DMPedia Lecture Notes in Multidisciplinary Research,
Volume 1, pp 854-867 2026

International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)
G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

Applications of Speech Recognition: A Survey

Shubham Pal, Mayank Vishwakarma, Anurag Yadav, Ankit Sahani, Mohammad Jeelani, Abhishek Saxena

Department of Computer Application, Future University, India

pal8887555053@gmail.com, glitzymayank7777@gmail.com, anuragyad979@gmail.com, amitkumarsahani1220@gmail.com, jeelani.0018@gmail.com, bhisheksaxena@futureuniversity.in

Abstarct

One important area of natural language processing (NLP) and artificial intelligence (AI) is speech recognition, which converts spoken language into text. Its evolution spans from early rule-based systems like Bell Labs' "Audrey" to modern deep learning-based architectures, significantly improving accuracy and usability. Applications extend across healthcare, education, telecommunications, defence, robotics, and consumer electronics, enhancing accessibility and human–machine interaction. Despite advancements, challenges such as accent variability, noisy environments, and ethical issues persist. Current research emphasizes deep neural networks, clustering, and statistical models, with continuous progress driving speech recognition toward greater efficiency, inclusivity, and real-world adoption.

Keywords: Speech Recognition, ASR, NLP, AI

1. Introduction

Automatic speech recognition (ASR), also known as voice recognition, is a rapidly developing area of natural language processing (NLP) and artificial intelligence (AI). It involves computers and other digital equipment that translate spoken language into text. What was once considered a futuristic concept is now an integral part of many technologies, we use every day. Advances in big data, cloud computing, and machine learning algorithms have allowed speech recognition to go beyond simple voice commands to sophisticated, context-aware systems that can comprehend and react to human language with astounding precision. Speech recognition is used across many industries, changing how people communicate with technology. Virtual assistants like Apple's Siri, Amazon's Alexa, Google Assistant, and Microsoft's Cortana are among the most well-known applications. With just vocal commands, users can use these AI-powered systems to send messages, set reminders, play music, and operate smart home appliances. The natural interaction they offer is driving their adoption in both personal and professional environments.

Speech recognition is proving revolutionary in the healthcare industry. Medical records and patient notes can be dictated by doctors straight into Electronic Health Record (EHR) systems, saving time and reducing administrative burden. This not only increases efficiency but also helps improve the accuracy of medical documentation. Some systems can now identify medical terminology, provide real-time transcription, and even assist with clinical decision-making. Customer service and call centres are another area where speech recognition is having a significant impact. Interactive Voice Response (IVR) systems powered by ASR can handle

routine customer queries without human intervention, improving response time and reducing operational costs. Moreover, combining ASR with sentiment analysis enables companies to gain insights into customer satisfaction and engagement levels. Voice-enabled solutions in the automotive sector improve driver safety by enabling hands-free operation of communication, entertainment, and navigation systems. This reduces outside distractions and makes driving safer. Automotive speech recognition systems are also becoming more context-aware, adapting to accents, languages, and background noise for improved performance. The education sector is also benefiting from speech recognition technologies. They are used to assist students with disabilities, such as those with visual impairments or dyslexia, by enabling them to interact with educational content through voice. Additionally, speech-to-text tools support language learning and lecture transcription, improving accessibility and learning outcomes. Another key application is in real-time translation and transcription services, which are increasingly used in global communications. Platforms like Zoom, Microsoft Teams, and Google Meet now offer live captioning powered by speech recognition, bridging language barriers in multilingual conversations.

As speech recognition continues to advance, its integration into everyday technologies is expected to deepen. It not only enhances convenience but also promotes inclusivity, allowing people of different abilities and backgrounds to engage with technology more naturally. Speech recognition is one of the most revolutionary technologies of the contemporary period, as the increasing use of voice as an interface heralds a move toward a more natural human-computer connection. Among the different areas of speech were identified, which include: speaker identification, speech emotion recognition, speech enhancement, speech recognition, speech transcription, among others. [6] More than a year ago, four research groups (a group at Google, plus the three groups represented by the current organisers) wrote an overview in which they presented their shared views on applying DNNs to acoustic modelling in speech recognition. [7] Modern speech recognition systems, including those described use a statistical approach based on Bayes' rule [4]. Clustering methods have been widely used in statistical data analysis to model a complex data set. Globally, the data set might be homogeneous and difficult to understand. However, if we cluster the data into homogeneous regions, each cluster is much simpler, allowing various models to be constructed. Many clustering algorithms have been developed in the literature, ranging from hierarchical methods, such as bottom-up (or agglomerative) methods and top-down (or divisive) methods, to optimisation methods such as the k-means algorithm [12].

