Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 11, No. 2, Desember 2021

ISSN 0216 – 0544
e-ISSN 2301– 6914

NEURAL NETWORK BACKPROPAGATION FOR KENDANG TUNGGAL TONE CLASSIFICATION

[a] I Putu Bayu Wira Brata, [b] I Dewa Made Bayu Atmaja Darmawan

[a, b] Computer Science Department, Faculty of Math and Science, Udayana University,
Bukit Jimbaran, Bali, Indonesia-803611, Tel: 081237787141
E-mail: bayuwirabrata@student.unud.ac.id, dewabayu@unud.ac.id

Abstract

Kendang Bali is one of the instruments incorporated in this karawitan art. Balinese kendang can be played alone, called a kendang tunggal, where this type of game has a high level of difficulty understanding the tone of the Balinese drums played because some variations of the tone have similar sounds to other tones. Knowing the tone that is in the kendang song automatically can make it easier to learn it. The first approach method used to classify the tone of a kendang tunggal song is segmentation. The onset detection method is used to segment a kendang song with a variation of the hop size parameter. The segmented tone of the punch will be classified using the Backpropagation method. Feature values of autocorrelation, ZCR, STE, RMSE, Spectral Contrast, MFCC, and Mel spectrogram will be used in the classification process. This study performed variations in hop size values in onset detection and obtained the proper configuration at a value of 110. The addition of the normalization process to the onset detection method also helps the segmentation process of kendang *songs correctly. The optimal backpropagation architecture obtained is learning rate 0.9, neuron hidden layer 10, and epoch 2000 produces an accuracy of 60.92%.*

Key words: Onset Detection, Neural Network Backpropagation, Kendang tunggal.

INTRODUCTION

The *Karawitan* art is one of the many diverse arts in Bali, Indonesia, recognized by the world. In this art, there is a unique musical instrument played by only one person called the *Kendang Tunggal*. *Kendang Tunggal* is a drum played only by a player with a very free pattern and pattern of strokes. If we want to learn this musical instrument, there are many things we have to learn, such as following basic patterns independently through professional playing from video or audio recordings [1]. However, it is not easy to do this if studying a *kendang tunggal* independently or practicing following a professional's game pattern from video or audio recordings because Balinese kendangs have a variety of tones with a sound color like other tones. This way of learning is certainly not very effective because not everyone has the same perspective in receiving voices. It could be that when we listen to the video or audio, there is a mishearing of the tone being played. Here the role of a machine learning model that can classify drumbeats into a notation form will help facilitate the learning.

Musical notation is a written documentation method of a song that stores all information about how music is played [2]. In this study, notation represented the kendang tone class at the classification stage. The first approach that can be taken in the classification of kendang tones is segmentation. Segmentation aims to get a constructive tone from a kendang song. The method used in segmentation in this study is onset detection.

Research on the onset of detection in audio has been conducted by [3] using the Short-time Fourier Transform (STFT) method to detect hit instruments, especially gamelan. STFT method can detect the onset of gamelan audio well by normalizing statistics against the onset reduction function. In addition, this research will also see if we eliminate the normalization process in the feature extraction process and combine it with other features. [4] researching different hop size widths in STFT resulted in a hop size usage of 220, resulting in an average f-measure of 76.37%. In research conducted by [5] and [6] performed calculations of onset strength operations with spectrum flux features through spectrograms without normalization, while [7] using the STFT method obtained satisfactory results by performing energy-weighted band splitting separation was able to increase recall from 85% to 94%.

This study will use various feature extraction methods for the extraction of kendang sound characteristics. Pitch is a subjective factor that involves a person's perception of the frequency with which a piece of music is composed. [8] obtained an estimated pitch accuracy of beating heart sounds with an Autocorrelation method of 75.00%. In comparison, [9] obtained 82.5% accuracy to detect gamelan sounds with the Classification of Neural Network Backpropagation combined with ZCR and STE feature features with a value of 10000 epoch and 100 Hidden Layer. In addition to the pitch feature, the MFCC and Mel Spectrogram features as used by [10] can represent the characteristics of the spectrum in more detail, and the Spectral Contrast used [11] will be used in this study as an interpretation of the low-level audio feature on kendang tones.

In this study, we proposed an amalgamation of many features of the Neural Network model that have not been carried out in previous studies. These audio features are combined with Overlapped STFT with Mel spectrogram, Autocorrelation, Zero Crossing Rate, Short Time Energy, Spectral Contrast, MFCC, and Root Mean Square Energy without changing the audio feature extraction method. However, this research will adjust to the size of the frame and windowing used. The method is used to generate the model has four stages: segmentation onset detection, voice feature extraction, classification, and matching. The results of this study are expected to bridge further research to build a broader tone transcription system and see how the combination of features from various points of view affects the classification results.

