Desktop Assistant Based on Voice Recognition and Face Detection

Article *in* International Journal of Grid and Utility Computing · July 2020

CITATIONS
0

READS
1,831

1 author:

Rajesh Kumar Patjoshi
National Institute of Science and Technology Berhampur
51 PUBLICATIONS **254** CITATIONS

SEE PROFILE

International Journal of Grid and Distributed Computing
Vol. 13, No. 1, (2020), pp. 2020 – 2030

Desktop Assistant Based on Voice Recognition and Face Detection

[1]Raju Shanmugam, [2]Rajesh Kumar Patjoshi, [3]Soumya Ranjan Jena, [4]Vishvaketan Gaur
[1] Professor and Dean, [2] Associate Professor, [3] Assistant Professor, [4] Under Graduate Student
[1] [3] [4] School of Computing Science & Engineering, Galgotias University, Greater Noida, U.P., India
[2] Department of ECE, National Institute of Science & Technology, Berhampur, Odisha, India
[1]srajuhere@gmail.com, [2]rajeshpatjoshi1@gmail.com, [3] soumyajena1989@gmail.com, [4]vishvaketan.1997@gmail.com

Abstract

In recent future all the electronic devices will be worked by utilizing the remote helper which is definitely not hard to get to yet it needs weakness. This structure affirmed the clients to get to the framework by the voice orders. User can request to the assistant that anything can be done by the framework, for example Music, Open Specified Application, Open Tabs, Open Websites and so forth. Voice associates are programming specialists that can decipher human discourse and react through orchestrated voices. Clients can pose their associate's inquiries, control home gadgets and media playback by means of voice, and oversee other essential errands, for example, email, daily agendas, and schedules with verbal orders.

Index Terms— *Voice Assistant, Voice Recognition Face Detection, Haar Cascade Algorithm, Fisher Face Algorithm.*

I. INTRODUCTION

Voice assistants are defined as the software agents who interpret or convert human speech and it responds through synchronized voices, i.e. Siri (Apple Inc.), Alexa (Amazon), Cortana (Microsoft), so forth these are most popular voice assistants. As the technology are getting advancer day-by-day, now we can see that the futuristic dream of talking to the computer is now comes true as from the day the computer invented, People wants to talk with the computer machine, can be possible with the technology development by the dedicated and devoted computer scientists. Several products deployed in the last few years which bring out the in-expensive use of the voice assistants to our day-to-day life, with regard to the time more features and platforms are being added and get updated from time to time.

.

II. MODULES AND DESCRIPTION

Speech To Text

In this module or layer a speech is to be converted into the text, which can be understandable by the system through the installed libraries.

Text Analysing

It converts the text for the system, it analyze and then make it into the readable form. Though, Computer understands the command, so virtual assistants like Siri, Cortona so forth converts the text to computer commands.

ISSN: 2005-4262 IJGDC

Interpret Commands:

In this layer, mapped PC sends requests to the server through the web. At the same time, as the discourse assessed locally. A nearby recognizer speaks with the server to decide whether the order will be ideal to deal with locally or not. For example Play Music, Restaurant Reservation, Movie Review and so forth.

The rest of the paper is written in following way. Section-III describes related research work. Section-IV discusses research challenges. Section-V tells us about different research requirements whereas Section-VI and VII highlight research set-up and system framework respectively. Proposed method has been given in Section-VIII. Several modules of this research project that are being used described in Section-IX. Section-X tells about the output results. Advantages, disadvantages and applications have been given in Section-XI. Finally conclusion has been drawn in Section-XII.

III. Related Work

This system has a long history with a few rushes of significant advancements. It acknowledges for transcription, search, what's more, voice orders has become a standard component on cell phones and wearable gadgets [1]. It is accomplished for utilizing a CTC based LSTM an acoustic model, which predicts setting free telephones so forth, It shrink to a 1/10 of its unique size utilizing a mixture of SVD-based pressure and quantization. It restricts the networks and on-the-fly language model rescoring to accomplish constant execution on present-day cell phones.

In [6] we likewise coordinate an inquiry parsing module among ASR and Search for multiple reasons. Also it [3] contains the methods for the proper implementation of computer generated voice search it administrates and proposed the versatile clients getting to the administrations over a scope of compact gadgets. Voice recognition is executed as 2-phase recognizing strategy where string up-and-comers created by a programmed Automatic Speech Recognition (ASR) framework. So as to distinguish the coordinating passage from a possibly extremely enormous application-explicit DB that gives a genuine case of how extra explicit information sources can be utilized with an ASR framework to encourage voice access to online hunt files.

