

Review on Multichannel HR-NMF Modeling for Non-Stationary Signals

Ashwini V.Nandurkar¹, Shrikant J. Honade²

¹G.H.Raisoni College of Engineering and Management, Sant Gadge Baba Amravati University, Amaravti India

²Assistant Professor, G. H.Raisoni College of Engineering and Management, Sant Gadge Baba Amravati University, Amaravti India

Abstract: *Engineering in the technosevia era concerns with the high speed with clarity. It may be either in the signals or in the Images. Golden point is to achieve the accuracy and purity of output. In this paper we studied the signal processing for the part of NMF and multichannel. Along with this we found that numbers of researchers are still working with single channel signal system. The way of implementations of the researchers is different but final aim to achieve the noise free signals. In this paper we are trying to study the multichannel signals and its separation for better output without noise. Most of the researchers use filters and NMF. The observations and study of the various researchers shows that nonnegative matrix factorization (NMF) is noticeable in the spectrograms. This study will take a short review of existing system and will find out the strength and drawbacks of the studied system. This study of review paper will help to minimize the problems of existing system in the implementation phase of our new approach to words the multichannel HR – NMF modeling for Non stationary signals.*

Keywords: non-stationary signal modeling, NMF, time-frequency analysis, variational EM algorithm. ICA, ARMA, HNM, HTFT, CNMF

1. Introduction

In this technosevia era engineering field is more conscious about the frequencies and the time period. Several models had been executed and implemented for modeling time-frequency (TF). The observations and study of the various researchers shows that nonnegative matrix factorization (NMF) is noticeable in the spectrograms. Representation of the audio signal is a spectrogram. This composition and decomposition of the audio signal is extended to horizon for study to the researchers. In this paper, we are trying to focus on drawbacks of exiting NMF multichannel system and represent new model multichannel signals with some advance feature including by overcoming the drawbacks of existing system. This methodology works with the combination of Time frequency, multichannel frequency and non-stationary signals. NMF was introduced as rank-reduction technique which is a non-negative matrix. It is widely used for decomposition of magnitude or power TF in the audio signal processing. Sometimes dissimilar audio events (both in their spectral and temporal characteristics) the decomposition on a single basis (such as the Fourier basis) is mostly not adequate to accurately justify the content of a large class of signals. sparse decomposition techniques use a redundant dictionary of vectors (called atoms) and take a look at to decompose a signal using few of them (much fewer than the dimension of the space) so, the signal may be accurately decomposed with few components

The decomposition of audio signals in terms of elementary atoms has been a large field of analysis for years. The technique known as factorization makes use of the natural redundancy of the signal, Mimicking human cognition that utilizes this redundancy to know visual and audio signals principal component analysis, independent component analysis, sparse coding or NMF are introduced each to reduce the dimensionality and to explain the total

information set by a few meaningful elementary objects. Every generation introduced a model with NMF. Still some loopholes are found in the existing models. In the next part literature review we studied several probabilistic models involving NMF for the best result purpose

2. Review of Literature

In this algorithm for the separation of sound sources in one-channel music signals presentation. It is noticed that Authors done the research By ignoring the phase information Wiener-like filtering [1]–[2] focus on the magnitude or a power TF, which works good but output signals with some noise as musical noise. It has been seen that frequency resolution was limited by that of the TF transform. It was unable to separate sinusoidal signals lying in the same frequency band, also found that it was not corresponds to TF transform of temporal signals. Thus for removing of this drawbacks reconstruction or enhancement as needed.[1].

Esfandiar Zavarehei and et;al ,[1] researchers addressed the problem of lost to noise or suppressed by noise reduction and tried to enhance the performance of conventional speech by using the harmonic noise model. They worked with amplitude, frequency, and harmonicity of sub bands of speech spectrum. They also use weighted codebook mapping for correction of harmonic amplitudes and harmonicies. Researchers also proved with implementation that enhancement is done with improvement of PESQ and LSD signals as compared to noisy speech. Experimental results also show that vocoders based on HNM model is best resultant for this model.

