

МОДИФІКАЦІЯ ЧАСОВОГО МАСШТАБУ ДИФТОНГІВ ПОЛЬСЬКОЇ МОВИ

© Фігура Р., Марцишин Р., 2007

Наведено результати дослідження зміни кепстрів дифтонгів польської мови залежно від зміни темпу мовлення. Отримано формули коректуючих вагових кепстральних функцій та розроблено метод модифікації часового масштабу польської мови на їх основі.

Results of research of change of cepstrums of diphthongs of Polish are offered depending on the change of speech rate are presented. The formulas of correcting cepstral functions are developed and the method of time-scale modification of Polish speech is developed on their basis.

Вступ

Використання систем аналізу-синтезу є одним із найпоширеніших методів перетворення часового масштабу мовних сигналів. Відомі успішні застосування систем аналізу-синтезу на основі короткочасового перетворення Фур'є та лінійного передбачення для прискорення або сповільнення відтворення мовних сигналів. У роботі [1] запропоновано технологію використання для цих цілей гомоморфних систем аналізу-синтезу мови, перевагою яких є максимальне збереження інформації про диктора, зокрема можливість оцінювання його фізичного та емоційного стану. В основу цієї технології покладено виконання модифікації кепстру мовного сигналу, отриманого на одному інтервалі часу, для синтезу сигналу на інтервалі, відмінному від початкового. Для забезпечення високої якості перетвореного сигналу алгоритми модифікації є залежними від особливостей перетворюваного сигналу, насамперед – від структури звукових одиниць, які своєю чергою, тісно пов'язані із мовою. Тому технології, опрацьовані для однієї мови, не завжди є придатними для іншої мови, особливо у випадках наявності інших недосліджених класів звуків. У цьому контексті польська мова порівняно із українською характеризується наявністю особливого класу звуків – дифтонгів (ą та ę), які містять звук η, який не є характерним для української мови і технології перетворення якого не є напрацьованими.

Метою статті є дослідження на основі експериментальних даних особливостей структури кепстрів звуку η при різних темпах мовлення з метою побудови алгоритму перетворення часового масштабу цього звуку та дифтонгу загалом.

Аналіз кепстрів звуку η при різних темпах мовлення

У відомих раніше роботах [1, 2] не було зовсім розглянуто можливості перетворення дифтонгів польської мови в системах аналізу-синтезу мови, зокрема в гомоморфних системах. Тому нами було проведено ряд експериментів для вивчення зміни кепстрів дифтонгів та особливостей їх перетворення в задачах перетворення часового масштабу мови. Для аналізу кепстрів звуку η як складової частини дифтонгу було сформовано словник із 20 слів, в яких дифтонг знаходився в однакових умовах наголошеності (під наголосом) в середині слова (наприклад, madra). При аналізі використовувалось зважування вікном Хеммінга та два темпи вимови диктора: нормальний та швидкий. При виборі дифтонгів для експериментів використовувалась класифікація фонем польської мови для задач регулювання темпу мови. Дифтонги виділялися досвідченим оператором на основі слухового сприйняття та візуального аналізу мовного сигналу з уточненням за динамічними спектрограмами. Для отримання миттєвого короткочасового кепстру фонем в

прискореному темпі приведені до середнього значення коефіцієнта прискорення 2, шляхом вирівнювання довжини фонему до довжини в двічі меншої довжини фонему в нормальному темпі. При проведенні експериментальних досліджень частота дискретизації мовного сигналу становила 12.5 кГц, частота зрізу вхідного фільтра – 5 кГц, розрядність аналого-цифрового перетворення – 16, представлення значення відліку мовного сигналу – двобайтне. Вікно аналізу становило 40 мс.

