Design and Practice of Meteorological New Media Broadcasting System Based on Dual Model Speech Cloning Architecture

Riyuan HAO, Kun MA*, Jianyong LIU, Xue HU, Haijiang ZHAO, Xinglu LI

Zhangjiakou Meteorological Bureau, Zhangjiakou 075000, China

Abstract In response to practical issues such as equipment dependence, personnel scheduling risks, and disaster warning timeliness in grassroots meteorological program production, a meteorological voice broadcast auxiliary process of online and offline dual model switching is proposed in this paper. The process integrates three core functions: message parsing, model switching, and speech generation. At the same time, it is connected to the cloud volcano engine speech replication large model and the local Index-TS 2.0 open source model. Operators can manually switch as needed. The system is equipped with an updatable mandatory dictionary of high-frequency meteorological words to improve the accuracy of pronunciation of terms and place names. Subsequent tests have shown that the system can serve as a backup solution to avoid downtime in the event of equipment failure or broadcaster absence, compress the production cycle of sudden warnings to the minute level, and support mass production of multiple versions of programs. This paper provides a solution for emergency voice program synthesis for meteorological new media by complementing cloud high precision and local high reliability.

Key words Meteorological new media; Speech cloning; Message parsing; Volcanic engine; Index-TTS 2.0

DOI 10.19547/j.issn2152 - 3940.2026.04.004

With the high-quality development of social economy and the rapid change of science and technology, the public tends to develop in a diversified way in the communication mode[1]. Meteorological new media (WeChat official account, Tiktok, Kwai, *etc.*) have become an important channel for the public to obtain weather information[2]. However, traditional meteorological program production heavily relies on recording equipment and professional broadcasters, with the following prominent problems: sudden malfunctions in microphones, sound cards, computers, or recording software that prevent the program from being recorded on time; the audio workstation or plugin is unstable, and the production process is interrupted; broadcasters may experience recording delays due to work arrangements, physical health, or other tasks, which may cause delays in the broadcast of the program; in case of sudden rainstorm, typhoon and other disasters, the whole process from writing, contacting the announcer to entering the studio to recording, editing and reviewing takes a long time, which is difficult to meet the needs of rapid early warning; due to limited manpower, it is unable to support the production of content with multiple weather manifestations and higher frequencies[3].

In recent years, significant progress has been made in AI speech cloning technology. In the field of new media, multiple industries have explored the integration and application of large-scale language models and AI speech cloning technology in radio program production. Through cloud and local programs and speech models, the automation process of program manuscript generation and speech synthesis has been achieved, significantly improving program production efficiency[4]. Lin Yichao *et al.* proposed a cross language film production process based on AI voice cloning using GPT-SoVITS as an example, and verified the feasibility of voice cloning in professional media content production[5]. Ma Yaohong elaborated on the technical principles, ethical risks, and response strategies of sound cloning based on practical experience in the field of broadcasting and television[6]. The above research has laid the technical and implementation foundation for the application of speech cloning in the field of meteorological new media.

However, the existing research mostly focuses on the performance optimization of a single model or fully automatic process design. In view of the requirements of "equipment failure", "emergency rapid response" and "controllable switching" in meteorological scenarios, there is still a lack of engineering solutions that combine online high quality and offline high reliability and are easy for operation and maintenance personnel to operate. This paper is based on the current research and development achievements in the field of AI, and the core ideas of voice cloning and efficiency improvement. Combined with the characteristics of meteorological business, a meteorological voice broadcast auxiliary process based on message parsing and supporting online and offline dual model switching is proposed. The purpose is to discuss three practical issues: continuity, timeliness, and output.

Received: June 1, 2026 Accepted: July 9, 2026

* Corresponding author.

1 Function design and expected effects of the system

1.1 Function design Message parsing: it could obtain the latest message from the meteorological data interface, extract key meteorological elements (temperature, wind speed, city name, *etc.*) with one click, and convert them into template data for operators to use.

Manual switching of speech model: the online volcano engine speech replication large model and the locally deployed Index-TTS 2.0 speech model are simultaneously connected. Operators can manually select the currently used model based on current network conditions, model availability, cost considerations, and other factors.

Meteorological voice program generation: based on the parsed meteorological information, the required city forecast text can be directly generated. Operators can use templates or manually edit to generate other program required documents, call the selected voice model to synthesize speech, and output audio files that can be used for new media.

