

Aplikacja do analizy częstotliwościowej sygnałów dźwiękowych

Dokumentacja projektu 2

Jakub Półtorak

11 maja 2025

Spis treści

1 Wstęp i cele projektu	5
1.1 Założenia projektu	5
1.2 Główne cele rozszerzenia	5
2 Analiza częstotliwościowa w teorii	6
2.1 Transformata Fouriera i jej zastosowanie w analizie dźwięku	6
2.1.1 Transformata Fouriera - wprowadzenie	6
2.1.2 Właściwości FFT w analizie dźwięku	6
2.2 Funkcje okienkowe	7
2.2.1 Rola okienkowania w analizie spektralnej	7
2.2.2 Matematyczna definicja funkcji okienkowych	7
2.2.3 Porównanie właściwości funkcji okienkowych	8
2.3 Spektrogram i jego interpretacja	8
2.3.1 Definicja i obliczanie spektrogramu	8
2.3.2 Parametry spektrogramu	9
2.3.3 Interpretacja spektrogramu	10
2.4 Analiza cepstralna i wyznaczanie częstotliwości podstawowej	10
2.4.1 Cepstrum - definicja i własności	10
2.4.2 Zastosowanie cepstrum do wyznaczania częstotliwości podstawowej	10
2.5 Parametry sygnału w dziedzinie częstotliwości	11
2.5.1 Głośność (Volume) w dziedzinie częstotliwości	11

2.5.2	Centroid częstotliwościowy (FC)	12
2.5.3	Szerokość pasma (BW)	12
2.5.4	Energia w pasmach (BE) i stosunek energii w pasmach (BER/ERSB)	12
2.5.5	Płaskość widma (SFM)	13
2.5.6	Współczynnik szczytu widma (SCF)	13
3	Nowe moduły i klasy	14
3.1	frequency_analysis.py	14
3.1.1	Klasa FrequencyAnalysisWindow	14
3.1.2	Klasa CepstrumAnalysisWindow	15
3.2	frequency_features_window.py	15
3.2.1	Klasa FrequencyFeaturesWindow	16
3.3	frequency_features.py	17
3.4	windowing.py	17
3.4.1	Implementowane funkcje okienkowania	17
3.5	cepstrum_analysis.py	17
3.5.1	Implementowane funkcje analizy cepstralnej	17
3.6	Inne	18
4	Eksperymenty	19
4.1	Analiza parametrów częstotliwościowych dla różnych głosów i głosek	19
4.1.1	Materiał badawczy	19
4.1.2	Volume (Głośność) a płeć	22
4.1.3	Centroid częstotliwościowy (FC) a płeć	22
4.1.4	Szerokość pasma (BW) a płeć	22
4.1.5	Płaskość widma (SFM) a płeć	22
4.1.6	Stosunek energii (ERSB) a płeć	23
4.1.7	Samogłoski	23
4.1.8	Spółgłoski nosowe (/m/, /n/)	23
4.1.9	Spółgłoski zwarte (/p/, /t/)	23
4.1.10	Spółgłoski szczelinowe (/j/)	24
4.1.11	Wnioski	24
4.2	Porównanie parametrów w dziedzinie częstotliwości dla różnych głosów i głosek	24
4.3	Analiza tonu podstawowego w głosie męskim	24
4.3.1	Częstotliwość podstawowa	24
4.3.2	Kontury F0	27
4.3.3	Charakterystyka widmowa	30
4.3.4	Wnioski	33

