Journal of Theoretical and Applied Information Technology
15th October 2011. Vol. 32 No.1

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

SARON MUSIC TRANSCRIPTION USING LPF-CROSS CORRELATION

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

[1]Senior Lecturer, Jurusan Teknik Elektro, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

[2]Doctorate Student in Jurusan Teknik Elektro, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

[3]Doctorate Student in Jurusan Sistem Informasi, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

E-mail: yoyonsuprapto@gmail.com, diah_basuki@yahoo.com, atjahyanto@gmail.com

ABSTRACT

Nowadays, mining of traditional music attracts people's interests in several aspects since the importance of archiving national heritage is emphasized. Unfortunately, there have been very few researches which analyzed traditional music instruments and their performances. Gamelan, one of Indonesia's traditional music instruments, has uniqueness in terms of its hand-made construction which in turn induces variations in frequency, amplitude, and resonance. These conditions have detained the exploration and development of gamelan music itself.

This research initiates gamelan sound extraction for music transcription as part of traditional music analysis. Spectral density model was constructed to extract the sound of an instrument from the gamelan ensemble performance by using LPF cross correlation (LPF-CC). The extracted sound was analyzed to generate notations. The experiment demonstrates 5-16% note error rate.

Keywords : *Saron Time And Frequency Model, LPF Cross-Correlation, Saron Extraction, Music Ranscription.*

1. INTRODUCTION

There are some differences between western music and eastern music. While western music perceives that good music is composed of stable tones, regulated frequency, and fixed amplitude; the eastern music such as gamelan has freely imposed tones with respect to its resonance, tone color, frequency, and amplitude [1]. Therefore traditional music is more difficult to analyze and its development is much lagged to that of modern music[2].

At the other side, gamelan is one of Indonesia's traditional music which its repetitive playing pattern has been increasingly accepted by international composers [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; in-depth research of gamelan sound is needed.

Gamelan consists of about fifteen groups of instruments. Their groups are saron, kenong, kempul, kendang, bonang, etc. Figure 1 shows saron group. Figure 2 shows another group called bonang. Both of these instruments, saron and bonang, have exactly the same fundamental frequency but they own different timbre. This research focuses on the extraction of saron sound. Each gamelan instrument is comprised of only an octave, which are 1, 2, 3, 5, and 6 in pentatonic or hexatonic tones [4]. Saron is constructed from several metal blades, where each blade represents a notation. Figure 3 shows a sample of gamelan notation.

Fig.1. The Saron family in a gamelan set.

Journal of Theoretical and Applied Information Technology
15th October 2011. Vol. 32 No.1

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

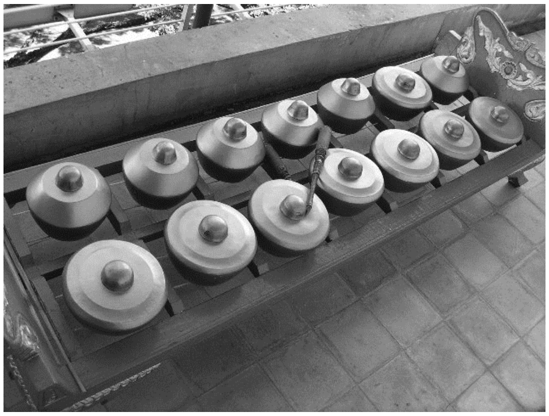


Fig 2 *Bonang* group in a gamelan set.

