

РОЗРОБКА МЕТОДУ ПОРІВНЯННЯ НА ІДЕНТИЧНІСТЬ ЗВУКОВИХ МОВНИХ СИГНАЛІВ

У статті запропоновано рішення науково-технічного завдання порівняння ідентичності звукових мовних сигналів з використанням методу, який базується на виділенні зі спектру сигналу інформативних частот мови.

Ключові слова: сигнал, звукова інформація, інформативні частоти, амплітудно-частотна характеристика (АЧХ), дискретизація.

In the article solution of scientific and technical assignment to compare the identity of the sound voice signals with the employment of the method, which rests on isolation from the spectrum of the signal the informative frequencies of the speech, is presented.

Keywords: signal, sound information, informative frequencies, amplitude-frequency description (AMD), discretization.

Вступ. У зв'язку з бурхливим розвитком технологій комп'ютерної обробки звуку часто постає завдання автоматичного визначення ступеню схожості звукових сигналів. Зважаючи на об'єктивну необхідність та переваги автоматичного визначення ступеню схожості звукових сигналів у таких сферах діяльності, як, наприклад, безпека, дистанційне навчання, системи автоматичного керування тощо, актуальність цієї теми дедалі зростає. Тому порівняння звукових сигналів – одне з найважливіших завдань у роботі з комп'ютерною технікою.

Постановка завдання. Звук зберігається на комп'ютері у цифровому вигляді, а його представлення залежить від способу і алгоритму оцифрування. Звук – це коливання повітря або іншого середовища, які уловлюються людським вухом.

На сьогоднішній день методів, за допомогою яких можна проводити аналіз схожості звукових сигналів, існує велика кількість. Серед них такі:

1. метод динамічного програмування. 2. нейронні мережі.

3. марківські моделі та ін.

Але не один з них не дає достатньо точних і якісних результатів. Тому постає завдання розробки якісного методу аналізу звукової інформації.

Виклад основного матеріалу. Необхідне порівняння двох звуків та проведення їх якісного аналізу. Звук, що надходить у мікрофон, аналізується та порівнюється з еталонним машинним сигналом. Весь процес аналізу можна поділити на 4 етапи:

Першим етапом аналізу нашого сигналу є усунення первинних спотворень, що вносить вхідний канал звуку (мікрофон).

Кожен мікрофон має певну АЧХ. Тому після проходження звукового сигналу крізь мікрофон наш сигнал спотворюється. Розглянемо два варіанти вирішення цієї проблеми:

1. Пронормувати еталонний сигнал у відповідності до АЧХ мікрофона.

2. Поставити смуговий фільтр, який вирізає діапазон частот еталонного сигналу комп'ютера не менший, ніж діапазон частот, що сприймає наш приймальний канал. Мікрофони ноутбуків приймають діапазон частот у межах 100-5000 Гц. Більш якісні мікрофони – діапазон частот у межах 60-18000Гц.

Другий етап полягає у виділенні характерних частот (для кожної мови вони різні). Виділення частот реалізується наступним чином: після того, як сигнал проходить крізь мікрофон, він потрапляє на систему смугових фільтрів, які виділяють характерні частоти відповідної мови. Аналогічно виділяємо характерні частоти з еталонного сигналу в комп'ютері. Для виділення характерних частот можна застосовувати також розклад в ряд Фур'є, для якого використовується швидке перетворення Фур'є (ШПФ).

Для якісного проведення аналізу сигналу потрібно коректно вибрати частоту дискретизації сигналу. Частоту дискретизації можна обрати відповідно до теореми Котельнікі:

$$2 * F_S \leq F_d ,$$

де F_d – частота дискретизації, F_S – максимальна частота сигналу.

Оскільки передусім нас цікавить цифрове представлення мовних сигналів, то розглянемо спектральні властивості мови (рис. 1).

