Applications and Innovations of Artificial Intelligence Voice Command Platforms in the Power Field

Yuxin Liu[1,*], Nan Lin[1], Yuanjun Zheng[2], Ying Zhang[1], Junxiang Xu[1], Huijian Liu[1], Dongling Jiang[1] and Yingjie Zhu[3]

[1]*State Grid Putian Power Supply Company, Putian, China*
[2]*State Grid Fuzhou Power Supply Company, Fuzhou, China*
[3]*State Grid Jilin Electric Power Research Institute, Changchun, China*

Keywords: Artificial Intelligence, Voiceprint Recognition, Authentication.

Abstract: With the growth of the electricity consumption scale, the volume of information and data rapidly increases in various segments such as power generation, transmission, transformation, and distribution. Consequently, the workload of grassroots teams in electric power has also surged, particularly in repetitive tasks. Simultaneously, the unique network architecture and security requirements in the electric power sector make addressing company affairs and enhancing work efficiency a pressing matter when operating outside the intranet environment. This paper constructs an artificial intelligence voice instruction platform, facilitating the interconnection between the voice platform and the company's intranet channel for office use. Identity authentication is ensured through voiceprint recognition and dual verification involving phone numbers. Moreover, leveraging artificial intelligence speech recognition technology, seamless conversion between spoken language and directives is achieved, thereby executing corresponding contextual tasks to elevate work efficiency and quality further.

1 INTRODUCTION

The current power system is rapidly transitioning towards a new type of power system that is more dynamic, flexible, and intelligent. This transition brings about significant challenges in various areas, such as integrating various distributed renewable energy sources, cybersecurity in the network space, demand-side management, and decision-making for system planning and operations. The new power system relies on underlying information and communication infrastructure and effective processing of large amounts of data generated from various sources such as smart meters, phasor measurement units, and various types of sensors (Li Y, 2022). As electricity consumption scales up, the information and data volume across the entire chain of generation, transmission, distribution, and consumption increases. This results in a growing workload for frontline teams, along with a substantial amount of repetitive tasks.

Moreover, due to the specialized network architecture and cybersecurity requirements in the power sector, breaking the spatial limitations and efficiently handling company matters to improve work efficiency outside the internal network environment presents a challenge. On the one hand, a secure authentication system is required, and on the other hand, the barriers between the internal and external networks need to be overcome.

Artificial Intelligence (AI) is currently a technology with a disruptive impact, encompassing the fields of computational intelligence, perceptual intelligence, and cognitive intelligence. Due to its potential for introducing new technological breakthroughs, AI is leading the way in the Fourth Industrial Revolution. As a core supporting technology for intelligent energy, AI possesses optimization and learning capabilities to address the challenges faced by energy systems, such as dealing with high-dimensional, time-varying, and nonlinear problems (Liu P, 2020).

Therefore, this paper constructs an AI-powered voice command platform that uses voice as a link to establish an instruction platform based on voiceprint authentication. This platform bridges the gap between the voice platform and the company's fixed electricity office through voice while ensuring identity authentication through voiceprint recognition and dual verification with phone

Liu, Y., Lin, N., Zheng, Y., Zhang, Y., Xu, J., Liu, H., Jiang, D. and Zhu, Y.
Applications and Innovations of Artificial Intelligence Voice Command Platforms in the Power Field.
DOI: 10.5220/0012278600003807

In *Proceedings of the 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (ANIT 2023)*, pages 228-232
ISBN: 978-989-758-677-4

numbers. This enables frontline teams to perform various operational tasks through voice commands, further enhancing work efficiency.

The remaining sections of this paper are organized as follows: Section 2 discusses an overview of key technologies, Section 3 introduces the platform's foundational architecture, and finally, Section 4 presents the conclusions.

2 KEY TECHNOLOGIES

The key technologies in this paper include voiceprint recognition and speech recognition, which are briefly introduced below.

