Theory and Practice of Science and Technology
2025, VOL. 6, NO. 7, 5-8
DOI: 10.47297/taposatWSP2633-456901.20250607

The Application and Innovation of Speech Recognition Technology in Intelligent Customer Service

Xiangyu Zhao

School of Communication Technology, Communication University of China, Nanjing, Jiangsu, 211172, China

ABSTRACT
Speech recognition technology, commonly referred to as automatic speech recognition (ASR), converts the vocabulary information contained in human speech into structured input that computers, such as text commands, can process. Intelligent customer service is an automated response system built using natural language processing, deep learning, and knowledge processing technologies, capable of operating 24/7 and handling multiple user requests simultaneously. Integrating ASR capabilities into such systems enables users to interact with the system via natural speech rather than traditional keypad input, thereby enhancing service accessibility and automation levels. This paper explores the application of speech recognition technology in intelligent customer service systems, aiming to provide a clear practical framework for developing more efficient and user-friendly intelligent service systems.

KEYWORDS
Speech recognition; Intelligent customer service; Technology application

1 Introduction

Automatic speech recognition (ASR) technology aims to solve the problem of machines understanding the meaning of words in human speech, with the output usually being standardized text or commands. Intelligent customer service, as an automated service solution, relies heavily on the system's ability to capture and respond to user intent accurately. In service scenarios, users tend to prefer the convenience of spoken communication over cumbersome text input, naturally leading to the need for ASR as an entry point. Integrating ASR into intelligent customer service can bring it closer to the convenience of human-to-human conversation. Therefore, this paper will analyze how speech recognition technology can be effectively integrated into the process and functional design of intelligent customer service to maximize the value of this technology in the field of innovative services.

2 Basic principles of speech recognition technology

2.1 Basic concepts of speech recognition

The goal of speech recognition technology is to accurately capture the lexical content of human speech and convert it into a structured data format that can be directly read and written by computer systems, such as text commands, operation codes, or character sequences [1]. The complete speech recognition process consists of two fundamental stages: feature extraction and pattern matching. In the feature extraction stage, the system first analyzes and processes the raw acoustic signal, extracting parametric representations that reflect the essential attributes of speech production (such as frequency energy distribution and phonetic duration characteristics), and converts them into digital feature vectors for subsequent calculations. The pattern matching stage is responsible for comparing and calibrating these feature vectors with predefined speech models, using pattern recognition algorithms to determine the corresponding word sequences and textual expressions. The complexity of this process lies in the fact that natural speech exists in a continuous stream, with unclear word boundaries and phenomena such as intonation changes and connected speech. Technologically, it is necessary to balance the distinguishability of speech details with the overall intelligibility of the language. The acoustic layer must accurately capture the physical characteristics of speech. In contrast, the semantic layer must restore the logical continuity across words and sentences to avoid intent deviations caused by fragmented outputs. Both layers collectively support the completeness and usability of the recognition results.

2.2 Core technologies of speech recognition

Speech recognition technology in modern times primarily relies on deep learning methods to achieve end-to-end automated modeling. Unlike traditional approaches that depend on manually designed feature engineering, modern systems train neural networks to extract feature representations from raw speech data autonomously. The core components of this technology encompass two functional modules: the acoustic model and the language model, which collaborate sequentially to convert signals into semantic meaning. The acoustic model processes digitized speech input using structures such as recurrent neural networks (RNNs) or their improved variants, long short-term memory networks (LSTMs), to identify basic phonetic units from acoustic waveforms. Through training on large-scale speech samples, this model gradually establishes a reliable mapping relationship between acoustic features and phoneme sequences. The language model assumes the role of semantic logic constraints, analyzing the preliminary phoneme sequence results generated by the acoustic model. Based on linguistic structural patterns, such as grammatical rules and collocation habits, it optimizes the text output to align more closely with real-world usage [2]. Statistical probability-based language models utilize massive text data to construct vocabulary distribution patterns, reducing sensitivity to specific noise or accent interference. For example, in banking systems, they enhance the recognition weight of professional expressions such as "annualized interest rate" and "funds transfer," forming optimized solutions tailored to business scenarios.

