Projekt 2 - Analiza i Przetwarzanie Dźwięku

Zofia Kamińska

Kwiecień 2025

Spis treści

1 Wprowadzenie

Celem projektu jest stworzenie interaktywnej aplikacji do analizy sygnałów dźwiękowych (plików .wav) z możliwością wizualizacji wybranych cech audio. Program pozwala na wczytywanie sygnału, wybór parametrów analizy (długość ramki, okno, nakładanie) oraz prezentuje wyniki w postaci wykresów fali dźwiękowej, widma, spektrogramu i cepstrum.
W ramach projektu przewidziana jest również analiza różnych cech dźwięku oraz porównanie sygnałów pochodzących od różnych głosów, co pozwoli na identyfikację charakterystycznych właściwości mowy oraz różnic pomiędzy użytkownikami.

2 Środowisko i biblioteki

Aplikacja została napisana w języku Python, z wykorzystaniem następujących bibliotek:

- **PyQt5** – interfejs graficzny
- **librosa** – przetwarzanie dźwięku
- **matplotlib** – wizualizacja danych
- **numpy** – operacje matematyczne
- **scipy.signal** – funkcje okna

3 Funkcjonalności aplikacji

Aplikacja oferuje zestaw narzędzi do interaktywnej analizy plików dźwiękowych, umożliwiając użytkownikowi szczegółowe przetwarzanie i wizualizację sygnału audio. Poniżej przedstawiono główne funkcjonalności:

- Wczytywanie plików audio (format WAV)
- Analiza ramkowa z wyborem długości: 10ms, 20ms, 30ms, 40ms
- Wybór funkcji okna: prostokątne, trójkątne, Hamming, Hanning, Blackman
- Nakładanie ramek (ang. overlapping)
- Obliczanie cech częstotliwościowych:
  - Głośność
  - Centroid częstotliwości
  - Szerokość pasma
  - ERSB1–3 (energia w podzakresach)
  - Płaskość i wypukłość widmowa
- Dynamiczna aktualizacja wykresów po kliknięciu i zoomie

4 Struktura aplikacji

4.1 Metody główne

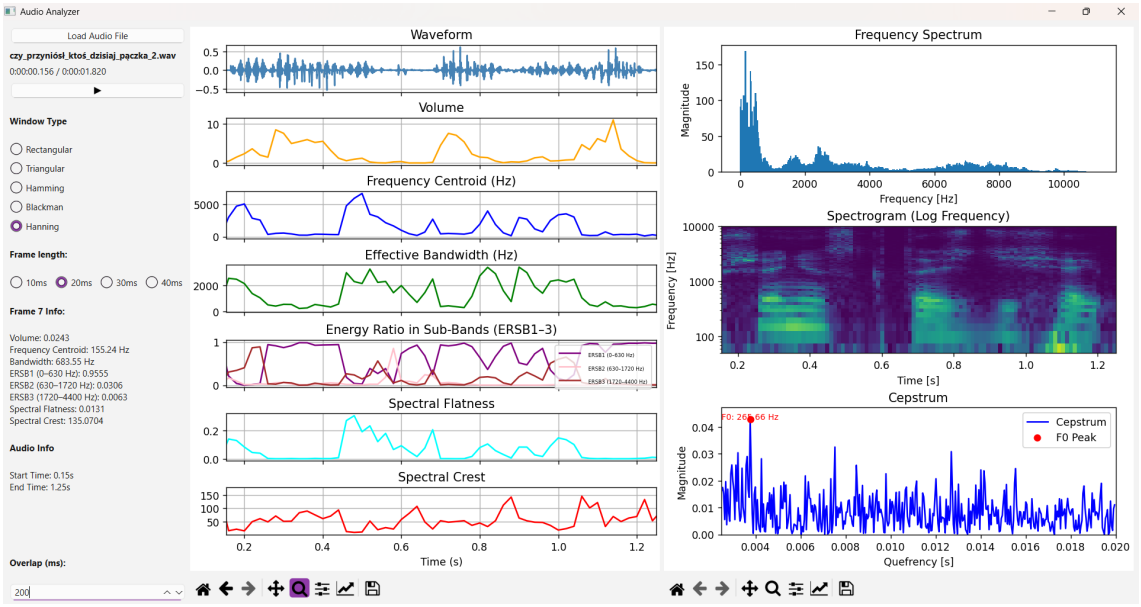
- `load_audio()` – ładowanie pliku dźwiękowego
- `process_audio()` – przetwarzanie sygnału i ekstrakcja cech
- `plot_audio()` – wykres sygnału, wywołanie pozostałych funkcji tworzących wykresy
- `plot_spectrum_for_selection()` – tworzenie widma wybranego fragmentu
- `plot_spectrogram_for_selection()` – tworzenie spektrogramu (z użyciem skali logarytmicznej)
- `plot_cepstrum_for_selection()` – tworzenie wykresu cepstrum i wyniku estymacji tonu podstawowego

