Space-Time-Frequency Processing of Acoustic Wave Fields: Theory, Algorithms, and Applications

Francisco Pinto, *Member, IEEE*, and Martin Vetterli, *Fellow, IEEE*

***Abstract*—Consider a nonparametric representation of acoustic wave fields that consists of observing the sound pressure along a straight line or a smooth contour $\mathcal{L}$ defined in space. The observed data contains implicit information of the surrounding acoustic scene, both in terms of spatial arrangement of the sources and their respective temporal evolution. We show that such data can be effectively analyzed and processed in what we call the space-time-frequency representation space, consisting of a Gabor representation across the spatio-temporal manifold defined by the spatial axis $\mathcal{L}$ and the temporal axis t. In the presence of a source, the spectral patterns generated at $\mathcal{L}$ have a characteristic triangular shape that changes according to certain parameters, such as the source distance and direction, the number of sources, the concavity of $\mathcal{L}$, and the analysis window size. Yet, in general, the wave fronts can be expressed as a function of elementary directional components—most notably, plane waves and far-field components. Furthermore, we address the problem of processing the wave field in discrete space and time, i.e., sampled along $\mathcal{L}$ and t, where a Gabor representation implies that the wave fronts are processed in a block-wise fashion. The key challenge is how to chose and customize a spatio-temporal filter bank such that it exploits the physical properties of the wave field while satisfying strict requirements such as perfect reconstruction, critical sampling, and computational efficiency. We discuss the architecture of such filter banks, and demonstrate their applicability in the context of real applications, such as spatial filtering, deconvolution, and wave field coding.**

***Index Terms*—Array signal processing, beamforming, directional filter banks, source localization, space-time-frequency analysis, spatial filtering, wave field coding.**

I. Introduction

A. Historical Perspective

SIGNAL processing has been used throughout history as a means to describe, manipulate, and reproduce physical phenomena occurring in nature. Many of these phenomena are governed by well known mathematical laws, as well as statistical properties that make it possible to predict the structure of the signals. The challenge is often how to find a suitable representation space where the signals are expressed in a more efficient and manageable way.

During the early ninenteenth century, Fourier suggested that the solution to the heat equation in a solid medium could be expressed as a linear combination of harmonic solutions, which simplified the problem in a notorious way [2]. This concept was reinforced by Dirichlet, who demonstrated that a similar transform could be obtained for a general class of signals [3], and hence be applied to many other fields of science. However, in the mid-twentieth century, Gabor realized that the Fourier transform was unable to represent the frequency variations along time that characterize nonstationary signals such as speech and music. To solve this limitation, Gabor modified the concept of frequency by representing it as a parametric function defined over a time-frequency representation space [4]. This new concept led to the development of the widely used short-time Fourier transform, and eventually to the wavelet transform [5] and the modulated lapped transform [6], along with the many practical applications in the areas of audio, speech, and image processing.

In this paper, we propose a generalization of Gabor's time-frequency representation such that it represents not only the frequency variations of sound along time but also the variations of the acoustic wave field across space. This consists of including a spatial dimension in the signal in order to identify the coordinates of multiple observation points, and extending the local Fourier analysis to this new dimension. This results in what we call the *space-time-frequency* representation space.

The proposed method allows to efficiently represent acoustic scenes composed of wideband point sources in the far-field, as well as more complex scenes that can be expressed as a function of known elementary solutions. The method finds applications in the areas of spatial audio and wave field processing in general.

B. Related Work and Contributions

A point source is typically defined as a singularity in space characterized by a source signal $s(t)$ and a spatial position $\mathbf{r}_o = (x_o, y_o, z_o)$. In free field, each point source contributes to the wave field with a sound pressure given by [7]

$$p(\mathbf{r}, t) = \frac{s\left(t - \frac{\|\mathbf{r}-\mathbf{r}_o\|}{c}\right)}{4\pi\|\mathbf{r} - \mathbf{r}_o\|} \qquad (1)$$

where $\mathbf{r} = (x, y, z)$ is the point of observation, c is the speed of sound, and $\|.\|$ is the usual vector 2-norm. In the case of a closed field, the result can be generalized to account for reflections and distortion. For instance, a general filter $h(\mathbf{r}, t)$ representing the acoustic path from position $\mathbf{r}_o$ to all other $\mathbf{r}$ can

Manuscript received November 08, 2009; accepted May 22, 2010. Date of publication June 07, 2010; date of current version August 11, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bogdan Dumitrescu. The project is funded by the Fundação para a Ciência e Tecnologia (Portugal) and the National Center of Competence in Research for Mobile Information and Communication Systems (Switzerland). The results of this research are reproducible [1] and can be found in the authors' repository at http://rr.epfl.ch/.

F. Pinto is with the LCAV, DSC, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland (e-mail: francisco.pinto@epfl.ch).

M. Vetterli is with the LCAV, DSC, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, and also with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, 94720 USA (e-mail: martin.vetterli@epfl.ch).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. The material includes Matlab scripts and instruction videos for generating the figures provided in the paper, as well as a Matlab graphical user interface for simulating and analyzing acoustic wave fields.

Digital Object Identifier 10.1109/TSP.2010.2052045

1053-587X/$26.00 © 2010 IEEE

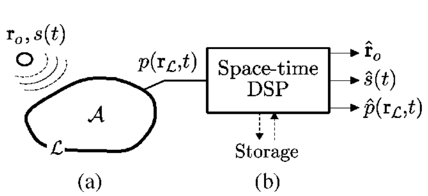


Fig. 1. Basic formulation a wave field processing problem: (a) the sound pressure is taken along an arbitrary contour $\mathcal{L}$, possibly enclosing a source-free area $\mathcal{A}$; (b) the resulting spatio-temporal signal $p(\mathbf{r}_\mathcal{L},t)$ goes through a digital signal processor operating in the spatio-temporal domain. Examples of output data include the source locations $\hat{\mathbf{r}}_o$, the source signals $\hat{s}(t)$, and the reconstructed sound pressure $\hat{p}(\mathbf{r}_\mathcal{L},t)$.

be defined such that $p(\mathbf{r},t) = h(\mathbf{r},t) * s(t)$, where $h(\mathbf{r},t) = \delta\big(t - \|\mathbf{r}-\mathbf{r_o}\|/c\big)/(4\pi\|\mathbf{r}-\mathbf{r}_o\|)$ yields the result in (1). The source signal $s(t)$ itself can be represented by a suitable parametric model, in case it has an efficient parametric form. In effect, many different parameters can be used to describe the acoustic scene depending on the amount of prior knowledge available and the desired accuracy when computing $p(\mathbf{r},t)$. This is called a *parametric* description of the wave field.

In a different scenario—the one we are interested in—it is assumed that no prior knowledge of the acoustic scene is available, or simply that no parametric description is desired. In this case, the wave field is characterized by $p(\mathbf{r},t)$ itself, and the description is said to be blind or *nonparametric*.

To describe the entire wave field using a nonparametric description $p(\mathbf{r},t)$ for all $\mathbf{r}$ would require a massive amount of data, impossible to handle in practice. However, in many situations of interest, it is possible to retain significant information about the wave field solely by considering $p(\mathbf{r},t)$ along an arbitrary contour $\mathcal{L}$ defined in space. The resulting signal $p(\mathbf{r}_\mathcal{L},t)$ can be used, for example, to localize the various sources in the acoustic scene [8], to filter the sources by generating directivity patterns (beamforming) [9], or to reconstruct the wave field in an enclosed area using wave field synthesis [10]. We formulate this as a digital signal processing (DSP) problem, where the system takes as input $p(\mathbf{r}_\mathcal{L},t)$ and returns a reconstructed or processed version of the input, $\hat{p}(\mathbf{r}_\mathcal{L},t)$, or any other data related to the wave field, as illustrated in Fig. 1.

A typical DSP system is composed of three stages: sampling, processing, and interpolation. In the case of Fig. 1, the system takes as input a spatio-temporal function $p(\mathbf{r}_\mathcal{L},t)$, representing the entire wave field outside $\mathcal{A}$. This means that the three stages have to operate not only in the time domain but also in the spatial domain.

The topic of *sampling* in the spatial domain is addressed in the work of Ajdler *et al.* [11] on the plenacoustic function (PAF). The authors show that the spatio-temporal representation of the wave field is essentially band-limited and can be reduced to a closed-form solution in many cases of interest, such as multiple near-field sources in a reverberant room. Moreover, the Nyquist sampling theory can be used to sample the acoustic wave field across space using different sampling patterns, while allowing subsequent reconstruction with minimum or no spatial aliasing.

The topic of *interpolation* in the spatial domain is addressed in the work of Berkhout *et al.* [10], [12] on wave field synthesis (WFS). The authors show how the spatial samples can be used to actually resynthesize the "analog" wave fronts such that the resulting wave field is physically equivalent to (or a processed version of) the original wave field. The theory of WFS is based on a combination of the Huygens–Fresnel principle and the divergence theorem [10] that implies that the wave field within a source-free area $\mathcal{A}$ is completely defined by the field values observed at the boundary $\mathcal{L}$ [see Fig. 1(a)] and can be replicated by driving a line source—coincident with $\mathcal{L}$—with the observed field values. For this reason, WFS is used as a technique for spatial audio playback, where a large loudspeaker array acts as the line source.

TABLE I
REFERENCE TABLE OF MATHEMATICAL NOTATION

Symbol	Description	Symbol	Description
$p(\mathbf{r},t)$, $p(x,t)$	4D, 2D sound pressure;	$\mathcal{L}$, $\mathcal{S}$	Contour, surface;
$P(\mathbf{\Phi},\Omega)$, $P(\Phi,\Omega)$	4D, 2D spectrum;	α, β	Far-field angles;
$P(\alpha,\Omega)$	Directional spectrum;	$s(t)$	Source signal;
$p[\mathbf{n}] = p_{n_x,n_t}$	Signal samples;	$B(\Phi,\Omega)$	Support function;
$P[\mathbf{b}] = P_{b_x,b_t}$	Transform coefficients;	$w_x(x)$	Spatial window;
υ_{b_x,n_x}, ψ_{b_t,n_t}	Orthogonal bases;	$w_t(t)$	Temporal window;
$\mathbf{P}$, $\mathbf{Y}$	Matrix forms of p, P;	$\alpha_{\mathrm{nf}}(x)$	Near-field angle;
$\mathbf{\Upsilon}$, $\mathbf{\Psi}$	Matrix forms of υ, ψ;	$M(\Phi,\Omega)$	Near-field mask.

The purpose of this paper is to address the topic of *processing* the wave field in discrete space and time. We take advantage of a powerful consequence of the work on the PAF and WFS: the fact that these allow the wave field to be processed using multidimensional signal processing theory—in particular, Fourier theory. In Section II, we review the spectral representation of $p(\mathbf{r}_\mathcal{L},t)$ when the Fourier transform is taken over space and time, and how it is affected by the characteristics of the acoustic scene, based on known results of acoustics theory. We also derive new results that establish a more comprehensive link between acoustics theory and signal processing, in particular by providing a definition of "frequency" on a space-time-frequency representation space. In Section III, we address the problem of processing the wave field in discrete space and time, based on Nyquist sampling theory and multidimensional filter banks theory. In particular, we discuss examples of orthogonal filter banks that effectively represent $p(\mathbf{r}_\mathcal{L},t)$ in terms of its elementary components while satisfying the requirements of critical sampling and perfect reconstruction of the input. Finally, in Section IV, we discuss potential applications of space-time-frequency processing that make direct use of the concepts discussed in this paper, with special emphasis on i) spatial filtering, ii) deconvolution, and iii) wave field coding.

II. Space-Time Analysis of the Wave Field

We begin our analysis with the characterization of the sound pressure generated by a point source over an infinite flat surface $\mathcal{S}$ and an infinite straight line $\mathcal{L}$, and show how $p(\mathbf{r}_\mathcal{S},t)$ relates to $p(\mathbf{r}_\mathcal{L},t)$. First, we consider that the point source is located in the far-field, where $\|\mathbf{r}_o\| \gg \|\mathbf{r}\|$, and then in the near-field, where $\|\mathbf{r}-\mathbf{r}_o\| \rightarrow 0$. In both cases, the acoustic scene is assumed to be in free field.

