

# CHAPTER I

## INTRODUCTION

### 1.1 Project Background

In daily life, it is very important for us to understand the emotional state of a person with certain emotions because emotions are one of the important things for life. Emotions are intense feelings directed toward someone or something (N.H. Frieda 1993). The types of emotions can be categorized as depression, anxiety, bored, frustration, fear, happiness, neutral, panic, sadness, stress, surprise, shock, and worried.

One form of interaction between humans and computers is speech. Speech consists of words that pronounced in various ways. If you only observe what is said without paying attention to the way pronunciation of the word, it is likely that important aspects of the speech will be lost, it can even happen misunderstanding. Such speech based systems can function either independently or as modules of more sophisticated techniques that combine other information sources such as facial expression and gesture (Gunes and Piccard, 2007). The challenge in the field of speech recognition is detecting the emotions of the speaker. Emotion is intense feelings directed toward someone or something. In addition, emotion can be interpreted as a reaction arising as a result of a person's actions or certain events.

To select the appropriate feature that carries information about the emotion of the sound signal is an important step in a voice-based emotion recognition system. Energy is a basic feature and most important in the sound signal. Most acoustic features that have been used for emotion recognition can be divided into two categories, that is prosodic and spectral. Prosodic features have been shown to deliver important emotional cues of the speaker (Cowie et al., 2001; Ververidis and Kotropoulos, 2006; Busso et al., 2009). To get the statistical value of the energy feature, we use the function short term to extract the energy value in each

speech frame. The emotion that a person feels is not only in the form of one form, but can be divided into various terms that are more suitable for describe how you felt at that time.

## **1.2 Research Problems**

A person can quickly change emotions due to changes in the circumstances around him. One way that a person's emotion can be identified is through their facial expressions, however sometimes human facial expressions do not match what they are experiencing or feeling. So at this time, someone will look for a more flexible and comfortable method in detecting the emotional state he is experiencing. If previously the human detection system was done through facial expressions, now the technology industry is increasingly looking for a new approach by developing a system for detecting human emotions through voice.

The need to know a significant set of emotions to be classified by automatic emotion recognition is a major concern in speech emotion recognition systems. A typical set of emotions contains 300 emotional states. Therefore to classify a large number of emotions is very complicated. Another problem is that emotion expression is depending on the speaker and his or her culture and environment. As the culture and environment gets change the speaking style also get change, which is another challenge in front of the speech emotion recognition system. There may be two or more types emotions, long term emotion and transient one, so it is not clear which type of emotion the recognizer will detect (M.E.Ayadi; M.S Kamel, F.Karray, 2011).

The performance of speech emotion recognition system is influenced many factors, especially the quality of the speech samples, the features extracted and classification algorithm. Actually, emotional feature extraction is the main problem in Speech emotion Recognition system. Surabhi and Saurabh (2016), have proposed important speech features which contain emotion information, such as energy pitch, format frequency, Linear Prediction Cepstrum Coefficients (LPCC), Mel Frequency Cepstrum Coefficients (MFCC).

So, some of the problem statement that related in this system are

- a. Difficulties to detect an individual's emotions by just listening to his/her

voice without looking at his/her facial expressions.

- b. Difficulties to predict all emotions, because speech can have more than one emotion at the same time, and that will obviously affect the accuracy of the prediction.
- c. A lack of the performance of speech pattern recognition when do for searching the vocabulary because of the number vocabulary increasing.

### 1.3 Objectives

This project will focus on developing an emotion detection system through voice with several type of emotions, such as angry, disgust, fear, happy, neutral, sad, and surprise. The objectives of this projects are :

1. To identify the basic emotions from user speech.
2. To calculate percentages types of the emotions from user speech.
3. To develop a speech emotion recognition system.

### 1.4 Scope

#### System Scope:

1. The system develop is a web-based application.
2. Record audio.
3. Database to store audios.
4. Analyze audio/speech.
5. Identify some of emotions , there are angry, disgust, fear, happy, neutral, sad, and surprise.

#### Admin Scope:

1. To register, edit and delete users.
2. To generate user reports.

#### User Scope:

1. Users can record their voice.
2. The users can see the result of predicting their emotions.
3. User can register an account.

### 1.5 Significance

In this project, the significance of this study include :

- a. The development of a web based application that is able to identify basic

emotion from user's speech.

- b. The development of a web based application that is able to calculate percentages types of the emotions from user speech.
- c. The development of a web based application that is able to recognizes speech users.

### **1.6 Assumption and Limitation**

There are 3 assumptions as per listed:

1. Speaker will record their voice in the system.
2. This system will analyze the recorded audio from speech speaker.
3. This system works by match the recorded audio with the specific emotions.

The limitation for this project as below:

1. The system will use input data that is pre-loaded into the system.
2. Results provided by the system will be based on the matching with existing input data only.