

Вінницький національний технічний університет

Зірнєєва Галина Валеріївна

УДК 534.4:681.3.01

Інформаційна технологія обробки та аналізу
характеристик мовленнєвої інформації

05.13.06 – Інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Вінниця – 2008

Дисертацією є рукопис.

Роботу виконано у Дніпропетровському національному університеті Міністерства освіти і науки України.

Науковий керівник: доктор технічних наук, професор

Карпов Олег Миколайович,

Дніпропетровський національний університет,

професор кафедри математичного забезпечення ЕОМ

Офіційні опоненти: доктор технічних наук, професор

Дубовий Володимир Михайлович,

Вінницький національний технічний університет,

завідувач кафедри „Автоматизованих систем

управління”

кандидат технічних наук

Сажок Микола Миколайович,

МННЦ Інформаційних технологій та систем,

старший науковий співробітник відділу

„Розпізнавання та синтез звукових образів”

Захист відбудеться “03” жовтня 2008 р. о 12³⁰ годині на засіданні спеціалізованої вченої ради Д 05.052.01 у Вінницькому національному технічному університеті за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГУК, 210.

З дисертацією можна ознайомитись у бібліотеці Вінницького національного технічного університету за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95.

Автореферат розісланий “02” вересня 2008 р.

Вчений секретар

спеціалізованої вченої ради

С.М. Захарченко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Одним з основних напрямків сучасних розробок та досліджень в галузі інформаційних технологій є аналіз, обробка та розпізнавання мовленнєвих сигналів. У наш час багато ресурсів спрямовується на створення систем, які б дозволили організувати взаємодію між людиною та комп'ютером у формі мовленнєвого діалогу, а завдяки технічному прогресові системи розпізнавання мовлення знаходять все нові й нові сфери застосування.

Актуальність теми підтверджує також велика кількість наукових досліджень та винаходів у галузі обробки мовленнєвих сигналів. Наукові дослідження з обробки, аналізу та розпізнавання мовленнєвих сигналів проводяться науковцями Інституту кібернетики НАН України, Вінницького національного технічного університету, МННЦ Інформаційних технологій та систем, Харківського національного університету радіоелектроніки, Національного технічного університету України «Київський політехнічний інститут», Дніпропетровського національного університету, Національного університету «Львівська політехніка» та Донецького інституту проблем штучного інтелекту. Вагомі результати в цій галузі здобуті відомими фахівцями, такими як Т.К. Вінцюк, М.Г. Загоруйко, В.Г. Величко, М.Ф. Бондаренко та ін.

Незважаючи на це, багато питань залишаються невирішеними. Для розпізнавання мовлення дуже важлива адекватна модель мовленнєвого сигналу. Існуючі на сьогодні методи розпізнавання мовлення недостатньою мірою забезпечують швидке та надійне розпізнавання мовленнєвих сигналів.

Проведений огляд технологій розпізнавання, а також обробки та аналізу мовленнєвих сигналів показує, що цей напрямок розвинений недостатньо, не відповідає потребам користувачів.

Виникає завдання створення нової інформаційної технології обробки та аналізу характеристик мовленнєвої інформації, яка б дозволяла проводити більш швидкісне та ефективне розпізнавання мовленнєвих сигналів (мовленнєвих команд). В основу інформаційної технології пропонується покласти модель мовленнєвого сигналу зі структурними змінами, а також методи, які дозволять проводити більш ефективну класифікацію мовленнєвих сигналів.

Зв'язок роботи з науковими програмами, планами, темами. Основний зміст роботи становлять результати наукових розробок та експериментальних досліджень, які проводились протягом 2003–2007 рр. на кафедрі математичного забезпечення ЕОМ Дніпропетровського національного університету. Протягом 2007 р. наукова робота проводилась відповідно до науково-дослідної держбюджетної теми Д/б №7-094-05 «Радіоспектроскопічні дослідження процесів рекомбінації в широкозонних напівпровідниках» (№ НДР держ. реєстрації 0105U000376).

Мета і завдання дослідження. *Метою дослідження* є підвищення швидкості та надійності розпізнавання мовленнєвих сигналів на основі побудови інформаційної технології аналізу, обробки та наближення спектрів мовленнєвих сигналів.

Для досягнення цієї мети в роботі виконуються такі основні завдання:

- аналіз існуючих систем розпізнавання мовлення, аналіз відомих методів апроксимації та розпізнавання мовленнєвих сигналів;
- розробка методу і моделей побудови аналітичного наближення спектрально-часового зображення мовленнєвого сигналу в класі знакозмінних функцій, що дозволить підвищити точність опису мовленнєвого сигналу та ефективність розпізнавання на його основі;
- розробка методів та алгоритмів розпізнавання мовленнєвих сигналів, які б дозволяли підвищити ефективність та швидкість розпізнавання;
- розробка методу зіставлення довжин реалізацій сигналу з кількістю входів нейронної мережі для підвищення ефективності розпізнавання;

- розробка багатокритеріальної ієрархічної системи прийняття рішення при розпізнаванні мовленнєвих сигналів для підвищення швидкості та ефективності розпізнавання;

- розробка програмного забезпечення для аналізу і розпізнавання характеристик мовленнєвої інформації та голосового керування.

Об'єкт дослідження – процеси обробки інформації про мовленнєві сигнали при аналізі та розпізнаванні.

Предмет дослідження – інформаційна технологія аналізу, обробки, наближення спектрів та прийняття рішень при розпізнаванні мовленнєвих сигналів.

У дисертаційній роботі використано такі *методи дослідження*:

- методи теорії обробки мовленнєвих сигналів, теорії спектрального аналізу для розробки математичної моделі мовленнєвого сигналу на рівні параметричної обробки;

- методи теорії розпізнавання для первинної обробки мовленнєвого сигналу, порівняння, оцінювання ефективності та надійності алгоритмів розпізнавання мовленнєвих сигналів;

- методи теорії оптимізації для побудови наближення спектрів мовленнєвих сигналів;

- методи теорії ймовірностей та математичної статистики для оцінювання адекватності результатів наближення спектрів мовленнєвих сигналів;

- теорія технології проектування інформаційних систем і теорія об'єктно-орієнтованого програмування для проектування та розробки інформаційної технології обробки, аналізу та розпізнавання характеристик мовленнєвої інформації.

