

Ю. Рашкевич, Т. Кватер*, З. Шиманські**
 Національний університет “Львівська політехніка”,
 кафедра автоматизованих систем управління,

*Університет Жешувський (Польща),

**Вища школа підприємництва та управління (Лодзь, Польща)

АЛГОРИТМ АВТОМАТИЧНОГО ВИДІЛЕННЯ ДИФТОНГІВ ПОЛЬСЬКОЇ МОВИ

©Рашкевич Ю., Кватер Т., Шиманські З., 2007

Розроблено алгоритм автоматичної сегментації та маркірування дифтонгів *a* та *e* польської мови в неперервному мовному потоці для задач перетворення часового масштабу мовного сигналу.

The algorithm of automatic segmentation and labeling of Polish diphthongs *a* and *e* in continuous speech signal is developed for use in speech time-scale modification.

Вступ. Дифтонги польської мови *a* та *e* належать до класу фонем, які не є типовими для інших мов, наприклад, української або англійської. Тому розроблені універсальні інформаційні технології класифікації та розпізнавання звуків не є придатними для опрацювання дифтонгів. Водночас деякі прикладні задачі (регулювання темпу мови, нормалізація сигналу в задачах верифікації та ідентифікації дикторів тощо) передбачають на етапі попередньої обробки класифікацію звукових одиниць у неперервному потоці мовлення. Більш того, розроблені в [1] адаптивні технології перетворення часового масштабу сигналів вимагають точного виділення та маркірування певних класів звуків, оскільки виконувані трансформації сигналу ґрунтуються на так званих функціях темпоральних перетворень (ФТП), розроблених для характерних для кожної мови класів звуків.

У зв'язку з цим проблема автоматизації сегментації та маркірування польських дифтонгів є актуальною науковою та прикладною задачею.

Постановка задачі. У роботі [2] показано, що у загальному випадку польський дифтонг складається із 5 основних ділянок: перехід від попередньої фонемі до голосної частини дифтонга (о або е), стаціонар голосної частини, перехід від стаціонару до вокалізованої приголосної частини, стаціонар вокалізованої приголосної частини та перехід до наступної фонемі. Особливістю польської мови є те, що залежно від наступної після дифтонга фонемі вокалізованою приголосною можуть бути фонемі *m* або *n*, при цьому фонема *n* може набувати вигляду двох різних алофонів – *n* та *ŋ*. Проведені польськими науковцями фонетичні та фонологічні дослідження звуків польської мови показують [3], що різноманіття є ще більшим: виділяються 4 алофони фонемі *m* та 11 алофонів фонемі *n*. Наші статистичні дослідження в контексті особливостей зміни структури фонемі залежно від темпу мовлення показують, що можна обмежитися трьома основними алофонами – *m*, *n* та *ŋ*. Алофонна транскрипція дифтонгів *a* та *e* залежить від такої фонемі:

- перед губними приголосними *a* та *e* набувають відповідно вигляду [om] та [em] (*dąb, wsep*);
- перед зубними приголосними та африкатами – вигляду [on] та [en] (*paczek, błędny*);
- перед середньо- та задньоязиковими приголосними – вигляду [oŋ] та [eŋ] (*mądra, potęga*).

Для адаптивних технологій перетворення тривалостей звуків цікавим є лише третій випадок, оскільки у перших двох дифтонг може бути адекватно описаний послідовністю класів звуків: вокалізований голосний – вокалізований приголосний, ФТП яких для української мови є розробленими і можуть бути успішно використані і для польської мови. Для виділення дифтонгів, що описуються транскрипцією [oŋ] та [eŋ], необхідно розробити алгоритм автоматичного розпізнавання звуку *ŋ* та правила маркірування дифтонга загалом.

Розроблення алгоритму автоматичної сегментації та маркірування дифтонгів

У загальному випадку процедура сегментації та маркірування звуків містить 6 основних етапів [1]:

- виділення мовного сигналу на фоні шуму та пауз;
- сегментацію на квазістаціонарні ділянки;
- аналіз сегментів та виділення параметрів для класифікації;
- класифікацію сегментів за ознаками вокалізованості;
- об'єднання сегментів у звукові одиниці;
- маркірування звукових одиниць.

