

Projekt zaliczeniowy z części audio

1 Cel projektu

Celem projektu jest opanowanie umiejętności analizowania plików dźwiękowych. Projekt wymaga przećwiczenia metod filtracji, parametryzacji i detekcji sygnałów dźwiękowych. Studenci na nagranych plikach dźwiękowych mają zliczyć ilość słów wypowiedzianych podczas rozmowy oraz przedstawić wnioski odnośnie wpływu zmęczenia na wartości parametrów. Projekt nie dotyczy rozpoznawania słów (tj. konwersji mowy na tekst).

2 Przebieg projektu

2.1 Akwizycja (w parach)

Studenci parami nagrywają dwie "rozmowy" pomiędzy sobą. Jedna rozmowa powinna być przeprowadzona na początku dnia, druga pod koniec (odstęp pomiędzy nimi powinien być od minimum 6 do maksimum 16 godzin). Obie rozmowy mają mieć identyczną treść. Rozmowa (może to być równie dobrze deklamacja poematu, odczytanie fragmentu książki, etc. - treść nie ma znaczenia) ma trwać ok. 3 minut i osoby mają mówić na przemian w ok. 20-30 sekundowych odcinkach czasowych zawierających min. 35 słów ("w", "i", etc. należy traktować jako osobne słowa). Wypowiedzi mają być prowadzone z "normalną" prędkością, tj. bez przesadnie długich przerw między słowami i uczestnikami, normalnym (zwykajnym) głosem oraz na podobnym poziomie głośności.

Analiza powinna być przeprowadzona w pakiecie MatLAB. Możliwe jest użycia innych narzędzi (np. python, R, C++) po wcześniejszym uzgodnieniu. W takim przypadku konieczne będzie dodatkowo umówienie się na pokazanie (zdalnie lub stacjonarnie) działania kodu.

UWAGA: Obowiązuje zakaz używania gotowych funkcji typu detectspeech i podobnych do wyznaczania ilości i początków słów w analizie z podrozdziału 2.2. Można je używać tylko przy analizie wpływu zmęczenia - podrozdział 2.3. Można bezpiecznie używać deskryptorów, energii w oknach, własnych parametrów, etc.

Nagranie powinno być zapisane w jednym następujących formatów: *.m4a, *.wav lub *.mp3 z częstotliwością próbkowania min. 8 KHz. Inne formaty proszę przekonwertować na *.mat (format binarny MatLAB).

2.2 Analiza (indywidualnie) pojedynczej rozmowy

W drugiej fazie każda osoba **samodzielnie** dokonuje następujących rzeczy:

- detekcji początku i końca każdego ze swoich fragmentów dla jednej, dowolnej rozmowy. Detekcja ma być automatyczna na podstawie wartości parametrów, a nie ręcznie (na sztywno) wpisana w kod;
- dla każdego fragmentu wypowiedzanego **przez siebie** zlicza ilość słów na dwa sposoby: ręcznie (kontrola - wypowiedziane) i przy użyciu własnoręcznie stworzonego kodu (obliczone);
- policzenie błędu detekcji dla każdego fragmentu oraz średniego (z całości) wg wzorów:

$$\text{Błąd_fragmentu} = \frac{|\text{ilość_słów_wypowiedzianych} - \text{ilość_słów_obliczonych}|}{\text{ilość_słów_wypowiedzianych}} \cdot 100\% \quad (1)$$

$$\text{Błąd}_{\text{średni}} = \frac{\sum_{k=1}^{\text{ilość_fragmentów}} [\text{Błąd_fragmentu}(k) \cdot \text{ilość_słów_wypowiedzianych}(k)]}{\sum_{k=1}^{\text{ilość_fragmentów}} \text{ilość_słów_wypowiedzianych}(k)} \quad (2)$$

2.3 Analiza obu rozmów - wpływ zmęczenia

Z obu nagranych rozmów należy wybrać jeden fragment (ten sam w obu nagraniach). Następnie policzyć wartości następujących parametrów:

- średni czas wypowiadania każdego ze słów (ile słów mówimy szybciej / wolniej i jaka to jest różnica - min, max, średnia);
- średni czas przerwy między słowami (czy wydłużamy/skracamy przerwy pomiędzy słowami/zdaniami);
- dla pierwszych 5 słów - czy istnieją zmiany tonu podstawowego dla poszczególnych fonemów (jeśli tak, to jakie).

Początki i końce słów można w ramach tej części wyznaczyć /poprawić ręcznie, jeśli analiza automatyczna okaże się mało dokładna.

2.4 Sprawozdanie

Sprawozdanie powinno zawierać następujące informacje:

- warunki przeprowadzenia akwizycji: kto i kiedy brał udział w nagraniu, jak wykonano eksperyment (parametry sprzętowe: ilość kanałów, format, rodzaj mikrofonu (wbudowany / zewnętrzny), częstotliwość próbkowania, etc.), które nagranie wykorzystano w analizie detekcji słów, a który fragment w analizie zmęczenia;
- opis schematu przetwarzania: rodzaj filtracji, parametry użyte do determinacji swojej części, parametry użyte do detekcji słów;
- tabela z opisem danych wejściowych dla każdego, własnego fragmentu osobno z wartościami wyznaczonymi ręcznie i przy użyciu algorytmu (tab. 1)
- wykresy podzielonych na ok. 30 sekundowe fragmentów z zaznaczeniem: początku i końca swojego fragmentu (lub różnymi kolorami: własny / obcy fragment), linii pionowych podziału między słowami;
- wyniki analizy zmęczenia wraz z omówieniem;
- listing wypowiedzi z podziałem na sekcje.

Tabela 1: Tabela do sprawozdania

numer fragmentu	początek [s]	koniec [s]	ilość słów wypowiedzianych	ilość słów obliczonych	błąd [%]
1					
2					
...	...	...	...	...	...
Suma	–	–	$\sum$	$\sum$	średni błąd

Sprawozdanie (w formacie PDF) wraz z plikami audio i kodem źródłowym należy wysłać spakowane na adres: **dwornik@agh.edu.pl**

3 Ocena projektu

Projekt jest oceniany na maksymalnie 100 punktów. Na ocenę z projektu ma wpływ termin oddania i jakość detekcji w następujący sposób:

1. Termin zaliczenia:

- do 4 maja 2025: mnożnik 1.00;
- od 5 maja do 8 czerwca 2025: mnożnik 0.75;
- od 9 czerwca do 6 lipca 2025: mnożnik 0.50;
- od 7 lipca 2025: brak zaliczenia.

2. Jakość detekcji (wszystkie wskaźniki mieszczą się w zadanym przedziale):

- średni błąd detekcji $\leq 5\%$, a dla każdego fragmentu błąd $\leq 10\%$: mnożnik 1.00;
- średni błąd detekcji $\leq 10\%$, a dla każdego fragmentu błąd $\leq 15\%$: mnożnik 0.80;
- średni błąd detekcji $\leq 15\%$, a dla każdego fragmentu błąd $\leq 20\%$: mnożnik 0.60;
- średni błąd detekcji $\leq 20\%$, a dla każdego fragmentu błąd $\leq 25\%$: mnożnik 0.40;
- średni błąd detekcji $> 20\%$ lub istnieje fragment z błędem $> 25\%$: brak zaliczenia.

UWAGA: Ocenę końcową z przedmiotu oraz wagę ocen poszczególnych składowych zaliczenia przedmiotu ustala Prof. Młynarczyk.