

Traditional Music Sound Extraction Based on Spectral Density Model using Adaptive Cross-correlation for Automatic Transcription

Yoyon K. Suprpto, *Member, IAENG*, Mochamad Hariadi and Mauridhi Hery Purnomo

Abstract—Nowadays, mining of the musical ensemble attracts the interests in several aspects since the importance of archiving traditional musical performance is emphasized. However, there are very few of them which take into account the Indonesian traditional instrument called Gamelan. While western music perceives that good music is composed by stable tones, the eastern music such as gamelan has freely imposed tones in terms of resonance and tone color. Exploration of the gamelan music is very rare, so its development is far lagged to western music. The in-depth development of gamelan music is needed to bring back the greatness of this music like the one its era (17th–18th century). This research initiates gamelan sound extraction for music transcription as part of traditional music analysis.

In this research we introduce a new method to generate music transcription for gamelan. The spectral density model is built to extract the sound of an instrument among the others by using Adaptive Cross Correlation (ACC). The experiment demonstrates 16% note error rate for gamelan performance.

Index Terms—Time and frequency model, saron extraction, adaptive cross-correlation, automatic transcription.

I. INTRODUCTION

There are some differences between western music and eastern music. Casey [1] provides an overview of the advances of audio-based feature extraction and classification methods applied to western music. In the present paper, we address the difficulties of handling eastern traditional music such as Gamelan. The main problem with ethnic music is that it does not always correspond to the western concepts that underlie the currently available content-based methods [2]. While western music perceives that good music is composed by 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 [3]. Fewer and

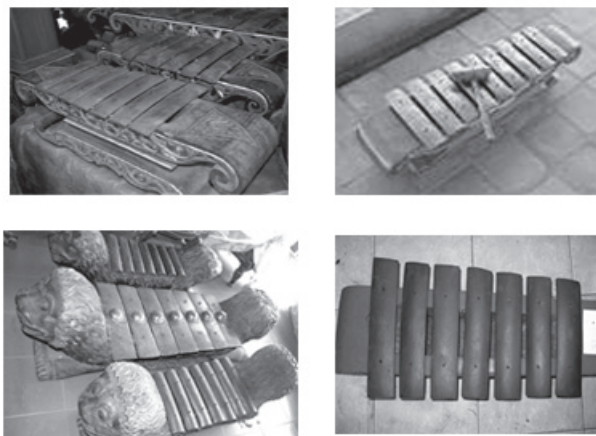


Fig. 1. The saron group from several gamelan sets.

fewer people cares about the traditional music, so that its development is increasingly lagged to that of western music. Gamelan still gets the stigma of being the art of traditional music and stuck in the notion of preservation of traditional arts. Therefore, gamelan sound in-depth analysis is needed.

Gamelan is one of Indonesia's traditional music which its repetitive playing pattern is increasingly accepted by international composers. Many world-class musicians have already accepted eastern music concepts such as Bella Bartok (Hungarian, 1923), Colin Mc Phee (U.S. 1930), Backet Wheeler (U.S. 1960), Claude Achille Debussy (French composer, 1910) [4].

Gamelan is constructed from about fifteen groups of different instruments. Figure 1 shows saron group from several gamelan sets. This instrument consists of only an octave. Its pitched tone is pentatonic or heptatonic. Each blade represents a gamelan notation. The other octave is belonged by the other instrument. The instruments in a gamelan set are played simultaneously, like the ones in an ensemble. Gamelan notations are very simple, they consist of 1, 2, 3, 5 and 6. Figure 2 shows a sample of gamelan notations.

Gamelan is manually constructed by hand. Constructors tune the instrument with their own sense based on

Manuscript received May 13, 2010, First revision Dec 31, 2010, Second revision Feb 27, 2011, Accepted Apr 11, 2011.

Yoyon K. Suprpto is with Electrical Engineering Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Email: yoyon-suprpto@ee.its.ac.id

Mochamad Hariadi is with Electrical Engineering Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Email: mochar@ee.its.ac.id

Mauridhi Hery Purnomo is with Electrical Engineering Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Email: hery@ee.its.ac.id

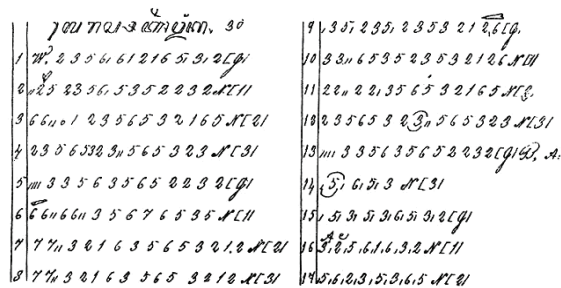


Fig. 2. Sample of gamelan notation belongs to Mangkunegaran Palace, 17th century.

