

<https://doi.org/10.31891/2219-9365-2026-86-21>

УДК 004.9

ОВЧАРЕНКО Максим

Національний технічний університет «Дніпровська політехніка»

<https://orcid.org/0009-0006-0730-0913>

e-mail: ovcharenko.m.a@nmu.one

КАШТАН Віта

Національний технічний університет «Дніпровська політехніка»

<https://orcid.org/0000-0002-0395-5895>

e-mail: kashtan.v.yu@nmu.one

МУЛЬТИМОДАЛЬНИЙ НЕЙРОМЕРЕЖЕВИЙ МЕТОД АНАЛІЗУ МОВНОЇ ВАРІАТИВНОСТІ ДІАЛОГІВ КОНТАКТ-ЦЕНТРУ НА ОСНОВІ АКУСТИЧНИХ І ЛЕКСИЧНИХ ОЗНАК

Запропоновано мультимодальний нейромережовий метод аналізу мовної варіативності діалогів контакт-центру з використанням акустичних та лексико-морфологічних ознак мовлення. Актуальність дослідження зумовлена необхідністю автоматизації моніторингу якості україномовного діалогу в умовах поширення змішаних мовних практик і емоційно забарвлених мовних конструкцій, які знижують точність стандартних систем автоматизованого розпізнавання мовлення та ускладнюють об'єктивне оцінювання якості комунікації. Запропонований в роботі метод передбачає використання мультимодальної нейромережової архітектури, що поєднує Transformer-енкодер для формування контекстно-залежних текстових представлень і згорткові енкодери для вилучення інформативних акустичних характеристик мовлення, зокрема мел-частотних кепстральних коефіцієнтів та спектральних параметрів. У межах методу введено формалізовані показники оцінювання мовної варіативності – коефіцієнт лінгвістичних аномалій та інтегральний індекс мовної чистоти діалогу, що забезпечують кількісне вимірювання відхилень від нормативного мовлення у режимі реального часу. Експериментальна апробація методу проведена на мультимодальному корпусі тривалістю 50 годин, сформованому з аудіозаписів контакт-центрів, інтерв'ю та медійних діалогів. Отримані результати підтверджують наявність статистично значущого зв'язку між рівнем мовної варіативності та показниками якості автоматизованого розпізнавання мовлення, зокрема метрикою Word Error Rate. Для практичного використання розроблено програмний інструментарій візуалізації «карти мовної чистоти», що відображає динаміку мовленнєвих характеристик діалогу та може бути використаний для автоматизованого моніторингу якості роботи операторів контакт-центрів.

Наукова новизна дослідження полягає у розробці мультимодального нейромережового методу аналізу україномовної варіативності діалогів контакт-центру, який базується на інтегрованій обробці акустичних характеристик мовлення та лексико-морфологічних ознак у спільному просторі представлень і передбачає використання формалізованих кількісних показників оцінювання – коефіцієнта лінгвістичних аномалій та інтегрального індексу мовної чистоти діалогу, що забезпечує автоматизоване виявлення змішаних мовних конструкцій і комплексну оцінку якості мовленнєвої взаємодії у режимі реального часу.

Ключові слова: мультимодальний аналіз, нейронні мережі, контакт-центр, індекс мовної чистоти, аудіо дані, лексико-морфологічні ознаки

OVCHARENKO Maksym, KASHTAN Vita

Dnipro University of Technology

MULTIMODAL NEURAL NETWORK METHOD FOR ANALYZING LINGUISTIC VARIABILITY IN CONTACT CENTER DIALOGUES BASED ON ACOUSTIC AND LEXICAL FEATURES

The paper proposes a multimodal neural network method for analyzing linguistic variability in contact center dialogues by integrating acoustic and lexical-morphological features. The relevance of the study stems from the need to automate the monitoring of Ukrainian-language communication quality in real conversational environments, characterized by mixed speech patterns and emotionally colored expressions, which degrade the performance of conventional automatic speech recognition systems. The proposed approach is based on a hybrid multimodal architecture that combines a Transformer encoder for context-aware textual representation learning and convolutional neural encoders for extracting informative acoustic characteristics, including Mel-frequency cepstral coefficients and spectral features. A unified feature fusion mechanism enables the formation of joint embeddings reflecting phonetic variability, lexical deviations, and discourse dynamics. To quantify deviations from normative speech, formalized evaluation indicators are introduced: the Linguistic Anomaly Coefficient and the Linguistic Cleanliness Index which enable real-time assessment of linguistic variability and communication quality. Experimental validation was conducted on a multimodal corpus comprising 50 hours of dialogue recordings collected from contact centers, interviews, and media interactions. The results demonstrate a statistically significant correlation between the degree of mixed speech usage and automatic speech recognition performance measured by Word Error Rate. A practical visualization tool, the "linguistic cleanliness map," was developed to represent temporal changes in dialogue quality through confidence-based color zoning.

