

*Research article*

УДК: 342

JEL: K2

DOI:10.17323/2713-2749.2024.4.28.45

# Human Voice: Legal Protection Challenges

---

---



**Ruslan Aleksandrovich Budnik<sup>1</sup>,**



**Evgenia Gennadievna Evpak<sup>2</sup>**

<sup>1, 2</sup> National Research University Higher School of Economics, 20 Myasnitskaya Street, Moscow 101000, Russia,

<sup>1</sup> rbudnik@hse.ru, <https://orcid.org/0000-0001-8076-1560>

<sup>2</sup> eevpak@hse.ru, <https://orcid.org/0000-0001-5961-5657>

---



## Abstract

Focused on the phenomenon of human voice, the paper purports to develop approaches to legal protection of ordinary people's voices and those of artists as the technologies for sound synthesis, music and vocal performance are becoming more sophisticated. It was demonstrated the problem of legal protection of human voice has become especially pressing one in the context of artificial neural networks such as vocaloid designed for sound synthesis and voice cloning. The study provides a systemic arrangement of legally meaningful knowledge of human voice useful for the development of legal provisions to protect this personal good. The legal substance of vocal impersonation as a way to simulate and manipulate a synthesized voice was explored. Theoretically applicable legal constructs for protection of voice and computerized cloning technologies were analyzed. A trend to use copyright rather than patent for protecting vocaloid-like generative neural networks and other technological solutions for vocal synthesis was identified. The concept of voice was critically analyzed to propose a viable legal provision for its protection. The primary and auxiliary features of the concept of voice were parametrized for possible use in disputes on the legitimacy of voice cloning or vocal impersonation. The concept of vocal identity of ordinary people and artists was proposed for legal protection of this personal good as a set of performative sonic features including the basic parameters of singing voice as well as acoustic-phonetic and articulatory features

of vocalization. A comprehensive legal and technological methodology for the protection of vocal identity was proposed.

---



## Keywords

human voice; voice as a personal good; protection of voice; voice cloning; vocal impersonation; musical neural networks; vocaloids; legal parametrization of voice; vocal identity.

---

---

**For citation:** Budnik R.A., Evpak E.G. (2024) Human Voice: Legal Protection Challenges. *Legal Issues in the Digital Age*, vol. 5, no. 4, pp. 28–45. DOI:10.17323/2713-2749.2024.4.28.45

## Background

Explosive progress of creative neural networks since the early 2010s has prioritized the study of legal nature and assessment of musical works created through the use of artificial intelligence as one of the topical and at the same time ambivalent problems [Bridy A., 2016]; [Levy P., 2005].

Musical products generated by weak AI-based neural networks are treated as compilation, iteration or algorithmic creative outcomes<sup>1</sup>. Meanwhile, recognizing a creative work and objective form of expression in these outcomes will suffice to make them copyrightable<sup>2</sup>.

By 2023, there were at least 17 popular brands of musical neural networks<sup>3</sup> on the global mass market including services such as Mubert and Suno that are tools for generative song prototyping based on the use of algorithmic vocal loop.

In 2023, the International Federation of the Phonographic Industry (IFPI) has published the *Engaging with Music* — annual Global Music Re-

---

<sup>1</sup> In his paper “Study of models and algorithmic methods for analysis of musical works”, L.A. Aznaurian with reference to A. Rozen and A. Vasiliev provides four types of algorithmic compositions including: (1) composition methods based on mathematical functions; (2) composition practices based on combinatorial methods; (3) practices based on natural processes and (4) practices using rule-based processes.

<sup>2</sup> Two creative patterns are normally identified: (1) spontaneous “divine” inspiration of irrational nature and (2) rational trial-and-error iterations of creative product.

<sup>3</sup> Music Cheat Sheet. A list of popular music apps and plugins (and tips on how to use them). Available at: <https://bobbyoswinski.com> (accessed: 16.10.2024)

port reflecting for the first time the listeners' opinion of neural network-generated music streamed live and listed in playlists. It was reported that 8 out of 10 respondents (almost 79% of 43,000 across 26 countries covered by the study) believed the human agent's creative potential to be sufficient to prop up musical creativity in the context of progress of electronic music and generative art. While 74% of respondents believed that using musical neural networks and AI technologies to clone or substitute for artistic personality was illegal and unacceptable one, 76% were convinced that music and voice timbres should not be processed by neural networks without approval of artists, and 73% agreed that neural network producers should disclose the information on whose music was used to train online music generators. More than half of those polled believed that governments should develop and approve legal provisions applicable to automatically generated musical content<sup>4</sup>.

Ten most popular genres in the IFPI's *Engaging with Music* report included were: pop, rock, hip-hop/rap, dance and electronic, Latino, rock-n-blues, classical and opera music, country, movie soundtracks and reggae<sup>5</sup>. The most popular works created in these genres are exposed to "digital activism", with fans using musical neural networks to create AI remixes and song covers. As a result, the authors, performers and record producers suffer losses from competing illegal remakes of popular works that pay no royalties to the actual copyright holders. This happened to Drake's *Heart On My Sleeve* featuring the Weekend: a neural network remix of this song was generated by Ghostwriter, anonymous user, and posted to TikTok. As an unprecedented fact, the remix was nominated for Grammy award and de-listed by the Recording Academy just in time before the event<sup>6</sup>.

From the perspective of legal assessment, streamed music can represent not only a musical work with added text, but also an audiovisual and other types of complex copyrighted works since a digital release promoted by the

---

<sup>4</sup> For details see: IFPI Global Music Report 2023. State of the Industry. 2023. P. 11. Available at: [https://www.ifpi.org/wp-content/uploads/2020/03/Global\\_Music\\_Report\\_2023\\_State\\_of\\_the\\_Industry.pdf](https://www.ifpi.org/wp-content/uploads/2020/03/Global_Music_Report_2023_State_of_the_Industry.pdf) (accessed: 30.05.2024)

<sup>5</sup> IFPI Engaging with Music 2023. Available at: [https://www.ifpi.org/wp-content/uploads/2023/12/IFPI-Engaging-With-Music-2023\\_full-report.pdf](https://www.ifpi.org/wp-content/uploads/2023/12/IFPI-Engaging-With-Music-2023_full-report.pdf) (accessed: 16.11.2023)

<sup>6</sup> AI-Generated Drake, The Weeknd Song Not Eligible for Grammy. Here's How It Could Have Been Though // Forbes Business. 8.09.2023. Available at: <https://www.forbes.com/sites/darreonnadavis/2023/09/08/ai-generated-drake-the-weeknd-song-not-eligible-for-grammy-heres-how-it-could-have-been-though/?sh=180db4102321> (accessed: 20.01.2024)

streaming service mandatorily includes, apart from the phonogram with or without text, the graphical cover and sometimes auxiliary copyrighted works that, together with the phonogram, make up an integral artistic form.

The paper presented is focused on multidisciplinary phenomenological analysis of human voice as exemplified by singing voice to propose, based on analytical findings, adequate and viable approaches to legal protection of this intangible personal good. Achieving this purpose will require to address a number of particular research objectives. The authors will need to analyze how vocaloid-like artificial neural networks and similar tools for cloning individual voice prints make the legal regulation of this problem highly relevant. There is a need to explore research publications that conclude on the open nature of voice parameters and argue that voice absorbs new characteristics resulting from the studies of articulatory and artistic expressive vocalization practices. There is a need to consider the view the basic list of voice parameters does not qualify this phenomenon clearly enough and therefore cannot provide a reliable basis for the underlying legal provision. Finally, authors will need to posit a hypothesis of *vocal identity* as a replacement for the legal concept of *voice* covering a wider parametric range of individual vocal expression, as well as propose a methodology for its legal and technological protection.

## **1. Findings**

### **1.1. Legally Meaningful Knowledge of Human Voice as a Phenomenon**

In 2006, Yamaha Corporation has marketed a novel technology under the name of *vocaloid* based on its proprietary method of algorithmic synthesis of artistic voices. Under this technology popularized by cartoon artist Hatsune Miku's concert shows, a target artist's voice print is digitized by the neural network across a number of preset parameters including the individual voice timbre, physiological peculiarities of vowel and consonant articulation, characteristic breathing and vocalization, as well as other vocal parameters associated with characteristic pronunciation, rhythmic and dynamic syllabic intonation of musical and speech phrases depending on emotional coloring and meaning of musical messages that are subsumed into the artist's recognizable individuality.

To the phenomenon of the Japanese hologram Hatsune Miku is addressed by a number of studies with the following interest for theme under

discussion [Drenten J., Brooks G., 2020: 1319–1323]; [March L., 2023: 894–910]; [Watt D., Harrison P., Cabot-King L., 2020: 172–180].

A. Jones and R. Bennett in their book *Digital Evolution of Live Music* explain the progress of voice synthesis technologies as a challenging trend of participatory culture behind the strife of anthropomorphic musicians to bring digital technologies to performance practices.

Apart from a popular J-pop performer, the book refers to Tupac Shakur's super-realistic hologram with the synthesized voice timbre recreated in 2012 by digital technologies at the Coachella Festival especially for the musician's fans. The authors consider Tupac Shakur's digital "resurrection" by voice synthesis technologies from existential standpoint of "life after death" while noting the technological promise of replicating the "Cher effect" in the algorithmic vocal synthesis and processing. The researchers find that musical artists willingly promote the introduction of digital technologies into performative practices, with fading boundaries between live and virtual performance to result in new legal and ethical implications [Jones A., Bennett R. et al., 2015: 14, 55–70].

In her book *Listening, Timbre and Vocality in African American Music*, Nina Eidsheim, following a subjectivist approach to perception of sound, argues that anyone listening to vocal performance acts as a listener and a vocalist at the same time [Eidsheim N.S., 2019: 186]. In support of the argument, she invokes the pedagogical practice of regular listening to reference records to develop a vocal style imitating the singer of choice, such as Billy Holliday's timbre, and observes, contrary to the generally acknowledged view of Deepfake as highly developed technology [Cheng H., Guo Y., Wang T. et al., 2023: 1–22], that under otherwise equal acoustic parameters a human agent's imitation of someone's vocal style will be always recognizable and different from a neural network's (vocaloid's) imitation due to complexities of harmonic and emotional nature of vocalization.

As a psychophysiological communicative function of human body, voice has been studied since antiquity in the context of biology, physiology and semiotics. Voice can be formally described as human ability to use the vocal tract for producing sound in conversation and singing [Watt D., Harrison P., Cabot-King L., 2020: 137-170]. In his monograph *Voice Development. Vocal control and Training*, V.V. Emelyanov, author of the phonopedic voice development methodology, identifies two types of voice: speaking voice and singing voice, and argues that singing voice will evolve in vocal

training from general forms of movement, singing patterns, intervals and simple vocalization sequences [Emelyanov V.V., 2023: 21]. It suggests that a neural network can perceive a reference timbre precisely in process of analyzing specific vocal and intoning patterns decomposed to identify the characteristics important for cloning the artist's original timbre.

In calling *voiceology* the discipline integrating multidisciplinary ideas of human voice, L.B. Rudin has identified a number of its conceptual branches such as anatomic functional arrangement of vocal apparatus; physiology and acoustic principles of phonation; notion of voice as a means of communication; voice genetics; notions of singing and speaking voice; classification of voices including children's; voice hygiene for vocal professionals; substance and basics of vocal methodology, rhetoric, scenic speech, oral speech technology; concept of vocal talent; basics of creative process psychology, etc. [Rudin L.B., 2009: 5–8]. The findings of Rudin's theory agree with those of other studies of human voice in identifying its main parameters such as pitch, timbre and tone. Voices pitch, timbre and tone were explored by a number of specialists and drew a parallel between vocal expression categories and generalized expression of musical language [Kolonei V., 2017: 18–31]; [Stulova G.P., 2014: 90–104]; [Toropova A.V., 2017: 394–405]; [Umetalieva-Bayalieva Ch. T., 2015: 204–207]. Sometimes they compare metaphorically singing voice with a musical instrument

It is noteworthy that, despite an expanded list of individualizing attributes of voice found in some thematic studies [Efthymiou et al., 2024: 117–134], most authors have accepted the said three main characteristics.

Facial expressions and voice are always among the parameters that neural networks assess to recognize human emotions. Thus, I. Makarevich in the article *Prospects of technologies for recognition of human emotions by facial expressions and voice* [Makarevich I.V., 2022] notes, in analyzing a number of technological solutions for recognition of emotions from voice including libraries (Open SMILE, Delta) and applications (Emotion AI, Beyond Verbal, Cogito, Sense NEMESYSKO), that software will recognize emotions from voice across a wider set of criteria than the three main characteristics.

In an experiment to assess personal character traits from voice recordings through blind listening to a fairy tale recited by different people, several specialists [Khrisofanova L.A., Diveeva A.S., 2017: 157–164] have proved the potential for aurally determining psychological characteristics like ex-

troversion/introversion; attachment/detachment; control/ naturalness; emotionality/impartiality; playfulness/pragmatism. The experiment reveals that human voice has a characteristically wide and open-ended range of attributes.

## 1.2. Hypothesis of Legal Protection of Vocal Identity

The above necessitates an evidence-based argument that a wider and more comprehensive concept of vocal identity embracing audio or vocal attributes is preferable to that of voice when discussing protection of voice as a personal good including for creative artistic use.

The evidence that this approach is viable can be found in legal practice and mass media that periodically report disputes on cloning the sound patterns of artists working in genres that cannot be fully qualified as vocal since they do not rely on unique vocal talent, ability to sing music notes, and mastery of what is considered to be the basic characteristics of voice (pitch, timbre and tone).

One such dispute involved Jay-Z and Roc Nation, an agency producer, that in 2020 demanded to delete deepfake video clips featuring the rap artist Jay-Z performing Billy Joel's song *We Didn't Start the Fire* and Hamlet's monologue *To Be or Not To Be* from William Shakespeare's eponymous play. The agency explained its protest in terms of illegal "vocal impersonation" of the artist they protected.

Rap and hip-hop artists can hardly be called singers in the classical sense meaning those who fully master singing skills and voice. They recite rather than sing, that is, rhythmically read verses, each in his unique manner of cadenced articulation. Vocal identity of these artists may be manifested and described in terms of other attributes auxiliary to the basic characteristics of singing voice that we will explore below. The same is true for other individuals such as politicians, movie stars and other celebrities who are not singers but whose vocal identity is exposed to cloning<sup>7</sup>.

In 2024, a 15-second original voice recording would suffice for Open AI's simulator to perform voice impersonation (also referred to as voice

---

<sup>7</sup> See for details: AI voice clones mimic politicians and celebrities, reshaping reality // The Washington Post. 15.10. 2023. Available at: <https://www.washingtonpost.com/technology/2023/10/13/ai-voice-cloning-deepfakes/> (accessed: 10.01.2024)



copying, cloning and synthesizing) <sup>8</sup>. Violations of individual rights by voice cloning are increasingly described as “vocal impersonation”.

The contemporary Russian Free Encyclopedia *Traditsiya* defines the term “impersonation” (from French em- + Lat. persona + -tion) as “imitation or theatrical cloning of a personality to pass for another person”. Impersonation is understood as a situation of taking someone’s name or personality without permission with the purpose of harming, cheating, intimidating or threatening. Therefore, “vocal impersonation” amounts to copying/cloning the characteristics of speaking or singing voice to replicate it.

The substance of such phenomenon as “vocal impersonation” underlying disputes on illegal manipulations of human voice is defined in terms of copying, replication and use of synthesized voice in various forms and for various purposes. It is noteworthy that this synonymic chain matches the terminology of copyright provisions. Under Article 1270 of the Civil Code of Russia, the right to copy, replicate and use creative outcomes, once voice or more exact definition of this phenomenon is acknowledged as such, is owned exclusively by the author whose consent is required for these actions to be legitimate.

It would be logical to draw a legal parallel between human voice or vocal identity as an unalienable and inseparable personal good, and the physical human appearance, the only difference being that, while the former is not governed by law, the latter is protected by Article 152.1 of the Civil Code. To proceed further, we well need to discuss legal constructs theoretically applicable for protecting voice and its cloning technologies.

### **1.3. Legal Constructs for Protecting Voice and its Cloning Technologies**

The first technological solution of voice synthesis — algorithmically driven vocal engine Yamaha Vocaloid — was an outcome of research performed by a group of students at the Pompeu Fabra University in Barcelona and patented as invention<sup>9</sup>. Based on Vocaloid software core licensed from

---

<sup>8</sup> Open AI’s Voice Cloning AI model only needs a 15-second sample to work // The Verge. Artificial Intelligence. Tech. Open AI. 30.03.2024. Available at: <https://www.theverge.com/2024/3/29/24115701/openai-voice-generation-ai-model> (accessed: 10.01.2024)

<sup>9</sup> Yamaha Corporation’s Voice Synthesizing Method And Voice Synthesizing Apparatus. U.S. Patent No. 10,002,604.



Yamaha, niche firms operating in audio and music markets created a variety of business applications. Other global technological giants in digital signal processing (DSP) of sound, speech and music also conducted research to obtain patents for similar successful solutions in this field.

As for legal protection, patents for vocaloids were awarded at the turn of the century simultaneously to several leading technological companies such as Yamaha Corp.<sup>10</sup>, Texas Instruments Inc.<sup>11</sup>, Sony Corp.<sup>12</sup>, and a number of less important firms and private inventors of technological solutions for computer-enabled synthesis of vocal musical performances. At the moment, the first wave patents are either expired or close to expiry. With explosive growth of technologies for computer-enabled and automatic creation of singing voice and music, these solutions have ceased to be breakthrough innovations, only to become widespread, standard and affordable — that is, commodified — products by 2024. The market abounds with alternatives that offer flexible terms of commercial use.

While still filing for patents to protectable solutions in this field, companies and individual inventors proceed to monetize vocaloids as copyrighted assets in the form of computer software under licensing agreements. To encourage legal fee-based use of vocaloids, patent holders are building up their functionalities and creating extensive plugin libraries for wider creative potential. Meanwhile, the problem of illegal use of “hacked” vocaloids, just like other applications, remains unresolved. Overall, it can be stated that creators and patent holders of vocaloids and related products increasingly prefer to protect and monetize them as computer software, that is, through copyright rather than patent law.

Progress of vocaloid-like neural networks has brought to the fore the legal problem of legitimate use of someone’s voice or vocal identity for training neural network generators. The most revealing situations are those where voices of popular actors or other celebrities are used to attract attention to and bolster up the rating of intelligent content synthesizers. Mass media increasingly report the claims by vocal artists and other owners of

---

<sup>10</sup> Yamaha Corp. Singing voice synthesis apparatus and method. Japan Patent JP3265995B2. Status: Expired.

<sup>11</sup> Texas Instruments Inc. Singing voice synthesis. U.S. Patent No. US6304846B1. Status: Expired.

<sup>12</sup> Sony Corp. Singing voice synthesizing method, singing voice synthesizing device, program, recording medium, and robot. European Patent Office EP1605436A1. Status: Expired.

recognizable voices against third parties alleging illegal vocal impersonation of creative works they produce, unauthorized use of voice or vocal articulation features.

A fresh example of such disputes is the actress Scarlett Johansson's claim against Sam Altman, OpenAI CEO, for allegedly unauthorized integration of her voice into AI-enabled software ChatGPT4.0. In his turn, Altman denied that the voice was that of Johansson claiming to have invited a different actress, whose name was not to be disclosed under the contractual terms, after Johansson had refused to grant her voice to ChatGPT4.0 [Elias, 2024]. Meanwhile, beyond celebrities and politicians, the voice cloning technology is increasingly used to fraudulently appropriate ordinary people's voice to delude interlocutors and compel them to a certain course of action.

In Russia the problem of protecting voice as intangible and unalienable personal good, has reached the level of legislative initiatives that in most cases assume the introduction of legal protection along the lines of what is afforded to human image under Article 151.1 of the Civil Code. However, it is worth noting that wider provisions of Article 150 can be used to address the problem since among intangible goods defined in part 1, Article 150, a person's voice or vocal identity can be rightly regarded as "unalienable and otherwise non-transferrable intangible good owned by the person by virtue of birth or by law". As such, it will be covered by "protection under the effective provisions of the Civil Code and other laws in cases and under procedure defined therein" pursuant to part 2 of this Article.

Efforts towards protecting voice or vocal identity as unalienable intangible good appears to be timely response of the legislator to the emerging problem. The chosen legal drafting methodology to regulate such relationships with like means looks quite logical. Obviously, claims against unauthorized use of voice will be regulated by courts. To provide and assess the evidence in legal proceedings, the parties will need to define what is voice in terms of specific parameters, features and characteristics in order to establish the fact of cloning or borrowing through examination and comparison of voice and vocal articulation samples.

#### **1.4. Parametrization of Human Vocal Identity**

Strength, pitch and timbre are associated with the main parameters of voice. Vocal strength is understood as the dynamic range and vocal sustainability that follow from natural abilities such as physical endurance, respira-

tory system operation and elasticity of vocal chords. While not widely used in professional vocal pedagogy, the term “vocal strength” is quite current in routine descriptions to provide more particular characteristics of “dynamic range” of voice. The latter can be variable depending on vocalization type (purpose and emotional coloring of human verbal communication or objectives and emotional nature of vocal intonation in singing). Mastering rhetoric and vocal performance techniques assumes control of dynamic range by changing vocalization loudness (high-moderate-low) and intensity.

As another important characteristic of voice, “pitch” means the main tonality of speech and ability to change it in process of speaking and singing. Pitch can vary depending on natural vocal abilities, develop in process of regular vocal exercising and alter over a lifetime as being physiologically conditioned, among other things, by hormone and age-specific changes. A distinction is made between speaking and singing pitch. Speaking pitch will normally cover one octave and, depending on intonation patterns, purpose-specific type of sentences and emotional charge, can include 3-4 notes. In contrast, singing pitch is determined, along with psychophysiological parameters, by the timbre characterized by unique frequency parameters of the singer’s individual sound for each specific note of different ranges. Quantitative characteristics of singing pitch are individual, with vocalists able to cover between 1.5 to 3 octaves to include, depending on the timbre, the notes of low, medium and high range.

Finally, “timbre” of voice means its inimitable individual frequency profile defined by the comprehensive structure of vocal tract comprising cavities and sound barriers in vocal apparatus that affect the overtones formed in inferior (trachea, bronchia) and superior (oral, nasal cavities and front zones) resonators. While vocal exercises and techniques can help to control superior resonators (oral and nasal cavities), trachea and bronchia are subordinated to higher nervous functions. Perfecting speaking and vocal techniques will make speech more expressive and convincing. Apart from the main parameters of voice, researchers also single out auxiliary emotional and functional attributes such as “airiness”, “fluidity”, “tonality”, “sonority” etc. In our view, the said parameters are quite measurable and, taken together, allow to compare effectively different voices to establish or deny their identity.

In vocal art where voice is a creative musical instrument, there is a need for clear and non-ambiguous understanding of the subject of regulation. For complex phenomenon such as singing voice, the task to identify a legal-

ly meaningful subject becomes considerably more complicated due to the following peculiarities. With singing voices divided into male and female in terms of range, these two distinctions alone will produce, according to Ivanov and Nikiforov [Ivanov A.P., Nikiforov V., 2006], six varieties further breaking down into more than twenty main types that can ramify without limit. Providing a parametric description of all singing voices for the judge to compare not abstract options of experts rushing to establish the case for either plaintiff or defendant but numeric values in search for the evidence of voice cloning is obviously a challenge.

While parametrization of singing voices for the purpose of comparison in the event of cloning claims appears challenging but feasible, the real problem is digitization of the vocal signature of those performers whose form of creative expression relies on vocal articulation techniques rather than vocal range and frequency — that is, where artistic expressivity is achieved not by conventional singing but peculiarities of pronunciation, speaking, recitation, original vowel combination or, conversely, phoneme division in the song's lyrics. Such vocal techniques dating back to early high gain microphones<sup>13</sup>, are now a major feature of musical art since they do not only underpin a number of popular genres such as rap and hip-hop but also fragmentarily permeate practically all musical and vocal styles.

Importantly, phonetics will identify vocalization properties as articulatory acoustic features not to be ignored in assessing someone's vocal individuality. Accounting for these features in combination with basic parameters of sound production will allow to more precisely characterize and distinguish the vocal identity of ordinary person or artist than voice would. Without going into details of these features, it has a sense to characterize some of them to catch an idea of their essence and relevance for this study.

Acoustic features of human articulation can subsume a specific way of pronouncing noise consonants (also called obstruents) produced by obstructed airflow causing higher air pressure in the vocal tract. In phonetics and phonology, articulation is divided into two main classes of noise and sonorant consonants.

Human persons also differ in the way they pronounce sonorants — sounds produced without causing turbulence in the vocal tract. These in-

---

<sup>13</sup> Gioia T. How Bing Crosby Made Silicon Valley Possible. *The Honest Broker*. 21.12.2023. Available at: <https://www.honest-broker.com/p/how-bing-crosby-made-silicon-valley> (accessed: 16.04.2024)

clude approximants, nasal consonants, single stress and trill consonants. There is a great deal of nuanced articulation of approximants or spirant sonorants with their specific way of production.

Nasal consonants produced by depressed soft palate with airflow passing through the nose are also widely specific. Practically each individual has his own way of articulating single stress consonants or flaps produced by a single flip of articulatory organ.

Society exhibits widely variable articulation of trill consonants (vibrants) produced by articulatory organ's vibration in the articulation zone. For example, the Spanish language sound transcribed by the digraph *rr* means a trill alveolar consonant, with the International Phonetic Alphabet featuring a variety of trill consonants such as alveolar trill, bilabial trill, uvular trill and epiglottal trill. Vibration patterns of trill consonants vary within a very broad range and amplitude as capable of comprising two to six periods with different fading dynamics.

### **1.5. Methodology for Protecting Vocal Identity**

It is the combination of these and other more specific features of acoustic articulation that underpins creative work in vocal, musical and audiovisual genres. Thus, "voice" may be a wrong concept to use for protecting the interests of its owner by legal means since it does not reflect the full scope of the problem and, therefore, will not provide a basis for efficient mechanism to regulate the relevant relationships. A legal provision relying on irrelevant concept will not work. For this reason, the category defined above as the artist's vocal identity appears more appropriate for vocal identification and effective resolution of voice cloning and unauthorized borrowing claims.

Vocal identity of an artist can be defined as a sum of performative vocal features including a combination of singing voice parameters and characteristic articulatory features of vocalization that make up the artist's creative technique.

It follows from this definition that singing and phonetic features could be interactive, mutually determining and intermixing to produce an original artistic technique. Another corollary is the extreme difficulty to identify and separately analyze the attributes relating to singing on the one hand and articulation on the other. Moreover, analytical approach appears counter-productive in this case. It is a combination of interdependent characteristics that makes up the unique artistic technique to be meaningfully registered in the form of the established industry standard — so-called audio signature

or frequency-time spectrogram obtained by the well-known Fourier transform method and reflecting all vocal patterns [Amiot E., 2016]. For comprehensive vocal portrait of an artist, several such patterns will be needed.

In a claim against unauthorized appropriation of what can be called in the law as “voice” or, as proposed in this paper, “vocal identity”, these patterns can be used to establish the identity and the fact of borrowing (impersonation or cloning of this unalienable personal good).

As regards the technology to establish or dismiss the vocal identity claim, this task could be best performed by AI neural network. Trained on original vocal signatures of artists, such neural network can take into account the whole set of interdependent pitch frequency and articulatory features of artistic performance techniques and reliably establish the degree of test sample identity.

## **Conclusion**

Current progress of AI technologies and generative neural networks for cloning and synthesis of human voice and automatic production of creative outputs in musical, vocal and sound arts have prioritized the protection of human voice and singing voice. In this regard, vocaloid and vocaloid-like tools have become especially relevant as opening up new creative opportunities based on performance methods that did not exist before and transforming the traditional forms of musical and vocal culture.

While singing voice is normally characterized by a combination of basic parameters, vocal musical genres do not rely on singing voice alone. In particular, relatively recent styles such as rap and hip-hop as well as other recitative arts that have won a large share of the music market are influencing classical musical genres. These performance approaches are based on peculiarities of phonetic articulation, pronunciation and rhythmical beat-driven recitation rather than vocal art described in terms of basic parameters of singing voices.

Because of this discourse, it does not appear adequate to use voice as a category for drafting legal provisions. The authors propose to use instead vocal identity as allowing to embrace more parameters of human vocal individuality for more efficient protection.

The study contains a methodology for resolution of disputes on cloning someone’s vocal identity (frequently defined as vocal impersonation)

by using specifically trained generative neural networks that can take into account a combination of multiple interdependent parameters of human vocal identity.

Authors hope their proposals supported by multidisciplinary analytical findings related to the phenomenon of human voice will produce purpose for addressing the problem of legal protection of this personal good in the context of explosive progress of generative neural networks for musical and audiovisual products.



## References

1. Abel J. E. (2022) *The New Real: Media and Mimesis in Japan from Stereographs to Emoji*. Minneapolis: University of Minnesota Press, 360 p.
2. Amiot E. (2016) *Music Through Fourier Space*. DOI:10.1007/978-3-319-45581-5.
3. Anisimova S.V., Golovaty A.I. (2021) Musical Fragments, Notes and Chord Recognition Algorithms. *Vestnik Polotskogo gosudarstvennogo univeristeta. Seria S. Fundamentalnye nauki*=Polotsk State University Bulletin. Series S. Fundamental Sciences, no. 4, pp. 45–49 (in Russ.)
4. Aznauryan L.A. (2017) Study in models and algorithmic methods for analysis of musical works. Papers of Annual Research Conference PhysicMinneapolis:al, Mathematical and Natural Sciences. Yerevan: Russian-Armenian (Slavonic) University, pp. 67–76.
5. Balashova A.I. (2022) AI in copyright and patent law: range of objects and subjects, legal protection term. *Zhurnal suda po intelektualnym pravam*=Journal of Intellectual Property Rights Court, no. 2, pp. 90–98 (in Russ.)
6. Beliaevskiy D.M., Darbinyan S.S. et al. (2016) Digital Identification of Objects: Beyond Technologies. Moscow: Nauchnoe obozrenie, 252 p. (in Russ.)
6. Bridy A. (2016) The evolution of authorship: works made by code. *Columbia Journal of Law and the Arts*, vol. 39, pp. 395, 397.
7. Boden M. (2004) *The creative mind. Myths and mechanisms*. London–New York: Routledge, 344 p.
8. Calo R. (2014) Robotics and the Lessons of Cyberlaw (2014) *California Law Review*, vol. 103, no. 3, pp. 513–563. University of Washington School of Law. Research Paper No. 2014-08.
9. Cheng H., Guo Y., Wang T. et al. (2023) Voice-face homogeneity tells deepfake. *ACM Transactions on Multimedia Computing, Communications and Applications*, no. 3, pp. 1–22.
10. Davis H. (2014) Creating music from literature. Papers of EACL Workshop on computer linguistics for literature, pp. 1–10. arXiv: 1403.2124. Bibcode: 2014arXiv1403.2124D. 10.3115/v1/W14-0901. S2CID 9028922. DOI: 10.3115/v1/W14-0901.S2CID9028922.



11. Davis H., Mohammad S. (2014) Generating Music from Literature. In *Proceedings of the Workshop on Computational Linguistics for Literature*. Gothenburg: Association for Computational Linguistics, pp. 1–10.
12. Deleuze G. (2011) *Logique du sens*. Moscow: Akademicheskii Proekt, 472 p. (in Russ.)
13. Denikin A.A. (2013) Model of diegetic analysis of the sound in screen media. *Mediamuzyka research e-journal=Media Music Research E-Journal*, no. 2. (in Russ.)
14. Drenten J., Brooks G. (2020) Celebrity 2.0: Lil Miquela and the rise of a virtual star system. *Feminist Media Studies*, no.8, pp. 1319–1323.
15. Eck D., Schmidhuber J. (2002) Finding temporal structure in music: Blues improvisation with LSTM recurrent networks. In: *Proceedings of the 2002 IEEE Workshop on Neural Networks for Signal Processing*, pp. 747–756.
16. Eidsheim N. S. (2018) Widening Rings of Being: Singer as Stylist and Technician. In: *The Race of Sound: Listening, Timbre, and Vocal in African-American Music*. Durham: Duke University Press, pp. 177–200. Available at: <https://doi.org/10.2307/j.ctv11hpntq.10> (accessed: 20.05.2024)
17. Emelianov V.V. (2023) *Development of Voice. Manual for Vocal Control and Training*. Saint Petersburg: Lan' Press, 168 p. (in Russ.)
18. Ivanov S.M. (2018) Neural network-generated polyphonic music. *Papers of Lomonosov Conference Computational mathematics and cybernetics panel*. Moscow: MGY Press, pp. 110–111 (in Russ.)
19. Ivanov A.P., Nikiforov V. (2006) *Vocal Art*. Moscow: Golos-Press, p. 37 (in Russ.)
20. Jones A., Bennett R., Cross S. (2015) 10 — Keeping it real? Life, death, and holograms on the live music stage. In: A. Jones, R. Bennett (eds.). *The Digital Evolution of Live Music*. Whitney: Chandos Publishing, pp. 123–138.
21. Khachatryan A.A. (2023) AI neural network as a tool for recognition of musical genres. *Applied mathematics: current problems of mathematics, computer science and simulation: papers of the All-Russia research workshop of young scientists*. Krasnodar: Russian Energy Agency Press, pp. 479–483 (in Russ.)
22. Khrisanfova L.A., Diveeva A.S. (2017) Perceiving the peculiarities of human psychology from face photos and voice recordings. *Yearbook of Psychological and Psychoanalytical Studies*. Moscow: Institute of Psychoanalysis, pp. 157–164 (in Russ.)
23. Kolonei V.A. (2017) Function of musical intonation components. *Muzykalnoe iskusstvo=Music Art*, no. 16, pp. 18–31 (in Russ.)
24. Levy D. (2005) *Robots Unlimited: Life in a Virtual Age*. Available at: <https://doi.org/10.1201/b10697> (accessed: 28.06.2022)
25. Lund J. (2012) *An Empirical Examination of the Lay Listener Test in Music Composition Copyright Infringement*. Available at: SSRN: <https://ssrn.com/abstract=2030509> or <http://dx.doi.org/10.2139/ssrn.2030509> (accessed: 16.05.2022)

