МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

І.А. Терейковський,
Л.О. Терейковська

РОЗПІЗНАВАННЯ ФОНЕМ В ГОЛОСОВОМУ СИГНАЛІ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня магістр
за освітньою програмою «Системне програмування та спеціалізовані комп'ютерні системи»
спеціальності 123 Комп'ютерна інженерія

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2022

Рецензент Толюпа С. В. доктор технічних наук, професор, професор кафедри кібербезпеки та захисту інформації факультету інформаційних технологій Київський національний університет імені Тараса Шевченка

Відповідальний редактор Тарасенко В.П., доктор технічних наук, професор

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № X від DD.MM.YYYY р.)
за поданням Вченої ради факультету прикладної математики
(протокол № 1 від 01.09.2022 р.)

Навчальний посібник містить матеріали для самостійної роботи здобувачів ступеня магістр за освітньою програмою «Системне програмування та спеціалізовані комп'ютерні системи» спеціальності 123 «Комп'ютерна інженерія» при вивченні розділу «Обробка голосових сигналів та зображень» з дисципліни «Цифрова обробка сигналів та зображень». Також стане у нагоді розробникам програмного забезпечення, аспірантам та студентам технічних снеціальностей.

Реєстр. № НП XX/XX-XXX. Обсяг X,X авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
https://kpi.ua
Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

ЗМІСТ

Список скорочень

АНМ – асоціативна нейронна мережа

БШП – багатошаровий персептрон

ВШ – вхідний шар

ГНМ – глибока нейронна мережа

ГС – голосовий сигнал

ДШП – двохшаровий персептрон

КС – комп’ютерна система

НМ – нейронна мережа

НМЗ – нейромережевий засіб

НММ – нейромережева модель

НМС – нейромережева система

РБФ – мережа радіальної базисної функції

СДН – система дистанційного навчання

СРГС – система розпізнавання голосового сигналу

СШН – схований шар нейронів

ШВ – вихідний шар

ШД – шар додавання

ШО – шар образів

PNN – ймовірнісна нейронна мережа

Вступ

Навчальний посібник присвячено опису підходів до вирішення актуальної науково-прикладної задачі розробки ефективних нейромережевих моделей, методів і засобів розпізнавання фонем в голосовому сигналі, пристосованих до умов використання в поширених веб-орієнтованих комп'ютерних системах. Обгрунтовано, що:

- Важливим напрямком підвищення ефективності модулю розпізнавання голосового сигналу в комп'ютерній системі є впровадження нейромережевих засобів розпізнавання фонем. Для цього необхідно розвинути методологічну базу і розробити методи побудови нейромережевих засобів розпізнавання фонем в голосовому сигналі, які забезпечать їх адаптацію до очікуваних умов застосування.

- Розвиток методологічної бази нейромережевого розпізнавання фонем за рахунок розроблених принципів та моделей забезпечує можливість створення ефективних нейромережевих методів розпізнавання фонем.

- Запропонований метод створення навчальної вибірки для нейромережевої моделі розпізнавання фонем в голосовому сигналі за рахунок врахування у вихідному сигналі близькості фонем та застосування експертних знань для визначення параметрів навчальних прикладів дозволяє приблизно в 2,4 рази зменшити кількість навчальних ітерацій та забезпечує можливість створення навчальної вибірки при відсутності доступу до баз даних фонем.

- Запропонований метод нейромережевого розпізнавання фонем в голосовому сигналі за рахунок визначення параметрів найбільш ефективної нейромережевої моделі, використання методу створення навчальної вибірки та визначення прогнозованого навантаження на модуль розпізнавання забезпечує достатню низьку похибку розпізнавання близько 0,01.

- Запропонований метод оцінки ефективності застосування нейромережевих засобів розпізнавання фонем за рахунок використання запропонованого принципу та критеріїв оцінки ефективності дозволяє обрати найбільш ефективний засіб.

РОЗДІЛ 1. АНАЛІЗ МОЖЛИВОСТЕЙ ВПРОВАДЖЕННЯ ЗАСОБІВ РОЗПІЗНАВАННЯ ГОЛОСОВИХ СИГНАЛІВ В КОМП'ЮТЕРНИХ СИСТЕМАХ

1.1. Науково-практична задача розпізнавання голосових сигналів

В загальному випадку задача розпізнавання ГС полягає в його автоматичній обробці з метою визначення послідовності слів [5, 9]. Складність розв'язку такої задачі пояснюється необхідністю врахування варіативності ГС, типу введення мови, розміру словника, рівня навколишнього шуму. Для вирішення задачі розпізнавання ГС розроблюються СРГС, створення яких ускладнюється необхідністю врахування різноманітних факторів, наприклад, розташування мікрофону [5, 11]. Сучасні СРГС, як правило, мають ієрархічну структуру [15, 23]. На першому акустичному рівні виконується попередня обробка та виділення акустичних ознак, які характеризують ГС.

Наступний рівень СРГС – лінгвістичний, в який входить процедура пошуку ГС по словниках еталонів. Крім того, СРГС можуть включати в себе фонетичний, фонологічний, морфологічний, лексичний, синтаксичний та семантичний рівні. Типова послідовність функціонування СРГС наведена на рис. 1.1.

Відповідно [23, 25], найбільш складним етапом роботи СРГС є реалізація процедури розпізнавання, результатом якої є визначення еталону, що відповідає невідомому ГС. Складність процедури розпізнавання пояснюється нелінійною зміною темпу мовлення слів та різною тривалістю пауз на початку і в кінці слова. Тому процедура розпізнавання розділяється на декілька етапів.

Вхідний ГС розділяється на елементи – фонеми, алофони, дифони, трифони, склади. Для вказаних елементів знаходяться еталони, а вже за допомогою еталонів елементів знаходяться еталони окремих слів [5, 19, 30]. Методи розпізнавання окремих слів на основі еталонів елементів також вважаються достатньо апробованими та надійними [33, 38, 40].

В той же час задача знаходження еталонів окремих елементів далека від свого вирішення [43, 45].

Рис. 1.1. Типова послідовність функціонування системи розпізнавання голосових сигналів

Результати [45, 47] дозволяють стверджувати про перспективність використання в якості окремих елементів фонем, що пояснюється їх відносною малочисельністю в порівнянні з кількістю складів, алофонів, дифонів та трифонів.

Процес визначення меж окремих фонем в ГС описано в [37-39]. Аналіз літературних джерел показав, що переважна більшість апробованих СРГС будуються на основі методів динамічного програмування, НМ та СММ. Перевагами методу динамічного програмування є простота встановлення часової відповідності між вхідним ГС та еталонним, а недоліками є висока обчислювальна складність та дикторозалежність.

Застосування НМ базується на їх здатності класифікувати ГС, задані за допомогою коефіцієнтів, які відповідають розрахованому вектору ознак ГС [47, 49]. Переваги нейромережевих методів: доведена ефективність при вирішенні важкоформалізуємих задач, стійкість до шумів у вхідних даних, висока швидкість розрахунків та низькі потреби обчислювальних ресурсів при прийнятті рішення, стійкість до часткових відмов при апаратній реалізації НММ. До недоліків відносять складність адаптації НММ до нестаціонарного вхідного сигналу, проблеми вибору параметрів НММ.

Використання СММ базується на постулаті, що ГС може бути представлений за допомогою схованого ланцюга Маркова. Перевагами методу є простота його застосування. До основних недоліків СММ відносять складність формування бази різноваріантних еталонних елементів слів та високу обчислювальну складність розрахунку параметрів марківської моделі.

Вказані недоліки дещо компенсуються за рахунок сумісного використання марківських моделей та НММ, що в свою чергу негативно впливає на складність моделі.

Крім того, відомі спроби застосування в СРГС динамічних мереж Байєса, машини опорних векторів, а також теорії несилової взаємодії. Широкому використанню цих методів заважає низька апробованість та необхідність адаптації до практичних аспектів застосування в СРГС.

Таким чином, одним із перспективних шляхів вирішення науково-практичної задачі розпізнавання ГС є розробка нейромережевих методів розпізнавання фонем, виділених із цього сигналу.

1.2. Аналіз сучасних нейромережевих моделей та методів розпізнавання голосових сигналів

В процесі аналізу робіт, присвячених використанню НМ для розпізнавання ГС, виявлено, що для більшості з них характерна деяка невідповідність назви та опису наведеної розробки. Наприклад, в назві роботи вказується на розробку НМС, хоча власне розробка полягає у визначенні алгоритму обробки вхідних параметрів для НММ. Тому аналіз вказаних робіт проведено з єдиних позицій визначення основних характеристик нейромережевих методів та моделей. Наведемо отримані дані.

Алгоритм автоматичного розпізнавання складів на основі лінійної авторегресивної НМ і теорії нечітких множин (**ААРС**) наведений в [37]. В роботі запропоновано визначення фонеми як нечіткої множини мінімальних мовних одиниць. Запропоновано алгоритм розпізнавання складів. На вхід алгоритму подається множина еталонних мінімальних мовних одиниць. Задачею алгоритму є віднесення невідомої мовної одиниці до однієї із еталонних. Показані приклади застосування алгоритму. Наведені результати вказують на недостатню точність розпізнавання, яка знаходиться в межах 3-35%.

В [17] наведено **спосіб розпізнавання голосових команд** за допомогою ТК (**СРГК**). Показано приклад проектування командної дикторозалежної системи. Описано побудову топології ТК, процес навчання якої визначається рівнянням наступного виду:

$$w_n = w_s + \alpha(x - w_s), \tag{1.1}$$

де w_n, w_s – нове та попередне значення ваги зв'язку між вхідною компонентою *x* та нейроном мережі; α – коефіцієнт швидкості навчання.

Для навчання такої мережі в роботі пропонується підготувати навчальну вибірку з різних варіантів наговорених визначених команд.

Нейромережу із затримкою часу (НЗЧ) для розпізнавання приголосних звуків «B», «D» і «G» представлено в [71]. Вхідними сигналами є 16 нормалізованих масштабованих спектральних коефіцієнтів. Масштабовані коефіцієнти розраховуються із спектра потужності сигналу. В **НЗЧ** використано метод оберненого розповсюдження помилки. Запропоновано корегування вагових коефіцієнтів для зсунутих у часі з'єднань, а саме оновлювати кожну вагу відповідного з'єднання до середнього значення із всіх затриманих у часі вагових змін. Така обчислювальна процедура достатньо складна, що пояснюється великою кількістю ітерацій, які необхідні для запам'ятовування комплексного багатовимірного простору вагових коефіцієнтів та навчальних образів.

В [10] запропоновано **підхід до навчання нейронної мережі із застосуванням генетичного алгоритму (ПНЗГА**). Наведено приклад розпізнавання десяти слів. Механізмом розпізнавання слів передбачено асоціацію кожного окремого слова з окремою НМ. Невідомий приклад подається на входи всіх НМ. Після цього визначається НМ з максимальним вихідним сигналом. Вважається, що слово, асоційоване з даною мережею, відповідає невідомому прикладу. Навчання НМ здійснювалось шляхом послідовного пред'явлення навчальних прикладів та налаштування вагових коефіцієнтів зв'язків, поки помилка навчання по всій множині не досягла прийнятно низького рівня. Помилка розраховувалась по формулі:

$$E = \frac{1}{2N}\sum_{i=1}^{N}(y_{ki} - d_i)^2, \tag{1.2}$$

де N – кількість навчальних прикладів НМ; y_{ki}, d_i – реальний та бажаний вихід НМ.

Розпізнавання голосових команд за допомогою **згорткових нейронних мереж (ЗНМ)** запропоновано в [23]. Такі мережі працюють з двохвимірними

даними – нейрони в кожному шарі утворюють площини. Вхідний шар представлено однією площиною, розмірність якої співпадає з розмірністю вхідних даних. Наступні шари мережі є згорнутими та складаються з площин нейронів (карти ознак). Кожен нейрон згорнутого шару з'єднаний з невеликою підобластю попереднього шару. Останні два шари мережі представляють собою НМ прямого розповсюдження. Наведено приклад практичного застосування. Показані переваги ЗНМ: невелика кількість параметрів навчання та висока точність навчання.

В [25] розроблено **алгоритм розпізнавання ізольованих слів** (**АРІС**) за допомогою РБФ, яка використовувалась для виділення найбільш інформативних ознак еталонів. Навчання РБФ відбувалось методами кластерного аналізу та градієнтного спуску. Також представлено алгоритм налаштування СРГС на нового диктора. Доводиться, що використання РБФ дозволяє значно підвищити частоту правильного результату розпізнавання проблемних (акустично схожих) слів та скоротити обсяг навчальної вибірки для процедури налаштування системи розпізнавання на нового диктора.

Метод розпізнавання слів на основі двохмірних векторів (МДСДВ) показано в [14]. Класифікація векторів здійснюється за допомогою НМ, на вхід якої поступає низькочастотні двохмірні вейвлет-перетворення інтервалів спектрограм окремих слів. Спектрограма масштабується до квадратного виду та проводиться двохмірне вейвлет-перетворення. Показано, що основною перевагою методу є зосередження найбільш шумостійкої частотно-часової інформації в компактному частотно-часовому векторі. Наведено приклад розпізнавання голосового потоку з обмеженою кількістю слів. Результати дозволяють визначити недостатню точність розпізнавання мови (30-35%).

В [22] описано **метод автоматичного розпізнавання ізольованих слів (МАРІС**). При побудові архітектури НМ та при виборі числа вхідних нейронів враховано фактичні аспекти сприйняття мови людиною. Тому НМ складається з 25 нейронів ВШ, трьох нейронів внутрішнього шару та одного вихідного нейрону, функція активації – сигмоїдальна. В експерименті наговорювалися десять чисельників, а також назви 40 міст. Число дикторів – 17 жінок та 15 чоловіків. Для виділення фонеми використано алгоритм мел-

кепстрального аналізу. Наведені результати вказують на недостатню точність розпізнавання мови, в межах 13-23%.

Метод розпізнавання фонем в голосовому сигналі (МРФГС) описано в [66]. Показано, що для використання методів розпізнавання фонем необхідна попередня сегментація ГС для отримання вектору властивостей окремих фонем. Вказано на недоліки, пов'язані із наперед заданим видом та кількістю коефіцієнтів перетворення, що може призвести до дублювання та надлишкової інформації в ГС. Запропонована структура системи розпізнавання фонем в ГС. Вхідний сигнал потрапляє на вхід лінії затримки, яка формує вхідний вектор НМ. Далі сигнал поступає в нейронний шар кореляції. Потім – на вхід шару аналізу. Виходами шару аналізу є ознаки окремої фонеми. До переваг системи відносять компактність її структури.

В засобах розпізнавання ГС, що застосовуються в пошуковій системі «Яндекс» [34], використана **нейронна мережа для акустичного моделювання (НМАМ)**. НМ використовувалась для розпізнавання сенонів, окремих частин фонем. В акустичній моделі використовується 48 фонем та близько 4000 сенонів. Входом є 13 мел-кепстральних коефіцієнтів звукового потоку, а виходом – розподіл ймовірностей по сенонах. В акустичній моделі використовувалась стохастична НМ з алгоритмом навчання без вчителя, що використана для побудови НММ на основі БШП.

Відзначається, що застосування стохастичної НМ підвищило якість розпізнавання в два рази. НМ не мають обмежень, які характерні для СММ, мають кращу узагальнюючу здатність, стійкіші до шуму та більш швидкодійні.

В нейромережевій системі розпізнавання ГС компанії Microsoft (**MSNET**) використовуються НММ виду БШП. Хоча деталізованого опису даної системи знайти не вдалось, але по опосередкованим даним [57, 58, 64, 65] та виходячи із практичного досвіду можливо зазначити, що для визначення оптимального виду та параметрів НММ використано багатокритеріальний підхід, метод навчання оптимізовано з позицій мінімізації терміну навчання. Крім того, при побудові системи враховані обмеження, що стосуються обчислювальних ресурсів.

Підхід сумісного використання (ПСВ) багатошарових НММ персептронного типу та СММ розглянуто в [4]. Враховані переваги НМ – можливість навчання за допомогою невеликої акустичної бази даних з фонетичною транскрипцією, тоді як параметри СММ оцінюються тільки на основі лексичної бази даних великого обсягу. Вказано, що недоліком БШП є погана пристосованість до роботи з часовими рядами. Також показано, що сумісний підхід дозволяє використати умовні ймовірності на етапі лінгвістичного моделювання і тим самим звузити коло пошуку.

В [9] описано процедуру розпізнавання ГС за допомогою **імпульсних нейронних мереж (ІНМ)**, спеціально створених для класифікації часових процесів. Відзначається більш якісне імітування природних нейронів у порівнянні з класичним персептроном. Також вказується на відсутність для імпульсних НМ апробованого математичного апарату.

Метод навчання персептрону без сегментації слів (МНБСС) запропоновано в [31]. Розглянуто використання ДШП, призначеного для розпізнавання окремих фонем в ГС. В якості вхідних параметрів використано 16 енергетичних характеристик ГС, отриманих на 15 часових вікнах. Для розрахунку енергетичних характеристик використано вираз:

$$x_i = \log \frac{E_i}{E_{\min}},\ i = 1, \ldots M, \tag{1.3}$$

де E_i – енергія в i-ій частотній смузі; $E_{\min}$ – найменша з енергій; $M = 16$ (кількість частотних смуг).

Кожен із вихідних нейронів БШП відповідає окремій фонемі. В процесі навчання на кожному окремому слові передбачено не коригувати вагові коефіцієнти зв'язків, що ведуть до вихідних нейронів, для яких в цьому слові немає відповідних фонем. За рахунок цього декларується можливість відмови від процедури попередньої сегментації ГС.

В [59] запропоновано **принцип підвищення швидкості навчання (ППШН)** та узагальнюючих можливостей НМ за рахунок використання вейвлет-перетворень. На основі даного підходу розроблена нова модифікація НМ – з модулем зворотньої вейвлет-декомпозиції вихідного сигналу. Запропоновано використання вказаного підходу та модифікованої НМ для

розпізнавання ГС. Наведено результати порівняльних експериментів розпізнавання ізольованого звуку «а» за допомогою звичайного БШП та модифікованої НМ. Відзначається значна перевага останньої, що виражається у десятикратному зменшенні терміну навчання та у зменшенні відносної помилки розпізнавання з 11,57% до 0,003%.

Нейромережевий алгоритм розпізнавання мови (**НАРМ)** описаний в [27]. Алгоритм розпізнавання складається із наступних етапів: введення акустичного сигналу в комп'ютер та виділення меж слів, виділення параметрів, які характеризують спектр сигналу, використання БШП для оцінки близькості акустичних параметрів, порівняння з еталонами в словнику. В якості 39 вхідних сигналів НМ використано результати спектрального аналізу ГС. ШВ складався із 19 нейронів, виходи яких порівнювались методами динамічного програмування із еталонами. Наведені результати доводять, що точність нейромережевого розпізнавання не поступається точності традиційних акустико-фонетичних методів.

В [7] описано процедуру **кластеризації простору ознак мовних сигналів** (**КПОМС**) за допомогою ТК. Як критерій кластеризації використано вираз:

$$d(x, y) = \sum \alpha_i |x_i - y_i|, \qquad (1.4)$$

де x, y – вектори властивостей навчальної вибірки; α_i – апріорні коефіцієнти.

Показано можливість використання ТК для кластеризації фонем і для кластеризації слів. Запропоновано різні модифікації методів навчання ТК. Доводиться, що точність розпізнавання ГС знаходиться в межах 96-99%.

В [14] наведено обґрунтування **вибору мовних ознак** (**ОВМО**) для навчання НМ. Доводиться, що початковими даними для навчання НМ можуть бути фонеми. Показано, що для формування вхідних даних НМ можливо застосувати результати: вейвлет-перетворення, віконного перетворення Фур'є або перетворення Гільберта-Хуанга для обробки енергетичних показників ГС. Наведені результати свідчать про однакову ефективність вказаних перетворень.

Метод нечіткого співставлення мовних образів (МНСМО) в нейромережевому базисі при розпізнаванні ізольованих слів російської мови описано в [54]. В методі мовний сигнал представлено у вигляді двохвимірного спектрально-часового образу, отриманого за допомогою віконного перетворення Фур'є. Невідомий образ y розглядається як чітке відношення між множиною частот і множиною часових інтервалів. Для нього розраховується ступінь схожості S_j з кожним нечітким відношенням r_j. Результатом розпізнавання являється слово j, котре належить множині еталонних слів J і для якого $j = \max_{j \in J} S_j$. Вказується, що при застосуванні даного методу точність розпізнавання недостатня і сягає всього 76%.

В [19] розроблено **нейромережевий підхід до інтегрованого представлення і обробки інформації** (**НПІПІ**) в інтелектуальних системах. Розроблено методи та алгоритми нейромережевої обробки мовної інформації та створена НМС розпізнавання ключових слів. Пропонується використовувати динамічні асоціативні запам'ятовуючі пристрої на основі рекурентних НМ. В якості вхідної інформації використано параметри, що характеризують звукову хвилю. В цілому система розпізнавання представляє ієрархічну структуру, яка аналізує мовний потік на акустико-фонетичному, морфологічному, лексичному та семантичному рівнях. Для аналізу на акустико-фонетичному рівні використана ТК, вхідні дані якої отримані за рахунок застосування до ГС методу перцептивного лінійного передбачення.

В [53] розглянуто **набір нейромережевих моделей для розпізнавання фонем** (**ННМРФ**) в ГС при управлінні текстовим редактором. Використано ДШП із структурою зв'язків 20-10-1, кількість входів якого відповідає періоду основного тону конкретного диктора.

В [25] описана система **автоматичного розпізнавання мови на базі нейромережевої технології (АРМНТ)**. Система має ієрархічну структуру, що відповідає традиційним рівням обробки мовної інформації – акустико-фонетичному, лексичному та фразовому. Розроблено специфічну структуру НММ, яка базується на використанні РБФ. Розглянуто особливості настройки системи розпізнавання на конкретного диктора. Також описуються загальні підходи до застосування динамічних АНМ на верхніх рівнях аналізу.

В [12] розглянуто метод **визначення найбільш інформативних ознак мовного сигналу (BIOMC).** Метод призначений для визначення вхідних параметрів ДШП, що використовується для виділення вокалізованих сегментів і класифікації голосних фонем. Доведено, що в задачі виділення вокалізованих сегментів високу інформативність має параметр кореляції між відліками. Для розрахунку критерію інформативності використано вирази:

$$M'(m_1, m_2) = \left(m_1^5 + k^5 m_2^5\right)^{1/s}, \tag{1.5}$$

$$M'(m_1, m_2) = \left(m_2^5 + k^5 m_1^5\right)^{1/s}, \tag{1.6}$$

де $k \in [1, \infty)$ – коефіцієнт значимості помилок 1-го та 2-го роду; $m_1 \in [0,1)$, $m_2 \in [0,1)$ – значення помилки 1-ого та 2-го роду; $s \in [0,1)$ – коефіцієнт значимості помилок.

Запропоновано використовувати для пошуку голосних формант параметри, які враховують структуру голосних звуків.

В [15] запропоновано **метод побудови пристрою розпізнавання мови** (**МППРМ**) на базі гібридної нейро-марківської моделі. Доведено, що НМ дозволяють створювати акустичні моделі, які є більш компактними відносно моделей, заснованих на теорії СММ. Задекларовано можливість створення такої моделі на основі БШП та рекурентних НМ. Однак акцент ставиться на використанні глобально-рекурентних БШП типу НМД. Описано два випадки об'єднання НММ та марківської моделі. В першому випадку НМ використовується тільки на етапі акустичного аналізу, а на етапі лексичного аналізу використовується СММ. В другому випадку НМ використовується для визначення перехідних ймовірностей СММ. Відзначено, що труднощі створення НММ пов'язані із нестаціонарним характером ГС.

В [54] описано **метод ковзаючого фонетичного аналізу** (**МКФА**) і структура багаторівневої системи автоматичного розпізнавання мови на основі НМ. Метод базується на постулаті про те, що вокалізований ГС складається із стаціонарних інтервалів, які характеризують фонеми, та нестабільних інтервалів, котрі відносяться до міжфонемних переходів. Для забезпечення нечуттєвості до зміни тривалості звучання фонем в методі використана ідея апроксимації стаціонарних інтервалів НММ еталонних елементів мови. Для опису еталону запропоновано використовувати НММ

виду БШП із структурою виду 80-10-10-1. Наведені результати розпізнавання фонем підтверджують перспективність запропонованого методу.

В [11] описана дикторонезалежна та стійка до акустичних шумів **система автоматичного пошуку ключових слів** (**САПКС**) в потоці неперервної мови. В системі використано ДШП та МЗР. ДШП використано безпосередньо для розпізнавання ключових слів, а МЗР – для зберігання та пошуку еталонів ключових слів. Вхідними параметрами системи є результати вейвлет-перетворень параметрів ГС. Оптимізація вагових коефіцієнтів ДШП проводилась по критерію мінімізації помилки пропуску ключового слова. Структура ДШП оптимізована з використанням теореми Колмогорова, що дозволяє розрахувати тільки мінімально-допустиму кількість схованих нейронів. Показано, що при відношенні сигнал/шум 30 дБ надійність розпізнавання 90%, що є достатнім для ІС загального призначення.

Нейромережева модель розпізнавання злитного мовлення (**НМРЗМ**) запропонована в [72]. НМРЗМ базується на концепції розділення ГС на окремі фонеми та переходи між ними. Модель складається із трьох модулів, що представляють собою НМ, здатні навчатись за допомогою правила Хебба. Входом моделі є мел-кепстральні коефіцієнти ГС. В першому модулі ГС сегментується на інтервали різної довжини. Кожен із інтервалів відповідає різноманітним варіантам фонем та переходам між ними. В другому модулі розпізнаються фонеми, а в третьому – окремі слова. Наведені результати вказують, що точність розпізнавання близько 90%.

В засобах розпізнавання ГС, що застосовуються в пошуковій системі «Google» та в операційній системі «Android» використовуються так звані **ГНМ** [13, 63]. Структурно ГНМ представляють собою БШП з великою кількістю СШН та двохетапним навчанням. На першому етапі відбувається попередня настройка вагових коефіцієнтів. Використовується метод «без вчителя», що базується на застосуванні обмеженої машини Больцмана. На другому етапі навчання реалізується метод «з вчителем» з використанням алгоритму зворотнього поширення помилки. Необхідність двохетапного навчання пов’язана з низькою ефективністю загальновживаних алгоритмів корекції вагових коефіцієнтів при великій кількості СШН. У випадку

невеликої кількості СШН (2-4) у двохетапному навчанні **потреби** не виникає. Використання ГНМ дозволяє оброблювати потужні голосові потоки, однак потребує значних обчислювальних потужностей. Також зазначимо, що цілісного детального опису методів побудови ГНМ не знайдено.

В [61] описана **система автоматичного розпізнавання фонем** (**САРФ**) в ГС. В системі передбачено двохетапну методику розпізнавання. На першому етапі ГС розділяється на фонеми. Для цього використовується аналіз частотних характеристик ГС. Розпізнавання фонем відбувається на другому етапі із використанням НММ або СММ. Описана підготовка її вхідних параметрів та алгоритм сегментації. Характеристики проаналізованих нейромережевих засобів наведено в табл. 1.1.

Таблиця 1.1

Характеристики проаналізованих нейромережевих моделей та методів

№	Назва	**Призначено для розпізнавання**				**Тип НМ**								
		складів	слів	фонем	сонем	ЗНМ	БШП	РБФ	ГНМ	ТК	МЗР	АНМ	НМЛ	ІНМ
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
1	ААРС	+										+		
2	СРГК		+							+				
3	НЗЧ			+									+	
4	ПНЗГА		+				+							
5	ЗНМ		+			+								
6	АРІС		+					+						
7	МДСДВ		+				+							
8	МАРІС		+				+							
9	МРФГС			+									+	
10	НМАМ				+		+			+				

Таблиця 1.1 (продовження)

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
11	MSNET			+	+		+							

12	ПСВ			+			+						
13	ІНМ		+	+									+
14	МНБСС			+			+						
15	НАРМ		+				+						
16	ППШН			+			+						
17	КПОМС		+	+						+			
18	ОВМО			+			+						
19	МНСМО		+				+						
20	НПІПІ		+								+		
21	ННМРФ		+				+						
22	АРМНТ	+	+	+				+					
23	ВІОМС			+			+						
24	МППРМ			+								+	
25	МКФА			+			+						
26	САПКС		+				+				+		
27	НМРЗМ			+	+		+						
28	ГНМ		+	+					+				
29	САРФ			+							+		

Аналіз даних табл. 1.1 вказує на те, що більшість відомих НМС спрямовані на розпізнавання фонем та окремих слів. При цьому використовуються БШП, ТК, ГНМ, МЗР та РБФ.

Крім того, в результаті аналізу, по аналогії з [20, 28] визначено, що ефективність НМЗ розпізнавання ГС в КС пов’язане з реалізацією в них множини базових процедур **G**:

- однокритеріального та багатокритеріального вибору виду НММ (G1, G2),
- однокритеріального та багатокритеріального вибору параметрів НММ (G3, G4),
- адаптації методу навчання (G5),
- ефективного кодування вихідних параметрів (G6),

- визначення допустимих видів НММ (G7),
- формування ефективної навчальної вибірки (G8),
- прогнозу достатності обчислювальних ресурсів (G9).

Наявність вказаних процедур в проаналізованих нейромережевих моделях та методах наведено в табл. 1.2.

Аналіз даних табл. 1.2 вказує на недостатню адаптацію відомих НМЗ до умов вітчизняних КС. Проведене дослідження результатів наукових робіт дозволяє стверджувати, що ефективне розпізнавання ГС в КС можливо реалізувати на основі нейромережевого методу розпізнавання виділених фонем, в якому передбачається виконання визначених базових процедур.

Таблиця 1.2

Наявність базових процедур

№	**Назва**	**Процедура**								
		G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
1	ААРС	-	-	-	-	-	+	-	-	-
2	СРГК	-	-	+	+	+	-	-	-	-
3	НЗЧ	+	-	+	+	+	-	-	-	-
4	ПНЗГА	-	-	+	-	-	-	-	-	-
5	ЗНМ	+	-	+	-	+	-	-	-	-
6	АРІС	+	+	+	+	+	-	-	+	-
7	МДСДВ	+	+	+	-	-	-	-	-	-
8	МАРІС	+	-	-	-	-	-	-	-	-
9	МРФГС	+	-	+	+	+	-	-	-	-
10	НМАМ	+	+	+	+	+	-	-	+	+
11	MSNET	+	+	+	+	+	-	+	-	+
12	ПСВ	+	+	-	-	+	-	-	-	-
13	ІНМ	+	+	+	+	-	-	-	-	-

Таблиця 1.4 (продовження)

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*

14	МНБСС	+	-	+	-	-	+	-	-	-
15	ППШН	+	-	+	+	-	+	-	-	-
16	НАРМ	-	-	+	-	-	-	-	-	-
17	КПОМС	+	-	+	-	-	-	-	-	-
18	ОВМО	+	+	+	-	-	+	-	-	-
19	МНСМО	+	-	+	-	-	+	-	-	-
20	НПІПІ	+	-	+	-	+	-	-	+	-
21	ННМРФ	+	-	+	-	-	-	-	-	-
22	АРМНТ	+	-	+	-	+	-	+	-	-
23	ВІОМС	+	-	-	-	+	-	-	-	-
24	МППРМ	+	+	+	-	+	-	-	-	-
25	МКФА	+	+	+	-	-	-	-	-	-
26	САПКС	+	+	+	-	+	-	-	-	-
27	НМРЗМ	+	+	+	-	+	-	-	-	-
28	ГНМ	+	+	+	+	+	+	+	+	+
29	САРФ	-	-	+	-	+	-	-	-	-

Розробка нейромережевого методу розпізнавання виділених фонем призводить до необхідності вдосконалення методологічних засад застосування НММ розпізнавання фонем в ГС в КС та прогнозування достатності обчислювальних ресурсів КС.

1.3. Аналіз підходів до прогнозування навантаження на модуль розпізнавання голосових сигналів в веборієнтованих комп'ютерних системах

Оскільки кількість розробок в області прогнозування навантаження на модуль розпізнавання ГС в веборієнтованих КС обмежена, то в процесі аналізу використано результати [28], присвячені схожій темі прогнозування навантаження на вебсервер інформаційних систем загального призначення. Зазначимо, що в [28] аналіз науково-практичних робіт проведено з позицій визначення підходів до розробки засобів прогнозування навантаження на КС

при застосуванні в ній технологій голосової взаємодії. Наведемо отримані результати аналізу.

Технологія моделювання завантаження каналу зв'язку для системи дистанційного навчання. Описані фактори, що впливають на пропускну здатність каналу зв'язку. Показано, що обсяг навчального курсу із звуковим супроводженням як мінімум в 20 разів більший, ніж обсяг аналогічного курсу без звуку. Визначено, що термін отримання клієнтом системи дистанційного навчання результатів запиту не повинен перевищувати 5 с. Для прогнозування навантаження використано пуасонівський потік запитів до ресурсів системи дистанційного навчання. Розроблена модель дозволяє оцінити час завантаження сторінки та каналу зв'язку. Визначно періодичний характер завантаження веб-сервера СДН.

Модель розподіленої обчислювальної системи обробки науково-освітніх ресурсів, яка базується на кластерній архітектурі веб-сервісів. Показано, що у вищих навчальних та наукових закладах в основному використовуються наступні веб-сервіси: Twitter, Delicious, SlideShare, Moodle, WikiSpaces, Elluminate, Lectora, TeacherTube, BaseCamp.

Розглянуто залежність характеристик кластеру від варіантів його реалізації. Розроблені математичні моделі для балансування навантаження кластеру в умовах стаціонарного та стохастичного потоку запитів. Наведено вирази для розрахунку оптимальної продуктивності серверів по критерію мінімізації терміну обробки запитів:

$$\begin{cases} \min\limits_{\alpha_i,\beta_i} \left\{ S(\alpha_1,...\alpha_N,\beta_1,...\beta_N) = \sum\limits_{i=1}^{N} \left\{ \dfrac{\alpha_i \alpha_i^2 \Lambda^2 \beta_i^2}{2(1-\alpha_i \Lambda \beta_i)} + \dfrac{\delta_i}{\beta_i}(1-\alpha_i \Lambda \beta_i) \right\} \right\}, \\ \sum\limits_{i=1}^{N} \alpha_i = 1,\ \alpha_i \geq 0,\ \beta \in \left(0; \dfrac{1}{\Lambda}\right) \end{cases} \quad (1.7)$$

де α_i – штраф за перебування запиту в черзі до і-го сервера на протязі одиниці часу; N – кількість серверів; β_i – середній час обслуговування запиту; Λ – інтенсивність загального потоку запитів на кластер; λ_i – інтенсивність потоку запитів на і-й сервер.

Книга Д. Менаске «Продуктивність Web-служб. Аналіз, оцінка та планування» носить оглядовий характер, в ній аналізуються наступні методи

прогнозування навантажень на веб-служби: регресійний, ковзаючого середнього, нелінійний та експоненційного згладжування. Відзначено необхідність вдосконалення існуючих методів для адекватного опису типового навантаження на веб-служби.

В роботі С.В. Ільніцького «Робота мережевого сервера при самоподібному (self-similar) навантаженні» розглянуто підхід до прогнозу використання потужностей веб-серверу для «пікових ресурсів» на основі екстраполяції. Показано, що завантаження мікропроцесора веб-сервера залежить від кількості активних процесів Apache з коефіцієнтом 1,1. При цьому коефіцієнт максимального завантаження процесора дорівнює 0,85. Відзначається періодичний характер вказаної залежності. В підході не враховуються пікові навантаження, що призводить до помилки прогнозування 20-30%.

