

Single Channel Source Extraction Using Adaptive Non- Negative Matrix Factorization (ADAPNMF)

Jasreen kaur¹, Shailaja Gaikwad²

P.G. Student, Department of Electronics Engineering, Terna Engineering College, Nerul, Maharashtra, India¹

Assistant Professor, Department of Electronics Engineering, Terna Engineering College, Nerul, Maharashtra, India²

ABSTRACT: In our day to day life, sound that are heard is not in seclusion but in fusion with background noise which depends on the surrounding like car noise, television noise, radio noise and crowd noise called as cocktail party effect. In order to segregate the mixed sources into underlying sound sources, a new method named adaptive non-negative matrix factorization (ADAPNMF) is proposed. In this approach, the each source is factorized as multiplication of two matrices, namely basis vector and weights. The basis vector matrix represents data sample and weight matrix represents data feature of each sound source. This methodology results in high values of signal-to-interference ratio (SIR), signal-to-artifact ratio (SAR) and signal-to-distortion ratio (SDR). In this paper, we have explained the methodology and also plotted SIR, SAR and SDR results. Furthermore, the proposed methodology is compared with Sparse REMML (SREMML) and Sparse RISRA (SRISRA) in terms of evaluating parameters. It is observed that the proposed algorithm is better than the existing approaches in few parameters.

KEYWORDS: NMF, Single channel, SIR, SDR, SAR, REMML, RISRA

I. INTRODUCTION

There exists large number of audio files which are composed of various sound sources. Our proposed methodology, adaptive non-negative matrix factorization is proposed is used for extracting single channel source. One of the major application of single channel extraction is noise cancellation. In cocktail party effect, it is very necessary to remove background noise from important source message. Another important application is music remixing. Music remixing is created by mixing different single channel sources as input. Single channel extraction is also used in audio restoration. In audio restoration the original work is restored into its original form by removing the noise imperfections. Speech recognition system also needs single source extraction of the user from background noise. The system can recognize the use only after removing the noise from the audio.

II. RELATED WORK

Previously, Nandakumar M et al. [1] separated the blind source using two of the multiplicative algorithms, i.e., Regularized Expectation Minimization Maximum Likelihood Algorithm (REMML) and Regularized Image Space Reconstruction Algorithm (RISRA) with sparseness constraint are taken to evaluate the performance of BSS. Three parameters namely, signal to distortion Ratio (SDR), signal to interference ratio (SIR) and signal to artifact ratio (SAR) are also evaluated. It was found that Sparse REMML algorithm outperforms the Sparse RISRA algorithm in NMF based single channel speech and music separation in terms of SDR, SIR, and SAR measures.

Wang et al. [2] has used method of single channel speech separation with the aid of a user-generated exemplar source that is recorded from a microphone. The proposed algorithm involves combining of general speaker-independent (SI) features with specifically generated utterance-dependent (UD) features in a joint probability model. This methodology is tested and compared with recent method using GRID database and the Mocha-TIMIT database. Mowlae et al. [3]

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

proposed approach is based on subband perceptually weighted transformation (SPWT), which is applied on magnitude spectrum, to improve the performance of single-channel separation scenario (SCSS). The proposed method along with the corresponding distortion measure in a single-channel separation framework significantly improves the speech quality of the speakers synthesized output.

Kristjansson et al. [5] proposed a method for segregating two speakers from a single microphone. The two speaker single microphone source separation problem is one of the most challenging source separation scenarios. Hence a method is proposed for separating two speakers from a single microphone channel. This method exploits the fine structure of male and female speech and relies on a strong high frequency resolution model for source signals. This algorithm is able to identify the correct combination of male and female speech and is able to reconstruct the component signals. This methodology was tested on the Aurora 2 data set. This method resulted in 6.59 dB average increase in SNR for female speakers and 5.51 dB for male speakers.

III. METHODOLOGY

Non-negative matrix factorization (NMF) methodology was first introduced by Lee and Seung as a method to decompose images [4]. NMF approach segregates any non-negative matrix N into basis vector matrix B and weight matrix W . All the elements of these matrices are non-negative. Basis vector or matrix basically represents data sample and dictionary elements of each source. Weight basically represents data feature and gains of each source. NMF methodology is represented as follows:

Data $[N] \approx$ Basic Vector $[B]$ Weights $[W]$ (1)

Where,

$$\begin{aligned} N &\in \mathbb{R}_+^{G \times H} \text{ -- original non -- negative data} \\ B &\in \mathbb{R}_+^{G \times K} \text{ -- matrix of basic vectors, dictionary element} \\ W &\in \mathbb{R}_+^{K \times H} \text{ -- matrix of activations, weights or gains} \end{aligned}$$

G – number of rows of the input matrix

H – number of columns of the input matrix

K – basis vector size

The major drawback in the previous methodology is that the basis vector and weight matrix are chosen randomly and the iteration keeps occurring till the desired B and W are achieved. The proposed methodology (ADAPNMF) is adaptive in nature. It means that w.r.t. the input sources the output also varies. This means that as we keep changing the audio samples the output varies accordingly. To obtain $N \approx BW$, cost function needs to be defined. The cost function used in this proposed methodology is linear cost function with marginal cost.

IV. ALGORITHM OF ADAPTIVE NMF

The algorithm for the proposed methodology is given as below

$$\text{Using } \| N - \frac{BW}{\sum B} \| = \| N^T - \frac{B^T W^T}{\sum B} \|^2$$

Now just iterate between

- Updating B
- Updating W
- Checking $\| N - \frac{BW}{\sum B} \|^2$. If the change since the last iteration is small, then declare convergence.

1. Initialize B, W
 $B = 1 + \text{Ps}(G, \text{sum}(K)); W = 1 + \text{Ps}(\text{sum}(K), H);$
2. Repeat
3. for $i=1:\text{MAXITER}$ do

$$B_{ij} \leftarrow B_{ij} \frac{\sum_{k=1}^H \left(\frac{N_{ik}}{B_{ik}(W+\epsilon)_{ik}} \right)}{\sum_{k=1}^H B_{ij}}$$

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

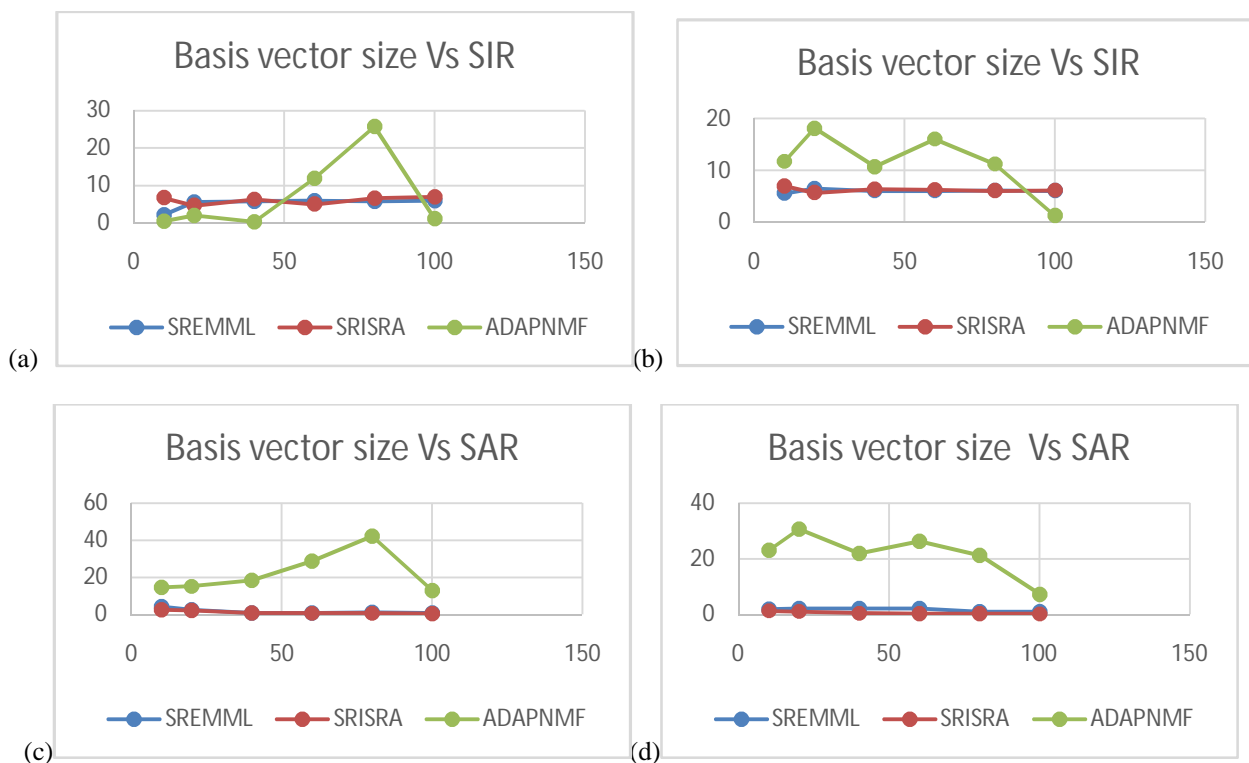
Vol. 6, Issue 8, August 2017

$$W_{kj} \leftarrow W_{kj} \frac{\sum_{k=1}^G B_{ij} \left(\frac{N_{ik}}{B_{ik} (W-\epsilon)_{ik}} \right)}{\sum_{i=1}^G W_{ij}}$$

4. end for
5. Until convergence returns B, W.

V. RESULTS AND DISCUSSIONS

We have evaluated performance of the proposed methodology and previously existing methodologies i.e. Sparse REMML (SREMML) and Sparse RISRA (SRISRA) in terms of signal-to-interference ratio (SIR), signal-to-artifact ratio (SAR) and signal-to-distortion (SDR). Figure 1 shows the evaluating parameters results of keyboard and drums mixture are plotted below. Figure 1: (a) Plot of SIR of keyboard source, (b) plot of SIR of drums source, (c) plot of SAR of keyboard source (d) plot of SAR of drums source, (e) plot of SDR of keyboard source (f) plot of SDR of drums source

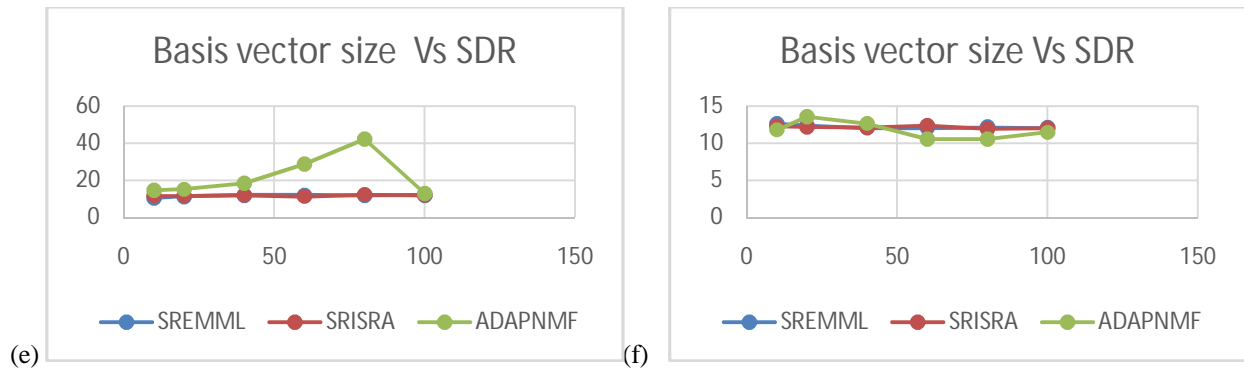


International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017



It is observed in figure 1. (a) that for high frequency signal SIR at basis vector size of 75 is significantly higher than any other method. For our system SIR peaks around 75, whereas decays as size increases above and below 75. In case of drums or lower frequency samples i.e. figure 1. (b) SIR peaks around basis vector 25 and there exists a secondary peak, which is 25% smaller at around vector size 60.

