

Gaussian model based Source Separation Implementation for Speech Separation

Ramjan Khatik, S M Shashidhar

Abstract: *This Paper is an attempt to develop the Independent Component Analysis (ICA) based source separation implementation on the speech signals. The blind source separation technique which work on the basis of the Gaussian process is developed and the performance is analyzed. Blind source separation is the process in which the source separation of the main signal and the noise is separated without any reference available. Matlab based implementation is carried out and the results are obtained. The results thus obtained are satisfactory.*

Keywords: *Blind Source Separation, Gaussian Model, Independent Component Analysis.*

I. INTRODUCTION

A human listener hear the sum of all the sound waves that are created by the varied sources. Although the mixture of the sounds reaches the auditory system there is an ability of the human being which would process that sound in such way that it can be heard individually. The sound from each sources would overlap both in frequency and time. The different sources of other sounds even overlap with both the temporal and spectral regions of the human speech also. In a verbal communication the noise that is created due to the crowd (like in a cocktail party) recognizing a single speaker would become a challenge. Few of the unpublished works that talks about the source separation in audio signals dates back to 1970s [1][4] which is followed by the adaptive variable comb filter based audio separation algorithms[5]. There are attempts made on the frequency domain based approach to segregate the voices those are concurrent recorded in an audiochannel [6-7].

Blind Source Techniques

Blind source separation (BSS) techniques are the important method that needs improvement in the way it is carried out as it has the advantage of not needing a reference signals for the source separation process like in the correlation of feature extraction based implementation. The BSS based works does not need a prior knowledge about the signal under study and as discussed in [8]. In literature [8] the speech separation is carried out from speech signal disturbed with reverberation.

Revised Manuscript Received on October 25, 2019.

Ramjan Khatik, PhD Research Scholar, Visvesvaraya Technological University, E.Mail: rak.ktc@gmail.com

Dr. S M Shashidhar, Principal, PDIT, Hospete, shashi.com@gmail.com

A novel source separation technique called Blind Deconvolution (BD) is applied on the speech in literature [9]. There is a dire need of these blind separation technology for the audio signals when applied on important applications like digital hearing aids and communication technology applications. BSS and BD problem though they work are source separation implementation perspective it can also add subsets like the speech separation and also separation of environmental sources like animal sounds and even music separation. The process of developing the source separation model is a modeling problem where modeling of the auditory system is developed. A term called Auditory Scene Analysis is coined on the problem where each source of the sounds are separated if those sounds emerge from different auditory objects. The book [10] deals with the computational details about the framework that involves the source separation process. When computers are used to computationally separate

The speech by solving the mixture problem then it is defined as computational auditory scene analysis (CASA). Segregation of voices by means of speakers is developed in [11] which is a CASA study on the monaural system. The noisy speech is separated using the Independent Component Analysis (ICA) technique when artificially mixed while the naturally mixed sources were separated with little limitations. While reverberation based source separation did not work well with ICA.

When co-channel speeches that is the speeches that are mixed using a single microphone is considered the traditional approaches would improve the target speech reducing the interfering speech. Though there is multiple peoples voices are there in a single audio source how can it be separated for a single speaker's voice? Thus question imposes a challenge to be addressed by the algorithms used for source separation. To resolve this problem both the spectral and the time domain based techniques are like some early implementation [12-14] that involved robust pitch detection algorithms. Fewer implementations like [15] were dependent on the PDAs while applying harmonic selection algorithms. The localization of different sound streams were involved while the refiltering implementation [16-17] that would apply on the spatio-temporal region. Speech separation algorithms usually used the spectral domain algorithms as each source is

considered the different spectral units in the spectrogram. Few literatures including [18] involved in training the time domain basis functions using ICA in order to predict the different channels in the mixed signals. The method worked satisfactorily for even the real world signals including blind source separation and de-noising. Head-related transfer-function (HRTF) is a method introduced in [19] that would apply spatial filtering on the audio sequence that are added with noises. A spectral learning approach that separate the one microphone speech mixtures using an adaptive segmentation algorithm which is speech specific is carried out [20]. Although segmentation carried out were on the spectrogram the distortions occurred on the outputs. The algorithm that used vector quantizer (VQ) a non-parametric model is used for source separation which performed with lower SNR even in ideal conditions [21]. Learning algorithm that would learn from the corpus of personalized speeches called Sparse Non-negative matrix factorization (SNMF) predicts the different components of the audio stream applied on the monaural speeches [22]. Co-channel speaker identification algorithm with no previous knowledge about the signal is carried out in few literatures [23-26]. A harmonic enhancement and suppression (HES) that separates the speaker at every frame is using the fact that the pitch would vary with the main speaker and the co-channel speaker [27]. These kind of algorithm is working better in even the reverberative environments [28]. The analysis on the co-channel speech signal at places where it is not disturbed by the interferer's speech is used for the speech separation algorithms defined in [29-31]. The signals are separated as usable and non-usable if the usable signal would provide adequate idea of the main speaker.

