

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

завідувача кафедри теоретичної кібернетики, Київський національний університет імені Тараса Шевченка, професора **КРАКА Юрія Васильовича** на дисертаційну роботу **ЛУЦА Владислава Євгеновича** на тему «Методи і засоби трансформації звукової інформації в жестову мову з використанням нейронної мережі», яка представлена на здобуття ступеня доктора філософії у галузі 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки»

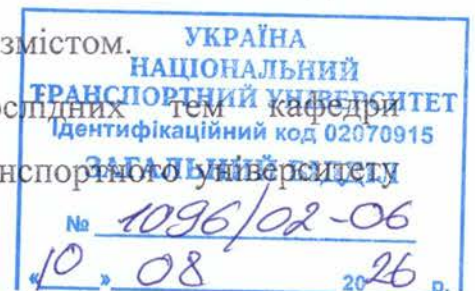
Актуальність теми дисертаційної роботи та її зв'язок з науковими програмами, планами, темами

Дисертаційне дослідження В.Є. Луца присвячене створенню методів і програмних засобів, що забезпечують автоматичне перетворення усного мовлення на текст, придатний для подальшого перекладу на Українську жестову мову, з опорою на нейромережеві технології обробки сигналів та природної мови. Забезпечення інформаційної доступності для осіб з порушеннями слуху є важливим соціальним питанням, яке й досі не має повного розв'язання ні в межах України, ні у світовій практиці.

Наявних перекладачів УЖМ недостатньо для покриття потреб глухих і слабочуючих громадян, а тому розробка автоматизованих технологічних рішень, спроможних частково компенсувати цей дефіцит, є нагальною потребою.

У дисертації автором обґрунтовано доцільність комплексного підходу до розв'язання поставленої задачі, що охоплює адаптивну фільтрацію звукового сигналу, автоматичне розпізнавання мовлення, багаторівневу автокорекцію помилок розпізнавання та концептуально-семантичну лематизацію тексту з урахуванням структури жестового словника. При цьому автору вдалося виявити принципову наукову проблему: традиційні підходи до граматичної лематизації не відповідають вимогам жестового словника УЖМ, оскільки не враховують його концептуально-семантичної організації — тоді як одна жестова лема здатна об'єднувати цілу групу словоформ, споріднених саме за змістом.

Дисертацію виконано у межах науково-дослідних тем кафедри інформаційних систем і технологій Національного транспортного університету



відповідно до планів наукової діяльності, зокрема в рамках теми «Математичні методи, моделі і технології дослідження складних систем», державний реєстраційний номер 0124U003679.

Наукова новизна і практичне значення отриманих результатів

Наукова новизна дисертаційної роботи полягає у розробці методів і засобів трансформації звукової інформації на жестову мову, що забезпечують підвищення якості транскрипції та коректність концептуально-семантичної лематизації для УЖМ.

У рамках дослідження **вперше**:

- запропоновано та реалізовано каскадний підхід до постобробки результатів розпізнавання мовлення у задачі перекладу на УЖМ, що поєднує словниковий пошук та нейромережевий fallback на основі character-level Seq2Seq;
- запропоновано технологію автоматизованого формування тренувального датасету реальних помилок розпізнавання через TTS-Whisper pipeline, що дозволяє отримувати предметно-специфічні дані для донавчання моделі без ручної обробки.

Удосконалено:

- технологію фільтрації аудіосигналів для ASR-систем шляхом розробки адаптивного багатокомпонентного каскаду з параметричною конфігурацією, оптимізованого під голосовий діапазон;
- підхід до лематизації флексивних мов шляхом впровадження character-level Seq2Seq архітектури з двонаправленим GRU-кодувальником та механізмом уваги, адаптованої до концептуально-семантичного відображення для УЖМ.

Набуло подальшого розвитку:

- технологія конфігурованих ланцюжків обробки для низькоресурсних мов із флексивною морфологією у задачах перекладу на жестову мову;
- підхід до контекстної корекції помилок розпізнавання на рівні речення шляхом адаптації байтової Transformer-моделі до специфіки українських ASR-помилки.

Практична значущість отриманих результатів полягає у розробці програмного комплексу для автоматичного перекладу українського мовлення на УЖМ у реальному часі. Каскад аудіофільтрації забезпечує ефективне зниження

рівня шуму при збереженні акустичних характеристик голосу. Підсистема автокорекції суттєво перевищує наявні аналоги за точністю виправлення помилок розпізнавання при прийнятному часі обробки. Підсистема концептуально-семантичної лематизації демонструє суттєво вищу точність порівняно з граматичними лематизаторами загального призначення. Комплексне тестування підтвердило практичну придатність системи для застосування в режимі реального часу.

Повнота опублікування основних положень дисертаційної роботи

За результатами дисертаційного дослідження підготовлено 23 наукових праць, в тому числі 5 статей у фахових наукових виданнях.