26. Makarevich I.V. (2022) Prospects of technologies for recognition of human emotions by facial expressions and voice. Papers of Regional Conference for young researchers. Volgograd: State Technological University Press, pp. 156–157 (in Russ.)
27. Malakhov V.E. (2022) Applying decomposition to study generative music creation processes. *Informatsionnye tekhnologii i telekommunikatsii*=Information and Communication, vol. 10, no. 4, pp. 69–78. DOI: 10.31854/2307-1303-2022-10-4-69-78 (in Russ.)
28. March L. (2023) Wrap You Up in My Blue Hair: Vocaloid, Hyperpop, and Identity in “Ashnikko Feat. Hatsune Miku — Daisy 2.0”. *Television and New Media*, no. 8, pp. 894–910. Available at: <https://doi.org/10.1177/15274764221093599> (accessed: 20.12.2023)
29. Markina Yu.Yu., Belov Yu.S. (2018) Spectral coefficients as a necessary parameter in creating human voice cloning systems using deep learning methods. *Mezhdunarodnyi studencheskiy nauchnyi vestnik*=International Bulletin of Student Research, no. 1. Available at: URL: <https://eduherald.ru/ru/article/view?id=18125> (accessed: 26.01.2023)
30. Mazzei L.A. (2016) Voice Without a Subject. Cultural Studies. Critical Methodologies, no. 2, pp. 151–161. Available at: <https://doi.org/10.1177/1532708616636893> (accessed: 16.07.2024)
31. Nikolskiy A.V., Alexeev E.E. et al. (2019) A Jew’s Harp Story: Jew’s harp and individual song as a basis of tone-oriented musical systems. *Yazyki i folklor korennykh narodov Sibiri*=Languages and Folklore of Aboriginal Siberian Peoples, no. 37, pp. 5–32. DOI: 10.25205/2312-6337-2019-1-5-32 (in Russ.)
32. Perduto P. (2022) NFT — Income for Creative Minds. How to Generate Additional Income as a Photographer, Graphic Designer, Composer, Musician or other Art Creator. S.L.: Books on demand Publishers, 156 p.
33. Rais M. (2016) Musical modes: classification, characteristics, properties, history, current status. *Israel XXI. Music Magazine*, no. 3, p. 9 (in Russ.)
34. Ramalho A. (2017) Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems. Available at: <http://dx.doi.org/10.2139/ssrn.2987757> (accessed: 12.04.2023)
35. Rinkerman G. (2023) Artificial Intelligence and evolving issues under US copyright and patent law. *Interactive Entertainment Law Review*, no. 2, pp. 48–65. Available at: <https://doi.org/10.4337/ielr.2023.0002> (accessed: 19.05.2024)
36. Rudin L.B. (2009) Principles of Voiceology. Moscow: Granitza, 104 p. (in Russ.).
37. Sivolap T.E. (2019) Audiovisual works: peculiarities of legal protection. Current issues of radio and cinema technologies: papers of International Conference of Technological Research to the Centenary of the Cinema and Television Institute. D.P. Barsukov et al. (eds.). Saint Petersburg: State Institute of Cinema and Television Press, pp. 237–242 (in Russ.)
38. Stulova G.P. (2014) Pedagogical factors affecting the singing voice timbre. UNESCO Chair of Musical Art and Education Bulletin, no. 4, pp. 90–104 (in Russ.)

39. Tikhomirova N.F. (2021) Conceptual principles of classification in the analysis of creative products of musical culture. *Vestnik Donetskogo natsionalnogo Universiteta*=Donetsk National University Bulletin, no. 1, pp. 123–128 (in Russ.)
40. Tormozova V.A. (2023) Current issues of audiovisual works as an object of copyright. *Trudy po intellektualnoy sobstvennosti*=Works on Intellectual Property, no. 4, pp. 54–67. DOI: 10.17323/tis.2023.18213 (in Russ.)
41. Santiago J. (2017) The “Blurred Lines” of Copyright Law: Setting a New Standard for Copyright Infringement in Music. *Brooklyn Law Review*, vol. 83, pp. 289–321.
42. Toropova A.V. (2017) Musical sensation as a basis of efficient musical therapy A study in cross-cultural interpretation of L.S. Vygotsky’s psychological framework in context of variable models of musical therapy. In: Art studies in the context of other sciences in Russia and elsewhere: parallels and interactions. Papers of the international research conference. Moscow: Soglasie Press, pp. 394–405 (in Russ.)
43. Umetalieva-Bayalieva Ch.T. (2015) Musical language as an object of cultural studies. The means and elements of musical expression (musical language). *Nauka i novye tehnologii*=Science and New Technologies, no. 2, pp. 204–207 (in Russ.)
44. Vanyukov P.D. (2023) The legal regime for AI and its outcomes. *Intellektualnaya sobstvennost*=Intellectual Property, no. 3, pp. 58–68 (in Russ.)
45. Vasilevskaya L.Yu. et al. (2021) Digitizing Civil Law Transactions: Legal Characteristics of “Artificial Intelligence” and “Digital” Subjects (a civil study in 5 vols.). Vol. 1. Moscow: Prospekt, 288 p. (in Russ.)
46. Watt D., Harrison P., Cabot-King L. (2020) Who owns your voice? Linguistic and legal perspectives of relationship between vocal distinctiveness and the rights of the individual speaker. *International Journal of Speech, Language and the Law*, no. 1, pp. 137–180.
47. Yuan J.E. (2024) Making virtual celebrity: Platformization and intermediation in digital cultural production. DOI:10.1177/13678779241230564.
48. Zakharov A.O. (2020) Neural networks in music. Week of Science and Creativity. Papers of international research workshop for students, postgraduate students and young scientists. Saint Petersburg: State Institute of Cinema and Television Publishers Press, pp. 36–39 (in Russ.)

---

#### **Information about authors:**

R.A. Budnik — Doctor of Sciences (Law), Professor.

E.G. Evpak — Researcher.

The article was submitted to editorial office 18.03.2024; approved after reviewing 14.05.2024; accepted for publication 24.06.2024.