

ВИДІЛЕННЯ ПІДКЛАСІВ НЕВОКАЛІЗОВАНИХ ПРИГОЛОСНИХ У ЗАДАЧАХ СПОВІЛЬНЕННЯ ТЕМПУ МОВЛЕННЯ

© Купчак М., 2010

Знайдено новий інформативний параметр мовного сигналу для автоматизованої класифікації підкласів невокалізованих приголосних.

Ключові слова: невокалізовані приголосні, аффрикати, щілинні, короткочасна енергія.

The new informing parameter of linguistic signal is found for the automated classification of the unvoiced consonants subclasses.

Keywords: unvoiced consonants, affricates, whispers, Short-Time Energy.

Вступ

Адаптивні технології перетворення тривалостей звуків із використанням методів трансформації структури звукових одиниць на основі функцій темпоральних перетворень (ФТП) вимагають розв'язання задачі класифікації звукових одиниць мовленнєвого сигналу. Це дає можливість забезпечити високу натуральність відтворення мовних сигналів при сповільненому темпі мовлення, що необхідне при вивченні іноземних мов, в задачах логопедії, судовокриміналістичній експертизі тощо.

На основі статистичного аналізу часової структури звукових одиниць української мови при різних темпах мовлення обгрунтовано [1], що для задач сповільнення темпу недостатньо виділення та маркірування основних 5 класів звуків (голосні під наголосом, ненаголошені голосні, вокалізовані приголосні, невокалізовані приголосні та вибухові звуки). Для забезпечення максимальної натуральності перетвореного сигналу доцільним є поділ класу невокалізованих приголосних на додаткові підкласи: щілинних (ф,с,ш,с',х,з,ж,з',г) та аффрикат (ц,ч,ц', дз, дж, дз') [2].

Оскільки близькі значення числових характеристик, подібні акустичні властивості та однакова трансформація внутрішньої структури щілинних та аффрикат при зміні темпу не дають змоги їх надійно поділити на підкласи, виникає необхідність визначення додаткових інформативних ознак мовного сигналу.

Основною метою статті є знаходження нових інформативних параметрів мовного сигналу для автоматизованої класифікації підкласів невокалізованих приголосних.

Основна частина

Розбиття сигналу на частини є однією із найважливіших задач аналізу розпізнавання та перетворення мовних сигналів. У контексті технологій адаптивного перетворення часової структури мовних сигналів метою такої задачі є поділ сигналу на ділянки, що відповідають окремим звукам (сегментація), та віднесення цього звуку до одного із наперед виділених класів (маркірування).

