

ЗАСОБИ ТА МЕТОДИ ВИМІРЮВАННЯ ПАРАМЕТРІВ МОВНИХ СИГНАЛІВ

Анна Домбровська¹, Леонід Ящинський², Аліна Третяк¹, Сергій Федосов^{1,2}

TOOLS AND METHODS FOR MEASURING SPEECH SIGNAL PARAMETERS

Anna Dombrovska¹, Leonid Yashchynskiy², Alina Tretiak¹, Serhii Fedosov^{1,2}

¹Волинський національний університет імені Лесі Українки,
просп. Волі, 13, Луцьк, 43025, Україна

²Луцький національний технічний університет, вул. Львівська, 75, Луцьк, 43018, Україна

Abstract. *The paper analyzes methods of processing speech signals that will contribute to increasing the efficiency of voice control systems, in particular in conditions of noise, multilingualism and complexity of speech structure. The basic tasks of speech command processing systems are determined, the classical structural diagram of the voice control system, the stages of processing in the SGC and the classification of speech signal processing methods are considered.*

У сучасних інформаційних технологіях мовлення дедалі частіше виступає основним інтерфейсом взаємодії людини з цифровими системами. Від ефективності та точності опрацювання мовних сигналів залежить якість роботи голосових помічників, систем біометричної ідентифікації, медичних діагностичних інструментів та адаптивних навчальних платформ. Незважаючи на суттєвий прогрес у цій сфері, залишається низка невирішених проблем, пов'язаних із точністю вимірювання ключових параметрів мовлення, зокрема формантних частот, висоти тону та спектральних ознак. Особливо гостро проблема стоїть для української мови, для якої майже відсутні системи розпізнавання з великим словником. Через це існуючі методи обробки мовних сигналів потребують вдосконалення, особливо у сфері сегментації, фільтрації та виділення інформативних ознак.

Складність полягає в тому, що мовні сигнали є нестационарними, індивідуальними для кожного мовця та дуже чутливими до зовнішніх впливів (шумів, ліній зв'язку, акцентів тощо). Більшість алгоритмів, що використовуються, мають певні обмеження в умовах реального середовища. Крім того, у науковому просторі України відчутний дефіцит комплексних досліджень, що аналізують ефективність сучасних методів обробки мовлення на прикладі реальних даних та зіставляють їх із міжнародними підходами.

Обробка мовлення є важливою складовою сучасних технологій, що застосовуються у різних галузях, від автоматичного розпізнавання мовлення до ідентифікації мовця. Вивчення методів вилучення ознак з мовних сигналів є критичним для покращення точності та ефективності таких систем. У роботі [1] автори пропонують новий метод для точного відстеження формант у мовних сигналах, що є важливим етапом у процесах аналізу та розпізнавання мовлення. Їхній метод поєднує кілька підходів, таких як часово-зважений аналіз квазізамкнутої фази, оптимізація залишкової розрідженості та використання лінійного прогнозування для покращення точності. Результати експериментів демонструють значні переваги цього підходу над традиційними інструментами, такими як Wavesurfer і Praat, а також сучасними методами на основі фільтра Калмана та глибоких нейронних мереж, що підвищує ефективність розпізнавання та синтезу мовлення.

Подібні методи значно розширюють можливості автоматичних систем розпізнавання мовлення. У дослідженні [2] обговорюється використання динамічного часового вирівнювання (англ. *Dynamic Time Warping* – DTW) та мел-частотних кепстральних коефіцієнтів (англ. *Mel-Frequency Cepstral Coefficients* – MFCC) для автоматичного

розпізнавання мовлення. Завдяки цим методам з'являється можливість точнішого ідентифікувати мовця, що є важливим для багатьох додатків, від систем безпеки до персональних помічників. DTW дозволяє зменшити вплив шуму та інших перешкод, роблячи процес розпізнавання більш стійким і точним.

Для покращення процесів класифікації даних, особливо у великих масивах інформації, автори [3] пропонують методику класифікації великих даних, що використовує характеристики мел-скали в поєднанні з параметризацією методом k-найближчих сусідів (k-NN). Цей підхід дозволяє оптимізувати процес класифікації звукових сигналів, що важливо для підвищення точності систем розпізнавання мовлення в реальних умовах. Метод k-NN ефективно працює з великими об'ємами даних, що є необхідним для побудови масштабованих систем, здатних обробляти численні мовні сигнали.

У контексті новітніх методів обробки мовлення, хвильові перетворення набувають все більшої популярності. У дослідженні [4] показано, як хвильові перетворення можуть бути використані для виділення характеристик мовних сигналів, що дозволяє значно покращити якість розпізнавання та синтезу мовлення. Це дає можливість ефективно обробляти нестационарні сигнали, що часто виникають під час реального мовлення, і застосовувати ці методи для підвищення точності та адаптивності систем розпізнавання.

Шумові перешкоди є одним із найскладніших викликів для розпізнавання мовлення. У своїй роботі [5] науковці досліджують методи оцінки спектральної щільності шуму, що базуються на оптимальному згладжуванні та мінімальних статистиках. Ці методи дозволяють знижувати вплив шуму на процеси розпізнавання, підвищуючи точність та якість обробки мовних сигналів, що, в свою чергу, дозволяє створювати більш стійкі системи розпізнавання в реальних умовах шуму.

Мовні сигнали є складними для аналізу через свою нестационарність, індивідуальні характеристики мовців і вплив зовнішніх шумів. Ефективне голосове керування можливе лише за умови якісної фільтрації, точної сегментації та виділення інформативних параметрів мовлення. Існуючі методи обробки часто не забезпечують необхідної точності в реальних умовах. Особливо гостро проблема стоїть у контексті української мови, для якої бракує адаптованих систем розпізнавання. Отже, необхідне подальше вдосконалення алгоритмів обробки мовних сигналів для розширення можливостей голосових технологій у різних сферах.

Бібліографія

1. Dhananjaya Gowda, Kadiri S. R., Story B., Alku P. Time-Varying Quasi-Closed-Phase Analysis for Accurate Formant Tracking in Speech Signals. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2020. Vol. 28. P. 1901–1914.
2. Yalova K., Babenko M., Yashyna K. Automatic Speech Recognition System with Dynamic Time Warping and Mel-Frequency Cepstral Coefficients. *CEUR Workshop Proceedings*. 2023. Vol. 3396. P. 141–151.
3. Osadchyy V., Skuratovskii R. V. Big data classification for the analysis mel scale features using knn parameterization. *International Journal of Circuits, Systems and Signal Processing*. 2020. Vol. 14. P. 978–989.
4. Romanyshyn Y., Hudym V. (2001). Wavelet transforms applications for speech signals processing. *The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM 2001)* : Proceedings of the 6th International Conference, 2001. 2001. P. 297–298.
5. Martin R. (). Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing*. 2001. Vol. 9, № 5. P. 504–512.