1.2 Expected effects Based on the above functions, the system aims to achieve the following effects.

1.2.1 Reducing program delays or shutdowns. when recording equipment malfunctions, software crashes, or the announcer is unable to be present, this system is used to generate voice to avoid broadcasting empty windows.

1.2.2 Shortening the emergency production cycle for sudden weather. In the face of sudden disasters, there is no need to wait for the announcer to enter the shed, and the system can quickly generate warning voices, compressing the production time to minutes.

1.2.3 Improving program output. Utilizing the system's ability to generate high-frequency and batch voice, more versions (such as regional and time-based) of weather programs can be produced without increasing the burden on broadcasters.

2 Key technology implementation

2.1 Meteorological message parsing module The meteorological message parsing module is responsible for parsing the real-time meteorological messages obtained from the data source into initial documents. The system can read professional meteorological messages, correct meteorological data, and convert them into text documents suitable for batch generation of voice.

Analysis process is as below: first, users use the system to obtain the latest messages. Second, the parser identifies key fields in the message, such as site name, observation time, temperature (℃), wind direction and speed (m/s), weather phenomena, *etc.* Third, the meteorological information is corrected according to the correction procedure, and is converted into a text document file of urban forecast, which is verified by the operators.

Based on this information, operators can directly use built-in templates to generate broadcast documents, or manually modify or rewrite them in text files. The flexibility of manual intervention is retained.

Just having a model is not enough, AI needs to understand the jargon. There are many easily mispronounced words in meteorological terminology. For example, the word "turn" in the "rainstorm turning to heavy rain" is easy to read by general AI too fast without pause. Some local place names have a lot of polyphonic characters. A "mandatory dictionary" of high-frequency meteorological words has been established, which can be continuously improved in the future. This step is crucial as it lays the foundation for AI to transition from "reading correctly" to "reading well".

2.2 Dual model speech model access

2.2.1 Online model: volcano engine speech reproduction large model. Volcano engine voice reproduction model is a commercial voice cloning service under ByteDance, which has the advantages of high sound quality, rich emotion, multilingual support, *etc.* Through API interface debugging, it supports fine control of speech speed and emotions (such as urgency and relief), ensuring that the generated speech has a highly realistic listening experience[7].

Access method: clean voice samples of the target broadcaster are uploaded in advance on the volcano engine console (recommended for more than 30 s), to complete tone reproduction, and obtain tone ID. The system encapsulates API request functions, inputs generated text documents, tone IDs, speech rates, emotions, and other parameters, and returns to synthesize audio files (WAV/MP3). Applicable scenarios: daily program production, good network conditions, pursuit of high-quality broadcasting effects.

2.2.2 Offline model: Index-TTS 2.0 local deployment. Index-TTS 2.0 is an open-source end-to-end speech synthesis model that supports localized deployment. Based on the Transformer architecture, it not only supports high-quality zero sample speech synthesis, but also significantly improves the realism and expressiveness of speech in the emotional expression dimension. It can perform real-time inference on consumer grade GPUs. The system deploys it on the intranet server and does not rely on the Internet[8].

Deployment steps: install the Index-TTS 2.0 framework and its dependencies (PyTorch, SoundFile, *etc.*); collect the voice of the target broadcaster for training, and train the tone of the broadcaster; start the local inference service, receive text requests through the HTTP interface, and return audio. Applicable scenarios: when the external network is interrupted, the volcano engine service is unavailable, or there are special requirements for data security. The operator can switch to the local engine by clicking the "offline model" button.

This system is positioned as an auxiliary tool for meteorological new media production, and does not replace all aspects of the existing production process. Instead, it provides speech synthesis capabilities at key nodes to meet broadcasting needs in scenarios such as equipment failures, personnel conflicts, and sudden

weather events.

3 Application scenarios

3.1 Recording equipment malfunction or software crash During the recording process of regular weather programs, hardware failures such as microphone, sound card, and mixing console may occur, or software crashes such as audio workstations and plugins may occur. At this point, the recording work cannot continue, and it will result in program delays or suspensions if waiting for repairs or software reinstallation. This system can serve as a backup tool: operators can use cloned announcer voice models to quickly generate current weather voice programs through the system, replacing traditional recording.

