Linearly Constrained Minimum Variance Source Localization and Spectral Estimation

Jacek Dmochowski, Jacob Benesty, *Senior Member, IEEE*, and Sofiène Affes, *Senior Member, IEEE*

***Abstract*—A signal's spectrum is a representation of the signal in terms of elementary basis functions which facilitates the extraction of desired information. For a temporal signal, the spectrum is one-dimensional and expresses the time-domain signal as a linear combination of sinusoidal basis functions. A space–time signal possesses a multidimensional Fourier transform known as the wavenumber–frequency spectrum, which represents the space–time signal as a weighted summation of monochromatic plane waves. The spatial and temporal frequencies are not separable, as spatial frequency is itself a function of the temporal frequency. Thus, it seems natural to analyze and estimate the spatial and temporal frequency components in tandem. It is therefore surprising that conventional spectral estimation methods focus on either the spatial or temporal dimension, without any regard for the other. Spatial spectral estimation is commonly referred to as source localization, as the direction of the wavenumber vector is indeed the direction of propagation. Conventional methods analyze a solely spatial aperture without accounting for the temporal structure of the desired signal. Conversely, temporal spectral estimation is performed using a single sensor, and thus the signal aperture is purely temporal. This paper proposes a spatiotemporal framework for spectral estimation based on the linearly constrained minimum variance (LCMV) beamforming method proposed by Frost in 1972. The aperture consists of an array of sensors, each storing a set of previous temporal samples. It is first shown that by taking into account the temporal structure of the desired signal, the ensuing source location estimate is more robust to the effects of noise and reverberation. Unlike conventional localizers, the LCMV steered beam *temporally* focuses the array onto the desired signal. The desired signal is modeled by an autoregressive (AR) process, and the resulting AR coefficients are embedded in the linear constraints. As a result, the rate of anomalous estimates is significantly reduced as compared to existing techniques. Moreover, it is then demonstrated that by employing multiple sensors and steering the array to the assumed source location, the estimate of the desired signal's temporal spectrum contains a lesser contribution from the unwanted noise and reverberation.**

***Index Terms*—Linearly constrained minimum variance (LCMV), microphone arrays, minimum variance distortionless response (MVDR), source localization, spectral estimation.**

I. Introduction

IN ANALYZING a given signal, it is common to evaluate the signal's spectrum: a representation of its underlying structure in terms of elementary basis functions. For a one-dimensional temporal signal, the standard Fourier transform expresses the time-domain signal as a linear combination of sinusoidal basis functions. In the multidimensional space–time domain, a space–time signal is decomposed into a weighted summation of monochromatic (i.e., occurring at a single temporal frequency) plane waves. In both cases, the task of *spectral estimation* is to determine the values of the weighting coefficients.

Classical temporal spectral estimation methods [1] analyze a purely temporal aperture; popular methods include the periodogram [2], the minimum variance distortionless response (MVDR) technique [3], [4], linear predictive-based methods [5], and Burg's maximum entropy principle [6]. Spatial spectral estimation is a more complex problem as the *wavenumber–frequency* spectrum [7] is a function of both spatial and temporal frequencies. As a result, conventional "narrowband" methods estimate a spatial spectrum for each temporal frequency [8]. The spatial spectral estimation problem is commonly referred to as source localization or direction-of-arrival estimation (DOA) [9]—this is due to the fact that the direction of the wavenumber vector, the spatial frequency variable, is indeed the direction of propagation of the signal. Spatial spectral estimation methods employ a spatial aperture, usually in the form of an array of sensors.

When the signal of interest is broadband in nature, the classical narrowband approach is inconvenient due to the need to assimilate all of the spatial spectra into a single location estimate. As a result, "broadband" source localization methods produce a single spatial spectrum which implicitly integrates the full temporal frequency range. The earliest methods are composed of two steps and are based on the time-differences-of-arrival (TDOAs) across the array which are then mapped to the source location using one of a number of existing techniques [10], [11]. Time delay estimation (TDE) [12], [13] is the process of estimating the TDOAs using the cross-correlation functions across the array of sensors. A more robust category of broadband source localization methods are based on parameterized spatial correlation and are detailed in [14]: this class of estimators includes the popular steered response power (SRP) method [15] as well as the broadband MVDR method [16].

This paper proposes a spectral estimation technique which is based on a spatiotemporal aperture: an array of sensors, each storing a finite number of previous temporal samples. The method is rooted in the linearly constrained minimum variance (LCMV) technique proposed by Frost [17]. We first present a novel source localization method which is based on a steered LCMV beamformer which is also temporally matched to the desired signal. By explicitly modeling the colored nature of the desired signal as an autoregressive (AR) process, a dramatic improvement in the source localization performance in noisy

Manuscript received December 17, 2007; revised July 22, 2008. Current version published October 17, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. George Tzanetakis.

The authors are with the Université du Québec, INRS-EMT, Montréal, QC H5A 1K6, Canada (e-mail: dmochow@emt.inrs.ca).

Digital Object Identifier 10.1109/TASL.2008.2005029

and reverberant conditions ensues. Conversely, it is also shown that knowledge of the source's location or DOA may be utilized to yield a more accurate temporal spectral estimate. To that end, we present a multichannel approach to temporal spectral estimation: we steer an array of sensors to the assumed source direction, and then pass the received and time-aligned signals through a multichannel filterbank which estimates the energy of the wavefield at each temporal frequency.

The paper is structured as follows. Section II presents the signal propagation model employed throughout the development. Section III briefly covers the wavenumber–frequency spectrum and explains the relationship between spatial and temporal frequency in propagating waves. Section IV describes the proposed source localization technique which takes the form of a steered delay-filter-and-sum beamformer (DFSB) that is matched to the temporal structure of the desired signal—the known temporal structure of the desired signal is utilized to improve the estimate of the spatial spectrum. Analogously, Section V proposes a method for temporal spectral estimation which incorporates known spatial information of the signal; the sensors are steered to the assumed source location prior to a filterbank which estimates the temporal frequency content—the known spatial information of the signal is utilized to improve the estimate of the temporal spectrum. The proposed methods are evaluated in a computer simulation in Section VI; concluding statements are given in Section VII.

II. Signal Model

Assume that an array of N sensors samples the wave field, and that the environment is anechoic. Assuming a single point source, the output of sensor n at time k is then modeled as

$$x_n(k) = \alpha_n(\mathbf{r}_\mathrm{s}) s[k - \tau - \mathcal{F}_{1n}(\mathbf{r}_\mathrm{s})] + v_n(k) \tag{1}$$

where $\alpha_n(\mathbf{r}_\mathrm{s}), n = 1, 2, \ldots, N$, models the attenuation of the source signal at sensor n as a function of the source location $\mathbf{r}_\mathrm{s} = (r_\mathrm{s}, \phi_\mathrm{s}, \theta_\mathrm{s})$, where r_s, ϕ_s, and θ_s denote the range, elevation, and azimuth, respectively, s is the source signal, τ is the propagation time (in samples) from the source to sensor 1, $\mathcal{F}_{ij}(\mathbf{r}_\mathrm{s})$ is a function that relates the source position to the relative delay between microphones i and j, and v_n is the additive noise at sensor n. In the free-field case

$$\alpha_n(\mathbf{r}_\mathrm{s}) \propto \frac{1}{d_{n,\mathrm{s}}(\mathbf{r}_\mathrm{s})} \tag{2}$$

where $d_{n,\mathrm{s}}(\mathbf{r}_\mathrm{s})$ is the distance from the source to the nth sensor. Typically, the attenuation is modeled as

$$\alpha_n(\mathbf{r}_\mathrm{s}) = \frac{d_{n_\mathrm{ref},\mathrm{s}}(\mathbf{r}_\mathrm{s})}{d_{n,\mathrm{s}}(\mathbf{r}_\mathrm{s})} \tag{3}$$

where n_ref is the index of a reference sensor.

The nature of the function $\mathcal{F}_{ij}$ depends on the array geometry. When the source lies in the far-field of the array, $\mathcal{F}_{ij}$ is well-approximated by a two-dimensional function of ϕ_s and θ_s; moreover, if we further assume that the source lies in the same plane as the microphones, $\mathcal{F}_{ij}$ reduces to a function of just the azimuth angle of arrival θ_s.

A. Three-Dimensional Near-Field Model

In the most general case, $\mathcal{F}_{ij}$ is related to the distances between the source and the sensors i and j

$$\mathcal{F}_{ij}(\mathbf{r}_\mathrm{s}) = \frac{d_{j,\mathrm{s}}(\mathbf{r}_\mathrm{s}) - d_{i,\mathrm{s}}(\mathbf{r}_\mathrm{s})}{c} \tag{4}$$

where c is the speed of propagation.

B. Far-Field Model

When the distance from the source to the array is large in comparison to the spatial aperture size, the far-field assumption is valid and the incoming wave front may be assumed to be planar. In that case, $\mathcal{F}_{ij}$ becomes independent of the source range

$$\mathcal{F}_{ij}(\phi_\mathrm{s}, \theta_\mathrm{s}) = \frac{\boldsymbol{\zeta}^T(\phi_\mathrm{s}, \theta_\mathrm{s})(\mathbf{z}_j - \mathbf{z}_i)}{c} \tag{5}$$

where

$$\boldsymbol{\zeta}(\phi_\mathrm{s}, \theta_\mathrm{s}) = [\sin\phi_\mathrm{s}\cos\theta_\mathrm{s} \quad \sin\phi_\mathrm{s}\sin\theta_\mathrm{s} \quad \cos\phi_\mathrm{s}]^T \tag{6}$$

is a unit vector which points in the direction of propagation of the source, and $\mathbf{z}_i = [z_{i,x}\ z_{i,y}\ z_{i,z}]^T$ is the position vector of the ith sensor.

