Acoustic Imaging With Circular Microphone Array: A New Approach for Sound Field Analysis

Marco Olivieri ⓘ, *Student Member, IEEE*, Amy Bastine ⓘ, *Student Member, IEEE*, Mirco Pezzoli ⓘ,
Fabio Antonacci ⓘ, *Member, IEEE*, Thushara Abhayapala ⓘ, *Senior Member, IEEE*,
and Augusto Sarti ⓘ, *Senior Member, IEEE*

Abstract—Acoustic imaging is powerful in collecting spatial information of acoustic sources into a visual representation. In this paper, we focus on the analysis of the exterior acoustic field captured by a circular array of microphones. With a proper parametrization based on angles, we map the directions of arrival of sources as a function of the microphone locations, thus obtaining an acoustic image called "angular space". Therefore, we introduce a linear transform to enable analysis and synthesis operations for mapping the microphone pressures onto the angular space using local space-time Fourier analysis. We prove the ability of this representation to combine global information coming from multiple arrays in a single acoustic image that can be processed and manipulated. Examples of source localization applications in simulated and measured scenarios show the effectiveness of the proposed method obtaining results comparable with state-of-the-art methods.

Index Terms—Acoustic imaging, circular arrays, soundfield analysis, source localization.

I. Introduction

THE problem of analyzing and reconstructing the sound field through acquisitions by multiple microphones has been widely studied in the literature. The ever-evolving trends in immersive multimedia communication have triggered a rapid evolution of computational acoustics, from the early solutions of spatial audio processing and rendering to the emerging paradigms of interactive extended reality [1]. Nowadays, indeed, sound field analysis and processing play a crucial role for the signal processing community which aims to develop new models and flexible representations to efficiently manipulate acoustic scenes. [2], [3].

An interesting approach to the analysis of the acoustic features is represented by acoustic imaging, which refers to a class of methods aiming at "visualizing" the sound. Such techniques

are able to encode the acoustic properties into suitable visual representations, which can be easily manipulated to understand the interactions between different entities of the environment [4], [5]. In the context of real-world sound fields, acoustic imaging is powerful in collecting spatial information of acoustic sources, e.g., for source localization, radiation pattern estimation, and signal extraction applications [6], [7], [8], [9]. These tasks require the adoption of microphone sensors arranged in clusters or arrays to analyze the directional components of the acoustic wave field [10], [11].

The geometry of the microphone array plays an important role in the formulation of the processing algorithms and it tends to rely on strong modeling assumptions that might limit the accuracy and flexibility of such solutions in specific applications. Moreover, depending on the relative positions between the acoustic sources and the microphone sensors, we can broadly classify the analysis into two different acoustic conditions, which imply the use of different assumptions: near field or far field [12]. In far field conditions, the emitter can be considered as a point source and the wavefront that impinges the receiver can be considered as planar. Therefore, in this case, the plane-wave decomposition is typically adopted to describe the sound field [10], [13], [14]. On the contrary, when the source is close enough to the receiver, the phase differences and the presence of evanescent waves have to be included in the model to properly capture all the sound field components [12], [15], [16]. However, in [17], authors represented sound fields produced by nearby sources with far-field components, thanks to the definition of the directional plenacoustic function [18], [19] that maps acoustic rays captured by a linear array onto the "ray space" domain using a non-linear parametrization.

Authors in [20] introduced the Ray Space Transform (RST) by exploiting the idea of estimating the sound field image with multiple beamforming operations performed on sub-arrays of Uniform Linear Arrays (ULAs). In particular, the RST and its inverse counterpart are efficient in representing spatial information of acoustic primitives such as sources and reflectors. However, with the RST framework, expanding the description of acoustic information captured with multiple microphone arrays requires an increase in the number of parameters, even for the analysis of 2D scenarios [7], [21], [22]. Moreover, the combination of local ray space images computed by distributed ULAs entails complex projection and interpolation operations [8], [23], thus increasing

the computational complexity and reducing the interpretability of the acoustic image [22].

In many applications, the use of different microphone array geometries is beneficial for the analysis of acoustic fields. In scenarios where the sources are placed in a predefined closed region, such as in conference meetings or teams working inside an office, circular arrays are particularly useful in capturing spatial cues of the wavefronts [24]. In this context, Spherical Harmonics (SH) provides an alternative solution to describe the sound field with a compact set of basis functions [25], [26], [27]. However, available techniques [28], [29], [30], [31] usually rely on a large number of high-order microphones (HOMs) and high computational power in order to fulfill the necessary spatial sampling requirements [32].

In this manuscript, we propose an acoustic imaging approach for the 2D sound field analysis using microphone arrays arranged in a circular geometry. In particular, we aim to represent the spatial information of acoustic entities of the exterior sound field surrounded by a circular array in a single complex-value 2D map. Therefore, we defined a parametrization that maps multiple beamforming operations computed in the Euclidean space for multiple observation windows of the circular array. The devised parametrization consists of two angles: the angle of each sub-portion of the circular array with respect to a predefined reference angle, and the local steering directions of the beamformer with respect to the specific observation window. For this reason, we called the devised representation as "angular space". For mapping the sound pressure acquired by the circular array onto the angular space, we defined analysis and synthesis operators as Angular Space Transform (AST) and its inverse counterpart (iAST). This transform follows a similar methodology as the RST, which utilized discrete Gabor frames [33], [34] in the space-time domain to effectively decompose the sound field onto a basis of wave field functions.

We validate the AST-based sound field representation by applying it to source localization and comparing its performance to the state-of-the-art SRP-PHAT algorithm [35], [36]. Our method achieves comparable localization accuracy to SRP-PHAT while also providing a full description of the sound field that can facilitate further audio processing tasks. Moreover, we present the application of the AST on source localization even in scenarios with variable reverberation time and sensor noise levels. We show the flexibility of the proposed approach to combine the spatial information of multiple arrays in the same angular space without increasing the number of parameters and complexity like previous imaging techniques [8], [22]. Finally, we performed acoustic measurements in a semianechoic room using four ULA placed along the circumference of a virtual circular array to prove the ability of the devised model to analyze the sound field in practical scenarios.

It is worth noticing that the main objective of this work is to represent the spatial information of acoustic entities of the exterior sound field surrounded by a circular array in a compact and unique complex-valued 2D map. Moreover, the devised model overcomes the limitations of previous acoustic imaging techniques by combining the local features captured from multiple and different positions in a single acoustic image

that can be directly exploited. Nevertheless, this manuscript serves as an initial proof of concept for devised AST, and we plan to implement further applications in the context of multi-zone speech separation and dereverberation, where we envision the potentiality of the angular space.

The rest of the paper is organized as follows. Section II provides the signal model and the background of the plenacoustic function. The definition of the setup along with the analytical parametrization and the AST formulation is presented in Section III. In Section IV simulated experiment for source localization applications of AST are presented. Experimental measurement scenarios using distributed arrays are reported in Section V. Finally, Section VI draws some conclusions and future works.

II. SIGNAL MODEL AND BACKGROUND

A. Notation

- Scalars are denoted by lowercase letters, column vectors by bold lowercase letters, and matrices by bold uppercase letters.
- Superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote transposition, conjugation, and conjugate transposition, respectively.
- $\langle \cdot, \cdot \rangle$ denotes inner product.
- j is the imaginary unit ($j = \sqrt{-1}$).
- $[\mathbf{A}]_{i,w}$ is the element in row i and column w of the matrix $\mathbf{A}$.
- $E\{\cdot\}$ is the expectation operator.
- Azimuth angles $\in [0, 2\pi)$ are measured counter-clockwise from the x-axis.
- $|\cdot|$ and $||\cdot||$ denote absolute value and Euclidean norm, respectively.

B. Signal Model

Consider a microphone array of M microphones located at the spatial coordinate $\boldsymbol{r}_m$, $m = 1, \ldots, M$, and an acoustic source located at $\boldsymbol{r}_s$. Let $h(\boldsymbol{r}_m | \boldsymbol{r}_s; \omega)$ be the acoustic transfer function between the source and the mth microphone, with ω being the angular frequency. Then the sound pressure received at the microphone can be expressed in frequency-domain as

$$p(\boldsymbol{r}_m, \omega) = h(\boldsymbol{r}_m | \boldsymbol{r}_s; \omega) s(\omega) + n(\boldsymbol{r}_m, \omega), \qquad (1)$$

where $s(\omega)$ is the source signal and $n(\boldsymbol{r}_m, \omega)$ is the additive sensor noise of zero mean and variance σ_n^2. Following (1), the output of the microphone array can be modeled as

$$\mathbf{p}(\omega) = \mathbf{h}(\omega) s(\omega) + \mathbf{n}(\omega), \qquad (2)$$

where

$$\mathbf{p}(\omega) = [p(\boldsymbol{r}_1, \omega), \ldots, p(\boldsymbol{r}_M, \omega)]^T \in \mathbb{C}^{M \times 1}, \qquad (3)$$

$$\mathbf{h}(\omega) = [h(\boldsymbol{r}_1 | \boldsymbol{r}_s; \omega), \ldots, h(\boldsymbol{r}_M | \boldsymbol{r}_s; \omega)]^T \in \mathbb{C}^{M \times 1}, \qquad (4)$$

and

$$\mathbf{n}(\omega) = [n(\boldsymbol{r}_1, \omega), \ldots, n(\boldsymbol{r}_M, \omega)]^T \in \mathbb{C}^{M \times 1} \qquad (5)$$

with a covariance matrix $\sigma_n^2 \mathbf{I}_M$, where $\mathbf{I}_M$ is the identity matrix of size $M \times M$.