The scientific contribution of the study lies in the development of a multimodal neural network framework for integrated acoustic-textual analysis of linguistic variability and in the introduction of quantitative indicators enabling automated evaluation of speech quality in service communication environments.

Keywords: multimodal learning, deep neural networks, speech analysis, contact center dialogue, linguistic variability, acoustic features.

Стаття надійшла до редакції / Received 21.03.2026
Прийнята до друку / Accepted 20.04.2026
Опубліковано / Published 31.05.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© ОВЧАРЕНКО Максим, КАШТАН Віта

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОКІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

У автоматизованих контакт-центрах якість мовленнєвої взаємодії між оператором та клієнтом є одним із чинників ефективного обслуговування, рівня довіри до сервісу та результативності виконання цільових сценаріїв діалогу [1]. Швидкий розвиток систем автоматичного розпізнавання мовлення, аналізу діалогів та інтелектуальної підтримки операторів зумовлює потребу у створенні методів, здатних адекватно враховувати реальні особливості мовної поведінки користувачів, зокрема варіативність вимови, лексики та стилю спілкування [2].

Однією з характерних особливостей мовленнєвої комунікації в Україні є поширення змішаного мовлення, що проявляється у варіюванні фонетичних, лексичних та граматичних ознак як між різними мовцями, так і в межах мовлення одного індивіда [3]. Така варіативність мови ускладнює автоматичний аналіз діалогів, знижує точність моделей розпізнавання та семантичного інтерпретування мовлення, а також впливає на об'єктивність оцінювання якості роботи операторів контакт-центрів. Попри значний інтерес до дослідження змішаних мовних практик, емпіричні роботи, спрямовані на аналіз їх акустичних характеристик у прикладних задачах інтелектуальних інформаційних систем, залишаються обмеженими.

Традиційні підходи до побудови систем діалогового аналізу базуються на модульних конвеєрних архітектурах, у яких етапи розпізнавання мовлення, виділення ознак, розуміння намірів та оцінювання якості діалогу оптимізуються незалежно. Хоча такі рішення забезпечують контрольованість окремих компонентів, вони не завжди гарантують узгоджене підвищення загальної ефективності системи. Натомість сучасні наскрізні нейронні моделі [4, 5], зокрема архітектури на основі механізму уваги та Transformer-підходу, демонструють високий потенціал у задачах моделювання діалогової взаємодії, однак у більшості випадків орієнтовані переважно на текстові представлення мовлення та недостатньо враховують його акустичну природу.

Незважаючи на значні досягнення у застосуванні рекурентних і згорткових нейронних мереж для аналізу мовленнєвих і текстових даних, сучасні дослідження дедалі більше орієнтуються на використання архітектур на основі механізму уваги, зокрема Transformer-моделей, які забезпечують ефективне моделювання довготривалих залежностей і формування контекстно-чутливих представлень. Водночас більшість існуючих підходів реалізує аналіз або текстової складової мовлення, або його акустичних характеристик, що обмежує можливості адекватного відображення складної структури реальної мовленнєвої взаємодії. Це є особливо важливим у випадку змішаних мовних практик, де варіативність вимови поєднується з лексичними та морфологічними відхиленнями від нормативного мовлення. За таких умов одномодальні моделі демонструють зниження стійкості та точності аналізу. У зв'язку з цим актуальним напрямом досліджень є розроблення мультимодальних методів, що передбачають інтегроване використання акустичних і лексичних ознак для комплексного аналізу мовної варіативності діалогів. Поєднання різномірних джерел інформації дозволяє підвищити робастність нейромережових моделей до варіацій мовлення та створює передумови для більш об'єктивного автоматизованого оцінювання якості професійної комунікації в контакт-центрах.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Розвиток методів машинного навчання сприяв появі ефективних підходів для автоматизованого аналізу текстових і мовленнєвих даних, зокрема задач класифікації, визначення емоційної тональності та виявлення мовних закономірностей. Сьогодні у наукових дослідженнях для аналізу природної мови широко застосовуються методи обробки природної мови та алгоритми машинного навчання, що дозволяють автоматизувати виділення інформативних ознак із великих масивів неструктурованих даних. Особливого поширення набули методи глибокого навчання, які демонструють високі результати у задачах класифікації текстів і мовлення. Зокрема, рекурентні нейронні мережі типу Long Short-Term Memory [6] ефективно моделюють послідовні залежності в текстових даних та забезпечують урахування контекстуальних зв'язків між мовними одиницями. Це робить їх придатними для аналізу діалогів, моделювання мовної динаміки та визначення емоційних характеристик висловлювань. З іншого боку, згорткові нейронні мережі (CNN), які спочатку використовувалися переважно для обробки зображень, були успішно адаптовані для аналізу текстових і мовленнєвих сигналів [7]. Завдяки здатності виявляти локальні закономірності та повторювані ознаки такі моделі показали високу ефективність у задачах класифікації текстів, аналізу тональності та розпізнавання мовленнєвих характеристик. У задачах обробки мовлення CNN застосовуються для автоматичного виділення спектральних та просодичних ознак, що дозволяє підвищити точність моделей розпізнавання та інтерпретації мовлення.

