Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi

Vol. 6, No. 3, Januari 2012 **ISSN 0216 – 0544**

SPECTRAL DENSITY BASED ON PHASE SHIFTING FOR MUSIC NOTATION

Yoyon K. Suprapto
Electrical Engineering Dept., Faculty of Industrial Technology
Sepuluh Nopember Institute of Technology, Surabaya 60111
E-Mail: yoyonsuprapto@ee.its.ac.id

Abstrak

Explorasi jenis musik tradisional diperlukan untuk mengetahui karakter music tersebut. Sementara perkembangan musik barat sudah sangat maju, musik timur, gamelan pada khususnya, masih mendapatkan stigma sebagai seni musik klasik dan terjebak dalam hal pelestarian seni tradisional, yang menghambat kreativitas untuk mengembangkan musik timur. Pengembangan mendalam musik timur diperlukan untuk membawa kembali kebesaran musik ini seperti yang ada di jamannya. Kami menerapkan pembuatan notasi musik guna memperluas pembelajaran bermain musik gamelan. Kami menggunakan *Adaptive Cross Correlation* (ACC). Percobaan menunjukkan metode yang diusulkan diberikan 2 - 7% untuk peningkatan kinerja gamelan nyata dibandingkan dengan metode konvensional. Dengan dimulainya riset dibidang musik gamelan, diharapkan pengembangan jenis musik ini dapat dilanjutkan seperti hubungan musik gamelan dengan musik barat, pembuatan musik gamelan elektronik, sehingga musik ini dapat mudah dibawa kemana saja sehingga menudahkan generasi muda untuk mempelajarinya.

Kata kunci: Ekstraksi peralatan musik, model suara Saron, pergeseran frekuensi dasar, Cross Correlation.

Abstract

Exploration traditional music is necessary to know the characters of this music. While the development of western music is very fast, eastern music, such as gamelan, in particular, still have the stigma as the art of classical music and get caught up in terms of preservation of traditional arts, which inhibit creativity to develop the eastern music. Therefore, the development of in-depth research of eastern music is needed to bring back the greatness of this music like the one in its era.We apply the Adaptive Cross Correlation (ACC) *method. Experiments show that the proposed method is given 2-7% for a real gamelan performance improvement compared to conventional methods. With the beginning of research in the field of gamelan music, the development is expected to continue this kind of music like gamelan music ties with western music, the manufacture of electronic gamelan, which can be easily taken and making it easier for young people to learn it.*

Key words: Extraction of Musical Instrument, Saron Spectral Density Model, The Pitch Shifting, Cross Correlation.

INTRODUCTION

Analysis of musical notation has attracted the attention of more than half a century, but major advances to the field of application has been made in the last two decades, mainly due to the expansion of computing capability. Several studies [1] using spectrum density analysis methods such as Fourier transformation, pitch analysis, heterodyne filter, phase vocoder, constant-Q transform, wavelet transformation, and distribution Weigner.

The paper focuses on observing the various methods of examination in the pitch, timbre, and review as a method for the extraction of various types of musical instruments, but only a few papers that mention references on traditional instruments Indonesia, gamelan. Musical instruments analysis is discussed in this paper [2].

Methods discussed in this study gives good results to analyze musical instruments such as the gamelan. Gamelan is a traditional musical instrument of Indonesia has several instrument families. Figure 1 shows a balungan family which consists of three instruments, namely demung, Saron and Peking. This instrument has several blades, and each blade indicates a gamelan notation.

Figure 1. Balungan Family from A Gamelan Set.

Table 1. The Pitch of Saron from Several Gamelan Sets.

SaronNotation	Pitch (Hz)	
	Min	Max
1	504	539
2	574	610
3	688	703
5	762	799
6	879	926

From Table 1, it can be seen that the saron pitch is not standard, it is influenced from the gamelan construction techniques which are carried out manually in accordance with the feelings of each of the instrument constructor, violence hit the gamelan, a position next to where the beat of the blades hitting the gamelan and style of musicians.

Thus we will too hard to make standardized forward to the set of gamelan music. In this study the authors try to present how to create a standard set of gamelan music by sampling frequency tone of Saron as a data base to conduct research.

