

РОЗПІЗНАВАННЯ І КЛАСИФІКАЦІЯ ЕМОЦІЙ ПО ГОЛОСУ ЛЮДИНИ

Роботу виконав:

Науковий керівник:

Рецензент:

Пилипенко Д.С.

ас. Караюз І.В.

к.т.н., доцент Безносик О.Ю.

АКТУАЛЬНІСТЬ ДОСЛІДЖЕННЯ ЗАДАЧІ РОЗПІЗНАВАННЯ ЕМОЦІЙ

- Розвиток систем розпізнавання мови та їх застосувань
- Необхідність покращення взаємодії людини та комп'ютера
- Безліч практичних задач, де використовується розпізнавання мови
- Необхідність обчислювальних потужностей досить мала, а тому задача має великий практичний інтерес
- Точність існуючих алгоритмів не є достатньо високою

ПОСТАНОВКА ЗАДАЧІ

Мета роботи:

- Аналіз та можливе покращення існуючих алгоритмів і методів розпізнавання і класифікації емоцій по звуковому сигналу

Об'єкт дослідження:

- Розпізнавання емоцій та ознаки, що використовуються для класифікації емоцій по голосу

Предмет дослідження:

- Програмні алгоритми та акустичні показники й ознаки, що застосовуються при розпізнаванні і класифікації емоційного стану людини за її мовою.

ЕМОЦІЇ



Відраза



Страх



Радість



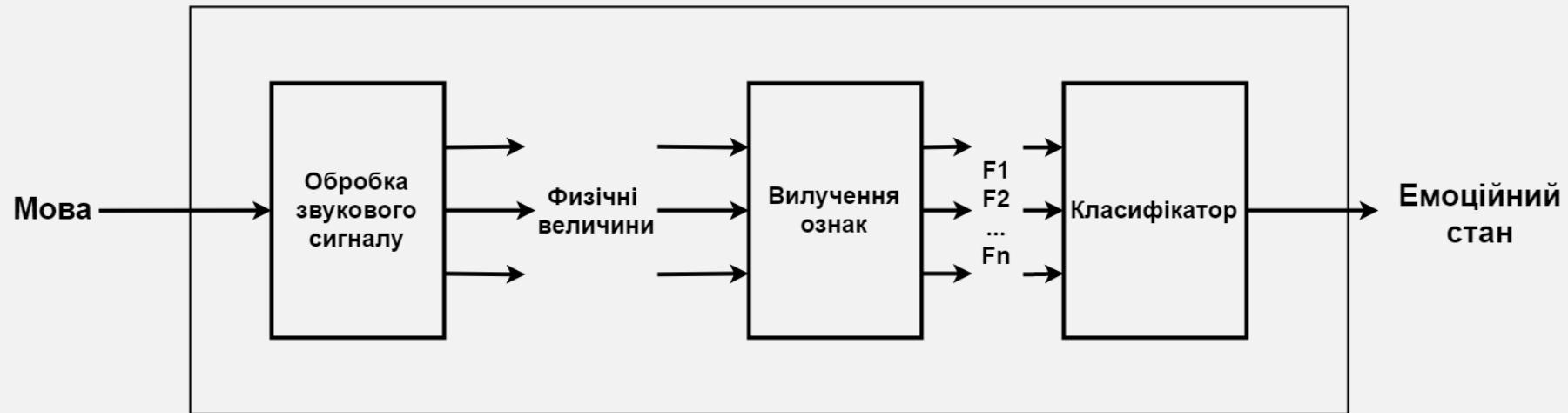
Смуток



Злість

А також, Здивування (Surprise) та Нейтральна емоція (Neutral)

АЛГОРИТМ РОЗПІЗНАВАННЯ ЕМОЦІЙ



Алгоритм розпізнавання емоцій містить чотири основні кроки:

1. Прийняття вхідного сигналу (Speech Input);
2. Вилучення ознак (Feature Extraction);
3. Класифікатор (Classifier);
4. Видача назви емоції як результату (Emotion Output).