Lancaran Manyar Sewu

Buka : . 1 . 6 . 1 . 6 . 5 . (3)

. 5 . 3^N . 5^P . 3^N . 5^P . 3^N . 6^P . [5]
. 6 . 5^N . 6^P . 5^N . 6^P . 5^N . 3^P . [2]
. 3 . 2^N . 3^P . 2^N . 3^P . 2^N . 1^P . [6]
. 1 . 6^N . 1^P . 6^N . 1^P . 6^N . 5^P . (3)

Fig.3 Sample of gamelan notation

Figure 3 shows the notation of Manyar sewu song. Gamelan music notation is not the same as the western musical notation. Gamelan musical notation is very simple. It consists of only 1, 2, 3, 4. 5, 6 and 7. This notation used by saron family such as saron, demung and peking by strking each written note once. Notes with superscript *N* (X^N) are stroke by kenong instrument while those with superscript *P* (X^P) are stroke by kempul instrument. Notes inside brackets (*X*) must be played by Gong Ageng instrument and those inside square brackets [*X*] must be played by Gong Suwuk instrument. So the notation of saron leads the notation for other instruments as well.
Gamelan is manually constructed by hand. Constructors tune the instruments with their own sense, based on experience. As a result, fluctuation of frequency inside the signal is not set correctly. Saron is played by striking the blade, so that its sound is basically impulsive [4]. The fundamental frequency of one *gamelan* performance could be slightly different from one *gamelan* to the other *gamelan*. Table 1 shows *saron* fundamental frequencies from several gamelan sets. Each note had varying frequency range. The other octave is belonged by the other instrument.

Table 1 Saron fundamental frequency from several gamelan sets.

Gamelan Notation	Gamelan Fundamental Frequency (Hz)					
	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	926	926	915	926

Music transcription of audio data is the process of taking a sequence of digital data corresponding to the sound waveform and extracting from it the symbolic information related to the high-level musical structures that might be seen on a score [5]. For music performance which comprises several instruments, like the one in an orchestra, the sound of specific instrument needs to be extracted first before generating the notations, because each instrument may be guided by different notations. Many algorithms have been applied to extract an instrument sound. Most of them use Short-time Fourier Transform (STFT). Barbancho et. al. used STFT and sliding windows to determine the onset and time duration of a notation's signal [6][7] shifted slightly the threshold to determine fundamental frequency [8]. Extraction was carried out based on fundamental frequency and its power density. J. P. Bello et. al. reported in their paper that for synthesis process, they used 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 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 Saron time-frequency characteristics. Previous researchers (Barbancho, McNab, Bello) mostly analyzed MIDI music that resulted from fabricated music instruments, which is well tuned and has uniform notation signal envelopes.
The target of this paper is to analyze the acoustic music such as gamelan music. The complexity of the playing style as well as the hand-made construction of gamelan causes the

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

conventional music transcription to be hardly adopted. In this paper, the model of spectral density was build to produce estimated saron sounds. These sounds were used as reference in extraction process using LPF cross correlation (LPF-CC), to generate estimated saron waveforms. The music transcription was established based on the extracted sounds. Saron was chosen as the target group of gamelan sound extraction due to the use of saron notation as the basic notation for the other instruments.

2. CONVENTIONAL METHOD

Previous works in music transcription mostly used STFT to extract the sound of particular instrument in MIDI music performance. We modified the STFT to apply it for acoustic music performance. This modified STFT was used in sound extraction process as a comparison with our proposed method, LPF-CC. Finally, both methods were evaluated by using the same data, gamelan music performance.
Music transcription using STFT: Other researchers applied non-overlapped STFT to analyze MIDI music. We modified STFT into an overlapped one because we tackle an acoustic music where there may occur variations in the sound signals, which might be influenced by external factors such as hammer stroke strengths and hammer stroke styles. Overlapped STFT apparently eliminated the wild peaks of magnitude in notation signal. This modification then only considered the fundamental frequency of the signal, while the harmonic and non-harmonic components were not addressed. Firstly we determined the window length and the hop length.

3. PROPOSED METHOD

LPF-CC is an advanced cross-correlation algorithm that utilizes various window lengths and is a pitch-shifting method which is used to reduce the errors associated with conventional music transcription. In this paper the spectral density model of saron sound was built and then compared to the real recorded gamelan performance by using LPF-CC, in order to extract the specific saron sound among the other instrument's sound. The LPF-CC algorithm is described in Figure 4.

