

APPLICATIONS OF SHORT SPACE-TIME FOURIER ANALYSIS IN DIGITAL ACOUSTICS

Francisco Pinto and Martin Vetterli

Ecole Polytechnique Fédérale de Lausanne
EPFL-IC-LCAV, Station 14, CH-1015 Lausanne, Switzerland
francisco.pinto@epfl.ch; martin.vetterli@epfl.ch

ABSTRACT

This paper presents a signal processing tool for analyzing and manipulating digitized acoustic wave fields, based on a spatio-temporal extension of the time–frequency representation space. The emphasis is on wave fields acquired with a 1-D linear array of equidistant microphones (representing the spatial samples), but the basic formulation is valid for the three dimensions of space. We start by defining a spatio-temporal extension of the short time Fourier transform, in both continuous and discrete space and time, and then focus on applications of the proposed representation. In particular, we show that acoustic scenes with multiple sources can benefit from the use of space–frequency analysis in applications such as source separation (spatial filtering) and spatial audio coding (wave field coding). The experiments suggest that there is a spatial window size for which the performance of filtering and coding is optimal.

Index Terms—Time–frequency analysis, space–frequency analysis, source separation, spatial audio coding

1. INTRODUCTION

Digital acoustics is a field of digital signal processing (DSP) where signals are spatio-temporal representations of digitized acoustic wave fields. The wave field is typically recorded with an array of microphones and reproduced with an array of loudspeakers, which are the space-domain equivalent of signal samples [1, 2]. By representing the wave field as a discrete function of space and time, we can use multidimensional Fourier analysis and DSP techniques to process the entire wave field as a whole.

The spatio-temporal representation of the wave field is given by the sound pressure $p(\mathbf{r}, t)$ at position $\mathbf{r} = (x, y, z)$ and time t . This function can be derived from the wave equation, given by

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) p(\mathbf{r}, t) = -\delta(\mathbf{r} - \mathbf{r}_p) s(t) \quad (1)$$

where $\delta(\mathbf{r} - \mathbf{r}_p) s(t)$ indicates that the wave field is generated by a point source located at $\mathbf{r}_p = (x_p, y_p, z_p)$, with source signal $s(t)$. A point source generates a spherical wave front that propagates with speed c . Alternatively, $p(\mathbf{r}, t)$ can be expressed in the spatio-temporal Fourier domain, using the definition

$$P(\Phi, \Omega) = \int_{\mathbb{R}^3} \int_{\mathbb{R}} p(\mathbf{r}, t) e^{-j(\Phi \cdot \mathbf{r} + \Omega t)} dt d\mathbf{r} \quad (2)$$

where $\Phi = (\Phi_x, \Phi_y, \Phi_z)$ are the spatial frequencies in rad/m and Ω is the temporal frequency in rad/s.

This research is funded by: Portuguese Science and Technology Foundation (SFRH/BD/27257/2006); Swiss National Science Foundation (200021-121935); ERC Advanced Grant: Support for Frontier Research (247006).

The radiation of sound, like many other phenomena in nature, manifests in the form of harmonic patterns—by inducing harmonic disturbances in the surrounding medium. This implies that $p(\mathbf{r}, t)$ behaves harmonically across both space and time, and is therefore efficiently represented by $P(\Phi, \Omega)$ [3]. One particular transform pair is known as the *plane wave*, characterized by the input $p(\mathbf{r}, t) = e^{j\Omega_0(t + \mathbf{u} \cdot \mathbf{r}/c)}$. This results in a single Dirac point centered at $(\Phi, \Omega) = (\mathbf{u}\Omega_0/c, \Omega_0)$, where Ω_0 is a fixed frequency and $\mathbf{u} = (u_x, u_y, u_z)$ is the direction of propagation. The plane wave is essentially a complex frequency that propagates as a plane wave-front (*i.e.*, a frequency with origin in the far-field); it is the basic harmonic element of acoustic wave fields. Thus, the same way signals are composed of frequencies, wave fields are composed of plane waves¹. The definition of plane wave, however, has an important limitation: it is unable to represent the curvature variations across space that characterize spherical wave-fronts. This is related to a problem encountered by the Hungarian mathematician Dennis Gabor in 1946, when he realized that the Fourier transform was not suited for representing the frequency variations that characterize non-stationary signals, such as speech and music [4]. As he pointed out, our everyday perception of sound is conditioned by the idea of “changing frequencies” (think, for example, of the sound of a police siren), which contradicts the strict definition of sinusoid as an infinitely long function with a fixed frequency. Gabor proposed the short time Fourier transform as a means to circumvent this limitation, by essentially redefining the Fourier basis function $e^{j\Omega t}$ into a windowed version $w(t)e^{j\Omega t}$. The short time Fourier transform has since become the standard representation of audio signals.