Independent Component Analysis based Source Separation:

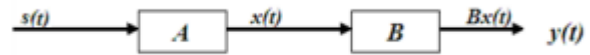
The ICA created BSS problem recovers unidentified source signals by creating an unpretentious supposition of n independent signals represented as $s(t) = s_1(t), \dots, s_n(t)$ and practical mix of signals represented as $x(t) = x_1(t), \dots, x_n(t)$, supposing the mixture of signals as linear and instantaneous, mixing equation can be given as :

$$x(t) = As(t) \quad (1)$$

where A is unidentified mixing matrix and t represents time instance. Here source signal s(t) and the process accountable for transformation of source signal into mixed signal is unidentified which replicates the "Blindness" stuff of the problem. If the mixing matrix A is invertible i.e. no. of sources N is less than or equal to no. of sensors P ($N \leq P$) then the source can be divided right and response y(t) can be articulated as :

$$y(t) = Bx(t) = BAs(t) \quad (2)$$

where B is a separating matrix or inverse of mixing matrix A. The common processing model for BSS can be characterized as



Components individuality is the key supposition of ICA model. It states that for a assumed two variables or source signals s1 and s2, are said to be independent if knowing the value of s1 does not provide any information of s2. In scientific term independence can be defined as : $P_{s1,s2}(s1,s2) = P_{s1}(s1) P_{s2}(s2)$ (3) Where $P_{s1,s2}(s1,s2)$ is the joint density of s1 and s2 , $P_{s1}(s1)$ and $P_{s2}(s2)$ are marginal probability densities of s1 and s2 correspondingly. Objective function to ICA is a gauge of independence principle of ICA which permits approximating ICA data model by design. Objective function is occasionally also known as dissimilarity function or cost function that can be minimized or maximized to obtain optimum results. The statistical properties and algorithmic properties highly be contingent upon the type of objective function being used. The different measures of independence used as ICA objective function are: Non-Gaussianity measure and Mutual Information (Idea about random variables influenced by other member variables of set)

Preprocessing is achieved in ICA to lessen noise from multidimensional dataset in order to escape deprivation of ICA performance. Furthermore pre-processing step before ICA allows shortening the algorithm and reduces the amount of parameters to be projected. Pre-processing steps that need to be achieved before application of ICA algorithm can be given as:

1) Centering : The procedure to "center" the observation x by deducting its mean vector i.e., $x_c = x - m$ (3)

where x_c is the centered observation vector and m is the mean vector. Centering permits to abridge ICA algorithm and un-mixing matrix can be projected by centered data.

Whitening: The procedure that lets to eliminate correlation in the data. Whitening process transforms the observed vector linearly so that a new vector got as a result of the process is white, i.e. components are uncorrelated and variances of the component is equal to unity. Whitening process decreases the number of parameters to be projected; in a way reduces complication of the problem.

II. RESULT AND DISCUSSION:

The separation of the source is the primary step to be carried out on the speech separation process .The Figure shows the ability of the ICA based source separation and PCA based speech separation process that would utilize its dominance in the speech separation paradigm.

Figure 1 depicts the input voice signal for speech separation and the corresponding spectrogram of the signal.

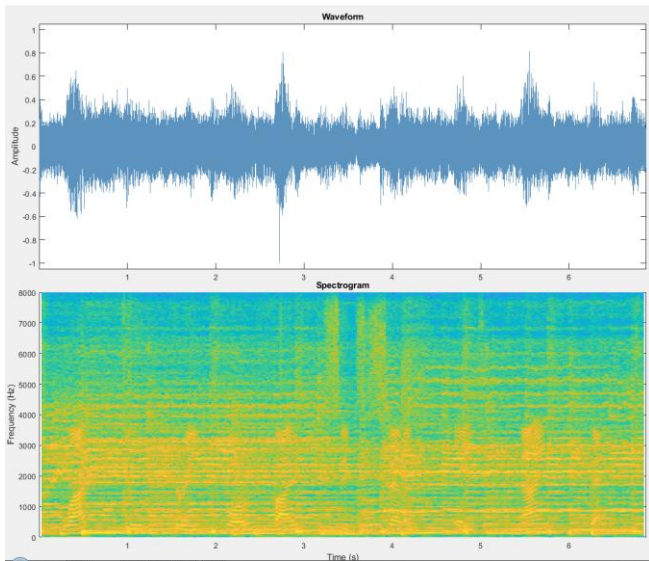


Figure 1. Input voice signal and corresponding spectrogram

The Figure 2 depicts the mixed signal that is under study and the Independent components under study.