4.2 Interakcje z użytkownikiem

Interfejs aplikacji umożliwia wygodną i intuicyjną analizę sygnału dźwiękowego. Poniżej przedstawiono dostępne formy interakcji użytkownika z aplikacją.

- Kliknięcie na wykresie – przeskok odtwarzania oraz wyświetlenie cech
- Zoom wykresu – aktualizacja widma, spektrogramu i cepstrum
- Zmiana opcji – ponowne przetworzenie sygnału

Poniżej znajduje się zrzut ekranu prezentujący graficzny interfejs użytkownika (GUI) aplikacji.



Rysunek 1: Graficzny interfejs użytkownika aplikacji

4.3 Opis metod w `feature_extractor.py`

Wszystkie metody opisane w tej sekcji zostały zaimplementowane zgodnie z wytycznymi i wzorami zawartymi w materiale źródłowym udostępnionym przez prowadzącego (pierwsza pozycja w bibliografii). Moduł `feature_extractor.py` zawiera funkcje służące do ekstrakcji wybranych cech sygnału dźwiękowego, ze szczególnym uwzględnieniem analizy w dziedzinie częstotliwości, obliczania cepstrum oraz szacowania częstotliwości podstawowej.

- **Ekstrakcja cech w dziedzinie częstotliwości**
  Funkcja `extract_frequency_features(audio, sr, frame_length, window_type)` przetwarza sygnał audio w ramkach o zadanej długości i ekstraktuje następujące cechy częstotliwościowe:
  - **Głośność (Volume)** – średnia energia widma mocy sygnału.
  - **Centroid częstotliwości (Frequency Centroid)** – ważona średnia częstotliwości, odzwierciedlająca środek ciężkości widma.
  - **Efektywna szerokość pasma (Effective Bandwidth)** – odchylenie standardowe częstotliwości względem centroidu, ważone mocą widmową.
  - **Energia w pasmach ERSB** – trzy współczynniki odpowiadające procentowemu udziałowi energii w przedziałach częstotliwości: 0–630 Hz, 630–1720 Hz, 1720–4400 Hz.
  - **Spłaszczenie widmowe (Spectral Flatness Measure)** – stosunek średniej geometrycznej do średniej arytmetycznej widma mocy, charakteryzujący „szorstkość” dźwięku.
  - **Współczynnik szczytowości widma (Spectral Crest Factor)** – stosunek maksymalnej wartości widma do jego średniej.

  Dla każdego okna stosowana jest funkcja okna (np. van Hanna, Hamminga) w celu ograniczenia efektów wycieku widmowego.

- **Funkcje okna**
  Funkcja `get_window(window_type, frame_length)` zwraca funkcję okna o określonym typie (np. prostokątne, Hamming, Blackman itp.) i długości. Funkcje okna są wykorzystywane w celu poprawy jakości analizy widmowej poprzez ograniczenie efektów nieciągłości na brzegach ramek.

- **Cepstrum rzeczywiste**
  Funkcja `compute_real_cepstrum(signal)` oblicza rzeczywiste cepstrum sygnału. Obliczenia obejmują:

  1. Transformację Fouriera sygnału,
  2. Obliczenie logarytmu wartości bezwzględnej widma,
  3. Zastosowanie odwrotnej transformacji Fouriera.

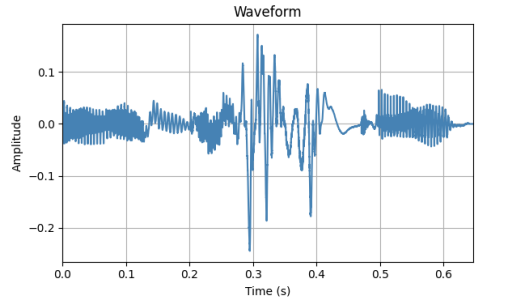
  Cepstrum rzeczywiste wykorzystywane jest m.in. do analizy struktury harmonicznej i detekcji tonu podstawowego (f0).

