Технічні науки

УДК 004.05

Климчук Ірина Олегівна

магістрант

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Климчук Ирина Олеговна

магистрант

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Klymchuk Iryna

Master of the

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

Потапова Катерина Романівна

кандидат технічних наук, доцент кафедри

системного програмування і спеціалізованих комп'ютерних систем

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Потапова Екатерина Романовна

кандидат технических наук, доцент кафедры

системного программирования и специализированных компьютерных систем

Национальный технический университет Украины

«Киевский политехнический институт имени Игоря Сикорского»

Potapova Kateryna

Candidate of Technical Sciences, Associate Professor of the

Systems Programming and Specialized Computer Systems Department

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

Івасенко Дмитро Віталійович
магістрант
Національної академії внутрішніх справ
Ивасенко Дмитрий Витальевич
магистрант
Национальной академии внутренних дел
Ivasenko Dmytro
Master of the
National Academy of Internal Affairs

ОСОБЛИВОСТІ ВИКОРИСТАННЯ MEL-КЕПСТРАЛЬНИХ КОЕФІЦІЄНТІВ ПРИ РОЗПІЗНАВАННІ МОВЛЕННЯ
ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ MEL-КЕПСТРАЛЬНЫХ КОЭФФИЦИЕНТОВ ПРИ РАСПОЗНАВАНИИ РЕЧИ
PECULIARITIES OF USING MEL-CEPSTRAL COEFFICIENTS IN SPEECH RECOGNITION

Анотація. *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. 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. У*

статті досліджується методи розпізнавання мови людей з порушенням мовного апарату по короткому словнику з використанням mel-кепстральних коефіцієнтів. У запропонованому методі застосовуються критерій, який використовується для неупередженої оцінки логарифмічного спектру, до спектральної моделі, представленої коефіцієнтами MEL. Для розв'язання задачі нелінійної мінімізації, задіяної в методі, вони дають ітераційний алгоритм, збіжність якого гарантована. Приклади аналізу мовлення та результати ізольованого експерименту з розпізнавання слів.

Ключові слова: *розпізнавання мови, мовний сигнал, короткий словник, Mel-кепстральні коефіцієнти програмний додаток, порушення мовного апарату.*

Аннотация. *Одной из основных форм взаимодействия людей – это речь. В настоящее время достаточное количество полезных программ для распознавания речи людей, имеющих определенные ограничения. Эти программы переводят текст, произнесенный голосом в текст, для четкого понимания, что человек хочет рассказать. Главной целью методов по распознаванию речи является получение информации как входящего голосового сигнала для последующего четкого перевода. В современных методах распознавания речи используется важная часть – моделирование языка и акустическое моделирование. В настоящее время достаточно большое количество методов для перевода голоса в тест. Самый оптимальный метод распознавания построен на базе скрытых моделей Маркова. Для вывода последовательностей символов используются статистические модели. Поэтому метод на базе скрытых моделей Маркова является одним из самых оптимальных для решения подобных проблем. В статье исследуются методы распознавания речи людей с нарушением речевого аппарата по короткому словарю с использованием Mel-кепстральных коэффициентов. В предложенном методе применяются*

критерий, используемый для беспристрастной оценки логарифмического спектра к спектральной модели, представленной коэффициентами MEL. Для решения задачи нелинейной минимизации, задействованной в методе, они дают итерационный алгоритм, сходимость которого гарантирована. Примеры анализа речи и результаты изолированного эксперимента по распознаванию слов.

Ключевые слова: *распознавание языка, речевой сигнал, краткий словарь, Mel-кепстральные коэффициенты программное приложение, нарушение речевого аппарата.*

Summary. *One of the main forms of human interaction is speech. Nowadays, there are enough useful programs for speech recognition of people who have certain limitations. These programs translate the text that was spoken aloud into the text, to clearly understand what the person wants to say. The main purpose of speech recognition methods is to obtain information as an input voice signal for further clear translation. Modern methods of speech recognition use an important part - language modeling and acoustic modeling. Nowadays, there are quite a number of methods for translating a voice into a test. The best method of recognition is bas on hidden Markov models. Statistical models are using to display sequences of characters. Therefore, the method based on hidden Markov models is one of the best for solving such problems. The article investigates the methods of speech recognition of people with speech disorders in a short dictionary using mel-keptral coefficients. The proposed method applies the criterion used for unbiased estimation of the logarithmic spectrum to the spectral model represented by MEL coefficients. To solve the nonlinear minimization problem involved in the method, they provide an iterative algorithm whose convergence is guaranteed. Examples of speech analysis and results of an isolated word recognition experiment.*

