PERSONAL VOICE ASSISTANT

K.VENKATESH[1] ,M.PREMANVI[2]

[1] **Assistant Professor MCA, DEPT,** Dantuluri Narayana Raju College, **Bhimavaram, Andhrapradesh**

Email id: -kornalavenkatesh@gmail.com

[2]**PG Student of MSC(CS),** Dantuluri Narayana Raju College, **Bhimavaram, Andhrapradesh**

Email id: - premanvi1999@gmail.com

ABSTRACT

Voice control is a major growing feature that change the way people can live. The voice assistant is commonly being used in smartphones and laptops. AI-based Voice assistants are the operating systems that can recognize human voice and respond via integrated voices. This voice assistant will gather the audio from the microphone and then convert that into text, later it is sent through GTTS (Google text to speech). GTTS engine will convert text into audio file in English language, then that audio is played using play sound package of python programming Language.

It will look at examples of intelligent programs with natural language processing that are currently available, with different categories of support, and examine the potential usefulness of one specific piece of software as a VPA. This engages the ability to communicate socially through natural language processing, holding (and analysing) information within the context of the user. It is suggested that new technologies may soon make the idea of virtual personal assistants a reality. Experiments conducted on this system, combined with user testing, have provided evidence that a basic program with natural language processing algorithms in the form of a VPA, with basic natural language processing and the ability to function without the need for other type of human input (or programming) may already be viable.

1 INTRODUCTION

Nowadays the Mobile Technology is being very famous for the User Experience, because it is very easy to access the applications and services from anywhere of your Geo location. Android, Apple, Windows, Blackberry, etc. are various famous and commonly used Mobile Operating Systems. All the Operating Systems provides plenty of applications and services for users. For an instance, the Contacts Applications is used to store the contact details of the user's contact and also helps user to connect a call or send an SMS to other person using the contents stored in this application. We can get similar types of application all around the world via Apple Store, Play Store, etc. All this features gives birth to various kinds of sensors or functionalities to be implemented in the mobile devices. The Most famous application of iPhone is "SIRI" which helps the end user to communicate end user to mobile with voice and it also responds to the voice commands of the user. Same kind of application is also developed by the Google that is "Google Voice Search" which is used for in Android Phones. But this Application mostly works with Internet Connections. But our Proposed System has capability to work with and without Internet Connectivity.

2. LITERATURE SURVEY AND RELATED WORK

2.1 Speech recognition with flat direct model

Authors:Geoffrey Zweig,Georg Heigold,Phuongtrang Nguyen

This paper describes a novel direct modeling approach for speech recognition. We propose a log-linear modeling framework based on using numerous features which each measure some form of consistency between the underlying speech and an entire sequence of hypothesized words. Since the model relates the entire audio signal to a complete hypothesis without necessarily positing any inherent structure, we term this a flat direct model (FDM). In contrast to a conventional hidden Markov model approach, no Markov assumptions are used, and the model is not necessarily sequential. We demonstrate the use of features based on both template-matching distances, and the acoustic detection of multi-phone units which are selected so as to have maximal mutual information with respect to word labels. Further, we solve the key problem of how to define features which can generalize to unseen word sequences. In the proposed model, template-based features improve sentence error rate by 3% absolute over the baseline, while multi-phone-based features improve by 2% absolute.

2.2 Accurate and compact large vocabulary speech recognition on mobile devices

Authors:Jeffrey Scott SorensenX. Lei, A. Senior,

In this paper we describe the development of an accurate, smallfootprint, large vocabulary speech recognizer for mobile devices. To achieve the best recognition accuracy, state-of-the-art deep neural networks (DNNs) are adopted as acoustic models. A variety of speedup techniques for DNN score computation are used to enable real-time operation on mobile devices. To reduce the memory and disk usage, on-the-fly language model (LM) rescoring is performed with a compressed n-gram LM.

2.3lattice-based optimization of sequence classification criteria for neural-network acostic modeling

Authors: Brian Kingsbury :Acoustic models used in hidden Markov model/neural-network (HMM/NN) speech recognition systems are usually trained with a frame-based cross-entropy error criterion. In contrast, Gaussian mixture HMM systems are discriminatively trained using sequence-based criteria, such as minimum phone error or maximum mutual information, that are more directly related to speech recognition accuracy. This paper demonstrates that neural-network acoustic models can be trained with sequence classification criteria using exactly the same lattice-based methods that have been developed for Gaussian mixture HMMs, and that using a sequence classification criterion in training leads to considerably better performance. A neural network acoustic model with 153K weights trained on 50 hours of broadcast news has a word error rate of 34.0% on the rt04 English broadcast news test set. When this model is trained with the state-level minimum Bayes risk criterion, the rt04 word error rate is 27.7%.

