Application of Speech Recognition Technology in Power Grid Dispatching Automation

Faqi Yan¹, Chunming Wang¹, Jianzhong Dou¹, Yang Liu¹ and Xusheng Yang²
¹Central China Power Dispatching and Communication Center, China
²Beijing Yongshang Technology Co., Ltd., China

Abstract. In recent years, with the rapid development of social economy, China's investment in power grid dispatching automation system has also been increasing year by year to meet the increasing demand for dispatching of modern power grids. Under this background, the smart grid system characterized by information, automation, intelligence, and interaction has become the trend of future electric power development. Voice human-machine interaction technology as an advanced interactive technology has also started in the electric power intelligent dispatch system. Used in the application. This article mainly expounds the development status of intelligent dispatching system and the application of human-computer interaction technology in intelligent dispatching system. It also analyzes speech recognition technology, speech synthesis technology and typical Chinese speech recognition system in speech human-computer interaction technology, and focuses on intelligence. The design and implementation of voice human-computer interaction technology in the dispatching system, with a view to provide theoretical guidance for the development of China's power dispatch system.

1. Introduction
The development status of intelligent dispatch systems the development direction of modern power grids is large-scale and complex. The amount of information that grid dispatch systems need to receive is extremely huge, which also brings great difficulties to power dispatch. How to ensure the stability and safety of power systems? The efficient operation has become a technical problem to be solved in the development of the power system. With the continuous development of computer information technology, intelligent software systems based on computer technology, information technology, automation technology, etc. are gradually applied in smart grid dispatch. In the smart grid dispatching system, through the new grid control technology, information and communication technology, automated management technology and other modules, can achieve intelligent communication of data and information in all aspects of the power grid [1], the current business scope of intelligent dispatch system mainly includes real-time monitoring, analysis and evaluation, Adjustments and controls, scheduling plans, dispatch management, etc., while grid status calculations, grid dispatch decisions, and system platform construction have become the main directions for future intelligent dispatch systems [2].
2. Definition of voice recognition technology in intelligent dispatching system

Model Design of Speech Recognition Technology Human-computer interaction technology refers to the process of information interaction that enables a person and a computer to accomplish a certain task through an interactive method, thereby realizing the general term of a type of technology in which a person interacts with a computer. Influencing Factors of Human-Computer Interaction Level in Intelligent Dispatch System In the intelligent dispatch system, the overall level of human-computer interaction is closely related to the service mode required by the request. Therefore, the influencing factors mainly include the following aspects: First, the request mode, after the intelligent scheduling system application program is started, the input device of the human-computer interaction system is also started. The input device will perform real-time monitoring of the input signal and perform the corresponding task according to the input request, so the operation of the request mode is performed. The situation is closely related to the level of human-computer interaction. The second is the sampling mode [3]. The human-computer interaction system needs to analyze various types of input requests based on the data information of the storage area, and can independently complete the data update, thereby significantly improving human-computer interaction efficiency; Event mode, this mode can establish a direct link between the human-computer interaction program and the input device. After accepting the input data, the mode is passed on the one hand and the program is executed on the other hand. On the other hand, it also automatically completes the storage and processing, which greatly improves the man-machine. Interactive system reliability.

3. Speech Recognition Technology Requirements

Typical Problems of Speech Recognition Analysis to meet the practical requirements of speech recognition, the following issues should be properly handled [4]. 1) Noise reduction. Clearly, speech recognition cannot avoid a noisy environment. The so-called noise, in addition to the background noise of the objective environment, the speaker's insensitivity due to emotional changes (compared with normal) is also a noise factor that cannot be ignored. At present, the methods to deal with noise mainly include spectral subtraction, environmental regulation techniques, and establishing a reasonable noise model. 2) Primitive selection. According to experience, to make the speech recognition system recognize more vocabulary, the selected primitive should be as small as possible. 3) Endpoint detection (determining the starting and ending point of speech). According to big data statistics, over 50% of speech recognition errors originate from endpoint detection. The key to improving the endpoint detection success rate is to find stable speech parameters. 4) Speed of recognition and rejection. Voice input should not use dialect or colloquial language as much as possible to improve the speed and success rate of speech recognition [5].