If the source lies in the same plane as the array of sensors (this assumes that the array itself is planar), $\boldsymbol{\zeta}$ becomes independent of the elevation angle, and as a result, $\mathcal{F}_{ij}$ loses its dependence on ϕ_s. In that case, $\mathbf{z}_i = [z_{i,x}\ z_{i,y}]^T$ and $\boldsymbol{\zeta}$ effectively become two-dimensional, with

$$\boldsymbol{\zeta}(\phi_\mathrm{s}, \theta_\mathrm{s})|_{2-\mathrm{D}} = [\cos\theta_\mathrm{s} \quad \sin\theta_\mathrm{s}]^T. \tag{7}$$

III. Wavenumber–Frequency Spectrum

Physically, no signal exists solely in the spatial or temporal domain; however, for ease of analysis, we choose to separate the two domains by analyzing the signal observed at, for example, a particular point in space. Nevertheless, real signals are most accurately modeled in the space–time domain: $f(\mathbf{z}, t)$, where f is the value of the signal, $\mathbf{z} = [x\ y\ z]^T$ is the observation point in Cartesian coordinates, and t denotes time. As mentioned in the introduction, Fourier analysis generalizes to the representation of multidimensional signals. In this case, the appropriate transform is termed the *wavenumber–frequency* spectrum, which decomposes an arbitrary space–time signal into a linear combination of monochromatic plane waves.

For a signal $f(\mathbf{z}, t)$, its wavenumber–frequency representation is given by the inverse transform [7]

$$f(\mathbf{z}, t) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\mathbf{k}, \omega) e^{j(\omega t - \mathbf{k}^T \mathbf{z})} d\mathbf{k} d\omega \tag{8}$$

where $e^{j(\omega t - \mathbf{k}^T \mathbf{z})}$ are the basis functions, $F(\mathbf{k}, \omega)$ are the weighting coefficients, ω is the angular temporal frequency variable, and $\mathbf{k} = [k_x\ k_y\ k_z]^T$ is termed the *wavenumber* vector with

$$\mathbf{k} = \frac{\omega}{c} \boldsymbol{\zeta}(\phi_\mathrm{s}, \theta_\mathrm{s}). \tag{9}$$

Notice that the spatial frequency variable is itself a function of the temporal frequency (in its magnitude). It is easy to see that at any fixed point in time, the value of the basis function $e^{j(\omega t-\mathbf{k}^T\mathbf{z})}$ is constant across a plane $\mathbf{k}^T\mathbf{z} = L$, where L is a constant. Moreover, at any given point in space, the basis function oscillates sinusoidally.

The wavenumber–frequency coefficients may be obtained using the transform [7]

$$F(\mathbf{k},\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\mathbf{z},t)e^{-j(\omega t-\mathbf{k}^T\mathbf{z})}d\mathbf{z}dt. \tag{10}$$

From (10), to determine the true wavenumber–frequency coefficients, the collection of the signal's values over an infinite spatiotemporal aperture is required. Conventional estimators obviously employ finite-length apertures; in addition, current spectral estimation methods focus on either the temporal or spatial aspect in isolation of the other. In the temporal case, a tapped delay line stores a set of previous samples. In the spatial case, an array of sensors samples the signal field in space.

From (8)–(10), it is evident that there is a relationship between the spatial and temporal frequency representations of a space–time signal. To that end, this paper proposes a spectral estimation framework which employs a spatiotemporal aperture. It is shown that the inclusion of both spatial and temporal dimensions in the aperture leads to benefits to both the resulting spatial and temporal spectral estimates.

IV. LCMV Spatial Spectral Estimation

The conventional delay-and-sum beamformer (DSB) is typically viewed as a signal enhancing spatial filter. We apply equalizing delays to the sensors to coherently sum the signal while attenuating the noise. Notice, however, that the DSB may also be viewed as a spatial spectral estimator: by steering the DSB to all candidate locations and determining the location which radiates the most energy, the source may be localized. Indeed, the latter forms the basis behind the SRP method. It is surprising that the steered beamformer approach to source localization has not been extended to more sophisticated beamforming structures.

The LCMV method of [17] is presented solely in the context of beamforming for signal enhancement: the location of the source is assumed to be known and the array is steered to the desired location. A temporal filter is then applied to each sensor and the filtered signals are summed to form the beamformer output. The nature of this temporal filtering is specified by the linear constraints of the LCMV scheme. Once the linear constraints are defined, a constrained optimization which minimizes the presence of noise in the beamformer output is carried out. The result is a cleaner estimate of the desired signal. In this section, we apply the LCMV technique to the source localization problem. It is shown that the temporal properties of the desired signal may be utilized to generate enhanced estimates of the spatial properties of the signal.

The proposed LCMV localization technique performs both spatial and temporal discrimination. The spatial processing consists of applying a steering delay to each sensor such that the propagation delays $\mathcal{F}_{1n}(\mathbf{r}_\mathrm{s})$ are equalized. This processing is done for each possible source location, leading to the parameterized output

$$x_{n,\mathrm{p}}(k,\mathbf{r}) = x_n[k+\mathcal{F}_{1n}(\mathbf{r})] \tag{11}$$

where $\mathbf{r}$ is the steered location (i.e., the parameter). When the steered location $\mathbf{r}$ matches the actual location $\mathbf{r}_\mathrm{s}$, the output becomes

$$x_{n,\mathrm{p}}(k,\mathbf{r}_\mathrm{s}) = \alpha_n(\mathbf{r}_\mathrm{s})s[k-\tau] + v_n[k+\mathcal{F}_{1n}(\mathbf{r}_\mathrm{s})] \tag{12}$$

where the noise component is potentially decorrelated by the steering delays. In vector notation, the received and time-aligned signals are written as

$$\mathbf{x}_\mathrm{p}(k,\mathbf{r}) = \mathbf{A}(\mathbf{r}_\mathrm{s})\mathbf{s}_\mathrm{p}(k-\tau,\mathbf{r}) + \mathbf{v}_\mathrm{p}(k,\mathbf{r}) \tag{13}$$

where the corresponding variables are defined at the bottom of the page, and $\mathrm{diag}(\bullet)$ is a diagonal matrix whose nonzero entries are indicated by the arguments.

To enable temporal processing, we add the previous $L-1$ samples of each sensor to form a spatiotemporal aperture

$$\overline{\mathbf{x}}_\mathrm{p}(k,\mathbf{r},L) = \overline{\mathbf{A}}(\mathbf{r}_\mathrm{s})\overline{\mathbf{s}}_\mathrm{p}(k-\tau,\mathbf{r},L) + \overline{\mathbf{v}}_\mathrm{p}(k,\mathbf{r},L) \tag{14}$$

where the corresponding variables are defined at the bottom of the next page, where $\mathbf{0}_{N\times N}$ is an N-by-N matrix of zeros and $\overline{\mathbf{A}}(\mathbf{r}_\mathrm{s})$ has size N-by-N. The LCMV technique is now applied in order to yield an array output whose temporal structure is constrained.

We form a multichannel finite-impulse response (FIR) filter

$$\mathbf{h}(\mathbf{r}) = [\mathbf{h}_{:1}^T(\mathbf{r}) \quad \mathbf{h}_{:2}^T(\mathbf{r}) \quad \cdots \quad \mathbf{h}_{:L-1}^T(\mathbf{r})]^T \tag{15}$$

where $\mathbf{h}_{:i}(\mathbf{r}) = [h_{1i}(\mathbf{r})\, h_{2,i}(\mathbf{r}) \cdots h_{Ni}(\mathbf{r})]^T$ is the subfilter applied to the spatial aperture corresponding to sample $k-\tau-i$. The multichannel filter is then applied to the spatiotemporal

$$\begin{aligned}
\mathbf{x}_\mathrm{p}(k,\mathbf{r}) &= [x_{1,\mathrm{p}}(k,\mathbf{r}) \quad x_{2,\mathrm{p}}(k,\mathbf{r}) \quad \cdots \quad x_{N,\mathrm{p}}(k,\mathbf{r})]^T\\
\mathbf{A}(\mathbf{r}_\mathrm{s}) &= \mathrm{diag}[\alpha_1(\mathbf{r}_\mathrm{s}),\ldots,\alpha_N(\mathbf{r}_\mathrm{s})]\\
\mathbf{s}_\mathrm{p}(k-\tau,\mathbf{r}) &= [s[k-\tau] \quad s[k-\tau-\mathcal{F}_{12}(\mathbf{r}_\mathrm{s})+\mathcal{F}_{12}(\mathbf{r})] \quad \cdots \quad s[k-\tau-\mathcal{F}_{1N}(\mathbf{r}_\mathrm{s})+\mathcal{F}_{1N}(\mathbf{r})]\,]^T\\
\mathbf{v}_\mathrm{p}(k,\mathbf{r}) &= [v_1(k) \quad v_2[k+\mathcal{F}_{12}(\mathbf{r})] \quad \cdots \quad v_N[k+\mathcal{F}_{1N}(\mathbf{r})]\,]^T
\end{aligned}$$

aperture such that a signal propagating from location $\mathbf{r}$ is coherently summed and then temporally filtered in some desired fashion

$$\mathbf{h}^T(\mathbf{r})\overline{\mathbf{A}}(\mathbf{r}_\mathrm{s})\overline{\mathbf{s}}_\mathrm{p}(k-\tau,\mathbf{r},L)=\sum_{l=0}^{L-1}f_l s(k-\tau-l) \tag{16}$$

where the coefficients $f_l, l=0,1,\ldots,L-1$ yield the desired temporal filtering. We perform a filtering for each candidate location, in that the multichannel FIR filter coefficients are a function of the steered location.

Assuming a source propagating from the steered location $\mathbf{r}$, the constraints follow from (16) as

$$\mathbf{c}_{\boldsymbol{\alpha},l}^T(\mathbf{r})\mathbf{h}(\mathbf{r})=f_l,\quad l=0,1,\ldots,L-1 \tag{17}$$

where

$$\mathbf{c}_{\boldsymbol{\alpha},l}(\mathbf{r})=\begin{bmatrix}\mathbf{0}_{N\times 1}^T & \cdots & \mathbf{0}_{N\times 1}^T & \underbrace{\boldsymbol{\alpha}^T(\mathbf{r})}_{l\text{th group}} & \mathbf{0}_{N\times 1}^T & \cdots & \mathbf{0}_{N\times 1}^T\end{bmatrix}^T$$

is a vector of length NL corresponding to the lth constraint, and

$$\boldsymbol{\alpha}(\mathbf{r})=[\,\alpha_1(\mathbf{r})\quad \alpha_2(\mathbf{r})\quad \ldots\quad \alpha_N(\mathbf{r})\,]^T.$$

The L constraints of (17) may be neatly expressed in matrix notation as

$$\mathbf{C}_{\boldsymbol{\alpha}}^T(\mathbf{r})\mathbf{h}(\mathbf{r})=\mathbf{f} \tag{18}$$

where

$$\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r})=[\,\mathbf{c}_{\boldsymbol{\alpha},0}(\mathbf{r})\quad \mathbf{c}_{\boldsymbol{\alpha},1}(\mathbf{r})\quad\cdots\quad \mathbf{c}_{\boldsymbol{\alpha},L-1}(\mathbf{r})\,] \tag{19}$$

is termed the *constraint matrix* and

$$\mathbf{f}=[\,f_0\quad f_1\quad\cdots\quad f_{L-1}]^T \tag{20}$$

is termed the *constraint vector*.