2. RELATED WORK

Rabiner [1] has reviewed the present state of speech recognition technology, shown how it has been used in today's services and applications, and outlined how it has evolved over time to produce the next generation of voice-enabled services. Additionally, far-improved performance across practically every voice recognition technology area is anticipated in the future, along with increased resilience to background noise and multiple speakers. Rabiner [2] has attempted to thoroughly and methodically examine the theoretical underpinnings of this kind of statistical modelling and demonstrate how they have been used to address specific voice recognition machine recognition issues. Russell, Brown, Skilling, Series, Wallace, Bonham and Barker [3] recognised the advantages of the technology, which needs to be broadly

accessible and reasonably priced for elementary schools. Only by utilizing the minimal processing power already present in the elementary school classroom would that be possible. Ganapathiraju, Hamaker, and Picone [4] showed that, based on the Deterding vowel data, SVMs significantly enhance performance in a static pattern classification task. Additionally, they discussed the use of SVMs for large-vocabulary voice recognition and demonstrated improvements in error rates on two tasks: the Switchboard large-vocabulary conversational speech task and the OGI Alphadigits continuous alphadigit task.

Vajpai and Bora [5] aimed to provide a comprehensive overview of the voice recognition technologies discussed in the literature, drawing on knowledge from individual studies and advancements. The real-world and industrial uses of speech recognition have also been described, with particular attention to the fields of medicine, industrial robots, forensics, defense, and aviation. Nassif, Shahin, Attili, Azzeh, and Shaalan [6] provided a comprehensive analysis of the research projects carried out for speech applications since 2006, when deep learning emerged as a novel field of machine learning. This review included a comprehensive statistical analysis, carried out by extracting specific data from 174 papers published between 2006 and 2018. The findings presented in this paper highlight new research areas and provide insight into the patterns of previous studies in this field. Deng, Hinton, and Kingsbury [7] demonstrated that DNN-based acoustic models are still evolving rapidly and that allied fields such as music can benefit from similar techniques. Gauvain and Lamel [8] reviewed large-vocabulary continuous voice recognition, the state-of-the-art in fundamental technology, with an eye toward highlighting recent developments. They then examine system efficiency, portability across languages and jobs, and improving the system output by adding tags and non-linguistic information. They also identify challenges with advancing toward applications. Arora and Singh [9] described an analysis of the literature on automatic speech recognition. That covered previous years' developments to present the advancements in this field of study. Gaikwad, Gawali, and Yannawar [10] provide a summary of key technological viewpoints, an appreciation of the basic advances in speech recognition, and an outline of the techniques developed at each stage of the process. This study discusses the relative benefits and drawbacks of each technique to assist in selecting a technique. Young and Mihailidis [11] presented an analysis of the clinical research literature looking at how people with dysarthria use commercially available speech-to-text automatic speech recognition technology. Shoabing and Gopalakrishnan [12] proposed to optimize the Bayesian information criterion (BIC), a model selection criterion in the literature on statistics, in order to determine the number of clusters. We create a termination criterion for hierarchical clustering techniques that greedily maximizes the BIC criterion. Mishaim Malik and Muhammad Kamran Malik [13] identified all factors that may affect an ASR's performance. As a result, they hypothesize that scholars interested in ASR research would find this work to be a suitable place to start.

3. Literature Review

Early speech analysis and recognition experiments in the 1950s laid the groundwork for voice recognition. Bell Labs' 1952 "Audrey" system, which could identify spoken numbers from a single speaker, is a noteworthy example. Speech recognition systems used rule-based

techniques in the 1970s and 1980s, recognizing speech by applying grammatical and phonetic criteria. The complexity and diversity of speech patterns were beyond the capabilities of these systems. Speech recognition was transformed with the advent of statistical modeling approaches like Hidden Markov Models (HMMs). Systems were able to increase recognition accuracy and understand patterns from audio data thanks to HMMs. Commercial uses for speech recognition technology started to emerge, such as:

- **Telecom Systems:** Automated customer service systems and interactive voice response (IVR) systems.
- **PC and Mobile Devices:** Speech recognition software enabled users to interact with devices using voice commands.

