p-ISSN: 2301-5373
e-ISSN: 2654-5101

Fast Independent Component Analysis (FastICA) in Separating Vocals and Instruments in the Art of *Geguntangan*

Angga Pramana Putra[a1], I Gede Arta Wibawa[a2]

[a]Informatics Department, Faculty of Mathematics and Natural Sciences,
Udayana University,
Jalan Raya Kampus Unud, Jimbaran, Bali, 80361, Indonesia
[1]sunghangga@gmail.com

Abstract

Geguntangan is pesantian in religious ceremonies in Bali accompanied by gamelan music. The human sense of hearing tends to have limitations, which causes not all vocals mixed with gamelan to be heard clearly. Therefore we need a system that can be used to separate vocals with gamelan in the geguntangan. Separation of sound sources is categorized as Blind Source Separation (BSS) or also called Blind Signal Separation, which means an unknown source. The algorithm used to handle BSS is the Fast Independent Component Analysis (FastICA) algorithm with a focus on separating the sound signal in a wav-format sound file. FastICA algorithm is used for the sound separation process with the value parameter used is Mean Square Error (MSE). From the simulation results show the results of MSE calculations using the mixing matrix [0.3816, 0.8678], [0.8534, -0.5853] obtained the results for the FastICA method, the MSE value is 3.60 x 10^{-5} *for the vocal and 1.71 x 10*$^{-6}$ *for the instrument.*

Keywords: *Blind Sources Separation, Fast Independent Component Analysis, Audio Signal Processing, Mean Square Error*

1. Introduction

Geguntangan is *pesantian* in religious ceremonies in Bali accompanied by *gamelan*. *Geguntangan* often used in religious traditions to accompany the ceremony and also for public entertainment. However, the voice mixed between vocal and *gamelan* causes difficulty in learning the vocal in the *geguntangan*.

The human sense of hearing tends to have limitations, which causes not all vocals mixed with *gamelan* to be heard clearly. Therefore we need a system that can be used to separate vocals with *gamelan* in the *geguntangan*. Separation of sound sources is categorized as Blind Source Separation (BSS) or also called Blind Signal Separation, which means an unknown source [1].

BSS is a way to separate the mixed signals into several forming signals, without information about the number of signal sources, or the process of mixing the signals. BSS utilizes the different signal characteristics before the sensor is detected and the information obtained due to differences in the angle of arrival and distance of the sensor. The sensor used is a microphone. There are two techniques in recording sound, namely, single-channel and multichannel. Single-channel is a recording technique that uses a single sensor, and multichannel is a recording technique that uses more than one sensor.

The algorithm used to handle BSS is the Fast Independent Component Analysis (FastICA) algorithm. This algorithm focuses on the separation of sound signals in a Wav format sound file that has two or more sounds mixed so that separate sound results are recorded. The reason for using the wav format is that the wav format file contains sounds that are not compressed [2].

This research uses the FastICA Algorithm. FastICA algorithm can distinguish the elements or components of the signal mixture independently [3], which in this research uses only two sound sources, namely vocal sound signal and instrument. Before doing the FastICA process, the first thing to do is to do some pre-processing. Pre-processing in question is centering and whitening.

Besides that, there is a Non-negative Matrix Factorization (NMF) algorithm, which can also be used in Blind Source Separation (BSS). NMF can process data on a large scale using a matrix factorization model compared to the classical algorithm. Some of the advantages of NMF are easy implementation, good interpretability of decomposition results, and small storage space [4]. Most of the research on the NMF method to overcome the limitations of the NMF method, namely the fact that NMF is a temporal magnitude model only [5].

Another algorithm commonly used to handle the Blind Source Separation (BSS) approach for signal processing problems is the Sparse Component Analysis (SCA) algorithm. SCA utilizes less signal to extract the source and has higher precision in terms of signal separation. If the signal source has a Gaussian distribution, which is a noise model that follows a standard normal distribution with zero averages, the SCA can still extract rare sources effectively. SCA is also a promising approach for BSS when there are fewer sensors from the source [6]. But, FastICA can sparate more effectively than SCA because the FastICA algorithm can define how many and what specifically source will be separate by using mixing matrix.