The AST formulation outlined in Section III primarily deals with the deterministic term $\mathbf{h}(\omega)s(\omega)$ of the signal model in (2). The stochastic noise term $\mathbf{n}(\omega)$ will be accounted and analyzed in the context of specific applications and experiments in Sections IV and V.

C. Directional Plenacoustic Function

According to the plane-wave decomposition principle [12], the sound field impinging on the mth sensor in (1) can be well approximated as a superposition of plane-waves $e^{j\langle \mathbf{k}_\theta, \mathbf{r}\rangle}$ propagating in all possible directions θ, where $\mathbf{k}_\theta = [k_x, k_y]^T = \omega/c[\cos(\theta), \sin(\theta)]^T$ is the wavenumber vector, with c being the speed of sound. Let $P(\theta, \omega)$ be the Herglotz Density function [37, p. 74] that characterizes the spatial spectrum with amplitude and phase of each plane-wave component. Therefore, we can represent the incident pressure in the point $\mathbf{r}$ as [12]

$$p(\mathbf{r}, \omega) = \frac{1}{2\pi} \int_0^{2\pi} P(\theta, \omega) e^{j\langle \mathbf{k}_\theta, \mathbf{r}\rangle} d\theta. \quad (6)$$

The integrand of (6) given by

$$\varphi(\mathbf{r}, \theta, \omega) = e^{j\langle \mathbf{k}_\theta, \mathbf{r}\rangle} P(\theta, \omega) \quad (7)$$

is called the Directional Plenacoustic Function [22]. It defines the sound field in terms of acoustic radiance in every direction at every point in space. The term $P(\theta, \omega)$, and thus $\varphi(\mathbf{r}, \theta, \omega)$, can be estimated through beamforming operations [13], [17], [20], [22], [38] performed on microphone array signals.

In [20], the $\varphi(\mathbf{r}, \theta, \omega)$ function was mapped onto the "ray space" domain, which is a parameter space that describes the sound field using linear patterns. The process involved transforming the microphone array signals through short-space Fourier transform with discrete Gabor frames, called Ray Space Transform (RST), resulting in coefficients that are parameterized in the ray space for sound field imaging. This approach overcame resolution issues and extended the validity of far-field assumptions [20]. The transformation facilitated the use of pattern-analysis techniques for different spatial sound analysis and manipulation tasks, such as source localization, source separation, and sound field reconstruction. However, the RST formulation is based on a linear observation window or aperture obtained through microphone arrays of linear geometry. Moreover, the adaptation of RST to represent global spatial information acquired by multiple and distributed linear arrays requires a more complex framework based on projective ray space and an additional number of parameters [22].

In Section III, we introduce a different approach for plenacoustic imaging adapted to circular microphone arrays. The microphone array processing follows a similar methodology as [20], but the plenacoustic function $\varphi(\mathbf{r}, \theta, \omega)$ will be parameterized based on an analytical formula that maps the acoustic scene geometry to a new angular parameter space, referred to as *angular space*, defined by the circular observation window. Note that this paper focuses on 2D geometries, where the microphone arrays and sound sources in the scene are situated on the same plane, allowing for a simpler and more practical implementation while still providing results for various applications. It
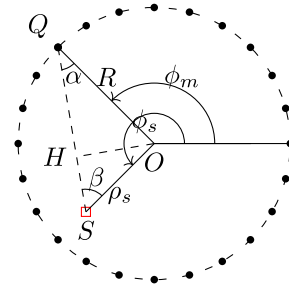

Fig. 1. Circular aperture of radius R centered at origin O depicting an observation point (black dot) at the polar coordinate (R, ϕ_m) and a sound source (red square) located at (ρ_s, ϕ_s).

serves as a foundation for future exploration in 3D sound field imaging.

III. DOA MAPPING WITH CIRCULAR ARRAY

This section presents the parametrization and the operations to map the DOAs of a circular microphone array to represent the exterior sound field in terms of plenacoustic functions. The computed acoustic image, called *angular space*, is established based on an analytical relation between the acoustic scene geometry and the steering angles of the circular observation window. Similar to RST formulation [20], plenacoustic image is generated by dividing the microphone array into sub-arrays and then estimating the amplitudes of the directional components for each sub-array through beamforming, thus having the proposed *Angular Space Transform* (AST). Unlike the linear patterns in RST, representing the beamformer outputs in *angular space* results in mapping the acoustic primitives as sinusoidal patterns.

A. Circular Array Setup

Consider a pressure-sensitive circular aperture of radius R centered at origin O with M microphones as depicted in Fig. 1.

Each sensor is located at the polar coordinate (R, ϕ_m), where $\phi_m \in [0, 2\pi)$ with respect to a reference axis for $m = 1, \dots, M$ with a uniform microphone spacing $d = R\bar{\phi}_m$, where $\bar{\phi}_m = 2\pi/M$. Moreover, the circular array captures the sound pressure generated by a point source located at (ρ_s, ϕ_s).

In this work, we consider only a source located inside the aperture, i.e., $\rho_s \leq R$, despite the generalization to the region outside the circle is possible.

B. Analytical Mapping

In order to map the Euclidean space to the angular space, we establish an analytical relation that expresses the DOAs of a point source in (ρ_s, ϕ_s) as a function of the observation point Q in the circular aperture placed in (R, ϕ_m).

Consider the geometric constructions illustrated in Fig. 1, particularly the triangle QOS formed between the observation point Q, the reference O, and the source S, respectively. Moreover, let $\overline{OH}$ be perpendicular with respect to $\overline{QS}$. As per the trigonometric properties, the altitude line segment $\overline{OH}$ holds the

(a) $\boldsymbol{r}_{s_1} = (0.707\,\text{m}, 225°)$

(b) $\Gamma(\boldsymbol{r}_{s_1})$

(c) $\boldsymbol{r}_{s_2} = (0.3\,\text{m}, 30°)$

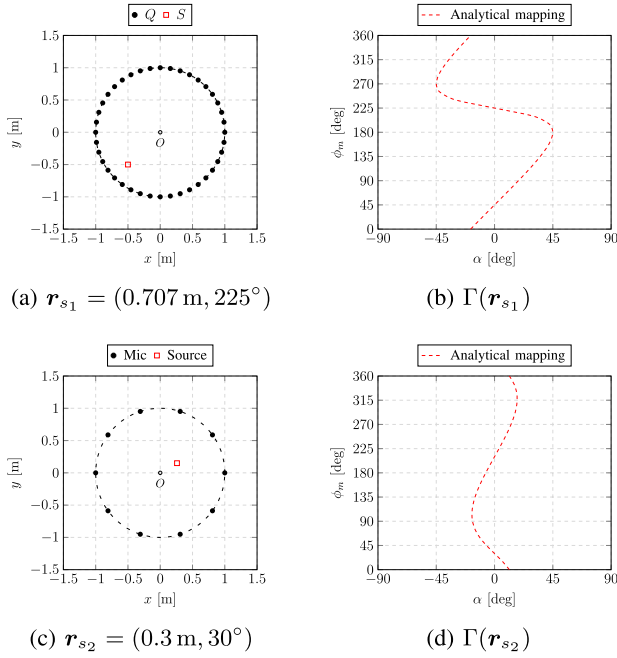(d) $\Gamma(\boldsymbol{r}_{s_2})$

Fig. 2. Mapping relation examples between the Euclidean space (left) and the angular space (right). The two point S are expressed in polar coordinates and observed in Q.

following relations

$$\overline{OH} = R\sin(\alpha), \qquad (8a)$$

$$\overline{OH} = \rho_s \sin(\beta), \qquad (8b)$$

where $\beta = \pi - \alpha - \gamma$ with $\gamma = \phi_s - \phi_m$.

