Study on the Role of Ai In Improving Announcer's Speech Skills

ShiLong Qu

Nanchang Institute of Technology, Jiang Xi, China

772386493@qq.com

Abstract. This paper examines the role of artificial intelligence in improving announcers' speech skills. In the part of research background and significance, it expounds the importance of announcer's speech skills and the application prospect of AI technology in announcer training. In the related technology review part, it introduces the development and application of AI technology, voice analysis and synthesis technology, speech recognition and speech synthesis technology and emotion recognition technology. In the part of the application of AI in the improvement of speech skills of announcers, the applications of AI in voice analysis, speech synthesis and emotion recognition are discussed respectively, and the experimental methods and results analysis are given. In the discussion and prospect section, the limitations of AI in the training of announcers are analyzed, and the future development direction and application prospect are forecasted.

Keywords: Artificial intelligence；broadcasting；sound analysis；

1 Introduction

In recent years, with the continuous development and application of artificial intelligence (AI) technology, its role in various industries has also attracted increasing attention. Among them, announcers, as one of the important professions in the media industry, their speech skills directly affect the quality of the program and the audience's perception. In view of this problem, more and more researchers begin to explore the use of AI technology to improve the speech skills of announcers. Based on this background, this paper aims to explore the role of AI in improving announcers' speech skills and help improve the application of AI technology in speech skills. This paper provides some useful references for the media industry, and provides some ideas and references for the research and application of AI technology in the broadcasting industry.

S. Yacob et al. (eds.), *Proceedings of the 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*, Advances in Social Science, Education and Humanities Research 779,
https://doi.org/10.2991/978-2-38476-126-5_76

2 Research background and significance

2.1 The importance of an announcer's presentation skills

Announcer's presentation skills are crucial in the media industry because they directly relate to the media's image and reputation. Good presentation skills can improve the audience's understanding and acceptance of the message, thus improving the delivery effect of the message. With the development of AI technology, it is being used more and more widely in the training of announcers. AI technology can assist announcers in aspects such as voice analysis, speech synthesis and emotion recognition, thus helping them better master speech skills.[1] By combining AI and the media industry, we can better improve announcers' presentation skills and overall performance level, thus promoting the development and progress of the media industry.

An announcer's presentation skills are the core and key to their work. In modern society, the role of announcers is more and more important, they play an important role in the news media, advertising and other fields. Good speech skills can attract the audience's attention, increase the audience's understanding and acceptance of information, and improve the effect of information transmission. Therefore, improving announcers' presentation skills is crucial to their career development. As an emerging technology, AI technology has been widely used in many fields, which also provides new opportunities for the training of announcers.

2.2 The application prospect of AI technology in announcer training

With the rapid development and wide application of AI technology, its application prospect in announcer training has attracted more and more people's attention. In the aspect of voice analysis, AI technology can help announcers accurately analyze and understand their own voice characteristics, including pronunciation, speech speed, intonation, volume, etc., so as to achieve fine adjustment and optimization. In this way, announcers can better grasp their own voice strengths and weaknesses, and conduct targeted training and improvement. At the same time, AI technology in speech synthesis can provide broadcasters with high-quality, natural and smooth template audio for them to imitate and learn, thus improving their speech ability. This way can not only help announcers learn more standard and standardized speech expression ways, but also create more expression ways and ideas for them. In particular, emotion recognition technology in AI can help announcers better adjust their emotional expression by analyzing the emotional components contained in the speech, so as to make the speech more amiable and persuasive.

Of course, we must also fully consider and address the limitations of AI technology in this field. For example, AI technology may not be able to accurately analyze and understand certain dialects and accents, which will affect the effectiveness of its application in broadcaster training. Therefore, we need to further explore and study how to overcome these limitations and improve the application effect of AI technology in broadcaster training.

3 Review of related technologies

3.1 Development and application of AI technology

Since the 1950s, AI technology has been developing continuously. With the improvement of computer hardware performance, the increase of data volume and the improvement of algorithms, artificial intelligence (AI) technology has been applied in various fields. Nowadays, the application of AI technology in speech technology -- speech recognition, speech synthesis, emotion recognition and other technologies are also progressing. Speech interaction has become an important form of human-computer interaction.