TABLE I
SARON FUNDAMENTAL FREQUENCY FROM SEVERAL GAMELAN SETS.

(c)	Gamelan Fundamental Frequency (Hz)					
saron notation	set 1	set 2	set 3	set 4	Min	Max
1	528	528	504	539	504	539
2	610	610	574	610	574	610
3	703	703	688	703	688	703
5	797	792	792	799	792	799
6	915	922	909	926	909	926
1'	1056	1056	1008	1078	1008	1078
2'	1220	1220	1148	1220	1148	1220

experience. As a result, fluctuation of frequency inside the signal is not set correctly. The fundamental frequency of gamelan instrument is slightly different from one gamelan to the other gamelan. Table I shows saron fundamental frequencies from several gamelan sets. Each notation has varying frequency range.

Gamelan is played by striking the blades, so that sound is basically impulsive [5]. Figure 3 shows the spectrum of gamelan which varies very much due to the hardness and the style of stroke although it still has the same fundamental frequency.

Transcription is transforming an acoustic signal into a symbolic representation[6]. In other words, transcription of music is defined to be the act of listening to a music recording and writing down the musical notation for the sounds[7].

Many algorithms extracted an instrument sound from the music performance using STFT. Barbancho et.al.[8] used STFT and sliding windows to determine onset and time duration of the signal. Rodger J McNab et.al. [9] in their papers shifted slightly the threshold to determine fundamental frequency. Extraction was carried out base on amplitude and fundamental frequency. A median filter was applied to the detection function to define a dynamic threshold function, and a note onset was detected whenever. JP Bello et.al. [10] reported that for the synthesis process on each frame, they used the harmonic combs of estimated notes to isolate the relevant signal 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 fundamen-

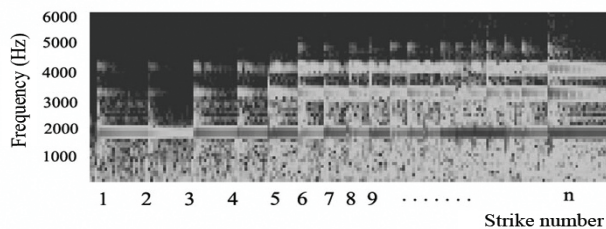


Fig. 3. The difference spectrum of the signal due to differences in hammer stroke strength.

tal frequencies. In normalization process, the short time Fourier transformation (STFT) was used by Barbancho et.al. [8] and Rodger J McNab et.al. [9] to obtain the fundamental frequency as well as time-frequency characteristics. The frequency and amplitude information were normalized according to the estimated fundamental frequency. Previous researchers [8], [9], [10] analyzed MIDI music or fabricated music instruments which are well tuned and have well uniform signal envelopes.

The target of this paper is to analyze the acoustic music such as gamelan. The complexity of the playing style causes conventional automatic transcription is hardly adopted. In this paper, the model of spectral density was built for generating simulated saron sounds. These sounds were used as sound reference on Adaptive cross correlation (ACC)[11] to generate estimated saron waveforms (extracted sounds). The automatic transcription is established by using the extracted sounds. Saron was chosen as the target group of gamelan extraction due to the use of saron notation as the basic notation for other instruments.

The remainder of the paper is organized as follows. Section II first briefly reviews the previous works, most related to our approach, which is short-time Fourier transform (STFT). Section III describes Adaptive cross correlation (ACC), an advanced cross-correlation algorithm that utilizes variable window's length and pitch shifting methods, that is used to reduce the errors associated with conventional music transcription. In this section we describe also the spectral density model which is constructed for generating simulated saron sounds. Section IV describes the performance evaluation. We show and investigate various types of gamelan playing, such as single synthetic gamelan, semi synthetic and gamelan ensemble. For performance evaluation, the conventional and our proposed methods are evaluated with our test data. Section V concludes the paper.

II. CONVENTIONAL METHODS

The musical transcription can be done with many methods. The previous work, which is mostly related to our approach is STFT. Each note signal has been extracted from gamelan ensemble recording.

Other researchers such as Barbancho et.al., Rodger J McNab et.al. used STFT and sliding windows to determine

the onset and time duration of the signal. We had to make some modifications on STFT for acoustic music sounds. Modified STFT was applied for comparison with our proposed method Adaptive Cross Correlation (ACC)[12] [13]. Both methods, the STFT and the ACC, are evaluated by using the same data, the sound of gamelan.

