ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

**Гладунов Сергій Анатолійович**

УДК 681.3.16

# АПАРАТНО-ПРОГРАМНІ ЗАСОБИ РОЗДІЛЬНОЇ ЛОКАЛІЗАЦІЇ ФОНЕМ В СИСТЕМАХ МОВНОЇ ВЗАЄМОДІЇ ЛЮДИНИ З ЕОМ

05.13.13 – Обчислювальні машини, системи та мережі

АВТОРЕФЕРАТ
дисертації на здобуття наукового ступеня
кандидата технічних наук

Донецьк – 2005

**Дисертацією є рукопис**

Робота виконана в Донецькому національному технічному університеті Міністерства освіти і науки України

| | |
|---|---|
| Науковий керівник: | кандидат технічних наук, доцент Федяєв Олег Іванович, доцент кафедри "Прикладна математика і інформатика" Донецького національного технічного університету. |
| Офіційні опоненти: | доктор технічних наук, професор Кривуля Геннадій Федорович, завідувач кафедри "Автоматичне проектування обчислювальної техніки" Харківського національного університету радіоелектроніки. |
| | кандідат технічних наук, доцент Вороний Сергій Михайлович, завідувач кафедри "Технічна інформатика" Донецького державного інституту штучного інтелекту. |
| Провідна установа: | Інститут космічних досліджень НАН України – НКА України, відділ космічних інформаційних технологій, м. Київ. |

Захист відбудеться 9 червня 2005 р. о 14 годині на засіданні спеціалізованої вченої ради К 11.052.03 Донецького національного технічного університету (адреса: 83000, м. Донецьк, вул. Артема, 58, 8 навч. корпус, ауд. 705).

З дисертацією можна ознайомитися в бібліотеці ДонНТУ за адресою: 83000, м. Донецьк, вул. Артема, 58, 2 навч. корпус.

Автореферат розісланий “05” травня 2005 р.

Вчений секретар
спеціалізованої вченої ради
К 11.052.03
к.т.н., доц.
Г. В. Мокрий

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Широке розповсюдження засобів обчислювальної техніки в багатьох областях людської діяльності зробило актуальною проблему створення засобів взаємодії людини з ЕОМ, що дозволяють підвищити зручність і ефективність вводу-виводу інформації. Один з підходів до розробки таких засобів заснований на мовній взаємодії як на найбільш звичному для людини способі спілкування. Організація мовного обміну інформацією вимагає ефективного рішення задачі автоматичного розпізнавання мови. Даній проблемі присвячено багато академічних досліджень (University California, Berkley; University Birmingham, University Edinburg та ін.), запропонований ряд комерційних рішень (лідерами тут є такі гіганти, як IBM, Phillips і Microsoft), в Україні серйозні дослідження за різних часів велися в Києві, Харкові, Дніпропетровську, Вінниці. В даний час в Інституті кібернетики ім. В. М. Глушкова в рамках програми "Образний комп'ютер" розв'язується проблема мовної взаємодії. В Донецькому державному інституті штучного інтелекту ведуться роботи по мовному управлінню роботами.

Не дивлячись на велику кількість робіт, всі відомі розв'язання володіють недоліками, що обмежують їхнє практичне застосування. Складність задачі витікає з різноманіття і істотної нестабільності мовних висловів по відношенню до їхніх основних характеристик – частотного складу, часової та енергетичної структури. Забезпечення високої якості розпізнавання вимагає значних обчислювальних витрат. В даний час існує ряд апаратних розв'язань, направлених на забезпечення необхідної швидкодії систем розпізнавання мови. В основному, такі системи орієнтовані на зниження шумів, фільтрацію і прискорення базових обчислювальних операцій. На ринку представлені окремі сигнальні процесори і материнські плати з функціями мовного введення інформації. В той же час, існують обмеження при використовуванні апаратних засобів розпізнавання мови, пов'язані з впливом особливостей дикторів і зовнішнього середовища. Хоча більшість відомих систем дозволяє проводити деяке пристосовування до особливостей конкретного диктора, настройка проводиться або тільки на рівні лінгвістичної моделі, або дозволяє перепрограмувати фонетичний словник цілком, що спричиняє істотні додаткові витрати часу користувача.

Функціональні характеристики апаратних систем розпізнавання мови можуть бути істотно поліпшені за рахунок розширення можливостей настройки таких систем на різні умови вимовлення на

рівні акустичної обробки сигналу. Для цього необхідно вирішити задачу незалежної локалізації фонем в мовному сигналі без порівняння з іншими фонемами.

**Зв'язок роботи з науковими програмами, планами, темами.** Основні дослідження по темі дисертації проводилися на кафедрі прикладної математики і інформатики ДонНТУ в рамках виконання держ. тем Н-13-2000 "Алгоритмічне і програмне забезпечення високопродуктивних і інтелектуальних обчислювальних мереж" і Д-1-05 "Розвиток теорії мережевих інформаційних технологій розподіленого моделювання і синтезу цифрових систем на програмованих пристроях" (номер ДР 0105U002288), в яких брав безпосередню участь як виконавець.

**Мета і задачі дослідження.**

*Мета дослідження:* підвищення гнучкості настройки системи мовної взаємодії людини з ЕОМ на довільний фонетичний склад словника.

*Задачі дослідження:*

– Аналіз існуючих програмно-апаратних засобів автоматичного виділення фонем з мовного сигналу.
– Розробка методу роздільної локалізації фонем з урахуванням індивідуального фонетичного словника диктора.
– Розробка обчислювальних структур, що дозволяють оцінювати наявність фонем в реальному масштабі часу.
– Дослідження ефективності організації обчислювального процесу локалізації фонем при розпізнаванні мови.
– Розробка і аналіз VHDL-моделі обчислювального модуля оцінки наявності фонем.

*Об'єктом дослідження* є програмно-апаратні засоби розпізнавання мовного сигналу.

*Предметом дослідження* є алгоритми та обчислювальні структури засобів обробки мовного сигналу в системах автоматичного розпізнавання мови.

*Методи дослідження*. Для вирішення поставлених задач використані методи теорії розпізнавання мовних образів, теорії штучних нейронних мереж, теорії проектування цифрових обчислювальних засобів.

**Наукова новизна отриманих результатів.** В роботі отримані наступні нові наукові результати:

– Запропонований метод, що дозволяє виконувати незалежну локалізацію фонем в мовному сигналі. Отримано оцінки оптимальних параметрів методу.