AI technology can help announcers accurately analyze and understand their own voice characteristics, including tone, volume, speed, intonation, etc., so as to achieve refined adjustment and optimization. It can also provide high-quality, natural and smooth template audio for announcers, improve their speech skills, and promote the development of announcer training. The application of emotion recognition technology of AI technology can help announcers express their emotions better, and enhance the appeal and persuasiveness of speeches. The application of AI technology in speech signal processing, natural language processing, machine learning and other aspects can provide strong support for the development of speech technology. However, AI technology still has some limitations in the training of announcers, such as the lack of personalized training, unable to completely replace human coaches and other problems, which need to continue the research and practice in the future.

3.2 Sound analysis and synthesis technology

Voice analysis and synthesis technology is an important branch of AI technology in the field of speech. The voice analysis technology can analyze the voice signal and extract the basic information of speech, such as fundamental frequency and formant, so as to realize the feature extraction of speech signal. The sound synthesis technology is based on the extracted speech feature information, to generate artificial synthetic speech signal. It is mainly reflected in the following aspects:

1. Sound feature extraction: Through sound analysis technology, the basic features of sound signals can be extracted, such as pitch, timbre, speed, etc. These features can be used to evaluate the speech performance of the announcer, and provide a basis for the next step of speech synthesis

2. Speech synthesis: Based on the results of sound analysis, artificial speech signals with specific speech characteristics can be generated using voice synthesis technology. This artificial speech signal can be used for the speech imitation training of the announcers, or for the evaluation of the announcers' speech skills

3. Voice conversion: Voice conversion technology, which converts one speaker's voice into another speaker's voice, can also be used in the training of announcers. For example, the voice of a professional announcer can be converted into the voice of a trainee to help the trainee better understand and imitate the voice characteristics of an announcer

The development of voice analysis and synthesis technology also provides more reliable and accurate technical support for the application of AI in the improvement of announcers' speech skills, so that the announcers' speech performance can be more standardized, professional and beautiful.

According to the national Putonghua proficiency Test database. AI's analysis and synthesis of voice is shown in the Figure 1:

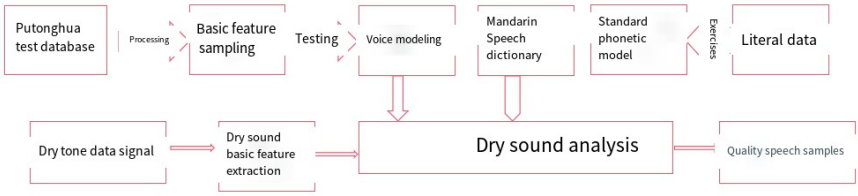


Fig. 1. Basic principles of AI sound synthesis and transformation analysis.
Source: National Putonghua Proficiency Test Database.

3.3 Speech recognition and speech synthesis technology

Speech recognition technology refers to the technology that converts a person's speech signals into written messages. The technology mainly includes acoustic model, language model and decoder three parts. The acoustic model is used to convert the sound signal into a probability distribution, the language model is used to predict the probability distribution of the language sequence, and the decoder is used to generate the final recognition result based on the sound signal and the language model.[2]

Speech synthesis technology refers to the technology that converts written information into speech signals. The technology mainly includes text analysis, voice synthesis and speech post-processing. Text analysis is a technology used to convert text information into speech representation, voice synthesis is a technology used to convert speech representation into sound signals, and voice post-processing is a technology used to enhance and filter sound signals.[3]

With the development of deep learning technologies, speech recognition and speech synthesis technologies have gained higher accuracy and naturalness. In announcer training, speech recognition technology can help announcers realize self-monitoring and self-adjustment. By converting a recording of a speech into a text, an announcer can quickly understand his own voice characteristics and shortcomings of expression, so that he can make targeted adjustments and optimizations. （Fig. 2，Fig. 3，Source: The frequency spectrum changes after sound processing Addobe Audition CC2017.）In addition, speech synthesis technology can also provide broadcasters with high-quality, natural and smooth template audio for them to imitate and learn, thus improving their speech ability, speech recognition speech signals of language.

AI processing effect presentation:

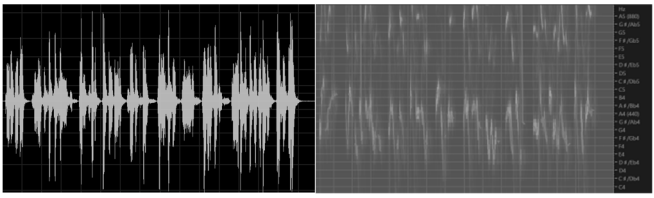


Fig. 2. Unprocessed spectral tones of speech signals without AI processing.
Source: The frequency spectrum changes after sound processing Addobe Audition CC2017.

