $\mathbf{v}(\omega)$ is an additive noise term with zero mean and covariance matrix $\Phi_{\mathbf{vv}} = \varphi_v \mathbf{I}_L$, $\mathbf{I}_L$ being the identity matrix of size $L \times L$ and $\varphi_v$ the noise variance at a single microphone.

In Sections III and IV we focus just on the deterministic part of the microphone signals, $\mathbf{p}(\omega)$. The model (5), including the noise term, will be adopted when introducing the examples of application of Section V.

### C. Directional Plenacoustic Function

Let us review the concept of directional plenacoustic function, as introduced in [41, Sec. II]. A useful representation for the signal in (1) consists in expressing it as an inverse Fourier transform with respect to the wavenumber vector $\mathbf{k} = [k_x, k_z]^T = \omega/c[\cos(\theta), \sin(\theta)]^T$ [41], where $\theta$ is the direction of propagation. The integral is taken over the propagating plane wave components

$$p(z, \omega) = \frac{1}{2\pi}\int_0^{2\pi} P(\theta, \omega)e^{j\omega z\cos(\theta)/c}\, d\theta. \tag{6}$$

This expression embodies the classical plane-wave decomposition of the acoustic wavefield, as $P(\theta, \omega)$ encodes the amplitude and phase of each plane-wave component. As discussed in [41], the *directional plenacoustic function* is formally defined as the integrand in (6)

$$\phi(z, \theta, \omega) := e^{jkz\sin(\theta)}P(\theta, \omega). \tag{7}$$

The estimation of $P(\theta, \omega)$ can be accomplished through a beamforming operation [31], [63], [64]. The function $P(\theta, \omega)$ is usually referred to as *Herglotz Density* in the literature [65, p. 54]. We remark that the representation (6) is valid only if the acoustic field is generated by far-field components [66].

Previous works in [35], [41] are based on the definition (7) of the directional plenacoustic function, where $\phi(z, \theta, \omega)$ is estimated through multiple beamforming operations. Conceptually, the estimation process can be divided into three steps: 1) the array is subdivided into (possibly overlapped) subarrays; 2) multiple beamforming is performed on subarray data to estimate $P(\theta, \omega)$; 3) taking the subarray center as a reference point, phase functions are associated to $P(\theta, \omega)$ to yield the estimate of $\phi(z, \theta, \omega)$ according to (7). This approach allows a local analysis without having to give up the plane-wave model. However, the approaches in [35], [41], [67] suffer from two limitations: 1) the reconstruction of the acoustic field from the plenacoustic function as described in [67] is not "perfect", as it only approximates the inverse of the analysis operation; 2) using a global representation for local analysis, as done in [35], [41], [67], introduces some performance limits in terms of source bandwidth and distance. Despite these limitations, the methods described in [35], [41] prove quite effective thanks to the adoption of a highly structured domain for the directional plenacoustic function: the *ray space*.

### D. The Ray Space and the Soundfield Map

In [35] a ray, or equivalently a planar wavefront, is described by the parameters of the line that it lies on. In the setting of Fig. 3 (left-hand side), acoustic rays sensed by the microphone array (lying on the $z$ axis) are oriented lines that cross the $z$ axis within the region occupied by the array (between $-q_0$ and $q_0$). The linear equation

$$z = mx + q, \tag{8}$$

parameterized by $(m = \tan(\theta), q)$ describes any line, except for those that are parallel to the $z$ axis. The space of such parameters is called *ray space*. Rays modeled this way are assumed to be coming from the half-space $x > 0$, which establishes an equivalence between acoustic rays and lines. The region of the ray space that corresponds to the rays that are *visible* from the microphone array is

$$V = \{(m, q) \in \mathbb{R}\times\mathbb{R} : -q_0 \le q \le q_0\}, \tag{9}$$

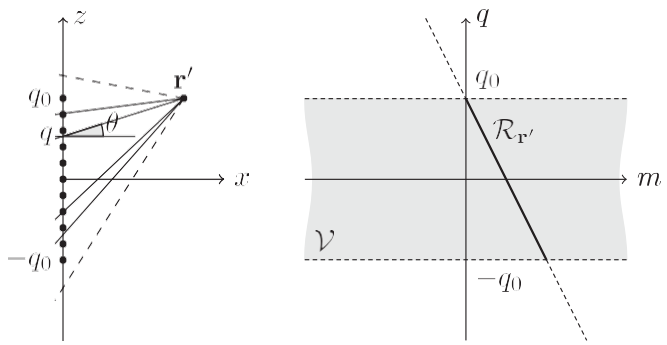


Fig. 3. Ray space representation of a point source [35, Fig. 1]. Acoustic rays emitted by the source and sensed by the array are drawn with solid lines. Rays that are not sensed by the array appear as dashed lines. Rays produced by the source and falling on the array form the region of interest $\mathcal{R}_{\mathbf{r}'}$, a line in the ray space.

and is therefore called *visibility region* [37].

In order to gain more insight on how the acoustic information is structured in the ray space, let us consider the ray space representation of a point source at $\mathbf{r}' = [x', y' = 0, z']^T$, $x' > 0$. In geometrical acoustics, the field emitted by the source is described by rays belonging to the bundle of lines crossing at $\mathbf{r}'$ [68, p. 26], [69, Chap. 11] (i.e., such that they are always perpendicular to the wavefronts). The parameters $(m, q)$ of this bundle of lines must satisfy the linear equation $q = -x'm + z'$. The region of the ray space corresponding to rays that are emitted by the source (in $x' > 0$) and are sensed by the microphone array is

$$\mathcal{R}_{\mathbf{r}'} = \{(m, q) \in \mathbb{R} \times \mathbb{R} : q = -x'm + z',\ -q_0 \le q \le q_0\}, \tag{10}$$

as shown in Fig. 3. In [35] the authors map the energy of rays in the ray space to obtain the *soundfield map*. We remark that more complex acoustic fields can be obtained as a superposition of fields generated by point sources [70, p. 31].

One advantage in mapping acoustic data onto the ray space is that acoustic primitives (point sources, reflectors, etc.) are all mapped onto the ray space as linear patterns, therefore many acoustic signal processing problems can be solved using pattern analysis techniques. For example, the localization of multiple acoustic sources can be accomplished by estimating the parameters of multiple lines [35]; the extraction of a desired source signal can be addressed by estimating the location of the source and then performing a spatial filtering that accounts for the location of the source [38]; other applications, such as the estimation of the radiance pattern of a source [39], [40] are accomplished by analyzing the soundfield map.

In Section III we will introduce the ray space transform and we will adopt the ray space parameters $m$ and $q$ as the parameters of the transform itself. We remark that this choice does not imply that the ray space transform is only valid under the geometrical acoustics approximation. To enforce this, in Section IV we include a simulation showing that the proposed ray space transform exhibits perfect reconstruction even when the geometrical acoustics approximation does not hold.

### *E. Local Fourier Analysis*

This paragraph introduces the local Fourier transform as a signal processing tool to cope with signals whose frequency content is only locally static [42], [60], [71]–[73]. Here we review local Fourier analysis over the time domain. The generalization to time-space is approached in the next section.

