International Journal of Engineering & Technology, 7 (2.32) (2018) 494-497

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Speech Processing Using FD Independent Component Analysis Algorithm

Dr. N. Shanmuga Priya[1]

[1]Asst. Professor & Head, Dept. of IT, DRSNSRCAS, Coimbatore.
**Corresponding Author Email: [*1]spriyanatrajan@gmail.com.*

Abstract

Speech is most mighty and usual medium to alternate the information amongst individuals. The speech method additionally goes underneath the system of some application tools to reward the know-how in effective approach. Audio supply separation is the crisis of computerized separation of audio sources gift in a room, making use of a set of differently placed microphones, shooting the auditory scene. On this paper, proposed a novel quick Frequency domain ICA algorithm using two viable implementations. Additionally, a effective possibility Ratio soar approach to the permutation difficulty of ordering sources along the frequency axis is provided. The suggestion of exploiting the additional geometrical expertise, such because the microphone spacing, so as to participate in permutation alignment using beam forming is then examined and ultimately discussed accuracy, sensitivity and Throughput ration values are evaluate with current and proposed ways.

Keywords: *frequency domain, speech processing, independent component analysis.*

1. Introduction

Humans showcase a tremendous potential to extract a sound object of curiosity from an auditory scene. The human brain can perform this everyday mission in real time utilizing only the knowledge bought from a pair of sensors, i.E. Our ears. Think the crisis of jogging down a busy avenue with a buddy. Our ears capture a tremendous style of sound sources: auto noise, different persons talking, a buddy talking, mobile telephones ringing. Nevertheless, we are able to focal point and isolate a designated supply that is of interest at this point. For instance, we may just listen to what our buddy is announcing. Becoming bored, we are able to overhear someone else's dialog, pay concentration to an disturbing cellular ringtone and even take heed to a passing vehicle's engine, handiest to recognize it is a Porsche. The human mind can automatically focal point on and separate a specific source of curiosity. Audio supply separation can be outlined as the problem of decomposing a real world sound blend (auditory scene) into man or woman audio objects. The automated evaluation utilizing a pc that captures an auditory scene through a number of sensors is the principal purpose of this thesis. Although this is a relatively easy assignment for the human auditory procedure, the automatic audio source separation may also be considered some of the difficult subject matters.

There are lots of functions the place an audio supply separation method can also be valuable: Noise Suppression for cell telephones/listening to aids. Having unmixed the sources that exist in an auditory scene, you'll eliminate the undesirable noise sources in a a couple of supply atmosphere. This will function a denoising utility for mobile phones, hearing aids or some other recording facility. Song transcription. Unmixing a recording to the exact devices that are taking part in within the recording is an incredibly priceless instrument for all track transcribers. Taking note of an instrument enjoying solo as a substitute the precise recording allows the transcription system. This applies to all automated polyphonic transcription algorithms that have appeared in study. Combining a supply separation algorithm with a polyphonic transcriber will lead to a very robust musical analysis tool. Effective coding of track. Each instrument has exceptional pitch, assault, timbre traits, requiring exclusive bandwidth for transmission. Decomposing a musical sign into sound objects (devices) will allow one-of-a-kind encoding and compression levels for each and every instrument, relying on its traits. The outcome will be a more efficient, excessive exceptional audio codec. This will likely be extra in keeping with the final framework of MPEG-four for video and audio. Medical applications. There are medical functions where an audio supply separation algorithm possibly useful, such as the separation of foetus's heartbeat from the mummy's within the womb. Surveillance purposes. The potential of discriminating between the audio objects of an auditory scene will enhance the efficiency of surveillance functions. Remixing of studio recordings. In the next day's audio functions, with all of the powerful tools that can seek for songs much like the ones we like or that sound like the artist we wish, a private remixing of a studio recording consistent with our liking will probably be possible with audio supply separation.

