American Journal Of Current Tendency and Innovation, 2024, 1-9
https://publishingjournals.org/ajcti
Volume: 1 | Number: 2 (2024) April

Speech Emotion Recognition System

Silviana Widya Lestari
School of Graduate Studies Management and Science University
Shah Alam,Selangor, Malaysia
silvianawidya46@gmail.com

AP. Ts. Dr. Saliyah Kahar
Faculty of Information Sciences and Engineering
Management and Science University Shah Alam,Selangor, Malaysia
saliyah_kahar@msu.edu.my

Trismayanti Dwi P.
Information Technology State Polytechnic of Jember Jember, Jawa Timur, Indonesia
Trismayanti@polije.ac.id

***Correspondence :**
Silviana Widya Lestari
silvianawidya46@gmail.com

Received: 12-02-2024
Accepted: 02-03-2024
Published: 24-03-2024

Abstract

Emotion is a reaction arising as a result of a person's actions or certain events. It is very important to understand the emotional state of a person with certain emotions because emotions are one of the important things for life. Emotion detection can be done in two ways, namely, through the face and through the voice. In this study, researchers used sound as a medium for detecting sound. System Development Life Cycle were used as a methodology where each phases are important to achieve the goals of the project. Each phase is critical to meet client requirements and achieve project objectives. Moreover, the development life cycle is a well-described method that has steps in standard phases which aim to regulate the development of the application. This application system is designed and implemented using the Python programming language in Visual Studio Code. There are 2 main features, namely real time for user voice recording and upload for users to upload their record files in wav format. Therefore, by implementing this system, he will be able to detect the main emotions, namely Angry, Disgust, fear, Happy, Neutral, Sad, and Surprised through the user's voice. Not only that, the

system also provides a real time system that can show the percentage of each type of emotion that is generated every 5 seconds.

Keywords: Recognition, Web based Application, Emotion

Introduction

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. 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. The types of emotions can be categorized as depression, anxiety, bored, frustration, fear, happiness, neutral, panic, sadness, stress, surprise, shock, and worried. 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.

LITERATURE REVIEW

In today's era, technological developments are undeniable, almost all fields have used technology. Many companies are creating engineering innovative solutions to facilitate human work, such as in the form of tools or applications.

Yixiong Pan, Peipei Shen, and Liping Shen (2012), focused on analyze the system accuracy on the first two aspects with large numbers of test and experiments. The performance of speech emotion recognition system is influenced by many factors, especially the quality of the speech samples, the features extracted and classification algorithm. For those study is like typical pattern recognition systems, their speech emotion recognition system contains four main modules: speech input, feature extraction, SVM based classification, and emotion output.

Seunghyun Yoon, Seokhyun Byun, and Kyomin Jung (2018), the system propose a novel deep dual recurrent encoder model that simultaneously utilizes audio and text data in recognizing emotions from speech. The proposed model outperforms previous state of the art methods by 68.8 % to 71.8% when applied to the IEMOCAP dataset, which is one of the most well-studied datasets. For the first step, they propose a multimodal approach that encodes both audio and textual information simultaneously via a dual recurrent encoder. The model are Audio Recurrent Encoder (ARE), Text Recurrent Encoder (TRE), and Multimodal Dual Recurrent Encoder (MIDRE).

Leila Kerkeni, Youssef Serrestou, Mohamed Mbarki, Kosai Raoof and Mohamed Ali Mahjoub (2018), this study is focuse to compare different approaches for emotions recognition task and propose an efficient solution based on combination of these approaches. In this experimental work, they have used Multivariate Linear Regression (MLR), Support Vectore Machine (SVM) and Recurrent Neural Networks (RNN) classifiers to identify the emotional state of spoken utterances. The emotional speech databases used in their experiments are Berlin Database and Spanish Database.

