Theory and Practice of Science and Technology
2025, VOL. 6, NO. 6, 146-149
DOI: 10.47297/taposatWSP2633-456929.20250606

Application of AI Technology in Film and Television Audio Production

Lingzi Li, Qingzhong Feng

Communication University of China, Nanjing, Radio and Television College, Nanjing, Jiangsu,210000, China

ABSTRACT
Artificial intelligence technology is causing profound changes in the field of film and television audio production. This study systematically analyzes the application mechanism and influence of AI technology in audio signal processing, content generation, and post-production. At the technical level, deep learning and generative AI provide a new algorithmic framework for extracting and synthesizing audio signal features, significantly enhancing the efficiency and accuracy of sound effects design, speech signal processing, and the audio mixing process. The study reveals the primary challenges faced by current AI audio technology, including strong data dependence, computational efficiency bottlenecks, and limited artistic expression. Therefore, the future development direction should focus on improving the algorithm's generalization ability, optimizing the human-computer cooperation mode, and exploring intelligent production schemes that utilize multi-modal fusion.

KEYWORDS
Generative AI; Film and television audio system; Sound effects design; Speech signal processing

1 Introduction

As a vital component of film and television art, the advancement of audio production technology has a significant impact on the artistic expression and market competitiveness of the work. In recent years, the rapid advancement of artificial intelligence technology has brought unprecedented technological opportunities to the traditional film and television audio production process, while also challenging the existing production mode. From the perspective of technological development, audio processing technology has undergone its first revolution, from analog to digital, and is now moving towards a second revolution led by artificial intelligence. Machine learning methods, represented by deep neural networks, have demonstrated strong feature extraction capabilities in the field of signal processing, while generative artificial intelligence has achieved breakthroughs in content creation [1]. The change in industrial demand is also promoting technological innovation. The rapid development of streaming media platforms has greatly shortened the content production cycle. The traditional audio production mode relying on human experience is difficult to meet the needs of large-scale and high-quality production. In short, the deep integration of artificial intelligence and film and television audio production has become an irreversible trend; however, this process requires careful technical assessment and thoughtful application planning. Only by accurately grasping the technical boundary and establishing a scientific evaluation system can we truly maximise the value of technology, empower creation, and promote the high-quality development of film, television, and audio art.

2 AI technology in the film and audio production technology

2.1 Overview of artificial intelligence audio processing technology

2.1.1 Deep learning and audio signal processing

The application of deep learning technology in the field of audio signal processing marks a major paradigm shift in traditional digital signal processing. Through the nonlinear mapping capabilities of multi-layer neural networks, deep learning offers a novel technical approach for audio feature extraction and pattern recognition. Due to the local receptive field and weight sharing characteristics, convolutional neural networks (CNN) exhibit unique advantages in frequency domain feature extraction, which can effectively capture local correlations in time-frequency domain features, such as the Mel frequency Spectrum. The recurrent neural network (RNN) and its improved architecture, the long short-term memory network (LSTM), are more suitable for processing audio signals with temporal dependencies and exhibit excellent performance in modeling context correlation in speech signals. At the concrete implementation level, the deep learning model can complete various tasks, such as noise suppression, speech enhancement, and music classification, by learning the deep representation of massive audio data [2]. The end-to-end training method significantly reduces the threshold for feature engineering in traditional methods, allowing the system to explore the optimal feature representation autonomously.

2.1.2 The application of generative AI technology in audio production

Generative artificial intelligence opens up new possibilities for the creative production of film, television, and audio, and its core technology is based on the probability generation model. The variational autoencoder (VAE) enables the controllable generation of audio features through latent space modelling, providing a flexible editing dimension for sound effect design. Generative adversarial networks (GAN) can synthesise high-quality natural sound samples through the adversarial training of the discriminator and the generator, which is of great value in the production of sound effects in simulated environments. In the process of film and television production, generative AI technology is mainly used in the following key aspects:①in the aspect of sound synthesis, the hybrid system based on physical modeling and neural network can dynamically generate complex sound effects that meet the needs of the scene;②in the field of speech synthesis, the combination of neuro-vocoder and text-to-speech (TTS) systems enables highly natural speech generation, providing an alternative to character dubbing and voice-over production; ③ the music creation dimension, the conditional generation model can automatically generate music clips adapted to film and television scenes based on emotional parameters.

