Beamformer Design For Nonstationary Signals by Means of Interfrequency Correlations

Tofigh Naghibi and Beat Pfister
Speech Processing Group, Computer Engineering and Networks Lab.
ETH Zurich, Switzerland
{naghibi, pfister}@tik.ee.ethz.ch

***Abstract*—**

The wide-sense stationary assumption has been frequently employed in array processing, since it results in uncorrelated frequency bins and consequently major simplifications arise. However, unlike stationary signals, significant interfrequency correlations are observable in nonstationary signals like speech. Here, we drop the stationarity assumption and will show that taking interfrequency correlations into account leads to higher noise reduction. We develop a framework to design nonstationary beamformers similar to the stationary Frost beamformers. Based on the noise model, it can be used to design both fixed and adaptive beamformers. The nonstationary beamformer derived here is a set of time-varying filters and hence can be seen as a set of time-frequency or wavelet transform filters.

I. INTRODUCTION

In array processing, a signal of interest is measured by an arbitrary manifold of sensors and it is very common to assume the measured signal obeys some kind of stationarity regardless of the true signal characteristics. In practice, this assumption can be heavily violated in some applications such as microphone arrays for speech acquisition. It was shown in [2] that unlike stationary processes, there is a correlation between different frequency components of nonstationary signals. Even when considering a stationary signal, e.g. an additive white noise, significant interfrequency correlations appear because of the finite length of the observation window [3]. Considering these interfrequency correlations may lead to some unconventional results. For example, Dmochowski et al. [1] show that the spatial Nyquist criterion has little importance for microphone arrays. Their work reveals that because of the interfrequency correlations, no spatial aliasing effect appear in microphone arrays and as a result, they do not need to be restricted in spatial extent. Interfrequency correlation can be seen as an additional information which can be used in array processing to achieve a higher noise reduction or more accurate direction of arrival (DOA) estimation.

Motivated by the unexpected results in [1] for nonstationary signals, in this work nonstationary beamformers are investigated and it is shown that taking the nonstationary assumption into account leads to higher performance at the expense of increased complexity. One source of this complexity is the time-variant filtering (see [6]) appearing in the nonstationary beamformers.

The structure of this paper is as follows: The next section provides the mathematical background required for the spectral characterization of nonstationary signals. Then in section III the nonstationary beamformer framework is derived and some simulation results are presented in the next section.

Notation: throughout this paper, we use bold upper case letters to denote matrices, and bold lower case letters to signify column vectors. Furthermore, $\{.\}^H$ and $E\{.\}$ will be used to denote the complex conjugate transpose and the expectation of the corresponding argument, respectively. Finally, $\otimes$ denotes the Kronecker product.

II. REVIEW OF NONSTATIONARY SIGNALS SPECTRAL THEORY

In this section we provide a short introduction to the so-called generalized spectrum which is a generalization of the power spectrum concept to nonstationary signals. Here, only essential mathematical concepts are described. More information can be found in [2] and the references therein.

A zero mean time series $x[n]$ is said to be second-order when its covariance function $R(n,m) = E\{x[n], x^*[m]\}$ exists and is finite for all m and n. A second-order time series is termed harmonizable when its covariance function is the Fourier transform of a complex-valued measure $\mu(f_1, f_2)$:

$$R(n,m) = \iint\limits_{D^2} e^{j2\pi(f_1 n - f_2 m)} d\mu(f_1, f_2) \tag{1}$$

This integral can be seen as a Lebegue integral with μ as the measure. Here, D^2 is $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$ the domain of integration. n, m and f_1, f_2 denote time instants and temporal frequencies, respectively. The harmonizable stochastic process $x[n]$ then has the stochastic integral representation:

$$x[n] = \int_D e^{j2\pi f n} d\zeta(f) \tag{2}$$

where $\zeta(f)$ is a complex valued random measure and is related to μ as follows:

$$E\{\zeta(f_1)\zeta^*(f_2)\} = \mu(f_1, f_2) \tag{3}$$

$\zeta(f)$ is a second order process whose covariance function is given by $\mu(f_1, f_2)$. The measure μ is called the general spectral distribution of $x[n]$ and under certain conditions, it is