Vokaturi - Android

Vokaturi is a software that can recognize the emotions of the

speaker's voice. The application shows the following emoji's based on the results detected from the speaker's voice. Currently the community version of the Vokaturi is able to detect the five different types of emotions, there are:

a. Neutrality
b. Happiness
c. Sadness
d. Anger
e. Fear

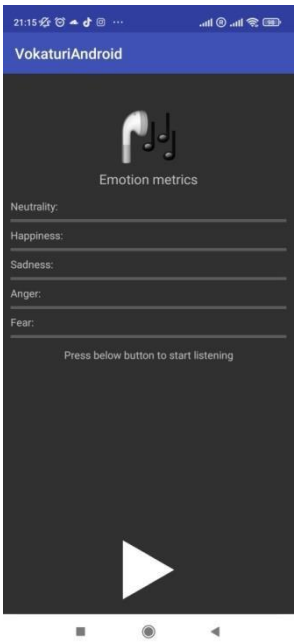

Figure 2.1: Vokaturi Android

EmoWatch

EmoWatch is a mental health app visualizing human mood from their voice, so the apps enables human to check their daily mood easily. By using "Empath," vocal emotion recognition sensor developed by Smartmedical Corp. The app is helpful for keeping a good mental wellbeing in various situations.

Figure 2.2 EmoWatch

RESEARCH & METHODOLOGY

This project will use the Agile project management methodology. This project is to help users recognize their speech to detect emotions and provide the type of emotions that is being recorded and help users in knowing their current emotions.

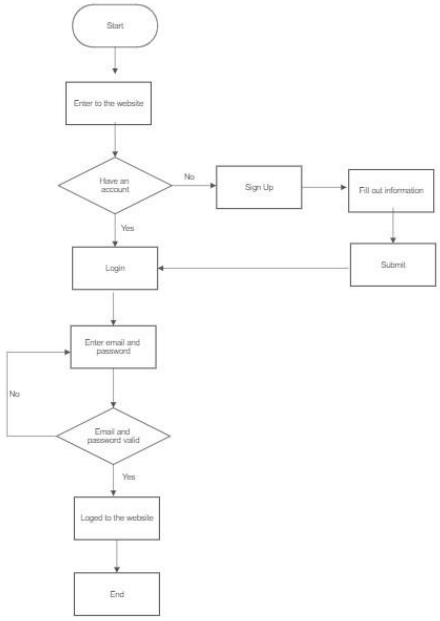

Figure 3.1: Login Flowcharts

Figure 3.1 shows login flowcharts for users. Users need to sign up first to register a new account, fill some data and then submit. Then, users can enter email and password for login to the website.

Detecting Flowchart from Recording

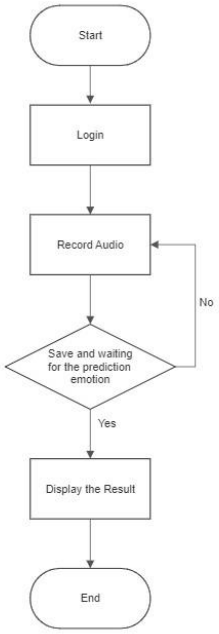

Figure 3.2: Detecting Flowchart from Recording

Figure 3.2 shows the detecting flowcharts from recording speech of user. Users are able to record their

voice and the user have to waiting for the result from the system, once they received the result for the type emotion so the detection process is done

Detecting Flowchart from Upload File

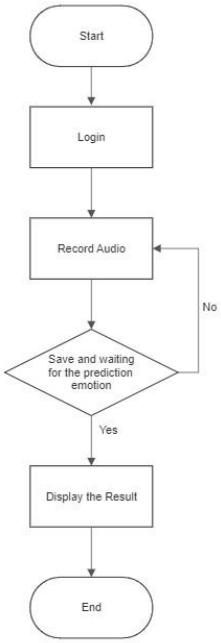

Figure 3.3: Detecting Flowchart from Upload File

Figure 3.3 shows the detecting flowcharts from upload file. Users are able to upload their file in format wav and the user have to waiting for the result from the system, once they received the result for the type emotion so the detection process is done.

Use Case Diagram

As shown in figure 3.4 below, the use case diagram illustrate how this system work. The actor for this system are user, admin, and also the system.

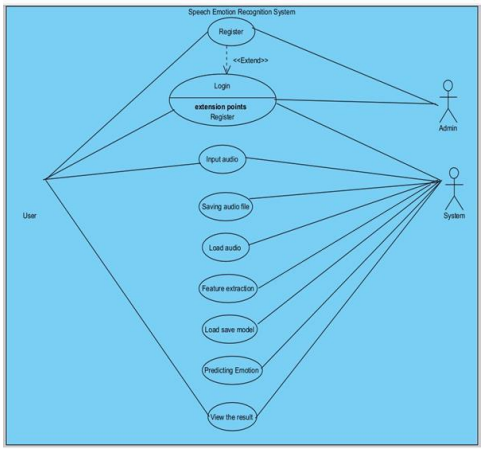

Figure 3.4: Use Case Diagram

Activity Diagram

As seen in figure 3.5 below, it illustrates an activity diagram which first starts with user to input the audio, then the system will load the trained model so the audio can be saved and extract to detect the type of emotions.

