УДК 057.087.1:621.391.26

ОБҐРУНТУВАННЯ ТА ВИБІР ПРОСТОРУ ПОПЕРЕДНЬОЇ ОБРОБКИ ГОЛОСОВОГО СИГНАЛУ В СИСТЕМІ АВТЕНТИФІКАЦІЇ

Н.Г.Б. КАМЕНІ, М.С. ПАСТУШЕНКО
Харківський національний університет радіоелектроніки

Abstract *– The article analyzes and investigates directions for improving the quality characteristics of voice authentication systems in various access systems. One of the main ways of improving the quality characteristics of these user authentication systems is related to the use of phase information of the voice signal. The urgent scientific task of researching new procedures for pre-processing the user's voice signal of the authentication system is being solved. Refinement of pre-processing procedures was carried out, considering the use of phase data of the voice signal. The developed pre-processing procedures make it possible to reduce the impact of random errors in the user's voice signal registration materials, which also appear due to the influence of external noise. The results are obtained in statistical analysis of simulation results using experimental voice data of the user of the authentication system. The phase space of the voice signal allows you to expand the possibilities of pre-processing due to the use of a priori information about the nature of the phase data change. The presented research results should be used in voice authentication systems, improving speech recognition systems, and solving speaker identification tasks.*

Анотація *– У статті аналізуються та досліджуються напрями підвищення якісних характеристик систем голосової автентифікації в різних системах доступу. Одним з основних напрямів підвищення якісних характеристик цих систем автентифікації користувача є використання фазової інформації голосового сигналу. Вирішується актуальне наукове завдання щодо дослідження нових процедур для проведення попередньої обробки голосового сигналу користувача системи автентифікації. Уточнення процедур попередньої обробки проводилося з урахуванням використання фазових даних голосового сигналу. Розроблені процедури попередньої обробки дозволяють зменшити вплив випадкових помилок у матеріалах реєстрації голосового сигналу користувача, які з'являються й за рахунок впливу зовнішнього шуму. Результати отримано в процесі статистичного аналізу результатів моделювання з використанням експериментальних голосових даних користувача системи автентифікації. Фазовий простір голосового сигналу дозволяє розширити можливості попередньої обробки за рахунок використання апріорної інформації про характер зміни фазових даних. Подані результати досліджень доцільно використовувати в системах голосової автентифікації, при вдосконаленні систем розпізнавання мови, а також під час вирішення завдань ідентифікації диктора.*

Вступ

Біометричні технології є перспективним напрямом у сфері інформаційної безпеки. Голосова біометрія на цей час широко поширена, і роботи над підвищенням якості голосових систем не втрачають своєї актуальності.

Водночас дослідження проводяться за такими основними напрямками:

- вдосконалення відомих і розробка нових методів вилучення мовних ознак;
- формування адекватних і надійних шаблонів користувача;
- розвиток процедур прийняття рішення на основі тих чи інших ознак.

Проте менше уваги приділяється попередній цифровій обробці матеріалів реєстрації голосового сигналу користувача. Зазвичай процедури попередньої цифрової обробки зводяться до нормалізації матеріалів реєстрації, що дозволяє істотно знизити вплив систематичних помилок. Однак в матеріалах реєстрації можуть бути випадкові й аномальні помилки, які істотно впливають на якість виконуваних у подальшому процедур і на результати автентифікації користувача.

Тому в даній роботі буде зроблено спробу усунення зазначеного пробілу в цифровій обробці голосового сигналу. Задача, що розглядається, досить складна, може, тому і відсутні фундаментальні роботи в цій галузі досліджень. З низки причин матеріали реєстрації будуть містити всі відомі типи помилок. Зараз на основі матеріалів реєстрації вирішуються всі завдання голосової автентифікації. Проте у сфері (просторі) амплітудно-частотних характеристик досить складно вирішити завдання попередньої обробки голосового сигналу.

Шляхи розв'язання зазначеної задачі можна спробувати шукати у фазовому просторі, оскільки, як буде показано нижче, фаза голосового сигналу має форму пилкоподібного сигналу невідомої тривалості. За такої умови випадкові помилки призводять до зміни форми пилкоподібного сигналу, а аномальні помилки (грубі промахи) призводять до різкої зміни тривалості. Аналогічні дослідження можна проводити у просторі аналітичної моделі голосового сигналу. Усунення випадкових і аномальних помилок з матеріалів реєстрації дасть можливість більш якісно вирішити завдання автентифікації при використанні навіть традиційних процедур, що базуються на основі обробки амплітудно-частотної інформації.

Таким чином, мета цієї роботи полягає в обґрунтуванні та виборі додаткового простору проведення попередньої обробки для зниження шуму в мовних сигналах стосовно мало досліджених в літературі принципів побудови процедур придушення шумової складової з урахуванням особливостей обробки матеріалів реєстрації в системах голосової автентифікації. Для досягнення зазначеної мети у роботі необхідно було вирішити такі завдання:

- проаналізувати роботи в галузі шумопридушення при обробці мовних сигналів;
- виконати аналіз і класифікацію шумової складової оброблюваних мовних сигналів;
- встановити простір, який є кращим для проведення попередньої обробки;
- розробити процедури для усунення основних складових шумових сигналів;
- провести математичне моделювання основних процедур компенсації шумових складових.

Теоретичні та експериментальні дослідження базуються на використанні апарату математичного аналізу, теорії та методів обчислювальної математики, теорії ланцюгів і сигналів, методів математичного моделювання.

