Pitch Shifting Based Phase Vocoder for Synthesizing Javanese Gamelan Gong Ageng

Muljono*, Y. Tyas Catur P*., Amiq Fahmi*, Khafiizh Hastuti*

Department of Computer Science
Dian Nuswantoro University,Semarang, Jawa Tengah, Indonesia
{muljono,tyas.catur.p,amiq.fahmi,afis}@dsn.dinus.ac.id

***Abstract* — This paper describes the analysis and synthesis of Javanese Gamelan Gong Ageng using the pitch shifting based Phase Vocoder. Javanese Gamelan is a traditional Indonesian musical instrument played with a repetitive pattern. Gong is one of a group of instruments in the Javanese Gamelan. Gong Ageng is a large gong used in Javanese Gamelan musical to mark the end of each cycle gong. In order to synthesize the tone Gong Ageng with a fundamental frequency, it is necessary to have a tone Gong Ageng selected as a reference signal. We take voice samples for the experiments by recordings the original Gong Ageng. The synthesis stages will be held until it reaches synthetic tones which are similar to the tone of Gong Ageng. To measure their similarities to the original tone, it is necessary to test the error in order to determine the error rate. The error rate is measured by using MSE (Mean Square Error) in order to compare the original signal with the synthetic signal from the tone of Gong Ageng in the frequency domain. The results obtained is MSE = 0.0172. Thus, the accuracies reach 99.98%.**

Key words - frequency, pitch, pitch shifting, phase vocoder, fast fourier transform.

I. INTRODUCTION

Gamelan is one of Indonesian traditional musical instruments that is widely grown in the area of Java and Bali. Gamelan consists of a set of musical instruments made of metal, wood, or bamboo. Javanese Gamelan is grouped into four kinds of musical instruments. The first type is chordophones gamelan which is a musical instrument that produces its sound by vibrating strings example, siter, rebab and celempung, the second type is idiophones which is a musical instrument that produces sound by vibrating themselves by being hit as saron, kenong, gambang and gong, the third type is an aerophones musical instrument which produces sound by vibrating the air column or tube like flutes and the fourth type is membranophones which produces sound by vibrating membrane, for example kendang [1].

Gamelan is played by hitting or swiping at the source of sound on the device to produce a signal which has a variation in terms of the fundamental frequency, harmonics frequency, and envelope signals. A gamelan has its own characteristics and it has different standard in each tone. This condition appears because the construction of gamelan is made by handmade and has different playing styles. Each instrument of gamelan is designed and is created manually so that when it becomes a complete entity it can be used together. In other words, each instrument is not recommended to be replaced from another set of gamelan. This is in contrast with western musical instruments that have the same standard for each tone, such as piano or guitar instrument [2].

One of Javanese Gamelan instrument is gong. Gong has the largest size and lowest tone among other gamelan instruments. It is called by Gong Ageng (Ageng in Javanese language means large) with about 34 inches (86 cm) diameter. Gong Ageng has an important role in Javanese Gendhing (Javanese Song) that is one of the instruments which describes the formal structure of a Javanese Gendhing. The gong marks the end of a long musical unit, to give a feeling of balance after the long melody of Javanese Gendhing. This research takes the voice samples for the experiments by recording the original Gong Ageng which is a part of the Gamelan Pusaka Kyai Talogo Muncar from Keraton Pura Paku Alam Yogyakarta. In order to synthesize the tone with a fundamental frequency, it is necessary to have a tone Gong Ageng selected as a reference signal.

