ІДЕНТИФІКАЦІЯ ПЕРЕДАТОЧНИХ ХАРАКТЕРИСТИК АКУСТИЧНОГО КАНАЛУ В СИСТЕМАХ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ

Биков М.М., Кузьмін І.В., Грищук Т.В., Ковтун В.В.
Вінницький державний технічний університет

Задачею системи автоматичного розпізнавання мови є визначення змісту вимовленого повідомлення. Однак інформаційна структура мовного сигналу, що підлягає розпізнаванню, окрім змісту повідомлення містить ще й такі складові, як характеристики індивідуальності диктора і характеристики емоціонального та фізичного стану диктора. При цьому одна і та ж складова інформаційної структури модулює не один, а декілька параметрів мовного сигналу, таких, як його інтенсивність, амплітуди і фази спектральних складових, частоту основного тону [1]. Семантичний зміст мови кодується зміною в часі параметрів джерела звукових коливань і форми миттєвих спектрів мовного сигналу. Інформація про індивідуальні особливості диктора кодується частотою основного тону і формою миттєвих спектрів. Фізичний і емоціональний стан диктора кодуються інтенсивністю сигналу, формою короткочасного спектра, частотою основного тону, тривалістю висловлення. Це вимагає розробки таких моделей утворення, передавання і сприйняття мови, які б дозволили описувати мовні образи інваріантними до диктора та завад і коваріантними до семантичного змісту ознаками. Результати проведених в роботі досліджень каналу передачі мови показують, що для одного і того ж повідомлення інформативні параметри мовного сигналу значно змінюються. Зокрема, ці зміни відтворюють різний вплив акустичного каналу на розповсюдження звуків різної фонетичної якості – вокалізовані звуки мають значно менший ступінь затухання, ніж фрикативні, а інтенсивність сигналу в паузі змінюється з відстанню незначно. Подалі в роботі наводяться математичні викладки, які дозволили встановити передаточну функцію акустичного каналу для вокалізованих звуків у вигляді обернено пропорційної залежності від відстані, для пауз – у вигляді одиничної. Для фрикативних звуків (таких, як “ф”, “ш”, “х”, “с” та ін.) не існує явної математичної залежності величини їх звукового тиску від відстані. Тому шляхом проведення експери-

ментальних вимірювань для них була проведена ідентифікація передаточної характеристики акустичного каналу у вигляді показникової функції за методом найменших квадратів.

На рис.1 наведені записи деяких параметрів мовного сигналу, які характеризують особливості його розповсюдження в акустичному каналі.

На рис.1а) і б) зображена інтенсивність мовного сигналу для слова "каша", вимовленого диктором на відстані 0 і 5 см відповідно, а на рис.1в) і г) – ті ж параметри для слова "кафе". Ці рисунки показують різний вплив передавального середовища на розповсюдження звуків різних фонетичних категорій – вокалізовані звуки мають значно менший ступінь затухання, чим шумні, а інтенсивність сигналу в паузі змінюється з відстанню незначно. Оскільки параметри передавального середовища інваріантні до багатьох видів інформації, і коваріантні до фонетичних характеристик звуків, що пов´язані з семантичною інформацією, то представляє інтерес розробка математичних співвідношень, які визначають залежність передаточної характеристики акустичного середовища від фонетичної якості звуків мови, що передаються.

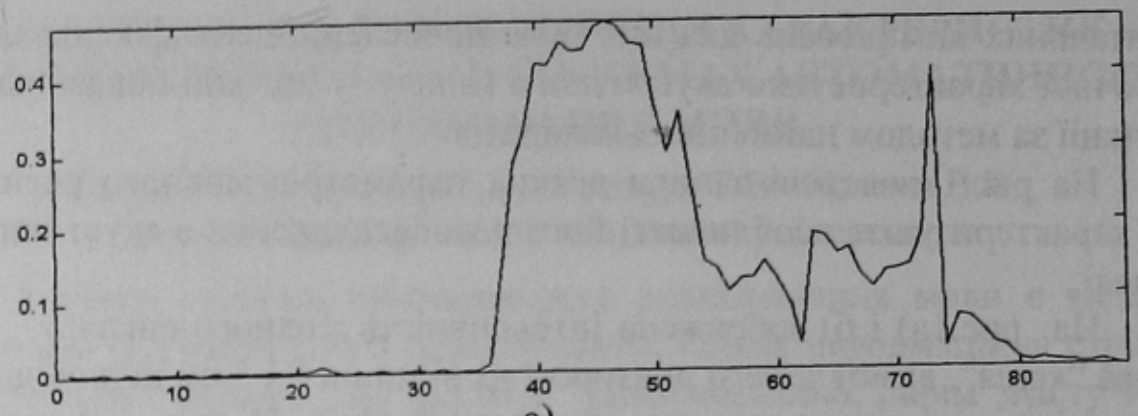
Отримані залежності дозволяють уточнити існуючі моделі каналу передачі мовної інформації і визначити ряд додаткових параметрів мовного сигналу, коваріантних до змісту вимовленого повідомлення.

