SPARSE RECOVERY METHOD FOR DEREVERBERATION

Nicolas Epain, Tahereh Noohi, Craig T. Jin

Computing and Audio Research Laboratory (CARLab)
School of Electrical and Information Engineering
The University of Sydney, NSW 2006, Australia

ABSTRACT

In this paper, we present our contribution to the REverberant Voice Enhancement and Recognition Benchmark (REVERB) challenge. We propose a multichannel dereverberation algorithm which enhances a target speech signal in a reverberant environment. The algorithm consists of two main steps. First, the directive component of the sound field is extracted from the microphone signals. Second, sparse recovery is employed to beamformed the extracted directive signals towards the target speaker. Experiments were carried out on the evaluation dataset of the challenge. The results of these experiments indicate that the method effectively de-reverberates the signals.

***Index Terms*—** Dereverberation, microphone array, sparse recovery, beamforming, subspace methods.

1. INTRODUCTION

Reverberation is one of the major challenges in acoustic signal processing problems. It degrades speech quality and speech intelligibility, and the performance of automatic speech recognition (ASR). The problem of speech dereverberation has received a lot of attention from the 1970s until now [1] and a number of systems have been developed. Dereverberation techniques can be grouped into the following three categories [2]. (1) Inverse filtering techniques: These methods design inverse filters to compensate the reverberation by estimating the room impulse responses (RIRs) blindly using the microphone signals [2]. (2) Beamforming techniques: These techniques estimate the Direction of Arrival (DOA) of the target speech signal and reflections, so as to form a beam in the direction of the target signal and direct nulls at reverberation signals [3]. (3) Model-based techniques: These techniques rely on the availability of parametric models that reflect the underlying speech process and acoustic systems. The parameters of model are estimated from the microphone signals, and then used to reconstruct the source signal [4, 5].

In this paper, we present our contribution to the Speech Enhancement task (SE) of the REVERB Challenge. Our approach can be described as an adaptive beamforming technique and is comparable to Asano's subspace method [6]. First, we separate the microphone signals into a diffuse and a direct component. The technique used for this step is that of [7]: the microphone signals are projected onto a subspace which corresponds to the directive part of the sound field. Second, the obtained directive signals are further beamformed towards the direction of the target. In order to calculate the beamforming weights, the signals are decomposed as a series of plane waves using a sparse-recovery algorithm. As demonstrated in our recent work [2, 7–10], sparse recovery can be used to localise multiple sound sources simultaneously. Note that our sparse recovery approach is based on the assumption that the sound field consists of a few dominant directional components. Therefore, it is generally not robust to the presence of noise or reverberation. The use of this sparse recovery approach is made possible here because it is applied to the directive component of the sound field.

Note that sparse recovery is not necessarily incompatible with reverberation, as shown in recent work by Asaei and colleagues [11]. Despite the apparent complexity of the reverberant sound field, sparse recovery methods may be employed for speech enhancement because the sound field originates from a few sound sources in the typical scenario. However, by contrast with the method described in this paper, one must then model how reverberation affects a source located in a particular location of the room. In [11], for instance, an image-source model is employed.

The structure of the paper is as follows. Section 2 begins by briefly describing the model of speech signals in the reverberant environment. The direct / diffuse separation method is presented in Section 2.1. Section 2.2 describes the sparse recovery approach to de-reverberate the microphone signals. The simulation results are described in Section 3 and conclusions are drawn in Section 4.

2. METHODOLOGY

We begin by introducing a model for the speech signal in the reverberant room. In the time-frequency domain, we express the L signals recorded with the microphone array as the combination of Q plane waves incoming from every direction in the horizontal plane. In other words, we apply the so-called narrowband approximation [12] and express the vector, $\mathbf{x}(m,f)$, of the microphone Short Term Fourier Transform (STFT) samples for time step m and frequency bin f as:

$$\mathbf{x}(m,f) = \mathbf{A}(f)\,\mathbf{s}(m,f)\,, \tag{1}$$

where

$$\mathbf{x}(m,f) = [x_1(m,f), x_2(m,f), \ldots, x_L(m,f)]^\mathsf{T}\;, \tag{2}$$

