UOT 34:343.1. Cinayət mühakimə icraatı, s.88-92

AFET MUSAYEVA

Expert of the Department of Manuscript and Audio
Expertise of the Forensic Center of the Ministry of Justice
E-mail: afet.musayeva@justice.gov.az
DOİ: https://doi.org/10.64498/IWLE9172

APPLİCATİON PROSPECTS OF ARTİFİCİAL İNTELLİGENCE AND NEURAL NETWORKS İN PHONOSCOPİC ANALYSİS

Açar sözlər: Neyron şəbəkə texnologiyaları, generativ modellər, məxfilik və audio nəzarət, RNN və LSTM modelləri, konvolyusiya neyron şəbəkələri, AI əsaslı ekspert rəylərinin qəbulunda problemlər, problemlər və risklər, gələcək perspektivlər.
Keywords: Neural network technologies, generative models, privacy and audio surveillance, RNN and LSTM models, convolutional neural networks, challenges in the acceptance of AI-based expert opinions, challenges and risks, future perspectives.
Anahtar kelimeler: Yapay sinir ağı teknolojileri, üretken modeller, gizlilik ve ses gözetimi, RNN ve LSTM modelleri, evrişimli sinir ağları, yapay zeka tabanlı uzman görüşlerinin kabulündeki zorluklar, zorluklar ve riskler, gelecek perspektifleri.
Ключевые слова: технологии нейронных сетей, генеративные модели, конфиденциальность и аудионаблюдение, модели RNN и LSTM, сверточные нейронные сети, проблемы принятия экспертных мнений на основе ИИ, проблемы и риски, перспективы на будущее.

For many years, phonoscopy has been applied as one of the key areas of forensic examination, particularly in the identification of voices, authentication of phonograms, and study of the acoustic features of speech. Traditionally, the methods employed in this field were mainly limited to measuring acoustic parameters, visually comparing spectrograms, and conducting instrumental analyses. However, the rapid advancement of information technologies, particularly the emergence of artificial intelligence (AI) and neural networks, has brought a completely new dimension to phonoscopy. AI technologies make it possible to process large volumes of audio data with high precision, recognize individual voice biometrics, and detect manipulated audio files. Neural networks, in turn, possess the ability to learn even the most complex phonetic and acoustic features of human speech. This makes them an invaluable tool not only in forensic examinations but also in the fields of national security and combating cybercrime.

At the same time, the application of AI in phonoscopy also raises a number of challenges and controversial issues. The development of deepfake technologies, the risk of erroneous outcomes from neural networks, as well as ethical and legal limitations, have become subjects of serious debate in this field.

Therefore, this article will analyze the role of AI and neural networks in forensic phonoscopy, outlining their advantages, risks, and future development prospects.

Artificial intelligence enriches the classical methods of phonoscopy with modern approaches that provide more accurate, faster, and more reliable results. Whereas traditional methods required manual measurement of acoustic parameters and visual comparison of spectrograms, AI technologies automate this process, reducing workload while increasing objectivity in results. Voice is regarded as a unique biometric characteristic inherent to each individual. AI algorithms, particularly neural networks, analyze personal voice parameters (such as timbre, pitch, intonation, and articulation features) with high accuracy. As a result, comparing numerous voice samples and identifying an individual becomes significantly easier. These methods greatly accelerate the forensic examination of telephone conversations and audio recordings in criminal cases.

One of the most common problems in phonoscopy is related to the editing and manipulation of phonograms. In classical methods, this often requires

lengthy analysis and does not always produce accurate results. Artificial intelligence, however, can detect even minor inconsistencies in the spectral, temporal, and amplitude indicators of sound. For instance, neural networks can immediately identify background noise changes, waveform interruptions, or rhythmic irregularities in edited sections.

One of the most complex challenges in phonoscopy is the accurate assessment of speaking style, dialects, and accents. Modern AI systems, trained on large datasets in various languages and dialects, are capable of determining the regional and social background of a speaker. This feature is invaluable both in criminological investigations and in linguistic-related phonoscopic analyses. AI systems can analyze large volumes of audio files within seconds, which is of exceptional importance in court cases involving numerous materials. Moreover, since neural networks minimize the influence of subjective factors, they increase the objectivity of results in disputed cases.