На рис. 1 зображено дифтонг [oŋ] (слово *mqdra*) та кепстр звуку *ŋ* в повільному темпі (верхня частина рисунку) та швидкому темпі (нижня частина рисунку). Дифтонг складається з 5 частинок – перша перехідна ділянка голосної частини, сама голосна частина, друга перехідна ділянка, звук *ŋ* та третя перехідна ділянка (на рисунку відокремлено вертикальними мітками). Аналіз кепстрів ділянок дифтонгу показав, що існують особливості кепстрів для звуку *ŋ*, які полягають в тому, що структура кепстру в області великого часу (після першого максимуму) змінюється як за амплітудами, так і за тривалістю. Зміна тривалостей кепстру при зміні темпу мовлення відповідає загальному підходу до зміни довжини ділянки сигналу при аналізі-синтезі мови, коли довжина вікна синтезованого сигналу відповідає довжині вікна аналізованого сигналу помноженій на коефіцієнт перетворення часового масштабу. Зміна амплітуд вимагатиме відповідно корекції кепстру для узгодження енергії. Зміну амплітуди першого коефіцієнта кепстру, який визначає енергію ділянки аналізу можна віднести до варіативності вимови диктора, оскільки під час аналізу нами не помічено жодної закономірності його зміни для зміни темпу вимови. На рисунку видно також, що період основного тону (ОТ) (відстань від нуля до максимуму в області можливого періоду ОТ) при пришвидшенні незначно зменшується (в середньому на 1–2 мс). Такі особливості кепстрів звуку *ŋ* вимагатимуть іншого методу перетворення ніж для інших звуків.

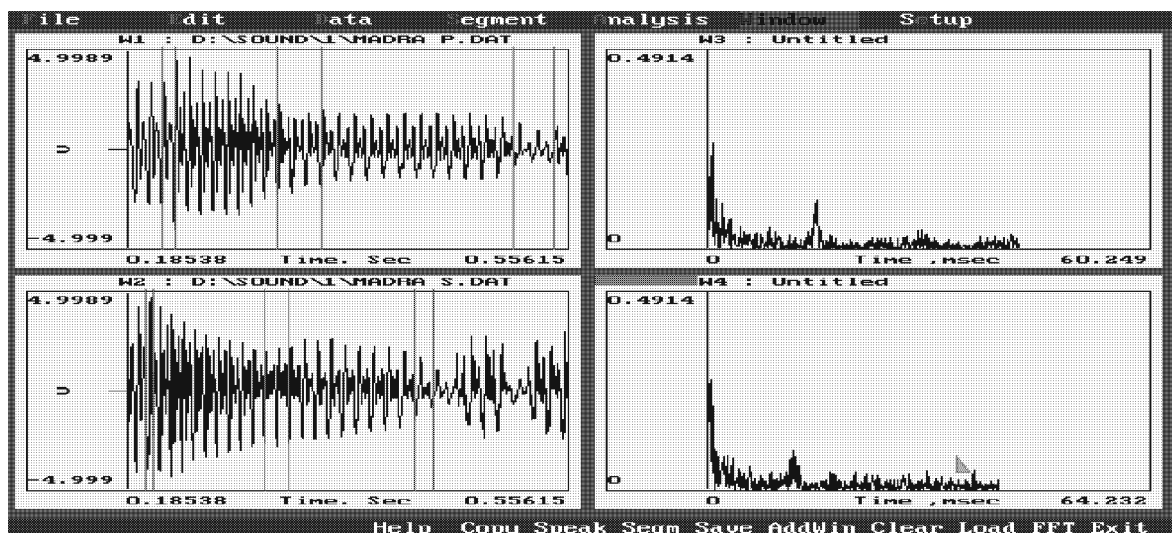


Рис. 1. Часовий сигнал та кепстри дифтонгу [oŋ] в слові *mqdra* для повільного (вверху) та швидкого (внизу) темпів вимови

З результатів досліджень можна зробити такі висновки:

для різних звуків *ŋ* із сформованого словника короткочасові кепстри, отримані з різними темпами вимови, відрізняються своєю тривалістю та незначною зміною ОТ, причому структура кепстру в області малого часу (2–40 коефіцієнти) характеризує сам звук, а зміна тривалості кепстру – зміну темпу;

зміна тривалості кепстру відбувається за рахунок зникнення відліків наприкінці області збурення кепстру із пришвидшенням темпу, або появи відліків кепстру в кінці області збурення із сповільненням;

амплітуди максимумів в області збурення кепстрів вокалізованих звуків для прискореного темпу відрізняються від аналогічних амплітуд для повільного темпу, що при розробці алгоритмів

перетворення вимагатиме використання коригуючого вікна перед синтезом для узгодження енергії синтезованих короткочасових сигналів мови;

в частині кепстру, що відповідає тракту, ні енергія, ні форма кепстру не залежать від зміни темпу, а визначаються самими звуками.