4.4	Analiza tonu podstawowego w głosie żeńskim	34
4.4.1	Częstotliwość podstawowa	34
4.4.2	Kontury F0	37
4.4.3	Charakterystyka widmowa	40
4.4.4	Wnioski	43
4.5	Wnioski dotyczące częstotliwości podstawowej u różnych mów- ców	44
4.6	Formanty w samogłoskach - analiza spektrograficzna	44
4.6.1	Definicja formantów	44
4.6.2	Analiza spektrogramów	45
4.6.3	Wartości formantów dla poszczególnych samogłosek . .	48
4.6.4	Zależności między wartościami formantów a cechami artykulacyjnymi	48
4.6.5	Dynamika formantów	49
4.6.6	Wnioski	49
4.7	Analiza zmienności formantów w różnych realizacjach tego sa- mego fonemu samogłoskowego	49
4.7.1	Metodologia	49
4.7.2	Wyniki analizy	50
4.7.3	Wnioski	53
4.8	Analiza porównawcza formantów samogłosek u różnych osób .	54
4.8.1	Metodologia	54
4.8.2	Analiza spektrogramów	55
4.8.3	Analiza centroidu częstotliwościowego	56
4.8.4	Analiza stosunków energii w pasmach częstotliwości . .	58
4.8.5	Wnioski	59
4.9	Różnice między samogłoskami a spółgłoskami w analizie wid- mowej	59
4.9.1	Charakterystyczne cechy widmowe samogłosek	59
4.9.2	Charakterystyczne cechy widmowe spółgłosek	60
4.9.3	Przykłady w analizowanym materiale	60
4.10	Analiza wpływu funkcji okienkowych na widmo częstotliwo- ściowe i spektrogram	61
4.10.1	Wprowadzenie	61
4.10.2	Widmo okna prostokątnego	62
4.10.3	Widmo okna trójkątnego (Bartletta)	63
4.10.4	Widmo okna Hamminga	64
4.10.5	Widmo okna Hanna	65
4.10.6	Widmo okna Blackmana	65
4.10.7	Spektrogram okna prostokątnego	66
4.10.8	Spektrogram okna trójkątnego (Bartletta)	67

4.10.9 Spektrogram okna Hamminga	68
4.10.10 Spektrogram okna Hanna	69
4.10.11 Spektrogram okna Blackmana	70
4.10.12 Wpływ funkcji okienkowych na parametry częstotliwo- ściowe	70
4.10.13 Wnioski	71
5 Podsumowanie i wnioski	71
5.1 Podsumowanie projektu	71
5.2 Kluczowe wnioski z przeprowadzonych eksperymentów	72
5.3 Kierunki dalszego rozwoju	72

1 Wstęp i cele projektu

1.1 Założenia projektu

Dokumentacja przedstawia rozszerzenie aplikacji z poprzedniego projektu, która umożliwiała odtwarzanie i podstawową analizę plików dźwiękowych w formacie WAV. W tej części dodano funkcjonalności związane z analizą częstotliwościową sygnałów dźwiękowych, zgodnie z wymaganiami postawionymi w zadaniu projektowym.

1.2 Główne cele rozszerzenia

1. **Implementacja analizy widmowej** - umożliwienie analizy sygnału w dziedzinie częstotliwości za pomocą Szybkiej Transformaty Fouriera (FFT)
2. **Zastosowanie różnych funkcji okienkowych** - implementacja okien: prostokątnego, trójkątnego, Hamminga, Hanna i Blackmana oraz wizualizacja ich wpływu na widmo częstotliwościowe
3. **Implementacja spektrogramu** - wizualizacja zmian częstotliwości w czasie z konfigurowalnymi parametrami (długość ramki, stopień nakładania, typ okna)
4. **Analiza cepstralna** - implementacja metod wyznaczania częstotliwości podstawowej (F_0) przy użyciu cepstrum
5. **Parametryzacja w dziedzinie częstotliwości** - obliczanie parametrów takich jak centroid częstotliwościowy, szerokość pasma, płaskość widma i inne
6. **Wizualizacja parametrów na poziomie ramki** - możliwość obserwacji zmian parametrów częstotliwościowych w czasie dla poszczególnych ramek sygnału

2 Analiza częstotliwościowa w teorii

2.1 Transformata Fouriera i jej zastosowanie w analizie dźwięku

2.1.1 Transformata Fouriera - wprowadzenie

Transformata Fouriera jest fundamentalnym narzędziem w analizie sygnałów, umożliwiającym dekompozycję sygnału na jego składowe częstotliwościowe. Dla sygnału ciągłego $x(t)$, transformata Fouriera jest zdefiniowana jako:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (1)$$

W przypadku sygnałów dyskretnych, które są przetwarzane w aplikacji, stosuje się Dyskretną Transformatę Fouriera (DFT):

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N}, \quad k = 0, 1, \dots, N-1 \quad (2)$$

gdzie $x[n]$ to próbki sygnału, N to liczba próbek, a $X[k]$ to zespolone współczynniki DFT reprezentujące amplitudę i fazę składowych częstotliwościowych.

