كلية الهندسة بشبرا
FACULTY OF ENGINEERING AT SHOUBRA
ERJ
PRINT ISSN 3009-6049
ONLINE ISSN 3009-6022

ENGINEERING RESEARCH JOURNAL (ERJ)
Volume (54),Issue (1)
January 2025, pp:95-102
https://erjsh.journals.ekb.eg

A Survey on Advancements in Voice Control Systems: Enhancing Human-Computer Interaction through Speech Recognition and AI

Amir Ahmed Mohamed El-Had El-Azazy [1,*] ,Raafat Abd-elfatah El-kammar [1] ,Ahmed Mohamed Fawzy [2] , Hala Mohamed Abd Elkader [1]

[1] *Electrical Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt.*
2 Nanotechnology Central Lab, Electronics Research Institute, Cairo, Egypt
**Corresponding author*

E-mail address: amir.azazy@moi.gov.eg, raafatelkammar@gmail.com , afawzy@eri.sci.eg, hala.mansour@feng.bu.edu.eg

Abstract: Voice control systems (VCS) have revolutionized how people interact with computers by providing voice commands and natural language. Modern technologies such as artificial intelligence, cloud computing, the Internet of Things, etc., are widely used in this era. When these technologies are combined with voice control systems, we find a wide range of applications that improve the quality of life for people in various categories, such, such as elderly people, children under four years old, people with physical disabilities... etc. This paper discusses the significance of voice control systems based on a collection of earlier research studies from 2013 to 2024 that describe the uses of control systems with cutting-edge technologies, the different kinds of voice control systems, the fundamental construction of voice control circuits, as well as the problems and difficulties associated with voice control systems..

Keywords: Voice Control System, Speech Recognition, Intelligent Voice Technology, Voice, Activation System Structure, and Artificial Intelligence (AI).

1. Introduction

Voice control systems (VCS) allow users to interact with computer interfaces using voice commands and natural language. Online services that utilize automatic speech recognition (ASR) enable users to send voice recordings, which are then converted into text. Improvements in voice recognition and natural language processing, fueled by advancements in deep learning, have greatly enhanced VCS technology. This progress highlights the tremendous potential and need for such technology, with predictions suggesting that the global VCS market will reach an impressive $95.41 billion by 2030 [1]. Traditional interfaces primarily rely on text and touch inputs. However, by employing speech recognition technology, users can save time and resources, as these systems provide quick services through an efficient interface without requiring extensive additional devices. Furthermore, VCS fosters voice communication between users and devices, enhancing user comfort and creating a more intuitive interface.

The adoption rate of voice-activated digital assistants has increased significantly, along with the emergence of new application areas for this technology [2]. As a result, a wide array of Internet of Things (IoT) devices now incorporates voice-activated user interfaces, offering simple ways to complete various tasks, such as creating grocery lists or making phone calls. Voice control is now also utilized for more complex activities as smart homes become more prevalent. Many of these tasks involve using voice-activated digital assistants found in smart speakers, smartphones, and other connected devices, allowing users to control multiple devices easily [3].

VCS technology has been applied in numerous areas, including smart gadgets, the Internet of Things (IoT), and automotive systems, all aimed at maximizing user experience and improving efficiency. By facilitating easier access to information and device control, VCS offers a straightforward and seamless solution for human-computer interaction.

AI plays a crucial role in how users engage with voice-based assistants (VAs), such as Amazon's Alexa and Apple's Siri. VAs have progressed significantly from their initial functions as simple news readers, music players, and timers. Recent advancements in AI technology have led to remarkable improvements in intelligent speech technology. This technology now enables quick access to information and efficient human-machine communication. It can convert text to speech through speech synthesis, translate speech recognition into text, comprehend the meaning of speech, and execute related commands and actions. Today, AI-driven speech technology is widely used across various industries, including banking, education, healthcare, and the automotive sector [4].

There are many problems facing elderly patients, people with physical disabilities, and children under four years old when they are at home alone. As well as challenges such as traffic congestion, epidemics like the Corona pandemic, and issues within healthcare systems. Voice-controlled systems offer a modern way of life where individuals can have complete control over their home or office appliances. Robotization has the potential to be especially important

today, with the real goal of improving our quality of life. It is necessary to develop easy and flexible systems that keep pace with the ongoing developments in the world to address the problems these groups face. There is a need to use voice control circuits with artificial intelligence to facilitate the management of home appliances, regulate traffic, prevent the spread of epidemics, and enhance healthcare systems to assist patients more efficiently.

Voice control systems are also used in automotive applications to prevent accidents on the roads, making it easier for drivers to focus on their surroundings. Although many people have concerns about using this technology due to issues related to security and data privacy, it remains easy and flexible for many others, as evidenced by its implementation in various applications. This will be further elaborated upon in the subsequent sections of the paper.