І. Аналіз робіт у галузі досліджень

Методи придушення шуму використовуються для очищення аудіосигналів від зайвих звукових подій для подальшого повторного відтворення. При монтажі фільмів, музики, подкастів та інших медіа найчастіше потрібно позбавлятися зайвих звукових подій. За таких завдань може знадобитися загальне поліпшення якості запису. Це містить як видалення шуму, так і модифікацію сигналу, що може поліпшити сприйняття записаної промови. Подібні інструменти зазвичай доступні в редакторах аудіо та програмах-мікшерах для створення треків.

Наприклад, в одному з найвідоміших аудіоредакторів Audacity використовується підхід, який називається «шумові ворота» (noise gate), вірніше, їх конкретний спектральний різновид, що використовується після швидкого перетворення Фур'є. Крім цього, в Audacity є віконні механізми згладжування сигналу та видалення його невеликих артефактів. Інструменти в Audacity по шумопридушенню особливо добре підходять для відновлення мікрокасетних записів [1].

Популярним і складним завданням є шумопридушення на льоту – придушення шуму та відтворення одночасно із записом мови. Переслідувана мета – це маскування звуків, які мають відношення до інформації, що вимовляється людиною, і заважають її сприйняттю. Найчастіше таке придушення шуму використовується для аудіоконференцій у Skype, Zoom, Discord тощо. Під час придушення шуму на льоту зазвичай використовуються ті ж принципи «шумових воріт», але крім цього застосовуються методи машинного навчання для очищення сигналу на льоту. Наприклад, компанія Microsoft за результатами змагання DNS-Challenge [2] адаптувала найкращі рішення під свої розробки Skype та Teams. Ці рішення засновані на рекурентних нейронних мережах з Long short-term memory (LSTM) блоками і на згорткових нейронних мережах [3, 4]. В результаті новітні версії Skype та Teams здатні в режимі реального часу транслювати чистий голос за наявності агресивних шумів: робота дриля, вентилятора чи вітру.

Третя важлива область використання методів придушення шуму – передобробка та очищення звукового сигналу перед застосуванням методів автоматичного розпізнавання мови, щоб результат генерувався правильно. У цій галузі багато підводного каміння, оскільки сигнал не повинен містити штучних артефактів мови, інакше таке «очищення» може погіршити результат. Наприклад, в [5] уточнюється, що системи шумопридушення на основі масок не здатні покращити результат розпізнавання мови і тільки погіршують метрики через неприродність спектральних характеристик підсумкового сигналу. З іншого боку, алгоритми щодо поліпшення сигналу на основі глибоких нейронних мереж показали непоганий результат в ході попередньої обробки в пайплайні розпізнавання мови.

Найпростіші традиційні методи шумопридушення, про які йшлося вище, використовуються в умовах, коли програмно не встановлено, який характер шуму та мови. Така відсутність інформації також спостерігається, коли хочемо позбавлятися шуму на льоту. При такому шумопридушенні використовуються звичайні або спектральні пороги – заглушуються будь-які звуки, якщо вони не перевищують певного порога гучності.

В основі інших традиційних методів лежить моделювання розподілу чистої мови або шуму. Робиться це за допомогою знаходження спектральної щільності потужності (гучності) сигналу. Щільність потужності сигналу – варіант опису розподілу значень сигналу в різні моменти часу. Спектральна щільність потужності сигналу – функція, яка визначає розподіл потужності сигналу залежно від частоти, саме – можливу потужність у різні одиниці частоти. У такому випадку, маючи спектральну

щільність потужності шуму, можна використовувати метод спектрального відніманння (spectral subtraction).

Серед відомих методів шумопридушення слід виділити вінеровське оцінювання. Вінерівське оцінювання (Wiener filter) використовується як один з традиційних способів шумопридушення, що навчаються. Він частково схожий на метод спектрального віднімання. Цей підхід заснований на оптимальному підборі такого фільтра, який би мінімізував різницю між чистим сигналом і поліпшеним сигналом. Подібно деяким алгоритмам машинного навчання, в процесі обчислення вінеровського фільтра мінімізується метрика Mean Square Error (MSE). Робота фільтра описується наступним співвідношенням

$$H(w) = \frac{P_{ss}(w)}{P_{yy}(w)} = \frac{P_{ss}(w) - P_{dd}(w)}{P_{yy}(w)}, \quad (1)$$

де $P_{ss}(w)$ – спектр чистого сигналу, $P_{yy}(w)$ – спектр зашумленого сигналу, $P_{dd}(w)$ – спектр шумового сигналу.

Таким чином, оптимальний вінерівський фільтр можна знайти у випадках, коли відома «чиста версія» зашумленого сигналу або якщо відомий конкретний шум, який зустрічається в аудіозаписах і який хочемо прибрати.

Найчастіше після операцій з фільтрації шуму застосовується згладжування, щоб позбавитися артефактів сигналу – «музичного» шуму – після чищення. Для згладжування застосовуються різні фільтри, наприклад, Гаусовий фільтр (або розмиття Гауса) [6].

Наведені нижче алгоритми використовуються як для розмежування спікерів або інструментів, так і для придушення шуму. При шумопридушенні важливо позначити, що шум і чиста мова – два незалежні процеси, які виникають одночасно, як два окремі інструменти в музичній композиції. Залежно від способу розв'язання задачі шумопридушення розмежування спікерів або поліпшення сигналу алгоритми машинного навчання можна розділити на дві категорії, які представлені в табл. 1.