The *local Fourier transform* of a signal $g(t), t \in \mathbb{R}$ is defined as the discrete Gabor expansion [44], [45]

$$[\mathbf{G}]_{i,w} = \int_{\mathbb{R}} g(t)\psi^*_{i,w}(t)\,dt, \tag{11}$$

where

$$\psi_{i,w}(t) = \psi(t - iT)e^{j\frac{2\pi}{W}w(t-iT)}, \tag{12}$$

$\psi(t)$ is the (real-valued) analysis window, $i \in \mathbb{Z}$ is the time frame index, $w = 0, \ldots, W-1$ is the frequency band index, $W$ is the number of frequency bins, and $T$ is the window's hop size. The *local inverse Fourier transform* enables the reconstruction of $g(t)$ from its discrete local Fourier representation $\mathbf{G}$ as

$$g(t) = \sum_{i} \sum_{w=0}^{W-1} [\mathbf{G}]_{i,w}\, \tilde{\psi}_{i,w}(t), \tag{13}$$

where

$$\tilde{\psi}_{i,w}(t) = \tilde{\psi}(t - iT)e^{j\frac{2\pi}{W}w(t-iT)}, \tag{14}$$

$\tilde{\psi}$ being the (real-valued) synthesis window. Any signal in $L_2$ can be perfectly reconstructed through (13) if the completeness condition

$$\sum_{i} \psi(t - iT)\tilde{\psi}(t - iT) = \frac{1}{W} \qquad \forall t \in \mathbb{R} \tag{15}$$

is satisfied. The windows $\psi$ and $\tilde{\psi}$, along with the parameters $T$ and $2\pi/W$ form a pair of dual discrete Gabor frames [58], [59]. In this view, $[\mathbf{G}]_{i,w}$ can be interpreted as the similarity between the shifted and modulated Gabor prototype function $\psi_{i,w}$ and $g(t)$. The discrete-time versions of all these expansion can be readily obtained through temporal sampling and by replacing the integral with a summation.

## III. Ray Space Transform

This section introduces the *ray space transform* as a structured redundant representation, based on the local Fourier transform. In the following we show that this analysis of acoustic signals reveals important characteristics of the acoustic field; more specifically, we show how to estimate directional components of the acoustic field in the vicinity of an observation point.

With the intent of adopting the ray space as the domain for our transformation, we parametrize directions $\theta$ by $m = \tan(\theta)$. With this parametrization, the phase shift at position $z$ due to a directional contribution from $\theta$ is given by

$$z\sin(\theta) = z\sin(\arctan(m)) = \sqrt{\frac{zm}{1+m^2}}, \tag{16}$$

where the first equality is obtained by substituting $\theta = \arctan(m)$ and the second equality follows from known interrelations among trigonometric functions [74, Tab. 4.16.3].

We will incorporate this expression for the phase shift into the spatial frequency analysis operation.

In the following, we adopt an uniform grid for sampling the $(m, q)$ plane, and we denote by $\bar{q}$ and $\bar{m}$ the sampling intervals on the $q$ and $m$ axes, respectively. In particular, we set $q_i = (i - (I-1)/2)\bar{q}$, $i = 0, \ldots, I-1$ and $m_w = (w - (W-1)/2)\bar{m}$, $w = 0, \ldots, W-1$, $I$ and $W$ being the number of samples on the $m$ and $q$ axes, respectively.

We consider all the acoustic quantities to be defined over a short time domain. For the sake of notational compactness, unless otherwise specified, we omit the temporal dependency.

### A. Ray Space Transform of a Continuous Aperture

In this paragraph we consider the case of a continuous pressure-sensitive aperture (an ideal continuous array) deployed along the $z$ axis between $z = -q_0$ and $z = q_0$. The aperture captures the acoustic field $p(z, \omega)$, $-q_0 \leq z \leq q_0$. This allows us to derive analytic expressions for the ray space transform of simple examples of acoustic fields. In the following, we consider the Gaussian function

$$\psi(z) = e^{-\pi \frac{z^2}{\sigma^2}}, \qquad \sigma \in \mathbb{R} \tag{17}$$

as the spatial window at the basis of our local Fourier analysis. The scalar $\sigma$ controls the width of the Gaussian window.

The *ray space transform* is defined as the local Fourier transform of the aperture data and it is obtained by computing the similarity between the captured acoustic field $P(z, \omega)$ and shifted and modulated copies of the window function (11). In the continuous setting considered in this section, it takes on the form

$$[\mathbf{Z}_{i,w}(\omega) = \int_{-q_0}^{q_0} p(z, \omega) e^{-\jmath k z \frac{m_w}{\sqrt{1+m^2}}} \psi^*_{w,i}(z)\, dz, \tag{18}$$

where $k = \omega/c$ denotes the magnitude of the wavenumber vector. The integer $w = 0, \ldots, W-1$ spans the grid of spatial frequencies, while the integer $i = 0, \ldots, I-1$ spans the spatial displacement along the $z$ axis. Notice that the array has, inevitably, a limited extension, therefore it is bounded within $-q_0$ and $q_0$. This means that, in addition to the moving spatial window (17), there is a fixed rectangular window of length $2q_0$.

The *inverse ray space transform* is defined as the inverse local Fourier transform of the ray space coefficients (cfr. (13))

$$\tilde{p}(z, \omega) = \sum_{i=0}^{I-1} \sum_{w=0}^{W-1} [\mathbf{Z}_{i,w}(\omega) e^{\jmath k z \frac{m_w}{\sqrt{1+m_w^2}}}] \tilde{\psi}_{i,w}(z), \tag{19}$$

where $\psi_{i,w}(z)$ is the synthesis window as defined in the context of (13). In Section III-B we will describe how to compute the synthesis window $\tilde{\psi}(z)$ given an analysis window $\psi(z)$.

*1) Interpretation:* The interpretation of (18) is immediate upon considering a single spatial window, i.e., upon fixing $i$. In this case, (18) is interpreted as the beamforming operation [75, Sec. 6.3.1] applied to the aperture data that have previously been weighed by a Gaussian spatial windowing function centered at $q_i$. In the light of this interpretation, $[\mathbf{Z}]_{i,:}$ provides a collection of the outputs of multiple beamforming operations, each computed from a specifically weighed portion of the aperture data.

