Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)
15/1 (2022), 9-19. DOI: http://dx.doi.org/10.21609/jiki.v15i1.1009

# Wavelet Transformation and Spectral Subtraction Method in Performing Automated Rindik Song Transcription

[1]Yuriko Christian, [2]I Dewa Made Bayu Atmaja Darmawan

[1,2]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana, Jalan Raya Kampus Unud, Badung, Bali, 80361, Indonesia

*E-mail: [1]yurikochristian@cs.unud.ac.id, [2]dewabayu@unud.ac.id*

**Abstract**

Rindik is Balinese traditional music consisting of bamboo rods arranged horizontally and played by hitting the rods with a mallet-like tool called "panggul". In this study, the transcription of Rindik's music songs was carried out automatically using the Wavelet transformation method and spectral subtraction. Spectral subtraction method is used with iterative estimation and separation approaches. While the Wavelet transformation method is used by matching the segment Wavelet results with the Wavelet result references in the dataset. The results of the transcription were also synthesized again using the concatenative synthesis method. The data used is the hit of 1 Rindik rod and a combination of 2 Rindik rods that are hit simultaneously, and for testing the system, 4 Rindik songs are used. Each data was recorded 3 times. Several parameters are used for the Wavelet transformation method and spectral subtraction, which are the length of the frame for the Wavelet transformation method and the tolerance interval for frequency difference in spectral subtraction method. The test is done by measuring the accuracy of the transcription from the system within all Rindik song data. As a result, the Wavelet transformation method produces an average accuracy of 83.42% and the spectral subtraction method produces an average accuracy of 78.51% in transcription of Rindik songs.

**Keywords:** *Rindik, Automatic Music Transcription, Wavelet Transformation, Spectral Subtraction*

## 1. Introduction

Rindik is one of the traditional musical instruments originating from Bali [1]. Rindik is a musical instrument made of several bamboo rods that are arranged horizontally and played by hitting it with two mallets in each player's hand called "panggul" [2]. Currently, Rindik gamelan players generally learn by imitating directly from a player by playing a song and there are not many attempts to transcribe Rindik songs so that they can be studied by Rindik players. The difficulty in playing this musical instrument is to use two hands which the playing pattern between the two hands is uncertain depending on the music being played. Visual representations such as guitar tabs can make it easier for guitar players to compose, share, and learn songs rather than using traditional musical notation [3]. In this study, we want to provide a solution to be able to transcribe the Rindik song into a visual representation of the Rindik song from the audio recording of the Rindik. The proposed system must be able to separate the two tones produced from the two Rindik rod in a signal.

Automatic music transcription (AMT) [4] is the capability to transcribe acoustic music signals into notation or other visual forms of music. This process involves perception, cognition, knowledge representation and inference.

Various previous studies have been conducted related to AMT. AMT [5] is a process that requires an effective solution because of the complexity of the musical sound. This research performs automatic transcription of the harmonic and melodic sound signal parts. The neural network model [6] was used to produce music scores. The results obtained show high accuracy in monophonic audio signals and also apply to polyphonic signals.

AMT on polyphonic signals can be done by detecting the onset and offset of the audio signal [7]. Onset and offset detection are used to perform segmentation prior to pitch detection. Sound

signals originating from multi-instrument forming polyphonic music signals can be identified using semantic segmentation [8].

AMT for traditional music has been performed on Hindustani instruments [9] by normalizing the cent scale, Multi-String PLCA on Chinese Traditional Plucked String Instruments [10], and Javanese Gamelan using STFT [11]. A research [12] has been carried out to identify the type of Rindik tone by using a wavelet transformation obtained an accuracy of 93.18%. However, this study only identified the type of Rindik, not the tone transcription. The challenge in this research is how to transcribe notations for multi-pitch like the Rindik instrument. Frequency detection with an iterative estimation and separation approach is used to estimate multi-pitch [13]. This approach is carried out with a method to remove a dominant part of the analyzed sound in the frequency domain which in this case is the spectrum. To remove the dominant part of the analyzed sound in the frequency domain, it can be done by signal subtraction or signal reduction approaches.

One of the signal subtraction approaches that have been carried out in previous studies is spectral subtraction [14]. In a study conducted by Christian and Darmawan, spectral subtraction was used to separate the sound of 2 Rindik rods that being hit simultaneously resulting an average MSE (mean squared error) of 0.0126 and a SIR (signal to interference ratio) of 55.68 dB [1]. In this study, the first pitch estimation for one segment of the rod hit was carried out and the Rindik tone with the nearest frequency to be used as noise in the spectral subtraction process. After spectral subtraction is performed, a second pitch estimation is performed. If a corresponding frequency is found (within the tolerance interval of the frequency difference) with one of the Rindik rods, then that frequency can be recognized and is a hit of 2 Rindik rods simultaneously. If there are no frequency matches one of Rindik's rods, then that frequency cannot be recognized and is only single rod sound.

Unlike modern musical instruments, traditional musical instruments do not have a fixed pitch [15]. Pitch differences can be found in some instrument craftsmen. However, the pitch interval of the instrument remains the same. This is a wealth from a cultural point of view, but a challenge in the process of automatic music transcription. Therefore applications built for traditional music transcription require a calibration feature to adjust the base note of each rod. The Spectral Subtraction method is expected to reduce the tone reference used as a comparison in the sound matching process. This reduces the space required to store the reference sound in the calibration process when compared to conventional pitch detection methods such as Wavelet.

In this paper, we describe the comparison of the Wavelet and Spectral Subtraction methods to create a transcription of traditional rindik music. We observed how the performance in terms of accuracy of the Spectral Subtraction method in transcribing Rindik song compared to the Wavelet Transformation method. We also want to know which method with each parameters performs the best. This system is expected to help users who are interested in traditional musical instruments, especially Rindik, to make learning easier.