ОЗНАКИ, ЩО ВИКОРИСТОВУЮТЬСЯ ДЛЯ РОЗПІЗНАВАННЯ ЕМОЦІЙ

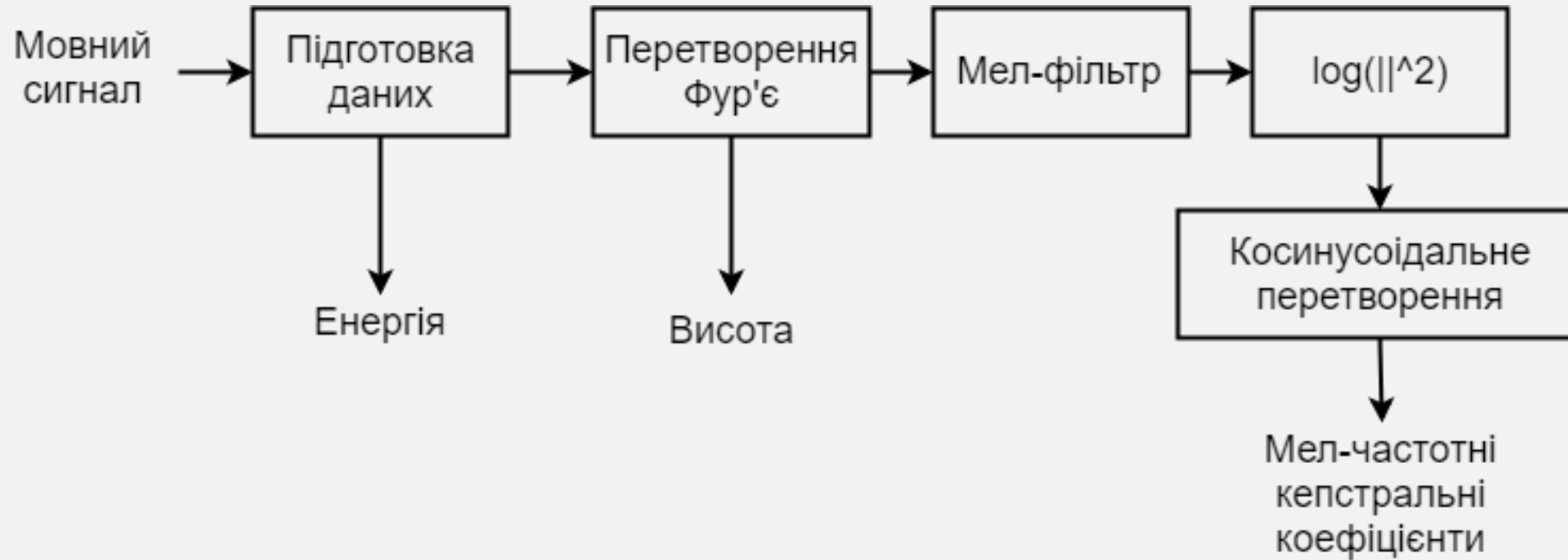
- **Основні:**
 - Енергія (Energy)
 - Висота (Pitch)
 - Мел-частотні кепстральні коефіцієнти (MFCC)
- **Інші:**
 - Лінійні прогностовні кепстральні коефіцієнти (LPCC)
 - Мел-енергетичні динамічні коефіцієнти спектру (MEDC)
 - Логарифмічно-частотні потужні коефіцієнти (LFPC)

ОГЛЯД ІСНУЮЧИХ МОДЕЛЕЙ КЛАСИФІКАЦІЇ

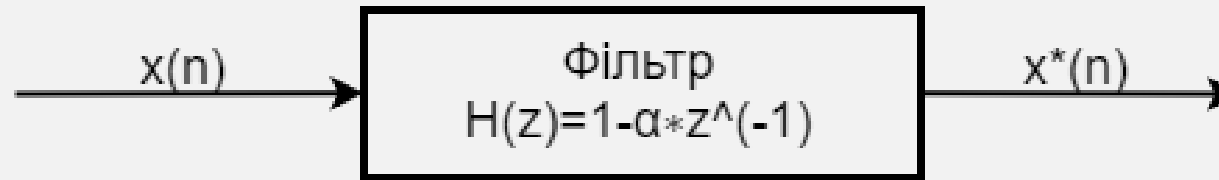
Серед існуючих класифікаторів емоцій можна виділити наступні:

- Прихована модель Маркова (Hidden Markov Model)
- Машина опорних векторів (Support Vector Machine)
- Багатошаровий перцептрон (Multilayer Perceptron)
- **Штучна нейронна мережа (Artificial Neural Network)**
- Обмежена машина Больтцмана (Restricted Boltzman Machine)
- Гаусівська модель суміші (Gauss Mixture Model)

МАТЕМАТИЧНІ ОСНОВИ



МАТЕМАТИЧНІ ОСНОВИ: ПІДГОТОВКА ДАНИХ



α – параметр фільтру. Як правило, береться 0.95.

