Time-frequency analysis on *gong timor* music using short-time fourier transform and continuous wavelet transform

Yovinia Carmeneja Hoar Siki [1], Natalia Magdalena Rafu Mamulak [2]

Department of Informatics Technology, Universitas Katolik Widya Mandira, Penfui, Kupang, Indonesia
*[1] cyovinia@yahoo.com *; [2] mamulak.natalia@gmail.com*
* corresponding author

ARTICLE INFO

ABSTRACT

Time-Frequency Analysis on Gong Timor Music has an important role in the application of signal-processing music such as tone tracking and music transcription or music signal notation. Some of Gong characters is heard by different ways of forcing Gong himself, such as how to play Gong based on the Player's senses, a set of Gong, and by changing the tempo of Gong instruments. Gong's musical signals have more complex analytical criteria than Western music instrument analysis. This research uses a Gong instrument and two notations; frequency analysis of Gong music frequency compared by the Short-time Fourier Transform (STFT), Overlap Short-time Fourier Transform (OSTFT), and Continuous Wavelet Transform (CWT) method. In the STFT and OSTFT methods, time-frequency analysis Gong music is used with different windows and hop size while CWT method uses Morlet wavelet. The results show that the CWT is better than the STFT methods.

I. Introduction

Gong is a traditional musical instrument of Nusa Tenggara Timur (NTT), Indonesia which is different in each region. The differences are used in the number of Gong such as the size of the gong, how to play and the tone of the gong. Gong music notation is diatonic which contains *Do*, *re*, *mi*, *sol*, *la* and *do*. Gong-making process has been done manually by forging of which the process causes free tone frequency to be applied so that it produces a variety of resonance, pitch and amplitude [1]–[3]. Some of these differences makes the development of eastern music behind, so it need a research that can encourage the development of eastern music.

Currently in Indonesia, traditional music research has been conducted in the groups of gamelan instruments. Signal extraction studies on gamelan music mentioned frequency diversity among the group of gamelan musical instruments [4]–[8]. This prompted researchers to examine musical instruments Gong to determine its diversity.

As mentioned in the preliminary study [2][3][8]–[10], this study analyzes the Gong music signal based on the frequency of time. This research uses the Gong instrument of North Central Timor District especially in North Bikomi, Indonesia to accompany a dance called Gong dance. In addition, the dance Gong used to receive a plenty of guests and other traditional celebrations. Notation songs played to accompany the dance Gong is a traditional regional song played by the players. Furthermore, Gong musical instruments is played by three people at Gong drum percussion set. Very fast tempo and instrument Gong tone spacing approaching each other. In addition, Gong dancers wear *Kerinci* as an accessory that adds a sound in Gong's music game.

Fig. 1 shows the Gong instrument consisted of five Gongs. Five Gongs show the number of Gong notation instruments including *do*, *re*, *mi*, *sol* and *la*. Five notations are divided into three parts: "tonu mese" is a small soprano like Gong that describes 'la' tone. "Ote" is two medium Gong as tenor for melody with Gongs at the top of the noodles and the bottom of Gong pitched in soles. While "kola" is

two large Gongs that are like bass with a Gong over a pitched *do* and Gong down pitched *re*. Notation Gong does not have the low and high tones.

Fig. 1. Gong Music Percussion

Time-frequency analysis of music signals is important in the development of other music signal processing applications such as a peak of detection, blow detection and tone tracking (called transcriptions) [3][9][11]. Comparison of the appropriate frequency time required to overcome short blow-density and amplitude. Thus the study of musical instruments Gong can be developed.

Short-Time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT) methods are used to display Gong music signals in the frequency domain of time. Spectrogram shows the frequency signal STFT time in the process, and schallogram shows CWT process. Thus, the difference is that the STFT process represents a time-frequency signal based on the size of the selected window so that the frequency in that window will be the same for all time [5][12][13], while the CWT process using the mother wavelet as a window that serves as a filter [14][15].

The purpose of this research is to get various frequency representation of Gong tone which can be used for development of other Gong music signal processing applications. Also, it can compare the performance of the method used for the development of further research. The sound of the gong which is used is a recording of two randomly chosen tones.

The performance evaluation of the signal extraction method is calculated by applying the Not Error Rate (NER) parameters. The accuracy of this NER method can be seen from the time frequency representation of the original signal from the analysis. The analysis result is expressed correctly if it produces a small NER.