## 2. Research Method

This section will discuss the methods, data, and stages carried out in the research. There are 2 methods that will be used in this research, which are wavelet transform and spectral subtraction. Tests were carried out to test several parameters, the parameters are the frame length of the wavelet transform method and the tolerance interval of the frequency difference in the estimation of the tone on the spectral subtraction method. The parameters of the Wavelet transform method are the length of the frame (512, 1024, and 2048 samples), while the spectral subtraction method is the tolerance interval for the difference in frequency (0.55, 1.10, and 1.66 Hz). The tolerance interval of the frequency difference is obtained by calculating the maximum and minimum values of the difference between the frequencies of the reference data using FFT and dividing them into 3 different interval.

### 2.1. Rindik

Gamelan Rindik is one of the traditional musical instruments originating from Bali [16]. Each bamboo rod has a different size from each other. Rindik consists of 11-14 bamboo rods are cut and then arranged on the left is a larger rod to a smaller size on the right. The bamboos are arranged in a row, tied together, and framed with decorated, carved and painted wood.

**Fig. 1.** Rindik instrument consisting 11 bamboo rod

## 2.2. System Design

The initial process begins by segmenting Rindik's song audio into several segments for each rod hit with onset segmentation. The tone of the segment will then be estimated and the results of the tone estimation are stored to be used as a result of Rindik song transcription.

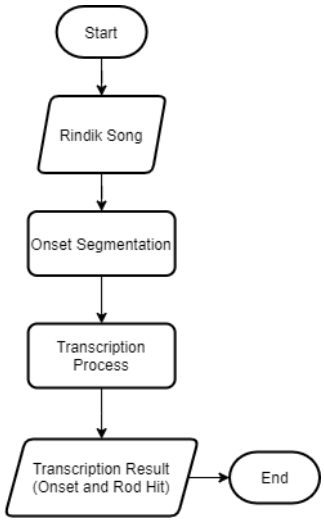


**Fig. 2.** Proposed system flowchart

In this research, two methods were used as a comparation for the transcription process. The methods are Wavelet Transformation and Spectral Subtraction method. The Spectral Subtraction method were applied with the iterative estimation and separation approach. After the system runs and produces the results in the form of music transcription, the system is tested.

System testing is carried out to measure the accuracy of the prediction of each audio segment according to the original label. The test is carried out by observing the accuracy of the frame length parameter on the Wavelet Transformation method and the frequency difference tolerance interval parameter on the Spectral Subtraction method.

## 2.3. System Application

The stages in the system design section will be implemented in a Python 3.7 application with its built in and external library. The system built as a desktop application with graphical user interfaces. The application could receive input as a form of an audio file. The audio file input will be a Rindik song. User could also choose the method of the transcription with the parameters. The transcription result can be both viewed as an application interface display and saved as a PDF file.

Interface design of the proposed system were drawn to visualize the application interface. The windows includes an initial setup window and the display of the transcription result. Figure 3 and 4 is the interface design of the application.

The first window when the application run shows a few field and button. The buttons is “Browse” button to select the Rindik song file. The “Transcription Method” contains two radio buttons to select the transcription method. After the user select one of the transcription method, the “Set Parameter” button should be click to show the “Set Parameter” window to select parameter according to the transcription method. Then, the user directed to the initial screen and able to begin the transcription process of the song by clicking the “Start Transcription Process” button.

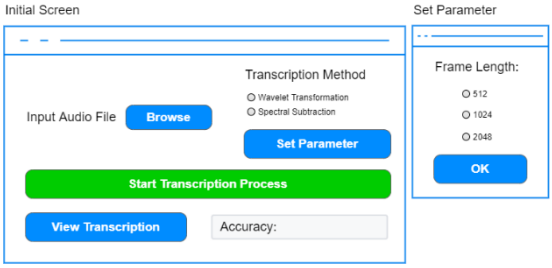


**Fig. 3.** Iterative estimation and separation flowchart

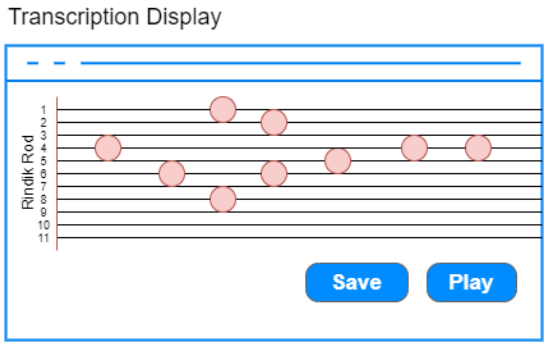


**Fig. 4.** Transcription Result Display

All the red dots in the display of Figure 4 will moving at the same speed towards the red line on the left side of the display when user clicks play. Right when a red dot or two red dots touches the red line, a Rindik rod or two Rindik rods supposed to be hit according to the order of the rods (top to bottom is left to right in Rindik instrument). The display is expected to help Rindik players to learn a Rindik song.

The library used to build the application of the system: Librosa [17], Numpy [18], PyWavelet [19], and Pygame [20]. Librosa is a digital audio signal processing library in Python. Numpy is an array manipulation and calculation library. PyWavelet is a library in python to do the Wavelet

Transformation. Pygame is a library in Python to build interactive interface.