Таблиця 1. Приклади алгоритмів розв'язання задач розрізнення спікерів

	На основі масок	Генеративні
Опис	Передбачають маски для кожного спікера, інструменту або чистого сигналу. Ці маски накладаються на оригінальний сигнал.	Передбачають новий сигнал кожного спікера, інструменту або чистого сигналу.
Приклади	Conv-TasNet	DEMUCS, Wave-U-Net, HiFi-GAN

Спочатку в нейромережних підходах також використовувалися методи накладання масок на спектрограму в поєднанні з прямим і зворотним перетвореннями Фур'є. Однак підходи, які ґрунтуються на маскування спектрограм, мають деякі недоліки. Наприклад, фаза хвилі в чистому сигналі може відрізнятись від фази хвилі в зашумленому сигналі. Тому навіть у разі обчислення ідеальної маски для спектрог-

рами відновлена із зашумленого сигналу фаза може вносити якісь елементи шуму і псувати підсумкову якість шумопридушення.

Ще одним недоліком такої системи є складність обчислення частотних характеристик сигналу за допомогою швидкого перетворення Фур'є. Вікно для такого перетворення має бути досить великим для кращої якості декомпозиції на частоти, що збільшує кількість обчислень. Велика кількість обчислень призводить до низької швидкості роботи алгоритму, його стає складно застосовувати у реальному часі.

Одним із «проривних» підходів до нейромережного придушення шуму та покращення мовного сигналу виявився підхід на основі згорткових нейронних мереж Conv-TasNet. Багато сучасних підходів шумопридушення часто порівнюються з його архітектурою, як з однією з найбільш робастних реалізацій. Він ґрунтується на накладанні 1D згорток на чистий сигнал без розкладання на частоти.

Попередник цієї архітектури – TasNet [7]. Архітектура TasNet складається зі згорткових енкодерів і декодера з деякими особливостями:

- вихід енкодера обмежений значеннями від нуля до нескінченності [0, ∞);
- лінійний декодер конвертує вихід енкодера в акустичну хвилю;
- подібно до багатьох методів-попередників на основі спектрограм на останньому етапі система апроксимує функцію, що зважує (в даному випадку LSTM), для кожного моменту часу.

Conv-TasNet – модифікація алгоритму TasNet, яка використовує як функцію, що зважує, згорткові шари з розширенням (dilation). Ця модифікація була зроблена після того, як згортки з розширенням показали себе ефективним алгоритмом при одночасному аналізі та генерації даних змінної довжини, зокрема, для синтезу в таких рішеннях, як WaveNet [8].

Підхід для поділу аудіо/шумопридушення Conv-TasNet складається з 3-х компонентів: енкодер, поділ, декодер.

Аналіз показує, що основний компонент у схемі – етап поділу. Цей етап вирішує проблему наближеного обчислення джерел, суміш яких розглядають як «брудні» приклади. Формально припущення про «змішаність» сигналу можна виразити наступним чином

$$x(t) = \sum_{i=t}^{C} s_i(t), \tag{2}$$

де $x(t)$ – суміш у певний момент часу t, C – кількість джерел, що здійснюють внесок у суміш, $s_1(t),...,s_C(t)$ – джерела в певний момент часу t.

Завдання алгоритму машинного навчання – визначити джерела $s_1(t),...,s_C(t)$, знаючи заздалегідь кількість джерел C і суміш $x(t)$.

Варто відзначити, що поділ в алгоритмі відбувається не відразу, а тільки після отримання ознак сигналу за допомогою «1D блоків». Більш детально ознайомитися з алгоритмом та результатами експериментів можна у роботі [10]. Деякі підходи до розв'язання цієї задачі розглянуті в [11].

Однак у всіх розглянутих роботах як інформаційні параметри (простір аналізу) голосового сигналу використовують його амплітуду і частоту. Останнім часом з'явилася низка робіт, які додатково використовують фазові дані (фазовий простір) голосового сигналу [12-17]. Саме такий підхід будемо використовувати в даній роботі, який, як буде показано нижче, дозволяє розширити можливості вирішення завдань попередньої обробки голосового сигналу. Тому нижче розглянемо поняття аналітичного сигналу та можливості використання фазових даних під час вирішення розглянутих завдань.

II. Постановка завдання у загальному вигляді

Людська мова є шумоподібним акустичним сигналом, що несе амплітудну і частотну модуляції. Основна енергія акустичних коливань мовного сигналу укладена в діапазоні від 70 Гц до 7 кГц, причому більше 95% смислової інформації розміщується у вужчому діапазоні від 200 Гц до 5 кГц. Акустичні коливання вище і нижче цих частот несуть інформацію про емоції та особистості того, хто говорить, сприяють пізнаванності та дещо підвищують розбірливість мови в умовах підвищення шумів.

Шумові сигнали у системах автентифікації мають різну фізичну природу. Насамперед, це зовнішні шумові впливи, які можуть бути присутніми під час реєстрації мовного сигналу користувача. Крім цього, мають місце помилки вимірювань, які пов'язані з роботою апаратних засобів реєстрації.

У загальному випадку всі шумові впливи за характером прояву можна класифікувати на випадкові, систематичні та аномальні (грубі промахи вимірювань). Водночас шумові дії призводять до похибок (помилок) вимірів. У загальному випадку похибкою виміру називається відхилення результату виміру від справжнього значення вимірюваної величини.

Випадкові похибки та грубі промахи можна значною мірою усунути з результатів реєстрації за допомогою статистичної обробки отриманих результатів на основі теорії ймовірностей та математичної статистики.

Шум у мовному сигналі окреслює безладні коливання акустичних хвиль. Формально взаємодія корисного сигналу та шуму зазвичай описується в літературі наступною формулою

$$Y_n = S_n + N_n, \tag{3}$$

де n – часовий індекс, який визначається часом реєстрації t; S_n – корисний сигнал; N_n – шум; Y_n – суміш корисного сигналу та шуму, тобто сигнал із реальних умов запису.