2.4Automatic question generation for decision tree based state tying

Authors: K. Beulen,H. Ney:Decision tree based state tying uses so-called phonetic questions to assign triphone states to reasonable acoustic models. These phonetic questions are in fact phonetic categories such as vowels, plosives or fricatives. The assumption behind this is that context phonemes which belong to the same phonetic class have a similar influence on the pronunciation of a phoneme. For a new phoneme set, which has to be used, for example, when switching to a different corpus, a phonetic expert is needed to define proper phonetic questions. In this paper a new method is presented which automatically defines good phonetic questions for a phoneme set. This method uses the intermediate clusters from a phoneme clustering algorithm which are reduced to an appropriate number afterwards. Recognition results on the Wall Street Journal data for within-word and across-word phoneme models show competitive performance of the automatically generated questions with our besthandcrafted question set.

3 EXISTING SYSTEM

In existing system user use to search by using keyboard typing.it takes lot of times to search.In recent times only in the Voice Assistants we can experience the major changes, the way user interacts and the experience of user. We are already using them for many tasks like switching on/off lights, playing music through streaming apps like Wynk Music, Spotify etc., This is the new method of interacting with the technical devices makes lexical communication as a new ally to this technology.

4 PROPOSED WORK AND ALGORITHM

The project will give a fair knowledge about the intelligent assistant which is capable of understanding the commands given by the user. Our assistant can easily understand the commands given by the user through vocal media and responds as required. Our assistant performs the most frequently asked requests from the user and makes their task easier. Our voice assistant listens to the command given by the user through the microphone. After listening it will says 'speaking'and displays what the user said and acts accordingly.

5 METHODOLOGIES

MODULES

Office Mathematical Operations

Online Operations

Exit

MODULES DESCRIPTION:

1.Office Mathematical Operations

In this module user will speak about Mathematical Operations.In this module all Mathematical Operations is carried out.

(Eg:1+2=3)

2.Online Operations

In this module user will speak about search related task. It will search in online using google search.

