

Protection of Voice Actors' Rights and Interests in the Context of AI Speech Synthesis Technology

Jiahua He

Law School, Hubei University of Economics, Wuhan, Hubei, 430205, China

Keywords: Dubbing Industry, AI Speech Technology, Voices Infringement, Voices Protection.

Abstract: The development of generative artificial intelligence (AIGC) technology is advancing rapidly, accelerating the transformation of digital audio content production models while also triggering a series of legal risks, among which voice infringement issues are particularly prominent. Given the unique industry characteristics of the dubbing field in China, it is imperative to regulate AIGC technology through legal means. Starting from the first judicial dispute over AI-generated voice infringement in China, this article attempts to deconstruct the phenomenon of voice infringement, finding that it is closely related to the transformation needs of the dubbing industry and the weak protection of the law. Based on this, it proposes relevant protection measures such as independent legislation, improving voice authorization, and establishing a voice evaluation mechanism, with the aim of exploring protection paths for dubbing actors' rights in the digital age and achieving a mutually beneficial interaction between AIGC technology and the dubbing industry.

1 INTRODUCTION

With the continuous maturity of Artificial Intelligence Generated Content (AIGC) technology in China, human-computer interaction scenarios have become increasingly rich. Technologies such as voiceprint recognition and speech synthesis have also found diverse application scenarios in daily life. However, as a representative personality right and an important commercial resource, the risk of infringement of voice's biological information is correspondingly increasing. Against the backdrop of infringement chaos caused by technology abuse, China has not yet issued laws with artificial intelligence as the main subject, which makes it more difficult to identify the responsibility and divide the rights and responsibilities of voice infringement in judicial practice.

This study aims to, within the existing judicial framework, through the rational transfer and application of similar systems, conduct an in-depth analysis of the issues related to the identification and protection of sound rights and interests from three dimensions: promoting the independent legislation of voice rights, improving the sound authorization system, and constructing a voice evaluation mechanism.

2 THE URGENCY OF PROTECTING AGAINST THE PHENOMENON OF AI VOICE INFRINGEMENT

2.1 The Particularity of the Dubbing Industry and the Need for Transformation

Sound is a unique acoustic phenomenon produced by the vibration of a person's vocal cords. Due to the differences in the structure of each person's vocal cords and oral cavity, everyone's voice has its own uniqueness (Schierholz, 2019). Voice actors can achieve the performance effect of playing multiple roles with one voice by changing various aspects such as the pitch and sound pressure of their voices. As can be seen from the judicial precedents of the first AI-generated voice personality right infringement case in China, the protection of voice rights and interests today not only has no systematic and professional legal support, but also the variable sound color and tone line creates a protection dilemma for the legislation of the voice itself. This undoubtedly further impedes the development of the industry. Also, allowing the proliferation of AI-infringing

products, with a large amount of unauthorized AI voiceover content flooding the market, will reduce the commercial trust of the audience in dubbing works, affect the reputation of the entire industry, and further reduce the demand for professional voice actors, thus impeding the sound development of the industry.

The transition of voice actors from behind the scenes to the spotlight is a symbol of occupational standardization and a necessary step for the transformation and development of the industry. This enables the listener to better establish the connection between a particular voice line and the sound source, realizing the “many-to-one” correlation effect, further laying a feasible foundation for the recognizability of the voice, and alleviating the difficulty of applying the extremity of variable voices. In addition, dubbing is a performance activity that integrates “emotion, intonation, and breath.” Hastily divested of human understanding and generalized digitization of the initiative itself has limitations. Because the volume of AI-generated sound is relatively constant, it is difficult for the listener to perceive the three-dimensional auditory space, and thus sample the immersive experience brought by dubbing performance. When the portability of voices technology is emphasized far more than the pursuit of aesthetics, the phenomenon of bad money driving out good will inevitably occur, squeezing the living space of practitioners and raising the threshold of entry for newcomers. In the long run, this will not only reduce the enthusiasm of the dubbing ecosystem, but will also be detrimental to the transformation and upgrading of the dubbing industry.

