# On the Noise Reduction Performance of a Spherical Harmonic Domain Tradeoff Beamformer

Daniel P. Jarrett and Emanuël A. P. Habets, *Senior Member, IEEE*

***Abstract*—In this letter, we derive an expression for the expected incoherent noise reduction factor of a spherical harmonic domain (SHD) tradeoff beamformer. The tradeoff beamformer attempts to reduce noise while minimizing speech distortion, and includes the minimum variance distortionless response (MVDR) and multichannel Wiener filters as special cases. For the open spherical microphone array, we find a number of analogies between the expressions for the expected noise reduction factor of the SHD and spatial domain MVDR beamformers. In an anechoic environment we find that the performance of the SHD MVDR beamformer with an open array depends almost entirely on the number of microphones, as in the spatial domain.**

***Index Terms*—Beamforming, incoherent noise reduction, minimum variance distortionless response (MVDR), Wiener.**

## I. Introduction

In distant speech acquisition scenarios, such as hands-free telephony, the captured speech is corrupted by noise and reverberation, thereby impairing its intelligibility. A common way of alleviating this problem is the use of microphone arrays combined with spatio-temporal filters, often called beamformers, which seek to attenuate noise while preserving the desired speech.

Beamforming techniques have been extensively studied for linear microphone arrays (see [1] and the references therein). More recently, spherical microphone arrays have been proposed to analyze the sound field in three dimensions [2], [3]. The sound field captured by these arrays can be efficiently represented and processed in the spherical harmonic domain (SHD). Spatial domain beamformers filter the signals received at each microphone; similarly SHD beamformers filter eigenbeams, which are obtained by applying the spherical Fourier transform to the spatial domain signals. An optimal SHD beamformer that provides control of performance measures such as sidelobe level and white noise gain was recently proposed in [4].

In this letter, we analyze the incoherent noise reduction performance of a previously proposed SHD tradeoff beamformer [5]. This beamformer achieves a balance between noise reduction and speech distortion, controlled by a tradeoff parameter. For specific choices of this parameter, SHD equivalents of the well-known minimum variance distortionless response (MVDR) and multichannel Wiener filters are obtained.

Knowledge of the noise reduction performance of the tradeoff beamformer as a function of various parameters such as the input signal to noise ratio, tradeoff parameter, array type and radius allow us to appropriately choose these parameters in order to achieve a certain level of noise reduction. In addition, we show that under certain circumstances, the expressions obtained for the performance of the SHD MVDR beamformer converge to expressions previously derived for the spatial domain MVDR beamformer.

Manuscript received July 02, 2012; revised August 21, 2012; accepted August 23, 2012. Date of publication September 24, 2012; date of current version October 03, 2012. D. P. Jarrett was a Visiting Researcher at the International Audio Laboratories Erlangen during the course of this work. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Constantine L. Kotropoulos.

D. P. Jarrett is with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K. (e-mail: daniel.jarrett05@imperial.ac.uk).

E. A. P. Habets is with the International Audio Laboratories Erlangen (a joint institution between the University of Erlangen-Nuremberg and Fraunhofer IIS), Erlangen, Germany (e-mail: emanuel.habets@audiolabs-erlangen.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2012.2220351

## II. Signal Models and Tradeoff Beamformer

### A. Spatial and Spherical Harmonic Domain Signal Models

We consider a conventional short-time Fourier transform (STFT) domain signal model in which a spherical microphone array captures $Q$ noisy signals $P(k, \mathbf{r}_q)$ at a discrete frequency index $k$ and microphone positions $\mathbf{r}_q = (r, \Omega_q)$ (in spherical coordinates), where $r$ is the radius of the sphere. The $q$th received signal, $P(k, \mathbf{r}_q)$, consists of a convolved speech signal $X(k, \mathbf{r}_q)$ and a noise signal $V(k, \mathbf{r}_q)$:

$$\begin{aligned} P(k, \mathbf{r}_q) &= G(k, \mathbf{r}_q) S(k) + V(k, \mathbf{r}_q) \\ &= X(k, \mathbf{r}_q) + V(k, \mathbf{r}_q), \quad q \in \{1, \ldots, Q\} \end{aligned} \tag{1}$$

where $G(k, \mathbf{r}_q)$ is the acoustic transfer function from the speech source $S(k)$ to the microphone at position $\mathbf{r}_q$, and is assumed to be time-invariant. For brevity, the dependency on the time frame index is omitted.

