

546. С. 23–29. 2. Поспелов И.Г. Применение принципов теории живучести многопроцессорных вычислительных систем для повышения их срока службы в условиях радиационных воздействий // Радиационная стойкость электронных систем “Стойкость-2000”: Науч.-тех. сб. – М.: Папрус, 2000. – Вып. 3. – С. 29–30. 3. Ваврук Є.Я. Організація контролю та діагностики ШПФ-схем у режимі реального часу в системах опрацювання сигналів // Вісн. Нац. ун-ту “Львівська політехніка”. – 2004. – 523. С. 24–29. 4. Ушаков И.А. Методы расчета эффективности систем на этапе проектирования. – М.: Знание, 1983. – 104 с. 5. Мультипроцессорные вычислительные системы / Под ред. Я.А. Хетагурова. – М.: Энергия, 1971. – 320 с. 6. Tinghuai Chen. Fault diagnosis and Fault Tolerance. A Systematic Approach to Special Topics. – Springer-Verlag Berlin Heidelberg, 1992. – P. 118–133. 7. А. с. 1679505 СССР, МКИ Н 05 К 10/00. Способ контроля правильности функционирования дискретных устройств / И.Б. Шубинский, В.И. Николаев, А.М. Заяц и др. – 1980. – 5 с.

УДК 681.84.087.4

Ю. Рашкевич, Р. Марцишин, З. Шиманські*

Національний університет “Львівська політехніка”,

кафедра автоматизованих систем управління

*Вища школа підприємництва та управління (м. Лодзь, Республіка Польща)

АВТОМАТИЧНА СЕГМЕНТАЦІЯ ТА МАРКІРУВАННЯ ДИФТОНГІВ ПОЛЬСЬКОЇ МОВИ

© Рашкевич Ю., Марцишин Р., Шиманські З., 2006

Запропоновано алгоритм автоматичної сегментації та маркірування дифтонгів польської мови для задач перетворення часового масштабу сигналів на основі адаптивного підходу із використанням функцій темпоральних перетворень. На основі експериментальних досліджень мовного сигналу сформовано набір інформативних параметрів для сегментації елементів дифтонгу та правила його маркірування.

An algorithm of automatic segmentation and labeling of Polish diphthongs for time-scale modification of speech signals on the base of adaptive approach with the use of temporal transforming functions is proposed. As a result of natural experiments a set of informative characteristics for segmentation as well as a labeling rules are developed.

Вступ

Перетворення часового масштабу є важливою задачею аналізу, розпізнавання та синтезу мовних сигналів. Його використовують як елемент попередньої обробки сигналів в задачах розпізнавання та синтезу мови, а також як основний елемент задач регулювання темпу, передачі мови каналами зв'язку, відновлення сигналу в умовах штучних атмосфер тощо. Найперспективнішим напрямком вирішення задач перетворення часового масштабу для мовних сигналів є виконання операцій розтягування/стиснення в часовій області на основі адаптивного підходу, що дає змогу отримати сигнал, максимально наближений за своєю внутрішньою структурою до сигналу, сформованого в умовах природного прискорення чи сповільнення мовлення людиною. В роботі [1] запропоновано технологію адаптивного перетворення мовного сигналу із використанням функцій темпоральних перетворень (ФТП), які задають послідовність виділення елементарних сегментів мовного сигналу в процесі його скорочення. ФТП будуються на основі експериментальних досліджень зміни структури мовного сигналу за різних темпів мовлення і є, очевидно, індивідуалізованими для кожної із мов як за кількістю видів ФТП, так і за їхньою внутрішньою структурою. Незважаючи на те, що в різних мовах є багато подібних за звучанням та

особливостями перетворення фонем, для яких можна використати існуючі апробовані технології, в кожній мові є свої особливі фонемні, які потребують індивідуалізованих підходів. У польській мові насамперед такими фонемами є дифтонги *q* та *ε*, які не мають аналогів в інших мовах.

Сьогодні за результатами досліджень [2–4] запропоновано класифікацію звуків польської мови для задач зміни темпу мовлення, з'ясовано особливості перетворення та побудовано ФТП для наголошених та ненаголошених дифтонгів. Але водночас ще не розроблено алгоритми класифікації характерних елементів дифтонгів та відповідні процедури маркірування, що обмежує можливості застосування вищеописаної технології адаптивного перетворення мовного сигналу.

Отже, метою статті є виявлення інформативних ознак для класифікації окремих елементів дифтонгів польської мови та розроблення правил маркірування відповідних фонем загалом.