### 2.4. Dataset

The data is recorded directly from the sound source of the Rindik musical instrument played by the Rindik instrument player himself and taking samples of every single hit of a single rod and a combination of two rods hit. A Rindik musical instrument was used in this study. Rindik used has 11 rods, each rod and every combination of two rods hit recorded 3 times. 33 records of single rod sound and 165 records of two rods sound. The data used for testing are recordings of 4 rindik songs with tempos between 60 - 70 BPM played by a rindik player and each song recorded 3 times, so there are 12 rindik song recordings for testing.

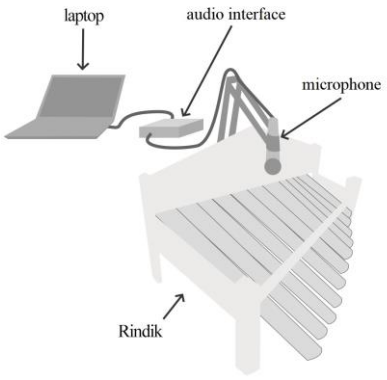


**Fig. 5.** Rindik recording setup

Audio data is captured directly using a microphone and an audio interface and recorded using Adobe Audition software in "*.wav" audio format with a sample rate of 44100 Hz. Figure 5 is a picture of the setup for the rindik audio recording when retrieving data using 1 microphone, 1 audio interface, and 1 laptop.

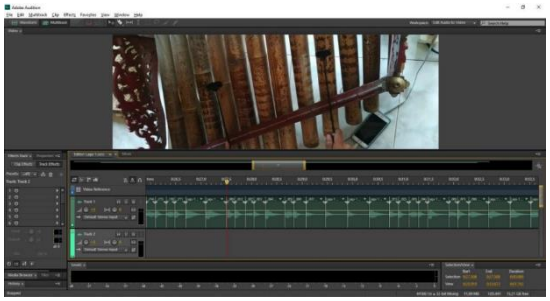

**Fig. 6.** Segment Labeling Process

The label of each segment result of the onset segmentation is manually labeled by matching each segment with a video of the performance of each song taken from Rindik's music. Labeling is performed by using the Adobe Audition application. Figure 6 shows how each segment is being labeled using the Adobe Audition application.

The dataset obtained by recording directly to the Rindik instrument with the setup in Figure 5. The dataset consists of 4 Rindik song and 11 rod sound with 55 two rod sound combination. All were recorded 3 times. The Rindik song 1 has a faster tempo compared to the other three. The Rindik song 1 has tempo about 70 beats per minute and the other 3 song has tempo about 60 beats per minute.

**Table 1.** Rindik Song Dataset Details

| Song | Recording No. | Duration (seconds) | Average Duration |
|---|---|---|---|
| 1 | 1 | 52 | 53 s |
| | 2 | 54 | |
| | 3 | 53 | |
| 2 | 1 | 31 | 31.33 s |
| | 2 | 32 | |
| | 3 | 31 | |
| 3 | 1 | 26 | 26 s |
| | 2 | 26 | |
| | 3 | 26 | |
| 4 | 1 | 22 | 22.33 s |
| | 2 | 22 | |
| | 3 | 23 | |

### 2.5. Onset Segmentation

Onset in audio signal is a moment or time right when an acoustic signal started. There are several ways to define onset. One can define it as the start of a note, the moment of an acoustic event start, or as an instant selected to mark a transient extended transient. Transients can be understood as short time intervals during which a significant energy change occurs in the signal [21].

Onset segmentation in this research is used to segmenting each note in the Rindik song. A note in Rindik song is a rod hit or two rods hit simultaneously. This segmentation process is used prior to pitch detection process. To segment each hit of the Rindik rod before processing, onset detection is used. The offset of a note will be the next note onset.

A segment of an onset segmentation result is considered one note in a Rindik song. The segment will be started from the onset of the note to the offset of the note (onset for the next note).

Each segment will be stored to be processed in the next stage of the system which is the transcription process.

Onset detection in this study is carried out by one of the modules from the Librosa library, onset_detect. This library first works by calculating the onset_strength which is the thresholded spectral flux of a spectogram and returns a one-dimensional array representing the amount of spectral energy that increases with each frame. After getting the onset_strength of the signal, onset_detect will look for the peaks of the onset_strength to be used as the onset time [17].

The onset detection in this study will be used to determine the time when a Rindik tone begins. From that time it will be used to segment according to the starting tone until the next tone starts. The segment will be analyzed to find out the Rindik tone that sounds at that time.

### 2.6. Transcription Using Spectral Subtraction with Iterative Estimation and Separation Approach

There are several tone detection algorithms that can be used to detect the tone of traditional Balinese musical instruments. The tone detection algorithms include ZCR (Zero Crossing Rate), HPS (Harmonic Product Spectrum), and FFT (Fast Fourier Transform) [22]. The algorithm is used to detect tones for musical instruments made of animal skins, metal, and bamboo. The algorithm that will be used in this research is the FFT algorithm.

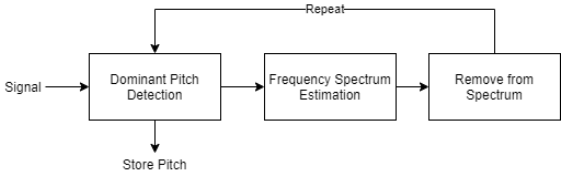


**Fig. 7.** Iterative estimation and separation flowchart

To detect more than one pitch, another approach is needed. One of these approaches is the iterative estimation and separation approach to detect more than one tone (polyphonic) in an audio signal piece [13]. This approach is carried out by finding the first dominant pitch on the spectrum. Once found, the pitch is removed from the spectrum and the next/second dominant pitch is searched again, which is called the separation process [13]. In this separation process, the spectral subtraction method will be used to eliminate partial tones in the process. Figure 7 explains the process for multi-pitch estimation with iterative and separation approach.

