Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій
(повна назва)
Кафедра Інфокомунікаційної інженерії імені В.В. Поповського
(повна назва)

АТЕСТАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Дослідження процедур оцінки частоти основного тону в системах голосової автентифікації
(тема)

Виконав:
студент 2 курсу, групи АМСЗІм-18-1
Куценко Є.Є.
(прізвище, ініціали)

Спеціальність: 125 Кібербезпека
(код і повна назва спеціальності)
Тип програми: освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма: Адміністративний менеджмент у сфері захисту інформації
(повна назва освітньої програми)

Керівник: професор кафедри ІКІ ім. В.В. Поповського Пастушенко М.С.
(посада, прізвище, ініціали)

Допускається до захисту
Зав. кафедри _____ Лемешко О.В.
(підпис) (прізвище, ініціали)

2020р.

Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій
(повна назва)
Кафедра Інфокомунікаційної інженерії імені В.В. Поповського
(повна назва)
Рівень вищої освіти другий (магістерський)
Спеціальність 125 Кібербезпека
(код і повна назва)
Тип програми освітньо- наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Адміністративний менеджмент у сфері захисту інформації
(повна назва)

ЗАТВЕРДЖУЮ
Зав. кафедри_____
(підпис)
«____» _____2020р.

ЗАВДАННЯ
НА АТЕСТАЦІЙНУ РОБОТУ

студентові Куценко Єгор Євгенович .
(прізвище, ім'я, по батькові)

1. Тема роботи: Дослідження процедур оцінки частоти основного тону в системах голосової автентифікації.
затверджена наказом по університету від « 17 » березня 2020р. № 465 Ст.
2. Термін подання студентом роботи до екзаменаційної комісії 10.05.2020р.
3. Вихідні дані до роботи: ISO/IEC TR 24741:2007 Information technology – Biometrics tutorial (ГОСТ Р 54412-2011), ISO/IEC/TR 24722:2007 Information technologies. Biometrics. Multimodal and other multibiometric fusion. (ГОСТ Р 54411-2011)
4. Перелік питань, що потрібно опрацювати в роботі:
1) Аналіз поточного стану біометричних систем автентифікації користувачів.
2) Процедури цифрової обробки, математична модель та схема проведення експериментальних досліджень голосових сигналів.
3) Результати експериментальної оцінки частоти основного тону голосового сигналу користувача системи автентифікації.
4) Пропозиції щодо використання фазової інформації при вдосконаленні систем голосової автентифікації.

5. Перелік графічного матеріалу із зазначенням креслень, плакатів, комп'ютерних ілюстрацій: Демонстраційний матеріал у вигляді ppt-презентації; структурна схема біометричної системи; методика експериментальних досліджень; результати оцінки частоти основного тону голосового сигналу.

6. Консультанти розділів роботи

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		(підпис)	(дата)
Основна частина	професор Пастушенко Микола Савелійович		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	05.02.2020	Виконано
2	Збір матеріалів для дослідження	01.03.2020	Виконано
3	Розробка 1 розділу	05.04.2020	Виконано
4	Розробка 2 розділу	25.04.2020	Виконано
5	Розробка 3 розділу	05.05.2020	Виконано
6	Оформлення атестаційної роботи	10.05.2020	Виконано

Дата видачі завдання 5 лютого 2020 року

Студент _____ Куценко Є.Є.
(підпис) (прізвище, ініціали)

Керівник роботи _____ професор Пастушенко М.С.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 82 с., 20 рис., 4 табл., 2 додатки, 35 джерел

АВТЕНТИФІКАЦІЇ, АНАЛІЗ, БІОМЕТРІЯ, НАДІЙНІСТЬ, ПАРОЛЬ, СИСТЕМА ДОСТУПУ, ФАЗА СИГНАЛА.

Об'єктом дослідження є процес цифрової обробки сигналів в голосових системах автентифікації.

Метою даної роботи є дослідження напрямків підвищення якості голосових систем автентифікації на основі використання фазових даних.

Методи досліджень – аналіз, спостереження, вимірювання, моделювання та експеримент, узагальнення результатів і формування висновків.

В роботі виконано аналіз поточного стану систем голосової автентифікації користувачів. Розглянуто їх переваги та недоліки. Особливу увагу приділено використанню голосових систем автентифікації користувачів при побудові сучасних інформаційно-комунікаційних систем і локальних мереж, які мають суттєві переваги в порівнянні з випадком, коли в якості системи доступу використовуються інші методи автентифікації користувача.

Обґрунтовано, що основним напрямком підвищення якості систем голосової автентифікації є використання фазових даних оброблюваних матеріалів реєстрації. На прикладі частоти основного тону проведено дослідження фазового спектру голосового сигналу користувача, отриманого в процесі модельного експерименту. Виконані дослідження і розроблені процедури більш ефективної оцінки частоти основного тону, в тому числі й з використанням фазових даних голосового сигналу користувача системи автентифікації.

ABSTRACT

The report contains: 82 p., 20 pictures, 4 tables, 2 application, 35 sources

AUTHENTICATION, ANALYSIS, BIOMETRICS, PASSWORD, ACCESS SYSTEM, RELIABILITY, SIGNAL PHASE.

The object of the study is the process of digital signal processing in voice authentication systems.

The purpose of this work is to investigate the directions of improving the quality of voice authentication systems based on the use of phase data.

Research methods – analysis, observation, measurement, simulation and experiment, generalization of results and conclusions.

The current state of users' voice authentication systems is analyzed. Their advantages and disadvantages are considered. Particular attention is paid to the use of voice authentication systems for users in the construction of modern information and communication systems and local networks, which have significant advantages over the case when other methods of user authentication are used as the access system.

It is substantiated that the main focus of improving the quality of voice authentication systems is the use of phase data of the processed registration materials. Based on the example of the fundamental frequency, the phase spectrum of the user's voice signal obtained during the model experiment was investigated. Research has been performed and procedures have been developed to more effectively estimate the pitch of the pitch, including the use of the phase data of the voice signal of the user of the authentication system.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК, СИМВОЛІВ, ОДИНИЦЬ І ТЕРМІНІВ

БД – база даних
КПВД – коефіцієнт помилкового відхилення доступу
КПД – коефіцієнт помилкового доступу
МО – математичне очікування
СГА – система голосового автентифікації
СКВ – середньоквадратичне відхилення
СКМ – система комп'ютерної математики
ЧОТ – частота основного тону
ШПФ – швидке перетворення Фур'є
AFIS – Automated fingerprint identification systems
ANSI – American national standards institute
IBIA – International Biometric Industry Association
FAR – False Acceptance Rate
FRR – False Rejection Rate
GMM – Gaussian Mixture Model
HMM – Hidden Markov Models
SVM – Support Vector Machine

ВСТУП

В останні десятиліття досягнення науки і новітні технології, як ніколи раніше, визначають динаміку економічного зростання, рівень добробуту населення, конкурентоспроможність держави в світовому співтоваристві, ступінь забезпечення її національної безпеки і рівноправної інтеграції в світову економіку. Стрімкий розвиток і широке використання сучасних інформаційно-телекомунікаційних систем ознаменували перехід людства від індустріального суспільства до суспільства інформаційного, в основі якого лежать новітні системи комунікації. Кількість, технічний рівень і доступність інформаційних систем вже зараз визначають ступінь розвиненості країни і її статус у світовому співтоваристві, а в недалекому майбутньому, безсумнівно, стануть вирішальним показником цього статусу.

Інформаційні технології та комп'ютерні мережі пронизують всі існуючі системи та пов'язують державні служби, охорону здоров'я, науку, транспорт, зв'язок, банківський сектор, енергетику, оборону і промисловість. Небезпека кібератак, наслідком впливу, яких можуть стати не тільки матеріальні втрати, але екологічні, соціальні, макроекономічні потрясіння, особливо велика для організацій і об'єктів життєвої важливості.

Разом з тим, процес інформатизації світової спільноти породжує комплекс негативних явищ. Дійсно, висока складність і одночасно уразливість всіх систем, на яких базуються регіональні, національне і світове інформаційні простори, а також фундаментальна залежність від їх стабільності державних інфраструктур призводять до виникнення принципово нових загроз.

Стрімкий розвиток сучасних комп'ютерних мереж і особливо мережі Інтернет призвело до повсюдного використання розподілених інформаційних систем різного призначення. У всьому світі активно йде процес децентралізації обчислень і хмарних додатків, що дозволяє створювати масштабовані системи, здатні обслуговувати величезну кількість користувачів з високою продуктивністю.

У зв'язку з широким розповсюдженням розподілених систем у всіх сферах людської діяльності гостро стоїть завдання забезпечення інформаційної безпеки в таких системах. Однією з основних заходів щодо захисту даних є забезпечення надійної автентифікації користувача.

На даний момент існує безліч підходів до автентифікації і ще більше реалізацій цих підходів. При цьому не всі класичні рішення задачі автентифікації підходять для використання в розподілених системах. А різні типи систем пред'являють свої унікальні вимоги до підсистем автентифікації. Крім того, активний розвиток комп'ютерної техніки дозволяє легко зламувати алгоритми автентифікації, які ще 10-15 років тому вважалися надійними. Унікальними є приклади з України. Чим активніше українці користуються банківськими картами, тим частіше їх обманюють шахраї. За 2019 рік сукупний розрахунковий дохід карткових шахраїв в Україні зріс з 245,8 млн. грн до 361,99 млн. грн (на 47,3%), повідомила заступник директора Української міжбанківської Асоціації членів платіжних систем ЕМА Олеся Дальніченко. Якщо говорити про середню суму шахрайської операції, то вона підросла з 3620 до 6200 гривень.

У зв'язку з цим ведеться безперервна робота в області дослідження і розробки методів автентифікації. Постійно з'являються нові і удосконалюються існуючі алгоритми, спрямовані на забезпечення захищеної автентифікації користувачів.

Тому все більш актуальною стає проблема автентифікації користувачів, що мають доступ до громадських та особистих фінансових, інформаційних і обчислювальних ресурсів. Особливо важлива ця проблема для відкритих, масових телекомунікаційних та інформаційних систем.

Одне з найбільш перспективних напрямків захисту подібних систем від несанкціонованих впливів – використання біометричних методів ідентифікації користувачів. Однак, незважаючи на всю привабливість, даний підхід пов'язаний з низкою серйозних проблем. Тому останнім часом багато досліджень проводиться в області застосування біометричних систем автентифікації.

В останні роки технологією автентифікації особистості за його біометричними даними вважається досить перспективною. Одна з найважливіших завдань біометрії – створення технічних пристроїв, здатних дізнаватися конкретної людини за його неповторними біометричними характеристиками і з ще більш високою ймовірністю розпізнавати зловмисників, які намагаються маскуватися під легальних користувачів.

Актуальність теми обумовлена зростанням цінності інформації, постійною появою нових загроз інформаційній безпеці і важливістю процесу автентифікації при побудові захищеної інформаційної системи. Першим бар'єром в забезпеченні інформаційної захисту є система автентифікації. У даній роботі досліджуються

голосові системи автентифікації, які за критерієм ефективність/вартість, як свідчать результати наукових досліджень, є найбільш перспективними.

На даний момент не існує універсального рішення в області голосових системи автентифікації, які можуть забезпечити необхідний рівень якості голосової автентифікації.

Сьогодні визнаним лідером розробки та впровадження біометричних технологій є США. Початок досліджень в даній області було покладено ще в середині 1980-х років. З метою підтримки програм по біометрії уряд США в 1995 році створило біометричний консорціум (www.biometrics.org), куди увійшли державні та приватні організації, університети, дослідницькі центри, лабораторії тестування і сертифікації продуктів біометричних технологій. Зараз в нього входять приблизно 500 різних організацій.

Урядом США створений також Національний біометричний тестовий центр при університеті Сан-Хосе (www.engr.sjsu.edu/biometrics), в період з 1998 року по теперішній час організована підготовка фахівців з біометрії в п'яти різних університетах країни. Національні інститути стандартизації США (ANSI та NIST) за останні 10 років розробили близько 40 національних біометричних стандартів, більшість з яких в даний момент використовується як основа при розробці міжнародних біометричних стандартів спеціально створеним підкомітетом ISO/IEC JTC1 SC37.

Поштовхом до бурхливого розвитку таких технологій стали трагічні події 11 вересня 2001 року, які відбулися в США.

Паралельно з державним біометричним консорціумом США за підтримки урядів провідних країн утворена Міжнародна асоціація виробників засобів біометрії (International Biometric Industry Association, www.IBIA.org), куди входять 26 найбільших виробників біометричних пристроїв. Уряд США в 1998 році підтримало і створення BioAPI Consortium для розробки промислового стандарту інтерфейсів зв'язку (API) різних біометричних програмно-апаратних додатків.

Означене вище свідчить про значну увагу до розвитку біометричних технологій.

Аналіз документальних джерел (літературних і патентів) свідчить, що в сучасних голосових системах автентифікації в якості інформаційних параметрів сигналу використовуються амплітуда і частота реєстрованих даних.

Разом з тим, давно відомо, що найбільш інформативним параметром сигналу, в тому числі і голосового, є фазова інформація. При цьому використання фазо-

вих даних голосового сигналу може дати можливість підвищити якість систем автентифікації на їх основі.

Слід зауважити, що процедури автентифікації користувача за його голосовим сигналом безпосередньо пов'язані з алгоритмами розпізнавання мови, які з історичної точки зору стали розвиватися раніше, і в цій області отримані більш істотні результати. Тут же зауважимо, що в частотної області в зазначених задачах використовуються різні частотні діапазони голосового сигналу. Більш того, в ряді програм при голосовій автентифікації доводиться вирішувати і завдання розпізнавання мови, наприклад, при автентифікації з обмеженим словником.

Метою даної роботи є дослідження напрямків підвищення якості голосових систем автентифікації на основі використання фазових даних.

Об'єкт дослідження – процес цифрової обробки в голосових систем автентифікації.

При цьому доцільно вирішити такі завдання:

- дослідження документальних джерел за темою роботи;

- аналіз поширених підходів до цифровій обробці голосових сигналів і конкретних алгоритмів, що реалізують ці підходи;

- розробка методики формування фазових даних голосового сигналу і їх попередню обробку;

- формування пропозицій щодо використання фазових даних для підвищення якості голосових систем автентифікації.

Результати досліджень опубліковані в роботах[1 – 4].

1 АНАЛІЗ ОСНОВНИХ БІОМЕТРИЧНИХ СИСТЕМ АВТЕНТИФІКАЦІЇ КОРИСТУВАЧІВ І ПОСТАНОВКА ЗАВДАНЬ ДОСЛІДЖЕНЬ

Системи автоматичної автентифікації і ідентифікації особистості за біометричними параметрами є одним з напрямків сучасної науки та техніки, який активно розвивається. Особливе місце серед них займають системи, засновані на голосових біометричних ознаках, тому що в даному випадку для проведення автентифікації не потрібно безпосереднього контакту користувача з апаратурою. Тому системи автентифікації по голосу застосовні там, де використання інших методів практично неможливо, наприклад, для надання віддаленого доступу до послуг і даними по телефонних каналах або через Internet.

1.1 Загальний аналіз структури біометричних систем автентифікації

Біометричні параметри (ознаки) користувача діляться на два великі класи: статичні (фізіологічні) і динамічні (поведінкові). Фізіологічні біометричні параметри, такі, як відбитки пальців або геометрія руки, є фізичними характеристиками, які зазвичай вимірюються в певний момент часу. Поведінкові біометричні параметри, наприклад підпис або голос, являють собою послідовність дій і тривають протягом певного періоду часу.

Фізіологічні біометричні параметри досить різноманітні й одного їх зразка зазвичай буває досить для порівняння. Що стосується поведінкових біометричних параметрів, то окремий зразок може і не давати достатніх для ідентифікації особи відомостей, але саме часова зміна сигналу (під впливом поведінки) містить необхідну інформацію.

Фізіологічні (статичні) та поведінкові (динамічні) біометричні параметри взаємно доповнюють один одного. Основна перевага статичної біометрії є відносна незалежність від психологічного стану користувачів, малі витрати їх зусиль і, отже, можливість організації біометричної ідентифікації великих потоків людей.

Основні біометричні параметри, які широко використовуються в даний час представлені в таблиці 1.1.

Таблиця 1.1 – Основні біометричні параметри

Фізіологічні	Поведінкові
Відбитки пальців	Голос
Райдужна оболонка	Підпис
Обличчя особи	Хода особи

Біометричні параметри повинні мати наступні властивості [5], які дозволяють застосовувати їх на практиці.

1) Загальність: кожна людина має біометричні характеристики.

2) Унікальність: для біометрії немає двох людей, що володіють однаковими біометричними характеристиками.

3) Сталість: біометричні характеристики повинні бути стабільні в часі.

4) Вимірність: біометричні характеристики повинні бути вимірювані будь-яким фізичним зчитувальних пристроїм.

5) Прийнятність: сукупність користувачів і суспільство в цілому не повинні заперечувати проти вимірювання / збору біометричних параметрів.

Сукупність цих властивостей визначає ефективність використання біометрії в цілях вирішення завдань автентифікації та захисту інформації.

Однак не існує біометричних параметрів, котрі абсолютно задовольняють будь-якої з цих властивостей, як і параметрів, які б поєднували в собі всі ці властивості одночасно, особливо якщо взяти до уваги п'яту властивість – прийнятність. Це означає, що не існує універсального біометричного параметру, і використання будь-якого біометричного методу захисту визначається призначенням і необхідними характеристиками інформаційної системи.

У всіх системах біометричної автентифікації можна виділити дві підсистеми, які наведені на рис. 1.1.

1) Реєстрації об'єкта (за допомогою декількох вимірів зі зчитувального пристрою формується цифрова модель біометричної характеристики (біометричний шаблон)).

2) Розпізнавання об'єкта (вимірювання, признаки сформовані при спробі автентифікації, перетворюються в цифрову форму, яка потім порівнюється з формою, яка отримана при реєстрації).