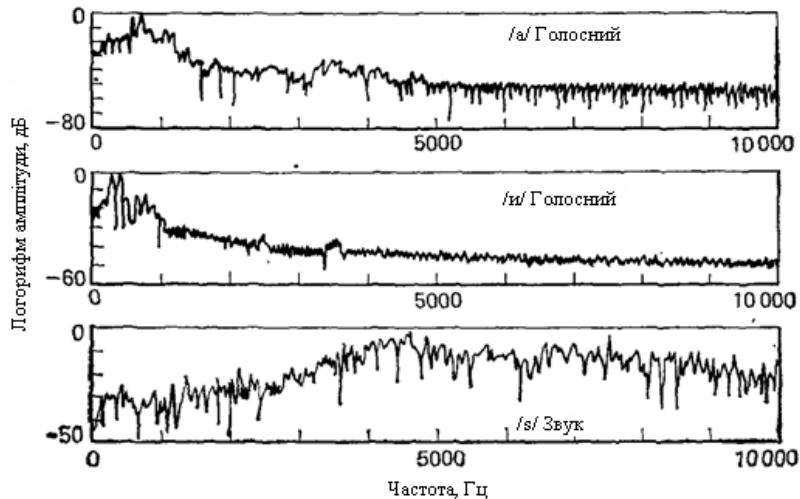


Рис. 1. Спектри типових звуків мови

Бачимо, що для вокалізованих звуків найвища частота, нижче якої максимуми спектра мають рівень 40 дБ, складає близько 4 кГц. З іншого боку, для невокалізованих звуків спектр не затухає навіть на частотах, вищих 8 кГц. Таким чином, для точного відтворення всіх звуків мови необхідна частота дискретизації близько 20 кГц. У більшості випадків така частота дискретизації не потребується. Наприклад, якщо дискретизація передус оцінюванню трьох перших формантних частот вокалізованої мови, то достатньо мати частину спектра до частоти

близько 3,5 кГц. Таким чином, якщо перед дискретизацією мовний сигнал пропускається через фільтр нижніх частот так, що частота Найквіста дорівнює 4 кГц, то частота дискретизації повинна складати 8 кГц. Але тут ми не враховуємо швидкість вимови звуків, залежно від якої потрібно обрати оптимальну узгодження ресурсів ЕОМ і ефективності аналізу сигналу частоту дискретизації.

Після аналізу спектрів однакових сигналів при різному часі вимови, було побудовано залежність частоти дискретизації від часу вимови (рис. 2.).

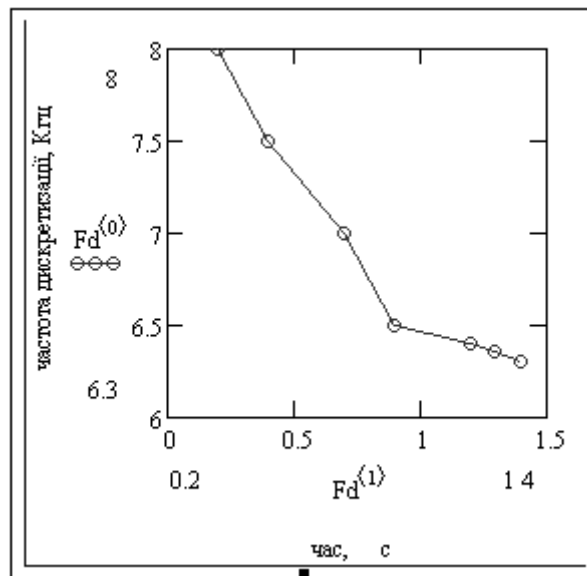


Рис. 2. Залежність частоти дискретизації від часу вимови

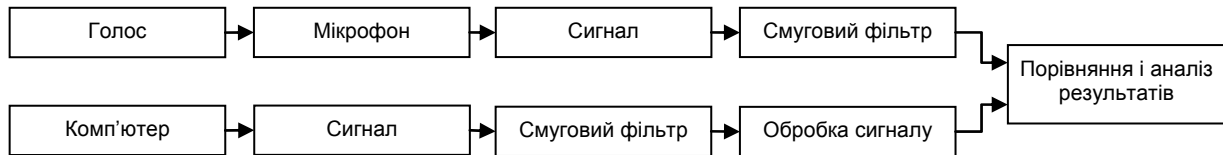
Цю залежність можна апроксимувати експоненціальною функцією. Завдяки отриманому графіку можна зробити висновок, що частота вимови та частота дискретизації є оберненопропорційними: при зменшенні часу вимови частота дискретизації збільшується, а при збільшенні часу вимови – зменшується і виходить на деяке стає значення. Тобто немає сенсу підвищувати F_d у випадку, коли звук вимовляється довго.

На третьому етапі розраховуємо інтегральний показник "I" кожної характерної частоти:

$$I = c_n * F_d,$$

де c_n – амплітудні значення дискет, F_d – частота дискретизації.