Good generalization ability with feature-based learning was proposed by Zhaozhang Jin and et;al[3]; He noted that there is a unstudied large space for unvoiced speech segregation in reverberant conditions, and has a major scope in the signal

processing area. In his method he learned map from a set of pitch-based auditory features to a grouping cue encoding the posterior probability of a time-frequency (T-F) unit being target dominant given observed features. They use Cooke's corpus and same subset of reverberant mixtures to compare their system with Wang. In the segmentation phase they classified the problem as for robust features and the distance between two classes. They proved by experimental proof that their system model for reverberant mixtures performs better, which was there expected. In addition to this their method can be handled together unit labeling problem that was observed in G. Hu and et;al; they given maximize SNR and reliable segments in the high-frequency range to the new researcher of signal processing.

Nancy Bertin;et;al[4] the problem of the harmonicity and temporal smoothness was observed by the researchers. They implement their algorithm within Bayesian framework with NMF. In this framework model researcher not only given a theoretical result but also supplements a statistical approach as a proof of implementation. In this model NMF base method was superimposed by state-of-the-art methods, in the statistical result they observed that the temporal smoothness constraint has not expected improvement. However result was satisfactory with respect to harmonicity. Limitation of this framework is 100% recall rate is never reached also Temporal smoothness has its drawbacks as far as transients' area unit involved specifically HS-NMF remains terribly slow.

Paul H. Peeling;et ;al[4] as like Nancy Bertin;et;al [3] they also introduce a framework of audio signals, but using NMF in combination with harmonicity and temporal activations and excitations. This paper contributes comparative study in the research zone of signal processing. The problem of selection of single rank matrices (sources) and labeling was still there. On the other hand they describe the adaption of EM algorithm in polyphonic frame-level transcription. The problem of Bayesian framework as there is no distinction between training and validation data is avoided and the timbral and dynamical variation are removed in some sort by using Poliner and Ellis .training and test data. Finally they present for NMF model for spectrogram coefficients

Michael Stark, and et;al[5] In this paper researchers target the single-channel speech separation. They also proposed computational efficient method for the likelihood estimation. The source- and model-driven approach was used for the base study and implementation purpose. This approach was also used by Radfar et al. [15]. On step ahead this researcher worked for spectrogram features of the speech mixture and models it by using Gaussian mixture models (GMMs). They used supervised data, as a raw input for GMM. Grid Corpus was used for result evaluation and for the extraction of the VTFs they use the spectral envelope estimation vocoder method. In the experimental result they observed that the max-sum algorithm performs worse than the junction tree algorithm for GD models applied to same gender mixtures. Overall complexity of the system is the cumulation of their references taken complexity for study and comparison. Drawback of this system is only sequential computations were performed.

Romain Hennequin, and et;al[6] temporal activation becomes frequency dependent model was proposed in this paper. It was an extension to NMF model. Researcher used autoregressive moving average (ARMA) filters for the research. This model was able to take efficiently strong spectral variations of the human voice into account. Drawback of Durrieu's work is overcome by this method. Decomposition of signal was more meaningful for the given spectrogram. This was an iterative algorithm and noticeable spectral variations.

Bin Gao, and etal[7] as like Michael Stark, and et;al he studied on single-channel speech separation. For problem of BSS they used a method of NFM which results in the non-negative matrices. This study of NMF is known with the label adaptive sparsity. Result analysis was done by separating music sources. In the experimental study they used 4-s polyphonic music containing trumpet and piano and plot the results. The resultant signals as also tested in wider types of music mixtures. They evaluate the performance of signals in the term of signal-to-distortion ratio. The challenge of signal separation is cleared with the help of "over-sparse" factorization. Remaining overlapped signals from the "over-sparse" was cleared in "under-sparse" section of this method. By obtaining a result the researcher proved and implements new adaptive sparsity non-negative matrix factorization.

Jonathan Le Roux, et.al; [8] Wiener filter was the root for this research. This researchers played with filters for the NMF and audio signals. In the experimental evaluation they showed separation of single-channel speech and real-world non stationary noise signals. This method was faster than other methods. The implementation of this paper showed that source separation performance under both oracle and blind conditions is improved.

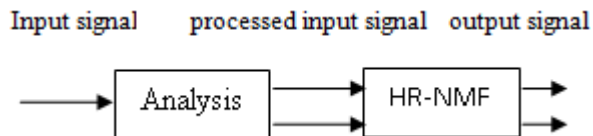
Joonas Nikunen and Tuomas Virtanen; [9] sound source separation problem from a multichannel microphone array was noticed by this researchers. SCM model with weighted sum of direction of arrival (DoA) was used as the base of implementation. Spatial properties of the sources were optimized over all frequencies this research aimed to exceed the performance of existing state of the art separation methods. STFT bin dependencies can be handled by this approach. Time delay dependency can be used across frequencies in a CNMF a set of room impulse responses were considered while evaluation. Approximately linear complexity is found in this algorithm. This method was a novel approach for spatial parameter estimation in frequency-domain. Blind source separation, which makes it interesting concept to be, utilized in different separation model structures. It is not useful in multi-channel case.