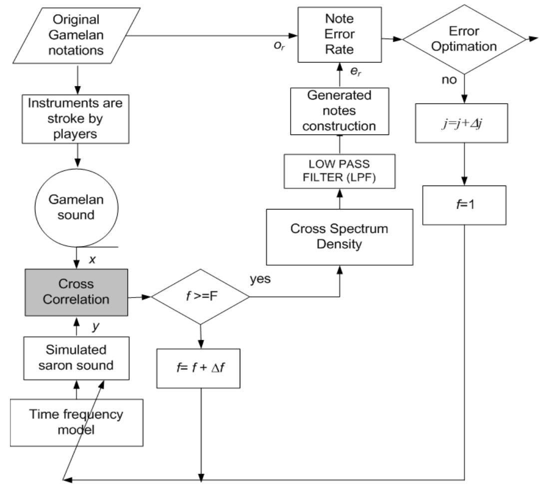


Fig. 4 Sound Extraction Based on Spectral Density Model using LPF Cross Correlation.

The actual gamelan sounds *x* were yielded by striking the instrument with a hammer which was guided by the original gamelan notations or. Signal *x* was then compared to the reference signal *y* using cross correlation to form the cross spectrum density [9][10]. Estimated notes were obtained from fundamental frequency of each musical note and were evaluated using note error rate (*ner*) [11][12]. *Ner* was resulted from note insertion, note substitution, and note deletion. Gamelan simulated sounds were produced by pitch-shifting method based on phase-vocoder theory [13]. Figure 5 describes simulated saron sound building using pitch-shifted.

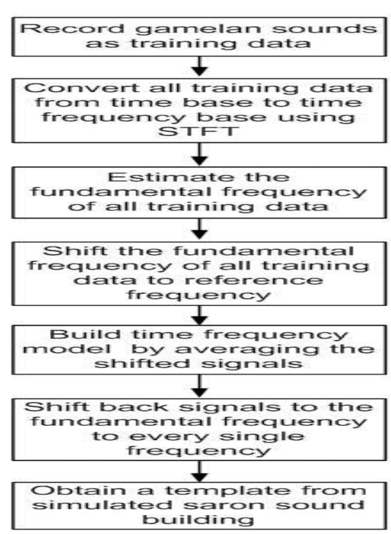


Fig. 5 The algorithm of simulated saron sound building using pitch-shifted

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

2.1 Saron 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 Saron 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.

Before we discuss the STFT, we evaluate how to convert time domain signal $x(n)$ to frequency domain $X(f)$ using discrete Fourier transform (DFT) which is described in Eq.(1),

$$X(f) = F[x(n)] = \sum_{n=0}^{N-1} x_n e^{-i2\pi f / f_s \frac{n}{N}} \quad (1)$$

where f is frequency, f_s is sampling frequency, n is time index, N is total sampling.

Due to gamelan characteristics, the estimated fundamental frequency of each of gamelan notes may vary. See table 1. The real fundamental frequency of each pre-recorded sound is determined by the maximum argument of the absolute value of the spectrum as described in Eq.(2),

$$f_{0b} = \underset{f=\min(f_0(c))}{\overset{\max(f_0(c))}{\arg\max}} (|X(f,b)|) + \min(c) \quad (2)$$

where f_{0b} is the fundamental frequency of pre-recorded sound, c is blade number which represents a note number, b is pre-recorded sound number. Maximum argument is the set of values of f for which $X(f,b)$ has the largest value. f is located between the minimum $min(f_0(c))$ and maximum $max(f_0(c))$ value of fundamental frequency in each notation c.

Normalized power density, X_{Nb} , can be obtained by absolute $X_b(f)$ divided by $max(|Xb(f_{0b})|)$ which is described in Eq.(3),

$$X_{Nb}(f) = \frac{|X_b(f)|}{\max(|X(f_{0b})|)} \quad (3)$$

where N is total pre-recorded sound number.

In order to build the Saron time frequency model, we applied 450 pre-recorded sounds of saron instrument which consisted of several combinations of hammer stroke strength and hammer stroke areas. The pre recorded Saron6 sound, the sixth note of saron instrument, was selected to be the standard tone for normalization [4].