III. PROPOSED METHOD

In this paper, the model of spectral density was built for generating simulated saron sounds. These sounds were used as sound reference on Adaptive cross correlation (ACC) to generate estimated saron waveforms. Adaptive cross correlation (ACC) is an advanced cross-correlation algorithm that utilizes various window's length and pitch shifting method which are used to reduce the errors associated with conventional music transcription. The ACC algorithm is described in Fig. 4. The simulated saron sound was applied as a reference signal on the cross correlation process to form the magnitude of cross power spectrum density.

Original gamelan sounds, x , were yielded by striking the instrument with a hammer which was guided by the original gamelan note, o_r . Signal x was compared with the simulated saron sound, y , using the cross correlation to form the cross spectrum density [14] [15]. Estimated notes, e_s , were obtained from the cross spectrum density by the fundamental frequency of each musical note and were evaluated using note error rate, ner [16] [17]. Ner was generated by the note insertion, note substitution and note deletion. Simulated saron sounds were produced by pitch-shifting method based on phase-vocoder theory [18]. Figure 5 shows three sides of tone database. The left-hand side is the real database obtained from observation. It leaves us with a database of a few detected notes and many

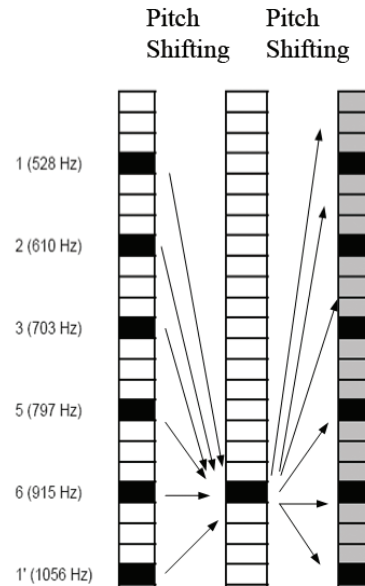


Fig. 5. Gaps in the database were filled by pitch-shifting the estimated notes.

gaps. The middle-hand side illustrates the pitch shifting process where pre-recorded sound was brought to Saron6 frequency as the reference and take the average of the spectrum of all shifted pre-recorded sounds. At the end, the average spectrum was shifted back to all possible saron frequencies to fill the gaps in the database on the right side.

Simulated saron sounds were organized in the database according to their fundamental frequency f_0 . The resulting database was incomplete, i.e. did not contain waveforms for all notes in the f_0 range. To do pitch shifting, we constructed a saron time-frequency model.

A. Time-Frequency Model based on the Spectral Density

To analyze gamelan performance, 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. The sounds are converted to time-frequency domain using STFT. 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 b , and its fundamental frequency estimation. We evaluate how to convert time domain signal $x(n)$ to frequency domain $X(f)$ using STFT which is described in Eq.(1)[8] [9],

$$\begin{aligned} STFT(x(n)) &\equiv X(t, f) \\ &= \sum_{n=0}^{N-1} x(n)w(n-t)e^{-i2\pi f/f_s \frac{n}{N}} \end{aligned} \quad (1)$$

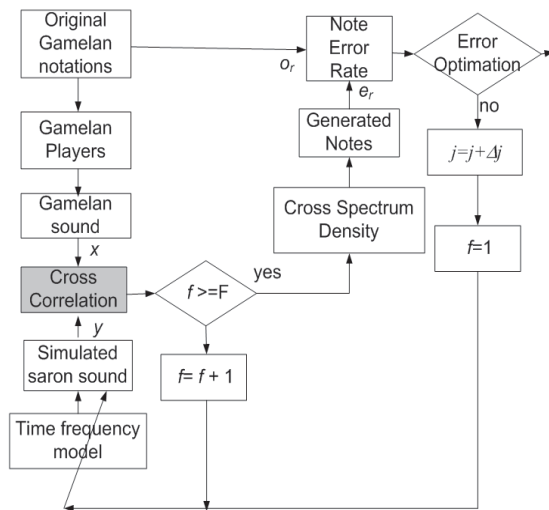


Fig. 4. Sound Extraction Based on Spectral Density Model using Adaptive Cross Correlation.

