

Analiza i Przetwarzanie Dźwięku Projekt 2 - Analiza częstotliwościowa

Karolina Dunal

24 kwietnia 2025

Spis treści

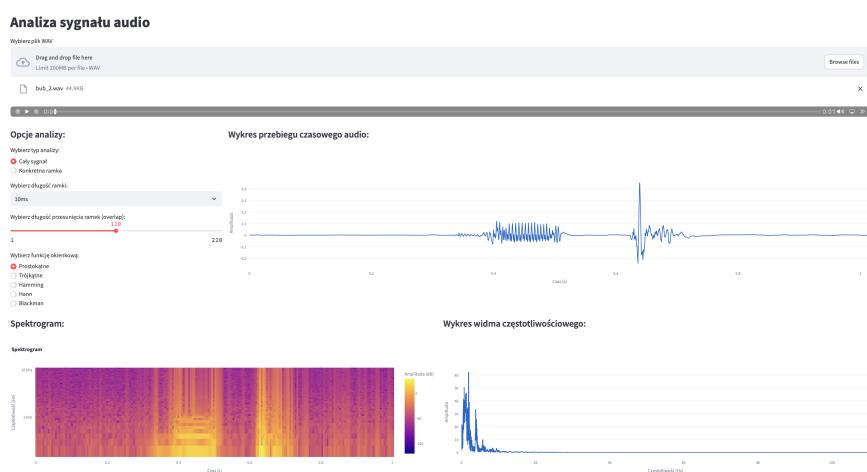
1	Opis Zadania	2
2	Opis Aplikacji	2
3	Metody Użyte w Aplikacji	3
3.1	Metoda <code>frequency_spectrum</code>	3
3.2	Metoda <code>apply_window</code>	3
3.3	Metoda <code>compute_volume</code>	4
3.4	Metoda <code>compute_frequency_centroid</code>	4
3.5	Metoda <code>compute_bandwidth</code>	4
3.6	Metoda <code>compute_band_energy</code>	4
3.7	Metoda <code>compute_band_energy_ratios</code>	4
3.8	Metoda <code>compute_spectral_flatness</code>	4
3.9	Metoda <code>compute_spectral_crest</code>	4
3.10	Metoda <code>compute_cepstrum</code>	5
3.11	Metoda <code>compute_spectrogram</code>	5
3.12	Metoda <code>extract_frame_features</code>	5
3.13	Metoda <code>get_frequencies</code>	5
4	Porównanie Wyników	6
4.1	Parametry w dziedzinie częstotliwości	6
4.2	Ton podstawowy	6
4.3	Formanty	7
4.4	Funkcje okna	8
5	Podsumowanie	9
6	Źródła	10

1 Opis Zadania

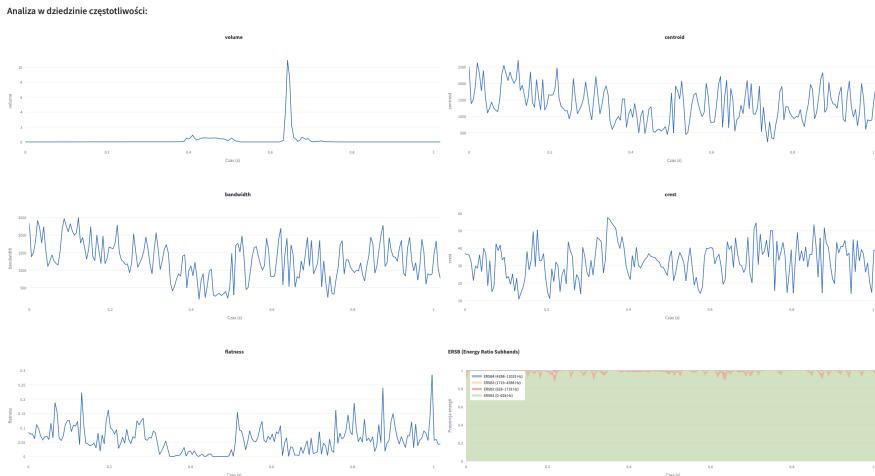
Celem projektu było stworzenie aplikacji do analizy plików audio, która umożliwia użytkownikowi wczytywanie, przetwarzanie oraz wizualizowanie wybranych parametrów plików audio w formacie WAV. Aplikacja ma na celu dostarczenie narzędzi do analizy dźwięku w dziedzinie częstotliwości.

