Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

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

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Дослідження використання коефіцієнтів лінійного передбачення за фазовою інформацією голосового сигналу в процедурах автентифікації
(тема)

Виконав:
студентка 2 курсу, групи ІКІм-21-1
Кіщенко М.І.
(прізвище, ініціали)

Спеціальність: 172 Телекомунікації та радіотехніка
(код і повна назва спеціальності)

Тип програми: освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма: Інфокомунікаційна інженерія
(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)

2022р.

Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій
(повна назва)

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

Рівень вищої освіти другий (магістерський)

Спеціальність 172 Телекомунікації та радіотехніка)
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Інфокомунікаційна інженерія
(повна назва)

ЗАТВЕРДЖУЮ

Зав.кафедри_____
(підпис)

«____» _____2022р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентці Кіщенко Маргариті Іванівні
(прізвище, ім'я, по батькові)

1. Тема роботи: Дослідження використання коефіцієнтів лінійного передбачення за фазовою інформацією голосового сигналу в процедурах автентифікації
затверджена наказом по університету від «24» жовтня 2021р. №1389_Ст.
2. Термін подання студентом роботи до екзаменаційної комісії 12.12.2022р.
3. Вихідні дані до роботи: вимоги стандарту ISO/IEC TR 24741:2007 Information technology – Biometrics tutorial; ISO/IEC/TR 24722:2007 Information technologies. Biometrics. Multimodal and other multibiometric fusion. Методика впровадження систем менеджменту інформаційної безпеки компанії IT-Grundshutz, аналітичні дані щодо аналізу впровадження та експлуатації систем менеджменту інформаційної безпеки створених на основі стандарту ISO/IEC 27001:2013, міжнародний стандарт ISO/IEC 27001:2013
4. Перелік питань, що потрібно опрацювати в роботі:
1) Аналіз наукових робіт за темою досліджень
2) Визначення фазових даних голосового сигналу та їх роль у цифровій обробці голосової інформації
3) Розробка імітаційної моделі для експериментального дослідження коефіцієнтів лінійного передбачення за фазовою інформацією
4) Розробка пропозицій щодо використання коефіцієнтів лінійного передбачення, які розраховані за фазовою інформацією

5) Перелік графічного матеріалу із зазначенням креслень, плакатів, комп'ютерних ілюстрацій: Демонстраційний матеріал у вигляді ppt-презентації; процедури формування фазових даних; схема експериментальної установки та імітаційної моделі; результати експериментальних досліджень

6. Консультанти розділів роботи

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

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	01.09.2022	Виконано
2	Збір матеріалів для дослідження	18.09.2022	Виконано
3	Розробка 1 розділу	19.10.2022	Виконано
4	Розробка 2 розділу	12.11.2022	Виконано
5	Розробка 3 розділу	30.11.2022	Виконано
6	Оформлення кваліфікаційної роботи	12.12.2022	Виконано

Дата видачі завдання 01 вересня 2022 року

Студентка _____ Кіщенко М.І.
(підпис) (прізвище, ініціали)

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

Робота не містить відомостей заборонених до відкритого опублікування

Студентка Маргарита КІЩЕНКО

Керівник Микола ПАСТУШЕНКО

РЕФЕРАТ

Пояснювальна записка: 76 с., 35 рис., 2 табл., 1 додаток, 38 джерел.

АВТЕНТИФІКАЦІЇ, АНАЛІЗ, БІОМЕТРІЯ, ГОЛОС, ПАРОЛЬ, МОДЕЛЮВАННЯ, КОЕФІЦІЄНТ ЛІНІЙНОГО ПЕРЕДБАЧЕННЯ, ФАЗА СИГНАЛА.

Об'єктом дослідження є процес голосової автентифікації користувача в інфокомунікаційних системах.

Мета роботи – розробка та дослідження процедур оцінки коефіцієнтів лінійного передбачення за рахунок фазової інформації для підвищення ефективності систем голосової автентифікації.

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

Одним із аспектів поліпшення якісних характеристик систем автентифікації користувачів є використання інформації про фазу голосового сигналу.

Актуальним науковим завданням є вивчення нових процедур підвищення надійності систем голосової автентифікації, дослідження коефіцієнтів лінійного передбачення за фазовою інформацією та порівняльний аналіз з даними отриманими за амплітудно-частотною інформацією.

За результатами досліджень запропоновано новий напрямок в покращенні автентифікації голосу, вдосконалення систем розпізнавання мовлення, а також при вирішенні проблем розпізнавання мовця.

ABSTRACT

The report contains: 76 p., 35 pictures, 2 tab., 1 application, 38 sources.

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

The object of research is the process of voice authentication of users in infocommunication systems and local networks.

The purpose of the work is development and research of procedures for estimating linear prediction coefficients due to phase information to improve the effectiveness of voice authentication systems.

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

One of the aspects of improving the quality characteristics of user authentication systems is the use of information about the phase of the voice signal.

An actual scientific task is the study of new procedures for increasing the reliability of voice authentication systems, the study of linear prediction coefficients based on phase information and comparative analysis with data obtained based on amplitude-frequency information.

The research results offer a new direction in improving voice authentication, improving speech recognition systems, and also in solving speaker recognition problems.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ І ТЕРМІНІВ

АЧХ – аплітудно-частотна характеристика
БД – база даних
ЕОТ – електронна обчислювальна техніка
КЛП – коефіцієнт лінійного передбачення
НМ – нейронна мережа
НКК – нормований коефіцієнт кореляції
СГА – система голосової аутентифікації
СКМ – система комп'ютерної математики
ССШ – співвідношення сигнал/шум
ЦОС – цифрова обробка сигналів
ШПФ – швидке перетворення Фур’є
AI – analog input
AR – autoregressive
ARMA – autoregressive moving average
CBEFF – common biometric exchange file format
EMD – empirical mode decomposition
ERR – equal error rate
FAR – false acceptance rate
FRR – false rejection rate
GMM – gaussian mixture model
HMM – hidden markov models
ID – identity document
LP – linear predictio
LTI – learning tools interoperability
MFA – multi-factor authentication
MFCC – mel frequency cepstral coefficient
NIST – national institute of standards and technology
SIIP – speaker identification integrated project
SVM – support vector machine

ВСТУП

У наш час важко уявити людство без інформаційних технологій. В такі важкі часи інтернет як ніколи потрібен. З початком коронавірусу, люди все більше почали звертатись до інформаційного простору, працюючи з дому і зберігаючи інформацію в комп’ютері. Бурхливий розвиток і широке поширення сучасних інформаційно-телекомунікаційних систем ознаменували перехід людства від індустріального суспільства до інформаційного суспільства, надійність якого базується на новітніх системах зв’язку, надійність яких не завжди відповідає зростаючим вимогам. У зв'язку з широким розповсюдженням розподілених систем у всіх сферах людської діяльності гостро постає завдання забезпечення інформаційної безпеки в таких системах [1]. Одним із основних заходів захисту інформаційних даних і обчислювальних ресурсів є забезпечення надійної аутентифікації користувачів.

В даний час існує багато підходів до автентифікації і ще більше застосувань цих підходів. Однак не всі класичні рішення проблеми автентифікації підходять для реалізації в розподілених системах. Доступ до інформаційних, фінансових та обчислювальних ресурсів здійснюється за допомогою парольної системи автентифікації користувача, яке не вважається надійним способом збереженням інформації.

Крім того, різні типи систем мають свої унікальні вимоги до підсистем автентифікації. Постійно розробляються нові алгоритми, а існуючі вдосконалюються, щоб забезпечити безпечну автентифікацію користувачів. Все більшої актуальності набуває проблема перевірки особистості користувачів, які мають доступ до публічних і персональних джерел інформації. Це питання особливо актуальне для відкритих масових телекомунікаційних та інформаційних систем.

Біометричні методи ідентифікації користувачів є одним із найперспективніших способів захисту таких систем від несанкціонованого доступу, цей спосіб прийшов з криміналістики. Для того щоб отримати біометричні дані потрібні біометричні сканери. Ці скановані дані потім мають бути порівняні з збереженими ключами, що зберігаються в база даних.

Перевагою цього методу автентифікації є те що всі необхідні дані вам вже даровані природою, форма і розмір обличчя, візерунок райдужної оболонки, сітківка очей – усі ці біометричні дані є унікальними.

Загалом існують такі види біометричної автентифікації:

- сканери відбитків пальців;
- сканери сітківки;
- сканери райдужки;
- розпізнавання мовця;
- розпізнавання обличчя;
- геометрія рук та пальців;
- геометрія жилки;
- на основі ДНК;
- поведінкові(біометрія ходи, розпізнавання набору тексту.

Біометричні системи використовуються зараз скрізь, наприклад, ваш телефон, скоріше за все, зараз можна розблокувати завдяки вашому обличчю або відбитку пальця, який ви раніше залишили у базі вашого телефону.

Розпізнавання по обличчю стало щось повсякденним для жителів Китаю, коли зараз після нововведених обмежень, після коронавірусу, почали блокувати виїзд із провінцій. Багато камер стоять в аеропортах усіх країн, так інтерпол та поліція боряться з злочинцями.

Але швидко цей шлях виявився неефективним, так як є велика вірогідність підробки даних та цей спосіб має обмежений ресурс для аналізу.

Саме через це фокус змістився на інші способи автентифікації на основі біометричних ознак. які складно підробити чи вкрасти.

Наприклад, розпізнавання набору тексту визначає людину на основі його унікальних характеристик набору тексту, по швидкості за якою людина друкує, тривалість часу, який необхідно для переходу від однією літери до іншої або рівень удару по клавіатурі.

Розпізнавання ходи – це тип поведінкової біометричної автентифікації, який розпізнає та перевіряє людей за стилем і темпом ходьби. Порівняно з іншими біометричними методами першого покоління, такими як розпізнавання відбитків пальців і райдужної оболонки ока, хода має перевагу ненав’язливості, оскільки не потребує контакту з об’єктом.

Системи голосової автентифікації також відносяться до тих що складно підробити, саме їх буде розглянуто в цій кваліфікаційній роботі.

Використання таких систем має перевагу над іншими способами:

- простота отримання біометричних даних, збір здійснюється за допомогою звичайного телефону або підключеним мікрофоном до комп'ютера.
- тому що системи голосової автентифікації не вимагають використання спеціального дорогого устаткування.
- управління доступом в голосовій автентифікації відбувається за допомогою телефонної мережі, Інтернету та корпоративної мережі – віддалено.
- можливість змінювати та нарощувати контрольну фразу.
- під вхідними даними людина представляє біометричний зразок, голосовий відбиток, який потім зберігається в базі даних.

Процес автентифікації в біометричних системах складається з певної послідновності дій, які контролюються на кожному з етапів. Першим етапом є введення біометричних даних, наприклад відбитку пальцю, обличчя, голос. Потім представлений матеріал формують в шаблон і перевіряють його наявність в своїй базі, щоб провести верифікацію, наприклад за голосом, необхідно у своїй базі вже мати шаблон голосу (зліпок голосу), цей зразок вже раніше був залишений ще на процесі реєстрації в системі.

В цілому в біометричних системах потрібно використовувати алгоритми прийняття рішень. Алгоритми існують для того щоб провести аналіз голосового відбитку і на основі аналізу дати результат автентифікації.

Існує багато методів процедур прийняття рішень, але зазвичай найпопулярнішими Gaussian Mixture Model (GMM) та Support Vector Machine (SVM) або також можна звернутись до використання прихованих марковських моделей (HMM) та .нейронних мережей,

Використання нейронних мереж в алгоритмі прийняття рішень базується на використанні нейромережевого класифікатора, який вирішує пройшов користувач автентифікацію чи ні.

При формуванні шаблону накладаються певні параметри, за якими пізніше буде складено відбиток голосу, такі як основний тон, форми імпульсів, тощо.

Шаблони можна порівнювати загалом або використовуючи особливості мовленнєвого сигналу, наприклад такі як: амплітуда та потужність, частота, енергетичні та фазові характеристики.

Лінійне передбачення (Linear Predictio, LP) є одним із найважливіших інструментів аналізу мовлення. Цей метод спочатку використовувався для аналізу сонячних плям, потім застосовували в нейрофізиці, сейсмології та голосовому з’язку.

Сутність методу в тому, що зразок мовлення може бути апроксимований як лінійна комбінація минулих зразків. Потім мінімізуючи суму квадратів відмінності між фактичними зразками мовлення та лінійно передбаченими скінченного інтервалу можна визначити унікальний набір коефіцієнтів предиктора. Аналіз LP розкладає мовлення на два дуже незалежні компоненти параметри голосового тракту (коефіцієнти LP) та голосового збудження (залишок LP).

Один з основних напрямів підвищення якісних характеристик систем автентифікації користувачів є використання фазової інформації голосу сигнал.

Фазовий простір голосового сигналу безпосередньо пов'язаний з аналітичним сигналом, який ефективно використовується в радіолокації та радіозв'язку.

Також використання фазових даних підходить для оцінки більшості властивостей шаблонів в голосовій автентифікації.

У даний час досліджені задачі формування за фазовими даними:

- формантних даних;
- кепстральних коефіцієнтів;
- мел-частотних кепстральних коефіцієнтів та інш.

Багато праць існує з використання фазових даних але використання саме коефіцієнтів лінійного передбачення(КЛП), що також можна розрахувати за фазовими даними, ще не так широко розповсюджене. Тому саме для цього пишеться ця магістерська робота, в якій проведемо це дослідження.

Для того щоб провести експеримент була використана імітаційна модель, де були виконані наступні дії:

1) Введення аудіо сигналу.
2) Вибір голосового сигналу користувача.
3) Формування аналітичного сигналу.
4) Формування фазових даних.
5) Розрахунок КЛП по амплітудно-частотним характеристика сигналу.
6) Розрахунок КЛП по фазовим характеристикам сигналу.

7) Порівнювальний аналіз.

В результаті отримаємо відповідь на нашу гіпотезу в використанні коефіцієнта лінійного передбачення по фазовій інформації.

По результатам досліджень було оприлюднено три наукові праці.

1 АНАЛІЗ ПОТОЧНОГО СТАНУ СУЧАСНИХ СИСТЕМ ГОЛОСОВОЇ АВТЕНТИФІКАЦІЇ

1.1 Актуальність задачі підвищення якості сучасних систем голосової автентифікації

В даний час особливої актуальності набувають питання кібербезпеки, через те що ставались випадки розкрадання, шифрування та зовсім знищення інформації. Для забезпечення безпечної і надійної процедури входу застосовуються додаткові заходи захисту: введення багатофакторної автентифікації, застосування ідентифікаторів пов'язаних з реальним світом і за допомогою біометричних даних людини.

Якщо говорити про парольний спосіб збереження інформації, то найбезпечніше користуватись багатофакторною автентифікацією (Multi-Factor Authentication, MFA).

Багатофакторна автентифікація – це метод автентифікації, який вимагає від користувача надавати два або більше факторів перевірки, щоб отримати доступ до ресурсу, наприклад програми, криптокошельку. Замість простого запиту імені користувача та пароля багатофакторна автентифікація вимагає один або кілька додаткових факторів перевірки, що зменшує ймовірність успішної кібератаки, через примусове використання двофактора, наприклад відбитка пальця або фізичного ключа, підвищує впевненість в надійності захисту.

В тому що потрібно декілька етапів підтвердження особи й полягає перевагою цього методу. Так як зменшується ризик втрати цінної інформації.

Гарною поєднанням в двофакторній автентифікації є біометрія:

- щось з біометричних даних
- пароль, апаратний маркер

В такому поєднанні цей вид автентифікації становить потужну комбінацію, завдяки такому поєднанню ваші дані становляться менш вразливі.

Але в свою чергу мінусом є те що потрібні додаткові програмно-апаратні комплекси, пристрої зчитування даних.

Зазначимо, що біометрія – це методи автоматичної ідентифікації та аутентифікації людини, на підставі його фізіологічних або поведінкових характеристиках. Для автентифікації найчастіше за все використовують відбитки пальців, фотографії райдужної оболонки ока та обличчя. Вибір цих матеріалів легко пояснити тим, що біометричні данні не будуть завдаватись сильним змінам і їх легко отримати. Але те що інші способи більш складні при автентифікації не означає що вони погано передають особливості людини. Навпаки, такі як голос, хода і почерк набагато складніше підробити зловмиснику, через те що це не статичний пароль, а дія, що є перевагою над іншими методами Саме тому одним із ефективних засобів захисту від цього є системи автентифікації, а саме система голосової автентифікації (СГА)

Метод розпізнавання особистості по голосу існує ще з часів коли людина почала говорити, ми можемо відрізнити своїх знайомих від сторонніх людей за допомогою голосу, так як у кожного свій тембр голосу. Метод голосової автентифікації має переваги перед іншими біометричними методами. Тому що більшість сучасних комп'ютерів мають уже вбудовану голосову систему і це вже зберігає кошти на дорогому обладнанні, також це ідеальний спосіб для телекомунікаційних додатків.

Технологія ідентифікації мовця може використовуватись не тільки в телекомунікаціях. Наприклад завдяки голосу можна розпізнати підозрюваних і ізолювати цікаві розмови правоохоронних органів , таких як договірні матчі, викуп або виклики терористів.

Так з 2014 по 2018 роки тривав інтегрований проект ідентифікації мовця (Speaker Identification Integrated Project, SIIP), проект сприяв ідентифікації злочинців і терористів завдяки розробці механізму ідентифікації мовців, який здатний працювати з різними джерелами, включаючи соціальні мережі або стаціонарні та мобільні телефони під час законного перехоплення. Окрім голосу система розпізнає стать, вік, мову та акцент, а також можна виявити клонування голосу.

Використання технології розпізнавання людського голосу базується на аналізі таких характеристик голосу, як тембр, спектр сигналу, акцент, інтонація, гучність, швидкість мови. Залежно від того, потрібна ідентифікація (розпізнавання) чи автентифікація (підтвердження) особи, використовуються різні методи ідентифікації. Біометрична технологія розбиває кожне вимовлене

слово на декілька сегментів і зберігається як математичний код. Системи можуть відрізнити голос людини з будь-якої фрази що була сказана.

З самого початку створення такого виду систем постало питання взлому, імітації голосу диктора. Методом експериментів з професійними імітаторами показало низький успіх таких підробок. Імітатори краще за все підроблювали інтонаційний контур і підкреслювали особливості мови, але більш глибокі фактори, які визначають індивідуальні якості голосу не удається скопіювати.

Останній час розроблюються методи трансформації одного голосу в інший, і це стає небезпекою для систем розпізнавання диктора. Формування голосу самозванця за допомогою перетворення до параметрів користувача збільшує помилку верифікації до 50%. Через це потрібно передивитись методи аналізу мови і використовувати параметри які складно повторити при трансформації голосу.

Підміна голосу виконується за допомогою попередньо записаним голосом або викраденим в каналі зв’язку голосових сигналів. Такий спосіб особливо небезпечний для систем верифікації з фіксований паролем. Можливо перевірити систему на проникнення за допомогою порівнювання двох однакових слів або фраз в системах з фіксованим паролем або там де вимова диктора контролюється самою системою. Якщо повністю ідентифікуються вимова, то це свідчить про проникнення в систему.

Саме через це потрібно вдосконалити СГА, щоб цінна інформація не витікала в маси. В поточній роботі спробуємо запропонувати спосіб завдяки якому голосові системи стануть більш захищеними.