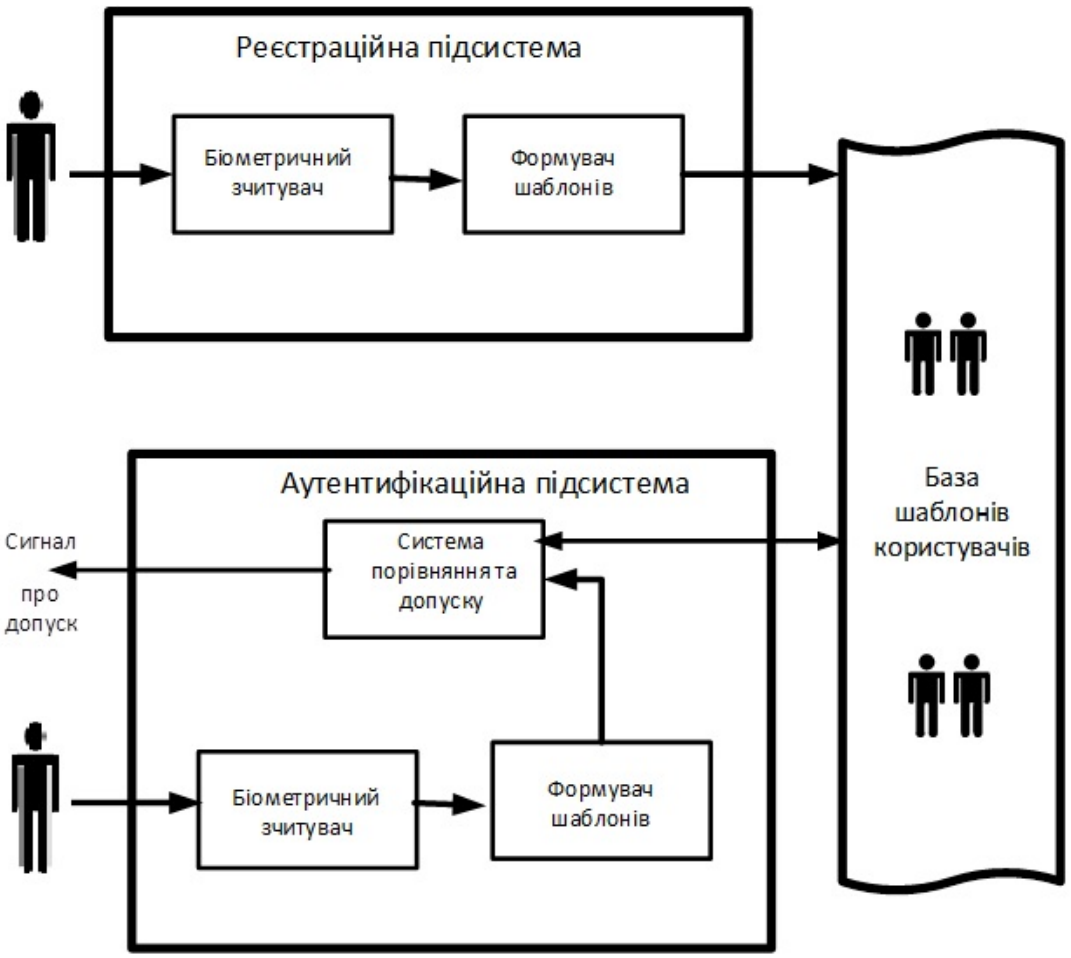

Рисунок 1.1 – Біометрична система автентифікації

Розрізняють два біометричних методу порівняння шаблонів.

1) Верифікація – порівняння з єдиним шаблоном, обраним на підставі певного унікального ідентифікатора, який виділяє конкретну людину (наприклад ідентифікаційний номер або код), або порівняння двох біометричних шаблонів один до одного (1: 1).

2) Ідентифікація – порівняння виміряних параметрів (біометричного шаблону людини) з усіма записами в бази зареєстрованих користувачів, а не з однією з них, обраної на підставі якогось ідентифікатора. На відміну від верифікації ідентифікація являє собою порівняння один до багатьох (1: m).

В основі будь-якої біометричної системи лежать зчитування (унікальну інформація виноситься з фізичного та / або поведінкового зразка і складається біометричний зразок-шаблон), зіставлення (представлений зразок порівнюється зі збереженим зразком з бази даних) і прийняття рішень (система визначає, чи збігаються біометричні зразки і виносить рішення про повторення, закінчення або зміну процесу автентифікації).

Біометрична автентифікація користувача стає важким завданням, коли потрібна висока точність, тобто низька ймовірність помилок. Крім того, користувач не повинен мати можливість згодом заперечувати проведену ним операцію і одночасно відчувати якомога менше незручностей при проходженні процедури автентифікації (можливість безконтактного зчитування, дружність інтерфейсу, розміри файлу-шаблона (чим більше розмір образу, тим повільніше йде розпізнавання) і т. д.). При цьому система автентифікації повинна відповідати також вимогам конфіденційності і бути стійкою до підробки (несанкціонованого доступу). Слід враховувати також стійкість біометричних систем автентифікації до навколишнього середовища (експлуатаційні якості можуть втрачати стабільність в залежності від оточуючих умов).

Таким чином, основні вимоги, що пред'являються до біометричних систем, наступні [5–6]:

- точність (чи завжди система приймає правильне рішення щодо об'єкту);
- швидкість обчислення і можливість масштабування баз даних;
- обробка виняткових випадків, коли біометричні параметри об'єкта не можуть бути зареєстровані (наприклад, в результаті хвороби або каліцтва);
- вартість (в тому числі витрат на навчання користувачів і персоналу);
- конфіденційність (забезпечення анонімності; дані, отримані під час біометричної реєстрації, не повинні використовуватися з метою, на які зареєстрований індивід не давав згоди);
- безпека (захист системи від загроз і атак).

1.2 Розпізнавання особи за відбитками пальців

Дактилоскопія – це встановлення особи людини за відбитками пальця, а точніше, по так званим папілярних узорів. Дактилоскопія ґрунтується на тому, що, по-перше, відбиток пальця унікальний (за всю історію дактилоскопії не було виявлено двох співпадаючих відбитків пальців, що належать різним особам), а по-друге, папілярний візерунок не змінюється протягом усього життя людини [5].

Шкірний покрив пальців рук має складний рельєфний малюнок (папілярний візерунок), утворений чергуючими валиками (висотою 0,1-0,4 мм і шириною 0,2-0,7 мм) і борозенками – заглибленнями (шириною 0,1-0,3 мм). Папілярний візерунок повністю формується на сьомому місяці розвитку плода. Більш того, в ре-

зультаті проведених досліджень було встановлено, що відбитки пальців різні навіть у однояйцевих близнюків, хоча показники ДНК у них ідентичні.

Крім того, папілярний узор неможливо видозмінити – ні порізи, ні опіки, ні інші механічні пошкодження шкіри не мають принципового значення, бо стійкість папілярного візерунка забезпечується регенеративної здатністю основного шару епідермісу шкіри. Тому можна стверджувати, що сьогодні дактилоскопія є найнадійніший спосіб ідентифікації особистості.

Незважаючи на різноманіття будови папілярних візерунків, вони піддаються чіткій класифікації, що забезпечує процес їх індивідуалізації та ідентифікації.

У кожному відбитку пальця можна визначити два типи ознак – глобальні та локальні. Глобальні ознаки – ті, які можна побачити неозброєним оком. Інший тип ознак – локальні. Їх називають мінуції – унікальні для кожного відбитку ознаки, що визначають пункти зміни структури папілярних ліній (закінчення, роздвоєння, розриви і т.д.), орієнтацію папілярних ліній і координати в цих пунктах.

Практика показує, що відбитки пальців різних людей можуть мати однакові глобальні ознаки, але абсолютно неможливо наявність однакових мікроузорів – мінуцій. Тому глобальні ознаки використовують для поділу бази даних на класи і на етапі автентифікації. На другому етапі розпізнавання використовують вже локальні ознаки.

Пристрою для читання відбитків пальців в даний час знаходять широке застосування. Їх встановлюють на ноутбуки, в миші, клавіатури, флешки, а також застосовують у вигляді окремих зовнішніх пристроїв і терміналів, що продаються в комплекті з системами AFIS (Automated fingerprint identification systems – система автоматизованої ідентифікації відбитків пальців). Всі сканери можна розділити на кілька видів: оптичні, напівпровідникові, ультразвукові.

Зараз в основному використовуються стандарти ANSI і ФБР США. У них визначені наступні вимоги до образу відбитка [6]:

- кожен зразок представляється у форматі нестислого TIF;
- зразок повинен мати дозвіл не нижче 500 dpi;
- зразок повинен бути напівтоновим з 256 рівнями яскравості;
- максимальний кут повороту відбитка від вертикалі не більше 15^{o};
- основні типи мінуцій – закінчення і роздвоєння.

Зазвичай в базі даних зберігають більше одного зразка, що дозволяє поліпшити якість розпізнавання. Зразки можуть відрізнятися один від одного зрушен-

ням і поворотом. Масштаб не змінюється, так як всі відбитки отримують з одного пристрою.

1.3 Розпізнавання особистості по райдужній оболонці

Райдужна оболонка за формою схожа на коло з отвором всередині (зіницею). Райдужка складається з м'язів, при скороченні і розслабленні яких розміри зіниці змінюються. Вона входить в судинну оболонку ока. Райдужка відповідає за колір очей (якщо він блакитний – значить, в ній мало пігментних клітин, якщо коричневий – багато). Виконує ту ж функцію, що діафрагма у фотоапараті, регулюючи світовий потік. Райдужка входить до складу ока. Вона знаходиться за рогівкою і водянистою вологою передньої камери. Унікальні структури райдужної оболонки обумовлені радіальної трабекулярної мережею (trabecular meshwork); її склад: поглиблення (крипти, лакуни), борінні стяжки, борозни, кільця, зморшки, веснянки, корони, іноді цятки, судини і інші риси [7].

Рисунок райдужки в великій мірі випадковий, а чим більше ступінь випадковості, тим більша ймовірність того, що конкретний рисунок буде унікальним. Математично випадковість описується ступенем свободи. Дослідження показали, що текстура райдужки має ступінь свободи рівній 250, що значно більшій ступені свободи відбитків пальців (35) і зображень осіб (20). Середні розміри райдужної оболонки: по горизонталі – R ≈6,25 мм, по вертикалі – R ≈ 5,9 мм; розмір зіниці становить 0,2 ... 0,7R.

Внутрішній радіус райдужки залежить від віку, стану здоров'я, освітлення та ін. Він швидко змінюється. Його форма може досить сильно відрізнятися від кола.

Центр зіниці, як правило, зміщений відносно центру райдужки у напрямку до кінчика носа.

По-перше, оболонка має дуже складний рисунок, в ній багато різних елементів. Тому навіть не дуже якісний її знімок дозволяє точно визначити особистість людини

По-друге, райдужна оболонка є об'єктом досить простої форми (майже плоске коло). Так що під час ідентифікації дуже просто врахувати всі можливі спотворення зображення, що виникають із-за різних умов зйомки.

По-третє, райдужна оболонка ока людини не змінюється протягом усього його життя з самого народження. Точніше, незмінною залишається її форма (ви-

няток становлять травми і деякі серйозні захворювання очей), колір же згодом може змінитися. Це надає ідентифікації по райдужній оболонці ока додатковий плюс в порівнянні з багатьма біометричними технологіями, які використовують відносно недовговічні параметри, наприклад геометрію особи або руки.

Райдужна оболонка починає формуватися на 3-й місяць внутрішньоутробного розвитку. На 8-й місяць вона є практично сформованою структурою. Крім того, вона формується випадково навіть у однояйцевих близнюків і гени людини не впливають на її структуру. Райдужна оболонка стійка після 1-го року життя – райдужка остаточно сформована і практично не змінюється аж до самої смерті, якщо немає травм або патологій ока.

Властивості райдужної оболонки як ідентифікатора:

- ізольованість і захищеність від зовнішнього середовища;

- неможливість зміни без порушення зору;

- реакція на світло і пульсація зіниці використовується для захисту від підробок;

- можливий ненав'язливий, безконтактний і потайний спосіб отримання зображень;

- висока щільність унікальних структур –3,2 біта / мм^2 або близько 250 незалежних характеристик (у інших методів близько 50), 30% параметрів досить, щоб прийняти рішення про збіг з ймовірністю не більше 10^{-6}.

У ідентифікації особистості по райдужній оболонці ока є ще одна серйозна перевага. Справа в тому, що деякі біометричні технології страждають таким недоліком. При установці в настройках системи ідентифікації високого ступеня захисту від помилок першого роду (ймовірність помилкового допуску – FAR) ймовірність появи помилок другого роду (помилковий не допуск в систему – FRR) зростає до недозволено високих величин – декількох десятків відсотків, в той час як ідентифікація за райдужною оболонкою ока повністю позбавлена цього недоліку. У ній співвідношення помилок першого та другого роду є одним з кращих на сьогоднішній день. Для прикладу можна навести кілька цифр. Дослідження показали, що при ймовірності виникнення помилки першого роду в 0,001% (відмінний рівень надійності) ймовірність появи помилок другого роду становить всього лише 1%.

Теоретична ймовірність того, що дві різні людини мають один і той же рисунок райдужної оболонки, приблизно дорівнює 10^{-78}, в той час як все населення Землі становить менше 10^{10}.

Недоліком цієї технології є відносно висока вартість обладнання. І дійсно, для проведення дослідження потрібна як мінімум камера, яка буде отримувати початкове зображення. А коштує це пристрій набагато дорожче, ніж, наприклад, сенсор відбитків пальців. Крім того, вона вимагає досить багато місця для розміщення. Все це обмежує область використання ідентифікації особистості по райдужній оболонці ока. На сьогоднішній день вона застосовується в основному в системах допуску на різні об'єкти як цивільного, так і військового призначення.

1.4 Розпізнавання обличчя

Для ідентифікації особистості найкраще підходять технології розпізнавання по обличчю. Вони ненав'язливі (розпізнавання людини відбувається на відстані, без затримок і відволікання уваги), як правило, пасивні (не вимагають будь-яких дій з боку людини), не обмежують користувача в свободі переміщень і відносно недорогі. Крім того, люди зазвичай легко впізнають один одного за обличчям, а значить, і автоматизовані системи не повинні відчувати труднощів (на практиці все інакше) [8].

За обличчям людини можна дізнатися його історію, симпатії і антипатії, хвороби, емоційний стан, почуття і наміри по відношенню до оточуючих. Все це представляє особливий інтерес для автоматичного розпізнавання осіб (наприклад, для виявлення потенційних злочинців).

До інформаційних ознаками обличчя особи відносяться:

- форма обличчя (кругла, квадратна, трикутна і т. д.);
- співвідношення частин особи між собою (лоб, середня і нижня частини обличчя особи);
- форма чола, скул і підборіддя;
- форма і розмір вуха, спосіб його прикріплення, форма частин вуха;
- симетрія / асиметрія особи;
- форма, величина, (кількість) і розташування очей, рота, носа;
- лінії зморшок і ін.

Залежно від того, який портрет використовується (в фас, профіль або обидва), ці методи комбінуються. У комерційних системах зазвичай використовують образи осіб в фас (з поворотом в бік до 15^{o}), а значить, не використовують інформаційні знаки вух і профіль особи (однак деяка інформація про профіль може бути отримана і з зображення в фас – за градієнтами яскравості).

Якість бази даних визначається на основі таких ознак:

- репрезентативність;
- спосіб структурування даних;
- якість образу:
- розмір кожного образу в пікселях;
- контраст і промальовування деталей особи;
- фон, на якому перебуває особа;
- відсутність перешкод на обличчі.

Бажано, щоб в базі даних були образи з різним поворотом голови, присутністю або відсутністю додаткових предметів (окуляри, сережки і т. д.) І з різними виразами.

Для оцінки системи розпізнавання зазвичай використовується спеціальна база даних (БД) ORL Database of Faces. Вона відповідає всім цим ознакам і доступна багатьом розробникам. Ця БД містить 400 образів по 10 в кожному класі (тобто всього 40 різних зображень осіб). Кожен образ має дозвіл 112x92 пікселя і 256 рівнів яскравості. Всі особи представлені на темному тлі. Репрезентативність даних забезпечується деякими змінами масштабу особи, кута спостереження і умов освітлення.

Найчастіше в літературі згадуються три методу розпізнавання особи.

1) Кореляційний (метод узгодженої фільтрації).

2) Метод на основі перетворень Карунена - Лоева і поняття «власних осіб» (EigenFace).

3) Метод на основі лінійного дискримінантного аналізу і поняття Fisherface (по імені Роберта Фішера).

Розвиваються зараз методи, орієнтовані: на репрезентативний характер вихідних даних – навчання системи в різних умовах; на зменшення розмірності вихідних даних; на розпізнавання в скороченому просторі ознак.

Переваги методу:

- низька ціна пристроїв отримання відео зображення;
- ненастирливої системи;
- безконтактність;
- непомітність.

Недоліки методу:

- складність реалізації системи;
- висока ціна пристроїв отримання термографічного образу;

- залежність відео зображення від перешкод.

1.5 Розпізнавання людини за голосом

Голос – це поведінковий біометричний параметр, що залежить від фізичних характеристик людини. Властивості голосу (такі, як частота, носової звук, модуляція, інтонація і т. д.) є унікальними особливостями людини [6].

Ідентифікація людини по голосу – один з традиційних способів розпізнавання, який застосовується повсюдно. Можна легко дізнатися співрозмовника по телефону, не бачачи його. Також можна визначити психологічний стан по емоційному забарвленню голосу.

Переваги технології:

- можливість розпізнавання на відстані (без затримок і відволікання уваги);

- пасивність; технології, як правило, пасивні (не вимагають будь-яких дій з боку людини),

- відсутні обмеження користувача в свободі переміщень.

Ідентифікація за голосом заснована на аналізі унікальних характеристик мови, обумовлених анатомічними особливостями (розмір і форма горла і рота, будова голосових зв'язок) і набутими звичками (гучність, манера, швидкість мовлення).

Голос схильний до суттєвих змін під впливом емоційних чинників (настрій людини) і стану здоров'я (ангіна, нежить, бронхіт і т. д.). На якості ідентифікації можуть позначатися зовнішні умови (наприклад, сторонні шуми від дорожнього руху, розмов інших людей). Якщо для передачі голосової інформації використовуються лінії зв'язку, перешкоди в них також здатні ускладнити розпізнавання користувача.

Голосові автентифікаційні системи поділяються на категорії залежно від вимог до розпізнавання мови:

- заданий текст (певні слова або фрази записуються при реєстрації. Слова можуть бути секретними, тоді вони діють як пароль);

- незалежність від тексту (системи обробляють будь-яку фразу особи, яку вона говорить. Спостереження може бути тривалим, і чим більше говорить людина, тим точніше система ідентифікує користувача. Такі системи можуть автентифікувати користувача, навіть в тому випадку, якщо він говорить іншою мовою);

- діалог (при цьому потрібно вимовити секретні слова або принаймні надати інформацію, яку не можна вгадати і дізнатися).

Одна з причин привабливості технологій розпізнавання мовця – це поширеність і низька вартість сенсора, необхідного для реєстрації мовного сигналу. Мікрофони зараз присутні практично в кожному пристрої: стаціонарних і стільникових телефонах, ноутбуках і настільних комп'ютерах. Всі вони можуть бути використані в якості сенсорів.

Розглянемо структуру мовного сигналу. Кожен сплеск голосового сигналу відповідає деякому фрагменту мови. Це може бути одна буква, поєднання букв (фонема) або коротке широко поширене слово. Всього в українській (російській) мові є 42 фонеми, але підходять для ідентифікації людини не все. Частина фонем огласована (вокалізованих). Саме їм притаманний індивідуальний характер. Це звуки «е», «о», «л», «а», «і» і ін. Інша частина фонем – шиплячі (шумоподібні). Це «ц», «ч», «ш», «щ» і т. д. Вони не є індивідуальними і їх використання при ідентифікації може привести до зниження якості розпізнавання.

На сьогоднішній день існує два підходи до ідентифікації людини по голосу, побудовані з урахуванням структури мовного сигналу.

Індивідуальні відмінності розподілу потужності сигналу по спектру покладені в основу першої категорії систем біометричної ідентифікації по голосу. Вони будуються на базі набору вузькосмугових фільтрів, які виділяють з голосу коливання різних частот.

Смуги пропускання фільтрів вибираються при проектуванні системи, але вони не повинні бути занадто вузькими, щоб не залежати від варіацій частотного спектра голосу. У той же час вони не повинні бути і дуже широкими. Потрібно підбирати оптимальну ширину, достатню для впевненої ідентифікації. Зазвичай використовують 16 фільтрів, смуга пропускання яких розширюється в міру зростання значень частот, які виділяються. Це пов'язано з нестабільністю високих частот по енергії (в порівнянні з низькими частотами). Підсумковий масив даних виходить дуже маленького розміру (потрібно записати тільки 16 координат вершин по одній осі).

