International Civil Aviation Organization

Sixth Meeting of the Asia/Pacific Air Traffic Management Automation System Task Force (APAC ATMAS TF/6)

Bangkok, Thailand 2-4 June 2025

Agenda Item 5: ATM Automation System Implementation Experience by States
5.6. Development of New Technology

RESEARCH ON SPEECH RECOGNITION TECHNOLOGY IN ATMAS

(Presented by China)

SUMMARY

This paper introduces the research progress of speech recognition technology in ATMAS in China, from the perspective of speech recognition processing mechanism, functional deployment and specific scenarios, expounds the plan of subsequent research work.

1. INTRODUCTION

1.1 As the first step in executing air traffic control instructions, speech errors can cause potential conflicts and affect the safety of operation. ATMAS introduces speech recognition technology, which can interpret the correct pilot's instructions for controllers and verify whether the pilot correctly understands and executes the instructions.

1.2 Procedures for Air Navigation Services》(ICAO) and 《Annex 10 to the Convention on International Civil Aviation Aeronautical Telecommunications》 clearly define the communication rules for air-to-ground calls, providing a standardized framework for speech recognition technology. In addition, ICAO Annex 11 《Air Traffic Services》 requires standardized expressions of control instructions, which serve as an important basis for the design of speech recognition algorithms.

1.3 The MALORCA project (Machine Learning of Speech Recognition Models for Controller Assistance, Number: PJ.16-04) is a key component of the European SESAR 2020 research and innovation programme, A universal and low-cost solution was designed, establishing ontological standards for voice command transcription to build basic automatic speech recognition (ASR) systems in Prague and Vienna.

1.4 In 2020, The Air Traffic Management Bureau of the Civil Aviation Administration of China conducted the "Research on Dynamic Monitoring and Error Correction System for Aircraft Approach and Landing on Closely Spaced Runways," which provides studies on voice command versus flight path/voice command conflicts and voice command versus flight path/flight plan inconsistency alerts, aiming to enhance the safety levels of aircraft approach and runway operations.

1.5 The speech recognition technology has been applied in relevant projects at Beijing Capital International Airport and the Xinjiang Air Traffic Management Bureau from 2021 to 2024.

2. DISCUSSION

Technical Architecture

2.1 Speech recognition technology in ATMAS applications primarily consists of per-processing model, language model, and intent model. As shown in Figure 1.

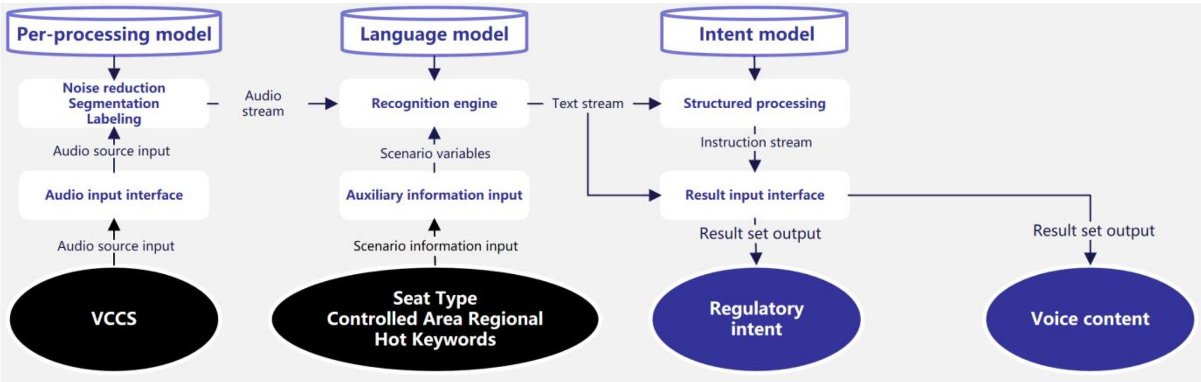

Figure 1: Technical Architecture of Speech Recognition in the ATMAS

2.2 Connect with the Voice Communication System (VCS) server to obtain air traffic control speech data through methods such as port mirroring.

2.3 The per-processing model performs noise reduction, endpoint detection, and segmentation on the speech data, then sends it to the language model in the form of an audio stream. When the language model incorporates information such as hot-words, scenarios, and auxiliary data for combined processing, neural network-based end-to-end recognition engines are responsible for performing speech recognition and converting the audio into text sequences.

