НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Факультет прикладної математики

Кафедра прикладної математики

«На правах рукопису»
УДК 004.093

«До захисту допущено»

Завідувач кафедри
_____ О. Р. Чертов
(підпис)

«___» _____2015 р.

Магістерська дисертація

на здобуття ступеня магістра

зі спеціальності 8.04030101 «Прикладна математика»

на тему: Удосконалена модель розпізнавання артефактів слабко структурованої української природної мови

Виконала: студентка 2 курсу, групи КМ-31М

Гузієнко Ірина Віталіївна _____ (підпис)

Науковий керівник	доцент, канд. техн. наук, ст. наук. співроб. Маслянко П. П.	_____ (підпис)
Консультант із нормоконтролю	старший викладач Мальчиков В. В.	_____ (підпис)
Консультант зі спеціальних питань	ст. наук. співроб., канд. техн. наук, ст. наук. співроб. Пилипенко В. В.	_____ (підпис)
Рецензент	професор, д-р техн. наук, проф. Панкратова Н. Д.	_____ (підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.
Студентка _____
(підпис)

Київ – 2015 року

Національний технічний університет України
«Київський політехнічний інститут»

Факультет прикладної математики

Кафедра прикладної математики

Рівень вищої освіти – другий (магістерський)

Спеціальність 8.04030101 «Прикладна математика»

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ О. Р. Чертов
(підпис)

«___» _____2015 р.

ЗАВДАННЯ
на магістерську дисертацію студентці
Гузієнко Ірині Віталіївні

1. Тема дисертації: Удосконалена модель розпізнавання артефактів слабко структурованої української природної мови,
науковий керівник дисертації Маслянко Павло Павлович, канд. техн. наук, ст. наук. співроб.,
затверджені наказом по університету від «20» березня 2015 року № 785-С.

2. Термін подання студентом дисертації: «18» червня 2015 р.

3. Об'єкт дослідження: особливості автоматизації стенографування, парадигми, моделі, методи і способи моделювання слів, алгоритми розпізнавання мовлення, методи і процеси розпізнавання української мови, засоби розпізнавання мовлення, програмні продукти та системи автоматизації стенографування.

4. Предмет дослідження: вдосконалена модель розпізнавання артефактів слабко структурованої української природної мови з урахуванням обмежень на основі застосування прихованих марковських моделей та гаусівських сумішей моделей.

5. Перелік завдань, які потрібно розробити:

– проаналізувати стан проблеми розпізнавання артефактів природної мови в Україні та за кордоном;

– проаналізувати існуючі рішення для задачі розпізнавання артефактів природної мови;

– розробити модель системи розпізнавання артефактів слабко структурованої української природної мови на базі структурного і динамічного представлення;

– розробити математичні моделі для розпізнавання артефактів української мови, що не ввійшли у словник системи розпізнавання;

– проаналізувати результати розпізнавання з використанням запропонованих моделей.

6. Орієнтовний перелік ілюстративного матеріалу:

– схема класифікації методів, моделей і систем розпізнавання мовлення;

– порівняльна таблиця існуючих засобів розпізнавання;

– діаграма компонентів системи розпізнавання артефактів слабко структурованої природної мови;

– діаграми діяльності системи розпізнавання артефактів слабко структурованої природної мови (процес навчання та процес розпізнавання);

– діаграми діяльності системи розпізнавання артефактів слабко структурованої природної мови у вигляді «плавальних доріжок» (процес навчання та процес розпізнавання);

– результати проведених експериментів.

7. Орієнтовний перелік публікацій:

– наукова конференція «Оброблення сигналів і зображень та розпізнавання образів – UkrObraz'2014»;

– VI наукова конференція магістрантів та аспірантів «Прикладна математика та комп'ютинг – ПМК'2015»;

8. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Вдосконалена модель розпізнавання артефактів природної мови	Пилипенко В. В., ст. наук. співроб.		

9. Дата видачі завдання «25» жовтня 2013 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Вибір напряму дослідження та узгодження тематики МД з керівником	15 вересня–30 жовтня 2013	
2	Ґрунтовне ознайомлення з предметною областю	30 жовтня 2013–15 лютого 2014	
3	Вивчення літератури, пошук додаткової інформації	15 лютого–1 вересня 2014	
4	Проведення дослідження, розроблення програмного забезпечення	1 вересня 2014–1 березня 2015	
5	Завершення роботи над основною частиною МД, переддипломна практика, робота над публікаціями	1 березня–1 травня 2015	
6	Оформлення текстової і графічної частини МД	1 травня–1 червня 2015	
7	Попередній захист МД	1 червня–15 червня 2015	

Студентка _____ І. В. Гузієнко
(підпис)

Науковий керівник дисертації _____ П. П. Маслянко
(підпис)

РЕФЕРАТ

Актуальність теми. Сучасні системи диктування тексту базуються на певній кількості слів у словнику, а при вирішенні реальних задач стенографування виникає задача розпізнавання слів, які зустрічаються в українській мові і які не входять у словник системи розпізнавання.

Тому, актуальною є розробка моделі розпізнавання артефактів, які відсутні у словнику системи розпізнавання з метою забезпечення точності та адекватності стенографування у відповідності до задокументованого голосового запису.

Об'єктом дослідження є особливості автоматизації стенографування, парадигми, моделі, методи і способи моделювання слів, алгоритми розпізнавання мовлення, методи і процеси розпізнавання української мови, засоби розпізнавання мовлення, програмні продукти та системи автоматизації стенографування.

Предметом дослідження є вдосконалена модель розпізнавання артефактів слабко структурованого українського мовлення з урахуванням обмежень на основі застосування прихованих марковських моделей та гаусівських сумішей моделей.

Мета роботи: автоматизація процесу стенографування для мінімізації часу документування тексту на паперових носіях, забезпечення точності і адекватності стенографування у відповідності до голосового запису, виключення помилок стенографування.

Методи дослідження. В роботі використовуються методи розпізнавання образів, методи математичного моделювання, статистичні методи, теорія машинного навчання.

Наукова новизна роботи полягає в моделюванні невідомих артефактів наступними способами:

— моделювання OOV слів у вигляді ГСМ з одним станом;

— моделювання OOV слів декількома станами ГСМ;

— моделювання груп невідомих слів за їх довжиною.

Практична цінність отриманих в роботі результатів полягає в тому, що запропонована модель:

1) дозволяє мінімізувати час документування тексту на паперових носіях;
2) підвищує точність та забезпечує адекватність стенографування у відповідності до голосового запису;
3) дозволяє виключити помилки стенографування.

Апробація роботи. Основні положення та ціль роботи були представлені та опубліковані на VII науковій конференції магістрантів та аспірантів «Прикладна математика та комп'ютинг – ПМК'2015» (Київ, 15–17 квітня 2015 р.) та на науковій конференції «Оброблення сигналів і зображень та розпізнавання образів – UkrObraz'2014» (Київ, 3–7 листопада 2014 р.), опубліковані в збірниках цих двох конференцій.

Структура та обсяг роботи. Магістерська дисертація складається з вступу, чотирьох розділів, висновків та додатків.

У вступі надано загальну характеристику роботи, розкрито сучасний стан наукової проблеми та її значущість, сформульовано мету і задачі досліджень, показано наукову новизну отриманих результатів і практичну цінність роботи.

У першому розділі розглянуто класифікацію методів, моделей та систем розпізнавання, проаналізовано методи та алгоритми розпізнавання мовлення, розглянуто та проаналізовано існуючі засоби розпізнавання артефактів природного мовлення, наведено таблицю основних характернистик існуючих засобів для вирішення задачі перетворення мовлення в текст.

У другому розділі наведено структурне та динамічне представлення системи розпізнавання артефактів природної мови, описано звязки між

компонентами системи, проаналізовано існуючу модель, показано відмінності вдосконаленої моделі від існуючої.

У третьому розділі продемонстровано застосування ПММ для розпізнавання мовлення та особливості їх використання для невідомих артефактів, описано використання ГСМ для моделювання невідомих артефактів.

У четвертому розділі представлено архітектуру програмного забезпечення системи розпізнавання та описано реалізацію окремих компонентів системи, проведено тестування розроблених моделей та представлено результати досліджень, надано інструкцію користувача та скріншоти інтерфейса користувача.

У висновках проаналізовано отримані результати досліджень та пропонується сфера застосування діючої системи.

У додатках наведено схему класифікації методів, моделей та систем розпізнавання мовлення, фрагменти коду програмного забезпечення та слайди презентації для супроводу доповіді.

Робота виконана на 101 аркуші, містить 3 додатки та посилання на список використаних літературних джерел з 41 найменування. У роботі наведено 14 рисунків та 1 таблиці.

Ключові слова: розпізнавання артефактів природної мови, математична модель, приховані марковські моделі, гаусівська суміш моделей, акустична модель, лінгвістична модель, корпус мовлення, корпус текстів.

SUMMARY

Scientific relevance. Modern dictation systems are based on a number of words contained in a dictionary. While solving real problems in transcription, the problem of recognition of words which occur in the Ukrainian language and are outside of the recognition system dictionary arises.

Therefore, design of the recognition model of language artifacts, which are absent from the recognition system dictionary aiming at accurate and adequate transcription according to a voice record, is relevant.

The object of research is stenography automation features, paradigms, models, methods and techniques of modeling words, speech recognition algorithms, methods and processes of recognition of the Ukrainian language, speech recognition devices, software and automation transcription systems.

The subject of research is an improved recognition model of the artifacts of poorly structured Ukrainian speech allowing for limitations based on the use of hidden Markov models and Gaussian mixture of models.

Objective of the research is automation of the transcription with the purpose of minimizing the time of text recording on paper, accuracy and adequacy in transcription according to a voice record, as well as elimination of transcription errors.

Research methods. Methods of pattern recognition, mathematical modeling methods, statistical methods, as well as machine learning theory are being used.

Scientific novelty is provided by modelling of the unknown artifacts in the following ways:

- simulating OOV words in the form of GMM with one condition;
- simulating OOV words in several states of GMM;
- simulating a group of unknown words according to their length.

Practical value of obtained results lays in the facts that the suggested model:

1) allows to minimize the time of text recording on paper;

2) increases accuracy and ensures adequacy of recording in accordance with a voice record;

3) allows to eliminate transcription errors.

Approvement. The main provisions and the purpose of research were presented and published at the VI Scientific Graduate and Postgraduate Student Conference "Applied mathematics and computing – PMK'2015" (Kyiv, April 15-17, 2015) and at the Scientific Conference "Signal/Image Processing and Pattern Recognition - UkrObraz'2014" (Kyiv, November 3-7, 2014), published in both conference information packages.

Structure and scope. Master's thesis consists of an introduction, four chapters, conclusion and appendices.

In the introduction, general characteristics are provided, current state of the scientific problem and its significance are revealed, the goal and objectives of the research are articulated, scientific novelty of the results and practical value are displayed.

In the first chapter, the classification of methods, models and systems of recognition are considered, the methods and algorithms for speech recognition are analyzed, existing means of natural speech recognition artifacts are viewed and analyzed, the table of main characteristics of existing tools for converting speech to text is provided.

In the second chapter, structural and dynamic representation of the system of natural speech artifacts recognition is provided, relations between system components are displayed, the existing model is analyzed, differences between the improved and existing models are shown.

The third chapter demonstrates the use of HMM for speech recognition, as well as features of applying them to unknown artifacts, and describes the use of GMM for simulation of unknown artifacts.

In the fourth chapter, software architecture of the recognition system is presented, implementation of certain system components is described, testing of designed models is conducted, results of the research are presented, user manual and screenshots of the software interface are provided.

In the conclusion, findings of the research are analyzed, as well as suggested sphere of usage for existing system is given.

Appendices present a classification scheme of methods, models and speech recognition systems, software code snippets and presentation slides to accompany the report.

The thesis is presented on 101 pages, contains 3 appendices and the list of used literature sources of 41 items. 14 figures and 1 table are provided by thesis.

Keywords: natural speech artifacts recognition, mathematical model, hidden Markov models, Gaussian mixture of models, acoustic model, language model, language body, text body.

ЗМІСТ

СПИСОК ТЕРМІНІВ, СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ

Артефакти – це структурні компоненти мови такі, як фонеми, звуки, букви, символи, слова, словосполучення, речення, розділові знаки і т.д. [1].

АМ – акустична модель.

ЕОМ – електронно-обчислювальна машина.

Невідомі слова – див. визн. *OOV* (невідомі у тому розумінні, що система розпізнавання не містить цих слів у словнику).

ЛМ – лінгвістична модель.

МННЦ ІТіС – Міжнародний науково-навчальний центр інформаційних технологій і систем.

ПММ – приховані марковські моделі.

ШНМ – штучні нейронні мережі.

Gaussian Mixture Model (GMM) – гаусівська суміш моделей (ГСМ).

HTK(Hidden markov model ToolKit) – інструментарій для побудови прихованих марковських моделей.

Out-of-Vocabulary (OOV) words – слова, які не входять у словник системи розпізнавання.

UML – уніфікована мова моделювання.

ВСТУП

Мовленнєві технології продовжують розвиватись і знаходять своє застосування в різних областях. Так, завдяки цим технологіям з'явилась можливість керувати комп'ютером за допомогою голосу, диктувати тексти, спілкуватись з комп'ютером на інтелектуальному рівні.

Основними напрямками досліджень в області мовленнєвих технологій є: розпізнавання та синтез мовлення, керування голосом та ідентифікація за зразком мови. Усі існуючі системи розпізнавання мовлення за призначенням розрізняються на системи диктування тексту, голосові інтерфейси, системи розшифрування записів, які попередньо збережені на цифрових носіях, тощо.

Перші програми, що забезпечують голосове введення даних, були розроблені за кордоном раніше вітчизняних. Слід зазначити, що розробка систем розпізнавання української мови є досить специфічним завданням. При розпізнаванні мовлення, сказаного українською мовою, виникає цілий ряд труднощів. Основними відмінностями української мови від, наприклад, англійської є наявність великої кількості словоформ для багатьох слів та вільний порядок слів у реченні. Велика кількість словоформ збільшує розмір словника, а довільний порядок слів збільшує кількість можливих речень. Крім того, найчастіше словоформи одного й того ж слова відрізняються лише закінченнями, які зазвичай вимовляються не так чітко як початок слова. Такі особливості української мови дещо ускладнюють задачу розпізнавання, тому й виникають помилки в розпізнаванні. Точність розпізнавання залежить ще і від індивідуальних характеристик того, хто говорить: специфіка вимови, акценти, наголоси, хезитації (мовленнєві коливання, пов'язані зі спонтанністю мови: мовленнєві збої, заминки при мовленні, вагання у виборі слова або конструкції). Також, в системах

розпізнавання української мови потрібно враховувати діалекти, які зустрічаються на території України.

В Україні основною установою, що займається розпізнаванням мовлення, є МННЦ інформаційних технологій і систем НАН України та МОН України. Ще з кінця 1960-их років у відділі розпізнавання та синтезу звукових образів під керівництвом Вінцюка Тараса Климовича проводяться роботи по розпізнаванню мовлення [2]. На даному етапі розвитку мовленнєвих технологій ще є не вирішені задачі, які можна і потрібно вирішувати.