Друга категорія систем ґрунтується на формуванні сигналу, що імітує голосову фразу з використанням апарату лінійного передбачення.

Вокалізовани (огласовании) коливання звуку імітуються періодичними впливами на цифровий фільтр. Період впливів повинен точно відповідати періоду (частоті) основного тону голосу. Динамічні характеристики цифрового фільтру

повинні змінюватися, щоб отримати форму, близьку до голосової фразі. Число коефіцієнтів фільтра коливається від 10 до 12 (a1, ..., a12). Цього достатньо для якісного відтворення мови зі збереженням індивідуальних особливостей. Коефіцієнти лінійного передбачення обчислюються на вибірці у 180-220 відліків. Обчислення параметрів передбачення (цифрового фільтру) знаходять рішенням системи з 10-12 лінійних рівнянь [8].

При імітації огласованних звуків на вхід цифрового фільтру подають періодичну послідовність імпульсів, промодулеваною за амплітудою. В такому випадку на виході фільтра з'являються періодичні перехідні процеси, що повторюють модельований звук. При моделюванні шиплячих на вхід фільтра подають випадковий шум потрібної амплітуди.

При навчанні системи на її вхід подають кілька зразків голосу користувача. Вони перетворюються в послідовність імпульсів основного тону і відповідною послідовність коефіцієнтів лінійного передбачення. Виходить масив даних, що описує індивідуальні особливості голосу людини для цієї фрази. Цей масив з коефіцієнтів і є тим біометричним еталоном (шаблоном), який записується в базу даних.

Більшість розроблених на сьогоднішній день систем ідентифікації особистості за голосом побудовані на основі одноразової перевірки відповідності необхідної ключової фрази і яку він виголосив у початковий момент доступу до обчислювальної системи.

Дані системи підтримують два основні режими роботи:

- навчання системи;
- перевірка справжності при доступі.

У першому режимі (реєстрація) користувачеві пропонується кілька разів вимовити ключову фразу (пароль), обмежену, як правило, за тривалістю (3-4 с). При цьому навчання системи ідентифікації проводиться на усереднених мовних відрізках за результатами записів кількох вимовив. Записаний ключ може зберігатися в повному обсязі або стискатися ефективними алгоритмом, які дозволяють зберігати індивідуальні параметри голосу без спотворень. Деякі системи видаляють з записаної ключової фрази слабко виражені мовні ділянки (паузи, шуми, сплески енергії) шляхом її поділу на відрізки, відповідно фонем базової мови, з яких потім виділяється сукупність необхідних параметрів.

Існуючі системи розпізнавання людини за голосом мають наступні характеристики: помилки першого роду (не допуск свого) складають 1-5% (хоча в залеж-

ності від реалізації програмного забезпечення можуть доходити до 40%). Кількість помилок другого роду (пропуск чужого) залежить від того, чи знає зловмисник ключову фразу (до 1%, якщо голоси близькі) чи ні (0,00000001%). Зараз можна використовувати голосову ідентифікацію спільно з іншими видами захисту. Наприклад, по геометрії особи. Тоді можна відстежувати рух губ і синхронізацію їх зі звуком.

Голосовий захист легко пройти, якщо перехоплена або записана ключова фраза. Тому розробники зараз намагаються створити систему, захищену від перехоплення, тобто впізнати людину за будь якою фразою.

Переваги розпізнавання за голосом:

- звичний для людини спосіб ідентифікації;
- низька вартість (найнижча серед всіх біометричних методів);
- безконтактність.

Недоліки:

- високий рівень помилок 1 і 2 роду;
- необхідність в спеціальному шумоізольованому приміщенні для проходження ідентифікації;
- можливість перехоплення фрази «магнітофоном»;
- якість розпізнавання залежить від багатьох факторів (інтонація, швидкість проголошення, психологічний стан, хвороби горла);
- необхідність підбору спеціальних фраз (з огласованнимі фонемами).

1.6 Верифікація підпису

Верифікація підпису – метод, який має довгу історію розвитку. Підпис використовувалася ще до появи комп'ютерів і широко застосовувалася при автентифікації документів і при проведенні транзакцій з використанням чеків і кредитних карт. Розпізнавання підпису – це приклад розпізнавання особи, який писав, яке приймалося як незаперечний доказ в суді. Підписи можуть мати різну форму, даючи можливість особи, яка підписувала, визначати «відмітні ознаки» і «унікальність» свого підпису, які будуть впливати на коефіцієнт помилкового доступу (КПД) і коефіцієнт помилкового відхилення доступу (КПВД) [5].

Для біометричних параметрів необхідними умовами є:

- універсальність;
- унікальність;

- сталість, тобто незмінність у часі;

- збирання.

Питання про сталість підписи досить спірне, тому що людина може змінити свій підпис в будь-який час. До певної міри руху м'язів руки визначаються генетикою і впливом середовища і перетворюються в візуальні і зчитувальні машиною знаки. Біометричний параметр (як особа і голос) підпадає під вплив хвороби, емоцій або віку, тому дані фактори знаходяться в процесі вивчення. Також не дуже ясно, чи пов'язані параметри, підраховувані в процесі верифікації підпису, з індивідуальними фізичними характеристиками особи, яка пише (які не можна підробити).

Як виявилося, підпис – такий же унікальний атрибут людини, як і його фізіологічні характеристики. Крім того, це і більш звичний для будь-якої людини метод ідентифікації, оскільки він на відміну від зняття відбитків пальців не асоціюється з кримінальною сферою. Одна з перспективних технологій автентифікації заснована на унікальності біометричних характеристик руху людської руки під час письма.

Технології автоматизованої верифікації підпису можна розділити по способам отримання зразків.

1) Офлайнові, або «статистичні» підписи скануються з документів і паперів. Офлайновий аналіз підпису може бути проведений з відсканованого зображення за допомогою камери або сканера.

2) Онлайнові, або «динамічні» підписи виходять за допомогою спеціальних пристроїв; динамічні характеристики (положення кінчика ручки в процесі письма) можна зчитувати з високою роздільною здатністю, навіть коли ручка не стосується паперу.

Перший спосіб досить ненадійний, тому що заснований на звичайному порівнянні введеної підписи з тими що зберігаються в базі даних графічними зразками. Через те, що підпис не може бути завжди однаковою, цей метод дає великий процент помилок. Спосіб динамічної верифікації вимагає набагато складніших обчислень і дозволяє в реальному часі фіксувати параметри процесу підписи, такі як швидкість руху руки на різних ділянках, сила тиску і тривалість різних етапах підпису. Це дає гарантії того, що підпис не зможе підробити навіть досвідчений графолог, оскільки ніхто не в змозі в точності скопіювати поведінку руки власника підпису. Тільки справжній користувач зможе повторити всі ці характеристики за той же час. Копіювальна машина або фахівець можуть з легкістю зробити дуб-

лікат вашого підпису і відтворити її, але дублювати час і все характеристики підписи практично неможливо.

Відпрацьована згодом манера особистого підпису людини є необхідною характеристикою для відтворення всіх необхідних параметрів, які потім і розглядає система. Кожен раз, підписуючи документи, в підписі можуть бути якісь невеликі варіації, але характеристики, які визначаються природними рухами і особливостями, виробленими протягом довгого часу, створюють впізнавані ознаки, які і роблять підпис об'єктом біометричної ідентифікації.

Користувач, використовуючи стандартний дігітайзер і ручку, імітує свою звичайну підпис, а система зчитує параметри руху і звіряє їх з тими, що були заздалегідь введені в базу даних.

Переваги методу:

- невисока вартість;

- відносна звичність для людини. Підпис вже давно є визнаним методом, здатним підтвердити особистість людини. Роботу даної системи легко пояснити людині, і люди їй вже довіряють, тому що даний спосіб ідентифікації є природним і ненав'язливим.

Недоліки:

- високий рівень помилок 1 і 2 роду;

- необхідність привчання до роботи з планшетом перед реєстрацією;

- тривалий час реєстрації користувача;

- користувачі можуть зображати нестабільний почерк, якщо опираються системі.

1.7 Порівняльна характеристика біометричних методів

Нижче (см. табл. 1.2) представлені результати порівняльної характеристики біометричних методів, наведеної на конференції VoiceBioCon 2007 (автори: Dan Miller, Senior Analist, Opus Reserch) [6].

Звернемо увагу на трудомісткість реалізації і точність біометричних методів, які схематично представлений на рис. 1.2 [5].

Таблиця 1.2 – Порівняльна характеристика біометричних систем

Порівняльна характеристика біометричних систем	Відбиток пальця	Голос	Райдужна оболонка	Обличчя
Надійність верифікації, %	96.7-98	99.14-99.9	95.4-95.9	95.9
Помилка реєстрації, %	4	2	7	0.1
Ймовірність «допуску чужого», %	2.5	0.75	6	4
Ймовірність «відмови своєму», %	0.1	0.75	0.001	10
Вартість системи	Висока	Низька	Дуже висока	Висока

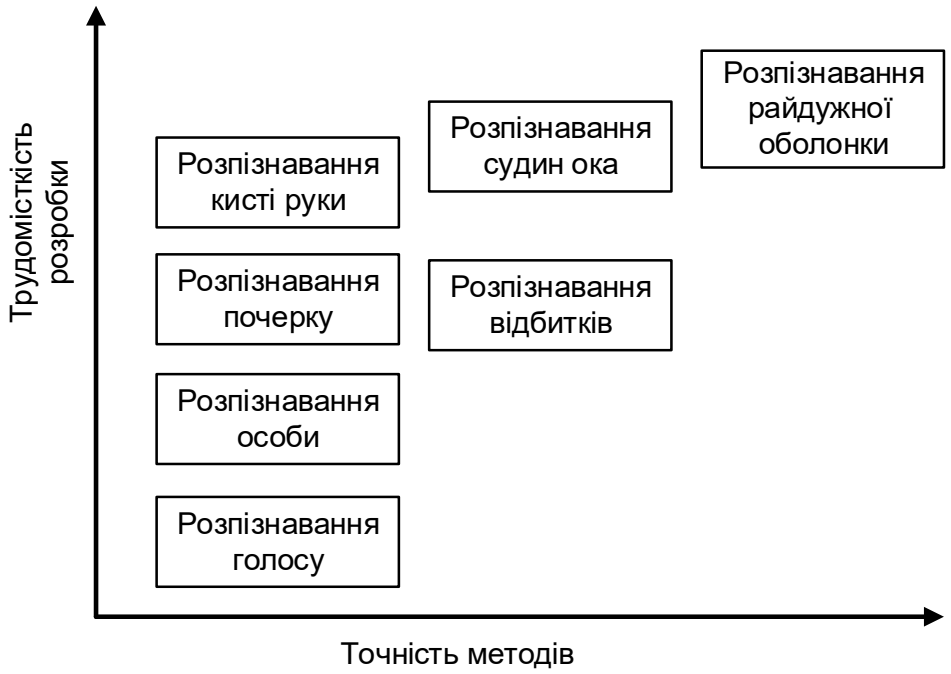

Рисунок 1.2 – Трудомісткість реалізації і точність біометричних методів

З аналізу рис. 1.2 випливає, що меншою трудомісткістю володіє голосова автентифікація. Наявність і простота обладнання для реалізації голосової автентифікації істотно спрощує її розробку. Тобто необхідно створити і встановити ві-

дповідний програмний модуль. Слід зауважити, що саме він сприяє кращої точності методу, яка в даному випадку поступається дактилоскопії і розпізнаванню за райдужною оболонкою ока. Але тут є ряд причин, про які піде мова нижче.

Таким чином, за критерієм ефективність / вартість найбільш перспективними є системи голосової автентифікації, яким і присвятимо основну увагу нижче.

1.8 Постановка завдань досліджень

У зв'язку з викладеним вище, можна запропонувати наступну схему досліджень:

- аналіз документальних джерел в області голосової автентифікації;
- патентний пошук в області обробки мовних сигналів;
- розробка методики формування фазової інформації голосового сигналу [9];
- дослідження інформативності фазових даних на основі оцінки частоти основного тону;
- розробка математичної моделі для дослідження процедур оцінки частоти основного тону;
- математичне моделювання та аналіз результатів досліджень.

2 АНАЛІЗ ДОКУМЕНТАЛЬНИХ ДЖЕРЕЛ І РОЗРОБКА МЕТОДИЧНИХ ОСНОВ ДЛЯ ПРОВЕДЕННЯ ДОСЛІДЖЕНЬ

2.1 Аналіз документальних джерел в галузі досліджень

Спочатку розвиток і впровадження біометричних систем пов'язували зі статичними біометричними ознаками користувача (зображення особи, папілярний узор пальця і райдужна оболонка ока), які добре зарекомендували себе в криміналістиці. Однак, до теперішнього часу ці надії зруйновані, в першу чергу, через простоту підробки.

Тому останнім часом багато досліджень проводиться в області застосування динамічних (поведінкових) біометричних систем автентифікації. Серед цих біометричних систем особливе місце займає голосова автентифікація, яка відрізняється простотою і зручністю. Але, як і всі біометричні системи, голосова автентифікація має низькі якісні характеристики. У зв'язку з цим в області голосової автентифікації проводяться інтенсивні дослідження, про що свідчать роботи [9– 12].

У сучасних системах голосової автентифікації (СГА) реєструється амплітудна інформація полігармонічного нестаціонарного голосового сигналу користувача. Автентифікація користувача здійснюється в основному в процесі аналізу амплітудно-частотного спектра матеріалів реєстрації [6]. Основні зусилля дослідників при цьому зосереджені на пошуку нових або вдосконалення існуючих процедур формування (оцінки) шаблонів (набору ознак – частоти основного тону, формантних даних, коефіцієнтів кепстра, мел-частотних кепстральних коефіцієнтів, коефіцієнтів лінійного передбачення і їх динамічні характеристики і ін.) користувача, а також на розробці вирішальних правил. Найбільш популярні серед останніх такі процедури прийняття рішень – методи гауссових сумішей (Gaussian Mixture Model, GMM) і опорних векторів (Support Vector Machine, SVM). Для цих цілей також використовуються штучні нейронні мережі і приховані Марковские моделі (Hidden Markov Models, HMM). Підвищення показників якості СГА може бути пов'язано з іншою парадигмою обробки матеріалів голосової реєстрації, яка пов'язана з доповненням процедур аналізу амплітудно-частотного спектру сучасними досягненнями цифрової обробки інформації, в тому числі, і з використанням моделі аналітичного сигналу, а також алгоритмами обліку та урахування фазових даних голосових сигналів.

В даний час існує інший шлях підвищення якісних показників СГА, який базується, в першу чергу, на використанні фазової інформації голосового сигналу користувача. Давно відомо [9], що фаза є більш інформативним параметром сигналу, проте в СГА вона традиційно ігнорується [6].

Обумовлено це тим, що для отримання фазової інформації необхідні додаткові обчислювальні і алгоритмічні ресурси, які не завжди були доступні в зазначених додатках. Зауважимо, що раніше в радіолокації і радіозв'язку для отримання фазових даних використовувалися спеціальні громіздкі пристрої – фазовращателі, які неможливо було застосовувати в області обробки голосових сигналів. В даний час існують спеціалізовані мікросхеми та цифрові сигнальні процесори, які можна застосувати й в області цифрової обробки голосових сигналів.

Крім цього, є деякі особливості оцінки, попередньої обробки та використання фазових даних голосових сигналів. Слід зазначити, що в даний час відсутній досвід і практика використання фази сигналу стосовно завдань голосовій автентифікації.

Підтвердженням цьому є те, що відомо лише обмежена кількість робіт, де фазові дані використовувалися при обробці мовних сигналів. Наприклад, в [10] зазначено на актуальність використання фазової інформації при обробці мовних даних, а в [11] використовувалася фаза для уточнення частотних характеристик оброблюваних голосових даних. В [12] виконано порівняльний аналіз процедур оцінки фазових співвідношень між коливаннями основного тону і обертонів мовних сигналів, які автори пропонують використовувати для вирішення завдань розпізнавання звуків мови та ідентифікації дикторів.

Зазначене вище підкреслює актуальність досліджень оцінки впливу фазових даних на якісні характеристики процедур голосовій автентифікації. Фазові дані в голосовій автентифікації можуть використовуватися за кількома напрямками:

- підвищення відносини сигнал / шум матеріалів реєстрації (відомий напрямок використання фази в радіолокації і радіозв'язку);

- підвищення якості формування ознак для традиційно використовуваних шаблонів, наприклад, частоти основного тону, формантної інформації і т.д.;

- розробка нових процедур формування елементів шаблонів на основі фазових даних [13].

Очевидно, технології ідентифікації особистості по голосу прийшли в системи автентифікації користувачів з криміналістики. Наукові основи застосування

технології ідентифікації голосу в криміналістиці досліджувалися і детально обговорювалися в [14].

Загальний висновок полягає в тому, що ідентифікація по голосу відрізняється від відбитків пальців, де варіації дуже малі, і немає абсолютно надійного методу для визначення того, чи належать мовні сигнали одної і той ж людині. У криміналістиці розпізнавання диктора може мати тільки імовірнісний характер, тобто із зазначенням правдоподібності того, що два мовних сигналу належать одній і тій же людині. В умовах аналогового телефонного каналу іноді ускладняється навіть розпізнавання статі або віку. В силу малої вибірки мовних сигналів довірчий інтервал оцінки правдоподібності приналежності двох записів мови одного й того ж диктора настільки великий, що однозначне рішення неможливо.

Досить близькою є завдання сегментації дикторів. Сегментація дикторів в потоці розмови різних дикторів (audio-indexing, diarization) необхідна при розмітці звукових стенограм, телеконференцій, радіо- і телепередач, інтерв'ю і т. д. Однак, як і в криміналістиці, якість виділення диктора є низьким і неприйнятним для вирішення завдань автентифікації користувача [15].

Індивідуальність акустичних характеристик голосу визначається трьома факторами: механікою коливань голосових складок, анатомією мовного тракту і системою управління артикуляцією. Акустично стиль реалізується у вигляді контуру частоти основного тону, тривалості слів і його сегментів, ритміки ударних сегментів, тривалості пауз, гучності [16].

Простір ознак, в якому приймається рішення про особистості диктора, має формуватися з урахуванням всіх факторів механізму мовотворення : голосового джерела, резонансних частот мовного тракту і їх загасання, а також динамікою управління артикуляцією. Зокрема в [16, 17], розглядаються наступні параметри голосового джерела: середня частота основного тону (ЧОТ), контур частоти основного тону, флуктуації частоти основного тону і форма імпульсу збудження. Спектральні характеристики мовного тракту описуються обвідною спектру і його середнім нахилом, формантними частотами і їх смугами, довготривалим спектром або кепстром [17].

В [18] показано, що найбільш важливий фактор індивідуальності голосу – це частота основного тону (F0), за нею йдуть формантні частоти, розмір флуктуацій F0 і нахил спектру. В [19] висловлюється думка, що ознаки, пов'язані з F0, забезпечують найкращу роздільність голосів, а за ними слідують енергія сигналу і тривалість сегментів.

У деяких роботах найбільш важливим фактором вважаються формантні частоти [20, 21]. Зокрема, четверта форманта практично не залежить від типу фонеми і характеризує тракт [21].

У роботах по розпізнаванню диктора домінує метод кепстрального перетворення спектра голосових сигналів, який вперше був запропонований в [22].