3.1. Contents of Speech Recognition Technology

Since speech signals are essentially non-stationary signals, the analysis of speech signals is currently based on the assumption of short-term stability. After the short-term steady assumption of the speech signal, the speech signal is windowed to implement feature extraction on the short-term speech segment. These short-term segments are called frames, and feature sequences in units of frames constitute the input to the speech recognition system. Since the Mel cepstrum coefficient and the
perceptual linear prediction coefficient can accurately describe the speech signal from the perspective of human auditory characteristics, it has become the current mainstream speech feature. In order to compensate for the inter-frame independence assumption, when people use the Mel cepstrum coefficients and the perceptual linear prediction coefficients, by speaking instruction series \( X \), To identify the language series \( P \).

\[
W = \arg \max \left\{ P(W | X) \right\} = \arg \max \left\{ \frac{P(W | X) P(W)}{P(X)} \right\}
\]

The windowed framed speech signal of the speech signal is a kind of uncertain non-stationary one-dimensional random signal as a whole. The formation process of the speech signal is closely related to the state change of the vocal organ, and can be considered in a very short time. The shape of the vocal organ such as the door remains unchanged, so the formed speech signal has a "short-time stable" characteristic.

\[
N = \frac{f \times t}{1000}
\]

It was found that the acoustic characteristics and signal spectrum are almost constant within 10-30 ms. According to this characteristic, we can cut the speech signal, that is, divide the signal into frames, and a small segment of speech within 10-30 ms as a frame. The length of a frame of speech is the length of the frame, and the speech signal in each segment can be considered as a stationary signal, so that a locally stable signal is obtained. In order to ensure a smooth transition between each frame of the speech signal and ensure the continuity of the speech characteristics, an overlapping portion is generally generated between the two frames of the speech signal before and after, and overlapping portions of the speech signals before and after the two frames are called frame shift. The ratio of frame shift to frame length is generally 0.5.

\[
S_m(n) = S(n) * W(n)
\]

According to the short-term stable characteristics of the speech signal within 10-30 ms, the length of each speech signal can be roughly calculated. The sampling frequency \( f \), the number of sampling points \( N \) in each frame, and the duration \( t \) of each frame of speech have a correspondence relationship. Therefore, in the scheduling automation, the following calculation sound function can be applied:

\[
W(n) = \begin{cases} 
1, & 0 \leq n \leq (N-1) \\
0, & n = \text{other}
\end{cases}
\]
After calculation, the signal expression wave forms after the simulation can obtain the sub-window of sound waves are as follows: The principle of selecting the window function is usually to ensure that the speech fragments intercepted by the window function can maintain the original information volume of the voice signals and concentrate the voice energy. To the main lobe, it has better out-of-band attenuation capability and better stop band attenuation capability. If the side lobe height in the spectrum tends to zero, the more the energy is concentrated in the main lobe, the closer the speech signal after the windowing function is to the original speech signal. There are many window functions for framing a speech signal, which are selected according to the spectral characteristics and research significance of the speech signal. In speech recognition technology, speech signals are usually processed. The rectangular window is a relatively simple window function. The rectangular window expression is as follows:

![Speech Recognition System Waveforms](image1)

**Figure 2.** Comparison of different identification system wave forms
They usually add their first-order and second-order differences to introduce the dynamic characteristics of the signal features. Acoustic models are the most important part of speech recognition systems.

