Blind Source Separation of Speech Signals using Kurtosis, Negentropy and Maximum Likelihood.

Mukul Kant
Dept. Of EEE
BITS Pilani, Rajasthan
2021H1240095P

Somnath Molio
Dept. Of EEE
BITS Pilani, Rajasthan
2021H1240091P

Yogi Patil
Dept. Of EEE
BITS Pilani, Rajasthan
2021H1240090P

***Abstract*— Independent component analysis is an excellent technique for extracting blind sources from their determined or overdetermined mixtures. FastICA is one of the state-of-the-art algorithms in ICA. The objective of this paper is to test the robustness of several contrast functions for speech signal analysis in noisy environments. The contrast function is a nonlinear function used to measure the independence of the estimated sources from the observed mixture signals in the FastICA algorithm. Kurtosis, negentropy and maximum likelihood functions are used as contrast functions in the FastICA algorithm. The FastICA algorithm based on these contrast functions is applied to both instantaneous synthetic mixtures as well as real recorded mixtures. The performance of contrast functions is evaluated using spectrograms and signal to distortion ratio. The maximum likelihood contrast function is found to yield better results for real time mixtures.**

I. Introduction

The aim of blind source separation is to extract desired sources from their mixtures without any prior information with regards to the source signals or the mixing environment involved. A real time example of BSS is the "cocktail party problem" where the listener is trying to follow a particular speaker from the mixture of speech that he/she is hearing.

Figure 1 illustrates general instantaneous BSS or blind source extraction (BSE) problems.

$\mathbf{x(n)} = [x_1(n), x_2(n)\ldots x_M(n)]^T$, **M** stands for observation of mixed signals from the sensors and **n** denotes discrete time index.

$\mathbf{s(n)} = [s_1(n), s_2(n)\ldots s_N(n)]^T$ denotes the source signals, where **N** is the number of sources and it is unknown in many situations.

v(n) designates uncorrelated additive white Gaussian background noise sources and contributes for the mixture signals **x(n)** apart from the contribution of **s(n)**.

Inverse of mixing matrix **A** is **W** of order **N × M**. **W** separates the desired sources from the mixture signals and it is known as the separation matrix. **y(n) =[y1(n), y2(n)…yN(n)]T** is an estimated version of the source signal s(n) after applying BSS technique to the mixture.

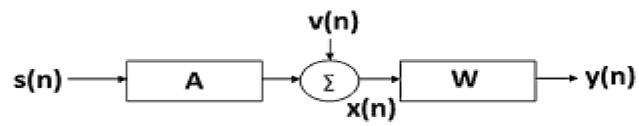


Fig. 1. General instantaneous BSS problem formulation or BSE problem

The general BSS problem may be stated as an estimation of the source signals **s(n)** from the observed mixture signals **x(n)** without knowing any information about **A**, **s(n)** and **x(n)** . The problem may seem quite irrational without any available information. Fortunately, statistical independence of the sources remains true in most applications and is a useful priori information for solving the BSS problem. However, we cannot estimate source signals up to their amplitude and their order. These indeterminate states are called as scaling and permutation ambiguity in the BSS respectively. Above example gives signal structure related useful information about the sources not the exact amplitude of the signals, also scaling ambiguity is not considered as a serious issue in this paper. BSS technique is used in various fields like speech enhancement and recognition, feature extraction, biomedical signal processing (EEG, fMRI), signal denoising, data mining and telecommunications. The base paper chosen is referenced as [1].

ICA as a statistical method for transforming an observed multidimensional random vector into components that are mutually as independent as possible; this was discussed in [2].

Every ICA neuron develops into a separator that finds one of the independent components. For reasons of computational and conceptual simplicity, the representation is often sought as a linear transformation of the original data was used in [3]. Application of a separation algorithm on each frequency separately, to solve the amplitude and permutation ambiguity properly to reconstruct the separated signals using matrix inversion for amplitude ambiguity and for the permutation ambiguity, a method based on the temporal structure of speech signals is discussed in [4].

II. Methods

A. Central limit theorem

Contrast functions are used to measure the degree of statistical independence. A measure of independence is equivalent to a measure of non-Gaussianity. According to the central limit theorem, the sum of non-Gaussian random variables is closer to Gaussian than that of the individual random variables .

$$y_i\ (n) = \sum_{j=1}^{M} w_{ij} x_j(n) \quad , \quad i = 1,2,\ldots,N$$

where, wij are elements of the separation matrix W. wij are chosen such that the linear transformation produces yi (n) as one of the independent components. Its non-Gaussianity is more than any other combinations of source signals. Because, if yi (n) are a combination of two or more sources then it would be closer to Gaussian distribution than the independent components. So searching for the combination which has maximum non-Gaussianity, it is likely that such a combination would have originated from a single source rather than more than one source.

B. Whitening

Typical preprocessing tools include Principal Component Analysis (PCA), Factor Analysis (FA), whitening, Fast Fourier Transform (FFT), Time Frequency Representation (TFR) and sparsification. Whitening is used as a preprocessing technique in the FastICA algorithms. A zero mean random vector is said to be white, if its elements are uncorrelated and have unit variance. One of the ways to make the signal white is to find a linear transformation matrix V such that mixture signals are turned white.

The transformation can be computed from Eigenvectors and Eigenvalues of the covariance matrix of mixture signals

$$z(n) = V_{M\times M}\ x(n)_{M\times 1}$$
$$C_{M\times M} = E_{M\times M}\ D_{M\times M}\ E^T_{M\times M}$$
$$V_{M\times M} = D^{-1/2}_{M\times M}\ E^T_{M\times M}$$

For obtaining Eigenvectors and Eigenvalues, Eigenvalue Decomposition (EVD) is used, where E has Eigen vectors as columns and D is a diagonal matrix containing Eigenvalues as its diagonal elements.

