

АЛГОРИТМ АВТОМАТИЗОВАНОЇ ВІЗУАЛІЗАЦІЇ МОВЛЕННЄВИХ СИГНАЛІВ ДЛЯ ЗАДАЧІ АВТЕНТИФІКАЦІЇ МОВЦЯ

Вінницький національний технічний університет;

Анотація

У тезах запропоновано процедуру попереднього оброблення мовленнєвого сигналу і алгоритм виділення з нього інформативних ознак для подальшого застосування у автоматизованій системі розпізнавання мовця. Наведено результати його тестування.

Ключові слова: автоматизована система розпізнавання мовця, інформативні ознаки, візуалізація

Abstract

The thesis proposes a procedure for preliminary processing of the speech signal and an algorithm for allocating informative features from it for further application in the automated speaker recognition system. The results of its testing are given.

Keywords: automated speaker recognition system, informative features, visualization.

Вступ

Мовленнєвий сигнал передає інформацію, пов'язану з фізіологічними властивостями мовного апарату мовця [1], у ньому відображається розмір і форма мовного тракту, ротової і носової порожнин тощо. Він також несе інформацію щодо поведінкових аспектів мовця, інформацію про акцент, мимовольні емоційні перетворення акустичних параметрів тощо. Процес автоматичної верифікації мовця – це процес автентифікації особи за її голосовими зразками [2]. Для верифікації мовця акустичний аналоговий мовний сигнал перетворюється у цифрову форму, тобто у послідовність чисел, для чого найчастіше використовують такі методи цифрової обробки мовних сигналів як вейвлет-аналіз, частотний Фур'є-аналіз, кепстральний аналіз.

Метою роботи є аналіз процесу виділення індивідуальних мовленнєвих ознак для автоматизованої комп'ютерно-інтегрованої системи верифікації мовця.

Результати дослідження

Вхідний мовний сигнал подається системі у вигляді WAV-файлу. На етапі попереднього оброблення мовний сигнал проходить такі послідовні стадії: розбиття сигналу на блоки; цифрова фільтрація; оброблення сигналу у вікні Хеммінга; спектральне перетворення; нормалізація частотного спектру. Відповідно, розбиваємо оцифрований мовний сигнал на блоки тривалістю 20 мс кожний. До сигналу у блоках застосовуємо високочастотне підсилення для компенсації спадання рівня гучності мовного сигналу, викликане розсіюванням від губ мовця. З цією метою блоки сигналу пропускають крізь фільтр першого порядку:

$$x(j) = y(j) - y(j-1) \% \text{Фільтр першого порядку};$$

Оброблення сигналу у блоках проводиться для зниження крайових ефектів, що з'являються внаслідок сегментації. Для пригнічення небажаних крайових ефектів необхідно помножити мовний сигнал на віконну функцію, в нашому випадку - вікно Хеммінга:

$$z(j) = (0.54 - 0.46 * \cos(2 * \pi * (j-1) / 408)) * x(j); \% \text{Вікно Хеммінга}$$

Для отримання спектральної оцінки мовного сигналу використовуються дискретне перетворення Фур'є. Для цього необхідно збільшити величину блоку до 512 відліків, за рахунок додаванням необхідною кількістю нулів. Після цього застосовуємо швидке перетворення Фур'є довжиною 512 точок і в результаті отримуємо 512 спектральних комплексних значень, які є попарно спряженими. Нам необхідні тільки 256 комплексних значень із 512 з перетворення. Для цих значень знаходимо їх

амплітуди для чого нормалізуємо частотний спектр сигналу - знаходимо найбільшу тривалість вектора і значення всіх векторів множимо на величину, обернену до цієї довжини. На рис. 1 зображено ілюстрацію етапів попереднього оброблення мовного сигналу.

