

---

## REALTIME PORTABLE MUSIC'S GENRE CLASSIFICATOR WITH THE KOHONEN (SOM) METHODS USING RASPBERRY PI

<sup>1</sup>WIWIET HERULAMBANG , <sup>2</sup>RIFKI FAHRIAL ZAINAL

<sup>1,2</sup>Department of Informatics Engineering, Universitas Bhayangkara Surabaya  
Jl. Ahmad Yani 114, Surabaya, Jawa Timur 60231, Indonesia  
e-mail: <sup>1</sup>herulambang@ubhara.ac.id , <sup>2</sup>rifkifz@ubhara.ac.id

### ABSTRACT

*Music genre is one of the digital music data that is determined to classify music based on all the character equations of each type. The characteristics in question are usually seen from the frequency of music, rhythmic structure, instrumentation structure, and harmony content that the music has. Classification of music genres in realtime (automatic / not manual), giving effect to the classification is no longer relative / subjective, because it is done based on predetermined parameters. In this study Raspberry Pi microcomputer is used, which is quite concisely used as a portable media and is quite powerful for realtime data processing. Raspberry Pi is used as a sound processing unit, music genre identifier, and information on the results of the introduction of the music genre. This system input is in the form of music sound (realtime), while the system output is information (text) about the music genre. Whereas for the process of recognizing the music genre, the Self Organizing Maps (SOM) type Neural Neural Network (JOM) method is also used, or also known as the Kohonen ANN Network. The feature extraction stage uses the Music Genre Recognition by Analysis of Text (MUGRAT) method, with nine features related to the spectral surface of music, and six features related to beat / rhythm of music. Mel Frequency Cepstral Coefficients (MFCCs) feature extraction process was carried out as input from the classification process using the Self Organizing Map (SOM) method. The classification results using the SOM method give an accuracy value of 74.75%. Accuracy of classification results using training data as many as 400 pieces which are divided into 4 musical genres amounting to 74.75%.*

**Keywords:** Music Genre Classifier, Kohonen, self-organizing map, RaspberryPi

### 1. RELATED RESEARCH

Today's daily activities cannot be separated from audio entertainment (music), which is increasingly rich in playback sources with the internet and streaming media. Often listeners want to know the type (genre) of music that he accidentally heard at that time, then the existence of portable equipment that can show the genre of music in realtime is clearly very helpful for the introduction of music.

Music genre is one of the digital music data that is determined to classify music based on all the character equations of each type. The characteristics in question are usually seen from the frequency of music, rhythmic structure, instrumentation structure, and harmony content that the music has. Classification of music genres in realtime (automatic / not manual), giving effect to the classification is no longer relative / subjective, because it is done based on predetermined parameters.

Research on the method of classification of music genres has been done a lot before, but is still in the scope of experimental and based on desktop PC computers so that it is less flexible. The various methods developed do indeed require quite large and high-end computational resources, considering that computational calculations carried out require a high level of processor and memory support (Schmadecke, 2013). One such classification method is artificial neural network, which has been proven to be very effective in conducting classifications with very large amounts of data (Herulambang, 2017). The higher the level of accuracy of the output of a method, the greater the hardware support it needs. This brings the consequence that the system is increasingly inflexible and portable.

On the other hand, the development of microcomputer technology supports the principle of mobility and flexibility, with increasing computational capabilities of microcomputer modules circulating in the public, one of which is Raspberry Pi types. With the support of extension modules that are quite diverse and ready to be embedded in the main Raspberry Pi module, it is possible to build a sophisticated system of microcomputers for the purpose of flexibility and mobility (Norris, 2017).

In this research, a portable and flexible prototype machine for classifying music genres was developed with realtime input recognition capabilities using the Kohonen / SOM JST classification method on Raspberry Pi microcomputers. Mel Frequency Cepstral Coefficients (MFCCs) feature extraction, which is widely used for speech recognition, to group music genres (Logan, 2000). Classification is done using the unsupervised type of neural network method, namely Kohonen neural network. or also called Self Organizing Maps (SOM). In this study, music will be grouped into 4 genres, namely classic, pop, rock, and jazz.

