УДК 004.934

В.І. Каркульовський[1], В.С. Ткаченко[2]
Національний університет "Львівська політехніка",
[1]кафедра систем автоматизованого проектування,
[2]кафедра електронних засобів інформаційно-комп'ютерних технологій

ОСОБЛИВОСТІ МЕТОДІВ СЕГМЕНТАЦІЇ МОВЛЕННЄВИХ СИГНАЛІВ

Розглянуті деякі питання аналізу методів попередньої сегментації мовленнєвих сигналів та їх особливостей для задач розпізнавання.

Ключові слова – сегментація сигналів, розпізнавання, метод

In this paper some questions of analysis of methods of preliminary segmentation of speech signals and their features for the tasks of recognition are considered.

Keywords – signal segmentation, recognition, method

Вступ

Мовленнєвий сигнал складається з певних структурних одиниць – речень, слів, складів, фонем. Фонема являє собою найменшу структурно-семантичну звукову одиницю, що здатна виконувати основні функції у мовленні.

Попередня сегментація мовленнєвих сигналів на структурні елементи (передусім на фонеми) може істотно підвищити ефективність подальшого розв'язання задач розпізнавання. Хоча в деяких підходах і методах розпізнавання мовленнєвих сигналів етап попередньої сегментації відсутній [1] (він безпосередньо пов'язується з етапом розпізнавання, і насамперед це стосується злитих мовленнєвих сигналів), однак у деяких публікаціях методам попередньої сегментації приділено значну увагу.

Особливості окремих методів сегментації мовленнєвих сигналів можна розглянути на прикладі фрази "Василь Стефаник", сигнал для якої наведений на рис. 1.

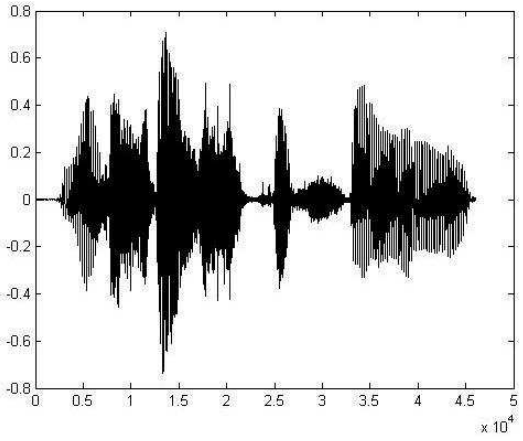


Рис. 1. Вигляд мовленнєвого сигналу для тестової фрази

Сегментація на основі обчислення значення енергії в заданому часовому вікні (фреймі)

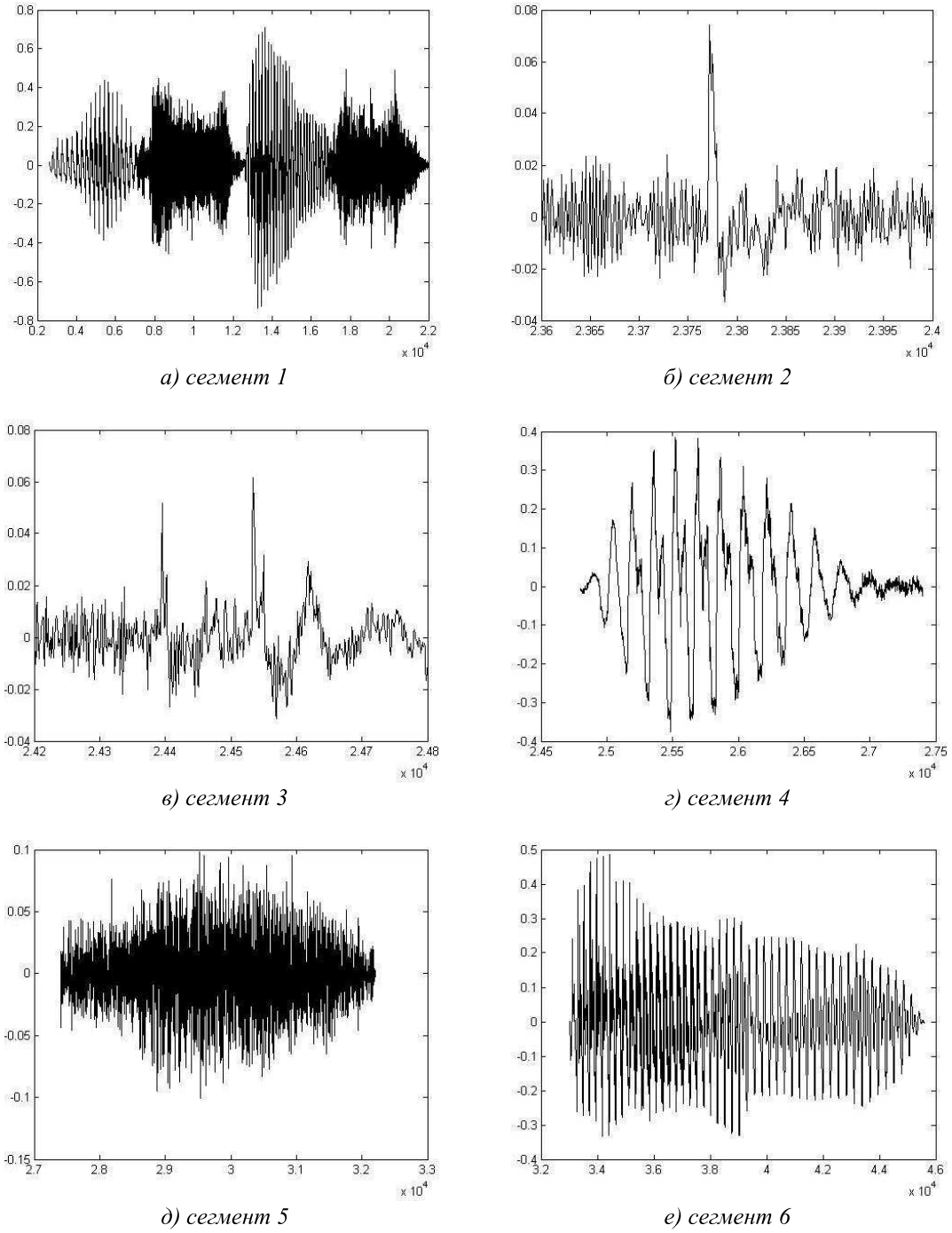
Цей метод є одним з найпростіших, однак він звичайно дає змогу визначати лише паузи між словами і малопридатний для сегментації злитої мови. При високому рівні шуму енергії невокалізованих фонем та шумів є близькими, подібними є і їхні спектри.