$\mathbf{s}(m,f)$ is the vector of the Q plane wave signals:

$$\mathbf{s}(m,f) = [s_1(m,f), s_2(m,f), \ldots, s_Q(m,f)]^\mathsf{T}\;, \tag{3}$$

$\mathbf{A}(f)$ is the matrix of the frequency responses from the plane waves, incoming from azimuthal directions θ_1, θ_2, ..., θ_Q, to the microphones:

$$\begin{aligned}\mathbf{A}(f) &= [\mathbf{a}(\theta_1,f), \mathbf{a}(\theta_2,f), \ldots, \mathbf{a}(\theta_Q,f)]\;,\\ \mathbf{a}(\theta_q,f) &= [a_1(\theta_q,f), \ldots, a_L(\theta_q,f)]^\mathsf{T}\,,\end{aligned} \tag{4}$$

and $(.)^\mathsf{T}$ denotes the matrix or vector transpose. The manifold matrix $\mathbf{A}(f)$ is calculated with the assumption that the microphone array is comprised of L ideal omnidirectional microphones distributed around a circle of radius R, *i.e.*:

$$a_l(\theta_q, f) = e^{i\frac{2\pi f}{c} R\cos(\theta_q - \theta_l)} \tag{5}$$

where c is the speed of sound, assumed to be 343 $\mathrm{m.s}^{-1}$. Note that the plane-wave directions are distributed evenly around the circle. As well, the number of plane-wave directions is chosen to be very large (Q=180 in this work), such that the target non-reverberant signal can be modelled as the contribution of only one or two plane-waves. On the contrary reverberant signals are modelled as a mixture of plane waves incoming from every direction. For brevity of notation we omit the frequency dependence in the following.

2.1. Direct / diffuse separation

The first step in our dereverberation method is the separation of the microphone signals into a directive and a diffuse components. This technique is similar to that described in [6] and [7]. The diffuse component of the sound field is expected to consist mostly of the late room reflections and microphone self noise, while the directive component is expected to consist of the direct sound and salient early reflections. Nevertheless, note that we make no formal hypothesis on the relative presence of early reflections in the directive component of the sound field.

2.1.1. Singular-beam domain

We begin by considering the case of a perfectly diffuse sound field. A perfectly diffuse sound field can be modelled as the sum of many plane waves, evenly distributed in space, the waveform of which have equal energies and are perfectly uncorrelated. The correlation matrix of the plane wave signals, $\mathbf{C}_s^{(\mathrm{dif})}$, is then given by:

$$\mathbf{C}_s^{(\mathrm{dif})} = \nu\,\mathbf{I}\,, \tag{6}$$

where ν is the power of the plane waves and $\mathbf{I}$ denotes the identity matrix.

In order to separate the directive and diffuse components of the sound field, we first transform the microphone signals to a domain in which the correlation matrix of the sensor signals is proportional to the identity matrix in the presence of a perfectly diffuse sound field. We refer to this domain as the *singular beam* domain, as it is related to the singular vectors of the array manifold, $\mathbf{A}$. The singular vectors of the manifold matrix may be regarded as the modes of the array, hence they are referred to as "array modes" in [13]. The vector of the singular-beam signals, $\mathbf{b}(m)$, is given by:

$$\mathbf{b}(m) = \mathbf{\Psi}^{-1}\mathbf{U}^{\mathsf{H}}\mathbf{x}(m) = \mathbf{\Phi}^{\mathsf{H}}\mathbf{s}(m)\ , \tag{7}$$

where $(\cdot)^{\mathsf{H}}$ denotes the conjugate-transpose operation and $\mathbf{U}$, $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are the matrices of singular vectors and values of matrix $\mathbf{A}$, *i.e.*:

$$\mathbf{A} = \mathbf{U}\mathbf{\Psi}\mathbf{\Phi}^{\mathsf{H}}\ . \tag{8}$$

It can be easily verified that, in the presence of a perfectly diffuse sound field, the correlation matrix of the singular-beam signals, $\mathbf{C}_b^{(\mathrm{dif})}$, is proportional to the identity:

$$\mathbf{C}_b^{(\mathrm{dif})} = \mathbf{\Phi}^{\mathsf{H}}\mathbf{C}_s^{(\mathrm{dif})}\mathbf{\Phi} = \nu\,\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Phi} = \nu\,\mathbf{I}\,. \tag{9}$$

