

Вінницький національний технічний університет
(повне найменування вищого навчального закладу)

Факультет комп'ютерних систем і автоматики
(повне найменування інституту, назва факультету (відділення))

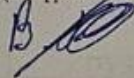
Кафедра комп'ютерних систем управління
(повна назва кафедри (предметної, циклової комісії))

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА
на тему:

«Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання»

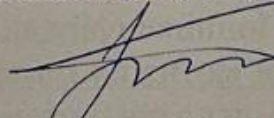
Виконав: магістр 2 курсу, групи 2АКІТ-22м
спеціальності 151- Автоматизація та
комп'ютерно- інтегровані технології

(цифр і назви напряму підготовки, спеціальності)



Вадим Кошельник
(ім'я ПРІЗВИЩЕ)

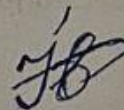
Керівник: к.т.н., професор каф. КСУ



Микола БИКОВ
(ім'я ПРІЗВИЩЕ)

«01» 12 2023 р.

Опонент: к.т.н., доцент каф. АІТ



Юрій ІВАНОВ
(ім'я ПРІЗВИЩЕ)

«05» 12 2023 р.

Допущено до захисту

Т.в.о. зав. кафедри КСУ

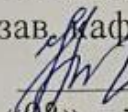
Марія ЮХИМЧУК

(прізвище та ініціали)

«07» грудня 2023 року

Вінницький національний технічний університет
 Факультет інтелектуальних інформаційних технологій та автоматизації
 Кафедра комп'ютерних систем управління
 Рівень вищої освіти II-й (магістерський)
 Галузь знань – 15 Автоматика та приладобудування
 Спеціальність - 151 Автоматизація та комп'ютерно-інтегровані технології
 Освітньо-професійна програма - Інтелектуальні комп'ютерні системи

ЗАТВЕРДЖУЮ

Т.в.о. зав. кафедри КСУ

 Марія ЮХИМЧУК
 «09» жовтня 2023 року

ЗАВДАННЯ
 НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
Кошельнику Вадиму Олександровичу

(прізвище, ім'я, по батькові)

1. Тема роботи «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання»

керівник роботи Биков Микола Максимович, к. т. н., доцент
 (прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом ВНТУ від «18» вересня 2023 року № 247

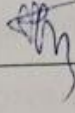
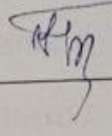
2. Строк подання студентом роботи «1» грудня 2023 року

3. Вихідні дані до магістерської кваліфікаційної роботи: об'єкт управління – процес комунікації між оператором і клієнтом в контакт центрі; регульовані параметри – темп мовлення; параметри, що оптимізуються в моделі – кількість вимовлених складів за хвилину; середовища моделювання – Матлаб, Visual Studio 2017.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) вступ, варіантний аналіз проблеми контролю темпу мови в комунікаційних центрах, розробка методу і пристрою виділення складів в сигналі мови, розробка алгоритмів і програмного забезпечення контролю темпу мови, машинний експеримент, економічна частина.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) структурна схема підсистеми контролю темпу мови, математичні моделі слухової системи і оцінки кількості градацій темпу мовлення, алгоритм визначення темпу мовлення, структурна схема пристрою управління темпом мовлення, схеми програм управління темпом мовлення, результати експерименту.

6. Консультанти розділів магістерської кваліфікаційної роботи

Розділ змістової частини роботи	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Розділ 4	Буреннікова Н.В. д.е.н., пофесор кафедри ЕПВМ		

7. Дата видачі завдання “ 09 ” жовтня 2023 року

КАЛЕНДАРНИЙ ПЛАН

1	Аналіз об'єкта автоматизації і обґрунтування доцільності розробки	02.10.2023 р.
2	Постановка задачі і розробка технічного завдання	04.10.2023 р.
3	Математичне обґрунтування методу вимірювання темпу мови	18.10.2023 р.
3	Розробка структури системи контролю темпу мови	20.10. 2023 р.
4	Розробка програмного забезпечення і машинний експеримент	5.11. 2023 р.
5	Підготовка економічної частини	12.11. 2023 р.
6	Оформлення пояснювальної записки, графічного матеріалу і презентації	20.11. 2023 р.
7	Апробація результатів дослідження	22.11.2023 р.
8	Публікації	22.11. 2023 р.
9	Графічні матеріали: Схема алгоритму управління Схема електрична принципова Схема автоматизації структурна Схема електрична функціональна Схема алгоритму зміни налаштувань Блок управління системою автоматичного поливу	23.11.2023 р. 23.11.2023 р. 23.11.2023 р. 23.11.2023 р. 26.11.2023 р. 30.11.2023 р.
10	Захист МКР	20.12. 2023 р.

Студент

Вадим КОШЕЛЬНИК

Керівник роботи

Микола БИКОВ

АНОТАЦІЯ

УДК 681.3.019:621.39

Кошельник В.О. Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання. Магістерська кваліфікаційна робота зі спеціальності 151 – автоматизація та комп'ютерно-інтегровані технології, освітня програма – інтелектуальні комп'ютерні системи. Вінниця: ВНТУ, 2023. 122 с.

На укр. мові. Бібліогр.: 116 назв; рис.: 20 ; табл. 15 .

Магістерська кваліфікаційна робота присвячена вирішенню актуальної проблеми підвищення ефективності голосового спілкування операторів комунікаційних центрів в системах критичного використання.

На основі проведеного аналізу факторів впливу на якість комунікації в контакт-центрах було зроблено висновок, що основним з них є темп мовлення. Тому в роботі запропоновано метод автоматизації контролю і регулювання темпу мовлення на основі інформації про кількість і тривалість складів в мовленнєвому сигналі. Виконано математичне обґрунтування запропонованого методу виділення складів в мовленнєвому потоці, розроблена структурна схема пристрою виділення складів.

В розділі алгоритмічного і програмного забезпечення розроблено алгоритми обробки сигналу мови, а також програмне забезпечення для їх реалізації. Проведене тестування програм підтвердило адекватність їх роботи для визначення темпу мовлення мовців.

В економічній частині проведено техніко-економічне огрунтування доцільності і ефективності впровадження результатів розробки в реальні проекти.

Ілюстративна частина роботи містить 12 плакатів і креслень з ілюстрацією результатів роботи.

Ключові слова: мовлення, сигнал мови, темп мовлення, автоматизація контролю, комунікаційний центр, критична система.

ABSTRACT

UDC 681.3.019:621.39

Wallek V.O. Automation of speech rate control in communication centers of critical use systems. Master's thesis on specialty 151 - automation and computer-integrated technologies, educational program - intelligent computer systems. Vinnytsia: VNTU, 2023. 122 p.

In Ukrainian speech Bibliography: 116 titles; Fig.: 20 ; table 15 .

The master's thesis is devoted to solving the current problem of increasing the effectiveness of voice communication of communication center operators in critical use systems.

Based on the analysis of factors influencing the quality of communication in contact centers, it was concluded that the main factor is the speed of speech. Therefore, the paper proposes a method of automating the control and regulation of speech tempo based on information about the number and duration of syllables in the speech signal. Mathematical substantiation of the proposed method of syllable selection in the speech stream was performed, and a structural diagram of the syllable selection device was developed.

Algorithms for speech signal processing, as well as software for their implementation, have been developed in the algorithmic and software section. The conducted testing of the programs confirmed the adequacy of their work for determining the speaking pace of speakers.

In the economic part, a technical and economic substantiation of the feasibility and effectiveness of implementing the development results into real projects was carried out.

The illustrative part of the work contains 12 posters and drawings illustrating the results of the work.

Key words: speech, speech signal, speech rate, control automation, communication center, critical system.

ЗМІСТ

ВСТУП.....	8
1 ВАРІАНТНИЙ АНАЛІЗ ПРОБЛЕМИ КОНТРОЛЮ ТЕМПУ МОВИ В КОМУНІКАЦІЙНИХ ЦЕНТРАХ.....	11
1.1 Суть технічної проблеми, що вирішується в роботі	11
1.2 Аналіз проблем, пов'язаних з темпом мовлення в контакт-центрах.....	13
1.3 Аналіз програмних рішень	18
1.4 Аналіз аналогів і прототипів.....	26
1.5 Аналіз методів обробки мовного сигналу з метою визначення швидкості мовлення.....	27
1.6 Висновки до розділу.....	35
2 РОЗРОБКА МЕТОДУ І ПРИСТРОЮ ВИДІЛЕННЯ СКЛАДІВ В СИГНАЛІ МОВИ	36
2.1 Моделювання обробки сигналу мови слуховою системою.....	36
2.2 Розробка методу і пристрою виділення складових сегментів слова.....	42
2.3 Розробка алгоритму виділення складових сегментів	49
2.4 Висновки до розділу.....	52
3 РОЗРОБКА АЛГОРИТМІВ І ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ.....	54
3.1 Вибір програмної платформи.....	54
3.2 Розробка алгоритмів і програм оброблення сигналу мовлення	54
3.3 Опис вхідних, вихідних та проміжних даних	65
3.4 Розробка програмної документації	67
3.5 Інструкція користувача.....	69
3.6 Висновки до розділу	70

4 ЕКОНОМІЧНА ЧАСТИНА.....	71
4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки.....	71
4.2 Розрахунок узагальненого коефіцієнта якості розробки	74
4.3 Розрахунок витрат на проведення науково-дослідної роботи.....	76
4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором.....	87
4.5 Висновки до розділу.....	92
ВИСНОВКИ.....	93
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	94
ДОДАТКИ.....	99
<u>Додаток А (обов'язковий) Протокол перевірки на плагіат.....</u>	<u>100</u>
Додаток Б (обов'язковий) Технічне завдання.....	101
Додаток В (довідниковий) Текст програми.....	105
Додаток Г Ілюстративна частина.....	109

ВСТУП

Сьогодні на порядку денному в багатьох дослідженнях, пов'язаних з мовними технологіями, вирішується проблема визначення параметрів голосового зв'язку в комунікаційних каналах, які здійснюють суттєвий вплив на якість комунікації. Серед них особливо актуальною є задача оцінювання такого показника, як темп мовлення. Аналіз темпу мови в критичних системах є важливою проблемою, пов'язаною з оцінкою швидкості та чіткості комунікації в ситуаціях, де навіть дрібна затримка або непорозуміння може мати серйозні наслідки. У критичних системах, таких як медичні процедури, авіаційна безпека, автоматизовані процеси виробництва або управління технологіями, точність та оперативність важливі для запобігання небезпеці та забезпечення ефективності.

Оцінка темпу мови полягає в здатності сприймати, аналізувати та відповідати на інформацію, що подається з мінімальною затримкою. У деяких випадках, наприклад у кризових ситуаціях, недоліки в чіткості або сприйнятті мовлення можуть призвести до помилок у виконанні інструкцій або прийняття неправильних рішень, що має потенційно небезпечні наслідки. Наприклад, ведення контролю над цим параметром у системах оброблення телефонних дзвінків у контакт-центрах дозволяє оптимізувати темп діалогу оператора з клієнтом, що призводить до підвищення ефективності управління трафіком та поліпшення комфорту обслуговування клієнтів.

На сьогодні існує декілька методів аналізу темпу мовлення, до основних з них можна віднести наступні.

Аудіоаналіз, який визначає швидкість мовлення на основі частоти слів або фраз у часовому проміжку шляхом аналізу звукових записів використанням спеціального програмного забезпечення (NLP – natural language processing, обробка природної мови). Реалізується не в реальному часі, а для записаних фонограм, що не дозволяє автоматизувати визначення темпу мови в реальному часі безпосередньо в комунікаційному центрі.

Метод вимірювань фізіологічних показників, таких як частота серцевих скорочень чи пульс, для визначення швидкості мовлення під час стресових ситуацій. Вимагає обладнання і аналізу результатів людиною, що не дозволяє автоматизувати процес контролю і управління темпом мовлення в реальному часі. Підходить більше для медичних досліджень патологій мовлення.

Експертна оцінка ґрунтується на оцінці експертів, які аналізують зразки мовлення та оцінюють їх темп. Зокрема, експертна оцінка темпу мовлення, яка виконувалась в роботах з розпізнавання мови з метою врахування впливу часової зміни швидкості мовлення на точність розпізнавання, потребує значних часових затрат і не підходить для задач он-лайн комунікації в контакт-центрах, і не може бути використано для управління темпом в реальному масштабі часу. Тому можна з цього аналізу зробити висновок, що дослідження і розробки, пов'язані автоматизацією процесу контролю і управління темпом мови в комунікаційних центрах, які вирішуються в даній магістерській роботі, є достатньо актуальними.

Метою даної роботи є розробка ефективного методу автоматичного контролю темпу мовлення з метою он-лайн управління швидкістю мовлення в комунікаційних центрах систем критичного використання.

Для досягнення поставленої цілі ставляться і вирішуються такі задачі:

- розробка надійного методу, алгоритму та пристрою для точного визначення компонентів у сигналі мовлення;
- розробка надійного методу виділення пауз в сигналі мовлення;
- розробка алгоритму комбінування розроблених методів для автоматичного обчислення темпу мови;
- розробка програмного забезпечення для реалізації розроблених алгоритмів автоматичного контролю темпу мови.

Об'єкт дослідження – сигнал мовлення в комунікаційному каналі.

Предмет дослідження – методи, алгоритми і пристрої для автоматичного контролю темпу мовлення.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконана відповідно до плану науково-дослідної роботи кафедри КСУ.

Методи дослідження. У процесі дослідження використовувалися такі теоретичні підходи та методи: теорія автоматичного управління; теорія системного аналізу; теорія алгоритмів; теорія інформації; теорія обробки цифрових сигналів. Також застосовувалось комп'ютерне моделювання для аналізу та перевірки достовірності отриманих теоретичних положень.

Наукова новизна роботи полягає в розробці надійного методу виділення складів на основі дослідження моделі слухової системи людини та пристрою підрахунку кількості складів в мовленнєвому потоці.

Практична цінність полягає в тому, що розроблені методи, алгоритми і програмне забезпечення дозволяють виконувати в реальному часі автоматичну оцінку темпу мовлення, а отже, і здійснювати управління ним в реальному масштабі часу.

Впровадження. Результати магістерської кваліфікаційної роботи використовуються в наукових дослідженнях і навчальному процесі кафедри КСУ.

Особистий внесок здобувача.

Здобувачем розроблений програмний комплекс для реалізації запропонованих методів і проведені експериментальні дослідження.

Апробація результатів роботи. Основні положення та результати досліджень були викладені та обговорені під час науково-технічної конференції молодих науковців ВНТУ "Молодь в науці" (2023).

Публікації: За результатами конференції опубліковані тези: М.М. Биков, В.О. Кошельник. Контроль і регулювання темпу мови в комунікаційних центрах систем критичного використання. – [Електронний ресурс]. Режим доступу:

<https://conferences.vntu.edu.ua/index.php/mn/mn2024/author/submission/19771>

1 ВАРІАНТНИЙ АНАЛІЗ ПРОБЛЕМИ КОНТРОЛЮ ТЕМПУ МОВИ В КОМУНІКАЦІЙНИХ ЦЕНТРАХ

1.1 Суть технічної проблеми, що вирішується в роботі

Автоматизація аналізу мовленнєвого темпу в комунікаційних центрах систем диспетчеризації різних організацій може відігравати важливу роль у підвищенні ефективності та безпеки цих систем. Особливо важливою ця задача є для таких критичних систем, як медична діагностика, авіаційна безпека, управління енергетичними системами, військовий зв'язок тощо

Автоматизація контролю мовленнєвої комунікації може бути зосереджена на реалізації таких задач.

- автоматичний аналіз аудіозаписів комунікації між диспетчерами та іншими учасниками системи. Вони можуть визначати тривалість висловлювань, інтервали між ними та загальний темп мовлення;

- аналіз інтонації, гучність та інших аспектів мовлення для виявлення емоційного стану диспетчера. Це може бути корисно для виявлення стресу або неприродної поведінки, що може вказувати на небезпеку чи недоліки у роботі;

- виявлення аномальних чи надто стрімких зміни у темпі мовлення або емоційному стані, автоматична фіксація цих подій, аналіз їх причин та надавання додаткової підтримки в екстрених ситуаціях;

- підвищення безпеки та ефективності роботи диспетчерів шляхом реагування на можливі проблеми швидше або надання рекомендацій для поліпшення якості комунікацій;

- збирання та аналіз аудіоданих протягом тривалого часу для виявлення тенденцій та покращення системи контролю на основі великого обсягу інформації;

- запису та аналіз дзвінки агентів з метою контролю якості обслуговування. В склад такої задачі можуть включати функції аналізу темпу

мовлення, визначення тривалості розмови та пауз, щоб забезпечити відповідність стандартам компанії.

Як показує перелік даних задач, що вимагають автоматизації їх виконання в комунікаційних системах різного призначення, для їх вирішення потрібно розробити програми NLP (обробки природної мови). На сьогодні існує декілька пакетів таких програм, але вони призначені для вирішення вказаних задач в автоматизованому режимі, який потребує участі людини-оператора. В системах критичного призначення, збої в роботі яких можуть привести до важких наслідків, такі функції повинні реалізовуватись автоматично в режимі реального часу. Тому метою даної магістерської роботи є розробка апаратного і програмного забезпечення для автоматичного контролю і регулювання темпу мовлення в комунікаційних центрах критичних систем та контакт-центрах систем диспетчеризації. Об'єктом розробки є методи, моделі, алгоритми і пристрої для реалізації автоматичного контролю темпу мови в реальному масштабі часу, що ґрунтуються на принципі визначення темпу мовлення мовця за кількістю складів, що промовляються ним на протязі заданого періоду часу. Суть технічної проблеми, яка виникає під час розв'язання даної задачі, полягає в необхідності розробки надійного і швидкісного методу виділення складів в мовленнєвому аудіо потоці.

В магістерській роботі ставиться технічна задача побудови програмного і апаратного забезпечення для системи автоматизації управління темпом мовлення операторів комунікаційних центрів критичних систем на основі розробки і дослідження методу автоматичного контролю темпу мовлення. Розглядається метод надійного виділення складових сегментів, призначений для контролю швидкості мовлення в комунікаційних центрах систем диспетчеризації різного типу.

В період інтенсивного розвитку інформаційних систем професія оператора контакт-центру набула великої популярності. Комунікаційний центр, або контакт центр (від англ. call center - центр обробки дзвінків) є спеціалізованою організацією, яка займається обробкою запитів та

інформуванням телефонними каналами зв'язку в інтересах головної організації або організації-замовника. Контакт центр може працювати і в режимі інтернет-чату [1], обробляючи також звернення по електронній пошті e-mail і звичайній пошті, обробку інформації з факсів, що працюють зі зверненнями.

1.2 Аналіз проблем, пов'язаних з темпом мовлення в контакт-центрах

Значна кількість організацій і компаній використовує центри обробки дзвінків для взаємодії зі своїми клієнтами, а також для супроводу та організації процесів продажу продуктів і послуг, управління сервісним обслуговуванням, керування критично важливими процесами, технічної підтримки тощо. На фірмах виробниках товарної продукції комунікаційним центрам може бути також поставлена функція взаємодії з постачальниками, підрядниками чи споживачами. Багато суспільних організацій, таких як державні, громадські та політичні використовують комунікаційні контакт-центри для забезпечення телефонної взаємодії з громадянами. Великі підприємства і організації покладають на контакт-центри виконання деяких внутрішніх функцій, наприклад, забезпечують технічну та офісну підтримку працівників [2].

Під час функціонування комунікаційних центрів різного призначення виникають проблеми різного типу, які потрібно вирішувати. Наприклад, телефонні дзвінки, за думкою спеціалістів, є однією з основних причин втрати часу керівника. До таких втрат призводить недосконала підготовка до телефонної розмови, невміння виділяти в ній головне, стисло та лаконічно викладати свої думки. Це призводить до значних втрат робочого часу, що складає від 20 до 30 відсотків всього робочого часу. А в комунікаційних центрах критичних системЮ таких як диспетчерські служби на залізниці, авіасполученні, на атомних електростанціях, у військових центрах зв'язку такі проблеми можуть приводити до важких наслідків.

Спілкування по телефону проводиться між двома суб'єктами, які поперемінно приймають позиції "слухача" і "мовця". Людям зазвичай, притаманна властивість більше говорити, ніж слухати, і це почасти перетворює діалог в монолог і призводить до неефективного використання часу комунікації. "Слухач" при цьому потребує великого терпіння, що в багатьох ситуаціях піддається серйозному випробуванню.

Також за оцінкою психологів, час телефонних розмов під впливом емоційного стану може значно збільшуватися, а значна тривалість спілкування є запорукою мовної нечіткості та не діловитості[3]. Також емоційний стан мовця досить часто стає причиною несприйнятливості до реакцій "слухача". Оскільки подібні реакції є не завжди однозначними, диспетчеру комунікаційного центру потрібно визначити ставлення "слухача", що викликає необхідність для нього займатися сторонніми міркуваннями, які не відносяться до суті ініціалізованої комунікації, наприклад:

- чи представляє для нього інтерес тема розмови;
- наскільки вірно він розуміє він зміст;
- чи має він бажання висловити свою думку;
- чи не затримує його затіяне спілкування..

У випадку, коли телефонна розмова триває довше, чим потрібно для уладнання справи, між сторонами виникає емоційна напруженість. Це значить, що існує визначений період часу спілкування, вихід за межі якого здатний привести до неприємних наслідків, зокрема, до втрати потреби в спілкуванні та навіть, до конфлікту. Тому оператор контакт-центру повинен враховувати, що для різних людей ця міра неодинакова [4].

Однією з найбільш поширених причин непорозуміння вважається такий показник, як темп мовлення. Оскільки різні люди схильні по різному (з різною швидкістю) спілкуватися та висловлювати при цьому свої думки та бачення, то даний факт призводить до конфліктів різного виду, а також непорозумінь. Тому для того, щоб спілкування з абонентом було ефективним, оператору

комунікаційного центру необхідно налаштовуватися на темп мовлення співрозмовника [5].

Розглянемо питання, що розуміють під темпом мовлення. 1.1.4 Поняття про темп мовлення.

В загальному випадку під темпом мовлення розуміють швидкість проголошення елементів мовлення (звуків, складів, слів) [6]. Темп мовлення відноситься до просодичних (не фонетичних, над сегментних) елементів [6], і є одним з компонентів інтонації. На абсолютний темп мови впливають багато факторів, таких, як індивідуальні риси мовця, стиль вимови, особливості його емоційного стану чи ситуації спілкування. Також існує залежність між довжиною мовленнєвої одиниці і швидкістю її вимовлення, що породжується прагненням до ізохронності: чим довші синтагми або слова, тим в них середня тривалість звуку (склади) менша [6].

Темп мовлення може вимірюватися :

- числом вимовлених в одиницю часу (зазвичай в секунду) елементів мови;
- середньою тривалістю елемента.

В математичному відношенні, швидкість мовлення в основному визначається кількістю вимовлених складів за секунду (іноді — кількістю слів за хвилину; також може використовуватися вимірювання швидкості за середньою тривалістю складів чи слів). Загальний діапазон швидкості мовлення людини оцінюється від 2 до 12 складів на секунду [7,8].

Збільшення швидкості мови досягається шляхом скорочення як голосних, так і приголосних звуків, тоді як зниження швидкості, головним чином, досягається подовженням тривалості голосних [9].

У ролі інтонаційного засобу швидкість мови виконує функцію виділення більш та менш важливих елементів висловлення. Слова чи фрази, які мають найбільший важливий зміст, характеризуються сповільненим темпом, тоді як відрізки, що несуть менш суттєву інформацію (наприклад, вступні слова), висловлюються прискореним темпом. Зміна темпу до кінця висловлення

служить одним з методів створення його цілісності [1], відображаючи синтаксичну [4] роль інтонації.

Темп може також тлумачитися як характеристика, яка визначає рівень напруженості артикуляції та виразності мовлення. Наприклад, при сповільненому темпі слова вимовляються в повних звукових формах [5].

Розрізнення темпу мовлення, темпу фрази, темпу синтагми (мовного сегмента) та темпу фонетичного слова є доцільним. Вимірювання темпу мовлення базується на підрахунку кількості звукових одиниць (звуків, складів, слів), вимовлених протягом одиниці часу. Іноді темп визначається за середньою тривалістю звукової одиниці. Отже, при характеристиці конкретного темпу мовлення можна вказати, що кількість промовлених складів за одну секунду становить 2,8 або що середня тривалість складу при такому темпі дорівнює 352 мс (мілісекунди). Також виділяють "абсолютний" темп мовлення (темп артикуляції) - кількість сегментів та тонованих звуків в одиницю часу у мовленні. Оптимізація мовного спілкування тісно пов'язана з вибором оптимального темпу мовлення, який може бути повільним, швидким або помірним, залежно від умов. Зазвичай виділяють повільний та швидкий темп мовлення, протиставляючи їх середньому темпу мовлення, який можна називати помірним. Важливо враховувати, що оптимальний темп мовлення може змінюватися в залежності від конкретних обставин. Середній темп мовлення для нашої мови приблизно складає 120 слів за хвилину. В освітньому процесі також існують встановлені норми: учитель в V-VI класах, як правило, вимовляє не більше 60 слів за хвилину, а в XI класі - не більше 75 слів за хвилину.

Швидкість мовлення залежить від індивідуальних характеристик мовця, його темпераменту, готовності до розгорнутого вислову, а також від контексту промови та ситуації спілкування. Темп артикуляції зростає при висловленні мовних одиниць, які є рутинними та легко використовуються в активному словниковому запасі мовця, таких як стійкі словосполучення, усталені вислови та вигуки. Приватні розмови, особливо на випадкові теми, часто

відбуваються на швидшому темпі, ніж публічні виступи, зокрема уроки. Цікаво, що чим важливіший контент, тим повільніше проходить мовлення. Темп мовлення, особливо під час лекцій, може значно сповільнюватися, особливо якщо аудиторія зайнята записами. Коли мова висловлюється в великому залі, ехо може призводити до накладання звукових слів, і в таких випадках рекомендується говорити повільніше.

Уповільнення темпу мовлення може бути реалізоване двома методами: через збільшення тривалості звуків та шляхом вимови фраз з паузами між словами [10].

На відміну від цього, прискорення темпу мовлення досягається шляхом скорочення тривалості звуків (збільшенням швидкості артикуляції); у випадку швидкого темпу мовлення паузи між словами стають коротшими або анімаційно зникають. Прискорення, як правило, впливає на приголосні більше, ніж на голосні. Це пояснюється тим, що при їх вимові артикуляційні рухи є більш компактними і, отже, можуть бути здійснені швидше. Одночасно з цим прискорення темпу мовлення може призводити до кількісної та якісної редукції голосних. Крім того, іншим наслідком прискорення мовлення може бути втрата деяких звуків, що є явищем, досить поширеним у повсякденному мовленні.

Багато осіб виражають своє мовлення в однаковому темпі у будь-яких обставинах. Однак мало хто приділяє увагу своїм слухачам в конкретний момент. Мовні особливості різних регіонів світу можуть впливати на швидкість мовлення людей, деякі можуть говорити повільніше, інші - швидше. У сфері бізнесу дуже важливо вміти адаптувати свій темп мовлення для підтримання рівної комунікації і забезпечення чіткості мовлення для аудиторії. Цей навичка передбачає вміння знаходити золоту середину між занадто великою швидкістю і занадто повільним темпом мовлення. Зберігаючи рівний темп, ви можете ефективно спілкуватися з різноманітними людьми. В той самий час, ця навичка підкреслює, що іноді важливо змінювати швидкість мовлення залежно від аудиторії, з якою ви взаємодієте.

Темп мовлення відіграє важливу роль у досягненні взаєморозуміння з іншими людьми, сприяючи побудові ефективних робочих відносин, заснованих на взаємній довірі та повазі. Здатність адаптувати свій темп до темпу мови співрозмовника свідчить про якісні міжособистісні відносини. Коли взаєморозуміння присутнє, імовірність правильного розуміння вашого повідомлення збільшується.

Однак помилки з темпом можуть виникнути, наприклад, якщо мова є надто швидкою, що може свідчити про схвильованість, але втрачає чіткість та переконливість у висловлюваннях, що ускладнює сприйняття вашого повідомлення.