where f is frequency, f_s is sampling frequency, t is time index, w is window, n is sampling index, N is total sampling. Due to gamelan characteristics, each power density spectrum from the gamelan notes may vary. Estimated fundamental frequency was obtained by the maximum argument of the absolute value of the spectrum as described in Eq.(2). Each pre-recorded sound has instrument name, note number c , pre-recorded sound number b and estimated fundamental frequency f_0 ,

$$f_{0b}(t) = \arg \max_{f=\min(f_0(c_b))}^{\max(f_0(c_b))} (|X_b(t, f)|) + \min(c_b) \quad (2)$$

and magnitude of fundamental frequency, $X(f_{0b}(t))$, can be described at Eq.(3),

$$X(f_{0b}(t)) = \max_{f=\min(f_0(c_b))}^{\max(f_0(c_b))} (|X_b(t, f)|) \quad (3)$$

where f_{0b} is the fundamental frequency of pre-recorded sound b , c is note number, b is pre-recorded sound number. See Table I. Maximum argument is the set of values of f for which $X_b(t, f)$ has the largest value. f is located between the minimum $\min(f_0(c_b))$ and maximum $\max(f_0(c_b))$ value of fundamental frequency in each notation c . Normalized power density, X_{Nb} , is obtained by absolute $X_b(t, f)$ divided by $X(f_{0b}(t))$ which is described in Eq.(4),

$$X_{Nb}(t, f) = \frac{|X_b(t, f)|}{X(f_{0b}(t))} \quad (4)$$

In order to build the time frequency model, we used 450 pre-recorded sounds of saron instrument which consisted of several combinations of hammer stroke strength, and several combinations of hammer stroke areas. A standard tone was selected for the pre-recorded sounds Saron6, the sixth note of saron instrument. It was chosen as the standard tone for normalization [5]. In our previous research [19], we evaluated a fundamental frequency relationship among gamelan notes. The slendro gamelan scale used in the Javanese gamelan has five equally-tempered pitches. The model is made by shifting all fundamental frequencies of pre-recorded sounds to the Saron6 fundamental frequency [17]. The pitch shifting Δf_{06} was calculated using Eq.(5), where f_{0b} is the fundamental frequency of a pre-recorded signal and f_{06} is the fundamental frequency of ideal Saron6 as the reference tone. Based on the pitch shifting Δf_b , all frequency components were shifted by same Δf_b and the shifted signal should be added by Δf_b zero paddings. Note: ideal Saron6 fundamental frequency f_{06} was obtained from the average of the sixth notation fundamental frequency of saron instrument from several gamelan sets,

$$\Delta f_b(t) = f_{0b}(t) - f_{06}(t) \quad (5)$$

where b is pre-recorded sound number, f_{0b} is fundamental frequency of pre-recorded sound b and f_{06} is the

fundamental frequency of Saron6. The model was made by shifting all fundamental frequencies of pre-recorded sounds to the Saron6 fundamental frequency [19]. The non-harmonic components are shifted by Δf_b which is shown in Eq.(6),

$$\hat{X}_{Nb}(t, f) = X_{Nb}(t, f + \Delta f_b(t)) \quad (6)$$

where $\hat{X}_{Nb}(t, f)$ is normalized shifted magnitude of pre-recorded b . The Pitch shifting algorithm is shown in Algorithm 1.

Algorithm 1: Pitch shifting.

- 1) $b \leftarrow 1$; b is pre-recorded sound index
 - 2) f_{0b} is fundamental frequency of b
 - 3) $f \leftarrow 1$; f is frequency index
 - 4) Shifted the power spectrum density by Δf_b using Eq.(6)
 - 5) $f \leftarrow f + 1$
 - 6) repeat 4) until $f \leq F$
 - 7) $b \leftarrow b + 1$; next pre-recorded sound
-

The time frequency model $A(k, f)$ was determined by averaging the power density $\hat{X}_{Nb}(k, f)$ for all of pre-recorded signals as shown in Eq. (7). In order to construct the time frequency model $A(k, f)$, we need to determine the average power density spectrum of each frequency index,

$$A(t, f) = \frac{\sum_{b=1}^S \hat{X}_{Nb}(t, f)}{S} \quad (7)$$

where S is total pre-recorded sounds.

The time frequency model is a discrete time frequency model Eq.(11). The time frequency model can be seen at Fig.6. The model is interpolated by using exponential curve fitting in Eq.(8) for filling the time interval gaps. Two parameters were added, α as amplitude and β as exponential parameters. Eq.(17),

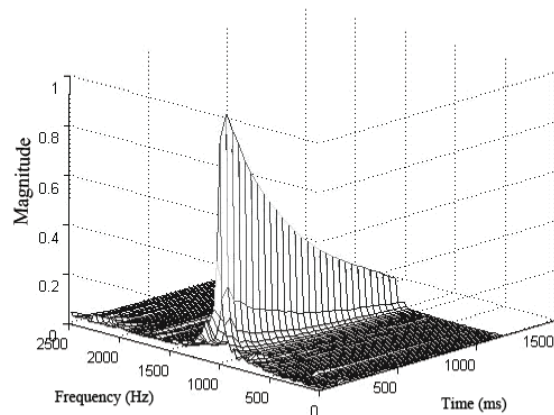


Fig. 6. Saron Time-frequency model.

