Міністерство освіти і науки України
Харківський національний університет
радіоелектроніки

Факультет *інфокомунікацій*
(повна назва)

Кафедра *інформаційно-мережної інженерії*
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка
рівень вищої освіти другий (магістерський)

Дослідження методів розпізнавання віку та статі людини за її голосом в умовах дії вузькосмугових завад
(тема)

Виконав:
студент 2 курсу, групи ІМІм-21-1
Клименко Д.Р.
(прізвище, ініціали)
Спеціальність 172 «Телекомунікації та радіотехніка»
(код і повна назва спеціальності)
Освітня програма Інформаційно-мережна інженерія
(повна назва освітньої програми)

Керівник доц.Омельченко С.В.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ Безрук В.М.
(підпис) (прізвище, ініціали)

2022 р.

Не містить відомостей, заборонених до відкритого публікування

Студент _____ *Клименко Д.Р.*
(підпис) (прізвище та ініціали)

Керівник _____ *Омельченко С.В.*
(підпис) (прізвище та ініціали)

Харківський національний університет радіоелектроніки

Факультет *інфокомунікацій*
Кафедра *інформаційно-мережної інженерії*
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 Інформаційно-мережна інженерія
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри ІМІ _____
(підпис)
«____» _____ 2022 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові *Клименко Дмитру Романовичу*
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів розпізнавання віку та статі людини за її голосом в умовах дії вузькосмугових завад
затверджена наказом по університету від « 21 » жовтня 2022 року № 1376 *Ст*

2. Термін подання студентом роботи до екзаменаційної комісії 2022р.

3.Вихідн іданідо роботи *Проаналізувати методи роспізнавання віку та статті людини за її голосом. Синтез методу роспізнавання людини за її голосом. Дослідити алгоритми роспізнавання голосу людини*

4. Перелік питань, що потрібно опрацювати в роботі
Вступ
1.Особливості підходів роспізнавання статі та віку людини за її голосом
2.Методи попередньої обробки мовних сигналів
3.Методи роспізнавання статті та віку людини
4.Експеременталні дослідження якості роспізнавання статті та віку
Висновки

5. Перелік графічного матеріалу із зазначенням креслень, схем, плакатів, комп'ютерних ілюстрацій (слайдів):
Слайди у форматі Power Point:

КАЛЕНДАРНИЙ ПЛАН

	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
	Ознайомлення із завданням. Уточнення ТЗ.	*24.10.22*	
	Підбір літератури за темою роботи	*30.10.22 -10.11.22*	
	Виконання розділу 1	*10.11.22 -21.11.22*	
	Виконання розділу 2	*22.11.22 -28.11.22*	
	Виконання розділу 3	*29.11.22 -05.12.22*	
	Виконання розділу 4	*07.12.22 -12.12.22*	
	Оформлення пояснювальної записки	*12.12.22 -15.12.22*	
	Оформлення презентаційного матеріалу,		
	підготовка до захисту у ЕК	*15.12.22 – 18.12.22*	

Дата видачі завдання __24__ __жовтня__2022 р.

Студент _____
(підпис)

Керівник роботи _____ ____доц. Омельченко С.В.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: с., рис., 1 додат.

Мета роботи: розробка та дослідження методів розпізнавання віку та статі людини за її голосом .

Об'єктом дослідження є методи розпізнавання віку та статі людини за її голосом .

Одними з найбільш проблемних місць є недостатня вивченість вибору ознак і вирішальних правил розпізнавання віку людини за її голосом в умовах дії завад.

Це необхідно для підвищення якості сегментації обличчя людини за його зображеннями.

Наведено результати дослідження алгоритмів розпізнавання віку людини за її голосом.

РОЗПІЗНАВАННЯ, ОБРОБКА, ОЗНАКИ, ФІЛЬТРАЦІЯ ШУМУ, АЛГОРИТМІВ.

THE ABSTRACT

Explanatory note: 67 p., 36 Fig., 1 Apendexs .

Purpose: development and research of methods for recognizing the age and sex of a person by his voice.

The object of the study are methods of recognizing the age and sex of a person by his voice.

One of the most problematic places is the insufficient study of the choice of signs and decisive rules for recognizing a person's age by his voice in the conditions of interference.

This is necessary to improve the quality of segmentation of the human face by its images.

The results of research of algorithms of recognition of age of the person by his voice are resulted.

RECOGNITION, PROCESSING, SIGNS, NOISE FILTRATION, ALGORITHM RESEARCH.

1 ОСОБЛИВОСТІ ПІДХОДІВ РОЗПІЗНАВАННЯ СТАТІ ТА ВІКУ ЛЮДИНИ ЗА ЇЇ ГОЛОСОМ 10
1.1 Огляд відомих методів розпізнавання віку та статі людини за її голосом 10
1.2 Особливості моделі мовотворення для різного віку та статі 12
1.3 Параметри моделі мовотворення, що є інформативними для розпізнавання віку та статі людини 15
2 МЕТОДИ ПОПЕРЕДНЬОЇ ОБРОБКИ МОВНИХ СИГНАЛІВ 18
2.1 Особливості попередньої обробки 18
2.2 Особливості віднімання спектрального шуму 20
2.3 Нормалізація середнього значення мовного сигналу 22
2.4 Особливості RASTA фільтрації 23
2.5 Застосування градчастого фільтру для захист від завад 23
3 МЕТОДИ РОЗПІЗНАВАННЯ СТАТІ ТА ВІКУ ЛЮДИНИ 26
3.1 Підготовка до оцінювання вектору ознак 26
4 РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ ЯКОСТІ РОЗПІЗНАВАННЯ САТІ ТА ВІКУ 39
4.1 Результати експериментальних досліджень де використовується в якості вирішую чого правила метод KNN (K-Найближчих сусідів) 39
4.2 Результати експериментальних досліджень з використанням класифікатора SVM 45
4.3 Результати класифікації за віковом 55
ВИСНОВОК 57
ПЕРЕЛІК ПОСИЛАНЬ 59
ДОДАТОК А ..... **Ошибка! Закладка не определена.**

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

АР– авторегресії
АРКС- авто регресії ковзного-середнього
КС ковзного-середнього
ОС – операційна система
ОТ – основний тон
ПЗ – програмне забезпечення
ПК – персональний комп’ютер
API – Application programming interface (прикладний програмний код);
GMM – Gauss Mixture Model (модель Гауссовых суміщів);

HMM – Hidden Markov Model (скриті Марковскі моделі);

IDE – Integrated Development Environment (інтегроване середовище розробки);

LPCC – Linear Predictive Cepstral Coefficients (кодування з лінійним предиктором);

MFCC – Mel-Frequency Cepstrum Coeffcents (мел-частотні кепстральні коефіцієнти);

SVM – Support vector machine (метод опорних векторів); Wav – Waveform Audio File Format (формат файлу-контейнера для зберігання записи оцифрованого аудіопотоку);

ІТ – інформаційні технології ;

Кепстр – спектр спектра сигнала;

Мел – одиниця виміру висоти сприйманого звуку; НМ – нейронна мережа;

ОС – операційна система;

ПЗ – програмне забезпечення;

Спектр – розподіл деякої фізичної величини за іншою величиною;

ЦП – цифрові пристрої;

ВСТУП

Автоматичне розпізнавання віку та статі для ряду мовних програм дуже важливою. Наприклад, реклама може бути специфічною для окремої статі та віку людини, а тому більш ефективною, що збільшує прорва товару. Але із за великих витрат на обчислення та через деякі неефективні алгоритми не було застосовано до розпізнавання за мовою. Автоматичне розпізнавання віку та статі може покращити взаємодію людини та різних додатків та ігор в телефоні. Це також може допомогти ідентифікувати підозрюваних у кримінальних справах. Нині дослідники почали застосовувати ці алгоритми до деяких реальних сфер як наприклад, розпізнавання мови, комп'ютерний зір, фінанси, банківська справа, робототехніка тощо. Також набір даних включав приклади з реального життя, так що підозрюваних. Деякі інші варіанти використання цієї системи можуть бути застосовано для адаптації музики відповідно до віку та статі людини.

Щоб мінімізувати вплив шуму та завад, та отримати добрі результати розпізнавання віку та статі необхідно використовувати стійкі до завад ознаки, та забезпечити придушення завад до розпізнавання

Мета роботи створити методи визначити статі та віку людини за її мовою, що є стійкими до дії завад.

В роботі для підвищення завадостійкості використовується попередня обробка та вибір для розпізнавання стійких до дій завад ознак.

Також важливимв вирішенні задачі розпізнавання статі та віку є правильний вибір методу класифікації.

1 ОСОБЛИВОСТІ ПІДХОДІВ РОЗПІЗНАВАННЯ СТАТІ ТА ВІКУ ЛЮДИНИ ЗА ЇЇ ГОЛОСОМ

1.1 Огляд відомих методів розпізнавання віку та статі людини за її голосом

Для автоматичного розпізнавання віку та статі людей може бути виконувана вилика кількість різних підходів.

Так, наприклад, Florian та ін. для оцінювання віку використовують кепстральні ознаки, а саме такі як кепстральні коефіцієнти частоти Мела (MFCC). З літератури відомим є те, що при класифікації із різними записаними сигналами, що використовували різні мікрофони застосування ознак MFCC дає погані результати для розпізнавання за статтю та віком. Для того щоб покращити результати та уникнути цієї проблеми, дослідники Cac та ін покращили застосування в якості ознак функції MFCC, експериментально проаналізувавши параметри, які впливають на процес вилучення ознак. Окрім цього додатково вивчався вплив вибору окремих параметрів та ознак на точність розпізнавання статі та віку. Деякі інші дослідники разом з MFCC використовували тонові або інші просодичні функції; завдяки чому покращувались результати розпізнаваня статі та віку (Harnsberger et al.,2008 p.).

Дослідники Голфер і Майкс, 2005 вивчали автоматичну гендерну класифікацію мовних сигналів для дорослих людей за їх голосом, використовуючи функції, які були пов’язані з частотою основного тону (F0) та оцінками перших чотирьох формантних частот. Цей підхід Potamianos and Narayanan в 2003році використали для класифікації трьох класів, включаючи дорослих чоловіків, дорослих жінок та дітей без гендерної дискримінації. Оскільки частота основного тону F0 та частота формантів (особливо другої та третьої форми) у дітей є вищую ніж у дорослих, то діапазон формантів зменшується з віком. При цьому зміни жіночого діапазону трансформуються в межах більшого часу ніж чоловічі, та основні

замітні відмінності стають найбільш значущими після дванадцяти років (Potamianos and Narayanan, 2003).

З метою побудови методів класифікації за статтю та віком можна використовувати різні класифікатори, включаючи моделі гауссових сумішей (GMM), приховані марківські моделі (HMM), регресія опорного вектора (SVR) та багатошарові персептрони (MLP), які зокрема були запропоновані та випробувані Hugo та ін., та результат правильного розпізнавання в його дослідженях для гендерної класифікації составляв 95%. Але в цьому дослідженні була виконана класифікація чоловіків та жінок, без врахування дітей (Hugo та Isabel, 2011). Для побудови методів оцінювання віку дослідники використовували різні класифікатори, що включали в себе векторну машину підтримки (SVM), метод класифікацій k-найближчий сусід (k-NN), імовірнісну нейронну мережу (PNN) та GMM, за яким в результаті було отримано 88% правильно розпізнаних класів (Sedaaghi, 2009). Відомим є використання одношарової нейронної мережі для оцінювання віку дітей, де вони застосували нечіткий синтез даних для прийняття загального рішення (Мірхассані, та ін., 2014). Томас та ін. поєднали в собі особливості трьох методів, включаючи MLP, що були підготовлені з перцептивним лінійним прогнозуванням (PLP) (Thomas, et al., 2014).

Як і інші вирішення прикладних задач розпізнавання образів, ефективність моделі залежить від надійності вибраних ознак та якості класифікаторів, особливост і яких було запропоновано попередніми дослідниками. Робот дослідників в основному зосереджені на використанні в якості ознак кепстральних коефіцієнтів MFCC та частот формантів.

Тому метою даної роботи є пошук завадостійких методів які використовують в якості ознак характеристики частоти основного тону, коефіцієнтів MFCC та формантних частот, які є релевантними ознаками для статевої та вікової класифікації. В роботі необхідно дослідити окремий внесок частоти основного тону, різних коефіцієнтів MFCC та формантних

частот у поліпшенні якості розпізнавання віку та статі. Зокрема особливо необхідно дослідити з точки зору внеску різних діапазонів частот, представлених у їхніх формантах та коефіцієнтах MFCC.

В якості класифікаторів у даній роботі будемо використовувати вектори підтримки SVM та класифікатор k-NN. Класифікатор SVM є одним з найвідоміших та популярних методів класифікації на завдяки його добрій здатності розділяти класи на оптимізовану гіпер-рівнину, так як він максимізує відстань на відстані від гіперплани сепаратора до підтримуючого вектора.

Дані, що використовуються як правило є сукупністю шумних і чистих мовних сигналів, тому необхідна також попередня обробка та ефективний вибір ознак що впливає процес якісного прийняття рішень та є темою для подальшого покращення результатів класифікатора за віком та статтю.

1.2 Особливості моделі мовотворення для різного віку та статі

Для формування окремих звуків мови необхідним є видихання повітря з легенів людини, наявність вібрацій голосових складок для вокалізованих звуків, та рух деяких м’яз, завдяки чому змінюється форма голосового тракту. Зуби, губи та язик приймають дуже активну участь при формуванні звуків мови та мовлення людей в цілому. Систематизований і впорядкований набір звуків мови сприймається людиною як мовлення та розмова в цілому.

Так як сигнал мови є аналоговим за своєю природою, то для того щоб можливо було працювати з інформацією яка міститься в мовних повідомленнях на ПК її необхідно перетворити з аналогової форми в цифрову. Завдяки цьому сигнали мови людини можуть бути легко перетворені з часової в частотну область.