The FastICA algorithm is used for the sound separation process with the value parameter used is Mean Square Error (MSE) or see the similarity between the output results with the input selected to test the sound output results. In speech signal recognition, the process of verbally listening to the output results compared to voice input is used to measure the parameters of success. Besides that, by comparing the input and output signals, it can be seen the results of the sound separation process.

2. Related Work

Blind Source Separation is one way to blindly separate a mixed-signal into several forming signals [7]. BSS is one of the techniques used to obtain sources from blind mixing. Because every mixed component can be reconstructed again into its constituent signals, many algorithms can be used to solve problems in Blind Source Separation.

FastICA algorithm in its testing of Blind Source Separation is better than the ability of the PCA algorithm, and NMF [8]. In testing the algorithm to determine a superior algorithm between FastICA, PCA, and NMF using the parameters of signal to interference ratio (SIR), signal to distortion ratio (SDR) and signal to artifact ratio (SAR). The greater the value of the parameter, the better the algorithm used for the BSS method, and vice versa. By using the three parameters, the FastICA method is superior to the other two methods tested.

There are two types of sound sources used in the FastICA Algorithm, namely single-channel and multichannel. FastICA can run using both kinds of sound sources by modifying it with FSS-Kernel (Finite Support Samples Kernel), where nonlinear functions are replaced with PDF (Probability Density Function). The results of the study are that FastICA with FSS-Kernel modification can effectively separate sound sources with more than one mixed-signal [9].

FastICA can be used to improve sound quality by reducing noise or noise in an audio signal. This can be done by separating noise with sound signals. Another method that is usually used to improve the quality of sound signals from noise is the Butterworth method. The results of this study found that the results of the separation using the FastICA algorithm are better in separating noise with sound signals compared to Butterworth [10].

3. Literature Review

3.1. *Geguntangan*

Geguntangan is a pesantian accompanied by *gamelan* which is usually used in religious events in Bali. In Bali, there is a diversity of *gamelan* instruments and the principles of playing them. Based on the criteria applied, the types of instruments, and the playing style can be divided into ± 25-30 genres of shaking. Bali has a strong character; the most prominent is the fast rhythm of music; this is because there is a small cymbal shaped device called Ceng-Ceng. Ceng-ceng is an instrument that sounds loud and is played very fast.

3.2. Noise

Noise can be defined as an unwanted signal that appears in communication, measurement, perception, or processing of a sign that contains information. The success of a noise processing method depends on its ability to characterize and model the noise process and to use noise characteristics advantageously to distinguish signals from noise.

 p-ISSN: 2301-5373
e-ISSN: 2654-5101

3.3. Sound Signal Processing of *Geguntangan*

In analyzing signals, several processes are needed, including the process of normalization, feature extraction, and classification. Feature extraction or identifier retrieval can be done in the time domain and frequency domain. The following is an explanation of several signal processing techniques.

3.3.1. Blind Source Separation

Blind Source Separation is one way to blindly separate a mixed-signal into several forming signals [7]. BSS is one of the techniques used to obtain sources from blind mixing. Because every mixed component can be reconstructed again into its constituent signals. The number of acoustic signals can be formulated:

$$Xi(t) = A^{*}Z_n(t) + n(t) \quad (1)$$

Where x = [x1, x2, ... xm] is a vector that represents the measured signal xi. z = [z1, z2, ... zn] are vectors that represent sources. A is the sum matrix that occupies the full column. BSS is used to find the A-1 matrix because A-1x is the same as the z source matrix with measured x.

