MANTECH Publications
Journal of Artificial Intelligence, Machine Learning and Soft Computing
Volume 8, Issue 1, January-June 2023

Pixie – A Speech and Gesture Enabled Virtual Assistant

Rahul Patil [1], Sandip Chavan [2], Ameya Bhupendra Deodhar [3], Chinmayi Kamalakar Juikar [4], Pradnya Sham Jagtap[5], Samruddhi Murlidhar Jadhav[6]

Assistant Professor [1, 2], Students[3, 4, 5, 6]

Department of Computer Science Engineering

Bharati Vidyapeeth College of Engineering, Navi Mumbai India

Email: *rahul.patil5@bharatividyapeeth.edu[1], sandip.chavan@bharatividyapeeth.edu[2], adeodhar04@gmail.com[3], chinmai.juikar@gmail.com[4], pradnyajagtap2001@gmail.com[5], 360jadhavsam@gmail.com[6]*

Abstract

In today's pace forward generation, it is convenient and functional to make daily tasks automated and digitized. Digitization opens many possibilities to make our repetitive tasks easier using assistive technology. Artificial assistants make use of machine learning, artificial intelligence and natural language processing to provide a personalized and conversational experience. This paper discusses about Pixie which is a desktop voice assistant aiming to provide a personalized, interactive and secure experience. Pixie also has face recognition system and is gesture enabled. Voice assistants are an emerging technology with a great future. In this paper, we discuss the development of a voice assistant in desktop.

Keywords: *Artificial Intelligence, Virtual Assistant, Natural Language Processing, Machine Learning.*

INTRODUCTION

In a generation where everyone is leaning towards automation, virtual assistant is a technology which has a huge scope and numerous possibilities to conquer. Virtual assistant is used to operate on voice command. A virtual assistant is an artificial intelligence program that can respond to voice and natural language commands to carry out tasks for users that have been set up in the software. The virtual assistant performs tasks such as open google, take a screenshot, open camera, search Wikipedia, send emails and much more. Virtual

assistants perform these tasks by voice command and hence save time and increases productivity. They are highly task-focused. Virtual assistants can comprehend and respond to requests. It is software that can interpret spoken and written commands and carry out user-defined inquiries. Virtual assistants can comprehend spoken language and answer with synthesized voices. To make it secure, pixie has a face recognition system which allows only authorized personnel to access the software. Virtual assistants as a technology have a lot of potential. It has a large scope of research and lot of challenges to cater to.

RELATED WORK

Microsoft Cortana – Microsoft created Cortana, a virtual assistant that leverages Bing to carry out activities like creating reminders and assisting users with enquiries. However, Cortana has some security flaws and is believed to be always listening to the user.

Alexa – The virtual assistant Alexa is a feature of the Amazon Echo range of smart speakers. However, Alexa has some issues like it does not work on batteries, it takes several days to get an update, others can access the echo conversation too.

Siri – Siri is the most widely known virtual assistant and is considered as Apple's brainchild. It is compatible with the iPhone, iPad, MAC, Apple Watch, Apple TV and the official speaker lineup of the business. Yet, it is said that Siri has trouble hearing your voice in noisy environments.

Google Assistant – Google Assistant is available on many android devices.

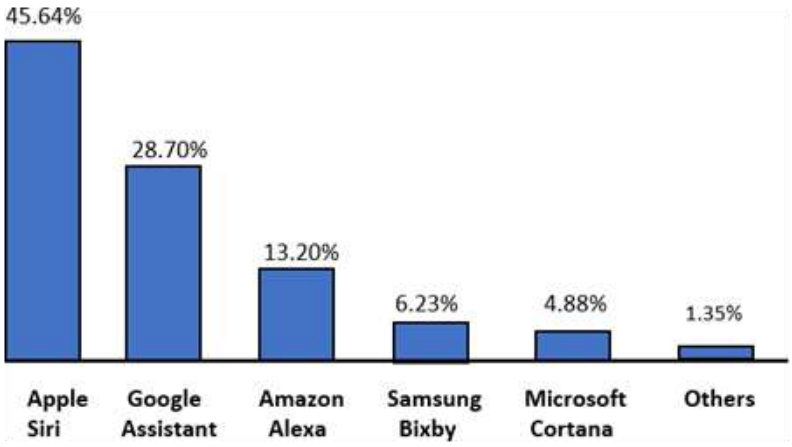

Figure 1: Virtual Assistant popularity

It also lends it services to some third-party applications such as JBL. However there have been issues like high data usage, increased heating of devices over a prolonged period of use.