RESEARCH METHOD

System Overview

This research has two main stages, specifically the training process and the testing process. The training process aims to create a classification model of the backpropagation used during the testing process. In the process of testing, songs kendang data will continue segmented using the onset detection method and run into the extraction process features with autocorrelation method, Zero Crossing Rate, Short Time Energy, Spectral Contrast, MFCC, and Root Mean Square Energy. The feature will be used in the training of backpropagation. The model applied in the testing process is the result of the training process. Figure 1 shows the flowchart process that occurred in this research

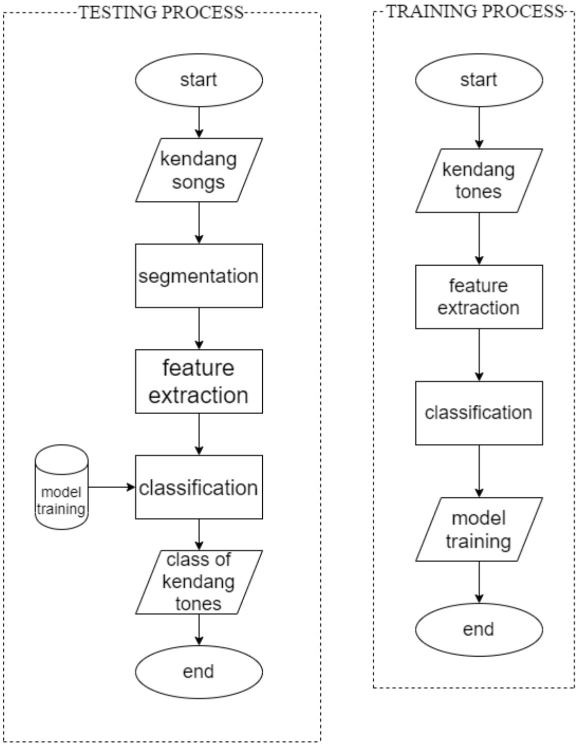

Fig 1. System overview

Kendang Tunggal

Kendang tunggal is a type of kendang game with a primary pattern and motive that is not bound by specific rules [12]. The *kendang tunggal* has six notes that sound in the right and left hand. If it is likened to the pronunciation of the voice, then the tone of the voice in the *Kendang tunggal* is *Pak*, *Pung*, *Teng* for the Left hand and *Cung*, *De*, *Tek* for the Right hand.

Pupuh kekendangan is a combination of several kendang tones to produce a pattern of *kekendangan*. *Pupuh Kakendangan* linguistically (etymology) comes from the word *pupuh* and *kakendangan*. *Pupuh* is a section of *karawitan* art that relates the human voice as a medium of expression. *Pupuh* is a kind of singing that is not based on gong gending but based on "*pada lingsa*." *Kakendangan* comes from the word kendang, which means Balinese gamelan tool made of wood and glass shaped [13]. According to experts about Balinese kendangs, *kendang tunggal* tones can classify the tone based on beginner and advanced classes in studying *kendang tunggal*s. In beginner classes, the tone used was only three, while the advanced classes were six. This study tried to detect using six tones.

Table 1. Kendang Tunggal Tones

No.	Symbol	Tones	Explanation
1	C.	*Cung*	Right Hand
2	D.	*De*	Right Hand
3	T.	*Tech*	Right Hand
4	p.	*plaque*	Left Hand
5	u.	*Pung*	Left Hand
6	t.	*Teng*	Left Hand

As seen in Table 1, the tones used are six tones, where each right and the left hand has three tones that are noted based on the algorithm used.

Segmentation with Onset Detection

Onset is a signal condition that undergoes a period of attack. Onset is commonly used to detect notes in a song and is not limited to pitch [4]. Onset helps segment a notation of a song then used for different purposes [6]. n general, the onset detection algorithm consists of a general pattern with preprocessing stages, feature extraction, and the latter is peak retrieval. Preprocessing uses the Short Time Fourier Transform algorithm approach. In this research, a *mel spectrogram* is used as a detection feature and a value for peak retrieval. Min-max normalization process in *mel spectrogram* will be added to help the peak retrieval process.

The mechanism used in STFT is to make non-stationary signals represent stationary signals using repeated window functions. STFT formed like equation one by calculating the Fourier value obtained by repeatedly performing FFT on each input signal frame.

$$S(n,k) = \sum_{wlen=0}^{N-1} x(wlen+nh) \cdot w(wlen) \cdot e^{-j2\pi kwlen} \quad (1)$$

Mel Spectrogram

[14] suggests that spectrograms are a set of FFT values stacked against each other, where

the spectrogram itself is a visual way to represent the loudness of signals, frequencies, and amplitudes. Spectrograms are defined as the square magnitude of the STFT, providing an overview of suitable strength for a given frequency and time. A *Mel Spectrogram* is a spectrogram that is converted into a range of *mel* scales. Humans can not feel the frequency on a linear scale, and then it needs to be changed to the mel scale with the equation:

$$m = 2595\ \log(1 + \frac{1}{f}) \qquad (2)$$

Normalization of Min-Max

The normalization process uses to facilitate the peak picking process on the *mel spectrogram* feature. This process is constructive to help detect the onset because the kendang's tone produces different signal energy—this study tested how it affects peak retrieval success for onset detection. The process of normalizing min-max counted with the equation:

$$A' = \frac{A - \min value\ of\ A}{\max value\ of\ A - \min value\ of\ A} \qquad (3)$$

Feature Extraction

There are two types of features used, specifically time-domain feature and frequency-domain feature. In the frequency domain, the change from the time domain to the frequency domain uses the concept of Short-Time Fourier Transform (STFT). The algorithms in these frequency-domains are Spectral Contrast, Mel Spectrogram, and MFCC, while the time-domains are Autocorrelation, Zero Crossing Rate, Short Time, and Root Mean Square Energy.