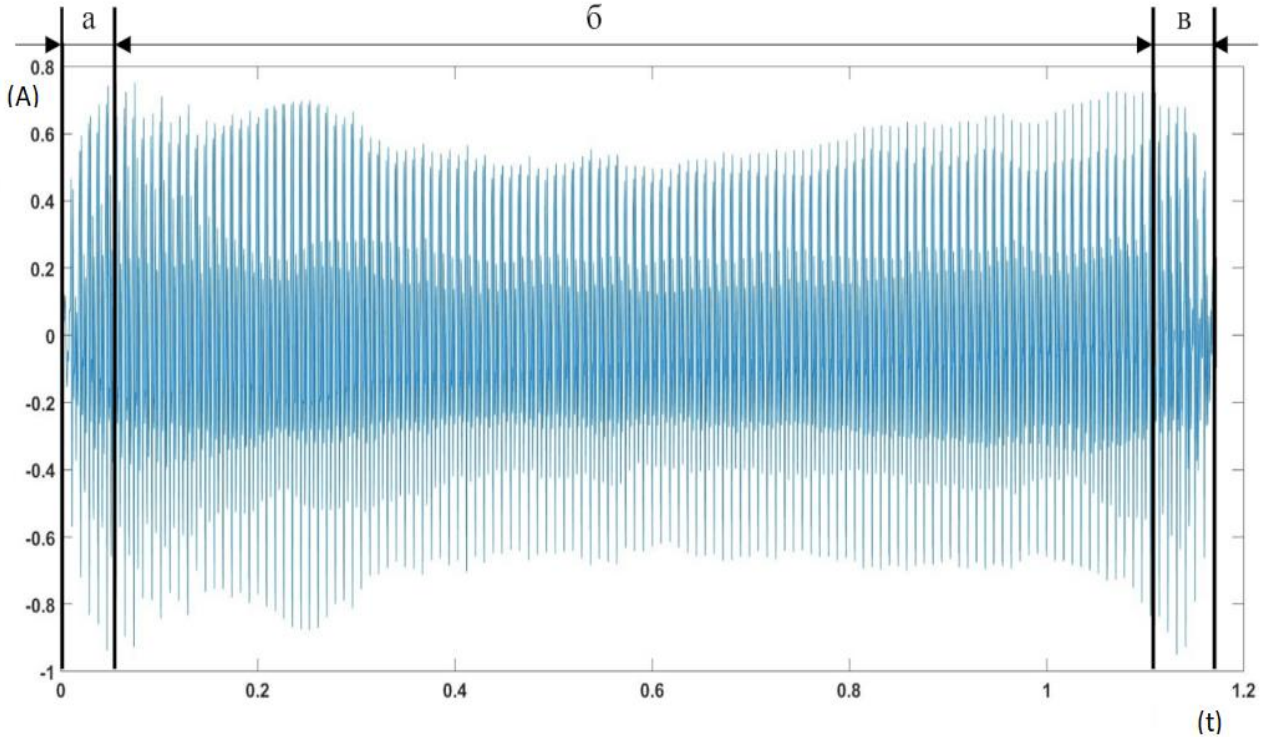

Рисунок 1.1- Фрагмент мовного сигналу

Мовленевий сигнал можливо поділити на три окремих групи, це зокрема:

- дзвінкі звуки мови,

- глухі звуки мови,

- відсутній мовлене вий сигнал (німа мова).

Глухі (невокалізовані) звуки є аперіодичними і шумоподібними. Для дзвінких (вокалізованих) звуків в порівнянні з глухими (невокалізованими) звуками.характерним є притаманна їм квазіперіодичність мовного сигналу та більш висока енергія яка сконцентрована в низькочастотній області. Під час відсутності корисного сигналу мови диктора в каналі зав’язку все рівно присутні шуми, різно характеру, або деякий рівень завад, що утворенні акустичним шумом навколишнього середовища та власними шумами мікрофонів, записуючих приладів.

Характеристики голосу дітей зуттево відмінні від голосу дорослих людей. Їхній голос слабкіше за силою, так як голосові складки дітей коливаються не всією поверхнею, а краями, тобто з меншою амплітудою. Дитячий голос відрізняється тим що частотній діапазон для них знаходиться вище ніж у дорослих, що пояснюється ти, що голосові складки дітей коротше, як і всі інші мовотворні частни організму, що приймають участь в формуванні мови, наприклад голосовий тракт. Тембральне забарвлення голосу дітей слабке що пояснюється не розвиненістю системи резонаторів у дітей. Формування голосу триває протягом тривалого періоду життя людини і зазнає зміни протягом усього життя. Окрім того звуки окремої особи можуть змінюватися із-за вплава зовнішніх факторів на організм та залежать від діяльності залоз внутрішньої секреції, включаючи статеві залози, гіпофіз, щитовидна залоза.

Особливо інтенсивні вікові зміни голосу, як правило, відбуваються в віці 12-15 років. При цьому вікова мутація обумовлена фізіологічними змінами в мовотворчих органах організму, а це зокрема гортань, яка збільшується в розмірі в 1,5-2 рази у чоловіків, та в розмірі на 1/3 у жінок. Зміни відбуваються таким чином що голосові складки збільшуються в розмірі за всими параметрами, включаючи їх довжину, ширину та товщину, та починають коливатися всією своєю масою. Збільшуються також други мовотворні органи включаючи і корінь язика, та при цьому голос підлітка не встигає пристосуватися до швидких анатомічних змін та звучить нестійко. При цьому голос хлопчиків знижується на цілу октаву, у дівчаток - на один або два тони.

Основними причинами зміни голосу підлітків в період мутації голосу є порушення функцій координації зовнішніх та внутрішніх м'язів гортані та відсутності узгодженості між диханням та фонацією підлітка.

1.3 Параметри моделі мовотворення, що є інформативними для розпізнавання віку та статі людини

Дослідження особливостей формування окремих звуків на основі моделі мовотворення, та й мови в цілому є необхідним першим етапом для синтезу автоматичних алгоритмів розрізнення вікових груп дикторів та статей. При цьому необхідно виявити статистичні відмінності в параметрах моделей за голосами мовців, що характерні для різних вікових груп та дослідити відмінності в параметрах мови за гендерною ознакою.

Звуки мови людини формуються таким чином що їх нижня частота 50 Гц. Голосовий тракт людини може формувати звуки мови в широкому діапазоні (існують складові ультразвука які не сприймає людське вухо), але основна енергія сконцентрована в діапазоні від 50 Гц і до 3400 Гц. В цьому діапазоні працюють також телефонні лінії. Для повсякденного спілкування як травило, використовують більш вузький діапазон від 300 Гц до 3000 Гц.

Щоб побудувати алгоритми розрізнення голосів людей за віком використовується така важлива ознака як період або частота основного тону (ОТ). Частота основного тону характеризує вокалізовані звуки, в формуванні який приймають участь голосві звязки Середні частоти та діапазони частот коливання голосових зв'язок, при проголошенні вокалізованих звуків, різняться для чоловіків та жінок, та змінюються з віком . Чоловіки, як правило, формують більш низькі частоти основного тону, ніж жінки. При цьому частоти основного тона динамічно змінюються в процесі формування звуків, формуючи інтонаційний контур (мелодику звуку), тому є моменти часу коли ці частоти співпадають для різних статей. Тому діапазони частот оновного тону для різних класів перетинаються, а в деяких випадках жіночому голосу може відповідати менша частота основного тону ніж у чоловіків.

Для чоловіків характерним є те, що в основному частота основного тону лежить в діапазоні від 85 гц до 155 гц. Частота основнного тону для дорослих жінок в середньому вища ніж для чоловіків та знаходиться в діапазоні від 165 гц до 255 гц. Частота основного тону пов’язана з розмірами голосових зв'язок, а тому діти, коли вони розмовлють, мають набагато більш високі частоти от. Для немовляти частота основного тону, як правило, змінюється від 250 гц до 650 гц. В середньому для десятирічного хлопчика або дівчинки частота основного тону приблизно 400 гц. Якщо людина розмовляє в природному голосі то динамічно змінюються параметри мовотворення і зокрема частота коливань голосових зв'язок, що призводить до зміни частоти основного тонну. Коли людина через структуру мови та емоційне піднесення змінює інтонацію або ритм мовлення то її частота основного тону також бистро змінюється. Але змінюється не тільки одна частота основного тону людини, а й другі параметри в мовотворенні.

Таблиця 1.1 – Вікові особливості мінімального, середнього та максимального значень частоти основного тона

Вік людини	Частота ОТ для чоловіків	Частота ОТ для жінок
8	182/255/422 Гц	181/281/419 Гц
15	95/124/248 гц	179/228/303 гц
20-39	87/127/189Гц	158/233/335 Гц

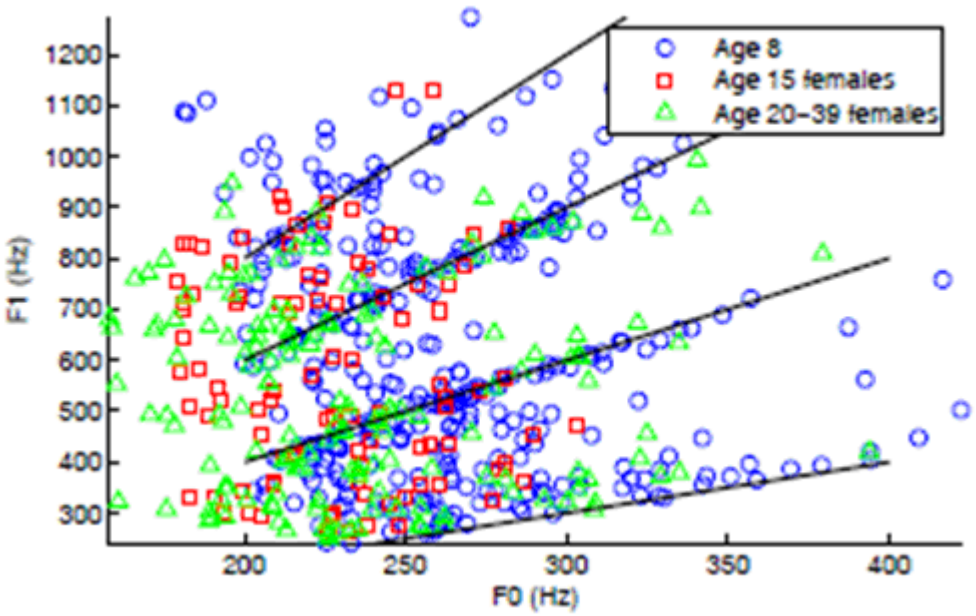

Рисунок 1.2 – Лінійні роздільні межі для компактного зосередження част формантів F1 та частоти основного тона F0 для дитини віком 8 років(позначка синій круг) , для дівчини 15 років (позначка червоний квадрат) та жінки 20-39 років (позначка зелений трикутник)

2 МЕТОДИ ПОПЕРЕДНЬОЇ ОБРОБКИ МОВНИХ СИГНАЛІВ

2.1 Особливості попередньої обробки

Для поліпшення якості розпізнаваня мови виконується попередня обробка сигналів мови, що здійснюється за допомогою ряду алгоритмів. Якість мови означає, що вона представлена чітко, зрозуміло та приємна для сприйняття. Тому використання методів попередньої обробки сигналів мови є дуже важливою для побудови методів автоматичного розпізнавання статі та віку. Так як у реальному світі присутні різнорідні всілякі шуми то без використання завадостійких методів з застосуванням попередньої обробки мови розпізнавання статі та віку стає неможливим здійснити в реальному світі. Можуть бути наявними різнорідні акустичні шуми та завади включаючи вуличний шум, також це може бути розмова других людей на задньому плані, а також може це може бути шум від музикального або іншого інструменту. Тому видалення цих фонових шумів настільки важливо що алгоритми попередньої обробки мови використовують в даний час для побудови багатьох інших методів розпізнавання мови. Яскравими прикладами його використання є відновлення історичних записей розмов, лекцій, інтерв'ю, для поліп нень прослуховування кримінальних розслідувань, а також для покращення якості звуку слухових апаратів для людей похилого віку, що використовується в Європі. З усіх цих причин для ефективного розпізнавання віку та статі будуть використані деякі методи попередньої обробки мови, як описано раніше.

Показник сигнал/шум характирезує мовний сигнал та є відношенням потужності реального звукового сигналу до шуму, виданого пристроєм запису. Кожен пристрої, включаючи мікрофон підсилювач, записуючий пристрій має в ньому якийсь шум (включаючи і теплові).

У багатьох випадках шум можно вирівняти, але в випадку якщо вхідний звуковий сигнал дуже слабкий то навіть невеликий шум від пристрою може спричинити сильні спотворення. Якість мови залежить від співвідношення сигнал-шум, чи воно вище тим краща якість.

Ввідношення сигнал-шум прийнято вимірювати в дБ, що суттєво полегшує роботу з великими числами.

Відношення сигнал-шум (SNR) можна виміряти, якщо використовувати наступне співвіднoршення (рис 2.1)

$$SNR = \frac{P_{signal}}{P_{noise}}$$

де Psignal - середня потужність мовного сигналу, а Pnoise - середня потужність завади або шуму.

В логарифмічному масштабі в dB можна виразити виразити за формулою :

$$SNR_{dB} = 10\ log_{10}(\frac{P_{signal}}{P_{noise}})$$

На рис. 2.1 наведено приклад фрагменту мовного сигналу із доданим до нього шуму, де Pn / Ps = 0,1.