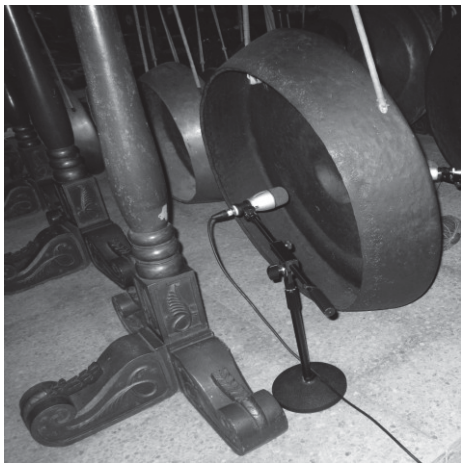

Figure 1. Gong Ageng

This research will perform signal analysis for Gong Ageng in order to determine the constituent components of Gong Ageng signal. Furthermore, the synthesis of the Gong

Ageng tone will be presented by using Pitch Shifting based Phase Vocoder method.

II. LITERATURE REVIEW

Some basic theories needed to perform the analysis and synthesis of signals from the tone gamelan. It includes the theoretical basis for evaluating the results of the synthesis of gamelan tones.

A. *Frequency Scale*

There are two units which are often used to measure the distances between musical notation, namely hertz and cent music [3]. In music, there are some of the terms such as octave, fifth, etc. An octave or perfect octave is the interval between one musical pitch and another with half or double its frequency. In diatonic music like piano or guitar, one octave has 12 semitones or 1200 cent music. The distance between each notation in cent music is 1200/12 cent or 100 cent. Displacement distance cent in diatonic music can be calculated by equation (1) and equation (2).

$$C = 1200 \times \log_2\left(\frac{f_j}{f_i}\right) \tag{1}$$

$$f_j = 2^{\frac{m}{12}} f_i \quad \text{or} \quad f_j = 2^{\frac{n}{1200}} f_i \tag{2}$$

where, C is the distance cent, fi and fj is the fundamental frequency with index i and j, fi is the initial fundamental frequency and fj is the target fundamental frequency, m and n is scaling factor.

Gamelan is a pentatonic music. It has five levels of notation. One octave in gamelan has 5 semitones, so that the distance between each notation in cent music is 1200/5 or 240 cents. Displacement distance in the gamelan music can be calculated by equation (3).

$$f_j = 2^{\frac{m}{5}} f_i \quad \text{or} \quad f_j = 2^{\frac{n}{1200}} f_i \tag{3}$$

B. *Pitch Shifting Algorithm*

Pitch shifting can be understood easily, such as a tone is played twice as fast, then all frequencies are doubled and the fundamental frequency will be shifted up one octave, but it causes the signal to two times shorter. If the length of a tone signal is doubled without affecting the fundamental frequency and then played two times faster then all the frequency will be doubled, will shift the fundamental frequency and the duration of the signal will be adjusted to the length of the initial signal [2][3].

To perform a pitch shifting of tone the first thing needs to be done is to change the duration of the signal without changing the fundamental frequency. The scale factor is defined as the factor used to stretch or compress to adjust the frequency spectrum so that the fundamental frequency is shifted. After this is done, it will be resampled to return to the initial duration with a shift in the fundamental frequency.

For example, if you want to shift the fundamental frequency of 1 step or 240 Cent, the scale factor needed is $\mathbf{2^{(1/5)}}$ or $\mathbf{2^{(240/1200)}}$. This means that first of all we need to stretch the signal without changing the fundamental frequency so that the duration of the current is multiplied by $\mathbf{2^{(1/5)}}$ or $\mathbf{2^{(240/1200)}}$. After this is done, then the signal is $\mathbf{2^{(1/5)}}$ or $\mathbf{2^{(240/1200)}}$ faster.

On the other hand, if it is desired to shift the fundamental frequency by decreasing 1 step or 240 Cent, the scale factor needed is $\mathbf{2^{(-1/5)}}$ or $\mathbf{2^{(-240/1200)}}$. This means that it is necessary to compress the signal without changing the fundamental frequency so that the duration of the current is multiplied by $\mathbf{2^{(-1/5)}}$ or $\mathbf{2^{(-240/1200)}}$. After this is done, then the signal is slower $\mathbf{2^{(-1/5)}}$ or $\mathbf{2^{(-240/1200)}}$ times from the initial velocity.