2 Opis Aplikacji

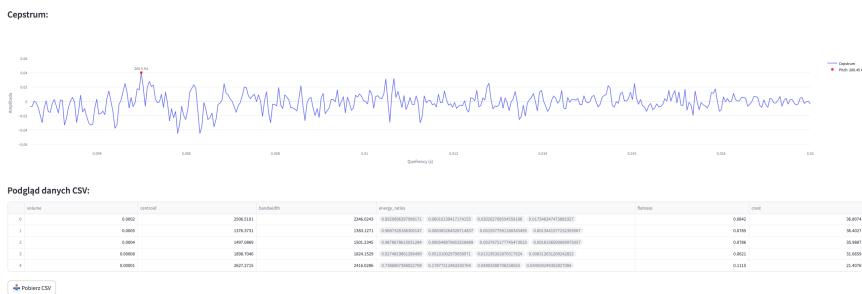
Aplikacja umożliwia analizę sygnałów audio poprzez obliczenie ich widma częstotliwościowego za pomocą transformacji Fouriera. Użytkownik może nałożyć funkcje okna na ramki sygnału, co redukuje efekt przecieków widmowych. Aplikacja oblicza także cechy spektroskopowe, takie jak głośność, centroid częstotliwościowy, szerokość pasma, energię w określonych pasmach oraz wskaźniki jakości widma. Dodatkowo, aplikacja generuje spektrogramy i umożliwia ekstrakcję cech dla wybranych fragmentów sygnału. Do realizacji tych funkcji użyto bibliotek Do realizacji wykorzystano biblioteki Pythonowe: NumPy do obliczeń numerycznych, SciPy do transformacji Fouriera, Librosa do analizy dźwięku oraz Plotly do generowania interaktywnych wizualizacji. W pliku `requirements.txt` znajdują się paczki wymagane do prawidłowego działania aplikacji. Plik `app.py` zawiera interfejs graficzny oraz logikę aplikacji, natomiast w pliku `audio_analysis` zaimplementowano funkcje odpowiedzialne za analizę audio.



Rysunek 1: Zrzut ekranu przedstawiający górną część interfejsu aplikacji. Na zdjęciu widoczne są przyciski umożliwiające wybór różnych funkcji oraz wykres przebiegu czasowego, spektrum i spektrogram.



Rysunek 2: Zrzut ekranu przedstawiający część przykładowych wykresów dla analizy w dziedzinie częstotliwości widocznych w interfejsie aplikacji.



Rysunek 3: Zrzut ekranu przedstawiający część interfejsu aplikacji z wykresem cepstrum, podglądem wyników analizy w formacie CSV oraz opcją zapisu do pliku.

3 Metody Użyte w Aplikacji

3.1 Metoda frequency_spectrum

Funkcja oblicza widmo częstotliwościowe sygnału czasowego y za pomocą dyskretnej transformaty Fouriera (DFT). Wynikiem są częstotliwości odpowiadające kolejnym binom oraz odpowiadające im amplitudy, czyli wartości bezwzględne zespolonych współczynników:

$$Y[k] = \sum_{n=0}^{N-1} y[n] \cdot e^{-j2\pi kn/N}$$

3.2 Metoda apply_window

Funkcja nakłada na ramkę sygnału wybraną funkcję okna $w[n]$, co pozwala redukować artefakty widmowe. Operacja ta tłumii próbki na krańcach ramki. Możliwe do wyboru funkcje okna to prostokątne, trójkątne, Hamminga, Hanninga lub Blackmana.

