

УДК 004.8

К.т.н., доцент Олефір О.С., магістрант Люльов М.В.

Національний технічний університет України  
«Київський політехнічний інститут»

## МОДЕЛЬ РОЗПІЗНАВАННЯ ГОЛОСОВИХ КОМАНД

### Abstract

**Aleksandr S. Olefir, assoc. prof., PhD; Maxim Lyulyov, student**

*The model of recognition of the voice commands*

*This article describes the model which finds out and distinguishes vocal commands. This model represents a sound in the digital form which will characterize the pledged word for the further recognition.*

### Вступ

Мовний інтерфейс, як більш природній для людини, набуває все більшої затребуваності у сучасних людино-машинних системах.

Діалог з комп'ютерами, роботами, автоматизованими системами управління за допомогою голосових повідомлень відкриває великі перспективи:

- простота спілкування з системою;
- доступність мовного інтерфейсу людям з порушеннями опорно-рухового та зорового апарату;
- можливість роботи користувачів в умовах перевантаженості тактильно-зорового каналу.

### Постановка задачі

Об'єкт дослідження – людський голос.

Предмет дослідження – математичні моделі аналізу та класифікації голосових команд.

Мета дослідження - розробка моделі розпізнавання голосових команд користувача персонального комп'ютера. Запропонована модель повинна представити вимовлену голосову команду в цифровій формі, яка буде характеризувати структуру звукового сигналу, а потім з найбільшою точністю класифікувати вимовлену користувачем команду до тієї або іншої групи команд.

## Модель розпізнавання голосових команд

Процес розпізнавання голосових команд можна розділити на два етапи - аналіз звукового сигналу та класифікація цієї команди.

Аналіз звукового сигналу призначений для отримання опису мовного сигналу – представлення мовного сигналу у вигляді набору значень ознак, які зберігають інформацію про зміст мовного повідомлення. Ці ознаки найчастіше представляють собою передаточну характеристику мовного тракту і характеристики джерел його збурення.

На етапі класифікації по отриманим даним голосова команда відноситься до того або іншого класу команд. Класом виступає одна команда.

Модель розпізнавання у вигляді UML діаграми представлена на Рис.1.

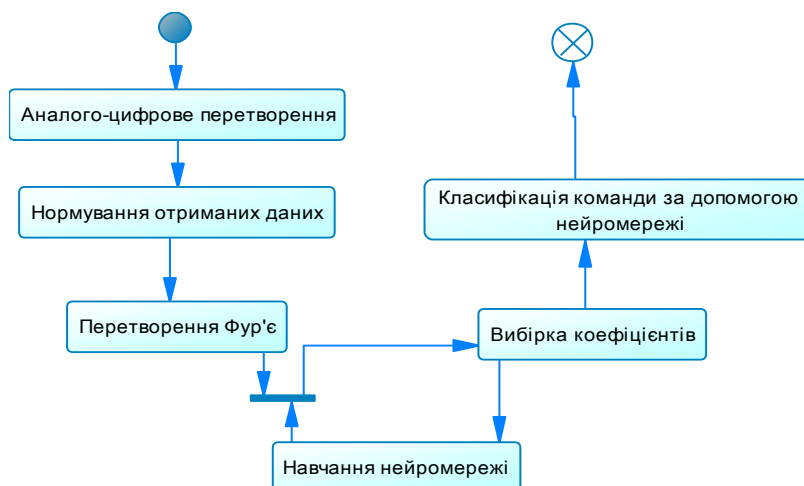


Рис. 1. Модель розпізнавання голосових команд

Аналіз звукового сигналу включає в себе аналого-цифрове перетворення, нормування отриманих даних та отримання з цих даних частотної характеристики звукового сигналу.

Сигнал подається на аналого-цифровий перетворювач, який з деякою частотою (частотною дискретизації), записує поточний рівень сигналу в цифровій формі, тобто квантує сигнал за часом і по амплітуді.

Згідно теореми Котельнікова інтервал часу між відліками  $\Delta t < \frac{1}{2 \cdot F_{\max}}$ , де

$F_{\max}$  - найбільша частота спектру сигналу [1].

Ряди Фур'є дозволяють виразити складну функцію сумою простих. Нехай звуковий сигнал це є функція  $p(t)$ . Так як вимовлені одні і ті ж команди будуть мати майже однакові частотні характеристики, то для нашої задачі найвигідніше взяти саме частотну характеристику звукового сигналу. Частотний аналіз дозволяє отримати розподіл амплітуди по частоті (амплітудо-частотні спектри) та розподіл фаз складових по частотам (фазо - частотні спектри).

При частотному аналізі звукового сигналу  $p(t)$  представляють сумою:

$$p(t) = \sum_n a_n \cos(2\pi f_n t + \varphi_n),$$

де  $a_n$  - амплітуда,  $f_n$  - частота,  $\varphi_n$  - початкова фаза. Набір чисел  $a_n$ ,  $f_n$  утворює амплітудно-частотний спектр, а  $\varphi_n$ ,  $f_n$  - фазо-частотний [2].

На Рис. 2. зображена звукова команда та її спектр.