Speech recognition has changed even further with the introduction of deep learning methods like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These models' ability to recognize intricate patterns in speech data results in notable gains in robustness and accuracy.

- **Improved Accuracy**: Notable gains in robustness and accuracy of recognition.
- **Virtual Assistants:** Speech recognition became widely used thanks to virtual assistants like Alexa, Google Assistant, and Siri.
- **Smart Home Devices:** Speech recognition integrated into smart home devices, enabling voice control.

In the papers we selected for our special session, we discuss recent findings on the use of DNNs for voice recognition [7].

4. Research Methodology

Systems that translate human voice into machine-readable text are designed, developed, and evaluated using a methodical approach in speech recognition research. It combines principles from linguistics, signal processing, and artificial intelligence. First, you've got to gather data. This involves collecting huge amounts of spoken language from a variety of sources. You need to make sure you have a wide range of speakers, different accents, and various background noises. It's also important to include multiple languages to make sure the model is robust. The quality and variety of these speech datasets, such as those in LibriSpeech or TIMIT, are absolutely critical for training models that are both accurate and reliable. After gathering the data, the next step is preprocessing. This is where the raw audio is cleaned and prepared for the model.

First, a number of techniques are applied, such as noise reduction, to eliminate background sounds, and normalization, to standardize the audio's volume. After that, the audio is segmented into more digestible, smaller pieces. Feature extraction from the processed audio is the final stage. These features are appropriate for machine learning algorithms because they provide a numerical depiction of the audio's salient aspects. The most popular techniques for this are generating spectrograms, which show the frequency and timing of the audio, or Mel-Frequency Cepstral Coefficients (MFCCs). The modeling stage is crucial to the study of voice recognition. In the past, systems used Gaussian Mixture Models (GMMs) and Hidden Markov Models

(HMMs) to recognize and order speech patterns. However, deep learning models like Transformers, Recurrent Neural Networks, and Convolutional Neural Networks (CNNs) are used in today's state-of-the-art systems. These modern architectures excel at uncovering intricate connections between sound features and the rules of language, enabling the transcription of speech to be fluid and natural.

Following the core modeling stage, the system then integrates a language model. This component is crucial for predicting the most likely sequence of words based on linguistic rules and context, helping resolve ambiguities in the audio. Older systems used statistical models such as N-grams, which estimate the probability of a word appearing after a given number of previous words. However, modern systems now use more sophisticated neural language models that can understand complex grammar and long-range context, significantly improving the accuracy of the final transcription. The process culminates with testing and evaluation. Metrics like accuracy and Word Error Rate (WER) are used to gauge the system's performance. For instance, great precision is indicated by a low WER. To ensure the system's dependability in practical conditions, it's rigorously tested under various conditions, including different accents, varying levels of background noise, and specific subject areas.

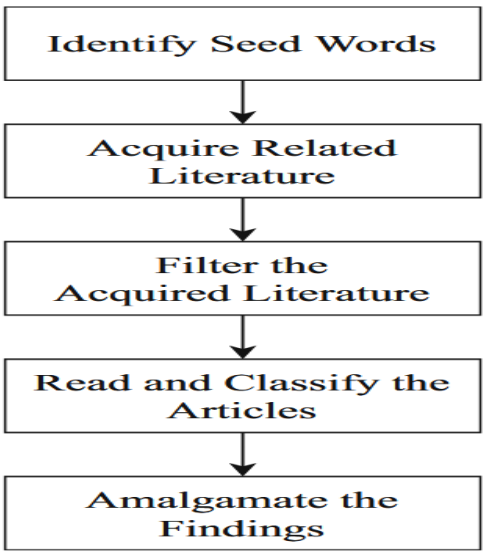


Figure 1: Overview of search method [13].

5. Evolution of speech recognition

Bell Laboratories' "Audrey" device, a single-speaker digit recognizer that could identify the numbers 0 through 9, marked the beginning of speech recognition in the 1950s. Other early accomplishments include IBM's "Shoebox" in the 1960s, which was able to recognize 16 words, and Carnegie Mellon University's "Harpy" system in the 1970s, which was partially funded by DARPA and was able to comprehend a considerably wider vocabulary of 1,000 words. As given in Table 1.