– Запропонована структурна декомпозиція модуля оцінки фонем системи розпізнавання, що дає можливість гнучкої настройки на довільний фонетичний склад словника.
– Запропонована схема організації паралельної обробки мовного сигналу, що дозволяє вирішувати задачу розпізнавання мови в реальному масштабі часу.
– Отримана оцінка ефективності нейромережевих обчислювальних засобів при апроксимації фонем і розпізнаванні спектрально-часових образів слів. Показано, що використовування методу апроксимації фонем дозволяє знизити помилку розпізнавання на 30-40%.
– Вибраний раціональний спосіб організації структур обчислювального модуля локалізації фонем при апаратній реалізації на основі ПЛІС.

**Практичне значення отриманих результатів.** Застосування запропонованих методів, алгоритмів і обчислювальних структур дозволяє підвищити ефективність експлуатації систем мовної взаємодії людини з ЕОМ за рахунок підвищення гнучкості і зниження часу настройки модуля розпізнавання фонем на конкретного диктора. Розроблені нейромережеві структури і алгоритми використовуються на кафедрі "Прикладна математика і інформатика" Донецького національного технічного університету в лабораторному практикумі по курсу "Організація функціонування ЕОМ і систем" та ін.

**Особистий внесок здобувача** полягає в розробці методів, алгоритмів і структур обчислювальної системи розпізнавання мови. Всі основні результати дисертації отримані автором самостійно. В приведених публікаціях здобувачеві належать: в роботах [2,3,7] – розробка структур і програмних моделей, проведення експериментів; в роботах [1,5,6,9,11,12,13,14,15] – розробка методів, алгоритмів і апаратно-програмних засобів розпізнавання; в роботах [4,8,10,16] – аналіз засобів взаємодії людини з ЕОМ, формалізація моделей системи мовного управління.

**Апробація результатів дисертації.** Основні наукові результати, теоретичні і практичні розробки, доповідалися на наступних конференціях і семінарах:
– конференція "Нейрокомпьютеры и их применение", м. Москва, 1998, 2001, 2002 р;
– конференції по штучному інтелекту "КИИ-2000", м. Переславль-Залеський, і "КИИ-2002", м. Коломна;
– семінар "Інтелектуальний аналіз інформації", м. Київ, 2002, 2003 р;

- конференція "Интерактивные системы: проблемы человеко-компьютерного взаимодействия", м. Ульяновськ, 2001, 2003 р;
- семінар "Практика і перспективи розвитку інституційного партнерства", м. Таганрог, 2002 р;
- конференція "Научная сессия МИФИ – 2002". Секція "Интеллектуальные системы и технологии", м. Москва, 2002 р;
- конференція "Нейромережеві технології та їх застосування", м. Краматорськ, 2002, 2004 р.

**Публікації.** Результати дисертаційної роботи опубліковані в 7 статтях у виданнях, що входять до переліку ВАК України, і в 9 збірках доповідей наукових конференцій і семінарів.

**Структура дисертації.** Дисертація складається з вступу, чотирьох розділів, висновків, списку використаної літератури з 82 найменувань. Обсяг дисертації – 127 сторінок основного тексту, ілюстрованого 48 рисунками і 12 таблицями.

## ЗМІСТ РОБОТИ

У **вступі** обгрунтована актуальність теми дисертаційної роботи, сформульовані цілі і задачі дослідження, відзначена наукова новизна і практична значимість отриманих результатів, надані відомості про їхню апробацію і публікацію матеріалів дисертації.

В **першому** розділі – "Аналіз методів та засобів взаємодії людини з ЕОМ", проведений аналіз переваг та недоліків мовних засобів введення інформації і зроблено висновок про доцільність розробки систем, що ними управляють з голосу. Наведений огляд сучасних методів і систем розпізнавання мови показав, що головною задачею при організації мовних інтерфейсів є поліпшення якості розпізнавання і зниження залежності від особливостей дикторського вимовлення. Постановка задачі розпізнавання формулювалася таким чином. Хай є мовний вислів $w_i$:

$$w_i=(\{A_j(w_i)\}, s(w_i)), w_i \in W,$$

$\{A_j(w_i)\}$ – сукупність всіх можливих форм вимовлення вислову $w_i$ в їхній акустичній формі;

s(wi) – символьне представлення інформації, що міститься у вислові $w_i$,

$s(w_i) \in S$;

S – словник;

W – множина допустимих висловів: s(W) = S.
Потрібно знайти відображення R:

$$\forall\ w_i \in W,\ j \qquad R[A_j(w_i)] = s(w_i).$$

Сформульована задача є окремим випадком проблеми класифікації і розпізнавання образів. В даний час одним з найбільш ефективних засобів розпізнавання образів є штучні нейронні мережі. Переваги нейромереж – можливість апроксимації практично довільних розділяючих поверхонь при класифікації, пристосовування до вирішуваної задачі, структура, що дозволяє значною мірою розпаралелювати обчислення, а також однорідність обчислювальних елементів. Однак проблемою є необхідність одночасної перебудови всієї нейромережі при настройці системи на конкретного диктора. Це питання може бути вирішено за рахунок створення модульної нейронної мережі шляхом декомпозиції алгоритму розпізнавання фонем.

Аналіз апаратних засобів моделювання нейромережевих обчислень показав, що найбільш часто при цьому використовують цифрові сигнальні процесори, нейрочіпи і програмовані логічні інтегральні схеми (ПЛІС). Використовування останніх бачиться доцільним зважаючи на їхню універсальність і гнучкість.

В **другому** розділі – “Розробка методів і нейромережевих алгоритмів розпізнавання ізольованих слів за спектральним образом”, досліджені методи мовної взаємодії на основі розпізнавання ізольованих слів без аналізу їхньої внутрішньої структури. Була розглянута типова схема розпізнавання: вхідний сигнал – цифрова обробка – нейромережеве розпізнавання. В якості способу цифрової обробки був використаний метод лінійного згладжування, заснований на оцінці частот формантних викидів по максимумах спектральної енергії. Метод дозволяє одержувати дискретний спектрально-часовий образ (ДСЧО) вхідного слова у вигляді матриці, елементи якої бінарні і визначають наявність формантного викиду в даному частотному діапазоні у відповідний період часу (10 мс).

ДСЧО використовувалися в якості вхідних образів для нейромережевого розпізнавання на основі мереж типу багатошаровий персептрон. Для навчання і тестування нейромереж був сформований словник малого обсягу, кожне із слів якого було представлено декількома варіантами. Результати тестування виявили високу точність розпізнавання – 92% (5 слів). Проте збільшення словника

приводить до зростання обсягів обчислень при навчанні нейромережі і знижує якість розпізнавання до 86% (10 слів).

З метою скорочення розмірів вхідного образу в роботі був запропонований підхід, заснований на декомпозиції спектрально-часового образу і названий методом інтегральної оцінки приналежності компонент спектру словам. В основі підходу лежить схожість структур низькочастотних спектральних складових. Отже, кожна окрема складова несе в собі інформацію про слово цілком і може бути використана як окремий образ при розпізнаванні. Проте розпізнавання по єдиній компоненті спектру не дозволяє врахувати структуру спектру і приводить до втрат точності, тому в роботі проводилося розпізнавання за 5 низькочастотними складовими спектру, а результати узагальнювалися і служили підставою для остаточного висновку. Обчислювальна структура методу інтегральної оцінки належності компонент спектру словам відображена на рис. 1.

