Check Analysis and identification Gamelan Bonang sound spectrum.pdf

by Arta Moro Sundjaja

Submission date: 24-Jan-2019 01:22PM (UTC+0700)
Submission ID: 624024364
File name: Analysis_and_identification_Gamelan_Bonang_sound_spectrum.pdf (323.14K)
Word count: 2281
Character count: 11835

Second International Conference on Computational Intelligence, Modelling and Simulation

Analysis and Identification Gamelan Bonang Sound Spectrum

Muhammad Nur Latief
Graduate Program in Informatics Engineering
Bina Nusantara University
Jakarta, Indonesia
dimas.latief@yahoo.com

Ford Lumban Gaol
Faculty of Computer Science,
Bina Nusantara University
Jakarta, Indonesia
fgaol@binus.edu

Bambang Heru Iswanto
Physics Department
Universitas Negeri Jakarta
Jakarta, Indonesia
bhionet@gmail.com

***Abstract*— In this research will show a method for sound-recognition with artificial neural network backpropagation concept. The artificial neural network use sigmoid activation function to all layer. Steps to the extraction, first, divide into ten component, second do the Fast Fourier Transform, third continue with Power Spectral Density, fourth count the average. The end result show that the pattern will recognize by the Artificial Neural Network. Extraction performed, as well as Fourier transform, and also calculate the power density, and then process them with artificial neural networks, allows the system to do the identification of voice data Gamelan Bonang using In this research will show a method for sound-recognition with artificial neural network backpropagation concept.**

***Keywords*— Identification, Component, Fourier Transform (FT), Power Spectral Density (PSD), Backpropagation**

I. INTRODUCTION

Gamelan Bonang is traditional musical instrument from Indonesia. It is one of Indonesian heritage. Gamelan has so many variations, and has different sound characteristics [6].

To analyze and identification data of the sound spectrum, the data need to convert from analog signal to digital signal. Studying the spectrum and look at the object relation. After analyzing, starting to find the identity of the object to do the identification [2].

Fourier transform is an operation that transforms one complex-valued function of a real variable into another. In such application as signal processing, the domain of the original function is typically time and accordingly called the *time domain*. The domain of the new function is typically called the frequency domain, and the new function itself is called the *frequency domain representation* of the original function [7].

Power Spectral Density, The unit of PSD is energy per frequency (width) and you can obtain energy within a specific frequency range by integrating PSD within that frequency range. Computation of PSD is done directly by the method called Fast Fourier Transform (FFT) or computing autocorrelation function and then transforming it [1].

FFT is a measurement technique that gives narrow-band filtering using post-processing of a digital time record. Modern digital processors can repeat the analysis fast enough to provide real-time measurement.

Back-Propagation (BP) network learns by example, that is, we must provide a learning set that consists of some input examples and the known-correct output for each case. So, we use these input-output examples to show the network what type of behavior is expected, and the BP algorithm allows the network to adapt [10].

II. LITERATURE REVIEW

A. *Sound Spectrum*

Sound spectrum is a representation of a sound – usually a short sample of a sound – in terms of the amount of vibration at each individual frequency. It is usually presented as a graph of either power or pressure as a function of frequency. The power or pressure is usually measured in decibels and the frequency is measured in vibrations per second (or hertz, abbreviation Hz) or thousands of vibrations per second (kilohertz, abbreviation kHz). Most people can hear frequencies in the range between 100Hz-15000Hz. Some people can hear very high frequencies above 19000Hz, but scientists *always* assume that the human ear is able to discern frequencies between 20Hz-20000Hz [2].

Sample rate, the sample rate of a piece of digital audio is defined as 'the number of samples recorded per second'. Sample rates are measured in Hz, or kHz (kiloHertz, a thousand samples per second) [2].