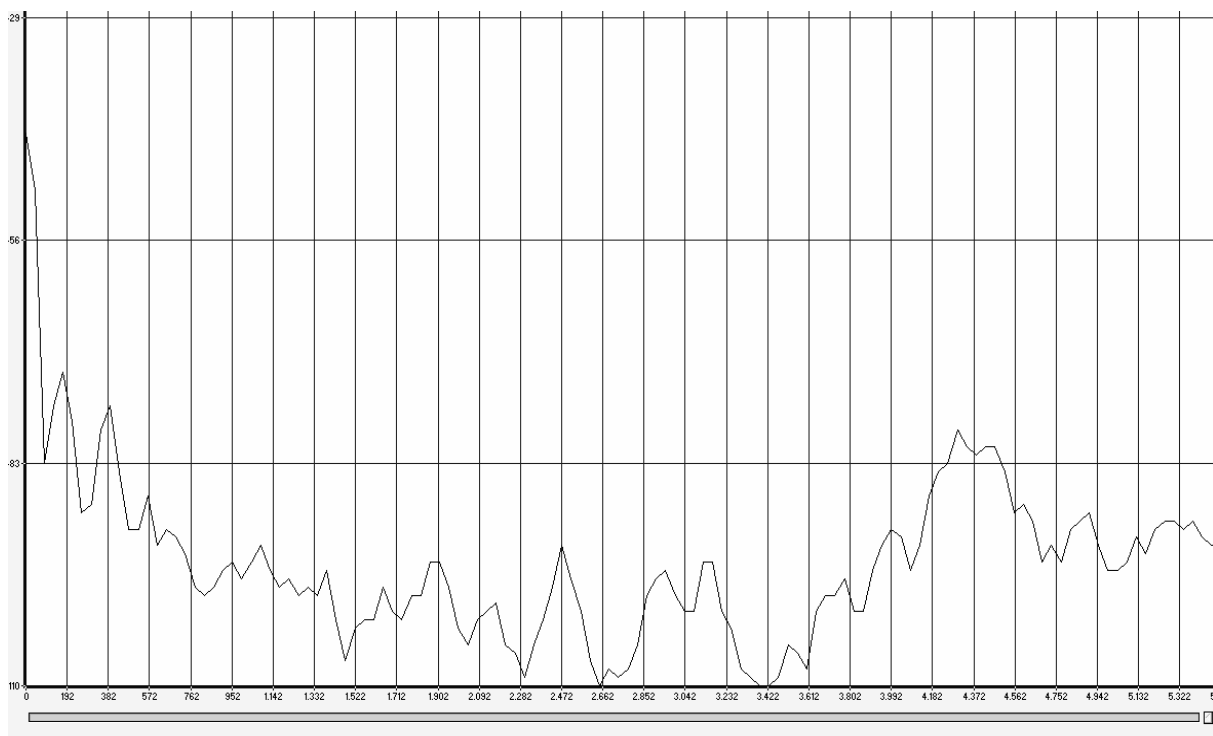
Виділення невокалізованих ділянок в даній роботі здійснювалось на основі запропонованого в [1] алгоритму із використанням значення першого коефіцієнта фільтру моделі авторегресії A_1 , ознаки вокалізованості O_v та короткочасної енергії E .

Невокалізованими вважали всі ділянки, для яких $A_1 > 0,8$; для сегментів з $-0,8 > A_1 > -1,5$ обчислювалася додаткова ознака вокалізованості $O_v = 0,03E - A_1$, і при $O_v < 2,2$ ділянка також вважалася невокалізованою.

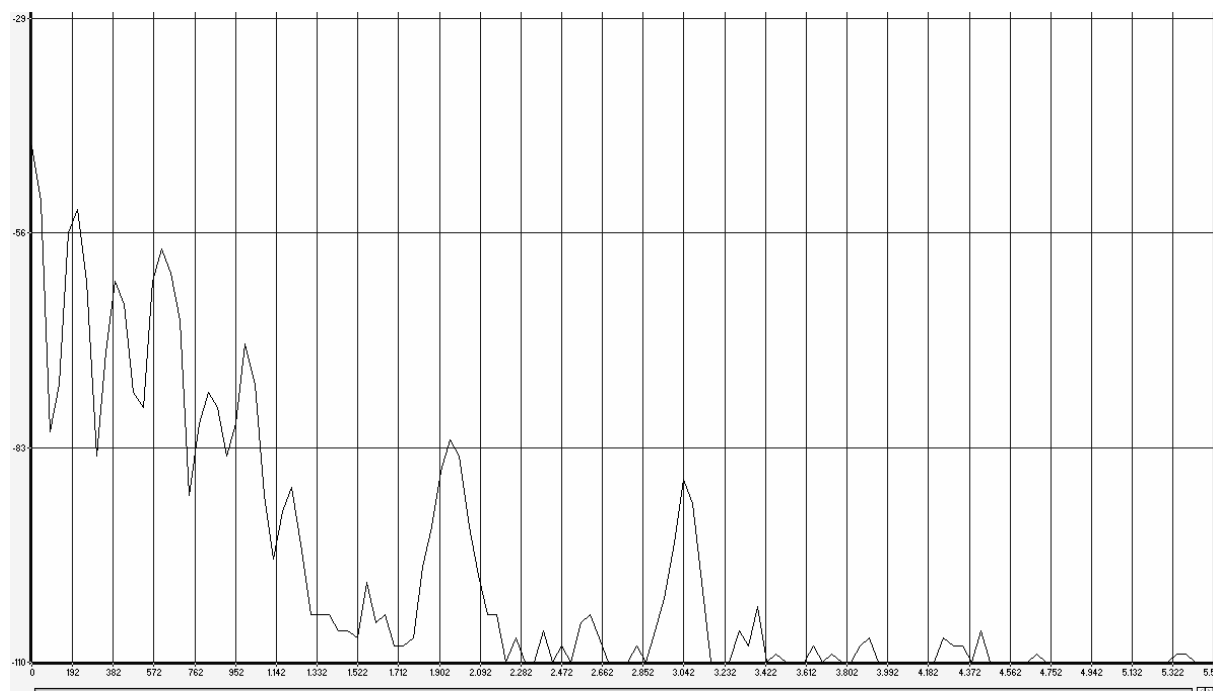
Наведені інформативні ознаки дали можливість надійно виділити невокалізовані приголосні. Але виділення підкласів щілинних та аффрикат вимагає знаходження нових інформативних ознак мовного сигналу.