W aplikacji zastosowano Szybką Transformatę Fouriera (FFT), która jest efektywnym algorytmem obliczania DFT, redukującym złożoność obliczeniową.

2.1.2 Właściwości FFT w analizie dźwięku

W analizie dźwięku, FFT dostarcza informacji o amplitudzie i fazie poszczególnych składowych częstotliwościowych sygnału. Dla sygnału próbkowanego z częstotliwością f_s , FFT daje widmo częstotliwościowe od 0 do $f_s/2$ Hz

Właściwości FFT istotne w kontekście analizy dźwięku:

- **Rozdzielczość częstotliwościowa** - zależy od długości analizowanej ramki: $\Delta f = f_s/N$, gdzie N to długość ramki
- **Symetryczność widma** - dla sygnałów rzeczywistych widmo jest symetryczne względem częstotliwości Nyquista, dlatego w implementacji używamy np.`fft.rfft()`, która zwraca tylko połowę współczynników (dla częstotliwości nieujemnych)
- **Przeciek widma (spectral leakage)** - zjawisko pojawiające się, gdy analizowana ramka nie zawiera całkowitej liczby okresów sygnału, prowadzące do rozmycia energii na sąsiednie częstotliwości

- **Efekt okna czasowego** - zastosowanie okna innego niż prostokątne modyfikuje widmo, zazwyczaj redukując przeciek widma kosztem rozdzielczości częstotliwościowej

2.2 Funkcje okienkowe

2.2.1 Rola okienkowania w analizie spektralnej

Okienkowanie to technika stosowana w analizie spektralnej w celu redukcji efektów związanych z analizą skończonych fragmentów sygnału. Gdy stosujemy FFT do fragmentu sygnału, zakładamy niejawnie, że ten fragment jest jednym okresem sygnału nieskońzonego (sygnał jest periodyzowany). Jeśli tak nie jest, pojawia się nieciągłość na brzegach analizowanego fragmentu, co prowadzi do przecieku widma.

Funkcje okienkowe są stosowane do wygładzania tych nieciągłości poprzez stopniowe zmniejszanie amplitudy sygnału na brzegach analizowanej ramki. Każda funkcja okienkowa oferuje inny kompromis między rozdzielczością częstotliwościową a poziomem przecieku widma.

2.2.2 Matematyczna definicja funkcji okienkowych

Sygnal okienkowany $y[n]$ jest iloczynem oryginalnego sygnału $x[n]$ i funkcji okienkowej $w[n]$:

$$y[n] = x[n] \cdot w[n], \quad n = 0, 1, \dots, N - 1 \quad (3)$$

W projekcie zaimplementowano następujące funkcje okienkowe:

- **Okno prostokątne (Rectangular):**

$$w[n] = 1, \quad n = 0, 1, \dots, N - 1 \quad (4)$$

- **Okno trójkątne (Triangular/Bartlett):**

$$w[n] = 1 - \left| \frac{2n}{N-1} - 1 \right|, \quad n = 0, 1, \dots, N - 1 \quad (5)$$

- **Okno Hamminga (Hamming):**

$$w[n] = 0.54 - 0.46 \cos \left(\frac{2\pi n}{N-1} \right), \quad n = 0, 1, \dots, N - 1 \quad (6)$$

- **Okno Hanna (Hann):**

$$w[n] = 0.5 - 0.5 \cos \left(\frac{2\pi n}{N-1} \right), \quad n = 0, 1, \dots, N - 1 \quad (7)$$

- **Okno Blackmana (Blackman):**

$$w[n] = 0.42 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right) + 0.08 \cos\left(\frac{4\pi n}{N-1}\right), \quad n = 0, 1, \dots, N-1 \quad (8)$$