Так, програми, що працюють з ізольованими словами, досягли високої точності в командних системах – в найбільш поширених сучасних додатках точність розпізнавання складає в середньому 95-99% і залежить в основному від рівня шуму. У той же час задача розпізнавання злитої української мови в достатній мірі не вирішена. Однією з причин є те, що українська природна мова є слабко структурованою. Крім того, не аби яку роль відіграє вміст словника системи розпізнавання, який лише на якийсь відсоток співпадає зі словником мовленнєвого сигналу. Програма ж може розпізнавати ті слова, які містяться в словнику системи розпізнавання. Якщо диктувати слова, відсутні в словнику, то програма підбере найближче за характеристиками слово зі словника.

Зокрема, сучасні системи диктування тексту базуються на певній кількості слів у словнику, а при вирішенні реальних задач стенографування виникає задача розпізнавання слів, які зустрічаються в українській мові і які не входять у словник системи розпізнавання. Це означає, що результат стенографування не буде у повній мірі відповідати змісту мовленнєвого сигналу. Навіть, якщо теоретично представити, що словник системи розпізнавання є абсолютно повним і містить всі необхідні слова, то при розпізнаванні слів слабко структурованої української природної мови традиційними методами [3-9], неминуче будуть виникати помилки.

Не зважаючи на значні результати при дослідженні та розробці систем диктування тексту [3-10], актуальною є розробка вдосконаленої моделі розпізнавання артефактів слабко структурованої української природної мови, які не ввійшли в словник системи розпізнавання.

МЕТА ДОСЛІДЖЕННЯ ТА ПОСТАНОВКА ЗАДАЧІ

Об'єкт досліджень – особливості автоматизації стенографування, парадигми, моделі, методи і способи моделювання слів, алгоритми розпізнавання мовлення, методи і процеси розпізнавання української мови, засоби розпізнавання мовлення, програмні продукти та системи автоматизації стенографування.

Предмет досліджень – вдосконалена модель розпізнавання артефактів слабко структурованої української природної мови з урахуванням обмежень на основі застосування ПММ та ГСМ.

Метою досліджень є автоматизація процесу стенографування для мінімізації часу документування тексту на паперових носіях, забезпечення точності і адекватності стенографування у відповідності до голосового запису, виключення помилок стенографування.

Науково-практична задача, що розв'язується в даній магістерській дисертації, включає наступні завдання:

1) аналіз стану проблеми розпізнавання артефактів природної мови в Україні та за кордоном;
2) аналіз існуючих рішень для задачі розпізнавання артефактів природної мови;
3) розробка моделі системи розпізнавання артефактів слабко структурованої природної мови на базі структурного і динамічного представлення;
4) розробка математичних моделей для розпізнавання артефактів української природної мови, що не ввійшли у словник системи розпізнавання;
5) отримання та аналіз результатів розпізнавання з використанням запропонованих моделей.

Створений програмний модуль повинен інтегруватись з існуючою в МННЦ ІТіС системою (яка базується на основі пакету програм НТК), здійснювати обробку вхідного сигналу та ідентифікувати в ньому артефакти, які не входять у словник системи розпізнавання.

1 ОГЛЯД ІСНУЮЧИХ РІШЕНЬ ДЛЯ ЗАДАЧІ РОЗПІЗНАВАННЯ АРТЕФАКТІВ ПРИРОДНОЇ МОВИ

1.1 Класифікація методів, моделей і систем розпізнавання мовлення

Кожна система розпізнавання має деякі завдання, які вона покликана вирішувати і комплекс підходів, які застосовуються для вирішення поставлених завдань. Далі представлені основні ознаки, за якими можна класифікувати всі системи розпізнавання природної мови та описано вплив цих ознак на роботу системи.

Однією з них є розмір словника. Словник голосового повідомлення (мовленнєвого сигналу) на якийсь відсоток співпадає зі словником системи розпізнавання. Стверджують, що чим більше розмір словника, який закладений в систему розпізнавання, тим більше частота помилок при розпізнаванні слів системою. Наприклад, словник з 10 цифр може бути розпізнаний практично безпомилково, тоді як частота помилок при розпізнаванні словника в 100 000 слів може досягати 45% [11]. З іншого боку, більший словник охоплює більше слів, які диктор може подати на систему розпізнавання, а розпізнавання невеликого словника може дати велику кількість помилок розпізнавання, якщо в цьому словнику багато подібних між собою слів.

Наступна ознака – в залежності від типу мовлення, – роздільна або злита мова. Якщо при мовленні кожне слово відділяється від іншого ділянкою тиші, то кажуть, що ця мова – роздільна. Злита мова – це природно виголошені речення. Розпізнавання злитої мови набагато важче у зв'язку з тим, що межі окремих слів не чітко визначені і їх вимова іноді сильно спотворена змазуванням вимовлених звуків [11].

Ще одна ознака – дикторозалежність або дикторонезалежність системи. За визначенням, дикторозалежна система призначена для

використання одним користувачем (людиною, яка навчала цю систему), в той час як дикторонезалежна система призначена для роботи з будь-яким диктором [11]. Дикторозалежні системи налаштовуються на параметри того диктора, на прикладі якого навчаються (тренуються). Налаштування на голос диктора дикторозалежних систем займає звичайно від 30 хвилин до декількох годин. Системи розпізнавання мови, яким властива відносна незалежність від диктора, дозволяють користувачу працювати без попереднього налаштування. Крім того, існують системи з адаптацією до голосу диктора. Надійність таких систем розпізнавання поліпшується після навчання. Незалежність від диктора в таких системах зазвичай досягається за рахунок збереження звукових еталонів для усіх найбільш типових голосів носіїв мови. Це, безумовно, вимагає в кілька разів більшої продуктивності й обсягу пам'яті.

За призначенням системи розпізнавання мови поділяються на системи диктування тексту, голосові інтерфейси (командні системи), системи розшифрування записів, які попередньо збережені на цифрових носіях, тощо [12]. Призначення системи визначає необхідний рівень абстракції, на якому відбуватиметься розпізнавання вимовленого людиною тексту. У командній системі (наприклад, голосовий набір в стільниковому телефоні) частіше за все, розпізнавання слова чи фрази відбувається як розпізнавання єдиного мовного елемента. А системи диктування тексту потребують більшої точності розпізнавання і тому при інтерпретації вимовленої фрази буде враховуватися і аналізуватися не тільки те, що було сказано в поточний момент, а й те, як воно співвідноситься з тим, що було вимовлено до цього. Також, в системі повинен бути вбудований набір граматичних правил, яким повинен задовольняти вимовлений і розпізнаний текст. Чим суворіші ці правила, тим простіше реалізувати систему розпізнавання і тим обмеженішим буде набір висловлювань, які вона зможе розпізнати [13].

Крім того, основні відмінності в структурі і процесі роботи різних систем розпізнавання мови можна класифікувати за наступними ознаками:

– за типом структурної одиниці. При аналізі мовлення, в якості базової одиниці можуть бути обрані окремі слова, фрази або частини вимовлених слів, такі як фонеми, дифони або трифони, алофони. В залежності від того, яка структурна одиниця обрана, змінюється структура, універсальність і складність словника системи розпізнавання [11, 13].

– по принципу виділення ознак. Сама послідовність відліків тиску звукової хвилі – є надлишковою для систем розпізнавання мовлення і містить багато зайвої інформації, яка при розпізнаванні не потрібна, або навіть може зашкодити. Таким чином, для представлення мовного сигналу з нього потрібно виділити деякі параметри (ознаки), які дозволяють адекватно представити цей сигнал для розпізнавання [13]. Найпоширеніший підхід заснований на перетворенні Фур'є, яке переводить вихідний сигнал з амплітудно-часового простору в частотно-часовий, а в часовій області – лінійне передбачення, яке описує мовний сигнал з допомогою моделі авторегресії. Інший підхід – вейвлет-аналіз, який розкладає вхідний сигнал в базис функцій, що характеризують як частоту, так і час. Таким чином, можна аналізувати властивості сигналу одночасно і у фізичному просторі, і в частотному. Окрім вейвлет- і Фур'є-аналізу в системах розпізнавання мови використовується кепстральний аналіз, тобто зворотне перетворення Фур'є від логарифма прямого перетворення [11].

– за механізмом функціонування. У сучасних системах широко використовуються різні підходи до механізму функціонування систем розпізнавання. Ймовірнісно-мережевий підхід полягає в тому, що мовний сигнал розбивається на певні частини – кадри, або за фонетичною ознакою, після чого відбувається ймовірнісна оцінка того, до якого саме елементу словника системи розпізнавання має відношення дана частина і (або) весь вхідний сигнал. Підхід, заснований на вирішенні оберненої задачі синтезу звуку, полягає в тому, що по вхідному сигналу визначається характер руху артикуляторів мовного тракту і, за спеціальним словником відбувається визначення вимовлених фонем [13]. Найпростіші (кореляційні) детектори

зручні для використання в системах з обмеженим словником (це можуть бути командні системи). Експертні системи з різними способами формування та обробки бази знань дозволяють оброблювати мовленнєвий сигнал з необмеженою тривалістю та без додаткової інформації про межі слів і розробляються для дослідження та оптимізації процесів розпізнавання мови і навчання [14].

– за класом методів розпізнавання. Найбільшого поширення набули системи розпізнавання мови на базі прихованих марковських моделей (ПММ). Окрім ПММ, в системах розпізнавання використовуються динамічне програмування та нейронні мережі. На базі нейронних мереж можна створювати системи розпізнавання мови, які попередньо потрібно навчати та які самонавчаються [11].

Узагальнюючи все вищесказане, можна наглядно представити класифікацію систем розпізнавання мовлення (додаток А).

1.2 Методи та алгоритми розпізнавання мовлення

1.2.1 Динамічне програмування

Загалом процес розпізнавання слів за методом динамічного програмування реалізовано наступним способом [15]: у процесі навчання користувач, на голос якого налаштовано систему розпізнавання мовлення, вимовляє роздільно всі слова з необхідного йому набору. Вимова повинна бути природною, виразною, у властивій диктору манері, так званим повним стилем. У пам'ять ЕОМ записують розрахунки параметричного опису кожного слова через 10 – 20 мс, які називають еталонами. Залежно від технічних можливостей комп'ютера можуть бути різні варіанти запису навчальної вибірки: одна реалізація на слово, кілька реалізацій на слово або

їхнє усереднене значення. Найефективнішим є запис не всіх розрахунків сигналу, а найбільш показових, що добре відображає відносно стаціонарні сегменти мовлення. У цьому випадку запам'ятовується й часова структура слова-еталона. Розмножуючи в допустимих межах кількість еталонних розрахунків на окремих сегментах слова, можна одержати різні варіанти запису того самого слова, що відрізняються часовою структурою.

У режимі розпізнавання проаналізований образ порівнюють за деякими спрощеними критеріями з усіма еталонами. Серед групи еталонів, схожих на аналізований сигнал, методом динамічного програмування (багатокроковим алгоритмом, який з усіх можливих варіантів вибирає найоптимальніший) визначають потрібний варіант еталонного сигналу.

В [16] показано, що такий підхід заснований на економному заданні множини еталонних сигналів за допомогою автоматних породжучих граматик, які складають (синтезують) еталонні сигнали з елементарних частин, що представляють собою фонеми або їх фази і називаються еталонними елементами. Порівняння спостережуваного мовного сигналу з еталонними і пошук (разом з розбором-аналізом) найбільш правдоподібного еталонного сигналу здійснюється з допомогою процедур динамічного програмування.

Розпізнавання мовленнєвих сигналів при такому підході реалізується шляхом направленого синтезу еталонних сигналів мовлення. Після визначення ймовірної послідовності еталонних елементів у вхідному сигналі необхідно відновити по ній невідому послідовність фонем, яка є найбільш правдоподібною по відношенню до мовного сигналу, що розпізнається.

В формулюваннях методів розпізнавання та смислової інтерпретації в рамках підходу з використанням динамічного програмування чільна роль відводиться засобам економного задання множин еталонних сигналів та встановленню зв'язків між еталонними і спостережуваними сигналами. Ці засоби та зв'язки встановлюються математичною моделлю мовленнєвих

сигналів, в якій, перш за все повинні бути відображені детерміновані закономірності, якими зв'язані спостережувані сигнали.

На рис. 1.1 подано загальне представлення процесу розпізнавання мовлення за допомогою динамічного програмування.

Рисунок 1.1 – Розпізнавання мовлення з допомогою динамічного програмування

Сегментація неперервного мовного сигналу і дискретне розпізнавання при такому підході здійснюється в єдиному взаємопов'язаному процесі, в якому з еталонних сигналів фонем та слів зіставляється найбільш правдоподібний еталонний сигнал злитого мовлення і рішення про послідовність дискретних елементів, що передаються неперервним мовним сигналом, здійснюється на основі аналізу цього найбільш правдоподібного еталонного сигналу з одночасним зазначенням, якщо це необхідно, і границь дискретних елементів в неперервному мовному сигналі [16].

Включення сегментації в єдиний з розпізнаванням процес є перевагою такого підходу. Однак, динамічний алгоритм програмування має поліноміальну складність, тому коли ми маємо справу з більшими послідовностями, виникають дві незручності: запам'ятовування більших числових матриць та виконання великої кількості розрахунків відхилень.

Повертаючись до системи розпізнавання мовлення з настроюванням на голос диктора відзначимо, що як показано в [15] такі системи автоматично адаптуються до різних типів і характеристик параметричного опису, до мовлення користувача, легко перебудовуються на нові словники – все це розцінюють як їхню позитивну властивість. Однак адаптивні

системи мають і багато недоліків. Так, надійність розпізнавання може значно зменшитися, якщо в режимі експлуатації зміниться голос диктора (через застуду, зміну емоційного стану), розташування мікрофона або акустика приміщення порівняно з режимом запису еталонів слів у процесі навчання. Досить втомлює сам процес навчання, коли диктору доводиться вимовляти всі слова з потрібного набору. Вважають, що 100 – 300 слів – це максимум, на що може погодитися користувач під час навчання і експлуатації подібної системи. Така кількість розпізнаваних слів сильно звужує можливості організації систем мовного діалогу людини з машиною. І нарешті, ще один недолік пов'язаний із необхідністю вимовляння слів повідомлення ізольовано, так, щоб між ними була розділова пауза. Така особливість спілкування з ЕОМ незручна для користувача.

Вирішення проблеми залежності від диктора може бути знайдене за рахунок статистичних алгоритмів, що ґрунтуються на обробці великої кількості звукових даних – записів голосів десятків і сотень дикторів. Зокрема, такий підхід використовують нейронні мережі й приховані марковські моделі [17].

1.2.2 Штучні нейронні мережі

Системи розпізнавання, в основі яких використовують апарат штучних нейронних мереж, мають високу швидкодію і не потребують додатковим витрат для інтерпретації і введення початкової інформації про реальні об'єкти, мають можливості до навчання і адаптації в умовах зміни навколишнього середовища. Для штучних нейронних мереж (ШНМ) характерна значна статистична потужність, оскільки вони дозволяють

автоматично налаштувати систему для ефективного розрізнення набору розпізнаних слів [15].