LITERATURE SURVEY

This article describes software architecture for developing portable, voice, vision, and automation applications for virtual assistants for smart homes. The authors of this paper designed a virtual assistant architecture for smart home automation systems using some of the most modern techniques in computer vision, deep learning, speech production and recognition and artificial intelligence. [1]

The creation of an open-source virtual assistant that can be taught new commands by voice in natural language is briefly discussed in the paper. The agent stated in a document can be fully operated from a distance. They can also gauge an agent's effectiveness and how well they are received by users.[2]

The voice assistant PARI was created specifically for blind people and responds to their spoken commands. Without an online connection, the assistant can understand voice instructions. As it contains essential features like Speech Pattern Detection, Keyword Learning, etc., it actively reacts to user voice queries faster than Online Voice Search tools. [3]

What are voice assistants, what can voice assistants do, history of voice assistants, different assistants available in the market, security and privacy issues in them, future uses of speech recognition technology all the factors are clearly mentioned by the authors of the paper.[4]

The researchers tried to examine the characteristics of usability and the pleasure felt by users of the VAs who speak English as a second language in comparison to native English speakers. Regardless of the dialect, the VAs can understand spoken English orders and reply accordingly. Initially, users of VAs can be set up in a range of demographic configurations, including English (Australia), English (UK). Even though neither Siri nor Alexa were native English speakers (Thailand), they had no trouble understanding English in a variety of accents. [5]

The paper provides brief idea about the smart personal assistants. The paper's major area of interest was natural language interaction with mobile smart personal assistant systems. They provide details about the many personal assistants in the market.[6]

This paper aimed to synthesize previous studies that addressed the security and privacy implications of well- known virtual assistants, examine their findings, and draw conclusions. Research has been done on user security issues, malicious attack risk, and authentication enhancement.[7]

This paper discusses the fundamentals of speech recognition, various types of speech recognition that have been proposed, issues with speech recognition, various practical methods for detecting the point of a speech signal, each of which has advantages and disadvantages, and vibrant pattern matching techniques for capturing the speech of various speakers.[8]

A wake word similar as 'Alexa' is spoken by the use to start conversation with the PVA. Authors used the audio recording of the wake word to determine the room in which interaction between user and assistant takes place. It includes data collected from 10 different apartments in which a user speaks the wake word at different locations. This dataset is used to estimate three different neural network- grounded algorithms for room identification.[9]

Because SPA research appears to be highly fragmented across disciplines such as computer science, human- computer interaction, and information systems, the authors have identified five functional principles and three research domains that appear capable of future study, particularly in the information systems field. The study includes a preliminary research summary as well as a methodical approach to generalizing findings from IS, HCI, computer science, and other domains. [10]

FUNCTIONALITIES

1. **Play songs –** Pixie can play songs on commands. The song can either be played directly from net or can also be played form a preinstalled application in the desktop.

2. **Setting alarms on desktop –** Alarms can be set, snoozed, or dismissed using Pixie.

3. **Display time and weather –** Pixie has the access to system's time and date; it fetches the time from the system's local time. At the same time, it has the access to system's location center and can easily access the location from there which is then used to give the weather updates for that location.

4. **Automation of Whats App messages –** There is no need to manually type in and send the messages. Pixie automates the process of sending messages for you.

5. **Automation of Gmail –** Currently we have automated the process of sending emails for Gmail only. It has a database of email ids of the recipients which is mutable at any given time. Hence, it fetches the email id of the recipient from the database and sends the mail automatically without the physical work.

6. **Taking Screenshots –** It is enables to take screenshots as well. It can take screenshots via both voice command as well as manually through keyboard.

7. **Running any system application –** It can invoke and run any system application. For example, it can run calculator, alarms, python application, Google chrome, to do lists, etc.

