

линейных сцепок // Контрольно-измерительная техника. – Львов: Вища школа. – 1989. – Вып.45. – С.23 – 25.
 2. Різник В.В. Синтез оптимальних комбінаторних систем. – Львів: Вища школа, 1989. – 168 с.

3. Шиманські́, Р.Фігура́, Ю.Рашкевич, Р.Марцишин
 *Вища школа підприємництва та управління м.Лодзь, Польща,
 Національний університет "Львівська політехніка"

УДК 681.84.087

ПОБУДОВА ФУНКЦІЙ ТЕМПОРАЛЬНОГО ПЕРТВОРЕННЯ ДИФТОНГІВ ПОЛЬСЬКОЇ МОВИ

© Шиманські З., Фігура Р., Рашкевич Ю., Марцишин Р., 2003

Пропонуються результати дослідження залежностей довжин ділянок дифтонгів польської мови від зміни темпу мовлення. Показано необхідність використання адаптивних методів перетворення часового масштабу мови. Побудовано функції зміни темпоральної структури дифтонгів для зміни темпу мовлення.

In this paper offered Polish language diphthongs areas of lengths dependence on the change of speech tempo research results. A necessity of adaptive methods timescale modification of speech is shown. Functions of change of temporal structure of diphthongs for the change of speech tempo are built.

Вступ

Серед важливих проблем оброблення мовних сигналів особливе місце займає проблема перетворення часової структури мови, оскільки такі перетворення надзвичайно важливі в задачах аналізу – синтезу мови, верифікації та ідентифікації дикторів, в системах кодування та передавання мови каналами зв'язку. Розроблення та використання ефетивних методів перетворення часового масштабу мови для задач зміни темпу подачі мовної інформації (регулювання темпу мови) є актуальними завданнями, оскільки особливості процесу мовотворення не дозволяють в загальному випадку забезпечити максимальне завантаження слухового каналу, який здатен суттєво швидше обробляти інформацію порівняно з мовним каналом [1]. Важливість проблематики досліджень у даному напрямку визначається роботами вчених Малаха, Портноффа, Гріффіна, Моулінеза, Кватієрі, а в Україні, зокрема, такими дослідженнями займалися наукові колективи, очолювані Т.Вінцюком, М.Бондаренком, В.Грициком [2].

Існуючі методи перетворення часової структури мови класифікуються за двома ознаками: областю виконання перетворень (часова чи частотна) та підходом до трансформації структури (лінійний чи адаптивний до сигналу мови) [2]. Адаптивні методи найбільш повно відповідають реальним процесам мовотворення і дозволяють значно розширити діапазон перетворення часової структури мовного сигналу та підвищити натуральність мови. Такі методи використовують кусково-постійну модель мовного сигналу, яка дозволяє відобразити нелінійності структури мови, що важливо при розробленні методів перетворення для високих коефіцієнтів.

Для використання адаптивних методів перетворення необхідно ідентифікувати внутрішню структуру для кожного із виділених класів звуків та виконувати перетворення на основі спеціально розроблених функцій зміни темпоральної структури звуку (ФЗТС). Такі функції є специфічними для кожної мови зокрема, і побудова їх вимагає індивідуального підходу.

Проблема розроблення ефективних методів перетворення часової структури польської мови сьогодні є маловивченою. Водночас попередні статистичні дослідження окремих звуків польської мови при різних темпах [3] показують, що, особливо для характерних для польської мови дифтонгів [4], не можна використовувати відомі типи ФЗТС.

Тому метою даної статті є побудова функцій змін темпоральних структур дифтонгів польської мови.

1. Адаптивне перетворення часової структури мови

Адаптивні перетворення часової структури мови виконуються на основі темпоральної моделі [2], яка подає сигнал мови у вигляді послідовно з'єднаних квазістаціонарних ділянок (кускова постійність структури) та подає по-новому відомості про ступінь та спосіб зміни часової структури мови.

Представимо модель мовного сигналу X у вигляді послідовного з'єднання ділянок:

$$X = X_1, X_2, \dots, X_p, \dots, X_n, \quad (1)$$

де кожна ділянка X_p являє собою звук, описаний темпоральною транскрипцією:

$$\tau_i = (l_i, k_i, w(k_i)), \quad (2)$$

де l_i – тривалість звуку чи паузи; k_i – клас звуку; $w(k_i)$ – функція зміни темпоральної структури звуку, яка задає алгоритм перетворення тривалості, що залежить від типу звуку, тривалості звуку, коефіцієнта перетворення часового масштабу та від результатів сегментації.

ФЗТС задається у вигляді неперервної, нормованої від 0 до 1 за часом та амплітудою, для кожного класу звуку.

Побудова ФЗТС для кожного класу звуків дозволяє розробляти алгоритми перетворення часової структури звуків, які максимально відповідають змінам структури мови при зміні диктором темпу мовлення. Використання ФЗТС для перетворення часової структури звуків дозволяє також значно зменшити шуми, викликані "склею-

ванням" мовного сигналу, оскільки точки склеювання знаходяться на квазістаціонарних ділянках. Також спрощується проблема забезпечення необхідної тривалості вихідної ділянки при невизначених результатах сегментації.

2. Побудова функцій темпоральних перетворень

Для побудови ФТП необхідно отримати спочатку темпоральні структури дифтонгів. Проведені нами дослідження дозволили спростити побудову темпоральних структур, оскільки, як показало дослідження статистичних параметрів дифтонгів, їх

Таблиця 1

Тривалості окремих ділянок ненаголошених дифтонгів для різних темпів мовлення (мс)

	Пов.	Норм.	Шв.
Звук	171	148	103
1перех.	7	7	7
1 стац.	44	41	31
2перех.	37	26	16
2 стац.	63	59	38
3перех.	20	15	11

Таблиця 2

Тривалості окремих ділянок наголошених дифтонгів для різних темпів мовлення (мс)

	Пов.	Норм.	Шв.
Звук	313	190	150
1перех.	12	10	9
1 стац.	102	66	39
2перех.	50	26	17
2 стац.	117	71	68
3перех.	32	17	17

переглядом динамічних спектрограм мовного сигналу. Виділяли наголошені та ненаголошені дифтонги та визначали статистичні параметри – середнє для 95% довірчого інтервалу з точністю до мілісекунди. Тривалості наголошених і ненаголошених дифтонгів та їх ділянок зведені в табл. 1 – 2. З таблиць побудовано залежності зміни довжин ділянок дифтонгів під наголосом та без наголосу при зміні темпу мовлення з повільного до швидкого.

Аналіз табл. 1 – 2 та рис. 1 показує, що наголошені дифтонги загалом змінюють свою довжину в 1,2 рази більше, ніж ненаголошені при прискоренні темпу мовлення. Окремі ділянки дифтонгів також ведуть себе по-різному: як в ненаголошених, так і в наголошених дифтонгах найбільших змін зазнає друга перехідна ділянка (відповідно

структури пов'язані з ознакою наголошеності. Тому функції зміни темпоральної структури будуть різними для наголошених і для ненаголошених дифтонгів. Для побудови таких функцій додатково досліджено структуру наголошених і ненаголошених дифтонгів та визначено їх тривалості та тривалості їх окремих ділянок для повільного та швидкого темпів мовлення (в мілісекундах).

Фонетично збалансований текст загальною тривалістю близько 250 секунд при частоті дискретизації 22050 Гц промовлявся чоловіком-диктором, носієм польської мови у трьох темпах: повільному, нормальному та швидкому. Різниця між повільним та швидким темпом для даного диктора становила 2.2 рази. Сегментація на ділянки, які відповідають окремим фонемам, та паузам здійснювалось вручну на основі слухового сприйняття з уточненням міток



Рис. 1. Залежності зміни довжин ділянок наголошених та ненаголошених дифтонгів при переході з повільного до швидкого темпу

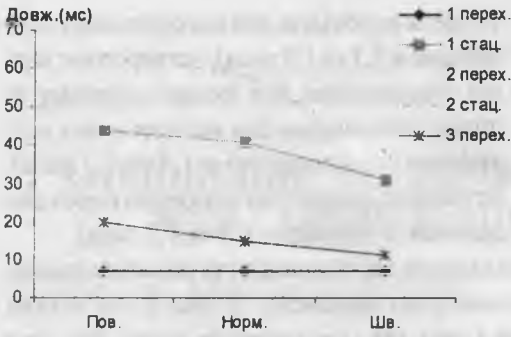


Рис. 2. Залежності зміни довжин ділянок ненаголошених дифтонгів при зміні темпу мови

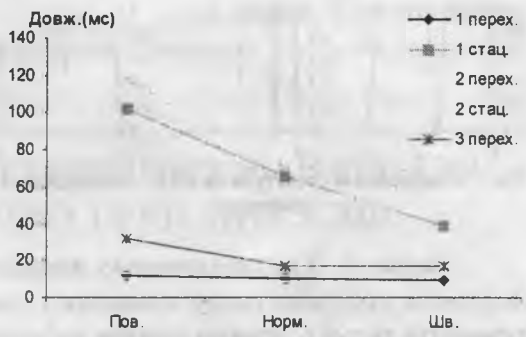


Рис.3. Залежності зміни довжин ділянок наголошених дифтонгів при зміні темпу мови

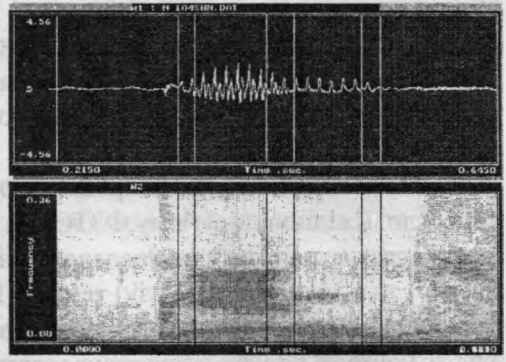
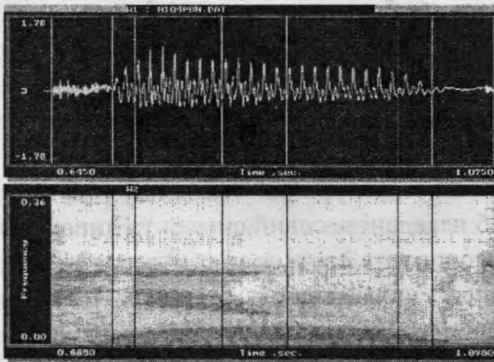


Рис. 4. Динамічні спектрограми ненаголошеного дифтонгу в повільному та швидкому темпах

в 3,2 та 5,1 раза). Далі ділянки змінюються залежно від наголошеності – для наголошених дифтонгів наступною скорочується перша стаціонарна ділянка, для ненаголошених – третя перехідна (відповідно в 2,6 та 1,8 раза). Третіми скорочуються друга стаціонарна ділянка для ненаголошених

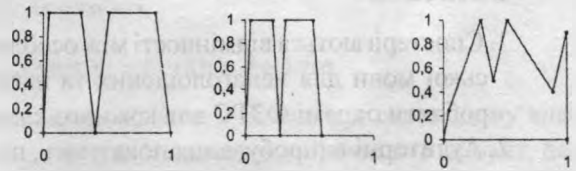


Рис.5. Темпоральні структури та ФТЗС ненаголошеного дифтонгу

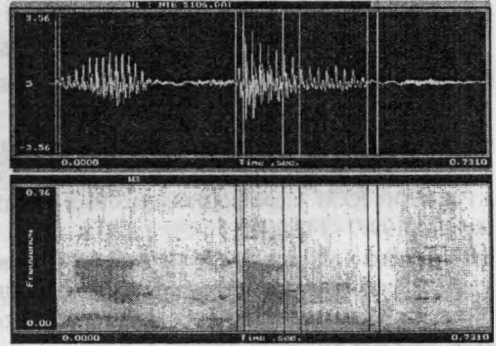
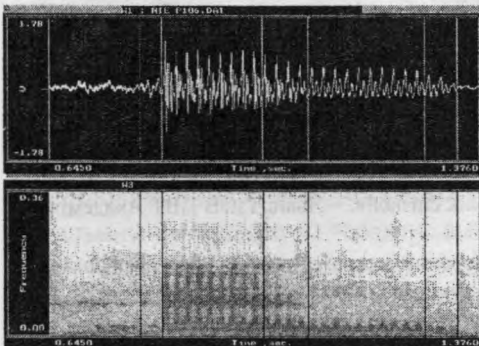


Рис. 6. Динамічні спектрограми наголошеного дифтонгу в повільному та швидкому темпах

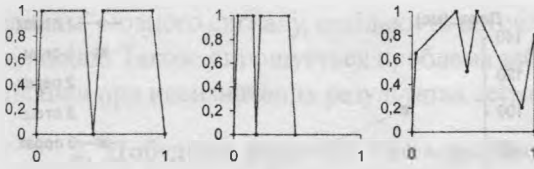


Рис.7. Темпоральні структури та ФЗТС наголошеного дифтонгу

та третя перехідна для наголошених (відповідно в 1,7 та 1,9 раза), четвертими: перша стаціонарна для ненаголошених та друга стаціонарна для наголошених всіх дифтонгів (відповідно в 1,4 та 1,7 раза), останніми скорочуються перші перехідні ділянки (відповідно в 1 та 1,3 раза).

На рис. 2 – 3 подано динаміку поведінки ділянок наголошених та ненаголошених дифтонгів при зміні темпу мовлення з повільного до швидкого. З рис. 1 – 3 можна визначити як зміну довжин ділянок дифтонгів в часі, так і пріоритет їх зміни. Рис. 4 та 5 ілюструють процес дослідження структури ненаголошеного дифтонгу та побудови відповідної йому функції зміни темпоральної структури. На рис.4 наведені часовий сигнал та динамічна спектрограма дифтонгу *e*, мовленому без наголосу в слові *napiętnować* в повільному (зверху) та в швидкому (внизу) темпах. На рис.5 зображена схематично усереднена темпоральна структура ненаголошеного дифтонгу в повільному (рис.5а) та швидкому (рис.5б) темпах, а також побудована узагальнена ФЗТС ненаголошених дифтонгів.

Рис. 6 та 7 ілюструють процес дослідження структури наголошеного дифтонгу та побудови відповідної йому ФЗТС. На рис.6 наведені часовий сигнал та динамічна спектрограма дифтонгу *e*, мовленому під наголосом в слові *mosięzny* в повільному (зверху) та в швидкому (внизу) темпах. На рис.7. зображена схематично усереднена темпоральна структура наголошеного дифтонгу в повільному (рис.7а) та швидкому (рис.7б) темпах, а також побудована узагальнена функція зміни темпоральної структури дифтонгів під наголосом (рис.7с).

Висновки

1. Спостерігаються відмінності між особливостями перетворення дифтонгів польської мови для ненаголошених та наголошених звуків. Тому необхідно розробляти окремі ФЗТС для кожного класу зокрема.
2. Аудиторні випробування показують, що запропонована в [2] технологія перетворень часової структури звуків з використанням запропонованих в даній роботі ФЗТС може бути успішно використана для прискорення темпу мовлення в діапазоні зміни коефіцієнта 1 – 2.5.
3. Подальші дослідження повинні бути зосереджені в напрямку побудови ФЗТС для інших класів звуків польської мови та вирішення проблеми зашумлення сигналу внаслідок процедури склеювання.

1. Вінцок Т.К. Интеллектуальні усномовні інформаційні технології та системи // Праці третьої всеукраїнської міжнародної конференції "Укробраз-96". – К.: ІК АН України, 1996. – С. 117 – 120.

2. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. – Львів:ТЗОВ НВТ Академічний Експрес, 1997. – 143с.

3. Шиманські З., Фігура Р., Марцишин Р. Особливості перетворення часової структури дифтонгів польської мови // Вісник Національного університету "Львівська політехніка". – 2002. – №468. – С. 170 – 174.

4. Ostaszewska D., Tambor J. Fonetyka i fonologia wspolczesnego jezyka polskiego. – W.: Wydawnictwo Naukowe PWN SA, 2000. – 142s.