








# Traditional Balinese Song Educational Game Application Based on Speech Recognition Using the MFCC-ANN and Its Effect on Cognitive Load

M A Raharja<sup>1,\*</sup>, K A Mogi<sup>2</sup>, I W Supriana<sup>3</sup>, Cokorda Pramatha<sup>4,5</sup> and I G N A C Putra<sup>6</sup>

<sup>1 2</sup> The Department of Informatics, Faculty of Mathematics and Natural Sciences, Udayana University, Badung, Bali – Indonesia

<sup>3 4 6</sup> Computer Science Departement, Udayana University, Indonesia

<sup>5</sup>Center for Interdisciplinary Research on the Humanities and Social Sciences, Udayana University, Indonesia

\*made.agung@unud.ac.id

**Abstract.** Bali's culture is in the form of literary arts which we must preserve. Tembang is a sound art that is built from various tunings and tones as singing materials. Tembang is one part of the literary arts that developed in Balinese society, and is divided into four, namely: Sekar Rare, Sekar Alit, Sekar Madya and Sekar Agung. As time goes by, the existence of tembang is increasingly fading, so learning media are needed that follow the current development of information technology and the existence of song teachers is increasingly difficult because singing songs must comply with the rules that bind the song. There is a need to digitize traditional Balinese songs to preserve their existence among the community. In this research, we developed an educational game application for traditional Balinese songs that can be used practically and theoretically using the Mel-Frequency Cepstrum Coefficients (MFCC) - Artificial Neural Network (ANN) algorithm and examined the cognitive learning load in terms of increasing speed, accuracy and consistency of use. It is hoped that this educational game application for traditional Balinese songs will be a solution where users can learn to sing traditional Balinese songs and find out where the errors are in the notes being sung, which will reduce the cognitive load of students learning. The results of this research showed that the average percentage of success in voice recognition using test data was 80.89%.

**Keywords:** speech recognition, cognitive load, Balinese songs, MFCC, educational games.

## 1. Introduction

Bali is an area that is famous for its arts and culture and has a diversity of tourism potential including natural tourism potential and cultural tourism potential accompa-

nied by the friendliness of the people making Bali a major tourist destination in Indonesia. Culture in the form of songs, dances and paintings which are the heritage and identity of Balinese culture [1]. An effective way to learn to sing pupuh is by hearing examples of how to play it as often as possible and being guided by someone who can play pupuh correctly. Therefore, it is necessary to digitize sekar alit to preserve the existence of sekar alit among the Balinese people [2]. It is hoped that in the future this sekar alit educational game will be a solution where users can learn to sing sekar alit and find out where the errors are in the notes being sung, which will reduce the cognitive load on students learning the sekar alit song [3].

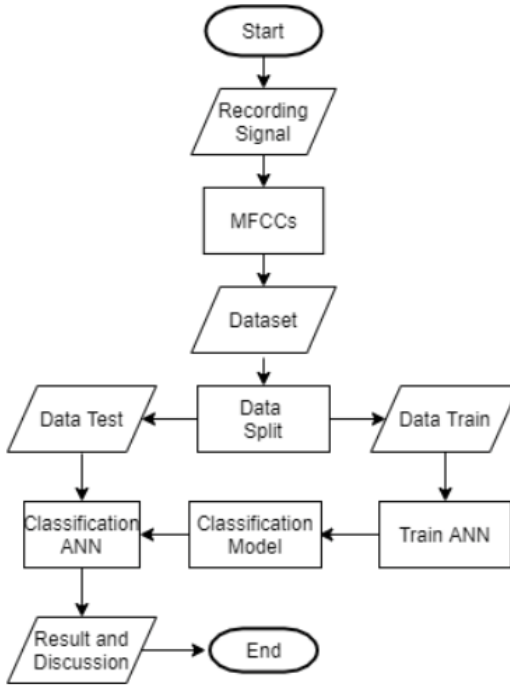
The term cognitive comes from cognition, meaning the process of thinking in the brain, using sensory input that goes to the brain plus information that has been stored as memory. Most motor work is the result of thoughts formed in the brain, namely a process called cognitive control of motor activity. The caudate nucleus plays a major role in the cognitive control of this motor activity [4]. So, it is hoped that using the Tembang Sekar Alit educational game application can reduce cognitive load and improve the cognitive learning outcomes of students who want to learn Tembang Sekar Alit using voice recognition methods or algorithms. One of the methods used for speech recognition is Mel Frequency Cepstral Coefficients (MFCC) and Artificial Neural Network (ANN). MFCC is a method that is widely used in the field of speech technology, both speaker recognition and speech recognition [5]. This method is used to perform feature extraction, a process that converts sound signals into several parameters [6].

