

# Магістерська кваліфікаційна робота

На тему: Інформаційна технологія для аудіодетекції  
звукових сигналів немовленнєвого походження

Виконав: Тютюнник Я. О.

Керівник: к.т.н., доц. кафедри ОТ Ткаченко О. М.

м. Вінниця -2020 рік

# Актуальність теми

Тема ідентифікації звукових сигналів є популярною на сьогоднішній день. Створюються застосунки за допомогою яких ми можемо розпізнавати музичні композиції, існують додатки які в змозі розпізнавати мову людини та на її основі керувати електронними пристроями. Також існують системи які в змозі виявляти постріли в містах та сповіщати поліцію про небезпечні ситуації. Усе це робить наше життя легшим, зручнішим та безпечнішим. Проте в них усіх є свої недоліки, що проявляються в вузькій спеціалізації, закритості програмних продуктів та високій ціні. Таким чином можна впевнено сказати що тема магістерської кваліфікаційної роботи є актуальною.

**Метою дослідження** є підвищення ймовірності правильного розпізнавання звуків немовленнєвого походження за рахунок розроблення нової інформаційної технології.

Для досягнення мети необхідно розв'язати такі **завдання:**

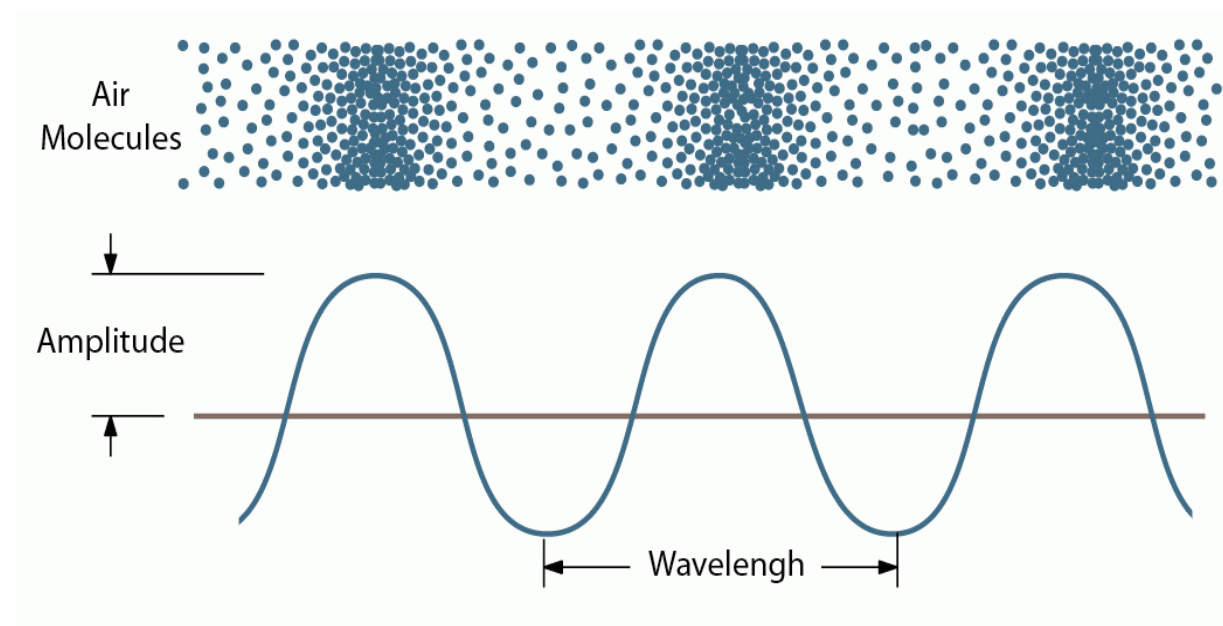
- ▶ провести аналіз існуючих на ринку систем розпізнавання звуку;
- ▶ дослідити існуючі моделі класифікації та визначити підходящої для розпізнавання звукового сигналу;
- ▶ дослідити сучасні інструменти які дозволили б реалізувати на практиці обрану модель класифікації;
- ▶ розробити програмне забезпечення;
- ▶ провести тестування розробленого програмного забезпечення.

- ▶ **Об'єкт дослідження** - є процес оброблення звукових сигналів у комп'ютерних системах.
- ▶ **Предмет дослідження** - методи та програмні засоби розпізнавання звукових сигналів немовленнєвого походження.