2. Literature Survey

Fuming Chen, Sheng Li, Chuantao Li, Miao Liu, Zhao Li, Huijun Xue, Xijing Jing and Jianqi Wang (2015), proposed for reinforcing the fine of usual speech is one of the important and effective manner for human communique, as a result, speech acquisition is particularly most important. There are some methods which can be utilized to gather speech signals, such as common air-borne microphones and non-air-borne contact detection. Nonetheless, usual microphones are easily disturbed by using history noise and their propagation distance is very quick, at the same time different methods utilizing non-air-borne contact detection akin to electroglottography and the bone conduction microphone constrain person's free motion and make customers suppose uncomfortable [13].

Pejman Mowlaee and Rahim Saeidi (2013), proposed an answer for the segment estimation concern utilising each geometry and workforce prolong deviation minimization. Desired speech signals are traditionally corrupted with some background noise the place the recording takes position, resulting within the requirement of a single-channel speech enhancement pre-processor for unique speech applications, to call a number of: mighty automated speech realization and speech transmission. The proposed procedure performs virtually two bounds on the speech enhancement performance accomplished via contemplating prior understanding about speech amplitude or phase spectra; in phrases of PESQ, at low SNRs, the proposed process asymptotically reaches the efficiency exhibited via the section-conscious amplitude estimator given by using the oracle section values even as at high SNRs, the proposed system performs practically that bought by exploiting the estimated phase given the oracle amplitude spectrum prior [14].

Ruiyu Liang, Ruxue Guo, Ji Xi, Yue Xie and Li Zhao (2017), proposed evolution algorithm. In this work, the updated parameters are the sub-band positive aspects. . The essence of self-becoming is that the set up, fitting and use of hearing aids are completed by means of the consumer by myself, without any listening to experts or equipment The design suggestion of the user-programmable listening to aid is just like that of the self-becoming listening to support, i.E., the person manually adjusts the parameters of the listening to aid to compensate for the deterioration of speech with alterations within the environment. The difference between these two categories of listening to aids is that the quantity and extent of exchange of the adjustable parameters for the user-programmable listening to help are restricted, whereas those of the self-becoming hearing aid are not [15].

Thijs van de Laar and Bert de Vries (2016), proposed a principled in-situ design method for HA algorithms that use binary performance feedback from patients. The process pursuits to cut down the burden-of interaction on the top consumer. Listening to loss is an primary main issue that affects the great of life of millions of individuals. About 15% of american adults (37.5 million) file issues with listening to. The proposed Bayesian model assessment procedure provides a principled process for choosing a great algorithm structure amongst choices, headquartered on a consumer-selected data set [16].

Sunita Dixit, Dr. MD Yusuf Mulge (2014), established on the provision of an auxiliary channel, known as reference direction, wherein a correlated pattern or reference of the contaminating noise is gift. The area noisy speech is praise. The adaptive noise cancellation (ANC) cancels the primary unwanted noise r(n) by introducing a canceling anti-noise of equal amplitude however contrary phase using a reference sign. This reference signal is derived from some of sensors located at functions near the noise and interference assets the region the interest signal is prone or undetectable [17].

Stephan Liwicki Georgios Tzimiropoulos Stefanos Zafeiriou Maja Pantic (2011), proposed a kernel PCA approach for rapid and mighty PCA, which we name Euler-PCA (e-PCA). In certain, our algorithm utilizes a strong dissimilarity measure established on the Euler illustration of problematic numbers. We show that Euler-PCA retains PCA's fascinating properties even as suppressing outliers. In addition, we formulate Euler-PCA in an incremental finding out framework which allows for effective computation. In our experiments we practice Euler-PCA to a few exceptional computer vision applications for which our procedure performs comparably with other stateof-the-art methods [18].

3. Proposed Work

3.1 Fast Frequency Domain ICA Algorithm

To switch the normal gradient scheme in the FD-ICA framework with a Newton-kind optimization scheme. Their common feature is that they converge so much rapid than gradient algorithms with the same separation satisfactory and while they are more computationally high-priced, the number of iterations for convergence is decreased. Within the following evaluation, we show that it is fundamental to prolong the algorithm proposed in to be applicable to the proposed time-frequency framework. In the context of FD-ICA, at a given frequency bin, the Unmixing matrix may also be interpreted as a null-steering beamformer that uses a blind algorithm (ICA) to place nulls on the interfering sources. The supply separation framework does no longer use any understanding related to the geometry of the auditory scene, however only the sources statistical profile. Inclusion of this extra knowledge can support in aligning the diversifications. Even though, we're coping with actual room recordings, we anticipate that there is a constant DOA alongside frequency for every source, belonging to the direct path signal. That is an identical of approximating the room's switch function with a single prolong. The permutations of the Unmixing matrices are flipped in order that the directivity sample of each and every beamformer is approximately "aligned".