Задачу визначення потреб у ресурсах для Інтернет-орієнтованих систем при встановленому навантажені розглянуто в роботі R. Hariharan «Web Server Performance Modeling». Визначено, що основною вимогою до таких систем є обмеження терміну відповіді веб-серверу на запит користувача 5 секундами. Додатковою вимогою є використання потужності процесора не більш, ніж на 90%. Описуються підходи щодо визначення потреб у ресурсах на основі максимального навантаження, для визначення якого пропонується провести додаткові дослідження.

В статті М.А. Бесараба «Аналіз мережевого трафіку корпоративної мережі університету методами нелінійної динаміки» представлено метод розрахунку середнього навантаження на веб-сервер в передумові пуасонівського потоку запитів. Показано, що підвищення середнього навантаження на веб-сервер більш, ніж на 75% від максимально допустимого призводить до миттєвого критичного збільшення середнього часу очікування вхідних запитів.

Розробці моделей СДН ВНЗ з використанням апарату теорії масового обслуговування присвячена робота І.А. Кузнєцова «Дистанційне навчання як система масового обслуговування». Показано, що основною задачею моделювання є прогнозування поведінки СДН в залежності від режимів

навантаження веб-сервера. Відзначається необхідність створення ефективного блоку прогнозування навантаження веб-сервера.

В статті Д. І. Парфьонова «Моделювання затребуваності ресурсів у розподіленій інформаційній системі дистанційної підтримки навчального процесу» розроблено алгоритм розподілення навантаження при доступі до даних хмарної системи зберігання, який передбачає прогнозування потреб мультимедійних даних. Зазначено, що навантаження на ключові ресурси СДН ВНЗ носить періодичний і нерівномірний характер. Запропоновано прогнозувати обсяги навантаження за допомогою розподілення Парето (відеотрафік), розподілення Вейбула (статичні дані – бінарний трафік), χ^2 розподілу (статичні дані – трафік невеликого обсягу).

Розробці графо-аналітичної моделі обліку навантаження веб-систем, яка включає граф переходів запитів клієнта та математичну формалізацію станів системи, присвячена стаття М. А. Керенцевої «Розробка графо-аналітичної моделі обліку навантажень веб-базованих систем». Кожен перехід характеризується часом завантаження сервера та ваговим коефіцієнтом, що відображує відносну частоту здійснення переходу. Для зниження завантаження веб-системи пропонується визначити критичні переходи, виділити класи, які потребують зменшення навантаження та вибрати засіб зменшення.

В статті Д.Л. Жусова «Модель та алгоритм обслуговування користувачів корпоративного web-сервера» розроблена модель обслуговування користувачів корпоративного веб-серверу. Показано, що основне навантаження на веб-сервер створюють запити користувачів з низьким приорітетом. Для прогнозування навантаження на веб-сервер запропоновано використання математичного апарату теорії часових рядів, що дозволило врахувати в моделі самоподібність вхідного потоку запитів.

Методологія розрахунку навантаження на веб-сервер розглянута в [50]. Запропоновано ряд показників, що характеризують навантаження.

В [64] проведено аналіз та моделювання продуктивності веб-сервера Apache в передумові пуасонівського процесу надходження запитів до серверу з частотою λ. Використана модель веб-сервера, яка складається із

розподіленого вузла процесора та прикріпленої до нього черги з *n*-завдань. Показано, що оптимальні показники середнього часу обслуговування кожного запиту та максимальної кількості запитів можуть бути визначені методом найшвидшого спуску, сполученого градієнта або прямим перебором.

Низька ефективність класичних схем використання веб-сервісів доводиться в [49]. З метою оптимізації навантаження на канали зв'язку пропонується використовувати систему кеш-серверів, що, крім іншого, дозволяє прискорити термін отримання інформації користувачем.

В статті С. А. Крашакова «Кешування інформаційних потоків та стратегія оптимізації маршрутів у розподілених системах» проаналізовано трафік центру дистанційного навчання та в термінах теорії масового обслуговування розроблено модель опису клієнт-серверного навантаження. Виявлено самоподібність реального трафіку та показано його суттєву відмінність від результатів, отриманих за допомогою класичних пуасонівських моделей. Показано недоцільність використання класичних ланцюгів Маркова та експоненційних моделей для моделювання навантаження веб-серверу навчального центру СДН. В статті О. О. Дружиніної «Підвищення ефективності функціонування веб-серверів з використанням технології прогнозування часових рядів на основі нейромереж» запропоновано підхід до підвищення ефективності функціонування веб-серверів з використанням технології прогнозування часових рядів на основі нейронних мереж радіально-базисного типу. При цьому для розв'язання задачі короткострокового прогнозування навантаження на сервер пропонується використовувати узагальнено-регресійну НММ виду GRNN. Низьку ефективність сучасних методів балансування та розподілу навантаження веб-серверу визначено в роботі Д. І. Парфьонова «Моделювання розподілу ресурсів та динамічного балансування навантаження в інформаційній системі дистанційної підтримки освітнього процесу». На підтвердження такого висновку наведено характерні графіки обслуговування клієнтських запитів (рис. 1.2). Аналіз графіків показує, що в середньому близько 40-60% запитів не обслуговуються. Доводиться, що однією із основних причин низької ефективності є неадекватність

загальнопоширених моделей прогнозу до типового характеру навантаження веб-серверу СДН.

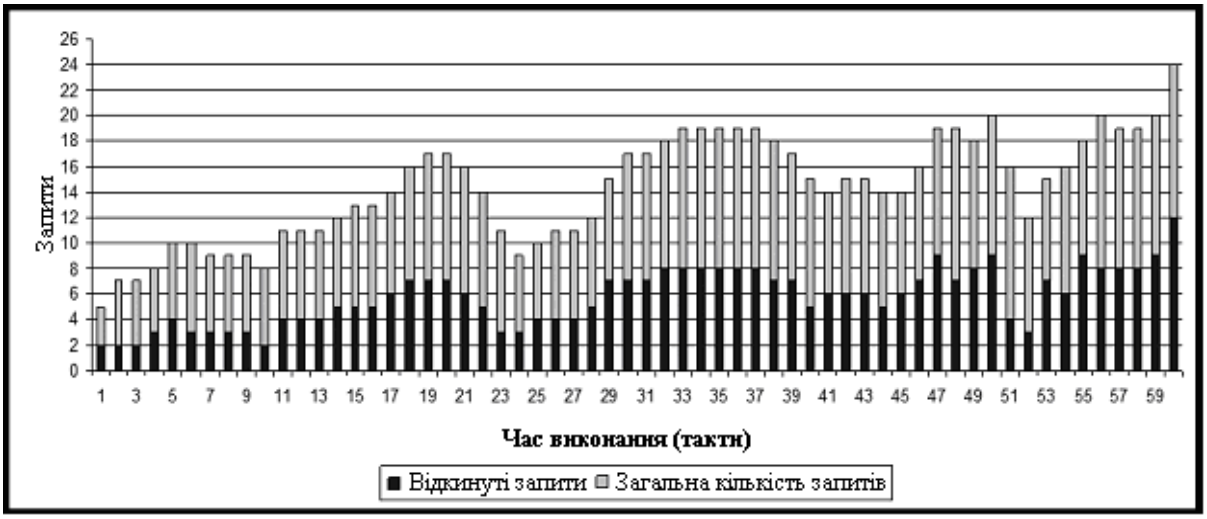

Рис.1.5. Графік обслуговування запитів без врахування приорітетів

Методика збору даних в веб-середовищах представлена в статті Горбатюка М. М. «Використання інформаційних технологій в навчально-виробничому процесі». Показано, що для опису поведінки веб-користувача можливо використовувати степеневі функції, популярність веб-об'єктів описується розподілом Зіпфа, а розміри файлів – розподілом Парето. Також доведена кореляція запитів користувачів, незалежність запитів різних користувачів до одного файлу та експоненційний закон розподілу часу між цими запитами. Зазначено, що веб-трафік містить пікові значення, які перевищують середнє значення в 5-10 разів та можуть перевищувати обчислювальні можливості сервера. Також показано, що навіть невелика пульсація трафіку має достатньо великий негативний вплив на продуктивність веб-сервера. В засобах компанії Cisco використовується алгоритм балансування навантаження CSM (Content Switching Module) для кластера серверів з метою перенаправлення мережевих підключень на сервер з найменшою кількістю з'єднань. Показано необхідність застосування циклічності прогнозування навантаження. Також використовується модель прогнозування стану сервера, яка базується на використанні теорії часових рядів та методів нечітких експертних оцінок. Описано алгоритм комплексної

оцінки стану сервера та метод регресійно-нечіткого моделювання для прогнозування стану сервера, заданого набором гетерогенних даних. Також відомо ряд досліджень [28], присвячених визначенню прогнозованого навантаження на веб-сервер на основі багатоперіодичної моделі клієнтських запитів до веб-серверу. Результати аналізу дозволяють сформулювати висновок про те, що методологія застосування засобів розпізнавання фонем в ГС в КС повинна враховувати складний багатоперіодичний характер прогнозованого навантаження на веборієнтований модуль розпізнавання. При цьому розрахунок прогнозованого навантаження можливо виконати за допомогою застосування теорії вейвлет-перетворень для прогнозування кількості клієнтських запитів до веб-серверу.

Разом з тим, результати проведеного аналізу вказують на те, що очікувані умови впровадження нейромережевих засобів розпізнавання фонем в ГС характеризуються варіативністю обмежень на термін розробки, залучення трудових ресурсів, акустичні параметри голосових сигналів та на обчислювальні ресурси КС. Окремо слід зазначити обмеження на використання баз даних прикладів аудіозаписів, необхідних для проведення навчання нейромережевих моделей, що в значній мірі впливає на точність розпізнавання засобів, які створені на їх основі. Також аналіз літературних джерел вказує на недостатню адаптацію сучасних нейромережевих моделей, методів розробки та застосування засобів розпізнавання ГС до варіативності означених умов впровадження. Відповідно, потребують теоретичного обґрунтування характеристики нейромережевих засобів розпізнавання. Виникає необхідність адаптації методологічної бази нейромережевого розпізнавання фонем в ГС до умов КС та розробки на цій базі методу створення навчальної вибірки для НММ та методу створення НМЗ розпізнавання фонем. Очевидно, що вказані методи потребують апробації, для чого доцільно розробити відповідну НМС та здійснити аналіз її ефективності.

РОЗДІЛ 2. ЕЛЕМЕНТИ МЕТОДОЛОГІЧНОЇ БАЗИ НЕЙРОМЕРЕЖЕВОГО РОЗПІЗНАВАННЯ ФОНЕМ В ГОЛОСОВОМУ СИГНАЛІ

2.1. Концептуальна модель забезпечення ефективності нейромережевого розпізнавання фонем

Результати досліджень [28] вказують на те, що важливим напрямком розвитку КС є застосування засобів автоматичного розпізнавання ГС користувачів. Для цього необхідно вирішити завдання нейромережевого розпізнавання фонем, попередньо виділених із ГС.

Особливістю сформульованого завдання є необхідність теоретичного обґрунтування характеристик нейромережевих моделей та методів, адаптованих до умов КС як у випадку дикторозалежного, так і у випадку дикторонезалежного розпізнавання. До вказаних умов відносяться:

- допустимий термін розробки,
- можливість залучення трудових ресурсів,
- наявність доступу до баз даних аудіозаписів, необхідних для навчання НММ,
- особливості акустичних параметрів ГС,
- доступний обсяг обчислювальних ресурсів КС.

Вирішення цього завдання дозволить розв’язати такі практичні задачі, як розпізнавання голосової відповіді в процесі комп’ютерного тестування, розпізнавання голосової команди та голосової аутентифікації користувачів КС за рахунок розпізнавання висловленого ним секретного слова (паролю).

У випадку дикторозалежного розпізнавання у вказаних практичних задачах враховуються особливості ГС конкретного користувача, а у випадку дикторонезалежного розпізнавання ці особливості не враховуються. При цьому задачі фільтрації ГС, виділення із ГС фонем та формування із розпізнаних фонем окремих слів вважаються вирішеними.

Відповідно рекомендацій [20, 28], відправним пунктом вирішення сформульованого завдання являється розробка концептуальної моделі

забезпечення ефективності нейромережевого розпізнавання фонем.

В загальному випадку концептуальна модель представляє собою модель предметної області, що складається з переліку взаємопов’язаних понять, котрі використовуються для опису цієї області разом з властивостями й характеристиками, класифікацією цих понять за типами, ситуаціями, ознаками в даній області, і законів протікання в ній процесів. Концептуальна модель являється відображенням концепції, під поняттям якої розуміють певний спосіб судження, трактовки деяких явищ, основну точку зору, керівну ідею для їх систематичного висвітлення.

Зазначимо, що розробка концептуальної моделі є загальноприйнятим відправним пунктом розвитку методологічної бази, яка представляє собою систему принципів і способів організації та побудови теоретичної і практичної діяльності, а також вчення про цю систему.

Оскільки практичний результат розробки системи розпізнавання передбачає створення програмно-апаратного забезпечення для розпізнавання фонем, то для визначення ефективності процесу нейромережевого розпізнавання фонем в ГС КС передбачено використовувати визначення з області комп’ютерної та програмної інженерії. Відповідно міжнародних стандартів цієї області, ефективність – це множина атрибутів, які визначають взаємозв'язок рівнів виконання програмної системи, використання ресурсів (засоби, апаратура, матеріали та ін.) і послуг, що виконуються штатним обслуговуючим персоналом та ін. До характеристик ефективності програмної системи належать:

- оперативність – атрибут, що вказує на час відгуку, обробки й виконання функцій;
- ресурсоємність – атрибут, що визначає кількість і тривалість використовуваних ресурсів при виконанні функцій програмної системи;
- погодженість – атрибут, що вказує на відповідність даного атрибута заданим стандартам, правилам та приписам.

Відповідно наведених визначень, на першому етапі створення концептуальної моделі було проведено гармонізацію термінології, що використовується в області застосування НМ для розпізнавання ГС.

Гармонізація проведена з позицій відображення сучасного стану науки і практики та підтримує вирішення задач дисертаційної роботи. В результаті визначені наступні терміни:

- ГС – складний акустичний сигнал, джерелом якого являється голос людини. В контексті даної дисертаційної роботи синонімом терміну ГС є мовний сигнал, хоча в загальному випадку між даними термінами є певні відмінності.
- Фонемоподібний елемент – виділений в ГС фрагмент, параметри якого відповідають окремій фонемі.
- Фонема – мінімальна структурно-функціональна звукова одиниця мови, яка служить для знаходження відмінностей та ототожнення значимих одиниць мови.
- НМ – мережа штучних нейронів, з'єднаних між собою синаптичними (зваженими) зв'язками.
- НММ – модель НМ, що характеризується методом навчання, способом розповсюдження сигналу, структурою зв'язків та типом штучного нейрону. Вказані параметри та їх комбінації визначають вид НММ [35]. Синонімом поняття виду НММ є архітектура НМ. Похідними від терміну НММ є нейромережеві методи, НМС та НМЗ, тобто це методи, системи та засоби, які базуються на НМ.

Оскільки в загальному випадку під поняттям засіб розуміють знаряддя (предмет, пристрій, сукупність пристроїв), то поняття НМЗ є збірним для НММ та НМС, що застосовуються для розпізнавання фонем в ГС КС. Апаратно-програмну реалізацію таких пристроїв будемо називати інструментальним НМЗ. Також концептуальна модель призначена для формалізації причинно-наслідкових зв'язків, які властиві процесу розпізнавання фонем в ГС, визначених необхідністю підвищення ефективності КС.

Крім того, в концептуальній моделі слід врахувати:

- умови функціонування НМЗ розпізнавання фонем, визначені характером взаємодії його окремих частин і компонентами КС;
- необхідність реалізації ефективного застосування НММ для

розпізнавання фонем та напрямок покращення його функціонування;

- можливості керування НМЗ та визначення його керованих змінних.

Враховуючи загальноприйняту технологію застосування НММ, побудовано показану на рис. 2.1 діаграму декомпозиції нейромережевого розпізнавання фонем.

Призначення складових частин цієї діаграми таке:

- Формування параметрів навчальних прикладів – визначення для кожної із фонем множини вхідних і вихідних параметрів та способу їх кодування до виду, придатного для використання в НММ.

- Формування навчальної вибірки – визначення такої множини навчальних прикладів, що відповідають еталонам фонем. Кількість, якість та номенклатура прикладів повинні бути достатніми для навчання НММ.

-

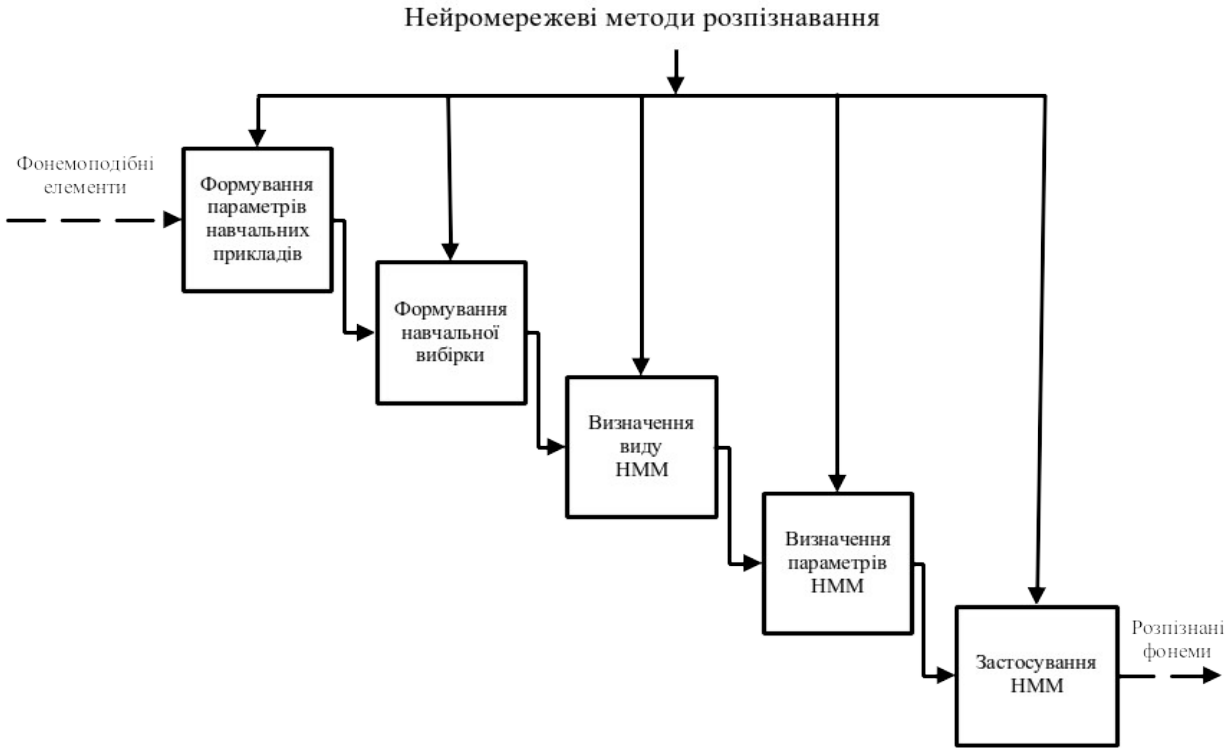

Рис. 2.1. Діаграма декомпозиції нейромережевого розпізнавання фонем

- Визначення виду та параметрів НММ – вибір для застосування такого виду НММ, з такими параметрами, що найбільш повно відповідають умовам задачі розпізнавання фонем в ГС конкретної КС.

- Застосування НММ – розпізнавання фонем в ГС. Слід враховувати,

що застосування НММ призводить до навантаження КС і може призвести до вичерпання її обчислювальних ресурсів.

Наступним етапом створення концептуальної моделі стала розробка показаної на рис. 2.2 схеми компонентів НМС розпізнавання фонем. В схемі враховані особливості реалізації НМС в умовах КС:

- недосконаліть методик формування параметрів навчальних прикладів для НММ, що призначені для розпізнавання фонем;
- тривалий термін формування навчальної вибірки для НММ у випадку обмеженого доступу до БД фонем;
- складність доступу до існуючих БД фонем;
- додаткове навантаження на КС за рахунок функціонування засобів розпізнавання.

Тому в схемі передбачена можливість формування параметрів навчальних прикладів та навчальної вибірки за допомогою експертних даних.

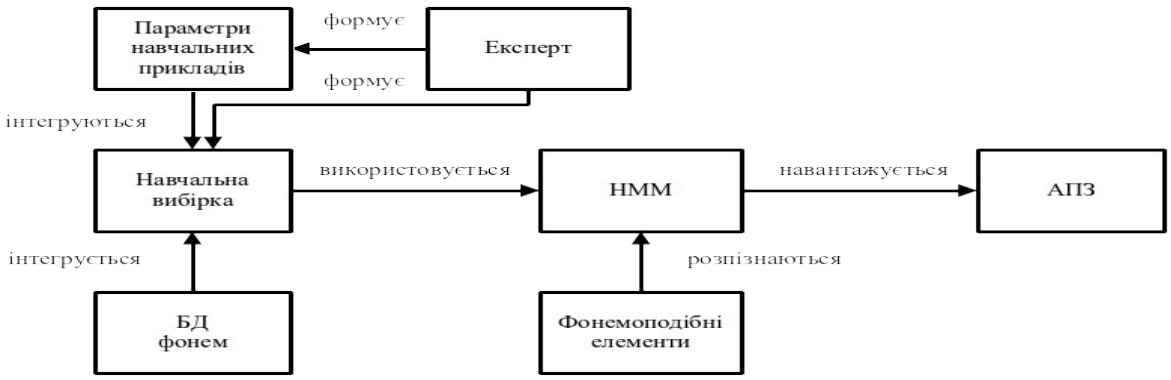

Рис. 2.2. Схема взаємодії компонентів НМС розпізнавання фонем в ГС

Аналіз даних, показаних на рис. 2.1 та рис. 2.2, дозволяє стверджувати, що на ефективність нейромережевого розпізнавання фонем впливають ряд факторів, показаних на рис. 2.3.

Крім того, можливо стверджувати, що ефективність нейромережевого розпізнавання доцільно оцінювати з точки зору ефективності процесу застосування НМЗ та з точки зору навчання НМЗ. Показники ефективності

повинні відображати тривалість, ресурсоємність та точність вказаних процесів. Таким чином, обґрунтовано показані на рис. 2.4 показники оцінки ефективності нейромережевого розпізнавання фонем.

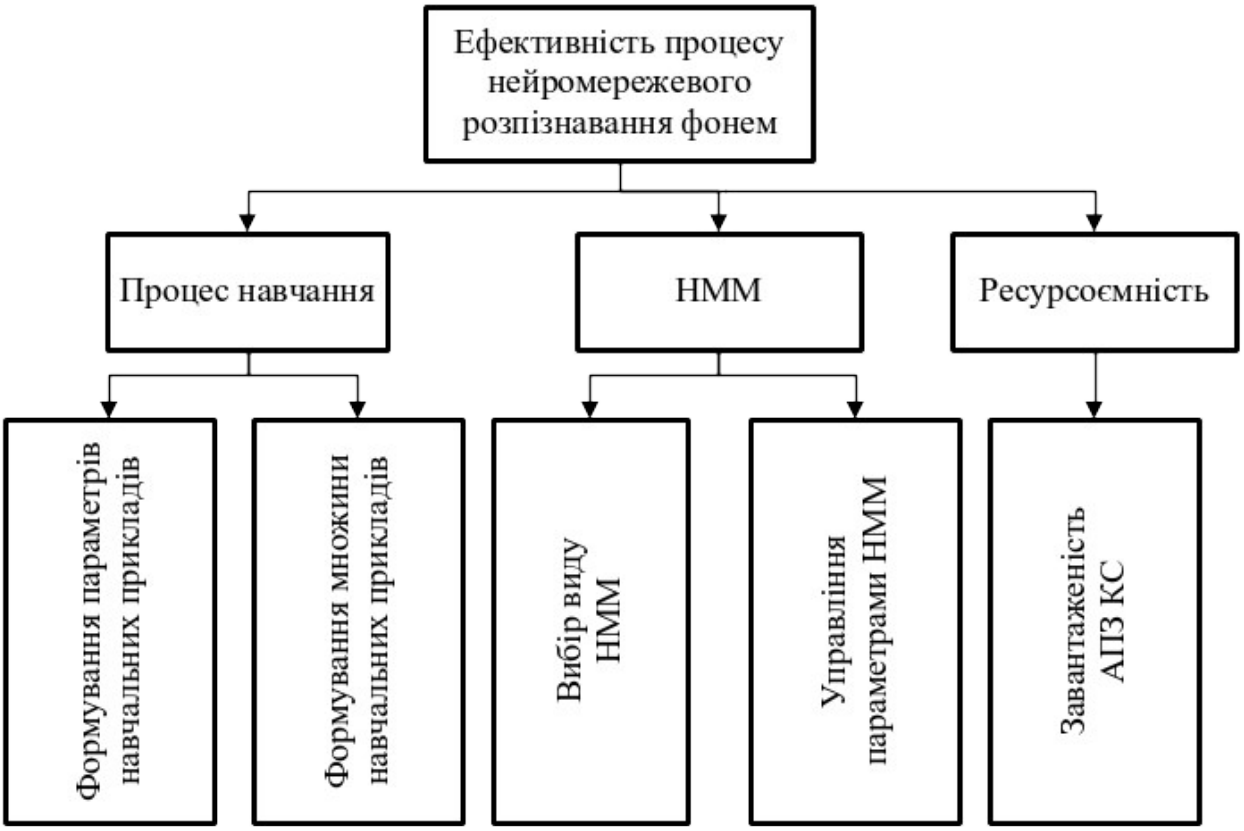

Рис. 2.3. Фактори впливу на ефективність розпізнавання

В підсумку визначено, що в аналітичному вигляді концептуальну модель забезпечення ефективності процесу нейромережевого розпізнавання фонем можливо відобразити за допомогою наступних виразів:

$$E_{\Sigma} = f(E_{НМЗ}, E_{НВ}), \tag{2.1}$$

$$E_{НМЗ} = f(e_1, e_2, e_3), \tag{2.2}$$

$$E_{НВ} = f(e_4, e_5), \tag{2.3}$$

де E_{Σ} – інтегральна ефективність процесу; $E_{НМЗ}$ – ефективність створення та застосування НМЗ; $E_{НВ}$ – ефективність створення навчальної вибірки; e_1 – визначення ефективних видів НММ; e_2 – визначення параметрів НММ, e_3 – ресурсоємність застосування НМЗ; e_4 – визначення параметрів навчальних прикладів; e_5 – формування навчальної вибірки.

Аналіз розробленої концептуальної моделі дозволяє стверджувати, що для ефективного застосування НММ для розпізнавання фонем в ГС в КС необхідно доповнити методологічну базу наступними принципами:

- допустимості застосування виду НММ,
- визначення множини ефективних видів НММ,
- оцінювання ефективності виду НММ,
- визначення очікуваного вихідного сигналу для еталонів фонем,
- прогнозу використання НМС розпізнавання фонем обчислювальних ресурсів КС,
- оцінки ефективності НМЗ,
- використання експертних знань для формування навчальної вибірки.

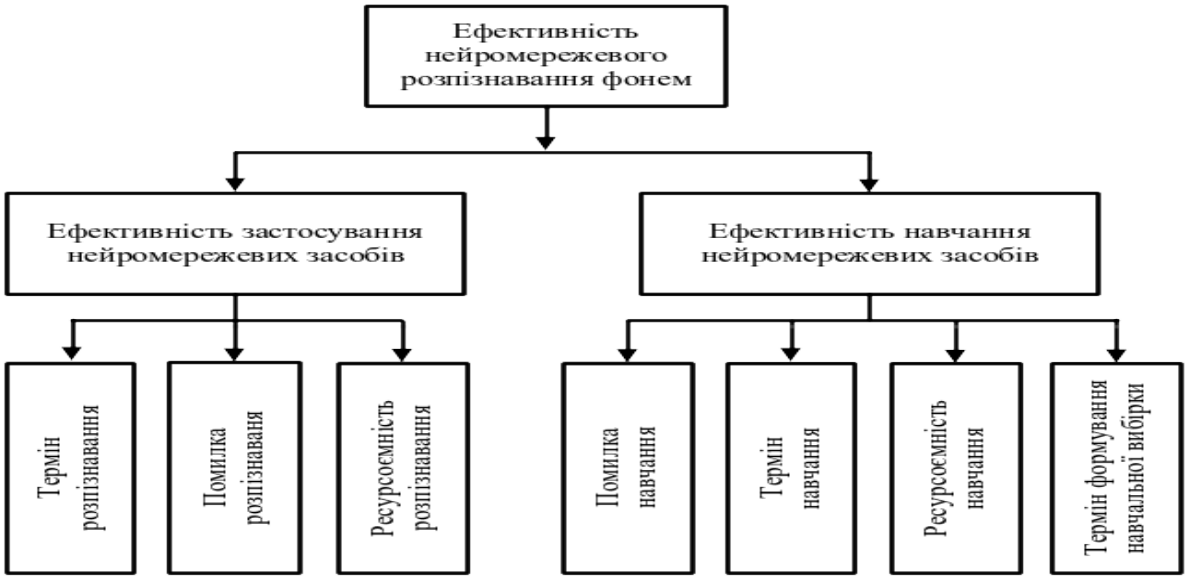

Рис. 2.4. Показники оцінки ефективності нейромережевого розпізнавання

2.2. Принципи застосування нейронних мереж

Принцип визначення множини ефективних видів НММ для

розпізнавання фонем в ГС.

Відповідно результатів [28, 32, 35], визначення множини видів НММ, що забезпечують ефективне розпізнавання фонем в ГС в КС, представлено таким чином:

$$Net_a \rightarrow Net_d \rightarrow Net_e, \tag{2.4}$$

де Net_a – множина доступних видів НММ; Net_d – множина допустимих видів НММ; Net_e – множина ефективних видів НММ.

Принцип допустимості застосування виду НММ для розпізнавання фонем в ГС КС.

Як свідчать результати [28, 35], основним фактором, який впливає на формування множини допустимих видів НММ, що використовуються для розпізнавання фонем в ГС в КС є забезпечення ефективного навчання НММ.

Для цього необхідно у прийнятний термін виконати такі процедури:

- визначити та провести кодування вхідних та вихідних параметрів НММ,
- створити навчальну вибірку,
- реалізувати процес навчання.

Виконання першої процедури реалізується на підготовчому етапі розробки НМЗ. Тому увагу звернено на виконання другої та третьої процедур. Прийнятний термін створення навчальної вибірки та навчання НМ визначається на основі вимог до створення ресурсів КС. Тобто:

$$t_\Sigma \le t_d, \tag{2.5}$$

де t_Σ – загальний термін навчання НМ розпізнавання фонем в ГС; t_d – прийнятний термін створення НМЗ розпізнавання фонем в ГС.

Таким чином, принцип допустимості застосування *i*-го виду НММ для розпізнавання фонем в ГС КС задається наступним правилом:

$$Якщо\ t_\Sigma(net_i) \le t_d \rightarrow net_i \in Net_d, \tag{2.6}$$

де net_i – *i*-ий вид НММ; Net_d – множина допустимих видів НММ.

Принцип оцінювання ефективності виду НММ, призначеної для розпізнавання фонем в ГС КС.

По аналогії з [20] вважається, що серед множини допустимих *i*-ий вид НММ є найбільш ефективним, якщо для нього функція ефективності

приймає максимальне значення:

$$\max_{V_i} = \{ V_1, V_2, .. V_I \}, \quad (2.7)$$

де I – кількість видів НММ; V_i – функція ефективності *i*-го виду НММ.

Розрахунок функції ефективності виконується так:

$$V_i = \sum_{k=1}^{K} \alpha_k R_k(net_i), \quad net_i \in \mathrm{Net}_d, \; i = 1,...I, \quad (2.8)$$

де $\alpha_k = [0..1]$ – ваговий коефіцієнт *k*–го критерію ефективності; net_i – *i*-ий вид НММ; $\mathbf{Net}_d$ – множина допустимих видів НММ; K – кількість критеріїв ефективності; $R_k(net_i)$ – значення *k*-го критерію для НММ *i*-го виду.

Відповідно результатів [20, 28, 35], критерії вибору найбільш ефективного виду НММ повинні відображати міру її пристосованості до поставленої прикладної задачі. Таким чином, під *k*-им критерієм визначення найбільш ефективного виду НММ будемо розуміти міру забезпечення в НММ *k*-ої вимоги задачі розпізнавання фонем в ГС в КС.

Прикладом такого критерію є міра забезпечення в НММ вимоги до безітераційного навчання. Очевидно, що для визначення критеріїв ефективності необхідно провести дослідження визначених в першому розділі найбільш перспективних видів НММ.

Принцип визначення очікуваного вихідного сигналу НММ для еталонів фонем.

Значення вихідного сигналу повинно відображати схожість начальних прикладів. Інакше навчання НММ може ускладнитись. Тому вихідний сигнал для еталонів фонем пропонується описати виразом:

$$\boldsymbol{Y_Ф} = f(\boldsymbol{d}_Ф), \quad (2.9)$$

де $\boldsymbol{Y_Ф}$ – очікувані вихідні сигнали НММ для еталонів фонем $\boldsymbol{Ф}$; $\boldsymbol{d}_Ф$ – множина мір схожості між компонентами $\boldsymbol{Ф}$.

Оскільки передбачено розпізнавати фонеми на основі аналізу їх акустичних характеристик, то пропонується, щоб ці характеристики були відображені в мірі схожості фонем між собою.

Принцип прогнозу використання НМС розпізнавання фонем обчислювальних ресурсів КС.

Відповідно результатів [18, 20], навантаження на модуль розпізнавання КС за рахунок використання НМЗ розпізнавання фонем безпосередньо залежить від кількості звернень до цієї системи. Подібні процеси ефективно описуються за допомогою теорії вейвлет-перетворень. Тому прогнозувати навантаження на модуль розпізнавання КС пропонується за допомогою вейвлет-моделі зміни кількості звернень:

$$Q_{web} = f(W), \tag{2.10}$$

де Q_{web} – навантаження на модуль розпізнавання КС; W – вейвлет-модель зміни кількості звернень до модулю розпізнавання КС.

Принцип оцінки ефективності НМЗ розпізнавання фонем.

По аналогії з [20] оцінювати ефективність НМЗ можливо на основі множини параметрів, що вказують на ступінь забезпечення в них виконання процедур, які вважаються необхідними для ефективного розпізнавання фонем в ГС в КС:

$$E_{НМЗ} = f(\boldsymbol{П}), \tag{2.11}$$

де $E_{НМЗ}$ – ефективність НМЗ; $\boldsymbol{П}$ – множина запропонованих параметрів.

Принцип використання експертних знань для формування навчальної вибірки.

По аналогії з [41, 45], даний принцип передбачає, що навчальні приклади можливо формувати на основі експертних знань стосовно фонем у вигляді продукційних правил виду:

$$Якщо\ x_1 \in \left\lfloor X_1^{min}, X_1^{max} \right\rfloor_I \wedge \ \ldots \wedge \ x_K \in \left\lfloor X_K^{min}, X_K^{max} \right\rfloor_I \rightarrow Y, \tag{2.12}$$

де $x_1, \ldots x_K$ – параметри, що ідентифікують фонему; $\left\lfloor X_1^{min}, X_1^{max} \right\rfloor, \ldots, \left\lfloor X_K^{min}, X_K^{max} \right\rfloor$ – задані діапазони $x_1, \ldots x_K$; K – кількість ідентифікуючих параметрів; Y – результат продукційного правила (фонема).

Розроблені принципи стали основою для створення моделей процесів застосування НМЗ для розпізнавання фонем.