$$\begin{aligned}
 A(t, f) &= \alpha(f)e^{\beta(f)t} \\
 \log(A(t, f)) &= \log(\alpha(f)e^{\beta(f)t}) \\
 &= \log(\alpha(f)) + \log(e^{\beta(f)t}) \\
 &= \log(\alpha(f)) + \beta(f)t
 \end{aligned} \tag{8}$$

If

$$\begin{aligned}
 A'(t, f) &= \log(A(t, f)) \\
 \alpha'(f) &= \log(\alpha(f)) \\
 A'(t, f) &= \alpha'(f) + \beta(f)t
 \end{aligned} \tag{9}$$

Linear regression coefficient [20] shows that estimate parameter $\hat{\alpha}(f)$ and $\hat{\beta}(f)$ is calculated using Eq.(10) and Eq.(11),

$$\begin{aligned}
 \hat{\beta}(f) &= \frac{K \sum_{k=1}^K k A(k, f)}{K \sum_{k=1}^K k^2 - (\sum_{k=1}^K k)^2} \\
 &- \frac{\sum_{k=1}^K k \sum_{k=1}^K A(k, f)}{K \sum_{k=1}^K k^2 - (\sum_{k=1}^K k)^2}
 \end{aligned} \tag{10}$$

$$\hat{\alpha}'(f) = \frac{\sum_{k=1}^K A(k, f) - \hat{\beta}(f) \sum_{k=1}^K k}{K} \tag{11}$$

from Eq.(11)

$$\alpha'(f) = \log(\alpha(f))$$

$$\hat{\alpha}(f) = e^{\alpha'(f)} \tag{12}$$

Based on the time frequency model at Fig. 6, each frequency has its envelope $A(k, f)$,

$$\hat{A}(t, f) = \hat{\alpha}(f)e^{\hat{\beta}(f)t} \tag{13}$$

where $\hat{A}(t, f)$ is estimated of envelope time frequency model.

Table II shows the value of $\alpha(f)$ and $\beta(f)$ for estimated of envelope time frequency model $\hat{A}(t, f)$. The refined time frequency model can be seen at Fig.7.

The simulated saron sounds were synthesized saron sounds which were organized in the database according to their f_0 . The resulting database is expanded by generating previously unavailable synthetic sounds using time-frequency model. The completeness of the database varies depending on the sound and on the parameter set the modified sounds are generated using Eq.(14),

$$\hat{x}(t, f_0) = \sum_{\Delta f=f+1}^F \cos(2\pi(f_0 + \Delta f)t/f_s) \hat{A}(t, f_0 + \Delta f) \tag{14}$$

We generate simulated saron sounds from $f_0 = 1, 2, 3, \dots, F$ Hz.

TABLE II
PARAMETERS FOR ESTIMATION ENVELOPE TIME FREQUENCY MODEL
 $\hat{A}(k, f)$.

Frequency (Hz)	α	β
:	:	:
f_0-4	0,2115	-0,5491
f_0-3	0,2766	-0,5472
f_0-2	0,4003	-0,5345
f_0-1	0,7422	-0,5150
f_0	1,1012	-0,5233
f_0+1	0,8715	-0,5775
f_0+2	0,5161	-0,6018
f_0+3	0,3381	-0,5935
f_0+4	0,2610	-0,5979
:	:	:

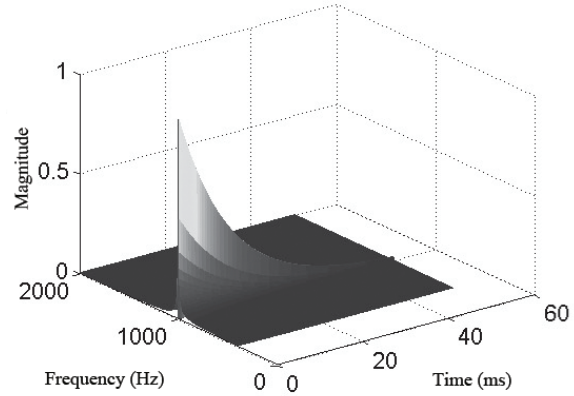


Fig. 7. Refined saron time-frequency model is interpolated by exponential curve fitting.