Because of the complexity that comes from his playing style and the gamelan construction, the conventional automatic transcription can hardly be adopted. In the process of normalization, short time Fourier transform (STFT) is used to obtain the pitch and frequency domain profiles, as well as time-frequency characteristics [3]. Frequency and amplitude information are normalized according to the estimates of the pitch, f_0.

In this paper, new spectrum density model was suitable for the proposed use gamelan type of template matching to normalize variations in playing style. To evaluate the performance of the proposed method, the Saron family was chosen as the target of gamelan instruments.

SHORT TIME FOURIER TRANSFORM

Many algorithms developed for the extraction of the performance of musical instrument sounds using STFT. Previous researchers [4-6] analyzed the MIDI music or musical instruments made by manufacturer. While the MIDI music or musical instruments made good and fine tuned the signal envelope, the gamelan instruments were built mostly by hand.

Rodger McNab J et al [5] in his paper, they do shift the threshold to determine the pitch. Barbancho et al [4] and use it STFT sliding window to determine the onset and duration of the signal. Segmentation on the basis of amplitude and frequency of the base a median filter is applied to the detection function to determine the dynamic threshold function. JP Bello et al [6] wrote that for the synthesis process on every window, they use a comb-harmonic, the notation is expected to isolate the

relevant signal components to fill the gaps from the data base was created and arranged for their pitch. In the process of normalization, short time Fourier transform (STFT) is used by Barbancho et al [4] and J Rodger McNab et al [5] to obtain the pitch and the frequency domain as well as profiles of time frequency characteristics [7]. Normalized frequency and amplitude information in accordance with the pitch estimates.

ADAPTIVE CROSS CORRELATION

In this paper, a model was built to generate the spectral density of noise simulation saron. This sound is used as a reference sound on Adaptive Cross Correlation (ACC). Where ACC is a cross-correlation algorithm that uses a variety of window length and frequency shift of the basic methods used to reduce transcription errors associated with conventional music. Diagram ACC algorithm described in Figure 2. Simulated Saron sounds as reference signals applied to the process of cross correlation to establish the amount of cross power spectrum density.

The gamelan sound *x* is played with by the gamelan notation guided which with be compared with the reference signal y using cross correlation to form the cross power density.

Frequency Scale

There are two units that are often used to measure the distance between the Hertz and musical notation Cent music [8]. In the harmonic music also found some terms such as octave, fifth, etc. Octave is found when the pitch of a musical notation twice the pitch when compared with other notations. The ratio is 1:2 and a factor of two times and distances cent music 1200 Cent. Basic tone of western musical instruments like piano and guitar has 12 levels then the distance the nearest music notations in cent is 1200/12 or 100 cents. Cent displacement range of music in piano or guitar can be calculated by Equation (1) and Equation (2).

