International Journal of Trend in Research and Development, Volume 13(3), ISSN: 2394-9333
www.ijtrd.com

Natural Language Processing in Speech Recognition

[1]Matthew N. O. Sadiku and [2]Janet O. Sadiku,
[1]Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX, USA
[2]Juliana King University, Houston, TX, USA

Abstract: Speech recognition, also known as automatic speech recognition (ASR) or voice recognition, basically means talking to a computer and getting it to understand and interpret your spoken words. It is a transformative process that converts auditory signals into textual transcripts. It is a technology that enables a machine or program to identify and understand words or phrases from spoken language and convert them into machine readable format. It is a subfield of computational linguistics that deals with technologies to allow spoken input into systems. From the voice assistants in our pockets to the automated transcription services powering global business meetings, ASR has become a cornerstone of modern technology. Once the spoken language is transcribed into text, NLP takes center stage. NLP's role is to interpret the meaning and intent embedded within this textual data. NLP's contributions to speech recognition extend far beyond basic text interpretation. This paper explores the integration of NLP within modern speech recognition systems.

Key words*: Natural Language Processing (Nlp), Computational Linguistics, Speech, Speech Recognition (Sr), Automatic Speech Recognition (Asr), Voice Recognition*

I. INTRODUCTION

Speech recognition, also known as automatic speech recognition (ASR), is the technology that enables computers to convert spoken language into text or commands. It is basically designed for a single user. It powers voice assistants like Siri and Alexa, real-time transcription services, accessibility tools, and countless enterprise applications. From the virtual assistants in our pockets to the automated transcription services in our hospitals, the synergy of sound and sense is reshaping how we interact with technology.

Natural language processing (NLP) and speech recognition (SR) are two closely intertwined fields of artificial intelligence that, when integrated, significantly enhance human-computer interaction. While speech recognition primarily focuses on converting spoken language into written text, NLP takes this textual output and imbues it with meaning and context. This synergistic relationship has led to a multitude of benefits, transforming various industries and improving the efficacy of voice-enabled technologies. For example, when a user speaks a command to a voice assistant, the automatic speech recognition (ASR) component transcribes the spoken words, and the NLP component then identifies the specific command and its parameters, enabling the system to execute the desired action [1].

The natural language processing (NLP) and speech recognition have transformed language learning by providing interactive and real-time feedback, enhancing oral English proficiency. These technologies facilitate personalized and adaptive learning, making pronunciation and fluency improvement more efficient. The integration of natural language processing (NLP) and speech recognition technologies has become one of the most perceptively effective means to assist language learning, especially in oral English practices. In the modern era of deep learning, NLP has become a core architectural component of speech systems. With the emergence of end-to-end models, deep learning has revolutionized the field of automatic speech recognition (ASR). The speech recognition community has made great progress toward building deep neural networks for speech recognition by utilizing enormous amounts of training data and high-quality test sets [2].

II. FUNDAMENTALS OF NLP

Natural language processing is a subfield of artificial intelligence that empowers computers to understand, interpret, and generate human language. It is a technique where machine can become more human and thereby making human to communicate with the machine easily. NLP seeks to make software intelligent enough to process a natural language as humans. For example, imagine a machine that takes instructions by voice.

NLP analysis generally consists of the following three levels [3]:

- *Syntax*, the study of sentence structure. Syntax deals with the formation of a sentence from individual words. Syntax alone suggests the proper interpretation of "Jimmy loves Lucy."
- *Semantics*, the study of context-independent meaning. This derives the meaning of a sentence based on the meanings of the words/phrases. For example, semantics determines whether the word "bank" refers to a river bank or to a financial institution.
- *Pragmatics*, the study of context-dependent meaning. Pragmatics deals with how meaning changes in the presence of a specific context and how the contexts affect the meaning of the sentences. This level is concerned with the purposeful use of language in situations.

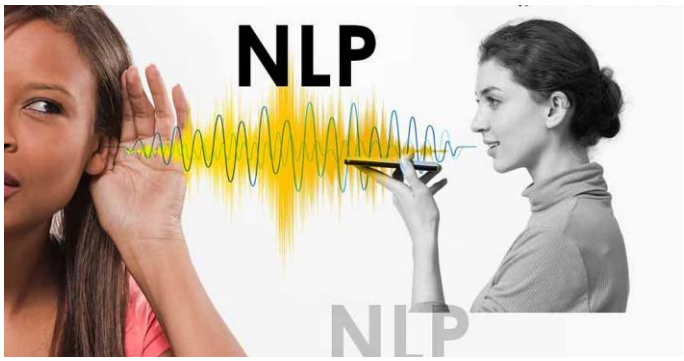

Figure 1: A representation of NLP [4].