Note that there are NL multichannel filter coefficients, and L linear constraints; we thus say that there are $NL-L$ "degrees of freedom." After forming the desired constraints, these remaining degrees of freedom are utilized to minimize the average output power

$$E\{y^2(k,\mathbf{r})\}=E\{[\mathbf{h}^T(\mathbf{r})\overline{\mathbf{x}}_\mathrm{p}(k,\mathbf{r},L)]^2\} \tag{21}$$

where

$$y(k,\mathbf{r})=\mathbf{h}^T(\mathbf{r})\overline{\mathbf{x}}_\mathrm{p}(k,\mathbf{r},L) \tag{22}$$

which corresponds to minimizing the contribution of noise and interference to the spectral estimate. The minimization problem is thus, for every steered location $\mathbf{r}$

$$\begin{aligned}\hat{\mathbf{h}}(\mathbf{r})&=\arg\min_{\mathbf{h}}\mathbf{h}^T\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k,\mathbf{r},L)\mathbf{h}\\ &\text{subject to }\mathbf{C}_{\boldsymbol{\alpha}}^T(\mathbf{r})\mathbf{h}=\mathbf{f}\end{aligned} \tag{23}$$

where $\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(\mathbf{k},\mathbf{r},\mathbf{L})$ is the *parameterized spatiotemporal correlation matrix* as defined in (24) at the bottom of the page, where $\mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(\mathbf{k},\mathbf{r},\mathbf{l})$ is defined in (25) at the bottom of the next page, and where

$$R_{x_ix_j}(\tau)=E\{x_i(k)x_j(k+\tau)\} \tag{26}$$

is the cross-correlation function for two jointly wide-sense stationary and real random processes.

The solution to the constrained optimization problem is well-known; using the method of Lagrange multipliers

$$\begin{aligned}\hat{\mathbf{h}}(\mathbf{r})=&\,\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}^{-1}(k,\mathbf{r},L)\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r})\\ &\left[\mathbf{C}_{\boldsymbol{\alpha}}^T(\mathbf{r})\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}^{-1}(k,\mathbf{r},L)\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r})\right]^{-1}\mathbf{f}.\end{aligned} \tag{27}$$

$$\begin{aligned}\overline{\mathbf{x}}_\mathrm{p}(k,\mathbf{r},L)&=[\,\mathbf{x}_\mathrm{p}^T(k,\mathbf{r})\quad \mathbf{x}_\mathrm{p}^T(k-1,\mathbf{r})\quad\cdots\quad \mathbf{x}_\mathrm{p}^T(k-L+1,\mathbf{r})\,]^T\\ \overline{\mathbf{A}}(\mathbf{r}_\mathrm{s})&=\begin{bmatrix}\mathbf{A}(\mathbf{r}_\mathrm{s}) & \mathbf{0}_{N\times N} & \cdots & \mathbf{0}_{N\times N}\\ \mathbf{0}_{N\times N} & \mathbf{A}(\mathbf{r}_\mathrm{s}) & \cdots & \mathbf{0}_{N\times N}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} & \cdots & \mathbf{A}(\mathbf{r}_\mathrm{s})\end{bmatrix}\\ \overline{\mathbf{s}}_\mathrm{p}(k-\tau,\mathbf{r},L)&=[\,\mathbf{s}_\mathrm{p}^T(k-\tau,\mathbf{r})\quad \mathbf{s}_\mathrm{p}^T(k-\tau-1,\mathbf{r})\quad\cdots\quad \mathbf{s}_\mathrm{p}^T(k-\tau-L+1,\mathbf{r})\,]^T\\ \overline{\mathbf{v}}_\mathrm{p}(k,\mathbf{r},L)&=[\,\mathbf{v}_\mathrm{p}^T(k,\mathbf{r})\quad \mathbf{v}_\mathrm{p}^T(k-1,\mathbf{r})\quad\cdots\quad \mathbf{v}_\mathrm{p}^T(k-L+1,\mathbf{r})\,]^T\end{aligned}$$

$$\begin{aligned}\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k,\mathbf{r},L)&=E\{\overline{\mathbf{x}}_\mathrm{p}(k,\mathbf{r},L)\overline{\mathbf{x}}_\mathrm{p}^T(k,\mathbf{r},L)\}\\ &=\begin{bmatrix}\mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},0) & \mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},-1) & \cdots & \mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},-L+1)\\ \mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},1) & \mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},0) & \cdots & \mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},-L+2)\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},L-1) & \mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},L-2) & \cdots & \mathbf{R}_{\mathbf{x}_\mathrm{p}\mathbf{x}_\mathrm{p}}(k,\mathbf{r},0)\end{bmatrix}\end{aligned} \tag{24}$$

The source location estimate follows as

$$\begin{aligned}\hat{\mathbf{r}}_{\mathrm{s}} &= \arg\max_{\mathbf{r}} S_{\mathrm{LCMV}}(\mathbf{r}) \\ &= \arg\max_{\mathbf{r}} \hat{\mathbf{h}}^T(\mathbf{r})\mathbf{R}_{\overline{\mathbf{x}}_{\mathrm{p}}\overline{\mathbf{x}}_{\mathrm{p}}}(k,\mathbf{r},L)\hat{\mathbf{h}}(\mathbf{r}) \qquad (28)\end{aligned}$$

where $S_{\mathrm{LCMV}}(\mathbf{r}) = \hat{\mathbf{h}}^T(\mathbf{r})\mathbf{R}_{\overline{\mathbf{x}}_{\mathrm{p}}\overline{\mathbf{x}}_{\mathrm{p}}}(k,\mathbf{r},L)\hat{\mathbf{h}}(\mathbf{r})$ is the estimate of the spatial spectrum at the spatial frequency corresponding to location $\mathbf{r}$.

The proposed source localization scheme is outlined in Fig. 1.

A. Particular Case: Minimum Variance Distortionless Response

The MVDR spectral estimate, proposed for narrowband signals by Capon [3], and later for the broadband case by Krolik and Swingler [16], is a particular case of the LCMV technique which employs $L = 1$ and $\mathbf{f} = f_0 = 1$. In the latter case, the constraint is utilized to pass a plane wave with unity gain; in other words, the spectral estimate simply passes $s(k-\tau)$ through to the output. As the method processes a purely spatial aperture, the second-order statistics are conveyed by the parameterized spatial correlation matrix [14], as shown in (29) at the bottom of the page. The methods which stem from the parameterized spatial correlation matrix [14] do not exploit any known temporal properties of the desired signal—the signal discrimination is only in the spatial domain. It will later be shown that this aspect limits their localization performance.

B. Temporal Filtering: Autoregressive Model

The MVDR technique attempts to estimate the incoming wavefront by passing the current sample with unity gain. There are two drawbacks to this spectral estimation approach. First of all, since the signal aperture is limited to the N sensors (one sample per sensor), the noise reducing minimization procedure is limited in the degrees of freedom. Second, when localizing a colored signal such as speech, the known temporal properties are not exploited to improve the spatial spectral estimation.

To that end, the proposed LCMV method alleviates the two limitations noted above. In LCMV spectral estimation, the idea is to estimate the present sample as a linear combination of the past samples. This naturally calls for the modeling of the desired signal as an autoregressive (AR) process. The desired signal is modeled as

$$s(k) = \sum_{l=1}^{p} a_l s(k-l) + w(k) \qquad (30)$$

where a_l are the predictive coefficients, p is the order of the AR model, and $w(k)$ is the white noise that drives the AR process. Note that $w(k)$ may also be interpreted as the prediction error.

Consider (16)—the goal of the constraint is to estimate $s(k-\tau)$ using a linear combination of $\{s(k-\tau), s(k-\tau-1), \ldots, s(k-\tau-L+1)\}$

$$\begin{aligned}\hat{s}(k-\tau) &= \mathbf{h}^T(\mathbf{r})\overline{\mathbf{A}}(\mathbf{r}_{\mathrm{s}})\overline{\mathbf{s}}_{\mathrm{p}}(k-\tau,\mathbf{r},L) \\ &= \sum_{l=0}^{L-1} f_l s(k-\tau-l) \qquad (31)\end{aligned}$$

where $\hat{s}(k-\tau)$ denotes the estimate of the desired present sample. The MVDR method chooses $f_0 = 1, f_l = 0, l = 1, \ldots, L-1$, yielding an errorless estimate but also meaning that temporal dependence is neglected.