Рис. 1. Схема розпізнавання: ЕСі –експертная система, що робить висновок
про розпізнаване слово по і-й спектральній компоненті.

На схемі елементи першого рівня, що здійснюють безпосередньо розпізнавання відповідних компонент спектру, представлені нейромережевими експертними системами, у функції яких входить оцінка приналежності вхідного образу тому або іншому слову словника. В якості таких експертів були використані нейромережі типу багатошаровий персептрон. На другому рівні здійснюється узагальнення результатів і вибирається найбільш відповідне слово.

В другому розділі запропонована нейромережева реалізація другого рівня схеми розпізнавання, що дозволяє використати однотипні обчислювальні пристрої при апаратній реалізації алгоритму в цілому. Структура нейромережевих обчислень другого рівня представлена на рис. 2.

Рис. 2 Реалізація другого рівня розпізнавання в нейромережевому базисі:

$X_{ij}$ – j-й вихід і-й нейромережевої експертної системи першого рівня розпізнавання; f1, f2, f3, f4 – функції активації відповідних шарів нейромережі;

Y – номер розпізнаного слова в словнику;

S – сигнал про неможливість розпізнавання

Запропонований спосіб декомпозиції вхідного образу призводить до зниження часу навчання (при обсязі словника в 15 слів кількість епох навчання нейромережі зменшилася на два порядки). Крім того, виросла точність розпізнавання (при словнику з 10 слів отримано якість розпізнавання 89%).

Описані підходи до розпізнавання цілих слів за їхніми спектральними властивостями виявили наступні недоліки:

– неможливість врахувати часову структуру сигналу, що істотно позначається на якості розпізнавання;
– необхідність використовування однієї нейромережі для класифікації всіх образів, що призводить до значного зростання обчислювальних витрат при збільшенні словника.

До переваг запропонованих методів можна віднести достатньо високу якість розпізнавання, стійкість до зміни диктора і невеликі обчислювальні витрати на розпізнавання.

В **третьому** розділі – “Нейромережеві обчислювальні структури модуля розпізнавання фонем”, пропонується нова структура системи мовного введення команд, заснована на апроксимації фонем. На відміну від інших методів розпізнавання, заснованих на параметризації мовного сигналу і порівнянні отриманих параметрів з еталонними, запропонований спосіб вимагає порівняння вхідного мовного сигналу з моделями фонем наступним чином.

Будь-яке мовне слово однозначно визначається складом і порядком фонем, які воно містить, і може розглядатися як точка в N-мірному символьному просторі $P^N$, де N – максимальне можливе число фонем в слові. Слово довжини менше N може бути доповнене умовною "порожньою" фонемою $p_0$. Таким чином, словник можна розглядати як набір точок в просторі $P^N$, а задачу розпізнавання – як задачу пошуку найближчої до вимовленого слова точки.

Будемо вважати, що амплітудне представлення мовного слова функціонально залежить від його фонетичної $P=(p_1, p_2, , p_N)$ і часової структури $T=(\tau_0, \tau_1, \tau_N)$:

$$A(t)=F(P,T,t), \qquad (1)$$

де t – дискретний час, $t \in (\tau_0, \tau_N]$; $\tau_{i-1}$ і $\tau_i$ визначають межі фонеми $p_i$.

Відомо, що амплітудне представлення поточної фонеми $p_т$ залежить від попередньої фонеми $p_п$:

$$A^{пт}(t) = f(p_т, p_п, t). \quad (2)$$

Рахуючи загальне число фонем в мові обмеженим, можемо пронумерувати їх і замінити кожну фонему відповідним номером, що дозволить перейти від символьного простору, до числового:

$$f(p_i, p_j, t) = f_{ij}(t). \quad (3)$$

Фонетичну транскрипцію P мовного слова також можна замінити набором номерів фонем $K=(k_1, k_2, k_N)$. З (1-3) отримаємо:

$$A(t) = \sum_{i=1}^{N} s_i(t) \cdot f_{k_i k_{i-1}}(t - \tau_{i-1}), \quad (4)$$

де $k_0$ – номер "порожньої" фонеми $p_0$;

$$s_i(t) = \begin{cases} 1, t \in (\tau_{i-1}, \tau_i] \\ 0, \text{иначе} \end{cases}. \quad (5)$$

Знаючи $A(t)$ і амплітудні функції фонем $f_{ij}(t)$, можна знайти найкращу фонетичну $K_{опт}$ і часову $T_{опт}$ структуру слова з рішення задачі мінімізації функціонала близькості E в просторі $P^N$:

$$E(T,K) = \int_0^{\tau_N} [A(t) - \sum_{i=1}^{N} s_i \cdot f_{k_i k_{i-1}}(t - \tau_{i-1})]^2 dt = \sum_{i=1}^{N} [\int_{\tau_{i-1}}^{\tau_i} (A(t) - f_{k_i k_{i-1}}(t - \tau_{i-1}))^2 dt), \quad (6)$$

$$(T_{опт}, K_{опт}) = \arg\min(E(T,K)). \quad (7)$$

Залежність $f_{ij}(t)$ в явному вигляді на сьогоднішній день невідома. Ідея запропанованого методу полягає у тому, щоб замінити $f_{ij}(t)$ апроксимуючими функціями $G_{ij}$. У зв'язку з тим, що початок фонеми не може бути точно визначеним, побудувати залежність амплітуди безпосередньо від часу не представляється можливим. Тому розглядається залежність поточного значення $f_{ij}(t)$ від ряду попередніх значень:

$$f_{ij}(t) = G_{ij}(f_{ij}(t-1), f_{ij}(t-2),..., f_{ij}(t-m)), \quad (8)$$

де $G_{ij}$ – деяка апроксимуюча функція. Ураховуючи, що $A(t) = f_{k_i k_{i-1}}(t - \tau_{i-1})$ при $t \in (\tau_{i-1}, \tau_i]$, можемо записати

$$A(t) \approx \sum_{i=1}^{N} s_i(t) \cdot G_{k_i k_{i-1}}(A(t-1), A(t-2),..., A(t-m)), \quad (9)$$

$$E^*(T,K) = \sum_{i=1}^{N} \left( \int_{\tau_{i-1}}^{\tau_i} [A(t) - G_{k_i k_{i-1}}(A(t-1), A(t-2),..., A(t-m))]^2 dt \right), \quad (10)$$

$$(T^*_{опт}, K^*_{опт}) = \arg\min(E^*(T,K)). \quad (11)$$

Основними перевагами такого підходу є позиційна незалежність і нечутливість до змін часової структури сигналу, а також можливість розпізнавання на одній нефромережі єдиної фонетичної одиниці. Це дозволяє істотно знизити обчислювальні витрати при навчанні нейромереж і, отже, розширити можливі об'єми словника.