Фреймування:

$$x_t(n), t = 1, \dots, T,$$

де T - кількість фреймів

Згладжування:

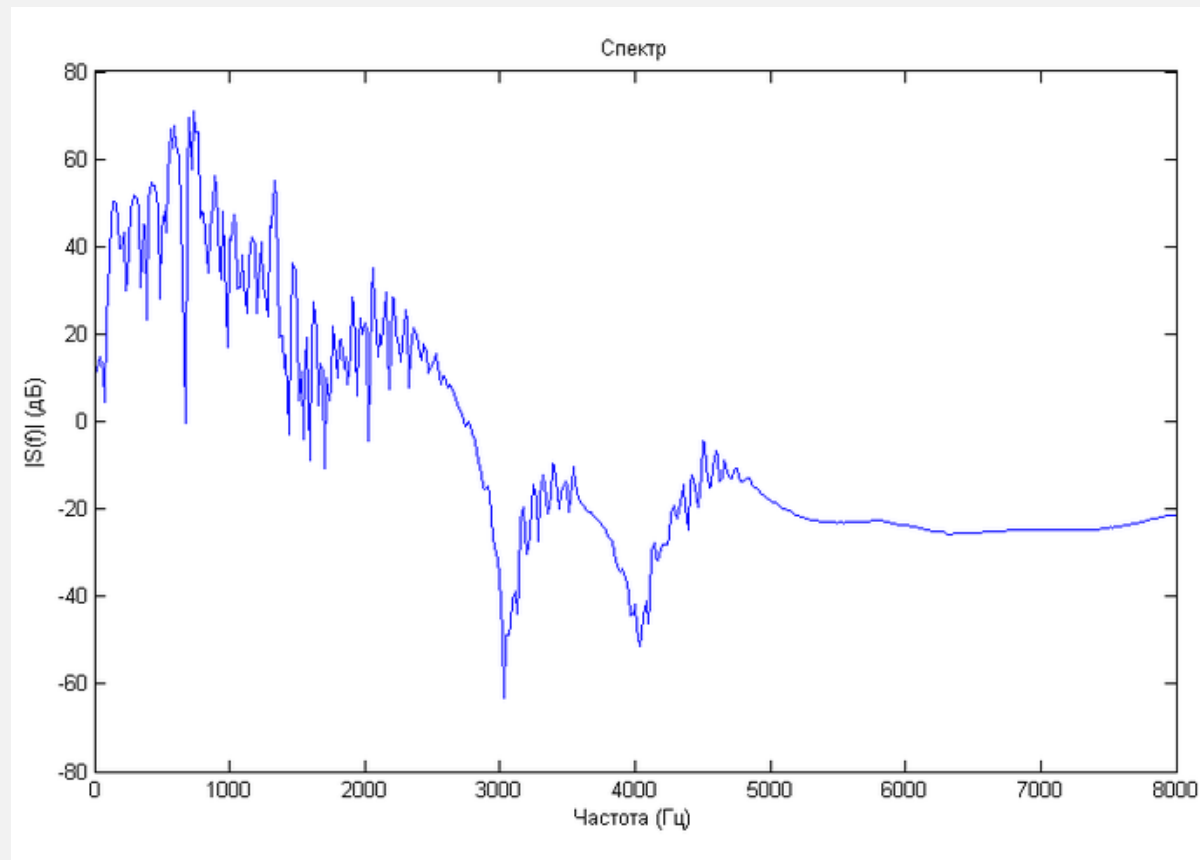
$$x_t(n) \equiv x_t(n) * w(n), n = 0, \dots, N - 1,$$

де $w(n) = 0.53836 - 0.46164 * \cos\left(\frac{2\pi n}{N-1}\right)$ – вікно Хеммінга

