**ПОРІВНЯЛЬНИЙ АНАЛІЗ ТЕХНОЛОГІЙ ЗВУКОВОГО ПОДІЛУ**

**НА БАЗІ НЕЙРОННИХ МЕРЕЖ**

*Гапей Максим Юрійович,*

*Державний університет інформаційно-комунікаційних технологій,*

*ПДМ-51,*

*121 Інженерія програмного забезпечення,*

*Фесенко Максим Анатолійович,*

*кандидат технічних наук, доцент,*

*доцент кафедри Штучного інтелекту*

Технології розпізнавання інформації усіх видів стрімко розвиваються, що відкриває нові можливості для інженерів, які можуть використовувати їх у власних системах. Ще недавно машини вивчили розпізнавати рукописні символи, класифікувати дані, прогнозувати, але зараз вони можуть також перекладати текст, відновлювати кольори на чорно-білих зображеннях. Однак, щодо розділення звукової інформації, ситуація трохи складніша. Машини легко розрізняють голоси людей, але якщо додати трохи шуму, наприклад, звуки дощу, крик немовляти, гру на кларнеті, справа стає складнішою. Чи буде завдання тепер вважатися тривіальним? Нажаль, ні, оскільки новоутворена спектрограма звуку буде містити декілька аудіоджерел. Як обійти цю проблему? Існують декілька технологій, які дозволяють це зробити.

*Постановка задачі.*

Проаналізувати три технології AudioSep, VoiceFilter, MusicSourceSeparation звукового розподілу, кожна із них має свої особливості, а також сильні та слабкі сторони.

*Мета дослідження.*

Визначити ключові можливості, відмінності та особливості кожної із технології та виділити найуніверсальнішу.

*Результати дослідження.*

AudioSep — це технологія розділення змішаних аудіоджерел на їх складові джерела, за допомогою їх текстового опису [1]. Вона працює за допомогою сліпого розділення джерел, що не потрібує попередні знання про ці джерела. AudioSep можна використовувати для звукового поділу на різні аудіоджерела, наприклад, ізоляції вокалу від фонової музики, виділення окремих інструментів із музичного твору або відокремлення мовлення в шумному середовищі. Алгоритми AudioSep часто покладаються на аналіз спектральних і часових характеристик аудіосигналу для ідентифікації та розділення різних джерел. Залежно від складності та вимог завдання поділу можуть застосовуватися такі методи, як аналіз незалежних компонентів, факторизація невід’ємної матриці або методи на основі глибокого навчання.

VoiceFilter — це спеціалізована технологія, розроблена спеціально для ізоляції та покращення людського мовлення в аудіозаписах, що містять кілька джерел звуку або фоновий шум [2]. На відміну від AudioSep, який націлений на загальне розділення джерел, VoiceFilter зосереджується саме на ізоляції людської мови, ігноруючи інші звуки (музика чи шум навколишнього середовища). VoiceFilter зазвичай використовує моделі на основі глибокого навчання, такі як згорткові нейронні мережі або рекурентні нейронні мережі, навчені на великих наборах даних змішаних аудіозаписів і відповідних чистих мовних сигналів. Модель вчиться ідентифікувати та виділяти мовні компоненти з шумного звуку, використовуючи собливості, специфічні для людської мови, такі як висота, спектральні характеристики та лінгвістичний контекст. Застосування VoiceFilter допоможе покращити мовлення в галасливих записах, корелювати мовлення спікерів під час аудіоконференцій і створювати пристрої з голосовим керуванням (наприклад смарт будинки) у шумному середовищі.

MusicSourceSeparation — ця технологія зосереджена на відокремленні окремих музичних компонентів, таких як вокал, ударні, бас та інші інструменти, від змішаних аудіозаписів [3]. Подібно до AudioSep, MusicSourceSeparation спеціально націлена на музичні аудіоджерела, а не на загальні. Методики, які використовуються в звуковому поділі, включають спектральне моделювання, обробку в частотній області та підходи на основі машинного навчання, адаптовані до унікальних характеристик музичних сигналів. MusicSourceSeparation застосувується у створенні музики, реміксуванні, автоматичній транскрипції музики та створенні інтерактивних музичних програм, де окремими аудіоджерелами можна маніпулювати незалежно.

*Висновки та перспективи.*

Хоча всі три технології мають спільну мету поділу звуку на аудіоджерела, вони відрізняються своєю специфікою взаємодії та сферами застосування. VoiceFilter спеціалізується на відокремленні людського голосу, від інших аудіоджерел, а MusicSourceSeparation розроблено спеціально для виділення музичних звуків. AudioSep — це, скоріше, найуніверсальніша технологія звукового поділу, оскільки дозволяє описати аудіоджерело в не менш універсальному форматі. Гармонійне поєднання усіх цих технології дозволяє ефективно вирішувати завдання звукового поділу з високим ступенем успішності.

Список використаних джерел:

1. Music Source Separation with Band-Split RoPE Transformer. *arXiv.org*. URL: <https://arxiv.org/abs/2309.02612>

2. Separate Anything You Describe. URL: <https://audio-agi.github.io/Separate-Anything-You-Describe/>.

3. VoiceFilter: Targeted Voice Separation by Speaker-Conditioned Spectrogram Masking. *arXiv.org*. URL: <https://arxiv.org/abs/1810.04826>