Ключовими етапами тут є аналіз сегментів та виділення параметрів, які дають змогу класифікувати сегменти за ознаками вокалізованості із подальшим об'єднанням у звукові одиниці виділених наперед класів.

Найпоширенішими параметрами мовного сигналу є :

- короткочасна енергія мовного сигналу, визначена на повному діапазоні частот E ;
- відношення E_n / E_e енергії сигналу в діапазоні 250–600 Гц до енергії в діапазоні 650–3000 Гц;
- тривалість квазістаціонарної ділянки;
- значення першого коефіцієнта моделі авторегресії A_1 .

На основі цих параметрів додатково обчислюються ознака голосності для виділення голосних та вокалізованих приголосних $O_2 = (17 E - 70 E_n / E_e) 1000 + 1.1 | A_1 |$ та ознака вокалізованості $O_e = 0.03 E - A_1$. У роботі [1] показано, що на основі цих ознак можна виділяти такі класи звуків: голосні (ВГ), вокалізовані приголосні (ВП), невокалізовані приголосні (НП) та вибухові звуки (ВЗ).

Отже, для класифікації дифтонгів залишається невирішеною задача розпізнавання η . Ця фонема за своїми параметрами є дуже близькою до фонем **n**. Тому необхідно лише знайти додаткову інформативну ознаку, яка б дала змогу розпізнати η в класі назальних вокалізованих приголосних.

Для цього були детально проаналізовані динамічні спектрограми дифтонгів a та e , розташованих на початку, в середині та наприкінці слова. Всього проаналізовано 60 спектрограм (по 30 для кожного із дифтонгів). Було визначено, що спектр звуку η порівняно із спектром звуку **n** характеризується підвищеним коефіцієнтом відношення короткочасної енергії в діапазоні 1200 – 1600 Гц до енергії в діапазоні 2200 – 3200 Гц. Експерименти показали, що при порозі 0,6 понад 90 % звуків η були розпізнані правильно, і лише у 15 % приголосна **n** була розпізнана як η . Для задач перетворення часового масштабу мовних сигналів на основі адаптивних технологій із використанням функцій темпоральних перетворень ці показники є задовільними.

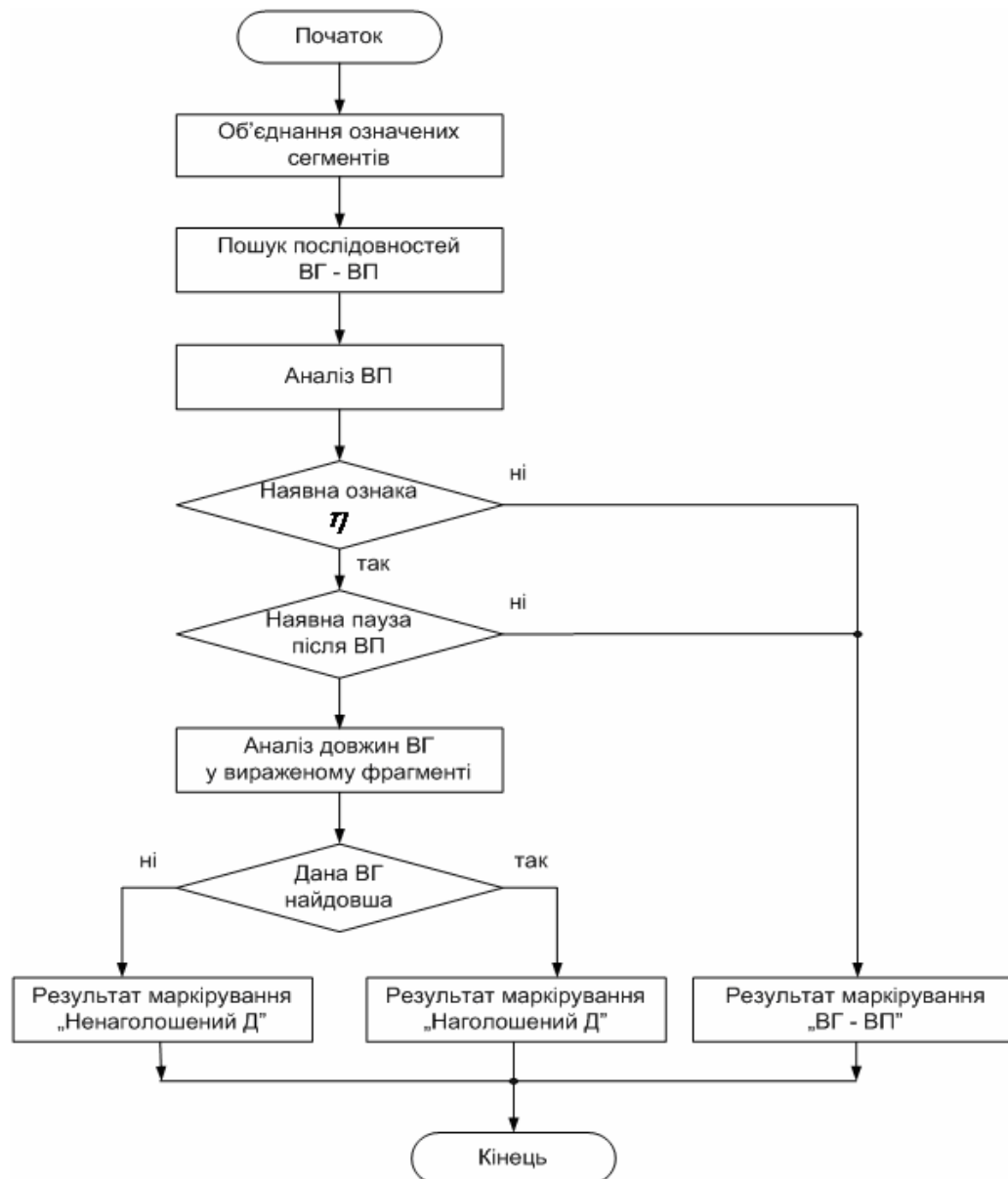
Дослідження динамічних спектрограм дифтонгів також показали, що додатковою ознакою для класифікування дифтонгів є обов'язкова наявність паузи завдовжки від 15 до 40 мілісекунд після приголосної частини дифтонга.