2.1 Voiceprint Recognition

Voice Print Recognition (VPR), also known as speaker recognition, is a technology that identifies unknown voices by analyzing the characteristics of one or more voice signals. It is a type of biometric technology. Due to the distinctiveness of each individual's vocal control organs, such as vocal cords, soft palate, pharyngeal cavity, oral cavity, nasal cavity, tongue, teeth, lips, lung volume, etc., their vocal frequency varies, giving rise to unique voiceprint features for each person, including pitch, intensity, duration, timbre, and various nuances. These elements can be decomposed into over 90 characteristics, revealing personality traits such as wavelength, frequency, intensity, and rhythm of different sounds. Voiceprints are distinct for any two individuals and can be observed, described, differentiated, and identified through spectrograms. In comparison to other identity authentication methods, voiceprints exhibit attributes of specificity, stability, universality, uniqueness, resistance to replication, and rapid recognition (Li, 2021).

Voiceprint recognition technology can be categorized into two directions: text-related and text-independent (Waibel A., 1989). In text-related voiceprint recognition methods, the speaker is required to utter predefined words, with both the training and testing voice containing identical text content. Although this recognition method can achieve solid training outcomes, its primary drawback is the necessity to adhere to fixed text during pronunciation.

Text-independent voiceprint recognition technology, on the other hand, imposes no rigorous constraints on the text content of the spoken words. Speakers need only to enunciate naturally, without the confines of fixed dialect or even the potential for mispronunciation. As long as the pronunciation is sufficiently clear, users can approximate real-world conditions during pronunciation. This method is employed in this study due to its user-friendly nature, independence from fixed text content, and reduced likelihood of user resistance.

2.2 Speech Recognition

Speech recognition (Kinnunen T, 2010) is an important biometric identification method. Its task is to identify someone's identity based on their speech signals. Speaker recognition is a valuable biometric technology that has been applied in various fields, such as secure access to high-security areas, voice dialling for devices, banking, databases, and computers. Due to the unique characteristics of speech signals, speaker recognition has gained increasing attention from researchers in the broad field of information security over the years (Ye, 2021).

There are two types of speech recognition: one is called a "speaker-dependent solution," and the other is a "speaker-independent system." In a speaker-dependent system, the solution is tailored for specific use cases where a limited vocabulary needs to be recognized with high accuracy. Speaker-dependent systems operate by identifying unique and specific characteristics of the speaker's voice, much like speech recognition methods. This system verifies the individual's voice, requiring initial training for someone using the system for the first time. This individual needs to read a few words or texts to the Automatic Speech Recognition (ASR) system. The system will then analyze the individual's specific speaking style, after which the person can use ASR. The system is designed to analyze the individual's voice. This is the approach taken in this paper.

Speaker-independent systems, on the other hand, are designed to recognize any voice and therefore do not require speaker-specific training. Speaker-independent systems often have lower accuracy compared to speaker-dependent systems. Typically, speech recognition engines handling speaker-independent systems cope with this fact by constraining grammar (Huang, 1991).

3 ARCHITECTURE DESIGN

The artificial intelligence voice command platform employs a flexible hierarchical structure, and its architectural design includes the access layer, command parsing layer, and command execution layer.As shown in Fig. 1.

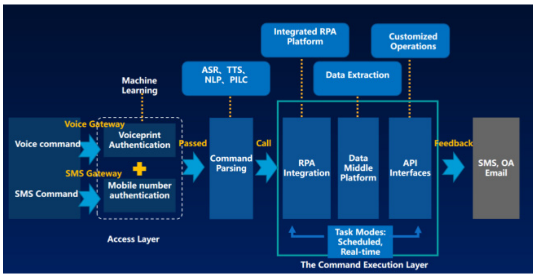

Figure 1: Platform Architecture.

3.1 Access Layer

The access layer comprises the establishment of the voice platform channel and authentication module. This involves creating a communication link between the voice command module and the company's office landline and mobile sides, utilizing voice gateways to facilitate interaction. Dual authentication through voiceprint recognition and phone number verification ensures the identity of the caller. Technologies like Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) are employed to achieve bidirectional conversion between spoken language and text. Artificial intelligence techniques are applied to interpret and understand the given instructions.

