Research on Artificial Intelligence-based Speech Recognition System and Its Applications

Zhiyuan Xu

Guangzhou Yunqu Information Technology Co., Ltd., Guangzhou 510630, Guangdong

Abstract: ***Artificial intelligence has played an important role in the field of speech recognition. This paper designs an efficient speech recognition system, which improves the speech recognition efficiency through data preprocessing, feature extraction and speech signal processing. The system hardware includes a host, a processor, a signal receiver and chip design; the system software design includes core word processing, audio program design and autonomous recognition. Experimental results show that the artificial intelligence speech recognition system improves the accuracy of speech recognition, meeting the actual application standards.***

Keywords: Artificial Intelligence; Audio Processing; Hardware Design.

1. INTRODUCTION

At the current stage, information dissemination mainly relies on channels such as the Internet and radio, which not only makes people's lives more convenient, but also poses higher requirements for information security. On this basis, a new voice mode has emerged, which can convert voice into text information in real time and efficiently through voice technology. The speech recognition system uses artificial intelligence technology to simulate computer operations to achieve the purpose of extracting and processing main data, while ensuring the accuracy of the recognized speech. Therefore, it is of great practical significance to design a speech recognition system using artificial intelligence technology.

2. OVERVIEW OF RELATED TECHNOLOGIES

2.1 Data Preprocessing and Feature Extraction

In a speech recognition system, data preprocessing and feature extraction are crucial parts, which are vital for the performance of the model. Data preprocessing cleans and standardizes speech data to improve the performance of subsequent models, mainly including noise reduction optimization, duration normalization, and speech segmentation. Feature extraction analyzes the preprocessed data through scientific methods and obtains a series of key features for characterizing the speech characteristics.

2.2 Speech Recognition and Synthesis Technologies

Speech recognition and synthesis technologies enable computers to convert human language into text or a machine - processable form. Specifically, they include voice assistants, caption display, and voice control. Speech synthesis is the conversion of text into natural speech to achieve operations such as reading aloud and voice answering. These two technologies have obvious advantages in dealing with multiple languages, accents, etc., while improving the interactive experience effect.

2.3 Features of Artificial Intelligence Speech Technology

The proposal of artificial intelligence aims to simulate the human intelligent thinking process. The artificial intelligence speech recognition technology combines artificial intelligence and speech technology, becoming an important research direction in the intelligent discipline. It can effectively acquire and understand human natural language, providing users with an intelligent interactive experience. The actual advantages include: 1) The language interaction is efficient, convenient, and very natural; 2) It can establish a high-standard service system; 3) After generating a strong scale effect, it reduces labor costs; 4) It fully applies the advantages of big data.

2.4 Relative Works

Peng (2022) explored multi-source and source-private cross-domain learning for visual recognition [1]; Peng et al. (2024) proposed a dual-augmentor framework for domain generalization in 3D human pose estimation [2]; Mao (2026) introduced a clustering-based approach to image compression using the K-means algorithm [3]; Dong and Li (2025) developed a face generation model based on DCGAN [4]; Jie (2024) applied AI technology to industrial defect detection [5]; Zhang Q. (2025) improved the steel surface defect detection algorithm of the YOLOv5 network [6]; Zhang J. et al. (2025) established a flow field evaluation system for polymer injection in thick reservoirs [7]; Wu et al. (2025) addressed small-sample object detection of surface cracks in concrete structures via multi-level transfer learning [8]; Zheng and Jiang (2025) proposed a new methodology for Chinese term extraction from scientific publications [9]; Zhang Y. and Shi (2026) optimized the thermal insulation performance of transport packaging boxes based on response surface algorithm [10]; Zhu et al. (2026) applied proximal policy optimization for motion control of flexible-joint robotic arms in warehouse sorting [11]; Zhang X. (2026) proposed a hybrid LSTM-GARCH model integrating US financial market volatility factors [12]; Shen et al. (2026) built a survival risk prediction model for colorectal cancer patients using graph convolutional networks and TCGA multi-omics data [13]; Shen et al. (2025) applied the whale optimization algorithm to financial payment fraud detection [14]; Shi et al. (2026) focused on multi-stage domain-grounded synthetic data generation for fine-tuning large language models in telecommunications [15]; Gao (2026) constructed an intelligent quality and operations control system for industrial electrical enterprises driven by ISO 14001 and ISO 9001 [16]; Zheng et al. (2023) improved the YOLOv5s algorithm for rebar cross-section detection [17]; and Zhao et al. (2023) proposed a smart warehouse track identification method based on Res2Net-YOLACT+HSV [18].

3. DESIGN AND IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE SPEECH RECOGNITION SYSTEM

3.1 Hardware Design

The hardware design can improve the quality of speech acquisition and suppress noise, making the recognition more accurate and enabling the speech system to respond faster.