На основі аналізу мовленнєвого сигналу $s[i]$ визначаються його основні параметри – частота дискретизації f_s, максимальне $s_{\max}$ та мінімальне $s_{\min}$ значення, середнє значення питомої енергії

сигналу $\overline{E} = \frac{1}{N}\sum_{i=0}^{N-1} s^2[i]$. На основі f_s визначається кількість дискретних значень m сигналу, які відповідають заданому часовому фрагменту Δt: $m = [\Delta t \cdot f_s + 1]$, де [ ] визначає цілу частину числа. Максимальне та мінімальне значення використовуються для нормування сигналу в заданому діапазоні амплітуд (наприклад, (-1; 1)).

Задачею сегментації є встановлення значень відліків сигналу, які відповідають початку та кінцю відповідного сегменту сигналу: (n_{p1}, n_{k1}), (n_{p2}, n_{k2}), … . Мовленнєвий сигнал розділяється на фрейми заданої тривалості (15 мс), яка відповідає мінімальній тривалості фонеми.

На рис. 2 наведено кілька виділених фрагментів сигналу, отриманих на основі сегментації за питомою енергією сигналу в часовому вікні.



а) сегмент 1 *б) сегмент 2*

в) сегмент 3 *г) сегмент 4*

д) сегмент 5 *е) сегмент 6*

Рис. 2. Сегменти мовленнєвого сигналу, виділені на основі питомої енергії в часовому вікні

Сегментація голосних звуків на основі спектрального аналізу (спектрограми)

Цей метод [2] дає змогу виділити в мовленнєвому сигналі квазістаціонарні фрагменти, що відповідають голосним звукам, однак виділити форманти невокалізованих звуків достатньо складно. Метод базується на дослідженні в спектральній області, обчисленні короткочасного спектра (з попереднім використанням спектральних вагових віконних функцій) та дослідженні формантних частот. При цьому можна досягнути 80–85 % правильної сегментації з наступним правильним детектуванням голосних [2]. На рис. 3 наведено спектрограму сигналу з 256 часовими фрагментами. Як очевидно з цього рисунку, в спектрограмі наявні достатньо чіткі межі формантних частот, за якими можна виділити фрагменти, що відповідають голосним звукам.

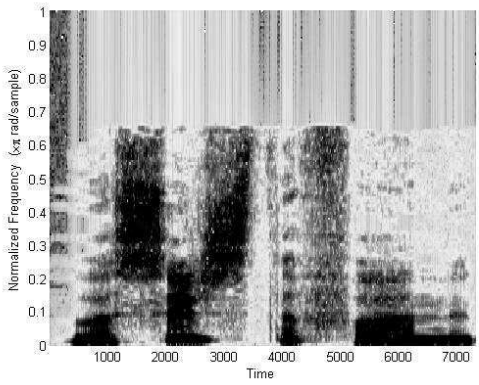


Рис. 3. Спектрограма мовленнєвого сигналу

Сегментація на основі кореляції між спектрами фрагментів сигналу однакової тривалості

У цьому методі [3], як і в деяких інших, тією чи іншою мірою структурні елементи виділяються на основі моделей формування мовленнєвих сигналів, в яких кожному типу елемента відповідають свої особливості формування. У кожному з цих випадків здійснюється пошук меж квазістаціонарних та перехідних процесів. На рис. 4 наведено коефіцієнти кореляції спектрів сигналів у сусідніх часових вікнах тривалістю 15 мс.

