DOI: https://doi.org/10.36910/6775-2524-0560-2022-47-18
УДК 004.05
Потапова Катерина Романівна, к.т.н., доцент
https://orcid.org/0000-0002-3347-6350
Наливайчук Микола Васильович, к.т.н., ст. викладач
https://orcid.org/0000-0002-8942-9844
Климчук Ірина Олегівна, здобувач
https://orcid.org/0000-0002-5315-2564
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна.

МЕТОД РОЗПІЗНАВАННЯ ДЕФЕКТНОГО МОВЛЕННЯ НА БАЗІ ТЕХНОЛОГІЇ MEL-CEPSTRAL

Потапова К.Р., Наливайчук М.В., Климчук І.О. Метод розпізнавання дефектного мовлення на базі технології Mel-cepstral. У статті досліджується методи розпізнавання мови людей з порушенням мовного апарату по короткому словнику з використанням mel-кепстральних коефіцієнтів.
Ключові слова:.

Potapova K.R., Nalyvajchuk M.V., Klymchuk I.O. Defect speech recognition method based on Mel-cepstral розпізнавання мови, мовний сигнал, короткий словник, Mel-кепстральні коефіцієнти, програмний додаток, порушення мовного апарату **technology.** The article investigates the methods of speech recognition of people with speech disorders in a short dictionary using mel-keptral coefficients.
Key words: mobile signal, mobile signal, short vocabulary, Mel-cepstral functionality, software add-on, destruction of the mobile device.

Постановка наукової проблеми. Одною з основних форм взаємодії людей – це мовлення. У наш час достатня кількість корисних програм для розпізнавання мовлення людей, які мають певні обмеження. Ці програми перекладають текст, який був промовлений голосом у текст, для чіткого розуміння, що людина хоче розповісти. Головною метою методів по розпізнаванню мовлення є отримання інформації як вхідного голосового сигналу для подальшого чіткого перекладу. Конкретні проблеми, що вирішує метод:

1) перетворення мовного сигналу на текст;
2) голосове введення інформації;
3) пошук головних слів;
4) правильне перетворення голосового повідомлень;
5) адаптація до голосу диктора;
6) розпізнавання мови, якою говорить диктор;
7) усний переклад з однієї мови на іншу.

Аналіз досліджень. У наш час існує велика кількість програм для розпізнавання мови, які можна використовувати як для домашніх цілей так і для роботи. В сучасних методах по розпізнаванню мовлення використовується важлива частина – моделювання мови та акустичне моделювання. У наш час досить велика кількість методів для перекладу голосу в тест. Найоптимальніший метод розпізнавання побудований на базі прихованих моделях Маркова. Для виводу послідовностей символів використовують статистичні моделі. Тому метод на базі приховані моделі Маркова є одним з найоптимальніших для вирішення подібних проблем. Використання прихованих моделей Маркова приводить до більш зручного і оптимальнішого налаштування. При використанні нейронних мереж значуще менше помітних припущень щодо статистичних властивостей ознак, ніж HMM. Для точної оцінки мовних сегментів нейронні мережі потрібно проводити дискримінаційне навчання природним та ефективним способом. Доведено, що творення мова має вигляд акустичної хвилі, яку можна представити у вигляді системи органів: легені, бронхи і трахеєю, а вже потім перетворюється в голосовий тракт[1]. Структура голосової форми приймає участь у витворення звуків мови. Це можна представити у вигляді сукупностей генераторів сигналів та шумів[2]. Є багато способів для покращення результатів роботи алгоритмів. У наш час складно винайти модифікацію в алгоритмі для збільшення ефективності та покращення результату та роботи алгоритму. Але якщо припустити, що вхідні слова можна розбивати на короткі відрізки дуже короткої тривалості і при цьому обчислювати коефіцієнти для кожного відрізку можна на декілька відсотків покращити

результат роботи методу [3].