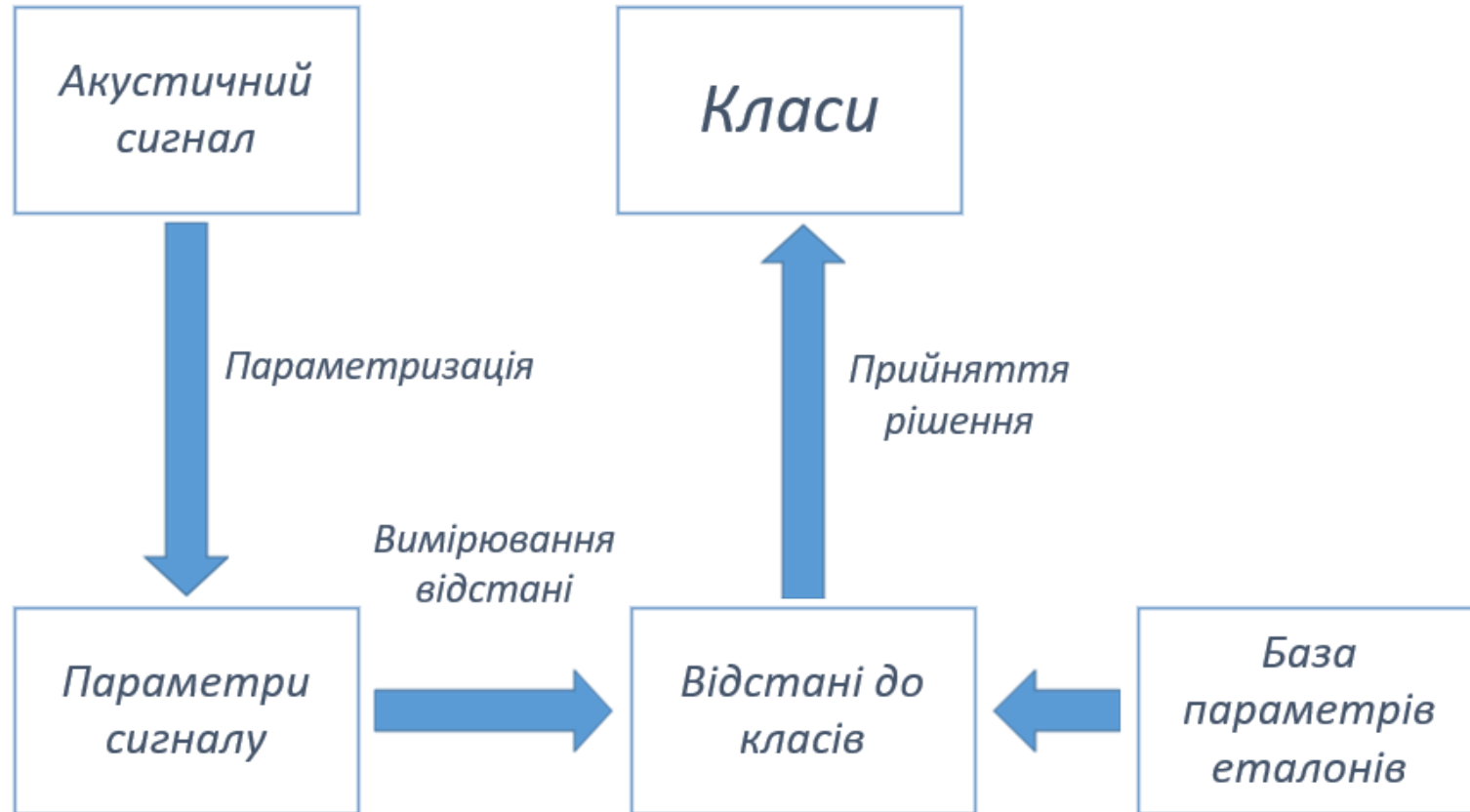
- ▶ **Наукова новизна отриманих результатів** магістерської роботи полягає в тому, що удосконалено метод розпізнавання звуків на основі змішаних Гаусівських моделей, в якому, на відміну від існуючого, на стадії прийняття рішення, враховується інформація про тривалість аудіо сигналу, що дало можливість підвищити ймовірність правильного розпізнавання звуків немовленнєвого походження.
- ▶ **Практична цінність** розроблених програмних засобів має широкий спектр застосування. Спеціальні застосунки виявлення та розпізнавання аудіо сигналів можуть бути використані у системах спостереження та безпеки, сприяючи спрацюванню сигналізації або виконанню перевірки.