In our previous research [14], we evaluated the relationship of fundamental frequencies among gamelan notes. The slendro gamelan scale, used in the Javanese gamelan, has five equally-tempered pitches. A musical equal temperament is most simply described as a mathematical series at Eq. (4),

$$F_m = 2^{m/k} F_r \quad (4)$$

where F is the fundamental frequency of a musical tone, k is a positive integer constant, m is an integer variable. k is equal to 5 for slendro gamelan scale. The cent is a logarithmic unit of measure used for musical intervals. 1200 cents are equal to one octave — a frequency ratio of 2:1. If we know the frequencies Fm and Fr of two notes, the number of cents measuring the interval from Fm to Fr may be calculated by the following formula Eq. (5),

$$\Delta c = 1200 \log_2 \left(\frac{F_m}{F_r}\right) \quad (5)$$

Likewise, if we know a note F_r and the number Δc of cents in the interval from F_m to F_r, then F_m may be calculated Eq.(6),

$$F_m = 2^{\Delta c / 1200} F_r \quad (6)$$

The Saron time-frequency model is made based on the pitch-shifting algorithm described in Fig. 5. Firstly, all fundamental frequencies of pre-recorded sounds are shifted to the Saron6 fundamental frequency. The pitch shifting Δf_{06} could be calculated using Eq.(7), where f_{0b} is the fundamental frequency of a pre-recorded sound and f_{06} is the fundamental frequency of ideal Saron6. Based on the pitch shift Δf_{0s}, all frequency components were shifted by same Δf_0 and the shifted signal should be added by Δf_b zero padding. Note that the fundamental

Journal of Theoretical and Applied Information Technology
15th October 2011. Vol. 32 No.1

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

frequency of ideal Saron6 f_{06} was obtained from the average of the sixth notation's fundamental frequency of saron instrument from several gamelan sets.

$$\Delta f_b = f_{0b} - f_{06} \qquad (7)$$

where b is pre-recorded sound number and f_{0b} is fundamental frequency of pre-recorded sound b and f_{06} is the fundamental frequency of Saron6. Figure 6 illustrates the spectral of pre-recorded sound shifted to Saron6.

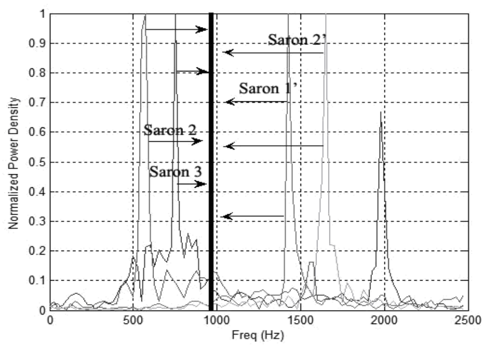


Fig.6 Pre-recorded pitch shifting to the referenced sound "Saron6"

The non-harmonic and harmonic components of pre-recorded sound spectrum are shifted by Δf which is shown in Eq.(8),

$$\hat{X}_{Nb}(f) = X_{Nb}(f + \Delta f) \qquad (8)$$

where $\hat{X}_{Nb}(f)$ is normalized shifted magnitude of pre-recorded b.

The Saron time-frequency model can be generated by adding k as time index parameter, which is equivalent with the fixed window's length. The Saron time frequency model $A(k,f)$ is determined by averaging the power density $\hat{X}_{Nb}(f)$ of each frequency index of all pre-recorded sounds as shown in Eq. (5).

$$A(k, f) = \frac{\sum_{b=1}^{S} \hat{X}_{Nb}(f)}{S} \qquad (9)$$

where S is total number of pre-recorded sounds.

Figure 7 shows the Saron time-frequency model of saron.

