

УДК 621.39

О. М. Ткаченко, к. т. н., доц.; Л. В. Крупельницький, к. т. н., доц.; С. В. Хрущак**АДАПТИВНЕ ВИЗНАЧЕННЯ ГОЛОСОВОЇ АКТИВНОСТІ ПРИ
ДЕЛЬТА-УЩІЛЬНЕННІ МОВЛЕННЄВИХ СИГНАЛІВ**

Розроблено метод дельта-ущільнення з прогнозуванням наступного індексу квантованого значення в кодовій книзі та вдосконалено метод визначення голосової активності, описаний у рекомендації G.729, Annex B. Проведено експериментальне дослідження запропонованих методів.

Ключові слова: *ущільнення мовленнєвих сигналів, коефіцієнти лінійного прогнозування, лінійні спектральні частоти, векторне квантування, визначення голосової активності.*

Актуальність

При передачі мовлення цифровими каналами одним з найбільш важливих питань є обсяг інформації, який доведеться передавати за одиницю часу для забезпечення якісного голосового зв'язку. Ущільнення мовлення зменшує обсяги даних, що передаються, та апаратні витрати.

У цій статті розглянуто процес ущільнення мовленнєвих сигналів за допомогою вокодера LPC та запропоновано низку підходів для подальшого зменшення обсягів даних.

Вступ

Переважна більшість алгоритмів ущільнення мовлення використовують модель лінійного прогнозування для опису та кодування сигналу, відповідно до якої спектральна інформація про сигнал описується 10-ма коефіцієнтами лінійного прогнозування (LPC). Для більш ефективного квантування та подальшої інтерполяції коефіцієнти LPC, зазвичай, перетворюють на лінійні спектральні пари (LSP) та квантують їх за допомогою кодових книг.

Коефіцієнти LSP, отримані в результаті кодування мови за методом лінійного прогнозування, мають залежності між елементами як у середині фрейму, так і між сусідніми фреймами, що свідчить про високий рівень надлишковості під час використання цих коефіцієнтів [1].

Урахування кореляції між коефіцієнтами дозволяє зменшити обсяг даних, необхідних для опису параметрів мовленнєвого сигналу, і таким чином зменшити вимоги до пропускної здатності каналу передачі інформації. У цьому дослідженні розглянуто параметричну систему ущільнення мовлення, яка враховує кореляцію між коефіцієнтами всередині фрейму сигналу та між послідовними фреймами й таким чином дозволяє досягти більшого ущільнення.

Подальшим етапом зменшення обсягів даних, які необхідно передавати в канал зв'язку, є використання дискретної передачі (Discontinuous transmission – DTX) – вимкнення передавача або передача комфортного шуму (Comfort noise – CN) на приймальну сторону під час пауз у розмові. Такий підхід, наприклад, використовується в мережі GSM. Основним завданням при цьому є визначення голосової активності (Voice Activity Detection – VAD). Більшість підходів базуються на енергетичних порогах, визначенні періоду основного тону, спектральному аналізі, частоті переходів через нуль або комбінації кількох методів. Складність завдання полягає в тому, що постійної високої точності визначення вокалізованості фрейму сигналу неможливо досягнути, оскільки більшість алгоритмів ґрунтується на певному пороговому значенні, яке є фіксованим або обчислюється з сигналу на шумових сегментах. Так, наприклад, під час використання для VAD методу найменших

квадратів шумові сегменти використовуються у фільтрі з кінцевою імпульсною характеристикою для його тренування. Але у випадках, коли фоновий шум не є стаціонарним або коли мовленнєвий сигнал є переважно вокалізованим, для досягнення високої точності алгоритму VAD необхідно постійно коригувати порогове значення незалежно від типу сегменту. Для цього потрібен метод, який міг би ефективно адаптуватися до зміни фонового шуму.

У цій роботі вдосконалено метод визначення голосової активності, запропонований міжнародною асоціацією ІТУ-Т – G.729 Annex B [3], який розроблено для мультимедіа та ІР-телефонії. Цей метод забезпечує адаптацію до високого рівня фонового шуму та використовує дані, які отримують у процесі кодування, що зменшує його обчислювальну складність. Але, незважаючи на це, він має низьку швидкодію, порівняно з іншими стандартами (ETSI AMR, статистичні методи), особливо при низькому співвідношенні сигнал/шум. Тому запропоновано деякі зміни в цьому методі для покращення його швидкодії та проведено його дослідження з використанням запропонованих змін.