Further on, we show how $p(\mathbf{r}_\mathcal{L},t)$ can be expressed as a function of plane waves and far-field components, and how such representations are influenced by the effects of windowing.

The main notation used in this paper is listed in Table I.

A. Far-Field Spectrum

The point source being located in the far-field (FF) implies that the wave front reaching $\mathcal{S}$ has negligible curvature (i.e., is nearly plane). Under the far-field assumption, $\|\mathbf{r}-\mathbf{r}_o\| = \|\mathbf{r}_o\| - \epsilon(\mathbf{r})$ where $\epsilon(\mathbf{r})$ is a residual term dependent on the point of observation $\mathbf{r}$. Thus, (1) is simplified to [7]

$$p(\mathbf{r},t) \approx \frac{s\left(t - \frac{\|\mathbf{r}_o\|}{c} + \Theta(\mathbf{r})\right)}{4\pi\|\mathbf{r}_o\|} \tag{2}$$

where $1/(4\pi\,\|\mathbf{r}_o\|)$ and $\|\mathbf{r}_o\|/c$ are fixed amplitude and phase values that we discard for the remainder of this paper, and $\Theta(\mathbf{r}) = \epsilon(\mathbf{r})/c$ is the residual phase term that depends on $\mathbf{r}$, and can not be discarded since it appears in the argument of $s(t)$. In addition, consider that the wave front hits the infinite surface $\mathcal{S}$ (defined as the xy-plane) with angles of arrival α and β, and the infinite straight line $\mathcal{L}$ (defined as the x-axis) with angle of arrival α, as shown in Fig. 2. The direction of propagation is given by the wave front normal, $\vec{u} = \mathbf{u}\cdot\vec{\mathbf{r}}$, where $\mathbf{u} = (u_x, u_y, u_z)$ contains the directional components in each axis and $\vec{\mathbf{r}} = (\vec{x}, \vec{y}, \vec{z})$ are the standard basis vectors $\vec{x}$, $\vec{y}$, and $\vec{z}$. Without loss of generality, consider that $\|\mathbf{u}\| = 1$. Since the wave front is plane, the expression $u_x x + u_y y + u_z z = c\Theta(\mathbf{r})$ defines a profile of constant phase in the three-dimensional space. Thus [7]

$$p(\mathbf{r},t) = s\left(t + \frac{\mathbf{u}\cdot\mathbf{r}}{c}\right). \tag{3}$$

The four-dimensional Fourier transform of $p(\mathbf{r},t)$ can be obtained with the regular multidimensional expression [13]

$$P(\mathbf{\Phi},\Omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\mathbf{r},t)e^{-j(\mathbf{\Phi}\cdot\mathbf{r}+\Omega t)}dt d\mathbf{r} \tag{4}$$

where $\mathbf{\Phi} = (\Phi_x, \Phi_y, \Phi_z)$ are the spatial frequencies in rad/m and Ω is the temporal frequency in rad/s. Plugging (3) into (4) and applying the transform over space and time, yields

$$P(\mathbf{\Phi},\Omega) = S(\Omega)\int_{-\infty}^{\infty} e^{-j\left(\mathbf{\Phi}-\mathbf{u}\frac{\Omega}{c}\right)\cdot\mathbf{r}}d\mathbf{r} = 8\pi^3 S(\Omega)\delta\left(\mathbf{\Phi}-\mathbf{u}\frac{\Omega}{c}\right) \tag{5}$$

where $\delta(\mathbf{\Phi}-\mathbf{u}(\Omega/c))$ is nonzero for $\mathbf{\Phi} = \mathbf{u}(\Omega/c)$. The respective projections on $\mathcal{S}$ and $\mathcal{L}$ are given by

$$P(\Phi_x,\Phi_y,\Omega) = 4\pi^2 S(\Omega)\delta\left(\Phi_x - u_x\frac{\Omega}{c}\right)\delta\left(\Phi_y - u_y\frac{\Omega}{c}\right) \tag{6}$$

and

$$P(\Phi_x,\Omega) = 2\pi S(\Omega)\delta\left(\Phi_x - u_x\frac{\Omega}{c}\right) \tag{7}$$

where, in polar coordinates, $u_x = \cos\alpha\sin\beta$ and $u_y = \cos\beta$. Accordingly, for the $\mathcal{S}$-projection case, depicted in Fig. 3(a), the spectrum is a Dirac function of partial derivatives $\partial\Phi_x/\partial\Omega = \cos\alpha\sin\beta/c$ and $\partial\Phi_y/\partial\Omega = \cos\beta/c$, weighted by the Fourier transform of the source signal $s(t)$. Given that $\alpha,\beta \in [0,\pi]$, the Dirac function is within a cone-shaped region defined by $\Phi_x^2 + \Phi_y^2 \le (\Omega/c)^2$. For the $\mathcal{L}$-projection case, depicted in Fig. 3(b), the result is again a weighted Dirac function, except that $u_y = 0$ and thus $\partial\Phi_x/\partial\Omega = \cos\alpha/c$ and $\partial\Phi_y/\partial\Omega = 0$. The delimited region is then given by $\Phi_x^2 \le (\Omega/c)^2$, representing a triangular-shaped region. Note that, for the general case in (5), this region is defined by $\|\mathbf{\Phi}\|^2 = (\Omega/c)^2$.

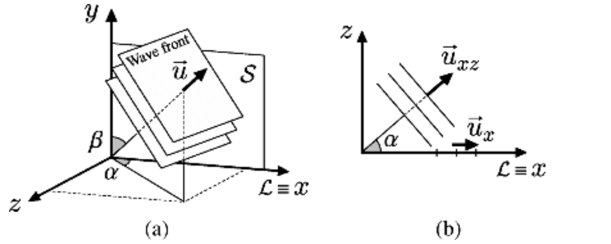


Fig. 2. Wave front radiated by a wideband point source located in the far-field: (a) in 3-D view and (b) in 2-D view. The wave front hits the plane $\mathcal{S}$ and the straight line $\mathcal{L}$ with angles $\alpha,\beta \in [0,\pi]$, generating the pressure signals $p(\mathbf{r}_{\mathcal{S}},t)$ and $p(\mathbf{r}_{\mathcal{L}},t)$. In both cases, the wave front leaves a trail of constant-phase profiles (not to confuse with a sinusoidal wave front).

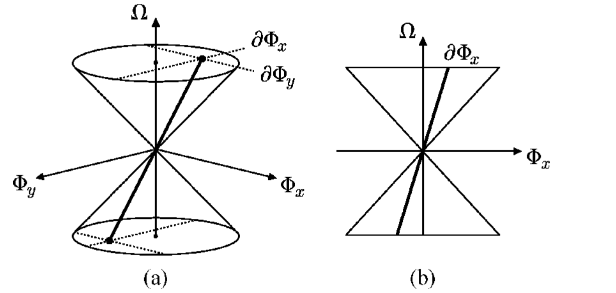


Fig. 3. Spectral representation of: (a) $p(\mathbf{r}_{\mathcal{S}},t)$ and (b) $p(\mathbf{r}_{\mathcal{L}},t)$ for a far-field source. The bold line represents a Dirac function weighted by the spectrum of the source signal $s(t)$. The slope of the line depends on the angle of arrival of the wave front.

To complete the intuition, consider two particular cases where $s(t) = e^{j\Omega_o t}$ and $s(t) = \delta(t)$. The respective Fourier transforms are given by $S(\Omega) = 2\pi\delta(\Omega-\Omega_o)$ and $S(\Omega) = 1$. Plugging $S(\Omega)$ into (6) and (7), it follows that a source signal with a complex frequency translates into a single point in the spectrum, whereas a source signal containing all the frequencies generates a "flat" line. A complex frequency in the far-field is known as a *plane wave*.[1]

We can conclude that the definition of "frequency" in the spatio-temporal Fourier analysis of the wave field is a plane wave, and that frequencies in the traditional context of signal processing are particular cases of the plane wave—when $\mathcal{L}$ is one point in space.

B. Near-Field Spectrum

A point source is considered to be in the near-field (NF) when its distance to the observation surface is very small compared to the size of the surface—in this case, infinite. For simplicity, as formulated in the introduction, we consider only the projection on the straight line $\mathcal{L}$, where $y = 0$ and $u_y = 0$.

To analyze the effects of near-field sources, it is important to understand the physical meaning of the region outside the spectral triangle shown in Fig. 3(b). For this purpose, let $\|\mathbf{u}\|^2 = 1$

[1] Note that any wideband signal $s(t)$ generates a flat wave-front as long as $\|\mathbf{r}_o\| \gg \|\mathbf{r}\|$, though historically a plane wave refers to a single complex frequency [7].

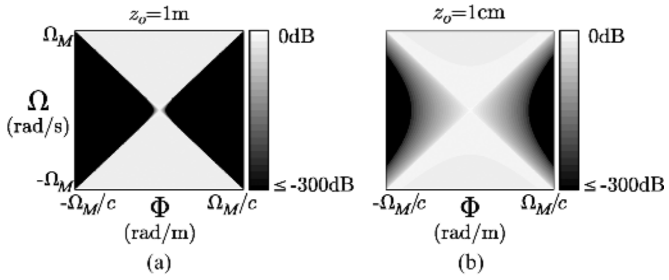


Fig. 4. Spectral representation of $p(\mathbf{r}_{\mathcal{L}}, t)$ for a near-field source at a distance of: (a) 1 m and (b) 1 cm relative to $\mathcal{L}$. The proximity of the source causes the energy to spread across Φ and intensify the evanescent-wave content.

be rewritten as $u_z = \pm\sqrt{1-u_x^2}$, and Φ_x as Φ. The outside region is given by $\Phi^2 > (\Omega/c)^2$, or equivalently $u_x^2 > 1$. This implies that $\sqrt{1-u_x^2}$ is complex, and thus $u_z = \pm j\sqrt{u_x^2-1} = \pm j|u_z|$. Taking the Fourier transform of (3) over time, yields [7]

$$p(\mathbf{r},\Omega) = S(\Omega)e^{j\frac{\Omega}{c}(u_x x + u_z z)} = S(\Omega)e^{\pm\frac{\Omega}{c}|u_z|z}e^{j\frac{\Omega}{c}u_x x} \quad (8)$$

for $z < 0$, where the sign $\pm$ is chosen such that $\pm\Omega z$ is negative, thus providing a physical solution [7]. Notably, the term $e^{\pm(\Omega/c)|u_z|z}$ represents an exponential decay that prevents the wave front from propagating towards z. This is called the *evanescent mode* of wave propagation, as opposed to the *propagating mode* analyzed in the previous section. In the special case of $\Phi^2 = (\Omega/c)^2$, the decay term disappears due to $u_z = 0$ and the result is a wave traveling in parallel to $\mathcal{L}$.

As mentioned before, a far-field source can only be defined inside the triangular region $\Phi^2 \leq (\Omega/c)^2$, since on the outside the wave front would require a complex angle of arrival α. However, this is not the case for sources in the near-field. When the source is closer to $\mathcal{L}$, there is an increase in the curvature of the wave front that spreads the spectral energy past the cut-off line $\Phi^2 = (\Omega/c)^2$, entering the evanescent region. It can be shown that, for a near-field source $s(t)$ located at $\mathbf{r} = (x_o, z_o)$, the Fourier transform over $\mathcal{L}$ is given by [11]

$$P(\Phi,\Omega) = \frac{1}{4}S(\Omega)H_0^{(1)*}\left(z_o\sqrt{\left(\frac{\Omega}{c}\right)^2 - \Phi^2}\right)e^{-j\left(x_o\Phi + \frac{\pi}{2}\right)} \quad (9)$$

where $H_0^{(1)}$ is the zeroth-order Hankel function of the first kind. The result for $S(\Omega) = 1$, illustrated in Fig. 4, shows that the near-field spectrum contains most of its energy inside the triangular region, except for some residual energy on the outside. For $\Phi^2 > (\Omega/c)^2$, the Hankel function is upper-bounded by $e^{-z_o\Phi}/\sqrt{z_o\Phi}$, and converges to the upper-bound for $\Phi^2 \gg (\Omega/c)^2$ [7], [11]. Thus, the farther away the source is from $\mathcal{L}$, the faster is the amplitude decay and the less evanescent waves emerge. On the contrary, as the source moves closer to $\mathcal{L}$, the balance between propagating and evanescent waves tips towards the evanescent waves.