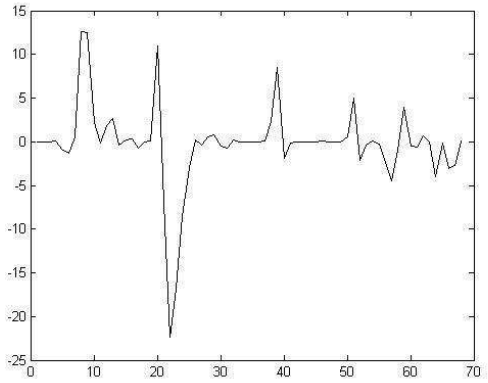


Рис. 4. Коефіцієнти кореляції спектрів сигналів у сусідніх часових вікнах (аргумент – номер часового вікна)

Різкі зміни коефіцієнтів кореляції спектрів сигналів у сусідніх часових вікнах визначають міжфонемні переходи.

Сегментація з використанням алгоритмів швидкого вейвлет-перетворення

Використання вейвлет-перетворення дає змогу визначати міжфонемні переходи по крайній мірі для фонем, які відповідають порівняно довготривалим квазістаціонарним ділянкам мовленнєвих сигналів [4]. На міжфонемних переходах сигнал сильно змінюється зразу на кількох масштабах дослідження, що проявляється у зростанні вейвлет-коефіцієнтів на кількох рівнях

деталізації, тоді як на стаціонарних ділянках фонем вейвлет-коефіцієнти проявляються лише на певних масштабах. Отже, знаходження міжфонемних меж може бути зведене до знаходження моментів зростання вейвлет-коефіцієнтів на значній кількості рівнів масштабування. Можливе використання кількох вейвлетних базисів для пошуку міжфонемних переходів в кожному з них з наступним об'єднанням результатів. На рис. 5 наведено структуру коефіцієнтів апроксимації та деталей при розкладі сигналу по 10 рівнях з вейвлет-функцією db1.

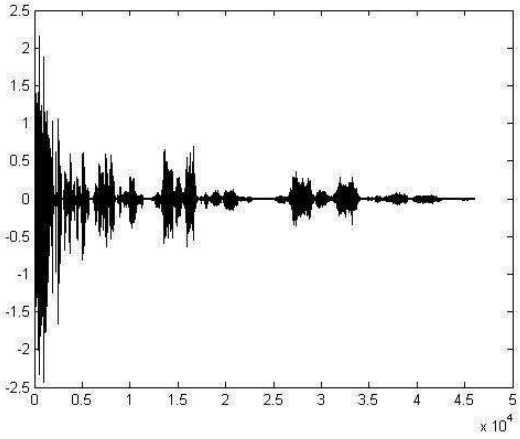


Рис. 5. Вейвлет-розклад сигналу по 10 рівнях

Сегментація базується та тому, що на міжфонемних переходах сигнал сильно змінюється зразу на кількох масштабах, що проявляється у зростанні вейвлет-коефіцієнтів на кількох рівнях деталізації.

Сегментація на основі використання штучних нейронних мереж

Різні підходи до застосування штучних нейронних мереж для розпізнавання та сегментації мовленнєвих сигналів [5] відрізняються, в основному, вибором вхідних параметрів для нейронної мережі, серед яких можна виділити: спектральні коефіцієнти, обчислені в окремих частотних смугах, коефіцієнти лінійного передбачення, кепстральні коефіцієнти для лінійного масштабу, мел- і барк-масштабу частот.

Крім того, можливий вибір різних архітектур штучних нейронних мереж, серед яких найпоширенішими є тришарові нейронні мережі прямого поширення зі зворотним поширенням похибки та нейронні мережі, які реалізують кластеризацію фрагменту мовленнєвого сигналу.

Висновки

Сегментація мовленнєвих сигналів є одним з важливих етапів при розв'язанні задач їх розпізнавання, правильна сегментація може підвищити ефективність розпізнавання[6]. Внаслідок значної складності мовленнєвих сигналів при їх сегментації доцільно використовувати комбінації розглянутих методів, які для певних видів фонем дають найкращі результати, що вимагає їх подальшого дослідження.

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К.: Наукова думка, 1987. – 264 с. 2. Сорокин В.Н., Цыплихин А.И. Сегментация и распознавание гласных // Информационные процессы. – 2004. – Т. 4, № 2. – С. 202–220. 3. Цыплихин А.И., Сорокин В.Н. Сегментация речи на кардинальные элементы // Информационные процессы. – 2006. – Т. 6, № 3. – С. 177–207. 4. Yermolenko T. Segmentation of a speech signal with application of fast wavelet transformation // Information Theories & Applications. – 2002. – Vol. 10. – 5 p. 5. Kamarauskas J. Automatic Segmentation of Phonemes using Artificial Neural Networks // Electronics and Electrical Engineering. – 2006. – No. 8 (72). – P. 39–42. 6. V. Pavlysh, Y. Romanyshyn, V. Tkachenko. Preliminary segmentation of speech signals for the tasks of their recognition. Perspective Technologies and Mathods in MEMS Design. Proceeding of the Vth International Conference of Young Scientists MEMSTECH'2009, Lviv. Publishing House of Lviv Polytechnic National University, 2007. – P. 144.