Відомо, що ефективним джерелом інформації, яку можна використати для побудови нових інформативних ознак, є короточасний спектр мовного сигналу. Різні класи звуків відрізняються між собою як різними рівнями енергії в певних діагонах частот, так і різними співвідношеннями енергій спектра у таких діапазонах.

На рисунку наведено приклади короточасних спектрограм невокалізованих приголосних: спектрограми щільної *с* у слові *Свобода* (рисунок, а) та спектрограми афrikати *ц* у слові *Цитадель* (рисунок, б).



а) щільного звуку *с*



б) афrikату *ц*

Короточасові спектри приголосних

Порівняння цих спектрограм показує, що істотна відмінність між ними полягає, окрім іншого, в значенні енергії на високих частотах (3,5 кГц і вище): для африкат середня енергія в діапазоні високих частот є невеликою, тоді як щілинні як раз і характеризуються високою концентрацією енергії на вказаних частотах.

Отже, достатньо обгрунтованим і логічним є формування нової інформативної ознаки для класифікації щілинних та африкат на основі відношення енергій короткочасного спектра в певним чином виділених діапазонах частот.

У проведених експериментах досліджувалися наступні інформативні ознаки: відношення $E1/E2$ середньої короткочасної енергії у діапазоні 0–1500 Гц до енергії в діапазоні 3500–5500 Гц, відношення короткочасної енергії в діапазоні 1500–3500 Гц до енергії в діапазоні 3500–5500 Гц, відношення короткочасної енергії в діапазоні 0–5500 Гц до енергії в діапазоні 3500–5500 Гц. Діапазон 3500–5500 Гц для оцінювання енергії $E2$ був вибраний, враховуючи, що високе значення енергії спектру в даному діапазоні є однією із основних характеристик щілинних звуків. Результати експериментів показали, що найбільш інформативною є перша ознака (відношення короткочасної енергії в діапазоні 0–1500 Гц до енергії в діапазоні 3500–5500 Гц) при порозі 1,5, оскільки із 20 випадків у 19, що становить 95%, невокалізована щілинна приголосна була класифікована як щілинна, а у 18 експериментах із 20 (90%) африката також класифікована правильно. Для інших ознак відповідні значення є такими: для другої ознаки – 75% і 80%, для третьої – 90% і 90%. Очевидно, що також найкращими у першому випадку є і результати помилкової класифікації: 7,5% для першої ознаки, 22,5% – для другої та 10% для третьої інформативної ознаки.

Висновки

Результати проведених досліджень та натурних експериментів показали, що введення додаткового інформативного параметра – відношення короткочасної енергії в діапазоні 0–1500 Гц до енергії в діапазоні 3500–5500 Гц дає можливість ефективно та надійно визначити щілинні та африкати в класі невокалізованих приголосних, що є необхідною умовою для реалізації високоефективних технологій сповільнення темпу відтворення мовних сигналів.

1. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. – Львів: Академічний експрес, 1997. – 140 с. 2. Рашкевич Ю.М., Шпак З.Я., Купчак М.І. Автоматична сегментація та маркірування звуків для задач сповільнення темпу відтворення мовних повідомлень // Інтелектуальні системи прийняття рішень та проблеми обчислюваного інтелекту. – Херсон: ХНТУ, 2009. – Т.2. – С. 417–418.