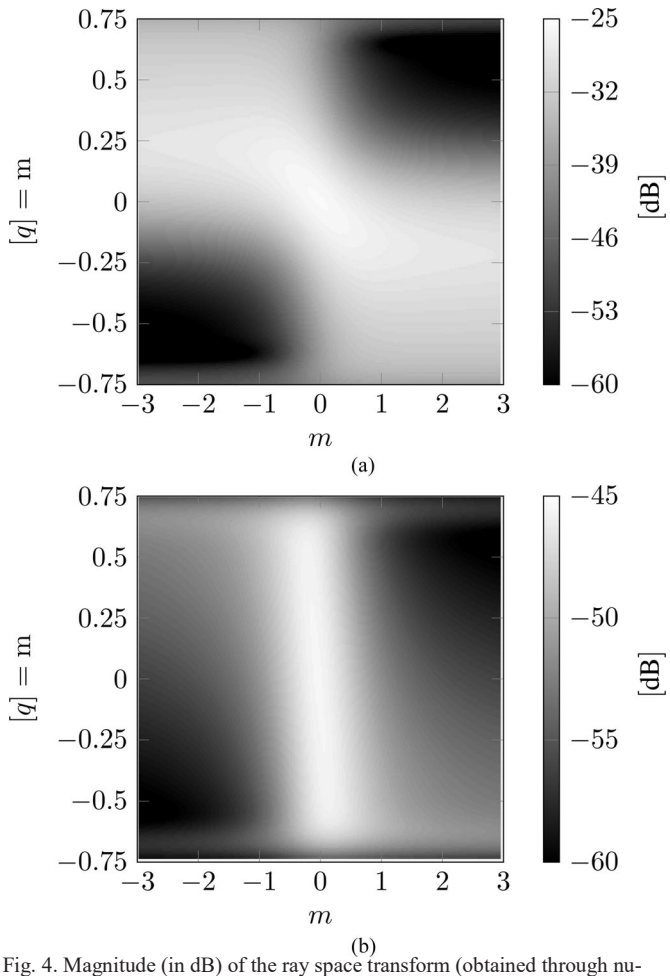


Fig. 4. Magnitude (in dB) of the ray space transform (obtained through numerical integration) of the acoustic field produced by a point source in two different positions, with $\bar{m} = 0.03$, $\bar{q} = 7.5$ mm, $\omega = 2\pi 1$ kHz, $c = 340$ ms$^{-1}$, $\sigma = 0.2$ m, and $z_0 = 0.75$ m. (a) $x^J = 0.3$, $z^J = 0$. (b) $x^J = 3$, $z^J = 0$.

In order to help clarify this interpretation, let us show an example of application of the ray space transform to a spherical wave acoustic field.

2) *Ray Space Transform of a Spherical Wave Acoustic Field:* Consider the acoustic field generated by a point source at position $\mathbf{r}^J = [x^J, 0, z^J]^T$ and observed on the continuous aperture defined on the $z$ axis between $z = -q_0$ and $z = q_0$. The observed acoustic field can be written as [3, p. 198]

$$p(z, \omega) = \frac{e^{-\jmath k \sqrt{x^{J2} + (z^J - z)^2}}}{4\pi \sqrt{x^{J2} + (z^J - z)^2}}. \tag{20}$$

The ray space transform of the acoustic field generated by a point source is obtained as a solution of the integral that we obtain by inserting (20) into (18). To the best of our knowledge, no closed form solution to such integral exists; for this reason, the following results are obtained through numerical integration in MATLAB, using the global adaptive quadrature method in [76].

Fig. 4 shows the magnitude (in dB) of the ray space transform, obtained through numerical integration in MATLAB, of the

acoustic field produced by a point source on the $x$ axis at 0.3 m (Fig. 4a), and at 3 m (Fig. 4b) from the array, respectively.

Similarly to what happens in soundfield imaging [35] the ray space transform maps a source onto a linear pattern. The slope and the intercept of that line are the coordinates of the point source [35].

### B. Ray Space Transform of a Discrete Array

Let us now consider a uniform microphone array of $L$ sensors on the $z$ axis, between $-q_0$ and $q_0$. Let us denote by $d$ the distance between adjacent microphones, so that the $l$th microphone will be in $z_l = (l - (L-1)/2)d$. After discretizing (18) with respect to $z$ we obtain

$$[\mathbf{Z}]_{i,w}(\omega) = d \sum_{l=0}^{L-1} p(z_l, \omega) e^{-j\frac{\omega z_l m_w}{1+m_w^2}} e^{-\frac{\pi (z_l - q_i)^2}{\nu^2}}. \quad (21)$$

The *discrete ray space transform* can be conveniently written in matrix form. For this purpose, we introduce the matrix $\mathbf{\Psi}(\omega) \in \mathbb{C}^{L \times IW}$ whose element in row $l$ and column $(i + wI + 1)$ is given by

$$[\mathbf{\Psi}]_{l,i+wI+1} = e^{j\frac{\omega z_l m_w}{1+m_w^2}} e^{-\frac{\pi (z_l - q_i)^2}{\nu^2}} d, \quad (22)$$

for $l = 0, \ldots, L-1$. We also introduce the *canonical dual matrix* $\tilde{\mathbf{\Psi}} \in \mathbb{C}^{IW \times L}$ [58] corresponding to the pseudo-inverse of $\mathbf{\Psi}$:

$$\tilde{\mathbf{\Psi}} = \mathbf{\Psi}\mathbf{\Psi}^H \left(\mathbf{\Psi}\right)^{-1}. \quad (23)$$

We finally introduce the vector $\mathbf{z} \in \mathbb{C}^{IW \times 1}$ obtained by rearranging the elements of $\mathbf{Z}$ as

$$[\mathbf{z}]_{i+wI+1} = [\mathbf{Z}]_{i,w}. \quad (24)$$

We can now write the discrete ray space transform as

$$\mathbf{z} = \mathbf{\Psi}^H \mathbf{p}. \quad (25)$$

By exploiting the properties of dual frames, we can also write the inverse of the ray space transform as

$$\mathbf{p}^{(\mathbf{Z})} = \tilde{\mathbf{\Psi}}^H \mathbf{z}. \quad (26)$$

In order for $\mathbf{p}^{(\mathbf{Z})}$ to represent a perfect reconstruction formula, we need the decomposition to be either complete or overcomplete, i.e., $IW \geq L$. For reasons that will be clearer later on, however, we will normally operate in conditions of overcompleteness, in which case there exist an infinite number of matrices that could play the role of dual matrices to $\mathbf{\Psi}$. The choice of the canonical dual, which leads to the pseudo-inverse of $\mathbf{\Psi}$ is the one that guarantees the coefficients to have minimum norm [58].

## IV. Relation with the Soundfield Map

This section investigates the relation between the ray space transform introduced in Section III and the soundfield map introduced in [35], [41], [67] and reviewed in Section II. We proceed in the following way: 1) we derive the expression for the coefficients of the soundfield map; 2) we derive the inversion formula for the soundfield map, that will express the observed acoustic field on the array given the soundfield map coefficients; 3) we compare the synthesized fields using soundfield map and ray space transform frameworks.

### A. Soundfield Map Analysis and Synthesis

We consider the soundfield map of a point source as defined in [35], where the length of each subarray (see Section II-C) is assumed as small compared to the distance of the acoustic source; and a far field propagation model can thus be adopted. The acoustic field observed by the $i$th subarray can therefore be written as

$$f_i(z, \omega) = e^{jkz \sin(\theta_i^J)}, \quad \text{for} \quad z \approx q_i \quad (27)$$

where $\theta_i^J$ is the angle under which the source is observed by the $i$th subarray. Coefficients in the soundfield map $\mathbf{F}(\omega) \in \mathbb{C}^{I \times W}$ are then obtained through beamforming operations, each steered towards a direction $\theta_w$. By adopting the ray space mapping (16), we obtain

$$[\mathbf{F}]_{i,w}(\omega) = d \sum_{l=0}^{L-1} e^{jkz\Theta_{i,w}} \mathrm{rect}\left(\frac{z - q_i}{\nu}\right), \quad (28)$$

where $\Theta_{i,w} = \sin(\theta_i^J) - \frac{m_w}{\sqrt{1+m_w^2}} = \sin(\theta_i^J) - \sin(\theta_w)$, and $\mathrm{rect}((z-q_i)/\nu)$ denotes a rectangular function of width $\nu$, centered in $q_i$. From (28) we notice that the soundfield map encodes the position of the acoustic source in the *locus* $q_i = -x^J + m_w + z^J$, which matches the locus that we would obtain through the ray space transform.