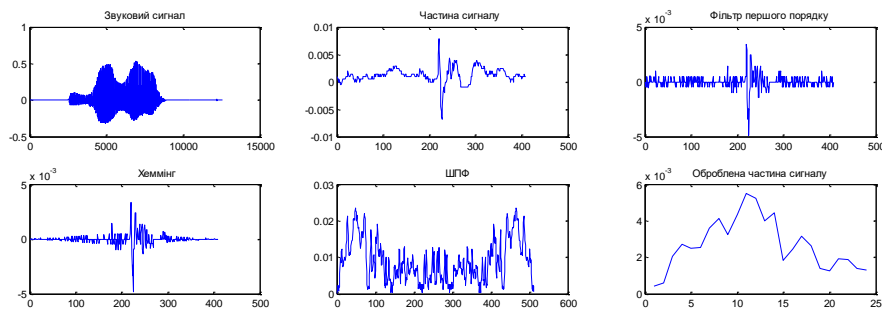


Рис. 1. Етапи попереднього оброблення мовного сигналу

В якості інформативної ознаки для верифікації мовця використаємо мел-частотні кепстральні коефіцієнти (Mel-Frequency Cepstral Coefficients, MFCC) [3]. Такі коефіцієнти базуються на ключових поняттях кепстру і мел-шкали. Кепстр – це результат дискретного косинусного перетворення логарифма амплітудного спектру сигналу. Мел-шкала описує частотну чутливість людського слуху – враховується, що зміна частоти в два рази в діапазоні низьких та високих частот людина сприймає по-різному. Мел-частотні кепстральні коефіцієнти (МЧКК) – це розподілені по мел-шкалі значення кепстру із використанням банку фільтрів.

Алгоритм знаходження МЧКК такий:

- 1) Попередньо оброблений мовний сигнал розбивається на певну кількість відрізків тривалістю 20 мс;
- 2) До кожного відрізка застосовується дискретне перетворення Фур'є;
- 3) Знаходиться спектральна потужність щільності отриманого у п.2 сигналу;
- 4) Застосовуючи банк фільтрів отриманий у п.3 спектр розподіляється за мел-шкалою (кількість каналів – 20, кількість кепстральних коефіцієнтів – 12, частотний діапазон – 300-3700 Гц);
- 5) Логарифмуємо результат п.4:

$$X_n[i] = \ln(X_n[i]), \quad i = 1, \dots, P; \quad (1)$$

- 6) Здійснюємо дискретне косинусне перетворення результату п.5.

$$C_n[j] = \sum_{k=1}^P X_n[k] \cos\left(j\left(k - \frac{1}{2}\right)\frac{\pi}{P}\right), \quad i = 1, \dots, P, \quad j = 1, \dots, J, \quad (2)$$

де $C_n[j]$ – масив кепстральних коефіцієнтів, k – кількість відрізків; P – кількість фільтрів; J – бажане число коефіцієнтів, $J < P$.

Для прикладу отриману матрицю кепстральних коефіцієнтів можна зобразити у вигляді бітової карти, зображену на рис. 2.