Навпаки, занадто повільне мовлення може викликати невпевненість та втрату інтересу. Загальна порада полягає в умінні змінювати темп в залежності від аудиторії, зберігаючи баланс між швидкістю та зрозумілістю. Також важливо уникати надмірного використання пауз, вигуків та слів-паразитів, щоб утримати плавність мовлення та уникнути втрати уваги слухача.

1.3 Аналіз програмних рішень

На сьогодні існує низка програмних рішень, спрямованих на аналіз просодичних параметрів мовлення, зокрема, мовленнєвого темпу в критичних ситуаціях, які можуть бути корисними в різних застосуваннях.

До них в першу чергу можна віднести наступні:

1) **Praat** - безкоштовне програмне забезпечення для аналізу звуку та мовлення [11]. Воно дозволяє вимірювати параметри мовлення, такі як темп, інтонація, паузи та інші аспекти акустичного сигналу.

Це досить потужний багатофункціональний інструмент, який доступний для засвоєння користувачами будь-якого рівня кваліфікації. Основним завданням програми Praat є аналіз, маніпуляція та синтез мови. Крім цього, у програмі реалізовано можливість створення високоякісних графічних об'єктів для подальшого їх експорту. Praat був створений як програма-помічник для

створення наукових праць. Проте висока якість цього аналізатора спричинила популярність софту.

Графічний інтерфейс програмного середовища Praat поданий на рисунку 1.1.

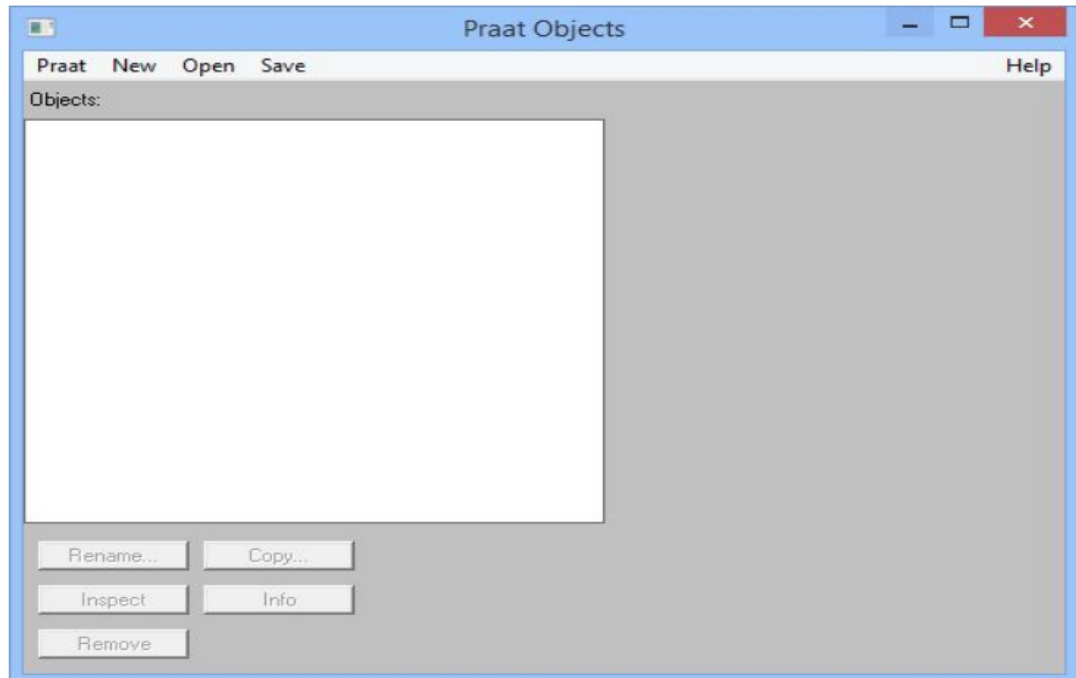


Рисунок 1.1 – Інтерфейс програмного середовища Praat

З рис. 1.1 видно, що інтерфейс інструменту відрізняється відсутністю зайвих елементів і строгістю. Це дає можливість в нескладному засвоєнні принципів роботи з ним, оскільки відволікаючих допоміжних опцій тут немає. Робоча область Praat поділена на дві частини: у першому вікні знаходяться всі інструменти для керування проектом, друге призначено для генерування зображення, яке стає доступним після аналізу та синтезу аудіо. Програмне середовище призначено для використання спеціалістами в обробці мовленнєвих сигналів, які розуміються на акустиці і цифровій обробці звукових сигналів. Програмний інструмент забезпечений керівництвом з використання, що дає зручність в освоєнні роботи з ним. Додатково програма допомагає експортувати файли у формати AIFF, NIST, FLAC, WAV та AIFC. Praat має інструменти для роботи з каналами, перетворення аудіо на стерео та моно, а

також багато інших можливостей з аналізу звуку і представлення його результатів. До основних функціоналів середовища Praat відносяться:

- Аналіз та синтез сигналу мовлення;
- Обробка сигналів мовлення за допомогою різних фільтрів;
- Сегментація та маркування;
- Наявність вбудованої мови програмування;
- Автоматичне створення графіків.

При всьому різноманітті можливостей даний інструмент не може бути застосований для автоматичного визначення темпу мовлення, оскільки працює з фонограмами, а аналіз характеристик сигналу мовлення проводиться користувачем, а не комп'ютером.

2) **Vosk**. Vosk - це інструментарій для розпізнавання мовлення [12, 13].

Програмний пакет Vosk призначений для обробки звуку в реальному часі, що дозволяє реалізувати миттєву реакцію на команди. Також підтримує кілька популярних мов програмування – Java, C#, Javascript.

Підтримує більше 20 різних мов та діалектів - українську, англійську, російську, індійську, німецьку, французьку, португальську, італійську, іспанську, китайську, турецьку, чеську, в'єтнамську, голландську, валенсійську, філіппінську, арабську, грецьку, перську, узбецьку, хінді, польську, корейську. В перспективі будуть додані й інші.

Може працювати без доступу до мережі, в тому числі і на таких мобільних пристроях, як Android, Raspberry Pi, iOS. Встановлюється за допомогою простої команди `pip3 install vosk` без додаткових кроків.

Моделі для кожної мови займають всього 50Мб, але є і більш точні великі моделі для більш точного розпізнавання.

Дозволяє швидко налаштовувати словник розпізнавання для покращення точності розпізнавання.

Додаткова функція - ідентифікація мовця.

Дане програмне середовище хоча і працює зі звуком в реальному масштабі часу, але автоматично визначити темп мовлення не дозволяє.

3) Програмний продукт **Acapela Group** призначений для синтезу мовлення з тексту. До основних його функціональних можливостей відносяться: навігація, розмова в Інтернеті, інтерактивна голосова відповідь, рішення для спеціальних можливостей, електронне навчання, аудіо книги. В його складі постійно з'являються нові програмні рішення, які розширюють його функціональні можливості. Сфера застосування даного програмного продукту – синтез мовлення, тому його використати в задачі автоматичного визації контролю і регулювання мовлення.

4) Програмний продукт **Voice Analytics** призначений для кол-центрів для реалізації процесів обробки природної мови (NLP), розпізнавання голосу та інших служб штучного інтелекту, призначених для запису голосових викликів, перекладу мови в текст і аналізу розмов.

Програмне забезпечення для аналізу голосу також визначає емоції та наміри мовця, вивчаючи його звукові шаблони. Ці аспекти роблять програмне забезпечення голосової аналітики корисним інструментом для контакт-центрів. Аналітики якості отримують глибоке розуміння потреб клієнтів, а також використовують це для моніторингу ефективності агентів.

Програма «Голосова аналітика» є розвинутою технологією, яка декодує, аналізує та робить висновки з голосових даних, зібраних під час взаємодії з клієнтом, щоб отримати розуміння та покращити взаємодію з клієнтами.

Розгортання програмного забезпечення голосової аналітики є неймовірно вигідним для будь-якого контакт-центру. При правильному та цілеспрямованому використанні компанії можуть дуже швидко вирішити навіть найскладніші проблеми. До речі, ШІ контакт-центру використовує автоматичне розпізнавання мовлення та обробку природної мови (NLP) для транскрибування та аналізу 100% взаємодії агента.

1) Забезпечує відповідність нормативним вимогам: визначаючи використання конкретних ключових слів у стенограмах, голосова аналітика відстежує дотримання процесу та відповідність для окремих агентів.

2) Вимірює продуктивність агента: програмне забезпечення голосової аналітики вимірює такі аспекти, як настрої та час утримання, надаючи розуміння якості агента.

3) Аналізує реакцію на нову політику: тон голосу запитань клієнта та його загальні відповіді допомагають оцінити ефективність, скажімо, нової кампанії бренду або зміни політики компанії.

4) Записує скарги: голосова аналітика висвітлює причини незадоволення клієнтів і пропонує рішення для вирішення їхніх проблем.

5) Знайдіть ідеї для нових продуктів: якщо є попит на нові продукти та послуги, голосова аналітика може сканувати точку зору клієнтів за допомогою аналізу мови та тону.

6) Перевага над конкуренцією. Аналіз відповідей або скарг клієнтів корисний для визначення стратегії конкурентів.

7) Економічна ефективність: довіра до аналізу голосу зводить нанівець необхідність для менеджерів і груп із забезпечення якості вручну виконувати завдання транскрипції, заощаджуючи величезну кількість часу та коштів, понесених компанією

Програма голосової аналітики зменшує ручну роботу QA та забезпечує високу точність, швидкість і підвищену ефективність. Завдяки програмному забезпеченню для перетворення мовлення в текст із чудовою точністю транскрипції, ця програма може змінити правила роботи для будь-якого контакт-центру та отримати перевагу перед конкурентами у створенні клієнтського досвіду світового рівня.

Хоча даний програмний продукт був розроблений спеціально для контакт-центрів диспетчерських служб різного призначення, в ньому не передбачена функція автоматичного контролю темпу мови.

Text Ananlyzer- програма для аналізу незначущого вмісту в тексті, так званої “води”. Може використовуватися в контакт-центрах опосередковано, оскільки потрібно спочатку перетворити аудіо записи телефонних розмов в

текст. Вона сприймає на вхід текст, призначений для аналізу, і по команді “Аналізувати” видає його характеристику.

Аналіз Технології **Google Cloud Speech-to-Text** - інструменти для розпізнавання мовлення стають необхідною складовою для інновацій та вдосконалення комунікаційних процесів. У цьому контексті Google Cloud Speech-to-Text виступає як ключовий гравець, забезпечуючи потужний інструментарій для перетворення аудіоданих у текст у реальному часі. Google Cloud Speech-to-Text надає унікальну можливість розпізнавання мовлення в реальному часі.

Це забезпечує високу ефективність у використанні, надаючи користувачам можливість отримувати текстовий вихід негайно, що є надзвичайно важливим у сферах, де швидка реакція визначає успіх. Однією з основних переваг є висока точність розпізнавання мовлення. Google Cloud Speech-to-Text використовує передові алгоритми машинного навчання та нейронні мережі для адаптації до різних мов, акцентів та індивідуальних варіацій у вимові, що робить його ефективним в різноманітних мовленнєвих сценаріях. Значущою перевагою є широкий спектр підтримуваних мов та діалектів, що визначає універсальність інструменту. З Google Cloud Speech-to-Text можна взаємодіяти у мовленнєвих процесах на різних географічних ринках, забезпечуючи глобальну доступність та використання. Однією з ключових особливостей є можливість інтеграції Google Cloud Speech-to-Text у різноманітні додатки та сервіси. Це робить його важливим компонентом для розробників, які прагнуть покращити можливості своїх продуктів у сферах, де обробка мовлення відіграє визначальну роль.

У висновку можна зазначити, що Google Cloud Speech-to-Text представляє собою технологічний прорив у сфері розпізнавання мовлення. Завдяки своїй здатності працювати в режимі реального часу, високій точності та гнучкості у використанні, він стає невід'ємною складовою для інноваційних проектів та розвитку сучасних комунікаційних систем.

Microsoft Azure Speech SDK - надає високорівневий інтерфейс для взаємодії з потужними засобами розпізнавання мовлення в хмарному середовищі Azure.

Microsoft Azure Speech SDK є ключовою складовою хмарної інфраструктури Azure, надаючи користувачам можливість легко інтегрувати розпізнавання мовлення у свої додатки та сервіси, що прискорює розробку та розгортання.

Однією з ключових переваг є широкий спектр підтримуваних мов та діалектів, що робить цей інструмент глобально придатним для використання в різних лінгвістичних спільнотах. Крім того, підтримка мовленнєвих функцій, таких як інтонація та темп мовлення, розширює можливості аналізу.

Microsoft Azure Speech SDK дозволяє виконувати розпізнавання мовлення в реальному часі, що визначає його застосування в сферах, де крайова швидкість і відгук важливі для досягнення успіху. Здатність обробляти різноманітні формати аудіоданих робить його універсальним для використання в різних сценаріях.

Важливим аспектом є легка інтеграція Microsoft Azure Speech SDK з іншими сервісами хмарного середовища Azure. Це дозволяє розробникам створювати комплексні рішення, поєднуючи розпізнавання мовлення з іншими функціональними можливостями платформи.

Microsoft Azure Speech SDK знаходить своє використання в різних секторах, включаючи технології розпізнавання голосових команд, реалізацію систем автоматизованого обслуговування клієнтів та багато іншого. Розвиток інтеграції з іншими службами та постійне оновлення функціональності роблять Microsoft Azure Speech SDK важливим інструментом для розробників та бізнес-аналітиків у сферах штучного інтелекту та розпізнавання мовлення.

Microsoft Azure Speech SDK визначається своєю високою продуктивністю, гнучкістю та відмінною інтеграцією з іншими хмарними послугами Azure. Використання цього інструменту сприяє ефективній розробці та впровадженню рішень, що базуються на розпізнаванні мовлення,

розширюючи горизонти комунікацій та автоматизації у різних сегментах індустрії.

Amazon Transcribe - стає ключовим інструментом, що пропонує ефективний та масштабований підхід до автоматизації процесу перетворення аудіоданих у текстовий формат.

Amazon Transcribe інтегрується в екосистему хмарних послуг Amazon Web Services (AWS), надаючи користувачам зручний та масштабований інтерфейс для розпізнавання мовлення. Завдяки хмарній інфраструктурі, використання Amazon Transcribe може легко масштабуватися від індивідуальних проєктів до великих корпоративних систем.

Однією з ключових переваг технології є її здатність працювати з різними мовами та діалектами. Amazon Transcribe підтримує широкий спектр мов, що робить його універсальним інструментом для глобально орієнтованих проєктів. Застосування передових алгоритмів машинного навчання дозволяє Amazon Transcribe адаптуватися до різних аспектів мовлення, включаючи акценти, темп та інтонацію. Це забезпечує високу точність розпізнавання в різних сценаріях.

Однією з важливих переваг Amazon Transcribe є його легкість інтеграції в різноманітні проєкти. За допомогою API та SDK, ця технологія може бути впроваджена у власні додатки, веб-сервіси чи корпоративні системи, надаючи флексібельність використання.

Amazon Transcribe знаходить своє використання у різних сферах, включаючи транскрипцію аудіо- та відеофайлів, розробку віртуальних асистентів та автоматизацію багатьох галузей бізнесу. З постійними оновленнями та додатковими функціональностями, Amazon Transcribe продовжує розширювати свої можливості.

Amazon Transcribe є важливим гравцем на ринку розпізнавання мовлення, пропонуючи рішення, що враховує потреби сучасного користувача. З його допомогою, організації можуть ефективно використовувати голосовий

контент, поліпшуючи комунікацію та забезпечуючи нові перспективи для розвитку інноваційних проектів.

Таким чином, проведений аналіз сучасних програмних продуктів для обробки мовленнєвого сигналу показує, що при всіх їхніх перевагах вони не здатні вирішити задачу автоматизації контролю і регулювання темпу мовлення. Тому в магістерській роботі ставиться задача пошуку ефективних методів обробки аудіосигналу для автоматичного визначення темпу мовлення. При розв'язанні цієї задачі будемо орієнтуватися на існуючі дослідження в даному напрямку.

1.4 Аналіз аналогів і прототипів

В якості аналога взято розробку співробітників і студентів Білоруського державного університету інформатики та радіоелектроніки, які запропонували систему контролю темпу мови на основі методу, алгоритму і пристрою, запропонованих і запатентованих проф. кафедри КСУ ВНТУ Биковим М.М.

В загальному цей алгоритм визначення темпу мови складається з наступних етапів:

- нормування мовного сигналу. Забезпечення вирівнювання слабких (тихих) сигналів для того, щоб виключити залежність результатів виміру від гучності вхідного мовного сигналу;
- надійне виділення сегментів складів у сигналі мовлення;
- оцінка тривалості складових сегментів, формування головних ознак;
- виділення і вимірювання тривалості пауз;
- прийняття рішення про темп мови [14].

Головний недолік такого алгоритму полягає в тому, що визначення темпу мовлення не ґрунтується виключно на середній кількості складів за одиницю часу, як це вимагають стандарти, а також враховує паузи.

Розглянемо методи обробки мовного сигналу, які потрібно використати для реалізації поставлених задач.

1.5 Аналіз методів обробки мовного сигналу з метою визначення швидкості мовлення

1.5.1 Поняття про сигнал мовлення

Основною функцією мови є забезпечення спілкування. Існують різні підходи до пояснення можливостей мови, одним із яких є теорія інформації, розроблена Шенноном. В рамках цієї теорії мову можна описати за її інформаційним вмістом. Інший підхід до опису мови полягає в розгляді її як сигналу, тобто акустичних коливань [15].

Процес мовленнєвого спілкування розпочинається тим, коли абстрактні повідомлення виникають у мозку диктора. Під час створення мовлення ці повідомлення перетворюються в акустичні мовленнєві коливання. Інформація, закладена в повідомленні, представлена у формі складних акустичних коливань. Спочатку повідомлення конвертується в послідовність нервових імпульсів, які керують артикулярним апаратом, таким як рух язика, губ, голосових зв'язок. Цей процес викликає активацію артикулярного апарату, що призводить до виникнення акустичних мовленнєвих коливань, які несуть інформацію про вихідне повідомлення [16].

Повідомлення, яке передається мовним сигналом, є дискретним, тобто може бути вираженим послідовністю символів певної кількості. Символи, які становлять мовний сигнал, отримали назву фонем [17]. У системах мовного зв'язку сигнали передаються, зберігаються і обробляються за допомогою різних методів. Проте у всіх випадках їх характеризують наступні особливості:

- подання мовленнєвого сигналу у зручній для передачі та збереження формі, або у формі, яка спрощує конвертацію мовного сигналу без великих втрат інформації.

1.5.2 Характеристика мовного сигналу як носія інформації

Мовний сигнал виступає як носій інформації, і його основною особливістю є здатність одночасно передавати різні аспекти інформації. Важливими компонентами інформаційної структури мовного сигналу є сутність повідомлення, характеристики емоційного та фізичного стану диктора та його індивідуальні особливості. Інформаційне кодування здійснюється шляхом модуляції різних параметрів, таких як інтенсивність мовного сигналу, амплітуди та фази його спектральних компонентів, частота першої гармоніки. Важливо враховувати, що один і той самий вид інформації може модулювати не один параметр, таким чином, смисловий зміст мови передається через зміну в часі ряду перерахованих параметрів мовного сигналу. Це включає в себе зміни в джерелі звукових коливань та формі миттєвих спектрів мовного сигналу. Інформація про індивідуальні особливості диктора, яка виражається в висоті тону, тембрі та особливостях вимови звуків, закодована частотою першої гармоніки (основною частотою) і формою короткочасних спектрів. Фізичний та емоційний стан, які характеризуються тембром, висотою тону, гучністю та швидкістю вимови, закодовані інтенсивністю, формою короткочасного спектру, частотою основної гармоніки та тривалістю сигналу. Записи звукового сигналу підтверджують, що для одного і того ж повідомлення параметри мовного сигналу значно коливаються. Наприклад, тривалість слова "слухай" при вимові другим диктором збільшилась на 100 мсек в порівнянні з першим диктором, а тривалість окремих звуків також варіюється в значній мірі. Амплітуда спектральної компоненти та її інтенсивність значно змінюються. Також спостерігається вплив особливостей середовища передачі на розподіл різних звуків: вокалізовані звуки затухають менше, ніж шуми, і інтенсивність сигналу в паузі змінюється слабше з відстанню. Отже, дослідження мовного сигналу, проведене в магістерській кваліфікаційній роботі, показує, що смислова інформація, призначена для декодування в системах автоматичного розпізнавання мови, подається в акустичних коливаннях дуже складно через

вплив параметрів сигналу та інших аспектів інформації, а також властивостей середовища передачі інформації. Оскільки параметри середовища передачі не зрозумілі для багатьох видів інформації та зрозумілі для фонетичних компонентів звуків, які мають відношення до сенсорної інформації, виникає потреба розробки математичних відношень, що описують залежність передаточних характеристик середовища від фонетичної природи звуків мовлення, що передаються. Отримані залежності дозволяють уточнити існуючі моделі мовотворення і, таким чином, визначити додаткові параметри мовного сигналу коваріантної та семантичної інформації.

1.5.3 Цифрова обробка мовлення

У випадку мовних сигналів джерелом інформації є людина. Зазвичай акустичні коливання піддаються спостереженню та вимірюванню. Схема обробки сигналів мовлення представлена на рисунку 1.2. Одним з ключових аспектів є перетворення сигналів мовлення в цифрову форму, що є фундаментальним аспектом теорії цифрової обробки.

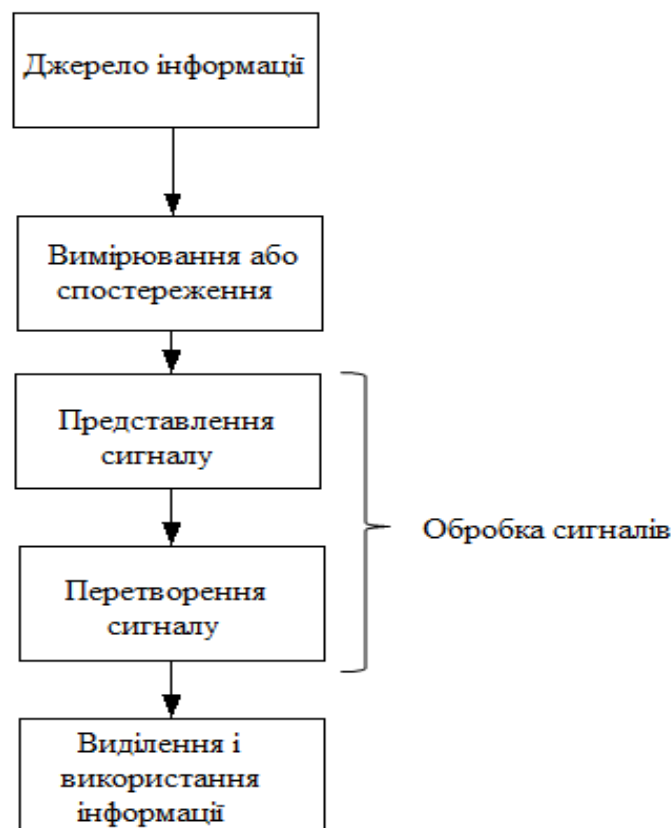


Рисунок 1.2- Схема оброблення сигналу мовлення

Існує низка методів дискретного представлення мовних сигналів. Як зображено на рисунку 1.3, ці способи можуть бути розбиті на групи [18].



Рисунок 1.3 –Методи представлення сигналу мовлення

На рисунку 1.4 зображено декілька напрямків застосування мовленнєвих технологій.

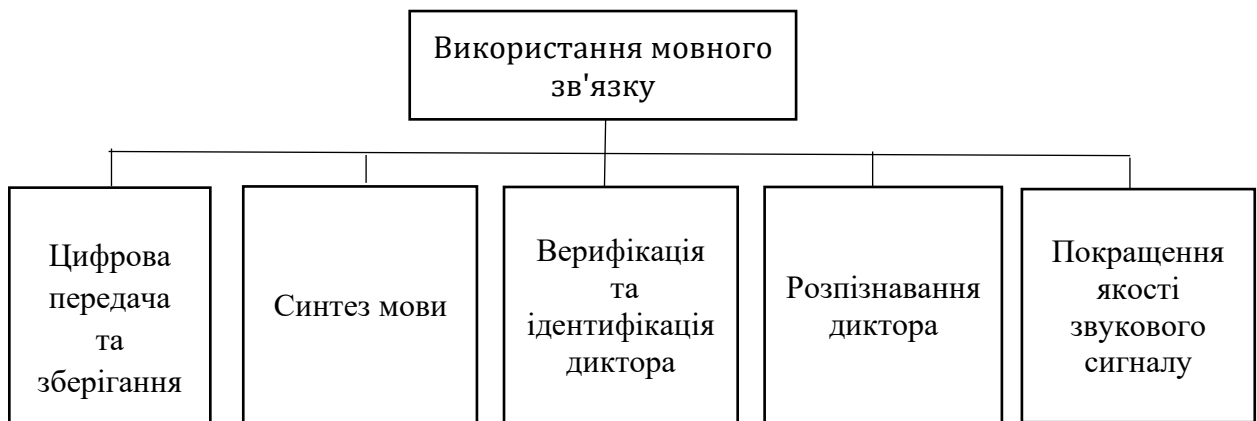


Рисунок 1.4 – Сфери застосування мовленнєвих технологій

1.5.4 Визначення темпу мови

Визначення темпу мови — це процес сегментації неперервного потоку мови на склади, визначення і вимірювання тривалості голосних звуків, а також виділення пауз в умовах шумів, характерних для телефонної лінії. Це один з напрямків використання мовленнєвих технологій. На сьогодні існує велика кількість алгоритмів, які базуються на традиційній обробці мовного сигналу в

частотній або часовій області, включаючи формантні характеристики. Також використовуються альтернативні методи, такі як приховані марківські моделі (ПММ). З метою досягнення максимальної швидкості роботи і можливості використовувати в багатоканальних системах обробки викликів, запропоновані алгоритми базуються на обробці мови в часовій області. Цей підхід дозволяє досягти прийнятної швидкості роботи при мінімальному використанні комп'ютерних ресурсів, що особливо важливо в багатоканальних системах обробки викликів з десятками каналів.

1.5.5 Аналіз методів виділення складових сегментів сигналу мови

Один з методів побудови систем розпізнавання, які не залежать від конкретного диктора, включає використання ієрархічної системи на верхніх рівнях. Ця система використовує елементи розпізнавання звукотипів, чий фонетичні характеристики слабо залежать від особливостей диктора і контексту, таких як склади, півсклади, фрикативні та дзвінки звуки, паузи і інші. Дослідження [24] демонструє, що використання лише інформації про склади на верхньому рівні розпізнавання дозволяє зменшити кількість кандидатів для класифікації в 2-4 рази. Одним із ключових параметрів для визначення складів у мовному сигналі є його енергія [21, 22]. У даному випадку ядро складу визначається в області, де максимуми енергії обмежені значними (40 дБ або 50 дБ) спадами енергії. Втім, деякі дослідження, такі як [7], вказують на часті помилкові виділення за цим параметром, особливо при наявності високоенергетичних фрикативних або сонорних звуків. Це підтверджується експериментально знятими записами. У роботі [24] в якості параметра для визначення ознак складу використовується функція "гучності", яка отримується зваженою сумою амплітуд сигналів у 22 частотних каналах, розміщених у критичних смугах. Однак великі апаратурні чи обчислювальні витрати та недостатня надійність є недоліками цього методу формування ознак складу.

З метою уникнення недоліків, що властиві зазначеним методам, у

цьому дослідженні ставиться завдання розробити новий метод. Цей новий підхід має на меті покращити надійність процесу виділення сегментів мовного сигналу, що відповідають складам мови. Також передбачається підвищення надійності пристрою та алгоритму, які використовують цей метод.

1.5.6 Аналіз методів виділення пауз у сигналі мовлення

- Аналіз літературних джерел [20, 22, 24] вказує на те, що у параметричному описі мовного сигналу використовують такі параметри:

- функція енергії сигналу;
- короткочасна функція середнього числа переходів через нуль
- короткочасна енергія і короткочасне середнє значення сигналу

Ці параметри використовуються в методах обробки мови, які базуються на припущенні, що властивості сигналу мовлення повільно змінюються з часом. Короткочасний аналіз враховує це припущення, виділяючи та обробляючи сегменти сигналу мовлення як короткі ділянки окремих звуків із різними властивостями. Ця процедура повторюється на кожному інтервалі (кадрі) аналізу, які частково перетинаються. Результатом обробки на кожному інтервалі є сукупність чисел або одне число, що стає новою характеристикою мовного сигналу залежно від часу.