Eg:Temperature

3.Exit

6 RESULTS AND DISCUSSION SCREENSHOTS

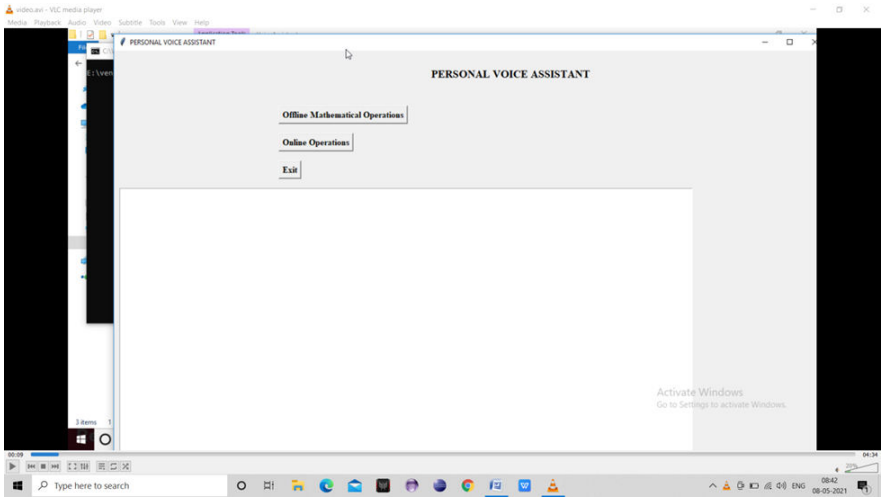

Fig 1:- HOME SCREEN

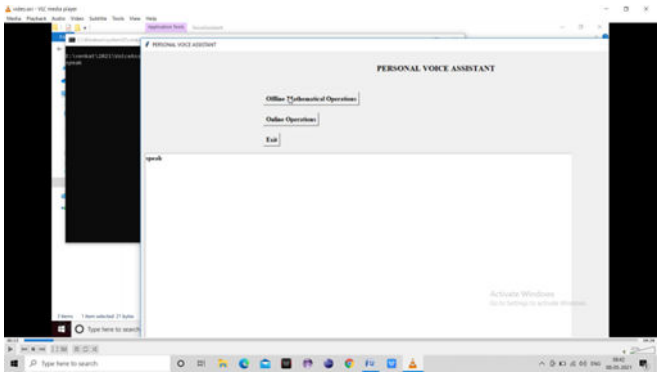

Fig 2:- MATHEMATICAL CALCULATIONS WITH VOICE

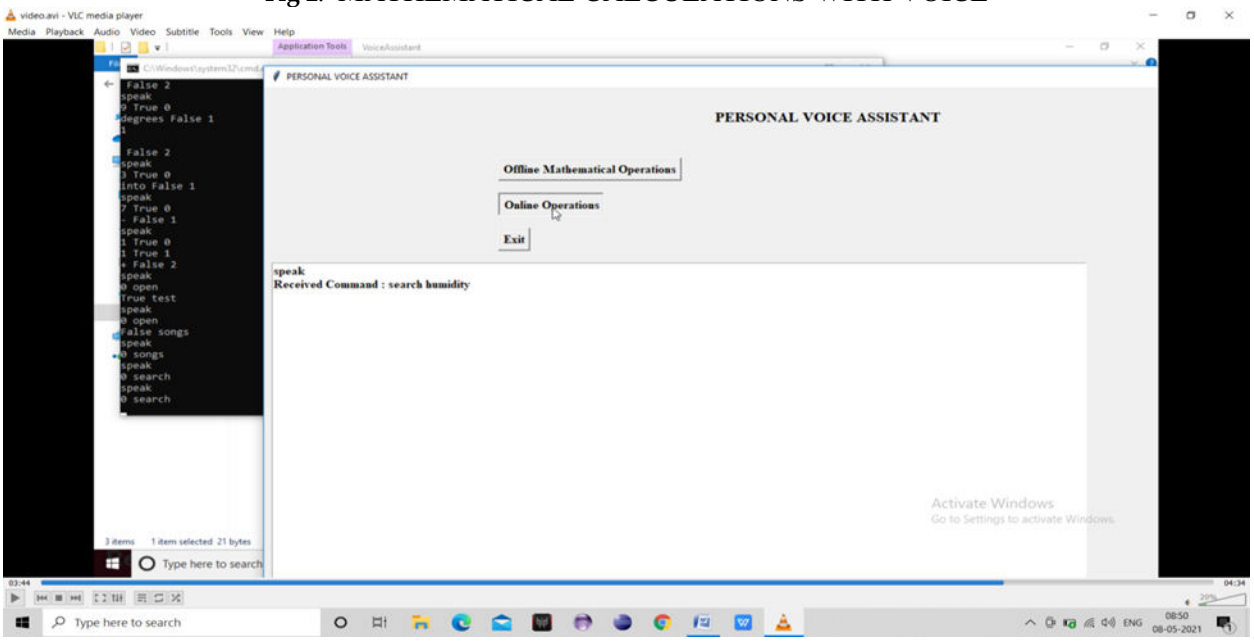

Fig 3:- Generating

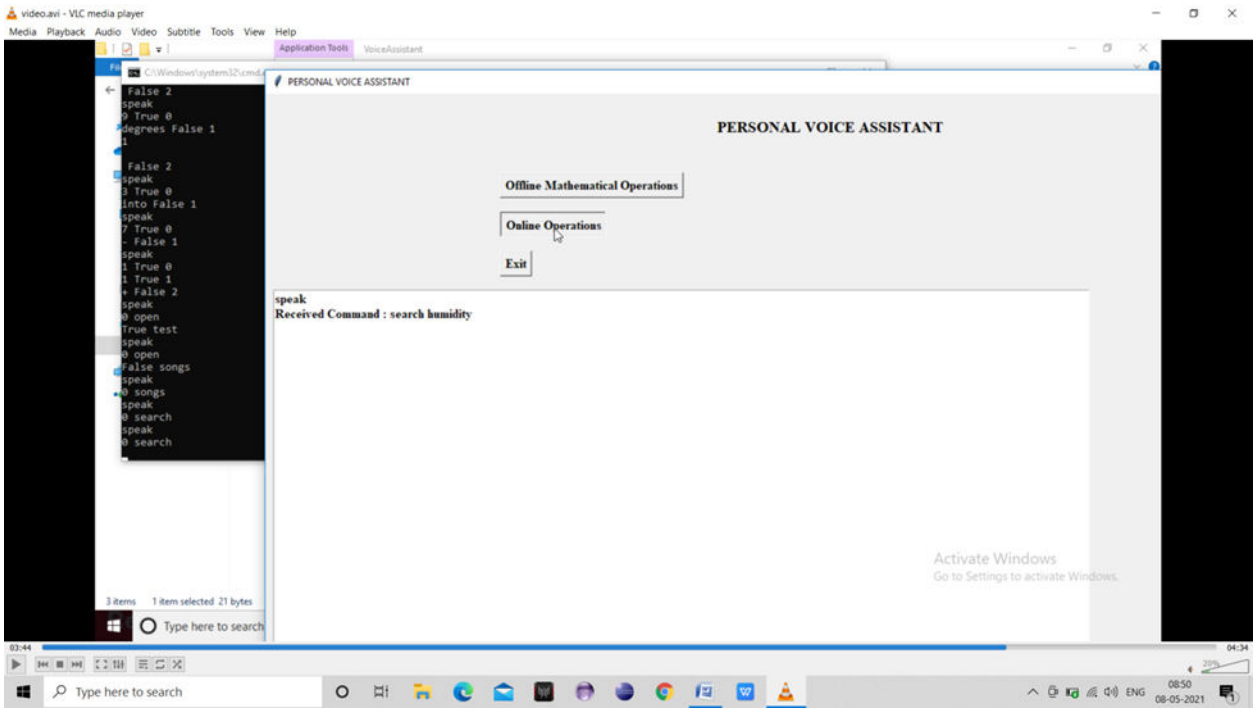

Fig 4:- PERSONAL VOICE TO KNOW CLIMATE CONDITION

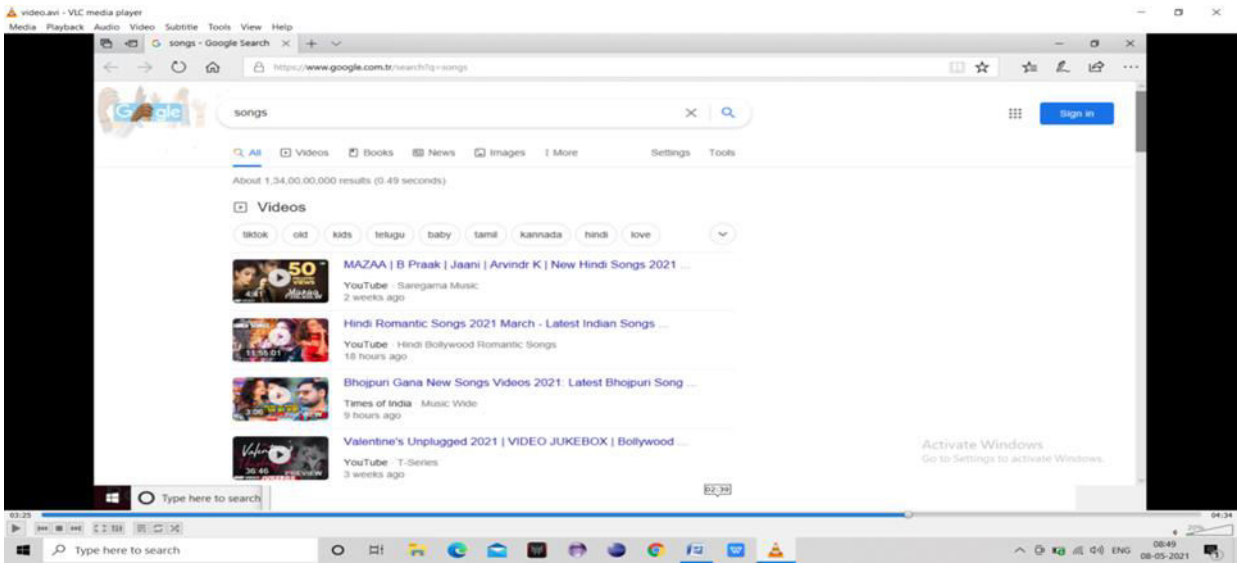

FIG 5 :- VOICE SEARCH FOR SONGS

6.CONCLUSION AND FUTURE SCOPE

CONCLUSION

In our project we have implemented many things compared to other assistants. Now a days it is very useful in human life because it is a hands-free application. It is a very simple application. As well as it is used in a business field also for example in laboratory, the person wears gloves and body suits for their safety purpose so it is difficult to type, through voice assistant they can get any information so that their work becomes easy. Voice assistants are useful in many fields such as education, daily life application, home appliances etc. and voice assistant is also useful for the illiterate people they can get any information just by saying to the assistant, luxury is available for people, thanks to AI based voice assistants.

FUTURE SCOPE:

In future voice assistants can be used for two developments: First quality of dialogue recognition will increase because broadband allows more complex data processing in powerful data centres. Second, from the users perspective, VAs aid for interaction. In the companies, voice assistants can be used to automate repetitive tasks for book meeting rooms and restaurants etc.

7 REFERENCES

[1] "Accurate and compact large vocabulary speech recognition on mobile devices," in INTERSPEECH. 2013, pp. 662–665, ISCA

[2] "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in ICML, 2006, pp. 369–376.

[3] "Lattice-based optimization of sequence classification criteria for neural-network acoustic modeling".

[4] "Language modeling for What-with Where on GOOG411," Proc. of INTERSPEECH, 2009.

[5] "DFirect construction of compact context-dependency transducers from data," Proc. of INTERSPEECH, pp. 218–221, 2010.

[6] "Automatic question generation for decision tree based state tying," Proc. of ICASSP, 1998.

[7] "Speech recognition with flat direct models," IEEE Journal of Selected Topics in Signal Processing, 2010.