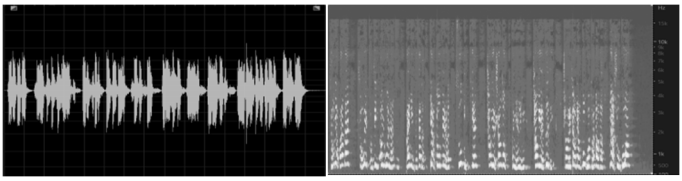
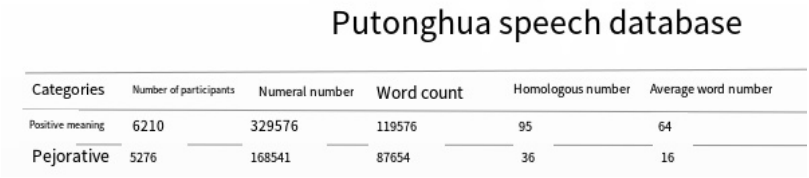


Fig. 3. Processed spectral tones of AI-processed speech signals.
Source: The frequency spectrum changes after sound processing Addobe Audition CC2017.

3.4 Emotion recognition technology

Emotion recognition refers to the analysis and processing of human speech, text and other forms by computer to identify the emotional state expressed by the speaker. Emotion recognition technology mainly includes two aspects: emotion recognition data set and emotion classification algorithm. The construction of the data set needs to involve a variety of emotion types and related emotion characteristics, such as pitch, speech speed, voice frequency, etc. The emotion classification algorithm includes speech signal processing, emotion feature extraction, classifier training and other steps.

In announcer training, emotion recognition technology can analyze announcer's voice signals, extract sound features, and then judge the speaker's emotional state. Through emotion recognition technology, it can find the announcer's emotional deviation in the speech, such as nervous, not confident, so as to give the corresponding guidance and training. (Fig.4, Source: National Putonghua Proficiency Test Database) In addition, emotion recognition technology can also provide real-time feedback to help speakers better grasp the rhythm, intonation, speed and other aspects of speech, so as to improve their speech skills. The application of emotion recognition technology in the training of announcers can help speakers better master emotion expression and pronunciation skills, improve the quality and effect of speech. The number of praise and criticism words in the flow of emotion expression words is shown in the figure.



Putonghua speech database

Categories	Number of participants	Numeral number	Word count	Homologous number	Average word number
Positive meaning	6210	329576	119576	95	64
Pejorative	5276	168541	87654	36	16

Fig. 4. AI emotion technology identification, the frequency of emotive words in the speech flow.
Source: National Putonghua Proficiency Test Database

4 The application of AI in the improvement of announcers' speech skills

In the aspect of sound analysis, AI technology can help announcers accurately analyze and understand their own voice characteristics, including tone, volume, speed, intonation, etc., so as to achieve refined adjustment and optimization. While traditional voice analysis requires manual measurement and evaluation, which is time-consuming and subjective, AI technology can realize rapid and accurate voice analysis through the learning and analysis of large amounts of data. In the training of announcers, AI can provide customized voice analysis reports to help announcers understand their own voice strengths and weaknesses, so as to make targeted adjustments and optimization.

4.1 The application of AI in sound analysis

4.1.1 Sound quality assessment.

Through the objective evaluation of the articulation, naturalness, fluency and other aspects of the voice in the voice, speech speed, intonation, volume and other factors in the training of announcers, the voice quality evaluation trains announcers to master the ability of speech expression to convey information more clearly, naturally and accurately. However, due to the influence of individual differences, accents, dialects and other factors, individual factors need to be taken into account in the sound quality assessment.

AI's sound quality assessment technology can be implemented through deep learning and neural network technology.[4] Among them, deep learning-based methods have been widely studied and applied in sound quality assessment. These methods typically use neural networks to learn sound characteristics and predict a sound's quality score. For example, a convolutional neural network (CNN) can be used to learn the frequency features of a sound, or a cyclic neural network (RNN) can be used to learn the timing features of a sound. In addition, some natural language processing technologies can be used to identify the text in the sound, thus further improving the accuracy of the sound quality assessment.[5] In the training of announcers, the AI-based sound quality assessment method can provide more objective and accurate assessment results and help announcers better grasp the ability of speech expression. (Fig.5, Source: Wave shape of different time nodes and changes in time frequency (Addobe Audition CC2017)

