Journal of Theoretical and Applied Information Technology
10th July 2013. Vol. 53 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# IDENTICAL SOUND DETECTION OF USING LEAST MEAN SQUARE – ADAPTIVE CROSS CORRELATION FOR TRANSCRIPTION

**[1]YOYON K. SUPRAPTO, [2]ARIS TJAHJANTO, [3]DIAH PUSPITO WULANDARI**

[1,3]Electrical Engineering Department in Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

[2]Information Systems Department in Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

**ABSTRACT**

Development of eastern music like gamelan is far lagged from that of western music because gamelan is still stigmatized as part of traditional arts which must be preserved instead of being analyzed and developed. Therefore development of in depth research concerning gamelan music is needed to bring back the greatness of this music like that in its era (17th-18th century).

This research initiates the gamelan sound extraction for music transcription. We applied Least Mean Square - Adaptive Cross Correlation (LACC). ACC was conducted to generate spectral density for music transcription while LMS was utilized to detect instruments which have identical fundamental frequency, in order to avoid over detection. Experiment demonstrates the test performance demonstrates that the proposed method provided 2 - 12 % improvement for real gamelan performance comparing to conventional methods such as STFT.

**Keywords:** *Saron time frequency model, Adaptive cross-correlation, saron extraction, music transcription, Least Mean Square***.**

## 1. INTRODUCTION

While western music perceives that good musicis composed with stable tones, regulated frequency, fixed amplitude, the eastern music such as gamelan has freely imposed tone in terms of resonance, tone color and amplitude or frequency [1]. Gamelan is one of Indonesia's traditional music whose repetitive playing pattern has been increasingly accepted by international composers [2,3] such as Claude Achille Debussy (French, 1910) Bella Bartok (Hungarian, 1923), Colin Mc Phee (U.S., 1930), Backet Wheeler (U.S, 1960). That is why deep research of gamelan sound is needed.
Gamelan consists of about ten groups of instruments. They are saron, kenong, kempul, kendang, bonang, etc. [4]. Figure 1 shows saron group. Figure 2 shows another group called bonang. Saron is constructed from several metal blades, Bonang is constructed from several small gongs. Bonang comprises of two sets of small gongs, where the first small gong set have identical fundamental frequencies with saron, and the second small gong set have fundamental frequencies which are one octave higher than those of the first set.
This research focuses on the extraction of saron sound. Most of gamelan instruments consist of only one octave [4]. Each blade or gong represents a notation.

*Fig.1. The Saron Family In A Gamelan Set.*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Fig 2 Bonang Group In A Gamelan Set.*

Gamelan instruments are manually constructed by hand. Constructors tune the instruments with their own sense, based on experience. As a result, fluctuation of frequency inside the sound is not set correctly. Therefore the fundamental frequency of one gamelan's instruments could be slightly different from one *gamelan* set to the others. Table 1 shows *saron* fundamental frequencies from several gamelan sets. Each note has varying frequency range. Saron has only one octave sound, and the other octave is belonged to the other instrument.

In gamelan ensemble, an instrument sound is always interfered by those of other instruments. For example, the extracted saron sound may still contain bonang sound since both instruments have the same fundamental frequency. But the presence of bonang sound can be distinguished from saron sound by comparing the spectral envelope of both sounds, since bonang sound (60 ms) has shorter envelope than that of saron (300 ms). See fig 3.

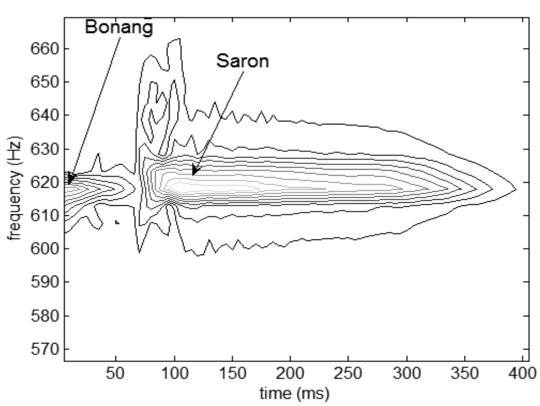


*Fig.3 Spectrum Bonang And Saron*

In gamelan music, saron and bonang are not sounded at the same time. Bonang is struck a half beat time before saron, where beat in this case is defined as the distance between two consecutive Saron sounds, See Fig. 4.