Autocorrelation

Autocorrelation [15] is an algorithm for estimating the fundamental frequency of sound signals (f0). This method works in the time domain based on the clipping center method, and the product crosses the signal by itself. This method is calculated as in the equation:

$$r'_t(\tau) = \sum_{j=t}^{t+W-1-\tau} x_j x_j + \tau \qquad (4)$$

Zero Crossing Rate

Zero-Crossing Rate (ZCR) measures the number of times a signal in the time domain changes past the 0-value limit on the horizontal axis, in which case it is a different value mark than each sample in the previous sample. Although calculated in the time domain, ZCR describes the amount of high-frequency energy in the signal and, in this case, is called the base frequency or f0. ZCR proved discriminatory for percussion instrument classes [16]. ZCR of domain-time signal x(n) is calculated by equation:

$$\boldsymbol{ZCR}(n) = \frac{1}{2} \sum_{k=t\cdot K}^{(t+1)\cdot K-1} \left| sign(s(k)) - sign(s(k+1)) \right| \qquad (5)$$

Short Time Energy

Short Time Energy (STE) is the energy that occurs in the sound signal segment [17]. STE is a simple and effective classification parameter for voiced and silent segments. Energy can detect the endpoint of the sound so that a fast-sounding kendang sound can easily be extracted. STE is calculated by the equation:

$$E_n = \sum_{m=n-N+1}^{n} [x(m)w(n-m)]^2 \qquad (6)$$

Root Mean Square Energy

Root Mean Square Energy (RMSE) is one of the time domain features obtained from calculating the average amount of energy from each sample in a single frame. RMSE is very well used to detect loudness of sound and obtained with equations [18] :

$$RMS_t = \sqrt{\frac{1}{K} \cdot \sum_{k=t\cdot K}^{(t+1)\cdot K-1} s(k)^2} \qquad (7)$$

Spectral Contrast

Spectral Contrast is a feature that estimates the strength of spectral peaks and signal valleys on each frame. The strength of spectral peaks and signal valleys is estimated with the average value of the difference between spectral peaks and signal valleys. The strength of the spectral peak and the signal valley are calculated by equations 6 and 7. Then Spectral Contrast describes the difference in value from spectral peaks and signal valleys in equation 10 [11]

$$Peaks_k = \log\{\frac{1}{\alpha N}\sum_{i=1}^{\alpha N} x_{k,i} \tag{8}$$

$$Valley_k = \log\{\frac{1}{\alpha N}\sum_{i=1}^{\alpha N} x_{k,N-i+1}\} \tag{9}$$

$$SC_k = Peaks_k - Valley_k \tag{10}$$

MFCC

Mel Frequency Cepstrum (MFC) is a linear cosine transformation representation of the short-time log energy spectrum of signals on a non-linear Mel frequency scale. MFCC (Mel-Frequency Cepstral Coefficient) has become processing [10]. MFCC represents the ear model and can produce voice recognition, especially when using many coefficients. Figure 2 displays the process of extracting the MFCC feature. The first step is to divide the signal into multiple frames, generally using windowing functions with fixed intervals. The goal is to model a small part (generally 20 ms) of the received signal. MFCC value will describe the frequency value that builds the kendang tones signal.

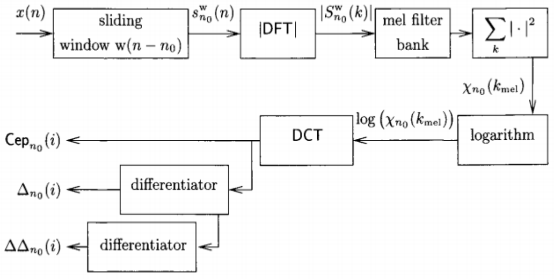

Fig 2. MFCC extraction process

Backpropagation NN

Backpropagation is a systematic method of artificial neural networks using supervised learning algorithms and is commonly used by perceptron's with multiple layer screens to change the weights present in their hidden layers. Backpropagation [19] is a controlled type of training whereby it uses a weight adjustment pattern to achieve a minimum error value between the predicted output and the real output. The activation function of each weight and unit by using the sigmoid activation function in equation 11.

$$f(x) = \frac{1}{1+\exp(-x)} \tag{11}$$

The error change values for each hidden unit and output and weight changes in backpropagation are calculated by equations 12 and 13.

$$\delta_j = O_j(1-O_j)(T_j-O_j) \tag{12}$$

$$W_{ij(t+1)} = W_{ij(t)} + \Delta W_{ji} \tag{13}$$

Dataset

The data used in this research is the primary data separated into two types, namely *kendang tunggal* punch tone data and *Kendang tunggal* Song Data (*Pupuh*). The data used was taken by recording a *kendang tunggal* game from three kendang players. There are three types of kendangs and two types of tempo, namely 65 bpm (medium tempo) and 90 bpm (fast tempo).