## 2. RESEARCH METHODS

### 2.1 Dataset

The GTZAN Genre Collection music file dataset is used for the training process. This dataset consists of 10 music genres, each consisting of 100 tracks. In this study only 4 genres were used, namely classic, pop, rock and jazz so that a total of 400 tracks were used. Each track has a length of 30 seconds with an audio file format of 22050Hz Mono 16-bit. This dataset was collected from 2000 - 2001 from various sources.

### 2.2 Features Extraction

Mel Frequency Cepstral Coefficients (MFCCs) are a feature that is often used in the speech recognition process because it is capable of displaying sound amplitude spectra in a concise form (Logan, 2000). The process for obtaining the MFCC feature is shown in Figure 1. To produce the MFCC feature, it is first performed to divide the sound signal into several frames using a window at certain intervals. In this study 200 frames of each data were used. Then the Discrete Fourier Transform (DFT) function is performed on each frame and a logarithmic function is performed to obtain the amplitude of the spectrum. The logarithmic function is chosen because the signal volume level has a shape similar to a logarithmic curve.

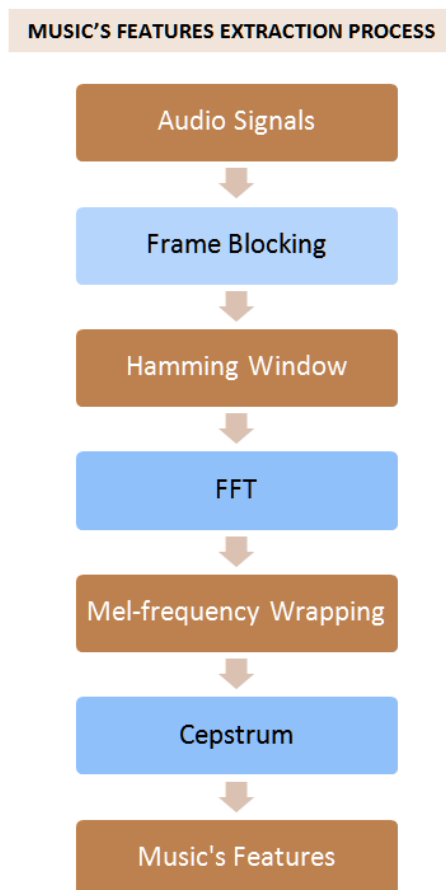


Figure 1. MFCC Features Extraction Process

Next, a selection of signal components is carried out to smooth the spectrum and increase the weight of the important frequency. The signal component will be divided into several bin where the distance between bin is in accordance with the frequency scale 'Mel'. The 'Mel' frequency scale is based on actual frequency mapping with frequencies accepted by humans because it turns out that the human hearing system does not receive frequencies linearly. The mapping function is linear for frequencies below 1 kHz and in the form of logarithms for frequencies above 1 kHz. The Mel function is shown in Figure 2. There is a high correlation between Mel vector components in each frame.

To eliminate the correlation between components while reducing the number of parameters, feature selection was carried out using the Discrete Cosine Transform (DCT) method (Marhav & Lee, 1993). In this study, the number of Mel components maintained for each data is 15 pieces. The output of this feature extraction process is a 200 x 15 matrix for each data where 200 is the number of frames of data, each frame consists of 15 components. Furthermore, the calculation of the average value of each component is obtained so that the features of the average value of the components amounting to 15 pieces for each data. This average value feature will be used as input from the next process.

