



**Politechnika  
Śląska**

**Wydział Automatyki, Elektroniki  
i Informatyki**

Systemy Interaktywne i Multimedialne  
Projekt  
Detekcja emocji w głosie

Natalia Stręk Jakub Kula, Paweł Wójtowicz,

Gliwice 2023

# 1 Przedstawić najważniejsze scenariusze użycia stworzonego oprogramowania.

Wsparcie w obsłudze klienta: Program może być zintegrowany z systemami call center, gdzie automatycznie analizuje emocje klientów podczas rozmów. Na podstawie wykrytych emocji, takich jak frustracja czy zadowolenie, system może przekazywać zgłoszenia do odpowiednich działów lub oferować automatyczne rozwiązania, poprawiając jakość obsługi klienta

Zarządzanie zasobami ludzkimi: W trakcie rekrutacji i rozmów kwalifikacyjnych program może analizować emocje kandydatów, co pomoże rekruterom w ocenie autentyczności i stresu kandydatów. Może także być używany do monitorowania stanu emocjonalnego pracowników podczas sesji feedbackowych, co pozwoli na szybszą reakcję na potencjalne problemy.

Interaktywne systemy rozrywki: Program może być używany w grach komputerowych i aplikacjach VR/AR, gdzie analizuje emocje graczy, aby dostosować poziom trudności gry, fabułę lub interakcje z postaciami w czasie rzeczywistym. Dzięki temu doświadczenie z gry staje się bardziej immersyjne i spersonalizowane.

## 2 Realizacja projektu

Projekt został zrealizowany zgodnie z założeniami, a wszystkie planowane etapy w projekcie zostały ukończone. Poniżej przedstawiono opis zrealizowanych etapów:

- Zgromadzenie i przetworzenie danych - dane w postaci krótkich nagrań dźwiękowych reprezentujące różne emocje,
- Wybór oraz ekstrakcja cech,
- Wybór architektury sieci, hiperparametrów, podział danych na zbiór uczący oraz treningowy,
- Trenowanie modelu i optymalizacja parametrów modelu.
- Ocena i walidacja modelu,
- Dostosowywanie parametrów modelu w celu poprawy jego wydajności - Podczas tworzenia pracy, aby poprawić wydajność zdecydowano się na wypróbowanie biblioteki Optuna,
- Stworzenie UI pozwalającego użytkownikowi dodanie pliku dźwiękowego oraz nagrania własnej wiadomości przy pomocy wbudowanego mikrofonu, a następnie analiza tych nagrań,
- Testowanie projektu.

Projekt jest realizowany zgodnie z harmonogramem

W trakcie trwania projektu nie wprowadzono żadnych nadzwyczajnych zmian ani w zakresie realizacji zadań projektowych, ani w składzie zespołu projektowego.

### 3 Doświadczenia uzyskane podczas realizacji projektu.

Realizacja projektu dotyczącego detekcji emocji w głosie była okazją do zdobycia wiedzy z zakresu przetwarzania sygnałów dźwiękowych. Podczas projektu dowiedzieliśmy się o technikach przetwarzania sygnałów dźwiękowych. Projekt pozwolił nam poznać w jakich postaciach można przedstawiać sygnały akustyczne. Dowiedzieliśmy się o cechach które są przydatne do analizy emocji w głosie, takie jak MFCC, Chroma STFT, Spektrogram, chroma, tonacja i inne.

Dodatkowo podczas realizacji projektu rozpoznawania emocji w głosie nauczyło się korzystania z biblioteki Librosa do szybkiej ekstrakcji cech dźwiękowych w dziedzinie częstotliwościowej oraz czasowej, co umożliwiło efektywne przetwarzanie danych audio. Kolejnym istotnym doświadczeniem podczas realizacji projektu była praca z metodami uczenia maszynowego przeznaczonymi do klasyfikacji emocji w głosie. W ramach projektu zagłębiliśmy wiedzy z zakresu różnych architektur sieci neuronowych takich jak: CNN czy RNN oraz nabyliśmy wiedzy z praktycznego wykorzystania biblioteki TensorFlow. Opanowano również użycie biblioteki Optuna do optymalizacji hiperparametrów, co pozwoliło na zwiększenie dokładności modelu z 77% do 80% na zbiorze testowym. Dodatkowo, zdobyto umiejętności tworzenia graficznych interfejsów użytkownika przy pomocy biblioteki Tkinter, co umożliwiło stworzenie intuicyjnej aplikacji prezentującej wyniki analizy głosu. W trakcie projektu zapoznano się także z biblioteką sounddevice, która pozwoliła na nagranie dźwięku poprzez wbudowany mikrofon.

### 4 Proponowane ulepszenia projektu.

Ulepszenie sieci neuronowej może być trudnym zadaniem, dlatego proponowane zmiany powinny skupić się na wykorzystaniu dodatkowej sieci do rozpoznawania emocji na podstawie mimiki twarzy, ponieważ mimika jest istotnym elementem komunikacji niewerbalnej. Można zastosować bibliotekę OpenCV lub FaceNet do ekstrakcji cech i klasyfikacji emocji. Integracja wyników z analizy dźwięku i obrazu w jednym systemie oraz stworzenie warstwy fuzji pozwoli na bardziej precyzyjne rozpoznawanie emocji. Projekt można by było również poprawić pod względem analizy sygnałów poprzez usunięcie ewentualnych szumów. Innym zaproponowanym ulepszeniem jest wykorzystanie innej architektury sieci neuronowej takiej jak Long Short-Term Memory (LSTM), która pozwala na skuteczne przetwarzanie sekwencyjnych danych. Architektura LSTM charakteryzuje się zdolnością do przechowywania informacji przez dłuższy okres, co jest kluczowe dla analizy długoterminowych zależności w sygnałach dźwiękowych

### 5 Lista dodatkowych plików

- src.rar - kody źródłowe służące do przetwarzania danych, nauki modelu ich ewaluacji oraz aplikacja służąca do rozpoznawania emocji w głosie
- prezentacja\_1.mp4 – przedstawienie możliwości odczytu pliku
- prezentacja\_2.mp4 – przedstawienie możliwości użycia mikrofonu