From the problems that have been described, the author wants to carry out sound recognition of Tembang Macapat. In this application, a technological approach is used with a song sound recognition method that applies the Mel Frequency Cepstral Coefficients-Artificial Neural Network (MFCC-ANN) algorithm. The MFCC-ANN algorithm consists of two algorithms with different functions [7]. MFCC functions to extract sound signal features. Meanwhile, ANN functions to classify sound signals. MFCC was chosen to extract voice signal features because this algorithm is not too complicated to implement and is the most effective in extracting varying voice signal features and in varying circumstances as well [8]. However, MFCC has a high computing time to extract voice signal features in real-time [9]. Therefore, the role of the ANN algorithm is needed to speed up the processing time that has been taken by MFCC by reducing the size of the voice signal features and classifying the reduced voice signal features [10].

Artificial neural networks (Artificial Neural Networks) are able to recognize activities based on past data. Past data will be studied by artificial neural networks so that they have the ability to provide decisions on data that has never been studied. The main advantage of the Artificial Neural Network (ANN) system is the ability to learn from the examples provided or training data [11].

## 2. Research Methodology

The system is designed to consist of a process to the data split or data separation process. Next, the process is divided into two parts, namely the training process and the testing process, as illustrated in Figure 1.



**Fig. 1.** Research Methodolog

Figure 1. In the first process carried out is to convert the footstep signal using the MFCCs feature extraction by dividing the signal into several frames and changing it into a cepstrum to get the vector value. The results of MFCCs will be labeled and used as a dataset in this study. In the next process, the separation of the dataset is grouped into two parts, namely training data and testing data. In the training process, the system will be trained using training data by applying the ANN algorithm to get the best classification model for speech recognition. Furthermore, after getting the model, then the testing process is carried out using test data by applying ANN classification that have been obtained in the training process. After carrying out these tests it will get the result of the accuracy of the speech recognition system process.

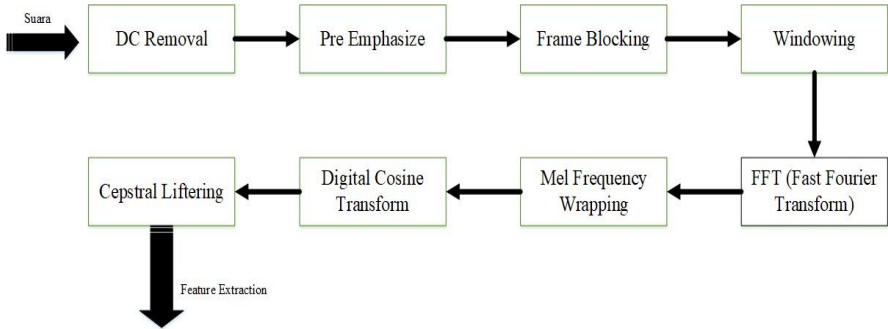
## 2.1 MFCC (Mel Frequency Cepstrum Coefficients)

In performing voice recognition, extraction of features from the voice itself is required. MFCC (Mel Frequency Cepstrum Coefficients) is a method that can be used to extract features from sound which is widely used in the field of speech technology, both speaker recognition and speech recognition [12]. Cepstral coefficient is a feature that is usually used in speech recognition systems. Some of the advantages of this method are :

- a. Able to capture sound characteristics which are very important for speech recognition, or in other words, able to capture important information contained in sound

signals.