Kendang tunggal Punch Tone data is a sample data of kendang base tone repeatedly recorded throughout 1 minute. The total tone data recorded amounted to 3700 tone files from six *Kendang tunggal* punch techniques, consisting of three left-handed tones and three right-handed tones. Data Song (*Pupuh*) Kendang is a *kendang tunggal* game data performed by kendang experts with a combination of predetermined tones. Each kendang expert plays three *pupuh* repeated five times so that the kendang *pupuh* amounts to 90 *pupuh* data.

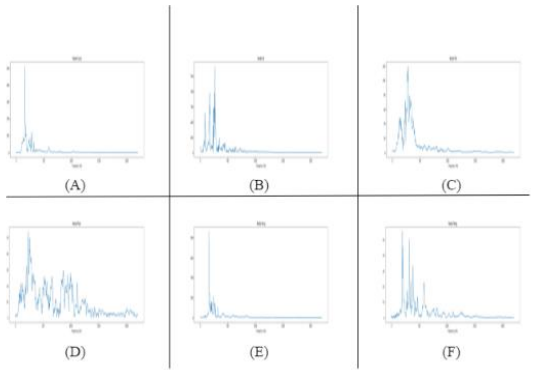

Fig 3. Six-tone spectrum data kendang punch

Figure 3 shows the spectrum of kendang tone data that has been taken i.e., *cung* tone (A), *de* tone (B), *tek* tone (C), *plak* tone (D), *pung* tone (E), *teng* tone (F).

Testing Scenario

The first test occurred to discover the optimal hop size value in detecting the onset of a *kendang tunggal*. Testing the effect of normalization on the *mel spectrogram* feature also coincides with the hop size value experiment. The hop size parameter values tested here are 110, 220, 440, and 512.

The second test observed three parameters of the Backpropagation architecture by determining three main variables that affect the accuracy of kendang tone recognition. The main variables applied for backpropagation architecture testing are the maximum epoch, learning rate, and neurons on the hidden layer. The maximum epoch value is determined from the Mean Square Error (MSE) analysis to obtain a convergent epoch value. Convergent values will show insignificant error changes in each epoch. Three test scenarios were performed on each of these parameters to determine the learning rate and the number of neurons in the hidden layer. Test scenarios for learning rate values used values 0.1, 0.5, and 0.9. The test scenario for the number of neurons used values 5, 10, and 15. The optimal architecture is determined by the test results of the system with parameters that have produced maximum accuracy.

RESULT AND DISCUSSION

The study results found that the detected onset amounted to more than the tone that the three kendang players should play. The kendang player briefly hits the tone of the long punch before hitting the original punch tone and closes the tone of the long punch briefly. The observations found that the players used to do this so that the kendang game is not stammering (*ngandet*). A *kendang tunggal* expert used to call it *nutup* or *mayas*. Figure 4 shows an example of one of the *nutup* and *mayas* that occurred in the kendang *pupuh* tested.

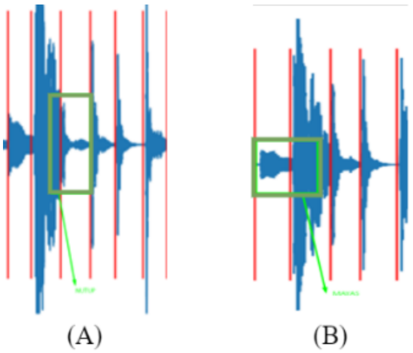

Fig 4. Sample Signal (A) Nutup, (B) Mayas

Determining the threshold value used is difficult to do because *mayas* or *nutup* can have the same height as the original tone, as in figure 5.

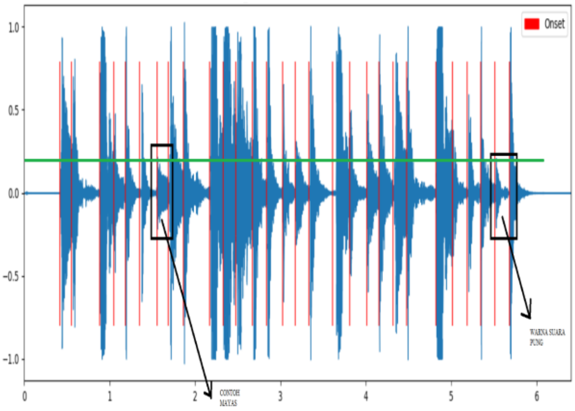

Fig 5. Example of peak mayas equals high with cung tone

Mayas and *Nutup* usually occur on long punch tones, examples of long punch tones are *Cung* tones and De on righthand. Table 2 shows the number of tone segmentations that should be obtained at the time of segmentation.

Table 2. Number of Segmentations on Kendang Song Data

No.	Kendang songs	Expected Segmentation
1	Song 1	27
2	Song 2	24
3	Song 3	23

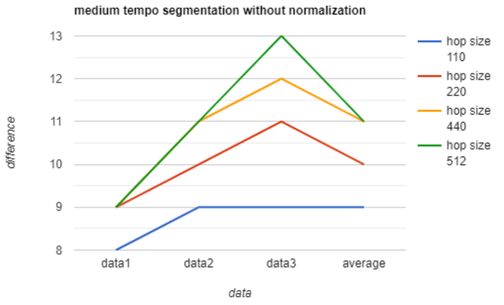

Fig 6. Graph of segmentation differences without Normalization at medium tempo

Segmentation results with onset detection without normalizing the *mel spectrogram* feature get poor results. The number of segments detected notably exceeds the total tone played by the kendang song, and this also happens at a fast tempo, as in figure 7.