Fast fourier transform is an algorithm to calculate discrete fourier transform (DFT) [23]. Discrete Fourier transform is a method used to convert signals from time domain to frequency domain. The FFT algorithm is a fast and efficient algorithm in calculating the discrete Fourier transform [24]. DFT has a complexity of $O(n^2)$, while calculations with FFT only have a complexity of O(n log n) [25]. The equation of the DFT is written in equation (1).

$$Y[k] = \sum_{n=0}^{N-1} X[n]\, e^{-j(2\pi k/N)n} \qquad (1)$$

Y[k] is the result spectrum of a frequency response k. X[n] is the signal input, n is signal sample index, and N is the total obeserved frame length. DFT is a function of k in the frequency domain. Each function of k is an impulse function with a magnitude, frequency, and phase shift [1]. In this study an iterative and separation estimation approach will be used and the Figure 8 and 9 is a flowchart of the implementation of the approach.

#### 2.6.1. Implementation of The Approach

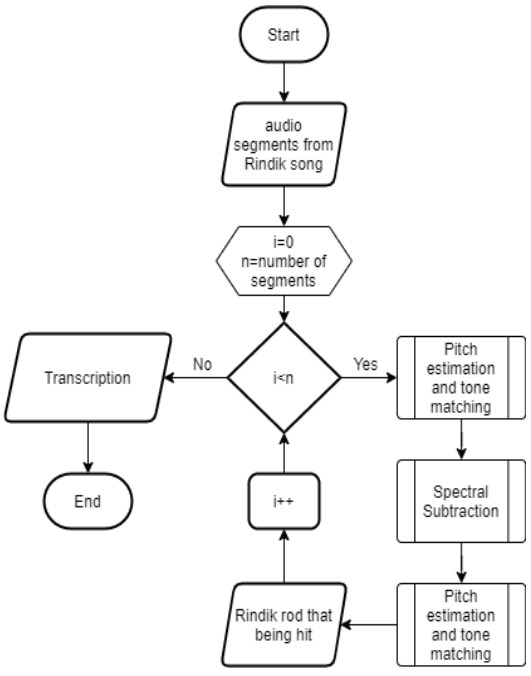


**Fig. 8.** Transcription with iterative estimation and separation process

Figure 8 shows how the iterative process being carried out with spectral subtraction process in between 2 pitch estimation and matching process. Only 2 process of pitch estimation needed to predict an audio segment. Figure 9 shows the pitch estimation and tone matching process. On

which the unseparated audio segment and the separated audio segment being treated with different process. The unseparated audio segment being pitch estimated and tone matched with the nearest frequency. While the separated audio being pitch estimated and tone matched with the difference of frequency in the interval of the tolerance value to know if an audio segment is 2 rods sound or only a single rod sound.

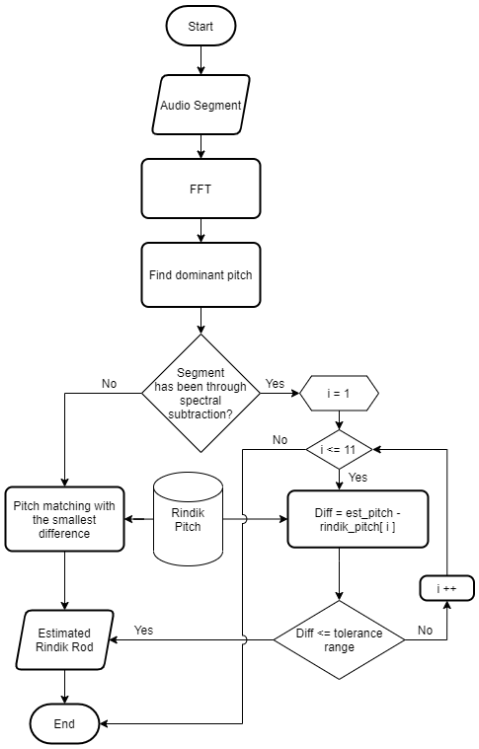


**Fig. 9.** Pitch estimation and tone matching process

### 2.6.2. Spectral Subtraction

Spectral subtraction algorithm applies if the source signal is independent of noise [26]. First, framing and windowing are carried out, then calculate the spectrum in the frequency domain and calculate the average magnitude of the noise spectrum, subtract the source signal with the average magnitude, half-wave rectification to change a negative value to 0, and Inverse FFT to change the sound signal from the frequency domain into the time domain. Figure 10 shows the process of a spectral subtraction that carried out to do the separation process the approach .

Because spectral subtraction works in the frequency domain, a Fourier transform is performed first with a fast Fourier transform (FFT) for each frame. After obtaining the frequency domain of the signal, the average magnitude of the noise spectrum will be calculated and squared [1] with the (2) formula.

$$\mu(k) = E\{|N[k]|\}^2 \qquad (2)$$

Where E is the mean value operator. After obtaining the average magnitude of the noise spectrum, it is necessary to reduce the average magnitude of the source spectrum with the (3) formula.

$$\tilde{S}[k] = \left|\tilde{X}[k]\right| - \mu(k) \qquad (3)$$

In certain cases, for each frequency, the average magnitude of the noise spectrum is greater than the magnitude of the source signal spectrum. This creates a negative value on the spectrum. Then half-wave rectification is performed to replace the negative value with a value of 0. Half wave rectification equation is shown on the (4) equation. $\tilde{S}[k]$ is the input signal, $\hat{S}[k]$ is the output of half wave rectification result, and $\mu(k)$ is the average magnitude of the noise.

$$\hat{S}[k] = \begin{cases} 0, & \tilde{S}[k] < \mu(k) \\ \tilde{S}[k], & \tilde{S}[k] \geq \mu(k) \end{cases} \qquad (4)$$

Conversion from frequency domain to time domain using IFFT of spectrum per frame.

$$\hat{S}[n] = \frac{1}{N} \sum_{k=1}^{N} \hat{S}[k]\, e^{j\theta_Y(k)} e^{j(2\pi k/N)n} \qquad (5)$$

Where $\theta_Y(k)$ is the phase of the audio signal at frequency Y(k) and j is an imaginary number.

