iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75881

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Voice-Controlled Robotics: Evolution and Trends

Harshvardhan Dwivedi[1], Sahil Ghorpade[2], Atharva Wadekar[3], Prasad Swami[4], Soham Lotlikar[5]
Electronics & Communication Department, Thakur College of Engineering and Technology

Abstract: ***The evolution of automated voice-controlled systems has transformed human–robot interaction from rudimentary mechanical devices into sophisticated, AI-driven platforms. Early developments in speech synthesis and recognition, such as Kratzenstein's vowel models and von Kempelen's speaking machines, provided the foundation for subsequent electronic and statistical approaches. Breakthroughs in probabilistic modeling with Hidden Markov Models and the advent of commercial products like Dragon NaturallySpeaking expanded the accessibility of speech recognition. More recently, the integration of deep learning, large language models, and edge computing has enabled near-human accuracy, multilingual adaptability, and real-time processing. This review paper traces the historical progression of voice-controlled systems, examines core enabling technologies including automatic speech recognition, natural language processing, and robotic integration frameworks, and highlights diverse applications in manufacturing, healthcare, defense, and space exploration. Current challenges such as noise resilience, accent variation, privacy risks, and interoperability are analyzed alongside emerging optimization strategies. Future directions emphasize multimodal interaction, ethical AI frameworks, and the potential of neuromorphic and quantum computing for next-generation robotics. By synthesizing these developments, the paper underscores the transformative role of voice-controlled robotics across industries and outlines research opportunities to advance their global deployment.***

Keywords: Voice-Controlled Robotics, Speech Recognition, Natural Language Processing, Deep Learning, Human–Robot Interaction, Edge Computing, Multimodal Systems, Ethical AI.

I. INTRODUCTION

The convergence of speech recognition technology, natural language processing (NLP), and robotic automation has fundamentally transformed human–machine interaction paradigms over the past two centuries. Early research efforts, such as Christian Kratzenstein's vowel synthesizers (1779) and Wolfgang von Kempelen's acoustic-mechanical speech machine (1791), demonstrated the feasibility of reproducing human speech through engineered systems. Later developments by Charles Wheatstone and Alexander Graham Bell established a direct technological lineage from mechanical synthesis to modern telecommunication and speech technologies [1].

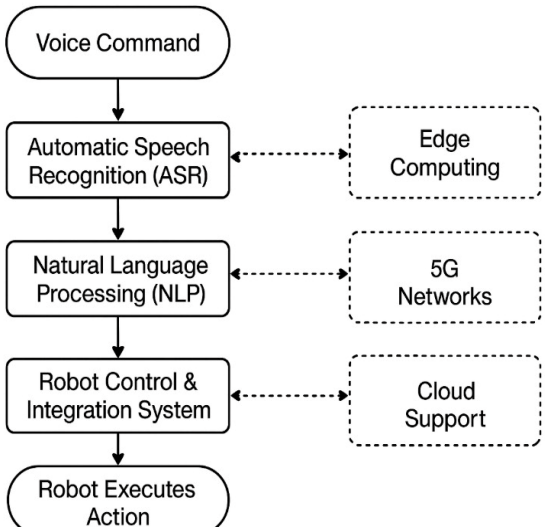


Fig. 1 Basic Flowchart of a Voice Controlled Robot

The transition to electronic speech processing was marked by Homer Dudley's Vocoder (1930s) and the Bell Labs "Audrey" system (1950s), which represented the first steps toward automatic speech recognition (ASR).

A major paradigm shift occurred in the 1970s with the adoption of statistical approaches, particularly Hidden Markov Models (HMMs), which enabled robust handling of variability in speech patterns and acoustic conditions. Supported by government-funded projects such as DARPA's Speech Understanding Research program, systems like Carnegie Mellon's Harpy achieved vocabularies of over 1,000 words, setting the foundation for future large-scale recognition systems [2].

By the 1990s, speech recognition entered commercial domains with products such as Dragon Dictate and Dragon NaturallySpeaking, which pioneered continuous speech recognition for consumer and industrial use. The field was further revolutionized in the 21st century by deep learning and cloud computing, exemplified by Google Voice Search and Apple's Siri, which achieved word error rates below 5% under optimal conditions while enabling multilingual and context-aware interaction [3].