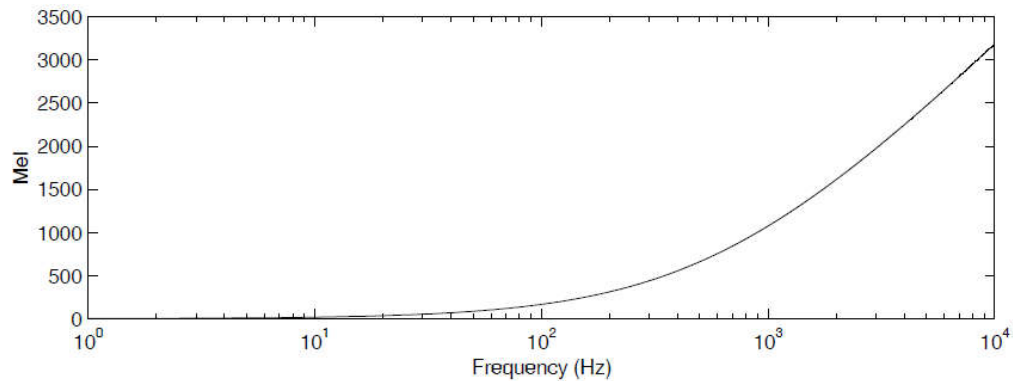


Figure 2. MEL Scale

### 2.3 Classification Process

The classification process is carried out using the Kohonen Neural Network method, namely Self Organizing Map (SOM). This method is able to organize unlabeled data into several clusters. This network is formed from a matrix of neurons that take input in the form of an input signal, which in this case is a feature obtained from the previous process. SOM networks do not have hidden layers and neurons in the output layer are usually in the form of two-dimensional rectangular grids. The vector in the input layer will be mapped nonlinearly to the output layer so that the topology of the data can be maintained as much as possible. This method is very useful for visualizing high dimensional data.

At first each neuron has a weight whose value is random. Furthermore, through the training process, the weights will be updated in value to produce the structure most similar to the distribution of the input data. The calculation of the similarity between input data and the weight of neurons is carried out. The neuron with the most suitable value will be chosen as the winner (winning neuron) and then the winning weight of the neuron and its neighbors will be updated. In SOM, there is no lateral connection between neurons in the output layer. But adjacent neurons will interact with each other using neighboring functions in the training process.

If  $x = [x_1, x_2, \dots, x_n]^T$  is an input vector and  $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$  is a weight vector of neuron  $i$  where  $i = 1, 2, \dots, m$  and  $m$  is som of output neuron. In the SOM training process, the calculation process is first carried out using Euclidean distance of input vector  $x$  for all of weight vectors  $w_i$ . Winning neuron are output neuron which has value of weight vector  $w_q$  and has shortest Euclidean distance to  $x$ . The winning neuron  $q(x)$  index is calculated using Equation (1).

$$q(x) = \arg \min_i \|x - w_i\|_2, i = 1, 2, \dots, m \quad (1)$$

Updating the weight of a  $i$  neuron of  $k+1$  iteration is using Equation (2). The  $i$  neuron which is adjacent to winning neuron  $q$  will be renewed using the equation (3) where  $\mu(k)$  is a learning rate parameter. Whereas the neighboring neurons will not be renewed, so  $n_{qi}(k) = 0$ .

In this research, SOM network was made with the number of neurons 4 to accommodate groups of music genres totaling 4 pieces. The first group represents the genre of classical music, the second group represents the genre of jazz, the third group represents the genre of rock music, and the fourth group represents the genre of pop music.

$$w_i(k + 1) = w_i(k) + n_{qi}(k)[x(k) - w_i(k)] \quad (2)$$

$$n_{qi}(k) = \mu(k) \quad (3)$$

### 3. DISCUSSION AND RESULTS

In this research, the implementation of music genre classification method was carried out using the neural network method using the Matlab software. The classification process is conducted with training data as many as 400 pieces which are divided into 4 music genres, namely classical, jazz, rock, and pop. Each music genre consists of 100 data. SOM training process was carried out with an epoch number of 500 and it was found that the results of classification accuracy were 74.75%.

The SOM network has 15 pieces of neuron input which are the average value of MFCC feature components and have a number of neuron outputs of 4 in accordance with the number of musical genres, as shown in Figure 3. SOM network structure in the process of randomly assigned neuron weighting in Figure 4, while Figure 5 shows the network structure after the training process is complete. In Figure 5 it appears that there are two groups that have close distances so that they can cause classification errors. Based on the results of the analysis shown in the confusion matrix in Table 1, the two groups that are close together are classical and jazz genres. It appears in Table 1 that a lot of data should be grouped in classical classrooms, by incorrect systems grouped into jazz classes.