The goal of this paper is to show that, in digital acoustics, we can generalize Gabor’s concept of time–frequency analysis into the plane wave domain, by redefining the spatio-temporal Fourier basis $e^{j(\Phi \cdot \mathbf{r} + \Omega t)}$ into a windowed version $w_r(\mathbf{r})w_t(t)e^{j(\Phi \cdot \mathbf{r} + \Omega t)}$, where $w_r(\mathbf{r})$ is a 3-D spatial window function. This allows the wave field to be represented in a space–frequency hyperspace, where frequency resolution can be sacrificed for better spatial resolution, thereby providing a more efficient representation of curved wave fronts. We call such representation the *short space-time Fourier transform* (SSTFT).

In addition to the theoretical concepts, this paper emphasizes applications of space–frequency analysis. In Section 2 we introduce the definitions of continuous and discrete SSTFT, and show some important examples; in Section 3 we show how to design a spatial filter in the Fourier domain, and how the size of the spatial window improves the filtering operation; in Section 4 we show how to compress the acoustic wave field based on plane wave encoding, and how the size of the spatial window improves the coding gain.

¹In rigour, wave fields are composed of both plane waves and evanescent waves, but evanescent waves have vanishingly small energy [1].

2. SHORT SPACE-TIME FOURIER TRANSFORM

2.1. Continuous case

In this section, we start by working with the three dimensions of space and then work solely on the x -axis. Given $\mathbf{r} = (x, y, z)$, the short space-time Fourier transform (SSTFT) of $p(\mathbf{r}, t)$ is given by the 8-D function $P(\mathbf{r}_0, t_0, \Phi, \Omega)$ defined as

$$P(\mathbf{r}_0, t_0, \Phi, \Omega) = \int p(\mathbf{r}, t) w_{\mathbf{r}}(\mathbf{r} - \mathbf{r}_0) w_t(t - t_0) e^{-j(\Phi \cdot \mathbf{r} + \Omega t)} dt d\mathbf{r}, \quad (3)$$

where \int denotes $\int = \int_{\mathbb{R}^3} \int_{\mathbb{R}}$, and, for simplicity, $w_{\mathbf{r}}(\mathbf{r}) = w_x(x)w_y(y)w_z(z)$ (though the spatial window can also be non-separable). The inverse SSTFT is defined as

$$p(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int P(\mathbf{r}_0, t_0, \Phi, \Omega) e^{j(\Phi \cdot \mathbf{r} + \Omega t)} d\Omega d\Phi dt_0 d\mathbf{r}_0. \quad (4)$$

where \int denotes $\int = \int_{\mathbb{R}^3} \int_{\mathbb{R}} \int_{\mathbb{R}^3} \int_{\mathbb{R}}$. The inverse formula can be interpreted as the continuous sum of all the “signal blocks” indexed by (\mathbf{r}_0, t_0) , and it requires that $w_{\mathbf{r}}(\mathbf{r})$ and $w_t(t)$ have unit area [3]. The definition in (3) is characterized by a partitioning of the $(\mathbf{r}_0, t_0, \Phi, \Omega)$ -hyperspace into uniform hypercubic regions spanning the dimensions of space and time and the respective frequencies.