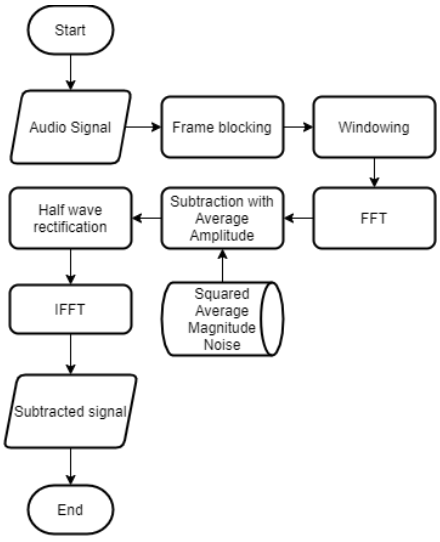


**Fig. 10.** Spectral subtraction process

The spectral subtraction process in this study is used in the separation process with an iterative estimation and separation approach. After the first dominant Rindik rod sound is recognized, it will be removed using spectral subtraction. So that if there is a next rod sound, it can be recognized again by using a pitch estimation.

### 2.7. Transcription Using Discrete Wavelet Transform

Wavelet transform is a technique used to calculate signals with shifts and stretches of the mother wavelet [27]. A wavelet is a short wave whose energy is concentrated at short intervals of time. One of the techniques used for wavelet transformation is to view the wavelet transform as a filter bank. This method is called the subband coding method where the signal is passed through the filter bank. The signal will be decomposed into detail coefficients and approximation coefficients by a series of high-pass filters, low-pass filters and downsampling. Figure 11 is an illustration of the subband coding method with a filter bank used in the wavelet transformation process.

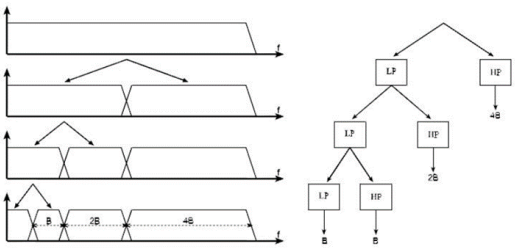


**Fig. 11.** Filter bank on subband coding

In this study, the Symlet Wavelet function will be used in the wavelet transformation process. The Symlet Wavelet function can be seen in Figure 12 [28].

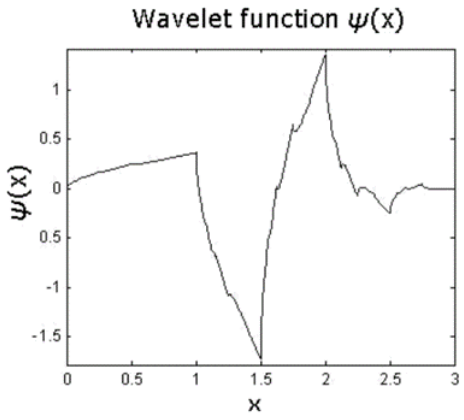


**Fig. 12.** Symlet wavelet function

The filter coefficients corresponding to the Symlet Wavelet function in Figure 11 are shown in Table 2 [28].

**Table 2.** Symlet Wavelet Decomposition Filter Coefficient

| Low-pass Filter Decomposition | High-pass Filter Decomposition |
|---|---|
| -0.1294 | -0.4830 |
| 0.2241 | 0.8365 |
| 0.8365 | -0.2241 |
| 0.4830 | -0.1294 |

In this study, the wavelet transformation will be applied to the spectrum of the FFT results for each frame to be matched with the results of the reference data wavelet transformation. Matching is done by calculating the difference between each pair of vector components resulting from the Wavelet transform. The smallest difference is the candidate of Rindik rods matching result.

## 3. Result and Discussion

This section discussed about the result of the experiment conducted. The experiment consists of testing the system with all the parameters. Accuracy of the method were measured by calculating the percentage of correct estimation of each segment in a Rindik song. As an example if there are 100 segments in a Rindik song and 80 segments were correctly estimated, the accuracy of the system will be 80%. With the formula of accuracy in equation (6).

$$acc = \frac{number\ of\ correct\ estimation}{number\ of\ segment} \quad (6)$$

Each segment represents a note in a Rindik song. To decide if a segment is estimated correctly or not, the label of the segments were compared to the estimated label from the transcription process.

The label is stored in string format for example “3-8” is the third and eighth rod voice. When the predicted label produced “8” label or “3” label only, the system will decide that the predicted label is wrong. The system need exact “3-8” label to have correct prediction count.

### 3.1 Application Interface

The system applcation is implemented with Python programming with its supported library.

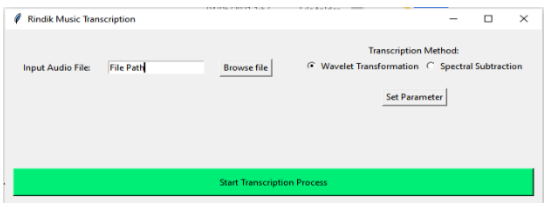


**Fig. 13.** Symlet wavelet function

Figure 14 shows the interface of the transcription. To help Rindik players learn the song on the transcript, the speed of the transcription can be slowed or accelerated as desired by pressing the "+" button to speed up and "-" to slow down. By slowing down the display, players can learn the Rindik song slowly, not at the speed of the original Rindik song. The transcription results can also be saved in the form of a PDF file by pressing the "Save" button in the upper right corner.


