APPLICATION OF VOICE ASSISTANT USING MACHINE LEARNING: A COMPREHENSIVE REVIEW

DEEPIKA PATEL and TORAN VERMA

Department of CSE

Rungta College of Engineering and Technology

Raipur, Chhattisgarh, India

E-mail: deepikapatel1209@gmail.com

toran.verma@rungta.ac.in

Abstract

Due to the fast growth of Artificial Intelligence and Machine Learning, the use of voice recognition technology has become more popular. Techniques such as Speech Recognition, Natural Language Processing, and Voice Synthesis are used to implement Voice Assistant systems. The goal of these systems is to provide a better service by translating information into a more interactive manner. The performance of a Voice Assistant depends on the speed and accuracy of human speech. Voice recognition is a challenging task due to the varying characteristics of a voice. This paper presents a comprehensive review of the various uses of Voice Assistant in various fields.

I. Introduction

A. Voice Assistant

The goal of a voice assistant is to perform tasks based on the user's spoken commands. Human uttering is a lot more natural than typing. An average person can type up to 35 words in a minute, but they can also say more than 100. The rapid development of AI and machine learning has allowed voice assistants to penetrate every part of our lives. They can perform various tasks without requiring the user to type in a keyboard. The main goal of a voice assistant is to make hands free and minimize the use of various input devices such as keyboards and mice. This paper aims to review a home automation system that uses speech recognition technologies to provide a personalized voice assistant for every user.

B. Speech Recognition

The process of speech recognition is the translation of a speech signal into a sequence of words that can be understood by the system. Voice assistant applications are based on the Automatic Speech Recognition (ASR) system. The process of speech recognition is performed by an ASR system, which records the speech and then breaks it down into phonemes. No matter which type of speech recognition software used, the performance depends in its implementation.

The generic model of the ASR depicted in Figure 1. The process of speech recognition begins with the device recording the audio. Then, the system performs an acoustic analysis to determine the meaning of the speech, which is performed on three different levels. An acoustic modeling, which takes into account the relationship between the phonemes and the words they contain. The pronunciation modeling algorithm is used to analyze the way phonemes are pronounced. It takes into account the various characteristics of the voice to produce a more accurate representation of the speech. The language modeling aimed at determining contextual probabilities relying on what phonemes were seized.

The decoder is used to find the most likely word sequence that was delivered by a speech transmission. It then transforms the data into text for future use, such as dictation or command. The low error rate is attributed to the use of AI. Speech Synthesis is used to allow the system or machine to speak. Natural Language Processing (NLP) is a branch of computer science which studies how to process and understand human language. It allows machines to understand and interact with speech. An automatic speech recognition system is used to convert a user's speech into text. Then, a natural language processing program is used to analyze the transcribed text and formulate a response. The voice assistant also includes the various capabilities of natural language generation, speech processing, and dialogue management. Speech processing and Natural Language Processing (NLP) using ML permit intelligent devices, such as smartphones, to interact with users via spoken language.
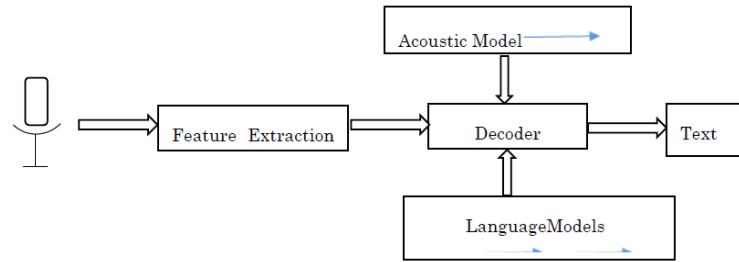
Figure 1. Generic Approach of an Automatic Speech Recognition.

II. Voice Assistant Using Machine Learning: A Review

The various aspects of the voice assistant are studied in this paper. Some of these include: AI-based voice assistant, voice assistant in education, voice assistant for gender recognition, and voice assistant for medical and health care.

A. Voice Assistant in Education

George Terzopoulos et al. [9] studied about abilities of voice assistants in the classroom. The authors believe that the lack of languages in the voice assistant makes it difficult for students to understand its functions. They also noted that the device does not have the necessary security measures to protect students. They also believe that teachers should be trained to use the voice assistant in their classes.