Наукова новизна одержаних результатів. У ході виконання поставлених завдань отримано нові наукові результати:

1. Вперше запропоновано нові методи розпізнавання мовленнєвих сигналів на основі принципу максимуму Понтрягіна та на основі локальних екстремумів, які дозволяють зменшити складність обчислень і підвищити швидкість розпізнавання.

2. Вперше запропоновано метод зіставлення реалізацій сигналу змінної довжини, який, на відміну від існуючих, встановлює відповідність входів нейронної мережі та елементів пред'явленої реалізації, що дає змогу підвищити ефективність розпізнавання.

3. Удосконалено метод побудови аналітичного наближення спектрально-часового зображення мовленнєвого сигналу, який базується на поліпшеній моделі спектрів мовленнєвих сигналів, ураховує знакозмінність залишків спектра в методі послідовного вилучення складових, а також враховує дійсну та уявну частину спектрально-часового зображення мовленнєвого сигналу, що дозволяє підвищити точність опису мовленнєвого сигналу й ефективність розпізнавання.

4. Удосконалено багатокритеріальну ієрархічну модель, що, на відміну від існуючих, створює ієрархію дій як за параметрами подання мовленнєвого сигналу, так і за запропонованими методами, і тим самим підвищує швидкість та ефективність розпізнавання слів.

Практичне значення одержаних результатів.

Створено інформаційну технологію на основі запропонованих моделей та методів. Розроблено програмне середовище «SpeechSAnalysis», що реалізує запропоновану інформаційну технологію, містить у собі модуль голосового керування і дозволяє підвищити швидкість на 20 % та ефективність розпізнавання слів на 6 % у порівнянні з класичними системами розпізнавання мовлення для економічного словника (навчальна вибірка – 1000 слів, контрольна вибірка – 500 слів). Також система «SpeechSAnalysis» може використовуватись для проведення порівняльного аналізу ефективності методів наближення спектрально-часового зображення мовленнєвого сигналу, для порівняння ефективності методів розпізнавання, дослідження реальних сигналів і методів їх обробки, а також для проведення лабораторних робіт. Розроблене програмне середовище було вбудовано, як модуль голосового керування, в системи автоматизації робочого місця фахівця економічного та інженерного напрямку.

Одержані результати можна застосувати для створення програмного забезпечення з аналізу, обробки й розпізнавання характеристик мовленнєвої інформації та модуля голосового керування.

Результати дисертаційних досліджень впроваджені у навчальний процес кафедри математичного забезпечення ЕОМ Дніпропетровського національного університету; модуль голосового керування впроваджено в ТОВ «УкрРадиоТел» та в дочірнє підприємство «Укрнафтогазкомплект» НАК «Нафтогаз України». Впровадження результатів дисертаційних досліджень підтверджено відповідними актами.

Особистий внесок здобувача. Усі результати, які становлять основний зміст дисертації, досягнуті здобувачем самостійно. Стосовно публікацій, написаних у співавторстві, особистий внесок здобувача полягає в такому: розроблено обчислювальну схему алгоритму на основі принципу максимуму Понтрягіна та проведено аналіз методів розпізнавання мовлення [21]; розроблено обчислювальну схему побудови аналітичного опису спектра мовленнєвого сигналу в класі функцій, що описує другу похідну від функції Гауса [12]; сформульовано граматику для модуля голосового керування веб-проектом [13]; розроблено алгоритм, обчислювальну процедуру та проведено експериментальні дослідження розпізнавання мовленнєвого сигналу на основі методу локальних екстремумів [2, 9, 16]; сформульовано обчислювальні процедури для побудови наближення спектрів мовленнєвих сигналів у класі знакозмінних функцій та проведено експериментальні дослідження щодо ефективності побудованого наближення та порівняння з наближенням у класі дзвоноподібних функцій [1, 3, 8, 14, 20]; розроблено обчислювальні процедури для модуля голосового керування інтернет-магазином [15]; розроблено алгоритм та обчислювальну схему методу прийняття рішення на основі нейронних мереж [5, 17]; спроектовано, розроблено і реалізовано програмне забезпечення обробки й розпізнавання мовленнєвих сигналів [4]; сформульовано обчислювальну процедуру визначення тональності звуків [7]; розроблено алгоритм та обчислювальну схему методу на основі принципу максимуму Понтрягіна [18, 19]; проведено аналіз розроблених методів розпізнавання мовлення [21].

Апробація результатів дисертації. Основні положення і результати дисертаційної роботи доповідалися й обговорювалися на 11 науково-технічних конференціях: на 3 міжнародних конференціях «Математичне та програмне забезпечення інтелектуальних систем MPZIS» (м. Дніпропетровськ, 2004–2007 рр.); на 4 міжнародних конференціях «Проблеми математичного моделювання» (м. Дніпродзержинськ, 2004–2007 рр.); на міжнародній конференції «Автоматика 2006», на міжнародній конференції «Комп'ютерні науки та інформаційні технології» (м. Львів, 2007 р.), на міжнародній конференції «Питання оптимізації обчислень» (м. Ялта, 2005 р.), на всеукраїнській конференції «Проблеми прикладної математики та інформатики» (м. Львів, 2007 р.), на щорічних конференціях професорсько-викладацького складу, співробітників і студентів ДНУ (м. Дніпропетровськ, 2003–2007 рр.) та семінарах кафедри математичного забезпечення ЕОМ, також лягли в основу навчального посібника. Крім того, матеріали дисертаційної роботи було представлено на семінарах «Актуальні проблеми інформаційних та транспортних систем та технологій» (м. Дніпропетровськ, 2008 р.) і «Аналіз, розпізнавання та інтерпретація сигналів у технічних системах» (м. Київ, 2008 р.).

Публікації. За темою дисертаційної роботи опубліковано 21 працю: 9 наукових статей, з них 7 надруковано у виданнях, затверджених ВАК України як фахові, а також 12 тез доповідей на конференціях.