2.1.2. Diffusivity Estimation

In the reverberant speech scenario, the sound field can be modelled as the sum of a perfectly diffuse component, the reverberation, and a perfectly directive component, the non-reverberant sound field emitted by the speaker. Hence, the relative amount of reverberation in the sound field can be estimated as the relative energy of the diffuse sound field component, which we refer to as the *diffusivity*, $\beta(m)$ [14, 15]. In order to estimate the diffusivity of the sound field, we first estimate the correlation matrix of the singular-beam signals as:

$$\mathbf{C}_b(m) = (1-\alpha_C)\,\mathbf{C}_b(m-1) + \alpha_C\,\mathbf{b}(m)\mathbf{b}^{\mathsf{H}}(m)\,, \tag{10}$$

where α_C is a forgetting factor. We then decompose $\mathbf{C}_b(m)$ in terms of its eigenvalues and eigenvectors as:

$$\mathbf{C}_b(m) = \mathbf{V}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathsf{H}}\,, \tag{11}$$

where $\mathbf{V}$ is the matrix of the eigenvectors and $\mathbf{\Sigma}$ is the diagonal matrix, the coefficients of which are the eigenvalues:

$$\begin{aligned}&\mathbf{\Sigma} = \mathrm{diag}\left([\sigma_1, \sigma_2, \sigma_3, ..., \sigma_L]\right)\\ &\text{where}\quad \sigma_1 \geq \sigma_2 \geq \sigma_3 \geq ... \geq \sigma_L\,.\end{aligned} \tag{12}$$

In the presence of a perfectly diffuse sound field, we have shown that $\mathbf{C}_b(m)$ is proportional to the identity matrix, therefore its eigenvalues are equal. On the other hand, in the presence of a perfectly directive sound field, only σ_1 is non zero. Hence, the more diffuse the sound field, the more homogeneous the eigenvalues σ_l. We therefore estimate $\beta(m)$ as:

$$\beta(m) = 1 - \frac{\gamma}{\gamma_0}\,, \tag{13}$$

where γ is the deviation of the eigenvalues σ_l from their mean, *i.e.*:

$$\gamma = \frac{1}{\langle\sigma\rangle}\sum_{l=1}^{L}|\sigma_l - \langle\sigma\rangle|\ ,\quad \text{where}\quad \langle\sigma\rangle = \frac{1}{L}\sum_{l=1}^{L}\sigma_l\,, \tag{14}$$

and γ_0 is the value of γ in the most directive case (in the presence of a single plane wave), given by:

$$\gamma_0 = 2\,(L-1)\ . \tag{15}$$

Note that, when the sound field is perfectly diffuse, $\gamma = 0$ and thus $\beta(m) = 1$. On the other hand, when the sound field is perfectly directive, $\gamma = \gamma_0$ and thus $\beta(m) = 0$.

2.1.3. Extraction of the directive signals

In the REVERB challenge scenario, we can assume that the largest eigenvalue, σ_1, corresponds to the target signal, while the other eigenvalues correspond to the reverberation. Therefore, in order to extract the directive component of the sound field, we project the singular-beam signals on the subspace defined by the first eigenvector, $\mathbf{v}_1$, of the matrix $\mathbf{C}_b$. Further, we apply a gain which depends on the diffusivity. The estimated directive signals are then given by:

$$\mathbf{b}^{(\mathrm{dir})}(m) = \left(1 - \beta^4(m)\right)\,\mathbf{v}_1\mathbf{v}_1^{\mathsf{H}}\,\mathbf{b}(m)\,. \tag{16}$$

The idea here is that, in the case where $\beta(m)$ is close to 1, the extracted signals mostly consist of undesirable reverberant noise. The term $1 - \beta^4(m)$ is then close to 0 and thus prevents noise from polluting the directive signals.

2.2. Sparse Recovery Method

In the second step of our dereverberation method, we employ sparse recovery to beamform the directive signals in the direction of the target.