2.2.3 Porównanie właściwości funkcji okienkowych

Każda z funkcji okienkowych ma charakterystyczne właściwości, które wpływają na jakość analizy spektralnej:

Tabela 1: Porównanie właściwości funkcji okienkowych

Okno	Szerokość listka głównego	Najwyższy listek boczny [dB]	Spadek listków bocznych [dB/okt]
Prostokątne	2	-13	-6
Trójkątne	4	-27	-12
Hanna	4	-32	-18
Hamminga	4	-43	-6
Blackmana	6	-58	-18

- **Szerokość listka głównego** - określa rozdzielcość częstotliwościową; węższy listek główny oznacza lepszą rozdzielcość
- **Najwyższy listek boczny** - określa maksymalny poziom przecieku widma; niższy poziom listków bocznych oznacza mniejszy przeciek
- **Spadek listków bocznych** - określa tempo zaniku amplitudy listków bocznych wraz z oddalaniem się od listka głównego

Okno prostokątne oferuje najlepszą rozdzielcość częstotliwościową, ale najgorsze właściwości w zakresie przecieku widma. Z kolei okno Blackmana oferuje najlepszą redukcję przecieku widma, ale kosztem znacznej utraty rozdzielcości częstotliwościowej. Okna Hamminga i Hanna stanowią dobry kompromis między tymi ekstremami.

2.3 Spektrogram i jego interpretacja

2.3.1 Definicja i obliczanie spektrogramu

Spektrogram to dwuwymiarowa reprezentacja sygnału w dziedzinie czas-częstotliwość, umożliwiająca obserwację zmian widma częstotliwościowego

w czasie. Jest szczególnie użyteczny w analizie sygnałów niestacjonarnych, takich jak mowa czy muzyka.

Spektrogram jest obliczany poprzez podział sygnału na krótkie, nakładające się ramki, zastosowanie funkcji okienkowej do każdej ramki, a następnie obliczenie FFT dla każdej okienkowanej ramki. Wartość bezwzględna wyników FFT (zazwyczaj w skali decybelowej) jest przedstawiana jako kolorowa mapa, gdzie czas jest reprezentowany na osi poziomej, częstotliwość na osi pionowej, a kolor odpowiada amplitudzie.

Matematycznie, krótkoczasowa transformata Fouriera (STFT), która jest podstawą spektrogramu, jest definiowana jako:

$$X(m, k) = \sum_{n=0}^{N-1} x[n + mH] \cdot w[n] \cdot e^{-j2\pi kn/N} \quad (9)$$

gdzie m to indeks ramki, H to przesunięcie (hop) między kolejnymi ramkami, $w[n]$ to funkcja okienkowa, a $X(m, k)$ to zespolone współczynniki STFT dla ramki m i częstotliwości k .

Spektrogram to kwadrat wartości bezwzględnej STFT:

$$S(m, k) = |X(m, k)|^2 \quad (10)$$

W implementacji często stosuje się skalę decybelową:

$$S_{dB}(m, k) = 20 \log_{10}(|X(m, k)| + \epsilon) \quad (11)$$

gdzie ϵ to mała wartość dodana, aby uniknąć logarytmowania zera.

2.3.2 Parametry spektrogramu

Kluczowe parametry wpływające na jakość spektrogramu:

- **Długość ramki (frame_length)** - determinuje rozdzielczość częstotliwościową; dłuższe ramki dają lepszą rozdzielczość częstotliwościową, ale gorszą rozdzielczość czasową
- **Przesunięcie między ramkami (hop_length)** - determinuje rozdzielczość czasową; mniejsze przesunięcie daje lepszą rozdzielczość czasową, ale zwiększa redundancję danych
- **Nakładanie się ramek (overlap)** - wyrażone jako procent długości ramki; typowo 50-75%, co odpowiada przesunięciu $hop_length = frame_length * (1 - overlap)$
- **Funkcja okienkowa (window_type)** - wpływa na wycieki widma oraz rozdzielczość częstotliwościową

2.3.3 Interpretacja spektrogramu

Spektrogram jest narzędziem do analizy sygnałów dźwiękowych, umożliwiającym obserwację:

- **Formantów** - wzmacnione obszary częstotliwości w sygnałach mowy, widoczne jako ciemniejsze pasma poziome
- **Tonów harmonicznych** - regularne, równo rozstawione linie poziome, będące wielokrotnościami częstotliwości podstawowej
- **Przejść między głoskami** - widoczne jako zmiany w strukturze formantów i energii
- **Szumów i trzasków** - widoczne jako pionowe linie lub obszary o rozproszonym rozkładzie energii
- **Zmian tempa** - widoczne jako rozciągnięcia lub kompresje wzorców w czasie

2.4 Analiza cepstralna i wyznaczanie częstotliwości podstawowej

2.4.1 Cepstrum - definicja i własności

Cepstrum to transformata odwrotna logarytmu widma amplitudowego sygnału. Cepstrum jest szczególnie użyteczne w analizie sygnałów, które mają naturę periodyczną lub quasi-periodyczną, takich jak sygnały mowy.

Matematycznie, rzeczywiste cepstrum jest definiowane jako:

$$c[n] = \text{IDFT}(\log(|X[k]|)) \quad (12)$$

gdzie $X[k]$ to transformata Fouriera sygnału, $|X[k]|$ to jej wartość bezwzględna (widmo amplitudowe), a IDFT to odwrotna dyskretna transformata Fouriera.

W dziedzinie cepstrum, czas jest zastąpiony przez tzw. "kwefrencję" (quefrency), która jest mierzona w jednostkach czasu (np. sekundach). Kwefrencja odpowiada opóźnieniu lub okresowi w oryginalnym sygnale.

2.4.2 Zastosowanie cepstrum do wyznaczania częstotliwości podstawowej

Cepstrum jest szczególnie przydatne do wyznaczania częstotliwości podstawowej (F_0) w sygnałach mowy. W sygnale mowy, częstotliwość podstawowa

powoduje pojawienie się w widmie regularnych, równo rozstawionych harmonicznych będących jej wielokrotnościami. Ta okresowość w widmie pojawia się jako pojedynczy pik w cepstrum na kwefrencji odpowiadającej okresowi podstawowemu.

Algorytm wyznaczania F0 za pomocą cepstrum:

1. Podziel sygnał mowy na ramki i zastosuj funkcję okienkową (np. Hamminga)
2. Oblicz FFT dla każdej ramki
3. Oblicz logarytm wartości bezwzględnej FFT (logarytmiczne widmo amplitudowe)
4. Oblicz odwrotną FFT logarytmicznego widma amplitudowego (cepstrum)
5. Znajdź pik w cepstrum w zakresie kwefrencji odpowiadającym oczekiwaniu zakresowi częstotliwości podstawowej (np. 50-500 Hz dla głosu ludzkiego)
6. Odwróć kwefrencję, aby uzyskać częstotliwość podstawową: $F0 = 1/\text{quefrency_peak}$

Zalety analizy cepstralnej w porównaniu do innych metod wyznaczania F0:

- **Odporność na harmoniczne** - metoda bazuje na okresowości harmonicznych, a nie na ich amplitudzie
- **Niezależność od fazy** - użycie widma amplitudowego eliminuje wpływ fazy
- **Separacja składowych** - w cepstrum składowe związane z częstotliwością podstawową są dobrze separowane od składowych związanych z formantami

2.5 Parametry sygnału w dziedzinie częstotliwości

2.5.1 Głośność (Volume) w dziedzinie częstotliwości

Głośność w dziedzinie częstotliwości jest miarą całkowitej energii spektralnej sygnału. Jest obliczana jako średnia kwadratów wartości współczynników widma:

$$\text{Vol}(n) = \frac{1}{N} \sum_{k=0}^{N-1} |S_n(k)|^2 \quad (13)$$

gdzie $S_n(k)$ to współczynniki widma dla ramki n , a N to liczba współczynników widma.