The received speech signals $X(k, \mathbf{r}_q)$ and the received noise signals $V(k, \mathbf{r}_q)$ are assumed to be uncorrelated. The received speech signals originate from a single source and are therefore, by definition, coherent across the array.

When dealing with spherical microphone arrays, it is convenient to work in the SHD. The spherical Fourier transform $F_{lm}(k)$ of a spatial domain signal $F(k, \mathbf{r}_q)$ involves an integral over all angles, however it can be approximated for a discretely sampled sound field using the expression [3]

$$F_{lm}(k) \approx \sum_{q=1}^{Q} c_q F(k, \mathbf{r}_q) Y_{lm}^{*}(\Omega_q), \tag{2}$$

where $Y_{lm}$ is the spherical harmonic of order $l$ and degree $m$, and $(\cdot)^*$ denotes the complex conjugate. The spherical Fourier transform coefficient $F_{lm}(k)$ for all values of $k$ is often called the *eigenbeam* of order $l$ and degree $m$, as the spherical harmonics are the eigensolutions of the acoustic wave equation in spherical

1070-9908/$31.00 © 2012 IEEE

coordinates. The weights $c_q$ are chosen such that the approximation in (2) is as accurate as possible [3]; with a sufficient number of microphones and appropriate positioning, the approximation error involved can be eliminated entirely. In this letter, we choose $c_q = 4\pi/Q$ which is suitable for $Q$ equidistant microphones.

We can now express our signal model in the SHD as

$$P_{lm}(k) = G_{lm}(k)S(k) + V_{lm}(k) = X_{lm}(k) + V_{lm}(k) \quad (3)$$

where $P_{lm}(k)$, $G_{lm}(k)$, $X_{lm}(k)$ and $V_{lm}(k)$ respectively denote the SHD representations of $P(k,\mathbf{r}_q)$, $G(k,\mathbf{r}_q)$, $X(k,\mathbf{r}_q)$ and $V(k,\mathbf{r}_q)$.

The SHD signals $P_{lm}(k)$, $G_{lm}(k)$, $X_{lm}(k)$ and $V_{lm}(k)$ are dependent on the mode strength $B_l(k)$ [3], [6], which is a function of the array properties (configuration, microphone type, radius). Mode strength expressions for various configurations (open, rigid, dual-sphere, etc.) can be found in [6]. To cancel the dependence on the array properties, we divide our eigenbeams by the mode strength to yield mode strength compensated signals:

$$\begin{aligned}\widetilde{P}_{lm}(k) &= \left[\sqrt{4\pi}B_l(k)\right]^{-1} P_{lm}(k)\\ &= \widetilde{G}_{lm}(k)S(k)+\widetilde{V}_{lm}(k)=\widetilde{X}_{lm}(k)+\widetilde{V}_{lm}(k)\end{aligned} \quad (4)$$

where $\widetilde{P}_{lm}(k), \widetilde{G}_{lm}(k), \widetilde{X}_{lm(k)}$ and $\widetilde{V}_{lm}(k)$ respectively denote the signals $P_{lm}(k)$, $G_{lm}(k)$, $X_{lm}(k)$ and $V_{lm}(k)$ after mode strength compensation. With the addition of the scaling factor $\sqrt{4\pi}$, the zero order eigenbeam $\widetilde{X}_{00}$ is equal to the speech signal which would be received by an omnidirectional microphone placed at the center of the sphere, i.e., at a position $\mathbf{r}_q = \mathbf{0}$. Hence $\widetilde{X}_{00}$ will hereafter be referred to as the *omnidirectional speech signal*.