C. *Windowing*

Windowing is one way to analyze a lengthy signal by taking a considerable portion representation. It is better known as Windowing process [4]. There are several examples of windowing process: *window* Hamming, *window* Hanning dan *window* Rectangular.

Window Hanning :

$$w[n] = \frac{1}{2}\left(1 - \cos\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1 \tag{4}$$

D. *Fast Fourier Transform (FFT)*

Fast Fourier Transform changes the waveform from the time domain into the frequency domain[4].

$$X(k) = \sum_{n=0}^{N/2-1} x(n)W^{kn} + \sum_{N/2}^{N-1} x(n)W^{kn} \tag{5}$$

where $W = e^{-j(2\pi/N)}$

E. *Invers Fast Fourier Transform (IFFT)*

Invers Fast Fourier Transform changes the waveform from the time domain into the frequency domain [4].

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{jk2\pi n/N}, \; n = 0, 1, \ldots, N-1 \tag{6}$$

F. *Evaluation*

The evaluation tests on the degree of similarity of the synthetic signal with the original signal from Gong Ageng is performed by using objective measurements. To measure their similarities to the original tone, it is necessary to test the error in order to determine the error rate. The error rate is measured by using MSE (Mean Square Error) in order to compare the original signal with the synthetic signal from the tone of Gong Ageng. MSE is the mean square error between the estimated data values with the actual data [5].

The formula of MSE can be presented as follows:

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}\left|X(n) - X'(n)\right| \quad (7)$$

Where N is the number of data, X (n) is the value of the elements of the original signal from the recording to the data to-n, while X '(n) is the value of the elements of the newly synthesized signals for the data to-n.

III. ANALYSIS OF SPECTRUM CHARACTERISTICS GONG AGENG

The authors performed an analysis of the Gong Ageng which is a part of the Heritage gamelan Kyai Talogo Muncar. This gamelan is a Javanese gamelan of Keraton Pura Paku Alam Yogyakarta is still often used in performances at Keraton Paku Alam Yogyakarta. The authors recorded Gong Ageng tone with a sampling rate of 48 KHz, mono-channel, 16-bit resolution with a microphone in the rear position Gong Ageng. The fundamental frequency of the Gong Ageng is 48 Hertz.

The author used Fast Fourier Transform (FFT) to perform an analysis in order to determine the frequency spectrum of the signal Gong Ageng. The parameter setting in the n-point DFT FFT is n = 48000. FFT (X) is equivalent to FFT (X, n) where X is the signal of Gong Gamelan X and n is the size of the first dimension non-singleton. If the length of X is less than n, then the length of the data sequence X is filled with zeros in sequence so that the length of X equal to a length of n. If the length of X is greater than n, then the length of the data sequence X will be cut to adjust the size of the n.

Figure 2. below shows the waveform Gong Ageng in Time Domain. The x-axis shows the duration of time from the start of Gong Ageng struck, so that the audible sound until the sound of the voice is lost. The duration is 14 seconds. While the y axis shows the amplitude of the Gong Ageng signal. The amplitude is the furthest distance / deviation from the equilibrium point which affects the sound when it sounded strong or weak. The waveform of the voice signal from a musical instrument is initiated at the Attack, a time in which the signal arises when the instrument is sounded at first. Decay, a time in which after it reaches the peak signal and then fell back before the Sustain. Sustain, a signal when the instrument resonates and Release, is a state of the Sustain to return to the position of the tool does not sound (to a state of zero).

Figure 3 shows a waveform signal of Gong Ageng in the frequency domain and it shows the fundamental frequency which includes harmonics and non-harmonics frequency of the Gong Ageng signal. The fundamental frequency of signal Gong Ageng is 48 Hertz. The x-axis shows the frequency of the Gong Ageng signal while the y-axis shows the magnitude of Gong Ageng signal.

