REVIEW ON VIRTUAL ASSISTANT WITH EMOTION RECOGNITION

Prof. Y.D.Choudhari*1, Aditya There*2, Mayuree Bawane*3, Ruchika Bais*4, Riya Rajgire*5, Harshal Balpande*6

*1Prof., Department of Information Technology, KDK College of Engineering, Nagpur, Maharashtra, India.
*2,3,4,5,6Department of Information Technology, KDK College of Engineering, Nagpur, Maharashtra, India.

ABSTRACT
Virtual Assistants became widely popular; they make a lot of tasks easier for us especially in our busy hours. It is one of the great results of advancing Artificial Intelligence. Enhancing them by adding features of emotion recognition by speech will make it more flexible. It reduces the time required to type as well as it will perform the task according to the emotion. This paper presents how we can use Emotion recognition, Speech recognition, with AI in Virtual assistant and improve its flexibility as well as how it is useful in daily life.

Keywords: Artificial intelligence, Python, Emotion recognition, Speech recognition, Virtual assistant.

I. INTRODUCTION
Artificial intelligence is advancing day by day with new features, new techniques and with new ways to make machines more intelligent like humans. AI is working on robotics and it is the biggest achievement in this field. But apart from this virtual personal assistant is another achievement which is used widely in many gadgets and devices like computers, televisions, washing machines, cars, etc.

With the development of voice recognition technology, automatic spoken dialogue system and natural language processing with machine learning and AI, Virtual Personal Assistant was born. VPA is software which is an assistant that works virtually via voice. Siri, Cortana, Alexa, Google assistant are all examples developed and used widely. As speaking covers more words than typing at the same time, it becomes faster. It takes more time to write a paragraph than reading it. This explains the wider use of VPA and more implementation by IT companies. Moreover it provides more flexibility and is very helpful for a busy world.

Here this is also a virtual personal assistant which performs basic tasks asked by us via voice. As well as answer certain basic questions like what’s the time? What’s the date? Etc. But the thing which makes it different from other VPA is that it can recognize the emotion. At every command it can recognize the user’s emotion and then perform tasks according to the emotion which VPA detects via speech or voice. In this way we get the specific information according to our current emotions.

II. LITERATURE STUDY
The first chatbot was made in 1960 which showed the beginning of development in virtual assistants. Considering existing product many new technology developed and keep advancing. Speech recognition is one of the techniques which helps to improve the functionality of virtual assistants. Now there are many different languages available to assist people and it keeps improving. For emotion recognition researchers work for decades and were able to successfully develop. Different methods and techniques are developed like Bayesian networks, Gaussian Mixture models and Hidden Markov models and deep neural networks. Speech recognition and emotion recognition both have been advancing in their own field. Many virtual assistants like Alexa, Siri are used in many places. Their latest need is to implement emotion recognition technique which will help to recognize our emotions. The development is called Emotion AI.

Prototypes and commercial products for emotion driven assistants already exist for example Beyond Verbal’s voice recognition app, Emotion AI for speech by Affectiva, and the connected home VPA Hubble.
III. PROBLEM FORMULATION

There are many virtual assistants which work on commands given by speech. They perform tasks according to the experience and data provided to it. But when it comes to the emotion part we need to specify the particular emotion in commanding the task to assist.

To overcome this problem we come out with this virtual assistant with emotion recognition via speech. It can be used as a normal assistant but we don’t need to specify our emotions. It is more useful when we are not in condition to specify our emotion and the task should be done by considering our emotion.

Virtual assistants will perform basic tasks and answer the basic questions which are added to its database. And it will also notify the emotion at the same time.

We are using python language for this project since it provides many libraries which makes it more reliable and handy, because of its object oriented dynamical nature.

Speech recognition technique is used here to identify the emotion, so input will be speech. For this purpose we are using different python libraries.

1. Librosa library, in order to understand and analyze the audio and music file. It is very useful for tasks which include music files.

2. scikit-learn is a very useful library for the python language, used for machine learning purposes. It includes many classification, regression and clustering algorithms. MLPClassifier is used here from the scikit-learn library.

3. Sound file is an audio library which is used to read and write the sound files.

For emotion recognition, we are using the Ravdess dataset, Ryerson Audio-Visual Database. It contains 7356 files which helps to identify the emotion via speech or song.

RNN, recurrent neural network is used here to enhance the result. In this, machines consider previous data to make the prediction for current output with provided input. Again when output is given it stores it for future tasks, so it goes in the loop. So while creating output it considers all the data it has. RNN includes LSTM, which is long short term memory and is well known for its classification, processing and making predictions based on time series data.

In this all the sequence is maintained, thus RNN is known for its sequence learning.

In this virtual assistant many features which consider emotions are included, one of the features is that if we ask the assistant “Play the song” then firstly it will recognize the emotion and according to it, it will play the song. For this purpose we are using a Ravdess dataset. Apart from this it is very useful when we are in a bad mood, it will pin the friend and inform them that we are not in a good mood by which the friend can contact and try to refresh the mood. It can also observe our health and fitness, by our voice it can tell about fatigue, overexertion, etc.

The basic functions like taking a photo, going to the browser, starting karaoke, etc. are also performed. The basic questions like what’s the time, what’s the date, etc. are already added to the database which helps the user in its busy time. By considering our emotions it can also start Netflix to refresh the mood or any movie.

IV. WORKING

A virtual assistant comprising: a user input for providing information about a user. An emotion detection module configured to detect a user’s emotion from input information. An input transform module is used for transforming information into normalized emotion data. And then a core module is used for producing a virtual assistant emotion for the virtual assistant based on detected user emotion.

When the speech is taken by an assistant it determines two things: first emotion and then task to be performed.

So, the speech recognition model will perform to identify emotion by analyzing the speech. In this model firstly speech signals are considered and then feature selection is performed to improve the learning algorithm performed on task by reducing the number of features used to characterize a dataset. And then classification of emotion is done.
And taking the result from the speech emotion model it performs a task and gives the result. For example if the user “say play song” then firstly it will go in speech emotion model and analyze it if for example the emotion is detected as “sad” then VPA will start playing sad songs stored in the database. This flow keeps going in loop until the user gives the end or bye command when it asks again.

V. CONCLUSION

Virtual assistants with emotion recognition by speech can respond to emotions correctly as well as show the emotions before performing tasks. Sometimes we aren’t able to express our emotions in words at such times. It is very useful. The time will come when an assistant will understand our emotions better than other people and be a friend of humans.

This assistant provides more interfaces for human interaction and improves the efficiency by reducing the time for accomplishing the task since typing takes more time compared to speaking.

Since we are applying AI its cost is very high and its development is slow. Additionally its accuracy is only 70 to 80%.

VI. REFERENCES