Neural networks are among the most powerful directions of artificial intelligence, and their application in phonoscopy allows the detection of even the subtlest acoustic nuances of sound. Different types of neural network models are selected for different analytical purposes, each offering unique advantages. Convolutional Neural Networks (CNNs), although primarily used in image processing, also play a significant role in phonoscopy. This is because sound signals are often converted into spectrograms and processed as images. CNNs can detect even the slightest differences in these spectrograms that may be imperceptible to the human ear. This method proves highly effective in voice identification, detection of edited segments, and the analysis of emotional states in speech.

Since Recurrent Neural Networks (RNNs) and LSTM models have speech sequencing (time-series) nature, RNNs and their advanced variant, LSTM (Long Short-Term Memory) models are widely applied in phonoscopy. They analyze the intonation sequence, rhythm, and tempo of speech. These models are particularly effective in studying dialect and accent differences, detecting repetitions and pauses in speech, and analyzing long-duration telephone conversations.

Transformer-based Models – in recent years, models built on Transformer architecture (such as wav2vec, Whisper, HuBERT) have come to the forefront in phonoscopy. Since these models are trained on large audio datasets, they demonstrate high accuracy both in transcription (converting speech to text) and in source voice recognition. Their advantages include the ability to operate in multilingual and multi-dialect environments, to produce reliable results even with low-quality or noisy recordings, and to establish automatic alignment between text and audio.

Generative Neural Networks and the Deepfake Problem – generative models (GAN – Generative Adversarial Networks) represent one of the greatest challenges of recent years. Through this technology, completely synthetic yet highly realistic voice deepfake audio files can be generated. This brings both opportunities and risks to forensic phonoscopic examination. On the positive side, such models can expand voice datasets and generate synthetic samples for training. On the negative side, detecting deepfake voices is becoming increasingly difficult, casting doubt on the reliability of audio files presented as evidence in court.

The application of AI and neural networks in phonoscopy is not only theoretical but also widely implemented in real court cases and forensic practice. These technologies facilitate the work of modern law enforcement agencies while enhancing the accuracy of evidence. One of the most common scenarios in legal proceedings is the identification of audio recordings. For instance, in telephone conversations, the suspect's voice can be compared through neural networks for identity verification. Covert voice recordings and audio files used in illegal activities are subject to expert examination. Manipulated evidence is detected by AI systems, which identify cut-and-joined or altered recordings. In such cases, artificial intelligence accelerates the work of experts and eliminates subjective biases. In the field of national security, analyzing the voice messages of terrorists, criminal groups, or foreign intelligence agencies is one of the main directions of phonoscopy. AI technologies are capable of automatically filtering large-scale audio databases, quickly identifying individuals with specific speaking styles, and extracting information from encrypted or poor-quality recordings.

Looking at the legal context, in some countries, AI-based phonoscopic opinions are already being presented in courts as evidence. However, a number of legal issues remain. Reliability – The complete objectivity of AI algorithms' results is still under debate. Accountability – The "black-box" nature of neural networks complicates the explanation of

their outcomes. Regulatory Gaps – In many legal systems, the legal status of AI-based expert opinions has not yet been fully regulated.

It is important to note that phonoscopy is not limited to forensic applications. Artificial intelligence technologies are also widely used in voice biometrics for customer identification, in call-center systems for recognizing customer voices and personalizing services, and in biometric authentication for securing access with voice passwords.

Alongside the innovations that AI and neural networks bring to forensic phonoscopy, their application also introduces certain risks and limitations. Ignoring these may negatively affect the reliability of expert opinions and the proper evaluation of evidence in legal proceedings. One of the most serious challenges in recent years has been related to deepfake technologies. Artificially generated voices created through neural networks resemble real human voices so closely that even the most advanced phonoscopic systems sometimes struggle to distinguish them. This can lead to falsification of court evidence, criminals evading responsibility, and a growing public mistrust in synthetic voices. The application of AI systems also raises concerns about privacy rights. Voice biometrics is a unique characteristic of each individual, and its unauthorized collection or use constitutes a violation of privacy rights. Furthermore, the automatic monitoring of phone conversations, creation of large-scale voice databases, and storage of voice samples without citizens' consent are ethically controversial practices.

In the legislation of various countries, the acceptance of AI-based expert opinions as legal evidence is still not fully regulated. The main risks here include the lack of transparency of AI algorithms, the challenge of how experts should explain AI-generated results, and the variations in regulatory frameworks across different jurisdictions.

