Harvesting Partially-disjoint Time-Frequency Information for Improving DUET (Degenerate Unmixing Estimation Technique)

Yudong He[1,2] He Wang[2] Qifeng Chen[1] Richard H.Y. So[1,2]

[1]The Hong Kong University of Science and Technology

[2]HKUST-Shenzhen Research Institute

IEEE ICASSP, May 2022

DUET

The degenerate unmixing estimation technique (DUET)(Rickard, 2007) [1] is one of the most efficient blind source separation (BSS) algorithms to separate any number of sources with two microphones. It is a **binary masking** based algorithm.

Reference
[1] Scott Rickard, *The DUET Blind Source Separation Algorithm,* pp. 217-241, Springer Netherlands, Dordrecht, 2007

Binary masking

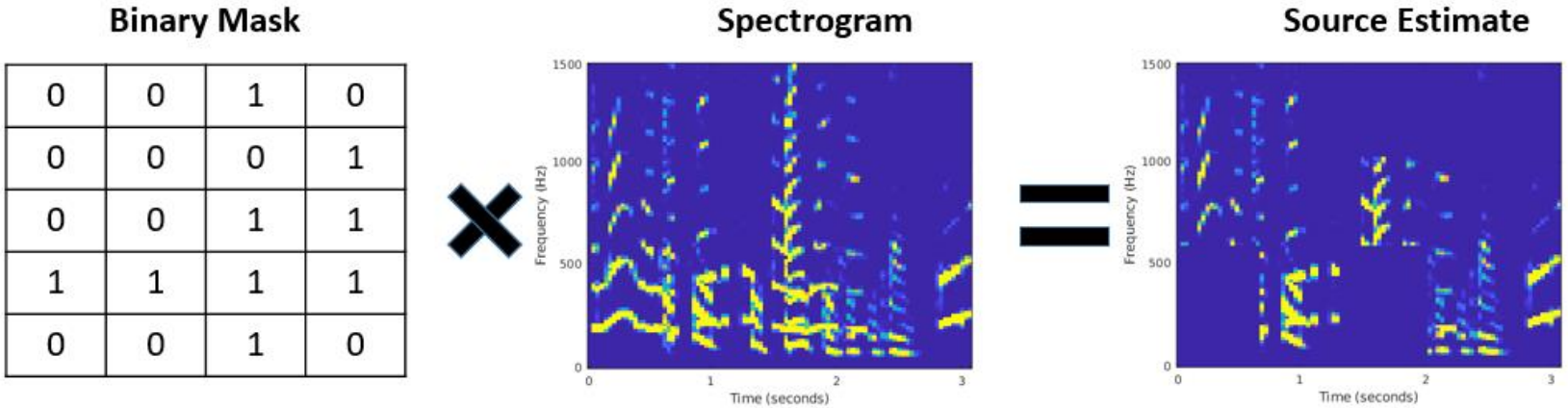


Figure: An exaggerated illustration of a binary masking process

But, binary masking has a problem

Research gap - problem of binary masking in duet

Binary masking always pollutes the target signal if the target signal and interfering signal are overlapped in both time and frequency domain. This problem severely limits the performance of DUET.

Let me illustrate this problem

Illustration of interference pollution of DUET

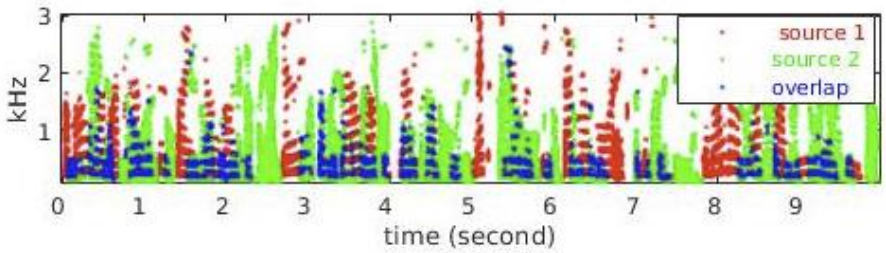


Figure: Spectrogram of two source signals

A binary masking based BSS algorithm, e.g., DUET, works well in the red and green region but fails in the blue region.