2.3. Модель правил визначення ефективних видів нейромережевих моделей

Деталізувавши вираз (2.6) отримаємо:

$$t_{\Sigma}(net_i) = t_v + t_l(net_i), \tag{2.13}$$

де t_v – термін створення навчальної вибірки; $t_l(net_i)$ – тривалість навчання для *i*-го виду НММ.

Зазначимо, що з точки зору визначення принципової можливості застосування певного виду НММ, створення навчальної вибірки означає формування такої мінімальної кількості навчальних прикладів, яка теоретично вважається достатньою для якісного навчання НММ. Відповідно даних [35], ця кількість в першому наближенні розраховується так:

$$P_{min} \approx 10 N_x, \tag{2.14}$$

де P_{min} – мінімально допустима кількість навчальних прикладів; N_x – кількість вхідних параметрів НММ.

Також можна прийняти, що:

$$t_v = \bar{t}_v P_{min}, \tag{2.15}$$

де $\bar{t}_v$ – середній термін створення одного навчального прикладу.

Очевидно, що величина $\bar{t}_v$ є індивідуальною для конкретної КС і залежить від багатьох факторів. Визначити величину $\bar{t}_v$ можливо шляхом експертного оцінювання. Після підстановки (2.14) в (2.15) отримаємо:

$$t_v = 10 \bar{t}_v N_x. \tag{2.16}$$

В свою чергу, відповідно результатів [35, 56], для приблизної оцінки терміну визначення вагових коефіцієнтів синаптичних зв'язків НММ необхідно врахувати обсяг навчальних прикладів, кількість вхідних параметрів, кількість вихідних параметрів, допустиму величину помилки навчання, швидкодію апаратно-програмної реалізації та вид НММ. При визначеній структурі для НММ *i*-го виду тривалість процесу визначення вагових коефіцієнтів можливо оцінити так:

$$t_l(net_i) = \tau \times L_i \times W_i \times K_{o,i}, \tag{2.17}$$

де τ – тривалість навчальної ітерації для синаптичного зв'язку; W_i – кількість синаптичних зв'язків НММ *i*-го виду; L_i – кількість нейронів; $K_{o,i}$ – кількість ітерацій в процесі навчання.

При цьому

$$K_{o,i} = f_{o,i}(\varepsilon, P), \tag{2.18}$$

де $f_{o,i}(\varepsilon, P)$ – залежність кількості ітерацій від похибки навчання та

кількості навчальних прикладів для НММ і-го виду; ε – допустима похибка навчання НМ; P – кількість навчальних прикладів.

Відповідно [46, 50], при приблизних розрахунках для множини видів НММ $\boldsymbol{net_1}$, яка складається із НММ на основі PNN, мереж адаптивної резонансної теорії, ТК, МЗР, РБФ, машини Больцмана, семантичних та асоціативних НМ, що навчаються шляхом безпосереднього запам'ятовування навчальних прикладів або на основі правил Гебба, Ойя, корелятивного або дельта-правил для оцінки терміну навчання можливо використовувати вираз (2.19). Для оцінки терміну навчання множини видів НММ $\boldsymbol{net_2}$, яка складається із НММ, що базуються на БШП та навчаються за допомогою градієнтних методів або генетичних алгоритмів, можливо використовувати вираз (2.20).

$$t_l(\boldsymbol{net_1}) \approx k_1 \tau e^{-\varepsilon} P(N_x + N_y), \tag{2.19}$$

$$t_l(\boldsymbol{net_2}) \approx k_2 \tau e^{-\chi\varepsilon} P^2 (N_x + N_y)^2, \tag{2.20}$$

де $t_l(\boldsymbol{net}_1)$ – термін визначення вагових коефіцієнтів для $\boldsymbol{net}_1$, $t_l(\boldsymbol{net}_2)$ – термін визначення вагових коефіцієнтів для $\boldsymbol{net}_2$, k_1 – коефіцієнт пропорційності для НММ, що належать до $\boldsymbol{net}_1$, τ – тривалість однієї обчислювальної операції процесу навчання, P – кількість навчальних прикладів, N_y – кількість вихідних параметрів, k_2 – коефіцієнт пропорційності для НММ, що належать до $\boldsymbol{net}_2$, χ – емпіричний коефіцієнт.

Як свідчать попередні результати монографії, з точки зору розпізнавання ГС найбільш перспективними видами НММ являються РБФ, ТК, МЗР, БШП та ГНМ. Для РБФ, ТК та МЗР приблизний термін навчання можливо розрахувати за допомогою виразу (2.19), а для БШП та ГНМ слід використати (2.20). Величини τ та ε можна визначити шляхом експертного оцінювання. При визначенні принципової можливості застосування НММ доцільно орієнтуватись на мінімально допустиму кількість навчальних прикладів. Врахувавши в (2.19) та (2.20) залежність (2.14), отримаємо:

$$t_l(\boldsymbol{net_1}) \approx 10 k_1 \tau e^{-\varepsilon} N_x (N_x + N_y), \tag{2.21}$$

$$t_l(\boldsymbol{net_2}) \approx 100 k_2 \tau e^{-\chi\varepsilon} N_x^2 (N_x + N_y)^2. \tag{2.22}$$

Підставивши (2.16, 2.21) та (2.16, 2.22) в (2.13) та провівши тривіальні спрощення, отримаємо:

$$t_{\Sigma}(\boldsymbol{net}_1) = 10N_x\left(\bar{t}_v + k_1\tau e^{-\varepsilon}\left(N_x + N_y\right)\right), \quad (2.23)$$

$$t_{\Sigma}(\boldsymbol{net}_2) = 10N_x\left(\bar{t}_v + 10k_2\tau e^{-\chi\varepsilon}N_x\left(N_x + N_y\right)\right), \quad (2.24)$$

де $t_{\Sigma}(\boldsymbol{net}_1)$ – термін навчання для НММ, що входять до $\boldsymbol{net}_1$; $t_{\Sigma}(\boldsymbol{net}_2)$ – термін навчання для НММ, що входять до $\boldsymbol{net}_2$.

Результати теоретичних робіт, присвячених НМ [2, 44, 50], дозволяють стверджувати, що $k_1 \approx 0{,}1$, $k_2 \approx 0{,}001$, $\chi \approx 1$, $\varepsilon \approx 0{,}01$. Підставивши вказані величини у (2.23, 2.24), отримаємо:

$$t_{\Sigma}(\boldsymbol{net}_1) \approx 10N_x\left(\bar{t}_v + 0{,}1\tau\left(N_x + N_y\right)\right), \quad (2.25)$$

$$t_{\Sigma}(\boldsymbol{net}_2) \approx 10N_x\left(\bar{t}_v + 0{,}01\tau N_x\left(N_x + N_y\right)\right). \quad (2.26)$$

Множина вхідних параметрів НММ може формуватись на основі вектору ознак (мел-кепстральних коефіцієнтів) квазістаціонарного фрагменту ГС. Оскільки тривалість однієї фонеми може складати $t_{\Phi} = 60$ мс [3, 5], а тривалість квазістаціонарного фрагменту $t_{cm} = 12$ мс, то t_{Φ} слід розділити на стаціонарні ділянки, які перекриваються між собою. При половинному перекритті стаціонарних ділянок, що відповідають одній фонемі, отримаємо:

$$n_{cm} = 2 \times t_{\Phi} / t_{cm} - 1 = 2 \times 60 / 12 - 1 = 9, \quad (2.27)$$

де n_{cm} – кількість стаціонарних ділянок.

Через те, що одну стаціонарну ділянку характеризує від 6 до 24 мел-кепстральних коефіцієнта [4], кількість таких коефіцієнтів дорівнює:

$$K_{mel} = (6..24) \times n_{cm} = (6..24) \times 9 = 54..216. \quad (2.28)$$

Враховуючи, що на вхід НММ, крім мел-кепстральних коефіцієнтів, можуть подаватись і інші вхідні дані, в першому наближенні будемо вважати, що кількість вхідних параметрів НМ $N_x = 60..220$. Базуючись на результатах [2, 8], вважатимемо, що для розпізнавання фонем в НММ достатньо одного вихідного нейрону. Оскільки (2.28) має приблизний характер, то $N_x + N_y \approx N_x \approx 220$. Це дозволяє модифікувати (2.25, 2.26):

$$t_{\Sigma}(\boldsymbol{net}_1) \approx 2200\left(\bar{t}_v + 22\tau\right), \quad (2.29)$$

$$t_{\Sigma}(\boldsymbol{net}_2) \approx 2200\left(\bar{t}_v + 484\tau\right). \quad (2.30)$$

Оскільки $t_{\Sigma}(\boldsymbol{net}_2) > t_{\Sigma}(\boldsymbol{net}_1)$, то з врахуванням (2.159, 2.160) правило визначення допустимості використання виду НММ для розпізнавання фонем в ГС в КС (2.6) можливо деталізувати за допомогою наступних двох виразів:

$$Якщо\ 2200(\bar{t}_v + 22\tau) \le t_d \rightarrow \boldsymbol{net}_1 \in \boldsymbol{Net}, \quad (2.31)$$

$$Якщо\ 2200(\bar{t}_v + 484\tau) \le t_d \rightarrow \boldsymbol{Net} = \{\boldsymbol{net}_1, \boldsymbol{net}_2\}. \quad (2.32)$$

Правило (2.31) визначає допустимість використовувати НММ на основі ТК, МЗР, РБФ, PNN, мереж адаптивної резонансної теорії, машини Больцмана семантичних та асоціативних НМ. Правило (2.32) доповнює допустиму множину НММ моделями на основі БШП та ГНМ.

Розглянемо оцінку прийнятного терміну створення НММ розпізнавання фонем в КС. Розробка НММ являється однією із складових процесу створення модулю розпізнавання в КС. Тому

$$t_d = k_{nmz} \times t_{max}, \quad (2.33)$$

де t_{max} – максимальний термін розробки модулю розпізнавання ГС в КС; k_{nmz} – коефіцієнт, що визначає пропорцію між t_d та t_{max}. В першому наближенні можна вважати, що:

$$k_{nmz} \approx 0{,}5. \quad (2.34)$$

Враховуючи (2.34) та практичний досвід щодо оновлення АПЗ КС, отримаємо

$$t_d \approx 0{,}5 \times 1\ рік = 1{,}5 \times 10^7\ с. \quad (2.35)$$

Використання виразу (2.35) дозволяє записати (2.31, 2.32) так:

$$Якщо\ 2200(\bar{t}_v + 22\tau) \le 1{,}5 \times 10^7 \rightarrow \boldsymbol{net}_1 \in \boldsymbol{Net}, \quad (2.36)$$

$$Якщо\ 2200(\bar{t}_v + 484\tau) \le 1{,}5 \times 10^7 \rightarrow \boldsymbol{Net} = \{\boldsymbol{net}_1, \boldsymbol{net}_2\}. \quad (2.37)$$

Вирази (2.36) та (2.37) є правилами визначення допустимих видів НММ в умовах типової КС. Застосування цих правил до множини доступних НММ дозволяє перейти до визначення ефективних НММ. Для цього, застосувавши принцип оцінювання ефективності виду НММ, призначеної для розпізнавання фонем в ГС КС (2.7, 2.8), проведена розробка критеріїв ефективності.

Відповідно [8, 16, 20, 28], з позицій задачі дослідження вимоги до НММ можливо розділити на групи, що характеризують:

- навчання,
- інтелектуальні можливості НМ,
- процес прийняття рішення.

Розглянемо вимоги кожної із груп.

1. Вимоги до навчання.

- В навчальних прикладах може бути різна кількість вхідних параметрів. Виконання цієї вимоги дозволяє застосувати НММ до класифікації елементів ГС без їх складної попередньої обробки..

- Мінімальна кількість навчальних прикладів може бути меншою від кількості вхідних параметрів, що дозволяє зменшити тривалість процесу формування навчальної вибірки. Стосовно задачі розпізнавання фонем в ГС виконання даної вимоги означає, що обсяг навчальної вибірки може бути менший, ніж 220 прикладів.

- Можливість не пропорційного представлення в навчальній вибірці даних, що описують класи, які повинна розпізнати НММ. Забезпечення даної вимоги дозволяє спростити процес формування навчальної вибірки фонем.

- Необхідність представлення в навчальній вибірці очікуваного вихідного сигналу НММ. Стосовно розпізнавання фонем в ГС, відсутність очікуваного вихідного сигналу означає відмову від складної процедури співставлення виділеного голосового фрагменту з відповідною фонемою.

- Відсутність в навчальній вибірці зашумлених та корельованих прикладів. Дана вимога пов’язана із забезпеченням однорідності класів, що мають бути розпізнані НММ. Її виконання дозволяє зменшити вимоги до якості формування навчальних прикладів фонем.

- Пристосованість до донавчання без втрати інформації, що була узагальнена. Забезпечення цієї вимоги дозволяє оперативно адаптувати НММ до нових умов застосування.

- Пристосованість до навчання окремими частинами, що визначає можливість розпаралелювання цього процесу при реалізації НММ.

- Забезпечення низької похибки навчання, що розраховується так:

$$\varepsilon = N_{true} / N_{\Sigma}, \qquad (2.38)$$

де N_{true} – кількість правильно розпізнаних навчальних прикладів; N_{Σ} – загальна кількість навчальних прикладів.

Для голосової ідентифікації користувача при вході в КС достатньою є похибка $\varepsilon \le 0{,}01$, а для визначення голосової відповіді в процесі комп’ютерного тестування та голосової команди користувача КС достатньою

є похибка $\varepsilon \leq 0{,}03$ [42, 43].

- Забезпечення мінімального терміну навчання, що в основному визначається кількістю навчальних ітерацій.

- Забезпечення високого рівня автоматизації навчання, що при використанні «однорідної» навчальної вибірки в основному залежить від кількості емпірично настроюваних параметрів НММ, які можуть істотно вплинути на його ефективність.

- Використання мінімального обсягу обчислювальних ресурсів, необхідних для реалізації процесу навчання.

- Забезпечення стабільності процесу навчання, що означає гарантоване досягнення заданої помилки навчання НММ за обмежену кількість навчальних ітерацій.

2. Вимоги до інтелектуальних можливостей.

- Максимальне відношення обсягу пам’яті (кількості прикладів) НММ до кількості синаптичних зв’язків. Дана вимога відображає можливість НММ узагальнювати навчальні дані, а не просто їх запам’ятовувати.

- Мінімальна похибка узагальнення, яка показує правильність класифікації на прикладах, які не ввійшли в навчальної вибірку.

3. Вимоги до процесу прийняття рішення.

- Мінімальна тривалість класифікації невідомого прикладу.

- Мінімальний обсяг обчислювальних ресурсів, який необхідний для класифікації невідомого прикладу.

Перелік критеріїв ефективності, що відповідають визначеним вимогам наведено в табл. 2.1

Таблиця 2.1

Критерії ефективності виду НММ

Критерій	**Вимога**
1	*2*
R_1	Варіативність кількості вхідних параметрів в прикладах вибірки
R_2	Обмеження мінімальної кількості навчальних прикладів

R_3	Можливість не пропорційного представлення в навчальній вибірці даних, що відповідають класам, які повинні бути розпізнані
R_4	Можливість використання навчальних прикладів, в яких відсутній очікуваний вихідний сигнал НММ
R_5	Можливість використання в навчальній вибірці зашумлених та корельованих прикладів
R_6	Пристосованість до донавчання
R_7	Пристосованість до навчання окремими частинами
R_8	Забезпечення низької похибки навчання
R_9	Забезпечення не тривалого терміну навчання
R_{10}	Забезпечення високого рівня автоматизації навчання
R_{11}	Мінімальний обсяг обчислювальних ресурсів, необхідних для навчання
R_{12}	Забезпечення стабільності навчання
R_{13}	Максимальне відношення обсягу пам'яті НММ до кількості синаптичних зв'язків
R_{14}	Мінімальна похибка узагальнення

Таблиця 2.1 (продовження)

1	*2*
R_{15}	Мінімальна тривалість класифікації невідомого прикладу
R_{16}	Мінімальний обсяг обчислювальних ресурсів, необхідний для класифікації невідомого прикладу
R_{17}	Апробованість в задачах розпізнавання ГС

Запропоновані критерії ефективності мають безрозмірний характер. Для *i*-го виду НММ значення *k*-го критерію дорівнює 1, якщо відповідна *k*-та вимога повністю забезпечується в даному виді НММ, та дорівнює 0, якщо не забезпечується.

Наприклад, оскільки БШП та РБФ можуть бути навчені тільки методом «з вчителем», то для цих видів НММ $R_4 = 0$. Значення складових множини критеріїв ($\boldsymbol{R}_a$) для доступних видів НММ, що використовуються

для розпізнавання ГС, наведено в табл. 2.2.

Відповідно положення про визначення критеріїв вибору найбільш ефективного виду НММ для розпізнавання фонем в ГС, кількість критеріїв ефективності K у виразі (2.8) дорівнює 17.

Також в першому наближенні будемо вважати:

$$\boldsymbol{Net}_a = \boldsymbol{Net}_d = \{ БШП, ГНМ, ТК, МЗР, РБФ \} . \quad (2.39)$$

Таблиця 2.2

Значення критеріїв ефективності

Критерій	**Вид НММ**				
	БШП	**ГНМ**	**ТК**	**МЗР**	**РБФ**
1	*2*	*3*	*4*	*5*	*6*
R_1	0	0	0	0	0
R_2	0,2	0,2	0,5	0,5	0,7
R_3	0,3	0,5	0,5	0,5	0,7
R_4	0	0	1	0	0
R_5	0,8	0,8	0,3	0,3	0,5

Таблиця 2.2 (продовження)

1	*2*	*3*	*4*	*5*	*6*
R_6	0,2	0,9	0,2	0,2	0,7
R_7	0,1	0,9	0,1	0,1	0,9
R_8	0,9	0,9	0,5	0,9	0,9
R_9	0,5	0,5	0,8	0,9	0,9
R_{10}	0,9	0,9	0,7	0,7	0,7
R_{11}	0,7	0,2	0,8	0,9	0,9
R_{12}	0,8	0,8	0,7	0,9	0,9
R_{13}	0,9	0,9	0,4	0,3	0,5
R_{14}	0,9	0,9	0,4	0,3	0,4
R_{15}	0,7	0,7	0,4	0,3	0,5
R_{16}	0,9	0,9	0,4	0,3	0,5

R_{17}	0,9	0,9	0,5	0,5	0,5

Таким чином, кількість допустимих видів НММ $I = 5$. Це дозволяє трансформувати (2.8) наступним чином:

$$V_i = \sum_{k=1}^{17} \alpha_k R_k(net_i), \quad net_i \in \{БШП, ГНМ, ТК, МЗР, РБФ\}, \; i = 1,...5. \qquad (2.40)$$

Відзначимо, що за допомогою вагового коефіцієнту α_k враховується значущість *k*-го критерію ефективності для конкретної прикладної задачі.

В підсумку запропоновано відображати правило визначення множини ефективних видів НММ за допомогою виразу (2.41). При цьому правило визначення найбільш ефективного виду НММ визначається за допомогою виразу (2.42):

$$Якщо\, V(net) \geq \Delta_V \;\wedge\; net \in \boldsymbol{Net}_d \rightarrow net \in \boldsymbol{Net}_e, \qquad (2.41)$$

$$Якщо \max_{V(net_u)} = \{V(net_1), \ldots V_I\}, net_i \in \boldsymbol{Net}_e \rightarrow net_i = net_e^{max}, \qquad (2.42)$$

де $V(net)$ – ефективність НММ, що розраховується за допомогою (2.40); Δ_V – мінімально допустима ефективність НММ.

2.4. Модель формування параметрів навчальних прикладів

В загальному випадку тривалість фонеми може змінюватись від $t_ф^{max} = 60$ мс до $t_ф^{min}$, що дорівнює тривалості квазістаціонарного фрагменту ГС [5, 47]. Вважається, що тривалість квазістаціонарного фрагменту ГС дорівнює $t_{cm} = 12$ мс.

Надалі виділений фрагмент ГС, який відповідає окремій фонемі, будемо називати фонемоподібним фрагментом.

Як показано на рис. 2.5, аналіз ГС реалізується частинами на квазістаціонарних фрагментах, що мають половинне перекриття.

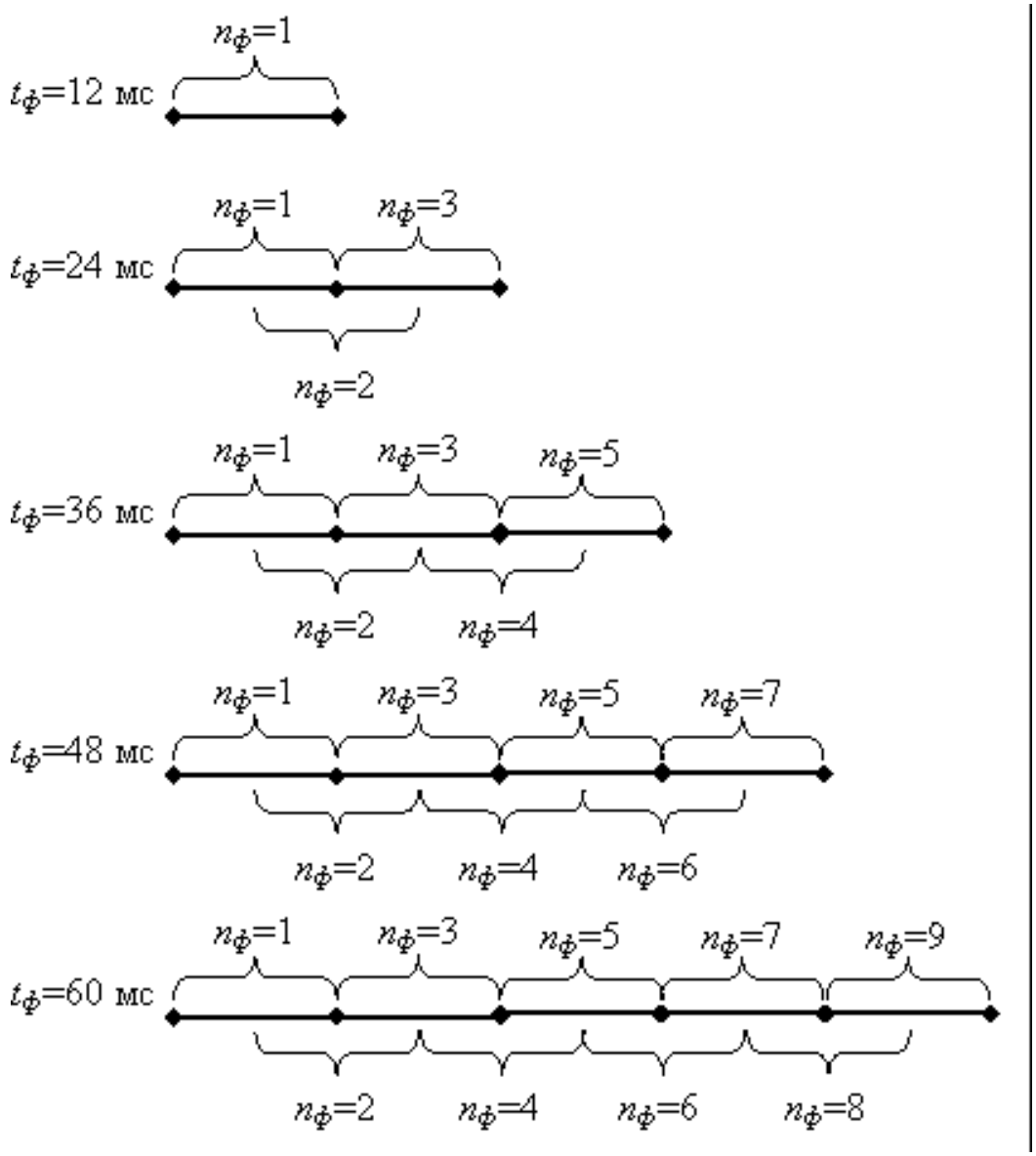

Рис. 2.5. Кількість можливих інтервалів аналізу при розпізнаванні фонем

Тому тривалість фонеми знаходиться в межах:

$$t_\phi = \left[t_\phi^{min} .. t_\phi^{max}, t_{cm} \right] = [12..60, 12]. \quad (2.43)$$

В залежності від тривалості фонеми можлива кількість відповідних квазістаціонарних фрагментів ГС дорівнює:

$$n_\phi = 2\frac{t_\phi}{t_{cm}} - 1. \quad (2.44)$$

Відповідно [6, 10], квазістаціонарному фрагменту відповідає від 6 до 24 мел-кепстральних коефіцієнта. Враховуючи (2.44), кількість таких коефіцієнтів, що будуть відповідати входам НММ, можливо визначити так:

$$N_x = (12..48)\frac{t_\phi}{t_{cm}} - (6..24), \quad (2.45)$$

де N_x – кількість входів НММ.

Підстановка (2.43) в (2.45) показує, що кількість вхідних параметрів НММ знаходиться в інтервалі:

$$N_x \in \left[54...217\right]. \tag{2.46}$$

При цьому відомі НММ не пристосовані до навчальних прикладів з варіативною кількістю вхідних параметрів. Тому безпосередньо використовувати мел-кепстральні коефіцієнти, що описують ГС, в якості вхідних параметрів НММ недоцільно.

Для виправлення цього недоліку використано результати [12, 28] про можливість якісного розпізнавання меж окремих фонем та принципову можливість нелінійного масштабування фрагменту ГС, що відповідає окремій фонемі. Для розрахунку масштабного коефіцієнту використано результати [38], в яких наведена загальна модель тривалості окремої фонеми:

$$T_j = \left[T_j^{(0)} - k_n(N - j)\right] \times \left[1 - (1 - k_T)\frac{N - j + 1}{N}\right] \times \frac{T}{N}, \tag{2.47}$$

де k_n – коефіцієнт стиснення; k_T – коефіцієнт темпу мовлення; N – кількість фонем у висловлюванні; T – тривалість висловлювання; $T_j^{(0)}$ – власна тривалість *j*-ої фонеми в ударній позиції; j – номер фонеми у висловлюванні. Після адаптації даної моделі до умов поставленої задачі дослідження отримаємо:

$$t_{\phi_j} = \left[t_{\phi_i}^e - k_n(N_\Phi - j)\right] \times \left[1 - (1 - k_T)\frac{N_\Phi - j + 1}{N_\phi}\right] \times \frac{T_{ГС}}{N_\phi},\ i \in \left[1,..K_\phi\right], \tag{2.48}$$

де t_{ϕ_j} – тривалість *j*-го фонемоподібного фрагменту піддослідному ГС; $t_{\phi_i}^e$ – тривалість еталону *i*-ої фонеми; N_ϕ – кількість фонемоподібних фрагментів у піддослідному ГС; $T_{ГС}$ – тривалість піддослідного ГС; j – номер фонемоподібного фрагменту у піддослідному ГС; K_ϕ – кількість фонем, на яку орієнтована система розпізнавання.

В цьому випадку кількість вхідних параметрів НММ буде відповідати кількості мел-кепстральних коефіцієнтів еталону фонеми, а виділений фрагмент ГС буде підлягати процедурі нелінійного масштабування.

При використанні еталону *i*-ої фонеми масштабний коефіцієнт для *j*-го фонемоподібного фрагменту розраховується так:

$$k_{\phi_{j,i}} = \frac{t^e_{\phi_i}}{t_{\phi_j}},\ i \in \left[1,..K_\phi\right]. \tag{2.49}$$

При мінімальній тривалості фонемоподібного фрагменту та максимальній тривалості еталону фонеми масштабний коефіцієнт становить:

$$k_{max} = 5/1 = 5, \tag{2.50}$$

де k_{max} – максимальна величина масштабного коефіцієнту.

При мінімальній тривалості еталону фонеми та максимальній тривалості фонемоподібного фрагменту ГС мінімальна величина масштабного коефіцієнту становить:

$$k_{min} = 1/5 = 0{,}2, \tag{2.51}$$

де k_{min} – мінімальна величина масштабного коефіцієнту.

Оскільки при розпізнаванні фонем в ГС t_{ϕ_j} та t_{ϕ_i} можуть змінюватись тільки дискретно, то і зміна масштабного коефіцієнту також носить дискретний характер. Остаточно функціонал діапазону змін масштабного коефіцієнту можливо описати таким чином:

$$Якщо\ t^e_{\phi_i} > t_{\phi_j} \rightarrow k_{\phi_{j,i}} \in \left]1..5\right], \tag{2.52}$$

$$Якщо\ t_{\phi_j} > t^e_{\phi_i} \rightarrow k_{\phi_{j,i}} \in \left[0{,}2..1\right[. \tag{2.53}$$

Процедуру масштабування виділеного фонемоподібного фрагменту можливо представити у вигляді наступного виразу:

$$\boldsymbol{y}_I \rightarrow \boldsymbol{Y}_J, \tag{2.54}$$

де $\boldsymbol{y}_I = \left| y_0, y_1, ... y_I \right|$ – вектор, що відповідає початковому фонемоподібному фрагменту, заданому за допомогою I відліків; $\boldsymbol{Y}_J = \left| Y_0, Y_1, ... Y_J \right|$ – вектор, що відповідає масштабованому фонемоподібному фрагменту, заданому за допомогою J відліків.

При цьому $J = k \times I$. При $k_{\phi_{j,i}} > 1$ для реалізації (2.54) необхідно виділити у векторі $\boldsymbol{Y}_J$ вектор опорних відліків $\boldsymbol{q}$ та вектор проміжних відліків $\boldsymbol{g}$. Для цього слід скористатись наступними виразами:

$$q_i = trunc(k \times i),\ i \in \left[0, I\right], \tag{2.55}$$

$$\boldsymbol{g} = \boldsymbol{Y}_J - \boldsymbol{q}, \tag{2.56}$$

де q_i – i-та опорна точка; *trunc* – операція округлення дійсного числа до найближчого цілого; k – масштабний коефіцієнт.

Зазначимо, що кількість опорних відліків дорівнює I, а кількість проміжних відліків дорівнює

$$g = J - I. \tag{2.57}$$

В кожному із опорних відліків необхідно встановити величини компонентів $\boldsymbol{Y}_J$ рівними відповідним компонентам $\boldsymbol{y}_I$:

$$Y_{q_i} = y_i, \tag{2.58}$$

де Y_{q_i} – значення $\boldsymbol{Y}_J$ в q_i-му опорному відліку.

По аналогії з методологією лінійної інтерполяції дискретної функції, для розрахунку компонентів $\boldsymbol{Y}_J$ в довільному g-му проміжному відліку, що знаходиться між q_i та q_{i+1} опорними відліками, скористаємось виразом:

$$Y_g = Y_{q_i} + \frac{Y_{q_{i+1}} - Y_{q_i}}{q_{i+1} - q_i} \times (q_{i+1} - g), \tag{2.59}$$

де Y_g – значення $\boldsymbol{Y}_J$ в g-ій проміжній точці.

У випадку $k_{\phi_{j,i}} < 1$, відповідно методології стиснення дискретного сигналу, компоненти $\boldsymbol{Y}_J$ можливо розрахувати за допомогою виразів:

$$Y_j = \frac{1}{k_{\phi_{j,i}}} \sum_{i=A}^{B} y_i, \; j = 1,\ldots J, \tag{2.60}$$

$$A = 1 + (j - 1) \times k_{\phi_{j,i}}, \tag{2.61}$$

$$B = j \times k_{\phi_{j,i}}. \tag{2.62}$$

Використання процедури масштабування призводить до необхідності додавання до вхідних параметрів НММ параметру, що відповідає величині масштабного коефіцієнту. Таким чином, у навчальних прикладах НММ, призначеної для розпізнавання деякої фонеми, кількість вхідних параметрів буде на одиницю перевищувати кількість мел-кепстральних коефіцієнтів, що характеризують еталон даної фонеми.

Розрахунок конкретних значень вхідних параметрів можливо реалізувати з використанням виразів (2.48, 2.51-2.62).

Крім того, ці дослідження вказують на доцільність використання в системі розпізнавання фонем декількох НММ. Мінімальна кількість таких НММ повинна дорівнювати кількості фонем, еталони яких мають однакову тривалість. Максимальна кількість НММ дорівнює кількості еталонів фонем, які мають бути розпізнані.

Розглянемо визначення очікуваного вихідного сигналу навчальних прикладів, що відповідають еталонам фонем. Скористаємось сформульованим принципом про необхідність врахування у очікуваному вихідному сигналі близькості фонем. Можливі два випадки формування вихідного сигналу. У першому випадку:

$$N_y = 1, \tag{2.63}$$

де N_y – кількість нейронів у вихідному шарі.

У другому випадку кількість нейронів у вихідному шарі дорівнює кількості фонем, які мають бути розпізнані

$$N_y = K_\phi, \tag{2.64}$$

де K_ϕ – кількість фонем, що мають бути розпізнані.

Деталізуємо перший випадок. При аналізі ГС кожній з фонем ставиться у відповідність деякий діапазон величин даного сигналу. В базовому випадку величини діапазонів для різних фонем однакові.

Крім того, для навчальних прикладів, які відповідають еталонам фонем, вихідний сигнал повинен дорівнювати середині вказаного діапазону. При використанні сигмоїдальної функції активації вихідний сигнал знаходиться в межах від 0 до 1:

$$y \in]0..1[. \tag{2.65}$$

За умови рівномірного квантування діапазону можливих значень y очікуваний вихідний сигнал для еталону довільної *i*-ої фонеми розраховується наступним чином:

$$y_{\phi_i} = \frac{1}{K_\phi} \times i - \frac{0{,}5}{K_\phi} = \frac{i - 0{,}5}{K_\phi}, \tag{2.66}$$

де i – номер фонеми.

Схожість фонем у (2.66) можливо врахувати тільки за рахунок того, що схожі фонеми повинні мати близькі номери.

Деталізуємо другий випадок. У навчальному прикладі для еталону *i*-ої фонеми вихідний сигнал відповідного *i*-го нейрону дорівнює 1. При цьому порядок нумерації вихідних нейронів може бути довільний. Разом з тим виникає необхідність визначення очікуваного вихідного сигналу для всіх інших вихідних нейронів, що не відповідають даному еталону.

Зазначимо, що нейрони, які відповідають фонемам, близьким до

еталону, повинні мати схожі величини вихідного сигналу.

По суті, визначення очікуваного вихідного сигналу НММ для другого випадку є дещо ускладненим варіантом першого випадку, який зводиться до розрахунку числової оцінки близькості фонем. При цьому відомі аналітичні методи такого розрахунку достатньо складні та погано описані.

В той же час, аналіз та розпізнавання фонем в ГС – це задачі, які достатньо ефективно вирішуються експертом. Тому представляється доцільним визначати числову оцінку міри схожості акустичних характеристик фонем на основі експертних даних.

Базуючись на результатах [1, 18, 28], пропонується використовувати статистичні методи обробки експертних даних. Для цього рекомендується, щоб кількість експертів була не меншою 10.

Процедура експертного оцінювання близькості фонем. Нехай в результаті опитування експертної групи, що складається з M членів, отримано наступні дані:

$$\begin{matrix} x_{1,1}, & x_{n,1}, & x_{N,1} \\ x_{1,m}, & x_{n,m}, & x_{N,m} \\ x_{1,M}, & x_{n,M}, & x_{N,M} \end{matrix}, \qquad (2.67)$$

де $x_{n,m}$ – оцінка міри схожості n-го об'єкту (фонеми) m-им експертом; N – кількість об'єктів (фонем).

Середня колективна оцінка n-ої фонеми знаходиться по формулі:

$$\bar{x}_n = \frac{1}{M} \times \sum_{m=1}^{M} x_{n,m}, \qquad (2.68)$$

де $x_{n,m}$ – оцінка міри схожості n-ої фонеми m-им експертом, n =1…N.

Дисперсія середньої колективної оцінки визначається так:

$$\sigma^2 = \frac{1}{M-1} \times \sum_{m=1}^{M} (x_{n,m} - \bar{x}_n)^2. \qquad (2.69)$$

Для визначення статистичної значимості отриманих результатів необхідно вказати інтервал довіри.