В роботі досліджена реалізація апроксимуючих функцій $G_{ij}$ за допомогою нейромереж типу багатошаровий персептрон, що обумовлено їхніми універсальними апроксимуючими властивостями і наявністю ефективного алгоритму навчання.

Експерименти з нейромережевого розпізнавання показали, що надійність локалізації фонем по мінімуму міри відмінності E недостатня, що вимагає залучення додаткових методів контекстної обробки. Це обумовило організацію розпізнавання відповідно до структурної схеми, показаної на рис. 3.

На схемі нейромережеві апроксиматори реалізують моделі відповідних фонем; інтегратори погрішності накопичують значення мір відмінності $Err_k(t)$ з метою компенсації впливу шумів; селектор призначений для вибору найбільш вірогідних фонем; граф ланцюгів фонем породжує всі слова, які можуть бути складені з вибраних фонем; семантико-синтаксичний аналізатор визначає ступінь можливості ланцюгів і вибирає серед них найкращу, яка і є результатом розпізнавання. Синтаксичний аналіз здійснювався методом динамічного програмування.

Рис. 3 Схема розпізнавання мови на основі нейромережевої апроксимації фонем

Експерименти, пов'язані з побудовою мір відмінності для ізольовано вимовлених фонем показали, що вокалізовані фонеми розпізнаються практично без помилок, тоді як надійність розпізнавання невокалізованих дуже низька, що зумовлено значною часткою шуму, який в них міститься. Отже, для організації процесу розпізнавання необхідний спосіб розділення сигналу на вокалізовані і невокалізовані участки. В третьому розділі запропонований алгоритм рішення задачі, заснований на аналізі енергії сигналу і включаючий наступну послідовність кроків.

Крок 1. Обчислюється енергія сигналу S(t) як набір сум абсолютних значень амплітуди сигналу в рамках ковзаючого вікна заданого розміру;

Крок 2. Визначається множина точок $B=\{b_1, b_2, \quad b_n\}$ и $K=\{k_1, k_2, \quad k_n\}$:

якщо $S(t)>P$ і $S(t_1)<P$, то $b_i = t$;

якщо $S(t)<P$ і $S(t_1)>P$, то $k_i = t_1$,

$i=i+1$

для всіх значень t, де P=const – деякий наперед визначений поріг;

Крок 3. Серед точок з B і K в якості меж вокалізованих інтервалів вибираються тільки ті точки $b_i$ і $e_i$, для яких справедливі умови:

- $e_i - b_i > L$, де L – задана величина, що визначає мінімальну можливу тривалість ділянки;
- $e_j - b_j < L$ для всіх $j<i$ або $b_i - e_{i-1} > L$.

Якщо $e_i - b_i > L$, $e_i\text{-}1 - b_i\text{-}1 < L$ і $e_{i-2} - b_{i-2} > L$, то в якості меж даної ділянки вибираються $b_{i-2}$ і $e_i$.

Надійність роботи алгоритму визначається наступними параметрами:

- $T_{min}$ – мінімальна довжина інформативної/неінформативної ділянки. Це значення може бути отримано виходячи з мінімальної довжени вимовлення однієї фонеми. Досвід ручної сегментації слів за іхнім графічним представленням показує, що раціональним значенням є $T_{min} = 350$ мс.
- $T_{win}$ – розмір ковзаючого вікна, в рамках якого обчислюється енергія сигналу;
- E – порогове значення енергії активного фрагмента.

Експериментальним шляхом були виявлені найбільш раціональні значення параметрів: $T_{min} = 350$ мс, $T_{win} = 30$ мс, E = 3% від максимально можливої енергії сигналу в даному вікні. Якість фрагментації при цих значеннях склала 94%. Проте цей результат може бути поліпшений за рахунок введення подвійної сегментації деяких слів. З 17 помилок сегментації (5,7%), отриманих в найбільш вдалому експерименті, неправильно сегментовані 12 слів, з них в 7 помилка зустрічається тільки на одній реалізації, а в 5 – на двох. Слово з 3 варіантами сегментації було єдиним. Таким чином, додавши в словник по додатковому варіанту сегментації 4 слів, можна збільшити точність алгоритму до 97%.

Висока точність фрагментації дозволяє проводити на її основі сегментацію словника по кількості активних ділянок в словах. Це знизить кількість альтернативних ланцюгів при синтаксичному аналізі, і, отже, підвищить точність розпізнавання. Загальна схема розпізнавання з урахуванням використовування алгоритму фрагментації показана на рис. 4.

Рис. 4 Схема включення модуля порогової обробки в систему розпізнавання мовних команд

Запропонований метод розпізнавання дозволив виконати структурну декомпозицію блоку розпізнавання фонем і організувати його відповідно до схеми, що наведена на рис. 5.

Рис. 5. Схема блоку розпізнавання фонем на основі нейромережевої апроксимації сигналу

На рис. 5 $A(t_i,t_{i+1})$ – амплітудно-часова форма сигналу; $p_i(f_k)$ – оцінка схожості і-го фрагмента сигналу з k-ю фонемою. Така організація обчислень дозволяє підвищити гнучкість настройки системи мовної взаємодії.

Реалізація запропонованої схеми вимагає визначення значень параметрів окремих блоків, при яких якість розпізнавання буде найбільш високою. Для цього були проведені наступні серії експериментів.

Обрано спосіб нормалізації сигналу. Необхідність нормалізації обумовлена нестабільністю енергетичної структури сигналу, а також обмеженістю функції активації нейромереж діапазоном [0; 1]. Розглядалися наступні варіанти:

- ортогональне проектування навчальних пар на відрізок [0;1];
- ортогональне проектування всього сигналу на відрізок [0;1];
- перетворення всіх значень вхідного сигналу за формулою

$$A^{/}_{w}(t) = \frac{1}{1 + e^{-A_w(t)}},$$

де $A_w(t)$ – поточне значення амплітуди сигналу; $A^{/}_{w}(t)$ – нормоване значення;

- нормувався всього сигналу за формулою

$$A^{/}_{w}(t) = \frac{\sin(A_w(t)) + 1}{2};$$

- перетворення всього сигналу за формулою

$$A^{/}_{w}(t) = \frac{arctg(A_w(t+1) - A_w(t)) + \pi}{2\pi}.$$

Найбільш раціональним виявився перший із запропонованих способів. Не звертаючи на його великі, порівняно з іншими, обчислювальні витрати, він дає результат, майже в 3 рази перевершуючий кращий з інших способів.