As a foundational pillar of modern artificial intelligence, NLP encompasses a wide array of tasks, including speech recognition, text classification, natural language understanding (NLU), and natural language generation (NLG). NLP encompasses a wide range of tasks, such as information

retrieval (IR), named entity recognition (NER), relation extraction, text classification, topic modeling, semantic textual similarity, machine translation, and question answering (QA). Figure 1 shows how NLP transforms raw acoustic data into meaningful interactions [4], while Figure 2 shows different components of NLP [5].

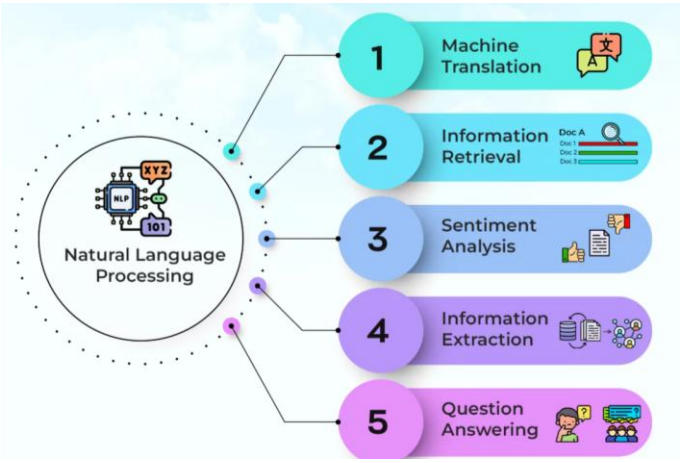

Figure 2: Different components of NLP [5].

Recently, large language models (LLMs) have shown their ability in learning universal language representations, text understanding and generation. LLMs refer to a model with a large number of parameters, vast training data, and substantial compute, enabling it to capture complex language patterns. In LLM-based NLP, pre-processing is followed by prompt engineering, which guides LLMs to produce outputs that align with extraction requirements during inference without altering the model's parameters. Models like GPT are pushing the boundaries of language understanding, enabling nuanced and context-aware applications. The GPT (Generative Pretrained Transformer) is a large-scale language model developed by OpenAI that consists of multiple layers of transformer blocks, each with a self-attention mechanism and a forward neural network [6]. GPT-based systems can summarize complex reports or generate creative content like essays, making them versatile in both academic and professional environments. ChatGPT uses NLP techniques to understand prompts. When you enter a prompt, the chatbot comprehends it and provides relevant replies. Figure 3 shows how NLP works [7].

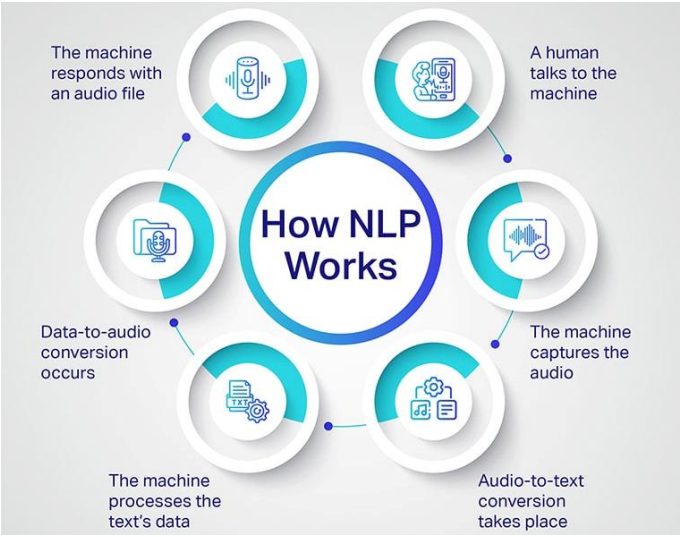

Figure 3: How NLP works [7].