Table 1: Evolution of speech recognition

Years	Revolution
1950	Bell Laboratories developed "Audrey," the first speech recognition system that can accurately identify spoken numbers 0 through 9 from a single speaker.
1960	IBM created the "Shoebox," which recognized 16 English words, and other Soviet systems also made progress in vocabulary size.
1970	The US government-funded "Harpy" program at Carnegie Mellon University developed a system that could recognize entire sentences with a 1,000-word vocabulary. With the advent of methods like Hidden Markov Models (HMMs), the groundwork for contemporary statistical models was also established.
1980	With the creation of statistical models and IBM's voice-activated typewriter, Tangora, speech recognition technology advanced further.
1990	Dragon Dictate was launched, and its successor, Dragon Naturally Speaking (1997), became a widely used consumer-grade speech recognition product.
2000	With the introduction of Google Voice Search (2007) and the development of the internet and cloud computing, enormous volumes of data were available for training, and accuracy was greatly improved by connecting user speech with search results.
2010-present	Speech recognition capabilities saw a "hyperspace jump" with the advent of deep learning, specifically Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. Voice-based interactions become even more common thanks to companies like Apple (Siri, 2011), Amazon (Alexa), and Google (Google Home).

5. Types of Speech Recognition

Speech falls into the following categories.

5.1. Isolated Words Recognition: As our first illustration, think about creating an isolated word recogniser with HMMs. [2] This class might be better known as Isolated Utterance. [9] Users must pause clearly between each spoken word in a system based on isolated-word utterances. [13] Because of this technology, a class of applications known as "command-and-control" applications became possible. In these applications, the system could identify a single-word command (from a limited vocabulary) and respond appropriately. [1]

5.2.Connected Words Recognition: Word recognition that is connected to both self-sufficient and trained speakers.[1] This recognition system employs pauses to distinguish words. [9] comprises a system that operates with linked utterances and will halt between two or more words, either nominally or not at all. These systems can process multiple words at once rather than one at a time. [13]

5.3. Continuous Speech: Continuous speech recognition, both speaker-independent and speaker-trained. [1] Consequently, the performance of continuous speech recognition systems is affected by unknown boundary information regarding words, co-articulation, creation of adjacent phonemes, and speech pace. [9]**.** As mentioned in the figure. 2.

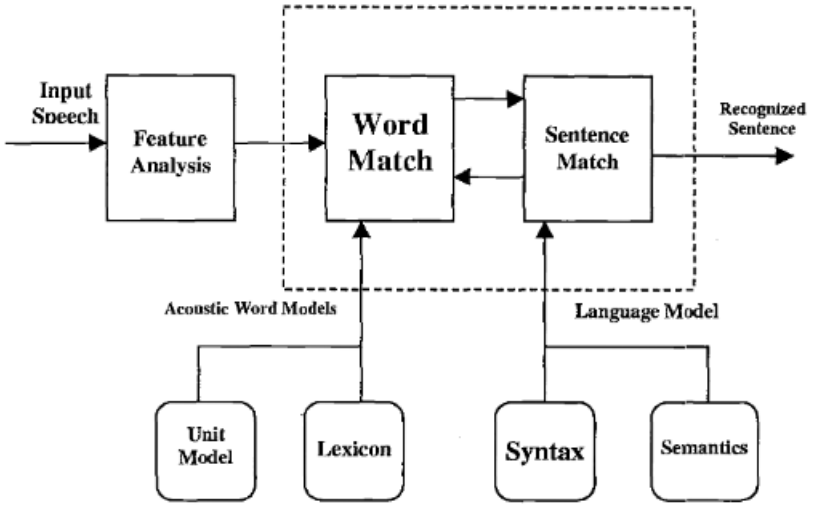


Figure 2: An example of an integrated continuous speech recognizer block diagram. **[1]**

7. Application of Speech Recognition

The ability to automatically identify the speaker rather than the content of what is being spoken is known as speaker recognition.[5] A type of biometric technology called speaker recognition uses a person's voice's distinctive qualities, such as pitch, tone, and vocal tract shape, to either identify or confirm them. The system creates a digital voiceprint from these features. Speaker verification confirms an individual's claimed identity, such as for device access, while speaker identification determines who is speaking from a known group. This technology is crucial for applications like voice-based banking, virtual assistants, and security. However, it faces challenges from background noise, vocal variations, and mimicry, and ongoing advancements in machine learning are working to improve its accuracy. Our goal is to group the statements based on the identities of the speakers. In the literature, most speaker clustering algorithms use hierarchical clustering with different distance metrics.[12]