1.2 Поточний стан систем і наукових робіт в області голосової автентифікації

Як уже зазначалося, до недавнього часу біометричні методи ідентифікації використовувалися лише в криміналістиці та суміжних професіях. Але ніщо не стоїть на місці, і біометрія людини, включаючи СГА, вже використовується для забезпечення інформаційної безпеки. Основними проблемами в сфері обробки мовленнєвих сигналів розділяють на такі завдання: розпізнавання особистості (самого диктора); розбірливість і розпізнавання мови.

Завдання розпізнавання мовлення має довгу історію, яка особливо розглядалася на початку 20 століття. Ще в 1929 році Д. Коллард представив формантний підхід, який використовувався до 1971 року, поки не був представлений підхід модуляції, який, крім шуму, також дозволяє враховувати реверберацийні перешкоди, нелінійні перетворення і навіть відлуння. Тут хотілося б виділити роботи [2, 3]. У першому джерелі можна прочитати чудову характеристику ролі та місця формантного підходу у вирішенні проблем розбірливості мовлення. Представником другої є класичною, в якій наводяться експериментальні методи визначення розбірливості шляхом артикуляційних вимірювань і чисельні методи обчислення розбірливості переданого мовлення.

Детальний аналіз робіт у галузі розпізнавання мовлення та розрізнення диктора наведено в роботі [4], виконаній Г. Рамішвілі. Охоплено період роботи з 1950 по 1980 роки. Автор одразу звертає увагу на те, що після введення мовного сигналу користувача необхідний етап попередньої обробки реєстраційних матриць, який передбачає боротьбу з внутрішніми та зовнішніми шумами та перешкодами. Але навіть зараз, через 50 років, у нас немає жодної роботи, спрямованої на збільшення відношення сигнал/шум (ССШ). Тут слід зазначити, що автор (див. [4], с. 67) спирається на теорему Котельникова для вибору частоти дискретизації.

У різних людей швидкість мовлення різна, і це може бути використовувати для розрізнення мовців. Лавер [5] стверджує, що середній темп мови для англійської мови становить близько 5,3 складів на секунду для мовця із середньою артикуляцією швидкісті.

У роботі [6] вперше підкреслено важливість високочастотного діапазону спектра мовного сигналу.

У статті [7] показано ефективність використання фазових даних для оцінки більшості властивостей патернів у СГА, а саме основної частоти, формантних частот, кепстральних і низькочастотних кепстральних коефіцієнтів.

Як відомо, джерелом мовного сигналу є [8] мовний тракт, який збуджує звукові хвилі в пружному повітряному середовищі. У просторі генерується і передається мовний сигнал у вигляді звукових хвиль. Приймачем сигналу є звуковий вібраційний датчик. Для цього прийнято використовувати мікрофон – пристрій що перетворює звукові коливання в електричні. Існує багато видів мікрофонів (електродинамічні, електростатичні, п'єзоелектричні та інші). Але в

мікрофонах будь-якого типу чутливим елементом є еластична мембрана, яка передає процес вібрації під впливом звукових хвиль. Мембрана з'єднана з елементом, який перетворює коливання мембрани в електричний сигнал.

Використання параметричної моделі мовного тракту вперше було запроваджено Г. Фантом [9], а також запропоновано ряд інших підходів до вирішення проблеми розпізнавання мови та розрізнення мовця.

У роботах [10, 11] відображено сучасний стан систем автентифікації голосу, в яких, нажаль, не відображені завдання боротьби з сигналами перешкод та проблема підвищення ССШ реєстраційних матеріалів.

В залежності від завдання використовують верифікацію мовця або ідентифікацію. При верифікації користувач надає свій ID (Ідентифікатор, Identity Document), котрий потрібно або підтвердити або відхилити. А в іншому випадку потрібно ідентифікувати диктора з усіх дикторів.

Механізм слухового сприйняття людини розрізняє різні звукові сигнали на основі трьох основних властивостей: висоти, гучності та тембру. У цьому розділі будемо визначать ці властивості та спробуємо дати їм кількісний опис. Це дуже складне завдання через опору на сприйняття.

Як буде зрозуміліше, це нелінійні зв’язки, які значною мірою залежать від наших суб’єктивних спостережень. Фактично, Стівенс [12] показав, що висота високих тонів зростає з інтенсивністю, висота низьких тонів зменшується з інтенсивністю, і точка, де відбувається зворотний рух, ручок залежить від рівня інтенсивності. Незважаючи на труднощі у відокремленні частки сприйняття, пов’язані з частотою та інтенсивністю, оскільки нас це стосується про систему автоматичного розпізнавання мовців, гарну кількісну модель для них вимірювання має вирішальне значення.

У більшості робіт параметри використовуються у вигляді кепстральних коефіцієнтів розпізнавання диктора, які розраховуються за огинаючою спектра, отриманої перетворенням Фур’є за допомогою фільтрів гребінки, або за передавальною функцією мовного тракту, яка, в свою чергу, була розрахована згідно з лінійним прогнозом, що було доведено методом [6]. Крім кепстральних коефіцієнтів використовуються також їх перша і друга різниці за часом. Це відбувається тому, що такий кепстр відображає індивідуальні особливості джерела голосу та анатомію мовного тракту.

Питання структури сучасних систем автентифікації голосу розглянуто в роботах [12, 13, 14].

Історично питання автентифікації диктора виникло в інтересах сфер правосуддя, криміналістики та контррозвідки. Тому спочатку проблему вважали схожою на проблему розпізнавання відбитків пальців, показану у відомій роботі [8], який ввів та популяризував термін «голосові відбитки».

Цей напрям досліджував контурні лінії однакових рівнів енергії на сонограмах, які корелюють з папілярними візерунками пальців. Швидко стало зрозуміло, що форми контурів піддаються багатьом варіаціям і не можуть використовуватися як ознаки для ідентифікації мовця.

Випробування на фіксованій базі даних Національного інституту стандартів і технологій США (National Institute of Standards and Technology, NIST) демонструють поступове підвищення продуктивності систем розпізнавання мовлення. Однак практичне застосування цих систем невелике, оскільки їх характеристики ще далекі від вимог користувачів. Тому необхідно регулярно вивчати ситуацію в цій сфері з метою визначення найбільш перспективних напрямків.

Головним недоліком усіх біометричних методів, крім лінгвістичних, є стабільність використовуваного біометричного коду, оскільки для людини незмінні відбитки пальців або долонь, малюнки райдужної оболонки ока та риси обличчя. Цей недолік перешкоджає застосуванню цих методів у випадках, коли потрібна особливо висока надійність ідентифікації особистості, оскільки незмінний біометричний код може бути обчислений зловмисним втручанням програми розпізнавання.

На відміну від біометрії з фіксованими параметрами, голосова перевірка має майже необмежений потенціал для зменшення помилок завдяки використанню дедалі довших голосових повідомлень. Голосову перевірку можна використовувати віддалено в темряві, особливо на звичайному телефонному каналі, коли неможливо отримати зображення обличчя.

Приклади конкретних застосувань верифікації диктора охоплюють широкий спектр додатків:

- ведення банківського рахунку, підтвердження права користування кредитною карткою);
- дозвіл на зміну пароля або PIN-коду;

- доступ до комп'ютера або окремих комп'ютерних програм (вихід в Інтернет, доступ до конфіденційних документів, баз даних і т.д.);
- дозвіл на те щоб потрапити в якесь приміщення, до якого мають доступ окремі обличчя.
- керування якимось механізмом або системами (наприклад, керування розумним дімом);
- контроль, хто мав доступ до чого, коли та ресурсів комп'ютера.

Додавання розпізнавання мовлення мовця багаторазово зменшує помилкове розпізнавання обличчя/фігури, але додавання візуальної інформації лише трохи покращує акустичне рішення [10].

У певних ситуаціях, таких як отримання команд, необхідно переконатися, що наказ віддає уповноважена особа. І в цій ситуації голосова автентифікація не є точною, оскільки одержувач цієї інформації може не впізнати голос, який видав цю команду. І саме в таких випадках є сенс автоматично ідентифікувати групу осіб, які мають право віддавати накази.

У сучасних системах автентифікації голосу використовується інформація про амплітуду і частоту записаного сигналу, а фазові дані, що важливіше, традиційно ігноруються.

Водночас не всі сучасні можливості цифрової обробки даних вичерпано. Тому далі ми проаналізуємо поточний стан систем голосової автентифікації та розробимо практичні рекомендації щодо покращення їх якісних характеристик.

У наступній роботі [15] описано, що як основні параметри, що складають шаблон користувача, прийнято використовувати наступні характеристики:

- амплітудно-частотна характеристика (АЧХ);
- основний тон;
- формантна інформація;
- відстань між обертонами;
- форми імпульсів збудження тощо.

Для компенсації часової нестабільності оголошення диктором парольних фраз існує два способи:

- шляхом стискання та розтягування областей адаптувати до стандарту за допомогою динамічного програмування;
- вибравши центр тонального діапазону та виявивши вимірювання поблизу центральної частини фонеми, тоді проблема тривалості більше не буде

суттєвою.

Одним із підходів до розпізнавання голосу є багатоканальний аналіз із застосуванням статичних методів і закінчуючи прийняттям рішення нейронною мережею та/або складною системою штучного інтелекту.

Підхід ідентифікації ознак можна було реалізувати багато років тому за допомогою аналогової фільтрації. З розвитком комп'ютерних технологій акцент зміщується на системи з лінійним предиктором мовного сигналу.

Лінійне передбачення [12, 7, 10] – це техніка аналізу часових рядів, яка виникає в результаті вивчення лінійних систем. Використовуючи лінійне передбачення, параметри такої системи можна визначити шляхом аналізу входів і виходів системи. Makhoul [7] каже, що цей метод вперше з’явився в статті 1927 року про аналіз сонячних плям, але з тих пір він застосовувався до проблем нейрофізики, сейсмології та голосового зв’язку.

Матриця Тепліца що проявляється в багатьох рішеннях різницевих рівнянь, особливо в системах керування і обробки сигналів. Структура матрици робить її досить простим для вирішення за допомогою використання алгоритму Левінсона-Дурбіна, який був представлений Левінсоном [16], а потім модифікований Durbin [17, 18, 19, 20]. Існує ще один алгоритм під назвою Schur рекурсія [20], яка більш ефективна для паралельних реалізацій.

Також є метод обчислення загальнополюсної оцінки який полягає у використанні дискретного косинусного перетворення за допомогою швидкого перетворення Фур’є (ШПФ), наприклад [21], який використовує для повної обробки спектральний шлях рішення лінійних коефіцієнтів прогнозування.

Функції емпіричного розкладу моди (Empirical Mode Decomposition, EMD) були представлені в [22] у 1998 році і з тих пір вони використовуються для розпізнавання мовлення.

В [23] представлена система EMD де завдяки використанню смуг проводиться фільтрація, на основі нейронної мереж виконання ідентифікації мовця.

Як описано в основу кодування мови цим методом покладена хвильова структура мовленнєвого сигналу. Метод лінійного передбачення побудований на апроксимації сусідніх хвиль в звуковій пачці перехідним процесом лінійного цифрового фільтру, який більш докладно буде розглянуто у другому розділі.

1.3 Загальна схема роботи систем голосової автентифікації

До основних переваг СГА відносяться:

- дешевизна, не потрібні спеціальні прилади, вистачить мікрофона на телефоні або підключеного мікрофону до ноутбука;
- часові ряди цифрових даних можливо швидко обробити;
- простота використання;

Недоліками в голосових системах автентифікації можна вважати:

- через стороні звуки в шаблоні мають місце хибні результати;
- вплив психо-емоційно-фізичного стану користувача на результат автентифікації(якщо людина, наприклад, захворіє і втратить голос, або охрипне, то система не пропустить далі, так як буде вважати що це інша людина);
- якщо текстозалежний вид автентифікації то парольну фразу важко зберегти в таємниці.

Голосова біометрія з технологією ідентифікації та верифікації мови це не одне й теж саме з технологією розпізнавання мови. Технологія розпізнавання відрізняє те що людина сказала, але не може визначити хто саме це каже. Тому цю технологію не безпечно використовувати для системи захисту. Для того щоб за голосом визначити хто є представлена людина використовують технології ідентифікації та верифікації особи.

Ідентифікація – перевіряє співпадіння одного зразку голосу з багатьма із бази голосів. У результаті отримаємо список осіб з схожими голосами в відсотковому співвідношенні. 100% співпадіння означає що зразок зійшовся повністю з голосом з бази даних.

Верифікація в свою чергу це процес порівняння двох зразків голосу: голосу людини чию особистість необхідно підтвердити з голосом що зберігається в базі даних і чия особистість уже повністю підтверджена. У результаті також отримаємо ступінь співпадіння у відсотковому співвідношенні.

Декілька факторів від яких залежить результат біометричного виміру:

- вхідних даних;
- математичних алгоритмів;
- обчислювальної потужності.

Вхідними даними являється біометричний зразок або голосовий відбиток, який зберігається в базі даних. В залежності від навколишнього середовища та

тип у обладнання(пристрою введення інформації) залежить якість біометричного зразку. Але якщо навіть у вас немає якісного типу обладнання, наприклад є тільки мікрофон телефону, то існують технології, які автоматично очищають зразок від шуму та в результаті отримати якісний голосовий відбиток.

Функціонування біометричної системи складається з наступних дій (рис. 1.1):

1) Запис – біометричні дані, які були записані пристроєм введення інформації.

2) Екстракція – з представлених біометричних даних витягується унікальна інформація і вигляді коду і це й становить образ.

3) Порівняння – проводиться порівняння представленого біометричного образу з одним або більшим числом еталонів, які зберігаються в базі даних системи.

4) Прийняття рішення – система вирішує, збігаються чи ні біометричні образи, і виносить судження про закінчення процедури ідентифікації.

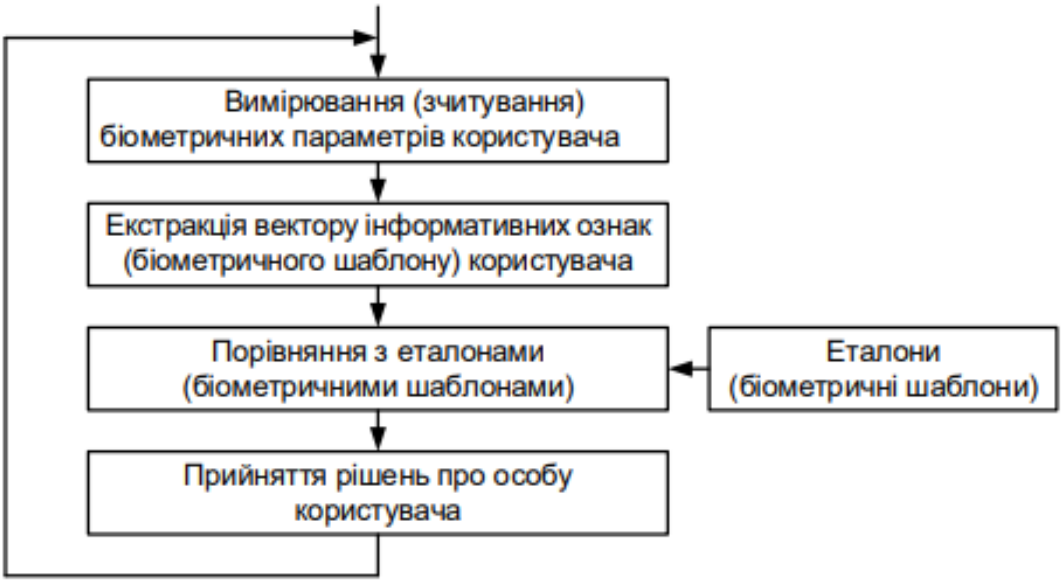

Рисунок 1.1 – Процедура біометричної ідентифікації

На етапі екстракції і порівняння біометричних даних використовуються біометричні шаблони. Шаблон це набор що містить закодовану інформацію про біометричні образи. Існує Міжнародний стандарт сервіс CBEFF (Common Biometric Exchange File Format), що встановлює єдиний формат подання біометричних даних.

Складності в біометричних системах можуть виникнути через такі причини:

- ймовірність «не визнати свого» і «визнати свого чужим», через те що біометрична верифікація має імовірнісний характер;
- можливості підміни біометричних даних.

Під обчислювальною потужністю розуміють швидкість та якість обробки біометричних ознак користувача.

На рис.1.2 наведено приклад застосування системи розпізнавання особистості за голосом в одному з Call-центрів.

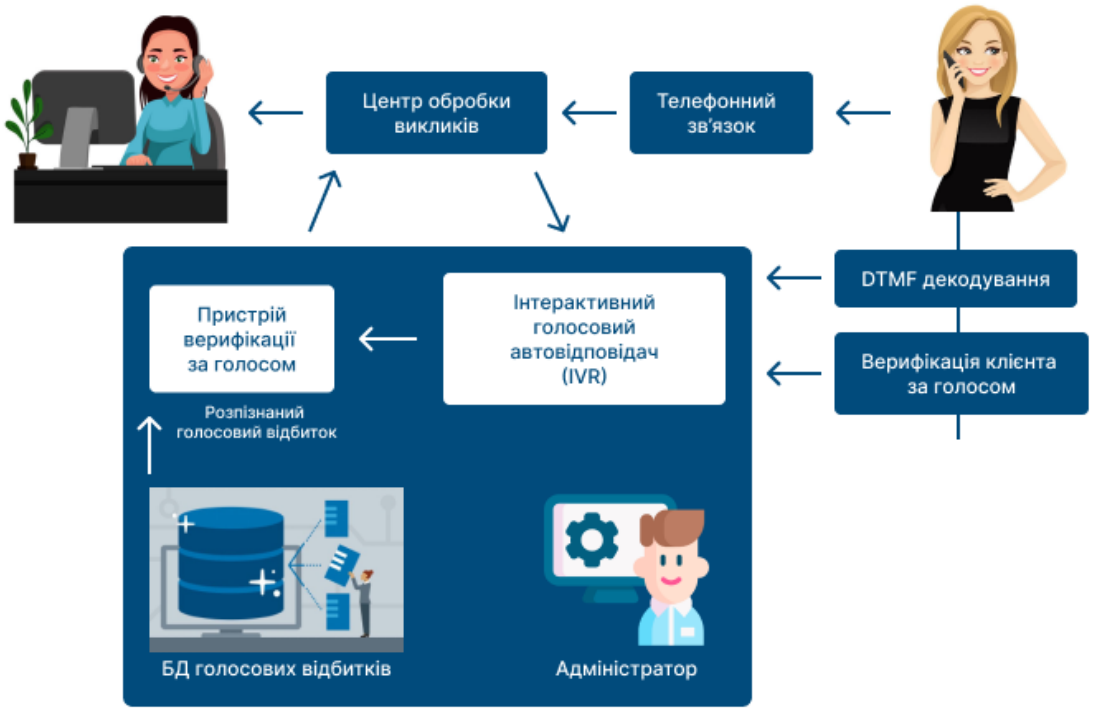

Рисунок 1.2 – Робота системи розпізнавання особистості

Залежно від типу мовленнєвих висловлювань завдання розпізнавання мовця поділяють на текстозалежні та текстонезалежні методи. У більшості відомих систем простір ознак, у якому виконується розпізнавання диктора, є рівномірним і не залежить від контексту та мови. Аналіз таких ознак наведено нижче.

У криміналістиці мовленнєві дані є довільними, тому дослідження в інтересах таких програм зосереджені на контекстно-незалежних методах розпізнавання. Делегування такого підходу на завдання авторизації доступу не є

дуже перспективним. У кожному акті впізнання вважається зручним говорити ті фрази, які забажає диктор. Насправді це вимагає від мовця свідомого створення щоразу нового тексту, що створює певне когнітивне навантаження.