Key words: *speech recognition, speech signal, short dictionary, Mel-cepstral coefficients software application, speech apparatus disorders.*

Постановка проблеми. Розпізнавання мовлення в текст — це здатність програми визначати слова, вимовлені вголос, і перетворювати їх у текст, який можна читати. Початкове програмне забезпечення для розпізнавання мовлення має обмежений словниковий запас і може розпізнавати слова та фрази лише тоді, коли вони промовляються чітко. Більш складне програмне забезпечення може обробляти природну мову, різні акценти та різні мови. Програмне забезпечення для розпізнавання мовлення може перекладати вимовлені слова в текст, використовуючи закриті субтитри, щоб людина з втратою слуху могла зрозуміти, що говорять інші. Розпізнавання мовлення також може дозволити людям з обмеженим використанням рук працювати з комп'ютером, використовуючи голосові команди замість введення тексту.

Аналіз останніх досліджень і публікацій. Mel-кепстральні коефіцієнти (MFCC) були запропоновані і введені S. Davis і P. Mermelstein [1]. Декілька параметричних уявлень акустичного сигналу порівнювалися з точки зору ефективності розпізнавання слів у системі розпізнавання безперервної мови, орієнтованої на склади [2]. Встановлено перевагу алгоритму симетричної форми. Успішно впроваджено новий метод, званий обмеженням нахилу, у якому нахил функції деформації обмежується, щоб покращити розрізнення слів у різних категоріях [3]. Декілька параметричних уявлень акустичного сигналу порівнювалися з точки зору ефективності розпізнавання слів у системі розпізнавання безперервної мови, орієнтованої на склади [4].

Метою роботи є аналіз та розкриття підходів існуючих алгоритмів та методів розпізнавання мовлення людей з дефектами мовного апарату.

Виклад основного матеріалу. Розпізнавання мовлення використовує широкий спектр досліджень з інформатики, лінгвістики та комп'ютерної інженерії. Багато сучасних пристроїв і програм, орієнтованих на текст,

мають функції розпізнавання мовлення, які дозволяють простіше використовувати пристрій або використовувати його без використання рук.

Розпізнавання мовлення використовується для визначення слів у розмовній мові.

Системи розпізнавання мовлення використовують комп'ютерні алгоритми для обробки та інтерпретації вимовлених слів і перетворення їх у текст. Програмна програма перетворює звук, записаний мікрофоном, на письмову мову, яку можуть зрозуміти комп'ютери та люди, дотримуючись цих чотирьох кроків:

1) проаналізувати аудіо;
2) розбити його на частини;
3) оцифрувати його у формат, читабельний комп'ютером; і
4) використовуйте алгоритм, щоб узгодити його з найбільш підходящим представленням тексту.

Нині системи розпізнавання мовлення використовуються в різних середовищах. За останні роки було розроблено багато систем розпізнавання мовлення для вирішення різноманітних проблем у реальному світі.

Найчастіше розпізнавання мовлення використовується для:

– Під час навчання мови використовується програмне забезпечення для розпізнавання мовлення. Програмне забезпечення чує мову користувача та пропонує допомогу з вимовою.
– Автоматичні голосові помічники слухають запити клієнтів і надають корисні ресурси.
– Лікарі можуть використовувати програмне забезпечення для розпізнавання мовлення, щоб транскрибувати нотатки в режимі реального часу в медичні записи.
– Програмне забезпечення можна використовувати для стенограми судових засідань, виключаючи потребу в транскрибаторах.

– Смартфони використовують голосові команди для маршрутизації викликів, обробки мовлення в текст, голосового набору та голосового пошуку. Користувачі можуть відповідати на текст, не дивлячись на свої пристрої. Розпізнавання мовлення також можна знайти в програмах обробки текстів, таких як Microsoft Word, де користувачі можуть диктувати слова, які потрібно перетворити на текст.

Одним з популярних методів виділення звукових характеристик є метод розпізнавання мовлення з використанням Mel-кепстральними коефіцієнтами.

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].