Особливу увагу привертає українсько-російське змішане мовлення, яке широко використовується у повсякденній комунікації та характеризується варіюванням фонетичних, лексичних і граматичних ознак. У роботі [8] досліджено лексичні особливості змішаного українського ідіому на основі кількісного аналізу великого мовного корпусу. Основна увага приділяється питанню впливу мови на формування лексичного складу змішаного мовлення, а також можливості зниження варіативності внаслідок стабілізації окремих мовних форм. Результати дослідження свідчать про те, що для більшості конкурентних українських лексем спостерігається тенденція до домінування одного варіанта, що вказує на поступову еволюцію змішаного мовлення у напрямі стабілізованої мовної системи. Для частини одиниць зберігається висока варіативність, що ускладнює формалізацію мовних закономірностей та їх використання у прикладних інформаційних системах.

Розроблення інструментів автоматизованого моніторингу мовлення розглянуто у [9], де запропоновано програмну систему аналізу усного мовлення, спрямовану на підвищення культури спілкування та дотримання мовленнєвих норм. У роботі представлено концептуальну модель системи, що включає визначення вхідних і вихідних даних, функціональних модулів та програмних засобів реалізації. Автори підкреслюють важливість інтеграції засобів мовного аналізу для підвищення усвідомлення користувачами мовних стандартів, однак запропонований підхід переважно орієнтований на текстову складову мовлення та не враховує акустичні параметри, що є суттєвими при аналізі змішаних мовних практик.

Значний внесок у розвиток методів автоматизованого аналізу змішаного мовлення зроблено у роботах, присвячених дослідженню code-mixing у багатомовних середовищах. Так, авторами в роботі [10] розглянуто задачу аналізу тональності код-змішаних текстів із використанням методів машинного навчання та глибоких нейронних мереж. Дослідження демонструє, що поєднання різних лінгвістичних ознак дозволяє підвищити точність класифікації емоційної тональності, однак обмеженням є використання лише текстових представлень. У роботі [11] запропоновано підхід до аналізу код-змішаних текстів на основі графових згорткових мереж із багатоголовою увагою. Така архітектура забезпечує ефективне моделювання залежностей між словами в умовах змішаного мовлення та покращує результати класифікації. Хоча дослідження зосереджене на обробці текстових даних, але не враховує варіативність акустичних характеристик мовлення. Автори в роботі [12] здійснили порівняльний аналіз підходів до навчання моделей аналізу тональності у код-змішаних текстах, показавши, що використання контекстно-залежних мовних представлень сприяє підвищенню якості класифікації. Проте результати дослідження підтверджують, що одномодальні текстові моделі мають обмежену стійкість до реальних варіацій мовлення, зокрема фонетичних і просодичних.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою роботи є розробка мультимодального нейромережевого методу аналізу мовної варіативності діалогів контакт-центру на основі інтеграції акустичних і лексичних ознак для автоматизованого виявлення змішаних мовних конструкцій та підвищення об'єктивності оцінювання якості мовленнєвої взаємодії операторів.

Для досягнення поставленої мети необхідно вирішити такі задачі:

- сформувати мультимодальний корпус мовленнєвих даних, що включає аудіозаписи діалогів різних типів комунікації та їх текстові транскрипції;
- виконати попередню обробку мовленнєвих сигналів, зокрема шумозаглушення, сегментацію мовленнєвої активності та нормалізацію параметрів аудіо даних;
- виконати виділення інформативних акустичних ознак мовлення, включаючи мел-частотні кепстральні коефіцієнти, спектральні характеристики та темп мовлення;
- забезпечити автоматичну транскрипцію мовлення з подальшим лексико-морфологічним аналізом тексту для виявлення мовних варіацій, маркерів змішаного мовлення та розмовних конструкцій;
- сформувати векторні представлення текстових ознак на основі сучасних мовних моделей для подальшої інтеграції з акустичними характеристиками;
- розробити мультимодальну нейромережеву архітектуру, що поєднує згортковий енкодер для акустичних ознак і трансформер-енкодер для текстових представлень із шаром об'єднання ознак;
- реалізувати механізм оцінювання подібності та залежності між модальностями, зокрема із використанням сіамських або інших метричних блоків навчання представлень;
- розробити модель класифікації рівня мовної варіативності та емоційного стану мовця, що впливають на якість діалогової взаємодії;
- сформувати інтегральну карту мовної чистоти діалогу, яка дозволяє здійснювати автоматизоване оцінювання професійної діяльності операторів контакт-центру;
- провести експериментальне дослідження ефективності запропонованого методу та оцінити його вплив на точність аналізу діалогів порівняно з одномодальними підходами.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Запропонований в даній роботі метод спрямований на автоматизований аналіз мовної варіативності діалогів контакт-центру шляхом інтеграції акустичних і лексичних ознак мовлення в єдиному нейронмережевому просторі представлень. Загальна структура методу представлена на рис.1 та включає етапи формування мультимодального набору даних, попередньої обробки, виділення ознак, побудови нейронмережевої моделі та формування інтегральної оцінки мовної чистоти діалогу.

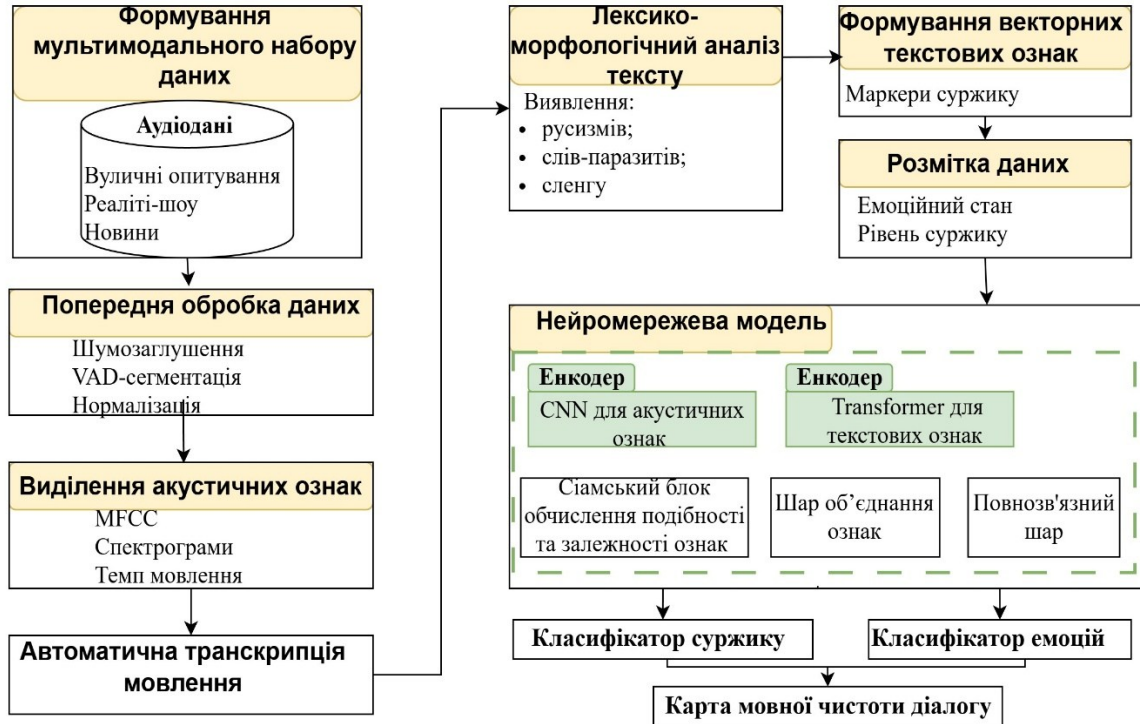


Рис. 1. Структурна схема мультимодального нейронмережевого методу аналізу мовної варіативності діалогів контакт-центру

На першому етапі здійснюється формування корпусу аудіоданих, що включає записи діалогів різних типів комунікаційних сценаріїв (контакт-центр, публічні інтерв'ю, реаліті-шоу, новинні матеріали тощо). Таке поєднання джерел забезпечує репрезентативність мовних варіацій та дозволяє врахувати різні стилі мовлення, темп комунікації та ступінь використання змішаних мовних конструкцій. Для кожного аудіофрагмента формується відповідна текстова транскрипція, що створює мультимодальне представлення діалогу у вигляді пари «аудіосигнал – текст». Нехай мультимодальний набір даних представлений множиною діалогових фрагментів:

$$D = \{(A_i, T_i, y_i^{(s)}, y_i^{(e)})\}_{i=1}^N, \tag{1}$$

де A_i – аудіосигнал i -го фрагмента діалогу, T_i – відповідна текстова транскрипція, $y_i^{(s)}$ – мітка рівня мовної варіативності (ступінь змішаного мовлення), $y_i^{(e)}$ – мітка емоційного стану мовця.