2.5.2 Centroid częstotliwościowy (FC)

Centroid częstotliwościowy reprezentuje "środek ciężkości" widma, czyli średnią ważoną częstotliwość, gdzie wagami są amplitudy widma. Jest związany z postrzeganą "jasnością" dźwięku:

$$\text{FC}(n) = \frac{\sum_{\omega} \omega \cdot |S_n(\omega)|}{\sum_{\omega} |S_n(\omega)|} \quad (14)$$

gdzie ω to częstotliwość, a $S_n(\omega)$ to widmo dla ramki n .

W implementacji dyskretnej:

$$\text{FC}(n) = \frac{\sum_{k=0}^{N/2} f[k] \cdot |S_n(k)|}{\sum_{k=0}^{N/2} |S_n(k)|} \quad (15)$$

gdzie $f[k]$ to częstotliwość odpowiadająca współczynnikowi widma k .

2.5.3 Szerokość pasma (BW)

Szerokość pasma to miara rozrzutu energii widma wokół centroidu częstotliwościowego. Jest definiowana jako odchylenie standardowe widma:

$$\text{BW}^2(n) = \frac{\sum_{\omega} (\omega - \text{FC}(n))^2 \cdot |S_n(\omega)|^2}{\sum_{\omega} |S_n(\omega)|^2} \quad (16)$$

W implementacji dyskretniej:

$$\text{BW}^2(n) = \frac{\sum_{k=0}^{N/2} (f[k] - \text{FC}(n))^2 \cdot |S_n(k)|^2}{\sum_{k=0}^{N/2} |S_n(k)|^2} \quad (17)$$

2.5.4 Energia w pasmach (BE) i stosunek energii w pasmach (BE-R/ERSB)

Energia w paśmie to miara energii sygnału w określonym zakresie częstotliwości:

$$\text{BE}_{[f_0, f_1]}(t) = \frac{\int_{f_0}^{f_1} |S_t(f)|^2 df}{\int w(\tau) d\tau} \quad (18)$$

gdzie f_0 i f_1 to granice pasma częstotliwości, S_t to widmo w czasie t , a w to funkcja okienkowa.

Stosunek energii w paśmie (BER/ERSB) to stosunek energii w danym paśmie do całkowitej energii sygnału:

$$\text{ERSB}_{[f_0, f_1]}(t) = \frac{\text{BE}_{[f_0, f_1]}(t)}{\text{Vol}(t)} \quad (19)$$

W analizie mowy i muzyki często używa się standardowych pasm częstotliwości odpowiadających krytycznym pasmom w układzie słuchowym człowieka:

- ERSB1: 0-630 Hz
- ERSB2: 630-1720 Hz
- ERSB3: 1720-4400 Hz
- ERSB4: 4400-Nyquist Hz

2.5.5 Płaskość widma (SFM)

Płaskość widma to miara regularności rozkładu energii w widmie. Jest obliczana jako stosunek średniej geometrycznej do średniej arytmetycznej mocy widma:

$$\text{SFM}(b, n) = \frac{\left(\prod_{i=il(b)}^{ih(b)} |S_n(i)|^2\right)^{\frac{1}{ih(b)-il(b)+1}}}{\frac{1}{ih(b)-il(b)+1} \sum_{i=il(b)}^{ih(b)} |S_n(i)|^2} \quad (20)$$

gdzie $il(b)$ i $ih(b)$ to dolny i górny indeks pasma częstotliwości b .

SFM przyjmuje wartości od 0 do 1, gdzie 1 oznacza płaskie widmo (biały szum), a wartości bliskie 0 oznaczają widmo tonalne (z dominującymi składowymi sinusoidalnymi).

2.5.6 Współczynnik szczytu widma (SCF)

Współczynnik szczytu widma to stosunek maksymalnej wartości widma do jego wartości średniej:

$$\text{SCF}(b, n) = \frac{\max_i(|S_n(i)|^2)}{\frac{1}{ih(b)-il(b)+1} \sum_{i=il(b)}^{ih(b)} |S_n(i)|^2} \quad (21)$$

SCF jest miarą "tonalności" sygnału - wysokie wartości wskazują na obecność silnych składowych sinusoidalnych, podczas gdy niskie wartości sugerują sygnał o charakterze szumowym.