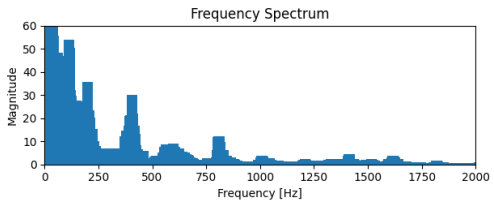
- **Szacowanie częstotliwości podstawowej (f0)**
  Funkcja `estimate_f0_from_cepstrum(cepstrum, sr, min_freq, max_freq)` szacuje częstotliwość podstawową sygnału na podstawie cepstrum. Przeszukuje wyznaczony zakres quefrencji odpowiadający zakładanemu przedziałowi częstotliwości [50 Hz, 400 Hz], a następnie znajduje maksymalną wartość w tym zakresie. Częstotliwość podstawowa obliczana jest jako odwrotność położenia tego piku.

5 Przykłady - wizualizacje i wnioski

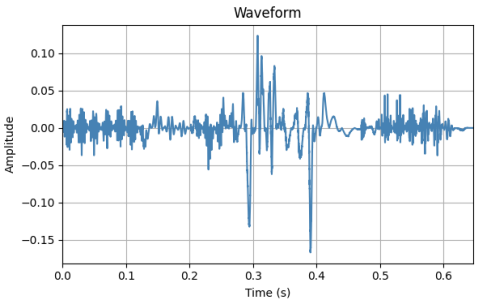
5.1 Wpływ funkcji okienkowych

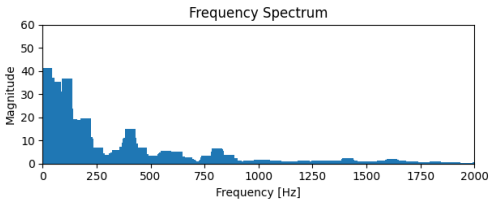
W celu zbadania wpływu różnych funkcji okna na wynik analizy widmowej, przeprowadzono eksperyment polegający na obliczeniu widma tego samego fragmentu sygnału mowy z zastosowaniem różnych okien: prostokątnego, trójkątnego, van Hanna oraz Blackmana. Długość ramki wynosiła 20 ms, bez nakładania ramek.
Poniżej przedstawione zostały wykresy przebiegu czasowego, oraz spektrum częstotliwościowe dla sygnału słowa "lodówka"pochodzącego z pliku *lodowka_ 1.wav*. Analizowane okna to: okno prostokątne (dźwięk oryginalny), trójkątne, van Hann oraz Blackman.