Більшість методів короткочасного аналізу можуть бути описані виразом:

$$Q_n = \sum_{m=-\infty}^{\infty} T[x(m)]\varpi(n-m). \quad (1.1)$$

Мовний сигнал може бути підданий перетворенню $T[\bullet]$, яке може бути як лінійним, так і нелінійним. Це перетворення може залежати від сукупності параметрів або від певного управляючого параметра. Результатом цього перетворення є послідовність, яку розширюється на значення тимчасового вікна. Це вікно розташоване в часі згідно з індексом n . Потім результати сумуються по всім ненульовим значенням. Зазвичай, хоча не завжди, послідовність значень тимчасового вікна має кінцеву протяжність. Q_n - це значення, що представляє "зважене" середнє значення послідовності $T[x(m)]$.

Короткочасна енергія обчислюється за наступним виразом:

$$E_n = \sum_{m=n-N+1}^n x^2(m). \quad (1.2)$$

Отже, короткочасна енергія в момент n є сумою квадратів N відліків від $n-N+1$ до n . З (1.1) видно, що в (1.2) $T[\bullet]$ є просто операція зведення в квадрат, а

$$\varpi(n) = \begin{cases} 1, & 0 \leq n \leq N-1, \\ 0, & \text{в протилежному випадку} \end{cases} \quad (1.3)$$

При опрацюванні сигналів у дискретному часі вважають, що коли двом послідовним відлікам присвоєно різні знаки, то відбувся перехід через нуль. У цьому випадку частота появи нулів у сигналі може служити найпростішою характеристикою його спектральних властивостей. Це найбільш справедливо для вузькополосних сигналів. Наприклад, синусоїдальний сигнал з частотою F_0 , підданий дискретизації з частотою F_s , має F_s/F_0 відліків за період. Кожний період містить два переходи через нуль, таким чином, середнє число нульових переходів за великий інтервал часу.

$$z = 2F_0 / F_s. \quad (1.4)$$

Середнє число нульових переходів можна вважати придатною оцінкою частоти синусоїдального коливання. Сигнал мовлення є широкосмуговим, а, отже, менш очевидна інтерпретація середнього числа переходів через нуль. Проте можна отримати приблизні оцінки спектральних властивостей сигналу, ґрунтуючись на використанні функції середнього числа переходів через нуль для мовного сигналу. Визначимо середнє число переходів через нуль:

$$Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| \varpi(n-m), \quad (1.5)$$

де

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0, \\ -1, & x(n) < 0 \end{cases} \quad (1.6)$$

і

$$\varpi(n) = \begin{cases} 1/2N, & 0 \leq n \leq N-1, \\ 0, & \text{в протилежному випадку} \end{cases} \quad (1.7)$$

Амплітуда мовного сигналу значно змінюється в часі. Зокрема, амплітуда нелокалізованих сегментів сигналу мовлення відчутно менше амплітуди локалізованих сегментів. Такі зміни в амплітуді ефективно описуються за допомогою функції короткочасної енергії сигналу. У загальному випадку, функцію енергії можна визначити як

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)\varpi(n-m)]^2 \quad (1.8)$$

Цей вираз може бути переписаний у наступний вигляд:

$$E_n = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m), \quad (1.9)$$

$$\text{де } h(n) = \varpi^2(n) \quad (1.10)$$

В роботах [24, 26] було показано, що велике число переходів через нуль (2.5) не є достатньо ефективним показником. Таким чином, використання окремо енергії звуку або кількості переходів через нуль не є надійним методом визначення пауз в мовному сигналі.

1.6 Висновки до розділу

Проведений в даному розділі аналіз проблеми автоматизації контролю і регулювання темпу мови в комунікаційних центрах критичних систем показав, що вирішення даною проблеми є достатньо актуальною задачею, оскільки існуючі програмні і апаратні засоби для обробки сигналів мовлення не мають засобів для реалізації даної функції.

Тому в роботі ставляться і вирішуються наступні задачі:

- Провести дослідження методів і моделей обробки мовленнєвого сигналу для реалізації надійного виділення складів в мовленнєвому потоці, запропонувати методи локалізації складових сегментів і підрахунку їх кількості;
- Провести дослідження аудіофайлів і запропонувати метод визначення оптимальної кількості градацій темпу мови;

- Розробка алгоритмів, пристроїв і програмного забезпечення для автоматичного визначення темпу мови та його управління в комунікаційних центрах критичних систем з метою оптимізації ефективності та забезпечення безпеки їх функціонування.

2 РОЗРОБКА МЕТОДУ І ПРИСТРОЮ ВИДІЛЕННЯ СКЛАДІВ В СИГНАЛІ МОВИ

2.1 Моделювання обробки сигналу мови слуховою системою

Можливість розкриття механізмів відображення спектральних параметрів мовного сигналу в просторі інваріантних ознак можлива шляхом розширення взаємодії процесів в периферійній слуховій системі за допомогою включення процесів в центральній слуховій системі, що відображається на наступній структурній схемі моделі слуху (рис. 2.1).



Рисунок 2.1 - Модифікована модель слухової системи

На рис. 2.2 зображена модель нейронної мережі МНМ.

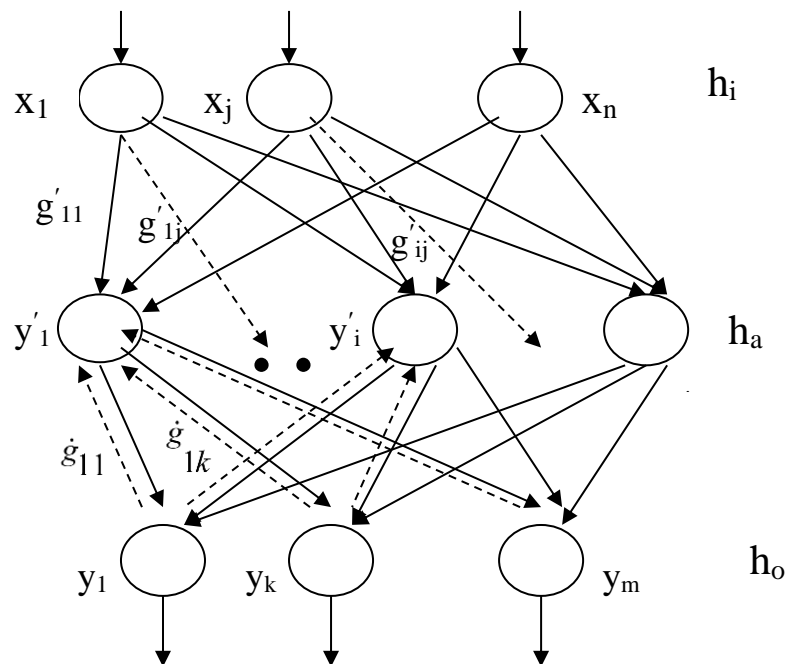


Рисунок 2.2 – Спрощена модель нейронної мережі слухової системи

В наведеній на рисунку 2.1 узагальненій слуховій моделі, блок спектрального аналізу (БСА) ілюструє частотно-вибіркові характеристики моделі та складається з набору третьоктавних фільтрів із центральними частотами, розташованими вздовж частотної осі відповідно до логарифмічного закону [27].

Модель сенсорних слухових нейронів (МССН) відтворює роботу слухових нейронів, які взаємодіють із волосковими клітками базилярної мембрани вуха. Механізми слухового сприйняття достатньо добре вивчені, однак спроби їх застосування не дали бажаних результатів.

Розглянемо особливості формування мовних образів моделлю нейронної мережі (МНМ). На рис. 2.2 приведена нейронна мережа, яка складається з трьох шарів: вхідного h_i , який відповідає нейронам-рецепторам, що пов'язані зі слуховими клітинами, аналізуючого h_o , та реагуючого h_a , що моделює процес прийняття рішення (розпізнавання образу).

Вихідні сигнали X_j , $j = \overline{1, n}$ нейронів вхідного шару пропорційні обвідним вхідних сигналів фільтрів БСА. Вихідний сигнал i -го нейрона Y'_i аналізуючого шару створюється інтегруванням наступного диференційного рівняння [28-30]:

$$\frac{d y'_i}{dt} = \sum_{j=1}^n q'_{ij} \cdot x_j - \gamma(y'_i), \quad (2.1)$$

де $\bar{X} = \{x_1, x_2, \dots, x_j, \dots, x_n\}$ – вхідний образ, вихідні сигнали фільтрів БСА;

g'_{ij} - сила зв'язку між i -им нейроном аналізуючого шару і j -им нейроном вхідного шару;

y'_i - вихідний сигнал i -го нейрона аналізуючого шару;

$\gamma(y'_i)$ - складова, яка робить передаточну характеристику нейрона нелінійною і враховує нелінійні ефекти. Вважається, що існує функція f_G , обернена γ , тобто $f_G = \gamma^{-1}$.

Стала інтегрування нейроної клітини має величину порядку декількох десятків мсек [31]. Враховуючи, що на такому проміжку вихідні сигнали x_i БСА для мови стаціонарні, можна записати:

$$\frac{dy'_i}{dt} \approx 0, \quad (2.2)$$

Звідки

$$\gamma(y'_i) = \sum_{j=1}^n q'_{ij} \cdot x_j, \quad (2.3)$$

$$y'_i = f_G \left(\sum_{j=1}^n q'_{ij} \cdot x_j \right). \quad (2.4)$$

У відповідності з рівняннями (2.11) і (2.14), при подачі на вхід рецепторних слухових нейронів спектра \bar{X} мовного сигналу величина початкового збудження i -того нейрона аналізуючого шару пропорційна добутку

$$\vec{g}_i^T \cdot \vec{x}, \quad (2.5)$$

де $\vec{g}_i^T = (g_{i1}, g_{i2}, \dots, g_{in})^T$ – вектор-стовбець ваг зв'язків i -того нейрона;

$\vec{x} = (x_1, x_2, \dots, x_j, \dots, x_n)$ – вектор спектральних параметрів сигналу.

Ядро активності в аналізуючому шарі формується для тих нейронів, для яких добутки (2.15) максимальний. Необхідний для розпізнавання мовних образів розподіл ваг зв'язків може бути встановлений в результаті навчання шляхом їх представлення на вхід нейроної мережі.

У відповідності з законом Хебба [32], зміна ваги зв'язків $\Delta g'_{ij}$ на кожному кроці ітерації повина бути пропорційна величинам вхідного та вихідного сигналів нейронів:

$$\Delta g'_{ij} = \alpha \cdot y_i \cdot x_k, \quad (2.6)$$

де α – стала, яка визначає величину кроку приросту.

Відомо [33], що передаточна функція нейрона має сигмоїдальну форму з нижнім та верхнім рівняннями насичення (рисунок 2.3).

На ділянці 1–2 активного навчання нейрона (рисунок 2.3) можна апроксимувати лінійною функцією з деяким нахилом k_f . Тоді формулу (2.16) можна переписати у вигляді:

$$\Delta g_{ik} = \alpha \cdot y_i \cdot x_k = \alpha \cdot k_f \cdot \sum_{j=1}^n g'_{ij} \cdot x_j \cdot x_k. \quad (2.7)$$

Ваги зв'язків g'_{ij} аналізуючого нейронного шару після навчання визначають структуру нейронної мережі, яка відповідає запам'ятованому образу. Через це (2.17) являється ознакою для перетворення вихідного вектора параметрів мовного сигналу в мовний образ.

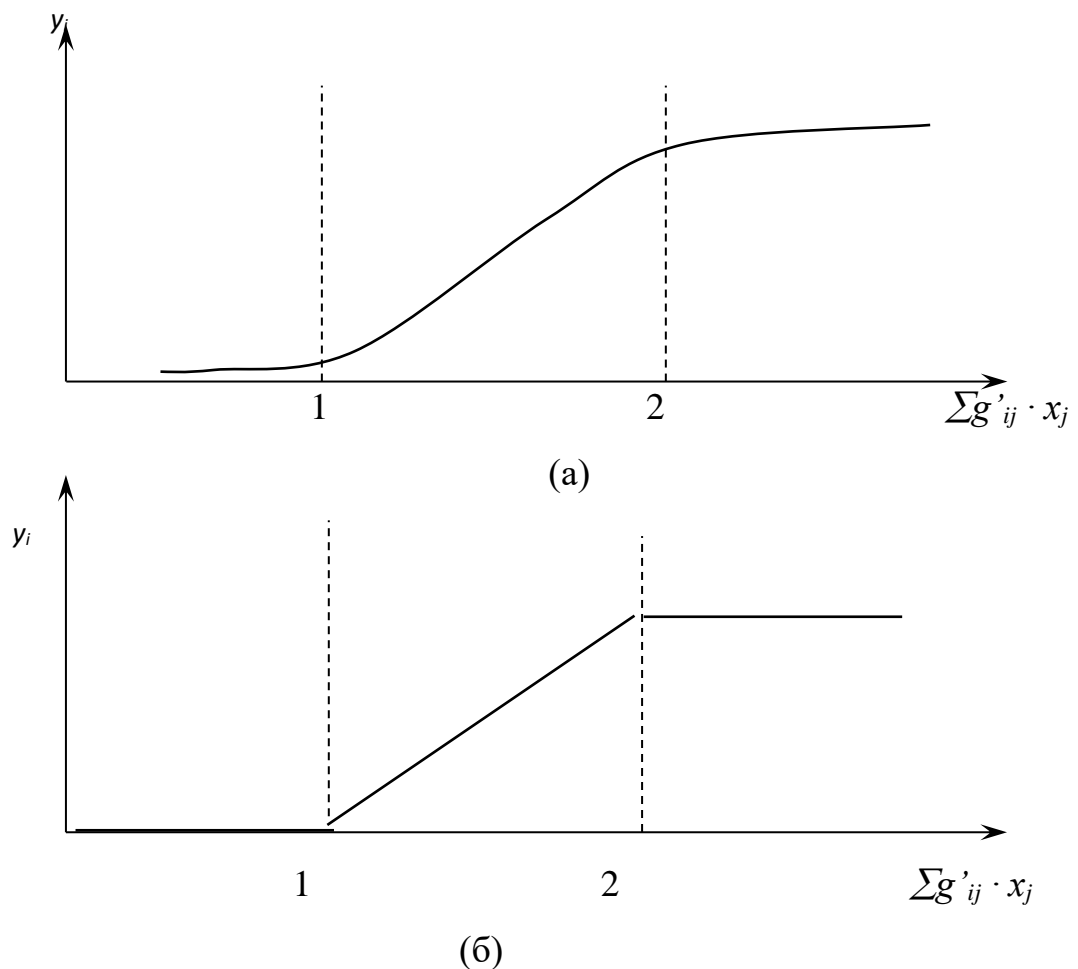


Рисунок 2.3 - Передаточна характеристика нейрона (а) та її апроксимації (б)

Аналізуючи (2.17) для всіх ваг зв'язків, можна зазначити, що ознаки мовного сигналу слід розглядати серед складових автокореляційної матриці спектральних параметрів мовного сигналу [34]:

$$\|y_x\| = x \cdot x^T = \begin{vmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & \cdots & x_1 \cdot x_n \\ x_2 \cdot x_1 & x_2 \cdot x_2 & \cdots & x_2 \cdot x_n \\ \cdots & \cdots & x_i \cdot x_j & \cdots \\ x_n \cdot x_1 & x_n \cdot x_2 & \cdots & x_n \cdot x_n \end{vmatrix}. \quad (2.8)$$

В [23] показано, що для стабілізації роботи нейронної мережі необхідна система гальмування (забування) збудження нейронів, яка регулює сумарний рівень активності нейронів. У відповідності з цим на рисунку 2.2 показані гальмівні зв'язки g'_{ij} , які ідуть від реагуючих нейронів до аналізуючих. Вони попереджують неконтрольоване збудження нейронів мережі, які не входять у збуджене вхідним образом ядро. При зміні диктора, який вимовляє слово, однакове з представленим на вхід слухової системи попереднім диктором, гальмівні зв'язки будуть зашкоджувати зміні вставлених ваг зв'язку через дикторську розбіжність в спектрі.

$$\frac{d g'_{ij}}{dt} = \alpha \cdot y_i \cdot x_i - \beta(y_i) \cdot g_{ij}. \quad (2.9)$$

Наявність границь обмеження в передаточній функції f_G нейрона (рисунок 2.2) стабілізує активність y_i нейрона на одному з рівнів – низькому або високому. При цьому функція гальмування (забування) $\beta(y_i)$ також може приймати одне з двох значень. Для наочності виведень і без втрати загальності розмірковувань перемасштабуємо змінні x_j та g'_{ij} таким чином, щоб $y_i \in \{0, 1\}$, а $\beta(y_i) \in \{0, \alpha\}$.

Тоді рівняння (2.9) можна представити у вигляді системи двох рівнянь:

$$\frac{dg'_{ij}}{dt} = \alpha \cdot (x_j \cdot g_{ij}), \text{ якщо } y_i = 1 \text{ то } \beta = \alpha, \quad (2.10a)$$

$$\frac{dg'_{ij}}{dt} = 0, \text{ якщо } y_i = \beta = 0, \quad (2.10б)$$

Рівняня (2.10a) описує зміну ваг зв'язків всередині активного ядра нейронного аналізуючого шару, який запам'ятав представлений на етапі навчання образ. Рівняня (2.10б) описує зміну ваг зв'язків поза цим образом. З рівнянь (2.10) слідує, що при зсуві спектра мовного образу через заміну диктора ваги зв'язку адаптивно відсліджують зміну спектральних складових вхідного сигналу, просуваючи тим самим по аналізуючому шару активне ядро нейронів. Усунення спектральних змін мовного образу, зумовлених дикторськими розбіжностями, можливе при цьому шляхом кореляції структур ядер аналізуючого шару, яка створюється нейронами реагуючого шару.

Оскільки сучасна технологія не дозволяє ефективно реалізувати нейронну мережу безпосередньо, то із розглянутих механізмів обробки мови в уточненій моделі слуху можна використати наступні принципи обробки інформації, реалізація яких в пристроях розпізнавання мови дозволила б збільшити їх дикторонезалежність.

По перше, інваріантні ознаки для опису мовних образів необхідно шукати на матриці зважених кореляцій спектральних складових або матриці зважених сум цих кореляцій (рівняня 2.18). По друге, інваріантним до диктора представленням ознак являється двійкова матриця, стовпці якої відповідають нормованим відрізкам часу, а стрічки – вибраним ознакам для розпізнавання вибраних підмножин образів Ω_i^h та W_j рівняня (2.20) із урахуванням двійкового характеру передаточної функції f_G нейронів.

2.2 Розробка методу і пристрою виділення складових сегментів слова

Використання інформації про складові сегменти в алгоритмах розпізнавання мовних команд дозволяє значно підвищити їхню ефективність, що було показано в розділі 2.1. Тому розробка методів, алгоритмів і пристроїв для надійного виділення ознак складів є важливою задачею при побудові систем розпізнавання мови. Один із ключових параметрів, що використовується для розпізнавання складів у мовному сигналі, - це його енергія [6, 24, 25, 26, 27]. При цьому ядро складу визначається на місці, де знаходяться піки енергії, обмежені значним (на рівні 40 або 50 дБ) зниженням енергії. Однак у деяких з наведених досліджень, наприклад, [28], відзначається часте невірне виділення за цим показником складів, що формуються внаслідок високоенергетичних фрикативних звуків. Це також підтверджується результатами експериментально записаних відомостей. Для виключення цих помилкових складів в [29] пропонується попередньо визначати гучність, що значно ускладнює алгоритм розпізнавання. У дослідженні [30] в якості параметра для виділення ознак складу використовується характеристика "гучності", яка визначається як взважена сума амплітуд сигналів 22 частотних каналів, розташованих у критичних смугах.

З метою виключення недоліків, властивим зазначеним методам, у даній роботі, з огляду на результати моделювання розділу 2.2, як параметри для формування ознаки складових сегментів запропоновано використати [24] огибаючі сигнали частотних діапазонів $\Delta_1 = 490 \dots 2500$ Гц і $\Delta_2 = 250 \dots 540$ Гц. Результируючий параметр отримується кореляційним методом і запишеться у вигляді:

$$U_c(t) = U_{\Delta_1}(t) \cdot U_{\Delta_2}(t), \quad (2.11)$$

де $U_{\Delta_1}(t)$ - огибаюча енергії в смузі частот Δ_1 , а $U_{\Delta_2}(t)$ - огибаюча енергії в смузі Δ_2 .

На рисунку 2.4 зображена структурна схема запропонованого пристрою для виділення складових сегментів в мовному сигналі, що реалізує запропонований метод.

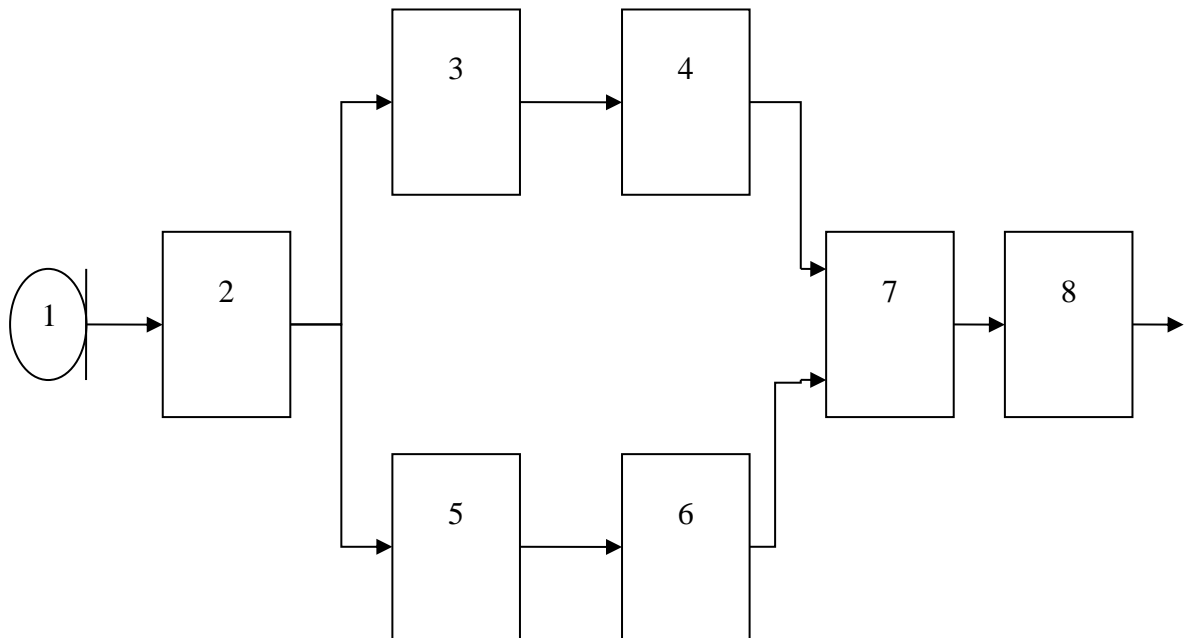


Рисунок 2.4 – Структурна схема пристрою виділення сегментів

На рисунку 2.5 зображені часові діаграми напруги в пристрої для виділення складових сегментів.

Пристрій для визначення елементів сегментів у мовному сигналі включає акустичний датчик 1, який з'єднаний з входом підсилювача 2. Вихід підсилювача 2 підключений до входу першого смугового фільтру 3, при цьому вхід першого детектора 4 пов'язаний з виходом першого смугового фільтру. Вхід другого смугового фільтру 5 з'єднаний з виходом підсилювача. Вхід другого детектора 6 пов'язаний з виходом другого смугового фільтру. Виходи першого і другого детекторів з'єднані з першим і другим входами перемножувача сигналу 7. Вихід перемножувача сигналу 7 підключений до входу генератора границь сегментів складів 8.

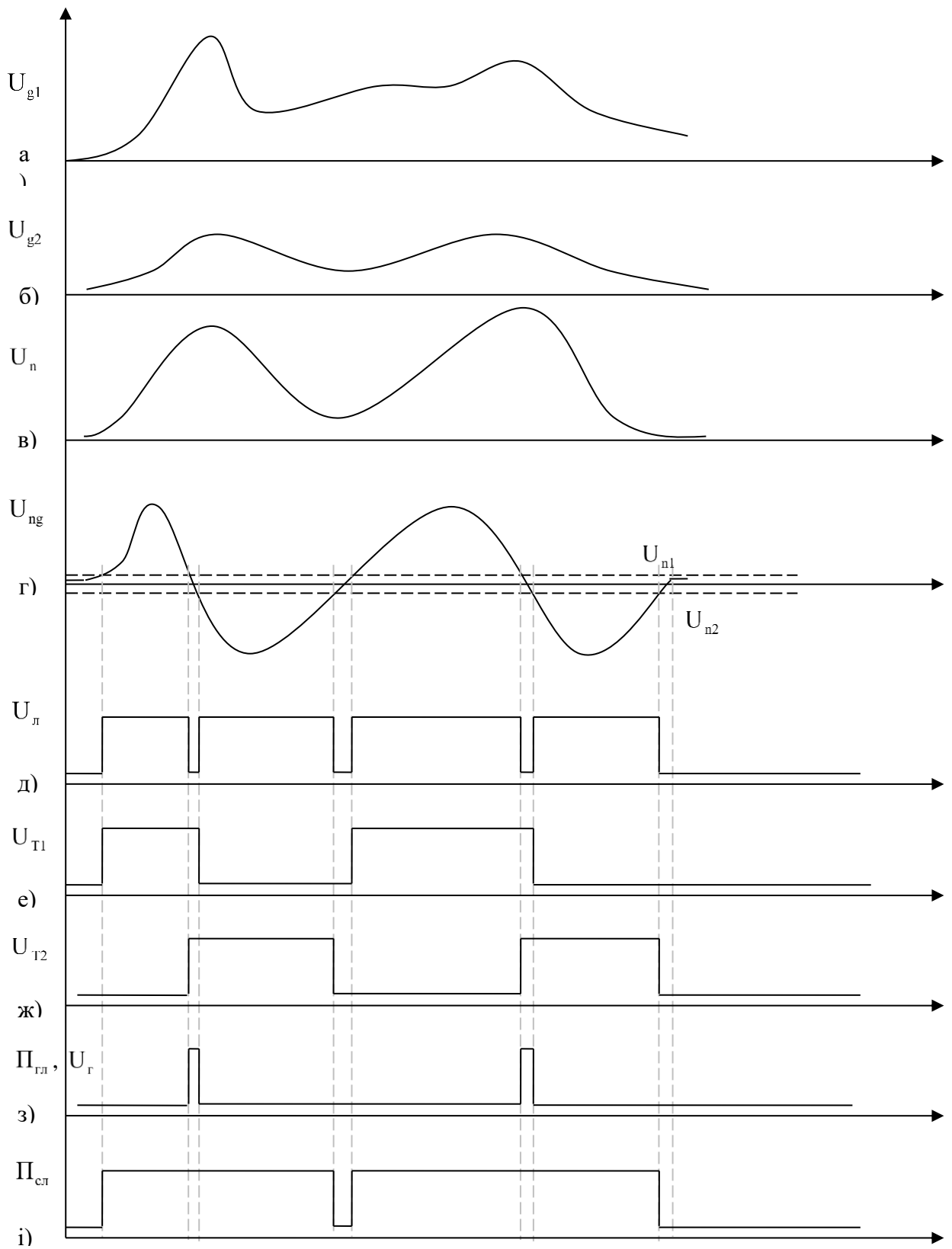


Рисунок 2.5 – Діаграми роботи пристрою виділення складів в часі

Рисунок 2.6 зображає функціональну схему формувача сегментних границь складів 8.