2.2.1. *Sparse plane-wave decomposition*

We first decompose the directive signals in terms of the Q plane-wave directions by solving the problem:

$$\mathbf{b}(m) = \mathbf{\Phi}^{\mathsf{H}}\,\mathbf{s}(m)\ . \tag{17}$$

As there are many more directions than observed signals ($Q \gg L$), this is an underdetermined system of equations and thus it has an infinite number of solutions. In order to solve this problem, we make the assumption that the directive signals $\mathbf{b}(m)$ can be explained by a small number of plane-wave directions. We thus solve Problem (17) for the *sparsest* vector of plane-wave signals.

There is a single stationary speaker in the room in the REVERB challenge scenario, therefore the solutions to Problem (17) share the same sparsity pattern over a short time interval. This allows us to solve the following sparse recovery problem:

$$\begin{aligned}&\text{minimize} && \|\mathbf{S}(m)\|_{p,2}\\ &\text{subject to} && \mathbf{B}^{(\mathrm{dir})}(m) = \mathbf{\Phi}^{\mathsf{H}}\,\mathbf{S}(m)\ ,\end{aligned} \tag{18}$$

where $\mathbf{B}(m)$ is a $L \times T$ matrix containing the T consecutive STFT samples of the microphone signals:

$$\mathbf{B}(m) = [\mathbf{b}^{(\mathrm{dir})}(m), \mathbf{b}^{(\mathrm{dir})}(m-1), \ldots, \mathbf{b}^{(\mathrm{dir})}(m-T+1)]\ , \tag{19}$$

$\mathbf{S}(m)$ is the $Q \times T$ matrix of the plane-wave source signals:

$$\mathbf{S}(m) = [\mathbf{s}(m), \mathbf{s}(m-1), \ldots, \mathbf{s}(m-T+1)]\ ,$$

and $\|.\|_{p,2}$ denotes the $l_{p,2}$-norm and is defined as:

$$\|\mathbf{S}(m)\|_{p,2} = \left(\sum_{q=1}^{Q}\left(\sqrt{\sum_{t=0}^{T-1} s_q(m-t)^2}\right)^{p}\right)^{1/p}\ . \tag{20}$$

The $l_{p,2}$-norm promotes sparsity along the plane-wave dimension when $p \leq 1$. In this work we use $p = 0.7$.

We use the IRLS algorithm [16] to solve (18). The IRLS algorithm iterates the following two steps until convergence:

$$\begin{aligned}&1.\ \ \omega_q = \left(\sum_{t=0}^{T-1} s_q(m-t)^2 + \mu\right)^{\frac{p-2}{2}},\\ &2.\ \ \mathbf{\Omega} = \operatorname{diag}\left([\omega_1,\ \omega_2,\ \ldots,\ \omega_Q]\right)\ ,\\ &\quad\ \mathbf{S}(m) = \mathbf{\Omega}\mathbf{\Phi}\left(\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Omega}\mathbf{\Phi} + \lambda\mathbf{I}\right)^{-1}\mathbf{B}^{(\mathrm{dir})}(m)\ ,\end{aligned} \tag{21}$$

where λ is a regularization factor and μ is a small-valued parameter that is incorporated into the weights, ω_i, to ensure that $\mathbf{\Omega}$ is defined when $s_i = 0$ [17]. Once the solver has converged, we obtain the unmixing matrix, $\mathbf{D}(m)$, which decomposes the directive signals into Q plane-wave components:

$$\mathbf{D}(m) = \mathbf{\Omega_0}\mathbf{\Phi}\left(\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Omega_0}\mathbf{\Phi} + \lambda\mathbf{I}\right)^{-1}, \tag{22}$$

where Ω_0 is the diagonal matrix of the weights obtained after convergence.

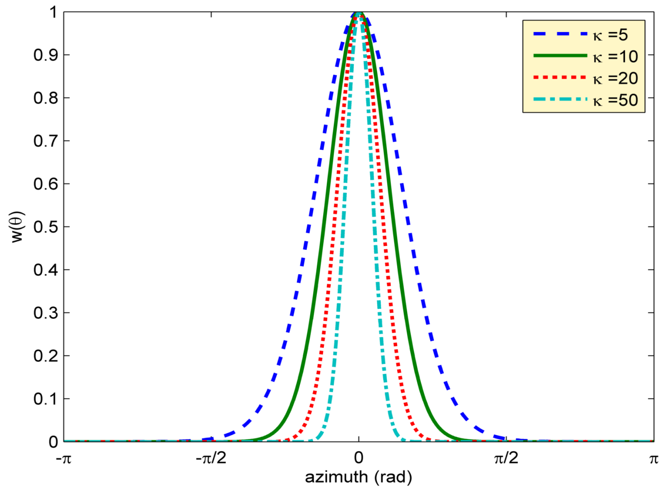

Fig. 1. This figure shows the spatial window with different κ in the case where $\theta_0 = 0$.