7.1. Medical Assistance

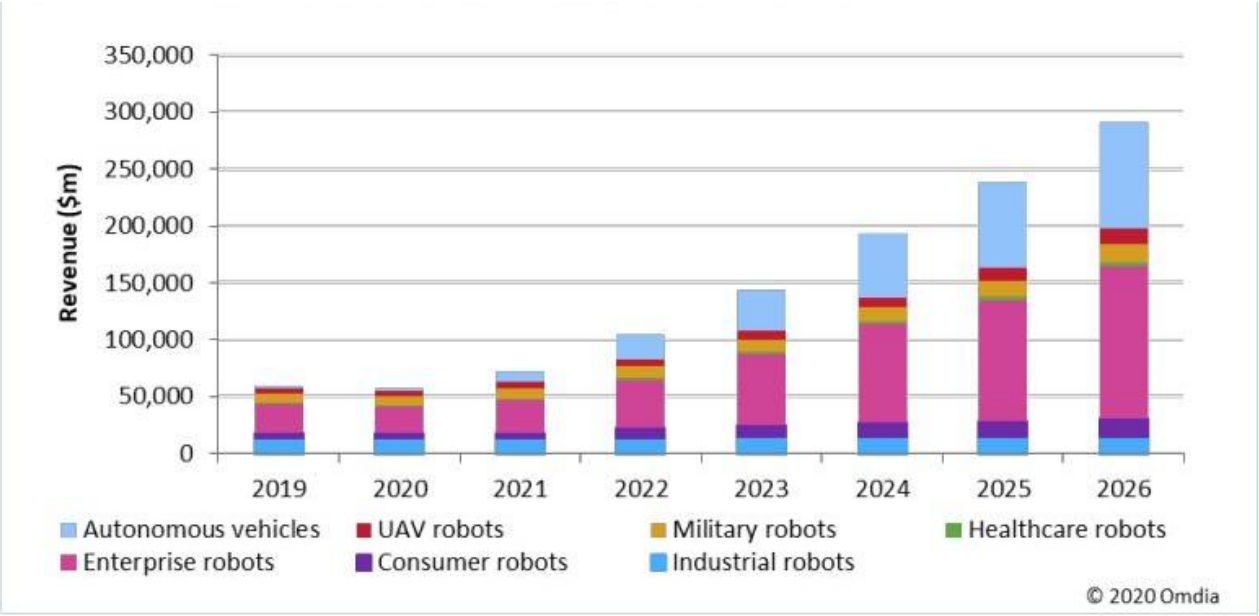
In the medical field, speech recognition technology is being utilized more and more to help patients and healthcare professionals. One of its main uses is in clinical documentation, where physicians and nurses dictate prescriptions, treatment plans, and patient notes straight into electronic health records (EHRs) using voice-to-text technology. This improves accuracy, reduces the time spent on manual data entry, and frees up healthcare workers to spend more time providing patient care. Speech recognition also enables hands-free operation of devices and systems in environments such as operating rooms, where maintaining sterility is crucial. By using voice commands, doctors can access medical records, imaging, or tools without

interrupting a procedure. For patients, especially those with physical disabilities or mobility limitations, speech recognition provides a vital communication channel. It enables them to interact with medical staff, control assistive devices, and manage aspects of their care through simple voice commands, thereby enhancing independence and quality of life. In telemedicine, speech recognition can transcribe real-time conversations between doctors and patients, ensuring accurate records and improving accessibility, particularly for those with hearing impairments, when combined with captioning tools.

7.2. Industrial robotics

Speech recognition is becoming an important component in industrial robotics, enabling more natural and efficient interaction between humans and machines in manufacturing and automation settings. Traditionally, industrial robots have been controlled through complex programming interfaces or physical controls. However, with the integration of speech recognition, operators can now give voice commands to robots, simplifying control processes and reducing the need for specialized training. Workers can operate machines while maintaining their hands and attention on the task at hand thanks to this, which is especially helpful in settings where hands-free operation is crucial, such assembly lines or dangerous work locations. The use of speech recognition in industrial robotics also enhances safety and productivity. Voice commands can quickly and efficiently start or stop machines, adjust settings, or trigger emergency shutdowns, especially in situations where reaching a control panel may be difficult or dangerous. In collaborative robotics (cobots), where robots work alongside human workers, speech recognition enables smoother coordination by allowing workers to communicate directly with the robot in real time, improving teamwork and reducing errors caused by miscommunication or delays.