Урахування кореляції між коефіцієнтами

Урахування кореляції між коефіцієнтами одного фрейму відбувається за рахунок використання векторного квантування коефіцієнтів LSP. При векторному квантуванні набір коефіцієнтів LSP розглядається як один вектор, для якого відбувається пошук найближчого квантованого значення у спеціальній таблиці – векторній кодовій книзі [4].

Для врахування кореляції між послідовними фреймами в [4] запропоновано метод упорядкування векторів згідно з відношенням мажорювання, що дозволяє здійснити перехід до передачі різниці між індексами (дельти) послідовних фреймів. Недоліком запропонованого методу є значна затримка (100 мс), необхідна для досягнення якісного сигналу, при зменшенні обсягу інформації для опису параметрів мовленнєвого сигналу з 24 до 20 бітів, що неприпустимо в деяких системах зв'язку. З огляду на це, запропоновано метод дельта-ущільнення з прогнозуванням наступного відліку, який дозволяє отримати меншу затримку (20 мс) за умови меншої точності передачі індексу.

Для зменшення вікна дельти між індексами сусідніх фреймів при досить малій затримці, порівняно з методом запропонованим в [5], запропоновано прогнозування індексу фрейму на основі попередніх значень. Прогнозування відбувається шляхом створення екстраполюючої функції для індексів кількох фреймів на приймальній та передавальній стороні. При ущільненні на приймальну сторону також передається дельта між прогнозованим і реальним значеннями. Ілюстрацію запропонованого методу представлено на рис. 1.

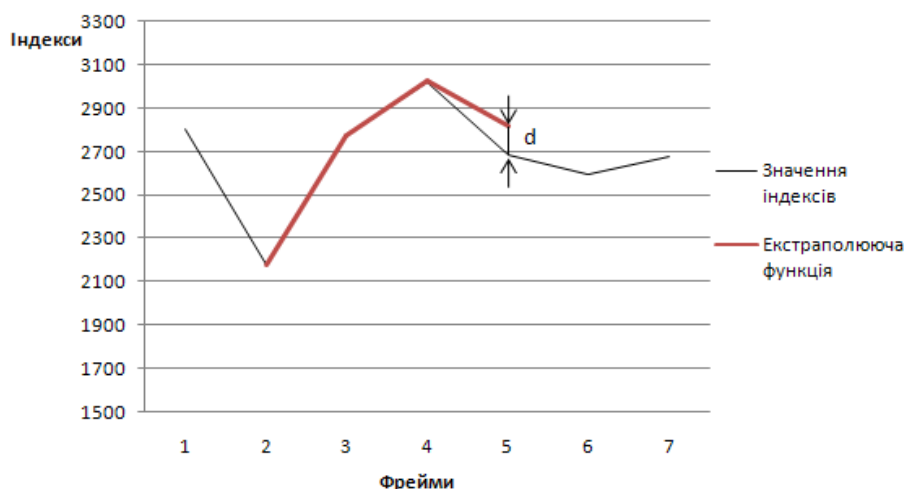


Рис. 1. Приклад роботи методу ущільнення з екстраполяцією

Екстраполюючу функцію будують за методом найменших квадратів. Суть цього методу полягає в тому, що функція для опису експериментальних даних g будується як лінійна комбінація M базисних функцій F_j [6]:

$$g = \sum_{i=0}^{M-1} c_j F_j(x). \quad (1)$$

При цьому коефіцієнти c_j обирають таким чином, щоб мінімізувати критерій $C_{\text{МКК}}$ – суму квадратів відхилень екстраполюючої функції від експериментальних значень:

$$C_{\text{МКК}} = \sum_{i=1}^n (y_i - g_i)^2 \rightarrow \min C, \quad (2)$$

де y_i – експериментальна функція.

Досліджено також альтернативні методи – метод найменших модулів (МНМ) та метод Чебишева. Для МНМ критерієм мінімізації є сума відхилення модулів (3):

$$C_{\text{МНМ}} = \sum_{i=1}^n |y_i - g_i|. \quad (3)$$

Для методу Чебишева критерієм мінімізації є максимальне відхилення [6]:

$$C_{\text{МЧ}} = \max |y_i - g_i|. \quad (4)$$

Проте тільки метод найменших квадратів дозволяє знайти найкращі c_j за кінцеве число операцій, зводячи завдання до вирішення системи лінійних рівнянь. Це означає, що він є найпростішим для обчислення, до того ж він точніше описує експериментальну функцію.

