A Brief Analysis of Intelligent Voice Technology for China's Mainstream Media Content Production and Distribution Implications
—— Take the Application of IFLYTEK Series Products in Mainstream Media as an Example

B X Song†,*,a
1Communication University of China, No.1 Dingfuzhuang East Street, Chaoyang District, Beijing, China
*a,song2018@cuc.edu.cn

Abstract. Intelligent speech technology including speech recognition (ASR) and speech synthesis technology (TTS), it is in China's mainstream media plays an important role in technological innovation, appropriate use of this technology can improve the efficiency of the mainstream media content production and distribution of precision, while fully give play to the role of the mainstream media propaganda, to give the mainstream media fresh vitality, so as to promote China's mainstream media communication effect. IFLYTEK, as a leader in the field of intelligent voice technology, has close cooperative relations with many mainstream media. By analyzing the application of IFLYTEK series products in mainstream media from the perspective of technology, we can better understand the impact of intelligent voice technology on the production and distribution of mainstream media content in China.

1.Introduction
In the current era, with the rapid development of media technology, a huge amount of information fills people's lives. Mainstream media often have a very wide coverage of the audience, assume an important propaganda role, and have the ability to influence public opinion. To play its role, mainstream media often need to keep up with The Times and innovate in the production and distribution of content.
Among them, artificial intelligence Figure 1, as the focus of technological innovation, largely helps the mainstream media to complete its functions under the background of The Times. Intelligent speech technology includes speech recognition technology (ASR) and speech synthesis technology (TTS). In media applications, it mainly plays the function of automatically recognizing speech content to form speech or synthetic voice Figure 2. As a kind of AI technology, although it is not as popular as intelligent anchor, intelligent recommendation and other functions, it still affects the production and distribution of mainstream media content with its unique functions.

2. The influence of intelligent voice technology on the content production of Mainstream Media in China

In the production process of mainstream media content, speech recognition technology is mainly able to transfer audio files and generate text manuscripts, which reduces labor demand and improves production efficiency. Speech synthesis technology can imitate all kinds of human voices, which enriches the types and forms of mainstream media content and improves the audience's experience Figure 3.

2.1. Transfer audio files to improve production efficiency

The application of voice recognition technology (ASR) in mainstream media content production mainly focuses on the identification and conversion of audio media content into text, so as to save labor cost and make it available for more flexible media content production. With the continuous enrichment of media content, this function also began to carry out various forms of innovation, such as synchronous generation of subtitles in live broadcast.

For example, Xunfei Heard M1 Transcribe Assistant (hereinafter referred to as IFLYTEK M1), developed by IFLYTEK, is a product whose main function is to transcribe speech into words. Compared with other similar products, it has the advantages of smaller volume and stronger functionality. In terms of sound quality, IFLYTEK M1 USES 24-bit sound pickup, which makes the sound quality of the received audio data better. More innovative is that IFI heard M1 is equipped with digital dynamic noise reduction technology based on digital signal processing (DSP) and algorithm technology, which can separate the noise through DSP digital noise reduction module and reduce the impact caused by the noise by using anti-phase sound wave [1]. This technique makes the parameters of the filter adjustable and greatly improves the accuracy of the audio information obtained. In appearance, IFLYTEK M1 adopts a circular design and is equipped with a "7+1" microphone array, that is, seven microphones are installed in the setting of the ring, and a microphone is set in the center of the ring. This design enables The Signal flight M1 to capture sound in all directions, avoiding the problem of too large a difference in sound volume due to sound orientation. With the support of IFLYTEK M1, the SMART conference system on THE PC side and the IFLYTEK App on the mobile side, the converted text can be directly edited and converted into Chinese and English.
Intelligent voice transliteration devices such as IFLYTEK M1 use voice recognition technology to improve the productivity of mainstream media content in a number of ways (Figure 4). First of all, such products can save the labor cost in the process of transforming media information from sound to text. The average salary of the current stenographer is 1400 yuan per day and 10,000 yuan per day for the same person [2]. IFLYTEK M1, by contrast, only need 999 yuan can be acquired, which provide mechanical transfer service only 0.33 yuan/minute, on the one hand to save the cost of the mainstream media to hire human, on the other hand, also make the more mainstream media staff can put more energy into the creation of content, rather than voice transfer text in the mechanical work. On the other hand, such products can improve the speed at which voice can be converted to digital, thus enhancing the timeliness of news and increasing the production speed of mainstream media content. For example, at the 2018 China International Intelligent Building Exhibition - Summit Forum, IFLYTEK M1, a voice-to-text system, was used in the forum. In the conference, IFLYTEK M1 can mainly convert the speech of the participants into words, and use the accompanying subtitle production function to make and add subtitles to the live webcast interview, so that the audience watching the webcast can also rely on words to better understand the speech content of the participants. This application in the mainstream media live broadcast enables the audience to grasp the news content more accurately and improves the production efficiency of the mainstream media content.