2.4 The intent model automatically determines speaker identity (pilot or air traffic controller) through semantic analysis of the linguistic content by the language model, identifies structured control instruction intents, and ultimately outputs the speech content.

Deployment and HMI display

2.5 Deploy an independent server in the ATMAS to achieve physical isolation between the speech recognition application functionality and other core ATMAS functions, ensuring the reliability of ATMAS core operations. Conduct intent calculation and matching processes using three categories of data: digital voice commands, air traffic control operations, and aircraft status.

2.6 Integrated display in ATMAS human-machine interface (HMI) reduces the number of screens deployed on the control console. Relevant prompts and alerts are provided through flight label colors and voice command windows. The HMI display configuration is shown as follows.

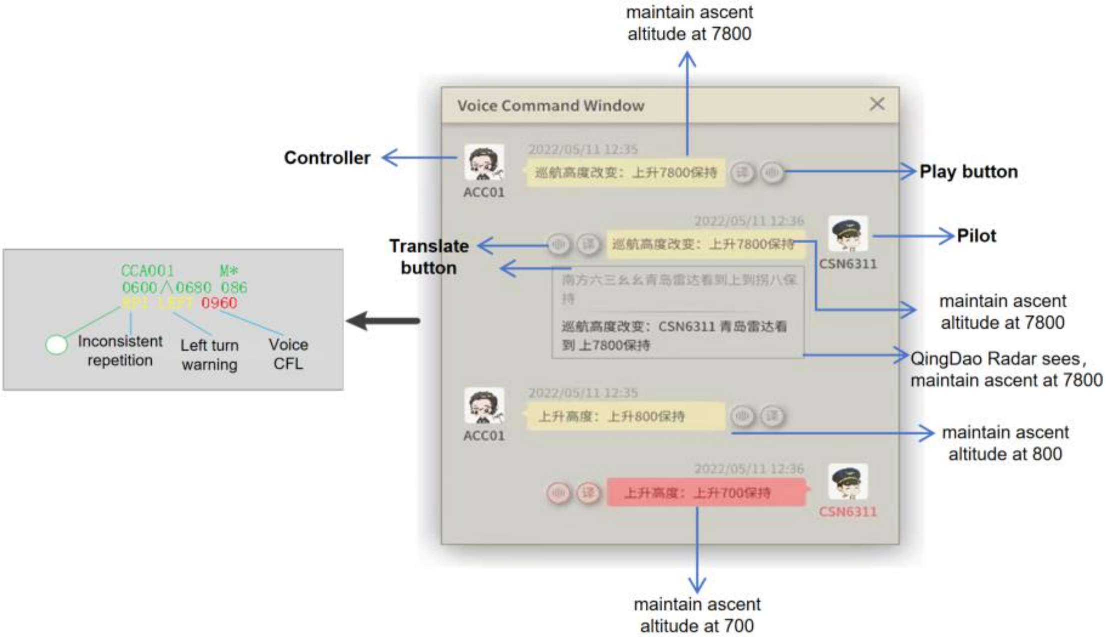

Figure 2: HMI Integrated Display

Alert and other function

2.7 Flight Communication Alert

When the workstation receives a voice message, it will search based on the flight number in the voice message. After locating the corresponding flight, it will trigger a color-changing flashing alert on that flight's identifier and duration of the alert is basis on offline VSP parameters. This helps busy sector controllers quickly locate the current communication flight on the HMI.

2.8 Command Check Inconsistency Alert

Extract air traffic control voice commands from ACC or APP control communications, compare them with the crew's read-back instructions, and provide alerts for any discrepancies in air traffic commands to help controllers determine whether pilots have correctly understood and executed the instructions. The main functions include:

a) Speed command discrepancy
b) Altitude command discrepancy
c) Offset and offset cancellation command discrepancy
d) Heading command discrepancy
e) Arrival procedure command discrepancy

Extract surface movement voice commands from Tower Control communication, compare them with crew read-backs, provide alerts for discrepancies in tower control instructions, and assist tower controllers in verifying whether pilots have correctly understood and executed surface movement commands. The main components include:

a) Hold short authorization vs line up and wait instructions discrepancy;
b) Takeoff clearance instructions discrepancy
c) Runway crossing clearance discrepancy
d) Landing clearance instructions discrepancy
e) Runway vacating instructions discrepancy
f) Departure procedure instructions discrepancy

2.9 Incorrect claim notification

Detect read-back content and promptly identify the incorrect claim notification flight numbers. Send that flight numbers and similar flights to the controller. The controller can then more quickly address issues caused by flight crews misidentifying and repeating incorrect flight numbers.