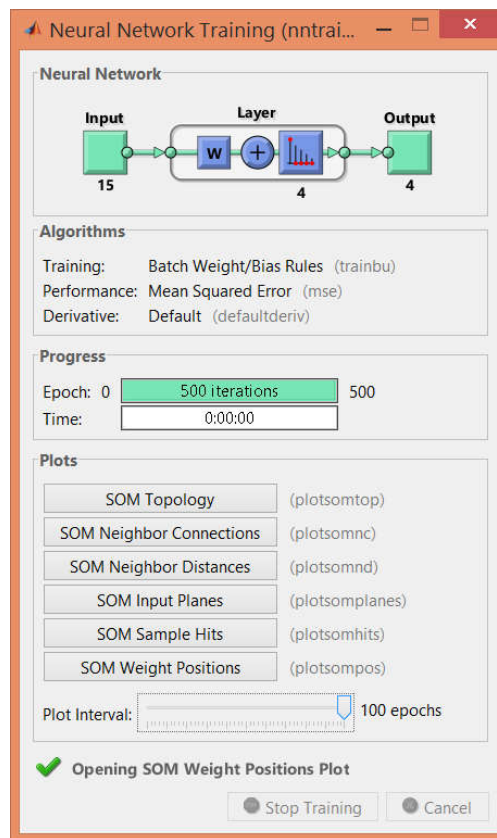


Figure 3. SOM Training Process

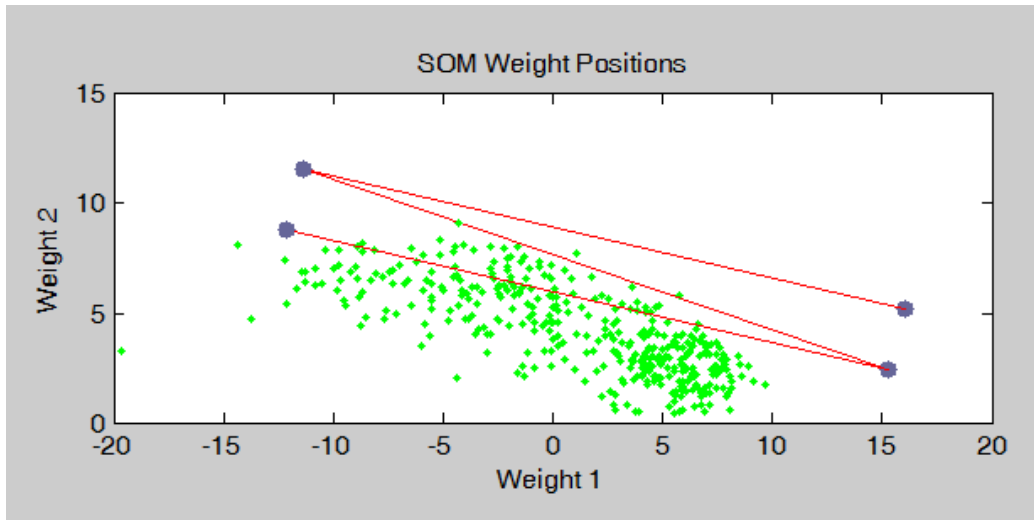


Figure 4. Network Structures : Before Training Process

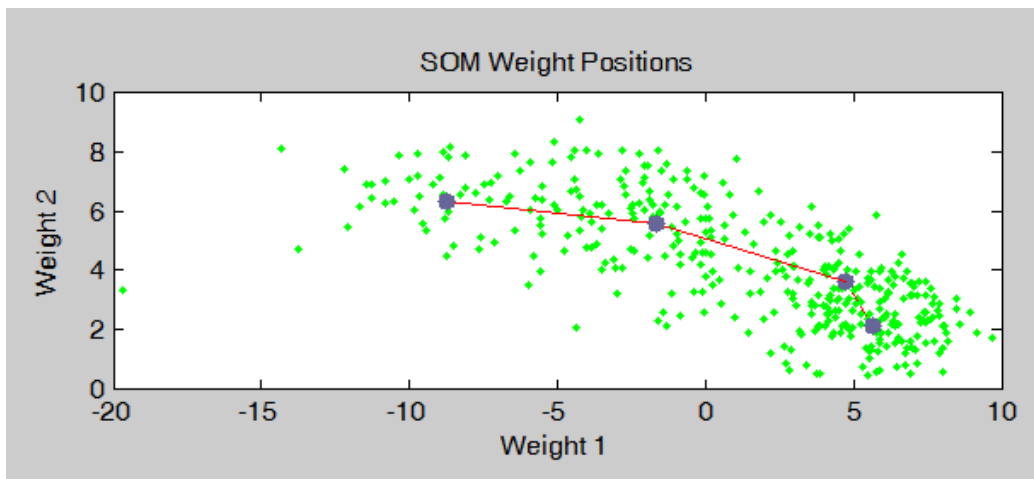


Figure 5. Network Structures : After Training Process



Figure 6. The Prototype

Table 1. Conclusion Matrix

		Classification Result			
		Classic	Jazz	Rock	Pop
Targets	Classic	47	46	6	1
	Jazz	12	64	9	15
	Rock	0	4	92	4
	Pop	0	5	1	94

#### 4. CONCLUSION

In this research used a dataset of GTZAN Genre Collection music files which amounted to 400 pieces of data and divided into 4 musical genres, namely classical, jazz, rock, and pop. Mel Frequency Cepstral Coefficients (MFCCs) feature extraction process was carried out in 15 pieces as input from the classification process using the Self Organizing Map (SOM) method. The SOM network structure consists of 4 neurons representing 4 musical genres and the training process is carried out until it reaches the epoch value of 500. The classification results using the SOM method give an accuracy value of 74.75%.

#### REFERENCES

- [1] Anan, Yoko, Kohei Hatano, Hideo Bannai, Masayuki Takeda, 2011, *Music Genre Classification Using Similarity Functions*, Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, October 24-28 2011, ISBN 978-0-615-54865-4, Miami, Florida, USA
- [2] Arjannikov, Tom, John Z. Zhang, 2014, *An Association-Based Approach To Genre Classification In Music*, Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014, October 27-31 2014, Taipei, Taiwan