3.1.1 Host Design

The VS78 host is the core structure of the automatic speech recognition system. It adopts an integrated box design and incorporates key components such as the main board. The VS78 host is a virtual host, and its greatest advantage is saving storage space and improving the sensitivity of audio signals. It realizes voice transmission through the host, so the two steps of encoding and compression are analyzed in key points. In this way, not only can the number of voice frequencies be effectively controlled, but also the system can operate with low load. At the same time, scientific technology is adopted to back up the recording content to prevent malicious damage by lawbreakers.

3.1.2 Processor Design

The TI processor is the core hardware of the system. It can process audio content in real time and accurately, and effectively execute processing commands. In addition, the TI processor can also achieve intelligent and efficient audio recognition. This processor has a frequency of 4.2GHz, and the internal main board has strong functions, which can provide an accurate and intelligent audio recognition environment.

3.1.3 Design of Signal Receiver

The VS78 signal receiver is based on wireless local area network (WLAN) and radio wave technologies to ensure compliance with communication specifications. Shielding technology is used to minimize external wireless signal interference, and frequency optimization is adopted to guarantee signal transmission stability. The voice analysis rate is adjusted within the standard recognition area, and the recognition accuracy of voice broadcast signals in different frequency bands is improved through frequency adaptive technology. The signal receiver can accurately identify external broadcast languages and also improve recognition stability.

3.1.4 Chip Design

The HI89 model chip improves the security of storing key data in the system. The chip introduces the JR7604

program and realizes intelligent audio design through a four-channel storage architecture. In addition, the chip has strong audio processing capabilities, can synchronously process more than 8 broadcast recognition channels, and maximally reduces the loss of audio recognition signals and recognition losses. The chip supports direct connection of audio modules, realizes host-independent automated real-time recognition, and improves system efficiency. The chip parameters are shown in Table 1.

Table 1: HI89 Model Chip Parameters

Parameter		Value
Operating Temperature / °C		-40~85
Input Forward Current / mA		29
Input Reverse Voltage / V		9
Low Level	Input Current / uA	300
	Strobe Voltage / V	1.2
	Delay Time / ns	80
	Input Current / mA	20
High Level	Strobe Voltage / V	9.0
	Delay Time / ns	50

3.2 System Software Design

To ensure the stability and effectiveness of the speech recognition system, the software design needs to give full play to the advantages of AI technology.

3.2.1 Core Word Processing Design

The core word processing in software design consists of two parts: 1) The voice core word library, which determines the recognition efficiency of the system and needs to be scientifically designed; extract high-frequency core words from the preset broadcast word library to ensure sufficient word quantity and close to actual needs. 2) The core word matching area, which filters voice audio based on the word library, with a single core word limited to within 7 bytes, avoiding word correlations. When encountering similar words, trigger the reverse verification mechanism of the word library.

3.2.2 Audio Processing Design

The core of audio processing is to preliminarily optimize the speech signal. Since speech has timeliness and the original audio often contains noise, which affects the recognition accuracy. Therefore, the audio quality is improved through noise reduction processing, and the optimized audio is stably input into the recognition system.

3.2.3 Automatic Recognition Program Design

The design process of the automatic speech recognition program is as follows: 1) Signal processing stage: Based on AI technology, scan the speech signal waveform and complete the conversion from acoustic features (such as frequency/time domain, etc.) to text. 2) Dual verification mechanism, including ① Primary recognition: Generate the initial text by processing the original audio through the acoustic model; ② Secondary comparison: Compare the deviation between the audio features and the recognition result; ③ Dynamic optimization: Iteratively adjust the recognition parameters to improve the accuracy. 3) Performance improvement: Optimize the acoustic model using the differential analysis method and improve the recognition logic through the error feedback mechanism.

3.3 Establishment of the System Operating Environment

3.3.1 Hardware Environment

Hardware environment setup plan: The core components include an application server, a database server, and client terminal devices; the storage configuration shows that the storage requirements of the server are significantly higher than those of the database server because it supports the operation services of the entire system, while the database server only handles data storage functions.

3.3.2 Software Environment

The construction of a reasonable software environment helps to improve the later design efficiency of the device function modules. The establishment of the software environment is inseparable from middleware, program development, databases, development tools, programming languages, etc. In the actual design process of this device, the SQL Server 2020 database, B/S architecture and Java programming language are adopted. Reasonable use of these tools can improve the overall operation level of the module and minimize development costs.