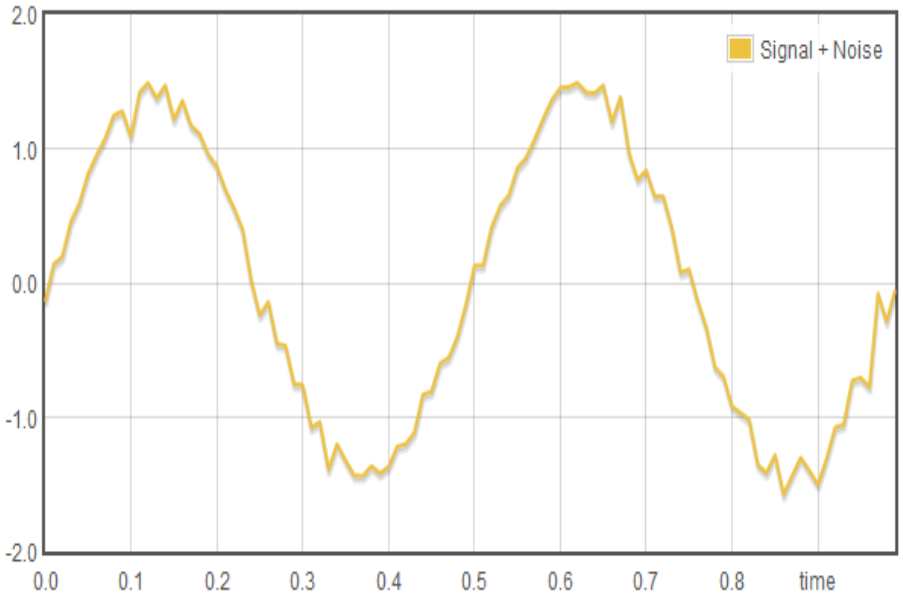

Рисунок 2.1 - Приклад мовних сигналів зі співвідношенням сигнал / шум 0,1 (Pn / Ps = 0,1)

2.2 Особливості віднімання спектрального шуму

Одним з найбільш широко застосовуваних на практиці методів видалення мовного шуму є віднімання спектрального шуму, що необхідно в випадку коли мова містить реальні звукові дані та деякий адитивний фоновий шум. Цей метод ефективно працює в частотній області в випадку коли відомо , про сптектр мовного сигналу та спектр фонового шуму. На рис.1.2 приведена структурна схема для методу віднімання спектрального шуму. Спочатку виконується оцінювання спектр шуму з мовних зразків, де присутній лише один шум, а потім він віднімається від зачумленого мовного спектру, з метою отримати чистий мовний спектр сигналу.

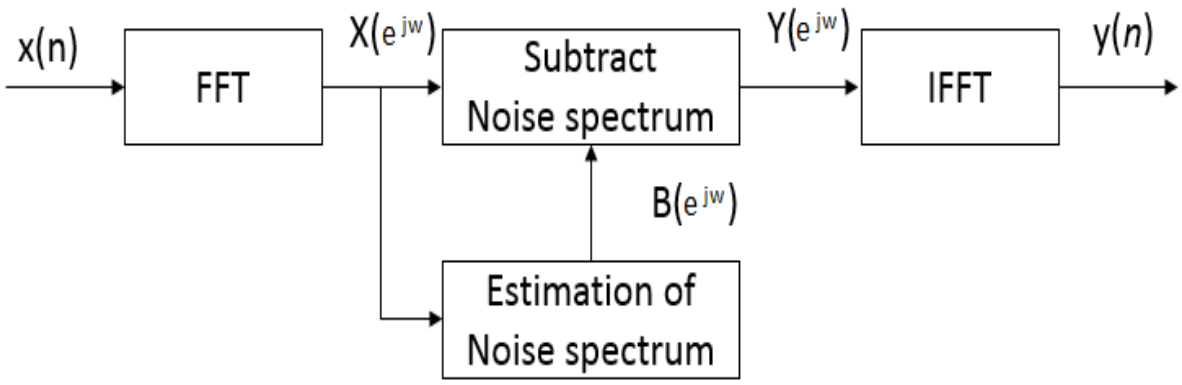

Рис 2.2

Рисунок 2.2 - Метод спектрального віднімання шуму

В схемі адитивна суміш описується виразом:

$$x(n) = s(n) + b(n),$$

де s (n) - чистий мовний сигнал, а b (n) – фоновий сигал шуму.

Ціллю спектрального віднімання є знаходження чистого мовного спектра шляхом віднімання середнього значення спектру шуму N [n] від вхідного спектру X [n], відповідно до формули:

$$|\hat{S}[n]|^b = |X[n]|^b - \overline{|N[n]|^b}$$

При цьому ля показника b може бути два різних значення. Якщо параметр b = 1, то це є величина спектрального віднімання за величиною, а коли b = 2, то це є величина для спектрального віднімання потужності. У випадку коли величина чистого мовного спектру окажеться негативною, тор тоді в цьому випадку результат прирівнюють до нуля.

тому з вразуванням зазначенного рівняння змінюється на вид:

$$|S[n]|^b = \begin{cases} |X[n]|^b - \overline{|N[n]|^b} & if\ |X[n]|^b - \overline{|N[n]|^b} > 0 \\ 0 & otherwise \end{cases}$$

Далі находять зворотне перетворення Фур'є, звдяки чому знаходять оцінку чистого мовного сигналу.

2.3 Нормалізація середнього значення мовного сигналу

Мовні сигнали, які записано з використанням різних типів мікрофонів в одному і тому ж акустичному середовищі, може мати різні частотні, амплітудні характеристики, так як кожен тип мікрофону має різні амплітудні та частотні характеристики зі своєю функцією передачі. Та навіть в тому вивадку коли однаковий тип мікрофонів використовується в одному акустичному середовищі запису, то все рівно звуки мови, що записано з різних мікрофонів, можуть мати різні звукові характеристики. Та це можливо також якщо якщо один і той же мікрофон знаходяться на різній відстані від диктору. При застосуванні того ж мікрофона в іншому примішенні із за зміши акустичних властивостей приміщення характеристики будуть відрізнятися.

CMN (нормалізація) є методом попередньої обробки що намагається вирішити цю актуальну проблему. Використовуючи мовні висловлювання кожного кадрів цей метод знаходить середні центровані вектори мови.

Алгоритм CMN описується за формулою:

$$y[n] = x[n] - \frac{1}{N}\sum_{n=1}^{N} x[n]$$

де відліки x [n] і y [n] –це поточні відліки вхідного та вихідного сигналу, які змінюються за часом до і після фільтрації, а N - дорівнює загальній кількість кадрів.

2.4 Особливості RASTA фільтрації

RASTA (відносна фільтрація спектрів) – це метод попередньої обробки, який використовується для видалиння шуму каналу з мовного звукового сигналу. Сигнали мова набагато сильніше змінюються ніж шум у каналі передачі. Тому з метою усунення цього шуму каналу з мовного сигналу, в частотній області може застосовуватися смуговий фільтр. За допомогою такого смуговововогоий фільтра можна видалити повільні зміни (що можно спостерігати на спектрограх) або постійні величини в мовному сигналі.

Функцію передачі фільтру RASTA можна придставити в виді:

$$H(z) = 0.1z^4.\frac{2 + z^{-1} - z^{-3} - 2z^{-4}}{1 - 0.98z^{-1}}$$

2.5 Застосування градчастого фільтру для захист від завад

Особливістю цього методу є те що коефіцієнти відбиття, які є основким параметром цього фільтру, обчислюють рекурентно без попереднього оцінювання кореляційної функції мовного сигналу.

За результатом обчислень помилок передбачення $d^{m}{}_{n}$, $b^{m}{}_{n}$ обчисляємо рекурентно кофіцієнти відбиття за формулою :

$$k_m = \frac{\sum_{n=1}^{N}(d^{m-1}{}_{n} \cdot b^{m-1}{}_{n-1})}{(\sum_{n=1}^{N}(d^{m-1}{}_{n})^2 \cdot \sum_{n=1}^{N}(b^{m-1}{}_{n-1})^2)^{1/2}} \quad , (2.2)$$

де значення m = 1,2, ..., p.

При використанні для обчислення коефіцієнтів відбиття методу Бурга отримаємо формулу:

$$k_m = \frac{2 \cdot \sum_{n=1}^{N} (d^{m-1}_n \cdot b^{m-1}_{n-1})}{\sum_{n=1}^{N} (d^{m-1}_n)^2 + \sum_{n=1}^{N} (b^{m-1}_{n-1})^2} \quad . (2.3)$$

Вибираємо начальні значення як:

$$d^{0}_n = x_n, \quad b^{0}_n = x_n, \quad n = 1, ..., N.$$

При цьому помилка прямого передбачення обчисляється наступним чином:

$$d^{m}_n = d^{m-1}_n - k_m \cdot b^{m-1}_{n-1}$$

Помилка зворотного передбачення знаходиться за формулою:

,

$$b^{m}_n = b^{m-1}_{n-1} - k_m \cdot d^{m-1}_n.$$

Пристрій для обчислення коефіцієнтів відбиття приведено на рис. 2.3.

Результат роботи фільтра з придушення завад буде

$$y_n = d^{p}_n .$$

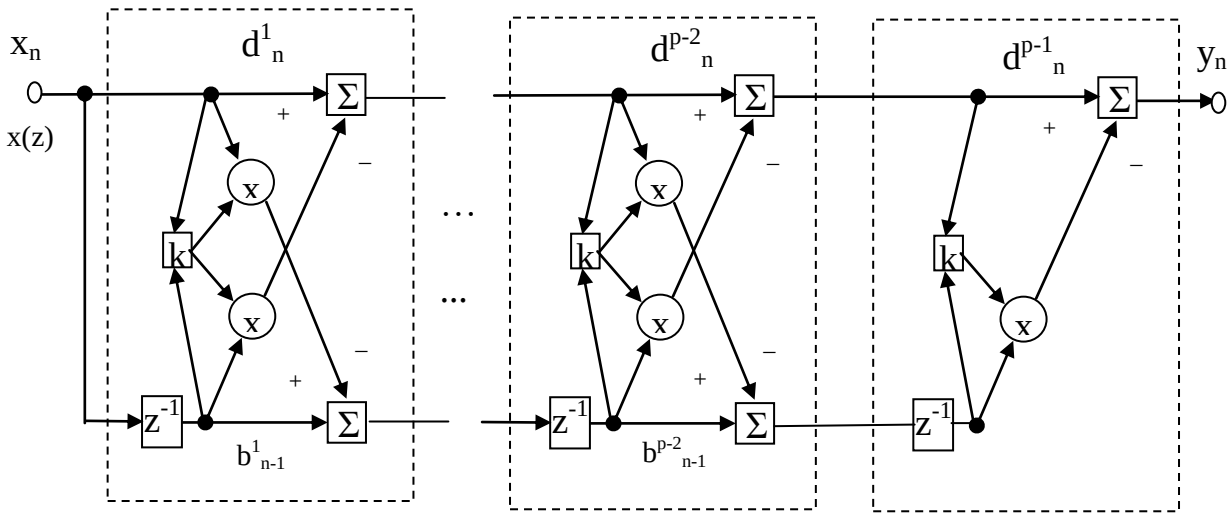

Рисунок 2.3 – Гратчатий фільтр для оцінювання

Після навчання в паузі сигналу раніше знайдені коефіцієнти відбиття $\vec{k} = (k_1, k_2, ..., k_p)$ для цифрового градчастого фільтра застосовують для придушення завад мовного сигналу фільтром, який приведений на рис.2.4.

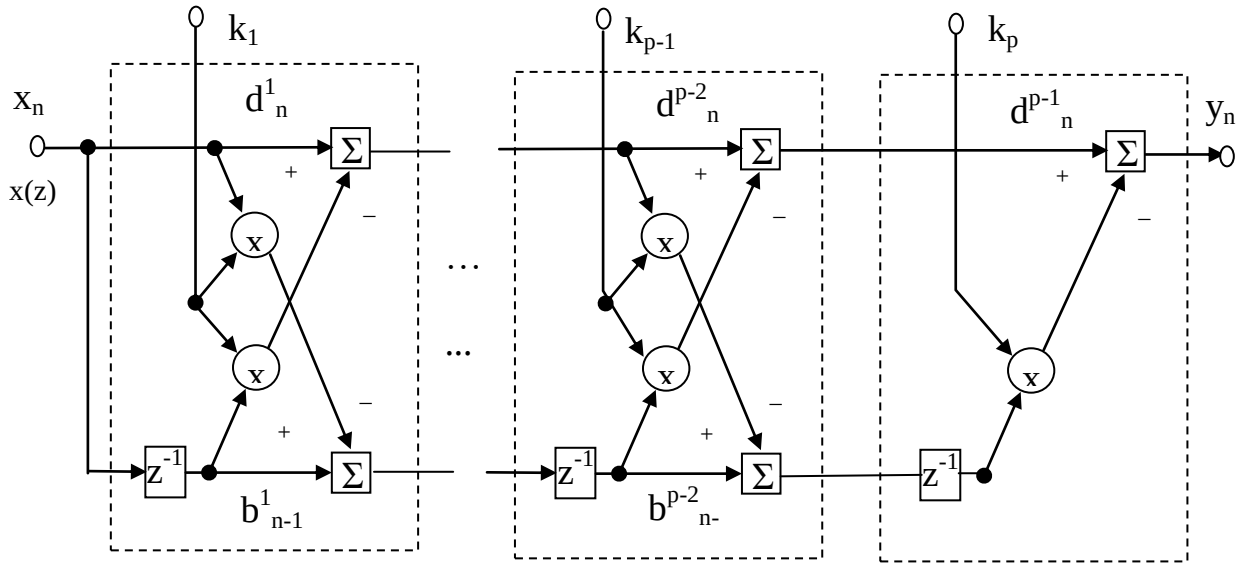

Рисунок 2.4 – Гратчастий фільтр для захисту від завад

3 МЕТОДИ РОЗПІЗНАВАННЯ СТАТІ ТА ВІКУ ЛЮДИНИ

3.1 Підготовка до оцінювання вектору ознак

Спектр сигналу мови для кожного сегменту має більш високий рівень енергії на нижніх частотах а ніж на високих, тому це явище називають спектральний нахил. Наявність спектрального нахилу обумовлено особливістю характеру глоттального імпульсу. Оскільки сигнал на більш високих частотах має менше енергії порівняно з низькочастотними сигналами то продуктивність розпізнавання покращується при посиленні високочастотної енергії, та надає для акустичної моделі більше інформації і. Тому перший етап процесу є попереднім наголосом.,

Різницеве рівняння, що описує вихідну реакцію попереднього фільтра має вид:

$$y(n) = x(n) - 0{,}95\ x\ (n- 1)$$

Мовний сигнал поділяється на окремі вибірки (кадри) після застосування фільтру (3.1). Мовний сигнал повинен бути квазістаціонарним в межах вибраних вибірок (кадрів). Кожен знайдений кадр містить в собі N вибіркових точок мовного сигналу. Можливий варіант вибору ширини кожної з вибірок, наприклад, це вибирають близько 30 мс з перекриттям їх близько 20 мс (змінна частина всього 10 мс). Час накладання кадрів друг на друга повинен становити від 30% до 75% довжини цих вибірок, M = 100 і N = 256, тому ми розподіляємо вибірку сигнала на 20-40 мс. В випадку коли вибірка є набагато коротшою, то у нас не буде достатньої кількості інформації для якісної оцінки ознак, а якщо він буде занадто довгий, то сигнал буде сильно змінюватися в межах кожної вибірки. Тому щоб отримати оптимальну достовірну спектральну оцінку необхідний її правильний вибір.