3 Nowe moduły i klasy

W ramach rozszerzenia aplikacji o funkcjonalności analizy częstotliwościowej, dodano szereg nowych modułów i klas, które współpracują z istniejącą infrastrukturą projektu 1. Poniżej przedstawiono szczegółowy opis nowych komponentów.

3.1 frequency_analysis.py

Zaimplementowano moduł analizy częstotliwościowej sygnałów dźwiękowych, który składa się z dwóch głównych klas: `FrequencyAnalysisWindow` i `CepstrumAnalysisWindow`. Obie klasy współpracują z wcześniej utworzonymi komponentami oraz zawierają specjalistyczne funkcje do przetwarzania sygnałów.

3.1.1 Klasa FrequencyAnalysisWindow

Klasa ta implementuje okno GUI dedykowane do analizy częstotliwościowej sygnału audio. Umożliwia wizualizację sygnału w dziedzinie czasu, analizę FFT oraz generowanie spektrogramu z konfigurowalnymi parametrami.

Najważniejsze metody klasy:

- `__init__(self, parent, audio_app)` - inicjalizuje okno analizy częstotliwościowej, tworzy interfejs użytkownika
- `create_control_panel(self)` - tworzy panel kontrolny z opcjami analizy (wybór ramki, funkcji okienkowej, parametrów spektrogramu)
- `create_plot_area(self)` - przygotowuje obszar wykresów z zakładkami dla różnych wizualizacji.
- `update_plots(self)` - aktualizuje wszystkie wykresy na podstawie bieżących ustawień
- `plot_time_domain(self)` - wizualizuje sygnał w dziedzinie czasu z zastosowaniem wybranej funkcji okienkowej
- `plot_frequency_domain(self)` - przedstawia reprezentację w dziedzinie częstotliwości przy użyciu FFT
- `plot_spectrogram(self)` - wizualizuje spektrogram sygnału
- `compute_spectrogram(self, signal, sample_rate)` - oblicza spektrogram sygnału z konfigurowalnymi parametrami (długość ramki, nakładanie)

- `update_stats(self)` - oblicza i wyświetla parametry częstotliwościowe dla analizowanej ramki sygnału

3.1.2 Klasa CepstrumAnalysisWindow

Klasa implementująca okno GUI dla analizy cepstralnej, która pozwala na estymację i śledzenie częstotliwości podstawowej (F0) w sygnałach mowy.

Najważniejsze metody klasy:

- `__init__(self, parent, audio_app)` - inicjalizuje okno analizy cepstralnej
- `create_control_panel(self)` - tworzy panel kontrolny z opcjami analizy cepstralnej (zakres F0, wybór ramki)
- `update_plots(self)` - aktualizuje wszystkie wykresy analizy cepstralnej
- `plot_log_spectrum(self)` - wizualizuje logarytmiczne widmo amplitudowe
- `plot_cepstrum(self)` - wizualizuje cepstrum z zaznaczonym pikiem częstotliwości podstawowej
- `plot_f0_over_time(self)` - śledzi i wizualizuje zmiany częstotliwości podstawowej (F0) w czasie
- `update_f0_info(self, f0_value)` - aktualizuje panel informacyjny o wartości wykrytej częstotliwości podstawowej

Zaimplementowany moduł pozwala na kompleksową analizę sygnałów dźwiękowych, umożliwiając obserwację charakterystyk zarówno w dziedzinie czasu jak i częstotliwości. Dodatkowo zapewnia funkcjonalność analizy cepstralnej dla sygnałów mowy ludzkiej, umożliwiającą identyfikację częstotliwości podstawowej i śledzenie jej zmian w czasie.

3.2 frequency_features_window.py

Moduł składa się z jednej klasy `FrequencyFeaturesWindow`.

3.2.1 Klasa FrequencyFeaturesWindow

Klasa ta implementuje interaktywne okno do wizualizacji parametrów częstotliwościowych sygnału dźwiękowego w funkcji czasu. Umożliwia obliczanie i wyświetlanie różnych parametrów sygnału w dziedzinie częstotliwości dla całego pliku audio, przedstawiając ich zmiany w czasie.