In speech signal, the blue region is common.

Challenge tackled in this paper

To remove the interference pollution meanwhile keep the computing efficiency of DUET.

Our solution to remove interference pollution in DUET

- Instead of binary masking, we propose soft filtering - a weighted summation of the observed signals.
- We use linear spatial filters, e.g., minimum variance distortionless response (MVDR)(Capon 1969) [2], as weighting coefficients.
- Spectrograms of source signals will not completely overlap. It allows us to construct spatial linear filters utilizing information embedded in the partially-disjoint region, i.e., the green and red region.
- How do we identify the partially-disjoint region?

[2] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of* the IEEE, vol, 57. no. 8, pp. 1408-1418, 1969

Partially-disjoint region identification

- Calculate mixing parameters for each time-frequency bin (local mixing parameters).
- Global mixing parameters are estimated by the peaks of the histogram of local mixing parameters. If there are N sources, there will be N sets of global parameters.
- A time-frequency bin is only contributed by the i-th source if the local mixing parameter is close enough to the i-th global mixing parameter.

Recap our solution - replace binary mask with soft filtering using linear spatial filters

- Instead of binary masking, we propose soft filtering - a weighted summation of the observed signals.
- We use linear spatial filters, e.g., minimum variance distortionless response (MVDR)(Capon 1969) [2], as weighting coefficients.
- Spectrograms of source signals will not completely overlap. It allows us to construct spatial linear filters utilizing information embedded in the partially-disjoint region, i.e., the green and red region.

[2] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of* the IEEE, vol, 57. no. 8, pp. 1408-1418, 1969

Main findings

- Our proposed soft filtering removes interference pollution >> outperforms DUET (using binary masking).
- Our soft filtering ALSO outperforms >> other mainstream BSS algorithms not using binary masking (IVA [3], ILRMA [4], MULTINMF [5], FULLRANK [6])
- Contrary to popular belief, our proposed spatial linear filter, interference suppression response (ISR), obtains a much better performance than the famous minimum variance distortionless response (MVDR).

Reference
see paper [3] (Kim, Eltoft, and Lee 2006); [4] (Kitamura et al. 2016); [5] (Ozerov and Fevotte 2010); [6] (Duong et al. 2010)

Third finding - ISR > MVDR

- Minimum variance distortionless response (MVDR) attempts to minimize the filtered interfering signal meanwhile introduce no distortion on the target signal.
- MVDR may be defiled by the estimation error of spatial parameters, i.e., the direction of the target source.
- Interference suppression response (ISR) has the same minimization objective as MVDR. It will not be defiled by the estimation error although it allows distortion on the target signal.

