

УДК 057.087.1:621.391.26

ВПЛИВ ФАЗОВОЇ ІНФОРМАЦІЇ НА ПОПЕРЕДНЮ ОБРОБКУ ГОЛОСОВОГО СИГНАЛУ В СИСТЕМІ АВТЕНТИФІКАЦІЇ



[М.О. ПАСТУШЕНКО](#)

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»



[М.С. ПАСТУШЕНКО](#), [М.О. ПЕТРАЧЕНКО](#)

Харківський національний університет радіоелектроніки



[Т.А. ФАЙЗУЛАЄВ](#)

Національний аерокосмічний університет ім. М.Є. Жуковського "Харківський авіаційний інститут"

Abstract – The article analyzes and investigates directions for improving the quality characteristics of voice authentication systems in various access systems by improving the procedures for pre-processing registration materials. One of the main ways of improving the quality characteristics of user authentication systems, which was studied in the work, is the use of phase information of the voice signal. The actual scientific task of researching new procedures for pre-processing the voice signal of the user of the authentication system is being solved. The purpose of this work is to develop additional preprocessing procedures to reduce noise in voice signals of the authentication system. Refinement of pre-processing procedures was carried out based on the use of phase data of the voice signal. The results are obtained in the process of statistical analysis of simulation results using experimental model data 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 changes in phase data. The scientific novelty of the obtained results lies in the fact that for the first time, a technique was developed, and experimental studies were carried out for the pre-processing of the user's voice signal using the space of phase data. The practical significance of the obtained results is as follows: the phase information approximation interval was selected taking into account a priori data on the nature of its changes; an original linear approximation of phase data containing one harmonic of a voice signal is proposed; a mechanism for determining two harmonics in the phase data of a voice signal when using the proposed linear approximation is developed; the conducted experimental studies allow to develop a mechanism for compensation of random errors in registration materials. The presented research results are advisable for use in voice authentication systems, improvement of speech recognition systems, and solving speaker identification tasks.

Анотація – У роботі аналізуються та досліджуються напрями підвищення якісних характеристик систем голосової автентифікації в різних системах доступу за рахунок удосконалення процедур попередньої обробки матеріалів реєстрації. Одним з основних напрямів підвищення якісних характеристик систем автентифікації користувачів, який досліджувався в роботі, є використання фазової інформації голосового сигналу. Вирішується актуальне наукове завдання щодо дослідження нових процедур для проведення попередньої обробки голосового сигналу користувача системи автентифікації. Метою цієї роботи є розробка додаткових процедур попередньої обробки для зменшення шуму в голосових сигналах системи автентифікації. Уточнення процедур попередньої обробки проводилося на основі використання фазових даних голосового сигналу. Результати отримані в процесі статистичного аналізу результатів моделювання з використанням експериментальних модельних даних системи автентифікації. Фазовий простір голосового сигналу дозволяє розширити можливості попередньої обробки за рахунок використання апріорної інформації про характер змін фазових даних. Наукова новизна отриманих результатів полягає в тому, що вперше розроблена методика і проведені експериментальні дослідження проведення попередньої обробки голосового сигналу користувача з використанням простору фазових даних. Практичне значення отриманих результатів полягає в наступному: вибрано інтервал проведення апроксимації фазової інформації з урахуванням апріорних даних про характер її зміни; запропоновано оригінальну лінійну апроксимацію фазових даних, що містять одну гармоніку голосового сигналу; розроблено механізм визначення двох гармонік у фазових даних голосового сигналу під час використання запропонованої лінійної апроксимації. Проведені експериментальні дослідження дозволяють розробити механізм компенсації випадкових помилок у матеріалах реєстрації. Представлені результати досліджень доцільно використовувати в системах голосової автентифікації, удосконалення систем розпізнавання мови, а також при вирішенні завдань ідентифікації диктора.

Вступ

Біометричні технології є основою для побудови сучасних систем контролю та управління доступом (СКУД). В основі сучасних СКУД лежать такі біометричні ознаки: відбитки пальців (до 60%); геометрія обличчя 2D та 3D (до 20%); голос (до 10%). Голосова біометрія сьогодні широко поширена, і роботи з покращення її якості не втрачають своєї актуальності.

Дослідження проводяться в таких основних напрямках:

- покращення відомих та розробка нових методів виділення мовних ознак;
- формування адекватних шаблонів користувачів;
- розробка процедур прийняття рішень на основі певних характеристик.

Однак менше уваги приділяється попередній цифровій обробці матеріалів голосових записів користувача. Як правило, процедури попередньої цифрової обробки зводяться до фільтрації та нормалізації реєстраційних матеріалів, що може суттєво зменшити вплив систематичних помилок. Проте реєстраційні матеріали можуть містити випадкові та аномальні помилки, які мають значний вплив на якість подальших процедур та на результати автентифікації користувача.