Equating the right-hand sides of (8b) and applying the sine sum formula, we get

$$R\sin(\alpha) = \rho_s \sin(\pi - \alpha - \gamma) = \rho_s \sin(\alpha + \gamma) \qquad (9a)$$

$$= \rho_s \sin(\alpha)\left(\cos(\gamma) + \frac{\cos(\alpha)}{\sin(\alpha)}\sin(\gamma)\right), \qquad (9b)$$

leading to

$$R = \rho_s \cos(\gamma) + \rho_s \cot(\alpha)\sin(\gamma). \qquad (10)$$

From (10) and fixing the radius R of the circular array, we can deduce the expression for the angle $\bar{\alpha}$ under which the source in (ρ_s, ϕ_s) is seen from the microphone in ϕ_m as

$$\bar{\alpha}(\phi_m, \rho_s, \phi_s) = \cot^{-1}\left(\frac{R - \rho_s \cos(\gamma)}{\rho_s \sin(\gamma)}\right). \qquad (11)$$

Moreover, let $\Gamma(\rho_s, \phi_s)$ be the function that maps source S in (11), defined as

$$\Gamma : [0, R] \times [0, 2\pi) \rightarrow \left(-\frac{\pi}{2}, \frac{\pi}{2}\right);$$

$$(\rho_s, \phi_s) \mapsto \Gamma(\rho_s, \phi_s) = \bar{\alpha}(\phi_m | \rho_s, \phi_s). \qquad (12)$$

In Fig. 2, we report two examples of the analytical mapping between a point source in the Euclidean space (Fig. 2(a) and (c)) and the angular space (Fig. 2(b) and (d) respectively) using (11).

The point S in the Euclidean space is mapped in the angular space as $\Gamma(\rho_s, \phi_s)$, resulting in a curve with sinusoidal-like shape. It is worth noticing that the position of S is related to $\Gamma(\rho_s, \phi_s)$. In particular, the source distance ρ_s from the origin O is mapped on the width of Γ as a function of α, whereas the angle position ϕ_s implies a phase shift of Γ with respect to ϕ_m. Indeed, by inspecting the examples in Fig. 2, we can observe that the closer the source is to the center O, the more the corresponding curve $\Gamma(\rho_s, \phi_s)$ tends to $\Gamma = 0$. This represents an extreme case where the source is seen under the same angle from each microphone.

C. Angular Space Transform (AST)

To define the AST, let us consider a uniform grid to sample the angular space formed by (α, ϕ_q), where $\alpha \in (-\pi/2, \pi/2)$ represents the steering directions of beamforming operations of local sub-arrays and $\phi_q \in [0, 2\pi)$ represents the observation angle for the AST. The two axes are sampled in α_w and ϕ_q^i points, with $w = 0, \ldots, W - 1$ and $i = 0, \ldots, I - 1$, where I and W represent the number of samples on the ϕ_q and α axes, respectively.

Analogous to the RST approach in [20], the AST can be formulated as a form of stacked beamforming performed in the angular space. This process entails applying a spatial window centered in I sub-arrays to perform local Fourier transforms using the Gabor expansion [34] on the aperture data $p(\phi_m, \omega)$, followed by gathering the W beamforming outputs steered to each α_w.

To perform the local Fourier analysis, we use a Gaussian spatial window of the form

$$\psi(q) = e^{-\pi \frac{q^2}{\sigma^2}}, \quad \sigma \in \mathbb{R}, \qquad (13)$$

where σ controls the width of the Gaussian window. Therefore, the AST can then be formulated in the discrete domain as

$$[\mathbf{Z}]_{i,w}(\omega) = d \sum_{m=1}^{M} p(\phi_m, \omega) e^{-j\langle \mathbf{k}_{\alpha_w}, \boldsymbol{r}_m \rangle} e^{-\pi \frac{(\phi_m - \phi_q^i)^2}{\sigma^2}}, \qquad (14)$$

where $\mathbf{k}_{\alpha_w} = \omega/c[\cos(\alpha_w), \sin(\alpha_w)]^T$ and $\boldsymbol{r}_m = R[\cos(\phi_m), \sin(\phi_m)]^T$ are the 2D Cartesian coordinates of the wavenumber vector and observation point at the mth microphone, respectively, and d the microphone spacing.

By introducing the matrix $\boldsymbol{\Psi}(\omega) \in \mathbb{C}^{M \times IW}$ given by

$$[\boldsymbol{\Psi}]_{m, i+wI+1} = e^{j\langle \mathbf{k}_{\alpha_w}, \boldsymbol{r}_m \rangle} e^{-\pi \frac{(\phi_m - \phi_q^i)^2}{\sigma^2}} d, \qquad (15)$$

we can conveniently write the discrete AST in (14) in matrix form as

$$\mathbf{z} = \boldsymbol{\Psi}^H \mathbf{p}, \qquad (16)$$

where the vector $\mathbf{z} \in \mathbb{C}^{IW \times 1}$ is formed from the elements of $\mathbf{Z}$ according to

$$[\mathbf{z}]_{i+wI+1} = [\mathbf{Z}]_{i,w}. \qquad (17)$$

We can reconstruct the sound field through *inverse Angular Space Transform* (iAST), which involves performing the local inverse Fourier transform of the angular space coefficients.
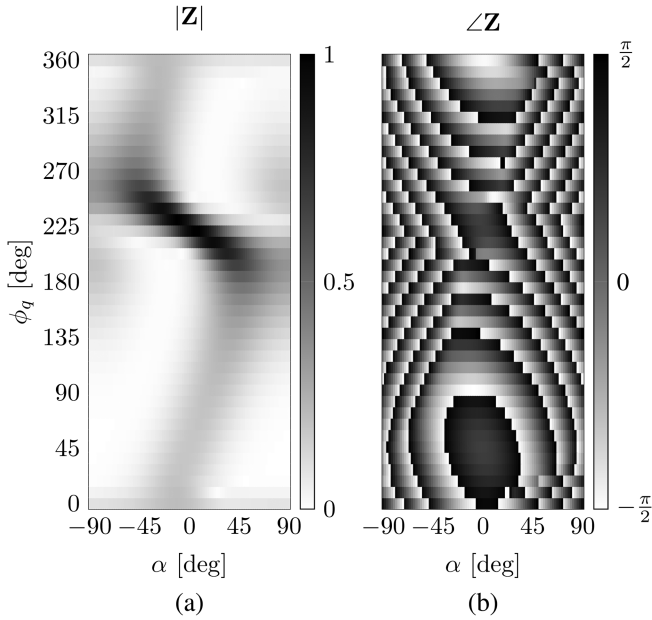
Fig. 3. Magnitude (left) and phase (right) of the AST related to the setup example of Fig. 2(a) with $r_{s_1} = (0.707\,\text{m}, 225°)$.

For this purpose, we utilize the canonical dual matrix [39] $\tilde{\boldsymbol{\Psi}} \in \mathbb{C}^{IW \times M}$ corresponding to the pseudo-inverse of $\boldsymbol{\Psi}$ given by

$$\tilde{\boldsymbol{\Psi}} = \left(\boldsymbol{\Psi}^H \boldsymbol{\Psi}\right)^{-1} \boldsymbol{\Psi}^H. \tag{18}$$

The *inverse Angular Space Transform* can be expressed using the properties of dual frames [39] as

$$\hat{\mathbf{p}} = \tilde{\boldsymbol{\Psi}}^H \mathbf{z}. \tag{19}$$

To achieve perfect reconstruction, $\boldsymbol{\Psi}$ should be either complete or over-complete, i.e., $IW \geq M$. In practice, the condition of over-completeness is favored for its flexibility in choosing dual matrices to $\boldsymbol{\Psi}$, which can guarantee a canonical dual with the minimum norm of the coefficients.

An example of the resulting AST considering the setup in Fig. 2(a) is depicted in Fig. 3. It is based on a single frequency point source signal computed at 1000 Hz in free field condition, considering a Signal-to-Noise Ratio (SNR) of $60\,\text{dB}$ in the received signal as per (2). We can evidently see the similarity between the $\Gamma(\rho_s, \phi_s)$ curve in Fig. 2(b) and the magnitude peaks of $|\mathbf{Z}|$ in Fig. 3(a). Additionally, $|\mathbf{Z}|$ exhibit maximum magnitude when ϕ_q is close to ϕ_s, particularly at $\alpha = 0$. This is expected as the beamformer is directly inline with the source and array center.

Fig. 3(b) shows the phase information of the AST, which depends on the relative position between the source and the sub-array centers. We can see symmetry in the phase pattern around the source angle $\phi_s = 225°$ for $\alpha = 0$, whereas at its supplementary angle the adjacent microphones are impinged with a similar phase, resulting in a larger pattern around $\phi_q = 45°$.

IV. EXAMPLES OF APPLICATION

One of the most relevant problems in the literature on sound field analysis is the localization of acoustic sources. In this section, we demonstrate some applications of AST to address this problem.

Mathematical solutions using the AST are presented to localize wideband signals. Moreover, the flexibility of the AST representation is emphasized by considering the case of distributed arrays. Finally, we provide comparison of localization performance between the AST and SRP-PHAT [36] methods for different reverberation and noise conditions.

A. Source Localization

Given the setup presented in Section III-A, we emulate a wideband source $s(\omega)$ with a white noise signal with zero mean and variance $\sigma_s^2 = 1$. From the acoustic image provided by the AST (14), we aim to estimate the source position represented by the polar coordinate $\hat{r}_s = (\hat{\rho}_s, \hat{\phi}_s)$.