III. WHAT IS SPEECH RECOGNITION?

Speech recognition, also known as automatic speech recognition (ASR) or speech-to-text (STT), is the computational process of converting an acoustic signal, captured by a microphone, into a sequence of words or text that a computer can process. It is a technology that enables a computer to identify and interpret words and phrases in spoken language and convert them into texts by computers. It is a subfield of computational linguistics that deals with technologies to allow spoken input into systems. Modern speech recognition is not a single process but a sophisticated pipeline that converts acoustic waves into meaningful text. The quest to teach computers to listen began decades ago, marked by significant milestones that transitioned from simple pattern matching to complex neural networks. Modern ASR systems process audio in several stages. The process begins by converting raw waveform into acoustic features that capture relevant sound characteristics. This way the captured analog audio is converted into a digital signal. This signal is then broken down into small "frames" (usually 10-25 milliseconds long) [1]. Figure 4 shows a voice or speech recognition system [8], while Figure 5 shows how the speech is produced by the vocal cords [9].

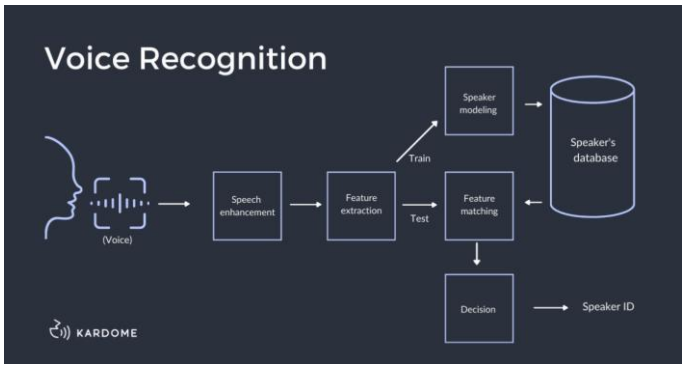

Figure 4: A voice or speech recognition system [8].

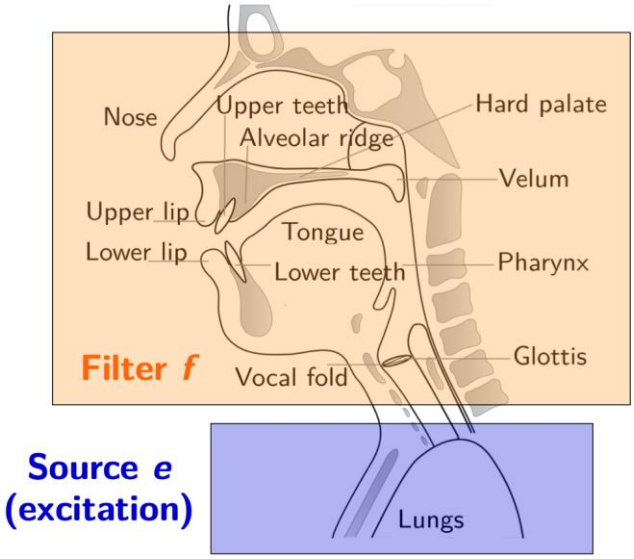

Figure 5: How the speech is produced by the vocal cords [9].

The roots of speech recognition trace back to the 1950s. In 1952, Bell Laboratories developed "Audrey," a system that could recognize spoken digits (0-9) from a single speaker with reasonable accuracy. The 1970s and 1980s saw significant government investment, particularly from DARPA in the U.S., which funded projects aiming for more natural, continuous speech recognition. The deep learning revolution transformed the field starting in the late 2000s and accelerating in the 2010s [10].

IV. NLP IN SPEECH RECOGNITION

The evolution of human-computer interaction has led to the quest to make machines understand our most natural form of communication: spoken language. In the rapidly evolving landscape of artificial intelligence, the ability of machines to understand and process human language has become a cornerstone of innovation. At the heart of this capability lies the powerful synergy between automatic speech recognition (ASR) and natural language processing (NLP). The two fields play distinct yet interconnected roles in transforming spoken words into actionable insights. While ASR is responsible for the mechanical conversion of acoustic signals into digital text, NLP provides the cognitive framework necessary to interpret that text, resolve ambiguities, and derive meaning. Historically, these systems operated in silos, with NLP serving as a mere post-processing layer. Figure 6 shows that an ASR-powered application involves the Speech-to-text, Natural Language Processing and Text-to-speech [9], while Figure 7 compares voice recognition and NLP [11].