Аналіз особливостей мовного сигналу дозволяє відзначити його складний характер. Це заважає розробці точних моделей мовотворення, тому всі відомі моделі основані на спрощеному поданні процесу мовотворення [2,3,4,5]. Найбільш розповсюдженою з них є лінійна модель [3,4], представлена на рис. 2.

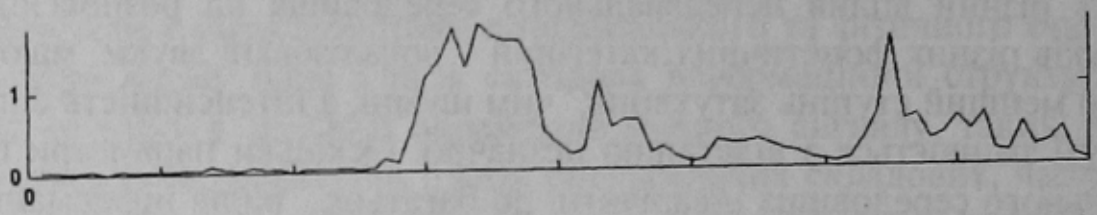
Згідно з цією моделлю звуковий тиск біля губ $Y(w)$ при відомій величині тиску $P_G(w)$ на виході джерела збуджувальних звукових коливань визначиться виразом

$$Y(w)=P_G(w)V_G(w)V_S(w)V_L(w), \qquad (1)$$

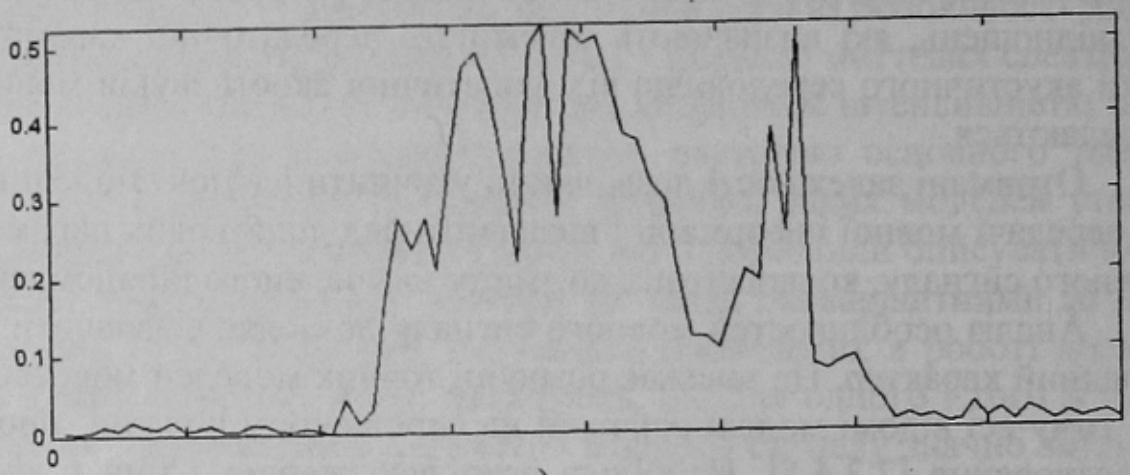
де w - частота коливання, а $V_G(w)$, $V_S(w)$, $V_L(w)$ – передаточні функції джерела звуку (голосової щілини), мовного тракту і випромінювача відповідно; $R_G(w)$, $L_G(w)$ - активний і індуктивний опори голосової щілини; $Z_L(w)$ – комплексний опір випромінювача звуку.