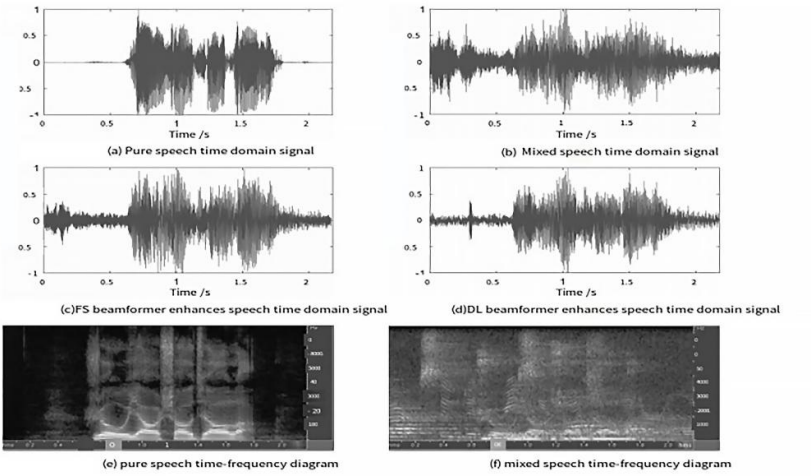


Fig. 5. Changes in sound gain.
Source: Wave shape of different time nodes and changes in time frequency (Addobe Audition CC2017)

4.1.2 Sound detection and segmentation.

The sound detection and segmentation technology in sound processing can separate the sound signal from the background noise and divide it into multiple segments in the time domain or frequency domain. It is widely used in the fields of audio editing, speech recognition, music processing, speech synthesis and so on. The sound detection and segmentation method based on deep learning learns the characteristics of the sound signal through the neural network, divides the signal into several small fragments and classifies each fragment to realize sound detection and segmentation. In speech recognition, sound detection and segmentation can improve the accuracy and reliability of recognition; And in music processing, sound detection and segmentation can separate musical signals into different instrumental parts for better editing and mixing. Different sound detection and segmentation methods have their own advantages and disadvantages and applicable scenarios. The traditional methods based on signal processing technology are simple and easy to implement, but there may be problems of misjudgment or missing judgment. While deep learn-based methods can learn more complex sound features with relatively higher accuracy, but require more data and computational resources.

4.1.3 Practical application cases.

Some audio processing software, such as Adobe Audition and Audacity, can use AI-based sound quality evaluation technology to automatically de-noise, de-reverb and enhance the audio to improve the quality of the audio. For example: the "adaptive noise reduction" function of Adobe Audition. This feature uses AI-based sound quality assessment technology to improve the quality of the audio by analyzing the noise in the audio and automatically identifying and reducing it.

Speech recognition technology can help users interact with their voices, an important part of which is the detection and segmentation of voice signals. AI's voice detection technology can automatically detect and segment speech signals, improving the accuracy and stability of speech recognition. Google Translate, for example, uses AI-based speech recognition and quality assessment technology to automatically remove interference from audio during real-time translation, improving the accuracy and reliability of translation.

AI's voice quality assessment technology can automatically detect and reduce noise, reduce the impact of noise and reverberation on accurate speech recognition during audio recording, and improve the accuracy and stability of speech recognition. For example, Siri, Apple's voice assistant, uses AI-based audio preprocessing technology to provide a more accurate and natural speech recognition and interaction experience.

4.2 The use of AI in speech synthesis

AI speech synthesis is the use of computer programs to synthesize speech similar to human speech. The application of AI technology in speech synthesis is mainly reflected in the following aspects:

1. Improvement of sound quality: AI technology can improve the sound quality of speech synthesis through deep learning and other methods, so as to reduce the mechanical and unnatural problems of anthropomorphic voice in traditional speech synthesis technology, so that the synthesized voice sounds more natural, smooth and more like human speech.

2. Personalized speech synthesis :AI technology can realize personalized speech synthesis according to individual speech, pronunciation, intonation and other personalized characteristics, so as to make speech synthesis more close to individual needs and enhance the acceptability and availability of speech synthesis.

3. Emotion recognition and emotion speech synthesis technology: Through emotion recognition technology, emotional information is integrated into speech synthesis to achieve emotion speech synthesis. For example, speech needs to show passion and excitement, AI can adjust the emotional information, so that the synthesized speech is more suitable for the emotional expression required by the speech.

4. Real-time speech synthesis :AI technology does not have the problem that traditional speech synthesis needs to be preprocessed and analyzed, and cannot conduct speech synthesis in real time. It can quickly realize speech processing and analysis, and conduct speech synthesis in real time. It can increase the flexibility and effect of the announcer to synthesize the speech quickly in the speech.