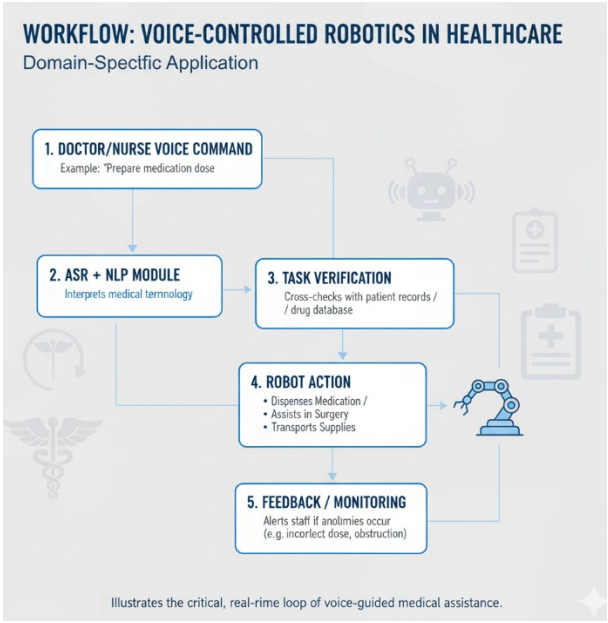


Fig. 2 Workflow of Voice Controlled Robotics in Healthcare

Today, voice-controlled robotics has been widely adopted across manufacturing, healthcare, defense, and space exploration. Modern implementations demonstrate measurable benefits, including up to 90% reductions in robot programming time, 25–60% improvements in healthcare productivity, and enhanced mission performance in military and aerospace applications. Despite these advances, challenges such as noise management, linguistic diversity, privacy, and latency remain critical obstacles. Future directions emphasize multimodal interaction, edge intelligence, and ethical AI frameworks, supported by emerging technologies such as 5G, neuromorphic computing, and quantum processing [4].

II. LITERATURE REVIEW

Research on automated voice-controlled systems has evolved over centuries, with each era introducing foundational advancements that shaped modern human–robot interaction. Early works in the late 18th and 19th centuries by Kratzenstein and von Kempelen established the mechanical foundations of artificial speech synthesis, demonstrating that engineered systems could reproduce vowel sounds and rudimentary words. Subsequent developments by Wheatstone and Bell linked mechanical speech devices to the later invention of the telephone, creating a bridge between synthetic speech and modern telecommunications [5].

The transition to electronic speech processing occurred in the 20th century with Homer Dudley's Vocoder (1935), which demonstrated the feasibility of electronic speech synthesis and encryption for secure communications. The introduction of Bell Labs' "Audrey" system in the 1950s and IBM's "Shoebox" in 1962 marked the first attempts at automatic speech recognition (ASR), though these were limited to digits and small vocabularies. The mathematical revolution of the 1970s, led by Baum's introduction of Hidden Markov Models (HMMs), transformed speech recognition from template matching to probabilistic modeling, enabling scalability to larger vocabularies and diverse speakers. DARPA-funded projects such as Carnegie Mellon's Harpy system further advanced vocabulary size and recognition accuracy [2].

By the 1990s, commercial systems such as Dragon Dictate and Dragon NaturallySpeaking brought continuous speech recognition to consumer markets, albeit with high costs and extensive training requirements. The late 1990s also saw the rise of voice portals like BellSouth's VAL and large-scale deployments in financial services, proving the business viability of speech recognition. However, traditional statistical methods plateaued around 80% accuracy by 2001, highlighting the need for a new paradigm.

The deep learning revolution marked a turning point. Google's Voice Search (2008) leveraged cloud computing and vast datasets to dramatically improve accuracy, while Apple's Siri (2011) introduced conversational interfaces combining ASR with natural language processing (NLP). Transformer-based neural architectures have since driven breakthroughs in accuracy, achieving word error rates below 5% under optimal conditions. Current research emphasizes edge AI, 5G connectivity, and multimodal systems that integrate speech with vision and haptics, reducing latency and enhancing contextual understanding [6].

Scholarly attention has also focused on application domains. In manufacturing, voice interfaces cut robot programming time by up to 90% and streamline quality assurance processes. Healthcare literature highlights applications in surgical robotics, patient care, and telemedicine, improving efficiency and reducing administrative burden. Defense and security studies emphasize tactical command of autonomous systems, surveillance, and multi-robot coordination. Space research demonstrates the utility of voice-controlled systems for astronaut support and autonomous planetary missions [7].

Across the literature, persistent research gaps are evident. Studies highlight ongoing challenges with noise resilience, accent and dialect variability, code-switching, and domain-specific jargon. Ethical concerns related to privacy, surveillance, and bias in speech datasets remain underexplored. Finally, while multimodal interaction and neuromorphic computing are promising directions, empirical studies validating their deployment in real-world robotics remain limited.

III. HISTORICAL EVOLUTION

The development of automated voice-controlled systems can be divided into five key stages. The earliest attempts at mechanical speech synthesis began in the late 18th century. In 1779, Christian Kratzenstein created vowel resonators that mimicked the human vocal tract. Wolfgang von Kempelen advanced this in 1791 with his "Acoustic-Mechanical Speech Machine," which simulated lungs, vocal cords, and the vocal tract to produce both vowels and rudimentary consonants. Charles Wheatstone later improved these machines in the mid-1800s, enabling them to articulate entire sentences, while Alexander Graham Bell's experimental devices linked mechanical synthesis to the later invention of the telephone [5].