II. Method

A. Research design

Fig. 2 shows the general purposed procedure. The input data is processed using a semi-synthetic process and then it separated by three methods namely STFT, OSTFT, and CWT. Spectrogram signal output is filtered to get the output of a separate recording. Furthermore, the output is evaluated by comparing the APM, and NER output with the smallest will be the choice.
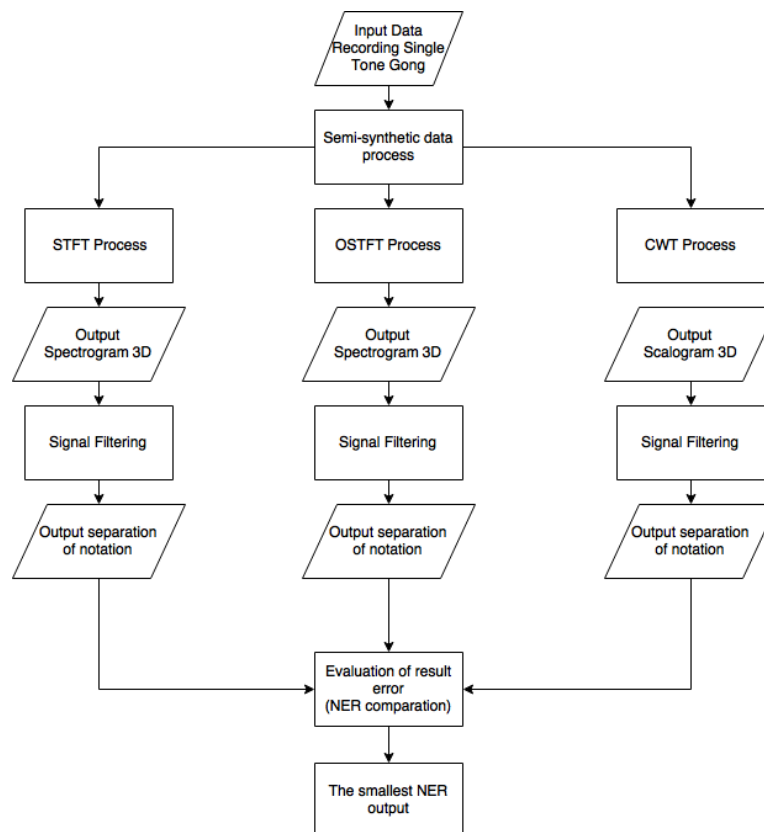
Fig. 2. General Purposed Procedures

B. Data Collection

1. Recording Data (Acoustics)

The process of recording data using Behringer ECM8000 Condherer sound card and a microphone as sound energy converter (acoustic) into the computer's power, this would be equivalent to the results of the computer recording software installed. Illustration of data recording can be seen in Fig. 3.
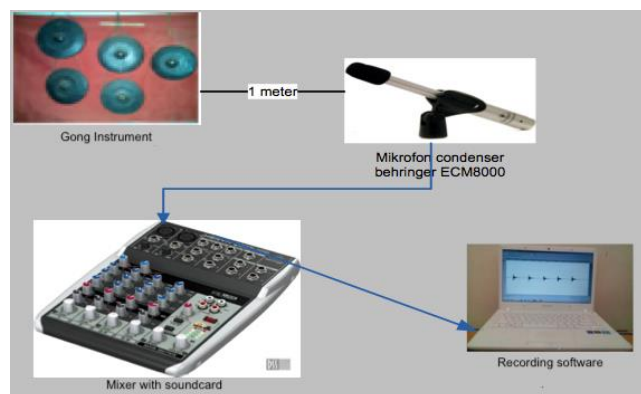