Rainer Winkler et al. [8] studied and answered two research questions. The goal of this study was to find out if using a variety of technology such as the SPA can improve student's problem-solving skills. They then conducted two experiments to see if this could happen. They found that the use of the technology over a period of five weeks significantly improved the student's skills. The second part of the study focused on how the use of the technology changes the learning process for students. They found that the students showed more interactive behavior when using the technology. The students also said that they benefited from the individualized support provided by the SPA. This study provides valuable evidence supporting the use of technology in the classroom. The findings of that study contribute to computer tutoring and technology-mediated learning research. Various use of Voice Assistant in Education Field are reflected in Table 1.

Table 1. Voice Assistant in Education.

Author Name and Year	AI Methods	Results
George Terzopoulos (2019) et al.[9]	NLP and cloud computing	Due to the lack of languages, voice assistants do not always speak the same language.
Rainer Winkler (2021) et al.[8]	SPA technology	The students were able to show more interactive learning behavior with the help of the SPA.

B. AI Based Voice Assistant

In their paper, Seema et al. [1] provide a comprehensive description of the various types and characteristics of automatic speech recognition. They also talk about the history of the technology and its future applications. In their paper, Yogendra Kumar Sharma et al. [2] presented a smart personal assistant that enables users to access e-mails and calendar updates using natural language. The user must first present the system and enable it to perform a dialogue in order to make the most of its capabilities. The SPA is designed to be used on an agent platform, which means it has a special controller that can handle the various tasks associated with the system. The agent-based dialogue model can be reused in different systems. Kshama et al. [3] presented a system that can work seamlessly without Internet connectivity. The system, which they referred to as the Personal Assistant with Voice recognition Intelligence, takes the user input and processes it. It returns various output options such as actions and search results. Sermakani (2021) et al. [6] talk about the main features of a desktop personal assistant that they built using AI technologies. One of these features is the ability to minimize the use of various input devices such as keyboards and mice.

Harshita Yadav et al. [7] has discussed that Artificial Intelligence-based Virtual Assistance is a software mechanism which can accomplish an individual's tasks according to pedagogy disseminated by a user. The term "chat bot is sometimes referred to assistants who are normally visible or especially accessible via online chat. A chat bot can be expressed as software

that disseminates with people using artificial intelligence. The summery of AI based Voice Assistant is shown in Table 2.

C. Voice Based Email System

Harsh et al. [15] proposed a system that allows users to login to their email account and send an email to a specific person using their voice commands and speech. The researchers focused on the use of artificial speech and a broad vocabulary to enable continuous speech in multiple languages. Rutuja et al. [4] discussed the various barriers that people with visual impairments face when it comes to communicating. They then developed a virtual assistant that can perform human voice commands without relying on others. This software allows users to receive and send emails, monitor their diary, and keep a record of their online activities. It can also recognize images using optical character recognition. Harivans et al. [13] developed a Voice-based email system that can be used by blind individuals. This system can provide them with the facility of communication and can also facilitate their access to other services. Rijwan et al. [14] developed an AI-based email system that will allow people with visual impairments to easily access their email accounts. The main objective of this project was to make the system more accessible. The summary of the review shown in Table 3.

Table 2. Voice Assistant in Education.

Name and Year	AI Methods	Results
SeemaRawat (2014) et al. [1]	Hidden Markov Model (HMM) and Mel Spectrum Cestrum Coefficient (MFCC)	Artificial intelligence (AI) based speech recognition improve the lives of people with physical limitations.
Yogendra Kumar Sharma (2016) et al. [2]	Natural language speaks through a PDA interface.	A smart personal assistant is a tool that combines the capabilities of AI and natural language interaction. It can perform various

		tasks and communicate with other agents.
Dr.Kshama V. Kulhalli (2017) et al. [3]	CTC based LSTM acoustic model	PARI is designed to manage various mobile applications by just a voice commands
Mrs. A. M. Sermakani (2021) et al. [6]	AI Technologies, ASR, TTS	AI based Digital Assistant is implemented which help visually impaired and physically challenged people.
Harshita Yadav (2021) et al. [7]	Open-source software module , modular model approach	The design and implementation of AI based Virtual Assistant is done which can accommodate additional feature.

