First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Voice-Based Intelligent Virtual Assistant for Windows using Speech Recognition and Speaker Identification Technology

Vrushali Kolte[1], Samidha Jadhav[1], Kalyani Kasar[1], Prof. Ashwini Pandagale[2]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India
[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Voice assistants developed by big companies like Cortana by Windows, Alexa by Amazon, Siri by Apple, and Google assistant by Google which performs the task ordered by the user. It can play audio-video, search for something, book flight tickets, can have fun conversations, etc. But the downfall of such technologies is the security issue as it stores its data on the cloud which can be retrieved by any technique and can be misused. Another issue is battery backup as most of the assistants do not have an inbuilt battery to work even in light cut-off situations. To overcome the above issues this project introduces Intelligent Virtual Assistant (IVA).

IVA – Intelligent Virtual Assistant which not only follows her boss's order but also gives the next response using her artificial brain. This response may advice, motivation, choices, legal actions, etc. It comprises of gTTS, pyttsx, AIML [Artificial Intelligence Markup Language], and Python-based state-of-the-art technology. The Internet has made information more accessible over a wide network, thus making it quicker and vaster, revolutionizing how people communicate in the world. The information available on the Internet about a given topic may be extensive, which helps in finding the solutions to day-to-day problems. And hence this project is based on forming more communication interactive models with the use of gTTs and AIML, facilitating the establishment of considerably smooth dialogues between the assistant and the users. Also, IVA stores its data in the user's PC this eliminates security problems.

IVA is also capable of recognizing the user's voice its gesture. This project not only gives logical or technical output but also an emotional one. IVA is a software agent that can assist people in many of their daily activities. It is capable of retrieving information from databases to give suggestions to people on performing different tasks, deploying a learning mechanism to acquire new information on user performance. It can make assistance more reliable and efficient by collecting information autonomously from objects that are available in the surrounding environment. This project also consists of voice-based user verification using a deep learning framework that recognizes the user's voice by its timbre and pitch. To make this idea feasible, IVA uses many searching protocols, artificial intelligence, machine learning, deep learning, etc.

Keywords: Cortana, Alexa, Siri, Google Assistant, gTTs, pyttsx, AIML, artificial intelligence, machine learning, deep learning.

I. INTRODUCTION

The term virtual assistant was coined way back in 1950s even before Siri which was developed by Apple as a virtual assistant for android. The term virtual assistant or virtual personal assistant is an application program that can understand natural human language, speak natural language and complete electronic task for the end user [11].The main aim is to design a voice-based intelligent virtual assistant (IVA) that acts as a digital organizer to provide variety of services to its master with the use of various machine learning algorithms, which accept voice input, process it and provide desired output to user[12].Some virtual assistants are able to interpret human speech and respond via synthesized voices, Users can ask their assistants questions, control home automation devices and media playback via voice, and manage other basic tasks such as email, to-do lists, and calendars with verbal commands, provide Profile management, Reminders [13], etc.

This intelligence system is classified into three generations: First Generation based on Pattern Matching; Second Generation including techniques of Artificial Intelligence such as deep neural network; Third Generation indulging higher ordered, sophisticated pattern matching techniques which is mostly based on AIML, a markup language for chat-bots constructions which is based on XML[3].

Following are the technologies used in Virtual assistant:

A. Speech Recognition

Speech Recognition is the ability of machine/program to identify words spoken aloud and convert them into readable text.

The main building blocks of speech recognition system are: Signal preprocessing Feature extraction, Language model, Decoder and Speech Recognition[5].

B. Speaker Identification

Speaker identification is the computing task of validating a user's claimed identity using characteristics extracted from their voices.

The two main modules of speaker identification are feature extraction and feature matching. Feature extraction is used to differentiate speaker according to their pitch and tone. In testing phase first features are extracted and then they are matched with the speaker templates using feature matching module[9].

II. LITERATURE SURVEY

Here we have discussed the literature survey of some of the existing techniques.
Giancarlo Iannizzotto et al. [1] proposed A vision and speech enabled, customizable, virtual assistant for smart environments. In this paper, an architecture is created that can assist the user not just only by voice but also by taking vision of the users. This paper introduces an architecture for building vision-enabled smart assistants, provided with expressive and animated graphical characters and speech recognition and synthesis. The limitation of this system was lack of accuracy in recognizing speaker and

system can be trained for fall detection and anomalous behaviour.