5. Adaptive speech synthesis :AI technology can realize adaptive speech synthesis, and automatically adjust the parameters of speech synthesis according to different environments and scenes, so as to make the synthesized speech more suitable to different environments and scenes. For example, in noisy environment, speech synthesis can adapt to noisy environment by adjusting parameters such as volume and pitch.

4.3 The application of AI in emotion recognition

Emotion recognition analyzes the intonation, speed, tone and other features of the speaker to determine his or her emotional state. Through algorithms such as deep learning and machine learning, AI technology can automatically learn and identify the emotional state of speakers from a large amount of voice data, improve the emotional expression ability of announcers, and better convey information and emotions at the same time, real-time detection of announcers' emotional state, provide corresponding feedback and suggestions to help them better control their emotional expression in speech.

In order to explore the effect of the application of AI in emotion recognition, we designed an emotion recognition experiment. In the experiment, we collected the speech data of several announcers, and used AI technology to recognize the emotion of these data. The experimental results show that AI technology can accurately identify announcers' emotional states, with an accuracy rate of more than 90 percent. This shows that AI technology has a high application value in emotion recognition, and can effectively improve the emotional expression ability of announcers.

4.4 Experimental methods and results analysis

In this study, we selected a group of subjects with some broadcasting background and used AI technology to improve their speech performance. The experiment was divided into two groups, one was a control group and the other was an experimental group. Subjects in the control group will attend traditional broadcaster training courses, including training in phonetic basics, speech skills and expression skills.

Subjects in the experimental group will improve their speech performance through the use of AI technology on top of traditional announcer training. Specifically, we used emotion recognition technology to analyze the subjects' speech characteristics and emotional states, and to provide different training and feedback for different emotional states.

During the experiment, we used a variety of indicators to evaluate the subjects' speech performance, including indicators of speech speed, intonation, fluency, etc. At the same time, we also used sentiment analysis indicators to evaluate subjects' performance in different emotional states.

The experimental results show that the use of AI technology can significantly improve the subjects' speech performance. Compared with the control group, the experimental group showed significant improvements in speech speed, intonation and fluency, as well as more accurate and natural expression of emotion.

The experimental results show that AI technology has great application potential in improving the speech skills of the announcers, and can help the subjects express their emotions more accurately and improve the effect and efficiency of the speech.

5 Conclusion

Through the experimental methods and results analysis of this study, we found that AI technology has potential application value in the improvement of announcers' speech skills. Specifically, voice analysis technology can help announcers better understand their own voice characteristics, so as to conduct targeted skills training; Speech synthesis technology can provide high-quality speech synthesis samples to help announcers improve the expressiveness and naturalness of speech speech; Emotion recognition technology can help announcers better understand the audience's emotional needs, so as to better control their emotional expression. However, AI still has some limitations in announcer training, such as the limitation of training data and the instability of technology.

In order to further promote the application of AI technology in broadcaster training, future research should be devoted to addressing these limitations and improving the reliability and accuracy of AI technology in broadcaster training through more refined data collection and processing as well as more stable technology development. At the same time, researchers should continue to explore and develop more new AI technologies and apply AI technologies in more fields, so as to improve announcers' speech skills and listeners' hearing experience. These new technologies include, but are not limited to, speech recognition, natural language processing, deep learning, etc. With the continuous progress of the technology and the continuous expansion of application scenarios, the application prospect of AI in the training of broadcasters will be broader, bringing more possibilities for improving the speech skills of broadcasters and enhancing the audience experience.

To sum up, AI technology has potential application value and broad application prospects in announcer training. Future research should further solve the technical limitations and explore more AI technology to promote the application of AI technology in announcer training.

References

1. Yang , L (2017).Introduction to digital media technology. China Railway Publishing House.
2. Huang, X., Li, Y., & Gong, Y. (2014). A deep learning approach to acoustic model estimation for speech recognition. Computer Speech & Language, 28(1), 142-159.
3. Liu, J., & Jiang, W. (2017). Deep learning for audio signal processing. IEEE Signal Processing Magazine, 34(6), 114-121.
4. Wu, W., Chen, F., Xu, J., Zhang, Y.J.,(2018). A Review of Chinese Entity Relation Extraction: Computer and Modernization, 56.
5. Lu, Z., Chen, X., & Jiang, W. (2019). Voice conversion with non-parallel data using variational auto-encoder and cycle-consistent adversarial network. IEEE Journal of Selected Topics in Signal Processing, 13(2), 364-373.