# Звук

- ▶ Звук - це особливий вид механічних коливань які поширюються в пружних середовищах і тілах, та здатний викликати слухові відчуття.
- ▶ Людина своїми слуховими органами може сприймати звук частота коливань яких лежать в межах від 20 Гц до 20 кГц



# Схема аудіодетекції звукового сигналу



# Обрахунок кепстральних коефіцієнтів

Одним з варіантів отримання кепстральних коефіцієнтів є їх обчислення із коефіцієнтів LPC. Для цього використовують такі перетворення

$$c_1 = -a_1$$
$$c_n = \begin{cases} -a_n - \sum_{m=1}^{n-1} \left(1 - \frac{m}{n}\right) \cdot a_m \cdot c_{n-m}, & (1 < n \leq M); \\ -\sum_{m=1}^M \left(1 - \frac{m}{n}\right) \cdot a_m \cdot c_{n-m}, & (M < n). \end{cases}$$

Де  $A = \{a_1, a_2, \dots, a_M\}$  - вектор LPC,

$C = \{c_1, c_2, \dots, c_M\}$  - вектор шуканих кепстральних коефіцієнтів



# Модель Гаусової суміші

- ▶ Модель Гаусової суміші (Gaussian Mixture Models) є імовірнісною моделлю, яка базується на припущенні, що всі спостереження, що належать до певних класів, утворюються з суміші скінченної кількості гаусових розподілів з невідомими параметрами.
- ▶ В цілому GMM модель цікава тим, що розглядає дані як результат лінійної комбінації певної кількості гаусових розподілів, в той час, як більшість інших методів статистичної класифікації використовують чи лише один певний розподіл, чи взагалі розглядають лише взаємовідносини між даними.
- ▶ З урахуванням того, що саме нормальний розподіл є дуже розповсюдженим у реальному світі, вибір саме цього розподілу є досить очевидним при побудові моделі.

## Модель Гаусової суміші

Дана модель являє собою зважену суму Гауссіан.

$$p(x \parallel \lambda) = \sum_{i=1}^M w_i p_i(x),$$

де  $\lambda$  – модель певного класу;

$M$  – кількість компонентів моделі;

