

CHAPTER 4

ANALYSIS AND DESIGN

4.1. Analysis

This chapter provides a detailed explanation and discusses one by one about the methods of solving problems that have been mentioned in several points in chapter 1. The main purpose of creating this project is how to implement a virtual assistant using speech recognition in carrying out commands from users. In addition, this project also discusses both methods used as a comparison in solving problems.

From the formulation of the problem that has been outlined, the first is how speech recognition can run the commands spoken by the user. In this case it needs to be discussed how the work of speech recognition itself. As explained below:

1. First the input data is received through a sample dataset that contains voice samples.
2. Once the input data is received the data will be extracted and converted into a form of frequency as well as creating a database for the template.
3. The next stage is the comparison or matching of the input of the received voice with the sound data that has been stored on the template.
4. After making a comparison it will be validated between the input and the available template. Such validation will result in a decision to determine what words the user is saying.

In the next problem formulation mentions about the algorithm. The difference from Deep Neural Network with Convolutional Neural Network if implemented into speech recognition. In the case of speech recognition classification, Deep Neural Network (DNN) is used. Deep Neural Network is an artificial neural network with several hidden layers between the input layer and output layer. DNN has many layers and also has a lot of connections and weights as well. DNN has a hidden layer that can learn higher levels. This DNN is better for modeling acoustics if using a convolutional layer in lowering the weight of nearby frequencies. Unlike DNN, CNN is a nerve that is devoted to processing data. It has a grid structure. For example, it is an image of two dimension. The name convolution is an operation

of linear algebra that multiplies matrix of the filter on the image to be processed. This process is called layers. Convolutional and is one of the many layers that can be had in a network. However, this convolution layer is a layer. The most important thing to use.

Last problem formulation mentions about the success rate from both algorithms. The design of the speech recognition system goes through two phases. Training phase and testing phase. In the training phase, the system will receive input in the form of samples that it will be used as training data. Training data will be stored in the database and will be used as reference in the testing phase. The next phase is the testing phase. In this phase the system will be tested with how to enter a sample and compare with existing training data, then decided the output based on similarities from training data. The two main modules needed in the design speech recognition, namely:

1. Feature Extraction

Characteristic extraction is the process of converting sound signals into several parameters, information it gets lower because it will eliminate some of the less important information. Without changing the real meaning.

2. Pattern Matching

In matching features will be compared or match data from input signals with training data already in the database. The result of matching this feature will be system output.

4.2. Desain

The use of flowchart aims to know the processes or procedures of a program that makes it easier to understand the program to be built. Flowchart system can be described as next:

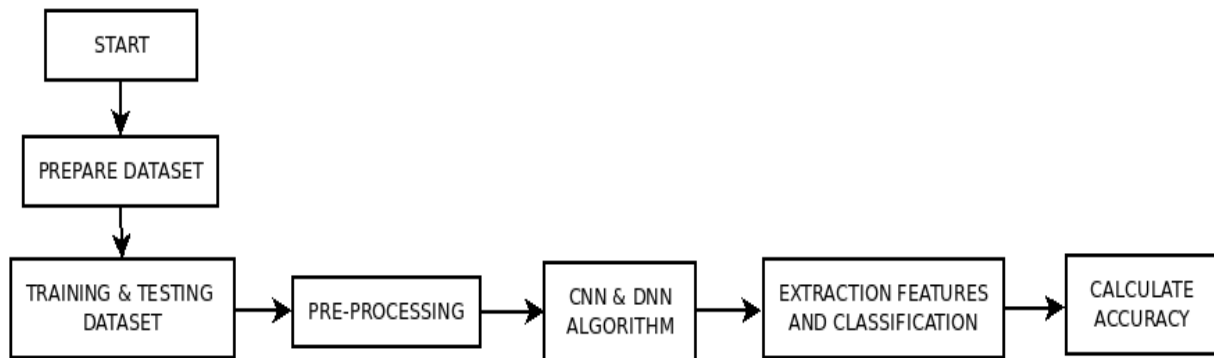


Figure 4.1: Flowchart of Speech Recognition Design

Flowchart 4.1 is used as an overview or flow of the system itself. Starting from preparing the dataset, then the dataset that has been prepared will go into the pre-processing stage where the dataset will be trained and tested using CNN and DNN algorithms. After that, the system does not directly execute the command but the system will start with the extraction of characteristics and classification simultaneously so that the spoken sentence can be understood by the system, after that the command is executed as the user says.

In general, speech recognition systems recognize and understand the word and matching the digital signal with a certain pattern stored in the system database. The stages in the process are:

1. **Prepare Dataset** : dataset contains a sample recording of human voices.
2. **Training & Testing Dataset** : at this stage, the dataset is divided into training, validation, and test sets using a 80:10:10 ratio.
3. **Pre-processing** : prepare and process the initial data so that the data used is data that is ready to use and can facilitate the processes in the next stage.
4. **DNN & CNN** : at this stage, after the dataset has gone through the pre-processing process, then the data that is ready will be reprocessed using DNN and CNN algorithms. In this process, the algorithms used play a role in maximizing the dataset and determining the accuracy level of the program itself.
5. **Extraction Features & Classification** : at this stage, extraction of characteristics from each data and store data in the form of sound as well as the creation of a database as a pattern. After that, the system will match the new data with the training data that is in the available database. Starting with the process of digital signal conversion resulting from

the extraction process into the form of sound spectrum that will be analyzed by clouding it with sound patterns. Previously input voice data was separated and processed one by one based the order.

6. **Calculate Accuracy** : calculate the accuracy of each algorithm and the results are compared between DNN and CNN algorithms.