Тому насправді користувачі в таких системах зазвичай говорять одну і ту ж фразу. Це перетворює систему розпізнавання в систему з фіксованим паролем, формально незалежною від контексту та найменш стійкою до втручання з боку шахраїв шляхом відтворення почутого або написаного пароля.

Недоліком систем фіксованих паролів є те, що пароль довільний для кожного користувача диктора, а база даних диктора створюється за допомогою іншого набору виразів. Така різниця в лінгвістичному матеріалі погіршує ефективність системи розпізнавання, і створити спеціальну довідкову базу для конкретного пароля практично неможливо.

Оптимально використовувати словник, в якому буде не багато слів, що буде легко запам’ятати, це можуть бути, наприклад, числівники від 0 до 8. Пароль в якому буде послідовність з таких слів повинен випадково змінюватись при кожному акті розпізнавання. За допомогою цього уникнемо небезпеки вторгнення за допомогою записаного коду, властива системам з фіксованим паролем.

Системи розпізнавання поділяються на системи індивідуального та колективного користування. При вході до операційної системи або будь-яким даними в персональному комп'ютері. При віддаленому доступі, наприклад, за телефонним каналом або за допомогою Інтернету, розпізнавання може здійснюватися на сервері з множинним доступом.

Найважливішим для успішного розпізнавання мовця є вибір інформативних ознак (параметрів мовлення), здатних ефективно відобразити інформацію про конкретні особливості мовлення.

До них застосовуються такі умови:

- ефективність подання інформації про специфічні характеристики мови оратора;
- простота вимірювання;
- стабільність у часі;
- повторюваність і природний характер мови;
- несприйнятливість до імітації.

На етапі вилучення ознак мовленнєвий сигнал сегментується на короткі ділянки і на кожній ділянці обчислюється набір ознак.

Проблема сегментації полягає в тому, що потрібно відокремити мовлення від інших навколишніх шумів, музичних сегментів, мови поверх музики тощо. Також варто зазначити, що більшість пристроїв розпізнавання виходять з ладу, якщо їх попередньо надіслано з музикою замість мови, тому важливо розділити музику з розпізнаваним мовленням це також допоможе заощадити місце в архіві.

Для характеристик розпізнавання мовлення в методах верифікації голосу використовуються різні параметри, які враховують як коефіцієнти мовлення (характеристики розподілу частот основного звуку, лінійна проекція, спектр Фур'є), так і сприйняття мовлення вейвлет-спектр, кепстральні частоти, мелчастотний кепстральний коефіцієнт (Mel-frequency cepstral coefficient, MFCC), та їх динамічні властивості [11].

Використання алгоритмів прийняття рішень в біометричних системах впроваджене для того щоб провести порівняння введеного образку голосу з тим що вже є в базі даних. Точність результату порівняння залежить від досконалості алгоритму.

Сама теорія прийняття рішень уявляє собою сукупність моделей та математичних методів, метою яких є обґрунтування рішень, які були прийняті на основі аналізу.

Етапи процесу прийняття рішень:

- виявлення проблеми та постановка задачі прийняття рішення
- формулювання поняття якості рішення
- прогнозування можливих результатів та альтернативних варіантів
- оцінка якості варіантів рішень, вибір оптимального
- аналіз та впровадження рішення

Першим етапом буде виявлення проблеми, зрозуміння що існує певна проблема. Потім потрібно встановити масштаб цієї проблеми і природу.

Коли проблему було визначено то потрібно провести далі дослідження що стало причиною цієї проблеми.

Далі потрібно поставити ціль для вирішення цієї проблеми, якими способами буде досягнене рішення.

Розробляється альтернативне рішення для того щоб знайти найкраще можливе рішення. Це робиться для того щоб відкинути варіант вибору першого

рішення, щоб краще подумати над усіма варіантами. Альтернативні рішення продумуються на етапі збору актуальної інформації та аналізу.

В результаті обирається альтернативний варіант і впроваджується в дію.

Далі проводиться спостереження за процесом реалізації. Та проводиться оцінка ефективності обраного варіанту, чи була вирішена проблема.

Одними з найпопулярніших процедур прийняття рішень являються методи Gaussian Mixture Model (GMM) дивитись на рис 1.3 та Support Vector Machine (SVM). Також підходить використання штучних нейронних мережей та прихованих марковських моделей (HMM).

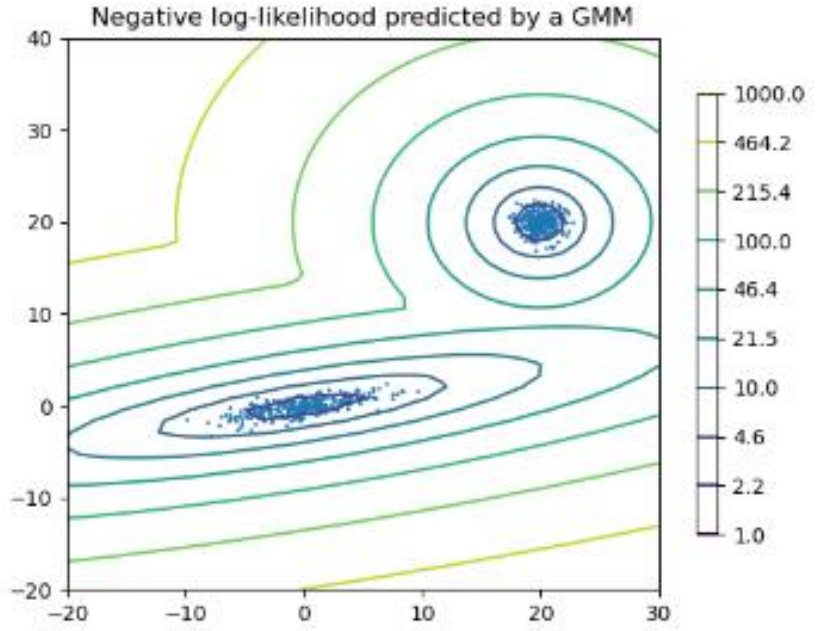

Рисунок 1.3 – Модель двокомпонентної суміші Гауса: точки даних та поверхні рівноймовірності моделі

Модель суміші Гауса – це імовірна модель, яка припускає, що всі точки даних генеруються із суміші кінцевого числа розподілів Гауса з невідомими параметрами.

Параметри моделі GMM можна оцінити за допомогою оцінка максимальної правдоподібності. Основна мета оцінки правдоподібності полягає в отриманні оптимальних параметрів моделі, які можуть максимізувати ймовірність GMM.

Однак значення ймовірності є дуже нелінійною функцією в моделі параметрів і пряма максимізація неможлива. Замість цього виконується максимізація через ітераційні процедури.

Опорні вектори – це точки даних, які розташовані ближче до гіперплощини та впливають на положення та орієнтацію гіперплощини (рис 1.4). Використовуючи ці опорні вектори ми максимізуємо запас класифікатора. Видалення опорних векторів змінить положення гіперплощини. Це точки, які допомагають нам побудувати SVM.

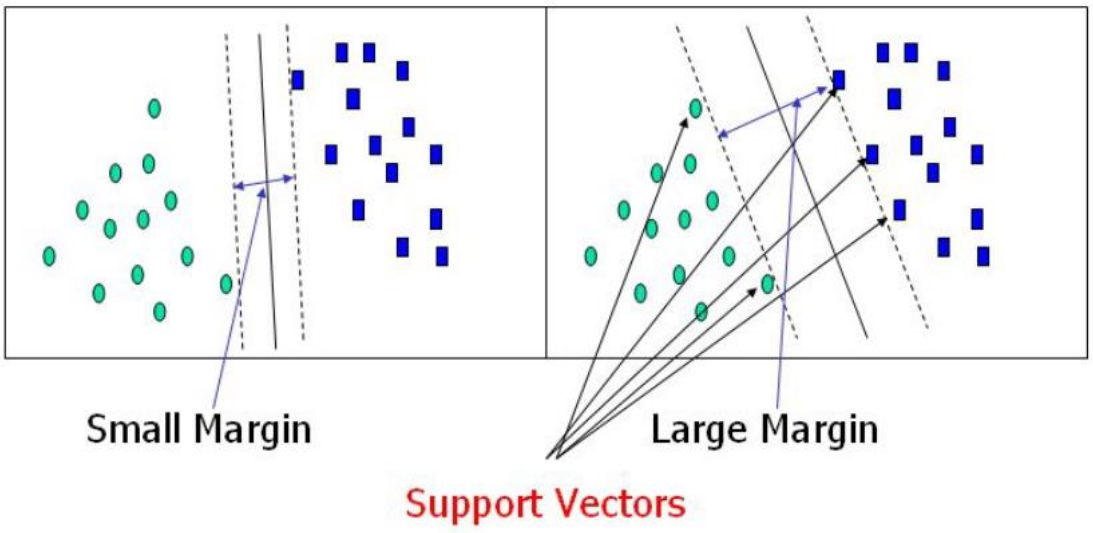

Рисунок 1.4 – Вектори опори

Прихована марківська модель (Hidden Markov Model, HMM) – це модель, в якій ви спостерігаєте послідовність викидів, але не знаєте послідовності станів, якими пройшла модель для генерації викидів (див. рис.1.5). Аналіз прихованих марківських моделей спрямований на відновлення послідовності станів із даних, що спостерігаються.

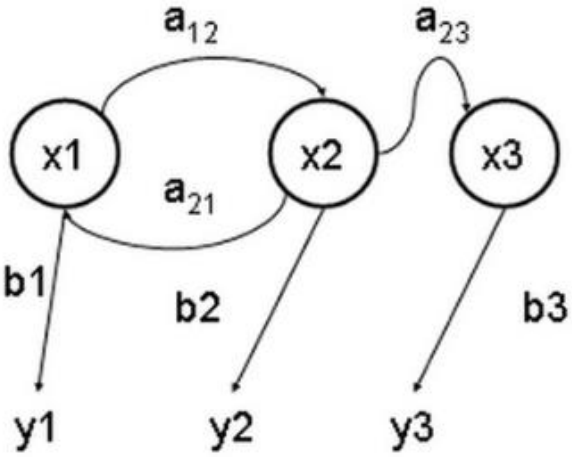

Рисунок 1.5 – Діаграма переходів у прихованій Марківській моделі

На рис. 1.5 наведені наступні позначення: x – приховані стани, y – результати, що спостерігаються, a – ймовірності переходів, b – ймовірність результату.

Для проведення верифікації за голосом, необхідно мати шаблон голосу(зліпок голосу) в БД системи, людини яка вже підтвердила особистість.

Для прикладу, найбільший український банк запровадив систему з голосовим доступом до особистих даних, за пів року кількість людей що скористувались голосовою автентифікацією виросло до мільйона людей.

Спочатку клієнт дзвонить у службу підтримки та пояснює свою причину звернення, в даний час з ним спілкується віртуальний асистент(робот), в процесі розмови відбуваються такі дії:

- робиться зліпок голосу та порівнюється з голосом який зберігається в БД цього банку.
- через декілька секунд система ідентифікує клієнта, що полегшує та пришвидшує процес взаємодії.

Через це завжди перший крок – це накопичення бази образами голосу. Для створення шаблонів спочатку клієнт проходить реєстрацію в системі. Він добровільно залишає відбитки свого голосу, який потім буде використовуватись кожен раз для верифікації. Прийнято брати поспіль три зліпки голосу, щоб була варіантність. Верифікація успішно пройдена, кожен вхід в систему оновлюються старі образи голосу, так система запобігає проблеми старіння голосу. У випадку якщо застосовується верифікація за динамічною парольною фразою, то можуть

продиктувати цифри від 0 до 9 тричі, після чого у системи буде 30 відбитків, завдяки цьому буде менше вірогідність помилки.

Потрібно враховувати те що, клієнт має залишити свій відбиток голосу, зареєструватись, по тому самому каналу де потім буде верифікуватися, бо так буде менше ймовірність помилок. Є випадки коли клієнти проходять реєстрацію по домашньому телефону, а потім проходять верифікацію з скайпу, звичайно тут фактор каналу зв'язку буде грати велику роль як і в надійності так і в роботі системи без помилок. Тому при побудові системи треба враховувати, що у людей постійно можуть змінюватись канали зв'язку, тому це можна протестувати окремо під кожен випадок і невілювати вплив іншого каналу майже повністю. Але якщо не зробити це на етапі розробки, то вже в працюючій системі буде складно це впровадити.

Важливо, що клієнт самостійно і усвідомлено пройшов реєстрацію (знав навіщо це і як це йому потім допоможе), через те що пройти потім верифікацію може лише лояльний абонент, якому потрібен результат і який приймає правила гри.

Якщо клієнта змушувати проходити верифікацію коли треба і коли лишній раз можна не просити, то він може підсвідомо змінювати голос, обманювати (не дружелюбний до сервісу) – це призводитиме до помилок і лояльність клієнта падатиме, хоча він сам у цьому опосередковано буде винен.

Статична парольна фраза – це набір слів що складає пароль що при автентифікації потрібно буде озвучити.

На рисунку 1.6 зображена схема реєстрації людини в біометричній системі (статична парольна фраза).

Рисунок 1.6 – Схема реєстрації людини у біометричній системі

1) Клієнт дзвонить у біометричну систему, що просить його придумати нескладну парольну фразу та вимовити її, повторюючи три рази.

2) Далі голос обробляється сервером біометрії і в кінці виходить три моделі голосу клієнта, для кожного повторення своя окрема модель.

3) На завершення заводиться картка на цього абонента (Михайло Михайлович), за якою закріплюються ці три моделі.

Автентифікація завдяки статичній парольній фразу не є безпечним способом захисту інформації. Через те що зараз дуже легко вкрасти, скопіювати кодові слова. Для того щоб зменшити вірогідність викрадення потрібно комбінувати з іншими видами біометричної ідентифікації.

Для того щоб ідентифікувати користувача за голосом в базі даних потрібно мати мовленнєвий шаблон, який дозволяє порівнювати введені данні з тими що вже змережені.

При виконанні процедури автентифікації в процесі розпізнавання мовлення клієнта, використовуючи мобільні пристрої, можуть бути помилки двух типів:

1) Хибні відхилення (False Rejection Rate, FRR).

2) Хибний допуск (False Acceptance Rate, FAR).

Помилка хибного відхилення (помилка 1-го роду), коли система верифікації відхиляє справжню ідентичність “свого”, характеризується вірогідністю хибної тривоги. А при допуску “чужого” має місце помилка хибного допуску (помилка 2-го роду), характеризующа вірогідністю пропуску “чужого”.

Тоді узагальненою характеристикою буде середня вірогідність похибка, визначається як напівсума вірогідностей похибок 1-го та 2-го родів. У результаті майже кожна система вміє перебудовуватися так, щоб похибка одного роду могла бути зменшена за рахунок збільшення похибок другого типу шляхом зміни порога прийняття рішення. Для оцінки якості систем аутентифікації можна використовувати критерій визначення рівного коефіцієнта помилок (Equal Error Rate, ERR), коли поріг прийняття рішення вибирається так, щоб забезпечити рівність обох похибок [27].

Ключ та шаблон можна порівнювати загалом або використовуючи такі особливості мовленнєвого сигналу як: амплітуди та потужності, частоти, енергетичні та фазові характеристики.

1.4 Нейромережеві алгоритми біометричної ідентифікації

Нейронні мережі (НМ) грають важливу роль при забезпеченні інформаційної безпеки через те що мають такі переваги:

- можливість відтворення з заданою точністю складних нелінійних залежностей;
- здатність до навчання і самонавчання, на навчальну вибірку не накладаються обмеження;
- здатність до узагальнення;
- архітектура НМ реалізується на паралельних обчислюваних засобах;
- потенційно висока завадо- і відмовостійкість.

Система що розпізнає образи, побудована на основі НМ, в загальному випадку складається з двох частин (рис. 1.7): підсистеми екстракції інформативних (інваріантних) ознак і нейромережевого класифікатора, що виконує функцію вирішального правила.

Рисунок 1.7 – Блок-схема розпізнавання образів на основі нейромережі

На першому етапі проводиться перетворення координат вхідного вектору (образа) x з метою вилучення з нього інформативних ознак, які зберігають найбільш важливу інформацію.

В результаті обертання вхідний образ точка x в m-мірному просторі даних переводиться в точку x^*в m^*-мірному просторі ознак.

Так як $m^* < m$, то дане перетворення доцільно назвати операцією стиснення даних, що полегшує завдання класифікації.

Класифікація являє собою перетворення, що відображає вектор x^*в один з класів m-мірного простору рішень. Вектор y уявляємо рівною m, припустимо j-я компонента вектору y приймає значення 1, це станеться тільки в випадку коли x^* належить відомому k-класу, усі інші складові вектору приймають нульові значення

Блок-схема нейромережевої системи біоідентифікації особи, наведена на рис. 1.8.

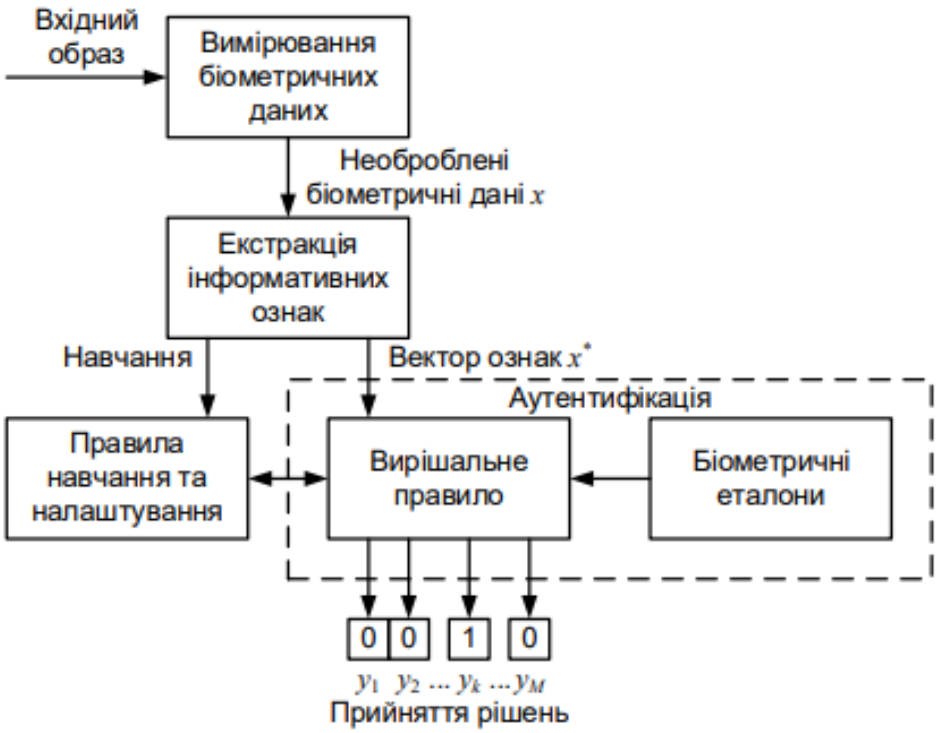

Рисунок 1.8 – Узагальнення блок-схема нейромережевої системи біометричної ідентифікації

На цій блок-схемі відображені основні етапи обробки інформації НМ:

- вимірювання біометрики користувача за допомогою сенсорів, що представляють собою вхідні перетворювачі;
- екстракція інформативних ознак;
- створення нейромережевого біометричного еталону користувача;
- реалізація останнього правила на основі НМ.