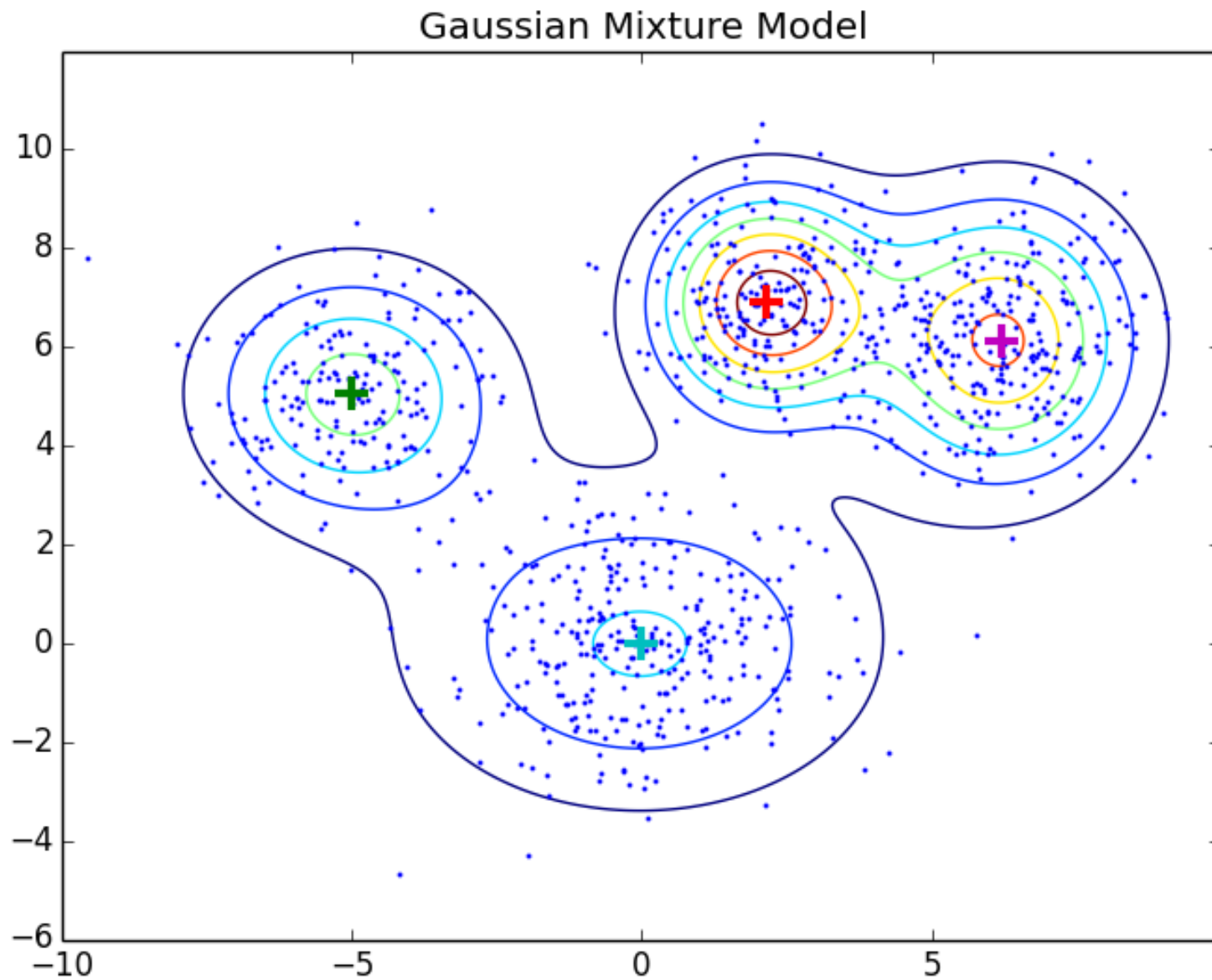
$w_i$  – ваги компонентів.

$$p_i(x) = \frac{1}{2\pi^{D/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right).$$

де  $D$  – розмірність простору ознак;

$\mu_i$  – вектор математичного очікування;

$\Sigma$  – матриця коваріації.



Стадія вимірювання відстані

## відстань Махалобіса

$$d(x, y) = (x - y)^T * W^{-1} * (x - y),$$

де  $d(x, y)$  - відстань до шаблонного класу;