3 Application of speech recognition technology in intelligent customer service

3.1 Basic functions and components of intelligent customer service

An intelligent customer service system is a solution that uses automation technology to simulate human service interaction capabilities. Its essence lies in converting unstructured user requests into executable operational instructions. The system consists of components such as a semantic understanding unit, a knowledge management hub, and a dialogue control engine, forming a continuous processing chain. The semantic understanding unit is responsible for parsing the true intent of user input, differing from traditional keyword matching mechanisms by identifying implicit needs within composite requests. For example, the phrase "Why was last month's deduction higher? " automatically associates with dual tasks: bill inquiry and fee explanation. The knowledge management hub continuously integrates the enterprise's business rule repository and common question knowledge graph, supporting real-time dynamic updates for changes in customer policies. The dialogue control engine manages interaction state logic, guiding users to supplement necessary parameters in multi-round dialogues. For example, when applying for a transfer, users must sequentially verify their identity, amount, and recipient. The basic service functions cover three scenarios: query response, transaction processing, and complaint handling. The technical challenges lie in the accuracy of intent recognition and maintaining consistency in context state during multi-step operations. This requires enhancing fault tolerance through intensive dialogue log training [3].

3.2 Core application areas of speech recognition in intelligent customer service

The application of speech recognition technology in intelligent customer service focuses on voice navigation control and voice data entry. In the field of voice navigation, the system maps users' spoken commands to operational commands through real-time speech transcription, replacing traditional keypad-based interactive menus. For example, in banking services, users can say "Check the USD exchange rate" to directly trigger the foreign exchange interface call, without going through the hierarchical menu selection process of "Press 3 → Press 1 → Press 5." This mode improves navigation efficiency while supporting natural language expressions, such as "Transfer to the last customer service representative I spoke with," which automatically links to historical conversation records. In the field of voice data entry, customer service personnel can quickly fill in business order fields by speaking. For example, in a repair center, an engineer reports, "The outdoor unit of the KFR-35GW variable frequency air conditioner is not starting." Voice recognition automatically extracts the product model (KFR-35GW) and fault phenomenon (outdoor unit not starting) and writes them into the repair system, avoiding model coding errors caused by manual input. The challenge in implementing this technology lies in improving the accuracy of specialized terminology recognition. For example, the term "on-load tap changer" in the power industry requires a dedicated acoustic model optimization, combined with industry-specific vocabulary enhancement and context-based semantic verification mechanisms to ensure usability.

3.3 The role of speech recognition technology in enhancing the efficiency of intelligent customer service

The role of speech recognition technology in enhancing the efficiency of intelligent customer service is reflected in three aspects: service response speed, human resource cost savings, and business process resilience. In terms of response speed, voice interaction enables direct communication between user requests and system responses, eliminating the time-consuming process of navigating through traditional menu options. When users provide complete operational instructions, such as "Check the status of my package using ID number 123456," the system simultaneously completes identity verification. It triggers the command, completing the service process in a single interaction. Labor cost optimization is achieved through the complete automation of basic consultation services, with routine inquiry requests no longer requiring human intervention. Call centers can thus reduce the number of entry-level customer service positions and instead have expert agents handle complex complaint scenarios. Process resilience is enhanced by the voice system's ability to tolerate non-standard expressions. When a user verbally enters an order number such as " 2024ABC," and there is confusion between letters, the system uses acoustic models and historical order patterns to generate the most likely correct encoding, such as correcting '202423BC' to "2024A2BC." The sustainability of efficiency optimization depends on improvements in acoustic environment adaptability. For example, in vehicle-mounted scenarios, strong noise suppression algorithms are required to ensure the simultaneous achievement of technical stability and business reliability [4].

4 Innovations in speech recognition technology for intelligent customer service

4.1 Innovations in multilingual and dialect recognition technology

Current advances in multilingual and dialect recognition are reflected in breakthroughs in comprehensive service capabilities. Technological innovations revolve around adaptive processing capabilities within a unified framework. A single interaction system shares underlying acoustic models to uniformly extract standard features across different languages, while the upper layer automatically matches language decoding rules based on input. This structure avoids the resource consumption issues associated with traditional parallel systems, reducing response latency. Dialect processing establishes a structured mapping library between dialect pronunciation and standard language, such as the Sichuan dialect term "yao de," precisely corresponding to the command meaning of "can." Probabilistic mapping models are used to complete semantic conversion. The breakthrough lies in the application of low-resource training schemes, where some systems require only hundreds of hours of dialect speech samples to build foundational recognition capabilities. This directly drives the upgrading of service capabilities in international customer service centers, addressing the recognition of English spoken with native accents by cross-border users (e.g., French-speaking users pronouncing "this" as "zis"), breaking through the geographical limitations of traditional voice interaction. Technical deepening requires attention to the interference of multilingual grammatical differences on intent judgment, such as subject-verb misjudgment caused by Arabic sentence inversion, necessitating the establishment of a language-neutral structural conversion mechanism at the semantic level [5].