Структура та обсяг дисертації. Дисертація складається з вступу, чотирьох розділів, висновків, списку використаних джерел, додатків, містить 117 сторінок основного тексту, 44 рисунки, 14 таблиць, чотири додатки. Загальний обсяг дисертації становить 153 сторінки.

ОСНОВНИЙ ЗМІСТ

У вступі обґрунтовано актуальність теми, сформульовано мету й завдання досліджень, їх зв'язок з науковими програмами, визначено методи дослідження, наведено загальну характеристику роботи і структуру дисертації, визначено її наукову новизну, практичне значення проведених у роботі досліджень та особистий внесок автора, наведено відомості про публікації та апробацію роботи.

Перший розділ дисертаційної роботи містить аналіз методів та інформаційних технологій розпізнавання мовленнєвих сигналів. Оцінено інформаційне забезпечення розпізнавання мовленнєвих сигналів. Виконано огляд публікацій існуючих методів розпізнавання мовленнєвих сигналів, методів наближення сигналів та моделей аналітичного опису мовленнєвого сигналу. Проведений аналіз показує, що існуючі системи розпізнавання не містять адекватної моделі мовленнєвого сигналу, яка б дозволяла аналітично описувати сигнал, та не існує систем розпізнавання, які б давали змогу використовувати сукупність методів для розпізнавання мовленнєвих сигналів та порівнювати ефективність методів між собою. Також існуючі методи розпізнавання мовленнєвих сигналів не забезпечують достатню швидкодію і водночас ефективність розпізнавання.

Існує модель аналітичного наближення спектрально-часового зображення мовленнєвого сигналу, запропонована О.М. Карповим, яка має такий вигляд: мовленнєвий сигнал $s(t)$ можна зобразити як сукупність компонентів де $g_i(t)$ компоненти, що містять інформацію про властивості мовленнєвого сигналу, а саме про фонетичну структуру повідомлення, індивідуальність та емоційний стан мовця. Проста модель сигналу – це його зображення у вигляді суми компонент, кількість і властивості котрих змінюються як функція часу

Компоненти визначаються параметрами, що характеризують їх частотні й часові властивості

де – функція генератора сигналу; – описує частотні, а – часові властивості компонентів $g_i(t)$.

де q_i – параметр генератора сигналу; – ступінь загасання компоненти;

β_i – резонансна частота; T_i^* – положення максимуму в часі; – коефіцієнти посилення частотної й часової складових компонента.

Задача опису сигналу є задачею вибору виду функцій, , і визначення параметрів компонент. Для розв'язання задачі вихідна функція $s(t)$ дискретизується з частотою f_d , формуючи сигнал $s(t_j)$, який потім розбивається на інтервали аналізу T_{IN} розміром N на інтервалі при – кількість інтервалів T_{IN} , зображаючи вихідний сигнал як двовимірну функцію часу при Задача визначення параметрів у часовій області розв'язується як найкраще наближення вигляду

Якщо відомі перетворення Лапласа функцій і знайдені спектри Фур'є – інтервал аналізу, k – кількість спектральних коефіцієнтів Фур'є) на кожному інтервалі аналізу T_{IN} (рис.1), то компоненти спектрально-часового опису представляються як

Задача визначення параметрів у частотній області вирішується як найкраще наближення виду

Для задачі зображення спектрально-часового опису мовленнєвих сигналів як функції багатьох змінних наведена математична модель побудови наближення спектрально-часового зображення мовленнєвого сигналу в класі дзвоноподібних функцій.

Модель зображення спектрально-часової функції мовленнєвого сигналу:

(1)

(2)

Проведений аналіз показує, що наведені моделі побудови аналітичного наближення спектрально-часового зображення мовленнєвих сигналів не є достатньо ефективними, у зв'язку з тим, що не враховують знакозмінність різниць та залишків спектра в методі послідовного вилучення складових. Становить інтерес розробка інформаційної технології, яка б містила адекватну модель аналітичного наближення спектрально-часового зображення мовленнєвого сигналу, а також би методи розпізнавання мовленнєвих сигналів, які б дозволили підвищити швидкість та ефективність розпізнавання; також проводила розпізнавання мовленнєвих сигналів на основі ранжирування критеріїв прийняття рішення і дозволяла б аналізувати якість побудови наближення спектрально-часового зображення мовленнєвого сигналу, порівнювати ефективність та швидкодію методів розпізнавання між собою та вбудовуватись як модуль голосового керування в інформаційні системи.

У кінці розділу, виходячи з актуальності задачі алгоритмічного та інформаційного забезпечення, сформульовано постановку завдання дисертаційної роботи.

Другий розділ дисертаційної роботи присвячено розробці методів і алгоритмів побудови наближення спектрально-часового зображення мовленнєвого сигналу в класі знакозмінних функцій. Обґрунтовано використання знакозмінних функцій. Запропоновано критерії для аналізу ефективності наближення спектрально-часового зображення мовленнєвого сигналу. Спектрально-часовий опис мовленнєвого сигналу обрано в ролі вхідних даних для побудови наближення, тому що в цьому зображенні можна в явному вигляді побачити всю структуру мовленнєвого сигналу в частотно-часовій області.

Припустимо, що задано прямокутну область, а в області R задано спектрально-часову функцію, де ω_k – дискретно задана частота, t_l – дискретно заданий час.

Функція в області R містить довільну кількість сплесків спектральної енергії, які розташовані довільним чином в заданій області. Необхідно найкращим чином описати функцію в класі функцій, що описують частотні – та часові – $\{h_i(t_l)\}$ властивості мовленнєвого сигналу, і визначити параметри сплесків функції як параметри функцій.

Завданням цього розділу є вибір функцій та побудова схем зображення мовленнєвого сигналу і його спектра в класі цих функцій для моделей наближення спектра, а також розробка критеріїв аналізу результатів наближення обраними функціями.

Як показав аналіз алгоритму побудови наближення спектрально-часової функції мовленнєвого сигналу в класі зазначених функцій, залишки спектра в методі послідовного вилучення складових є знакозмінними, але при побудові наближення в класі функцій Гауса та модифікованого локона Аньєзі ця особливість не враховується. Звідси випливає, що функції, повинні бути знакозмінними з дзвоноподібним сплеском.