Figure 4 shows a waveform of Gong Ageng signal in the time-frequency domain. It is presented in the form of 3-dimensional. The x-axis shows the time duration of the signal, the y-axis shows the frequency of the signal and the z-axis shows the magnitude of the signal.

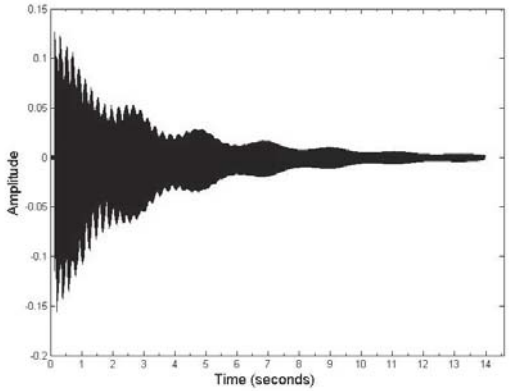

Figure 2. Signal Gong Ageng in Time Domain

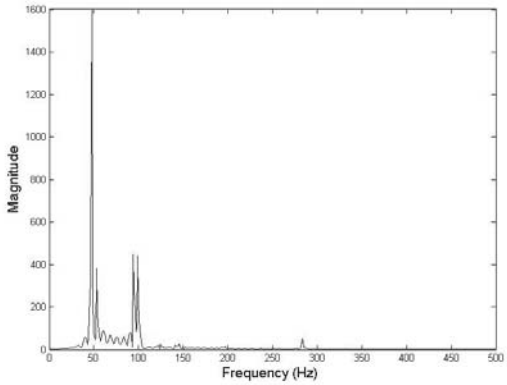

Figure 3. Signal Gong Ageng in Frequency Domain

.

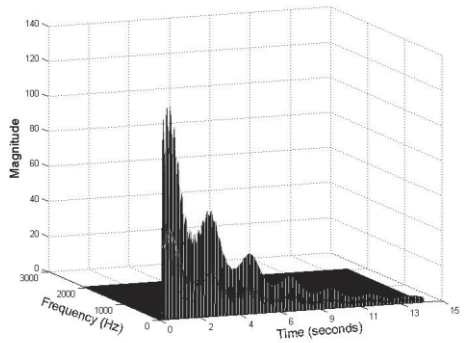

Figure 4. Signal Gong Ageng in Frequency-Time Domain

There are some visible signal envelopes which refer to the fundamental frequency, harmonics frequency and non-harmonics frequency. We use Short Time Fourier Transform (STFT) to analyze the Gong Ageng signal in this domain.

Figure 5 shows the envelope of the Gong Ageng signal. Hilbert transformation is used to detect the envelope of the signal. The x-axis shows the time duration and the y-axis shows the amplitude of the signal Gong Ageng.

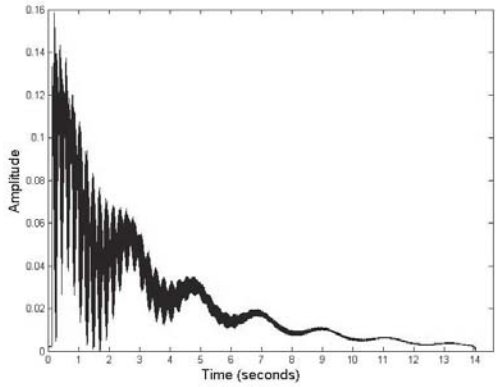

Figure 5. Envelope Signal Gong Ageng

From Figure (2, 3, 4 and 5), we can see the spectral characteristics of Gong Ageng signals including the waveform signal in the time domain, frequency domain, time-frequency domain, the fundamental frequency, harmonics frequency, non-harmonic frequency and the envelope signal.

IV. SYNTHESIS MODEL GONG AGENG

There are several steps performed to synthesise Gong Ageng tone by using pitch shifting method based phase vocoder :

Step 1: The selection of Gong Ageng signal as the base material to be used in this research.