Це математичні моделі, а також їхні програмні або апаратні реалізації, побудовані за принципом організації й функціонування біологічних нейронних мереж – мереж нервових клітин живого організму. Це поняття виникло при вивченні процесів, що відбуваються в мозку при мисленні, і при спробі змоделювати ці процеси. Як і всяка модель, вони являються наближенням. ШНМ являють собою систему з'єднаних між собою простих процесорів (однотипних елементів). Такі процесори зазвичай досить прості, особливо порівняно з процесорами, що використовуються в персональних комп'ютерах. Вони імітують роботу біологічного нейрона, і зазвичай називаються штучними нейронами. Кожний процесор подібної мережі має справу тільки із сигналами, які він періодично одержує, і сигналами, які він періодично посилає іншим процесорам. Тобто, кожен з нейронів, в кожен момент часу знаходиться, як і біологічний нейрон, в деякому поточному стані. Він має групу односпрямованих вхідних зв'язків – синапсів, що йдуть від входу в мережу або від інших нейронів. Крім того він має один односпрямований вихідний зв'язок – аксон. З'єднані в досить велику мережу з керованою взаємодією, такі локально прості процесори разом здатні виконувати досить складні завдання [18].

Найпростішою моделлю нейронної мережі є одношаровий персептрон. Одношаровість означає, що вхідний сигнал входів $(x_1, x_2, \dots x_n)$ подається на одну групу нейронів, іменованих шаром нейронної мережі, а вихідні сигнали цих нейронів надходять відразу на вихід мережі. Для двохшарової мережі вихідні сигнали подавалися б не на вихід мережі, а на другу групу – шар нейронів, а звідти на вихід. Зрозуміло, що тришаровий нейрон має вже три групи – шари, N-шаровий – N груп – шарів і т.д. Здатність до розпізнавання у багатошарових мереж значно перевершує ту ж здібність у одношарового персептрона. Проте дещо ускладнюється процес навчання цієї мережі. Під навчанням мережі ми

розуміємо процес налаштування ваг синапсів, так щоб вихід мережі був очікуваним [18].

Нейронні мережі не програмуються у звичайному розумінні цього слова, вони навчаються. Алгоритми, що реалізують навчання мережі, у випадку коли вона є багатошаровою, отримали назву алгоритмів зворотного поширення. Технічно навчання полягає в знаходженні коефіцієнтів зв'язків між нейронами. У процесі навчання нейронна мережа здатна виявляти складні залежності між вхідними даними і вихідними, а також робити узагальнення. Що означає, що в разі успішного навчання мережа може повернути правильний результат на підставі даних, які були відсутні в навчальній вибірці.

Отож, використання ШНМ для розпізнавання мовлення передбачає насамперед навчання мережі на значній кількості звукових еталонів. Наприклад, мережа попередньо опрацьовує 100 варіантів вимови звука «а» ста різними дикторами, унаслідок чого встановлюють зв'язки між штучними нейронами, що зафіксують середньостатистичні акустичні характеристики (форматні частоти, силу тощо) цього звуку. Надалі така ШНМ може не тільки правильно визначати звук «а», вимовлений кожним зі ста дикторів, із якими вона навчалася, а й розпізнавати мовлення інших, не знайомих їй дикторів, однак ланцюжки слів, а також слова, вимовлені з різним темпом, нейронні мережі ідентифікують не так добре.

1.2.3 Приховані марковські моделі

Математичний апарат прихованих марковських моделей (ПММ) являє собою універсальний інструмент моделювання стохастичних процесів, для опису яких не існує точних математичних моделей, а їх властивості

змінюються з плином часу відповідно з деякими статистичними законами. Найбільш широке застосування ПММ знайшли при вирішенні таких завдань, як розпізнавання мови, аналізу послідовностей ДНК і ряду інших [19].

Структура ПММ являє собою умовно залежні змінні з випадковим значенням [20]. Випадкова змінна $x(t)$ представляє собою значення прихованої змінної в момент часу t. Випадкова змінна $y(t)$ – це значення спостережуваної змінної в момент часу t. Значення прихованої змінної $x(t)$ (в момент часу t) залежить лише від значення прихованої змінної $x(t-1)$ (тобто в момент часу $t-1$). Це називається властивістю Маркова. Хоча в той же час значення спостережуваної змінної $y(t)$ залежить лише від значення прихованої змінної $x(t)$ (обидві в момент часу t). Ймовірність спостерігати послідовність $Y = y(0), y(1), \dots, y(L-1)$ довжиною L дорівнює

$$P(Y) = \sum_X P(Y|X)\, P(X), \tag{1.1}$$

тут сума пробігає по всіх можливих послідовностях прихованих вузлів $X = x(0), x(1), \dots, x(L-1)$.

Робота з прихованими марковськими моделями здійснюється в два етапи:

- навчання – визначення параметрів моделі – алгоритм Баума-Уелча;
- визначення – яка ймовірність того, що спостережувана послідовність векторів була згенерована даною моделлю – алгоритм максимума правдоподібності (Вітербі).

Перевагою ПММ являється можливість обробки послідовностей і сигналів різної довжини, що складно зробити при роботі зі штучними нейронними мережами.

Приховані марковські моделі, на відміну від нейронних мереж, успішно моделюють послідовності з декількох слів і практично не залежать від темпу вимови. Ще один плюс марковських моделей – висока швидкодія. Крім того, вони дозволили вченим підійти до вирішення складнішого завдання – розпізнавання довільного злитого мовлення.

Відомо, що наша мова будується з обмеженого набору мінімальних звукових складників – фонем, а отже, кожне слово можна представити у вигляді послідовності декількох фонем. Таким чином, не потрібно зберігати записи кожного слова – досить створити значний корпус записів мовлення достатньої кількості дикторів, який би дозволив одержати статистично достовірний опис усіх звуків, що трапляються в мовленні.

Під час застосування методу ПММ невідомою (прихованою) постає вимовлена диктором послідовність звуків, представлених у вигляді марковського ланцюга зі скінченного набору фонем, а спостережуваним значенням є параметри звукового мовлення (середня частота звучання, сила звуку, форматні частоти, представлені в схематичному вигляді – на спектрограмі, сонограмі, тощо). Сам процес розпізнавання полягає в аналізі спостережуваних звукових параметрів, на основі якого роблять припущення, що за фонема може ховатися за таким набором параметрів. Нарешті, формується ланцюжок фонем із найвищим ступенем ймовірності – імовірне слово, який або перевіряють за словником, або, якщо словникового контролю немає, подають користувачеві як найбільш ймовірний результат розпізнавання.

1.3 Існуючі засоби розпізнавання

Системи диктування тексту є дуже привабливими на даному етапі розвитку суспільства в силу новизни наданих користувачу можливостей.

Такі системи дозволяють користувачам мовленнєвий сигнал, записаний у звуковому файлі, перетворювати в звичайний текст.

На сьогоднішній день на ринку можна знайти достатньо комерційних систем розпізнавання мови, від систем дискретного диктування тексту до систем, що здатні розпізнавати злите мовлення, наприклад:

- Dragon NaturallySpeaking, ViaVoice, Voice_PE (Voice Personal Edition) – для англійської мови;
- «Горинич» – для російської мови;
- Існуюча в МННЦ ІТіС система для розпізнавання української мови.

Крім того, відомим і популярним на сьогоднішній день є голосовий пошук Google, який має можливість розпізнавати запити в пошукову систему українською мовою і показує хороші результати по якості розпізнавання. Однак він не призначений для введення великих текстів, а обмежений розпізнаванням ізольованих слів або невеликого речення.

Характеристики існуючих програмних та апаратно-програмних засобів, які призначені для диктування тексту чи комп'ютерного документування усних виступів на конференціях, нарадах та інших подібних заходах представлений в таблиці 1.1.

Таблиця 1.1 – Порівняльна таблиця існуючих засобів розпізнавання

Назва фірми / Назва системи	Для кого призначена	Метод	Словник	Точність розпізнавання	Вартість	Особливості (переваги та недоліки)
Scansoft / Dragon NaturallySpeaking [21]	Диктування будь-яких текстів англійською, французькою, німецькою, іспанською, японською мовами	ПММ	Можна експортувати/ імпортувати будь-які списки слів до 230 000 слів.	94 – 98%	Від 150$ до 600$	Переваги: зручна, надійна система, має досить широкий функціонал. Недоліки: не зручне введення чисел.
IBM / IBM ViaVoice [22]	Диктування будь-яких текстів англійською, іспанською та французькою мовами	ПММ	64 000 слів	92 – 96%	Від 149$	Переваги: хороша якість розпізнавання. Можна диктувати в будь-який текстовий редактор. Недоліки: складності з деінсталяцією
IBM / VoiceType [23]	Диктування будь-яких текстів англійською, іспанською та французькою мовами	ПММ	25 000 слів	82 – 90%	7 днів безкоштовна версія, далі 45$ реєстрація	Переваги: хороше розпізнавання простих слів. Недоліки: низька точність розпізнавання власних назв і скорочених слів

Продовження таблиці 1.1

Назва фірми / Назва системи	Для кого призначена	Метод	Словник	Точність розпізнавання	Вартість	Особливості (переваги та недоліки)
IBM / MedSpeak [24]	Для диктування звітів лікарів-радіологів	ПММ	25 000 слів	95 – 98 %	4495$	Переваги: сама висока безпомилковість розпізнавання, зручність у використанні, дикторонезалежність. Недоліки: словник системи обмежений набором специфічних медичних термінів.
Kurzweil / Voice_PE (Personal Edition) [25]	Для диктування текстів англійською мовою.	ПММ, ШНМ	30 000 слів	90 - 97%	295$ (ціна, рекомендована виробником)	Переваги: дикторонезалежність, високий відсоток розпізнавання навіть без навчання системи, простота використання. Недоліки: –
Lernout & Hauspie / Voice Xpress Professional [26]	Для диктування текстів англійською мовою.	–	30 000 слів	80%	150 $	Переваги: висока якість розпізнавання чисел, зручність використання. Недоліки: дикторозалежність, нерівномірна якість розпізнавання.
Sakrament / "Сакрамент" [27]	Для диктування текстів білоруською, російською, українською мовами	ДП, ПММ	до 10 000 слів	95-98%	–	Переваги: дикторонезалежність, зручність використання. Недоліки: розпізнавання коротких фраз, додаткові словники створюються лише по замовленню.

Продовження таблиці 1.1

Назва фірми / Назва системи	Для кого призначена	Метод	Словник	Точність розпізнавання	Вартість	Особливості (переваги та недоліки)
VoiceLock / "Горинич" ПРОФ 3.0 [28]	Для диктування текстів російською мовою.	ПММ	10 000 слів з можливістю поповнення	75 - 85%	–	Переваги: можливість поповнювати словник. Недоліки: для прийнятної якості розпізнавання мови необхідно тривале навчання (начитування мовної бази). Говорити потрібно чітко, монотонно.
РеалСпікер Лаб / RealSpeaker [29]	Для диктування текстів англійською, іспанською, українською та іншими мовами.	–	–	Близько 70% для українських слів	Від 5$ до 50$ в залежності від терміну дії ліцензії	Переваги: можна вводити тексти в будь-який текстовий редактор. Недоліки: розпізнавання ізольованих слів з явно вираженими паузами між словами.
Існуюча в МННЦ ІТіС система	Для диктування будь-яких текстів, зокрема, текстів на суспільно-політичну тематику.	ПММ	100 000 слів	80 – 95 %	–	Переваги: дикторонезалежність, розпізнавання української мови.

Варто відзначити, що системи диктування тексту на Заході знайшли своє практичне застосування в медицині. Це в першу чергу пов'язано з тим, що наукові дослідження в цій галузі добре фінансуються. Крім того задача тут спрощується тим, що словники медичних термінів в вузькій предметній області мають менший об'єм.

Усі вищезгадані системи в принципі використовують одні і ті ж методи та алгоритми. Різниця в об'ємі словника, типі мовлення та інших характеристиках обумовлена лише специфікою конкретної задачі та обмеженнями на швидкість обчислень та об'єм необхідної пам'яті.

Більшість з цих систем зручні в експлуатації і мають досить непогані можливості, але не вміють працювати з українською мовою. А от розроблена в МННЦ ІТіС система розпізнаваня мовлення дозволяє користувачам мовленнєві сигнали записані українською мовою перетворювати в текст (створювати стенограми українською мовою). Точність розпізнавання цієї системи дещо відстає від провідних закордонних розробок, однак це пояснюється специфікою української мови та тим, що українська мова є слабко структурованою.

Висновки до розділу

У розділі здійснено огляд існуючих рішень для задачі розпізнавання артефактів природної мови. Для кращого розуміння області досліджень проведено класифікацію методів, моделей і систем розпізнавання за основними ознаками і встановлено вплив цих ознак на роботу системи. Таку класифікацію наведено в додатку А.

Досліджено методи розпізнавання мовлення, моделі та способи моделювання слів. Аналіз існуючих методів розпізнавання мовлення показав, що широке застосування отримали алгоритми, в яких для

моделювання застосовуються приховані марковські моделі (ПММ), оскільки використання такого методу дає кращі результати в рамках запланованих досліджень.

Досліджено існуючі на ринку системи диктування текстів в Україні та за кордоном. Аналіз основних характеристик цих систем показав, що в даний час не існує універсальної системи, яка була б здатна до самонавчання, була б дикторонезалежною, стійкою до шумів, розпізнавала б злите мовлення, була б здатна працювати зі словниками великих розмірів і при цьому мала б низьку частоту появи помилок. Крім того, було встановлено, що більшість існуючих систем не вміють працювати з українською мовою, оскільки вона має свої особливості та є слабко структурованою.

Існуючу в МННЦ ІТіС систему, яка працює з українською мовою, планується вдосконалити з метою підвищення точності та забезпечення адекватності стенографування у відповідності до голосового запису.

2 ВДОСКОНАЛЕНА МОДЕЛЬ РОЗПІЗНАВАННЯ АРТЕФАКТІВ ПРИРОДНОЇ МОВИ

2.1 Структурне представлення системи розпізнавання

Для існуючої системи розпізнавання було виділено основні компоненти та встановлено зв'язки між ними. Таку модель системи описано в нотації UML на рис. 2.1.