Розглянуто два варіанти способу формування повчальної множини для нейромережевих апроксиматорів – навчання по ізольованих вимовах фонем і по фонемах, що їх виділено із слів словника суб'єктивно при аналізі графіків. При другому способі отримано якість розпізнавання на 10% вище, ніж при першому.

Визначені параметри нейромережевих апроксиматорів. Розглядалися наступні параметри: кількість шарів, розподіл нейронів по шарах, розмір вхідного образу. Найбільш успішною з погляду якості розпізнавання показала себе тришарова мережа на 60 входів з 20, 10 і 1 нейронами в шарах.

Визначено такт інтеграторів погрішності. Найбільш ефективне значення такту склало 30 мс.

Досліджено доцільність згладжування сигналу. Розглянута можливість поліпшення якості апроксимації за рахунок зниження енергії високочастотної складової сигналу шляхом його послідовного усереднювання. Експерименти показали, що таке зниження є недоцільним, оскільки призводить до втрати якості.

Проведено порівняння апроксимації сигналу нейромережами і методом групового врахування аргументів (МГУА). Було розглянуто

два варіанти МГУА – апроксимація за допомогою багаточлена і деревовидні нейронні мережі, що базуються на формуванні моделі у вигляді суперпозиції елементарних функцій одного аргументу. Окремо проводилася апроксимація вокалізованих і невокалізованих фонем. Всі фонеми вимовлялися ізольовано. Результати показали, що якість розпізнавання у всіх випадках значно поступається якості, яка отримана за допомогою нейромереж.

Проведено порівняння результатів розпізнавання за спектром і за параметричним описом, отриманим методом нейромережевої апроксимації фонем. Перший результат склав 63%, другий – 79%.

На основі отриманих в результаті експериментів параметрів був створений інтерпретатор мовних команд управління інформаційною системою нейромережевого аналізу даних.

Набір мовних команд управління системою включає 60 слів і словосполучень, що покривають основні її функції, за винятком введення даних. Точність розпізнавання, отримана на цьому наборі, склала близько 90%.

**Четвертий** розділ – “Апаратна реалізація нейромережевих обчислень при розпізнаванні мови”, присвячений створенню засобів апаратного прискорення роботи методу. Програмна реалізація схеми розпізнавання на основі нейромережевої апроксимації фонем виявила необхідність великого об’єму обчислень при розпізнаванні однієї команди. Зокрема, розпізнавання команд із словника блоку мовного управління інформаційною системою нейромережевого прогнозування (60 слів і словосполучень), в середньому, вимагає 5.718 сек. на комп'ютері з процесором AMD-K6-3DNow! (тактова частота 500 МГц), що неприйнятне для організації діалогу в реальному часі. Структурна схема мовного каналу показана на рис. 6, а час виконання окремих етапів методу приведений в табл. 1.

Рис. 6 Структурна схема мовного каналу введення команд

Таблиця 1. Розподіл часових витрат при розпізнаванні на послідовній ЕОМ з процесором AMD K6-2

| Етап розпізнавання | Об'єм обчислень, |
|---|---|
| Порогова | 0,17 |
| Обчислення мір | 99,74 |
| Лінгвістична обробка | 0,05 |

| | |
|---|---|
| Інші обчислення | 0,04 |

Дані таблиці показують, що найбільш доцільним є прискорення роботи нейромережевих апроксиматорів, що здійснюють обчислення мір відмінності. В якості апаратної бази в роботі використані типові (24576 КЛБ) програмовані логічні інтегральні схеми (ПЛІС).

Аналіз обчислювальної ієрархії в нейромережевих обчисленнях дозволив виділити наступні можливі шляхи рішення задачі:

1) реалізація алгоритму в цілому у вигляді комбінаційної схеми;

2) реалізація мереж у вигляді автоматів з комбінаційною частиною, що реалізує один шар (нейромережі працюють паралельно);

3) реалізація нейромереж комбінаційно (нейромережі працюють послідовно);

4) реалізація нейромереж у вигляді автоматів з комбінаційною частиною, що реалізує один шар (нейромережі працюють послідовно);

5) реалізація нейрона у вигляді комбінаційної схеми (нейрони в шарі працюють послідовно);

6) реалізація нейрона у вигляді автомата (нейрони в шарі працюють паралельно);

7) реалізація нейрона у вигляді автомата (нейрони в шарі працюють послідовно).

Найбільш швидким є перший спосіб, але він вимагає найбільших апаратних витрат. Сьомий спосіб відповідає реалізації на послідовній ЕОМ і є найбільш повільним.

Принциповим питанням при апаратній реалізації нейромереж є вибір способу зберігання вагових коефіцієнтів. Канал вводу-виводу у ПЛІС Virtex має місткість від 180 до 512 біт. Зберігання вагових коефіцієнтів в зовнішній пам'яті вимагає 300-800 операцій обміну з ПЛІС між тактами надходження нового значення мовного сигналу (125 мкс). Більш вдалим є підхід, при якому вагові коефіцієнти будуть розміщені безпосередньо на ПЛІС. В цьому випадку можливі два варіанти:

- значення терезів будуть побічно зберігатися в блоках множення числа на константу;
- значення терезів будуть зберігатися безпосередньо на ПЛІС.

Вибір раціонального способу залежить від способу представлення вагових коефіцієнтів. В програмній моделі на універсальній ЕОМ вагові коефіцієнти представлялися 32-бітовими числами з плаваючою комою, проте в ході експериментів було встановлено, що для їхньої цілої частини достатньо 8 біт (один біт

знаковий), а дробова також може бути уявлена у вигляді 8-бітових чисел з фіксованою комою. Таке представлення призводить до зниження точності розпізнавання на 0,33%. Зниження розрядності ще на 1 біт погіршує якість на 4%, що було визнано неприйнятним.

Обидва способи зберігання вагових коефіцієнтів було промодельовано на VHDL. Моделювання нейромереж з помножувачами на константу показало, що на ПЛІС Virtex XCV1000-FG680-4 можна розмістити 81 нейрон з 20 входами (без функцій активації), тоді як в розробленому алгоритмі потрібне обчислення вихідних сигналів 420 нейронів з 20 входами і 14 нейронів з 10 входами. Отже, апаратна реалізації в такому варіанті неможлива.

Варіант безпосереднього зберігання вагових коефіцієнтів був реалізований у вигляді ПЗП. В результаті моделювання отримана оцінка апаратних витрат: 13270 конфігуруємих логічних блоків (КЛБ), що складає 54% поверхні використаних ПЛІС. Отже, схема із зберіганням вагових коефіцієнтів на ПЛІС у вигляді ПЗП може бути реалізована. Далі розглядалася можливість реалізації нейромережевих обчислень на залишених вільними 46% ПЛІС.