Example 1 (far-field source). A point source located in the far-field generates a plane wave-front with direction of propagation \mathbf{u} . If the source signal is $s(t)$, the sound pressure is given by $p(\mathbf{r}, t) = s(t + \mathbf{u} \cdot \mathbf{r}/c)$. Plugging $p(\mathbf{r}, t)$ into (3), yields [3]

$$P(\mathbf{r}_0, t_0, \Phi, \Omega) = S(\Omega) W_{\mathbf{r}}\left(\Phi - \mathbf{u} \frac{\Omega}{c}\right) * W_t(\Omega) e^{-j(\Phi \cdot \mathbf{r}_0 + \Omega t_0)} \quad (5)$$

where $S(\Omega)$ is the Fourier transform of $s(t)$, and $*$ denotes convolution over Ω . The functions $W_{\mathbf{r}}(\Phi)$ and $W_t(\Omega)$ are the Fourier transforms of $w_{\mathbf{r}}(\mathbf{r})$ and $w_t(t)$, and $W_{\mathbf{r}}(\Phi - \mathbf{u} \frac{\Omega}{c})$ is defined such that $W_{\mathbf{r}}(\Phi - \mathbf{u} \frac{\Omega}{c}) = W_x(\Phi_x - u_x \frac{\Omega}{c}) W_y(\Phi_y - u_y \frac{\Omega}{c}) W_z(\Phi_z - u_z \frac{\Omega}{c})$.

The result can be further simplified if the source signal is wide-band and $W_t(\Omega)$ is a nascent Dirac function (which is typically the case in practice). Then, [3]

$$P(\mathbf{r}_0, t_0, \Phi, \Omega) = S(\Omega) W_{\mathbf{r}}\left(\Phi - \mathbf{u} \frac{\Omega}{c}\right) e^{-j\Phi \cdot \mathbf{r}_0}. \quad (6)$$

This shows that the windowing effects caused by $w_{\mathbf{r}}(\mathbf{r})$ tend to be dominant over the effects caused by $w_t(t)$. Thus, we focus our attention on the space-frequency representation of $p(\mathbf{r}, t)$.

Additionally, the emphasis of this paper is on wave fields observed along the 1-D array axis, which, for simplicity, is represented by the x -axis. For $\mathbf{r}_0 = 0$, it follows that

$$P(\Phi_x, \Omega) = S(\Omega) W_x\left(\Phi_x - \cos \alpha \frac{\Omega}{c}\right) \quad (7)$$

where $u_x = \cos \alpha$ represents the direction of arrival of the plane wave-front to the x -axis. This result is illustrated in Fig. 1(a).

Example 2 (near-field source). A point source located in the near-field at position $\mathbf{r} = \mathbf{r}_p$ and with source signal $s(t)$ generates a spherical wave-front defined by $p(\mathbf{r}, t) = s(t - \|\mathbf{r} - \mathbf{r}_p\|/c)/(4\pi \|\mathbf{r} - \mathbf{r}_p\|)$. On the x -axis, the SSTFT is approximated by [3]

$$P(\Phi_x, \Omega) = S(\Omega) \max \left\{ W_x\left(\Phi_x - \cos \alpha \frac{\Omega}{c}\right), M(\Phi_x, \Omega) \right\}, \quad (8)$$

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

$$M(\Phi_x, \Omega) = \begin{cases} W_x(0) & , (\Phi_x, \Omega) \notin \mathcal{U} \\ 0 & , (\Phi_x, \Omega) \in \mathcal{U}, \end{cases} \quad (9)$$