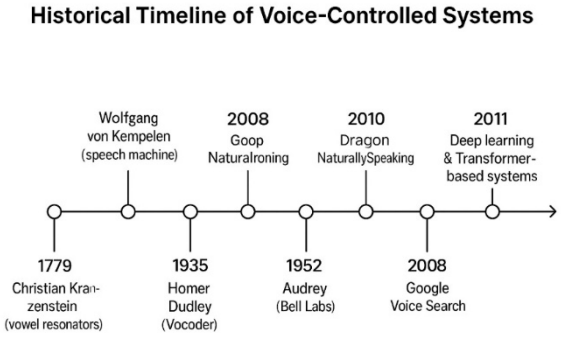


Fig. 3 History of Voice Controlled Systems

The transition to electronic speech synthesis and recognition occurred in the 20th century. Homer Dudley developed the Vocoder at Bell Labs in 1935, the first successful electronic speech synthesizer, later applied in the SIGSALY secure communication system during World War II. The first automatic speech recognition (ASR) system, Audrey, was introduced at Bell Labs in 1952, capable of recognizing spoken digits. IBM further demonstrated speech recognition with its "Shoebox" machine at the 1962 World's Fair, which recognized 16 English words [2].

The statistical modeling era emerged in the 1970s with the introduction of Hidden Markov Models (HMMs) by Lenny Baum, providing probabilistic approaches that overcame limitations of template matching. The U.S. DARPA Speech Understanding Research Program (1971) funded large-scale research, including Carnegie Mellon's Harpy system, which achieved a vocabulary of 1,011 words and pioneered efficient search methods. Commercial applications appeared as well, such as Texas Instruments' Speak and Spell (1978), which introduced digital speech technology to consumers.

Through the 1980s and 1990s, HMM-based approaches dominated, enabling vocabulary expansion into the thousands and improved robustness across speakers and environments.

In the 1990s and early 2000s, speech recognition reached mass-market adoption. Dragon Systems, founded in 1982, launched Dragon Dictate in 1990 and Dragon NaturallySpeaking in 1997, which introduced continuous speech recognition, allowing users to speak naturally without pausing. SpeechWorks, launched in 1984, applied ASR in telephone-based services, while BellSouth's VAL (1996) pioneered the first interactive voice portal. By 2001, recognition accuracy approached 80% in laboratory settings, though Internet-driven language diversity exposed the limitations of traditional statistical models [8].

The most recent stage, the deep learning revolution, began in the 2000s. Google Voice Search (2008) leveraged cloud computing and massive datasets to improve recognition accuracy dramatically, while Apple's Siri (2011) integrated cloud-based ASR with contextual natural language understanding and conversational interaction. Modern systems employ deep neural networks and transformer-based architectures, achieving word error rates below 5% in optimal conditions while supporting multiple languages and accents. These advances, combined with cloud and edge computing, have enabled real-time, large-scale deployment of voice interfaces, representing the culmination of progress from Kratzenstein's vowel resonators to today's AI-driven conversational systems [9].

IV. CORE TECHNOLOGIES

At the foundation of voice-controlled robotics lies Automatic Speech Recognition (ASR), which transforms raw acoustic signals into text representations. Modern ASR systems employ deep neural network architectures and transformer-based models that extract linguistic meaning from speech with word error rates below 5% under optimal conditions. These systems integrate advanced preprocessing techniques such as Gaussian masking, Wiener filtering, and beamforming microphone arrays to mitigate environmental noise and improve recognition accuracy in challenging industrial, healthcare, and defense settings. End-to-end ASR frameworks further simplify the recognition pipeline by directly mapping audio inputs to textual outputs, achieving superior performance compared to earlier hybrid approaches [10].

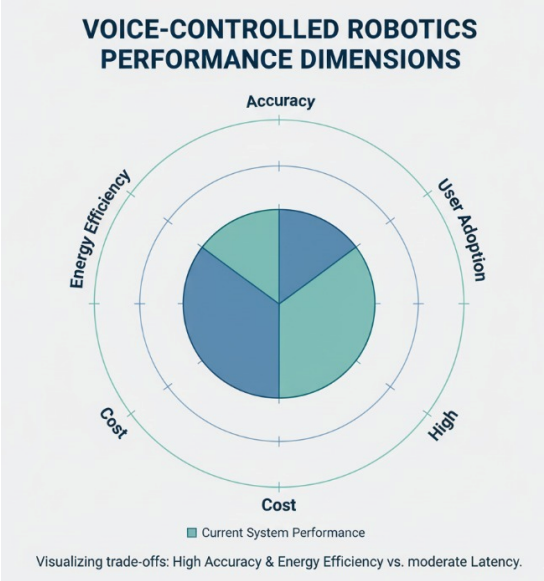


Fig. 4 Performance Dimensions

Building on ASR outputs, Natural Language Processing (NLP) plays a pivotal role in interpreting user intent and generating actionable robotic commands. NLP modules leverage large language models (LLMs) trained on massive corpora to understand semantics, context, and idiomatic expressions. These systems support intent recognition, named entity extraction, semantic parsing, and dialogue management, enabling robots to process commands such as location-based tasks, operational modes, or sequential actions with contextual continuity. Modern NLP frameworks incorporate few-shot learning and memory-based models to ensure adaptability to new vocabularies and real-world conversational complexities [11].