3.3 Metoda compute_volume

Oblicza poziom głośności dla danej ramki sygnału. Jest to średnia moc sygnału w oknie czasowym:

$$V = \frac{1}{N} \sum_{n=0}^{N-1} |S_n[n]|^2$$

gdzie $S_n[n]$ to próbka sygnału w ramce n , a N to rozmiar okna analizy.

3.4 Metoda compute_frequency_centroid

Wyznacza centroid częstotliwościowy (środek ciężkości widma amplitudowego), który odzwierciedla położenie „środka” energii częstotliwościowej:

$$f_c = \frac{\sum_k f[k] \cdot |S_n[k]|}{\sum_k |S_n[k]|}$$

gdzie $S_n[k]$ to wartość widma dla binu k w ramce n , a $f[k]$ to odpowiadająca mu częstotliwość.

3.5 Metoda compute_bandwidth

Oblicza efektywną szerokość pasma, czyli rozrzut energii wokół centroidu:

$$B = \sqrt{\frac{\sum_k (f[k] - f_c)^2 \cdot |S_n[k]|^2}{\sum_k |S_n[k]|^2}}$$

3.6 Metoda compute_band_energy

Wyznacza sumę energii sygnału zawartą w określonym przedziale częstotliwości $[f_0, f_1]$:

$$E_{[f_0, f_1]} = \sum_{f \in [f_0, f_1]} |S_n[f]|^2$$

3.7 Metoda compute_band_energy_ratios

Dzieli widmo na cztery zakresy częstotliwości i oblicza udział energii każdego z nich względem całkowej energii:

$$\text{ERSB}_i = \frac{E_i}{E_{\text{total}}}$$

gdzie E_i to energia w i -tym paśmie.

3.8 Metoda compute_spectral_flatness

Mierzy „płaskość” widma — stosunek średniej geometrycznej do arytmetycznej mocy widma.

$$\text{SF} = \frac{\left(\prod_k |S_n[k]|^2 \right)^{1/N}}{\frac{1}{N} \sum_k |S_n[k]|^2}$$

3.9 Metoda compute_spectral_crest

Oblicza stosunek najwyższego piku mocy do średniej mocy widma, co pozwala ocenić „tonalność” sygnału:

$$\text{SC} = \frac{\max_k |S_n[k]|^2}{\frac{1}{N} \sum_k |S_n[k]|^2}$$

3.10 Metoda compute_cepstrum

Oblicza cepstrum sygnału, które pozwala analizować okresowości w widmie, np. wysokość tonu:

$$C[n] = \mathcal{F}^{-1} \{\log(|S_n[k]|)\}$$

gdzie $S_n[k]$ to widmo amplitudowe ramki n , a \mathcal{F}^{-1} to odwrotna transformata Fouriera.

3.11 Metoda compute_spectrogram

Tworzy spektrogram — dwuwymiarową reprezentację zmian widma sygnału w czasie. Dla każdej ramki n obliczane jest widmo:

$$\text{Spectrogram}[n, k] = |S_n[k]|$$

3.12 Metoda extract_frame_features

Dzieli cały sygnał na ramki i oblicza dla każdej zestaw cech: głośność, centroid, szerokość pasma, energię w podzakresach, płaskość i szczytowość widma.

3.13 Metoda get_frequencies

Zwraca częstotliwości odpowiadające binom FFT:

$$f_k = \frac{k \cdot sr}{N}, \quad k = 0, 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor$$

gdzie sr to częstotliwość próbkowania, a N — rozmiar transformaty. Pozwala to przekształcić biny widma $S_n[k]$ na odpowiadające im częstotliwości w Hz.