При цьому враховується те, що хоча сигнал мови є нестаціонарним, але він дозволяє здійснити якісну оцінку протягом невеликого інтервалу часу.

3.2 Особливості обчислення MPCC

Етапи обчислення MPCC наведено на рис. 3.1.-

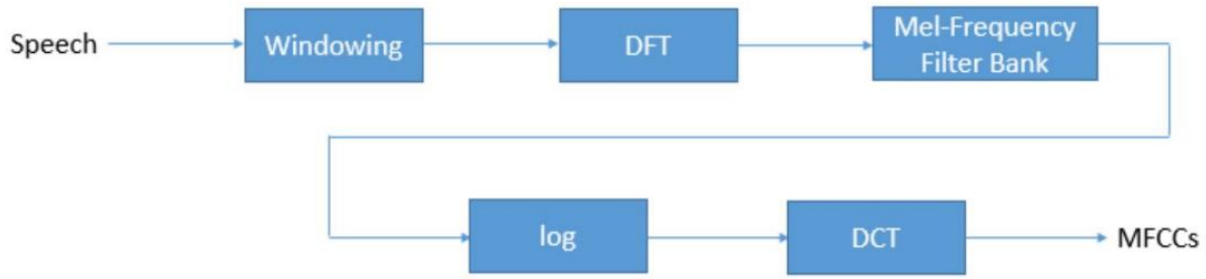

Рисунок 3.1 - Етапи обчислення MPCC

Першим кроком у обчисленн MPCC є застосування вікна.

Функції заходження віконна необхідні для згладжування сигналу до обчислення дискретного перетворення Фур'є (DFT). При цьому існує вилика кількість різних вікон, кожне з яких має свої особливості. Одими з найбільш вживаних залишаються вікна Хеммінга, Блекмена, Гаусове, прямокутне та трикутне. Прямокутне вікно дуже часто використовується в мовленнєвому аналізі, оскільки воно досить просте, хоча має широкий спектр.

Більш ефективним є вікна Хеммінга, спектр якого спадає досить швидко, а тому воно краще підходить для виявлення формантів. Рівняння вікна Хеммінга в часовій області має вид:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{M}\right)$$

де M – визначає довжину вікна, n – дескретний час.

Вікно та результати застосування вікна Хеммінга до звукового сигналу наведене на рисунку 3.2.

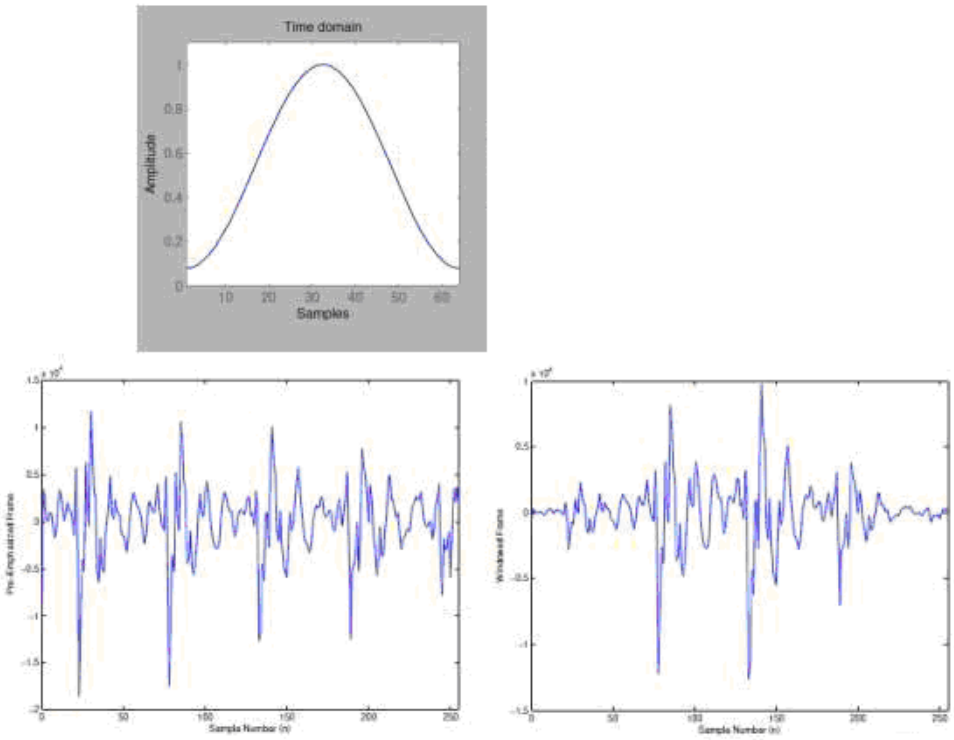

Рисунок 3.2 – Вікно та вхідний звуковий сигнал , та результат застосування вікна Хеммінга

Для кожного блоку, що включено N відліків за якими виконується перетворення з часової області в частотну область. Для цього застосовується перетворення Фур'є. Для цьогое перетворення використовується наступний вираз:

$$F(x) = \sum_{n=0}^{N-1} f(n).\, e^{-j2\pi(x\frac{n}{N})}$$

де N – кількість відліків дискретного сигналу, що визначає розмір FFT, який може становити 512, 1024 або 2048, F (x) – результат перетворень сигналу в частотну область, f (n) – відліки сигналу вчасовій області.

$$yt(k) = \sum_{m=1}^{M} \log(|Yt(m)|) \cos\left(k(m-0.5)\right)\frac{\pi}{M} \ , k = 0, \ldots . J$$

Людське вухо менш чутливе до більш високих звукових частот, а це приблизно буде більше 1000 Гц. Окрім того наше сприйняття слуху нелінійне. Ралізований банк Mel -фільтрів враховує особливості сприйняття людського вуха , та є сукупністю фільтрів, шкала частот яких концентрується лише на певних змінних частотних полосах. Залежність Mel –шкали від рівномірної для частотної області наведена на на рис.3.3.

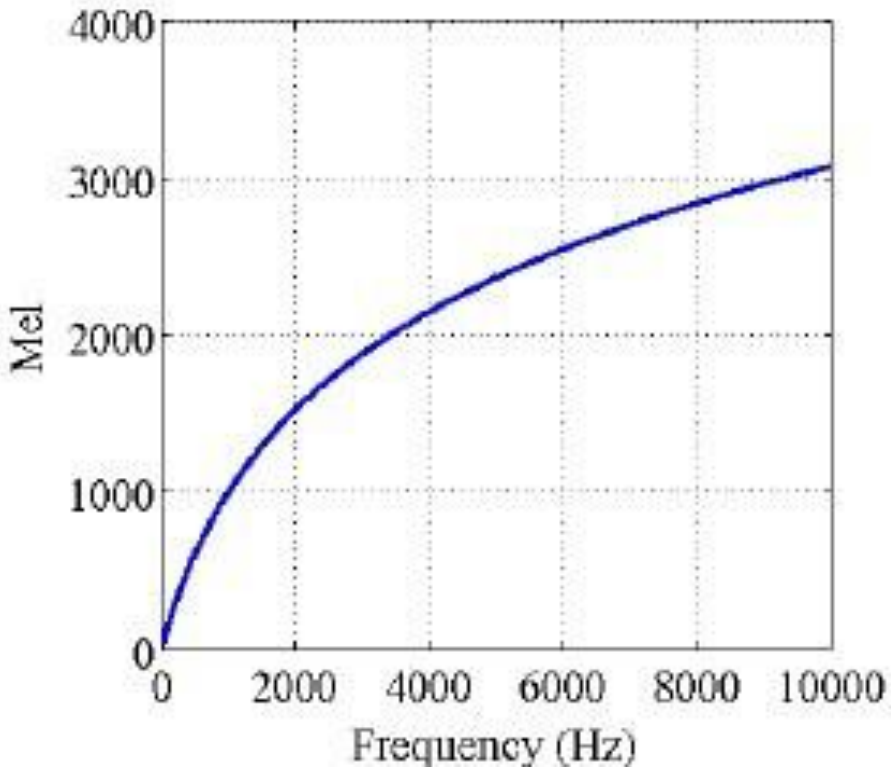

Рисунок 3.3 –Перерахунок лінійної шкали частот в частоти за шкалою Мела

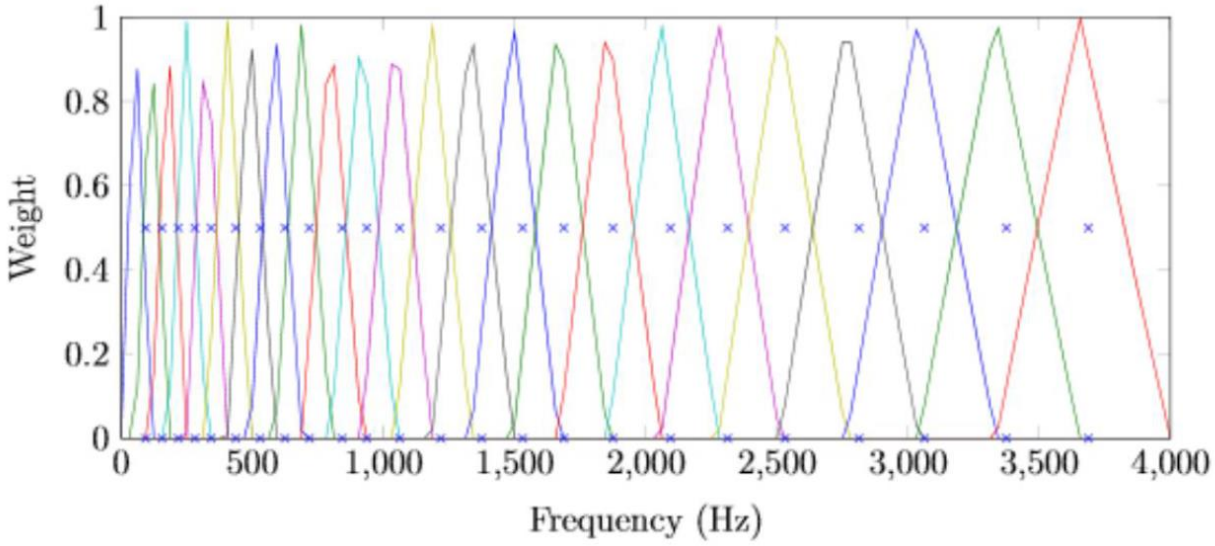

Рисунок 3.4 – Банк Мел фільтрів для 24 полос

На рис 3.4 наведено для шкали за рис.3.3 банк Мел фільтрів для 24 полос.

Особливості чутливості слуху людини в тому, що від адаптивно змінює свою чутливість, при цьому він менш чутливий до незначних фіксованих змін у амплітудах для великих загальних амплітудах, чім при менших амплітудах з тими ж самими значеннями змін вхідних даних, а це приводить до того що реакція людини на сигнали – логарифмічна. На рис. 3.5 наведено сигнал до і після використання логарифму:

Так як реакція людини на сигнали – логарифмічна, то далі динамічний діапазон значень стискається з застосуванням функції логарифму. В наступному кроці обчислюється логарифм квадрату величини з виходу банку Mel –фільтра.

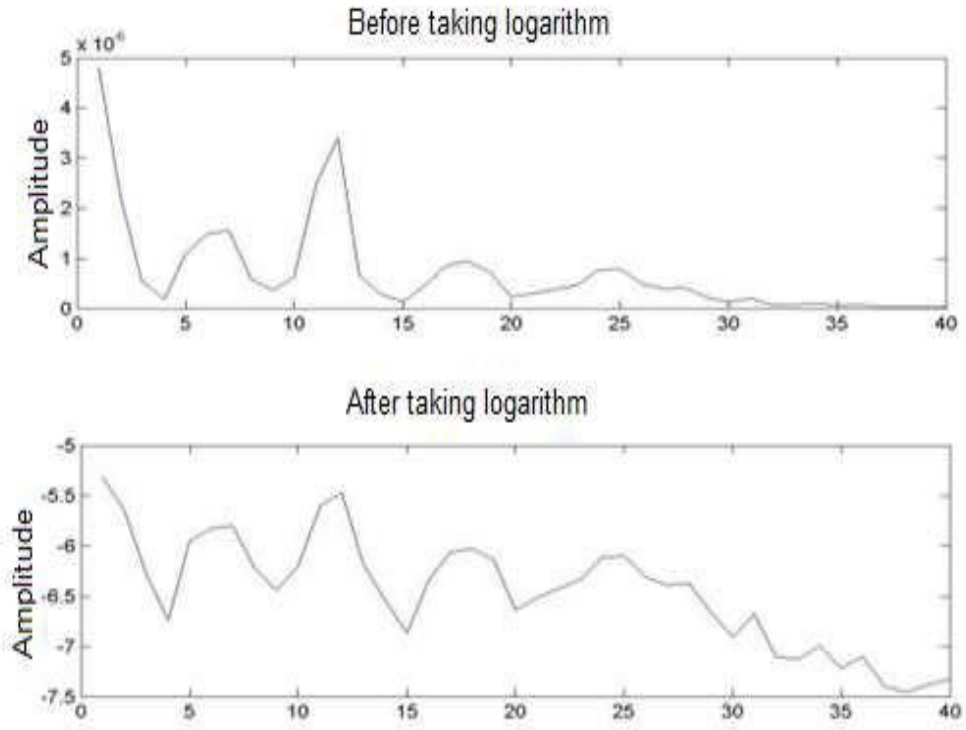

Рисунок 3.5 - Амплітуда мовного сигналу до і після обчислення логарифму

На останньому етапі алгоритму MFCC застосовується дескретне косінусне претворення Фурє (DCT). Таким чином відбувається завдяки цьому процес перетворення спектру сигналі в методі Mel в часову область за допомогою DCT. Результат такого перетворення називаються кепстральними коефіцієнтами Mel частоти, а набір таких коефіцієнтів називають акустичними векторами.