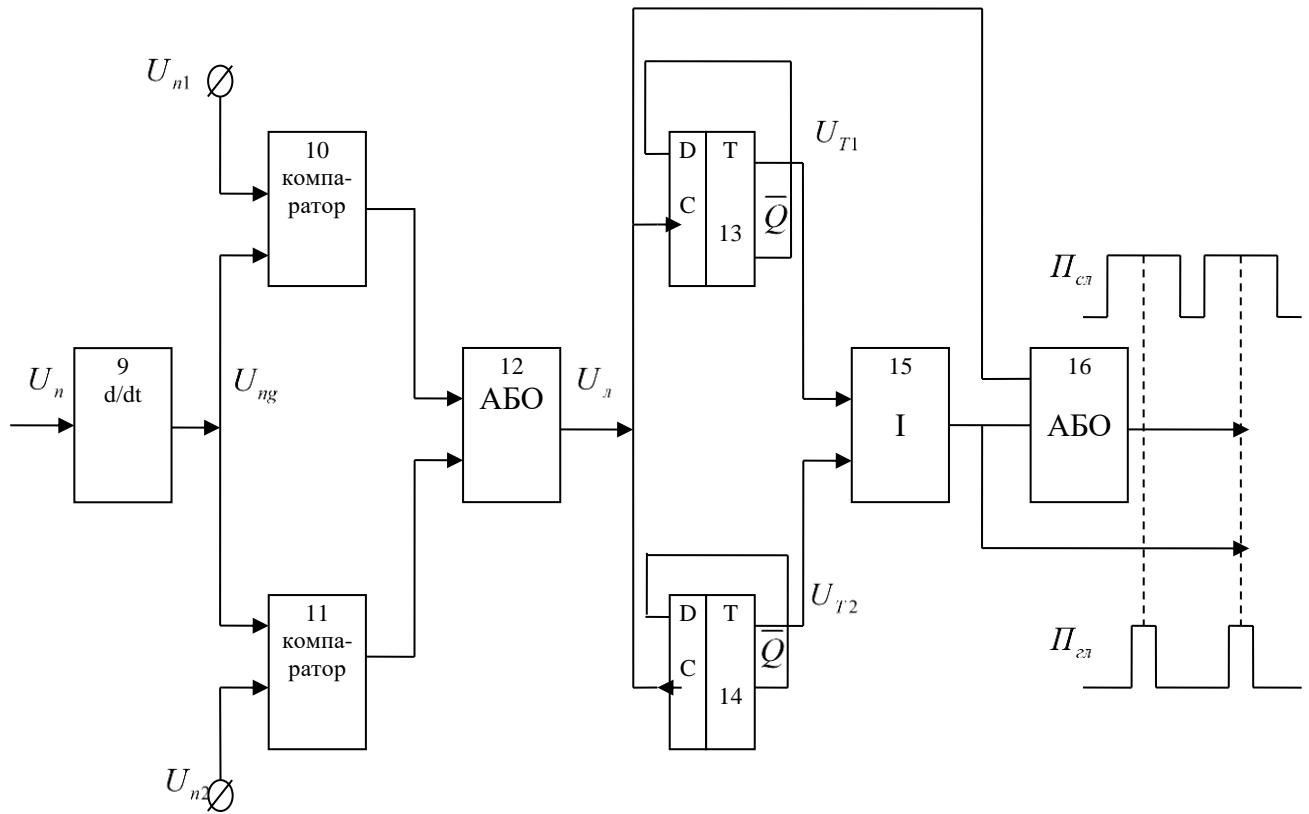


Рисунок 2.6 - Схема функціональна формувача сегментних границь слів

На рисунку 2.6 позначено: 9 – диференціатор; 10, 11 – перша і друга порогові схеми відповідно; 12 – логічна схема АБО на два входи; 13 – лічильний тригер з прямим динамічним управлінням; 14 - лічильний тригер із механізмом зворотного динамічного управління; 15 – логічна схема І на два входи; 16 - логічна схема АБО на два входи.

Диференціатор 9 є практичною схемою диференціювання на операційному підсилювачі [33]. Перша і друга порогові схеми 10 і 11 є компараторами [34]. Логічна схема АБО на два входи 12 є інтегральною схемою диз'юнктора [35]. Лічильний тригер 13 з прямим динамічним управлінням є інтегральною схемою D-тригера в лічильному включенні [35].

Лічильний тригер 14 із зворотним динамічним управлінням є інтегральною схемою JK-тригера в лічильному включенні [33]. Логічна схема І на два входи 15 є інтегральною схемою кон'юнктора [35]. Логічна схема АБО на два входи 16 аналогічна описаній вище схемі 12.

На рисунку 2.5 наведені часові діаграми напруги в пристрої для виділення складових сегментів на прикладі слова «Миша». Рисунок 2.5 а) зображає огинаючу мовного сигналу U_{g2} в діапазоні 490...2500 Гц, одержану за допомогою амплітудного детектора 6 на виході смугового фільтру 4, а рисунок 2.5 б) – огинаючу мовного сигналу U_{g1} в діапазоні 250...540 Гц на виході смугового фільтру 3, одержану за допомогою амплітудного детектора 5. Рисунок 2.5 в) зображає сигнал U_n на виході помножувача сигналу 7. На рисунку 2.5 г) зображені сигнал U_{ng} на виході диференціатора 9 (суцільна лінія) і порогові напруги U_{n1} і U_{n2} , що подаються на другі входи порогових схем 10 і 11 відповідно (штрихова лінія). На рисунку 2.5 д) зображена напруга U_n , що одержується на виході схеми АБО 12. Рисунок 2.5 е) і рисунок 2.5 ж) зображають напруги логічних сигналів U_{T1} і U_{T2} на прямих виходах тригерів 13 і 14 відповідно. На рисунку 2.5 з) зображений логічний сигнал U_z (P_{z1}) на виході схеми збігу 15, що визначає положення складових ядер в мовному сигналі. На рисунку 2.5 і) зображений логічний сигнал P_{c1} , що сегментує мовний сигнал на відрізки, відповідні складам мови.

Пристрій для виділення складових сегментів в мовному сигналі працює таким чином.

Мовний сигнал сприймається акустичним датчиком (мікрофоном) 1, перетворюється в електричний і поступає на вхід підсилювача 2. Електричний сигнал, посилений до величини, достатньої для роботи подальших каскадів, поступає на входи смугових фільтрів 3 і 4. Амплітудний детектор 5, вхід якого підключений до виходу смугового фільтру 3, виділяє огинаючу U_{g2} (рисунок 2.5 б) мовного сигналу в діапазоні 250...540 Гц. Амплітудний детектор 6, вхід

якого підключений до виходу смугового фільтру 4, виділяє огинаючу U_{g1} (рисунок 2.5 а) мовного сигналу в діапазоні 490...2500 Гц. Напруги U_{g1} і U_{g2} , що поступають на перший і другий входи помножувача сигналів 7, перемножаються, внаслідок чого на його виході з'являється напруга U_n (рисунок 2.5 в), рівна:

$$U_n = U_{g1} \cdot U_{g2}. \quad (2.12)$$

Діапазони частот перших і других смугових фільтрів підібрані так, що під час виконання операції множення вони корелюють лише з енергією голосних звуків. Це призводить до усунення максимумів, що виникають на високоенергетичних фрикативних ділянках, які представлені на рисунку 2.5 в. Напруга, що виникає при цьому, подається на вхід диференціатора 9 формувача ознак 8, де утворюється напруга, пропорційна похідній від цієї напруги (рисунок 2.5 г). Наступною кроком є подання цієї напруги на вхід порогових схем 10 і 11. Перші входи порогової схеми 10 отримують додатню порогову напругу, а другі входи порогової схеми 11 - негативну порогову напругу. Параметри вибрані таким чином, щоб уникнути помилкових спрацювань через фоновий шум при переході через нульові значення. Сигнали з виходів порогових схем 10 і 11 поступають на входи схеми АБО 12, де утворюється цифровий сигнал, представлений на рисунку 2.5 д. Фронті вхідного сигналу лічильного тригера з прямим динамічним управлінням 13 перемикають його в протилежний стан, формуючи вихідний цифровий сигнал (рисунок 2.5 е). Лічильний тригер із зворотнім динамічним управлінням 14 видає сигнал, представлений на рисунку 2.5 ж, і перемикається спадами вхідного сигналу. Сигнали і подаються на входи схеми збігу 15, де вони генерують короткі імпульси (рисунок 2.5 з), локалізуючи центри складових ядер у слові. Ці імпульси подаються на перший вхід схеми АБО 16 та на вихід пристрою для виділення складових сегментів у слові, що є ознакою положення

центрів голосних звуків у складі. На другий вхід схеми АБО 16 подається цифровий сигнал, який разом із сигналом об'єднується на виході схеми 16, утворюючи сигнал, який виділяє сегменти у слові, відповідні положенню складів у слові.

Був вибраний діапазон частот від 250 до 540 Гц для першого смугового фільтру 3 з урахуванням відсутності енергії високоенергетичних фрикативних звуків, таких як /ш/ і /ч/. Це рішення прийнято для уникнення утворення помилкових складових ядра. Також врахована значна концентрація енергії всіх дзвінких звуків, включаючи голосні. Однак в цьому діапазоні енергія сонорних звуків, таких як /л/, /м/, /н/, збігається за значенням з енергією голосних. Тому визначення складових сегментів тільки за огинаючою мовного сигналу в цьому діапазоні може призвести до численних помилок. З цією метою для другого смугового фільтру 4 був обраний діапазон від 490 до 2500 Гц, де енергія голосних звуків є принаймні в два рази вищою, ніж енергія сонорних звуків.

У запропонованому методі для визначення сегментів складів використовується часова функція, яка формується шляхом множення перемножувачем сигналу на огинальних мовних сигналах, отриманих від смугових фільтрів. Під час операції множення відбувається посилення областей кривих, що відповідають голосним звукам, завдяки кореляції їх енергії в обох діапазонах. Помилкові максимуми енергії, обумовлені фрикативними звуками в діапазоні 490-2500 Гц, усуваються множенням на практично нульові амплітуди фрикативних звуків у діапазоні 250-540 Гц. В пристрої-прототипі для виділення складових сегментів використовується функція гучності, яка формується шляхом сумування огинаючих сигналів у 22-х частотних діапазонах. Це підходить для визначення помилкових максимумів на ділянках високоенергетичних фрикативних звуків, що може спричинити появу неточностей у визначенні складових сегментів на виході пристрою-прототипу. Додатково, застосування центрального кліпування сигналу за допомогою змін у схемі формувача ознак покращило надійність

визначення складових сегментів, особливо в мінімумах огинаючих, шляхом зменшення впливу фонових шумів. Таким чином, введення першого і другого смугових фільтрів, налаштованих на конкретні діапазони частот, у пристрій, що розглядається, сприяло покращенню його продуктивності 250...540 Гц. У пристрої-прототипі для виділення складових сегментів використовується функція гучності, яка отримана підсумовуванням оцифрованих сигналів у 22 частотних діапазонах. Такий підхід не дозволяє усунути помилкові максимуми на ділянках високоенергетичних фрикативних звуків, що може призводити до виникнення невірних складових сегментів на виході пристрою-прототипу. Додатково, використання центрального кліпування сигналу через зміни в схемі формувача ознак допомогло підвищити надійність визначення складових сегментів, особливо в мінімумах оцифрованих сигналів, за рахунок виключення впливу фонових шумів. Таким чином, внесення змін у пристрій, що пропонується, першого і другого смугових фільтрів, налаштованих на діапазони частот 250...540 Гц і 490...2500 Гц, амплітудних детекторів, перемножувача сигналів, диференціатора, двох порогових схем, використання першої або другої схеми "АБО", лічильного тригера з прямим динамічним управлінням, лічильного тригера зі зворотнім динамічним управлінням та схеми збігу "Г" дозволило підвищити надійність визначення складових сегментів у порівнянні із відомим пристроєм. Також спостерігається зменшення складності пристрою, оскільки в блоку спектрального аналізу використовується приблизно в 10 разів менше смугових фільтрів і амплітудних детекторів.

2.3 Розробка алгоритму виділення складових сегментів

Після аналізу результатів моделювання розроблено алгоритм виділення ознак складових сегментів, де в якості первинних параметрів використовуються оригінальні сигнали з частотними діапазонами $\Delta 1 = 800 - 2500$ Гц і $\Delta 2 = 250 - 540$ Гц. Застосовуючи кореляційний метод, отримується результуючий параметр, який в подальшому використовується для виділення

ознак складів і подається у вигляді:

$$U_c(t) = U_{\Delta_1}(t) \cdot U_{\Delta_2}(t), \quad (2.13)$$

де $U_{\Delta_1}(t)$ – огибаюча енергії в смузі частот Δ_1 , а $U_{\Delta_2}(t)$ – огибаюча енергії в смузі Δ_2 .

Діапазон частот першого смугового фільтру, який варіюється від 250 до 540 Гц, був обраний через відсутність енергії високоенергетичних фрикативних звуків, таких як /ш/ і /ч/, які можуть призводити до утворення неточних складових ядра. Крім того, в цьому діапазоні значна частина енергії всіх дзвінких і голосних звуків. Оскільки енергія сонорних звуків, таких як /л/, /м/, /н/, в цьому діапазоні корелює з енергією голосних, визначення складових сегментів лише за огибаючою сигналу мовлення в цьому діапазоні призведе до численних помилок. З цією метою другий смуговий фільтр вибраний в діапазоні від 800 до 2500 Гц, де енергія голосних звуків, як мінімум, вдвічі перевищує енергію сонорних звуків.

При виконанні операції множення огиначальних сигналів в часовій функції спостерігається значне посилення областей кривих, що відповідають голосним звукам, через кореляцію їх енергій в обох діапазонах. Поява помилкових максимумів енергії пояснюється наявністю значної частини енергії фрикативних звуків у діапазоні від 800 до 2500 Гц, які усуваються множенням на практично нульове значення амплітуди фрикативних звуків у діапазоні від 250 до 540 Гц.

Згідно з описом процедури виділення складових сегментів, алгоритм роботи включає наступні етапи:

- а) Зафіксування мовного сигналу за допомогою мікрофона.
- б) Проведення фільтрації сигналу двома смуговими фільтрами Баттерворта четвертого порядку в діапазонах 250 Гц – 540 Гц і 800 Гц – 2500 Гц.
- в) Детектування вихідних сигналів фільтрів для отримання обвідних сигналів.

- г) Проведення множення огинаючих вихідних сигналів фільтрів.
- д) Диференціювання отриманого сигналу.
- е) Порівняння отриманого сигналу з додатньою та від'ємною пороговими напругами, а також виділення логічного сигналу.
- ж) Створення логічних сигналів для додатніх і від'ємних півперіодів диференційованого сигналу на основі отриманих результатів порівняння.
- з) Визначення сегментів складів і центрів складів за допомогою логічного додавання і множення відповідних отриманих логічних сигналів.

Схема алгоритму виділення складових сегментів зображена на рис.2.7

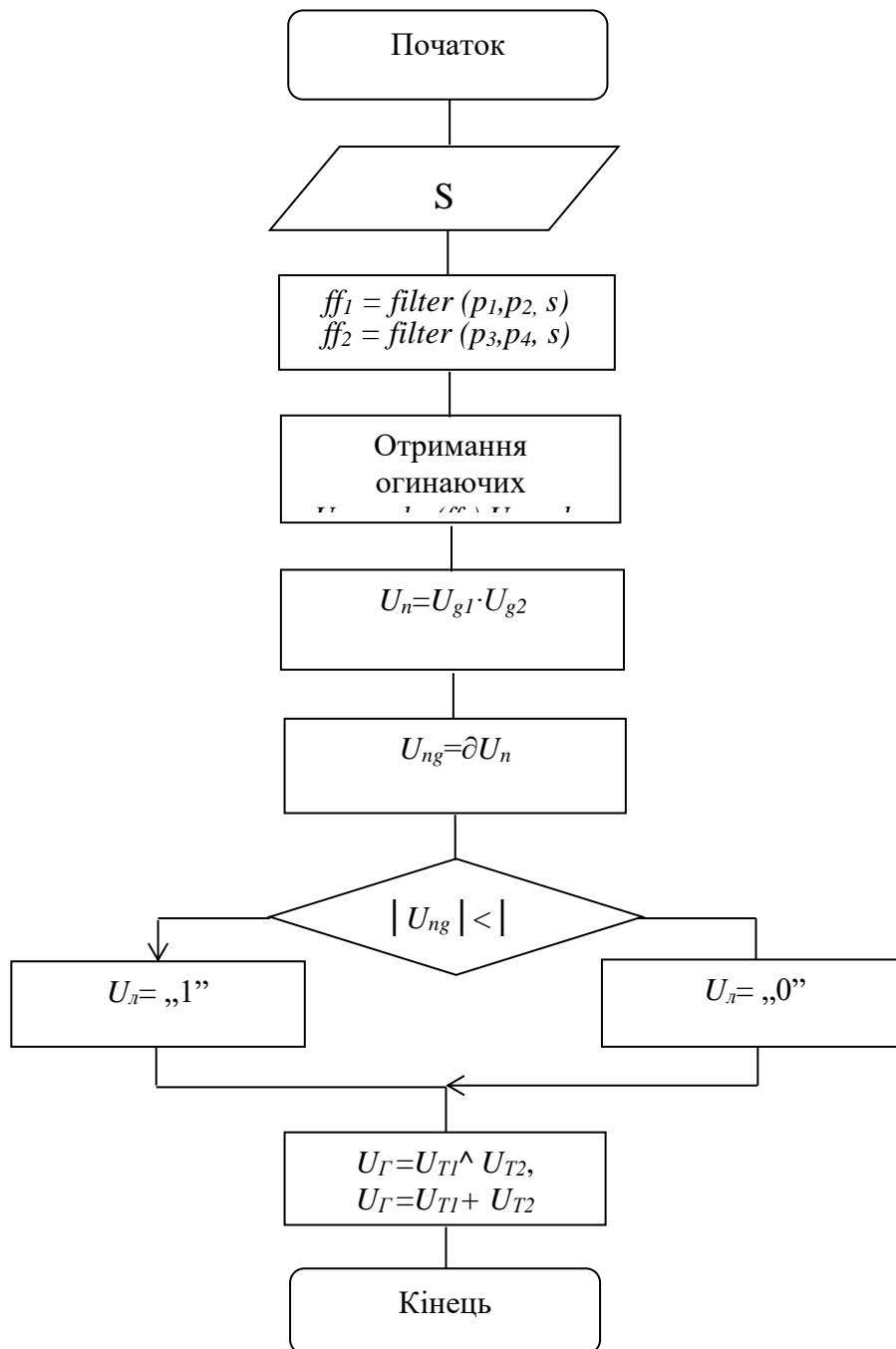


Рисунок 2.7 – Схема алгоритму виділення складів

2.4 Висновки до розділу

В даному розділі розроблена математична модель слухової системи, яка показала, що надійні ознаки для виділення складових ознак потрібно шукати серед кореляцій енергій сигналу мовлення в вибраних частотних діапазонах.

На основі моделі розроблено пристрій і алгоритм виділення складових сегментів в мовленнєвому потоці. Пристрій дозволяє реалізувати апаратно автоматичну систему контролю темпу мови в контакт-центрах, а алгоритм – програмним шляхом.

3 РОЗРОБКА АЛГОРИТМІВ І ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

3.1 Вибір програмної платформи

Розроблений в магістерській роботі програмний продукт реалізований в середовищі Matlab 6.1 [36]. Цей вибір є доцільним, оскільки в Matlab 6.1 міститься спеціальний модуль роботи із звуковими файлами. В цьому пакеті можна розглядати спектральну характеристику мовного сигналу. Середовище Matlab 6.1 призначене для інженерних і наукових розрахунків і важливу роль в усіх версіях виконує підсистема технічних розрахунків. Слід відмітити ту унікальну особливість системи Matlab, що всі розрахункові операції в ній виконуються над векторами чи наборами векторів. Також середовище містить великий набір графічних інтерфейсів, при використанні яких робота з розробленим програмним продуктом буде зрозумілою і корисною навіть для роботи з ним не досить обізнаним в цьому напрямку користувачем. Оскільки в роботі ми працюємо із звуковим файлом і операції які слід виконувати пов'язані із операціями над векторними величинами, тому використання математичного пакету Matlab 6.1 є доцільним як з точки зору легкості написання, так і зручності представлення результатів програми.

Розроблене програмне забезпечення представлено у вигляді однієї основної програми Main, яка містить шість зовнішніх підпрограм: Granusa, Obrs, Flt_ogib, Koreliatsia, Difer, Logic і функцію $fn_z(x,n)$;

3.2 Розробка алгоритмів і програм оброблення сигналу мовлення

3.2.1 Визначення границь слова в мовному сигналі

Для вирішення поставленої задачі запропоновано використовувати енергію сигналу (ознака E) та кількість нульових перетинів (ознака ZC) для кожного інтервалу фонового шуму [8]. Для впровадження цього алгоритму

необхідно мати статистику шумів, яка формується шляхом вирізання інтервалів шуму з різних звукових сигналів

- знаходимо максимальну енергію шуму:

$$E_{шуму\max} = \max(E_j), j = 1..n; \quad (3.1)$$

- знаходимо математичне очікування енергії шуму:

$$M_E = \frac{\sum_{j=1}^n E_j}{n}; \quad (3.2)$$

- знаходимо дисперсію енергії шуму:

$$D_E = \frac{\left(\sum_{j=1}^n E_j\right)^2}{n} - \frac{\sum_{j=1}^n E_j^2}{n}; \quad (3.3)$$

- знаходимо СКВ енергії шуму за формулою:

$$\sigma_E = \sqrt{D_E}; \quad (3.4)$$

- 3) визначаємо границі детекції:

$$E_{\max} = E_{шуму\max} + 3\sigma_E; \quad (3.5)$$

$$E_{\min} = M_E + 3\sigma_E; \quad (3.6)$$

$$ZC_{noise} = \left(\frac{\sum_{j=1}^n ZC_j}{n} \right) \cdot \frac{t_{int_мови}}{t_{int_шуму}}, \quad (3.7)$$

де $t_{int_мови}$ - тривалість інтервалів, на які розбивається мова (10 мс);

$t_{int_шуму}$ - тривалість інтервалів, на які розбивається шум, становить 100

мс.

Після отримання інтервалів аудіоданих довжиною в 10 мсек визначається енергетична характеристика для кожного проміжку та кількість переходів через нуль.

$$M_k = \frac{1}{m} \sum_{i=1}^m a_i, \quad (3.8)$$

$$E_k = \frac{1}{m} \sum_{i=1}^m a_i^2, \quad (3.9)$$

$$E_k' = E_k - M_k^2, \quad (3.10)$$

де m – кількість амплітуд на проміжку (для 11 кГц - 128).

Розглянутий нижче алгоритм заснований на вимірі двох простих характеристик – енергій і числа переходів через нуль.

Пояснимо алгоритм за допомогою рисунку 3.1.

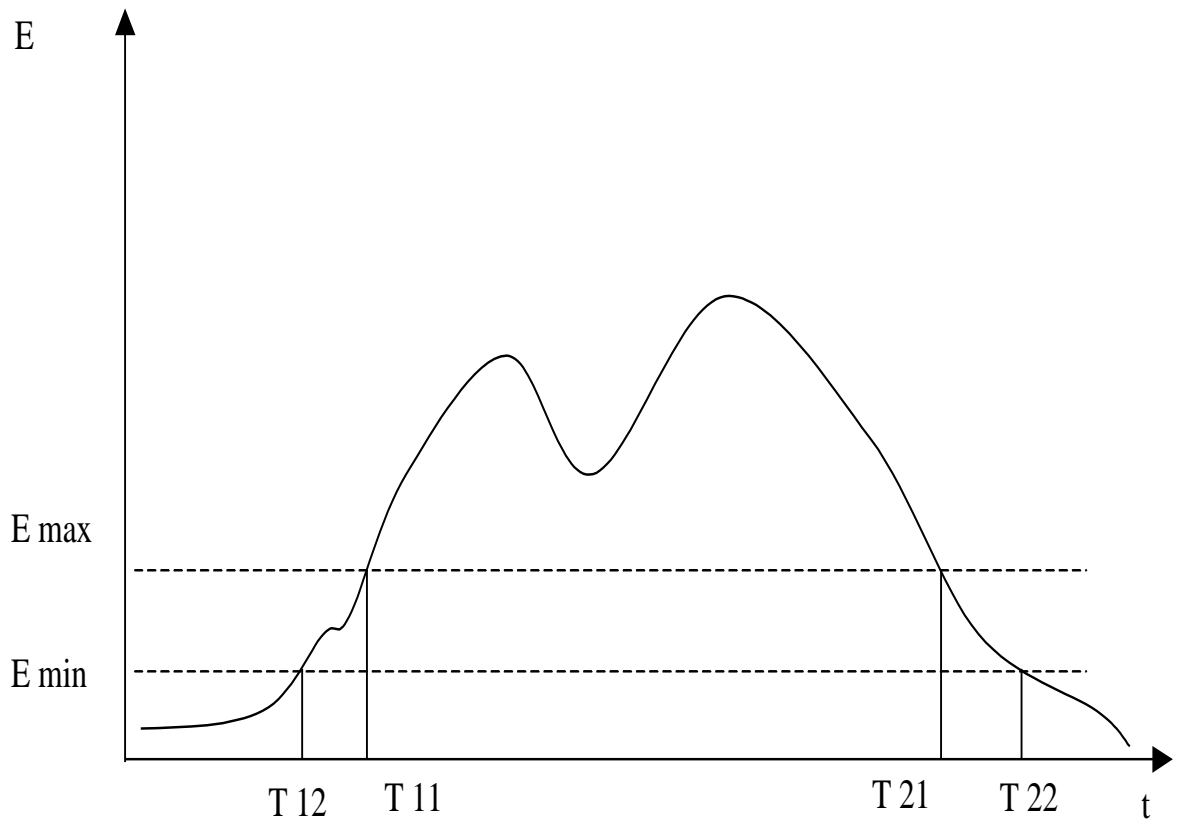


Рисунок 3.1 – Виділення мови з звукового сигналу

Рухаючись від початку звукового сигналу вправо (рисунок 3.1), чекаємо на появу інтервалу, на якому $E_k > E_{\max}$, де k – номер часового сегмента тривалістю 10 мсек. Попередньо встановлюємо початок сигналу мови (позначено як T_{11} на рисунку 3.1).

Послідовно розглядаючи відрізки від T_{11} вправо по вісі часу, чекаємо на той відрізок, на якому $E_k \leq E_{\max}$. Це точка T_{21} на рисунку 3.1. Якщо протягом деякого часу T_{sp} від точки T_{21} ($T_{sp} = 400$ мксек) не було проміжку, на якому $E_k > E_{\max}$, послідовно розглядаючи відрізки від T_{11} вправо по вісі часу, чекаємо на той відрізок, на якому встановлюємо попереднє закінчення мови в точці T_{21} . У протилежному випадку продовжуємо пошук точки T_{21} . Таким чином, ми визначили попередньо часові моменти початку і закінчення сигналу мови. При цьому необхідно враховувати те, що при умові $E_k \leq E_{\max}$ - мінімальна тривалість мовного сигналу, необхідно вважати знайдений відрізок не мовою і необхідно продовжувати пошук точки T_{11} . Рухаючись вліво по вісі часу від точки T_{11} , шукаємо перший відрізок, на якому виконується умова $E_k \leq E_{\max}$. Це точка T_{12} (рисунк 3.1). Після цього вліво по вісі часу від точки T_{12} шукаємо перший відрізок, на якому виконується умова $E_k \leq E_{\max}$. Пошук такого відрізка не повинен віддалятися від точки T_{12} на час, більший (200 мсек). Якщо ми змістилися до $E_k \leq E_{\max}$, то далі ми не рухаємося і вважаємо цю точку початком мови. Рухаючись вправо по вісі часу від точки T_{21} , шукаємо перший відрізок, на якому виконується умова $E_k \leq E_{\max}$. Це точка T_{22} . Після цього вправо по вісі часу від точки T_{22} шукаємо перший відрізок, на якому виконується умова $E_k \leq E_{\max}$. Пошук такого відрізка не повинен віддалятися від точки T_{22} на час, більший (200 мсек). Якщо ми змістилися на $E_k \leq E_{\max}$, то далі ми не рухаємося і вважаємо цю точку закінченням мови.

Для створення моделі виділення сегмента мови з аудіосигналу була зібрана статистика різних фонових шумів (з різних дикторів і різною гучністю вимовлення). Це було здійснено шляхом вирізання відповідних фрагментів з аудіофайлів. Загальний обсяг статистичної вибірки складав 37 значень.