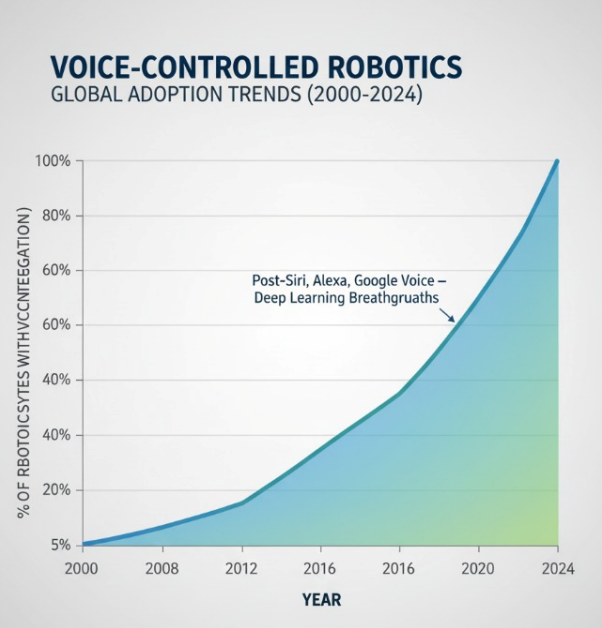


Fig. 5 Global Adoption Trends

To bridge human language with physical action, robot control and integration systems translate parsed commands into executable behaviors. These frameworks ensure that high-level natural language inputs are mapped into low-level control signals that govern motion, manipulation, and sensing tasks. Advanced planning algorithms optimize robot trajectories while maintaining safety, obstacle avoidance, and real-time responsiveness. Furthermore, integrated safety monitoring systems employ redundant control layers, hardware interlocks, and human override options, ensuring that voice-directed robotic actions remain reliable and secure in mission-critical scenarios [12].

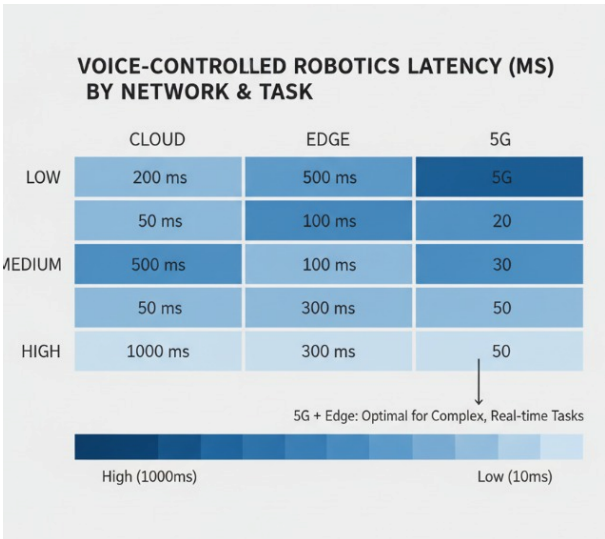


Fig. 6 Latency by Network & Task Chart

Finally, the integration of edge computing and 5G networks has significantly enhanced the efficiency and responsiveness of voice-controlled robotics. Edge architectures minimize latency by performing recognition and processing locally, reducing response times from hundreds of milliseconds in cloud-based systems to below 50 milliseconds. Optimization techniques such as model quantization and hardware acceleration allow advanced neural networks to run on embedded devices while conserving power. Meanwhile, 5G networks provide ultra-low latency (under 10 ms), high bandwidth, and network slicing capabilities that enable seamless multi-robot coordination and remote operations such as telesurgery, hazardous material handling, and space exploration. Together, edge computing and 5G form the backbone of next-generation, real-time voice-controlled robotic systems [13].

V. APPLICATIONS

The integration of voice-controlled robotics has significantly transformed industrial applications, particularly in manufacturing, logistics, and quality control. In production environments, voice interfaces reduce robot programming time by up to 90% compared to conventional methods, enabling faster reconfiguration for diverse tasks. Manufacturing floors benefit from noise-resilient systems employing multi-microphone arrays and beamforming techniques that maintain recognition accuracy above 95% even in environments exceeding 85 dB. Logistics and warehouse operations leverage voice-directed robotics for inventory management, order picking, cross-docking, and returns processing, reducing picking errors by up to 40% while improving throughput. Similarly, quality control applications employ hands-free inspection, defect documentation, and compliance verification through spoken interaction, enhancing efficiency and regulatory compliance in industries such as aerospace and pharmaceuticals.

