Paper #: 4907

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Harvesting Partially-disjoint Time-Frequency Information for Improving 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

Introduction

The degenerate unmixing estimation technique (DUET)[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 (see Figure 1) based algorithm. However, DUET always pollutes the target signal if the target signal and interfering signal are overlapped in both time and frequency domain and this is the downside of binary masking (see Figure 2). The challenge we tackled in this paper is to remove the interference pollution meanwhile keep the computing efficiency of DUET. In comparison with the conventional DUET, our method achieved an impressive **improvement greater than 5 dB** in the source-to-interference ratio (SIR) and **2 to 5 dB improvement** in the source-to-distortion ratio (SDR), respectively. Findings are substantiated by unmixing simulation using live-recorded mixture signals from up to four sources. Audio examples can be found on the web page: **https://ydcnanhe.github.io/demo-icassp2022/**

Method

1. Instead of binary masking, we propose soft filtering – a weighted summation of the observed signals.
2. We use linear spatial filters, e.g., minimum variance distortionless response (MVDR) [2], as weighting coefficients.
3. 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.

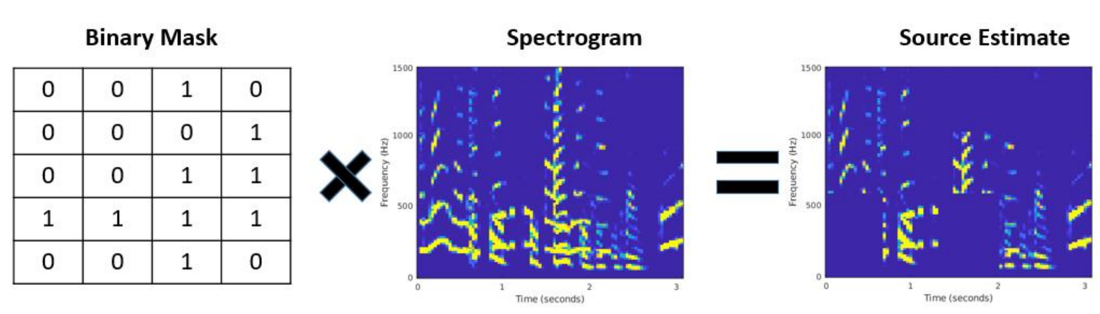

Figure 1. An exaggerated illustration of a binary masking process

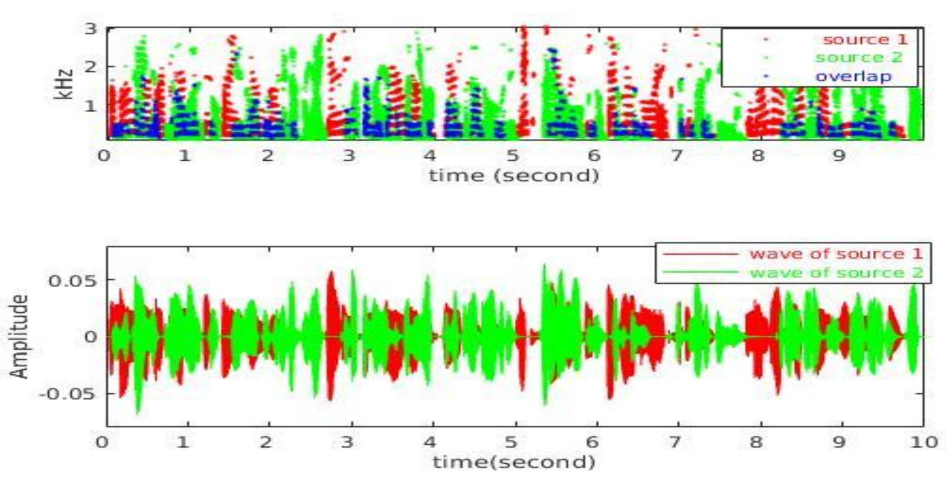

Figure 2. Spectrogram of two source signals to illustrate the problem of binary masking