Розроблення методу модифікації кепстру дифтонгів

Як вже було сказано, для перетворення часового масштабу мови необхідно виконати модифікацію тривалості короткочасового кепстру, яка залежатиме від коефіцієнта перетворення. Для випадку прискорення з коефіцієнтом β вікно аналізу буде в β раз більшим за вікно синтезу, а для випадку сповільнення – навпаки. Наприклад, якщо при аналізі вікно аналізу мало довжину N , то при синтезі для скорочення часу звучання (випадок прискорення) вдвічі вікно синтезу буде $N/2$, та навпаки, якщо вікно аналізу N , то при синтезі для продовження часу звучання вдвічі (випадок сповільнення) вікно синтезу повинно бути $2N$.

Оскільки гомоморфні перетворення часового масштабу мови не вимагають класифікації, то для зміни темпу потрібно перетворювати тривалість кепстру залежно від коефіцієнта перетворення β та використовувати корекцію амплітуд максимумів в області збурення кепстру для вокалізованого сигналу.

Як видно з аналізу кепстрів дифтонгів для повільного та швидкого темпів, у разі прискорення права частина кепстру, яка не потрапляє до вікна синтезу, просто відкидається. У випадку сповільнення вікно синтезу буде більше за вікна аналізу, тому в кепстрі з'явиться пусте місце, яке необхідно заповнити. Це можна здійснити шляхом повтору ділянки кепстру з області збурення. При цьому виникає питання вибору точки спряження ділянок кепстру.

З проведеного вище аналізу випливає, що в обох випадках необхідно використовувати коригуючу кепстральну функцію для врахування зміни енергії кепстру у разі зміни темпу, а також для зменшення спотворень при спряженні ділянок кепстру у разі сповільнення.

Як точки спряження в вокалізованих звуках доцільно використовувати максимуми в області збурення кепстру, кратні періоду ОТ, а в невокалізованих звуках точки в області збурення кратні періоду ОТ попередньої вокалізованої ділянки аналізу. Це дасть змогу мінімізувати спотворення синтезованого сигналу та спростити алгоритм модифікації завдяки повторенню при сповільненні ділянки кепстру між передостаннім та останнім максимумами в області збурення для вокалізованого сигналу. Для невокалізованого мовного сигналу приймають, що точки розташування передостаннього та останнього максимумів відповідають розташуванню максимумів у кепстрі попередньої ділянки вокалізованого сигналу.

Для визначення виду кепстрального вікна досліджувався дифтонг [оη], вимовлений в швидкому та сповільненому темпах. Оскільки коефіцієнт сповільнення $\beta=0.5$, вибрано вікна аналізу $N=512$ та $N=1024$ для отримання достатньої кількості максимумів в області ОТ.

Для отримання коректуючої кепстральної функції виділено характерні точки обвідних кепстру (див. табл. 1, 2) та використано стандартні апроксимаційні функції Microsoft Excel для побудови апроксимованих (згладжених) кривих. Для вибору однієї з отриманих кривих обчислювали величину достовірності апроксимації R^2 .

Таблиця 1

Характерні точки обвідної кепстру для $N=512$

t, мс	0	0.05	0.12	0.29	6	12	19.7
Ампл.	4.61	0.12	0.02	0.023	0.24	0.07	-0.01

Таблиця 2

Характерні точки обвідної кепстру для $N=1024$

t, мс	0	0.05	0.12	0.29	6	12	23.78	39
Амп.	4.77	0.14	0.04	0.015	0.25	0.1	0.06	0

Проведені дослідження показали, що у разі використання всіх точок обвідної всі апроксимовані криві не відповідали реальній формі обвідної та мали низькі значення величини достовірності ($R^2 < 0.6$). В зв'язку з цим побудовано апроксимаційні криві тільки для характерних точок обвідної, що належать області збурення кепстру. Для них знову використовувалось обчислення величини достовірності. В результаті експериментів найвищу величину достовірності $R^2 > 0.99$ отримали поліноміальні апроксимаційні криві, показані на рис. 2, де верхня крива ілюструє сповільнення, а нижня – прискорення. На рис. 2 показано також формули, використовувані при побудові коригуючої кепстральної функції для модифікації кепстру у разі перетворення часової структури мови.

Отже, враховуючи проведені експерименти, коригуюча вагова кепстральна функція для прискорення матиме вигляд:

$$l(q) = \begin{cases} 1, & \text{для } q = \overline{0, N_1 - 1} \\ ax^2 + bx + c, & \text{для } q = \overline{N_1, N - 1} \end{cases}, \quad (1)$$

де N_1 – дорівнює періоду ОТ (обчислюється через визначення номеру відліку, де знаходиться перший максимум кепстру в області збурення),

a, b, c – експериментально визначені коефіцієнти, що залежать від коефіцієнта перетворення часового масштабу β (для $\beta=0.5$ коефіцієнти $a=-2 \cdot 10^{-5}$, $b=0.0029$, $c=0.9396$)

$$l(q) = \begin{cases} 1, & \text{для } q = \overline{0, N_1 - 1} \\ ax^3 + bx^2 + cx + d, & \text{для } q = \overline{N_1, N - 1} \end{cases} \quad (2)$$

де N_1 – дорівнює періоду ОТ, a, b, c, d – експериментально визначені коефіцієнти, що залежать від коефіцієнта перетворення часового масштабу β (для $\beta=2$ коефіцієнти $a = -7 \cdot 10^{-7}$, $b = 0.0002$, $c = 0.0082$, $d = 1.0345$).

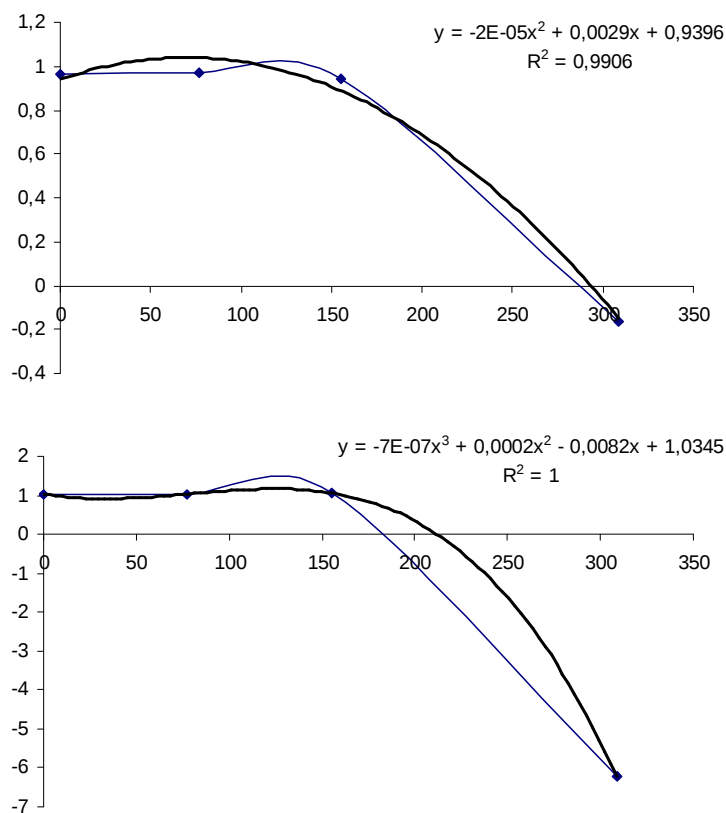


Рис. 2. Коригуючі кепстральні функції для прискорення (нагорі) та сповільнення (внизу)