Step 2: Analysis stage. In this stage signal, the analysis is performed in the time domain. This analysis uses Short Time Fourier Transform (STFT). The first process is the frame blocking and windowing of the signal. In the next process, each signal performed FFT to find the magnitude and phase of the signal. The result of the analysis step is a time-frequency domain signal.

Step 3: Signal processing in the Time-Frequency Domain. In this step the process is done to shift the fundamental frequency.

Step 4: Synthesis stage. The stages aim to convert the signal from the frequency domain to the time domain by using Invers Fast Fourier Transform (IFFT). The results of this stage are signal in the time domain.

Step 5: Resampling stage is the stage to unify the output frame that has been formed and stored and return the signal to the initial duration.

After we performed the pitch shifting based phase vocoder above, starting from determining Gong Ageng signal as a reference signal until the last stage, finally we get a synthetic signal Gong Ageng. Figure 7 and Figure 8 below, show the similarity between Gong Ageng synthetic signal with the original signal. The similarities can be seen from the fundamental and harmonic frequencies, as well as the envelope at the each signal.

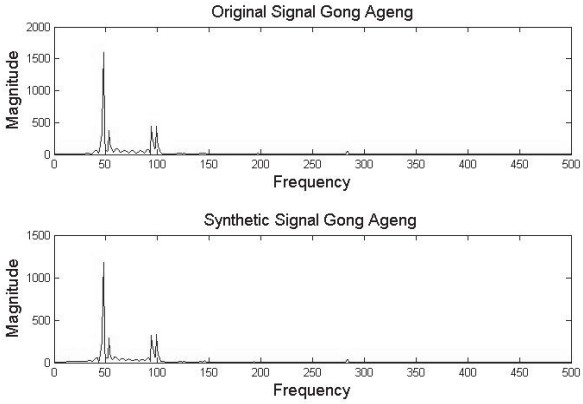

Figure 7. Comparison of the frequency of the original signal Gong Ageng with the synthetic signal Gong Ageng

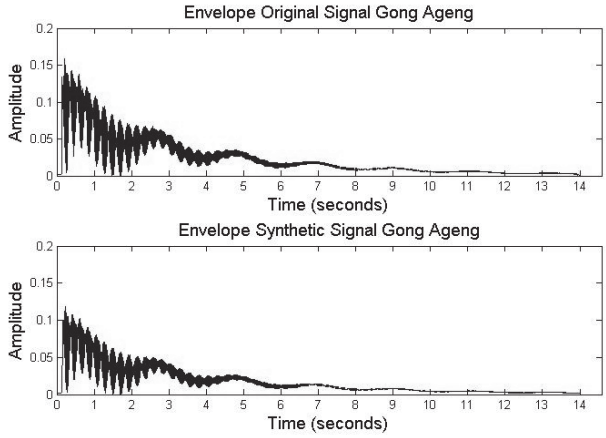

Figure 8. Comparison of the envelope original signal Gong Ageng with the envelope synthetic signal Gong Ageng

The evaluation tests on the degree of similarity of the synthetic signal with the original signal from Gong Ageng is performed by using objective measurements. Objective measurements are done automatically by calculating the voice data signal using a computer. This measurement is typically faster and cheaper to compute because it does not involve human experiments. In contrast to the subjective measurements which are based on the opinions expressed by human in listening and evaluation of the synthesis results, the subjective measurements are based on human opinions and analysis. The advantage is that this measurement is directly related to human perception, which is usually the standard for judging the quality of synthesized speech. The disadvantage is that they are time consuming, expensive, and difficult to interpret the results of their opinions. We use objective measurements to test the level of signal similarity by using the Mean Square Error (MSE).