C. Kurtosis

Kurtosis is the measure of non-Gaussianity, specifically the ‘tailedness’ of a distribution. Only the samples that lie outside at least one standard deviation from the mean are significant contributors to Kurtosis. It is defined as the fourth order cumulant of a random variable.

kurt[y(n)] = E { y^4(n)} - 3[E^2{y(n)}]^2

Since **y(n)** is an estimate of the source signals, **E{y^2(n)}** is assumed to be unit variance and hence the kurtosis of the source signal is just a normalized version of the fourth moment. Kurtosis is zero for Gaussian random variables and nonzero for most non-Gaussian random variables. Kurtosis is positive for super gaussian random variables and negative for subgaussian random variables. To measure non-Gaussianity of the random variable, the absolute value of the kurtosis is taken as the measure of non-Gaussianity. Consider the estimation of source signals from whitened signals by using separation matrix **W N×M**

$$\mathbf{y(n)_{N\times 1} = W_{N\times M}\ z(n)_{M\times N}}$$

$$\mathbf{y(n)_{N\times 1} = W_{N\times M}\ V_{M\times N}\ x(n)_{M\times 1}}$$

$$\mathbf{W_{N\times M}\ V_{M\times M}\ M_{M\times N} s(n)_{N\times 1}} = B_{N\times N}\ s(n)_{N\times 1}$$

The estimated source signal **yi (n)** is equal to one of the source signals if only one element of matrix B in each and every row and column is unity. This shows that source signals can be estimated upto permutation.

D. Negentropy

Negentropy is an information theoretic measure of non-Gaussianity. Negentropy measures the difference in entropy between a given distribution and a Gaussian distribution with the same mean and variance.

$$\mathbf{J[y_i(n)] = H[y_G(n) - H[y_i(n)]}$$

Where, **H(yG)** is the entropy of a Gaussian random variable with the same mean and variance as **yi (n)** and **H[yi (n)]** is entropy of a random variable **yi (n)**. Property of entropy suggests that a Gaussian distributed random variable has maximum entropy than any other distributions. So negentropy is always non-negative and it is zero only if the distribution is Gaussian. The advantage of negentropy is that the linear transformation of a random variable does not change its negentropy value, whereas the kurtosis varies with linear transformation of a random variable. Since the estimation of probability density function is difficult, some approximation of negentropy is used in practice. A reliable and flexible approximation proposed in Pearlmutter and Parra (1997).

$$J[y_i(n)] = \rho[E\{f_i(y_i)\} - E\{f_i(y_G)\}]$$

where, ρ is constant and f_i**(.)** is a non-quadratic function. If $y_i 4(n)$ is to be considered as that non-quadratic function, then it again leads to kurtosis based measurement of nonGaussianity. The following choices of non-quadratic functions proved to be robust estimators (Hyvarinen et al. 2001).

$$f_i[y_i(n)] = log\ cosh[y_i(n)]$$
$$f_i[y_i(n)] = -exp[-y_i^2(n)/2]$$

We have chosen the second for estimation.

E. Maximum likelihood estimation

Maximum likelihood is a popular method in the estimation of independent components. Consider the basic ICA model, **x(n) = A*s(n)**. According to basic ICA model, the density of observed mixture signals can be represented as

$$\mathbf{x(n) = A^*\ s(n)}$$

$$\mathbf{P_x[\ x(n)] = (1/det(A))^*\ p_s(x(n))}$$

$$\mathbf{P_x[\ x(n)] = det(B)^*\ p_s(x(n))}$$

where **B = A-1.**

Since calculating the joint distribution is a complicated procedure. The estimator proposed by Hyavarinen(Based on material from the book Independent Component Analysis Copyright Wiley Interscience,2001) is used.

E{tanh(S)*S – (1 – tanh(S)^2)}

F. FastICA algorithm

The FastICA algorithm uses any contrast function which maximizes the non-Gaussianity to find the independent components from the observed mixture signals after proper preprocessing. The next step in the FastICA algorithm is solving constrained optimization problems. Here the contrast function is optimized with the constraint that bounds the norm of the separation vector to unit. Newton's method is used for solving constrained optimization problems. FastICA algorithm steps are as follows:

1. Center the observed sensor signals to make its mean zero.

   $\mathbf{x_i(n) = x(n) - E[x_i(n)]}$ **for i = 1, 2, … M**

2. Linearly transform the observed sensor signals so that it is white.

   z(n) = V x(n)

3. Choose an initial separating matrix randomly
4. Choose an initial separating matrix randomly

   y(n) = W z(n)

5. Update separation matrix based on the contrast function calculation.
6. Normalize the separation matrix so that to avoid repeated extraction of independent components.

W ← (WCW')^(-0.5) . W

7. If the separation matrix has not converged, then continue from step 4.

III. Results

In the base paper, Noisy speech data sets from Araki and Vincent (2019) are used to evaluate the efficiency of various algorithms. Data sets containing a synthetic instantaneous mixture as well as live recording mixture signals used. Two channel mixtures of two speech sources and real background noises sampled at 16 KHz are considered for testing our algorithm. The mixture and source signals of length 10 s are used.

Because we had difficulty obtaining the data set mentioned above, we have used the data set available in Brian Moore (2021). PCA and ICA Package (https://www.mathworks.com/matlabcentral/fileexchange/38300-pca-and-ica-package), MATLAB Central File Exchange. Retrieved December 15, 2021.

It contains both synthetic as well real mixtures. For synthetic mixture, 2 channel recording of 3 sources (source 1,2,3), sampled at 8 KHz of 6 second duration in above mentioned package was used. For real mixtures, 2 channel recording of two sources (rsm2_mA and rsm2_mB) with similar parameters we used.

Base paper(synthetic mixtures)

SDR(Db)	Kurtosis	Negentropy	Maximum likelihood
Source 1	49.7015	45.5588	19.4378
Source 2	45.0607	43.0009	19.2976

Our results(synthetic mixtures)

SDR(Db)	Kurtosis	Negentropy	Maximum Likelihood
Source1	17.69	17.69	17.69
Source2	16.27	18.88	18.88
Source3	25.551	30.551	30.58

Base paper (Real mixtures)