Given the soundfield map analysis formula (28), we obtain its inverse as

$$p^{(\mathbf{F})}(z, \omega) = \sum_{i=0}^{I-1} \sum_{w=0}^{W-1} [\mathbf{F}]_{i,w}(\omega) e^{jk\frac{\sqrt{m_w}}{1+m_w}} \widetilde{\mathrm{rect}}\left(\frac{z - q_i}{\nu}\right), \quad (29)$$

where $\widetilde{\mathrm{rect}}(\cdot)$ denotes the dual window of the rect function within the setting of Gabor frames. This dual window $\widetilde{\mathrm{rect}}(\cdot)$ does not exhibit a finite length, therefore it needs to be properly truncated.

### B. Discussion

We compare the normalized mean squared error of the acoustic field from (26) (with $\mathbf{Z}(\omega)$ given in (21)) and the one of (29) (with $\mathbf{F}(\omega)$ given in (28)) with respect to the ideal acoustic field (20). The normalized mean squared errors are computed as

$$\mathrm{NMSE}^{(\mathbf{Z})} = 10 \log_{10} \left( \frac{1}{L} \sum_{l=0}^{L-1} \frac{|p(z_l, \omega) - p^{(\mathbf{Z})}(z_l, \omega)|^2}{|p(z_l, \omega)|^2} \right), \quad (30)$$

$$\mathrm{NMSE}^{(\mathbf{F})} = 10 \log_{10} \left( \frac{1}{L} \sum_{l=0}^{L-1} \frac{|p(z_l, \omega) - p^{(\mathbf{F})}(z_l, \omega)|^2}{|p(z_l, \omega)|^2} \right). \quad (31)$$

Fig. 5 shows $\mathrm{NMSE}^{(\mathbf{Z})}$ and $\mathrm{NMSE}^{(\mathbf{F})}$ as a function of temporal frequency, for a source located at (1 m, 0 m). We observe that the ray space transform is accurate at all temporal frequencies. On the other hand, the soundfield map is not able to accurately

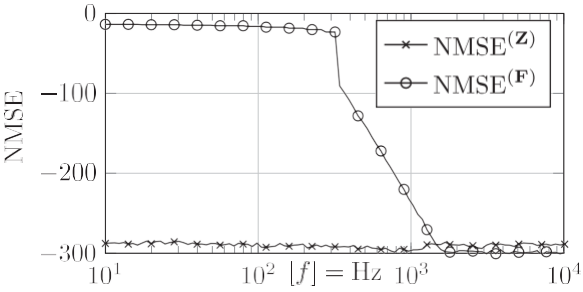


Fig. 5. Normalized mean-squared error of (19) and (29) as a function of temporal frequency. $\bar{m} = 0.03$, $W = 201$. $\bar{q} = 8$ mm, $I = 201$. $c = 340$ ms$^{-1}$, $\sigma = \nu = 0.3$ m, $q_0 = 1$ m. The acoustic source is located at $\mathbf{r}^J = [1\text{ m}, 0\text{ m}]^T$.

represent an acoustic field at mid-low temporal frequencies. We attribute this fact to the mismatch between the selected and the actual dual of the rectangular window. In the applications that include spatial filtering, we are therefore forced to filter out frequencies at which NMSE$^{(\mathbf{F})}$ is not negligible, thus introducing an unwanted limitation of the signal bandwidth.

## V. Examples of Application

We now discuss two examples of acoustic signal analysis and processing in the ray space. These examples are not to be taken as optimized solutions for the problems at hand but rather are presented with the purpose of showing how acoustic analysis and processing problems can be effectively formulated in the ray space domain. In both cases we assume operating conditions to be time-varying, which requires us to perform a short space-time analysis (11) on the microphone signals (5), by subdividing them into time frames. For notational simplicity, time dependence is only made explicit where needed.

### A. Localization

The idea of acoustic source localization in near field conditions based on soundfield image analysis in the ray space was first proposed in [35]. Localization was based on the acoustic field model described in Section II-C, remapped onto the ray space. The same localization approach was adopted in [67]. We now show how the same localization problem can be approached using the discrete ray space transform of Section III-B. The purpose, however, is not to improve localization performance, but to show how the formalism introduced in this manuscript can be used for that purpose.

*1) Methodology:* Localization is performed considering a wideband extension of the ray space coefficients. In particular, given the magnitude of ray space coefficients $|\mathbf{Z}(\omega_a)|$ at discrete temporal frequencies $\omega_a$, the wideband magnitude estimate is defined as

$$\bar{\mathbf{Z}} = \sum_{a=0}^{N/2} E\{|\mathbf{Z}(\omega_a)|\} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}_{1 \times N/2}, \quad \bar{\mathbf{Z}} \in \mathbb{R}^{I \times W}, \tag{32}$$

where $N/2$ is the number of considered frequencies. In the simulations the expectation is approximated by an average over $J = 10$ time frames:

$$E\{\mathbf{Z}(\omega_a)\} = \frac{1}{J} \sum_{j=0}^{J-1} \mathbf{Z}^{(j)}(\omega_a), \tag{33}$$

where $\mathbf{Z}^{(j)}(\omega_a) = \mathbf{\Psi}^H(\omega_a)\mathbf{u}^{(j)}(\omega_a)$ are the ray-space coefficients computed at the $j$th time frame.

Localization is accomplished by exploiting specific features of the ray space, in particular through (10). From $\bar{\mathbf{Z}}$, peaks are identified for each row $\bar{\mathbf{Z}}_{i,:}$ at values $\hat{m}_i$. The position of the acoustic source is estimated through a least-squares regression of the identified peak locations. In particular, let us introduce $\hat{\mathbf{M}} = [-\hat{\mathbf{m}}, \mathbf{1}]$, with $\hat{\mathbf{m}} = [\hat{m}_0, \ldots, \hat{m}_{I-1}]^T$ and $\mathbf{1} = [1, \ldots, 1]^T \in \mathbb{R}^{I \times 1}$; we also introduce the vector $\mathbf{q} = [q_0, \ldots, q_{I-1}]^T$ collecting the positions of the centers of the spatial windows on the $z$ axis. Using these definitions and (10), we can easily derive $\hat{\mathbf{M}}\mathbf{r}^J = \mathbf{q}$. The source position can then be estimated as

$$\hat{\mathbf{r}}^J = \left(\hat{\mathbf{M}}^T \hat{\mathbf{M}}\right)^{-1} \hat{\mathbf{M}}^T \mathbf{q}. \tag{34}$$

*2) Performance measure:* In order to validate the localization approach, we measure the performance with the localization error, defined as the distance between the actual source position and the estimated one:

$$\varepsilon_S^J = \left\| \hat{\mathbf{r}}^J - \mathbf{r}^J \right\|, \tag{35}$$

where $\hat{\mathbf{r}}^J$ is obtained through (34).