2.2. Simulate the application of sound imitation, enrich the production content
Speech synthesis technology (TTS) plays a contrary role to speech recognition technology in a certain level. Its application in the production of mainstream media content mainly focuses on synthesizing the media content in the form of text into audio form to broadcast with simulated speech. IFLYTEK's AI anchors, for example, use voice synthesis technology to deliver news. By calling IFLYTEK's SDK package (software development Kit) function, this technology can output the existing text in the form of voice information through background processing [3]. The specific flow chart is shown in the figure below Figure 5.
At present, this technology has been widely used in the SPEECH synthesis of AI anchors and AI voice broadcast of news. The Shenyang evening news, which USES the speech synthesis technology, founded the function of voice "AI headlines", provide the audience with two periods in the morning and evening news content of speech, speech need saving labor costs, facilitate faster to users present media content, makes the production of media content more rich, vivid, ordinary text into speech information, enrich the dimensions of the mainstream media content from the auditory, users are easier to accept the media content.

The another advantage of this technology is that it can broaden the scope of the mainstream media audience, also take IFLYTEK products as an example, it introduced the "reading listening solution" can be customized for clients of speech synthesis and sound scheme, users can customize free AI the content of the speech and voice types, choose, such as "XiaoPing", "XiaoWei" and "Ma Shu" voice of AI broadcasters, and through voice to interact with them. This function has been applied to many mainstream media, such as People's Daily and Xinhuanet.com, which enables users who are busy with work or have visual impairment to receive mainstream media content without barriers and become a member of their target groups.

3. The influence of intelligent voice technology on content distribution of Mainstream Media in China

In the current development of intelligent media, the speech recognition technology and speech synthesis technology covered by intelligent speech technology are mainly used in the production of media content in mainstream media. Although less mentioned, this technology can also be applied to the content distribution process of mainstream media in China.

Taking IFLYTEK products as an example, AIUI is a set of voice-centric human-computer interaction solutions launched by IFLYTEK in 2015 Figure 6, which aims to enable applications and devices to quickly acquire the ability to listen, speak, understand and think [4]. Multiple platforms including Android, iOS, Windows can access AIUI through its built-in acoustic VAD (voice activity monitoring), can intercept from the voice of the received signal flow before and after the user's voice endpoint, in order to improve the speech recognition technology to user natural speech recognition accuracy, so as to more accurately meet the needs of users. In the Android system with high degree of freedom, AIUI can be deeply integrated with the system to achieve linkage with apps of various mainstream media, so that media content meeting users' requirements can be accurately distributed to the audience, and the accuracy and efficiency of content distribution of mainstream media can be improved.

4. The defects and development prospect of intelligent speech technology

At present, the development of intelligent voice technology is still not perfect, so it is unrealistic to entrust the mainstream media content production with the whole voice transfiguration into text. China
is a country with a vast territory and numerous ethnic groups. Different ethnic groups and regions often have different dialects, and some dialects also have high similarity. Under such conditions, it is still difficult for ARTIFICIAL intelligence to distinguish different dialects.