Тому ця робота намагається заповнити цю прогалину в цифровій обробці голосових сигналів в системах автентифікації. Розглянута проблема є досить складною, що може бути причиною нестачі фундаментальних робіт у цій галузі досліджень. З ряду причин реєстраційні матеріали будуть містити всі відомі типи помилок, і наразі всі проблеми голосової автентифікації вирішуються на основі реєстраційних матеріалів. Однак у галузі амплітудно-частотних характеристик голосового сигналу, де проводиться попередня цифрова обробка, досить складно вирішувати її проблеми.

У роботі зроблено спробу пошуку шляхів вирішення розглянутої проблеми з урахуванням даних фазового простору, оскільки, як буде показано нижче, фаза голосового сигналу має форму пилоподібного сигналу невідомої тривалості. У цьому випадку випадкові помилки призводять до зміни форми пилоподібного сигналу, а аномальні помилки (грубі помилки) призводять до різкої зміни тривалості. Подібні дослідження можна проводити у просторі аналітичної моделі голосового сигналу.

Вилучення випадкових та аномальних помилок з реєстраційних матеріалів дозволить краще вирішувати завдання автентифікації при використанні навіть традиційних процедур, заснованих на обробці амплітудно-частотної інформації.

Таким чином, метою цієї роботи є розробка додаткових процедур попередньої обробки для зменшення шуму в голосових сигналах відносно принципів побудови процедур придушення шумової компоненти, які слабо вивчені в літературі, з урахуванням особливостей обробки реєстраційних матеріалів в системах голосової автентифікації. Для досягнення цієї мети необхідно було вирішити такі завдання:

- проаналізувати роботи в галузі зменшення шуму в обробці мовних сигналів;
- виконати аналіз та класифікацію шумової компоненти оброблених мовних сигналів;
- розробити процедури для вилучення основних компонентів шумових сигналів

з використанням фазових даних голосового сигналу;

–виконати математичне моделювання основних процедур компенсації шумових компонентів.

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

І. Аналіз робіт, пов'язаних з темою досліджень

Найпростіший спосіб боротьби з шумами, який може застосовуватися для СКУД, це використання додаткового мікрофона для реєстрації шумової складової. У цьому випадку можна з матеріалів реєстрації користувача відняти шумову складову. Але цей спосіб застосовується не у всіх випадках.

При реалізації процедури придушення шуму в масштабі реального часу, як правило, відсутня інформація про голосовий і шумовий сигнал. У такому випадку використовуються звичайні або спектральні пороги – будь-які відлуння приглушуються, якщо вони не перевищують певний поріг гучності.

Інші традиційні методи ґрунтуються на моделюванні розподілу чистого мовлення або шуму. Це здійснюється шляхом визначення спектральної щільності потужності (гучності) сигналу. Спектральна щільність потужності сигналу – це варіант опису розподілу значень сигналу на різних частотах в різні моменти часу. Спектральна щільність потужності сигналу, у свою чергу, є функцією, що описує розподіл потужності сигналу залежно від частоти, а саме можливої потужності в різних частотних одиницях. У цьому випадку, маючи спектральну щільність потужності шуму, можна використовувати метод спектрального віднімання.

Серед відомих методів зниження шуму варто виділити оцінку Вінера [1]. Фільтр Вінера використовується як один із традиційних методів зниження шуму, дещо подібний до методу спектрального віднімання. Цей підхід ґрунтується на оптимальному виборі фільтра, який би мінімізував різницю між чистим сигналом і покращеним сигналом. Як і деякі алгоритми машинного навчання, розрахунок фільтра Вінера мінімізує метрику середньої квадратичної помилки (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, w – кругова частота.

Таким чином, оптимальний фільтр Вінера може бути знайдений у випадках, коли відома "чиста версія" зашумленого сигналу, або якщо відомий конкретний шум, що виникає в аудіозаписах і який потрібно видалити.

Часто після операцій фільтрації шуму застосовується згладжування для позбавлення артефактів сигналу – «музичного» шуму – після очищення. Для згладжування

використовуються різні фільтри, наприклад, гаусівський фільтр (або Gaussian Blur) [2].

Останнім часом перевага надається автоматизованим системам обробки звукових файлів. Техніки зменшення шуму використовуються для видалення непотрібних звукових подій з аудіосигналів для подальшого відтворення. При редагуванні фільмів, музики, подкастів та інших медіа часто виникає потреба позбутися непотрібних звукових подій. Такі завдання можуть також вимагати загального покращення якості запису. Це включає не лише видалення шуму, але й модифікацію сигналу, що може покращити сприйняття записаного мовлення. Подібні інструменти зазвичай доступні в аудіоредакторах і програмах для мікшування треків.