*3) Simulation Setup:* For the simulations we use a uniform linear microphone array made of $L = 16$ sensors, with a spacing of $d = 0.1$ m. The temporal sampling frequency is set to $F_s = 8$ kHz. For the source signal $s(\omega)$ we use a complex Gaussian white random process of power spectral density $\varphi_s = 1$, constant for all frequencies. An additive white noise process according to (5) is considered, where the noise power spectral density $\varphi_v$ is constant with frequency.

We remark that the microphone spacing $d$ chosen for the simulations dictates aliasing for frequencies above 1.7 kHz, while the upper frequency limit of all our analysis is $F_s/2 = 4$ kHz.

We introduce the input signal to noise ratio as the ratio between the variance of the source signal received by the first microphone and the variance of the noise on a single microphone. More formally,

$$\text{iSNR}(\omega) = \frac{E\{|h(\mathbf{r}_0|\mathbf{r}^J;\omega)s|^2\}}{\varphi_v} = \frac{|h(\mathbf{r}_0|\mathbf{r}^J;\omega)|^2 \varphi_s}{\varphi_v}. \tag{36}$$

The short time analysis is implemented considering time frames of duration 8 ms (i.e., 64 samples) and the short time Fourier transform is evaluated at frequencies $\omega_a = 2\pi a F_s/(N-1)$, $a = 0, \ldots, N-1$.

*4) Simulations:* Fig. 6 shows the localization performance (35) as a function of the iSNR. For this simulation, the acoustic source is placed at $\mathbf{r}^J = [1\text{ m}, 0\text{ m}]^T$. From a qualitative analysis of Fig. 6 we deduce that the localization is achieved and its performance is stable (in the sense that the localization error $s$

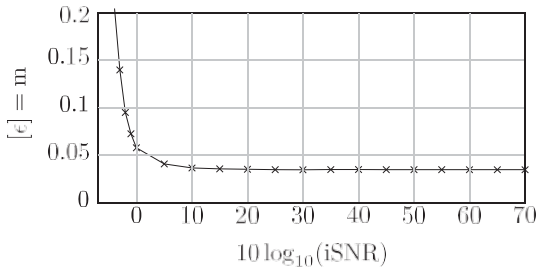


Fig. 6. Localization error $s$ as a function of iSNR. $W = 201$, $\bar{m} = 0.03$. $I = 101$, $\bar{q} = 0.012$ m. $\sigma = 0.3$ m. $c = 340$ ms$^{-1}$ .

is constant) above an input signal to noise ratio iSNR $\geq 0$ dB. Below this threshold, the localization error increases largely. We remark that the choice of the sampling intervals $\bar{m}$ and $\bar{q}$ is rather arbitrary and mainly depends on the specific application scenario. For the simulative scenario considered here, the choice $\bar{m} = 0.03$, $\bar{q} = 0.012$ m gives good results.

Furthermore, we remark that this simulation result is obtained considering also frequencies above the aliasing frequency of the array. Localization results are not affected by aliasing thanks to the property of the ray space mapping, which maps aliased components as quadratic patterns [35]. Those patterns are easily discarded by the methodology adopted in this section.

### B. Filtering in the Ray Space

Let us now consider an application of ray space processing to the case of filtering of acoustic signals, as captured by a microphone array. The filtering operation is designed in order to reduce the effect of sensor noise on the acoustic data.

*1) Filter design:* Given the array model (5), we target an estimate $\hat{s}(\omega)$ of the source signal $s(\omega)$ in the form $\hat{s}(\omega) = \mathbf{b}^H \mathbf{z}(\omega)$, being $\mathbf{z}(\omega) = \mathbf{\Psi}^H(\omega)\mathbf{u}(\omega)$ and $\mathbf{b} \in \mathbb{C}^{IW \times 1}$ a suitable ray space filter. In the following we omit the dependence on $\omega$ for notational convenience.

We adopt a minimum mean square error criterion

$$E\{|s - \mathbf{b}^H \mathbf{z}|^2\}, \tag{37}$$

to derive the optimal filter. The solution to this minimization problem is the *Wiener filter* (e.g., [4, Chap. 6] and references therein)

$$\mathbf{b} = \Phi_{\mathbf{zz}}^{-1} \Phi_{s*\mathbf{z}}, \tag{38}$$

where

$$\Phi_{\mathbf{zz}} = E\{\mathbf{z}\mathbf{z}^H\} = \mathbf{\Psi}^H \Phi_{\mathbf{uu}} \mathbf{\Psi} \tag{39}$$

is the covariance matrix of the transformed array data, and

$$\Phi_{s*\mathbf{z}} = \varphi_s \mathbf{\Psi}^H \mathbf{h} \tag{40}$$

is the cross-covariance between the source signal and the transformed array data. The equality of (40) is valid if the source signal and the additive noise signals are statistically independent.

Equation (38) returns the filter in a form that is not convenient for practical applications, as it requires knowing the source variance, which is not easy to extract from real-world array data. A more practical formulation is obtained by rewriting the transformed array covariance matrix as[1]

$$\Phi_{\mathbf{zz}} = \Phi_{s*\mathbf{z}} \mathbf{h}^H \mathbf{\Psi} + \mathbf{\Psi}^H \Phi_{\mathbf{vv}} \mathbf{\Psi}. \tag{41}$$

In order to obtain $\Phi_{s*\mathbf{z}}$ from (41), we introduce the vector $\mathbf{g}$ as the collection of the inverse propagation functions from microphones to the source location, which, for the temporal frequency $\omega$, can be written as

$$\mathbf{g} = [g(\mathbf{r}_0|\mathbf{r}^J;\omega), \ldots, g(\mathbf{r}_{L-1}|\mathbf{r}^J;\omega)]^T \in \mathbb{C}^{L\times 1}, \tag{42}$$

where

$$g(\mathbf{r}_l|\mathbf{r}^J;\omega) = 4\pi\|\mathbf{r}_l - \mathbf{r}^J\| e^{-\jmath \frac{\omega}{c}\|\mathbf{r}_l - \mathbf{r}^J\|}. \tag{43}$$

With this definition of $\mathbf{g}$, we have $\mathbf{h}^H \mathbf{g} = 1$. Then we right multiply each member of (41) by $\tilde{\mathbf{\Psi}}\mathbf{g}$. Upon considering (from (Section III-B)) that $\mathbf{\Psi}\tilde{\mathbf{\Psi}} = \mathbf{I}_L$ , we can rewrite the filter $\mathbf{b}$ in the following form, which requires the knowledge of the noise variance,

$$\mathbf{b} = \tilde{\mathbf{\Psi}}(\mathbf{I}_L - \Phi^{-1} \Phi_{\mathbf{uu}} \mathbf{v})\mathbf{g}. \tag{44}$$

Notice that the formulation (44) is important for applications as it only relies on the knowledge of the noise variance; which can be accurately estimated from array data (see [77]). Notice also that the proposed filter in (44) requires the knowledge of the source position. As this information can be extracted from array data using the ray space approach presented in Section V-A, in what follows we investigate on the impact of the localization error on the filtering performance.