Постановка задачі

У роботі [4] наведено статистичні дані про структуру дифтонгів за різного розташування в словах та різних темпах мовлення. У загальному випадку дифтонг складається із 5 основних ділянок: перехід від попередньої фонемі до голосної частини дифтонга (*o* або *e*), 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 фонемні *m* або *n*, при цьому фонема *n* може набувати вигляду двох різних алофонів – *n* та *ŋ*. Треба зауважити, що фонетичними та фонологічними дослідженнями звуків польської мови доведено, що різноманіття є ще більшим – у роботі [5] виділено 4 алофони фонемі *m* та 11 алофонів фонемі *n*. Результати статистичних досліджень в контексті особливостей зміни структури фонемі залежно від темпу мовлення свідчать, що можна обмежитися трьома основними алофонами – *m*, *n* та *ŋ*. Алофонна транскрипція дифтонгів *q* та *ε* залежить від такої фонемі:

перед губними приголосними *q* та *ε* набувають відповідно вигляду *[om]* та *[em]*;

перед зубними приголосними та африкатами – вигляду *[on]* та *[en]*;

перед середньо- та задньоязиковими приголосними – вигляду *[oŋ]* та *[eŋ]*.

З певним ступенем імовірності (аудитори під час тестових випробувань не зазначили істотних відмінностей) можна вважати, що під час операцій вилучення окремих ділянок звуків з метою скорочення загальної тривалості (прискорення темпу мовлення) часову структуру дифтонга в перших двох випадках можна перетворювати, трактуючи дифтонг як послідовне сполучення двох фонем – голосної (*o* або *e*) та вокалізованої приголосної (*m* або *n*) з відповідними їм перехідними ділянками. При цьому можна використовувати розроблені для української мови ФТП відповідних фонем. У третьому випадку відсутність ФТП фонемі *ŋ* вимагає проведення додаткових досліджень.

Отже, для розв'язання поставленої задачі нами було проведено ряд експериментів на реальному сигналі з метою вивчення відмінностей в сполученнях фонем типу *on(en)* та дифтонгів *q(ε)*, аналізу стаціонарних голосних та стаціонарних приголосних ділянок, визначення інформативних параметрів для класифікації фонемі *ŋ* та побудови відповідних правил маркірування дифтонга.

Акустичним матеріалом для досліджень вибрано пари слів типу *szalony – mađra*, в яких сполучення фонем *on(en)* та дифтонги *q(ε)* були в однакових умовах наголосу.

Результати експериментальних досліджень

При проведенні експериментальних досліджень частота дискретизації мовного сигналу становила 12.5 кГц, частота зрізу вхідного фільтра – 5 кГц, розрядність аналого-цифрового перетворення – 16, подання значення відліку мовного сигналу – двобайтне. Для розділення сигналу на квазістаціонарні ділянки (сегменти) використано алгоритм сегментації на основі критерію правдоподібності [1] з такими параметрами: порядок моделі – 6, довжина вікна аналізу – 10 мс, порогове значення – 260. Класифікація виділених сегментів містила два основні етапи:

I Розподіл сегментів на вокалізовані та невокалізовані ділянки.

Розподіл проводили за двома інформативними параметрами – першого коефіцієнта фільтра моделі авторегресії A_1 та короточасної енергії E . Сегменти, для яких $A_1 > -0.4$, класифікували як невокалізовані, а сегменти з $A_1 < -1.5$ – як вокалізовані. Для решти сегментів визначали додаткову ознаку вокалізованості $O_e = 0.03E - A_1$, а сегменти вважали вокалізованими, якщо $O_e > 2.2$.

II Розподіл вокалізованих сегментів на голосні та приголосні.

Додатковим інформативним параметром на цьому етапі було E_H/E_e – відношення енергії сигналу в області низьких частот (250–600 Гц) до енергії в діапазоні високих частот (650 – 3000 Гц). Вокалізованими голосними сегментами вважали ті, для яких $A_1 < -1.6$ або $E > 65$. Для решти сегментів обчислювали ознаку голосності $O_g = (17E - 70E_H/E_e)/1000 + 1.1|A_1|$. Сегменти, для яких $O_g > 2.2$, класифікували як голосні, решту – як вокалізовані приголосні.

Об'єднували сегменти у звукові одиниці (голосні та вокалізовані приголосні) за запропонованими у [1] правилами коригування.

Отримані так результати автоматичної сегментації та класифікації додатково коригували вручну на основі динамічних спектрограм; корекція становила не більше ніж 5 мс.