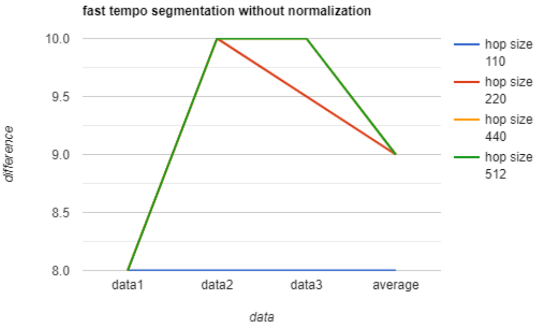

Fig 7. Graph of segmentation differences without normalization at fast tempo

The onset increases due to the detected onset that should be one segment can be detected up to three segments. Noise signals such as wind rustling in Figure 8 can also be detected as an onset.

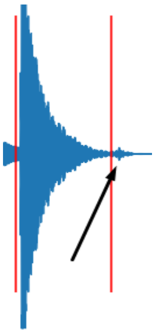

Fig 8. Noise detected as onset

The increase in the number of segments that occur if not normalizing the *mel spectrogram* as much as 8-11 onset pieces.

The observations also obtained small hop size parameters that can detect the onset at the right time. The exact time in the onset is when the signal experiences an Attack period. The difference in onset time with different hop size parameter values can be seen in figure 9.

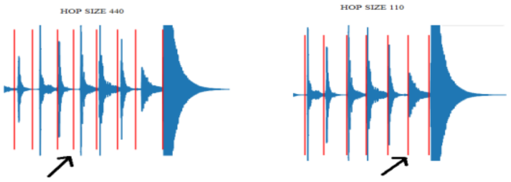

Fig 9. Comparison of hop sizes 440 and 110 in onset detection

It appears that the longer the hop size parameter, the onset of which is detected away from the exact signal period is attacked. Therefore, the parameter used next in the onset segmentation is 110.

The addition of normalized proses performed on the *mel spectrogram* feature results in an excellent average difference in segmenting the kendang song. The subsequent research is to compare the onset detection process by adding to the normalization process. Graphs of onset detection accuracy with normalization can be seen in figures 9 and 10.

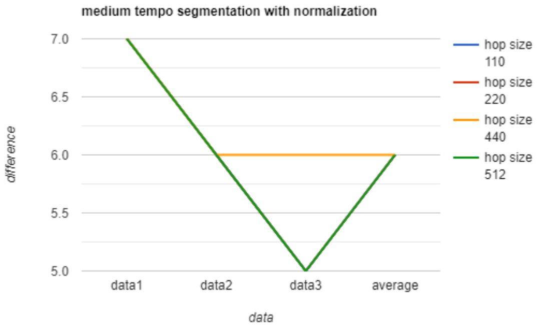

Fig 10. Graph of segmentation differences with normalization at medium tempo

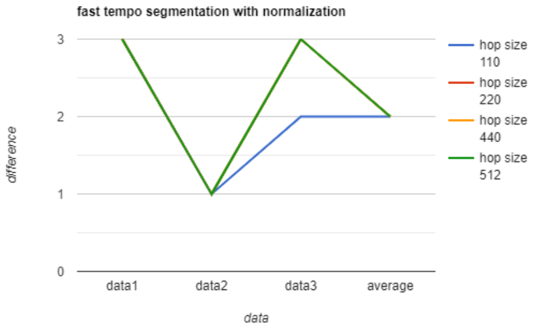

Fig 11. Graph of segmentation differences with normalization at fast tempo

The difference in the number of onsets detected is reduced compared to without normalizing. The addition of the normalization process will help to segment the kendang pupuh well.

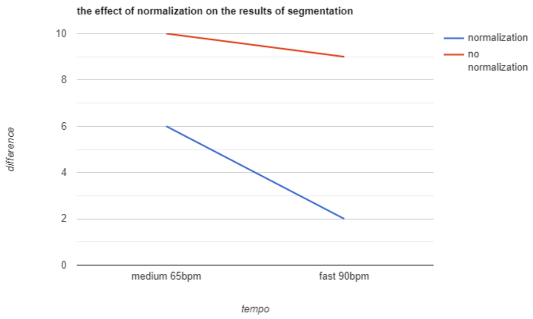

Fig 12. Comparison Graph of differences between normalization and no normalization

Based on the segmentation with onset detection results, the output class of the backpropagation classification will experience additions. Class of kendang punch tones numbering six, namely *cung*, *de*, *tek*, *plak*, *pung*, *teng*, will be added a new class of *mayas* / *nutup*. The new kendang punch tone class can be seen in table 3

Table 3. Backpropagation Output Class

Class	**Symbol**	**Tones**	**Explanation**
1	C.	Cung	Right Hand
2	D.	De	Right Hand
3	T.	Tech	Right Hand
4	P.	plaque	Left Hand
5	U.	Pung	Left Hand
6	T.	Teng	Left Hand
7	^	-	*Mayas/Nutup*

The determination of the maximum epoch is a stage to find a convergent point where the error value that each epoch has obtained does not change. The method used to calculate errors for the backpropagation method in this study was MSE. Figure 13 is an MSE graph showing an error value with a maximum epoch value of 500).