Similar to the wideband DOA estimation proposed in [40], we employ the geometric mean to average the magnitude of angular space coefficients $|\mathbf{Z}(\omega_a)|$ across discrete temporal frequencies ω_a. As a matter of fact, the geometric mean is based upon the product operation where the lower frequencies remove any sidelobes of the beamformer, while the higher frequencies narrow the beamwidth and hence give better overall resolution [40]. Therefore, we define the wideband magnitude of AST as

$$\mathcal{Z} = \left(\prod_{a=1}^{A} E\{|\mathbf{Z}(\omega_a)|\}\right)^{1/A}, \quad \mathcal{Z} \in \mathbb{R}^{I \times W}, \tag{20}$$

where A is the number of considered frequency bins and the expectation is approximated by an average over J time frames as given by

$$E\{|\mathbf{Z}(\omega_a)|\} \approx \frac{1}{J} \sum_{j=0}^{J-1} \mathbf{Z}^{(j)}(\omega_a), \tag{21}$$

where $\mathbf{Z}^{(j)}(\omega_a)$ are the angular space coefficients computed at the jth time frame.

From the wideband magnitude coefficients $\mathcal{Z}$ (20), we can perform point-wise localization through a pattern-matching algorithm. The first step consists of a peak picking process applied to $\mathcal{Z}$. For each row $\mathcal{Z}_i$, the maximum peak is identified at $\hat{\alpha}_i$, retrieving the overall vector of maxima $\hat{\boldsymbol{\alpha}} = [\hat{\alpha}_0, \ldots, \hat{\alpha}_{I-1}]^T$. Notice that the curve traced by $\hat{\boldsymbol{\alpha}}$ in the angular space represents the source location based on its correspondence to the Γ function (12). Therefore, in the second step, source position estimate $\hat{r}_s$ can be computed by solving a minimization problem between the computed $\hat{\boldsymbol{\alpha}}$ and the analytical expression of the point source (11), namely

$$(\hat{\rho}_s, \hat{\phi}_s) = \arg \min_{\rho_s, \phi_s} ||\hat{\boldsymbol{\alpha}} - \Gamma(\rho_s, \phi_s)||^2. \tag{22}$$

We solved the ℓ_2 problem (22) with a gradient-based optimization consisting of a minimum constrained nonlinear multivariable function such that $\hat{\rho}_s \in [0, R)$ and $\hat{\phi}_s \in [0, 2\pi)$.
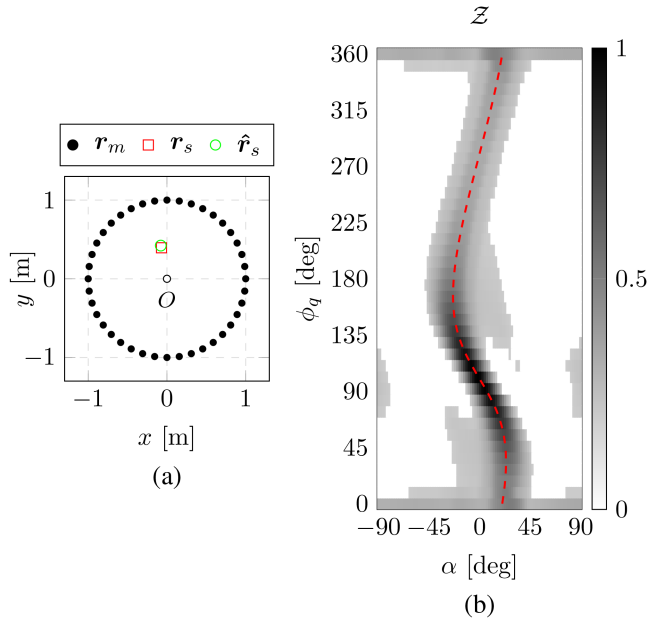
Fig. 4. Localization example for a wideband source signal. (a) shows the setup in the Euclidean space with a circular array of $M = 40$ microphones and $R = 1$ m along with the ground truth $\boldsymbol{r}_s = (0.4\,\text{m}, 100°)$ and the estimate $\hat{\boldsymbol{r}}_s$ of source position. The wideband AST $\mathcal{Z}$ is depicted in (b) along with the analytical mapping relation $\Gamma(\boldsymbol{r}_s)$ represented with a red dashed line.
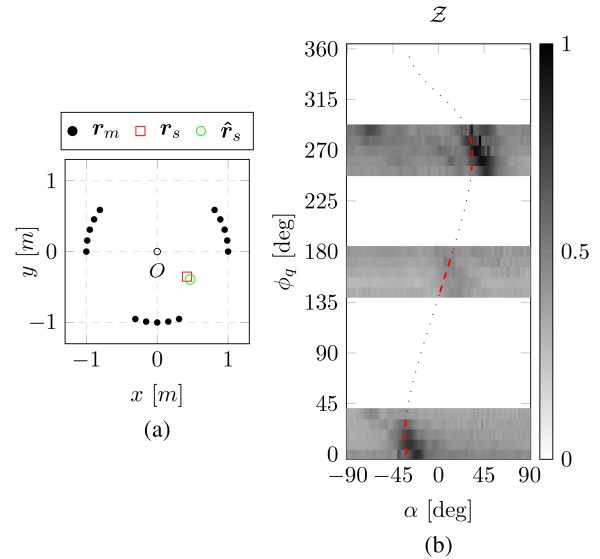

Fig. 5. Discontinuous circular array example for a wideband source signal. (a) shows the setup with three 5-microphone arc-shaped arrays of radius $R = 1$ m and a source at $\boldsymbol{r}_s = (0.55\,\text{m}, 320°)$. (b) shows the combined wideband AST $\mathcal{Z}$ along with the analytical mapping $\Gamma(\boldsymbol{r}_s)$ represented with a dotted black line and the mapping regions used in the minimization algorithm (22) emphasized as dashed red line.

In Fig. 4, we report an example of source localization application. We simulated a white noise point source signal with SNR $= 60\,\text{dB}$ in an anechoic room (i.e. $T_{60} = 0\,\text{s}$) with dimensions $6 \times 8 \times 3$ meters, recorded by $M = 40$ microphones placed on a circular array with $R = 1$ m. Considering the center of the circular array as the origin of the Cartesian coordinate system, the source is placed at $\boldsymbol{r}_s = (\rho_s, \phi_s) = (0.4\,\text{m}, 100°)$.

The wideband magnitude $\mathcal{Z}$ (20) of the sound field has been computed by considering a short-time analysis on the microphone signal (2). In particular, we applied a 512 point STFT with 32 ms of Hamming window, 50% overlap and 8 kHz of sampling frequency. We computed the L2 minimization algorithm (22) estimating $\hat{\boldsymbol{r}}_s$ at $(\hat{\rho}_s, \hat{\phi}_s) = (0.429\,\text{m}, 100.57°)$, hence with an error distance of $||\boldsymbol{r}_s - \hat{\boldsymbol{r}}_s|| = 0.029\,\text{m}$.

It is worth noticing that harmonics-based soundfield analysis approaches using circular microphone arrays face numerical instabilities, i.e., the *forbidden frequency problem* [41], [42], [43]. Nevertheless, the presented application of AST for source localization circumvents this problem using wideband processing (20). Furthermore, in the next section, we show the potentiality of AST to work with multiple linear arrays where the constraints traditionally associated with the forbidden frequency problem are not applicable. Therefore, potential concerns in the narrowband application of AST using circular arrays will be investigated in future works.

B. Discontinuous Circular Arrays for Localization

The proposed AST enables the simultaneous use of discontinuous circular microphone array to observe the same acoustic scene from multiple locations.

Microphones placed along the circumference of a circle with a radius R can be considered as multiple arrays arranged in an arc shape. The transform matrix (15) can also be used for distributed arc arrays that share the same radius R from a common origin reference. This requirement enables the application of the devised analytical model (11) to represent the spatial recordings in a unified encoded representation. Hence, the acoustic scene, as captured by different arc circle, can be analyzed through the AST (14) in a single angular space image without increasing the number of parameters.

Fig. 5(a) shows an example of discontinuous setup with three arrays and a source located at $\boldsymbol{r}_s = (\rho_s, \phi_s) = (0.55\,\text{m}, 320°)$. Each array is composed of 5 microphones with $\bar{\phi}_m = 9°$ and the central microphones of each array are at $18°$, $162°$, and $270°$ azimuth angles, respectively. Therefore, the resulting AST depicted in Fig. 5(b) stacks the output of the wideband extension of three different arrays using (15).

Notice that the location of the maxima of the wideband AST magnitude $\mathcal{Z}$ and the analytical function match in the angular space. Hence, the localization procedure can be applied for the valid $\Gamma(\rho_s, \phi_s)$. Using (22), we computed an estimate of the source position with an error distance of $||\boldsymbol{r}_s - \hat{\boldsymbol{r}}_s|| = 0.058\,\text{m}$. It is worth noticing that by approximating the arc of the circumference with a segment, it is possible to combine the measurements of multiple distributed ULAs into a single acoustic image by applying the AST . Experimental measurements are reported in Section V to assess the robustness of the proposed framework.