The function of "IFLYTEK M1 transcribe Assistant" has been introduced in the previous article. Although it has reached a very advanced level in the application of speech recognition technology, it is still difficult to overcome the recognition problem of different dialects. For example, IFLYTEK M1 provides 0.33 yuan/min of intelligent audio to text function, but also provides 80 yuan/min of high precision manual transliteration service, which often deals with the situation that the machine cannot recognize the dialect. However, further identification by humans also consumes human capital and reduces the productivity of mainstream media content. According to the principle of speech recognition, this paper proposes a possible solution to this problem, that is, to improve the learning ability of artificial intelligence. At present, the main way for speech recognition artificial intelligence to distinguish languages is to recognize the features of speech, words and grammar in audio data, extract the characteristic parameters of speech signals, such as LPC predictive coding parameters, and match them with the data in the database [5]. This kind of speech recognition method has the defects of relatively rigid and relying on the richness of database. Meanwhile, it is easy to make mistakes when facing similar dialect.

Back Propagation neural network is based on Back Propagation algorithm Figure 7. It is a common method used to train artificial neural networks in combination with optimization methods (such as gradient descent). [6] The operation principle of the neural network can be briefly summarized as: input value and expected known results, calculated by artificial intelligence output, and compared with the difference between the known results and input value, so as to train the artificial neural network. The application of this technology in speech recognition, especially in language and dialect recognition, can gradually improve the accuracy of ARTIFICIAL intelligence in distinguishing different languages or dialects through continuous learning, and finally reach an ideal level. However, BP neural network also has its disadvantages, such as slow learning speed and large amount of training. In the process of the development of artificial intelligence, there are still many problems to be solved. The continuous development of artificial intelligence can often further integrate artificial technology into the application of media, especially mainstream media, so that the mainstream media can better play its role in communication and provide people with more rich media content.

5. Conclusion
As a part of AI technology, intelligent voice technology can be deeply integrated with mainstream media in the field of content production and distribution at multiple levels, thus having a beneficial impact on the production and distribution of mainstream media content. [7] Among them, speech recognition technology (ASR) mainly focuses on transforming speech into text to improve the production efficiency of content. It can also transform text into other forms of content and play a role that cannot be achieved manually in live broadcast and other occasions. Its recognition of natural
speech also makes the distribution of media content more precise and better meets the needs of users. [8] Speech synthesis technology (TTS) is widely used in AI anchors, AI readers and other fields, and plays a role in enriching the form of mainstream media content, making it vivid and broadening the audience scope. At present, there are still some problems in intelligent voice technology, such as insufficient recognition ability and unnatural sound synthesis. However, with the improvement of hardware and algorithm, intelligent voice technology is also developing along with the needs of the audience. Looking to the future, intelligent voice technology will have deeper connections with mainstream media in more fields, enabling mainstream media to play a further role.

References
[1] Qi Zhang, Shitu Luo, Guofu Liu 2000 China. J. Electroacoustic Technology: Principle of dynamic noise reduction and its Application in speech Recognition System. 07 21-23
[2] Yongrun Zhu, 2018 China. J. China Exhibition: What you hear is what you see: speech recognition transcription system. 11 66-67
[3] Lin Zhang 2019 China. J. Computer and digital engineering: Design of intelligent voice chat robot in virtual environment. 10 2617-2621
[4] Xi Wang 2019 China. J. Innovation and application of science and technology: New development: design mode of intelligent voice interaction. 29 35-36
[5] Qi Xu 2019 China. J. Research on publication and distribution: How to integrate American traditional media into intelligent voice ecology. 09 63-67
[6] Huaping Xie 2019 China. J. media: New media quality content production and distribution strategy. 20 43-45
[7] Feng Pan 2019 China. J. Electronic design engineering: Design and research of intelligent voice control system. 14 6-9
[8] Hanati Reyizha, Nurbuil 2019 China. J. Modern electronic technology: Quantitative analysis of the research status and development trend of speech synthesis technology. 124 116-119