Задавшись ймовірністю помилки P_n (рівнем значимості) можливо визначити інтервал, в який величина, що оцінюється, попадає з ймовірністю $(1 - P_n)$:

$$I_{x_n} = (\bar{x}_n - \varepsilon_{pn}, \bar{x}_n + \varepsilon_{pn}). \qquad (2.70)$$

Величина ε_{pn} визначає границі інтервалу довіри та розраховується так:

$$\varepsilon_{pn} = t_p \times \frac{\sigma_n}{\sqrt{M}}, \tag{2.71}$$

де t_P – коефіцієнт, який залежить від заданої ймовірності довіри *P*.

Вважається, що величина, що оцінюється, розподілена нормально з центром x_i та дисперсією σ. Коефіцієнт t_P має розподілення Ст'юдента з $(N - 1)$ ступенями свободи.

Ступінь узгодженості експертів визначається за допомогою коефіцієнта варіації виду:

$$\gamma_n = \frac{\sigma_n}{\bar{x}_n}. \tag{2.72}$$

Вважається, що узгодженість думок експертів задовільна, якщо всі $\gamma_n < 0{,}3$ та достатня, якщо всі $\gamma_n < 0{,}2$.

Оцінка компетентності експертів може виконуватися по об'єктивному коефіцієнту компетентності та коефіцієнту відносної самооцінки експерта.

Коефіцієнти компетентності дозволяють скорегувати колективну оцінку експертів. При цьому середня колективна оцінка об'єкту буде дорівнювати:

$$\bar{x}_n = \frac{1}{M} \times \sum_{m=1}^{M} \lambda_m x_{n,m}, \tag{2.73}$$

де λ_m – коефіцієнт компетентності *m*-го експерта.

2.5. Модель прогнозування кількості запитів до модулю розпізнавання

Необхідність розробки даної моделі пояснюється тим, що використання в КС модулю розпізнавання ГС призводить до виникнення додаткового навантаження на КС. Результати [1, 28] вказують на те, що неадекватність розповсюджених моделей прогнозування до складного характеру навантаження модулю розпізнавання ГС в КС можливо виправити за рахунок застосування теорії вейвлет-перетворень.

З урахуванням запропонованого принципу прогнозування використання обчислювальних ресурсів, побудова вейвлет-моделі зводиться до вибору базисної функції та обробки статистичних даних для розрахунку

вейвлет-коефіцієнтів [1, 28, 48]. Залежність кількості звернень до модулю розпізнавання ГС можливо представити так:

$$s = s(t, \Theta), \tag{2.74}$$

де Θ – вектор випадкових параметрів.

По причині того, що реєстрація статистичних даних відбувається через відповідні інтервали часу, то на кінцевому відрізку їх відносять до фінітних дискретних сигналів. З точки зору вейвлет-аналізу, такий сигнал характеризується енергією та моментами. Енергія сигналу на заданому інтервалі часу $t \in [t_1, t_2]$:

$$E = \int_{t_1}^{t_2} s(t)^2 dt. \tag{2.75}$$

Початковим моментом *v*-го порядку функції $s(t)$ називається вираз виду

$$m_v = m_0^{-1} \int_{-\infty}^{\infty} t^v s(t) dt, \tag{2.76}$$

де m_0 – нульовий момент, який розраховується так:

$$m_0 = \int_{-\infty}^{\infty} s(t) dt. \tag{2.77}$$

Центральним моментом *v*-го порядку функції $s(t)$ називається вираз виду

$$\overset{0}{m}_v = m_0^{-1} \int_{-\infty}^{\infty} (t - m_1)^v s(t) dt, \tag{2.78}$$

де m_1 – перший початковий момент.

Як правило, сигнал можна передбачити у вигляді суми простих коливань, сукупність інтенсивностей яких називається спектром. Іншими словами спектр – це набір чисел, що визначає долю кожного коливання у сигналі. Теорія вейвлет-перетворень дозволяє не тільки визначити спектр сигналу, але й локалізувати частотні характеристики спектру у часі.

Формально інтегральне вейвлет-перетворення функції $f(t) \in L^2(R)$ записується наступним чином:

$$W(a,b) = |a|^{-0,5} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t - b}{a} \right) dt, \tag{2.79}$$

де ψ – базовий вейвлет (базисна функція); * – процедура комплексного спряження; *a* – масштаб вейвлету; *b* – зсув вейвлету.

Базисна функція вейвлету повинна відповідати загальним вимогам:

- Повинен дорівнювати 0 нульовий момент функції s, що розраховується відповідно (2.77).

- Енергія функції (2.75) повинна бути кінцевою, концентруватися всередині деякого фінітного інтервалу $\Delta t=[t_1, t_2]$ та швидко спадати до нуля поза цим інтервалом. Крім цього, для аналізу рядів з поліноміальним трендом в базисних вейвлетах повинні рівнятись нулю центральні моменти v-го порядку. В інженерних задачах широке розповсюдження отримали вейвлет-функції, що побудовані на основі похідних функцій Гауса виду

$$\psi(t)=\frac{(-1)^k}{\sqrt{2\pi}}\frac{\partial^k}{\partial t^k}e^{-0,5t^2}, \tag{2.80}$$

де k – ціле число, $k\geq 1$.

Збільшення k збільшує кількість центральних моментів, що дорівнюють 0, це дозволяє визначити із сигналу інформацію про особливості більш високого порядку. З позицій обробки статистики запитів до КС, це представляє можливість більш детально аналізувати високочастотні складові.

Зазначимо, що підвищена деталізація може негативно вплинути на узагальнюючу сторону аналізу, а також значно підвищити об'єм обчислювальних операцій. В той же час деталізація параметрів вейвлет-моделі повинна відповідати значимим періодам зміни кількості звернень до модулю розпізнавання ГС.

Особливістю дискретного вейвлет-перетворення дискретної функції є використання дискретних значень масштабу та зсуву вейвлета. Вказані величини задаються у вигляді степеневих функцій виду

$$a=a_0^{-m}, \tag{2.81}$$

$$b=k\times a_0^{-m}, \tag{2.82}$$

де m – параметр масштабу; k – параметр зсуву; a_0 – початковий масштаб.

З врахуванням (2.81, 2.82) вираз (2.79) запишемо так:

$$W(m,k)=|a_0|^{0,5m}\int_{-\infty}^{\infty}f(t)\psi^*(a_0^m\times t-k)dt. \tag{2.83}$$

Доволі часто a_0 приймають рівним 2. Таке дискретне вейлет-перетворення називають діадним [28]. Для діадного вейлет-перетворення вирази (2.79-2.83) трансформуються так:

$$a = 2^{-m}, \tag{2.84}$$

$$b = k \times 2^{-m}, \tag{2.85}$$

$$W(m,k) = 2^{0,5m} \int_{-\infty}^{\infty} f(t)\psi^*(2^m \times t - k)dt. \tag{2.86}$$

На початку аналізу вейвлет розміщується в початок сигналу (t=0), перемножується з сигналом, інтегрується на інтервалі свого визначення та нормалізується на $|a_0|^{0,5m}$. Результат обчислення $W(a,b)$ розміщується в точці $(a = a_0^{-m}, b = 0)$ масштабно-часового спектру перетворення. Далі вейвлет масштабу a_0^{-m} зсувається вправо на значення

$$b = k \times a_0^{-m}. \tag{2.87}$$

Після цього процедура перемноження з сигналом, інтегрування на інтервалі свого визначення та нормалізації повторюється. На частотно-часовій площині отримаємо значення, що відповідає $t = b$ и $a = a_0^{-m}$. Процедура повторюється до тих пір, поки вейвлет не досягне кінця сигналу. Для обчислення наступного масштабного рядка значення a дискретно збільшується на деяке значення, що визначається параметром m. Тим самим здійснюється дискретизація масштабно-часової площини.

Максимальне значення масштабу a відповідає тривалості всього ряду даних, що аналізується. Для деталізації самих високих частот сигналу мінімальний розмір вікна вейвлету не повинен перевищувати період найбільш високочастотної гармоніки.

З точки зору проведення дискретного вейвлет-аналізу, статистичні дані мають наступні особливості:

- Обмежені інтервалом $t \in \lfloor t_{min}, t_{max} \rfloor$.
- Реєстрація даних відбувається з визначеною дискретністю Δt.
- Для діадного перетворення кількість точок ряду N повинна дорівнювати

$$N = 2^z, \tag{2.88}$$

де z – ціле число.

З врахуванням цих особливостей (2.83, 2.86) трансформуються так:

$$W(m,k) = a_0^{0,5m} \sum_{i=1}^{N} \left(f(t_i)\psi^*\left(a_0^m \times t_i - k\right)\right), \tag{2.89}$$

$$W(m,k) = 2^{0,5m} \sum_{i=1}^{N} \left(f(t_i) \psi^* \left(2^m \times t_i - k \right) \right), \qquad (2.90)$$

де N – кількість точок ряду; $f(t_i)$ – значення ряду даних у момент часу t_i; t_i – i-ий момент реєстрації.

Зазначимо, що в (2.90) при фіксованому масштабі зсув змінюється в інтервалі від 0 до 2^{z-m}, а максимальний масштаб дорівнює $z - 1$.

Результатом вейвлет-перетворення одномірного ряду динаміки, якому відповідає статистика звернень до модулю розпізнавання ГС, являється множина значень коефіцієнтів W, що визначені у просторі a (масштаб вейвлету) та b (зсув вейвлету).

Таким чином, побудова вейвлет-моделі кількості звернень до модулю розпізнавання ГС зводиться до розрахунку виразу (2.89) або (2.90).

Обернене дискретне вейвлет-перетворення, яке являється відновленням дискретної функції по набору вейвлет-коефіцієнтів, визначається виразом:

$$f(t) = \frac{\pi}{\ln a_0} \sum_{m=0}^{N-1} \sum_{k=0}^{L-1} \psi(t)^* W(m,k), \qquad (2.91)$$

де L – кількість масштабів.

Використання (2.91) дозволяє прогнозувати кількість звернень до модулю розпізнавання ГС.

2.6. Нейромережева модель розпізнавання фонем за допомогою експертних знань

Базуючись на запропонованому принципі використання експертних знань для формування навчальної вибірки, визначено, що відповідно [28], в першому наближенні в якості параметрів фонем, що застосовуються в продукційних правилах виду (2.12), можливо використати амплітудні характеристики перших 7 формант фонемоподібного елементу.

За рахунок такого припущення вираз (2.12) модифікується так:

$$Якщо\ x_1 \in \left[X_1^{min}, X_1^{max} \right]_l \wedge \ \ldots \ \wedge \ x_7 \in \left[X_7^{min}, X_7^{max} \right]_l \rightarrow Y_l, \qquad (2.92)$$

де $x_1, \ldots x_7$ – характеристики перших семи формант ГС; $\left[X_1^{min}, X_1^{max} \right], \ldots, \left[X_7^{min}, X_7^{max} \right]$ – задані діапазони $x_1, \ldots x_7$; l – номер продукційного правила; Y_l – результат продукційного правила (очікувана фонема).

У випадку визначення продукційних правил виразом (2.92) завдання експертів полягає у визначенні меж діапазонів $\left\lfloor X_1^{min}, X_1^{max} \right\rfloor, \ldots, \left\lfloor X_7^{min}, X_7^{max} \right\rfloor$. Кожну із формант можливо характеризувати за допомогою одного мел-кепстрального коефіцієнту.

При цьому аналіз [20, 35, 41] показав, що найбільш повно пристосована для навчання за допомогою продукційних правил HMM типу MPNN. Структура MPNN, що призначена для класифікації фонем [а] та [б], показана на рис. 2.6.

Структурно MPNN складається із 5 нейронних шарів: вхідного (ВШ), фільтрації (ШФ), образів (ШО), додавання (ШД) та вихідного (ШВ). На нейрони ВШ подають ідентифікуючі параметри.

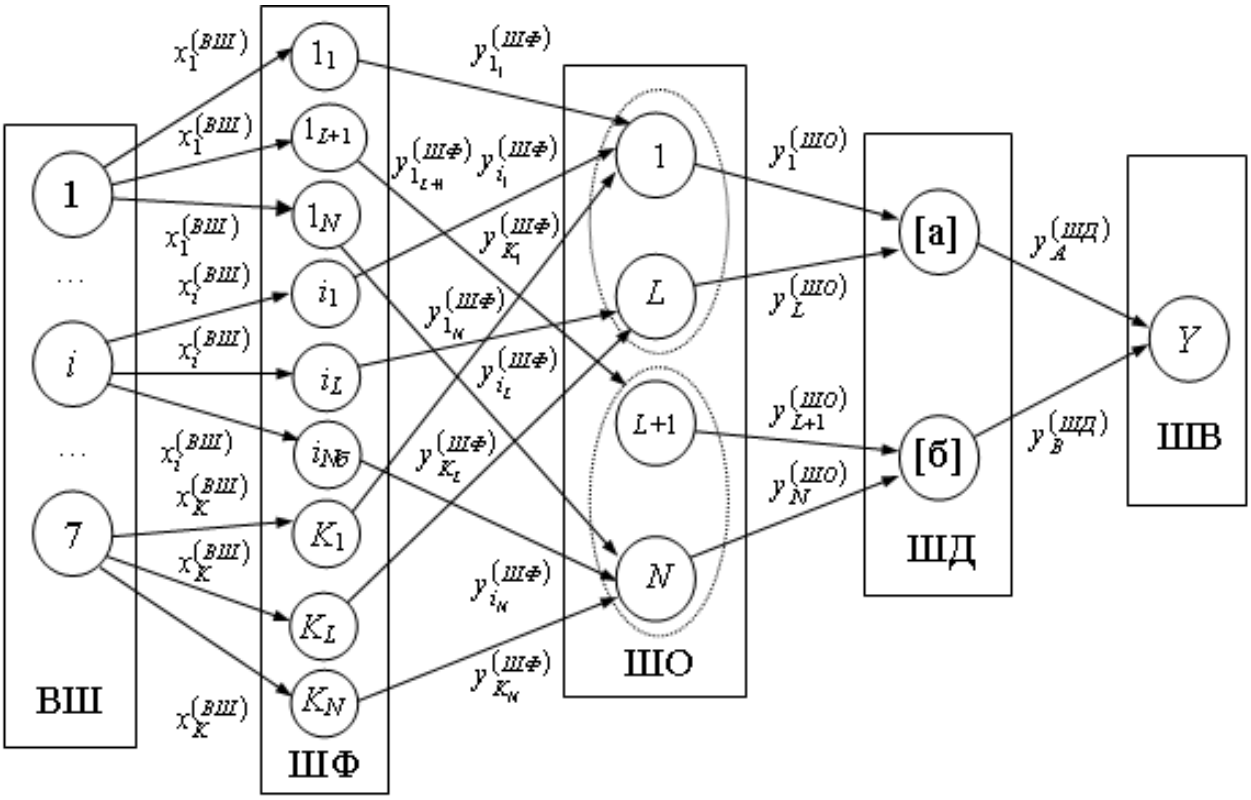

Рис. 2.6. Структура MPNN

Завданням i_l нейрону ШФ є фільтрація i-го вхідного параметру a_i відповідно l–го продукційного правила. Для цього застосовується функція активації виду:

$$\exists x_i^{(ВШ)} \in \left\lfloor X^{min}, X^{max} \right\rfloor_l \rightarrow y_{j_l}^{(ШФ)} = x_i^{(ВШ)}, \exists x_i^{(ВШ)} \notin \left\lfloor X^{min}, X^{max} \right\rfloor \rightarrow y_{j_l}^{(ШФ)} = 0, \quad (2.93)$$

де $x_i^{(ВШ)}$ – значення амплітудної характеристики *i*-ої форманти ГС; $y_{j_l}^{(ШФ)}$ – вихідний сигнал j_l нейрону ШФ.

Вихідний сигнал *l*-го нейрону ШО розраховується так:

$$y_l^{(ШО)} = \sum_{k=1}^{K} exp\left(- (w_{k_l,l} - y_{k_l}^{(ШФ)})^2 / 2\sigma^2\right), \quad (2.94)$$

де $w_{k_l,l}$ – вага зв'язку між k_l-им нейроном ШФ та *l*-им нейроном ШО; K=7 - кількість вхідних параметрів; σ - радіус функції Гауса.

В нейронах ШД використовується лінійна функція активації. Вихідний сигнал *j*-го нейрону ШД ($y_j^{(ШД)}$) розраховується так:

$$y_j^{(ШД)} = \sum_{i=1}^{N} y_i^{(ШО)}, \quad (2.95)$$

де N - кількість нейронів ШО, пов'язаних з *j*-им нейроном ШД; $y_i^{(ШО)}$ - активність *i*-ого нейрону ШО, пов'язаного з *j*-им нейроном ШД.

Завданням єдиного нейрону ШВ є визначення максимального вихідного сигналу нейронів ШД. Для внесення в MPNN знань про правило класифікації фонеми [а] або [б] виду (2.93) достатньо:

- внести в ШО новий нейрон;
- внести в ШФ додаткові нейрони, що відповідають продукційному правилу і співвіднести для них вагові коефіцієнти вхідних зв´язків з величинами ідентифікуючих параметрів, які відповідають заданій фонемі;
- встановити зв'язки нових нейронів ШФ і ШО;
- встановити для нового нейрону ШО вихідний зв´язок з відповідним нейроном ШД *А* або *Б*.

Таким чином в даному розділі наведено результати вдосконалення методологічного забезпечення вирішення задачі нейромережевого розпізнавання фонем в ГС КС, що стосуються:

- розробки концептуальної моделі, що за рахунок конкретизації параметрів оцінювання та факторів впливу на ефективність процесу нейромережевого розпізнавання фонем, яка дозволить деталізувати напрямки подальших досліджень;
- розробки принципів застосування нейронних мереж для розпізнавання фонем, що за рахунок врахування ступеню забезпечення

нейромережевою моделлю характеристик поставленої задачі, співвідношення очікуваних вихідних сигналів нейромережевої моделі з подібністю фонем між собою, за рахунок застосування теорії вейвлет-претворень для прогнозування кількості запитів модулю розпізнавання ГС та застосування продукційних правил при формуванні навчальних прикладів, які забезпечують можливість підвищення ефективності нейромережевих моделей;

- розробки моделей процесів застосування нейромережевих засобів, в яких за рахунок реалізації запропонованих принципів та розробленого математичного апарату забезпечується можливість формалізації визначення ефективних видів нейромережевих моделей, підвищення точності прогнозу завантаження обчислювальних ресурсів КС, що обслуговує нейромережеві засоби розпізнавання, та оперативність створення нейромежевих засобів розпізнавання;
- побудуви моделі формування навчальних прикладів, що за рахунок використання запропонованих принципів визначення очікуваного вихідного сигналу для еталонів фонем забезпечить можливість зменшення терміну навчання нейронних мереж.

Запропоновані принципи та моделі являються основою розвитку методологічної бази нейромережевого розпізнавання фонем в ГС в КС, що знайшло відображення в інфологічній схемі, яка показана на рис. 2.7.

Також на основі запропонованих принципів та моделей розроблено наступні методи:

- створення навчальної вибірки НММ для розпізнавання фонем,
- нейромережевого розпізнавання фонем,
- оцінки ефективності застосування НММ для розпізнавання фонем.

Необхідність розробки пояснюється тим, що у відомих нейромережевих методах запропоновані принципи та моделі (які дозволяють підвищити ефективність розпізнавання) не використовуються.

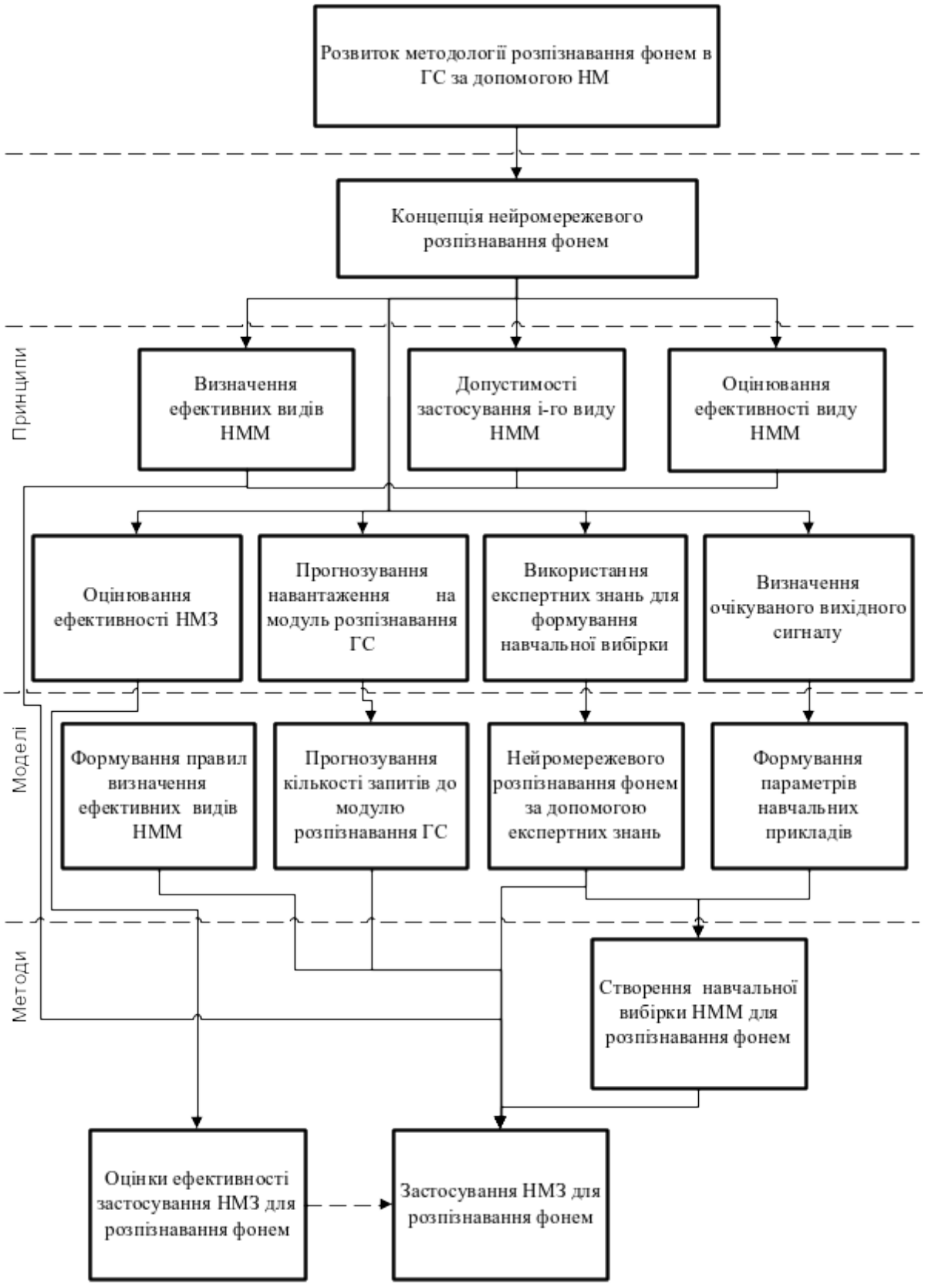

Рис. 2.7. Інфологічна схема методології нейромережевого розпізнавання ГС

РОЗДІЛ 3. РОЗРОБКА НЕЙРОМЕРЕЖЕВИХ МЕТОДІВ РОЗПІЗНАВАННЯ ФОНЕМ В ГОЛОСОВОМУ СИГНАЛІ

3.1. Метод створення навчальної вибірки для нейромережевої моделі розпізнавання фонем

Відправною точкою розробки методу створення навчальної вибірки стали запропоновані в другому розділі навчального посібника принципи визначення очікуваного вихідного сигналу та використання експертних знань для формування навчальної вибірки, і розроблена на їх основі модель формування параметрів навчальних прикладів.

Прийнято до уваги, що БД фонем можуть бути недоступними. Тому передбачена можливість формувати навчальні приклади на основі експертної оцінки фонем в ГС. Завданням експертів буде визначення, яка фонема відповідає виділеному фонемоподібному елементу.

В першому наближенні визначено, що вихідний сигнал НММ реалізується за допомогою одного нейрону. Крім того, передбачено випадок, коли в навчальній вибірці очікуваний вихідний сигнал буде відсутній. В загальному вигляді перетворення інформації, що реалізується даним методом, можливо представити за допомогою наступних виразів:

$$\langle \boldsymbol{\Phi}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3, N_x, N_y, \boldsymbol{Z}_{\text{нв}}, W_n \rangle \rightarrow \langle \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3 \rangle, \tag{3.1}$$

$$\boldsymbol{Z}_1 = \left| z_1^{(\phi_k)} \right|_{K_\phi} = \left| \left(x^{(\phi_k)}, y^{(\phi_k)} \right) \right|_{K\phi}, \tag{3.2}$$

$$\boldsymbol{Z}_2 = \left| z_2^{(\phi_k)} \right|_{K\phi} = \left| r^{(\phi_k)} \right|_{K_\phi}, \tag{3.3}$$

$$\boldsymbol{Z}_3 = \left| z_3^{(\varphi_l)} \right|_{L_\phi} = \left| x^{(\varphi_l)} \right|_{L_\phi}, \tag{3.4}$$

де $\boldsymbol{\Phi}$ – алфавіт фонем, що мають бути розпізнані, $\boldsymbol{\Omega}_1$ – множина фонем, отриманих із БД фонем, $\boldsymbol{\Omega}_2$ – множина фонемоподібних елементів, $\boldsymbol{D}_1$ – множина експертних даних, що стосуються співвідношення очікуваного вихідного сигналу НМ з еталоном фонеми, $\boldsymbol{D}_2$ – множина експертних даних, що стосуються продукційних правил розпізнавання фонем, $\boldsymbol{D}_3$ – множина експертних даних, що стосуються розпізнавання фонем серед фонемоподібних елементів, $\boldsymbol{Z}_1$ – навчальна вибірка, приклади

якої містять очікуваний вихідний сигнал, Z_2 – навчальна вибірка з прикладами у вигляді продукційних правил, Z_3 – навчальна вибірка, приклади якої не мають очікуваного вихідного сигналу, $Z_{нв}$ – множина обмежень на формування навчальної вибірки, W_n – обчислювальна потужність апаратних засобів, що використовуються НМС розпізнавання фонем, $z_1^{(\phi_k)}$, $z_2^{(\phi_k)}$ – навчальні приклади виду Z_1, Z_2 для *k*-ої фонеми, $z_3^{(\varphi_l)}$ – навчальні приклади виду Z_3 для *l*-го фонемоподібного елементу, $x^{(\phi_k)}$ – множина вхідних параметрів для *k*-ої фонеми, $y^{(\phi_k)}$ – очікуваний вихідний сигнал НММ для *k*-ої фонеми, $r^{(\phi_k)}$ – множина продукційних правил виду (2.222) для *k*-ої фонеми, K_ϕ – кількість фонем, які мають бути розпізнані, $x^{(\varphi_l)}$ – вхідні параметри для *l*-го фонемоподібного елементу, L_φ – кількість фонемоподібних елементів.

Основним джерелом даних для формування навчальної вибірки є Ω_1 та Ω_2, що представляють собою акустичні фрагменти ГС. Різниця між Ω_1 та Ω_2 полягає в тому, що кожному акустичному фрагменту ГС, який входить до Ω_1, поставлена у відповідність певна фонема, а в Ω_2 така відповідність відсутня. Φ представляє собою множину фонем, що мають бути розпізнані, яка впорядкована за певним принципом (по алфавіту). При розробці моделі формування параметрів навчальних прикладів визначено, що множина $x^{(\varphi_l)}$ складається із 217 компонентів. Крім того, при розробці НММ розпізнавання фонем за допомогою експертних знань показано, що кількість параметрів в продукційних правилах, які входять до складу $r^{(\phi_k)}$, дорівнює 7.

Аналізуючи (3.1-3.4) встановлено, що для їх реалізації необхідні такі етапи: розрахунок допустимого обсягу навчальної вибірки, визначення очікуваного вихідного сигналу НММ для кожного із еталонів фонем, формування Z_1, Z_2 та Z_3.

Також доцільно застосувати етап визначення можливості формування навчальної вибірки у вигляді множин Z_1, Z_2 чи Z_3. Зазначимо, що наявність навчальної вибірки у вигляді Z_1 надає можливість застосування більш потужних видів НММ, однак її формування є найбільш складним. Тому можливо вважати:

$$Якщо\ \boldsymbol{Z}_1 \in \boldsymbol{Z} \rightarrow \boldsymbol{Z}_2, \boldsymbol{Z}_3 \notin \boldsymbol{Z}. \quad (3.5)$$

При цьому формування навчальної вибірки виду $\boldsymbol{Z}_2$ чи $\boldsymbol{Z}_3$ також може бути ускладнене проведенням експертного оцінювання, а використання відповідних їм НММ може бути обмежене максимально допустимою помилкою розпізнавання. Крім того, базуючись на загальноприйнятій методології розробки НММ [20, 28], зроблено висновок про необхідність етапу нормалізації вхідних і вихідних параметрів та етапу перевірки якості попередньої обробки навчальної вибірки. Також вважалось, що ГС, який є джерелом створення навчальної вибірки, відповідає мінімальним вимогам стандарту GSM та вже пройшов попередню фільтрацію.

Структурна схема методу створення навчальної вибірки показана на рис. 3.1. Наведені на рис. 3.1 етапи методу деталізуються так.

Етап 1 – розрахунок обсягу навчальної вибірки. На даному етапі розраховується мінімально та максимально допустима кількість навчальних прикладів. Вхідними даними етапу є – W_n, N_x та $K_ф$. Для розрахунку мінімально допустимої кількості прикладів використовується вираз

$$L_{\Sigma}^{min} = \sum_{k=1}^{K_{\phi}} L_{\phi_k}^{min}, \quad (3.6)$$

де L_{Σ}^{min} – загальна мінімально допустима кількість прикладів, $L_{\phi_k}^{min}$ – мінімально допустима кількість навчальних прикладів для кожної k-ої фонеми, що має бути розпізнана.

Враховуючи розроблену модель формування параметрів навчальних прикладів та [73, 103], для навчальної вибірки виду $\boldsymbol{Z}_1$ та $\boldsymbol{Z}_3$ мінімально допустима кількість навчальних прикладів для кожної із фонем дорівнює:

$$L_{\phi_k}^{min}(\boldsymbol{Z}_1, \boldsymbol{Z}_3) = 10 \times N_x = 10 \times (55...217) = 550...2170, \quad (3.7)$$

де N_x – кількість вхідних параметрів НММ.

Відповідно розробленої НММ виду MPNN та результатів [84], для навчальної вибірки виду $\boldsymbol{Z}_2$ мінімально допустиму кількість навчальних прикладів для кожної із фонем можливо розрахувати так:

$$L_{\phi_k}^{min}(\boldsymbol{Z}_2) = 20 \times N_x = 20 \times 7 = 140. \quad (3.8)$$

Загальну мінімально допустиму кількість навчальних прикладів для кожного виду вибірки отримаємо, підставивши (3.7, 3.8) в (3.5).

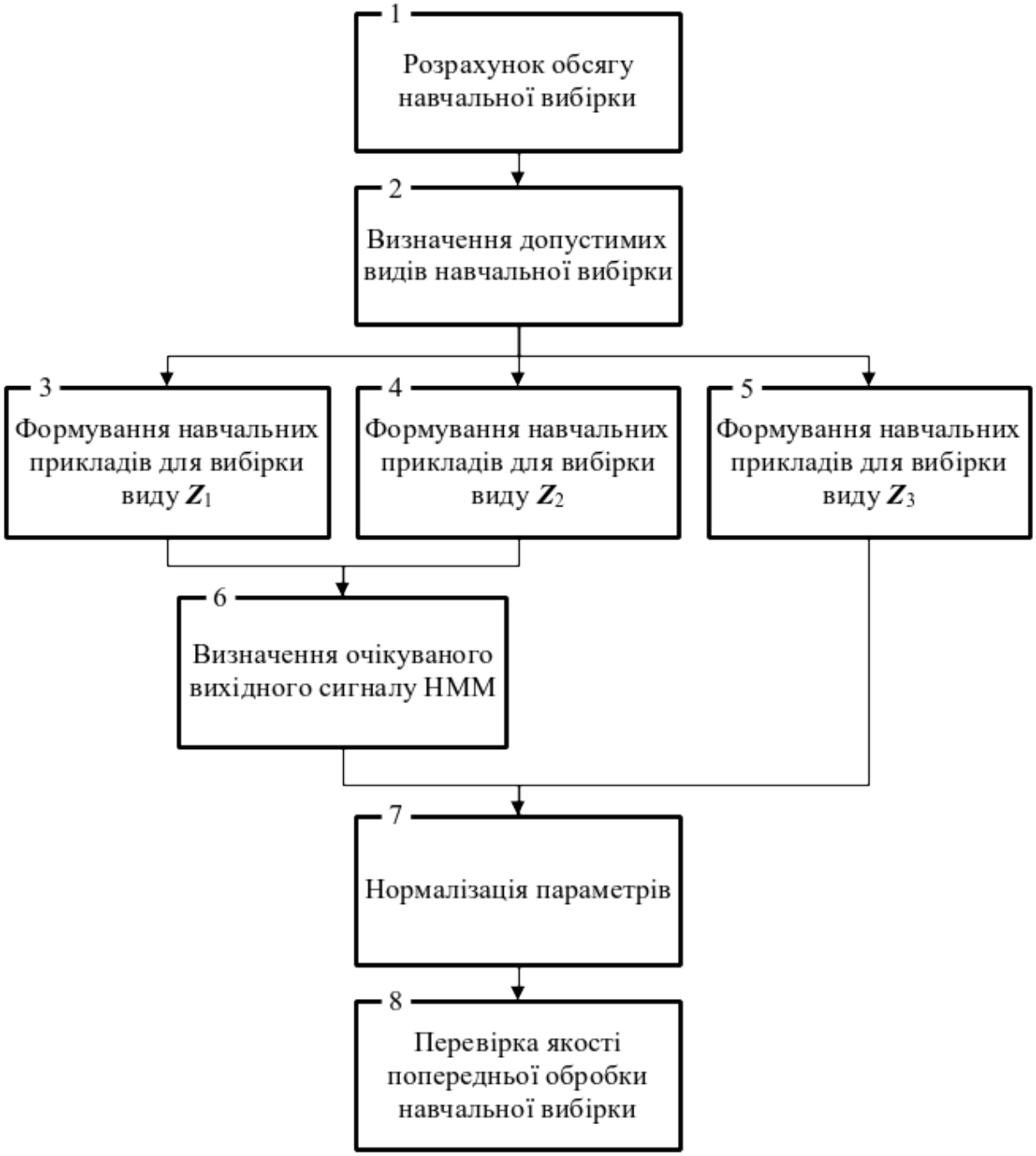

Рис. 3.1. Структурна схема методу створення навчальної вибірки

Максимально допустимий обсяг навчальної вибірки визначається, виходячи з можливості забезпечення безперебійного навчання НММ при її реалізації на загальнодоступному апаратно-програмному забезпеченні. В першому наближенні вважалось, що така реалізація дає змогу безперебійного навчання на протязі 1 доби (86400 с) [46, 47].

Використання вказаної величини у виразах (2.25, 2.26), що отримані в

розробленій моделі правил визначення ефективних видів НММ, дозволяє записати наступні рівняння:

$$8640 \approx k_1 \tau e^{-\varepsilon} L_{\phi_k}^{max}(\boldsymbol{Z}_2, \boldsymbol{Z}_3)(N_x + N_y), \tag{3.9}$$

$$8640 \approx k_2 \tau e^{-\chi\varepsilon} L_{\phi_k}^{max}(\boldsymbol{Z}_1)^2 (N_x + N_y)^2, \tag{3.10}$$

де $L_{\phi_k}^{max}(\boldsymbol{Z}_1)$ – максимально допустима кількість прикладів виду $\boldsymbol{Z}_1$, $L_{\phi_k}^{max}(\boldsymbol{Z}_2, \boldsymbol{Z}_3)$ – максимально допустима кількість прикладів виду $\boldsymbol{Z}_2$, $\boldsymbol{Z}_3$, τ – тривалість однієї обчислювальної операції процесу навчання.