Nyquist sampling, In essence, the theorem shows that a band limited analog signal that has been sampled can be perfectly reconstructed from an infinite sequence of samples if the sampling rate exceeds $2B$ samples per second, where B is the highest frequency in the original signal. If a signal contains

978-0-7695-4262-1/10 $26.00 © 2010 IEEE
DOI 10.1109/CIMSiM.2010.78

a component at exactly B hertz, then samples spaced at exactly seconds do not completely determine the signal [7].

Formula of sampling rate:

$$S = F_s * T$$

Where:

S = vector length

F_s = Sampling rate (Hz)

T = length of time

Tansig is a neural transfer function. Transfer functions calculate a layer's output from its net input. Here is the formula [7],

$$f(x) = \frac{1}{1+e^{-2x}} - 1$$

The formula of sigmoid activation function is,

$$f(x) = \frac{1}{1+e^{-x}}$$

Formula to calculate each unit inside the hidden layer,

$$Net_{(t)} = \sum_{i=1}^{n} w_i p_i + b_i + OutH_{(t-1)}$$

$$OutH_{(t)} = f(Net_{(t)})$$

Where,

$OutH_{(t)}$ = Output T time

W_t = Weight

P_i = Input Vector

B_i = Bias from neuron

$Net_{(t)}$ = Sum of the input

F(x) = Activation function

Formula to calculate the output layer,

$$Net_{(t)} = \sum_{i=1}^{n} w_i OutH_i + b_i$$

$$Out_{(t)} = f(Net_{(t)})$$

Where,

$OutH_{(t)}$ = Output T time

W_t = Weight

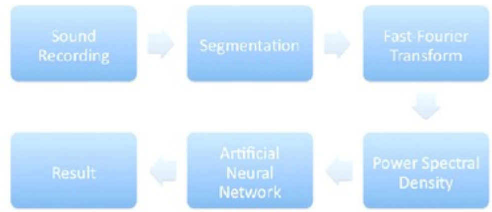
P_i = Input Vector

B_i = Bias from neuron

$Net_{(t)}$ = Sum of the input

F(x) = Activation function

III. METODOLOGY



A. *Sound Recording*

First, sound will be recorded in a computer. By using, normal microphone and Audacity software. Placing computer, microphone, and the noise is important to consider. Less noise, can result in better recorded sound.

B. *Segmentation*

In the segmentation, sound will be taken exactly 1 second. The sound taken, can use audacity software. From the starting sound, until the first second.

C. *Fast-Fourier Transform (FFT)*

FFT use a mathematical function to do calculation of the number. It can transform from time-domain into frequency-domain.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \qquad k = 0, \dots, N-1.$$

Evaluating this definition directly requires O (N^2) operations: there are N outputs X_k, and each output requires a sum of N terms. An FFT is any method to compute the same results in O (N log N) operations. More precisely, all known FFT algorithms require Θ (N log N) operations (technically, O only denotes an upper bound) [7].

D. *Power Spectral Density*

Power Spectral Density (PSD) is a positive real function of a frequency or a deterministic function of time, which has dimensions of power per Hz, or energy per Hz. It is often called simply the spectrum of the signal.

The PSD is the Fourier transform of the autocorrelation function, $R(\tau)$, of the signal, which describes how the power of a signal or time series is distributed with frequency [1].

This results in the formula,

$$S(f) = \int_{-\infty}^{\infty} R(\tau)\, e^{-2\pi i f \tau}\, d\tau = \mathcal{F}(R(\tau)).$$

E. *Artificial Neural Network*

A neural network (NN) is a massive processing system that consists of many processing entities connected through links that represent the relationship between them. MLP Neural Network architecture [5].

Here are, the value for the boundaries.

Table 3.1. Boundaries

Boundaries	Values
Epochs	1000
Time	Infinite
Goal	0
Max_fail	6

Epochs will be limited up to 1000 times. Epochs here can be interpreted as iteration. Thus limit the training to be conducted by the neural network has a maximum iteration times as many as in 1000, when more than that, the training will be terminated. If seen in the time limit, i.e. infinite, so the time required in training the network is not restricted. The goal constraints, achieved the final value. The value specified is 0, but, in some experiments rarely up to 0, a very small value close to 0 even suffice. Max fail to show the limits, if the errors made during training a network of more than six, then the training will be discontinued.