Таблиця 1

Фрагмент статистичних ознак первинних аудіо даних

ID файлу	Тип сценарію	Тривалість (с)	К-сть сегментів
test1.m4a	Реаліті-шоу / Побутовий	230	48
test2.m4a	Вуличне опитування	30	6
test3.m4a	Новинний матеріал	100	25
test4.m4a	Контакт-центр (Конфлікт)	150	32

Для забезпечення високої точності класифікатора суржику та емоцій було проаналізовано загальний масив аудіо даних обсягом 50 годин, що пройшли етапи попередньої обробки: процедури шумозаглушення, нормалізації рівня сигналу та сегментації мовленнєвої активності (Voice Activity Detection). Це дозволяє зменшити вплив зовнішніх перешкод, забезпечити однорідність вхідних даних і підвищити стабільність подальшого виділення ознак.

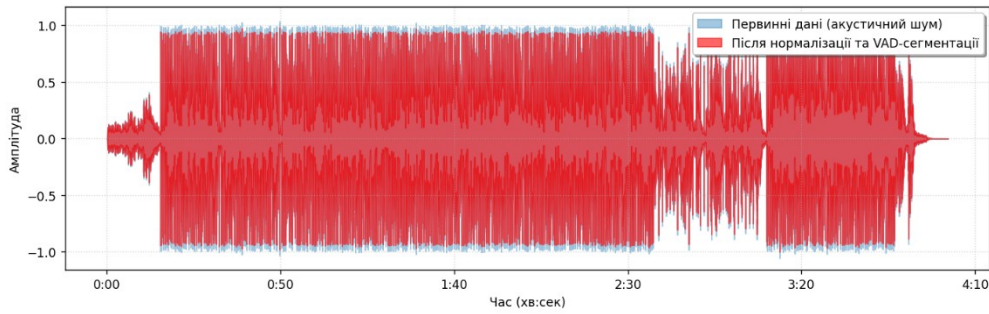


Рис. 2. Результат попередньої обробки аудіоданих для файлу test1

Після попередньої обробки з аудіосигналу формується матриця акустичних ознак:

$$X_i^{(a)} \in \mathbb{R}^{F \times L}, \quad (2)$$

де F – кількість спектральних ознак (наприклад, MFCC), L – кількість часових фреймів.

Із сегментованих мовленнєвих сигналів виділяються інформативні акустичні характеристики, зокрема мел-частотні кепстральні коефіцієнти (MFCC), спектральні параметри та показники темпу мовлення. Отримані ознаки відображають фонетичні особливості вимови, просодичні характеристики та динаміку мовлення, що є важливими для виявлення мовної варіативності.

Подальший етап передбачає автоматичне перетворення мовлення у текст із використанням систем розпізнавання мовлення. Отримані транскрипції піддаються лексико-морфологічному аналізу, у процесі якого визначаються маркери мовної варіативності, розмовні конструкції, запозичення та інші індикатори змішаного мовлення.

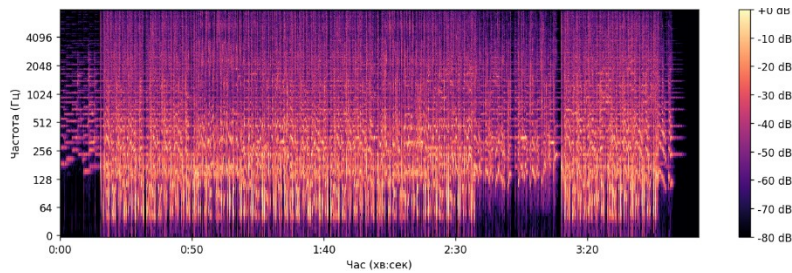


Рис. 3. Результат етапу виділення акустичних ознак для файлу test1

На основі результатів аналізу формується набір текстових ознак, що кодуються у вигляді векторних представлень для подальшого використання в нейромережевій моделі.

Центральним компонентом методу є мультимодальна нейромережева архітектура, що поєднує два енкодери. Перший енкодер, реалізований на основі згорткової нейронної мережі, здійснює навчання представлень акустичних ознак. Згортковий енкодер виконує нелінійне перетворення [7]:

$$h_i^{(a)} = f_{\theta_a}(X_i^{(a)}), \quad (3)$$

де f_{θ_a} – CNN-модель з параметрами θ_a , $h_i^{(a)} \in \mathbb{R}^{d_a}$ – вектор акустичного представлення.