3. Proposed Work

- 1) Firstly analyze frequency components.
- 2) Then the HR-NMF model is introduced in the time domain. The HR-NMF model of a multichannel signal is defined for all channels and times as the sum of source images plus a Gaussian noise.

- 3) The multichannel HR-NMF model in the TF domain is presented.
- 4) In order to make the estimation of HR-NMF faster, then two different strategies are used. Firstly improve the convergence rate, by replacing the M-step of the EM algorithm by multiplicative update rules. The second approach aimed to reduce the Computational cost, by using a variational EM algorithm.
- 5) This model is applied to audio signal.

3.1 Flow diagram



4. Conclusion

The proposed high-resolution (HR) NMF is able to model both phases and correlations over time frames (within frequency bands) in a principled way. It shows that this model offers an improved frequency resolution able to separate sinusoids within the same frequency band, an improved synthesis capability and able to restore missing TF observations.

References

- [1] Esfandiari Zavarehei, Saeed Vaseghi, and Qin Yan, "Noisy speech enhancement using harmonic-noise model and codebook-based post-processing," MAY 2007 VOL. 15, NO. 4, pp. 1194-1203.
- [2] Zhaozhang Jin and DeLiang Wang, "A Supervised learning approach to monaural segregation of reverberant speech," May 2009 VOL.17.NO.4, pp. 625-638.
- [3] N. Bertin, R. Badeau and E. Vincent "Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription," Mar. 2010, vol. 18, no. 3, pp.538-549.
- [4] Paul H. Peeling, A. Taylan Cemgil and Simon J. Godsill, "Generative Spectrogram Factorization Models for Polyphonic Piano Transcription" March 2010, VOL18 NO.3, pp519-527.
- [5] Michael Stark, Michael Wohlmayr and Franz Pernkopf, "Source-Filter-Based Single-Channel Speech Separation Using Pitch Information," February 2011, VOL19, NO. 2, pp. 242-255.
- [6] Romain Hennequin, Roland Badeau, and Bertrand David, "NMF with Time-Frequency Activations to Model Non stationary Audio Events," May 2011, VOL. 19, NO. 4, pp. 744-753.
- [7] Bin Gao, W. L. Woo and S. S. Dlay, "Adaptive Sparsity Non-Negative Matrix Factorization for Single-Channel Source Separation," September 2011, VOL.5, NO.5, pp. 989-1001.
- [8] Jonathan Le Roux and Emmanuel Vincent, "Consistent Wiener Filtering for Audio Source Separation", MARCH 2013, VOL. 20, NO. 3, pp. 217-220.

- [9] Joonas Nikunen and Tuomas Virtanen, "Direction of arrival based spatial covariance model for blind sound source separation," MARCH 2014 VOL. 22, NO. 3, pp. 727-739.
- [10] Roland Badeau, Senior and Mark D. Plumbley, "Multichannel High-Resolution NMF for Modeling Convulsive Mixtures of Non-Stationary Signals in the Time-Frequency Domain," November 2014, VOL. 22, NO. 11, pp. 1670-1680.
- [11] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 3, pp. 1066-1074, Mar. 2007.
- [12] A. Liutkus, R. Badeau, and G. Richard, "Informed source separation using latent components," in Proc. 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA), Saint Malo, France, Sep. 2010, pp. 498-505.
- [13] G. Hu and D. L. Wang, "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. Neural Netw., vol.15, no. 5, pp. 1135-1150, Sep. 2004.
- [14] G. Hu and D. L. Wang, "An auditory scene analysis approach to monaural speech segregation," in Topics in Acoustic Echo and Noise Control, E. Hansler and G. Schmidt, Eds. New York: Springer, 2006, pp. 485-515.
- [15] M. H. Radfar, R. M. Dansereau, and A. Sayadiyan, "A maximum likelihood estimation of vocal-tract-related filter characteristics for single channel speech separation," J. Audio, Speech, Music Process., vol. 1, pp. 15, 2007.