- b. Produce as little data as possible, without eliminating the important information it contains.
- c. Replicating the human hearing organ in perceiving sound signals.



**Fig. 2.** Mel Frequency Cepstrum Coefficients Diagram

### 3. Results and Discussion

#### 3.1 Dataset

The research steps began with exploring the traditional Balinese song dataset to understand the distribution of sound classes, duration and other characteristics. After data exploration, a preprocessing stage is carried out to clean and prepare the data, including normalization and handling of missing or invalid values. After data preprocessing, a feature extraction technique was carried out using MFCC. MFCC aims to convert sound signals into a more abstract feature representation, enabling analysis of the frequency and time characteristics of sound. Librosa's mfcc library is used to generate MFCC from audio data in time series form.

#### 3.2 Training and Testing

The next step is to label the data according to the correct sound class and divide the dataset into two parts with a ratio of 80-20, where 80% of the data is used for the training process and 20% for testing. This ensures the model being built will learn from the most data but also be tested on data it has never seen before.

#### 3.3 System Design

In the following sub-chapter, the system design is carried out starting with the process of entering the thresholding process in the form of frame length (N), frame shift length (M), number of MFCC coefficients, and the path where the training data folder is stored. The training data has a data type (.wav) and has been grouped based on wirama. Next, the sound file will go through a feature extraction process. The feature extraction process is a process for extracting the features of each training file contained in the training

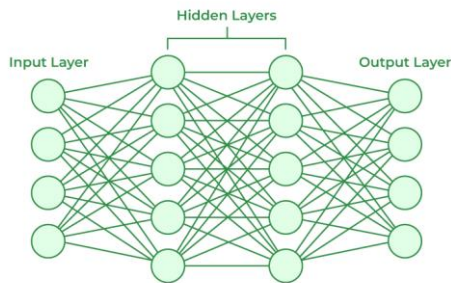
data folder so that the data can go through the next process. The results of the feature extraction process contain a feature vector in the form of a list with size  $N \times K$ , where  $N$  is the number of frames formed and  $K$  is the number of MFCC coefficients determined.

### 3.4 Matching Process Design

The matching process is a process for matching data between expert voice reference data and test voice data. Users are asked to choose what they want to match. After that, the expert voice reference data will be loaded. The reference data stores the expert voice characteristic vector values and the selected threshold range. After the expert voice is loaded, the user can then upload the test voice and what will be extracted. The extraction results of the test voice will be matched with the expert's characteristic vector to determine the resulting DTW distance. After the DTW distance is obtained, the distance will be compared with the threshold range resulting from the range measurement process. The test data will be considered correct if the DTW distance is within the range that has been obtained, and vice versa.

### 3.5 Training and Testing

Voice signals training data will be used to build the model. In this process, ANN will study all training data and optimize the classification process. The ANN classification module uses with the following architecture :



**Fig. 3.** Layer Structure of ANN Classification

1. Input layer : Dense layer with 512 neurons and Using ReLU (Rectified Linear Unit) activation function. Input as many shapes as the number of features in the MFCC data (39 features).
2. Hidden Layers : Dense layer with 256 neurons, Using ReLU activation function, Batch normalization layers and Dropout layer with dropout rate 0.5.
3. Output Layer : Dense layer with 'num\_labels' neurons, Using the Softmax activation function, because it is suitable for classification with classes more than 1, and to make it easy to make predictions or find probabilities.

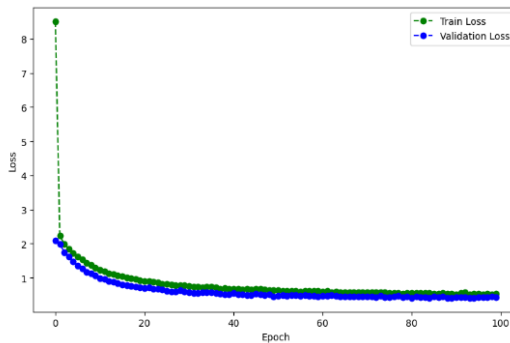


Fig. 4. Model Training Graph.

