

# Використання інтелектуальних систем для голосового керування комп'ютером

Довбиш А.С., Нонко Д.Ю., Ободяк В.К., Шелехов І.В.  
СумДУ, vobodyak@id.sumdu.edu.ua

*Existing programs for voice control of a computer. It is shown that the known solutions have certain drawbacks. Elaboration of software for voice control computer using intelligent technologies. Software implementation is made in the programming environment Borland Delphi 7 using the audio library bass.dll to receive and process the input audio.*

## ВСТУП

Проблема доступності комп'ютерних технологій для користувачів з обмеженнями зору останній час набуває все більшої актуальності. Це обумовлено збільшенням навантаження на зоровий апарат в сучасному світі і, як наслідок, підвищення числа людей з такими вадами здоров'я.

Допоміжні технології для незрячих і людей зі слабкий зором, що забезпечують доступ до комп'ютерної техніки, існують вже не одне десятиліття. Під «допоміжними технологіями» будемо розуміти програмне, апаратне або апаратно-програмне рішення, покликане підвищити, зберегти або оптимізувати функціональні здібності людей з обмеженими можливостями.

В роботі розглядаються особливості застосування інформаційно-екстремальної інтелектуальної технології (ІЕІТ) для розробки системи голосового керування комп'ютером, яке є одним із можливих варіантів такої допоміжної технології.

## ПОСТАНОВКА ЗАДАЧІ

Відомі алфавіт класів розпізнавання  $\{X_m^o | m = \overline{1, M}\}$ , який характеризує  $M$  можливих голосових команд і навчальна матриця типу «об'єкт-властивість»  $\|y_{m,i}^{(j)}\|$ ,  $i = \overline{1, N}$ ,  $j = \overline{1, n}$ , де  $N$ ,  $n$  – кількість ознак розпізнавання та реалізацій кожної команди відповідно.

Необхідно на етапі навчання інтелектуальної системи побудувати оптимальне (тут і далі в інформаційному розумінні) нечітке розбиття  $\mathfrak{R}^M$  простору ознак розпізнавання  $\Omega$  на класи і на етапі екзамену за максимальним значенням функції належності ідентифікувати голосову команду і виконати відповідні їй дії.

## МАТЕМАТИЧНІ МОДЕЛІ

Математичні моделі функціонування системи голосового керування в режимі навчання та екзамену реалізують класичну задачу розпізнавання образів за відповідними базовими методами ІЕІТ [1]. Як критерій функціональної ефективності (КФЕ) системи в роботі використовується інформаційна міра Кульбака:

$$J_m^{(k)} = 0,5 \log_2 \left( \frac{D_1^{(k)} + D_2^{(k)}}{\alpha^{(k)} + \beta^{(k)}} \right) [(D_1^{(k)} + D_2^{(k)}) - (\alpha^{(k)} + \beta^{(k)})]$$

де  $D_1^{(k)}$ ,  $D_2^{(k)}$ ,  $\alpha^{(k)}$ ,  $\beta^{(k)}$  - точнісні характеристики процесу навчання: перша та друга достовірності, помилки першого та другого роду, отримані на  $k$ -тому кроці навчання.

## ФОРМУВАННЯ НАВЧАЛЬНОЇ МАТРИЦІ

Для формування навчальної матриці  $\|y_{m,i}^{(j)}\|$  використовуються зображення спектру голосової команди, який синтезується на основі швидкого перетворення Фур'є [2]. Кожне зображення оброблюється в полярних координатах, що робить їх інваріантним до зсуву, повороту та масштабу.

## РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Реалізуємо програму на об'єктно-орієнтованій мові програмування Object Pascal у середовищі програмування Borland Delphi 7 з використанням аудіо-бібліотеки bass.dll для отримання та обробки аудіо-потоків отриманого з мікрофону. Бібліотека bass.dll працює має в основі швидке перетворення Фур'є (FFT). Якщо для прямого обчислення дискретного перетворення Фур'є з  $N$  точок даних потрібно  $O(N^2)$  арифметичних операцій, то FFT дозволяє обчислити такий самий результат використовуючи  $O(N \log N)$  операцій. Алгоритм FFT використовується для цифрової обробки сигналів для перетворення дискретних даних з часового у частотний діапазон.

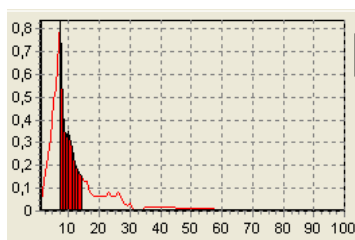
Основні принципи роботи алгоритму програми полягають в тому, щоб отримати аудіо-потік з мікрофону, побудувати для нього спектр та на базі отриманих спектрів проводиться навчання інтелектуальної системи. В результаті виконання алгоритму екзамену програма запускає на виконання файл, шлях до якого був вказаний при навчанні програми.

### ПРИКЛАД ОПТИМІЗАЦІЇ

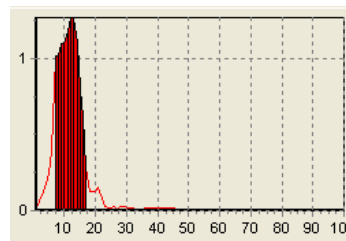
#### ФУНКЦІОНАЛЬНИХ ПАРАМЕТРІВ

#### СИСТЕМИ ГОЛОСОВОГО КЕРУВАННЯ

Для оцінки працездатності розробленої системи голосового керування комп'ютером було розв'язано задачу розпізнавання двох голосових команд. При побудові нечіткого розбиття простору зі 100 ознак розпізнавання було сформовано контейнер гіперсферичної форми для кожного з класів розпізнавання. При цьому виконувалася ітераційна процедура оптимізації їх геометричних параметрів. На рис. 1 подано динаміку зміни КФЕ при оптимізації радіусів контейнерів.



а)



б)

Рисунок 1 – Графік залежності КФЕ від радіусу контейнеру: а) першого класу, б) другого класу

Аналіз графіків показує, що КФЕ обох класів не досягає свого максимально можливого значення, що вказує на необхідність застосування більш складних і ресурсоемних алгоритмів ІЕІТ.

## ВИСНОВКИ

В роботі досліджена можливість створення прикладного забезпечення з застосуванням інтелектуальних систем для голосового керування комп'ютером і розроблено відповідне програмне забезпечення. Проведені дослідження показали, що обраний метод вирішення проблеми є більш надійним та надає більш коректний результат.

На підставі того, що системи підтримки прийняття рішень надають достатньо високий результат порівняння зображень, то реалізовано алгоритм, який отримує аудіо-потік з мікрофона, зберігає його в файл та будує для нього спектр, в якому відображається тембр голосу з файлу в який збережена команда користувача.

## ЛІТЕРАТУРА

- [1] А.С. Краснопопсовський Інформаційний синтез інтелектуальних систем керування, що навчаються // Видавництво СумДУ Суми – 2003. А.Б.Сергиенко Цифровая обработка сигналов // СПб, Питер, 2002. — 608 с.