3.2 Announcer's schedule conflicts or inability to attend Due to the broadcaster's temporary sick leave, business trip, urgent tasks, *etc.*, they are unable to enter the studio for recording on time. This system allows operators to generate program speech using a trained speech model without relying on the presence of a broadcaster, ensuring that the broadcast is not affected by personnel factors.

3.3 Emergency report on sudden weather When the meteorological observatory issues rainstorm, lightning, gale and other sudden disaster warnings, the traditional process requires temporary writing, recording, editing and review, which takes a long time. This system can respond quickly: operators can directly parse the latest messages, edit short warning documents, select available models (use online volcano engine when the network is normal, offline Index-TTS 2.0 when the network is interrupted or it needs local processing), and complete speech synthesis and export within minutes, greatly reducing the production cycle of emergency information.

3.4 High-frequency multi-version program production In addition to regular daily programs, additional content such as zone forecasts, agricultural meteorological reports, and tourism meteorological alerts should be added. Operators can use this system to generate these voice programs in bulk during off peak hours, increasing program output without increasing labor costs.

3.5 Workflow The typical process for operators to generate meteorological voice programs using the system is as follows (Fig. 1).

3.5.1 Reading message. Entering the latest message document, and clicking text generation, the system automatically fills in meteorological data, and a document for each city is generated.

3.5.2 Editing the document. The broadcast content is confirmed or modified in the text document.

3.5.3 Model selection. The desired speech model is selected based on the current situation.

3.5.4 Generating speech. Selecting the desired broadcaster tone that has been trained, audio for all city forecasts is batch generated. Then, it clicks on "synthesize all". The time interval requirement between two city forecasts is met by generating blank audio and adding it to the middle of the city forecast audio. Finally, clicking on "synthesize all", a complete city forecast audio is stitched together.

3.5.5 Trial listening and export. Playing the forecast audio of each city, wrong audio can be generated separately to save time in speech synthesis. After confirming that there are no errors, the audio file is exported for new media release.

The entire process usually takes no more than 10 min. For repetitive programs, the program can be edited to achieve one-click generation. This system retains manual intervention in document generation and model selection, making it more suitable for grassroots meteorological departments with high requirements for content accuracy and controllability.

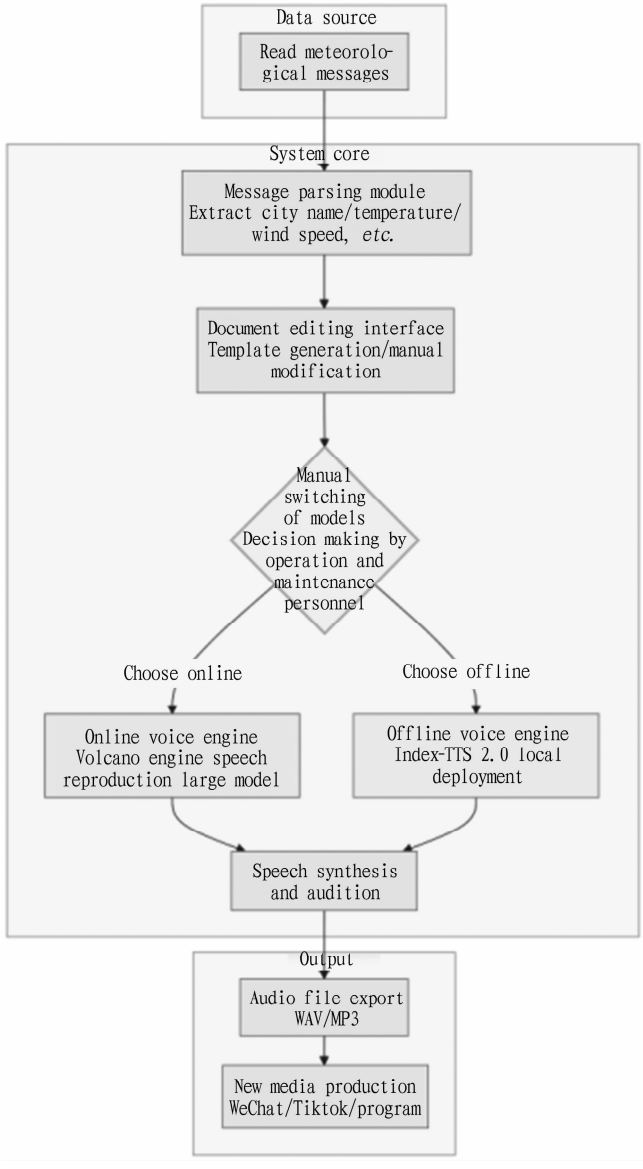

Fig. 1 System architecture diagram

In practical use, operators decide which speech model to choose based on the following situations (Table 1).