4 Porównanie Wyników

4.1 Parametry w dziedzinie częstotliwości

W kontekście parametrów w dziedzinie częstotliwości porównane zostały próbki głosów damskego i męskiego dla słowa *dziewięć* (Rysunek 4). Głośność była bardziej zróżnicowana w obu próbkach, z innymi punktami akcentu w każdym przypadku. Centroid częstotliwości miał podobny przebieg, ale większe wahania występowały w głosie damskim. Współczynnik crest był bardziej poszarpany w próbce męskiej. Bandwidth i flatness mają podobny kształt, ale w głosie damskim można zauważać większe różnice między kolejnymi ramkami. Wartości ERSB wskazują na wyraźniejsze występowanie ERSB2 i ERSB4 u głosu damskego na początku wypowiedzi, podczas gdy głos męski charakteryzował się silniejszym ERSB3 na początku i pod koniec próbki.



Rysunek 4: Porównanie wykresów parametrów w dziedzinie częstotliwości dla głosów damskego i męskiego dla słowa *dziewięć*. Wiersze 1 i 3 przedstawiają wykresy dla głosu damskego, natomiast wiersze 2 i 4 dla głosu męskiego.

4.2 Ton podstawowy

Porównując nagrania słów *dziewięć* i *chrząszcz* wypowiadane przez lektorów o głosie damskim i męskim, zauważono wyraźne różnice w częstotliwości podstawowej. W obu przypadkach głos kobiecy osiągał wyższe wartości – średnio około 190 Hz, natomiast głos męski oscylował w okolicach 100 Hz. Taka zależność jest zgodna z ogólną tendencją, zgodnie z którą głosy żeńskie z reguły są wyższe, co znajduje odzwierciedlenie w wyższej częstotliwości podstawowej. Dodatkowo, niezależnie od płci mówcy, słowo *dziewięć* charakteryzowało się nieco wyższym tonem podstawowym niż słowo *chrząszcz*, co może wynikać z różnic artykulacyjnych i intonacyjnych między tymi wyrazami. Dla zachowania porównywalności, próbki głosu żeńskiego pochodzą od jednej kobiety, a próbki głosu męskiego od jednego mężczyzny.



Damskie próbki dla słów *dziecięć* i *chrząszcz* odpowiednio.



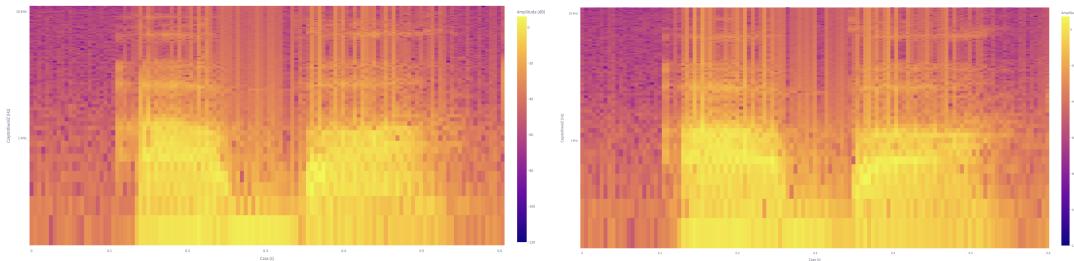
Męskie próbki dla słów *dziecięć* i *chrząszcz* odpowiednio.

Rysunek 5: Porównanie tonów podstawowych dla głosów damskiego i męskiego.

