УДК 681.142.2; 622.02.658.284; 621.325

Ю.М. Рашкевич, *д-р техн. наук, професор,* ***Д.Д. Пелешко,*** *д-р техн. наук, професор,*
Ю.М. Пелех *(Національний університет «Львівська політехніка»)*
М.З. Пелешко, *канд. техн. наук*
(Львівський державний університет безпеки життєдіяльності)

ВИДІЛЕННЯ КВАЗІСТАЦІОНАРНИХ ДІЛЯНОК МОВНОГО СИГНАЛУ НА ОСНОВІ ПСЕВДООБЕРТАННЯ МАТРИЦІ ВІДНОСНИХ НЕСИМЕТРИЧНИХ МІР КОНВЕРГЕНЦІЇ

Запропоновано методи сегментації мовного сигналу квазістаціонарними ділянками, які будуються за Хаусдорфовою метрикою у векторному просторі алгебраїчних характеристик. Характеристичні вектори отримуємо шляхом побудови з елементарних ділянок мовного сигналу матриці самоподібності (матриці відносних несиметричних мір конвергенції) і подальшого розв’язання лінійної виродженої системи за методом Мура-Пенроуза. Псевдообертання здійснюється на основі сингулярного розкладу матриці самоподібності. Наводяться порівняння результатів сегментації на основі розробленого методу та методу DELCO. На основі цих результатів встановлено ефективність розробленого методу при сегментації ділянок високої енергії мовного сигналу.

Ключові слова: мовний сигнал, квазістаціонар, матриця самоподібності, псевдообернена матриця, псевдорішення Мура-Пенроуза, сингулярний розклад, метрика Хаусдорфа.

Вступ та аналіз літературних джерел

Побудова сучасних систем штучного інтелекту залежить від ефективних методів інтелектуального аналізу різноманітних даних, у тому числі і мовних сигналів. Обробка мовних сигналів становить окремий підклас, що зумовлене природою виникнення мови. На сьогодні стосовно цього підкласу сигналів існує велика кількість не розв’язаних повністю задач. Наприклад, розпізнавання мови, зміна часового масштабу та ін.

Пошуки стійких і якісних розв’язків класичних задач інтелектуального аналізу стосовно мовних сигналів зумовлені стрімким розвитком робото-технічних систем. Але такі системи накладають два важливих обмеження. Перше – це потреба роботи в режимі реального або наближеного до нього часу. Друге обмеження є наслідком першого – у мовного аналізатора (обробника) на вході не завжди є увесь часовий ряд [11]. Тому можна опиратись на аналіз інформативності лише частини сигналу. А це на початкових етапах може призвести до значних помилок, які, передусім, викликані малим за розміром вхідним набором даних.

Такі обмеження, водночас, унеможливлюють використання спектральних методів, які базуються на класичних операційних перетвореннях (так звані методи лінійного спектрального аналізу) [10]. Це пояснюється тим, що якісні представлення гармонійними функціями вимагають знання сигналу на усьому проміжку аналізу. Більше цього, використання синусоїди як базової функції операційного перетворення не завжди дає можливість ефективно охарактеризувати локальні властивості (характеристики) досліджуваного сигналу.

Описані недоліки стали поштовхом до появи спектральних методів, які базуються на використанні віконного (короткочасного) перетворення. Це покращило аналіз локальних характеристик, але обмеженням вже стало саме вікно, що призводить до часової локалізації сигналу [7, 10, 11].

Інший напрям розвитку спектральних методів – це, фактично, параметричне моделювання [8, 10] на основі авторегресії. Наприклад, метод лінійного передбачення мови і його варіації, кепстральний аналіз та ін. Очевидною перевагою методів цього напряму є можливість прогнозування (передбачення) сигналу, або його характеристичних трендів. Серед недоліків – складність підбору порогових значень, неефективність методів на ділянках низьких енергій.

Статистичні та статистично-мережеві методи (приховані марківські процеси, змішані гаусівські моделі і ін.) є методами, які базуються на формулюванні моделі і організації вірогідного попереднього “навчання” (наприклад, у [6]) для визначення параметрів моделі, якщо метод аналізу розглядається у поєднанні з мережевими методами. Саме в останньому і полягає основна слабкість таких моделей. Відсутність якісних наборів для попереднього визначення параметрів моделі призводить до того, що такі моделі є нечутливими до випадкових чи тимчасових мовних викидів, які можуть виникати у мовному сигналі внаслідок різноманітних причин. Але у випадку статистичних методів, наприклад, при використанні прихованих марківських моделей без нейронних мереж, цих проблем немає. Проте якість роботи методу загалом, наприклад, при розпізнаванні мови, є залежною від якості стаціонарних чи квазістаціонарних ділянок, що забезпечує правильне визначення можливих станів марківського процесу та імовірностей переходу між станами та спотворення стану [5]. Тому важливим завданням в обробці мовних сигналів є задача виділення стаціонарних чи квазістаціонарних ділянок.