D. Voice Assistant for Gender Recognition

Monica et al. [5] discussed the various techniques used in speech analysis to perform gender recognition. They also proposed a practice that would allow gender to be taken into account in the design of emotion recognition systems. Shivangee et al. [12] developed a system that can detect the frequency, amplitude, and other characteristics of a speaker by combining the vocal folds and articulations. They then used four models to classify the different genders. They habituated four models to categorize the genders namely CART, XGBoost, SVM and Random Forest. An ensemble of models is used to improve the accuracy of the system. This system can be used in various applications such as voice recognition and gender detection. Brief description about Voice Assistant for Gender Recognition is shown in Table 4.

E. Voice Assistant for Medical and Health Care

Arul et al. [10] presented a paper that focused on the use of machine learning and the cloud to improve the personal healthcare of patients. One of the main issues that people face while making an appointment is the time it

takes them to make the call. With the help of a smart personal assistant, they can easily schedule an appointment with a doctor at any time. This project aims to develop a smart personal assistant that can provide users with the ability to schedule an appointment with various doctors at anytime and anywhere. It will help users save time and improve their satisfaction. Alena et al. [11] developed a plausible scenario that will allow decision-makers to make more effective use of the virtual patient assistants (VIPAs) in health care. The project carried out a two-stage study, which involved the recruitment of experts from various backgrounds, including medical professionals, academicians, and health authorities. They then formulated and evaluated various projections. The summary of the review shown in Table 5.

Table 3. Voice Based Email System.

Author Name and Year	AI Methods	Results
Harsh D. Shah (2021) et al.[15]	Automatic Speech Recognition(ASR)	This project proposes a system that allows users to send email using their voice commands and the address book of the person they want to communicate with.
Rutuja V. Kukade (2018) et al. [4]	Speech to Text Engine, Text to speech Engine, OCR (Optical character recognition)	Implemented model perform tasks by taking human voice commands
Harivans Pratap Singh (2021) et al. [13]	Speech Recognition using AI, NLP	The success of this project could also encourage developers to create more effective and useful

		products for people with low vision.
Rijwan Khan (2020) et al. [14]	Artificial Intelligence for Speech Recognition	The project is to create a system that will allow people with limited mobility to grow along with society.

Table 4. Voice Assistant for Gender Recognition.

Author Name and Year	AI Methods	Results
Monica La Mura (2020) et al. [5]	Speech Recognition	Performed well in terms of gender recognition rate.
Shivangee Kushwah (2019) et al. [12]	CART, XGBoost, SVM and Random Forest	The project's success with 89% was attributed to the use of a stacked model, which was able to classify the voice as either male or female.

Table 5. Voice Assistant for Medical and Health Care.

Author Name and Year	AI Methods	Results
Arul Srinivasan (2018) et al. [10]	Natural Language Processing (NLP)	The proposed system provides quick confirmation of Appointment and easy interface where user can book appointment at any time with the doctor.

Alena Ermolina (2021) et al. [11]	Two-stage Delphi study	A panel of experts evaluated the project's projections and came up with a plausible scenario for the future of voice assistants.

III. Discussion

In real word environments, audio signals can contain channel and background noise. This can affect the accuracy of the system. To reduce the background noise and improve the audio signal's quality, a noise reduction technique is used. The system can then convert the words into strings of text by analyzing the difference between the background noise and the words. It can then send those words to the Voice Assistant.

A. Voice Assistant evolution.

In 1922, the first voice activated toy was released. It was called Speech Radio rex. In 1952 Bell Labs introduced the "Audrey" machine, which could recognize the basic units of speech. However, it was limited to recognizing digits that were spoken by designated talkers. In 1962, IBM introduced the first voice assistant at the World's Fair in Seattle. The company's product was called the IBM Shoebox, which was able to recognize numbers 0-9 and 6 simple commands. In the 1960s, a chatbot known as ELIZA was developed by Joseph Weizenbaum, a professor at MIT. The creation of the first machine learning system was motivated by the idea that the communication between humans and machines was superficial. During the 1970s, a team of researchers at Carnegie Mellon University developed a machine learning system known as the Harpy, which was able to learn about 1,000 words. It was able to reduce its errors by following pre-defined grammar structures and vocabulary. In 1986, the successor to the Shoebox was called the Tangora. It had a built-in vocabulary of 20,000 words and was able to predict the outcome based on what was said in the past. In 1997, a software program called Dragon's Naturally Speaking was able to recognize and transcribe human speech at a rate of about 100 words per minute [16].