3.2. Methods for Establishing a Speech Recognition Technology Model

Acoustic modeling involves many aspects such as modeling unit selection, model state clustering, and model parameter estimation. In the current power grid dispatching automation system, a context-dependent model is generally used as a basic modeling unit to characterize the coordinated pronunciation of continuous speech. After considering the influence of the context, the number of acoustic models increases dramatically. The system usually adopts a state clustering method to compress the number of acoustic parameters to simplify the training of the model. In the training process, the system prepossess several training speeches and obtains feature vector sequences through feature extraction. Then, the feature modeling module establishes a reference pattern library for training speech. Search is the process of finding the optimal word sequence according to certain optimization criteria in the specified space. The essence of search is problem solving, which is widely used in various fields of artificial intelligence and pattern recognition such as speech recognition and machine translation. It uses the acquired knowledge (acoustic knowledge, phonetic knowledge, dictionary knowledge, language model knowledge, etc.) to find the optimal state sequence in the state (words, acoustic model, HMM state from top to bottom). The final word sequence is an optimal description of the input speech signal under certain criteria. In the recognition phase, the feature vector parameters of the input speech are compared with the patterns in the reference template library obtained by the training, and the category to which the pattern with the highest degree of similarity belongs is output as the recognition intermediate candidate result. In order to improve the accuracy of the recognition, the candidate recognition results obtained above are further processed in the post-processing module, including the fusion of higher-ranking language models by re-scoring, the reliability degree of the recognition results obtained by the confidence measure, and the like. Finally, by adding constraints, more reliable recognition results are obtained.
3.3. Acoustic modeling methods commonly used acoustic modeling methods

Matching grid scheduling automation through this method is an earlier method of pattern matching. Based on the idea of dynamic programming, it solves the problem of template matching with different lengths of speech signal feature sequences in isolated word speech recognition.

![Diagram](image)

**Figure 4.** Language recognition device access SCADA system

In practical applications, this method calculates the similarity between the processed and framed speech signal and the reference template, calculates the similarity between the templates according to some distance measure, and selects the best path.

This method is a statistical model established for the time series structure of speech signals. It is developed on the basis of this method. It is a statistical identification method based on parametric models. HMM can imitate human speech process and can be regarded as a double stochastic.

One is to simulate the implicit stochastic process of the change of statistical characteristic of speech signal with this method with finite state number, and the other is with each state of the method. The random process of the associated observation sequence, which simulates neuronal activity in a mathematical model, fully applies the principle of parallel distributed operation of a large number of neurons in an artificial neural network, an efficient learning algorithm, and the ability to imitate human cognitive systems. The speech recognition field, combined with the neural network and the identification algorithm implying the method model, overcomes the shortcomings of the method in describing the temporal dynamic characteristics of the speech signal, and further improves the accuracy of the method of speech recognition. The successful method is to replace Gaussian mixture model in the mixed model to estimate the posterior probability of phoneme or state. Power dispatch automation system greatly improves the accuracy of speech recognition.

3.4. Voice Interaction Technology Advantages

The specialization and standardization of the terminology of voice interaction and contouring technology provide good conditions for the application of voice recognition technology. The speech technology is integrated into the power dispatching voice recognition technology to realize voice recognition information based on voice recognition technology. Screen statistics, calls and management. Speech interaction includes speech recognition technology and speech synthesis technology. Contour lines can be used to set color gradients that reflect changes in line flow, voltage, and other data. Visualization of data such as load levels, voltage levels, etc., is applied to the entire network, from brightness, color, threshold, color layer spacing, and saturation. Finally, a clear and easy-to-observe effect can be achieved. Contour coloring technology is based on geographic
information systems. Through its combination with the node operating data of the power system, it establishes and optimizes the interpolation and search of triangular meshes, is-value points, fitting curves, and drawing colors in layers. Generates a node run data contour method.

3.5. Technical requirements for voice recognition of power grid dispatch automation
Voice recognition of power grid dispatching automation systems with the continuous expansion of the power grid scale, the requirements for the statistical accuracy and speed of the dispatching data are also increasing. Therefore, this decentralized upgrade and transformation can no longer meet the requirements of the overall voice recognition under intelligent control. Dispatch voice recognition breaks the inherent boundary between existing systems, considers scheduling business model as the main, analyzes and discusses every link from power grid dispatching production, and provides interactive technology window platform through voice recognition structural module group to make scheduling Personnel can study in detail and control the operation process of the power grid before, now and in the future, and exert its control over the power grid.