C. Towards Real Scenarios

In order to assess the performance of the devised acoustic imaging approach, we simulated different acoustic situations compatible with real scenarios.
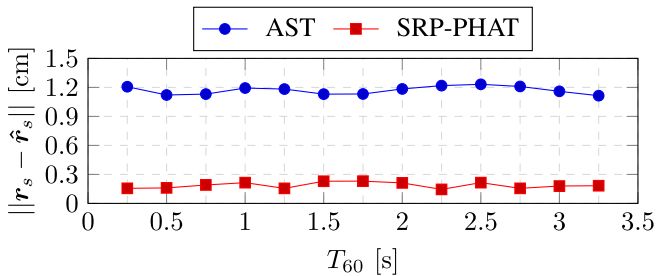
Fig. 6. Comparison of localization error between AST and SRP-PHAT as function of T_{60} with SNR = 50 dB fixed.

Fig. 8. Equipment used for the experimental scenario in the semi-anechoic room at Politecnico di Milano.
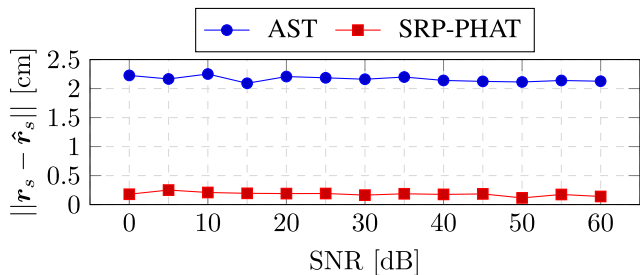

Fig. 7. Localization error between AST and SRP-PHAT as function of SNR with $T_{60} = 0.5$ s fixed.

We performed an analysis of localization errors at 13 different levels of reverberation time and thermal noise at the microphones. Therefore, we considered a room with dimensions $6 \times 8 \times 3$ meters, $T_{60} \in [0.25, 3.25]$s with a step of 0.25 s, and SNR $\in [0, 60]$dB with a step of 5 dB.

For the analysis with respect to the reverberation time, we fixed the SNR to 60 dB. The analysis of localization error as function of SNR was performed with $T_{60} = 0.5$ s.

For each configuration of SNR and T_{60}, we computed the RIRs through the Image Source Method (ISM) [44], [45] of 10 sources randomly placed inside a circular array with radius $R = 1$ and at the same elevations of the $M = 40$ microphones. Notice that, the formulation of AST has been derived for a general 3D sound field propagation. Nevertheless, in this preliminary study we analyze its applications in the specific case where the microphones and the sources lie on the same plane. From the computed wideband AST (20), we estimated each source position and we averaged the error distance for each source, namely $||\mathbf{r}_s - \hat{\mathbf{r}}_s||$.

In order to assess the performance of the devised approach with a state-of-the-art method, we compared the localization errors with the SRP-PHAT algorithm [35], [36]. Results highlight that both methods are robust for different reverberation and noise conditions. In particular, localization error as a function of T_{60} for the AST and SRP-PHAT are stable around the average values of 1.17 cm and 0.19 cm, respectively, as depicted in Fig. 6. Moreover, inspecting Fig. 7 we can notice that the two methods achieve a constant trend for the localization error also by varying the SNR at the microphones. In particular, we obtained an average value of 2.16 cm and 0.18 cm for the localization error of the AST and SRP-PHAT, respectively.
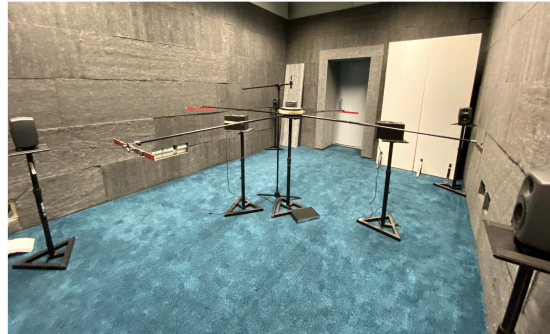
Differently from the SRP-PHAT, the devised method is more affected by the noise at the microphone rather than the reverberation time. However, we can notice that AST reaches good performance inline with SRP-PHAT with a difference in localization error of 0.98 cm and 1.98 cm varying T_{60} and SNR, respectively.

V. Experiments

In this section, we present the applications of AST in an experimental scenario. We employed ULA microphones in order to assess the localization robustness of the devised model with respect to linearization of the microphone arrays and the SRP-PHAT algorithm, thus proving to be compatible with more practical setups.

A. Setup

The acquisitions have been conducted in a semi-anechoic room at the Cremona campus of the Politecnico di Milano. The dimensions of the room are $3.64 \times 6.22 \times 3.84$ meters with measurement error of ± 0.01 m, measured reverberation time $T_{60} \approx 0.08$ s and temperature of $18\,°\mathrm{C}$.

Fig. 8 shows the equipment used for the experimental scenario. Two Yamaha MSP3[1] loudspeakers have been used as sound sources placed at elevation $z_m = 1.08$ m ± 0.01 m.

Measurements have been performed with 4 eSticks [46] arrays connected through ethernet port and synchronized using Dante[2] protocol. Each eStick is a ULA consisting of 16 MEMS microphones arranged in a rigid structure with $d = 3$ cm of linear spacing. The elevation of the sensors was fixed to $z_m = 1.08$ m ± 0.01 m from the ground with each eStick placed tangentially on a circumference with $R = 1.58$ m. In addition, we placed an omnidirectional Beyerdynamic MM1[3] microphone in the center of the circle in order to relate both the positions of microphones and sources to the circular array. Notice that this microphone is not used in the localization process, but only for preliminary calibration purposes. Notice that, the adopted setup considers sources and linear arrays placed on the same plane.

---

[1][Online]. Available: https://USA.yamaha.com/products/proaudio/speakers/msp3/index.html
[2][Online]. Available: https://www.audinate.com
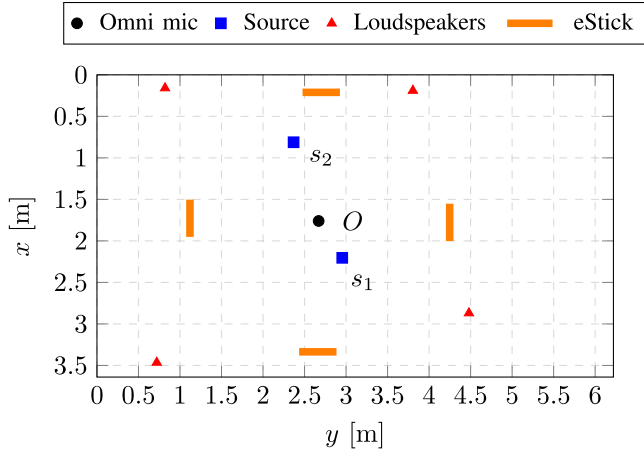[3][Online]. Available: https://global.beyerdynamic.com/mm-1.html

Fig. 9. Experimental setup using 4 eStick arrays depicted with orange thick lines and two sources s_1 and s_2 in blue square mark. The four red triangles represent the loudspeakers used for the calibration phase and the black dot shows the position of the origin reference measured by an omnidirectional microphone.

Fig. 10. eStick localization example for source s_1 in the semi-anechoic room. (a) depicts the setup with the estimated positions. (b) shows the wideband AST computed from the acquisitions along with the analytical solution of s_1 in the angular space represented in black dotted line. The dashed red line in (b) emphasizes the data used in the minimization algorithm for the source localization.

Although the microphone sensors capture 3D components of the acoustic propagation, they do not analyze the height dependency of the sound field. Therefore, we show the application of the devised method in the context of source localization by focusing on the horizontal plane where sources and microphones lie.

a) Calibration: a proper calibration phase has been performed in order to retrieve the actual locations of the eStick arrays and the ground truth positions of the sources. For this stage, we used four Genelec 8020C[4] loudspeakers $\{G1, G2, G3, G4\}$ placed around the room at the same elevation z_m of the eSticks. In particular, $\boldsymbol{r}_{G1} = (3.465, 0.72)$m, $\boldsymbol{r}_{G2} = (2.87, 4.48)$m, $\boldsymbol{r}_{G3} = (0.19, 3.805)$m, and $\boldsymbol{r}_{G4} = (0.16, 0.82)$m are the x and y coordinates of the loudspeakers used for the calibration by considering the origin of the Cartesian system to the bottom-left corner of the room. Fig. 9 reports the complete setup in the 2D plane at elevation $z_m = 1.08$ m.

RIRs have been acquired through the Exponential Sine Sweep (ESS) method [47] using a signal length of 10 s, sampled at 48 kHz with frequencies ranging from 50 Hz to 22 kHz. The signals have been driven with the sound card PreSonus Firestudio-Project[5] and recorded using REAPER[6].

