Comparison of Wavelet and Hybrid (ICA and Wavelet) Methods in Separation of Sound Frequency in Forensic Audio Models

Ane Prasetyowati[1]*, Noor Suryaningsih[1] and Vector Anggit Pratomo[1]

[1]Department of Electrical Engineering, Faculty of Engineering, Universitas Pancasila, Indonesia

*Corresponding author: ane_prast@univpancasila.ac.id

Abstract. Audio forensic is a process to improve an authenticity of a voiceprint evidence. But we must understand that voiceprint has a lot of noise inside it. In order to get a better quality of a voiceprint evidence, we need to reduce the amount of noise. There are several noise reduction methods that we can use. In this thesis, we will make a program with MATLAB by using two methods. First is wavelet transformation and then the hybrid process between Independent Component Analysis (ICA) and wavelet transformation. Wavelet transformation use high pass filter and low pass filter to reduce noise from a voiceprint. And ICA has a principle to estimate individual signal from mixtures of signal. After we use these methods to reduce noise from a voiceprint, we will see which methods work better for noise reduction process. The result from these two methods will be presented in SNR (Signal to Noise Ratio) output. From this research, wavelet transformation SNR output is higher than hybrid method SNR output. We can conclude that wavelet transformation is better than hybrid method for a noise reduction process. The best SNR output from this research is at level 5 with average value of 6.7274 dB for 30 seconds sample, 6.1256 dB for 60 seconds sample, and 6.0296 dB for 90 seconds sample.

Keywords: Audio forensic, noise reduction, transformasi wavelet, Independent Component Analysis (ICA), SNR.

1. Introduction

Audio forensics is an application of science used to investigate or investigate the authenticity of audio digital evidence in criminal or civil cases [1]. Audio digital evidence may be in the form of sound recordings. In the sound recording there is still noise that must be removed. In investigating various legal cases related to reconnaissance and or wiretapping issues in the audio forensic process, an identification process is required. In the process of identification, sound print conditions must be good, but in reality there must be a lot of noise contained therein. The noise contained in this sound print can interfere with the identification process. To be able to get optimal identification results, the noise in the sound print must be reduced by the denoising process. The same research is done by eliminating noise at the sound source using the wavelet method [2]. while other studies use the Blind Source Separation technique to separate the signal source with acoustic signals under water using the Natural Gradient ICA method [3]. In 2015, Mohanaprasad K. and Sankar Ganesh conducted research to solve the cocktail party problem because it reduced the quality of sound sources by using wavelet methods based on the Maximum Likelihood Estimation technique and ICA estimation-based wavelet methods to obtain a more independent signal [4]. By using wavelet transformation method based on Orthagonal Daubechies which will then be matched with Neural Network methods to be able to identify or recognize human voices. From various studies that have been carried out above, as well as looking at several methods that have also been used in research in an effort to reduce noise at the source of the sound signal, then in this study two methods will be used, namely the wavelet transformation method and the hybrid method (ICA and wavelet). This hybrid method is expected to make a better noise reduction process. On method and the hybrid method (ICA and wavelet).

2. Wavelet and Decomposition

Wavelet transform was first discovered around 1980, where this wavelet transformation was used as an alternative to the Short Time Fourier Transform for signal analysis. Wavelets are "short waves" with energy concentrated at any given moment. Wavelet has been used in the analysis of frequency time signal region in signal processing, function approximation, approximation in the completion of partial differential equations and etc. One function of the wavelet can be used for noise reduction in sound signals [5]. Wavelet is a mathematical function that divides data into several different frequency components, then an analysis is performed for each component using a resolution that is appropriate to the scale [6].

Wavelet decomposition process is a process carried out on a signal that is a signal entered into a system that has a high pass filter and a low pass filter that has a threshold value so that it can be divided into two parts simpler one. The wavelet decomposition process is simple, that is, the source signal will enter a series of systems in which the filter passes through and the filter passes down. Then you will get two outputs from each filter used. The two signals have different frequency domains. If you want to carry out the advanced decomposition process, then the output used is the output from the filter passes down. The decomposition process can be repeated as many times as needed. Figure 1 below explains this procedure, where x [n] is a signal that will undergo decomposition, h [n] is a low-pass filter, g [n] is a high-pass filter, and f represents the bandwidth of the signal at each level [7].