Перші дві частини обробки інформації працюють за однаковим алгоритмом, що не залежить від режиму роботи біометричної системи. В процесі режиму роботи системи визначається сукупність операцій, де використовуються вектори інформативних біометричних ознак:

$$x_t^* = (x_{1,t}^*, x_{2,t}^*, ..., x_{m^*,t}^*)^T, (t = 1, 2, ..., T) \quad (1.1)$$

У режимі навчання вектори ознак $x_t^*, (t = 1,2,\ldots,T)$ надходять в блок правил та налаштування, який формує біометричні еталони користувачів. Через те що біометричні образи одного користувача неідентичні один до одного, то для формування еталонів потрібно декілька прикладів таких образів для кожного користувача.

Далі в режимі автентифікації вектор ознак x^* порівнюється з біометричним еталоном, якщо виявляється що вони схожі то автентифікація пройдена. Але якщо відрізняються, то отримаємо відмову користувачеві в автентифікації, у деяких випадках можуть запропонувати повторну автентифікацію.

На рис. 1.9 наведено приклад найбільш поширеною схеми НМ – багатошарового персептрона.

Мережа складається з таких прошарок нейронів:

- вхідний прошарок, на який подаються надсилаються компоненти вектору ознак x^* і котрий потім далі відправляє на нейрони до іншого по черзі прошарку, при цьому не здійснюючи перетворень;
- прихований прошарок, який здійснює перетворення координат (x_1^* , $x_2^*, \ldots, x_{m^*}^*$) до виходів нейронів прихованого прошарку ($z_1, z_2, \ldots, z_n$);
- вихідний прошарок, на якому формується вектор вихідних реакцій ($y_1, y_2, \ldots, y_M$), який утворюється з M нейронів (за кількістю класів розпізнавання).

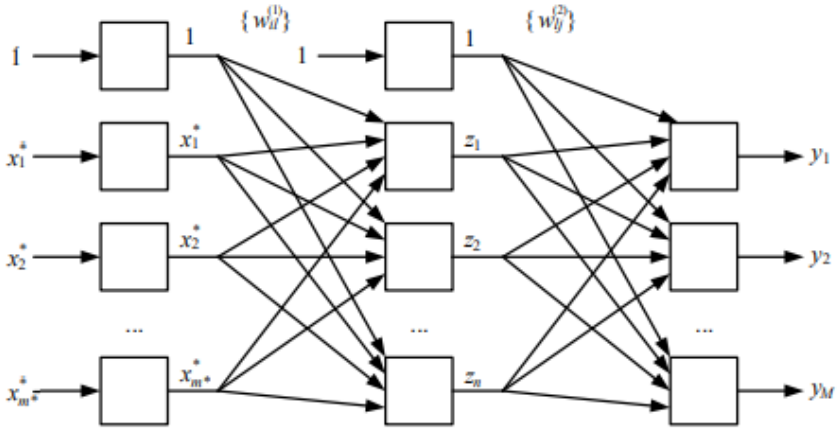

Рисунок 1.9 – Структурна схема тришарового персептрону

Вихідні значення нейронів прихованого та вихідного прошарків обчислюються за такими рівняннями:

$$z_l = f\left(\sum_{i=1}^{m^*} w_{il}^{(1)} x_i^* + w_{0l}^{(1)}\right), l = 1,2,...,n$$

$$y_j = f\left(\sum_{l=1}^{n} w_{lj}^{(2)} z_l + w_{0j}^{(2)}\right), j = 1,2,...,M\ , \qquad (1.2)$$

де $f(s)$ нелінійна передатна функція нейрона; $w_{il}^{(1)}$ і $w_{lj}^{(2)}$ це матриці ваг синаптичних зв'язків; $w_{0l}^{(1)}$та $w_{0j}^{(2)}$ це зміщення.

На рисунку 1.10 зображено передатну функцію де а) порогова (логічна), б) сігмоїдна, в) порогова, г) у вигляді гіперболічного тангенсу.

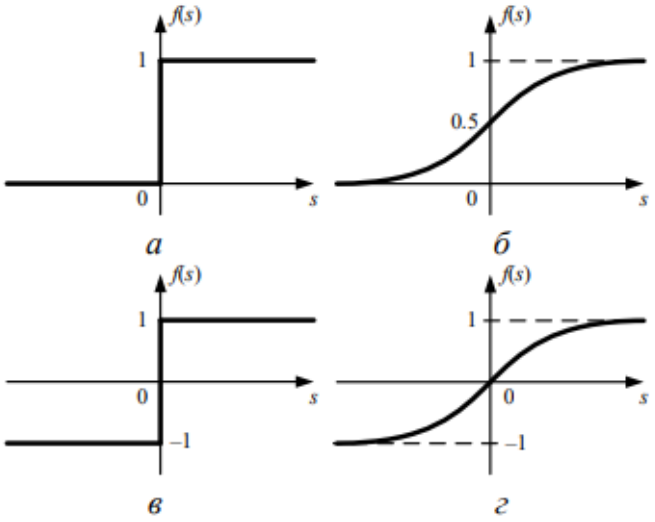

Рисунок 1.10 – Передатні функції: а) порогова (логічна), б) сігмоїдна, в) порогова, г) у вигляді гіперболічного тангенсу

Мета навчання НМ полягає в поданні на входи мережі вектору ознак w_t^*, що належить k-му класу де вихідний сигнал би вказував номер класу.

1.5 Постановка задач на проведення наукових досліджень

Спільним для розглянутих робіт і методів голосової автентифікації є те що в якості вхідної інформації використовуються амплітуда і частота зареєстрованого голосового сигналу.

Останнім часом з'явилися ряд робіт, які додатково використовують фазові дані (фазовий простір) голосового сигналу, наприклад [28-38]. Саме такий підхід був використаний у даній роботі, який, як буде показано нижче, дозволяє розширити можливості вирішення завдань формування признаків шаблонів за голосовим сигналом.

Зауважимо, що фазовий простір голосового сигналу безпосередньо пов'язаний з поняттям аналітичного сигналу, який ефективно і продуктивна використовується в радіолокації та радіозв'язку.

Що стосується області голосових сигналів сформувати фазові дані можна програмно за допомогою перетворення Гільберта.

Як показали результати досліджень фазові дані голосового сигналу мають форму пилкоподібного сигналу, амплітуда якого змінюється за лінійним законом від 0 до 360 градусів, а тривалість невідома. Тут суттєвим є те, що амплітуда змінюється за лінійним законом. За наявності неточностей в реєстрації або зовнішнього шуму будуть з'являтися випадкові помилки, які призводитимуть до відхилення змін амплітуди фазового сигналу від лінійного закону.

К даному часу досліджені задачі формування за фазовими даними:

- формантних даних;
- кепстральних коефіцієнтів;
- мел-частотних кепстральних коефіцієнтів та інш.

Як свідчать результати досліджень сформовані ознаки в багатьох випадках мають кращі характеристики, чим ти, які розраховані за амплітудно-частотною інформацією.

Однак до цього часу не досліджені коефіцієнти лінійного передбачення, які розраховані за фазовою інформацією та не виконаний порівняльний аналіз з даними отриманими за амплітудно-частотною інформацією. Це основне завдання даної магістерської роботи.

Для цього необхідно розробити математичну модель для цифрової обробки голосових сигналів, формування фазових даних аналізованого сигналу, а також

здійснити розрахунки й порівняльний аналіз коефіцієнтів лінійного передбачення в амплітудно-частотним та фазовим просторі.

2 КОЕФІЦІЄНТИ ЛІНІЙНОГО ПЕРЕДБАЧЕННЯ – ЗАГАЛЬНА ХАРАКТЕРИСТИКА ТА ПОРЯДОК РОЗРАХУНКУ

2.1 Теорія лінійних систем

Лінійне передбачення є одним з найефективніших методів для оцінки основних параметрів мовленнєвих сигналів. Важливість цього методу полягає у високій точності отриманих оцінок і відносній простоті розрахунків.

В [17] 10 коефіцієнтів лінійного передбачення перераховуються на 12 коефіцієнтів кепстру. В цьому випадку частотний масштаб розрахованого спектру є лінійним, що призводить до втрати порівняно з нелінійний масштаб. Недоліком цього методу є те, що він найкраще зарекомендував себе методом передбачення рядків, що представляють рух мовленнєвого тракту, як дробово-раціональна функція, що містить лише полюси.

Підхід в порівнюванні результатів прогнозу для різних методів який традиційно оцінює втрати прогнозу в порівнянні з деякою ідеальною моделлю. Спочатку оцінюються параметри статичної моделі на основі спостерігання, а потім проводиться прогноз на основі цієї моделі при оцінюваних параметрах.

Прогнозуючий алгоритм може спостерігати прогнози експертних стратегій і оцінювати їх ефективність в минулому, після якого дає свій прогноз. Все відбувається в порівнянні з прогнозами експертних алгоритмів.

Порівняння може використовувати алгоритм рандомізації де може оцінка проводиться, як в середньому, так і в найгіршому випадку.

В якості основі для визначення функції втрат можуть використовуватись багатокількісні методи оцінки якості класифікації або передбачення.

Лінійна система являється такою системою, що виробляє свій вихід як лінійну комбінацію своїх поточних та попередніх входів та його попередні виходів. Це можна писати як інваріантний у часі, якщо параметри системи не змінюються з часом.

Математично лінійні незмінні в часі системи можна представити наступним рівнянням:

$$y(n) = \sum_{j=0}^{q} b_j x(n-j) - \sum_{k=1}^{p} a_k y(n-k) \qquad (2.1)$$

Це загальне рівняння для будь-якої лінійної системи з вихідним сигналом y і вхідний сигнал x, а також скаляри b_j і a_k для $j = 1 \ldots q$ і $k = 1 \ldots p$ де максимум p і q є порядком системи.

Система може бути представлена графічно (див. рис. 2.1).

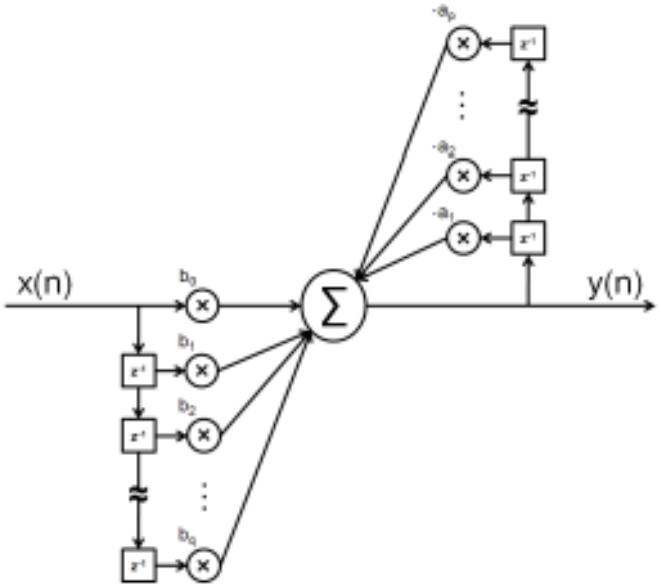

Рисунок 2.1 – Графічне представлення загального рівняння для лінійно-незмінної в часі системи

Переставляючи рівняння (2.1) і перетворюючи його в Z-область, ми можемо виявити передаточну функцію $H(z)$ такої системи:

$$y(n)+\sum_{k=1}^{p} a_k y(n-k)=\sum_{j=0}^{q} b_j x(n-j),$$

$$\sum_{k=0}^{p} a_k y(n-k)=\sum_{j=0}^{q} b_j x(n-j) \text{ , де } a_0 = 1$$

$$\sum_{k=0}^{p} a_k z^{-k} Y(z)=\sum_{j=0}^{q} b_j z^{-j} X(z) \text{ ,} \qquad (2.2)$$

$$\Rightarrow H(z)=\frac{Y(z)}{X(z)}=\frac{\sum_{j=0}^{q} b_j z^{-j} X(z)}{\sum_{k=0}^{p} a_k z^{-k}}$$

Коефіцієнти вибірок вхідного та вихідного сигналів у рівнянні (2.1) розкривають полюси і нулі передатної функції. Лінійне передбачення природно

випливає із загальної математики лінійного системи. Оскільки вихід системи визначається як лінійна комбінація минулих зразків, майбутній результат системи можна передбачити, якщо масштабні коефіцієнти b_j і a_k відомі.

Таким чином, ці скаляри також відомі як прогностичні коефіцієнти системи.

Загальна передавальна функція лінійної системи породжує три різні типи лінійної моделі, що залежить від виду передатної функції $H(z)$, наведеної в рівняння (2.2).

1) Коли чисельник передатної функції постійний, однополюсний або визначено модель авторегресії (AR).

2) Модель повного нуля або ковзного середнього передбачає, що знаменник передатної функції є константою.

3) Найбільш загальний випадок - це змішана модель полюса/нуля, яку також називають модель авторегресійного ковзного середнього (Autoregressive Moving Average, ARMA), де нічого не передбачається щодо функції передачі.

Багатополюсна модель для лінійного прогнозування є найбільш широко вивченою та реалізованою з трьох підходів з ряду причин.

По-перше, вхідний сигнал, який необхідний для моделювання ARMA та повністю нуля, часто є невідома послідовність. Таким чином, вони недоступні для використання в наших похідних.

По-друге, рівняння, отримані на основі підходу всеполюсної моделі відносно простий у розв’язанні, різко контрастуючи з нелінійними рівняннями, отриманими з ARMA або повністю нульового моделювання.

Нарешті, і, можливо, найбільше важлива причина, чому всеполюсне моделювання є кращим вибором інженерів, багато додатків реального світу, включаючи більшість типів виробництва мови, можуть бути точно змодельованим за допомогою означеного підходу.

2.2 Всеполюсна модель лінійного прогнозування

Виходячи з рівняння лінійної системи (2.1), можна сформулювати рівняння та необхідні для визначення параметрів всеполюсної лінійної системи, так звані нормальні рівняння лінійного прогнозу.

По-перше, продовжуючи всеполюсну модель (див. рисунок 2.2), оцінка лінійного прогнозу $\hat{y}$ при номері вибірки n для виходу сигнал y за допомогою p^{th} фільтр передбачення порядку можна задати наступним чином

$$\hat{y}(n) = -\sum_{k=1}^{p} a_k y(n-k) \ , \tag{2.3}$$

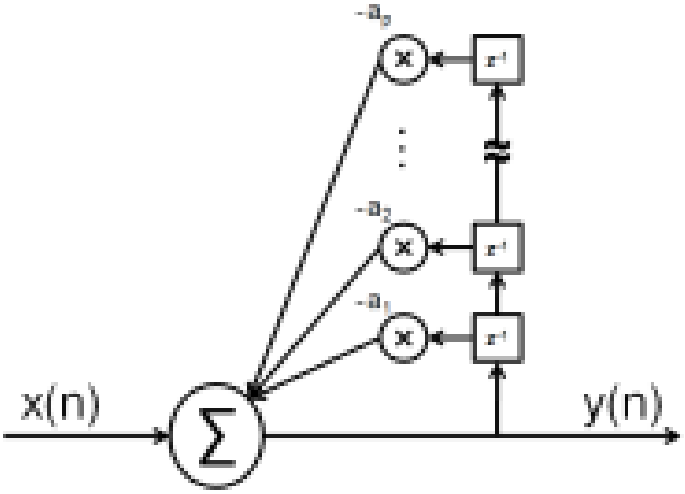

Рисунок 2.2 – Графічне представлення повнополюсної лінійної системи

На рис. 2.2 показано, що вихід є лінійною функцією масштабованих попередніх виходів і вхідних даних x.

Помилка або залишок між вихідним сигналом і його оцінкою у вибірці n , або як різницю між двома сигналами.

$$e(n) = y(n) - \hat{y}(n) \tag{2.4}$$

Загальна квадратична помилка для ще невизначеного діапазону вибірок сигналу визначається наступним рівнянням

$$E = \sum_n [e(n)]^2 = \sum_n [y(n) - \hat{y}(n)]^2 =$$

$$= \sum_n [y(n)]^2 - \sum_n [y(n) - \hat{y}(n)]^2 \qquad (2.5)$$

Рівняння (2.5) дає значення, що вказує на енергію в сигналі помилки. Очевидно, що коефіцієнти предикторів бажано підібрати так, щоб значення E мінімізується протягом невизначеного інтервалу.

Оптимальні мінімізуючі значення можна визначити за допомогою диференціального числення, тобто шляхом отримання похідної від рівняння (2.5) по відношенню до кожного коефіцієнта предиктора та встановлення рівного значення до нуля.

$$\frac{\partial E}{\partial a_k} = 0, \text{ де } 1 \le k \le p$$

$$\Rightarrow \frac{\partial}{\partial a_k}\left(\sum_n [y(n)]^2 - 2 \cdot y(n) \cdot \hat{y}(n) + [\hat{y}(n)]^2\right) = 0$$

$$-2\sum_n y(n) \cdot \frac{\partial}{\partial a_k}\hat{y}(n) + 2\sum_n \hat{y}(n) \cdot \frac{\partial}{\partial a_k}\hat{y}(n) = 0$$

$$\sum_n y(n) \cdot \frac{\partial}{\partial a_k}\hat{y}(n) = \sum_n \hat{y}(n) \cdot \frac{\partial}{\partial a_k}\hat{y}(n)$$

$$y(n) = -y(n-k) \text{ з рівняння (2.1)}$$

$$\Rightarrow \sum_n y(n) \cdot (-y(n-k)) = \sum_n \hat{y}(n) \cdot (-y(n-k))$$

$$\sum_n y(n) \cdot (-y(n-k)) = \sum_n \left(-\sum_{i=1}^{p} a_i y(n-i)\right) \cdot (-y(n-k))$$

$$-\sum_n y(n) \cdot y(n-k) = \sum_{i=1}^{p} a_i \sum_n y(n-i) \cdot (-y(n-k)) \qquad (2.6)$$

Для стислості та майбутньої корисності визначено кореляційну функцію φ. Розширення цього підсумовування описує те, що буде називатися кореляцією матриця

$$\varphi(i,k) = \sum_n y(n-i) \cdot y(n-k) \qquad (2.7)$$

Підстановка кореляційної функції в рівняння (2.6) дозволяє її записати більш компактно.

$$-\varphi(0,k)=\sum_{i=1}^{p}a_i\varphi(i,k) \tag{2.8}$$

Отриману систему рівнянь називають нормальними рівняннями лінійного передбачення.

2.3 Розв’язки нормальних рівнянь

Обмеження на підсумовування повної квадратичної енергії були опущені в рівнянні (2.5). Але їхньому вибору треба приділити особливу увагу. Тут буде показано, що два різні, але логічні інтервали підсумовування призводять до двох різних наборів нормальних рівнянь і призводять до різних коефіцієнтів предиктора.

З огляду на достатню кількість точок даних і відповідні межі, нормальні рівняння визначити p рівнянь з p невідомими, які можна розв’язати за допомогою будь-яких загальних алгоритмів одночасного розв’язування лінійних рівнянь, наприклад, гаусове усунення, краут декомпозиція тощо.

Однак певні обмеження призводять до надлишковості матриці та дозволяють ефективні рішення, які можуть значно зменшити обчислювальне навантаження.

Існує два підходи до оцінки цих значень: метод автокореляції та коваріаційний метод.

Автокореляційний метод лінійного передбачення мінімізує сигнал помилки весь час, від $-\infty$ до $+\infty$. Коли мова йде про кінцеві цифрові сигнали, сигнал формується у вікнах таким чином, що всі вибірки за межами цікавого інтервалу приймаються рівними 0 (див. рисунок 2.3).