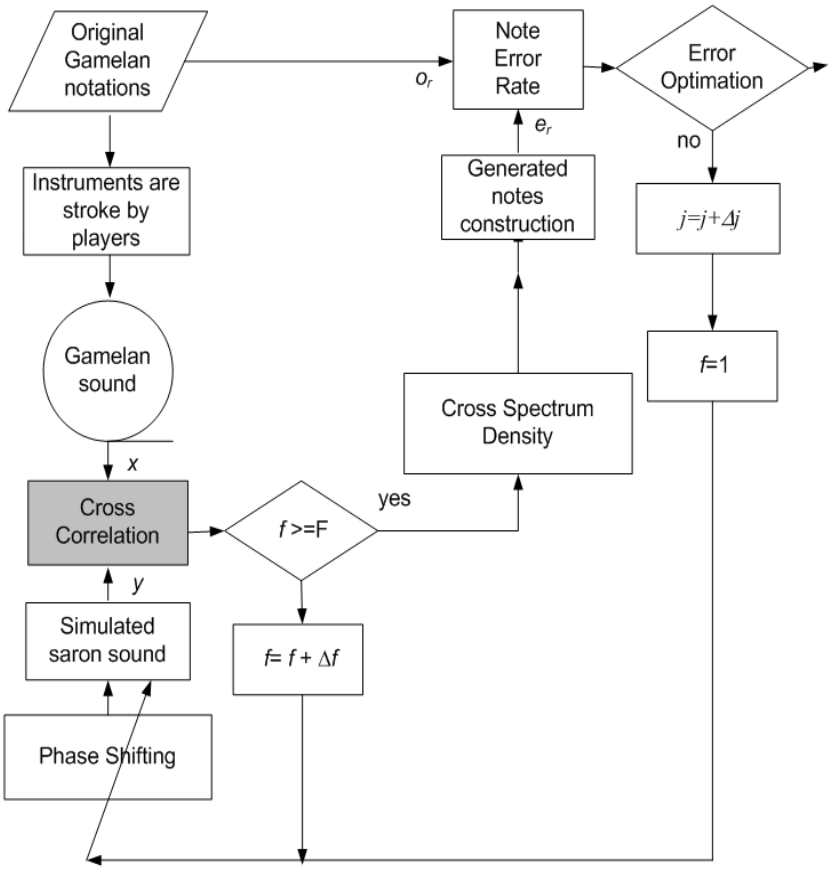

Figure 2. Diagram Sound Extraction Using Adaptive CrossCorrelation Sound.

$$C = 1200 \log_2(\frac{f_j}{f_i}) \quad (1)$$

or

$$f_j = 2^{\frac{m}{12}} f_i$$

$$f_j = 2^{\frac{n}{1200}} f_i \quad (2)$$

in this case *C* is the displacement cent, f_i and f_j is the pitch and the notation *j* to *i*, f_i and f_j are the pitch of both notes, m is the distance scale of notation *i* to *j*, while *n* is the increment / decrement the desired frequency.

For example, pitch shifting from tone *C* to tone *E*. There are twelve semi tones per octave and tones within the scale of *C* to *E* is a 4 step tone. This means that the upward shift of the pitch can be calculated by Equation (3).

$$f_E = 2^{(4/12)} \times f_C \quad (3)$$

or

$$f_E = 2^{((4\times100)/1200)} \times f_C \quad (4)$$

Gamelan music has different tones calculation cent music. Gamelan has Slendro and Pelog. While slendro has five levels of tone, Pelog has 8 levels [7,9]. Because the slendro has five levels of the distance the nearest music notation in cent is 1200/5 or 240 cents. Displacement distance of the slendro cent music can be calculated with Equation (5) and Equation (6).

$$C = 1200 \log_2(\frac{f_j}{f_i}) \quad (5)$$

$$f_j = 2^{\frac{m}{5}} f_i$$

$$f_j = 2^{\frac{n}{1200}} f_i \quad (6)$$

For example, pitch shifting of the basic saron1 to tone base saron2. There are 5 semi tones per octave. This means that the upward the pitch shifting can be calculated by Equation (7).

$$f_{saron2} = 2^{(1/5)} \times f_{saron1} \quad (7)$$

or

$$f_{saron2} = 2^{((1\times240/1200)} \times f_{saron1} \quad (8)$$

While the tone peloggained 8 levels so we get the Equation (9) and (10).

$$f_j = 2^{\frac{m}{8}} f_i \quad (9)$$

or

$$f_j = 2^{\frac{n}{1200}} f_i \quad (10)$$

the pitch shifting can be understood easily, such as atone is played twice as fast, then all the frequency will be doubled and the pitch will be shifted up one octave, but this causes the signal to be two times shorter. If the length of a tone signal is doubled without affecting the pitch and then played two times faster then all the frequency will be doubled, will shift the pitch and duration of the signal will be adjusted to the length of the initial signal.

To performthe pitch of atone shift, the tone must first be done to change the duration of the signal without changing the pitch. Scale factoris definedas the factorused to stretch or compress to adjust the frequency spectrum so that the pitch is shifted.

For example, if you want to shift the pitch of 240 Hertz, is required scale factor 2(240/1200). This means first of all we need to stretch the signal without changing the pitch so that the duration of the current multiplied by 2(240/1200). Once this is done, then the signal 2 (240/1200) times faster.