4.2 Personalized development of emotion recognition and voice interaction

Emotion recognition is driving the evolution of intelligent customer service toward an emotion-responsive service model. The system combines acoustic parameter analysis and semantic feature extraction to establish an emotion judgment framework. Physical characteristics of the voice are analyzed to identify basic emotional tendencies, such as a low, slow tone indicating sadness, while keywords in the conversation content reinforce the confidence level of emotion recognition. Emotional state serves as a new decision-making dimension influencing service path selection. When users are in an angry state, the system automatically upgrades service priority and provides compensation plans, replacing standard response workflows. Personalized services establish a continuous binding between user voiceprint features and historical interaction data. For example, in insurance service cases, the system automatically enables amplification mode when identifying voice features of elderly users and extends the pause time for voice announcements. Technical implementation relies on a continuous self-evolution mechanism, where each round of dialogue automatically updates the user's emotional response record database. For example, if a user frequently rejects promotional recommendations, the weight of marketing strategies is reduced, gradually forming an individualized interaction plan. To ensure service ethical compliance, emotional label application boundaries must be set, prohibiting the reduction of service levels due to frequent anxiety, and such actions must be constrained through manual review permissions.

4.3 The convergence of artificial intelligence and speech recognition

The deep integration of artificial intelligence and speech recognition is giving rise to a cognitive intelligence-based interaction system. The main areas of development encompass three key capabilities. First, the enhancement of long-term contextual modeling capabilities enables the system to maintain the overall state of the conversation through a mechanism that remembers continuous dialogue data. In express delivery service scenarios, after a user initially inquires about delivery timelines, they may subsequently provide updated address information. The system automatically traces historical parameters to calculate a new delivery plan. Second, the integration of speech reception and command execution. A new architecture enables direct conversion from acoustic signals to business operations without intermediaries. Users can trigger work order creation directly by speaking an activation command. Third, an active service system based on behavior prediction. Relevant business modules are preloaded before users initiate inquiries. For example, when reporting a power outage, the system immediately pushes the location of the nearest emergency repair station. The technological innovation lies in restructuring the traditional voice processing chain to form an end-to-end closed-loop system from voice input to business output. In large-scale applications, the issue of decision transparency must be addressed to prevent critical operations from becoming untraceable system behaviors. For example, unauthorized changes to subscription packages must be subject to mandatory manual verification to ensure business security.

5 Conclusion

This article analyzes the progress of speech recognition technology in the field of intelligent customer service, including breakthroughs in multilingual and dialect recognition, improvements in service experience through emotional interaction, and efficiency transformations brought about by intelligent deep integration. These advancements indicate that speech interaction is transforming the operational models and execution requirements of traditional customer service. The industry should prioritize enhancing the ability to analyze dialects and non-standard pronunciations, establish usage boundaries for emotion recognition technology to protect user rights, and ensure the secure integration of recognition results with application functions. In practical implementation, priority should be given to foundational business scenarios such as account inquiries and logistics tracking to establish a reliable technical implementation framework gradually. By continuously refining human-machine collaborative workflows, enhancing decision-making reliability in complex service scenarios, and integrating intelligent voice services into a cohesive system, the goal is to elevate overall service standards and create a more seamless and efficient customer service operation system.

About the Author

Xiangyu Zhao, master's degree, lecturer, research interests: computer science and technology.

References

[1] Yang Wei, Wang Maohuan. Application of Deep Learning-Based Speech Recognition Technology in the Digitization of Audio Archives [J]. Zhejiang Archives, 2024, (2): 52-54+60.

[2] Wang Feifei, Ben Kerong, Zhang Xian. Research on Speech Recognition Robustness Enhancement Technology Based on Domain Knowledge [J]. Computer Engineering and Science, 2023, 45(12): 2155-2164.

[3] Zhong Lijun, Zhou Chuny. The Influence of Intelligent Customer Service Types on Consumer Recommendation Persuasiveness: A Discussion on the Mediating Role of Human-Computer Trust and the Moderating Role of Anthropomorphism [J]. Business Economics Research, 2024, (24): 71-74.

[4] Liu Ying. The Impact of Artificial Intelligence Customer Service on Consumers' Online Purchase Intentions: A Mediating Perspective Based on Consumer Experience [J]. Business Economics Research, 2024, (13): 73-76.

[5] Peng Guochao, Wu Siyuan. A Study on the Mechanism of User Avoidance Behavior in Intelligent Customer Service Contexts: An Exploration Based on Grounded Theory and Narrative Research [J]. Information Theory and Practice, 2024, 47(3): 104-111.