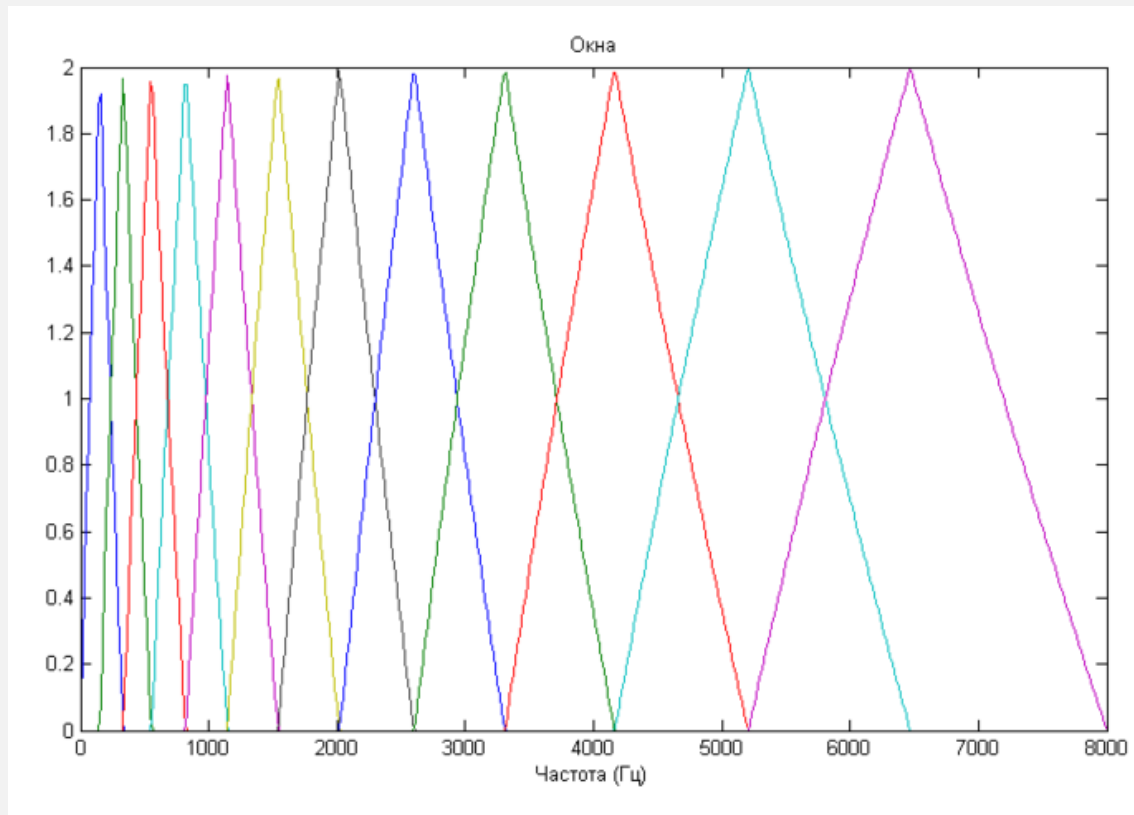
МАТЕМАТИЧНІ ОСНОВИ: ПРОЦЕС ВИЛУЧЕННЯ ОЗНАК

Дискретне перетворення Фур'є:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-2\pi i k \frac{n}{N}}, k = 0, \dots, N - 1$$



МАТЕМАТИЧНІ ОСНОВИ: ПРОЦЕС ВИЛУЧЕННЯ ОЗНАК



Вікна:

$$Y_1(m) = \sum_{k=b_m-\Delta_m}^{b_m+\Delta_m} X_1(k) * U_{\Delta_m}(k + b_m),$$

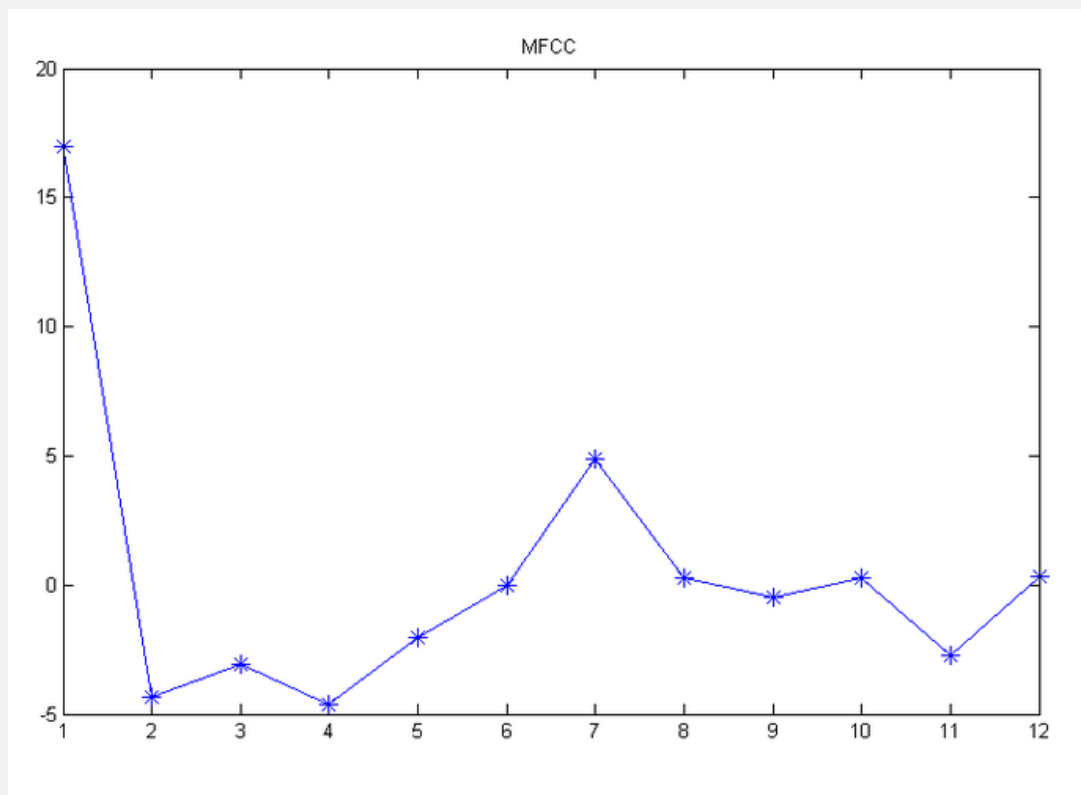
$$U_{\Delta_m}(k) = \begin{cases} 1 - |k|/\Delta_m, & \text{якщо } |k| < \Delta_m \\ 0, & \text{якщо } |k| \geq \Delta_m \end{cases}$$