Завдяки застосуванню DFT ми перетворюємо сигнал з частотної області назад у часову область відповіно до наступної формули:

$$yt(k) = \sum_{m=1}^{M} \log(|Yt(m)|) \cos\left(k(m-0.5)\right)\frac{\pi}{M} \ , k = 0, \ldots . J$$

3.3 Особливостіоцінювання формантних частот

Окремі форманти це згладжені вершини спектру сигналу мови.

З точки зору фізичної трактовки, формантні частоти відповідають акустичному резонансу та оцінюються як пік амплітуди в сглаженому частотному спектрі звукового сигналу. На рис.3.6 наведено алгоритм оцінювання формантних частот, що застосовує використання лінійних коефіцієнтів прогнозування (LPC).

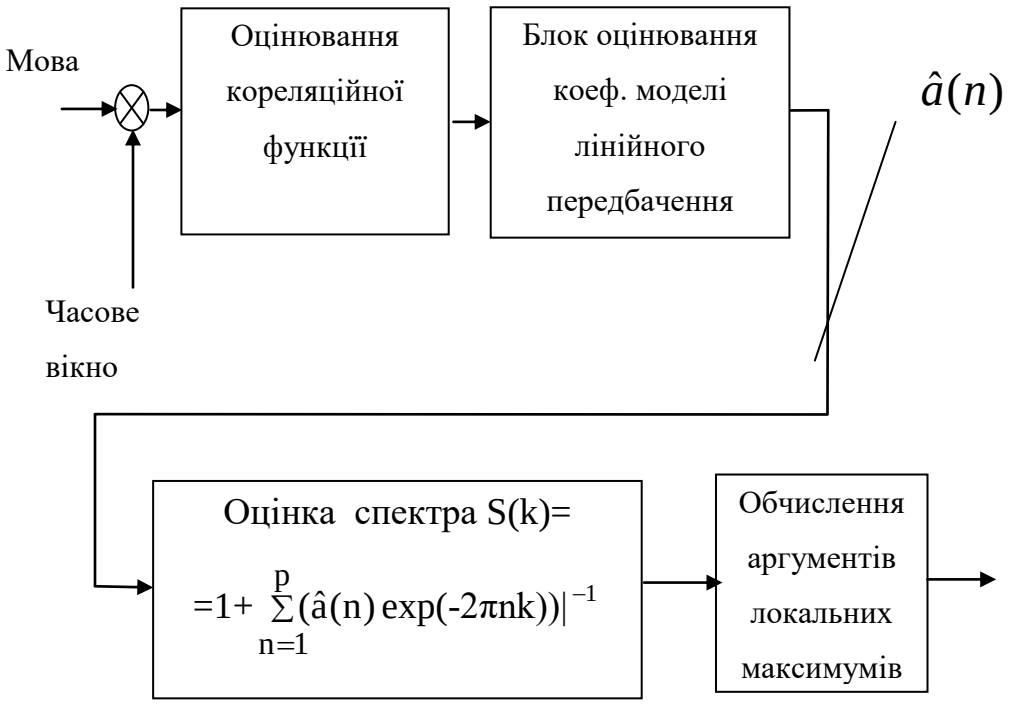

Рисунок. 3.6 – Структурна схема оцінювання формант

3.4 Побудова методу опорних векторів SVM для розділення класів

Особливістю лінійного методу SVM, є те що для випадку тільки двух класів та лінійної межі необхідно утворіати лінійні роздільні межі між множинами відомих еталонів (це наприклад двох статей або при бінарному поділі людей за віком).

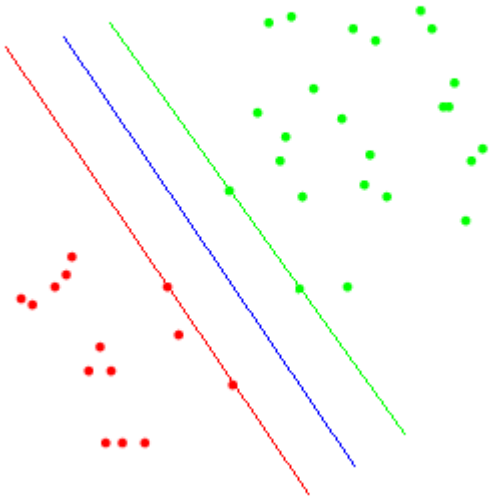

Рисунок 3.7 – Побудова роздільної межі між класами для лінійного методу SVM

За цим алгоритмом метод машинного навчання SVM присвоює позитивний або негативний знак в залежності від місця знаходження навчальних елементів безлічі точок відносно межі двох підмножин. Використання цього алгоритм дозволяє знайти родільню поверхню між точками двух множин. В термінах SVM такі точки називаються опорними векторами.

Від вибіра ядра SVM залежать основні характеристики класифікатора людини за віком та статтю.

Для класифікатора SVM справедливим є формула:

$$h(x) = sign(u(x))$$

$$u(x) = \sum_{i=1}^{m} \lambda_i y_i K(x_i, x) - w_0$$

де K(xi,x) - ядро перетворення; w0- значення порогу; параметр λ i- коефіцієнт; xi, yi- пара з навчального набору (при цьому вектори xi, для яких λ i>0 називаються опорними).

Відображення вхідного вектора в простір більш високої розмірності здійснюється за допомогою функції ядра. Завдяки відображенню вхідного вектору в простір більш високої розмірності, машина опорних векторів знаходить гіперплоскость. Ця гіперплоскість розділює класи в просторі ознак з максимально можливим для цього метода зазором між точками різних класів.

Самим першим SVM було створено лінійний класифікатор, який міг вирішувати завдання тільки з лінійною межою розділення. Алі не всі задачі можливо вирішити застосовуючи такий лінійний класифікатор.

У випадку використання нелінійного ядра можливо відобразити вихідні дані в простір більшої розмірності. При застосуванні нелінійного ядра в SVM буде існувати більш точна складна нелінійна роздільна поверхність.

При цьому для функції ядра повинні виконуватися спеціальні умова [2],

Відомі ряд ядер які були застосовувані при вирішенні задач класифікації та показали свою ефективність. Серед них відомими є наступні.

Лінійна функція ядра має вид:

$$K(x_i, x) = x_i^T x,$$

Полиномиальне ядро (поліном ступеня k) обчислюється наступним чином:

$$K(x_i, x) = \left(x_i^T x + c\right)^k.$$

Гаусівське ядро (Gaussian kernel) має вид:

$$K(x_i, x) = exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right).$$

де параметр σ =const.

Параметри класифікатора знаходять на етапі навчання SVM.

Це завдання можливо вирішувати з застосуванням різні методів.

Відомим є методом, що може бути використано це є метод невизначених множників Лангранжа.

Задача квадратичної оптимізації може бути знайдена методом невизначених множників Лангранжа що сформульовано наступним чином:

$$L(\lambda) = \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j K(x_i, x_j) \to \max_{\lambda}$$

при запровадженні обмеження в виді

$$\begin{cases} \sum_{i=1}^{l} \lambda_i y_i = 0 \\ 0 \le \lambda_i \le C \ ; \ C = const \end{cases}$$

3.5 Розпізнавання за методом UBM

Створення супервекторів GMM складається з кількох частин: спочатку універсальна фонова модель (UBM) навчається на великій базі даних мовлення, потім виконується адаптація GMM для кожного мовця, щоб отримати специфічна модель динаміка, на основі якої створюється супервектор GMM.

Модель UBM — це модель модель Гауссовых суміщів (GMM), яка навчена на акустичних характеристиках мел-частотних коефіцієнтів кепстра (MFCC), витягнутих із мовних висловлювань, сегментованих із записанихдля сеансів мовлення. Модель GMM визначається набором параметрів λ, що визначають ваги w_i , засоби μ_i і коваріаційні матриці Σ_i M гаусових компонент ($1 \le i \le M$), які створюють разом функцію ймовірності набору спостережень. На етапі навчання виконується алгоритм EM

(expectation-maximization) для оцінки параметрів GMM. Компоненти Гауса можна вважати моделюючими основними широкими фонетичними звуками, які характеризують голос людини. Оскільки GMM визначає функцію розподілу, вона використовується для оцінки ймовірності нового спостереження o_t , який у нашому випадку є вектором ознак MFCC, що відповідає мовному кадру в момент часу t.

Формула оцінки ймовірності одного вектора ознак:

$$p(o_t|\lambda) = \sum_{i=1}^{M} w_i p_i(o_t) \tag{3.1}$$

де $p_i(o_t)$ це ймовірність, отримана за допомогою i^{th} Гаусс розраховується таким чином:

$$p_i(o_t) = \frac{1}{(2\pi)^{D/2}\sqrt{|\mathbf{\Sigma}_i|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)} \tag{3.2}$$

де D — розмірність вектора ознак.

Процес навчання здійснюється за допомогою ітераційного алгоритму, який включає два кроки на ітерацію: оцінку та максимізацію.

Етап оцінки полягає в обчисленні ймовірності приналежності до кожного компонента всіх навчальних вибірок.

Розрахунок такий:

$$p(o_t|\lambda_i) = \frac{w_i p_i(o_t)}{\sum_{i=1}^{M} w_i p_i(o_t)} \tag{3.3}$$

Потім у максимізації параметри моделі оновлюються.

Компонент означає наступне:

$$\hat{\mu}_i = \frac{\sum_{t=1}^{n} o_t p(o_t|\lambda_i)}{\sum_{t=1}^{n} p(o_t|\lambda_i)} \tag{3,4}$$

Коваріації оновлюються таким чином:

$$\hat{\sigma}_i^2 = \frac{\sum_{t=1}^{n} (o_t - \hat{\mu}_i)^2 \, p(o_t | \lambda_i)}{\sum_{t=1}^{n} p(o_t | \lambda_i)} \tag{3,5}$$

І ваги:

$$\hat{w}_i = \frac{1}{n} \sum_{t=1}^{n} p(o_t | \lambda_i) \tag{3,6}$$

Ці кроки повторюються до досягнення збіжності або максимальної кількості кроків.

Промови, які використовуються для навчання UBM, повинні проводити диктори, рівномірно розподілені за віком і статтю. Модель UBM також називають «світовою моделлю», оскільки вона представляє великий і різноманітний набір дикторів.

Після того, як модель UBM побудована, оцінка MAP (maximum aposteriori) використовується для адаптації моделі UBM для представлення моделі конкретного мовця. Адаптація виконується за допомогою функцій MFCC, отриманих під час сеансу доповідача, як показано на малюнку 3-4. Формули такі:

$$\bar{\mu}_i = \frac{\sum_{t=1}^{n} p(o_t | \lambda_i)}{\tau + \sum_{t=1}^{n} p(o_t | \lambda_i)} \hat{\mu}_i + \frac{\tau}{\tau + \sum_{t=1}^{n} p(o_t | \lambda_i)} \mu_i \tag{3,7}$$

де τ - це зважування апріорних знань до адаптованих мовних даних. Адаптація MAP виконується лише на основі гаусових середніх, залишаючи їхні ваги та коваріаційну матрицю незмінними.

Побудова супервектора

Кожна модель GMM представлена супервектором GMM v, утвореним об’єднанням усіх середніх Гаусса:

$$\mathbf{v} = \begin{vmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \dots \\ \boldsymbol{\mu}_M \end{vmatrix} \tag{3,8}$$

Де $\boldsymbol{\mu}_i$ є середнім вектором i^{th} гаусівський. Навчальні супервектори формуються з використанням моделей GMM, адаптованих до MAP. У базовій системі супервектори використовуються як вектори ознак.

4 РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ ЯКОСТІ РОЗПІЗНАВАННЯ САТІ ТА ВІКУ

4.1 Результати експериментальних досліджень де використовується в якості вирішую чого правила метод KNN (K-Найближчих сусідів)

В проведенних еЕкспериментальних дослідженях якості розпізнавання віку та статі людини за її голосом використовували 40 зразків чистої мови англійської мови. Для випадку використання класів: літні, дорослі та діти обох статей якість правильного розпізнавання статі досягла 66%, а . правильне розпізнавання віку людини досягало 55%.

Таблиця 4.1 – Показник правильного розпізнавання в процентах за віком та статю тільки при використанні ознак MFCC

Ознаки	Процент правильного розпізнавання за статю	Процент правильного розпізнавання за віком
MFCC 1 2 3	55,55%	44,44%
MFCC 4 5 6 7 8 9	66,66%	44,44%
MFCC 10 11 12	22,22%	33,33%
Усі MFCC	66,66%	44,44%

Таблиця 4.2- Проценти правильних розпізнавань людей одночасно за віком та статтю (вікове та стате злиття) лише за допомогою ознак MFCC

Ознаки	Процент правильного розпізнавання за віком і статтю разом
MFCC 1 2 3	33,33%
MFCC 4 5 6 7 8 9	44,44%
MFCC 10 11 12	11,11%
Усі MFCC	44,44%

При використанні в якості ознак коефіцієнтів MFCC найкращий результат було отримано за допомогою 4-9 коефіцієнтів MFCC.

Наступні результати, що наведено таблицях - це результати, які отримано за використанням в якості ознак частот формант.

