

ВІДГУК

офіційного опонента на дисертаційну роботу

Борисова Гліба Олександровича

за темою

«Адаптивні системи оброблення акустичної інформації для створення персоналізованого медіаконтенту»,

представлену на здобуття ступеня доктора філософії
в галузі знань 17 «Електроніка та телекомунікації»

за спеціальністю 171 «Електроніка»

Актуальність теми дисертації.

Сьогодення спостерігається стрімкий розвиток технологій з сфери масової комунікації. Це сучасна медіасистема. Вона безперервно розвивається, вводяться нові терміни та поняття. Відбуваються зміни у процесах медіа виробництва. Потреба має суть в тому, що сьогодні засоби мультимедіа стали основним джерелом отримання нової інформації для користувачів. Медіаконтент застосовується в різних сферах діяльності, від галузі освіти до маркетингу. Дане поняття використовується як загальновживане, не актуалізується і не конкретизується з урахуванням його своєрідних властивостей, що існують у межах сучасного комунікаційного середовища. Саме тому так важливо з'ясувати значення медіаконтенту та його види. При створенні, наприклад, систем клонування голосу, які знаходять своє застосування у медицині, освіті, інформаційних системах, навіть, у розвагах та засобах комунікації. Для клонування оригіналу необхідно мати якісний персоналізований медіаконтент. Одним з доступних та популярних на сьогодні шляхів для забезпечення цієї якості – використання різноманітних технологій оброблення акустичної інформації, до числа котрих можна віднести застосування нейромережових алгоритмів. Так можна отримати точну ідентифікацію голосу, враховуючи індивідуальні характеристики мовця, реалізувати синтез природного мовлення. При реалізації цих процесів треба забезпечити ефективно зменшення шуму і реверберації сигналів. З іншого боку, задача підвищення розбірливості мовлення є важливою та актуальною задачею для поліпшення комунікації між користувачами при перебуванні останніх у різних акустичних середовищах, серед яких це лекційні аудиторії, офіси або будь-які відкриті локації. Питання забезпечення якості аудіосигналів дуже часто виникають і у випадках створення систем з розпізнавання образів та української мови, при проведенні процедур дереверберації записаних сигналів в приміщенні для запису. При записі аудіокниг з текстового джерела (системи озвучування розмовного тексту), при налаштуванні та створенні голосових асистентів. Ці системи аналогічні світовим розробкам від компаній Google чи Amazon. Нейронні мережі дозволяють при аудіоаналізі виділити важливі ознаки, характерні для мовного сигналу: акцент, інтонація, артикуляція тощо. Застосування нейронних мереж навіть допомагає виявляти аномалії в аудіо даних. Наприклад при аналізі звуків, пов'язаних з хворобами у людині,

насамперед, легень (стетоскопічні звуки), серця (аритмія). Адаптивні технології оброблення акустичної інформації з використанням нейронних алгоритмів, знаходять широке використання і у випадках створення інтелектуальних систем безпеки, в інтерфейсах "розумного будинку", слухових апаратах, а також у медіа-індустрії для забезпечення автоматичного дубляжу в режимі реального часу, створення віртуальних асистентів та різноманітних прикладних інтерактивних систем. Цим визначається *актуальність* обраної автором дисертації теми.

Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни.

Дисертаційна робота присвячена дослідженню адаптивних систем оброблення акустичної інформації для створення персоналізованого медіаконтенту. Нові наукові результати отримані автором на базі попередніх досліджень відомих вчених: Продеус А. М, Наконечний Р. А., Бабак В. П., Міхаель Ферлендер, Девід Хейвелок, Джон Картер, Майкл Джонсон та інші.

Об'єктом дослідження є різноманітний аудіо контент з записом живого або синтетично створеного голосового повідомлення на українській та англійській мовах.

Предметом дослідження є методи та алгоритми використання нейронних мереж для обробки мовних сигналів, включаючи розпізнавання, зменшення шуму, дереверберацію і клонування голосу, які спрямовані на покращення якості та забезпечення персоналізації медіаконтенту.

Мета і завдання дослідження. *Метою дисертаційної роботи* є дослідження нейромережевих алгоритмів для створення персоналізованого медіаконтенту на основі підходів оброблення акустичної інформації. Дослідження спрямоване на отримання покращення якості мовних сигналів та врахування індивідуальності медіа контенту, шляхом застосування методів навчання нейронних мереж, які в свою чергу, здатні адаптуватись в процесі навчання до голосових особливостей мовця, і які також дозволяють ефективно здійснювати дереверберацію, синтез і розпізнавання мовних сигналів.

Для досягнення поставленої мети були сформульовані *наступні завдання*:

1. Розробити систему ідентифікації за голосом, яка характеризується задовільною точністю спрацювання при п'яти різних варіантах вимовлення контрольної фрази-перевірки.
2. Розробити модель системи розпізнавання мовних сигналів з досягненням мінімального значення функції втрат.
3. Розробити алгоритм застосування нейронної мережі для розв'язання задач дереверберації та зменшення шуму.
4. Розробити схему клонування голосу з використанням нейронних мереж для збереження інтонацій, тембру та інших специфічних фонетичних характеристик мовлення, що забезпечить у свою чергу високий рівень персоналізації контенту.