As $X(k,\mathbf{r}_q)$ and $V(k,\mathbf{r}_q)$ are uncorrelated and the STFT, spherical Fourier transform and division by the mode strength are linear operations, $\widetilde{X}_{lm}(k)$ and $\widetilde{V}_{lm}(k)$ are also uncorrelated. It can be shown that as the spatial domain signals $X(k,\mathbf{r}_q)$ are coherent across $\mathbf{r}_q$, the eigenbeams $\widetilde{X}_{lm}(k)$ are also coherent across $l$ and $m$.

### B. Tradeoff Beamformer

In this work, our desired signal is chosen to be $\widetilde{X}_{00}(k)$, the omnidirectional speech signal. For convenience, we rewrite the SHD signals in a vector notation, where each of the vectors has length $N = (L+1)^2$, the total number of eigenbeams up to order $L$:

$$\begin{aligned}\widetilde{\mathbf{p}}(k) &= \widetilde{\mathbf{g}}(k)S(k) + \widetilde{\mathbf{v}}(k)\\ &= \widetilde{\mathbf{x}}(k) + \widetilde{\mathbf{v}}(k)\\ &= \mathbf{d}(k)\widetilde{X}_{00}(k) + \widetilde{\mathbf{v}}(k),\end{aligned} \quad (5)$$

where

$$\widetilde{\mathbf{p}}(k) = \left[\widetilde{P}_{00}(k)\widetilde{P}_{1(-1)}(k)\widetilde{P}_{10}(k)\widetilde{P}_{11}(k)\cdots\widetilde{P}_{LL}(k)\right]^T, \quad (6)$$

$$\mathbf{d}(k) = \left[1\,\frac{\widetilde{G}_{1(-1)}(k)}{\widetilde{G}_{00}(k)}\,\frac{\widetilde{G}_{10}(k)}{\widetilde{G}_{00}(k)}\,\frac{\widetilde{G}_{11}(k)}{\widetilde{G}_{00}(k)}\cdots\frac{\widetilde{G}_{LL}(k)}{\widetilde{G}_{00}(k)}\right]^T, \quad (7)$$

$(\cdot)^T$ denotes the vector transpose, and $\widetilde{\mathbf{x}}(k)$, $\widetilde{\mathbf{g}}(k)$ and $\widetilde{\mathbf{v}}(k)$ are defined similarly to $\widetilde{\mathbf{p}}(k)$.

The eigenbeams $\widetilde{X}_{lm}(k)$ are coherent, therefore the signal vector $\widetilde{\mathbf{x}}(k)$ can also be written as $\widetilde{\mathbf{x}}(k) = \boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)\widetilde{X}_{00}(k)$, where

$$\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k) = \frac{E\left[\widetilde{\mathbf{x}}(k)\widetilde{X}^*_{00}(k)\right]}{E\left[\left|\widetilde{X}_{00}(k)\right|^2\right]} \quad (8)$$

is the partially normalized [with respect to $\widetilde{X}_{00}(k)$] coherence vector between $\widetilde{\mathbf{x}}(k)$ and $\widetilde{X}_{00}(k)$ and $E[\cdot]$ denotes mathematical expectation. Using (8), (5) can be expressed as

$$\widetilde{\mathbf{p}}(k) = \boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)\widetilde{X}_{00}(k) + \widetilde{\mathbf{v}}(k). \quad (9)$$