Використовуючи вищеописане, пропонується такий метод модифікації кепстру для дифтонгу: прискорення ($N_a > N_s$):

- $N_s > 3$ періоди ОТ;
- Знаходження розташування першого максимуму N_1 в області збурення, що перевищує заданий поріг (дозволяє встановити факт вокалізованості сигналу та період основного тону).
- згладження ваговою кепстральною функцією кепстру $X(mS_a, q) = l(q) X(mS_a, q)$, де $l(q)$ – обчислюється за формулою (1).
- Копіювання ділянки згладженого $X(mS_a, q)$ включно від позиції максимуму N_1 до кінця N_a та зсув вліво на 1 мс.
- Відкидання $N_a - N_s$ відліків кепстру $X(mS_a, q)$ з області збурення до досягнення відповідності вікну N_s , тобто $Y(mS_s, q) = X(mS_a, q)$, де $q = 1, \dots, N_s$.

сповільнення ($N_a < N_s$):

- $N_a > 3$ періоди ОТ;
- Знаходження розташування першого максимуму N_1 в області збурення, що перевищує заданий поріг (дозволяє встановити факт вокалізованості сигналу та період основного тону).
- згладження ваговою кепстральною функцією кепстру $X(mS_a, q) = l(q) X(mS_a, q)$, де $l(q)$ обчислюється за формулою (2).
- Копіювання ділянки згладженого $X(mS_a, q)$ включно від позиції максимуму N_1 до кінця N_a та зсув вправо на 1 мс.
- копіюємо кепстр $X(mS_a, q)$ в $Y(mS_s, q)$ та, використовуючи факт повтору решти максимумів в кепстрі з періодом ОТ, доповнюємо кепстр $Y(mS_s, q)$ до досягнення відповідності вікну N_s шляхом вставляння ділянки між передостаннім та останнім максимумами області збурення $X(mS_a, q)$, де $q = N_{k-1}, \dots, N_k$, а $k = \text{int}(N_a / N_1)$, з позиції розташування останнього максимуму N_k . У випадку наявності одного максимуму в області збурення $k=1$ (величина вікна менша за чотири періоди ОТ) кепстр $Y(mS_s, q)$ доповнюється вставленням ділянки $X(mS_a, q)$ включно від позиції максимуму N_1 до кінця N_a синхронно з ОТ (утворюються періодичні максимуми, кратні періоду ОТ) в $Y(mS_s, q)$, де $q = mN_1, \dots, (m+1)N_1$, а $m = 1, \dots, M-1$, де відповідно $M = \text{int}(N_s / N_1)$, та заповнення пустих місць копіюванням попередніх відліків кепстру.

Висновки

Показано, що для дифтонгів польської мови короткочасові кепстри дифтонгів для різних темпів мовлення відрізняються своєю тривалістю, тобто перетворення часового масштабу мови відбувається за рахунок відкидання-додавання відліків кепстру в області параметрів збурення без зміни форми кепстру в області параметрів тракту мовотворення. Енергія кепстрів дифтонгів в області параметрів збурення для різних темпів та період ОТ відрізняються, тому під час розроблення алгоритмів модифікації необхідно використовувати коригуючу вагову кепстральну функцію для узгодження енергії синтезованих короткочасових ділянок мовного сигналу та корекції ОТ. Як точки спряження при модифікації кепстру доцільно використовувати місцерозташування максимумів в області збурення кепстру, кратних періоду ОТ. Метод перетворення часового масштабу кепстру може бути використаний в гомоморфних системах аналізу-синтезу мови при високоякісному нелінійному перетворенні часового масштабу без класифікації фонем.

1. Марцишин Р.С. Модифікація часового масштабу мови через кепстральний аналіз-синтез // Вісн. Держ. ун-ту “Львівська політехніка”. – 1998. – № 351. – С. 190–196. 2. Шиманські З., Фігура Р., Марцишин Р. Особливості перетворення часової структури дифтонгів польської мови // Вісн. Нац. ун-ту “Львівська політехніка”. – 2002. – № 468. – С. 170–174.