Main findings

1. Our proposed soft filtering algorithm significantly outperforms DUET by successfully removing interfering pollution.
2. Our proposed soft filtering also outperforms other mainstream BSS algorithms, like IVA[3], ILRMA[4], MULTINMF[5], FULLRANK[6], which are not based on binary masking.
3. Contrary to popular belief, our proposed spatial linear filter, interference suppression response (ISR), obtained a much better performance than the famous minimum variance distortionless response (MVDR)

Results

The proposed algorithm was evaluated in terms of the standard audio source separation metrics: source-to-distortion ratio (SDR), source-to-interference ratio (SIR), and source-to-artifacts ratio (SAR) [7]. The larger the three metrics, the better the performance. The mixtures used in all separation tasks were from the underdetermined speech and music mixtures in SiSEC2011 [8]. Mixture signals from up to four sources are two-channel live recordings obtained for a meeting room of 130 ms or 250 ms reverberation time. Benchmark algorithms were DUET, independent vector analysis (IVA) [3], and state-of-the-art methods, independent low-rank matrix analysis (ILRMA) [4], multichannel non-negative matrix factorization (MULTINMF) [5], and full-rank model (FULLRANK) [6]. Results are shown in Figure 4-7.

Conclusion & Future work

We significantly improved DUET by replacing simple masking by soft filtering using multiple linear spatial filters optimized for individual sources utilizing information embedded in the TF points occupied by a single source. For future work, we can

1. develop a real-time version of the proposed algorithm.
2. combine MVDR and ISR to construct a new linear spatial filter to reduce the distortion on the target signal meanwhile improve robust to the error of spatial parameters estimation.