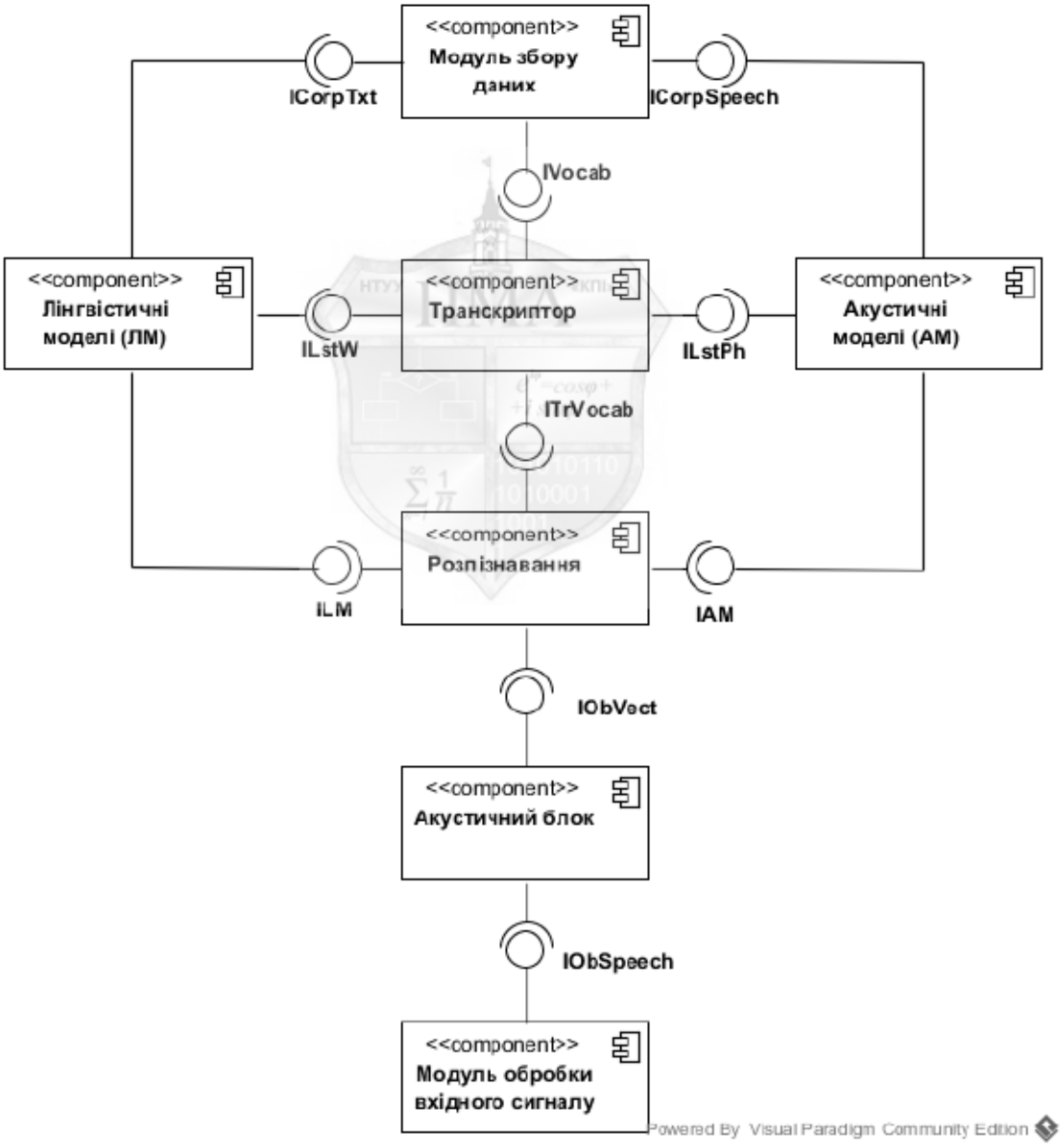

Рисунок 2.1 – Модель системи розпізнавання артефактів слабко структурованої природної мови. Діаграма компонентів в нотації UML.

Загалом процес розпізнавання проходить в два етапи: навчання і власне розпізнавання.

Модуль збору даних призначений для збору інформації, фонограм, стенограм та текстів необхідних для навчання системи.

Компонент «Транскриптор» переводить орфографічне представлення слів зі словника в фонетичну послідовність (послідовність фонем), таким чином створюючи словник транскрипцій.

Компоненти «Акустичні моделі» та «Лінгвістичні моделі» використовують дані, що містяться в модулі збору даних та словник транскрипцій. Таким чином, для навчання акустичних моделей використовується корпус мовлення та список фонем зі словника транскрипцій, а для навчання лінгвістичних моделей – корпус текстів та список слів зі словника транскрипцій.

Модуль обробки вхідного сигналу призначений для отримання вхідного сигналу та здійснення його попередньої обробки.

Компонент «Акустичний блок» відповідає за представлення мовленнєвого сигналу у вигляді послідовності векторів спостереження.

Компонент «Розпізнавання» приймає дані від акустичного блоку та аналізує їх на основі даних отриманих на етапі навчання від транскриптора, акустичних та лінгвістичних моделей. Цей компонент відповідає за перетворення вхідного мовленнєвого потоку в текст та подання результату розпізнавання користувачу.

2.2 Динамічне представлення системи розпізнавання

2.2.1 Діаграма діяльності. Процес навчання

Діаграма діяльності являє собою блок-схему, що показує потік управління від однієї діяльності до іншої [30].

Процес навчання системи розпізнавання злитої мови показано на рис. 2.2.

Рисунок 2.2 – Модель процесу навчання системи розпізнавання артефактів слабко структурованої природної мови. Діаграма діяльності в нотації UML.

Для навчання системи необхідно сформувати дані на яких будуть навчені основні компоненти системи. Так, перш за все, відбувається збір

інформації: наповнення корпусу текстів, наповнення корпусу мовлення, вибір словника.

Процедура підготовки корпусу мовлення полягає у встановленні точної відповідності між звуком фонограми і текстом стенограми. Кожен запис прослуховується спеціалістом і переводиться в текст – створюється експертна розмітка. Усі фрази, слова і абревіатури повинні бути прописані так, як їх вимовив диктор.

Процедура підготовки корпусу текстів полягає в наступному:

1) виключення з тексту службової інформації;
2) приведення тексту до канонічного вигляду;
3) формування у вигляді речень (явна вказівка початку і кінця).

Формування орфографічного словника полягає в наповненні словника словами, які будуть розпізнані системою. Для подальшого використання цей словник необхідно перетворити у словник транскрипцій, де кожне слово з орфографічного словника буде представлене у вигляді послідовності фонем.

Для навчання акустичної моделі необхідна оцінка ймовірності кожного стану на кожній ділянці сигналу. В процесі навчання відбувається оптимізація акустичних ознак та параметрів акустичної моделі. Дані для навчання беруться з підготованого корпусу мовлення. Під час навчання акустичних моделей відбувається, відповідно до обраного критерію оптимальності, налаштування (підбір) значень параметрів моделі за даними спостережень, в результаті чого створюється модель, яка найкращим чином відповідає реальному явищу. Тут кожна базова одиниця, що в подальшому буде розпізнана, представляється деяким типом прихованої марковської моделі, налаштування параметрів якої здійснюється по навчальній безлічі мовленнєвих даних.

Формування статистичної мовленнєвої моделі відбувається на основі аналізу особливостей української мови та збору статистики появи різних пар

фонем. Тут розраховуються ймовірності появи всіх пар фонем, які зустрічаються в українській розмовній мові.

В результаті навчання, на виході формуються навчені акустичні та лінгвістичні моделі і словник системи розпізнавання.

2.2.2 Діаграма діяльності. Процес розпізнавання

Попередня обробка вхідного сигналу починається з оцінки якості мовленнєвого сигналу, очистка його від шумів, а далі, виділяються фрагменти мовлення з вхідного сигналу. Результат поступає в акустичний блок, де відбувається розрахунок параметрів мовлення, які необхідні для розпізнавання. При цьому якість розпізнавання нерідко залежить саме від того, як вихідний сигнал буде підготовлений до виділення з нього інформативних ознак. Потім вектори ознак поступають в найважливіший компонент системи – «Розпізнавання».

Однією з важливих умов для розпізнавання мовлення є наявність порівняльних образів або еталонних записів голосу і мовлення. Такі еталони отримуються на етапі навчання і потім використовуються в процесі розпізнавання.

В компоненті «Розпізнавання» вхідний мовленнєвий потік зіставляється з інформацією, яка зберігається в акустичних і лінгвістичних моделях і далі визначається найбільш ймовірна послідовність слів, яка і є кінцевим результатом.

Процес розпізнавання злитого мовлення у вигляді діаграми діяльності представлений на рис. 2.3.

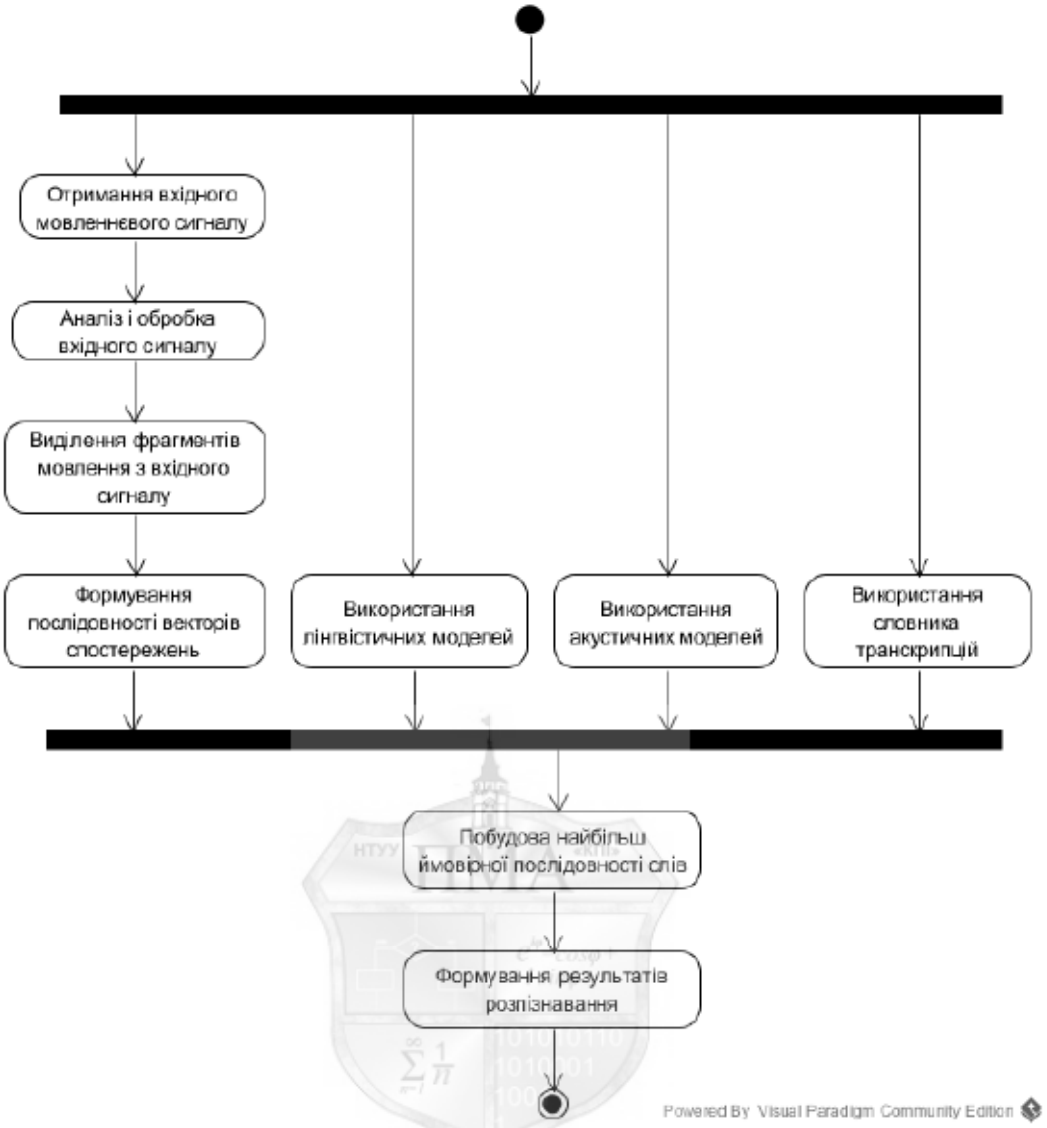

Рисунок 2.3 – Модель процесу розпізнавання артефактів слабко структурованої природної мови. Діаграма діяльності в нотації UML.

Збережені в пам'яті комп'ютера еталони вимови, що отримані на етапі навчання, по черзі порівнюються з поточною ділянкою послідовності десяти мілісекундних векторів, які описують вхідний мовленнєвий сигнал. В залежності від ступеня співпадіння вибирається найкращий варіант і формується гіпотеза про зміст висловлювання. На цьому етапі є така проблема, як необхідність нормалізації сигналу за часом.

Темп мовлення, тривалість вимови окремих слів і звуків навіть для одного диктора варіюється в дуже широких межах. Таким чином, можливі значні розбіжності між окремими ділянками наявного еталону і теоретично

співпадаючим з ним вхідним сигналом за рахунок їх неузгодженості по часу. Достатньо ефективно вирішувати цю проблему дозволяє алгоритм Вітербі та його різновиди.

Особливістю таких алгоритмів є можливість динамічного стиснення і розтягнення сигналу по часовій осі безпосередньо в процесі порівняння з еталоном. А застосування прихованих марковських моделей дозволяє на основі багаторівневого імовірнісного підходу до опису сигналу виробляти нормалізацію по часу та прогнозування продовжень, що прискорює процес перебору еталонів і підвищує надійність розпізнавання.

2.2.3 Діаграма діяльності у вигляді «плавальних доріжок». Процес навчання

Діаграма діяльності у вигляді «плавальних доріжок» дозволяє добре зобразити які процеси відбуваються в кожному компоненті системи. Для етапу навчання діаграма діяльності матиме наступний вигляд (рис. 2.4).

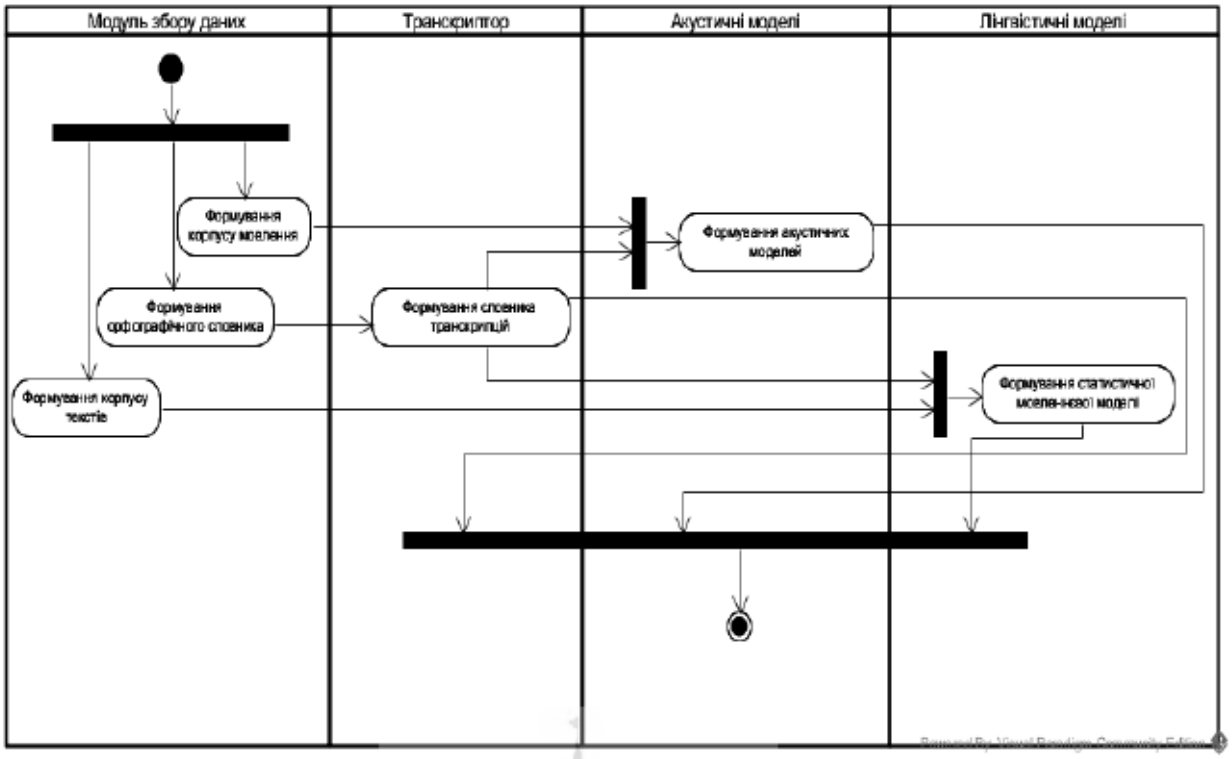

Рисунок 2.4 – Модель процесу навчання системи розпізнавання артефактів слабко структурованої природної мови. Діаграма діяльності у вигляді «плавальних доріжок» в нотації UML.

