Advances in Computer Science Research, volume 93
2nd International Conference on Mathematics, Modeling and Simulation Technologies and Applications (MMSTA 2019)

Innovation and Application of Speech Recognition Technology in Automatic Fare Collection of Rail Transit

Xiaoting Dong
Technical Center of Building 1, 909 Guilin Road, Xuhui District, Shanghai, China

***Abstract*—Based on the strategic planning of Shanghai Metro Smart Station and the main problems existing in the Automatic Fare Collection System of rail transit, the application requirements of speech recognition technology in the Automatic Fare Collection System of rail transit are put forward. On the basis of construction background, research and development of key technology points, software and hardware architecture of the scheme and pilot analysis, a voice ticket purchase scheme is proposed. On the basis of the original ticket purchase process of Ticket Vending Machine (TVM), voice recognition technology is used to increase the functions of wake-up-free voice input, Chinese phonetic alphabet input and fuzzy location inquiry, so as to reduce the queue pressure of stations with large passenger flow, improve the efficiency of passenger ticket purchase, and the degree of equipment modernization.**

Keywords—speech recognition technology; rail traffic; automatic fare collection system

I. PROJECT BACKGROUND

Since 2011, Shanghai has made every effort to promote the construction of future-oriented smart cities. Relying on the background of smart city, Shanghai Metro strives to build smart metro, focusing on efficiency, efficiency, reliability, improving safety and service level. As the node of rail transit network and the concrete carrier of passenger service, the intelligent level of station will directly map the intelligent construction level of the whole line and the overall network. Smart Station is the concrete manifestation and externalization of Smart Subway. More and more new technologies have been applied in AFC system of rail transit. As an important automation terminal, ticket vending machine is an important window for passengers. The traditional ticket vending machine of rail transit basically uses finger touch input mode to interact with human beings [1]. With the network operation of urban rail transit, more and more information needs to be displayed on the passenger display screen. Many passengers are not familiar with the use of ticket vending machine or the ticket vending interface is not friendly enough. They have made wrong operation, and the input efficiency of finger touch control has become low.

In order to realize the strategic planning of Shanghai Smart Metro and to solve the problem that some passengers can not choose the destination station in the complex network of Shanghai Metro. Shanghai Metro has developed the International Pioneering "wake-up-free" and high accuracy far-field voice recognition technology adapted to the high traffic density and strong noise environment, and the combination of natural semantics understanding technology and cloud-based electronic map, automatic and accurate matching of sites, to achieve a new type of ticketing function of voice input and alphabet input of designated sites and fuzzy locations in complex scenarios of rail transit, and further enhance the intelligence of automatic ticketing and checking equipment.

II. RESEARCH OF KEY POINTS OF SPEECH RECOGNITION TECHNOLOGY

In metro stations with high traffic density, the most difficult point of applying voice technology is to resist noise interference. Compared with the general space, the noise in metro stations is often great intensity and complexity. Traditional speech signal processing method based on single microphone has a speech recognition error rate of over 67% in the strong noise environment of subway station. In order to solve the problem of speech recognition accuracy in the strong noise environment of metro station, the project team originally developed a high accuracy, wake-free far-field speech recognition technology in the high traffic and strong noise environment, and successfully solved the above problems. The main innovative technologies are as follows:

A. Multimodal Speech Signal Processing Technology

Speech signal processing technology based on microphone arrays often fails to achieve the desired results in strong noise environment. In order to avoid noise interference, the speech signal processing and face detection are combined by using the information of microphone array and camera. Firstly, the accurate location of passengers is carried out by face detection. When multiple faces appear in the camera picture, the only passenger can be located according to the angle of the face, the relative position of the camera and the movement of the lips. Then, according to the location information of the speaker, the noise suppression can be effectively carried out based on the microphone array technology. At the same time, the passengers can be automatically detected approaching the TVM. The speech signal processing technology actively initiates interaction to realize wake-up-free voice interaction experience. In order to further improve the effect of speech recognition in

noisy environment, a multi-channel voice front-end signal processing engine is invented in the back-end algorithm. The engine integrates many signal processing technologies based on physical modeling, such as multi-microphone spatial filtering , speech separation, de-reverberation and sound source localization, and integrates the data modeling mechanism based on machine learning. The engine has certain Inhibitory ability to background noise, non-stationary interference, equipment echo, room reverberation and other kinds of noise.

FIGURE I. TVM WITH SPEECH RECOGNITION

B. *Speech Recognition Technology in Strong Noise Environment*

In order to maximize the speech recognition rate in ticketing scenarios, the project team used real business data to optimize the acoustic model and language model respectively. After collecting tens of thousands of hours of voice data at the subway station, the acoustic model is optimized after manual labeling. At the same time, a large number of business-related text data such as subway station names and place names are collected, and the language model is optimized by using hot word loading and interpolation technology, which makes the speech recognition technology highly business-specific, solves the problem that the general model can not correctly recognize the proper nouns, and the actual speech recognition error rate was reduced to less than 7%.In the decoding engine, the low frame rate decoding technology is adopted. In the decoding process, the soft phoneme decision benchmark is used. With the error rate unchanged, the decoding speed can be increased three times, so that the real-time rate of speech recognition process can be maximized [3]. Compared with traditional CTC method, word-level CTC method can reduce the error rate by 10% and the computational load by 30%. It can improve the concurrent ability of single machine support by more than twice.