2.2.2. *Source Direction Estimation*

In order to estimate the target source direction we first calculate the energy corresponding to every plane-wave direction. Note that, at this stage, the result of the sparse plane-wave decomposition could be used. Instead, we arbitrarily chose to estimate the energy of the plane-wave signals directly from the directive singular-beam signals. The energy, $e_q(m,f)$, corresponding to direction θ_q, is given by:

$$e_q(m,f) = \left|\Phi_q^{\star}(f)\,\mathbf{b}^{(\mathrm{dir})}(m,f)\right|^2, \tag{23}$$

where $(.)^{\star}$ denotes the conjugate and Φ_q is the q-th column of the matrix $\mathbf{\Phi}^{\mathsf{H}}$. We then estimate the vector of the target source cartesian coordinates, $\mathbf{v}_0 = [x_0, y_0]^{\mathsf{T}}$, as the energy-weighted average plane-wave direction, *i.e.*:

$$\mathbf{v}_0 = \sum_{f=f_{\text{low}}}^{f_{\text{high}}} \mathbf{e}(m,f)\,\mathbf{v}\ , \tag{24}$$

where:

$$\begin{aligned}\mathbf{v} &= [v_1, \ldots, v_q, \ldots, v_Q]^{\mathsf{T}}\ ,\\ v_q &= [\cos(\theta_q), \sin(\theta_q)]^{\mathsf{T}}\ ,\\ \mathbf{e}(m,f) &= [e_1(m,f), \ldots, e_q(m,f), \ldots, e_Q(m,f)]\ ,\end{aligned}$$

and f_{low} and f_{high} denote the lower and upper cutoff frequency indices, respectively.

Lastly, we estimate the target source direction, $\theta_0(m)$, as:

$$\theta_0(m) = (1-\alpha_\theta)\,\theta_0(m-1) + \alpha_\theta \arctan(y_0/x_0)\ \ , \tag{25}$$

where α_θ is the forgetting factor for the estimation of the target direction.

2.2.3. *Beamforming*

In order to further isolate the target signal, we steer a beam towards the target direction. The beamforming weights, $\mathbf{d}(m)$, for the m^{th} time window is given by:

$$\mathbf{d}(m) = (1-\alpha_d)\mathbf{d}(m-1) + \alpha_d\,\mathbf{w}^{\mathsf{T}}\,\mathbf{D}(m), \tag{26}$$

Table 1. Parameters used in the speech enhancement task.

Common	Sampling rate=16 kHz Window=hanning Frame length=128 points (8 ms) Frame shift=frame length / 2 $Q = 180$, L=8
Sparse Recovery	Number of iterations = 10 $p = 0.7$, $\alpha_d = 0.2$, $T = 4$ $\mu = \max_i(\frac{\sum_{l=0}^{T-1} s_i(m+l,f)^2}{Q})$ $\lambda = \frac{r}{1-r}\langle \mathrm{diag}\left(\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Omega}\mathbf{\Phi}\right)\rangle$ $r = \beta^{0.15}(m)$
Spatial window	$\kappa = 50$
Source direction estimation	$f_{\text{low}} = 500$ Hz, $f_{\text{high}} = 3500$ Hz $\alpha_\theta = 0.2$
Correlation matrix ($\mathbf{C}$)	$\alpha_C = 0.01$

where $\mathbf{w}$ is a spatial window, the maximum of which is located in the target source direction. We calculate the weights of the spatial window as the von Mises distribution with parameter κ:

$$\mathbf{w} = [e^{\kappa[\cos(\theta_1-\theta_0)-1]}, e^{\kappa[\cos(\theta_2-\theta_0)-1]}, \ldots, e^{\kappa\left[\cos(\theta_Q-\theta_0)-1\right]}]^{\mathsf{T}} . \tag{27}$$

Figure 1 shows the spatial windows for different κ values. Lastly, the target signal, $s_0(m, f)$ is estimated as:

$$s_0(m) = \mathbf{d}(m)\,\mathbf{b}^{(\mathrm{dir})}(m) \ . \tag{28}$$

The inverse STFT is then applied to $s_0(m)$ to recover the time-domain estimated source signal.