2.2 The Weak Protection of Voices by Existing Laws Abstract Frame

Compared with representative portraits, auditory sounds have long been on the marginal position in terms of legislative protection. Although previously, relevant provisions regarding sound rights and interests have been made in different forms in legal norms such as the Trademark Law and the Anti-Unfair Competition Law, on the whole, these provisions tend to protect the economic rights and interests generated by sounds, with relatively little protection for the sounds themselves. With the increasing frequency of voice infringement cases recently, and the judicial practice facing the embarrassing situation of having no laws to rely on, there is an urgent and realistic need for legislation on the right to voices.

Throughout the world, the United States protects voices' interests through a dual legislative model of the right to privacy and the right of publicity. However, Liming Wang on the protection of the migration of the model of the application of the denial of the right of publicity as the United States as the original concept of the rule of law, from the concept of the creation of the object it protects are not applicable to the Chinese system. Article 9 of the French Civil Code stipulates that, as one of the personality characteristics, when the voices meet a certain degree of subject recognizability, they can be protected by an independent right to voices. According to Article 36 of the Civil Code of Quebec, Canada, in this region, names, portraits, sounds, etc., all fall within the scope of the extended rights and interests of the right to privacy, and they are protected by safeguarding the right to privacy in judicial practice. Germany adopts a criminal legislation model, protecting the voices as an independent right of personality through criminal law (Wang, 2024; Chen, 1981).

In summary, in addition to the Canadian province of Quebec and other geographical areas, most countries or regions of the law to a certain extent, recognized the voice of the status of independent personality rights, effectively demonstrating the inevitable development trend of the legalization of voice rights and interests on a global scale.

Looking back on the research by Chinese scholars, Guodong Xu believes that the “portrait and voice rights” should be combined to create the same legislative protection for sound and similar portraits. However, the author of this paper holds the view that the establishment through combination implies the acknowledgement that the legal interests of the two are different and that the existing laws are imperfect. Given the irreversible development of AIGC in today's era, this view precisely corroborates the theory of independent legislation for voices proposed by Lixin Yang. Liming Wang once advocated that the voice is not an independent personality right, since the promulgation of the Civil Code, changed his view that the voice is a special legal personality interest, not a specific personality right; and scholars represented by Lixin Yang believe that the right to the voice is a natural person to independently dominate their own voice interests, decide to use and dispose of their own voice of the specific personality right, the right to the voice should be independent (Xu, 2004; Wang, 2018; Yang, Yuan, 2005). Article 1023 (2) of the Civil Code for the first time on the protection of voice “reference to the application of” portrait rights of the quasi-legislative technology. But the rights and

interests of the voice and the protection of the right to portrait in the infringement of the elements and forms of different, and the voice does not need to be a carrier to break through the limitations of the right to portrait the need for carriers, which gives rise to a variety of insufficient to be "reference to the application" of the right to voice. The differences are not sufficient to be covered by the "application by reference", so people need a strong relevance of the legislative guarantee.

3 APPROACHES TO SAFEGUARD THE RIGHTS AND INTERESTS OF VOICES

Indeed, the widespread application of AI in the field of voices will indeed squeeze the survival space of dubbing practitioners. However, in the context of technological empowerment, people cannot be afraid of the pain of change in the industry to stand still. The key contradiction is that the relations of production represented by the law have not been able to adapt to the productivity changes brought about by artificial intelligence promptly, and stopping it at this time is tantamount to holding on to the past. It is not wise to curb the development of productivity, but it is necessary to proactively introduce relevant policies and measures. The following three solutions are proposed in order to realize the protection of the rights and interests of voices.

3.1 Promote the Independent Legislation of the Right to Voice

3.1.1 What Is the Legal Attribute of the Right to Voice?

The importance of voice legislation has already been discussed, but since China has not yet introduced an enforceable statutory law, it needs to be explained from the perspective of feasibility. Many countries define image rights as covering the right of individuals to prohibit the unauthorized use of their names, portraits and voices. However, at this stage, the EU lacks a unified and coordinated image rights framework, and there are significant differences and fragmentation in the image rights systems of countries around the world. Such as some legislative initiatives in Russia, which analogize the protection of voice to the protection of the human image. However, the legal definition and protection scope of voices remain unclear. This makes it difficult for voice actors to find clear and powerful legal bases to

safeguard their own rights and interests when their voices are infringed upon (Baris, 2024; Ruslan, & Evgenia, 2024).

