ABSTRACT

A method for dereverberation and noise reduction is presented. The method is designed for a spherical microphone array, and formulated in the spherical harmonics domain, based on an acoustic model that is also formulated in the spherical harmonics domain. The novelty in the proposed method is the dereverberation process, which exploits the useful formulation in the spherical harmonics domain, which facilitates dereverberation by employing DOA estimation, rather than room impulse response identification. Noise reduction is further performed by a linearly constrained minimum variance filter, where the array output power is minimized with constraint of distortionless response to the direct sound. The method achieves partial dereverberation and minimizes the noise in two separate stages, as described in sections 4 and 5. After presenting the method, the performance is analyzed and compared to the algorithm in [1].

2. EXISTING METHODS

Various models for acoustic environments are described in the literature. This section presents three models and methods for broadband signals in different acoustic environments as presented in [1].

When the acoustic environment is considered as reverberant, the received signal are expressed as:

\[ y_i(n) = g_i * s(n) + v_i(n), \]

where \( g_i \) is the impulse response from the unknown source \( s(n) \) to the \( i \)th microphone, and \( n \) is the time index. It is assumed that \( s(n) \) is uncorrelated with the additive noise \( v_i(n) \). In a matrix form, the signal model (1) can be rewritten as:

\[ y_i(n) = g_i^T \cdot s(n) + v_i(n), \]

where \( g_i = [g_{i,0}, g_{i,1}, \ldots, g_{i,L_g-1}]^T \), \( s(n) = [s(n), s(n-1), \ldots, s(n-L_g+1)]^T \), and \( L_g \) is the length of the longest acoustic impulse responses among the \( I \) channels, \( g_{i}, i = 1,2,\ldots,I \). If the \( I \) impulse responses are known, or can be estimated, and are stationary, an LCMV filter can be performed with constraint of dereverberation, minimizing the output power [1]. In that case, the filter achieved full dereverberation with a significant amount of noise reduction. This filter seems to be very attractive, but it may not be very practical since the acoustic impulse responses from the unknown source to the \( I \) microphones are difficult to estimate in real-world applications [1].

A simpler model is the anechoic model. In that case, it is assumed that the signal picked up by each microphone is a delayed and attenuated version of the original source signal plus...
some additive noise. Mathematically, the received signals, at
time $n$, are expressed as
\[ y_i(n) = \alpha_i s[n - t - F_i(\tau)] + v_i(n), \] (3)
where $\alpha_i, i = 1, 2, \ldots, I$ are the attenuation factors due to
propagation effects, $s(n)$ is the unknown source signal, $t$ is
the propagation time from the unknown source to sensor 1, $v_i(n)$ is an additive noise signal at the $i$th microphone, $\tau$ is the
relative delay between microphones 1 and 2, and $F_i(\tau)$ is
the relative delay between microphones 1 and $i$ with $F_1 = 0$ and
$F_2 = \tau$. Here only noise reduction is required due to the ba-
sic assumption of the model of no reverberation. By using the
LCMV filter with constraint of distortionless or specific spec-
trum shaping to $s(n)$, noise minimization can be obtained. In
order to implement the filter, the relative delays $[F_i(\tau)]$ need
to be known or accurately estimated [4]. In addition, the at-
tenuation factors $[\alpha_i]$ are required.

Another model presented in [1] is the spatio-temporal
model. This model exploits the spatial information of the
unknown source as well as its temporal signature. It is as-
sumed here that the signals $x_i(n), i = 2, 3, \ldots, I$ are related
to $x_1(n)$ by an FIR filter:
\[ x_i(n) = \mathbf{W}_i x_1(n), i = 2, 3, \ldots, I, \] (4)
where $x_i(n) = [x_i(n), x_i(n-1), \cdots, x_i(n-L_h+1)]^T$ con-
tains the most recent $L_h$ samples of the signal without ad-
ditive noise at the $i$th microphone, and $\mathbf{W}_i$ is an $L_h \times L_h$
matrix. The LCMV filter using the spatio-temporal model
performs noise reduction at one of the microphone signals,
without trying to recover the desired source $s(n)$ but with no
further distortion. This model does not deal with the rever-
eration but reduces the noise in a reverberant environment;
however, it is more practical than the reverberant model, due
to the ability to estimate $\mathbf{W}_i$.