SDR	Kurtosis	Negentropy	Maximum Likelihood

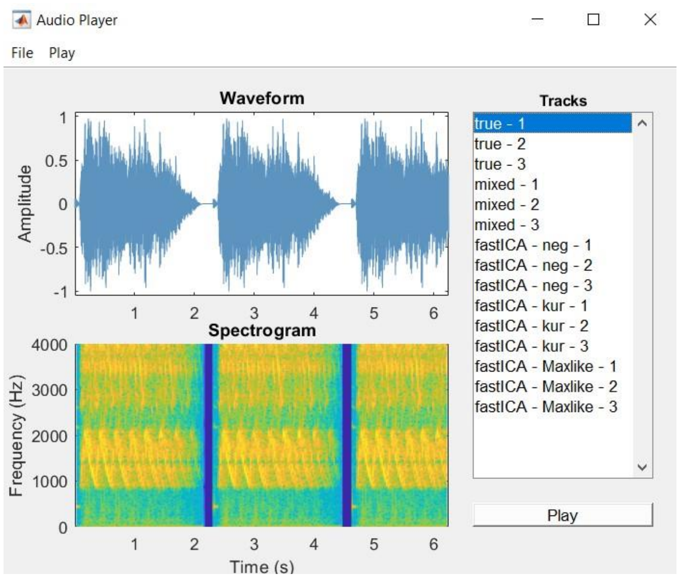
Source 1	-13.9451	-13.9581	-18.3843
Source2	-18.6029	-18.6478	-13.94

Our results (Real mixtures)

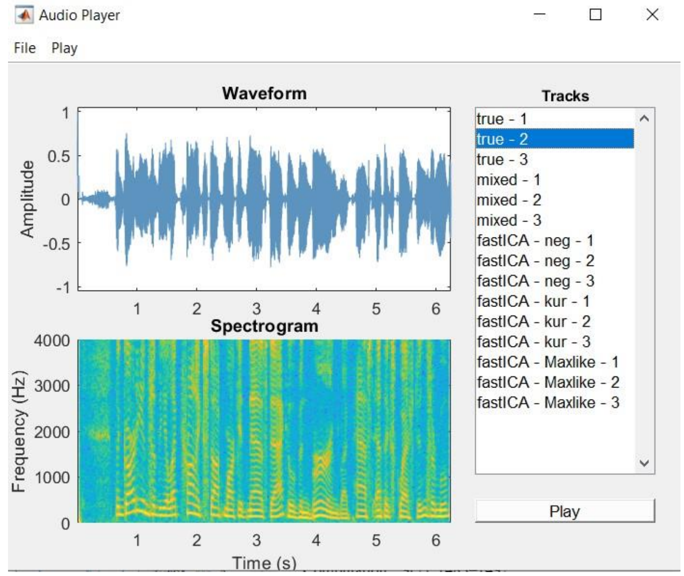
SDR	Kurtosis	Negentropy	Maximum Likelihood
SOURCE 1	-23.922	-23.67	-24.495
SOURCE 2	-25.87	-26.48	-22.0013
SOURCE3	-23.84	-20.12	-24.461

Spectrograms for our results(synthetic mixtures)

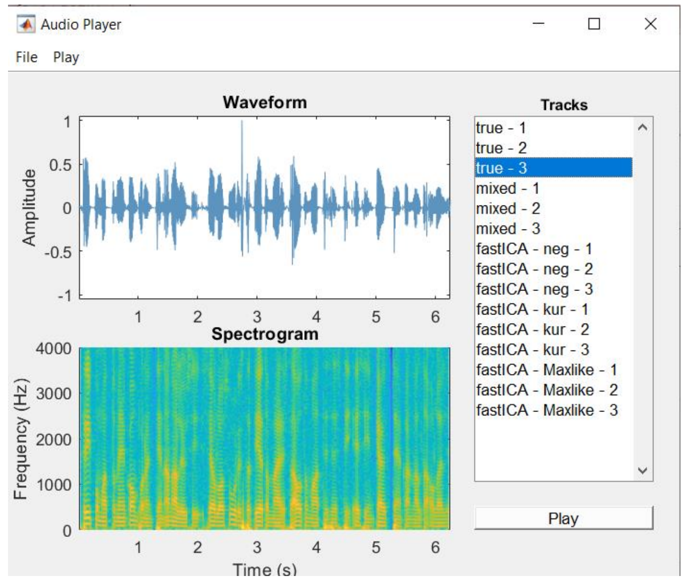
True source 1 :



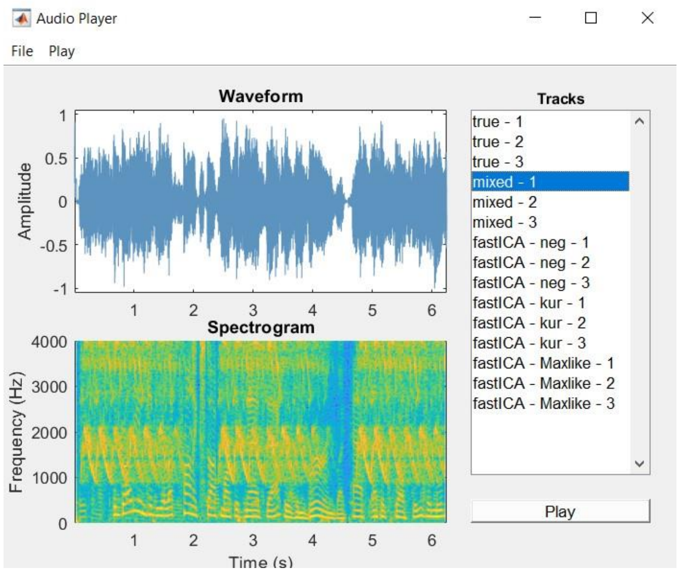
True source 2:

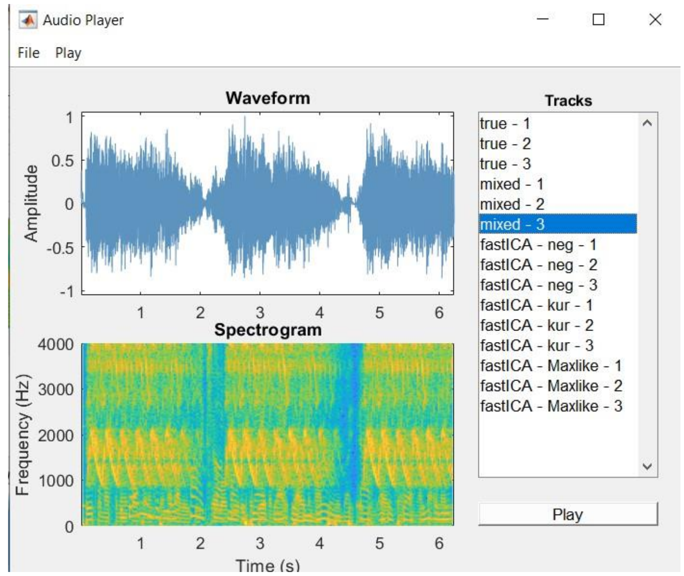


True source 3:

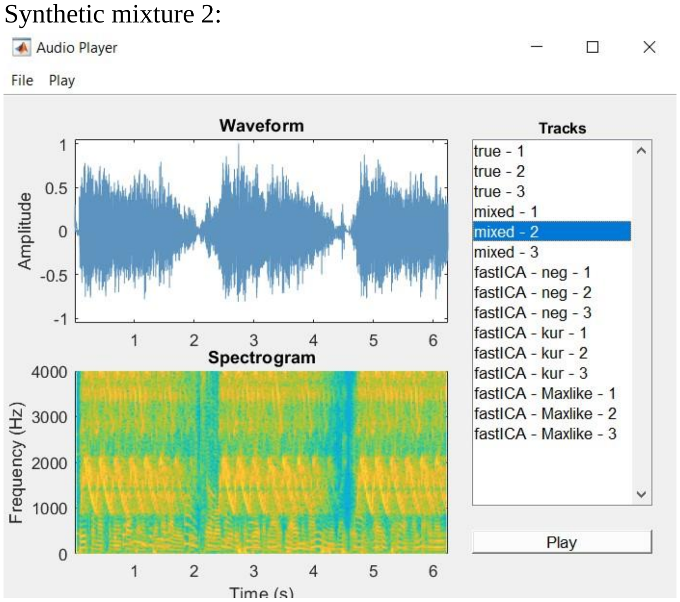


Synthetic mixture 1:

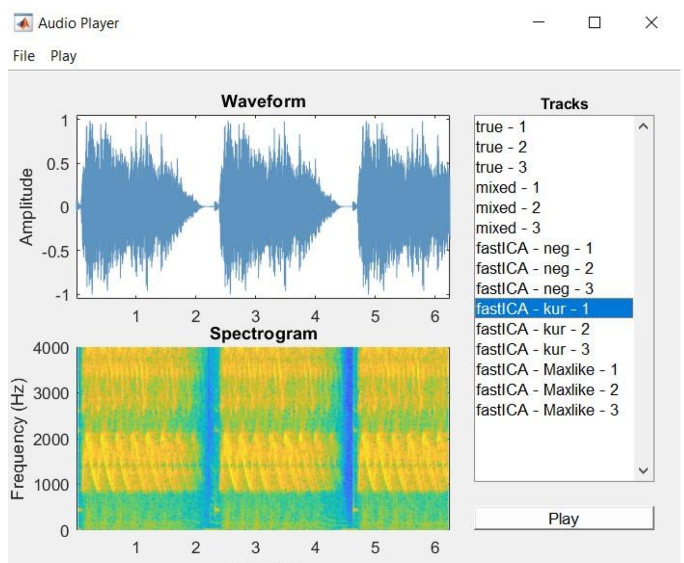




Synthetic mixture 2:

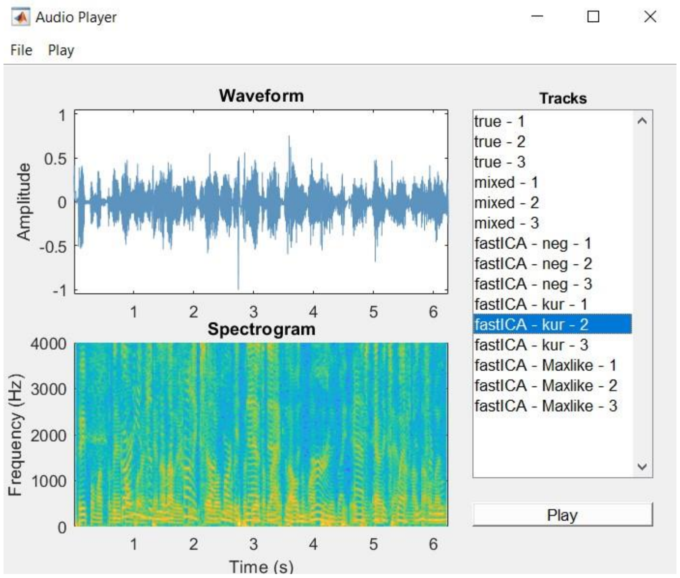


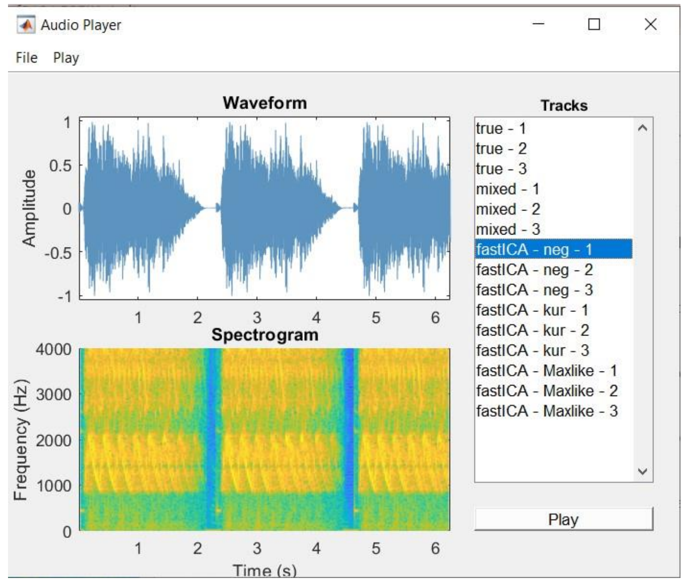
Kurtosis Source1



Synthetic mixture 3:

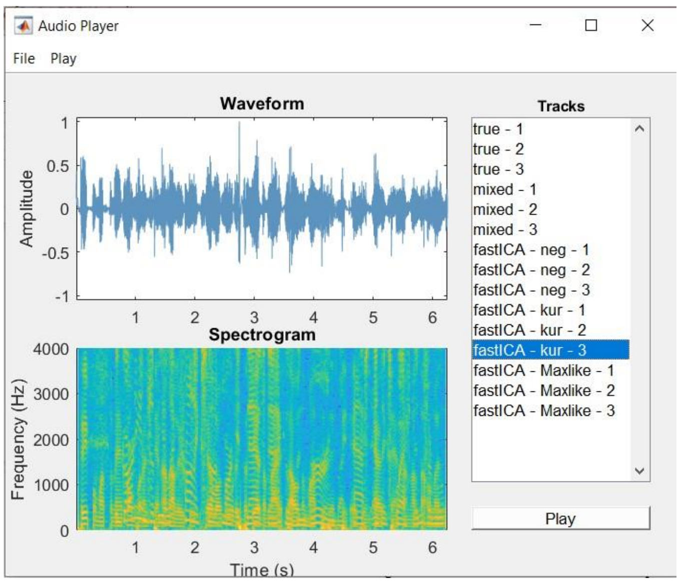
Kurtosis Source 2:

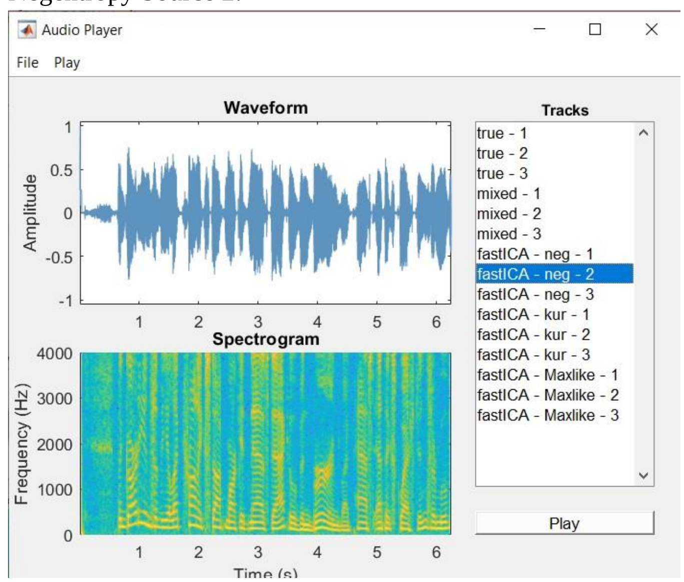




Kurtosis Source 3:

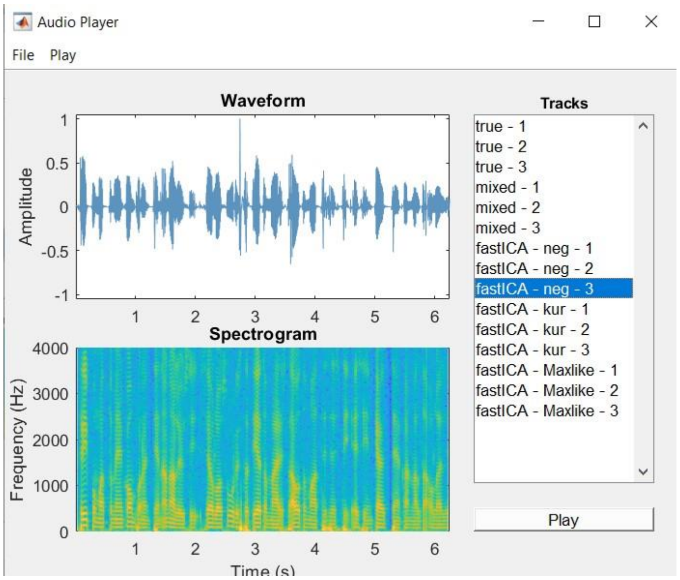
Negentropy Source 2:



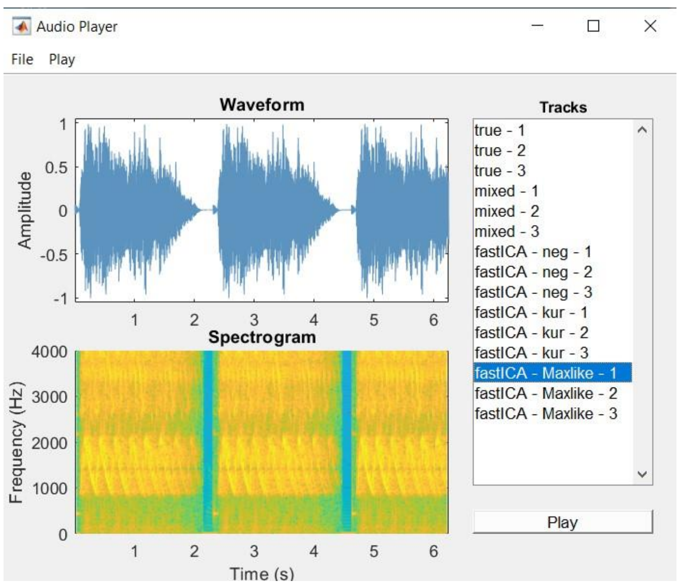


Negentropy Source 3:

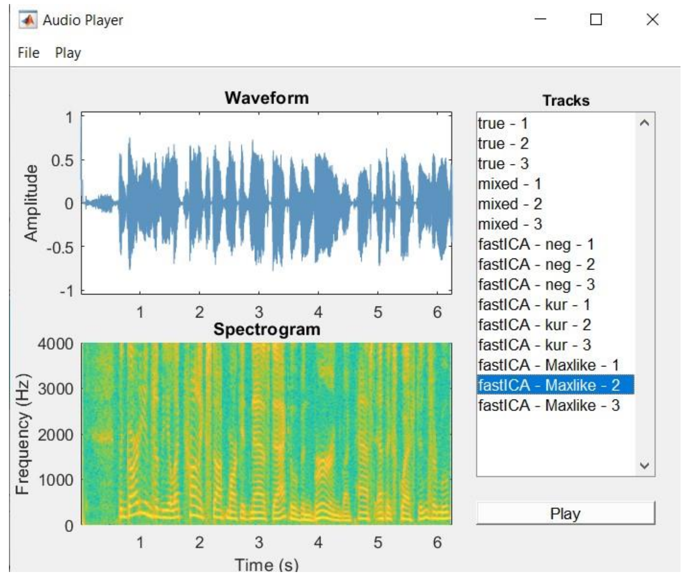
Negentropy Source 1 :