The following sections comprise the remaining text of the work: Section 2 provides a summary of related works on voice control systems with AI for various important applications from 2013 to 2024. Section 3 offers an overview of voice control systems. Section 4 includes a classification of intelligent voice technology. Section 5 presents the structure of the voice activation system. Section 6 discusses the challenges associated with voice control systems. Section 7 summarizes the conclusions of the work.

2. RELATED WORKS

Numerous studies utilized AI-powered voice control systems for important applications from 2013 to 2024, which we present as follows:

The research work in [5] involved a speech-activated robotics system for early childhood education. It enables children as young as four to communicate with robots using voice commands. The research leverages recent advances in artificial intelligence, including large language models, to enhance the intelligence and user-friendliness of robots. It acknowledges the challenges young learners face when trying to grasp programming languages and robotics concepts. This innovative approach significantly lowers barriers to entry and enhances the educational potential of robotics in the classroom by fostering a natural and intuitive interaction between children and robots. Within this framework, a software pipeline is proposed to translate voice instructions into actions for the robot. The most suitable deep learning models and cloud services are evaluated for each component, and the best-performing options are selected. Finally, an integration test involving children aged four to six years old validates the chosen arrangement.

The research work in [6] involved a novel distributed framework for smart home systems that features a voice user interface, specifically designed for languages with limited resources. This framework combines IoT, Fog, and Cloud architectures with advanced speech recognition techniques. It effectively employs speech recognition within the specific domain of smart homes by integrating existing Speech-To-Text (STT) and Text-To-Speech (TTS) cloud services.

The research work in [7] involved the integration of AI voice control with IoT within a secure infrastructure that spans multiple platforms and remote areas. The study highlighted an application model for a smart home automation system, addressing several key issues. These include the need for various underlying technologies to enable user-friendly voice-based control, concerns about security and privacy that lead to a lack of confidence and usability, and the general lack of awareness among users regarding how to effectively utilize machine intelligence to maximize the capabilities of IoT devices in a home environment.

The research work in [8] examined how perioperative services (Periop) personnel can utilize mobile devices, specifically through speech recognition technologies, to record workflow milestones. If voice recognition technology can be improved, it would facilitate the deployment of mobile technology to enhance patient flow and care quality. The goal of this initiative is to allow Periop staff to concentrate more on patient care instead of data entry and query processes. The findings from this study could also apply to other situations where an engineering manager seeks to improve communication effectiveness through mobile technology. Additionally, the study employed post-processing classifiers, including maximum entropy, support vector machines, and bag-of-sentences, to enhance the performance of Google's speech recognition. The trials were conducted at three different levels—zero training repetitions, five training repetitions, and ten training repetitions—and examined three factors: original wording, reduced phrasing, and personalized phrasing. The results indicated that maximum accuracy was achieved with personalized phrasing, and accuracy further improved when the device was trained to recognize a specific individual's voice.

The research work in [9] involved proposed a smart traffic system based on the Arduino UNO that adjusts traffic light signals based on spoken commands. The plan involves developing a Voice-Controlled Smart Traffic (VCST) system that enables the management of LED street lights according to traffic flow. Specific voice commands can be used to monitor the lighting system. Utilizing a low-power ZigBee network, the system is designed to be cost-effective, reliable, and user-friendly, contributing to energy optimization in smart grid architectures. Common voice commands like "red," "green," "yellow," and "stop" are employed to operate the system. These voice commands are transmitted to a Bluetooth module through an Android application, which accurately recognizes previously saved commands to control the traffic light system and street lights.

The research work in [10] discussed the integration of voice control, ZigBee technology, and the most widely used operating systems. It also suggests several feasible alternatives for voice control systems while considering cost factors. The intelligent terminal processes the speech signal through a series of steps, including preprocessing, feature extraction, pattern matching, and post-processing. This allows it to interpret the speech as a specific voice command, which is then sent to the main control center for further analysis. The voice command is transmitted via radio frequency communication to the appropriate control node. The control node identifies the command from the voice

input, executes the instruction to operate the household appliance, and simultaneously sends feedback back to the main control center.

The research work in [11] explained the overall architecture of a wireless home automation system that allows users to manage lights and other electrical appliances at home or work using voice commands and touchscreen interactions. The system has undergone verification and testing, which includes indoor ZigBee communication tests, touchscreen response tests, and speech recognition response tests. Home appliances can be controlled by tapping an icon on a touchscreen that operates using ZigBee technology, enabling remote operation of household devices. This system is suitable for use in various settings, including residences, hotels, retail establishments, factories, and process control systems. Special consideration was given to the needs of disabled users, as it utilizes both touchscreen technology and speech recognition for ease of access.

The research work in [12], the authors described the design of a low-cost voice recognition-based proposed home automation system designed for individuals with paraplegia or quadriplegia—those who are unable to move their limbs but can still speak and hear. This system allows them to raise or lower their bed using simple voice commands, tailored to their comfort and needs. The system consists of an Arduino Uno microcontroller, an adjustable bed, a relay circuit, and a voice recognition module. Before use, the voice recognition module must be trained to accurately recognize commands. The Arduino utilizes the relay circuit to control the bed's position based on the correctly detected voice commands.