3.4 Experimental Research

In this experiment, to efficiently verify the stability of the system operation, this system was compared with the traditional system, which is an automatic recognition system driven by time-varying sequences and data mining. The purpose of this experiment was to effectively verify the operation stability of the system and evaluate its speech recognition effect by comparing it with the traditional system. The EEG signal was selected as the test sample for the system. To efficiently test and study the speech recognition effect of the above recognition system, the operator randomly selected 5-second EEG signal samples and focused on detecting the recognition effect of OA interference signals. The results showed that the recognition ability coefficient of this system was significantly better than that of the automatic recognition system driven by time-varying sequences and data mining. This system had a higher recognition accuracy rate, verifying the stability advantage of the system in the speech recognition process.

4. APPLICATIONS OF ARTIFICIAL INTELLIGENCE SPEECH RECOGNITION SYSTEMS

4.1 Application of Speech Recognition System in the Scenario of Benefit Giving

In customer operation, the favorability of users is often enhanced through preferential treatment and giving to users, etc., to improve user stickiness. In the customer operation of operators, traffic giving can not only increase user activity and expand market share, but also reduce input costs.

For example, in a certain traffic operation, based on rewarding old users, it is necessary to deeply explore the needs of potential users, select users with a relatively high traffic saturation and give them a certain amount of traffic every month to improve user perception. In the first push, only 5% of the users participated in the activity, failing to reach the expected goal. The reason is analyzed as the defect of the publicity channel: for 100,000 users, the text message sending method was adopted, and 100,000 users were called manually for 1 week, with huge investment costs.

After analysis, the artificial intelligence speech recognition system was adopted for secondary promotion, and the conversion rate was significantly improved. It can be found that in the early stage of the application of the artificial intelligence speech recognition system, the intentional connection rate reached more than 10%, exceeding the text message application rate. The initial response rate of the artificial intelligence voice system was relatively low, so the multi-number sequential call measure was adopted, that is, if the call was not connected for the first time, other numbers would be called sequentially after a preset time interval. Based on this part of the statistical data, the number with the highest call-through rate was pushed to the user-level data source management system, and the system could iterate autonomously, first calling the number with the highest call-through rate. Due to the continuous expansion of the data volume, the artificial intelligence speech recognition system was optimized, improving the invitation effect of customer operation and reaching a 20% conversion rate within two weeks. The anthropomorphic level and call effect of the speech recognition system continue to enhance, thus maintaining double the effect of the initial push. From the user's perspective, a large number of users can accept the activity and actively participate in the short call process, enhancing their favorability towards the enterprise while gaining benefits.

4.2 Application of Speech Recognition System in Customer Care Scenarios

Conducting satisfaction surveys through user follow-up visits is a crucial means of getting close to users, improving services, and consolidating the market. It runs through multiple aspects such as service optimization, business expansion, and brand building, and plays an irreplaceable role in the long-term development of enterprises. However, for operators, conducting satisfaction follow-up visits for all served users incurs relatively high costs. Therefore, only some users are selected, and user satisfaction is evaluated based on partial data. The artificial intelligence speech recognition system resolves this dilemma, enabling comprehensive coverage of a large amount

of user data and effectively controlling the user experience.

Table 2 shows a set of data on user satisfaction in follow-up visits, that is, users who received services recently were followed up to conduct research on whether they received services in a timely manner, whether they were satisfied, and whether they had other opinions, so as to obtain user perception data, and unsatisfied users were appeased through secondary follow-up visits.

Table 2: Statistical Results of Service Satisfaction Survey

Outbound call time	Number of outbound calls / piece	Number of connected calls / piece	Connection rate	Investigation situation				Completion rate of questionnaire	Connection and completion rate
				Not connected	Satisfied	Average	Dissatisfied		
The first batch	1105	883	80.09%	27	471	8	11	46.79%	57.60%
The second batch	1692	1327	78.39%	68	671	11	19	45.49%	57.00%
The third batch	980	780	80.70%	29	397	20	10	46.60%	57.80%

Combined with data analysis and standard satisfaction return visits, the artificial intelligence speech recognition system can perform multi-round interaction operations to accurately obtain user satisfaction. After a series of trainings, the speech recognition system has higher application efficiency and stronger stability, with a connection rate of over 80% and a completion rate of 45%, ultimately grasping the actual situation of users.

However, satisfaction return visits should focus on analyzing the constraints of the artificial intelligence itself. In the above case, users are already dissatisfied with the service; during the satisfaction survey, problems such as the understanding ability constraints of the artificial intelligence speech recognition system, as well as the influence of user accents and environments, are very likely to lead to inappropriate answers, intensify user dissatisfaction, and trigger complaints. Therefore, technical means should be used to optimize the application effect of artificial intelligence and reduce such problems. The first solution is to add an emotion recognition engine to identify user emotions and transfer special users to the manual mode to complete appeasement. Another solution is to combine outbound expansion with artificial intelligence, refer to user big data, link user complaint traces, personalized data, outbound call numbers, etc., obtain user do-not-disturb related data, and further provide solutions, including reducing the outbound call frequency, setting periodic do-not-disturb, etc.