Таблиця 4.3 – Результати розпізнавання віку та статі людини лише за допомогою форматних частот

Використанні форманти	Оцінка правильного розпізнавання віку в процентах	Оцінка правильного розпізнавання статі в процентах
F3, F4	22,22%	22,22%
F3	38,88%	27,77%
F4	27,77%	44,44%
F1, F4	38,88%	55,55%
F1, F2, F3	50%	55,55%

Продовження таблиці 4.3 – Результати розпізнавання віку та статі людини лише за допомогою форматних частот

Використанні форманти	Оцінка правильного розпізнавання віку в процентах	Оцінка правильного розпізнавання статі в процентах
F1, F2, F4	44,44%	50%
F1, F2	55,55%	61,11%
F1, F3, F4	33,33%	44,44%
F1, F3	50%	55,55%
F1	38,88%	38,88%
F2, F3	55,55%	50%
F2, F3, F4	38,88%	50%
F2, F4	50%	38,88%
F2	38,88%	38,88%
F1,F2, F3, F4	50%	66,66%

Таблиця 4.4 - Результати сумісного розпізнавання які об’єднують вік та стать людини лише за допомогою одніх формантних частот

Використанні форманти	Оцінки правильного розпізнавання віку та статі людини за голосом людини разом в процентах
F1, F2, F3, F4	50%

Продовження таблиці 4.4 - Результати сумісного розпізнавання які об’єднують вік та стать людини лише за допомогою одніх формантних частот

Використанні форманти	Оцінки правильного розпізнавання віку та статі людини за голосом людини разом в процентах
F1, F4	33%
F1, F3	39%
F1, F2	55%
F1	28%
F1, F2, F3	39%
F1, F2, F4	39%
F1, F3, F4	28%
F2, F3, F4	33%
F2, F4	33%
F2	33%
F4	17%
F3	22%
F2, F3	39%
F3, F4	22%

Таблиця 4.5- Процент правильного розпізнавання віку та статі людини за голосом лише з використанням сукупності формантних частот та кепстральної ознаки MFCC

Використані формантні частоти	За ознакою MFCCs	Вік та стать	Стать	Вік
Усі	Усі	50%	66,66%	50%
F1 F2	Усі	50%	55,55%	50%
F1 F2	MFCC 1 2 3	50%	55,55%	50%
F1 F2	MFCC 10 11 12	50%	55,55%	50%
F1 F2	MFCC 4 5 6 7 8 9	50%	55,55%	50%

Для методів розпізнавання, в яких використовується лише частоти формант найкращим результат був у випадку застосування формантів 1 і 2, а найкращий результат для розпізнавання статі, було отримано при використанні всіх частот формантів разом.

У випадку коли було виключено з експерименту дітей, то якість розпізнавання статі при застосовуванні різних формантів була суттєво кращою, що наведено в таблиці 4.6. При цьому якість розпізнавання статі мала значно кращий (вищий) показник –це 91,6 %.

Таблиця 4.6 – Якість розпізнавання статі (без урахування дітей) за оценками формантних частот

Оцінка показника правильного розпізнавання в процентах (без дітей)	
Формантні частоти	Стать людини
F1 F2 F3 F4	83,33%
F1 F2	83,33%

Продовження таблиці 4.6 – Якість розпізнавання статі (без урахування дітей) за оценками формантних частот

Оцінка показника правильного розпізнавання в процентах (без дітей)	
Формантні частоти	Стать людини
F3 F4	91,66%
F3	66,66%
F4	66,66%
F1 F4	66,66%

Таким чином, оцінка віку та статі можуть бути знайдено з мовних сигналів.

Найкращі результати було знайдено для випадку використання першої та другої формантних частот, оскільки основна інформація мовного сигналу зосереджена в низькочастотній частині спектру, що справедливо і для ознак MFCC. При використанні ознак MFCC сумістно з ознаками основного тону якість розпізнавання не покращелась суттєво.

Кращі результати можливо отримати за рахунок поєднання кепстральних ознак MFCC з такою акустичною характеристикою як частота основного тона. Якщо замість простого класифікатора KNN використовувати класифікатор SVM, то можна додатково покращити результати розпізнавання віку та статі.

4.2 Результати експериментальних досліджень з використанням класифікатора SVM

Сумістні оцінки ознак формантних частот та оцінок MFCC використовуються як ознаки для класифікації віку та статі людини.

Хоча ознаки MFCC є дуже добрими ознаками для визнання віку та статі, їх недолік полягає в тому, що вони не враховують особливостей сигналів вокалізованих звуків. Для очищення від шуму мовленевих сигналів використовується попередня цифрова обробка яка спеціалізується на очищенні мовних сигналів від шумів та завад. Для дослідження алгоритмів використано базу даних, що складаються з 82 еталонів та включає в себе дорослих чоловіків та жінок. Результати класифікації статі за двома класами (дорослих жінок і чоловіків) наведено в таб. 4.7 та рис. 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, де використовуються оцінки різних частот формантів (F1, F2, F3 і F4) в якості ознак та класифікатор SVM з функцією лінійного поділу як класифікатора.

Таблиця 4.7- Розпізнавання статі з використанням оцінок формантних частот та лінійний класифікатор SVM

Процент правильного розпізнавання статі	
Формантні частоти	Процент правильного розпізнавання
F2, F4, F3, F4	85 %
F3	89%
F2,F3	90%
F3,F4	89%
F1 F2 F3	90%

Продовження таблиці 4.7- Розпізнавання статі з використанням оцінок формантних частот та лінійний класифікатор SVM

Процент правильного розпізнавання статі	
Формантні частоти	Процент правильного розпізнавання
F1 F3	90%
F1 F4	81%
F1 F2	83%
F1 F2 F3 F4	88%

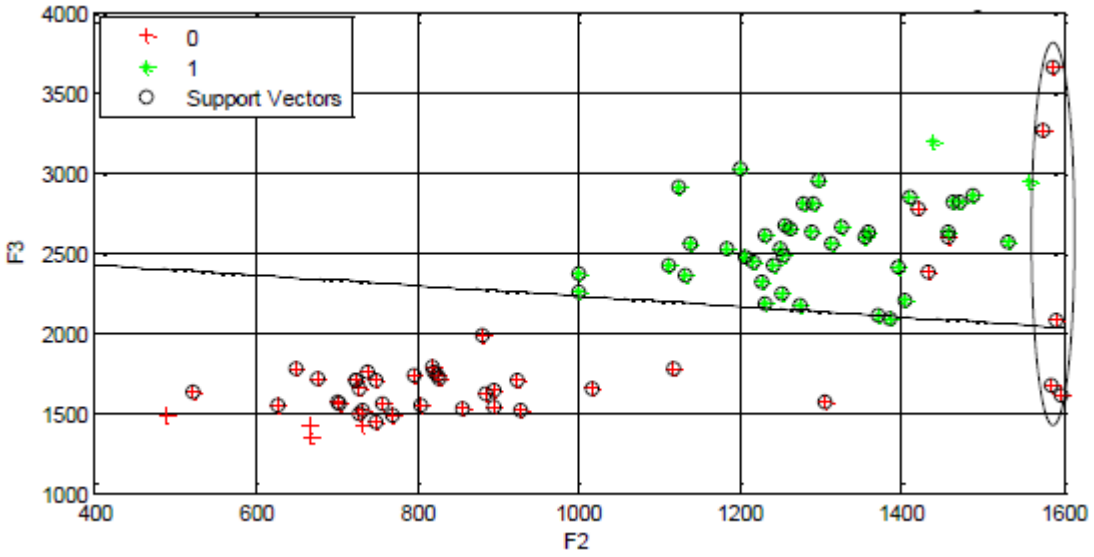

Рисунок 4.1- Результати класифікації статі за формантними частотами F2, F3 та лінійний класифікатор SVM

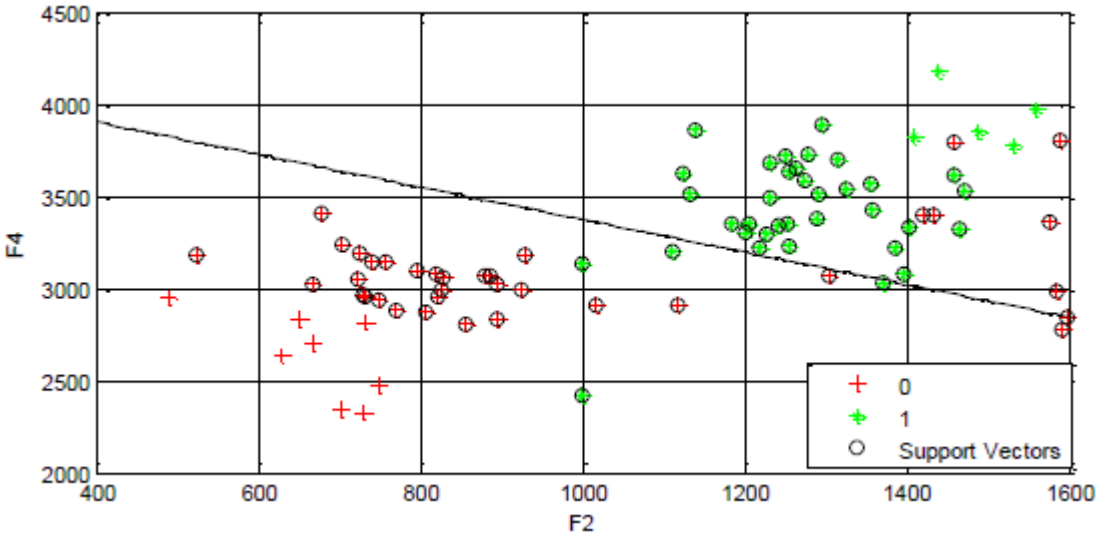

Рисунок 4.2- Результати класифікації статі за формантами F2, F4 та лінійний класифікатор SVM.

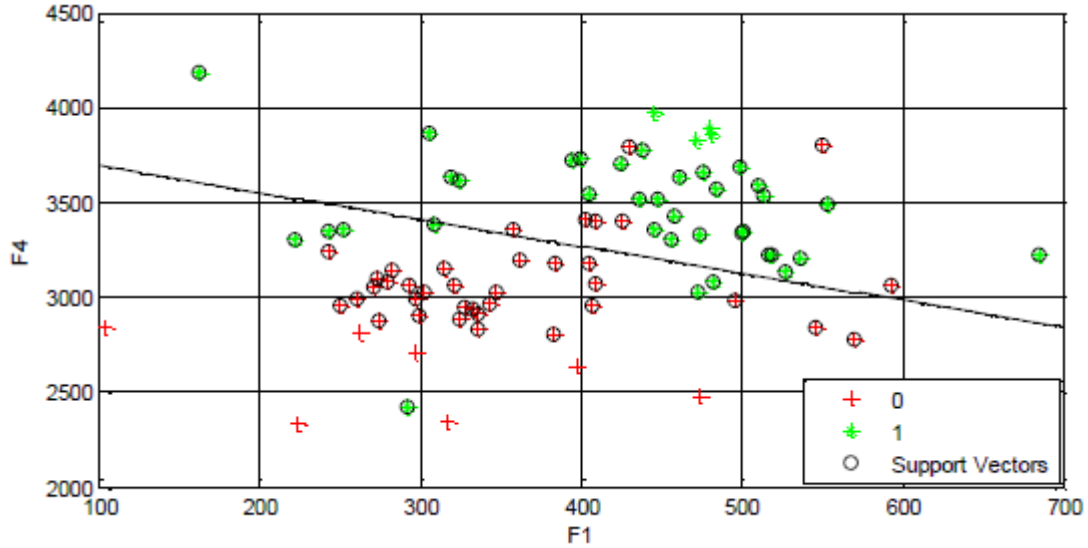

Рисунок 4.3- Результати класифікації статі за формантними частотами F1, F4 та лінійний класифікаторSVM

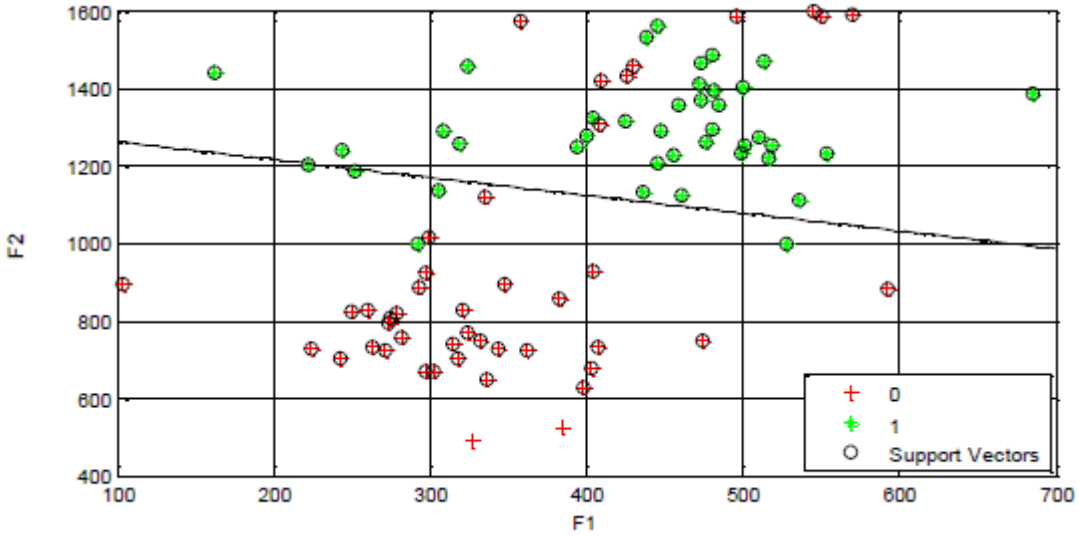

Рисунок 4.4- Результати класифікації статі за формантними частотами F1, F2 та лінійний класифікатор SVM