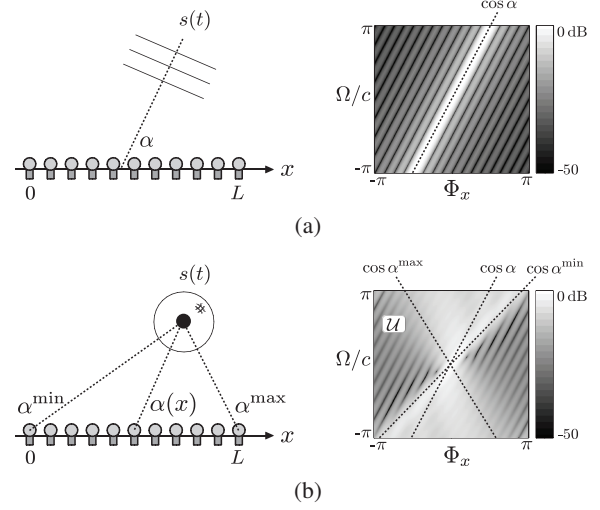


Fig. 1. Spatio-temporal Fourier transform of the sound pressure generated by a point source with $s(t) = \delta(t)$ on the windowed array axis. (a) A far-field source with angle of incidence α generates a sinc-type function oriented towards the slope $d\Phi_x/d\Omega = \cos \alpha/c$. (b) A near-field source with angle of incidence $\alpha(x)$ generates a sinc-type function with its main lobe spread into a triangular region. The minimum and maximum angles determine the aperture of the “main lobe”, and the average of $\cos \alpha(x)$ from $x = 0$ to L determines the orientation of the notches.

with $\mathcal{U} = \mathbb{R}^2 \setminus \{(\Phi_x, \Omega) : \Phi_x^{\min} \leq \Phi_x \leq \Phi_x^{\max}, \Omega \geq 0\}$, and point-symmetric for $\Omega < 0$. The parameters $\cos \alpha$, Φ_x^{\min} , and Φ_x^{\max} are given by $\cos \alpha = \frac{1}{L} \int_0^L \cos \alpha(x) dx$, $\Phi_x^{\min} = \cos \alpha^{\max} \frac{\Omega}{c}$, and $\Phi_x^{\max} = \cos \alpha^{\min} \frac{\Omega}{c}$. The result is illustrated in Fig. 1(b).

2.2. Discrete case

In discrete space and time (again, considering solely the x -axis), the SSTFT can be implemented as a lapped 2-D block transform. For this purpose, define $p[\mathbf{n}] = p[n_x, n_t]$ as the uniformly sampled version of $p(x, t)$, where n_x and n_t are the sample indexes in space and time. The decomposition of $p[\mathbf{n}]$ into 50%-overlapped blocks $p_i[\mathbf{n}]$ can be written as

$$p_i[\mathbf{n}] = p[\mathbf{n}], \quad \mathbf{n} = \frac{\mathbf{N}}{2} \mathbf{i}, \dots, \frac{\mathbf{N}}{2} (\mathbf{i} + 2) - \mathbf{1}, \quad \mathbf{i} \in \mathbb{I}^2, \quad (10)$$

where $\mathbf{i} = \begin{bmatrix} i_x \\ i_t \end{bmatrix}$ is the block index and $\mathbb{I}^2 \subset \mathbb{Z}^2$ is the respective set of block indexes. The matrix $\mathbf{N} = \begin{bmatrix} N_x & 0 \\ 0 & N_t \end{bmatrix}$ contains the number of samples N_x and N_t in space and time. The notation $\mathbf{n} = \frac{\mathbf{N}}{2} \mathbf{i}, \dots, \frac{\mathbf{N}}{2} (\mathbf{i} + 2) - \mathbf{1}$ means that $n_x = \frac{N_x}{2} i_x, \dots, \frac{N_x}{2} (i_x + 2) - 1$ and $n_t = \frac{N_t}{2} i_t, \dots, \frac{N_t}{2} (i_t + 2) - 1$. The vector integers are defined as $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and so on. Note 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.

Defining $P[\mathbf{b}] = P[b_x, b_t]$ as the discrete SSTFT of $p[\mathbf{n}]$, the direct and inverse transforms of each block are given by

$$P_i[\mathbf{b}] = \sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{N1}-1} p_i[\mathbf{n}] \varphi[\mathbf{b}, \mathbf{n}], \quad \mathbf{b} = \mathbf{0}, \dots, \mathbf{N1} - \mathbf{1} \quad (11)$$

$$p_i[\mathbf{n}] = \sum_{\mathbf{b}=\mathbf{0}}^{\mathbf{N1}-1} P_i[\mathbf{b}] \varphi[\mathbf{b}, \mathbf{n}], \quad \mathbf{n} = \mathbf{0}, \dots, \mathbf{N1} - \mathbf{1}. \quad (12)$$