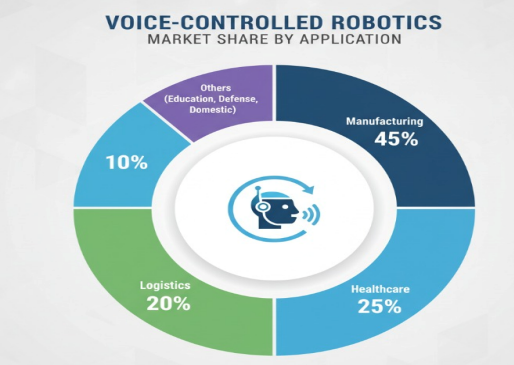


Fig. 7 Market Share by Application

In healthcare, voice-controlled robotics are revolutionizing surgical, clinical, and administrative operations. Surgical systems allow hands-free control of robotic instruments, imaging displays, and auxiliary equipment while maintaining sterile fields, with applications in neurosurgery, orthopedics, and minimally invasive procedures. Patient care robots integrate voice interfaces for medication management, rehabilitation assistance, vital sign monitoring, and cognitive assessments, improving outcomes while reducing healthcare staff workload. Telemedicine platforms further expand healthcare access by enabling remote consultations, medication reviews, and follow-up care through natural voice interaction. Administrative tasks such as patient transport, pharmacy operations, and hospital logistics are also streamlined via voice-driven mobile robots, improving efficiency and reducing errors in critical hospital workflows [14].

In the realm of defense and security, voice-controlled robotics have been deployed in combat systems, surveillance, and intelligence operations. The U.S. Army, for instance, has tested voice-controlled robotic tanks and unmanned ground vehicles capable of executing tactical maneuvers through natural language commands. Surveillance applications employ drones and ground robots for border patrol, perimeter security, and infrastructure protection, enabling real-time coordination with human operators. Moreover, voice-controlled systems enhance intelligence gathering, counter-surveillance operations, and force protection by providing secure and responsive command execution even in acoustically challenging combat environments. These applications demonstrate the strategic importance of natural language interfaces in modern military and security operations [15].

Finally, space exploration presents unique challenges where voice-controlled robotics offer significant advantages. NASA has developed voice management systems for spacecraft operations, life support, and experiment coordination in microgravity environments. Astronauts on the International Space Station (ISS) use voice-controlled robots for maintenance, cargo handling, and experiment management, enabling multitasking while reducing cognitive load. For planetary missions such as Mars exploration, voice-controlled rovers must autonomously interpret high-level voice commands due to communication delays of up to 20 minutes, facilitating geological surveys and sample collection. Future lunar and Martian missions envision extensive reliance on voice-enabled robotics for habitat construction, resource extraction, and scientific operations in extreme conditions [16].

VI. CURRENT CHALLENGES

One of the foremost challenges in voice-controlled robotics is managing noise and acoustic environments. Industrial facilities often exceed 85 dB, while outdoor military and mobile applications face interference from vehicles, engines, and environmental sounds. In such conditions, recognition accuracy can degrade drastically, with word error rates rising from under 5% in quiet conditions to over 50% in noisy environments.

Reverberation in large spaces such as warehouses and hospital corridors further complicates recognition, requiring advanced echo cancellation and adaptive filtering algorithms. Additionally, cross-talk and multi-speaker interference remain difficult to resolve in collaborative environments [17].

Another critical limitation arises from linguistic diversity, accents, and code-switching. With over 7,000 global languages and hundreds of dialects, no system can fully generalize across all speech variations. Current models exhibit bias toward dominant training datasets, often underperforming with non-native accents or regional pronunciations. Code-switching, where users alternate between languages within the same command, poses further complexity, particularly in multilingual populations and global deployments. Domain-specific jargon in technical fields such as medicine, law, and science also creates recognition barriers when absent from training datasets [18].

Privacy and security risks are another major concern, as voice data inherently contains biometric identifiers that cannot be revoked if compromised. Cloud-based ASR introduces vulnerabilities during transmission, processing, and storage, while always-listening devices raise concerns over surveillance and unintended recording. Adversarial attacks using spoofed or synthesized speech have demonstrated the potential to bypass authentication and trigger unauthorized robot commands. Furthermore, workplace monitoring risks emerge when voice-controlled systems continuously record employee communications.