де $b_m = b_{m-1} + \Delta_m$ - «центральна» частота,

$2\Delta_m$ - довжина m -го фільтру

Використовується апроксимація: $\Delta_m = 1,2 * \Delta_{m-1}$

МАТЕМАТИЧНІ ОСНОВИ: ПРОЦЕС ВИЛУЧЕННЯ ОЗНАК

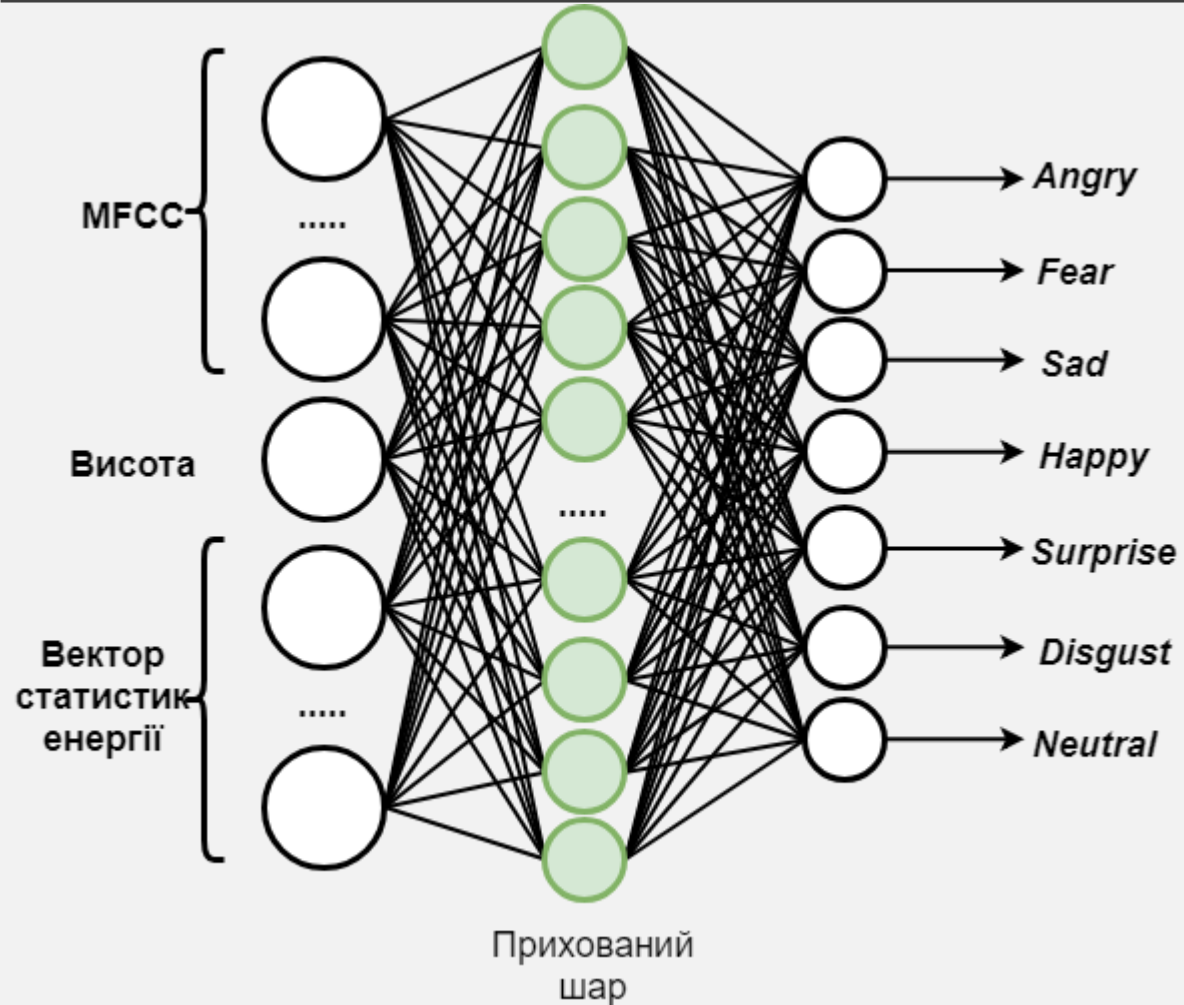


Мел-частотні кепстральні коефіцієнти:

$$y_t^{(m)}(k) = \sum_{m=1}^M \log\{|Y_t(m)|\} * \cos\left(\frac{k\left(m - \frac{1}{2}\right)\pi}{M}\right), k = 0, \dots, L$$

де L - кількість коефіцієнтів

НЕЙРОННА МЕРЕЖА ПРЯМОГО ПОШИРЕННЯ



Функція активації:

$$OUT = \frac{1}{1 + e^{-\sum_{i=0}^n w_i * x_i}}$$

РЕЗУЛЬТАТИ НАВЧАННЯ МОДЕЛІ



РЕЗУЛЬТАТИ НА ТЕСТОВІЙ ВИБІРЦІ

Тестові записи

Прогнозовані емоції	Емоція	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
	Angry	0.816	0.005	0.150	0.055	0.121	0.023	0.222
	Disgust	0.000	0.678	0.000	0.001	0.001	0.021	0.001
	Fear	0.005	0.011	0.061	0.000	0.005	0.036	0.000
	Happy	0.113	0.000	0.243	0.072	0.104	0.001	0.105
	Neutral	0.027	0.056	0.189	0.641	0.112	0.243	0.110
	Sad	0.001	0.078	0.023	0.091	0.025	0.676	0.031
	Surprise	0.038	0,172	0.334	0.090	0.542	0.000	0.541

ПОРІВНЯННЯ РЕЗУЛЬТАТІВ РІЗНИХ МЕТОДІВ РОЗПІЗНАВАННЯ ЕМОЦІЙ