Мета дослідження

Основним завданням роботи є розробка методів виділення квазістаціонарних ділянок на основі побудови метричних просторів характеристичних векторів елементарних ділянок. Метод виділення цих векторів базується на розв'язанні лінійної виродженої системи на основі методу Мура-Пенроуза [1]. Вироджена квадратна матриця-оператор лінійної системи визначається як матриця відносних несиметричних мір конвергенції, псевдообертання якої здійснюється за її сингулярним розкладом.

1. Постановка задачі виділення квазістаціонарних ділянок мовного сигналу

Нехай задано часовий проміжок $\tau = [0; T]$, $T \in R^{1,+}$. На проміжку τ системою відкритих множин

$$\Gamma = \{T_i\}_{i = 1, 2, \ldots}, \; T_i = [t \mid t_{i-1} \le t < t_i, \, t_{i-1}, t_i \in \tau], \tag{1}$$

з діаметром $\left|T_i\right| = \sup\limits_{t_a, t_b \in T_i} d(t_a, t_b)$ (де d – метрика простору R^1), визначимо топологію Γ. Оскільки проміжок τ є замкненою обмеженою множиною, то у топології Γ завжди існує диз'юнктивне (з тривіальним перетином) скінченне покриття $\chi = \{T_i \mid i = 1..n\}$, тобто має місце

$$\forall i, j \in [1; n]: T_i, T_j \in \chi \wedge i \ne j \to T_i \cap T_j = \varnothing; \tag{2}$$

$$\tau = \bigcup_{i=1}^{n} T_i . \tag{3}$$

Вважатимемо, що усі елементи покриття χ мають однаковий діаметр, що дорівнює l: $\forall i \in [1, n]: \left|T_i\right| = l$. Це визначає χ як l-покриття проміжку τ.

Нехай на проміжку τ визначено мовний сигнал $x(t)$ як неперервне сюр'єктивне відображення

$$x: \tau \to R^1. \tag{4}$$

Оскільки $x(t)$ є неперервним відображенням, то, за умови існування покриття χ, отримаємо покриття η (необов'язково диз'юнктивне) області значень функції $x(t)$, а саме

$$\eta = \{X_i \mid i = 1..n \}, \tag{5}$$

де $X_i = \{x(t) | t \in T_i\} = \{x(t) | t \in T_i\}$ – елемент покриття η.

Квазістаціонарні ділянки Y_i мовного сигналу $x(t)$ є об'єднанням елементів диз'юнктивного покриття η на основі заданого критерію належності

$$Y_i = \bigcup_{j=a_i}^{b_i} X_j , \tag{6}$$

де $X_j \in \eta$ – послідовні елементи покриття η, які задовольняють критерію належності; a_i, b_i – індекси в послідовності $\{X_i\}$, які визначають підмножину елементів, об'єднання яких утворюють Y_i. Різниця значень індексів $m_i = b_i - a_i$ визначає кількість елементарних ділянок X_i, яка визначає потужність квазістаціонарної ділянки Y_i: $|Y_i| = m_i l$.

За (5) і (6) множина $\{X_i\}$ утворює нове диз'юнктивне покриття η' простору сигналу $x(t)$, а тому має місце рівність

$$x(t)=\bigcup_{i=1}^{m} Y_i\ , \tag{7}$$

де $m \le n$ – кількість квазістаціонарних ділянок Y_i.

Зважаючи на (6) і (7) розмірність сигналу $x(t)$ (кількість точок дискретизації на проміжку T) визначається так

$$\dim\,(x(t)) = nl = l\sum_{i=1}^{m} m_i\ . \tag{8}$$

З урахуванням (5), (7) і (8) задачу, виділення квазістаціонарних ділянок сигналу $x(t)$ можна розглядати як задачу побудови оператора перетворення покриттів f: $\chi \to \chi'$ або F: $\eta \to \eta'$.

2. Побудова матриці відносних несиметричних мір подібності

Після дискретизацій на проміжку τ сигнал $x(t)$ зсунемо в додатну область $x(t) = x(t) + \max\limits_{t\in\tau} x(t) - \min\limits_{t\in\tau} x(t)$ і пронормуємо зсунутий сигнал $x(t)$ за схемою оберненого мінімаксного множника K [2]

$$K=\frac{1}{\sup\limits_{i\in[1,n]} x(t)}. \tag{9}$$

У результаті надалі розглядатимемо $x(t)$ як дискретний нормований сигнал $K\,x(t)$, $t \in \tau$.

Оператор ∇_i: $T_i \to X_i$ перетворення l-вимірних векторів, визначених на проміжку T_i, який відповідає елементу X_i покриття η, будуємо як один із видів матриці подібності, а саме – як матрицю відносних несиметричних мір конвергенції (матриця включення), елементи якої є несиметричними мірами подібності об’єктів [9]:

$$\nabla_i=\begin{pmatrix} 1 & \dfrac{x(t_2)}{x(t_1)} & \dots & \dfrac{x(t_l)}{x(t_1)} \\ \left(\dfrac{x(t_2)}{x(t_1)}\right)^{-1} & 1 & \dots & \dfrac{x(t_l)}{x(t_2)} \\ \dots & \dots & \dots & \dots \\ \left(\dfrac{x(t_l)}{x(t_1)}\right)^{-1} & \left(\dfrac{x(t_l)}{x(t_2)}\right)^{-1} & \dots & 1 \end{pmatrix},\ \forall i\in[1;l]: t_i\in T_i\,. \tag{10}$$

Розмірність оператора (10) становить: $\dim \nabla_i = l$? l.

Оператор ∇_i є квадратною виродженою невід’ємною матрицею. Невід’ємність випливає із зсуву та нормування мінімаксним множником (8). А невиродженість – із методу оточення мінорів [4] за такими формулами та міркуваннями

$$M_1 = 1;$$

$$M_2^1=\begin{pmatrix} 1 & \dfrac{x(t_2)}{x(t_1)} \\ \left(\dfrac{x(t_2)}{x(t_1)}\right)^{-1} & 1 \end{pmatrix}=0;\ M_2^2=\begin{pmatrix} 1 & \dfrac{x(t_3)}{x(t_1)} \\ \left(\dfrac{x(t_2)}{x(t_1)}\right)^{-1} & \dfrac{x(t_3)}{x(t_2)} \end{pmatrix}=0;\dots M_2^{l-1}=\begin{pmatrix} 1 & \dfrac{x(t_l)}{x(t_1)} \\ \left(\dfrac{x(t_2)}{x(t_1)}\right)^{-1} & \dfrac{x(t_l)}{x(t_2)} \end{pmatrix}=0, \tag{11}$$

де M – мінори матриці ∇_i. Оскільки усі мінори другого порядку ($M_2^1,\dots,M_2^{l-1}$) є нульовими, то M_1 є базисним мінором [4]. Це означає, що ранг матриці ∇_i буде становити

$$\operatorname{rang}\,(\nabla_i) = 1. \tag{12}$$

Оскільки $\operatorname{rang}\,(\nabla_i) < l$, то матриця ∇_i є виродженою.

3. Виділення алгебраїчних характеристик на основі псевдообертання Мура-Пенроуза

Для вирішення завдання виділення векторів, які виступатимуть характеристиками елементарної ділянки мовного сигналу для кожного елемента покриття χ, розглянемо рівняння

$$\nabla_i y_i = x_i \,, \tag{13}$$

де $x_i = \left\{ x\left(t_p\right) \in X_i \mid p = 1..l \right\}$; $y_i = (y_{i,1}, \ldots, y_{i,l})$ – l-вимірний вектор амплітудних значень мовного сигналу $x(t)$ на i-му інтервалі.