Additionally, speech recognition in industrial robotics supports adaptability and customization in manufacturing processes. For example, in environments where products or tasks change frequently, voice commands can help reconfigure robotic actions without the need for reprogramming, saving time and improving flexibility. It also supports multi-language functionality, making it easier for international or multilingual workforces to interact with robotic systems. However, challenges remain in deploying speech recognition in industrial environments. Factors like high levels of background noise, varied accents, and specialized vocabulary can impact recognition accuracy. Moreover, ensuring system reliability and preventing unintended commands is critical in high-risk industrial applications. The performance of voice recognition in industrial robotics is continuously improving despite these obstacles, to developments in machine learning, contextual language processing, and noise-cancelling technologies. It is anticipated that these systems will increasingly contribute to the development of intelligent, responsive, and user-friendly industrial automation solutions as they grow stronger.



Source: Omdia

Figure 3: Estimated yearly exports of industrial robots worldwide

7.3. Forensic and law enforcement

Speech recognition technology is becoming more and more important in law enforcement and forensic investigations because it makes it easier to identify suspects, confirm identities, and analyse audio evidence more effectively. One of the primary applications is speaker recognition, which involves analyzing a person's voice to either confirm their identity (speaker verification) or determine who is speaking from a group of known individuals (speaker identification). This can be critical in cases involving threatening phone calls, ransom demands, intercepted communications, or covert recordings. By comparing a suspect's voice with recordings from a crime scene, forensic experts can use speech recognition systems to support criminal investigations and provide evidence in court. In law enforcement, speech recognition is also used to enhance surveillance capabilities. Police and security agencies can process large volumes of intercepted audio data, such as phone calls or recordings from surveillance devices, more quickly using automated speech-to-text systems. This allows investigators to search transcriptions for keywords, patterns, or names relevant to an investigation, significantly speeding up the evidence review process. Additionally, real-time speech recognition can be integrated into emergency response systems to transcribe and analyze 911 calls or radio communications, improving response time and situational awareness. Another important application is in body- and dashboard-mounted cameras used by police officers. Speech recognition can be used to automatically transcribe spoken interactions between officers and the public, creating detailed records of encounters that can be reviewed for accountability, training, or legal proceedings. In some cases, voice analysis tools may also help detect stress, deception, or emotional states in speakers, potentially offering additional investigative leads. Despite its advantages, the use of speech recognition in forensic and law enforcement settings faces challenges. Legal and ethical concerns around privacy, consent, and the admissibility of

voice evidence must be carefully managed. Furthermore, background noise, voice disguise, and poor-quality recordings can reduce accuracy, and false positives or errors in speaker identification could have serious legal consequences. Nevertheless, as technology advances, speech recognition is expected to become an increasingly valuable tool for forensic analysis and public safety, complementing traditional investigative methods with faster, data-driven insights.

7.4. Defence & Aviation

The defence and aviation industries are increasingly adopting speech recognition technology to enhance communication, improve operational effectiveness, and support mission-critical tasks. In military operations, where speed, precision, and situational awareness are crucial, speech recognition enables hands-free and real-time interaction with equipment, systems, and vehicles. For example, pilots, soldiers, or operators in command centres can issue voice commands to control weapons systems, navigate digital maps, or access mission data without diverting attention from their primary tasks. This hands-free feature is especially useful in combat or high-stress situations where quick reaction is crucial. In aviation, speech recognition is used to assist pilots by automating certain cockpit functions and reducing workload. Pilots can use voice commands to control flight systems, communicate with onboard computers, retrieve checklists, or adjust navigation settings. This helps improve situational awareness, particularly during crucial phases of flight such as takeoff, landing, and emergency situations. Voice-activated systems also reduce the need for physical interaction with control panels, making the cockpit environment safer and more efficient. Despite its benefits, speech recognition in defense and aviation must overcome significant challenges. These include dealing with noisy environments, handling multiple accents or languages, and ensuring system reliability under extreme conditions. Security is also a major concern, as voice-command systems must be protected from spoofing or unauthorised access. Nonetheless, continuous developments in secure voice authentication, noise cancellation, and machine learning are progressively enhancing the resilience and dependability of speech recognition systems in these challenging domains. Because of this, the technology is set to become increasingly significant in improving the safety and efficacy of defence and aviation operations.