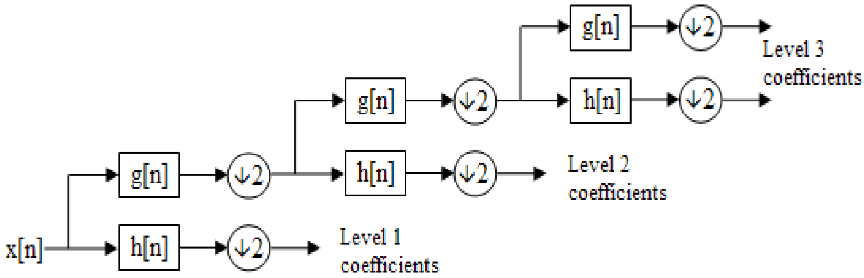

FIGURE 1. Filter banks on 3 level

Wavelets are formed using a low pass filter and a high pass filter. The outputs of these filters can be expressed mathematically by:

$$y_{HP}[k] = \sum u[m] \cdot g[2k - m] \tag{1}$$
$$y_{LP}[k] = \sum u[m] \cdot h[2k - m] \tag{2}$$

where m is the wavelet level and k is the wavelet constant. There are two kinds of threshold functions, namely soft threshold and hard threshold. The function we will use is the soft threshold:

$$T_{\lambda}^{soft}(d) = \begin{cases} d - \lambda \text{ jika } d \geq \lambda \\ d + \lambda \text{ jika } d \leq -\lambda \\ 0 \text{ jika } |d| \leq \lambda \end{cases} \tag{3}$$

where d is the coefficient detail and is the threshold value (> 0). Determining the threshold value is the main one in the noise reduction process. There are several ways to determine the threshold value, including universal threshold, SureShrink threshold, minimum threshold, and penelized threshold. In this research only Penelized threshold will be used. Penelized threshold is a type of threshold introduced by Birge and Massart (2001). In this method the detail coefficient can be sorted descending. Then to find the threshold value you can use the equation:

$$\lambda = arg_t^{min}\left[-\sum_{k=1}^{t} d_k^2 + 2\sigma^2 t\left(\alpha + \ln\frac{n}{t}\right)\right]; t = 1 \ldots n \tag{4}$$

where > 1 is the setting parameter for the penalty term in this method. The inverse DWT is then used to rebuild the signal to the original signal with the remaining information. The output of the filter passes below and the filter passes through in the signal reconstruction stage can be expressed by:

$$u[m] = \sum_{-\infty}^{\infty}\{y_{HP}[k] \cdot g[2k - m] + y_{LP}[k] \cdot h[2k - m]\} \tag{5}$$

3. Independent Component Analysis (ICA)

In the ICA model, we assume that each mixture as well as each independent component is a random variable. The observed value (λ), for example, the microphone signal from the cocktail party problem is a sample of this random variable. We can assume the mixed variable and also the independent component have zero average: if this is not true, then the observable variable can always be centered by reducing the sample average, which makes the model a zero average. The starting point of the assumptions for the ICA method is to assume the component is statically independent. It will be shown below that we must assume that the independent component must be in a non-gaussian distributed state. But in the basic model we cannot assume these distributions are known. To put it simply, we can also assume that the mixture of unknown matrix is square. Then, after matrix A is estimated, we can calculate the inverse value, say W, where we get a simple independent component, which is:

$$s = Wx \tag{6}$$

ICA is closely related to a process called Blind Source Separation (BSS). A "source" is interpreted as a genuine signal, for example independent components, such as speakers in cocktail party problems. "Blind" means that if there is a mixed signal, make a few assumptions on the source signal. ICA is one of the methods, perhaps the most widely used in doing blind source separation. In many applications, it would be more realistic to assume noise at the time of measurement, which means adding noise conditions to the model. To put it simply, we are eliminating the noise requirement, because the estimation of the noise-free model is already quite difficult and seems sufficient for various applications.

4. Pre-Processing of ICA

1. Centering (center average) The most basic and important thing in pre-processing is centering on the source signal value, we call x. For example, the average vector m = E {x} which makes x a zero-average variable. This has an impact on s which becomes zero. The way this process works is by reducing the signal by the average of the signal itself, the result of the reduction is a new signal that has a zero-mean average (*zero-mean*).

$$x_c = x - \bar{x} \tag{7}$$

This pre-process is only made to simplify the ICA algorithm. After estimating the mixture of matrix A with centered data, we can complete the estimate by adding back the average of vector s to the intermediate estimate of s.