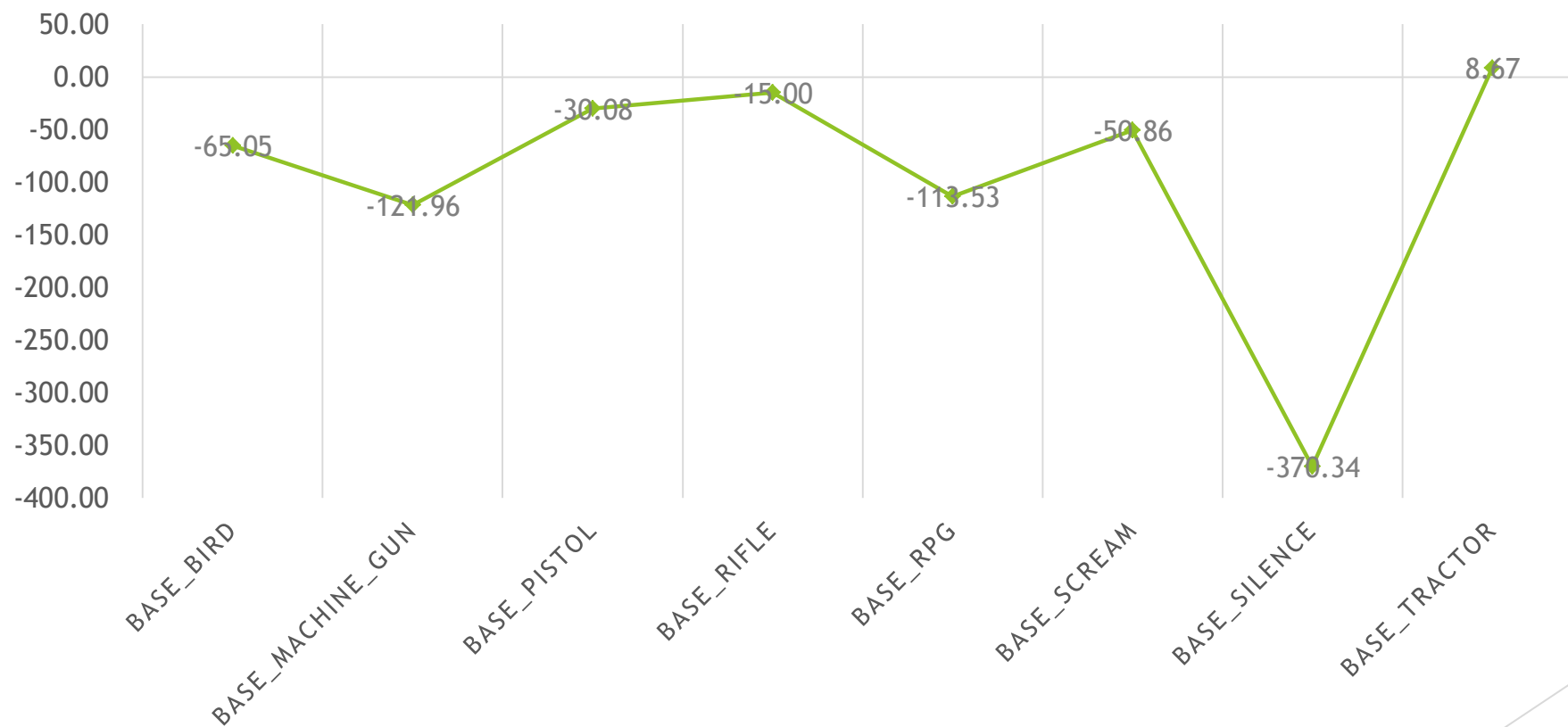
$x$  - вектор, який проходить перевірку;

$y$  - центроїд гаусіана;

$W^{-1}$  - матриця коваріації.