Although the legal protection systems for the right to voice in various countries are not yet perfect at the present stage, some legal measures specifically targeting AI-generated voices are already being piloted and promoted. For example, the U.S. state of Tennessee enacted the Ensuring the Safety of Portraits, Voices, and Images Act (ELVISAct) in 2024, which expanded the scope of the state's statutory right of publicity, exposing the behaviors of artificial intelligence services, internet platforms and so on, that use artists' voices and portraits to new liability risks (McCarthy, 2024)

Scholars and legislators in various countries have produced a wealth of arguments on the legal attributes of sound. Taking into account the current development of AIGC technology and the outcome of China's first AI sound infringement case, the authors of this article believe that the development of sound rights and interests has become a booming trend, and Yang Lixin's view on separate legislation for sound rights and interests is more realistic and feasible in the contemporary era.

3.1.2 Does the Voice Have Recognizability?

The recognizability of the voice is a prerequisite for legal protection. Recognizability can be further understood as whether the AI work is creative or not, and whether it can cover the original human voice. The Beijing Internet Court pointed out on its official platform that the recognizability of a natural person's voice means that a specific natural person can be identified through the characteristics of that voice on the basis of repeated multiple or long-term listening by others. Not coincidentally, before the introduction of AI statute law, the standard of recognizability for natural person's voice can be analogized to AI-generated voices. If the voice synthesized by artificial intelligence enables the general public or the public in relevant fields to associate it with a specific natural person based on its timbre, intonation, and pronunciation style, it can naturally be determined to be identifiable (Beijing Internet Court Research Group, 2024). Patel suggests granting copyrights to AI voice-over models and regarding the outputs as original works, as a way to fulfill the conditions for copyright protection (Patel, 2024). However, the authors of this paper believe that this suggestion is not applicable to China's national conditions. AI technology is just at the initial stage in China and there are relatively few relevant judicial precedents.

Before the official artificial intelligence act is issued, AI should not be arbitrarily identified as the creative subject.

In addition, Wang Shaoxi also proposed that “when determining identifiability, a distinction should be made between celebrities and ordinary people, and the cognitive standards of specific groups should be taken into account (Wang, 2023)”. In the author's opinion, the starting point of the argument is reasonable, but it fails to clarify how to apply it. Subsequent research is needed to conduct in-depth interpretations from aspects such as regionality, social circles, and other directions.

Guaranteeing voice rights and interests through independent legislation can fundamentally curb the chaos of infringement and avoid the identification problems caused by the dilution of voice data. Only by promptly improving the identification and protection of the right to voice at the legislative and judicial levels can the ability to protect the right to voice in a normalized manner be enhanced.

3.2 Improve the Voice Authorization System

The AI technology collects the voices of dubbing practitioners and enters them as training data into the information database, which will make it more difficult to define the voices products that are accused of infringement and more difficult to promote the process of providing evidence and safeguarding rights.

Thus, in addition to legislation, people can also improve the chaos by improving the authorization system. Unilateral authorization is a relatively common authorization method nowadays. For example, on April 24, the voice actor Qianjing Zhao announced that he had authorized his voice to the AI audio series A Record of Mortal's Cultivation to Immortality produced by TME. Another example is that after the huge popularity of Ne Zha 2, the voice actor Yanting Lü used her exceedingly distinctive voice of Ne Zha to record commercials in exchange for commercial benefits. As can be seen from the above, legal authorization is undoubtedly a good proof of voiceprint protection.

Conversely, legal authorization can not only bring commercial benefits but also turn the tide in critical moments. For example, when the famous voice actor Guangtao Jiang was unable to participate in voice recording work due to suspected criminal offenses, the game project team used the "Anti-Entropy AI" technology to generate the voice of that character.