2. Whitening Another useful pre-process for ICA is whitening all observed variables. This means that before applying the ICA algorithm (after the centering process), we change the observed vector x in a linear fashion so that we get a new vector˜ which is white, meaning that the components do not have the same correlations and variances (energies). In other words, the covariance matrix of˜ is the same as the identity matrix:

$$E\{\tilde{x}\tilde{x}^T\} = \mathrm{I} \tag{8}$$

Whitening transformation is always possible. The most popular method for the whitening process is to use eigenvalue decomposition (EVD) of the covariance matrix:

$$E\{xx^T\} = EDE^T \tag{9}$$

Where E is the orthogonal matrix of the eigenvector of $E\{xx^T\}$ and D is the diagonal matrix of eigenvalues, **D** = diag (d1, ..., dn). Note that E {xxT} can be estimated by standard means from existing samples x (1), ..., x (T). This whitening process can be done with the equation:

$$\tilde{x} = ED^{\frac{1}{2}} * E^T * x \tag{10}$$

Where the matrix $D^{-(1/2)}$ is calculated with a simple operating component as $D^{-(1/2)}$ = diag ($d_1^{-(1/2)}$, ..., $d_n^{-(1/2)}$,).

5. Algorithm of FastICA

FastICA is based on a one-point processing scheme to find the maximum non-gaussian noise value of wTx. It can also be derived as Newton's approach to measurement. Symbolized by g inheritance from non-quadratic G functions; for example the functions of equation (11) with their derivatives in equation (12) are:

$$g_1(u) = \tan h(a_1 u) \tag{11}$$
$$g_1'(u) = a_1\,(1 - \tan h(a_1 u)^2) \tag{12}$$

where $1 \leq a_1 \leq 2$ is the corresponding constant, the most commonly used a_1 = 1. The basic form of the FastICA algorithm is as follows:

1. Select the vector loading w with initial values, for example random values
2. Use the following equation to update the value of w :
   $$w^+ = E\{xg(w^T x)\} - E\{g'(w^T x)\}w \tag{13}$$
3. Normalize the value of w with
   $$w^+ = w^+ / \|w^+\| \tag{14}$$
4. If not centered, return to 2

It should be noted that the old convergence and new values of w go in the same direction, for example their dot products (almost) are equal to 1. Vectors do not have to be centered at one point, because w and –w are defined to have a direction the same one. This is because independent components can be defined only by multiplicative signs. Also keep in mind that it is assumed that this data has passed the prewhitened process.

6. The Proposed Model of Forensic Audio

The block diagram below explains the noise reduction process in the source signal mixed with the noise signal. There are two outputs from the two methods used. The results of the process show the same output, namely signal with reduced noise (noise reduction signal). The difference is in the series of processes carried out. At the first output, the method used is only wavelet decomposition. This method reduces noise by dividing the signal into high pass filter and low pass filter, then reduces the existing noise by using a threshold as a reference. Then reconstructed the remaining signals. While the other method is to use Independent Component Analysis (ICA) before it is processed by the wavelet method. The ICA method performs the process of separating the signal source with noise first with a process called Blind Source Separation (BSS) so that the noise in the source signal can be further reduced

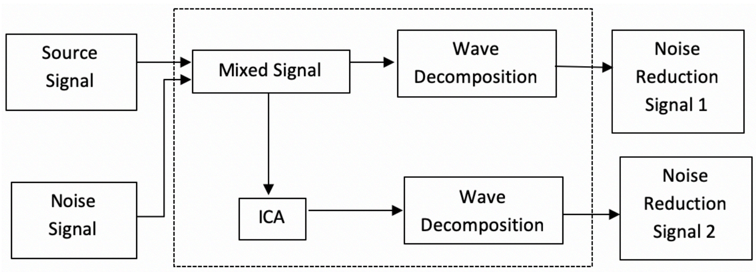

FIGURE 2. Block diagram system

7. Analysis and Result of The Models

7.1. Comparison of The Source Signal Form of Each Test Material with The Wavelet Process

There are 2 test materials used as test results, namely the source of the human voice signal and noise noise. But what will be used as a comparison is only the human voice. To test the ability of the method in performing the noise reduction process, each audio file will be given a noise of 10 dB. The data in this test consists of primary data and secondary data. Primary data is data obtained from sound sources that are mixed with direct recorded noise, while secondary data is data obtained from sound sources mixed with noise originating from

Youtube. Each audio file that is tested also has a different time, namely: 30 seconds, 60 seconds, and 90 seconds. Here is a picture of the audio file signal 1 which has been given noise.