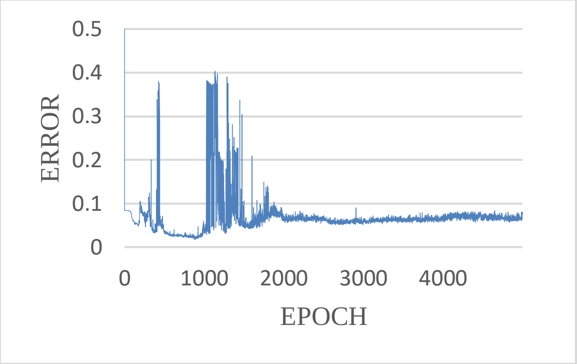

Figure 13 shows the effect of the maximum epoch on the decreased error value. In the epoch to 2000, the error value changes from being at the convergent point to the epoch to 5000. Therefore, the maximum epoch value that is used for subsequent trial scenarios is 2000.

In addition to the number of epoch parameters of the learning rate (a) and the number of neurons in the hidden layer (h), it also needs to be determined to produce a good hit tone recognition. Learning rate is a parameter used in artificial neural networks that affect learning speed to obtain optimal solutions. The optimal learning speed of several test scenarios is not precisely the same because of the initial random value initialization. The initial weight random initialization affects the initial conditions of the learning process to be random as well. The initial weight is used in the training process continuously optimal conditions or maximum epoch are reached.

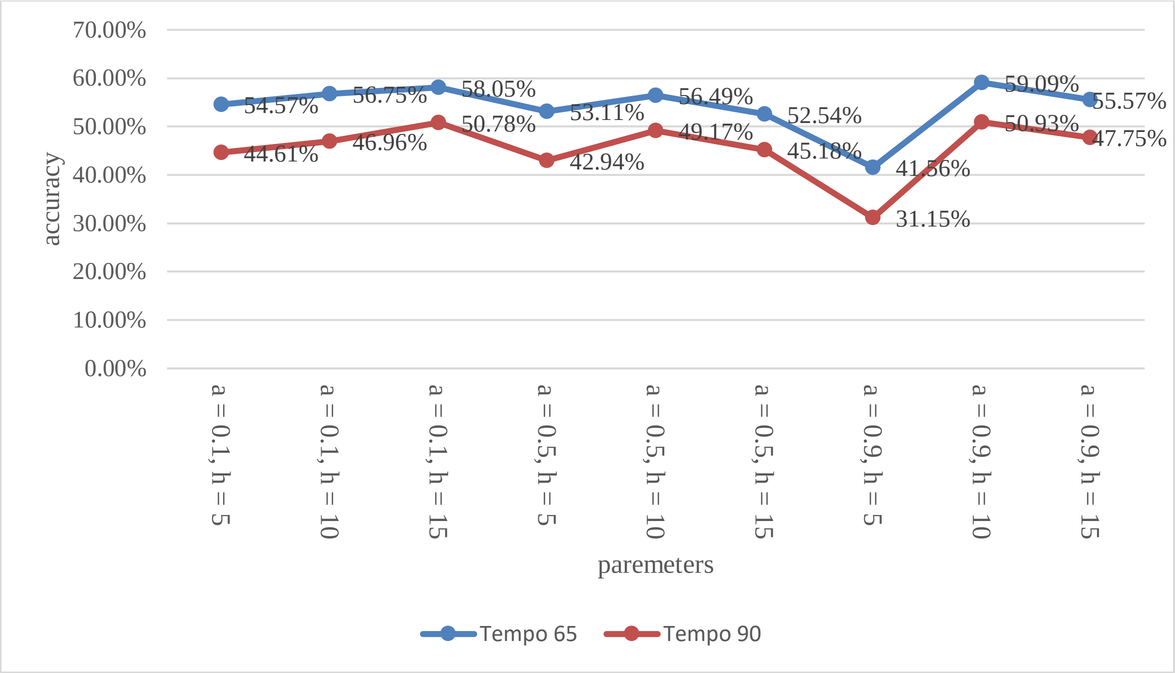

Fig 13. 7-Class test accuracy graph

The test results of the nine system test scenarios are in Figure 14. From the test results in the learning rate scenario 0.1, its accuracy will increase along with the number of neurons used. However, in learning rate scenarios 0.5 and 0.9, the number of neurons with maximum accuracy is in the number of neurons 10. Consequently, the highest accuracy scenario is the number of hidden layer neurons ten and the learning rate 0.9 with an average accuracy of 53.23%. Further testing is performed by eliminating mayas/nutup signals during segmentation and extraction of onset features. Mayas/Nutup will be manually removed from each test data and see how it affects the resulting accuracy.

