



DOI: 10.22363/2313-2299-2026-17-2-392-400

EDN: LGKZBF

Research Article / Научная статья

Forensic Linguistic Research of Actor-Based AI-Generated Speech in Audio-Video Games Post-Release

Andrey V. Dzhunkovskiy  

Moscow State Linguistic University, *Moscow, Russian Federation*

 Vetinari01@gmail.com

Abstract. Since 2023, the audio-video game industry has witnessed the active implementation of generative artificial intelligence for voice-acting game characters. In some cases, AI completely replaces the role of voice actors; in others, actors and developers enter into an agreement whereby the developer reserves the right to use neural networks and the actor's voice samples to synthesize new lines as needed. With proper algorithm training and a sufficient training dataset, the results of speech synthesis based on an actor's voice can be indistinguishable from natural speech except through deep phonoscopic analysis. Another trend in the modern audio-video game industry is noteworthy, which, in combination with the previous one, creates the foundation of the problem we aim to research: “games-as-a-service” (GaaS), a phenomenon in which a game is not a static, finished product, but evolves over time, receiving updates with new content. Under these conditions, a situation arises where it is possible to synthesize speech indistinguishable at first glance from human speech by a specific individual who has given legal consent for this, yet lacks control over which phrases will be publicly voiced. The research methodology relies on works in the field of specialized forensic linguistics of text and speech. The aim of the research is to formulate an algorithm for the linguistic analysis of such utterances, which has been accomplished.

Keywords: AI-speech synthesis, forensic linguistics, speaker attribution, steganography

Conflict of interests: The author states that there were no conflicts of interest.

Article history: Received: 17.01.2026; Accepted: 27.02.2026.

For citation: Dzhunkovskiy, A.V. (2026). Forensic Linguistic Research of Actor-Based AI-Generated Speech in Audio-Video Games Post-Release. *RUDN Journal of Language Studies, Semiotics and Semantics*, 17(2), 392–400. <https://doi.org/10.22363/2313-2299-2026-17-2-392-400> EDN LGKZBF

© Dzhunkovskiy A.V., 2026



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Юрислингвистическое исследование сгенерированной ИИ речи на базе голосов актеров озвучки в аудио-видеоиграх

А.В. Джунковский  

Московский государственный лингвистический университет, Москва, Российская Федерация
 Vetinari01@gmail.com

Аннотация. Актуальность обусловлена активным внедрением разработчиками развлекательного программного обеспечения генеративного искусственного интеллекта для озвучки персонажей в индустрии аудио-видеоигр. В некоторых случаях ИИ полностью заменяет роль актеров озвучки, в других случаях актеры и разработчики заключают соглашение, согласно которому разработчик оставляет за собой право использовать нейронные сети и образцы голоса актера для синтеза новых реплик по мере необходимости. При должном обучении алгоритма и достаточной обучающей выборке результаты синтеза речи на основе голоса актера могут быть неотличимы от естественного речепроизводства без глубокого фонетического анализа. Дополнительно необходимо отметить другой тренд современной индустрии аудио-видеоигры, в совокупности с предыдущим, формирующим проблему нашего исследования: «игры-как-сервис» (англ. GaaS; Games as a Service), явление, при котором игра не статична, а развивается с течением времени, получая обновления с новым контентом. В этих условиях возникает ситуация, при которой возможен синтез неотличимой на первый взгляд человеческой речи конкретным человеком, который дал на это юридическое согласие, однако не имеет контроля над тем, какие фразы будут озвучены его голосом в публичном пространстве. Методология исследования опирается на юрислингвистических работах в сфере специализированного изучения текста и речи. Цель исследования заключается в том, чтобы сформировать алгоритм лингвистического анализа таких высказываний, что было осуществлено.

Ключевые слова: ИИ-синтез речи, юрислингвистика, атрибуция говорящего, стеганография

Конфликт интересов: автор заявил об отсутствии конфликтов интересов.

История статьи: дата поступления: 17.01.2026; дата приема в печать: 27.02.2026.

Для цитирования: *Dzhunkovskiy A.V. Forensic Linguistic Research of Actor-Based AI-Generated Speech in Audio-Video Games Post-Release // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2026. Т. 17. № 2. С. 392–400. <https://doi.org/10.22363/2313-2299-2026-17-2-392-400> EDN LGKZBF*

Introduction

The audio-video game industry (colloquially simply referred to as “The game industry”) is not a new phenomenon, but it is becoming increasingly impossible to ignore its global impact on lifestyle, economy, politics and society. Despite this, modern Russian methods of Forensic Linguistics are not yet reworked to be able to analyze the products of this industry: audio-video games.