Latency constraints significantly impact user experience, as natural conversation requires system responses within 200 milliseconds to feel intuitive. Yet, sequential pipelines of ASR, NLP, intent analysis, and robotic execution often introduce multi-stage delays. Network dependencies exacerbate the issue, with unpredictable latencies ranging from milliseconds to several seconds depending on connectivity and server load. Mobile and battery-powered robots face additional trade-offs between processing speed, accuracy, and energy consumption.

Finally, integration and interoperability challenges limit large-scale deployment. Legacy industrial and institutional systems often lack the APIs and protocols necessary for voice control, requiring costly retrofits. Standardization gaps between ASR platforms create vendor lock-in and hinder cross-platform compatibility. Real-time synchronization between recognition systems and robotic control loops further complicates deployment, as microsecond-level timing mismatches can destabilize safety-critical operations. Additionally, distributed architectures spanning edge, cloud, and robotic hardware increase maintenance complexity, while scalability issues arise when expanding from pilot projects to enterprise-wide implementations [19].

VII.FUTURE DIRECTIONS

The future of voice-controlled robotics will be shaped by advances in next-generation artificial intelligence (AI) and large language models (LLMs). Foundation models with hundreds of billions of parameters, trained on multimodal datasets, are enabling systems that can interpret nuanced instructions, perform contextual reasoning, and sustain extended dialogues with unprecedented accuracy. These models integrate command understanding with real-world knowledge, reducing the need for extensive task-specific training and facilitating rapid adaptation to specialized domains through few-shot learning techniques [20].

Another major trajectory is the rise of multimodal interaction, combining speech with visual perception, tactile sensing, and haptic feedback. By fusing modalities, robots can ground voice commands in environmental context, verify ambiguous instructions with vision-based recognition, and adapt to dynamic conditions more effectively. For example, a voice command such as “pick up the red tool on the left” can be validated by integrating visual recognition and proprioceptive sensing, thereby minimizing errors and enhancing safety [21].

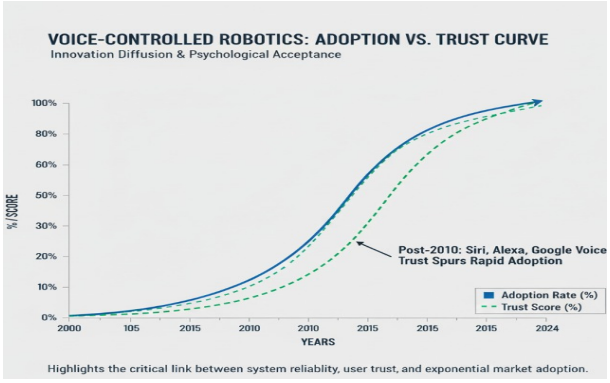


Fig. 8 Adoption vs Trust Curve

Emerging computational paradigms like quantum and neuromorphic computing promise transformative breakthroughs in processing capabilities. Quantum processors are expected to deliver instantaneous pattern recognition and optimization, overcoming the scaling bottlenecks of classical architectures. Similarly, neuromorphic chips modeled on biological neural systems offer microsecond response times with drastically reduced power consumption, making real-time conversational robotics viable even in mobile and resource-constrained platforms . These technologies hold the potential to eliminate latency, expand system scalability, and support increasingly complex robotic tasks.

Finally, the adoption of ethical AI frameworks will be essential for responsible deployment. Issues such as data privacy, surveillance risks, and bias in speech recognition models necessitate governance structures that ensure fairness, transparency, and accountability. Ethical frameworks will also guide the integration of autonomous voice-controlled systems in sensitive domains like defense, healthcare, and workplace environments, balancing innovation with safety and societal trust. As voice-controlled robotics become more autonomous and pervasive, embedding ethical principles in design and regulation will be critical for long-term sustainability and public acceptance.

VIII. CONCLUSION

The evolution of voice-controlled robotics highlights a remarkable journey from early mechanical speech synthesis experiments to today's deep learning-powered conversational systems. Over two centuries of innovation have transformed rudimentary vowel generators and template-based recognizers into intelligent platforms capable of natural, multimodal human-robot interaction. Modern systems integrate automatic speech recognition, natural language processing, and real-time robotic control to achieve hands-free operation with near-human accuracy, enabling seamless interaction across industrial, healthcare, defense, and space application. These contributions have redefined human–machine interaction paradigms by combining intuitive accessibility with technical sophistication [22].

The impact across industries is substantial. In manufacturing and logistics, voice-controlled robotics reduce programming time by up to 90% and improve warehouse efficiency through error reduction and optimized material flow. In healthcare, they support surgical precision, patient care, and telemedicine, reducing administrative overhead by 25–60% while enhancing quality of care. Defense and security implementations enable tactical command of autonomous systems and advanced surveillance capabilities in combat and critical infrastructure protection. In space exploration, voice-driven systems assist astronauts in multitasking operations aboard the ISS and enable autonomy in planetary missions where communication delays preclude real-time human control. Collectively, these applications demonstrate the wide-ranging societal and economic benefits of adopting voice interfaces in robotics [23].