In the FD-ICA approach for instantaneous mixtures, we form and try to maximize the following likelihood with respect to the un-mixing matrix W:

$$\mathbf{F = \log L(x|W) = \mathcal{E}\ \{\log p(u)\} + \log|\det(W)|}$$

In, Hyvarinen tries to solve the following optimization problem:

$$\mathbf{\max \mathcal{E}\ \{G(W_x)\}}$$
$$\textbf{subject to } \mathbf{\mathcal{E}\ (uu^T) = I}$$

where G(u) is a non-quadratic function. The solution for this problem can be estimated by finding the maximum of the following function:

$$\mathbf{K(W) = \mathcal{E}\{G(W_x)\} - \alpha(\mathcal{E}\ \{uu^T\} - I)}$$

where α is the Lagrange multiplier. Performing a gradient ascent on K(W), we get:

$$\mathbf{K = \mathcal{E}\ \{G^d\ (W_x)_x{}^{T\}} - \alpha CW}$$

where $C = E\{xx^T\}$. If we choose G(u) = log p(u), then this update law is almost identical to the Bell-Sejnowski law and the natural gradient, with a different term controlling the scaling of the unmixing matrix W. After a series of steps and using G(u) = log p(u), we end up to the following learning rule:

$$\mathbf{W = D[diag(-\alpha_i) + \mathcal{E}\ \{\ \varphi(u)u^{T\}}\}]\ W}$$

where $\alpha_i = \mathcal{E}\{u_i\varphi(u_i)\}$, $D = \text{diag}(1/(\alpha_i - \mathcal{E}\{\varphi'(u_i)\}))$. In practice, we observed that this algorithm converges at a faster rate than the gradient based update rules, as it will be demonstrated further on. Evaluating the update rule with the long-established common gradient legislation, we can see that they are similar. Instead of a constant finding out price, there's a studying fee (the D matrix) that adapts to the sign. Hence, the algorithm is less stylish on sign levels and hence extra steady. Hyvarinen states that replacing I with the adaptive time period diag(−αi) is also useful for convergence speed. If we use pre-whitened data x, then the formula is an identical to the common fixed-factor algorithm, while it is still expressed in terms of the average gradient algorithm. The fundamental outcome for us, nonetheless, is that the nonlinear activation function φ(u) in has exactly the identical interpretation as within the ML technique.

4. Experimental Results

4.1 Accuracy

Table 1: Comparison table of Accuracy

Beam forming	Normalized Least Mean Square	Proposed Frequency Domain Independent Component Analysis
0.21	0.145	0.426
0.289	0.175	0.5
0.3	0.202	0.542
0.345	0.281	0.599
0.391	0.35	0.644

This table describes the comparison of accuracy of two existing methods that is Beamforming, Normalized Least Mean Square method and proposed method frequency Domain Independent Component Analysis. Comparing these three methods we assume that proposed method shows highest values of accuracy from 0.426 to 0.644. Whereas the other two shows a minimum accuracy ratio than proposed method.