This measure can, to the greatest extent, avoid the subsequent lack of in-game voice resources, reduce the operational risks of the project, and mitigate the company's losses. However, the one-way authorization mechanism has always been restricted by efficiency and is insufficient to meet the needs of economic and social development. Therefore, we need to draw on advanced domestic and foreign experiences and create diverse authorization methods that suit different development models. Concerning the relatively complete portrait rights authorization models internationally, it can be specifically divided into individual authorization or the agency of industry organizations and the coordination measures of some authorization agencies. Also, as in the case of the China Music Copyright Association, the industry organization agent mode, in various parts of the rights protection management agencies, is unified management. Through centralized management, sound rights holders can save time and energy while efficiently obtaining commercial remuneration, allowing industry organizations to drive economic and social employment and optimize the efficiency of resource allocation in their business, and ultimately achieving a “mutually beneficial” authorization system.

3.3 Construct a Voice Evaluation Mechanism

3.3.1 What Are the Specific Judgment Criteria for the Recognizability of a Voice?

The question of by what criteria the recognizability of sounds should be judged is a matter that the judiciary urgently needs to address. The research group of the Beijing Internet Court believes that it should be comprehensively considered from two aspects: subjective criteria and usage methods, supplemented by whether the general public or the public within a certain scope can recognize it as the judgment standard (Research Group of Beijing Internet Court, 2024). Budnik and Evpak proposed a hypothesis for the legal protection of voice identity, that is, to create a data identity covering multiple aspects such as voice parameters and vocal characteristics. On this basis, trained generative neural networks are used for identification and comparison, which can provide more effective legal bases and technical means for resolving disputes over voice cloning and the unauthorized use of voices.

In the first judicial case regarding AI voice infringement in China, the judge determined through an on-site inspection in court that the AI voice had a high degree of consistency with Ms. Yin's timbre, intonation, pronunciation style and so on. Thus, it was inferred that the plaintiff's voice rights and interests extended to the AI voice involved in the case (Beijing Internet Court, 2023). The above-mentioned method takes the judge as the subject of appraisal, and the fluctuating influence on the judgment result due to the lack of professional knowledge of the lay audience cannot be excluded. Therefore, the authors of this article believe that when determining whether a work that infringes upon others' voices is an act of infringement, the expert witness system is indeed necessary. Experts can analyze the similarities in aspects such as voice characteristics and expression techniques from a professional perspective, thereby obtaining more convincing appraisal results.

In addition, in terms of matching the similarity between the production content and the suspected source, voiceprint comparison is a common means of identifying infringement. Voiceprint collection equipment is introduced to verify the similarity, and set threshold standards to test the recognizability of the voice. For example, if the similarity reaches more than 80%, it may be determined by the judicial authorities as a substantial similarity, thus constituting an infringement.

3.3.2 How Should the Responsibilities Be Determined After the Occurrence of an Infringement?

Sound products generated through voice processing are, to a certain extent, similar to musical works. Therefore, the judgment methods in music infringement cases can be transferred applications. Manuelian pointed out that in music copyright infringement cases, to determine infringement, the plaintiff needs to prove three elements: having a valid copyright, the defendant's replication of the protected materials, and the defendant's replication constituting "inappropriate appropriation" (Manuelian, 1988). The author of this paper believes that the above-mentioned judgment ideas can also be applied by reference in the dubbing industry. Moreover, based on the practical problem of difficulty in the burden of proof, it is necessary to refer to the provisions of the Tort Liability Law of China and reverse the burden of proof under specific circumstances. That is, the infringing party shall prove that the data resources used for training AI do not contain the biometric information of the infringed party. If the infringing party is unable to prove itself or has no training materials, the similarity between the materials

provided by the prosecutor and the generated content will be used to judge how much to penalize.