Другий енкодер, побудований на базі Transformer-архітектури, формує контекстно-залежні текстові представлення. Текстова транскрипція перетворюється у послідовність токенів

$$T_i = \{w_1, w_2, \dots, w_M\}. \quad (4)$$

Transformer-енкодер формує контекстно-залежне представлення:

$$h_i^{(t)} = f_{\theta_t}(T_i), \quad (5)$$

де f_{θ_t} – модель Transformer, $h_i^{(t)} \in \mathbb{R}^{d_t}$.

Для узгодження інформації між модальностями використовується шар об'єднання ознак, який інтегрує акустичні та текстові вектори у спільний простір представлень. Для об'єднання інформації використовується шар злиття (fusion layer) [7]:

$$h_i = \phi(W_a h_i^{(a)} \oplus W_t h_i^{(t)}), \quad (6)$$

де \oplus – операція конкатенації, W_a, W_t – матриці проєкції у спільний простір, $\phi(\cdot)$ – нелінійна активація (ReLU), $h_i \in \mathbb{R}^{d}$ – мультимодальне представлення.

Додатково застосовується сіамський блок, що дозволяє моделювати подібність та залежності між різними типами ознак і підвищує дискримінаційну здатність моделі.

На виході нейромережевої моделі формується узагальнене представлення мовленнєвого фрагмента, яке використовується для подальших задач класифікації. На основі сформованих мультимодальних представлень реалізуються два класифікаційні модулі: класифікатор рівня мовної варіативності [13] та

класифікатор емоційного стану мовця. Такий підхід дозволяє враховувати не лише мовні характеристики, але й психоемоційний контекст діалогу, що може впливати на якість комунікації.

Фінальним етапом методу є побудова інтегральної карти мовної чистоти діалогу, яка відображає рівень використання змішаних мовних конструкцій у поєднанні з емоційними характеристиками мовлення. Отримана оцінка може застосовуватися для автоматизованого моніторингу якості роботи операторів контакт-центрів, підтримки прийняття управлінських рішень та вдосконалення сценаріїв комунікації.

На основі проведеного аналізу 50 годин аудіоданих та детального вивчення чотирьох репрезентативних фрагментів, представлено результати дослідження. У таблиці 2 представлено метрики: лексико-морфологічного аналізу (*LAC*), точності автоматичного розпізнавання (*WER*) та загальної кількості лінгвістичних аномалій для кожного сценарію.

Аналіз даних табл.2 свідчить про наявність вираженої кореляції між типом контенту та рівнем лінгвістичних відхилень. Мінімальне значення коефіцієнта лексико-морфологічного аналізу ($LAC=0,043$) зафіксовано у сценарії «Новини», що відповідає найвищому показнику точності транскрибування ($WER=37,41\%$). Найвищий рівень специфічних термінів, емоційних маркерів (суржик), що відхиляються від стандартної україномовної моделі ($LAC=0,288$) виявлено у роботі контакт-центрів.

Візуалізація результатів у форматі «Карти мовної чистоти» (рис. 4) дозволяє простежити часову динаміку якості мовлення. Карта мовної чистоти (Language Purity Map) – це інструмент візуального моніторингу, розроблений для автоматизованого аналізу якості усного мовлення в режимі реального часу. На рис.4 відображено зміну індексу Language Purity Index (*LCI*) протягом часу розмови. Індекс *LCI* обчислюється на основі параметрів: впевненості моделі та коефіцієнта аномалій.

Таблиця 2

Метрики аудіо даних

ID файлу	Сценарій комунікації	Avg LAC (Індекс аномалій)	WER (%) (Помилка розпізнавання)	Кількість виявлених аномалій
test1.m4a	Реаліті-шоу	0.160	44.21%	7
test2.m4a	Вуличне опитування	0.148	50.91%	2
test3.m4a	Новини (Норма)	0.043	37.41%	4
test4.m4a	Контакт-центр	0.288	43.47%	2