Upon entering the system through audio-capturing devices, the voice signals undergo preliminary processing. Preprocessing involves tasks such as endpoint detection and noise elimination. Endpoint detection analyzes the incoming audio stream to automatically remove silence or non-vocal parts, retaining only meaningful speech. Noise elimination filters out background noise to meet user needs in various environments. Processed voice signals then enter the feature extraction phase, where spectral feature parameters that represent specific vocal organ structures or behavioural habits of the speaker are extracted from the voice signals. These parameters exhibit relative stability for the same speaker, remaining consistent across time and environmental changes and demonstrating resistance to noise and imitation. Extracted personal voiceprint feature parameters are used for training within the voiceprint recognition system, generating unique voiceprint models specific to each user. These models are stored in the voiceprint model database, corresponding to user IDs. For a given user, the larger the volume of input speech, the more refined the resulting voiceprint model becomes.

During recognition, the voiceprint recognition system preprocesses the collected voice signal and extracts features, obtaining the parameters for recognition. These parameters are matched for similarity against the voiceprint model of a specific user or all users in the database. Similarity distances between feature patterns are measured using an appropriate distance metric as a threshold to determine recognition results, which are then outputted.

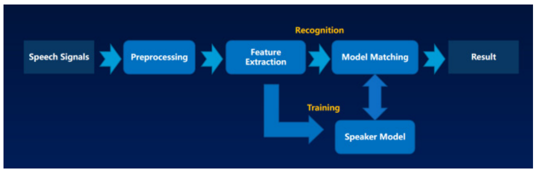

Figure 2: Voiceprint Recognition Workflow.

3.2 The Command Parsing Layer

Speech recognition is crucial in the command parsing layer, and its basic principle includes the following steps:

1). Audio data collection: First, collect voice signals from microphones or other recording devices.

2). Feature extraction: convert the collected speech signal into a feature representation that the computer can understand.

3). Acoustic model: Use a large number of labelled speech data sets to train the acoustic model. The model learns to map acoustic features to units of speech such as phonemes, syllables, or words.

4). Language model: To improve accuracy, add a language model, which helps to predict the next possible word based on previous words and grammatical structures.

5). Decoding: According to the output of the acoustic model and the language model, an algorithm is used to find the most likely text sequence.

Another core component of the command parsing layer is the power industry's specialized command library. Based on foundational configuration information such as voltage levels and equipment types, a basic command library is generated. Additionally, utilizing artificial intelligence technology and continuous corpus training, a specialized command library for the power industry is developed to accommodate specific contexts within various departments, including equipment maintenance, customer service, and performance analysis. The following steps further elaborate on this process:

1). Generation of Basic Command Library: Using foundational configuration information such as voltage levels and equipment types, a basic command library is generated.

2). Context and Domain Annotation: Data preprocessing involves tasks like text cleaning, tokenization, and stopword removal to prepare data for model training. To enable command recognition and understanding within specific contexts, training data needs to be annotated to indicate in which context the data is valid. The data is categorized, and labels are added for each category to indicate the specific context to which the command applies (e.g., equipment maintenance, customer service, performance analysis).

3). Model Training and Optimization: Leveraging annotated data, modern deep learning techniques are used for pre-training, followed by fine-tuning according to specific tasks.

4). Iterative Training: Model training is an iterative process. The trained model is used to infer new data, and inference results are compared to real labels. The model is continuously optimized through backpropagation.

3.3 The Command Execution Layer

The command execution layer mainly consists of RPA, Data Middleware, and API interfaces, as follows:

1). RPA Integration: The first integrated RPA platform within Fujian Province achieves integration by utilizing API interfaces of RPA.

2). Data Middle Platform: In response to the central platform strategy of State Grid Fujian Province Company, the data middle platform directly extracts data from various specialized systems, breaking down silos.

(3). API Interfaces: Primarily involves custom operations through Python scripts and others, executing various tasks such as one-click disconnection.