In the proposed LCMV method, the desired signal's temporal properties are taken into account via AR modeling. The AR parameters of the desired signal are encoded in the constraint vector $\mathbf{f}$ which in turn shapes the multichannel filter $\mathbf{h}(\mathbf{r})$. Connecting (30) to (31), the LCMV method chooses

$$f_0 = 0 \qquad (32)$$

$$f_l = a_l, \quad l = 2, \ldots, p \qquad (33)$$

resulting in an estimation error given by

$$s(k-\tau) - \hat{s}(k-\tau) = w(k). \qquad (34)$$

By modeling the present sample as a weighted combination of the previous samples, a zero-mean estimation error is incurred. However, it is expected that the extra degrees of freedom in the multichannel filter $\mathbf{h}(\mathbf{r})$ will lead to a more accurate spectral estimate. Moreover, the filter $\mathbf{h}(\mathbf{r})$ *temporally* focuses the steered

$$\mathbf{R}_{\mathbf{x}_{\mathrm{p}}\mathbf{x}_{\mathrm{p}}}(k,\mathbf{r},l) = \begin{bmatrix} R_{x_1x_1}(l) & R_{x_1x_2}[l+\mathcal{F}_{12}(\mathbf{r})] & \cdots & R_{x_1x_N}[l+\mathcal{F}_{1N}(\mathbf{r})] \\ R_{x_2x_1}[l-\mathcal{F}_{12}(\mathbf{r})] & R_{x_2x_2}(l) & \cdots & R_{x_2x_N}[l+\mathcal{F}_{2N}(\mathbf{r})] \\ \vdots & \vdots & \ddots & \vdots \\ R_{x_Nx_1}[l-\mathcal{F}_{1N}(\mathbf{r})] & R_{x_Nx_2}[l-\mathcal{F}_{2N}(\mathbf{r})] & \cdots & R_{x_Nx_N}(l) \end{bmatrix} \qquad (25)$$

$$\begin{aligned}\mathbf{R}_{\overline{\mathbf{x}}_{\mathrm{p}}\overline{\mathbf{x}}_{\mathrm{p}}}(k,\mathbf{r},1) &= E\left\{\overline{\mathbf{x}}_{\mathrm{p}}(k,\mathbf{r},1)\overline{\mathbf{x}}_{\mathrm{p}}^T(k,\mathbf{r},1)\right\} \\ &= \mathbf{R}_{\mathbf{x}_{\mathrm{p}}\mathbf{x}_{\mathrm{p}}}(k,\mathbf{r},0) \\ &= \begin{bmatrix} R_{x_1x_1}(0) & R_{x_1x_2}[\mathcal{F}_{12}(\mathbf{r})] & \cdots & R_{x_1x_N}[\mathcal{F}_{1N}(\mathbf{r})] \\ R_{x_2x_1}[-\mathcal{F}_{12}(\mathbf{r})] & R_{x_2x_2}(0) & \cdots & R_{x_2x_N}[\mathcal{F}_{2N}(\mathbf{r})] \\ \vdots & \vdots & \ddots & \vdots \\ R_{x_Nx_1}[-\mathcal{F}_{1N}(\mathbf{r})] & R_{x_Nx_2}[-\mathcal{F}_{2N}(\mathbf{r})] & \cdots & R_{x_Nx_N}(0) \end{bmatrix} \qquad (29)\end{aligned}$$

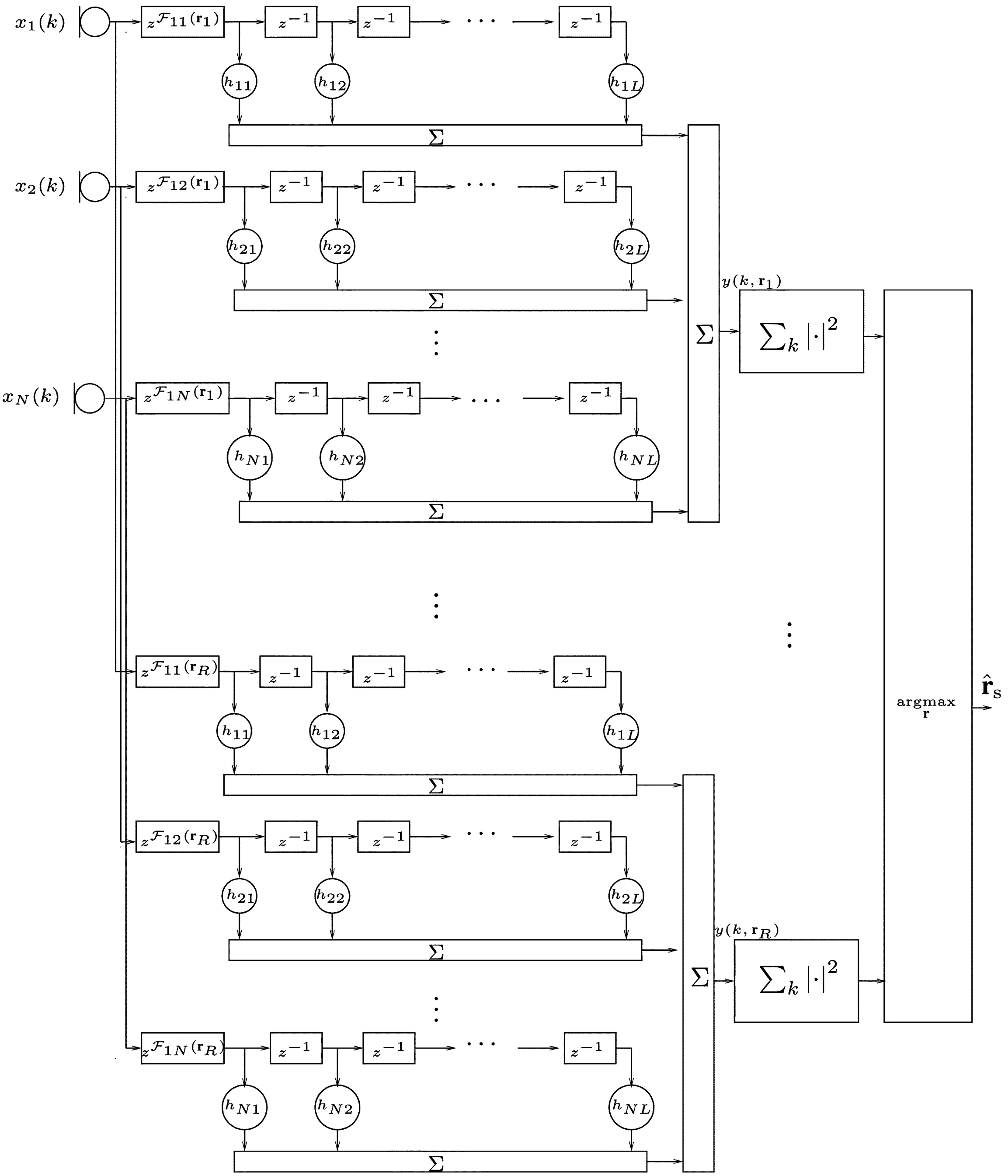

Fig. 1. Proposed source localization scheme where it is assumed that there are R candidate locations. The FIR filters temporally focus the steered beams onto the desired signal.

beam to pick up a signal with the temporal structure contained in $\mathbf{f}$. Any noise or interfering signals with a different temporal structure should be attenuated by this temporally focused filter.

Note that in practice, the AR parameters need to be estimated from the observed signals using either a classical single-channel method such as solving of the Yule–Walker equations [18], or a multichannel method that somehow incorporates the data from all sensors [19].

C. Modeling the Attenuation Constants

By characterizing the energy distribution of the wavefront across the spatial aperture (especially for signals in the near-field of the array), the resulting LCMV constraints yield a finer estimate of the present sample.

The parameterized attenuation model accounts for the location dependence of the vector of attenuation constants. Each candidate location is characterized by the attenuation constants that are experienced from a spherical wavefront originating from that location. The drawback of this scheme is that the location space is now necessarily multidimensional (because the attenuation constants are clearly a function of the range).

For each candidate location $\mathbf{r}$, the hypothetical attenuation vector is given by

$$\boldsymbol{\alpha}(\mathbf{r}) = \begin{bmatrix} \frac{d_{n_\mathrm{o},\mathrm{s}}(\mathbf{r})}{d_{1,\mathrm{s}}(\mathbf{r})} & \frac{d_{n_\mathrm{o},\mathrm{s}}(\mathbf{r})}{d_{2,\mathrm{s}}(\mathbf{r})} & \cdots & \frac{d_{n_\mathrm{o},\mathrm{s}}(\mathbf{r})}{d_{N,\mathrm{s}}(\mathbf{r})} \end{bmatrix}^T \quad (35)$$

where $d_{n_{\mathrm{o,s}}}(\mathbf{r})$ is the distance from location $\mathbf{r}$ to the closest sensor—n_o denotes the index of the sensor nearest location $\mathbf{r}$.

In some cases, a multidimensional search is undesirable; moreover, when operating in the far-field of the array, the attenuation levels experienced by the desired signal at the various sensors do not differ greatly. In that case, fixed values for the attenuation constants may be assumed

$$\alpha_n(\mathbf{r}_\mathrm{s}) \approx 1, \quad n = 1, 2, \ldots, N. \tag{36}$$

As a result, the constraint matrix is easily constructed and is independent of the steered location.

D. Complexity and the Generalized Sidelobe Canceller

The drawback of incorporating a spatiotemporal aperture is the resulting algorithmic complexity: the LCMV method requires the inversion of the NL-by-NL matrices $\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L)$ and $\mathbf{C}_{\boldsymbol{\alpha}}^T(\mathbf{r})\mathbf{R}^{-1}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L)\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r})$—this is computationally burdensome. To that end, it is well known that the LCMV optimization problem may be alternatively implemented by the generalized sidelobe canceller (GSC); that is, the LCMV and GSC filters are strictly equivalent [20]. The GSC decomposes the LCMV solution into two orthogonal components

$$\hat{\mathbf{h}}(\mathbf{r}) = \hat{\mathbf{g}}(\mathbf{r}) - \mathbf{B}(\mathbf{r})\hat{\mathbf{w}}(\mathbf{r}) \tag{37}$$

where

$$\hat{\mathbf{g}}(\mathbf{r}) = \arg\min_{\mathbf{g}} \left\| \mathbf{C}_{\boldsymbol{\alpha}}^T(\mathbf{r})\mathbf{g} - \mathbf{f} \right\|^2 \tag{38}$$

which leads to the minimum-norm solution

$$\hat{\mathbf{g}}(\mathbf{r}) = \mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r}) \left[\mathbf{C}_{\boldsymbol{\alpha}}^T(\mathbf{r})\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r}) \right]^{-1} \mathbf{f}. \tag{39}$$

Additionally, $\mathbf{B}(\mathbf{r})$ is termed the *blocking matrix* as it forms a null in the direction of the desired signal. The blocking matrix is of size NL-by-$(N-1)L$ as its columns span the null space of the constraint matrix

$$\mathbf{C}_{\boldsymbol{\alpha}}^T(\mathbf{r})\mathbf{B}(\mathbf{r}) = \mathbf{0}_{L \times (N-1)L} \tag{40}$$

where $\mathbf{0}_{L \times (N-1)L}$ is an L-by-$(N-1)L$ matrix of zeros.