*Table 1 Saron Fundamental Frequency From Several Gamelan Sets.*

| Saron Notation number | Fundamental Frequency (Hz) | |
|---|---|---|
| | Min | Max |
| 1 | 514 | 544 |
| 2 | 589 | 620 |
| 3 | 681 | 716 |
| 5 | 776 | 820 |
| 6 | 896 | 942 |

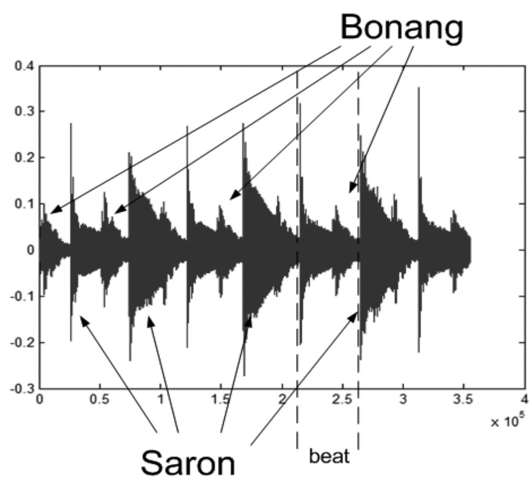


*Fig. 4 How Saron And Bonang Were Played*

Table 2 shows *bonang* fundamental frequencies from several gamelan sets. Bonang has two octave sounds.

Gamelan can be played by applying different strength according to the rhythm of the song. This leads to the variations of magnitudes of each note as well as the emergence of harmonic frequencies within a note. Figure 5 shows the spectrum of gamelan sound which varies very much due to the strength and the style of stroke although it still has the same fundamental frequency. This figure shows a fundamental frequency of 600 Hz. The harder stroke resulted in the emergence of harmonic sound which has lower frequency than that of gamelan sound. Since there is a number of gamelan instruments have the same fundamental frequency and the complexity of gamelan playing style, it

Journal of Theoretical and Applied Information Technology
10th July 2013. Vol. 53 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

causes conventional music transcription could be hardly adopted.

*Table 2 Bonang Fundamental Frequency From Several Gamelan Sets.*

| Bonang Notation | Fundamental Frequency (Hz) | |
|---|---|---|
| Number | Min | Max |
| 1 | 252 | 270,5 |
| 2 | 290 | 312,5 |
| 3 | 344 | 357,5 |
| 5 | 390 | 410 |
| 6 | 458 | 464 |
| 1 | 504 | 541 |
| 2 | 580 | 625 |
| 3 | 688 | 715 |
| 5 | 780 | 820 |
| 6 | 916 | 928 |

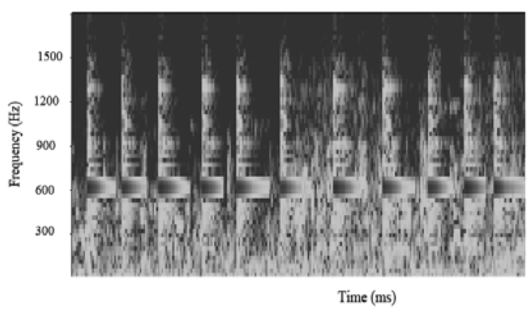


*Fig.5 The Difference Spectrum Of The Signal Due To Differences In Hammer Stroke Strength*

In this paper we introduced the suitable method, Least Mean Square Adaptive Cross Correlation (LACC). *Saron* instrument was chosen as the target group of *gamelan* extraction due to the use of *saron* note as the basic note for other instruments.
The remainder of the paper is organized as follows. Section 2 briefly reviews the previous works, most related ones to our approach which apply STFT. Section 3 describes our proposed method. We also describe the spectral density model which is constructed for generating simulated saron sound. Section 4 describes the performance evaluation. We show and investigate the accuracy of the model for various types of gamelan playing styles. We also show the conventional methods and our proposed one which were evaluated with our test data. Section 5 concludes the paper.

## 2. CONVENTIONAL METHOD

Previous works [5] [6] [7] in music transcription mostly used STFT to extract the sound of particular instrument in Musical Instrument Digital Interface (MIDI) music performance. We apply STFT for acoustic music performance. The STFT was used in sound extraction process as a comparison with our proposed method LACC. Finally, both methods were evaluated by using the same data, gamelan music performance.