C. *Stream Multi-Round Multi-Intention Semantic Understanding Technology*

In the actual ticketing scenario, there are often multiple intentions in a voice instruction. For example, "go to the Oriental Pearl TV Tower, no, go to Lujiazui, two tickets." To solve this kind of multi-intention problem, the project team developed a multi-round multi-intention semantic understanding technology. In the process of passenger input, dynamic semantic understanding is achieved by combining context information, and real-time correction is made according to the latest passenger input. Stream Multi-Round and multi-intention semantic understanding involves many sub-tasks, including: entity information extraction, long sentence semantics segmentation, intention recognition, multiple relationship extraction, entity link, entity reference resolution and so on. Traditional methods model these tasks separately, and then solve the whole task in series. In contrast, the project team proposed a new end-to-end solution, which skipped the dependence on subsystems and directly modeled passenger input and intention, instead of relying on subsystems. This method avoids the error accumulation and transmission among subsystems, reduces the complexity of the system, reduces the dependence of training data, and greatly reduces the cost of research and development.

III. COMPOSITION AND FUNCTION OF SYSTEM SCHEME

A. *System Software Solution*

1) *Ticket Purchase Mode Service*

The TVM with speech recognition technology includes three modes: voice query, spell query and touch screen selection. Among them, the selection of touch screen is the existing selection of traditional TVM, which will not be introduced here. The service modes of voice query and spell query are as follows:

(1) Voice query mode: When passengers say their destination, the system will judge and query the nearest subway station to the passenger's destination based on map data. When there are multiple possible destinations, route selection lists will be provided for passengers to choose, and detailed transfer routes for all destinations will be displayed, including the route where the transfer station is located, the name of the transfer station and the number of transfers.

1) Mode activation: When passengers are detected approaching the TVM or clicking the voice ticket button on the main interface, they will automatically enter the voice query mode.

2) Mode closure: When a passenger is detected leaving the area outside the voice window of the TVM screen, or the passenger clicks the close voice buying button on the main interface, the voice query mode will be automatically closed.

3) Mode disablement: When the voice recognition module works abnormally or fails to start, the TVM opportunity automatically disables the voice query mode and hides the voice ticket purchase button until the voice ticket purchase module returns to normal.

(2) Spell inquiry mode: Passengers enter the destination Spell initials by clicking on the alphabetic keyboard on the TVM screen, and the screen will prompt candidate locations verbatim with input. When passengers select their destination,

they display the destination list in the form of a list and replace it with the path details.

1) Mode activation: When passengers click the Spell ticket button on the main interface, they will automatically enter the Spell query mode.

2) Mode closure: When the passenger clicks the button to close the Spell ticket purchase on the main interface, the Spell query mode will be automatically closed.

3) Mode disablement: When the Spell query module works abnormally or fails to start, the TVM automatically disables the Spell query mode and hides the Spell query button until the spell query module returns to normal.

2) Background Query Service

Background query service is mainly used to realize voice recognition, dialogue and semantic understanding, voice blurred location query, Spell blurred location query and transfer path query.

(1) Speech Recognition Service

The voice of passengers in noisy places is converted into text, and the names of places and metro stations are recognized accurately.

(2) Dialogue and Semantic Understanding Service

Passenger's intention can be accurately identified by analyzing the text content of passenger's voice conversion. The passenger destination is accurately judged and the destination name is extracted.

(3) Voice Fuzzy Location Query Service

According to the location name, the nearest subway station is inquired, and the distance, exit and other information are obtained. When there are multiple candidates, they are returned in the form of a list for passengers to choose.

(4) Spell Fuzzy Location Query Service

According to the initials or complete spell of the places entered by passengers, the corresponding place names are queried, and then the nearest metro station is queried, and the distance and exit information of the metro station are obtained. When there are multiple candidate locations, the passengers can select in the form of a list.

(5) Transfer Path Query Service

According to the passenger's current station, the best transfer route to the destination station is inquired, and information such as transfer station, transfer route is acquired.

B. System Hardware Solution

The hardware of the ticket vending machine for speech recognition is mainly composed of server module, industrial computer module, camera module and microphone array module. The server is used for real-time background data processing and supports the running of online real-time speech recognition daemon and map query daemon. After the microphone array and camera collect the sound and image data, they are processed by speech signal processing and video signal processing respectively, and processed by the multi-mode engine in the industrial computer module. The effective user voice is recognized from the noise environment. Through the identification, understanding and query of three service modes of cloud voice recognition service, dialogue understanding service and Gaud path query service, the main program display interface of TVM is finally transmitted through serial port or inter-process communication.