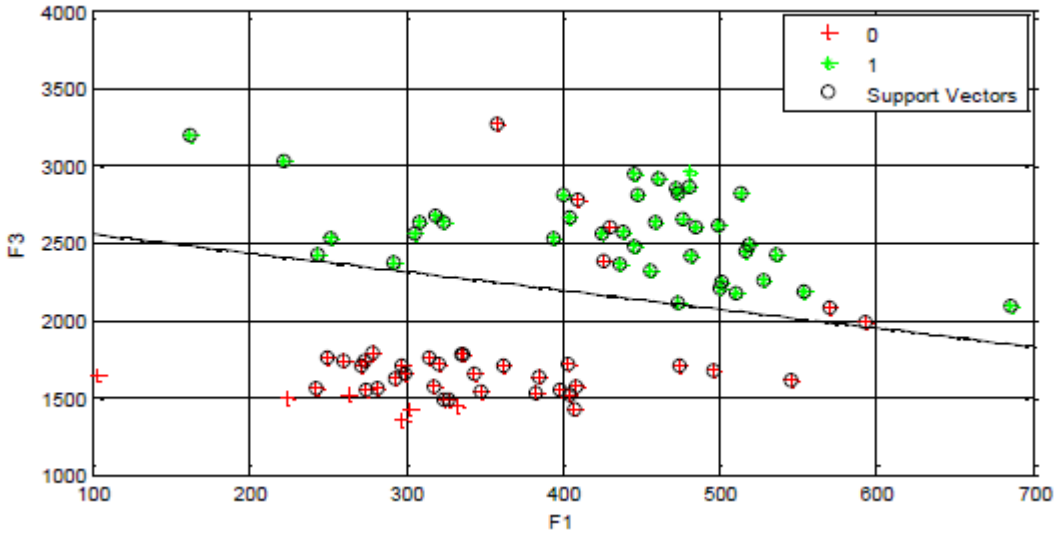

Рисунок 4.5- Результати класифікації статі за формантними частотами F1, F3 з лінійним класифікатором SVM

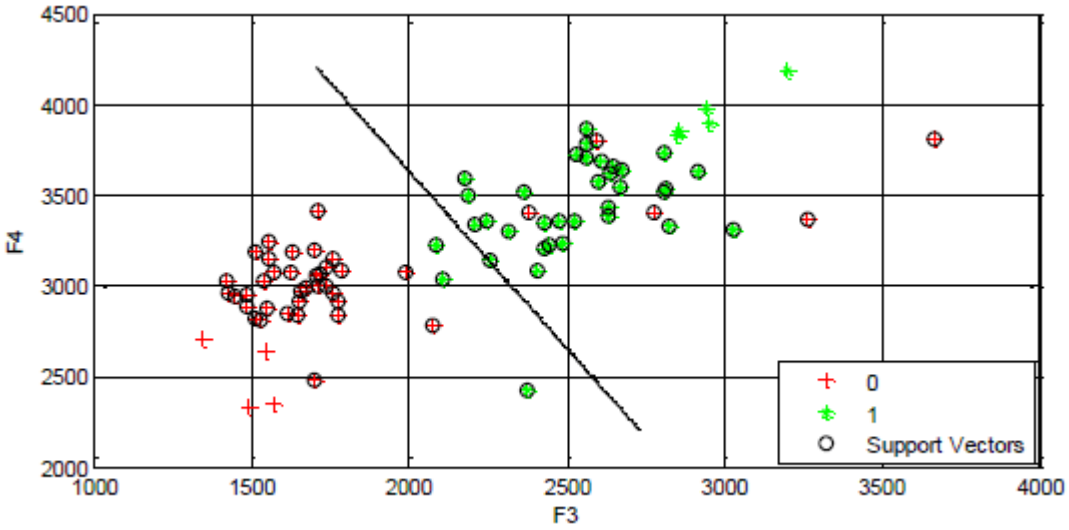

Рисунок 4.6- Результати класифікації статі за формантними частотами F3, F4 з лінійним класифікатором SVM

Вибір надійних ознак для класифікації після використання процедури попередньої обробки також має важливе значення для отримання найкращих результатів.

Класифікатор K-Найближчих-сусідів (k-NN) це дуже простий класифікатор з простими мірами для порівняння і тому він є широко вживаним як класифікатор для вирішення задач розпізнавання. Такий класифікатор використовується у багатьох випадках та цей класифікатор розділює кожен набір даних у вибірці в одну з груп знайдених на етапі тренування.

В роботі виконано дослідження для різних варіантів побудови пристрою розпізаваня за різною сукупністю ознак. Класифікація статі та віку за кількома ознаками призводить до отримання кращих результатів розпізавання.

Форманти 2 та 3 за даними попередніх дослідників є специфічними для дітей та дорослих, та є різними для дорослих чоловіків та дорослих жінок, а тому ці дві формантні частоти (друга та третя) містять особливості голосу людей з різною статтю та віком і можуть бути обрані для

покращення результатів розпізнавання. Роздільна інформація про статть та вік міститься також в функції MFCC з тієї ж причини.

Інформацію про вік міститься також в смугах середньої частоти; смугах низької та високої частоти мовного сигналу. При цьому високочастотні компоненти людського голосу слабкішають з її віком.

Для випадків використання нелінійних класифікаторів SVM результати гендерного розпізнавання для двох класів наведені у таблиці 4.8 та рис. 4.7, 4.8, 4.9, 4.10, 4.11, 4.12.

Таблиця 4.8- Розпізнавання статі за допомогою формантних частот та нелінійним класифікатором SVM

Показник процент правильного розпізнавання	
Використані формантні частоти	Результат
F2 F4	90 %
F3	94%
F2,F3	96%
F3,F4	92%
F1 F2 F3	91%
F2 F3 F4	94%
F1 F3	87%
F1 F4	83%
F1 F2	85%
F1 F2 F3 F4	90%

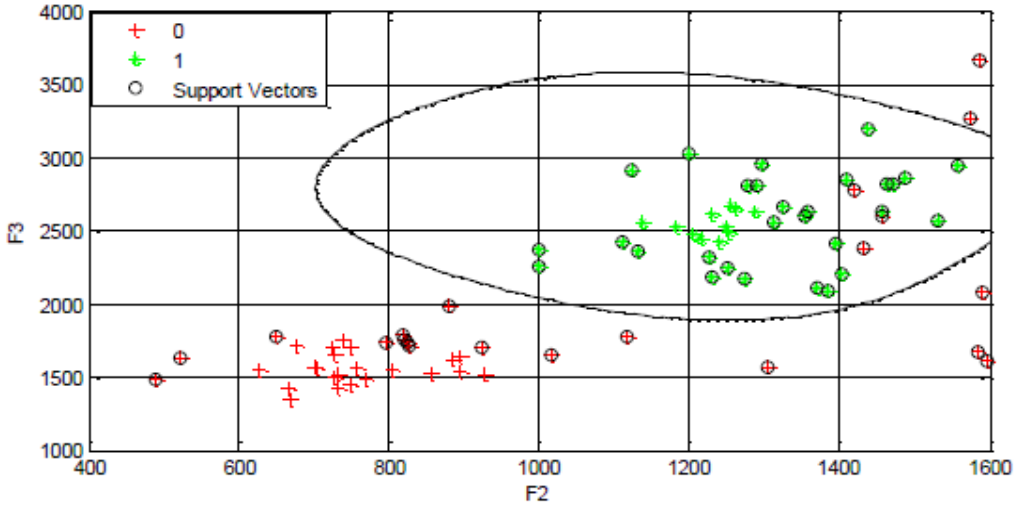

Рисунок 4.7- Результати класифікації статі за формантними частотами F2, F3 та нелінійний класифікатор SVM

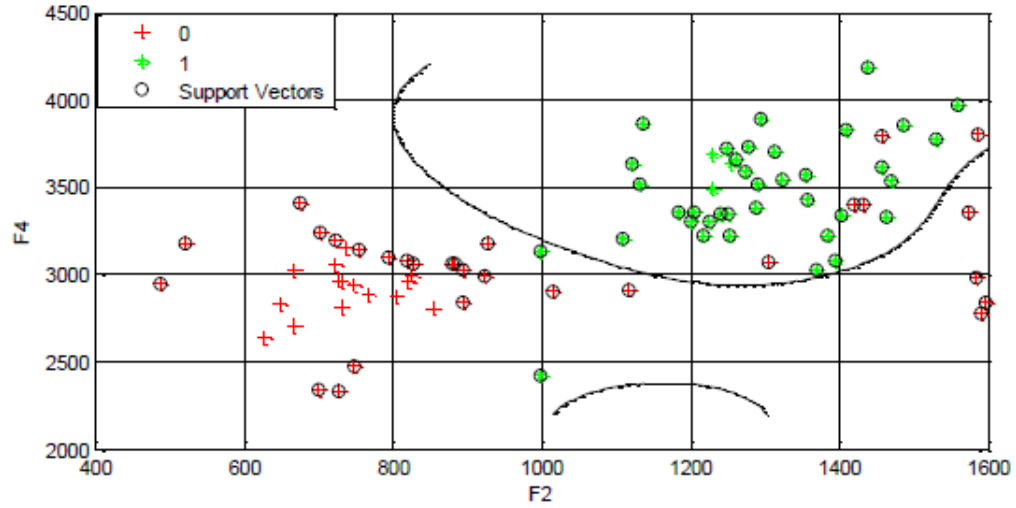

Рисунок 4.8- Результати класифікації статі за формантними частотами F2, F4 та нелінійним класифікатором SVM

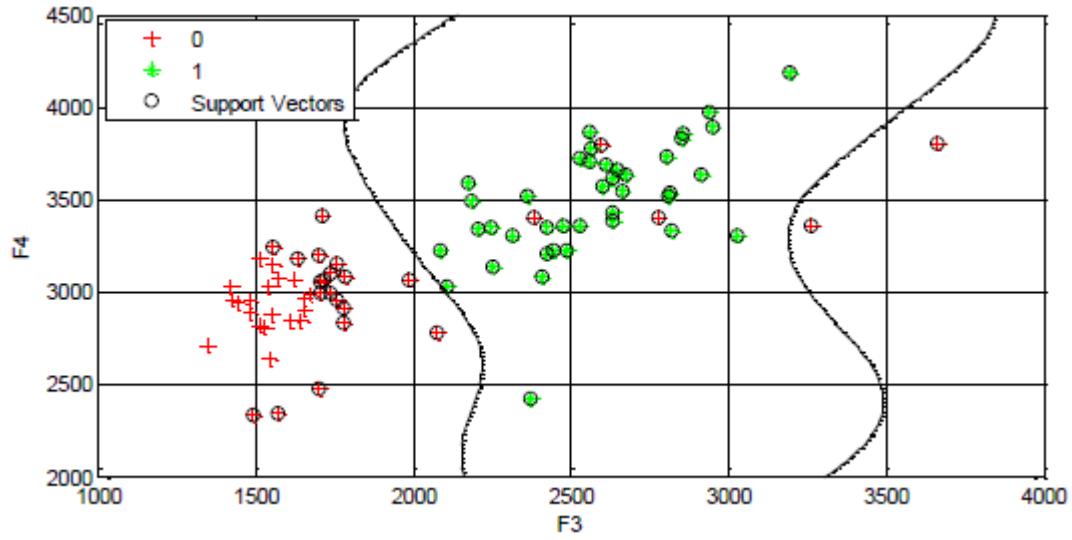

Рисунок 4.9- Результати класифікації статі за формантними частотами F3 F4 та нелінійним класифікатором SVM

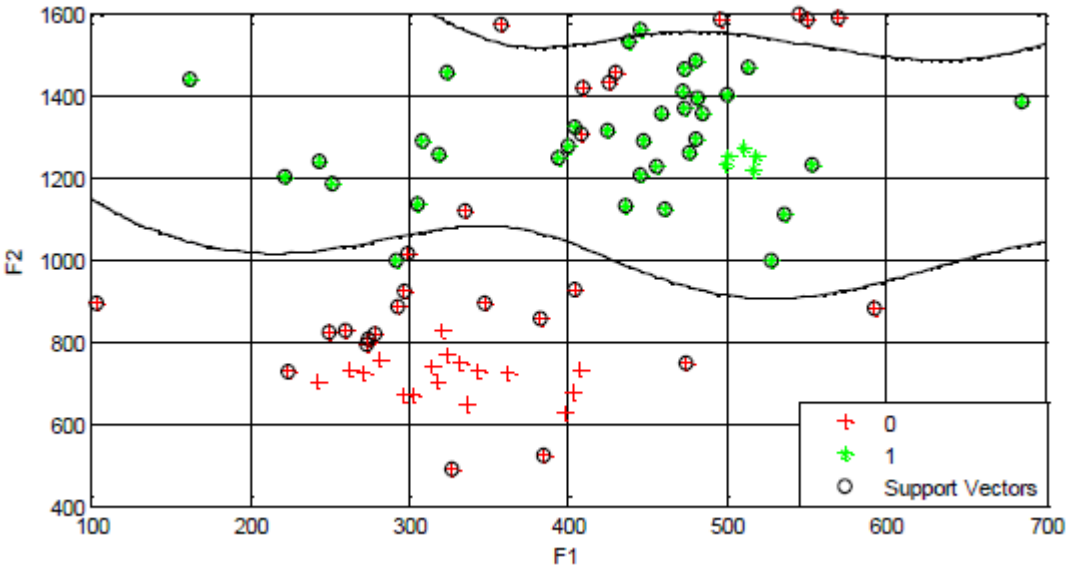

Рисунок 4.10- Результати класифікації статі за формантними частотами F1, F2 та нелінійним класифікатором SVM

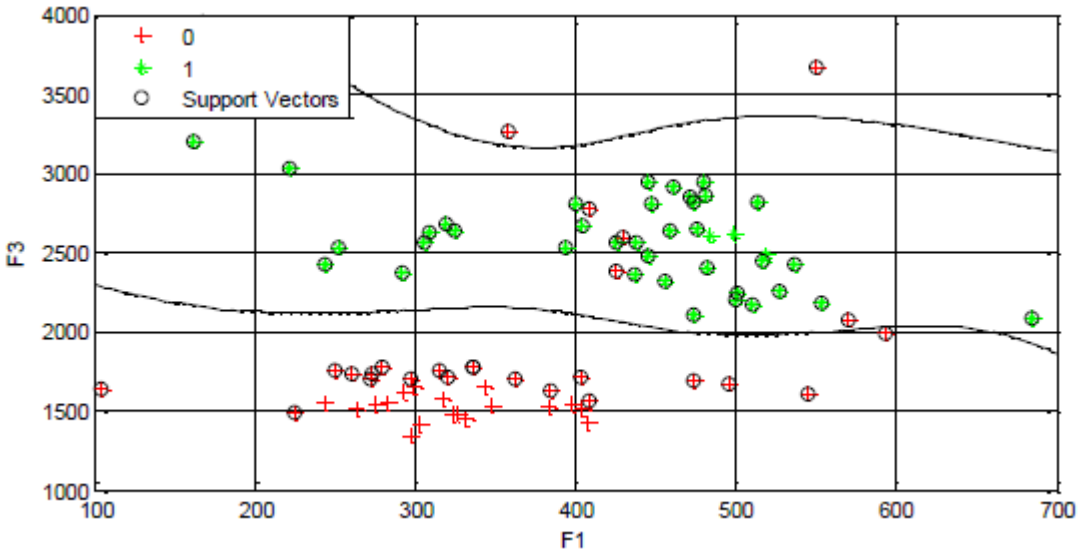

Рисунок 4.11- Результати класифікації статі за формантами F1, F3 та нелінійним класифікатором SVM