Тепер можемо сформулювати завдання передобробки матеріалів реєстрації (шумопридушення або поліпшення голосового сигналу користувача) системи автентифікації: маючи зашумлений Y_n, потрібно знайти його оцінку, максимально наближену до вихідного сигналу S_n.

Аналізуючи формулу зворотного перетворення Фур'є, приходимо до висновку, що довільний сигнал $S(t)$ з відомою спектральною щільністю $S(w)$ можна записати як суму двох складових, кожна з яких містить або тільки позитивні, або тільки негативні частоти

$$Z(w) = S(w) + jS_m(w), \tag{4}$$

а у часовій області

$$Z(t) = S(t) + jS_m(t), \tag{5}$$

де S_m – уявна складова аналітичного сигналу.

На комплексній площині сигнал відображається вектором, модуль і фазовий кут якого змінюються в часі. Слід зауважити, що цей вектор обертається і швидкість обертання визначається частотою сигналу, що реєструється. Проекція аналітичного сигналу на мовну вісь у будь-який момент часу дорівнює вихідному сигналу $S(t)$, який реєструється в системах автентифікації за допомогою мікрофона. Щоб практично отримати сполучений сигнал, необхідно вихідне коливання $S(t)$ подати на вхід деякої системи, яка здійснює поворот фаз всіх спектральних складових на кут -90° області позитивних частот і кут 90° області негативних частот, не змінюючи його амплітуду.

Якщо раніше ці процедури в радіолокації та радіозв'язку реалізували апаратно, то зараз виконують програмно, використовуючи співвідношення (перетворення Гільберта)

$$S_m(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{S(\tau)d\tau}{t-\tau}, \tag{6}$$

де τ – змінна інтегрування.

У межах методу перетворення Гільберта огинаюча $U(t)$ довільного сигналу S(t) визначається як модуль відповідного аналітичного сигналу

$$U(t) = |Z(t)| = \sqrt{S^2(t) + S_m^2(t)}. \tag{7}$$

За визначенням повна фаза будь-якого сигналу $S(t)$ дорівнює аргументу аналітичного сигналу $Z(t)$

$$\varphi(t) = arctg\frac{S_m(t)}{S(t)}. \tag{8}$$

Як показано у [12-17], фаза голосового сигналу має пилкоподібну форму невідомої тривалості, амплітуда якого змінюється за лінійним законом у межах від 0 до

360 градусів. За наявності випадкових помилок вимірювання фази відхилятимуться від лінійного закону, а за наявності аномальних помилок фаза змінюється різко (більш ніж 10 градусів). Лінійний характер зміни фази можна використовувати в процесі попередньої обробки голосового сигналу системи автентифікації, як апріорну інформацію.

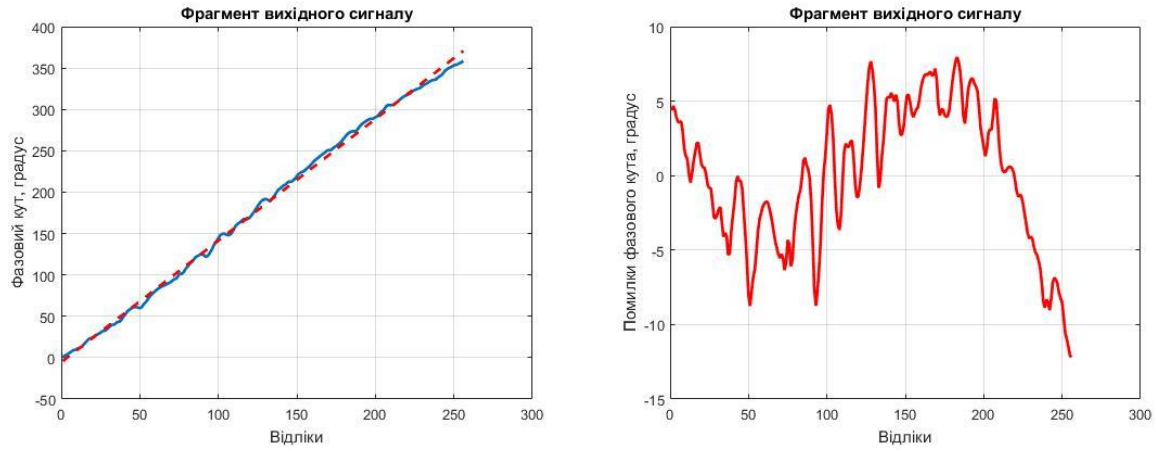
Тут же слід зазначити, що огинаюча (модуль аналітичного сигналу) на період зміни фази також змінюється незначно, а на комплексній площині траєкторія цього вектору може бути апроксимована кривою, близькою до кола.

У зв'язку з цим нижче проведемо цифрову попередню обробку голосового сигналу у фазовому просторі та просторі аналітичного сигналу, що огинає.

III. Результати досліджень у галузі фазового простору та у просторі огинаючої аналітичного сигналу

Як вихідні дані будемо використовувати матеріали реєстрації голосового сигналу з частотою дискретизації 64 кГц, який оброблявся за допомогою співвідношень (6)-(8). Далі з отриманих результатів виділявся фазовий пилкоподібний сигнал, один з періодів якого представлений на рис. 1, а суцільною лінією.

Для апроксимації даних у фазовому просторі природно використовувати метод найменших квадратів (МНК) з лінійною моделлю зміни фази. Результати такої апроксимації представлені на рис. 1. Графік вихідної (суцільна крива) і розрахованої залежності представлений на рис. 1, а, а помилки апроксимації на рис. 1, б.