Despite this progress, future research opportunities remain central to advancing the field. Addressing unresolved challenges such as code-switching, domain-specific jargon, and accent diversity will be vital for global accessibility. Developing robust multimodal frameworks that integrate speech with vision, haptics, and environmental context will enhance command accuracy and adaptability in dynamic environment. Emerging computational paradigms such as neuromorphic and quantum processors promise to eliminate latency bottlenecks and expand scalability. Finally, ethical frameworks are urgently needed to mitigate risks related to privacy, surveillance, and bias in deployment, ensuring responsible and equitable adoption of voice-controlled robotics. By addressing these research gaps, the next generation of systems will achieve even greater levels of intelligence, trustworthiness, and global impact [24].

IX. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Thakur College of Engineering and Technology for providing me with the opportunity to delve into the captivating realm of voice-controlled robotics. The institution's support and resources have been instrumental in facilitating my research, and I am thankful for the academic environment that encourages exploration and innovation.

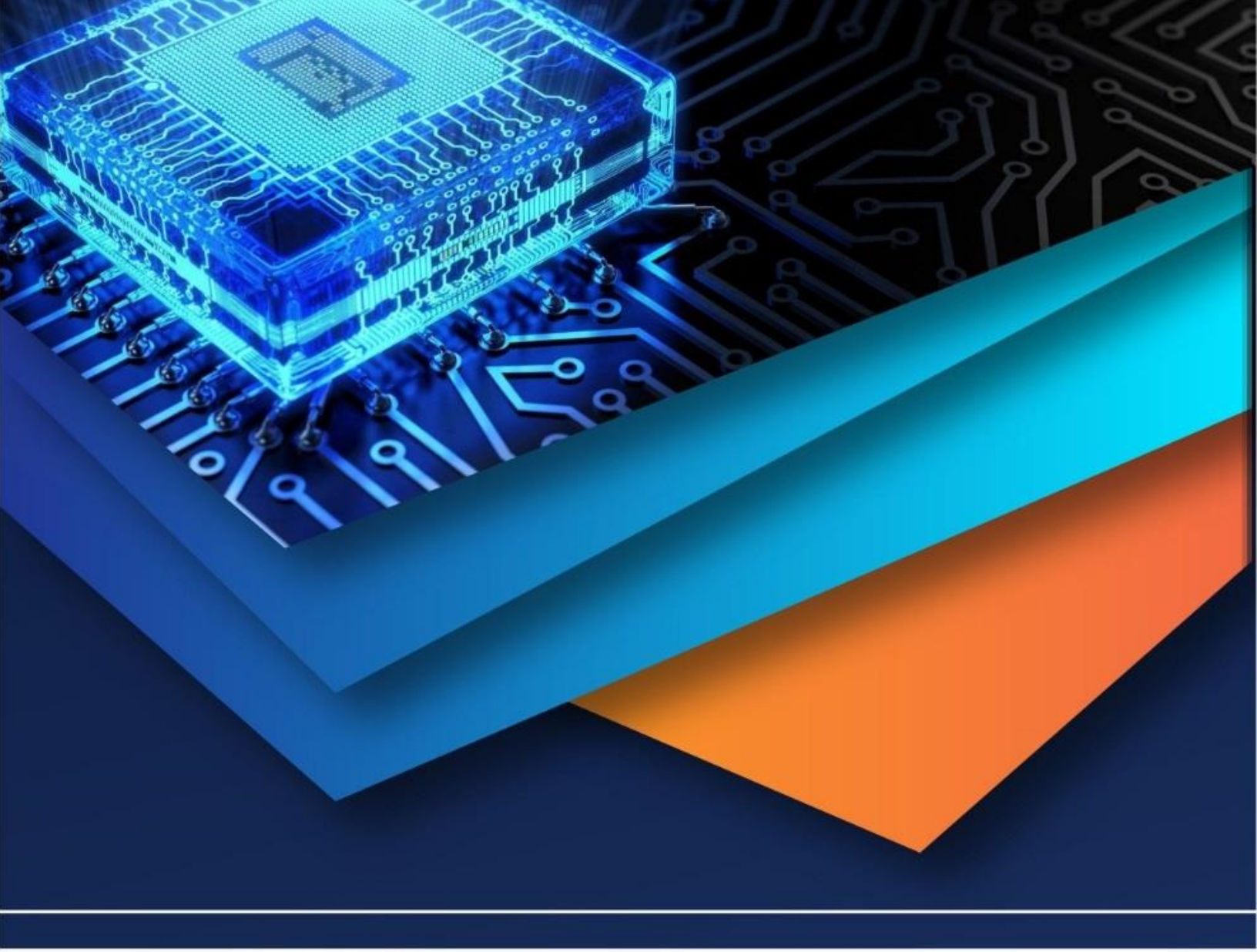
I am also deeply thankful to my professors and mentors who have guided me throughout this journey, sharing their vast expertise and insights that have helped shape this work. Their unwavering support, valuable feedback, and constructive criticism have been invaluable in refining my understanding and approach to this topic. I am grateful for the time they dedicated to discussing ideas, addressing queries, and providing direction, all of which have significantly contributed to the development of this review.