На підставі отриманих вимог до функцій, для побудови наближення спектрально-часового зображення мовленнєвого сигналу в класі знакозмінних функцій обрано такі функції: друга похідна від функції модифікованого локона Аньєзі, друга похідна від функції Гауса, функція ядра $\sin(x)/x$ та комплексна функція.

Запропоновано модифіковану модель представлення (3) та (4), яка відрізняється від розглянутої вище моделі (1) і (2) тим, що не містить модулів, тобто враховує знакозмінність представлених у ній функцій.

(3)

(4)

Функція збурення зі спектром:

(5)

В роботі запропоновано загальну схему для побудови наближення спектрально-часового зображення мовленнєвого сигналу в класі знакозмінних функцій. Запропоновано математичну модель наближення спектрально-часового зображення мовленнєвого сигналу в класі функції, яка описує другу похідну від функції Гауса:

Складова спектра шукається у вигляді:

де

Далі використовується різницева схема вилучення складових – СЕТ (метод послідовного вилучення складових) у прямому вигляді для спектральної функції.

Задачу розв'язано методом покоординатного спуску при цьому мінімізувалась функція

У результаті виконання алгоритму обчислюються сукупності.

Запропоновано математичну модель наближення спектрально-часового зображення мовленнєвого сигналу в класі функції, яка описує другу похідну від функції модифікованого локона Аньєзі:

Також в роботі запропоновано математичну модель наближення спектрально-часового зображення мовленнєвого сигналу в класі функції ядра $\sin x/x$ (рис. 1):

Рис. 1. Графік функції ядра $\sin x/x$

Обчислення параметрів виконано аналогічно до того, як і в класі функції, яка описує другу похідну від функції Гауса.

Математична модель наближення спектрально-часового зображення мовленнєвого сигналу в класі комплексних функцій (рис. 2):

Рис. 2. Графіки комплексної функції: дійсна та уявна частини.

Комплексна функція – це природна функція для опису параметрів мовотворчого тракту, фізична й математична моделі збігаються, тому вона найкраще підходить для побудови наближення спектрально-часового зображення мовленнєвого сигналу. Комплексні функції використовуються в моделі тільки для частотної функції, часова функція зображена другою похідною від функції Гауса.

Модифікована модель для пошуку складової спектра має вигляд:

В зображеній схемі враховується дійсна та уявна частина функції $W_{(i)}(\omega_k)$ мовотворчого тракту. Модифікована різницева схема вилучення складових – СЕТ у прямому вигляді для спектральної функції при побудові наближення в класі комплексних функцій матиме такий вигляд:

Далі задача розв'язується методом покоординатного спуску. У результаті виконання алгоритму обчислюються сукупності.

Запропоновані моделі дозволяють отримувати параметри моделей мовотворчого тракту і будувати аналітичний опис спектрально-часового зображення мовленнєвого сигналу та враховують знакозмінність залишків спектра в методі послідовного вилучення складових, що дозволяє підвищити точність наближення спектрально-часового зображення мовленнєвого сигналу та ефективність розпізнавання на основі побудованого опису.

Третій розділ дисертаційної роботи присвячено розробці методів та алгоритмів розв'язання задачі нелінійного розтягання, зіставлення вибірок довільної довжини та розпізнавання мовленнєвих сигналів.

Метод розпізнавання мовленнєвих сигналів на основі локальних екстремумів. У роботі пропонується підхід в кутових координатах для розв'язання задачі нелінійного розтягання та зіставлення вибірок довільної довжини. Припустимо, задано дві вибірки X довжиною n_1 і Y довжиною n_2 . Припустимо, що всі ці довжини відповідають інтервалу 2π – вибірки визначені на колі.

Задача пошуку відповідності між вибірками X довжиною n_1 і Y довжиною n_2 полягає в пошуку відповідності їхніх ділянок на кроці k між X_k і Y_k при перегляді на кроці 1 із точок X_1 і Y_1 . Критерієм закінчення пошуку на кожному кроці визначення точок ix для X на інтервалі $1, \dots, n_1$ і iy для Y на інтервалі $1, \dots, n_2$ є положення екстремуму відповідності між X_k і Y_k .

У цьому випадку шукаємо локальний екстремум у деякій невеликій частині вибірок X_k і Y_k , послідовно переміщуючись від першої точки до точок n_1, n_2 , реалізуючи спрямоване перебирання.

Зазначений алгоритм, так як і метод динамічного програмування, намагається встановити загальну довжину й у тому випадку, коли, наприклад, одна вибірка за фонемним складом міститься в іншій. У цьому випадку можливий другий алгоритм, що закінчує роботу не на інтервалі $0, \dots, 2\pi$, а на інтервалі $0, \dots, \phi$, де $\phi < 2\pi$. Третій алгоритм обробляє послідовності, якщо, наприклад, одна вибірка розташована всередині іншої. У цьому випадку інтервал пошуку відповідності міститься в межах ϕ_1, \dots, ϕ_2 , де $\phi_1 > 0$, $\phi_2 < 2\pi$.

Для розв'язання задачі розпізнавання мовленнєвих сигналів пропонується алгоритм на основі методу локальних екстремумів. Методом локальних екстремумів шукаються точки, де проходять можливі межі сегментів (рис. 3).

рис. 3. Схема алгоритму на основі локальних екстремумів

Для кожної точки зберігаються всі можливі шляхи до неї. Шукається шлях до діагоналі, який містить локальний екстремум. Якщо такий шлях знайдено, то відтинається частина простору, і пошук починається від знайденої точки.

Метод розпізнавання мовленнєвих сигналів на основі принципу максимуму Понтрягіна. Перехід до принципу максимуму Понтрягіна відбувається так: спектрально-часове зображення мовленнєвого сигналу – це похідна від інтегральної функції сенсу (розуміння висловлювання), і її можна зіставити зі швидкістю зміни якого-небудь процесу. При висловленні відбувається накопичення рівня розуміння.