Мета, задачі, об'єкт і предмет дослідження автором визначені вірно.

Наукова новизна результатів дисертаційного дослідження полягає в наступному:

- вперше розроблено систему ідентифікації за голосом, яка є стійкою до штучно підробленого голосу і показує високу точність схожості з еталонним записом відразу за 4 критеріями;

- вперше побудовано акустичну модель розпізнавання мовних сигналів з підтримкою нейронної мережі, яка дозволяє в якості вхідної інформації використовувати українські словосполучення. Для її реалізації розроблено змінену рекурентну нейронну мережу, яка вирізняється тим, що за рахунок вбудованої пам'яті в структурі етап навчання та тестування нейронної мережі моделі можна проводити одночасно;

- удосконалено програмний алгоритм дереверберації записаних аудіо сигналів з адитивним додаванням шуму, де використано згорткову нейронну мережу за архітектурою U-Net і яка адаптована до запису не тільки тестових сигналів типу 'сплеск' або "постріл", але й словосполучень українською мовою;

- набуло подальшого розвитку створення систем клонування голосу за рахунок введення послідовно трьох попередньо навчених нейронних мереж. Такий підхід дозволив зберегти акцент, інтонаційні та інші фонетичні особливості у синтезованих фразах як англійської, так і української мов.

Методи дослідження. Для досягнення поставленої мети використано методи аналітичного та комп'ютерного моделювання з використанням інструментарію Matlab, а також методи експериментального дослідження та метод експертних оцінок. Експериментальні дослідження включали кілька етапів, зокрема тестування розроблених систем на різних наборах даних, оцінювання точності та адаптації сигналів у складних акустичних умовах. Вирішення проблем, що поставлені у дисертації, ґрунтуються на відомих досягненнях фундаментальних і прикладних наукових дисциплін.

У дисертації автор грамотно використовує методи математичної фізики, автор коректно формулює постановку задач.

Практичне значення отриманих результатів. Практичне значення отриманих результатів полягає у розробці та впровадженні ефективних підходів обробки акустичної інформації, які базуються на принципах функціонування сучасних нейронних мереж. Отримані результати можуть бути використані для створення систем автоматичного розпізнавання мовлення, синтезу персоналізованого голосу, адаптації аудіосигналів до різних акустичних умов та зменшення впливу шумів і реверберації. Запропоновані алгоритми та підходи є універсальними, їх можливо використання її результатів для підвищення точності, стійкості та адаптивності сучасних технологій персоналізованого медіаконтенту.

Зв'язок роботи з науковими програмами, планами, темами.

Викладені у дисертації нові теоретичні та практичні результати досліджень можуть використовуватися при розробці систем оброблення акустичної інформації та створення аудіо контенту з застосування технологій Інтернету речей, а також у освітньому процесі при підготовці нових навчальних дисциплін за спеціальністю 171 Електроніка.

Отже, в дисертаційній роботі поставлене наукове завдання виконано повністю, здобувач повною мірою оволодів методологією наукової діяльності.

Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності.

За своїм змістом дисертаційна робота здобувача Борисова Г.О. повністю відповідає Стандарту вищої освіти зі спеціальності 171 «Електроніка» та напрямкам досліджень відповідно до освітньої програми «Електроніка».

Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у науковий напрям «Електроніка».

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадіння, можна зробити висновок, що дисертаційна робота Борисова Гліба Олександровича є результатом самостійних досліджень здобувача і не містить елементів фальсифікації, компіляції, фабрикації, плагіату та запозичень. Використані ідеї, результати і тексти інших авторів мають належні посилання на відповідне джерело.

Мова та стиль викладення результатів

Дисертаційна робота написана українською мовою. Робота викладена на високому науково-технічному рівні з використанням стандартної наукової термінології, яка прийнята в галузі електронних засобів та прикладної акустики.

Дисертація складається з вступу, трьох розділів, висновків, списку літератури та трьох додатків. Робота містить 29 рисунків та 21 таблицю. Загальний обсяг дисертації складає 137 сторінок.

У вступі визначено актуальність дослідження, її мету та завдання. Також наведено об'єкт та предмет дослідження і сформульовано завдання та методи дослідження. Окремо, зазначено наукову новизну та практичне значення отриманих результатів.