Furthermore, I would like to acknowledge the pioneering researchers and developers in the field of voice-controlled robotics, whose groundbreaking work has laid the foundation for this review.

Their contributions have not only enabled the growth of this field but also inspired further innovation and exploration. I am thankful for the wealth of knowledge and insights they have shared through their publications, presentations, and projects, which have served as a constant source of inspiration and motivation.

REFERENCES

[1] H. Zhou et al., "Language-conditioned Learning for Robotic Manipulation: A Survey," arXiv (Cornell University), 2023.
[2] S. Furui, "History and Development of Speech Recognition," Speech Technology, pp. 1-18, 2010.
[3] L. Rabiner, and B. Juang, "Historical Perspective of the Field of ASR/NLU," Springer Handbook of Speech Processing, pp. 521-538, 2008.
[4] Y. Kim et al., "A survey on integration of large language models with intelligent robots," Intelligent Service Robotics, vol. 17, no. 5, pp. 1091-1107, 2024.
[5] Z. Fagyal, "Phonetics and speaking machines," Historiographia Linguistica, vol. 28, no. 3, pp. 289-330, 2001.
[6] S. Latif et al., "Transformers in Speech Processing: A Survey," arXiv (Cornell University), 2023.
[7] M. Z. Iqbal et al., "Untitled," Physiology and molecular biology of plants : an international journal of functional plant biology, vol. 31, no. 10, pp. 1755-1774, 2025.
[8] R. Pieraccini, and D. Lubensky, "Spoken Language Communication with Machines: The Long and Winding Road from Research to Business," Lecture Notes in Computer Science, pp. 6-15, 2005.
[9] M. Z. Iqbal et al., "Untitled," Physiology and molecular biology of plants : an international journal of functional plant biology, vol. 31, no. 10, pp. 1755-1774, 2025.
[10] Y. Zhang et al., "Google USM: Scaling Automatic Speech Recognition Beyond 100 Languages," arXiv (Cornell University), 2023.
[11] B. Li et al., "Interactive Task Planning with Language Models," arXiv (Cornell University), 2023.
[12] K. Lin et al., "Text2Motion: From Natural Language Instructions to Feasible Plans," arXiv (Cornell University), 2023.
[13] Z. Lin et al., "Pushing Large Language Models to the 6G Edge: Vision, Challenges, and Opportunities," arXiv (Cornell University), 2023.
[14] J. Schreiter et al., "Multimodal human–computer interaction in interventional radiology and surgery: a systematic literature review," International Journal of Computer Assisted Radiology and Surgery, vol. 20, no. 4, pp. 807-816, 2024.
[15] S. G. Hill, D. Barber, and A. W. Evans, "Achieving the Vision of Effective Soldier-Robot Teaming," Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, pp. 177-178, 2015.
[16] K. Hambuchen, J. Marquez, and T. Fong, "A Review of NASA Human-Robot Interaction in Space," Current Robotics Reports, vol. 2, no. 3, pp. 265-272, 2021.
[17] V. Pratap et al., "Scaling Speech Technology to 1,000+ Languages," arXiv (Cornell University), 2023.
[18] W. Seymour et al., "A Systematic Review of Ethical Concerns with Voice Assistants," Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, pp. 131-145, 2023.
[19] S. Gallo, F. Paterno, and A. Malizia, "Conversational Interfaces in IoT Ecosystems: Where We Are, What Is Still Missing," Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia, pp. 279-293, 2023.
[20] Y. Kim et al., "A survey on integration of large language models with intelligent robots," Intelligent Service Robotics, vol. 17, no. 5, pp. 1091-1107, 2024.
[21] H. Li et al., "See, Hear, and Feel: Smart Sensory Fusion for Robotic Manipulation," arXiv (Cornell University), 2022.
[22] T. Mező, "Robots Communicate at the Speed of Light: Revolutionary Milestones in the Development of Human Speech," American Journal of Information Science and Technology, vol. 9, no. 2, pp. 69-78, 2025.
[23] M. D. Vu et al., "GPTVoiceTasker: LLM-Powered Virtual Assistant for Smartphone," arXiv (Cornell University), 2024.
[24] H. Zhou et al., "Language-conditioned Learning for Robotic Manipulation: A Survey," arXiv (Cornell University), 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY