

Обґрунтування критерію для автоматизованого розпізнавання шиплячих звуків української мови дикторнезалежними системами

Джичка Н.І., Шадріна Г.М.

Кафедра біотехнічних систем
Тернопільський державний технічний університет імені Івана Пулюя
magnanimous@ukr.net

Abstract

This article is devoted to the problem of phoneme recognition of speech. The informing signs of sibilants of Ukrainian language for recognition are chosen. The criterion for recognition of sibilants on the basis of the chosen informing signs is grounded.

Вступ

Актуальність розроблення систем розпізнавання мови полягає у спрощенні процесу набору текстів, особливо для користувачів із вадами зору, можливості перекладу фраз з однієї мови на іншу в реальному режимі часу, можливості управління технікою за допомогою голосових команд.

Більшість таких систем використовують алгоритми, на базі шаблонів слів або звуків. Оскільки голос кожного диктора має характерні особливості невластиві для інших, то такі системи потребують створення наборів шаблонів для кожного диктора окремо. Розпізнавання здійснюють шляхом порівняння двох сигналів – вимовленого та шаблону. У такому випадку багато часу займає і створення шаблонів, і процедура порівняння.

Для спрощення процесу розпізнавання мови доцільно було б використовувати шаблони окремих звуків єдині для всіх дикторів. На даний час таких шаблонів не існує через те, що не виявлено інформативних ознак звуків, які не залежать від характерних особливостей голосу.

Тому для реалізації ефективних дикторнезалежних систем автоматизованого розпізнавання мови необхідно виділити інформативні ознаки звуків мови, розробити математичні методи їх опрацювання з метою створення єдиних для всіх дикторів шаблонів.

У такому випадку система розпізнавання не буде потребувати навчання (створення набору шаблонів окремо для кожного диктора), її швидкодія збільшиться, оскільки відпаде необхідність створення набору шаблонів слів і з'явиться можливість розпізнавати мову незалежно від характерних особливостей голосу диктора.

Найкраще піддаються виділенню та розпізнаванню голосні [а], [о], [у], [е], [и], [і] та сонорні [р], [л], [м], [н], [й], [в], оскільки вони наближаються за тоном звучання до голосних. Інші групи звуків виявити і розпізнати важко через присутність шумів при їх вимові.

Проте, шиплячі звуки [ж], [ч], [ш] та [дж], що відносяться до шумних приголосних, відрізняються від інших відсутністю коливань голосових зв'язок і при їх творенні не проявляються особливості голосу конкретного диктора, тобто, для них можливо створити

єдиний шаблон. З цією метою з мовного сигналу потрібно виділити ділянки шумів, які відповідають шиплячим звукам, виконати певні перетворення для відшукування інформативних ознак цих звуків, створити шаблони та розробити критерій для їх автоматизованого розпізнавання.

Вибір інформативних ознак шиплячих звуків

Інформативні ознаки шиплячих звуків повинні бути інваріантними до часу і містити в собі інформацію про відповідний звук, достатню для розпізнавання [1].

Однією з ознак шиплячих звуків, яка задовольняє вимогу інваріантності до часу і містить інформацію про звук, є набір формант та антиформант у спектрі [2].

З метою створення шаблонів шиплячих звуків було отримано амплітудні спектри звуків [ж], [ч], [ш] та [дж] (див. рис. 1-4). Для зручності проведення досліджень масштабуємо їх до діапазону від 0 мВ/Гц до 1мВ/Гц

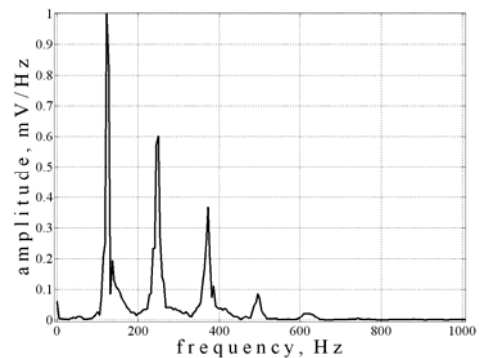


Рисунок 1. Амплітудний спектр звуку [ш].

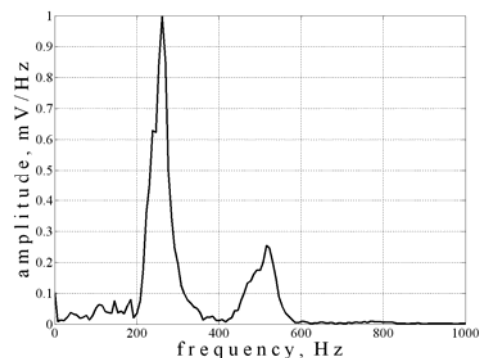


Рисунок 2. Амплітудний спектр звуку [ж].

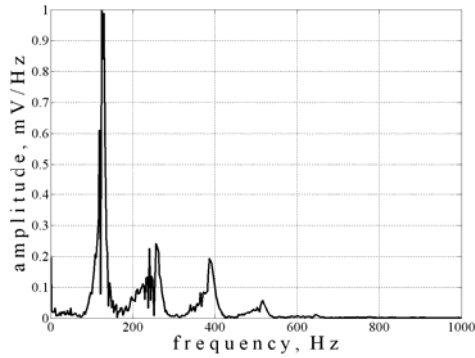


Рисунок 3. Амплітудний спектр звуку [дж].

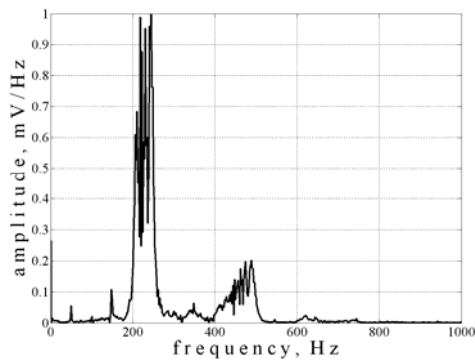


Рисунок 4. Амплітудний спектр звуку [ч].

Як видно з рисунків, у амплітудних спектрах шиплячих звуків проявляється три форманти та дві антиформанти, які належать до одного і того ж вузького діапазону частот [1]. Оскільки, для жіночих голосів характерні вищі частоти ніж для чоловічих, тому перша та друга форманти жіночого голосу можуть накладатися одна на одну.

Враховуючи це, побудуємо шаблон в частотній області для виокремлення шиплячих звуків з інших звуків мови, показаний на рисунку 5. У діапазоні частот від 650 Гц до 22 кГц амплітуда не перевищує 0,1 мВ/Гц, тому цей частотний діапазон показаний частково.

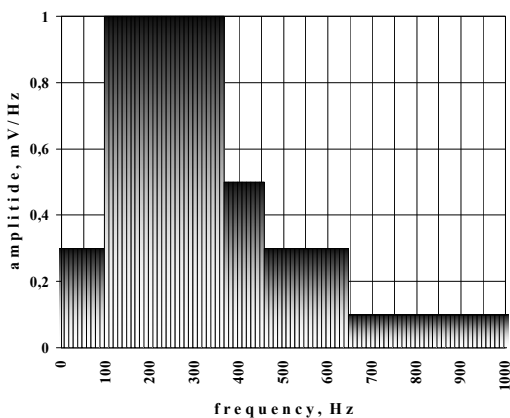


Рисунок 5. Шаблон амплітудного спектру шиплячих звуків.

На рисунку чітко видно п'ять частотних діапазонів, які відрізняються за амплітудою.

Для створення критерію розпізнавання введемо наступні позначення: i – номер частотного діапазону, Δv_i – частоти, які потрапляють у цей діапазон; A_i – максимальна амплітуда i -го частотного діапазону.

Опис набору формант та антиформант, для всіх шиплячих звуків, наведено в таблиці 1.

Таблиця 1: Інформативні ознаки звуків [ж], [ч], [ш] та [дж] у частотній області

i	Δv_i , Гц	A_i , мВ/Гц	Інформативна ознака
1	1÷100	$A_1 < 0.3$	антиформанта
2	100÷360	$0.1 < A_2 < 1$	форманта
3	360÷450	$A_3 < 0.5$	форманта
4	450÷650	$A_4 < 0.3$	форманта
5	650÷900	$A_5 < 0.1$	антиформанта

Позначимо через v_i , $i = \overline{1,5}$ параметри, які визначають присутність формант та антиформант у діапазоні частот Δv_i , v_i можуть набувати значень 1 або 0 (1 – в частотному спектрі присутня форманта або антиформанта, 0 – відсутня).

$v_i = 1$, якщо максимальна амплітуда i -го діапазону відповідає значенню, наведеному в таблиці 1.

$v_i = 0$, якщо максимальна амплітуда i -го діапазону не відповідає табличному значенню.

Параметрів v_1, v_2, v_3, v_4, v_5 , достатньо щоб повністю описати амплітудні спектри шиплячих звуків, проте недостатньо для того, щоб розрізнити звуки між собою.

Для розрізнення шиплячих звуків [ж], [ч], [ш] та [дж] використаємо їх динамічні портрети в часовій області. Особливістю шиплячих звуків є те, що вони не піддаються коартикуляції, тому їх динамічні портрети можна розпізнати у будь-якому слові [3].

Динамічний портрет звуку в часовій області – це обвідна реєстрограми звукового сигналу. Оскільки реєстрограми мовних сигналів симетричні відносно осі абсцис (див. рис 6), тому для досліджень обмежимося верхньою обвідною кривою.

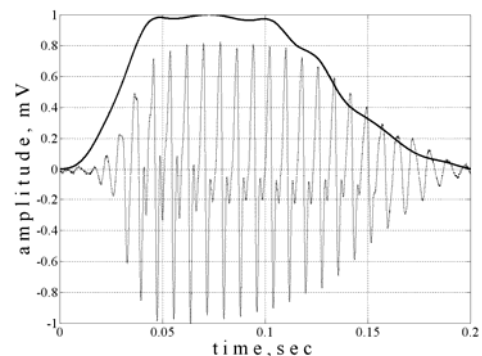


Рисунок 6. Реєстрограма звуку [ш] та його динамічний портрет.

Динамічні портрети шиплячих звуків будують таким чином:

1) масштабуємо амплітуду зареєстрованого звуку до діапазону від -1 мВ до 1 мВ, оскільки динамічний портрет не залежить від гучності вимови;

2) розділяємо проміжок, на якому зареєстровано звук, на 20 однакових часових діапазонів;

3) знаходимо максимальне значення амплітуди для кожного з діапазонів;

4) приймаємо значення амплітуд 1-го та 21-го діапазону за нуль;

5) за отриманими максимальними значеннями амплітуд будують динамічний портрет звуку.

У результаті експерименту було побудовано динамічні портрети шиплячих звуків для семи дикторів (чотири зразки жіночих голосів та три зразки чоловічих голосів). На рисунках 7-10 показано динамічні портрети кожного звуку для всіх дикторів (пунктирною лінією позначено динамічні портрети для жіночих голосів, а суцільною лінією – для чоловічих).

Як видно з рисунків форми обвідних кривих для одного і того ж звуку для жіночих голосів схожі між собою, як і форми обвідних кривих для чоловічих голосів, проте чітко видно різницю між динамічними портретами звуків для жіночих та чоловічих голосів.

Отже, динамічні портрети шиплячих звуків, як інформативні ознаки, є інваріантними до диктора одної статі.

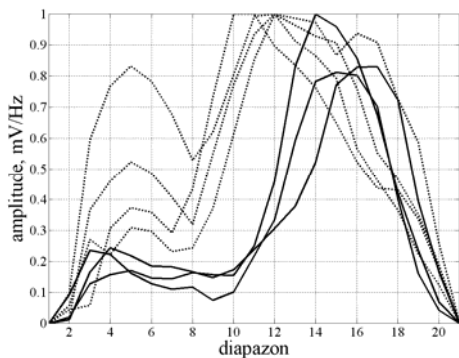


Рисунок 7. Динамічні портрети звуку [ж]

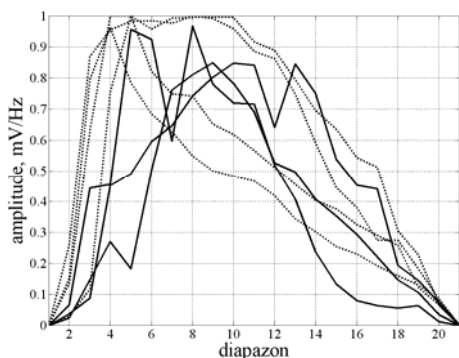


Рисунок 8. Динамічні портрети звуку [ч]

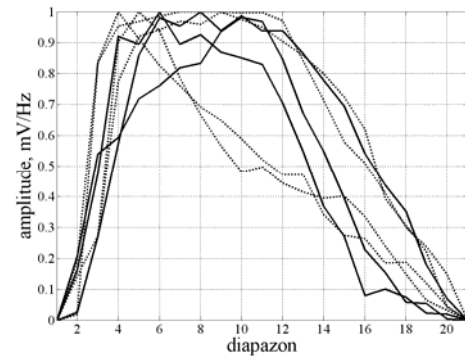


Рисунок 9. Динамічні портрети звуку [ш]

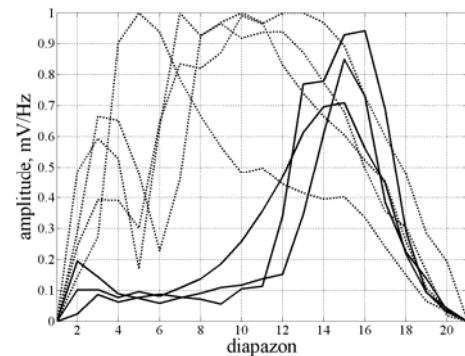


Рисунок 10. Динамічні портрети звуку [дж]

Тому, на їх основі будують шаблони для розпізнавання окремого звуку для жінок та чоловіків.

Шаблони динамічних портретів шиплячих звуків для жіночого та чоловічого типів голосів подано на рисунках 11-14.

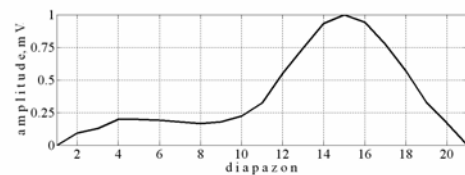
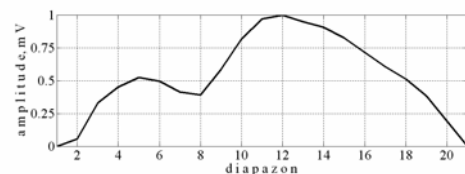


Рисунок 11. Шаблони динамічних портретів звуку [ж] для жіночого голосу (вверху) та для чоловічого голосу (внизу).

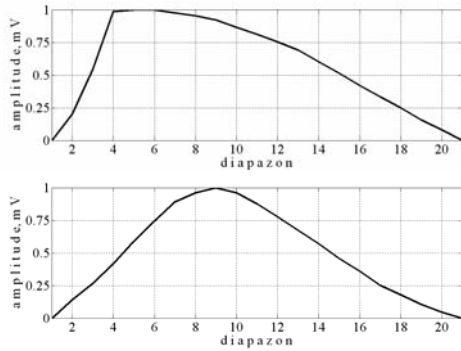


Рисунок 12. Шаблиони динамічних портретів звуку [ч] для жіночого голосу (вгорі) та для чоловічого голосу (внизу).

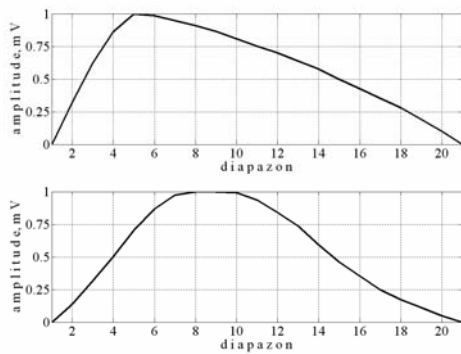


Рисунок 13. Шаблиони динамічних портретів звуку [ш] для жіночого голосу (вверху) та для чоловічого голосу (внизу).

