WP3: Signal Separation & Broadband Distributed Beamforming

WP Leaders:Wenwu Wang and John McWhirter
Researcher: Swati Chandna

Introduction

Extracting signals of interest and suppression of interference from corrupted sensor measurements remain fundamental challenges in many networked battlespace applications. Mathematically,

$x(t) = A(t) \star s(t) + n(t),$

where ★ denotes the convolution operator, s denotes the signal of interest, x denotes the recorded mixture measurements, A denotes the mixing matrix, and n denotes the noise vector.

Objectives

The objective of this work package is to develop robust and low-complexity algorithms for source separation (SS) and broadband distributed beamforming. We aim to achieve the above by developing --

- algorithms based on Polynomial Matrix Eigenvalue Decomposition (PEVD) techniques – this has the advantage of only requiring second-order statistics thereby reducing the computational load associated with higher order statistics
- Sparse representations and T-F masking techniques robust to noise/incomplete measurements for underdetermined SS.

Current Focus

Let N_m denote the number of mixtures, and, N_s denote the number of sources. The case $N_s > N_m$ characterizes the underdetermined SS problem.

Since the mixing matrix is an N_m x N_s matrix, traditional matrix inversion demixing techniques are not applicable in the underdetermined case.
Techniques for underdetermined CBSS are based on the fact that speech signals satisfy the W-disjoint orthogonality (WDO) condition i.e., given speech signals $s_1(t)$ and $s_2(t)$

$$s_1(\omega,\tau)s_2(\omega,\tau) = 0 \text{ for all } (\omega,\tau),$$

i.e. signals have a disjoint support in the time-frequency domain.

Techniques relying on the sparsity of speech signals in the time-frequency domain proceed by assigning either a binary or probabilistic weight to the dominant source at each time-frequency point. The matrix of such weights at each T-F point is known as the T-F mask.

Our Aim: To improve the performance of model based expectation-maximization SS methods utilizing interaural and mixing vector cues, e.g. [Mandel et. al. 2010], [Sawada et. al. 2007], and the combined method [Atiyeh et. al. 2011] for highly reverberant mixtures