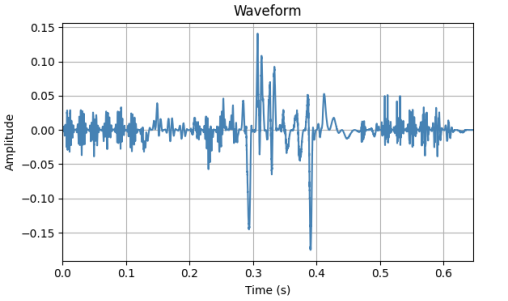


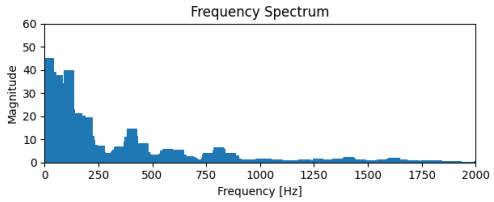
(a) okno prostokątne



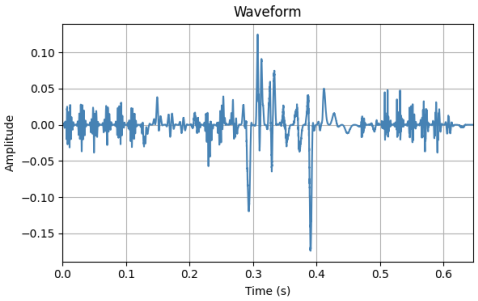


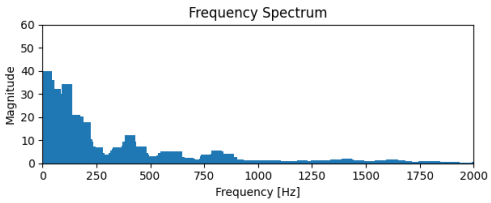
(b) okno trójkątne





(c) okno van Hanna





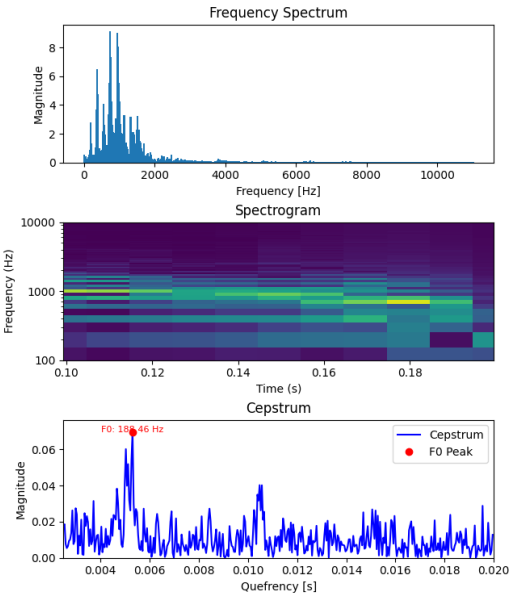
(d) okno Blackmana

Rysunek 2: Porównanie wpływu różnych funkcji okienkowych na przebieg czasowy i widmo sygnału dla słowa "lodówka".

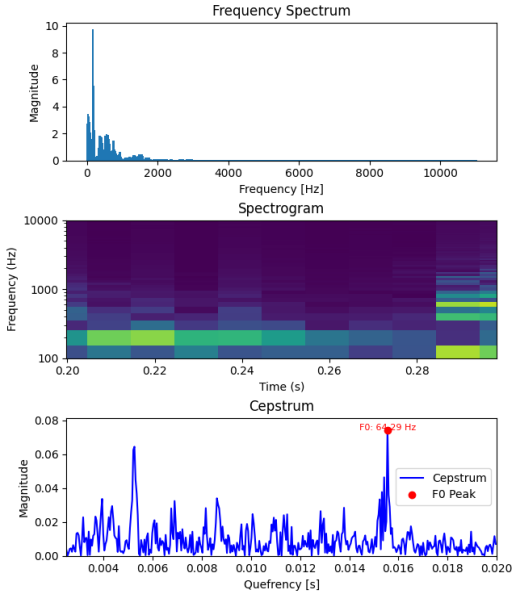
Zastosowanie różnych funkcji okna na całym sygnale poprzez ramkowanie z długością 20 ms, wpływa na ogólny kształt przetworzonego przebiegu czasowego. W przypadku okna prostokątnego sygnał pozostaje niezmieniony w strukturze i wszystkie zakłócenia w dźwięku są wyraźnie widoczne.
Dla okien wygładzających, takich jak van Hann, Hamming czy trójkątne, poszczególne ramki są łagodniej zintegrowane w całościowym przebiegu. Kształt sygnału staje się bardziej płynny, z mniejszą ilością gwałtownych przejść. Cały sygnał wygląda bardziej spójnie i naturalnie, co może sprzyjać dalszej analizie cech, np. detekcji tonu podstawowego czy estymacji formantów.
Najsilniejszy efekt wygładzenia występuje przy użyciu okna Blackmana, co może ułatwiać analizę ogólnej struktury, ale jednocześnie prowadzi do utraty drobnych zmian lokalnych w sygnale.
Na wykresach częstotliwościowych (widmach) okna wygładzające – trójkątne, van Hanna oraz Blackman – działają w sposób bardzo podobny, zmniejszając zmienność w widmie i prowadząc do ogólnego spłaszczenia całego widma. Dzięki temu widmo staje się gładsze, co może ułatwić identyfikację głównych składowych częstotliwościowych. Okno Blackmana powoduje najsilniejsze spłaszczenie, ale w kontekście czytelności widma różnice między tym oknem a trójkątnym czy van Hannem są niewielkie.

