Comprehensive Survey on Single Channel Source Separation (SCSS)

[1]Jasreen kaur,[2] Shailaja Gaikwad
[1]M.E. Student, [2] Assistant Professor
[1]Electronics Department,
[1]Terna Engineering College, Navi Mumbai, Maharashtra.

Abstract - **In applications such as audio denoising, music transcription, music remixing, and audio based forensics, Speech denoising and enhancement, it is desirable to segregate a single-channel recording into its respective sources. The aim of getting the desired source from a mixture of sources and noise poses a great difficulty for researchers, especially when each source is highly correlated with one another. The research has been ongoing about the various solutions to be given for the single channel source separation. These works are motivated by the fact that real world sounds are inherently constructed by many individual sounds (e.g., human speakers, musical instruments, background noise, etc.). So, this paper has described various types of methods that have been used for Single Channel Source Separation (SCSS).In addition to this various model has been discussed in the paper. A brief introduction and literature review on the various methodologies is given in this paper. Different designs of experiment techniques available for the optimization of experimental runs have been reviewed.**

Keywords: SCSS, Blind Source Separation, BSS, NMF, Single channel.

I. INTRODUCTION

Blind Source Separation (BSS) of underdetermined mixture has obtained a huge attention in signal processing Environment. Single Channel Source Separation is a subdivision of BSS and has found usage in many applications such as music transcript, music remixing, audio based forensics, etc [7].There has been a large amount of research regarding the task of separating a single mixture recording into its respective sources .It is highly motivated by the fact that real world sounds are inherently constructed by many individual sounds (e.g., human speakers, musical instruments, background noise, etc.). The difficulty of the problem is stretched when only a single recording of the mixture is available. This topic is known as single-channel speech separation (SCSS)[11].While source separation is a difficult and ill-defined mathematical problem, the topic is highly motivated by many outstanding problems in audio signal processing and machine learning, including the following:

- Speech denoising and enhancement – the task of removing background noise (e.g., wind, babble, etc.) from recorded speech and improving speech intelligibility for human listeners and/or automatic speech recognizers.
- Audio restoration – the task of removing imperfections such as noise, hiss, pops, and crackles from (typically old) audio recordings.
- Music remixing and content creation – the task of creating a new musical work by manipulating the content of one or more previously existing recordings[8]

Earlier researchers has utilized many techniques in this context like Principal Component Analysis(PCA) , Independent Component Analysis(ICA) , Microphone Array , Non –Negative Analysis(NMF). Each technique has its own advantages and drawbacks according to the database used.

II. METHODS FOR SINGLE CHANNEL SOURCE SEPERATION

The single channel speech separation (SCSS) has found its application in speech and audio denoising, music transcription, music and audio based forensics etc. In all these application it is desirable to decompose a single-channel recording into its respective sources. [3]

1. Single Microphone source separation using high resolution signal reconstruction

Trausti and Hagai proposed a method [1] 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.

2. Group Delay Based Methods for Speaker Segregation and its Application in Multimedia Information Retrieval

Karan Nathwani, Pranav Pandit, and Rajesh M. Hegde [2] proposed an innovative method of single channel speaker segregation using the group delay cross correlation function. The group delay function, which is the negative derivative of the phase spectrum, generates robust spectral estimates. Hence the group delay spectral estimates are first computed over frequency sub-bands after passing the speech signal through a bank of filters. The spacing of filter bank is based on a multi-pitch algorithm that computes the

pitch estimates of the competing speakers. An affinity matrix is then computed from the group delay spectral estimates of each frequency sub-band. This affinity matrix represents the correlations of the different sub-bands in the mixed broadband speech signal. The grouping of correlated harmonics present in the mixed speech signal is then carried out by using a new iterative graph cut method. The signals are reconstructed from the respective harmonic groups which represent individual speakers in the mixed speech signal. Spectrographic masks are then applied on the reconstructed signals to refine their perceptual quality. It was observed that the segregating performance of the group delay cross correlation method is reasonably higher than several conventional algorithms.

3. Performance evaluation of single channel speech separation using non-negative matrix factorization

Mona Nandakumar M, Edet Bijoy K [3] proposed method to separate the blind source is NMF. Two of the multiplicative algorithms, 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. The proposed method works as follows: input mixture is first Short Time Fourier Transformed (STFT) and separated in to its magnitude and phase components. On the magnitude spectrum of the input signal NMF decomposition is performed and it decompose the magnitude of input mixture into basis vector W and activation vector H. Using masking filters at the reconstruction side followed by Inverse Short Time Fourier Transform (ISTFT) along with the multiplication of phase components, the mixture is separated back into its underlying sources at reconstruction stage 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.