В результаті проведення описаних вище розрахунків були отримані наступні значення основних параметрів:

$$E_{\text{шумуmax}} = 1.695 E - 05;$$

$$M_E = 4.469E - 0.6;$$

$$D_E = 6.992E - 10;$$

$$\sigma_E = 2.644E - 05;$$

$$E_{\max} = 9.628E - 05;$$

$$E_{\min} = 8.379E - 05;$$

$$ZC_{\text{noise}} = 5.991.$$

Для ілюстрації роботи алгоритму представимо результат роботи моделі для слова “слава”. На рисунку 3.2 представлено вигляд звукового сигналу для даного слова:

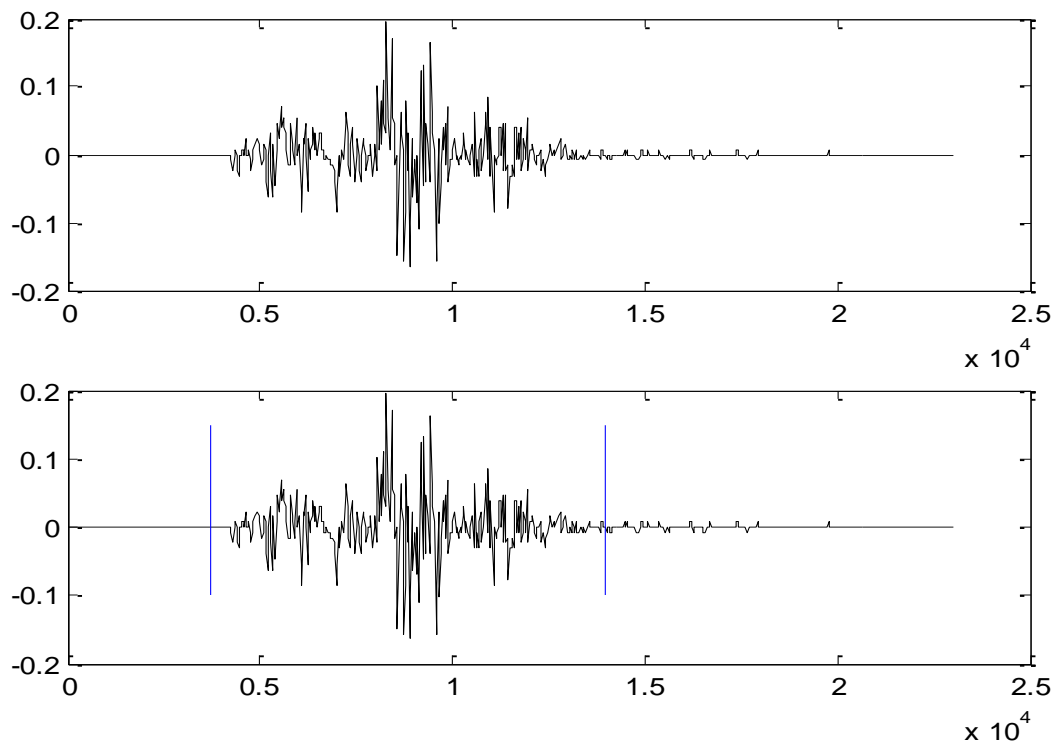


Рисунок 3.2 – Запис сигналу слова “слава”

В результаті звуковий сигнал був зменшений у 2 рази.

За результатами дослідження рівнів шумів, яке проведено в цьому розділі, і спектральних характеристик звукотипів української мови, в [37, 38] зроблено висновок про те, що не інваріантність ознак, що виділяються до змін у рівні зовнішніх акустичних шумів, відстані між акустичним приймачем та джерелом мовного сигналу, та до зміни дикторів. Це призводить до

підвищення надійності виокремлення ознак фонетичних класів звуків, може досягатись за рахунок проведення запису мовних сигналів за допомогою двох акустичних датчиків, один з яких знаходиться на фіксованій відстані. Програма granusa, яка знаходить інтервали початку і кінця мови у звуковому файлі використовує функцію Gran_word(og), і Ogib(x,n), схеми алгоритму яких представлені в Додатку Б, а текст в Додатку В. Схему алгоритму виділення границь мови в звуковому сигналі представлено на рисунку 3.3, а її текст – в Додатку В.

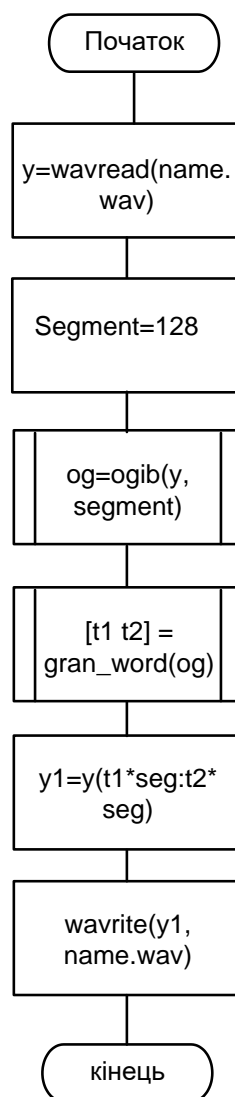


Рисунок 3.3 – Схема алгоритму програми виділення границь слова

3.2.2 Програма фільтрації звукового сигналу

Програма Flt_ogib при передачі їй управління розраховує параметри цифрового смугового фільтра, проводить фільтрування вхідного сигналу, який отримується при читанні файлу вбудованою функцією wavread() і формує вектор енергії звукового сигналу з періодом дискретизації 10 мсек.

Схему алгоритму підпрограми Flt_ogib зображено в Додатку Г. Результатом роботи функції буде рядок матриці, що представляє собою вектор енергії звукового сигналу із періодом дискретизації 12 мс, виокремлений конкретним смуговим фільтром. Номер смугового фільтра відповідає номеру рядка вихідної матриці. Текст функції "Flt_ogib" представлений в Додатку В.

3.2.3 Алгоритм і програма знаходження огинаючої енергії сигналу

Друга зовнішня підпрограма – Obrs. Вона завантажує файл і запускає для кожного з двох фільтрів функцію Flt_ogib. Частотні діапазони 1-го і 2-го фільтрів задані в розділі 2.4: $\Delta F_1 = 250 \dots 540$ Гц, $\Delta F_2 = 490 \dots 2500$ Гц.

Схема алгоритму підпрограми Obrs представлена на рисунку 3.4.

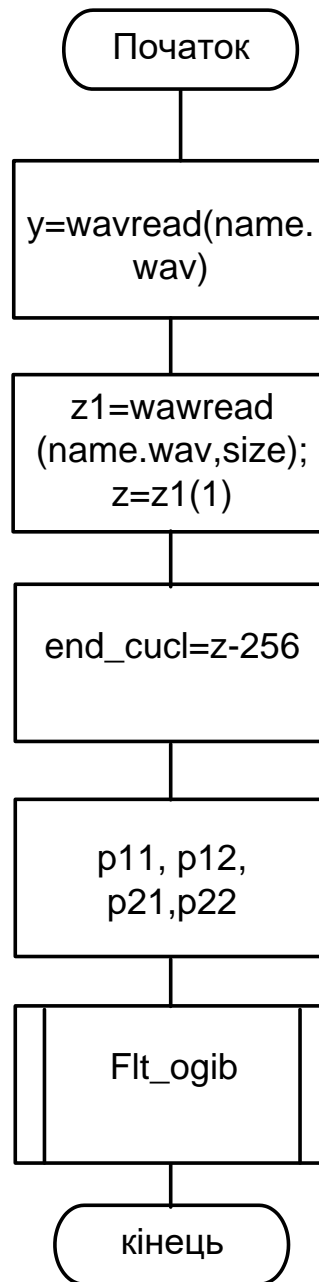


Рисунок 3.4 – Схема алгоритму підпрограми Obrs

Результатом роботи підпрограми буде матриця, яка містить значення енергій звукового сигналу, виділених двома смуговими фільтрами, з періодом дискретизації 12 мс. Текст підпрограми Obrs представлений в Додатку В. Потім функцією $fn_z(x,n)$ відбувається згладжування огинаючих енергій отриманих після двох фільтрів. Опис цієї функції наведено нижче в цьому розділі.

3.2.4 Алгоритм і програма розрахунку кореляції

Програмний модуль Koreliatsia реалізує представлення мовного сигналу в просторі кореляційних ознак. Програма використовує дані, отримані в модулі Obrs, а саме матрицю енергій мовного сигналу, розкладеного смуговими фільтрами і згладжених.

Результатом роботи програми опису мовного сигналу в просторі кореляційних ознак являється кореляційна матриця розміром $[1,N]$, де N залежить від розміру вхідного звукового файлу. Кожен елемент кореляційної матриці характеризує добуток сигналів від двох смугових фільтрів в визначений проміжок часу, який складає 12 мс. Текст підпрограми Koreliatsia представлений в додатку В. Схема алгоритму підпрограми Koreliatsia представлена на рисунку 3.5:

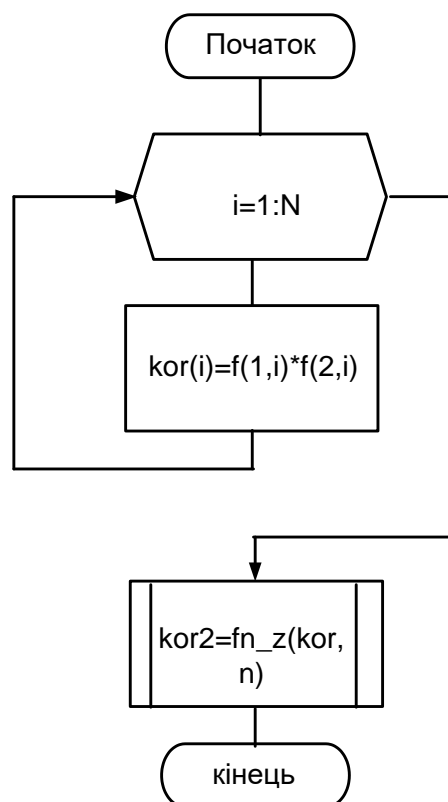


Рисунок 3.5 – Схема алгоритму підпрограми Koreliatsia

Функція $fn_z(x,n)$ повертає вектор, який має довжину рівну вектору x . Елементи цього вектора є згладженими методом середнього значення. функція

має два аргументи: перший з яких – вектор (x), а другий непарне число, яке має бути набагато меншим числа, яке дорівнює кількості елементів вектора x . Схема алгоритму функції $fn_z(x,n)$ наведена на рисунку 3.6, а її текст в Додатку В.

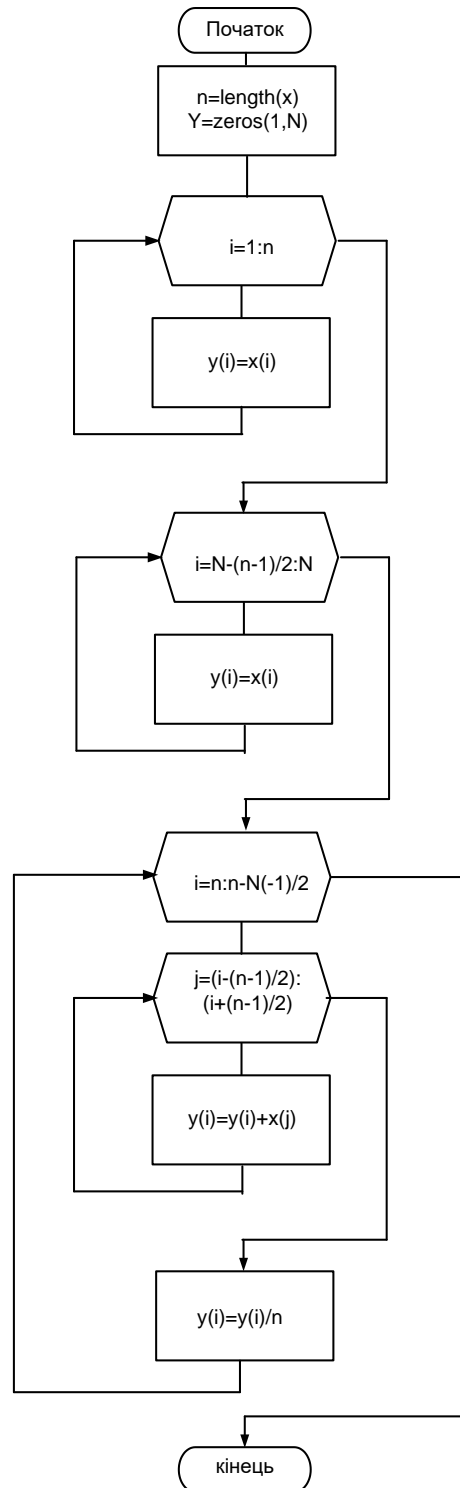


Рисунок 3.6 – Схема алгоритму функції $fn_z(x,n)$

3.2.5 Алгоритм і програма диференціювання обвідних

Диференціювання кореляції здійснюється в підпрограмі Difer. В підпрограмі реалізоване різницеве диференціювання, яке отримується внаслідок різниці наступного значення кореляційної величини до поточного. Також в програмі Difer знаходиться порогове значення величини, яке необхідне для подальших розрахунків. Схема алгоритму підпрограми диференціювання представлена на рисунку 3.7. Текст програми наведено в Додатку В.

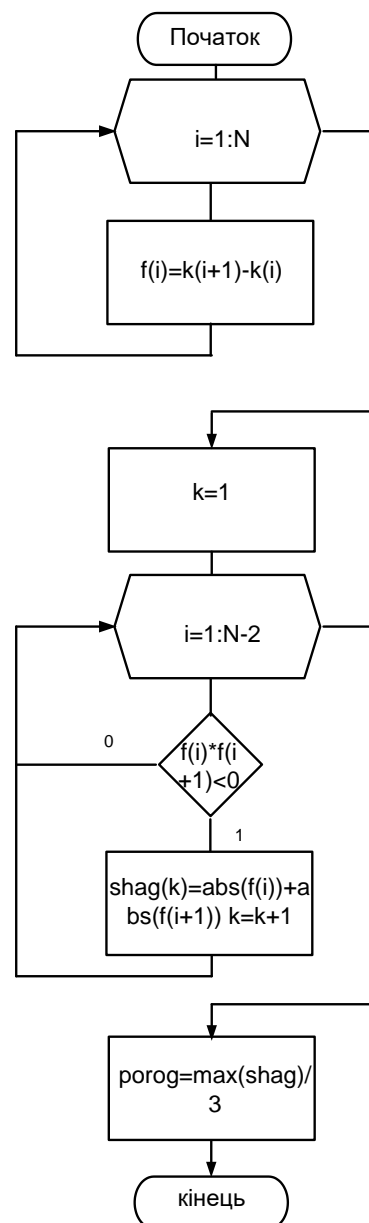


Рисунок 3.7 – Схема алгоритму підпрограми диференціювання

3.2.6 Алгоритм і програма виділення центрів і сегментів складів

Програма Logic працює із вихідними даними програми диференціювання. Принцип її роботи аналогічний роботі пристрою, робота якого розглядається в пункті 2.3. Спочатку формується логічний сигнал $U_{\text{л}}$, значення якого рівне “1”, якщо абсолютне значення диференційованої кореляції в поточному часі буде більшим за порогове, а в протилежному випадку поточне значення сигналу $U_{\text{л}}$ прирівнюється до “0”. Потім розраховуються значення U_{T1} і U_{T2} на основі $U_{\text{л}}$. При непарних значеннях зміни величини $U_{\text{л}}$ із “1” на “0” U_{T1} має значення протилежне значенню $U_{\text{л}}$ в іншому випадку дорівнює. U_{T2} розраховується аналогічно але тут враховується зміна величини $U_{\text{л}}$ із “0” на “1”. Центри складів обчислюються множенням поточних значень U_{T1} і U_{T2} , а границі складів операцією виключення або цих самих величин. Алгоритм її роботи представлено схематично в Додатку Д, а її текст в Додатку В.

3.3 Опис вхідних, вихідних та проміжних даних

Початкові дані –.wav- файли з записами акустичних сигналів слів.

Вихідні дані всієї системи – логічні сигнали, тобто часові інтервали, на яких представлено у відповідних проміжках часу центри та склади слова, яке відображає звуковий файл. Центри та склади відповідають одиничним значенням.

Опишемо структуру всіх даних, які виводять всі функції та програми.

Програма granуса, яка призначена для знаходження границь слова в початковому звуковому файлі, разом із функціями gran_word() ogib(), оперує даними, які описані в таблиці 3.1.

Таблиця 3.1 – Опис даних програми Granusa

Назва	Опис	Коментар
seg	Кількість амплітуд на проміжку 10 мс	Для файла частотою 11 кГц для інтервалів, на які розбивається мова (10 мс) значення амплітуд буде 128

Продовження таблиці 3.1

y	Звуковий файл	Масив, який містить дані файла, який зчитується функцією wavread()
og	Енергія і кількість нульових перетинів	Вихідна матриця функції ogib(), використовується функцією gran_word() для знаходження значень енергії і переходів через нуль, порівнюючи ці значення із критичними
t1, t2	Початок і кінець слова	Вихідні аргументи функції gran_word(), які показують початок і кінець слова в файлі
y1	Звуковий файл	Масив, який зменшився в результаті обрізання його в відповідних інтервалах (t1*seg:t2*seg)

Програма Obrs, і підпрограма Flt_ogib призначені для фільтрації сигналу і знаходження огинаючих енергії. Опис даних цих програм наведено в таблиці 3.2.

Таблиця 3.2 – Опис даних програми Obrs і Flt_ogib

Назва	Опис	Коментар
z	Розмір звукових даних	Використовується для організації цикла розрахунку енергій
f	Значення енергій	Матриця $nf \times len/128$, де len – розмір файла
pr1	Верхня межа частоти пропускання смугового фільтра	Використовується фільтр Баттерворта
pr2	Нижня межа частоти пропускання смугового фільтра	Використовується фільтр Баттерворта
nf	Номер фільтра	Застосовуємо два фільтри
Am	Сума амплітуд	Проміжне значення для знаходження енергії
i	Індексна змінна	Номерується енергія в часі

Програма Koreliatsia перемножує значення енергій отриманих програмою Obrs. При цьому відбувається подавлення низьких значень енергій і виділення високих. Потім її згладжують для подальшої роботи і проводиться

диференціювання програмою Difer. Опис даних програм наведений в таблиці 3.3.

Таблиця 3.3 – Опис даних програми Koreliatsia і Difer

Назва	Опис	Коментар
f	Енергія	Матриця $2 \times N$
kor	Корельована енергія	Матриця $1 \times N$, Отримується шляхом перемноження енергій – $f(1,i) * f(2,i)$
N	Розмірність	Залежить від величини файлу
f _{yd}	Диференційована кореляція	Отримується шляхом різниць попереднього значення корельованої енергії від поточної
porog	Поріг	Порогове значення енергії

Програма Logic формує логічні ознаки складів. Дані, які використовуються в програмі наведені в таблиці 3.4.

Таблиця 3.4 – Опис даних програми Logic

Назва	Опис	Коментар
f _{yd}	Диференційована кореляція	Використовується різницеве диференціювання
U1	Логічний сигнал	Масив $1 \dots N$, отримується шляхом порівняння
p1,p2	Кількість переходів	p1 – кількість переходів з “1” в “0” на часовому проміжку, p2 – з “0” в “1”.
Ut1	Логічний сигнал	Масив $1 \dots N$
Ut2	Логічний сигнал	Масив $1 \dots N$
Pgl	Центри складів	Масив $1 \dots N$
Psl	Виділені склади	Масив $1 \dots N$

3.4 Розробка програмної документації

3.4.1 Опис логічної структури програмного забезпечення

Логічну структуру програми можна представити схемою на рисунку 3.8.

Із рисунку 3.8 можна побачити як між собою взаємодіють програмні модулі. Пояснимо обрану схему. Звуковий файл відкривається за допомогою програми Granusa, і в процесі виконання цієї підпрограми визначаються межі

початку та кінця слова. Після цього формується файл, в якому міститься відокремлене від шуму слово. Отриманий файл завантажується в програмі Obrs.

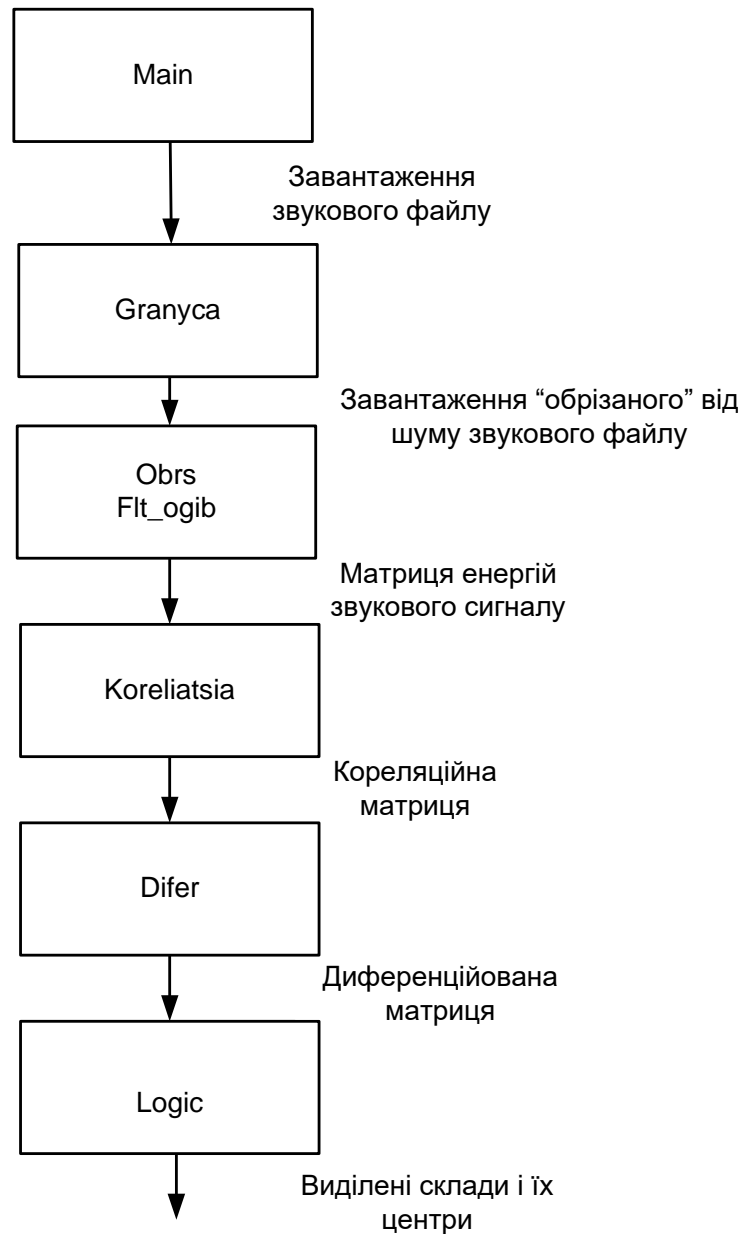


Рисунок 3.8 – Схема взаємодії програм

За допомогою програми Flt_ogib сигнал фільтрується двічі. В результаті ми отримуємо відфільтрований сигнал, який пройшов обидва фільтри. Цей сигнал представляє собою матрицю енергій звукового сигналу з часовими відліками із періодом дискретизації 12 мс. Далі, у програмі Koreliatsia, енергії перемножуються, і ми отримуємо кореляційну матрицю для всіх енергій і

відліків часу. Після цього застосовується метод різницевого диференціювання у програмі Difer, що призводить до обчислення порогу. Після цього в програмі Logic відбуваються логічні перетворення диференційованої матриці кореляції, в результаті чого отримуємо склади і центри складів слова. Результати машинного експерименту наведені в Додатку Д.

3.5 Інструкція користувача

Даний програмний продукт реалізовано на мові високого рівня Matlab

6.1. Необхідні умови для нормальної роботи програмного забезпечення:

- тип процесора – Pentium II – 233 і вище;
- об’єм ОЗУ – 32 МВ і більше;
- монітор – VGA, SVGA;
- відеоадаптер АТІ або аналогічний;
- вільне місце на НЖМД – не менше 12 МВ.

Для зручності і розуміння в користуванні програмою розроблений інтерфейс, що показаний на рисунку 3.9. Він з’явиться після того, як запустити головну програму Main. Для того, щоб отримати очікуваний результат слід покроково виконати такі дії:

- в першому пункті меню “Finding Boundary” вибрати файл формату “wav” для завантаження. Програма визначить границі слова і графічно представить результат;
- вибрати меню “Signal processing”, де відбувається обробка сигналу і отримання його огинаючих енергій;
- натиснути меню “Correlations”. Програма розрахує кореляційну матрицю;
- потім натиснути на вікні меню “Differentiation”;
- після чого слідує остання команда меню – “Logical signal”.

Розрахунки проходять швидко і кожен крок графічно показує результат роботи відповідної підпрограми.

Вихід з програми здійснюється як в більшості Windows вікнах - натисканням курсора миші на значок панелі “Закрити”.

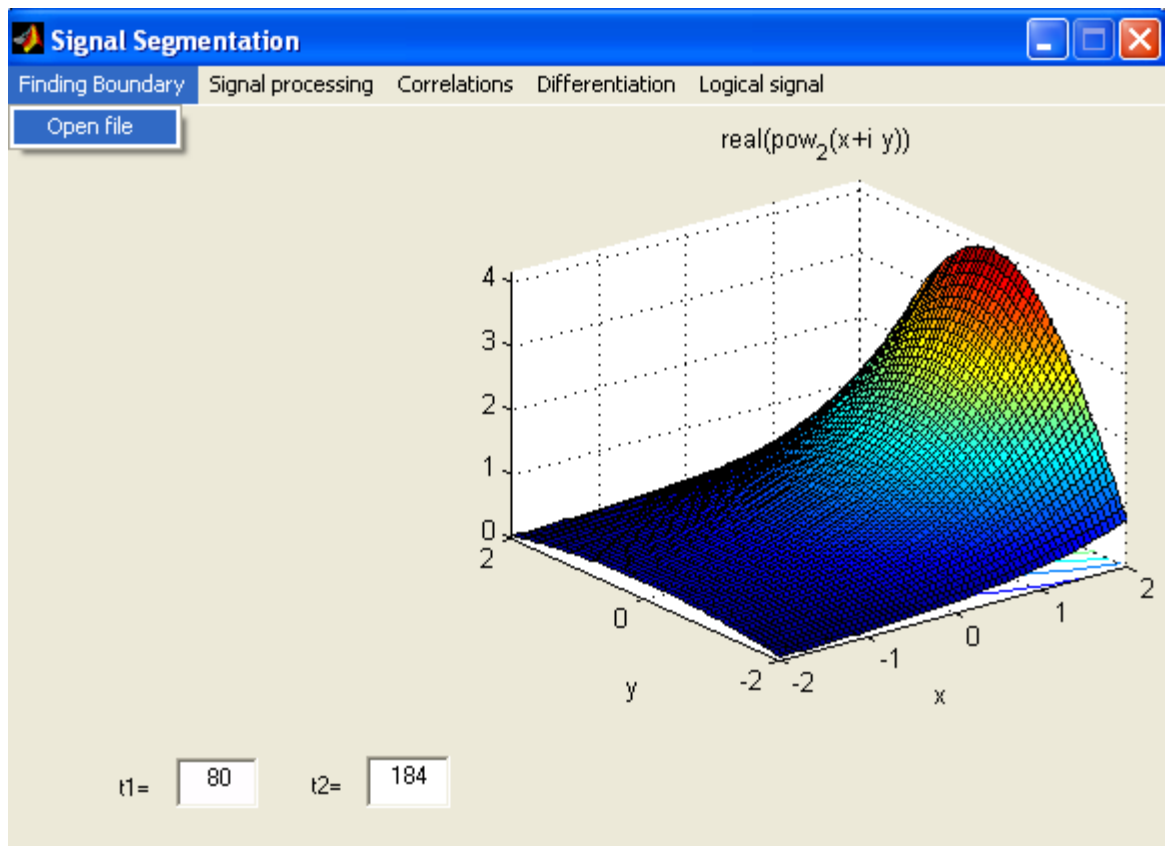


Рисунок 3.9 – Інтерфейс програми

3.6 Висновки до розділу

В даному розділі розроблено алгоритми роботи та програмне забезпечення для системи автоматизації контролю темпу мови в комунікаційних центрах систем критичного використання. В якості платформи обрано програмне середовище Матлаб, оскільки вон має широкий набір бібліотек для обробки сигналів. Тестування розробленого програмного забезпечення показало надійність його роботи для виділення і підрахунку кількості складів в сигналі мовлення, що дає можливість застосувати його для автоматичного контролю темпу мовлення в контакт-центрах різного призначення.