From the Time Of Arrival (TOA) information extracted from the first peak of the RIRs corresponding to the acoustic direct path and the known positions of the Genelec loudspeakers $\boldsymbol{r}_\kappa$, we applied a calibration algorithm to estimate the unknown positions $\boldsymbol{r}_m$ of the 64 MEMS microphones by solving an optimization least-squares problem in 2D as

$$(\boldsymbol{r}_m, b) = \arg \min_{(\boldsymbol{r}_m, b)} \sum_{m, \kappa} \left[||\boldsymbol{r}_m - \boldsymbol{r}_\kappa|| - (d_{m,\kappa}^{\text{TOA}} - b) \right]^2, \quad (23)$$

where $d_{m,\kappa}^{\text{TOA}} = \text{TOA} \cdot c$ is the distance between the mth microphone and κ^{th} loudspeaker estimated from the measured RIR and b represents an estimate of the signal delay caused by the

buffer of the overall acquisition system expressed in meter. Therefore, from (23) we computed the positions of the four eStick arrays.

In a second step, after computing RIR measurements between the sources s_1 and s_2 and each eStick, we retrieved the positions of the two sources from (23) using the 64 microphones as known positions. It is worth noticing that this calibration phase is common between SRP-PHAT baseline and the proposed technique. Therefore, the resulting source estimates will be used as ground truth during the experiments as $\boldsymbol{r}_{s_1}$ and $\boldsymbol{r}_{s_2}$.

b) Measurement: we performed six different acquisitions of the two sources by changing the microphone positions. We rotated the eSticks with a step of $15°$ with respect to the circle origin O, thus having measurements along all $360°$ circular aperture. For each setup, we calibrated the microphone positions through (23) from the RIRs of the Genelec loudspeakers. Then, we computed the RIRs of the two sources and the relative ground truth positions.

From (23), we found that the overall signal delay caused by the buffer is 3.7 ms, and $\boldsymbol{r}_{s_1} = (2.204, 2.954)$m and $\boldsymbol{r}_{s_2} = (0.812, 2.369)$m are the two source positions in Cartesian coordinates.

In the next section, we present the results of localization by comparing the devised AST approach with the SRP-PHAT baseline in terms of distance error between the estimated and ground truth source position, namely

$$e = ||\boldsymbol{r}_s - \hat{\boldsymbol{r}}_s||. \quad (24)$$

We will use $e^{(\text{AST})}$ and $e^{(\text{SRP-PHAT})}$ for referring to the localization error coming from the devised method and the reference, respectively. It is worth noticing that that the experiment setup
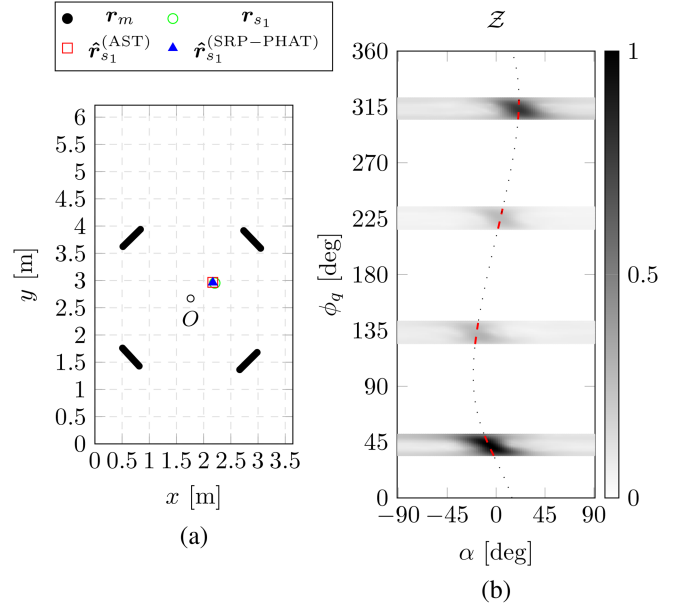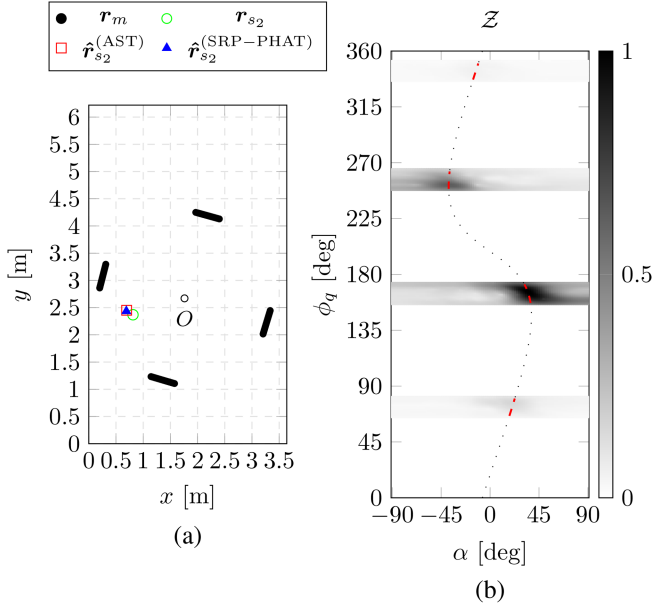
Fig. 11. eStick localization example for source s_2 in the semi-anechoic room. (a) depicts the setup with the estimated positions. (b) shows the wideband AST computed from the acquisitions along with the analytical solution of s_2 in the angular space represented in dotted line and the data used in the minimization algorithm emphasized as dashed red line.

adopts linear arrays, so they don't analyze the height dependency of the sound field. This implies that the microphone sensors capture the 3D components of the acoustic propagation, but the multiple beamforming operations adopted to compute the AST (14) work only in the horizontal plane.

B. Results

a) Experimental equiangular configurations: examples of localization performed using the measurements is depicted in Fig. 10 and in Fig. 11 for different microphone configurations and considering s_1 or s_2, respectively.

Fig. 10(a) shows the setup of the eSticks along with the ground truth position of source s_1 and the positions $\hat{r}_{s_1}^{AST}$ and $\hat{r}_{s_1}^{SRP-PHAT}$ obtained through proposed and reference approaches, respectively. In Fig. 10(b), we report the wideband AST $\mathcal{Z}$ (20) computed from the eStick acquisitions along with the analytical curve Γ, depicted in dotted line.

After a proper peak picking process, we applied the minimization algorithm (22) for valid observation windows of $\mathcal{Z}$, thus obtaining $\hat{r}_{s_1}^{AST} = (2.167, 2.966)$m. Meanwhile, we estimated the s_1 position with the SRP-PHAT algorithm yielding $\hat{r}_{s_1}^{(SRP-PHAT)} = (2.168, 2.963)$m. Notice that for both methods, we get a distance error e (24) less than 4 cm. In particular, $e^{(AST)} = 0.0389$ m and $e^{(SRP-PHAT)} = 0.0371$ m.

Fig. 11 shows a localization example for s_2 with a different configuration of microphones. In this setup, we obtained higher localization error, i.e., $e^{(AST)} = 0.142$ m and $e^{(SRP-PHAT)} = 0.138$ m. This is due to the position of the source. Indeed, s_2 is closer to the circumference on which the eSticks lie, i.e., $\rho_{s_2} = 0.96$ m and the linearization of the microphone arrays introduces
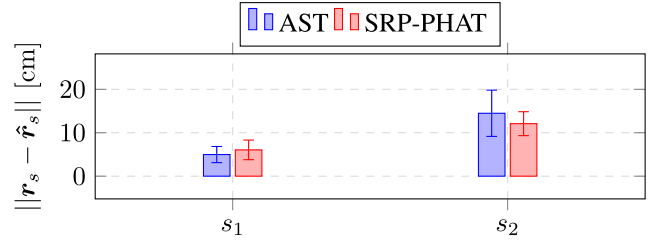

Fig. 12. Mean and standard deviation comparison of s_1 and s_2 localization errors between AST and SRP-PHAT computed from the six different measurement configurations.
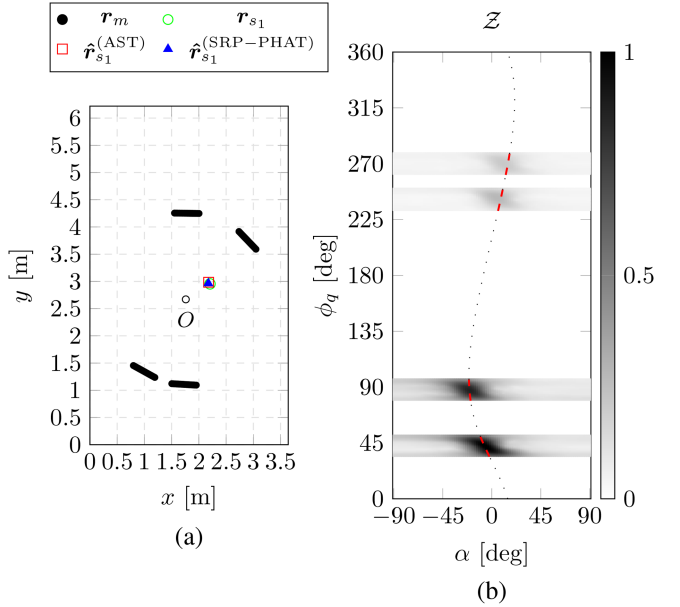

Fig. 13. Localization example for source s_1 in the semi-anechoic room with a nonequiangular configuration of eSticks. (a) depicts the setup. (b) shows the wideband AST along with the analytical solution.

a non negligible distortion in the model of the direction of arrival (11).