The Fast-ICA fits to the kinfolk of fix-point procedures for ICA, that is grounded on the reiteration to examine for the maximum of the non-Gaussianity of variables. The foremost improvement of the fixed point algorithms is that it can converge in a fast manner. (Cubic or at least quadratic). Merging the good statistical possessions (e.g. robustness) of the novel contrast functions, and the decent algorithmic possessions of the fixed point algorithm, Fast-ICA delivers a precise attractive technique for ICA.

The Figure 2. Displays the original signal that is taken for source separation is shown in the first waveform as true signals, and the observed mixed signal are shown in the observed mixed signals, and followed by the independent and principal components.

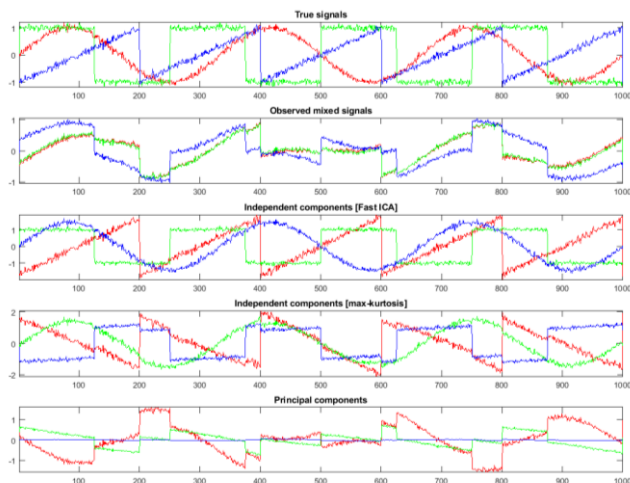


Figure 2. Mixed Signal, Extraction of Independent and Principle Components from the Signal

The PCA implementation defines the PCA extraction from the mixed signals that are as shown in the following Figure 3. The Gaussian separation algorithms that shows the principal components separately in three dimensional views are as shown in Figure3.

First 2 principal directions of 3D Gaussian samples

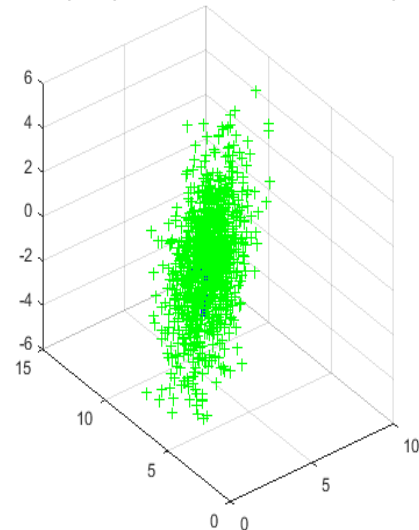
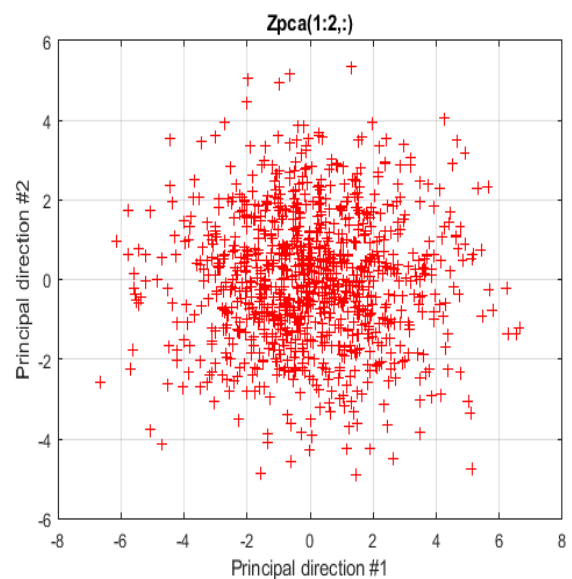


Figure 3. Principal Directions in three Dimensional view



The principal components that define the separation of the source from the noise can be observed.

III. CONCLUSION:

The Gaussian dissimilarity based implementation of source separation is carried out on the speech signal and the source separation is observed. The independent component analysis based implementation is carried out for source separation on speech signals. The Gaussian dissimilarity is use as the measure of separation in the objective function and the results are found to be Satisfactory.

REFERENCES

[1] V.C.Shields,Jr., "Separation of added speech signals by digital comb filtering", S.M.thesis ,Department of Electrical Engineering, MIT, 1970, unpublished.

[2] B.McA.Sayers and E.C.Cherry, "Mechanism of binaural fusion in the hearing of speech," J.Acoust.n Soc.Am, 29, 1957, pp. 973-987.

[3] O.M.M.Mitchell et al., "Signal Processing for a cocktail party effect", J.Acoust. Soc.Am, Vol. 50, 1971, pp.656-660.

[4] J.K.Evertone, Sr., "The separation of voice signals of simultaneous speakers", Ph.D.Thesis (Department of Computer Science, University of Utah, 1975),unpublished.