4 ЕКОНОМІЧНА ЧАСТИНА

Науково-технічна розробка має право на існування та впровадження, якщо вона відповідає вимогам часу, як в напрямку науково-технічного прогресу та і в плані економіки. Тому для науково-дослідної роботи необхідно оцінювати економічну ефективність результатів виконаної роботи.

Магістерська кваліфікаційна робота, яка стосується розробки та дослідження "Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання", віднесена до категорії науково-технічних проектів, орієнтованих на введення на ринок. Рішення щодо комерціалізації науково-технічної розробки може бути ухвалено під час проведення роботи. Головний акцент робиться на тому, що результати цієї розробки можуть знайти широке застосування серед інших споживачів, приносячи при цьому економічний вигаш. Проте для реалізації проекту необхідно знайти потенційного інвестора і переконати його в економічній доцільності такого підходу.

Для даного випадку необхідно провести наступні етапи робіт:

- 1) здійснено аналіз комерційний оцінки науково-технічної розробки для визначення її науково-технічного рівня та комерційного потенціалу;
- 2) проведено оцінку витрат на науково-технічну розробку;
- 3) розрахована економічна вигідність впровадження та комерціалізації науково-технічної розробки потенційним інвестором, а також обґрунтовано економічну доцільність комерціалізації інвестором.

4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Метою проведення комерційного і технологічного аудиту дослідження за темою «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання» є оцінювання науково-технічного рівня та

рівня комерційного потенціалу розробки, створеної в результаті науково-технічної діяльності.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями, наведеними в табл. 4.1 [43].

Таблиця 4.1 – Рекомендовані критерії оцінювання науково-технічного рівня і комерційного потенціалу розробки та бальна оцінка

Бали (за 5-ти бальною шкалою)					
	0	1	2	3	4
Технічна здійсненність концепції					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено працездатність продукту в реальних умовах
Ринкові переваги (недоліки)					
2	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому ринку
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів
4	Технічні та споживчі властивості продукту значно гірші, ніж в	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в	Технічні та споживчі властивості продукту значно кращі, ніж в
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
Ринкові перспективи					
6	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкурентів немає
Практична здійсненність					

8	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витратити значні кошти та час на навчання наявних фахівців	Необхідне незначне навчання фахівців та збільшення їх штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї
9	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово-промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуються у виробництві
11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу	Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Результати оцінювання науково-технічного рівня та комерційного потенціалу науково-технічної розробки потрібно звести до таблиці.

Таблиця 4.2 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

Критерії	Експерт (ПІБ, посада)		
	1	2	3
	Бали:		
1. Технічна здійсненність концепції	4	4	4

2. Ринкові переваги (наявність аналогів)	3	3	3
3. Ринкові переваги (ціна продукту)	1	1	1
4. Ринкові переваги (технічні властивості)	3	3	3
5. Ринкові переваги (експлуатаційні витрати)	2	2	2
6. Ринкові перспективи (розмір ринку)	2	2	2
7. Ринкові перспективи (конкуренція)	2	2	2
8. Практична здійсненність (наявність фахівців)	4	4	4
9. Практична здійсненність (наявність фінансів)	2	3	2
10. Практична здійсненність (необхідність нових матеріалів)	2	2	2
11. Практична здійсненність (термін реалізації)	3	4	4
12. Практична здійсненність (розробка документів)	4	4	4
Сума балів	32	34	33
Середньоарифметична сума балів CB_c	33,0		

За результатами розрахунків, наведених в таблиці 4.2, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в табл. 4.3 [43].

Таблиця 4.3 – Науково-технічні рівні та комерційні потенціали розробки

Середньоарифметична сума балів CB_c , розрахована на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
41...48	Високий
31...40	Вище середнього
21...30	Середній
11...20	Нижче середнього
0...10	Низький

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання» становить 33,0 бала, що, відповідно до таблиці 4.3, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки вище середнього).

4.2 Розрахунок узагальненого коефіцієнта якості розробки

Окрім комерційного аудиту розробки доцільно також розглянути технічний рівень якості розробки, розглянувши її основні технічні показники. Ці показники по-різному впливають на загальну якість проектної розробки.

Узагальнений коефіцієнт якості (B_n) для нового технічного рішення розрахуємо за формулою [43]:

$$B_n = \sum_{i=1}^k \alpha_i \cdot \beta_i, \quad (4.1)$$

де k – кількість найбільш важливих технічних показників, які впливають на якість нового технічного рішення;

α_i – коефіцієнт, який враховує питому вагу i -го технічного показника в загальній якості розробки. Коефіцієнт α_i визначається експертним

способом і при цьому має виконуватись умова $\sum_{i=1}^k \alpha_i = 1$;

β_i – відносне значення i -го технічного показника якості нової розробки.

Відносні значення β_i для різних випадків розраховуємо за такими формулами:

- для показників, зростання яких вказує на підвищення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ni}}{I_{ai}}, \quad (4.2)$$

де I_{ni} та I_{ai} – чисельні значення конкретного i -го технічного показника якості відповідно для нової розробки та аналога;

- для показників, зростання яких вказує на погіршення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ai}}{I_{ni}}; \quad (4.3)$$

Використовуючи наведені залежності можемо проаналізувати та порівняти техніко-економічні характеристики аналогу та розробки на основі отриманих наявних та проектних показників, а результати порівняння зведемо до таблиці 4.4.

Таблиця 4.4 – Порівняння основних параметрів розробки та аналога.

Показники (параметри)	Одиниця вимірювання	Аналог	Проектований пристрій	Відношення параметрів нової розробки до аналога	Питома вага показника
Швидкість реакції	мс	50 мс	30 мс	1,67	0,2
Точність розпізнавання	%	0,95	0,98	1,03	0,25
Сумісність інтерфейсів	-	X версія, Y API	Z версія, W API	2	0,2
Витрати енергії	Вт	10 Вт	8 Вт	1,25	0,25
Робоча температура	°C	-10°C - 50°C	0°C - 40°C	0,66	0,1

Узагальнений коефіцієнт якості (B_n) для нового технічного рішення складе:

$$B_n = \sum_{i=1}^k \alpha_i \cdot \beta_i = 1,67 \cdot 0,2 + 1,03 \cdot 0,25 + 2 \cdot 0,2 + 1,25 \cdot 0,25 + 0,66 \cdot 0,1 = 1,37.$$

Отже за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,37 рази.

4.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов'язані з проведенням науково-дослідної роботи на тему «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання», під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

4.3.1 Витрати на оплату праці

До статті «Витрати на оплату праці» належать витрати на виплату основної та додаткової заробітної плати керівникам відділів, лабораторій, секторів і груп, науковим, інженерно-технічним працівникам, конструкторам, технологам, креслярам, копіювальникам, лаборантам, робітникам, студентам, аспірантам та іншим працівникам, безпосередньо зайнятим виконанням конкретної теми, обчисленої за посадовими окладами, відрядними розцінками,

тарифними ставками згідно з чинними в організаціях системами оплати праці.

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників (Z_o) розраховуємо у відповідності до посадових окладів працівників, за формулою [42]:

$$Z_o = \sum_{i=1}^k \frac{M_{ni} \cdot t_i}{T_p}, \quad (4.4)$$

де k – кількість посад дослідників залучених до процесу досліджень;

M_{ni} – місячний посадовий оклад конкретного дослідника, грн;

t_i – число днів роботи конкретного дослідника, дн.;

T_p – середнє число робочих днів в місяці, $T_p=22$ дні.

$$Z_o = 15650,00 \cdot 36 / 22 = 25609,09 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.5 – Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Число днів роботи	Витрати на заробітну плату, грн
Керівник науково-технічної роботи	15650,00	711,36	36	25609,09
Інженер-розробник автоматизованих систем	15000,00	681,82	36	24545,45
Лаборант	6700,00	304,55	10	3045,45
Всього				53200,00

Основна заробітна плата робітників

Витрати на основну заробітну плату робітників (Z_p) за відповідними найменуваннями робіт НДР на тему «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання» розраховуємо за формулою:

$$Z_p = \sum_{i=1}^n C_i \cdot t_i, \quad (4.5)$$

де C_i – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

t_i – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду C_i можна визначити за формулою:

$$C_i = \frac{M_M \cdot K_i \cdot K_c}{T_p \cdot t_{зм}}, \quad (4.6)$$

де M_M – розмір прожиткового мінімуму працездатної особи, або мінімальної місячної заробітної плати (в залежності від діючого законодавства), прийmemo $M_M=6700,00$ грн;

K_i – коефіцієнт міжкваліфікаційного співвідношення для встановлення тарифної ставки робітнику відповідного розряду (табл. Б.2, додаток Б) [42];

K_c – мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати.

T_p – середнє число робочих днів в місяці, приблизно $T_p = 22$ дн;

$t_{зм}$ – тривалість зміни, год.

$$C_1 = 6700,00 \cdot 1,10 \cdot 1,35 / (22 \cdot 8) = 56,53 \text{ грн.}$$

$$З_{р1} = 56,53 \cdot 6,50 = 367,45 \text{ грн.}$$

Таблиця 4.6 – Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Тарифний коефіцієнт	Погодинна тарифна ставка, грн	Величина оплати на робітника грн
Підготовка робочого місця розробника автоматизованій системи	6,50	2	1,10	56,53	367,45
Підготовка бази даних аудіозаписів	5,00	3	1,35	69,38	346,90
Всього					714,35

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 ... 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$Z_{\text{доо}} = (Z_o + Z_p) \cdot \frac{H_{\text{доо}}}{100\%}, \quad (4.7)$$

де $H_{\text{доо}}$ – норма нарахування додаткової заробітної плати. Прийmemo 10%.

$$Z_{\text{доо}} = (53200,00 + 714,35) \cdot 10 / 100\% = 5391,43 \text{ грн.}$$

4.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуемо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$Z_n = (Z_o + Z_p + Z_{\text{доо}}) \cdot \frac{H_{zn}}{100\%} \quad (4.8)$$

де H_{zn} – норма нарахування на заробітну плату. Приймаємо 22%.

$$Z_n = (53200,00 + 714,35 + 5391,43) \cdot 22 / 100\% = 13047,27 \text{ грн.}$$

4.3.3 Сировина та матеріали

До статті «Сировина та матеріали» належать витрати на сировину, основні та допоміжні матеріали, інструменти, пристрої та інші засоби і предмети праці, які придбані у сторонніх підприємств, установ і організацій та витрачені на проведення досліджень за темою «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання».

Витрати на матеріали (M), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$M = \sum_{j=1}^n H_j \cdot C_j \cdot K_j - \sum_{j=1}^n B_j \cdot C_{\text{в}j}, \quad (4.9)$$

де H_j – норма витрат матеріалу j -го найменування, кг;

n – кількість видів матеріалів;

C_j – вартість матеріалу j -го найменування, грн/кг;

K_j – коефіцієнт транспортних витрат, ($K_j = 1,1 \dots 1,15$);

B_j – маса відходів j -го найменування, кг;

$C_{\text{в}j}$ – вартість відходів j -го найменування, грн/кг.

$$M_1 = 1 \cdot 259,00 \cdot 1,05 - 0 \cdot 0 = 271,95 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.7 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
USB флеш накопичувач Transcend 64Gb JetFlash 700 (TS64GJF700)	259,00	1	0	0	271,95
Диск оптичний CD-RW	25,50	5	0	0	133,88
Картридж для принтера	980,00	1	0	0	1029,00
Набір канцелярський офісний FAX	207,00	2	0	0	434,70
Папір А4 500 аркушів клас-С Crystal Print&Copy UPM	187,00	3	0	0	589,05
Папір офісний Офіс Центр А5 80г/м2 500 аркушів клас С	116,00	3	0	0	365,40
Прибор настільний 13 предметів 6300-01 Виготах чорний	235,00	2	0	0	493,50
Всього					3317,48

4.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі (K_6), які використовують при проведенні НДР на тему «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання», розраховуємо, згідно з їхньою номенклатурою, за формулою:

$$K_6 = \sum_{j=1}^n H_j \cdot C_j \cdot K_j \quad (4.10)$$

де H_j – кількість комплектуючих j -го виду, шт.;

C_j – покупна ціна комплектуючих j -го виду, грн;

K_j – коефіцієнт транспортних витрат, ($K_j = 1,1 \dots 1,15$).

$$K_6 = 1 \cdot 12450,00 \cdot 1,05 = 13072,50 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.8 – Витрати на комплектуючі

Найменування комплектуючих	Кількість, шт.	Ціна за штуку, грн	Сума, грн
Відеокарта Asus PCI-Eх GeForce RTX 3060 Dual OC V2 LHR 12GB GDDR6 (192bit) (1837/15000) (1 x HDMI, 3 x DisplayPort) (DUAL-RTX3060-O12G-V2)	1	12450,00	13072,50
Звукова карта Asus Xonar U7 МКП	1	3599,00	3778,95
Всього			16851,45

4.3.5 Спецустаткування для наукових (експериментальних) робіт

До статті «Спецустаткування для наукових (експериментальних) робіт» належать витрати на виготовлення та придбання спецустаткування необхідного для проведення досліджень, також витрати на їх проектування, виготовлення, транспортування, монтаж та встановлення.

Балансову вартість спецустаткування розраховуємо за формулою:

$$B_{\text{спец}} = \sum_{i=1}^k C_i \cdot C_{\text{пр.і}} \cdot K_i, \quad (4.11)$$

де C_i – ціна придбання одиниці спецустаткування даного виду, марки, грн;

$C_{\text{пр.і}}$ – кількість одиниць устаткування відповідного найменування, які придбані для проведення досліджень, шт.;

K_i – коефіцієнт, що враховує доставку, монтаж, налагодження устаткування тощо, ($K_i = 1,10 \dots 1,12$);

k – кількість найменувань устаткування.

$$B_{\text{спец}} = 4199,00 \cdot 1 \cdot 1,04 = 4366,96 \text{ грн.}$$

Отримані результати зведемо до таблиці:

Таблиця 4.9 – Витрати на придбання спецустаткування по кожному виду

Найменування устаткування	Кількість,	Ціна за	Вартість,
---------------------------	------------	---------	-----------

	шт	одиночку, грн	грн
Мікрофони високої якості для запису аудіозаписів	1	4199,00	4366,96
Датчики для вимірювання акустичних параметрів (Мікрофон HyperX Quadcast (HX-MICQC-BK/4P5P6AA) Black)	2	689,00	1433,12
Всього			5800,08

4.3.6 Програмне забезпечення для наукових робіт

До статті «Програмне забезпечення для наукових (експериментальних) робіт» належать витрати на розробку та придбання спеціальних програмних засобів і програмного забезпечення, (програм, алгоритмів, баз даних) необхідних для проведення досліджень, також витрати на їх проектування, формування та встановлення.

Балансову вартість програмного забезпечення розраховуємо за формулою:

$$B_{\text{прог}} = \sum_{i=1}^k C_{\text{инрг}} \cdot C_{\text{прог.і}} \cdot K_i, \quad (4.12)$$

де $C_{\text{инрг}}$ – ціна придбання одиниці програмного засобу даного виду, грн;

$C_{\text{прог.і}}$ – кількість одиниць програмного забезпечення відповідного найменування, які придбані для проведення досліджень, шт.;

K_i – коефіцієнт, що враховує інсталяцію, налагодження програмного засобу тощо, ($K_i = 1, 10 \dots 1, 12$);

k – кількість найменувань програмних засобів.

$$B_{\text{прог}} = 6249,00 \cdot 1 \cdot 1,01 = 6311,49 \text{ грн.}$$

Отримані результати зведемо до таблиці:

Таблиця 4.10 – Витрати на придбання програмних засобів по кожному виду

Найменування програмного засобу	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
---------------------------------	---------------	----------------------	---------------

Інструменти для розробки інтерфейсів інтеграції з комунікаційними центрами	1	6249,00	6311,49
ОС Windows	1	4299,00	4341,99
Прикладний пакет Microsoft Office	1	5240,00	5292,40
Спеціалізовані програми для автоматичного аналізу темпу мови	1	7800,00	7878,00
Всього			23823,88

4.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{обл} = \frac{Ц_б}{T_в} \cdot \frac{t_{вик}}{12}, \quad (4.13)$$

де $Ц_б$ – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$ – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_в$ – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$$A_{обл} = (16510,00 \cdot 2) / (2 \cdot 12) = 1375,83 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.11 – Амортизаційні відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Ноутбук ASUS Vivobook 15 X1500EA-BQ3733 (90NB0TY6-M040W0) Transparent Silver / 15.6" IPS Full HD / Intel Core i3-1115G4 /	16510,00	2	2	1375,83

RAM 12 ГБ / SSD 512 ГБ				
Обчислювальний комплекс та комп'ютеризована система проектування ARTLINE D31 (D31v05) Дисплей 7" / 1280*800 / Intel Core i5-13400F / Водяне охолодження CPU / Asus PRIME B660M-A D4 / 32GB DDR4-3600 / SSD 1TB NVMe / nVidia GeForce RTX 4070 12GB	65899,00	2	2	5491,58
Робоче місце розробника автоматизованої системи	6520,00	5	2	217,33
Пристрій графічного виводу інформації	6750,00	4	2	281,25
Офісна оргтехніка	7300,00	4	2	304,17
Приміщення лабораторії	350000,00	25	2	2333,33
Обладнання відтворення звуку	6899,00	5	2	229,97
Всього				10233,47

4.3.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію (B_e) розраховуємо за формулою:

$$B_e = \sum_{i=1}^n \frac{W_{yi} \cdot t_i \cdot C_e \cdot K_{eni}}{\eta_i}, \quad (4.14)$$

де W_{yi} – встановлена потужність обладнання на визначеному етапі розробки, кВт;

t_i – тривалість роботи обладнання на етапі дослідження, год;

C_e – вартість 1 кВт-години електроенергії, грн; (вартість електроенергії визначається за даними енергопостачальної компанії), прийmemo $C_e = 7,50$ грн;

K_{eni} – коефіцієнт, що враховує використання потужності, $K_{eni} < 1$;

η_i – коефіцієнт корисної дії обладнання, $\eta_i < 1$.

$$B_e = 0,05 \cdot 280,0 \cdot 7,50 \cdot 0,95 / 0,97 = 105,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.12 – Витрати на електроенергію

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Ноутбук ASUS Vivobook 15 X1500EA-BQ3733 (90NB0TY6-M040W0) Transparent Silver / 15.6" IPS Full HD / Intel Core i3-1115G4 / RAM 12 ГБ / SSD 512 ГБ	0,05	280,0	105,00
Обчислювальний комплекс та комп'ютеризована система проектування ARTLINE D31 (D31v05) Дисплей 7" / 1280*800 / Intel Core i5-13400F / Водяне охолодження CPU / Asus PRIME B660M-A D4 / 32GB DDR4-3600 / SSD 1TB NVMe / nVidia GeForce RTX 4070 12GB	0,42	280,0	882,00
Робоче місце розробника автоматизованої системи	0,08	280,0	168,00
Пристрій графічного виводу інформації	0,22	3,3	5,45
Офісна оргтехніка	0,45	1,6	5,47
Обладнання відтворення звуку	0,05	100,0	37,50
Всього			1203,41

4.3.9 Службові відрядження

До статті «Службові відрядження» дослідної роботи на тему «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання» належать витрати на відрядження штатних працівників, працівників організацій, які працюють за договорами цивільно-правового характеру, аспірантів, зайнятих розробленням досліджень, відрядження, пов'язані з проведенням випробувань машин та приладів, а також витрати на відрядження на наукові з'їзди, конференції, наради, пов'язані з виконанням конкретних досліджень.

Витрати за статтею «Службові відрядження» відсутні.

4.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації

Витрати за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації» розраховуємо як 30...45% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{cn} = (Z_o + Z_p) \cdot \frac{H_{cn}}{100\%}, \quad (4.15)$$

де H_{cn} – норма нарахування за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації», прийmemo $H_{cn} = 30\%$, $B_{cn} = (53200,00 + 714,35) \cdot 30 / 100\% = 16174,30$ грн.

4.3.11 Інші витрати

До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками.

Витрати за статтею «Інші витрати» розраховуємо як 50...100% від суми основної заробітної плати дослідників та робітників за формулою:

$$I_g = (Z_o + Z_p) \cdot \frac{H_{ig}}{100\%}, \quad (4.16)$$

де H_{ig} – норма нарахування за статтею «Інші витрати», прийmemo $H_{ig} = 50\%$.

$$I_g = (53200,00 + 714,35) \cdot 50 / 100\% = 26957,17$$
 грн.

4.3.12 Накладні (загальновиробничі) витрати

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов'язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов'язані з набором робочої сили; витрати на оплату послуг банків; витрати, пов'язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін.

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100...150% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{нзв} = (Z_o + Z_p) \cdot \frac{H_{нзв}}{100\%}, \quad (4.17)$$

де $H_{нзв}$ – норма нарахування за статтею «Накладні (загальновиробничі) витрати», прийmemo $H_{нзв} = 100\%$.

$$B_{нзв} = (53200,00 + 714,35) \cdot 100 / 100\% = 53914,35 \text{ грн.}$$

Витрати на проведення науково-дослідної роботи на тему «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = Z_o + Z_p + Z_{доп} + Z_n + M + K_e + B_{спец} + B_{прз} + A_{обл} + B_e + B_{св} + B_{сп} + I_e + B_{нзв}. \quad (4.18)$$

$$B_{заг} = 53200,00 + 714,35 + 5391,43 + 13047,27 + 3317,48 + 16851,45 + 5800,08 + 23823,88 + 10233,47 + 1203,41 + 0,00 + 16174,30 + 26957,17 + 53914,35 = 230628,65 \text{ грн.}$$

Загальні витрати $ЗВ$ на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$ЗВ = \frac{B_{заг}}{\eta}, \quad (4.19)$$

де η - коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, прийmemo $\eta=0,95$.

$$ЗВ = 230628,65 / 0,95 = 242767,00 \text{ грн.}$$

4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором

В ринкових умовах узагальнюючим позитивним результатом, що його може отримати потенційний інвестор від можливого впровадження результатів тієї чи іншої науково-технічної розробки, є збільшення у потенційного інвестора величини чистого прибутку.

Результати дослідження проведені за темою «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання» передбачають комерціалізацію протягом 4-х років реалізації на ринку.

В цьому випадку майбутній економічний ефект буде формуватися на основі таких даних:

ΔN – збільшення кількості споживачів пристрою, у періоди часу, що аналізуються, від покращення його певних характеристик;

Показник	1-й рік	2-й рік	3-й рік	4-й рік
Збільшення кількості споживачів, осіб	300	500	500	400

N – кількість споживачів які використовували аналогічний пристрій у році до впровадження результатів нової науково-технічної розробки, прийmemo 2800 осіб;

C_o – вартість пристрою у році до впровадження результатів розробки, прийmemo 8000,00 грн;

$\pm \Delta C_o$ – зміна вартості пристрою від впровадження результатів науково-технічної розробки, прийmemo 960,00 грн.

Можливе збільшення чистого прибутку у потенційного інвестора $\Delta \Pi_i$ для кожного із 4-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою [41]:

$$\Delta \Pi_i = (\pm \Delta C_o \cdot N + C_o \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\rho}{100}\right), \quad (4.20)$$

де λ – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2023 році ставка податку на додану вартість складає 20%, а коефіцієнт $\lambda = 0,8333$;

ρ – коефіцієнт, який враховує рентабельність інноваційного продукту).
Прийmemo $\rho = 45\%$;

ρ – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2023 році $\rho = 18\%$;

Збільшення чистого прибутку 1-го року:

$$\Delta\Pi_1 = (960,00 \cdot 2800,00 + 8960,00 \cdot 300) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 1646507,52 \text{ грн.}$$

Збільшення чистого прибутку 2-го року:

$$\Delta\Pi_2 = (960,00 \cdot 2800,00 + 8960,00 \cdot 800) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 3018597,12 \text{ грн.}$$

Збільшення чистого прибутку 3-го року:

$$\Delta\Pi_3 = (960,00 \cdot 2800,00 + 8960,00 \cdot 1300) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 4390686,72 \text{ грн.}$$

Збільшення чистого прибутку 4-го року:

$$\Delta\Pi_4 = (960,00 \cdot 2800,00 + 8960,00 \cdot 1700) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 5488358,40 \text{ грн.}$$

Приведена вартість збільшення всіх чистих прибутків $\Pi\Pi$, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$\Pi\Pi = \sum_{i=1}^T \frac{\Delta\Pi_i}{(1 + \tau)^t}, \quad (4.21)$$

де $\Delta\Pi_i$ – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

T – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, $\tau = 0,14$;

t – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$$\begin{aligned} \Pi\Pi &= 1646507,52/(1+0,14)^1 + 3018597,12/(1+0,14)^2 + 4390686,72/(1+0,14)^3 + \\ &+ 5488358,40/(1+0,14)^4 = 1444304,84 + 2322712,47 + 2963588,47 + 3249548,76 = \\ &= 9980154,54 \text{ грн.} \end{aligned}$$

Величина початкових інвестицій PV , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{инв} \cdot 3B, \quad (4.22)$$

де $k_{инв}$ – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо $k_{инв} = 2$;

$3B$ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 242767,00 грн.

$$PV = k_{инв} \cdot 3B = 2 \cdot 242767,00 = 485534,00 \text{ грн.}$$

Абсолютний економічний ефект $E_{абс}$ для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = III - PV \quad (4.23)$$

де III – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 9980154,54 грн;

PV – теперішня вартість початкових інвестицій, 485534,00 грн.

$$E_{абс} = III - PV = 9980154,54 - 485534,00 = 9494620,54 \text{ грн.}$$

Внутрішня економічна дохідність інвестицій E_e , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_e = T_{жс} \sqrt{1 + \frac{E_{абс}}{PV}} - 1, \quad (4.24)$$

де $E_{абс}$ – абсолютний економічний ефект вкладених інвестицій, 9494620,54 грн;

PV – теперішня вартість початкових інвестицій, 485534,00 грн;

$T_{жс}$ – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримання позитивних результатів від її впровадження, 4 роки.

$$E_г = \sqrt[T_{жс}]{1 + \frac{E_{абс}}{PV}} - 1 = (1 + 9494620,54/485534,00)^{1/4} = 1,13.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій $\tau_{мін}$

:

$$\tau_{мін} = d + f, \quad (4.25)$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні $d = 0,12$;

f – показник, що характеризує ризикованість вкладення інвестицій, приймемо 0,4.

$\tau_{мін} = 0,12 + 0,4 = 0,52 < 1,13$ свідчить про те, що внутрішня економічна дохідність інвестицій $E_г$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки вища мінімальної внутрішньої дохідності. Тобто інвестувати в науково-дослідну роботу за темою «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання» доцільно.

Період окупності інвестицій $T_{ок}$ які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_г}, \quad (4.26)$$

де $E_г$ – внутрішня економічна дохідність вкладених інвестицій.

$$T_{ок} = 1 / 1,13 = 0,89 \text{ р.}$$

$T_{ок} < 3$ -х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

4.5 Висновки до розділу

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання» становить 33,0 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки вище середнього).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,37 рази.

Також термін окупності становить 0,89 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою «Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання».

ВИСНОВКИ

Вході виконання магістерської кваліфікаційної роботи отримані такі результати.

В першому розділі роботи проведено аналіз факторів впливу на якість комунікації в контакт-центрах і зроблено висновок, що основним з них є темп мовлення. Тому в роботі запропоновано метод автоматизації контролю і регулювання темпу мовлення на основі інформації про кількість і тривалість складів в мовленнєвому сигналі.

В другому розділі роботи виконано математичне обґрунтування запропонованого методу виділення складів в мовленнєвому потоці, розроблена структурна схема пристрою та алгоритму виділення складів в сигналі мовлення. .