**Fig. 14.** Symlet wavelet function

## 3.2 Result of Transcription with Wavelet Transformation

The system testing using the Wavelet transformation method was carried out by calculating the transcription accuracy of each Rindik song. Several frame length parameters used in the Wavelet transform method were tested in this study.

The frame lengths used are 512, 1024, and 2048. The transcription accuracy of each Rindik song was recorded 3 times and the results are shown in the table 3.

**Table 3.** Accuracy result of transcription with wavelet transformation

| Data | Accuracy (%) | | |
|---|---|---|---|
| Frame Length | 512 | 1024 | 2048 |
| Song 1-1 | 51.63 | 67.39 | 75 |
| Song 1-2 | 60.73 | 76.96 | 86.91 |
| Song 1-3 | 40.43 | 61.17 | 74.47 |
| Song 2-1 | 68 | 84 | 90.67 |
| Song 2-2 | 40.74 | 74.07 | 83.95 |
| Song 2-3 | 38.75 | 66.25 | 78.75 |
| Song 3-1 | 63.24 | 76.47 | 86.76 |
| Song 3-2 | 54.41 | 73.53 | 79.41 |
| Song 3-3 | 52.11 | 73.24 | 78.87 |
| Song 4-1 | 62.96 | 77.78 | 81.48 |
| Song 4-2 | 72.22 | 85.19 | 90.74 |
| Song 4-3 | 82.35 | 94.12 | 94.12 |
| Average | 57.29 | 75.84 | **83.42** |

Accuracy is measured by counting the number of correct guesses from the transcription results. From the results in table 3, it can be seen that the highest transcription accuracy of each song is obtained by using a frame length of 2048. An average accuracy of 83.42% is obtained for transcription using the Wavelet transform method using a frame length of 2048. While transcription using a frame length of 512 does not have good accuracy results in transcribing Rindik song, where the average accuracy was only 57.29%.

## 3.3 Result of Transcription with Spectral Subtraction

The system testing using the spectral subtraction method was carried out by calculating the transcription accuracy of each Rindik song with predetermined parameters. There are several parameters of the tolerance interval of the frequency difference used in the spectral subtraction method tested in this study. The tolerance interval used are 0.55 Hz, 1.10 Hz, and 1.66 Hz. The results of the transcription accuracy of each Rindik song were observed which were recorded 3 times and the results are shown in the table 4.

**Table 4.** Accuracy result of transcription with spectral subtraction

| Data | Accuracy (%) | | |
|---|---|---|---|
| Tol. Interval | 0.55 Hz | 1.10 Hz | 1.66 Hz |
| Song 1-1 | 35.33 | 58.15 | 64.13 |
| Song 1-2 | 39.27 | 51.31 | 60.73 |
| Song 1-3 | 51.06 | 60.11 | 62.23 |
| Song 2-1 | 76 | 81.33 | 90.67 |
| Song 2-2 | 66.67 | 81.48 | 81.48 |
| Song 2-3 | 70 | 81.25 | 81.25 |
| Song 3-1 | 51.47 | 69.12 | 83.82 |
| Song 3-2 | 35.29 | 72.06 | 85.29 |
| Song 3-3 | 35.21 | 69.01 | 77.46 |

| | | | |
|---|---|---|---|
| Song 4-1 | 35.33 | 58.15 | 85.19 |
| Song 4-2 | 39.27 | 51.31 | 77.78 |
| Song 4-3 | 51.06 | 60.11 | 92.16 |
| Average | 35.21 | 69.01 | **78.51** |

Accuracy is measured by counting the number of correct guesses from the transcription results. From the results in table 4, it can be seen that the highest transcription accuracy of each song is obtained by using a tolerance interval of 1.66 Hz frequency difference. An average accuracy of 78.51% was obtained for transcription using the spectral subtraction method using a tolerance interval of 1.66 Hz frequency difference. The highest average accuracy is obtained when using a tolerance range of 1.66 Hz frequency difference, but the accuracy of Rindik 2-2 and 2-3 songs has the same transcription accuracy both with a tolerance range of 1.10 Hz and 1.66 Hz.

The result of Song 4-2 and 4-3 being observed specifically. The accuracy of the Song 4-2 and Song 4-3 have a quite high difference. This is the result of a speed difference of the Rindik player while performing the song. The player has a little bit inconsistency while performing both of the song. But when the result of the Wavelet Transform is observed, they have a slight difference between Song 4-2 and Song 4-3. We assume this happen because the Wavelet Transform method has a slightly better performance on recognizing the one rod voice with a remnant voice of previous segment that is predicted two rods voice with Spectral Subtraction method.

### 3.4 Further Analysis on The Result

After testing the transcription system, there are some estimation errors in the transcription process with both methods. Some of the errors in the transcription results and their causes were observed. Table 5 shows some errors in the method of wavelet transformation and spectral subtraction.

**Table 5.** Result Snippet of Estimation Error in Transcription Process

| Actual Note | Wavelet Transformation Estimation | Spectral Subtraction Estimation |
|---|---|---|
| 8 | 3-8 | 3-8 |
| 7 | 3-7 | 3-7 |
| 1 | 1-2 | 1 |
| 2 | 2 | 2-5 |
| 2 | 2 | 2-7 |

After observing some of the estimation errors from both methods, the error that occurred was caused by the sound of the remnants of the previous rod hit still being heard in the next note segment. For example, in Table 5, the eighth Rindik rod sound are estimated for combination of third and eighth Rindik rod sound because the previous segment had the third Rindik rod sound so that it can still be heard in the next segment. This is expected to have resulted in the emergence of a sufficiently high magnitude of frequency from the previous Rindik rod hit on the spectrum that the Rindik rod estimated so that the estimation was incorrect.