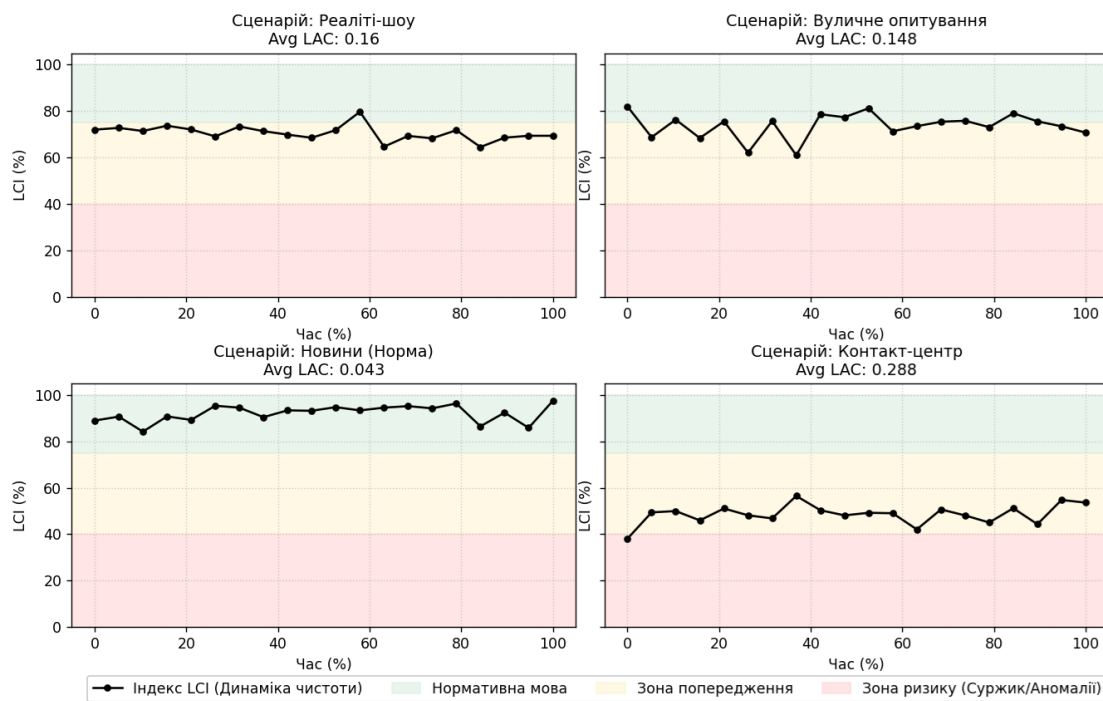


Рис. 4. Карта мовної чистоти аудіо даних

Візуалізація результатів на рис. 4 дозволяє простежити часову динаміку якості мовлення, а саме:

- зона нормативного мовлення (75–100%) характерна для офіційних повідомлень (test3), де лінія тренду демонструє стабільність без різких амплітудних коливань;
- зона попередження (40–75%) охоплює більшість сегментів побутового мовлення (test1, test2), повідомляє про періодичне вживання суржиків та слів-паразитів;

– зона ризику (нижче 40%) локальні мінімуми на графіках відповідають моментам ідентифікації грубих морфологічних аномалій або русизмів.

Для верифікації роботи запропонованого методу було проведено детальний якісний аналіз розпізнаних сегментів, що дозволило класифікувати виявлені аномалії за трьома категоріями: морфологічні, лексичні та стилістичні (сленг/паразити). Найбільше словесних аномалій зафіксовано у сценарії розмов контакт-центрів (test4.m4a) (табл.3).

Таблиця 3

Класифікація виявлених мовних аномалій

Категорія аномалій	Приклади виявлених дериватів та запозичень	Лінгвістична характеристика
Морфологічні	розкажуть, прийдували, бачим, злазить	Порушення парадигми дієвідмінювання, гібридні форми на стику мов.
Лексичні (русизми)	вообщем, ето, походу, последний, сахар	Прямі лексичні запозичення та деформовані лексеми, що порушують чистоту коду.
Сленг та слова-паразити	тіпа, кароче, шо, поняв	Дискурсивні маркери, що знижують офіційно-діловий статус комунікації.

Отримані дані з табл.3 підтверджують гіпотезу про те, що індекс *LCI* є чутливим не лише до акустичного шуму, а й до семантичної та морфологічної структури висловлювання.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

У роботі представлено мультимодальний нейромережевий метод аналізу мовної варіативності діалогів, адаптований до специфіки української мови в умовах контакт-центрів та медіа-платформ. Запропонована мультимодальна архітектура за рахунок інтеграції акустичних та лексико-морфологічних ознак забезпечує комплексний моніторинг якості мовлення. Використання сіамської структури та трансформерних енкодерів дозволило досягти стабільної ідентифікації суржиків та емоційних аномалій у реальному часі. Встановлено, що лінія тренду індексу мовної чистоти є ефективним діагностичним інструментом: значення у «зеленій зоні» ($LCI > 75\%$) відповідають літературній нормі (сценарій «Новини», $LAC=0,043$), тоді як значення у «червоній зоні» ($LCI < 40\%$) чітко маркують вживання лексичних русизмів та морфологічних деформацій (сценарій «Контакт-центр», $LAC=0,288$). Експериментально підтверджено залежність точності розпізнавання від лінгвістичної чистоти сигналу. Мінімальна похибка $WER = 37,41\%$ зафіксована на еталонних зразках, тоді як наявність суржиків та акустичних завад (вуличний шум) підвищує значення WER до $50,91\%$, що обґрунтовує необхідність впровадження блоків шумозаглушення та нормалізації для етапу попередньої обробки даних. Класифіковано типові мовні аномалії, серед яких найбільш деструктивний вплив на індекс чистоти мають гібридні морфологічні форми («розкажуть») та лексичні запозичення («вообщем», «ето»). Це дозволяє використовувати метод для автоматизованого аудиту роботи операторів контакт-центрів та оцінки культури мовлення в публічному просторі.