In general, the spatio-temporal spectrum over $\mathcal{L}$ can be defined as the spectrum of the sound source $S(\Omega)$ based on a support function $B(\Phi,\Omega)$ such that

$$P(\Phi,\Omega) = S(\Omega)B(\Phi,\Omega) \quad (10)$$

where, depending on whether the source is in the far-field (FF) or the near-field (NF),

$$B_{\text{ff}}(\Phi,\Omega) = 2\pi\delta\left(\Phi - \cos\alpha\frac{\Omega}{c}\right) \quad (11)$$

and

$$B_{\text{nf}}(\Phi,\Omega) = \frac{1}{4}H_0^{(1)*}\left(z_o\sqrt{\left(\frac{\Omega}{c}\right)^2 - \Phi^2}\right)e^{-j\left(x_o\Phi + \frac{\pi}{2}\right)}. \quad (12)$$

C. Decomposition Into Plane Waves and Far-Field Components

Up to this point, we have considered the space-time representation of the wave field over a straight line $\mathcal{L}$, given some parametric specification of the acoustic scene. In the introduction, however, we have stated the problem as being nonparametric. Thus, the main challenge is how to identify and exploit the characteristics of the acoustic scene given $p(\mathbf{r}_{\mathcal{L}}, t)$. For this purpose, we introduce two theoretical tools of central importance to this paper—i) the decomposition of the wave field into plane waves and ii) the decomposition of the wave field into far-field components—enunciated in the following two propositions. A third proposition is also derived, with particular importance to the discussion of directional filter banks in Section III.

Definition 1: The directional spectrum is a function $P(\alpha,\Omega) \in \mathbb{C}^2$ such that

$$P(\alpha,\Omega) = \frac{d\Phi}{d\alpha}P(\Phi,\Omega)\Big|_{\Phi=\cos\alpha\frac{\Omega}{c}}$$

and

$$\int_0^{\pi} |P(\alpha,\Omega)|d\alpha < \infty, \quad \forall\Omega.$$

Proposition 1: An acoustic wave field generated by an arbitrary number of sources with $z_o \gg 0$ over a straight line $\mathcal{L}$ can be expressed as a function of plane waves traveling towards $\vec{u} = \cos\alpha(\Omega/c)\vec{x} + \sin\alpha(\Omega/c)\vec{z}$ with complex amplitude $P(\alpha,\Omega)$ within the interval $\alpha \in [0,\pi]$, plus a residual evanescent wave component $p_{\mathcal{E}}(x,\Omega)$, where

$$p(x,\Omega) = \frac{1}{2\pi}\int\limits_0^{\pi} P(\alpha,\Omega)e^{j\cos\alpha\frac{\Omega}{c}x}d\alpha + p_{\mathcal{E}}(x,\Omega). \quad (13)$$

Proof: Take the inverse Fourier transform of $P(\Phi,\Omega)$ over Φ keeping the propagation term in z,

$$p(\mathbf{r},\Omega) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} P(\Phi,\Omega)e^{j\frac{\Omega}{c}u_z z}e^{j\Phi x}d\Phi$$

where $\mathbf{r} = (x,z)$. The integration region can be split into the propagating region $\mathcal{P}$ and the evanescent region $\mathcal{E}$, depending on whether $u_z = \pm\sqrt{1-u_x^2}$ is a real value $u_z^{\mathbb{R}}$ or a complex value $u_z^{\mathbb{C}}$,

$$p(\mathbf{r},\Omega) = \frac{1}{2\pi}\int\limits_{\mathcal{P}} P(\Phi,\Omega)e^{j\frac{\Omega}{c}u_z^{\mathbb{R}}z}e^{j\Phi x}d\Phi + \underbrace{\frac{1}{2\pi}\int_{\mathcal{E}} P(\Phi,\Omega)e^{\pm\frac{\Omega}{c}|u_z^{\mathbb{C}}|z}e^{j\Phi x}d\Phi}_{=p_{\mathcal{E}}(\mathbf{r},\Omega)}.$$

The evanescent term is left unchanged, and denoted $p_{\mathcal{E}}(\mathbf{r},\Omega)$. In the propagating term, the spatial frequency can be expressed as $\Phi = u_x(\Omega/c)$, where $u_x = \cos\alpha$ and $u_z^{\mathbb{R}} = \sin\alpha$, and thus

$$\frac{1}{2\pi}\int_{-\frac{\Omega}{c}}^{\frac{\Omega}{c}} P(\Phi,\Omega)e^{j\left(\Phi x+u_z^{\mathbb{R}}\frac{\Omega}{c}z\right)}d\Phi$$

$$=\frac{1}{2\pi}\int_{0}^{\pi}\underbrace{\frac{d\cos\alpha\frac{\Omega}{c}}{d\alpha}P\left(\cos\alpha\frac{\Omega}{c},\Omega\right)}_{\overset{\text{def}}{=}P(\alpha,\Omega)}e^{j(\cos\alpha x+\sin\alpha z)\frac{\Omega}{c}}d\alpha.$$

The result in (13) is obtained by taking $z = 0$. Since $z_o \gg 0$, the evanescent region has residual energy compared to the propagating region, due to the exponential decay in (9) for $\Phi^2 > (\Omega/c)^2$ [see Fig. 4(a)]. ■

Proposition 2: The propagating wave fronts of an acoustic wave field observed on a straight line $\mathcal{L}$ can be expressed as a function of far-field components (virtual sources in the far-field) with source spectrum $P(\alpha,\Omega)$ and angle of arrival $\alpha \in [0,\pi]$, where

$$P(\Phi,\Omega)=\int_{0}^{\pi}P(\alpha,\Omega)\delta\left(\Phi-\cos\alpha\frac{\Omega}{c}\right)d\alpha. \quad (14)$$

Proof: The result can be obtained by taking the Fourier transform over x of the propagating term in (13),

$$\int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{0}^{\pi}P(\alpha,\Omega)e^{j\cos\alpha\frac{\Omega}{c}x}d\alpha\right)e^{-j\Phi x}dx$$

$$=\int_{0}^{\pi}P(\alpha,\Omega)\left(\underbrace{\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j\cos\alpha\frac{\Omega}{c}x}e^{-j\Phi x}dx}_{=2\pi\delta\left(\Phi-\cos\alpha\frac{\Omega}{c}\right)}\right)d\alpha.$$

The order of integration with respect to x and α can be exchanged under the Fubini theorem [14], given that $P(\alpha,\Omega)$ is absolutely integrable. Note that the far-field sources are characterized by $P(\Phi,\Omega) = 2\pi S(\Omega)\delta(\Phi - \cos\alpha(\Omega/c))$, and thus $P(\alpha,\Omega)$ can be interpreted as the source spectrum of a virtual source in the far-field with direction α. We use the term "far-field components" to avoid confusion with the true sources in the acoustic scene, which are not necessarily in the far-field. ■

Corollary 1: The directional spectrum $P(\alpha,\Omega)$ for any given $\alpha \in [0,\pi]$ is obtained by sampling $P(\Phi,\Omega)$ across $\Phi = \cos\alpha(\Omega/c)$ and normalizing it with respect to α and Ω, such that

$$P(\alpha,\Omega)=-\sin\alpha\frac{\Omega}{c}\int_{-\frac{\Omega}{c}}^{\frac{\Omega}{c}}P(\Phi,\Omega)\delta\left(\Phi-\cos\alpha\frac{\Omega}{c}\right)d\Phi. \quad (15)$$

Proof: This follows directly from Definition 1 and the definition of Dirac function, where $\int_{-\Omega/c}^{\Omega/c}P(\Phi,\Omega)\delta(\Phi - \cos\alpha(\Omega/c))d\Phi = P(\cos\alpha(\Omega/c),\Omega)$. ■

The result in Proposition 2 can be interpreted as the continuous counterpart of a discrete superposition of far-field sources. Equivalently, it can be seen as a projection of the directional spectrum onto the "space" formed by Dirac support functions, where $P_\Phi = \langle P_\alpha, \delta_{\Phi,\alpha}\rangle$ and $P_\alpha = (d\Phi/d\alpha)\langle P_\Phi, \delta_{\Phi,\alpha}\rangle$. This results in another important property described next.

Proposition 3: Given $\alpha_A, \alpha_B \in [0,\pi]$ such that $0 \le \alpha_A < \alpha_B \le \pi$, the spectral energy within the interval $\alpha \in [\alpha_A, \alpha_B]$ satisfies

$$\int_{\alpha_A}^{\alpha_B}P(\alpha,\Omega)d\alpha=\int_{\cos\alpha_B\frac{\Omega}{c}}^{\cos\alpha_A\frac{\Omega}{c}}P(\Phi,\Omega)d\Phi. \quad (16)$$

Proof: Plugging (14) into the right term of (16), yields

$$\int_{\cos\alpha_B\frac{\Omega}{c}}^{\cos\alpha_A\frac{\Omega}{c}}\int_{0}^{\pi}P(\alpha,\Omega)\delta\left(\Phi-\cos\alpha\frac{\Omega}{c}\right)d\alpha d\Phi$$

$$=\int_{0}^{\pi}P(\alpha,\Omega)\int_{\cos\alpha_B\frac{\Omega}{c}}^{\cos\alpha_A\frac{\Omega}{c}}\delta\left(\Phi-\cos\alpha\frac{\Omega}{c}\right)d\Phi d\alpha$$

$$=\int_{\alpha_A}^{\alpha_B}P(\alpha,\Omega)d\alpha$$

where the exchange of integrals comes from the Fubini theorem [14], and the last equality is due to

$$\int_{\cos\alpha_B\frac{\Omega}{c}}^{\cos\alpha_A\frac{\Omega}{c}}\delta\left(\Phi-\cos\alpha\frac{\Omega}{c}\right)d\Phi=\begin{cases}1, & \alpha\in[\alpha_A,\alpha_B]\\ 0, & \text{otherwise.}\end{cases}$$

■

Proposition 3 essentially says that the energy of $P(\alpha,\Omega)$ over the interval $\alpha_A \le \alpha \le \alpha_B$ is equivalent to the energy of $P(\Phi,\Omega)$ within the triangular region $\cos\alpha_B(\Omega/c) \le \Phi \le \cos\alpha_A(\Omega/c)$. This is the theoretical motivation for the discussion of directional filter banks in Section III.

The results presented in this section show how a blind (nonparametric) description of the acoustic wave field can acquire parametric characteristics when decomposed into elementary wave fronts, such as plane waves and far-field components. These wave fronts are not only sparse in the spatio-temporal Fourier domain but often appear concentrated within a limited region of the spectrum, which results in a potentially compact description of the wave field based, for example, on a limited number of far-field components $P(\alpha_0,\Omega), P(\alpha_1,\Omega),\ldots$ and the respective angles $\alpha_0, \alpha_1, \ldots$.