Розглядається поведінка об'єкта, стан якого в кожен момент часу характеризується n дійсними числами x^1, x^2, \dots, x^n . Для зіставлення реалізацій в задачі розпізнавання мовлення ці числа є координатами точки, тобто $n = 2$.

Нехай задана динамічна система

з початковою умовою $x(t_0) = x^0$ й областю керування U .

Закон зміни стану описується такою динамічною системою:

(7)

тобто (7) описує динаміку переходу від однієї чарунки до іншої, де x_1 – описує положення у рядку, x_2 – у стовпці.

Відомі початковий і кінцевий стан системи та інтервал керування.

(8)

Область керування, тоді величини керування u_1 та u_2 відповідають таким вимогам

(9)

Подібні значення функцій керування дозволяють об'єктові рухатися в двох напрямках – праворуч або вниз. Необхідно побудувати таке керування u_1, u_2 , щоб на інтервалі керування $n = 0, N$ виконувалась умова мінімуму цільової функції.

(10)

де $d(x_1(n), x_2(n))$ – це відстань між станами $x_1(n)$ пред'явленої та $x_2(n)$ еталонної реалізаціями.

Для розв'язання задачі розпізнавання мовленнєвих сигналів в роботі було запропоновано такий вираз для гамільтоніана:

У результаті аналітичних викладок отримано такі вирази:

Метод розпізнавання мовленнєвих сигналів на основі нейронної мережі. Для зіставлення довжин реалізацій при розпізнаванні мовленнєвих сигналів на основі нейронної мережі побудовано модульну нейронну мережу, в якій кожен нейрон-модуль є також нейронною мережею та являє собою розподілену мережу прямого розповсюдження із затримкою в часі, яка ґрунтується на використанні просторово-часової моделі нейрона –

нейронного фільтра з декількома входами. На вході нейронної мережі – слово. Для кожного нейрона-модуля навченої нейронної мережі зберігається інформація про кількість сегментів мовленнєвого сигналу, на якій він навчався. При розпізнаванні вхідного мовленнєвого сигналу до нейронної мережі додається ще один нейрон-модуль зі стандартною структурою. Для цього нейрона проводиться навчання за вхідним сигналом і за стандартним алгоритмом навчання (зворотного розповсюдження в часі).

Подальше прийняття рішення відбувається таким чином: для прийняття рішення зіставляються синаптичні ваги кожного нейрона-модуля за будь-яким з наведених вище алгоритмів прийняття рішення для розпізнавання мовленнєвого сигналу, наприклад методом на основі принципу максимуму Понтрягіна, методом локальних екстремумів і приймається рішення щодо мовленнєвого сигналу.

При використанні методу розпізнавання на основі лінгвістичного опису на першому етапі проводиться первинна обробка та побудова лінгвістичного опису мовленнєвого сигналу. Для опису мовленнєвого сигналу використовуються такі ознаки: «максимум», «середнє» та «мінімум».

Багатокритеріальний ієрархічний підхід. Задача розпізнавання полягає у виборі оптимальної послідовності розпізнавання та звуженні підмножини слів, що розпізнаються за часом T і за ефективністю розпізнавання. Фізичне (акустичне) трактування полягає в тому, що на кожному рівні визначається деяка підмножина слів $U_{k,l}$, для котрих властиві деякі спільні ознаки (або одна ознака). Коефіцієнт розподілу поточної підмножини

де n_o – обсяг словника, n_l – кількість слів підмножини U_l , якій властива дана ознака при $l = 1 \div m$, коли Q_l задає коефіцієнт розподілу вихідної множини U для незалежної класифікації за окремими рівнями. Оптимальна послідовність розпізнавання буде мати вигляд:

де t_m , на відміну від існуючих послідовностей, відповідає як властивостям параметрів так і складності методу.

У разі багатокритеріального підходу на основі ранжирування критеріїв прийняття рішення метою є пошук найкращого рішення, тобто пошук еталонного сигналу, який більш за все відповідає вхідному. Критерії оцінки ефективності рішення у запропонованому багатокритеріальному підході для кожного методу розпізнавання використовуються різні. У методі розпізнавання на основі лінгвістичного опису – це найбільша відповідність лінгвістичних описів еталонної та пред'явленої реалізації. У методі на основі принципу максимуму Понтрягіна – це максимум гамільтоніану. У методі на основі нейронних мереж – це значення виходів нейронної мережі. У методах на основі локальних екстремумів та методі динамічного програмування – це найбільша відповідність між представленою реалізацією та еталонною.

Розпізнавання мовленнєвого сигналу відбувається за наступною послідовністю методів: розпізнавання на основі лінгвістичного опису, метод на основі принципу максимуму Понтрягіна (за ССЗ), метод на основі нейронних мереж (за ССЗ), метод на основі локальних екстремумів (за СЧЗ), метод динамічного програмування (за СЧЗ); де ССЗ – спектрально-смугове зображення мовленнєвого сигналу, СЧЗ – спектрально-часове зображення мовленнєвого сигналу.

Запропоновані методи та алгоритми розв'язання задачі нелінійного розтягання, зіставлення вибірок довільної довжини та розпізнавання мовленнєвих сигналів, які дозволяють підвищити швидкість та ефективність розпізнавання.

Четвертий розділ присвячено опису, практичній реалізації та аналізу ефективності розробленої інформаційної технології. Розроблена інформаційна технологія може бути використана для проведення порівняльного аналізу ефективності методів наближення спектрально-часового зображення мовленнєвого сигналу, для порівняльного аналізу ефективності методів розпізнавання, дослідження реальних сигналів, методів їх обробки та як модуль голосового керування. У ролі вхідних даних використовується тип мовленнєвої

одиниці – слова. Побудовано та реалізовано структуру граматики і словників для системи голосового керування системою «Ведення та облік персоналу». Проведені дослідження за розробленими методами та алгоритмами.

На базі запропонованої обчислювальної технології автором створено автоматизовану систему «SpeechSAnalysis» для аналізу, обробки й розпізнавання мовленнєвих сигналів.

Проектування системи виконано на уніфікованій мові моделювання (UML) за допомогою засобу Microsoft Office Visio 2007. Специфікацію загальних властивостей функціонування системи наведено на діаграмі використання (рис. 4).