Четвертим етапом є порівняння інтегральних показників еталонного сигналу та сигналу, що досліджувався. Представимо весь процес обробки у вигляді схеми:



Математичний апарат, що використовується в розробленій методиці для аналізу звуків. Для якісного аналізу звукового спектру здійснюємо його дискретизацію. Суть дискретизації аналогових сигналів полягає в тому, що неперервність в часі аналогової функції $s(t)$ замінюється послідовністю коротких імпульсів, амплітудні значення яких c_n визначаються з допомогою вагових функцій, або безпосередньо вибірками (відліками) миттєвих значень сигналу $s(t)$ в моменти часу t_n . Представлення сигналу $s(t)$ на інтервалі T сукупністю дискретних значень c_n записується в вигляді:

$$(c_1, c_2, \dots, c_N) = A[s(t)],$$

де A – оператор дискретизації.

Запис операції відновлення сигналу $s(t)$:

$$s(t) = B[(c_1, c_2, \dots, c_N)].$$

Вибір операторів A і B визначається потрібною точністю відновлення сигналу. Найпростішими є лінійні оператори. В загальному випадку:

$$c_n = \int q_n(t) * s(t) * dt \quad (7.1.1),$$

де $q_n(t)$ – система вагових функцій.

Для подальшого виділення характерних частот мови здійснюється розклад спектру ШПФ.

Швидке перетворення Фур'є:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) * e^{-i\omega t} * dt,$$

$F(\omega)$ – називається спектральною густиною сигналу $f(t)$ і дає змогу виявити частоти, які містяться в деяко-

му залежному від часу сигналу. Таким чином виявляємо наші характерні частоти.

Основну ідею швидкого перетворення Фур'є (ШПФ) складає проріджування по частоті та пірамідальний алгоритм, який виключає повторні розрахунки членів ряду Фур'є, що періодично повторюється. Припустимо, що масив чисел f_k містить $N = 2^Y$ відліків. Розділимо цей масив на два проміжні масиви з парними та непарними відліками:

$$f'_k = f_{2k} \quad f''_k = f_{2k+1} \quad 0 \leq k \leq \frac{N}{2} - 1.$$

Якщо тепер виконати ДПФ кожного з масивів з урахуванням того, що крок дискретизації функції збільшився вдвічі, то отримаємо два проміжних спектри, період яких буде дорівнювати $N/2$

$$f'_k \rightarrow F'_n \quad f''_k \rightarrow F''_n \quad 0 \leq n \leq \frac{N}{2} - 1.$$

Для отримання однієї половини спектра F_k складемо отримані спектри із врахуванням того, що відліки функції F''_k зсунуті відносно функції F'_k на один крок дискретизації:

$$F_n = F'_n + F''_n * e^{-i\pi n/N}.$$

Інша половина спектру, що комплексно спряжена з першою, із врахуванням періоду повторення $N/2$ проміжних спектрів визначається виразом:

$$F_{n+N/2} = F'_n + F''_n * e^{-i2\pi(n+N/2)/N} = F'_n - F''_n * e^{-i2\pi n/N}$$

Можна побачити, що для розрахунку повного спектру в даному випадку знадобиться $N^2/4$ операцій, для розрахунку проміжних спектрів – плюс ще N операцій комплексного додавання та множення, що створює досить відчутний ефект порівняно із звичайним ДПФ. Але розділення масивів на дві частини може бути застосованим і до перших проміжних масивів, і до других, і т.д. до тих пір, доки в масивах не залишиться по одному відліку. Таким чином, алгоритм перетворення стає пірамідальним алгоритмом перестановок зі складанням/відніманням та з одиничним множенням на значення $e^{-i2\pi n/N}$ що відповідає рівню піраміди. Виходячи з цього, загальна кількість операцій, що необхідна для повного розрахунку спектру, дорівнює $N * \log_2 N$. Отже, порівняно с ДПФ, швидкість розрахунку збільшується в $N/\log_2 N$ раз. Причому, чим більше N , тим більший ефект від ШПФ. Оцінка точності порівняння відповідних інтегральних показників проводиться через імовірність відхилення за допомогою довірчого інтервалу.