Proposed Method

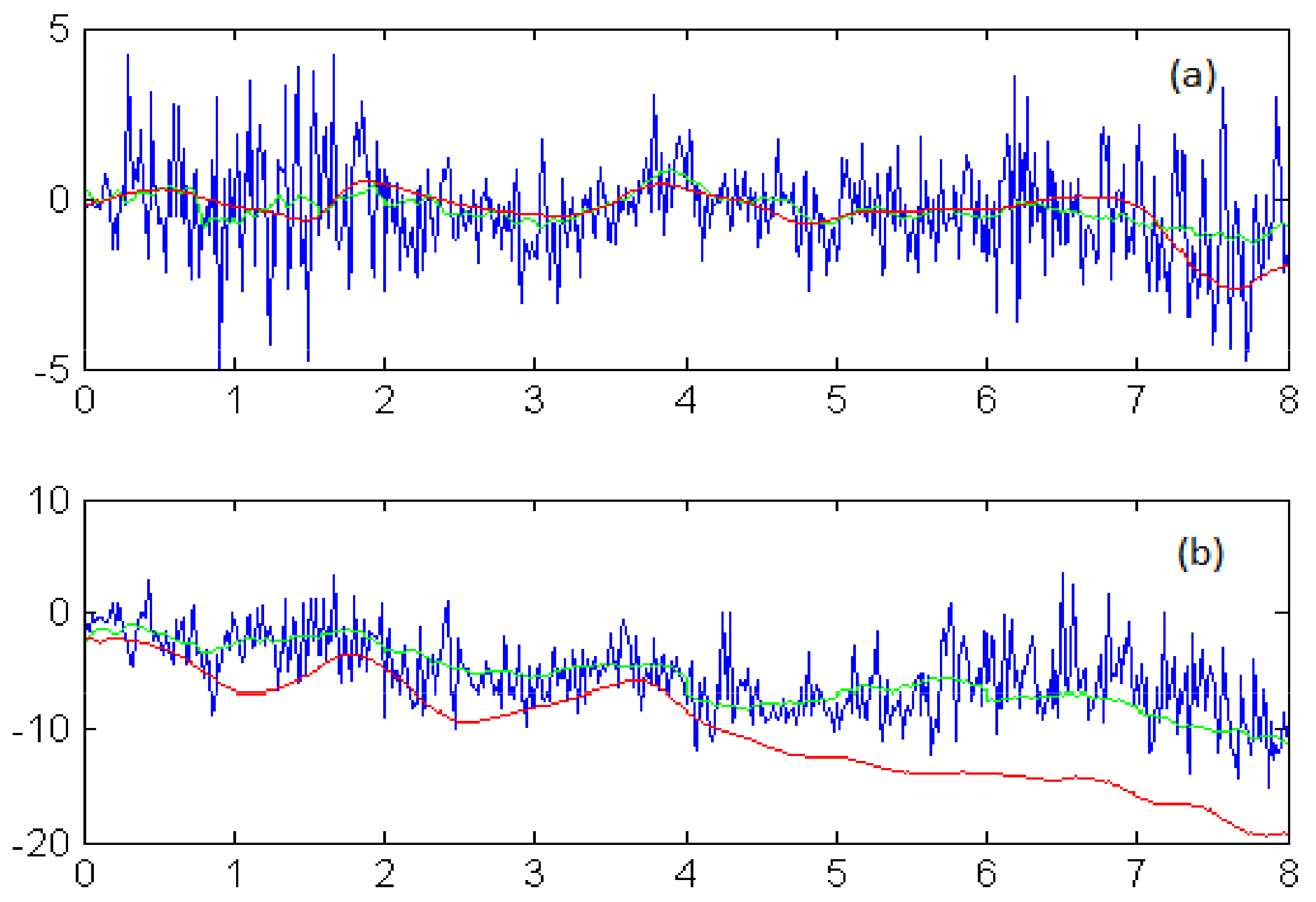
- Given $\underline{x}(t) = [x_1(t),x_2(t)]$, the ratio $X_1(\omega,\tau)/X_2(\omega,\tau)$ is the Interaural Spectrogram denoted by $IS(\omega,\tau)$.
- $IS(\omega,\tau)$ is expressible in terms of the ILD and IPD.
- A probabilistic T-F mask is obtained as a by product of the EM algorithm that is used to obtain max likelihood estimates of unknown parameters of the assumed ILD and IPD models.
- We propose the idea of bootstrap averaging to improve parameter estimates which in turn determine the T-F mask –
  - ○ Generate bootstrap samples $\underline{x}_1(t)$,....., $\underline{x}_B(t)$ using an appropriate simulation methodology.
  - ○ Obtain the ILD parameter estimates $\alpha_j(\omega,\tau)$ and IPD parameter estimates $\phi_j(\omega,\tau)$ for each of the $\underline{x}_j(t)$ using Mandel's algorithm.
  - ○ Use the averaged parameter estimates, i.e.

$$\alpha_{avg}(\omega,\tau) = <\alpha_j(\omega,\tau)>/B \text{ and } \phi_{avg}(\omega,\tau) = <\phi_j(\omega,\tau)>/B$$

to reconstruct the target source.