2.10 Extended alarm

In addition to the basic alert functions mentioned above, the system provides the following extended alert capabilities by integrating operational status data and other auxiliary information.

a) Conflict Ere-warning (Altitude/Direction): Identifies the controller's instructed altitude (direction) and the crew's read-back altitude (direction), then calculates potential flight conflicts based on other aircraft's situational data to assist in air traffic control decision-making and prevent conflicts.

b) Runway Incursion Warning: Detects controller intent data (such as landing, takeoff, entering the runway, or crossing the runway) and correlates it with surface surveillance data to proactively prevent runway incursion incidents. Warning is triggered when landing instructions, runway entry instructions, takeoff instructions, or crossing instructions conflict with another flight.

c) Handoff Alert: Extracts the control sector name and frequency value from the controller's instructions and compares them with the offline defined control sector frequency list. If discrepancies are detected, an alert is issued to assist controllers in promptly addressing abnormal handoff situations.

2.11 Other auxiliary applications

a) Voice Playback: The voice playback window displays real-time speech-to-text information of air traffic controllers and pilots, while providing historical playback functionality.

b) Historical Voice Data Retrieval: Controller voice communications are converted to text and recorded after speech recognition. Users can query radio communication text records by date, time, radio frequency channel, flight number, or keywords, with synchronized voice playback and exportable results to facilitate incident investigation.

c) Conducting call quality analysis on speech rate, number of call instruction entries, proportion of valid/invalid expressions, and non-compliant instructions to help controllers promptly correct improper speech habits.

PROGRESS AND PLAN

2.12 In 2021, at Beijing Capital International Airport, speech and semantic recognition technologies were utilized to provide visual display boards, command repetition inconsistency alerts, and conflict warning assistance for apron control, preventing potential conflicts caused by verbal communication. The accuracy of speech-to-text recognition was enhanced through text annotation, achieving an overall character recognition rate exceeding 95%.

Type	Speech data	Marking speech duration	Recognition accuracy rate
Chinese	2692.35 hours	420.12 hours	97.54%

English	849.05 hours	72.68 hours	94.60%

2.13 In 2024, for the Xinjiang Big Data Project, speech recognition technology was utilized to provide management personnel with post-event recitation inconsistency checks, voice call quality analysis (speech rate), and speech saturation statistics. This enabled the measurement of air traffic controllers' workload through speech recognition to facilitate work allocation.

2.14 In 2025, at the North China Air Traffic Management Bureau site, on-site voice and integrated flight track data were connected for installation, deployment, and debugging. Different manufacturers conducted comparative tests on-site to advance the trial operation of speech recognition technology in the field.

PROSPECTS

2.15 Currently, voice recognition systems in air traffic control face challenges such as noise sensitivity (e.g., VHF communication interference), rapid speech commands, and code-switching between Chinese and English, which still require optimization. The on-going efforts in China's civil aviation voice recognition technology for ATMAS applications include:

2.16 Data augmentation and noise suppression to enhance model robustness. Presently, noise reduction can be applied to voice data prior to recognition processing, resulting in a nearly 50% improvement in word error rate (WER) after denouncing.

Duration of data samples	WER （before noise reduction）	WER （after noise reduction）
20 minutes	3.3%	1.8%
100 hours	4%	2.0%
960 hours	8.2%	4.8%

2.17 High-quality datasets and model training improve the accuracy of the model. Currently, a total of 1,700 hours of annotation have been completed, and the introduction of large models for training is being considered to enhance recognition accuracy and reduce computational costs.

3. ACTION BY THE MEETING

3.1 The meeting is invited to:

a) note the information contained in this paper; and
b) discuss any relevant matter as appropriate

– – – – – – – – – – – – –