Two mics with two sources

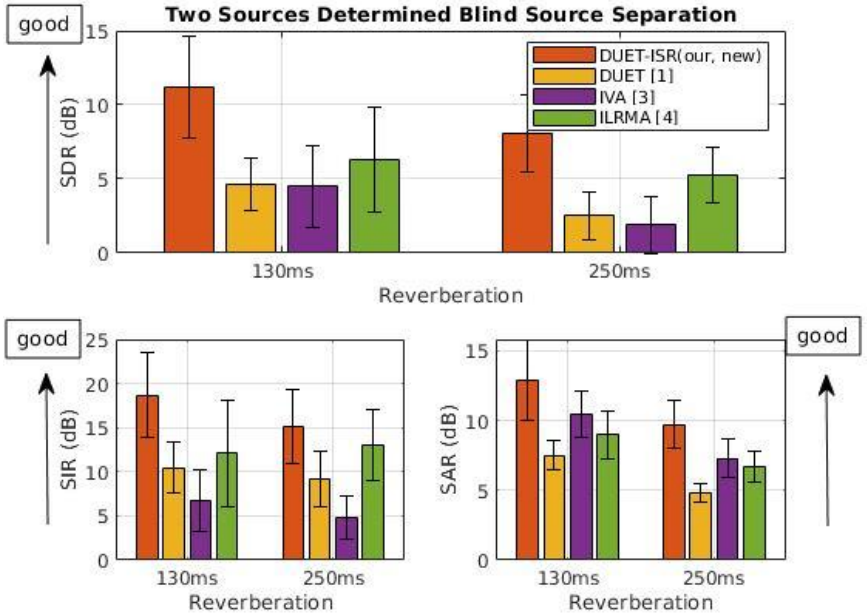


Figure: Average BSS performance comparison for two mixtures of two sources in the presence of 130 ms and 250 ms reverberations.

Reference

[1] (Rickard, 2007); [3] (Kim, Eltoft, and Lee 2006); [4] (Kitamura et al. 2016)

Two mics with three sources

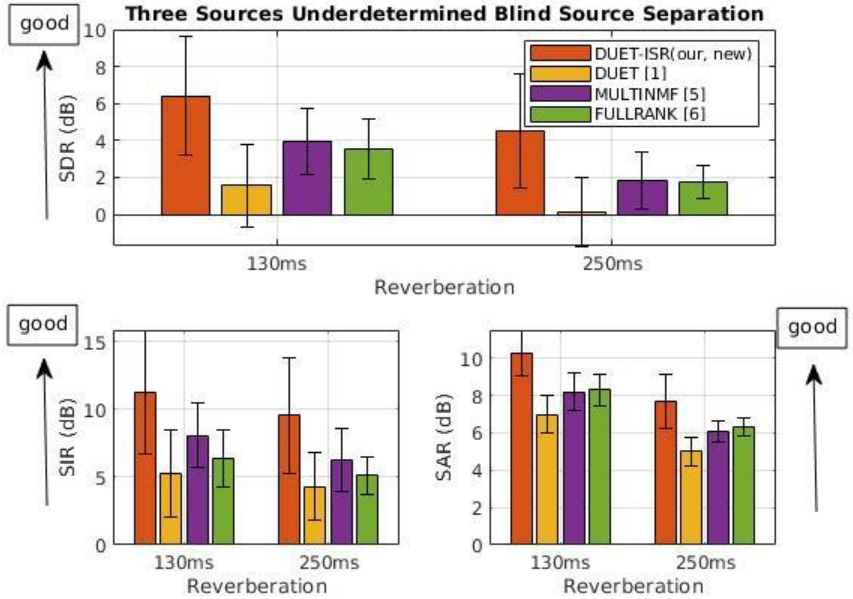
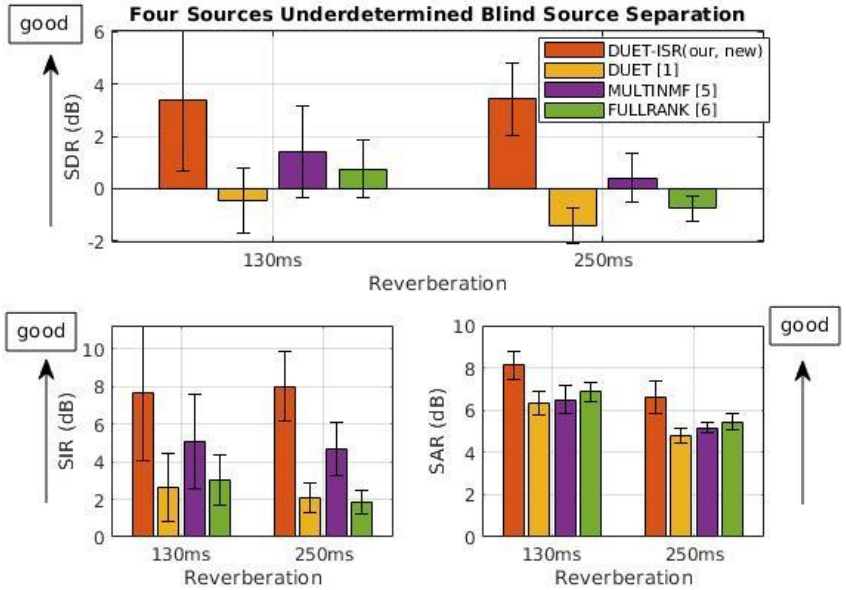


Figure: Average BSS performance comparison for two mixtures of three sources in the presence of 130 ms and 250 ms reverberations.