978-1-4673-1071-0/12/$31.00 ©2012 IEEE

absolutely continuous. In this case, $\mu(f_1, f_2)$ has a generalized spectrum[1] $S(f_1, f_2)$ given by:

$$S(f_1, f_2)df_1 df_2 = d\mu(f_1, f_2) \quad (4)$$

and the associated random measure $\zeta(f)$ is also absolutely continuous and is related to the Fourier transform of $x[n]$ as follows:

$$d\zeta(f) = X(f)df \quad (5)$$

With these results, we can now rewrite the time-frequency domain relations of harmonizable (potentially nonstationary) signals as follows:

$$X(f) = \sum_{n=-\infty}^{\infty} e^{-j2\pi fn} x[n] \quad (6)$$

$$S(f_1, f_2) = E\{X(f_1)X^*(f_2)\} \quad (7)$$

$$R(n, m) = \iint_{D^2} e^{j2\pi(f_1 n - f_2 m)} S(f_1, f_2) df_1 df_2 \quad (8)$$

The generalized spectrum $S(f_1, f_2)$ plays a key role in developing the nonstationary beamformers. The generalized spectrum of a nonstationary white noise (white noise with time-varying power) has been illustrated in Figure 1. It is not difficult to show that unlike nonstationary signals, the generalized spectrum values of wide-sense stationary (WSS) signals are zero for $f_1 \neq f_2$. That means:

$$S(f_1, f_2) = \begin{cases} S(f) & f_1 = f_2 = f \\ 0 & f_1 \neq f_2 \end{cases} \quad (9)$$

where $S(f)$ is the power spectral density of the random process $x[n]$. Thus for stationary processes, Fourier coefficients of the distinct frequency bins are uncorrelated. Other way around for nonstationary signals, $E\{X(f_1)X^*(f_2)\} = S(f_1, f_2) \neq 0$ for $f_1 \neq f_2$ which is somewhat of a misunderstanding in literature. For instance the nonstationary noise and the statistical independence of the DFT coefficients assumptions are common contradictory assumptions appearing in some papers. However, while for WSS signals each frequency bin can be processed individually, for nonstationary signals it may lead to some performance degradation when the interfrequency correlations are ignored.

III. Nonstationary beamformers

Consider a K elements microphone array with an arbitrary manifold measuring the harmonizable (nonstationary) signal. The microphone outputs can be expressed in the frequency domain as:

$$\mathbf{y} = X(f)\mathbf{d} + \mathbf{n} \quad (10)$$

where $X(f)$ is the spectrum of the desired signal and $\mathbf{n} = [N_1(f), \ldots, N_K(f)]$ a vector of the additive sensor noises $N_i(f)$. The steering vector is defined by $\mathbf{d} = [e^{-j2\pi f\tau_1}, \ldots, e^{-j2\pi f\tau_K}]$, where $\tau_1, \ldots, \tau_K$ are delays relating the desired source location to the microphones. It should be noted that both signal and noise are harmonizable (potentially nonstationary) processes and they are assumed to be statistically independent. Using (7), the generalized spectrum of the array data elements $\mathbf{y}$ can readily be written as

$$\mathbf{S}_{yy}(f_1, f_2) = S_{xx}(f_1, f_2)\mathbf{d}^H(f_1)\mathbf{d}(f_2) + \mathbf{S}_{nn}(f_1, f_2) \quad (11)$$

where $\mathbf{S}_{yy}(f_1, f_2)$ is a $K \times K$ matrix termed the generalized covariance matrix function[2] of the array data. Similarly, $\mathbf{S}_{nn}(f_1, f_2)$ denotes the generalized noise covariance matrix function. The goal of the beamformer is to obtain an estimate of the desired signal spectrum $X(f)$ by filtering and summing the microphone outputs:

$$Z(f) = \mathbf{w}^H(f)\mathbf{y}(f) \quad (12)$$

where $\mathbf{w} = [W_1(f), \ldots, W_K(f)]^T$ is a weight vector containing the beamforming weights $W_i(f)$, $i = 1, \ldots, K$ and $Z(f)$ the beamformer output. Later, it will be revealed that both $\mathbf{w}$ and Z are time-variant and have to be denoted as $Z(f, n)$ and $\mathbf{w}(f, n)$. However, here we drop n to avoid notational distraction. The generalized power spectra of $Z(f)$ and $\mathbf{y}(f)$ are related as:

$$S_{zz}(f_1, f_2) = \mathbf{w}(f_1)^H \mathbf{S}_{yy}(f_1, f_2)\mathbf{w}(f_2). \quad (13)$$