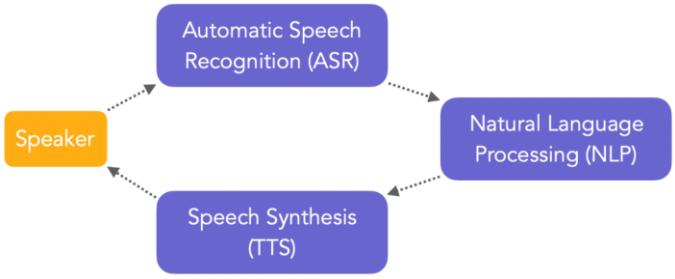

Figure 6: An ASR-powered application [9].

Speech Recognition
VS
Natural Language Processing
Comparison Chart

Speech Recognition	Natural Language Processing
It is a technology that enables a computer to identify and interpret words and phrases in spoken language and convert them into texts by computers.	It is a branch of artificial intelligence that investigates the use of computers to process or to understand human languages for the purpose of performing useful tasks
It is a subfield of computational linguistics that deals with technologies to allow spoken input into systems.	It is a technology that devel-ops methodologies and algorithms that take as input or produce as output unstructured, natural lan-guage data.
Speech recognition tools are used in different types of dictation tasks, such as composing a text message, playing music through a home-connected device, or text-to-speech applications with virtual assistants.	NLP is used to perform tasks such as automatic summarization, topic segmentation, relationship extraction, information retrieval, and speech recognition.
Popular speech recognition software include Windows Speech Recognition, Google Assistant, and Dragon.	Alexa, Siri, and Cortana are some of the best examples of natural language processing.

Figure 7: Comparing voice recognition and NLP [11].

In an era where technology is reshaping the way we communicate, natural language processing (NLP) in speech has emerged as a transformative force. From virtual assistants like Siri and Alexa to real-time language translation tools, NLP in speech is revolutionizing industries and enhancing human-machine interactions. Natural language processing (NLP) in speech refers to the intersection of linguistics, computer science, and artificial intelligence, enabling machines to understand, interpret, and respond to spoken language. At its core, NLP in speech relies on advanced algorithms, machine learning models, and linguistic databases to process and analyze human speech [12].

V. EXAMPLES OF SPEECH RECOGNITION SYSTEMS

Existing system which uses speech recognition include the following [13]:

- *Apple (Siri):* Siri is a virtual assistant and a part of Apple Inc. It is designed to offer you a multiple way of interaction with your phone by speak up.
- *Google Assistant:* Google Assistant is virtual assistant of google Inc's. Google assistant control your devices and smart phone. Some important features of Google assistant are: it controls your device and your smartphone and accesses information from calendar.
- *Microsoft Cortana:* Like Siri and Google Assistant, Cortana is also a voice assistant developed and created by Microsoft. Basically it is designed for window devices. Nowadays, it is available in various devices. It can perform a multiple task for users, like remainder setting, as well as it can also scheduling the calendar events.
- *Amazon Alexa:* Alexa is a virtual assistant technology designed and created by Amazon. This technology is based upon machine learning and NLP(natural language processing. Alexa can perform various task such as it can acknowledge the user about weather.
- *Samsung Bixby:* Bixby is a virtual assistant developed by Samsung Electronics. With the help of this system you can send message from one device to another, as well as you can check the cricket and any other game score. Bixby also supports some fancy features.

VI. APPLICATIONS OF NLP SPEECH RECOGNITION

Speech recognition has become integral to daily life and many industries. The integration of NLP and ASR has birthed a wide array of applications that have become ubiquitous in daily life. Common applications include the following [1,10,14]:

- *Voice Assistants:* Perhaps the most visible applications are virtual assistants like Siri, Alexa, and Google Assistant. Siri, Alexa, Google Assistant, and others handle queries, control smart homes, and provide hands-free interaction. These systems do not just transcribe speech; they perform intent recognition.
- *Healthcare:* In medical settings, NLP transforms dictated notes into structured electronic health records (EHRs), reducing administrative burden and improving data accuracy. This allows healthcare providers to focus more on patient care. Dictation systems allow doctors to document patient encounters efficiently while reducing administrative burden.
- *Education:* Educational institutions have already started using speech recognition and NLP technologies in several different settings. NLP technology is increasingly accepted in academic

institutions. These technologies have also been usefully embedded into computer-assisted language learning (CALL) systems, which allow learners to analyze their pronunciation, fluency, or practice conversations with a virtual agent. It is expected that adding ideas from corpus linguistics to NLP-based language learning tools will make them more useful for education.