	Прихована модель Маркова (MFCC)	Машина опорних векторів (MFCC)	Багатошаровий перцептрон (MFCC)	Гаусівська модель суміші (MFCC, енергія)	Штучна нейронна мережа (MFCC, енергія)	Власна реалізація (MFCC, енергія, висота)
Leave One Speaker Out	61.77 %	63.52 %	51.65 %	64.11 %	51,19 %	64.32 %
Leave One Text Out	64.12 %	65.24 %	63.10 %	67.54 %	52.87 %	69.14 %

ВИСНОВКИ

- Реалізовано нейронну мережу для розпізнавання і класифікації емоцій по голосу людини, з використанням ефективно підібраних ознак (мел-частотні кепстральні коефіцієнти, енергія та висота)
- Наочно продемонстровано вищу точність розпізнавання емоцій, ніж у стандартних варіантах реалізації алгоритму класифікації
- Проведено порівняльний аналіз сучасних методів рішення задачі розпізнавання і класифікації емоцій по голосу людини

ШЛЯХИ ПОДАЛЬШОГО РОЗВИТКУ

- Використання різних нейронних мереж, наприклад, конвуляційної НМ, обмеженої машини Больцмана та наївного Баєсового класифікатора.
- Включення інших акустичних ознак (Лінійні прогнозовані кепстральні коефіцієнти (*LPCC*), Мел-енергетичні динамічні коефіцієнти спектру (*MEDC*), Логарифмічно-частотні потужні коефіцієнти (*LFPC*)) та розпізнавання гендеру людини, що, теоритично, має підвищити точність розпізнавання і класифікації емоцій по голосу.

Дякую за увагу!