3.3.2. Mean Square Error (MSE)

MSE is a method for measuring the difference between the estimator (reconstruction signal) and the true value (baseline signal) of estimated strength. By calculating the value of MSE, the difference between the original signal and the reconstruction signal will be obtained, which can be shown in the equation below [11].

$$MSE = \frac{1}{n}\int_i^t (S - S_n)^2 dt \quad (2)$$

3.4. Pre-processing

Before doing the FastICA process, the first thing to do is to do some pre-processing so that the FastICA process can run well. Two types of pre-processing in question are centering and whitening. The centering process is the process of centralizing the data, which makes the mixed-signal value (X) into a mixed-signal which has a mean of zero. Centering formula:

$$X_c = x - m \quad (3)$$

Xc is the observation signal from the centering result, x is the observation signal, while m is the mean. Next, from whitening, we get a new vector whose variance is equal to one. The formula for whitening is as follows:

$$X_w = V_x \quad (4)$$

Where Xw is the whitening mixing matrix, V is the whitening matrix with the equation:

$$V = ED^{-1/2}E^T \quad (5)$$

E is the orthogonal matrix of the eigenvector E {xT}, and D is the diagonal matrix of the eigenvalue.

3.5. Fast Independent Component Analysis (FastICA)

Fast Independent Component Analysis (FastICA) is the most widely used technique for solving problems from Blind Source Separation (BSS). The general model of FastICA is a source generated through a linear transformation in the presence of additional noise. Suppose there are independent statistical signals N, si (t), i = 1, ..., N. with the assumption that the source cannot be directly observed and each signal si (t) is a realization of the probability distribution at time t. Through observation on the N sensor then obtained a set of N observation signals xi (t), i = 1, ..., N. which is the result of mixing from the source. A fundamental aspect of the mixing process is that sensors must be separated so that each sensor records mixing differently. That way the mixing process according to Ganesh R. Naik and Dinesh K Kumar can be modeled with the following matrix:

$$(t) = As(t) \quad (6)$$

In the equation above, A is the mixing matrix, and x (t), s (t) are two vectors representing the observed signal and the source signal. Indirectly this is what is meant by the blind, i.e., and there is no information on the mixing matrix and each source. The main goal is to get the original signal si (t), only from the observed signal xi (t). Furthermore, to estimate the source can be obtained by first using unmixing matrix W, where W = A − 1.

Therefore, the estimated source signal, separate (t) can be obtained from the equation below:

$$\hat{s}(t) = (t) \tag{7}$$

4. Methods

4.1. Research Design

The research conducted is an experimental study comparing MSE on the FastICA algorithm in vocal separation with *gamelan* in the *geguntangan*. The shaking file obtained in the form of wav will be centered and whitened to get a new vector with the same variance as one, finally doing the FastICA process for the centering and whitening vector.

4.2. Research Location and Time

This research was conducted at the Network Laboratory of the Informatics Department, Faculty of Mathematics and Natural Sciences, Udayana University for three months.

4.3. Data Collection Methods

The tools used in this study include a recorder using a Tascam DR-40 recorder. The first recording is done on the vocal sound, followed by recording the sound of the instrument. The programming language used to process the sound signals of the *geguntangan* is Python, and laptops as a tool to use the software to be used. The data used in this study include vocal recording and its shaking instruments in the wav file format. The amount of data used is 80 data. There are 40 vocal data with different sources, namely two men and two women. Instrument data as much as 40, which is an accompaniment from the vocal Data is collected from vocal retrieval, followed by data retrieval of instruments that have been adjusted to their vocals. Adjust the vocals with the instrument by using a headset on the drummer, and the accompaniment can be harmony with the vocal.

4.4. Research Stages

Table 1. Research Stages

Description	Output	Performance Indicator
Collection and determination of recorded data.	Information from a recording of a shaking that has a slight other sound disturbance.	The availability of recording data of the *geguntangan* which has a few different sound disturbances.
Pre-processing centering on the recording data of the shaking.	Data recording with minimal noise.	Obtained a matrix of data recording with minimal noise.
Pre-processing whitening on the centering result matrix.	The identity matrix results from the transformation of the recorded signal.	Obtained an identity matrix from the transformation of the recorded signal.
FastICA process on the pre-processing matrix.	Separate recording signal between vocal and instrument.	Obtained a separate recording signal between vocal and instrument.
The process of testing the data of FastICA results with MSE.	MSE value as a parameter of the success of the recording signal separation.	Obtained MSE value as a parameter of success in recording signal separation.
Data analysis.	MSE value information.	Obtained the best algorithm information in signal separation.