2.2 The core process of film and television audio production

Film and television audio production is far from a simple recording and synthesis, but a fusion of precision technology and artistic creativity of the system engineering. Its core process is rigorous and orderly, usually divided into three key stages: pre-planning, mid-recording, and post-production. Each stage is linked to the others and plays a decisive role in the auditory sensation and artistic appeal of the final work. The preliminary planning is the foundation period of the whole work. The core work involves formulating a detailed, sound design scheme and completing thorough recording preparation. The sound designer or sound director needs to interpret the script in depth, accurately grasp the director's artistic intent, and conduct a detailed sound needs analysis by scene to establish the sound style, emotional tone, and key sound concepts throughout the film. [3] The primary goal of intermediate recording is to capture high-quality, high-availability raw sound material on set. The simultaneous recording group is the executive body. They need to use professional equipment (such as directional microphones, all-directional microphones, wireless microphone systems, recorders, and mixers) to accurately capture the real sound of the moment in the complex environment of the set. Post-production is the creative process of transforming the original material into the final screen auditory sensation. The sound team (including sound editors, sound engineers, sound mixers, etc.) plays a central role here.

3 The application of AI technology in film and television audio production

3.1 The application of AI in sound effects design and synthesis

Artificial intelligence is fundamentally reconstructing the logical chain of sound effects creation, promoting the field from an experience-driven to an intelligent generation paradigm. The traditional production model, which relies on physical recording and manual editing, is being gradually replaced by a generative model based on deep learning. AI realises intelligent scene construction of environmental sound effects by analysing the physical attributes and perceptual characteristics of sound. The system can understand abstract semantic instructions, automatically synthesize composite sound effects with spatial hierarchy and dynamic changes, and significantly reduce the cost limit of field sampling. In the field of special sound effects, the fusion of generative adversarial networks and physical modelling technology enables the approximation of complex acoustic phenomena, such as explosion shock waves and material fragmentation processes, through virtual simulation, and even the creation of unnatural sound effects that do not exist in reality. In addition, AI has opened up an unprecedented creative dimension for sound designers. Through a parametric control interface, the creator can adjust the material hardness, impact energy, or reverberation characteristics of the sound effect in real-time, and the system generates corresponding variants in real-time.

3.2 The application of AI in speech processing

Artificial intelligence demonstrates significant value in the entire process of film and television speech processing, and its core breakthrough lies in resolving the acoustic dilemma that traditional technology struggles to overcome. Through the deconstruction of human voice features, deep neural networks achieve pure speech extraction in environments with strong noise interference. Even in the face of complex reflections and background noise on the set, they can still effectively separate and enhance the clarity of the dialogue. Aiming to address the inevitable transient defects (such as sonic boom and tooth sound distortion) in recordings, the intelligent repair algorithm can seamlessly reconstruct the damaged waveform and significantly reduce the need for post-recording adjustments. Speech synthesis technology has

entered a high-fidelity stage. A neural vocoder can not only clone the timbre and pronunciation habits of specific actors but also generate intonation changes that conform to the plot tension through emotional parameter adjustment, providing new possibilities for multilingual version production and character voice extension. This ability has also given birth to a new proposition of voice copyright management, prompting the industry to establish a voiceprint confirmation mechanism.

3.3 The application of AI in mixing and post-production

Artificial intelligence is reshaping the decision-making system of audio post-production, transforming the process of audio mixing, which is highly dependent on subjective experience, into an intelligent engineering that can be quantified and optimized. By analysing the acoustic characteristics of massive professional audio mixing works, the system establishes an intelligent benchmark model for multi-track balance, which can automatically complete the level matching and frequency band allocation of dialogue, music, and sound effects tracks. At the same time, the dramatic dynamic range of key elements is preserved. The multi-modal analysis technology, based on computer vision and semantic understanding, enables AI to actively identify the emotional turning point and scene type of the picture, and dynamically adjust parameters such as reverberation time and low-frequency enhancement. This enables the accurate synchronization of sound and picture emotions. In the final master tape processing step, the adaptive algorithm can automatically optimise the overall dynamic curve and frequency response according to the acoustic characteristics of different playback scenes, such as cinemas and streaming media [4].

4 Challenges and future trends of AI technology application

4.1 Limitations of current technologies

4.1.1 Data dependency and lack of generalization

The core problem with current AI audio technology is its high dependence on a specific distribution of training data. In contrast, the complexity and artistic diversity of film and television sound far exceed the generalization boundary of conventional machine learning models. The scarcity of professional-grade audio samples for film and television results in a sharp degradation of model performance in the face of unknown acoustic scenarios. For example, when the automatic noise reduction system encounters an untrained industrial noise spectrum or the generative model needs to synthesise cross-cultural ritual sound effects, it often produces distortion or style mismatch due to data blind spots. Although transfer learning and small sample learning can alleviate some data anxiety, it is difficult to resolve the fundamental contradiction between sound and narrative depth in film and television creation. Algorithms cannot reconstruct the logic of sound that fits the plot's tension based on limited cues, as human designers can. At the same time, existing models of sound"understanding" remain at the level of acoustic feature statistics, which not only fail to capture the subtle variations of sound as a carrier of emotion (such as the implied psychological state of a character reflected in breathing rhythms), but also struggle to adapt to the differentiated aesthetic demands of different directors regarding sound treatment in the same scene.