Кепстр описує форму обвідної спектру сигналу, в якій інтегруються характеристики джерел збудження (голосового, турбулентного і імпульсного) та форми мовного тракту. В експериментах щодо суб'єктивного розпізнавання диктора було встановлено, що огинаюча спектру сильно впливає на пізнаваність голосу [23]. Тому використання того чи іншого способу аналізу обвідної спектру з метою розпізнавання диктора виправдано.

Замість обчислення спектру мовного сигналу з використанням дискретного перетворення Фур'є на короткому інтервалі часу, може використовуватися також амплітудно-частотна характеристика сигналу, знайдена за коефіцієнтами лінійного передбачення мови [24].

В роботі [25] було знайдено три інформативні області: 100-300 Гц (вплив голосового джерела), 4-5 кГц (грушоподібні порожнини) і 6.5-7.8 кГц – (можливо, вплив приголосних). Невелика область – в районі 1 кГц.

В силу того, що в переважній більшості систем розпізнавання диктора використовується один і той же простір ознак, наприклад, у вигляді кепстральних коефіцієнтів, їх перших і других різниць, велика увага приділяється побудові вирішальних правил, про котре йшлося вище.

Розробка і застосування методу GMM розглянуто в роботах [26, 27]. Метод GMM може розглядатися, як розширення методу векторного квантування [27]. Векторне квантування є найпростішою моделлю в системах розпізнавання диктора незалежно від контексту.

Метод опорних векторів (SVM) активно використовується в різних системах розпізнавання образів після публікації монографії [28]. Цей метод дозволяє побудувати гіперплощину в багатовимірному просторі, що розділяє два класи, наприклад, параметрів цільового диктора і параметрів дикторів з референтної бази. Гіперплощина обчислюється з використанням не всіх векторів параметрів, а тільки спеціально обраних. Ці вектори і називаються опорними. Оскільки розділяє поверхню в вихідному просторі параметрів не обов'язково відповідає гіперплощина, то виконується нелінійне перетворення простору виміряних параметрів в деякий простір ознак більш високої розмірності. Це нелінійне перетворення має

задовольняти вимогі лінійної роздільності в новому просторі ознак. Якщо ця умова виконується, то розділяюча поверхня в гіперплощині будується методом опорних векторів. Очевидно, що успіх застосування методу опорних векторів залежить від того, наскільки вдало підібрано нелінійне перетворення в кожному конкретному випадку при розпізнаванні дикторів.

Метод опорних векторів застосовується для верифікації дикторів часто в комбінації з методом GMM або HMM.

До розпізнаванню дикторів застосовується і метод прихованих Марківських моделей (HMM), добре зарекомендував себе в задачах автоматичного розпізнавання мови [29, 30]. Зокрема передбачається, що для коротких фраз тривалістю в кілька секунд для контекстна залежного підходу найкраще застосовувати фонемна залежні HMM, а не моделі на основі ймовірностей переходу від кадру до кадру тривалістю 10 - 20 мс. Метод прихованих Марківських моделей може використовуватися в сукупності з методом GMM.

Загальний висновок з аналізу відомої літератури – шаблони при автентифікації (розпізнаванні диктора) формуються на основі цифрової обробки амплітудно-частотного спектру голосового сигналу користувача. У той же час ігнорується більш інформативний параметр голосових даних користувача, а саме, фазочастотний спектр. Це може бути перспективним напрямком підвищення надійності голосовій автентифікації.

В [31] запропоновано змінити існуючу парадигму обробки голосових сигналів, яку доцільно доповнити аналізом фазових даних, які до теперішнього часу не використовуються.

2.2 Результати патентних досліджень за темою роботи

Патенти, які відносяться до даної теми зазначені у класифікації до розділу фізики (G). Одночасно вказані патенти включені до підрозділу G10 – музичні інструменти; акустика і клас G10L – аналіз або синтез мови; розпізнавання мови.

Деякі результати патентного пошуку представлені в додатку А (див. таблицю А1). Розглянемо їх і виконаємо деякий аналіз.

На жаль, дослідження проводилося за відкритими базами даних.

Аналіз результатів патентного пошуку свідчить, що в процесі розв'язання задачі ідентифікації особистості за голосовим сигналом в основному використовується [32 – 34]:

- частота основного тону;
- аналіз амплітудного спектру;
- дослідження формантної інформації;
- оцінка кепстральних коефіцієнтів і коефіцієнтів лінійного передбачення;
- аналіз фонемної інформації.

Спільним є те, що в основу всіх методів покладені процедури аналізу амплітудного спектру. В окремих випадках здійснюють нормалізацію спектру та враховують частотну характеристику мікрофонів.

Тільки в одному патенті (US 4624011) проводився аналіз фазового спектру для вирішення задачі розпізнавання мови.

Таким чином, фазові дані голосового сигналу не використовуються у сучасних СГА та можуть бути перспективним напрямком їх застосування для підвищення якісних характеристик голосової автентифікації.

2.3 Модель аналітичного сигналу

Для того щоб сигнали були об'єктами теоретичного вивчення і аналізу, необхідно мати їх математичні моделі. Математична модель сигналу – це формалізоване його представлення у вигляді певного математичного об'єкту. Фізичною величиною, що визначає характер радіосигналу, зазвичай є напруга або струм, що змінюються в часі за певним законом. Тому найбільш часто в якості моделі сигналу використовується функціональна залежність, аргументом якої є час, тобто функція часу.

Відомо, що будь-яке комплексне число можна представити у вигляді точки на комплексній площині або вектору, який виходить з 0 до цієї точки, а комплексний сигнал можна трактувати як комплексну функцію часу, тобто вектор який описує на комплексній площині деяку траєкторію в плині часу, як це показано на рисунку 2.1.

В області радіотехніки, радіолокації і радіозв'язку широке застосування знаходить модель аналітичного сигналу, яка була введена Д. Габором в 1946 році [32]. Ця модель дала можливість визначити поняття миттєвих амплітуди, фази і частоти сигналу. Введені параметри і зазначена модель дозволили істотно спростити вивчення і підвищити ефективність процедур цифрової обробки.

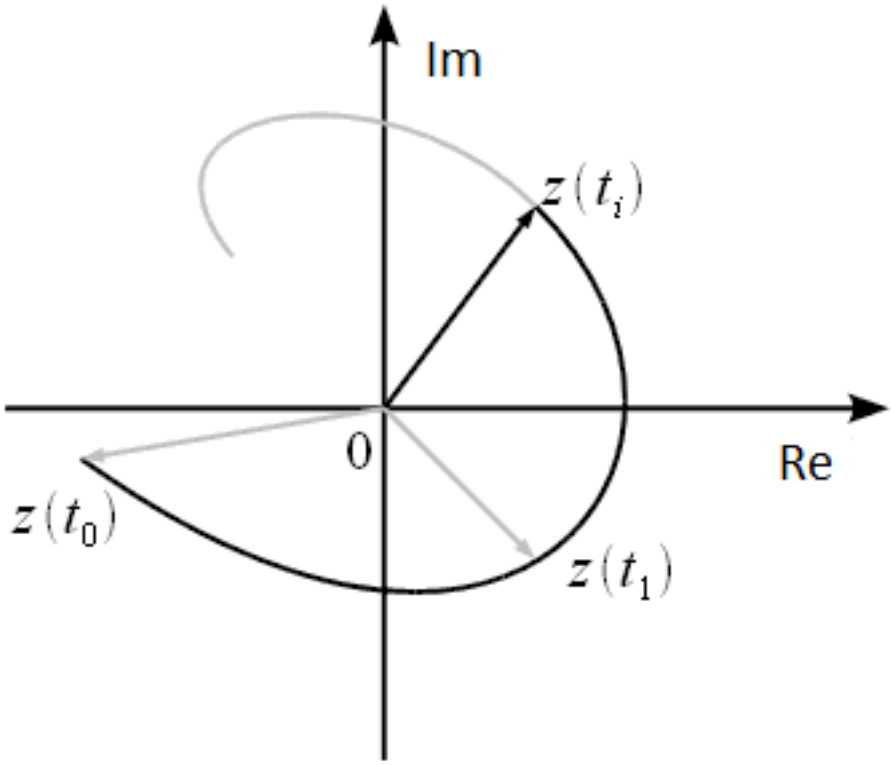

Рисунок 2.1 – Векторне подання комплексного сигналу

Тут можна згадати і пов'язані з цією моделлю формули Ейлера. Для введення комплексного зображення синусоїдальної величини формула Ейлера має вигляд

$$e^{j\cdot(\omega\cdot t+\varphi)} = \cos(\omega\cdot t+\varphi) + j\cdot\sin(\omega\cdot t+\varphi), \qquad (2.1)$$

де ω – кругова частота;

φ – початкова фаза;

t – змінна часу.

У зв'язку з цим, для невідомої амплітуди $(A(t))$ та довільного сигналу $x(t)$ будемо мати такий вираз

$$x(t) = A(t)\cdot e^{j\cdot(\omega\cdot t+\varphi)} = A(t)\cdot\cos(\omega\cdot t+\varphi) + j\cdot A(t)\cdot\sin(\omega\cdot t+\varphi)\ , \qquad (2.2)$$

При векторному поданні аналітичного сигналу величина, пов'язана з косинусом (a) відноситься до дійсної осі, а синусоїдальна складова (b) відповідно до уявної осі (див. рис.2.2).

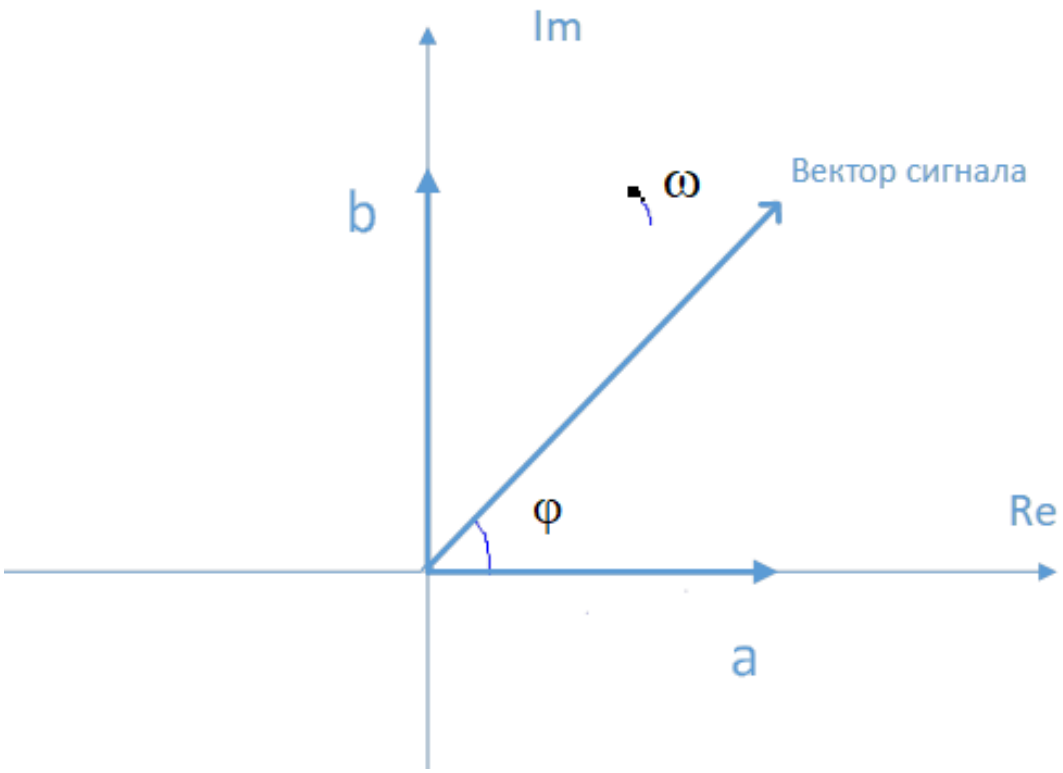

Рисунок 2.2 – Векторне подання сигналу

Тепер введемо поняття миттєвих амплітуди, фази і частоти сигналу. Для цього скористаємося рисунком 2.2, на якому зображений вектор деякого сигналу для певного моменту часу. Як відомо, представлений вектор можна розкласти на дві складові a і b.

При цьому a має назву синфазної складової і визначається наступним виразом

$$a = A(t) \cdot \cos(\omega \cdot t + \varphi), \tag{2.3}$$

a b відповідно

$$b = A(t) \cdot \sin(\omega \cdot t + \varphi), \tag{2.4}$$

яка має назву квадратурної складової.

В даному випадку (див. вирази наведені вище) φ має назву початкової фази сигналу. Вектор сигналу обертається навколо початку координат зі кутової (кругової) швидкістю ω проти годинникової стрілки.

Можна значення фази (φ) розраховувати і для кожного моменту часу (t), використовуючи наступне співвідношення

$$\varphi(t) = \text{arctg}(b/a), \tag{2.5}$$

тобто для визначення поточної фази необхідно знати синфазну і квадратурну складові.

В якості математичної моделі сигналу використовується також функціональна залежність, аргументом якої є циклічна ω або кутова частота, тобто сигнал розглядається як функція частоти. Ця функціональна залежність, що є по суті спектральним поданням сигналу, отримала назву спектру сигналу. Таке уявлення сигналу частіше розглядають не як власне сигнал, а як характеристику сигналу в частотній області.

Широке використання в даний час дискретних і цифрових систем привело до необхідності застосовувати дискретизовані сигнали. При цьому розрізняють сигнали: дискретні за часом; квантовані за рівнем; цифрові (дискретні за часом і квантовані за рівнем).

Наприклад, в цифровому поданні зареєстрований голосовий сигнал, а саме його синфазну складову, можна представити в наступному вигляді

$$u_i = A_i \cdot \exp\{j \cdot \lfloor 2 \cdot \pi \cdot f_o \cdot (i-1)/f_д + \varphi_i \rfloor\}, \tag{2.6}$$

де A_i – поточна амплітуда голосового сигналу;

f_o – частота несучого коливання;

$f_д$ – частота дискретизації голосового сигналу;

φ_i – фазові дані;

$i = 1,..., N$ – номер відліку аналізованого сигналу;

N – кількість аналізованих відліків.

При цьому, зв'язок кругової частоти і частоти несучого коливання пов'язані наступним співвідношенням

$$f_o = \omega/(2 \cdot \pi). \tag{2.7}$$

Таким чином, геометричні методи в теорії сигналів засновані на представленні сигналу як вектору в просторі векторів, які відповідають певним умовам (лінійності, ортогональності). При цьому можливе використання поняття лінійного простору дійсних або комплексних сигналів з властивостями лінійного простору векторів.

Причиною об'єднання сигналів в безліч, що утворить простір сигналів, є наявність загальних властивостей, що задовольняють принципам лінійності. При цьому є можливість одні елементи безлічі висловити через інші. Дослідження властивостей сигналів в рамках векторного уявлення виявляється корисним для синтезу пристроїв, які відповідають принципу суперпозиції.

Періодичний сигнал можна представити у вигляді суми нескінченного числа гармонійних складових (синусоїдальної і косинусоїдальної), кожна з яких характеризується своєю амплітудою і частотою. Сукупність цих складових називають спектром сигналу, а сукупність їх амплітуд – амплітудним спектром сигналу, який будемо використовувати нижче.

Аналіз вище наведених співвідношень дозволяє виділити наступні інформаційні параметри голосового сигналу: амплітуда, частота і фаза. На жаль, останній інформаційний параметр радіосигналів – поляризація, пов'язаний зі складнощами реєстрації для голосових сигналів. При цьому в сучасних системах голосової автентифікації використовується амплітуда і частота матеріалів реєстрації.

Тепер коротко зупинимося на методиці формування фазової інформації за голосовими даними користувача.

У радіолокації і радіозв'язку, де широко і ефективно використовується фазова інформація, спочатку для формування фази сигналу використовувалися фазові вращатели, які неможливо застосовувати в системах автентифікації. Очевидно, це було однією з причин ігнорування фазових даних в системах голосової автентифікації.

Останнім часом стан справ у формуванні фазових даних істотно змінилося, вони обчислюються програмно-алгоритмічно, за допомогою цифрових сигнальних процесорів або спеціалізованих мікросхем. В основу програмно-алгоритмічних процедур покладено перетворення Гільберту, яке має наступний вигляд [35]

$$u_m(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t - \tau} d\tau, \qquad (2.8)$$

де τ – змінна інтегрування;

$u(\tau)$ – синфазна складова.

В результаті буде сформована квадратурна (уявна) складова аналітичного сигналу.

Формування фазових даних в цифровому варіанті має такий вигляд

$$\varphi(i) = \text{arctg}((u_m(i)/u(i)). \qquad (2.9)$$

Реалізація перетворення Гільберту в системі комп'ютерної математики MatLab виконується за допомогою функції heming, яка буде використовуватися нижче.

На жаль, функція arctg видає значення кута в діапазоні від $-\pi/2$ до $\pi/2$. Для визначення правильного значення фазового кута, який у голосового сигналу змінюється в межах від 0 до 2π, необхідно кут $\varphi(t)$ відповідним чином відкоригувати з урахуванням знаків чисельника і знаменника в співвідношенні функції arctg. В іншому випадку фазовий спектр буде некоректним. Після корекції отримаємо фазовий кут, який має форму пилоподібного сигналу невідомої тривалості.

Як показали результати попередніх досліджень [31, 33], після формування фазових даних необхідно виконати процедури їх попередньої обробки. Це обумовлено деякими факторами, серед яких виділимо наступні:

- полігармонічний характер голосового сигналу, який обробляється перетворенням Гільберту. Останнє орієнтоване на роботу з гармонійними стаціонарними даними;

- некоректні дані при рівності нулю складових $u_m(i)$ або $u(i)$ у функції arctg;

- при малих значеннях складових $u_m(i)$ або $u(i)$, останні можуть губитися в шумах округлення.

Зазначені фактори призводять до того, що в пилкоподібних фазових сигналах можуть мати місце як випадкові помилки, так і аномальні вимірювання. Цим обумовлена необхідність попередньої обробки, як голосового сигналу, так і фазо-

вих даних. Попередня обробка може базуватися на апріорних даних щодо характеру зміни фази голосового сигналу і дозволить поліпшити якість формування характеристик як існуючих, так і перспективних складових шаблонів.

Таким чином, фазові дані в голосовій автентифікації можуть використовуватися за кількома напрямками:

- підвищення відносини сигнал/шум матеріалів реєстрації (відомий напрямок використання фази в радіолокації і радіозв'язку);

- підвищення якості формування ознак для традиційно використовуваних шаблонів, наприклад, частоти основного тону, формантної інформації і т.д.;

- розробка нових процедур формування елементів шаблонів на основі фазових даних.

2.4 Методичні засади для проведення досліджень

В основу методологічних засад досліджень в рамках магістерської роботи використовувалися експериментальний метод і метод моделювання. У зв'язку з цим була розроблена експериментальна установка, представлена на рис. 2.3.

Основу експериментальної установки становив ноутбук з операційною системою. До звукової карти ноутбука підключався мікрофон. Всі процедури формування звукових файлів і подальшої обробки, в рамках проведених досліджень, виконувалися в системі комп'ютерної математики (СКМ) MatLab.

При цьому виконувалися процедури управління звуковою картою, які включали завдання частоти дискретизації, кількості біт для квантування амплітуди і час реєстрації голосового сигналу.

Формований файл голосових даних представлявся в форматі wav. Для обробки голосових даних використовувалися як стандартні функції, так і алгоритмічну мову для розробки m-файлу СКМ MatLab.