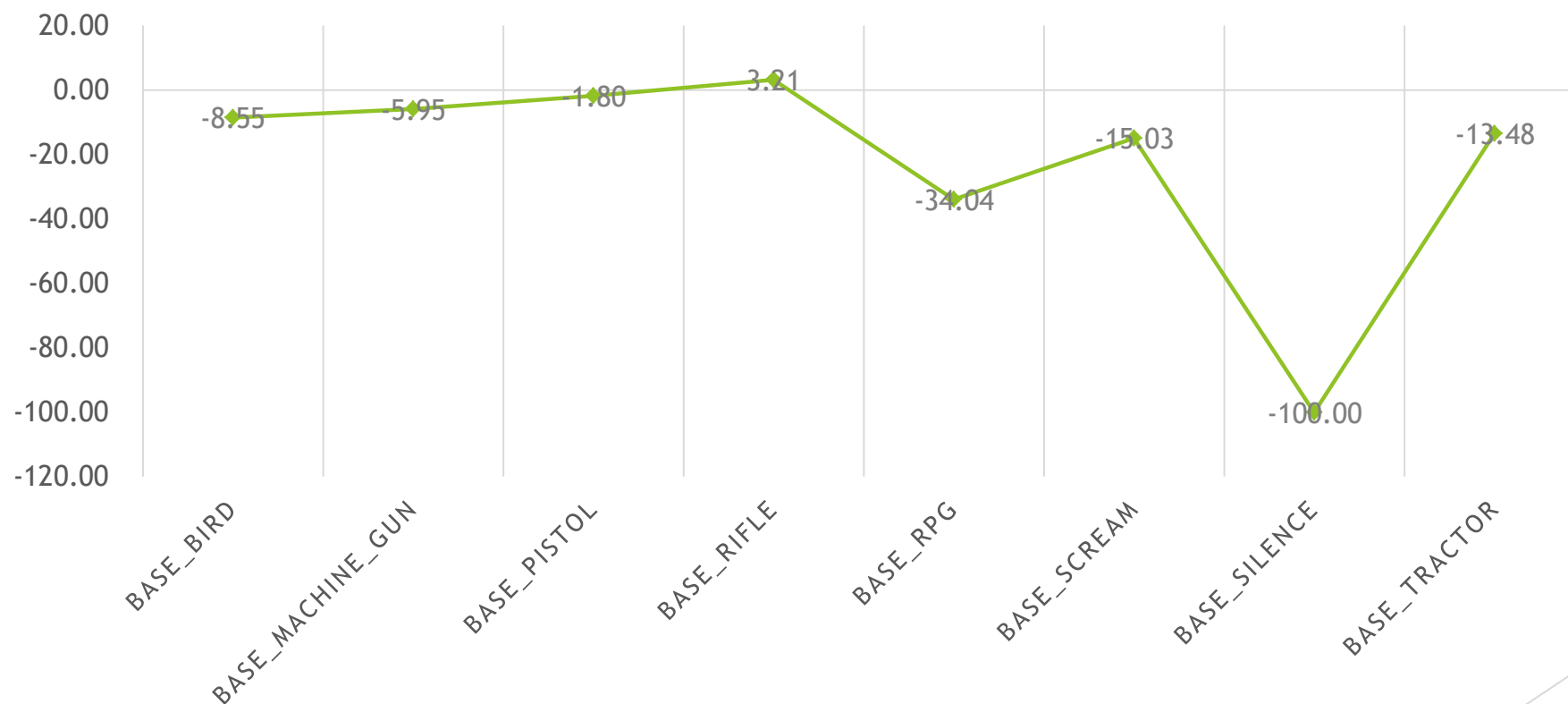
Відстані до класів

# Приклад відстані фрейму звуку трактора до наявних класів



Відстані до класів

# Приклад відстані фрейму звуку пістолета до наявних класів



## Класифікація

Це один з розділів машинного навчання, присвячений розв'язанню наступної задачі:

Задана певна обмежена множина об'єктів, для яких відомо, до якого класу вони відносяться. Ця множина називається навчальною вибіркою. Класова приналежність інших об'єктів не відома. Вимагається побудувати алгоритм, здатний класифікувати будь-який об'єкт початкової множини.