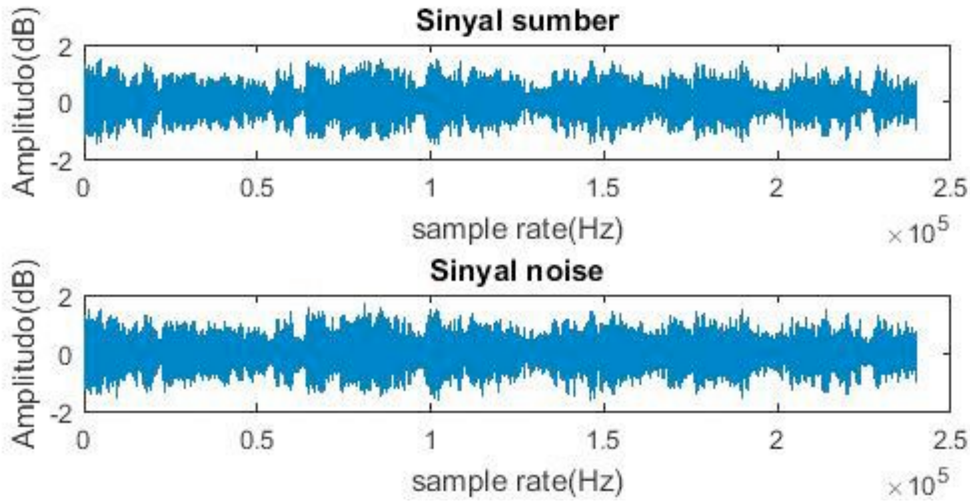

FIGURE 3. Signal source for audio file 1 given 10 dB noise

The picture below is Comparison of the signal shape between the original source signal and the noise reduction source signal for the audio file 1.

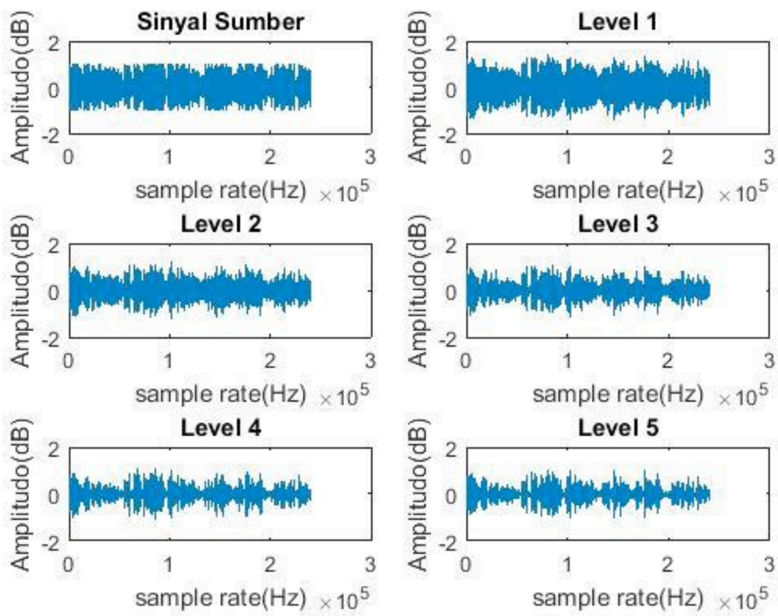

FIGURE 4. Reduction of noise in the audio file source signal 1 is performed by the Wavelet method for level 1 to level 5

Subjectively, it can be seen in the picture that the signal shape looks the same, then the application of the wavelet method to reduce noise causes the amplitude of the sound source signal to decrease.