Considering the stability and security of TVM devices and modules, voice recognition services are deployed locally, and TVM with fuzzy query function are connected to AFC private network by wired way. Considering the concurrency and reliability requirements, the background servers in the MCC room are accessed by deploying servers in the multi-line central system (MCC) room. All kinds of services are physically isolated from the Internet through offline CD-ROM upgrade, playing a safe and controllable role.

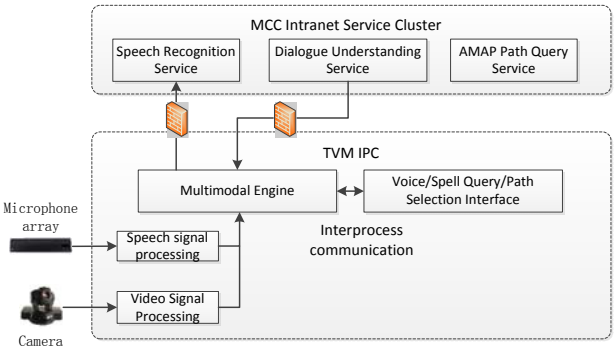

FIGURE II. HARDWARE SCHEME OF SPEECH RECOGNITION SYSTEM

IV. EXPERIMENTAL APPLICATION

By formulating the technical standard of speech recognition, the project team quantitatively stipulated the index of inquiry destination accuracy, coverage of geographical coordinates, anti-interference requirements and processing speed, designed the ticket purchasing process in line with the composite consumption of Shanghai Metro, and developed a speech recognition system in strong noise environment, which was used in the simulation test environment. At present, the speech recognition technology has been successfully applied to 44 TVM equipments in Shanghai Railway Station, Hongqiao Terminal 2, Shanghai South Station and Pudong Airport Station. The application results are as follows:

A. The Time Consumed for Selecting Stations is Reduced by a Large Margin

TABLE I. TIME-CONSUMING COMPARISON TABLE OF SELECTED STATIONS

Ticket Purchase Type	Time-consuming	Average time consumed
Station selection through the speech recognition (From passengers clicking on the route to choosing destination stations)	2.2s-3.5s	3s
The original manual selection station (Input voice from passenger to interface to display the optional station and transfer path of destination)	5s-20s	14.86s

B. Ticket Sales Increased by More Than 8%

Without voice recognition, the average number of tickets sold per device per day is 884. The average number of tickets sold is about 1025 when the speech recognition function is used.

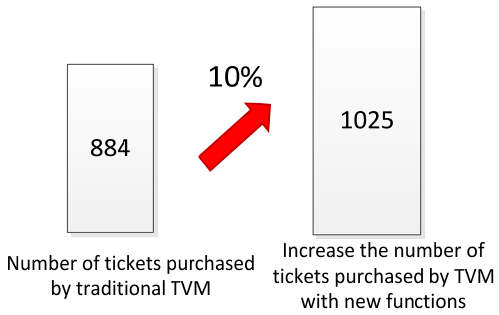

FIGURE III. CONTRAST CHART OF TICKET SALES GROWTH

Speech recognition function can provide personalized voice prompt in the key steps of passenger ticket purchasing, which improves passenger's ticket purchasing accuracy, service support and ticket purchasing experience, reduces equipment failure rate effectively, reduces the investment of maintenance manpower and material resources, and improves passenger's ticket purchasing accuracy, service support and ticket purchasing experience. Intelligent and modern level of automatic ticket-selling and checking equipment, give full play to the effectiveness of the TVM, provide better services for passengers, achieve the original intention of providing more convenient services for passengers. In the future, it can be popularized in an all-round way and applied to more fields.

V. CONCLUSIONS AND PERSPECTIVES

With the concept of "Internet + Traffic" gradually win support among the people, voice recognition technology, as a new method of ticketing, is different from the traditional ticketing mode of TVM. The application in rail transit can not only increase the ticketing experience of passengers， improve payment efficiency, expanding the marketing model of subway ticketing and also an important measure to improve the level of rail transit services and promote the construction of smart subways and smart cities. Through its unique technological advantages, it will achieve seamless connection between voice recognition technology and the AFC of rail transit. Using technologies such as big data, information technology, smart terminals, and the Internet of Things, the concept of "energy and efficiency +" is advocated to bring people a more convenient and efficient life experience, and to help rail transit adapt to the new normal of economic development.

REFERENCES

[1] Zhang Ning, Hetiejun, Wang Jian, A study on the interchangeability of automatic ticket Vending Machine for rail transit[J]. Urban Rail Transit Research, 2007(11) : 37.

[2] Jiao kejie, Research and Application of Speech Recognition Technology in Intelligent Ticketing Terminals of Rail Transit, Broadcasting Communications and Television, Issue 2 ~ 3, 2018.

[3] Acero A. Acoustical pre-processing for robust speech recognition[C] ,Stern R M, Association for Computational Linguistics, 1989: 311-318.

[4] Ji zhendong, Research on the Application of Big Data Analysis Cloud Platform Technology in Intelligent Transportation[J], Silicon Valley.