The results of training and testing the model can be seen in Figure 4. The model succeeded in providing predictions with an accuracy of around 80.89% on the test data. This shows that the model does not experience significant overfitting and can generalize well to data that has never been seen before.

### 3.6 Implementation of the APTASARI Tembang Sekar Alit Educational Game Application

The Sekar Alit song sound matching application uses the Mel Frequency Cepstral Coefficients (MFCC-ANN) method. Android mobile and web-based server side applications are built using the Python programming language. The sub-chapter on the process of creating the Sekar Alit song sound matching application is discussed in four discussion points consisting of the initial appearance of the application, the login process, registration, the main menu which consists of : selecting the Sekar Alit song category, the song input process by the user to be matched, and the application output is a match of the song sound can be seen in Figures 5.

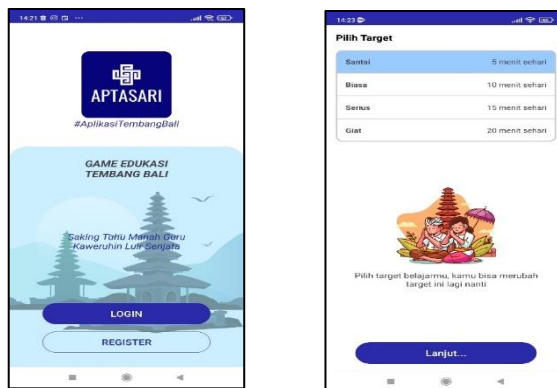


Fig. 5. APTASARI Application Menu Display

### 3.7 Results of Cognitive Load Analysis in Terms of Speed, Accuracy and Constancy

The fatigue of students was also measured objectively using the Bourdon-Wiersma test to determine the effect of using the APTASARI Tembang Sekar Alit educational game application based on voice recognition on Android on students' accuracy, speed and consistency in the Tembang Sekar Alit learning process. Testing of cognitive load in terms of the speed of research subjects was carried out by measuring and analyzing the normality of speed data. Data collection was carried out before learning and after learning ended and was repeated five (5) days for each sample in the Control Group (CG) and Treatment Group (TG). Based on the normality test in Table 3.1, to obtain a mean difference test, a difference test was carried out using an independent-sample t test which can be seen in Table 3.1.

**Table 1.** Differential Test of Cognitive Load Data in Terms of Speed, Accuracy and Constancy

Variables	CG	TG	<i>p</i>
	Mean± SB	Mean ± SB	
Speed:			
Before	8,03±0,99	8,29±1,41	0,618
After Learning	10,85±1,55	9,02±0,79	0,003
Accuracy:			
Before	4,15±1,69	4,49±0,90	0,554
After Learning	6,77±1,71	4,75±0,78	0,002
Constancy:			
Before	2,74±0,83	2,69±0,69	0,708
After Learning	5,09±1,02	3,58±1,38	0,003

Based on Table 3.1, it shows that the mean of KK and KP for cognitive load in terms of speed, accuracy and constancy before work on KK and KP is  $p > 0.05$ , which means the data is comparable, while cognitive load in terms of speed after work on KK and KP is different. significant with  $p < 0.05$ . In other words: (a) the average cognitive load in terms of speed, accuracy and constancy before work in the two groups is not significantly different or comparable so that this condition does not affect the research results; (b) the average cognitive load in terms of speed, accuracy and constancy in KP is lower than the average cognitive load in terms of speed in KK and is significantly different; (c) reduction in cognitive load in terms of speed, accuracy and consistency in the use of the Tembang Sekar Alit educational game application based on voice recognition on Android improving the quality of health and learning of Tembang Sekar Alit in students as seen from the reduction in cognitive load in terms of of speed.

## 4. Conclusion

Based on the research results, the design of the sekar alit song learning application uses the Mel Frequency Cepstral Coefficients (MFCC-ANN) method. From the results of the tests that have been carried out, the following conclusions can be drawn :

Tembang sekar alit educational game application software based on voice recognition on Android APTASARI using the Mel Frequency Cepstral Coefficients-Artificial Neural Network (MFCC-ANN) method can improve health and learning as seen from the reduction in cognitive load in terms of increasing student speed by 8.33%, an increase in student accuracy of 30.05% and an increase in student constancy of 16.91%.

