

До проблеми голосу в судовій експертизі відео-, звукозапису

Ігор Римар

кандидат історичних наук, завідувач відділу дослідження звуко- та відеозапису лабораторії досліджень у сфері інформаційних технологій, Київський науково-дослідний експертно-криміналістичний центр Міністерства внутрішніх справ України, м. Київ, Україна, ORCID: <https://orcid.org/0000-0001-6153-0900>, e-mail: qwerty14531204@gmail.com

Окреслено проблему голосу, його сприйняття, розуміння та дослідження у контексті розвитку генеративного штучного інтелекту протягом першої чверті XXI ст. у межах судової експертизи відео-, звукозапису.

Ключові слова: *судова експертиза відео-, звукозапису; голос; фонема; акустатичний голос; штучний інтелект.*

Regarding the Issue of Voice in Forensic Examination of Video and Audio Recordings

Ihor Rymar

The issue of voice, its perception, understanding, and research has been outlined in the context of the development of generative AI during the first quarter of the 21st century within the scope of forensic video and audio examination.

Keywords: *audio and video forensics; voice; phoneme; acousmatic voice; artificial intelligence (AI).*

Голос інтелекту є тихим, однак він не вмовкає, аж поки його не вислухають.
3. Фрейд «Майбутнє однієї ілюзії» (1927)

Ми зіштовхуємося з голосами щодня та щохвилини. Голоси оточують нас, проходять крізь нас, звучать у нас самих. Голоси перехожих, рідних та близьких, голоси музики, медіа, атопічний голос, голос Бога, сумління, розуму, істини, голос без тіла, голос страху. Усі вони кричать, погрожують, благають, наказують, спокушають, молять, проголошують, закликають, шепочуть, і все це, ба навіть більше (!), одночасно, на фоні все наростаючого шуму цивілізації.

Голоси не лише оточують нас, ми самі наділені голосом та активно їм користуємось. Характеризуючи го-

лос як невід'ємний аспект соціального словенський філософ М. Долар, соціалізує людину «...за допомогою голосу та через голос; здається, що голос розміщений на самій осі наших соціальних зв'язків, і що голоси — це сама текстура соціального...» [1, с. 25].

Для експерта голос є безпосереднім об'єктом дослідження та роздумів. Стикаючись з голосом у практичному вимірі, ми постійно піддаємо його аналізу, вивченню та пізнанню. Для нас голос перетворюється на інструмент переступу, засіб учинення

протиправних дій. Виходячи за межі соціального, ми препаруємо голос, розкладаючи його на компоненти, виділяючи особливості, порівнюючи, вивчаючи кожен з них, розділяючи та об'єднуючи. Щодня ми працюємо з цією одночасно незбагненою та зрозумілою частиною індивіда — нематеріальною, і через це такою привабливою. Як нам вдається візуалізувати голос, який пробивається крізь шуми та перешкоди, маніфестуючи свою присутність? Голос підозрюваного, який губиться на фоні інших, змішаний із шумом автостради на якісь заправці, із музичними ритмами, які линути із дешевої акустичної системи? Наділити голос тілом? Ми можемо розкласти його на компоненти та дослідити не лише на математичному, фізичному, рівні, але й на лінгвістичному, розкласти на фонеми.

Фонетика та фонологія надають інструменти, які дають змогу помістити «живий голос» у логічну таблицю, засновану на бінарній опозиції: присутності чи відсутності диференційних ознак. Зробити видимим. Спіймати, розкласти на фонеми та помістити в матрицю, додати диференційних ознак, таких як акцент, інтонація, просодія, звівши до набору символів, змусивши голос замовкнути. Перетворити нематеріальний голос на матеріальні символи, таблицю зрозумілих та впізнаваних елементів лінгвістичного вивчення. Але такий голос лінгвістики упускає індивідуальність. Будучи чимось на кшталт відбитків пальців, голос дає змогу одній людині безпомилково ідентифікувати іншу за тембром, темпом, висотою

голосу, мелодикою, специфічним вимовлянням певних звуків, артикуляцією тощо. Усе це неможливо описати лінгвістично, незначні відхилення, що не порушують загально визначеної норми. Для кожної мови такий «відбиток голосу» буде унікальним, не лише в бінарних опозиціях, але й в просодії.

На противагу живому голосу механічний голос безособовий, без побічних ефектів, дотримується абсолютної норми. Позбавлений елементів людського, голос перетворюється на частину машини, із жахливим реалізмом втілюючи давню мрію людини про механічну ляльку, здатну говорити. Ми досягли межі, де проблема знаку оприявнюється найбільше — нездатність включити до матриці кожне відхилення, кожную варіацію, вийти за самі межі матриці перетворює лінгвістику на простий набір нулів та одиниць. Це не означає, що варто відкинути лінгвістику як один із методів чи навіть інструментів дослідження та ідентифікації особи за голосом та мовленням. Це вказує на поріг, за яким таке дослідження буде ставити більше запитань, ніж давати відповідей.