3. ACOUSTIC MODEL IN THE SPHERICAL
HARMONICS DOMAIN

This section presents the system model formulation in the
spherical harmonics domain. This model is then used in the
following sections.

A spherical microphone array is constructed of $I$ micro-
phones. The position of each microphone is $(r; \theta_i, \phi_i)$, where
$r$ is the array radius, and $\theta_i$ and $\phi_i$ are the elevation angle and
the azimuth angle of the $i$th microphone, respectively. The
acoustic environment is assumed to be reverberant, and the
signals can be expressed as in Eq. (2). Here it is assumed that
the impulse responses $g_i$ contain $L$ reflections. This assump-
tion will result in model mismatch, but it simplifies the for-
mulation at the cost of achieving partial dereverberation only.
Actually, the method will separate only $L$ reflections out of
the full impulse response, while all other reflections will be
considered as noise. The reflections can be considered as co-
erherent plane waves incident the array. $\Psi_i = (\theta_i, \phi_i)$ denotes
the angle of arrival of the $i$th plane-wave.

The pressure function on the sphere is assumed to be band
limited. In that case the spherical Fourier transform can be
computed by the following summation [5]:
\[ x_{nm}(k, r) = \sum_{i=1}^{I} c_i x(k, r, \Omega_i) Y_{n}^{m*}(\Omega_i), \] (5)
where $c_i$ are the weights of the sampling scheme, $x(k, r, \Omega_i)$
is the input signal at the $i$th microphone, $\Omega_i = (\theta_i, \phi_i)$ are
the samples on the sphere representing sensor positions, and $Y_{n}^{m}$
is the spherical harmonic function of order $n$ and degree $m$. It
is assumed that the pressure is band limited, so the spherical
Fourier coefficients for $n > N$ are negligible [5]. The inverse
transform in this case is
\[ x(k, r, \Omega) = \sum_{n=-N}^{N} \sum_{m=-n}^{n} x_{nm}(k, r) Y_{n}^{m*}(\Omega). \] (6)

The spherical Fourier transform coefficients of a unit-
amplitude plane-wave in the frequency domain are given by
[6]:
\[ p_{nm}(k, r) = b_n(kr) Y_{n}^{m*}(\Psi_i), \] (7)
where $k$ is the wave-number, and $b_n$ is a function that depends
on the array configuration; for an open sphere it is given by:
\[ b_n(kr) = 4\pi i^n j_n(kr) \] (8)
where $j_n$ is the spherical Bessel function. $b_n$ for a rigid sphere
and other array configurations can be found in [7]. Now, the
cpherical Fourier transform of the pressure composed of $L$ co-
erherent plane-waves, each with different phase and amplitude,
can be written as:
\[ x_{nm}(k, r) = \sum_{l=1}^{L} a_l(k) \cdot p_{nm}(k, \Psi_l) \cdot s(k) + v_{nm}(k) \]
\[ = \sum_{l=1}^{L} a_l(k) \cdot b_n(kr) \cdot Y_{n}^{m*}(\Psi_l) \cdot s(k) + v_{nm}(k), \] (9)
where $a_l(k)$ contains the amplitude and phase of the $l$th wave
at wave number $k$, $s(k)$ is the unknown source signal, and
$v_{nm}(k)$ is the spherical Fourier coefficients of the noise. The
summation of Eq. (9) can be written in a matrix form:
\[ x_{nm} = b_n \cdot \mathbf{Y}_{nm} H \cdot \mathbf{a} \cdot s + v_{nm}, \] (10)
where $\mathbf{a} = [a_1, a_2, \cdots, a_L]^T$ and
$\mathbf{Y}_{nm} = [Y_{n}^{m}(\Psi_1), Y_{n}^{m}(\Psi_2), \cdots, Y_{n}^{m}(\Psi_L)]^T$. The depen-
dence on $k$ and $r$ have been omitted for simplicity.
4. DEREVERBERATION