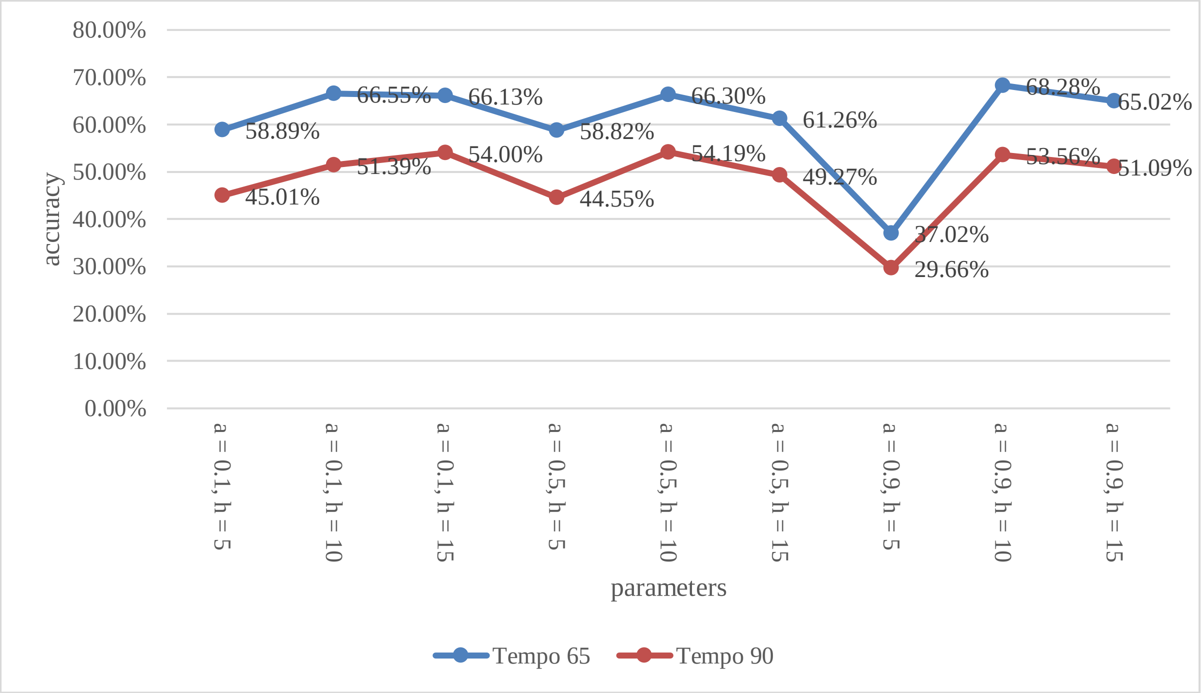

Fig 14. 6-Class test accuracy graph

Figure 15 shows an increase in accuracy across scenarios where the mayas/nutup is removed at the testing process. The parameter test results in this scenario are identical to the seven classes in the previous scenario, where the highest accuracy was obtained with hidden layer ten neurons and a learning rate of 0.9.

Table 4. Significance Test 6 classes and 7 classes

No.	**Scenario**	**Tempo scenario 65**		**Tempo Scenario 90**		**Difference at tempo 65**	**Difference at tempo 90**
		Testing 6 classes	**Testing 7 classes**	**Testing 6 classes**	**Testing 7 classes**		
1	a = 0.1, h = 5	58.89%	54.57%	45.01%	44.61%	4.32%	0.40%
2	a = 0.1, h = 10	66.55%	56.75%	51.39%	46.96%	9.80%	4.43%
3	a = 0.1, h = 15	66.13%	58.05%	54.00%	50.78%	8.08%	3.22%
4	a = 0.5, h = 5	58.82%	53.11%	44.55%	42.94%	5.71%	1.61%
5	a = 0.5, h = 10	66.30%	56.49%	54.19%	49.17%	9.81%	5.02%
6	a = 0.5, h = 15	61.26%	52.54%	49.27%	45.18%	8.72%	4.09%
7	a = 0.9, h = 5	37.02%	41.56%	29.66%	31.15%	-4.54%	-1.49%
8	a = 0.9, h = 10	68.28%	59.09%	53.56%	50.93%	9.19%	2.63%
9	a = 0.9, h = 15	65.02%	55.57%	51.09%	47.75%	9.45%	3.34%
					Mean	6.73%	2.58%
					Standard Deviation	4.64%	2.09%

95% confidence interval

scenario	**Lower Limit**	**Upper Limit**
Tempo 65	3.70%	9.76%
Tempo 90	1.22%	3.95%

Table 4 shows the significance test of the use of 6 classes against seven classes to classify kendang tones. The upper and lower limit values in both tempo scenarios are entirely positive. With a confidence level of 95%, it proves that the scenario with six classes is better than seven classes. The parameters with the highest accuracy for the introduction of *kendang tunggal* punch tones without *mayas*/*nutup* are with a learning rate of 0.9, the number of hidden layer neurons is ten, and the maximum epoch is 2000 resulting in an average accuracy of 60.92%. The architectural results will be used to introduce kendang punch tones at the time of kendang class prediction.

From the optimal architecture tests conducted on 90 kendang pupuh test data with a total of 2220 tones, the system was able to recognize 1352 drum punch tones without mayas /nutup, so that the accuracy of the resulting introduction was 60.92%.

CONCLUSION

The research results get the optimal parameter hop size to detect the exact onset when the signal is undergoing an attack period of 110 with a difference of 6 onset detection from the initial onset. The higher the value of the hop size parameter used, can make the segmentation results less good.