By invoking (8) and (13) the beamformer output power at each time instant n is given by

$$E\{z^2[n]\} = R_{zz}(n, n) \quad (14)$$
$$= \iint_{D^2} \mathbf{w}(f_1)^H e^{j2\pi(f_1 - f_2)n} \mathbf{S}_{yy}(f_1, f_2)\mathbf{w}(f_2) df_1 df_2,$$

where $z[n]$ is the inverse Fourier transform of $Z(f)$ and R_{zz} its covariance function. The conventional MVDR (minimum variance distortionless response) beamformer chooses its weight

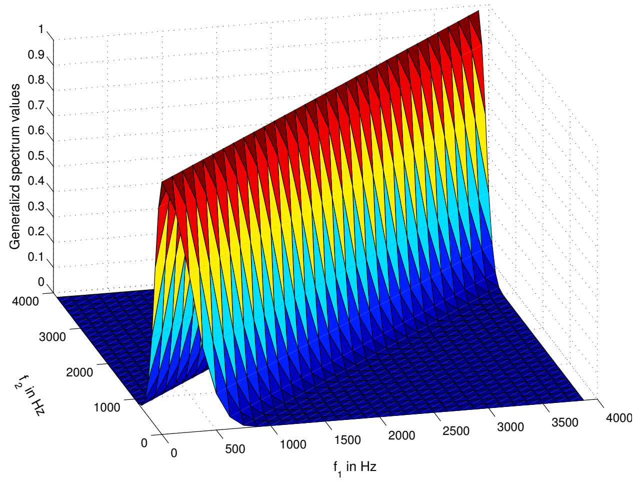

Fig. 1: Generalized spectrum of 64 white noise samples multiplied with a Hamming window. The sampling frequency is 8 kHz.

[1]The generalized spectrum is also referred to as the bifrequency spectral correlation, Loeve bifrequency spectrum and co-intensity spectrum in literature.

[2]The term *matrix function* is used to emphasize on the fact that it is a matrix-valued function of f_1 and f_2.

vector $\mathbf{w}(f)$ to minimize the output power while maintaining the signal from a specified direction of arrival. However, minimizing the output power of WSS signals is equivalent to minimizing the output power spectrum at each frequency point since for WSS signals, (14) reduces to

$$E\{z^2[n]\} = R_{zz}(0) = \int_D \mathbf{w}^H(f)\mathbf{S}_{yy}(f)\mathbf{w}(f)df \tag{15}$$

and thus,

$$\begin{aligned} &\underset{\mathbf{w}(f)}{\text{argmin}} \int_D \mathbf{w}^H \mathbf{S}_{yy}(f)\mathbf{w}(f)df \\ &= \underset{\mathbf{w}(f)}{\text{argmin}}\ \mathbf{w}^H(f)\mathbf{S}_{yy}(f)\mathbf{w}(f) \qquad \forall f \in D \end{aligned} \tag{16}$$

where the last equality follows from the fact that the integrand is non-negative for all f. For nonstationary signals, the time-domain optimization can not be reduced as for WSS signals in (16). Hence, to design the optimal beamformer one has to deal with the double integral appearing in (14). Considering (14), the nonstationary MVDR beamformer may be expressed as follows:

$$\begin{aligned} &\underset{\mathbf{w}(f)}{\text{argmin}} \iint_{D^2} \mathbf{w}(f_1)^H e^{j2\pi(f_1-f_2)n}\mathbf{S}_{yy}(f_1,f_2)\mathbf{w}(f_2)df_1 df_2 \\ &\qquad \text{s.t} \quad \mathbf{w}^H(f)\mathbf{d}(f) = 1 \qquad \forall f \in D \end{aligned} \tag{17}$$