5. CONCLUSION

In conclusion, artificial intelligence technology has significantly promoted the progress of the field of speech recognition. Through actual measurement, it is found that the speech recognition system can accurately process speech information and optimize the context understanding ability by combining natural language processing technology. This paper explores the application and development of the artificial intelligence speech recognition system, demonstrating its important role in customer management. In the future, with the continuous progress of technology, speech recognition will be extended to more application fields, bringing more profound impacts to social development.

REFERENCES

[1] Peng, Qucheng. Multi-source and source-private cross-domain learning for visual recognition. Purdue University, 2022.
[2] Peng, Qucheng, Ce Zheng, and Chen Chen. "A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.
[3] Mao, H. (2026). A Clustering-Based Approach to Image Compression Using the K-Means Algorithm. International Journal of Advance in Applied Science Research, 5(3), 26-30.
[4] Dong, Y., & Li, Z. (2025). Face Generation Model Based on DCGAN. International Journal of Advance in Applied Science Research, 4(6), 59-64.
[5] Jie, Z. (2024). Application of Artificial Intelligence Technology in Industrial Defect Detection. International Journal of Advance in Applied Science Research, 3, 43-48.
[6] Zhang, Q. (2025). Improving the Steel Surface Defect Detection Algorithm of YOLOv5 Network. International Journal of Advance in Applied Science Research, 4(7), 22-28.

[7] Zhang, J., Liu, Y., Chen, X., Zheng, B., Li, J., & Ye, L. (2025). Establishment and Application of Flow Field Evaluation System after Early Polymer Injection in Thick Reservoir. International Journal of Advance in Applied Science Research, 4(12), 54-63.
[8] Wu, J., Luo, L., & Liao, N. (2025). Small-Sample Object Detection of Surface Cracks in Concrete Structures of High-Rise Buildings via Multi-Level Transfer Learning. Innovation & Technology Advances, 3(2), 57–72. https://doi.org/10.61187/ita.v3i2.262
[9] Zheng, H., & Jiang, T. (2025). A New Methodology for Chinese Term Extraction from Scientific Publications. Innovation & Technology Advances, 3(2), 19–45. https://doi.org/10.61187/ita.v3i2.222
[10] Zhang, Y., & Shi, R. (2026). Optimal Design of Thermal Insulation Performance of Transport Packaging Box Based on Response Surface Algorithm. Innovation & Technology Advances, 4(1), 17–30. https://doi.org/10.61187/ita.v4i1.253
[11] Zhu, X., Pan, X., & Gao, X. (2026). Motion Control of Flexible-Joint Robotic Arms for Variable-Station Warehouse Sorting Based on Proximal Policy Optimization. Innovation & Technology Advances, 4(1), 1–16. https://doi.org/10.61187/ita.v4i1.296
[12] Zhang, X. (2026, March). A Hybrid LSTM-GARCH Model Integrating Volatility Factors from the US Financial Markets. In Proceedings of the 2026 International Conference on AI Decision-Making and Management (pp. 183-189).
[13] Shen, Z., Lin, S., Wang, Y., Saunders, E., & Dai, Y. (2026, May). Survival Risk Prediction Model for Colorectal Cancer Patients Based on Graph Convolutional Network and TCGA Multi-Omics Data Integration. In 2026 7th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) (pp. 370-373). IEEE.
[14] Shen, Zepeng, et al. "Research on Application of Whale Optimization Algorithm in Financial Payment Fraud Detection." 2025 4th International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID). IEEE, 2025.
[15] C. Shi, G. Macdonald, B. Jalli, W. Lei, J. Zou, M. Jain, and J. Philip, “Think Less, Label Better: Multi-Stage Domain-Grounded Synthetic Data Generation for Fine-Tuning Large Language Models in Telecommunications,” in 2026 IEEE International Conference on Communications (ICC), 2026, pp. 1–6, doi: 10.1109/ICC59461.2026.11587874.
[16] Gao, W. (2026). Construction of an Intelligent Quality and Operations Control System for Industrial Electrical Enterprises Driven by Dual Standards ISO 14001 and ISO 9001. Journal of Industrial Engineering and Applied Science, 4(1), 1-7.
[17] Zheng, Y., Zhou, G., & Lu, B. (2023). Rebar Cross-section Detection Based on Improved YOLOv5s Algorithm. Innovation & Technology Advances, 1(1), 1–6. https://doi.org/10.61187/ita.v1i1.1
[18] Zhao, X., Zhang, L., & Hu, Z. (2023). Smart warehouse track identification based on Res2Net-YOLACT+HSV. Innovation & Technology Advances, 1(1), 7–11. https://doi.org/10.61187/ita.v1i1.2