Veton Këpuska and Gamal Bohouta [2] proposed Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In this Proposal, an approach is proposed that will be used to design Next generation of Virtual personal assistants, increasing interaction between users and the computers by using the Multi-model dialogue system with techniques including gesture recognition, image/video recognition, speech and conversational knowledge base, General knowledge based. The advancement of the system is to achieve more accurate result and customer service home automation.

Neha Madhavi et al. [3] proposed JARVIS: An interpretation of AIML with integration of gTTS and Python. The system comprises of three modules first is based on Pattern Matching, second includes techniques of Artificial Intelligence and third includes indulging higher ordered, sophisticated pattern matching techniques, based on eXtensible Markup Language (XML). All this module has worked on achieving high accuracy of the system. But there are some limitations of this system that is AI should be transformed from stable AI to dynamic self-learning AI system which will be more potent to the system itself.

Artur Dovbysh and Vladyslav Alieksieiev [4] proposed Development and integration of speech recognition tools into software applications and an approach to improve of speech recognition quality. The purpose of the current research paper is to improve existing free to use tools of speech recognition and develop the approach to achieve both speech recognition engine and modules than bring recognition improvements with custom application. This architecture mainly includes speaker's diarization, Word correction text filter which includes Levenshtein's distance technique. To improve speech recognition efficiency, a primitive Python Neural network was implemented and using the TensorFlow training Library. The weak point of the system was differentiation of speakers' accents.

Ashok Kumar and Vikas Mittal [8] proposed Speech Recognition: A Complete Perspective. This paper introduces building blocks for general speech recognition system i.e. Signal pre-processing Feature extraction, Language model, Decoder and Speech Recognition. Challenges in speech recognition like environment factors, Vocabulary solves where overcome using above building blocks. Fields like local language or foreign language recognition with efficiency has to work still.

Sanket Shah et al. [6] proposed Speech recognition using neural network. This paper focuses on deep neural network which is a modern approach for recognizing the speech. Various neural network such as DNN, RNN and LSTM is discussed in the paper. The LSTM algorithm complexity per weight and time step is O(1). This is the extreme advantageous for high speed working of system and outruns the other approaches such as RTRL. The advancement in this field is that it can be used in various languages depending upon the region being used.

Nisha [7] proposed Voice Recognition Technique: A Review paper. In this paper biometric technology is used to recognize a particular individual's voice. The speech waves of particular voice form the basis of identification of speaker. Voice has its unique characteristics called feature & the process of extracting these features from the individual voice is called feature extraction. Techniques used in this system are Linear Predictive Coding, Perceptual Linear Prediction, Relative Spectral Filtering and Mel Frequency Cepstral Coefficient. Features are used to train a classifier so that it can classify the words which are spoken by the subject. Various classifiers which are used in voice recognition system are Hidden Markov Model, Neural Network Model, Dynamic Time Warping and Vector Quantization. This paper provides review of various voice and speaker recognition systems.

David Snyder et al. [8] proposed Speaker Recognition for Multi-Speaker Conversations using X-Vectors. This paper introduces x-vectors, an embedded neural network that is very effective for both speaker recognition and diarization. Speaker diarization is the process of grouping segments of speech according to the speaker, and is sometimes referred to as the "who spoke when" task. This system solves problem of speaker recognition on multi-speaker conversation

Satyam P. Todkar et al [9] proposed Speaker Recognition Techniques: A review paper. This paper has aimed to identify the speaker from different speakers available or verify a particular speaker. The voice of every individual sound is different as they are attributed to different features that create the voice, this may be- pitch, length of the vocal tract, sound frequency etc. Algorithms used are Vector Quantization (VQ), Gaussian Mixture Model (GMM), Neural Networks (NN), Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), GMM & Pitch Detection Algorithm, Hidden Markov Model (HMM). The system was able to cut off background noise using MFCC algorithm.

Garima Sood et al [10] proposed A Robust Speaker Identification System for Natural and Whispered Speech. This paper solves the main challenge lies in improving the robustness of the system in highly noisy environment. using, different identification algorithms for both normal and whispered speech have been compared to check the robustness. The extracted features have been classified with various classifiers such as Support Vector Machine, Fine K Nearest Neighbour (KNN) and Weighted KNN. Again, the limitation of the system was acquiring high accuracy.