The value of $\hat{\mathbf{w}}(\mathbf{r})$ follows as the solution to the unconstrained optimization problem given by

$$\begin{aligned}\hat{\mathbf{w}}(\mathbf{r}) = \arg\min_{\mathbf{w}} &[\hat{\mathbf{g}}(\mathbf{r}) - \mathbf{B}(\mathbf{r})\mathbf{w}]^T \\ &\times \mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L)[\hat{\mathbf{g}}(\mathbf{r}) - \mathbf{B}(\mathbf{r})\mathbf{w}].\end{aligned} \tag{41}$$

The solution to (41) is given by

$$\hat{\mathbf{w}}(\mathbf{r}) = [\mathbf{B}^T(\mathbf{r})\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L)\mathbf{B}(\mathbf{r})]^{-1}\mathbf{B}^T(\mathbf{r})\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L)\mathbf{f}. \tag{42}$$

The matrix to be inverted in (42) has size $(N-1)L$-by-$(N-1)L$, a reduction of L in row and column from the original LCMV solution.

E. Matrix Regularization

The inversion of the NL-by-NL block matrix $\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L)$ may lead to numerical stability issues in practice. Even in the GSC implementation, the inversion of $\mathbf{B}^T(\mathbf{r})\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L)\mathbf{B}(\mathbf{r})$ may pose problems. To alleviate this, some form of matrix regularization is required. In the simulation study that follows, the Tikhonov regularization method is employed [21]. The inversion of the parameterized spatiotemporal correlation matrix is performed as

$$\mathbf{R}^{-1}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L) \leftarrow [\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L) + \delta\mathbf{I}_{NLh \times NL}]^{-1} \tag{43}$$

where $\leftarrow$ denotes assignment, $\mathbf{I}_{NLh \times NL}$ is the NL-by-NL identity matrix, and δ is the regularization parameter, which is taken in the simulations as

$$\delta = \frac{1}{NL}\mathrm{trace}[\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k, \mathbf{r}, L)]\Delta \tag{44}$$

where Δ is the normalized regularization constant, with typical values being $\Delta = 0.1$, $\Delta = 1$, and $\Delta = 10$. The selection of Δ is a tradeoff between the alleviation of the ill-condition of the desired matrix and the resulting accuracy of the inversion calculation.

V. LCMV Temporal Spectral Estimation

Consider a sinusoidal signal of the form

$$s(k) = e^{j\Omega k} \tag{45}$$

where $\Omega = (\omega/f_s)$ is the digital frequency with f_s denoting the sampling rate. The output of sensor n with such a propagating signal is then given by

$$x_n(k) = \alpha_n(\mathbf{r}_\mathrm{s})e^{j\Omega[k - \tau - \mathcal{F}_{1n}(\mathbf{r}_\mathrm{s})]} + v_n(k). \tag{46}$$

The idea behind LCMV temporal spectral estimation is to estimate the energy of the desired signal at a particular temporal frequency while minimizing the contribution of the noise and interference to the resulting estimate. Conventional temporal spectral estimation involves passing the signal through a bank of filters—one for each temporal frequency. If the source location is known, the conventional temporal aperture may be extended to include multiple sensors, thus increasing the available degrees of freedom in the applied filters. It is important to understand that the location space is no longer being scanned: the array is parameterized with respect to the assumed location, and the scanning is performed across the temporal frequency range. The multichannel filters that are applied to the spatiotemporal aperture are parameterized by the temporal frequency. By "scanning the temporal frequency range," we mean to say that we have a bank of multichannel filters: one multichannel filter for each temporal frequency; we pass the incoming space–time signal through each filter, and observe the output power. The power of the output of a given multichannel filter corresponds to the strength of the desired signal at the temporal frequency to which the multichannel filter is tuned.

In order to tune each multichannel filter in the filterbank to a desired temporal frequency, we constrain the response of the multichannel filter to have a unity gain response to a sinusoid at the desired frequency: thus, the constraint vector $\mathbf{f}$ varies with the temporal frequency. The taps of each multichannel filter in the filterbank are computed by solving a constrained optimization problem where the temporal frequency of interest is embedded into the constraint vector $\mathbf{f}$.

For the time-aligned case of $\mathbf{r}=\mathbf{r}_\mathrm{s}$, the parameterized output is then written as

$$x_{n,\mathrm{p}}(k,\mathbf{r}_\mathrm{s}) = \alpha_n(\mathbf{r}_\mathrm{s})e^{j\Omega(k-\tau)} + v_n[k+\mathcal{F}_{1n}(\mathbf{r}_\mathrm{s})]. \tag{47}$$

The signal portion of the spatiotemporal aperture follows as

$$\begin{aligned}&\overline{\mathbf{A}}(\mathbf{r}_\mathrm{s})\overline{\mathbf{s}}_\mathrm{p}(k-\tau,\mathbf{r}_\mathrm{s},L)\\ &\quad= \Big[\alpha_1(\mathbf{r}_\mathrm{s})e^{j\Omega(k-\tau)} \quad \cdots \quad \alpha_N(\mathbf{r}_\mathrm{s})e^{j\Omega(k-\tau)} \quad \cdots \\ &\qquad \alpha_1(\mathbf{r}_\mathrm{s})e^{j\Omega(k-\tau-L+1)} \quad \cdots \quad \alpha_N(\mathbf{r}_\mathrm{s})e^{j\Omega(k-\tau-L+1)}\Big]^T.\end{aligned} \tag{48}$$

The goal of the spectral estimation in this case is to pass the sinusoidal signal through to the output

$$\mathbf{h}^H(\Omega)\overline{\mathbf{A}}(\mathbf{r}_\mathrm{s})\overline{\mathbf{s}}_\mathrm{p}(k-\tau,\mathbf{r}_\mathrm{s},L) = e^{j\Omega(k-\tau)} \tag{49}$$

where ^{H} denotes Hermitian transposition as all quantities in the temporal spectral estimation case are complex. The desired constraint of (49) is accomplished by

$$\mathbf{h}_{:i}^H(\Omega)\boldsymbol{\alpha}(\mathbf{r}_\mathrm{s}) = \frac{1}{L}e^{j\Omega i}, \quad i=0,1,\ldots,L-1. \tag{50}$$

Thus, the constraint vector is now a function of the temporal frequency Ω

$$\mathbf{f}(\Omega) = \frac{1}{L}\left[1 \quad e^{j\Omega 1} \quad \cdots \quad e^{j\Omega(L-1)}\right]^T. \tag{51}$$

The constraint matrix $\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r}_\mathrm{s})$ remains exactly in the form of (19), where it should be noted that the argument $\mathbf{r}_\mathrm{s}$ is fixed because we are not scanning the location space.

The optimization problem is then to choose, for each temporal frequency, a filter that minimizes the output power while passing a sinusoidal signal at that frequency with unity gain

$$\begin{aligned}&\hat{\mathbf{h}}(\Omega) = \arg\min_{\mathbf{h}} \mathbf{h}^T\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k,\mathbf{r}_\mathrm{s},L)\mathbf{h}\\ &\text{subject to } \mathbf{C}_{\boldsymbol{\alpha}}^T(\mathbf{r}_\mathrm{s})\mathbf{h} = \mathbf{f}(\Omega).\end{aligned} \tag{52}$$

The solution to (52) then follows as

$$\begin{aligned}\hat{\mathbf{h}}(\Omega) &= \mathbf{R}^{-1}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k,\mathbf{r}_\mathrm{s},L)\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r}_\mathrm{s})\\ &\quad\times\left[\mathbf{C}_{\boldsymbol{\alpha}}^H(\mathbf{r}_\mathrm{s})\mathbf{R}^{-1}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k,\mathbf{r}_\mathrm{s},L)\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r}_\mathrm{s})\right]^{-1}\mathbf{f}(\Omega).\end{aligned} \tag{53}$$

The estimate of the temporal spectrum at frequency Ω is then given by

$$S_{\mathrm{LCMV}}(\Omega) = \hat{\mathbf{h}}^H(\Omega)\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k,\mathbf{r}_\mathrm{s},L)\hat{\mathbf{h}}(\Omega). \tag{54}$$

Notice that with $N=1$ (a single sensor), the LCMV temporal spectral estimate of (53) and (54) simplifies to the classical MVDR estimate, as shown in equations (55) and (56) at the bottom of the page. Fig. 2 illustrates the proposed temporal spectral estimation scheme.

VI. Simulation Evaluation

A. Simulation Model: Spatial Spectral Estimation

The LCMV spatial spectral estimator is evaluated in a computer simulation using the image method model of [22]. A six-microphone uniform circular array with a 4.25-cm radius is simulated. The array radius is chosen to fulfill the spatial Nyquist sampling criterion with the maximum frequency of interest being 4 kHz. The simulated room is rectangular with plane reflective boundaries (walls, ceiling and floor). The reflection coefficients of the boundaries are independent of frequency. The room dimensions in centimeters are (304.8, 457.2, 381). The center of the array sits at (152.4, 228.6, 101.6). The speaker is located at (152.4, 406.4, 101.6). The reverberation times are measured using the method of [23]: the reverberation times range from $T_{60}=300$ ms to $T_{60}=900$ ms, where T_{60} is the time for the impulse response's energy to decay by 60 dB. The source signal is convolved using the synthetic impulse responses. Appropriately scaled temporally and spatially white Gaussian noise is then added at the microphones to achieve the required SNR: SNRs of 10, 20, and 30 dB are simulated. Two signal types are examined: a stationary AR process generated with Gaussian noise ($a_0=1, a_1=-0.5, a_2=0.4, a_3=-0.2, a_4=0.1, a_l=0$ for $l\geq 5$) and female English speech. The sampling rate is 48 kHz. Due to the planar array geometry and far-field source, the location space is limited to the set of azimuth angles in the range 0–360°, with a resolution of 1°. Note that the LCMV algorithm is general in that the parametrization may be one-, two-, or three-dimensional, depending on the desired propagation model. In this paper, the one-dimensional parametrization (i.e., azimuth only) was chosen for simplicity. The DOA estimates are computed once per 64-ms frame over a one-minute signal. The algorithms are evaluated in terms of the percentage of anomalous estimates—those that vary from

$$\begin{aligned}\mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k,\mathbf{r},L)|_{N=1} &= \mathbf{R}_{\overline{\mathbf{x}}_\mathrm{p}\overline{\mathbf{x}}_\mathrm{p}}(k,L)\\ &= \begin{bmatrix} R_{x_1x_1}(0) & R_{x_1x_1}(-1) & \cdots & R_{x_1x_1}(-L+1)\\ R_{x_1x_1}(1) & R_{x_1x_1}(0) & \cdots & R_{x_1x_1}(-L+2)\\ \vdots & \vdots & \ddots & \vdots\\ R_{x_1x_1}(L-1) & R_{x_1x_1}(L-2) & \cdots & R_{x_1x_1}(0)\end{bmatrix}\end{aligned} \tag{55}$$