In 2011, Apple's Siri became the first digital assistant to be featured on a smartphone. In 2018, the company introduced the HomePod, which became a smart speaker. Google Now, on the other hand, only works on the iOS platform. In 2016, Google introduced Google Now as its digital assistant. In 2013, Microsoft launched its Cortana voice assistant. In 2014, Amazon introduced the Echo, a smart speaker that can be controlled by a user's voice. The Alexa marketplace has been regarded as the most significant contribution to the evolution of the digital assistant space. Over the years, thousands of skills have been created for it, and developers have started making money from it. Despite the growing pains, Amazon's platform is well-positioned for the future. In 2018, Samsung launched its own voice assistant called the Bixby [17].

IV. Conclusion

The goal of this study was to analyze the various applications of the voice assistant and how it can be utilized in human-computer interaction. It found that it can help people save time and money. The study also discussed the various features and capabilities of the voice assistant. It also noted that the speed and accuracy of the voice assistant's performance are the most critical factors that determine its success.

References

[1]   Rawat, Seema, Parv Gupta and Praveen Kumar, Digital life assistant using automated speech recognition, In 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH), IEEE (2014), 43-47.

[2]   Sharma Yogendra Kumar and Neeraj Sharma, A review paper on smart personal assistant, International Journal of Engineering Research and Technology (IJERT) ISSN: 22780181 V-IMPACT 4(32) 2016.

[3]   V. Kulhalli Kshama, Kotrappa Sirbi and Abhijit J. Patankar, Personal assistant with voice recognition intelligence, International Journal of Engineering Research and Technology 10(1) (2017).

[4]   V. Kukade Rutuja, Ruchita G. Fengse, Kiran D. Rodge, Siddhi P. Ransing and Vina M. Lomte, Virtual personal assistant for the blind, International Journal of Computer Science and Technology(IJCST) 9(4) October - December 2018.

[5]   La Mura, Monica, and Patrizia Lamberti, Human-machine interaction personalization: a review on gender and emotion recognition through speech analysis, In 2020 IEEE International Workshop on Metrology for Industry 4.0 and IoT, IEEE (2020), 319-323.

[6]   A. M. Sermakani, J. Monisha, G. Shrisha and G. Sumisha, Creating desktop speech recognition and using python programming, International Journal of Advanced Research in Computer and Communication Engineering 10(3) (2021).

[7]   Harshita Yadav, Ravi Teja Nandam and Sonu Lamba, Artificial intelligence-based virtual assistant, International Journal of Computer Science and Wireless Security (IJCSWS) 08(02) (2021), 39-43.

[8]   Winkler Rainer, Matthias Söllner and Jan Marco Leimeister, Enhancing problem-solving skills with smart personal assistant technology, Computers and Education 165 (2021), 104148.

[9]   Terzopoulos George and Maya Satratzemi, Voice assistants and artificial intelligence in education, In Proceedings of the 9th Balkan Conference on Informatics (2019), 1-6.

[10] Srinivasan, Arul and A. Neela Madheswari, The role of smart personal assistant for improving personal healthcare, International Journal of Advanced Engineering, Management and Science 4(11) (2018), 268-274.

[11] Ermolina Alena and Victor Tiberius, Voice-controlled intelligent personal assistants in health care: International Delphi study, Journal of Medical Internet Research 23(4) (2021), e25312.

[12] Kushwah Shivangee, Shantanu Singh, Kshitij Vats and Varsha Nemade, Gender identification Via voice analysis, International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT) 5(2) (2019), ISSN : 2456-3307.

[13] Harivans Pratap Singh, Aman Pratap Kushwaha, Aayushmaan and Harendra Singh, Voice-Based Email System, International Journal of Innovative Science and Research Technology ISSN No:-2456-2165, 6(7) (2021).

[14] Rijwan Khan, Pawan Kumar Sharma, Sumit Raj, Sushil Kr. Verma and Sparsh Katiyar, Voice based E-mail system using artificial intelligence, International Journal of Engineering and Advanced Technology (IJEAT) 9(3) (2020).

[15] Harsh D. Shah, Amit Sundas and Shabnam Sharma, Controlling Email system using audio with speech recognition and text to speech, In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), IEEE, (2021) pp. 1-7.