Якщо сигнал відмінний від нуля від 0 до $N-1$, то результуюча помилка сигналу буде відмінна від нуля від 0 до $N-1+p$. Таким чином, підсумовуючи загальну енергію за цей інтервал математично еквівалентно підсумовуванню за весь час

$$E = \sum_{n=-\infty}^{\infty} [e(n)]^2 = \sum_{n=0}^{N-1+p} [e(n)]^2 \qquad (2.9)$$

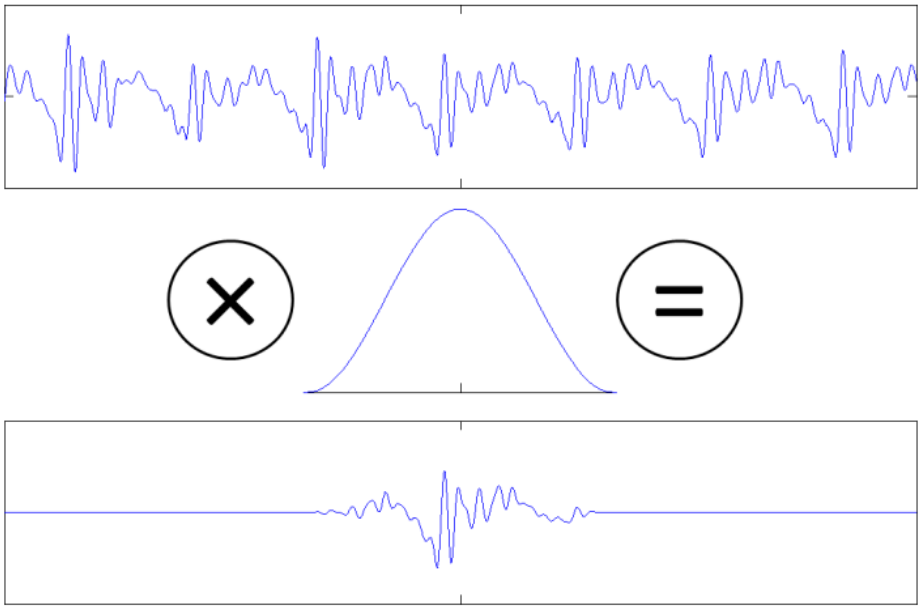

Рисунок 2.3 – Вікно сигналу шляхом множення з відповідною функцією, у цьому випадку вікно Хеммінга

Коли ці обмеження застосовуються до рівняння (2.7), виникає корисна властивість, оскільки сигнал помилки кореляції дорівнює нулю поза інтервалом аналізу. Тому функція нормального рівняння може бути тотожним чином виражена в більш зручній формі:

$$\varphi_{covar}(i,k) = \sum_{n=0}^{N-1+p} y(n-i) \bullet y(n-k) \; 1 \le i \le p \quad 1 \le k \le p$$

$$= \sum_{n=0}^{N-1+(i-k)} y(n) \bullet y(n+(i-k)) \; 1 \le i \le p \quad 1 \le k \le p \qquad (2.10)$$

Ця форма кореляційної функції є просто короткочасовою автокореляцією функцію сигналу, оцінена із затримкою ($i - k$) вибірок. Цей факт дає цей метод розв'язування нормальних рівнянь його назва.

Наслідки цієї зручності такі, що кореляційна матриця, визначена нормальними рівняннями, демонструє подвійну симетрію, яку можна використати у комп'ютерному алгоритмі.

Враховуючи, що $a_{i,j}$ є членом кореляційної матриці на ith рядок і j^{th} у стовпчику кореляційна матриця демонструє:

- стандартну симетрію, де $a_{i,j} = a_{j,i}$,

$$\begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} & \cdots & a_{m,1} \\ a_{2,1} & a_{2,2} & a_{3,2} & \cdots & a_{m,2} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{m,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{pmatrix}$$

- симетрію Тепліца, де $a_{i,j} = a_{i-1,j-1}$

$$\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,m} \\ a_{2,1} & a_{1,1} & a_{1,2} & \cdots & a_{m,2} \\ a_{3,1} & a_{2,1} & a_{1,1} & \cdots & a_{m,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{1,1} \end{pmatrix}.$$

Ці надмірності означають, що нормальні рівняння можна розв’язувати за допомогою методу Левінсона-Дурбіна. Це рекурсивна процедура, яка значно зменшує обчислювальне навантаження.

На відміну від автокореляційного методу, коваріаційний метод – лінійне передбачення мінімізує загальний квадрат енергії лише в інтервалі, що цікавить

$$E = \sum_{n=0}^{N-1} [e(n)]^2 \qquad (2.11)$$

Використовуючи ці межі, перевірка рівняння (2.7) показує, що значення сигналу, необхідні для розрахунку, виходять за межі інтервалу аналізу (див.

рис.2.14). Метод коваріації потребує вибірок − p (тут показано червоним кольором), а інтервал аналізу від 0 до N − 1 (показано синім кольором).

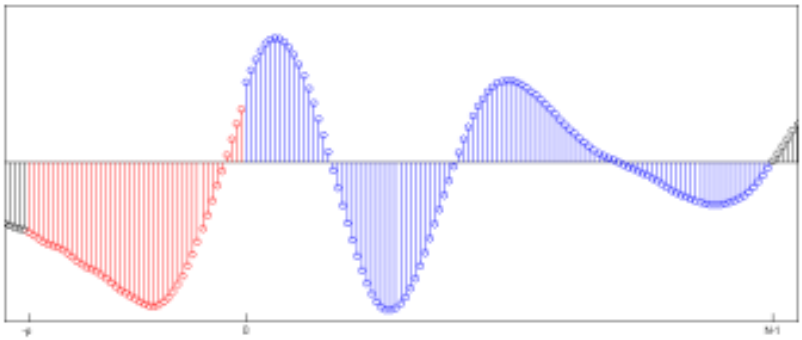

Рисунок 2.4 – Метод коваріації

А саме:

$$\varphi_{covar}(i,k) = \sum_{n=0}^{N-1} y(n-i) \bullet y(n-k) \; 1 \le i \le p \quad 1 \le k \le p \tag{2.12}$$

Таким чином, потрібні вибірки від $-p$ до $N - 1$. Отримана кореляційна матриця демонструє стандартну симетрію, на відміну від матриці, яка визначена автокореляцією.

Отримана матриця методом коваріації не демонструє симетрію Тепліца. Це означає, що для вирішення нормальних рівнянь необхідно використовувати інший метод, наприклад розкладання Холецького або метод квадратного кореня.

Кожне з цих розв'язків нормальних рівнянь лінійного прогнозу має свої власні сильні і слабкі сторони. При цьому, все значною мірою визначається сигналом, що аналізується. Коли сигнали аналізу довгі, два різних рішення практично ідентичні. Через більшу надмірність у матриці, визначеній методом автокореляції, це трохи легше обчислювати.

Експериментальні дані вказують на те, що коваріаційний метод є точніше для періодичних звуків мови, тоді як автокореляційний метод кращий для фрикативних звуків.

2.4 Представлення мовленнєвого тракту людини як лінійної системи

Спочатку проаналізуємо, як створюється людська мова. Як показано на рис. 2.5, переважна більшість звуків людської мови виробляється таким чином. Легені ініціюють процес мовлення, діючи як сильфон, який випускає повітря вгору в інші області системи. Повітряний тиск підтримується міжреберними та черевними м’язами, що дозволяє злагоджену роботу мовних механізмів.

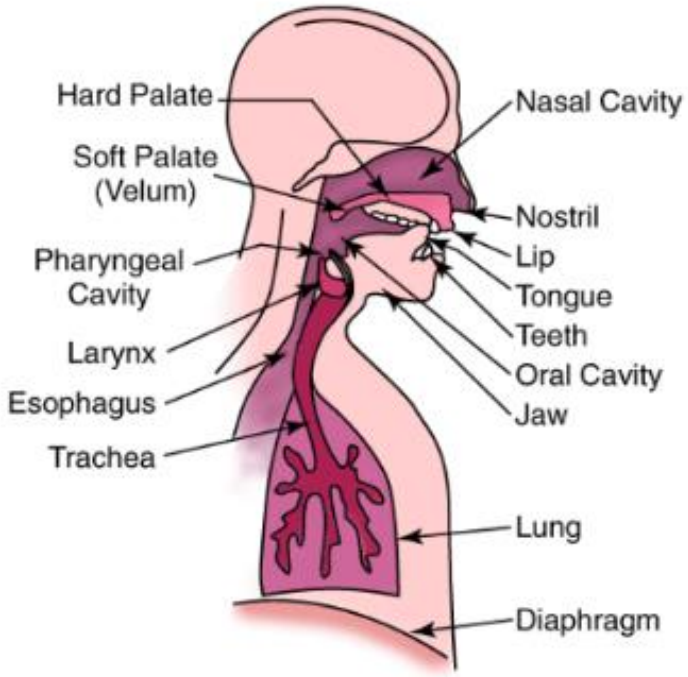

Рисунок 2.5 – Система виробництва людської мови

Повітря, що виходить з легень потім надходить до решти областей системи виробництва мови через трахеї. Це система органів, що складається з легенів, трахеї і сполучних між собою каналів, відомий як легеневий тракт.

Приводиться в рух турбулентний повітряний потік вгору по трахеї в гортань. Гортань – коробчатий апарат, що складається з м'язів і хрящів. Дві мембрани, відомі як голосові складки, охоплюють структура, що підтримується спереду щитовидним хрящем, а ззаду – черпалоподібні хрящі. Аритеноїди прикріплені до м’язів, які їх забезпечують для зближення і розділення голосових складок.

Дійсно, основна функція гортані, не пов'язана з процесом мовлення, полягає в ущільненні трахеї шляхом підтримки голосових складок при закриті. Це має подвійну перевагу – можливість захистити легеневого тракту та сприяти зростанню тиску в грудній порожнині необхідні при певних навантаженнях і кашлі

Простір між голосовими складками називається голосовою щілиною. Звук мовлення – це класифікується як дзвінкий або глухий залежно від поведінки гортані під час проходження повітря крізь це.

Відповідно до мієлопружинно-аеродинамічної теорії, вібрація голосових складок виникає в результаті взаємодії двох протилежних сил. Апроксимовані складки розсуваються через підвищення підгортанного тиску повітря, коли повітря проходить через голосову щілину. Це явище всмоктування, відомо як ефект Бернуллі.

Цей ефект пояснюється зниженням тиску на звужені апертури разом з цією складкою. Взаємодія між цими силами зашкоджує вібрації голосових складок, утворюючи голосовий звук.

Ця фонація має основну частоту, яка пов'язана з частотою вібрації складки. Під час глухого мовного звуку голосова щілина залишається відкритою і струмінь повітря безперешкодно проходить через гортань. Отримана голосова хвиля збудження демонструє плоский частотний спектр.

Фонація з гортані потім надходить у різні голосові камери шляхи: глотка, носова порожнина і ротова порожнина. Глотка – це камера, що тягнеться по всій довжині горла від гортані до ротової порожнини. Доступ до носової порожнини залежить від положення велюма, шматка м’якої тканини, яка утворює задню частину ротової порожнини.

Для виробництва певних фонем велум опускається вниз, з’єднуючи носову порожнину з інші камери голосового тракту.

Акустична теорія мовлення передбачає процес мовлення являти собою лінійну систему, що складається з джерела і фільтра [6]. Ця модель захоплює основи процесу виробництва мови. Динамічний фільтр голосового тракту, що змінюється, представлений на рисунку 2.6.

Відповідно до теорії джерело-фільтр, короткі часові кадри мови можна охарактеризувати шляхом виявлення параметрів джерела і фільтра.

Сигнал джерела знаходиться в одним із двох станів: серія імпульсів, яка має певну основну частоту для дзвінких звуків і білий шум для глухих звуків. Це джерело з двома станами досить добре відповідає справжній глотальній поведінці, хоча моменти змішаного збудження не можуть бути добре представлені.

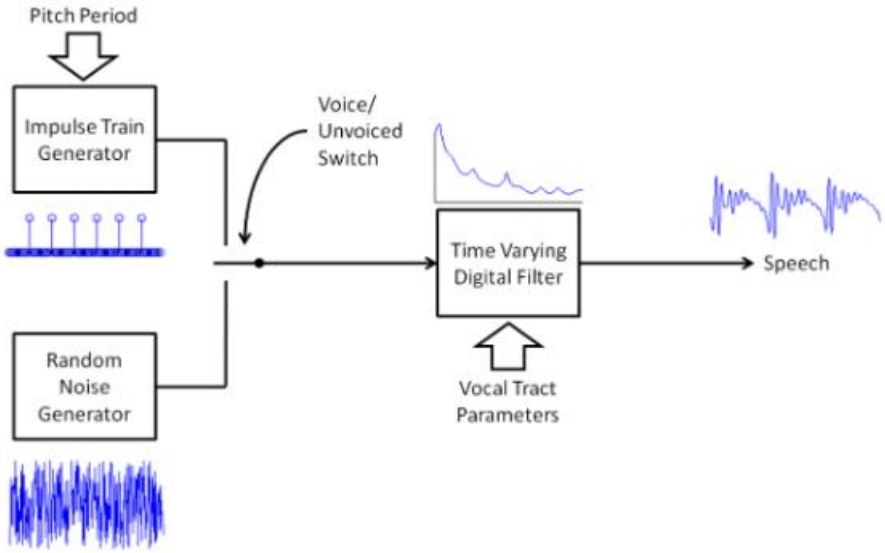

Рисунок 2.6 – Спрощена модель мовлення, запропонована акустичною теорією виробництва мови

Голосовий тракт параметризований його резонансами, які називаються формантами. Усі акустичні трубки мають природний резонанс параметри якого є функцією його форми.

Хоча голосовий тракт постійно змінює свою форму, а отже, і резонанс з плинним мовленням нерозумно вважати його статичним протягом короткого часу інтервали порядку 20 мілісекунд. Таким чином, можна переглянути мовленнєву продукцію як систему взаємодії засобів навчання (Learning Tools Interoperability, LTI) і до неї можна застосувати лінійне передбачення.

Відомо, що система виробництва мови має деякі нелінійні ефекти, а голосове джерело та фільтр голосового тракту не повністю роз’єднані.

Іншими словами, було помічено, що акустичні ефекти голосового тракту модулюють залежно від поведінки джерела таким чином, що лінійні системи не можуть повністю описати. Крім того, голосовий тракт відхиляється від поведінки все полюсного фільтру під час виробництва певних голосових звуків.

Лінійні системи за визначенням припускають, що вхідні дані системи не впливають на параметри системи. У випадку процесу виробництва мови це

означає, що вібраційна поведінка голосової щілини не має відношення до формантних частот і ширини – припущення, яке іноді порушується. Особливо в ситуації де висота голосу висока, а перша форманта центральної частоти низька, імпульс збудження може впливати на загасання попереднього імпульсу.

Описаний метод лінійного прогнозування працює в припущенні, що частотна характеристика голосового тракту складається тільки з полюсів. Це припущення прийнятно для більшості дзвінких звуків мови, але не підходить для носових і фрикативних звуків. Під час створення цих типів висловлювань у спектрі утворюються нулі завдяки захоплення певних частот у тракті.

Використання моделі, у якій відсутні представлення нулів на додаток до полюсів, не повинно викликати зайвого занепокоєння, тому що якщо p має досить високий порядок, моделі з усіма полюсами достатньо для майже всіх звуків мови.

Незважаючи на ці обмеження, всеполюсне лінійне передбачення залишається дуже корисним методом для аналізу мовлення.

Вибір порядку передбачення є важливим, оскільки він визначає характеристики фільтра голосового тракту. Якщо передбачення буде надто низьким, ключові області резонансу будуть упущені, оскільки недостатньо полюсів для їх моделювання – якщо порядок передбачення надто високий, специфічні характеристики джерела, напр. гармонік, визначаються (див. рис. 2.7). Форманти потребують двох комплексно спряжених полюсів. Спектральні огинаючі звукові труби, як визначено аналізом лінійного прогнозування, потрібують сукупний збільшений порядок прогнозування.

Таким чином, порядок передбачення повинен бути подвійним числом формант, які присутні у смузі сигналу. Як правило, в середньому припадає одна форманта на кілогерц смуги пропускання. Порядок передбачення приблизно визначається наступним чином:

$$p = \frac{f_s}{1000} + y \qquad (2.13)$$

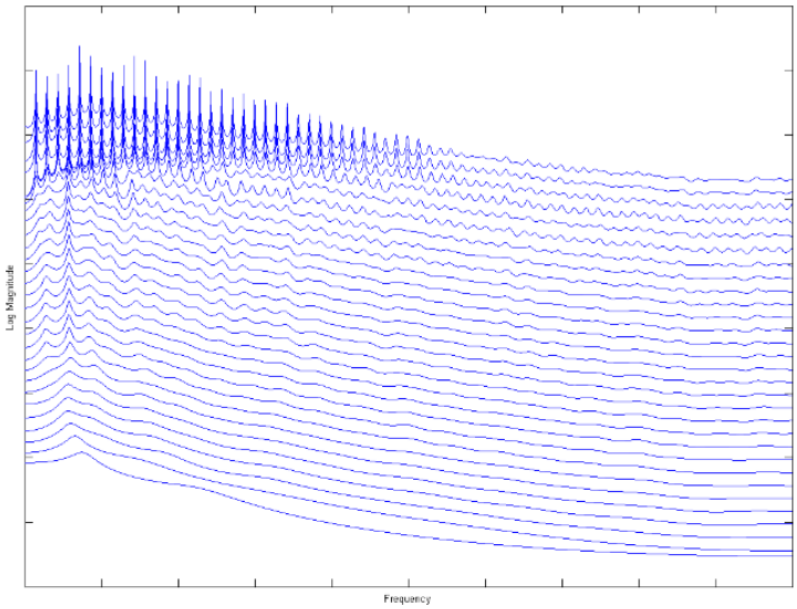

Рисунок 2.7 – Спектральні огинаючі звукової труби

Де p представляє порядок передбачення, а fs частоту дискретизації сигналу. Значення у, описане в літературі як «коефіцієнт помилки», необхідне для компенсації глотального скочування та гнучкості предиктора, зазвичай дається як 2 або 3.

3 РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕННЯ КОЕФІЦІЄНТІВ ЛІНЕЙНОГО ПЕРЕДБАЧЕННЯ

3.1 Аналітичний сигнал та його роль у цифровій обробці сигналів

Цифрова обробка сигналів (ЦОС) є однією з найпотужніших технологій, яка сформує науку та інженерію в двадцять першому столітті. Революційні зміни вже відбулися в широкому діапазоні різних сфер: зв’язок, медичне зображення, радар і гідролокатор, музика високої якості, відтворення та розвідка нафти, щоб назвати лише деякі.

Цифрова обробка сигналу передбачає розробку алгоритмів, які можна використовувати для покращення сигналу певним чином або вилучення з нього деякої корисної інформації

Для комп’ютера використовуються два підходи генерована мова: цифровий запис і симуляція голосового тракту. У цифровому запису голос людини, що говорить, оцифровується та зберігається, як правило, в стислому вигляді. Під час відтворення збережені дані розтискаються та перетворюється назад в аналоговий сигнал. Ціла година записаного виступу вимагає лише близько трьох мегабайт пам’яті, що цілком відповідає можливостям навіть невеликі комп'ютерні системи. Це найпоширеніший цифровий метод генерації мови, який використовується сьогодні.