There are also a little amount of another kind of errors which occur in a little amount (~ 1-2 occurrence). One of the errors are two rods voice being predicted as one rod voice. This kind of error occur in the two rods voice with one octave difference. The another higher octave note will have same harmonic pattern with the exact one lower octave. This causes the pattern of the frequency spectrum of a two rods voice looks almost similar with the one rod voice of the lower octave. As the Rindik traditional music uses pentatonic scale note [29], the one octave higher note will be heard in 5 rod apart from the lower octave note [30]. For example, the lower octave note of the sixth rod voice is the first rod voice.

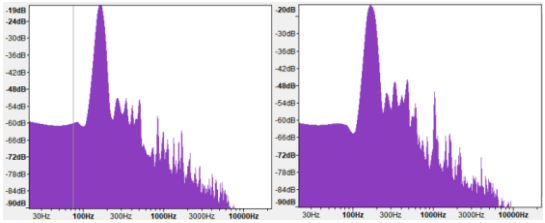

**Figure 15.** Spectrum Plot of Rindik Rod 1 (left) and Rindik Rod 1 & 2 (right)

The other error is misprediction of rod voice i.e. fifth rod voice being predicted as sixth rod voice. This happen in the case when a rod voice is adjacent with the predicted Rindik rod. When observing at a segment in the frequency domain, there is a spectral pattern of a Rindik rod that is similar to that of the another adjacent Rindik rod. Figure 15 shows the spectrum of the first Rindik rod sound segment (estimated sound of Rindik rod 1-2) in the song and the actual Rindik rod 1-2 sound segment spectrum. The spectrum plot is analyzed by Audacity software with a frame length of 2048 and uses a Hamming window. This error can be overcome by using a longer frame size during the FFT process.

The spectrum in Figure 16 is obtained by analyzing the segment with 8192 sample in a frame. The spectrum peaks of Rindik rod 1 and 2

start to look separated in Figure 16, but by using a frame length of 8192 there is a Rindik rod segment in the song which is less than 8192 samples long and also more time and computational cost. Therefore a different approach is needed such as adding a silent signal to the observed segment in order to overcome limitation of the length of the frame on the segment. From these errors, further research is expected to be able to take additional step to reduce the sound from the previous stroke segment that is still present in the next stroke and also use a longer frame size.

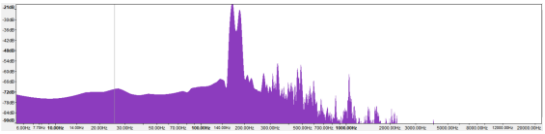

**Figure 16.** Spectrum Plot of Rindik Rod 1 (left) and Rindik Rod 1 & 2 (right)

The transcription process with Wavelet Transformation produces more accuracy than the Spectral Subtraction method. This happened because of the the Spectral Subtraction method only relies with only a scalar feature that obtained from the spectrum which is the highest magnitude pitch. Unlike the Wavelet Transform method, the feature used to estimate are patterns of the Wavelet Transformation result. It should have more accurate result on estimating the Rindik rod sound.

## 4. Conclusion

After testing the transcription system using the wavelet transform and spectral subtraction methods, the accuracy results for each parameter of each method were obtained. The frame length parameter in the wavelet transformation method that produces the highest accuracy is 2048. Meanwhile, the frequency difference tolerance interval parameter in the spectral subtracion method which produces the highest accuracy is 1.66 Hz. The highest accuracy was obtained by transcription using the wavelet transform method. The highest average accuracy in the Wavelet Transform method obtained with a 2048 frame length is 83.42%. The Spectral Subtraction method is less accurate than the Wavelet Transform method because the Spectral Subtraction method only relies on the highest magnitude scalar compared with Wavelet Transform that use a spectrum pattern for matching.

By looking at the estimation errors in the transcription process in both method, there is a sound left over from the previous Rindik rod hit in one audio segment. It is expected in the future that the method able to eliminate or reduce the sound from the previous segment which is the sound of a Rindik rod at that time. In addition, it is also expected a longer frame length but can overcome the limited number. So from these suggestions, it can improve the accuracy of Rindik's song transcription to make it even better.

As a learning tool, this transcription application is considered easier to use on a smartphone device. It is hoped that the Rindik song transcription application can be developed on the Android or iOS platform.

## References

[1] Y. Christian and I. D. M. B. A. Darmawan, “Rindik rod sound separation with spectral subtraction method,” *J. Phys. Conf. Ser.*, vol. 1810, no. 1, 2021, doi: 10.1088/1742-6596/1810/1/012018.

[2] I. P. J. Aristana, I. K. Rinartha, Y. Negara, and I. N. R. Hendrawan, “Aplikasi Permainan Alat Musik Perkusi Tradisional Rindik Bali dengan Augmented Reality Berbasis Android,” *Eksplora Inform.*, 2015.

[3] A. Wiggins and Y. Kim, “Guitar tablature estimation with a convolutional neural network,” 2019.

[4] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, “Automatic Music Transcription: An Overview,” *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 20–30, Jan. 2019, doi: 10.1109/MSP.2018.2869928.