The research work in [13] explored how artificial intelligence-based speech recognition technologies could help reduce educational disparities during COVID-19. It presents a comparative study of real-world examples of AI being utilized in education. Both teachers and students find that incorporating artificial intelligence into specialized software makes the educational process more convenient.

The research work in [14] explored the implementation of a developed general design in practice. The Wireless Home Automation System (WHAS) has become easier to use and more cost-effective to install, thanks in part to the Voice Recognition Application. The second section of this study, titled "Foot Step Counter," addresses automated light switching. This feature ensures that when someone approaches the door and there is no time to launch the application or establish a Bluetooth connection, the room lights will automatically turn on. The speech recognition application's effective range is reported to be 1.5 meters, and with four access group characteristics, an accuracy of 85.25% in speech recognition was achieved.

Table 1 summarizes recent studies conducted from 2021 to 2024 on voice control systems, highlighting the most prevalent technologies such as Cloud Computing (C.C), Edge Computing (E.C), Fog Computing (F.C), Artificial Intelligence (AI), and the Internet of Things (IoT).

TABLE 1. Summary of related works of voice control system with the most common technologies

Ref	Techniques	Technologies				
		AI	IoT	C.C	E.C	F.C
[15]	Examples of Safe and Consolidated Voice-Control System Use in Smart Home Automation	✓	✓	✓	x	x
[16]	Artificial intelligence and Internet of Things-based voice-activated smart home automation system	✓	✓	✓	x	x
[17]	Voice Recognition Control System Using IoT Sensors and Cloud Computing	✓	✓	✓	x	x
[18]	A system for Galician and mobile opportunistic scenarios that focuses on voice recognition for IoT home automation design, implemented, and practically evaluated. It is suitable for low-resource languages and resource-constrained edge IoT devices.	✓	✓	✓	✓	x
[19]	Enhancing IoT-Based Fog-Based Speech Recognition Remote Patient Monitoring Systems	✓	✓	X	x	✓
[20]	Voice assistants in the hotel industry: using AI to customer support	✓	✓	X	x	x
[21]	A Voice-Assisted Method for Enquiring About Vehicle Data from Automotive Internet of Things Databases	x	✓	✓	x	x
[22]	Voice-based AI in contact center customer support: An organic field test	✓	X	X	x	x
[23]	Voice-Activated Aid for the Elderly: Combining IoT with Speech Recognition	✓	✓	X	x	x
[24]	Voice Guidance System Using Internet of Things for Colour Recognition	✓	✓	✓	x	x

3. OVERVIEW OF VOICE CONTROL SYSTEMS

Voice control systems [25] rely on speech recognition technology and natural language processing, enabling operations to be performed in response to spoken commands from users. Today, voice interaction is utilized in a variety of systems, sometimes serving as the primary means of interaction and other times complementing other methods. Intelligent home assistants, often referred to as conversational agents or voice user interfaces (VUIs), are examples of interfaces that primarily operate through voice commands [26]. These systems typically employ synthetic speech interfaces and do not include a graphical user interface. However, voice interaction is more commonly integrated into mobile devices as an enhancement to touch controls. The hands-free capability of voice control allows for improved user experience and greater convenience.

4. CLASSIFICATION OF INTELLIGENT VOICE TECHNOLOGY

Intelligent voice technology can be categorized into three types: Speech transcription technology, Speech recognition technology, and Speech synthesis technology, as illustrated in Figure 1.

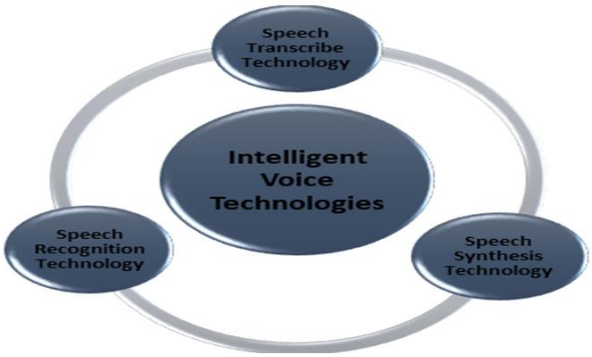

FIGURE 1. Intelligent voice technologies.

4.1 SPEECH SYNTHESIS TECHNOLOGY

The term "speech synthesis" refers to the process by which a computer generates spoken language from text. This technology can be applied to various tasks, including playback, voice prompting, and language navigation. The primary functions involved in language synthesis include determining the order of words, processing word data, converting language into sound waves, arranging units into a sequence of waveforms, and producing the final audio output [27].