In [9] authors' precision is to prepare bigger acoustic models. There is a nonparametric system, the exact model that abuses rich preparing information to legitimately learn elocution variety. Introducing the exact model with a parametric model performs efficiently, with an overall improvement of 5.2% in WER [2]. There are various methods [7] in which this work could be expanded. In the first place, it reconciles with the acoustic model, preparing is probably going to be a keener appropriations and a more tightly fixed to the information. Secondly, evaluating word articulation co-event includes in semi-managed style (for example through word acknowledgement rather than constrained arrangement) would widen its relevance to a wide scope of various types and errands.

The observations from an overlooking literature review in order to present universal knowledge (theory and concepts) about voice control, digital assistants, [4] fields of use and more (Stufflebeam and Shinkfield, 2007). Computerized database key word looking became used on this review, As the technique is fast and efficient. The 3-Fundamental DB used for the studies changed into Google, Google Scholar and the KTH Publication database DiVA. Google turned into to discover general information as Google Scholar and DiVA were used to locate most research facts.

Toward the end, it is to be acceptable to change our models to factor with the unmistakable marvels that influence the elocution (for example complement, lingo, and acknowledgment blunders) [5]. This paper features AT&T Speak4it R voice search application which brings up the turn of events and advances on programmed discourse acknowledgment. Security is the significant issue with the voice-helped gadgets while anybody can access to a voice-actuated gadget. He can ask it inquiries, gathers data about the records and administrations related with the dynamic gadget, and request that it perform errands which can be unsafe and can cause misfortune as well. It comprises of high security hazard as these gadgets may

ISSN: 2005-4262 IJGDC

uncover the schedule substance, messages, and other individual subtleties.

It was accounted [6] for by an individual that, his iPad in his lounge would open the front entryway for any individual who remained outside and asked Siri to give them access. As of late Google has overhauled its Assistant delicate product. The updated form incorporates different highlights like voice printing, which extraordinarily recognizes every client by voice [8].

IV. Research Challenges

The purpose of this study is to research existing or conceivable controlling of voice, which specialized capacities to figure about that what potential impacts those establishments will create on both home and living. Few issues proclamations need to be replied: Are there any current or conceivable savvy home administrations actualized by voice control? Which potential impacts could execute voice order in brilliant homes have on the client?

V. Research Requirements

The different requirements of this research work are outlined as follows:

Hardware Requirements:

- i3 Processor or higher version
- 4GB RAM or higher
- Monitor

Software Requirements:

- Windows 7 or higher version of OS
- Kernel Version 3.0.16 or higher version
- Active Internet Connection (24 * 7)

VI. Research Set-Up

Virtual assistant permits making sure about the framework by Haar Cascade algorithm. The voice is to perceived by the mice and the face is to be recognized with the assistance of camera, which to be dealt with utilizing OpenCV module which assists with identifying the appearances and puts the articles before the camera and the picture conversely by Haar Cascade algorithm. After location of the face it grants to partners with the virtual framework.

It keeps the things under control for our voice orders. Here to the remote helper, the information is given in the position through Bluetooth associated gadget then framework changes over the system into the content configuration and comprehends the information and experiences handling and gives the suitable output in the content arrangement, Later it is changed over to the design as the output by the utilization of speech synthesizer and the output in the audio format.

VII. System Framework

System Design

It is characterized as the way towards defining the segments for example designs, modules, interface, and information for an offered framework to fulfill determined necessities. Framework configuration could be viewed as the use of framework hypothesis to item advancement. It covers with the control of a framework investigation, framework design and framework building.

Principles of System Design

Try not to be a human:

ISSN: 2005-4262 IJGDC

Playing lure switch with a client can cause them to feel that they been tricked, or that they don't see how a framework functions; both are terrible encounters. Maintain a strategic distance from, indicators or fake postponements to cause the UI to appear to be increasingly human interaction between the client and the bot framework plainly marked in a manner that conveys progressively human.

Keep It Simple:
Discussion ought to be limited to specific subjects and follow straight discussions streams and do maintain a strategic distance from muddled stretching ways. It's alright to uncover and clarify confinements ass users will tire of complicated passage of dialogue.

Utilize structured information when conceivable:
Try not to put clients in a circumstance where they have to figure the right mantra required to continue. Custom soft consoles license a restricted scope of information and can spare a lot of composing. For example as opposed to requesting that the end client type "YES" or "NO".