D. Space–Time–Frequency Analysis

Consider the case where $p(x,t)$ is analyzed in smaller blocks along the x-axis, such that each block represents the wave field at a given region in space. If $w_x(x)$ and $w_t(t)$ are the spatial and temporal window functions, a spectral block can be defined as

$$P(\Phi,\Omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_x(x)w_t(t)p(x,t)e^{-j(\Phi x+\Omega t)}dtdx. \quad (17)$$

For $p(x,t) = s(t + (\cos\alpha/c)x)$, the result can be shown to be

$$P(\Phi,\Omega) = \frac{2\pi c W_x(\Phi)}{|\cos\alpha|} *_{\Phi} \left(S\left(\frac{c}{\cos\alpha}\Phi\right) W_t\left(\Omega - \frac{c}{\cos\alpha}\Phi\right)\right) \quad (18)$$

where $W_x(\Phi)$ and $W_t(\Omega)$ are the Fourier transforms of $w_x(x)$ and $w_t(t)$, and $*_{\Phi}$ denotes convolution over Φ. To visualize the behavior of (18), we consider the following cases:

(i) for a complex exponential defined by $s(t) = e^{j\Omega_o t}$ and $S(\Omega) = 2\pi\delta(\Omega - \Omega_o)$, (18) simplifies to

$$P(\Phi,\Omega) = 2\pi W_x\left(\Phi - \cos\alpha\frac{\Omega_o}{c}\right) W_t(\Omega - \Omega_o); \quad (19)$$

(ii) for a Dirac pulse defined by $s(t) = \delta(t)$ and $S(\Omega) = 1$, (18) is given by

$$P(\Phi,\Omega) = 2\pi W_x\left(\Phi - \cos\alpha\frac{\Omega}{c}\right) *_{\Phi} \frac{cW_t\left(\frac{c}{|\cos\alpha|}\Phi\right)}{|\cos\alpha|}. \quad (20)$$

The result in (20) can be further simplified if the right term of the convolution is of type $(1/a)\mathrm{sinc}((1/a)\Phi)$ or $(1/a)\mathrm{sinc}^2((1/a)\Phi)$, such that $\lim_{a\to 0}(1/a)W_t((1/a)\Phi) = \delta(\Phi)$, in which case

$$P(\Phi,\Omega) = 2\pi W_x\left(\Phi - \cos\alpha\frac{\Omega}{c}\right). \quad (21)$$

Window functions of this type include the rectangular, the triangular, and the cosine windows in general. An example with rectangular windows is depicted in Fig. 5(a) and (b).

In the near-field case, the spectral pattern generated as a result of windowing is difficult to express mathematically, since it is given by a convolution between (9) and the Fourier transform of a shifted window function. However, the result can be intuitively understood as a combination of the results in Fig. 4 and Fig. 5(b), and is well approximated by the empirical model illustrated in Fig. 5(c). This model is parametric and given by

$$P(\Phi,\Omega) = S(\Omega)\max\left\{W_x\left(\Phi - \cos\alpha\frac{\Omega}{c}\right), M(\Phi,\Omega)\right\} \quad (22)$$

where $M(\Phi,\Omega)$ is a triangular mask given by

$$M(\Phi,\Omega) = \begin{cases} W_x(0), & (\Phi,\Omega) \notin \mathcal{U} \\ 0, & (\Phi,\Omega) \in \mathcal{U} \end{cases} \quad (23)$$

with $\mathcal{U} = \mathbb{R}^2 \setminus \{(\Phi,\Omega) : \Phi^{\min} \le \Phi \le \Phi^{\max}, \Omega \ge 0\}$, and point-symmetric for $\Omega < 0$. As we show next, the parameters α, $\Phi^{\min}$, and $\Phi^{\max}$ can be optimized for any given source.

Proposition 4: Define $\alpha_{\mathrm{nf}}(x) = \angle(x_o - x + jz_o)$ as the angle of incidence at point x, where $\alpha_{\mathrm{nf}}^{\min} = \alpha_{\mathrm{nf}}(0)$ is the smallest angle and $\alpha_{\mathrm{nf}}^{\max} = \alpha_{\mathrm{nf}}(L)$ is the largest angle, such that $\alpha_{\mathrm{nf}}^{\min} \le \alpha_{\mathrm{nf}}(x) \le \alpha_{\mathrm{nf}}^{\max}$. The ripples within the region $\mathcal{U}$ are oriented towards $\cos\alpha = \mathbb{E}_x[\cos\alpha_{\mathrm{nf}}(x)]$, where $\mathbb{E}_x$ denotes expectation over x, and the triangular mask is delimited by $\Phi^{\min} = \cos\alpha_{\mathrm{nf}}^{\max}(\Omega/c)$ and $\Phi^{\max} = \cos\alpha_{\mathrm{nf}}^{\min}(\Omega/c)$.

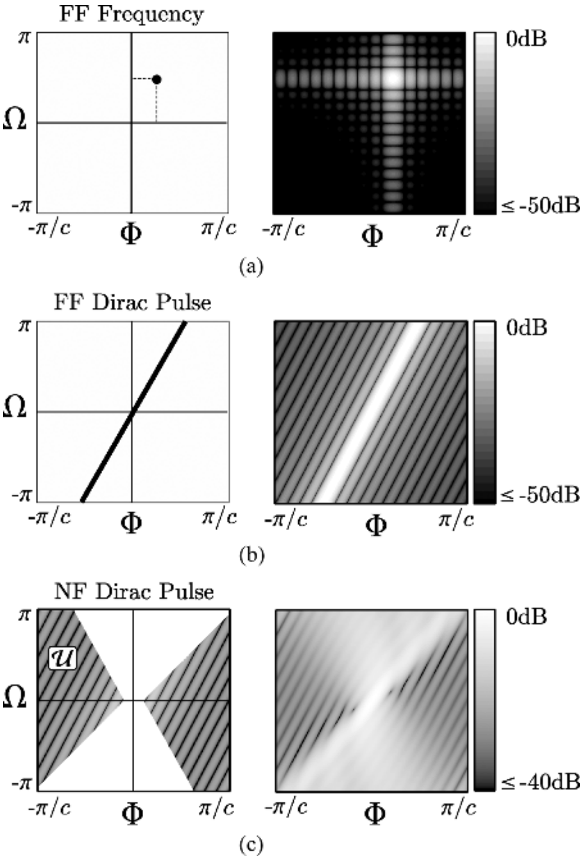


Fig. 5. Effects of rectangular windowing on $P(\Phi,\Omega)$ for: (a) $s(t) = e^{j(\pi/2)t}$ in the far-field, (b) $s(t) = \delta(t)$ in the far-field, and (c) $s(t) = \delta(t)$ in the near-field. The window functions are given by $W_x(\Phi) = \mathrm{sinc}(L_x\Phi/(2\pi))$ and $W_t(\Omega) = \mathrm{sinc}(L_t\Omega/(2\pi))$, where $L_t = L_x/c = 2\pi N_0$ with $N_0 = 8$ being half the number of zeros in the sinc function. The angle of arrival is $\alpha = \pi/4$. In the cases of (a) and (b), the theoretical results given by (7) are illustrated on the left, whereas the windowing effects given by (19) and (21) are shown on the right. In the case of (c), the theoretical result on the left represents the parametric spectral model given by (22), where $\mathcal{U}$ is the region that resembles the far-field spectral pattern, and the result on the right is a Matlab simulation of a near-field source.

Proof: Divide the window function $w_x(x)$ of length L into M segments $w_x^{(m)}(x)$ of length L/M, such that $w_x(x) = \sum_{m=0}^{M-1} w_x^{(m)}(x - m(L/M))$. For M large enough, the near-field wave fronts become increasingly far-field in the range of each segment (by definition, $\|\mathbf{r}_o\| \gg \|\mathbf{r}\|$). Thus

$$P(\Phi,\Omega) = \sum_{m=0}^{M-1} \frac{2\pi}{r^{(m)}} S(\Omega)\delta\left(\Phi - \cos\alpha^{(m)}\frac{\Omega}{c}\right) *_{\Phi} \left(W_x^{(m)}(\Phi)e^{-jm\frac{L}{M}\Phi}\right)$$

where $\alpha^{(m)} = \alpha_{\mathrm{nf}}(m(L/M))$ and $r^{(m)} = \|(m(L/M),0) - \mathbf{r}_o\|$. Note also that, as M increases, $W_x^{(m)}(\Phi)$ becomes increasingly flat with magnitude $(L/M)w_x(m(L/M))$. This simplifies the result to

$$P(\Phi,\Omega) = S(\Omega)\frac{2\pi}{M}\sum_{m=0}^{M-1} \frac{Lw_x\left(m\frac{L}{M}\right)}{r^{(m)}} e^{-jm\frac{L}{M}\left(\Phi - \cos\alpha^{(m)}\frac{\Omega}{c}\right)}. \quad (24)$$

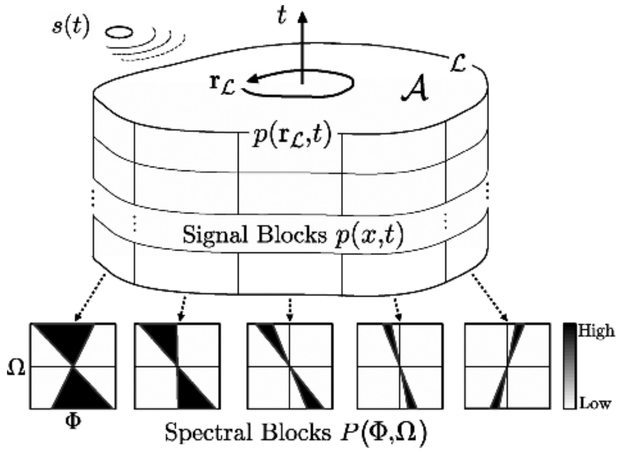

$s(t)$
t
$\mathcal{L}$
$\mathbf{r}_{\mathcal{L}}$
$\mathcal{A}$
$p(\mathbf{r}_{\mathcal{L}},t)$
Signal Blocks $p(x,t)$
High
Low
Ω
Φ
Spectral Blocks $P(\Phi,\Omega)$

Fig. 6. Space-time manifold generated on a smooth contour $\mathcal{L}$, where $p(\mathbf{r}_{\mathcal{L}}, t)$ is locally approximated by several spatio-temporal blocks $p(x, t)$. The resulting spectral blocks $P(\Phi, \Omega)$ have different plane-wave content depending on the local properties of the wave field. In this example, the blocks on the left have more near-field characteristics than the blocks on the right, where the energy is more concentrated around the dominant direction.

The notches in $\mathcal{U}$ occur when the sum in (24) is minimized, i.e., when the exponential vectors are in maximum phase opposition. This requires a minimization of $\mathbb{E}[L(\Phi - \cos\alpha^{(m)}(\Omega/c)) - k2\pi]$ for $k \in \mathbb{Z} \setminus 0$, which occurs at $\Phi = \mathbb{E}[\cos\alpha^{(m)}](\Omega/c) + (k2\pi/L)$; hence the orientation towards $\mathbb{E}[\cos\alpha^{(m)}]$. When $\Phi = \cos\alpha^{(m)}(\Omega/c), \forall m$, at least one exponential vector in the sum equals 1. Since $\alpha_{\text{nf}}(x)$ is a smooth function, the other $M - 2$ vectors are also concentrated in the vicinity of 1. On the contrary, as $|\Phi - \cos\alpha^{(m)}(\Omega/c)|$ increases, the vectors become more dispersed in the complex plane, and the sum decreases in magnitude. This places the optimal limits at $\Phi^{\min} = \cos\alpha_{\text{nf}}^{\max}(\Omega/c)$ and $\Phi^{\max} = \cos\alpha_{\text{nf}}^{\min}(\Omega/c)$. ■

Similarly to Gabor's generalization of the concept of frequency [4], the results in (19), (21), and (22) can be viewed as generalizations of the concepts of plane wave, far-field component, and near-field component in the space-time-frequency representation space. The near-field component itself generalizes the far-field component, obtained when $\alpha_{\text{nf}}(x) \to \alpha$. The various parameters regulate the tradeoff between spatio-temporal resolution and frequency resolution, thus allowing a selective discrimination of the local properties of the wave field across space and time (e.g., "directional versus diffuse" relative to space and "harmonic versus noisy" relative to time).

Furthermore, the use of space-time-frequency representation allows an extension of the Fourier analysis to the more general case where $\mathcal{L}$ is not a straight line but a curved contour, as in the case of Fig. 1. If $\mathcal{L}$ is smooth enough, the space-time manifold defined by $\mathbf{r}_{\mathcal{L}} = (x_{\mathcal{L}}, z_{\mathcal{L}})$ and t can be locally approximated by an Euclidean space where z is dropped and x represents the local tangent to $\mathcal{L}$. This is illustrated in Fig. 6. Note, however, that the approximation of the curved contour by a straight line reduces the sharpness of the spectrum across Φ. This can be compensated, for example, by varying the window size according to the local smoothness of $\mathcal{L}$.