Введення затримки дозволяє коригувати значення індексів, що передаються; це необхідно в ситуаціях, коли прогноз виявляється неточним (рис. 2). Наприклад, у цьому разі ми маємо спрогнозувати індекс 5-го фрейму на основі попередніх – 2-го, 3-го та 4-го. Проте прогнозування тільки погіршило результат – значення дельти зросло.

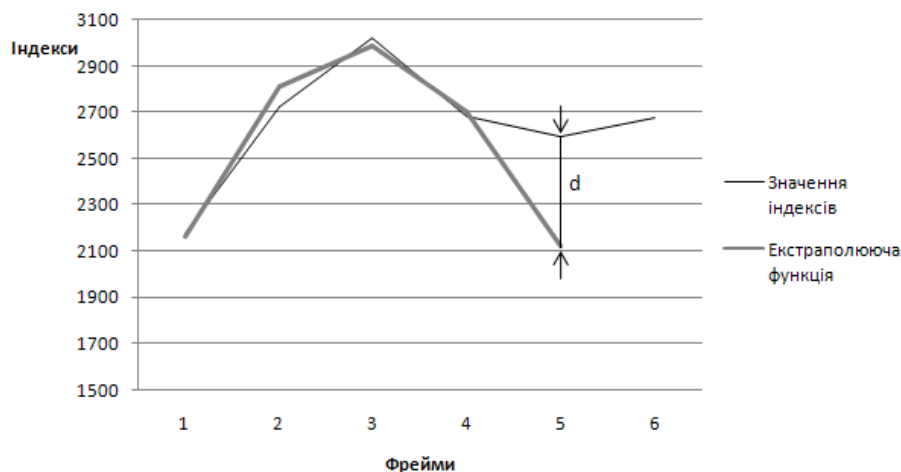


Рис. 2. Приклад неточного прогнозу

У випадках, коли при передачі параметрів i -го фрейму виконується нерівність (5):

$$\Delta_{i+1} > S_w, \quad (5)$$

де Δ_{i+1} – значення дельти для $i+1$ фрейму, S_w – розмір вікна застосовуємо такий алгоритм:

1. Здійснюється апроксимація значень експериментальної функції, урахувавши такі значення індексів $i+1$, $i+2$, ..., $i+k$, де k – величина затримки у фреймах. Для апроксимації також використовуємо метод найменших квадратів.

2. Здійснюється корекція значення i -го індексу відповідно до апроксимуючої функції, щоб

отримати кращий прогноз для наступного фрейму.

3. Здійснюється прогноз $i+1$ фрейму на основі попередніх фреймів $i, i-1, i-2, \dots, i-p$, де p – кількість фреймів для створення прогнозу, яка визначається порядком моделі прогнозування.

4. Якщо різниця між прогнозованим і реальним значенням виходить за межі вікна, то корекція проводиться знову: відбувається перехід до п. 2. У граничному випадку, коли при максимально припустимому відхиленні i -го фрейму прогноз є неточним, то відбувається перехід до п. 5.

5. Різниця між коригованим значенням та прогнозованим, урахувавши значення індексів для $i-1, i-2, \dots, i-(p+1)$ фреймів, передається в канал зв'язку.

Результат роботи цього алгоритму можна бачити на рис. 3. Після корекції значення індексу для 4-го фрейму прогноз відбувається успішно – значення дельти залишається незначним.