Рис. 4. Діаграма варіантів використання системи «SpeechSAnalysis»

Система «SpeechSAnalysis» реалізована на мові Object Pascal в середовищі Borland Delphi 7.0.

Побудова аналітичного опису спектрально-часового зображення мовленнєвого сигналу проводилась для двох випадків, наближення при отриманні параметрів частотної функції (рис. 5б – у класі функцій, що описують другу похідну від функції Гауса, рис. 6а – у класі функцій $\sin x/x$) та наближення при отриманні параметрів мовотворчого тракту (рис. 6б – у класі функцій, що описують другу похідну від функції Гауса).

а) б)

Рис. 5. Вихідна спектрально-часова структура слова «чотири» (а), аналітично описана спектрально-часова структура (б)

Аналіз ефективності побудови аналітичного наближення спектрально-часового зображення мовленнєвого сигналу проводився за двома критеріями: найкраще наближення до початкового сигналу та найкраща ефективність розпізнавання. Найбільш ефективною за розглянутими критеріями виявилася побудова наближення в класі комплексних функцій. Отримані результати ще раз доводять, що комплексна функція – це природна функція для опису параметрів мовотворчого тракту.

Рис. 6. Аналітично описана спектрально-часова структура слова «чотири»

Автором було проведено аналіз ефективності запропонованих методів розпізнавання мовленнєвих. Також було проведено аналіз ефективності запропонованої модифікації багатокритеріального ієрархічного підходу (табл. 1). Аналіз проводиться за допомогою розробленої нової інформаційної технології, системи «SpeechSAnalysis» на економічному словнику (навчальна вибірка – 1000 слів, контрольна вибірка – 500 слів).

Таблиця 1

Аналіз запропонованого багатокритеріального ієрархічного підходу

ССЗ – спектрально-смугове зображення мовленнєвого сигналу, СЧЗ – спектрально-часове зображення мовленнєвого сигналу.

Ефективність розпізнавання розробленої системи на економічному словнику з навчальною вибіркою 1000 слів та контрольною вибіркою 500 слів (частина навчальної) – 98 %. За результатами проведеного експерименту розпізнавання командної мови для систем «Горыныч» та «Диктограф» можна зробити висновок, що ефективність розпізнавання в цих системах на тому ж самому словнику становить 92 та 90 % відповідно. Швидкість розпізнавання для розглянутих систем становить в середньому 1,2 с на одне слово.

За проведеним аналізом можна зробити висновок, що застосування різних методів розпізнавання в запропонованій ієрархічній багатокритеріальній системі в середньому дозволяє підвищити ефективність розпізнавання на 6 % і скоротити час обробки на 20 % у порівнянні з класичними системами розпізнавання мовленнєвих сигналів.

Розроблене програмне середовище впроваджено на дочірньому підприємстві «Укрнафтогазкомплект» НАК «Нафтогаз України» та в ТОВ «УкрРадиоТел».

У висновках сформульовано основні результати досліджень.

Додатки містять результати побудови лінгвістичного опису мовленнєвих сигналів для різних реалізацій; список слів тестового матеріалу зі словника; організацію інтерфейсу

системи «SpeechSAnalysis»; результати обчислювального експерименту з аналітичного опису мовленнєвих сигналів і документи про впровадження результатів дисертаційної роботи.

ОСНОВНІ РЕЗУЛЬТАТИ РОБОТИ

У результаті виконаних досліджень розв'язано актуальне завдання створення інформаційної технології обробки та аналізу характеристик мовленнєвої інформації, на основі якої досягнуто мети підвищення швидкості та надійності розпізнавання мовленнєвих сигналів (мовленнєвих команд).

Після дослідження вибраного об'єкта отримано основні наукові й практичні результати дисертаційної роботи:

1. Вперше запропоновано нові методи розпізнавання мовленнєвих сигналів на основі принципу максимуму Понтрягіна та на основі локальних екстремумів, які дозволяють зменшити складність обчислень і підвищити швидкість розпізнавання.

2. Вперше запропоновано метод зіставлення реалізацій сигналу змінної довжини, який, на відміну від існуючих, встановлює відповідність входів нейронної мережі та елементів пред'явленої реалізації, що дає змогу підвищити ефективність розпізнавання.

3. Удосконалено метод побудови аналітичного наближення спектрально-часового зображення мовленнєвого сигналу, який базується на вдосконаленій моделі спектрів мовленнєвих сигналів, враховує знакозмінність залишків спектра в методі послідовного вилучення складових, враховує дійсну та уявну частину спектрально-часового зображення мовленнєвого сигналу, що дозволяє підвищити точність опису мовленнєвого сигналу та ефективність розпізнавання.

4. Удосконалено багатокритеріальну ієрархічну модель що, на відміну від існуючих, створює ієрархію дій як за параметрами подання мовленнєвого сигналу, так і за запропонованими методами, і тим самим дозволяє підвищити швидкість та ефективність розпізнавання.

5. Створено інформаційну технологію обробки та аналізу характеристик мовленнєвої інформації, яка дозволяє будувати аналітичний опис спектра мовленнєвого сигналу і проводити розпізнавання на основі ранжирування критеріїв прийняття рішення. Розроблена інформаційна технологія може бути використана як модуль голосового керування для автоматизованого робочого місця фахівця економічного профілю, а також для порівняльного аналізу ефективності методів побудови аналітичного опису спектрів мовленнєвих сигналів і методів розпізнавання.

6. Розроблено програмне забезпечення «SpeechSAnalysis» для аналізу та розпізнавання мовленнєвих сигналів на основі ранжирування критеріїв прийняття рішення, в якому реалізовано запропоновані методи й моделі.

7. Результати дисертаційної роботи впроваджені в ТОВ «УкрРадиоТел», використовуються в навчальному процесі Дніпропетровського національного університету. Запропонована інформаційна технологія використовується ТОВ «УкрРадиоТел» як модуль голосового керування в системі «Розрахунок санітарних зон та зон обмеження забудови» та дочірнім підприємством «Укрнафтогазкомплект» НАК «Нафтогаз України» як модуль голосового керування в системі «Ведення та облік персоналу».