Barbancho et. al. used STFT and sliding windows to determine the onset and time duration of a notation's sound [5]. Rodger J McNab et. al. shifted slightly the threshold to determine fundamental frequency [6]. Extraction was carried out based on fundamental frequency and its power density [7]. Bello reported in their paper that for synthesis process, they used harmonic combs of estimated notes to isolate the relevant sound components. They also created a database of an instrument sound for diverse frequencies and filled the gaps of the database by synthesizing an instrument sound for particular fundamental frequencies. In normalization process, STFT was used by Barbancho, McNab, and Witten to obtain the fundamental frequency of a notation as well as to acquire spectral density characteristics. Previous researchers (Barbancho, McNab, Bello) mostly analyzed MIDI music that resulted from fabricated music instruments, which is well tuned and has uniform notation sound envelopes.
Using the STFT, the frequency resolution and time resolution depends on the length of the window. With sampling frequency 48 kHz, if we use STFT 8192, the frequency resolution is 5.85 Hz. Thus, by using this window length, we cannot distinguish two sounds whose fundamental frequencies differ from each other with less than 5.85 Hz difference. For example fundamental frequency of bonang and saron are 617 Hz and 614 Hz respectively. Both sounds were recognized as 615 Hz which were shown in Fig. 6.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

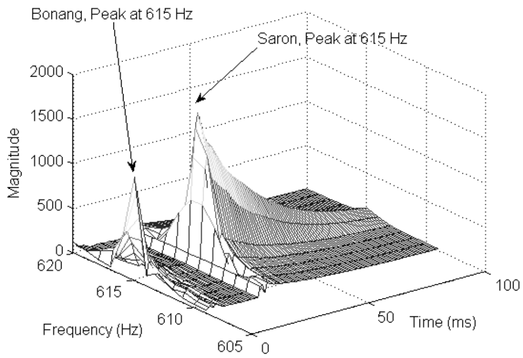


*Fig. 6 Stft 8192 Spectrum Density*

## 3. PROPOSED METHOD

Our proposed method can be divided into two sub methods, namely the ACC to generate notes candidate, and the LMS to detect instruments which have identical fundamental frequency that may cause over detection in music transcription. Block diagram of our method can be seen in Figure 7.

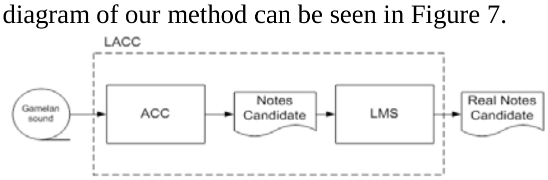


Fig.7 The block diagram of LACC Algorithm

### 3.1 Adaptive Cross Correlation

ACC is a modified cross-correlation algorithm [8] which is described in Figure 8. Where $J$ is window length, $f_0$ is fundamental frequency, $\Delta J$ is increment of window length, and $\Delta f_0$ is increment of fundamental frequency.

Gamelan sound $x$ was generated by hitting the instrument with a hammer which was guided by the original gamelan notation. Gamelan sound was recorded in WAV format. The simulated saron waveform $\bar{y}$ was applied as a reference signal on the cross correlation process to form the magnitude of cross power spectrum density. Table 1 shows that saron sound fundamental frequencies are between 500 Hz to F, where F is 1000 Hz. We constructed a time-frequency model of saron instrument by pitch shifting. The model generated various simulated saron sounds by adjusting the fundamental frequency by 1 Hz and changing the window length by $\Delta J$. This process could be performed through several iterations. In each iteration, the window length was enlarged by $\Delta J$ until the spectral density was convergent.

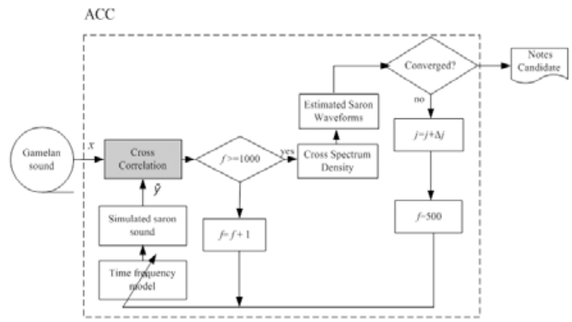


*Fig. 8 Block Diagram Of Adaptive Cross Correlation (ACC)*

The output of this method is note candidates which contain the onset, estimated pitch, and the length of the sound.