Напрями подальших досліджень у межах розробленого мультимодального нейромережевого методу передбачають інтеграцію методів глибинного навчання для автоматичного виявлення нових неологізмів та гібридних форм суржиків, що дозволить системі самостійно оновлювати бази даних лексичних аномалій без ручного втручання лінгвістів. Це підвищить адаптивність моделі до регіональних діалектів та молодіжного сленгу.

References

1. Kashtan, V., Hnatushenko, V., Ovcharenko, M., & Ivanko, A. (2025) A hybrid neural network architecture for semantic-contextual analysis of emotions in social media. *MoDaST 2025: Modern Data Science Technologies Doctoral Consortium*, pp. 15-28. <https://ceur-ws.org/Vol-4005/paper2.pdf>
2. Chakravarthi, B. R., Priyadharshini, R., Thavareesan, S., Chinnappa, D., Thenmozhi, D., Sherly, E., McCrae, J. P., Hande, A., Ponnusamy, R., Banerjee, S., & Vasantharajan, C. (2021) Findings of the Sentiment Analysis of Dravidian Languages in Code-Mixed Text. *Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation*, CEUR.
3. Zeller, J. P. (2022) Іkannia and akannia in Ukrainian-Russian mixed speech (“surzhyk”). *Ukrains'ka Mova*, vol. 2(82), pp. 38–59. <https://doi.org/10.15407/ukrmova2022.01.038>
4. Priyadharshini, R., Chakravarthi, B. R., Thavareesan, S., Chinnappa, D., Thenmozhi, D., & Sherly, E. (2021) Overview of the DravidianCodeMix 2021 shared task on sentiment detection in tamil, malayalam, and kannada. *Forum for Information Retrieval Evaluation, FIRE 2021*, Association for Computing Machinery.
5. Mahata, S., Das, D., & Bandyopadhyay, S. (2021) Sentiment classification of code-mixed tweets using bi-directional RNN and language tags. *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, Association for Computational Linguistics, pp. 28–35. <https://aclanthology.org/2021.dravidianlangtech-1.4>
6. Sultan, D., Mendes, M., Kassenkhan, A., & Akylbekov, O. (2023) Hybrid CNN-LSTM Network for Cyberbullying Detection on Social Networks using Textual Contents. *International Journal of Advanced Computer Science and Applications*, vol. 14(9).
7. Choi, J., & Zhang, X. (2022) Classifications of restricted web streaming contents based on convolutional neural network and long short-term memory (CNN-LSTM). *J. Internet Serv. Inf. Secur.*, vol. 12(3), pp. 49-62.

8. Hentschel, G. (2024) Ukrainian and Russian in the lexicon of Ukrainian Surżyk: reduced variation and stabilisation in central Ukraine and on the Black Sea coast. *Russ Linguist*, vol. 48(2). <https://doi.org/10.1007/s11185-023-09286-9>
9. Shestakevych, T., & Kobylukh, L. (2021) Designing an Application for Monitoring the Ukrainian Spoken Language. *Proceedings of the 1st Workshop on Interactive Systems and Information Society Technologies (ISIST 2021)*.
10. Mandalam, A. V., & Sharma, Y. (2021) Sentiment analysis of Dravidian code mixed data. *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, Association for Computational Linguistics, pp. 46–54. <https://aclanthology.org/2021.dravidianlangtech-1.6>
11. Dowlagar, S., & Mamidi, R. (2021) Graph convolutional networks with multi-headed attention for code-mixed sentiment analysis. *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, Association for Computational Linguistics, pp. 65–72. <https://aclanthology.org/2021.dravidianlangtech-1.8>
12. Balouchzahi, F., & Shashirekha, H. L. (2021) LA-SACo: A study of learning approaches for sentiments analysis in Code-mixing texts. *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, Association for Computational Linguistics, pp. 109–118. <https://aclanthology.org/2021.dravidianlangtech-1.14>
13. Gmdn. (n.d.) *Specific rules for Surzhyk*. GitHub. https://github.com/gmdn/Surzhyk/blob/master/tables/specific_rules.docx