Moreover, we collected the source position estimates for all the six experimental microphone configurations. In Fig. 12, we report the overall results in terms of the mean and standard deviation of $e = ||r_s - \hat{r}_s||$. We can notice that for s_1 AST reaches better results with respect to SPR-PHAT, both in terms of mean and standard deviation of the localization error. In particular, AST estimates are below 7 cm of error for all the different setups. However, this behavior is the opposite for the case of source s_2 since the discrepancy between the measurement setup and the AST model introduces non-negligible errors.

b) Simulations of nonequiangular configurations: in order to further investigate the localization error between the two sources and the two methods, we simulated different eStick configurations from the measurements collected in the semi-anechoic room at Politecnico di Milano. In particular, we randomly selected 20 different setups by considering nonequiangular placement of the eSticks. In Figs. 13 and 14, we show two examples of simulated configurations along with the localization results for s_1 and s_2, respectively, with the proposed AST application
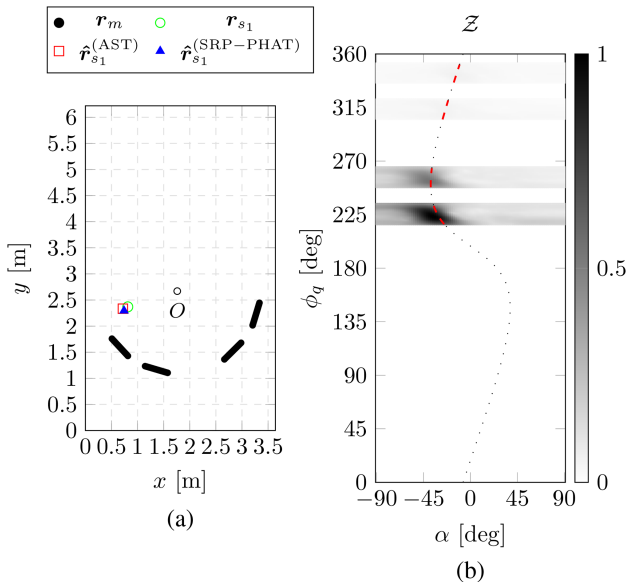
Fig. 14. Localization example for source s_2 in the semi-anechoic room with a nonequiangular configuration of eSticks. (a) depicts the setup. (b) shows the wideband AST along with the analytical solution.

Fig. 15. Mean and standard deviation comparison of s_1 and s_2 localization errors between AST and SRP-PHAT computed from 20 different nonequiangular placement of eSticks and different source signals. (a) shows results when considering white noise signal with zero mean and variance $\sigma^2 = 1$, and (b) with a speech signal.

and SRP-PHAT method. Notice that, in both examples, the localization error computed with AST is below $10\,\mathrm{cm}$.

Furthermore, to prove the generalization with respect to different distributed configurations and different source signals, we evaluate the mean and standard deviation of the localization errors between the 20 nonequiangular eStick setups. Moreover, we considered two different signal data sets for s_1 and s_2: white noise with zero mean and variance $\sigma^2 = 1$ and speech signals. Results are depicted in Fig. 15(a) and (b), respectively.

Observing the results in Fig. 15 and comparing the localization errors with respect to the equiangular and non-speech cases in Fig. 12, it is possible to note that the performance is aligned between the different scenarios. As a matter of fact, the average localization errors considering white noise sources are within $14\,\mathrm{cm}$ for both s_1 and s_2. Moreover, when speech signal is considered, the localization performance is reduced with a maximum difference in the average error of $2.7\,\mathrm{cm}$ with respect to the equiangular, white noise case for s_1.

Although the results highlight larger localization error for s_2 for both the devised AST model and the SRP-PHAT algorithm, the error is less than $15\,\mathrm{cm}$ on average, thus proving the effectiveness of AST also in experimental scenarios. Moreover, it is worth noticing that even if we are assuming that the source is placed inside the circle, differently from the SRP-PHAT, the devised localization algorithm in the angular space does not require information about a bounded region of interest to estimate the source positions. Hence, future extensions will be able to fully benefit from this property for example, in contexts where it is not possible to assume a priori position of the source.

VI. Conclusion

In this paper, we proposed an approach for sound field analysis acquired by circular arrays based on acoustic imaging. We defined the "curved space" domain that maps acoustic entities of
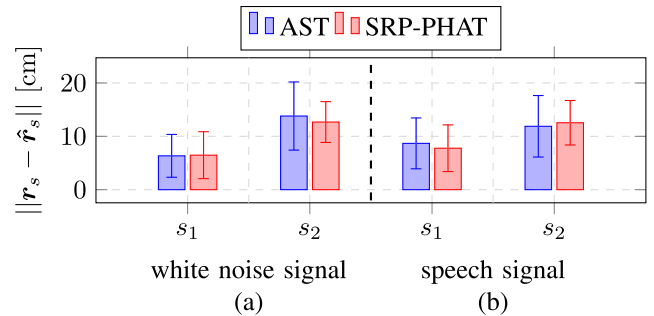
the exterior acoustic field captured by the microphone sensors using a parametrization based on the steering angles and the angle position of the local observation window of the circular arrays. The Angular Space Transform (AST) is introduced to efficiently encode the spatial information onto the angular space using discrete Gabor frames. We show the efficiency of such methodology to globally represent in a unified image the information acquired by a discontinuous circular array, thus from multiple arrays without introducing further parameters.

Pattern-matching techniques facilitate the exploitation of the angular space, particularly for applications like sound source localization. The theoretical derivations are substantiated with simulations varying the acoustic conditions of reverberation time and noise level sensors and with experimental measurements. Especially, we analyzed the flexibility of the AST when multiple Uniform Linear Arrays are employed in an arrangement along the circumference of a virtual circular array. We assessed the performance of such technique by comparing the localization errors with respect to the SRP-PHAT algorithm, thus proving to achieve comparable results with state-of-the-art methods.

The examples presented in this paper serve as an initial proof of concept for acoustic imaging techniques based on angular space using different microphone array geometries, and future works will expand on this idea. In particular, we plan to investigate multiple source localization and the mapping of acoustic reflectors, such as walls or windows, onto the angular space. Moreover, due to the intrinsic ability of the angular space to encode the acoustic information of the sound field bounded by the microphone array, we aim to apply the devised method in the context of speech separation and dereverberation applications. Hence, discarding undesired components coming from the external regions of the microphone array. Additionally, taking advantage of the circular array setup, we foresee the integration of the devised acoustic imaging technique with Spherical Harmonic representation and to combine high-order descriptors with the spatial information derived by the AST. Nonetheless, an analogous approach can be explored for external viewing of the sound field, which can be developed for spherical microphone array analysis.

REFERENCES

[1] B. Rafaely et al., "Spatial audio signal processing for binaural reproduction of recorded acoustic scenes–review and challenges," *Acta Acustica*, vol. 6, 2022, Art. no. 47.

[2] M. Olivieri et al., "Audio information retrieval and musical acoustics," *IEEE Instrum. Meas. Mag.*, vol. 24, no. 7, pp. 10–20, Oct. 2021.

[3] L. McCormack, A. Politis, T. McKenzie, C. Hold, and V. Pulkki, "Object-based six-degrees-of-freedom rendering of sound scenes captured with multiple ambisonic receivers," *J. Audio Eng. Soc.*, vol. 70, no. 5, pp. 355–372, 2022.

[4] W. S. Gan, *Acoustical Imaging: Techniques and Applications for Engineers*. Hoboken, NJ, USA: Wiley, 2012.

[5] N. Epain and C. T. Jin, "Super-resolution sound field imaging with subspace pre-processing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 350–354.

[6] M. Pezzoli, J. J. Carabias-Orti, M. Cobos, F. Antonacci, and A. Sarti, "Ray-space-based multichannel nonnegative matrix factorization for audio source separation," *IEEE Signal Process. Lett.*, vol. 28, pp. 369–373, 2021.

[7] F. Borra, F. Antonacci, A. Sarti, and S. Tubaro, "Extraction of acoustic sources for multiple arrays based on the ray space transform," in *Proc. Hands-Free Speech Commun. Microphone Arrays*, 2017, pp. 146–150.

[8] M. Pezzoli, F. Borra, F. Antonacci, S. Tubaro, and A. Sarti, "A parametric approach to virtual miking for sources of arbitrary directivity," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 2333–2348, 2020.

[9] D. Marković, F. Antonacci, L. Bianchi, S. Tubaro, and A. Sarti, "Extraction of acoustic sources through the processing of sound field maps in the ray space," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 12, pp. 2481–2494, Dec. 2016.

[10] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, vol. 1. Berlin, Germany: Springer, 2008.

[11] K. Imoto and N. Ono, "Spatial cepstrum as a spatial feature using a distributed microphone array for acoustic scene analysis," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 6, pp. 1335–1343, Jun. 2017.

[12] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, U.K.: Academic Press, 1999.

[13] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2005, pp. 150–153.