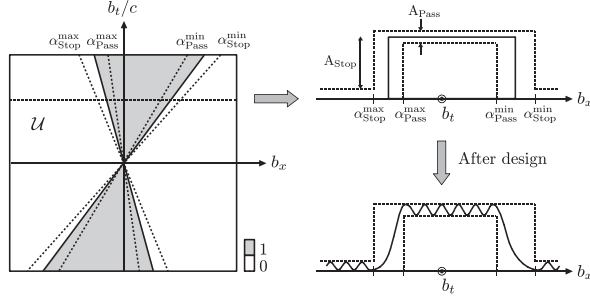


Fig. 2. Design of spatial filters in the 2-D Fourier domain.

where $\varphi[\mathbf{b}, \mathbf{n}] = w[\mathbf{n}]e^{j2\pi\mathbf{b}\cdot\mathbf{N}^{-1}\mathbf{n}}/\sqrt{\det\mathbf{N}}$ is the windowed 2-D Fourier basis, with $w[\mathbf{n}] = w_x[n_x]w_t[n_t]$ being the product of window functions in space and time.

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

$$p[\mathbf{n}] = \sum_{\mathbf{i} \in \mathbb{Z}^2} p_{\mathbf{i}}\left[\mathbf{n} - \frac{1}{2}\mathbf{N}\mathbf{i}\right], \quad \mathbf{n} \in \mathbb{Z}^2. \quad (13)$$

The SSTFT requires that $w_x[n_x]$ and $w_t[n_t]$ satisfy the conditions $w[n] = w[N-1-n]$ and $w^2[n] + w^2[n+N/2] = 1$ [3].

3. APPLICATION I: SPATIAL FILTERING

3.1. Filtering frequencies vs plane waves

One of the advantages of the Fourier transform in the design of digital filters is that it allows the interpretation of convolutional filtering in terms of 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). In digital acoustics, the concept of filtering is more general, since the goal is to filter plane waves rather than frequencies. Since plane waves are the basic elements of acoustic wave fields, spatio-temporal filters can, in principle, be designed to enhance or suppress entire sources in space, rather than only specific frequencies. Using a similar reasoning, a spatio-temporal filter can be sketched 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. If the goal is to filter an entire source, the ideal filter is one that preserves the spectral region where most of the energy is contained. Using the result in (8) and (9), the filter can be defined as

$$H[\mathbf{b}] = \begin{cases} 1 & , \mathbf{b} \notin \mathcal{U} \\ 0 & , \mathbf{b} \in \mathcal{U}. \end{cases} \quad (14)$$

Spatio-temporal filters of this type are non-separable, and therefore require 2-D filter design techniques (e.g., Dudgeon *et al.* [5]). The design can be performed in two steps: (i) specification of the ideal filter, and (ii) approximation by a realizable filter. The filter can be specified with parameters such as the cut-off angles of pass-band and stop-band regions, the width of transition bands, and the order of the filter. The filter design algorithm then translates the filter specifications into a realizable filter, usually through error minimization. This procedure is illustrated in Fig. 2.

3.2. Spatial filtering in the SSTFT domain

In a scene with two (or more) sources, the design of spatio-temporal filters in the 2-D Fourier domain is conditioned by a fundamental

