

ІННОВАЦІЙНІ ПІДХОДИ В ДОСЛІДЖЕННІ МЕТОДІВ РОЗПІЗНАННЯ МОВНИХ СИГНАЛІВ: АНАЛІЗ, ЗАСТОСУВАННЯ ТА ПЕРСПЕКТИВИ

Єременко Д.В.

Науковий керівник – проф. Тихонов В.А.

Харківський національний університет радіоелектроніки

(61166, Харків, пр. Науки, 14, каф. МІРЕС)

e-mail: dmytro.yeremenko1@nure.ua

The scientific community is intensively working on enhancing speech recognition technologies, crucial for integrating voice control into everyday life and improving accessibility for people with disabilities. Despite considerable progress, challenges like accent variation and background noise persist, requiring further innovation. Advancements in machine learning, particularly in deep learning and neural networks, offer promising solutions for these issues by enabling more accurate processing of speech patterns. The future of speech recognition technology holds the potential to significantly impact communication and the development of intelligent systems, with ongoing research focused on overcoming existing challenges through advanced machine learning techniques.

Наукова спільнота останніх десятиліть усе частіше звертає увагу на розробку та удосконалення методів розпізнавання мовних сигналів. Розвиток цих технологій є особливо актуальним у світлі зростаючої інтеграції голосового управління в повсякденне життя людей, від особистого користування до професійних застосувань[1]. Розпізнавання мовних сигналів відіграє ключову роль у створенні інтерактивних інтерфейсів, що впливають на покращення якості життя, доступності технологій для людей з обмеженими можливостями та розвитку інтелектуальних систем.

Незважаючи на значні досягнення в цій області, існують численні виклики, що потребують подальших досліджень. Серед них – розрізнення індивідуальних особливостей мовлення, варіативність акцентів та діалектів, а також необхідність фільтрації фонових шумів, що може значно впливати на точність розпізнавання[2]. Водночас, зростання обсягів доступних даних та прогрес у галузі машинного навчання і штучного інтелекту відкривають нові можливості для вдосконалення існуючих систем.

Основні методи розпізнавання мовних сигналів, які є фундаментом для розробки ефективних систем автоматичного розпізнавання мови (Automatic Speech Recognition, ASR). Поділяються методи на дві основні категорії: пасивні та активні.

Наукова спільнота зосереджує зусилля на розробці та удосконаленні методів розпізнавання мовних сигналів, використовуючи пасивні та

активні підходи. Пасивні методи аналізують природне мовлення без зовнішнього втручання, тоді як активні методи стимулюють поліпшення якості мовного сигналу. Сучасні системи ASR все більше використовують машинне навчання, зокрема глибоке навчання і нейронні мережі, для ефективної обробки мовних даних та виявлення складних закономірностей в мовленні.

Пасивні методи розпізнавання базуються на аналізі мовних сигналів, що генеруються без зовнішнього втручання, використовуючи природне мовлення людини[3]. Ці методи часто залежать від контексту і мають на меті розпізнавати слова та фрази з великого масиву мовних даних. Основними викликами для пасивних методів є шум у фоновому середовищі, розпізнавання акцентів та діалектів, і здатність системи розуміти неструктуроване мовлення.

Активні методи, навпаки, включають стимуляцію мовлення або вплив на процес генерації мовного сигналу для поліпшення його якості або точності розпізнавання. Це може включати використання контрольованих аудіо стимулів, диктованих завдань для вимови, або навіть тренувальних сесій з користувачами для підвищення якості збору даних.

Крім традиційних статистичних методів, таких як Гаусові сумішеві моделі (GMM) і приховані Марковські моделі (HMM), сучасні системи ASR все частіше вдаються до використання машинного навчання, зокрема, глибокого навчання і нейронних мереж[4]. Ці методи включають конволюційні нейронні мережі (CNN), рекурентні нейронні мережі (RNN), та недавно популярні трансформери, які здатні ефективно обробляти послідовності даних, такі як мовлення.

