University Defence Research Collaboration (UDRC)
Signal Processing in a Networked Battlespace
L_WP3: Signal Separation & Broadband Distributed Beamforming

WP Leaders: Wenwu Wang, University of Surrey; John McWhirter, Cardiff University
Researcher: Swati Chandna, University of Surrey

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 $\star$ 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

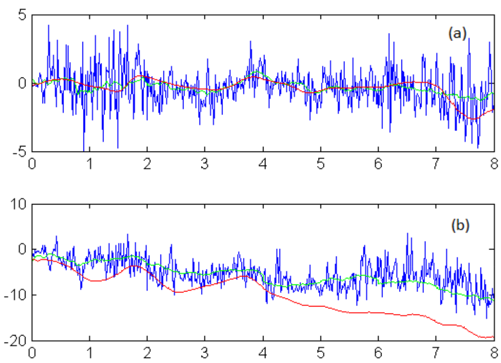
- Given $\underline{x}(t) = [x_1(t),x_2(t)]$, the ratio $X_1(\omega,\tau)/X_2(\omega,\tau)$ is the Interaural Spectrogram denoted by $IS(\omega,\tau)$.
- $IS(\omega,\tau)$ is expressible in terms of the ILD and IPD.
- A probabilistic T-F mask is obtained as a by product of the EM algorithm that is used 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 $\underline{x}_1(t),\ldots\ldots, \underline{x}_B(t)$ 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.

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

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 above plots suggest that the averaged parameter estimates may lead to better separation performance.

Our aim is to incorporate the smoothed parameter estimates of the ILD, IPD and/or mixing vector cue models appropriately in the model based SS algorithms.

Signal-distortion-ratio (SDR) and Perceptual Evaluation of Speech Quality (PESQ) will be used as measures to compare the performance of our proposed method with the hybrid method of [Atiyeh et. al.11] which combines interaural cues of [Mandel et al.] with the mixing vector cue of [Sawada et al.].

Conclusion

a) The idea of bootstrap averaging (bagging) is used to improve T-F mask estimates obtained from model based EM SS methods.
b) The performance of bootstrap averaging heavily relies on:
   I. The simulation technique used to obtain copies of the mixture vector, and
   II. Variance of the parameter estimates with respect to different input mixture vectors $\underline{x}(t)$.
c) Since $\underline{x}(t)$ is a bivariate time series vector, with components $x_1(t)$ and $x_2(t)$ possibly correlated, it is important to recreate samples of $\underline{x}(t)$ with the correct second-order statistical structure. This is achieved using the estimated power spectral density of $\underline{x}(t)$.

References

M.I. Mandel et. al., "Modelbased expectation-maximization source separation and localization," *IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, pp. 382–394, 2010.*

H. Sawada et. al., "A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2007, pp. 139–142.*

A. Alinaghi, W. Wang, and P.J.B Jackson, "Integrating binaural cues and blind source separation method for separating reverberant speech mixtures," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011.*

S. Chandna and A. Walden, "Simulation methodology for inference on physical parameters of complex vector-valued signals," *IEEE Transactions on Signal Processing,* vol. 61, pp. 5260–5269, 2013.

UDRC [dstl] EPSRC Engineering and Physical Sciences Research Council

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