III. Subband Analysis of the Wave Field Spectra

In the previous section, we have introduced the theoretical background of wave field processing in the continuous space-time domain. The following step, enabling us to target real applications, is understanding how these theoretical concepts can be translated into computational algorithms that operate in the discrete space-time domain, i.e., with discretized versions of x and t. For this purpose, we present three filter bank structures that achieve the desired subband partitions in the space-time-frequency representation space. The first is a basic orthogonal transform filter bank that can be customized to perform a plane-wave expansion of the input signal—for example, by using a DFT basis function. The second is a variation of the transform filter bank that performs plane-wave expansion with block overlapping in space and time while preserving critical sampling and perfect reconstruction, making the transform more suitable for coding applications. Finally, we present a nonseparable filter bank that decomposes the input signal into far-field components, based on a tree-structured implementation of the quincunx filter bank.

A. Sampling and Reconstruction

From a signal processing perspective, the spatial axis is just another dimension of the input signal, which must be sampled in order to be processed. Similarly to the temporal axis, the sampling pattern can be chosen freely according to the specifications of the problem. Uniform sampling, compressed sensing, and finite rate of innovation are examples of sampling techniques that can be applied to the spatial domain. In this paper, we assume uniform sampling in both dimensions.

The Nyquist conditions in the space-time domain are given by $\Phi_S \geq 2(\Omega_M/c)$ and $\Omega_S \geq 2\Omega_M$, where Φ_S and Ω_S are the spatial and temporal sampling frequencies, and Ω_M is the maximum temporal frequency. As a result of sampling, an infinite number of spectral repetitions show up at multiples of Φ_S and Ω_S. Aliasing effects can thus occur in either dimension if the respective Nyquist conditions are not satisfied. Conversely, if the two conditions are met, the signal can be perfectly reconstructed in both dimensions.

For a more detailed analysis on spatio-temporal sampling, the reader may refer to the work of Ajdler *et al.* [11] on the plenacoustic function.

B. Orthogonal Transform Filter Banks

A spatio-temporal transform can be obtained through any combination of orthogonal bases applied separately to the spatial and temporal dimensions. Examples of transforms that can be used to exploit the temporal evolution of the sound field include the discrete Fourier transform (DFT), the discrete cosine transform (DCT), and the discrete wavelet transform (DWT). In general, both the DFT and the DCT are better suited for audio and speech sources, due to their harmonic nature, whereas the DWT is better suited for impulsive and transient-like sources (e.g., shot events).

In the spatial domain, the choice of basis takes into account other factors, such as the position of the sources and the geometry of the acoustic environment (which influence the diffuseness of sound and the curvature of the wave field), as well as the geometry of the observation contour $\mathcal{L}$. The Fourier transform, as we have shown, provides an efficient representation of the wave field on a straight line or a smoothly curved contour. However, under some particular conditions, a different choice of

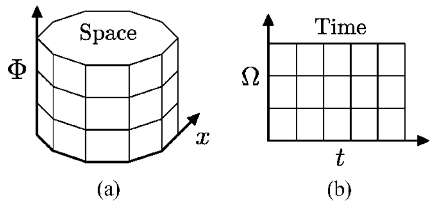


Fig. 7. Space-time-frequency tiling that characterizes a Fourier-based uniform filter bank, illustrated here as two separate 2-D tilings: (a) the space dimension is mapped onto the spatial frequency dimension; (b) the time dimension is mapped onto the temporal frequency dimension. The x axis is bent to emphasize that it may be defined on a curved contour, and not necessarily on a straight line.

basis may prove to be more efficient. For instance, if $\mathcal{L}$ is perfectly circular [11], a Fourier transform for circularly symmetric functions, such as the Hankel transform, can be used instead. In contrast, if $\mathcal{L}$ is sharply curved, the wavelet transform is more able to represent the sharp transitions across space in the wave field representation. This is an interesting topic to be addressed in future research.

Consider that $p[\mathbf{n}] = p_{n_x,n_t}$ represents the uniformly sampled version of $p(x,t)$ such that $p[\mathbf{n}] = p(n_x(2\pi/\Phi_S), n_t(2\pi/\Omega_S))$, where $\mathbf{n} = (n_x, n_t)$ are the space and time sample indexes, and $P[\mathbf{b}] = P_{b_x,b_t}$ the respective transform-domain coefficients with indexes $\mathbf{b} = (b_x, b_t)$. A spatio-temporal transform filter bank is characterized by a four-dimensional tiling that spans the variables x, t, Φ, and Ω, representing the mapping of spatio-temporal partitions onto the transform-domain subbands. In the case of Fourier-based transforms, the mapping is separable and uniform, as illustrated in the tiling of Fig. 7.

Consider two general orthogonal bases υ_{b_x,n_x} and ψ_{b_t,n_t} such that

$$P[\mathbf{b}] = \sum_{\mathbf{n}\in\mathbb{Z}^2} p[\mathbf{n}]\upsilon^*_{b_x,n_x}\psi^*_{b_t,n_t}, \quad \mathbf{b}\in\mathbb{Z}^2 \tag{25}$$

and

$$p[\mathbf{n}] = \sum_{\mathbf{b}\in\mathbb{Z}^2} P[\mathbf{b}]\upsilon_{b_x,n_x}\psi_{b_t,n_t}, \quad \mathbf{n}\in\mathbb{Z}^2. \tag{26}$$

In matrix notation, (25) and (26) can be written as

$$\mathbf{Y} = \boldsymbol{\Upsilon}\mathbf{P}\boldsymbol{\Psi}^H \tag{27}$$

and

$$\mathbf{P} = \boldsymbol{\Upsilon}^H\mathbf{Y}\boldsymbol{\Psi} \tag{28}$$

where $\mathbf{P}$, $\mathbf{Y}$, $\boldsymbol{\Upsilon}$, and $\boldsymbol{\Psi}$ are the matrix expansions of $p[\mathbf{n}]$, $P[\mathbf{b}]$, υ_{b_x,n_x}, and ψ_{b_t,n_t}, respectively. The same matrix operations can be expressed in the filter bank structure of Fig. 8, where the input signal $p[\mathbf{n}]$ of size $N_x \times N_t$ is decomposed into a transform matrix $P[\mathbf{b}]$ of equal size.

To perform a spatio-temporal DFT, the basis functions are defined as $\upsilon_{b_x,n_x} = e^{j(2\pi/N_x)b_xn_x}$ and $\psi_{b_t,n_t} = e^{j(2\pi/N_t)b_tn_t}$, which implies that $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Psi}$ are DFT matrices of size $N_x \times N_x$ and $N_t \times N_t$ respectively. The DFT filter bank performs a plane-wave expansion in the discrete space-time domain.

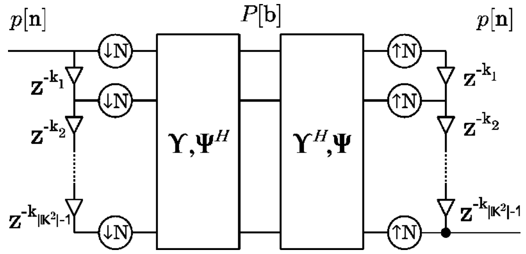

$p[\mathbf{n}]$
$P[\mathbf{b}]$
$p[\mathbf{n}]$
$\downarrow$N
$\mathbf{z}^{-\mathbf{k}_1}$
$\mathbf{z}^{-\mathbf{k}_2}$
$\mathbf{z}^{-\mathbf{k}_{|\mathbb{K}^2|-1}}$
$\boldsymbol{\Upsilon},\boldsymbol{\Psi}^H$
$\boldsymbol{\Upsilon}^H,\boldsymbol{\Psi}$
$\uparrow$N

Fig. 8. Spatio-temporal uniform filter bank that decomposes $p[\mathbf{n}]$ into $N_x \times N_t$ frequency subbands. The delay chain that generates a signal block $\mathbf{P}$ is composed of a series of z-transform delay factors [15] defined by $\mathbf{z}^{-\mathbf{k}} = z_x^{-k_x} z_t^{-k_t}$, followed by a downsampling matrix $\mathbf{N} = \begin{bmatrix} N_x & 0 \\ 0 & N_t \end{bmatrix}$, where $\mathbf{k} = \left[\begin{smallmatrix} k_x \\ k_t \end{smallmatrix}\right]$ are the coset vectors in the coset space $\mathbb{K}^2 \subset \mathbb{Z}^2$ generated by $\mathbf{N}$. The polyphase blocks $\boldsymbol{\Upsilon}, \boldsymbol{\Psi}^H$ and $\boldsymbol{\Upsilon}^H, \boldsymbol{\Psi}$ represent the separable matrix operations of (27) and (28).

C. Spatio-Temporal Lapped Orthogonal Transforms

In many applications, the usage of short-time Fourier analysis requires that consecutive blocks are overlapped in time in order to avoid discontinuities in the reconstructed signal. The same argument applies to the spatial dimension, where overlapping helps to preserve the curvature of the wave field. Looking at Fig. 8, it is clear that overlapping can be obtained by applying the resampling matrix $\mathbf{N} - \mathbf{O}$ instead of $\mathbf{N}$, where $\mathbf{N} = \begin{bmatrix} N_x & 0 \\ 0 & N_t \end{bmatrix}$ and $\mathbf{O} = \begin{bmatrix} O_x & 0 \\ 0 & O_t \end{bmatrix}$ contains the number of overlapping samples O_x and O_t in each dimension. Without loss of generality, we assume that $\mathbf{O} = (1/2)\mathbf{N}$, representing 50% of overlapping in both dimensions.

An additional requirement, typically related to audio coding, is that the output of a lapped transform is critically sampled and, yet, perfectly reconstructible. In most cases, these two conditions can not be met simultaneously, since making $P[\mathbf{b}]$ critically sampled implies subsampling the transform by a factor of 2 in both dimensions. However, perfect reconstruction can be achieved if the aliasing generated by the inverse transform is canceled out in the overlap-and-add operation—a technique known as time-domain aliasing cancelation [16]. This depends on a proper choice of the basis functions υ_{b_x,n_x} and ψ_{b_t,n_t}.

There are many examples in the literature of lapped orthogonal transforms that meet the above requirements, both in the 1-D case (e.g., Princen *et al.* [16], Malvar [17], and Schuller *et al.* [18]) and the 2-D case (e.g., Kovacevic *et al.* [19] and Johnson *et al.* [20]). In this work, we focus on the so-called modified discrete cosine transform (MDCT) [16], used in current state-of-art audio coders. In effect, it is possible to apply an MDCT separably to the spatial and temporal dimensions by defining the basis functions as

$$\upsilon_{b_x,n_x} = w_x[n_x]\sqrt{\frac{4}{N_x}}\cos\left(\frac{2\pi}{N_x}\left(n_x + \frac{N_x}{4} + \frac{1}{2}\right)\left(b_x + \frac{1}{2}\right)\right),$$
$$b_x = 0,\ldots,\frac{N_x}{2} - 1 \text{ and } n_x = 0,\ldots,N_x - 1 \tag{29}$$

and

$$\psi_{b_t,n_t} = w_t[n_t]\sqrt{\frac{4}{N_t}}\cos\left(\frac{2\pi}{N_t}\left(n_t+\frac{N_t}{4}+\frac{1}{2}\right)\left(b_t+\frac{1}{2}\right)\right),$$
$$b_t=0,\dots,\frac{N_t}{2}-1 \text{ and } n_t=0,\dots,N_t-1 \quad (30)$$

where $w_x[n_x]$ and $w_t[n_t]$ are the window functions in space and time, satisfying the conditions $w[n] = w[N-1-n]$ and $w^2[n] + w^2[n+(N/2)] = 1$ [6].

The decomposition of $p[\mathbf{n}]$ into overlapped blocks $p_{\mathbf{i}}[\mathbf{n}]$ can be written as

$$p_{\mathbf{i}}[\mathbf{n}] = p[\mathbf{n}], \quad \mathbf{n} = \frac{\mathbf{N}}{2}\mathbf{i},\dots,\frac{\mathbf{N}}{2}(\mathbf{i}+\mathbf{2})-\mathbf{1}, \quad \mathbf{i}\in\mathbb{I}^2 \quad (31)$$

where $\mathbf{i} = \left[\begin{smallmatrix} i_x \\ i_t \end{smallmatrix}\right]$ is the block index and $\mathbb{I}^2 \subset \mathbb{Z}^2$ is the respective set of block indexes. The notation $\mathbf{n} = (\mathbf{N}/2)\mathbf{i},\dots,(\mathbf{N}/2)(\mathbf{i}+\mathbf{2})-\mathbf{1}$ means that $n_x = (N_x/2)i_x,\dots,(N_x/2)(i_x+2)-1$ and $n_t = (N_t/2)i_t,\dots,(N_t/2)(i_t+2)-1$. The vector integers are defined as $\mathbf{0} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, $\mathbf{1} = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, and so on. Note also that, in order to handle the blocks that go outside the boundaries of $\mathbf{n}$, we consider the signal to be circular (or periodic) in both dimensions. This presents an advantage over zero-padding, in particular, when $\mathcal{L}$ is closed.

