

Ćwiczenie 2

Celem zadania jest stworzenie programu umożliwiającego przetwarzanie dźwięku na tekst, z uwzględnieniem wykrywania szumów, ich redukcji oraz porównania różnych metod transkrypcji. Interfejs graficzny nie jest wymagany.

1. Przetwarzanie pliku audio:

- Program powinien umożliwiać wczytanie pliku dźwiękowego (w formacie WAV lub MP3), wykrywać poziom zakłóceń i stosować filtrację.
- Po przetworzeniu dźwięku program powinien zwrócić rozpoznany tekst.
- Powinna być możliwość ręcznej regulacji parametrów filtru (np. częstotliwości odcięcia, typu filtru).
- Należy wykonać testy na dwóch plikach: jednym bez zakłóceń oraz drugim z zakłóceniami.
- Plik audio zaleca się stworzyć samodzielnie, powinien trwać 10-15 sekund. Można wykorzystać np. fragment wiersza (np. *Inwokację z Pana Tadeusza*).

2. Przetwarzanie dźwięku z mikrofonu w czasie rzeczywistym:

- Program powinien umożliwiać przechwytywanie dźwięku z mikrofonu systemowego.
- Należy zaimplementować wykrywanie poziomu szumu i zastosowanie filtracji przed przekazaniem dźwięku do modelu transkrypcji.
- Powinna być możliwość rozpoczęcia i zatrzymania nagrywania.

3. Porównanie dwóch metod transkrypcji:

- Pierwsza metoda powinna wykorzystywać model wav2vec2 z biblioteki transformers.
- Druga metoda powinna wykorzystywać inny model AI do przetwarzania mowy na tekst, np. Whisper od OpenAI lub facebook/mms-1b-all dostępny na Hugging Face.
- Jeśli któryś z modeli nie obsługuje języka polskiego, test należy przeprowadzić w języku angielskim.
- Wyniki transkrypcji powinny zostać porównane pod względem dokładności i odporności na zakłócenia.
- Transkrypcję należy porównać z samodzielnie przestuchanym plikiem i wykonanym transkryptem, aby ocenić dokładność rozpoznania.
- Porównanie metod transkrypcji należy przeprowadzić zarówno dla pliku audio, jak i dla dźwięku nagrywanego na żywo z mikrofonu.

Wskazówki techniczne:

- Można wykorzystać bibliotekę pyaudio do przechwytywania dźwięku z mikrofonu.
- Do obróbki dźwięku i filtracji można użyć librosa, scipy.signal lub torchaudio.
- Do transkrypcji można użyć modeli wav2vec2 oraz Whisper lub facebook/mms-1b-all.
- Wyniki należy zapisać i porównać (np. procent poprawnych słów w transkrypcji).

Minimalne wymagania zaliczeniowe:

- Program poprawnie przetwarza pliki audio i mikrofon.
- Wykrywa i raportuje poziom szumów.
- Oferuje podstawowe opcje filtracji i ich regulacji.
- Przetwarza mowę na tekst za pomocą dwóch różnych modeli AI.
- Wyniki transkrypcji są porównane w krótkim raporcie podsumowującym jakość rozpoznania.