[5] M. O. Faruqe, S. Ahmad, M. A.-M. Hasan, and F. H. Bhuiyan, “Template music transcription for different types of musical instruments,” in *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, Feb. 2010, vol. 5, pp. 737–742, doi: 10.1109/ICCAE.2010.5451347.

[6] R. G. C. Carvalho and P. Smaragdis, “Towards end-to-end polyphonic music transcription: Transforming music audio directly to a score,” in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2017, pp. 151–155, doi: 10.1109/WASPAA.2017.8170013.

[7] E. Benetos and S. Dixon, “Polyphonic music transcription using note onset and offset detection,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 37–40, doi: 10.1109/ICASSP.2011.5946322.

[8] Y.-T. Wu, B. Chen, and L. Su, “Polyphonic Music Transcription with Semantic Segmentation,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 166–170, doi: 10.1109/ICASSP.2019.8682605.

[9] P. Dhara, P. Rengaswamy, and K. S. Rao, “Designing automatic note transcription system for Hindustani classical music,” in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2016, pp. 899–903, doi: 10.1109/ICACCI.2016.7732159.

[10] Y. Wang, Y. Huang, W. Wei, D. Cazau, O. Adam, and Q. Wang, “Automatic Music Transcription dedicated to Chinese Traditional Plucked String Instrument Pipa using Multi-string Probabilistic Latent Component Analysis Models,” in *11th International Conference of Pattern Recognition Systems (ICPRS 2021)*, Mar. 2021, vol. 2021, pp. 223–230, doi:

10.1049/icp.2021.1460.

[11] 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)*, May 2015, pp. 279–284, doi: 10.1109/ISITIA.2015.7219992.

[12] I. G. M. M. Utama Yasa, L. Linawati, and N. Paramaita, "Penentuan Notasi Gamelan Rindik Menggunakan Metode Transformasi Wavelet," *Maj. Ilm. Teknol. Elektro*, 2018, doi: 10.24843/mite.2018.v17i03.p03.

[13] a. P. Klapuri, T. Virtanen, and J. M. Holm, "Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals," *Proc. COST-G6 Conf. Digit. Audio Eff.*, 2000.

[14] M. Karam, H. F. Khazaal, H. Aglan, and C. Cole, "Noise Removal in Speech Processing Using Spectral Subtraction," *J. Signal Inf. Process.*, 2014, doi: 10.4236/jsip.2014.52006.

[15] I. D. M. B. A. Darmawan, "PERBANDINGAN METODE ZCR DAN AUTOCORRELATION UNTUK MENGHITUNG FREKUENSI PADA GAMBELAN GENDER WAYANG," *J. Ilmu Komput.*, vol. 8, pp. 1–6, 2015.

[16] I. M. Widiartha and A. Muliantara, "Rindik Voice Synthesis Using Modified Frequency Modulation as Bali Cultural Preservation Efforts," *Kursor*, 2017, doi: 10.28961/kursor.v8i3.90.

[17] B. McFee *et al.*, "librosa: Audio and Music Signal Analysis in Python," 2015, doi: 10.25080/majora-7b98e3ed-003.

[18] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*. 2020, doi: 10.1038/s41586-020-2649-2.

[19] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A Python package for wavelet analysis," *J. Open Source Softw.*, 2019, doi: 10.21105/joss.01237.

[20] A. Sweigart, *Making Games with Python & Pygame*. 2012.

[21] A. Klapuri and M. Davy, *Signal processing methods for music transcription*. New York: Springer, 2006.

[22] I. P. B. W. Brata and I. D. M. B. A. Darmawan, "Comparative study of pitch detection algorithm to detect traditional Balinese music tones with various raw materials," 2021, doi: 10.1088/1742-6596/1722/1/012071.

[23] R. Dianputra, P. Diyah, and E. Ernawati, "Implementasi Algoritma Fast Fourier Transform Untuk Pengolahan Sinyal Digital Pada Tuning Gitar Dengan Open String," *J. Teknol. Inf.*, 2015.

[24] S. Riyanto, A. Purwanto, and Supardi, "Algoritma Fast Fourier Transform (FFT) Decimation In Time (DIT) dengan Resolusi 1/10 Hertz," *Semin. Nas. Penelitian, Pendidikan, dan Penerapan MIPA*, 2009.

[25] R. Alfina, I. Arifianto, D. Astharini, and P. Wulandari, "Mendisain GUI Untuk Menampilkan Nilai FFT dan IFFT Menggunakan LabVIEW," *TESLA J. Tek. Elektro*, 2019, doi: 10.24912/tesla.v21i1.3250.

[26] D. Cao, Z. Chen, and X. Gao, "Research on noise reduction algorithm based on combination of LMS filter and spectral subtraction," *J. Inf. Process. Syst.*, 2019, doi: 10.3745/JIPS.04.0123.

[27] I. G. A. Wibawa, "PENDUGAAN NILAI REFLECTANCE MENGGUNAKAN TRANSFORMASI WAVELET UNTUK MENENTUKAN USIA DAN KANDUNGAN PIGMEN DAUN JATI BELANDA," Institut Pertanian Bogor, 2015.

[28] A. Glowacz, "Diagnostics of Direct Current machine based on analysis of acoustic signals with the use of symlet wavelet transform and modified classifier based on words," *Eksploat. i Niezawodn.*, 2014.

[29] K. Stepputat, "Nice 'n' easy - The Balinese gamelan rindik: Its music, musicians, and value as tourist art," *Asian Music*. 2006, doi: 10.1353/amu.2007.0012.

[30] D. Wu, C. Y. Li, and D. Z. Yao, "An ensemble with the chinese pentatonic scale using electroencephalogram from both hemispheres," *Neurosci. Bull.*, 2013, doi: 10.1007/s12264-013-1334-y.