B. Saron sound extraction for Automatic Transcription using Template

To transcribe the gamelan music, saron sound waveforms were extracted from gamelan ensemble using adaptive cross-correlation is describe in Eq.(15). Simulated saron sounds were used as the template for cross-correlation to extract the saron sound. Figure 8 illustrates the estimation process of saron note generating. Original gamelan waveform is generated by striking gamelan instrument using original gamelan note.

$$r(t, n, f) = \frac{1}{J} \sum_{m=0}^{J-1} x(t, m+n) \hat{x}(m, f) \tag{15}$$

where n is lag, J is the window's length of the x and $\hat{x}$. If f is frequency scanning from 1 to F Hz, $r(n, f)$ becomes the magnitude of cross power spectral density of observed sound $x(k)$. The estimated saron waveforms are extracted from gamelan ensemble using range fundamental frequency of each saron note, c ,

$$p(t, c) = \max_{f=\min(f_0(c))}^{\max(f_0(c))} (|r(t, n, f)|) \tag{16}$$

where $c = 1, 2, 3, 5$ and 6 are gamelan notes, p is estimate of saron waveform based on the template. It is necessary

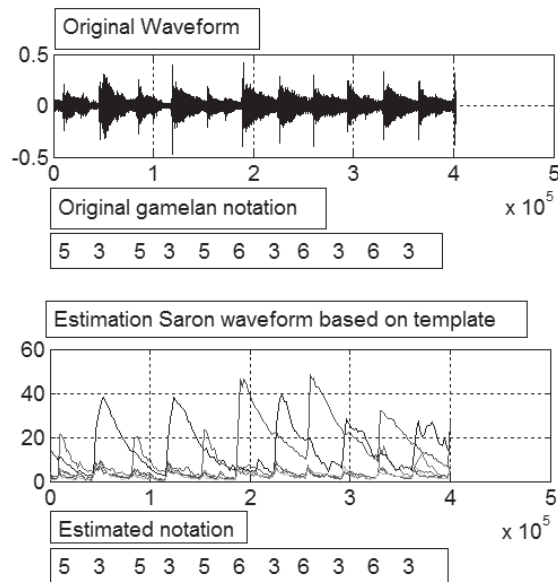


Fig. 8. Estimated saron note generating.

to eliminate the noise using threshold. In gamelan performance, each note may have different magnitude, so each note may have its own threshold. The simplest way to segment notes is to set a threshold 20%. These values were achieved through experiment. The candidate of the notes are obtained by determining the peak of each sound. Each note candidate has its note number, the magnitude of cross power density and the onset.

All note candidates were sorted by the onset. More than one note candidate, Saron1 and Saron1', were evaluated at the same time interval, 10 ms areas, to determine the note. The real note was determined by the highest magnitude among all sorted note candidates. Unfortunately, gamelan had a lot of instrument groups. Besides saron group, gamelan had fifteen groups. Both, saron and bonang, had the same fundamental frequency but they had different

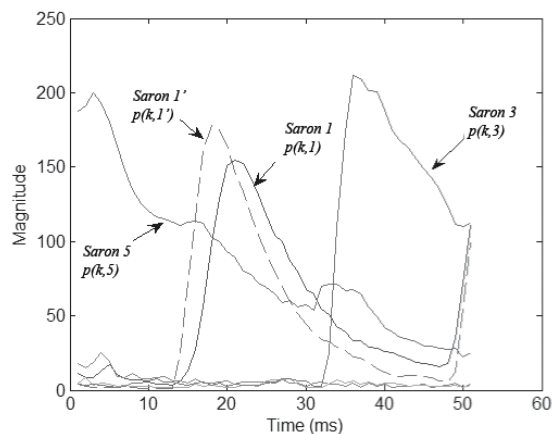
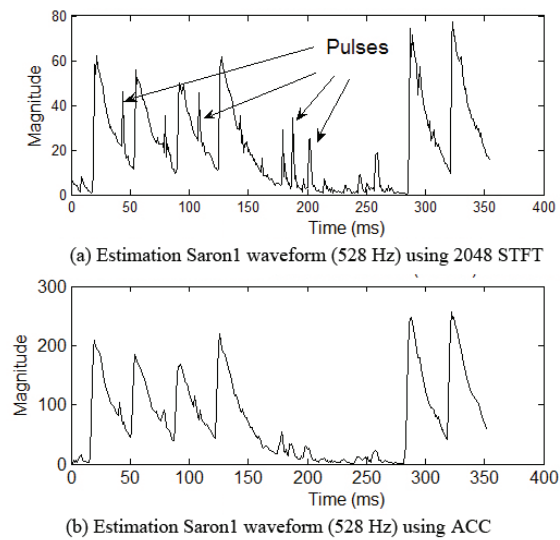

 Fig. 9. Estimated Saron waveforms for $c= 1, 3, 5$ and $1'$.