4.5. Pre-processing Data

4.5.1. Pre-processing Centering of the Recording *Geguntangan*

 p-ISSN: 2301-5373
e-ISSN: 2654-5101

Signal recording shakes with the average signal itself to make the average of a data zero. Calculations use functions from python. The results of the centering process are in the form of a matrix.

4.5.2. Pre-processing Whitening of the Recording *Geguntangan*

The matrix from the centering results will be carried out the process of transforming the data into new data forms in order to obtain the value of the covariance matrix in the form of an identity matrix using the functions contained in python.

4.6. FastICA Process

4.6.1 Mixing Matriks Process

Before the signal is processed using the FastICA algorithm, the previous signal is mixed using a mixing matrix. Matrix mixing is restored in a random manner which is later used to mix vocal signals and instrument signals. The output generated from this process is in the form of vocal signals with mixed instruments.

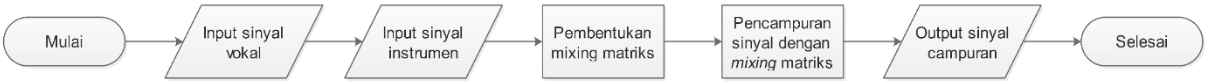

Figure 1. Flow chart of mixing matrix

4.6.2 FastICA

The FastICA process can be illustrated with a flow chart, as shown in Figure 1, with mixed-signal input that has been made with a mixing matrix.

Figure 2. Flow chart of FastICA

In this process, Y values are obtained as estimation signals and W as matrix unmixing. The sound separation process is carried out with the following equation:

$$Y(t) = W\, Xn(t) \tag{8}$$

To determine the components of the FastICA algorithm, the asymmetrical method is required, the following steps:

1. Determine the number of independent components or variables m.
2. Choose the initial value of the complex vector w, where y is the product of w with xn.
3. Find the values of g (nonlinearity) and g 'using the following equation:

$$g(y) = y\, exp\left(-\frac{y^2}{2}\right)$$

$$g(y) = (1 - y^2)\ exp\left(-\frac{y^2}{2}\right) \tag{9}$$

4. Calculate the new w value with the equation:

$$w \leftarrow E\{zg(w^t z)\} - E\{g'(w^t z)\}w \tag{10}$$

5. Perform the iteration process as in the following equation:

$$W \leftarrow \frac{W}{\|W\|}$$

$$W \leftarrow \frac{3}{2}W - \frac{1}{2}WW^T W \tag{11}$$

6. Obtained the value of W or unmixing matrix.

Next, to the equation Y (t) = W Xn (t) is carried out by entering the value of W to get the estimation signal or Y.

4.7. Testing Method

4.7.1. Testing the Quality of the Results of the FastICA Process with MSE

MSE is a method for measuring the difference between the estimator (reconstruction signal) and the real value (baseline signal) of estimated strength. By calculating the amount of MSE, the difference between the original signal and the reconstruction signal will be obtained, which can be shown in the equation below [11]. The variables needed for MSE testing are the amount of data, the original signal, and the estimation signal.

$$MSE = \frac{1}{n}\int_i^t (S - S_n)^2 dt \quad (12)$$

4.7.2. Data Analysis

Data obtained from the results of the MSE calculation will determine the quality of the results of the separation. If the MSE value is getting closer to zero, the signal from the separation results is more similar to the original signal, and vice versa.