Table 3.2. Network

Description	Used
Network	*Feed-forward Backpropagation*
Training Function	*TrainLM*
Adaption Learning Function	*LearnGD*
Number of Layers	2
Transfer Function	*Tansig*

Type of network used is Feed-Forward Backpropagation, training function Lavernberg-Marquadt using two layers, and transfer function using TanSigmoid. The selected network has two layers and using the activation function commonly used to model this network has no over fitting, whereby if there is over fitting the data will be tested will issue an inaccurate result. However, over fitting will provide an excellent accuracy results if the data being examined is a trained data.

IV. 'RESULTS AND DISCUSSION

First, the sound will be divided into 10 segments. Each segment calculated with FFT, and continues with Power-Spectral Density. There will be three different sound spectrum, which is, Gamelan Bonang, Gamelan Kempyang, and Gamelan Gong.

Here are, the 10-segment value of the Gamelan Bonang.

Table 4.1. Table Value Gamelan Bonang

Segment	1	2	3
I	0.0639	0.0448	0.0314
II	0.0497	0.0271	0.0194
III	0.0315	0.0155	0.0111
IV	0.0280	0.0123	0.0095
V	0.0177	0.0075	0.0058
VI	0.0083	0.0037	0.0028
VII	0.0046	0.0022	0.0015
VIII	0.0023	0.0011	0.0008
IX	0.0010	0.0006	0.0004
X	0.0007	0.0004	0.0003
Target	**1**	**1**	**1**

If seen in table 4.1. Every stroke has a ten-value, the first showing the first gamelan was split into ten sections, these values can be seen from the first Roman numeral to roman tenth. Initial value greater than last value indicates that the frequency is greater at the beginning in accordance with the image shown in the frequency domain. Ten of this value indicates that there will be ten input values for neural network. Target located at the bottom of the table refers to the value you want is found. So, from the tenth that value, the middle screen (hidden layer) neural network will calculate the weight value that will remove the output value of one, from ten input values provided.

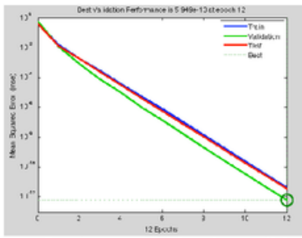
Next table is value that not Gamelan Bonang.

Table 4.2. Table Value Not Gamelan Bonang (Gong and Kempyang)

Segment	1	2	3
I	0.0267	0.0290	0.0133
II	0.0398	0.0337	0.0048
III	0.0274	0.0243	0.0007
IV	0.0229	0.0216	0.0007
V	0.0182	0.0175	0.0003
VI	0.0144	0.0139	0.0002
VII	0.0125	0.0120	0.0002
VIII	0.0093	0.0086	0.0000
IX	0.0084	0.0081	0.0000
X	0.0060	0.0053	0.0000
Target	**0**	**0**	**0**

Gamelan is not bonang, i.e. consisting of Gamelan Gong (for first and second data) and Gamelan Kempyang (for the third data) input into neural network training in order, the network knows what kind of data that is not true. Target given for these data is 0. That is, data that have a range of values as the data is otherwise 0, or to the definition given here is the wrong data.

Once these values are obtained, the next can be incorporated into neural network. Backpropagation network is used. Experiments performed adequately, by using backpropagation with the number of neurons; 30, 40, 60, 100, 150, 200, 250, and 300 neurons. Here is a picture of the network training performance for neurons 30.



Picture 4.1. Performance Number of Neurons 30

From the picture above, the best performance is obtained at the 12th iteration. Indicated by the line that point, and there is a green circle. That is, at the 12th iteration, the obtained data that has the least error (minimum).