*2) Performance measures:* We validate the filtering approach by considering the signal to noise ratio at the output of the filter [62], defined as the ratio between the variance of the filtered transformed signal over the variance of the filtered transformed noise, i.e.

$$\text{oSNR} = \frac{E\{|\mathbf{b}^H \mathbf{\Psi}^H \mathbf{p}|^2\}}{E\{|\mathbf{b}^H \mathbf{\Psi}^H \mathbf{v}|^2\}}. \tag{45}$$

*3) Simulations:* In the following, the filter $\mathbf{b}$ in (44) is designed based on the sample estimate of the covariance matrices $\Phi_{\mathbf{uu}}$ and $\Phi_{\mathbf{vv}}$ . In order to regularize the inversion of $\Phi_{\mathbf{uu}}$ , we apply the eigenvalue decomposition and we set to zero those eigenvalues that are smaller than −50 dB [11]. We adopt here the same simulative setup described in Section V-A3.

Fig. 7 shows the output SNR (45) as a function of the temporal frequency $\omega$. For this simulation we assume that the source position $\mathbf{r}^J$ is known with infinite accuracy, so that the filter $\mathbf{b}$ is computed by exploiting this knowledge, while in Section V-B4 we show the impact of the localization error. Results are shown for different values of the input signal to noise ratio iSNR. We notice the the oSNR, for all the iSNR, is constant throughout frequencies, even above the spatial aliasing frequency of the array. Moreover, we notice that the ratio oSNR/iSNR is approximately equal to 15 dB, irrespective of the value of iSNR.

[1]The formulation in (41) is obtained from (39) starting with [75, p. 281] $\Phi_{\mathbf{uu}} = \varphi_s \mathbf{h}\mathbf{h}^H + \Phi_{\mathbf{vv}}$ . This result is transformed through $\mathbf{\Psi}$ and then, after using (40), (41) follows.

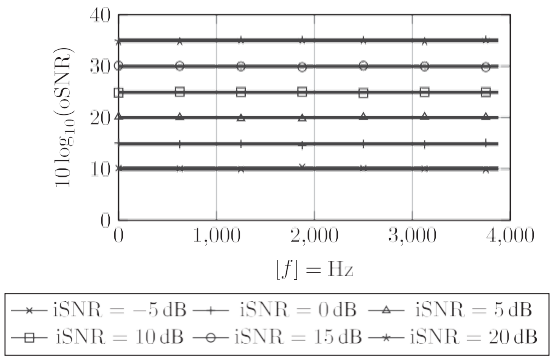


Fig. 7. Output signal to noise ratio oSNR as a function of temporal frequency for different values of the iSNR. $W = 201$, $\bar{m} = 0.03$. $I = 101$, $\bar{q} = 0.012$ m. $\sigma = 0.3$ m. $c = 340$ ms$^{-1}$. The source position $\mathbf{r}^{\prime}$ is assumed to be known.

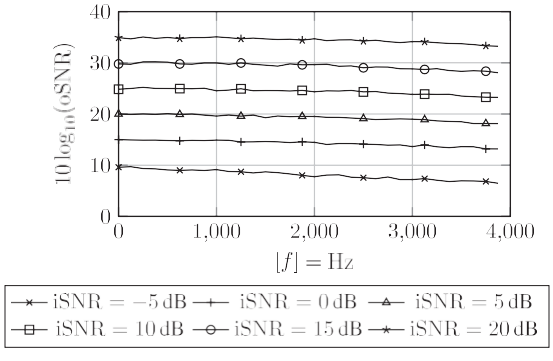


Fig. 8. Output signal to noise ratio oSNR as a function of temporal frequency for different values of the iSNR. $W = 201$, $\bar{m} = 0.03$. $I = 101$, $\bar{q} = 0.012$ m. $\sigma = 0.3$ m. $c = 340$ ms$^{-1}$. The source position $\hat{\mathbf{r}}^{\prime}$ is estimated from ray space data.

*4) Impact of localization error:* In practical applications, where the source position is not known in advance, the filter can be computed based on an estimate of the source position obtained through the approach presented in (V-A). The estimate $\hat{\mathbf{r}}^{\prime}$ is used, in particular, to compute the estimated propagation vector $\hat{\mathbf{h}} = [h(\mathbf{r}_0|\hat{\mathbf{r}}^{\prime};\omega), \ldots, h(\mathbf{r}_{L-1}|\hat{\mathbf{r}}^{\prime};\omega)]^T$, or its inverse $\hat{\mathbf{g}} = [g(\mathbf{r}_0|\hat{\mathbf{r}}^{\prime};\omega), \ldots, g(\mathbf{r}_{L-1}|\hat{\mathbf{r}}^{\prime};\omega)]^T$, which are used, in place of the ideal $\mathbf{h}$ or $\mathbf{g}$, for the computation of $\mathbf{b}$ in (38) or (44), respectively.

Fig. 8 shows the output signal to noise ratio (45) as a function of the temporal frequency $\omega$, when the source position is estimated. Results are shown for different values of the input signal to noise ratio iSNR. We observe that the localization error does not affect the oSNR at low frequencies; on the other hand, with increasing frequencies, the oSNR slightly decreases. This is due to the mismatch between the actual propagation vector $\mathbf{h}$ that generates array data and the estimated one (both in the form of $\hat{\mathbf{h}}$ or $\hat{\mathbf{g}}$). We notice, moreover, that the reduction in filtering performance is more relevant for low values of the iSNR, while for high values of iSNR the filtering performance is almost constant over frequencies.

We limited our analysis of spatial filtering performance to that of the impact of the localization error. This analysis, however, is more general than it seems, as other sources of uncertainty are shown in [78], [79] to exhibit a similar impact.

## VI. Conclusions

In this manuscript we introduced the ray space transform in the context of acoustic array processing. We formulated the ray space transform as a local Fourier analysis of array signals in the spatial domain. We also showed that the parameterization of the transform coefficients in the ray space enables array processing operations to be carried out through pattern analysis techniques. In particular, we showed applications of the ray space transform to problems of acoustic source localization in the near field of the array and for spatial spot filtering. We also supported the theoretical derivations with simulations.

We found the ray space transform to be a valuable tool for multiple reasons. 1) It inherits all the advantages of the ray space-based representation [35], [41]; in particular, it turns acoustic analysis problems into pattern analysis problems, thus enabling the use of solutions from the rich literature on pattern analysis and multidimensional signal processing. 2) Ray space analysis and synthesis operations are both data-independent and application-independent; thus, we can foresee the production of dedicated hardware to perform such tasks. 3) The theory of Gabor frames provides us with an analytical tool to assess the quality of the transform, in terms of resolution issues, numerical stability, inversion. 4) The quality of the approximation provided by the ray space transform is constant over temporal frequency. This is not the case of the soundfield map, which provides a good approximation of the acoustic field only for mid-high frequencies.

The examples of application presented in this manuscript represent an initial proof of concept of these idea and they will be extended in future works.

## References

[1] *Audio Signal Processing for Next Generation Multimedia Communication Systems*, Y. A. Huang and J. Benesty, Eds. Boston, MA, USA: Kluwer Academic, 2004.

[2] S. J. Elliott and P. A. Nelson, "Active noise control," *IEEE Signal Process. Mag.*, vol. 10, no. 4, pp. 12–35, Oct. 1993.

[3] E. G. Williams, *Fourier Acoustics*. London, U.K.: Academic, 1999.