Denoting $\varphi[\mathbf{b},\mathbf{n}] = \upsilon_{b_x,n_x}\psi_{b_t,n_t}$, the direct and inverse transforms for each block are given by

$$P_{\mathbf{i}}[\mathbf{b}] = \sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{N}\mathbf{1}-\mathbf{1}} p_{\mathbf{i}}[\mathbf{n}]\varphi[\mathbf{b},\mathbf{n}], \quad \mathbf{b} = \mathbf{0},\dots,\frac{\mathbf{N}}{\mathbf{2}}\mathbf{1}-\mathbf{1} \quad (32)$$

and

$$\hat{p}_{\mathbf{i}}[\mathbf{n}] = \sum_{\mathbf{b}=\mathbf{0}}^{\frac{\mathbf{N}}{2}\mathbf{1}-\mathbf{1}} P_{\mathbf{i}}[\mathbf{b}]\varphi[\mathbf{b},\mathbf{n}], \quad \mathbf{n} = \mathbf{0},\dots,\mathbf{N}\mathbf{1}-\mathbf{1}. \quad (33)$$

Finally, the reconstruction of $p[\mathbf{n}]$ through overlap-and-add is given by

$$p[\mathbf{n}] = \sum_{\mathbf{i}\in\mathbb{I}^2}\hat{p}_{\mathbf{i}}\left[\mathbf{n}-\frac{1}{2}\mathbf{N}\mathbf{i}\right], \quad \mathbf{n}\in\mathbb{Z}^2. \quad (34)$$

In matrix notation, the complete mechanism of the spatio-temporal MDCT can be expressed as

$$\begin{bmatrix} \cdots & \vdots & \cdots \\ \mathbf{Y}_{\mathbf{i}} & \mathbf{Y}_{\mathbf{i}+\begin{bmatrix}0\\1\end{bmatrix}} & \\ \mathbf{Y}_{\mathbf{i}+\begin{bmatrix}1\\0\end{bmatrix}} & \mathbf{Y}_{\mathbf{i}+\mathbf{1}} & \\ \cdots & \vdots & \cdots \end{bmatrix} = \begin{bmatrix} \ddots & & \\ \underset{L}{\boldsymbol{\Upsilon}}\;\underset{R}{\boldsymbol{\Upsilon}} & & \\ & \underset{L}{\boldsymbol{\Upsilon}}\;\underset{R}{\boldsymbol{\Upsilon}} & \\ & & \ddots \end{bmatrix} \begin{bmatrix} \cdots & \vdots & \cdots \\ \mathbf{P}_{\mathbf{i}} & \mathbf{P}_{\mathbf{i}+\begin{bmatrix}0\\1\end{bmatrix}} & \\ \mathbf{P}_{\mathbf{i}+\begin{bmatrix}1\\0\end{bmatrix}} & \mathbf{P}_{\mathbf{i}+\mathbf{1}} & \\ \cdots & \vdots & \cdots \end{bmatrix} \begin{bmatrix} \ddots & & \\ \underset{L}{\boldsymbol{\Psi}}\;\underset{R}{\boldsymbol{\Psi}} & & \\ & \underset{L}{\boldsymbol{\Psi}}\;\underset{R}{\boldsymbol{\Psi}} & \\ & & \ddots \end{bmatrix}^T$$

and

$$\begin{bmatrix} \cdots & \vdots & \cdots \\ \mathbf{P}_{\mathbf{i}} & \mathbf{P}_{\mathbf{i}+\begin{bmatrix}0\\1\end{bmatrix}} & \\ \mathbf{P}_{\mathbf{i}+\begin{bmatrix}1\\0\end{bmatrix}} & \mathbf{P}_{\mathbf{i}+\mathbf{1}} & \\ \cdots & \vdots & \cdots \end{bmatrix} = \begin{bmatrix} \ddots & & \\ \underset{L}{\boldsymbol{\Upsilon}}\;\underset{R}{\boldsymbol{\Upsilon}} & & \\ & \underset{L}{\boldsymbol{\Upsilon}}\;\underset{R}{\boldsymbol{\Upsilon}} & \\ & & \ddots \end{bmatrix}^T \begin{bmatrix} \cdots & \vdots & \cdots \\ \mathbf{Y}_{\mathbf{i}} & \mathbf{Y}_{\mathbf{i}+\begin{bmatrix}0\\1\end{bmatrix}} & \\ \mathbf{Y}_{\mathbf{i}+\begin{bmatrix}1\\0\end{bmatrix}} & \mathbf{Y}_{\mathbf{i}+\mathbf{1}} & \\ \cdots & \vdots & \cdots \end{bmatrix} \begin{bmatrix} \ddots & & \\ \underset{L}{\boldsymbol{\Psi}}\;\underset{R}{\boldsymbol{\Psi}} & & \\ & \underset{L}{\boldsymbol{\Psi}}\;\underset{R}{\boldsymbol{\Psi}} & \\ & & \ddots \end{bmatrix}$$

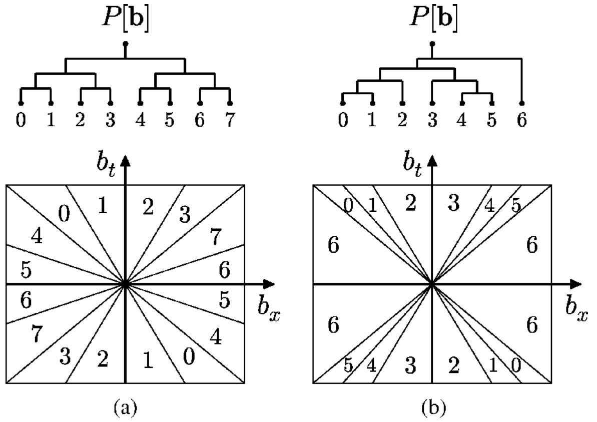


Fig. 9. Decomposition of the space-time spectra into directional subbands using a tree-structured directional filter bank (DFB). The subband partitioning can be either: (a) uniform or (b) nonuniform. A partitioning similar to (b) is more suitable for wave field spectral patterns, since it gathers all the evanescent waves into a single subband and has variable directional resolution in the plane-wave region.

where $\boldsymbol{\Psi}$ and $\boldsymbol{\Upsilon}$ are split into left and right halves in order to enforce overlapping directly in the transformation matrix. The resulting block-bidiagonal matrices are orthogonal.

D. Nonseparable Directional Filter Banks

An important property derived in Section II-C is the decomposition of the wave field into far-field components. According to Proposition 2 and 3, the far-field components can be obtained from the spectrum by properly choosing the integration limits. If the integration area is increased, covering a larger range of angles, there is a loss of directional resolution. This also reduces the sensitivity to proximity and movement of the source.

Conceptually, one can think of a filter bank that decomposes the spectrum into directional subbands defined in the range $\cos(\alpha+(\Delta/2))(\Omega/c) \leq \Phi \leq \cos(\alpha-(\Delta/2))(\Omega/c)$, for $\Omega \geq 0$ (and point-symmetric for $\Omega < 0$), where α is the central direction and Δ is the directional bandwidth, as illustrated in Fig. 9(a). Due to Definition 1, this partitioning can also be interpreted as a four-dimensional tiling that spans the variables x, t, α, and Ω. If we assume that each pair of subbands is obtained by slicing a larger band in half, the filter bank can be designed as an iterated two-channel structure, providing more flexibility to obtain nonuniform directional decompositions such as the one in Fig. 9(b). This way, the biggest effort goes into designing a two-channel filter bank that can be used in all nodes of the tree.

Such a filter bank has been extensively studied [21]–[24], and is known as quincunx filter bank (QFB). The QFB is a nonseparable perfect reconstruction filter bank, maximally decimated, defined by two diamond-shaped half-band filters, $H_0(\mathbf{z})$ and $H_1(\mathbf{z})$, preceded by a parallelogram resampler $\mathbf{R}$ and followed by a quincunx resampler $\mathbf{Q}$ (see Do *et al.* [23]). The directional filter bank (DFB) is then obtained with a tree-structure of multiple QFB, as illustrated in Fig. 10.

The intuition behind the iterated QFB structure is that, instead of using different filters to obtain different subbands, we use the resampling matrices to rotate and skew the subbands before slicing them with a fixed filter. The design of the half-band filters $H_0(\mathbf{z})$ and $H_1(\mathbf{z})$, and the respective synthesis filters, can

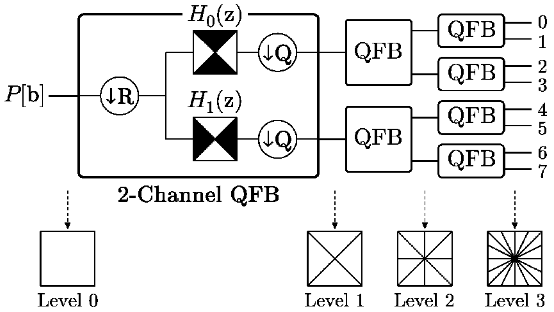


Fig. 10. Three-level iterated quincunx filter bank that performs a uniform decomposition of the input spectra into directional subbands. The matrices $\mathbf{R}$ and $\mathbf{Q}$ represent a parallelogram resampler followed by a quincunx resampler, and the filters $H_0(\mathbf{z})$ and $H_1(\mathbf{z})$ represent two diamond-shaped half-band filters.

be done using two-dimensional filter design techniques. In the simulations shown in this paper, we use a technique introduced by Phoong *et al.* [22].

IV. Applications of Space–Time–Frequency Processing

We briefly present some of the applications of space-time-frequency processing that make use of the theory and algorithms derived in this paper. In particular, we focus on three classes of applications: spatial filtering, deconvolution, and wave field coding.

A. *Spatial Filtering*

In digital signal processing, filtering is the cornerstone operation when it comes to manipulating signals, images, and ultimately acoustic wave fields. The use of spatial filtering to separate sources from each other is not a new concept, and several innovative techniques (commonly known as beamforming) have been proposed in the past (e.g., Frost [25], Widrow *et al.* [26] and Griffiths *et al.* [9]). However, from a conceptual point of view, these techniques rely on a nonintuitive interpretation of spatial filtering based on convolution—with a few exceptions (e.g., Start *et al.* [27] explore the idea of frequency-domain spatial filtering in the context of wave field synthesis).

One of the advantages of the Fourier transform is that it allows the interpretation of convolutional filtering in terms of more intuitive concepts. For instance, a filter can be sketched in the Fourier domain such that it has a unitary response for a given range of frequencies (pass-band) and a high attenuation for the remaining frequencies (stop-bands), plus an equiripple magnitude response and a linear phase. Using existing algorithms [28], the ideal filter can be translated into a realizable filter that optimally obtains the desired response. In the context of space-time analysis, we have seen that the Fourier transform provides a sparse representation of the elementary components of the wave field. Using the same reasoning, we can sketch a spatial filter in the Fourier domain such that it has a unitary response for every plane wave within a given range of directions (pass-band) and a high attenuation for the remaining plane waves (stop-bands), plus any additional magnitude and phase constraints. The ideal magnitude response of such a filter is given by the spectral mask defined in (23).