As $\widetilde{\mathbf{p}}(k)$ is the sum of two uncorrelated components $\widetilde{\mathbf{x}}(k)$ and $\widetilde{\mathbf{v}}(k)$, the correlation matrix of $\widetilde{\mathbf{p}}(k)$ is given by

$$\boldsymbol{\Phi}_{\widetilde{\mathbf{p}}}(k) = E\left[\widetilde{\mathbf{p}}(k)\widetilde{\mathbf{p}}^H(k)\right] = \boldsymbol{\Phi}_{\widetilde{\mathbf{x}}}(k) + \boldsymbol{\Phi}_{\widetilde{\mathbf{v}}}(k), \quad (10)$$

where $\boldsymbol{\Phi}_{\widetilde{\mathbf{x}}}(k) = \phi_{\widetilde{X}_{00}}(k)\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)\boldsymbol{\gamma}^H_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)$ and $\boldsymbol{\Phi}_{\widetilde{\mathbf{v}}}(k) = E[\widetilde{\mathbf{v}}(k)\widetilde{\mathbf{v}}^H(k)]$ are respectively the covariance matrices of $\widetilde{\mathbf{x}}(k)$ and $\widetilde{\mathbf{v}}(k)$, $\phi_{\widetilde{X}_{00}}(k) = E[|\widetilde{X}_{00}(k)|^2]$ is the variance of $\widetilde{X}_{00}(k)$, and $(\cdot)^H$ denotes the Hermitian transpose.

The beamformer's output is obtained by applying a complex weight to each eigenbeam, and summing across all eigenbeams:

$$\begin{aligned}Z(k) &= \mathbf{h}^H(k)\widetilde{\mathbf{p}}(k) = \mathbf{h}^H(k)\widetilde{\mathbf{x}}(k) + \mathbf{h}^H(k)\widetilde{\mathbf{v}}(k)\\ &= \widetilde{X}_{\mathrm{fd}}(k) + \widetilde{V}_{\mathrm{rn}}(k),\end{aligned} \quad (11)$$

where $\widetilde{X}_{\mathrm{fd}}(k) = \mathbf{h}^H(k)\widetilde{\mathbf{x}}(k) = \mathbf{h}^H(k)\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)\widetilde{X}_{00}(k)$ is the filtered desired signal and $\widetilde{V}_{\mathrm{rn}}(k) = \mathbf{h}^H(k)\widetilde{\mathbf{v}}(k)$ is the residual noise.

We design a tradeoff beamformer which seeks to minimize the speech distortion while maximizing the noise reduction according to the following optimization criteria [5]:

$$\min_{\mathbf{h}(k)}\left|\mathbf{h}^H(k)\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)-1\right|^2 \text{ s.t. } \mathbf{h}^H(k)\Phi_{\widetilde{\mathbf{v}}}(k)\mathbf{h}(k) = \beta\phi_{\widetilde{V}_{00}}(k),$$

where $\phi_{\widetilde{V}_{00}}(k) = E[|\widetilde{V}_{00}(k)|^2]$, and $0 < \beta < 1$ ensures that we get some noise reduction. Using a Lagrange multiplier, $\mu \geq 0$, to adjoin the constraint to the cost function, we deduce the tradeoff filter [5]:

$$\mathbf{h}_{\mathrm{T},\mu}(k) = \frac{\phi_{\widetilde{X}_{00}}(k)\boldsymbol{\Phi}^{-1}_{\widetilde{\mathbf{v}}}(k)\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)}{\mu + \phi_{\widetilde{X}_{00}}(k)\boldsymbol{\gamma}^H_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)\boldsymbol{\Phi}^{-1}_{\widetilde{\mathbf{v}}}(k)\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)}, \quad (12)$$

where $\mu \geq 0$ is the tradeoff parameter. Increasing the value of $\mu$ increases noise reduction at the expense of higher speech distortion. It has been shown [7] that for $\mu = 0$, this corresponds to the MVDR beamformer, while for $\mu = 1$, this corresponds to the multichannel Wiener filter.