- *Language Learning*: NLP-powered speech recognition tools offer real-time feedback on pronunciation and grammar, providing an interactive and personalized learning experience for language students. Translation-integrated systems break down language barriers.
- *Customer Service:* Voice-enabled chatbots and interactive voice response systems use NLP to understand customer queries, route calls efficiently, and provide automated support, enhancing customer experience. Automated call centers use ASR for routing, sentiment analysis, and virtual agents.

VII. BENEFITS

With speech recognition, computers can understand and interpret spoken words of phrases and convert them into text. NLP in speech minimizes errors in transcription, translation, and data entry, reducing the costs of rectifying mistakes. Industries like healthcare, education, customer service, and entertainment benefit significantly from NLP in speech. Other benefits include the following [1,12]:

- *Automation:* Automated customer service systems lower the costs associated with hiring and training support staff. By automating tasks that traditionally required human intervention, NLP in speech reduces response times, minimizes errors, and enhances productivity.
- *Cost Savings:* Implementing NLP in speech can lead to significant cost savings for businesses. Speech-based NLP solutions can handle large volumes of interactions without additional costs, making them ideal for growing businesses.
- *Accessibility*: For individuals with disabilities, these technologies provide crucial accessibility features, enabling hands-free control of devices and facilitating communication. Real-time captioning and voice control empower people with hearing or motor impairments.
- *Enhanced Accuracy:* One of the most significant benefits of NLP in speech recognition is the substantial improvement in accuracy. ASR systems can introduce errors, especially with homophones or ambiguous phrasing. NLP models are crucial in mitigating these errors by leveraging contextual understanding. By analyzing surrounding words and the overall discourse, NLP can correct misheard words or phrases that are phonetically similar but semantically incorrect. This semantic error correction is vital for reliable speech-to-text applications.
- *Intent Recognition:* Even when an ASR system achieves a perfect literal transcription, it may still fail the user if it misses the underlying intent. Beyond mere transcription, NLP empowers speech recognition systems with the ability to understand the user's intent and the broader context of their communication. This moves the interaction from a simple conversion of speech to text to a more profound comprehension of what the user wants to achieve. NLP techniques such as intent classification, entity extraction, and sentiment analysis allow systems to discern the purpose behind spoken words, identify key information, and even gauge the user's emotional state. By analyzing not just the words but the linguistic patterns, intent, and tone, systems can detect a speaker's mood—frustration, satisfaction, or urgency. This is particularly valuable in call centers, where automated systems can escalate a call to a human supervisor if they detect a customer's rising anger.
- *Multilingual Systems:* Multilingual systems have demonstrated a dominance over the monolingual systems with the help of shared learning of model components in several languages, particularly for those languages with limited data. By supporting multiple languages with just a single speech model instead of various distinct models, multilingual systems considerably simplify infrastructure.
- *Personalization:* Future systems will likely focus on "personal language models" that adapt to an individual's unique speech patterns and preferences, leading to a truly personalized and intuitive human-machine partnership.
- *Emotional Intent:* Modern systems are beginning to look beyond what is said to how it is said. By analyzing pitch, pace, and tone, AI can now detect a speaker's emotional state or underlying intent, enabling more empathetic and effective customer service bots.

VIII. CHALLENGES

In spite of the remarkable progress, the integration of NLP in speech recognition still faces notable challenges. The challenges of NLP in speech recognition are deeply rooted in the inherent variability of human language, the nuances of social context, and the systemic biases present in data collection. A primary challenge in speech recognition is not just hearing the words, but selecting the correct words from a sea of acoustic possibilities. Another major challenge is the integration of temporal and prosodic information into modern learning architectures. The challenges of NLP in speech recognition represent a cross-section of linguistics, computer science, and sociology. Other challenges include the following [1,15]:

- *Privacy Concerns:* Collecting and processing speech data raises concerns about user privacy and data security. Privacy concerns arise because processing often involves cloud servers, raising questions about data storage and surveillance. For reasons of privacy and latency, there is a massive shift toward on-device ASR. Instead of sending your voice data to a cloud server, modern smartphones now use dedicated AI chips to process speech locally, ensuring your conversations remain private. Organizations must implement robust data security measures and obtain user consent.
- *Ethical Concerns:* As these technologies become more pervasive, there is a growing emphasis on developing ethical AI practices, including bias reduction in speech recognition systems to ensure fairness and inclusivity.
- *Hallucinations:* Hallucinations (fabricating words) can occur in some models. Latency is critical for real-

time applications, requiring trade-offs between accuracy and speed. Robustness to code-switching (mixing languages) and far-field audio (distant microphones) also needs improvement.