Одразу можна виділити такі переваги:

- реалізація оптимальних алгоритмів обробки з високою точністю, програмованість і функціональну гнучкість,
- можливість адаптації до оброблюваних сигналів;
- можливість апаратної реалізації з використанням спеціальних процесорів і чіпсетів ЦОС.

На відміну від аналогових схем, цифрові схеми менш чутливі до змін компонентів і перешкод. Цифрові схеми також більш гнучкі і придатні для застосування математичних функцій.

Сфери застосування включають цифрові системи зв'язку для радіо та телебачення, системи стільникового зв'язку, комп'ютерні мережі, обробку зображень тощо.

Цифрова обробка сигналу зазвичай підходить до проблеми розпізнавання голосу в два етапи: виділення ознак з подальшим зіставленням ознак. Кожне слово у вхідному аудіосигналі ізольовано, а потім проаналізовано що б визначити тип збудження та резонансні частоти. Ці параметри потім порівнюють із попередніми прикладами вимовлених слів, щоб виявити найбільш близькосхожі.

Часто ці системи обмежені лише кількома сотнями слів, може сприймати мову тільки з чіткими паузами між словами і треба перенавчати для кожного окремого динаміка.

ЦОС набув найбільшого поширення в області обробки та передачі голосового сигналу завдяки розвитку IP-телефонії та інших форм цифрового зв'язку.

У процесі цифрової обробки важливу роль відіграє аналітична модель сигналу. Кожен справжній обмежений смугою сигнал разом зі своєю парою, перетвореною Гільбертом, формує аналітичний сигнал. Перетворення Гільберта полегшує формування аналітичного сигналу. Аналітичний сигнал корисний у сфері зв'язку, зокрема в смуговій обробці сигналу. Існує два типи компонентів: реальні, які найчастіше використовуються на практиці, та уявні, без яких неможливо представити цифрову обробку. Ці компоненти показані на рис. 3.1.

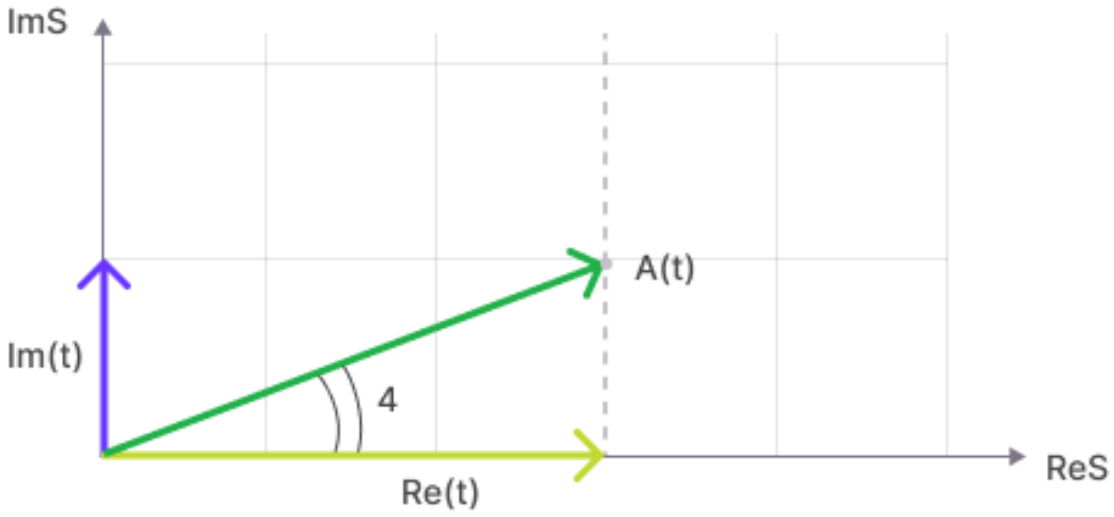

Рисунок 3.1 – Уявна і реальна складові аналітичного сигналу

Для позначення реальної та уявної складових вводимо такі позначення:

$$S_r = Re\, S(t) \tag{3.1}$$

$$S_m = Im\, S(t) \tag{3.2}$$

Ураховуючи ці вирази створюємо формулу аналітичного сигналу:

$$S(t) = S_r(t) + jS_m(t) \tag{3.3}$$

При цьому уже можемо визначити амплітудну огинаючу, завдяки тому що у нас є уявна та реальна складова, записуємо у вигляді:

$$A(t) = |S(t)| = \sqrt{S_r^{\,2}(t) + S_m^{\,2}(t)} \tag{3.4}$$

Якщо проаналізувати формулу то можна побачити, що шляхом введення уявної складової можна збільшити відношення сигнал/шум як мінімум на 40%, а зі збільшенням значення відношення буде збільшуватися.

Можемо визначити $\varphi(t)$ – фазову функцію сигналу, для цього беремо цю формулу:

$$\varphi(t) = arctg\frac{S_m(t)}{S_r(t)} \tag{3.5}$$

Дослідження фазової функції в першу чергу призначене для створення когерентне накопичення сигналу, яка значною мірою впливає на співвідношення сигнал/шум.

Швидкість зміни несучого коливання можна розрахувати з використанням співвідношення:

$$\omega(t) = \frac{S_r(t)\cdot \dot{S}_m(t) - \dot{S}_r(t)\cdot S_m(t)}{A^2(t)} \tag{3.6}$$

За рахунок використання уявної складової аналітичного сигналу значно підвищується якість процедур обробки, що суттєво впливає на достовірність і

швидкість обробки різних даних.

Виходячи з цих результатів, можна сказати, що, реєструючи дійсну та уявну складові аналітичного сигналу, ми дозволяємо використовувати переваги сучасної цифрової обробки та досягати вищої якості в ряді практичних застосувань.

Співвідношення (3.5) дозволяє розрахувати новий інформаційний параметр голосового сигналу, який раніше не використовувався при оцінці КЛП, а саме фазові дані.

Як відомо, фазові дані дуже продуктивна та ефективно використовуються в радіолокації, радіозв'язку, сейсмології тощо.

Понад те, як свідчать роботи [28-36], фазові дані можуть дуже плідна використовуватися й у системах автентифікації.

Мета даного розділу дослідити можливості фазових даних для розрахунку коефіцієнтів лінійного передбачення та виконати порівняльний аналіз з результатами, отриманими в процесі обробки амплітудно-частотної інформації.

У цьому необхідно вибрати метод дослідження, розробити експериментальну установку, і навіть синтезувати математичну модель.

3.2 Обґрунтування методу досліджень

Моделювання – метод наукових досліджень, котрий ґрунтується на використанні моделі як засобу дослідження явищ і процесів різного характеру.

Під моделями розуміють системи, що замінюють об'єкт дослідження і служать джерелом інформації стосовно нього. Моделі – це такі аналоги, подібність яких до оригіналу суттєва, а розбіжність – несуттєва.

Моделювання та імітація використовуються для дослідження систем. Система визначається як сукупність сутностей, які складають об'єкт або процес, що цікавить. Полегшувати експериментування та оцінку системи, представлення, створюється аналог, які називають моделлю. Фізичні моделі представляють системи як фактичне обладнання, тоді як математичні моделі представляють системи, як набір обчислювання або логічні асоціації.

Існують статичні моделі що представляють систему в певній точці часу, або динамічними, що представляється, як система що може змінитись з часом. Набір змінних, що описують систему у якийсь момент часу називається вектором стану. В цілому, вектор змінюється у відповідь на подію в системі.

У безперервних системах вектори стану постійно змінюються, але в дискретних системах вони змінюються лише кінцеву кількість разів. Коли хоча б одна випадкова величина є в даний час, то модель називається стохастичною; коли випадково змінні відсутні, модель називається детермінованою.

Іноді математичні співвідношення описують систему досить просто для аналітичного вирішення. Аналітичні рішення надають точну інформацію щодо продуктивності моделі. Коли модель неможливо розв'язати аналітично, він часто може імітувати системні операції, процес, який називається імітацією – можливість оцінити продуктивність системи чисельно.

Загалом системи моделюються як такі, що мають число входів (або стимулів), $x_1, x_2, \ldots . x_r$; ряд виходів (або відповіді), $y_1, y_2, \ldots . y_s$;; і ряд систем параметри (або умови), $p_1, p_2, \ldots . p_t$;. Хоча кожена система унікальна, вхідні дані часто непередбачувані, а системні параметри часто виникають як засіб для налаштування відповідей певним чином. Таким чином, вхідні дані зазвичай моделюються як випадкові процеси, і параметри системи як регульовані умови. Наприклад, вхід моделі для комунікаційної мережі може включати час надходження повідомлень і розмір кожного повідомлення.

Системні параметри можуть включати протоколи черги, кількість каналів передачі і пропускну спроможність каналів передачі на кожному комутаційної станції. Результат може включати характеристику затримок, викликаних повідомленнями в чергах, і загальний час доставки. Таким чином, система часто розглядається як функція *f*, яка створює вихідні дані у входи x і параметри системи *p*; тобто *y = f(x, p),* як показано на рис. 3.2.

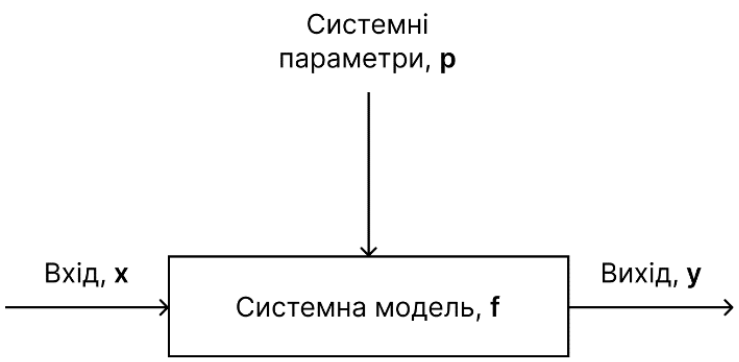

Рисунок 3.2 – Функціональний вигляд системних моделей, $y = f(x, p)$

Моделі поділяють на два види: матеріальні й ідеальні.

Матеріальні моделі втілюються у певному матеріалі – дереві, металі, склі та ін. Ідеальні моделі фіксуються в таких наочних елементах, як креслення, рисунок, схема, комп'ютерна програма тощо.

Метод моделювання має таку структуру(рис 3.3):

- постановка завдання;
- створення або вибір моделі;
- дослідження моделі;
- перенесення знань та результатів із моделі на об’єкт дослідження.

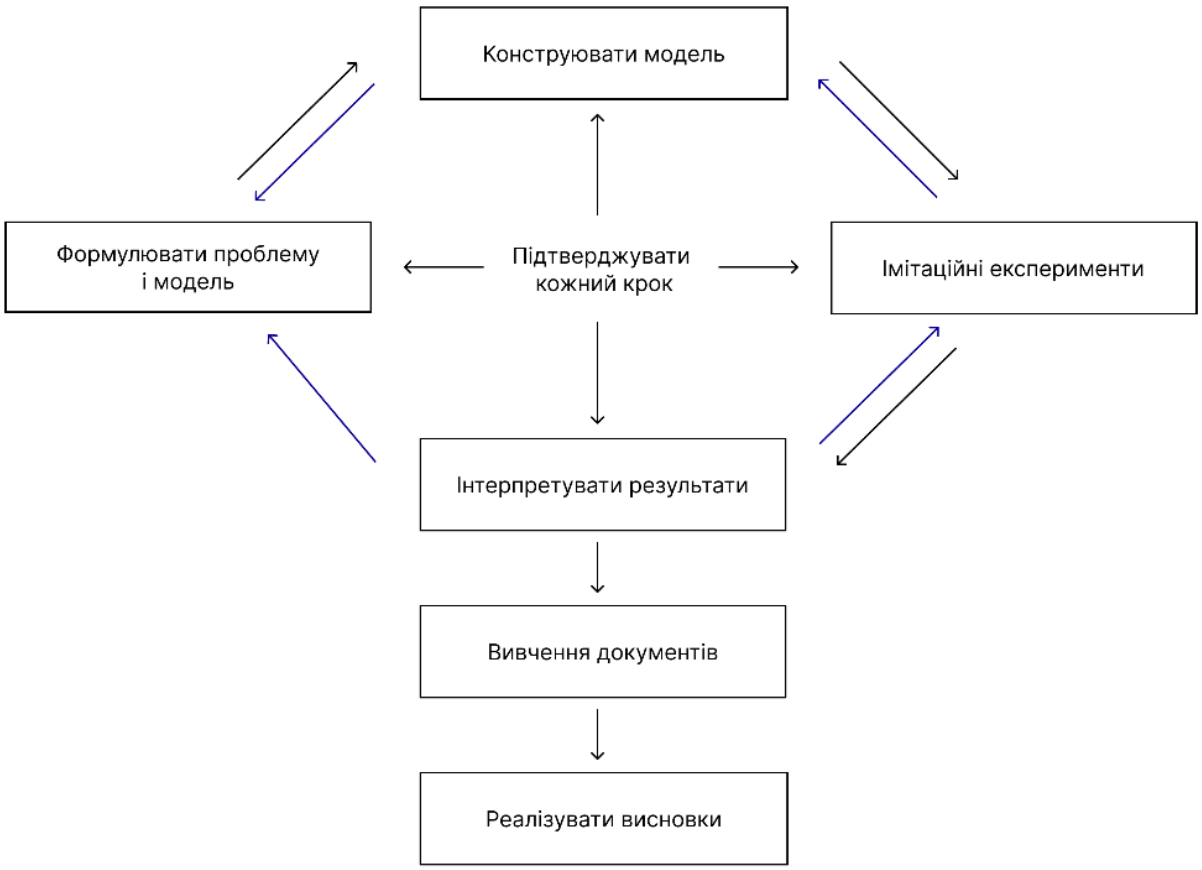

Рисунок 3.3 – Процес розв’язування задач за допомогою моделювання та імітації

Перевірка має бути це постійний процес, і, як показано на малюнку, він застосовується на всі кроки. Крім того, знання, отримані на будь-якому кроці

може знадобитися хід "назад", щоб скорегувати попередній працювати. Сині лінії на рис. 3.3 показують можливість перегляд попередньої роботи. Кожен крок у проблемі, процес вирішення детально розглядається в розділах, які слідувати.

Поява моделювання пов'язане з областю теоретичної і експериментальної фізики і відноситься до XVII століття. В основному це феноменологічні (є продуктом теоретичного узагальнення) моделі. Яскравими прикладами таких моделей є моделі, пов'язані з законами Кеплера і законом всесвітнього тяжіння.

У міру розвитку фізики в ній були розроблені моделі, що відображають внутрішні сторони об'єктів (систем) і їх структуру. Прикладом такої моделі є модель атома. При цьому широко використовувалися індуктивні методи (наведення, висновок від часткового до загального), що встановлюють зв'язок між елементами системи і системою в цілому.

Справді революційний стрибок в математичному моделюванні стався в зв'язку з початком застосування електронної обчислювальної техніки (ЕОТ), розвитком кібернетичного моделювання і моделювання в дослідженні операції, що збіглося з початком широкого використання ЕОТ при проведенні досліджень на математичних моделях (50-е ... 60-ті роки XX століття).

У сучасній науковій літературі, присвяченій різним областям знань, міцно зайняло місце вираз "математичне моделювання". Прагнення до математичної формалізації особливо проявляється в тих областях знання, де прямий експеримент, що дає можливість отримати досить повну і об'єктивну інформацію про досліджувану реальність, практично неможливий. Особливо актуальними ці питання є в галузі телекомунікацій, космічної техніки і ін.

Таким чином, моделлю називають процес, явище, предмет, установку, знаковий або умовний образ (опис, схему), які знаходяться в деякому співвідношенні відповідно до досліджуваного об'єкту і здатні заміщати його в процесі дослідження, даючи про нього інформацію. А сам процес розробки та тестування моделі і дослідження з її допомогою реального об'єкта називається моделюванням.

Моделювання по відношенню до інших відомих методів наукових досліджень має низку переваг: універсальність, економічність, гнучкість, наочність. Особливо актуальними є питання моделювання систем є для навчальних закладів. Оскільки за допомогою математичної моделі можна відтворювати структуру системи, а також фізичних процесів які протікають в

ній, аналізувати вплив окремих чинників, що вивчаються, дозволяючи, таким чином, поєднувати експериментальний і розрахунковий підхід до дослідження, об'єднуючи наочність експерименту з гнучкістю обчислювальної програми моделі.

Використання моделювання дозволяє в значній мірі вирішити одну з основних проблем сучасної науки і техніки – проблему складності.

Таким чином, математична модель повинна володіти такими основними властивостями:

- знаходиться в об'єктивної відповідності до пізнаваного (досліджуваного) об'єкту (системою);
- заміщати в певному відношенні даний об'єкт (систему);
- давати при цьому інформацію про даний об'єкт, що отримується на основі дослідження даної моделі і відповідних правил переходу модель-об'єкт (прототип).

У 60-х роках попереднього століття було введено поняття імітаційного моделювання, яке відрізняється від аналітичного двома істотними обставинами. По-перше, імітаційна модель повинна з необхідною повнотою відтворювати будову прототипу з тим, щоб висновки, одержувані при моделюванні якогось елементу моделі можна було віднести до відповідного елементу прототипу. По-друге, і в цьому головна ознака, імітаційне моделювання орієнтується на отримання знань про прототипі не шляхом аналітичного дослідження або одноразових чисельних розрахунків, а шляхом експериментів на імітаційної моделі.

Означеним вимогам можуть задовольняти як спеціально сконструйовані матеріальні імітаційні моделі, так і математичні імітаційні моделі, для реалізації яких необхідно залучати ЕОТ.

Основними достоїнствами імітаційних моделей є можливість відображення адекватним чином різних властивостей елементів системи, таких, як нелінійність, дискретність роботи, ймовірність спрацьовування, часова логіка функціонування і ін. В той же час основним недоліком імітаційної моделі в порівнянні з аналітичною є недостатність теоретичного опрацювання та принципова необхідність (для отримання достовірної інформації про властивості системи) здійснення багаторазових експериментів, які потребують калібрування результатів моделі по еталонним даним. Тому фундаментальність висновків,

отриманих даним методом моделювання буде нижче, ніж у аналітичного моделювання.

Як основний метод досліджень будемо застосовувати метод імітаційного моделювання. Для цього, в першу чергу, повинні розробити експериментальну установку та математичну модель.

3.3 Розробка експериментальної установки та математичної моделі

В основу експериментальної установки був покладений ноутбук. Це дуже зручно є багато необхідних вбудованих апаратних та програмних засобів.

Експериментальна установка для проведення досліджень включала ноутбук із операційною системою та системою комп'ютерної математики (СКМ) MatLab. Ноутбук мав убудований мікрофон та звукову карту, також була можливість підключення зовнішнього мікрофона. Як перешкодовий сигнал мав місце акустичний шум роботи вінчестера, а також внутрішні шуми мікрофона та звукової карти. Частота дискретизації сигналу становила 64 кГц. Відношення сигнал/шум аналізованої послідовності складало більше 25 дБ.

Тепер розглянемо процедури керування роботою звукової карти. Спочатку необхідно створити об'єкт із аналоговим входом (Analog Input, AI) для звукової карти. Потім додамо до AI один канал.

