**PROPOSAL**

**A Chinese Speech Recognition Application for Enhancement in Chinese Speaking Education**



By :

**Lukita Iswara**

**NRP: 160919005**

Information Technology Program

University of Surabaya Engineering Department

2022

**TITLE**

A Chinese Speech Recognition Application for Enhancement in Chinese Speaking Education.

**INTRODUCTION**

Speech recognition is a translation of human speech into an understandable written format. Much research on speech recognition is already well developed, and speech recognition is already widely used by many language education enterprises. The problem is that the used-on voice recognition in the Chinese spoken language is still underdeveloped. For this reason, many languages education companies struggled to process spoken Chinese into a word format.

Today's speech recognition, such as Duolingo, Mondly, and Babbel, still have problems recognizing the proper speech in Chinese voices. Moreover, Chinese have four different tones, which means if the word is the same, but the intonation is different, it translated into a different meaning. Additionally, many applications do not process any word with intonation.

Most typical language speech recognition application only tries to compare similar wavefront; in other words, the users can use any word with a parallel wavefront. Additionally, the application can interpret it differently if the users speak the correct word, but the wavefront is different. It causes some users to repeat the words several times to get the correct wavefront.

There is a need for a better tool in Chinese speech recognition with these existing problems. The problem can be realized by using a Convolutional Neural Network (CNN). CNN is a part of an artificial neural network used to analyze images. It is expected if CNN is used correctly, then the spoken Chinese characters can be transcribed into words. CNN will be used to analyze the spectrogram from the recorded Chinese speeches.

**Research Question**

This question of the problem can be summarized as follows, “How to make application understand Chinese spoken language.”

**Research Objective**

This project objective is to enhance speech recognition application in identifying spoken Chinese characters.

**IMPLICATION**

The Implication that is expected from the project is as follows:

1. It can help speech recognizers to recognize spoken Chinese words better.
2. To add some additional method for speech recognition application
3. For finding the best accuracy in CNN speech recognition methods.

**SCOPE**

The scope of the project are:

1. Native-based application and run on a local server
2. Data obtained from YouTube
3. The data used Chinese language
4. The output is in the form of a written word format.
5. Using the CNN method for analyzing the spectrogram image.

**Activity Plan**

The plan of activities that will be carried out in making the final project can be structured as follows

1. Preparation

At this stage, data collection through YouTube video is carried out and translated into a spectrogram image. There is also a need for further studies and experiments on CNN through existing journals.

1. Analysis

At this stage, an analysis is carried out for determined the required featured and implemented the features on the application. The analysis process is carried out by reading the previous studies.

1. Design

We will design the dataset, database, and user interface for speech recognition at this stage. The data will be labeled in order to be used for classification.

1. Implementation

At the implementation stage, we will ensure that the available features can function as expected and ensure there are no bugs in the application. We will also make sure that the data is already trained.

1. Experiment and Evaluation

At this stage, we will try to experiment and evaluate the application. The experiment will try to test data that are not labeled yet. Here we will also try to find the best testing accuracy from the CNN; if the accuracy does not fulfill the requirement, the CNN method will be renewed.

1. Report

In this section, the preparation of the report is carried out based on the steps carried out in the previous activity plan.

**Proposed Solution**

It was started when the users select the word from the interface. The user then started to say the word they chose and speak it in front the microphone. The inputted voice will be translated into the MFCC coefficient. The data which are already translated will be processed in the CNN algorithm to find the best similar MFCC from the datasets. Before it tries to find the best match, the dataset will be trained to get the best training accuracy. When the word is match from the dataset, the user will receive a correct answer but if its not then it will receive incorrect.

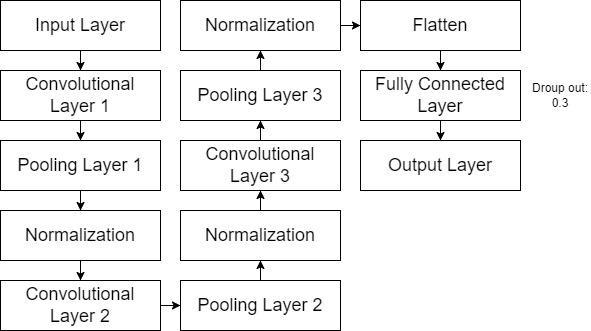


Figure 1: The proposed CNN methods

We will try the method that is shown in figure 1. If the training and testing accuracy in figure 1 is not fulfilling the requirement, we will change the CNN models. The CNN models will not be modified until they achieve the highest accuracy in testing and training.