- *Error Propagation*: The propagation of errors remains a primary concern. Error propagation from ASR to NLP is a persistent issue, where initial transcription errors can lead to incorrect interpretations. A misheard word by the ASR system can lead to a completely incorrect interpretation by the NLP component, resulting in unintended actions or misunderstandings.
- *Ambiguity:* Human language is inherently ambiguous, with homophones and context-dependent meanings posing a constant challenge. The diversity of human speech is perhaps the most visible challenge in ASR. NLP models must be robust enough to resolve these ambiguities accurately. In spoken language, homophones—words that sound identical but possess different meanings and spellings—present a significant obstacle. For example, an ASR system must distinguish between "there," "their," and "they're" based solely on the surrounding text. NLP is instrumental in error correction and disambiguation.
- *Domain Specificity:* Achieving high accuracy in specialized domains (e.g., medical, legal, financial) often requires extensive domain-specific training data, which can be costly and time-consuming to acquire.
- *Digital Divide*: The "digital divide" in speech technology is most apparent when looking at low-resource languages. The vast majority of ASR and NLP research is concentrated on a handful of high-resource languages, such as English, Mandarin, and Spanish. For thousands of other languages—including many African, indigenous, and regional tongues—there is a severe lack of annotated audio data and text corpora. With over 7100 languages in the world, one of the most pressing concerns for the speech and language community is to swiftly design and implement speech processing systems in unsupported languages at an acceptable cost and to bridge the gap between linguistic and technological expertise.
- *Interpretability:* Another important challenges concerns model interpretability and trustworthiness. As SR and NLP systems increasingly influence decision-making in sensitive domains, such as healthcare, education, and human–computer interaction, the ability to explain model behavior becomes essential. The contrast between high empirical accuracy and limited explanatory insight suggests that interpretability should be treated as a core research objective rather than an afterthought.

IX. FUTURE OF NLP SPEECH RECOGNITION

Speech recognition is moving beyond simple transcription toward true "speech understanding." The future of ASR is not just about audio. End-to-end systems that not only transcribe but comprehend and act on speech will become standard. By 2030, speech interfaces may feel as natural as talking to another person, driving broader adoption in augmented reality, robotics, and ambient computing. The technology will continue democratizing access to information and tools while raising important ethical questions about consent, bias, and employment impacts.

The integration of NLP and speech recognition continues to evolve at a rapid pace, with several key trends shaping the future. Future systems will increasingly integrate speech with other modalities, such as visual cues, to enhance understanding and context. This could lead to more intuitive and human-like interactions with AI. Speech recognition is moving toward seamless integration with multimodal AI. Multimodal AI combines speech recognition with computer vision (lip reading) and contextual text data. This is particularly useful in noisy environments where audio alone might be insufficient for high accuracy. Figure 8 shows multimodal data collection [14]. As deep learning continues to mature and multimodal models become the standard, the boundary between human and machine communication will continue to blur [1].

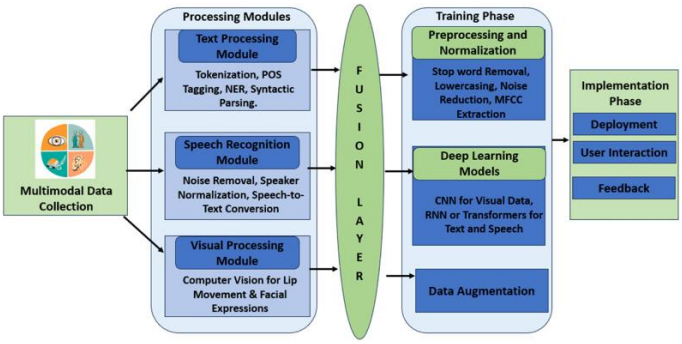

Figure 8: Multimodal data collection [14].

CONCLUSION

Speech recognition has evolved from a laboratory curiosity into an essential tool for accessibility, productivity, and global communication. It exemplifies the rapid progress of artificial intelligence. The integration of natural language processing (NLP) into automatic speech recognition (ASR) has transformed how humans interact with machines, moving from rigid command-based systems to fluid, conversational interfaces. At its core, NLP aims to bridge the gap between human communication and computer understanding, making it possible for machines to interpret, analyze, and generate human language in a way that feels natural and intuitive. NLP combines the strengths of computational linguistics, machine learning, and deep learning to process and make sense of vast amounts of language data.