4. Single microphone wind noise PSD estimation using signal centroids

C.M. Nelke, N. Chatlani, C. Beaugeant, and P. Vary [4] proposed an efficient technique for enhancement of speech signals by removing the wind noise. The proposed estimation technique consists of three steps: a feature extraction followed by wind noise detection and the calculation of the current wind noise power spectral density (PSD). It is assumed that the noisy signal x (k) is the superposition of the clean speech signal s (k) and the wind noise signal n (k). The wind noise reduction system is realized as a short-time frequency domain overlap-add structure. The noisy input signal x (k) is first segmented into frames of 20 ms with 50% overlap on which a hann window is applied. The frames are transformed into the frequency domain via a fast Fourier transform (FFT) yielding X (λ, μ) where λ and μ are the discrete frame index and frequency bin. The enhanced signal ˆ S (λ, μ) is obtained by multiplying X (λ, μ) with spectral gains G (λ, μ). The enhanced time domain signal s (k) is obtained by using the IFFT and overlap add. The proposed algorithm has low complexity and low memory consumption as compared to other wind noise reduction techniques.

5. Wind noise short term power spectrum estimation using pitch adaptive inverse binary masks

Wind noise can severely degrade the speech quality and intelligibility. The wind noise is generated by turbulences in an air stream close to the microphone which picks up the desired speech signal. Conventional algorithms for background noise estimation fail in the case of wind noise due to its non-stationary characteristics Christoph M. Nelke and Peter Vary [5] proposed method which exploits the spectral characteristics of speech and noise to estimate the wind noise short term power spectrum (STPS). The spectral power distribution of wind noise and the pitch frequency of speech are used to generate a binary mask for the noise STPS estimations. The comparison with other wind noise reduction techniques shows that the proposed method efficiently removes noise from the speech which leads to improved speech enhancement.

6. Informed single-channel speech separation using HMM-GMM user generated exemplar source

Qi Wang,W.L.Woo and S.S. Dlay[6] combined general speaker-independent (SI) features with specifically generated utterance-dependent (UD) features in a joint probability model. The UD features are initially extracted from the user-generated exemplar source and represented as statistical estimates. These estimates are calibrated based on information extracted from the mixture source to statistically represent the target source. The UD probability model is subsequently generated to target problems of ambiguity and to offer better cues for separation. The proposed algorithm is tested and compared with recent method using the GRID database and the Mocha-TIMIT database. The proposed method has been compared with recent methods and shown to be on-par performance with SD results in terms of the TMR and SDR criteria. The fusion of SI-HMM the UD-HMM to form the FHMM has contributed in reducing the permutation errors.

7. New algorithms for non –negative matrix factorization in applications to blind source separation

Andrzej Cichocki, Rafal Zdunek and Shun-ichi Amari [7] proposed several algorithms for non-negative matrix factorization (NMF) to segregate the underlying sources. The method is focused for sources which are generally statistically dependent under when additional constraints are imposed such as nonnegativity, sparsity, smoothness, lower complexity or better predictability. Although standard NMF (without any auxiliary constraints) provides sparseness of its component, some control of this sparsity as well as smoothness of components can be achieved by imposing additional constraints to natural non-negativity constraints. There are several ways by which incorporate smoothness or sparsity constraints can be incorporated. One of the simplest approaches is to apply in each iteration step a nonlinear projection which can increases sparseness and/or smoothness of the estimated components. An alternative approach is to add to the loss function suitable regularization or penalty terms. The proposed relaxed forms of the NMF algorithms have a higher convergence speed with the desired constraints. The results indicate that the proposed NMF multiplicative algorithms are efficient and robust for extracting and separation of statistically dependent sparse and/or smooth sources.

8. Moving sound source parameter estimation using a single microphone and signal extrema samples

Neeraj Sharma, Sai Gunaranjan Pelluri and Thippur V. Sreenivas [9] proposes a method of speech separation of moving sound sources using Doppler Effect. This method can be applicable for contact less source monitoring applications like industrial robotics and bio –acoustics. The mixture of time-varying sinusoids is analyzed using Doppler Effect. The proposed algorithm uses non-uniformly spaced signal extrema samples of received signal which results in smooth instantaneous frequency (IF) profile. An accurate estimation of moving source parameters can be done using this smooth IF profile. The IF profile is basically composed of IF and its first two derivatives. The numerical implementation of traditional methods, like analytic signal and energy separation approach, results in non-smooth IF profile, which is overcome by this new approach. The proposed approach was evaluated along with two other features, namely conventional log-spectrum and magnitude spectrum. The proposed method together with the consequent distortion measure in a single-channel separation framework significantly improves the speech quality of the speakers synthesized output.

9. Performance evaluation for transform domain model-based single-channel speech separation

Due to the weak quantization performance of the Short time Fourier transformation (STFT) feature vectors, the resulting quality of the separation method degrades. To improve the quantization behavior of the model based single channel source separation (SCSS) Pejman Mowlaee and Abolghasem Sayadiyan[10] proposes sub band perceptually Weighted transformation(SPWT). This transformation also tries to minimize the spectral distortion between the original magnitude spectrum and its reproduction. At the core of this methodology are two trained codebooks of the quantized feature vectors of speakers, whereby the main estimation for separation is performed. The simulation results show that the proposed transformation tends to improve the separation performance of model-based SCSS. The proposed feature results in lower –error bound in terms of the spectral distortion (SD) and also in higher SSNR in comparison with other features.