За (13) вектор $y_i = (y_{i,1}, \ldots, y_{i,l})$ буде визначатись так

$$y_i = \nabla_i^{-1} x_i \,. \tag{14}$$

Оскільки матриця ∇_i є виродженою ($\det(\nabla_i) = 0$), то обернена матриця ∇_i^{-1} не існує. Тому для розв'язання задачі (13) за теоремою про мінімізацію нев'язки $\|x_i - \nabla_i y_i\|^2$ лінійної системи (14) [3] пропонується такий спосіб визначення вектора y_i

$$y_i = \nabla_i^+ x_i + \left(1 - \nabla_i^+ \nabla_i\right) r_i \,, \tag{15}$$

де ∇_i^+ – узагальнена обернена матриця Мура-Пенроуза (псевдобернена до ∇_i матриця [1, 3]); $\left(1 - \nabla_i^+ \nabla_i\right)$ – оператор проектування на ядро оператора ∇_i; r_i – випадковий вектор розмірності l. Перший доданок у (15) виступає псевдооберненим рішенням, а другий є розв'язком однорідної системи $\nabla_i y_i = 0$. Наведений через (15) спосіб визначення вектора характеристик i-го елемента покриття χ є можливим оскільки, згідно з [3], матриця $\nabla_i^+ \nabla_i$ не є виродженою.

Матриця Мура-Пенроуза ∇_i^+ визначається за сингулярним розкладом матриці ∇_i у такий спосіб [1]:

$$\nabla_i^+ = V_i \Sigma_i^+ U_i^T \,, \tag{16}$$

де U_i, V_i – унітарні матриці порядку $l?l$ сингулярного розкладу матриці ∇_i; Σ_i^+ – матриця порядку $l?l$, яка є псевдоберненою до діагональної матриці Σ_i сингулярного розкладу матриці ∇_i. Оскільки матриця Σ_i є також виродженою, то матрицю Σ_i^+ отримуємо з Σ_i шляхом заміни усіх ненульових сингулярних чисел $\sigma_{i,q}$ ($\sigma_{i,1} \geq \sigma_{i,2} \geq \ldots \geq \sigma_{i,l} \geq 0$) на відповідно обернені до них $1/\sigma_{i,q}$.

В ітераційному процесі знаходження y_i^{j+1} за (15), випадковий вектор r_i^{j+1} – визначався за нев'язкою: $r_i^{j+1} = \|x_i - \nabla_i y_i^j\|_l$, тут $\|\cdot\|_l$ – l-норма [4].

3. Виділення квазістаціонарних ділянок мовного сигналу шляхом побудови метричних просторів

З вектора $y_i = (y_{i,1}, \ldots, y_{i,l})$, отриманого за (15), сформуємо такі набори даних

$$\Lambda = \left\{ \Lambda_i \mid \Lambda_i = (y_{1,i}, \ldots, y_{l,i}) \right\}_{i=1..l} \,, \tag{17}$$

Якщо для елементів цієї множини ввести метрику, то задача виділення квазістаціонарів зводиться до об'єднання близьких за метрикою елементів отриманого метричного простору. Якщо метрикою вибрати метрику Хаусдорфа, то за нею близькість двох елементів Λ_i і Λ_j, $i, j \in [1; l]$, простору Λ буде визначатись так

$$\rho_H\left(\Lambda_i, \Lambda_j\right) = \max\left\{ \sup_{y_i \in \Lambda_i} \inf_{y_j \in \Lambda_j} \rho\left(y_i, y_j\right), \sup_{y_j \in \Lambda_j} \inf_{y_i \in \Lambda_i} \rho\left(y_i, y_j\right) \right\}_{i=1..l} \,, \tag{18}$$

де $\rho(y_i, y_j)$ – відстань між точками множин Λ_i і Λ_j.

За метрикою (16) квазістаціонар формується за таким правилом: будь-яка множина $\{X_n..X_m\}$ послідовних елементів покриття η утворює квазістаціонар Y_i (6), якщо для будь-яких двох елементів цієї множини метрика (18) не перевищує наперед заданого відхилення $\Delta > 0$

$$\exists \varepsilon \geq 0, \forall i, j \in [n; m] : \rho_H\left(\Lambda_i, \Lambda_j\right) \leq \Delta \,. \tag{19}$$

Оскільки квазістаціонари утворюють диз'юнктивне покриття, то побудову наступної квазістаціонарної ділянки треба розпочинати із m+1-го елемента покриття η.

4. Результати експериментів та висновки

Розроблений метод сегментації мовного сигналу на квазістаціонарні ділянки апробовано на прикладі первинного поділу слова "миша", хвильове представлення якого наведено на рисунку.

Характеристики мовного сигналу є такими: тривалість слова – 1,03 секунди, частота дискретизації 11 025 Гц, довжина елементарної ділянки – l = 120 відліків. Мовний сигнал містить основні види фонем: наголошену та ненаголошену голосні, а також вокалізовану та невокалізовану приголосні.