System Architecture Design

System Architecture Design divided into two phases i.e. Face Detection and Voice Recognition

Face Detection:

For face detection we create a database in which all the training images are stored and the captured images to be stored in database. The face detection is completed by two algorithms i.e. Haar Cascade Algorithm and Fisher Face Algorithm.

The Haar Cascade algorithm relies upon Course classifiers which contain Haar features. The course classifier is interlinked with heaps of classifiers which assist with distinguishing the human face it subject to the most important highlights like eyes, eyebrows and lips. Haar features are recognized and relies upon the strong computation, wherein we dole out a pixel power to every single pixel related to grayscale values inside the degree of 0 to 255 where 0 for the white and 255 for the dull concealing.

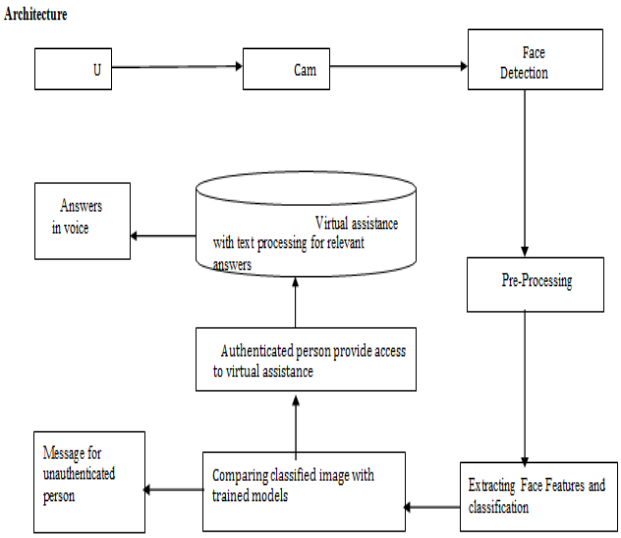


Fig.1 Face Detection System Architecture

Voice Recognition:

ISSN: 2005-4262 IJGDC

Personal Assistant consists of various modules services, but the salient feature of the assistant is that Voice Recognition can function without active internet.

VIII. PROPOSED METHOD

Algorithms:

Voice Recognition: *Python Speech Recognition Module*

This algorithmic program are the modules or the packages for the recognizing the voice.

1. "sudo pip install SpeechRecognition" PyAudio(For Linux Users)
2. "sudo apt-get install python-pyaudionpython3-pyaudion"PyAudio(For Windows User)
3. " pip install pyaudio**"**

Face Detection: Haar Cascade Algorithm

For Face discovery we use Haar Cascade calculation which includes in the accompanying advances:

1. In the Haar Cascade estimation relies upon Course classifiers which contain Haar features which are in Haar record helps for the acknowledgment of face.

2. The course classifiers are the blend of a lot of frail classifiers used to make a solid classifier.

3. This mix frames a square shape which comprises of highly contrasting recognizable proof lines on the face or the picture.

4. By utilizing Cascade Classifiers it additionally distinguishes grin, eyes regardless of countenances.

Syntax of Haar Cascade Algorithm:

- face_cascade = cv2. CascadeClassifier('haarcascade_frontalface_default.xml')
- eye_cascade = cv2. CascadeClassifier('haarcascade_eye.xml')
- smile_cascade = cv2. CascadeClassifier('haarcascade_smile.xml')

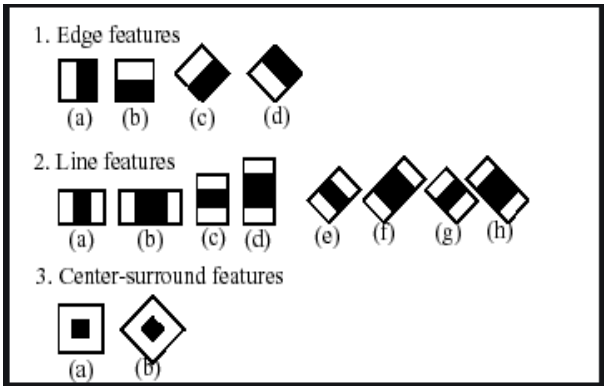


Fig. 2 Components Utilized and their Features

This section utilized in a specific classifier is exhibited by its shape (1a, 2b, and so forth.), position inside the zone of intrigue and the scale (this scale isn't equivalent to the scale utilized at the affirmation stage; at any rate these two scales are duplicated). For instance, by ethicalness of the third line include (2c) the response is settled as the contrast between the total of picture pixels under the square shape covering the entire part (counting the two white stripes and the dull stripe in the center) and the whole of the picture

ISSN: 2005-4262 IJGDC

pixels under the diminish stripe extended by 3 so as to make up for the separations in the size of zones. The aggregates of pixel respects over rectangular districts are settled quickly utilizing pivotal pictures.