В третьому розділі розроблено алгоритми обробки сигналу мови, а також програмне забезпечення для їх реалізації. Проведене тестування програм підтвердило адекватність їх роботи для визначення темпу мовлення мовців.

В економічній частині проведено техніко-економічне огрунтування доцільності і ефективності впровадження результатів розробки в реальні проекти.

Розроблені апаратні і програмні засоби можуть бути застосовані як для підвищення надійності і оперативності роботи контакт-центрів різних диспетчерських систем, так і для підвищення точності і швидкості розпізнавання мови.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Стецюк В. І. Системи автоматизованого управління телекомунікаційних мереж / В. І. Стецюк, В. В. Мішан // Вісник Хмельницького національного університету. Серія: технічні науки. – 2018. – № 6. Том 2. – С. 178–182.
2. Smith B. Effects of speech rate on personality perception // Language and Speech. — 1975. — Vol. — 18 (2). — P. 28—35.
3. Bykov N.M. Development of effective strategy of pattern recognition / Bykov N.M., Kuzmin I.V., Jakovenko A.I. // Proceedings of SPIE. – 2000. - Vol.4425. - P. 76-82. - ISBN 0-8194-4807-9.
4. Биков Микола. Дикторонезалежне описання образів звуків в системах розпізнавання сигналів мови / Биков Микола, Раїмі Абдурахман, Биков Максим // Вісник технічної метрології та метрології. - 2006. – Випуск 66. - С.13-17. – ISSN 0368-6418.
5. Биков М.М. Моделювання процесу аналізу і класифікації голосових команд / М.М. Биков, Т.В. Гришук // Монографія. — Вінниця: ВНТУ, 2009. — 129 с.
6. Темп мовлення. [Електронний ресурс]. – Режим доступу https://uk.wikipedia.org/wiki/Темп_мовлення
7. Темп мовлення // Літературознавча енциклопедія : у 2 т. / авт.-уклад. Ю. І. Ковалів. — Київ : ВЦ «Академія», 2007. — Т. 2 : М — Я. — С. 473.
8. [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/Темп_мовлення
9. Pellegrino F., Coupé C., Marsico E. A cross-language perspective on speech information rate // Language. — 2011. — Vol. 87 (3). — P. 539—558
10. Іщенко О. С. Голосні звуки української мови залежно від темпу мовлення. — К.: Ін-т української мови НАН України, 2012 — 220 с.

11. [Електронний ресурс]. – Режим доступу: <https://relizua.com/besplatnaya-multimedia/1802-praat.html>
12. VOSK Offline Speech Recognition API [Електронний ресурс] – Режим доступу до ресурсу: [Електронний ресурс]. – Режим доступу: <https://alphacephei.com/vosk/>
13. [Елекпронний ресурс]/ - Режим доступу: <https://dou.ua/forums/topic/31961/>
14. Вербич Н. С. Інтонація переконування в публічному мовленні. — К.: Ін-т української мови НАН України; Луцьк: Терен, 2011. — 208 с.
15. Типологія інтонації мовлення / Відп. Ред. А. Й. Багмут. К.: Наукова думка, 1977. — 494 с.
16. Ладощко О. М. Дослідження характеристик вокалізованих пауз спонтанної мови / О. М. Ладощко // Акустичний симпозиум «Консонанс 2011», 27-29 вересня 2011 р.: тези доп. – К. – 2011. – С.188- 193.
17. Ладощко О. М. Дослідження впливу характеристик телефонного каналу зв'язку на надійність розпізнавання фонем / О. М. Ладощко // III Міжнародна науково-технічна конференція студентів, аспірантів та молодих вчених. Інформаційні управляючі системи та комп'ютерний моніторинг (ІУС і КМ-2012), 16- 18 квітня, 2012р.: тези доп. – Донецьк. – 2012. – С. 143-148
18. Sazhok N.N. Clustering words for construction of a linguistic model for the automatic recognition of the speech signal // Cybernetics and Computer Science: Interagency collection of scientific papers. – Issue. 170. – Kyiv, 2012.
19. T. Van Pham. Wavelet analisys for robust speech processing and applications. – 2007. –171 p.
20. Vintsiuk T.K. The analysis, pattern recognition and semantic interpretation of the speech signals. – Kiev : Naukova Dumka, 1987. – 263 p
21. Рашкевич Ю. Маркування мовного сигналу на основі використання спектральної метрики / Ю. Рашкевич, Д. Пелешко, Ю. Пелех, І. Ізонін // Матеріали міжнародної наукової конференції [“Інтелектуальні системи

прийняття рішень і проблеми обчислювального інтелекту”, (ISDMCI-2013)], (Херсон, 20-24 травня 2013) / ХНТУ. – Херсон, 2013. – С.258-260.

22. Rashkevych Yu. Speech signal marking on the base of local magnitude and invariant segmentation / Yu. Rashkevych, D. Peleshko, I. Pelekh, I. Izonin // Journal Mathematical Modeling and Computing. – Lviv, 2014. – Vol. 1, No. 2. – P. 234–244.

23. Пелешко Д.Д. Виділення псевдоінваріантів та квазістаціонарних ділянок мовних сигналів на основі сингулярних розкладів / Д.Д. Пелешко, А.М. Ковальчук, Ю.М. Пелех, В.І. Киричук // Вісник Національного університету “Львівська політехніка”: Комп’ютерні науки та інформаційні технології. – Львів. –2012. – №732. – С. 58-66.

24. Ruske C., Schotola F. An approach to speech recognition using syllabic decision units. – Proc. 1978, IEEE ICASSP, Tulsa, 1978. – N.Y., 1978, pp. 772 – 725.

25. Jelinek F. Development of an experimental device that recognizes separately pronounced words. // Proceeding of IEEE. - 1985. – Т. 73. – No. 11. – P. 91 – 100.

26. Rabiner L.R., Shafer R.V. Digital Processing of Speech Signals. – Prentice Hall: New Jersey, 1978 – 496 p.

27. Биков М.М. Надійний метод виділення складових сегментів у мовному сигналі / М. М. Биков, В. В. Ковтун, Н. Г. Савінова; Наукові праці ВНТУ: Автоматика і інформаційно-вимірювальна техніка. – 2007, №1.– с.1-6

28. Speech Recognition for Ukrainian. URL: <https://github.com/egorsmkv/speech-recognition-uk>

29. Zwakman DS et al. Voice usability scale: measuring the user experience with voice assistants. In: 2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS). IEEE; 2020.

30. Voice Assistant Use Cases: Business Implementations of VUIs in 2023. URL: <https://masterofcode.com/blog/voice-assistant-use-cases-businessimplementations-of-vuis-in-2021>

31. Speech Recognition — Feature Extraction MFCC & PLP. [Електронний ресурс] - Режим доступу до ресурсу:<https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9>.

32. Speech Recognition. Wikipedia [Електронний ресурс] - Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Speech_recognition

33. Speech and Language Processing. Daniel Jurafsky & James H. Martin [Книга] - Режим доступу до ресурсу: <https://web.stanford.edu/~jurafsky/slp3/3.pdf>

34. Speech Recognition — GMM, HMM. Jonathan Hui [Електронний ресурс] - Режим доступу до ресурсу: <https://jonathan-hui.medium.com/speech-recognition-gmm-hmm-8bb5eff8b196>

35. Acoustic Model Selection for Recognition of Regional Accented Speech. Maryam Najafian [Стаття] - Режим доступу до ресурсу: https://www.researchgate.net/figure/Structure-of-a-DNN-HMM-system-5_fig8_292365466

36. Забара О.С. Моделювання в середовищі Матлаб / О.С. Забара, О.О. Гагарін, І.М. Кузьменко, Ю.Д. Щербашин. – Київ, 2011. – 300 с.

37. Лазарєв Ю. Ф. Л17 Довідник з MATLAB / Електронний навчальний посібник з курсового і дипломного проектування. – К.: НТУУ "КПІ", 2013. – 132 с.

38. Yu D. Automatic Speech Recognition / D. Yu, L. Deng. – London: Springer, 2015. – 321 с.

39. ARPA Language models [Електронний ресурс] – Режим доступу до ресурсу: <https://cmusphinx.github.io/wiki/arpaformat/>

40. A time delay neural network architecture for efficient modeling of long temporal contexts. Vijayaditya Peddinti, Daniel Povey, Sanjeev Khudanpur [Електронний ресурс]. - Режим доступу до ресурсу: https://www.danielpovey.com/files/2015_interspeech_multisplce.pdf

41. Zou J., Han Y., So S.S. (2008) Overview of Artificial Neural Networks. In: Livingstone D.J. (Eds) Artificial Neural Networks. Methods in Molecular Biology™, vol/ 458. pp. 14-22

42. Neural network. Wikipedia. URL: https://en.wikipedia.org/w/index.php?title=Neural_network

43. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.

ДОДАТКИ

Додаток А
ПРОТОКОЛ
ПЕРЕВІРКИ КВАЛІФІКАЦІЙНОЇ РОБОТИ
НА НАЯВНІСТЬ ТЕКСТОВИХ ЗАПОЗИЧЕНЬ

Назва роботи: « Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання»

Тип роботи: Магістерська кваліфікаційна робота

(БДР, МКР)

Підрозділ КСУ,
ФШТА

(кафедра, факультет)

Показники звіту подібності Unicheck

Оригінальність 89,8%

Схожість 10,2%

Аналіз звіту подібності (відмітити потрібне)

Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.

Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її автора. Роботу направити на розгляд експертної комісії кафедри.

Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Особа, відповідальна за перевірку

ДУБОВОЙ

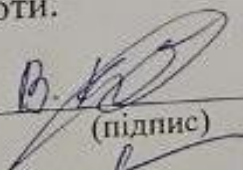
(підпис)

Володимир

(прізвище, ініціали)

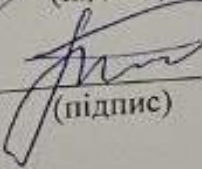
Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

Автор роботи


 (підпис)

Вадим Кошельник
 (прізвище, ініціали)

Керівник роботи



 (підпис)

Микола БИКОВ
 (прізвище, ініціали)

Додаток Б
(обов'язковий)
ВНТУ

ЗАТВЕРДЖУЮ

В.о. зав. кафедри КСУ
д.т.н., проф. Марія ІОХИМЧУК


« 6 » 10 2023 р.

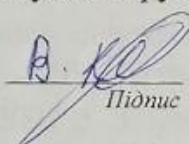
ТЕХНІЧНЕ ЗАВДАННЯ

на виконання магістерської кваліфікаційної роботи

«Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання»

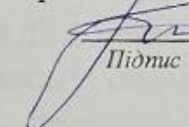
08-33.МКР.003.00.000 ТЗ

Студент групи 2 АКІТ-22 м


Підпис

Вадим КОШЕЛЬНИК
Ім'я ПРІЗВИЩЕ

Керівник к.т.н., проф. каф. КСУ


Підпис

Микола БИКОВ
Ім'я ПРІЗВИЩЕ

1. Назва та галузь застосування

1.1. Назва – Автоматизація контролю темпу мови в комунікаційних центрах систем критичного використання

1.2. Галузі застосування – Автоматика та приладобудування, інтелектуальні інформаційні системи

2. Підстава для проведення розробки.

Тема магістерської кваліфікаційної роботи затверджена наказом по ВНТУ від “18” вересня 2023 року №247

3. Мета та призначення розробки.

Метою магістерської кваліфікаційної роботи є підвищення ефективності і надійності спілкування між оператором та користувачем комунікаційних центрів систем критичного використання за рахунок розробки методу і програм автоматичного контролю темпу мовлення.

4. Джерела розробки.

Магістерська кваліфікаційна робота виконується вперше. В ході проведення розробки повинні використовуватись такі документи:

1. N.M. Vykov, I.V. Kuzmin, A.I. Yakovenko. Development of effective strategy of pattern recognition. – Proceedings of SPIE, 2001, Vol. 4225, pp.76 – 83.
2. Нікіфоров С.Н. та інш. Практичний алгоритм визначення темпу мови для використання в контакт-центрах / С.Н. Нікіфоров // Мовні технології. – 2010, №1.
3. Биков М.М., Грищук Т.В. Методи підвищення дикторнезалежності опису і розпізнавання мовної інформації в мережі INTERNET // “Інтернет – Освіта – Наука – 2002”, третя міжнародна конференція ІОН – 2002, 8 – 12 жовтня 2002 р. Збірник матеріалів конференції. – Вінниця: УНІВЕРСУМ – Вінниця, 2002. – Том 2.– С. 329 – 332.
4. М. М. Биков, В. В. Ковтун, Н. Г. Савінова. Надійний метод виділення складових сегментів у мовному сигналі // Наукові праці ВНТУ: Автоматика і інформаційно-вимірювальна техніка. – 2007, №1. – С. 1 – 6.

5. Вимоги до розробки.

5.1. Перелік головних функцій:

- функція wavread('name_file.wav') являється стандартною функцією Matlab, яка зчитує в змінну якій присвоюється функція файл з

записами акустичних сигналів. Значення амплітуд знаходиться в межах $[-1; 1]$

- функція `Filtr_ogib` при передачі їй управління проводить фільтрування сигналу смуговим фільтром і формує вектор енергії звукового сигналу з частотою дискретизації 11 кГц
- функція `Gran_word()` визначає границі вимовленого слова, завантаженого функцією `wavread()`, тобто відділяє слово від шуму для подальшої роботи
- функція `fn_z()` згладжує енергії шляхом знаходження середнього значення енергії в кожній точці.

5.2. Основні технічні вимоги до розробки.

5.2.1. Вимоги до програмної платформи:

- WINDOWS 7/10;
- MATLAB 7.0;

5.2.2. Умови експлуатації системи:

- робота на стандартних ПЕОМ в приміщеннях зі стандартними умовами;
- можливість цілодобового функціонування системи;
- текст програмного забезпечення системи є цілком закритим.

6. Стадії та етапи розробки.

6.1 Пояснювальна записка:

- | | |
|--|--------------|
| - Аналіз методів обробки мовного сигналу з метою визначення швидкості мовлення. Постановка задач дослідження | 12.09.2023р. |
| - Удосконалення технології контролю темпу мови а комунікаційних центрах систем критичного використання | 22.09.2023р. |
| - Розробка алгоритмічного та програмного забезпечення | 3.10.2023р. |
| - Підготовка економічної частини | 12.10.2023р. |
| - Апробація результатів дослідження | 02.11.2023р. |
| - Публікації | |

- Оформлення пояснювальної записки, графічного матеріалу і презентації 20.11.2023р.
- Захист МКР 19.12.2023р.

6.2 Графічні матеріали:

- структурна схема системи: 11. 11. 2023 р.
- UML діаграма варіантів використання: 12. 11. 2023 р.
- UML-діаграма діяльності: 13. 11. 2023 р.
- UML-діаграма класів: 14. 11. 2023 р.
- плакати демонстраційні 17. 11. 2013 р.

7. Порядок контролю і приймання.

- 7.1. Хід виконання магістерської кваліфікаційної роботи контролюється керівником роботи, консультантами з економічної частини. Рубіжний контроль провести до «12» грудня 2023.
- 7.2. Атестація проекту здійснюється на попередньому захисті. Попередній захист магістерської кваліфікаційної роботи провести до «14» грудня 2023р.
- 7.3. Підсумкове рішення щодо оцінки якості виконання магістерської кваліфікаційної роботи приймається на засіданні ДЕК. Захист магістерської кваліфікаційної роботи провести «19» грудня 2023 р.

Додаток В

(довідниковий)

Лістинги програм

Текст функції gran_word(og)

```

function [t12, t22]=gran_word(og);
t11=0;t12=0;t21=0;t22=0;
Emax=0.0000962844;
Emin=0.0000837963;
Pnoise=5.991;
E=og(2,:).^2;
int=find(E>=Emax);
t11=int(1);
t21=int(length(int));
int1=find(E(1:t11)>=Emin);
t12=int1(length(int1));
int2=find(E(t21:length(E))>=Emin);
t22=int2(length(1))+t21;
Per=og(1,:);
int3=find(Per(1:t12-1)>Pnoise);
if isempty(int3)==[]
    int3=t12-10;
end
t12_new=int3(length(int3));
if (t12_new-t12)>10
    t12=t12-10;
else
    t12=t12_new;
end
int4=find(Per(t22+1:length(Per))>Pnoise);
if isempty(int4)==[]
    int4(1)=10;
end
t22_new=t22+int4(1);
if (t22_new-t22)>10
    t22=t22+10;
else
    t22=t22_new;
end
end

```

Текст функції ogib(x,n)

```

function [S]=ogib(x,n);
ny=x;
j=1; A(j)=0; Am(j)=0;
n=round(length(ny)/(n*2)); k=0; m=0; Per(j)=0;
ii=0:n-1;h=1;
w=0.54-0.46*cos((ii.*2*pi)/n);
for i=1:length(ny)-1
    A(j)=abs(ny(i))*w(h)+A(j);
    Am(j)=ny(i)+Am(j);
    k=k+1;
    if sign(ny(i))~=sign(ny(i+1))
        m=m+1;
    end
    if k>=n-1
        X(j)=i-n;
        A(j)=A(j)/k;
        Am(j)=Am(j)/k;
        k=0;
    end
end

```

```

    Per(j)=m; m=0;
    j=j+1; A(j)=0; Am(j)=0;
    h=1;
end
h=h+1;
end
X(j)=i-k/2;
if k~=0
    A(j)=A(j)/k;
    Am(j)=Am(j)/k;
end
Per(j)=m;
S(1,:)=Per;
S(2,:)=A;
S(3,:)=Am;

```

Текст програми виділення границь слова granуса

```

seg=128;
Flag_all='True';
while Flag_all=='True'
fid=uiigetfile('*.wav','Open wave file');% -----
if fid==0
    msgbox('Please select *.wav file');
break
end
y=wavread(fid);
%y=wavread('four.wav');
og=ogib(y,seg);
[t1 t2]=gran_word(og);
A=og(2,:);
y1=y(t1*seg:t2*seg);
wavwrite(y1,'fr.wav');
figure(1);
subplot(2,1,1)% -----
plot(1:30:length(y),y(1:30:length(y)), 'k');
subplot(2,1,2)
plot(1:30:length(y),y(1:30:length(y)), 'k'), hold on
line([t1*seg t1*seg],[-0.1 0.15]), hold on
line([t2*seg t2*seg],[-0.1 0.15])
pause(1)
q_ans=questdlg('Do you want to continue?');
Flag_all='No ';
if strcmp(q_ans,'No')==1
    Flag_all='No ';
else
    Flag_all='True';
end
end
end

```

Текст програми Flt_ogib

```

[b,a]=butter(4,[pr1 pr2]*2/z);
ff=filter(b,a,y);
fff=abs(ff);
figure(2);
iii=-127;
i=0;
E=0;
while iii < end_cycl
    iii=iii+128;
    iii_e=iii+127;
    i=i+1;
    N=i;
    Ampl=0;
    for ii=iii:iii_e
        Ampl=Ampl+fff(ii);
    end
end

```

```

end
f(nf,i)=Ampl/128;
end

```

Текст програми Obrs

```

z1=wavread('fr.wav','size');
z=z1(1);
end_cycl=z-256;
y=wavread('fr.wav');
p1_1=250; p1_2=540;
p2_1=490; p2_2=2500;
pr1=p1_1; pr2=p1_2; nf=1; FLT_OGIB;
pr1=p2_1; pr2=p2_2; nf=2; FLT_OGIB;
figure(2);
subplot(2,1,1)%-----
plot(f(1,:))
subplot(2,1,2)
plot(f(2,:))

```

Текст програми Koreliatsia

```

kor=f1.*f2;
kor2=fn_z(kor,3);
figure(3);
plot(kor, axis([0 N+5 min(kor)-0.5*max(kor) max(kor)+0.5*max(kor)]));
title('Energia pislia koreliatsii');

```

Текст функції fn_z(x,n)

```

function[arn]=fn_z(ar,ni);
N=length(ar);
arn=zeros(1,N);
for i=1:ni
    arn(i)=ar(i);
end
for i=(N-(ni-1)/2):N
    arn(i)=ar(i);
end
for i=ni:(N-(ni-1)/2)
    for j=(i-(ni-1)/2):(i+(ni-1)/2)
        arn(i)=arn(i)+ar(j);
    end
    arn(i)=arn(i)/ni;
end

```

Текст програми Difer

```

liniya=zeros(N,1);
for i=1:N-1
    fyd(i)=kor2(i+1)-kor2(i);
end
k=1;
for i=1:N-2
    if (fyd(i)*fyd(i+1)<0)
        shag(k)=abs(fyd(i+1))+abs(fyd(i));
        k=k+1;
    end;
end
porog=max(shag)/3;
for i=1:N
    lin(i)=porog;
end

```



```

end
figure(4);
subplot(2,1,1)
plot(kor2), axis([0 N+5 min(kor2) max(kor2)+0.5*max(kor2)]), grid on;
title('Energia pislia korelyatsii');
subplot(2,1,2)
plot(fyd), hold on
plot(1:N-1,liniya(1:N-1),'k'),axis([0 N+5 min(fyd)+0.5*min(fyd) max(fyd)+0.5*... max(fyd) ]);
title('dyferenciuyvannya korelyatsii'), grid on, hold on
plot(1:N,lin(1:N),'--k'), hold on
plot(1:N,-lin(1:N),'--k');

```

Текст програми Logic

```

for i=1:N-1
    if abs(fyd(i)) > porog
        l(i)=1;
    else l(i)=0;
    end
end
kp=0;
for i=1:N-2
    if l(i)>l(i+1)
        kp=kp+1;
    end
end
p1=0;
p2=0;
k=1;
while k <= N-2
    if(mod(p1,2)==1) & (p1<(kp+1))
        tone(k)=not(l(k));
    else tone(k)=l(k);
    end
    if l(k)>l(k+1)
        p1=p1+1;
    end
    if(mod(p2,2)==1)
        ttwo(k)=not(l(k));
    else ttwo(k)=l(k);
    end
    if l(k)<l(k+1)
        p2=p2+1;
    end
    k=k+1;
end
tone(N-1)=l(N-1);
ttwo(N-1)=l(N-1);
pgl=and(tone,ttwo);
psl=or(pgl,l);
end
figure(5);
subplot(6,1,1)
plot(fyd), hold on
plot(1:N-1,liniya(1:N-1),'k'), axis([0 N+5 min(fyd)+0.5*min(fyd) max(fyd)+0.5*... max(fyd) ]);
title('dyferenciuyvannya korelyatsii'), hold on
plot(1:N,lin(1:N),'--k'),
plot(1:N,-lin(1:N),'--k');
subplot(6,1,2)
stairs(l), axis([0 N+5 -1 2]);
subplot(6,1,3)
stairs(tone), axis([0 N+5 -1 2]);
subplot(6,1,4)
stairs(ttwo), axis([0 N+5 -1 2]);
subplot(6,1,5) stairs(pgl), axis([0 N+5 -1 2]); subplot(6,1,6) stairs(psl), axis([0 N+5 -1 2]);

```

Додаток Г
(обов'язковий)

ІЛЮСТРАТИВНА ЧАСТИНА

АВТОМАТИЗАЦІЯ КОНТРОЛЮ ТЕМПУ МОВИ В КОМУНІКАЦІЙНИХ
ЦЕНТРАХ СИСТЕМ КРИТИЧНОГО ВИКОРИСТАННЯ

Студент групи 2 АКІТ-22м

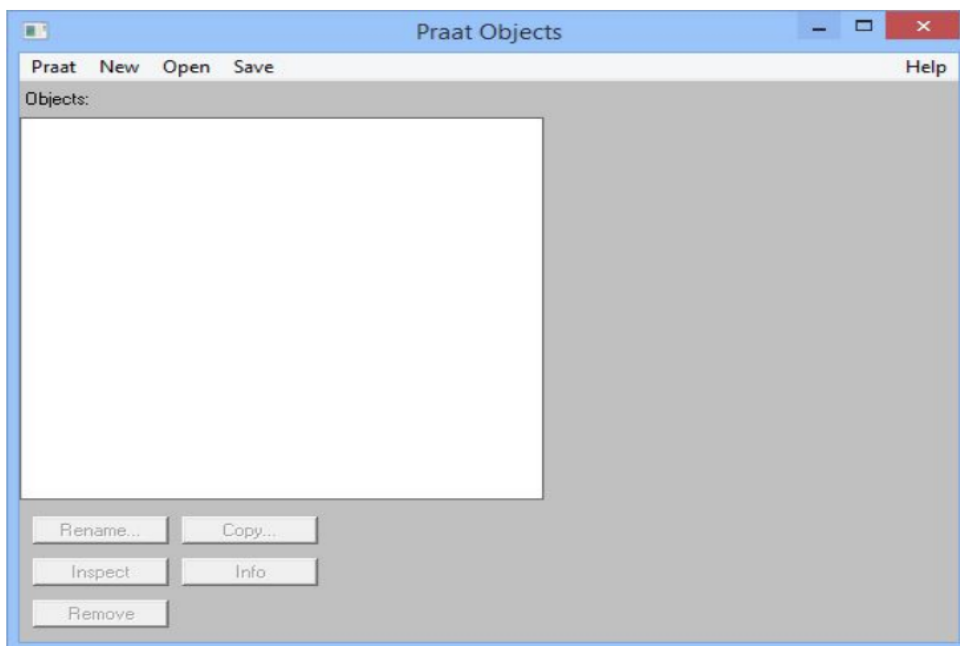
Підпис Вадим КОШЕЛЬНИК
Ім'я ПРІЗВИЩЕ

Керівник к.т.н., доцент, проф. кафедри КСУ

Підпис Микола БИКОВ
Ім'я ПРІЗВИЩЕ

ПРОГРАМНІ ПЛАТФОРМИ ДЛЯ АНАЛІЗУ ПАРАМЕТРІВ МОВЛЕННЯ

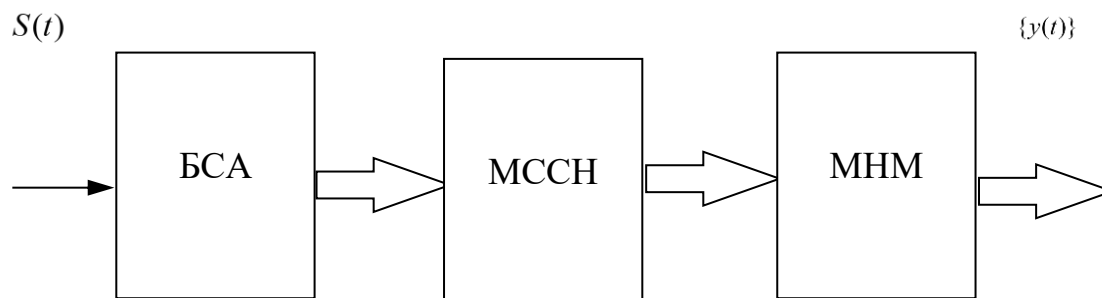
1. PRAAT



2. VOSK

3. VOICE ANALYTICS

УЗАГАЛЬНЕНА МОДЕЛЬ СЛУХОВОЇ СИСТЕМИ



КОРЕЛЯЦІЙНА МАТРИЦЯ СПЕКТРАЛЬНИХ СКЛАДОВИХ СИГНАЛУ МОВЛЕННЯ

$$\|y_x\| = x \cdot x^T = \begin{vmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & \dots & x_1 \cdot x_n \\ x_2 \cdot x_1 & x_2 \cdot x_2 & \dots & x_2 \cdot x_n \\ \dots & \dots & x_i \cdot x_j & \dots \\ x_n \cdot x_1 & x_n \cdot x_2 & \dots & x_n \cdot x_n \end{vmatrix}$$

$\|y_x\|$ - автокореляційна матриця спектральних параметрів мовного сигналу

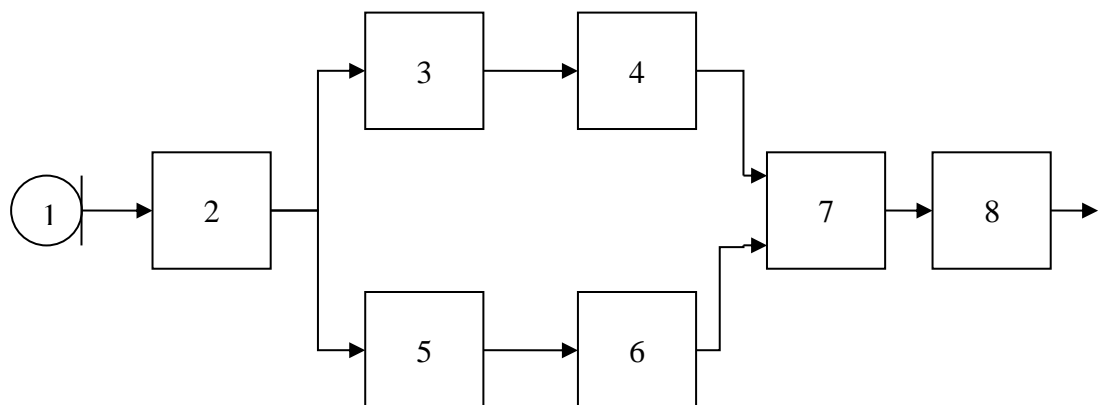
$$U_c(t) = U_{\Delta 1}(t) \cdot U_{\Delta 2}(t)$$

$U_c(t)$ - результуючий сигнал;

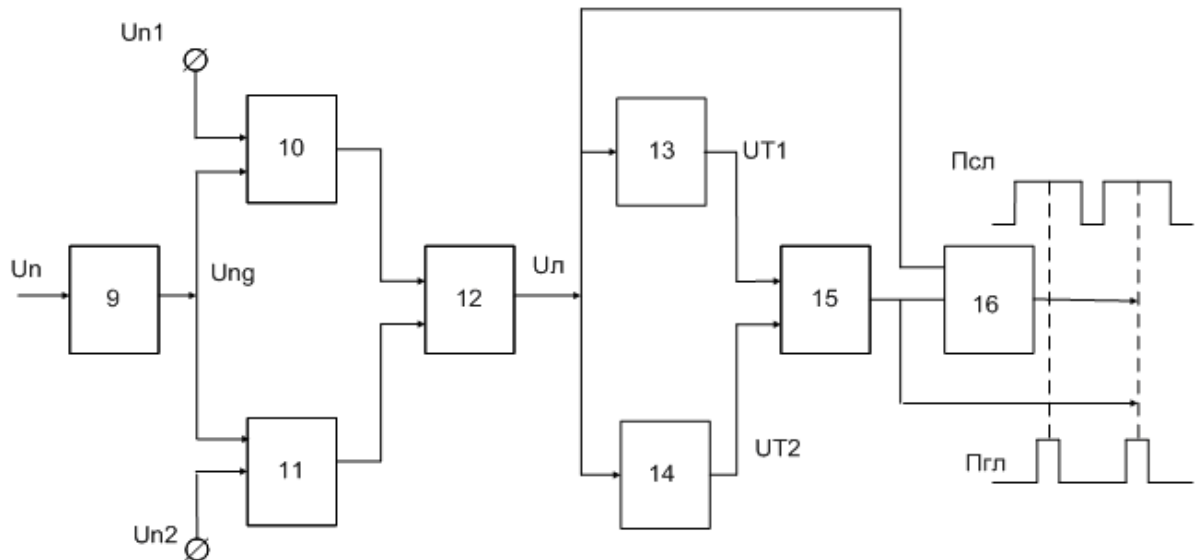
$U_{\Delta 1}(t)$ - огибаюча енергії в смузі частот $\Delta 1$;

$U_{\Delta 2}(t)$ - огибаюча енергії в смузі частот $\Delta 2$.