5.2 Głoski: spółgłoski vs samogłoski

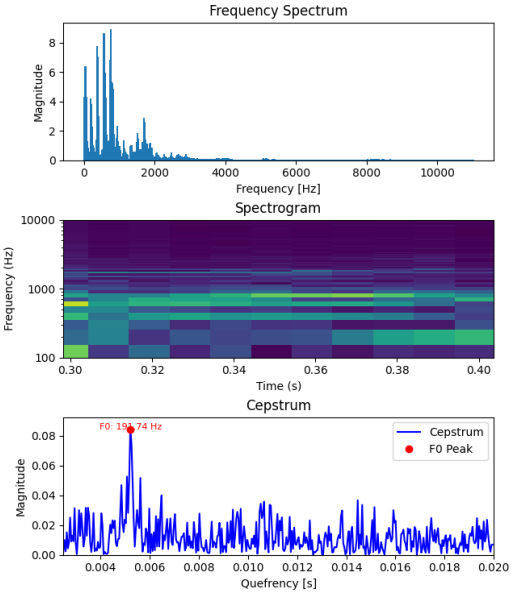
Analizując różnice między spółgłoskami a samogłoskami, wyodrębniono krótkie fragmenty dźwięków odpowiadające głoskom ze słowa: *"abe"* oraz osobno głosce *"sz"* w celu efektywniejszego porównania. Dla każdej głoski obliczono widmo, spektrogram oraz wybrane cechy widmowe: centroid częstotliwościowy, płaskość widmową oraz szerokość pasma.



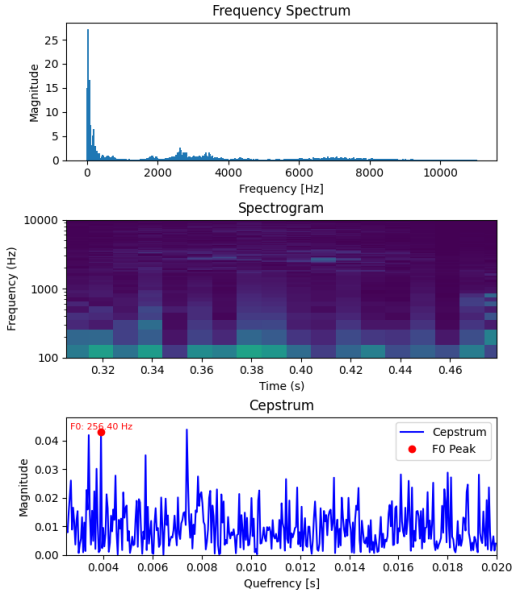
(a) samogłoska "a"



(b) spółgłoska "b"
(przechodząca na koniec w "e")



(c) samogłoska "e"



(d) spółgłoska "sz"

Rysunek 3: Widma, spektrogramy oraz cepstrum wybranych głosek – samogłosek ("a", "e") i spółgłosek ("b", "sz").

Na spektrogramach samogłosek „a" i „e" widać wyraźne poziome pasma (formanty). W „a" najintensywniejsze pasmo znajduje się w okolicach 900–1000 Hz, a w „e" niżej – ok. 700-800 Hz, co jest zgodne z ich typową strukturą. „b" pokazuje energię skupioną w niskich częstotliwościach (200–300 Hz) i brak formantów, a „sz" ma rozlane widmo – od niskich po wysokie częstotliwości – typowe dla szumu.
Na wykresach widma częstotliwościowego „a" i „e" mają wyraźniejsze piki (formanty), „b" – niższy zakres energii, a „sz" – szerokie, rozproszone widmo z dominacją szumu.
Cepstrum samogłosek zawiera wyraźny pik (oznaka tonów harmonicznych), natomiast dla „sz" piki są aż tak wyróżniające się potwierdzając brak wyraźnej okresowości. Ton podstawowy (F0),

liczony na podstawie cepstrum, był nieco wyższy dla „e” (191.74 Hz) niż dla „a” (189.46 Hz), co może wynikać z różnic artykulacyjnych.

Głoska	Centroid [Hz]	Płaskość widmowa	Szerokość pasma [Hz]
a	928.4 ± 142.7	0.0093 ± 0.0041	428.9 ± 102.3
e	765.2 ± 130.6	0.0151 ± 0.0056	711.7 ± 148.5
b	381.6 ± 87.4	0.0378 ± 0.0113	954.2 ± 161.9
sz	1988.3 ± 312.5	0.5847 ± 0.0892	1882.6 ± 239.8

Tabela 1: Średnie cechy widmowe i odchylenia standardowe dla 3 losowo wybranych osób