3. SIMULATION

The REVERB challenge consists of two tasks: one for speech enhancement (SE) and the other for automatic speech recognition (ASR). For the SE task, 1ch, 2ch, and/or 8ch speech enhancement algorithms can be used. We addressed the speech enhancement (SE) task with 8 channels. The challenge dataset consists of real recordings (RealData) and simulated data (SimData). SimData contains a set of reverberant speech signals simulated by convolving clean speech signals with measured room impulse responses (RIRs) and subsequently adding measured noise signals. It simulates 6 different reverberation conditions: 3 rooms with different volumes (small, medium and large size), 2 types of distances between a speaker and a microphone array (near=50cm and far=200cm). RealData contains a set of real recordings made in a reverberant meeting room which is different from the ones used for SimData. It contains 2 reverberation conditions: 1 room, 2 types of distances between a speaker and a microphone array (near= 100cm and far= 250cm)[18].

For each test utterance, the following quality measures were calculated (please refer to [19] for more details):

- Cepstrum distance (CD)
- Log likelihood ratio (LLR)
- Frequency-weighted segmental SNR (FWSegSNR)
- Speech-to-reverberation modulation energy ratio (SRMR)
- Perceptual Evaluation of Speech Quality (PESQ).

Note that the CD, LLR, FWSeqSNR and PESQ require a reference signal, which is available only in the case of the simulated set of data. Therefore, the performance of the algorithm in the case of the measured dataset (RealData) was evaluated using the SRMR only. Smaller values of CD and LLR and larger values of FWSeqSNR, SRMR, and PESQ are assumed to indicate better speech quality.

The parameters used in our algorithm are summarized in Table 1. The parameter values were adjusted by applying the method on the development set data and listening to the result. In particular, the parameter r was set to $\beta^{0.15}(m)$ with the objective of removing as much reverberation as possible. In other words, this setting leads to a rather aggressive processing which may affect the quality of speech, as perceived by a human listener. Note that the rather complicated definition of the regularization parameter, λ, arises from the normalisation of this parameter with regard to the energy of the directive signals. This aspect will be explained in detail in a forthcoming publication.

Table 2 compares the values of the various speech quality measures obtained with: 1) the signals enhanced using the method presented in this paper (enh) ; and 2) the original reverberant signals (org). The values presented in the table correspond to the quality scores averaged over every utterance for each test condition of the challenge. In the case of the simulated data, the proposed method improved the speech quality in terms of the CD, FW-SeqSNR and PESQ. However, it degraded the LLR and did not improve the SRMR significantly. In the case of the measured data, the SRMR scores of the enhanced signals are significantly higher than that of the original signals, which indicates that the reverberation was greatly reduced.

4. CONCLUSION

In this paper, we propose a dereverberation algorithm which consists of two steps. First the microphone signals are processed such that only the directive component of the sound field remains. Second, the directive signals are decomposed in plane-wave components using sparse-recovery. The plane-wave signals are then beamformed in the target source direction. Experiments were carried out on the evaluation dataset of the challenge. The proposed algorithm enhanced the signals in terms of the CD, FW-SeqSNR and PESQ.

References

[1] K. Lebart and J.M. Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acustica*, vol. 87, no. 3, pp. 359–366, May 2001.

[2] P.K.T. Wu, N. Epain, and C.T. Jin, "A dereverberation algorithm for spherical microphone arrays using compressed sensing techniques," in *Proceedings of the 2012 ICASSP*, Kyoto, Japan, March 2012.

[3] M.S. Brandstein and D.B. Ward, "Microphone arrays: signal processing techniques and applications," *Springer*, 2001.

[4] M. Jeub, M. Schafer, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Transactions on Audio, Speech, and Language Processing,*, vol. 18, pp. 1732 – 1745, 2010.