Для введення мовного сигналу необхідно встановити частоту часової дискретизації і кількість біт для квантування, а також час реєстрації. Для цього у СКМ MatLab використовувалася наступна послідовність команд:

```
AI = analoginput('winsound');    % Об'єкт з аналоговим входом (AI)
addchannel(AI, 1);               % Додамо до AI один канал реєстрації
Fs = 64000;                      % Частота дискретизації 64000 Гц
set (AI, 'SampleRate', Fs):
duration = 2;                    % Зйом триватиме 2 секунди
set(AI, 'SamplesPerTrigger', duration*Fs);
start(AI); % Запуск знімання даних за командою start
data = getdata(AI); % Перемістити дані до змінної data
delete(AI); % Зупинити знімання даних, скидання звукової карти
Bits = 16; % Кількість біт
```

wavwrite(data,Fs,Bits, 'd1.wav'); % Запис результатів у файл 'd1.wav'

Структурну схему експериментальної установки представлено на рис. 3.4. Для читання доцільно виконати такі команди:

wavesize = wavread('d1.wav','size'); % Виявлення довжини та числа каналів запису

N1 = 500; % Номер відліку з якого буде розпочато читання

N2 = 7800; % Номер відліку, на якому буде закінчено читання

[res,FS,BITS]=wavread('d1.wav',[N1 N2]); Читання у змінну res.

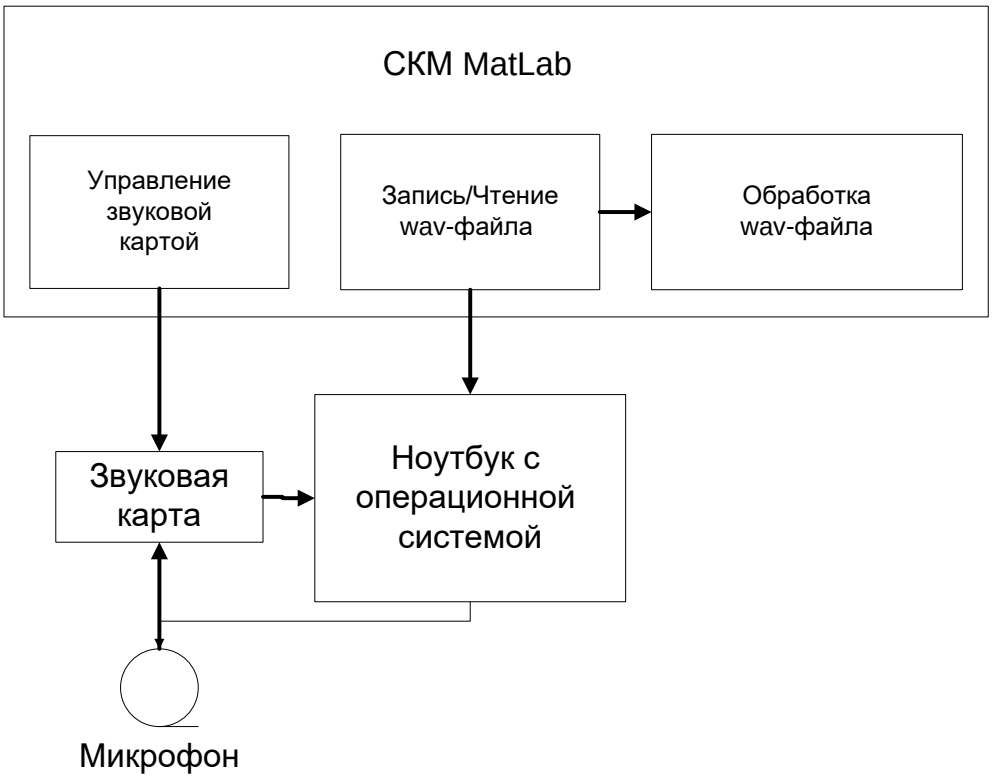

Рисунок 3.4 – Структурну схему експериментальної установки

Як процедури цифрової обробки використовувалися:

- перетворення Гільберту;
- процедури розрахунку та корегування фазових даних;
- функція оцінки коефіцієнтів лінійного передбачення по заданим даним;
- процедури графічного відображення голосового сигналу та статистичної обробки результатів розрахунків.

Вся цифрова обробка була реалізована середовищі комп'ютерної математики MatLab.

Аналізу піддавався мовленнєвий сигнал користувача цифр, який називав цифри від 0 до 9. Введення мовленнєвого сигналу здійснювалося з відстані 0,7,…,1 м нормалі до осі мікрофона в замкнутому приміщенні.

При цьому основну увагу приділятимемо аналізу діапазону спектру до 8 кГц, що обумовлено наявністю відмітних ознак користувача в його мовленнєвому сигналі (в діапазоні від 0,1 кГц до 8 кГц). Для цього розрахований спектр, діапазон зміни якого визначається половиною частоти часової дискретизації, обмежуватимемо частотою 8 кГц (далі, «короткий» спектр).

Структурна схема програми цифрової обробки голосового сигналу наведена на рис. 3.5, а текст у додатку А.

Аналітичний сигнал формується у комп’ютерної програмі за допомогою перетворення Гільберта, яке у СКМ MatLab виконується за допомогою функції sq2 = hilbert(x). Звернімо увагу на розрахунок фазового кута, який виконується за допомогою функції арктангенсу.

У зв'язку з тим, що множина значень функції арктангенс знаходиться в інтервалі $(-\pi/2;\ \pi/2)$ результат (φ) для більшості чвертей необхідно відкоригувати, як показано у табл. 3.1. При корекції враховуються знаки (позитивний або негативний) речовинної та уявної складової аналітичного сигналу, які беруть участь у розрахунку фазового кута.

Таблиця 3.1 – Корегування фазового кута

№ чверті	1	2	3	4
Значення кута (результат функції)	$0< \varphi<\pi/2$	$\pi/2< \varphi<\pi$	$\pi< \varphi<3\pi/2$	$3\pi/2< \varphi<2\pi$
Справжнє значення (результат корекції)	$\varphi_i = \varphi$	$\varphi_i = \varphi + \pi$	$\varphi_i = \varphi + \pi$	$\varphi_i = \varphi + 2\pi$

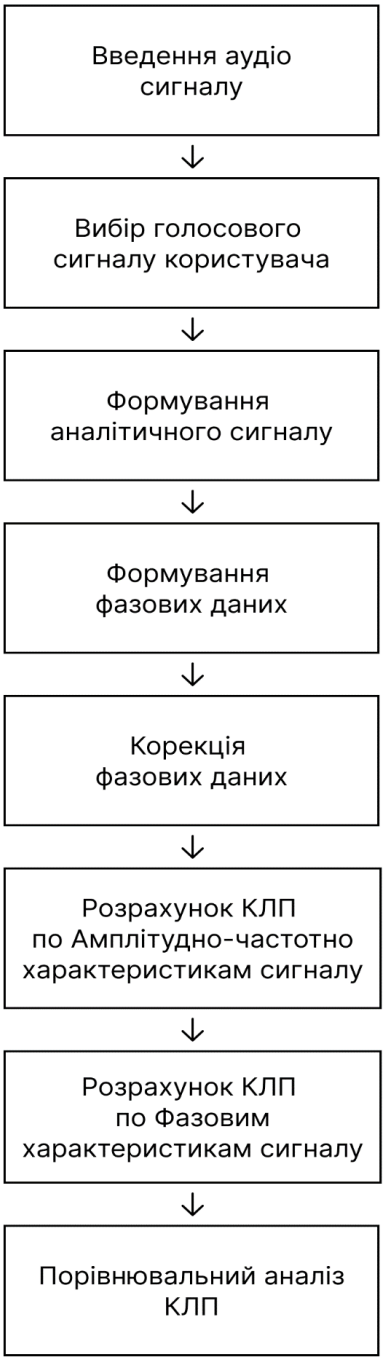

Рисунок 3.5 – Схема імітаційної моделі

Покажемо викладене вище з прикладу. На рисунку 3.6 представлений сигнал користувача, який піддавався цифровій обробці. Попередньо виконувалося перетворення Гільберта, а потім обчислювалася функція арктангенсу для формування фазових даних.

На рис. 3.7 представлені результати розрахунку функції арктангенс для фрагмента сигналу, що аналізується. Як і слід очікувати фазові дані не коректні. Тому надалі здійснювалися процедури корекції фазових даних, що розглянуті.

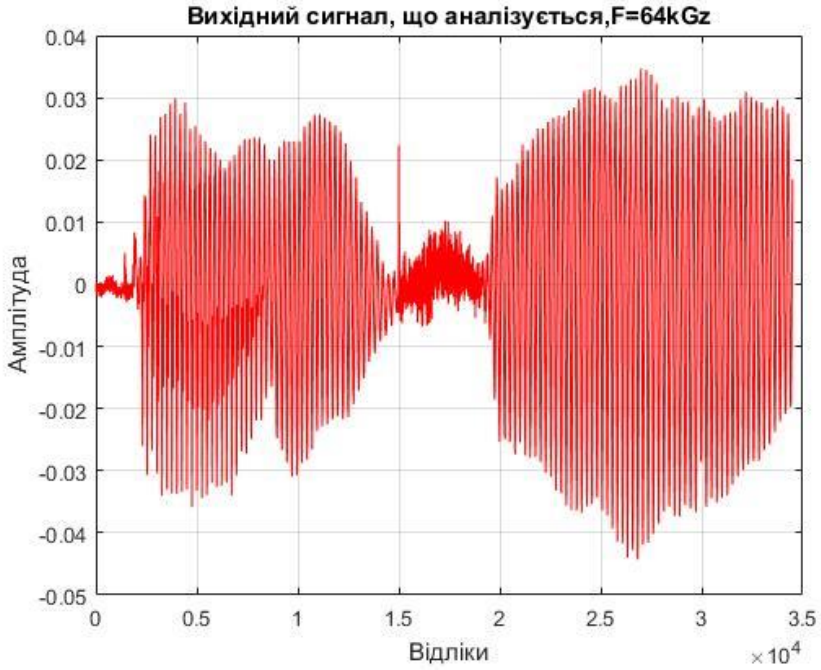

Рисунок 3.6 – Сигнал користувача системи автентифікації

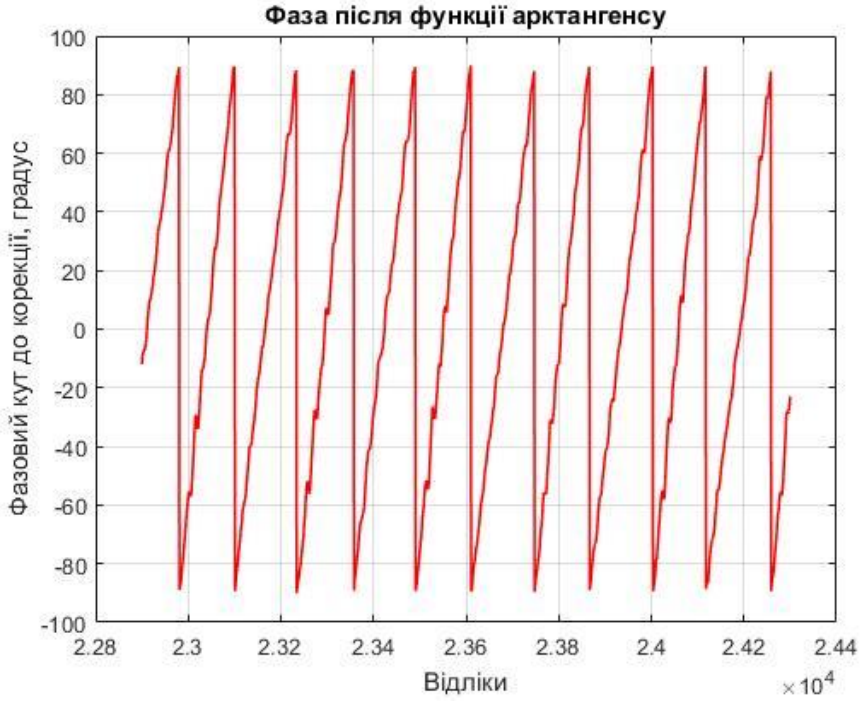

Рисунок 3.7 – Фазовий сигнал після розрахунку функції арктангенс

Результати корекції фазового кута представлені на рисунку 3.8.

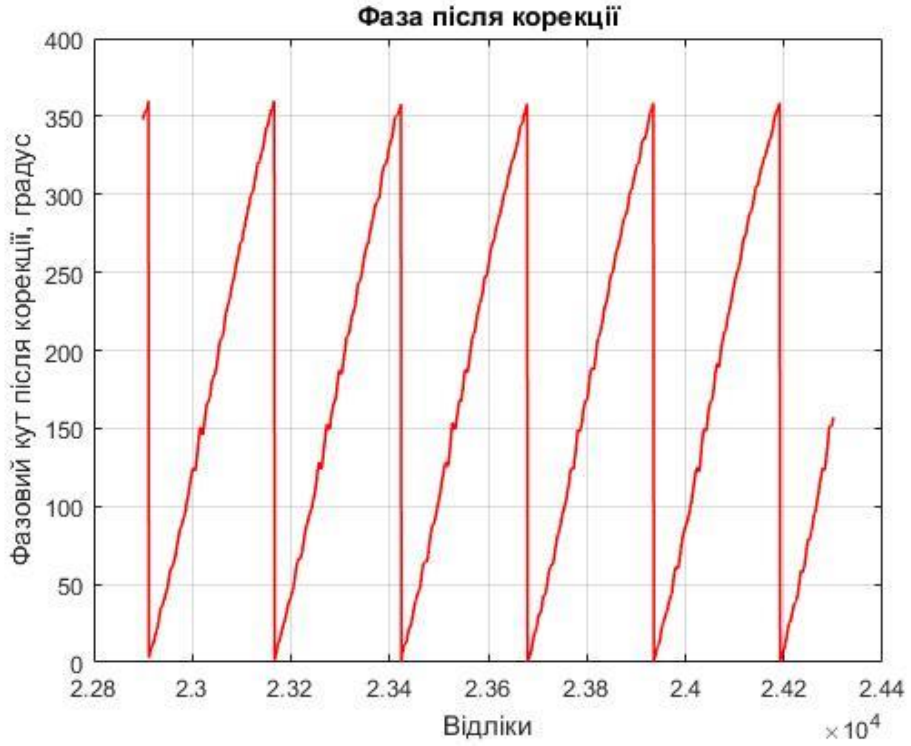

Рисунок 3.8 – Результат коригування фазового кута

3.4 Результати дослідження коефіцієнтів лінійного передбачення

Як відомо, у радіозв'язку, де вперше використовувалися КЛП дають можливість стиснення інформації при передачі та відновлення на приймальній стороні. У системах аутентифікації аналогічні процедури стиснення інформації необхідні для формування ознак шаблону і, як показала практика, можуть характеризувати голос користувача. При цьому розрахунок КЛП у практичних додатках виконується за амплітудно-частотною інформацією голосового сигналу користувача. Наша гіпотеза – оцінка КЛП щодо фазової інформації голосового сигналу та порівняння отриманих даних з тими, які розраховані за амплітудно-частотною інформацією.

Для цього візьмемо фрагмент аналізованого голосового сигналу, що дорівнює одному періоду зміни фази (див. рис. 3.9. та 3.10)

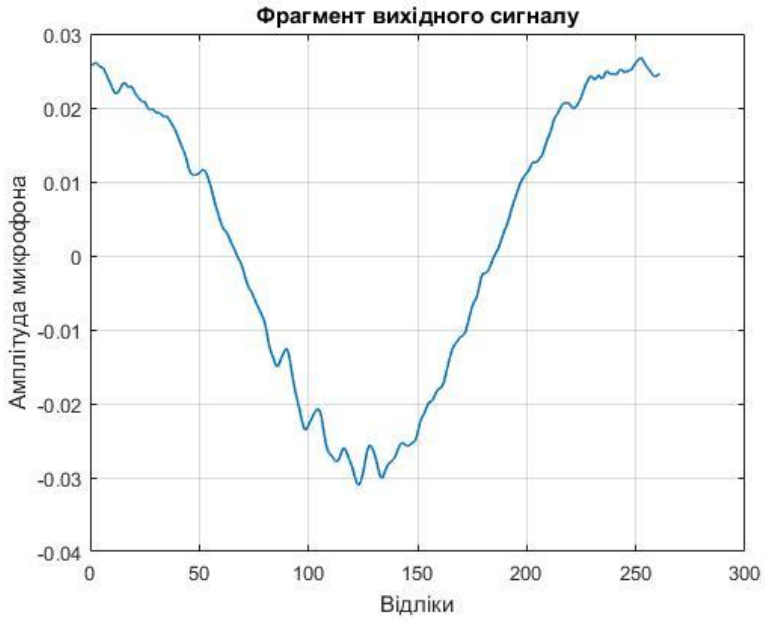

Рисунок 3.9 – Аналізований фрагмент голосового сигналу

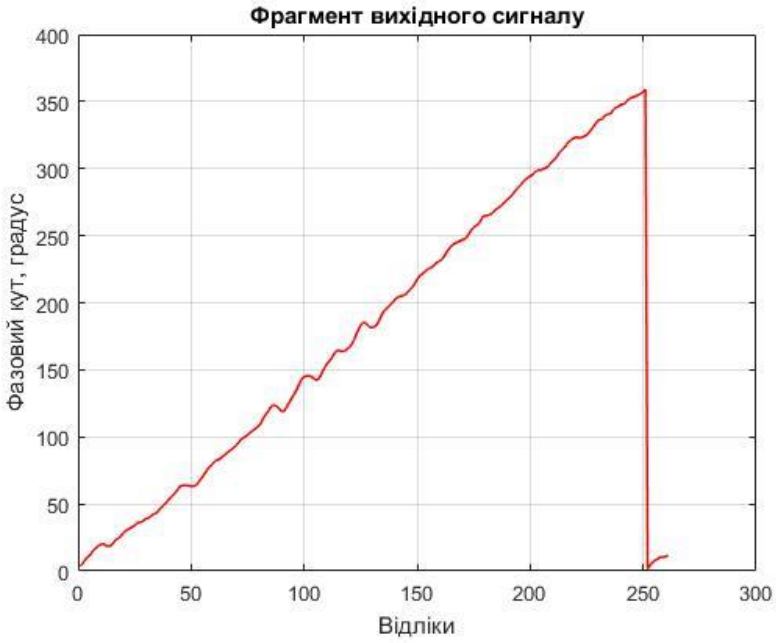

Рисунок 3.10 – Аналізований фрагмент фазового сигналу

За даними фрагментами розрахуємо КЛП за допомогою функції $[af, gf] = lpc\ (ys, p)$. Результати розрахунку подано на рис. 3.11 – 3.13.