7.2. Comparison of input SNR values with SNR outputs from the wavelet Haar level 5 method and the hybrid method at each SNR level

After testing, the results obtained compare the SNR values between the two methods with different times, namely: 30 seconds, 60 seconds, and 90 seconds. The test results can be seen in the following table:

TABLE 1. Results of comparison between SNR input and SNR output in audio file 1 with a sample time of 30 seconds

Level Wavelet Haar	SNR *Input*	SNR *Output with* method	
		Wavelet	ICA and *Wavelet*
1	-1.8734	-0.6729	-10.2062
2	-1.8734	1.9751	-7.7913
3	-1.8734	4.9298	-4.6349
4	-1.8734	6.0381	-3.0251
5	-1.8734	6.1961	-2.5925

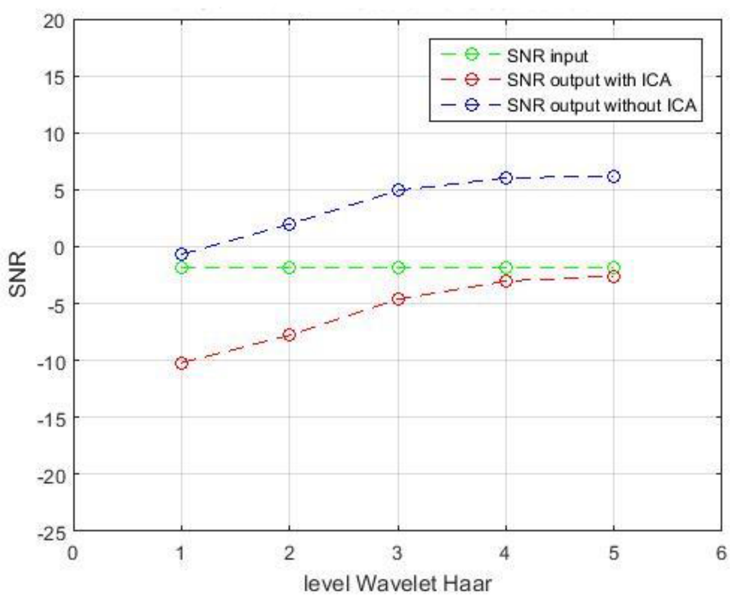

FIGURE 5. Comparison graph of SNR input and SNR output in audio file 1 with a sample time of 30 seconds

It can be seen that the input SNR remains in the same position as the decomposition level rises by the value -1.8734 dB. Then an increase in the value of the SNR output using the wavelet method. The highest SNR value was obtained at Level 5 with a value of 6,1961 dB. Likewise for ICA and wavelet methods an increase occurs along with an increase in the level of decomposition. The highest value of the SNR output from the ICA and wavelet methods is obtained at level 5 with a value of -2.5925 dB.

7.3. Comparison of Haar wavelet levels and SNR outputs from wavelet and hybrid methods (ICA and wavelet)

Following are the results of comparison of the Haar wavelet level with the SNR output of the wavelet method and the hybrid method (ICA and wavelet) that use different Haar wavelet levels. Level 1 The following is a comparison value of level 1 Haar wavelet with the SNR output of the wavelet method and hybrid method with the time for each audio file tested is 30 seconds.

TABLE 2. The Comparison Result of Haar level 1 wavelet and SNR output from wavelet method and hybrid method with 30 seconds sample time

Sampel	SNR Input (Db)	SNR Output	
		Wavelet (dB)	ICA + Wavelet (Db)
File Audio 1	-1.8734	-0.6729	-10.2062
File Audio 2	-1.7398	0.5295	-11.1262
File Audio 3	-1.5966	-0.0021	-14.0297
File Audio 4	-1.2227	1.2831	-11.0248
File Audio 5	-1.3522	-0.1305	-14.8576
File Audio 6	-1.7733	0.0719	-9.9289
File Audio 7	-1.7615	-0.2895	-13.2335
File Audio 8	-1.5039	0.0980	-12.0093
File Audio 9	-1.7175	-0.2642	-12.5054
File Audio 10	-1.0928	0.2250	-15.1923
Rata-rata	-1.5634	0.0848	-12.4114

From the results of table 2 it can be seen that the wavelet method produces a SNR that is greater than the value of the input SNR and SNR output of the hybrid method. The average value of the SNR output of the wavelet method is 0.0848 dB.