# Алгоритм прийняття рішення



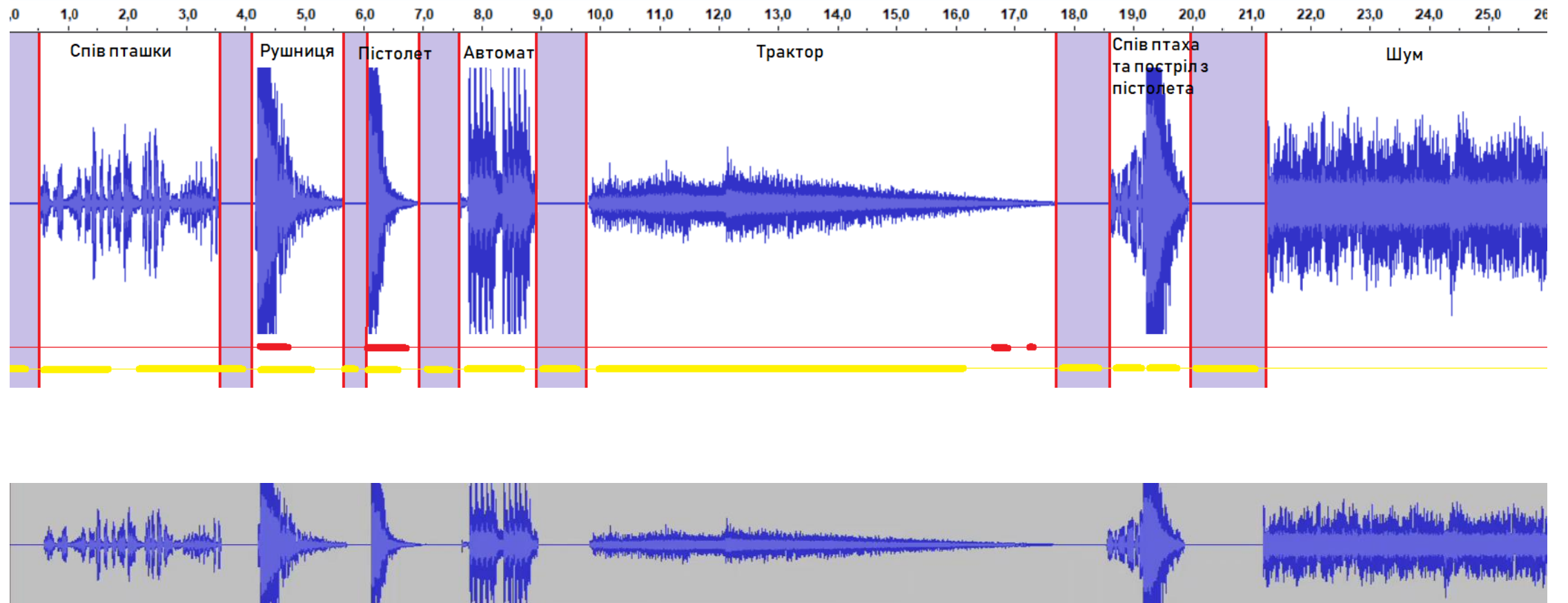


# Результати експериментальних досліджень

Для тестування роботи системи була створена база звуків на яких проводилось навчання моделі. Для кожного шаблону було взято по 3 різних зразки.

- ▶ спів пташки;
- ▶ автомат;
- ▶ рушниця;
- ▶ пістолет;
- ▶ ручний протитанковий гранатомет (RPG);
- ▶ крик людини;
- ▶ звук роботи трактора;
- ▶ тиша.

# Результати експериментальних досліджень



## Результати сегментації по кожному фрагменту звуку

Середній результат вірної сегментації по усіх звуках складає 65%

№	Назва звуку	Усього фреймів	Розпізнано	Розпізнано неправильно	Не розпізнано	Відсоток розпізнавання
1	Тиша	52	30	0	22	58%
2	Спів пташки	310	195	0	115	63%
3	Тиша	53	34	0	19	64%
4	Рушниця	157	83	25	74	53%
5	Тиша	38	19	0	19	50%
6	Пістолет	90	40	58	50	44%
7	Тиша	70	45	0	25	64%
8	Автомат	120	92	0	28	77%
9	Тиша	90	55	0	35	61%
10	Трактор	792	639	41	153	81%
11	Тиша	84	62	0	22	74%
12	Спів пташки	62	59	0	3	95%
13	Пістолет	82	45	0	37	55%
14	Тиша	131	101	0	30	77%
15	Шум	400	0	0	400	0%

# Висновки

- ▶ Було проведено аналіз існуючих на ринку систем розпізнавання звукових сигналів та моделей класифікації. На основі цього аналізу було обрано гаусівські змішані моделі, які найбільше підходять для розв'язання поставленої задачі.
- ▶ Було розроблено програмний продукт мовою C++ який дає змогу проводити розпізнавання звукових сигналів наявних у базі у реальному масштабі часу.
- ▶ Також проведено тестування розробленого продукту. Для тесту були взяті окремі звуки, звуковий файл із послідовністю різних звуків та послідовність різних звуків із шумовим фоном. Усі ці звуки були розпізнані системою. При цьому імовірність правильної сегментації звуків склала 65%.
- ▶ Був зроблений економічний аналіз розробки, в результаті якого було підтверджено перспективність подальшої розробки.

# Публікації

За результатами проведених досліджень було написано три публікації та подано статтю в науковий журнал «Інформаційні технології та комп'ютерна інженерія»:

- ▶ Тютюнник Я. О. Інформаційна технологія для аудіодетекції звукових сигналів немовленнєвого походження. Тези: 2020 рік : матеріали Всеукр. наук.-практ. конф. Молодь в науці 2020 р. Вінниця : ВНТУ, 2020. С. 1-2.
- ▶ Тютюнник Я. О., Чорний Д. С., Ткаченко О. М. Програмні засоби для аудіодетекції звукових сигналів немовленнєвого походження. Тези: 2019 рік : матеріали Всеукр. наук.-практ. конф. Молодь в науці, 11-30 травня. 2019 р. Вінниця : ВНТУ, 2019. С. 1-2.
- ▶ Чорний Д. С., Тютюнник Я. О., Ткаченко О. М. Програмні засоби формування словника параметрів звукових сигналів для аудіодетекції. Тези: 2019 рік : матеріали Всеукр. наук.-практ. конф. Молодь в науці, 11-30 травня. 2019 р. Вінниця : ВНТУ, 2019. С. 1-2.

Дякую за увагу!