10. New EM algorithms for source separation and deconvolution with a microphone array

Hagai Attias [11] proposes a new algorithm for source separation with a microphone array. The proposed method is the framework of statistical models. This framework is used to create models for sources and for noise, amalgamation with the reverberant mixing transformation in an upright manner, and compute parameter and source estimates from the data which are Bayes optimal. This methodology results in improved SNR for speech sources. This model tries to overcome issues of ignorance of background and sensor noise.

III. OBSERVATIONS

The brief overviews of all the research papers along with their key features have been tabulated as follows.

Sr. No	Name of Research Paper	Key features
1.	Single Microphone source separation using high resolution signal reconstruction	This algorithm was able to identify the correct combination of male and female speech and was also able to reconstruct the component signals. This methodology has been tested on the Aurora 2 data set. The end result of this methodology was 6.59 dB average increase in SNR for female speakers and 5.51 dB for male speakers.
2.	Group Delay Based Methods for Speaker Segregation and its Application in Multimedia Information Retrieval	It has been concluded that the segregating performance of the group delay cross correlation method was reasonably better than several conventional algorithms.
3.	Performance evaluation of single channel speech separation using non-negative matrix factorization	In terms of SIR, SDR and SAR measures the Sparse REMML algorithm outperforms the Sparse RISRA algorithm in NMF based single channel speech and music separation.
4.	Single microphone wind noise PSD estimation using signal centroids	This methodology shows low complexity and low memory consumption when compared with other wind noise reduction techniques.
5.	Wind noise short term power spectrum estimation using pitch adaptive inverse binary masks	The proposed approach was able to efficiently remove noise from the speech, which leads to improved speech enhancement.
6.	Informed single-channel speech separation using HMM-GMM user generated exemplar source	On comparison with other recent methods, the proposed method gives on-par performance with SD results in terms of the TMR and SDR criteria.
7.	New algorithms for non –negative matrix factorization in applications to blind source separation	The results have indicated that the proposed NMF multiplicative algorithms were efficient and robust in terms of extracting and separation of statistically dependent sparse and/or smooth sources.
8.	Moving sound source parameter estimation using a single microphone and signal extrema samples	The proposed methodology along with the consequent distortion measure in a single-channel separation framework has note worthily improved the speech quality of the speakers synthesized output.
9.	Performance evaluation for transform domain model-based single-channel speech separation	The proposed feature has resulted in lower –error bound in terms of the spectral distortion (SD).When compared with other

		features, this methodology has resulted in higher SSNR.
10.	New EM algorithms for source separation and deconvolution with a microphone array	This methodology has resulted in improved SNR for speech sources. This model has tried to overcome issues of ignorance of background and sensor noise.

IV. CONCLUSION

The paper highlights the importance of single channel speech separation systems critically reviewed its growth in the last few decades, raising an awareness of the challenges faced by the researchers in the development of new theory and algorithms. This paper overviewed different single channel source separation (SCSS) methodologies proposed. The major drawback seen in the proposed methodologies is that a priori knowledge of the underlying sources are required to estimate the sources. Hence a system should be designed for separating two sources without the prior knowledge of the source signals. This paper summarizes by suggesting new direction of research to be focused in future by researchers.

REFERENCES

[1] Trausti kristjansson, Hagai Attias "Single Microphone source separation using high resolution signal reconstruction" ICASSP 2004

[2] Karan Nathwani, Pranav Pandit, and Rajesh M. Hegde "Group Delay Based Methods for Speaker Segregation and its Application in Multimedia Information Retrieval"IEEE2013

[3]Mona Nandakumar M ,Edet Bijoy K "Performance evaluation of single channel speech separation using non-negative matrix factorization"IEEE 2014.

[4] 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

[5] Christoph M. Nelke, and Peter Vary" Wind noise short term power spectrum estimation using pitch adaptive inverse binary masks" Institute of Communication Systems and Data Processing RWTH Aachen University, Germany ,IEEE 2015

[6]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.

[7] Andrzej Cichocki ,Rafal Zdunek and Shun-ichi Amari"New algorithms for non –negative matrix factorization in applications to blind source separation" ICASSP,2006.

[8] Nicholas J. Bryan," Interactive sound source separation" PhD thesis.

[9] Neeraj Sharma,Sai Gunaranjan Pelluri and Thippur V. Sreenivas"Moving sound source parameter estimation using a single microphone and signal extrema samples"ICASSP2015.

[10] Pejman Mowlaee and Abolghasem Sayadiyan "Performance evaluation for transform domain model-based single-channel speech separation"IEEE, 2009

[11] Hagai Attias "New EM algorithms for source separation and deconvolution with a microphone array"in proc.ICASSP 2003.