Face Detection: Fisher Face algorithm:

This involves in training the images as it is stored datasets. It creates the different records one for pictures/images and other ones images with their comparing names. It encircle the picture and their id for every single sub-directory in the dataset. Then create a numpy array common for both the lists. OpenCv helps to train models for images with respective id using function. Picture affirmation using Fisher Face procedure relies upon the reduction of face space estimation using Principal Component Analysis (PCA) strategy. By then we apply Fisher's Linear Discriminate (FDL) system or in any case Linear Discriminate Analysis (LDA) procedure to get the picture recognized..

Different Features of Fisher Face Algorithm:

1. It tells about a class-explicit change grid.
2. It analyzes discriminately the facial features to compare between the persons.
3. It heavily depends on the input data as well.
4. It also allows for a reconstruction of the projected image.

There are three built-in face recognizers for OpenCV among which we can use any of them by a single line of code. Different recognizers are:

- EigenFaces-cv2.face.createEigenFaceRecognizer()
- FisherFaces – cv2.face.createFisherFaceRecognizer()
- Local Binary Patterns Histograms (LBPH) – cv2.face.createLBPHFaceRecognizer()

Stage 1: Retrieve information

Collection of information is done in type of face pictures. Though, collection should be possible utilizing photos effectively spared or from a webcam. Face must be completely obvious and must look ahead.

Stage 2: Processing of Image

a) Pre-processing stage: Getting pictures using camera or saved pictures and change from RGB to grayscale. Picture data is apportioned into getting ready and test data.

b) Processing stage : Fisher Face technique will be applied to deliver feature vector of facial picture data used by system and a short time later to facilitate vector of properties of planning picture with vector typical for test picture using Euclidean formula.

Stage 3: Feature generation

In this stage highlights of the images are extracted.

Recognition process

After the preparation is done, the following stage is picture acknowledgment process. The objective is to effectively perceive the test picture.

ISSN: 2005-4262 IJGDC

- In this phase of training of the picture is comparable to the testing picture. For this situation the system can successfully recognize the test picture precisely up to 100%.

If the training picture isn't equal to the testing picture at that point testing picture and the training picture must begin from the image of a comparable person's face. Framework can now viably perceive the test picture precisely up to 90%.

IX. MODULES FOR IMPLEMENTATION

There are several modules that are used in the project which can be given as follows.

datetime: In python this module supplies classes to work with date and time. And helps to deal with the dates, times, and time intervals.

OpenCv: This module sorts the image and analysis the video like face detection, license plate reading, optical character recognition, and so forth.

Httplib: This module helps for the implementation of the client side of the Hyper Text Protocols (HTTP and HTTPs). Generally, It can't be accessed directly but uses URLLIB which uses to handle URLs.

Urllib2: This module defines the functions and helps to open the URLs but mostly HTTP i.e. Authentication, Redirections, Cookies and more.

Json: It's a JavaScript Object Notation which is information exchange format determined by JS. It uncovered an API natural to clients of the standard library marshal and pickle modules.

Subprocess: This module allows spawning new processes, which connect to their onput/output/errors and obtain their return codes. This module used to replace several older functions and modules as well.

Cv2: This python method loads an image from the specified file and read it. It consists of three flags which are as follows i.e.
cv2.IMREAD_COLOR, cv2.IMREAD_GRAYSCALE, cv2._UNHANGED

Numpy: Numpy is a broadly useful cluster preparing bundle which gives a multidimensional exhibit item, and devices for working with these clusters. It's an essential bundle for logical figuring with the python.

OS: This module provides a way of using operating system dependent functionality. OS module allows us to interface with the underlying operating system that on python runs.

X. OUTPUT SCREENS

In this section we describe the output results of the desktop assistant.

Voice Recognition:
In this phase a user speaks to the desktop assistant, which recognizes the voice command, encode it and answers to the user.