#### 3.1.1 Gamelan Frequency Scale

There are two units that are often used to measure the distance between two notations, there are the Hertz and Cent music. Octave is used when the fundamental frequency of a musical notation is twice the fundamental frequency of other notation. Gamelan has different scales called Slendro and Pelog. While Slendro has five level of notes in an octave (Pentatonic), Pelog has seven level of notes in an octave (heptatonic) [4]. The distance between nearest notes for Slendro gamelan instruments is 1200/5 or 240 cents. Displacement distance of the slendro cent music can be calculated with equation (1) and equation (2).

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

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

where $f_i$ is fundamental frequency of Saron $i^{th}$ and $f_j$ is the fundamental frequency of saron $j^{th}$, while $m$ is saron notation increment / decrement and $n$ is the increment / decrement of the desired frequency in cent.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**3.1.2 Saron spectral density model Building**

Simulated saron sounds were produced by pitch-shifting method [9]. The spectral density model is built as the template to generate simulated saron sound which is used as reference sound on cross correlation process[10][11].

Since the saron sound was assumed to be sinusoid with several harmonics which are also sinusoid, these sinusoids could be observed by taking the cross-correlation $r(l,f)$ between the sound block $x(k)$ and sinusoids sound $y(k,f)$. The cross correlation of two finite-length sequences $x(k)$ and $y(k)$ [11] each of length $J$ is defined by Eq.(3),

$$r(l,f)=\frac{1}{J}\sum_{k=0}^{J-1}x(k+l)y(k,f) \qquad (3)$$

where $f$ is frequency, $k$ is sampling index, $l$ is lag and $y(k,f)$ is $\sin(2\pi\ kf/f_s)$ and $f_s$ is sampling frequency. The magnitude of cross power spectral density $p$ is determined by using the peak value of the magnitude $r(l,f)$ using Eq.(4) for all frequencies,

$$P(f)=\max_{l}\left|r(l,f)\right| \qquad (4)$$

Fundamental frequency estimation of sound $x(k)$ is calculated using Eq.(5),

$$f_0(c)=\mathop{\arg\max}_{f=\min(f_0(c))}^{\max(f_0(c))}P(f)+\min(f_0(c)) \qquad (5)$$

where $c$ is saron notation number. arg max stands for the argument of the maximum gives a position at which $P(f)$ is maximized in each notation $c$. See Table 1. Simulated saron sounds are important for sound extraction. To construct the simulated saron sound, we need a saron time frequency model. The model was constructed from several single strokes of saron sounds, called saron pre-recorded sounds. Our database of Pre-recorded saron sounds consist of 600 single stroke sounds from 30 gamelan sets. Since gamelan are percussive, the playing style applied to gamelan instruments also contributes some variations, for example, the strength of the stroke (hard, medium, soft), and the hammer stroke areas (upper edge, lower edge, center). These three conditions induces variations in terms of fundamental frequency, signal envelope, and harmonic contents of an instrument's sound among different gamelan sets.

The sounds are converted to Saron time-frequency domain using cross correlation. The process continues by registering the pre-recorded sounds as training data. Each label of pre-recorded sound contains notation name, instrument number, pre-recorded sound number and its fundamental frequency estimation. "*Saron 6*", the sixth notation of saron instrument, was chosen as the reference note for normalization [4].

The frequency shift of pre-recorded sound $s$ to the saron 6 can be calculated using Eq. (6).

$$\Delta c_s=1200\log_2\left(\frac{f_{0s}}{f_{06}}\right) \qquad (6)$$

where $s$ is pre-recorded sound number, $f_{0s}$ is fundamental frequency of pre-recorded sound $s$, $f_{06}$ is the fundamental frequency of Saron 6 as the reference tone. Based on the pitch shifting method, all frequency components are shifted by same $\Delta c_s$. Note that ideal Saron 6 fundamental frequency $f_{06}$ was obtained from the average of the sixth notation fundamental frequency of saron instrument from several gamelan sets.

Since the spectrum of sound during attack time is different from that during sustain time, the pre-recorded sound was divided into several short period windows. In order to obtain the spectral density model cross correlation $r_{bm}(l,t,f)$ between $x_{bm}(t+l,f_{0m})$ and $y(t,f)$ can be seen in Eq. (7),

$$r_{bm}(l,t,f)=\frac{1}{J}\sum_{j=t}^{t+J-1}x_{bm}(j+l,f_{0m})y(j,f) \qquad (7)$$

where $b$ is blade index, $m$ is pre-recorded sound index, $t$ is time index, $J$ is window's length, $f_{0m}$ is estimated fundamental frequency of pre-recorded sound.