As with any technology, AI-based systems also have certain technical limitations. If the training dataset is of poor quality, the model's results may lack accuracy. In noisy or low-quality recordings, the probability of misidentification increases. Neural networks may occasionally produce false classifications, which can, in turn, lead to unjust court decisions.

Although the use of AI and neural networks in phonoscopy is still at an early stage, it is expected that this field will undergo a complete transformation in the coming years. The directions of development will manifest themselves in both technological and legal-ethical dimensions. In the future, the most effective approach in phonoscopy will be a combination of classical acoustic methods with AI technologies. That is, experts will analyze traditional parameters (frequency, amplitude, intonation) while also relying on the refined results provided by AI systems. This hybrid approach will ensure that outcomes are both scientifically grounded and technologically precise.

For the training of neural networks, the creation of large-scale and high-quality audio databases at both national and international levels is essential. In the future, the establishment of national phonoscopic databases in each country, as well as the development of shared international voice banks, may be planned. This would support the unification of methodologies and strengthen cooperation in the investigation of cross-border crimes. Alongside the development of synthetic voice technologies, the creation of systems designed to detect them will also become a priority. AI-based "deepfake detectors" will become an integral part of future forensic phonoscopy. These systems will be capable of automatically detecting spectral inconsistencies, artificial alterations in harmonic structures, and unnatural characteristics in speech dynamics. Since voice biometrics represent highly sensitive data, future efforts will focus on strengthening ethical standards and legal frameworks. This will involve clarifying the conditions under which AI-based expert opinions are accepted in courts, developing international standards for the protection of personal voice data, and improving codes of professional ethics for experts. In the near future, the implementation of semi-automated or fully automated forensic systems will also be possible. Such systems will allow an audio file to be uploaded, conduct a preliminary analysis, generate likely outcomes, and guide the expert in their assessment. This will significantly minimize the influence of the human factor and greatly accelerate the overall process.

Reference list:

1. Musayev H., Əliyev L., Rəhmanov C. "Məhkəmə-fonoskopik ekspertizası", Bakı, 2007.
2. Идентификация звукозаписывающих устройств. Методические рекомендации / под ред. Г.Л. Грановского. – М.: ВНИИСЭ, 1986, 51с.
3. Каганов А.Ш. Судебная экспертиза звукозаписей. М.: Юрлитиинформ, 2005, 272 с.
4. Иванов И.Л. Экспертное исследование формата GSM. Экспертная практика. //http://www.illidiy.orel.ru/Pages/publ.htm
5. Галяшина Е.И. Судебная фоноскопическая экспертиза. М.: Триада; Лтд., 2001, 302 с.
6. ГОСТ 13699-91. Запись и озвучивание информации. Термины и понятия.

ANNOTATION

The application of artificial intelligence and neural networks in phonoscopy has opened new horizons in this field. Compared to traditional methods, it has enabled the achievement of more accurate, faster, and more objective results, while also bringing significant progress in detecting edited or manipulated phonograms. With the use of CNN, RNN, LSTM, and Transformer-based models, biometric identification of human voices, analysis of dialects and accents, and the examination of low-quality recordings are now conducted with greater reliability. Nevertheless, the development of deepfake technologies, issues of privacy and ethics, and the lack of sufficient legal regulations present serious challenges for forensic phonoscopy. For this reason, not only technological advancements but also the refinement of ethical and legal frameworks are essential. The application of hybrid methods, the creation of national and international voice databases, the enhancement of deepfake detection algorithms, and the development of automated forensic systems stand as priority tasks for the future of phonoscopic examination. Altogether, these advancements will ensure that forensic phonoscopy becomes a more powerful and reliable source of evidence, both scientifically and legally.