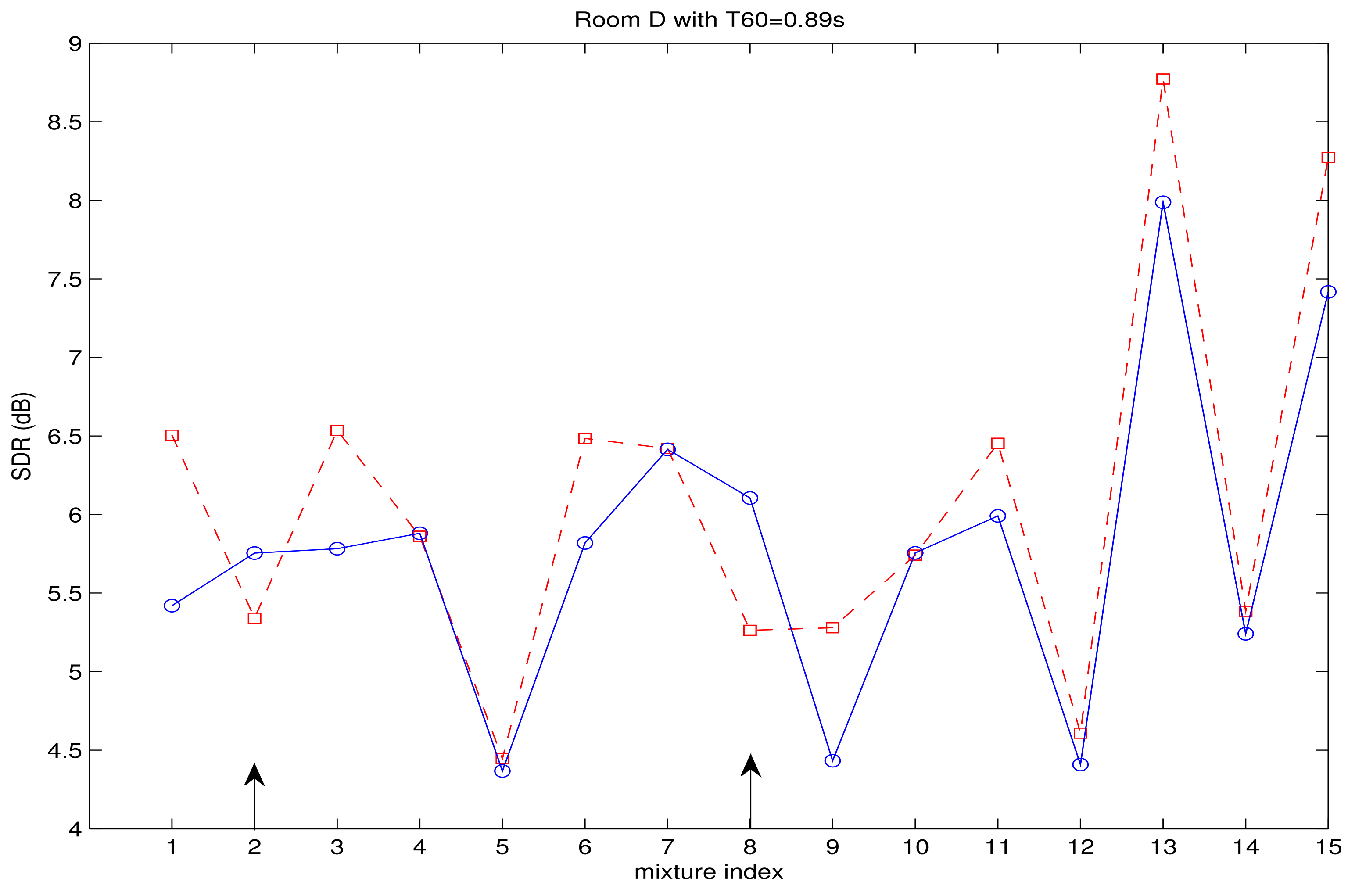
Results

A comparison of the averaged ILD mean estimates (green) with the ground-truth direct response estimates (red) as well as the original estimates (blue) from Mandel's algorithm is shown below:



Clearly, the bootstrap averaged ILD model parameter estimates (green) follow the ground-truth direct response estimates (red) very closely.
We use the bootstrap averaged T-F mask which corresponds to the smoothed parameter estimates of the ILD, IPD and/or mixing vector cue models to separate the target source from interference.

Signal-distortion-ratio (SDR) of the target source is used to compare the performance of our proposed method (red) with the hybrid method of [Atiyeh et. al.11] (blue) which combines interaural cues of [Mandel et al.] with the mixing vector cue of [Sawada et al.]. The figure below compares the SDR of the target source for a set of 15 two-channel stereo mixtures recorded in a room with reverberation time of 0.89s when the interference source is placed at ϴ=30°.



Remarks

a) On an average, we see an improvement of 0.31 decibels in the SDR over 15 mixtures.
b) We note that the performance of our method based on bootstrap averaging relies heavily on:
   I. The simulation technique used to simulate the highly non-stationary mixture vector, and
   II. Variance of the parameter estimates with respect to input mixture vector $\underline{x}(t)$.
c) Our simulation technique mimics the spectral density estimate obtained from smaller stationary time blocks. We note that choosing blocks of the same size can miss information when the spectrum changes quickly .
d) Future work: To further improve our results by performing an automatic statistical analysis of bivariate non-stationary time series vector based on the Best Bias Algorithm to select stationary blocks.

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) and Dstl