Signal reflections from the room boundaries result in superposition of coherent signals at each microphone. The decomposition of the coherent waves is a difficult problem. This section presents a method for decomposition of the coherent waves, based on the knowledge of the waves DOA. A method for DOA estimation for broadband coherent waves using a spherical array that may be used is the Frequency-smoothing MUSIC, presented in [3]. The proposed method for the dereverberation is performed in the spherical harmonics domain.

First, Eq. (10) is divided by \( b_n \):
\[
y_{nm} = \frac{x_{nm}}{b_n} = Y_{nm} H \cdot a \cdot s + \tilde{v}_{nm},
\]
(11)
where \( \tilde{v}_{nm} = \frac{v_{nm}}{b_n} \). This operation cancels the dependence on the array configuration. Note that \( b_n \) vanishes at some frequencies, but this can be solved by using certain array configurations [7].

Now, a vector containing all the coefficients is defined:
\[
y_{nm} = Y H \cdot a \cdot s + \tilde{v}_{nm},
\]
(12)
where
\[
y_{nm} = \begin{bmatrix} y_{00}, y_{1(-1)}, y_{10}, y_{11}, \ldots, y_{NN} \end{bmatrix}^T
\]
is \((N + 1)^2 \times 1\) vector,
\[
Y = \begin{bmatrix} Y_{00}, Y_{1(-1)}, Y_{10}, Y_{11}, \ldots, Y_{NN} \end{bmatrix}
\]
is the \( L \times (N + 1)^2 \) spherical harmonics matrix, and
\[
\tilde{v}_{nm} = \begin{bmatrix} \tilde{v}_{00}, \tilde{v}_{1(-1)}, \tilde{v}_{10}, \tilde{v}_{11}, \ldots, \tilde{v}_{NN} \end{bmatrix}^T
\]
is \((N + 1)^2 \times 1\) vector. Based on the DOA estimation, an estimated \( Y \) matrix can be formulated. Then Eq. (12) can be multiplied by the pseudo-inverse of the estimated spherical harmonics matrix:
\[
y = (\tilde{Y}^H)^+ \cdot Y^H \cdot a \cdot s + v
\]
\[
\approx a \cdot (s + v),
\]
(16)
where \( \tilde{Y} \) is the estimated spherical harmonics matrix, and \( v = (\tilde{Y}^H)^+ \cdot \tilde{v}_{nm} \). A necessary condition in order to obtain \((\tilde{Y}^H)^+\) is that the number of spherical harmonics is equal to or greater than the number of the plane waves:
\[
(N + 1)^2 \geq L,
\]
(17)
and that each wave comes from a different direction. Recall that \( L \) represents the most significant reflections in the sound field. The accuracy of the equality in (16) depends on the DOA estimation.

The presented algorithm exploits the useful model formulation in the spherical harmonics domain where the effect of array configuration can be separated from the array response, and the incident directions of the plane waves can be readily estimated. In order to employ the algorithm the incident directions have to be estimated, which is usually simpler than blind room impulse response estimation as required for the LCMV filter in the reverberant model [1].

The latter term in Eq. (16) represents the decomposition. Each element of vector \( y \) represents a single reflection. Now, noise minimization can be performed due to information redundancy in the multiple coherent signals.

5. NOISE REDUCTION

The output of the dereverberation stage (Eq. 16) is appropriate to the spatio-temporal model as presented in section 2 due to the relation between the signals which depends on \( a_l(k) \). Note that the noise components in Eq. (16) are correlated with each other due to the derverberation stage. The LCMV filter here is designed to achieve noise reduction with distortionless response constraint on the direct sound. Note that when \(|a_l|\) in Eq. (16) is frequency independent and \( \angle a_l \) is linear, the LCMV with the anechoic model can be used to reduce the noise.