7.4. Analysis of Test Results with Comparison of SNR from Wavelet method with Hybrid method (ICA and Wavelet)

After testing it can be seen the pattern of audio from the test results. Figure 4 shows the higher level of wavelet decomposition used in audio, the greater the SNR output value. This is due to the reduction in noise in the source signal, so that the source signal generated by the wavelet method produces an amplitude value per sample which decreases at each level increase in the test. By referring to the SNR equation (2.20), we can know that the greater the total addition of the difference between the average value of the source signal with the source signal that has been reduced noise, the SNR value becomes larger. Then in Figure 4 it can be seen that the hybrid ICA and wavelet methods, which are marked with a red line, produce the same pattern. The pattern of the figure shows that the higher the level of wavelet decomposition used, the greater the SNR output value. For example as in test material 1 with a sample time of 30 seconds, the largest SNR output value from the hybrid method is at level 5 with a value of -2.5925 dB. It can also be seen that the value of the SNR output from the hybrid method is always less than the value of the Input SNR and also the SNR Output from the wavelet method. This is because the ICA method performs the Blind Separation Source (BSS) process, which is to find the value of the mixing matrix of the source signal mixed with noise to be returned back to the source signal. The value of the mixing matrix of the ICA process results is not exactly the same as the value of the original mixing matrix, but rather the estimated value that can separate the sound source signal that has been mixed with the noise source signal, so that the amplitude value of each sample signal source is getting bigger. As a result of the amplitude value that is enlarged, the total of the sum of the source signal values resulting from the hybrid ICA and wavelet methods is also getting bigger. Based on the SNR equation (2.20), if the total sum of the difference between the average value of the source signal and the source signal value that has been done the reduction in noise is greater, then the SNR value becomes larger. The sample time used in each test has no effect on the test results. The success of this noise reduction process can be seen from the comparison graph of the SNR wavelet method output with the SNR output hybrid method. The greater the SNR value, the better noise reduction process. From this it can be seen that the wavelet method is still better than the hybrid method with the best SNR output value obtained at level 5 at each time of the test sample that is 6.7274 dB at 30 seconds sample time, 6.1256 dB at 60 sample times

8. Conclusions

In the process of noise reduction, the ability of the wavelet method is better when compared to the hybrid method because the SNR value of the wavelet method output is always higher than the hybrid ICA and wavelet methods, the SNR value of the output will be greater if the decomposition level is also higher. The best SNR output value each time the test sample is on the SNR output of the wavelet method at level 5 that is with an average value of 6.7274 dB at 30 seconds sample time, 6.1256 dB at 60 seconds sample time and 6.0296 dB at 90 seconds sample time. Output from the process The ICA cannot be determined, this is due to the free component calculated by the ICA method to get the approximate value of the mixing matrix not fully in accordance with the original mixing matrix value, so that the order of outputs changes.

References

1. Wicaksono, G. & Prayudi, Y,Teknik Forensika Audio Untuk Analisa Suara Pada Barang Bukti Digital Semnas Unjani, 2013.
2. K. Prakash dan Hepzibha Rani, Blind Source Separation for Speech Music and Speech Mixtures, ,International Journal of Computer Applications (0975-8887), vol. 110, pp. 40-43, 2015.
3. Grubesa, Sanja & Grubesa, Tomislav ,Speaker Recognition Method Combining FFT, Wavelet Functions and Neural Network, AES Conference: 26th International Conference: Audio Forensics in the Digital Age, Number: 2-2, 2015
4. Daubenchies,K. Agustini, Biometrik Suara dengan Transformasi Wavelet Berbasis Orthogonal, 50 GEMATEK JURNAL TEKNIK KOMPUTER, vol. 9, no. 2, 2007.
5. Kurniawan, Reduksi Derau Pada Sinyal Suara dengan Menggunakan Transformasi Wavelet, 2002.
6. A. Grasp,An Introduction to Wavelets,IEEE Computational Science and Engineering, vol. 2, no. 2, 1995.
7. Mayo Ama Kela Loing dkk,Desain dan Implementasi Sistem Peningkatan Kualitas Perekaman Audio dengan Wavelet Derau Reduction dan Automatic Gain Adjustment, Konferensi Nasional Sistem dan Informatika, Bali, 2008.