For the purposes of our research, we shall specify that from a linguistic viewpoint we consider an audio-video game to be an interactive, multimodal, polycode, creolized intertextual text with variable content. The last attribute requires additional explanation: audio-video game content is seldom linear — certain parts may not be encountered, may be skipped by choice or due to

circumstances of a particular playthrough, hence shaping a unique perception of the product for different users. This is further complicated by the ability of developers to add new content via updates.

The totality of what we described creates a unique challenge when it comes to analyzing linguistic contents of audio-video games. Studies indicate that, as of 2025, the average age of a gamer is 41 years old, with a gender balance of 51% to 49% slightly favoring males¹, and a total of approximately 3.57 billion consumers of such works worldwide. Concurrently, certain estimates placed the value of the industry at \$189 billion USD in 2025². In the same year, the music industry's value amounted to \$29.6 billion USD³, while the film industry's reached \$33.5 billion USD⁴. Thus, the audio-video game industry is approximately three times more valuable than the film and music industries combined in 2025. This state of affairs is not novel: the assertion that the audio-video game industry is valued higher than the film and music industries combined held true each year from 2007 onward⁵. We have not discovered indicators that state that this trend will change.

Narratives, themes, and dialogue in audio-video games frequently address the topics of philosophy, politics, religion, social issues, ideology. Moreover, these works achieve instantaneous global reach through a sophisticated system of digital distribution. Under these circumstances, audio-video games can only be classified as powerful tools with the capacity for linguistic, ideological, political, psychological, and other forms of influence, as well as prospective repositories for propaganda and extremism.

We have to touch upon two important audio-video game industry trends. The first one has to do with the advent of AI-generated voice acting in audio-video games. It is possible to use fully AI-generated voices to reduce game development costs, as well as decrease the time it takes to produce a new game. Another scenario has become not uncommon since at least 2023: game developers hire real voice actors and negotiate in their contract the continued use of their voice with AI-generated voice lines for future content. This future content is required for the “games-as-a-service” (GaaS) distribution model, which entails continual addition of new content to already existing games via updates. Hence, a modern audio-video game is often not static, but a changing body of text, parts of which may be added, reworked or even removed over time. Analysis of audio-video games is

¹ Power of Play: 2025 Global Video Games Report. In: *Entertainment Software Association*. URL: <https://www.theesa.com/resources/the-global-power-of-play-report> (accessed: 17.02.2026)

² Global games market to hit \$189 billion in 2025 as growth shifts to console. In: *Newzoo*. URL: <https://newzoo.com/resources/blog/global-games-market-to-hit-189-billion-in-2025> (accessed: 17.02.2026).

³ How Big Is the Music Industry in 2025? In: *Badenstock*. URL: <https://artists.badenstock.com/how-big-is-the-music-industry-in-2025/> (accessed: 17.02.2026).

⁴ Global Box Office Estimated At \$33.5B; Anime Explodes In Up-And-Down Year: Studio Report Cards. In: *Deadline*. URL: <https://deadline.com/2026/01/global-box-office-2025-report-hollywood-studio-rankings-1236660512> (accessed: 17.02.2026).

⁵ Newzoo Global Games Market Report 2019. In: *Newzoo*. URL: <https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2019-light-version> (accessed: 17.02.2026).

only possible via version control where researchers reach conclusions regarding a particular static temporal slice of an evolving body of text under the same name.

Theoretical Framework

Forensic linguistic research of speech activity artifacts represents a branch of applied linguistics within which a wide range of issues concerning semantic and stylistic text analysis is addressed for the purposes of regulating legal relations in society and establishing linguistic facts. In a broader sense, forensic linguistic research also encompasses similar analytical activities conducted by a linguist that are not regulated by Russian Federal Law № 73-FZ “On State Forensic Expert Activity in the Russian Federation.”⁶ The objective of studying speech activity products in this context is not the pursuit of new knowledge but rather the evaluation of linguistic facts pertaining to a speech product.