4.2 SPEECH TRANSCRIBES TECHNOLOGY

Voice transliteration technology is a process that recognizes speech and voice signals and converts them into written text. It can also achieve real-time speech interpretation by utilizing extensive databases and user history. The significant semantic context, including pauses, tone, and other elements, is summarized, and phrases and paragraphs are identified and closely examined to address issues such as background noise [28]. There are various types of speech transcription technologies, as illustrated in Figure 2.

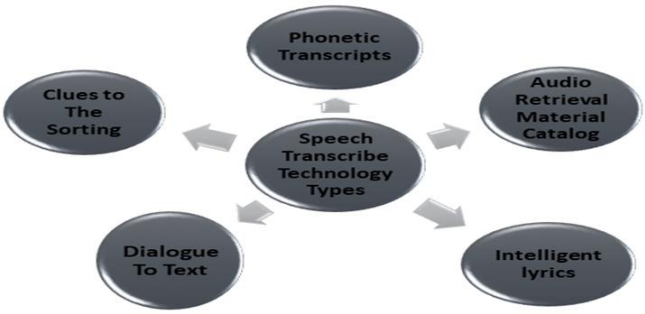

FIGURE 2. Types of speech transcribe technology.

A. Phonetic Transcripts

Phonetic transcriptions are symbols that represent the sounds of human speech. This sort of technology may be utilized to properly transcribe spoken words, especially when the pronunciation is uncertain or inconsistent.

B. Clues to The Sorting:

This form of speech transcription technology sorts and organizes spoken words using indications such as keywords, context, or tone of voice. It aids in categorizing and organizing enormous amounts of spoken stuff.

C. Audio Retrieval Material Catalog

This technique is used to construct catalogs or databases of spoken information, which can then be readily searched and retrieved. It facilitates the management and organization of audio data for easy reference and access.

D. Dialogue To Text:

This sort of technology records chats or discussions as printed text. It is widely used for transcribing interviews, meetings, and other vocal engagements.

E. Intelligent lyrics

This form of voice transcribe technology is designed to properly transcribe song lyrics, as well as recognize and transcribe background vocals and ad-libs. It facilitates the creation of accurate and complete lyrics for music streaming services and websites.

4.3 SPEECH RECOGNITION TECHNOLOGY

Speech recognition is an innovative technology that focuses on converting spoken words into text for input purposes. This technology identifies speech patterns and determines when the audio ends before starting the recognition process. It takes into account the conversational context of the sentence while intelligently analyzing punctuation and other relevant elements of the incoming data. During the input process, the system preferentially selects suitable phrases by recognizing user-defined terms [29]. A diagram illustrating speech recognition patterns is presented in Figure 3.

5. VOICE ACTIVATION SYSTEM STRUCTURE

The majority of voice activation systems consist of the following components (voice Signal acquisition, Feature Extraction, Acoustic Modelling, Language, and Recognition) as presented in Figure 4.

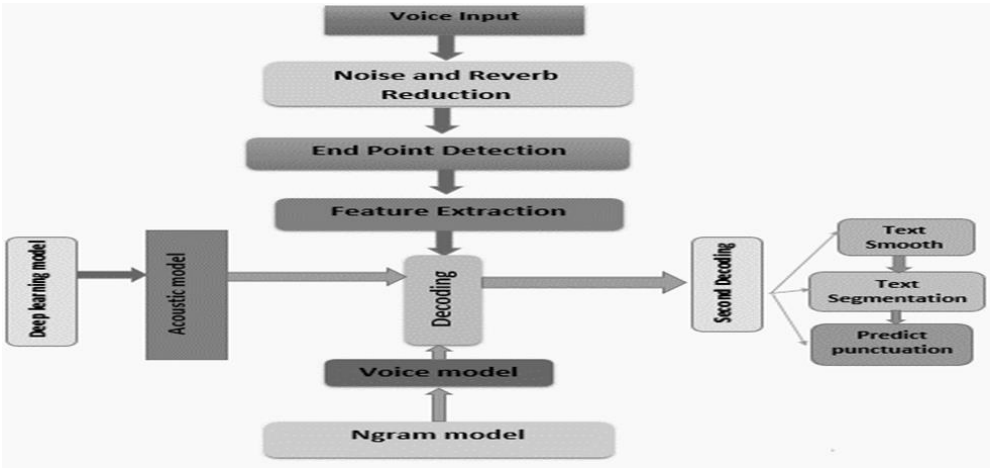

FIGURE 3: Diagram of speech recognition patterns.

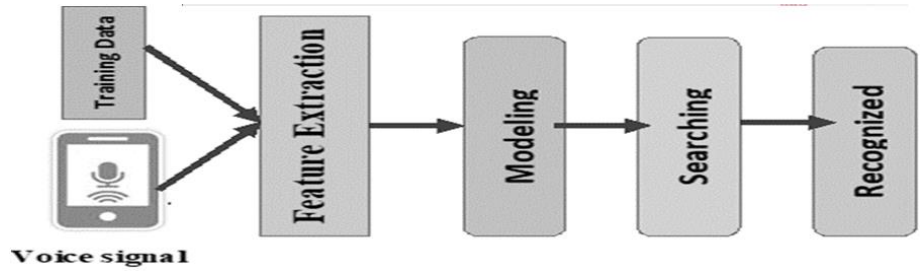

FIGURE 4. Voice Activation System Structure.