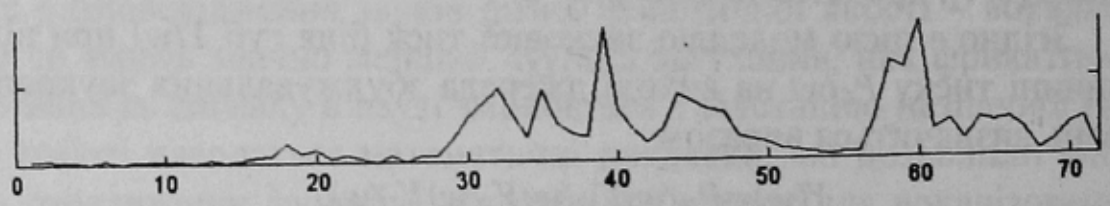
а)



б)



в)



г)

Рисунок 1 – Інтенсивність мовного сигналу : а), б) – для слова "каша" на відстані 0 см і 5 см від мікрофона відповідно; в), г) – для слова "кафе" на тих же відстанях

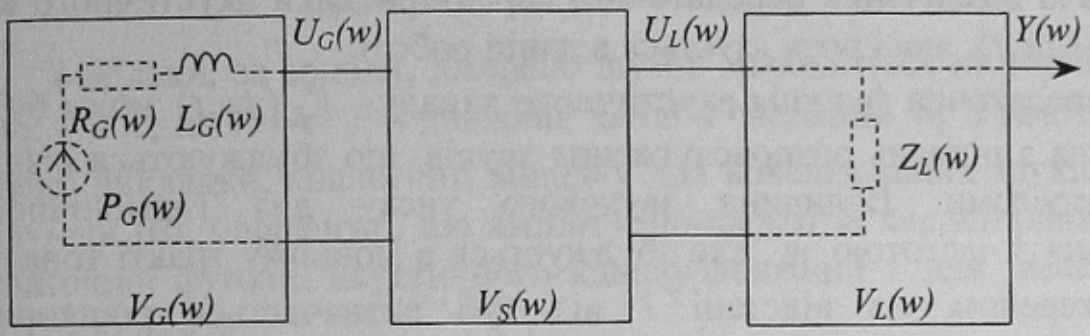


Рисунок 2 - Лінійна модель мовотворення

Передаточні функції голосової щілини і випромінювача при відомих значеннях їх комплексних опорів є повністю визначеними. Передаточна функція голосового тракту описує його резонансні властивості і визначається шляхом чисельного інтегрування хвильових рівнянь для тієї чи іншої конфігурації мовного тракту. Відповідно з моделлю (1) інформативними параметрами для розпізнавання звуків мови є форма миттєвого спектра сигнала, формантні параметри, їх часові характеристики. Однак розгляд цих параметрів не дає пояснення явища різного затухання сигналів вокалізованих і шумних звуків мови та пауз, відзначеного на рис. 1. Для уяснення цього явища в роботі розглядається узагальнена передаточна функція каналу мовного спілкування, який складається з джерела мови, акустичного каналу та приймача мови:

$$V(jw,t) = \frac{U_s(jw,t)}{P_G(jw,t)} = V_M(jw,t)V_K(jw,t)V_R(jw,t), \qquad (2)$$

де $V_M(jw,t)$ - комплексна передаточна функція моделі мовотворення, $V_K(jw,t)$ - комплексна передаточна функція передавального середовища (акустичного каналу), $V_R(jw,t)$ - комплексна передаточна функція приймача, $U_S(jw,t)$ - сигнал на виході приймача. Передаточна функція моделі мовотворення визначається з рівняння (1) і може бути знайдена для поточного моменту часу t безпосередньо з мовного сигналу [6,7]. Частотна характеристика приймача постійна в часі та є рівномірною для всього діапазону частот мовного сигналу для більшості типів сучасних мікрофонів і теж є відомою. Теоретичний і практичний інтерес представ-

ляє задача визначення передаточної характеристики акустичного каналу $V_K(jw,t)$, яка і розв´язується в даній роботі.

Передаточна функція акустичного каналу $V_K(jw,t)$ може бути визначена з рівнянь розповсюдження звуків, що збуджуються різними джерелами. Величина звукового тиску для гармонічного коливання з частотою w_i, яке збуджується в мовному тракті тональним джерелом, на відстані l від губ визначається рівнянням сферичної хвилі [3,4,5]:

$$Y_l(jw_i,t) = \frac{1}{l} Y(jw_i,t) e^{-jw_i(l/c)}, \qquad (3)$$

де $Y(jw_i,t) = Y(jw)e^{-jw_i}$ - комплексна амплітуда акустичного тиску біля губ. Тоді, відповідно з (3), спектр вокалізованого звука на віддалі l від губ визначиться виразом

$$F_{lv}(jw,t) = \frac{1}{l} F_v(jw,t) e^{-jwt_d}, \qquad (4)$$

де $t_d = d\psi(w)/dw = l/c$ - час затримки в акустичному каналі довжини l. Вираз (4) з урахуванням теореми зміщення [8] дозволяє перейти від спектра вокалізованного звука до його представлення у вигляді часової функції:

$$Y_{lv}(t) = \frac{1}{l} Y_v(t - t_d) \quad , \qquad (5)$$

звідки амплітудно-частотна характеристика передаточної функції акустичного каналу для вокалізованих звуків визначиться виразом:

$$\left|V_{Kv}\right| = \frac{Y_{lv}(t)}{Y_{lv}(t - \mathrm{t_d}) -} = \frac{1}{l} \quad . \qquad (6)$$