On the other hand, if it is desired to shift the pitch of 240 Hertz down, it takes a scale factor 2(-240/1200). This means it is first necessary to compress the signal without changing the pitch so that the duration is now multiplied by 2(-240/1200). Once this is done, then the signal is slower at 2(-240/1200) times the initial velocity.

Window Superposition

Signal is converted into the frequency domainusing multiple windows, STFT. This window is taken from the initial signal that 75% window overlap each other. See Figure 3. Window that is formed the nare spaced differently to stretch or compress the output signal.

If the distance window is changed, the signal is inverse, in the time domain, will create a discontinuity. This is shown in Figure 4.

This will create a sound that will seem foreign to the human ear. To eliminate the

foreign vote is necessary to eliminate the discontinuity.

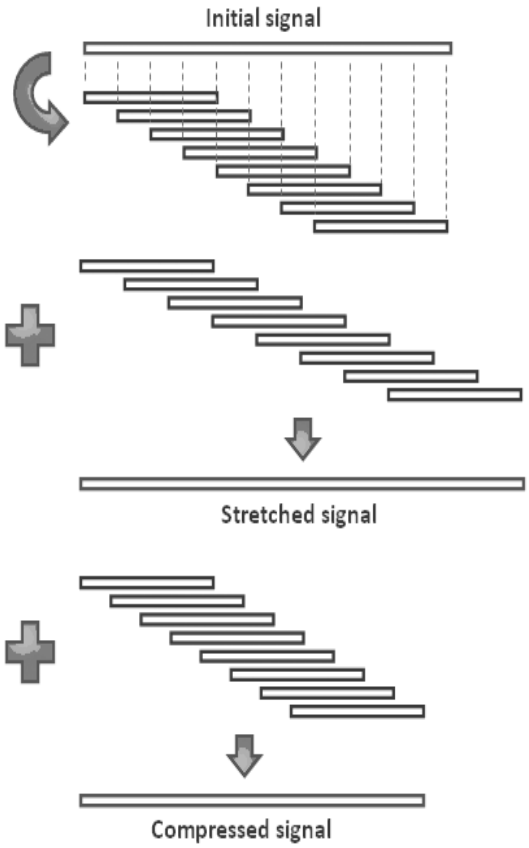

Figure 3. Stretched dan Compressed the Windows Signal to Preferm Pitch Shifting.

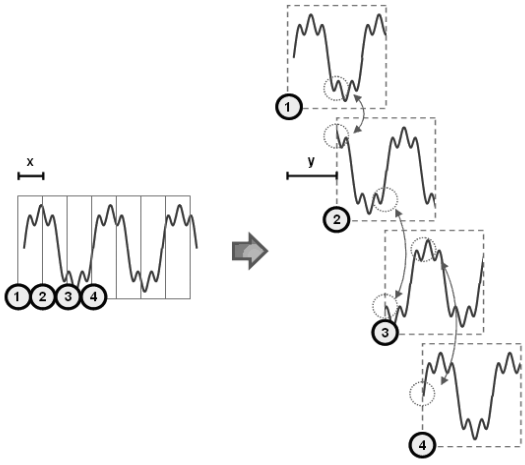

Figure 4. Discontinuous Signals when Stretching or Compression in The Time Domain.

PITCH SHIFTING

Pitch can be shifted by several algorithms:

--- ---

Algorithm 1 conversion time domain to the frequency domain

---

1. Read the input file (in WAV format)
2. Call the function to shift the frequency and content of the input parameters and the function
3. Perform FFT on Signals

Algorithm 2 shift the location of window

---

1. Give the value of the parameter function (the starting signal, the size of the window, the space between the window (hop), the desired number of shifts).
2. Determine the scale factor multiplying the pitch to the desired frequency 2 ^ (number ofshift / 1200).
3. Create a Hanning window.
4. Perform matrix for the manufacture of processed sound signal file.
5. Initialization.
   a. Create a window for the input signal matrix.
   b. Create a matrix to receive the window which has been processed.
   c. Initial value to the cumulative phase.
   d. Window gives the initial value of the previous phase.