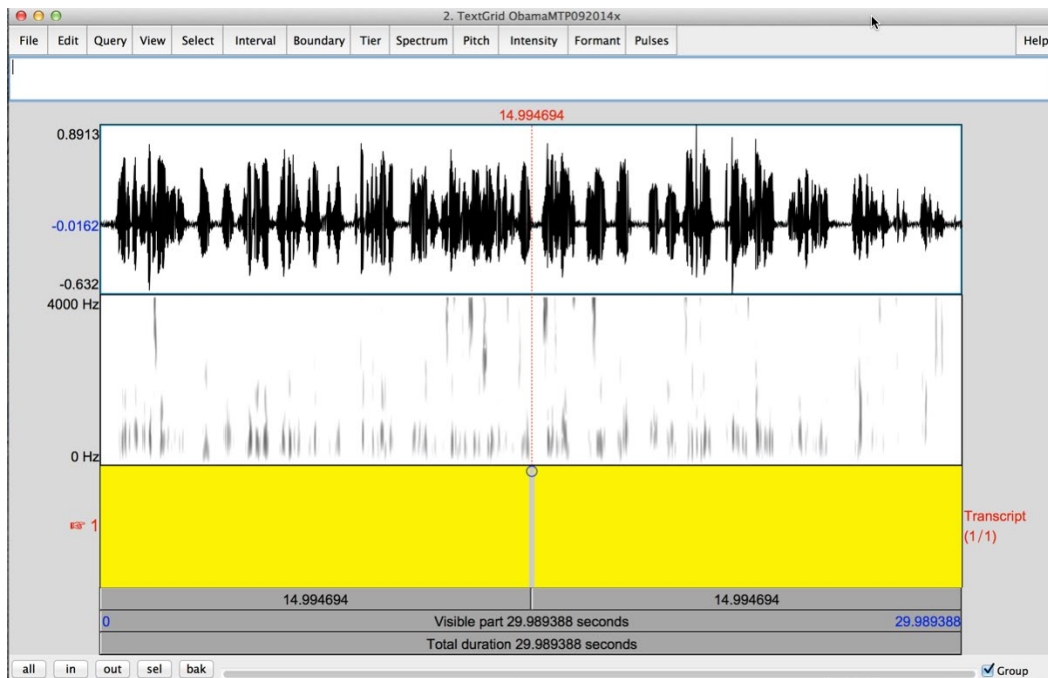
At present, such research in Russia is frequently conducted on trademarks, media texts, correspondence, social media messages, literary works, music, cinematography, as well as audio and video recordings of speech activity [1]. The latter four types of research typically involve the engagement of an acoustics expert. Meanwhile, the substantive analysis of the linguistic object is still performed by a linguist. Thus, a gap in the list of speech activity products amenable to analysis is revealed: interactive audiovisual works, commonly referred to as computer or video games (or, by some researchers, as included, audio-videogames) [2–4].

Let us propose a thought experiment. In the process of creating an audio-video game, a voice actor reaches an agreement with the developers where they not only voice one or multiple characters, but also grant the rights to use their voice samples to train a generative AI speech synthesis model. Their direct engagement with the development process ceases, but as the game evolves in a “games as a service” framework, the developers create new voice lines nigh-indistinguishable from real speech production without complicated phonoscopic analysis. These new AI-synthesized voice lines include content that is illegal to be voiced openly in certain jurisdictions. In an even more complex scenario, the trained AI speech synthesis model may leak, allowing anyone the ability to synthesize any utterance desired using the voice of a real person.

In contemporary forensic linguistics, experts use specialized software to determine if recorded speech has been tampered with, or simply to determine if recorded speech is indeed uttered by the same individual (using comparison to a controlled sample guaranteed to contain speech by the same individual). While more sophisticated software is used professionally, we may easily demonstrate what such a process looks like using open software (Figure).

⁶ Russian Federal Law № 73-FZ “On State Forensic Expert Activity in the Russian Federation.” In: *Garant*. URL: <https://base.garant.ru/12123142/> (accessed: 17.02.2026). (In Russ.).

Above we may see a spectrogram of human speech. For the purposes of this research, it is not necessary to explain in depth the machinations on display. Spectral analysis allows forensic linguists to uncover a host of data that can be compared to control samples, thus allowing one to produce data-based conclusions that are later used in court.



PRAAT audio analysis software

Source: compiled by Andrey V. Dzhunkovskiy.

The tool pictured above is PRAAT⁷, developed by Paul Boersma and David Weenink at the University of Amsterdam. It stands as a cornerstone in the field of audio signal analysis, offering a sophisticated, open-source platform for phonetic and speech research. Primarily designed for the visualization, manipulation, and quantitative assessment of acoustic signals, PRAAT enables researchers to extract precise measurements from waveforms, spectrograms, and derived parameters such as formant frequencies, pitch contours, and intensity envelopes. Its scripting language further extends its utility, automating complex analyses across large corpora of audio data.