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\alpha}}(\mathbf{r}_\mathrm{s})|_{N=1} &= \mathbf{C}_{\alpha_1}\\ &= \alpha_1\mathbf{I}_{L\times L}.\end{aligned} \tag{56}$$

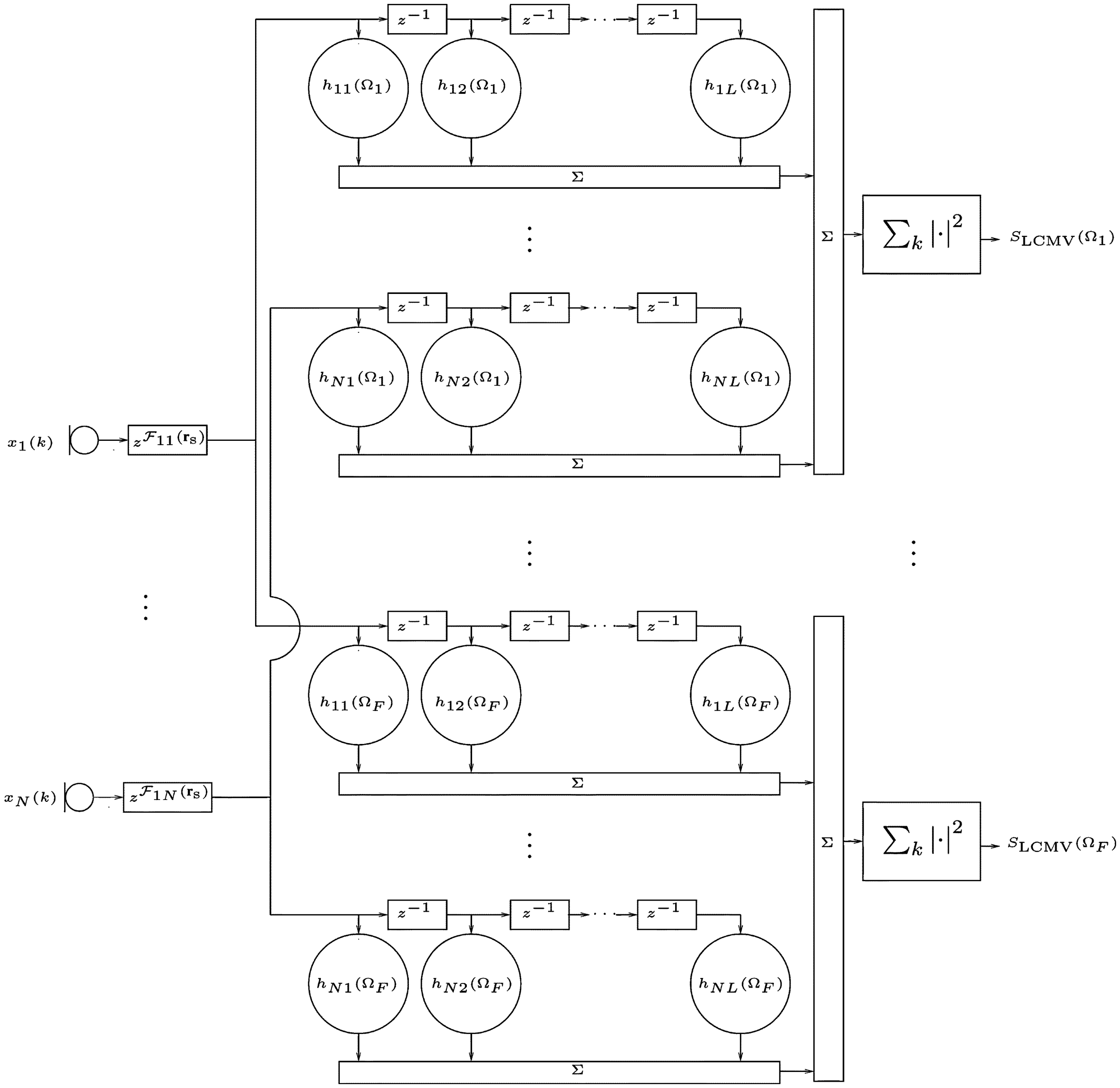

Fig. 2. Proposed temporal spectral estimation scheme, where we have assumed that there are F temporal frequencies. The sensor array is first steered to the source location $\mathbf{r}_{\mathrm{s}}$. We then feed the outputs of the steered array to a filterbank of F multichannel filters, where the taps of each multichannel filter are dependent on the desired temporal frequency.

the true azimuth by more than 5°, and by the root-mean-square (rms) error for the nonanomalous estimates

$$\mathrm{e}_{\mathrm{rms}} = \sqrt{\frac{1}{L_{\mathrm{na}}} \sum_{l \in \chi_{\mathrm{na}}} (\hat{\theta}_l - \theta_l)^2} \tag{57}$$

where χ_{na} is the set of all nonanomalous estimates, L_{na} is the number of elements in χ_{na}, and $\hat{\theta}_l$ and θ_l are the estimated and actual azimuth angles of the source for frame l.

For comparison, the proposed estimators are compared to the SRP and MVDR methods. In the case of speech signals, the generalized cross-correlation (GCC) phase transform (PHAT) method is employed to whiten the observed cross-correlations. This is applied to all three algorithms: SRP, MVDR, and LCMV—all three are compatible with the GCC family of methods. Some literatures make a distinction between "SRP" and "SRP-PHAT"—the only difference between the two is that the cross-correlations are whitened in the latter. Thus, in this paper, the terms are used interchangeably.

To estimate the AR coefficients of the desired signal, the Yule–Walker or "autocorrelation" method is employed using data collected from the first sensor. Two sets of simulations are run: one modeling the parameterized attenuation signal model, and the other utilizing the fixed attenuation model. As previously mentioned, the parameterized attenuation model requires a multidimensional search involving the range dimension; in the forthcoming simulations, the location space consists of the set $\{(r_{\mathrm{s}}, \theta) | 0 \leq \theta < 360\}$. In practice, the range dimension needs to be scanned as well.

The parameterized spatiotemporal correlation matrix is regularized using the Tikhonov method with $\Delta = 0.1$ for the simulations involving a speech source signal, and $\Delta = 10$ for the simulations with a stationary AR source. The SRP method does not involve a matrix inversion and thus it does not need regularization. Even though the MVDR method does invert the parameterized spatial correlation matrix, this matrix is substantially smaller (N-by-N), and thus regularization is needed only

TABLE I
SOURCE LOCALIZATION PERFORMANCE OF CONVENTIONAL AND PROPOSED ESTIMATORS WITH A STATIONARY AR PROCESS; FIXED ATTENUATION CONSTANTS

		SRP		MVDR		LCMV	
SNR (dB)	T_{60} (ms)	$\%_{\text{anomalies}}$ (%)	rms (degrees)	$\%_{\text{anomalies}}$ (%)	rms (degrees)	$\%_{\text{anomalies}}$ (%)	rms (degrees)
10	300	0	0	0	0	0*	0*
	600	1.39	0	1.28	0	0.11*	0*
	900	39.70	0	39.70	0	20.28*	0.073*
20	300	0	0	0	0	0*	0*
	600	1.39	0	1.39	0	0.11*	0.09*
	900	39.81	0	41.09	0	19.53*	0*
30	300	0	0	0	0	0*	0*
	600	1.71	0.066	1.60	0.066	0.21 *	0*
	900	40.34	0	40.66	0	19.42*	0.073*

*regularized with $\Delta = 10$

TABLE II
SOURCE LOCALIZATION PERFORMANCE OF CONVENTIONAL AND PROPOSED ESTIMATORS WITH A SPEECH SIGNAL; FIXED ATTENUATION CONSTANTS

		SRP		MVDR		LCMV	
SNR (dB)	T_{60} (ms)	$\%_{\text{anomalies}}$ (%)	rms (degrees)	$\%_{\text{anomalies}}$ (%)	rms (degrees)	$\%_{\text{anomalies}}$ (%)	rms (degrees)
10	300	43.12	1.82	42.69	1.82	19.96*	0.85*
	600	48.45	1.78	48.67	1.78	31.91*	1.01*
	900	53.36	1.81	53.68	1.80	40.88*	1.10*
20	300	18.78	1.48	18.57	1.49	11.21*	0.85*
	600	27.32	1.45	27.21	1.44	19.74*	1.02*
	900	37.46	1.39	37.46	1.38	29.24*	1.10*
30	300	10.14	1.03	9.82	1.03	10.46*	0.81*
	600	21.34	0.99	21.34	0.99	21.99	0.97*
	900	30.31	1.03	30.10	1.04	30.42*	1.10*

*regularized with $\Delta = 0.1$

TABLE III
SOURCE LOCALIZATION PERFORMANCE OF CONVENTIONAL AND PROPOSED ESTIMATORS WITH A STATIONARY AR PROCESS; PARAMETERIZED ATTENUATION CONSTANTS