Our aim is to reformulate this continuous objective function in the discrete-frequency domain and represent the nonstationary MVDR beamformer in matrix notation in analogy to the stationary MVDR beamformer. By discretizing the double integral in (17) and estimating it with its Riemann sum[3], we have:

$$\underset{\mathbf{w}(f)}{\text{argmin}}\ (\Delta f)^2 \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{w}^H(f_i)\mathbf{S}_{yy}(f_i,f_j)e^{j2\pi\Delta f(i-j)n}\mathbf{w}(f_j) \tag{18}$$

where N is the number of the frequency points used in the above Riemann sum and Δf the Riemann sum step size. By stacking all weight vectors $\mathbf{w}(f_i), i = 1,\ldots,N$ to the $NK\times 1$ vector $\boldsymbol{w} = [\mathbf{w}^T(f_1), \mathbf{w}^T(f_2), \ldots, \mathbf{w}^T(f_N)]^T$ and ignoring Δf, (18) can be restated in matrix notation as follows:

$$\underset{\boldsymbol{w}}{\text{argmin}}\ \boldsymbol{w}^H \boldsymbol{\mathcal{S}}_{yy}\boldsymbol{w} \tag{19}$$

where $\boldsymbol{\mathcal{S}}_{yy}$, termed generalized covariance matrix, is the $NK \times NK$ block matrix composed of blocks $\mathbf{A}_{i,j} = \mathbf{S}_{yy}(f_i,f_j)e^{j2\pi\Delta f(i-j)n}$ for $i,j = 1,\ldots,N$. As long as off-diagonal blocks of $\boldsymbol{\mathcal{S}}_{yy}$ are non-zero, distinct frequencies are correlated. However, in case of having a block-diagonal $\boldsymbol{\mathcal{S}}_{yy}$, different frequencies are uncorrelated and (19) reduces to the stationary MVDR beamformer.

To avoid converging to incorrect minima, one has to show that the objective function in (19), which is the discrete approximation of the positive function in (14), is bounded from below by zero. In other words, $\boldsymbol{\mathcal{S}}_{yy}$ has to be a positive (semi) definite matrix (which is a necessary condition for a covariance matrix). Actually, as it can be seen in Appendix A, it is indeed a positive definite matrix regardless of the complex time-dependent terms $e^{j2\pi\Delta f(i-j)n}$.

[3] Under some continuity assumptions

To complete the nonstationary MVDR beamformer representation, the distortionless response constraint in (17) has to be rewritten in the discrete-frequency domain. This can be done by defining an $NK \times N$ steering matrix $\mathbf{D}$ as:

$$\mathbf{D} = \begin{pmatrix} \mathbf{d}(f_1) & \mathbf{0}_{K\times1} & \cdots & \mathbf{0}_{K\times1} \\ \mathbf{0}_{K\times1} & \mathbf{d}(f_2) & \cdots & \mathbf{0}_{K\times1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{K\times1} & \mathbf{0}_{K\times1} & \cdots & \mathbf{d}(f_N) \end{pmatrix} \tag{20}$$

Then, the distortionless constraint follows as:

$$\mathbf{D}^H\boldsymbol{w} = \mathbf{1}_{N\times1} \tag{21}$$

Invoking (21) and (19), the nonstationary MVDR beamformer in the discrete-frequency domain is the solution of:

$$\underset{\boldsymbol{w}}{\text{argmin}}\ \boldsymbol{w}^H\boldsymbol{\mathcal{S}}_{yy}\boldsymbol{w} \quad \text{s.t} \quad \mathbf{D}^H\boldsymbol{w} = \mathbf{1} \tag{22}$$

which, by using Lagrange multipliers, can readily be written as:

$$\boldsymbol{w} = \boldsymbol{\mathcal{S}}_{yy}^{-1}\mathbf{D}(\mathbf{D}^H\boldsymbol{\mathcal{S}}_{yy}^{-1}\mathbf{D})^{-1}\mathbf{1} \tag{23}$$