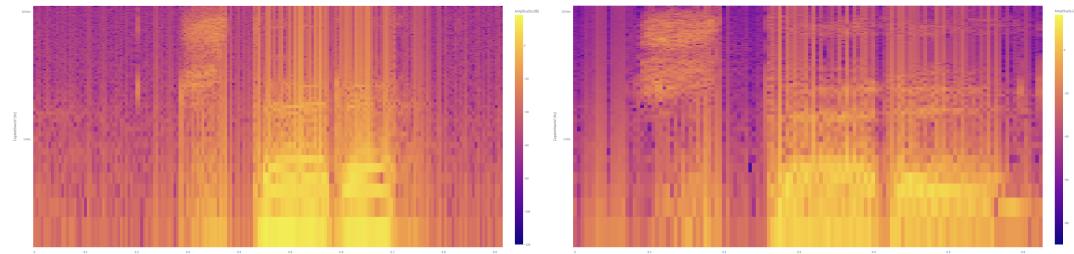
4.3 Formanty

Formanty to pasma częstotliwości, które są wzmacniane podczas przechodzenia dźwięku przez jamę ustną oraz gardło. Pełnią kluczową rolę w różnicowaniu samogłosek – każda samogłoska ma charakterystyczny układ formantów, w szczególności pierwszego (F_1) i drugiego (F_2). Wartości te zależą od ułożenia języka oraz stopnia otwarcia ust w trakcie artykulacji. Formanty można zaobserwować na spektrogramie jako poziome pasma w zakresie od kilkuset do kilku tysięcy herców.

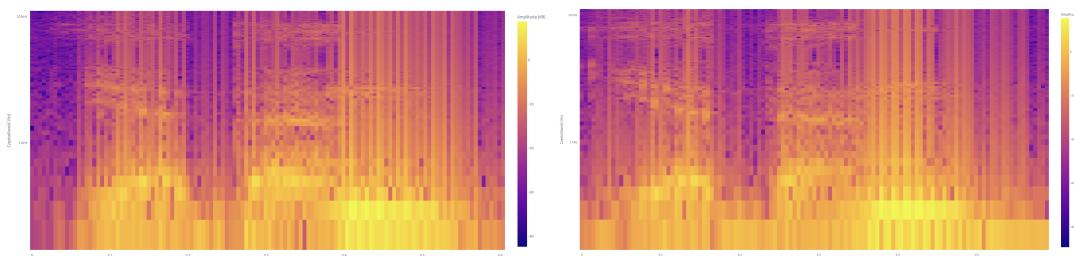
Analizowane słowa pochodzą z nagrań dwóch różnych osób co umożliwia zaobserwowanie wpływu indywidualnych cech głosu na obraz spektrogramu. Jak można zauważyć na spektrogramach (Rysunek 9, 10, 8) dla słów *aba*, *cztery* oraz *jeden*, w miejscach odpowiadających samogłoskom pojawia się wyższa amplituda w zakresie wyższych częstotliwości. Na spektrogramie przejawia się to jako jaśniejsze pasma, które – choć subtelne – układają się poziomo wzdłuż trwania samogłosek. Jest to związane z obecnością formantów charakterystycznych dla samogłosek. W zależności od rodzaju samogłoski, położenie tych pasm (czyli wartość formantów F_1 i F_2) nieco się różni, co odzwierciedla różnice w artykulacji. Dodatkowo, widoczne są pewne różnice w wysokości występowania tych pasm między osobami. Można także zauważyć drobne zmiany położenia formantów nawet dla tej samej samogłoski (np. *a* czy *e*) zarówno między różnymi słowami, jak i w ramach ich powtórzeń w tym samym wyrazie.



Rysunek 6: Spektrogramy dla słowa *aba* pochodzące z dwóch próbek.



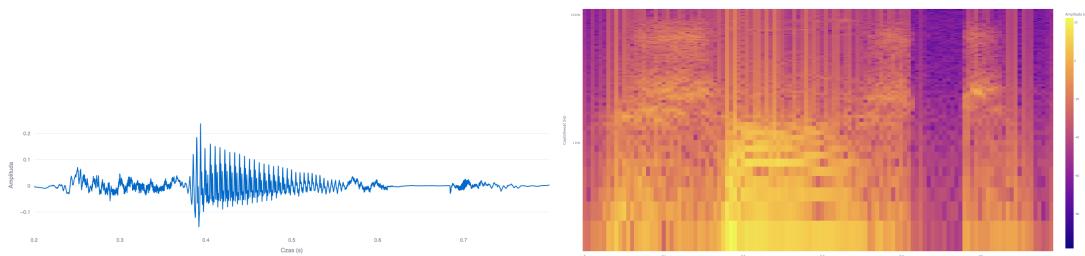
Rysunek 7: Spektrogramy dla słowa *cztery* pochodzące z dwóch próbek.