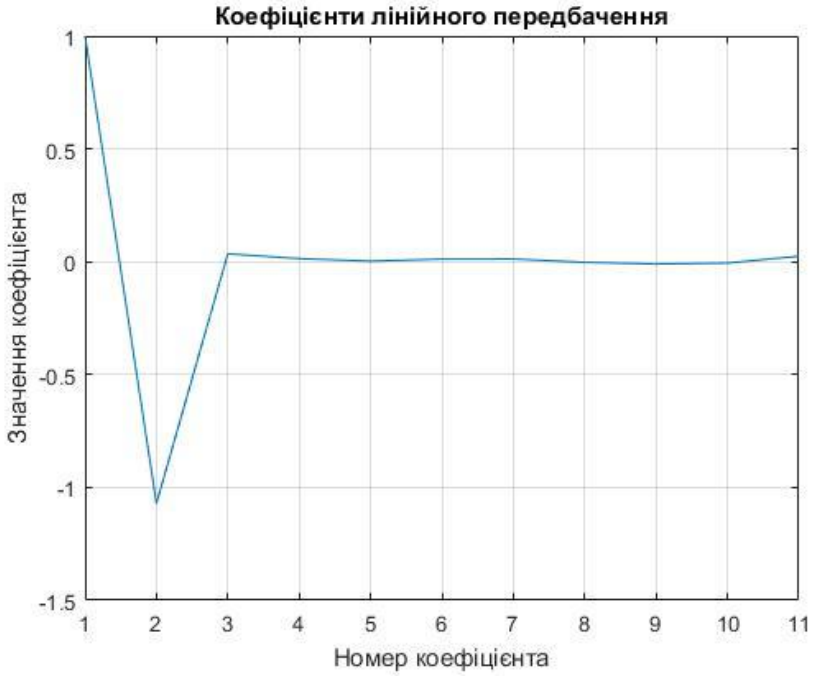

Рисунок 3.11 – Залежність КЛП (амплітудно-частотні дані)

Рисунок 3.12 – Залежність КЛП (фазові дані)

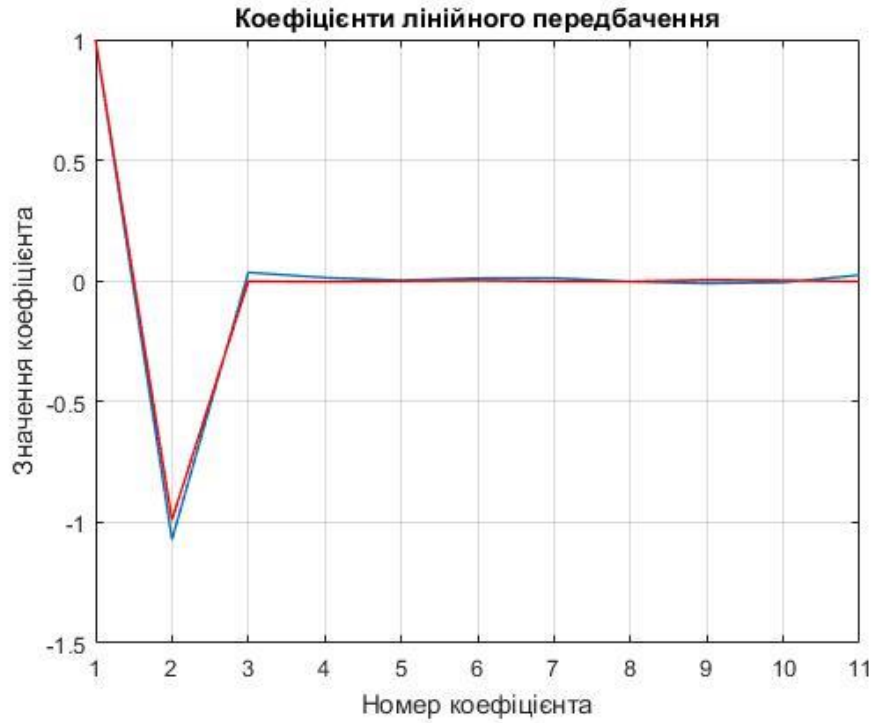

Рисунок 3.13 – Порівняльна характеристика залежностей

Різниця у величині значень коефіцієнтів представлена на рис. 3.14.

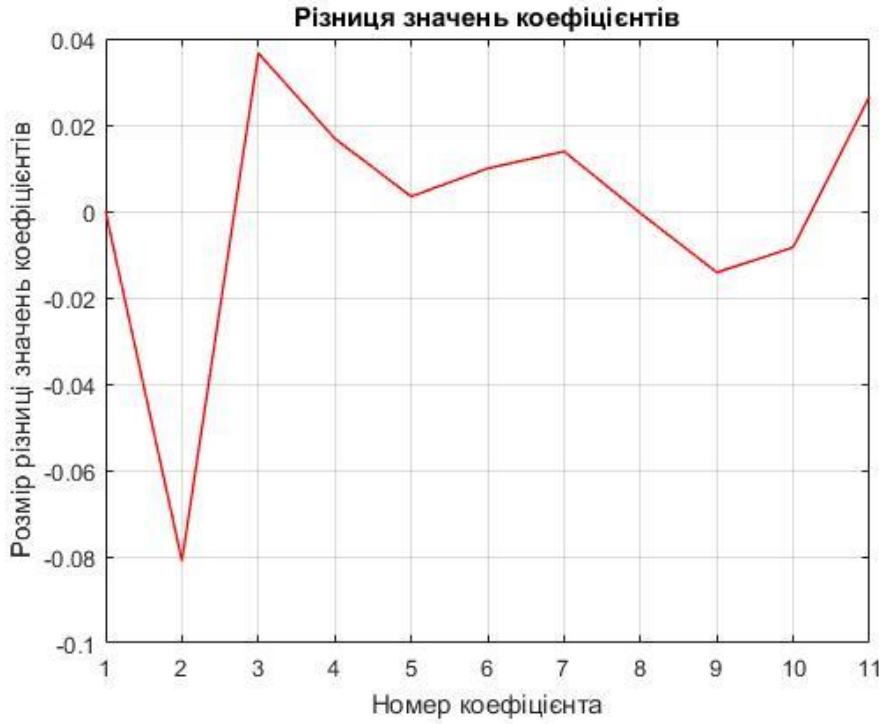

Рисунок 3.14 – Різниця у величині значень КЛП

Як бачимо, характер змін коефіцієнтів збігається, а величина помилок незначна. Абсолютна величина значень коефіцієнтів значно зменшується, починаючи з третього. По малюнках виконали якісний аналіз.

Тепер перейдемо до кількісного аналізу.

Значення КЛП, які розраховані за амплітудно-частотною інформацією, представлені нижче.

1 -1,07126 0,03499 0,01376 0,00254 0,01160 0,01205 -0,00272 -0,01019 -0,00592 0,02435.

Значення КЛП, які розраховані по фазовій інформації мають наступний вигляд:

1 -0,99027 -0,00162 -0,00308 -0,00086 0,00166 -0,00181 -0,00226 0,00398 0,00242 -0,00232.

Порівняльний аналіз показує їхню схожість. Однак абсолютна величина значень КЛП значно менша. Зауважимо, що раніше розраховані дані, наприклад, формантні частоти, мали протилежний характер.

Для більш детального кількісного аналізу скористаємося нормованим коефіцієнтом кореляції (НКК), для розрахунку якого використовується наступне співвідношення

$$x = \left| \frac{\sum_i (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_i (x_i - M_x)^2 * \sum_i (y_i - M_y)^2}} \right| \qquad (3.7)$$

де Мх і Му математичні очікування для масивів X[] та Y[] відповідно (КЛП, розраховані по різним інформаційним параметрам голосового сигналу), обчислюються за такою формулою

$$M_Z = \frac{1}{N} \sum_1^N Z_i \qquad (3.8)$$

При розрахунку нормованого коефіцієнта кореляції змінюваною величиною була довжина вибірки, якою здійснювався розрахунок. Результати розрахунків представлені у таблиці 3.2.

Таблиця 3.2 – Результат оцінки НКК для КЛП

Довжина вибірки	50	150	260	300	350
Величина НКК	0.9997	0.9986	0.9985	0.9965	0.9974
Довжина вибірки	450	550	750	950	1250
Величина НКК	0.9919	0.9923	0.9946	0.9814	0.98

Аналіз розрахованих даних дає право зробити висновок, що КЛП отримані за аналізованими інформаційними параметрами голосового сигналу (амплітудно-частотні та фазові дані) добре узгоджуються. Це дає право включати до шаблону користувача системи автентифікації додаткову ознаку – коефіцієнти лінійного передбачення, отримані на основі фазових даних. Такий підхід може підвищити ефективність та надійність голосової системи автентифікації.

ВИСНОВКИ

Багато дослідників пов'язують засоби підвищення надійності систем аутентифікації з впровадженням біометричних характеристик користувача. Однак з часом стало зрозуміло, що не всі засоби біометричної аутентифікації ефективні та надійні. Такі системи як двофакторна автентифікація користується популярністю серед способів захисту інформації. Але вона так як і інші способи автентифікації виявилась ненадійною, так як високий ризик викрадення ключів. Тому серед багатьох, система голосової автентифікації виявилася одним із найнадійніших засобів.

Перевагами системи є:

- простота використання(клієнту не потрібні якісь спеціальні навички або знання)
- дешевизна (не потрібне дороге обладнання, його постійне обслуговування, із засобів вводу інформації потрібен мікрофон телефону або на комп’ютері).

У цій галузі ще тривають роботи над вдосконаленням цієї системи, наприклад проблема підвищення якісних характеристик систем автентифікації голосу. Як варіант вдосконалення було запропоноване використання коефіцієнту лінійного передбачення з використанням фазових даних аналізованого голосового сигналу в процесі цифрової обробки, результати якого були наведені в розділі 3.4.

Лінійне передбачення – це техніка обробки сигналів, яка широко використовується в аналізу мовленнєвих сигналів.

КЛП дають можливість стиснення інформації при передачі та відновленні на приймальній стороні, ці процедури в системах автентифікації необхідні для формування ознак шаблону і можуть характеризувати голос користувача.

Зауважимо, що фазовий простір голосового сигналу безпосередньо пов'язаний з поняттям аналітичного сигналу, який ефективно і продуктивна використовується в радіолокації та радіозв'язку.

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

В основу експериментальної установки був покладений ноутбук. Далі була виставлена частота 64кГц та налаштування відношення сига налаштування відношення сигнал/шум аналізованої послідовності, що складало більше 25дБ. Завдяки СКМ MatLab за допомогою команд було встановлено частоту часової дискретизації і кількість біт для квантування, час реєстрації.

Як процедури цифрової обробки було використовувано перетворення Гільберта, процедури розрахунку та корегування фазових даних, функція оцінки КЛП по заданим даним та процедури графічного відображення голосового сигналу та статичної обробки результатів розрахунків.

Було сформовано аналітичний сигнал за допомогою перетворення Гільберта, фазовий кут який виконується за допомогою функції арктангенс, множина значень якого знаходився в інтервалі $(-\pi/2; \pi/2)$ тому φ потрібно буде коригувати на більшості чвертей. Після цього було запущено програму і, як і очікувалось, дані були не точними, тому далі проводилась корекція фазових даних.

Для розрахунку КЛП по фазовій інформації було взято аналізований голосовий сигнал та за допомогою функції $[af, gf] = lpc\,(ys, p)$ в MatLab перевірено гіпотезу.

Після аналізу розрахованих даних робимо висновок що КЛП отримані по аналізованим інформаційним параметрам голосового сигналу(амплітудно частотним та фазовим даним) добре узгоджуються.

Характер змін коефіцієнтів збігається, величина помилок незначна. Після проведення порівняльного аналізу можна зазначити схожість цих двох методів, але абсолютна величина значень КЛП значно менша, ніж розраховують, наприклад, формантні частоти, де вони навпаки більші.

Для більш детального кількісного аналізу скористались нормованим коефіцієнтом кореляції, де змінювали величину довжини вибірки. При аналізі результатів цих розрахунків робимо висновок, що при збільшені значення довжини вибірки величина НКК зменшується.

Отже, наукова новизна отриманих результатів полягає в тому, що вперше розроблено методику та проведено експериментальні дослідження щоб сформувати оцінку КЛП щодо фазової інформації голосового сигналу та порівняння отриманих даних, з тими які зазвичай розраховуються за амплітудно-частотною інформацією.

Фазові дані мовленнєвого сигналу дозволяють отримати адекватні та надійні оцінки в процесі аналізу.

Підхід який був запропонований може повисити ефективність і надійність голосової системи автентифікації. Доцільно проводити подальші дослідження в цьому напрямку з використанням КЛП на основі фазових даних.

Результати досліджень оприлюднені в трьох наукових роботах та доповідались на міжнародних конференціях, де отримали позитивну оцінку.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Методи наукових досліджень про телекомунікаціях. У 2-х томах: навчальний посібник за ред. В.В. Поповський. Харків: Компанія ЗМІТ, 2013. – 720 с.
2. Вимірювання розбірливості мови: формантний підхід [Електронний ресурс] Режим доступу: http://habrahabr.ru/post/128213/
3. Покровський Н.Б. Розрахунок та вимір розбірливості мови / Н.Б. Покровський // М.: Держ. вид. літ. з питань зв'язку та радіо. 1962. - 390 с.
4. Рамішвілі Г.С. Автоматичне впізнання того, хто говорить по голосу / Г.С. Рамішвілі // М.: Радіо та зв'язок, 1981. - 224 с.
5. Laver, J.: Principles of Phonetics / J. Laver // Cambridge Press, New York (1994). ISBN: 0-521-45031-4
6. Pollack L. On the Identification of Speakers by Voice / L.Pollack, J.M. Pickett, W.H. Sumby // JASA. 1954. Vol.26. No.3.p. 403-406.
7. Камені Нгалані Г. Б., Кіщенко М. І., Пастушенко М. С. Напрямки підвищення якості систем голосової Аутентифікації 2022
8. Радзишевський А. Ю. Основи аналогового та цифрового звуку / А. Ю. Радзишевський // М: Вид. Будинок "Вільямс" 2006. - 288с.
9. Фант Г. Акустична теорія речетворення: пров. з англ. Л.А. Варшавського. / Г. Фант // М.: Наука, 1964. - 284 с.
10. Hollien H. Forensic Voice Identification / H. Hollien // San Diego, CA: Academic Press, 2002. – 256 с.
11. Rose P. Forensic Speaker Identification / P.Rose // London: Taylor & Francis, 2002. – 360 с.
12. Stevens S.S. The Relation of Pitch to Intensity / S.S Stevens // Journal of the Acoustical Society of America 6(3), 150–154 (1935)
13. Болл Р.М. Посібник з біометрії: пров. з англ. Н.Є. Агаповій / Р.М. Болл // М.: Техносфера, 2007. - 368 с.
14. ГОСТ Р 54412 – 2011/ ISO/IEC/TR 24741:2007 Інформаційні технології. Біометрія // М.: Стандартінформ, 2012. - 58 с.
15. Pollack L. On the Identification of Speakers by Voice / L.Pollack, J.M. Pickett, W.H. Sumby // JASA. 1954. Vol.26. No.3.p. 403-406.
16. Levinson, N.: The Wiener RMS (Root-Mean-Square) Error Criterion in

Filter Design and Prediction/ N. Levinson // Journal of Mathematics and Physics 25, 261–278 (1947)

17. Durbin, J.: Efficient Estimation of Parameters in Moving Average Models / J. Durbin // Biometrika 46, 306–316 (1959)

18. Durbin, J.: The Fitting of Time Series Models / J. Durbin // Revue Institute International de Statistic 28, 233–243 (1960)

19. Broersen, P.: Accurate ARMA models with Durbin's second method. / P. Broersen // In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1999) vol. 3, pp. 15–19 (1999)

20. Rabiner, L., Juang, B.H.: Fundamentals of Speech Recognition / L. Rabiner, B.H. Juang // Prentice Hall Signal Pro-cessing Series. PTR Prentice Hall, New Jersey (1990). ISBN: 0-130-15157-2

21. Makhoul, J.: Spectral Linear Prediction: Properties and Applications / J. Makhoul // IEEE Transactions on Audio, Speech and Signal Processing 23(3), 283–296 (1975)

22. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Lui, H.H.: The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis / N.E Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Lui // Proceedings of the Royal Society of London 454(1971), 903–995 (1998)

23. Molla, M., Hirose, K., Minematsu, N.: Robust speaker identification system using multi-band dominant features with empirical mode decomposition./ M. Molla, K. Hirose, N. Minematsu // In: 10th International.

24. Ідентифікація голосу [Електронний ресурс] Режим доступу: http://www.des-crypto.ru/itsecur/voice.

25. Сорокін В.М. Розпізнавання особистості за голосом: аналітичний огляд / В. Н. Сорокін, В. В. В'югін, А. А. Тананікін // Інформаційні процеси. М: РАН. 2012. Т. 12. № 1. С. 1-30.

26. Сучасні біометричні методи ідентифікації [Електронний ресурс] Режим доступу: http://www.polyset.ru/article/st327.php.

27. Голубинський О.М., Булгаков О.М. Математичні моделі мовних сигналів для верифікації та ідентифікації особи за голосом./О.М. Голубинський, О.М. Булгаков // Вороніж: Видавничо-поліграфічний центр Воронезького державного університету 2010 року.

28. Пастушенко Н.С. Експериментальне дослідження інформативності амплітудного спектра голосового сигналу для автентифікації користувача [Електронний ресурс]/Н.С. Пастушенко, Б.Д. Малонга, О.М. Файзулаєва // Проблеми телекомунікацій. - 2015. - № 2 (17). - С. 3-11.

29. Пастушенко Н.С. Експериментальні дослідження амплітудного та фазового спектрів мовного сигналу користувача систем голосової автентифікації [Електронний ресурс]/О.М. Файзулаєва, Н.С. Пастушенко // Проблеми телекомунікацій. - 2016. - № 2 (19). – С. 28 – 34. – Режим доступу до журн.: http://pt.journal.kh.ua/2016/2/1/162_pastushenko_speech.pdf

30. M. Pastushenko and O. Faizulaieva, "Employment of phase characteristics of user voice signal in authentication systems," 2016 Third International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T), Kharkiv, 2016, pp. 205-206.

31. Пастушенко Н.С. Дослідження фазових характеристик голосового сигналу користувача системи аутентифікації [Електронний ресурс]/І.С. Павленко, Н.С. Пастушенко, О.М. Файзулаєва // Проблеми телекомунікацій. - 2017. - № 2 (21). – С. 52 – 60. – Режим доступу до журн.: http://pt.journal.kh.ua/2017/2/1/172_pavlenko_signal.pdf

32. Пастушенко Н.С. Дослідження інформативності фазових даних голосового сигналу користувача системи аутентифікації. [Електронний ресурс]/Н.С. Пастушенко, В. Педро, О.М. Файзулаєва // Проблеми телекомунікацій. – 2018. – № 1(22)

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

34. Pastushenko M. Analysis of voice signal phase data informativity of authentication system / Mykola Pastushenko, Yana Krasnozheniuk, Oleksandr Lemeshko // Zaporizhzhia, Ukraine, April 27-May 1, 2020. Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020). PP 1040-1053

35. M.Pastushenko, Ya.Krasnozheniuk, M. Zaika, "Investigation of Informativeness and Stability of Mel-Frequency Cepstral Coefficients Estimates based on Voice Signal Phase Data of Authentication System User," International Conference

"Problems of Infocommunications. Science and Technology" (PIC S&T′2020), 2020, pp. 1-5.

36. Кіщенко М. І., Пастушенко М.С. Напрямки підвищення ефективності голосових систем автентифікації. Сема Міжнародна науково-технічна конференція «Проблеми електромагнітної сумісності перспективних безпроводових мереж зв'язку», EMC-2021. pp. 1-2

37. Камені Нгалані Г. Б., Кіщенко М. І., Пастушенко М. С. Напрямки підвищення якості систем голосової аутентифікації. Міжнародна науково-практична конференція “Застосування інформаційних технологій у підготовці та діяльності сил охорони правопорядку” 15 березня 2022 року, м. Харків. Pp 87-88.

38. Кіщенко М. І., Камені Нгалані Г. Б., Пастушенко М.С. Напрямки удосконалення процедур попередньої обробки голосових сигналів в системах автентифікації. Восьма Міжнародна науково-технічна конференція «Проблеми електромагнітної сумісності перспективних безпроводових мереж зв'язку», EMC-2022. pp. 1-2