

Вінницький національний технічний університет
Факультет інформаційних технологій та комп'ютерної інженерії
Кафедра обчислювальної техніки

ПРОГРАМНИЙ ЗАСІБ РОЗПІЗНАВАННЯ МУЗИЧНИХ ФРАГМЕНТІВ

Виконав: студент групи 1КІ-17М
Баліцький В. В.
Керівник роботи:
к. т. н., доц. Цирульник С. М.

АКТУАЛЬНІСТЬ

Проблема розпізнавання та розподілу мовлення від музики стає дедалі важливішою в системах автоматичного розпізнавання мовлення (ASR – Automatic Speech Recognition), які все більш широко застосовуються в реальних сферах роботи з мультимедіа. Наприклад, якщо ми хочемо побудувати системи, що виконують ASR даних музичного файлу, важливо вміти розрізняти, які сегменти звукової доріжки міститимуть мовлення. Людина з легкістю може відокремити мовлення від музики без належного впливу змішаної музики. Завдяки новим технікам аналізу та синтезу мовленнєвих сигналів обробка музичного сигналу набрала особливого значення, і тому класичні методи аналізу звуку використовуються при обробці музичних сигналів.

МЕТА І ЗАВДАННЯ ДОСЛІДЖЕННЯ

Розширення функціональних можливостей існуючих систем розпізнавання музики та мовлення за рахунок розробки засобів попередньої сегментації звукових файлів на музичні та мовленнєві фрагменти.

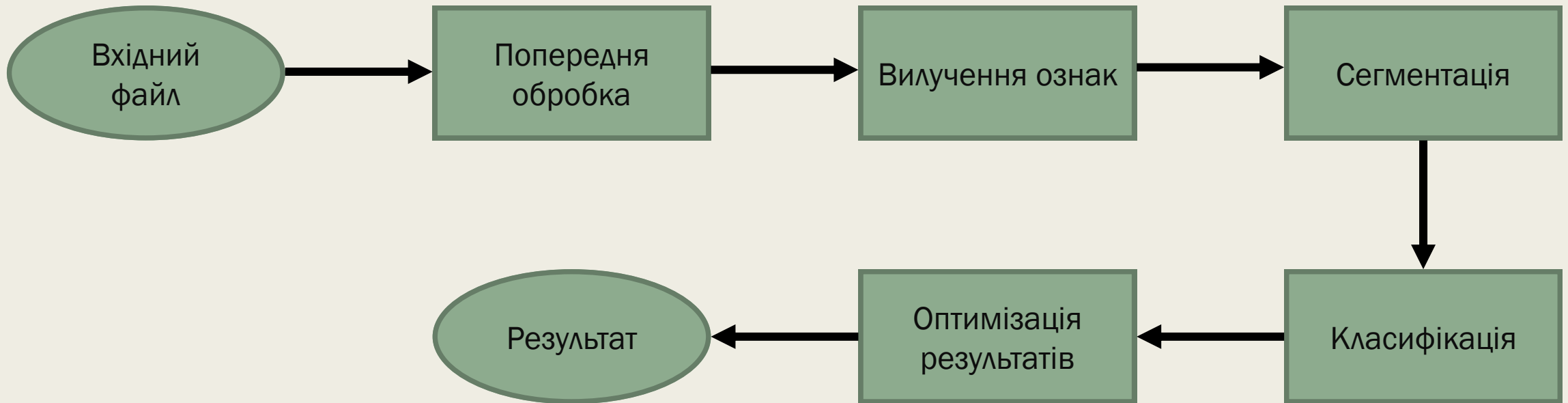
ВІДПОВІДНО ДО ПОСТАВЛЕНОЇ ВИРІШУЮТЬСЯ ТАКІ ЗАДАЧІ

- проаналізувати музичні сигнали та сигнали мовлення;
- розробити метод та алгоритм дихотомічної класифікації;
- розробити програмне забезпечення для реалізації запропонованого алгоритму;
- провести тестування розробленого продукту;
- провести розрахунок економічної ефективності розробки.