На етапі навчання акустичні моделі для кожного слова навчаються на наборах вхідних даних, що містяться в модулі збору даних. У компоненті «Акустичні моделі» містяться акустичні моделі для фонем мови і створюються акустичні зразки (еталони) кожного слова. Тут, перш за все, для кожного звука будується статистична модель, яка як найточніше описує вимову даного звука при мовленні.

Формування статистичної моделі мови відбувається шляхом збору статистики появи різних пар фонем, що присутні в навчальних даних отриманих з модуля збору даних. Тут, з урахуванням правил та особливостей української мови, розраховуються ймовірності появи всіх пар фонем.

2.2.4 Діаграма діяльності у вигляді «плавальних доріжок». Процес розпізнавання

Діаграма діяльності для етапу розпізнавання представлена на рис. 2.5.

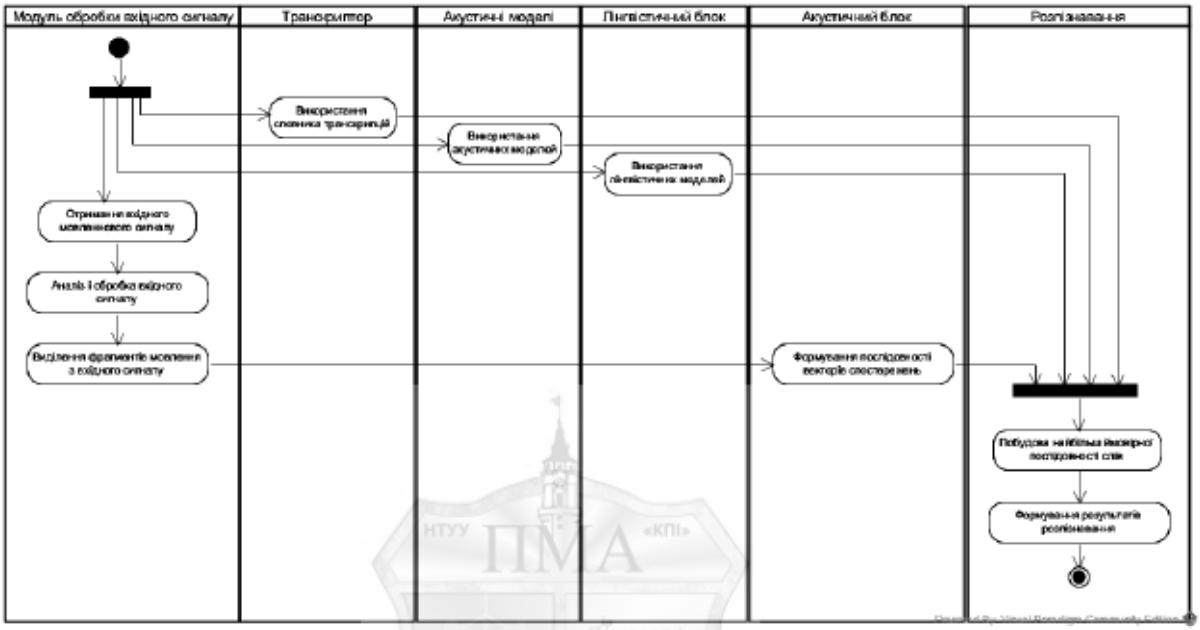

Рисунок 2.5 – Модель процесу розпізнавання артефактів слабко структурованої природної мови. Діаграма діяльності у вигляді «плавальних доріжок» в нотації UML.

В першу чергу звуковий потік потрапляє в модуль обробки вхідного сигналу. Тут здійснюється його попередня обробка, яка містить автоматичне регулювання посилення, придушення еха, виявлення наявності / відсутності мовлення та виявлення інтонаційного кінця фрази. Цей модуль включає також виділення фрагменту мови з вхідного мовленнєвого сигналу.

Існує декілька алгоритмів визначення початку і кінця промови. В одному з них визначається деякий пороговий рівень сигналу. Початкова точка промови в цьому випадку відповідає моменту, коли вхідний сигнал починає перевищувати граничний рівень, а кінцева точка – моменту, де амплітуда вхідного сигналу менше порогової. Основний недолік цього методу полягає в неможливості точного визначення мовного відрізка в разі сильного шуму, або, навпаки, тихої промови. Для уникнення цього недоліку

використовується граничний рівень, який обчислюється в залежності від рівня співвідношення «сигнал-завади».

Акустичний блок виконує частотний аналіз сигналу. Акустично-фонетичний потік даних розбивається на короткі кадри, або вектори, тривалістю близько 10 мс. Для кожного кадру визначається ряд параметрів, використовуючи mel-частотні кепстральні коефіцієнти з їх першою та другою похідною.

Компонент «Розпізнавання» здійснює акустичне порівняння: кожен кадр, або вектор, порівнюється з наявними акустично-фонетичними зразками, що зберігаються в спеціальній базі даних. При цьому порівнюватися можуть як окремі фонеми, так і слова, і навіть фрази. При невеликій кількості слів, що використовуються диктором, більш високу надійність і швидкість можна очікувати від розпізнавання цілих слів, але при збільшенні словника швидкість різко падає, і оптимальним стає розпізнавання окремих фонем. В цьому компоненті вирішується завдання динамічного програмування з метою знайти найкраще розбиття отриманого від лінгвістичних моделей потоку на слова і фрази. Залежно від обсягу використовуваного словника і діючих синтаксичних правил, застосовуються різні стратегії пошуку та відсіву. У даному блоці з розпізнаних фонем формуються слова, а зі слів фрази. При цьому використовується алгоритм променевого пошуку по Вітербі для визначення найбільш вірогідної фрази.

2.3 Відмінності вдосконаленої моделі від існуючої

В існуючій системі розпізнавання набір акустичних моделей будується лише для слів, які є в словнику системи розпізнавання. Однак, словник системи розпізнавання лише на якийсь відсоток співпадає зі словником мовленнєвого сигналу. Тому, слова, яких немає в словнику існуюча система розпізнає помилково, що в більшості випадків тягне за собою помилки в розпізнавані всього речення.

Відмінність вдоскналеної моделі полягає у додаванні до набору акустичних моделей існуючої ситеми таких моделей, що дозволяють ідентифікувати слова, які не ввійшли в словник системи розпізнавання (OOV слова). В залежності від розміру та вмісту словника кількість та різноманітність OOV слів може коливатись в достатньо широких межах. Моделі цих слів створюються таким чином, щоб уся різноманітність невідомих слів враховувалась.

Отож, модифікуючи компонент «Акустичні моделі» ствоюється вдосконалена модель розпізнавання артефактів слабко структурованої української природної мови.

2.4 Підготовка даних для навчання системи

Оскільки, аналіз існуючої системи розпізнавання показав, що необхідно розробити моделі слів, що не входять у словник системи розпізнавання, то для того, щоб ці моделі давали бажаний результат їх необхідно навчити.

Перший етап, необхідний для навчання моделей – підготовка даних. Мовленнєві дані необхідні як для навчання, так і для тестування. В існуючій

системі, всі ці мовленнєві дані записані в оперативній пам'яті. Для початку необхідно визначити фонемний набір, скласти словник так, щоб охопити етапи навчання і тестування, а також визначити граматику.

Перший крок при складанні словника – це створення сортованого списку необхідних слів. Цей список слів залежить від області в якій планується використовувати систему. В даному випадку словник містить слова на соціально-політичну тематику. Для створення стійких акустичних моделей необхідно при навчанні використовувати великий набір висловлювань, що містять багато слів, яких немає в словнику системи розпізнавання і бажано, щоб вони були фонетично збалансовані.

Є стандарт опису граматик розпізнавання мовлення. Цей стандарт визначає правила створення і синтаксис граматик, з якими працює система розпізнавання мовлення, тобто дозволяє вказувати, які саме слова мають бути розпізнані системою та визначати синтаксичні конструкції в рамках яких будуть розпізнаватися ці слова.

Для побудови тренувального файлу необхідно стенограми, які містяться в корпусі мовлення, розмітити таким чином, щоб явно позначити слова, яких немає в словнику системи. Цей тренувальний файл в подальшому використовується для навчання акустичних моделей.

Процес навчання ПММ здійснюється шляхом послідовного виконання наступних процедур:

— ініціалізація ПММ для OOV слів;

— навчання ПММ невідомих слів на навчальному корпусі мовлення (алгоритм Баума- Уелча);

— послідовне збільшення числа компонент гауссових сумішей моделей з одночасним навчанням на корпусі мовлення.

В результаті навчання формуються еталони кожного слова зі словника та еталони для OOV слів.

Висновки до розділу

Аналіз існуючої системи розпізнавання показав, що набір акустичних моделей будується лише для слів, які є в словнику системи розпізнавання. Однак, в мовленнєвому сигналі можуть зустрічатися слова, яких немає в словнику. В таких випадках точність розпізнавання падає, оскільки помилка в розпізнаванні одного слова може негативно вплинути на розпізнавання всього речення.

Пропонується вдосконалити існуючу систему за рахунок моделювання слів, які не входять у словник системи розпізнавання.

Модифікуючи компонент «Акустичні моделі» ствоюється вдосконалена модель розпізнавання артефактів слабко структурованої української природної мови.

3 МОДЕЛЮВАННЯ АРТЕФАКТІВ СЛАБКО СТРУКТУРОВАНОЇ УКРАЇНСЬКОЇ ПРИРОДНОЇ МОВИ, ЩО НЕ ВХОДЯТЬ У СЛОВНИК СИСТЕМИ РОЗПІЗНАВАННЯ

3.1 Застосування ПММ для розпізнавання

Прихованою марковською моделлю (ПММ) називається модель, яка складається з N станів, в кожному з яких деяка система може приймати одне з M значень якого-небудь параметра. Ймовірності переходів між станами задаються матрицею ймовірностей $A = \{a_{ij}\}$, де a_{ij} – ймовірність переходу з i-го в j-ий стан. Ймовірності появи кожного з M значень параметра в кожному з N станів задається вектором $B = \{b_j(k)\}$, де $b_j(k)$ – ймовірність появи k-го значення параметра в j-му стані. Ймовірність настання початкового стану задається вектором $\pi = \{\pi_i\}$, де π_i – ймовірність того, що в початковий момент система опиниться в i-му стані [31].

Таким чином, прихованою марковською моделлю називається трійка $\lambda = \{A, B, \pi\}$. Використання прихованих марковських моделей для розпізнавання мови базується на двох наближеннях:

1) мовленнєвий сигнал може бути розбитий на фрагменти, що відповідають станам в ПММ, параметри мовлення в межах кожного фрагмента вважаються постійними.

2) імовірність кожного фрагмента залежить тільки від поточного стану системи і не залежить від попередніх станів.

Модель називається «прихованою» оскільки нас, як правило, не цікавить конкретна послідовність станів, в якій перебуває система. Ми або подаємо на вхід системи послідовності типу $O = \{o_1, o_2, \ldots, o_\tau\}$, де кожне o_i – значення параметра (одне з M), прийняте в i-й момент часу, а на виході очікуємо модель $\lambda = \{A, B, \pi\}$, яка з максимальною ймовірністю генерує таку послідовність, або навпаки подаємо на вхід параметри моделі і

генеруємо породжену нею послідовність. І в тому і в іншому випадку система виступає як «чорний ящик», в якому сховані справжні стани системи, тому пов'язана з нею модель називається прихованою [19].

Щодо прихованих марковських моделей вирішуються, як правило, три задачі [32]:

1) дана послідовність спостережень $O = \{o_1, o_2, \ldots, o_\tau\}$ і модель $\lambda = \{A, B, \pi\}$. Необхідно обчислити ймовірність появи зазначеної послідовності для даної моделі. Тобто рішення цієї задачі безпосередньо пов'язано із завданням розпізнавання мовлення. Якщо, наприклад, стани моделі відповідають відрізкам часу, в які знімаються параметри мовного сигналу, і в кожному з цих станів (відрізків) параметри мовного сигналу приймають деякі значення, які ми представляємо у вигляді вектора спостережень $O = \{o_1, o_2, \ldots, o_\tau\}$, то вирішивши задачу знаходження ймовірності появи цієї послідовності для кожної з наявних у нас моделей $\lambda = \{A, B, \pi\}$, які відповідають, наприклад, фонемам (звукам мови) або словам, ми можемо вибрати ту з фонем чи те слово, яке найбільшою мірою відповідає вихідному відрізку мовного сигналу. А це і означає розпізнати мовну одиницю (фонему або слово).

2) дана послідовність спостережень $O = \{o_1, o_2, \ldots, o_\tau\}$ і модель $\lambda = \{A, B, \pi\}$. Необхідно вибрати послідовність станів $Q = \{q_1, q_2, \ldots, q_\tau\}$, яка з найбільшою ймовірністю породжує вказану послідовність спостережень. Дані, які є результатом рішення цієї задачі використовуються для вивчення поведінки отриманої моделі.

3) дана послідовність спостережень $O = \{o_1, o_2, \ldots, o_\tau\}$ і модель $\lambda = \{A, B, \pi\}$. Необхідно підібрати параметри моделі так, щоб максимізувати ймовірність даної послідовності спостережень. Це в чистому вигляді задача навчання моделі на наборах вхідних даних, для того щоб надалі використовувати цю модель для вирішення задачі 1, тобто розпізнавання. Знову ж таки, стани моделі відповідають відрізкам часу (як правило 10 мс), в яких знімаються значення параметрів мовного сигналу, а прийняті на

деякому часовому відрізку значення параметрів і утворюють послідовність спостережень O.

Для розв'язання задачі 1 використовують алгоритм прямого і зворотного ходу – це дві модифікації алгоритму обчислення ймовірностей, рівноцінні по обчислювальним витратам.

Розглянемо алгоритм прямого ходу, в якому водиться змінна $\alpha_t(i)$ – ймовірність того, що до моменту часу t система буде перебувати в i-му стані, а послідовність породжених нею до цього моменту спостережень дорівнює $o_1, o_2, \ldots, o_\tau$ [31].

Розглянемо даний алгоритм поетапно.