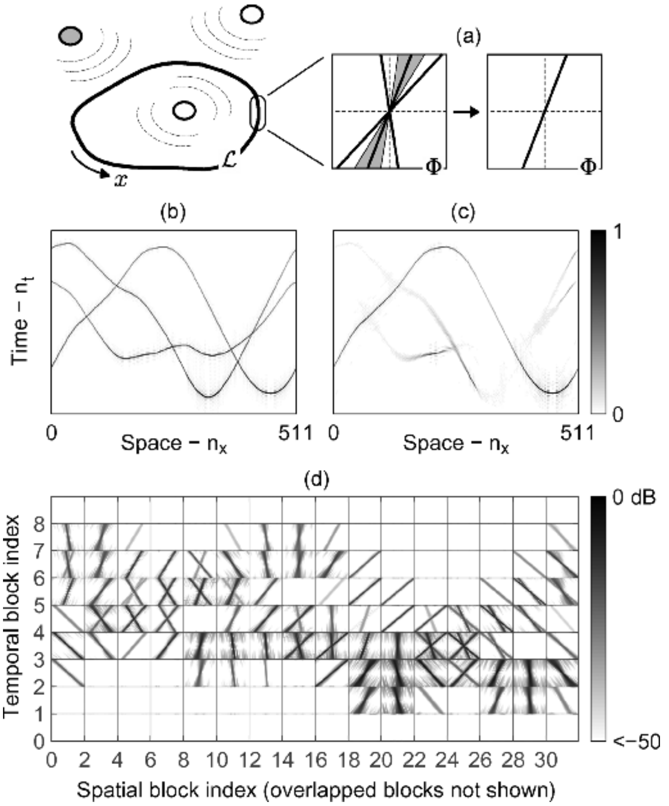


Fig. 11. Spatial filtering on a curved contour $\mathcal{L}$ (the same used in previous figures) with 512 spatial samples satisfying Nyquist, in a scene with three Dirac sources. The goal is to apply a bandpass filter to the shaded source, in order to eliminate the other two. For each spatial block, the filter takes a different shape according to the relative direction of the desired source. An example for a random block is shown in (a), where the pass-band frequency range is given by the shaded region. The signals at the input and output of the complete filtering process are shown in (b) and (c), respectively. Note that the result in (c) would not be possible to obtain by simply taking the Fourier transform along the entire contour, which would result in a severely blurred spectrum. Instead, the spatial filters are applied on the Gabor decomposition shown in (d), which provides a sharper separation of the three sources in most of the blocks.

Example 1: Consider a sampling line defined between $x = 0$ and 1 m. The goal is to filter the wave fronts radiating from $(x_o, z_o) = (0.75$ m, 0.433 m$)$. It can be easily verified that the angle of incidence at point $x \in [0, 1]$ varies within the range $(\pi/6) \leq \alpha_{\mathrm{nf}}(x) \leq (2\pi/3)$. Thus, from (23), the desired filter has an ideal pass-band region defined by $\cos(2\pi/3)(\Omega/c) \leq \Phi \leq \cos(\pi/6)(\Omega/c)$ for $\Omega \geq 0$, and point-symmetric at $\Omega < 0$.

In the general case where the wave field is sampled along a curved contour $\mathcal{L}$, the tangential lines of $\mathcal{L}$ are likely to be facing the sources from many different angles. Thus, if the Fourier transform is taken along the entire contour, the resulting spectrum may be too blurred to allow any distinction between sources. One solution is to perform a localized Gabor-style analysis, as proposed in this paper, by windowing the signal along the spatial dimension, filtering each block individually, and reconstructing the signal back to its original size. An example is shown in Fig. 11.

B. *Deconvolution*

Another operation that can be conveniently performed in the space-time domain is signal deconvolution. The idea of deconvolution is particularly useful in the parametrization of the acoustic scene, where the goal is to estimate parameters such as the position of the sources or the characteristics of the room. In general, the problem consists of estimating a number

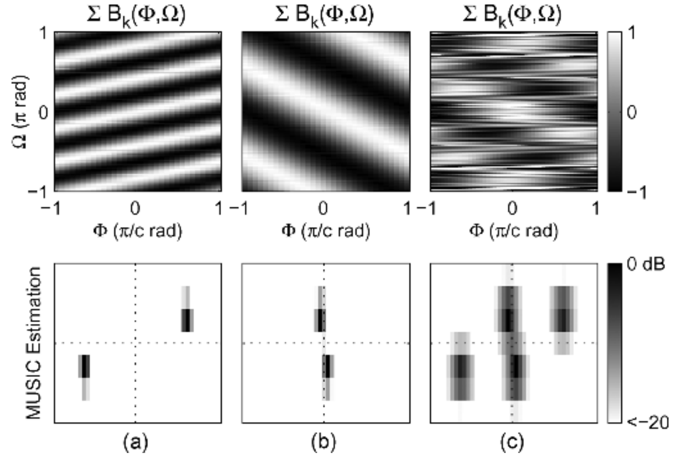


Fig. 12. Estimation of the parameters α_k of $B_k(\Phi, \Omega)$ in a wave field with two far-field white noise sources at $\alpha_1 = \pi/4$ and $\alpha_2 = 2\pi/3$, using two spatial samples separated by 8 times the Nyquist spacing (this adds more periods to the support functions). In the three cases shown, the acoustic scene consists of (a) a single source at $\pi/4$, (b) a single source at $2\pi/3$, and (c) the two sources at $\pi/4$ and $2\pi/3$. The result, as expected, is a perfect sinusoid in (a) and (b), and a sum of distorted sinusoids in (c). The respective frequencies are estimated with a spatio-temporal version of the MUSIC algorithm, from which the parameters α_k can be directly obtained by peak detection and proper scaling.

of impulse responses affecting each source signal at the scene, having only the knowledge of $p(x, t)$. Equivalently, it consists of estimating a number of support functions $B_k(\Phi, \Omega)$ such that $P(\Phi, \Omega) = \sum_k S_k(\Omega) B_k(\Phi, \Omega)$.

Assuming that the signal components $S_k(\Omega)$ carry no useful information for this purpose, they can be attenuated by filtering $P(\Phi, \Omega)$ with $1/\bar{S}(\Omega)$, where $\bar{S}(\Omega) = \sum_k S_k(\Omega) \approx p(x, \Omega)|_{x=0}$. An estimation of $\sum_k B_k(\Phi, \Omega)$ is then given by

$$\frac{P(\Phi, \Omega)}{\bar{S}(\Omega)} = \frac{1}{1 + \frac{\bar{S}(\Omega)}{S_0(\Omega)}} B_0(\Phi, \Omega) + \frac{1}{1 + \frac{\bar{S}(\Omega)}{S_1(\Omega)}} B_1(\Phi, \Omega) + \cdots \tag{35}$$

where $(1 + (\bar{S}(\Omega)/S_k(\Omega)))^{-1}$ is a frequency-dependent distortion function that equals one if $S_k(\Omega) \rightarrow \bar{S}(\Omega)$ and zero if $S_k(\Omega) \rightarrow 0$. This suggests that, for each frequency, the estimation of $B_k(\Phi, \Omega)$ is more accurate for the most dominant source. In general, the support functions $B_k(\Phi, \Omega)$ can be modeled as parametric functions given by (22) and (23), where the angle α and the distance $\|\mathbf{r}_o\|$ are the parameters to estimate (see, e.g., Pinto *et al.* [29]). In some cases, however, the problem can be simplified.

Example 2: Suppose the sources are in the far-field, and $p(x, t)$ is weighted by a rectangular window of length 2. If x is discretized, as is the case in practice, the support function for each source k is given by a sinc function repeated at integer multiples of 2π, and hence $B_k(\Phi, \Omega) = \sum_{r \in \mathbb{Z}} 2\text{sinc}(((\Phi - 2\pi r)/\pi) - (\cos\alpha_k/(c\pi))\Omega)$. Using the properties of the cotangent function $\cot(a/2) = \sum_{r \in \mathbb{Z}} (2/(a + 2\pi r))$ and $\cot(a/2) = (\cos a + 1)/\sin a$ [30], it can easily be shown that $B_k(\Phi, \Omega) = \cos(\Phi - (\cos\alpha_k/c)\Omega) + 1$. Thus, using only two spatial samples, (35) is a sum of distorted sinusoids with one degree of freedom, α_k, which can be estimated efficiently using the annihilating filter method [31] or the MUSIC algorithm [8]. An example is shown in Fig. 12.

C. Wave Field Coding

The third problem we address in this paper is related to the amount of information that is contained in the spatio-temporal representation of the wave field. We elaborate on a concept introduced in previous work [32], [33] which consists of compressing the wave field through plane-wave encoding, and obtaining a rate-distortion function at the output.

As discussed in previous sections, plane waves are the elementary components in the spatio-temporal analysis of the wave field, the same way frequencies are the elementary components in the temporal analysis of signals. In continuous space and time, assuming $\mathcal{L}$ to be the x-axis, the spectral representation of a given number of plane waves consists of an equal number of Diracs points. The same is true in the directional representation of the wave field, as $P(\alpha, \Omega)$ is simply a scaled version of $P(\Phi, \Omega)$ in polar coordinates. Thus, we can conclude that having a spatial dimension in the signal does not increase the amount of information in the spectrum, other than the position of the Diracs across the Φ or α-axis.

In discrete space and time, however, if the spatial axis is windowed, the plane waves are not based on ideal Diracs but on smooth support functions. This means that, in order to reconstruct the original signal with low distortion, the entire support functions must be encoded, and the number of transform coefficients to encode increases with the number of spatial points. On the contrary, if a higher distortion is tolerated, we can take advantage of the fact that most of the energy is concentrated in the main lobe of the support functions, in which case the coarse quantization will set to zero most of the side-lobe values. Thus, we can expect that as the bit-rate decreases the amount of spectral information to encode tends to a single point for each plane wave, as in the continuous space-time case.

The example of Fig. 13 compares the rate-distortion functions obtained by encoding a wideband source in the MDCT and DFB domains, using the mean square error (MSE) distortion metric. As the figure shows, the two filter banks have a similar behavior in terms of coding gain. However, the use of these filter banks in the context of a coding application has different advantages. On the one hand, the MDCT is a critically sampled Gabor-style transform that allows overlapping between blocks in both space and time domains, as opposed to the DFB which does not preserve critical sampling if overlapping is used. This gives the MDCT an advantage in terms of coding gain. On the other hand, in a perceptual audio coding application [34], the DFB provides a more suitable representation of the wave field in terms of directional sound, which simplifies the use of psychoacoustic models. For instance, one could combine a frequency masking model on the Ω axis (see Bosi *et al.* [34]) with a directional masking model on the α axis (e.g., Blauert [35]). A cascade of the two filter banks is also an interesting case to consider (see, e.g., Pinto *et al.* [36] and Eslami *et al.* [37]).

Note in particular that, for a high number of spatial points, assuming Nyquist is satisfied, it is likely that at least half of the entire spectrum has very low energy values (see Fig. 4). This