## III. Noise Reduction Performance

In this study, we assume spatially incoherent noise (e.g., sensor noise) of equal power $\phi_V$ at each of the $Q$ microphones:

$$\phi_V(k) = E\left[\left|V(k,\mathbf{r}_q)\right|^2\right], \quad q = 1,\ldots,Q. \quad (13)$$

Using [8, eqn 7.31], it can be shown that the power of the noise eigenbeams $\phi_{\widetilde{V}_{lm}}$ is related to the noise power $\phi_V$ at the microphones via the expression

$$\phi_{\widetilde{V}_{lm}}(k) = E\left[\left|\widetilde{V}_{lm}(k)\right|^2\right] = \frac{\phi_V(k)}{Q\left|B_l(kr)\right|^2}, \quad (14)$$

and that the noise eigenbeams are incoherent over $l$ and $m$. The noise eigenbeam covariance $\mathbf{\Phi}_{\widetilde{\mathbf{v}}}$ is therefore a diagonal matrix given by

$$\mathbf{\Phi}_{\widetilde{\mathbf{v}}}(k) = \operatorname{diag}\left\{\phi_{\widetilde{V}_{00}}(k), \phi_{\widetilde{V}_{1(-1)}}(k), \ldots, \phi_{\widetilde{V}_{LL}}(k)\right\}. \quad (15)$$

The power of the residual noise at the beamformer's output is given by $\phi_{\widetilde{V}_{\mathrm{rn}}}(k) = E[|\widetilde{V}_{\mathrm{rn}}(k)^2]$.

We measure the noise reduction performance of our beamformer using the *noise reduction factor* $\xi_{\mathrm{nr}}$ [9], which is defined here as the ratio of the power of the noise at the microphones over the power of the residual noise $\widetilde{V}_{\mathrm{rn}}$:

$$\xi_{\mathrm{nr}}\left[\mathbf{h}(k)\right] = \frac{\phi_V(k)}{\phi_{\widetilde{V}_{\mathrm{rn}}}(k)} = \frac{\phi_V(k)}{\mathbf{h}^H(k)\mathbf{\Phi}_{\widetilde{\mathbf{v}}}(k)\mathbf{h}(k)}. \quad (16a)$$

It should be noted that part of this noise reduction is obtained from the combination of the spatial domain signals via the spherical Fourier transform and mode strength compensation [providing a noise reduction factor of $\phi_V(k)/\phi_{\widetilde{V}_{00}}(k) = Q|B_0(kr)|^2$], and the remainder is obtained from the beamformer itself [providing a noise reduction factor of $\phi_{\widetilde{V}_{00}}(k)/\phi_{\widetilde{V}_{\mathrm{rn}}}(k)$].

## A. *Tradeoff Beamformer*

By substituting (12) into (16a), we find the noise reduction factor $\xi_{\mathrm{nr}}[\mathbf{h}_{\mathrm{T},\mu}(k)]$ for the tradeoff beamformer:

$$\xi_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{T},\mu}(k)\right] = \frac{\phi_V(k)}{\mathbf{h}^H_{\mathrm{T},\mu}(k)\mathbf{\Phi}_{\widetilde{\mathbf{v}}}(k)\mathbf{h}_{\mathrm{T},\mu}(k)} \quad (17a)$$

$$= \frac{\phi_V(k)\left[\mu + \lambda(k)\right]^2}{\phi_{\widetilde{X}_{00}}(k)\lambda(k)}, \quad (17b)$$

where $\lambda(k)$ is the *a priori* multichannel signal-to-noise ratio (SNR) given by

$$\lambda(k) = \phi_{\widetilde{X}_{00}}(k)\boldsymbol{\gamma}^H_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k)\mathbf{\Phi}^{-1}_{\widetilde{\mathbf{v}}}(k)\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k) \quad (18a)$$

$$= \operatorname{tr}\left[\mathbf{\Phi}^{-1}_{\widetilde{\mathbf{v}}}(k)\mathbf{\Phi}_{\widetilde{\mathbf{x}}}(k)\right], \quad (18b)$$

and $\operatorname{tr}[\cdot]$ is the matrix trace operator. We also define the subband input SNR as the ratio of the omnidirectional speech power to the power of the noise at any of the microphones:

$$\mathrm{iSNR}(k) = \frac{\phi_{\widetilde{X}_{00}}(k)}{\phi_V(k)}. \quad (19)$$