Наприклад, один із найвідоміших аудіоредакторів, Audacity, використовує підхід, що називається «шумовий гейт», точніше, специфічний спектральний різновид цього підходу, який застосовується після швидкого перетворення Фур'є. Крім того, Audacity має механізми вікон для згладжування сигналу та видалення його невеликих артефактів. Інструменти зниження шуму Audacity особливо добре підходять для відновлення записів на мікрокасетах [3].

Популярним та складним завданням є зниження шуму в режимі реального часу – зниження шуму та відтворення одночасно з записом мовлення. Мета полягає в тому, щоб замаскувати звуки, які не пов'язані з інформацією, що вимовляється людиною, та заважають її сприйняттю. Найчастіше таке зниження шуму використовується для аудіоконференцій у Skype, Zoom, Discord тощо. При зниженні шуму в реальному часі, як правило, використовуються ті ж принципи «шумових гейтів», але додатково застосовуються методи машинного навчання для очищення сигналу в реальному часі. Наприклад, компанія Microsoft на основі результатів конкурсу DNS-Challenge [4] адаптувала найкращі рішення для своїх розробок Skype і Teams. Ці рішення базуються на рекурентних нейронних мережах з блоками довгої короткочасної пам'яті (Long Short-Term Memory, LSTM) та згорткових нейронних мережах [5, 6]. У результаті останні версії Skype та Teams здатні передавати чистий голос у реальному часі в умовах агресивного шуму, наприклад, під час роботи дрилі, вентилятора чи вітру.

Третя цікава сфера впровадження технік зниження шуму – попередня обробка та очищення аудіосигналу перед застосуванням автоматичних методів розпізнавання мовлення, щоб результат був сформований правильно. У цій сфері є багато підводних каменів, оскільки сигнал не повинен містити штучних мовних артефактів, інакше таке "очищення" може погіршити результат. Наприклад, у [7] уточнюється, що системи зниження шуму на основі масок не здатні покращити результат розпізнавання мовлення і лише погіршують метрики через неприродні спектральні характеристики кінцевого сигналу. З іншого боку, алгоритми покращення сигналу на основі глибоких нейронних мереж показали хороші результати при попередній обробці в системах розпізнавання мовлення.

На закінчення розглянемо нещодавні роботи [8-12]. У [8] розглядаються питання розділення мовлення та музики. Робота [9] присвячена розробці програмного забезпечення для обробки голосових сигналів, а [11] присвячена поділу обробленої послі-

довності на аудіо та не аудіо сегменти. У [11] аналізуються питання ідентифікації користувачів. Робота [12] має більш фундаментальний характер, у якій, як і у [8-11], детально розглядаються процедури попередньої обробки:

- виділення меж мовленнєвих сигналів;
- цифрова фільтрація;
- нарізка сигналу на кадри, що перекриваються;
- обробка сигналу у вікні;
- перетворення спектра;
- нормалізація спектра.

Однак всі розглянуті роботи використовують його амплітуду та частоти як інформаційні параметри (простір аналізу) голосового сигналу. Нещодавно з'явилася низка робіт, які додатково використовують фазові дані (фазовий простір) голосового сигналу [13-14]. Саме цей підхід будемо використовувати в цій роботі, що, як буде показано нижче, дозволяє розширити можливості для вирішення завдань попередньої обробки голосового сигналу. Тому нижче розглянемо концепцію аналітичного сигналу та можливість використання фазових даних при вирішенні розглянутих завдань.

II. Загальне формулювання проблеми

Людське мовлення є акустичним сигналом, подібним до шуму, який несе інформацію про амплітуду та частоту. Основна енергія акустичних коливань голосового сигналу знаходиться в діапазоні від 70 Гц до 8 кГц, причому понад 95% семантичної інформації розташовується в більш вузькому діапазоні від 200 Гц до 5 кГц. Акустичні коливання вище та нижче цих частот несуть інформацію про емоції та особистість мовця, сприяють впізнаванню та дещо підвищують розпізнавання мовлення в умовах підвищеного шуму.

Шумові сигнали в системах автентифікації мають різну фізичну природу. Насамперед це зовнішні шумові впливи, які можуть бути присутніми під час запису голосового сигналу користувача. Крім того, існують також похибки вимірювань, пов'язані з роботою апаратури запису.

Загалом, всі шумові впливи можна класифікувати за характером їх прояву на випадкові, систематичні та аномальні (помилки вимірювання). У цьому випадку шумові впливи призводять до неточностей вимірювання (похибок). У загальному випадку неточність вимірювання є відхиленням результату вимірювання від істинного значення вимірюваної величини.