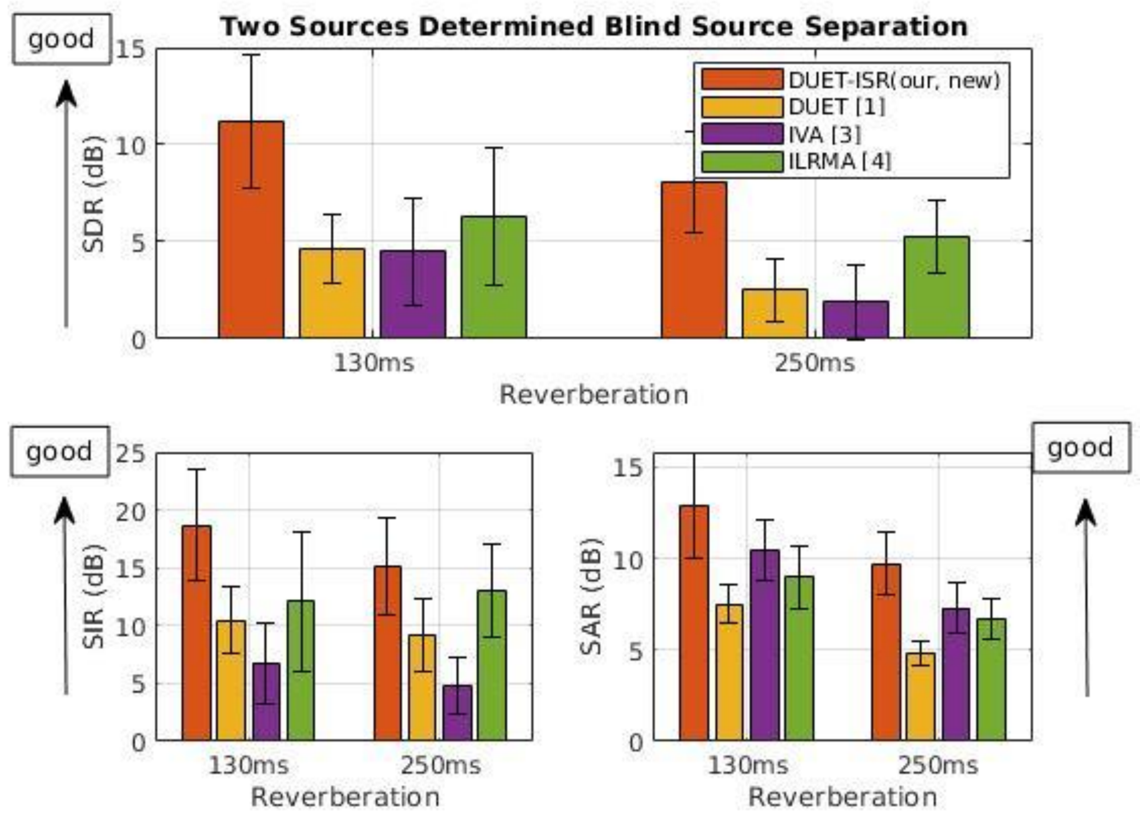

Figure 3. BSS performance for two mixtures of two sources

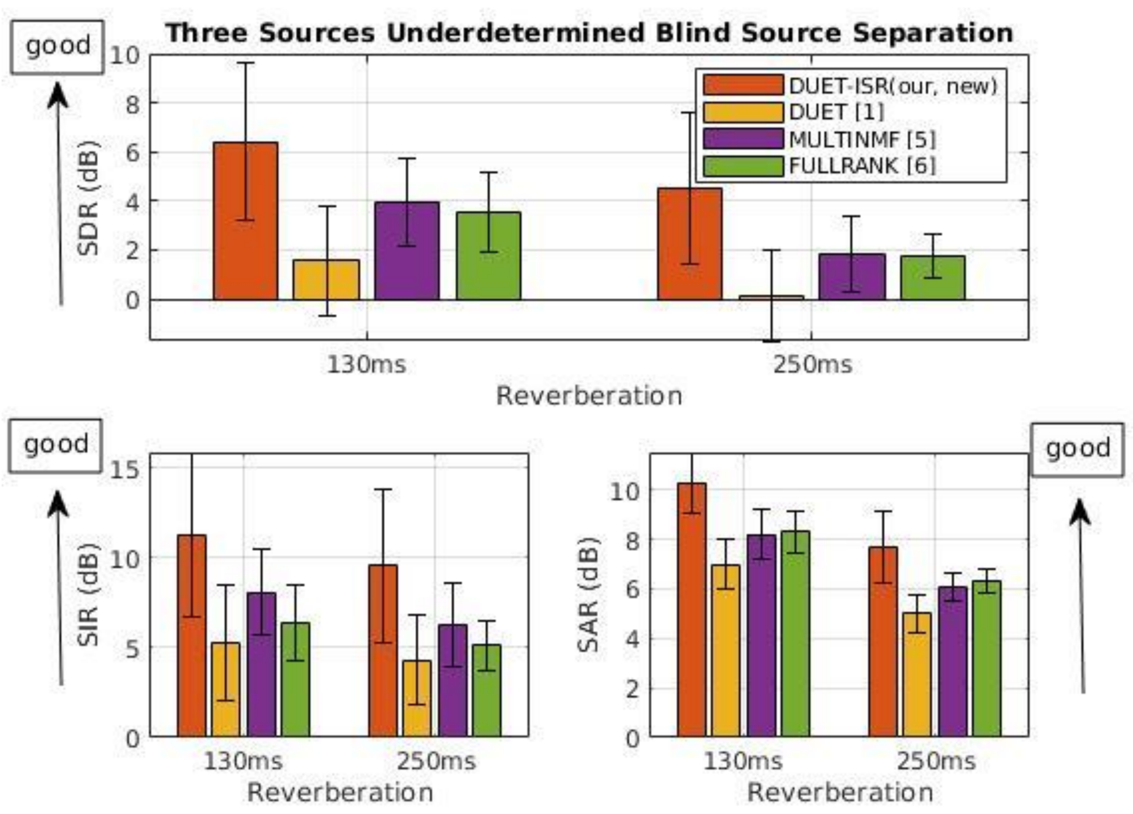

Figure 4. BSS performance