[14] J. Ahonen, M. Kallinger, F. Küch, V. Pulkki, and R. Schultz-Amling, "Directional analysis of sound field with linear microphone array and applications in sound reproduction," in *Audio Engineering Society Convention*, vol. 124. New York, NY, USA: Audio Engineering Society, 2008.

[15] M. Olivieri, M. Pezzoli, F. Antonacci, and A. Sarti, "A physics-informed neural network approach for nearfield acoustic holography," *Sensors*, vol. 21, no. 23, 2021, Art. no. 7834.

[16] L. Kumar and R. M. Hegde, "Near-field acoustic source localization and beamforming in spherical harmonics domain," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3351–3361, Jul. 2016.

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

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

[19] D. Markovic, G. Sandrini, F. Antonacci, A. Sarti, and S. Tubaro, "Plenacoustic imaging in the ray space," in *Proc. Int. Workshop Acoust. Signal Enhancement*, 2012, pp. 1–4.

[20] L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro, "The ray space transform: A new framework for wave field processing," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5696–5706, Nov. 2016.

[21] F. Borra et al., "A fast ray space transform for wave field processing using acoustic arrays," in *Proc. 28th Eur. Signal Process. Conf.*, 2021, pp. 186–190.

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

[23] L. Comanducci, F. Borra, P. Bestagini, F. Antonacci, A. Sarti, and S. Tubaro, "Ray space transform interpolation with convolutional autoencoder," in *Proc. 16th Int. Workshop Acoust. Signal Enhancement*, 2018, pp. 261–265.

[24] A. M. Torres, J. J. Lopez, B. Pueo, and M. Cobos, "Room acoustics analysis using circular arrays: An experimental study based on sound field plane-wave decomposition," *J. Acoust. Soc. Amer.*, vol. 133, no. 4, pp. 2146–2156, 2013.

[25] T. D. Abhayapala and A. Gupta, "Spherical harmonic analysis of wavefields using multiple circular sensor arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1655–1666, Aug. 2010.

[26] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acoustical Soc. Amer.*, vol. 118, no. 5, pp. 3094–3103, 2005.

[27] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "PSD estimation and source separation in a noisy reverberant environment using a spherical microphone array," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 9, pp. 1594–1607, Sep. 2018.

[28] P. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 3, pp. 647–658, Mar. 2014.

[29] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, vol. 2, pp. II–1781.

[30] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 135–139, Jan. 2018.

[31] F. Borra, I. D. Gebru, and D. Markovic, "Soundfield reconstruction in reverberant environments using higher-order microphones and impulse response measurements," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 281–285.

[32] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Sound field separation in a mixed acoustic environment using a sparse array of higher order spherical microphones," in *Proc. Hands-Free Speech Commun. Microphone Arrays*, 2017, pp. 151–155.

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

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

[35] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2007, vol. 1, pp. I–121–I–124.

[36] H. Do and H. F. Silverman, "SRP-PHAT methods of locating simultaneous multiple talkers using a frame of microphone array data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2010, pp. 125–128.

[37] D. L. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, vol. 93, 4th ed. Cham, Switzerland: Springer, 1998.

[38] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. Amer.*, vol. 116, no. 4, pp. 2149–2157, 2004.

[39] T. Werther, Y. Eldar, and N. Subbanna, "Dual Gabor frames: Theory and computational aspects," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4147–4158, Nov. 2005.

[40] M. R. Azimi-Sadjadi, A. Pezeshki, and N. Roseveare, "Wideband DOA estimation algorithms for multiple moving sources using unattended acoustic sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 4, pp. 1585–1599, Oct. 2008.

[41] S. Koyama, G. Chardon, and L. Daudet, "Optimizing source and sensor placement for sound field control: An overview," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 696–714, 2020.

[42] K. SongGong, H. Chen, and W. Wang, "Indoor multi-speaker localization based on Bayesian nonparametrics in the circular harmonic domain," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1864–1880, 2021.

[43] X. Han, M. Wu, Z. Han, and J. Yang, "Sound source localization using multiple circular microphone arrays based on harmonic analysis," *J. Acoust. Soc. Amer.*, vol. 149, no. 5, pp. 3517–3523, 2021.

[44] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[45] E. Habets, "ehabets/rir-generator: Rir generator," Oct. 2020, doi: 10.5281/zenodo.4117640.

[46] M. Pezzoli et al., "A dante powered modular microphone array system," *Audio Eng. Soc. Conv.*, vol. 145, New York, NY, USA: Audio Engineering Society, 2018.

[47] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*. New York, NY, USA: Audio Engineering Society, 2000.

Marco Olivieri (Student Member, IEEE) received the M.Sc. degree (*cum laude*) in music and acoustic engineering from Politecnico di Milano, Milan, Italy, in 2020. In 2020, he joined the Ph.D. Program in information technology. His main research interests include computational acoustics and space-time audio processing using deep learning techniques. In particular, focusing on the characterization of vibrating surfaces of complex shape with novel combinations of Nearfield Acoustic Holography analysis and Convolutional Neural Networks. He is currently working on speech recognition and enhancement in noisy environments through a combination of classical signal processing methods and deep learning and on new model representations for soundfield processing.

Amy Bastine (Student Member, IEEE) received the Master of Engineering degree in digital systems and telecommunications from Australian National University (ANU), Canberra, ACT, Australia, in 2019. She is currently working toward the Ph.D. degree working with the Audio and Acoustic Signal Processing Group, ANU. Her research interests include room acoustic parameterization for soundfield synthesis and binaural soundfield translation using headwearable microphone arrays.

Mirco Pezzoli received the M.S. degree (*cum laude*), in computer engineering and the Ph.D. degree in information engineering from Politecnico di Milano, Italy, in 2017 and 2021, respectively. He is currently a Postdoctoral Researcher with the Politecnico di Milano. His main research interests include space-time audio signal processing and musical acoustics.

Fabio Antonacci (Member, IEEE) was born in Bari, Italy, in 1979. He received the Laurea degree in telecommunication engineering and the Ph.D. degree in information engineering from Politecnico di Milano, Milan, Italy, in 2004 and 2008, respectively. He is currently an Associate Professor with Politecnico di Milano. He is the author of more than 130 articles in proceedings of international conferences and peer-reviewed journals. His research focuses on musical acoustics, in particular on the development of innovative non invasive measurement methodologies. He is also active in the areas of space-time processing of audio signals, for both speaker and microphone arrays (source localization, acoustic scene analysis, rendering of spatial sound) and on modeling of acoustic propagation (visibility-based beam tracing).

Thushara Abhayapala (Senior Member, IEEE) received the B.E. degree in engineering and the Ph.D. degree in telecommunication engineering from Australian National University (ANU), Canberra, ACT, Australia, in 1994 and 1999, respectively. He is currently the Professor of audio and acoustic signal processing with ANU. He held a number of leadership positions, including the Deputy Dean with the ANU College of Engineering and Computer Science from 2015 to 2019, Head of the ANU Research School of Engineering from 2010 to 2014, and the Leader of the Wireless Signal Processing Program with the National ICT Australia, Sydney, NSW, Australia, from 2005 to 2007. He has supervised 44 Ph.D. students and co-authored more than 300 peer-reviewed papers. His research interests include the areas of spatial audio and acoustic signal processing, and multichannel signal processing. Among many contributions, he is one of the first researchers to use spherical harmonic based Eigen-decomposition in microphone arrays and to propose the concept of spherical microphone arrays, and was one of the first to show the fundamental limits of spatial sound field reproduction using arrays of loudspeakers and spherical harmonics this is now termed as higher order Ambisonics. He also made outstanding contributions to the problem of the multizone sound field reproduction. He worked in industry for two years, before his doctoral study and has active collaboration with a number of companies. He was the Co-chair of IEEE WASPAAA 2021. He was an Associate Editor for IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING. From 2011 to 2016, he was the Member of the Audio and Acoustic Signal Processing Technical Committee of the IEEE Signal Processing Society. He is the Fellow of Engineers Australia.

Augusto Sarti (Senior Member, IEEE) received the Ph.D. degree in information engineering from the University of Padova, Padua, Italy, (along with a joint Graduate Program from the University of California, Berkeley, Berkeley, CA, USA) in 1993. In 1993, he joined the Politecnico di Milano (PoliMI), Milan, Italy, where he is currently a Full Professor. From 2013 to 2017, he held a Professorship with the University of California, Davis, Davis, CA. At PoliMI, he currently coordinates the research activities of the Musical Acoustics Lab and Sound and Music Computing Lab, and M.Sci. Program in Music and Acoustic Engineering. He has coauthored more than 300 scientific publications on international journals and congresses, and numerous patents in the multimedia signal processing area. His research interests include audio and acoustic signal processing, with particular focus on audio and acoustic signal processing, music information retrieval, and musical acoustics. He was two terms with the IEEE Technical Committee on Audio and Acoustics Signal Processing. He also was an Associate Editor for IEEE/ACM TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING, and the Senior Area Editor of IEEE SIGNAL PROCESSING LETTERS. He is currently with the EURASIP Board of Directors.