In order to implement the LCMV filter with the spatio-temporal model [1], consider \( L \) filters \( h_l \), of length \( L_h \). the noise reduction filter \( h \) of length \( L \times L_h \) in the time domain is:
\[
h = [h_1^T, h_2^T, \ldots, h_L^T]^T.
\]
(18)
The filter output power is
\[
h^T R_{yy} h = h^T R_{xx} h + h^T R_{vv} h,
\]
(19)
where \( R_{xx} = E[x(n)x^T(n)] \) is the correlation matrix of the signal
\[
x(n) = [x_1^T(n), x_2^T(n), \ldots, x_L^T(n)]^T.
\]
(20)
Substituting Eq. (4) in Eq. (19) gives
\[
h^T R_{yy} h = h^T W R_{x_1x_1} W^T h + h^T R_{vv} h,
\]
(21)
where \( R_{x_1x_1} \) and
\[
W = [I_{L_h \times L_h}, W_2^T, \ldots, W_L^T]^T
\]
is a matrix of size \( L \times L_h \times L_h \).

The optimization problem is now:
\[
\min_h h^T R_{yy} h \quad \text{subject to } W^T h = u,
\]
(23)
where
\[
u = [1, 0, 0, \ldots, 0]^T
\]
(24)
is a vector of length \( L_h \) whose first component is equal to 1 while all others are zero. Under that constraint the output power becomes:
\[
h^T R_{yy} h = \sigma^2_x + h^T R_{vv} h,
\]
(25)
where $\sigma_{x_1}^2$ is the variance of $x_1(n)$. Eq. (25) shows that it is possible to recover $x_1(n)$ undistorted while reducing the noise. The optimal solution for the optimization problem in (23) is:

$$h = R_{yy}^{-1}W(W^TR_{yy}W)^{-1}u,$$

(26)

where $W$ can be estimated by:

$$W_1 = (R_{y_1y_1} - R_{v_1v_1})(R_{y_1y_1} - R_{v_1v_1})^{-1}.$$

(27)

The statistics of the noise signals can be estimated during silences, where it is assumed that the noise is stationary.

6. SIMULATION STUDY

The method was analyzed by a simulation study. The image method [8] was used to simulate room impulse responses, in a room of dimensions $15.6 \times 9.1 \times 5.2$ m$^3$, the distance from the speaker to the array is 4.58 m, and the reverberation time is 0.5 sec. The room impulse response is presented in Figure 1(a).

A speech signal of duration 1 sec was convolved with the room impulse responses to generate array input. The algorithm was performed with the model composed of only 50 reflections ($L = 50$). A rigid spherical array of the order of $N = 8$, having 162 microphones was employed in order to acquire the signal, and the noise reduction filter length was $L_h = 30$. The implementation of the method did not include the DOA estimation stage (the directions of the first 50 reflections were assumed to be known). White noise producing an SNR of 10 dB was added to the array input. The impulse response from the source to the array output is presented in Figure 1(b). It can be seen that the method achieved significant dereverberation. Note that the additive noise at the array output was artificially eliminated from the figure to show more clearly dereverberation performance. Note that the residual reverberation that results from the model mismatch of limited plane-waves appears as noise in the impulse response. The impulse response at the array output using the LCMV method with the reverberant model, which assumed known room impulse responses is presented in Figure 1(c). In this case complete dereverberation is achieved.

The proposed method achieved 4 dB additional noise reduction, while the LCMV method achieved no additional noise reduction. This can be explained by the relatively short filter length of $L_h = 30$ which had a more significant effect on the LCMV method.

7. CONCLUSION

This paper presented an acoustic model in the spherical harmonics domain. The model is used to achieve dereverberation by plane-wave decomposition, based on DOA estimation. The separated waves are filtered to minimize noise in the direct sound signal. Method analysis shows significant dereverberation and noise reduction.

8. REFERENCES