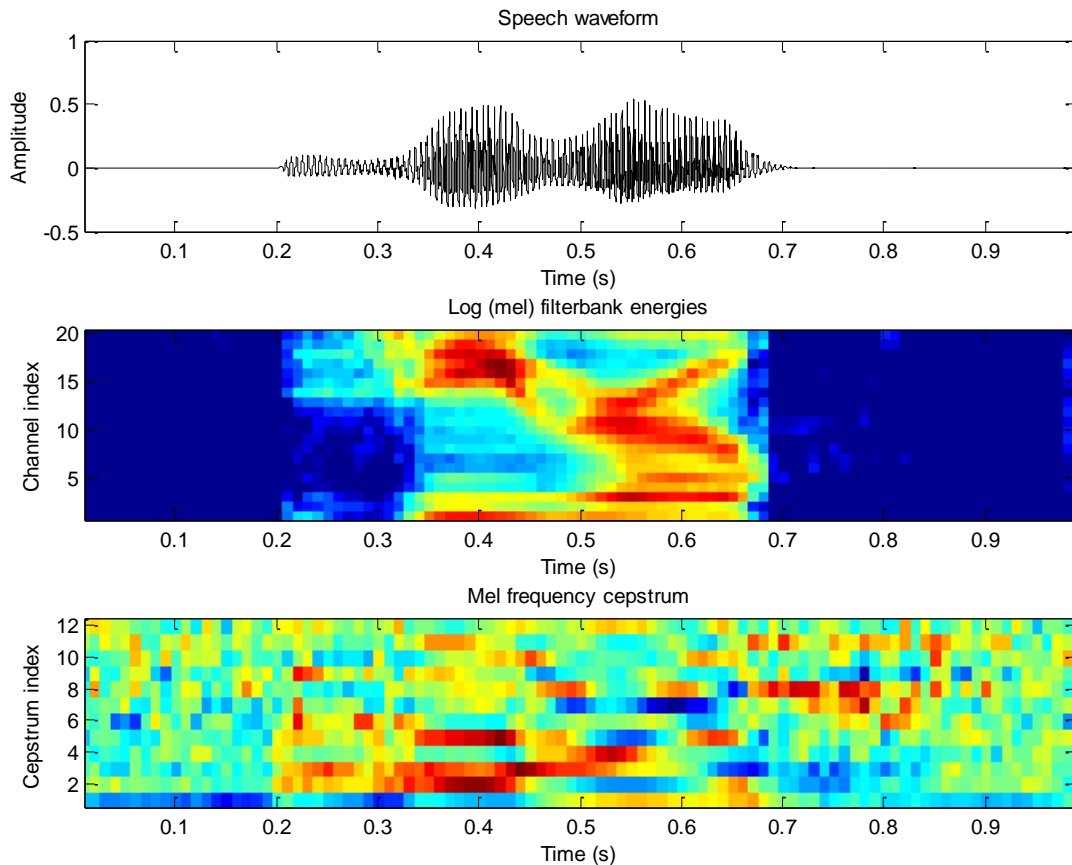


Рис.2. Виділення кепстральних коефіцієнтів з мовного сигналу

Висновки

Розроблено алгоритм для виділення інформативних ознак для верифікації мовців автоматизованою комп'ютерно-інтегрованою системою верифікації мовця на основі кепстральних коефіцієнтів. Також описано процедуру попереднього оброблювання мовного сигналу для подальшого аналізу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ковтун В.В. Оптимізація алфавіту інформативних ознак для автоматизованої системи розпізнавання мовців критичного застосування. / А.О. Береза, М.М. Биков, А.Д. Гафурова, В.В. Ковтун // Вісник Хмельницького національного університету, серія: Технічні науки, Хмельницький. – 2017. - №3(249). – 222-228 с.
2. Ковтун В.В. Метод представлення ознак у автоматизованій системі розпізнавання мовця критичного застосування. / М.М. Биков, В.В. Ковтун, М.С. Фурман // Вісник Хмельницького національного університету, серія: Технічні науки, Хмельницький. – 2017. - №5(253). – 112-120 с.
3. Ковтун В.В. Дослідження ефективності ознак розпізнавання мовців при використанні згортальних нейромереж / М.М. Биков, В.В. Ковтун // Оптико-електронні інформаційно-енергетичні технології, Вінниця. – 2016. - №2(32). – 22-28 с.

Таранюк Юлія Ярославівна — студентка групи КІВ-166, факультет комп'ютерних систем і автоматики, Вінницький національний технічний університет, Вінниця, e-mail: tysvnyr@ukr.net

Ковтун В'ячеслав Васильович — канд. техн. наук, доцент кафедри т комп'ютерних систем управління, Вінницький національний технічний університет, Вінниця, e-mail: kovtun_v_v@vntu.edu.ua

Науковий керівник: **Ковтун В'ячеслав Васильович** — канд. техн. наук, доцент кафедри т комп'ютерних систем управління, Вінницький національний технічний університет, Вінниця

Tanyuk Yulia Yaroslavivna — Student of the Group KIB-166, Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia, e-mail: tysvnyr@ukr.net

Kovtun Vjatcheslav Vasilievich — Cand. Sc. (Eng), Assistant Professor for the Computer Control Systems Department, Vinnytsia National Technical University, Vinnytsia, e-mail: kovtun_v_v@vntu.edu.ua

Supervisor: **Kovtun Vjatcheslav Vasilievich** — Cand. Sc. (Eng), Assistant Professor for the Computer Control Systems Department, Vinnytsia National Technical University, Vinnytsia