Як експертну оцінку обрано результати сегментації цього ж сигналу за алгоритмом DELCO із пороговим значенням 1,6 при тому ж розмірі елементарної ділянки.

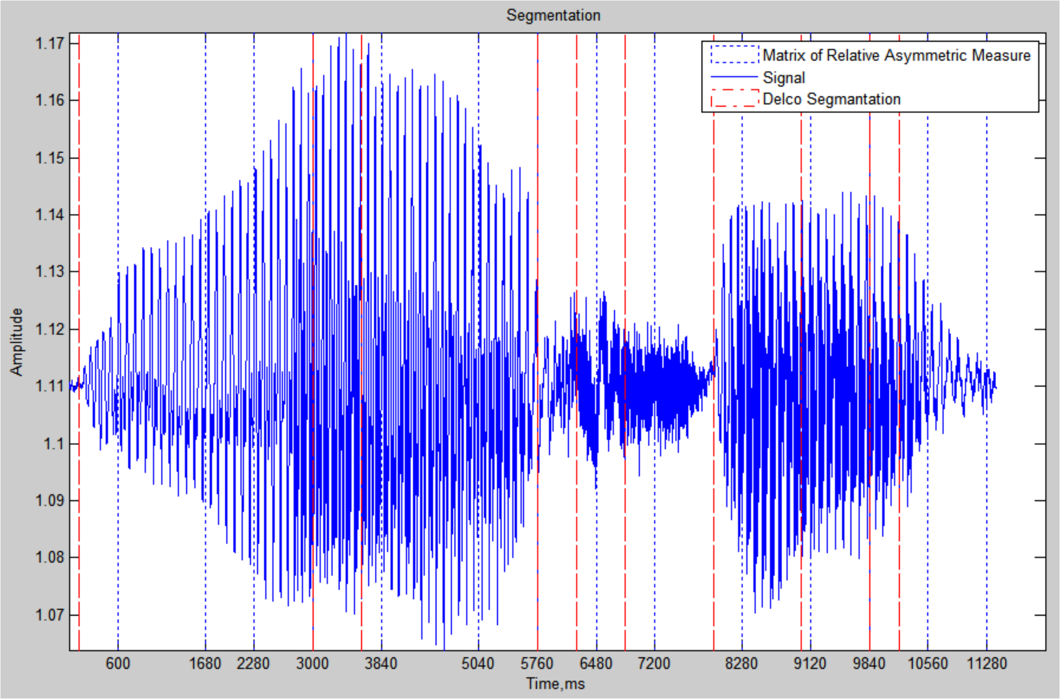
Вибір методу DELCO зумовлений тим, що він позиціонується як один із найкращих параметричних методів сегментування в областях високої енергії.

За розробленим методом виділено 13 сегментів, а за методом DELCO – 10. При цьому, в точках t = 0,3с, t = 0,57с, t = 0,98с результати сегментації повністю збігаються. На решті частини сигналу результати є близькими з різними ступенями близькості.

На основі цих результатів можна констатувати, що розроблений метод порівняно з алгоритмом DELCO є більш чутливим до змін енергії в часі. Це особливо добре видно на відповідних проміжках [0,06с; 0,57с] і [0,57с; 0,82с].

Так перевага отриманого методу може бути дуже корисною в задачах, ефективне розв'язання яких залежить від змін енергетичних характеристик мовного сигналу. Наприклад у задачах кодування для процесів передачі мовних сигналів телекомунікаційними каналами.

Аналіз результатів порівняння роботи розробленого методу та алгоритму DELCO на прикладах використання інших мовних сигналів лише підтвердив сформульований висновок і підтвердив можливість використання цього методу для первинної сегментації на квазістаціонарні ділянки.



Результати сегментації мовного сигналу (слово "миша") за алгоритмом DELCO та за методом на основі матриці відносних асиметричних мір конвергенції

Список літератури:

1. **Penrose R. A** generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society 51, 406-413 (1955).
2. **Rashkevych Y.,** Peleshko D., Kovalchuk A., Kupchak M., Pelekh Y. Speech signal pseudoinvariants. Computer Science and Information Technologies: Materials of the Vith International Scientific and Technical Conference CSIT 2011. - Lviv: Publishig House Vezha&Co, 2011. 2011, с.21-22.
3. **Алберт А.** Регрессия, псевдоинверсия и рекуррентное оценивание/ А.Алберт// пер.с англ. – М.: Наука. – 1977. – 224с.
4. **Гантмахер Ф.Р.** Теория матриц . – М.: Наука. – 1967. – 576 с.
5. **Добрушкін Г. О.,** Данилов В. Я. Основні підходи до розпізнавання мовленнєвої інформації/ Г. О. Добрушкін, В. Я. Данилов// Інформаційні технології та комп’ютерна техніка. Вісник Вінницького політехнічного інституту. – 2009. – № 4. – С. 50-64.
6. **Дорохин О.А.,** Старушко Д.Г., Федоров Е.Е., Шелепов В.Ю. Сегментация речевого сигнала. «Искусственный интеллект». – 2000. – №3. – С. 450-458.
7. **Дремин И.М.,** Иванов О.В., Нечитайло В.А. Вейвлеты и их использование/ И.М.Дремин, О.В.Иванов, В.А.Нечитайло// Успехи физических наук. – 2001. – Т. 171. – №5. – С.465-501.
8. **Жиляков Є.Г.,** Прохоренко Є.І., Болдишев А.В., Фірсова А.А., Фатова М.В. Сегментація мовних сигналів на основі аналізу особливостей розподілу часток енергії за частотним інтервалом, Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2011. – № 17. – С. 44 – 50.
9. **Матрица мер** конвергенции [Електронний ресурс] – Режим доступу до журналу: http://ru.wikipedia.org/wiki/Матрица_мер_конвергенции.
10. **Сорокин В.Н.,** Цыплихин А.И. Сегментация речи на кардинальные элементы, Информ. процессы. – 2006. – Т. 6. – № 3. – С. 1772-207.
11. **Фаніна Л. О.** Алгоритми відновлення вимовленої послідовності в системах розпізнавання мови/Л. О. Фаніна // Автоматика. Автоматизация. Электротехнические комплексы и системы. – Херсон: Изд-во Херсон. нац. техн. ун-та. – 2004. – № 2. – С.131-137.

Ю.М. Рашкевич, Д.Д. Пелешко,
Ю.М. Пелех, М.З. Пелешко

ОПРЕДЕЛЕНИЕ КВАЗИСТАЦИОНАРНЫХ УЧАСТКОВ РЕЧЕВОГО СИГНАЛА НА ОСНОВЕ ПСЕВДООБРАЩЕНИЯ МАТРИЦЫ ОТНОСИТЕЛЬНЫХ НЕСИММЕТРИЧНЫХ МЕР КОНВЕРГЕНЦИИ

Предложены методы сегментации речевого сигнала квазистационарными участками, которые строятся за метрикой Хаусдорфа в векторном пространстве алгебраических характеристик. Характеристические векторы получены путем построения из элементарных участков речевого сигнала матрицы сходства (матрицы относительных несимметричных мер конвергенции) и последующего решения линейной вырожденной системы по методу Мура-Пенроуза. Псевдовращение осуществляется на основе сингулярного разложения матрицы сходства. Приводятся сравнения результатов сегментации на основе разработанного метода и метода DELCO. На основе этих результатов установлена эффективность разработанного метода при сегментации участков высокой энергии речевого сигнала.

Ключевые слова: речевой сигнал, квазистационар, матрица сходства, псевдообратная матрица, псевдорешение Мура-Пенроуза, сингулярное разложение, метрика Хаусдорфа.

Y.M. Rashkevych, D.D. Peleshko, Y.M. Pelekh, M.Z. Peleshko

IDENTIFICATION OF THE QUASISTATIONARY AREAS OF THE VOICE SIGNAL BASED ON PSEUDOINVERSE MATRIX OF RELATIVE ASYMMETRIC CONVERGENCE VALUES

Methods of speech signal segmentation based on the quasistationary areas that are constructed by the Hausdorff's metric in the vector space of algebraic characteristics are suggested. Characteristic vectors are obtained by constructing the self-similarity matrix from the elementary parts of the speech signal (matrix of the relative asymmetric convergence values) and then solving a linear degenerated system by Moore–Penrose technique. Pseudoinverse is performed basing on the singular value decomposition of the self-similarity matrix. The comparison of the segmentation results by the proposed method and the method DELCO has been carried out. Based on these results the efficiency of the developed method for the segmentation of high-energy parts of the speech signal was estimated.

Key words: voice signal, quasistationary, self-similarity matrix, pseudoinverse matrix, Moore–Penrose pseudoinverse, singular value decomposition, Hausdorff's metric.

❖ ❖ ❖