[4] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[5] S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[6] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1627, Aug. 2001.

[7] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1053–1065, Mar. 2007.

[8] J. Chen, J. Benesty, and Y. Huang, "A minimum distortion noise reduction algorithm with multiple microphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 481–493, Mar. 2008.

[9] M. Souden, J. Chen, J. Benesty, and S. Affs, "An integrated solution for online multichannel noise tracking and reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2159–2169, Sep. 2011.

[10] E. A. P. Habets and J. Benesty, "Multi-microphone noise reduction based on orthogonal noise signal decompositions," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 6, pp. 1123–1133, Jun. 2013.

[11] G. Huang, J. Benesty, T. Long, and J. Chen, "A family of maximum SNR filters for noise reduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 12, pp. 2034–2047, Dec. 2014.

[12] Y. Lacouture-Parodi, E. A. P. Habets, J. Chen, and J. Benesty, "Multichannel noise reduction in the Karhunen–Loève expansion domain," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 5, pp. 923–936, May 2014.

[13] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical analysis of linearly constrained multi-channel Wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2384–2397, Dec. 2015.

[14] S. Doclo and M. Moonen, "Multimicrophone noise reduction using recursive GSVD-based optimal filtering with ANC postprocessing stage," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 53–69, Jan. 2005.

[15] J. Benesty, P. Duhamel, and Y. Grenier, "A multichannel affine projection algorithm with applications to multichannel acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 3, no. 2, pp. 35–37, Feb. 1996.

[16] Y. A. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[17] A. Levy, S. Gannot, and E. A. P. Habets, "Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1540–1555, Aug. 2011.

[18] I. Cohen, "Multichannel post-filtering in nonstationary noise environments," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1149–1160, May 2004.

[19] P. A. Gauthier, C. Camier, Y. Pasco, A. Berry, E. Chambatte, R. Lapointe, and M. A. Delalay, "Beamforming regularization matrix and inverse problems applied to sound field measurement and extrapolation using microphone array," *J. Sound Vibrat.*, vol. 330, pp. 5852–5877, 2011.

[20] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 35–47, Jan. 2011.

[21] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[22] Y. Bucris, I. Cohen, and M. A. Doron, "Bayesian focusing for coeherent wideband beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1282–1296, May 2012.

[23] O. Thiergart, M. Taseska, and E. A. P. Habets, "An informed parametric filter based on instantaneous direction-of-arrival estimates," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 12, pp. 2182–2196, Dec. 2014.

[24] K. Kowalczyk, O. Thiergart, M. Taseska, G. D. Galdo, V. Pulkki, and E. A. P. Habets, "Parametric spatial sound processing," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 31–42, Mar. 2015.

[25] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.

[26] E. H. Adelson and J. R. Bergen, *Computational Models of Visual Processing*. Cambridge, MA, USA: MIT Press, 1991, ch. The Plenoptic Function and the Elements of Early Vision, pp. 3–20.

[27] M. Levoy, "Light fields and computational imaging," *Computer*, vol. 39, no. 8, pp. 46–55, Aug. 2006.

[28] T. Ajdler and M. Vetterli, "The plenacoustic function, sampling and reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2003, vol. V, pp. 616–619.

[29] T. Ajdler, L. Sbaiz, A. Ridolfi, and M. Vetterli, "On a stochastic version of the plenacoustic function," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2006, vol. IV, pp. 1125–1128.

[30] R. Mignot, G. Chardon, and L. Daudet, "Compressively sampling the plenacoustic function," in *Proc. SPIE*, 2011, vol. 8138.

[31] F. Pinto and M. Vetterli, "Space-time-frequency processing of acoustic wave fields: Theory, algorithms, and applications," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4608–4622, Sep. 2010.

[32] F. Pinto and M. Vetterli, "Near-field adaptive beamforming and source localization in the spacetime frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2010, pp. 14–19.

[33] R. A. Kennedy, T. D. Abhayapala, and D. B. Ward, "Broadband nearfield beamforming using a radial beampattern transformation," *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2147–2156, Aug. 1998.

[34] F. Pinto, M. Kolundzija, and M. Vetterli, "Digital acoustics: processing wave fields in space and time using DSP tools," *APSIPA Trans. Signal Inf. Process.*, vol. 3, pp. 1–21, 2014.

[35] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Soundfield imaging in the ray space," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 12, pp. 2493–2505, Dec. 2013.

[36] D. Marković, G. Sandrini, F. Antonacci, A. Sarti, and S. Tubaro, "Plenacoustic imaging in the ray space," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Aachen, DE, Germany, 2012.

[37] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 812–824, May 2008.

[38] L. Bianchi, F. D'Amelio, F. Antonacci, A. Sarti, and S. Tubaro, "A plenacoustic approach to acoustic signal extraction," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, 2015.

[39] L. Bianchi, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Deconvolution of plenacoustic images," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, 2013.

[40] A. Canclini, L. Mucci, F. Antonacci, A. Sarti, and S. Tubaro, "Estimation of the radiation pattern of a violin during the performance using plenacoustic methods," in *Proc. AES 138th Conv.*, Warsaw, PL, Poland, 2015.

[41] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Multiview soundfield imaging in the projective ray space," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 6, pp. 1054–1067, Jun. 2015.

[42] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Electr. Eng.—Part III: Radio Commun. Eng.*, vol. 93, no. 26, pp. 429–441, Nov. 1946.

[43] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.

[44] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Trans. Signal Process.*, vol. 41, no. 7, pp. 2429–2438, Jul. 1993.

[45] S. Qiu and H. G. Feichtinger, "Discrete Gabor structures and optimal representations," *IEEE Trans. Signal Process.*, vol. 43, no. 10, pp. 2258–2268, Oct. 1995.

[46] M. J. Bastiaans and M. C. Geilen, "On the discrete Gabor transform and the discrete Zak transform," *Signal Process.*, vol. 49, no. 3, pp. 151–166, Mar. 1996.

[47] P. L. Sondergaard, "Finite discrete Gabor analysis," Ph.D. dissertation, Kongens Lyngby, DK, Denmark, Danish Tech. Univ., 2007.

[48] S. Moreno-Picot, M. Arevalillo-Herráez, and W. Díz-Villanueva, "A linear cost algorithm to compute the discrete Gabor transform," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2667–2674, May 2010.

[49] M. Taseska and E. A. P. Habets, "Spotforming using distributed microphone arrays," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, 2013.

[50] M. Taseska and E. A. P. Habets, "Informed spatial filtering for sound extraction using distributed microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 7, pp. 1195–1207, Jul. 2014.

[51] D. Lugara, C. Letrou, A. Shlivinski, E. Heyman, and A. Boag, "Frame-based gaussian beam summation method: Theory and applications," *Radio Sci.*, vol. 38, no. 2, pp. 1–15, 2003.

[52] T. Heilpern and E. Heyman, "MUSIC imaging using phase-space gaussian-beams processing," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1270–1281, Mar. 2014.

[53] J. J. Maciel and L. B. Felsen, "Discretized Gabor-based beam algorithm for time-harmonic radiation from two-dimensional truncated planar aperture distributions—I: Formulation and solution," *IEEE Trans. Antennas Propag.*, vol. 50, no. 12, pp. 1751–1759, Dec. 2002.