а) б)

Рис. 1. Фрагменти вихідного сигналу до апроксимації фазового кута з використанням МНК

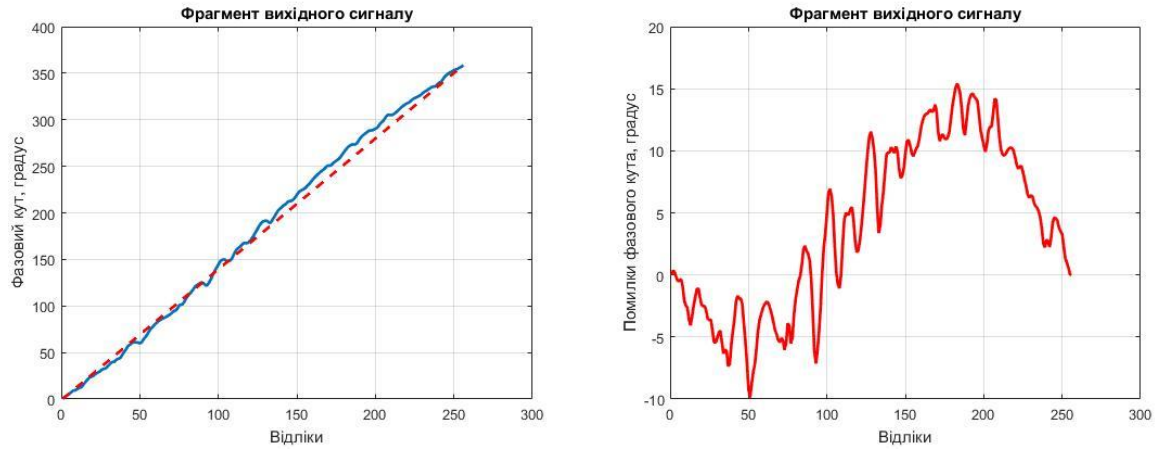
Середньоквадратичне відхилення досліджуваної апроксимації становить $4{,}75^{\circ}$, що є досить великим значенням. Особливо великі помилки мають місце на краях досліджуваної залежності, які виходять за межі зміни фази. Останнє є неприйнятним. Нормований коефіцієнт кореляції між вихідною та згладженою залежністю дорівнює 0,999. Однак дана процедура обробки має істотний недолік – результат обробки може виходити за межі інтервалу від 0° до 360°.

Тому нижче розглянемо апроксимацію фазового кута за допомогою прямої, яку проведемо через дві крайні точки (спочатку і наприкінці пилкоподібного фазового сигналу). Рівняння прямої в цьому випадку, як відомо, має такий вигляд

$$y = kx + b\,, \tag{9}$$

де $\frac{y_2 - y_1}{x_2 - x_1}$ – кут нахилу прямої, а $b = y_1 - kx_1$ – вільний член лінійного рівняння. Тут x_1, y_1 та x_2, y_2 – координати першої та другої точок у фазовому просторі, що використовуються у розрахунках.

Результат зазначеної апроксимації представлений на рис. 2, а, а помилки апроксимації на рис. 2, б. При цьому, як і раніше, суцільною кривою показані фазові дані за матеріалами реєстрації, а штриховою – дані апроксимації.



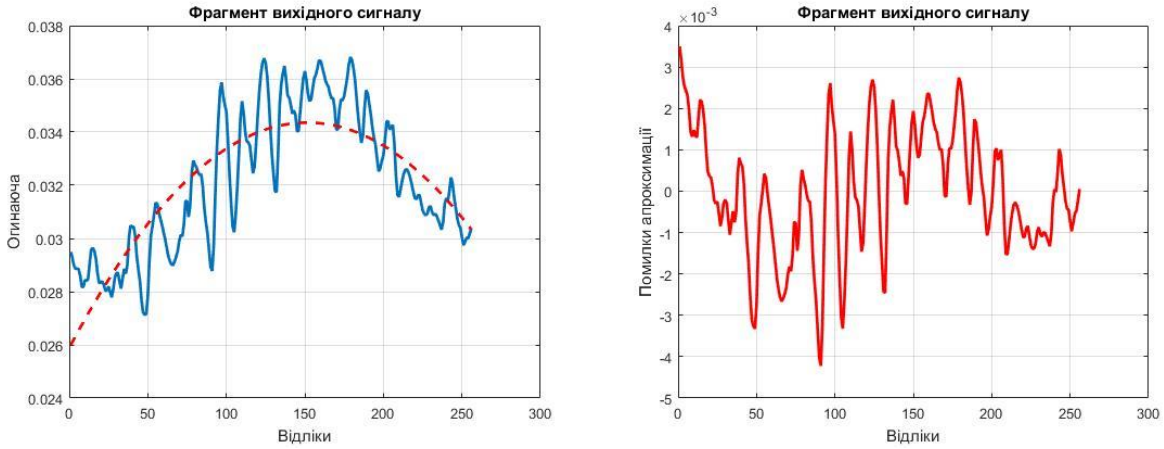
а) б)

Рис. 2. Апроксимація фазового кута лінійною моделлю

Як випливає з аналізу отриманих результатів, помилки на кінцях сигналу відсутні. Мають місце значні відхилення в першій і другій половинах сигналу з протилежними знаками. Середньоквадратичне відхилення досліджуваної апроксимації становить $3{,}9^\circ$, що менше порівняно з апроксимацією за методом найменших квадратів. Нормований коефіцієнт кореляції між вихідною і згладженою залежністю, як і раніше, дорівнює 0,999.