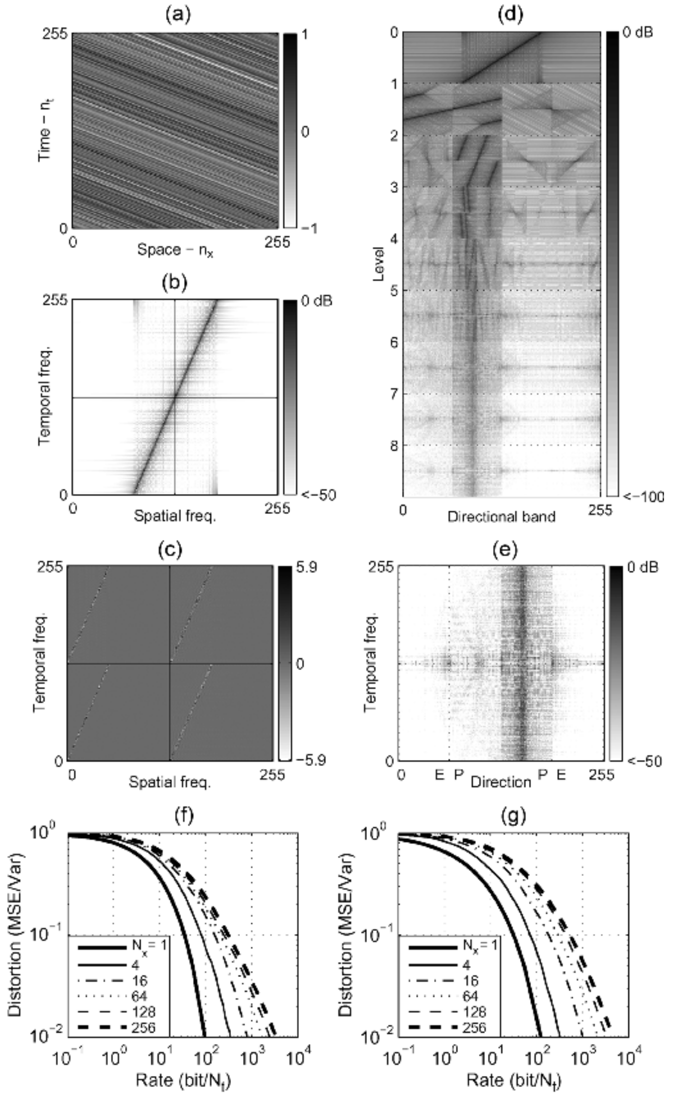


Fig. 13. Example of encoding of a far-field white noise source at $\alpha = \pi/3$, in both the MDCT and uniform DFB domains. The input signal of (maximum) size 256 × 256 and the Fourier transform are shown in (a) and (b) respectively. The MDCT with circular block overlapping is shown in (c). The DFB decomposition levels are shown in (d), and the output level is shown in (e) after a correct rearrangement of the subbands; the propagating and evanescent regions are denoted by P and E in the direction axis. Finally, the rate-distortion curves obtained for a different number of spatial points N_x (using a fixed codebook) are shown in (f) and (g) for the MDCT and DFB cases, respectively. In both cases, the increase of N_x also increases the number of bits required, but the curve eventually converges to an upper-bound. The reason is that, even though doubling N_x also duplicates the number of transform coefficients, the support functions are narrowed to half the width, and the tradeoff tends to balance itself out. Thus, increasing N_x pass a certain limit does not increase the spectral information. It can also be observed that for lower bit-rates—in the order of those used by perceptual audio coders [34]—the difference between one channel and a large number of channels is low in terms of MSE. Comparing the coding results in the MDCT and DFB domains, the only difference is that the DFB has a slightly higher upper-bound, mostly due to the smoothing effect of the half-band filters.

results in an effective low-MSE coding gain of 2 when transforming the signal to the MDCT or DFB domains. In the case of the DFB, this inherent gain can be exchanged by a 50% overlapping in one of the domains.

V. Conclusion

We have presented a signal processing framework that is suitable for analyzing, processing, and encoding acoustic wave fronts, based on a multidimensional spectral representation of the wave field over a spatio-temporal manifold. Based on previous works [10], [11] which show that the wave field can be sampled and interpolated using Nyquist theory and wave field synthesis, we derived new methods of expressing the wave field in a way that is intuitive in the context of Fourier theory and signal processing. In particular, we have shown that the wave field can be expressed as a function of elementary "spatio-temporal frequencies" called *plane waves* and elementary directional components called *far-field components*. We generalized these into the windowed case, where a local analysis of the wave field requires a parametrization of the elementary components to account for space-time-frequency resolution tradeoffs. In the discrete domain, where real applications operate, a special emphasis was given to the discussion of spatio-temporal orthogonal filter banks, which are characterized by a four-dimensional space-time-frequency mapping. In particular, we have shown that a plane-wave expansion can be obtained with a separable uniform filter bank (e.g., a DFT filter bank), whereas a decomposition into directional far-field components requires a nonseparable 2-D filter bank (e.g., an iterated quincunx filter bank). We also presented a spatio-temporal lapped orthogonal transform that obtains a form of plane-wave expansion while satisfying the requirements of block overlapping, critical sampling, and perfect reconstruction. Finally, we discussed applications that make use of space-time-frequency processing, such as i) filtering a source in a wave field sampled on a curved contour, ii) parametrizing the acoustic scene through spatial deconvolution, and iii) compressing the wave field through plane-wave encoding.

References

[1] P. Vandewalle, J. Kovacevic, and M. Vetterli, "Reproducible research in signal processing—What, why, and how," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 37–47, 2009.

[2] J. Fourier, "Mémoire sur la propagation de la chaleur dans les corps solides [Memoir on the propagation of heat in solid bodies]," (in French) *Nouveau Bulletin des Sciences par la Société Philomatique de Paris*, vol. 6, pp. 215–221, 1807.

[3] P. Dirichlet, "Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données [On the convergence of trigonometric series which serve to represent an arbitrary function between given limits]," (in French) *J. Für Die Reine und Angew. Math.*, vol. 4, pp. 157–169, 1829.

[4] D. Gabor, "Theory of communication," *J. Inst. Electr. Eng.*, vol. 93, pp. 429–457, 1946.

[5] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. Anal.*, vol. 15, pp. 723–736, 1984.

[6] J. Princen, A. Johnson, and A. Bradley, "Subband/transform coding using filter bank designs based on time domain aliasing cancellation," in *Proc. IEEE Inter. Conf. Acoustic, Speech, Signal Process. (ICASSP)*, 1987, vol. 12, pp. 2161–2164.

[7] E. Williams, *Fourier Acoustics*. New York: Academic Press, 1999.

[8] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[9] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas, Propag.*, vol. 30, no. 1, pp. 27–34, 1982.

[10] A. Berkhout, D. de Vries, J. Baan, and B. van den Oetelaar, "A wave field extrapolation approach to acoustical modeling in enclosed spaces," *J. Acoust. Soc. Amer.*, vol. 105, no. 3, pp. 1725–1733, 1999.

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

[12] M. Boone, "Acoustic rendering with wave field synthesis," presented at the ACM SigGraph Campfire: Acoustic Render, Snowbird, UT, May 26–29, 2001.

[13] D. Dudgeon, *Multidimensional Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[14] P. Bremaud, *Mathematical Principles of Signal Processing*. New York: Springer, 2002.

[15] P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[16] J. Princen and A. Bradley, "Analysis/synthesis filter bank design based on time domain aliasing cancellation," *IEEE Trans. Acoustic, Speech, Signal Process.*, vol. 34, no. 5, pp. 1153–1161, 1986.

[17] H. Malvar, *Signal Processing With Lapped Transforms*. Norwood, MA: Artech House, 1992.

[18] G. Schuller and T. Karp, "Modulated filter banks with arbitrary system delay: Efficient implementations and the time-varying case," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 737–748, 2000.

[19] J. Kovacevic, D. L. Gall, and M. Vetterli, "Image coding with windowed modulated filter banks," in *IEEE Inter. Conf. Acoustic, Speech, Signal Process.*, 1989, pp. 1949–1952.

[20] A. Johnson, J. Princen, and H. Chan, "Frequency scalable video coding using the MDCT," in *IEEE Inter. Conf. Acoustics, Speech, Signal Process.*, 1994, vol. V, pp. V/477–V/480.

[21] R. Bamberger and M. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Process.*, vol. 40, pp. 882–893, 1992.

[22] S. Phoong, C. Kim, P. Vaidyanathan, and R. Ansari, "A new class of two-channel biorthogonal filter banks and wavelet bases," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 649–665, 1995.

[23] M. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, 2005.

[24] V. Chappelier, C. Guillemot, and S. Marinkovic, "Image coding with iterated contourlet and wavelet transforms," in *IEEE Inter. Conf. Image Process.*, 2004, vol. 5, pp. 3157–3160.

[25] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, 1972.

[26] B. Widrow, K. Duvall, R. Gooch, and W. Newman, "Signal cancellation phenomena in adaptive antennas: Causes and cures," *IEEE Trans. Antennas Propag.*, vol. 30, no. 3, pp. 469–478, 1982.

[27] E. Start, V. Valstar, and D. de Vries, "Application of spatial bandwidth reduction in wave field synthesis," presented at the Audio Eng. Soc. 98th Conv., Paris, France, Feb. 25–28, 1995.

[28] A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[29] F. Pinto and M. Vetterli, "Near-field adaptive beamforming and source localization in the spacetime frequency domain," in *IEEE Inter. Conf. Acoustic, Speech, Signal Process.*, 2010, pp. 2734–2737.

[30] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*. New York: Academic, 2000.

[31] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, pp. 1417–1428, 2002.

[32] F. Pinto and M. Vetterli, "Wave field coding in the spacetime frequency domain," in *IEEE Inter. Conf. Acoustic, Speech, Signal Process.*, 2008, pp. 365–368.

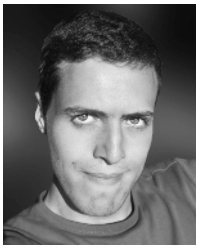
[33] F. Pinto and M. Vetterli, "Bitstream format for spatio-temporal wave field coder," presented at the Audio Eng. Soc. 124th Conv., Amsterdam, The Netherlands, May 17–20, 2008.

[34] M. Bosi and R. Goldberg, *Introduction to Digital Audio Coding and Standards*. New York: Springer, 2002.

[35] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1996.

[36] F. Pinto and M. Vetterli, "Coding of spatio-temporal audio spectra using tree-structured directional filterbanks," in *IEEE Work. Appl. Signal Process. Audio and Acoustic*, 2009, pp. 277–280.

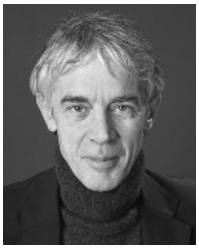
[37] R. Eslami and H. Radha, "A new family of nonredundant transforms using hybrid wavelets and directional filter banks," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1152–1167, 2007.



Francisco Pinto (M'07) received the Dipl. El.-Eng. degree from the University of Porto, Portugal, in 2004.

From 2004 to 2006, he was a Senior Researcher at the Institute for Systems and Computer Engineering, Portugal, working in the area of digital equalization of room acoustics. Since 2006, he is a Research Assistant at the Audiovisual Communications Laboratory at EPF Lausanne (EPFL), Switzerland, where he is currently working towards the Ph.D. degree in the area of Fourier acoustics and signal processing.

Mr. Pinto was the recipient of the Calouste Gulbenkian Fellowship in 2006 and the IEEE Best Student Paper Award at the IEEE International Conference on Acoustics, Speech, and Signal Processing in 2008.



Martin Vetterli (S'86–M'86–SM'90–F'95) received the Dipl. El.-Ing. degree from ETH Zurich (ETHZ), Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, CA, in 1982, and the Doctoratès Sciences degree from EPF Lausanne (EPFL), Switzerland, in 1986.

He was a Research Assistant at Stanford University and EPFL and has worked for Siemens and AT&T Bell Laboratories. In 1986, he joined Columbia University, New York, where he was an Associate Professor of electrical engineering and Co-Director of the Image and Advanced Television Laboratory. In 1993, he joined the University of California at Berkeley, where he was a Professor in the Department of Electrical Engineering and Computer Sciences until 1997, and currently holds an Adjunct Professor position. Since 1995, he has been a Professor of Communication Systems at EPFL, where he chaired the Communications Systems Division from 1996 to 1997, and heads the Audiovisual Communications Laboratory. From 2001 to 2004, he directed the National Competence Center in Research on mobile information and communication systems. He has also been a Vice-President at EPFL since October 2004 in charge of international affairs and computing services, among others. He has held visiting positions at ETHZ (1990) and Stanford (1998). He is the coauthor of three books, one with J. Kovacevic titled *Wavelets and Subband Coding* (1995), one with P. Prandoni titled *Signal Processing for Communications* (2008), and one with J. Kovacevic and V. K. Goyal titled *Fourier and Wavelet Signal Processing* (2010). He has published about 140 journal papers on a variety of topics in signal/image processing and communications and holds a dozen patents.

Dr. Vetterli is a fellow of the ACM, a fellow of EURASIP, and a member of SIAM. He is on the editorial boards of *Applied and Computational Harmonic Analysis*, the *Journal of Fourier Analysis and Application*, and the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. He received the Best Paper Award of EURASIP in 1984, the Research Prize of the Brown Bovery Corporation, Switzerland, in 1986, the IEEE Signal Processing Society's Senior Paper Awards in 1991, in 1996 and in 2006 (for papers with D. LeGall, K. Ramchandran, and Marziliano and Blu, respectively). He won the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, the IEEE Signal Processing Technical Achievement Award in 2001 and is an ISI highly cited researcher in engineering. He was a member of the Swiss Council on Science and Technology from 2000 to 2003. He was a plenary speaker at various conferences (e.g., IEEE ICIP, ICASSP, and ISIT).