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

Шум у голосовому сигналі визначається як випадкові коливання акустичних хвиль. Формально взаємодія корисного сигналу та шуму зазвичай описується в літературі наступною формулою

$$Y_n = S_n + N_n, \quad (2)$$

де n – це деякий індекс, визначений часом реєстрації t , S_n – корисний сигнал, N_n – шум, Y_n – суміш корисного сигналу і шуму, тобто сигнал з реальних умов запису.

Тепер можемо сформулювати задачу попередньої обробки матеріалів реєстрації (зменшення шуму або покращення голосового сигналу користувача) системи автентифікації: маючи зашумлений процес Y_n , потрібно знайти його оцінку, яка була б якомога ближчою до оригінального сигналу S_n .

Аналізуючи формулу оберненого перетворення Фур'є, приходимо до висновку, що довільний сигнал $S(t)$ з відомою спектральною щільністю $S(w)$ може бути записаний як сума двох компонентів, кожна з яких містить лише додатні або лише від'ємні частоти:

$$Z(w) = S(w) + jS_m(w), \quad (3)$$

і в часовій області

$$Z(t) = S(t) + jS_m(t), \quad (4)$$

де $S_m(w)$ ($S_m(t)$) – уявна складова в частотній (часовій) області аналізованого голосового сигналу.

На комплексній площині цей сигнал відображається як вектор, модуль та кут фази якого змінюються з часом. Варто зазначити, що цей вектор обертається, а швидкість обертання визначається частотою записаного сигналу. Проекція аналітичного сигналу на вісь реальних чисел у будь-який момент часу дорівнює оригінальному сигналу $S(t)$, який записується в системах автентифікації за допомогою мікрофона.

Для отримання спряженого сигналу на практиці необхідно застосувати оригінальне коливання $S(t)$ на вхід якоїсь системи, яка обертає фази всіх спектральних компонентів на кут -90° в області позитивних частот та на кут 90° в області від'ємних частот, не змінюючи їх амплітуду. Якщо раніше ці процедури в радіолокації та радіозв'язку реалізувалися в апаратурі, то зараз вони виконуються в програмному забезпеченні за допомогою відношення (перетворення Гільберта)

$$S_m(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{S(\tau) d\tau}{t-\tau}, \quad (5)$$

де τ – змінна інтегрування.

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

$$U(t) = |Z(t)| = \sqrt{S^2(t) + S_m^2(t)}. \quad (6)$$

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

$$\phi(t) = \arctg \frac{S_m(t)}{S(t)}. \quad (7)$$

Як показано раніше [13 - 14], фаза голосового сигналу має форму пилоподібного сигналу невідомої тривалості, амплітуда якого лінійно змінюється в діапазоні від 0° до 360° . За наявності випадкових помилок вимірювання фаза відхиляється від лінійного закону, а в разі наявності аномальних помилок фаза різко змінюється (більш ніж на 10°). Лінійний характер зміни фази може бути використаний в процесі попередньої обробки голосового сигналу в системі автентифікації.

Варто зазначити, що огинаюча (модуль аналітичного сигналу) під час періоду зміни фази також трохи змінюється, і на комплексній площині траєкторія цього вектору може бути апроксимована кривою, близькою до кола.

III. Результати попередньої обробки модельного сигналу в області фазового простору

У цій роботі основна увага буде приділена дослідженню гіпотези, яка висунута в [15]. Як досліджувані дані будуть використані матеріали модельних сигналів з частотою дискретизації 64 кГц, які були оброблені за допомогою співвідношень (5) – (7). Такий підхід дає можливість досліджувати різні випадки: аналізу піддається модельна послідовність, яка містить одну або дві частоти.

Надалі було виділено пилоподібний сигнал фази з отриманих результатів, один період якого показаний на рис. 1. При цьому частота модельного сигналу складала 400 Гц. Для виділення пилоподібних сигналів використовувалися процедури, описані в [13].

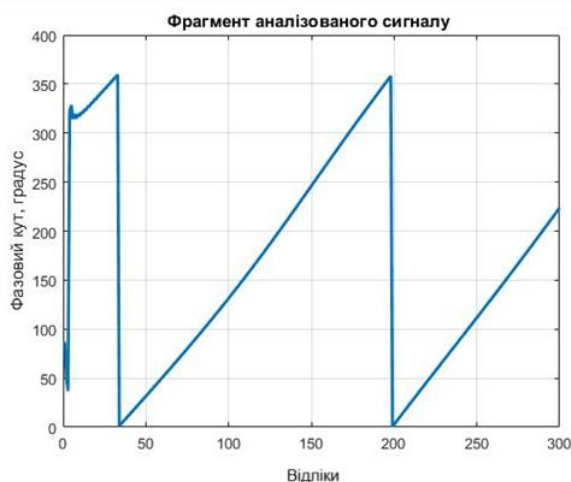


Рис. 1. Фрагмент даних фази (одна гармоніка)