Наступним питанням, розглянутим в роботі, стала реалізація функції активації нейронів $f(g)=\frac{1}{1+e^{-g}}$ . Для відтворення цієї функції розглянуто алгоритми на основі лінійного згладжування, побітового представлення ступеня експоненти і табличне визначення функції. Моделювання показало, що при лінійній апроксимації погрішність обчислень дуже велика, а при двох інших способах досить незначно відображається на якості розпізнавання.

Розглянуто два варіанти реалізації функції активації другим способом: у вигляді автомата або комбінаційної схеми. В результаті моделювання встановлено, що комбінаційний варіант дозволить обчислити функцію активації за 248 нс і вимагає 1698 КЛБ, а автоматний – 861 нс і 294 КЛБ. При порівнянні результатів необхідно врахувати особливості нейроалгоритму: на ПЛІС повинно бути розміщено 20 функцій активації, а загальний час обчислень не повинен перевищувати 125 мкс. З урахуванням розміщення вагових коефіцієнтів на ПЛІС, вільна місткість складає 12490 КЛБ, що дозволяє реалізувати усього 8 комбінаційних функцій активації, майже не залишивши місця для реалізації інших операцій нейрона. Автоматний спосіб вимагає для 20 функцій 5846 КЛБ, що складає близько 23% від ресурсів ПЛІС. Часові витрати автоматного способу визначаються необхідністю виконання 42 нелінійних перетворень, що потребує 36,162 мкс. Це не протиречить вимогам реального часу розпізнавання.

При дослідженні табличного способу розглядалася найбільш ефективна розрядність аргументу і значення функції. Дослідження показали, що 8-бітне представлення обох параметрів дозволяє проводити обчислення без втрат якості розпізнавання. В табл. 2 показані апаратні витрати і затримки при обчисленнях трьома способами. Оптимальним є табличний спосіб, на якому базуються подальші структури і алгоритми.

Таблиця 2. Результати моделювання пристрою обчислення сігмоїдальної функції різними способами

| **Спосіб** | **Час** | **Апаратні витрати** |
|---|---|---|
| Комбінаційний | 248 нс | 1698 КЛБ |
| Автоматний | 861 нс | 294 КЛБ |
| Табличний | 28 нс | 76 КЛБ |

Аналогічні питання розв'язувалися при виборі способу апаратної реалізації нейрона в цілому. Вільна частина ресурсів ПЛІС, що залишилася не зайнятою під вагові коефіцієнти і функції активації, дозволяє реалізувати або два комбінаційні нейрони, або для 20 автоматних. Максимальне розпаралелювання обчислень в шарі вимагає використовування другого варіанту.

Результати проведених досліджень дозволили побудувати схему нейрона, що задовольняє вимогам реалізації нейроалгоритму на ПЛІС (рис. 7).

Рис. 7 Схема апаратної реалізації нейрона

На схемі X – вектор входів нейрона, що складається з 20 компонент, кожна з яких представлена 9-розрядним числом з фіксованою комою в діапазоні [0; 1]; CLK – синхросигнал керуючого автомата; Reset – сигнал встановлення у початкове положення; блок "Вагові коефіцієнти" організовано у вигляді ПЗП з адресним входом Addr, що визначає номер вагового коефіцієнта (від 0 до 19); блок "Мультиплексор" пропускає компоненту вхідного вектора X з номером Addr; блок "Обчислювач" проводить накопичення зваженої суми входів і її нелінійне перетворення; Y – сигнал нейрона виходу.

Схема блоку "Обчислювач", показана на рис. 8, включає основні операційні елементи штучного нейрона. Входи і вихід схеми описані вище. Блок "Помножувач" обчислює добуток поточної вхідної

компоненти вектора X та відповідного йому вагового коефіцієнта W з блоку "Вагові коефіцієнти". Блоки "Суматор" і "Регістр" призначені для накопичення елементів скалярного добутку $X \cdot W$. Зважена сума, отримана за 20 тактів роботи пристрою, подається на блок "Функція активації", який формує вихідний сигнал нейрона.

Рис. 8 Функціональна схема операційного автомата штучного нейрона

Розроблена схема нейрона була використана при апаратній реалізації всього нейроалгоритму апроксимації фонем. Операційна частина отриманого пристрою показана на рис. 9. Модуль "Набір нейронів" складається з 20 схем, що реалізують штучний нейрон (рис. 7).

Робота керуючого автомата визначається наступним алгоритмом:

1. Встановити у початковий стан значення зважених сум в нейронах модуля "Набір нейронів" (сигнал Reset). Встановити в 0 значення адреси вагових коефіцієнтів Addr. Встановити в 1 сигнал First, внаслідок чого "Мультиплексор шарів" на цьому кроці пропустить вхідний вектор X;
2. Записати вихід блоку "Мультиплексор шарів" в "Набір регістрів" (сигнал RegSetClock). Встановити в 0 сигнал Code;
3. Циклічно (протягом 20 тактів) накопичувати зважені суми в нейронах шара, інкрементуючи сигнали Addr і Code;
4. Обчислити значення функцій активації в блоці "Набір функцій активації" (сигналу ResetPorog);
5. Якщо сигнал Addr досяг значення 60, тобто були задіяні всі вагові коефіцієнти даної мережі, то завершити роботу пристрою, встановити в 0 сигнал First і перейти до кроку 2 для обчислення наступного шара.

Рис. 9 Операційна частина пристрою нейромережевого розпізнавання фонем

Тактування роботи пристрою здійснюється сигналом CLK. Сигнали X (вхідний вектор), NetNum (номер нейромережевого апроксиматора) і Reset (встановлення у початковий стан) є вхідними для пристрою.

Моделювання алгоритму за допомогою САПР Xilinx ISE 4 показало, що задача апаратного прискорення алгоритму розпізнавання може бути успішно вирішена з використовуванням ПЛІС Xilinx Virtex XCV1000-FG680. При реалізації пристрою було задіяно 24094 КЛБ ПЛІС, що склало близько 98% від їхнього загального числа. Час роботи однієї нейромережі дорівнюється 4,816 мкс, і, отже, виходи 14 нейромереж можуть бути обчислені за допомогою розробленого пристрою приблизно за 70 мкс. Для розпізнавання в реальному масштабі часу достатньо одержувати результат за 125 мкс. Таким чином, проведені дослідження показують можливість організації мовного введення інформації на основі методу нейромережевої апроксимації фонем з використовуванням ПЛІС для прискорення нейромережевих обчислень.