The peak of cross-correlation $r_{bm}(l,t,f)$ at time $t$ and frequency $f$ is defined by Eq.(8),

$$P_{bm}(t,f)=\max_{n}\left(\left|r_{bm}(l,t,f)\right|\right) \qquad (8)$$

The spectral density model $A(t,f)$ is determined by the average value of cross-correlation $P_{bm}(t,f)$ for all of pre-recorded sounds as shown in Eq. (9),

$$A(t,f)=\frac{\sum_{m=1}^{M}P_{bm}(t,f)}{M} \qquad (9)$$

where $A(t,f)$ is spectral density model, $M$ is total pre-recorded sounds for each blade index $b$. By using ACC, the frequency resolution is 1 Hz, so we can distinguish the sound of bonang from that of saron using STFT, since bonang and saron may have slight difference of fundamental frequency which ranges from 0 to 3Hz , which was shown in Fig. 9.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

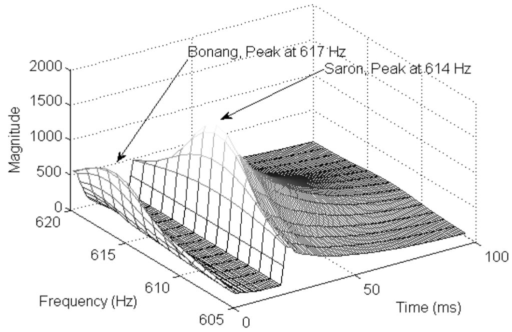


*Fig. 9 Cross Power Density Using ACC*

### 3.2 Least Mean Square (LMS) filter

LMS filter is different from fixed point filter [10][11]. Figure 10 shows LMS method[12].

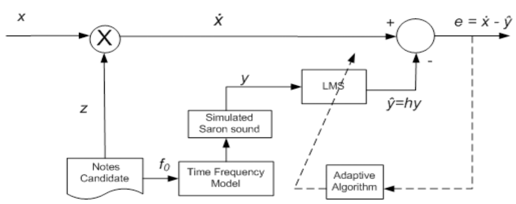


*Fig.10 LMS Filter Block Diagram.*

In order to extract the number of identical pitch instruments, we utilized the note candidates resulted from ACC each of which owns its onset time, fundamental frequency estimation and the signal length. The length of each note represents the note's onset and offset time and denotes as sub interval [t1, t2] time. In the next step we measure the number of onsets that occur within [t1, t2]. We consider it as an estimation of the number of identical pitch instruments. The estimated fundamental frequency of a note candidate is employed to generate simulated saron sound as a reference signal being used in LMS process.

If $x$ is input signal, $y$ is reference signal, $h$ is filter weight, $\mu$ is step size, $e$ is error signal, $n$ is iteration number and $p$ is filter order.

$$y(n) = [y(n)\ \ y(n-1) \ldots\ y(n-p+1)]^T$$

$$h(n) = [h(n,0)\ h(n,1) \ldots\ldots h(n,p-1)]^T$$

$$h^H(n) = [h^*(n,0)\ h^*(n,1) \ldots\ldots h^*(n,p-1)]^T$$

$$\hat{y}(n) = h^H(n).y(n) \qquad (10)$$

where $h^H(n)$ [13] is conjugate transpose of $h(n)$, $h^*$ is complex conjugate[14].

$$e(n) = x(n) - \hat{y}(n) \qquad (11)$$

$$\xi = E\left[e^2(n)\right] = \text{minimum} \qquad (12)$$

Minimum when

$$\frac{\partial \xi(h)}{\partial(h)} = 0 \qquad (13)$$

$$\begin{aligned} h(n+1) &= h(n) + \Delta h(n) \\ &= h(n) + \mu \frac{\partial \xi(h)}{\partial(h)} \\ &= h(n) + 2\mu e(n) y(n) \qquad (14) \end{aligned}$$

## 4. PERFORMANCE EVALUATION

Gamelan sound is influenced by many external factors such as hammer stroke area and stroke strength, so it has various different magnitude of each note [14]. In real gamelan performance, since each note might have different maximum magnitude of cross power density, so each note has its own threshold. The magnitude of the input sound was calculated over 50 ms time frames, and the resulting sound was used to segment notes in the input stream. The simplest way to segment notes is to set a threshold, denoting a note start when the magnitude of power density exceeds it, and a note end when the magnitude drops below it. Having calculated the magnitude over 50 ms frames, an overall magnitude figure was calculated for the entire input sound and the threshold for note start and note end was set to 20 % of this value. These values were arrived through experiments.