Using (19), we can simplify the expression in (17b):

$$\xi_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{T},\mu}(k)\right] = \frac{\left[\mu + \lambda(k)\right]^2}{\mathrm{iSNR}(k)\lambda(k)}. \quad (20)$$

We are interested in tradeoff beamformers which do not amplify the noise, i.e., beamformers for which the noise reduction factor is at least $Q|B_0(kr)|^2$.

**Reverberant environment**: Assuming the coherence vector $\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}$ is perfectly estimated, $\boldsymbol{\gamma}_{\widetilde{\mathbf{x}}\widetilde{X}_{00}}(k) = \mathbf{d}(k)$. Substituting (14) and (15) in (18a), and using (7), we then find

$$\lambda(k) = \frac{\phi_{\widetilde{X}_{00}}(k)Q}{\phi_V(k)\left|\widetilde{G}_{00}(k)\right|^2}\sum_{l=0}^{L}\sum_{m=-l}^{l}|B_l(kr)|^2\left|\widetilde{G}_{lm}(k)\right|^2 \quad (21a)$$

$$= \frac{\mathrm{iSNR}(k)Q}{4\pi\left|\widetilde{G}_{00}(k)\right|^2}\sum_{l=0}^{L}\sum_{m=-l}^{l}|G_{lm}(k)|^2, \quad (21b)$$

where $\widetilde{G}_{00}(k) = G(k, \mathbf{0})$ denotes the acoustic transfer function between the source and an omnidirectional microphone placed at the center of the sphere.

It can be seen from (20) and (21) that the noise reduction performance depends on the tradeoff parameter $\mu$, the number of microphones $Q$ and order $L$ of the array, the subband input SNR (except for $\mu = 0$), and the array properties (including the radius).

*Property 3.1:* The noise reduction factor of the tradeoff beamformer is an increasing function of the parameter $\mu$.

*Proof:* The derivative of the noise reduction factor with respect to $\mu$ is given by

$$\frac{\mathrm{d}\xi_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{T},\mu}(k)\right]}{\mathrm{d}\mu} = \frac{2\left[\mu + \lambda(k)\right]}{\mathrm{iSNR}(k)\lambda(k)}. \quad (22)$$

As $\lambda(k)$ is the *a priori* multichannel SNR, $\lambda(k) \geq 0$. In addition, $\mu \geq 0$ and $\mathrm{iSNR}(k) \geq 0$, therefore the derivative of the noise reduction factor with respect to $\mu$ is positive. ■

*Property 3.2:* For all $\mu > 0$, the noise reduction factor of the tradeoff beamformer is a decreasing function of the subband input SNR.

*Proof:* It can straightforwardly be shown that the derivative of $\xi_{\mathrm{nr}}$ with respect to iSNR is negative for $\mu > 0$. ■

**Anechoic environment**: Assuming plane wave incidence from a direction $\Omega_0$, the mode strength compensated SHD acoustic transfer functions are given by [3]

$$\widetilde{G}_{lm}(k) = Y^*_{lm}(\Omega_0). \quad (23)$$

Using (19), the *a priori* multichannel SNR in (21a) can then be expressed as

$$\lambda(k) = \frac{\mathrm{iSNR}(k)Q}{|Y^*_{00}(\Omega_0)|^2}\sum_{l=0}^{L}\sum_{m=-l}^{l}|B_l(kr)|^2\left|Y^*_{lm}(\Omega_0)\right|^2. \quad (24)$$