Below are some of the highlighted researches in speaker recognition and speaker identification field:

Sr. No.	Author	Techniques used	Challenges	Constraints
1	Giancarlo Iannizzotto et al.	Mary TTS module Flite2 module	To make interaction more natural and attractive.	Fall detection or anomalous behaviour.
2	Gamal Bohouta et al	Next-generation of VPA model	To establish a lengthy conversation with user.	System can be more accurate and can be used in customer service home automation.
3	Neha Madhavi et al.	AIML and Google API	Interpretation of AIML.	To transform stable AI assistant into a dynamic self-learning AI system.
4	Artur Dovbysh et al.	Speaker Diarization	To improve speech recognition quality.	Differentiation in speaker's accent.
5	Ashok Kumar et al.	Linear predictive coding (LPC), PLP and Mel Cepstral Coefficients (MFCC)	Challenges in speech recognition like environment factors, Vocabulary solves.	Fields like local language or foreign language recognition with efficiency has to work still.
6	Sanket Shah et al.	Speech classification techniques.	To obtain highest accurate results.	Speech recognition can be used for various languages.
7	Nisha	ASR System	To gain highest result from various feature extraction techniques.	Access control or Transaction Authentication.

8	David Snyder et al.	Speaker diarization, x-vector, AHC.	Speaker recognition for multi-speaker conversation.	AHC can be replaced with alternative method.
9	Satyam P. Todkar et al.	Feature matching and extracting techniques.	To identify speaker from different speakers.	Useful in noisy environment.
10	Garima Sood et al.	SVM AND K-NN	Speaker identification for natural and whispered speech.	Real time training and useful with IOT devices.

III.CONCLUSION

This paper introduces voice-based intelligent virtual assistant specifically designed for windows operating system. In this system we have integrated both speech recognition and speaker identification technology. This IVA system uses speech, audio, video and other mode of communication to interact with human. Further this system can be used in different fields such as home automation, medical assistance, car automation, robotics and security access.

IV. REFERENCES

[1]. Giancarlo Iannizzotto, Lucia Lo Bello, Andrea Nucita, Giorgio Mario Grasso “A vision and speech enabled, customizable, virtual assistant for smart environments” Dept. for Cognitive Sciences, Psychology, Education and Cultural Studies (COSPECS).

[2]. Veton Këpuska, Gamal Bohouta Electrical & Computer Engineering “ Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home) ” Department Florida Institute of Technology Melbourne, FL, USA

[3]. Ravivanshikumar Sangpal, Tanvee Gawand, Sahil Vaykar, and Neha Madhavi, of Computer Technology, Government Polytechnic Pen “JARVIS: An interpretation of AIML with integration of gTTS and Python ” 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT).

[4]. Artur Dovbysh, Vladyslav Alieksieiev “Development and integration of speech recognition tools into software applications and an approach to improve of speech recognition quality ” 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET).

[5]. Ashok Kumar, Vikas Mittal “Speech Recognition: A Complete Perspective” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6C, April 2019

[6]. Hardik Dhudrejia, Sanket Shah “Speech Recognition using Neural network” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2278-0180, Volume-7 Issue-10, October 2018

[7]. Nisha “Voice Recognition Technique: A Review” International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653 Volume 5 Issue V, May 2017.

[8]. D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, “X-vectors: Robust dnn embeddings for speaker recognition,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

[9]. Satyam P. Todkar, Snehal S. Babar, Rudrendra U. Ambike, Prasad B. Suryakar Department of Computer Engineering Sinhgad College of Engineering Pune, India “Speaker Recognition Techniques: A review” 2018 3rd International Conference for Convergence in Technology (I2CT), Apr 06-08, 2018.

[10]. Garima Sood, Sidharth Pancholi, Amit M. Joshi Electronics and Communication Engineering Department National Institute of Technology, Hamirpur Himachal Pradesh, India “A Robust Speaker Identification System for Natural and Whispered Speech” 2017 IEEE.

[11]. Virtual assistant: What is it, 10 2017 onlineAvailable: www.searchcustomerexperiences.techtarget.com

[12]. V.Lalitha, A.Dinesh , L.Parameswaran, S.Dinesh Kumar, Department of CSE, Sri Sairam Engineering College, Chennai “ML Based Virtual Personal Assistant” International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 6, Issue 7, July 2019.

[13]. Prajyot Mane, Shubham Sonone, Nachiket Gaikwad and Prof. Jyoti Ramteke Computer Engineering Dept., Sardar Patel Institute of Technology, Mumbai, “Smart Personal Assistant using Machine Learning” International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).