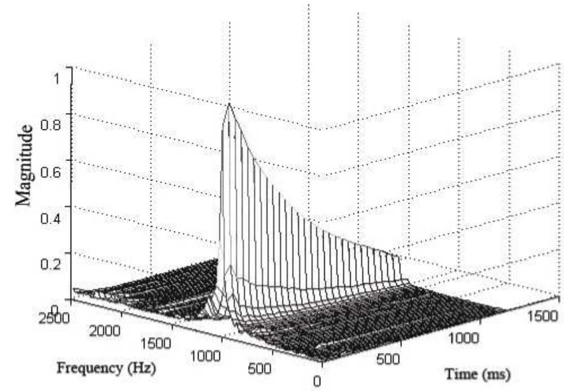


Fig. 7 Saron Time-frequency model.

Saron time-frequency model is used to construct the simulated saron sound. Each model produces a simulated saron sound. The other models are obtained by shifting the fundamental frequency to a specific frequency. The total number of models is F unit. This means that there will have F saron simulated sounds which each has a fundamental frequency of 1 to F Hz. All simulated saron sounds will form a database in the process of cross correlation. The simulated saron sound, template, can be generated using Eq.(10).

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

f_0= 1,2,3 … F Hz (10)

2.2 Saron sound extraction for music transcription using template

The simulated saron sound $\hat{x}(k, f_0)$ was applied as a reference signal on the cross-correlation process to form the cross power spectrum which would generate saron estimation waveforms. Figure 8 illustrates the estimation process of saron transcription. Actual gamelan waveform is generated by striking gamelan instrument using original gamelan note. Estimated saron waveform is extracted by comparing the template to the actual gamelan sound waveform using LPF cross-correlation, which can be calculated using Eq.(11).

Journal of Theoretical and Applied Information Technology
15th October 2011. Vol. 32 No.1

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

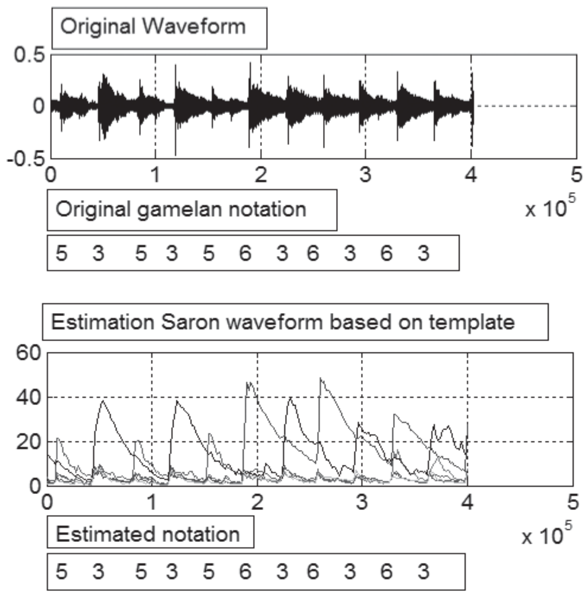


Fig.8 Estimated saron note generating

$$r(k,n,f)=\frac{1}{J}\sum_{m=0}^{J-1}x(k,m+n)\hat{x}(m,f) \quad (11)$$

where n is lag, k is sampling index and J is the length of the x and $\hat{x}$. If f is frequency scanning from 1 to F Hz, $r(n,f)$ becomes cross spectral density of observed signal x. The cross correlation results in separated saron sound among the other instrument sounds, which then be recognized as estimated saron waveforms. Equation (12) shows that by setting the value of f to be in particular interval $min(f_0(c)) \le f \le max(f_0(c))$ we obtain the occurrences of estimated saron waveforms which are note candidates across time. The next step is to translate the waveforms into candidate notations.

$$p(k,c)=\max_{f=\min(c)}^{\max(c)}\left(\left|r(k,n,f)\right|\right) \quad (12)$$

where c = 1, 2, 3, 5, 6 is each gamelan note, and p are estimated *saron* waveforms. It is necessary to eliminate the noises occur in notation waveform using threshold. In real gamelan performance, each note may have different magnitudes due to the difference of stroke strength during playing. That is why different thresholds can be applied to separate the real note-onset from noises. The simplest way to segment notes is to set a threshold to be 20 % of the maximum magnitude. This value is arrived through experiments.