Після підстановки у (3.9, 3.10) визначених у п. 2.4 величин $k_1 \approx 0{,}1$, $k_2 \approx 0{,}001$, $\chi \approx 1$, $\varepsilon \approx 0{,}01$ та тривіальних спрощень отримаємо:

$$L_{\phi_k}^{max}(\boldsymbol{Z}_2, \boldsymbol{Z}_3) \approx 86400 / \tau (N_x + N_y), \tag{3.11}$$

$$L_{\phi_k}^{max}(\boldsymbol{Z}_1) \approx 8640000 / \tau (N_x + N_y)^2. \tag{3.12}$$

Враховуючи в (3.11, 3.12), що для НММ, які відповідають вибіркам $\boldsymbol{Z}_1$, $\boldsymbol{Z}_3$, $N_x + N_y \approx (60 \ldots 220)$, а для НММ, що відповідають $\boldsymbol{Z}_2$, $N_x + N_y \approx 10$ та орієнтуючись на максимізацію терміну навчання отримаємо:

$$L_{\phi_k}^{max}(\boldsymbol{Z}_1) \approx 2400 \tau^{-1}, \tag{3.13}$$

$$L_{\phi_k}^{max}(\boldsymbol{Z}_2) \approx 8640 \tau^{-1}, \tag{3.14}$$

$$L_{\phi_k}^{max}(\boldsymbol{Z}_3) \approx 1440 \tau^{-1}. \tag{3.15}$$

Зазначимо, що $\tau = f(W_n)$, а її величину слід визначити експериментальним шляхом. Таким чином, виходом першого етапу являється L_{Σ}^{min} та $L_{\phi}^{min}(\boldsymbol{Z}_1, \boldsymbol{Z}_3)$, $L_{\phi}^{min}(\boldsymbol{Z}_2)$, $L_{\phi}^{max}(\boldsymbol{Z}_1)$, $L_{\phi}^{max}(\boldsymbol{Z}_2)$, $L_{\phi}^{max}(\boldsymbol{Z}_3)$.

Етап 2 – визначення допустимих видів навчальної вибірки. Вхідними даними етапу являються K_{ϕ}, t_d, L_{Σ}^{min}, $L_{\phi}^{min}(\boldsymbol{Z}_1, \boldsymbol{Z}_3)$, $L_{\phi}^{min}(\boldsymbol{Z}_2)$, $\boldsymbol{\Phi}$, $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{D}_2$, $\boldsymbol{D}_3$. Виходом етапу є множина допустимих видів навчальної вибірки $\boldsymbol{Z}$. Етап розділено на три кроки, кожен із яких співвідноситься з перевіркою входження до $\boldsymbol{Z}$ кожного із видів навчальної вибірки $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$ чи $\boldsymbol{Z}_3$.

Крок 1 – <u>перевірка допустимості</u> $\boldsymbol{Z}_1$. Навчальна вибірка виду $\boldsymbol{Z}_1$ вважається допустимою при виконанні будь-якої із двох умов.

Умова 1. Доступність БД, які дозволяють сформувати $z_1^{(\phi_k)}$ з достатньою кількістю елементів. Для цього в означених БД для кожної із фонем повинна бути достатня кількість відповідних акустичних фрагментів,

записаних з частотою дискретизації не менше, ніж 8 кГц або спектральні характеристики таких фрагментів. Аналітичний вираз умови 1 наступний:

$$Якщо\ \boldsymbol{\Omega}_1 \notin \varnothing \wedge\ L_{\phi_k}^{\Omega_1} \geq 550, k = 1,..K_\phi \rightarrow \boldsymbol{Z}_1 \in \boldsymbol{Z}\ , \qquad (3.16)$$

де $L_{\phi_k}^{\Omega_1}$ – кількість елементів $\boldsymbol{\Omega}_1$, що стосуються *k*-ої фонеми.

Умова 2. За допомогою експертного оцінювання можливо в прийнятний термін сформувати із доступних фонемоподібних елементів множину $\boldsymbol{z}_1^{(\phi_k)}$. В першому наближенні дану умову можливо записати так:

$$Якщо\ \boldsymbol{\Omega}_2 \notin \varnothing \wedge\ \mathrm{E}_1(\boldsymbol{\Phi}, \boldsymbol{\Omega}_2, \boldsymbol{D}_3) = \boldsymbol{z}_1^{(\phi_k)}, L_{\phi_k}^{z_1} \geq 550 \wedge\ t_{Z_1} \leq t_d \rightarrow \boldsymbol{Z}_1 \in \boldsymbol{Z}\ , \qquad (3.17)$$

де E_1 – процедура формування $\boldsymbol{z}_1^{(\phi_k)}$ із $\boldsymbol{\Omega}_2$, $L_{\phi_k}^{z_1}$ – кількість елементів $\boldsymbol{z}_1^{(\phi_k)}$, t_d – допустимий термін формування навчальної вибірки, тривалість якого можливо оцінити за допомогою (2.39), t_{Z_1} – термін формування навчальної вибірки виду $\boldsymbol{Z}_1$, тривалість якого можливо оцінити за допомогою (2.30).

Крок 2 – <u>перевірка допустимості</u> $\boldsymbol{Z}_2$. Навчальна вибірка виду $\boldsymbol{Z}_2$ вважається допустимою при виконанні умови – за допомогою експертного оцінювання в прийнятний термін для кожної *k*-ої фонеми можливо розробити множину $\boldsymbol{r}^{(\phi_k)}$ з кількістю елементів не меншою, ніж 140:

$$Якщо\ \mathrm{E}_2(\boldsymbol{\Phi}, \boldsymbol{D}_2) = \boldsymbol{r}^{(\phi_k)}, L_{\phi_k}^{D_2} \geq 140 \wedge\ t_{Z_2} \leq \bar{t}_\partial \rightarrow \boldsymbol{Z}_2 \in \boldsymbol{Z}\ , \qquad (3.18)$$

де E_2 – процедура формування продукційних правил, t_{Z_2} – термін формування навчальної вибірки виду $\boldsymbol{Z}_2$, який можливо оцінити за допомогою (2.29).

Крок 3 – <u>перевірка допустимості</u> $\boldsymbol{Z}_3$. Навчальна вибірка виду $\boldsymbol{Z}_3$ вважається допустимою при виконанні умови:

$$Якщо\ L_\varphi \geq k_\varphi L_\Sigma^{min} \rightarrow \boldsymbol{Z}_3 \in \boldsymbol{Z}\ , \qquad (3.19)$$

де k_φ – коефіцієнт очікуваного розподілу фонем в доступній $\boldsymbol{\Omega}_2$.

Використавши [28, 48], визначено, що при застосуванні в якості джерела даних для формування $\boldsymbol{\Omega}_2$ загальнодоступних телевізійних звукових записів $k_\varphi = 10$. Це дозволило записати (3.19) так:

$$Якщо\ L_\varphi \geq 10 L_\Sigma^{min} \rightarrow \boldsymbol{Z}_3 \in \boldsymbol{Z}\ . \qquad (3.20)$$

Етап 3 – формування навчальних прикладів для вибірки виду $\boldsymbol{Z}_1$.

Базою етапу являється розроблена модель формування параметрів навчальних прикладів. Враховано, що джерелом формування $\boldsymbol{Z}_1$ можуть бути або БД фонем, або множини фонемоподібних елементів. Для цього випадку передбачено застосувати процедуру експертного оцінювання відповідності фонемоподібного елементу до деякої фонеми. Вхідними даними етапу є K_ϕ, $\boldsymbol{\Phi}$, $L_\phi^{min}(\boldsymbol{Z}_1)$, $L_\phi^{max}(\boldsymbol{Z}_1)$, $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{D}_3$, а виходом – множина навчальних прикладів $\boldsymbol{Z}_{1,a} = \left\{ \boldsymbol{z}_{1,a}^{(\phi_k)} \right\}_{K_\phi} = \left\{ \left(\boldsymbol{x}_a^{(\phi_k)}, y_a^{(\phi_k)} \right) \right\}_{K\phi}$, параметри яких потребують нормалізації. При цьому вихідний сигнал представлено у символьному вигляді. Етап розділяється на шість кроків.

Крок 1 – визначення квазістаціонарних фрагментів. Для визначення застосовуються вирази (2.59, 2.60). Виходом кроку є від 1 до 5 квазістаціонарних фрагментів ГС, кожен із яких представляє собою 128 відліків амплітуд ГС.

Крок 2 – обробка квазістаціонарних фрагментів. Результатом обробки являються величини мел-кепстральних коефіцієнтів кожного із квазістаціонарних фрагментів. Виконання кроку полягає у накладенні на фрагмент віконної функції Хемінга (3.21), застосуванні дискретного перетворення Фур'є (3.22) та розрахунку мел-кепстральних коефіцієнтів (3.23-3.27).

$$w(n) = 0{,}53836 - 0{,}46164 \times cos(2\pi n/(N-1)) \times s(n), \tag{3.21}$$

де n – номер відліку дискретизованого фрагменту ГС, N – кількість відліків у фрагменті, $s(n)$ – початкова величина дискретизованого ГС у n-му відліку, $w(n)$ – величина дискретизованого ГС у n-му відліку після обробки віконною функцією Хемінга:

$$x_k = \sum_{n=0}^{N-1} w(n) \times cos(2\pi nk / N), 0 \le k < N, \tag{3.22}$$

де x_k – амплітуда на частоті k.

$$m = 1127{,}01048 \times ln\left(1 + \frac{f}{700}\right), \tag{3.23}$$

де m – кількість мелів ГС, що відповідає частоті f.

$$f = 700 \times \left(e^{\frac{m}{1127{,}01048}} - 1 \right), \tag{3.24}$$

$$n_{mel} = \frac{m_{max} - m_{min}}{k_{mel}}, \tag{3.25}$$

де $k_{mel} = 24$ – кількість мел-фільтрів, n_{mel} – ширина мел-фільтру, m_{max}, m_{min} – максимальне та мінімальне значення ГС в мелах.

$$c_k = \alpha_k \sum_{n=0}^{k_{mel}-1} x_n \cos\left[\frac{\pi}{k_{mel}}(n + 0{,}5)k\right], k = 0,..(k_{mel} - 1), \tag{3.26}$$

$$\alpha_k = \begin{cases} \sqrt{\frac{1}{N}}, якщо\ k = 0 \\ \sqrt{\frac{2}{N}}, якщо\ k \neq 0 \end{cases}, \tag{3.27}$$

де c_k – *k*-ий мел-кепстральний коефіцієнт.

Крок 3 – <u>експертне оцінювання фонемоподібних елементів</u>. Даний крок виконується тільки при використанні в якості джерела даних множини фонемоподібних елементів ГС.

Результатом такого виконання є визначення очікуваних фонем, що відповідають кожному із фонемоподібних елементів. Для цього пропонується застосовувати процедуру експертного оцінювання, подібну до процедури експертної оцінки близькості фонем, що використовується в розробленій моделі формування параметрів навчальних прикладів, а математичний апарат якої задано виразами (2.67–2.73). Відмінність означених процедур полягає тільки у тому, що при оцінюванні фонемоподібних елементів експерт для кожного з них виставляє ймовірність співвіднесення з кожною із фонем, що входять до алфавіту $\boldsymbol{\Phi}$.

Тобто у виразі (2.67) величина $x_{n,m}$ представляє собою виставлену експертом ймовірність того, що даний фонемоподібний елемент є *n*-ою фонемою. Відповідно, у виразі (2.68) величина $\bar{x}_n$ інтерпретується як середня колективна оцінка того, що даний фонемоподібний елемент є *n*-ою фонемою. Також вважається

$$Якщо\ x_n \leq \Delta_ф \rightarrow x_n = 0, \tag{3.28}$$

де $\Delta_ф$ – емпіричний коефіцієнт (в першому наближенні $\Delta_ф = 0{,}1$).

Запропонована процедура експертного оцінювання передбачає співвіднесення фонемоподібного елементу з декількома фонемами.

Крок 4 – визначення вихідного сигналу в символьному вигляді. Крок адаптовано до джерела навчальних прикладів. Якщо джерелом даних є Ω_1, то значенням вихідного параметру прикладу є комбінація символів:

$$y^{(\phi_k)} = 1 * \phi_k, \tag{3.29}$$

де *k*– номер фонеми в алфавіті.

Якщо джерелом даних являється Ω_2 і в результаті виконання кроку 4 визначено, що є ймовірність віднесення фонемоподібного елементу до декількох фонем, то даний елемент є основою формування декількох навчальних прикладів. Вихідний сигнал для цих прикладів визначається так:

$$y^{(\phi_k)} = \bar{y}_k * \phi_k, \tag{3.30}$$

де k – номер фонеми в алфавіті, Φ_k – k-та фонема, $\bar{y}_k$ – середня колективна оцінка того, що даний фонемоподібний елемент є *k*-ою фонемою.

Крок 5 – масштабування. На даному кроці, відповідно до розробленої моделі формування навчальних прикладів, розраховується масштабний коефіцієнт k_ϕ та проводиться масштабування фонемоподібного ГС. Для цього використовуються вирази (2.43– 2.62).

Крок 6 – визначення вхідних параметрів. Крок орієнтовано на визначення значень вхідних параметрів навчальних прикладів. Значення першого вхідного параметру дорівнює масштабному коефіцієнту k_ϕ:

$$x_1 = k_\phi. \tag{3.31}$$

Значення всіх інших вхідних параметрів дорівнюють відповідним значенням мел-кепстральних коефіцієнтів, визначених в результаті застосування процедури масштабування:

$$x_q = c_q,\ q = i + 24(j - 1) + 1, \tag{3.32}$$

де j – номер квазістаціонарного інтервалу фонемоподібного елементу після масштабування, i – номер мел-кепстрального коефіцієнту в *j*-му фрагменті.

Етап 3 виконується, доки обсяг навчальної вибірки не перевищить $L_{\phi_k}^{max}(\boldsymbol{Z}_1)$ (3.13) або доки термін її формування не перевищить t_d (2.33).

Етап 4 – формування навчальних прикладів для вибірки виду $\boldsymbol{Z}_2$. Етап базується на розробленій НММ розпізнавання фонем за допомогою експертних знань. Вхідними даними етапу являються $\boldsymbol{\Phi}$, $\boldsymbol{D}_2$ та K_ϕ.

Виходом є множина навчальних прикладів виду $\boldsymbol{Z}_{2,a} = \left| z_{2,a}^{(\phi_k)} \right|_{K\phi} = \left| r_a^{(\phi_k)} \right|_{K_\phi}$, параметри яких потребують нормалізації.

При цьому вихідний сигнал представлено у символьному вигляді. Суть етапу полягає в тому, що для кожної із фонем алфавіту $\boldsymbol{\Phi}$ реалізується процедура експертного оцінювання даних, що стосуються продукційних правил розпізнавання фонем на основі аналізу перших семи формант ГС.

Як і у випадку експертного оцінювання фонемоподібного елементу, що проводиться на третьому кроці другого етапу даного методу, в якості бази обрана процедура експертної оцінки близькості фонем, що використовується в розробленій моделі формування параметрів навчальних прикладів (2.67–2.73).

В першому наближенні вважається, що на відміну від (2.67) експертні дані, які стосуються деякої *k*-ої фонеми із алфавіту $\boldsymbol{\Phi}$, представляють собою матрицю виду:

$$D_{2,k} = \begin{vmatrix} \left(X_{k,1,1}^{min}, X_{k,1,1}^{max}\right), & \dots & \left(X_{k,n,1}^{min}, X_{k,n,1}^{max}\right), & \dots & \left(X_{k,N,1}^{min}, X_{k,N,1}^{max}\right), y_{k,1} * \phi_k \\ \dots & \dots & \dots & \dots & \dots \\ \left(X_{k,1,m}^{min}, X_{k,1,m}^{max}\right), & \dots & \left(X_{k,n,m}^{min}, X_{k,n,m}^{max}\right), & \dots & \left(X_{k,N,m}^{min}, X_{k,N,m}^{max}\right), y_{k,m} * \phi_k \\ \dots & \dots & \dots & \dots & \dots \\ \left(X_{k,1,M}^{min}, X_{k,1,M}^{max}\right), & \dots & \left(X_{k,n,M}^{min}, X_{k,n,M}^{max}\right), & \dots & \left(X_{k,N,M}^{min}, X_{k,N,M}^{max}\right), y_{k,M} * \phi_k \end{vmatrix}, \quad (3.33)$$

де N – кількість компонент в продукційному правилі для *k*-ої фонеми; $X_{k,n,m}^{min}, X_{k,n,m}^{max}$ – нижня та верхня межа *n*-го діапазону продукційного правила для *k*-ої фонеми, визначені *m*-им експертом, $y_{k,m}$ – визначена *m*-им експертом ймовірність того, що при виконанні продукційного правила фонемоподібний елемент є *k*-ою фонемою, M – кількість експертів.

Відповідно моделі MPNN, для всіх фонем алфавіту $\boldsymbol{\Phi}$ *n*-та компонента продукційного правила відповідає *n*-ій форманті ГС, а кількість таких компонент $N = 7$. Враховуючи (3.33), $\boldsymbol{D}_2$ можливо представити так:

$$\boldsymbol{D}_2 = \left| D_{2,1}, \dots D_{2,K_\phi} \right|. \quad (3.34)$$

Власне процедура експертного оцінювання полягає у застосуванні виразів (2.67-2.73) до компонент (3.33, 3.34). Виходом даного етапу є навчальні приклади, що входять до складу $\boldsymbol{Z}_2$ і визначаються виразами виду:

$$\boldsymbol{z}_2^{(\phi_k)} = \left| \left(\left(X_{k,1}^{\min}, X_{k,1}^{\max} \right), \ldots \left(X_{k,7}^{\min}, X_{k,7}^{\max} \right), \overline{y}_k \times \phi_k \right)_i \right|_I, \qquad (3.35)$$

де I – кількість навчальних прикладів для *k*-ої фонеми.

Етап 4 виконується доки обсяг навчальної вибірки не перевищить $L_{\phi_k}^{max}(\boldsymbol{Z}_2)$ (3.14), або доки термін її формування не перевищить t_d (2.33).

Етап 5 – формування навчальних прикладів для вибірки виду $\boldsymbol{Z}_3$. Реалізація даного етапу схожа на реалізацію етапу 3 за виключенням того, що фонемоподібний елемент не співвідноситься з жодною фонемою із алфавіту $\boldsymbol{\Phi}$, відповідно вихідний сигнал в начальних прикладах не розрахується.

Вхідними даними цього етапу є $\boldsymbol{\Omega}_2$, $L_{\phi_k}^{max}(\boldsymbol{Z}_3)$, t_d, K_ϕ та L_Φ, а виходом – $\boldsymbol{Z}_{3a}$. Відсутність можливості співвіднести фонемоподібний елемент з еталоном фонеми призводить до неможливості визначити масштабний коефіцієнт виду (2.49). Тому пропонується проводити масштабування фонемоподібного елементу для всіх можливих значень масштабного коефіцієнту від 0,2 до 5.

У відповідності з розробленою моделлю формування параметрів навчальних прикладів, якщо $k_\phi \in [0{,}2 \ldots 1]$, то крок зміни $\Delta k_\phi = 0{,}2$. У випадку, коли $k_\phi \in [1 \ldots 5]$, крок зміни $\Delta k_\phi = 1$. Таким чином, один фонемоподібний елемент буде джерелом даних для 9 прикладів.

В підсумку реалізація етапу полягає у виконанні трьох кроків.

Крок 1 – визначення квазістаціонарних фрагментів. Цей крок ідентичний першому кроку третього етапу даного методу.

Крок 2 – обробка квазістаціонарних фрагментів. Цей крок ідентичний другому кроку третього етапу даного методу.

Крок 3 – визначення вхідних параметрів прикладів. На даному кроці для кожного із можливих значень $k_\phi \in \{0{,}2; 0{,}4; 0{,}6; 0{,}8; 1; 2; 3; 4; 5\}$ розраховуються значення елементів множини $\boldsymbol{x}^{(\varphi_l)}$, які і задають навчальні приклади виду є $\boldsymbol{Z}_3$. Для розрахунку використовуються вирази (3.31, 3.32).

Етап 5 виконується доки обсяг навчальної вибірки не перевищить $L_{\phi_k}^{max}(\boldsymbol{Z}_3)$ (3.15), або доки термін її формування не перевищить t_d (2.33).

Етап 6 – визначення вихідного сигналу. Етап орієнтовано на визначення для кожного навчального прикладу виду $\boldsymbol{Z}_1$ та $\boldsymbol{Z}_2$ величини очікуваного вихідного сигналу НММ, що враховує схожість акустичних характеристик фонем.

Вхідними даними етапу являються $\boldsymbol{\Phi}$, $\boldsymbol{D}_1$, K_ϕ, а також множини навчальних прикладів, що входять до складу $\boldsymbol{Z}_{1,a}$ та $\boldsymbol{Z}_{2,a}$. Виходом етапу являються модифіковані множини навчальних прикладів вибірки $\boldsymbol{Z}_{1,b}$ та $\boldsymbol{Z}_{2,b}$.

Базою етапу являється запропонований принцип визначення очікуваного вихідного сигналу НММ для еталонів фонем та розроблена з використанням вказаного принципу модель формування параметрів навчальних прикладів.

Відповідно вказаної моделі, для визначення вихідного сигналу НММ слід обрахувати міру схожості кожної з фонем та врахувати цю схожість у вихідному сигналі, розрахованому на третьому та четвертому етапах даного методу. Тому етап розділено на два кроки.

Крок 1 – <u>розрахунок міри схожості фонем.</u> Розрахунок полягає в проведенні експертного оцінювання множини фонем, що входять до множини $\boldsymbol{\Phi}$, та обробці експертних даних за допомогою математичного апарату.

Виходом кроку є множина мір схожості фонем між собою:

$$\boldsymbol{O} = \{o_k\}_{K_\phi}, \tag{3.36}$$

де o_k – оцінка міри схожості *k*-ої фонеми із алфавіту $\boldsymbol{\Phi}$.

Крок 2 – <u>врахування схожості фонем.</u> Для врахування оцінки міри схожості у вихідному сигналі навчального прикладу у (3.30, 3.35) компонент $y_k * \phi_k$ змінюється на $y_k \times \overline{o}_k$.

В результаті *i*-ий навчальний приклад, що стосується *k*-ої фонеми виду $\boldsymbol{Z}_1$ або $\boldsymbol{Z}_2$, отримає наступний вигляд:

$$z_{1,i}^{(\phi_k)} = (x_1, \ldots x_{217})_i, y_{k,i} \times \overline{o}_{k,i}, \tag{3.37}$$

$$z_{2,i}^{(\phi_k)} = \left(\left(X_{k,1}^{\min}, X_{k,1}^{\max} \right), \ldots \left(X_{k,7}^{\min}, X_{k,7}^{\max} \right) \right)_i, y_{k,i} \times \overline{o}_{k,i}. \quad (3.38)$$

Етап 7 – нормалізація параметрів. Етап призначено для того, щоб вхідні та вихідні величини параметрів навчальних прикладів входили в інтервал допустимих значень.

Входом етапу є навчальні приклади, що входять до складу $Z_{1,b}$, $Z_{2,b}$, $Z_{3,a}$. Нормалізація параметрів навчальних прикладів виконується за допомогою наступного виразу:

$$\overline{a} = \frac{a - a^{min}}{a^{max} - a^{min}}, \quad (3.39)$$

де a – початкова величина параметру, $\overline{a}$ – нормалізована величина вхідного параметру *a*, a^{max}, a^{min} – максимальна та мінімальна величина *a* у вибірці.

Тому для кожного із параметрів етап розділяється на два кроки.

Крок 1 – визначення екстремумів. На основі аналізу всіх прикладів навчальної вибірки визначається a^{max} та a^{min}.

Крок 2 – виконання нормалізації. Величина параметру нормалізується із застосуванням виразу (3.39).

Таким чином, виходом етапу являються Z_1, Z_2, Z_3, в яких параметри навчальних прикладів адаптовані до застосування в НММ. Якщо виходом являються Z_2, Z_3, то виконання методу завершується.

Етап 8 – перевірка якості попередньої обробки навчальних прикладів. Даний етап виконується лише для навчальної вибірки виду Z_1. Для перевірки використовується константа Ліпшіца виду:

$$Lp = \max_{i \neq j, x_i \neq x_j} \frac{|y_i - y_j|}{\|x_i - x_j\|}, \quad (3.40)$$

де x_i, x_j – вектори вхідних параметрів для *i*-ої та *j*-ої фонем; y_i, y_j – вихідні параметри для *i*-ої та *j*-ої фонем.

Якість попередньої обробки вважається достатньою, якщо $Lp \leq K_{Lp}$. Інакше слід змінити методику нормалізації, структуру та обсяг навчальної вибірки. В першому наближенні вважається, що $K_{Lp} = 50$.

Виходом етапу є сигнал Q, що свідчить про завершення формування навчальної вибірки або про необхідність її модифікації.

Структурно-аналітична схема методу створення навчальної вибірки показана на рис. 3.2.

Таким чином, в результаті проведених досліджень розроблено метод формування навчальної вибірки для нейромережевих засобів розпізнавання фонем в голосовому сигналі в системі дистанційного навчання.

На відміну від відомих в розробленому методі за рахунок застосування створеної оригінальної моделі формування навчальних прикладів та процедури експертного оцінювання вихідного сигналу навчальних прикладів, який формується на основі множини фонемоподібних елементів, забезпечується можливість зменшення кількості навчальних ітерацій нейромережевої моделі. Крім того, забезпечується можливість створення навчальної вибірки з очікуваним вихідним сигналом (метод навчання з вчителем) у випадку відсутності доступу до баз даних фонем.

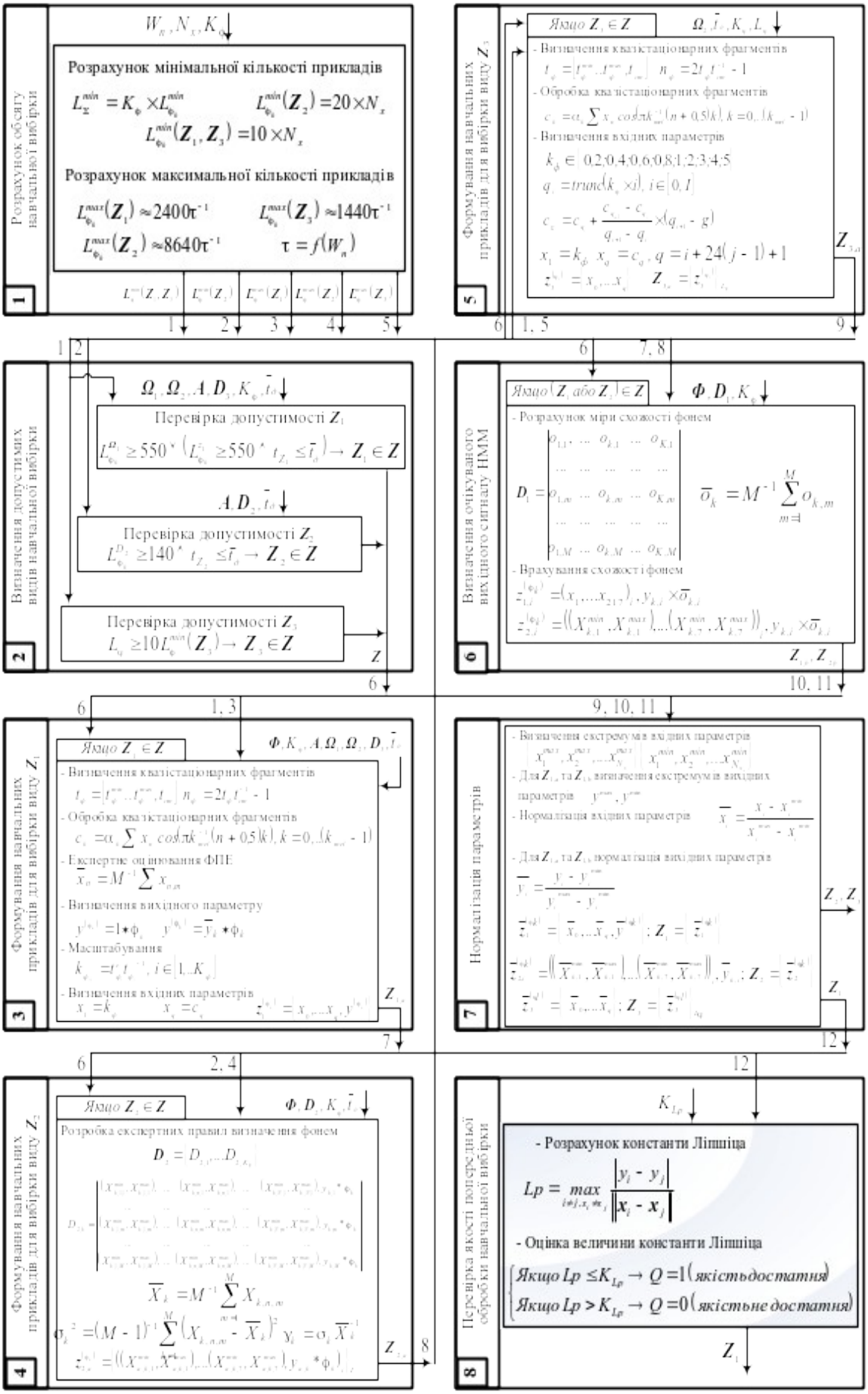

Рис. 3.2. Структурно-аналітична схема методу створення навчальної вибірки

3.2. Метод нейромережевого розпізнавання фонем

Даний метод є результатом інтеграції розроблених елементів методологічної бази процесу нейромережевого розпізнавання, методу створення навчальної вибірки НММ для розпізнавання фонем з відомими нейромережевими моделями та методами розпізнавання ГС. Метою його виконання є визначення таких параметрів НМЗ, що дозволяють найбільш ефективно розпізнавати фонеми в ГС.

Тому в першому наближенні вхідними даними методу є множини умов задачі розпізнавання фонем в ГС, експертних даних та параметрів доступних видів НММ, а виходом є кортеж характеристик та параметрів ефективних НМЗ, який складається із:

– множини параметрів ефективних НММ,

– навчальної вибірки (множини навчальних прикладів),

– помилок розпізнавання,

– ресурсоємності реалізації.

Вказані параметри відповідають означеним ефективним НММ.

При цьому до множини умов задачі розпізнавання фонем в ГС в КС відносяться:

– умови представлення ГС,

– обмеження на формування навчальної вибірки,

– множина статистичних даних кількості запитів до модулю розпізнавання ГС,

– характеристики клієнтського та серверного забезпечення КС,

– обмеження на помилку розпізнавання НМЗ.

Таким чином, аналітичну модель виконання даного методу можливо відобразити за допомогою наступного виразу:

$$\left\langle \boldsymbol{\Phi}, \boldsymbol{У}, \boldsymbol{E}, \boldsymbol{Net_a}, \boldsymbol{H}_{Net_a}, \boldsymbol{S} \right\rangle \rightarrow \left\langle \boldsymbol{Net}_{ve}, \boldsymbol{H}_{Net_{ve}}, \boldsymbol{Z}_{ve}, \boldsymbol{\varepsilon}_{Net_{ve}}, \boldsymbol{\xi}_{Net_{ve}} \right\rangle, \quad (3.41)$$

де $\boldsymbol{У}$ – множина умов задачі розпізнавання фонем в ГС, $\boldsymbol{\Phi}$ – алфавіт фонем; $\boldsymbol{E}$ – множина експертних даних, $\boldsymbol{Net_a}$ – множина доступних видів НММ, $\boldsymbol{H}_{Net_a}$ – параметри доступних видів НММ, $\boldsymbol{S}$ – множина зареєстрованих запитів до модулю розпізнавання, $\boldsymbol{Net}_{ve}$ – множина

ефективних видів НММ, які пройшли верифікацію, $\boldsymbol{H}_{Net_e}$ – параметри верифікованих ефективних видів НММ, $\boldsymbol{Z}_{ve}$ – множина навчальних прикладів, відповідних, $\boldsymbol{\varepsilon}$ – множина помилок розпізнавання для $\boldsymbol{Net}_{ve}$, $\boldsymbol{\xi}$ – множина обсягів використання обчислювальних ресурсів для $\boldsymbol{Net}_{ve}$.

Враховуючи розроблені моделі процесів застосування НМЗ та метод створення навчальної вибірки, для реалізації виразу (3.41) в даному методі необхідно передбачити виконання наступних етапів:

– визначення умов створення та застосування НМЗ;

– створення навчальної вибірки;

– визначення ефективних видів НММ;

– визначення параметрів НММ ефективних видів;

– навчання НММ;

– верифікації НМЗ.

Структурна схема методу нейромережевого розпізнавання фонем в ГС в КС показана на рис. 3.3. Наведені на рис. 3.3 етапи методу нейромережевого розпізнавання деталізуються так.

Етап 1 – визначення умов створення та застосування НМЗ. Основним призначенням даного етапу є визначення умов створення та застосування НМЗ розпізнавання фонем в ГС. Відповідно результатів [20, 28], ці умови можливо охарактеризувати за допомогою:

– параметрів фіксації та попередньої обробки ГС ($\boldsymbol{\Theta}_{гс}$);

– параметрів, що відповідають умовам формування навчальної вибірки ($\boldsymbol{\Theta}_{нв}$), розробки НММ ($\boldsymbol{\Theta}_{ннм}$) та використання апаратно-програмного забезпечення НМЗ ($\boldsymbol{\Theta}_{апз}$);

– часових параметрів застосування НМЗ ($\boldsymbol{\Theta}_{чп}$).

Зазначимо, що до множини $\boldsymbol{\Theta}_{гс}$ відносяться відстань до мікрофону, напрям до мікрофону, частота дискретизації ГС, якість запису, рівень шуму. Кортежі $\boldsymbol{\Theta}_{нв} = \langle \boldsymbol{\Phi}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3, \boldsymbol{Z}_{нв} \rangle$, $\boldsymbol{\Theta}_{ннм} = \langle N_x, N_y, \boldsymbol{Net}_a, \varepsilon_{\max}, \tau, \boldsymbol{R}_a, \boldsymbol{\alpha} \rangle$, $\boldsymbol{\Theta}_{апз} = \langle \boldsymbol{S}_з, \xi^o, k_{нмс}, \boldsymbol{k}, W_п \rangle$ та $\boldsymbol{\Theta}_{чп} = \langle t_d, \bar{t}_v, \boldsymbol{T}_p, T_з \rangle$ деталізовано у розроблених моделях правил визначення ефективних видів НММ, формування параметрів навчальних прикладів, прогнозування кількості запитів до модулю розпізнавання ГС, розпізнавання фонем за допомогою експертних знань.