8. **Open websites on internet –** It can open different websites such as "python.org", www.google.com, www.youtube.com, www.stackoverflow.com.

9. **Tell jokes –** It can tell jokes from various libraries which contain a set of pre compiled jokes.

10. **Gesture Enabled –** It is gesture enabled virtual assistant. To use some selected features, the user has a choice of using voice command, manual opening, or gestures, etc.

11. **Facial Recognition –** It is a safety feature of Pixie. The user must be verified before being able to have access to the features of Pixie. It is the primary feature and the first step to use the Pixie.

METHODOLOGY

Technologies Used

1. **Python:** Python is a dynamic programming language used for data analysis, Machine learning and can be termed as an important aspect of Artificial Intelligence.

2. **Pyttsx3:** Python Text to Speech library is referred to as Pyttsx3. It enables text to speech synthesis by acting as a wrapper for the same. It provides a crucial benefit by working in offline mode.

3. **PyAuto GUI:** It automates the devices like mouse, keyboard, etc. and enables their use without human intervention.

4. **Speech Recognition:** It mostly recognizes the speech input for executing the corresponding output. The use of speech recognition gives PIXIE a touch of automation.

5. **Mediapipe:** It offers configurable, cross-platform solutions for live and streaming data. Support for Android, iOS, Windows, MAC, and other platforms is included in cross-platform support.

6. **MTCNN:** Multi-Task Cascaded Convolutional Neural Networks is a neural network for facial landmark recognition in pictures. It is predominantly used for Face Detection, Face Extraction and Face Classification.

7. **Haar Cascade Classifier:** Specially used for Object Detection. Haar Cascade classifiers can detect objects from images as well.

8. **Tensorflow –** Used as a subsidiary to complement and help in Image Recognition in Facial and Gesture recognition.

9. **Convolutional Neural Network:** CNN layers will be used to select the most appropriate model.

10. **Qt Designer:** For designing interactive GUIs for the virtual assistant.

11. **Wolfram Alpha:** Using Wolfram's algorithms, knowledge base, and AI technology, it computes expert-level responses to any command.

12. **Pyjokes:** Pyjokes is a library for collecting jokes online by using Python as a medium.

13. **Pyaudio:** PyAudio makes use of PortAudio which acts as a cluster which holds together various audios.