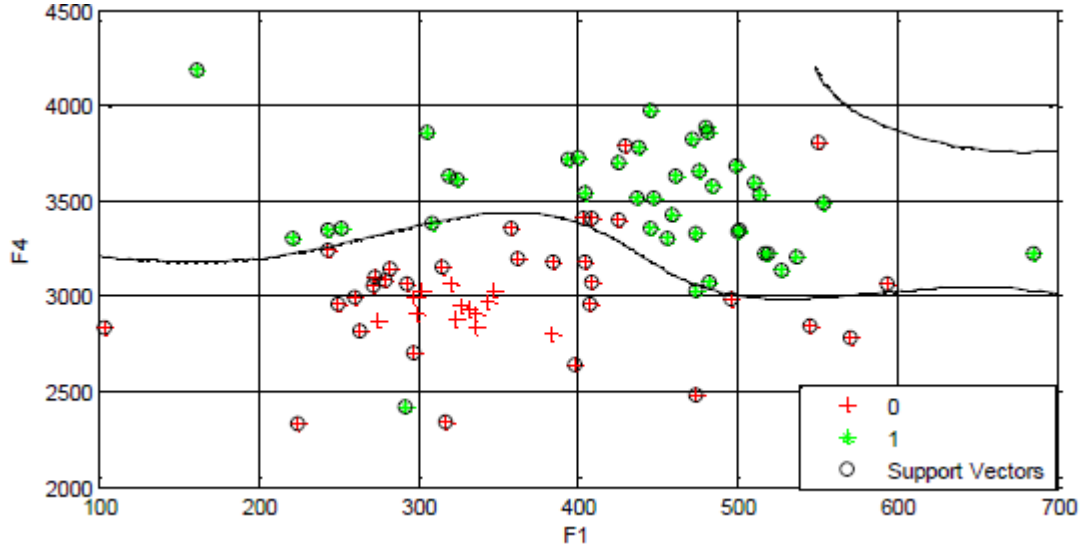

Рисунок 4.12- Результати класифікації статі за формантними частотами F1, F4 та нелінійним класифікатором SVM

Дія шум впливає на оцінки формант обох статей, але для формант жінок вплив є набагато більшим, ніж для чоловіків. Така проблема впливає особливо на розпізнавання при використанні SVM з лінійним розділеннями. Хоча при цьому при використанні нелінійних розділених функцій більш активна для галасливих сигналів, особливо при використанні оцінок формантних частот F2 і F3, що показано на рис. 4.7.

Проведено дослідження гендерного розпізнавання трьох класів де включено як клас дітей з використанням різних частот форматів та різних коефіцієнтів функції MFCC та застосуванні k-NN класифікатора. Для дослідження цього алгоритму використано до 114 еталонів (мовців), включаючи дітей без гендерної дискримінації, та дорослих чоловіків та жінок.

При цьому найкращий результат було отримано при використанні формант F2, F3 та середніх коефіцієнтів ознаки MFCC.

Таблиця 4.9- Результати дослідження якості розпізнавання статі за допомогою ознак MFCC, формантних частотах та нелінійному класифікаторі SVM

Оцінки в процентах якості правильного розпізнавання статі	
Формантні частоти	Результат в процентах
12 MFCC+ F1, F2, F3, F4	88 %
4-9 MFCC+ F1, F2, F3, F4	92%
4-9 MFCC	96%
4-10 MFCC	93%
3-9 MFCC	92%
12MFCC+ F2, F3	88%
4-9 MFCC+ F1, F2, F3, F4	96%

Таблиця 4.10- Результати досліджень розпізнавання статі за допомогою ознак MFCC та формантних частотам з класифікатором K-NN

Оцінки в процентах правильного розпізнавання	
Формантні частоти	Стать
12 MFCC+ F1, F2, F3, F4	85 %
F2, F3	94%
12 MFCC+ F2, F3	85%
F1, F2, F3	93%
F1, F4	79%
12MFCC+ F2, F3	94%
4-9 MFCC	94%
F1, F2, F3, F4	92%

4.3Результати класифікації за віковом

В таблиці 4.11 зведено результати досліджень класифікатора за віком людини з використанням в якості ознак частоти першої форманти та 12 мел кепстральних коефіцієнтів MFCC як функцій та k-NN в якості класифікатора.

Таблиця 4.11- Розпізнавання віку людини за допомогою за допомогою ознак MFCC, оцінок формантних частот та класифікатора K-NN

Процент правильного розпізнавання статі	
Формантні частоти	Результат
F1, F4	66 %
F2, F3	52%

Продовження таблиці 4.11- Розпізнавання віку людини за допомогою за допомогою ознак MFCC, оцінок формантних частот та класифікатора K-NN

Процент правильного розпізнавання статі	
Формантні частоти	Результат
F2, F3, F4	52%
F3,F4	59%
F1, F2, F3	54%
F1, F3, F4	68%
12 MFCC+ F1, F2, F3, F4	69%
12 MFCC+ F1, F4	57%
1-4,8-12 MFCC+ F1, F4	75,3%

Для дослідження в цьому алгоритмі використовується для досліджень базу даних із 114 дикторів різного віку насинаючи від п’яти до шістдесяти п’яти років. За результатами експериментальних досліджень можливо зробити висновок що формантні частоти F1 та F4 є важливими ознаками для класифікації віку людини и є дуже ефективним при поєднанні цих ознак з низько- та високочастотними характеристиками MFCC. З результатів цієї частини досліджень можна зробити висновок, що форманти для низької та високої частоти спектру як і ознаки MFCC є інформативними для розділення людей за віком, а зокрема це можна помітити з результатів гендерного розпізнавання людей, які наведено у таб. 4.9. З проведених результатів експериментальних досліджень слідує що 4, 8 та 9 коефіцієнти MFCC є корисною роздільною інформаціїю як для статевої, так і для вікової класифікації.

Результати придушення завад фільтром свідчит про ефективність його використання та правильний вибір ознак та класифікатора, що наведено в таблиці 4.10. У цьому випадку формантні частоти F1 і F4 також мають головне значення для розпізнавання, а високочастотні коефіцієнти MFCC мають великий вплив на результати классифікацій.

ВИСНОВОК

При використанні формантних частот, найкращі результати розпізнавання віку та статі було отримано при використанні пешої та другої формантних частоти. Це пояснюється тим, що основна роздільна інформація мовного сигналу зосереджена в низькочастотній частині спектру. При застсуванні в якості ознак MFCC найкращі результати було отримано для алгоритма першими трьом коефіцієнтами MFCC. Результати обєднання формантних частоти та ознак MFCC не покращували суттєво якість розпізнавання.

Алгоритми в яких є поєднання ознак MFCC з іншою акустичною характеристикою, такою як частота основного тону показал свою ефективність.

Найкращими результати класифікації за віком та станттю було експериментально отримано для випадку двох класів для статі та трьох класів для віку при використанні другої та третьої формантю Це пояснюється тем , ці форманти відрізняються для вікових категорій та жінок і чоловіків. В середньочастотних функціях MFCC також є інформація про стать, а тому ці функції MFCC раціонально використовувати для покращення якості класифікації за статтю.

Застосування в якості класифікатора SVM з нелінійною розділеною функцією дає кращі результати ніж використання лінійної розділеної функції.

З проведених експериментальних досліджень, щ проведені в цій роботі, випливає, що формантні частоти F1 і F4 мають більше значення для визнання віку людини, ніж формантні частоти F2 і F3, а тому вони були обрані для визнання віку людини.

Важливими для визнання віку є 12 оцінок MFCC, та серед них особливо низьких та високих номерів MFCC, оскільки в них містеться

інформація про вік людини. Однак недоліком такого вибору є те, що на результат розпізнавання впливає шум. Застосування методу попередньої обробки мови та вибору надійних функцій, якими є дванадцять MFCC на додаток до функцій F1 і F4 призводять до кращих результатів розпізнавання віку та статі.

Завадостійкі ознаки вибираються в цій роботі за результатами експериментальних досліджень, а не застосовуються автоматично за схемою попереднього знання.

ПЕРЕЛІК ПОСИЛАНЬ

1. M. Li, K. Han, and S. Narayanan, "Automatic speaker age and gender recognition using acoustic and prosodic level information fusion," Computer speech and language, Vol. 27, No. 1, pp. 151-167, Jan. 2013

2. H Meinedo and I Trancoso, “Age and Gender Classification Using Fusion of Acoustic and Prosodic Features”, Proc. INTERSPEECH, pp. 2818-2821, 2010

3. R. Nisimura, A. Lee, H. Saruwatari, and K. Shikano, "Public speech-oriented guidance system with adult and child discrimination capability,"Proc. ICASSP2004, vol. 1, pp. 433-436, 2004.

4. H. Kim, K. Bae, H. Yoon, “Age and gender classification for a home-robot service” Proc. 16th IEEE International Symposium on Robot and Human Interactive Communication, pp. 122–126, 2007

5. W. Li, D. J. Kim, C. H. Kim, and K. S. Hong, "Voice-Based Recognition System for Non-Semantics Information by Language and Gender" Electronic Commerce and Security (ISECS), 2010.

6. P. Nguyen, D. Tran, X. Huang, and D. Sharma, "Automatic classification of speaker characteristics" Communications and Electronics (ICCE), 2010.

7. G. Dobry, R. M. Hecht, M. Avigal, and Y. Zigel, "Supervector dimension reduction for efficient speaker age estimation based on the acoustic speech signal." Audio, Speech, and Language Processing, 2011

8. M. H. Bahari, and H. V. Hamme, "Speaker age estimation and gender detection based on supervised non-negative matrix factorization" Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2011.

9. M. H. Sedaaghi, "A comparative study of gender and age classification in speech signals" Iranian Journal of Electrical & Electronic Engineering, 2009

10. T. Bocklet, G. Stemmer, V. Zeissler, and E. Nöth, "Age and gender recognition based on multiple systems-early vs. late fusion" INTERSPEECH, 201011. S. Schötz, "Acoustic analysis of adult speaker age" Speaker Classification I. Springer Berlin Heidelberg, 2007

12. M. K. Wolters, V. Ravichander, and R. Steve,"Age recognition for spoken dialogue systems: Do we need it?" INTERSPEECH, 2009

13. M. Feld, F. Burkhardt, and C. A. Müller, "Automatic speaker age and gender recognition in the car for tailoring dialog and mobile services" INTERSPEECH, 2010

14. F. Metze, J. Ajmera, R. Englert, and U. Bub, "Comparison of four approaches to age and gender recognition for telephone applications" Acoustics, Speech and Signal Processing, 2007

15. C. A. Müller, F. Wittig, and J. Baus, "Exploiting speech for recognizing elderly users to respond to their special needs" INTERSPEECH, 2003

16. M. W. Lee, and K. C. Kwak. "Performance Comparison of Gender and Age Group Recognition for Human-Robot Interaction" International Journal of Advanced Computer Science & Applications, 2012

17. F. Wittig, and C. Müller, "Implicit Feedback for User-Adaptive Systems by Analyzing the Users' Speech.", 2003

18. T. Bocklet, A. Maier, and E. Nöth, "Age determination of children in preschool and primary school age with GMM-based supervectors and support vector machines/regression" Text, Speech and Dialogue, 2008

19. M. H. Bahari, and H. V. Hamme, "Speaker age estimation using Hidden Markov Model weight supervectors" Information Science, Signal Processing and their Applications (ISSPA), 2012.

20. Voice activity detection online. Available: http://en.wikipedia.org/wiki/Voice_activity_detection#Applications

21. J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection" Signal Processing Letters, 1999

22. T. Carrasquillo, P. A., E. Singer, and M. A. Kohler, "Approaches to language identification using Gaussian mixture models and shifted delta cepstral features"INTERSPEECH, 2002

23. W. M. Campbell, E. Singer, P. A., and T. Carrasquillo, "Language recognition with support vector machines" ODYSSEY04-The Speaker and Language Recognition Workshop, 200424. F. Allen, E. Ambikairajah, and J. Epps, "Language identification using warping and the shifted delta cepstrum" Multimedia Signal Processing, 2005

25. A. Russell, L. Penny, and C. Pemberton, "Speaking Fundamental Frequency Changes Over Time in Women A Longitudinal Study" Journal of Speech, Language, and Hearing Research, 1995

26. H. Hollien, and T. Shipp, "Speaking fundamental frequency and chronologic age in males" Journal of Speech, Language, and Hearing Research, 1972

27. X. Sun, "A pitch determination algorithm based on subharmonic-to-harmonic ratio", 2000

28. M. Brookes, "Voicebox: Speech processing toolbox for matlab", Mar 2011 29. M. Sahidullahl, “Mfcc to SDC Matlab Library”, 2011

30. X. Sun, "A pitch determination algorithm based on subharmonic-to-harmonic ratio", 2000

31. C. C. Chang, and C. J. Lin, "LIBSVM: a library for support vector machines." ACM Transactions on Intelligent Systems and Technology, 2011

32. L. Feng, “Speaker Recognition, Informatics and Mathematical Modelling”, Technical University of Denmark, 2004

33. H. Hermansky, and N. Morgan, “RASTA processing of speech.” Speech and Audio Processing, 1994

34. S. Davis, and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences” Acoustics, Speech and Signal Processing, 1980