# SPARSE RECOVERY METHOD FOR DEREVERBERATION

*N. Epain, T. Noohi and C.T. Jin*

CARLab, The University of Sydney, Australia

http://www.ee.usyd.edu.au/carlab/index.htm

## 1. INTRODUCTION

In this paper, we present a multichannel dereverberation algorithm which enhances a target speech signal in a reverberant environment. The algorithm consists of two main steps. First, the directional component of the sound field is extracted from the microphone signals. Second, sparse recovery is employed to beamform the directional signals towards the target speaker.

It has been shown that Sparse Recovery can be employed to provide a sharp image of the incoming sound field, *i.e.*, locate multiple sound sources and estimate their strength. In this work we attempt to take advantage of this capability to derive beamforming weights that isolate the target speaker from the ambient noise.

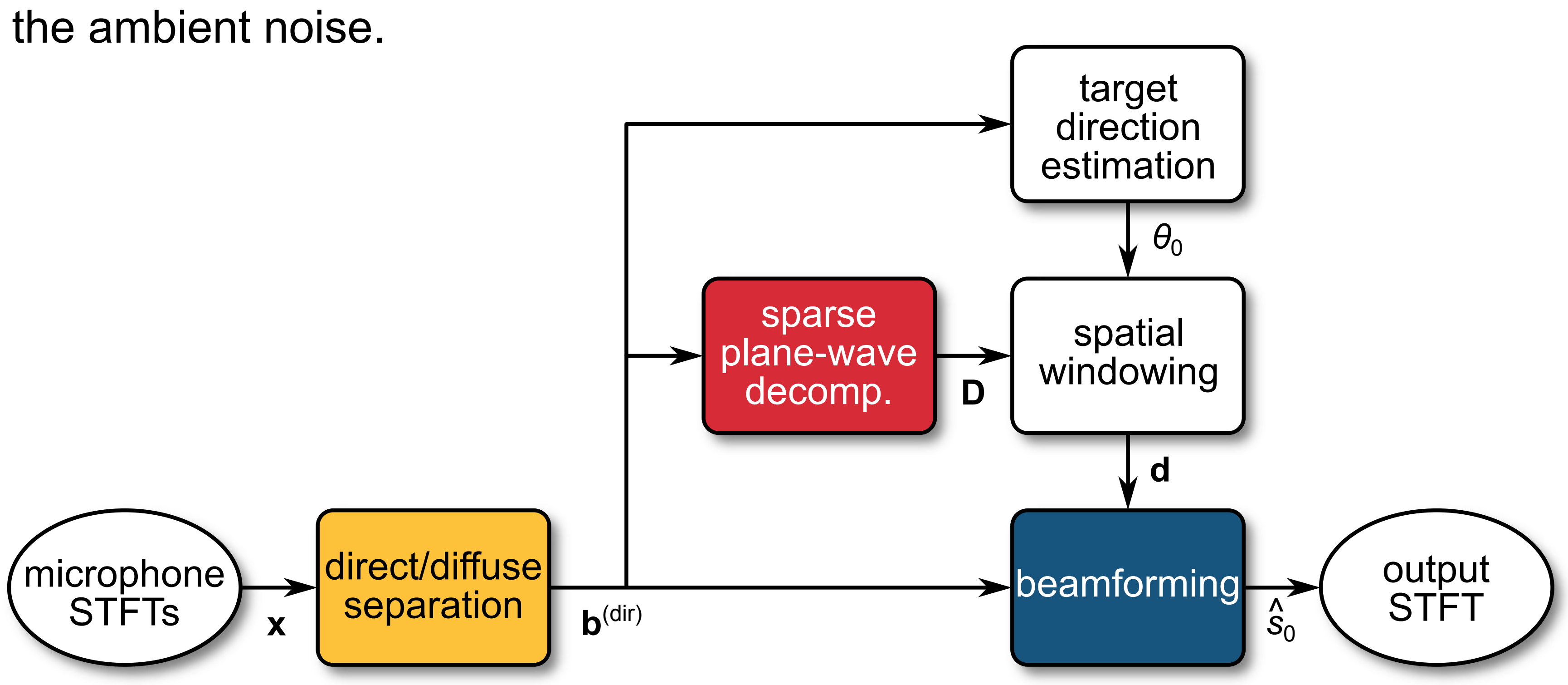
target direction estimation
$\theta_0$
sparse plane-wave decomp.
D
spatial windowing
d
microphone STFTs
x
direct/diffuse separation
$\mathbf{b}^{(\mathrm{dir})}$
beamforming
$\hat{s}_0$
output STFT

***Fig. 1:*** *Flow diagram of our sparse-recovery dereverberation method.*

## 2. DIRECT/DIFFUSE SEPARATION

The first step in our method is the separation of the microphone signals in two components: a) a **diffuse** component, which mostly consists of noise and reverberation ; and b) a **directional** component, which mostly consists of the target speaker, interfering sources and strong specular reflections in the room.

Throughout this work we model the microphone signals as a combination of *Q* plane-wave components incoming from “every” direction in the horizontal plane. The vector, **x**(*m*), of the microphone Short-Term Fourier Transforms (STFTs) for time step *m* and frequency *f* is given by:

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

where **s**(*m*) denotes the vector of the plane-wave STFTs and **A** is the array manifold matrix. Note that we omit the frequency dependence for brevity.

