

Protection of Acoustic Rights and Responses to Sound Infringement by Smart Energy Devices

Ziyang Luo^{1,a,*}

¹University College London, Empire Way, Wembley, London, United Kingdom
a. ucbvz75@ucl.ac.uk
*corresponding author

Abstract: With the advancement of artificial intelligence (AI) technology, the infringement of personal acoustic rights has become an increasingly severe issue. Objective factors, such as the ease of collecting voice data, have contributed to the frequent emergence of cases involving the infringement of personal acoustic rights through generative AI. In particular, the collection of voice data by smart energy devices has led to numerous legal and ethical issues. This paper examines the development and current status of legal frameworks protecting acoustic rights in Chinese law. It explores how the rise of AI and smart devices has challenged existing regulations and created new risks for personal acoustic rights. In response, this study advocates for the establishment of a robust voice licensing system, the incorporation of watermarking technology in AI-generated audio, and stricter regulatory oversight on usage scenarios and generated content. By implementing these measures, the study aims to strengthen the legal and technological safeguards for acoustic rights, ensuring effective protection in the age of AI.

Keywords: Acoustic right, generative artificial intelligence, Chinese law, smart energy devices

1. Introduction

As artificial intelligence (AI) technology continues to emerge and improve, it has become capable of collecting and analyzing the voiceprints of specific individuals, ultimately mimicking their voices with near-perfect accuracy. This technology, which simulates or synthesizes the voices of specific individuals using AI, is referred to as AI-generated voice synthesis. While this technology has become readily accessible to the public, the illegal collection of natural persons' voice data and the infringement of voice rights have become increasingly severe. The misuse of AI to imitate individuals' voices raises numerous issues. Legally, it can distort the victim's image, cause others to mistakenly associate the voice with the victim and thereby infringe upon their rights to reputation and image.

On April 23, 2024, the Beijing Internet Court issued a civil judgment ((2023) Jing 0491 Min Chu No. 12142), marking the first case of AI-generated voice infringement on personality rights in China. In this case, the plaintiff, Yin, a voice actor, discovered that her voice had been synthesized using AI technology and sold on a platform operated by the defendant, a smart technology company. Yin filed a lawsuit against the company and four other defendants, claiming that their actions had seriously infringed upon her voice rights. She demanded an immediate cessation of the infringement, a public apology, and compensation for both economic and emotional damages.

The defendants denied the infringement. The smart technology company argued that the voice products on its platform originated from another software company and were obtained legally. The software company claimed that its voice source was provided by a cultural media company, which argued that it owned the copyrights to the recordings based on a prior agreement with the plaintiff. Meanwhile, the platform operator and the product distributor asserted that they were not liable for infringement.

Upon review, the court found that natural people's voices, distinguished by voiceprints, timbre, and frequency, possess uniqueness, individuality, and stability. These attributes can evoke thoughts or emotions related to the individual, serving as identifiers of their behavior and identity. AI-generated voices, if identifiable to the public or specific audiences based on their timbre, tone, and pronunciation style, can be considered as linked to the individual and therefore recognizable. In this case, the software company had used the plaintiff's voice to develop a text-to-speech product, and forensic examination showed that the AI-generated voice highly matched the plaintiff's timbre, tone, and pronunciation style. This alignment enabled the public to associate the voice with the plaintiff, thereby identifying her. As a result, the plaintiff's voice rights extended to the AI-generated voice.

Although the cultural media company held copyrights to the recorded works, these rights did not include the authority to permit the AI synthesis of the plaintiff's voice. The unauthorized AI synthesis by the cultural media and software companies, conducted without the plaintiff's knowledge or consent, constituted an infringement of her voice rights. The court ruled that these actions caused damage to the plaintiff's voice rights, and the responsible parties were held liable. Consequently, the smart technology company and the software company were ordered to issue a public apology, while the cultural media company and the software company were jointly required to compensate the plaintiff with 250,000 RMB.

This case clearly demonstrates the judicial application of the provisions on the protection of natural persons' voice interests under the Civil Code since its enactment. The court accurately understood the spirit of the Civil Code, fully safeguarding the voice rights of natural persons. It clarified that voice, as a marker of personal identity, reflects human dignity and contains certain economic value, thereby requiring a heightened duty of care when using others' voices.