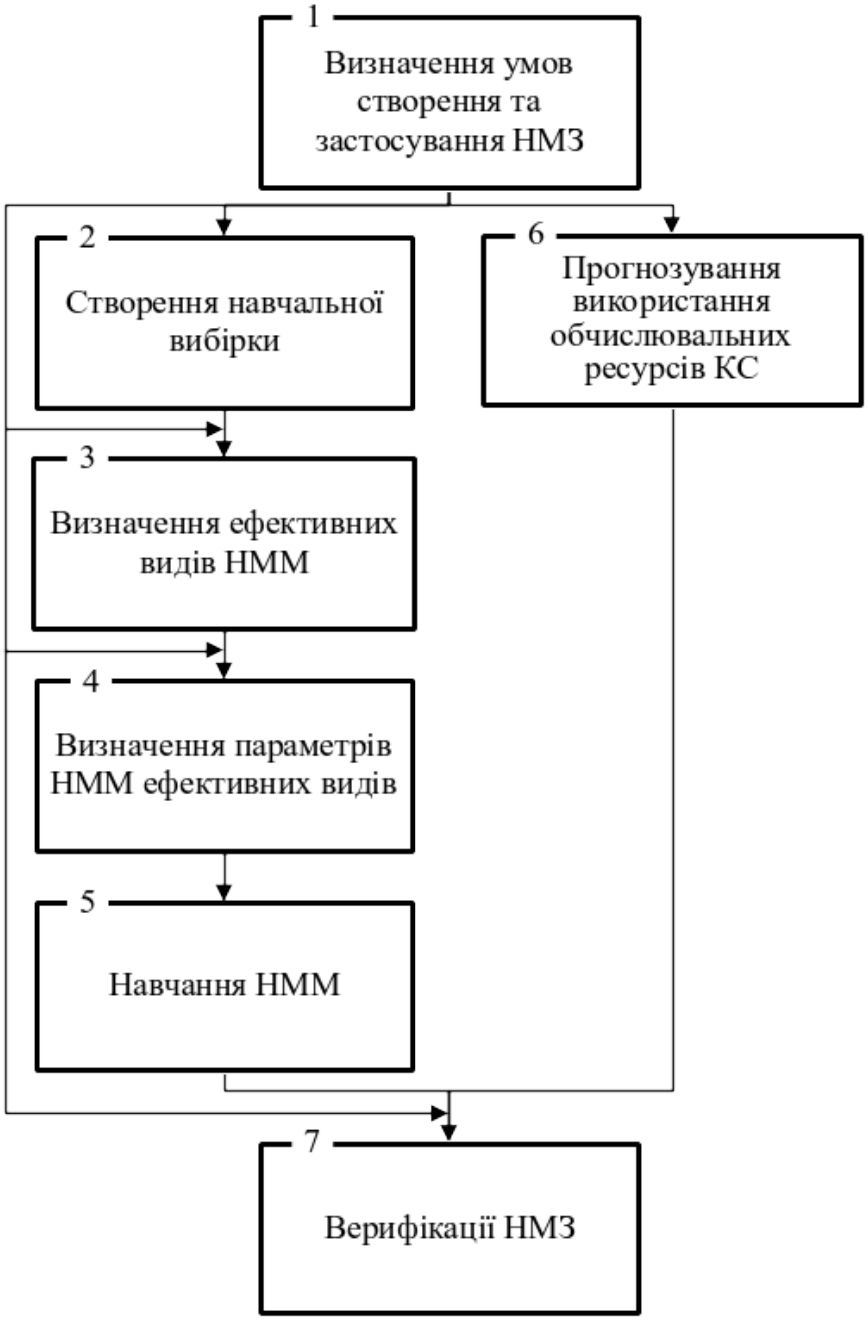

Рис. 3.3. Структурна схема методу нейромережевого розпізнавання фонем

Підґрунтям зазначених умов визначення являються вимоги до КС ($\boldsymbol{Q}_{сдп}$), характеристики програмно-апаратного забезпечення КС ($\boldsymbol{X}_{ПАЗ}$) та матеріально-технічні і трудові ресурси ($\boldsymbol{R}_{сдп}$), які виділяються на її розробку. Оскільки зв'язок між наведеними категоріями обмежень та означеним підґрунтям слабко формалізований, то в базовому варіанті визначення обмежень пропонується реалізувати за допомогою експертних даних ($\boldsymbol{E}_{сдп}$). Обробку експертних даних пропонується проводити подібно до процедури

експертної оцінки близькості фонем, що використовується в розробленій моделі формування параметрів навчальних прикладів, а математичний апарат якої задано виразами (2.67–2.73).

В узагальненому вигляді реалізацію даного етапу можливо представити за допомогою наступного виразу:

$$f(\boldsymbol{Q}_{сдн}, \boldsymbol{R}_{сдн}, \boldsymbol{X}_{ПАЗ}, \boldsymbol{E}_{сдн}) = \langle \boldsymbol{\Theta}_{гс}, \boldsymbol{\Theta}_{нв}, \boldsymbol{\Theta}_{нмм}, \boldsymbol{\Theta}_{апз}, \boldsymbol{\Theta}_{чп} \rangle . \quad (3.42)$$

Етап 2 – створення навчальної вибірки. Виконання даного етапу є реалізацією методу створення навчальної вибірки. Вхідними даними даного етапу є кортеж $\langle \boldsymbol{\Phi}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3, N_x, N_y, \boldsymbol{\Theta}_{нв}, W_n \rangle$, а виходом $\langle \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3 \rangle$. Математичне забезпечення етапу складають вирази (2.47, 2.48, 2.52, 2.53, 2.55-2.62, 2.67-2.73, 3.1-3.40).

Етап 3 – визначення ефективних видів НММ. Етап базується на запропонованих принципах визначення множини ефективних видів НММ, допустимості застосування виду НММ, оцінюванні ефективності виду НММ, призначеної для розпізнавання фонем в ГС та створеній на їх основі моделі правил визначення ефективних видів НММ.

Вхідними даними етапу являється прийнятний термін створення НМЗ (t_{cl}), множина доступних видів НММ ($\boldsymbol{Net}_a$), середній термін створення одного навчального прикладу ($\bar{t}_v$), допустима помилка навчання НММ (ε), тривалість однієї обчислювальної операції процесу навчання (τ), кількість вхідних параметрів НМ (N_x).

Виходом етапу є множина ефективних видів НММ ($\boldsymbol{Net}_e$). Виконання етапу розділено на два кроки, що співвідносяться із визначенням множини допустимих видів НММ та визначенням множини ефективних видів НММ.

Крок 1 – визначення допустимих видів НММ. В загальному випадку виконання даного кроку полягає у розрахунку розроблених виразів (2.31, 2.32). У випадку використання типових вимог до розробки вітчизняних КС можливо використовувати спрощені вирази (2.36, 2.37).

Крок 2 – визначення ефективних видів НММ. Для цього з використанням виразів (2.39-2.42) та даних табл. 2.1 за допомогою експертних даних для кожного допустимого виду НММ визначаються значення критеріїв ефективності. Після цього, базуючись на формалізованих

умовах задачі розпізнавання, для кожного k-го критерію визначаються значення вагового коефіцієнту α_k, за допомогою якого враховується значущість цього критерію.

Далі за допомогою виразів (2.36, 2.37) проводиться визначення множини ефективних видів НММ. Розрахунок найбільш ефективного виду НММ реалізується за допомогою виразу (2.42).

Етап 4 – визначення параметрів НММ ефективних видів. Етап призначений для визначення таких параметрів ефективних видів НММ, що забезпечують їх максимальну обчислювальну потужність (Θ) та мінімальну похибку розпізнавання (ε).

Традиційно Θ оцінюється за допомогою відношення внесених в пам'ять навчальних прикладів до кількості синаптичних зв'язків. Для визначення параметрів використано вираз

$$\begin{cases} \Theta(\Lambda) \to \max \\ \varepsilon(\Lambda) \to \min \end{cases}, \tag{3.43}$$

де $\Lambda = \{\lambda_1, \lambda_2, ...\}$ – множина параметрів НММ.

Розрахунок величин $\{\lambda_1, \lambda_2, ...\}$ пропонується проводити методами, специфічними для виду НММ. Наприклад, при використанні БШП до вказаних параметрів відносяться кількість СШН, кількість нейронів у кожному СШН, структура зв'язків між нейронами.

Для розрахунку цих параметрів можливо використовувати результати [2, 35]. Виходом етапу є множина ефективних видів НММ з визначеними параметрами (Net_e^o).

Етап 5 – навчання НММ. Етап призначено для розрахунку вагових коефіцієнтів синаптичних зв'язків НММ, що входять до множини Net_e^o.

Для цього використовується навчальна вибірка, сформована в результаті виконання етапу 2. Розрахунок реалізується за допомогою методів, характерних для виду НММ.

Для навчання MPNN застосовується розроблений математичний апарат (2.93-2.95). Виходом етапу є $H_{Net_e^o}$ – множина параметрів ефективних видів НММ, що входять до складу Net_e^o.

Етап 6 – верифікації НМЗ. Пропонується проводити верифікацію

розроблених НМЗ з позицій допустимості помилки розпізнавання та достатності обсягу обчислювальних ресурсів КС.

Вхідними даними етапу являються множина ефективних НММ з визначеними параметрами ($\boldsymbol{Net}_e^o$), параметри вказаних НММ ($\boldsymbol{H}_{Net_e^o}$), максимально допустима помилка розпізнавання ($\varepsilon_{\max}$), множина максимально допустимих величин використання обчислювальних ресурсів ($\boldsymbol{\xi}^o$), очікуваний термін використання НМЗ ($T_з$).

Виходом етапу є $\boldsymbol{Net}_{ve}$, $\boldsymbol{H}_{Net_e}$, $\boldsymbol{Z}_{ve}$, $\boldsymbol{\varepsilon}_{Net_{ve}}$ та $\boldsymbol{\xi}_{Net_{ve}}$.

Виконання етапу полягає у реалізації правила:

$$Якщо\ для\ net_i \in \boldsymbol{Net}_e^o\ \varepsilon_i \le \varepsilon_{\max} \wedge \xi_1^{\max} \le \xi_1^o \wedge \ldots \xi_J^{\max} \le \xi_J^o \rightarrow net_i \in \boldsymbol{Net}_{ve}, \quad (3.47)$$

де ε_i – помилка розпізнавання net_i, $\xi_j^{\max}$ – максимальний обсяг використання net_i j-го виду обчислювальних ресурсів на протязі $T_з$, ξ_j^o – максимально допустимий обсяг використання обчислювальних ресурсів.

Зазначимо, що множини $\boldsymbol{H}_{Net_e}$, $\boldsymbol{Z}_{ve}$, $\boldsymbol{\varepsilon}_{Net_{ve}}$ та $\boldsymbol{\xi}_{Net_{ve}}$ формуються на основі відповідних даних $net_i \in \boldsymbol{Net}_{ve}$.

Також для нівелювання ситуації $\xi_i > \xi_{\max}$ можливо застосовувати НМЗ в той часовий період, коли прогнозована кількість запитів до модулю розпізнавання ГС мінімізується.

Отже, проведені дослідження дозволили розробити метод нейромережевого розпізнавання фонем в ГС в КС, який, базуючись на розроблених моделях застосування НМЗ і розробленому методі створення навчальної вибірки, забезпечує достатню точність розпізнавання фонем з врахуванням обмежень щодо створення навчальної вибірки та щодо обмежень, які стосуються обчислювальних ресурсів КС.

3.3. Метод оцінки ефективності застосування нейромережевих засобів для розпізнавання фонем

Розробка методу базується на визначеній множині процедур, виконання яких впливає на ефективність застосування НМЗ для розпізнавання фонем в ГС, визначених показниках оцінки ефективності процесу нейромережевого

розпізнавання фонем та розробленому принципі оцінки ефективності НМЗ розпізнавання фонем.

Відповідно означеного принципу оцінки, відправною точкою розробки методу стало визначення множини критеріїв ефективності $\boldsymbol{U}$, величини яких дозволяють оцінити ефективність застосування НМЗ. Враховуючи визначену множину процедур, запропоновано використовувати наступні критерії, що відображають ступінь виконання процедур:

- u_1 – однокритеріального вибору виду НММ,
- u_2 – багатокритеріального вибору виду НММ,
- u_3 – однокритеріального вибору параметрів НММ,
- u_4 – багатокритеріального вибору параметрів НММ,
- u_5 – адаптації методу навчання,
- u_6 – визначення допустимих видів НММ,
- u_7 – ефективного кодування вихідних параметрів,
- u_8 – застосування різних видів навчальної вибірки,
- u_9 – визначення можливості використання НМЗ при прогнозованому навантаженні КС.

Зазначимо, що застосування критеріїв u_7, u_8 та u_9 являється специфікою даного методу, що відображає особливості застосування НМЗ розпізнавання ГС в КС.

Величини запропонованих критеріїв в першому наближенні можливо оцінити за допомогою експертних даних, подібно до розробленої процедури експертного оцінювання міри близькості фонем.

При цьому вважаємо, що критерій дорівнює 0, коли відповідна можливість в НМЗ не забезпечується, 0,5 – коли забезпечується опосередковано і 1 – коли забезпечується безпосередньо.

Необхідність вдосконалення НМЗ визначається за допомогою процедури виду:

$$u_k \le u_k^d, \tag{3.48}$$

де u_k – величина *k*-го критерію ефективності; u_k^d – мінімально допустима величина *k*-го критерію ефективності.

Підсумкову оцінку ефективності пропонується розрахувати, базуючись на формулі (2.8), а якщо серед декількох доступних НМЗ слід визначити найбільш ефективний, то можливо використати вираз (2.7). Умову достатньої ефективності *i*-го НМЗ можливо сформулювати у наступному вигляді:

$$Якщо\ V_i^{nz} \geq \Delta_{\Sigma}^{nz} \rightarrow V_i \in \boldsymbol{V}_e^{nz}, \tag{3.49}$$

де V_i – ефективність *i*-го НМЗ; Δ_{Σ}^{nz} – коефіцієнт мінімально допустимої ефективності НМЗ; $\boldsymbol{V}_e^{nz}$ – множина ефективних НМЗ.

Таким чином, в узагальненому аналітичному вигляді реалізацію даного методу можливо відобразити за допомогою наступних виразів:

$$\langle \boldsymbol{Nz}_d, \boldsymbol{G}, \boldsymbol{U}, \boldsymbol{D}_e, \boldsymbol{Y} \rangle \rightarrow \langle \boldsymbol{G}_v^{\Sigma}, \boldsymbol{V}_e^{nz}, nz^{max} \rangle, \tag{3.50}$$

$$\boldsymbol{G}_v^{\Sigma} = \{ \boldsymbol{G}_v(nz_i) \}, nz_i \in \boldsymbol{Nz}_d, \tag{3.51}$$

де $\boldsymbol{Nz}_d$ – множина доступних НМЗ, $\boldsymbol{G}$ – множина процедур, реалізація яких визначає ефективність НМЗ, $\boldsymbol{U}$ – множина критеріїв ефективності НМЗ, $\boldsymbol{D}_e$ – множина експертних даних, $\boldsymbol{Y}$ – множина умов задачі розпізнавання фонем в ГС в КС, $\boldsymbol{G}_v(nz_i)$ – множина процедур, реалізація яких дозволяє вдосконалити *i*-ий НМЗ, $\boldsymbol{V}_e^{nz}$ – множина ефективних НМЗ, nz^{max} – найбільш ефективний НМЗ.

Етапи даного методу повинні відображати: визначення величин критеріїв ефективності, визначення коефіцієнтів значимості критеріїв в умовах поставленої прикладної задачі, розрахунок функції ефективності, визначення мінімально допустимих величин критеріїв ефективності, визначення напрямків вдосконалення НМЗ, оцінки достатньої ефективності та визначення найбільш ефективного НМЗ.

Структурна схема методу показана на рис. 3.4. Наведені на рис. 3.4 етапи деталізуються так.

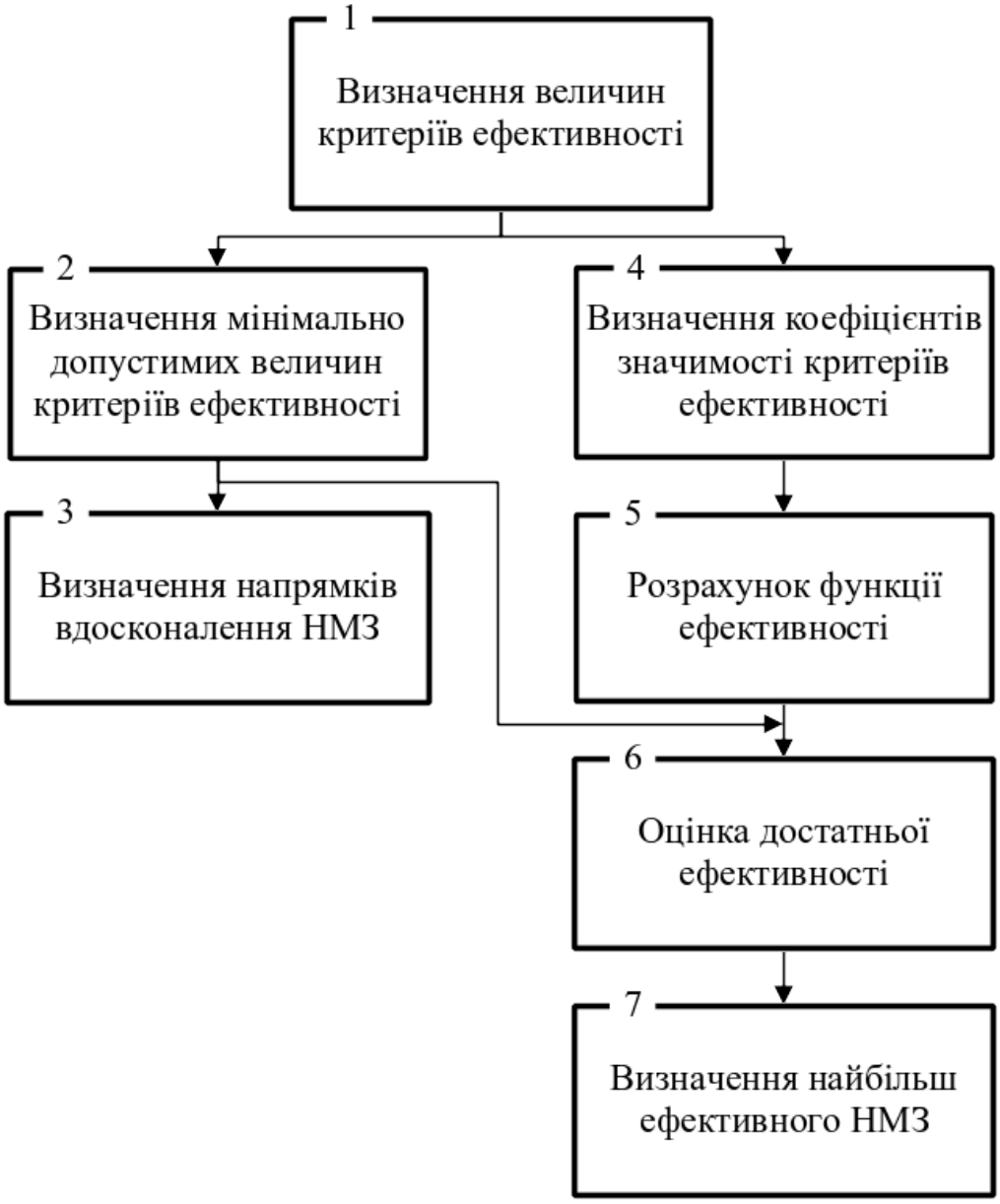

Рис. 3.4. Структурна схема методу оцінки ефективності застосування нейромережевих засобів розпізнавання

Етап 1 – визначення величин критеріїв ефективності. На даному етапі для кожного допустимого НМЗ проводиться експертне визначення величин критеріїв ефективності, що входять до $\boldsymbol{U}$. Для цього використовується процедура експертного оцінювання, задана (2.67-2.73), в яких складова $x_{n,m}$ інтерпретується як оцінка величини *n*-го критерію ефективності, виставлена *m*-им експертом.

Вхідними даними етапу являються множина допустимих НМЗ $\boldsymbol{Nz}_d$, множина експертних даних щодо оцінки величин критеріїв ефективності, а виходом – множина $\boldsymbol{U}^{Nz}$, компонентами якої є множини $\boldsymbol{U}(nz_i)$ для кожного

$nz_i \in \boldsymbol{Nz}_d$.

Етап 2 – визначення мінімально допустимих величин критеріїв ефективності. Вхідними даними етапу являється множина умов конкретної задачі розпізнавання фонем в ГС в КС $\boldsymbol{У}$ та множина експертних даних щодо оцінки мінімально допустимих величин критеріїв ефективності.

Для визначення множини мінімально допустимих величин критеріїв оцінки ефективності НМЗ $\boldsymbol{U}^d$ та коефіцієнту мінімально допустимої ефективності НМЗ застосовується процедура, задана виразами (2.67-2.73), в яких складова $x_{n,m}$ інтерпретується як оцінка мінімально допустимої величини *n*-го критерію ефективності, виставлена *m*-им експертом та величина зазначеного коефіцієнту.

Виходом етапу є множина $\boldsymbol{U}^d$ та коефіцієнт Δ_{Σ}^{nz}.

Етап 3 – визначення напрямків вдосконалення НМЗ. Вхідними даними етапу є множина $\boldsymbol{U}^d$. Для формування процедур, що визначають напрямки вдосконалення кожного $nz_i \in \boldsymbol{Nz}_d$ використовується вираз:

$$Якщо\ u_k(nz_i) \le u_k^d \rightarrow g_k(nz_i) \in \boldsymbol{G}_v(nz_i), u_k(nz_i) \in \boldsymbol{U}(nz_i), u_k^d \in \boldsymbol{U}^d, g_k \in \boldsymbol{G}, \quad (3.52)$$

де $\boldsymbol{G}_v(nz_i)$ – процедури, застосування яких дозволяє вдосконалити nz_i.

Виходом етапу є множина $\boldsymbol{G}_v^{\Sigma}$, компонентами якої є $\boldsymbol{G}_v(nz)$ для кожного доступного НМЗ.

Етап 4 – визначення коефіцієнтів значимості критеріїв ефективності. Етап орієнтовано на визначення множини коефіцієнтів значимості критеріїв ефективності.

Його виконання подібне до другого етапу за виключенням того, що складова $x_{n,m}$ інтерпретується як оцінка коефіцієнту значимості *n*-го критерію ефективності, виставлена *m*-им експертом. Входом етапу є множина умов конкретної задачі розпізнавання фонем в ГС $\boldsymbol{У}$ та множина експертних даних щодо оцінки мінімально допустимих величин критеріїв ефективності.

Виходом етапу є множина

$$\boldsymbol{A}^{nz} = \left| \alpha_k^{nz} \right|_{K^{nz}}, \quad (3.53)$$

де $K^{nz} = 9$ – кількість критеріїв ефективності НМЗ.

Етап 5 – розрахунок функції ефективності. Вхідними даними етапу

являються елементи множин $\boldsymbol{U}^{Nz}$ та $\boldsymbol{A}^{nz}$. Для розрахунку ефективності $nz_i \in \boldsymbol{Nz}_d$ використовується дещо модифікований вираз (2.8):

$$V_i^{nz} = \sum_{k=1}^{K^{nz}} \alpha_k^{nz} u_k(nz_i). \tag{3.54}$$

Виходом етапу являється множина значень функцій ефективності для всіх доступних НМЗ $\boldsymbol{V}^{nz} = \left| V_i^{nz} \right|, nz_i \in \boldsymbol{Nz}_d$.

Етап 6 – оцінка достатньої ефективності. Для оцінювання використовується вираз (3.49). Оцінювання реалізується на основі значень Δ_{Σ}^{nz} та $\boldsymbol{V}^{nz}$, що являються вхідними даними методу.

Виходом є множина ефективних НМЗ $\boldsymbol{V}_e^{nz}$.

Етап 7 – визначення найбільш ефективного НМЗ. Етап виконується за рахунок визначення максимального елементу множини $\boldsymbol{V}_e^{nz}$:

$$\max_{V_i^{nz}} = \left| V_1^{nz}, V_2^{nz}, .. V_I^{nz} \right| \rightarrow nz_i = nz^{max}, \tag{3.55}$$

де I – кількість елементів $\boldsymbol{V}_e^{nz}$, nz^{max} – найбільш ефективний НМЗ .

Таким чином, створено метод оцінки ефективності розробки та застосування НМЗ розпізнавання фонем в ГС, який за рахунок використання запропонованого принципу, критеріїв та функції оцінки ефективності дозволяє, відповідно до визначених показників, обрати найбільш ефективний засіб. Крім того, створений метод дозволяє порівняти ефективність розробки та застосування НМЗ розпізнавання фонем в ГС та визначити напрямки вдосконалення таких засобів.

РОЗДІЛ 4. РОЗРОБКА НЕЙРОМЕРЕЖЕВОЇ СИСТЕМИ РОЗПІЗНАВАННЯ ФОНЕМ ТА ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

4.1. Архітектура нейромережевої системи

Архітектуру НМС розпізнавання фонем в ГС побудовано, базуючись на розробленому методі нейромережевого розпізнавання фонем. Структурно система складається із 22 модулів, об'єднаних в 5 підсистем та окремого модулю управління. Структура НМС показана на рис. 4.1.

Об'єднання модулів в підсистеми проведене з позицій інтеграції компонент, необхідних для виконання етапів нейромережевого методу розпізнавання фонем. Таким чином, призначенням підсистем являється:

- ПВУ – підсистема визначення умов створення та застосування НМЗ, призначена для формування параметрів, що характеризують вказані умови та використовуються як базис подальших розрахунків в інших підсистемах;
- ПСАНВ – підсистема формування адаптивної навчальної вибірки, що призначена для визначення її мінімального та максимального обсягу, допустимих видів, визначення вихідного сигналу НММ для еталонів фонем, визначення та нормалізації вхідних та вихідних параметрів НММ;
- ПРНМ – підсистема розробки НММ, основним призначенням якої є визначення параметрів НММ, що дозволяють найбільш ефективно розпізнавати фонеми в ГС;
- ППВН – підсистема прогнозування навантаження на КС, що дозволяє на основі прогнозованої кількості запитів до модул розпізнавання ГС визначити достатність/недостатність обсягу його обчислювальних ресурсів;
- ПРФ – підсистема розпізнавання фонем, яка, базуючись на результатах спрацювання інших підсистем, дозволяє проводити розпізнавання фонем, виділених із ГС.

Призначенням модулю управління системою МУС є переведення

системи в наступні режими експлуатації: РВУФ – визначення умов функціонування, РВН – визначення налаштувань, РРФ – розпізнавання фонем, РЗ – зупинки.

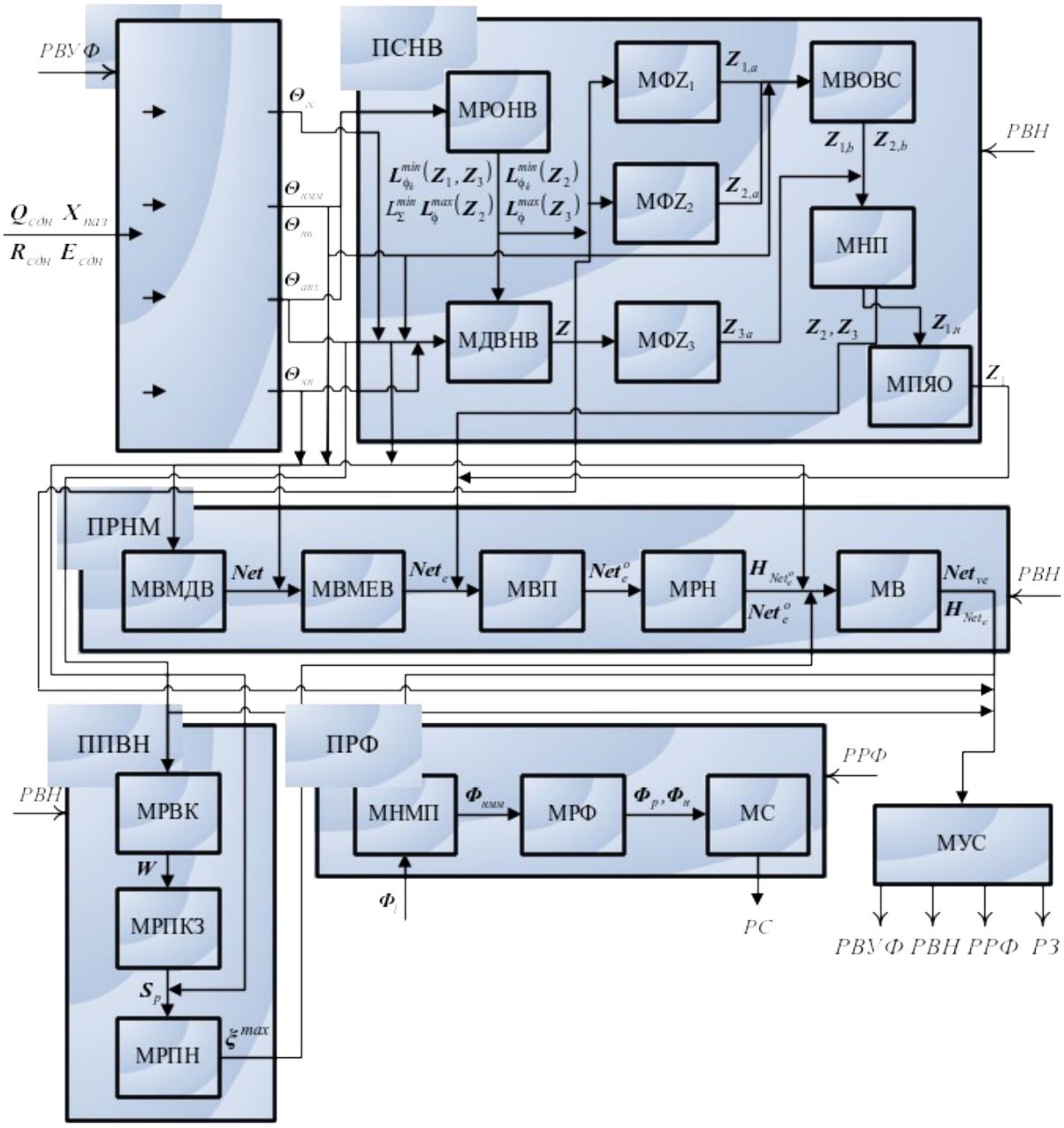

Рис. 4.1. Структура нейромережевої системи розпізнавання фонем

Призначення окремих модулів розробленої НМС, що входять до складу означених підсистем, наведені в табл.4.1.

Розроблена НМС починає функціонувати в режимі визначення умов функціонування, призначенням якого є визначення параметрів, що характеризують вказані умови. Для реалізації такого визначення одночасно спрацьовують модулі МПФГС, МУФНВ, МВАПЗ, МЧП, що входять до

складу ПВУ.

Таблиця 4.1

Склад нейромережевої системи розпізнавання фонем

Назва підсистеми	Назва модулю	Призначення модулю
1	*2*	*3*
ПВУ	МПФГС	Визначення параметрів фіксації ГС
	МУФНВ	Визначення умов формування навчальної вибірки та розробки НММ
	МВАПЗ	Визначення умов використання апаратно-програмних засобів, що реалізують НМЗ
	МЧП	Визначення часових параметрів застосування НМЗ
ПСНВ	МРОНВ	Розрахунку обсягу навчальної вибірки
	МДВНВ	Визначення допустимих видів начальної вибірки
	МФZ_1	Формування навчальної вибірки виду Z_1
	МФZ_2	Формування навчальної вибірки виду Z_2
	МФZ_3	Формування навчальної вибірки виду Z_3
	МВОВС	Визначення очікуваного вихідного сигналу навчальних прикладів виду Z_1 та Z_2
	МНП	Нормалізації параметрів навчальних прикладів
	МПЯО	Перевірка якості обробки прикладів навчальної вибірки
ПРНМ	МВМДВ	Визначення множини допустимих видів НММ
	МВМЕВ	Визначення множини ефективних видів НММ
	МВП	Визначення параметрів НММ, що входять до множини ефективних видів
	МРН	Реалізації навчання НММ
	МВ	Верифікації НММ
ППВН	МРВК	Розрахунок вейвлет-коефіцієнтів на основі статистичних даних кількості запитів до модулю розпізнавання ГС

Таблиця 4.1 (продовження)

1	*2*	*3*
ППВН	МРПКЗ	Розрахунок прогнозованої кількості запитів до модулю розпізнавання ГС
	МРПН	Розрахунок прогнозованого навантаження на КС
ПРФ	МНМП	Нейромережевого порівняння
	МРФ	Порівняння розпізнавання фонем
	МС	Сигналізації про розпізнані фонеми

Джерелом даних для ПВУ являються вимоги до КС ($\boldsymbol{Q}_{сдн}$), характеристики програмно-апаратного забезпечення КС ($\boldsymbol{X}_{паз}$) та матеріально-технічні і трудові ресурси ($\boldsymbol{R}_{сдн}$), які виділяються на її розробку. Для їх обробки в ПВУ також надходить множина відповідних експертних даних ($\boldsymbol{E}_{сдн}$).

Результатом спрацювання модулів ПВУ являються:

– МПФГС – кортеж параметрів фіксації та попередньої обробки ГС ($\boldsymbol{\Theta}_{гс}$), який визначається з використанням розробленої моделі параметрів навчальних прикладів.

– МУФНВ – кортеж параметрів, що характеризують умови створення навчальної вибірки $\boldsymbol{\Theta}_{нв} = \langle \boldsymbol{\Phi}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3, \boldsymbol{Z}_{нв} \rangle$, компоненти якого визначені у виразі (3.1), та кортеж $\boldsymbol{\Theta}_{нмм} = \langle N_x, N_y, \boldsymbol{Net}_a, \varepsilon_{max}, \tau, \boldsymbol{R}_a, \boldsymbol{\alpha} \rangle$, компоненти якого визначені у виразах (2.14, 2.17-2.20, 2.40).

– МВАПЗ – кортеж параметрів, що характеризують умови застосування апаратно-програмного забезпечення $\boldsymbol{\Theta}_{апз} = \langle \boldsymbol{S}_з, \boldsymbol{\xi}^{\partial}, k_{нмс}, \boldsymbol{k}, W_п \rangle$, компоненти якого визначені у виразах (3.44-3.47, 3.1).

– МЧП – кортеж параметрів, що характеризують часові характеристики застосування НМС розпізнавання фонем $\boldsymbol{\Theta}_{чп} = \langle t_d, \bar{t}_v, \boldsymbol{T}_p, T_з \rangle$, компоненти якого визначені у виразах (2.5, 2.15, 3.44).

Якщо параметри, що характеризують умови застосування апаратно-програмних засобів, не відповідають певним обмеженням, то МУС переводить НМСРФ в режим зупинки. В протилежному випадку МУС

переводить систему в режим визначення налаштувань. В цьому режимі одночасно паралельно спрацьовують підсистеми ППВН та ПСНВ.

Модулі, які входять до складу ППВН, спрацьовують послідовно. Вхідними даними ППВН є $\boldsymbol{T}_p, T_з \in \boldsymbol{\Theta}_{чп}$ та $\boldsymbol{S}_з, k_{нмс}, \boldsymbol{k} \in \boldsymbol{\Theta}_{апз}$. Першочергово в ППНВ спрацьовує МРВК, де проводиться розрахунок вейвлет-коефіцієнтів розробленої моделі прогнозування кількості запитів до модулю розпізнавання. Математичне забезпечення даного розрахунку складають вирази (2.84, 2.85, 2.90). Вхідними даними МРВК є $\boldsymbol{T}_p, T_з, \boldsymbol{S}_з$, а виходом являється множина вейвлет-коефіцієнтів $\boldsymbol{W}$, яка передаються в модуль МРПКЗ.

В цьому модулі за допомогою виразу (2.91) розраховується прогнозована кількість запитів до модулю розпізнавання на протязі очікуваного терміну застосування НМЗ. Виходом МРПКЗ є множина значень прогнозованої кількості запитів $\boldsymbol{S}_p = \{s_p(t_i)\}$. $\boldsymbol{S}_p$ разом з $k_{нмс}, \boldsymbol{k}$ передається в модуль МРПН для розрахунку прогнозованого навантаження на веб-сервер. Цей розрахунок реалізується за допомогою виразу (3.46).