ISSN: 2005-4262 IJGDC

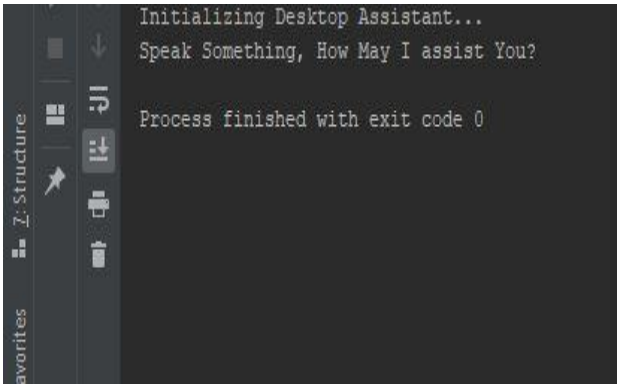
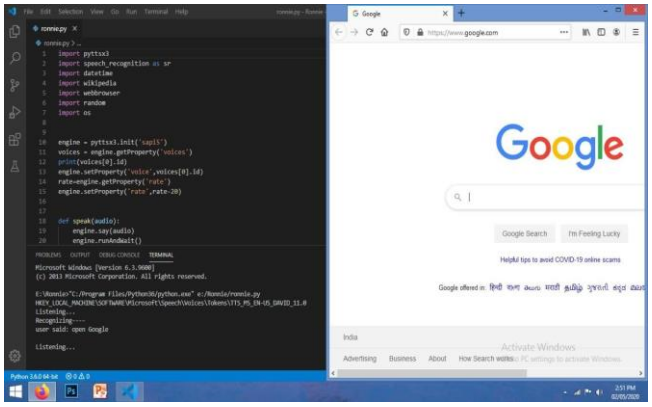


Fig.3 Desktop Assistant (Voice Recognition)

Flow Chart of the Voice Recognition:

Fig.4 Flow-Chart of Voice Recognition



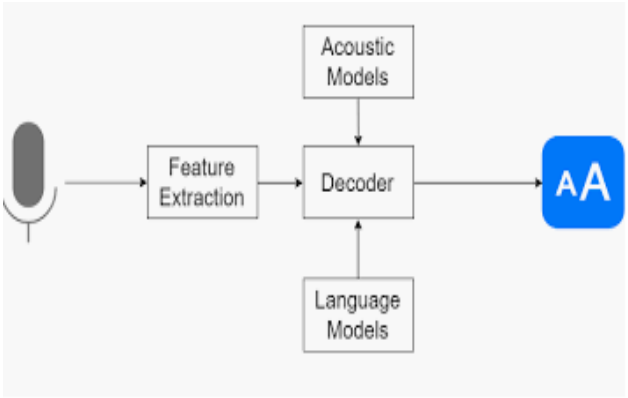


Fig.5 Desktop Assistant Voice Recognition (Open Google)

Face Detection:

ISSN: 2005-4262 IJGDC

In this phase, camera clicks the picture of the user at that point then it trains the module for the picture and put in the DB and

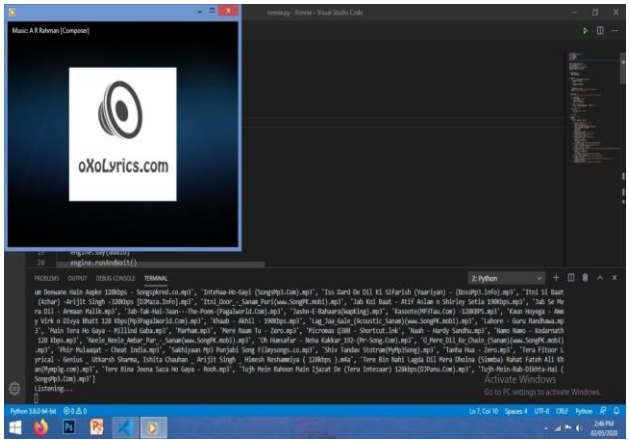


Fig. 6 Desktop Assistant Voice Recognition (Open Music System)

matches the highlights of that individual whose image will detected. With the help of Haar Cascade Algorithm it detects the face of the person on the Haar xml file and then matches the image with the trained models.