5.1 FEATURE EXTRACTION

The success of recognition systems heavily relies on the feature extraction stage, which deserves significant attention. The human voice is a time-varying digital signal that can be analyzed mathematically using digital signal processing tools. It's utilized in various authentication applications, including speaker recognition, language identification, speech synthesis, voice analysis, and speech coding. Several techniques have been developed to extract similar properties from human speech, as illustrated in Figure 5.

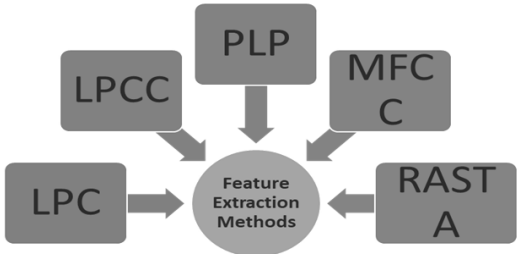

FIGURE 5. Voice feature extraction methods.

5.1.1 LINEAR PREDICTIVE COEFFICIENTS (LPC)

One popular low-bit rate coder used to determine the power spectrum of digital voice signals is the linear predictive coding (LPC) method. Formant analysis relies heavily on the LPC approach [30]. Instead of converting the entire digital speech signal, this method utilizes the differences between samples to predict future voice samples. These predicted samples are then statistically processed to extract the features of the digital speech signal. By combining the "$\boldsymbol{m}$ " number of predicted samples, it's possible to reconstruct the original digital voice signal.

To illustrate the prediction model, consider the current voice sample represented as $\boldsymbol{x}$**(m)** and the previous sample as $\boldsymbol{x}(\boldsymbol{m}-\mathbf{1})$. The future sample $\boldsymbol{x}^{-}(\boldsymbol{m})$ can be predicted using a specific equation as presented in equation (1).

$$\boldsymbol{x}^{-}(\boldsymbol{m}) = \sum_{i=1}^{m=\infty} \boldsymbol{k_i x}(\boldsymbol{m}-\boldsymbol{i}) \pm \boldsymbol{e}(\boldsymbol{m}) \quad \textbf{(1)}$$

Where the $\boldsymbol{k_i}$ is a factor of prediction and $\boldsymbol{e}(\boldsymbol{m})$is an error of prediction can calculated as presented in equation (2)

$$\boldsymbol{e(m)= x\ (m)\ -\ x^{-}(m)} \quad \textbf{(2)}$$

5.1.2 LINEAR PREDICTIVE CEPSTRAL COEFFICIENTS (LPCC)

An improved variant of LPC, the Linear Predictive Cepstral Coefficients (LPCC) approach [31] uses Cepstral Mean Subtraction (CMS) to address channel effects. Pitch tracking uses the Cepstral, a series of power spectral densities taken from the periodogram. The cepstrum can be classified as real, complex, phase, or power, depending on whether it is used for speech analysis or not. It is created by executing an Inverse Fourier Transform on the power spectrum of voice sounds.

Let $\boldsymbol{x\ (m)}$ represent the time-domain audio signal that can be analytically analyzed using the Discrete Fourier Transform (DFT). Squaring this result produces the power spectrum $\boldsymbol{p}$ as presented in equation (3).

$$\boldsymbol{p} = \left|\boldsymbol{DFT^{-1}}(\boldsymbol{m})\right)|^{\mathbf{2}} \quad \textbf{(3)}$$

Instead of producing $\boldsymbol{x(m)}$, the inverse $\boldsymbol{DFT}$ of $\boldsymbol{p}$ produces the autocorrelation $\boldsymbol{U}(\boldsymbol{m})$for the time domain speech signal $\boldsymbol{x(m)}$ as presented in equation (4)

$$U(m) = DFT^{-1}(p) \quad (4)$$

The cepstrum $C(m)$ is obtained by processing the Power spectrum with inverse DFT after logarithmic compression as presented in equation (5).

$$C(m) = DFT^{-1}(log(p) \quad (5)$$

5.1.3 PERCEPTUAL LINEAR PREDICTIVE COEFFICIENTS (PLP)

An adapted model of LPCC for mitigating noise and training/test voice sample mismatches is called Perceptual Linear Predictive (PLP) Coefficients.

$$p(W) = Re[S(W)]^2 + lm[S(W)]^2 \quad (6)$$

$p(W)$ Is the power spectrum distorted for the barking frequency Ω and can calculated $p(W)$ and $\Omega(W)$ as presented in equation (6&7).W Is the angular frequency

$$\Omega(W) = 6ln\{[(\frac{W}{1200\pi})^2 + 1]^{-1}\} \quad (7)$$

5.1.4 MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

The MFCC utilizes a linear cosine transform of the spectrum. Its uniformly spaced frequency bands distinguish it from the Cepstrum coefficient approach.