1) для всіх i від 1 до N

$$\alpha_0(i) = \pi_i b_i(o_1) \tag{3.1}$$

2) для всіх t від 1 до τ і для всіх j від 1 до N

$$\alpha_t(j) = b_j(o_t) \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} \tag{3.2}$$

3) за формулою 3.3 визначається $P(O|\lambda)$ – ймовірність того, що дана спостережувана послідовність побудована саме для даної моделі.

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_t(i) \tag{3.3}$$

Розглянемо алгоритм зворотного ходу, в якому водиться змінна $\beta_t(i)$ – ймовірність того, що до моменту часу t система буде перебувати в i-му стані, а послідовність породжених нею після цього спостережень дорівнює $o_{t+1}, o_{t+2}, \ldots, o_{\tau-1}, o_\tau$

Розглянемо даний алгоритм поетапно.

1) для всіх i від 1 до N

$$\beta_\tau(i) = 1 \tag{3.4}$$

2) для всіх t, що йдуть у зворотному порядку від τ-1 до 1 і для всіх i від 1 до N

$$\beta_\tau(i) = \sum_{j=1}^{N} a_{ij}\beta_{\tau+1}(j)b_j(o_{t+1}) \tag{3.5}$$

3) аналогічно п.3 алгоритму прямого ходу визначається $P(O|\lambda)$:

$$P(O|\lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1)\beta_1(i) \tag{3.6}$$

Для здійснення розпізнавання на основі прихованих моделей Маркова використовується безліч еталонних наборів для характерних ознак мови. Для цього еталонні мовні фрагменти записуються, розбиваються на елементарні складові (відрізки мовлення, на яких можна вважати параметри мовного сигналу постійними) і для кожного з них обчислюються значення характерних ознак. Одній елементарній складовій буде відповідати один набір ознак з безлічі наборів ознак словника [33].

Далі необхідно налаштувати модель розпізнавання. Одна прихована модель Маркова $\lambda = \{A, B, \pi\}$ ставиться у відповідність деякій одиниці мови, як правило, слову, що розпізнається.

Фрагмент мовлення розбивається на відрізки, протягом яких параметри мови можна вважати постійними. Для кожного відрізка обчислюються характерні ознаки і підбирається еталон з найбільш відповідними характеристиками. Кожному слову словника відповідає одна послідовність векторів спостереження $O = \{o_1, o_2, \ldots, o_\tau\}$. Матриця A – матриця ймовірностей переходів з одного мінімального відрізка мовлення в інший мінімальний відрізок мовлення. Елементи матриці B – ймовірності появи в кожному стані конкретного спостереження [34].

На етапі налаштування моделей Маркова застосовується алгоритм Баума-Уелча для наявного словника і кожному слову зі словника зпівставляються матриці A і B.

При розпізнаванні, мовлення розбивається на відрізки і застосовується алгоритм прямого або зворотного ходу для обчислення ймовірності відповідності даного звукового фрагмента певному слову зі словника. Якщо ймовірність перевищує деяке порогове значення – слово вважається розпізнаним.

Для розв'язання задачі 2 використовується, заснований на динамічному програмуванні, алгоритм Вітербі [35]. Суть алгоритму полягає в тому, щоб по заданій послідовності векторів параметрів визначити найбільш ймовірну послідовність станів (шлях Витербі), що породжуються розглянутою моделлю. Порівнюючи отримані оцінки ймовірностей для різних моделей, можна визначити модель, якій з найбільшою ймовірністю належить послідовність векторів параметрів, що розпізнається.

Алгоритм робить декілька припущень:

- спостережувані і приховані події повинні бути послідовністю;
- дві послідовності повинні бути вирівняні: кожна спостережувана подія має відповідати рівно одній прихованій події;

- обчислення найбільш вірогідної прихованої послідовності до моменту t повинно залежати тільки від спостережуваної події в момент часу t, і найбільш вірогідної послідовності до моментуt-1.

Отож, необхідно вибрати послідовність станів $Q = \{q_1, q_2, \ldots, q_\tau\}$, яка з найбільшою ймовірністю породжує вказану послідовність векторів ознак [36].

Вводяться змінні:

$$\delta_t(i) = maxP(q_t = S_i | q_1 q_2 \ldots q_{t-1}, o_1 o_2 \ldots o_t, \lambda) \quad (3.7)$$

тобто $\delta_t(i)$ – максимальна ймовірність того, що при заданих спостереженнях до моменту t послідовність станів завершиться в момент часу t в стані S_i, а також вводиться змінна $\psi_t(\mathrm{i})$ для зберігання аргументів, які максимізують $\delta_t(i)$.

1 крок алгоритму Вітербі. Для всіх i від 1 до N

$$\delta_1(i) = \pi_i b_i(o_1) \quad (3.8)$$

$$\psi_1(\mathrm{i}) = 0 \quad (3.9)$$

2 крок. Для всіх j від 1 до N і t від 2 до T

$$\delta_t(i) = \max\left|\delta_{t-1}(i)\mathrm{a_{ij}}\right|\mathrm{b_j}(\mathrm{o_t}) \quad (3.10)$$

$$\psi_t(\mathrm{j}) = \arg\max\left|\delta_{t-1}(i)\mathrm{a_{ij}}\right| \quad (3.11)$$

3 крок. Отримуємо найбільшу ймовірність спостереження послідовності $\mathrm{o_1}, \mathrm{o_2}, \ldots, \mathrm{o_T}$, яка досягається при проходженні деякої

оптимальної послідовності станів $Q^* = \{q_1^*, q_2^*, \ldots, q_T^*\}$, для якої на цей момент відомо тільки останній стан:

$$P^* = \max_{i=1\ldots N} |\delta_T(i)| \tag{3.12}$$

$$q_T^* = \arg\max_{i=1\ldots N} |\delta_T(i)| \tag{3.13}$$

4 крок. Відновлення оптимальної послідовності станів (зворотний прохід):

Для всіх t від T-1 до 1 (крок = -1)

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \tag{3.14}$$

В результаті отримується оптимальна послідовність станів.

Для розв'язання задачі 3 застосовується Алгоритм Баума-Уелча. Даний алгоритм (також називається, як ЕМ-метод – Expectation-Maximization) являє собою ітеративний алгоритм, який намагається підібрати параметри прихованої моделі Маркова так, щоб максимізувати ймовірність даної послідовності спостережень [37].

Кожна ітерація алгоритму складається з двох кроків. На Е-кроці (expectation) обчислюється очікуване значення функції правдоподібності, при цьому приховані змінні розглядаються як спостережувані. На М-кроці (maximization) обчислюється оцінка максимальної правдоподібності, таким чином збільшується очікувана правдоподібність, що обчислюється на Е-кроці. Потім це значення використовується для Е-кроку на наступній ітерації. Алгоритм виконується до збіжності.

Алгоритм Баума-Уелча гарантує, що модель з новими параметрами буде дорівнювати або буде кращою за попередню модель з точки зору критерію максимальної правдоподібності [31].

Отож, необхідно підібрати параметри прихованої моделі Маркова так, щоб максимізувати ймовірність даної послідовності спостережень. Для цього вводяться змінні

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \tag{3.15}$$

які показують імовірність того, що при заданій послідовності спостережень O система в моменти часу t і t+1 буде знаходитися відповідно в станах S_i і S_j. Використовуючи пряму і обернену змінні:

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)} \tag{3.16}$$

Вводяться змінні ймовірності того, що при заданій послідовності спостережень O система в момент часу t буде перебувати в стані S_i:

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \tag{3.17}$$

При цьому ми можемо обчислити очікуване число переходів зі стану S_i:

$$\sum_{t=1}^{T-1} \gamma_t(i) \tag{3.18}$$

а очікуване число переходів зі стану S_i в стан S_j:

$$\sum_{t=1}^{T-1} \xi_t(i,j) \tag{3.19}$$

Виходячи з цього можна отримати формули для переоцінки параметрів моделі Маркова:

$$\pi_i^* = \gamma_t(i) \tag{3.20}$$

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{3.21}$$

$$b_{ij}^*(k) = \frac{\sum_{t=1,o_t=k}^{T-1} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{3.22}$$

Вираз $\sum_{t=1,o_t=k}^{T-1} \gamma_t(i)$ у формулі для $b_{ij}^*(k)$ означає що підсумовуються тільки ті $\gamma_t(i)$, для яких значення стану дорівнює k, тобто $o_t = k$.

Після переоцінки параметрів моделі або з'ясовується, що вона вже була оптимальною до переоцінки або обов'язково поліпшуються її параметри (тобто правдоподібність моделі після переоцінки вище, ніж до переоцінки у всіх випадках, коли модель можна оптимізувати).

3.2 Особливості використання ПММ для невідомих артефактів

Система розпізнавання мовлення складається з процесу аналізу та обробки аналогового сигналу і процесу розпізнавання. При аналізі аналогового сигналу з мовлення виділяються властивості, які використовуються далі в процесі розпізнавання для того, щоб визначити, що було сказано.

Задача розпізнавання мовлення зазвичай розв'язується шляхом задання (представлення) еталонних слів словника і подальшого порівняння звукових сигналів з цими еталонами (розв'язується задача знаходження відповідності між звуковими сигналами і еталонами слів словника). Для моделювання артефактів, яких немає в словнику системи розпізнавання можна використати спосіб аналогічний тому, як це робилося у [38] для моделювання екстралінгвістичних явищ. Складність цієї задачі полягає в тому, що різні частини звукового сигналу в різних вимовах одного і того ж слова відрізняються ступенем зжатості та розтягнення. Окрім того, необхідно підібрати такі параметри моделі, які б враховували всі слова, яких немає в слонику системи розпізнавання.

Для задання фонетичних еталонів використовуються статистичні методи, які припускають, що акустичні параметри фонем розподілені по нормальному закону. В реальності, точна модель еталонів звуків та слів повинна включати в себе множину еталонних елементів (по одному на кожен варіант вимови). В якості акустичних моделей використовуються ПММ з гаусівською функцією розподілу ймовірностей появи векторів.

3.3 Використання ГСМ для моделювання артефактів

В роботі [39] слова та інші артефакти, які відсутні у словнику системи розпізнавання, моделюються довільною послідовністю фонем. Але застосування цього підходу показує, що такі моделі не є достатньо ефективними і можуть давати велику похибку в залежності від якості голосового сигналу.

Для усунення встановленого недоліку пропонується три етапи моделювання невідомих слів та артефактів [40]:

- моделювання OOV слів у вигляді ГСМ з одним станом, що дозволить визначити необхідну кількість сумішей;
- моделювання OOV слів декількома станами ГСМ, що дозволить визначити необхідну кількість станів;
- моделювання груп невідомих слів за їх довжиною, а саме: для моделювання коротких слів потрібно брати менше станів, для довших слів – більше станів.

Акустичні моделі будуються на основі використання прихованих марковських моделей та гаусівських сумішей. Для навчання системи всі OOV слова у НВ замінюємо на слово «unknown». Модель для цих слів спрощуємо:

- вважаємо, що модель слова складається з однієї фонеми;
- всі невідомі слова представляються однією фонемою «unknown».

На рис. 3.1 показано представлення фонеми за допомогою ПММ лінійного типу з трьома емісійними станами.

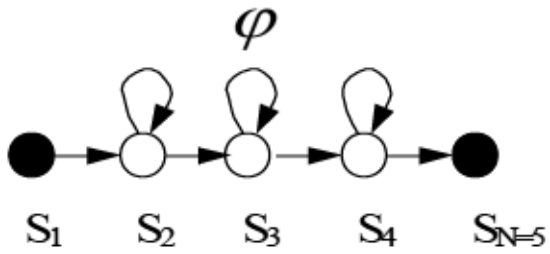

Рисунок 3.1 – ПММ лінійного типу

Для опису ПММ використовуються наступні основні параметри:

1) $S = \{S_1, S_2, \dots, S_N\}$ – множина вершин моделі;

2) N – кількість станів моделі;

3) $O = \{o_1, o_2, \dots, o_T\}$ – послідовність векторів спостереження (обсервацій мовного сигналу);

4) $o_t = \{x_{t1}, x_{t2}, \dots, x_{tn}\}$ – вектор спостереження, що представляє собою ознаковий опис мовного сигналу на певному стаціонарному проміжку;

5) T – кількість різних векторів спостереження;

6) A_{nn} – матриця перехідних ймовірностей, між станами;

7) N функцій щільності ймовірності $f_i(x)$.

Як правило, в системах, що базуються на ПММ, класифікацію ділянки сигналу, що відноситься до того чи іншого стану ПММ, здійснює гаусова суміш [41].

Параметр ПММ – функція щільності ймовірності $f_i(x)$ описується зваженою гаусівською сумішшю:

$$f(X) = \sum_{i=1}^{M} w_i p_i(x), \qquad (2.23)$$

де M – кількість компонент суміші; w_i – ваги компонентів суміші; $p_i(x)$ – нормальний розподіл ймовірностей. Кожен компонент є D-вимірною гаусівською функцією розподілу вигляду:

$$p_i(x) = \frac{1}{(2\pi)^{D/2}|\sigma_i|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu_i)^T \sigma_i^{-1}(x-\mu_i)\right)}, \quad (2.24)$$

де μ_i – вектор математичного очікування; σ_i – матриця коваріації. Для ваг суміші повинна виконуватись умова $\sum_{i=1}^{M} w_i = 1$.

Повністю модель гаусівської суміші визначається ймовірностями переходу, векторами математичного очікування, коваріаційними матрицями й вагами сумішей (по алгоритму Баума-Уелча) для кожного компонента моделі [41]. Всі разом ці параметри записуються у вигляді

$$\lambda = \{p_i, \mu_i, \sigma_i, w_i\}, i = 1, \dots, M. \quad (2.25)$$

.

Таким чином, функція щільності ймовірності асоціюється з кожним станом ПММ.

При побудові математичних моделей для опису OOV слів з використанням ПММ та ГСМ йдемо від найпростіших моделей до більш складних на кожному етапі визначаючи необхідні характеристики. Параметри моделі коригуються при навчанні.

Висновки до розділу

При детальному аналізі кожного з етапів моделювання артефактів, які не входять у словник системи розпізнавання встановлено, що моделювання груп невідомих слів за їх довжиною є найбільш логічним способом моделювання. Це пояснюється тим, що всі слова відрізняються довжиною. Тому для моделювання коротких слів потрібно брати менше станів, а для довших слів – більше станів. Попередньо, для моделювання артефактів, що не входять у словник розпізнавання можна використовувати гаусівську суміш моделей з одним та декількома станами для всіх невідомих слів для визначення необхідної кількості сумішей та кількості станів відповідно.

4 ПРОГРАМНА РЕАЛІЗАЦІЯ МОДЕЛЕЙ НЕВІДОМИХ АРТЕФАКТІВ

4.1 Архітектура ПЗ системи розпізнавання артефактів природної мови

Ядром системи розпізнавання є програмний пакет HTK – інструментарій для побудови ПММ. Звязки між модулями системи розпізнавання артефактів природної мови встановлюються інструментальними засобами HTK. Загальна схема роботи програмного комплексу показана на рис. 4.1.