## ВИСНОВКИ

В дисертації приведено нове рішення наукової задачі структурної декомпозиції засобів локалізації фонем в системах мовної взаємодії людини з ЕОМ. Головні наукові і практичні результати роботи полягають в наступному:

1. Запропонований метод інтегральної оцінки приналежності спектральних складових словам, заснований на декомпозиції спектрального образу і порівнянні результатів розпізнавання окремих спектральних складових. Метод дозволяє знизити помилку розпізнавання порівняно з обробкою неподільного образу на 30-40%.
2. Запропонований метод розпізнавання мови, заснований на нейромережевій апроксимації фонем. Метод є позиційно-незалежним і дозволяє організувати незалежне розпізнавання фонетичних одиниць.
3. Досліджена залежність якості розпізнавання фонем методом нейромережевої апроксимації сигналу від параметрів оцифровки сигналу, розмірностей нейромереж, способів нормалізації і попередньої обробки сигналу. Отримані субоптимальні значення параметрів. Проведено порівняння нейромережевої апроксимації з апроксимацією на основі МГУА. Показано, що якість розпізнавання окремих фонем при використовуванні нейромереж вища на 40-50%.
4. Розроблено обчислювальну структуру модуля розпізнавання фонем і схему його включення в систему розпізнавання мови. Запропонована структура має високу ступінь модульності і

дозволяє проводити гнучку настройку системи на довільний фонетичний склад словника.

5. Розроблено алгоритм розподілу слова на інформативні і неінформативні ділянки за енергією сигналу, а також запропонована сегментація словника за кількістю інформативних ділянок в словах. Сегментація дозволяє знизити об'єм обчислень при розпізнаванні в 1,5-2 рази.
6. Запропонована паралельна апаратна реалізація нейромережевих апроксиматорів фонем на ПЛІС, що дозволяє організувати розпізнавання в реальному масштабі часу.
7. На основі методу нейромережевої апроксимації фонем реалізована програмна система аналізу даних з елементами мовного управління. Набір мовних команд управління системою складає 60 словарних одиниць, а точність розпізнавання – близько 90%.

## СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ


1. О.И. Федяев, С.А. Гладунов. Распознавание речевых слов с помощью искусственных нейросетей // Науч. тр. Донецкого гос. тех. университета. Серия: Информатика, кибернетика и вычислительная техника, вып. 6, 1999 – С. 145-150.
2. Федяев О.И., Гладунов С.А., Прокофьев А.В. Прогнозирование временных рядов на основе нейросетевых и нечетких моделей // Науч. тр. Донецкого гос. тех. университета. Серия: Проблемы моделирования и автоматизации проектирования динамических систем, вып. 10, 1999. – С. 38-43.
3. Федяев О.И., Гладунов С.А. Исследование эффективности нечеткого группового метода обработки данных в задачах прогнозирования // Науч. тр. Донецкого гос. тех. университета. Серия: Информатика, кибернетика и вычислительная техника, вып. 15, 2000. – С. 187-191.
4. Федяев О.И., Гладунов С.А. Речевая компонента в интерфейсах информационных систем // Науч. тр. Донецкого гос. тех. университета. Серия: Проблемы моделирования и автоматизации проектирования динамических систем, вып. 29, 2001. – С. 100-105.
5. Федяев О.И., Гладунов С.А. Многоуровневая нейросетевая структура распознавания речевых слов по низкочастотным гармоникам. – Науч. тр. Донецкого гос. тех. университета. Серия: Информатика, кибернетика и вычислительная техника, вып. 39, 2002. – С. 30-35.

6. Гладунов С.А., Федяев О.И. Нейросетевой метод фонетической сегментации речевого сигнала. – Науч. тр. Донецкого гос. тех. университета. Серия: Проблемы моделирования и автоматизации проектирования динамических систем, вып. 52, 2002. – С. 125-130.
7. Федяев О.И., Гладунов С.А. Оценка параметров метода нейросетевой аппроксимации фонем. – Науч. тр. Донецкого гос. тех. университета. Серия: Информатика, кибернетика и вычислительная техника, вып. 70, 2003. – С. 220-227.
8. С. А. Гладунов, О.И. Федяев. Речевое управление программными системами с помощью нейросетей // КИИ-2000. Труды конференции. – М.: Издательство физико-математической литературы, 2000. – Том 2, С. 464-471.
9. Федяев О.И., Гладунов С.А. Нейросетевой интерпретатор речевых команд для управления программными системами // Труды 7-й всероссийской конференции “Нейрокомпьютеры и их применение”. – М.: ИПУ РАН, 2001. – С. 283-288.
10. Fedyaev O.I., Gladunov S.A. Usage of a vocal component in interfaces of programmed systems // Interactive Systems: The Problems of Human-Computer Interaction. Proceedings of the International Conference. – Ulyanovsk: UlSTU, 2001. – P. 26-28.
11. Федяев О.И., Гладунов С.А. Иерархическая нейросетевая структура распознавания слов на основе низкочастотных гармоник // Сборник научных трудов “Научная сессия МИФИ – 2002”. В 14 томах.– М.: МИФИ, 2002. – Т.3. Интеллектуальные системы и технологии. С. 115-116.
12. Федяев О.И., Гладунов С.А. Распознавание речевых слов по низкочастотным гармоникам с помощью нейросетей // Труды 8-й всероссийской конференции “Нейрокомпьютеры и их применение”. – М.: Век книги, 2002. – С. 156-161.
13. Федяев О.И., Гладунов С.А. Фонетический анализ речи на основе нейросетевой аппроксимации сигнала // Труды 8-й всероссийской конференции “Нейрокомпьютеры и их применение”. – М.: Век книги, 2002. – С. 150-155.
14. Федяев О.И., Гладунов С.А. Распознавание слитной речи методом нейросетевой аппроксимации сигнала // Известия ТРТУ-ДонНТУ. Материалы 3-го Международного научно-практического семинара “Практика и перспективы развития институционного партнерства”. В 2-х кн. – Таганрог: Издательство ТРТУ, 2002. – Кн. 1. С. 140-145.
15. Федяев О.И., Гладунов С.А. Распознавание речи на основе нейросетевой аппроксимации фонем // КИИ-2002. Труды

конференции. В 2 томах. – Коломна: Коломенская типография, 2002. – Т.2, С. 187-192.
16. Oleg I. Fedyaev., Sergey A. Gladunov. Organizing a speech input of information based on neural phonemes approximation // Interactive Systems: Problems of Human-Computer Interaction. Proceedings of the International Conference, 23-27 September 2003 – Ulyanovsk: UlSTU, 2003. – P. 198-203.

## АНОТАЦІЯ

**Гладунов Сергій Анатольевіч.** Апаратно-програмні засоби роздільної локалізації фонем в системах мовної взаємодії людини з ЕОМ.

Дисертація на здобуття вченого ступеня кандидата технічних наук за спеціальністю 05.13.13 – обчислювальні машини, системи та мережі. – Донецький національний технічний університет. Донецьк, 2005.