Наукові положення та новизна дисертації були презентовані та обговорені на 18 наукових заходах всеукраїнського й міжнародного рівня, що засвідчено відповідними публікаціями тез.

Обґрунтованість наукових положень, висновків і рекомендацій, та достовірність отриманих у роботі результатів підтверджується коректним вибором математичних моделей і методів, детальним порівняльним аналізом з відомими аналогами та верифікацією на реальних мовленнєвих даних із застосуванням стандартних метрик оцінки якості.

Структура та зміст дисертаційного дослідження, рівень виконання поставленого наукового завдання

Дисертаційна робота складається зі вступу, чотирьох основних розділів та висновків.

Перший розділ присвячено аналізу соціального контексту та актуальності проблеми, огляду вітчизняних та міжнародних досліджень у сфері автоматичного розпізнавання мовлення і лематизації. Виявлено принципову прогалину в існуючих підходах і сформульовано мету, задачу та наукову новизну дослідження.

Другий розділ містить теоретичні основи та методи трансформації мовлення на жестову мову. Розглянуто теоретичні засади автоматичного розпізнавання мовлення, методи корекції помилок розпізнавання, методи концептуально-семантичної лематизації, методи цифрової фільтрації аудіосигналів та синтезу жестового відображення. Обґрунтовано вибір

компонентів системи та описано програмні засоби реалізації.

Третій розділ присвячено практичній реалізації та верифікації підсистем: адаптивної аудіофільтрації, автокорекції помилок розпізнавання та концептуально-семантичної лематизації. Для кожної підсистеми наведено результати порівняльного тестування.

Четвертий розділ описує інтеграцію підсистем у єдиний ланцюжок обробки реального часу з графічним інтерфейсом. Наведено результати комплексного тестування, аналіз розподілу обчислювального часу та оцінку досягнутих показників якості. Узагальнені результати підтверджують ефективність запропонованих методів і засобів для вирішення задачі трансформації звукової інформації в жестову мову.

Узагальнені результати підтверджують ефективність запропонованих методів і засобів для вирішення задачі трансформації звукової інформації в жестову мову.

Зауваження та побажання до дисертаційної роботи

У цілому дисертаційна робота Луца Владислава Євгеновича заслуговує позитивної оцінки, проте необхідно зазначити деякі зауваження.

1. У роботі не описано, яким чином забезпечується якість і надійність автоматично згенерованого навчального набору даних (створеного через синтез мовлення й повторне розпізнавання) - чи перевірялися отримані пари "помилка-виправлення" людиною, чи весь процес повністю автоматичний і покладається на якість вихідних алгоритмів синтезу й розпізнавання.
2. Не висвітлено, як нейромережеві моделі поведуться з мовленням кількох людей одночасно чи з мовленням на тлі інших голосів - чи розрахована система виключно на один чіткий голос диктора.
3. У роботі недостатньо обговорено потенційні обмеження застосування нейромережевих моделей у контексті жестової мови - зокрема, чи може система коректно передавати емоційне забарвлення чи інтонаційні відтінки усного мовлення, які в жестовій мові часто виражаються мімікою та невербальними засобами, а не лише вибором конкретного жесту.

Загальні висновки

Аналіз дисертаційної роботи свідчить про те, що вона є цілісним науковим дослідженням, виконаним на високому теоретичному та прикладному рівні. Всі результати, що мають елементи наукової новизни, обґрунтовані автором і підтверджені експериментальними дослідженнями, що дало змогу сформулювати практичні рекомендації щодо розробки систем автоматичного перекладу мовлення на жестову мову.

Тематика дослідження повністю відповідає спеціальності 122 «Комп'ютерні науки» галузі знань 12 «Інформаційні технології» та сприяє подальшому розвитку як прикладних, так і теоретичних положень у сфері нейромережевої обробки мовлення та інклюзивних інформаційних технологій.

Беручи до уваги рівень наукової новизни, достовірність отриманих результатів, використання сучасного методичного інструментарію, а також наявність належної публікаційної активності, можна зробити висновок, що дисертаційна робота й оприлюднені за її результатами наукові праці повністю відповідають вимогам, визначеним пунктами 6-9 Постанови Кабінету Міністрів України від 12 січня 2022 року №44 «Про затвердження Порядку присудження ступеня доктора філософії». Автор дисертації, Луц Владислав Євгенович, продемонстрував достатній рівень володіння методологією наукових досліджень і заслуговує на присудження наукового ступеня доктора філософії в галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки».

Завідувач кафедри теоретичної кібернетики,

Київський національний університет

імені Тараса Шевченка,

доктор фізико-математичних наук, професор



Юрій КРАК

Завідувач
Ю. Крак



Заст. декана
Тетяна ШАХОВКО