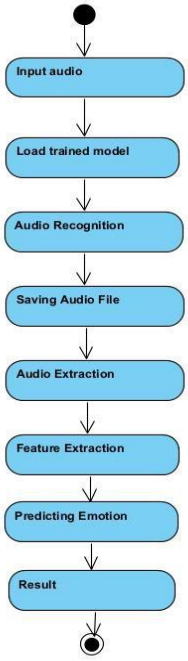

Figure 3.5: Activity Diagram

Sequence Diagram

As seen in the image 3.6 below, it depicts a sequence diagram of this system where we have actors as users.

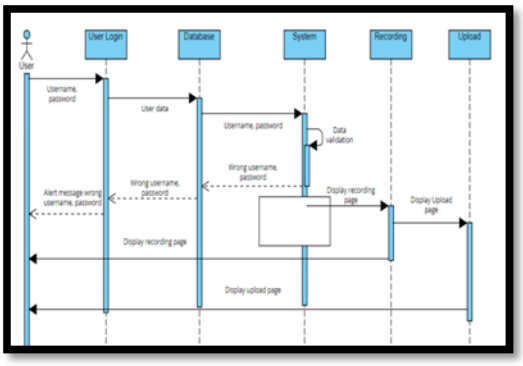

Figure 3.6: Sequence Diagram

Communication Diagram

In this diagram. there are several symbols that represent the function. An actor is a type of role played by an entity that interacts with a subject. Next is the call as a message symbol that defines a certain communication between the Lifelines of an interaction. Figure 3.7 below illustrates how the communication diagram works for this project.

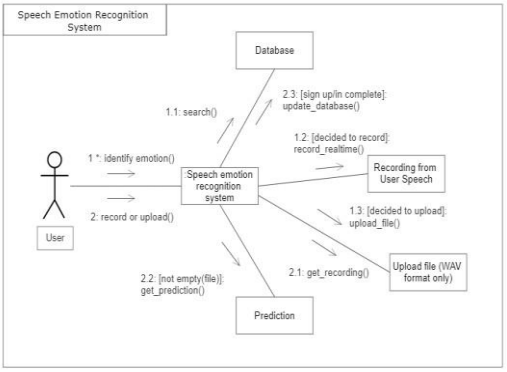

Figure 3.7: Communication Diagram

FINDING AND DISCUSSION

a. Register page

As in general, if the user has not registered then he will not get user access rights to be able to access the system.

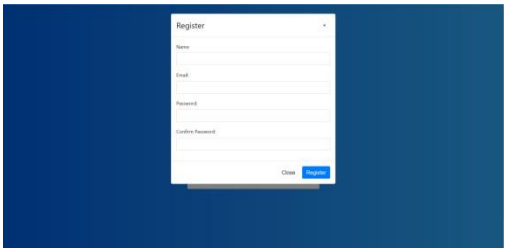

Figure 4.1: Register page

b. Login page

This page shows the first page when the system is opened. Users will be faced with a login page for users who have registered.

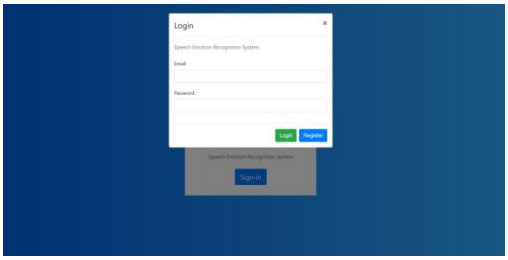

Figure 4.2: Login page

c. Home page

Page On this home page it has 2 main menus, namely the upload menu and voice recording as shown in the image above.

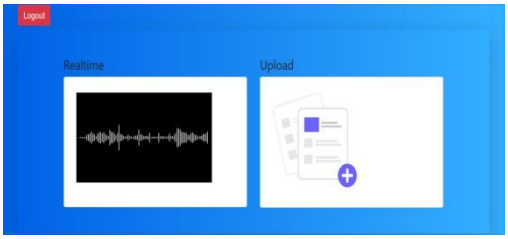

Figure 4.3: Home page

d. Record

Page On the record page, there is a record button for users to click before they record their voice. Voice recording in this menu is limited to 3 seconds. In addition, on this page there is also a real time menu, so that the voice that enters the device will be detected with a different percentage of emotion in each type of emotion, depending on the frequency of the sound captured by the device.