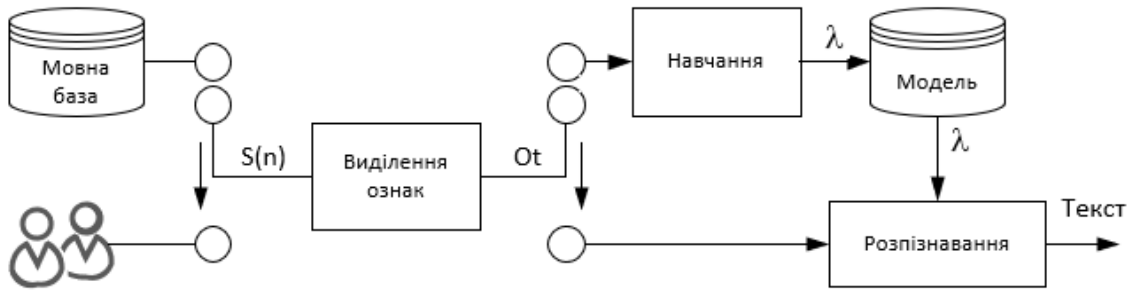

Рис. 1. Загальна схема САРМ

Векторні ознаки формуються у деяку послідовність довжиною T, яку називають акустичною. Також можемо побачити, що формується деяка послідовність O=(o1, o2, ..., oT). За допомогою цієї послідовності можна передати точну послідовність слів W=(w1, w2, ..., wN). Сама задача розпізнавання мови така: отримання послідовності слів W, який аналогічно відповідає деякій акустичній послідовності O [6].

У статті розглянемо нову систему розпізнавання мовлення з використанням покращеного кепстрального коефіцієнта частоти Mel з методом вікон та кадрування. Метод вікон і кадрування використовується для видалення білого шуму Гауса у вхідному мовному сигналі. Блок видалення шумів ефективно використовує алгоритм невід'ємної матриці факторів для спектрів Mel -величини шумного вхідного звукового сигналу. Крім того, кепстральні коефіцієнти Mel -частот (MFCC) використовуються для знаходження більш важливих ознак, які існують у мовному сигналі.

При обробці звуку кепстральні частоти Mel (MFC) є представленням короткочасного спектру потужності звуку, заснованого на лінійному косинусному перетворенні логарифмічного спектру потужності на нелінійній шкалі частоти Mel.

Mel -частотні кепстральні коефіцієнти (MFCC) — це коефіцієнти, які разом складають MFC [7]. Вони є похідними від типу кепстрального представлення аудіо кліпу (нелінійний «спектр спектру»). Різниця між кепстром і кепстром Mel -частоти полягає в тому, що в MFC смуги частот рівномірно розташовані за шкалою Mel, що наближає реакцію слухової системи людини більш точно, ніж лінійно-рознесені смуги частот, які використовуються в нормальному спектрі. Таке викривлення частоти може дозволити краще відображати звук, наприклад, під час стиснення звуку.

Нижче наведено процес вилучення функцій MFCC. MFCC зазвичай отримують наступним чином:

1. Обирають перетворення Фур'є (витримку з вікна) сигналу.
2. Зображення потужності спектра, отриманого вище, на шкалу Mel, використовуючи трикутні вікна, що перекриваються, або, як альтернатива, вікна, що перекриваються косинусами.
3. Вибір журнали потужностей на кожній з частот Mel.
4. Дискретне косинусне перетворення списку потужностей Mel log, як ніби це сигнал.