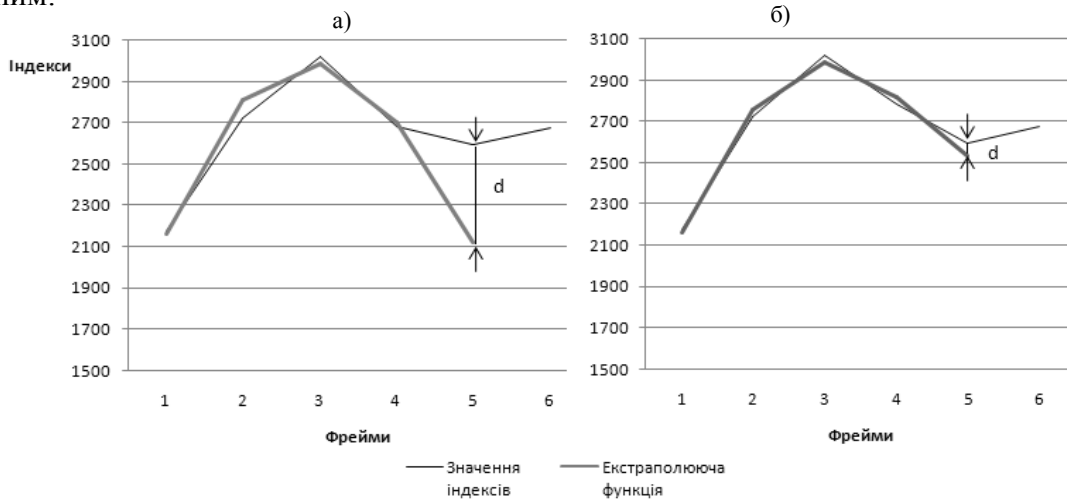


Рис. 3. Графіки експериментального та екстрапольованого значення індексів:
а) до виконання корекції;
б) після виконання корекції

Покращення алгоритму визначення голосової активності

Рекомендацію ITU-T G.729 Annex B розроблено як додаток до вокодера G.729. Рішення про вокалізованість приймають для фреймів тривалістю 10 мс на основі таких характеристик:

1. Диференційна енергія сигналу для всього діапазону спектру – ΔE_f ;
2. Диференційна енергія сигналу для низькочастотного спектру (0 – 1 кГц) – ΔE_l ;
3. Диференційна частота переходів через нуль – ΔZCR ;
4. Спектральне спотворення – ΔLSF .

Схему цього VAD наведено на рис. 4.

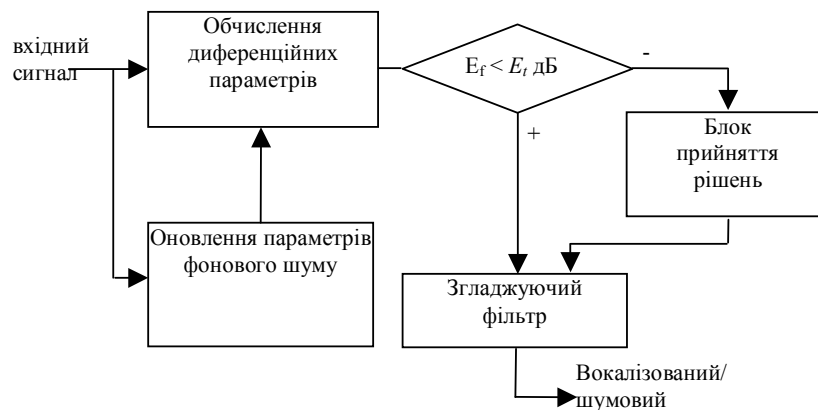


Рис. 4. Блок-схема алгоритму ITU-T G.729 Annex B

Вхідні параметри отримують з мовленнєвого сигналу використовуючи авторегресійну модель першого порядку, якщо виконується умова $\Delta E_f < E_t$, де E_t – задане порогове значення. Однак під час розрахунку цих параметрів виникають додаткові ускладнення. А використання частоти переходів через нуль і енергетичного порогу вносить похибки при високій зашумленості сигналу.

Для покращення якості запропонованого методу пропонуємо прискорити оцінку спектральної обвідної за рахунок згладжування за допомогою МНК. Завдяки цьому значно легше обчислити пологість спектру, яка є ключовою при визначенні вокалізованості фрейму сигналу (для невокалізованого фрагменту – спектральна обвідна буде більш пологою). Приклад такого згладжування спектральної обвідної зображено на рис. 5.