На рис. 1 наведено часовий сигнал, динамічну спектрограму та короточасний спектр ділянок, які відповідають голосній у сполученнях фонем [on] (слово *szalony*) та [oŋ] (слово *mądra*). Близькість інформативних параметрів та аналогічність структури динамічних спектрограм дають змогу зробити висновок, що наведену вище методику сегментації та класифікації можна однаково успішно застосовувати як для класифікації голосних фонем сполученнях голосна–приголосна, так і для класифікації голосної частини дифтонга, але не можна вважати критерієм розподілу сполучення голосна–приголосна та дифтонга загалом. Останнє є важливим, оскільки, як показано в [2,4], у разі зміни темпу мовлення перетворення структури сполучення голосна–приголосна та структури дифтонга істотно відрізняються.

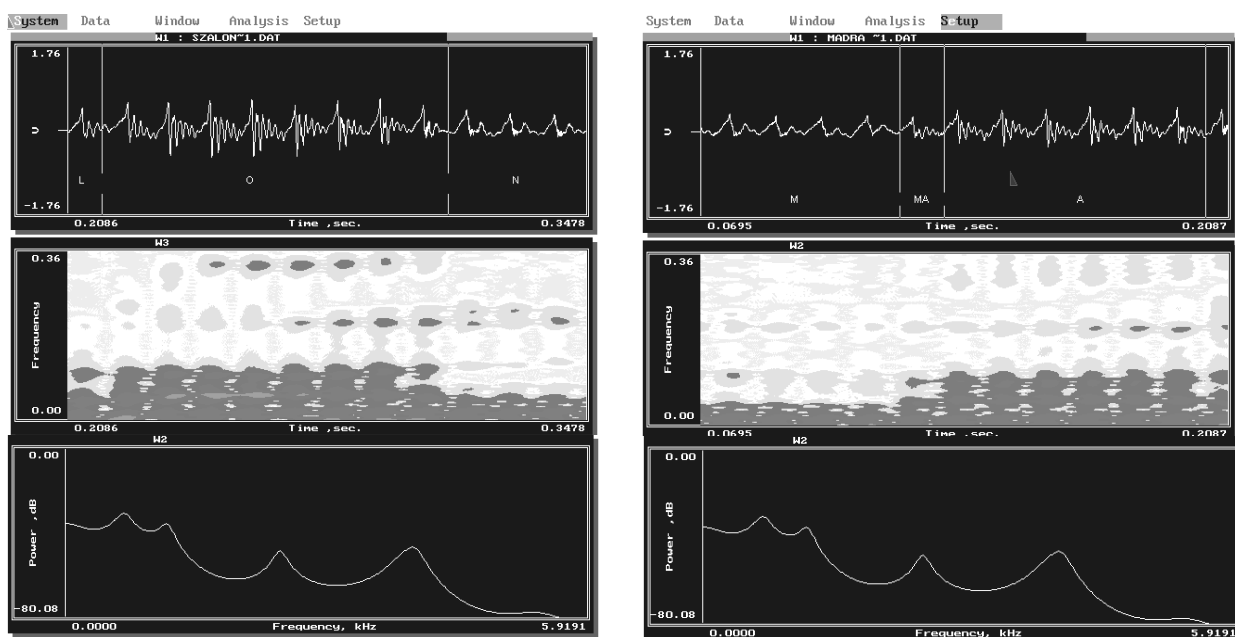


Рис. 1. Голосна о та голосна частина дифтонга q

На рис. 2 наведено часовий сигнал, динамічну спектрограму та короточасний спектр ділянок, які відповідають вокалізованій приголосній п у сполученні фонем [on] (слово *szalony*) та вокалізованій приголосній η у дифтонзі q (слово *mądra*).

За результатами досліджень, носові вокалізовані приголосні m та n характеризуються спектром із відносно невисокою енергією в діапазоні частот 1–2 кГц порівняно із енергією в діапазоні частот 2–3 кГц. Разом з тим, носова вокалізована приголосна η має в першому діапазоні

чітку формантну складову (див. рис. 2). Отже, інформативна ознака E_1/E_2 – відношення енергій сигналу в області 1200–1600 Гц та в області 2200–3200 Гц може бути основою для класифікації. Вказані границі областей отримані нами в результаті експериментальних досліджень 20 пар $n - \eta$, де було 10 реалізацій під наголосом і 10 ненаголошених. Ті самі результати дослідження дають змогу рекомендувати значення E_1/E_2 , що дорівнює 0.6, як порогове для визначення η .