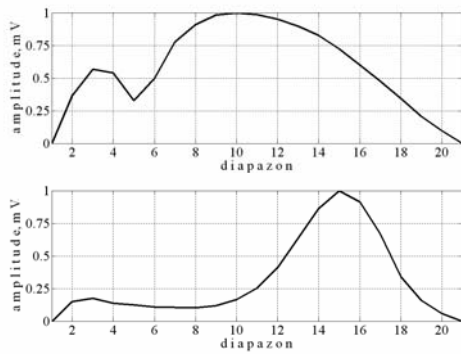


Рисунок 14. Шаблиони динамічних портретів звуку [дж] для жіночого голосу (вверху) та для чоловічого голосу (внизу).

Обґрунтування критерію для розпізнавання шиплячих звуків

Використовуючи побудовані шаблиони у частотній (набір формант та антиформант) та часовій областях (динамічні портрети) будуємо критерій для розпізнавання шиплячих звуків серед інших звуків:

$$Y = v_1 \cdot v_2 \cdot v_3 \cdot v_4 \cdot v_5, \quad (1)$$

де, v_1, v_2, v_3, v_4, v_5 - параметри, які визначають присутність формант та антиформант у діапазоні частот Δv_i .

Умова належності звуку до групи шиплячих:

$$\begin{cases} Y = 1, \dots \text{звук} \dots \text{належить} \dots \text{до} \dots \text{групи} \\ Y = 0, \dots \text{звук} \dots \text{не} \dots \text{належить} \dots \text{до} \dots \text{групи} \end{cases} \quad (2)$$

Критерієм розрізнення шиплячих звуків між собою є суміщення динамічного портрету шиплячого звуку та відповідного шаблону динамічного портрету цього ж звуку. як показано на рисунку 14 (суцільною лінією позначено динамічний портрет звуку [ш], пунктиром – шаблон).

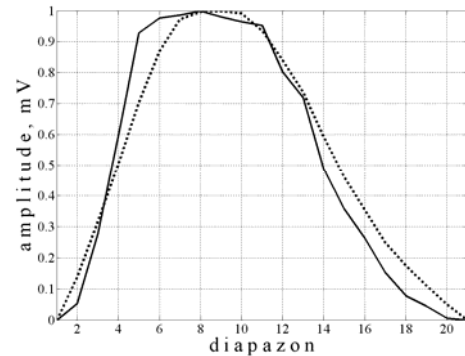


Рисунок 14. Динамічні портрети шаблону та звуку [ш].

Очевидно, що шаблон динамічного портрету звуку [ш] для чоловічого голосу та побудований динамічний портрет звуку [ш] схожі між собою. Хоча цей критерій має суб'єктивний характер, він може бути придатним для застосування в логопедії, зокрема в реабілітаційних системах та системах тренування вимови у людей із сигматизмом, коли пацієнт може візуально оцінити правильність своєї вимови.

Висновки

Запропонований критерій може бути використаний для створення дикторнезалежних автоматизованих систем розпізнавання розмовної української мови та систем для корекції вимови шиплячих звуків при сигматизмі.

Література

1. Джичка Н.І. Розробка математичної моделі шиплячих приголосних звуків для диктор – незалежних систем розпізнавання мовних сигналів. // Тези до XI студентської конференції «Гуманітарні науки. Актуальні питання». – ТДТУ: 2008. – с. 66.
2. Джеймс Л. Фланаган Анализ, синтез и восприятие речи перевод с английского под редакцией А.А.Пирогова издательство "Связь" Москва-1968
3. Бондаренко М.Ф., Чугун А.І., Николаенко В.Л. Анализ речевых сигналов по их динамическим портретам. Речевая информатика: Сб. науч. тр. / АН УССР. Ин-т кибернетики им. В.М.Глушкова; Редкол.: Винцюк Т.К. и др. – Киев, 1989. – 159с.