4.1.2 Computing resources and processing efficiency constraints

The strict requirement for an efficient workflow in the film and television industry is in sharp conflict with the high computational complexity of AI audio processing. In the case of generative models, synthesizing a single piece of high-fidelity ambient sound can take tens of minutes of GPU rendering time, and the hours-long audio volumes of movies make such techniques challenging to incorporate into actual production cycles. The challenge of real-time processing scenarios is even more severe: although the lightweight neural network reduces the theoretical calculation amount through model pruning and quantification techniques. However, a systematic bottleneck remains in memory accumulation, and latency increases in long-term audio stream processing. Cloud computing solutions offer flexibility, but are constrained by the confidentiality requirements of film and television production. Transmitting the original audio to a third-party server involves the risk of copyright leakage, and locally deployed edge computing keeps small and medium-sized studios out due to the high cost of specialised equipment.

4.1.3 The balance between artistry and algorithmic constraints

The essence of film and television sound creation lies in the integration of subjective aesthetic decision-making and technical realisation. The biggest limitation of AI at present is that it cannot bridge the cognitive gap between artistic intention and algorithmic rules. When the director requires a mechanical sound to be a metaphor for the character's

anxiety, or wants the environmental sound to carry a metaphorical auditory sensation during the transition, the existing model can only respond to the surface-level acoustic parameter commands, however, they can not deconstruct the emotional symbols and cultural codes in the narrative language environment. This disconnect is particularly pronounced in music-picture collaborations: AI can match the beat precisely, but it cannot capture the emotional resonance between camera movement and chord progression, nor can it employ advanced techniques for creating psychological tension through dissonance. The reason is that the core goal of algorithm optimisation remains the optimisation of acoustic indicators in a statistical sense (such as signal-to-noise ratio improvement or spectrum smoothness), rather than the creation of artistic value.

4.2 Outlook

Artificial intelligence will promote film and television audio production into a new stage of deep intelligence and creative paradigm reconstruction. The core of its evolution lies in breaking through the attribute of tools and becoming a creative variable that reshapes sound aesthetics and workflow. The breakthrough of adaptive learning architecture is expected to completely deconstruct the dependence of current technology on massive annotated data, allowing the system to capture the director's auditory aesthetic signature with only a few stylised samples. In addition, AI will evolve from an automated tool to a "Cognitive partner" with creative awareness, actively suggesting a dynamic range adjustment strategy that conforms to the narrative rhythm in the audio mixing process, or generating a multi-level environmental sound evolution scheme based on the psychological arc of the character, or even challenging the creator's preset sound logic to stimulate new artistic possibilities. At the same time, the lightness of edge computing and neural rendering engines will enable the integration of film-grade AI audio processing into mobile terminals. Independent creators can complete intelligent sound field simulation and adaptive master processing, supported by a large mixer, through tablet devices. Under the premise of ensuring security, the cloud collaboration framework can reconstruct the distributed production process, allowing multinational teams to share the same intelligent audio mixing environment in real-time.

5 Conclusion

The reconstruction of film and television audio production by artificial intelligence has gone beyond simple technical iteration. It is catalysing the historical migration of sound creation from an empirical paradigm to an intelligent paradigm. Although current technology is limited by data dependence, computational constraints, and artistic expression bottlenecks, the breakthrough of adaptive learning and semantic interaction is promoting the evolution of AI from an execution tool to a creative collaborator with aesthetic understanding, thereby deconstructing the physical limitations of sound production. The essence of this revolution lies in redefining the relationship between creators and sound. As algorithms take over the technical work of spectrum balancing and noise filtering, humans are freed from the shackles of traditional craft, and creativity is focused on the possible boundaries of sound as a carrier of narrative entropy.

Funding

Funding Communication University of China, Nanjing 2025 Teaching Reform Project, "Innovative Research on the 'AI-Human Collaboration' Dual-Track Teaching Model for Audio-Visual Content Production Courses in Film and Television Directing During the Digital-Intelligence Era" (JG20250608)

About the Author

Lingzi Li, master, lecturer, research direction: sound art, digital media, film, and television recording.

References

[1] Liu Lang, Zhu Shiling. Film and television sound creation driven by AIGC: technology empowers thinking expansion and challenges risk [J]. Film and television production, 2025,31(4):38-42.

[2] He Ying. Innovative models and challenges of audio creation empowered by AIGC technology [J]. Journalism and communication, 2025,(7): 64-66.

[3] Hao Kuo, Wang Kexin. Innovative application of AI technology in film and television short film production [J]. Home Theater,2025,(2):100-102.

[4] Zhang Qinghong. The application of AI in the late field of film and television audio [J]. Modern television technology,2024,(10):107-111.