Розглянемо схему проведення експериментальних досліджень, яку будемо використовувати нижче.

В якості голосового сигналу використовувалися цифри від 0 до 9 за допомогою яких, наприклад, можна задати пароль користувача. Для виключення імітації пароля іншим користувачем, він може формуватися і видаватися на екран системою голосової автентифікації.

Частота дискретизації становила 64 кГц. Кількість біт квантування амплітуди – 8.

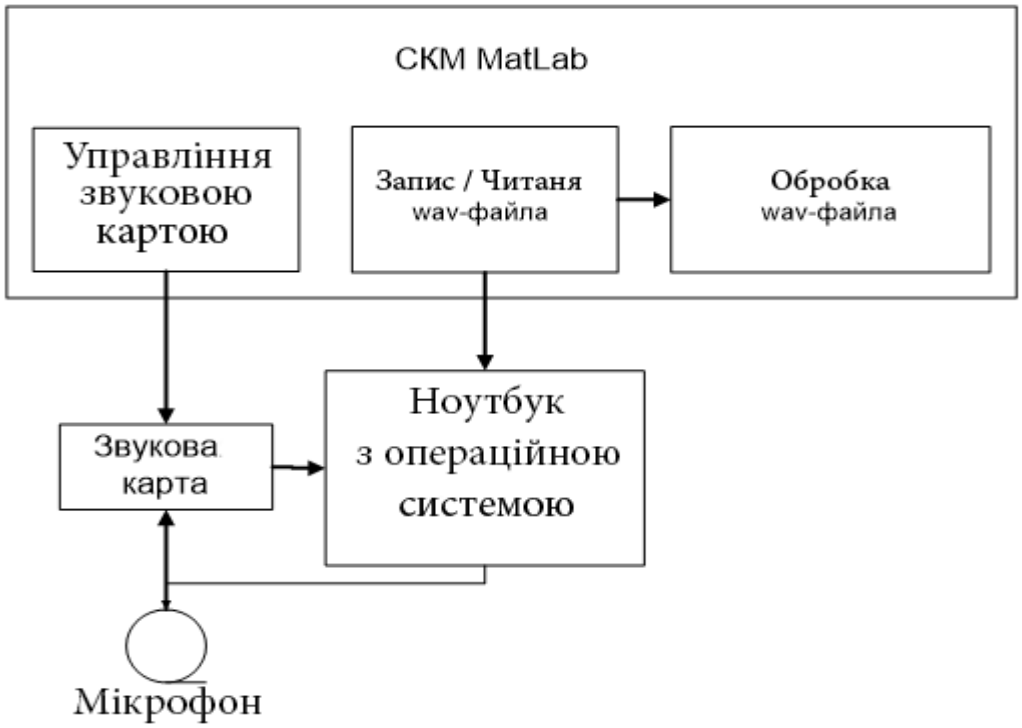

Рисунок 2.3 – Експериментальна установка для проведення досліджень

Ефективність запропонованого методу підвищення якісних показників систем голосової автентифікації буде проводитися в процесі оцінки частоти основного тону, яка, як правило, входить до складу всіх шаблонів користувачів. Відомо, що величина частоти основного тону (F0) є індивідуальною характеристикою диктора. Вона може змінюватися в залежності від емоційного забарвлення мови, але в досить вузьких межах. При параметричному кодуванні мови припускають, що частота основного тону людини лежить в межах 80 - 400 Гц, а більшість формантних частот кратні F0.

Оцінка частоти основного тону може застосовуватися для вирішення широкого кола завдань:

- розпізнавання емоцій;
- визначення статі користувача;
- сегментація аудіо з декількома голосами або поділу мови на фрази.

Крім цього, частоту основного тону можна використовувати в медицині для визначення патологічних характеристик голосу або виявлення при-знаків захворювання Паркінсона.

2.5 Аналіз відомих методів оцінки частоти основного тону

Одним з найбільш важливих етапів, який, в кінцевому рахунку, буде визначати якість автентифікації, є вибір класифікаційних ознак для формування шаблонів.

Як правило, в якості класифікаційної ознаки, за якими проводиться автентифікація користувача (розпізнавання диктора), використовують частоту основного тону. Однак, як показує практика, однієї частоти основного тону недостатньо для достовірної автентифікації користувача. Тому додатково використовується формантна інформація, яка базується на зазначеній частоті. Формантна інформація доповнюється кепстральними ознаками, коефіцієнтами лінійного передбачення та ін.

Оцінка частоти основного тону – одна з основних задач цифрової обробки мовних сигналів. Результати її рішення використовуються при рас-пізнавання і стисненні мови, ідентифікації / верифікації диктора і т.д.

Методи, призначені для вирішення даного завдання, аналізують мовний сигнал в частотній і часовій областях.

У часовій області оцінка частоти здійснюється за кількістю перетинів сигналом нульового рівня, за автокореляційною функцію, за функцію середнього значення різниці амплітуд.

Алгоритми, що обробляють дані в частотної області, аналізують гармоніки спектру, використовують кепстральний аналіз.

В [29] використовується підхід до оцінки частоти основного тону, який виконує одночасну обробку сигналу в частотній і часовій областях. Коротко розглянемо деякі алгоритми оцінки частоти основного тону з зазначених груп.

Метод амплітудної селекції в часовій області полягає в тому, що на стаціонарній ділянці вокалізованого сигналу при незначному рівні шумів, форма мовного коливання майже так само повторюється на кожному черговому періоді основного тону. Відстань між максимумами мовного сигналу можна приблизно вважати рівним періоду основного тону. Основна проблема алгоритмів амплітудної селекції полягає в необхідності придушення помилкових локальних максимумів. Цього можна домогтися за допомогою підвищення порогового значення виявлення в пошуку максимумів. Однак при цьому збільшується ймовірність помилкового визначення за рахунок пропуску істинного максимуму. Пропуск або втрата максимуму може привести до істотних спотворень звуку після обробки. Додавання

другого каналу амплітудної селекції, що виділяє положення мінімумів мовного сигналу, підвищує надійність визначення періоду основного тону.

Головним достоїнством пристроїв часової селекції є простота в реалізації. Основний недолік – невисока точність і нестійке визначення основного тону навіть при відносно невеликому рівні шумів.

Як відомо, в спектрі звукового сигналу присутні піки на частотах, кратних частоті основного тону. Якщо побудувати дискретне перетворення Фур'є з досить малим кроком дискретизації по частоті, то в якості оцінки частоти основного тону можна використовувати частоту, відповідну максимальному значенню енергії спектра. Поняття енергії і потужності в теорії сигналів не відносяться до характеристик будь-яких фізичних величин сигналів, а є їх кількісними характеристиками, що відображають певні властивості сигналів і динаміку зміни їх значень в часі. Для довільного, в загальному випадку комплексного, енергія сигналу дорівнює інтегралу від потужності по всьому інтервалу існування або завдання сигналу. Потужність по визначенню дорівнює квадрату функції його модуля, для речових сигналів – квадрату функції амплітуд. Пошук максимуму енергії спектру слід проводити в інтервалі 80 - 400 Гц. Однак має місце ситуація, коли в зазначеній смузі лежить і друга гармоніка основного тону, яка іноді має більшу енергією.

Кореляційні методи можуть використовуватися як в часовій, так і в частотній області. В основі кореляційних методів визначення періоду ЧОТ мовного сигналу закладені принципи оцінки середнього значення періоду пульсацій квазіперіодичної кореляційної функції. Як відомо, кореляційна функція є Фур'є-перетворенням енергетичного спектру, і положення її піків відповідають відстаням між рівномірно розташованими гармоніками спектру. В окремому випадку обчислюється перший глобальний максимум кореляційної функції. Однак область частот першої форманти (область посилених частот) досить відчутно впливає на якість роботи кореляційного аналізу. Найважливішим параметром, що характеризує спектр (розподіл енергії або амплітуди за частотами) мовного сигналу, є форманти, що представляють собою концентрації енергії в обмеженою частотної області.

Форманта характеризується частотою, шириною і амплітудою. За частоту форманти приймають частоту максимальної амплітуди в її обмежених межах. Іншими словами, форманта – це деякий амплітудний сплеск на графіку спектру, а його частота – частота піку цього сплеску. Голосовий тракт в силу своїх резонансних властивостей вносить в формований сигнал набір характерних для кожної

людини частотних складових, які називаються формантами. Частоти і смуги цих формант можуть управлятися зміною форми голосового тракту, наприклад, зміною положення язика. Для вирівнювання спектра може бути використана або зворотна фільтрація на основі лінійного передбачення, або поділ сигналу на кілька частотних смуг з обчисленням кореляційної функції в кожній смузі з нормування і підсумовуванням.

Відомо, що в кепстрі присутні яскраво виражені максимуми в діапазоні від 2 мс до 20мс і вони дуже точно вказує на те, що даний кадр є вокалізованим, а положення максимуму визначає період сигналу, що аналізується. Для визначення кепстру виконують зворотне перетворення Фур'є комплексного логарифму спектру потужності сигналу на кадрі аналізу. Для збільшення швидкості обчислень перетворення Фур'є виконують за допомогою алгоритму швидкого перетворення Фур'є (ШПФ). Тривалість аналізованого кадру повинна перевищувати тривалість принаймні двох-трьох найбільш довгих для даної фонограми періодів основного тону і бути кратною ступеня двох, що становить зазвичай 512 відліків для низьких чоловічих голосів і 256 для жіночих і високих чоловічих голосів (при частоті дискретизації, рівної 10 кГц). У той же час, чим більше вікно, тим довше вважається основний тон і тим гірше відслідковуються швидкі зміни ЧОТ. Для зняття ефекту накладення частот використовують вікна, наприклад, Ханна.

Виходячи з викладеного вище далі для отримання оцінки ЧОТ будемо використовувати спектральний і кепстральний аналіз голосового сигналу користувача.

Для цього побудуємо амплітудно-частотний спектр аналізованого голосового сигналу користувача, який виголошував цифру «один». Максимум рівня спектральної щільності буде відповідати оцінці ЧОТ, яку будемо вважати еталонною. Після цього виконаємо оцінку інформативності фазових даних аналізованого голосового сигналу. Для цього сформуємо фазові дані і проведемо їх попередню обробку, відповідно до процедур, які розглянуті вище.

Далі побудуємо фазовий спектр аналізованого сигналу і проведемо його попередню обробку, пов'язану з виключенням впливу мовного тракту. Після чого можемо визначити максимум рівня спектральної щільності, який буде оцінкою ЧОТ. Природно очікувати, що отримана оцінка повинна збігатися або бути близькою з оцінкою, отриманою в процесі аналізу амплітудно-частотного спектра.

2.6 Методика оцінки частоти основного тону при формуванні кепстральних коефіцієнтів

Тут же слід звернути увагу ще на один факт. При формуванні шаблону, як правило, розраховуються і використовуються кепстральні або мел-кепстральні коефіцієнти. Зазначені коефіцієнти припускають розрахунок спектру і кепстру за фрагментами (семплами) аналізованого сигналу. Кількість коефіцієнтів становить декілька десятків. У цих фрагментах можна оцінювати ЧОТ для вокалізованих звуків аналізованого голосового сигналу. Таким чином, можна отримати кілька десятків оцінок ЧОТ як по амплітудним, так і по фазовим даними аналізованого сигналу. Подальша статистична обробка дасть можливість отримати додаткову інформацію щодо ЧОТ користувача.

Розглянемо запропонований метод оцінки ЧОТ більш докладно. Як відомо, кепстральний аналіз дозволяє більш якісно досліджувати спектр аналізованого сигналу. Досягається це за рахунок того, що кепстр дозволяє «згладити» спектр і більш якісно виділити максимуми останнього. Саме це необхідно при визначенні оцінки ЧОТ.

У зв'язку з цим при оцінці ЧОТ необхідно реалізувати схему розрахунку кепстру (див. рис.2.4), яка виконується в семплах. Розмір семплу (nfft) вибирають тривалістю кілька десятків мілісекунд і кратним ступеню 2, що дозволяє скористатися швидким перетворенням Фур'є.

Рисунок 2.4 – Загальна схема кепстрального аналізу сигналу

Семпли вибираються з деяким перекриттям. Нижче будемо вважати, що коефіцієнт перекриття k = 0.75. Кожен семпл вихідного сигналу піддається попередній обробці (поелементному множенню) вікном Хеммінга для зменшення обчислювальних шумів [35]

$$w(n) = 0.54 - 0.46 \cdot \cos(\frac{2 \cdot \pi \cdot n}{N}), \quad (2.10)$$

де N – ширина вікна в відліках;

$n = 1,..., N$.

Обробка може бути реалізована в матричному вигляді, або в «ковзному» вікні. І в тому і іншому вигляді необхідно для аналізованого масиву даних (nx) визначити кількість оброблюваних елементів в семплі (nfft), номери відліків початкових елементів для кожного семплу (j) і їх загальна кількість (ncol).

Для цього спочатку необхідно визначити число елементів масиву, які «перекриваються» в семплі за допомогою наступного співвідношення

$$nlap = k \cdot nfft. \quad (2.11)$$

Загальна кількість оброблюваних семплів (стовпців при матричному варіанті обчислень) визначається, як ціла частина від наступної дроби

$$ncol = \frac{n_x - nlap}{nfft - nlap}. \quad (2.12)$$

Далі необхідно визначити масив початкових елементів оброблюваних семплів

$$j = 1 + i \cdot (nfft - nlap), \quad (2.13)$$

де $i = 1,...., ncol$.

Як зазначено вище, обраний семпл поелементна множиться на масив відліків вікна Хеммінга, а потім виконується ШПФ результатів множення. З результатів ШПФ виділяється матеріальна частина перетворення (y), яка піддається наступній обробці

$$r = \log(abs(y)). \quad (2.14)$$

Потім результат логарифмування r піддається зворотному перетворенню Фур'є. Зазначена процедура дасть можливість згладити спектр і більш якісно виділити максимуми спектра, тобто сформувати кепстр аналізованого семплу (R).

Як відомо, початкові значення як спектру, так і кепстру містять інформацію про мовний тракт користувача, які доцільно виключити з оцінки ЧОТ. Для цього можна речову частину спектру (кепстру) від першого до 0.002 $f_д$_ відліків виключити з подальшого аналізу, де $f_д$ – частота дискретизації аналізованого голосового сигналу. З решти елементів масиву R вибирається максимальне значення

$$R_{max} = \max(R), \tag{2.15}$$

яке порівнюється з деяким порогом P.

При перевищенні порогу (аналізу піддається вокалізований звук) фіксується значення частоти R_{max}, яке є поточною оцінкою ЧОТ в даному семплі. Аналогічно формується і обробляється наступний семпл. Отримані оцінки піддаються статистичній обробці (оцінюється середнє значення ЧОТ і середньоквадратичне відхилення).

Зауважимо, що зазначена обробка може проводитися, як при аналізі амплітудних, так і фазових даних голосового сигналу.

Аналогічна методика оцінки ЧОТ може застосовуватися і при формуванні мел-кепстральних коефіцієнтів.

3 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ ГОЛОСОВОГО СИГНАЛУ КОРИСТУВАЧА СИСТЕМИ АВТЕНТИФІКАЦІЇ

3.1 Методика проведення експериментальних досліджень

Структурна схема методики проведення досліджень представлена на рис. 3.1. Текст програми моделі досліджень наведений у додатку Б.

Рисунок 3.1 – Методика проведення досліджень голосового сигналу користувача

Аналізу будемо піддавати експериментальний голосовий сигнал користувача системи автентифікації, який виголошував цифру «один». Частота дискретиза-

ції 64 кГц і співвідношення сигнал / шум більше 20 дБ. Аналізований голосовий сигнал представлений на рис. 3.2.

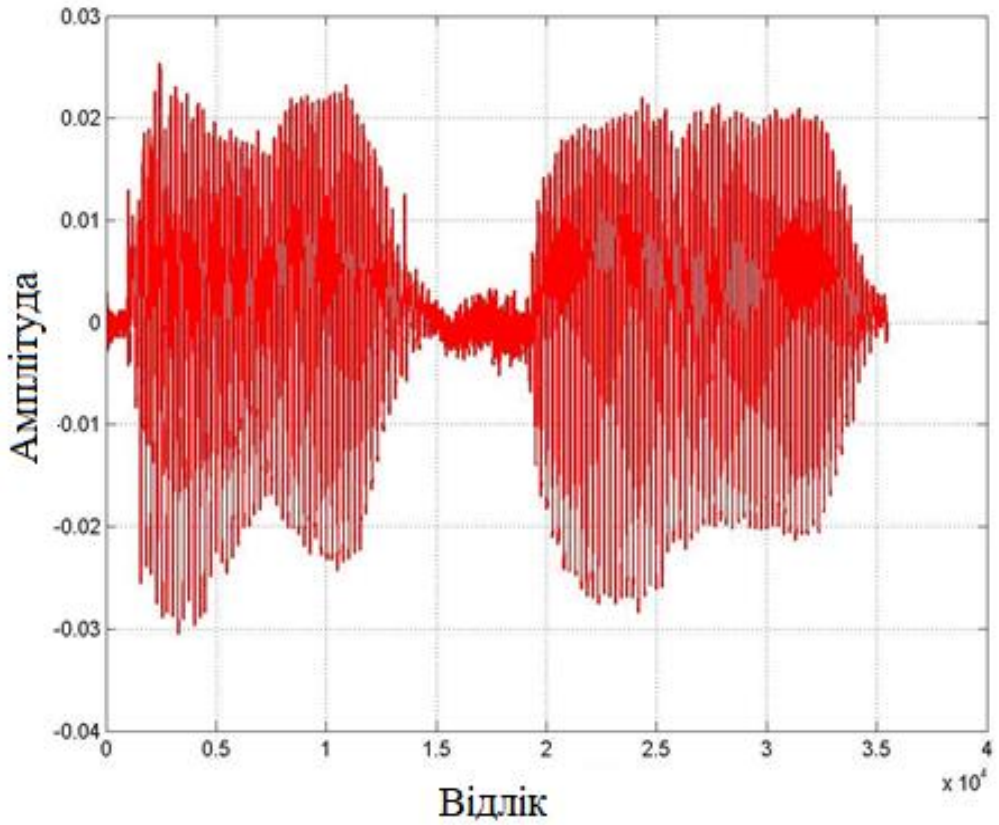

Рисунок 3.2 – Аналізований голосовий сигнал цифри «один»

Як і в відомих сучасних СГА, спочатку розрахуємо амплітудно-частотний спектр з експериментального голосового сигналу і виконаємо його аналіз. При цьому, як зазначено вище, основну увагу будемо приділяти області низьких частот, де знаходяться ознаки користувача системи автентифікації, зосередивши особливу увагу на частоті основного тону. Деяка увагу приділимо і форматним частотам, які, як відомо, пов'язані з ЧОТ.

3.2 Результати експериментальних досліджень амплітудного та фазового спектрів голосового сигналу

Відомо, що величина ЧОТ є індивідуальною характеристикою користувача системи автентифікації. Вона може змінюватися в обмежених межах в залежності

від емоційного забарвлення мови. Дослідники припускають, що ЧОТ користувача системи автентифікації знаходиться в межах 80 - 400 Гц, а формантні частоти кратні *F0*.

Амплітудно-частотний спектр аналізованого сигналу представлений на рис. 3.3.

