

В.І. СОЛОВЙОВ

Компанія «Сілентіум Систем», Ванкувер, Канада, e-mail: *edemsvi@gmail.com*.

О.В. РИБАЛЬСЬКИЙ

Національна академія внутрішніх справ, Київ, Україна, e-mail: *rov_1946@ukr.net*.

В.В. ЖУРАВЕЛЬ

Київський науково-дослідний експертно-криміналістичний центр МВС України, Київ, Україна, e-mail: *fonoscopia@ukr.net*.

О.М. ШАБЛЯ

Одеський науково-дослідний інститут судових експертиз Міністерства юстиції України, Одеса, Україна, e-mail: *alikhshablya@gmail.com*.

Є.В. ТИМКО

Київський науково-дослідний інститут судових експертиз Міністерства юстиції України, Київ, Україна, e-mail: *e.tymko@kndise.gov.ua*.

НАДЛИШКОВІСТЬ ІНФОРМАЦІЇ У ПОБУДОВІ СИСТЕМ ЕКСПЕРТИЗИ ЗВУКОВИХ СИГНАЛІВ НА НЕЙРОННИХ МЕРЕЖАХ ГЛИБОКОГО НАВЧАННЯ

Анотація. Описано методи попереднього оброблення сигналів, які використовують для створення нового інструментарію експертизи матеріалів і засобів цифрового звукозапису. Показано, що застосування надлишковості інформації у створенні бази навчання нейронних мереж глибокого навчання, які використовуються для цієї експертизи, забезпечує підвищення ефективності ідентифікації диктора за параметрами характеристик голосу приблизно на 15 %. Підтверджено, що запропоновані методи оброблення забезпечили можливість ідентифікації диктора за фонограмами тривалістю від 1 с.

Ключові слова: вейвлет Морле, часове вікно, часо-частотне перетворення, диктор, ідентифікація, надлишковість, нейронна мережа, спектр, фонограма.

ВСТУП

Сучасні системи, розроблені у галузі мовних технологій, здебільшого ґрунтуються на нейронних мережах глибокого навчання [1–4]. Зокрема, в основу практично всіх ефективних систем Speech-Text нині покладено машинне навчання. Не менш широко технології нейронних мереж, що навчаються, застосовують і в інших галузях оброблення звукової інформації, наприклад, для розпізнавання дикторів [5–7]. Проте у спільноті фахівців з побудови систем контролю доступу за різними біометричними показниками заведено вважати, що сучасні системи ідентифікації людини за характеристиками голосу диктора не мають ефективності, достатньої для практичних цілей [8–10]. Це пояснюється тим, що хоча під час випробувань найкращих таких систем на тривалих фонограмах точка перетину графіків помилок першого та другого роду знаходиться у межах декількох відсотків, їхнє практичне застосування на коротких фонограмах (характерних для завдань, розв'язуваних такими системами) дає значно гірші результати. Це, зокрема, стосується експертизи матеріалів та засобів цифрового звукозапису, бо численні фахівці у цій галузі дуже скептично ставляться до можливості автоматичної ідентифікації диктора [8–10].

Уважний розгляд принципів побудови таких систем на сучасному етапі їхнього розвитку (незалежно від використовуваних технологій) дає змогу досить впевнено виділити всього дві основні індивідуальні ознаки голосу, покладені в їхню основу. Ними є відомі з часів Фанти та Фланагана частота основного тону та формантні ознаки спектра, що, проте, у жодному разі не зменшує цінності багатьох інших досліджень у цій галузі [11, 12].

Наведені міркування також стосуються використання технологій нейронних мереж, де на вхід під час їхнього навчання подають різні варіанти функцій