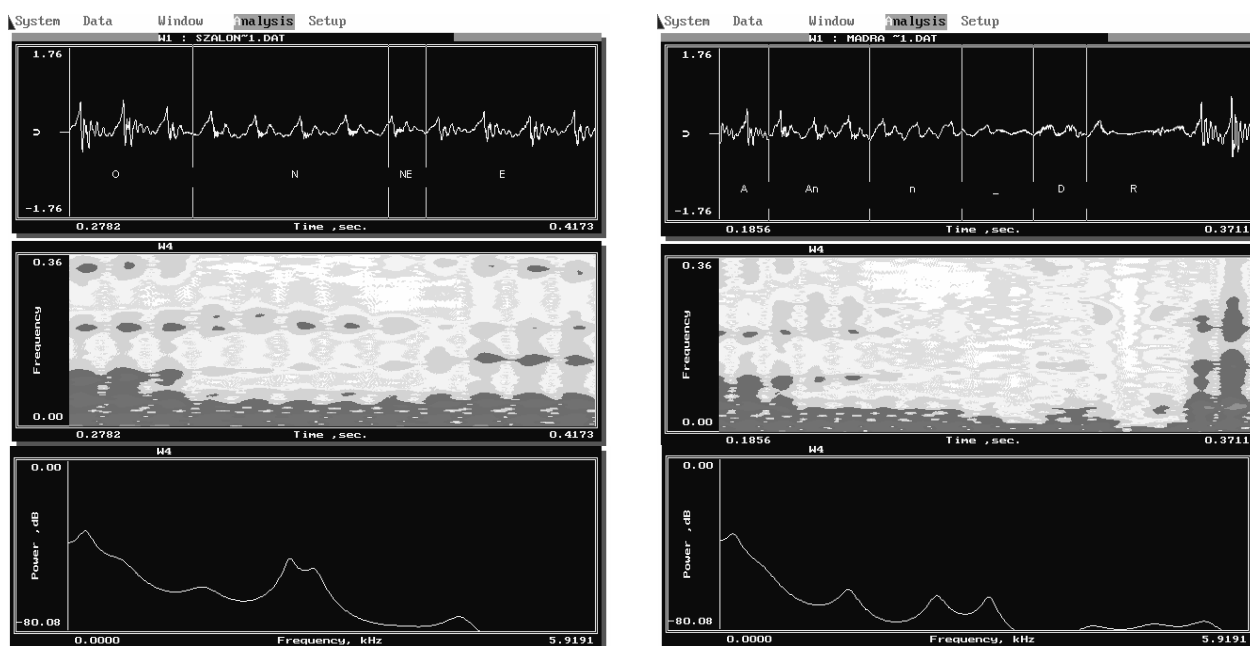


Рис. 2. Приголосна o та приголосна частина дифтонга a

Ще однією особливістю часової структури дифтонгів є обов'язкова наявність паузи завдовжки принаймні 15–20 мс перед наступною фонемою. Тому для маркування дифтонга можна сформулювати таке правило: голосна (o або e) – вокалізована приголосна із $E_1/E_2 > 0.6$ – пауза (> 15 мс).

Висновки

У процесі досліджень проаналізовано особливості часової структури сполучень фонем типу $on(en)$ та дифтонгів $a(e)$, виділено інформативні ознаки та експериментально встановлено порогове значення для класифікації носової вокалізованої приголосної η , сформульовано правило маркування дифтонга, що дає змогу підвищити якість розбиття звукових одиниць на класи та ефективно використовувати технології адаптивного перетворення часового масштабу сигналів.

1. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. – Львів: Акад. експрес, 1997. – 140 с.
2. Шиманські З., Фігура Р., Рашкевич Ю., Марцишин Р. Побудова функцій темпорального перетворення дифтонгів польської мови // Вісн. Нац. ун-ту “Львівська політехніка”. – 2003. – № 481. – С. 152–156.
3. Szymanski Z., Figura R. Klasyfikacja dźwięków języka polskiego dla zadan zmianę tempa // Вісн. Нац. ун-ту “Львівська політехніка”. – 2003. – № 496. – С. 263–268.
4. Шиманські З., Фігура Р., Марцишин Р. Особливості перетворення часової структури дифтонгів польської мови // Вісн. Нац. ун-ту “Львівська політехніка”. – 2002. – № 468. – С. 170–174.
5. Ostaszewska D., Tambor J. Fonetyka i fonologia współczesnego języka polskiego. – Warszawa: Wydawnictwo naukowe PWN, 2000. – 142 с.