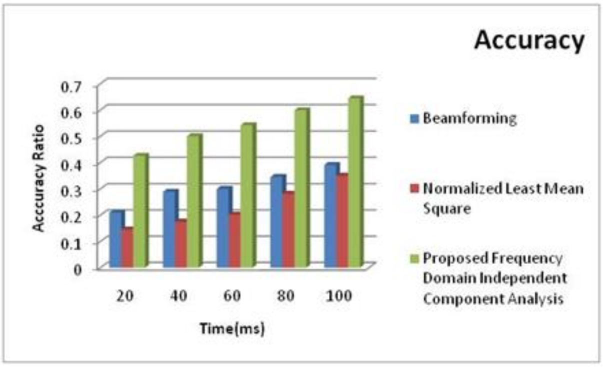

Fig. 1: Comparison chart of accuracy

This chart explains about the accuracy ratio of two existing methods and one proposed method. The accuracy ratio of its range is been explained using number of packets in X axis and the accuracy ratio of the process in Y-axis. While analyzing and comparing proposed method with existing method, proposed Method shows maximum value of accuracy ratio from 0.426 to 0.644. Whereas, the other two existing methods involves more accuracy ratio.

4.2 Sensitivity

Table 2: Comparison table of sensitivity

Beam forming	Normalized Least Mean Square	Proposed Frequency Domain Independent Component Analysis
0.037	0.089	0.149
0.066	0.102	0.188
0.089	0.145	0.16
0.101	0.189	0.256
0.123	0.2	0.223

This table describes the comparison of sensitivity of two existing methods that is Beam forming, Normalized Least Mean Square method and proposed method frequency Domain Independent Component Analysis. Comparing these three methods we assume that proposed method shows highest values of sensitivity from 0.149 to 0.223. Whereas the other two shows a minimum sensitivity ratio than proposed method.

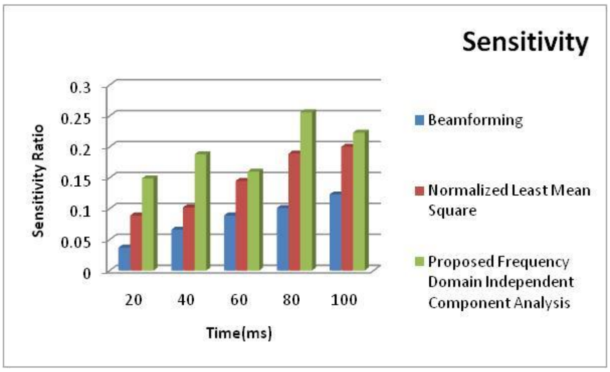

Fig. 2: Comparison chart of sensitivity

This chart explains about the sensitivity ratio of two existing methods and one proposed method. The sensitivity ratio of its range is been explained using number of packets in X axis and the sensitivity ratio of the process in Y-axis. While analyzing and comparing proposed method with existing method, proposed Method shows maximum value of sensitivity ratio from 0.149 to 0.223. Whereas, the other two existing methods involves more sensitivity ratio.

4.3 Throughput

Table 3: Comparison table of Throughput Ratio

Beamforming	Normalized Least Mean Square	Proposed Frequency Domain Independent Component Analysis
0.037	0.089	0.099
0.066	0.102	0.128
0.089	0.145	0.16
0.101	0.189	0.206
0.123	0.2	0.213

This table describes the comparison of throughput of two existing methods that is Beamforming, Normalized Least Mean Square method and proposed method frequency Domain Independent Component Analysis. Comparing these three methods we assume that proposed method shows highest values of a throughput from 0.099 to 0.213. Whereas the other two shows a minimum throughput ratio than proposed method.

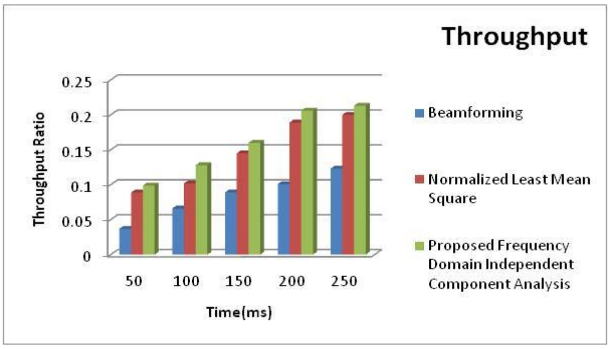

Fig. 3: Comparison chart of throughput

This chart explains about the throughput ratio of two existing methods and one proposed method. The throughput ratio of its range is been explained using number of packets in X axis and the sensitivity ratio of the process in Y-axis. While analyzing and comparing proposed method with existing method, proposed Method shows maximum value of throughput ratio from 0.149 to 0.223. Whereas, the other two existing methods involves more throughput ratio.