Fig. 10. Estimated saron waveform influenced by bonang waveform.

timbre so bonang sounds influence the saron sounds. They were detected as pulses which is shown in Fig.10. Pulses were generated from other instrument like bonang. To eliminate the pulse, the length of the sound J in Eq.(15) was varied. Adaptive cross-correlation is applied by varying the frequency f and the window's length J .

IV. PERFORMANCE EVALUATION

A. The Gamelan Songs for Testing

We generated three types of gamelan sound for testing:

- 1) Full synthetic. The gamelan sounds were generated by the computer. The ensemble were played by using computer with gamelan note direction.
- 2) Semi synthetic. Each gamelan note was recorded and the ensemble were played using computer with gamelan note direction.
- 3) Full acoustic. Gamelan ensemble was played by the players and was recorded. It was recordings of gamelan ensemble performances which was consisted of nine simultaneously played instruments. It was 90 seconds of duration and it contained 129 original notes.

B. Automatic Transcription

In order to show the effectiveness of template matching for automatic transcription, various types of playing, such as single synthetic gamelan, mixture of three synthetic gamelan, single semi synthetic, mixture of three semi synthetic and gamelan ensemble were investigated. As the basic automatic transcription, the cross-correlation method was used. To evaluate the estimated generated notes, we used the Note Error Rate [16] [17]. Recognition of error rates were often reported at Eq.(17),

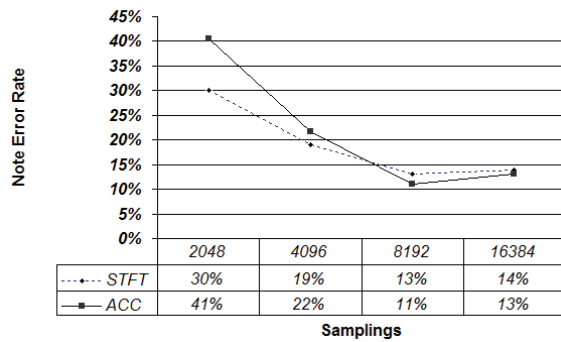


Fig. 11. Note error rate ner against various windows lengths for STFT and ACC methods.

TABLE III

PERFORMANCE OF SARON EXTRACTION FOR GAMELAN TRANSCRIPTION BY CONVENTIONAL METHOD STFT AND ADAPTIVE CROSS-CORRELATION (ACC) WITH MATCHING TEMPLATE.

Test Type	Total notations	Total instruments	8192 STFT	ACC
Full synthetic	30	3	0%	0%
Semi synthetic	30	1	4%	3%
Semi synthetic	30	3	5%	4%
Full acoustic	30	2	6%	4%
Full acoustic	30	2	8%	6%
Full acoustic	129	9	18%	16%

$$ner = \frac{\text{deletion} + \text{insertion} + \text{substitution}}{\text{totaltruesentence}} \quad (17)$$

To evaluate sound extraction using STFT [8], [9], [10], the sampling frequency was 48000 Hz. The fastest gamelan beat time was 250 ms or 12000 samplings. In STFT, we had to decide how frequent it was to perform DFT computations on the sound. For evaluating the performance, we varied window's length. The result is shown in Fig.11. The smallest *ner* occurred at 8192 window's length. The overall results, 8192 STFT was compared with our proposed method ACC. Table III shows the results as the ratio of *ner*.

The experiment results showed that instrument numbers did not affect the performance of instrument extraction. Two instruments, saron and bonang, were played simultaneously, the performance was not always better than five instruments. Saron and bonang have the same f_0 , so bonang influences the saron sounds.

V. CONCLUSION

In this study the Adaptive Cross Correlation (ACC) method proposed for automatic notation of Saron instrument. The performance test demonstrates the proposed method provided 2 - 4 % improvement for analyze the acoustic music such as gamelan comparing to conventional method such as STFT.

The complexity of the playing style causes conventional automatic transcription is hardly adopted. These results

show the effectiveness of template matching for picked up specified instrument and for automatic transcription.