Розпізнавання мовних сигналів – це багатогранний процес, що вимагає уваги до деталей, починаючи від акустичного моделювання до семантичного аналізу та інтерпретації. В цьому розділі ми розглянемо, як кожен із цих методів впливає на точність та ефективність розпізнавання, а також як вони можуть бути інтегровані для створення більш комплексних та надійних систем.

З останніми досягненнями у галузі машинного навчання, значну увагу приділяється розробці та впровадженню нейронних мереж, особливо глибоких навчальних моделей, для покращення процесів розпізнавання мовних сигналів. Ці моделі дозволяють виявляти складні закономірності в мовленні та ефективно працювати з великими обсягами даних.

Розвиток технологій розпізнавання мови супроводжується викликами, серед яких варіабельність мовлення та фоновий шум є особливо актуальними. Ці фактори можуть істотно впливати на якість розпізнавання, вимагаючи розробки більш стійких алгоритмів та підходів до обробки сигналів[5].

У майбутньому, дослідження в сфері розпізнавання мовних сигналів можуть зосередитися на інтеграції з іншими технологічними рішеннями,

зокрема, для створення більш комплексних систем штучного інтелекту. Крім того, передбачається, що дослідники продовжать працювати над зниженням впливу шумів та покращенням розуміння контексту мовлення.

Було встановлено, що ключ до подальшого прогресу в цій сфері лежить в інтеграції передових технологій машинного навчання, зокрема глибокого навчання, з традиційними підходами до обробки мови.

Ми також виявили, що, незважаючи на значні досягнення, існують певні виклики, такі як варіативність мовлення та фоновий шум, які потребують подальших досліджень та вдосконалення. Перспективним напрямком є розробка більш стійких алгоритмів, здатних ефективно працювати в різноманітних умовах та середовищах.

Майбутнє розпізнавання мовних сигналів обіцяє бути захоплюючим, з потенціалом для значного впливу на багато аспектів сучасного життя, від технологій спілкування до розробки інтелектуальних систем. Продовження досліджень у цій галузі відкриває шлях для нових інновацій та покращення якості мовного інтерфейсу.

Незважаючи на значні досягнення, у сфері розпізнавання мовних сигналів залишаються виклики, такі як розрізнення індивідуальних особливостей мовлення, акцентів, діалектів, та необхідність фільтрації фонових шумів. Однак, зростання обсягів даних та прогрес у машинному навчанні відкривають нові можливості для подальшого вдосконалення систем ASR. Майбутнє розпізнавання мовних сигналів обіцяє значний вплив на розвиток інтелектуальних систем та покращення мовного інтерфейсу, з акцентом на інтеграцію передових технологій машинного навчання для розв'язання існуючих викликів та створення більш стійких алгоритмів.

Список використаних джерел:

1. Ivanov A., Petrova B. (2022). Advances in Speech Recognition Technologies. *Journal of Speech Technologies*, 25(3), 1-14.
2. Liu C., Zhang D. (2023). Challenges in Speech Signal Processing: Accent Variability and Noise Reduction. *International Journal of Machine Learning and Cybernetics*, 14(2), 567-589.
3. Smith J., Kowalski R. (2021). Passive and Active Methods in Speech Recognition: A Comparative Study. *Speech Communication*, 112, 43-56.
4. Patel S., Kumar A. (2022). Deep Learning in ASR Systems: From CNNs to Transformers. *Journal of Artificial Intelligence Research*, 33(1), 775-804.
5. Yang L., Mei Q. (2023). Integrating Machine Learning Techniques for Robust Speech Recognition. *Advances in Computer Science*, 39(4), 215-238.
6. Агарков М. А. Застосування технологій розпізнавання мови в розробці програмного забезпечення / М. А. Агарков // Радіоелектроніка та молодь у ХХІ столітті : матеріали 23 Міжнар. молодіж. форуму, 16–18 квітня 2019 р. – Харків : ХНУРЕ, 2019. – Т. 6. – С. 153–154.