		SRP		MVDR		LCMV	
SNR (dB)	T_{60} (ms)	$\%_{\text{anomalies}}$ (%)	rms (degrees)	$\%_{\text{anomalies}}$ (%)	rms (degrees)	$\%_{\text{anomalies}}$ (%)	rms (degrees)
10	300	0	0	0	0.22	0*	0*
	600	1.39	0	1.07	0	0.11*	0*
	900	39.70	0	39.59	0	20.38*	0.073*
20	300	0	0	0*	0*	0*	0*
	600	1.39	0	1.39*	0*	0.11*	0.065*
	900	39.81	0	41.09*	0*	19.96*	0*
30	300	0	0	0*	0*	0*	0*
	600	1.71	0.066	1.60*	0.066*	0.21*	0*
	900	40.34	0	40.66*	0*	19.74*	0.073*

*regularized with $\Delta = 10$

TABLE IV
SOURCE LOCALIZATION PERFORMANCE OF CONVENTIONAL AND PROPOSED ESTIMATORS WITH A SPEECH SIGNAL; PARAMETERIZED ATTENUATION CONSTANTS

		SRP		MVDR		LCMV	
SNR (dB)	T_{60} (ms)	$\%_{\text{anomalies}}$ (%)	rms (degrees)	$\%_{\text{anomalies}}$ (%)	rms (degrees)	$\%_{\text{anomalies}}$ (%)	rms (degrees)
10	300	43.12	1.82	42.58	1.82	19.21*	0.85*
	600	48.45	1.78	48.77	1.77	31.59*	1.02*
	900	53.36	1.81	53.90	1.80	41.09*	1.09*
20	300	18.78	1.48	18.78	1.49	11.21*	0.83*
	600	27.32	1.45	27.53	1.44	19.64*	1.03*
	900	37.46	1.39	37.67	1.38	29.78*	1.11*
30	300	10.14	1.03	9.82	1.03	10.35*	0.80*
	600	21.34	0.99	21.13	0.99	22.31*	0.97*
	900	30.31	1.03	30.42	1.03	30.84*	1.10*

*regularized with $\Delta = 0.1$

in some of the high SNR cases. The LCMV method requires regularization in all cases.

Notice that the purpose of the simulation study is to evaluate the proposed estimator through a comparison to the present-day methods. Factors such as moving sources, multiple sources, or uncalibrated microphones are not considered as these are separate issues which deserve their own treatment.

B. Simulation Model: Temporal Spectral Estimation

To evaluate the LCMV temporal spectral estimator, the image method model is again employed to simulate the propagation of the desired temporal signal to the array. Two reverberation levels are considered: $T_{60} = 0$ ms and $T_{60} = 300$ ms. Uniform circular arrays of $N = 3$, $N = 6$, and $N = 9$ sensors are simulated. For each array, the radius corresponds to the maximum value that does not lead to spatial aliasing (assuming that the maximum frequency of interest is again 4 kHz). This corresponds to arrays of radius 2.45, 4.25, and 6.21 cm, respectively. The SNR at the array varies from -10 to 0 dB. The desired signal is a single digital tone at $\Omega = 2\pi \cdot 0.15$. The tone is buried in the Gaussian noise. The sampling rate is again 48 kHz. The length of the temporal portion of the aperture is $L = 100$.

To evaluate the effect of employing a spatiotemporal aperture on the resulting temporal spectral estimate, the LCMV spectral estimate $S_{\mathrm{LCMV}}(\Omega)$ is compared to that of the MVDR estimate $S_{\mathrm{MVDR}}(\Omega)$, which as already mentioned, is a particular case of the LCMV method. The MVDR method employs only the first sensor from the array. Tikhonov regularization with $\Delta = 100$ is applied to the parameterized spatiotemporal correlation matrix prior to inversion for both MVDR and LCMV methods.

C. Results

Tables I and II present the source localization results for an AR process and a speech signal, respectively, for the fixed attenuation model. Tables III and IV present the corresponding results for the parameterized attenuation model. The results indicate the presence of regularization, where appropriate.

Consider first the simulations with a stationary AR process. Several factors make these simulations less challenging than the ones employing a speech signal. The estimation of the AR parameters is less problematic as the true AR values do not change over the duration of a frame. Second, the estimation error is expected to be lower since the signal inherently follows the AR model. Lastly, the negative effect of reverberation on the resulting cross-correlation measurements is limited since the selected AR parameters lead to a less colored signal than speech. It takes a reverberation time of $T_{60} = 900$ ms for the prevalence of anomalies to become significant. At this value and a 10-dB SNR, the LCMV method achieves approximately 20% less anomalies than the leading conventional method for both the fixed and parameterized attenuation models. It becomes apparent that the increased number of degrees of freedom in the LCMV spatial spectral estimate leads to a greater robustness to the effects of noise and reverberation. By explicitly modeling the temporal nature of the desired signal, the LCMV filters are better able to estimate the signal power emanating from all directions than the SRP and MVDR methods. Notice that the MVDR technique does not significantly outperform SRP in any case—this is because with only one temporal sample per sensor in the aperture, the minimization of noise and reverberation in the constrained optimization problem is severely limited.

Turning now to the simulations involving speech signals, it is evident that the colored nature of speech generally leads to increased levels of anomalies for all methods. For the lower SNR cases (i.e., 10 and 20 dB), the LCMV method significantly outperforms the SRP and MVDR techniques, with the improvement reaching 24% in the SNR $= 10$ dB, $T_{60} = 300$ ms, and parameterized attenuation case. The rms error is also reduced by approximately 1°. As the level of reverberation is increased, the performance improvement provided by employing the LCMV method's spatiotemporal aperture is somewhat reduced. This seems to suggest that the additional degrees of freedom in the multichannel LCMV filter are better suited at combating the effects of uncorrelated rather than colored noise. Viewed in another manner, notice that since the reverberant components are coherent with the desired signal, the AR coefficients of any reverberant component are expected to be quite similar to that of the desired signal. Thus, from this standpoint, there is less discrimination between the signal and additive noise. Nevertheless, for the SNR $= 10$ dB and SNR $= 20$ dB cases, the LCMV method still produces substantially less anomalies than the SRP and MVDR methods.

For the speech simulations with an SNR of 30 dB, the LCMV method yields accuracies comparable to the less complex SRP and MVDR techniques. This may be attributed to one or more factors: the estimation error associated with modeling speech as an AR process results in a corresponding error in the LCMV constraints. Furthermore, the LCMV method necessitates the regularization of the parameterized spatiotemporal correlation matrix—as previously mentioned, this introduces an error in the obtained matrix inverse, thus biasing the spectral estimate. With a high SNR, the regularization becomes even more critical as uncorrelated sensor noise naturally regularizes the parameterized spatiotemporal correlation matrix. Lastly, with such a high SNR, there is less noise for the proposed LCMV method to combat.

Nevertheless, the results of spatial spectral estimation point to a tremendous advantage of the LCMV method: its ability to combat the effects of noise and reverberation using a spatiotemporal aperture which performs both spatial and temporal signal discrimination. The tradeoff here is between localization accuracy and algorithm complexity. The LCMV method processes an aperture L times longer than the MVDR and SRP methods. The resulting correlation matrix has $N^2(L^2-1)$ more elements. Moreover, for a dynamic signal such as speech, the AR coefficients must be recomputed with every frame. Thus, the LCMV method is more computationally burdensome.

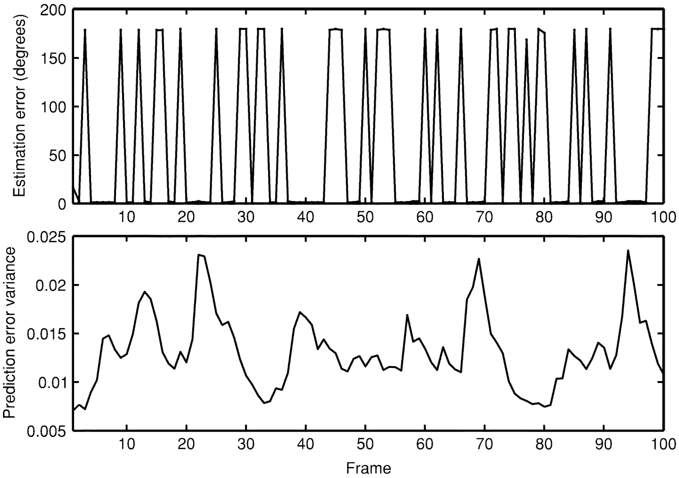

Fig. 3. Prediction error versus localization accuracy over 100 frames.

For certain frames of speech such as unvoiced sounds, the signal is not quite predictable and the AR model may not be a good fit. It is interesting to investigate the relationship between the linear prediction error (i.e., the validity of the AR model) and the ensuing localization accuracy. To that end, Fig. 3 plots the localization accuracy in parallel with the variance of the prediction error $w(k)$ for 100 frames of speech data ($T_{60} = 900$ ms, SNR $= 10$ dB). It is evident that there is little correlation between the linear prediction error and the resulting location estimate accuracy. In fact, the square of the correlation coefficient between the variance of the prediction error and the estimation error is equal to $\rho^2 = 0.0525$. It appears that even for unvoiced sounds, there is enough temporal redundancy in the signal for the algorithm to exploit via the linear constraints.

Turning now to the results of the temporal spectral estimation simulations, Fig. 4 shows the various spectra produced by the conventional MVDR and proposed LCMV methods. It is clear that the inclusion of additional sensors reduces the level of noise and reverberation present in the spectral estimates for all combinations of parameters. Moreover, as the number of sensors is increased, the noise level is decreased. By taking into account the spatial properties of the desired signal, the LCMV method's multichannel filter bank is able to provide a greater discrimination between the digital tone and the corrupting noise and reverberation.