Виходом МРПН є множина максимальних обсягів використання кожного виду обчислювальних ресурсів КС $\boldsymbol{\xi}^{max} = \{\xi_1^{max}, \ldots \xi_J^{max}\}$, яка передається в МВ, що входить до складу ПРНМ. На цьому ППВН закінчує своє функціонування.

В ПСНВ першочергово спрацьовує МРОНВ. Вхідними даними цього модулю являються W_n, N_x та $K_ф$, описані у виразі (3.1). Виходом МРОНВ є L_Σ^{min} та множини $\boldsymbol{L}_ф^{min}(\boldsymbol{Z}_1, \boldsymbol{Z}_3)$, $\boldsymbol{L}_ф^{min}(\boldsymbol{Z}_2)$, $\boldsymbol{L}_ф^{max}(\boldsymbol{Z}_1)$, $\boldsymbol{L}_ф^{max}(\boldsymbol{Z}_2)$, $\boldsymbol{L}_ф^{max}(\boldsymbol{Z}_3)$.

Для розрахунку L_Σ^{min} використовується вираз (3.6), а для розрахунку $\boldsymbol{L}_ф^{min}(\boldsymbol{Z}_1, \boldsymbol{Z}_3)$, $\boldsymbol{L}_ф^{min}(\boldsymbol{Z}_2)$, $\boldsymbol{L}_ф^{max}(\boldsymbol{Z}_1)$, $\boldsymbol{L}_ф^{max}(\boldsymbol{Z}_2)$, $\boldsymbol{L}_ф^{max}(\boldsymbol{Z}_3)$ вирази (3.7, 3.8, 3.13-3.15) відповідно. Визначені в МРОНВ параметри L_Σ^{min}, $\boldsymbol{L}_ф^{min}(\boldsymbol{Z}_1, \boldsymbol{Z}_3)$, $\boldsymbol{L}_ф^{min}(\boldsymbol{Z}_2)$, що характеризують мінімально допустимий обсяг навчальної вибірки та визначені в ПВУ $K_ф$, $\bar{t}_\partial$, $\boldsymbol{\Phi}$, $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{D}_2, \boldsymbol{D}_3$ передаються в МДВНВ, де формується множина допустимих видів навчальної вибірки $\boldsymbol{Z}$, яка і є виходом данного модулю.

Для формування $\boldsymbol{Z}$ використовуються вирази (3.16-3.20). Множина $\boldsymbol{Z}$

визначає спрацювання модулів, в яких реалізується формування навчальної вибірки – МФZ$_1$, МФZ$_2$ та МФZ$_3$. Тобто модуль МФZ$_1$, МФZ$_2$ чи МФZ$_3$ спрацьовує тільки в тому випадку, коли відповідна йому навчальна вибірка належить до допустимого виду. Якщо множина $\boldsymbol{Z}$ порожня, то система переводиться у режим зупинки. Для визначення такої ситуації множина $\boldsymbol{Z}$ також передається в МУС.

На вхід МФZ$_1$ поступає K_ϕ, $\boldsymbol{\Phi}$, $\boldsymbol{L}_\phi^{min}(\boldsymbol{Z}_1)$, $L_{\phi_k}^{max}(\boldsymbol{Z}_1)$, $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{D}_3$. Виходом є множина навчальних прикладів $\boldsymbol{Z}_{1,a} = \left\{ \boldsymbol{z}_{1,a}^{(\phi_k)} \right\}_{K_\phi} = \left\{ \left(\boldsymbol{x}_a^{(\phi_k)}, y_a^{(\phi_k)} \right) \right\}_{K\phi}$. Математичне забезпечення МФZ$_1$ складають вирази (2.33, 2.43, 2.44, 2.48, 2.49, 2.52, 2.53, 2.55-2.62, 2.67-2.73, 3.13, 3.21-3.32). На вхід МФZ$_2$ поступає $\boldsymbol{\Phi}$, $\boldsymbol{D}_2$ та K_ϕ. Виходом є множина навчальних прикладів $\boldsymbol{Z}_{2,a} = \left\{ \boldsymbol{z}_{2,a}^{(\phi_k)} \right\}_{K\phi} = \left\{ \boldsymbol{r}_a^{(\phi_k)} \right\}_{K_\phi}$.

Математичне забезпечення МФZ$_2$ складають вирази (2.33, 2.47-2.73, 3.14, 3.33-3.35). На вхід МФZ$_3$ поступає $\boldsymbol{\Omega}_2$, $L_{\phi_k}^{max}(\boldsymbol{Z}_3)$, t_d, K_ϕ та L_φ. Виходом є множина навчальних прикладів $\boldsymbol{Z}_{3a}$. Математичне забезпечення МФZ$_3$ складають вирази (2.33, 2.43, 2.44, 3.15, 3.22-3.27, 3.31, 3.32). Якщо хоча б одна із навчальних вибірок виду $\boldsymbol{Z}_1$ та $\boldsymbol{Z}_2$ є допустимою, то спрацьовує МВОВС. Крім $\boldsymbol{Z}_{1,a}$ та $\boldsymbol{Z}_{2,a}$, на вхід цього модулю надходять $\boldsymbol{\Phi}$, $\boldsymbol{D}_1$, K_ϕ.

Виходом являються множини навчальних прикладів виду $\boldsymbol{Z}_{1,b}$ та $\boldsymbol{Z}_{2,b}$. Зазначимо, що на відміну від $\boldsymbol{Z}_{1,a}$ та $\boldsymbol{Z}_{2,a}$ в прикладах даного виду є чисельна величина очікуваного вихідного сигналу. Для розрахунку цієї величини використовуються вирази (2.47-2.73, 3.30, 3.35-3.38).

Наступним спрацьовує модуль МНП, який реалізує нормалізацію параметрів навчальних прикладів, що входять до складу $\boldsymbol{Z}_{1,b}$, $\boldsymbol{Z}_{2,b}$ чи $\boldsymbol{Z}_{3a}$. Для нормалізації використовується вираз (3.39). Виходом модулю є множини $\boldsymbol{Z}_{1,н}$, $\boldsymbol{Z}_2$, $\boldsymbol{Z}_3$.

Якщо множина $\boldsymbol{Z}_{1,н}$ входить до складу допустимих, то спрацьовує МПЯО, в якому перевіряється якість попередньої обробки навчальних прикладів. Для цього використовується вираз (3.40).

Якщо якість достатня, то виходом МПЯО є множина $\boldsymbol{Z}_1 = \boldsymbol{Z}_{1,н}$, в протилежному випадку $\boldsymbol{Z}_1 = \varnothing$ Після цього ПСНВ передає функціонування до ПРНМ. Модулі ПСНВ спрацьовують в порядку

МВМДВ→МВМЕВ→МВП→ МРН→МВ.

Першим спрацьовує МВМДВ, призначенням якого є визначення множини допустимих видів НММ ($\boldsymbol{Net}$). Вхідними даними цього модулю є величини $\bar{t}_v, \tau, t_d, \boldsymbol{net}_1, \boldsymbol{net}_2$. Зазначимо, що $(\bar{t}_v, t_d) \in \boldsymbol{\Theta}_{чп}$, $\tau \in \boldsymbol{\Theta}_{нмм}$, $(\boldsymbol{net}_1, \boldsymbol{net}_2) \in \boldsymbol{Net}_a \in \boldsymbol{\Theta}_{нмм}$.

Виходом є множина допустимих видів НММ – $\boldsymbol{Net}$. Для визначення $\boldsymbol{Net}$ використовуються вирази (2.161, 2.162, 2.166, 2.167). Отримавши на вхід $\boldsymbol{Net}$, $(\boldsymbol{R}_a, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_{нмм}$ в МВМЕВ проводиться визначення множини ефективних видів НММ – $\boldsymbol{Net}_e$ та найбільш ефективного виду НММ – $net_e^{max} \in \boldsymbol{Net}_e$.

Для цього використовуються вирази (2.40–2.42). $\boldsymbol{Net}_e$ та net_e^{max} передаються в МВП, де за допомогою виразу (3.43) та результатів [28] реалізується визначення параметрів НММ, що входять до множини ефективних видів.

Виходом МВП є $\boldsymbol{Net}_e^o$ – множина ефективних видів НММ з адаптованими параметрами, що разом з множинами $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, $\boldsymbol{Z}_3$ подається на вхід МРН. В цьому модулі реалізується навчання НММ, які входять до складу $\boldsymbol{Net}_e^o$. Для цього використовуються результати [2, 35], що відображають методи навчання видів НММ.

У випадку використання MPNN для навчання використовується розроблене математичне забезпечення (2.92-2.95). Виходом МРН є $\boldsymbol{H}_{Net_e^o}$ – множина параметрів ефективних видів НММ, що входять до складу $\boldsymbol{Net}_e^o$.

Останнім в ПРНМ спрацьовує МВ, що призначений для верифікації розроблених НММ. Вхіднимим даними цього модулю є $\boldsymbol{Net}_e^o$, $\boldsymbol{H}_{Net_e^o}$, $\varepsilon_{max} \in \boldsymbol{\Theta}_{нмм}$, $\boldsymbol{\xi}^o \in \boldsymbol{\Theta}_{апз}$, $T_з \in \boldsymbol{\Theta}_{чп}$ та $\boldsymbol{\xi}^{max}$, а виходом $\boldsymbol{Net}_{ve}$ – множина верифікованих НММ та $\boldsymbol{H}_{Net_e}$ – множина параметрів верифікованих НММ. Верифікація відбувається відповідно сьомого етапу розробленого методу розпізнавання фонем за допомогою виразу (3.47).

Якщо множина $\boldsymbol{Net}_{ve}$ порожня, то відповідний сигнал передається в МУС, який переводить НМС в режим зупинки, в протилежному випадку МУС переводить НМС в режим розпізнавання фонем. Якщо до множини $\boldsymbol{Net}_{ve}$ входить декілька видів НММ, то з метою спрощення побудови НМС в

першому наближенні для подальшого використання вибирається той вид моделі $\boldsymbol{net}_{ve}$, у якої помилка розпізнавання найменша.

Зазначимо, що в базовому випадку до складу $\boldsymbol{net}_{ve}$ входить тільки одна НММ, призначена для нейромережевого порівняння фонемоподібних елементів із всіма еталонами фонем, які входять до алфавіту $\boldsymbol{\Phi}$.

В подальшому $\boldsymbol{net}_{ve}$ може складатись із декількох НММ, кожна із яких призначена для розпізнавання окремої фонеми (групи фонем з однаковою тривалістю еталонів), що входить до $\boldsymbol{\Phi}$. В режимі розпізнавання фонем отримана із ГС множина фонемоподібних елементів $\boldsymbol{\Phi}_I$ подається в МНМП ПРФ.

В цьому модулі за допомогою $\boldsymbol{net}_{ve}$ формується множина результатів нейромережевого порівняння $\boldsymbol{\Phi}^{НММ}$. Якщо до складу $\boldsymbol{net}_{ve}$ входить тільки одна НММ, то складові $\boldsymbol{\Phi}^{НММ}$ мають наступний вигляд:

$$\phi_i^{НММ} = (i, y), \tag{4.1}$$

де i –номер фонемоподібного елементу; y – величина вихідного сигналу НММ.

Якщо до складу $\boldsymbol{net}_{ve}$ входить тільки декілька НММ, то складові $\boldsymbol{\Phi}^{НММ}$ визначаються так:

$$\phi_i^{НММ} = (i, y_1, y_2, \ldots y_N), \tag{4.2}$$

де y_n – вихід НММ, призначеної для розпізнавання n-ої фонеми.

$\boldsymbol{\Phi}^{НММ}$ предається в МРФ, де на її основі формується множина параметрів розпізнаних фонем $\boldsymbol{\Phi}_P$ та множина фонемоподібних елементів, котрі розпізнати не вдалось $\boldsymbol{\Phi}_H$. При цьому фонемоподібний елемент вважається не розпізнаним, якщо вихідний сигнал НММ не знаходиться в межах, що відповідають еталонам фонем алфавіту $\boldsymbol{\Phi}$.

Якщо для нейромережевого порівняння використовується декілька НММ, то фонемоподібний елемент розпізнається як фонема, що відповідає НММ з найбільшим значенням вихідного сигналу. $\boldsymbol{\Phi}_P$ та $\boldsymbol{\Phi}_H$ предають в МС, де формується сигнал PC про результати розпізнавання. Надалі PC може використовуватись для розпізнавання окремих слів.

Таким чином, в результаті проведених досліджень розроблено архітектуру нейромережевої системи розпізнавання фонем в ГС, в якій на

відміну від відомих передбачено використання:

– підсистеми визначення умов застосування,

– підсистеми прогнозування навантаження на модуль розпізнавання ГС,

– підсистеми формування навчальної вибірки,

– модулю визначення допустимих видів нейромережевих моделей,

– модулю визначення ефективних видів нейромережевих моделей.

Запропоновані архітектурні рішення забезпечують достатню точність розпізнавання фонем, а також забезпечують адаптацію системи розпізнавання до умов розробки і застосування.

4.2. Експериментальна установка

Експериментальна установка представляє собою апаратно-програмний комплекс, призначений для проведення експериментальних досліджень запропонованих моделей і методів та створеної на їх основі нейромережевої системи розпізнавання фонем в ГС. Обчислювальні потужності та конфігурація апаратного забезпечення експериментальної установки визначались з позицій забезпечення мінімально допустимих вимог до серверної частини розповсюджених КС.

Враховано, що можливості аудіосистеми повинні відповідати стандарту GSM. Тому використано універсальний комп’ютер на основі процесора Intel(R) Core(TM)2 Quad CPU Q6600 @ 2.40GHz з оперативною пам’яттю обсягом 3 ГБ, жорстким диском з обсягом 500 ГБ і аудіосистемою Realtek HD Audio.

Вимоги до функціональності програмного забезпечення експериментальної установки визначені з позицій забезпечення виконання розроблених методів. Для цього використано декілька програмних додатків, основні характеристики яких наведено в табл. 4.2.

Таблиця 4.2

Характеристики програмних додатків експериментальної установки

Назва	Джерело отримання	Функціональність
Audacity	У вільному доступі (http://audacity.sourceforge.net)	Запис ГС з мікрофону у файл.

		Попередня обробка ГС.
Google Chrome	У вільному доступі (http://www.google.com.ua)	Розпізнавання окремих слів та фонем.
Sphinx	У вільному доступі (http://cmusphinx.sourceforge.net)	Виділення окремих слів та фонемоподібних елементів.
WavTest	Власна розробка	Виділення окремих слів та фонемоподібних елементів.
AnalizWav	Власна розробка	Аналіз фонемоподібних елементів. Розпізнаваня фонем. Розпізнавання окремих слів.
MathCAD	Ліцензована версія (http://www.ptc.com/go/ mathsoft/)	Прогнозування кількості запитів за допомогою вейвлет-моделі.
NeuroPro	У вільному доступі (http://www.neuropro.ru)	Розпізнавання фонем
Decibel Meter	У вільному доступі (http://www.microsoft.com)	Вимірювання рівня шуму

Основою частиною експериментальної установки є розроблений у Windows-застосунок AnalizWav, який функціонує відповідно до рішень другого та третього розділів.

Головне вікно AnalizWav показано на рис. 4.2. Верхня частина вікна «Розрахунок параметрів звукових записів» призначена для відображення параметрів поточного піддослідного ГС, записаного в wav-файл, вибір якого реалізується за допомогою кнопки «Аналіз звукового файлу».

Для прикладу на рис. 4.2 показані параметри ГС, записаного в файл 2.wav. Нижня частина вікна «Порівнянн звукових записів» призначена для розпізнавання фонем шляхом пошуку найбільш схожого еталону.

Для забезпечення сервісу прослуховування звукових записів в цій же частині вікна знаходяться кнопки «Відтворити». Ідентифікація запису проводиться по назві файлу.

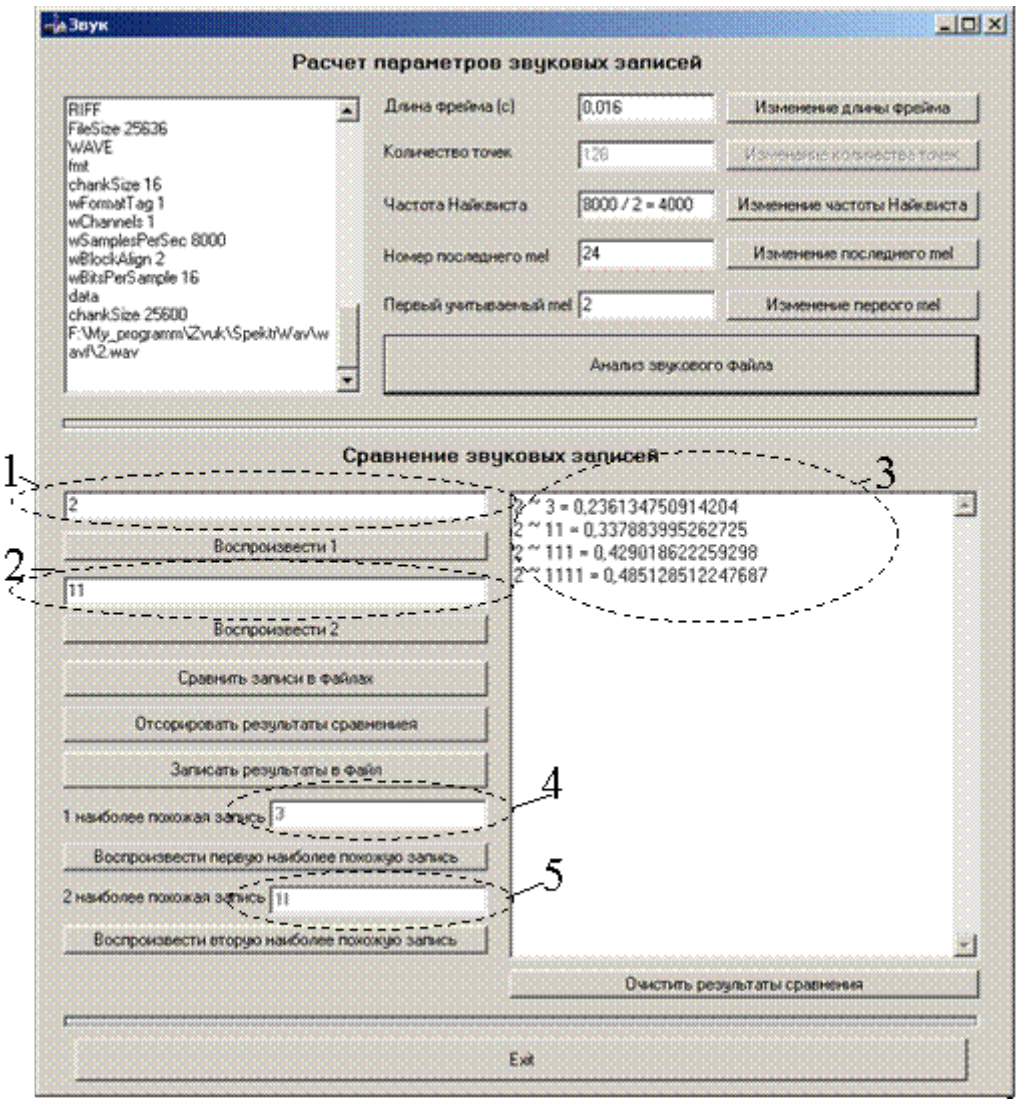

Рис. 4.2. Головне вікно Windows-застосунку AnalizWav

Результати порівняння відображаються в полі 3. Наприклад, запис «2 ~ 3 = 0,236134750914204» означає, що різниця між ГС, записаними в файлах 2.wav та 3.wav, становить 0,236134750914204. Мінімальна різниця, коли ГС співпадають, дорівнює 0.

Для реалізації процедури розпізнавання необхідно виконати такі операції:

1. Скориставшись кнопкою «Аналіз звукового файлу», визначити параметри навчальної вибірки еталонів ГС (фонем), що мають бути розпізнані, та параметри піддослідного ГС (фонемоподібного елементу).

2. Ввести в поле 1 назву файлу, який відповідає піддослідному фонемоподібному елементу.

3. Послідовно ввести в поле 2 назви файлів з еталонами фонем. Після кожного вводу слід натиснути кнопку «Порявняти записи в файлах», що призводить до відображення в полі 3 запису з результатами порівняння.

4. Для сортування результатів порівняння слід натиснути кнопку «Відсортувати результати порівняння». Крім того, що в полі 3 відобразяться результати сортування, в полі 4 і 5 відобразяться назви двох файлів з записами найбільш схожими до піддослідного фонемоподібного елементу.

5. Використання кнопки «Записати результати в файл» призводить до запису результатів порівняння, показаних в полі 3, в текстовий файл.

Таким чином, визначено склад експериментальної установки, призначеної для дослідження розробленої системи розпізнавання фонем в ГС, яка є практичною реалізацією запропонованих теоретичних рішень.

4.3. Експериментальні дослідження

У відповідності з [28, 42], основною метою експериментальних досліджень було підтвердження достовірності основних результатів теоретичних досліджень. Для цього перевірено гіпотези про те, що застосування розроблених методів та системи дозволить забезпечити ефективне розпізнаваня фонем в ГС. У відповідності з [20] проведені дослідження розділені на три етапи.

На першому етапі була перевірена гіпотеза про те, що застосування розробленого методу дозволить забезпечити ефективність процесу навчання НМЗ розпізнавання фонем. При цьому забезпечення ефективності буде реалізоване за рахунок забезпечення допустимої похибки, терміну і ресурсоємності навчання та допустимого терміну формування навчальної вибірки для НММ, що застосовується для розпізнавання.

Відправним пунктом дослідження стало визначення умов формування навчальної вибірки. На основі практичного досвіду встановлено, що в цьому випадку допустимий термін створення модулю розпізнавання складає 1 рік, можливе залучення висококваліфікованих спеціалістів, а виділення додаткових ресурсів, необхідних, наприклад, для доступу до баз даних фонем, не передбачено.

Також визначено, що в КС передбачене розпізнавання елементарних голосових відповідей на тестові запитання типу «Так» або «Ні». Тому кількість фонем K_{ϕ} =5, а до алфавіту входять фонеми $[m],[a],[к],[и],[i]$. Для

формування навчальної вибірки використано один універсальний комп'ютер, характеристики якого наведені в п. 4.2.

Основні результати виконання розробленого методу формування навчальної вибірки наступні. Використавши вирази (3.6-3.8), розраховано, що мінімально допустима кількість навчальних прикладів для навчальної вибірки виду $\boldsymbol{Z}_1$ та $\boldsymbol{Z}_3$ становить $L_{\Sigma}^{min}(Z_1, Z_3) = 2750$, а для навчальної вибірки виду $\boldsymbol{Z}_2$ становить $L_{\Sigma}^{min}(Z_2) = 700$.

При цьому, відповідно виразів (3.13-3.15) та визначеної на основі експериментальних даних тривалості однієї обчислювальної операції процесу навчання $\tau = 0{,}01\,\text{с}$, максимально допустима кількість навчальних прикладів для вибірки виду $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$ та $\boldsymbol{Z}_3$ становить $L_{\phi_k}^{max}(\boldsymbol{Z}_1) = 240000$, $L_{\phi_k}^{max}(\boldsymbol{Z}_2) = 864000$, $L_{\phi_k}^{max}(\boldsymbol{Z}_3) = 144000$.

Враховуючи, що допустимий термін формування навчальної вибірки $t_d \approx 0{,}5 \times 1\,рік = 1{,}5 \times 10^7\,c$ та використавши (3.17-3.20) визначено, що до множини допустимих відносяться навчальні вибірки виду $\boldsymbol{Z}_1$ та $\boldsymbol{Z}_3$. Недоцільність формування вибірки виду $\boldsymbol{Z}_2$ визначена з позицій занадто тривалого терміну розробки достатньої кількості продукційних правил розпізнавання фонем, що частково наведені в табл. 4.3.

Зазначимо, що в табл. 4.3 символом c_k позначено k-ий мел-кепстральний коефіцієнт. Визначено, що середній термін створення експертами одного навчального прикладу (продукційного правила) $\bar{t}_v(Z_3)$ =45000 c, а термін створення вибірки з мінімально допустимою кількістю прикладів $t_v(Z_2) = \bar{t}_v(Z_2) \times L_{\Sigma}^{min}(Z_2) = 45000 \times 700 = 3{,}15 \times 10^7\,c$, що перевищує t_d.

Таблиця 4.3

Приклади продукційних правил для розпізнавання фонем

Фонема	Продукційне правило
$[a]$	$c_1 \in [-99.8, -99.2]$ ^ $c_2 \in [0.65, 0.6$ $c_4 \in [3.3, 3.5]$ ^ $c_5 \in [7.1, 7.2]$ ^ $c_6 \in$
$[m]$	$c_1 \in [-181, -179]$ ^ $c_2 \in [2.9, 3.1]$ ^ $c_3 \in [5.5, 5.7]$ ^ $c_4 \in [1.1, 1.4]$ ^ $c_5 \in [11.9, 12.1]$ ^ $c_6 \in [-0.6, -0.4]$ ^ $c_7 \in [0.4, 0.5]$
$[\kappa]$	$c_1 \in [-174.1, -172.8]$ ^ $c_2 \in [-29.5, -$ $c_4 \in [8.3, 8.5]$ ^ $c_5 \in [-5.4, -5.1]$ ^ c_6

Процедура формування навчальних прикладів виду Z_1 та Z_3 реалізована відповідно 3-5 етапів методу створення навчальної вибірки. Визначено, що середній термін формування одного навчального прикладу виду Z_3 складає $\bar{t}_v(Z_3)=300\,c$. Тому термін створення вибірки виду Z_3 з мінімально допустимою кількістю навчальних прикладів становить $t_v(Z_3)=\bar{t}_v(Z_3)\times L_\Sigma^{min}(Z_3)=300\times 2750=8{,}25\times 10^5\,c$, що менше, ніж t_d.

Відповідно розробленої моделі формування параметрів навчальної вибірки, алфавіт фонем впорядковано наступним чином: $\Phi=\{\lfloor a\rfloor,\lfloor m\rfloor,\lfloor к\rfloor,\lfloor i\rfloor,\lfloor и\rfloor\}$. При цьому відправним пунктом впорядкування прийнята фонема $\lfloor a\rfloor$, найбільш схожою до неї визначена фонема $\lfloor m\rfloor$, після неї $\lfloor к\rfloor$, $\lfloor i\rfloor$ та $\lfloor и\rfloor$.

Зазначимо, що експерти оцінювали схожість фонем та визначали продукційні правила для їх розпізнавання за допомогою амплітудно-часових характеристик, приклад якої наведено на рис. 4.3. На рис. 4.3 по осі абцис відкладено номер відліку ГС (частота дискретизації 8000 Гц), а по осі ординат амплітуда ГС.

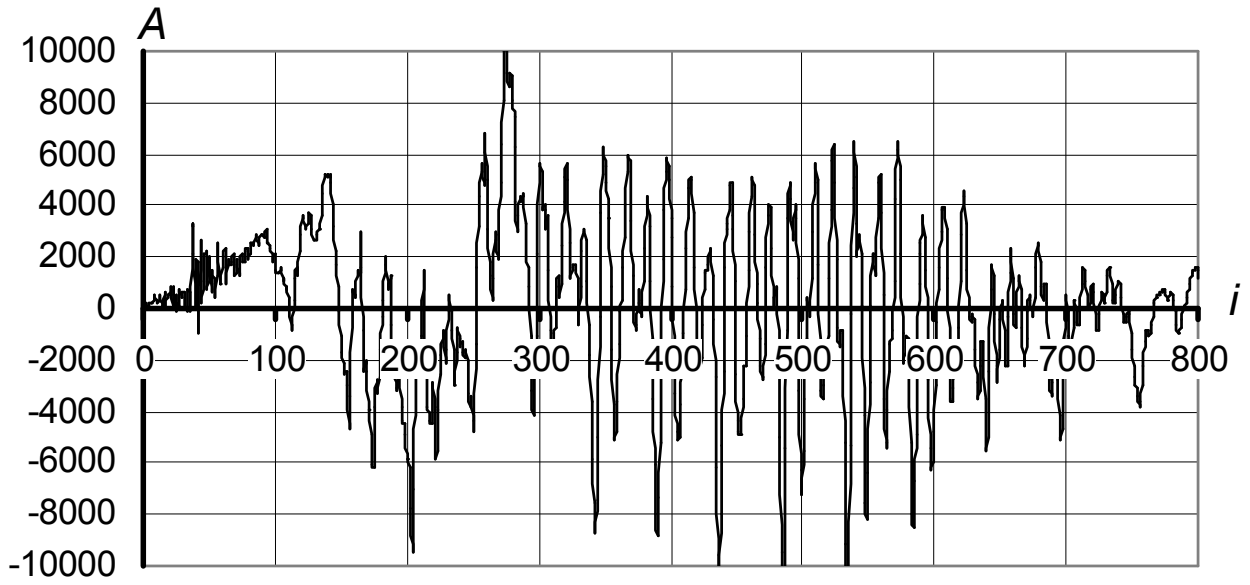

Рис. 4.3. Графік амплітудно-часової характеристики фонеми [*m*]

Визначено, що середній термін формування одного навчального прикладу виду Z_1 складає $\bar{t}_v(Z_1)=360\,c$. Тому мінімальний термін створення навчальної вибірки Z_1 становить $t_v(Z_1)=360\times 2750=9{,}9\times 10^5\,c$, що менше, ніж t_d.

Параметри навчальних прикладів для навчальної вибірки виду Z_1 та Z_2 частково наведені в табл. 4.4

Для перевірки якості попередньої обробки навчальних прикладів виду Z_1 за допомогою виразу (3.40) була розрахована константа Ліпшіца. Величина цієї константи $Lp = 37{,}7$ менша від порогового значення $K_{Lp} = 50$, що вказує на задовільну якість попередньої обробки.

Для перевірки ефективності розробленого методу створення навчальної вибірки були проведені експерименти з метою навчити БШП розпізнавати фонеми $[m],[a],[к],[н],[i]$.

Таблиця 4.4

Фрагмент параметрів навчальних прикладів виду Z_1 та Z_2

Фонема	Вихід НММ	Вхідні параметри НММ			
	y	x_1	x_2	x_3	x_4
[*а*]	0	0,440872442	0,835787164	0,723201374	0,846993373
[*м*]	0,03	0,131313804	0,841316833	0,851257859	0,834405861
[*к*]	0,07	0,156029915	0,715385081	0,814637881	0,862401439
[*і*]	0,09	0,003448567	0,942252523	0,950911538	0,880790742
[*н*]	0,12	0,049386689	0,922795287	0,930338382	0,895085228

БШП навчався як за допомогою створеної навчальної вибірки виду Z_1, так і за допомогою навчальної вибірки виду Z_g, що була створена відповідно загальновідомих методів [56]. Основною відмінністю навчальних прикладів, які входили до Z_g, стало кодування очікуваного вихідного сигналу НММ відповідно номеру букви в українському алфавіті.

Експерименти проводились за допомогою додатку NeuroPro та пакету Mathcad [37]. Результати експериментів показали, що для досягнення допустимої середньої похибки навчання $\varepsilon \approx 0{,}01$ при використанні вибірки виду Z_1 потрібна кількість навчальних ітерацій приблизно в 2,4 разів менша, ніж при використанні вибірки виду Z_g, що підтверджує справедливість сформульованої гіпотези про ефективність запропонованого методу створення навчальної вибірки.

На другому етапі досліджень була перевірена гіпотеза про те, що застосування розробленого методу нейромережевого розпізнавання фонем дозволить створити на його основі НМС розпізнавання фонем, ефективність якої вища, ніж у відомих НМС та достатня для застосування в поширених КС.

Для перевірки гіпотези використано метод оцінки ефективності застосування НМЗ. Результати досліджень наведено в табл. 4.5. В процесі досліджень прийняті наступні величини коефіцієнтів значимості критеріїв ефективності $A^{nz} = \{0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 0.7, 1, 1\}$. Збільшення коефіцієнтів значимості критеріїв ефективності u_8 та u_9 пояснюється тим, що виконання відповідних процедур визначає принципову можливість застосування НМЗ для розпізнавання ГС. Прийнято, що мінімально допустимі величини всіх критеріїв ефективності дорівнюють 0,5.

Аналіз даних табл. 4.5 дозволяє стверджувати, що використання розробленого методу дозволяє приблизно в 1,2 рази підвищити ефективність застосування НМЗ розпізнавання фонем в ГС по відношенню до найкращих відомих НМЗ. Також визначено, що ефективність створеної НМСРФ достатня для застосування у поширених КС, а основним напрямком її вдосконалення являється підвищення ефективності методу навчання.

Таблиця 4.5

Оцінка ефективності НМЗ розпізнавання ГС

№	Назва	Критерії та функція ефективності									
		u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9	V^{nz}
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1	ААРС	0	0	0	0	0	0,5	0	0	0	0,35
2	СРГК	0	0	1	0,5	0,5	0	0	0	0	1,4
3	НЗЧ	1	0	1	0,5	1	0	0	0	0	2,45
4	ПНЗГА	0	0	1	0	0	0	0	0	0	0,7
5	ЗНМ	1	0	1	0	1	0	0	0	0	2,1
6	АРІС	1	0,5	1	0,5	1	0	0	0,5	0	3,3
7	МДСДВ	1	0,5	1	0	0	0	0	0	0	1,75

8	МАРІС	1	0	0	0	0	0	0	0	0	0,7
9	МРФГС	1	0	1	1	0,5	0	0	0	0	2,45
10	НМАМ	1	0,5	1	1	1	0	0	0,5	1	4,65
11	MSNET	1	0,5	1	1	1	0	1	0	1	4,85
12	ПСВ	1	0,5	0	0	1	0	0	0	0	1,75
13	ІНМ	1	0,5	1	0,5	0	0	0	0	0	2,1

Таблиця 4.5 (продовження)

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
14	МНБСС	1	0	1	0	0	1	0	0	0	2,1
15	ППШН	1	0	1	0,5	0	1	0	0	0	2,45
16	НАРМ	0	0	1	0	0	0	0	0	0	0,7
17	КПОМС	1	0	1	0	0	0	0	0	0	1,4
18	ОВМО	1	0,5	1	0	0	1	0	0	0	2,45
19	МНСМО	1	0	1	0	0	1	0	0	0	2,1
20	НПІПІ	1	0	1	0	1	0	0	0,5	0	2,6
21	ННМРФ	1	0	1	0	0	0	0	0	0	1,4
22	АРМНТ	1	0	1	0	1	0	1	0	0	2,8
23	ВІОМС	1	0	0	0	1	0	0	0	0	1,4
24	МППРМ	1	0,5	1	0	1	0	0	0	0	2,45
25	МКФА	1	0,5	1	0	0	0	0	0	0	1,75
26	САПКС	1	0,5	1	0	1	0	0	0	0	2,45
27	НМРЗМ	1	0,5	1	0	1	0	0	0	0	2,45
28	ГНМ	1	1	1	1	0,5	1	1	0,5	0,5	5,55
29	САРФ	0	0	1	0	1	0	0	0	0	1,4
30	**НМСРФ**	**1**	**1**	**1**	**1**	**0,5**	**1**	**1**	**1**	**1**	**6,55**

Третій етап досліджень був спрямований на перевірку гіпотези про адекватність розробленої нейромережевої системи розпізнавання до варіативності очікуваних умов застосування в КС. Адекватність оцінювалась за допомогою відповідності величин помилки розпізнавання фонем та навантаження на модуль розпізнавання ГС до встановлених вимог. Варіативність умов застосування визначалась зміною рівня шуму в приміщенні, зміною відстані користувача до мікрофону та зміною кута між віссю мікрофону і користувачем.

З використанням методу нейромережевого розпізнавання фонем визначено, що найбільш ефективним видом НММ є ДШП, у якого кількість вхідних нейронів $N_x = 217$, кількість вихідних нейронів $N_y = 1$, а кількість схованих нейронів $N_s = 20$.