This research utilizes MFCC and ANN to classify the sounds of Sekar Alit songs in educational game applications with a significant level of accuracy. Through experiments on the dataset, this method succeeded in identifying five types of sekar alit songs, although there are still several challenges in correctly classifying several sound classes. The ANN model used shows good generalization capabilities with a final total accuracy of 80.89%. This research makes additional contributions to the fields of speech signal processing and deep learning and opens up opportunities for further development.

## References

- [1] C. Pramatha and J. Davis, "Digital Preservation of Cultural Heritage for Small Institutions," *Digital Cultural Heritage*, pp. 109–117, 2019, doi: 10.1007/978-3-030-15200-0\_8.
- [2] M. A. Raharja, S. Purnawati, I. P. G. Adiatmika, I. N. Adiputra, and I. B. A. Swamardika, "Usability Analysis of Tembang Sekar Alit Learning (SekARAI) Applications Using The Human Computer Interaction (HCI) Model In Bali Students," *Proceedings of the Second Asia Pacific International Conference on Industrial Engineering and Operations Management*, pp. 2870–2879, 2021.
- [3] M. A. Raharja, I. P. G. Adiatmika, I. N. Adiputra, and S. Purnawati, "The Effect of Balinese Traditional Song Learning Software with Artificial Intelligence Based on Total Ergonomic Approach to Reduce Cognitive Load and Fatigue Load For Balinese Students," vol. 12, no. 2, pp. 1104–1111, 2022.
- [4] A. C. Guyton and J. E. Hall, *Fisiologi Kedokteran II*. Jakarta: EGC Buku Kedokteran, 2018.
- [5] I. Gusti *et al.*, "Implementasi Metode Convolutional Neural Network Pada Pengenalan Aksara Bali Berbasis Game Edukasi," *SINTECH (Science and Information Technology)*, vol. 6, no. 1, pp. 1–15, 2023, [Online]. Available: <https://doi.org/10.31598>
- [6] C. Sunitha and E. Chandra, "Speaker Recognition using MFCC and Improved Weighted Vector Quantization Algorithm," *International Journal of Engineering and Technology (IJET)*, vol. 7, pp. 1685–1692, 2015.
- [7] D. S. Mistry and Prof. A. V. Kulkarni, "Overview: Speech Recognition Technology, Mel- frequency Cepstral Coefficients (MFCC), Artificial Neural Network (ANN)," *International Journal of Engineering Research & Technology*, vol. 2, no. 10, pp. 1994–2002, 2013.



- [8] J. Ilmiah and S. Teknika, “Sistem Pengenal Wicara Menggunakan Mel-Frequency Cepstral Coefficient (Speech Recognition System Using Mel-Frequency Cepstral Coefficient),” vol. 20, no. 1, pp. 75–80, 2017.
- [9] C. Asmita, T. Savitha, and K. Upadhya, “Voice Recognition Using MFCC Algorithm,” 2014.
- [10] M. Nabil Aljufri and B. Henryranu Prasetio, “Sistem Deteksi Tingkat Stress Menggunakan Suara dengan Metode Jaringan Saraf Tiruan dan Ekstraksi Fitur MFCC berbasis Raspberry Pi,” vol. 6, no. 11, pp. 5278–5285, 2022.
- [11] M. Gilke, P. Kachare, R. Kothalikar, and V. P. Rodrigues, “MFCC-based Vocal Emotion Recognition Using ANN,” *2012 International Conference on Electronics Engineering and Informatics (ICEEI 2012)*, vol. 49, no. Iceei, pp. 150–154, 2012, doi: 10.7763/IPCSIT.2012.V49.27.
- [12] F. N. Suciani, E. C. Djamal, and R. Ilyas, “Identifikasi Nama Surat Juz Amma dengan Perintah Suara Menggunakan MFCC dan Backpropagation,” *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2018*, pp. 18–23, 2018.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