Fig. 3. Illustration of data recording

2. Semi-synthetic Data

Semi-synthetic data obtained by recording the data incorporation Gong tone to the same song with the original notation of the song. Semi-synthetic illustration of data can be seen in Fig. 4.
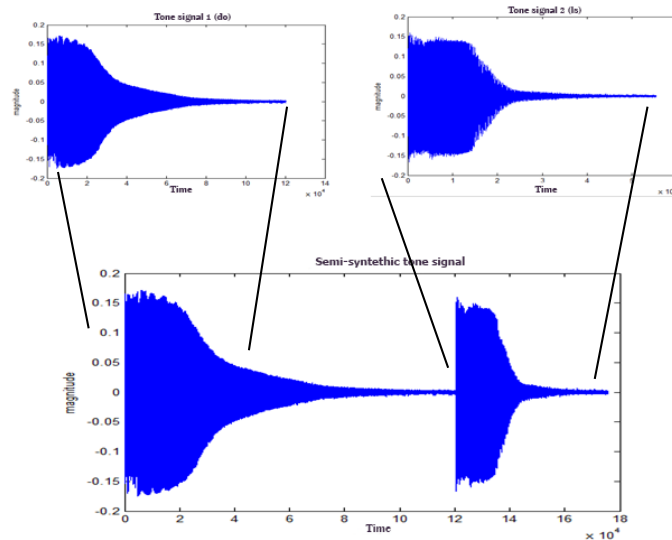
Fig. 4.Illustration of semi-synthetic data

3. Research Data

The research data used in this study is an instrument Gong. Details of the research data is shown in Table 1. Total data used 9 data with the number of tones is 17 units.

Table 1. Research Data

No.	Tone	Number of tones
1	Single acoustics (*do*)	1
2	Semi Synthetic (*do sol*)	2
3	Semi Synthetic (*do mi*)	2
4	Semi Synthetic (*re sol*)	2
5	Semi Synthetic (*re la*)	2
6	Semi Synthetic (*do la*)	2
7	Semi Synthetic (*sol do*)	2
8	Semi Synthetic (*sol re*)	2
9	Semi Synthetic (*la re*)	2
	Total	17

C. Performance Evaluation

Evaluation of error in the result of the division aims to determine the accuracy of the analysis method track of time spent. If a small percentage-error-rate is generated, the accuracy of the method used is better. On the other hand, the accuracy of the method is stated worse if the percentage error rate is greater than the accuracy of the method used. Performance measurement is calculated by using (1).

$$NER = 100\ \% \ x\ \frac{insertion+deletion+substitution}{number\ of\ notes} \tag{1}$$

The insertion represents a tone that should not exist but is recognized as a tone, deletion represents an unknown tone deleted, and the substitution represents a misidentified tone.

III. Results and Discussion

A. Short-Time Fourier Transform (STFT) Process

Signal feature extraction using STFT generate frequency domain signal time (Fig. 5). The initial process had been done by placing the selected window on the input signal. This window will divide the signal into several sections based on the width of the selected window into a frame. For example,

the selected window is N = 8000 and input signal has signal length l = 48000 sampling per second. The feature extraction will start from the signal length l = 1 - 8000 for sampling the first N window as the first frame. Next, feature extraction is performed again for the next signal length with signal length l = 8001 - 16000 for sampling the second N window as the second frame. The extraction continues until it reaches the last N sampling window located at the length of the signal l = 40001 - 48000.
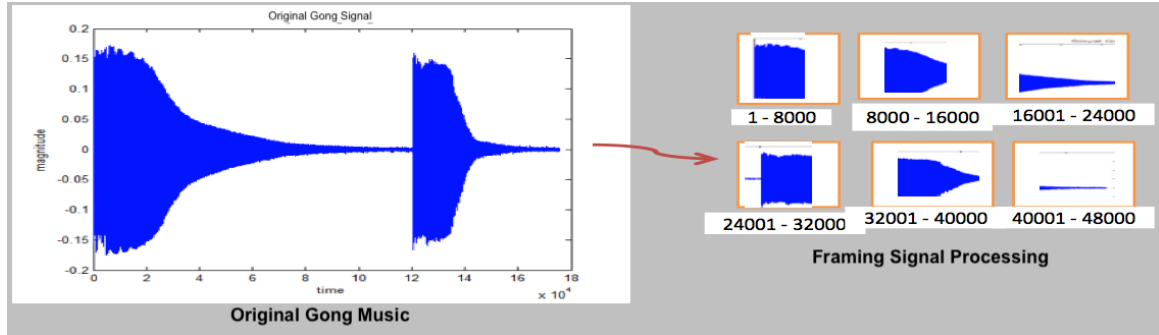

Fig. 5. Ilustration of STFT process

B. Overlap Short-Time Fourier Transform (OSTFT) Process

OSTFT process is STFT development methods to produce a smoother signal spectrum. The initial distance from the next frame will be processed at the beginning of the previous frame (called h-size). The hop size *h* is arranged so that the frames overlap with *h* greater than N (h < N) [5][7][10]. Hop size is calculated by using (2).

$$h = 0.25 \; x \; N \tag{2}$$

Hop Size initialized by *h*, a constant value in determining the size of hops using 0.25 and N is initialized as the width of the window.