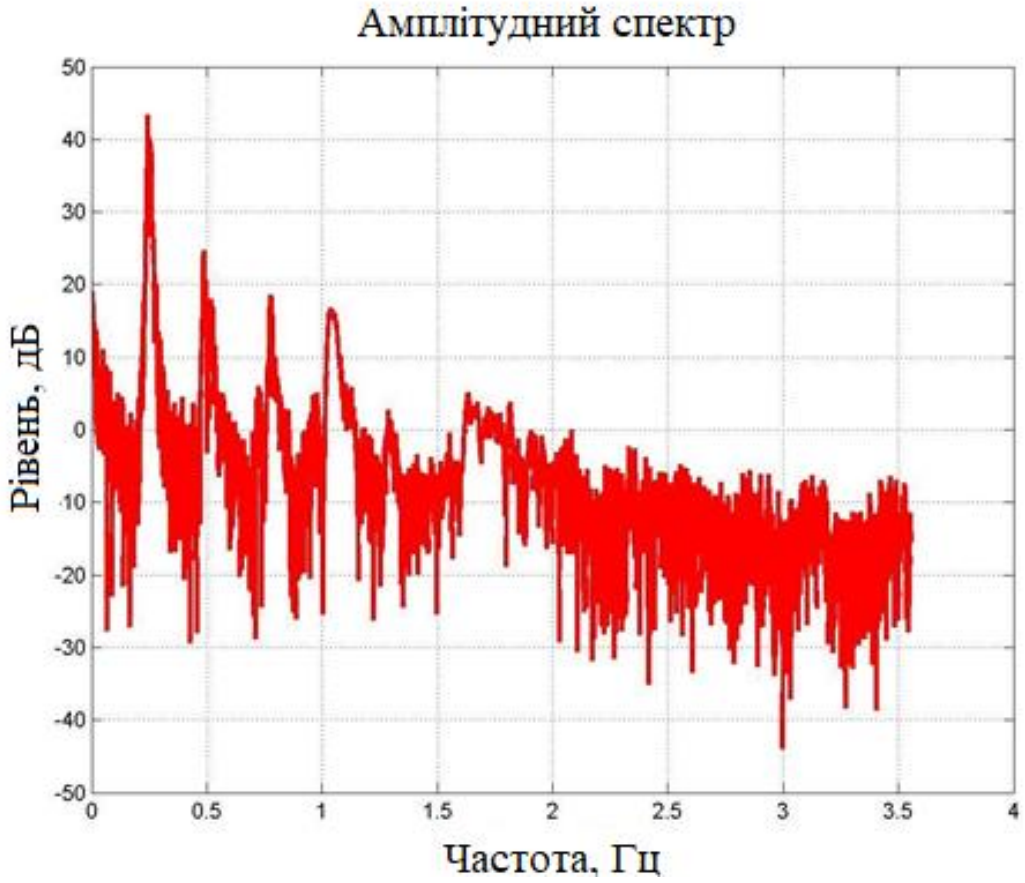

Рисунок 3.3 – Амплітудно-частотний спектр (короткий) голосового сигналу цифри «один»

Аналіз короткого амплітудно-частотного спектра досліджуваного голосового сигналу користувача дає можливість визначити оцінку ЧОТ, яка дорівнює 243 Гц. На уже згадуваному спектрі яскраво виражені три формантних частоти (див. табл.3.1), а більш високі мають низький рівень інтенсивності.

Тепер досліджуємо розглянуті характеристики стосовно фазової інформації голосового сигналу користувача системи автентифікації. Для цього необхідно сформувати фазові дані, які для голосового сигналу не реєструються мікрофоном.

Таблиця 3.1 – Характеристики ЧОТ і формантних частот амплітудного спектру

Рівень, дБ	43.4	24.6	18.6	14.2
Частота, Гц	243	486	776	1025

Тому фазові дані, як правило, розраховують програмно - алгоритмічно. Для цього необхідно відновити за матеріалами реєстрації квадратурну (уявну) складову голосового сигналу. Зазначені процедури пов'язані із застосуванням перетворення Гільберту, яке розглянуто вище.

Як зазначено вище, функція видає значення кута в діапазоні від $-\pi/2$ до $\pi/2$. Результати розрахунку фазового кута представлені на рис. 3.4 (темна лінія).

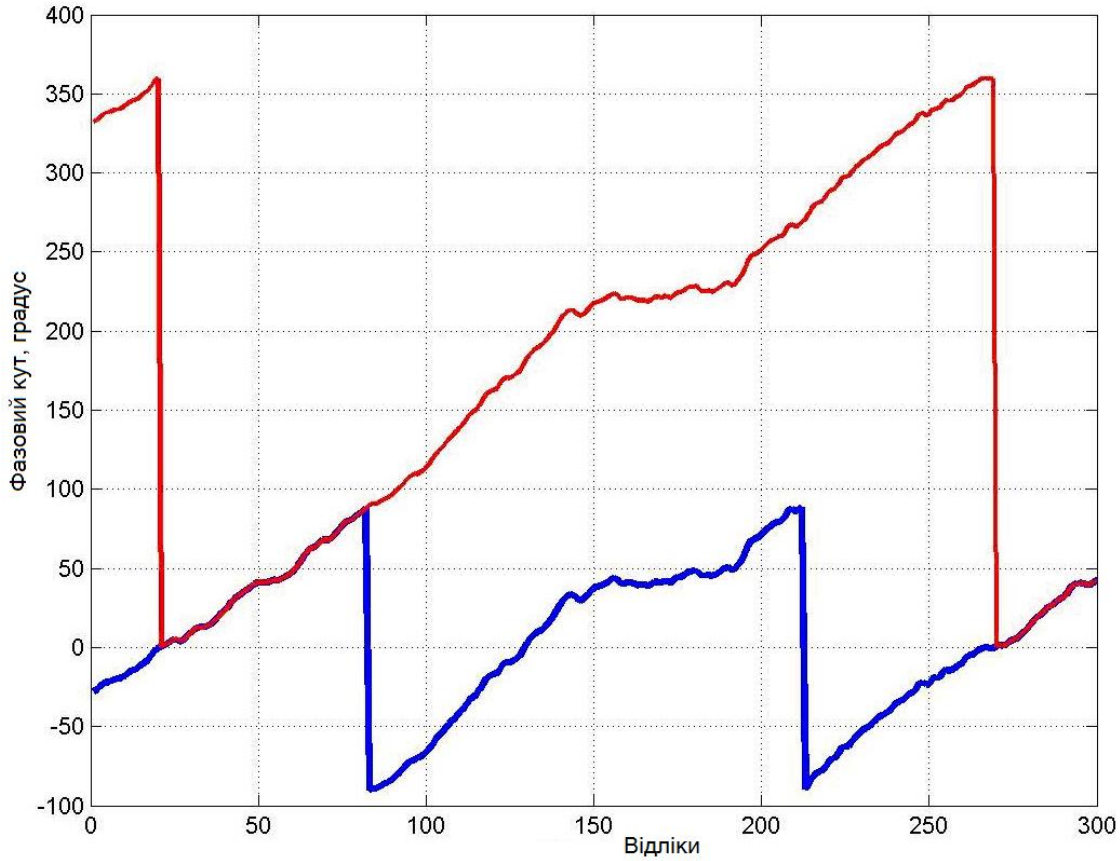

Рисунок 3.4 – Фазова інформація фрагмента аналізованого голосового сигналу

Для визначення правильного значення фазового кута, який у голосового сигналу змінюється в межах від 0 до 2π, необхідно кут відповідним чином відкоригувати з урахуванням знаків чисельника і знаменника в співвідношенні функції (див. рис. 3.4, світліша лінія). В іншому випадку фазовий спектр буде некоректним. Після

корекції отримаємо фазовий кут, який має форму пилоподібного сигналу невідомої тривалості.

Крім цього, відомо, що голосовий сигнал є нестаціонарним, характеристики якого часто змінюються в часі. Це також призводить до неякісних результатів, що отримуються за допомогою перетворення Гільберту при малому співвідношенні сигнал/шум (спочатку і наприкінці голосового сигналу) або коли дійсна складова аналізованого сигналу наближається до нуля.

Проілюструємо сказане за допомогою експериментальних даних. На рис. 3.5 представлений фрагмент аналізованого сигналу.

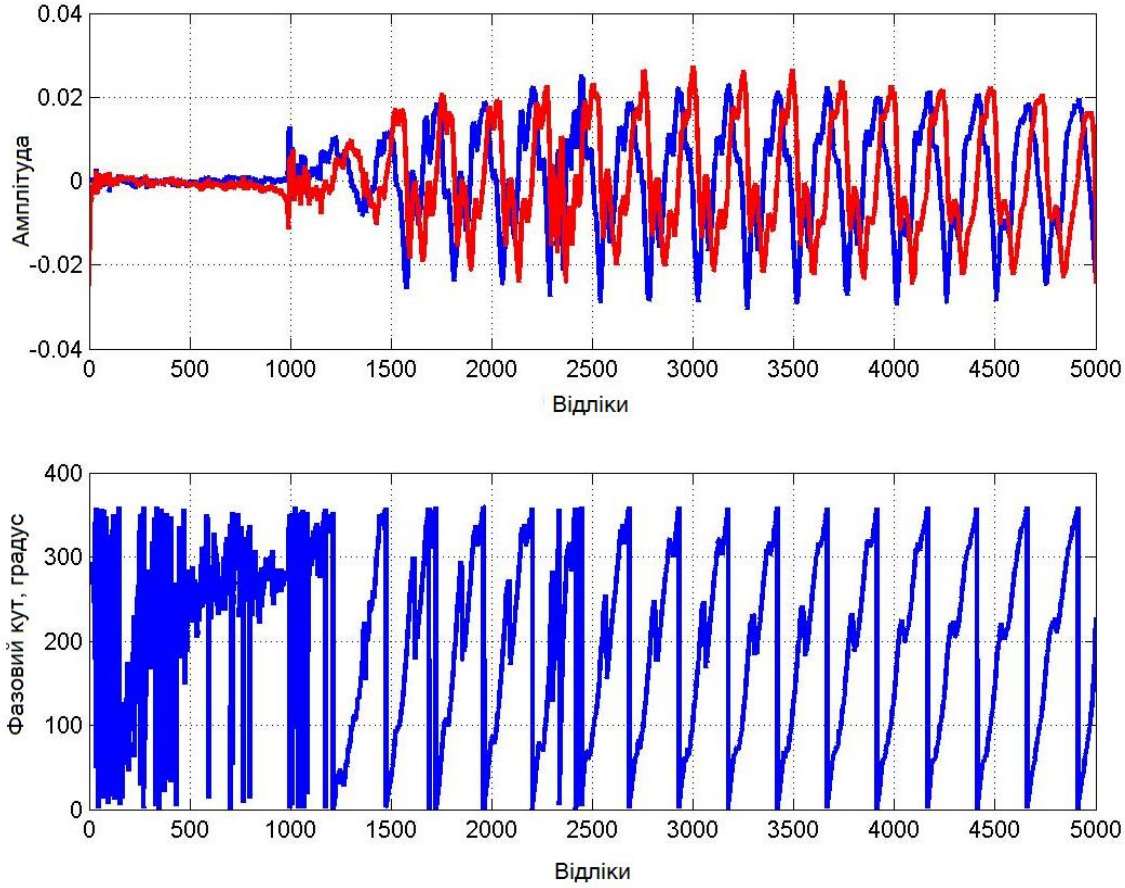

Рисунок 3.5 – Фрагмент реєстрації та обробки голосового сигналу цифри один (низька якість)

У верхній частині цього рисунку представлені залежності (речова і уявна складова) голосового сигналу і шуму. У нижній частині представлена залежність фазового кута. Аналіз представлених залежностей підтверджує низьку якість визначення фазового кута на початку сигналу. Крім того, мають місце помилки у визначенні фазового кута в діапазоні від 2250 до 2500 відліків.

У той же час слід зазначити періодичність зміни фазового кута в діапазоні від 0^o до 360^o. Зауважимо, що при реєстрації цифри відбуваються невеликі коливання речової складової. Останнє призводить до відповідних змін у фазовому куті (див. рис. 3.5).

На рис. 3.6 представлений фрагмент тієї самої цифри, яка була зареєстрована у того ж користувача системи автентифікації.

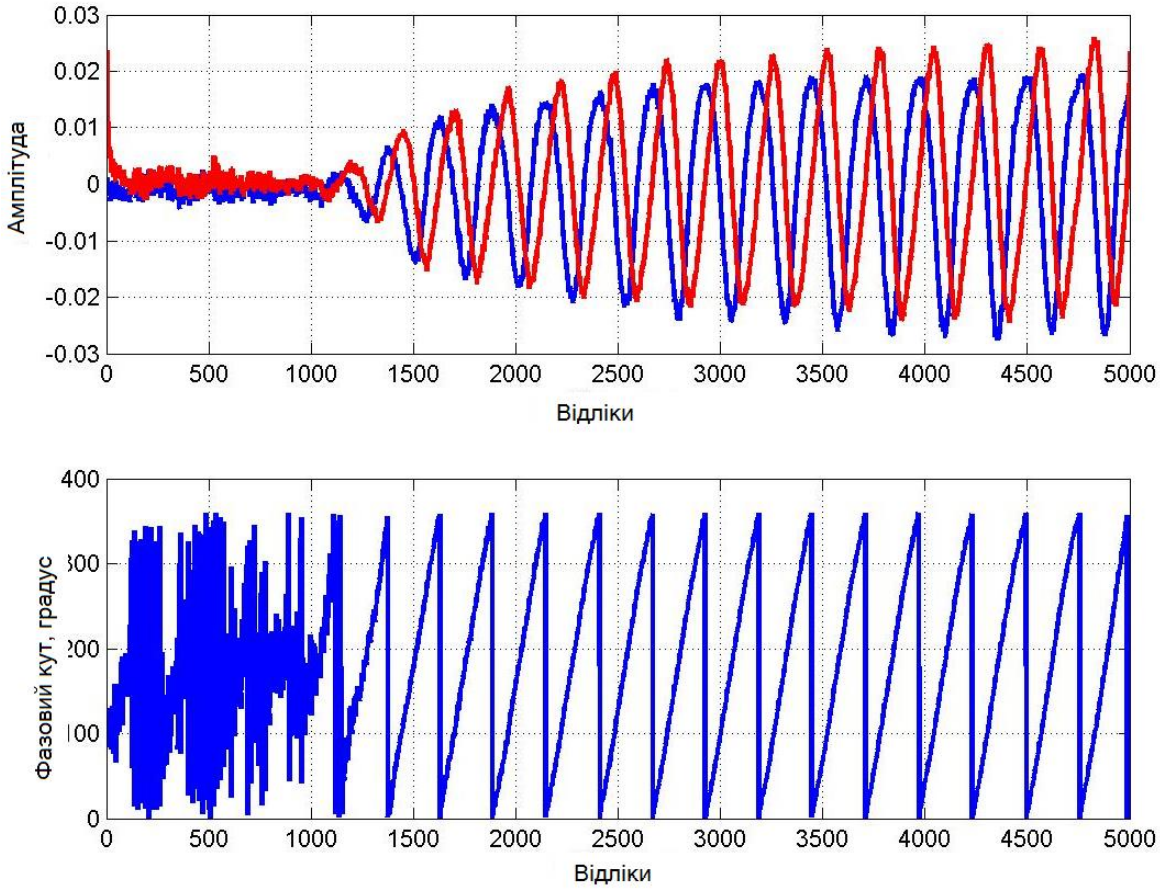

Рисунок 3.6 – Фрагмент реєстрації та обробки голосового сигналу цифри один (висока якість)

Якісна реєстрація голосового сигналу призводить до відсутності коливань фазового кута. Слід зазначити, що фазові дані мають форму пилоподібного сигналу невідомої тривалості. Цей факт (апріорну інформації про форму очікуваного сигналу) доцільно використовувати в процедурах виділення, попередньої обробки та аналізу фазових даних.

Таким чином, використання процедур попередньої обробки на основі використання апріорної інформації щодо фазових даних надає можливість усунуть не-

доліки реєстрації голосового сигналу. Цей факт дає реальну можливість підвищити якість автентифікації користивачів.

На рис. 3.7 представлені два періоди зміни фазового кута, які отримані в результаті розрахунку з використанням зареєстрованого голосового сигналу. Ці залежності показані чорним кольором. Червоним кольором на цих рисунках представлені залежності очікуваних значень фазового кута відповідно до висунутої гіпотезою про пилкоподібний зміну сигналу фази.

Аналіз представлених залежностей підтверджує достовірність висунутої гіпотези про зміну фазового кута у вигляді пилоподібного сигналу. Тривалість періоду зміни фазового кута залежить від характеристик голосового тракту користувача і змісту сигналу, що може використовуватися в системах автентифікації.

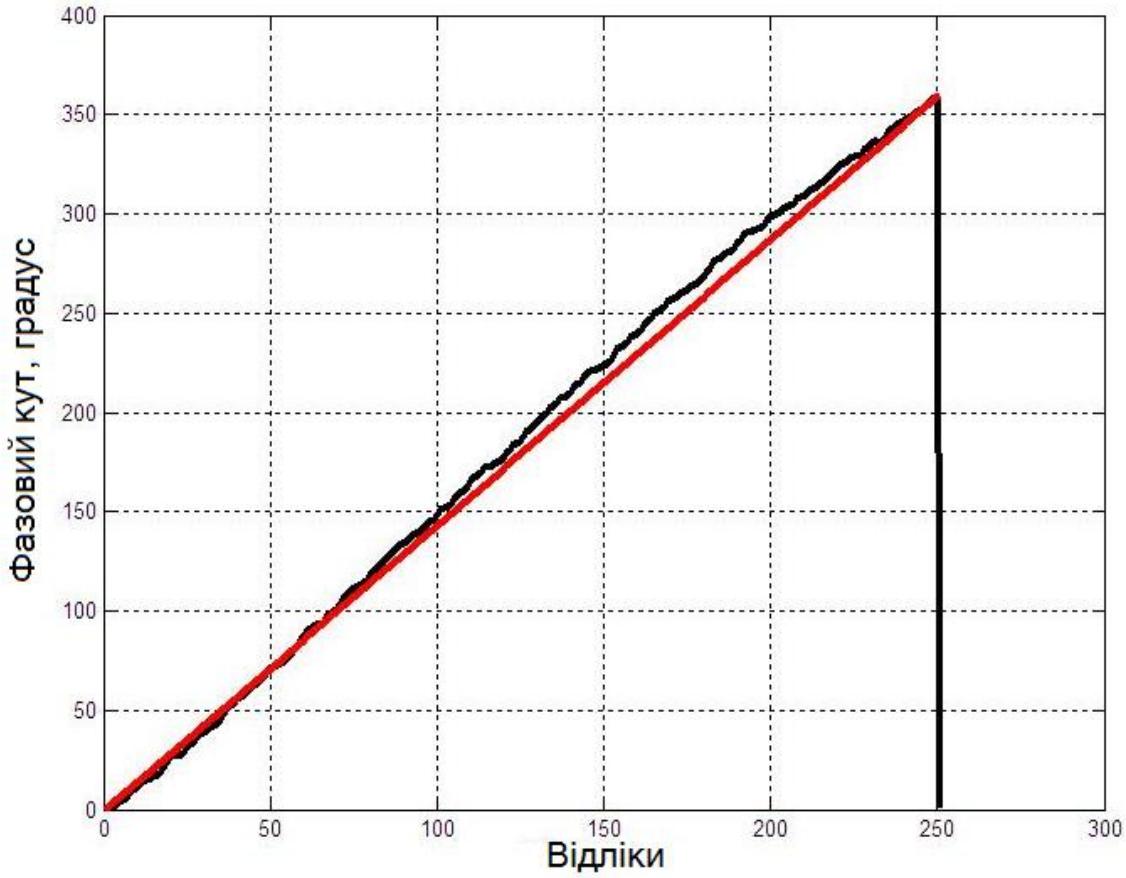

Рисунок 3.7 – Очікувана і розрахункова залежності фазового кута при відсутності помилок

Однак, в процесі реєстрації та аналізу голосового сигналу мають місце відхилення від очікуваної форми залежності фазового кута (див. рис. 3.8). При цьому, як і раніше, червоним кольором зображено очікувана залежність фазового ку-

та. Виявлені помилки у визначенні фази сигналу доцільно відкоригувати з урахуванням апріорної інформації, а потім на основі цих даних уточнити дійсну і уявну складові голосового сигналу.