In the post-AI era, the Personal Information Protection Law has explicitly regulated the collection of voice data, and documents such as the Provisions on the Administration of Deep Synthesis for Internet Information Services have indirectly advanced legislative protection for personal voice rights [1]. However, the increasingly severe situation has reignited discussions on the protection of voice rights and posed new challenges to the inherent lag of the law.

This paper explores the concept and forms of protection for voice rights, examines the current practices of voice data collection by smart energy devices, and proposes recommendations to prevent voice rights infringement and enhance their protection in mainland China.

2. The concept of acoustic rights and the development of their legal protection in China

2.1. The concept of acoustic rights

In China, it is widely acknowledged in academic circles that acoustic rights constitute a statutory personality right, with their fundamental nature defined as the "right to respect." Regarding the subject of this right, acoustic rights are exclusively attributed to natural persons, excluding legal entities and non-legal organizations. Due to the inherently personal nature of acoustic rights, they cannot be restricted, nor can they be transferred to other entities.

The object of acoustic rights is the acoustic interest, representing the personality interests embodied by sound [2]. This encompasses both economic and non-economic (spiritual) interests. As for the content of acoustic rights, "sound" refers solely to the sound itself, distinct from the

"expression of sound" protected under copyright law. Drawing a parallel to portrait rights, individuals holding such rights are entitled to produce, use, publicly display, or authorize others to publicly use their portraits under the law. Accordingly, acoustic rights grant natural persons the lawful authority to produce, use, publicly display, or authorize others to publicly use their own voices.

2.2. The development of legal protection for acoustic rights in China

Before the implementation of the Civil Code in 2021, voice was primarily protected as an object of intellectual property under the Trademark Law and the Copyright Law. During the era of the General Principles of Civil Law, there were no explicit provisions for the protection of voice rights within the personality rights framework closely related to the voice rights of natural persons. The voice rights of natural persons were only marginally protected through the provisions on the right to reputation. Article 101 of the General Principles of Civil Law stipulated that citizens and legal persons enjoy the right to reputation, and citizens' personal dignity is protected by law, prohibiting insults, defamation, and other actions that damage the reputation of individuals or legal entities.

Theoretically, reputation, as a form of social evaluation, can be understood as an objective assessment by society or others regarding the morality, talent, credibility, qualifications, prestige, and image of specific natural persons, legal persons, or unincorporated organizations. If voice is regarded as an identifiable characteristic of a natural person and as an aspect of others' evaluation of that person's character, imitating and distorting someone's voice could constitute an infringement of their right to reputation. However, such cases were extremely rare prior to this legal framework. Before the Civil Code, voice was predominantly viewed as having economic value rather than as an attribute of personality, and it was not directly protected as an object of personality rights.

Nevertheless, as natural persons are the fundamental components of society, their voice rights, as an identifying characteristic, should receive greater attention compared to non-natural persons. Protecting voice rights solely from the perspective of copyright is evidently insufficient. On one hand, the voices of non-public figures are also at risk of infringement, but their value is difficult to quantify, making it challenging to provide effective protection under these frameworks, thereby threatening the principle of equality among civil subjects. On the other hand, if voice-related protections are afforded only to non-natural persons while ignoring the reputation-based protection of natural persons' voices, this would undermine the human-centered legislative philosophy of private law.

The promulgation of the Civil Code elevated the protection of voice rights to a new level, bringing unprecedented attention to the protection of natural persons' voice rights and providing guidance for subsequent legislative efforts. Comparative legislation generally protects voice rights by expanding the interpretation of specific personality rights, such as privacy rights and portrait rights. For instance, Article 36 of the Quebec Civil Code establishes protection for voice rights from a privacy perspective [3]. Similarly, China's Civil Code expanded the scope of portrait rights, explicitly including voice as one of its objects of protection.

However, as the Civil Code has only been in effect for a relatively short period, many of its provisions remain unclear and have yet to be fully implemented. For instance, Article 1023 states that "the protection of natural persons' voices shall refer to the relevant rules for the protection of portrait rights." However, the term "refer to" does not imply full applicability. As noted in the interpretive text by the Legislative Affairs Commission of the National People's Congress, the protection of natural persons' voices "can only refer to, but not entirely apply, the rules governing portrait rights [4]." Nor can these rules be analogously applied. At present, there is no judicial interpretation clarifying issues such as the liable parties for voice rights infringement, principles of liability, and specific standards for liability [5].