Reference

[1] (Rickard, 2007); [5] (Ozerov and Fevotte 2010); [6] (Duong et al. 2010)

Two mics with four sources (very challenging)



Demo:

mix of 4 src

duet

the proposed algorithm

Figure: Average BSS performance comparison for two mixtures of four sources in the presence of 130 ms and 250 ms reverberations.

Reference

[1] (Rickard, 2007); [5] (Ozerov and Fevotte 2010); [6] (Duong et al. 2010)

ISR >> the well known MVDR

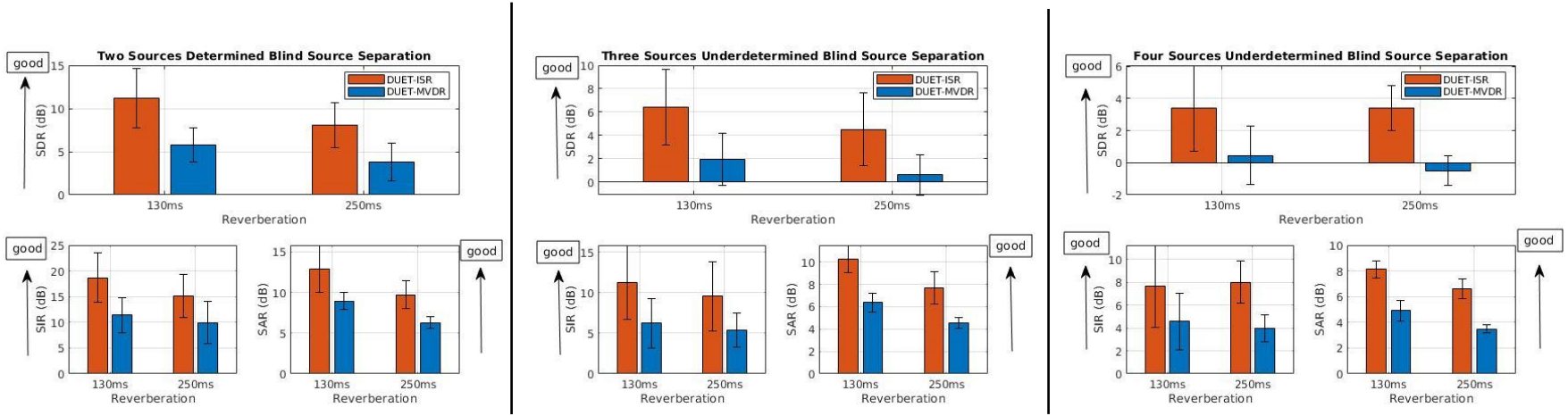


Figure: Average BSS performance comparison between the soft filtering with the proposed ISR (red) and with the well known MVDR [2] (blue) in the presence of 130 ms and 250 ms reverberations.

Reference

[2] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of* the IEEE, vol, 57. no. 8, pp. 1408-1418, 1969

Time consumption

	2 sources	3 sources	4 sources
DUET-MVDR	0.74 s	0.84 s	0.96 s
DUET-ISR	0.73 s	1.00 s	1.01 s
DUET [1]	0.1 s	0.14 s	0.18 s
IVA [3]	3.40 s	\	\
ILRMA [4]	8.39 s	\	\
MULTINMF [5]	\	32.85 s	43.54 s
FULLRANK [6]	\	575.32 s	585.39 s

Demo is provided

Audio examples can be found on the web page
https://ydcnanhe.github.io/demo-icassp2022/

Thank you for listening!

A bird-view of HKUST campus