Довірчим інтервалом (ДІ) для параметра ϑ , відповідним довірчій імовірності γ ($\gamma = 0,95$), називається інтервал $J_\gamma(\vartheta) = (\underline{\vartheta}, \bar{\vartheta})$, де $\underline{\vartheta}$ і $\bar{\vartheta}$ – нижня і верхня грани-

ця, які визначаються за вибірковими даними так, щоб $P(\underline{\vartheta} < \vartheta < \bar{\vartheta}) = \gamma$, імовірність "накриття" інтервалом невідомого значення параметра ϑ рівна довірчій імовірності γ . Нижня і верхня границі довірчого інтервалу є СВ, так як визначаються за результатами спостереження. Напівширина довірчого інтервалу визначає точність оцінки, а $\gamma = 1 - \alpha$ – її достовірність, α – рівень значимості. Нехай СВ $X \sim N(\mu, \sigma)$. По виборці об'єму n потрібно побудувати ДІ для параметрів μ і σ з рівнем довіри γ :

$J_\gamma(\mu) = (\underline{\mu}, \bar{\mu})$ и $J_\gamma(\sigma) = (\underline{\sigma}, \bar{\sigma})$. Випадок I ($\sigma - \sigma_0$ – відома величина). Будемо шукати симетричний ДІ у вигляді $J_\gamma(\mu) = (\underline{\mu}, \bar{\mu}) = (\bar{x} - \varepsilon, \bar{x} + \varepsilon)$, где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ – вибіркове середнє параметра μ . Тоді залишається знайти $\varepsilon > 0$, відповідає довірчій імовірності γ , щоб $P\{\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon\} = \gamma$. Якщо $X \sim N(\mu, \sigma_0)$, тоді СВ

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \sim N(\mu, \sigma_0/\sqrt{n}) \quad \text{Відповідно} \quad P\left\{ \left| \bar{X} - \mu \right| < \delta \frac{\sigma_0}{\sqrt{n}} \right\} =$$

$= 2 \cdot \hat{O}(\delta) - 1$. Тоді $\varepsilon = \delta \frac{\sigma_0}{\sqrt{n}}$ і $2 \cdot \hat{O}(\delta) - 1 = \gamma$. Визначимо з цих двох рівнянь δ і ε . Зрозуміло, що $\delta = t_{(\gamma+1)/2}$ – квантиль рівня $(\gamma+1)/2$ нормального розподілу, який знаходиться за відповідними таблицями. Тоді $\varepsilon(\gamma) = t_{(\gamma+1)/2} \cdot \sigma_0 / \sqrt{n}$. Отримані ДІ для параметра μ наступного виду: $J_\gamma(\mu) = (\underline{\mu}, \bar{\mu}) = (\bar{x} - \varepsilon, \bar{x} + \varepsilon)$,

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ і $\varepsilon(\gamma) = t_{(\gamma+1)/2} \cdot \sigma_0 / \sqrt{n}$.

Висновки і аналіз отриманих результатів. Отже, на основі запропонованої ідеї аналізу звукових сигналів було зроблено моделювання у математичній програмі Матлаб. Записані два різних аудіо-сигнали аналізувалися в математичному пакеті. Зробивши ШПФ спектрів обох сигналів, виділили одну характерну частоту на 7 кГц.

На цій частоті отримали огинаючу сигналу, вибравши частоту дискретизації в 3.5 кГц з міркувань, що були зазначені вище. Наступний етап полягав в обчисленні інтегральних показників обох сигналів та їх порівняння.

Наведемо отримані результати у вигляді таблиць та графіків:

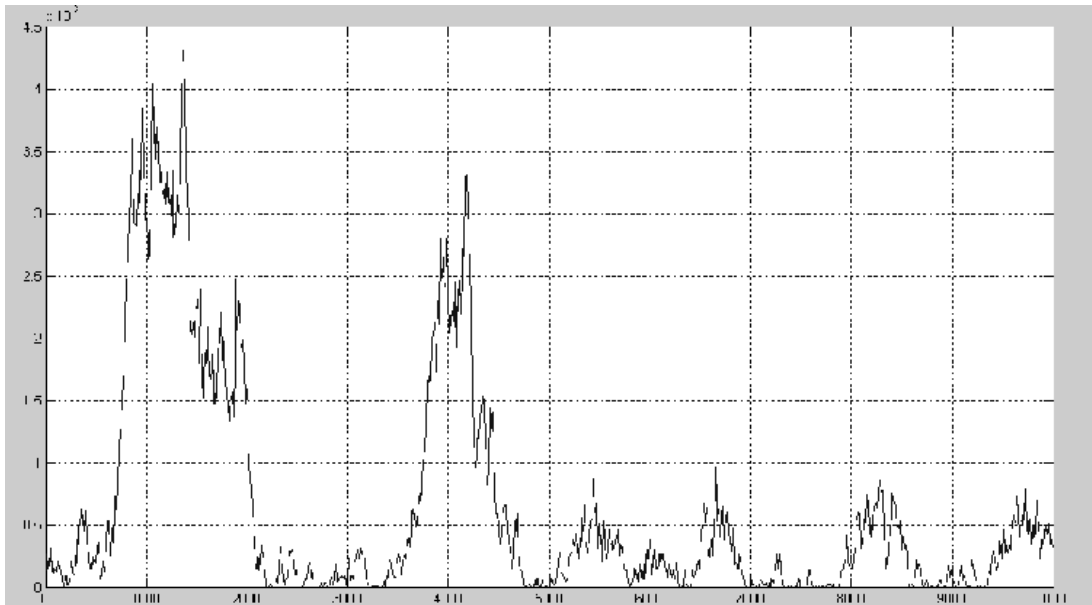


Рис.3. Спектр першого сигналу

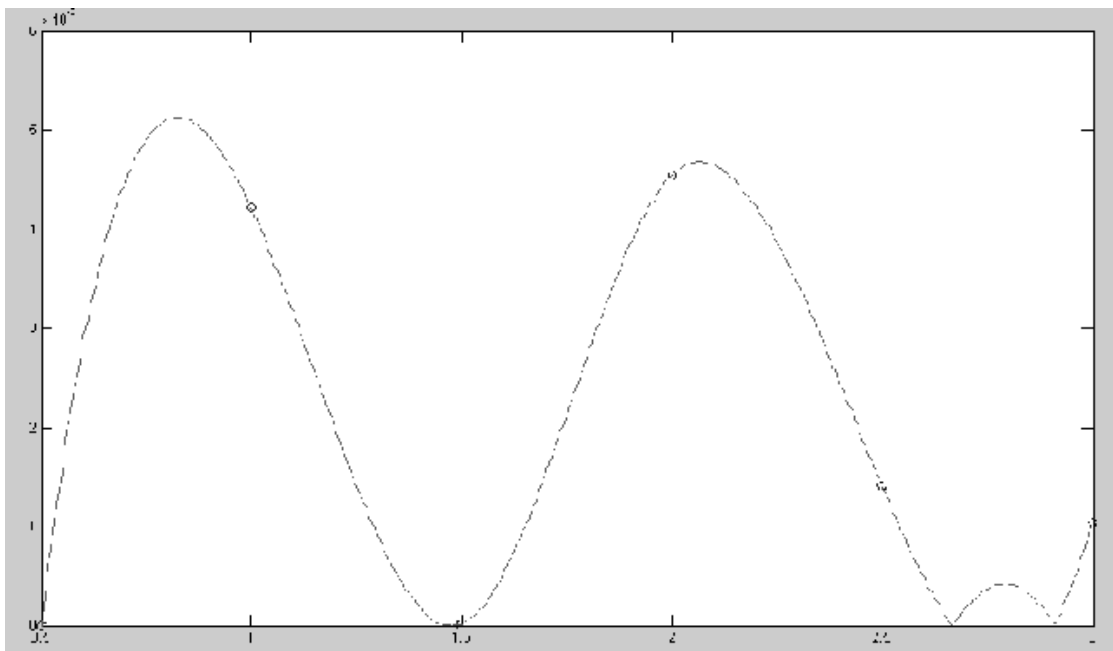


Рис. 4. Огинаюча на частоті 7 кГц з $F_d = 3.5$ кГц

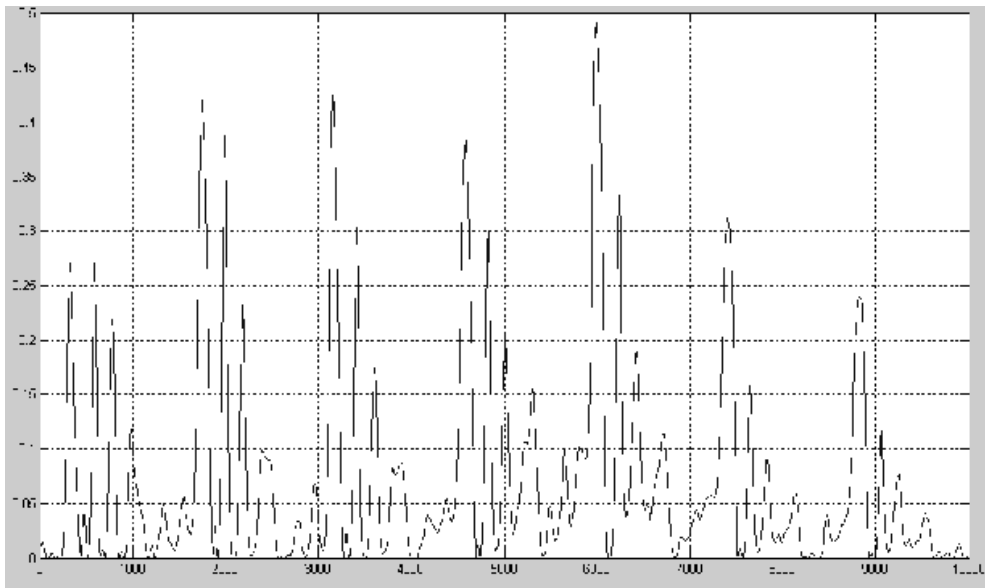


Рис. 5. Спектр другого сигналу

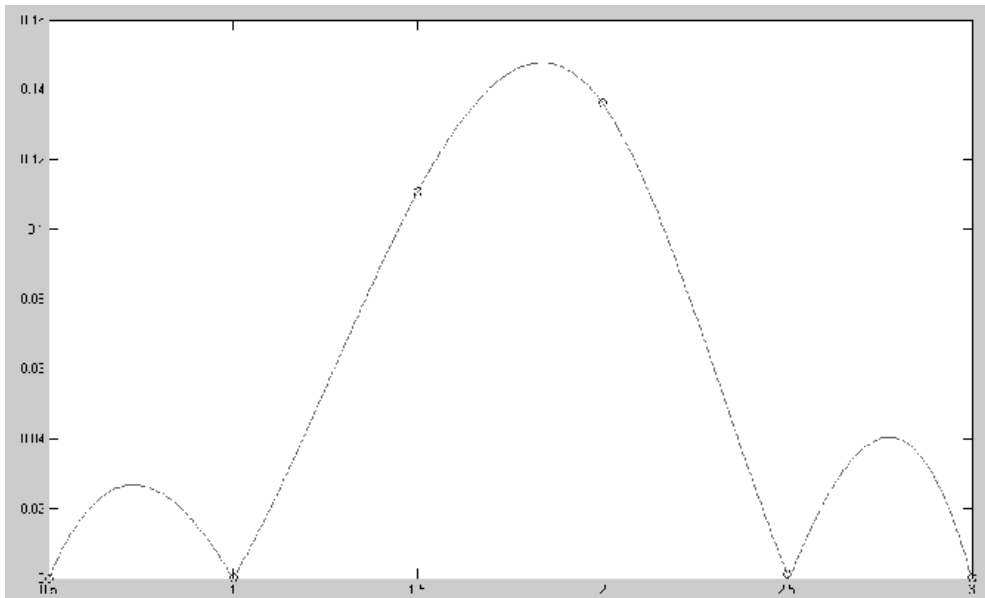


Рис. 6. Огибающая на частоте 7 кГц з $F_d = 3.5$ кГц

Таблиця 1. Дані обчислених інтегральних показників. Взято перші 5 значень

сигнал №2	0.00138821	0.00273562	0.0040425	0.00530921	0.00653632
сигнал №1	0.00020015	0.00039561	0.0005864	0.00077257	0.00095417

Згідно з отриманими даними інтегральних показників сигнали у нас різні.

Критерієм оцінки сигналу на схожість запропоновано точність значення "I" до десятого знаку.

Дана методика дозволяє аналізувати звукові сигнали з досить великою точністю. Використання ШПФ дозволяє економніше використовувати ресурси комп'ютера. Тому даний алгоритм можна впровадити для створення відповідного програмного забезпечення для аналізу звукових сигналів.

1. Гольденберг Л.М. и др. Цифровая обработка сигналов: Справочник. – М.: Радио и связь, 1985.
2. Гольденберг Л.М. и др. Цифровая обработка сигналов: Учебное пособие для вузов. – М.: Радио и связь, 1990. – 256 с.
3. Кибзун А.И., Наумов А.В. Теория вероятности и математическая статистика. – Москва, 2002. – 224 с.
4. Коленов С.О. Методичні вказівки до курсу лабораторних робіт по цифровій обробці інформації.
5. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. – М. Радио и связь, 1981. – 160-162 с.
6. http://ru.wikipedia.org/wiki/Динамическое_программирование.
7. http://uk.wikipedia.org/wiki/Штучні_нейронні_мережі.
8. http://ru.wikipedia.org/wiki/Скрытая_марковская_модель

Надійшла до редколегії 15.09.11