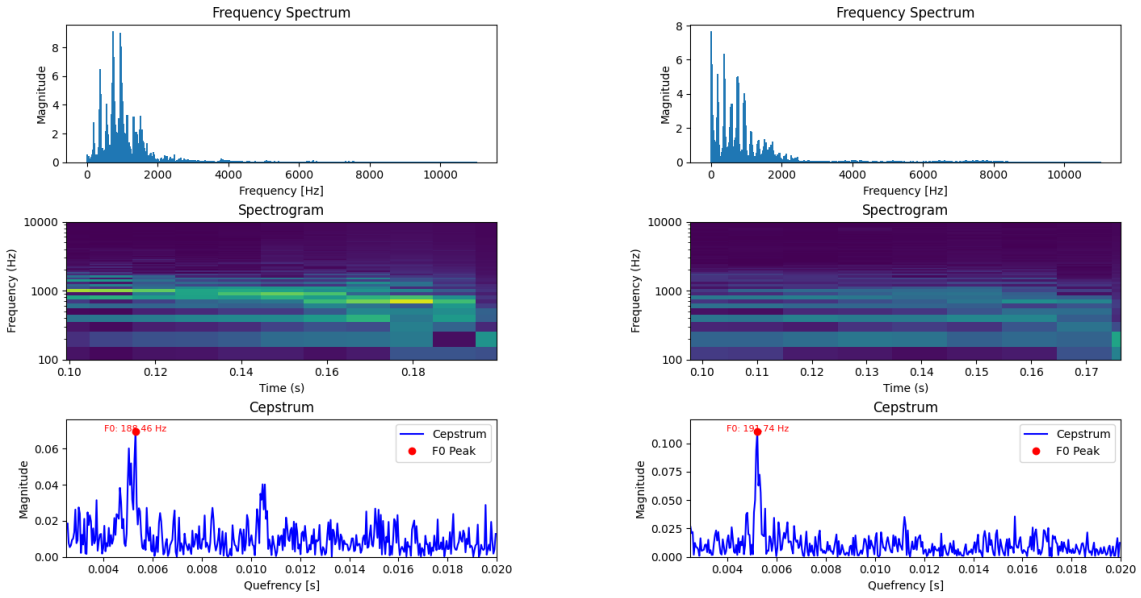
Zgodnie z oczekiwaniami, samogłoski ("a", "e") charakteryzują się niższym centroidem częstotliwościowym, mniejszą płaskością widmową i węższym pasmem w porównaniu do spółgłosek. Głoska "sz"jako spółgłoska szumowa wykazuje znacznie wyższy centroid i bardzo wysoką płaskość widmową, typową dla sygnałów szumowych. Różnice w odchyleniach standardowych wynikają z naturalnej zmienności głosów męskich i żeńskich oraz jakości nagrań.

5.3 Formanty i ton podstawowy

Formanty to rezonanse toru głosowego, które wzmacniają określone zakresy częstotliwości w dźwiękach mowy – zwłaszcza w samogłoskach. Każda samogłoska ma charakterystyczny układ formantów, z których dwa najniższe (F1 i F2) są najważniejsze dla jej rozróżnienia. Na spektrogramie formanty objawiają się jako jasne, poziome pasma, utrzymujące się przez czas trwania dźwięku. Na wykresie widma częstotliwościowego (frequency spectrum) formanty widoczne są jako wyraźne piki amplitudy w określonych zakresach częstotliwości.

5.3.1 Próbki tej samej osoby

Dla zobrazowania tego zjawiska, porównano dwie próbki głoski „a” wypowiedziane przez tę samą osobę (autora). Obie próbki zawierają ten sam fonem, ale mogły różnić się nieznacznie sposobem artykulacji, intonacją lub głośnością.
W przypadku dwóch nagrań tej samej głoski od tej samej osoby spodziewamy się podobnego układu formantów i zbliżonej częstotliwości podstawowej (F0).



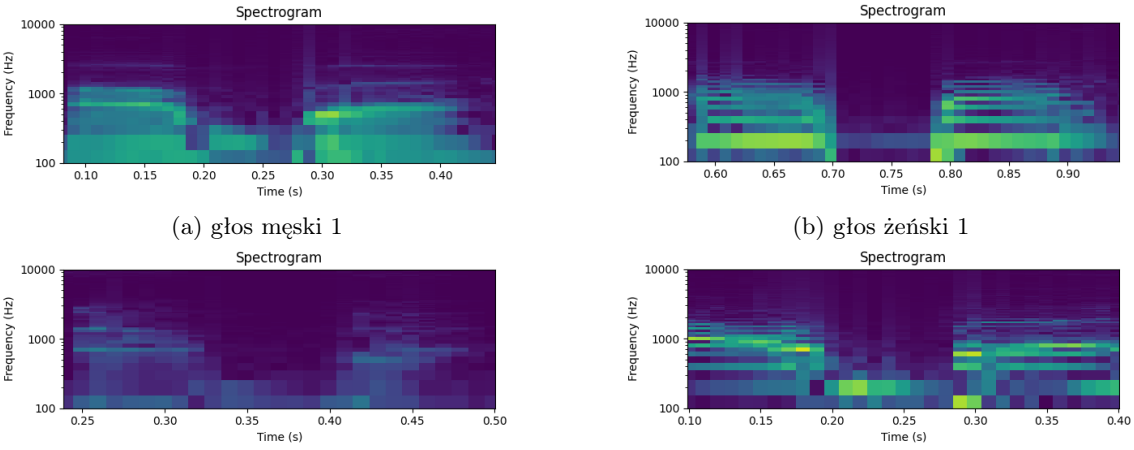
Rysunek 4: Widma, spektrogramy oraz cepstrum samogłoski 'a' - ta sama osoba