[54] M. Katsav and E. Heyman, "Gaussian beam summation representation of beam diffraction by an impedance wedge: A 3D electromagnetic formulation within the physical optics approximation," *IEEE Trans. Antennas Propag.*, vol. 60, no. 12, pp. 5843–5858, Dec. 2012.

[55] V. Cerveny, "Gaussian beam synthetic seismograms," *J. Geophys.*, vol. 58, p. 44.72 , 1985.

[56] T. George, J. Virieux, and R. Madariaga, "Seismic wave synthesis by Gaussian beam summation: A comparison with finite differences," *Geophys.*, vol. 52, no. 8, pp. 1065–1973, Aug. 1987.

[57] R. L. Nowack, "Calculation of synthetic seismograms with gaussian beams," *Pure Appl. Geophys.*, vol. 160, pp. 487–507, 2003.

[58] T. Werther, Y. C. Eldar, and N. K. Subbanna, "Dual Gabor frames: Theory and computational aspects," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 4147–4158, Nov. 2005.

[59] T. Werther, E. Matusiak, Y. C. Eldar, and N. K. Subbana, "A unified approach to dual Gabor windows," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1758–1768, May 2007.
[60] J. B. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-25, no. 3, pp. 235–238, 1977.
[61] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1305–1319, May 2007.
[62] J. Benesty, J. Chen, and E. A. P. Habets, *Speech Enhancement in the STFT Domain*, ser. Springer Briefs in Electrical and Computer Engineering. Heidelberg, Germany: Springer, 2012.
[63] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, 2005.
[64] M. Guillaume and Y. Grenier, "Sound field analysis based on analytical beamforming," *J. Adv. Signal Process.*, vol. 2007, pp. 1–15, 2007.
[65] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*. Berlin, Germany: Springer-Verlag, 1992.
[66] G. C. Sherman, "Diffracted wave fields expressible by plane-wave expansions containing only homogeneous waves," *J. Opt. Soc. Amer.*, vol. 59, pp. 697–711, 1969.
[67] L. Bianchi, V. Baldini, D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro, "A linear operator for the computation of soundfield maps," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2016, pp. 410–414.
[68] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. New York, NY, USA: Wiley, 1991.
[69] M. Vorländer, *Auralization*, 1st ed. Berlin, Germany: Springer-Verlag, 2008.
[70] N. A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Oxford, U.K.: Elsevier Ltd., 2004.
[71] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.
[72] P. Brémaud, *Mathematical Principles of Signal Processing*. New York, NY, USA: Springer Science+Business Media, 2002.
[73] M. Vetterli, J. Kovacevic, and V. K. Goyal, *Foundations of Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
[74] F. W. J. Olver, Ed., *NIST Handbook of Mathematical Functions*. New York, NY, USA: Nat. Inst. of Standards and Technol., 2010.
[75] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.
[76] L. F. Shampine, "Vector adaptive quadrature in Matlab," *J. Comput. Appl. Math.*, vol. 211, pp. 131–140, 2008.
[77] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–175, Sep. 2003.
[78] S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2511–2526, Oct. 2003.
[79] S. Doclo and M. Moonen, "Design of broadband beamformers robust against microphone position errors," in *Proc. Int. Workshop Acoustic Echo Noise Control (IWAENC)*, Kyoto, JP, Japan, 2003.

**Lucio Bianchi** (M'15) received his B.Sc. degree in electronic engineering (2010), the M.Sc. degree in computer engineering (2012), and the Ph.D. degree in information technology (2016) from Politecnico di Milano, Milan, Italy. He then joined the Department of Electronics, Information and Bioengineering of the Politecnico di Milano as a Post-Doctoral researcher. His research interests are mainly focused on applications of signal processing to acoustics.

**Fabio Antonacci** (M'15) received his "Laurea" B.S. and M.S. degrees in 2004 and his Ph.D. degree in 2008, from the Politecnico di Milano, Italy. He has been with DEIB—Politecnico di Milano since then, first as a Post-Doctoral researcher, then (in 2014) as an assistant professor.

His research interests are in acoustic scene analysis and rendering through microphone and loudspeaker arrays, with particular focus on source localization/tracking, reflector localization, estimation of acoustic parameters, rendering of the acoustics of virtual environments using wavefield synthesis techniques. He has authored over 70 journal and conference papers on these topics, and is currently a member of the EURASIP Special Area Team (SAT) on Acoustic, Sound and Music Signal Processing (ASMSP).

**Augusto Sarti** (SM'04) received his M.S. degree in electronic engineering in 1988 from the University of Padova, Italy; and his Ph.D. degree in information engineering in 1993, from the University of Padova in 1993, a joint program with the University of California, Berkeley. In 1993, he joined the faculty of the Politecnico di Milano, Italy. In 2013, he also joined the University of California, Davis, as an Adjunct Professor. His research interests are in the area of multimedia signal processing, with particular focus on sound analysis, synthesis and processing; space-time audio processing; geometrical acoustics; and music information extraction. He has also worked on problems of multi-dimensional and nonlinear signal processing, as well as 3D vision. He coauthored over 250 scientific publications on international journals and congresses as well as numerous patents in the multimedia signal processing area. He coordinates the activities of the Musical Acoustics Lab and of the Sound and Music Computing Lab of the Politecnico di Milano. He has been the promoter/coordinator and/or contributor to numerous (20+) European projects. He is an active member of the IEEE Technical Committee on Audio and Acoustics Signal Processing, and is in the Editorial Board of the IEEE.

**Stefano Tubaro** (SM'01) was born in Novara in 1957. He received his M.S. degree in electronic engineering in 1982 from the Politecnico di Milano, Italy. He then joined the Dipartimento di Elettronica, Informazione e Bioingegneria of the Politecnico di Milano, first as a researcher of the National Research Council, and then (in November 1991) as an Associate Professor. Since December 2004 he has been appointed as Full Professor of Telecommunication at the Politecnico di Milano.

His current research interests are on advanced algorithms for video and sound processing. He authored over 180 scientific publications on international journals and congresses and he is co-author of more than 15 patents. In the past few years he has focused his interest on the development of innovative techniques for image and video tampering detection and, in general, for the blind recovery of the processing history of multimedia objects. He coordinates the research activities of the Image and Sound Processing Group (ISPG) at the Dipartimento di Elettronica, Informazione e Bioingegneria of the Politecnico di Milano. He had the role of Project Coordinator of the European Project ORIGAMI: A new paradigm for high-quality mixing of real and virtual and of the research project ICT-FET-OPEN REWIND: REVerse engineering of audio-VIsual coNtent Data. This last project was aimed at synergistically combining principles of signal processing, machine learning and information theory to answer relevant questions on the past history of such objects. He is a member the IEEE Multimedia Signal Processing Technical Committee and of the IEEE SPS Image Video and Multidimensional Signal Technical Commitee. He was in the organization committee of a number of international conferences: IEEE-MMSP-2004/2013, IEEE-ICIP-2005, IEEE-AVSS-2005/2009, IEEE-ICDSC-2009, IEEE-MMSP-2013, IEEE-ICME-2015 to mention a few. From May 2012 to April 2015 he was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING. Currently he is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.