Перший розділ (літературний огляд). Окреслено у дослідженнях різні підходи до оброблення аудіо сигналів, і окремо визначено існуючі способи аналізу сигналів мовлення на основі проведення, як приклад, частотного та кепстрального аналізу. Також в цьому розділі наведено співвідношення, які дозволяють кількісно оцінити розбірливість та провести дереверберацію мовних сигналів. Показано яким чином проводиться синтез мовлення і яку роль при цьому відіграють MEL-спектрограми. В окремій частині розділу описано два види нейронних мереж, які можна застосувати для реалізації адаптивних технологій обробки акустичної інформації. Аналіз літератури, що привів автор, свідчить про те, що автор володіє сучасною термологією в області акустичного аналізу мовлення, глибоко розбирається в сучасних системах кепстрального та мел-аналізів.

Другий розділ. Автор розглянув теоретичні особливості трьох систем оброблення акустичної інформації, які поділено структурно. Це він використав для реалізації подальших досліджень. Використані система ідентифікації за голосом, система підвищення розбірливості мовлення, система клонування голосу. Для кожної з цих систем наведено відповідну структурну схему,

показано роль використання при цьому нейронних мереж, як одної так і декількох, побудовано технологічний ланцюжок виконання етапів оброблення акустичної інформації. Наведена реалізація цього проекту на прикладі створення системи клонування голосу.

Третій розділ дослідження охоплює практичну частину роботи автора. Тут наведено вирішення таких ключових *практичних* задач: забезпечення надійної перевірки та контролю доступу до приміщення особи за голосом; розроблення мовного автомату з розпізнавання слів (іменник, дієслово, прикметник). Автор використав цю методику для розпізнавання речі як українською, так і англійською мовами. Тут враховується різна швидкість та інтонація проголошення слів. Автором запропонована методика відновлення записаного мовного сигналу в приміщенні, яке неналежно акустично оброблено. Актуальне вирішення проблеми формування синтезованого голосу для створення аудіо додатків. Наприклад, для створення голосових асистентів чи для створення системи переведення текстів у голосові файли. Автор пропонує всі згадані задачі вирішувати на основі використання інструментарію нейронних мереж різних видів. Відповідно до цього, автором в цьому розділі визначено особливості навчання нейронних мереж і запропоновано використання цих особливостей на практиці. Автор провів перевірку якості функціонування запропонованих технічних рішень на реалізації у таких системах: системи ідентифікації за голосом, системи розпізнавання мови, системи підвищення розбірливості мови з функцією дереверберації та системи клонування голосу. Для перевірки якості функціонування систем клонування голосу та підвищення розбірливості мови було залучено експертну групу з 10 осіб та використано метод експертних оцінок. Для системи ідентифікації за голосом додатково проаналізовано метрики оцінки прогнозованої сили моделей ROC та AUC.

Дисертаційна робота оформлена відповідно до вимог наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

Оприлюднення результатів дисертаційної роботи

Наукові результати дисертації висвітлені у 4 наукових публікаціях здобувача, серед яких: 4 статті у наукових виданнях, включених на дату опублікування до переліку наукових фахових видань України. Також результати дисертації були апробовані на 1 міжнародній науково-практичній конференції.

Таким чином, наукові результати описані в дисертаційній роботі повністю висвітлені у наукових публікаціях здобувача.

Недоліки, зауваження та побажання до дисертаційної роботи.

До недоліків та зауважень до роботи слід віднести:

1. Пункт 3.5 роботи відноситься до опису метрик, які стосуються системи ідентифікації за голосом, і тому логічніше було б його розмістити після п.3.1 роботи.
2. На стор. 81 автором визначена група експертів з 10 людей для оцінки якості проведеної дереверберації сигналів в приміщенні, і в цьому

контексті бажано було б окремо зазначити, на основі яких ознак експерти були включені в цю групу.

3. Бажано було б у дисертації дати абзац такого характеру: Дисертаційна робота виконана на кафедрі Національного технічного університету України «КПІ імені Ігоря Сікорського».
4. До побажань можна віднести. «Отримані в роботі результати можуть використовуватись у навчальному процесі вищих навчальних закладів України, у тому числі НТУУ «КПІ» імені Ігоря Сікорського при підготовці інженерів-акустиків та в Дніпровському національному університеті імені Олеся Гончара на фізико-технічному факультеті на кафедрі кібербезпеки та комп'ютерно-інтегрованих технологій».

Вважаю, що висловлені зауваження не є визначальними і не зменшують загальну наукову новизну та практичну значимість результатів та не впливають на позитивну оцінку дисертаційної роботи.

Висновок про дисертаційну роботу

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Борисова Гліба Олександровича на тему «Адаптивні системи оброблення акустичної інформації для створення персоналізованого медіаконтенту» виконана на високому науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для галузі знань «Електроніка та телекомунікації». Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що передбачені в п.6 – 9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувач Борисов Гліб Олександрович заслуговує на присудження ступеня доктора філософії в галузі знань 17 «Електроніка та телекомунікації» за спеціальністю 171 «Електроніка».

Офіційний опонент:

Професор кафедри механотроніки,
Дніпровський національний
університет імені Олеся Гончара
д.т.н., професор,

Відмінник освіти України



Галина СОКОЛ

М.П.

«___» _____ 20__ року