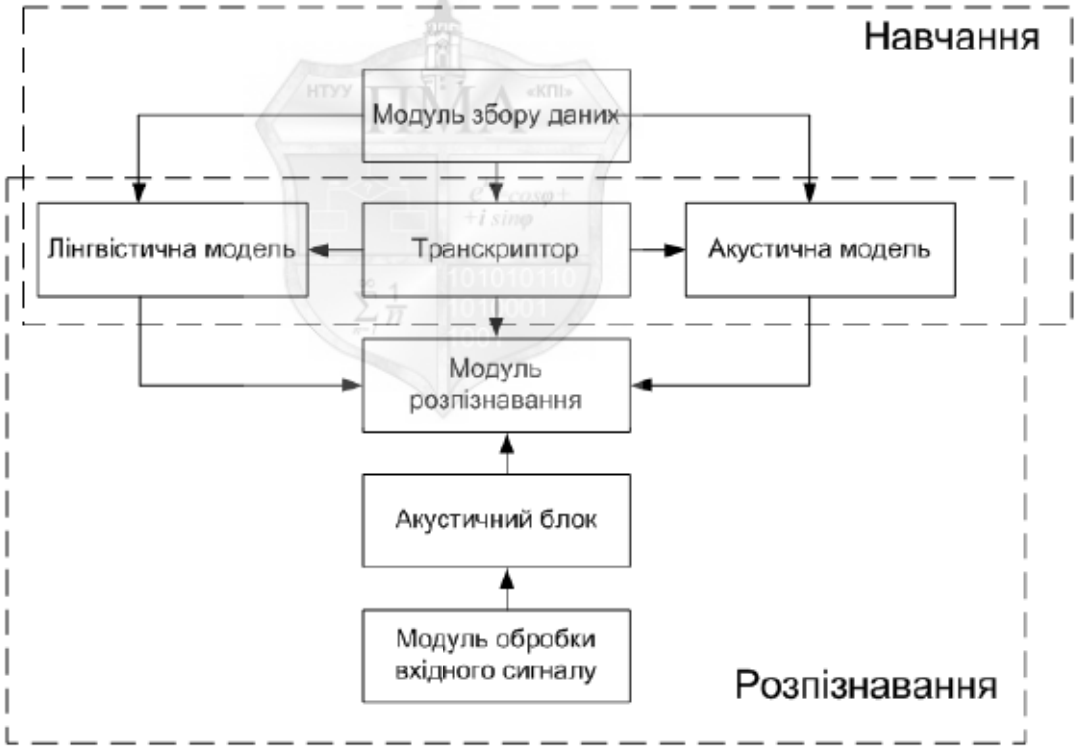

Рисунок 4.1 – Схема роботи програмного комплексу

HTK (Hidden markov model ToolKit) [20] – це комплекс програмних засобів, призначений для побудови прихованих марковських моделей. В роботі інструментарію можна виділити дві основних пов'язаних стадії обробки. По-перше, навчаючі інструментальні засоби HTK застосовуються

для оцінки параметрів безлічі ПММ, з використанням навчальних зразків вимови та відповідних їм транскрипцій. По-друге, невідомі зразки вимови транскрибуються за допомогою засобів розпізнавання HTK.

Вдосконалення полягає у зміні наступних компонентів: «Модуль збору даних» та «Акустичні моделі». Розробка моделей для слів, які не ввішли в словник системи розпізнавання була здійснена на мові програмування C, з використанням бібліотек існуючої сиситеми. Для навчання ПММ використовувались бібліотеки HTK. Розпізнавання, з урахуванням розроблених моделей для невідомих слів також здійснювалось за допомогою бібліотек HTK.

В HTK не існує інструменту, який безпосередньо реалізує алгоритм Вітербі. Замість цього використовується інструмент HVite, який, разом з його бібліотеками підтримки, HNet і HREC, призначений для розпізнавання злитого мовлення з наявним словником, крім того враховуючи можливість появи слів, яких немає в словнику. Оскільки така система розпізнавання є синтаксично орієнтованою, з її допомогою можна вирішувати і часткову задачу розпізнавання ізольованого слова. Розпізнавання на випадок злитого мовлення отримується простим послідовним з'єднанням декількох ПММ. Кожна модель в цій послідовності прямо відповідає передбачуваному базовому символу. Це можуть бути або цілі слова при так званому розпізнаванні зв'язного мовлення, або фрагменти слів, такі як фонеми, при розпізнаванні злитого мовлення.

4.1.1 Реалізація компонента «Модуль збору даних»

Першим етапом обробки є модифікація словників і граматик з метою внесення в них даних про OOV слова, а також створення навчальних файлів

розмітки аудіо сигналу. Дані файли використовуються для навчання моделей за допомогою пакету HTK і отримання файлів результату.

Використовуючи корпус мовлення та словник формується фонетична транскрипція (рис. 4.2).

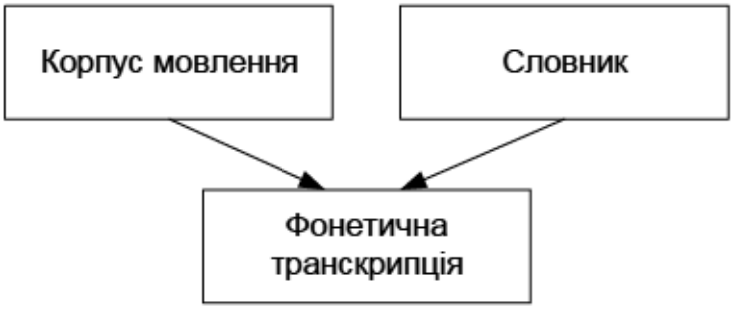

Рисунок 4.2 – формування даних для навчання ПММ

Щоб навчити набір ПММ, кожен файл навчання повинен мати пов'язану з ним фонетичну транскрипцію. Відправною точкою для наборів фонетичних транскрипцій є орфографічна транскрипція у форматі розмітки HTK. Така транскрипція отриується в результаті виконання функції `prompts2mlf`. Результатом є конвертування сценарію висловлювань до вигляду, який представлений на рис. 4.3.

```
#!MLF!#
"'*'/a04-8192010_LR_SPK00056_000000.lab"
в
рамках
цієї
справи
обговорюється
питання
про
тлумачення
пункту
третього
статті
чотириста
сто
третьої
закону
про
банкрутство
.
"'*'/a04-8192010_LR_SPK00056_000001.lab"
на
підставі
наказу
.
```

Рисунок 4.3 – вигляд файлу міток (MLF)

Тут мітки сценарію перетворюються в назви шляхів, кожне слово записується в окремому рядку, і кожне висловлювання закінчується крапкою. Перший рядок файлу ідентифікує його як головний файл міток (Master Label File (MLF)). Це окремий файл, що містить повний набір транскрипцій.

У випадку появи слова, якого немає в словнику системи розпізнавання на його місці у файлі транскрипцій виводиться «BGNDW» (рис 4.4)

```
2788  "'*'/a05-80032010_LR_SPK00076_000063.lab"
2789  було
2790  винесене
2791  BGNDW
2792  BGNDW
2793  рішення
2794  від
2795  двадцять
2796  шостого
2797  нуль
2798  третього
2799  дві
2800  тисячі
2801  десятого
2802  року
2803  у
2804  справі
2805  про
2806  порушення
2807  антимонопольного
2808  законодавства
2809  .
2810  "'*'/a05-80032010_LR_SPK00076_000064.lab"
2811  межі
2812  використання
```

Рисунок 4.4 – вигляд файлу міток (MLF)

Форма назви шляху, що використовується в даному файлі, потребує деякого пояснення, оскільки насправді це не ім'я, а шаблон (pattern). Коли HTK обробляє звукові файли, він очікує знайти транскрипцію (або файл розмітки) з тим же ім'ям, але з іншим розширенням. Використання в шаблоні символа «*» дозволяє використовувати однакові транскрипції з різними варіантами мовленнєвих даних, записаних у різних місцях, що дозволяє навчити ПММ на різних варіантах вимови різними дикторами.

Після того як створений MLF на рівні слів, генерується MLF на рівні фонем командою

HLEd -l '*' -d dict -i phones0.mlf mkphones0.led words.mlf

де words.mlf – файл MLF на рівні слів, а опція -l потрібна, щоб згенерувати шлях '*' в результуючих шаблонах.

Останнім етапом підготовки даних є параметризація мовленнєвих сигналів в послідовності векторів характеристик. Для цього створюється конфігураційний файл (config), що визначає всі параметри перетворення:

```
# Coding parameters
TARGETKIND = MFCC_0
TARGETRATE = 100000.0
SAVECOMPRESSED = T
SAVEWITHCRC = T
USEHAMMING = T
PREEMCOEF = 0.97
NUMCHANS = 26
NUMCEPS = 13
 ENORMALISE = F
```

Такий конфігураційний файл означає, що результатом перетворення повинні бути мел-кепстральні коефіцієнти (MFCC), протяжність кадру, де характеристики мовлення вважаються постійними, 10 мс (HTK використовує одиниці по 100 нс), результат повинен бути збережений в стислому вигляді, повинно використовуватись вікно Хемінга, а сигнал попередньо перетворюється фільтром першого порядку з коефіцієнтом 0.97, фільтри повинні мати 26 каналів, а результатом є 13 коефіцієнтів MFCC. Змінна ENORMALISE забезпечує нормалізацію енергії в записаних аудіо файлах і за замовчуванням має значення true. Нормалізація неможлива при роботі з живим звуком, і оскільки система в кінцевому рахунку призначена для роботи з живим звуком, цій змінній повинно бути присвоєно значення false.

4.1.2 Реалізація компонента «Акустичні моделі»

Структура взаємодії компонентів вдосконаленої моделі представлена на рис. 4.5.

Рисунок 4.5 – Структура взаємодії компонентів

Дані, отримані від компонента «Модуль збору даних» використовуються для навчання акустичних моделей. На кожному з трьох етапів моделювання кількість станів ПММ для невідомих слів буде різною.

Ініціалізація ПММ для OOV слів здійснюється шляхом задання довільних параметрів, які в процесі навчання коригуються. Початкові параметри задаються в подібній конструкції:

```
<BeginHMM>
  <NumStates> 5
  <State> 2
          ...
  <State> 3
          ...
  <State> 4
          ...
  <TransP> 5
          ...
 <EndHMM>
```

Вектор ознак описується 39-ма коефіцієнтами. Така кількість коефіцієнтів визначається як сума підрахованих мел-кепстральних коефіцієнтів (13 коефіцієнтів), коефіцієнтів дельта (13 коефіцієнтів) та

коефіцієнтів прискорення (13 коефіцієнтів) і задається як <VecSize> 39 <MFCC_0_D_A>.

На першому етапі моделювання OOV слів визначається необхідна кількість ГСМ. Число компонент гаусівських сумішей моделей послідовно збільшується з одночасним навчанням на корпусі мовлення. Коли модель перестане покращуватись – зупиняємось.

На другому етапі, з урахуванням визначеної кількості ГСМ, визначається необхідна кількість станів для моделювання невідомих слів.

На третьому етапі для моделювання коротких слів беремо менше станів, для довгих слів – більше станів.

Навчання ПММ невідомих слів здійснюється на навчальному корпусі мовлення з використанням бібліотек НТК. В навчанні беруть участь всі невідомі слова, які є в тренувальному файлі. В результаті формуються еталони для OOV слів.

4.2 Тестування. Адекватність моделей

При дослідженнях використовується мовленнєва база даних «Записи Верховної Ради України» – записи виступів депутатів Верховної Ради України, записані через телевізор. Мовний матеріал, що використовується для побудови АМ, складається з аудіозаписів тривалістю близько 52 годин, в яких міститься промова близько 400 дикторів. Текстовий матеріал, що використовується для побудови лінгвістичних моделей, складається з текстів, завантажених з Інтернету (2 Гб текстів українською мовою). Для корпуса характерні наступні особливості: спонтанне слабко структуроване мовлення та різний темп вимови.

Контрольна вибірка вміщує 1900 слів. Словник системи розпізнавання налічує 71 000 слів.

На таких тестових даних існуюча система розпізнавання давала точність 70,3 %. Після використання запропонованої моделі точність розпізнавання виросла до 74, 2 %.

На рис. 4.6 та рис. 4.7 показано результати роботи системи до вдосконалення та після.