OSTFT process (Fig. 6) is begun by taking a window of N in the signal then determines the size of the size of h hops to make the process of feature extraction. Signal extraction step is the same as the STFT process by dividing the input signal into a window frame according to the length of N, the frame signal is given hop with size *h*. For example, the used hop size is 1000 and the input signal has a signal length l = 48000 sampling per second. The feature extraction process will start from the signal length l = 1 - 8000 for sampling the first N window with the hop size of 1000 for *h*. Furthermore, the extraction starts from signal length l = 1000. This means that the extraction process will begin on the long signal l = 1000 to the end of the first frame signal and a proxy of the second frame to the last frame.
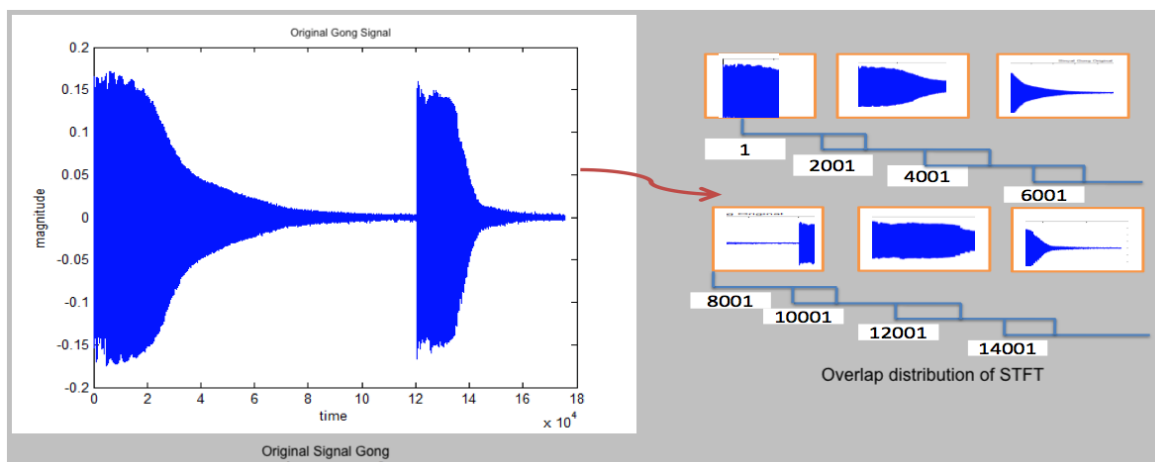

Fig. 6. Illustration of OSTFT process

C. Continous Wavelet Transform (CWT) Process

The extraction process is begun by placing the wavelet at the beginning of the signal according to the time scale $\tau = 0$ on the scale $a = 1$. Then it is multiplied by the signal until everything is integrated. Furthermor, the integration is multiplied by the constant $\frac{1}{\sqrt{s}}$. Multiplication with constants aims for the normalization of energy so that the convoluted signal has the same energy in each scale. The end of the result shows the CWT transformation value at time $\tau = 0$ and a scale = 1. Wavelets on a scale = 1 then shift to the right at time $\tau = 1$. Over time τ on a scale = 1 is calculated and obtained the transformation value at scale = 1. Furthermore, this procedure is repeated for each scale a by increasing it. The calculation results for a given value will fill one line according to the time scale field. This calculation process is performed for all desired values until all signals are calculated. Fig. 7 illustrates the CWT signal calculation.
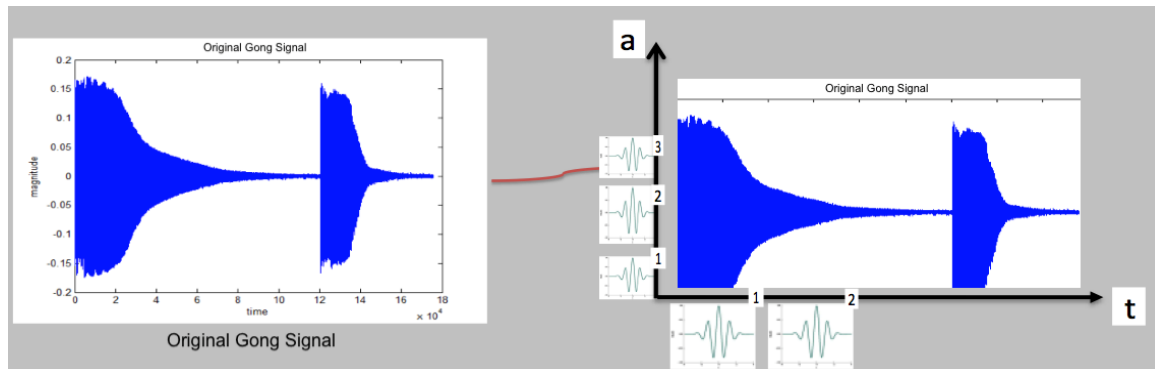