Основна частина. Доведено, що мова - це звична форма спілкування людей за допомогою мовних конструкцій та певних правил [4]. Допустимо, що звичне середовище для того щоб передавати інформацію – це повітря, тоді при спілкуванні звукове коливання можна охарактеризувати звуковою частотою і амплітудою. Усім відомо, що мова – це носій інформації, що використовується людиною для передачі повідомлень - сигналом. Мова - це акустичний сигнал, що безперервно змінюється в часі.

Майже всі сигнали, що існують мають схожість походження. Тому при обробці сигналів потрібно перетворювати в дискретні сигнали за допомогою аналого-цифрового перетворення (АЦП). У наш час велика кількість різних обчислювальних потужностей, але у поєднанні з розпізнаванням мови є і залишається досить складною проблемою. Висока складність обчислювального алгоритмів у нашому житті додає деякі дуже важливі факти для автоматичного розпізнавання мови, а саме на обсяг оброблюваного словника, швидкість отримання відповіді і точність. Описані вище проблеми можна об'єднати і на виході отримати надійний, самостійний та максимально швидкий пристрій. Питання пошуку нових архітектурних рішень по розпізнаванню мови набуває все більш актуального значення. Одним з перспективних та нових напрямків є дослідження і використання методу Mel-кепстральних коефіцієнтів по короткому словнику.

У системі автоматичного розпізнавання мови є такі фази: виділення ознак, навчання і розпізнавання (рис. 1). Після виділення ознак, ми отримуємо вектор ознак, в якому стислий опис сигналу з корисною інформацією для подальшого розпізнавання. Для того, щоб отримати результат прийнято використовувати методи, які можуть працювати в частотній області та в тимчасовій, при цьому проблема подання мови не вирішена до кінця і дослідження ведуться до теперішнього часу [5].

Векторні ознаки формуються у деяку послідовність довжиною T, яку називають акустичною. Також можемо побачити, що формується деяка послідовність $O=(o_1, o_2, ..., o_T)$. За допомогою цієї послідовності можна передати точну послідовність слів $W=(w_1, w_2, ..., w_N)$. Сама задача розпізнавання мови така: отримання послідовності слів W, який аналогічно відповідає деякій акустичній послідовності O [6].

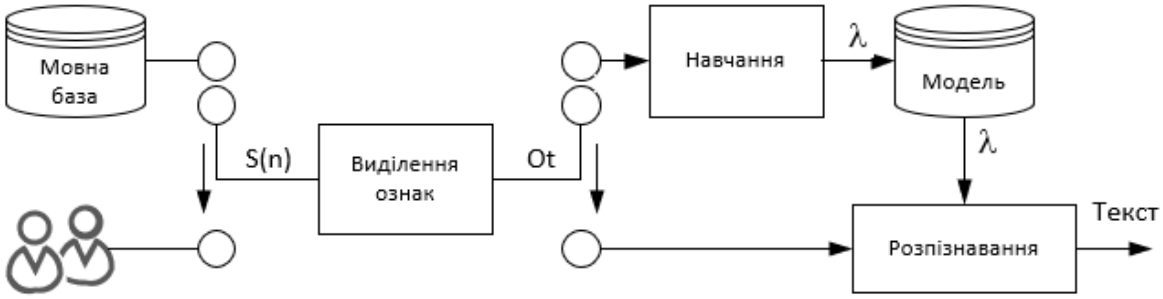

Рисунок 1 – Загальна схема САРМ

Потрібно перевірити всі ланцюжки слів W, але на практиці це майже неймовірно. Для того, щоб полегшення це завдання вводять різні обмеження, наприклад, розпізнавання тільки виділених слів.

Викладення основного матеріалу. У порівнянні з тимчасовою областю частотна область найкраще представляє мову ознаками. Для частотної області ця лінійна система і є головним завданням для знаходження ознак. Метод лінійного передбачення представляє собою можливість наближеність поточного мовного сигналу за допомогою лінійної комбінації відліків, які були використані до цього [7]:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k), \quad (1)$$

де $\{a_k\}$ – коефіцієнти лінійного передбачення.