5. Result and Discussion

5.1. Analysis of Blind Source Separation Simulation Results

To get the results in the form of signal separation from the application of Blind Source Separation is done by python simulation. Figure 3 shows the source signal, the mix, and the results of the separation of the first female voice with the instrument using the FastICA method. The X-axis is time, and the Y-axis is the amplitude of the signal.

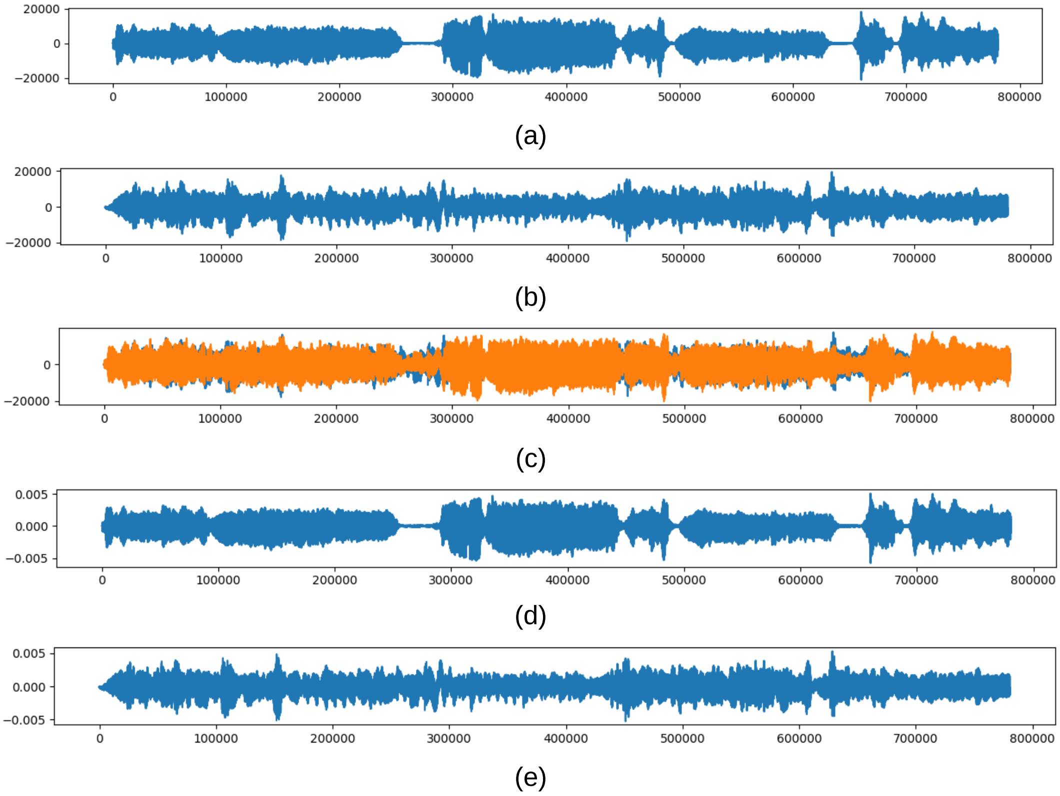

Figure 3. Results of plot of vocal source signal (a), instrument source signal (b), mixed signal (c), separation vocal signal (d) and separation instrument signal (e)

5.2. MSE analysis on the application of the BSS Algorithm

Tests of the two methods use the MSE parameter to test the error between signals by comparing the source signal with the signal separation. The smaller the MSE value, the better the method used in signal separation, and vice versa. MSE calculation results using the mixing matrix [0.3816, 0.8678], [0.8534, -0.5853]. Then it is implemented on 80 data, and conclusions are obtained by finding an average of each test result obtained. The following is presented an average of 80 test data results:

Table 2. The average MSE calculation results of the FastICA method

	Mixed Source	MSE	
		Vocal	Instrument
Person 1	Source1	2.35861E-05	3.80157E-06
	Source 2	3.10381E-05	5.34357E-06