- [3] Herulambang, Wiwiet, Retantyo Wardoyo, 2017, *3D Simulation of Plant Growth Modeling Using Neuro-Fuzzy, Lindenmayer System, and Turtle Geometry*, JEECS (Journal Of Electrical Engineering And Computer Sciences) Vol 2 No 2 2017, Surabaya, Indonesia
- [4] Li, Tom LH., Antoni B. Chan, Andy HW. Chun, 2010, *Automatic Musical Pattern Feature Extraction Using Convolutional Neural Network*, Proceedings of The International MultiConference of Engineers and Computer Scientists (IMECS) Vol 1, March 17-19 2010, pp546-550, Hongkong
- [5] Norris, Donald J., 2017, *Beginning Artificial Intelligence with the Raspberry Pi*, ISBN-13 (pbk): 978-1-4842-2742-8, DOI 10.1007/978-1-4842-2743-5, Apress Publishing, Barrington, New Hampshire, USA
- [6] Schmadecke, Ingo, Holger Blume, 2013, *High Performance Hardware Architectures for Automated Music Classification Algorithms from and for Nature and Life*, Studies in Classification, Data Analysis, and Knowledge Organization, DOI 10.1007/978-3-319-00035-0 55, Springer International Publishing, Switzerland
- [7] Schreiber, Hendrik, 2015, *Improving Genre Annotations For The Million Song Dataset*, Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015, October 26-30 2015, ISBN 978-84-606-8853-2, Málaga, Spain
- [8] Silla Jr, Carlos N., Celso A. A. Kaestner, Alessandro L. Koerich, 2010, *Improving automatic music genre classification with hybrid contentbased feature vectors*, Proceedings of the 2010 ACM Symposium on Applied Computing (SAC'10), pages 1702–1707, New York, USA