Методом MFCC можна виділити сигнал. Для цього використовують дискретне перетворення Фур'є. При використанні дискретного перетворення в частотній області можна обчислити логарифм спектру сигналу на вході. Після чого можна обчислити косинусне перетворення. Основною перевагою MFCC перед LPC і PLP з подібною якістю розпізнавання є простота впровадження. Цей метод є більш швидким, за рахунок ефективної процедури знаходження ДПФ і ОДПФ – а саме швидкого перетворення Фур'є (ШПФ). Аналізуючи методи було виявлено, що MFCC застосовується найбільш широко.

Якщо на вхідний сигнал, який ми отримуємо на виході застосувати метод Фур’є, ми отримаємо такий графік (Рис. 2).

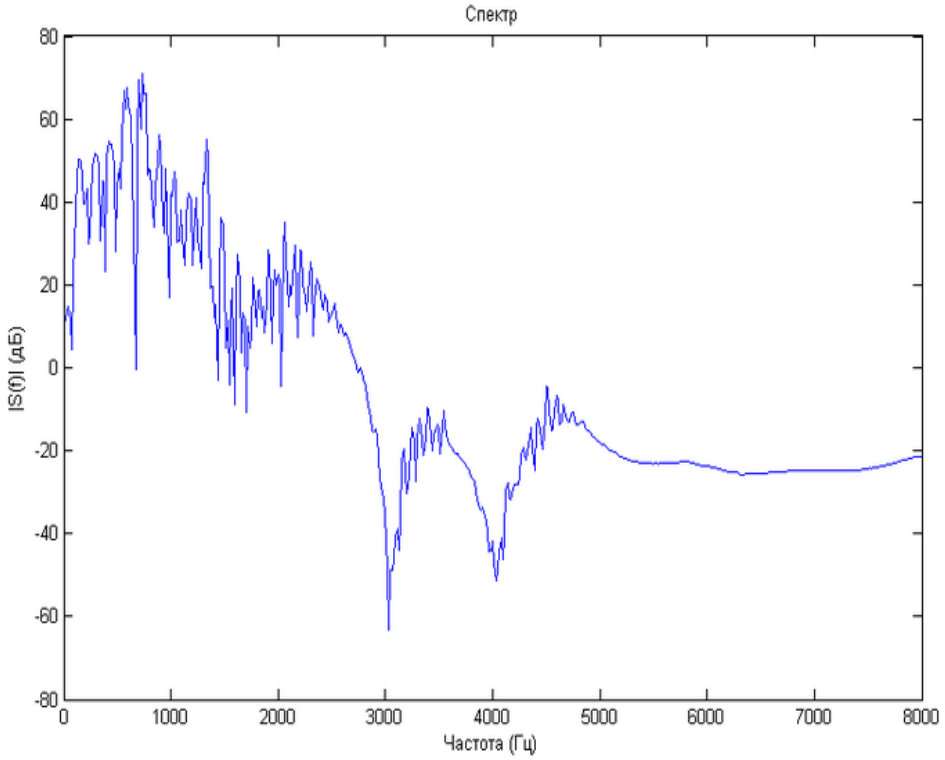

Рисунок 2 – Перетворення вхідного сигналу

Після того, як ми отримали перетворення можемо продемонструвати вид на Mel-осі (Рис. 3).