7.5. Telecommunications Industry

Speech recognition technology has significantly transformed the telecommunications industry by enhancing the way service providers interact with customers and manage communication services. One of its most prominent applications is in **automated customer service systems**, where voice recognition allows users to navigate phone menus, request support, and perform transactions simply by speaking. Long wait periods and touch-tone inputs are eliminated, providing a more effective and user-friendly experience. Telecom firms today frequently deploy voice-enabled interactive voice response (IVR) systems to manage a high frequency of standard client inquiries, which eliminates the need for human agents and lowers operating expenses.

Another important application is in “call transcription and analysis”. Speech recognition systems are used to transcribe customer service calls in real time, enabling telecom providers to monitor conversations for quality assurance, regulatory compliance, and sentiment analysis. These transcriptions help identify trends in customer complaints, measure service performance, and ensure that agents follow communication guidelines. Advanced speech analytics can even detect emotions such as frustration and satisfaction, enabling companies to respond more effectively and improve customer satisfaction. In addition to improving operations, speech recognition is playing a key role in “accessibility and inclusivity”. In order to enable those who are deaf or hard of hearing to engage more completely in voice communications, telecom providers are utilizing the technology to provide real-time captioning services. This makes telecommunications services more inclusive and aligns with regulatory requirements that ensure equal access to communication technologies. Moreover, speech recognition is being integrated into “virtual assistants, smart devices, and telecom apps”. Customers can now use voice commands to check their data usage, pay bills, troubleshoot issues, or control connected home devices. Many telecom companies are also implementing “voice biometrics” for user authentication, allowing customers to verify their identities securely using their unique voice patterns, thereby enhancing both convenience and security.

7.6. Home automatic and security access

Speech recognition technology has become a vital part of modern home automation and security systems, enabling homeowners to conveniently and intuitively control their living environments. With voice-activated smart home devices, residents can effortlessly manage lighting, heating, entertainment systems, and appliances simply by speaking commands. In addition to being more convenient, this hands-free control improves accessibility for people with disabilities or mobility issues. By using commands like "Turn off the lights" or "Set the thermostat to 72 degrees," users can easily control many aspects of their homes without physically engaging switches or apps. In terms of security, speech recognition provides an additional layer of protection by enabling voice-based authentication for access control. Instead of relying solely on traditional methods like keys, PINs, or biometric scans, homeowners can use unique voiceprints to unlock doors or disable alarms. This voice biometric approach helps prevent unauthorized entry, as the system recognizes only registered voices. Furthermore, voice commands can be used in emergencies to quickly alert authorities or activate security measures, ensuring faster response times. Despite challenges such as background noise or potential voice spoofing, ongoing advances in speech recognition technology continue to improve reliability, making voice-controlled home automation and security systems an increasingly popular and effective solution for modern households.

7.7. Information technology and consumer electronics

Speech recognition technology has become a fundamental component of the information technology (IT) industry, transforming how users interact with software and digital platforms. In IT, speech recognition enables more natural and efficient communication with computers, allowing users to dictate documents, send emails, and control applications using voice

commands. This reduces dependency on traditional input methods like keyboards and mice, improving accessibility for users with disabilities and increasing productivity for professionals who benefit from hands-free operation. Additionally, speech recognition powers virtual assistants and AI-driven chatbots that help automate customer service, provide personalized support, and streamline business processes, making interactions faster and more intuitive.

In the realm of consumer electronics, speech recognition has revolutionized everyday devices, making them smarter and easier to use. Smartphones, smart speakers, televisions, and even household appliances now come equipped with voice-activated assistants like Google Assistant, Apple's Siri, and Amazon Alexa. These assistants allow users to do simple voice commands to search the internet, play music, control smart home appliances, and set reminders. This technology enhances convenience and accessibility, especially for people who find traditional interfaces challenging. Continuous improvements in natural language processing and noise reduction have made speech recognition more accurate and reliable, even in noisy environments. As a result, speech recognition is becoming an essential feature that enriches user experience and drives innovation across both the IT sector and consumer electronics.