REFERENCES

- [1] Michael A. Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney, "Content-Based Music Information Retrieval: Current Directions and Future Challenges," *Proceedings of the IEEE*, Vol. 96, No. 4, April 2008
- [2] Olmo Cornelis, Micheline Lesaffre, Dirk Moelants, Marc Leman, "Access to ethnic music: Advances and perspectives in content-based music information retrieval," *Signal Processing 90 Elsevier, Amsterdam*, pp. 1008-1031, 2010
- [3] Sutton, Anderson, R., "Central Javanese gamelan music: Dynamics of a steady state," *Northern Illinois University in DeKalb, IL*, pp. 278-288, 1993.
- [4] Tamagawa, Kiyoshi, "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*, 1998.
- [5] Sumarsam, "Cultural Interaction and Musical Development in Central Java," *The University of Chicago Press, ISBN 0-226-78011-2*, 1992- 1995.
- [6] Klapuri, A. and Davy, M., "Signal Processing Methods for Music Transcription," *Springer-Verlag*, New York, 2006.
- [7] Eric Scheirer, "Extracting expressive performance information from recorded music," *Master's thesis, MIT*, 1995.
- [8] Barbancho, A. Jurado, L.J. Tardo, "Transcription of piano recordings," *Applied Acoustic* 65, pp. 1261- 1287, 2004.
- [9] Rodger J. McNab, Lloyd A. Smith and Ian H. Witten, "Signal Processing for Melody Transcription," *Proceedings of the 19th Australian Computer Science Conference, Melbourne, Australia, January 31-February 2 1996*.
- [10] J. P. Bello, L. Daudet and M. B. Sandler, "Automatic piano transcription using frequency and time-domain information," *IEEE Transaction on Audio, Speech and Language Processing*, vol. 14 no 6, pp. 2242-2251, 2006.
- [11] M. Arezki, A. Benallal, P. Meyrueis and D. Berkani, "A New Algorithm with Low Complexity for Adaptive Filtering," *Engineering Letters, IAENG*, 18:3, EL 18 3 01, Volume 18, Issue 3, 2010.
- [12] Farshad Arvin, Shyamala Doraisamy, "Real-Time Pitch Extraction of Acoustical Signals Using Windowing Approach," *Australian Journal of Basic and Applied Sciences*, vol. 3(4), pp. 3557-3563, 2009.
- [13] Bokyoung Sung, Jungsoo Kim, Jinman Kwun, Junhyung Park, Jihye Ryeo, and Ilju Ko, "Practical Method for Digital Music Matching Robust to Various Sound Qualities," *World Academy of Science, Engineering and Technology*, 2009.
- [14] Willam J. Pielemeier, Gregory H.W. and Mary H. Simoni, "Time-Frequency Analysis of Musical Signals," *Proceedings of The IEEE*, vol.84, No.9, pp. 1216-1230, 1996.
- [15] David Havelock, Sonoko Kuwano, Michael Vorlander, "Handbook of Signal Processing in Acoustics," *Springer New York*, 2008.
- [16] Christopher Raphael, "Automatic Transcription of Piano Music," in *Proc. ISMIR*, pp.15-19, 2002
- [17] Anssi P. Klapuri, "Automatic Transcription of Music," *Proceedings of the Stockholm Music Acoustics Conference, Sweden, August 6-9, 2003*.
- [18] Mark Dolson, "The Phase Vocoder: A Tutorial," *Computer Music Journal*, vol. 10 No. 4, pp. 14-27, Winter, 1986.
- [19] Yoyon K Suprpto, T Usagawa, Mochamad Hariadi, "Time frequency modelling of gamelan instrument based on spectral density for automatic notation," *the Third International Student Conference on Advanced Science and Technology, Seoul, Korea*, pp. 15-19, 2009.
- [20] Jaan Kiusalaas, "Numerical Method in Engineering with Mathlab," *Cambridge University Press, New York*, 2005.



Yoyon K Suprpto received the bachelor degree in Electrical Engineering from Institut Teknologi Bandung, Bandung, Indonesia in 1977. He received his Master of Science Computer Science from The University of Missouri, Columbia, Missouri, USA in 1981. He joined Electrical Engineering Department in Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia since 1977. His current interests are Data Mining, Sound Signal Processing and Traditional Music. He is currently pursuing the

Ph.D. degree at Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia since 2007. He is a student member of IEICE. He is a student member of IEEE. He is a member of IAENG.



Mochamad Hariadi received the B.E. degree in Electrical Engineering Department of Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 1995. He received both M.E. and Ph. D. degrees in Graduate School of Information Science Tohoku University Japan, in 2003 and 2006 respectively. Currently, he is the staff of Electrical Engineering Department of Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. He is the project leader in joint research with PREDICT JICA project

Japan and WINDS project Japan. His research interest is in Video and Image Processing, Data Mining and Intelligent System. He is a member of IEEE, and member of IEICE.



Mauridhi Hery Purnomo received the bachelor degree from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia in 1985. He received his M.S., and Ph.D degrees from Osaka City University, Osaka, Japan in 1995, and 1997, respectively. He has joined ITS in 1985 and has been a Professor since 2004. His current interests include intelligent system applications an electric power systems operation, control and management. He is a Member of IEEE.