Розглянуто питання, пов’язані із підвищенням ефективності засобів мовного введення інформації в ЕОМ за рахунок структурної декомпозиції модуля розпізнавання фонем. Розроблено метод інтегральної оцінки приналежності спектральних складових словам, заснований на декомпозиції спектрального образу, що дозволив скоротити час навчання і знизити помилку розпізнавання на 30-40%. З метою підвищення гнучкості настройки системи розпізнавання мовних команд запропоновано метод фонетичного аналізу мовного сигналу, заснований на апроксимації фонем. Розглянуто нейромережеву реалізацію апроксиматорів фонем і запропоновано алгоритм розподілу мовного слова на інформативні і неінформативні ділянки.

З метою підвищення швидкості розпізнавання запропонований алгоритм апаратного прискорення нейромережевих обчислень. Показано, що використовування типових ПЛІС дозволяє організувати розпізнавання в реальному масштабі часу. На основі запропонованого методу апроксимації фонем розроблений мовний інтерпретатор команд управління програмною системою нейромережевого аналізу даних зі словником у 60 команд і точністю розпізнавання приблизно 90%.

Ключові слова: структурна декомпозиція, розпізнавання мови, модуль розпізнавання фонем, штучні нейронні мережі, спектральний образ, ПЛІС, паралельна реалізація нейроалгоритму, мовний інтерпретатор команд.

# АННОТАЦИЯ

**Гладунов Сергей Анатольевич.** Аппаратно-программные средства раздельной локализации фонем в системах речевого взаимодействия человека с ЭВМ.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.13 – вычислительные машины, системы и сети. – Донецкий национальный технический университет. Донецк, 2005.

Диссертация посвящена повышению эффективности средств речевого ввода информации в ЭВМ за счет структурной декомпозиции модуля распознавания фонем.

Сделан обзор аппаратно-программных средств автоматического распознавания речи, а также выполнен анализ нейросетевых структур и алгоритмов решения задачи. Показано, что эффективным средством фонетического анализа речевого сигнала являются нейронные сети. Рассмотрены аппаратные средства поддержки нейросетевых вычислений. Исследован типовой алгоритм распознавания с помощью нейросетей, основанный на нейросетевом распознавании двоичных спектрально-временных образов слов. Алгоритм показал недостатки такого подхода: небольшой словарь и значительное время на обучение нейронных сетей.

Разработан метод интегральной оценки принадлежности спектральных составляющих словам, основанный на декомпозиции спектрального образа. Метод позволил сократить время обучения и снизить ошибку распознавания на 30-40% в сравнении с типовой схемой. Предложена схема реализации метода в нейросетевом вычислительном базисе.

С целью повышения гибкости настройки системы распознавания речевых команд предложен метод фонетического анализа речевого сигнала, основанный на аппроксимации фонем. Идея метода состоит в оценке сходства фрагмента сигнала с фонемами по ошибке прогноза развития сигнала, полученного на моделях каждой из фонем словаря. Рассмотрены модели фонем, построенные с помощью нейросетей типа многослойный персептрон, настроенных путем минимизации квадратичного отклонения выхода сетей от эталонной реализации речевого сигнала для каждой из фонетических единиц. Эксперименты показали, что эффективная оценка может быть получена не для всех фонем. Предлагается алгоритм разбиения речевого слова на информативные и неинформативные наборы фонем по энергии сигнала. На основе такой сегментации выполнено разделение словаря

на группы слов по количеству активных участков, что позволяет существенно снизить объем вычислений при лингвистической обработке.

Предложена программная реализация алгоритма распознавания речи на основе нейросетевой аппроксимации фонем в системе нейросетевого анализа данных с элементами речевого управления. Словарь системы составил 60 слов, точность распознавания которых приблизительно равняется 90%.

Эксперименты показали, что время распознавания одной команды слишком велико даже при монопольном использовании процессора. Для сокращения времени распознавания и снижения нагрузки на центральный процессор предложены средства аппаратного ускорения процесса распознавания фонем. Поскольку общее время распознавания команды, складывающееся из этапов акустической и лингвистической обработки, на 99% состоит из времени вычисления выходов нейросетей, рассматривались варианты аппаратного ускорения работы нейросетевых аппроксиматоров фонем. В качестве аппаратной базы выбраны типовые ПЛИС, являющиеся мощным и универсальным средством моделирования цифровых устройств.

Рассмотрены различные варианты аппаратной реализации нейросетевых аппроксиматоров фонем на различных уровнях нейросетевых вычислений: на уровне базовых операций, нейронов, слоев и модулей. Выбран вариант, удовлетворяющий временным ограничениям, связанным с распознаванием в реальном масштабе времени, и реализуемый на рассмотренной модели ПЛИС.

Исследован вопрос о цифровой реализации функции активации нейронов. Показано, что наиболее эффективным является табличное представление функции с разрядностью аргумента и значения 8 бит. Рассмотрены варианты представления весовых коэффициентов. Выбран вариант хранения весов в ПЗУ непосредственно на ПЛИС. Общая схема реализации нейросетевых вычислений позволила получать выходы всех сетей за время между моментами поступления значений речевого сигнала. Таким образом, распознавание речи на основе нейросетевой аппроксимации фонем может быть организовано в реальном масштабе времени.

Ключевые слова: структурная декомпозиция, распознавание речи, модуль распознавания фонем, искусственные нейронные сети, спектральный образ, ПЛИС, параллельная реализация нейроалгоритма, речевой интерпретатор команд.

SUMMARY

**Gladunov Sergiy Anatoliyovich.** Hardware and software tools of separated phoneme localization in speech communication systems of human and computer.

Dissertation on taking the candidate degree of technical science in specialty 05.13.13 – computing machines, systems and networks. – Donetsk national technical university. Donetsk, 2005.

There are questions discussed, connected with speech input of information into computer efficiency inducing by mean of structural decomposition of phoneme recognition module. A new method of integral estimation of spectral components to words correspondence is designed. The method is based on spectral pattern decomposition, which allows to reduce training time and decrease a recognition error by 30-40%**.** To increase a flexibility of speech command recognition system adaptation it was proposed a phonetic analysis of speech signal method, which is based on phoneme approximation. There was researched a neural based realization of phoneme approximators and proposed a speech signal division into informative and non-informative part.

To increase a recognition speed there was proposed an algorithm of neural computations hardware acceleration. There's shown that using standard FPGA a real time recognition may be achieved. Based on the neural phoneme approximation method it was designed a speech command interpretation unit to conduct a programmed neural data analysis system with 60 commands vocabulary. A recognition accuracy is about 90%.

Keywords: structural decomposition, speech recognition, phoneme recognition unit, artificial neural networks, spectral pattern, FPGA, parallel neural algorithm realization, speech command interpretation system.