The addition of the normalization process of the *mel spectrogram* feature on onset detection affects the success of onset segmentation. The testing conducted without normalization of the segmentation process will encounter an increase in the number of onsets that should be detected up to 8-1 1 onset.

The neural network architecture of neural backpropagation that obtained the highest accuracy in recognizing drum punch tones was obtained with a learning rate of = 0.9, the number of hidden layer neurons = 10, and the maximum epoch = 2000 with the accuracy obtained was 60.92%.

REFERENCES

[1] I. P. D. Pryatna, I. G. A. Sugiartha, and N. M. Arsiniwati, “Metode mengajar *Kendang tunggal* I Ketut Widianta,” *Kaji. Seni*, vol. 06, no. 01, pp. 25–37, 2019, [Online]. Available: https://doi.org/10.22146/jksks.51868.

[2] A. Suryarasmi and R. Pulungan, “Penyusunan Notasi Musik dengan Menggunakan Onset Detection pada Sinyal Audio,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 7, no. 2, p. 167, 2013, doi: 10.22146/ijccs.3357.

[3] D. P. Wulandari, A. Tjahyanto, and Y. K. Suprapto, “Gamelan music onset detection based on spectral features,” *Telkomnika*, vol. 11, no. 1, pp. 107–118, 2013, doi: 10.12928/TELKOMNIKA.v11i1.521.

[4] E. Amasaro, “Deteksi Onset Pada Beberapa Genre Musik Menggunakan Metode Deviasi Fasa,” 2017.

[5] B. McFee *et al.*, “librosa: Audio and Music Signal Analysis in Python,” *Proc. 14th Python Sci. Conf.*, no. Scipy, pp. 18–24, 2015, doi: 10.25080/majora-7b98e3ed-003.

[6] S. Böck, F. Krebs, and M. Schedl, “Evaluating the online capabilities of onset detection methods,” *Proc. 13th Int. Soc. Music Inf. Retr. Conf. ISMIR 2012*, no. Ismir, pp. 49–54, 2012.

[7] K. Subramani, S. Sridhar, R. Ma, and P. Rao, “Energy-Weighted Multi-Band Novelty Functions for Onset Detection in Piano Music,” *2018 24th Natl. Conf. Commun. NCC 2018*, pp. 1–6, 2019, doi: 10.1109/NCC.2018.8599955.

[8] K. Kolhatkar, M. Kolte, and J. Lele, “Implementation of pitch detection algorithms for pathological voices,” *Proc. Int. Conf. Inven. Comput. Technol. ICICT 2016*, vol. 1, no. September, 2016, doi: 10.1109/INVENTIVE.2016.7823210.

[9] F. Firdausillah *et al.*, “Implementation of Neural Network Backpropagation Using Audio Feature Extraction for Classification of Gamelan Notes,” *Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018*, pp. 570–574, 2018, doi: 10.1109/ISEMANTIC.2018.8549805.

[10] D. Lionel, R. Adipranata, and E. Setyati, “Klasifikasi Genre Musik Menggunakan Metode Deep Learning Convolutional Neural Network dan Mel-Spektrogram,” *J. Infra Petra*, vol. 7, no. 1, pp. 51–55, 2019, [Online]. Available: http://publication.petra.ac.id/index.php/teknik-informatika/article/view/8044.

[11] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, and L. H. Cai, “Music type classification by spectral contrast feature,” *Proc. - 2002 IEEE Int. Conf. Multimed. Expo, ICME 2002*, vol. 1, pp. 113–116, 2002, doi: 10.1109/ICME.2002.1035731.

[12] I. M. Bandem, *Gamelan Bali : di atas panggung sejarah*. Denpasar: STIKOM Bali : Denpasar., 2013, 2013.

[13] I. W. Suweca, “Karawitan Bali Dalam,” *Mudra*, vol. 20, no. 1, 2007.

[14] F. Eyben, *Real-time speech and music classification by large audio feature space extraction*, no. 1975. Springer Science & Business Media, 2016.

[15] A. Von Dem Knesebeck and U. Zölzer, “Comparison of pitch trackers for real-time guitar effects,” in *13th International Conference on Digital Audio Effects, DAFx 2010 Proceedings*, 2010, pp. 1–4.

[16] 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,” *J. Phys. Conf. Ser.*, vol. 1722, p. 012071, 2021, doi: 10.1088/1742-6596/1722/1/012071.

[17] M. Jalil, F. A. Butt, and A. Malik, “Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals,” *2013 Int. Conf. Technol. Adv. Electr. Electron. Comput. Eng. TAEECE 2013*, no. m, pp. 208–212, 2013, doi: 10.1109/TAEECE.2013.6557272.

[18] C. Panagiotakis and G. Tziritas, “A speech/music discriminator based on RMS and zero-crossings,” *IEEE Trans. Multimed.*, vol. 7, no. 1, pp. 155–166, 2005, doi: 10.1109/TMM.2004.840604.

[19] R. Hecht-Nielsen, “Theory of the backpropagation neural network,” *Acad. Press*, pp. 65–93, 1992, doi: 10.1109/ijcnn.1989.118638.

.