7.8. The Challenge

There is significant promise in using automatic speech recognition in computer-based tools for children's speech and language development. Even if these resources can't replace the human connection that occurs when a parent or teacher helps a child learn to read, they could significantly improve the individualised support a child receives and enable more efficient use of precious time spent with the parent or instructor [3]. It is widely known that there can be a significant performance difference between laboratory systems (as far as we know, no performance metrics exist for commercial dictation systems). For example, the word error rates for the best (1%–2%) and worst (25%–30%) speakers can differ by a factor of 20 or more. There are several reasons for this, but the primary ones are the speaking rate and manner [8].

7.9. Performance of Systems

Accuracy and speed are typically used to describe the performance of voice recognition systems. While speed is measured using the real-time factor, accuracy can be determined using performance accuracy, which is often graded using word error rate (WER). Command Success Rate (CSR) and Single Word Error Rate (SWER) are additional accuracy metrics [10].

7.10. Human Speech Perception Versus ASR

Ferrier et al. (1995) and Doyle et al. (1997) hypothesized that speaker-adaptable, discrete word ASR technologies (e.g., Dragon Dictate) may outperform the human listener in recognizing dysarthric speech when speech intelligibility reaches moderate to severe levels of dysarthria. This is because the two distinct cases in which the severely dysarthric speakers were unintelligible to a casual listener were consistent enough for an ASR system to recognize them with a relatively high degree of accuracy [11].

8. Conclusion

Speech recognition has evolved from simple rule-based systems like Bell Labs' Audrey to sophisticated deep learning–powered models that drive today's virtual assistants, smart devices, and industrial solutions. Its applications span across diverse domains such as healthcare, education, telecommunications, defence, robotics, and consumer electronics, demonstrating its transformative impact on human–computer interaction. By enabling natural, voice-based communication, ASR enhances accessibility, efficiency, and inclusivity, especially for individuals with disabilities. Despite remarkable progress, challenges persist, including variability in accents, background noise, speaking styles, and ethical concerns surrounding privacy and security. Performance gaps between human and machine recognition, though narrowing, still exist in complex environments. Research continues to focus on improving accuracy through advanced neural architectures, clustering methods, and robust language modeling techniques. As integration deepens, speech recognition is set to become even more pervasive, redefining interaction across personal, professional, and industrial domains. Its future lies in achieving greater adaptability, reliability, and contextual awareness, cementing its role as one of the most impactful technologies of the modern digital era.

Reference

[1] Rabiner, L. R. (1997, December). Applications of speech recognition in the area of telecommunications. In *1997 IEEE workshop on automatic speech recognition and understanding proceedings* (pp. 501-510). IEEE.

[2] Rabiner, L. R. (2002). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286.

[3] Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., & Barker, P. (1996, October). Applications of automatic speech recognition to speech and language development in young children. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* (Vol. 1, pp. 176-179). IEEE.

[4] Ganapathiraju, A., Hamaker, J. E., & Picone, J. (2004). Applications of support vector machines to speech recognition. *IEEE transactions on signal processing*, *52*(8), 2348-2355.

[5] Vajpai, J., & Bora, A. (2016). Industrial applications of automatic speech recognition systems. *International Journal of Engineering Research and Applications*, *6*(3), 88-95.

[6] Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., & Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE access*, *7*, 19143-19165.

[7] Deng, L., Hinton, G., & Kingsbury, B. (2013, May). New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 8599-8603). IEEE.

[8] Gauvain, J. L., & Lamel, L. (2002). Large-vocabulary continuous speech recognition: advances and applications. *Proceedings of the IEEE*, *88*(8), 1181-1200.

[9] Arora, S. J., & Singh, R. P. (2012). Automatic speech recognition: a review. *International Journal of Computer Applications*, *60*(9).

[10] Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition technique. *International Journal of Computer Applications*, *10*(3), 16-24.

[11] Young, V., & Mihailidis, A. (2010). Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assistive Technology*, *22*(2), 99-112.

[12] Chen, S. S., & Gopalakrishnan, P. S. (1998, May). Clustering via the Bayesian information criterion with applications in speech recognition. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)* (Vol. 2, pp. 645-648). IEEE.

[13] Malik, M., Malik, M. K., Mehmood, K., & Makhdoom, I. (2021). Automatic speech recognition: a survey. *Multimedia Tools and Applications*, *80*(6), 9411-9457.