Тривалість голосної частини дифтонга є інформативною ознакою для розподілу наголошених та ненаголошених дифтонгів.

Отже, алгоритм автоматичної сегментації дифтонгів польської мови (рис. 1) можна подати у вигляді такої послідовності етапів:

1. Після виконання процедури об'єднання означених сегментів відбувається пошук послідовностей ВГ–ВП.
2. Для кожної пари ВГ–ВП аналізуються та виділяються інформативні параметри ВП.
3. Перевірка наявності інформативної ознаки η : у випадку її відсутності результатом маркірування пари є рішення про два окремі звуки – ВГ та ВП, у випадку наявності ознаки проводиться аналіз наступної після ВП ділянки.
4. Якщо наступною ділянкою не є пауза, то результатом маркірування пари є аналогічне рішення про два окремі звуки – ВГ та ВП; якщо є пауза тривалістю понад 15 мс, то проміжний результат маркірування – дифтонг.

5. Тривалість ВГ порівнюється із тривалостями інших ВГ виділеного фрагмента, якщо тривалість цієї ВГ є найбільшою, то кінцевий результат маркірування – наголошений дифтонг, якщо ж ні, то кінцевий результат маркірування – ненаголошений дифтонг.



Блок-схема алгоритму автоматичного виділення дифтонгів

Висновки. Запропонований алгоритм був експериментально апробований на наборі 30 слів, що містили дифтонги обидвох видів. У 26 випадках класифікація дифтонгів була правильною, у 4-х випадках дифтонг був класифікований як послідовність ВГ–ВП. Ці результати дають можливість рекомендувати запропонований алгоритм для використання в задачах перетворення часового масштабу мовних сигналів.

1. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. – Львів: ТЗОВ НВТ Академічний експрес, 1997. – 143 с. 2. Шиманські З., Фігура Р., Рашкевич Ю., Марцишин Р. Побудова функцій темпорального перетворення дифтонгів польської мови // Вісник Нац. ун-ту “Львівська політехніка”: Комп’ютерна інженерія та інформаційні технології. – 2003. – №481. – С. 152–156. 3. Ostaszewska D., Tambor J. Fonetyka i fonologia współczesnego języka polskiego. Warszawa: Wydawnictwo naukowe PWN SA, 2000. – 142 s.