[5] T. Yoshioka, H. Kameoka, T. Nakatani, and H.G. Okuno, "Statistical models for speech dereverberation," *WASPAA*, 2009.

Table 2. Results of the proposed algorithm for the Speech Enhancement task.

	SimData														RealData					
	Room1				Room2				Room3				Ave.		Room1				Ave.	
	Near		Far		Near		Far		Near		Far		-		Near		Far		-	
	org	enh	org	enh	org	enh	org	enh	org	enh	org	enh	org	enh	org	enh	org	enh	org	enh
Cepstral distance (dB)	1.99	1.45	2.67	2.42	4.63	2.84	5.21	3.89	4.38	2.85	4.96	4.14	3.97	2.93	-	-	-	-	-	-
SRMR	4.50	3.61	4.58	4.03	3.74	3.69	2.97	4.69	3.57	3.82	2.73	3.89	3.68	3.96	3.17	5.48	3.19	5.66	3.18	5.57
Log likelihood ratio	0.35	0.58	0.38	0.69	0.49	0.46	0.75	0.84	0.65	0.72	0.84	0.96	0.58	0.71	-	-	-	-	-	-
Freq-weighted seg.SNR(dB)	8.12	9.87	6.68	8.34	3.35	8.77	1.04	6.55	2.27	7.37	0.24	4.64	3.62	7.59	-	-	-	-	-	-
PESQ	2.14	3.42	1.61	2.23	1.40	2.30	1.19	1.43	1.37	2.14	1.17	1.37	1.48	2.15	-	-	-	-	-	-

[6] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 497–506, 2000.

[7] N. Epain and C.T. Jin, "Super-resolution sound field imaging with sub-space pre-processing," in *Proceedings of the 2013 ICASSP*, Vancouver, Canada, May 2013.

[8] A. Wabnitz, N. Epain, A. McEwan, and C.T. Jin, "Upscaling ambisonic sound scenes using compressed sensing techniques," in *Proceedings of the 2011 WASPAA*, New Paltz, NY, USA, October 2011.

[9] A. Wabnitz, N. Epain, and C.T. Jin, "A frequency-domain algorithm to upscale ambisonic sound scenes," in *Proceedings of the 2012 ICASSP*, Kyoto, Japan, March 2012.

[10] T. Noohi, N. Epain, and C.T. Jin, "Direction of arrival estimation for spherical microphone arrays by combination of independent component analysis and sparse recovery," in *Proceedings of the 2013 ICASSP*, Vancouver, Canada, May 2013.

[11] A. Asaei, M. Golbabaee, H. Bourlard, and V. Cevher, "Structured sparsity models for reverberant speech separation," *EEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, vol. 22, no. 3, pp. 620–633, 2014.

[12] M. Kowalski, E. Vincent, and R. Gribonval, "Beyond the narrowband approximation: Wideband convex methods for underdetermined reverberant audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1818–1829, 2010.

[13] F. Kaiser, F. Zotter, and Fazi. F., "Array modes of compact rigid microphone arrays with unconventional shapes," in *Proceedings of the AIA-DAGA 2013 Conference on Acoustics*, Merano, Italy, March 2013.

[14] Hodgson, "On measures to increase sound?field diffuseness and the applicability of diffuse?field theory," *J. Acoust. Soc. Am.*, vol. 95, pp. 3651–3653, 1994.

[15] B.N. Gover, J.G. Ryan, and M.R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound fields," *J. Acoust. Soc. Am.*, vol. 112, pp. 1980–1991, 2002.

[16] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proceedings of the 2008 ICASSP*, Las Vegas, USA, May 2008.

[17] I. Daubechies, R. DeVore, M. Fornasier, and Güntürk C.S., "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

[18] "The REVERB challenge," `http://www.reverb2014.dereverberation.com/data`.

[19] K. Kinoshita, M. Delcroix, T. Yoshioka, T. Nakatani, E. Habets, R. Haeb-Umbach, V. Leutnant, A. Sehr, W. Kellermann, R. Maas, S. Gannot, and B. Raj, "The REVERB challenge: A common evaluation framework for dereverberation and recognition of reverberant speech," in *Proceedings of the 2013 WASPAA*, New Paltz, NY, USA, October 2013.