At its core, PRAAT's audio signal analysis workflow begins with signal loading and preprocessing. Users import digitized audio files typically sampled at rates exceeding 16 kHz to capture human speech harmonics and apply filtering techniques, such as bandpass or pre-emphasis, to mitigate noise and enhance spectral clarity. The software's autocorrelation-based pitch detection algorithm,

⁷ PRAAT: doing phonetics by computer. In: *Praat*. URL: <https://www.fon.hum.uva.nl/praat/> (accessed: 17.02.2026).

which computes periodicity via time-domain correlations, yields fundamental frequency (F0) tracks with sub-millisecond resolution, crucial for prosodic studies. For instance, in analyzing intonation patterns, researchers select voiced segments and invoke the “To Pitch” command, specifying autocorrelation parameters like time step (e.g., 0.001 s) and frequency range (75–600 Hz for adult speech), generating pitch objects amenable to statistical modeling.

Spectral analysis constitutes another pillar of PRAAT’s capabilities. Fast Fourier Transform (FFT) spectrograms, configurable with window lengths from 0.005 to 0.1 s and dynamic ranges up to 70 dB, reveal formant structures, resonant peaks in the vocal tract transfer function. Burg’s linear predictive coding (LPC) method, implemented in the “To Formant (burg)” procedure, estimates formant frequencies (F1–F5) by minimizing prediction error over frames, with order typically set to 2.5 times the number of formants desired. This facilitates vowel quality mapping in acoustic spaces like the F1–F2 plane, where normalization schemes (e.g., Lobanov or Watt-Fabricius) account for speaker-intrinsic variation.

Intensity and duration metrics further enrich PRAAT’s analytical repertoire. Root-mean-square (RMS) energy computations, smoothed via Gaussian kernels, delineate stress and phrasing boundaries, while tiered TextGrid annotations synchronize phonetic transcriptions with temporal landmarks. Advanced users leverage scripts for batch processing, such as extracting F0 means, standard deviations, and formant dispersions from aligned corpora, enabling multivariate analyses via exported CSV matrices interfaced with R or Python.

Empirical validation underscores PRAAT’s reliability; inter-rater agreement studies report intraclass correlation coefficients exceeding 0.95 for formant tracking across diverse accents. Limitations persist, notably in handling non-stationary signals or severe distortions, where extensions like deep learning plugins (e.g., via PRAAT’s plugin architecture) are emerging. Nonetheless, PRAAT remains indispensable for rigorous audio signal analysis, bridging raw acoustics to perceptual linguistics in experimental paradigms.

When AI-synthesized speech is concerned, this becomes considerably more complicated and contentious, especially in cases when the AI is trained on a body of work created by the same individual, like in the thought experiment presented above.

It is evident that the voice actor is no longer the author of these new utterances, but they may face a reality where they would need to prove in a court of law or public opinion that the utterance is a product of AI, and not themselves. Parallels may be easily drawn with the AI deepfake problem.

Discussion and Results

The issue presented above is currently becoming more important and prominent. While use of generative AI in audio-video games is often criticized by users and creatives whose jobs are ostensibly at risk, the audio-video games industry is heavily invested in promoting the use of generative AI. Prominent

examples include the free 2013 audio-video game “The Finals”⁸ [5] and the 2025 audio-video game “ARC Raiders”⁹. These are not small, relatively unknown entities. Data aggregators estimate ownership of the former at 18 million and the latter at 9.38 million users worldwide¹⁰.

A question naturally arises: with such outreach and impact of audio-video games, what steps may be taken to protect voice actors from potential misuse of generative AI voice synthesis models trained with their voices? A potential solution may be found in the field of linguistic steganography.

Linguistic steganography is a branch of steganography dealing with how verbal, non-verbal, and paraverbal variables may be used to embed secret information into text and voice recordings. Steganography in and of itself deals with the issue of covertly relaying secret messages without third parties becoming cognizant of confidential information being “in transit”. Unlike cryptography, the goal of steganography is obfuscation of the very fact that additional information is being relayed between parties [6–9].

One prominent use of linguistic steganography is watermarking which helps protect copyright and determine sources of potential leaks. Watermarking entails embedding hidden markers that allow one to claim ownership of a digital entity.

There is research that details the algorithm of how additional information may be imbedded into audio. This additional information is added via altering the audio signal by altering wavelengths at peaks. This is not perceptible to the human ear and is only detectable on a spectrogram via forensic linguistic analysis [10].