Na podstawie porównania dwóch próbek głoski „a” wypowiedzianych przez tę samą osobę można zauważyć kilka różnic. W drugiej próbce częstotliwość podstawowa (F0) była nieco wyższa, co może wynikać z delikatnie wyższej intonacji lub napięcia głosu przy jej wypowiedzeniu. Cepstrum w obu przypadkach miało zbliżoną strukturę, jednak w pierwszej próbce piki były wyraźniejsze, co może świadczyć o lepszej jakości tonalnej lub mniejszym udziale szumu.

Spektrogramy wykazały wyraźniejszą intensywność (jaśniejsze pasma) w pierwszej próbce, co może być efektem mocniejszej artykulacji lub bliższego ustawienia mikrofonu. Mimo różnic w intensywności, układ formantów pozostał niemal identyczny – co potwierdza stabilność właściwości artykulacyjnych samogłoski „a" u tej samej osoby.

5.3.2 Porównanie dla różnych osób

Aby zbadać, jak zmieniają się właściwości formantów i tonu podstawowego w zależności od mówcy, przeanalizowano nagrania słowa „abe" wypowiedziane przez cztery różne osoby: dwie kobiety i dwóch mężczyzn. Każdy z nagranych sygnałów został poddany analizie widmowej oraz spektrogramowej.
Ze względu na różnice anatomiczne (długość i kształt toru głosowego) oraz naturalne cechy głosu, spodziewane są wyraźne różnice w wyglądzie spektrogramów.



(a) głos męski 1

(b) głos żeński 1

(c) głos męski 2

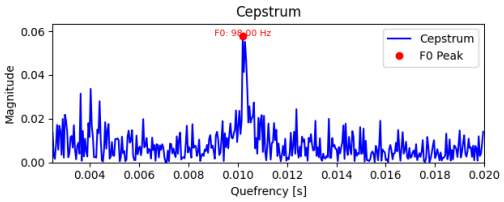
(d) głos żeński 2

Rysunek 5: Spektrogramy dla słowa *"abe"* - porównanie różnych głosów.

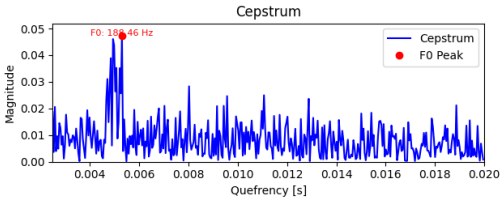
Na podstawie spektrogramów zauważalne są różnice w rozmieszczeniu i intensywności formantów pomiędzy głosami kobiet a mężczyzn. U kobiet pasma formantowe są wyraźniejsze i częściej występuje ich większa liczba, szczególnie w wyższych częstotliwościach. U mężczyzn formanty są położone niżej, co jest zgodne z krótszą długością toru głosowego, jednak ich intensywność jest mniejsza, co prawdopodobnie wynika z różnic w jakości nagrań (np. odległości od mikrofonu).

Ton podstawowy

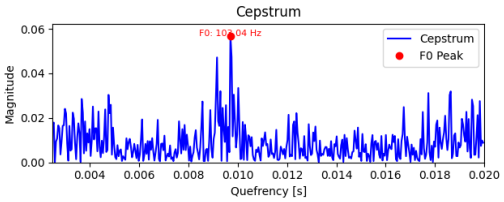
Ton podstawowy (F0) odpowiada za wysokość odbieranego dźwięku i stanowi podstawę dla powstawania wyższych harmonicznych. W mowie ludzkiej F0 zależy od wielu czynników, takich jak płeć, wiek, emocje czy sposób artykulacji. W przypadku tej samej osoby ton podstawowy powinien utrzymywać się na zbliżonym poziomie, choć mogą występować niewielkie naturalne różnice między kolejnymi realizacjami tych samych wyrazów.
Dla dorosłych mężczyzn typowy zakres tonu podstawowego mieści się w przedziale 85–180 Hz, natomiast dla kobiet wynosi 165–255 Hz. Różnice te wynikają głównie z budowy anatomicznej krtani i długości fałdów głosowych.