### 4.1 Notes Candidate construction using ACC

#### 4.1.1 Saron Sound Separation

The simulated saron sound $\bar{y}(t,f)$ is applied as a reference sound on the cross-correlation process to form the cross power spectrum which would generate saron estimation waveforms using Eq. (15).

Journal of Theoretical and Applied Information Technology
10th July 2013. Vol. 53 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\bar{y}(t,f) = \sum_{\Delta i=-f+1}^{2f} \cos(2\pi(f+\Delta i)t/f_s)A(t,f+\Delta i) \quad (15)$$

Transcription could be generated by using the extracted sound. The cross-correlation between the observed sound and reference sound is calculated using Eq. (16)

$$r(l,t,f) = \frac{1}{J}\sum_{j=t}^{t+J-1} x(j+l)\,\bar{y}(j,f) \quad (16)$$

where $f$ is frequency scanning from 500 to F Hz , $r(l,t,f)$ becomes cross spectral density of observed sound $x$. The cross correlation results in separated saron sounds among the other instrument sounds. The next step is to translate the separated saron sounds into note candidates. Equation (17) shows that by setting the value of $f$ to be in particular interval $min(f_0(c)) \leq f \leq max(f_0(c))$ we obtain the occurrences of estimated saron waveforms which are note candidates across time.

$$pb_{bm}(t,f) = \arg\max_{m=\min(f_0(c))}^{\max(f_0(c))} \left(\frac{\sum_L |r_{bm}(l,t,f)|}{L}\right) + \min(f_0(c)) \quad (17)$$

where $c$ = 1,2,3,5,6 is each gamelan note, $L$ is the length of observed signal $x$.

Normalized $p(t,c)$ can be calculated using Eq.(18),

$$p_N(t,c) = \frac{p(t,c)}{\max(p(t,c))} \quad (18)$$

### 4.1.2 Convergent Spectral Density

This process must be performed through several iterations until the spectral density is convergent. To determine whether the process has been convergent or not we applied the Mean Square Error ($MSE$), which can be calculated using Eq.(19),

$$MSE(n) = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{J}\sum_{k=1}^{J}(p(n,k,c)-p(n-1,k,c))^2 \quad (19)$$

where $n$ is iteration index, $C$ is total gamelan notes, for pentatonic $C$ =5. On each iteration we modified window size $J$ by $\Delta J$.

Convergence is achieved if $MSE(n) - MSE(n-1) <= v$, where $v$ is a value. The result can be seen in Fig. 11. During experiment $v$ = 0.0001. Figure 8 shows that the larger the window length (ms) the smaller the MSE.

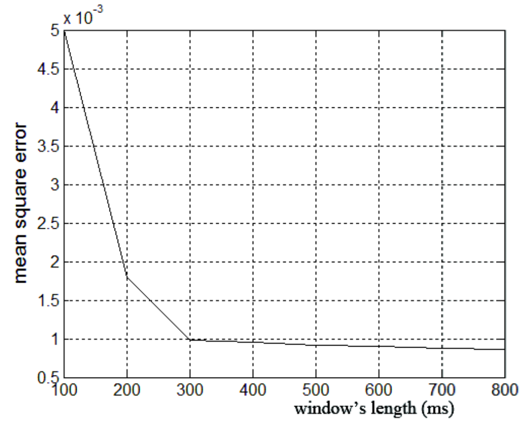

*Fig. 11 MSE Against Various Window's Lengths*

Besides Saron, we also considered other instruments which have the same fundamental frequency, such as Bonang, which generate pulses for saron sound. These pulses may cause an over detection in transcription.

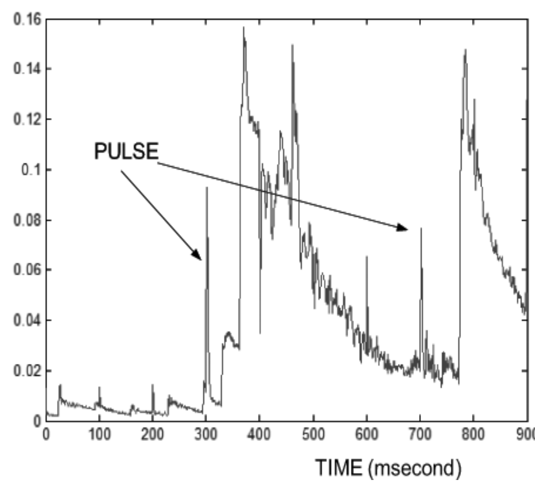

*Fig. 12 Instruments Which Have An Identical Fundamental Frequency Can Generate Pulses*