Using the spherical harmonic addition theorem [8] $\sum_{m=-l}^{l}|Y_{lm}(\Omega_0)|^2 = (2l+1)/4\pi\ \forall\Omega_0$, and the fact that $|Y_{00}(\Omega_0)|^{-2} = 4\pi\ \forall\Omega_0$, we find

$$\lambda(k) = \mathrm{iSNR}(k)Q\sum_{l=0}^{L}|B_l(kr)|^2(2l+1). \quad (25)$$

For an open sphere (i.e., microphones suspended in free space) $B_l(kr) = i^l j_l(kr)$, where $j_l$ is the spherical Bessel function of order $l$, and $\sum_{l=0}^{\infty}|j_l(kr)|^2(2l+1) = 1$ [10], therefore as $L \to \infty$, $\lambda(k)$ converges to

$$\lim_{L\to\infty}\lambda(k) = \mathrm{iSNR}(k)Q. \quad (26)$$

## B. *MVDR Beamformer*

In the special case of the MVDR beamformer, $\mathbf{h}_{\mathrm{MVDR}} = \mathbf{h}_{\mathrm{T},0}$. The noise reduction factor in (20) then simplifies to

$$\xi_{\mathrm{nr}}(k)\left[\mathbf{h}_{\mathrm{MVDR}}(k)\right] = \frac{\lambda(k)}{\mathrm{iSNR}(k)}. \quad (27)$$

From (21b) and (27) it can be seen that the noise reduction factor of the MVDR beamformer is independent of the subband input SNR, and is, as one would expect, proportional to the number of microphones $Q$ and an increasing function of the array order$L$.

**Reverberant environment**: Using (21b), (27) and Parseval's relation in the SHD [3] $\sum_{l=0}^{\infty}\sum_{m=-l}^{l}|G_{lm}(k)|^2 = \sum_{q=1}^{Q} c_q |G(k,\mathbf{r}_q)|^2$ with $c_q = 4\pi/Q$, we find that as $L \to \infty$, the noise reduction factor converges to

$$\lim_{L\to\infty} \xi_{\rm nr}(k)\left[\mathbf{h}_{\rm MVDR}(k)\right] = \sum_{q=1}^{Q} \frac{|G(k,\mathbf{r}_q)|^2}{|G(k,\mathbf{0})|^2}, \tag{28}$$

which can be compared to the expression for a spatial domain MVDR beamformer derived in [11, eqn 43]

$$\xi_{\rm nr}(k)\left[\mathbf{h}_{\rm MVDR}(k)\right] = \sum_{q=1}^{Q} \frac{|G(k,\mathbf{r}_q)|^2}{|G(k,\mathbf{r}_1)|^2}, \tag{29}$$

where the desired signal is the signal received at microphone $q = 1$, instead of the signal received at an omnidirectional microphone at the center of the sphere ($\mathbf{r}_q = \mathbf{0}$).

**Anechoic environment**: Using (25), the noise reduction factor of the MVDR beamformer is given by

$$\xi_{\rm nr}(k)\left[\mathbf{h}_{\rm MVDR}(k)\right] = Q \sum_{l=0}^{L} |B_l(kr)|^2 (2l+1). \tag{30}$$

For an open sphere, using (26), as $L \to \infty$ the noise reduction factor converges to the simple expression

$$\lim_{L\to\infty} \xi_{\rm nr}(k)\left[\mathbf{h}_{\rm MVDR}(k)\right] = Q, \tag{31}$$

which is the well-known result for the maximum incoherent noise reduction achievable without speech distortion using $Q$ omnidirectional microphones [12, p. 66]. For a finite $L$ and an open sphere $\xi_{\rm nr}(k)[\mathbf{h}_{\rm MVDR}(k)] < Q$, however $\sum_{l=0}^{L} |B_l(kr)|^2(2l+1)$ converges relatively quickly to 1.