Algorithm 3 Analysis

---

1. Get window of the signal to be processed Perform windowing on the signal.
2. Perform FFT on the windowing of the signal.
3. Find the magnitude by taking the value of the windowing absolute magnitude.
4. Find the angle (phase).

Algorithm 4 Processing

---

1. Find a phase difference between the previous phase.
2. Eliminate the phase difference.
3. Map to range between-π to π.
4. Find the true frequency.
5. Find the final phase.

Algorithm 5 Synthesis
---
1. Find the magnitude.
2. Produces the output window.
3. Save the window that has been in the process.
4. Additional overlap on a vector.
5. Resample with linear interpolation.

SARON SOUND EXTRACTION

Transcription can also mean rewriting a piece of music, either solo or orchestra. To write music notation automatically, several steps must be performed. Notation can be recognized and is usually done by changing from the time domain into the frequency domain. To write a gamelan musical notation, the sound of saron was extracted from the gamelan ensemble using cross correlation. The reason for choosing saron instrument, saron works as abased notation for other instruments.

Figure 5 shows the illustration of the process of the notations construction.

Desirable to the pitch shifting of the tone of 480 Hertz saron up. The first call is made saron tone through the main program,

```
d=wavread('SARON5SL.wav')
```

Then enter the value parameter to the function *pitchShift* with signal *init_input= d*, window's length= 500, hop size=50, frequency displacements=480.

```
y = pitchShift(d,500,50,480)
```

Referenced saron sound were constructed, *y*, was applied for as referenced signal on CrossCorrelation which ca be calculated using Equation (11).

$$R(l, f) = \frac{1}{J}\sum_{k=0}^{J-1} x(k+l)y(k, f) \qquad (11)$$

Where *l* is the lag time, the sampling index *k* and *J* is the length of *x* and *y*. If *f* is the frequency of 500-1100Hz scanning, *R(l, f)* be the crosss pectral density of the observed signal *x*.

Estimated saron wave form can be separated from the gamelan ensembles using pitch range of each notation saron, *N*.

$$p(N) = \max_{f=min(N)}^{max(N)} (|R(l, f)|) \qquad (12)$$

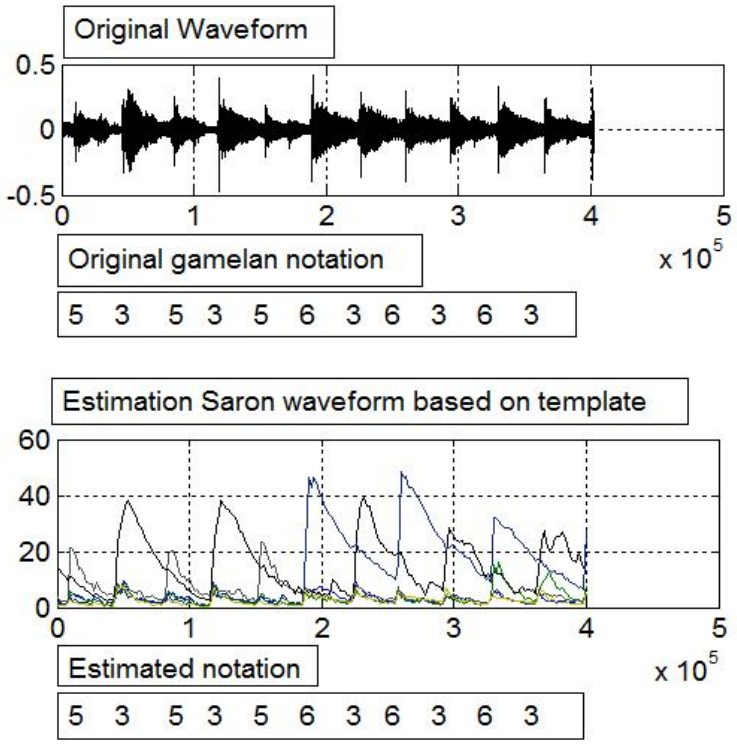

Figure 5. Illustration of The Process of the Notations Construction.

Table 2. Performance for Transcription Gamelan Saron Extraction by The Conventional Method of STFT and ACC.