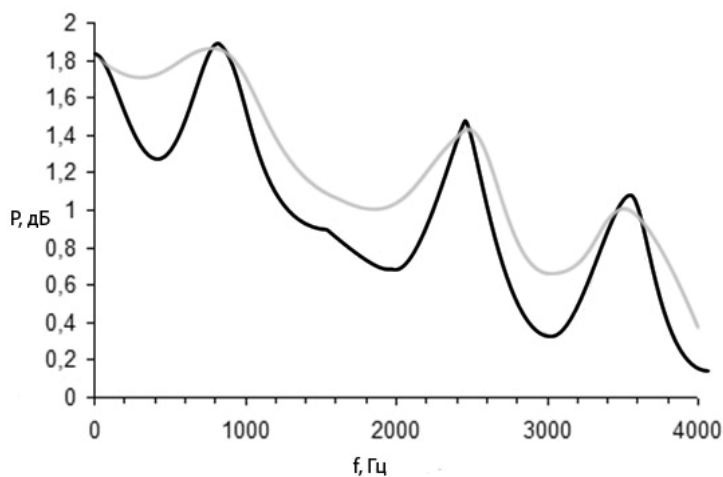


Рис. 5. Приклад згладжування спектральної обвідної сигналу

Отже, це дозволяє зменшити кількість обчислень, оскільки пологість спектру вираховується тільки за кількома точками, та підвищити точність алгоритму, особливо в низькочастотній смузі сигналу.

Для обчислення оптимального порогового значення вокалізованості сигналу застосовуємо новий підхід – геометрично адаптивний енергетичний поріг, запропонований Єзером та Таньєром [7]. Цей підхід базується на використанні ймовірнісного розподілу амплітуд і дозволяє встановлювати порогове значення адаптивно, використовуючи не лише шумові сигнали, але й вокалізовані. Цей метод є стійким до нестационарного шуму, але іноді може давати неправильний результат при різких й коротких стрибках шуму. Завдяки простоті обчислення адаптивного енергетичного порогу також зменшується складність алгоритму VAD.

Дослідження методу дельта-ущільнення

Тестування методу прискореного пошуку векторів відбувалося на основі фонетичного матеріалу, який складався з двох фонетично повних текстів [4]. Під час використання методу дельта-ущільнення з прогнозуванням значення індексу в кодовій книзі для наступного фрейму прогнозування відбувається шляхом створення екстраполюючої функції для індексів кількох фреймів. Досліджувалися декілька методів побудови екстраполюючої функції, а саме: метод найменших квадратів (МНК), метод Чебишева та метод найменших модулів (МНМ).

За критерії оцінювання методів використовувались дисперсія розподілу ймовірності дельти D між послідовними фреймами та спектральне спотворення SD . Результати оцінювання наведено в табл. 1.

Таблиця 1

**Результати дослідження методів екстраполяції для
кодової книги розміром 4096 векторів**

Назва методу	D	Середнє SD, дБ	Відсоток фреймів, для яких SD>2 дБ, %	Відсоток фреймів, для яких SD > 4 дБ, %
МНК	21531	0,882	1,78	0
Метод Чебишева	31095	0,912	1,96	0
МНМ	141950	0,956	2,15	0

Як можна бачити з таблиці 1, найкращим за показниками дисперсії кінцевого розподілу та спектральним спотворенням є МНК. Для МНМ кількість фреймів зі спектральним спотворенням у межах 2 – 4 дБ перевищує припустиме значення у 2%, прийняте для систем низькошвидкісного ущільнення мовлення.

Дослідження методу визначення голосової активності

Загальноприйнятою практикою оцінювання якості методів VAD є використання одного з двох видів тестів – об'єктивних або суб'єктивних. До суб'єктивних відносять методи, у яких робота алгоритму визначається за допомогою оцінок слухачів. Проте така схема не дозволяє визначити кількість неточних спрацювань при високому рівні шуму, оскільки це виражатиметься в збільшенні кількості даних у каналі, а не якості результуючого мовлення, тому було обрано об'єктивний метод оцінювання VAD. Для тестів було обрано мовленнєву базу ТІМІТ, у записи якої додано випадковий шум та поділено на фрейми, які потім розмічено вручну на вокалізовані та шумові, тобто фактично порівняння проводилось з ідеальним варіантом. Ці файли обробляли за допомогою різних алгоритмів визначення активності – запропонованого, оригінального, ETSI AMR 1 та ETSI AMR 2, після чого підраховували кількість фрагментів, на яких було обрізання мовлення та на яких шум було розпізнано як мовлення. Результати тестування наведено в таблиці 2.

Таблиця 2

Результати тестування якості

Назва методу	Фреймів з обрізанням мовлення, %	Шумових фреймів розпізнаних як вокалізовані, %
Покращений ITU-T G.729B	83%	2%
ITU-T G.729B	68%	5%
AMR 2	87%	0,3%
AMR 1	76%	5,5%

Результати тестування показали, що покращений метод ITU-T G.729B дозволяє отримати кращу якість розпізнавання мовлення на зашумлених фреймах, ніж звичайний метод, але дещо гіршу, ніж досягається в методі з використанням статистики високого порядку, який, проте, є більш обчислювально складним. Якщо враховувати, що в середньому при двосторонній бесіді кожен із співрозмовників розмовляє близько 40% [2], то цей метод може зменшити навантаження на канал зв'язку до 60%, хоча в реальних умовах цифра не перевищує 45%.