Fig. 7.Illustration of CWT process

D. Data Filtering

Tone filter processes a different tone according to the frequency band tone into the path notation, notation first for the first record, notation-2 for the second record, notation 3rd to the third record, notation 5th to record the fifth, and the notation to 6 for The sixth record. Division notation is based on the fundamental frequency for STFT and OSTFT method, and based on the basic scale cwt method. The basic frequency and base scale of the Gong tones used in the signal filtering process are shown in Table 2. The basic frequencies and basic scales are obtained from observations of three Gong instruments.

Table 2. Frequency and scale of Gong tone on three Gong instruments

Tone	Set 1		Set 2		Set 3		Min		Max	
	Scale	*Freq*	*Scale*	*Freq*	*Scale*	*Freq*	*Scale*	*Freq*	*Scale*	*Freq*
Do	90	428	100	450	108	386	90	386	108	450
Re	80	490	86	438	99	390	80	390	99	490
Mi	71	549	80	481	85	453	71	453	85	549
Sol	59	639	69	549	75	540	59	540	75	639
La	53	760	52	645	68	640	52	640	68	760

In the process of STFT and OSTFT, the signal filtering process is performed by dividing the frequency into the maximum path and minimum signal path by using the basic frequency range in Table 2. For example, the maximum frequency of the *Do*'s tone is 450 and the minimum is 386. Thus, signal with frequency band range 386 to 450. The filtering process in CWT is performed by dividing the frequency into the maximum path and minimum signal path by using the base scale range in Table 2. Supposing the maximum scale on the *Do*'s tone is 108 and the minimum is 88. Thus, the search process will get a signal whose range of paths is 88 to 108.

E. Time Accuracy

The comparison of time accuracy shows the comparison of time-frequency accuracy in semi-synthetic *mi do* of research results. Comparison of experimental data while others are not shown because their basic similarity frequency accuracy of time for each tone in the experimental data,

However, the method used will make differentiation. Accuracy of time obtained through the equation (3).

$$X = \frac{Fs \; x \; (t_{2(STFT;OSTFT)} - t_{1(STFT;OSTFT)})}{t_{2(CWT)} - t_{1(CWT)}} \; x \; \frac{1}{1000} \tag{3}$$

Comparison is done by comparing the accuracy of CWT method with STFT and OSTFT methods. Based on the observations, the timeline comparison of three methods is shown in Table 3.

The comparison results show that the CWT method has an accuracy of less than 1 millisecond (<1 ms). The smallest accuracy obtained after making a comparison between STFT and OSTFT methods that have the best time duration.

Table 3. Comparison of CWT method time accuracy between STFT and OSTFT

STFT	OSTFT	CWT	Accuracy (Delta t)	
			STFT/CWT	*OSTFT/CWT*
15 ms	23 ms	80000 ms	0.01	0.02
3.5 ms	12 ms	80000 ms	0.00289	0.0099
2 ms	6 ms	80000 ms	0.00165	0.00496

F. Filtering Data

Evaluation of the accuracy obtained from the screening aimed to determine the accuracy of the method of dividing the signal into the path notation. The evaluation was done by calculating (NER). Best accuracy has been produced by the method that has the smallest error. The accuracy of the performance results is shown in Table 4.