### 4.2 Detection of instruments with identical pitch

The Main purpose of this method is to detect the number of instruments which have similar fundamental frequency, played simultaneously to avoid over detection during transcription process.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

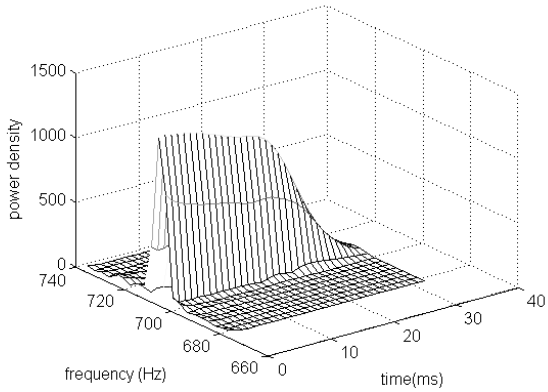


*Fig.13 Spectral Density Of A Note Candidate Which Have An Identical Fundamental Frequency, Bonang And Saron,,Played Simultaneously.*

Each note candidate, that is shown in Fig. 13, gives no guarantee that it contains only one instrument, so we should detect number of instruments that have the same fundamental frequency are struck simultaneously. This detection can be performed using LMS which can be followed using algorithm (1).

---

Algorithm 1 LMS .

---

1. Set first note candidate, Crossed to observed signal $x$
2. Set step size $\mu = 0.001$
3. Set filter order $p$ = length of a note candidate signal.
4. Set $h(0) = 0$
5. Set first iteration, $n = 1$
6. Set filter output $\hat{y}(n)=h^H(n).y(n)$
7. Set estimation error $e(n) = x(n)- \hat{y}(n)$
8. The next filter weight $h(n+1) = h(n) + 2\mu e(n)y(n)$
9. $m(n)= max(h(n))$
10. Set next iteration $n = n + 1$
11. The process , goto 5, will be continue until $(m(n+1) – m(n))<0.01$
12. Set next note candidate, get another onset, estimated pitch, signal length
13. IF not eof() goto 2

---

Any onset of instruments that have the same fundamental frequency will be detected as a peak in the weight filter $h(n)$ which can be seen in Fig. 14. Weights filter coefficients $h(n)$ performed after a several time iterations will appear a few peaks that indicate how many instruments are contained in the note candidate signal.

The Gamelan Songs for Testing

We generated three types of gamelan sound for testing :

1. Synthetic. The sound of synthetic gamelan can be made using our gamelan emulator. Gamelan notation is considered as the input description for gamelan emulator, number of instruments and type of instrument that can be set. Synthetic gamelan sound can be recorded in WAV format.
2. Acoustic. Gamelan ensemble was played by the players which was guided by the original gamelan notation and was recorded.
3. Commercial Recorded from CD or cassette. The notations must be evaluated by the experts to determine the original notes.

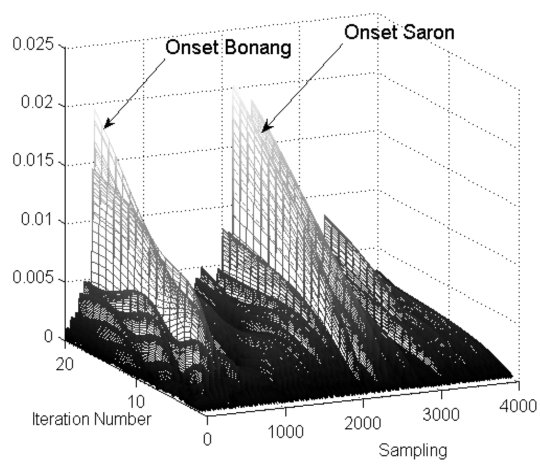


*Fig.14 Filter Weight H(T) Shows Detected Instrument Number Which Have An Identical Fundamental Frequency.*

To evaluate the estimated generated notes, we used the note error rate [11][12]. Recognition of error rates is stated by Eq.(20),

$$ner = \frac{deletion + insertion + substitution}{totaltruesentence} \quad (20)$$

The overall results were then compared with our proposed method. Table 3 shows the results as the ratio of *ner*. Where *ner* is note error rate, *total true sentence* is *original notation - (deletion +substitution +insertion)* .