- **Об'єктом дослідження** є процес обробки звукових файлів в існуючих системах розпізнавання музики та мовлення.
- **Предмет дослідження:** методи та засоби автоматичної сегментації звукових файлів.
- **Методи дослідження:**
 - методи цифрових обробки сигналів для отримання математичних параметрів із звукових файлів;
 - математичної статистики для обробки результатів експериментальних досліджень;
 - теорії штучного інтелекту для підвищення достовірності прийняття рішення.

- **Наукова новизна:** отримав подальшого розвитку метод сегментації звукових файлів, в якому на відміну від інших, запропоновано класифікувати використовуючи дихотомічну класифікацію фрагментів, що дозволило отримати результати наближені до сегментації, що проводиться людиною оператором.
- **Практичне значення отриманих результатів:**
 - модифіковано метод сегментації для заощадження витрат на обчислення та для мінімізації часу на зчитування звукових файлів.
 - розроблено програмний засіб для автоматичної сегментації та класифікації звукових файлів.

Огляд загального алгоритму обробки аудіо-файлу.



Для реалізації алгоритму та розробки програмного засобу було використано модель TensorFlow VGGish як інструмент для вилучення ознак з аудіо-файлів.

VGGish залежить від таких пакетів Python:

- numpy;
- scipy;
- resampy;
- tensorflow;
- six;
- pysoundfile;

Модель декодує файл у моно-формат з частотою 16KHz.

Процес навчання моделі зайняв 1 годину 45 хвилин та 53 секунди.

Щоб вказати (набір) ознаки, які потрібно визначати, необхідно встановити прапорці

```
--feature_names;  
--feature_sizes.
```

Аудіо-знаки називаються "audio" і мають розмірність 128. Два прапорці приймають розділений комами список значень у рядку. Наприклад, для використання функцій аудіовізуального відео-рівня прапори повинні бути встановлені таким чином:

```
--feature_names = "rgb, audio" --feature_sizes = "1024128"
```

Загалалом було протестовано 29 різних ознак.
Кінцевий набір ознак для роботи пз складається з 5.

	MLER	SC SDD	ZCR SD	RMS SD	PC
MLER	-				
SC SDD	0.93	-			
ZCR SD	0.97	0.97	-		
RMS SD	0.97	0.93	0.97	-	
PC	0.87	0.77	0.77	0.83	-

Модифікована версія методу сегментації полягає в тому, що він використовує ту ж базову ознаку, амплітуду RMS, що і класифікація. Це означає, що подальшого вилучення ознак не потрібно, це заощаджує час обчислення та мінімізує читання звукових файлів. Амплітуда RMS використовується іншим способом, оскільки ознака MLER, використовувана для класифікації, вимагає довших кадрів аналізу, а середнє значення та дисперсія RMS змінюються так само швидко, як і звукові класи.

РЕЗУЛЬТАТИ РОБОТИ АЛГОРИТМУ ПО КОЖНОМУ З КЛАСІВ

	Мовлення	Музика
Мовлення	95.4%	4.6%
Мовлення	1.9%	98.1%

ПІДСУМКОВІ ДАНІ РОБОТИ АЛГОРИТМУ

	Коректність	Похибка
Загалом	97.3%	2.7%

ВИСНОВКИ

- На основі проведеного аналізу визначено, що для сегментації звукових файлів доцільно використовувати такі параметри, як швидкість перетину нуля ZCR, RMS, SC і MLER.
- Модифіковано метод сегментації та за результатами роботи алгоритму дихотомічної класифікації отримали точність розпізнавання музичних фрагментів та фрагментів мовлення понад 97% в тестах, виконаних на основі матеріалу знайдених радіо-передач.
- Розроблено програмний засіб для реалізації запропонованого алгоритму. Проведено тестування продукту на ОС Windows 10.
- Проведений розрахунок економічної ефективності показав, що вона складає 103912,31 грн. Термін окупності розробки складає 0,45 року.
- В результаті виконання магістерської кваліфікаційної роботи розроблено програмний засіб для сегментації звукових файлів, який може використовуватися в системах розпізнавання музики та мовлення

ДЯКУЮ ЗА УВАГУ