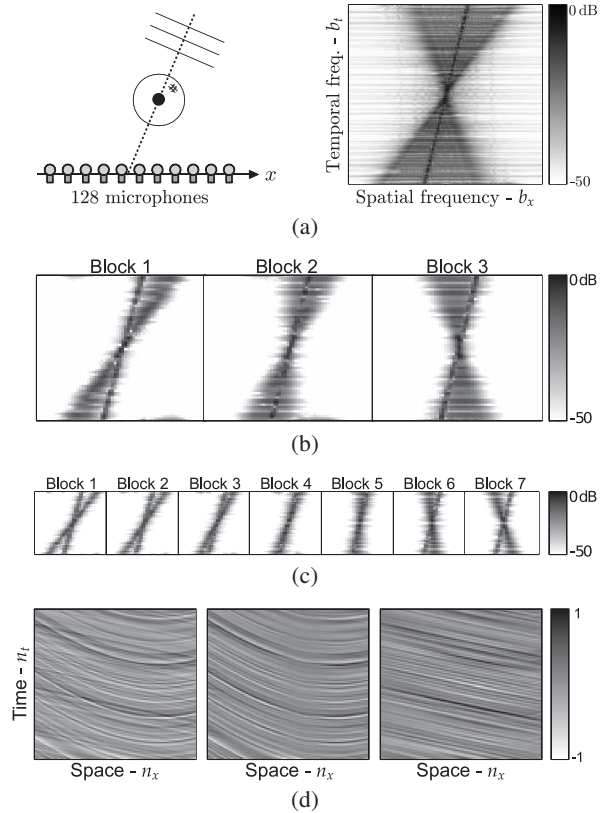


Fig. 3. Filtering co-linear sources in the SSTFT domain. (a) The scene is composed of two co-linear white noise sources (one behind the other)—one in the far-field, the other in the near-field—and a linear array with 128 microphones. Co-linearity means that the parameter $\cos \alpha$ of the two sources is equal. The resulting spectrum is a sum of the spectral patterns in (7) and (8). (b) Using a spatial window of size $L = 32$, the aperture of the triangular region decreases, resulting in less spectral overlapping. (c) With $L = 16$, the spectral overlapping is further decreased, but the spatial frequency resolution is considerably degraded. (d) The window size $L = 32$ produces the best results: on the left is the input $p[\mathbf{n}]$; in the middle is the filtered near-field source; on the right is the filtered far-field source.

trade-off: if there is overlapping between the two spectral patterns, and the goal is to suppress the interfering source, the overlapping region can not be filtered without affecting the target source as well (in particular, without affecting the *curvature* of the wave-front). One solution to this problem is to split the array axis into smaller segments. This effectively reduces the aperture of the overlapping spectral region, at the expense of a lower spatial frequency resolution. Since the quality of the spatial filtering operation depends on these two factors, we can expect that there is an optimal spatial window size that properly balances the amount of spectral overlapping with the spatial frequency resolution. An example is shown in Fig. 3.

4. APPLICATION II: WAVE FIELD CODING

4.1. Coding frequencies vs plane waves

Since the invention of MP3, the standard way of compressing audio data has become mostly based on encoding frequency-domain

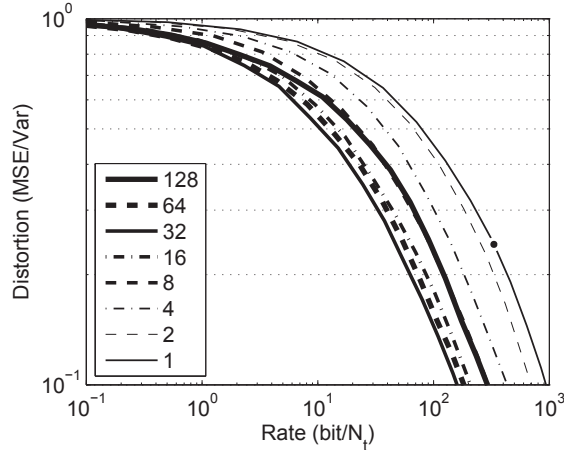


Fig. 4. Coding of co-linear sources in the SSTFT domain. The scene is the same as in Fig. 3, and $P[b]$ is encoded using (15) and spatial window sizes from $L = 1$ to 128. The results show that the rate-distortion curve improves at first as we decrease the window size, but then starts to degrade all the way down to $L = 1$ (which corresponds to coding each channel independently). As in the filtering operation, the best coding gain is obtained for $L = 32$. Notice the black dot on the $L = 1$ curve, which represents the operating point of MP3. For the same distortion, the bit-rate achieved simply by changing the window size to $L = 32$ is 7 times lower.

coefficients, rather than the time-domain waveform. This strategy was heavily influenced by the fact that the human auditory system analyzes sound in terms of frequencies, and that certain frequencies are more relevant than others on a perceptual level. If the audio data contains multiple channels, the typical assumption is that the channels are highly correlated and therefore can be jointly encoded. The correlation criteria can be either mathematically motivated or perceptually motivated (*e.g.*, based on the ability of humans to localize sound sources in space [6]).