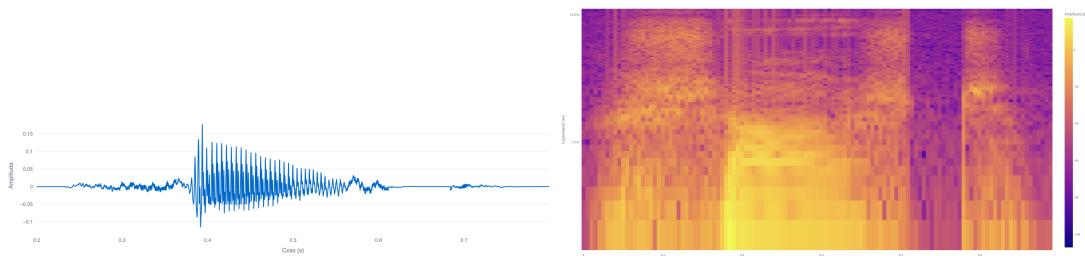
Rysunek 8: Spektrogramy dla słowa *jeden* pochodzące z dwóch próbek.

4.4 Funkcje okna

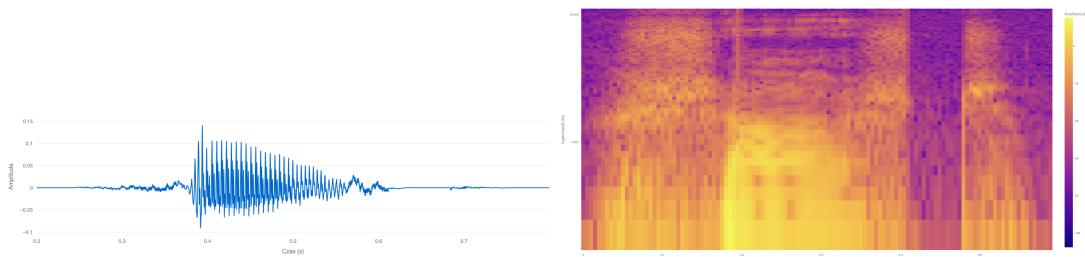
Jak można zauważyć na wykresach (Rysunek 11), zmiana funkcji okna wpływa na wygląd zarówno przebiegu czasowego, jak i spektrogramu – szczególnie w zakresie wygładzenia krawędzi sygnału oraz precyzji odwzorowania częstotliwości. Porównano trzy typy okien: prostokątne, Hamminga oraz Blackmana. Zastosowanie funkcji okna innych niż prostokątna powoduje delikatne rozmycie widma na spektrogramie, jednak poprawia separację między częstotliwościami. W przypadku okna Blackmana można także zaobserwować wyraźniejsze różnice między amplitudami pasm, co wpływa na czytelność prezentacji danych w domenie częstotliwości.



Rysunek 9: Widmo częstotliwościowe i spektrogram dla okna prostokątnego.



Rysunek 10: Widmo częstotliwościowe i spektrogram dla okna Hamminga.



Rysunek 11: Widmo częstotliwościowe i spektrogram dla okna Blackmana.

5 Podsumowanie

Projekt skutecznie realizuje swoje założenia, umożliwiając podstawową analizę sygnałów w dziedzinie częstotliwości. Choć rozwiązań nie jest pozbawione pewnych ograniczeń i nie osiąga perfekcji, działa na satysfakcyjnym poziomie, dostarczając użytkownikowi wartościowych informacji na temat charakterystyki częstotliwościowej analizowanych próbek.

6 Źródła

- dr inż. J. Rafałko, *Wykłady i materiały z Analizy i Przetwarzania dźwięku*, Politechnika Warszawska, 2025
- OpenAI, ChatGPT (GPT-4), <https://openai.com/chatgpt>
- https://mkramarczyk.zut.edu.pl/?cat=M&l=D_07