VII. Conclusion

This paper has presented a novel spectral estimation technique based on the LCMV beamformer proposed by Frost; the proposed method utilizes a spatiotemporal aperture. It was shown that by accounting for the temporal properties of the desired signal in the linear constraints via AR modeling, source localization performance is dramatically improved. Moreover, by employing multiple sensors, the signal's temporal spectral

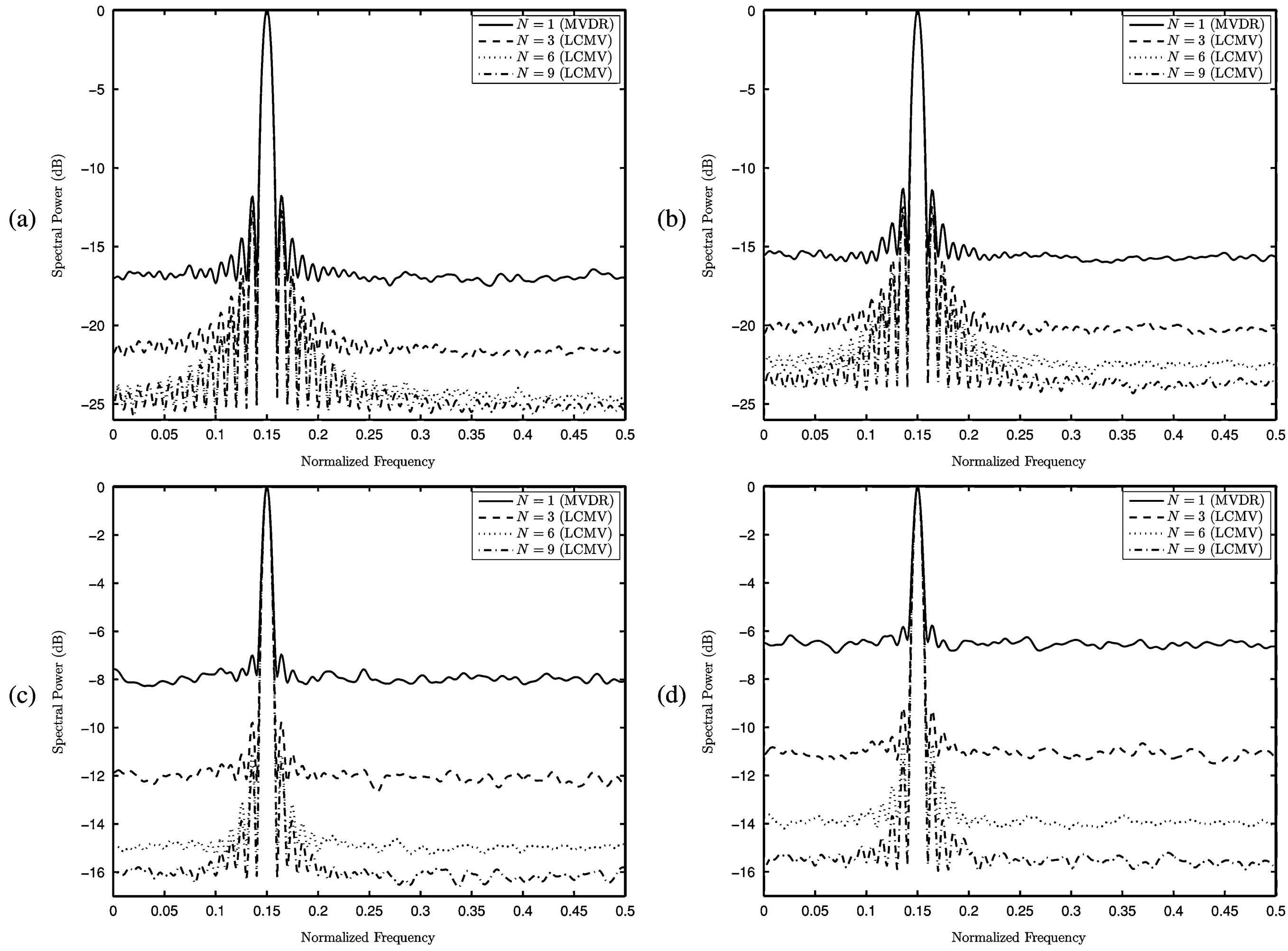

Fig. 4. Temporal spectra for a tone at $\Omega = 2\pi \cdot 0.15$ buried in Gaussian noise and reverberation. (a) SNR $=$ 0 dB, $T_{60} = 0$ ms. (b) SNR $=$ 0 dB, $T_{60} = 300$ ms. (c) SNR $= -10$ dB, $T_{60} = 0$ ms. (d) SNR $= -10$ dB, $T_{60} = 300$ ms.

estimate reveals a lesser contribution from noise and reverberation, a benefit which increases as additional sensors are added.

The localization of speech represents a major application of spatial spectral estimation; while human speech has very distinct characteristics such as quasi-periodicity of voiced phonemes, localization algorithms have not yet exploited these speech-specific temporal properties. The presented algorithm provides one way of accounting for the nature of speech in localization applications. Moreover, the LCMV framework is general in the sense that other temporal filtering schemes may also be applied; for example, aspects such as pitch and formants may be encoded into the constraint vector. Lastly, since microphone arrays are widely deployed to capture speech, it makes sense to utilize this spatial diversity to enhance the speech's temporal spectral estimate.

References

[1] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[2] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," *Terrestrial Magnetism and Atmospheric Electricity*, vol. 3, pp. 13–41, 1898.

[3] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[4] J. Benesty, J. Chen, and Y. Huang, "A generalized MVDR spectrum," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 827–830, Dec. 2005.

[5] J. Makhoul, "Linear prediction: A tutorial overview," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[6] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Dept. Geophys., Stanford Univ., Stanford, CA, 1967.

[7] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[8] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[9] J. Dibiase, H. F. Silverman, and M. S. Brandstein, , M. S. Brandstein and D. B. Ward, Eds., "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001, pp. 157–180.

[10] P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Process. Mag.*, pp. 63–6–69, Nov. 2006.

[11] Y. Huang, J. Benesty, and J. Chen, , J. Benesty, M. M. Sondhi, and Y. Huang, Eds., "Time delay estimation and source localization," in *Springer Handbook of Speech Processing*. Berlin, Germany: Springer-Verlag, 2007.

[12] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 4, pp. 320–327, Aug. 1976.

[13] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Appl. Signal Process.*, vol. 2006, p. 19, 2006, article ID 26503.

[14] J. Dmochowski, J. Benesty, and S. Affes, "Direction of arrival estimation using the parameterized spatial correlation matrix," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1327–1339, May 2007.

[15] M. Omologo and P. G. Svaizer, "Use of the cross-power-spectrum phase in acoustic event localization," ITC-IRST Tech. Rep. 9303–13, Mar. 1993.

[16] J. Krolik and D. Swingler, "Multiple broad-band source location using steered covariance matrices," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1481–1494, Oct. 1989.

[17] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[18] S. L. Marple Jr., *Digital Spectral Analysis With Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[19] N. D. Gaubitch, P. A. Naylor, and D. B. Ward, "Statistical analysis of the autoregressive modeling of reverberant speech," *J. Acoust. Soc. Amer.*, vol. 120, pp. 4031–4039, Dec. 2006.

[20] B. R. Breed and J. Strauss, "A short proof of the equivalence of LCMV and GSC beamforming," *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 168–169, Jun. 2002.

[21] A. N. Tikhonov, "On the stability of inverse problems," *Dokl. Akad. Nauk SSSR*, vol. 39, pp. 195–198, 1943.

[22] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.

[23] M. R. Schroeder, "New method for measuring reverberation time," *J. Acoust. Soc. Amer.*, vol. 37, pp. 409–412, 1965.

Jacek Dmochowski was born in Gdansk, Poland, in December 1979. He received the B.Eng. degree with high distinction in communications engineering and the M.A.Sc. degree in electrical engineering from Carleton University, Ottawa, ON, Canada, in 2003 and 2005, respectively. He is currently pursuing the Ph.D. degree at the University of Quebec, INRS-EMT, Montreal, QC, Canada.

His research interests include microphone array beamforming and source localization, blind source separation, as well as frequency-domain uncertainty analysis.

Mr. Dmochowski is the recipient of the National Sciences and Engineering Research Council (NSERC) Post Graduate Scholarship at the Doctoral Level (2005–2007).

Jacob Benesty (M'98–SM'04) was born in 1963. He received the M.S. degree in microwaves from Pierre and Marie Curie University, Paris, France, in 1987, and the Ph.D. degree in control and signal processing from Orsay University, Paris, in April 1991.

While studying for the Ph.D. degree (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris. From January 1994 to July 1995, he worked at Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In May 2003, he joined the University of Quebec, INRS-EMT, Montreal, QC, Canada, as an Associate Professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He coauthored the books *Microphone Array Signal Processing* (Springer-Verlag, 2008), *Acoustic MIMO Signal Processing* (Springer-Verlag, 2006), and *Advances in Network and Acoustic Echo Cancellation* (Springer-Verlag, 2001). He is the Editor-in-Chief of the reference *Springer Handbook of Speech Processing* (Springer-Verlag, 2007). He is also a coeditor/coauthor of the books *Speech Enhancement* (Springer-Verlag, 2005), *Audio Signal Processing for Next Generation Multimedia communication Systems* (Kluwer, 2004), *Adaptive Signal Processing: Applications to Real-World Problems* (Springer-Verlag, 2003), and *Acoustic Signal Processing for Telecommunication* (Kluwer, 2000).

Dr. Benesty received the 2001 Best Paper Award from the IEEE Signal Processing Society. He was a member of the editorial board of the *EURASIP Journal on Applied Signal Processing* and was the cochair of the 1999 International Workshop on Acoustic Echo and Noise Control.

Sofiène Affes (M'94–SM'04) received the Diplôme d'Ingénieur degree in electrical engineering and the Ph.D. degree with honors in signal processing, both from the École Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1992 and 1995, respectively.

He has been with INRS-ÉMT, University of Quebec, Montreal, QC, Canada, as a Research Associate from 1995 until 1997, then as an Assistant Professor until 2000. Currently, he is an Associate Professor in the Personal Communications Group. His research interests are in wireless communications, statistical signal and array processing, adaptive space–time processing and MIMO. From 1998 to 2002, he was leading the radio-design and signal processing activities of the Bell/Nortel/NSERC Industrial Research Chair in Personal Communications at INRS-ÉMT. Currently he is actively involved in a major project in wireless of PROMPT-Québec (Partnerships for Research on Microelectronics, Photonics, and Telecommunications).

Prof. Affes is the corecipient of the 2002 Prize for Research Excellence of INRS and currently holds a Canada Research Chair in High-Speed Wireless Communications. He served as a General Co-Chair of the IEEE VTC'2006-Fall conference, Montréal, Canada, and currently acts as a member of Editorial Board of the *Wiley Journal on Wireless Communications and Mobile Computing*.