It is worthwhile to note that, since $\boldsymbol{\mathcal{S}}_{yy}$ is time dependent because of the $e^{j2\pi\Delta f(i-j)n}$ terms, beamforming weights are time-varying filters and have to be designed for all time instants. For instance, for a frame with N samples, (23) has to be N times evaluated. Note that the same number of executions is necessary for conventional beamformers since the beamformers have to be designed for N frequency points. However, unlike conventional adaptive beamformers, nonstationary beamformers do not converge to the Wiener filters (unless the generalized spectrum is available in advance) and thus have to be estimated for all frames.

As can be seen in (23), the knowledge of the generalized spectrum is necessary to evaluate the nonstationary beamformers. Either it can be estimated from the array data (though it is a difficult task, see [4]) or a predetermined model based on the nature of the noise is assumed for it. In the following section, we consider the second approach and assume a model for the nonstationary noise.

IV. Simulation Results

To evaluate the proposed beamformer, we assume an additive white noise with time-varying power. Thus, the noise at each sensor is $n_i[n]\sqrt{q[n]}, i = 1,\ldots,K$ where $q[n]$ is a positive time-varying function describing the power variation of the noise. Without loss of generality, the $n_i[n]$ variance is assumed to be 1. A uniform linear array consisting of 8 microphones with 9 cm interspace and a sound source placed 3 meters away from the array center on the array axis are assumed in this scenario. By using (6-8) and considering the fact that the autocorrelation function of the stationary white noise $n_i[n]$ is the Dirac function $\delta[n-m]$, the generalized spectrum of the additive noise at each microphone can easily be written as:

$$S(f_1, f_2) = Q(f_1 - f_2) \quad (24)$$

where Q is the discrete Fourier transform of $q[n]$. Let us define a Toeplitz matrix $\mathbf{Q}$ with the first row:

$$[Q(0), e^{-j2\pi\Delta f n}Q(\Delta f), \ldots, e^{-j2\pi\Delta f(N-1)n}Q(\Delta f(N-1))]. \quad (25)$$

Then, the generalized covariance matrix in (23) is simply

$$\boldsymbol{\mathcal{S}}_{yy} = \mathbf{Q} \otimes \mathbf{I}. \quad (26)$$

By using the Kronecker product property $(\mathbf{B} \otimes \mathbf{C})^{-1} = \mathbf{B}^{-1} \otimes \mathbf{C}^{-1}$ and the fact that $\mathbf{Q}$ is a Toeplitz matrix, the inversion of $\boldsymbol{\mathcal{S}}_{yy}$ can be calculated with about $O(N \log N)$ operations [5]. Thus the weight vector can be found from (23) with reasonable computational complexity. For several $q[n]$ functions, the beamformer output power has been computed from (14) and depicted in Figure 2 for the conventional delay-and-sum (DS) beamformer and the nonstationary (NS) beamformer. As can be seen in Figure 2, by considering the power variations of the noise, the NS beamformer can suppress noise more efficiently. The $q_i[n]$ functions are chosen to have different bandwidths to evaluate the NS beamformer performance under different noise conditions. One can argue from (24) that the larger the bandwidth of $q_i[n]$ (in this case rectangular pulses) is, the more frequency bins are correlated and hence, a larger performance difference between DS and NS beamformers is expected. Obviously, by tending $q[n]$ to a constant function, the NS beamformer converges to the conventional DS beamformer (not shown here). Therefore, the proposed beamformer can be seen as the nonstationary DS beamformer.

V. Conclusion

We proposed a framework to design the nonstationary fixed and adaptive beamformers. In contrast to the traditional beamformers, the proposed algorithm uses interfrequency correlations to achieve higher noise reduction and thus more precise estimate of the desired signal. The proposed nonstationary beamformer is a set of time-varying filters. We have shown that they do not impose any additional computational cost on the beamforming design procedure. However, they require higher computational efforts to reconstruct the time-domain signals than time-invariant filters. The algorithm suggested here, needs to estimate the generalized spectrum of the array data or assume a predetermined model for it. Both approaches have their own difficulties. A possible solution can be to use a combination of these approaches, i.e., a noise model with several parameters that have to be estimated from the array data.