Type of Test	STFT	ACC
1 synthetic	100%	100%
3 synthetic	100%	100%
1 real gamelan	98%	100%
3 real gamelan	95%	96%
Gamelan orchestra	85%	92%

Where *N*=1,2,3,5,6 is any gamelan notation, *p* is the approximate wave form of each saron notation. After all of the above process, it is necessary to eliminate the noise using a threshold. In areal gamelan performance, each record has a different maximumvalue of cross correlation, so that each record has its own threshold.

Input signal power is calculated over 50ms time window, and the resulting signal is used to record the segment in the input stream. The easiest way to record the segment is to set a threshold, which in dicates the beginning of a record electric current is exceeded, and the final note when power falls below it. Once the power is calculated over 50 ms windows length. Theen tire record of the beginning and end of the threshold set to 20% of this value. These values are obtained through experiments.

Candidate of the music notation scan be obtained by determining the peak of each signal. Each record has are cord number of offices, meeting power cross and on set.

RESULT AND DISCUSSION

In order to demonstrate the effectiveness of the templates that are suitable for extraction saron, various types of play, a single synthetic gamelan, a mixture of three synthetic gamelan, gamelan single concrete, a mixture of three real gamelan and gamelan. Extracted sound is used as the building of transcription. Table 2 shows the results as the ratio of the correct notation.

Notation of prospective candidates can be obtainedby determining the peak of the signal. Eachsignal is set a number of records a case of on set and magnitude.

To evaluate the note produced is estimated, we apply the *aon*, see Equation(13).

$$aon = 100\% \frac{totaltruthsentence}{original\_notation} \quad (13)$$

Where *aon* is the accuracy of manufacture of notation, *totaltruthsentence* is the correct total is detected the total original notation - (deletion + substitution + insertion). Insertion of undesired notation, deletion or notation is missing or incorrect notation defined, substituted, an example notation 5 should be recognized as a notation 3.

With many different types of gamelan songs being played, the results obtained are listed in Table 2.

CONCLUSION

Proposed method ACC shows better precision, about 2-7%, of musical transcription compared to conventional STFT method. This is due to ACC spectrum density can be more frequency scale dense than by STFT.

Of the experiment, the number of instruments is not sufficient influence the extraction performance of the instrument. Two instruments such as saron and bonang were played simultaneously, the performance is not always better than the five instruments. Saron and bonang have the same *f0*, so that it can affect the other instrument sound.

On the future will be developed using the pitch shifting technique phase vocoder, onset detection more detail.

REFERENCES

[1] Pielemeier WJ, WakefieldGH, and Simoni MH. Time-Frequency Analysis of Musical Signals. *Proceedings of The IEEE*.84: 1216-1230. 1996.

[2] Havelock D, Sonoko K, Vorlander M. *Handbook of Signal Processing in Acoustics*. NewYork:Springer. 2008.

[3] Alm JF, Walker JS. Time Frequency Analysis of Musical Instruments. *SIAM*

REVIEW, Society for Industrial and Applied Mathematics. 44:457-476. 2002.

[4] Barbancho LT and Jurado A. Transcription of piano recordings. *Applied Acoustic*. 65:1261-1287. 2004.

[5] Rodger LAS, McNab J, and Witten IH. Signal processing for melody transcription. *Proceedings of the 19th Australian ComputerScience Conference, Melbourne*. Australia. January 31-February 2. 1996.

[6] Daudet L, Bello JP, and Sandler MB, Automatic piano transcription using frequency and time-domain information. *IEEE Transaction on Audio, Speech and Language Processing*. 14:2242-2251.2006.

[7] Usagawa T, Yoyon K Suprapto, and Hariadi M. Time frequency modelling of gamelan instrument based on spectral density for automatic notation. *Proceedings Of The Third International Student Conference on Advanced Science and Technology*. 2009.

[8] Carterette EC and Kendall RA. On the Tuning and Stretched Octave of Javanese Gamelans. *Leonardo Music Journal*. 4:59-68. 1994.

[9] Sumarsam. *Cultural interaction and musical development in central java*. The University of Chicago Press. 1995.