3.6. Speech recognition of production real-time monitoring system
The real-time monitoring content of the power grid mainly includes the trend of the whole network, the voltage of the whole network, the power angle of the engine, and the cross section of the stability section. Voice recognition technology for real-time production monitoring systems mainly includes two aspects of voltage and voice recognition applications in factories and station nodes. The level of voltage and its distribution in the voice recognition applications of plant and station node voltage are important basis for measuring whether the grid can operate safely. Voice recognition for production of real-time monitoring systems includes multiple technologies in real-time operations: Voice recognition technology for voltage monitoring; Voice recognition applications for power flow monitoring; Voice recognition applications for spin standby; Voice recognition for other operational status monitoring Application Technology.

3.7. Real-time observation and control of speech recognition in power grids
The research objects of speech recognition mainly include: substation transformer load, equipment overload, power plant output, line active status, and so on. Through a variety of known speech recognition methods and calculation methods, intuitive display of scheduling data is achieved. For the analysis results and discussion of automatic voltage control, automatic power generation control and other control applications, according to the requirements of voice recognition, various intuitive voice recognition interfaces can be set up, which facilitates the dispatcher to control and monitor the power grid more conveniently, and helps the dispatcher to move from the whole Azimuth and mulch-angle grasp the real-time operation of the power grid, which in turn can reduce the pressure on dispatchers and improve work efficiency. The voice recognition of power grid early warning and auxiliary decision-making scans the power system for faults and performs power flow calculations, prevention control or anticipation of accident analysis. The results of the warning analysis are displayed through various voice recognition widgets. Voice recognition indicators have voltage over-limit margins, predictive fault transient warnings, dynamic stability damping margins, and so on, as well as early warning analysis results from both the severity of the fault and the vulnerability of components in the system. Voice recognition applications in prevention, emergency, and recovery control refer to the intuitive display of calculation results and analysis results in prevention and control. There are many supplementary analysis methods to adjust the operating mode of the system in the event of an accident or abnormality. There are mainly emergency control and recovery control. Voice recognition applications in prevention, emergency and recovery control help prevent emergencies.

3.8. Speech recognition technology of a power wiring diagram
The power regulation one-time wiring diagram voice recognition is the basis of the application technology of speech recognition. Many applications of the speech recognition technology are based
on the power regulation one-time wiring diagram in the power dispatch system as the base plate, and on this basis, the layered information can be displayed. In the power dispatching, a wiring diagram for power regulation is generally applied to the voice recognition technology, and the geographic information system and the power dispatching system are generally combined organically. For example, the alignment of the line and the site of the plant usually rely on the power regulation of a wiring diagram. In addition to providing the inherent voice recognition means, the power dispatching system also needs to provide the necessary voice recognition tools to ensure flexibility and compatibility.

4. Conclusion
Speech recognition technology is widely used in power dispatching with the development of science and technology. The grid dispatch voice recognition is based on the transient, dynamic and steady-state data in the operation of the power grid, as well as real-time, early-warning, prior and other perspectives. Through the combination with the power system application, various related grid dispatch data are organized and integrated. Then use a good voice recognition interface to provide a very effective help for the dispatcher to deal with the problems in the grid work process.

References
[1] Yuan Sandong, et al. Research on speech recognition in the process of power dispatching operation. China New Technology and New Products, Vol. 5 (2013) No.26, p.316–321.
[2] Deng Xiaoqing, Zhou Wei, et al. How to Do a Good Job in Power Dispatching Operation. Science and Technology, Vol. 9 (2012) No.32, p.129-135.
[3] Lei Yi, Hu Xiaofei, Li Duanchao, et al. Research and application of power dispatching voice recognition technology. Electric Power Informatics, Vol. 11 (2011) No.18, p.187~192.
[4] Yao Jianguo, Yang Shengchun, Gao Zonghe, et al. Power grid dispatch automation system development trend forecast. Power System Automation, Vol. 5 (2007) No.45, p.117-123.
[5] Lei Yu, Huang Taigui, Yuan Lin, et al. Research and Practice of Power Dispatching Data Integration Application. Electric Power Informatization, Vol. 1 (2010) No.22, p.118-122.