5. MFCC – це амплітуди результуючого спектру.

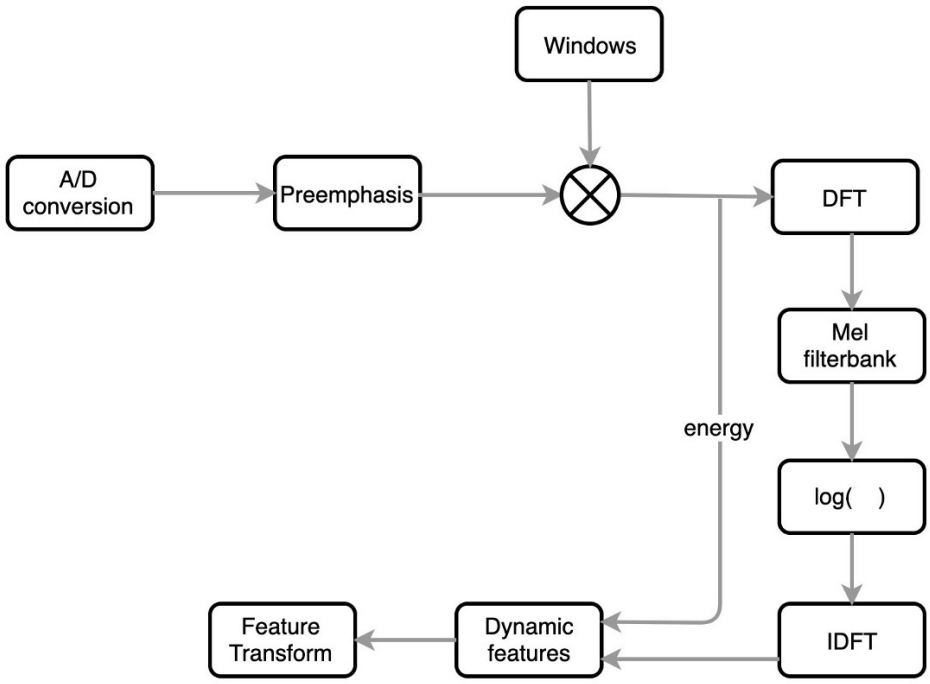

Рис. 2. Процес вилучення функцій MFCC

Використовуючи MFCC можна краще представити людський голос, зазвичай LPCC використовується в цифровому спілкуванні, тому основна мета цієї техніки не представляє голос, а полягає в стиснення та передачі інформації, яка містить голос. Оскільки MFCC використовує шкалу Mel, наближення до поведінки людського голосу є хорошим, він краще представляє голос.

Подібні одиниці виміру часто використовують при вирішенні задач розпізнавання, так як вони дозволяють наблизитися до механізмів людського сприйняття, яке поки що лідирує серед відомих систем розпізнавання мови.

При збереженні промови на амплітуду звукового сигналу впливає низка чинників: гучність голосу диктора, його віддаленість від мікрофона тощо. Всі ці перелічені фактори призводять до великої варіативності гучності мовного сигналу [8]. Особливо сильно це явище помітне при використанні різнорідної звукозаписної апаратури.

Висновки. Таким чином, в даній статті дано опис методу розпізнавання мови людей з порушенням мовного апарату з використанням Mel-кепстральних коефіцієнтів. Проаналізувавши всю інформацію можна зробити висновок, що запропонований метод має ряд значних переваг, тому для розпізнавання мови людей з порушенням мовного апарату рекомендовано використовувати саме його.

Література

1. Davis S., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Processing. 1980. № 28. P. 357-366.
2. IEEE Transactions on Acoustics, Speech, and Signal Processing. July, 1989. Vol. 37. Iss. 7.
3. IEEE Transactions on Acoustics, Speech, and Signal Processing. Февраль 1978. Т. 26. Вып.1.
4. IEEE Transactions on Acoustics, Speech, and Signal Processing. Август 1980. Т. 28. Вып. 4.
5. Ognev I.V., Ognev A.I., Paramonov P.A., Sutula N.A. 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". 2013. Vol. 1. P. 114-117.
6. Claudio Becchetti, Lucio Prina Ricotti, Speech Recognition. Theory and C++ Implementation – Wiley. 1999. 428 p.
7. Min Xu; et al. (2004). "HMM-based audio keyword generation" (PDF). In Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh (eds.). Advances in Multimedia Information Processing – PCM 2004: 5th Pacific Rim Conference on Multimedia. Springer. ISBN 978-3-540-23985-7. Archived from the original (PDF) on 2007-05-10.

8. Огнев И. В., Парамонов П.А. Предварительная обработка речевого сигнала для построения базы произношений одиночных слов // Информационные средства и технологии: труды Международной научно-технической конференции (20 – 22 октября 2012 г.): в 3 т. М.: МЭИ, 2012. 1 т. С. 53-58.

References

1. Davis S., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Processing. 1980. № 28. P. 357-366.
2. IEEE Transactions on Acoustics, Speech, and Signal Processing. July, 1989. Vol. 37. Iss. 7.
3. IEEE Transactions on Acoustics, Speech, and Signal Processing. February, 1978. Vol. 26. Iss. 1.
4. IEEE Transactions on Acoustics, Speech, and Signal Processing. August, 1980. Vol. 28. Iss. 4.
5. Ognev I.V., Ognev A.I., Paramonov P.A., Sutula N.A. 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". 2013. Vol. 1. P. 114-117.
6. Claudio Becchetti, Lucio Prina Ricotti, Speech Recognition. Theory and C++ Implementation – Wiley. 1999. 428 p.
7. Min Xu; et al. (2004). "HMM-based audio keyword generation" (PDF). In Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh (eds.). Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia. Springer. ISBN 978-3-540-23985-7. Archived from the original (PDF) on 2007-05-10.
8. Ognev I.V., Paramonov P.A. Preliminary processing of a speech signal to build a database of pronunciations of single words // Information tools and

technologies: Proceedings of the International Scientific and Technical Conference (October 20 - 22, 2012): in 3 vols. M .: MPEI, 2012. 1 Vol. P. 53-58.