Звуковий тиск $Y_{\ln}(jw_i,t)$ шумового коливання з частотою w_i в паузі мовного сигналу, який створюється в навколишньому середовищі джерелом акустичних шумів на відстані l від приймача, визначається рівнянням розповсюдження плоскої хвилі

$$Y_{\ln}(jw_i,t) = Y_n(jw_i,t)e^{-jw_i(l/c)}, \quad (7)$$

оскільки, за звичай, джерело шумів знаходиться на значній відстані: $l \succ \lambda$, де $\lambda = c/f_i$ - довжина хвилі з частотою w_i в повітрі. Зробивши викладки, аналогічні випадку для вокалізованих звуків, і враховуючи (6), одержимо, що амплітудно-частотна характеристика передаточної функції акустичного каналу довжини l для завад визначиться виразом:

$$|V_{Kn}| = \frac{Y_{\ln}(t - t_d + \Delta t_d)}{Y_{l_a n}(t - t_d)} = 1, \quad (8)$$

де $t_d = l/c$ - час затримки сигналу завади при його розповсюдженні від джерела шуму до приймача мовного сигналу, $\Delta t_d = l_a/c$ - час розповсюдження сигналу завади при його поширенні в акустичному каналі довжиною l_a . Оскільки в разі відсутності мовного сигналу на мікрофон діє тільки сигнал завади $Y_n(t)$, то вираз (8) можна ототожнювати з передаточною функцією акустичного каналу для пауз мовних сигналів.

Величина акустичного тиску для фрикативних звуків, породжуваних в голосовому тракті джерелом шуму, на відстані l_a від губ визначається виразом [2,4] :

$$Y_{lf} = \alpha_1(\mathrm{Re}^2 - \mathrm{Re}_{кр}^2) -, \quad (9)$$

де α_1 - константа, Re - число Рейнольдса, $\mathrm{Re}_{кр}$ - критичне число Рейнольдса.

Вираз (9) не визначає в явному вигляді залежність величини звукового тиску фрикативних звуків від відстані. З метою її визначення були зняті експериментальні залежності затухання інтенсивності фрикативних звуків від відстані до мікрофона. Експеримент проводився з протяжними фрикативними звуками / с /, /ш /, / ф /, / х /, інтенсивність яких вимірювалась мікрофонами МК-5А на відстані l_a=0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.14 і 0.20 м від губ. Для кожної з вказаних відстаней проводилась серія із десяти вимірів, результат вимірювань представлявся усередненим значен-

ням $\overline{Y}_{lf}$. На основі одержаних результатів вимірювань проведена ідентифікація передаточної функції акустичного каналу для фрикативних звуків, представленої у вигляді показникової функції $g^{-\alpha l_a}$. Значення параметрів g і α даної функції, знайдені по методу найменших квадратів, дорівнюють відповідно 2.5 і 0.35, звідки амплітудно-частотна характеристика передаточної функції акустичного каналу для фрикативних звуків визначиться виразом

$$\left|V_{Kf}\right| = 2.5^{-\alpha l_a}. \qquad (10)$$

Висновки

Відмінність математичних виразів (5), (8) і (10), які описують залежність звукового тиску для вокалізованих, фрикативних звуків і пауз від відстані, дозволяє використати передаточну характеристику акустичного каналу за інваріантну до диктора і завад ознаку для розрізнювання мовних елементів даного типу. Необхідною умовою отримання цієї ознаки є вимірювання сигналу не менше ніж у двох точках акустичного каналу, з чого виникає необхідність двоканальної обробки сигналу.

Список джерел

1. Быков М.М. Методы и средства измерения и преобразования информации в системах машинного распознавания речи. – Дисс. на соискание уч. степени канд. техн. наук. – Винница, 1985. – 243 с.
2. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ. / Под ред. М.В. Назарова и Ю.Н. Прохорова. - М.: Радио и связь, 1981. – 496 с.
3. Маркел Д. Дж., Грей А.Х. Линейное предсказание речи. – М.: Связь, 1980. – 308 с
4. Фант Г. Акустическая теория речеобразования. – М.: Наука,1964. – 284 с.
5. Фланаган Дж. Анализ, синтез и восприятие речи. – М.: Связь, 1968. – 392 с.