Із тілом складніше. Переважно експерт або індивід стикається саме з голосом, позбавленим тіла, акустичним голосом. До того ж на практиці покарання за проступок або ж учинений злочин повинно понести саме тіло, чи скоріше душа, яка, за вдалим висловом французького філософа М. Фуко, є *«наслідком та знаряддям політичної анатомії»* [2, с. 44], свого роду в'язницею тіла. Такий безтілес-

ний об'єкт дослідження тільки посилює подив, враження експерта, коли зазнає деакузматизації, отримує тіло. Тіло, що йому зовсім не підходить або ж уражає своєю будовою, недосконалістю, вразливістю або ж буденністю. Голос завжди репрезентує тіло у його найбільш інтимній формі; голос, що несе подих, указує на душу, що є незвіданою до тіла [1, с. 103] та на яку спрямовані усі механізми покарання та виправлення в межах сучасних карних систем.

Сьогодні «акузматична якість голосу стала універсальною, а відтак — тривіальною. Джерело голосу замінилося пристроєм, який його маскує, починаючи функціонувати як його замітник» [1, с. 92]. Замітник викликає необхідність дослідження. Експерт ніколи не досліджує безпосередньо голос як нематеріальну субстанцію. Він завжди досліджує замітник, копію, що лунає з механічного пристрою. Сьогодні ми перебуваємо на межі переходу від голосу, що має тіло, до не просто голосу, що лунає з механічного пристрою, а голосу машинного тіла, здатного говорити, як людина. Наступний крок — наділення машини власним, повноцінним людським голосом, з усією його індивідуальністю та недосконалістю, що аж ніяк не додає впевненості, навпаки, породжує проблему ідентифікації такого голосу, розділення та класифікації голосів, наділених тілом від природи та голосів, для яких його створено (фізично, програмно чи віртуально).

На сьогоднішній день у мережі «Інтернет» доступні десятки програм для клонування та генерації голосу

на основі генеративного штучного інтелекту (далі — ШІ): *ElevenLabs* [3], *Resemble AI* [4], *Synthesys* [5], *Murf* [1], *Altered Studio* [7] та ін.; програм, з якими можна просто поспілкуватися (*Google Gemini* [8], *ChatGPT* [9]) або які уже працюють на державу (Вікторія ШІ [10]). Після роботи з ними розумієш: розробники наблизилися до межі найбільшої подібності, переступивши яку, відрізнити живий голос від генерованого чи клонованого за допомогою ШІ буде надзвичайно важко, якщо взагалі можливо.

З-поміж перспективних проєктів, що розробляють, варто виокремити проєкт ШІ, над яким працює компанія *Nvidia* — *Fugatto* [11—13]. Доступні в мережі аудіогенерованих голосів за допомогою попередніх версій *Fugatto* вказують на високу якість та дуже хорошу роботу алгоритмів ШІ. Голос оживає, з'являється просодія, диференціюється інтонація, з'являється більше відтінків та акцентів. Голос машини починає звучати по-людськи.

Чи зможе експерт із наявними на сьогодні знаннями, матеріально-технічною базою відповісти на такий стрімкий розвиток генеративного ШІ? Чи достатньо у нашому розпорядженні інструментів, щоб відрізнити голос ШІ від людського? На моє переконання — ні, що породжує необхідність створення повноцінної полілінгвістичної текстонезалежної системи ідентифікації дикторів за голосом та мовленням, у якій експерт посяде місце оператора та невід'ємної складової частини. Системи, яка зможе ефективно розрізняти та класифікувати

голоси, обробляти значні масиви даних, самонавчатися, розкладати голос на позначники та досліджувати на лінгвістичному рівні тощо. Разом із тим, удосконалення наявних та

розроблення нових методик, пошук інструментів дослідження та аналізу, підвищення професійного рівня експертів забезпечить їхню компетентність перед обличчям нових викликів.

Перелік джерел посилання

1. Долар М. Голос і більше нічого / пер. з англ. Павла Шведа. Київ, 2022. 304 с.
2. Фуко М. Наглядати й карати. Народження в'язниці / пер. з фр. Петра Тарашука. Київ, 2023. 452 с.
3. ElevenLabs. URL: <https://elevenlabs.io/> (дата звернення: 08.04.2025).
4. Resemble AI. URL: <https://www.resemble.ai/> (дата звернення: 08.04.2025).
5. Synthesys. URL: <https://synthesys.io/> (дата звернення: 08.04.2025).
6. Murf. URL: <https://murf.ai/> (дата звернення: 08.04.2025).
7. Altered Studio. URL: <https://www.altered.ai/> (дата звернення: 08.04.2025).
8. Google Gemini. URL: <https://gemini.google.com/app> (дата звернення: 08.04.2025).
9. ChatGPT. URL: <https://chatgpt.com/g/g-mzFm1dKjW-chat-gpt> (дата звернення: 08.04.2025).
10. Вітальне слово представниці МЗС з консульських питань Вікторії Ші. URL: <https://mfa.gov.ua/news/vitalne-slovo-predstavnici-mzs-z-konsulskih-pitan-viktoriyi-shi> (дата звернення: 08.04.2025).
11. Fugatto. URL: <https://fugatto.pro/> (дата звернення: 08.04.2025).
12. Fugatto 1 — Foundational Generative Audio Transformer Opus 1. URL: https://research.nvidia.com/publication/2025-04_fugatto-1-foundational-generative-audio-transformer-opus-1 (дата звернення: 08.04.2025).
13. Nvidia представила ШІ-модель Fugatto, яка «розуміє і генерує звук так само, як це роблять люди». URL: <https://itc.ua/ua/novini/nvidia-predstavyla-shi-model-fugatto-yaka-rozumiye-i-generuye-zvuk-tak-samo-yak-tse-roblyat-lyudy/> (дата звернення: 08.04.2025).