The candidate of note can be obtained by determining the peaks occur in each estimated saron waveform. Each note candidate has its note number, the magnitude of cross power density, the offset and the onset.

All note candidates are sorted by the onset. When there are more than one note candidates which occur at the same time, the real note is determined by choosing the one with highest magnitude of power density.

Gamelan sound signal is influenced by many external factors such as hammer stroke styles and stroke strength, so it has various sound qualities [15]. Figure 9 shows the spectrum of gamelan sound which varies very much due to the hardness and the style of stroke although it still has the same fundamental frequency. This figure shows a fundamental frequency of 2000 Hz. The harder stroke resulted in the emergence of harmonic sound which has lower frequency than that of gamelan sound signal. The length of harmonic sound is also much shorter compared with the fundamental frequency.

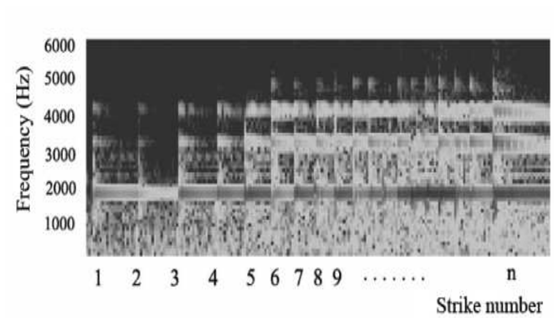


Fig.9 The difference spectrum of the signal due to differences in hammer stroke strength

Moreover, the sound of an instrument in gamelan ensemble 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. In fact, the length of bonang sound has shorter compared to that of saron. While the length of saron sound has 500 ms, the length of bonang sound only has 150 ms. Figure 10 shows several estimated saron waveform with various window's length. The presence of bonang sound appears as small pulses inside the extracted saron sound for 40 ms of window's length.

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

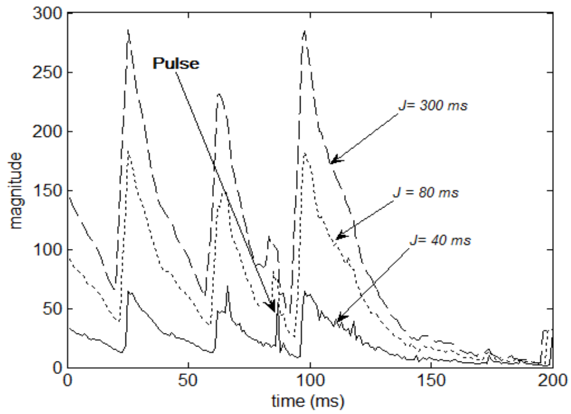


Fig. 10 Estimated Saron waveforms due to various window's length, *J*

The harmonic sound which is caused by the interference of another instrument in gamelan ensemble can be eliminated in two ways. First, the presence of pulses can be suppressed by altering the window's length. Larger window makes the power density of bonang sound to be much smaller than that of saron sound. The second way is the use of low-pass filter to suppress any ripples or pulses inside saron sound. Figure 11 shows how an LPF (with different values of the window's length *J*) deprived pulses inside saron's notation signal.

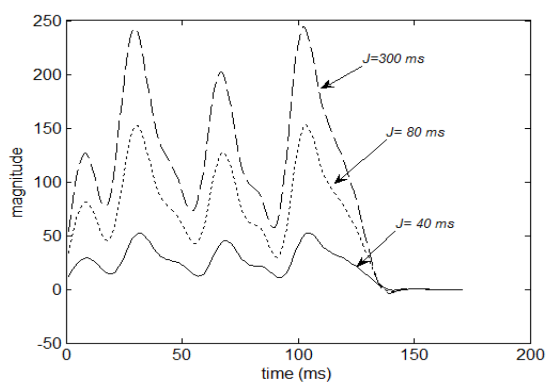


Fig. 11 Estimated Saron waveform after using Low Pass Filter (LPF)

Since the frequency distance between successive notes in saron instrument is only about 100 Hz, an LPF with sharp transition is needed. The Kaiser window was chosen to create a low pass filter, because it generates the smallest number of filter orders to obtain a sharp low pass filter [16]. Several parameters in the low pass filter including the the passband (*Fpass*), and the stopband (*Fstop*). See fig. 12.