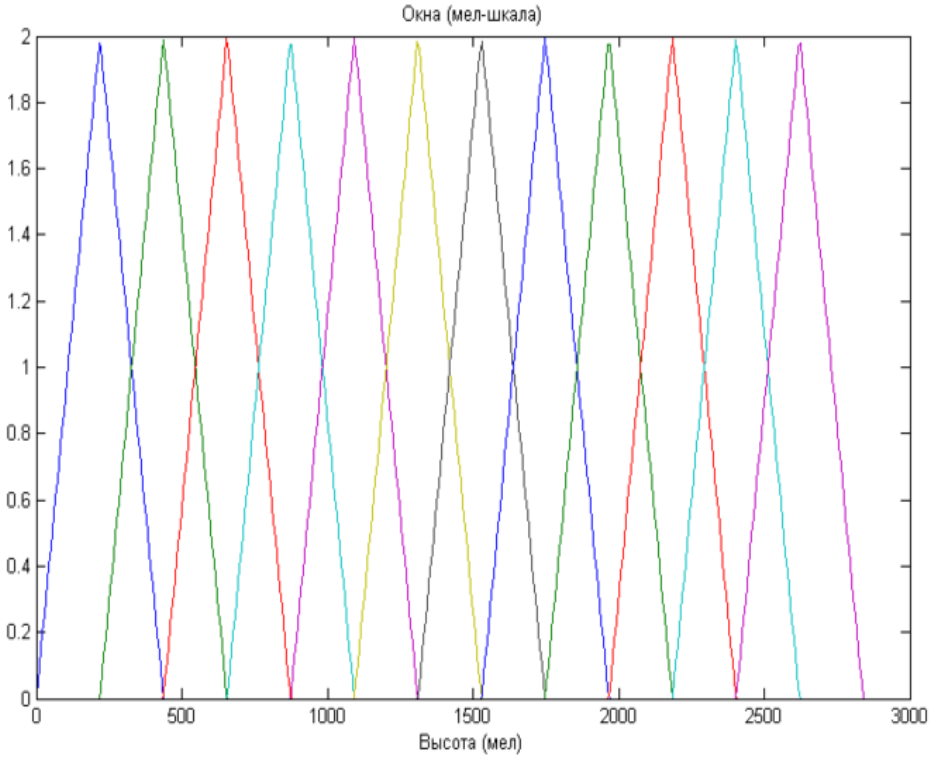

Рисунок 3 – Накладення вхідного сигналу на Mel-осі

Проведемо детальний опис послідовностей знаходження MFCC. Кожний етап розглянуто на малюнку з додатковим описуванням кожного процесу (Рис. 4).

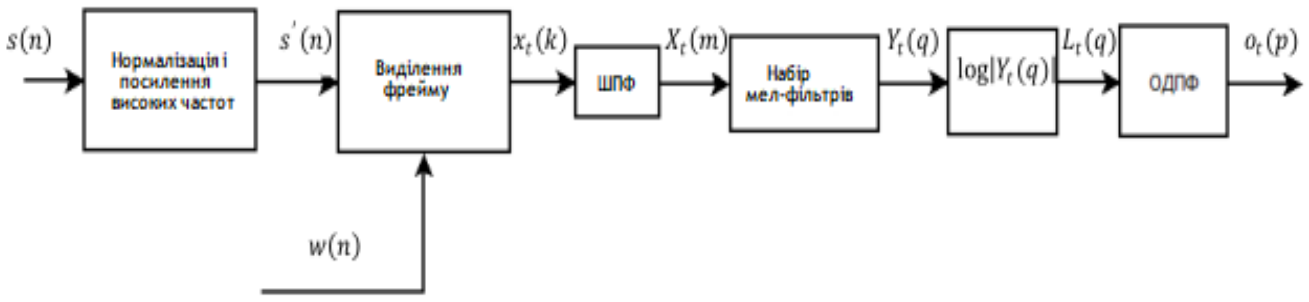

Рисунок 4 – Знаходження Mel-кепстральних коефіцієнтів

1. Нормалізація початкового сигналу може вирівняти сигналу і збільшення частотних висот. Низькочастотні форманти мають більші амплітуди, ніж високочастотні форманти, хоча останні також несуть важливу ідентифікаційну інформацію. Тому до вхідного сигналу застосовують фільтр:

$$s'(n) = s(n) - 0.95 \cdot s(n-1), \quad (2)$$

2. Сигнал виділяє короткочасну ділянку і накладає віконну функції $w(k)$ для мінімізації витоку спектра. Результатом є фрейм, довжина якого K відліків. Надалі всі процедури ведуться в межах саме цього фрейму:

$$x_t(k) = s'(k + t \cdot K) \cdot w(k), 0 \le k \le K - 1, \quad (3)$$

Таким чином, вектор ознак отримують для кожного $0 \le t \le T$ фреймам вхідного сигналу $s'(n)$. В якості віконної функції зазвичай використовують функцію Хемінга:

$$w(k) = 0.54 - 0.46 \cdot \cos(\frac{2\pi k}{K-1}), \quad (4)$$

3. Обчислення ДПФ для фрейм:

$$X_t(m) = \sum_{k=0}^{K-1} x_t(k) \cdot e^{\frac{-j2\pi mk}{K}} \quad , \quad (5)$$

ДПФ розраховується за допомогою алгоритму швидкого перетворення Фур'є (FFT).