The symbiotic relationship between NLP and speech recognition is a driving force behind the next generation of intelligent systems, promising a future where human-computer interaction is more natural, intuitive, and effective. NLP's ability to interpret meaning, understand context, and correct errors significantly elevates the utility and accuracy of speech recognition systems. As these technologies continue to evolve, their combined benefits will lead to more seamless, intuitive, and powerful voice-enabled applications, further blurring the lines between human and artificial intelligence. More information about the integration of NLP in speech recognition can be found in [16-22] and the following related journals:

- *Natural Language Processing Journal*
- *Journal of Emerging Technologies and Innovative Research*
- *International Journal of Future Generation Communication and Networking*

References

[1] https://manus.im
[2] H. Ahlawat, N. Aggarwal, and D. Gupta, "Automatic speech recognition: A survey of deep learning techniques

and approaches," *International Journal of Cognitive Computing in Engineering*, vol. 6, December 2025, pp. 201-237.

[3] J. Hirschberg, B. W. Ballard, and D. Hindle, "Natural language processing," *AT&T Technical Journal*, Jan./Feb. 1988, vol. 67, no. 1, 1988.

[4] A. Jain, "What is the role of NLP in voice assistants?" August 2024, https://www.analyticsinsight.net/nlp/what-is-the-role-of-nlp-in-voice-assistants#google_vignette

[5] "How Google uses NLP to improve SERPs, featured Snippets & UX," https://digitalguider.com/blog/what-is-google-nlp/

[6] X. Jiang et al., "Applications of natural language processing and large language models in materials discovery," *NPJ Computational Materials*, vol. 11, no.79, 2025.

[7] "What is NLP? How it works, benefits, challenges, examples," June 2025, https://www.shaip.com/blog/what-is-nlp-how-it-works-benefits-challenges-examples/

[8] A. P. Sudhakar, "How does natural language understanding work?" January 2026, https://botpenguin.com/blogs/how-does-natural-language-understanding-work

[9] M. Fabien, "Introduction to automatic speech recognition (ASR)," https://maelfabien.github.io/machinelearning/speech_reco/#

[10] https://grok.com

[11] "Difference between speech recognition and natural language processing," https://www.differencebetween.net/technology/difference-between-speech-recognition-and-natural-language-processing/

[12] "Natural language processing in speech," February 2026, https://www.meegle.com/en_us/topics/speech-recognition/natural-language-processing-in-speech

[13] R. Pahwa, H. Tanwar, and S. Sharma, "Speech recognition system: A review," *International Journal of Future Generation Communication and Networking*, vol. 13, no. 3, 2020, pp. 2547–2559.

[14] P. Dubey et al., "Bridging language gaps: The role of NLP and speech recognition in oral English instruction," *MethodsX*, May 2025.

[15] A. Abdi and F. Meziane, "Advances and challenges in speech recognition and natural language processing," *Applied Sciences*, vol. 16, no. 2, January 2026.

[16] S. Vajjala et al., *Practical Natural Language Processing.* O'Reilly Media, 2020.

[17] H. Lane and M. Dyshel, *Natural Language Processing in Action*. Manning Publications, 2nd edition, 2025.

[18] D. Jurafsky and J. Martin, *Speech and Language Processing.* Prentice Hall, 2nd edition, 2008.

[19] J. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Pearson India, 2014.

[20] D. Jurafsky, *Speech & Language Processing*. Pearson Education, 2000.

[21] U. Kamath, J. Liu, and J. Whitaker, *Deep Learning for NLP and Speech Recognition.* Springer, 2019.

[22] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Youcanprint, 2023.

ABOUT THE AUTHORS

Matthew N.O. Sadiku is a professor emeritus in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics, computer networks, engineering education, and marriage counseling. He is a Life Fellow of IEEE and a fellow of NAE.

Janet O. Sadiku holds bachelor degree in Nursing Science in 1980 at the University of Ife, now known as Obafemi Awolowo University, Nigeria and doctoral degree from Juliana King University, Houston, TX in December 2023. She has worked as a nurse, educator, and church minister in Nigeria, United Kingdom, Canada, and United States. She is a co-author of some papers and books.