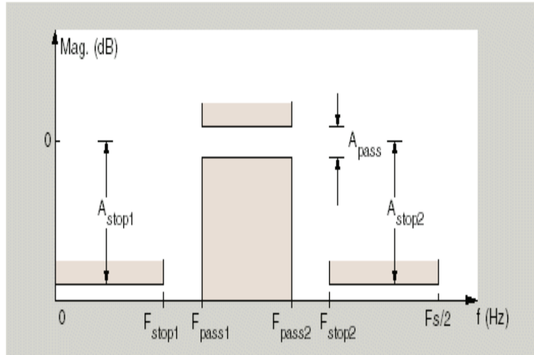


Fig. 12 Bandpass filter parameters

Smoother waveform in notation signal will increase the accuracy of music transcription because it will minimize errors of transcription related to the presence of pulses inside the signal.

3. PERFORMANCE EVALUATION

3.1 The Gamelan Songs for Testing

We generated two types of gamelan sound for testing [17]:

1. Synthetic. Each gamelan note was recorded and the ensemble was played using computer with gamelan note direction.
2. Full acoustic. Gamelan ensemble was played by the players and was recorded. We played two songs, song1 and song2. Both songs were recordings of gamelan ensemble performances which were consisted of nine simultaneously played instruments. Song1 was 110 seconds and song2 was 90 seconds of duration. Song1 contained 161 original notes, and song2 contained 129 original notes.

3.2 Gamelan Transcription:

In order to show the effectiveness of template matching for gamelan transcription, various types of playing, such as single synthetic gamelan, mixture of three synthetic gamelan and gamelan ensemble were investigated. To evaluate the estimated generated notes, we used the note error rate [11][12]. Recognition of error rates is stated by Eq.(13),

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

$$ner = \frac{deletion + insertion + substitution}{totaltruesentence} \quad (13)$$

For evaluating the performance, gamelan orchestra was investigated to show the effectiveness of template matching for music notation. The conventional methods STFT, Barbancho, Rodger, Bello, were investigated with our test data. For the smallest *ner* we did STFT by using varying window's length and a fix hop length [18]. Figure 13 shows that by the window's length of 8192 samplings and hop length of 2048 sampling, the performance produced the smallest note error rate (*ner*). Using overlap STFT, we obtained *ner* 9% for song1 and 18% for song2. LPF-CC generated *ner* 6% for song1 and *ner* 16% for song2. LPF-CC and LPF generated *ner* 5% for song1 and *ner* 14% for song2. The overall results were then compared with our proposed method Table 2 shows the results as the ratio of *ner*.

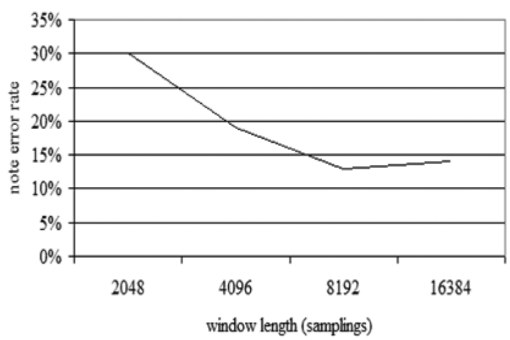


Fig.13 Note error rate *ner* against various window's length.

Table 2 Performance of saron extraction for gamelan transcription by conventional method STFT and LPF cross-correlation (LPF-CC).