We posit the following: the selfsame steganography algorithm for relaying additional info by altering the audio signal may be used to watermark AI-synthesized speech that was trained with the use of real human voice actors. Such a procedure would provide a powerful, subtle and cost-effective protection for voice actors, and for developers if the embedding of this additional info is built into the AI model itself in case it becomes available to a malicious third party.

Final Remarks

While linguistic forensic analysis of audio-video games is in its infancy, it is both possible and necessary to predict and preemptively solve potential problems that may arise from not having the sufficient methods and tools that allow experts to conduct such analysis when it comes to problematic content in other forms of media: deepfakes in news broadcasts, fabricated celebrity endorsements, courtroom audio tampering. Without robust methodologies, audio-video games risk becoming vectors for disinformation: AI-generated propaganda disseminated via in-game voiceover, indistinguishable from voice actor-created or even developer-sanctioned

⁸ The Finals Criticised by Actors and Devs for Using AI-Generated Voice Overs. In: *IGN*. URL: <https://www.ign.com/articles/the-finals-criticised-by-actors-and-devs-for-using-ai-generated-voice-overs> (accessed: 17.02.2026).

⁹ Arc Raiders and the Ethical Use of Generative AI. In: *AI and Games*. URL: <https://www.aiandgames.com/p/arc-raiders-and-the-ethical-use-of> (accessed: 17.02.2026).

¹⁰ Steam Database. URL: <https://steamdb.info/> (accessed: 17.02.2026).

content, eroding player trust and inciting real-world unrest. Preemptive development of tools—such as real-time spectrographic watermarking can avert these pitfalls by embedding verifiable markers during production, enabling post-hoc authentication even in live-service environments where content changes over time.

The problem is additionally compounded by the rise of generative AI that creates unique and novel challenges. Generative AI exacerbates these challenges exponentially. Models like those powering dialogues or procedural narratives produce hyper-realistic, contextually adaptive speech that evades conventional phonetics: prosody mimics human variability and synthesis artifacts are minimized through training. This leads us to “uncanny valley” forensic linguistics: subtle indicators like unnatural micro-pauses or spectral inconsistencies that elude human perception without additional expert analysis. In “games as a service” ecosystems, speaker attribution fractures. Proactive countermeasures may entail integrating forensic linguistics into audio-video game development pipelines from inception. Developers could mandate “audio passports” containing metadata schemas logging synthesis parameters, training data provenance, and edit histories.

References

1. Kukushkina, O.V., Safopova, U.A., & Sekerazh, T.P. (2011). *Theoretical and Methodical Basis for Forensic Psychological and Linguistic Text Research for Extremism*. Moscow: Russian Ministry of Justice publ. (In Russ.).
2. Kamaruddin, N., et al. (2018). Review of Text Watermarking: Theory, Methods, and Applications. *IEEE Access*, 6, 8011–8028. <https://doi.org/10.1109/ACCESS.2018.2796585>
3. Kuleshov, S.V., Zaytseva, A.A., Aksenov, A.Yu. (2024). Analysis of Statistical Characteristics of Artificially Generated Texts. *Journal of Instrument Engineering*, 67(11), 958–968. (In Russ.). <https://doi.org/10.17586/0021-3454-2024-67-11-958-968> EDN: HIYQAL
4. Ogorelkov, I.V., & Makhaeva, E.V. (2025). Artificial Neural Network Text Generation as a New Type of Masking the Authorship. *Virtual Communication and Social Networks*, 4(2), 163–171. (In Russ.). <https://doi.org/10.21603/2782-4799-2025-4-2-163-171> EDN: HIYQAL
5. Prokhorov, A.I., & Asadchaya, K.V. (2023). Tools for Determining the Text Generated Using a Neural Network. In: *Nauchny Vektor*, proceeding (Vol. 9, pp. 250–253). Rostov-on-Don. (In Russ.). EDN: MGTGIE
6. Potapova, R.K., & Dzhunkovskiy, A.V. (2021). Trigger-Containers in Linguistic Steganography. *Russian Linguistic Bulletin*, 26(2), 19–21. (In Russ.). <https://doi.org/10.18454/RULB.2021.26.2.10> EDN: DAKDOO
7. Potapova, R.K. (2015). *Speech: Communication, Information, Cybernetics*. Moscow: URSS. (In Russ.).
8. Potapova, R. (2002). Linguistic Knowledge and New Technologies. *Acoustic Physics*, 48(4), 486–493. (In Russ.). EDN: NXXVUV
9. Potapova, R., & Dzhunkovskiy, A. (2020). Preliminary Investigation of Potential Steganographic Container Localization. In: *Lecture Notes in Artificial Intelligence* (12335, pp. 389–398). Springer. https://doi.org/10.1007/978-3-030-60276-5_38 EDN: RLZNGS
10. Salagaj, M.O. (2011). *Prosodic Means of Information Protection (Experimental Phonetic Inquiry in the Field of Steganography)* [PhD Thesis]. Moscow: MSLU publ. (In Russ.). EDN: QFOJLJ