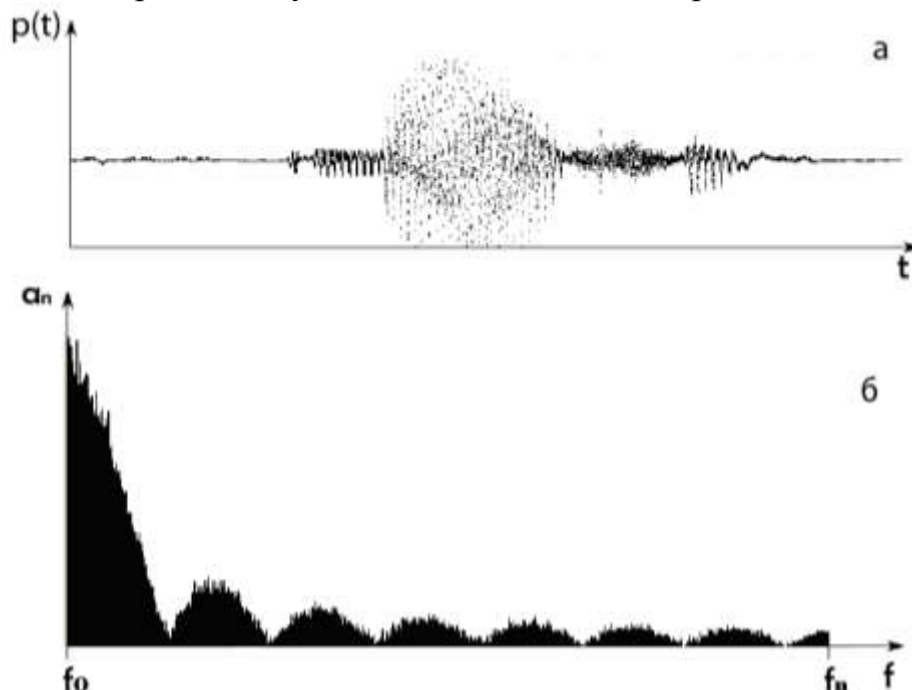


Рис. 2. Звуковий сигнал (а) та його спектр (б). На осях ординат відображені відповідно звуковий тиск  $p(t)$  та амплітуди спектральних складових  $a_n$ , по осям абсцис - час  $t$  та частота  $f$ .

Частоти  $f_0 \dots f_n$  утворюють гармонійний ряд. Для вибірки необхідних коефіцієнтів необхідно розбити звуковий сигнал на декілька частин. Як показують досліди сто інтервалів підходять для розв'язання поставленої задачі. В цих утворених інтервалах береться середні значення, які і будуть подаватися на вхід класифікатора.

Процес класифікації можна виконувати за допомогою машини опорних векторів, але нажаль, цей метод є чутливим до шумів і стандартизації даних. З задачею розпізнавання чудово справляється апарат штучних нейронних мереж.

Системи розпізнавання, що реалізуються на основі використання апарату штучних нейронних мереж, мають високу швидкодію і не потребують додаткових витрат для інтерпретації і введення початкової інформації про реальні об'єкти, мають можливості до навчання і адаптації в умовах зміни навколишнього середовища.

Для моделювання звукових образів можна використовувати нейронну мережу типу багатошаровий перцептрон [3]. Для організації взаємодії нейронної мережі з зовнішнім середовищем служать блоки фільтрації і декодування інформації про звуковий об'єкт.

Для опису обчислювальної моделі нейронної мережі використовують матриці ваг міжнейронних зв'язків. Ваги міжнейронних зв'язків представляються як матриці такого виду:

$$W^\mu = \{w_{ij}^\mu\}, \mu = \overline{1, K}, i = \overline{1, N_\mu}, j = \overline{1, N_{\mu-1}}$$

де  $\mu$  - номер матриці міжнейронних зв'язків;  $w_{ij}^\mu$  - вага зв'язку між  $j$ -им нейроном вихідного шару та  $i$ -го нейрона вхідного шару для  $\mu$ -тої матриці зв'язків;  $N_\mu$  - кількість нейронів  $\mu$ -го шару;  $N_{\mu-1}$  - кількість нейронів  $\mu-1$ -го шару.

Багат шарова нейронна мережа функціонує в такий спосіб. Вектор вхідних даних  $X = \{x_1, x_2, \dots, x_n\}$  подається на вхідний шар нейронної мережі. В нашому випадку вхідними параметрами  $x_1, \dots, x_n$  будуть отримані значення частот  $f_0 \dots f_n$ . Після цього здійснюється по чергове обчислення вихідних сигналів для нейронів кожного наступного шару з використанням уже відомих виходів нейронів попереднього шару і значень ваг міжнейронних зв'язків. У підсумку інформація, отримана з вихідного шару мережі, видається як вихід нелінійного об'єкта [3]

$$Y = F(W^1, W^2, \dots, W^K, x_1, x_2, \dots, x_n).$$

В нашому випадку на виході ми отримуємо номер класу до якого належить вимовлена команда.

## Висновок

Розглянута модель розпізнавання голосових команд призначена для створення мовного інтерфейсу, який дозволить істотно підвищити ефективність роботи людино – машинних систем. Ця модель базується на використанні частотного аналізу мовного сигналу, зокрема, перетворення Фур'є, що забезпечить високу швидкодію програмних засобів.

Класифікація мовних команд виконується на основі математичного апарату штучних нейронних мереж. Розглянута модель розпізнавання дає можливість навчати новим, необхідним для користувача, голосовим командам.

## Література

1. <http://ru.wikipedia.org/wiki/Дискретизация>
2. [http://dic.academic.ru/dic.nsf/enc\\_physics/ЗВУКА](http://dic.academic.ru/dic.nsf/enc_physics/ЗВУКА)
3. Биков М.М., Грищук Т.В. Моделювання процесу аналізу та класифікації голосових команд// Вінниця ВНТУ 2009. С.23-24