Person 2	Source 1	5.14622E-05	1.12641E-06
	Source 2	2.58266E-05	5.46144E-07
Person 3	Source 1	2.34617E-05	7.79372E-07
	Source 2	2.52862E-05	5.18586E-07
Person 4	Source 1	6.41834E-05	1.05842E-06
	Source 2	4.34488E-05	5.36542E-07

From table 2, it can be concluded for the FastICA method, and the MSE value is 3.60 x 10^{-5} for the vocal and 1.71 x 10^{-6} for the instrument. If the MSE value is getting closer to zero, it shows that the signal from the separation is more similar to the original signal, which indicates the better quality of the signal from the separation. For the grade of MSE value, show in table 3.

Table 3. Grade MSE

MSE	**Akurasi Error**
< 0.15	Very good
0.15 sampai 0.30	Good
0.30 sampai 0.60	Buruk
0.40 sampai 0.60	Very bad
> 0.60	Unaccepted

(Source: Corwin & Lesch, 2016)

6. Conclusion

Obtained MSE calculation results by using a mixing matrix [0.3816, 0.8678], [0.8534, -0.5853] obtained results for the FastICA method, the MSE value is 3.60 x 10^{-5} for the vocal and 1.71 x 10^{-6} for the instrument. The results of the separation depend on the mixing matrix used. If the MSE value is getting closer to zero, it shows that the signal from the separation is more similar to the original signal, which indicates the better quality of the signal from the separation. Can alse be seen in table 3, the resulting MSE value can be categorized very good. Because the value of MSE from separating is less than 0.15.

References

[1] X. Yu, D. Hu, and J. Xu, *Blind Source Separation Theory and Applications*, First. Solaris South Tower: John Wiley & Sons Singapore Pte. Ltd., 2014.

[2] A. Westner and V. M. Bove Jr., "Blind separation of real world audio signals using overdetermined mixtures," *Proc. First Int. Conf. Indep. Compon. Anal. Blind Source Sep. ICA'99*, pp. 251–256, 1999.

[3] C. Tony and J. Shen, *Image Processing and Analysis Variational, PDE, Wavelet, and Stochastic Methods*, Second., no. c. Philadelphia: The Mathworks, Inc., 2005.

[4] Ganesh R. Naik and W. Wang, *Blind Source Separation Advances in Theory, Algorithms and Applications*, Signals an. Sydney: Springer Heidelberg New York Dordrecht London, 2014.

[5] B. Schuller, *Audio Source Separation*, Signals an. Ibaraki: Springer International Publishing, 2013.

[6] F. Georgiev, F. Theis, and A. Cichocki, "Blind source separation and sparse component analysis of overcomplete mixtures," pp. V-493–6, 2004.

[7] D. N. Hayati and D. Arifianto, "Penerapan Independent Component Analysis (ICA) Untuk Pemisahan Sinyal Suara Mesin Berputar di PT. Gresik Power Indonesia (The Linde Group)," no. 2009, pp. 1–9, 2010.

[8] N. Hassan and D. A. Ramli, "A Comparative study of Blind source separation for Bioacoustics sounds based on FastICA, PCA and NMF," *Procedia Comput. Sci.*, vol. 126, pp. 363–372, 2018.

[9] P. He, T. She, W. Li, and W. Yuan, “Single channel blind source separation on the instantaneous mixed signal of multiple dynamic sources,” *Mech. Syst. Signal Process.*, vol. 113, pp. 22–35, 2018.

[10] S. H. Chiu and C. P. Lu, “Noise separation of the yarn tension signal on twister using FastICA,” *Mech. Syst. Signal Process.*, vol. 19, no. 6, pp. 1326–1336, 2005.

[11] A. A R, “Penggunaan Frekuensi Sesaat Untuk Deteksi Pola Suara Kerusakan Motor Listrik,” ITS, Surabaya, 2006.

[12] D. Corwin and S. Lesch, “Validation of the ANOCOVA model for regional-scale ECa-ECe calibration,” *Soil Use Manag.*, 2016.