Термін навчання такого ДШП на навчальний вибріці виду Z_1 з мінімально допустимою кількістю навчальних прикладів дорівнює приблизно 200 с. Також визначено, що допустима величина похибки розпізнавання дорівнює $\varepsilon = 0{,}01$, допустимий рівень шуму 30 дБ, обслуговування одного клієнтського запиту пов'язане з використанням 1 % обчислювальних потужностей КС, максимально допустимий обсяг використання обчислювальних ресурсів КС $\xi^{\partial} = 80\%$. Проведені експерименти спрямовані на дослідження можливості розробленої НМСРФ по прогнозуванню кількості запитів до модулю розпізнавання ГС. Графіки залежності реальної та прогнозованої кількості переглядів веб-сайту одного із вищих вітчизняних навчальних закладів від часу показано на рис. 4.4. На рис. 4.4 цифрою 1 позначено графік, побудований на статистичних даних, а цифрою 2 – графік, побудований за допомогою моделі прогнозування кількості запитів модулю розпізнавання. По осі ординат цього графіку відкладено кількість запитів, а по осі абцис – термін в днях.

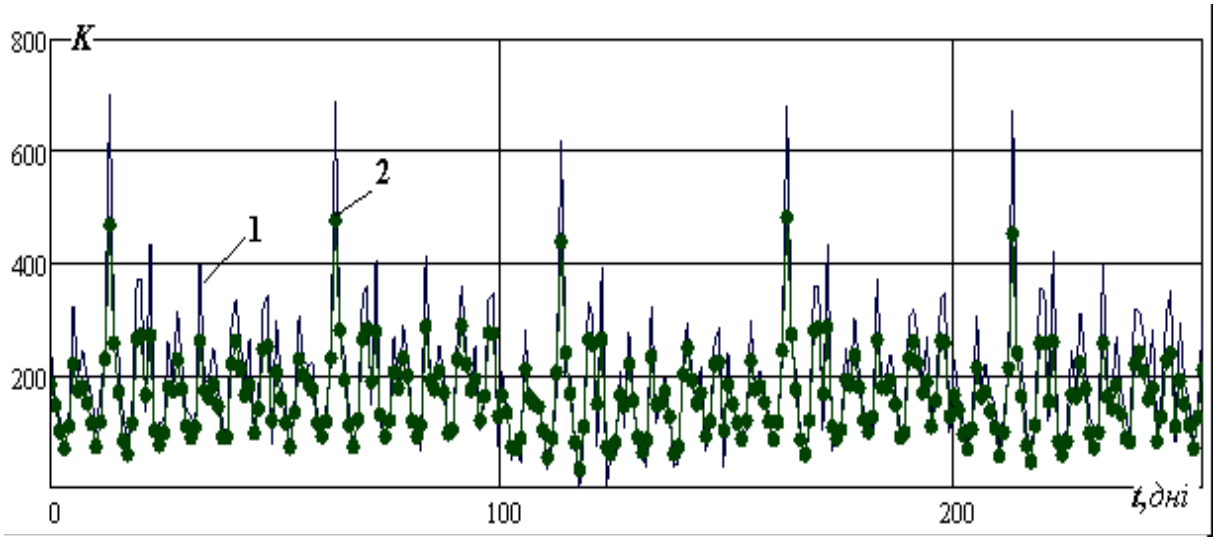

Рис. 4.4. Графіки залежності переглядів веб-сайту від часу

Аналіз графіків рис. 4.4 вказує на те, що величина максимальної приведеної похибки прогнозування не перевищує 0,1. В той же час максимальна приведена похибки розпізнавання, отримана за допомогою загальнопоширених поліноміальних моделей, становить приблизно 0,12. Таким чином, застосування розробленої моделі прогнозування кількості запитів дозволяє зменшити в 1,2 приведену похибку прогнозування. Крім того, аналіз експериментів вказує на те, що очікувана максимальна кількість одночасних звернень до веб-оієнтованого модулю розпізнавання ГС не перевищить 70.

Даний результат отримано, виходячи з того, що максимальна кількість всіх звернень до КС не перевищує $k_{\Sigma} = 700$, до НМС відносяться тільки 0,1 звернень від k_{Σ}. Таким чином, можливо зробити висновок, що при очікуваній кількості звернень розроблена НМС розпізнавання ГС здатна виконувати свої функції.

Відповідно до визначеної варіативності умов застосування, проведені експерименти по розпізнаванню фонем з алфавіту $[m],[a],[к],[и],[i]$.

Варіативність визначена зміною рівня шуму в межах від 10 до 80 дБ, зміною відстані від користувача до мікрофону від 0,01 до 1 метра та зміною кута між віссю мікрофону та джерелом мовлення від 0 до 180 градусів.

Межі варіативних змін визначені відповідно результатів [13, 22] та практичного досвіду. Наприклад, рівень шуму 10 дБ відповідає звуку від спокійного дихання людини, а шум 80 дБ відповідає рівню шуму від

кухонного комбайну.

Запис ГС проводився послідовно 6 дикторами: 3 чоловіками та 3 жінками віком від 18 до 22 років, що не мали дефектів голосу та слуху. ГС був озвучений рівним голосом 10 разів кожним диктором. Рівень гучності ГС в 50 дБ зберігався на протязі всього запису.

Для розпізнавання фонем використано програмні додатки AnalizWav та Google Chrome. Це пояснюється тим, що AnalizWav реалізує запропоновані рішення. Google Chrome використовувався для порівняння отриманих результатів. Графіки експериментально отриманих залежностей помилки розпізнавання ε фонем від рівня шуму A наведено на рис. 4.5-4.6. Також результати експериментів дозволили побудувати показані на рис. 4.7-4.8 графіки залежності величини похибки розпізнавання від співвідношення гучності ГС до шуму A'.

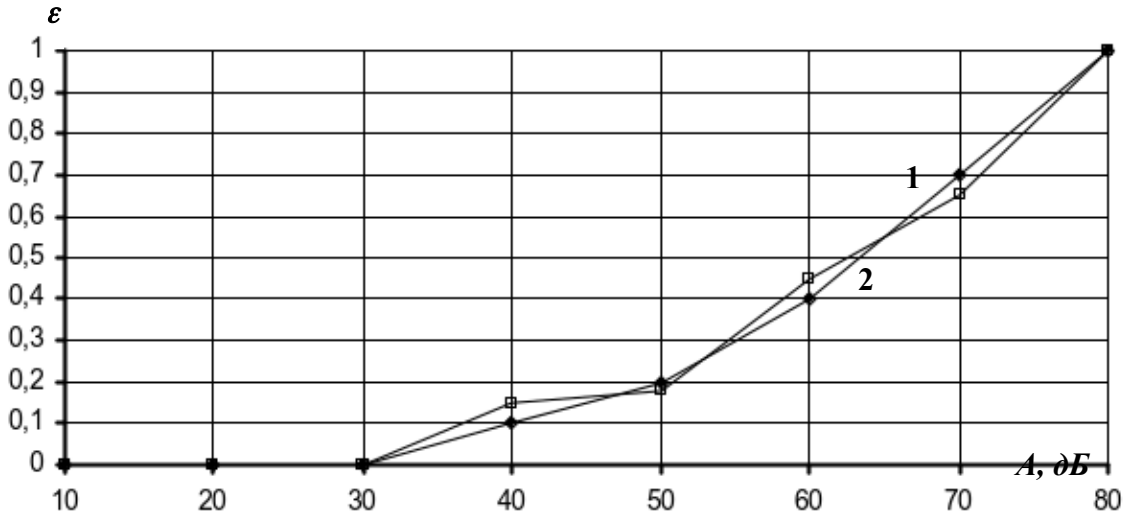

Рис. 4.5. Графік залежності помилки розпізнавання фонем від рівня шуму, викликаного кондиціонером

Рис. 4.5, 4.6 відповідають залежностям, коли шум викликаний кондиціонером. Залежності, показані на рис. 4.7, 4.8, побудовані при шумі, що викликаний трансляцією новин.

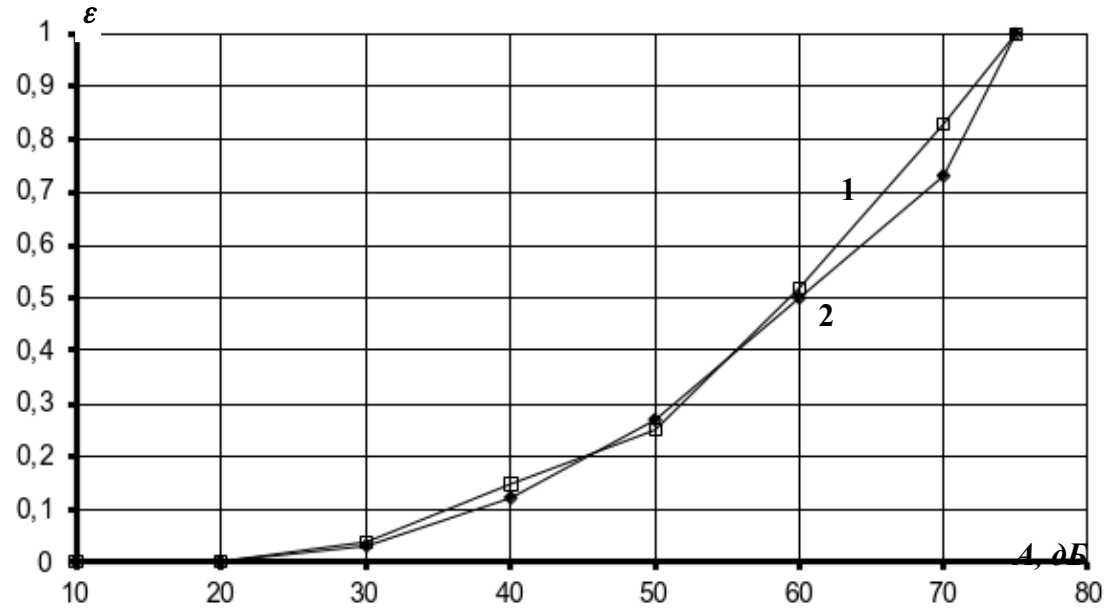

Рис. 4.6. Графік залежності помилки розпізнавання фонем від рівня шуму, викликаного телевізійною трансляцією новин.

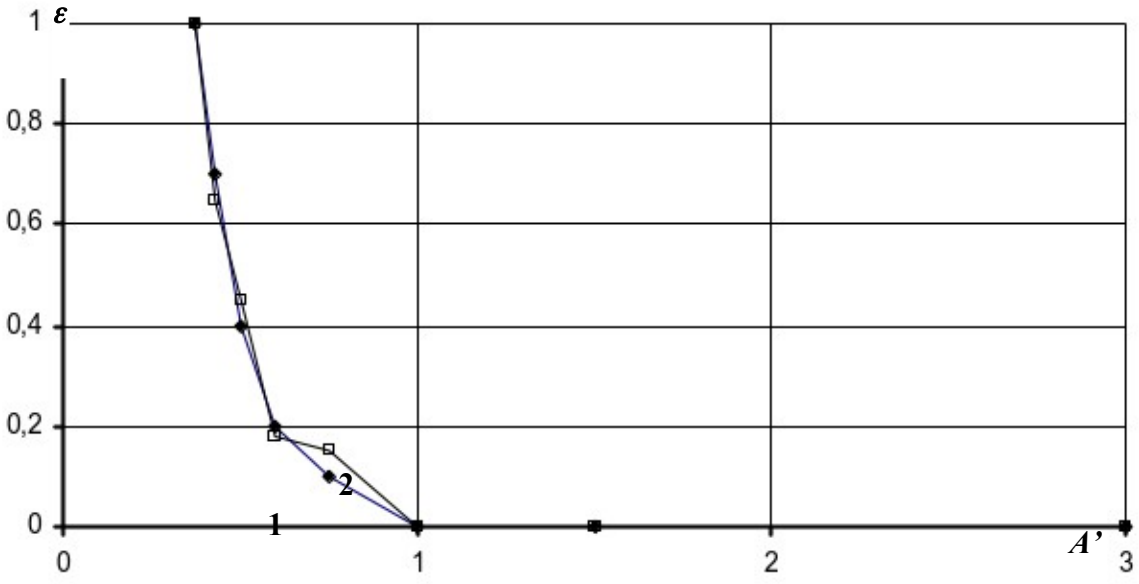

Рис. 4.7. Графік залежності помилки розпізнавання від співвідношення рівня ГС до шуму, що викликаний кондиціонером

На рис. 4.5-4.8 цифрою 1 позначено графік, що відповідає Google Chrome, а цифрою 2 позначено графік, що відповідає AnalizWav.

Також слід зазначити, що Google Chrome не зміг розпізнати ізольовано вимовлені фонеми $[m]$ та $[к]$, однак здатен розпізнавати слова «Так» та «Ні». Тому на графіках показані результати розпізнавання Google Chrome не окремих фонем, а саме цих слів.

ε

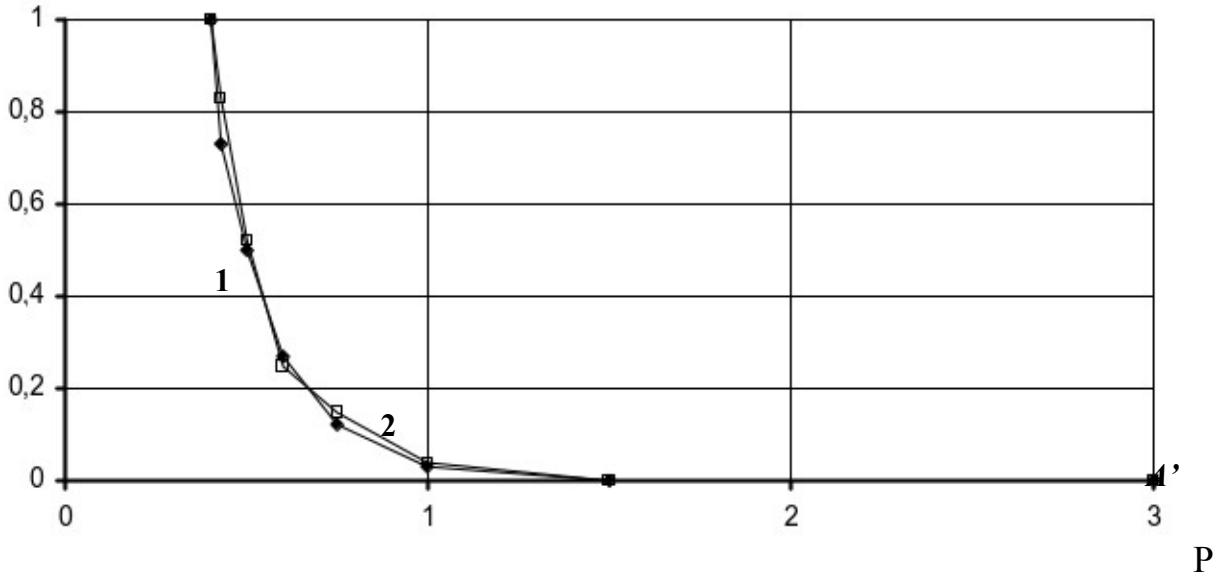

Рис. 4.8. Графік залежності помилки розпізнавання від співвідношення рівня ГС до шуму, що викликаний телевізійною трансляцією новин

Аналіз наведених графіків показує, що в очікуваних умовах застосування помилка розпізнавання фонем за допомогою AnalizWav знаходиться в допустимих межах і не суттєво відрізняється від похибки розпізнавання за допомогою Google Chrome.

Крім того, проведені експерименти показали, що зміна у визначених межах відстані та напрямку до мікрофону не впливає на точність розпізнавання фонем. При цьому в очікуваних умовах створена НМС здатна обслуговувати до 90 клієнтів, що задовольняє вимогам допустимого навантаження на модуль розпізнавання ГС.

Таким чином, результати досліджень, наведених в навчальному посібнику, підтверджують можливість підвищення ефективності модулю розпізнавання голосового сигналу за рахунок застосування розроблених нейромережевих моделей та методів розпізнавання фонем в голосовому сигналі.

В підсумку аналіз результатів проведених експериментальних досліджень дозволяє стверджувати:

-використання методу створення навчальної вибірки дозволяє в 2,4 рази зменшити кількість навчальних ітерацій нейромережевої моделі для досягнення допустимої помилки навчання;

- використання методу нейромережевого розпізнавання фонем дозволяє приблизно в 1,2 разів підвищити ефективність нейромережевих засобів розпізнавання фонем в голосовому сигналі;

-при очікуваних умовах застосування розроблена нейромережева система дозволяє забезпечити похибку розпізнавання фонем в голосовому сигналі в межах 0,01, що достатньо для практичного застосування.

Список використаних джерел

1. Андерсон Т. Статистический анализ временных рядов / Андерсон Т. ; пер. с. англ. И. Г. Журбенко. - М. : МИР, 1976. - 757 с.

2. Барский А. Б. Нейронные сети: распознавание, управление, принятие решений / Барский А. Б. - М. : Финансы и статистика, 2004. - 176 с.

3. Бондарко Л. В. Основы общей фонетики / Л.В. Бондарко, Л.А. Вербицкая, М.В. Гордина. – СПб.: Академия, 2004. – 160 с.

4. Веренич И. В. Анализ методов построения систем распознавания речи на основе нейросетевых и скрытых марковских моделей / И. В. Веренич, О. И. Федяев // III международная научно-техническая конференция молодых ученых и студентов «Информатика и компьютерные технологии», Донецк, ДонНТУ, 11-13 декабря 2007 г. : тезисы докл. – Донецк., 2007 – С. 126 – 129.

5. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк – К. : Наукова думка. – 1987. – 262 с.

6. Вишнякова О. А. Автоматическая сегментация речевого сигнала на базе дискретного вейвлет-преобразования / О. А. Вишнякова, Д. Н. Лавров // Математические структуры и моделирование. – 2011. – № 23. – С. 43 – 48.

7. Галунов Г. В. К вопросу об использовании самоорганизующихся карт Кохонена для квантования пространства признаков речевого сигнала / Г. В. Галунов, Долгобородов А. Н. // IV Всероссийская конференция «Нейрокомпьютеры и их применение» : науч.-практ. конф., 16-18 февр. 2000 г. : сб. докл. – М., 2000. – С.111 – 115.

8. Галушкин А. И. Теория нейронных сетей / А. И. Галушкин. - М. : ИПРЖР, 2000. - 416 с.

9. Головинов М. В. Об автоматическом распознавании речи / М. В. Головинов // III Международная научно-практическая конференция «Информатизация общества» : междунар. науч.-практ. конф., 28-30 июня 2012 г. : тезисы докл. – М., 2012. – С. 132 – 136.

10. Гребнов С. В. Аналитический обзор методов распознавания речи в системах голосового управления / С. В. Гребнов // Вестник ИГЭУ. – 2009. –

№3. – С. 22 – 25.

11. Дмитриев В.Т. Дикторонезависимая система автоматического поиска ключевых слов в потоке слитной речи, устойчивая к акустическим шумам / В. Т. Дмитриев, И. В. Баландин // Вестник РГРТУ. – 2008. – № 2 (выпуск 24). – С. 17 – 22.

12. Елистратов С. А. Сравнение параметров для выделения вокализованных сегментов и классификации гласных фонем / С. А. Елистратов, М. А. Косенко, А. А. Чичерин // Доклады ТУСУР. – 2012. – № 1 (25). – С.171 – 174.

13. Жиляков Е. Г. Исследование сервиса компании google inc. по распознаванию русской речи / Е. Г. Жиляков, С. Л. Бабаринов, П. В. Чадюк // Научные ведомости Белгородского Государственного Университета. – 2013. №15 (158). – С. 247 – 255.

14. Жирков А.О. Нейросетевой анализ и сопоставление частотно-временных векторов на основе краткосрочного спектрального представления и адаптивного преобразования Эрмита / Жирков А.О., Корчагин Д.Н., Лукин А.С., Крылов А.С., Баяковский Ю.М. – М. ИПМ им. М.В.Келдыша РАН, 2001. – 5 с. – (Препринт / РАН, Ин-т прикл. матем.; ИПМ 2001).

15. Иванов А. В. Методы построения устройств распознавания речи на базе гибрида нейронная сеть/скрытая марковская модель / А. В. Иванов, А. А. Петровский // Нейрокомпьютеры: разработка, применение. – 2007. – № 12. – С. 26 – 36.

16. Каллан Р. Основные концепции нейронных сетей / Каллан Р. ; пер. с англ. А. Г. Сивака. - М. : Вильямс, 2003. - 288 с.

17. Кладов С. А. Распознавание голосовых команд с помощью самоорганизующейся нейронной сети Кохонена. [Электронный ресурс] / С. А. Кладов // Молодежный научно-технический вестник – 2012 – № 05. – Режим доступу до журн. : http://sntbul.bmstu.ru/doc/ 458.html.

18. Козловський В. В. Нечіткі нейронні продукційні мережі в системах захисту інформації / В. В. Козловський, Ю. І. Хлапонін, А. В. Міщенко // Безпека інформації. – 2014. – № 3 – С. 231 – 235.

19. Конев А. А. Сегментация речевого сигнала / А. А. Конев, А. А.

Пономарёв // Сборник трудов XVI сессии Российского акустического общества.– 2005. – № 3. – С. 44 – 47.

20. Корченко А. Нейросетевые модели, методы и средства оценки параметров безопасности Интернет-ориентированных информационных систем: монографія / А. Корченко, И. Терейковский, Н. Карпинский, С. Тынымбаев. - К. : ТОВ «Наш Формат». - 2016. – 275 с.

21. Костюченко Е. Ю. Обработка естественной информации на основе аппарата нейронных сетей / Е. Ю. Костюченко // Сборник научных трудов ТГУСУР. – 2009. – № 1(19). – С. 54 – 56.

22. Костюченко Е. Ю. Критерии информативности при обработке биометрических сигналов при помощи нейронных сетей / Е. Ю. Костюченко, Р. В. Мещеряков, А. Ю. Крайнов // Сборник научных трудов ТГУСУР. – 2010. – № 1(21). – С. 118 – 120.

23. Котомин А. В. Распознавание речевых команд с использованием сверточных нейронных сетей / А. В. Котомин // Наукоемкие информационные технологии. – 2012. – № 1 . – С.12 – 24.

24. Кручинин А. Ю. Распознавание фонем речи на основе анализа формы спектров сигналов / А. Ю. Кручинин // Математическое моделирование, обратные задачи, информационно-вычислительные технологии : междунар. науч.-техн. конф., 26-27 дек. 2007 г. : тезисы докл. – Пенза, 2007. – С. 169 – 171.

25. Ле Н. В. Распознавание речи на основе искусственных нейронных сетей / Н. В. Ле, Д. П. Панченко // Технические науки в России и за рубежом: междунар. науч. конф., 8-9 окт. 2011 г. – М., 2011. – С. 8 – 11.

26. Мещеряков Р.В. Система оценки качества передаваемой речи // Доклады ТУСУРа. – 2010. – № 2(22). – С. 324 – 329.

27. Мисюрёв А. В. Использование искусственной нейронной сети для оценки близости векторов акустических параметров / А. В. Мисюрёв, А. Я. Подрабинович, А. В. Брухтий // Интеллектуальные технологии ввода и обработки информации. – 1999. – С. 94 – 98.

28. Міхайленко В. М., Терейковська Л. О., Терейковський І. А., Ахметов Б. Б. Нейромережеві моделі та методи розпізнавання фонем в

голосовому сигналі в системі дистанційного навчання : монографія. Київ : Компринт, 2017. 252 с.

29. Музычук Д. С. Сегментация, шумоподавление и фонетический анализ в задаче распознавания речи / Д. С. Музычук, М. С. Медведев // Молодой ученый. – 2013. – № 6 (53). – С. 86 –95.

30. Музычук Д. С. Использование преобразования Гильберта-Хуанга для формирования моделей фонем русского языка в задаче распознавания речи / Д. С. Музычук, М. С. Медведев // Молодой ученый. – 2013. – № 6 (53). – С. 75 – 85.

31. Овчинников П.Е. Обучение персептрона без сегментации слов из обучающей выборки в задаче распознавания звуков / П. Е. Овчинников, Ю. А. Сёмин // Нейроинформатика. – 2006. – № 3. – С. 212 – 216.

32. Осовский С. Нейронные сети для обработки информации / С. Осовский. - М. : Финансы и статистика, 2002. – 344 с.

33. Павлов А. А. Многокритериальный выбор в задаче обработки данных матрицы парных сравнений / А. А. Павлов, Е. И. Лищук, В. И. Кут // - Вісник національного технічного університету України “КПІ”. Інформатика, управління та обчислювальна техніка. – 2007. – № 46. – С. 42 – 48.

34. Пат. 2 268 504 Российская Федерация, МПК G10L 15/06 (2006.01) G10L 11/04 (2006.01). Способ распознавания фонем речи и устройство для реализации способа / Гиголо Л. А., Сахаров В. О.; заявитель и патентообладатель Открытое акционерное общество «Корпорация Фазотрон». – № 2004109253/09 ; заявл. 30.03.2004 ; опубл. 20.01.2006, Бюл. № 02.

35. Рабинер Л. Р. Цифровая обработка речевых сигналов / Л. Р. Рабинер, Р. В. Шафер. – М. : Радио и связь. – 1981. – 496 с.

36. Распознавание речи от Яндекса. Под капотом у Yandex.SpeechKit. [Электронный ресурс] // Хабрахабр – 2013. – Режим доступа к журн.: http://habrahabr.ru/company/yandex/blog/198556.

37. Резник А. М. О природе интеллекта / А. М. Резник // Математические машины и системы. – 2008. - № 1. – С.23 – 45.

38. Руденко О.Г. Штучні нейронні мережі. / О. Г. Руденко, Є. В.

Бодянський. – Харків: ТОВ "Компанія СМІТ", 2006. – 404 с.

39. Саакян А. А. Исследование свойств показателей качества систем распознавания речи / А. А. Саакян // Проблемы управления. – 2009. – № 4. – С. 66 – 73.

40. Савченко Л. В. Автоматическое распознавание слогов на основе линейной авторегрессионной нейронной сети и теории нечетких множеств / Л. В. Савченко // Нейроинформатика. – 2013. – № 1. – С. 52 – 62.

41. Савченко В. В. Фонема как элемент информационной теории восприятия речи // Известия вузов России. Радиоэлектроника. – 2008. – Вып.4. – С.3 – 11.

42. Савченко В. В. Метод фонетического декодирования слов в задаче автоматического распознавания речи на основе принципа минимума информационного рассогласования. // Известия вузов. Радиоэлектроника. – 2009. – Вып. 5. – С. 31 – 41.

43. Сорокин В. Н. Сегментация и распознавание гласных / В. Н. Сорокин, А. И. Цыплихин // Информационные процессы. – 2010. – № 2. – С. 202 – 220.

44. Тарасенко В. П. Метод застосування продукційних правил для подання експертних знань в нейромережевих засобах розпізнавання мережевих атак на комп'ютерні системи / В. П. Тарасенко, О. Г. Корченко, І. А. Терейковський // Безпека інформації. – 2013. – Т. 19, № 3. – С. 168 – 174.

45. Тейлор Дж. Введение в теорию ошибок / Тейлор Дж. ; пер. с англ. Л. Г. Деденко. - М. : Мир, 1985. - 272 с.

46. Темников В. А. Параметризация автоматического контроля доступа операторов к ресурсам информационных систем по голосу / В. А. Темников, Е. Л. Темникова // Вестник Восточноукраинского национального университета им.В.Даля. - №9 (151). – Ч.1. – 2010. – С.143 – 148.

47. Терейковская Л. А. Определение оптимальной архитектуры нейронной сети, предназначенной для распознавания голосовых сигналов / Л. А. Терейковская // Всеукраїнська науково-практична конференція «Стратегії розвитку інформаційного культурно-освітнього та економічного простору України»: наук.-практ. конф., Київ, 20-21 травня 2014 р. : зб. тез допов. –

С.132 – 134.

48. Терейковська Л. О. Нейромережева модель розпізнавання фонем за допомогою експертних знань / Л. О. Терейковська, І. А. Терейковський // Друга міжнародна науково-практична конференція «Управління розвитком технологій»: міжнар. наук.-практ. конф., Київ, 21–23 травня 2015 р. : тези допов - С. 90 – 92.

49. Терейковська Л.О. Визначення найбільш ефективної архітектури нейронної мережі, призначеної для розпізнавання голосових сигналів в Moodle / Л.О. Терейковська, І.А. Терейковський // Теорія і практика використання системи управління навчанням Moodle: матер. 2 міжнар. наук.-практ. конф. "MoodleMoot Ukraine-2014", (22-23 травня 2014 р.). – К.: КНУБА, 2014. – С.36.

50. Терейковська Л.О. Використання голосової взаємодії в Moodle / Л. О. Терейковська, І. А. Терейковський // Теорія і практика використання системи управління навчанням Moodle: матер. 1 міжнар. наук.-практ. конф. “MoodleMoot Ukraine-2013”, (30–31 травня 2013 р.).– К. : КНУБА, 2013. – С. 48.

51. Терейковська Л. О. Використання схованих марківських процесів для розпізнавання голосових сигналів в системі дистанційного навчання / Л. О. Терейковська // XXI Всеукраїнська наукова конференція «Сучасні проблеми прикладної математики та інформатики», Львів, 24–25 вересня 2015 р.: зб. наук. праць - С. 304 – 306.

52. Терейковська Л. О. Обмеження використання схованих марківських моделей для розпізнавання голосових сигналів в системі дистанційного навчання Moodle / Л. О. Терейковська // Буд-майстер-клас-2015, Київ, 26-27 листопада 2015 р.: зб. тез допов. - С. 197.

53. Терейковський І. Нейронні мережі в засобах захисту комп’ютерної інформації: монографія / І. Терейковський. - К. : ПоліграфКонсалтинг. - 2007. – 209 с.

54. Титов Ю.Н. Математическая модель органа слуха для автоматического распознавания речи / Ю.Н. Титов // Науч.-техн. вестн. СПбГУ ИТМО – 2016. – Т. 16, №1. – С. 307 – 310.

55. Федяев О. И. Анализ эффективности метода нечёткого сопоставления образов для распознавания изолированных слов / О. И. Федяев, И. Ю. Бондаренко // VI междунар. науч. конф. “Интеллектуальный анализ информации ИАИ-2006”, Киев, 16-19 мая 2006 г : тезисы докл. – К., 2006. – С. 112 – 117.

56. Федяев О.И. Нейросетевой интерпретатор речевых команд для управления программными системами / О. И. Федяев, С. А. Гладунов // VII Всероссийская конференция “Нейрокомпьютеры и их применение”, Москва, 14-16 февраля 2001 г. : тезисы докл. – М., 2001 – С. 298 – 301.

57. Федяев О. И. Фонетический анализ речи на основе нейросетевой аппроксимации сигнала / О. И. Федяев, С. А. Гладунов // VIII Всероссийская конференция «Нейрокомпьютеры и их применение», Москва, НКП-2002, 21-22 марта 2002 г.: тезисы докл. – М., 2002 – С. 435 – 438.

58. Фролов А. В. Синтез и распознавание речи. Современные решения [Электронный ресурс] / А.В. Фролов, Г.В. Фролов // Режим доступа: http://www.frolov-lib.ru/books/hi/.

59. Хайкин С. Нейронные сети: полный курс, 2-е изд., испр. / Хайкин С. ; пер. с англ. Н. Н. Куссуль - М. : Вильямс, 2006. - 1104 с.

60. Цюцюра С. В. Застосування нейронних мереж для розпізнавання «ідеального співрозмовника» серед користувачів соціальних мереж / С. В. Цюцюра, І. А.Терейковський, С. В. Палій // Системи навігації та управління. – 2013. – Випуск 4(28). – С.123 – 127.

61. Цюцюра С. В. Модифікація класичної нейронної мережі ймовірнісного типу для розпізнавання "ідеального співрозмовника" серед користувачів соціальних мереж / С. В. Цюцюра, І. А.Терейковський, С. В. Палій // Науково-технічний збірник “Управління розвитком складних систем” Київського національного університету будівництва і архітектури. – 2014. – Випуск 19. – С 118 – 123.

62. Червяков Н. И. Использование вейвлетов для улучшения параметров нейронных сетей в задачах распознавания речи / Н. И. Червяков, К. А. Астапов // Инфокоммуникационные технологии. – 2009. – № 4, С. 9 – 12.

63. Щербина А. А. Метод определения нейросетевой архитектуры в

задачах голосового взаимодействия дистанционного обучения / А. А. Щербина, Л. А. Терейковская // Науково-технічний збірник “Управління розвитком складних систем” Київського національного університету будівництва і архітектури. – 2014. – Вип. 17. – С. 148 – 155.

64. Ali A. An acoustic-phonetic feature-based system for automatic phoneme recognition in continuous speech / A. Ali // Journal of the Acoustical Society of America/ – 1998. – 103(5). – P. 2777 – 2778.

65. Cucu H. SMT-based ASR domain adaptation methods for under-resourced languages: Application to Romanian / H. Cucu, A. Buzo, L. Besacier, C. Burileanu. // Speech Communication archive. – North-Holland, 2014. – Vol.56. – P.195 – 212.

66. Geoffrey Hinton Deep Belief Nets. [Електронний ресурс]. – Режим доступу: http://www.cs.toronto.edu/~hinton/nipstutorial/nipstut3.pdf.

67. Hariharan R. Web Server Performance Modeling / R. Hariharan, Van der Mei, P. K. Reeser. // Telecomunication Systems. – 2001. – vol. 16. – P. 361 –378.

68. Hahn S. Comparison of Grapheme-to-Phoneme Methods on Large Pronunciation Dictionaries and LVCSR Tasks / S. Hahn, P. Vozila, M. Bisani. // Interspeech 2012, ISCA. – Portland, 2012. – P.2537 – 2540.

69. Karpov A. Large vocabulary Russian speech recognition using syntactico-statistical language modeling / A. Karpov, K. Markov, I. Kipyatkova, D. Vazhenina, A. Ronzhin. // Speech Communication. – North-Holland, 2013. – Vol.56. – P.213 – 228.

70. Kozlovskyi V. Generalized Algorithm for Synthesis of Protective Coating against / Kozlovskyi V., Aleksander M., Karpinski M., Mischenko A. Chlaponin Y. // Electromagnetic Radiation Technika transportu scynovego. – 2015. – № 12 (261) – P. 54 – 56.

71. Kulyk M. Using Of Fuzzy Cognitive Modeling in Information Security Systems / Kulyk M., Khoma V., Kozlovskyi V., Chlaponin Y. // Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications : International Conference, 24-26 September, 2015, Warsawa, 2015. – P. 408 – 411.

72. Tereykovskaya L. A. Using the expertise in the development of neural network model for recognition of phonemes in the voice signal / L. A.

Tereykovskaya, Tereykovskiy I. A. // The proceedings of the II International scientific-practical conference «Information and telecommunication technologies: education, science and practice». 3-4 December, 2015, Almaty, Kazakhstan. – 2015. - P. 258 – 261.

73. Tereykovskaya L. Prospects of neural networks in business models / L. Tereykovskaya , O. Petrov , M Aleksander . // TransComp 2015. 30 November – 3 December, 2015, Zakopanem, Poland. – P. 1539 – 1545.

74. Waibel A. Phoneme Recognition Using Time-Delay Neural Networks / A. Waibel, T. Hanazawa, G. Hinton // Transaction on acoustics, speech and signal processing – 1989. – V. 37, №3. – P. 328 – 339.

75. Yu H. A Neural Network using Non-Uniform Unit for Continuous Speech Recognition / Yu H., Oh Y. // EU-Rospeech'95, Madrid, Spain, September 1995. – Vol 3. – P. 1677 – 1680.