In digital acoustics, the coding strategy is considerably different: instead of encoding the multichannel audio data parametrically as correlated signals, the data is encoded non-parametrically in the spatio-temporal Fourier domain. In other words, instead of jointly encoding frequencies, we are encoding plane waves, since they are the basic harmonic elements of the acoustic wave field. Plane waves can also be encoded using perceptual criteria, based on a combination of frequency masking models and spatial masking models [7]. The purpose here, however, is not to discuss the advantages of different coding criteria, but the influence of space-frequency analysis in the overall coding efficiency—in particular, the effects of the spatial window size. Therefore, in our experiments, we use a coding strategy based on traditional rate-distortion analysis.

We define the quantization and de-quantization operations as

$$P_Q[b] = \varepsilon \left\lfloor \text{SF}[b] |P[b]| \right\rfloor \quad \text{and} \quad \hat{P}[b] = \varepsilon_Q \frac{|P_Q[b]|}{\text{SF}[b]} \quad (15)$$

where $\lfloor \cdot \rfloor$ denotes rounding to the closest lower integer, $\text{SF}[b]$ are the scale factors of each coefficient, and $\varepsilon = \text{sign}\{P[b]\}$. The purpose of the scale factors is to scale the coefficients of $P[b]$ such that the rounding operation yields the desired quantization noise in the reconstructed coefficients $\hat{P}[b]$.

To determine the number of bits $R[b]$ required to encode the quantized coefficients, we associate the amplitude values to a Huffman code book similar to the one used by the MPEG standard [8],

where code words are organized such that less bits are used to describe lower amplitude values. Using the MSE distortion metric, we estimate the rate-distortion function $D(R)$ with the parametric pair

$$R = \sum_{b=0}^{N1-1} R[b] \quad \text{and} \quad D = \frac{1}{\det \mathbf{N}} \sum_{n=0}^{N1-1} (p[n] - \hat{p}[n])^2 \quad (16)$$

where R and D are functions of $\text{SF}[b]$, with $0 \leq \text{SF}[b] < \infty$.

4.2. Wave field coding in the SSTFT domain

Similarly to what happens in the spatial filtering operation, the process of encoding plane waves is conditioned by the size of the spatial window. The reason is that the wave-front generated by each source has a different curvature along the array axis. According to (9), window segments that are closer to the source exhibit a wider (less compact) spectral pattern compared to segments that are farther away from the source. This means that far-field spectral patterns contain less information than near-field spectral patterns, and therefore can be encoded more efficiently. Using space-frequency analysis along the array axis, the curvature variations of the wave-front can be better exploited in order to obtain more compact spectral patterns. On the downside, reducing the spatial window size also degrades the spatial frequency resolution. So, again, we can expect that there is an optimal spatial window size that results in the highest coding gain. An example is illustrated in Fig. 4.

5. CONCLUSIONS

In this paper, we show through a worst-case scenario (co-linear sources) that the use of space-frequency analysis in digital acoustics improves the performance of spatial filtering and wave field coding. In both experiments, we found that a spatial window with 25% the size of the microphone array produces the best results: (i) in spatial filtering, the two sources are better separated; (ii) in wave field coding, a coding gain of 7 is achieved for an MSE similar to MP3.

6. REFERENCES

- [1] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Sig. Proc.*, vol. 54, pp. 3790–3804, 2006.
- [2] A. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoustic Soc. America*, vol. 93, pp. 2764–2778, 1993.
- [3] F. Pinto and M. Vetterli, "Space-time-frequency processing of acoustic wave fields: Theory, algorithms, and applications," *IEEE Trans. Sig. Proc.*, vol. 58, pp. 4608 – 4620, 2010.
- [4] D. Gabor, "Theory of communication," *Journal I.E.E.*, vol. 93, pp. 429–457, 1946.
- [5] D. Dudgeon, *Multidimensional Digital Signal Processing*, Prentice Hall, 1984.
- [6] J. Herre, C. Faller, C. Ertel, J. Hilpert, A. Hoelzer, and C. Spenger, "Mp3 surround: Efficient and compatible coding of multi-channel audio," in *Audio Eng. Soc. 116th Conv.*, 2004.
- [7] F. Pinto and M. Vetterli, "Wave field coding in the spacetime frequency domain," in *IEEE Inter. Conf. Acoustic, Speech, Signal Process.*, 2008.
- [8] ISO/IEC, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s - part 3: Audio," 1993, JTC1/SC29/WG11.