5.1.5 RELATIVE SPECTRA FILTERING OF LOG DOMAIN COEFFICIENTS (RASTA)

The suppression of zero frequency and slowly variable components in the speech signal when feature extraction is carried out using the MFCC technique is overcome by the Relative spectra filtering of log domain coefficients (RASTA) approach. By reducing the influence of noise in speech signals, the RASTA filter outperforms the MFCC and offers a high rate of resilience.

5.2 ACOUSTIC MODEL

The responsibility of the acoustic model is to simulate the acoustic characteristics of the chosen acoustic unit. A probability distribution across the vectors of MFCC features, for instance, may be obtained from an auditory model when a certain phrase is spoken. In practical terms, P (S|u) [TeX:] $ P (S|u) $ is calculated using the acoustic model, where S represents sound and u is an acoustic unit. Finding P(S|u) [TeX:] after computing P(u|S [TeX:] $ P(u|S)$ is frequently more intuitive or simpler. Using the Bayes theorem, $ P(S|u)$. It is very common to use this method with Hidden Markov Models (HMM). This is the construction of the most widely used acoustic model for voice activation. The collection of HMM states may be rationally separated into two parts: a trash model (a model of the remaining sound, such as background speech, noise, and real voice request) and a component that reflects the audio event of keyword pronunciation. An example of HMM utilized in Amazon's spotter for the term "Alexa".

5.3 DECODING

The procedure of ascertaining the state sequence is based on acoustic observations and an acoustic model to establish the utterance of a keyword. To determine whether a keyword was uttered, Chen et al [32] describe voice activation systems that use a deep neural network-specified acoustic model to extract Log Mel-filter bank (feature extraction). The outputs of the deep neural network are then smoothed and compared with a threshold (decoding). When the acoustic model is applied to an audio stream, we obtain the values that describe the likelihood that this or that acoustic unit was spoken at a specific time. Based on the acquired one or more numeric series, the voice activation system must determine if the keyword was stated in an audio stream. This is accomplished by using various decoding techniques. To make a choice, the simplest scenario requires merely comparing the resulting number with the threshold value. For instance, if the entire keyword is the acoustic unit, the calculated probability is compared to 0.5 to make the decision.

5.4 VOICE RECOGNITION

In recent decades, the appeal of voice recognition technology has resurfaced due to technological advancements. Innovations in various fields of computer science and technology are now being applied in cutting-edge research related to machine learning and deep learning. This research involves developing the capability to classify sounds and predict the categories to which they belong. Numerous innovative applications of deep learning-based voice recognition are transforming everyday life. Voice recognition serves several purposes, including providing voice assistance, identifying individuals based solely on their speech, classifying music clips to determine their genres, and much more.

6. CHALLENGES OF VOICE CONTROL SYSTEM

Researchers must promptly address the following questions regarding concerns:

a. The fact that VCS is so diverse and includes both speaker verification (SV) and automated speech recognition (ASR) means that these two aspects are frequently researched separately [33, 34]. Of course, a lot of VCS systems combine ASR and SV features, and attacks on these systems frequently include features that overlap. Hence, when creating VCS, designers need to take into account the possible weaknesses of both ASR and SV simultaneously.
b. VCS designers must be aware of efficient defense strategies in addition to attack techniques. Adversarial training is one example of a current defense strategy that is frequently customized to target certain attack types [35, 36, 37, 38, and 39]. This makes it difficult to choose the right defense combinations to effectively strengthen VCS against a variety of threats.
c. Computer systems are vulnerable to various attacks, primarily due to two factors: the advancement of technology, which provides attackers with increasingly sophisticated tools and algorithms, and the inherent complexity of Visual Control Systems (VCS). VCS consists of multiple hardware and software components, each with its security weaknesses. Designers need to understand both the mechanisms behind these attacks and the specific vulnerabilities of each component in the VCS.

The main challenges of voice control systems.

- ***Noise Disturbance***

The presence of ambient noise can affect both attackers and defenders in the security environment of Voice Command Systems (VCS). For attackers, ambient noise may reduce the reach and effectiveness of harmful audio. Conversely, as shown in reference [40], noise can interfere with defense mechanisms, such as liveness detection systems, making them less accurate for defenders. As a result, the security landscape of VCS becomes more complex, as both sides must take into account the impact of noise when developing their strategies.

- ***Hardware Enhancement***

Security threats often exploit hardware vulnerabilities, particularly those related to microphones. However, these weaknesses are not present in all microphone types. For example, the iPhone 6 Plus has been shown to effectively protect against voice synthesis attacks due to its unique microphone design [41]. This variation in hardware vulnerabilities significantly complicates the execution of consistent attacks at the physical layer.