Список литературы

1. Кукушкина О.В., Сафонова Ю.А., Секераж Т.П. Теоретические и методические основы судебной психолого-лингвистической экспертизы текстов по делам, связанным с противодействием экстремизму. М. : Министерство Юстиции РФ, 2011.
2. Kataruddin N. et al. Review of Text Watermarking: Theory, Methods, and Applications. // IEEE Access. 2018. Vol. 6. P. 8011–8028. <https://doi.org/10.1109/ACCESS.2018.2796585>
3. Кулешов С.В., Зайцева А.А., Аксенов А.Ю. Анализ статистических характеристик искусственно сгенерированных текстов // Известия вузов. Приборостроение. 2024. Т. 67. № 11. С. 958–968. <https://doi.org/10.17586/0021-3454-2024-67-11-958-968> EDN: HIYQAL
4. Огорелков И.В., Махаева Е.В. Искусственная нейросетевая генерация текста как новый вид маскировки письменной речи автора // Виртуальная коммуникация и социальные сети. 2025. Т. 4. № 2. С. 163–171. <https://doi.org/10.21603/2782-4799-2025-4-2-163-171> EDN: HIYQAL
5. Прохоров А.И., Асадчая К.В. Инструментальные средства определения текста, сгенерированного при помощи нейросети // Научный вектор: сб. научных трудов / ред. Е.Н. Макаренко. Т. 9. Ростов-на-Дону : Ростовский государственный экономический университет «РИНХ», 2023. С. 250–253. EDN: MGTGIE
6. Потапова Р.К., Джунковский А.В. Понятие «Триггер-контейнер» в лингвистической стеганографии // Russian Linguistic Bulletin. 2021. № 26 (2). С. 19–21. <https://doi.org/10.18454/RULB.2021.26.2.10> EDN: DAKDOO
7. Потапова Р. К. Речь: коммуникация, информация, кибернетика. М. : Изд-во УРСС, 2010.
8. Потапова Р.К. Лингвистические знания и новые технологии // Акустический журнал. 2002. Т. 48. № 4. С. 552–559. EDN: NXXVUV
9. Potapova R., Dzhunkovskiy A. Preliminary Investigation of Potential Steganographic Container Localization // Lecture Notes in Computer Science. № 12335. Springer, 2020. P. 389–398. EDN: RLZNGS
10. Салагай М.О. Просодические средства защиты смысловой информации (экспериментально-фонетическое исследование в области стеганографии) : дис. ... канд. филол. наук. М. : МГЛУ, 2011. EDN: QFOJLJ

Information about the author:

Andrey V. Dzhunkovskiy, PhD in Philology, Applied and Experimental Linguistics Department Chair, Lead Researcher at Experimental Phonetics and Forensic Linguistics Laboratory, Moscow State Linguistic University (38 build 1 Ostozhenka str., Moscow, Russian Federation, 119034); *Research interests*: corpus linguistics, experimental linguistics, pragmalinguistics, forensic linguistics; *e-mail*: Vetinari01@gmail.com

ORCID: 0000-0002-3761-8010, Scopus ID: 57210805837; ResearcherID: J-6081-2017, SPIN-code: 8541-6226, AuthorID: 930082.

Сведения об авторе:

Джунковский Андрей Владимирович, кандидат филологических наук, заведующий кафедрой прикладной и экспериментальной лингвистики, ведущий научный сотрудник экспериментально-фонетической лаборатории криминалистики по речеведению, Московский государственный лингвистический университет (119034, Российская Федерация, г. Москва, ул. Остоженка, д. 38 с. 1); *сфера научных интересов*: корпусная лингвистика, экспериментальная лингвистика, лингвопрагматика, юрислингвистика; *e-mail*: Vetinari01@gmail.com

ORCID: 0000-0002-3761-8010, Scopus ID: 57210805837; ResearcherID: J-6081-2017, SPIN-код: 8541-6226, AuthorID: 930082.