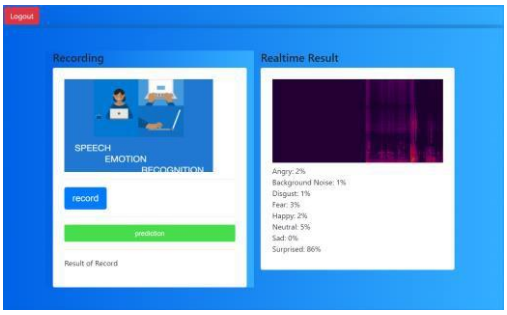

Figure 4.4: Record page

e. Upload File

On the upload page, the user is required to upload a voice recording in wav format so that the type of emotion can then be predicted.

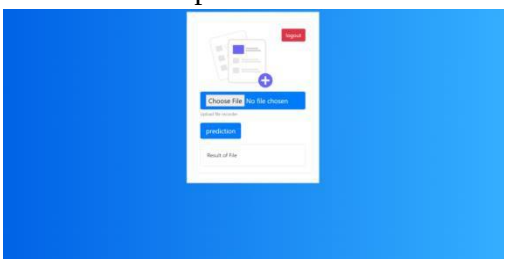

Figure 4.5: Upload File Page

CONCLUSION AND RECOMMENDATION

The purpose of using this application is to make it easier for users to detect the 7 basic emotions. The 7 basic emotions are angry, disgust, fear, happy, neutral, sad and surprised. The main feature of this application is the feature of recording in real time and also uploading files in wav form. This application system is designed and implemented using the Python programming language and Visual Studio Code. The proposed system has succeeded in identifying basic emotions from user speech and displaying calculated percentages in real time. This can be utilized by the mental health sector, especially for counselors or psychologists in helping to analyze the state or emotional level of their clients.

The results obtained from this study are not perfect, but in general, the system can run well because it can detect 7 basic emotions as in the objectives of this study described previously. Therefore, to improve it, further project development can be carried out by adding more types of emotions other than the 7 basic emotions that are already common.

REFERENCES

Reakaa, S. D., & Haritha, J. (2021, May). Comparison study on speech emotion prediction using machine learning. In *Journal of Physics: Conference Series* (Vol. 1921, No. 1, p. 012017). IOP Publishing.

Han, K., Yu, D., & Tashev, I. (2014, September). Speech emotion recognition using deep neural network and extreme learning machine. In *Interspeech 2014*.

Shen, P., Changjun, Z., & Chen, X. (2011, August). Automatic speech emotion recognition using support vector machine. In *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology* (Vol. 2, pp.
621-625). IEEE.

Wootaek Lim, Daeyoung Jang dan taejin Lee. (2016). Speech Emotion Recognition using Convolutional and Recurrent Neural Networks.

Yoon, S., Byun, S., & Jung, K. (2018, December). Multimodal speech emotion recognition using audio and text. In *2018 IEEE Spoken Language Technology Workshop (SLT)* (pp. 112-118). IEEE.

Kerkeni, L., Serrestou, Y., Mbarki, M., Raoof, K., & Mahjoub, M. A. (2018). Speech Emotion Recognition: Methods and Cases Study. *ICAART (2)*, *20*.

Han, K., Yu, D., & Tashev, I. (2014, September). Speech emotion recognition using deep neural network and extreme learning machine. In *Interspeech 2014*.

Mirsamadi, S., Barsoum, E., & Zhang, C. (2017, March). Automatic speech emotion recognition using recurrent neural networks with local attention. In *2017 IEEE International conference on acoustics, speech and signal processing (ICASSP)* (pp. 2227-2231). IEEE.

Jiang, W., Wang, Z., Jin, J. S., Han, X., & Li, C. (2019). Speech emotion recognition with heterogeneous feature unification of deep neural network. *Sensors*, *19*(12), 2730.

Ghosh, S., Laksana, E., Morency, L. P., & Scherer, S. (2016, September). Representation Learning for Speech Emotion Recognition. In *Interspeech* (pp. 3603-3607).