Robotic Process Automation (RPA) is an automation technology (Ribeiro J, 2021) that uses software robots to simulate human operations. It automates repetitive and standardized tasks, enhancing efficiency, reducing human errors, and freeing up human resources. RPA is often used for structured data and repetitive business processes. Currently, the State Grid Fujian Electric Power RPA platform in Putian Company has deployed over a hundred RPA processes.

Data middle platform (Wu H, 2020) is a comprehensive data management platform that collects, stores, integrates, and processes both internal and external enterprise data. It consolidates data from different sources and provides standardized data models, allowing various business departments convenient access and use of the data. Data middle platform emphasizes data sharing, consistency, and quality, enabling data to be a crucial support for business decisions and optimization.

API interfaces enable customized operations for various scenarios. Developers can use API interfaces to implement Python scripts, applications, and even system integrations. These interfaces facilitate data exchange between applications, which can be on different platforms, programming languages, or devices. API interfaces allow these applications to connect and communicate, enabling data sharing and collaborative work.

By integrating RPA, data middle platform, and API interfaces, the voice command platform can achieve more efficient automation processes, leveraging the advantages of data-driven optimization for business processes and decision-making.

4 CONCLUSION

This article applies artificial intelligence voiceprint recognition and speech recognition technologies in human-computer interaction processes to achieve user authentication during login. It simplifies entering passwords and passphrases, enabling various scenario tasks to be executed through voice commands. This includes remote startup, shutdown, and more, freeing up the hands of workers, enhancing human-computer interaction experiences, facilitating efficient and secure interaction with complex information, and boosting work efficiency and productivity. Currently, it has been promoted within the State Grid Putian Company, with 61 scenarios developed involving areas such as supply instructions, operation inspection, marketing, etc. Since its deployment, the platform has completed 6,214 tasks, resulting in a saving of approximately 9,000 hours of manual work, showing remarkable results.

However, the platform still has some areas that need further improvement in subsequent development:

1). Optimize the voiceprint recognition model to enhance recognition accuracy.

2). Expand the dialect feature library to improve recognition accuracy for dialect commands.

REFERENCES

Li Y, Xiao T, Liu W, et al. Information Collection System of Power Integrated Energy in Artificial Intelligence Environment[C]//*2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*. IEEE, 2022: 380-386. https://doi.org/10.1109/icdsca56264.2022.9988606

Liu P, Jiang W, Wang X, et al. Research and application of artificial intelligence service platform for the power field[J]. *Global Energy Interconnection*, 2020, 3(2): 175-185. https://doi.org/10.1016/j.gloei.2020.05.009

Waibel A. Modular construction of time-delay neural networks for speech recognition[J]. *Neural computation*, 1989, 1(1): 39-46. https://doi.org/10.1162/neco.1989.1.1.39

Kinnunen T, Li H. An overview of text-independent speaker recognition: From features to supervectors[J]. *Speech Communication*, 2010, 52(1): 12-40. https://doi.org/10.1016/j.specom.2009.08.009

Ye F, Yang J. A deep neural network model for speaker identification[J]. *Applied Sciences*, 2021, 11(8): 3603. https://doi.org/10.3390/app11083603

Ribeiro J, Lima R, Eckhardt T, et al. Robotic process automation and artificial intelligence in industry 4.0–a literature review[J]. *Procedia Computer Science*, 2021, 181: 51-58. https://doi.org/10.1016/j.procs.2021.01.104

Li J, Zhang J. A study of voice print recognition technology[C]//*2021 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2021: 1802-1808. https://doi.org/10.1109/iwcmc51323.2021.9498681

Huang X. A study on speaker-adaptive speech recognition[C]//*Speech and Natural Language: Proceedings of a Workshop Held at Pacific Grove*, California, February 19-22, 1991. https://doi.org/10.3115/112405.112458

Wu H, Shen L, Chen X, et al. Research on the application of data middle platform technology in integrated energy business system[C]//*2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)*. IEEE, 2020: 316-319. https://doi.org/10.1109/iicspi51290.2020.9332448