### Singular-Beam Domain

We first transform the microphone signals to a domain in which, in the presence of a perfectly diffuse sound field, the sensor correlation matrix is proportional to the identity. We refer to this domain as the **singular-beam** domain as the transformation consists in projecting the microphone signals on the singular vectors of the manifold matrix. The vector of the singular-beam signals, b(m), is given by:

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

where **U**, **Ψ** and **Φ** are the matrices of the singular vectors and values of **A**, i.e., **A** = **UΨΦ^H**. It can easily be verified that, in the presence of a perfectly diffuse sound field, the correlation matrix of **b**(*m*) is proportional to the identity.

### Diffusivity Estimation

We use the property of the singular-beam domain described above to estimate the relative amount of ambient noise in the sensor signals. Intuitively, we can measure **how diffuse** the signals are by estimating **how close to the identity** the correlation matrix is. In practice, we estimate the sound field diffusivity by looking at the distribution of the correlation matrix eigenvalues, $\sigma_l$. The diffusivity, $\beta$(m) is estimated as:

$$\beta(m) = 1 - \frac{\gamma}{\gamma_0}\,, \qquad \text{where} \qquad \gamma = \frac{1}{\langle\sigma\rangle}\sum_{l=1}^{L}|\sigma_l - \langle\sigma\rangle|\ ,$$

and $\gamma_0$ is the value of $\gamma$ in the most directional case (one plane-wave source).

### Extraction of the Directional Signals

The dominant source at a given instant corresponds to the largest eigenvalue of the correlation matrix. Therefore, in order to extract the directional component of the signals **b**(*m*), we project them on the first eigenvector, $\mathbf{v}_1$. Further, we apply a gain which decreases when the diffusivity value increases. The directional signals are given by:

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

## 3. SPARSE PLANE-WAVE DECOMPOSITION

We perform a sparse plane-wave decomposition of the directional signals in order to locate the various sources that are active at a given time. This decomposition is done by solving the following optimisation problem

$$\text{minimize} \quad \|\mathbf{S}(m)\|_{p,2} \quad \text{subject to} \quad \mathbf{B}^{(\mathrm{dir})}(m) = \mathbf{\Phi}^{\mathsf{H}}\,\mathbf{S}(m)\,, \qquad (1)$$

where **S**(*m*) and $\mathbf{B}^{(\mathrm{dir})}(m)$ are matrices that concatenate, respectively, plane-wave and directional signal vectors over *T* consecutive time windows, *p* is comprised between 0 and 1 and $||.||_{p,2}$ denotes the $l_{p,2}$ norm, given by:

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

We solve Problem (1) by applying the Iteratively-Reweighted Least-Square (IRLS) algorithm. The result is a de-mixing matrix, **D**(*m*), which decomposes the directional signals into *Q* plane-wave. The de-mixing matrix is given by:

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

where $\mathbf{\Omega}_0$ is a (diagonal) matrix of weights, only a few of which are non-zero, and λ is a regularisation parameter.

## 4. BEAMFORMING

### Target Direction Estimation

In order to estimate the direction of the target speaker, we estimate the energy incoming from every plane-wave direction and calculate the energy-weighted average direction. In other words the vector, $\mathbf{v}_0$, of the target source cartesian coordinates is given by:

$$\mathbf{v}_0 = \sum_{f=f_{\mathrm{low}}}^{f_{\mathrm{high}}}\sum_{q=1}^{Q} e_q(m,f)\,\mathbf{v}_q\ ,$$

Where $\mathbf{v}_q$ is the vector pointing to the *q*-th plane wave direction, $\mathbf{e}_q(m,f)$ is the energy for direction *q* at time step *m* and frequency *f* and $f_{\mathrm{low}}$ and $f_{\mathrm{high}}$ denote the lowest and highest frequency considered.

### Spatial Windowing and Beamforming

Lastly, the beamforming weights to be applied to the directional signals are given by:

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

where $\alpha$ is a forgetting factor and **w** is a spatial window (values of a Von Mises distribution centred around the target direction). The output STFT is then:

$$\hat{s}_0(m) = \mathbf{d}(m)^{\mathsf{T}}\,\mathbf{b}^{(\mathrm{dir})}(m)\ .$$

## 5. RESULTS

We used the technique described above to address the REVERB Challenge’s Speech Enhancement task with eight microphone channels. The speech quality scores obtained with the challenge’s evaluation dataset are shown in Table 1 below. In the case of the simulated reverberant signals, the segmented SNR and PESQ were improved significantly, while the other measures were only slightly improved (CD, SRMR) or slightly degraded (LLR). In the case of the measured (real) data, the SRMR was significantly higher, which indicates that reverberation was greatly reduced.

| | **Simulated data (average)** | | **Real data (average)** | |
|---|---|---|---|---|
| **Measures** | original | enhanced | original | enhanced |
| Cepstral distance (CD) | 3.97 | 2.93 | - | - |
| SRMR | 3.68 | 3.96 | 3.18 | 5.57 |
| Log likelihood ratio (LLR) | 0.58 | 0.71 | - | - |
| Freq. weighted seg. SNR | 3.62 | 7.59 | - | - |
| PESQ | 1.48 | 2.15 | - | - |

***Table 1:*** *Quality scores obtained with the REVERB Challenge evaluation dataset, averaged over every room and distance condition.*