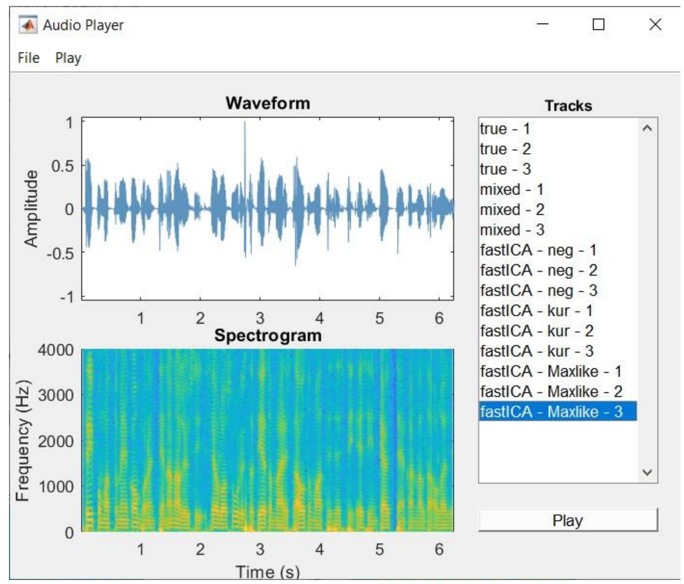
Max Likelihood Source 1:



Max Likelihood Source 2:

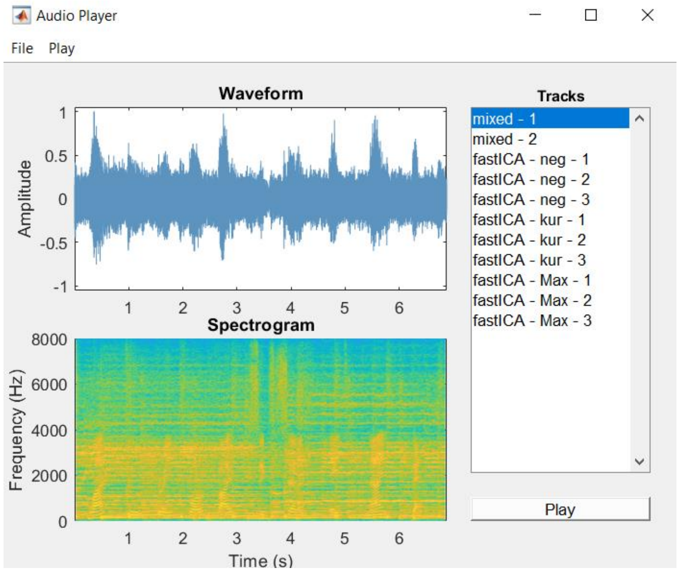


Max Likelihood Source 3:

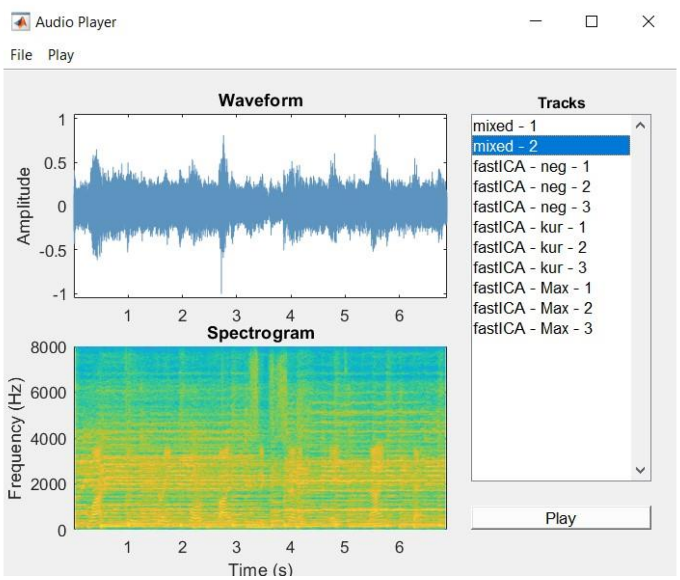


Real time mixed signals

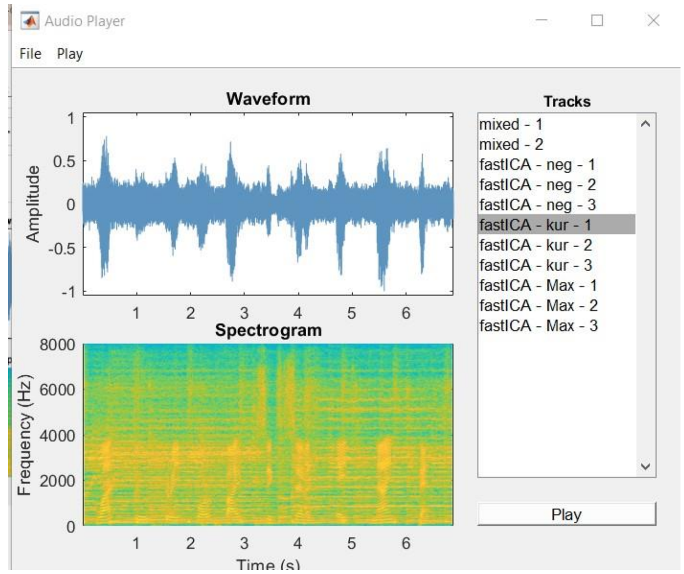
Mixture 1:



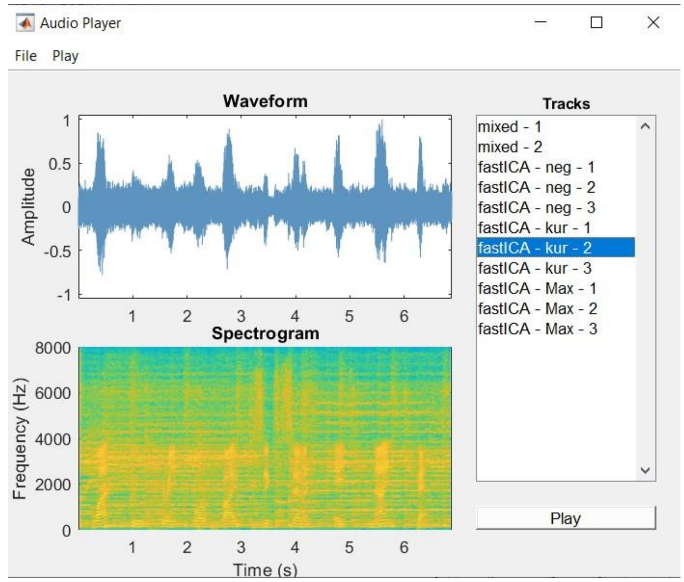
Mixture 2:



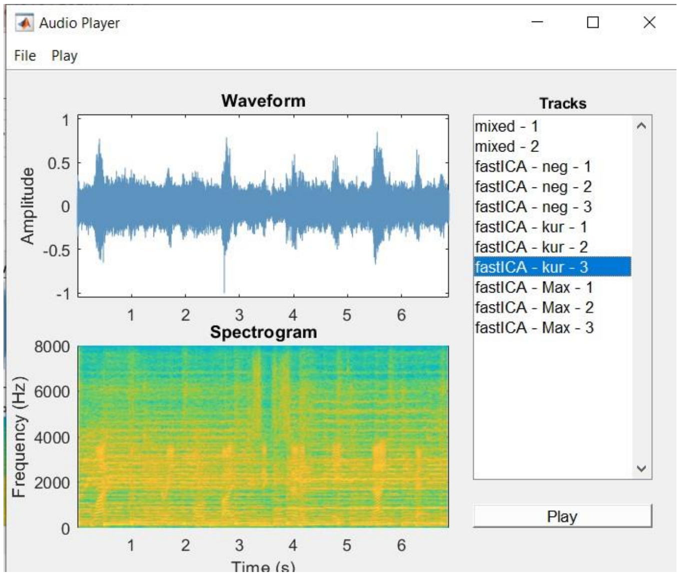
Kurtosis 1:

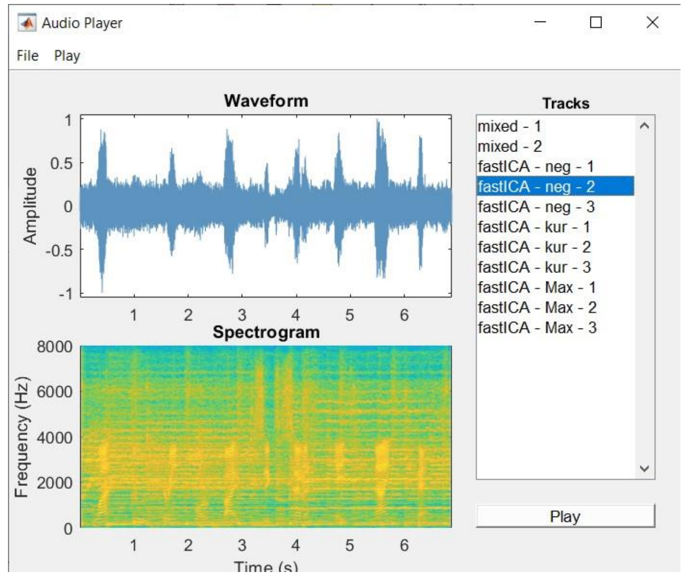


Kurtosis 2:



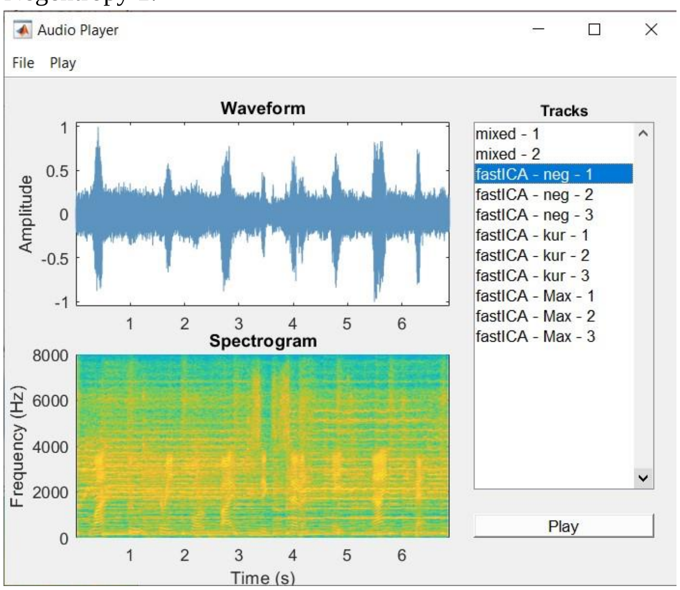
Kurtosis 3:

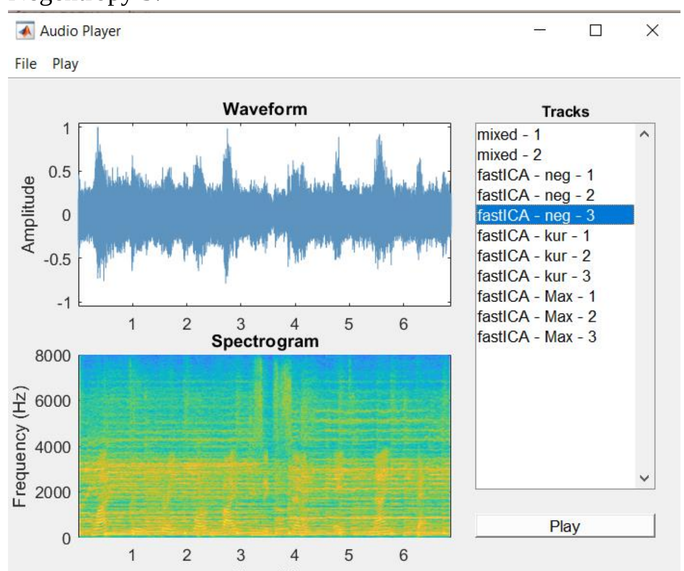




Negentropy 1:

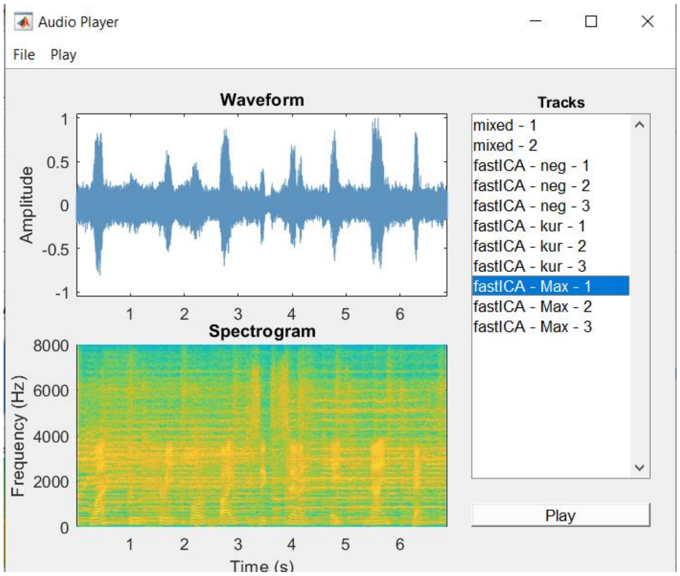
Negentropy 3:

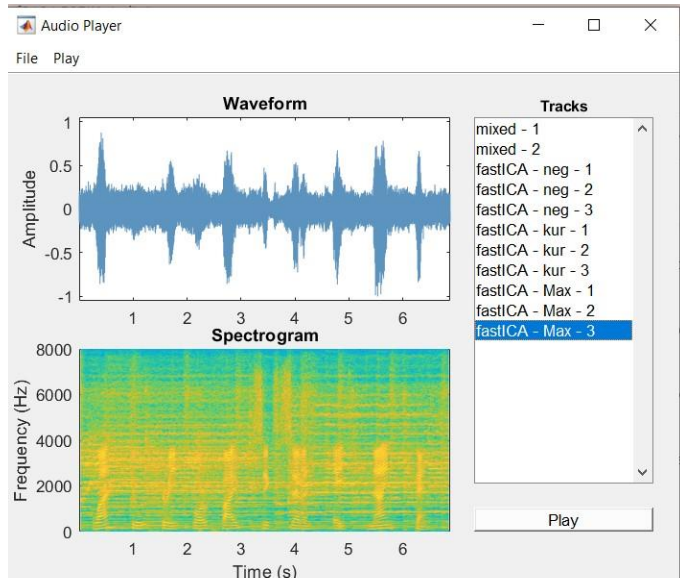




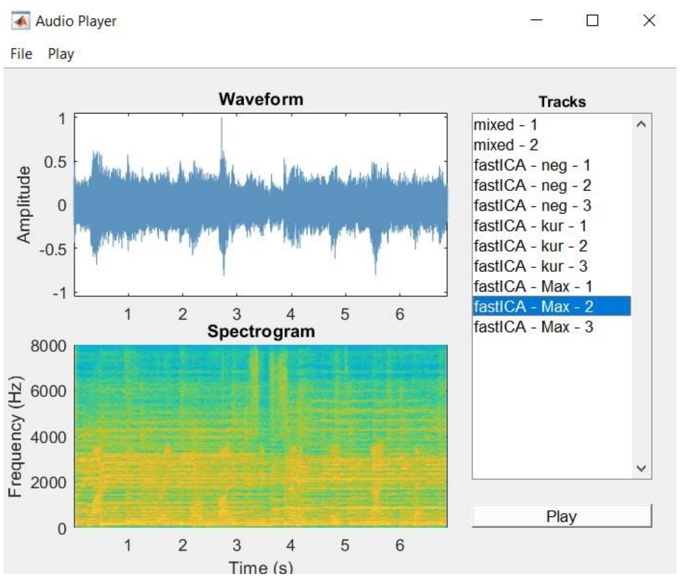
Negentropy 2:

Max Likelihood 1:





Max Likelihood 2:



Max Likelihood 3:

IV. Novel Idea And Further Scope

- We had referred to the paper regarding "Consistent ICA: Determined BSS Meets Spectrogram Consistency" by Yatabe(2020)[6] which resolves the permutation problem of BSS and we intend to implement it here for the same.
- Although Spectrogram Consistency has been used to resolve the issue, we intended to use statistical correlation in frequency domain to reorder the outputs of the BSS corresponding to another set of segregated voice samples of the same sources.
- Although the present algorithm is efficient for artificially mixed sources; we intend to make it more efficient by usage of methods with variable windowing in frequency domain such as STFT or using wavelet transforms on the resultant outputs of the FICA algorithm.
- An alternative approach to increase real time efficiency of maximum likelihood FICA would be to use the following functions for computation:
  G = (Sk).*(tanh(cos(Sk)));
  Gp = 1 - (tanh (cos(Sk.^2)));
  And then applying post conditioning methods.

V. CONCLUSION

Although the base paper concludes that the Maximum Likelihood as the most efficient technique for real time signals followed by Negentropy followed by Kurtosis; our observations concur with it (of the order of efficiency of Negentropy being greater than Kurtosis) except that the Max Likelihood is not the most efficient in our observations as FastICA for Max likelihood contrast function remains to be optimized by us.(The contrast function for Max Likelihood has not been mentioned in the base paper).

Thus, the proposed changes may reflect better efficiency of FICA based on Max Likelihood.

Acknowledgment

We are extremely grateful to our course instructor Dr. Anantha Krishna Chintanpalli for inspiring us to take up this project and guiding us throughout the course.

REFERENCES

[1] M. Kumar, V. E. Jayanthi, "Blind source separation using kurtosis, negentropy and maximum likelihood functions", International Journal of
Speech Technology, 23, 13 - 21, 2020.

[2] Bingham, E., & Hyvarinen, A. (2000). "A fast fixed point algorithm for independent component analysis of complex valued signals", International Journal of Neural Systems, 10(1), 1–8.

[3] Hyvarinen, A., & Oja, E. (1997a). "A fast fixed point algorithm for independent component analysis. Neural Computation", 9(7), 1483–1492.

[4] Hyvarinen, A., & Oja, E. (1997b)." One-unit learning rules for independent component analysis". In Proceedings of advances in neural information processing systems, 9.

[5] Hyvarinen, A., & Oja, E. (2000). "Independent component analysis: Algorithms and applications. Neural Network", 13(4–5), 411–430.

[6] Kohei Yatabe, "Consistent ICA: Determined BSS Meets Spectrogram Consistency",IEEE SIGNAL PROCESSING LETTERS, VOL. XX, NO. X, XXXXX 2020.