Table 4. Accuracy STFT comparison method, OSTFT, and CWT maintain tone based on screening results

Method	NER (%)
STFT (N = 24000)	35.29 %
STFT (N = 16000)	35.29 %
STFT (N = 8000)	35.29 %
OSTFT (N = 24000)	47.05 %
OSTFT (N = 16000)	35.29 %
OSTFT (N = 8000)	35.29 %
CWT	5.88 %

The obtained accuracy shows that the STFT method has the same precision value for all window lengths. OSTF methods produce the best accuracy on the length of the window 16000 and 8000, while the length of the window to 24000 results have an accuracy value of 47.05% due to the large window length. The length of the large windows known to provide low resolution and low frequency so that the signal amplitude widened. However, the three methods used by CWT provide the best accuracy of 5.88%.

IV. Conclusion

The study compared the frequency of two-tone Gong Timor using the Short-Time Fourier Transform (STFT), Overlap Short-Time Fourier Transform (OSTFT), and Continuous Wavelet Transform (CWT). Experiments of 9 data showed that CWT has the best accuracy with an error rate of 5.88%. These results indicate that the CWT method is more appropriate to represent the time-frequency domain tone signals in a variety of tones Gong, such as the detection of onset and the introduction of musical notation or transcription.

References

[1] M. H. Budhiantho and G. Dewantoro, "Javanese gong wave signals," in *Proceedings of Meetings on Acoustics*, 2014, p. 035003.

[2] M. H. W. Budhiantho and G. Dewantoro, "Homomorphic filtering for extracting Javanese Gong wave signals," in *2014 8th International Conference on Telecommunication Systems Services and Applications (TSSA)*, 2014, pp. 1–6.

[3] M. H. W. Budhiantho and G. Dewantoro, "The spectral and temporal description of Javanese Gong Kempul," in *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2013, pp. 300–304.

[4] A. P. Klapuri, A. J. Eronen, and J. T. Astola, "Analysis of the meter of acoustic musical signals," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 1, pp. 342–355, Jan. 2006.

[5] K. Hastuti and K. Mustafa, "A method for automatic gamelan music composition," *Int. J. Adv. Intell. Inform.*, vol. 2, no. 1, pp. 26–37, 2016.

[6] L. Fitria, Y. K. Suprapto, and M. H. Purnomo, "Music transcription of Javanese Gamelan using Short Time Fourier Transform (STFT)," in *2015 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2015, pp. 279–284.

[7] D. P. Wulandari, Y. K. Suprapto, and M. H. Purnomo, "Gamelan music onset detection using Elman Network," in *2012 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings*, 2012, pp. 91–96.

[8] T. P. Tomo, G. Enriquez, and S. Hashimoto, "Indonesian puppet theater robot with gamelan music emotion recognition," in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2015, pp. 1177–1182.

[9] H. Katayose, M. Imai, and S. Inokuchi, "Sentiment extraction in music," in *[1988 Proceedings] 9th International Conference on Pattern Recognition*, 1988, pp. 1083–1087 vol.2.

[10] M. Muller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal Processing for Music Analysis," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 6, pp. 1088–1110, Oct. 2011.

[11] M. N. Latief, F. L. Gaol, and B. H. Iswanto, "Analysis and Identification Gamelan Bonang Sound Spectrum," in *2010 Second International Conference on Computational Intelligence, Modelling and Simulation*, 2010, pp. 335–338.

[12] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A Tutorial on Onset Detection in Music Signals," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035–1047, Sep. 2005.

[13] A. Tjahyanto, D. P. Wulandari, and Y. K. Suprapto, "Time-Frequency Analysis of Gamelan Sound Signals using Continuous Wavelet Transform," in *Seminar on Intelligent Technology and Its Applications*, 2014.

[14] P. De Gersem, B. De Moor, and M. Moonen, "Applications of the continuous wavelet transform in the processing of musical signals," in *Proceedings of 13th International Conference on Digital Signal Processing*, 1997, vol. 2, pp. 563–566 vol.2.

[15] A. Paradzinets, H. Harb, and L. Chen, "Use of continuous wavelet-like transform in automated music transcription," in *Signal Processing Conference, 2006 14th European*, 2006, pp. 1–4.

[16] Y. Ge and D. G. Daut, "Bit error rate analysis of digital communications signal demodulation using wavelet denoising," in *34th IEEE Sarnoff Symposium*, 2011, pp. 1–6.