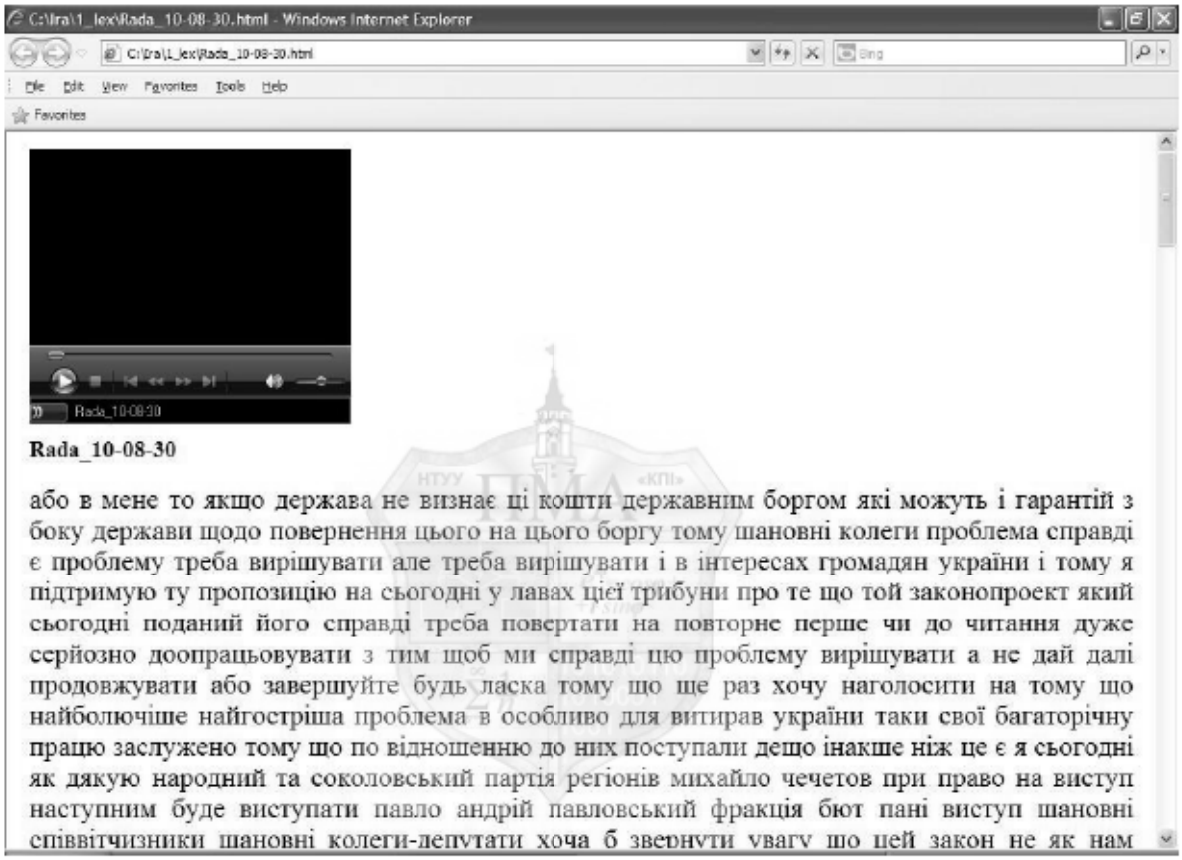

Рисунок 4.6– Результат роботи системи до вдосконалення

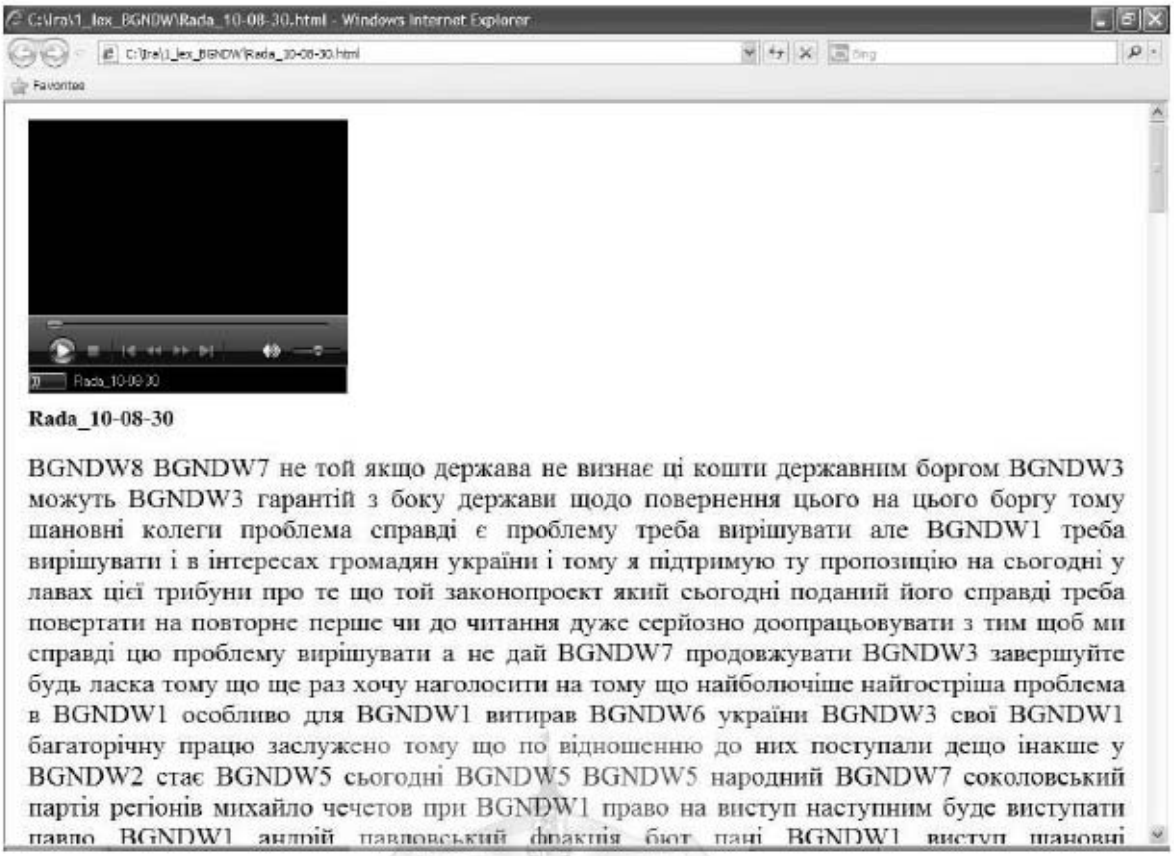

Рисунок 4.7 – Результат роботи системи після вдосконалення

4.3 Керівництво користувача

Оскільки система розпізнавання є дикторонезалежною, то не потрібно здійснювати ніяких налаштувань на голос конкретного диктора, не потрібно проводити додаткове навчання системи.

Для старту процесу розпізнавання необхідно запустити файл tst.bat – розпізнавання або в командному рядку викликати набір наступних команд:

HVite-o ST-T 1-l '*'-C config-a-H hmm1/hmmdefs-i hmm1/recout1.mlf-p 0.0-s 5.0-S test.scp-w wdnet dict.trn monophones

...

HVite-o ST-T 1-l '*'-C config-a-H hmm7/hmmdefs-i hmm7/recout7.mlf-p 0.0-s 5.0-S test.scp-w wdnet dict.trn monophones

де:

config - файл конфігурації {(додається)},

hmm1/hmmdefs - файли результатів навчання одного кроку {(конкретно, після першого кроку навчання, бо після hmm стоїть цифра 1)},

recout1.mlf - результат розпізнавання, записується в директорію hmm1,

test.scp - список файлів для розпізнавання,

wdnet - мережа допустимих слів і виразів (в даному випадку - словосполучення),

dict.trn - словник розпізнавання,

monophones - список фонем.

Результат розпізнавання аудіо файлу після першого кроку навчання розпізнаванню записується у файл recout1.mlf. Результати розпізнавання тестової вибірки слів після другого, третього і т.д. кроків навчання розпізнаванню записуються у файли recout2.mlf – recout7.mlf.

Результати навчання, взяті на будь-якому кроці навчання і записані у файлі hmmdefs (hmmdefs – файл визначень ПММ для фонем), можуть бути використані для подальшої роботи по розпізнаванню слів. Конкретний вибір залишається за користувачем, і залежати цей вибір має від результатів розпізнавання так, щоб досягалася хороша надійність розпізнавання (якщо не найкраща).

Запустити файл res.bat - оцінка результату розпізнавання (у відсотках). Якщо такого файлу немає, то в командному рядку викликати команди наступного вигляду:

mkdir results

HResults-I words_test.mlf monophones hmm1/recout1.mlf> results/output1

HResults-I words_test.mlf monophones hmm2/recout2.mlf> results/output2

…

HResults-I words_test.mlf monophones hmm6/recout6.mlf> results/output6

HResults-I words_test.mlf monophones hmm7/recout7.mlf> results/output7

де:

recout1.mlf - результат розпізнавання після першого кроку навчання, записаний раніше відповідно в директорію hmm1,

output1 - результат аналізу розпізнавання, записується відповідно в директорію results,

monophones - список фонем,

words_test.mlf - файл тестових слів.

Результати аналізу розпізнавання для кожного кроку навчання відображаються у файлах output1-output7.

Висновки до розділу

Отримані результати дозволяють суттєво розширити область застосування системи розпізнавання в різних галузях. При наявній розмірності словника системи розпізнавання, без додаткових затрат на його розширення значно уточнюється розпізнавання для стенографування важливих нарад особливо в політичній діяльності.

Така модель враховує наявність слів, що не входять у словник системи розпізнавання, «виловлює» ці слова і таким чином в результаті розпізнавання замість іншого подібного по звучанню слова стенографісту надається позначка в якому місці зустрілось слово, що не входить у словник

і надається можливість прослухати аудіо запис цього слова і відредагувати результат розпізнавання в текстовому редакторі.

Окрім того, можна накопичувати статистику слів, що зустрічаються і потім доповнювати словник системи розпізнавання з подальшим навчанням системи розпізнавання.

ВИСНОВКИ

У роботі розв'язано наукову задачу пов'язану з удосконаленням математичної моделі для розпізнавання артефактів слабко структурованої української природної мови. Зокрема, досліджено основні методи та алгоритми, що застосовуються для розпізнавання мовлення, здійснено огляд існуючих засобів розпізнавання, запропоновано вдосконалену модель розпізнавання артефактів української природної мови, встановлено, що моделювання груп невідомих слів за їх довжиною є найбільш логічним способом моделювання, проведено аналіз результатів розпізнавання при використанні запропонованої моделі, встановлено, що точність розпізнавання системи вцілому підвищилась.

Сфера застосування технології в комерційних цілях досить широка: медицина (введення даних в електронні картки пацієнтів), субтитри (підготовка субтитрів для ТБ), стенографування (документування усних виступів і заходів із збереженням дослівного змісту). Таку систему можна використовувати для синхронного документування (протоколювання) усних виступів, засідань, зборів та конференцій, що проводяться органами виконавчої, законодавчої та судової влади, а також громадськими та комерційними організаціями. Окрім того, корисною буде для людей з обмеженими можливостями.

Використання такої моделі дозволяє спростити роботу стенографістів при документуванні усних виступів на засіданнях, симпозіумах та подібних заходах. Така модель підвищує швидкість і комфортність документування звукових записів усного мовлення.

ПЕРЕЛІК ПОСИЛАНЬ

1. Киевская городская библиотека (КГБ) [Електронний ресурс]. – Режим доступу:http://lib.misto.kiev.ua/
2. Сайт з розпізнавання та синтезу мовлення в Україні [Електронний ресурс]. – Режим доступу: http://speech.com.ua/ukraine.html
3. Центр Речевых Технологий [Електронний ресурс]. – Режим доступу: http://www.speechpro.ru
4. Audio-Visual Speech Recognition (AVSR) [Електронний ресурс]. – Режим доступу:http://www.intel.com
5. J. K. Baker. Stochastic modeling for automatic speech understanding. In D. R. Reddy, editor, Speech Recognition // Academic Press, New York. – 1975. – P. 521-541.
6. S. Das, R. Bakis, A. Nadas, D. Hahamoo, M. Picheny. Influence of background noise and microphone on the performance of the IBM tangora speech recognition system. Proc. of ICASSP, Vol. II. – 1993. – P. 71-74.
7. Dragon NaturallySpeaking Solutions [Електронний ресурс]. – Режим доступу: http://www.dragonsys.com
8. Speeding Medical Documentation [Електронний ресурс]. – Режим доступу:http://www.provox.com
9. Фролов А., Фролов Г. Синтез и распознавание речи. Современные решения [Електронний ресурс] / Александр Фролов, Григорий Фролов. – Электрон. журн. – 2003. – Режим доступу: http://www.frolov-lib.ru
10. Burger S., Sloane Z., Yang. J. Competitive Evaluation of Commercially Available Speech Recognizers in Multiple Languages / Susan Burger, Zachary Sloane, Jie Yang. – Pittsburgh: Carnegie Mellon University, 2006. – P. 809-814.

11. Федосин С.А., Еремин А. Ю.Классификация систем распознавания речи // электронное научное периодическое издание: электронный журнал / Электроника и информационные технологии [Електронний ресурс]. – Режим доступу: http://fetmag.mrsu.ru/2010-2/pdf/SpeechRecognition.pdf
12. М.М.Биков, Т.В.Грищук. Моделювання процесу аналізу і класифікації голосових команд: Монографія – Вінниця: ВНТУ, 2009. – 128 с. ISBN 978-966-641-322-5
13. Распознавание речи. Часть 1. Классификация систем распознавания речи [Електронний ресурс]. – Режим доступу: http://geektimes.ru/post/64572/
14. Фаніна Л.О. Аналіз тенденцій побудови систем мовного інтерфейсу // электронное научное периодическое издание: электронный журнал / Информационно-измерительные системы [Електронний ресурс]. – Режим доступу: http://aaecs.org/fanna-lo-analz-tendenci-pobudovi-sistem-movnogo-nterfeisu.html
15. Мовознавчий вісник [Текст] : зб. наук. пр. / Черкаський нац. ун-т імені Богдана Хмельницького. – Черкаси, 2006. – 17-21 с.
16. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – Киев, 1987г. – с. 264.
17. Винцюк Т.К. Сравнительный теоретический анализ ИКДП- и НММ-методов распознавания речи // Автоматическое распознавание слухових образов. Тезисы докл. 15-й Всесоюзного семинара (Таллин-89). – Таллин: ИК АН ЭССР. – 1989. – С 18–24.
18. Jurafsky D., Martin J.H.Speech and Language processing. 2nd ed. Englewood Cliffs, New Jersey: Pretince Hall Inc., 2008. 302 p.
19. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition // Proceedings of the IEEE. – 1989. – vol. 77, №2. – P. 257–286.

20. Young, S. The HTK Book (for HTK Version 3.4) [Text] / Steve Young, Gunnar Evermann, Mark Gales [and other] / Cambridge University Engineering Department, 2006. – 359 p.
21. Система розпізнавання NaturallySpeaking [Електроний ресурс] — Режим доступу: http://www.dragonsys.com
22. Система розпізнавання IBM ViaVoice [Електроний ресурс] — Режим доступу: www.ibm.com/viavoice
23. Система розпізнавання VoiceType [Електроний ресурс] — Режим доступу: http://www-4.ibm.com/software/speech/
24. Система розпізнавання MedSpeak [Електроний ресурс] — Режим доступу: http://www.ibm.com/
25. Система розпізнавання Voice_PE [Електроний ресурс] — Режим доступу: http://www.voicerecognition.com/kurzweil/voicedes.html
26. Система розпізнавання Voice Xpress Professional [Електроний ресурс] — Режим доступу: www.lhs.com
27. Система розпізнавання Sakrament [Електроний ресурс] — Режим доступу: http://www.sakrament.com/
28. Система розпізнавання "Горинич" ПРОФ 3.0 [Електроний ресурс] — Режим доступу: http://www.upspecial.ru/gorynych-prof-3-0.html
29. Система розпізнавання RealSpeaker [Електроний ресурс] — Режим доступу: http://www.realspeaker.net/ua/
30. Програмний продукт Visual Paradigm [Електроний ресурс] — Режим доступу: http://www.visual-paradigm.com/support/documents/vpuserguide/94/2580/6713_creatingacti.html
31. David Barber. Bayesian reasoning and machine learning, Cambridge University Press, 2014.
32. Чесебиев И.А. Компьютерное распознавание и порождение речи / И.А. Чесебиев. – М.: Спорт и культура, 2008 – 128 с.

33. Kevin Murphy. Machine Learning – A probabilistic perspective, MIT Press, 2012.
34. Stadermann J., Rigoll G. Ahybrid. SVM/HMM acoustic modeling approach to automatic speech recognition, In Proc. Of INTERSPEECH-2004 ICSLP. – 2004. – P. 661–664.
35. A. Viterbi. Error bounds for convolutional codes and anasymptotically optimum decoding algorithm // IEEE Trans. Information Theory. – 1967. – vol. 13, no. 2, pp. 260–269.
36. Er Liu. Convolutional Convolutional Coding & Coding & Viterbi Viterbi Algorithm, Helsinki University of Technology, 2004.
37. Rabiner, Lawrence. First Hand: The Hidden Markov Model. // IEEE Global History Network. – Retrieved 2 October. – 2013
38. Пилипенко В., Гузієнко І. Кластеризація екстралінгвістичних явищ на прикладі мовленнєвих записів Верховної Ради України // Збірник праць Дванадцятої Всеукраїнської міжнародної конференції з оброблення сигналів і зображень та розпізнавання образів УкрОбраз'2014, Київ, 2014. С 91–94.
39. Пилипенко В.В. Використання фонетичного стенографа при розпізнаванні мовлення з великих словників // Тези 12-ї міжнародної конференції «Автоматика – 2005», Харків, 2005, с. 73.
40. Маслянко П. П., Гузієнко І. В. Вдосконалена модель розпізнавання артефактів слабко структурованої української природної мови // Збірник тез доповідей VII наукової конференції магістрантів та аспірантів «Прикладна математика та комп'ютинг – ПМК'2015», Київ, 2015. С 191–196.
41. Reynolds D.A., Rose R.C. Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models // IEEE Trans. Speech Audio Process – 1995.– 3.– P. 72–83.