[5] R.H.Frazier et al, "Enhancement of speech by adaptive filtering", in Procedures of the IEEE International Conference on Acoustics, Speech , and Signal Processing,1 976.

[6] P.W.Parsons and M.R.Weiss., "Enhancig intelligibility in noise or multitalker environments," Rome Air Development Center report, RADC- TR-75-155,1975.

[7] T. W. Parsons, "Separation of speech from interfering speech by means of harmonic selection ", J.Acoust. Soc.Am, Vol. 60,No.4, 1976,pp. 911-918.

[8] S.Haykin, "Unsupervised Adaptive Filtering: Blind Source Separation", Volume 1,2000, John Wiley & Sons.

[9] S.Haykin, "Unsupervised Adaptive Filtering: Blind Deconvolution ", Volume 1I,2000, John Wiley & Sons.

[10] A.S.Bregman.AuditorySceneAnalysis,MITPress,Cambridge, MA1990.

[11] M.Weintraub. AtheoryandcomputationalmodelofAuditoryMonaural Sound Separation.Ph.D Thesis,Stanford University, August 1985.

[12] M.Weintraub,"SoundSeparationandauditoryperceptualorganization", in The psychophysics of speech perception, M.E.H.Schouten(ed.), Dordrecht:MartinusNijhoff, 1987,pp125-134..

[13] R.J.Stubbsandand Q.Summerfield,"Algorithms for separating the speech of interfering talkers: Evaluation with voiced sentences, and normal-hearing and hearing-impaired listeners", J.Acoust. Soc.Am, Vol.84, 1990,pp. 1236-1249.

[14] P.N.DenbighandJ.Zhao,"Pitchextractionandseparationof overlapping speech", Speech Communication, vol.11,1992,pp.119-125.

[15]R.I.Damper,J.R.ThorpeandC.H.Shadle,"SeparationofSpeechfrom simultaneous talkers",X11th International Congress of Phonetic Sciences, 1995

[16] S.T.Roweis,"Onemicrophonesourceseparation",inNIPS,13,2001, pp.793-799.

[17] S.T.Roweis,"Factorialmodelsandfilteringforspeechseparationand denoising", in Eurospeech, 2003,pp.1009-1012.

[18] G.J.JangandT.W.Lee,"Amaximumlikelihoodapproachtosingle channel source separation", JMLR,Vol.4,2003, pp.1365-1392.

[19] B.A.PearMutterandA.M.Zedor,"Monauralsourceseparationusing spectral cues", in ICA, 2004, pp.478-485.

[20] F.BachandM.I.Jordan,"Blindone-microphonespeechseparation:A spectral learning approach", in NIPS,2005,pp.65-72.

[21] D.P.W.Ellis and R.J.Weiss, "Model-based monaural source separation using a vector-quantized phase-vocoder representation", ICASSP, 2006.

[22] M.N.Schmidt and R.K.Olsson., "Single-Channel Speech Separation using Sparse Non-Negative Matrix Factorization", Interspeech 2006, 2006.

[23] J. A. Naylor and S.F.Boll, "Techniques for suppression of an interfering talker in co-channel speech ",Proc.ICASSP,1987, pp. 205-208.

[24] J. A. Naylor and J.Porter, "An effective speech separation system which requires no a priori information",Proc.IEEE ICASSP,1991, pp. 937-940.

[25] C.K.Lee., and D.G.Childers., "Co-channel speech separation", J.Acoust,Soc.Am.,Vol.83,N0.1,1988,pp.274-280.

[26] D.P.Morgan., E.B.George.,L.T.Lee., and S.M.Kay, "Co-channel speaker separation", ICASSP- 95,Vol.1,1995,pp828-831.

[27] T.F.Quatieri and R.G.Danisewicz., "An approach to co-channel talker interference suppression using a sinusoidal model for speech", IEEE transactions on acoustics,speech and signal processing,Vol:38 Jan.1990, pp.56-69.

[28] D.P.Morgan., E.B.George.,L.T.Lee., and S.M.Kay, "Co-channel speaker separation by harmonicenhancement and suppression", IEEE transactions on speech and audio processing, ,Vol:5 Issue:5,Sept.1997,pp407-424.

[29] R.E., Yantorno., "Co-channel speech and speaker identification study", Final report for summer research faculty program, Air Force Office of Scientific Research, Speech Processing Lab, Rome labs, New York, 1998.

[30] R.E.Yantorno., "Co-channel speech study", Final report for research faculty program, Research Laboratory AFRL/IF, Speech Processing Lab, Rome Labs, New York,1999.

[31] B.Y.Smolenski., R.E.Yantorno., D.S.Benincasa., and S.J. Wendnt., "Co-channel speaker segment separation", ICASSP, Vol.1, 2002, pp.125-128.