Acknowledgment

This work has been supported by Swiss National Science Foundation (SNSF).

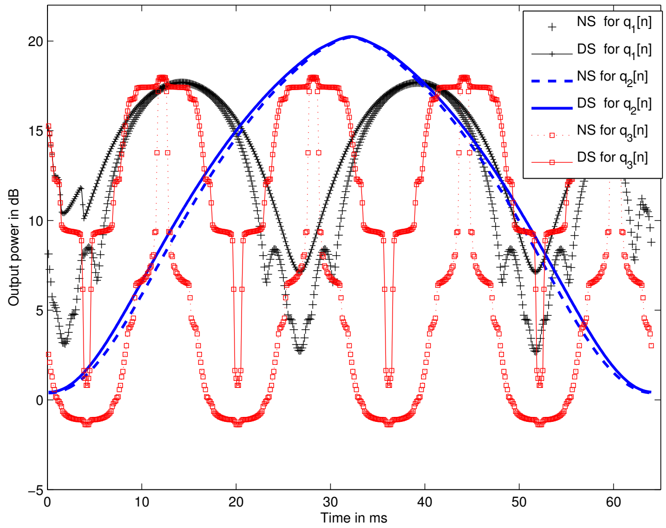

Fig. 2: Output power as a function of time for NS and conventional DS beamformers. Output powers have been calculated for 3 different nonstationary noises with $q_1[n] = (|cos(40\pi n/8000)| + 0.1)^2$, $q_2[n] = e^{-2(0.004n-1)}(\frac{1}{2} + \frac{1}{2}cos(32\pi(n-256)/8000) + 0.1)^2$ and $q_3[n] = \left((\sum_k \Pi(n/64 - 2k)) + 0.1\right)^2$, respectively. In $q_3[n]$, $\Pi(t)$ is a rectangular function that is equal to 1 for $0 \le t \le 1$ and 0 otherwise.

Appendix A

To prove the positive definiteness of $\boldsymbol{\mathcal{S}}_{yy}$, we use the fact that the Kronecker product, denoted by $\otimes$, of two positive definite matrices is a positive definite matrix. In (19), $\boldsymbol{\mathcal{S}}_{yy}$ can be written as a product of the positive definite matrices as follows:

$$\boldsymbol{\mathcal{S}}_{yy} = E\{\boldsymbol{y}_c \boldsymbol{y}_c^H\} \odot (\mathbf{e}\mathbf{e}^H \otimes \mathbf{I})$$

where $\boldsymbol{y}_c = [\mathbf{y}^T(f_1), \ldots, \mathbf{y}^T(f_N)]^T]$ and $\mathbf{e} = [e^{j2\pi f_1 n}, \ldots, e^{j2\pi f_N n}]^T$. In the expression above, $\odot$ denotes the Hadamard (pointwise) product which is the special case of the Kronecker product and hence holds all its properties.

References

[1] J. Dmochowski, J. Benesty, and S. Affes. On spatial aliasing in microphone arrays. *IEEE Trans. on Signal Processing*, 57:1383–1395, 2009.

[2] W. Gardner, A. Napolitano, and L. Paura. Cyclostationarity: Half a century of research. *IEEE Trans. on Signal Processing*, 86:639–697, April 2006.

[3] N.L. Gerr and J.C. Allen. The generalised spectrum and spectral coherence of a harmonizable time series. *Digital Signal Processing*, 4:222–238, 1994.

[4] H.L. Hurd. Spectral coherence of nonstationary and transient stochastic processes. *Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling*, pages 387–390, 1988.

[5] P.G. Martinsson, V. Rokhlin, and M. Tygert. A fast algorithm for the inversion of general Toeplitz matrices. *Computers Mathematics with Applications*, 50:741–752, 2005.

[6] L.A. Zadeh. Frequency analysis of variable networks. *Proceedings of the IRE*, 38:291–299, 1950.