Результати досліджень показали, що краще апроксимувати пилоподібний сигнал за допомогою рівняння прямої, яка проходить через дві крайні точки (початок і кінець пилоподібного сигналу фази) [15]. Рівняння прямої в цьому випадку має наступний вигляд

$$y = kx + b, \quad (8)$$

де $k = (y_2 - y_1)/(x_2 - x_1)$ – кут нахилу прямої, а $b = y_1 - kx_1$ – вільний член лінійного рівняння. Тут x_1, y_1, x_2, y_2 – координати першої і другої точок у фазовому просторі, які використовуються у обчисленнях (рис. 2).

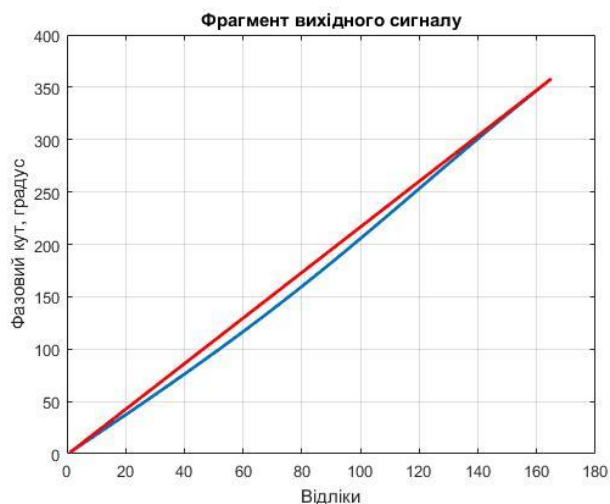


Рис. 2. Апроксимація фазового кута (одна гармоніка)

На рис. 2 червоним кольором наведена апроксимація фазового кута за допомогою рівняння (8). Далі обчислюється різниця між реальним пилоподібним сигналом (синя залежність) та його апроксимацією. Наступним кроком є оцінка помилок апроксимації. Для вирішення цього завдання можна використовувати критерій χ^2 , оцінку математичного очікування та середнє квадратичне відхилення (СКВ) помилок апроксимації.

У випадку, коли критерій χ^2 вказує на збіг між аналізованими помилками та рівномірним (гаусовським) розподілом, результати апроксимації можна використовувати замість матеріалів реєстрації. Це буде підтверджувати, що гіпотеза апроксимації за допомогою прямої вірна. Одночасно слід зауважити, що наявність незначної систематичної помилки приводить до того, що за критерієм χ^2 дані не збігаються.

Для набору статистичних даних було оброблено двадцять дев'ять реалізацій означених пилоподібних сигналів. В результаті отримано залежності критерію χ^2 , оцінок математичного очікування та СКВ помилок апроксимації від числа реалізацій. Отримані залежності наведені на рис. 3 – 5. Залежність оцінки критерію χ^2 отримані для числа ступенів свободи 29 і рівня значимості 0,001. Синім кольором позначено розрахункові значення, а червоним – табличні.

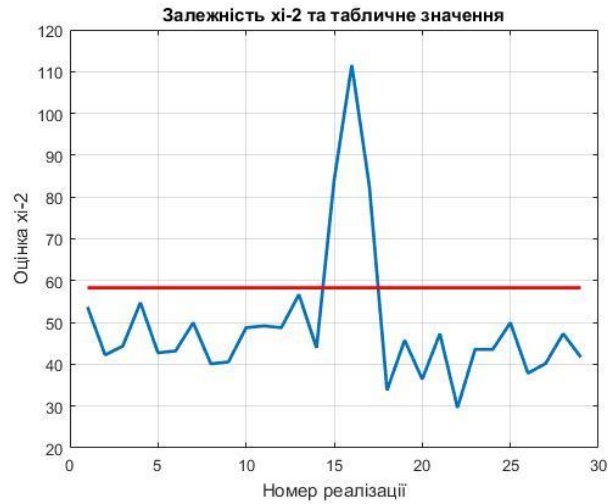


Рис. 3. Залежність оцінки критерію χ^2 (одна гармоніка)