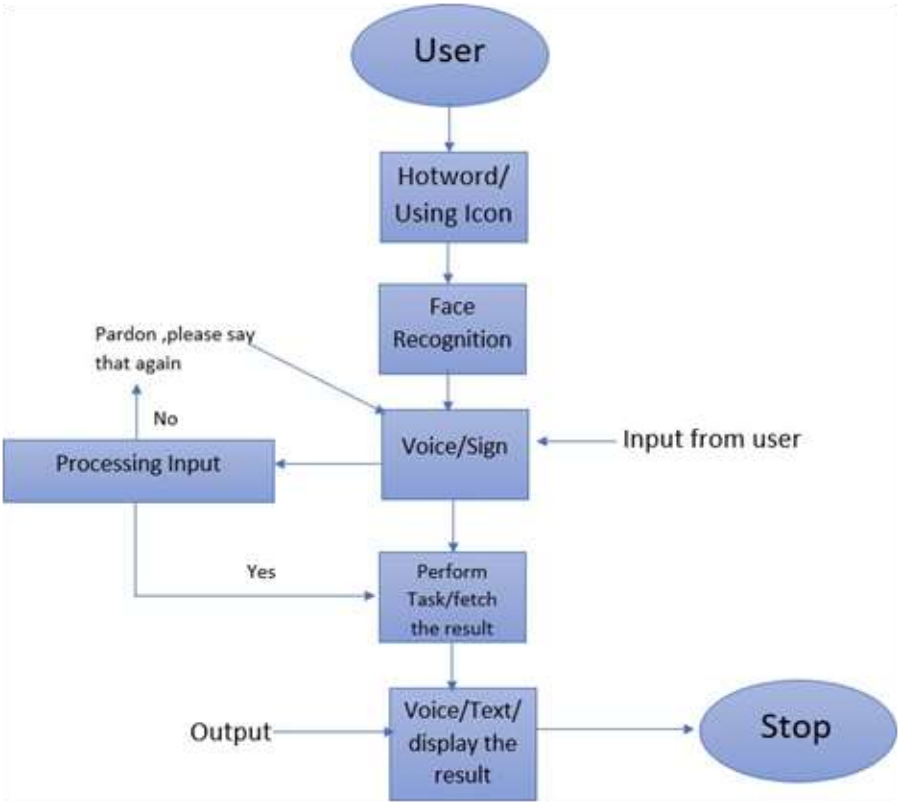

Figure 2: Workflow of Pixie

The user will have to initiate Pixie with by saying the hot word “Pixie” at first or by double tapping the icon for Pixie on their device. After entering Pixie, user will have to confirm identity by going through a facial recognition process. The facial records of legitimate user will be pre saved into the system. Once the user successfully goes past the facial recognition stage, all the services of Pixie are accessible to the user. They can be accessed either by voice or by sign or gestures. Depending upon the users’ commands, Pixie will execute the task and fetch the result. The result or output will be given in two ways parallelly, i.e., Pixie will give an audio output as well as it will be displayed on screen as captions.

CONCLUSION

The adoption of virtual assistant technology benefits people. A virtual assistant offers the freedom to just hire for the services they need. Due to their greater portability, loyalty, and availability at all times, virtual personal assistants are also more dependable than real personal assistants. We created a virtual assistant in Python for all Windows versions. It is similar to Alexa, Cortana, Siri, and Google Assistant. "Pixie" is a very useful virtual assistant based on cutting-edge technologies and algorithms that adapt to changing needs. Because it accepts input via keyboard, mouse, voice, and gesture, it is more effective and efficient.

It is intended to shorten the time required for people to communicate with numerous different subsystems that would otherwise have to be done manually. By doing so, the system will make human life more comfortable.

It currently works online to perform basic tasks such as searching Wikipedia, opening YouTube, viewing weather reports, receiving news updates, capturing photos, and soon. Pixie recognizes speech and gestures, which is extremely beneficial to the elderly, the blind, physically challenged people, children, and others. "Pixie" is integrated with devices such as desktops, making desktop usage compatible with the changing needs of different age groups.

REFERENCES

1. Iannizzotto, Giancarlo, et al. "A vision and speech enabled, customizable, virtual assistant for smart environments." 2018 11th International Conference on Human System Interaction (HSI). IEEE, 2018.
2. Chkroun, Merav, and Amos Azaria. "Lia: A virtual assistant that can be taught new commands by Speech. "International Journal of Human– Computer Interaction 35.17 (2019): 1596-1607.
3. Kulhalli, Kshama V., Kotrappa Sirbi, and Mr Abhijit J. Patankar. "Personal assistant with voice recognition intelligence." International Journal of Engineering Research and Technology 10.1 (2017): 416-419.
4. Hoy, Matthew B. "Alexa, Siri, Cortana, and more: an introduction to voice assistants." Medical reference services quarterly 37.1 (2018): 81-88.
5. Pal, Debajyoti, et al. "User experience with smart voice assistants: the accent perspective." 2019 10th International Conference on Computing, Communication and

Networking Technologies (ICCCNT). IEEE, 2019.

6. Ahmed, Shimaa, et al. "Towards more robust keyword spotting for voice assistants." 31st USENIX Security Symposium (USENIX Security 22). 2022.
7. Alepis, Efthimios, and Constantinos Patsakis. "Monkey says, monkey does: security and privacy on voice assistants." IEEE Access 5 (2017): 17841- 17851.
8. Saksamudre, Suman K., P. P. Shrishrimal, and R.R. Deshmukh. "A review on different approaches for speech recognition system." International Journal of Computer Applications 115.22 (2015).
9. Azimi, Mohammadreza, and Utz Roedig. "Wake word-based room identification with Personal Voice Assistants." 1st Workshop on Hot Trends in Embedded Systems Privacy (HTESP 2022), Linz, Austria, October 03, 2022. Association for Computing Machinery (ACM), 2022.
10. Knote, Robin, et al. "The what and how of smart personal assistants: principles and application domains for IS research." Multik onferenz Wirtschafts in formatik (MKWI) (2018).