## IV. Simulation Results

We evaluated the noise reduction performance of a third order tradeoff beamformer in an anechoic environment. We simulated a rigid and an open spherical array of radius 4.2 cm with $Q = 32$ microphones. The source signal consisted of 60 s of male and female speech. The noise signals consisted of spatially white noise with a fullband input signal to noise ratio of 0 dB at the microphone closest to the source. The noise power was set based on active speech levels, computed according to ITU-T Rec. P. 56. Processing was performed in the STFT domain at a sampling frequency of 12 kHz. The source and noise statistics were estimated based on the signal vectors, in order to neglect the influence of a voice activity detector, and were averaged across the entire signal length.

In Fig. 1, we plot the simulated noise reduction factor of the tradeoff beamformer as a function of frequency for various values of $\mu$ and for both an open and rigid array. In addition, we plot the theoretical results given by (20) and (25).

We find that the theoretical expressions accurately predict the beamformer's performance, except at 4.1 kHz, where one of the disadvantages of the open sphere configuration is revealed: the open sphere mode strength contains sharp nulls at certain frequencies, and the open configuration therefore suffers from poor robustness [3]. The performance of the rigid array is generally higher than that of the open array, most particularly for $\mu = 0$ where an improvement of approximately 2 dB is obtained in the 3–4 kHz range.

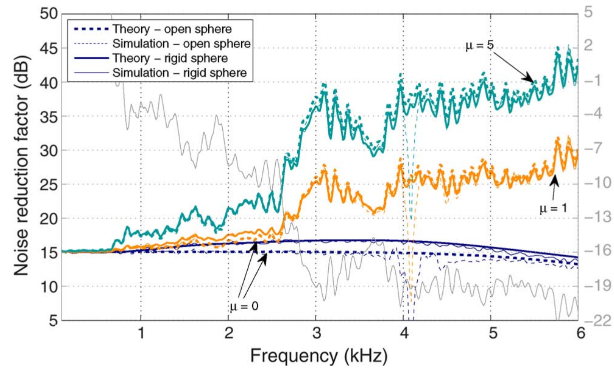


Fig. 1. Simulated and theoretical noise reduction factor as a function of frequency, for an open and rigid sphere and $\mu \in \{0, 1, 5\}$.

In this example, $\mathrm{iSNR}(k)$ is high at low frequencies, where most of the speech energy is concentrated. As predicted, the noise reduction is independent of the subband input SNR $\mathrm{iSNR}(k)$ for $\mu = 0$ and decreases with $\mathrm{iSNR}(k)$ for all $\mu > 0$.

## V. Conclusion

In this letter, we analyzed the expected incoherent noise reduction performance of a spherical harmonic domain tradeoff beamformer, and showed that it is an increasing function of the number of microphones and array order. We also made analogies between the spherical harmonic domain and spatial domain MVDR beamformers. Finally we performed simulations, the results of which closely match those obtained from the expressions derived in this letter.

## References

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

[2] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2002, vol. 2, pp. 1781–1784.

[3] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[4] S. Yan, H. Sun, U. P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 2, pp. 361–371, Feb. 2011.

[5] D. P. Jarrett, E. A. P. Habets, J. Benesty, and P. A. Naylor, "A tradeoff beamformer for noise reduction in the spherical harmonic domain," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Sep. 2012.

[6] B. Rafaely, "Spatial sampling and beamforming for spherical microphone arrays," in *Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, May 2008, pp. 5–8.

[7] J. Benesty, J. Chen, and E. A. P. Habets, *Speech Enhancement in the STFT Domain*, ser. in Electrical and Computer Engineering. Berlin, Germany: Springer-Verlag, 2011.

[8] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, 1st ed. London, U.K.: Academic, 1999.

[9] *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Berlin, Germany: Springer, 2005.

[10] *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and I. A. Stegun, Eds. New York: Dover, 1972.

[11] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, pp. 158–170, Jan. 2010.

[12] H. L. van Trees, *Detection, Estimation, and Modulation Theory*. Hoboken, NJ: Wiley, Apr. 2002, vol. IV, Optimum Array Processing.