- ***Mode Knowledge***

Models are becoming increasingly common as VCS technology continues to advance. To protect their intellectual property and prevent competitors from copying their work, businesses often use proprietary models that are not open-sourced. This secrecy makes it much more difficult for hostile attacks to succeed, as attackers must operate within a black-box environment. However, a significant challenge remains in the field of adversarial attacks: developing universal adversarial perturbations that can produce similar attack outcomes across different models.

7. CONCLUSION

This paper presents a comprehensive study of the significant applications of voice control systems. Where Section 2 provided a summary of related works on voice control systems with AI for various important applications from 2013 to 2024. Section 3 offered an overview of voice control systems. Section 4 included a classification of intelligent voice technology. Section 5 presents the structure of the voice activation system. Section 6 discusses the challenges associated with voice control systems. It highlights the importance of these systems and how, when integrated with the latest technologies, they can address various challenges. For instance, voice control systems facilitate the use of home appliances for elderly individuals, children under the age of four, paralyzed patients, and people with physical disabilities. Additionally, they have been effective in alleviating traffic congestion and were utilized during the COVID-19 pandemic to help prevent the spread of infection. The research also covers the primary types of voice control systems and their essential components. Furthermore, we discuss the issues and challenges associated with these systems. A summary of numerous previous studies is provided, illustrating the significance of voice control systems and their integration with artificial intelligence and other technologies.

REFERENCES

[1] Wang, Y., Yan, Q., Ivanov, N., & Chen, X. (2023). A Practical Survey on Emerging Threats from AI-driven Voice Attacks: How Vulnerable are Commercial Voice Control Systems? arXiv preprint arXiv:2312.06010.

[2] Han, S., & Yang, H. (2018). Understanding the adoption of intelligent personal assistants: A parasocial relationship perspective. Industrial Management & Data Systems, 118(3), 618-636.

[3] Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018, April). Voice interfaces in everyday life. In proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1-12).

[4] Kumar, M., Sharma, S., Chaudhary, D., & Prakash, S. (2021, March). Image recognition using artificial intelligence. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 760-763). IEEE.

[5] Aguilera, C. A., Castro, A., Aguilera, C., & Raducanu, B. (2024). Voice-Controlled Robotics in Early Education: Implementing and Validating Child-Directed Interactions Using a Collaborative Robot and Artificial Intelligence. Applied Sciences, 14(6), 2408.

[6] Iliev, Y., & Ilieva, G. (2022). A framework for a smart home system with voice control using NLP methods. Electronics, 12(1), 116.

[7] Venkatraman, S., Overmars, A., & Thong, M. (2021). Smart home automation—use cases of a secure and integrated voice-control system. Systems, 9(4), 77.

[8] Uddin, M. M., Huynh, N., Vidal, J. M., Taaffe, K. M., Fredendall, L. D., & Greenstein, J. S. (2015). Evaluation of Google's voice recognition and sentence classification for health care applications. Engineering Management Journal, 27(3), 152-162.

[9] Biswal, A. K., Singh, D., & Pattanayak, B. K. (2021). IoT-based voice-controlled energy-efficient intelligent traffic and street light monitoring system. In Green Technology for Smart City and Society: Proceedings of GTSCS 2020 (pp. 43-54). Springer Singapore.

[10] Yang, C. (2021). Design of smart home control system based on wireless voice sensor. Journal of Sensors, 2021(1), 8254478.

[11] Kirankumar, T., & Bhavani, B. (2013). A Sustainable automated system for elderly people using voice recognition and touch screen technology. International Journal of Science and Research (IJSR), 2(8), 265-267.

[12] Kumar, M., & Shimi, S. L. (2015). Voice recognition-based home automation system for paralyzed people. International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), 4(10).

[13] Tregubov, V. (2021). Using Voice Recognition in E-Learning System to Reduce Educational Inequality During COVID-19. International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol, 11.

[14] Sravanthi, G., Madhuri, G., Sharma, N., Tiwari, A., Kashyap, A., & Suresh, B. (2018, October). Voice recognition application-based home automation system with people counter. In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 574-578). IEEE.

[15] VenkUltraman, S., Overmars, A., & Thong, M. (2021). Smart home automation—use cases of a secure and integrated voice-control system. Systems, 9(4), 77.

[16] Torad, M. A., Bouallegue, B., & Ahmed, A. M. (2022). A voice-controlled smart home automation system using artificial intelligence and the Internet of Things. TELKOMNIKA (Telecommunication Computing Electronics and Control), 20(4), 808-816.

[17] Song, X., & Sun, S. (2022). [Retracted] Voice Recognition Control System Based on Cloud Computing and IoT Sensors. Wireless Communications and Mobile Computing, 2022(1), 4489452.

[18] Froiz-Míguez, I., Fraga-Lamas, P., & Fernández-CaraméS, T. M. (2023). Design, Implementation, and Practical Evaluation of a Voice Recognition Based IoT Home Automation System for Low-Resource Languages and Resource-Constrained Edge IoT Devices: A System for Galician and Mobile Opportunistic Scenarios. IEEE Access, 11, 63623-63649.