Для дослідження швидкодії запропонованого покращеного методу визначення голосової активності використовували матеріали з набору ТІМІТ, оскільки це дозволяє порівнювати отримані результати із закордонними аналогами. Результати порівняння покращеного методу ITU-T G.729 Annex B та відомих методів для різного розміру аналізованих блоків

наведено в таблиці 3.

Як видно з таблиці, удосконалений метод працює швидше, ніж оригінальний, і швидше, ніж метод, що базується на статистиці високого порядку. Метод з використанням адаптивних енергетичних рівнів є найшвидшим, але й найменш точним.

Таблиця 3

Середня кількість операцій для запропонованого та відомих методів при різному розмірі аналізованого блоку

Назва методу	Кількість математичних операцій для заданого розміру аналізованого блоку, мс			
	10	20	50	100
Покращений ITU-T G.729B	$1,7 \cdot 10^4$	$2,4 \cdot 10^4$	$5,2 \cdot 10^4$	$14,3 \cdot 10^4$
ITU-T G.729B	$2,2 \cdot 10^4$	$2,7 \cdot 10^4$	$6,1 \cdot 10^4$	$18,6 \cdot 10^4$
AMR 2	$2,9 \cdot 10^4$	$2,6 \cdot 10^4$	$4,8 \cdot 10^4$	$9,8 \cdot 10^4$
AMR 1	$1,6 \cdot 10^4$	$2,0 \cdot 10^4$	$3,9 \cdot 10^4$	$7,2 \cdot 10^4$

Висновки

Дослідження методу дельта-ущільнення з прогнозуванням показало, що його використання дозволяє зменшити кількість бітів для передачі мовленнєвих параметрів з 24 до 20 при затримці в 1 фрейм та незначному ($SD = 0,882$ дБ) спотворенні сигналу. Застосування дискретної передачі пакетів дозволяє додатково зменшити навантаження на канал на 45%. Запропонований удосконалений метод ITU-T G.729B дозволяє отримати кращу якість розпізнавання зашумлених фреймів, ніж звичайний, і меншу обчислювальну складність, ніж метод ETSI AMR 2.

СПИСОК ЛІТЕРАТУРИ

1. Chu W. C. Speech Coding Algorithms – Foundation and Evolution of Standardized Coders. / W. C. Chu. New Jersey: Wiley. – 2003. – 553 p.
2. Ramirez J. Voice Activity Detection. Fundamentals and Speech Recognition System Robustness, Robust Speech Recognition and Understanding. / J. Ramirez, J. M. Gorriz and J. C. Segura. Vienna: I-Tech Education and Publishing. – 2007. – 460 p.
3. Benyassine A. ITU-T recommendation G.729 annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data application. / A. Benyassine, E. Shlomot, and H.-Y. Su // IEEE Communications. – 1997. – № 35. – P. 64 – 73.
4. Ткаченко О. М. LSF-вокодер на основі векторного квантування / О. М. Ткаченко, Н. О. Біліченко, О. Д. Феферман, С. В. Хрущак // Реєстрація, зберігання і обробка даних. – 2007. – № 1. – С. 35 – 41.
5. Ткаченко О. М. Метод дельта-ущільнення мовних сигналів / О. М. Ткаченко, О. Д. Феферман, С. В. Хрущак // Інформаційні технології та комп'ютерна інженерія. – 2008. – № 1(11). – С. 8 – 13.
6. Коваленко І.Н. Теория вероятностей и математическая статистика: Учебное пособие. / И. Н. Коваленко, А. А. Филиппова. – М.: Высшая школа, 1982. – 256 с.
7. Özer H. A geometric algorithm for voice activity detection in nonstationary Gaussian noise / H. Özer and S. G. Tanyer // EUSIPCO, – 1998. – № 1. – P. 23 – 26.

Ткаченко Олександр Миколайович – доцент кафедри обчислювальної техніки.

Крупельницький Леонід Віталійович – доцент кафедри обчислювальної техніки.

Хрущак Сергій Вікторович – аспірант кафедри обчислювальної техніки.

Вінницький національний технічний університет.