По ряду причин в пилкоподібних фазових сигналах можуть мати місце, як випадкові помилки, так і аномальні вимірювання. Цим обумовлена необхідність попередньої обробки, як реєстрованого голосового сигналу, так що формуються квадратурної складової і фазових даних.

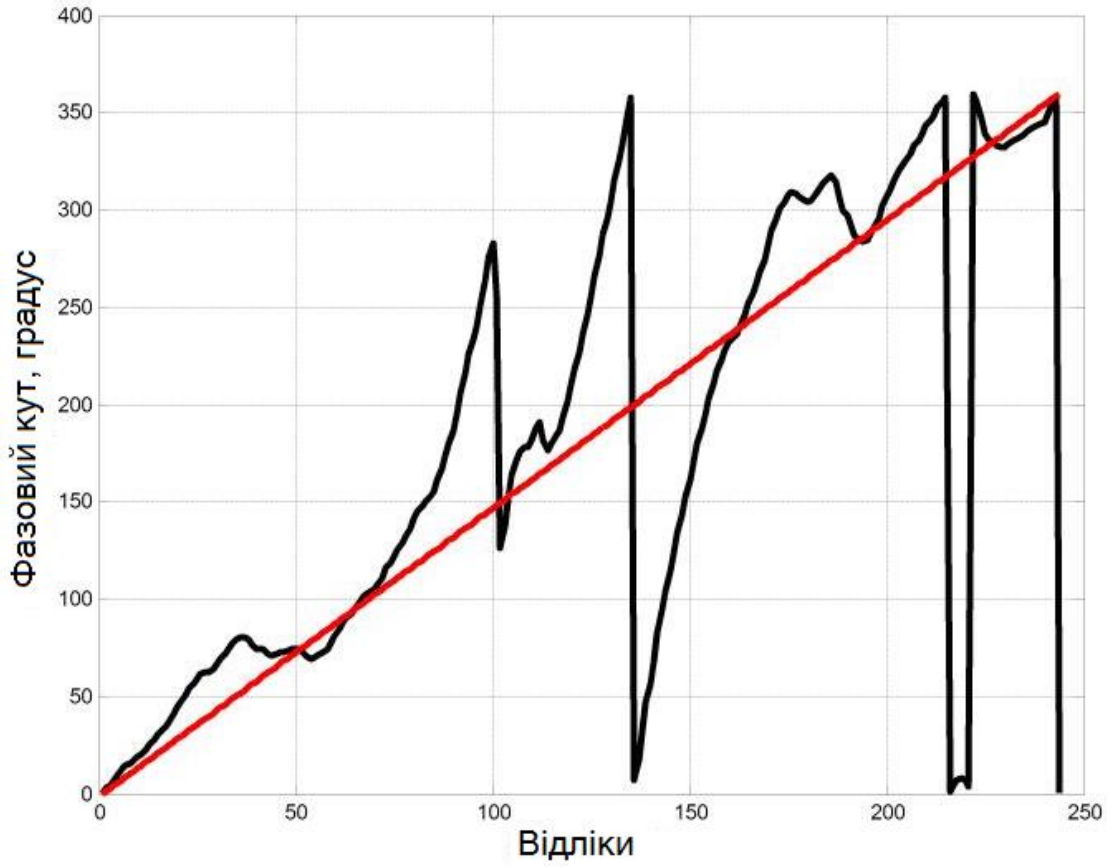

Рисунок 3.8 – Очікувана і розрахункова залежності фазового кута при наявності помилок

Попередня обробка може базуватися на апріорних даних щодо характеру зміни фази голосового сигналу і дозволить поліпшити якість формування характеристик як існуючих, так і перспективних складових шаблонів.

Нижче виконаємо аналіз фазового спектру досліджуваного голосового сигналу. На рис. 3.9 представлений розрахований фазовий спектр відкоригованих фазових даних, який будемо аналізувати нижче.

Аналіз результатів обробки формантної інформації фазового спектру досліджуваного голосового сигналу представлені в табл. 3.2. Частота основного тону досліджуваного голосового сигналу, як і в амплітудному спектрі, становить 243 Гц.

У фазовому спектрі цього сигналу можна виділити шість формант, які в більшості випадків кратні ЧОТ, а сьома і восьма мають незначну енергетичну відмінність.

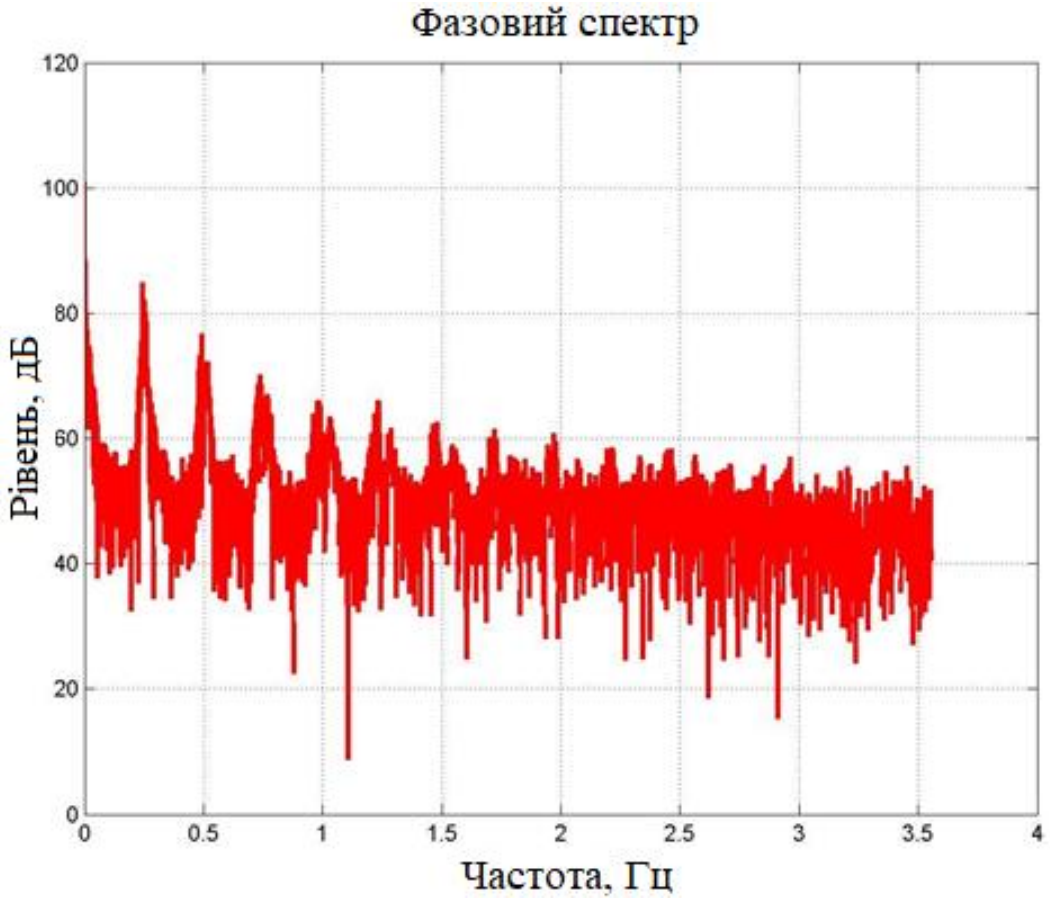

Рисунок 3.9 – Короткий фазовий спектр голосового сигналу

Таблиця 3.2 – Характеристики фазового спектра аналізованого сигналу

Рівень, дБ	84.9	76.7	70.3	65	64	62
Частота, Гц	243	492	738	990	1217	1450

Рівень спектральної щільності виділених максимумів частотного спектру в кілька разів перевищує рівень максимумів амплітудного спектру. Останнє істотно спрощує процедуру їх виділення і обробки. Кількість цих формант у фазового

спектра в півтора рази більше, що свідчить про більшу інформативність фазового спектру голосового сигналу.

В подальшому основну увагу приділімо досліджуванню оцінці частоти основного тону на основі аналізу амплітудно-частотного і фазочастотного спектрів і кепстрального аналізу.

3.3 Результати експериментальних досліджень оцінки частоти основного тону при формуванні кепстральних коефіцієнтів

Спочатку проведемо аналіз модельного сигналу, який включав послідовність двох гармонік різної амплітуди і частоти. При цьому перша гармоніка мала несучу частоту 100, а друга 500 Гц. Фрагмент аналізованого сигналу наведений на рис. 3.10.

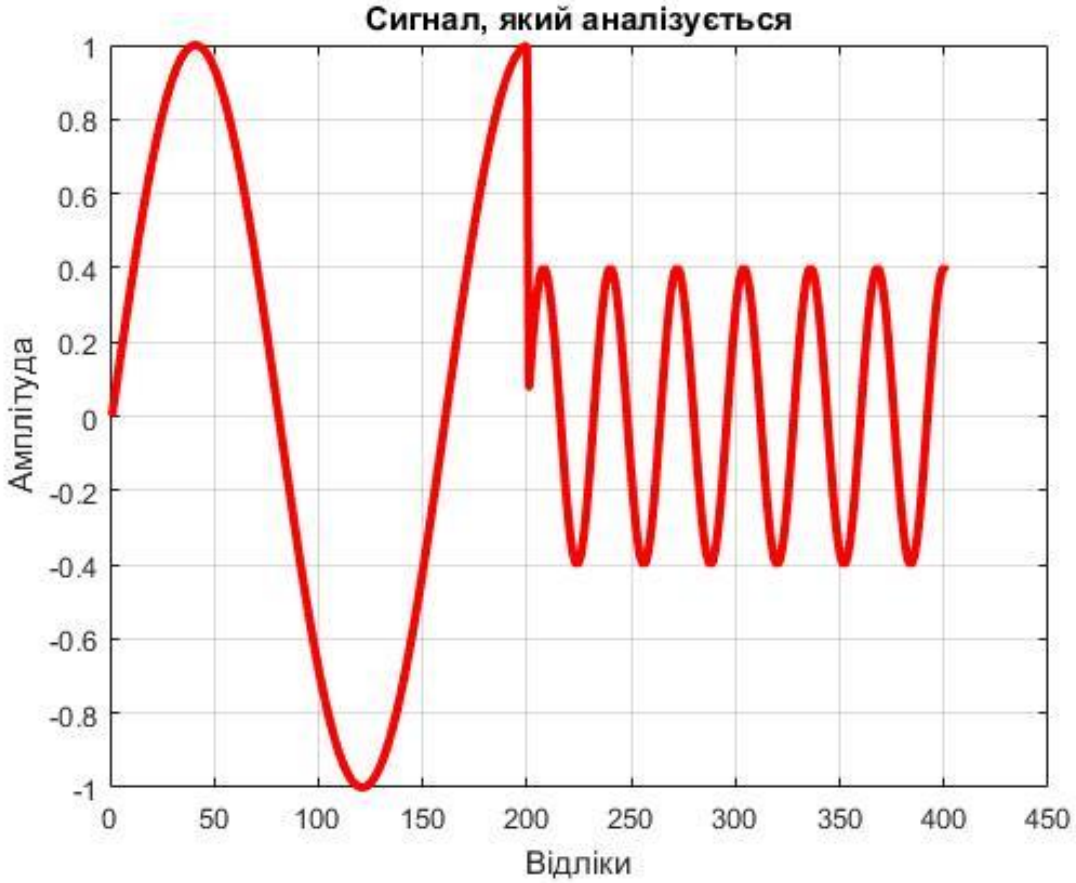

Рисунок 3.10 – Фрагмент аналізованого сигналу

Як показує аналіз амплітудного (див. рис. 3.11) та фазового (див. рис. 3.12) спектрів свідчить, що вони не дозволяють виділити перший максимум (аналог частоти основного тону).

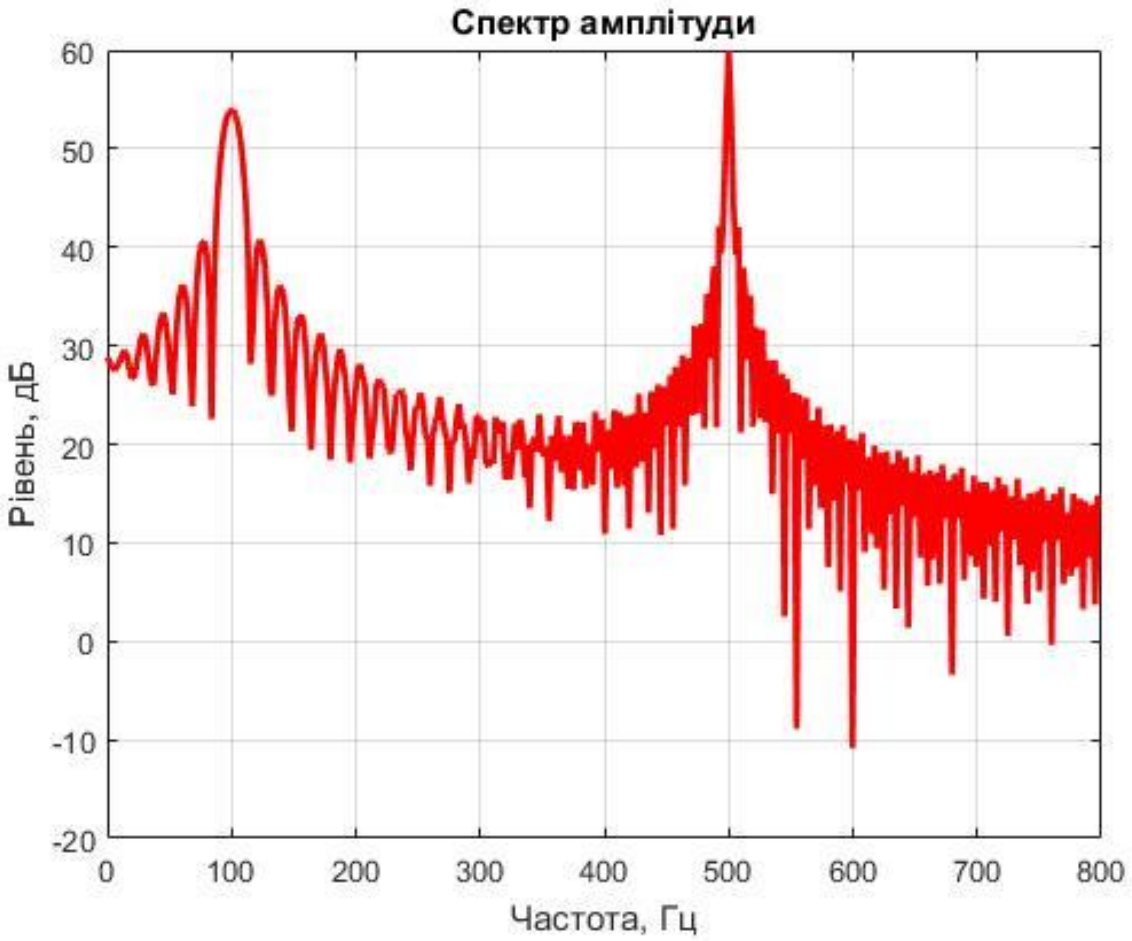

Рисунок 3.11 – Амплітудний спектр аналізованого сигналу

Нижче наведемо результати іншого шляху отримання наближеної оцінки частоти основного тону, який розглянутий вище та пов'язаний з розрахунком кепстральних або мел-частотних кепстральних коефіцієнт (Mel Frequency Cepstral Coefficient – MFCC), які, як правило, входять ознаками до шаблону користувача.

Схема розрахунку кепстральних коефіцієнтів розглянута у другому розділі та передбачає розрахунок зворотного швидкого перетворення Фур'є від амплітудного або фазового спектру аналізованого голосового сигналу.

Таким чином, кепстральні коефіцієнти є результатом застосування зворотного перетворення Фур'є до логарифму енергетичного спектру. Розрахунок цих коефіцієнтів здійснюється на семплах сигналу, які по тривалості становлять кілька десятків мілісекунд. При цьому семпли вибираються з деяким перекриттям. Ре-

зультат кожного зворотного перетворення дозволяє отримати оцінку максимуму частоти в семпли.

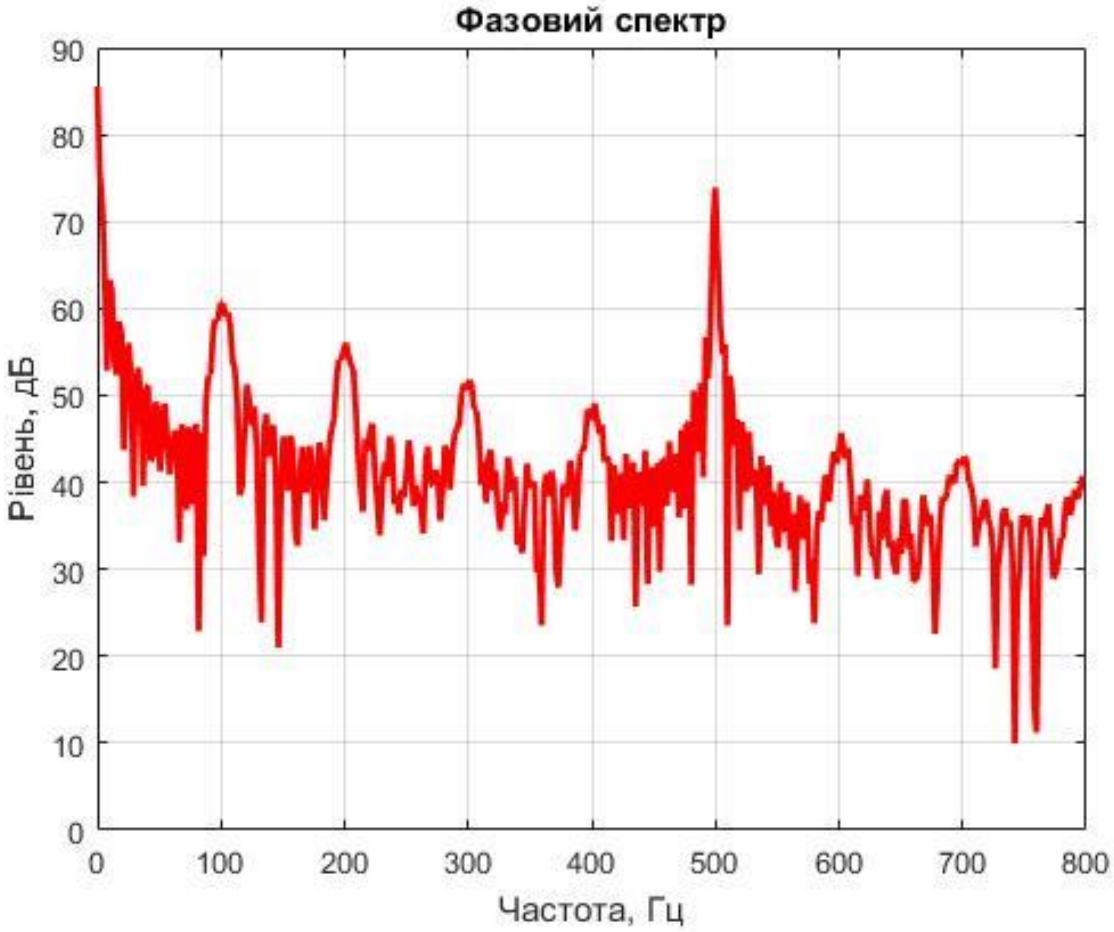

Рисунок 3.12 – Фазовий спектр аналізованого сигналу

Як правило, розраховується приблизно 40 коефіцієнтів, а значить і стільки-ж оцінок частоти. Усереднюючи результати можемо отримати більш точну оцінку максимуму частоти. Зазначені процедури обробки модельного сигналу дозволили отримати оцінку приблизно 200 Гц, що ближче до першого максимуму.