Test Type	Total Nota-tions	Instru-ments	8192 STFT	ACC	LPF-CC
Synthetic	30	1	2%	0%	0%
Synthetic	30	3	5%	3%	2%
Acoustic sounds, Song 1	161	9	9%	6%	5%
Acoustic sounds, Song 2	129	9	18%	16%	14%

4. CONCLUSION

In order to construct a robust instrument extraction from music ensemble, a smoother Saron time-frequency model and a template matching scheme were proposed. According to the performance test, the proposed method provides 2 - 4% improvement from the conventional method. This result shows the effectiveness of template matching for picking up specified instrument for music transcription. The proposed method can also be applied to other acoustic instruments apart from gamelan.

REFRENCES:

[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, 1992, 1995. Cultural interaction and musical development in central java, The University of Chicago Press, ISBN 0-226-78011-2.

[5] Eric Scheirer, 1995, “Extracting expressive performance information from recorded music,” Master’s thesis, MIT, 1995

[6] Barbancho, A. Jurado, 2004. Transcription of piano recordings, Applied Acoustic 65 (12 SPEC. ISSN).

[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] Rodger J. McNab and I. H. Witten, 1996. Signal processing for melody transcription, Proceedings of the 19th Australian Computer Science Conference, Melbourne, Australia.3

[9] Pielemeier Willam J. and M. H. Simoni, 1996. Time-frequency analysis of musical signals, Proceedings of The IEEE, vol. Vol.84, No.9.

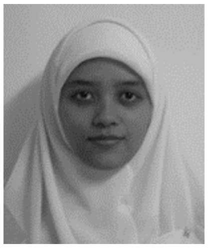
Journal of Theoretical and Applied Information Technology
15th October 2011. Vol. 32 No.1

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

[10] David Havelock, Sonoko Kuwano, 2008. Handbook of signal processing in acoustics, Springer New York.
[11] Raphael C., 2002. Automatic transcription of piano music, In Proc. ISMIR.
[12] Klapuri A P, 2003. Automatic transcription of music, Proceedings of the Stockholm Music Acoustics Conference, Sweden.
[13] Dolson, 1996. The phase vocoder: A tutorial, Computer Music Journal, vol. Vol. 10, No. 4.
[14] Yoyon K Suprapto, Mochamad Hariadi, Tsuyoshi Usagawa, 2009. Time frequency modeling of gamelan instrument based on time frequency for automatic notation, the Third International Student Conference on Advanced Science and Technology, Seoul, Korea.
[15] Bokyung Sung, Jungsoo Kim, Jinman Kwun, Junhyung Park, Jihye Ryeo, and Ilju Ko, 2009. Practical Method for Digital Music Matching Robust to Various Sound Qualities, World Academy of Science, Engineering and Technology 60
[16] Oppenheim, RW Schaefer & J Buck, 1999, Discrete Time-Signal Processing, 2nd ed., Prentice Hall.
[17] Bokyung 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.
[18] 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.

AUTHOR PROFILES

Yoyon K Suprapto received the bachelor degree in Electrical Engineering from Bandung Institute of Technology, Bandung, Indonesia in 1977. He received his Master of Science Computer Science in The University of Missouri, Columbia, Missouri, USA in 1981, and Doctor degrees in in Electrical Engineering Department of Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 2010 respectively. He joined Electrical Engineering Department in Institut Teknologi Sepuluh Nopember since 1977. His current interests research area are Data Mining, Sound Signal Processing and Traditional Music.



Diah Puspito Wulandari received her bachelor degree in electrical engineering from Institut Teknologi Bandung (ITB) Indonesia and her master degree in artificial intelligence from University of Edinburgh United Kingdom in 2004 and 2006 respectively. She has been attached as a lecturer in Department of Electrical Engineering Institut Teknologi Sepuluh Nopember Surabaya Indonesia since 2005 and she is now a doctoral student in the same department. Her research interest focuses in speech and audio signal processing.

Aris Tjahyanto received the B.E. degree in Electrical Engineering Department of Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 1989. He received M.E. degrees Computer Science Universitas Indonesia in 1995. Currently, he is the staff of Information Systems Deparment of Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia and he is now a doctoral student in the same department.. His research interest is in Information Systems, Speech recognition and Signal Processing.