[19] Baucas, M. J., & Spachos, P. (2023). Improving Remote Patient Monitoring Systems Using a Fog-Based IoT Platform With Speech Recognition. IEEE Sensors Journal, 23(15), 17611-17618.

[20] Buhalis, D., & Moldavska, I. (2022). Voice assistants in hospitality: using artificial intelligence for customer service. Journal of Hospitality and Tourism Technology, 13(3), 386-403.

[21] Andrade, M., Wanderley, E., Azevedo, M., Medeiros, T., Silva, M., Silva, I., & Costa, D. G. (2023, October). A Voice-Assisted Approach for Vehicular Data Querying from Automotive IoT-Based Databases. In 2023 Symposium on Internet of Things (IoT) (pp. 1-5). IEEE.

[22] Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2023). Voice-based AI in call center customer service: A natural field experiment. Production and Operations Management, 32(4), 1002-1018.

[23] Chumuang, N., Ganokratanaa, T., Pramkeaw, P., Ketcham, M., Chomchaiya, S., & Yimyam, W. (2024, January). Voice-activated assistance for the elderly: Integrating speech recognition and iot. In 2024 IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-4). IEEE.

[24] Sung, W. T., Chen, G. R., & Hsiao, S. J. (2023). Voice Guidance System for Color Recognition Based on IoT. Computer Systems Science & Engineering, 45(1).

[25] Jansson, M., & Nelderup, E. (2023). Introducing Voice Control in a Graphical User Interface Using a Keyword-Based Approach.

[26] Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018, April). Voice interfaces in everyday life. In proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1-12).

[27] Ran, D., Yingli, W., & Haoxin, Q. (2021). Artificial intelligence speech recognition model for correcting spoken English teaching. Journal of Intelligent & Fuzzy Systems, 40(2), 3513-3524.

[28] Lu, J., Xiong, S., & Wang, M. (2019). Application of Image Recognition Technology Based on Artificial Intelligence in Traffic Meteorological Service [J]. China Computer & Communication.

[29] Zhong, X., & Shih, F. Y. (2021). Automatic Image Pixel Clustering based on Mussels Wandering Optimization. International Journal of Pattern Recognition and Artificial Intelligence, 35(02), 2154005.

[30] Xing, X., Lin, J., Wan, C., & Song, Y. (2017). Model predictive control of LPC-looped active distribution network with high penetration of distributed generation. IEEE Transactions on Sustainable Energy, 8(3), 1051-1063.

[31] Gupta, H., & Gupta, D. (2016, January). LPC and LPCC method of feature extraction in Speech Recognition System. In 2016 6th international conference-cloud system and big data engineering (confluence) (pp. 498-502). IEEE.

[32] Chen, N. F., Sivadas, S., Lim, B. P., Ngo, H. G., Xu, H., Ma, B., & Li, H. (2014, May). Strategies for Vietnamese keyword search. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4121-4125). IEEE.

[33] Abdullah, H., Warren, K., Bindschaedler, V., Papernot, N., & Traynor, P. (2021, May). Sok: The faults in our answers: An overview of attacks against automatic speech recognition and speaker identification systems. In 2021 IEEE symposium on security and privacy (SP) (pp. 730-747). IEEE.

[34] Chen, Y., Zhang, J., Yuan, X., Zhang, S., Chen, K., Wang, X., & Guo, S. (2022). Sok: A modularized approach to study the security of automatic speech recognition systems. ACM Transactions on Privacy and Security, 25(3), 1-31.

[35] Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and language processing, 22(10), 1533-1545.

[36] Gong, Y., & Poellabauer, C. (2018). An overview of vulnerabilities of voice-controlled systems. arXiv preprint arXiv:1803.09156.

[37] Hai, J., & Joo, E. M. (2003, December). Improved linear predictive coding method for speech recognition. In the fourth International Conference on Information, communications and Signal Processing, 2003, and the fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 joint (Vol. 3, pp. 1614-1618). IEEE.

[38] Iter, D., Huang, J., & Jermann, M. (2017). Generating adversarial examples for speech recognition. Stanford Technical Report.

[39] Wang, J. (2021, August). Adversarial Examples in Physical World. In IJCAI (pp. 4925-4926).

[40] Meng, Y., Li, J., Pillari, M., Deopujari, A., Brennan, L., Shamsie, H., ... & Tian, Y. (2022). Your microphone array retains your identity: A robust voice liveness detection system for smart speakers. In 31st USENIX Security Symposium (USENIX Security 22) (pp. 1077-1094).

[41] He, Y., Bian, J., Tong, X., Qian, Z., Zhu, W., Tian, X., & Wang, X. (2019, October). Canceling inaudible voice commands against voice control systems. In The 25th Annual International Conference on Mobile Computing and Networking (pp. 1-15).