5. Conclusion

Addressed the problem of convolutive mixtures source separation using Frequency-Domain Independent Component Analysis. A method to solve the scale ambiguity was proposed. A novel method to solve the permutation ambiguity using a time-frequency source model along with a Likelihood Ratio jump solution was proposed. The methods seemed to rectify the permutation problem in the majority of the cases. Two fast “fixed-point” algorithms were adapted to work with complex numbers and incorporate the solution for the permutation problem. All these modules were put together to form a unifying frequency domain ICA framework that manages to perform fast and robust source separation in the majority of the cases. Several tests were performed to test the efficiency of the proposed framework with encouraging results.

References

[1] Yi Hu and Philipos C. Loizou, “Subjective comparison and evaluation of speech enhancement algorithms” IEEE Trans.Speech Audio Proc.2007:49(7): 588–601.

[2] Yariv Ephraim, Hanoch Lev-Ari and William J.J. Roberts “A Brief Survey of Speech Enhancement” IEEE Sig. Proc. Let., vol. 10,pp. 104-106, April 2003 s.

[3] Yoon. B-Y.; Tashev, I. & Acero, A. (2007) Robust Adaptive Beamforming Algorithm Using Instantaneous Direction Of Arrival With Enhanced Noise Suppression Capability. IEEE International Conference on Acoustics, Speech and Signal Processing 1:I-133–I136.

[4] Reddy, D.R, “Speech recognition by machine: A review”, Proceedings of IEEE (Volume: 64, Issue: 4) ISSN: 0018-9219.

[5] Young, S.J, “Robust continuous speech recognition using parallel model”, IEEE Transactions on Speech and Audio Processing (Volume: 4, Issue: 5).

[6] Bassam A. Q. Al-Qatab , Raja N. Ainon, “Arabic Speech Recognition Using Hidden Markov Model Toolkit(HTK)”, 978-1-4244-6716-711 0/$26.00 ©2010 IEEE.

[7] Chibelushi, C.C. “A review of speech-based bimodal recognition”, 2002 IEEE Transactions on Multimedia, (Volume: 4, Issue: 1).

[8] Iosif Mporas,”Dynamic Selection of A Speech Enhancement Method For Robust Speech Recognition In Moving Motorcycle Environment”, ICASSP 2011 978-1-4577-0539- 7/11©2011 IEEE.

[9] Ciira wa Maina,”Joint Speech Enhancement and Speaker Identification Using Approximate Bayesian Inference”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING 1558-7916© 2010 IEEE.

[10] Fuming Chen, Sheng Li, Chuantao Li, Miao Liu, Zhao Li, Huijun Xue, Xijing Jing and Jianqi Wang, “A Novel Method for Speech Acquisition and Enhancement by 94 GHz Millimeter-Wave Sensor”, Published: 31 December 2015 Academic Editor: Vittorio M. N. Passaro.

[11] Pejman Mowlaee and Rahim Saeidi, “Iterative Closed-Loop Phase-Aware Single-Channel Speech Enhancement”, IEEE SIGNAL PROCESSING LETTERS, VOL. 20, NO. 12, DECEMBER 2013.

[12] Ruiyu Liang, Ruxue Guo, Ji Xi, Yue Xie and Li Zhao, “Self-Fitting Algorithm for Digital Hearing Aid Based on Interactive Evolutionary Computation and Expert System”, Published: 9 March 2017.

[13] Thijs van de Laar and Bert de Vries, “A Probabilistic Modeling Approach to Hearing Loss Compensation”, Copyright (c) 2016 IEEE.

[14] Sunita Dixit, Dr. MD Yusuf Mulge, “Review on Speech Enhancement Techniques”, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.8, August- 2014, pg. 285-290.

[15] Stephan Liwicki Georgios Tzimiropoulos Stefanos Zafeiriou Maja Pantic Barkatullah Qureshi, “Euler Principal Component Analysis”, Received: 5 December 2011 / Accepted: 8 August 2.