When compared SPREMML and SRISRA systems with our system, ADAPNMF proves to be better than other two when SIR ratio is considered. In very few cases such as basis vector size 90 SIR is less than the SIR achieved with the existing two systems.

With respect to SAR, our algorithm is better than the existing algorithm for all the sizes of vector. It is observed in figure 1. (c) that SAR with high frequency audio is peaking at basis vector size 75. On the other hand in figure 1. (d) shows that for low frequency sounds SAR peaks at basis vector size 25. As mentioned in the earlier case of SIR, SAR also has a secondary peak around basis vector of size 20.

When SDR was computed, it is observed that our system ADAPNMF is slightly better than the existing system for low basis vector size. It is observed in figure 1. (e) that performance at high frequency is significantly improved whereas in figure 1.(f) it is observed that at lower basis vector the performance at low frequency is marginally better.

Although, the higher basis vector size 75 performs exceptionally well at high frequencies, majority of the times the audio files we are targeting are in close approximation of lower frequencies. Hence, for lower frequency performance basis vector size of 25 proves to be best size because it has very good SIR, SAR and SDR values respectively.

VI. CONCLUSION

This paper highlights the significance of single channel speech separation (SCSS) systems, which is required in various applications such as speech and audio denoising, music remixing, music transcription, audio based forensics, etc. It is observed that our proposed algorithm gives better separation results than existing methodologies. It is also observed that basis vector size 25 for all audio files gives best results in terms of SIR, SAR and SDR. It is also observed that as the basis vector size is increased the computational time also increases. One of the unique feature of this proposed methodology is being adaptive in nature.

REFERENCES

- [1] Mona Nandakumar M, Edet Bijoy K "Performance evaluation of single channel speech separation using non-negative matrix factorization" IEEE 2014
- [2] Qi Wang, W.L. Woo and S.S. Dlay "Informed single-channel speech separation using HMM-GMM user generated exemplar source" IEEE/ACM Transactions Vol.22, No.12, Decembere 2014
- [3] Pejman Mowlae and Abolghasem Sayadiyan "Performance evaluation for transform domain model-based single-channel speech separation" IEEE, 2009
- [4] Andrzej Cichocki, Rafal Zdunek and Shun-ichi Amari "New algorithms for non-negative matrix factorization in applications to blind source separation" ICASSP, 2006.
- [5] Trausti Kristjánsson, Hagai Attias "Single Microphone source separation using high resolution signal reconstruction" ICASSP 2004
- [6] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte, "Performance measurement in blind audio source separation" IEEE transactions on audio, speech, and language processing, vol. 14, no. 4, 2006.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor & UGC Approved Journal)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

- [7] F. Cong, Z. Zhang, I. Kalyakin, T. Huttunen-Scott, H. Lyytinen, and T. Ristaniemi, "Non-negative Matrix Factorization vs. FastICA on Mismatch Negativity of Children", proceedings of International Joint Conference on Neural Networks, pp 586-600, June 2009.
- [8] G. Logeshwari and G S. Anandha Mala, "A survey on single channel speech separation" V.V. Das and J. Stephen (Eds.): CNC 2012, LNICST 108, pp. 387–393, 2012. © Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2012
- [9] Bin Gao, W.L. Woo and S.S. Dlay "Single -channel source separation using EMD subband variable regularized sparse features" IEEE transactions on audio, speech and language processing , vol 19, No4 , May 2011.
- [10] Karan Nathwani, Pranav Pandit, and Rajesh M. Hegde "Group Delay Based Methods for Speaker Segregation and its Application in Multimedia Information Retrieval" IIEEE2013
- [11] L. Benaroya, F. Bimbot, and R. Gribonval, "Audio source separation with a single sensor," IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 1, pp. 191–199, 2006.
- [12] C.M. Nelke, N. Chatlani, C. Beaugeant, and P. Vary, "Single microphone wind noise PSD estimation using signal centroids," in Proc. of IEEE Intern. Conf. on Acoustics, Speech and Signal Process. (ICASSP), Florence, Italy, May 2014