Hutiri and Wiebke pointed out that it is necessary to explore an accountability system for training data. For example, authenticate the source of training data and provide creators with an exit mechanism, such as canceling training data, so as to regulate the application of voice generation technology from the source and protect the rights and interests of relevant personnel (Hutiri, Papakyriakopoulos, and Xiang, 2024). By clarifying the data source and granting creators control, it is possible to effectively reduce the illegal collection and use of data and promote the development of voice generation technology on a legal and compliant track.

4 CONCLUSION

In the digital era, as an important biological and commercial resource, the significance of the relevant norms for the protection of the rights and interests of voices has become increasingly prominent. This paper preliminarily explores three measures to protect the rights and interests of voice through the analysis of the reality specificity of the dubbing industry and the theoretical deconstruction of the domestic and international scope of voice legislation. Protecting voice rights through legislation is the inevitable requirement of the artificial intelligence era, and the possibility provided by the legal and open nature of personality rights. Improving the sound authorization mechanism is a due course of action in line with the development of the times. The construction of a voice evaluation mechanism is to empower the legal system with scientific and technological means. The above thoughts provide a little idea for China's budding sound right provisions and even the AI draft, but this paper has yet to go deeper into the technical aspects of voice identification in the sound assessment mechanism. In the future, it is expected that the AI technology and the sound protection draft will continue to develop and improve to better meet the diversified development of the dubbing industry and bring more wonderful interpretations to the audience.

REFERENCES

Baris, A. 2024. AI covers: legal notes on audio mining and voice cloning. *Journal of Intellectual Property Law & Practice*, 19(7), 571-576.

Beijing Internet Court Research Group. First-instance Judgment on the First Case of Infringement of Personality Rights by AI - AI-generated Voice in China, published on the WeChat official account "Beijing Internet Court" , https://mp.weixin.qq.com/s/_GxGaG6Q2NYHJWQuOtMyrQ (uploaded on April 23, 2024)

Civil Judgment, 2023. Jing 0491 Min Chu No. 12142 of the Beijing Internet Court.

Der Manuelian, M. 1988. The Role of the Expert Witness in Music Copyright Infringement Cases. *Fordham L. Rev.*, 57, 127.

Guodong Xu. Draft of the Green Civil Code. Beijing: Social Sciences Academic Press, 2004: 92.

Hutiri, W., Papakyriakopoulos, O., & Xiang, A. (2024, June). Not My Voice! A Taxonomy of Ethical and Safety Harms of Speech Generators. In The 2024 ACM Conference on Fairness, Accountability, and Transparency (pp. 359-376).

McCarthy, C. R., 2024. Artificial Intelligence and the Entertainment Industry. *Studies in Law and Business*.

Patel, P. 2024. AI Voice Enters the Copyright Regime: Proposal of a Three-Part Framework. *Fordham Intellectual Property, Media and Entertainment Law Journal*, 34(2), 451.

Research Group of Beijing Internet Court, 2024. Legal Recognition of the Infringement of Voice Rights and Interests by AI-generated Voices—Taking the Case of Yin Mou v. Beijing Certain Intelligent Technology Company and Others for Infringement of Personality Rights as an Example. *Law Application*, (09): 123-133, 129.

Ruslan, B., & Evgenia, E. 2024. Human Voice: Legal Protection Challenges. *Legal Issues in the Digital Age*, (4), 28-45.

Schierholz, V., 2019. *Götting/Schertz/Seitz, Handbuch Persönlichkeitsrecht*, München.

The Civil Code of Czechoslovakia (in italics), 1981. translated by Hanzhang Chen, Law Press, 10-11.

Wang, L. M., 2018. On the Legality and Openness of Personality Rights. *Economic and Trade Law Review*, (01): 17-27.

Wang, L. M., 2024. On the Legal Protection Model of Voice Rights and Interests. *Journal of Financial and Economic Law*, (01): 3-20, 5.

Wang, S. X., 2023. Interpretation and Application of Voice Protection in the Era of the Civil Code, *Law Application*, 6(44).

Yang, L. X., Yuan, X. S., 2005. On the Independence of the Right to Voice and Its Protection in Civil Law. *Studies in Law and Business*, (04): 103-109.