. СТРУКТУРНА СХЕМА ПРИСТРОЮ ВИДІЛЕННЯ СЕГМЕНТІВ СКЛАДІВ

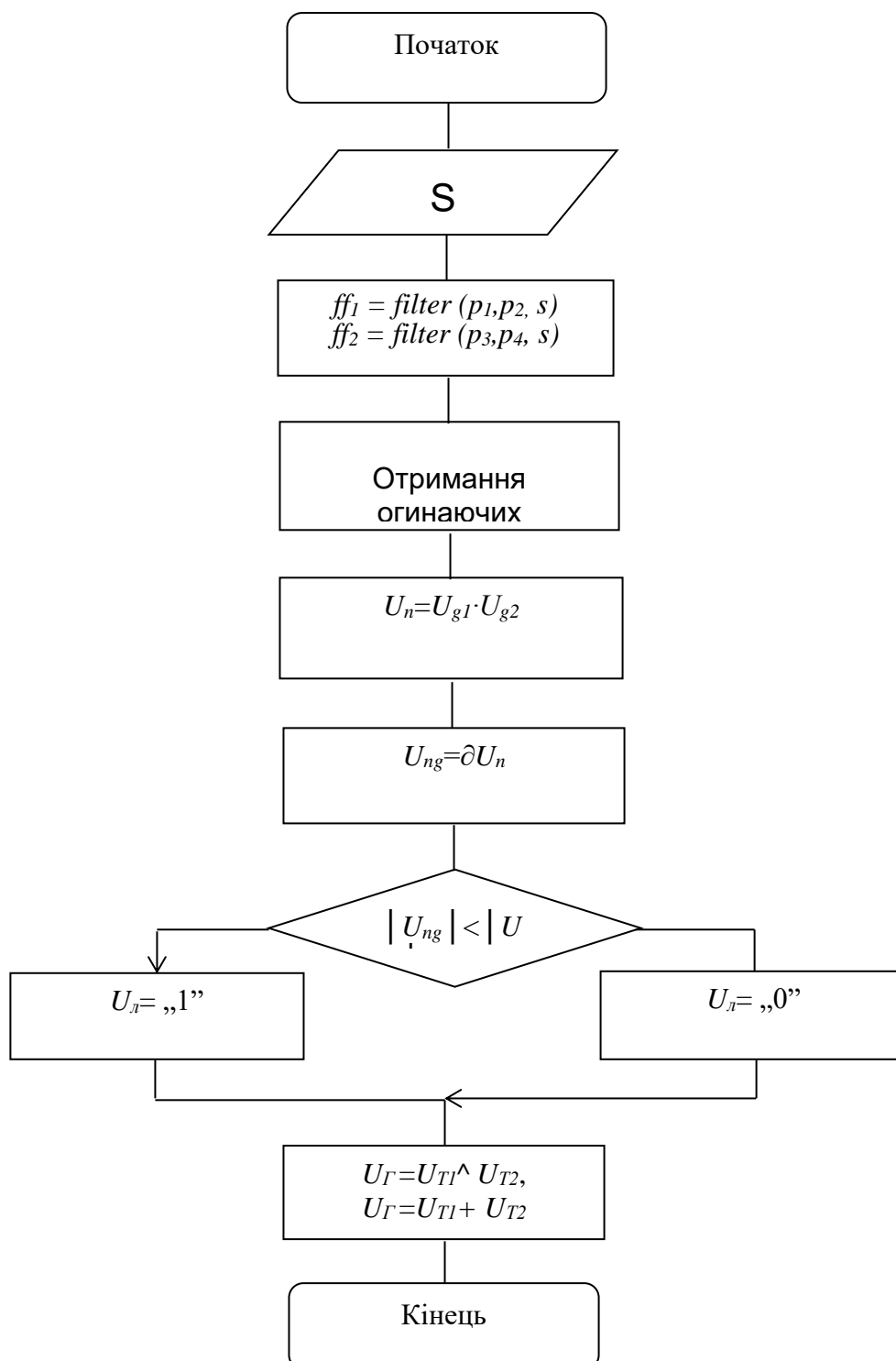


ФУНКЦІОНАЛЬНА СХЕМА ФОРМУВАЧА СЕГМЕНТНИХ ГРАНИЦЬ СКЛАДІВ



- 1 – акустичний датчик;
- 2 – підсилювач;
- 3, 5 – перший і другий смуговий фільтр;
- 4, 6 – перший і другий детектор;
- 7 – перемножувач сигналу;
- 8 – формувач сегментних границь складів

- 9 – диференціатор;
- 10, 11 – перша і друга порогові схеми;
- 12 – логічна схема “АБО” на два входи;
- 13 – лічильний тригер з прямим динамічним управлінням;
- 14 – лічильний тригер з зворотним динамічним управлінням;

АЛГОРИТМ ВИДІЛЕННЯ ОЗНАК СКЛАДІВ

ЧАСОВІ ДІАГРАМИ РОБОТИ ПРИСТРОЮ ВИДІЛЕННЯ СКЛАДІВ

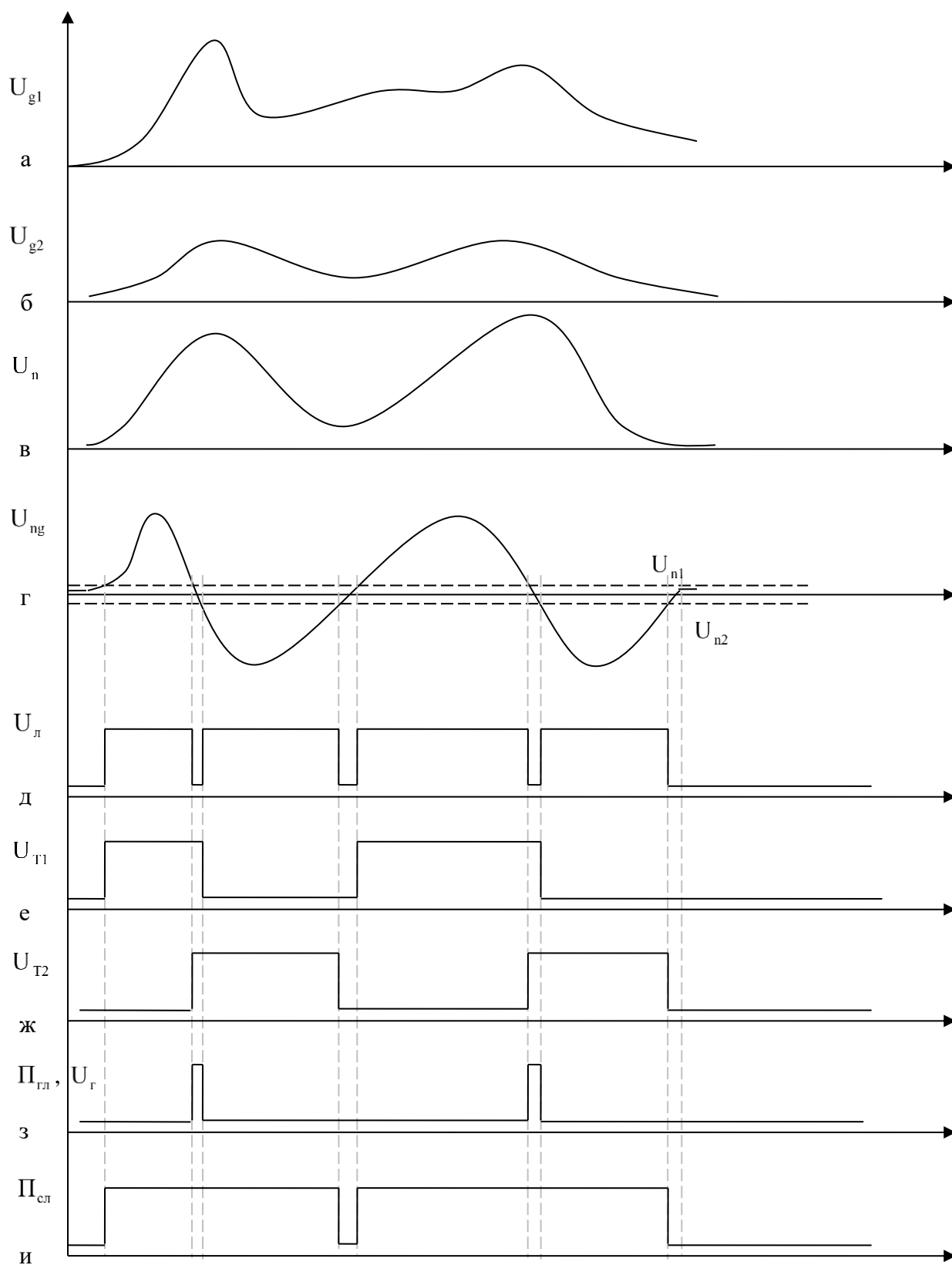


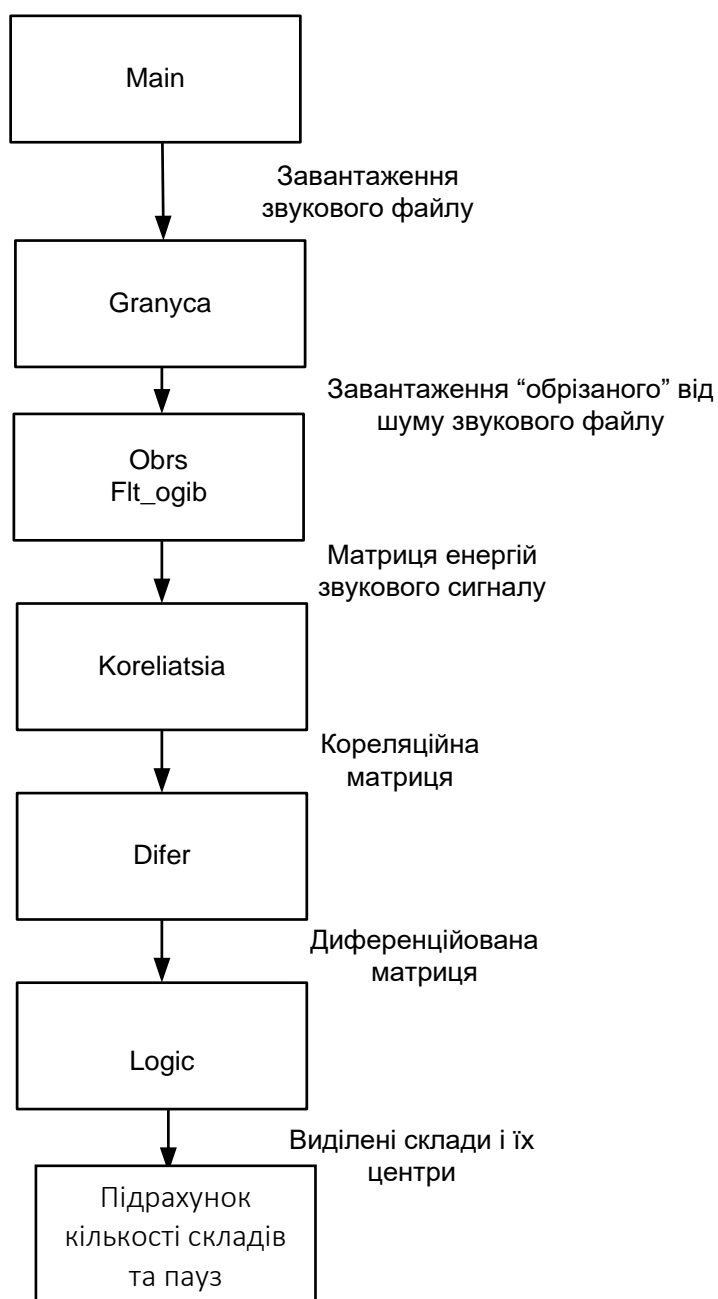
СХЕМА АЛГОРИТМУ ВИЗНАЧЕННЯ ТЕМПУ МОВИ

СХЕМА АЛГОРИТМУ ФУНКЦІЇ GRAN_WORD(OG)

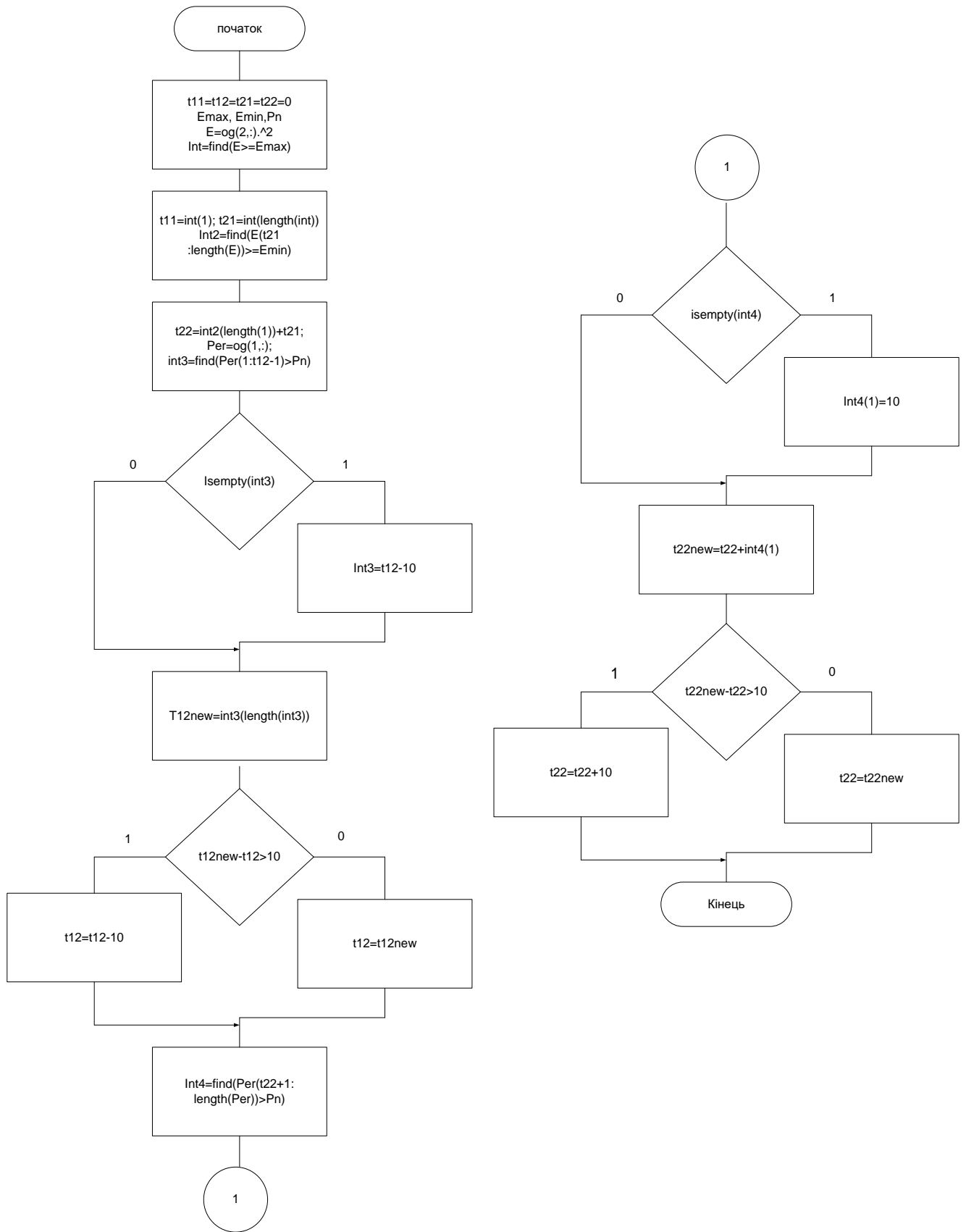
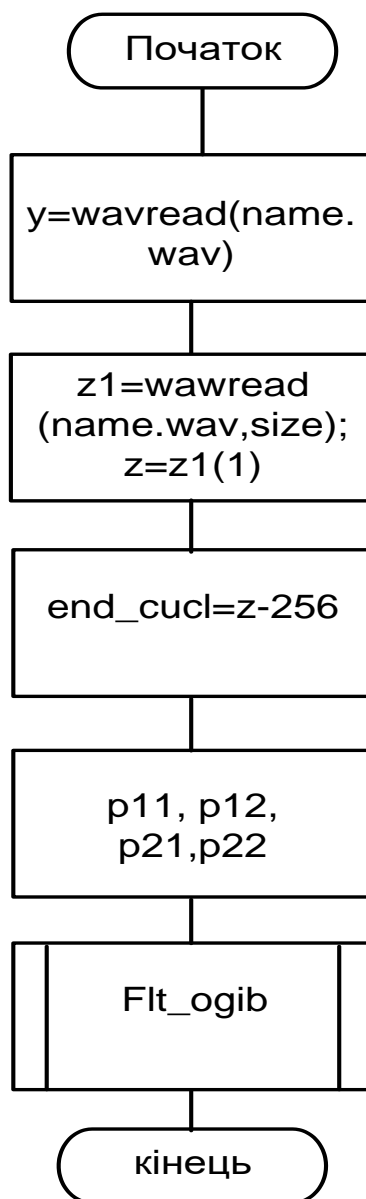
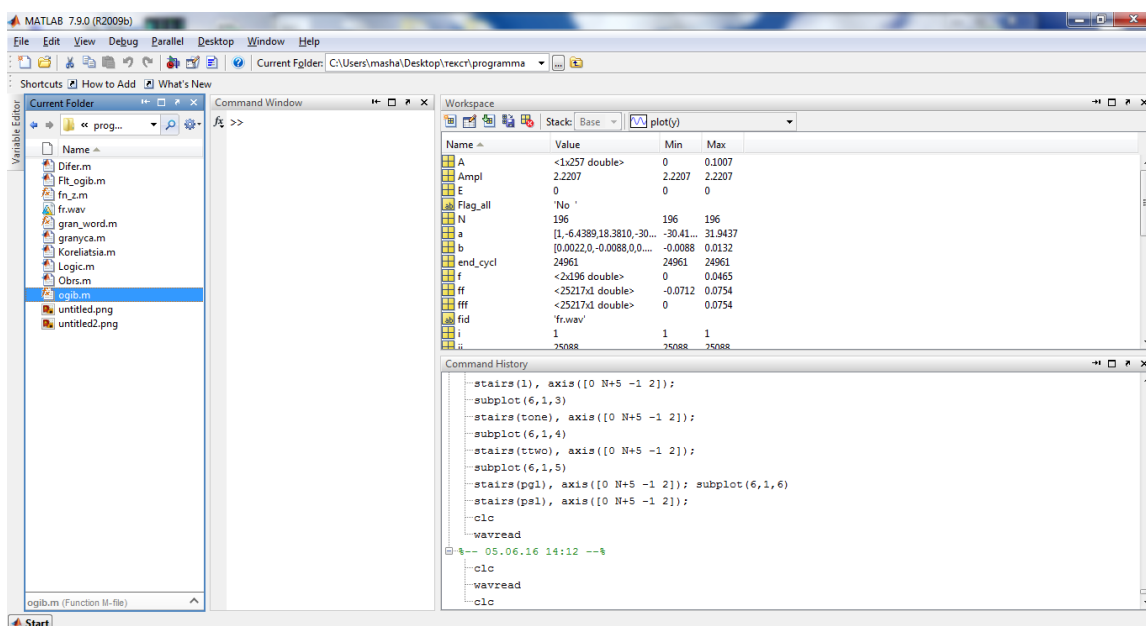


СХЕМА ПІДПРОГРАМИ OBRS



СТВОРЕННЯ ФАЙЛІВ ПРОГРАМИ В СЕРЕДОВИЩІ МАТЛАВ



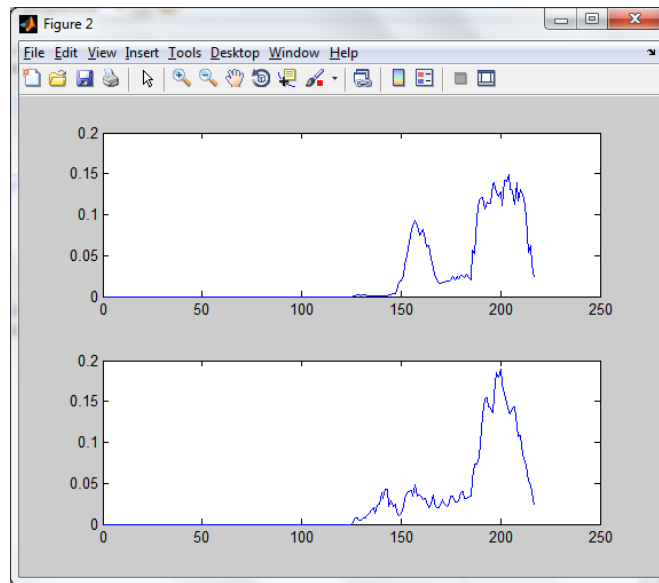
The screenshot displays the MATLAB 7.9.0 (R2009b) environment. The interface is divided into several panes:

- Current Folder:** Shows a list of files and folders, including 'oglib.m' which is currently selected.
- Command Window:** Contains the MATLAB prompt 'f> >>'.
- Workspace:** A table listing variables and their properties. The variables and their values are as follows:

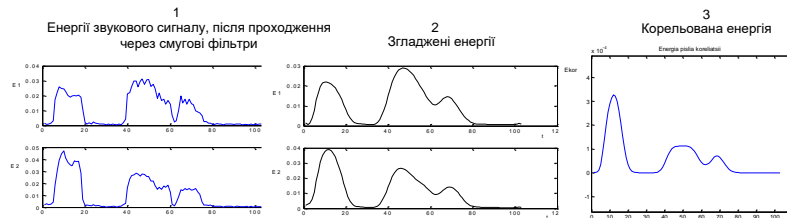
Name	Value	Min	Max
A	<-1x257 double>	0	0.1007
Ampl	2.2207	2.2207	2.2207
E	0	0	0
Flag_all	'No '		
N	196	196	196
a	[1,-6.4389,18.3810,-30...	-30.41...	31.9437
b	[0.0022,0,-0.0088,0,0...	-0.0088	0.0132
end_cycl	24961	24961	24961
f	<-2x196 double>	0	0.0465
ff	<-25217x1 double>	-0.0712	0.0754
fff	<-25217x1 double>	0	0.0754
fid	'fr.wav'		
i	1	1	1
z	75088	75088	75088
- Command History:** Shows the executed commands:

```
--stairs(1), axis([0 N+5 -1 2]);  
--subplot(6,1,3)  
--stairs(tone), axis([0 N+5 -1 2]);  
--subplot(6,1,4)  
--stairs(ttwo), axis([0 N+5 -1 2]);  
--subplot(6,1,5)  
--stairs(pgl), axis([0 N+5 -1 2]); subplot(6,1,6)  
--stairs(psl), axis([0 N+5 -1 2]);  
--clc  
--wavread  
05.06.16 14:12 -->  
--clc  
--wavread  
--clc
```

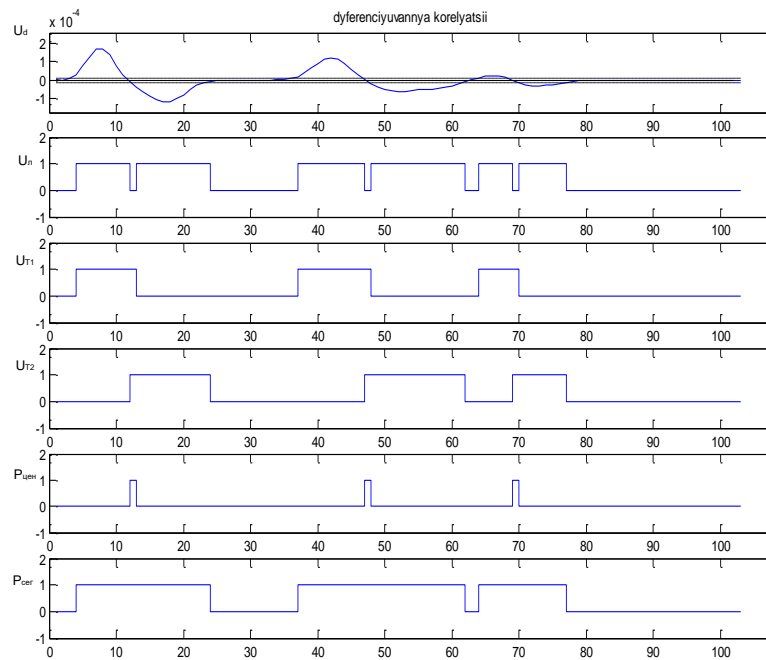
ЕНЕРГІЇ ДЛЯ ЗАВАНТАЖЕНОГО ФАЙЛУ В ДВОХ ДІАПАЗОНАХ



Результат виділення сегментів складів в акустичному сигналі на прикладі слова "чотири"



4
Формування логічних ознак складів



ПРОЕКТ РЕАЛІЗАЦІЇ СИСТЕМИ В МОБІЛЬНОМУ ДОДАТКУ

