

Ст. викладач Т.С.Ладогубець, студент Ю.В. Боровик

Національний технічний університет України  
«Київський політехнічний інститут»

## РОЗПІЗНАВАННЯ МОВЛЕННЄВОГО СИГНАЛУ НА ОСНОВІ АУДИТОРНОЇ МОДЕЛІ

### Abstract

*Tetyana S. Ladogubets; Yuri Borovik , student  
Recognizing speech signal based audio model*

*This work is devoted to converting speech into electronic text. It introduces basic methods of automatic speech recognition information. This work also developed and implemented algorithm, model and software for recognition of isolated words.*

### Вступ

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 [1].

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 [1].

Проведений огляд технологій розпізнавання, а також обробки та аналізу мовленнєвих сигналів показує, що цей напрямок розвинений недостатньо, не відповідає потребам користувачів [1].

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 [1].

## **Постановка задачі**

Метою даної роботи є створення інтелектуальних мовленнєвих інформаційних технологій для усного спілкування людини з машиною природною мовою.

Об'єкт інформатизації — мовленнєвий сигнал, мовленнєві інформаційні технології, мовленнєве спілкування між людиною та машиною.

Предмет інформатизації — моделі розпізнавання мовленнєвого сигналу, огляд технологій розробки систем, які стосуються заданої проблеми.

## **Огляд існуючих рішень**

Розробка систем розпізнавання голосу протягом останніх років стала важливим напрямом у сфері інформаційних технологій. Це пов'язано у першу чергу із зростанням обчислювальних здатностей комп'ютерної техніки, широким поширенням звукоприймальних засобів та збільшенням кількості мобільних пристроїв, розмір яких ускладнює введення інформації.

У дослідженнях цієї області задіяні водночас наукові установи та великі корпорації. Зокрема, питаннями розпізнавання мовленнєвого сигналу займаються Міжнародний науково-навчальний центр інформаційних технологій та систем, Національний технічний університет України «Київський політехнічний інститут» та Інститут кібернетики Національної академії наук України [2].

Серед основних напрямків досліджень варто виділити розробку алгоритмів точного опису мовленнєвого сигналу, створення нових методик розпізнавання ізольованих слів та злитого мовлення, вдосконалення вже існуючих методів та зменшення похибки розпізнавання.

Для розпізнавання слів в злитій мові випробувано два різні підходи. У першому випадку слово, яке необхідно розпізнати, порівнюється з акустичним еталоном кожного слова у словнику. При цьому використовується математичний опис мовленнєвого сигналу у деякому просторі ознак, наприклад, на основі спектрального представлення кожного слова. Серед різних методів порівняльного типу добрі результати дав метод динамічного програмування.

У другому випадку кожне слово або група слів спочатку сегментується на менші одиниці. Сегментами є складоподібні або фонемоподібні одиниці. Це дозволяє проводити розпізнавання або на

складовому, або на фонемному рівні і одночасно зберігати в пам'яті параметри (тривалість, енергію і тому подібне) [3,4].

Для розпізнавання мовлення використовують в основному алгоритми ELVIRS і ELVIRCOS, що складаються з двох частин [5]:

Підготовчий етап:

- підготувати словник для розпізнавання;
- вибрати множину фонем і створити транскрипції слів із словника за допомогою правил.

Етап розпізнавання:

- застосувати фонетичний стенограф до вхідного сигналу для отримання послідовності фонем;
- поділити отримані фонем на трійки і створити запити до БД з трійок фонем;
- одержати списки транскрипцій за допомогою запитів до індексу бази даних;
- алгоритм ELVIRCOS створює граф слів злитого мовлення, а алгоритм ELVIRS переходить зразу на наступний крок;
- упорядкувати транскрипції до їх рангу;
- вибрати перші  $N$  транскрипцій з найвищими рангами як підсловник для розпізнавання;
- далі для алгоритму ELVIRS розпізнати вхідний мовний сигнал в умовах обмеженого підсловника, а для алгоритму ELVIRCOS розпізнати вхідний мовний сигнал для графа зливої мови в умовах обмеженого підсловника [5].

Початкові етапи цих алгоритмів співпадають, але перший алгоритм використовується для розпізнавання окремих слів, другий можна використовувати як для окремих слів, так і для злитого мовлення.

### **Аналіз результатів роботи**

Після розгляду методик розпізнавання мовленнєвого сигналу було створено систему розпізнавання мовленнєвих сигналів. За основу даної системи було взято базову систему розпізнавання мовленнєвого сигналу, яка використовується при реалізації аудиторної моделі розпізнавання мовлення. А для попереднього опису мовленнєвого сигналу було взято опис мовленнєвих сигналів за допомогою кепстральних коефіцієнтів.

Після створення даної системи було проведено ряд тестів для оцінки точності розпізнавання окремих слів. Дані тестові дослідження, а саме розпізнавання 1000 різних слів для кількох дикторів (два чоловічі голоси і два жіночі голоси), показали що при використанні опису мовленнєвого сигналу за допомогою кепстральних коефіцієнтів ми маємо:

- 96% точність розпізнавання чоловічого голосу;
- 82% точність розпізнавання жіночого голосу.

### **Висновок**

Система, в якій був застосований зазначений опис мовленнєвих сигналів за допомогою кепстральних коефіцієнтів, показала високі результати розпізнавання мовлення, оскільки в подібних системах точність розпізнавання чоловічого голосу складає близько 93%, а жіночого – близько 80% [5]. Це свідчить про ефективність використання даної методики розпізнавання. Але реалізовані методи не є ідеальними, тому є можливість їх вдосконалення, розробки їх модифікацій для того, щоб покращити якість розпізнавання мовлення, створення інтелектуальних мовленнєвих інформаційних технологій для усного спілкування людини з машиною природною мовою.

### **Література**

1. «Розроблення принципів, методів та інформаційної технології автоматичного розпізнавання, смислової інтерпретації та синтезу мовленнєвих сигналів в умовах необмежених словників» – Звіт про науково-дослідну роботу / МННЦТіС. – Київ. – 2003-2006. – С 12-107.
2. *Винцюк Т.К.* Анализ, распознавание и смысловая интерпретация речевых сигналов. – Киев: Наукова думка, 1987. – С 267.
3. *Сажок М.М.* Автоматизовані засоби формування баз даних і знань для озвучення українських текстів. – Дис. канд. техн. наук /МННЦТіС. – Київ, 2004. – С 142.
4. *Винцюк Т.К.* Сравнительный теоретический анализ ИКДП- и НММ-методов распознавания речи // Автоматическое распознавание слуховых образов. Тезисы докл. 15-й Всесоюзного семинара (Таллин-89). – Таллин: ИК АН ЭССР. – 1989. – С 18–24.
5. *Сажок М.М., Селюх Р.А., Юхименко О.А.* // Адаптация акустических моделей фонем на голос диктора для фонемного распознавания изолированных слов украинского языка. – Киев, «Искусственный интеллект». – 2009'4. – С 230-233.