4. Накладення в частотній області послідовностей Q Мел-фільтрів на фрейм. В результаті процедури фільтра q на основі триманих результатів знімають енергію $Yt(q)$. Таким чином моделюють сприйняття мови людиною: роздільна здатність слуху зростає при русі по спектру від низьких частот до високих. Центральні частоти Fq Мел-фільтрів вибираються за так званою Мел-шкалою, яка залежить від звичайної по логарифмічному закону (рис. 5):

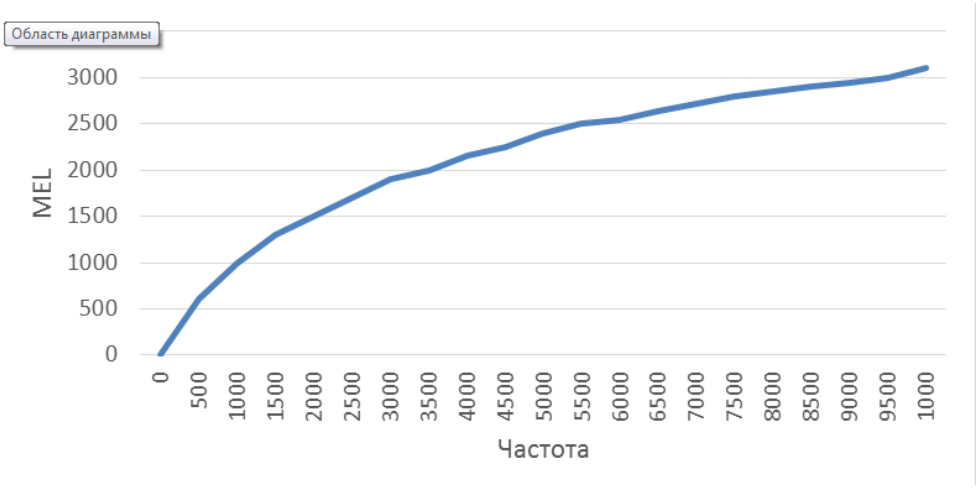

Рисунок 5 – Відповідність частот звичайної шкали частотам Мел-шкали

$$F_{mel}2095 \cdot log_{10}(1 + \frac{F_{Hz}}{700}), \quad (6)$$

5. Логарифмування $Yt(q)$. На цьому кроці виконується гомоморфні перетворення. Наведена нижче формула дає нам змогу відділити сигнал і фільтри $Yt(q)$:

$$L_t(q) = \log |Y_t(q)| \, , \quad (7)$$

6. Обчислення ОДПФ. На останньому етапі отримують Мел-кепстральні коефіцієнти шляхом обчислення ОДПФ для $Lt(q)$. При цьому, так як $Lt(q)$ дійсний і симетричний, ОДПФ буде еквівалентно дискретному косинусному перетворенню:

$$Q_t(p) = \sum_{q=1}^{Q} L_t(q) \cdot \cos\left(p \cdot \left(q - \frac{1}{2}\right) \cdot \frac{\pi}{Q}\right), p = 0, \dots, P \, , \quad (8)$$

В результаті виходить коефіцієнт P, хоча P може дорівнювати Q, але зазвичай береться лише половина значення MFCC. Це пояснюється тим, що кепстра сигналу збудження зазвичай знаходиться «праворуч» від кепстри мовного каналу.

Отже, розпізнавач на виході отримує p-вимірний вектор ot, що містить Мел-кепстральні коефіцієнти для t фрейму [8].

Висновки та перспективи подальшого дослідження. У даній роботі детально розглянуто метод використання Mel-кепстральних коефіцієнтів з точки зору організації звукового інтерфейсу для людей із дефектами мовлення.

Основні результати роботи полягають в наступному:

1. Проведено дослідження поставленої задачі: спосіб розпізнавання мови з виділенням основних компонентів для систем автоматичного розпізнавання мови.