Таким чином, представлені результати дають право обрати як основний метод апроксимації фазового кута за допомогою прямої, яку проводимо через дві крайні точки (спочатку й у кінці пилкоподібного фазового сигналу). Зумовлено це тим, що МНК не забезпечує зміну фазового кута в заданих межах (від 0° до 360°). У той же час обраний метод апроксимації забезпечує виконання даної вимоги, але на результати апроксимації будуть впливати помилки, які мають місце у вимірах координат, що використовуються для розрахунку.

Тепер проаналізуємо простір зміни огинаючої (модуля вектору) аналітичного сигналу. На періоді зміни фазового кута огинаючу можна досить добре апроксимувати за допомогою полінома другого ступеня (параболи), характеристики якої отримані за МНК. На рис. 3 представлені результати такої апроксимації і помилки, що мають місце.



а) б)

Рис. 3. Апроксимація огинаючої (а) та помилки апроксимації (б)

На рис. 3, а суцільною лінією показана огинаюча аналітичного сигналу на одному періоді зміни фазового кута, розрахована з урахуванням матеріалів реєстрації голосового сигналу. Штриховою лінією показані результати апроксимації за допомогою полінома другого ступеня, коефіцієнти якого отримані з використанням МНК. Як випливає із результатів, представлених на рис. 3, б помилки незначні.

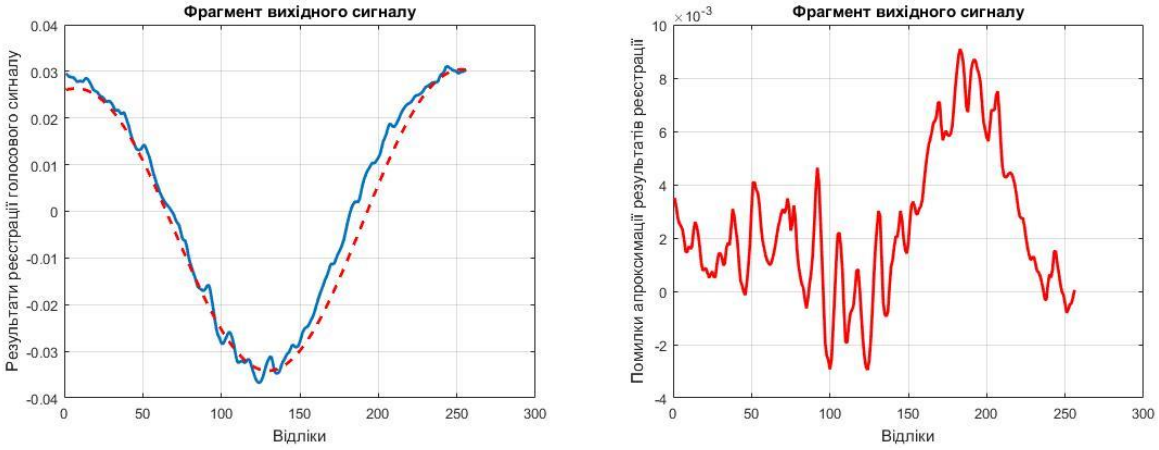
Можна надалі дослідити й апроксимацію траєкторії вектору аналітичного сигналу на комплексній площині. Але це вимагатиме значних обчислювальних ресурсів і буде надмірним, оскільки траєкторія вектору є наслідком раніше досліджуваних величин. А саме, в розрахунку траєкторії (координат x та y кінці поточного положення вектору аналітичного сигналу) використовується модуль цього вектору та фазовий кут, а саме

$$x(t) = U(t)\cos(\varphi(t)), \tag{10}$$

$$y(t) = U(t)\sin(\varphi(t)). \tag{11}$$

Тепер можна оцінити вплив згладжених даних на матеріали реєстрації та уявну складову голосового сигналу. Фактично, це реалізація співвідношень (10) і (11), і навіть порівняння їх з аналізованими залежностями на періоді зміни фази сигналу.

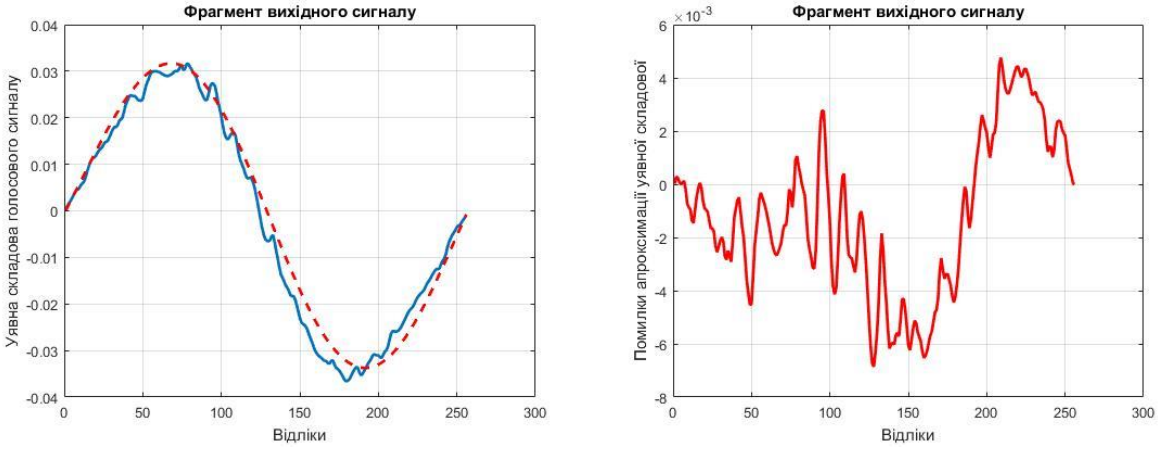
На рис. 4 представлені результати реєстрації голосового сигналу та апроксимація матеріалів реєстрації, а також помилки, що мають місце. Суцільною кривою на рис. 4, а показані матеріали реєстрації, а штриховою – результати розрахунків.



а) б)

Рис. 4. Апроксимація матеріалів реєстрації (а) та помилки апроксимації (б)

На рис. 5 представлені аналогічні результати для уявної складової.

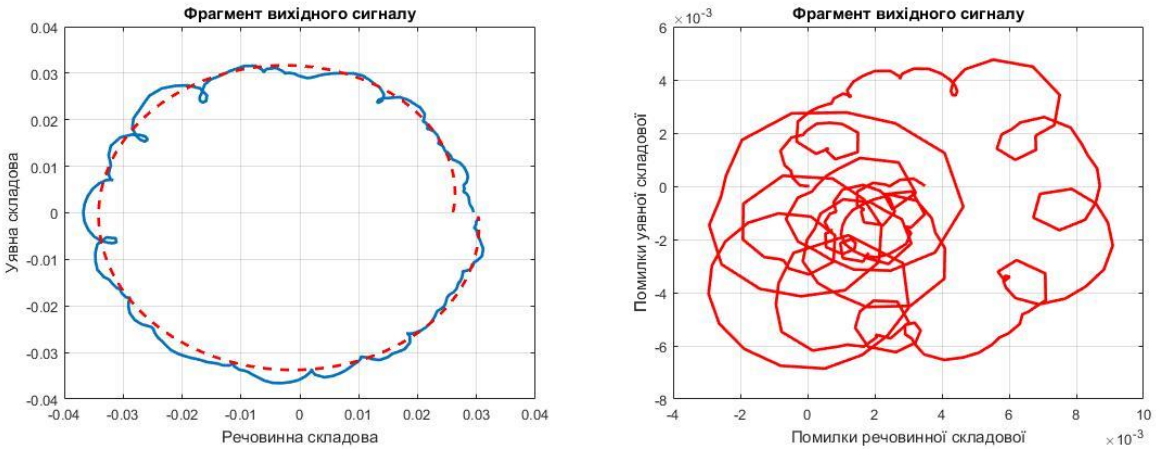


а) б)

Рис. 5. Апроксимація уявної складової (а) та помилки апроксимації (б)

На закінчення представимо результати аналізу траєкторії вектору аналітичного сигналу комплексної площині (див. рис. 6). Як і раніше, на рис. 6, а суцільною лінією показана траєкторія, розрахована за матеріалами реєстрації, а апроксимація – штриховою кривою. Звернімо увагу, що траєкторія апроксимуючої кривої в точці початку траєкторії не збігається за значеннями в точці її завершення. Очевидно, була велика фазова помилка або помилка в оцінці модуля огинаючої аналітичного сигналу на початку пи-

лкоподібного сигналу. Це дає право уточнити значення на початку пилкоподібного сигналу та повторити обчислення.



а) б)

Рис. 6. Апроксимація траєкторії вектору аналітичного сигналу на комплексній площині (а) та помилки апроксимації (б)

Використання отриманих результатів пов'язано, насамперед, з необхідністю формування фазової інформації голосового сигналу. Методика попередньої обробки голосового сигналу в цьому випадку може зводитися до наступного:

- формування уявної складової та фазових даних;
- виділення пилкоподібних фазових сигналів;
- проведення лінійної апроксимації пилкоподібного фазового сигналу та квадратичної апроксимації модуля (огинаючої) вектору аналітичного сигналу;
- уточнення матеріалів реєстрації та уявної складової аналітичного сигналу з урахуванням результатів розрахунків;
- перевірка правильності уточнення отриманих результатів розрахунку за допомогою траєкторії кінця вектору аналітичного сигналу.

Висновки

У статті розглядається проблема підвищення якості проведення попередньої обробки матеріалів реєстрації систем голосової автентифікації. В даний час, як показано в статті, попередня обробка проводитися у просторі амплітудно-частотних характеристик. Досягнення прийнятних результатів застосовуються нейромережні підходи.

Як основний напрямок вирішення зазначеної проблеми запропоновано в процесі цифрової обробки використовувати фазові дані аналізованого голосового сигналу. Достовірність запропонованого варіанта вирішення зазначеної проблеми та аналіз інформативності фазових даних голосового сигналу досліджується у процесі експериментальної обробки матеріалів реєстрації голосового сигналу. У зв'язку з цим у роботі розглядалася актуальна наукова задача щодо дослідження нових процедур

для проведення попередньої обробки голосового сигналу користувача з метою компенсації випадкових помилок.

Наукова новизна отриманих результатів полягає в тому, що вперше розроблено методику, процедури та проведено їх експериментальні дослідження протягом попередньої обробки голосового сигналу користувача з використанням простору фазових даних, а також простору зміни модуля та траєкторії вектору аналітичного сигналу. Крім цього, розроблені нові методи апроксимації фазових даних і огинаючої аналітичного сигналу на періоді пилкоподібної зміни фази. Результати отримані у процесі статистичного аналізу результатів моделювання з використанням експериментальних голосових даних користувача системи автентифікації.

Практичне значення отриманих результатів полягають у наступному:

- обґрунтовано та обрано додаткові простори проведення попередньої обробки голосового сигналу користувача;

- вибрано інтервал проведення апроксимації фазової інформації з урахуванням апріорних даних про характер її зміни;

- розроблено методику та виявлено особливості апроксимації фазової інформації досліджуваного голосового сигналу;

- проведено експериментальні дослідження компенсації випадкових помилок у матеріалах реєстрації голосового сигналу користувача.

Подальші дослідження доцільно проводити у напрямку оцінки якості формування ознак для традиційно використовуваних шаблонів (наприклад, кепстральних коефіцієнтів, мел-частотних кепстральних коефіцієнтів, коефіцієнтів лінійного передбачення тощо) з урахуванням проведення попередньої обробки на основі фази голосового сигналу.

Список літератури

1. How Audacity Noise Reduction Works – Audacity Wiki. URL: https://wiki.audacityteam.org/wiki/How_Audacity_Noise_Reduction_Works (дата звернення 15.09.2022)

2. *Chandan, K. A., Gopal, V., Cutler, R.* (2020), "The INTERSPEECH 2020 Deep Noise Suppression Challenge: Datasets, Subjective Testing Framework, and Challenge Results", P. 1-5. **DOI**: https://doi.org/10.48550/arXiv.2005.13981

3. *Hao, X., Su, X., Horaud, R., Li, X.* (2021), "FullSubNet: A Full-Band and Sub-Band Fusion Model for Real-Time Single-Channel Speech Enhancement", Proceedings of the ICASSP 2021 – IEEE International Conference on Acoustics, Speech, and Signal Processing, Jun 2021, Toronto, Canada. P. 1-5. **DOI**: https://doi.org/10.48550/arXiv.2010.15508

4. *Hu, Y., Liu, Y., Xing, M.* (2020), "DCCRN: Deep Complex Convolution Recurrent Network for Phase-Aware Speech Enhancement", INTERSPEECH, 20, October 25–29, 2020, Shanghai, China. P. 2472-2476. **DOI:** https://doi.org/10.48550/arXiv.2008.00264

5. *Kinoshita, K., Ochiai, T., Delcroix, M., Nakatani, T.* (2020), "Improving Noise Robust Automatic Speech Recognition with Single-Channel Time-Domain Enhancement Network", Proceedings of the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and

Signal Processing (ICASSP), 2020, P. 7009-7013. **DOI:** https://doi.org/10.1109/ICASSP40776.2020.9053266

6. Размытие по Гауссу. URL: https://ru.wikipedia.org/wiki/Размытие_по_Гауссу (дата звернення 15.09.2022)

7. *Luo, Y., Mesgarani, N.* (2018), "TASNET: Time-Domain Audio Separation Network for Real-Time, Single-Channel Speech Separation", Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), P. 696-700. **DOI:** https://doi.org/10.1109/ICASSP.2018.8462116

8. *Dieleman, S., Zen, H., Simonyan, K.* (2016), "WAVENET: A Generative Model for Raw Audio", 19 Sep 2016. P. 1-14. **DOI:** https://doi.org/10.48550/arXiv.1609.03499

9. *Luo, Y., Mesgarani, N.* (2019), "Conv-TasNet: Surpassing Ideal Time-Frequency Magnitude Masking for Speech Separation", Proceedings of the IEEE/ACM Transactions on Audio, Speech, and Language Processing. P.1-12. **DOI**: https://doi.org/10.1109/TASLP.2019.2915167

10. *Defossez, A., Synnaeve, G., Adi, Y.* (2020), "Real Time Speech Enhancement in the Waveform Domain", INTERSPEECH 20, October 25–29, 2020, Shanghai, China. P.1-5. **DOI:** https://doi.org/10.48550/arXiv.2006.12847

11. *Su, J., Jin, Z., Finkelstein, A.* (2020), "HiFi-GAN: High-Fidelity Denoising and Dereverberation Based on Speech Deep Features in Adversarial Networks", INTERSPEECH 20, October 25–29, 2020, Shanghai, China. P.1-5. **DOI:** https://doi.org/10.48550/arXiv.2006.05694

12. *Файзулаева, О.Н., Пастушенко, Н.С.* (2016), "Экспериментальные исследования амплитудного и фазового спектров речевого сигнала пользователя систем голосовой аутентификации", Проблеми телекомунікацій, No. 2(19), С. 28-34. URL: https://pt.nure.ua/wp-content/uploads/2020/01/162_pastushenko_speech.pdf

13. *Pastushenko, M., Faizulaieva, O.* (2016), "Employment of phase characteristics of user voice signal in authentication systems", Proceedings of the Third International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T), Kharkiv, P. 205-206. **DOI:** https://doi.org/10.1109/INFOCOMMST.2016.7905382

14. *Пастушенко, Н.С., Педро, В.Г., Файзулаева, О.Н.* (2018), "Исследование информативности фазовых данных голосового сигнала пользователя системы аутентификации", Проблеми телекомунікацій, No. 1(22), P. 67-74. **DOI**: https://doi.org/10.30837/pt.2018.1.05

15. *Pastushenko, M., Pastushenko, V., Pastushenko, O.* (2019), "Specifics of Receiving and Processing Phase Information in Voice Authentication Systems", Proceedings of the International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kyiv, Ukraine, 2019, P. 621-624. **DOI**: https://doi.org/10.1109/PICST47496.2019.9061260

16. *Pastushenko, M., Krasnozheniuk, Ya., Lemeshko, O.* (2020), "Analysis of voice signal phase data informativity of authentication system", Proceedings of the Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), Zaporizhzhia, Ukraine, April 27-May 1, 2020, P. 1040-1053. URL: https://ceur-ws.org/Vol-2608/paper78.pdf

17. *Pastushenko, M., Krasnozheniuk, Ya., Zaika, M.* (2020), "Investigation of Informativeness and Stability of Mel-Frequency Cepstral Coefficients Estimates based on Voice Signal Phase Data of Authentication System User", Proceedings of the 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), 2020, P. 467-471. **DOI**: https://doi.org/10.1109/PICST51311.2020.9468083