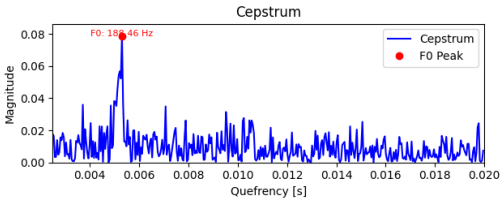
(a) głos męski 1



(b) głos żeński 1



(c) głos męski 2



(d) głos żeński 2

Rysunek 6: Wykresy cepstrum dla słowa *"abe"* - porównanie różnych głosów.

Na podstawie wykresów cepstrum można zauważyć wyraźne różnice w położeniu tonu podstawowego (F0) pomiędzy głosami męskimi a żeńskimi. W próbkach głosów męskich główne piki pojawiały się w wyższych wartościach quefrency, co odpowiada niższym częstotliwościom F0 typowym dla tej grupy. W przypadku głosów żeńskich piki były przesunięte w stronę niższego quefrency (wyższych częstotliwości) zgodnie z oczekiwanym zakresem dla kobiet.
Dokładne wartości średnie i odchylenia standardowe tonu podstawowego dla każdej osoby zostały przedstawione w kolejnej tabeli.

Osoba	**Ton podstawowy [Hz]**
Głos męski 1	107.7 ± 23.5
Głos męski 2	118.4 ± 31.7
Głos żeński 1	215.6 ± 47.2
Głos żeński 2	202.3 ± 31.1

Tabela 2: Średnia wartość tonu podstawowego oraz odchylenie standardowe dla 5 próbek słów oraz próki zdania wypowiedzianych przez każdą osobę.

Porównując różne głosy, zauważono wyraźne różnice zarówno w tonie podstawowym, jak i w rozmieszczeniu formantów. Głosy różniły się wysokością tonu oraz charakterystyką widmową, co odzwierciedla naturalne indywidualne cechy mówców. Nawet w obrębie jednej osoby zaobserwowano istotne wahania tonu podstawowego, co może wynikać z różnic w intonacji lub spontanicznej artykulacji podczas nagrania.

6 Podsumowanie

Projekt stanowił połączenie wiedzy z zakresu analizy sygnałów dźwiękowych oraz praktycznych umiejętności programistycznych. Powstała aplikacja umożliwia wczytywanie plików dźwiękowych, analizę ich struktury oraz wizualizację kluczowych cech akustycznych. Dzięki interaktywnemu interfejsowi możliwe było łatwe porównywanie wyników analizy dla różnych parametrów, co ułatwia lepsze zrozumienie zjawisk zachodzących w sygnałach mowy.
Przeprowadzone eksperymenty dotyczyły m.in. porównania właściwości spółgłosek i samogłosek, tonu podstawowego w głosach różnych osób oraz rozmieszczenia formantów dla wybranych fonemów. Wyniki tych analiz potwierdzają istnienie wyraźnych różnic akustycznych między badanymi przypadkami i dobrze obrazują charakterystyczne cechy mowy.
Realizacja projektu pozwoliła również na praktyczne zastosowanie teorii z zakresu przetwarzania sygnałów oraz rozwój umiejętności analitycznych i programistycznych. Ostateczny efekt może być przydatny zarówno jako narzędzie edukacyjne, jak i jako punkt wyjścia do bardziej zaawansowanych analiz fonetycznych czy badań nad sygnałem mowy.

7 Źródła (audio)

Zasoby własne (nagrane podczas zajęć)

Literatura

[1] J. Rafalko, "Cechy sygnału audio w dziedzinie częstotliwości,"2024.
URL: https://pages.mini.pw.edu.pl/ rafalkoj/www/?Dydaktyka:2024

[2] J. Rafalko, "Cepstrum,"2024.
URL: https://pages.mini.pw.edu.pl/ rafalkoj/www/?Dydaktyka:2024

[3] ChatGPT, OpenAI, 2025.
URL: https://www.openai.com/chatgpt