2. Розглянуто існуючі методи обробки і виділення головних ознак сигналу мовлення, серед яких було обрано підхід, який найкраще підходить під поставлене завдання.

Список бібліографічного опису

1. The 11th International scientific and practical conference "European scientific discussions" (September 12-14, 2021) Potere della ragione Editore, Rome, Italy, 2021, 337 p., UDC 001.1, ISBN 978-88-32934-02-1, P. 64-68 URL: https://sci-conf.com.ua/xi-mezhdunarodnaya-nauchno-prakticheskaya-konferentsiya-european-scientific-discussions-12-14-sentyabrya-2021-goda-rim-italiya-arhiv/
2. S. Davis and P. Mermelstein Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, 1980.
3. UDC 001.1 The 7 th International scientific and practical conference "Results of modern scientific research and development" (September 19-21, 2021) Barca Academy Publishing, Madrid, Spain. 2021. 336 p. ISBN 978-84-15927-33-4, P. 105-108 URL: https://sci-conf.com.ua/vii-mezhdunarodnaya-nauchno-prakticheskaya-konferentsiya-results-of-modern-scientific-research-and-development-19-21-sentyabrya-2021-goda-madrid-ispaniya-arhiv/
4. Вінцюк Т.К., Аналіз, розпізнавання і інтерпретація мовних сигналів, - Київ: Наукова думка, 1987. - 264 с.
5. I.V. Ognev, A.I. Ognev, P.A. Paramonov, N.A. Sutula, The use of extrema distribution as a feature vector for speech patterns recognition // The 11th International Conference "Pattern Recognition and Image Analysis: New Information Technologies", Vol. 1, 2013. – pp. 114-117.
6. Claudio Becchetti, Lucio Prina Ricotti, Speech Recognition. Theory and C++ Implementation – Wiley. – 1999, 428 p.
7. Л. Рабинер, Р. Шафер, Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – 496 с.
8. Т.В. Шарий, О проблеме параметризации речевого сигнала в современных системах распознавания речи // Вісник Донецького Національного Університету, 2008, вип. 2, 536-541 с.
9.

References

1. The 11th International scientific and practical conference "European scientific discussions" (September 12-14, 2021) Potere della ragione Editore, Rome, Italy, 2021, 337 p., UDC 001.1, ISBN 978-88-32934-02-1 , P. 64-68 URL: https://sci-conf.com.ua/xi-mezhdunarodnaya-nauchno-prakticheskaya-konferentsiya-european-scientific-discussions-12-14-sentyabrya-2021-goda-rim-italiya-arhiv/
2. S. Davis and P. Murmelstein Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, 1980.
3. UDC 001.1 The 7th International scientific and practical conference "Results of modern scientific research and development" (September 19-21, 2021) Barca Academy Publishing, Madrid, Spain. 2021. 336 p. ISBN 978-84-15927-33-4, P. 105-108 URL: https://sci-conf.com.ua/vii-mezhdunarodnaya-nauchno-prakticheskaya-konferentsiya-results-of-modern-scientific-research-and-development-19-21-sentyabrya-2021-goda-madrid-ispaniya-arhiv/
4. Vintsyuk TK, Analysis, recognition and interpretation of speech signals, - Kyiv: Naukova Dumka, 1987. - 264 p.
5. I.V. Ognev, AI Ognev, P.A. Paramonov, NA Sutula, The use of extrema distribution as a feature vector for speech patterns recognition // The 11th International Conference "Pattern Recognition and Image Analysis: New Information Technologies", Vol. 1, 2013. - pp. 114-117.
6. Claudio Becchetti, Lucio Prina Ricotti, Speech Recognition. Theory and C ++ Implementation - Wiley. - 1999, 428 p.
7. L. Rabiner, R. Schaefer, Digital processing of speech signals. - M .: Radio and communication, 1981. - 496 p.
8. T.B. Sharii, On the problem of speech signal parameterization in modern speech recognition systems // Visnyk of Donetsk National University, 2008, vol. 2, 536-541 c.