СПИСОК ОПУБЛІКОВАНИХ ЗА ТЕМОЮ ДИСЕРТАЦІЇ ПРАЦЬ

1. Зирнеева Г.В. Сравнение свойств колебательных функций в задаче анализа спектров речевых сигналов / Зирнеева Г.В., Карпов О.М. //Актуальні проблеми автоматизації та інформаційних технологій. – Д.: ДНУ, 2004. – Т. 8. – С. 57–62.

2. Зирнеева Г.В. Метод вариаций в задаче распознавания речи / Зирнеева Г.В., Карпов О.М. //Актуальні проблеми автоматизації та інформаційних технологій. – Д.: ДНУ, 2005. – Т. 9. – С. 57–62.

3. G. Zirneeva Representation of the spectral-time description in the class of bell-shaped and bell-shaped-oscillatory functions of speech recognition / G. Zirneeva, O. Karпов // Актуальні проблеми автоматизації та інформаційних технологій. – Д.: ДНУ, 2007. – Т. 11. – С. 74–90.
4. Зірнєєва Г.В. Інформаційна технологія побудови системи розпізнавання мовних сигналів та модуля голосового керування / Зірнєєва Г.В., Карпов О.М. // Вісник Академії митної служби України. – Д.: Академія митної служби України, 2008. – № 1 (37). – С.61–68
5. Зірнєєва Г.В. Прийняття рішення при розпізнаванні мови на основі нейронних мереж / Зірнєєва Г.В., Карпов О.М. //Математичне моделювання. – Дніпродзержинськ. : ДДТУ, 2007. – № 2 (17). – С. 106–109.
6. Зірнєєва Г.В. Порівняльний аналіз алгоритмів прийняття рішень в задачі розпізнавання мовлення / Зірнєєва Г.В. // Комп'ютерні науки та інформаційні технології. – Львів, 2007. – № 604. – С. 60–64.
7. Зирнеева Г.В. Алгоритм формирования параметров и оценки функционального состояния говорящего по его речи / Карпов О.М., Чугай А.А., Зирнеева Г.В., Асадулин В.А. //Актуальні проблеми автоматизації та інформаційних технологій. – Д.: ДНУ, 2006. – Т. 10. – С. 62–71.
8. Зирнеева Г.В. Описание спектрально временного представления речевых сигналов в классе производных функций Гаусса второго порядка, / Зирнеева Г.В., Карпов О.М. // Питання прикладної математики та математичного моделювання. – Д.: ДНУ, 2004. – Т. 11. – С. 88–97.
9. Зирнеева Г.В. Модели согласования длин речевых реализаций в угловых координатах при распознавании речи / Зирнеева Г.В., Карпов О.Н., Басюк А.К. // Питання прикладної математики і математичного моделювання. – Д.: ДНУ, 2005. – С. 95–103.
10. Зирнеева Г.В. Анализ спектров речевых сигналов в классе комплексных функций / Зирнеева Г.В. // Проблеми прикладної математики та комп'ютерних наук: тематична наук. конференція за підсумками наук.-дослідн. роботи ДНУ за 2004–2005 рр. – Дніпропетровськ, 2006. – С. 27–28.
11. Зирнеева Г.В. Архитектура модуля голосового управления web-проектами / Зирнеева Г.В. // Проблеми математичного моделювання: міжнар. наук.-метод. конф., 23–25 трав. 2007 р.: тези доп. – Дніпродзержинськ, 2007. –С. 174–175.
12. Зирнеева Г.В. Речевые технологии в системе компьютерных технологий / Зирнеева Г.В. Басюк А.К. // Единое информационное пространство: III междунар. конф., 8–9 декабря 2005 г.: тезисы докл. – Днепропетровск, 2005. – С. 14–15.
13. Зирнеева Г.В. Использование грамматик в построении системы речевого управления веб-приложением / Зирнеева Г.В., Карпов О.Н. //Математичне та програмне забезпечення інтелектуальних систем MPZIS-2006, 15–17 лист. 2006 р.: тези доп. – Дніпропетровськ, 2006. – С. 170–171.
14. Зирнеева Г.В. Экспериментальные исследования и обработка спектров речевых сигналов в классах колебательных и колоколообразных функций / Зирнеева Г.В., Карпов О.М. // Математичне та програмне забезпечення інтелектуальних систем MPZIS 2004, 17–19 лист. 2004 р.: тези доп. – Дніпропетровськ. –2004. – С. 47–48
15. Зирнеева Г.В. Использование Интернет технологий в построении системы распознавания речи / Зирнеева Г.В., Карпов О.Н. //Интернет – Образование – Наука 2006: 5- междунар. практ. конф.: тези доп. – Винница, 2006. – С. 57–58.
16. Зирнеева Г.В. Построение деревьев решений методом локальных экстремумов в задаче распознавания речи / Зирнеева Г.В., Карпов О.Н. // Проблеми математичного моделювання: міжнар. наук.-метод. конф., 24–26 трав. 2006 р.: тези доп. – Дніпродзержинськ, 2006. – С. 161–162.
17. Зирнеева Г.В. Принятие решения при распознавании слов речи на основе нейронных сетей / Зирнеева Г.В., Карпов О.Н., Палагута И.В. // Проблеми математичного моделювання:

міжнар. наук.-метод. конф., 23-25 трав. 2007 р.: тези доп. – Дніпродзержинськ, 2007. – С. 180–181.

18. Зірнеєва Г.В. Задача розпізнавання мовлення принципами максимуму Понтрягіна та правдоподібності / Зірнеєва Г.В., Карпов О.М. // Сучасні проблеми прикладної математики та інформатики: всеукр. наук. конф., 2–4 жовт. 2007: тези доп. – Львів, 2007. – С.66–67.

19. Зірнеєва Г.В. Застосування принципу максимуму Понтрягіна в задачі прийняття рішення на основі нейронних мереж / Зірнеєва Г.В., Карпов О.М. // Математичне та програмне забезпечення інтелектуальних систем MPZIS-2007, 14–16 листопада 2007 р.: тези доп. – Дніпропетровськ, 2007. – С. 63.