Table 1 Selection of speech model

Situation	Recommended model	Reason
Daily programs and smooth internet connectivity	Online (volcano engine)	Higher sound quality and richer emotional expression
Volcano engine API flow restriction or return error	Offline (Index-TTS 2.0)	Locally available, not restricted by cloud
External network interruption	Offline (Index-TTS 2.0)	Independent of the Internet
Sensitive data is involved and should not be uploaded to the cloud	Offline (Index-TTS 2.0)	Local processing of data throughout the entire process
Rapid emergency response is required, and the response delay of online models is high	Offline (Index-TTS 2.0)	Local inference has lower and more stable latency

4 Discussion and prospect

Although this system solves the risk of suspension caused by equipment failures and personnel conflicts, it still has the following limitations: model switching relies on manual labor, and operators need to constantly pay attention to the availability of online models. In the future, simple health check prompts (such as displaying "online API delay too high") can be added to assist manual decision-making, but the final choice is still retained. The manuscript still requires manual editing. For complex weather processes, operators need to have certain meteorological knowledge. In terms of emotional expression, it is relatively limited, and the current speech model lacks the sense of urgency when broadcasting disaster warnings compared to real people. Operators need to manually select appropriate emotional labels based on the warning level.

In the future, it should focus on the following areas: upgrading model capabilities, continuously tracking updates to open-source models such as Index-TTS, and enhancing the naturalness and emotional expression of offline models; consider adding a large model interface for writing documents in the future to improve document generation efficiency; template library construction, accumulating broadcast templates for common weather types, further reducing document editing time[9-10].

5 Conclusions

The dual model speech cloning solution based on "volcano engine + Index-TTS" proposed in this paper aims to solve the business problems of high cost, complicated process, and high risk in meteorological program production through technical means. This solution not only achieves cost reduction and efficiency improvement in meteorological broadcasting, but also provides solid technical support for the digital transformation of meteorological media through the complementary advantages of "cloud high precision" and "local high reliability". In the future, with the continuous iteration of the model and the introduction of manuscript models, this system is expected to further improve the level of meteorological broadcasting and promote the development of meteorological science popularization.

References

[1] YAN N, YING C. Discussion on the advantages of new media short videos in meteorological service work[J]. Meteorology Journal of Inner Mongolia, 2023(4): 25-28.

[2] WANG RJ. Research on meteorological short video production and communication operation in new media era[J]. Journal of News Research, 2023,14(8): 71-73.

[3] MA SR, LIU JX. Reflections on improving the service ability and level of meteorological film and television in Yanbian Prefecture[J]. Journal of Agricultural Catastrophology, 2022, 12(2): 59-61.

[4] ZHANG XM, MA XW, WU JT. Full-chain transformation of radio program production enabled by AI[J]. Audio Engineering, 2025, 49(12): 124-126, 133.

[5] LIN YC, WANG P, ZHENG Y. Exploration and research on cross-language application of AI voice cloning technology in films: Taking GPT-SoVITS as an example[J]. Advanced Motion Picture Technology, 2025(6): 59-65.

[6] MA YH. Implementation and application of sound cloning technology in the field of broadcasting and television[J]. Radio & Television Information, 2025, 32(12): 37-39.

[7] Volcano Engine. Voice reproducing large model product document[EB/OL]. 2025.

[8] ZHOU SY, ZHOU YQ, HE Y, *et al.* IndexTTS2: An autoregressive zero sample text to speech model with breakthrough progress in emotional expression and duration control[J]. arXiv, 2025(2506.21619): 1-9.

[9] LIU C, LIU QY, LIANG LN, *et al.* Exploration and application of automatic processing system for meteorological service short video[J]. Chinese Informatization, 2023(9): 67-68.

[10] ZHANG XM. Specific practice of empowering broadcasting program production with AI technology[J]. Modern Audio-Video Arts, 2025(2): 82-84.