To meet the growing legislative demands of the information age, other legal frameworks have indirectly provided protection for voice rights. For example, the Personal Information Protection Law

and the Provisions on the Administration of Deep Synthesis for Internet Information Services address voice-related issues. Article 4 of the Personal Information Protection Law defines personal information as "various information recorded electronically or otherwise that is related to identified or identifiable natural persons, excluding anonymized information." This definition implies that the widespread practices of collecting others' voices and analyzing, altering, or using them through AI technologies fall under its regulation. Misusing others' voices may result in penalties such as fines or credit record entries under the law. Furthermore, Article 13 of the Personal Information Protection Law outlines several legitimate grounds for processing personal information, including individual consent, necessity for fulfilling agreements or statutory obligations, emergency situations, and reasonable use in the public interest.

3. A review of voice data collection by smart energy devices in the intelligent society

With the continuous advancements and investments by algorithm developers in the field of artificial intelligence (AI) voice synthesis, the quality of synthetic voice technology has significantly improved. Consequently, the production costs of synthetic voice have decreased substantially, and production cycles have been greatly shortened. Most smart energy devices are now equipped with the necessary hardware and software capabilities to support the operation of such technologies. As a result, AI voice synthesis algorithms have rapidly gained widespread acceptance and application among the general public. However, the reduced operational costs and the ease of use of these technologies have inadvertently lowered the barriers for malicious actors to exploit this field. This has led to a surge in cases involving voice rights infringements, highlighting the dual-edged nature of technological progress in the context of smart energy devices and voice data collection.

3.1. Objective perspective: The ease of access to training samples for synthetic voice

The development of artificial intelligence (AI) voice synthesis technology relies heavily on the construction of extensive voice databases, compelling technology companies to accumulate vast amounts of user voice data. In practice, however, many companies fail to obtain such data through legal and compliant means. Instead, they exploit devices such as voice assistants embedded in smartphones, smart speakers, social media platforms, and various mobile applications to clandestinely monitor and record users' daily conversations, thereby unlawfully expanding their voice databases. Similarly, on an individual level, some users construct voice databases by extracting audio recordings of specific individuals available on the internet. A commonality between these practices is the unauthorized use of such recordings for training deep learning models without obtaining the explicit consent of the individuals involved. This process often leads to the unauthorized exposure and processing of highly sensitive personal information.

The increasing penetration of the internet has further facilitated the ease of obtaining training samples for synthetic voice generation. Whether through voice assistants, smart speakers, social media platforms, mobile applications, or by extracting audio data of specific individuals online, acquiring samples for voice synthesis has become relatively straightforward. Public figures such as business leaders like Lei Jun or popular celebrities frequently have their voice data readily accessible in the online space. Even the voices of ordinary individuals can be obtained through various means. Once these voice samples are uploaded to AI systems, advanced algorithms can accurately replicate their "original voices," significantly heightening the risk of voice rights infringement for the general public.

Moreover, the ongoing advancements in voice synthesis technology have made the process increasingly precise and efficient. It now requires only minimal voice samples—sometimes mere seconds of audio—to generate highly realistic synthetic voices. This rapid progress has exacerbated

the potential threats of voice forgery and misuse, posing severe risks to the protection of voice rights in society.

3.2. Individual perspective: The prevalence of unauthorized voice recording

The ability to record and reproduce sound through specific mediums is a fundamental aspect of voice rights. On a positive note, voice rights holders can grant consent for the recording of their voices. Conversely, in the absence of fair use conditions, voice rights holders are entitled to prohibit unauthorized recording of their voices [6]. In other words, recording someone's voice without their explicit consent constitutes a typical form of voice rights infringement.

However, the widespread adoption of smart energy devices has made such unauthorized recording ("stealth recording") increasingly common. Nearly all smart devices are equipped with highly convenient recording functions, enabling users to record voices anytime and anywhere with remarkable ease and concealment, making detection extremely challenging. Once an individual's voice is recorded, it can be processed using artificial intelligence technologies. Objectively, the processed voice can be manipulated to distort the original intent, making it feasible for use in illicit activities such as fabricating evidence. This presents significant risks and challenges to the protection of voice rights in the context of modern technology.

3.3. Institutional perspective: Deficiencies in the regulatory framework

China's current legal system lacks specific provisions dedicated to the protection of voice rights. Although Article 1023 of the Civil Code states, "The protection of natural persons' voices may refer to the relevant rules for the protection of portrait rights," this approach has inherent limitations. Portrait rights are designed to safeguard an individual's image-related interests, whereas voice rights focus on the protection of an individual's vocal attributes. These two rights address fundamentally different objects of protection. As such, simply applying measures designed for portrait rights is insufficient for providing comprehensive and effective safeguards for voice rights [7].

Additionally, platforms that host and distribute synthetic voice works—those most capable of directly regulating such content—often fail to enforce adequate oversight. Synthetic voice works, due to their non-authenticity and subjective creative nature, frequently simulate target individuals making statements they have never actually made. These statements often include provocative, satirical, novel, or sensational elements designed to capture public attention. As a result, online platforms, driven by their own interests, tend to lower the standards for reviewing and moderating synthetic voice content during its storage and dissemination. This leniency inadvertently creates opportunities for illicit activities, further exacerbating the issue of inadequate regulation.

4. Recommendations for strengthening the protection of voice rights in the information age

To ensure the lawful and compliant use of AI-generated voices while enhancing the protection of voice rights, it is essential to refer to relevant provisions under the Copyright Law and the Civil Code. Obtaining proper authorization or consent from the relevant copyright holders or the natural persons whose voices are involved is imperative; otherwise, there is a risk of infringing on others' copyright or personality rights.

Moreover, given the close connection between AI-generated voices and personal information as well as cybersecurity, the use of such technologies should also comply with regulations such as the Personal Information Protection Law, the Provisions on the Administration of Deep Synthesis for Internet Information Services, and the Interim Measures for the Management of Generative Artificial Intelligence Services. Additionally, the recently drafted Measures for the Identification of AI-

Generated Synthetic Content (Draft for Comments) by the Cyberspace Administration of China highlights the future direction for voice rights protection.

For individual AI users and service providers, adherence to these regulations is crucial to mitigating potential legal risks associated with the use of AI-generated voices. These measures collectively provide a robust framework to safeguard voice rights in the evolving landscape of the information age.

4.1. Implementation and improvement of the voice licensing system

According to Article 1023(2) of the Civil Code, the protection of voice rights can refer to the relevant rules governing portrait rights. If a person infringes on another's voice rights without permission, the injured party is entitled to claim civil liability under these rules. Additionally, the Personal Information Protection Law (PIPL) and the Copyright Law both stipulate that prior authorization must be obtained before using someone else's voice. However, if the voice data constitutes publicly available information, the PIPL does not require explicit consent but mandates that the processor fulfills its duty of due diligence and limits its actions to reasonable use [8]. Furthermore, Article 14 of the Provisions on the Administration of Deep Synthesis for Internet Information Services specifies: "Providers of deep synthesis services and technical support for editing biometric information, such as facial or vocal features, shall inform the individuals being edited and obtain their separate consent." Accordingly, when operators of virtual digital humans use deep synthesis or other technologies to collect or utilize biometric features like voices or faces, they and the users of their services are obligated to fully inform the individuals involved and obtain their explicit and separate consent in advance [9].

However, the current voice licensing mechanism requires further refinement. For instance, the lack of convenient channels for obtaining consent presents significant challenges, such as difficulties in contacting the rights holders, especially for public figures, where direct communication with the individuals or their agents is often infeasible through ordinary means. Additional questions also arise, such as whether voice licensing should adhere to formal procedures like those for trademark licensing, or whether administrative intervention is necessary for processing voice licenses.

Another issue, widely debated in academic circles, is whether all recordings require the consent of the voice rights holders or if exceptions exist. According to Article 1020 of the Civil Code, recordings made for personal learning, appreciation, classroom teaching, or scientific research purposes do not require the consent of the voice rights holder. However, how to define the scope of "personal learning and appreciation" remains unclear. In the absence of detailed legal provisions, some argue that, as with portrait rights, personal use of portraits for learning or appreciation must involve publicly available images [10]. By analogy, does this imply that recordings must also be limited to publicly accessible audio? This paper contends that such a limitation would be inappropriate, as it could expose the public to excessive liability risks and destabilize individual rights under such a restrictive framework. For instance, recording a performance in the atrium of the National Centre for the Performing Arts for personal enjoyment without commercial use should be considered fair use.

Furthermore, other scenarios of fair use, such as recordings made by law enforcement agencies for investigative purposes, should not constitute voice rights infringement. Similarly, individuals recording conversations with another party to protect their legitimate rights should not be deemed as infringing voice rights. Therefore, when determining whether a voice recording constitutes infringement, it is essential to consider the purpose of the recording. Properly addressing these nuances would enhance the clarity and effectiveness of the voice licensing system.

4.2. Identification in AI-generated synthetic voices

The Measures for the Identification of AI-Generated Synthetic Content (Draft for Comments) highlights the growing trend of requiring identification for the use of AI-generated synthetic voices. Furthermore, Article 10 of the Provisions on the Administration of Deep Synthesis for Internet Information Services stipulates: ‘Providers of deep synthesis services shall establish and improve feature databases for identifying illegal and harmful information, refine the standards, rules, and procedures for database entries, and record and retain relevant network logs. Upon discovering illegal or harmful information, providers shall take appropriate measures in accordance with the law, preserve relevant records, and promptly report to cyberspace authorities and other competent departments. Providers must also warn, restrict, suspend, or terminate the accounts of users who violate these rules as required by law and relevant agreements.’

The significance of identifying AI-generated synthetic voices can be understood from several perspectives. First, it addresses the challenges of traceability. The pervasive nature of the internet makes it exceedingly difficult to trace the origins of specific content, especially when an infringement gains widespread attention. For service providers, implementing technologies that effectively identify, distinguish, and trace the origins of electronic data—such as embedding specific voice watermarks into synthesized audio—can enable the tracking of users who engage in unlawful use of synthetic voices. If such misuse is detected, providers can take immediate remedial actions, such as suspending the accounts of the offending users.

Second, marking synthetic voices ensures that audiences are properly informed about the nature of the content. This prevents them from mistakenly if the voice originates from the rights holder. Such transparency not only protects the audience’s right to know but also safeguards the personality interests of the voice owner, including their reputation and dignity, preventing violations stemming from impersonation or misuse.

In summary, establishing mandatory identification practices for AI-generated synthetic voices is a critical step toward enhancing accountability, protecting individual rights, and fostering ethical standards in the age of artificial intelligence.

4.3. Enhancing the regulation of legality in AI-generated content

Article 6 of the Provisions on the Administration of Deep Synthesis for Internet Information Services explicitly states: ‘No organization or individual may use deep synthesis services to produce, reproduce, publish, or disseminate information prohibited by laws and administrative regulations, or engage in activities prohibited by laws and administrative regulations, such as those that endanger national security and interests, harm the national image, infringe upon social public interests, disrupt economic and social order, or infringe upon the lawful rights and interests of others. Providers and users of deep synthesis services must not use such services to produce, reproduce, publish, or disseminate false news information. Reposting news information created using deep synthesis services must be limited to content originally published by authorized internet news information sources.’

Based on this provision, users must avoid generating content that is illegal when using AI voice technologies. For service providers and relevant network regulatory authorities, measures such as keyword detection can be implemented to pre-screen user-generated content for potential violations.

Furthermore, for users who repeatedly generate illegal content using AI technologies or engage in severe violations, appropriate punitive measures should be enacted. These may include restricting account functionality, suspending or permanently banning accounts, and, in cases of particularly severe violations resulting in significant harm, pursuing corresponding civil or criminal liability.

By enhancing regulatory oversight and ensuring compliance with existing laws, these measures aim to balance the innovative potential of AI-generated content with the imperative of safeguarding social order, individual rights, and legal standards.

4.4. Enhancing the regulation of legality of AI-generated content usage

AI-generated voices have the potential to be misused for unlawful activities or even criminal offenses. Both the Provisions on the Administration of Deep Synthesis for Internet Information Services and the Interim Measures for the Management of Generative Artificial Intelligence Services emphasize that criminal liability will be pursued in cases where generative AI is used to commit crimes. In the realm of criminal activities, the feasibility of voice imitation introduces opportunities for innovative criminal methods, such as impersonation fraud or the fabrication of evidence in litigation or forensic investigations.

For example, in the case involving the "Three Sheep Company" on September 20, 2024, the company reported false audio and video content allegedly featuring its employee Lu being circulated online. Subsequent investigations revealed that Wang had used AI tools to train and generate counterfeit audio mimicking Lu's voice based on existing audio and video materials. The falsified audio was then disseminated online, resulting in widespread rumors. Wang was ultimately subjected to criminal coercive measures. This incident underscores the necessity of strengthening regulations governing the intended use of AI-generated voices.

Additionally, advancements in forensic technologies for electronic data authentication are essential. These technologies should focus on accurately detecting signs of human fabrication or tampering in audio recordings. Such innovations not only support stricter regulatory oversight of AI-generated voice technologies but also better meet judicial needs by providing reliable tools to verify the authenticity of audio evidence.

5. Conclusion

While advancements in AI technology have significantly improved daily life, the accompanying risks of legal systems being circumvented must not be overlooked. Voice rights, as a particularly vulnerable category of rights, face heightened challenges in the AI era. The ease of acquiring voice samples and the strong concealment of infringement activities highlight the urgency of improving related legal frameworks.

Under the framework of the Civil Code, the protection of voice rights has undergone significant reform. However, due to the relatively short period of practical implementation, certain gaps in specific provisions still require further clarification through judicial interpretation. The publication of the Provisions on the Administration of Deep Synthesis for Internet Information Services and the draft Measures for the Identification of AI-Generated Synthetic Content has, to some extent, enhanced the mechanisms for protecting voice rights established by the Civil Code. These developments suggest that the legal frameworks and enforcement mechanisms for protecting voice rights in the AI domain will continue to improve in the future.

To further strengthen the protection of voice rights, it is necessary to implement a robust voice licensing system and establish a more comprehensive protective framework to ensure the lawful use of voiceprint data. In safeguarding voice rights, measures such as embedding watermarks in AI-generated voices and reinforcing regulatory standards for usage scenarios and generated content should also be adopted. These steps will provide a more secure and structured environment for the use and protection of voice rights in the age of artificial intelligence.

References

- [1] Wang Huiquan, Liu Lu. (2024) *Challenges and Responses: Legal Regulation of Personal Information Protection in the Era of Artificial Intelligence*. *Economic Responsibility Audit*, 3, 50-57.
- [2] Yang Lixin, Yuan Xueshi. (2005) *On the Independence of Voice Rights and Their Protection under Civil Law*. *Studies in Law and Business*, 4, 107.
- [3] Wang Liming. (2024) *On the Legal Protection Model of Voice Rights*. *Journal of Law and Economics*, 1, 11.
- [4] Huang Wei. (2020) *Interpretation of the Personality Rights Section of the Civil Code of the People's Republic of China*. Beijing: China Legal Publishing House.
- [5] He Xiong, Tang Fei. (2024) *Civil Law Protection of Natural Persons' Voice Rights in the Era of Artificial Intelligence*. *Journal of Yangtze University (Social Sciences Edition)*, 47(4), 119.
- [6] Wang Shaoxi. (2023) *Interpretation and Application of Voice Protection in the Era of the Civil Code*. *Journal of Law Application*, 6, 38.
- [7] Wang Liming, Cheng. (2020) *Commentary on the Civil Code of China: Personality Rights Section*. Beijing: China Renmin University Press, 298.
- [8] Wang Nijie. (2024) *On the Legal Nature and Protection Pathways of Voice Rights*. *Law Science*, 10, 132-146.
- [9] Guo Dezhong, Zhang Yunwei. (2024) *Infringement Risks of Training Data in Generative Artificial Intelligence and Legal Responses*. *Journal of Xiangtan University (Philosophy and Social Sciences)*, 48(05), 78-86.
- [10] Wang Liming, Cheng. (2020) *Interpretation and Commentary on the Civil Code of China: Personality Rights Section*. Beijing: China Renmin University Press, 298.