for two mixtures of three sources

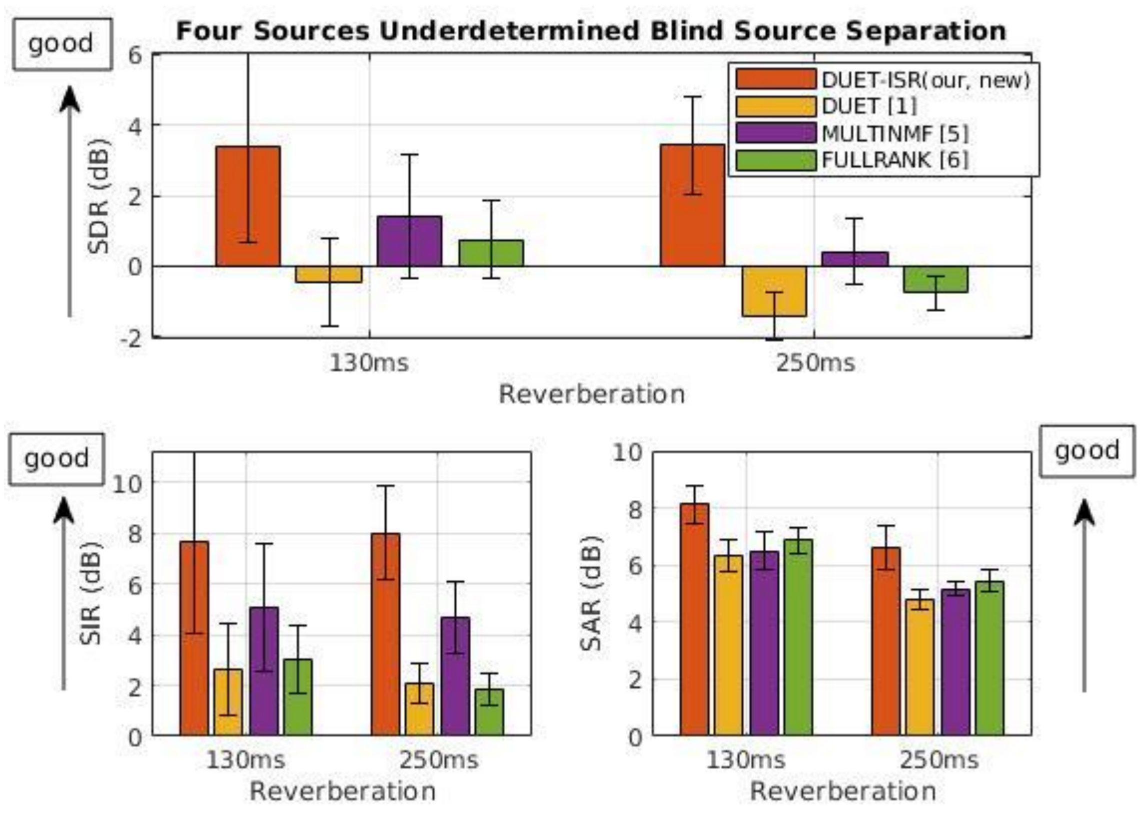

Figure 5. BSS performance for two mixtures of four sources. This is a very challenging case.