Table 4.3. Average Overall Value of Performance Test Data

Neuron	Gamelan Bonang	Gamelan Bukan Bonang
30	1	0.1461
40	1	2.4549e-05
50	1	0.0042
60	1	0.0694
100	1	3.2063e-06
150	1	0.5092
200	1	0.0023
250	1	0.0011
300	1	0.0181

Table 4.3. Shows the overall average of test data. If seen from the data, which can be selected the best network is a network that has the number of neurons as many as 100 and 40.

V. CONCLUSION

The conclusion shows that, using backpropagation artificial neural network with two layers, then use the training function trainLM, adaption LearnGD learning function, transfer function Tansig, with the maximum epoch training data requirements in 1000, goals 0, maximum 6 fail, without limitation of time training, and has a number of neurons as many as 100 and 40 have a good identification accuracy.

Extraction performed, as well as Fourier transform, and also calculate the power density, and then process them with artificial neural networks, allows the system to do the identification of voice data Gamelan Bonang.

REFERENCES

[1] Antonisfia Y., Wiryadinata R., *Ekstrasi Ciri Pada Isyarat Suara Jantung Menggunakan Power Spectral Density Berbasis Metode Welch.* Media Informatika, Vol. 6, No. 1, Juni 2008, 71-84. ISSN: 0854-4743.

[2] Boomkamp, J., 2004. *The Theory of Digital Sound,* [Online] Available: http://www.hammersound.net/audiobasics/audiobasics.html [2010, Januari 20]

[3] Biyanto, T. R., *Over Fitting pada Neural Network,* [Online] Available: http://www.waliallah.org/index.php?option=com_content&view=article&id=107%3Aoverfitting-pada-neural-network&catid=38%3Ateknologi&Itemid=60&lang=in [2010, Juni 18]

[4] Danikos, 2010. *Neural Networks: A Requirement for Intelligent Systems,* [Online] Available: http://www.learnartificialneuralnetworks.com/ [2010, Juni 18]

[5] Emerald, 2010. *Multilayer Preceptron,* [Online] Available: http://emerald2010.cjb.net/fig/1340170102011.png [2010, Juni 18]

[6] Kuo, H., 2010. *Bonang.* [Online] Available: http://www.seasite.niu.edu/Indonesian/Budaya_Bangsa/Gamelan/Javanese_Gamelan/gongs/bonang.htm [2010, Januari 20]

[7] Mathwork Inc., 1991. *Neural Network Toolbox for use with Matlab.* USA: The Mathwork Inc. Natick.

[8] Martin D.K., *Sound-Source Recognition: A Theory and Computational Model.* Massachusetts Institute of Technology, 1999.

[9] Mohamad M., Farokhi F., *An Advanced Method for Speech Recognition,* World Academy of Science, Engineering and Technology 49 2009.

[10] Rich E., Kevin K., 1991. *Artificial Intelligent, Second Edition.* USA: McGraw Hill, Inc.

[11] Setiawan F.B., et al. *Model Sinusoida Secara Segmental Untuk Pengkodean Sinyal Suara,* Makara Teknologi, Vol. 10, No.2, November 2006: 61-66.

[12] Siang, J. J., 2009. *Jaringan Syaraf Tiruan & Pemrogramannya Menggunakan Matlab.* Jakarta: Penerbit Andi.

[13] The Sonic Spot., 2005. *The Waves Files,* [Online] Available: http://www.sonicspot.com/guide/wavefiles.hmtl [2010, Januari 20]

[14] Vesaintwo10, 2009. *Gambaran Secara Matematik,* [Online]

[15] Wolfe J., 2010. *What is Sound Spectrum,* [Online] Available: http://www.phys.unsw.edu.au/jw/sound.spectrum.html [2010, Januari 20]

Check Analysis and identification Gamelan Bonang sound spectrum.pdf

ORIGINALITY REPORT

97%	38%	97%	26%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Latief, Muhammad Nur, Ford Lumban Gaol, and Bambang Heru Iswanto. "Analysis and Identification Gamelan Bonang Sound Spectrum", 2010 Second International Conference on Computational Intelligence Modelling and Simulation, 2010. Publication	97%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		