Flow Chart for Face Detection:

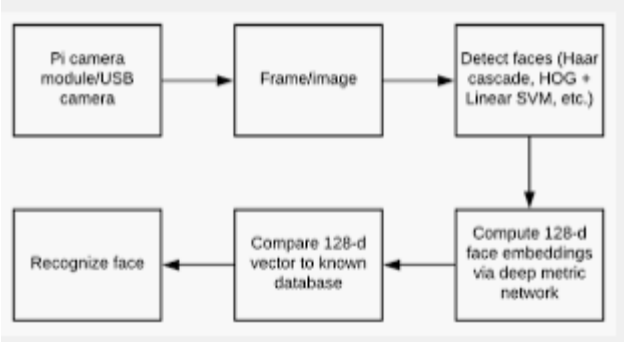


Fig.7 Flow Chart for Face Deteaction

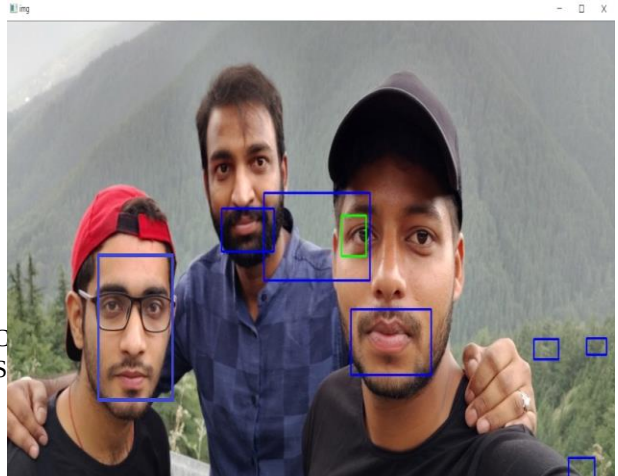


Fig.8 Face Reorganization using Desktop Assistant

XI. ADVANTAGES, DISADVANTAGES AND APPLICATIONS

Advantages

- It helps to save the time by assigning the tasks, also can assign the tasks repeatedly.
- Helps to deal with daily routine, for example: by assigning the task for ringing the alarm at certain time, also can schedule the appointment and so forth.
- It makes the work faster easier and in the efficient manner.

Disadvantage

- Language issue may arise.
- Strong risk of downtime.
- May sometime cause network glitches.

Applications

- Desktop Assistant can be used in the day-to-day routine which can save the time.
- This assistant can likewise be installed on the site of the corporate areas, lodging and the executives businesses for questionnaires and so on.

XII. FUTURE PROSPECTS AND CONCLUSION

How rapidly the time changed? If we look back around twenty year ago, Voice Recognition was in its infant stage. As when the computer system came into the existence, it was the dream to fully fledged interaction with the computer machine. Now, we can eventually talk, ask and as well order to do the assigned task.

This technology advancement is taking the world to the next level. If we think about the future competence of the voice recognition as well as the face detection, it can help the security agencies which may help to verify the details of the criminal and so forth. If we compare, two decades ago our words may very carried as we could have imagined.

Desktop Assistant has various functions as like of mobile phone like managing various applications on the voice commands. It helps to access the system hands-free and to get rid of typing chaos. By the use of facial recognition an individual can access the system, with the help of the face detection helps to secure the data, as no other person can access the system. It strategies the machine learning and help the user to access securely.

REFERENCES

[1] Xin Lei, Andrew Senior, Alexander Gruenstein and Jeffrey Sorensen “Accurate and Compact Large Vocabulary Speech Recognition on Mobile Devices,” in INTERSPEECH. 2013, pp. 662–665, ISCA.

ISSN: 2005-4262 IJGDC

[2] Alex Graves, Santiago Fern´andez , Faustino Gomez , and J¨urgen Schmidhuber "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks," Proceedings of the 23 rd International Conference on Machine Learning, Pittsburgh, PA, 2006 .
[3] Brian Kingsbury, "Lattice-based Optimization of Sequence Classification Criteria for Neural-Network Acoustic Modeling" 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan.
[4] Daniel L. Stufflebeam, and Anthony J. Shinkfield, "Evaluation Theory, Models, and Applications", John Wiley & Sons, 2007.
[5] David Rybach, Michael Riley, and Chris Alberti, "Direct Construction of Compact Context-dependency Transducers from Data," Proc. of INTERSPEECH, pp. 218–221, 2010.
[6] K. Beulen , and H. Ney, "Automatic Question Generation for Decision Tree Based State Tying," Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98.
[7] Patrick Nguyen, Georg Heigold, Geoffrey Zweig, "Speech Recognition with Flat Direct Models," IEEE Journal of Selected Topics in Signal Processing, Volume: 4 , Issue: 6 , Dec. 2010.
[8] Klaus Beulen, Hermann Ney, "Automatic Question Generation for Decision Tree Based State Tying," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98.
[9] Efthimios and Constantino"Monkey says, Monkey Does-Security and Privacy on VoiceAssistants", IEEE journal, 2017, ISSN: 2169-3536.

ISSN: 2005-4262 IJGDC