FONOSKOPİK ANALİZDƏ SÜNİ İNTELLEKT VƏ NEYRON ŞƏBƏKƏLƏRİN TƏTBİQİ PERSPEKTİVLƏRİ

XÜLASƏ

Süni intellektin və neyron şəbəkələrin fonoskopiyada tətbiqi bu sahədə yeni üfüqlər açıb. Ənənəvi üsullarla müqayisədə bu sistemlər daha dəqiq, daha sürətli və daha obyektiv nəticələr əldə etməyə imkan verib, eyni zamanda redaktə edilmiş və ya manipulyasiya edilmiş fonoqramların aşkar edilməsində əhəmiyyətli irəliləyişlər əldə edilib. CNN, RNN, LSTM və Transformer əsaslı modellərin istifadəsi ilə insan səslərinin biometrik identifikasiyası, dialekt və ləhcələrin təhlili, keyfiyyətsiz yazıların yoxlanılması artıq daha etibarlı şəkildə aparılır. Buna baxmayaraq, "deepfake" texnologiyalarının inkişafı, məxfilik və etika məsələləri və kifayət qədər hüquqi tənzimləmələrin olmaması məhkəmə fonoskopiyası üçün ciddi problemlər yaradır. Bu səbəbdən təkcə texnoloji irəliləyişlər deyil, həm də etik və hüquqi çərçivələrin təkmilləşdirilməsinə ehtiyac vardır. Hibrid üsulların tətbiqi, milli və beynəlxalq səs məlumat bazalarının yaradılması, dərin saxtakarlığın aşkarlanması alqoritmlərinin təkmilləşdirilməsi və avtomatlaşdırılmış məhkəmə ekspertiza sistemlərinin inkişafı fonoskopik ekspertizanın gələcəyi üçün prioritet vəzifələr kimi dayanır.

FONOSKOPİK ANALİZDE YAPAY ZEKA VE SİNİR AĞLARININ UYGULAMA BEKLENTİLERİ

AFET MUSAYEVA

Adalet Bakanlığı Adli Merkezin El Yazma ve Ses Ekspertiz Dairesinin Uzmanı

ÖZET

Yapay zekâ ve sinir ağlarının fonoskopide uygulanması bu alanda yeni ufuklar açmıştır. Geleneksel yöntemlere kıyasla bu sistemler daha doğru, daha hızlı ve daha objektif sonuçlara olanak sağlarken, düzenlenmiş veya değiştirilmiş fonogramların tespitinde önemli ilerlemeler kaydedilmiştir. CNN, RNN, LSTM ve Transformer tabanlı modeller kullanılarak insan seslerinin biyometrik olarak tanımlanması, lehçe ve aksanların analizi ve düşük kaliteli kayıtların doğrulanması artık daha güvenilirdir. Ancak "deepfake" teknolojilerinin gelişimi, gizlilik ve etik sorunları ve yeterli yasal düzenlemelerin olmaması adli fonoskopi için ciddi zorluklar yaratmaktadır. Bu nedenle, yalnızca teknolojik gelişmelere değil, aynı zamanda etik ve yasal çerçevelerin iyileştirilmesine de ihtiyaç vardır. Hibrit yöntemlerin uygulanması, ulusal ve uluslararası ses veri tabanlarının oluşturulması, derin sahtecilik tespit algoritmalarının iyileştirilmesi ve otomatik adli inceleme sistemlerinin geliştirilmesi fonoskopik incelemenin geleceği için öncelikli görevlerdir.

ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И НЕЙРОННЫХ СЕТЕЙ В ФОНОСКОПИЧЕСКОМ АНАЛИЗЕ

АФЕТ МУСАЕВА

Эксперт Отдела Почерковедческой и голосовой экспертизы
Судебного Центра Министерства Юстиции

РЕЗЮМЕ

Применение искусственного интеллекта и нейронных сетей в фоноскопии открыло новые горизонты в этой области. По сравнению с традиционными методами, эти системы позволяют получать более точные, быстрые и объективные результаты, а также достигнут значительный прогресс в выявлении отредактированных или модифицированных фонограмм. Биометрическая идентификация человеческого голоса, анализ диалектов и акцентов, а также проверка некачественных записей стали более надежными благодаря использованию моделей на основе сверточных нейронных сетей (CNN), рекуррентных нейронных сетей (RNN), LSTM и Transformer. Однако развитие технологий «дипфейка», вопросы конфиденциальности и этики, а также отсутствие адекватного правового регулирования создают серьезные проблемы для судебной фоноскопии. Следовательно, существует необходимость не только в развитии технологий, но и в совершенствовании этических и правовых основ. Применение гибридных методов, создание национальных и международных звуковых баз данных, совершенствование алгоритмов глубокого обнаружения подделок и разработка автоматизированных систем судебной экспертизы являются приоритетными задачами будущего фоноскопической экспертизы.

Rəyçi: h.e.d. F.Xəlilov
Təqdim edən: A.Q.Musayeva
Daxil olma tarixi: 01.09.2025
Təkrar işlənmə tarixi: göndərilməyib.
Çapa imzalanma tarixi: 15.09.2025