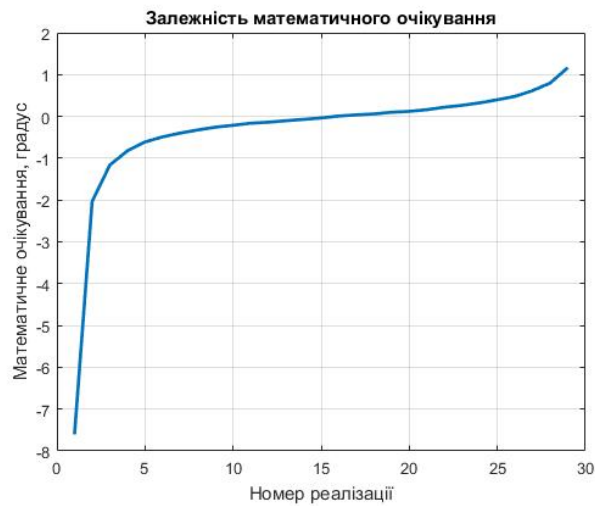


Рис. 4. Залежність оцінки математичного очікування (одна гармоніка)

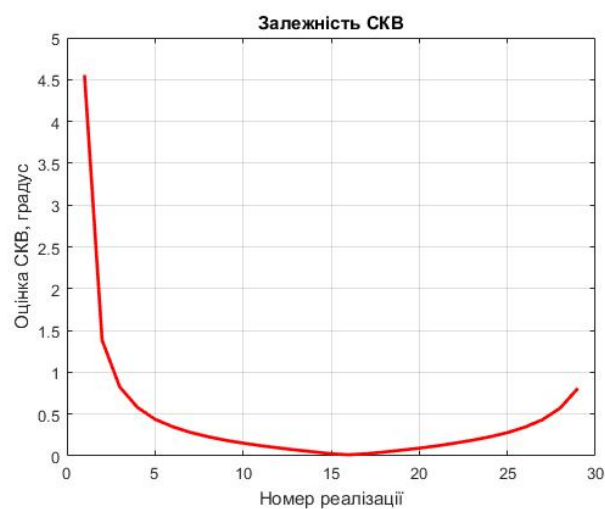


Рис. 5. Залежність оцінки СКВ (одна гармоніка)

На рис. 3 показано випадок, який було зазначено вище (дивись шістнадцяту реалізацію). У 16 реалізації помилки апроксимації за критерієм χ^2 не задовольняють прийнятним вимогам. Звернемо увагу, що оцінки математичного очікування та СКВ в означеному випадку близькі до нуля. Аналіз помилок апроксимації свідчить про наявність незначної систематичної помилки, яка впливає на результат розрахунку критерію χ^2 .

Тепер проаналізуємо залежності оцінок математичного очікування та СКВ. Звернімо увагу на початок і кінець аналізованих залежностей, де їхні величини зростають. Це зумовлено неточностями в результатах перетворення Гільберта на початку та в кінці аналізованої послідовності.

Тут же доцільно звернути увагу, що іноді на якість оцінки χ^2 впливає частота дискретизації. Наприклад для аналізованої частоти 300 Гц доцільно знизити частоту дискретизації до 32 кГц.

Тепер перейдемо до аналізу випадку, коли пилоподібний фазовий сигнал містить дві гармоніки 300 і 700 Гц. У цьому випадку у модельного сигналу амплітуда першої гармоніки дорівнювала одиниці, а другої гармоніки – 0.3. Форма фазового сигналу (синя залежність) та його лінійна апроксимація (червона залежність) представлена на рис. 6.

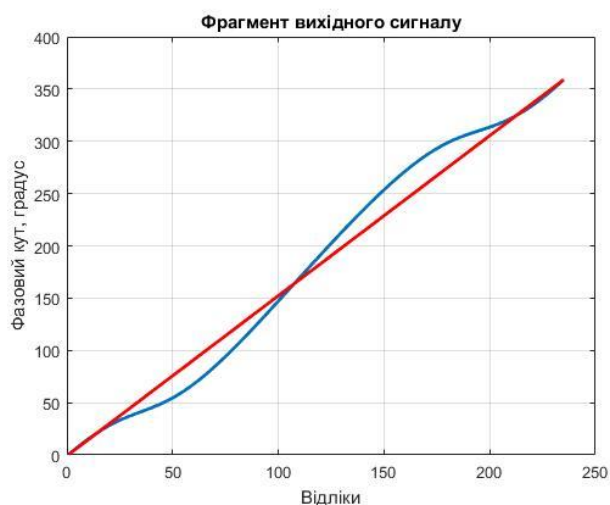


Рис. 6. Пилоподібний сигнал фази та його апроксимація (дві гармоніки)

В даному випадку фундаментальна гармоніка, яка має високу енергію, визначає тривалість пилоподібного сигналу. Друга гармоніка, як правило, має принаймні вдвічі менше енергії. Друга гармоніка призводить до відхилення форми сигналу від пилоподібної, як показано на рис. 6. У той же час, реєстраційні матеріали (сигнал двох гармонік) показані синім кольором, а фазовий сигнал фундаментальної гармоніки показаний червоним кольором (апроксимація).

Природно, що в цьому випадку критерій χ^2 не може допомогти в ідентифікації при використанні лінійної апроксимації. Це можна побачити на рис. 7. Розрахункові значення критерію, як і раніше, представлені синім кольором.

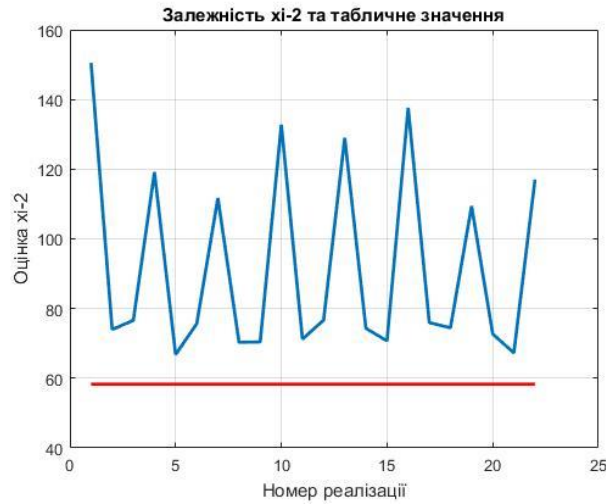


Рис. 7. Залежність оцінки критерію χ^2 (дві гармоніки)

Оцінки математичного очікування та СКВ представлені на рис. 8 та 9.

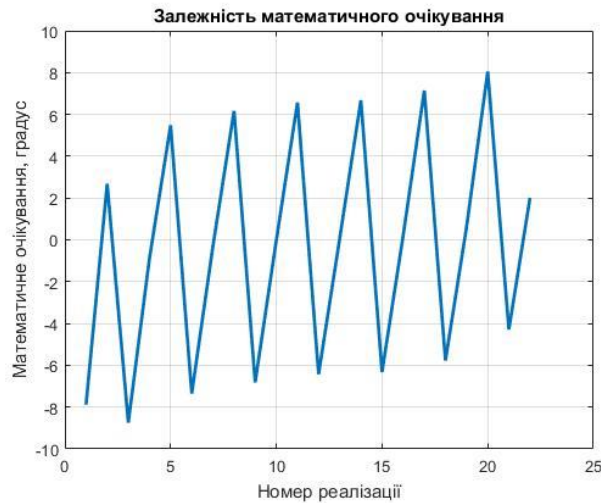


Рис. 8. Залежність оцінки математичного очікування (дві гармоніки)

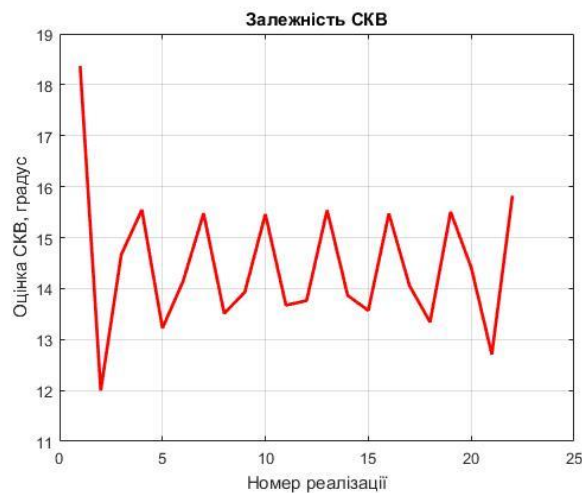


Рис. 9. Залежність оцінки СКВ (дві гармоніки)

Аналіз двох останніх залежностей та їхній порівняльний аналіз із залежностями, представлених на рис. 4 та 5, дає право зробити наступний висновок: для ідентифікації випадку наявності двох гармонік у фазовому сигналі при лінійній апроксимації можна використовувати оцінку СКВ. Якщо оцінка СКВ більше 10, то маємо випадок наявності у фазовому сигналі двох гармонік, інакше фазовий сигнал має одну гармоніку.

Висновки

Розглянуто актуальну проблему покращення якості попередньої обробки реєстраційних матеріалів для систем голосової автентифікації. На даний момент попередня обробка проводиться у просторі амплітудно-частотних характеристик.

Запропоновано використовувати фазові дані голосового сигналу як додатковий напрямок для вирішення цієї проблеми. Надійність запропонованого рішення та аналіз інформаційного вмісту фазових даних досліджено в процесі експериментальної обробки модельного голосового сигналу.

Наукова новизна отриманих результатів полягає в тому, що вперше була розроблена методологія та проведені експериментальні дослідження з попередньої обробки голосового сигналу користувача з використанням простору фазових даних. Результати були отримані в процесі статистичного аналізу результатів моделювання з використанням експериментальних голосових даних для системи автентифікації.

Практичне значення отриманих результатів таке: вибрано інтервал для апроксимації фазової інформації з урахуванням апріорних даних про характер її зміни; розроблена методологія та виявлені особливості апроксимації фазової інформації голосового сигналу, який аналізувався; визначено порядок ідентифікації в даних, які обробляються, однієї або двох гармонік сигналу.

Рекомендується проводити подальші дослідження у напрямку оцінки якості формування ознак для традиційно використаних шаблонів з урахуванням попередньої обробки на основі фазового сигналу голосу.

Список літератури

1. Lange, J., Lange, T. (2022), "Optimal Receive Filter (Wiener Filter)", *Focierier Transformation for Signal and System Description, Essentials*, Springer, Wiesbaden, P. 51–56. DOI: https://doi.org/10.1007/978-3-658-33817-6_9
2. Ifeachor, E.C., Jervis, B.W. (2001), *Digital Signal Processing: A Practical Approach, Second Edition*, Prentice Hall, 933 p.
3. How Audacity Noise Reduction Works, Audacity Wiki, P.1. URL: <https://support.audacityteam.org/troubleshooting/missing-features#audacity-wiki>
4. Reddy, C.K., Gopal, V., Cutler, R., Beyrarni, E., Cheng, R., Dubey, H., Matusovych, S., Aichner, R., Aazami, A., Braun, S., Rana, P. (2020), The INTERSPEECH 2020 Deep Noise Suppression Challenge: Datasets, Subjective Testing Framework, and Challenge Results. arXiv preprint arXiv:2005.13981. DOI: <https://doi.org/10.48550/arXiv.2005.13981>
5. 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, Toronto, Canada, 06-11 June 2021, P. 6633–6637. DOI: <https://doi.org/10.1109/ICASSP39728.2021.9414177>

6. Hu, Y., Liu, Y., Lv, S., Xing, M., Zhang, S., Fu, Y., Wu, J., Zhang, B., Xie, L. (2020), “DCCRN: Deep Complex Convolution Recurrent Network for Phase-Aware Speech Enhancement”, Proceedings of the Interspeech 2020, Shanghai, China, 25–29 October 2020, P. 2472-2476. DOI: <https://doi.org/10.21437/Interspeech.2020-2537>

7. 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 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), P. 1-5. DOI: <https://doi.org/10.1109/ICASSP40776.2020.9053266>

8. Nema B.M., Abdul-Kareem A. A. (2017), “Preprocessing signal for Speech Emotion Recognition”, Al-Mustansiriyah Journal of Science, No. 28(3), P. 157–165. DOI: <https://doi.org/10.23851/mjs.v28i3.48>

9. Niyozmatova, N, Mamatov, N, Nurimov, P, Samijonov, A., Samijonov, B. (2021), “Development Software for Preprocessing Voice Signals”, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, P. 53–66. URL: https://www.grinrey.com/bookstock/0501/02_0501/04_0502_Grinrey.pdf

10. Nematov, S., Kamolova, Y. (2022), “Preprocessing of speech signals for the systems of recognition and spectral analysis of speech,” Technical science and innovation, No. 1, 8, P. 209–216. DOI: <https://doi.org/10.51346/tstu-01.22.1-77-0163>

11. Nawas, K. Kh., Barik, M. K., Khan, A. N. (2021), “Speaker Recognition using Random Forest”, Proceedings of the International Conference on Innovative Technology for Sustainable Development (ICITSD-2021), No. 37, P. 1–5. DOI: <https://doi.org/10.1051/itmconf/20213701022>

12. Esquivel Jaramillo, A. (2021), “Pre-processing of Speech Signals for Robust Parameter Estimation”, Aalborg Universitetsforlag, PhD thesis, Aalborg Universitet, 188 p. DOI: <https://doi.org/10.54337/aau456472165>

13. Pastushenko, M., Pastushenko, V., Pastushenko, O. (2019), “Specifics of Receiving and Processing Phase Information in Voice Authentication Systems”, Proceedings of the 2019 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kyiv, Ukraine, 08 – 11 October 2019, P. 621–624. DOI: <https://doi.org/10.1109/PICST47496.2019.9061260>

14. 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, 27 April – 1 May 2020, P. 1040–1053. URL: <https://ceur-ws.org/Vol-2608/paper78.pdf>

15. Pastushenko M, Pastushenko M, Petrachenko M. (2023), “Clarification of Procedures for Voice Signal Pre-Processing in Authentication System”, Proceedings of the 2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT), Lviv, Ukraine, 21-25 November 2023, P. 1–5. DOI: <https://doi.org/10.1109/AICT61584.2023.10452693>