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Tabel 3 Performance Of Saron Extraction For Gamelan Transcription From Several Type Of Gamelan Sounds.*

| Test Type | Note Error Rate (ner) | | | | |
|---|---|---|---|---|---|
| | Duration (s) | Instru Ment # | 8192 STFT | ACC | LACC |
| Synthetic | 10 | 2 | 2% | 0% | 0% |
| Synthetic | 10 | 3 | 6% | 4% | 4% |
| Acoustic | 21 | 2 | 2% | 2% | 2% |
| Acoustic | 21 | 2 | 2% | 1% | 1% |
| Acoustic | 63 | 8 | 8% | 5% | 3% |
| Acoustic | 43 | 10 | 18% | 16% | 6% |
| Commercial Recorded | 60 | 10 | 12% | 17% | 8% |

The following is a list of musical instruments: plucked Instruments (guitar, piano), wind instruments (flute, trumpet), bowed sting instruments (violin, cello), percussion instruments (xylophone, gamelan). The envelope of audio signals obtained wind instruments and bowed string instruments have slightly different shape compared to that of bowed string instruments and percussion instruments: the envelope has much longer attack area depending on the length of excitation time and it also has much shorter decay and release area since the signal fades almost directly after the end of excitation time. LACC is suitable for evaluating monophonic music. Beside percussion instruments (xylophone, gamelan), LACC method is also applied to evaluate either acoustic music or synthetic music, plucked Instruments (guitar, piano) for music transcription. LACC has not been able to evaluate the type of polyphonic music.

## 5. CONCLUSION

In this research, we employ the ACC method for gamelan transcription of saron instrument. The LMS is used to detect instruments which have same fundamental frequency and are played simultaneously. This detection is applied to avoid over detection during transcription process.

The test performance demonstrates that the proposed method provided 2 - 12 % improvement for real gamelan performance comparing to conventional methods such as STFT. These results show the effectiveness of template matching for picking up specified instrument for music transcription.

## 6. FUTURE WORKS

In the future, LACC can be applied to several research areas such as accurate onset detection, Source number estimation, Sound Tagging and Sound Watermark.

## REFERENCES

[1] Sutton, Anderson, 1993. Central Javanese gamelan music: dynamics of a steady state, Northern Illinois University in DeKalb, Il.

[2] Olmo Cornelis, Micheline Lesaffre, Dirk Moelants, Marc Leman,“ Access to ethnic music: Advances and perspectives in contentbased music information retrieval,” Signal Processing 90 Elsevier, Amsterdam, pp. 1008-1031, 2010

[3] Tamagawa K., 1998. Echoes from the east: The Javanese gamelan and its influence on the music of Claude Debussy, D.M.A. document. The University of Texas at Austin.

[4] Sumarsam, 1995. Cultural interaction and musical development in central java, The University of Chicago Press, ISBN 0-226-78011-2.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[5] Barbancho, A. Jurado, 2004. Transcription of piano recordings, Applied Acoustic 65 (12 SPEC. ISSN).
[6] Rodger J. McNab and I. H. Witten, 1996. Signal processing for melody transcription, Proceedings of the 19th Australian Computer Science Conference, Melbourne, Australia.3
[7] Bello and M. B. Sandler, 2006. Automatic piano transcription using frequency and time-domain information, IEEE Transaction on Audio, Speech and Language Processing, vol. 14 no 6.
[8] Yoyon K. Suprapto, Hariadi and Hery Purnomo, "Traditional Music Sound Extraction Based on Spectral Density Model using Adaptive Cross-correlation for Automatic Transcription"; IAENG International Journal of Computer Science, VOL 38, 2011
[9] Dolson, 1996. The phase vocoder: A tutorial, Computer Music Journal, vol. Vol. 10, No. 4.
[10] Pielemeier Willam J. and M. H. Simoni, 1996. Time-frequency analysis of musical signals, Proceedings of The IEEE, vol. Vol.84, No.9.
[11] David Havelock, Sonoko Kuwano, 2008. Handbook of signal processing in acoustics, Springer New York.
[12] Sophocles J. Orfanidis, 1988, Optimum Signal Processing, MacMillan Publishing Company, New York, 2nd edition.
[13] Golub, G. H. and Van Loan, C. F. Matrix Computations, 3rd ed. Baltimore, MD: Johns Hopkins University Press, p. 14, 1996.
[14] Courant, R. and Hilbert, D. Methods of Mathematical Physics, Vol. 1. New York: Wiley, 1989.