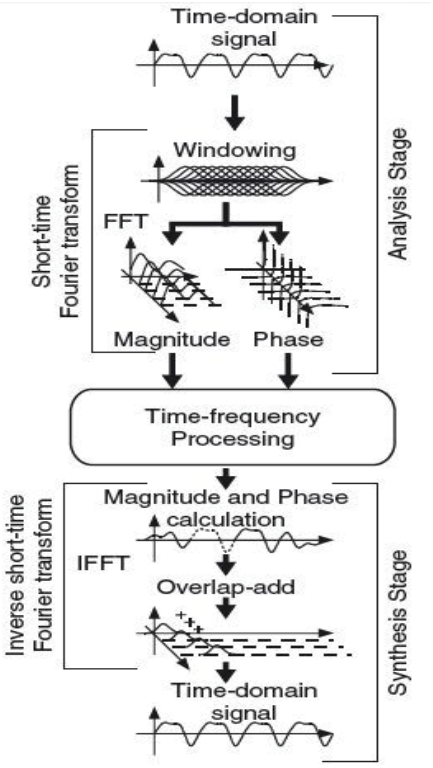

Figure 6. Model Phase Vocoder [6]

MSE calculation is performed by comparing the original signal with the synthetic signal from the tone of Gong Ageng in the frequency domain. The results obtained is MSE = 0.0172. Thus, the accuracies reach 99.98%..

V. CONCLUSIONS

We have done the analysis and synthesis of signals from Javanese Gamelan Gong Kyai Ageng Taloga Muncar. Gamelan is a Gamelan Pusaka Paku Alam Yogyakarta. The result of the analysis shows that the fundamental frequency of the tone Gong Ageng is 48 Hertz. It also obtains frequency harmonics, waveform in the time domain, frequency domain, time-frequency domain and the envelope of the Gong Ageng signal.

By using the pitch shifting based phase vocoder method that created the synthetic signals, the synthetic signals obtained from Gong Ageng are almost similar to the original signal from the gong. To measure their similarities to the original tone, it is necessary to test the error in order to determine the error rate. The error rate is measured by using MSE (Mean Square Error) in order to compare the original signal with the synthetic signal from the tone of Gong Ageng in the frequency domain. The results obtained is MSE = 0.0172. Thus, the accuracies reach 99.98%.

ACKNOWLEDGMENT

We would like to thank to the Directorate of Higher Education, Ministry of Education and Culture of the Republic of Indonesia, which has provided the financial support through the Competitive Research Grants for Fiscal Year 2014. We also deliever our gratitude to Dian Nuswantoro University for all the help and support that are poured empirically through the contract agreements No. 013/A.35-02/UDN.09/V/2014.

REFERENCES

[1] Sumarsam," *Cultural Interaction and Musical Development in Central Java*", 1992-1995, The University of Chicago Press, ISBN 0-226-78011-2

[2] Muljono, Suprapto, Y., Hariadi, M., 2012, "Sintesis Nada Saron Menggunakan *Pitch Shifting Phase Vocoder* untuk Standarisasi Suara Saron", ISBN : 9786029876802, Pebruari 2012, Proceeding KNSI 2012 Stikom Bali, Bali

[3] Laroche, Jean. , Dolson, Mark., *"New Phase-Vocoder Techniques For Pitch-Shifting, Harmonizing and Other Exotic Effect",* Oct. 17-20, 1999**,** Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York.

[4] Richards G. Lyons, 2004, "*Understanding Digital Signal Processing*", Prentice-Hall

[5] Duhamel, P. & Vetterli M. "Fast Fourier Transforms: A Tutorial Review and a State of the Art" *Digital Signal Processing Handbook,* Ed. Vijay K. Madisetti and Douglas B. Williams, Boca Raton: CRC Press LLC, 1999

[6] Zolzer, Udo, 2011, "*DAFX: Digital Audio Effects*, Second Edition", John Wiley & Sons , Ltd. Published 2011 by John Wiley & Sons , Ltd. ISBN: 978-0-470-66599-2.