Тепер зробимо обробку реального голосового сигналу, який аналізувався вище. Спектральний аналіз амплітудних і фазових спектрів реального голосового сигналу користувача дозволив отримати оцінку частоти основного тону в 243 Гц.

Результати обробки цього сигналу за допомогою пропонованих процедур оцінки частоти основного тону представлені на рис. 3.13. Коефіцієнт перекриття дорівнював 0.75. На рисунку 3.13 представлені оцінки математичного очікування (МО). При цьому суцільною лінією показані результати обробки амплітудних, а штриховий відповідно фазових даних.

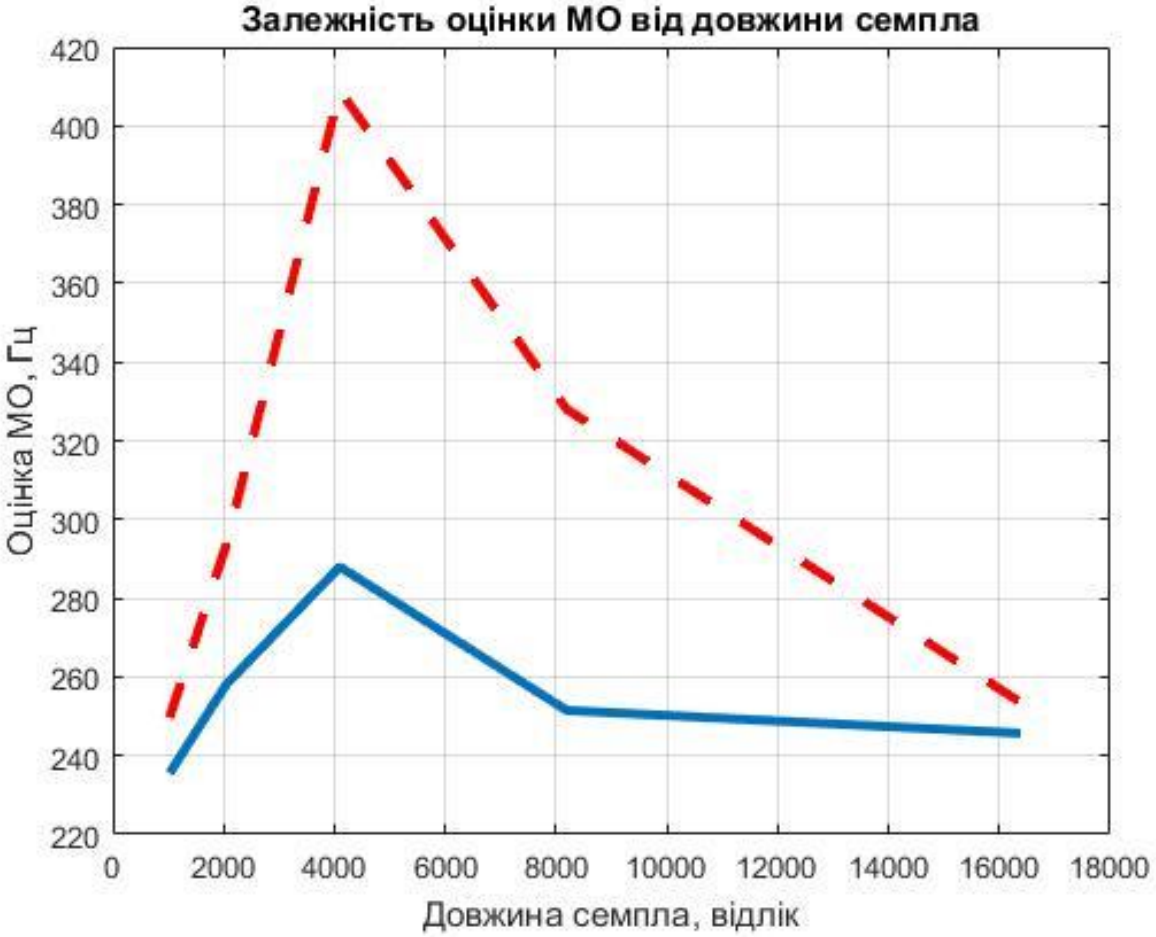

Рисунок 3.13 – Залежність оцінки МО від довжини семпла

На рисунку 3.14 представлені оцінки середнього квадратичного відхилення (СКВ) для амплітудного та фазового кепстрів.

Аналіз наведених рисунків дозволяє зробити висновок, що даний варіант оцінки частоти основного тону дозволяє вибрати параметри для отримання надійної оцінки. Досліджено вплив перекриття семплів в діапазоні від 0.5 до 0.85. Характер залежностей оцінок МО і СКВ відповідає результатам, представленим на рис. 3.13 і 3.14 відповідно. При цьому суцільною лінією показані результати обробки амплітудних даних, а штриховою –фазових.

Представлені результати можна істотно поліпшити. Для цього необхідно враховувати не всі оцінки ЧОТ, отримані в процесі розрахунку кепстральних коефіцієнтів.

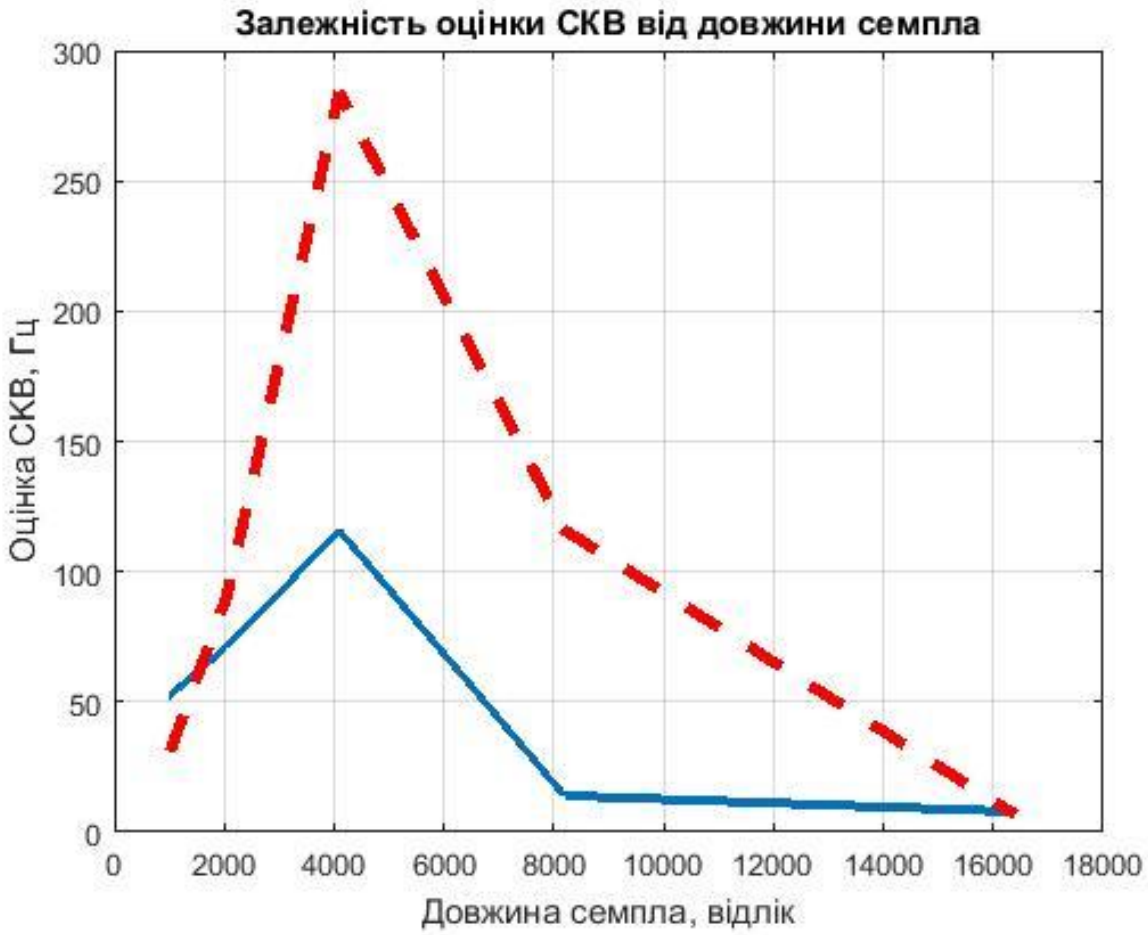

Рисунок 3.14 – Залежність оцінки СКВ від довжини семпла

Обґрунтуємо це. При аналізі голосового сигналу семпли будуть включати різні звуки – вокалізовані або невокалізовані. Крім цього, семпли можуть включати і шумові ділянки, наприклад, між складами. Відомо, що ЧОТ необхідно оцінювати на вокалізованих звуках [30].

Виділення вокалізованих звуків, в найпростішому випадку, можна виконати за допомогою порогової обробки.

ВИСНОВКИ

У магістерській роботі розглядається актуальне завдання підвищення якісних характеристик систем голосової автентифікації.

В якості основного напрямку вирішення зазначеної проблеми запропоновано в процесі цифрової обробки використовувати фазові дані аналізованого голосового сигналу.

Достовірність запропонованого варіанту вирішення зазначеної проблеми і аналіз інформативності фазових даних голосового сигналу досліджується в процесі експериментальної оцінки частоти основного тону і формантної інформації, які входять до складу більшості шаблонів користувача в якості обов'язкових параметрів.

Крім цього, до складу шаблону включені кепстральних або мел-частотні кепстральни коефіцієнти і ряд інших ознак. Частота основного тону дозволяє додатково вирішувати такі завдання: сегментація аудіо з декількома голосами і поділі мови на фрази, розпізнавання емоцій, визначення статі та ін..

У зв'язку з цим в роботі розглядалася актуальне наукове завдання розробки та дослідження нових процедур для уточнення оцінок частоти основного тону, отриманих на основі аналізу амплітудно-частотного спектру. Уточнення оцінок проводилося на основі використання фазових даних голосового сигналу, а також оцінки частоти основного тону в процесі отримання кепстральних коефіцієнтів.

Результати отримані в процесі статистичного аналізу результатів моделювання з використанням експериментальних голосових даних користувача системи автентифікації.

Фазові дані голосового сигналу дозволяють отримувати адекватні і достовірні оцінки в процесі спектрального аналізу. Однак, при наявності помилок, пов'язаних з грубими промахами, наприклад, прийняття за оцінку частоти основного тону максимумів частот першої або другої формант, перевагу слід віддавати оцінці, отриманої в процесі розрахунку кепстральних коефіцієнтів.

Розроблено та досліджено методика отримання адекватної оцінки частоти основного тону в процесі формування кепстральних коефіцієнтів. Достовірність методики підтверджена в процесі модельного експерименту.

Результати наукових досліджень оприлюднені в чотирьох наукових працях.

Подальші дослідження доцільно проводити в напрямку оцінки якості формування ознак для традиційно використовуваних шаблонів з урахуванням фази голосового сигналу, а також розробки нових процедур формування елементів шаблонів на основі фазових даних.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Куценко Є.Є. Використання фазової інформації мовного сигналу користувача в системах голосової автентифікації / Є.Є. Куценко, В.О. Тертичний, Р.А. Сердюк. // Одеса, ОНАЗ, Інфокомунікації – сучасність та майбутнє: тези доповідей 9-ій міжнародній науково-практичній конференції. – 2019. – С.198–200.

2. Куценко Є.Є. Оцінка частоти основного тону голосового сигналу користувача системи автентифікації. / Є.Є.Куценко, М.С. Пастушенко. // Проблеми телекомунікацій. – 2019. – № 2(25). – С. 97–103.

3. Куценко Є.Є. Оцінка частоти основного тону голосового сигналу користувача системи автентифікації / Є. Є. Куценко, М. С. Пастушенко. // Харків, НАНГУ, Міжнародна науково-практична конференція «Застосування інформаційних технологій у підготовці та діяльності сил охорони правопорядку». Збірник тез доповідей. – 2020. – С. 24.

4. Куценко Є.Є. Оцінка частоти основного тону голосового сигналу користувача системи автентифікації / Є. Є. Куценко, М. С. Пастушенко.. // Харків, ХНУРЕ, Матеріали 24-го Міжнародного молодіжного форуму «Радіоелектроніка і молодь в XXI столітті. Том 4.– 2020. – С. 88–89.

5. Прудник А.М. Биометрические методы защиты информации / А.М. Прудник, Г. А. Власова, Я. В. Рощупкин. – Минск: БГУИР, 2014. – 123 с.

6. Beigi H. Fundamentals of Speaker Recognition. / H. Beigi. – NY: Springer, 2011.–1029 p.

7. ISO/IEC 2382-37:2012 Information technology – Vocabulary – Part 37: Biometrics.

8. Болл Р.М. Руководство по биометрии / Р. М. Болл, Дж. Х. Коннел, Ш. Панканти, Н. К. Ратха, Э. У. Сеньор. – М. Техносфера, 2007. – 368 с.

9. Oppenheim A.V., Lim J.S. The Importance of Phase in Signals: Article *in* Proceeding of the IEEE, t. 69(5), 1981. – P. 529–541.

10. Paliwal K. Usefulness of phase in speech processing. Proc. IPSJ Spoken Language Processing Workshop, Gifu, Japan. – 2003. – P. 1–6.

11. Paliwal K., Atal B. Frequency-related representation of speech / In Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH-2003), 2003. – P. 65–68.

12. Борисенко С.Ю. Сравнение некоторых способов анализа фазовых соотношений между квазигармоническими составляющими речевых сигналов / С.Ю. Борисенко, В.И. Воробьев, А.Г. Давыдов // Сборник трудов 1-ой Всероссийской акустической конференции. – 2004. – С. 2–7.

13. Wu Z. A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case. Kinnunen T., Chng E., Li H., Ambikairajah E // Proc. Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC). – 2012.

14. Broeders, Ton, Forensic Speech and Audio Analysis Forensic Linguistics 1998-2001. Proceedings 13th INTERPOL Forensic Science Symposium, Lyon, France, D2, 54-84, 16-19 October 2001. [Електронний ресурс]. –2001. – Режим доступу до ресурсу: https://ssrn.com/abstract=2870568

15. Fergani B. Speaker diarization using one-class support vector machines / B. Fergani, M. Davy, A. Houacine // Speech Communication. – 2008 Vol. 50. – P. 355–365.

16. Kuwabara H. Acoustic characteristics of speaker individuality: Control and Conversion / H. Kuwabara, Y. Sagisaka // Speech Communication. –1995. - Vol. 16. – P. 165–173.

17. SorokinV.N. Speaker verification using the spectral and time parameters of voice signal / V.N. Sorokin, A.I. Tsyplikhin // Journal of Communications Technology and Electronics. –2010. - Vol.55, No. 12. – P. 1561–1574.

18. Matsumoto H. Multidimensional representation of personal quality of vowels and its acoustical correlates / H. Matsumoto, S. Hiki, T. Sone, T. Nimura // IEEE Trans. AU. –1973. - Vol. AU- 21. – P. 428–436.

19. Shriberg E. Modeling prosodic feature sequences for speaker recognition / E. Shriberg, L. Ferrer, S. Kajarekar, A. Venkataraman, A. Stolcke // Speech Communication. –2005. - Vol.46, No.3–4. – P. 455–472.

20. Lavner Y. The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels / Y. Lavner, I. Gath, J. Rosenhouse // Speech Communication. – 2000. - Vol.30. – P. 9–26.

21. Takemoto H. Acoustic roles of the laryngeal cavity in vocal tract resonance / H. Takemoto, S. Adachi, T. Kitamura, P. Mokhtari, K. Honda // J. Acoust. Soc. Am. – 2006. – Vol. 120. – P. 2228–2239.

22. Davis S. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences / S. Davis, P. Mermelstein // IEEE Trans.

Acoustics, Speech, Signal Process. – 1980. – Vol.28, No. 4 – P.357–366.

23. Itoh K. Perceptual analysis of speaker identity / K. Itoh // In: Saito, S. (Ed.), Speech Science and Technology. IOS Press. – 1992. – P.133–145.

24. Huang X. Spoken Language Processing: a Guide to Theory, Algorithm, and System Development / X. Huang, A. Acero, H.-W. Hon // Prentice-Hall, New Jersey. – 2001. – 935 p.

25. Lu X. An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification / X. Lu, J.Dang // Speech Communication. – 2007. - Vol. 50, No. 4. – P. 312–322.

26. Reynolds D. Speaker identification and verification using Gaussian mixture speaker models / D. Reynolds // Speech Communication. – 1995. - Vol. 17. – P. 91–108.

27. Reynolds D. Speaker verification using adapted gaussian mixture models / D. Reynolds, T. Quatieri, R. Dunn // Digital Signal Process. – 2000. - Vol. 10, No.1. – P. 19–41.

28. Vapnik V.N. Statistical Learning Theory / V.N. Vapnik // New York: Wiley. – 1998. –740 p.

29. BenZeghiba M. On the combination of speech and speaker recognition / M. BenZeghiba, H. Bourlard // In: Proc. Eighth European Conf. on Speech Communication and Technology (Eurospeech). –2003. – P. 1361–1364.

30. Bimbot F. An overview of the CAVE project research activities in speaker verification / F. Bimbot, M. Blomberg, L. Boves, D. Genoud, H.-P. Hutter, C. Jaboulet, J. Koolwaaij, J. Lindberg, J.-B. Pierrot // Speech Communication, v. 31. – 2000. –P. 155–180.

31. Pastushenko M. "Specifics of Receiving and Processing Phase Information in Voice Authentication Systems" / M. Pastushenko, V. Pastushenko, O. Pastushenko // International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kyiv, Ukraine. – 2019 – P.621–624.

32. Gabor D. Theory of communications / D. Gabor // The Journal of the Institute of Electrical Engineers, Part III (Radio and Communication Engineering) – 1946. - Vol. 93, No. 26. – P. 429 – 457.

33. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу / Г.С. Рамишвили. – М.: Радио и связь, 1981. – 224 с.

34. Фант Г. Акустическая теория речеобразования / Г.Фант. – М.: Наука, 1964. – 284 с.

35. Бендат Д. Прикладной анализ случайных данных / Д. Бендат А. Пирсол. – М.: Мир, 1989. – 540 с.