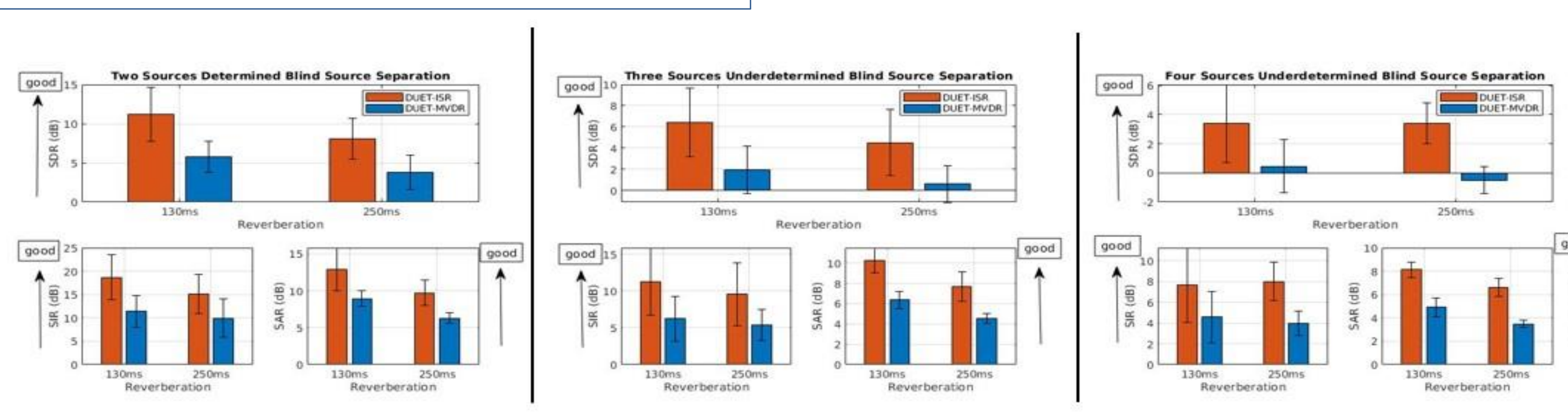

Figure 6. BSS performance comparison between soft filtering using our proposed ISR (red) and with the well known MVDR [2] (blue).

Contact

Yudong He
The Hong Kong University of Science and Technology
Email: yhebh@connect.ust.hk

Acknowledgement

This work was partially supported by the Foshan-HKUST Project (FSUST20-SR107D and FSUST20-HKUST08D) and the National Natural Science Foundation of China (NSFC) (No. 31900709).

References

1. Scott Rickard, The DUET Blind Source Separation Algorithm,pp. 217–241, Springer Netherlands, Dordrecht,2007.
2. J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of* the IEEE, vol, 57. no. 8, pp. 1408-1418, 1969
3. Taesu Kim, Torbjorn Eltoft, and Te-Won Lee, "Independent vector analysis: An extension of ica to multivariate components," in Independent Component Analysis and Blind Signal Separation, Justinian Rosca, Deniz Erdogmus, Jose C. Principe, and Simon Haykin, Eds., Berlin,Heidelberg, 2006, pp. 165–172, Springer Berlin Heidelberg.
4. Daichi Kitamura, Nobutaka Ono, Hiroshi Sawada, Hirokazu Kameoka, and Hiroshi Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 9, pp. 1626–1641, 2016.
5. Alexey Ozerov and C edric Fevotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 3, pp.550–563, 2010.
6. Ngoc Q. K. Duong, Emmanuel Vincent, and Remi Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 7, pp. 1830–1840, 2010.
7. E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," IEEE Transactions on Audio, Speech, and Language Process- ing, vol. 14, no. 4, pp. 1462–1469, 2006.
8. Shoko Araki, Francesco Nesta, Emmanuel Vincent, Zbynˇek Koldovsḱy, Guido Nolte, Andreas Ziehe, and Alexis Benichoux, "The 2011 signal separation evaluation campaign (sisec2011): - audio source separation," in Latent Variable Analysis and Signal Separation, Fabian Theis, Andrzej Cichocki, Arie Yeredor, and Michael Zibulevsky, Eds., Berlin, Heidelberg, 2012, pp.414–422, Springer Berlin Heidelberg.