20. Зирнеева Г.В. Экспериментальные данные анализа свойств речевых сигналов в классе колебательных и колоколообразных функций / Зирнеева Г.В., Карпов О.Н. // Питання оптимізації обчислень (ПОО-XXXII): міжнар. конф., 19–23 вересня 2005 р.: тези доп. – Київ, 2005. – С. 92–93.

21. G. Zirneeva Comparison methods of making decision for speech recognition system / G. Zirneeva, O. Karpov // Computer Science and Information Technologies: intern. conf., septemb. 27th–29th 2007. : thes. pap. – Lviv, 2007. – P. 82–83.

АНОТАЦІЇ

Зірнеєва Г.В. «Інформаційна технологія обробки та аналізу характеристик мовленнєвої інформації». – Рукопис. Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Вінницький національний технічний університет, Вінниця - 2008.

Дисертацію присвячено розробці інформаційної технології обробки та аналізу характеристик мовленнєвої інформації для побудови систем розпізнавання мовленнєвих сигналів, систем голосового керування. Запропоновано нові методи розпізнавання мовленнєвих сигналів на основі принципу максимуму Понтрягіна та на основі локальних екстремумів. Запропоновано метод зіставлення реалізацій сигналу змінної довжини, який, на відміну від існуючих, встановлює відповідність входів нейронної мережі та елементів пред'явленої реалізації. Удосконалено метод побудови аналітичного наближення спектрально-часового зображення мовленнєвого сигналу, який базується на вдосконаленій моделі спектрів мовленнєвих сигналів, враховує знакозмінність залишків спектра в методі послідовного вилучення складових, враховує дійсну та уявну частину спектрально-часового зображення мовленнєвого сигналу. Удосконалено багатокритеріальну ієрархічну модель розпізнавання мовленнєвих сигналів.

Ключові слова: наближення спектрально-часового зображення мовленнєвих сигналів, обробка та аналіз мовленнєвих сигналів, розпізнавання мовленнєвих сигналів, багатокритеріальний ієрархічний підхід.

Зирнеева Г.В. «Информационная технология обработки и анализа характеристик речевой информации». – Рукопись. Диссертация на соискание научной степени кандидата технических наук по специальности 05.13.06. – информационные технологии. – Винницкий национальный технический университет, Винница - 2008.

Диссертация посвящена разработке информационной технологии обработки и анализа характеристик речевой информации для построения командных систем распознавания речевых сигналов, систем голосового управления.

Предложены новые методы распознавания речевых сигналов на основе принципа максимума Понтрягина и на основе локальных экстремумов, что позволило увеличить скорость распознавания речевых сигналов для командных систем. Проведен анализ сложности вычислений предложенных методов и экспериментально проанализирована их эффективность.

Предложен метод сопоставления реализаций сигнала переменной длины, который в отличие от существующих, устанавливает соответствие входов нейронной сети и элементов

предъявленной реализации. В методе используется модульная нейронная сеть, где каждый нейрон-модуль также является нейронной сетью и представляет собой распределенную сеть прямого распространения с задержкой во времени, которая основана на использовании пространственно-временной модели нейрона – нейронного фильтра с несколькими входами.

Усовершенствован метод построения аналитического приближения спектрально-временного представления речевого сигнала, который базируется на усовершенствованной модели спектров речевых сигналов, учитывает знакопеременность остатков спектра, в методе последовательного извлечения составляющих, учитывает действительную и мнимую часть представления речевого сигнала. Использование предложенного метода позволяет повысить точность построения аналитического представления спектров речевых сигналов и эффективность распознавания. Введены три критерия для сравнения предложенных моделей. Это критерии – качество приближения, эффективность распознавания на основе построенного приближения и качество фильтрации.

Для систем распознавания речи предлагается использовать усовершенствованный многокритериальный иерархический подход, который, в отличие от существующего, создает иерархию действий как по параметрам представления речевого сигнала, так и по методам распознавания.

На основе предложенных методов и моделей разработана информационная технология. Составной частью информационной технологии является программная среда «SpeechSAnalysis» для анализа, обработки и распознавания речевых сигналов. Проектирование системы выполнено на унифицированном языке моделирования UML с помощью Microsoft Office Visio 2007. Разработанная информационная технология может быть использована как модуль голосового управления для автоматизированного рабочего места специалиста экономического профиля, а также для сравнительного анализа эффективности методов построения аналитического описания спектров речевых сигналов и методов распознавания.

Система «SpeechSAnalysis» позволяет вводить, анализировать, обрабатывать речевой сигнал, строить приближение спектрально-временного представления речевого сигнала, проводить распознавание различными методами и на основе многокритериального иерархического подхода. Разработанная информационная технология используется как модуль голосового управления в ТОВ «УкрРадиоТел» и дочерним предприятием «Укрнафтогазкомплект» НАК «Нафтогаз України».

Ключевые слова: приближение спектрально-временного представления речевых сигналов, обработка и анализ речевых сигналов, распознавание речевых сигналов, многокритериальный иерархический подход.

Zirneeva G.V. «Information technology of processing and the analysis of characteristics of the speech information». - a manuscript. The dissertation on competition of a scientific degree of Cand.Tech.Sci. on a speciality 05.13.06. - information technologies. – Vinnytsia National Technical University, Vinnytsia - 2008.

The dissertation is dedicated to information technology of processing and the analysis of characteristics of the speech information of speech signals and voice control systems development.

The new methods are offered on the basis of a principle of Pontryagin's maximum and in the basis of a local extrema. The method of comparison of lengths of realizations for a neural network with quantity of entrance is offered. It is modified the method of spectral-time presentation of speech signals which is based on advanced model of spectra of speech signals which as against existing takes into account the oscillatory nature of a speech signal and also takes into account the method of consecutive extraction of components and takes into account real and imaginary part of spectral-time presentation of speech signals. It is modified multicriterion the hierarchical model of recognition of speech signals.

Key words: approaching spectral-time presentation of speech signals, processing and analysis speech signals, recognition of speech signals, multicriterion the hierarchical approach.