- `__init__(self, parent, audio_data, sample_rate, frame_size=256, window_type='hamming', overlap=0.5, frame_step=None)` - inicjalizuje okno i parametry analizy, takie jak rozmiar ramki, typ okna i stopień nakładania ramek
- `create_control_panel(self)` - tworzy panel kontrolny z checkboxami umożliwiającymi wybór parametrów do wyświetlenia (głośność, centroid częstotliwościowy, szerokość pasma, stosunki energii w pasmach, płaskość widma, współczynnik szczytu)
- `create_plot_area(self)` - inicjalizuje obszar wykresów dla wizualizacji parametrów częstotliwościowych
- `compute_all_features(self)` - kluczowa metoda obliczająca wartości wszystkich parametrów częstotliwościowych dla każdej ramki sygnału, wykonuje podział sygnału na nakładające się ramki, stosuje funkcję okienkową, oblicza FFT i na jej podstawie wyznacza parametry
- `update_plots(self)` - aktualizuje wykresy na podstawie wybranych przez użytkownika parametrów, tworzy odpowiednią liczbę podwykresów dla zaznaczonych opcji i wypełnia je danymi z analizy

Klasa ta umożliwia obserwację zmian następujących parametrów w dziedzinie częstotliwości:

- **Volume (Głośność)**
- **Centroid częstotliwościowy (FC)**
- **Szerokość pasma (BW)**
- **Stosunki energii w pasmach (ERSB)**
- **Płaskość widma (SFM)**
- **Współczynnik szczytu widma (SCF)**

3.3 frequency_features.py

Moduł ten zawiera implementację wszystkich wcześniej opisanych parametrów w dziedzinie częstotliwości, które są podlegane analizie i wizualizacji w module `frequency_features_window.py`

3.4 windowing.py

Moduł implementuje funkcje do aplikowania różnych funkcji okienkowych na ramki sygnału.

3.4.1 Implementowane funkcje okienkowania

- `apply_window(frame, window_type)` - główna funkcja modułu, która aplikuje wybraną funkcję okienkową do ramki sygnału.

Obsługuje pięć różnych typów okien:

- **Prostokątne** ('rectangular')
- **Trójkątne Bartletta** ('triangular') - implementowane za pomocą `np.bartlett`
- **Hamminga** ('hamming') - implementowane za pomocą `np.hamming`
- **Hanna** ('hann') - implementowane za pomocą `np.hanning`
- **Blackmana** ('blackman') - implementowane za pomocą `np.blackman`

- `get_window_type_name(window_type)` - funkcja pomocnicza, która mapuje techniczne nazwy funkcji okienkowych używane w kodzie na ich polskie odpowiedniki. Jest wykorzystywana w interfejsie użytkownika do wyświetlania czytelnych nazw.

3.5 cepstrum_analysis.py

Moduł implementuje funkcje do analizy cepstralnej sygnału dźwiękowego, umożliwiając estymację częstotliwości podstawowej (F0) głosu.

3.5.1 Implementowane funkcje analizy cepstralnej

- `compute_cepstrum(frame, sample_rate, window_type='hamming')` - funkcja obliczająca rzeczywiste cepstrum sygnału:
 - Przyjmuje ramkę sygnału, częstotliwość próbkowania oraz opcjonalnie typ funkcji okienkowej

- Stosuje wybraną funkcję okienkową do ramki wykorzystując funkcję `apply_window`
 - Oblicza FFT ramki za pomocą `np.fft.rfft`
 - Wyznacza logarytm widma amplitudowego z dodanym małym offsetem (1e-10) dla uniknięcia logarytmowania zera
 - Oblicza odwrotną FFT logarytmicznego widma amplitudowego, otrzymując rzeczywiste cepstrum
 - Generuje oś kwefrencji (wymiar czasu w cepstrum) w sekundach
 - Zwraca trójkę: (cepstrum, kwefrencia, logarytmiczne widmo)
- `estimate_f0_from_cepstrum(cepstrum, quefrency, min_f0=50, max_f0=500)`
 - funkcja estymująca częstotliwość podstawową na podstawie cepstrum:
 - Przyjmuje cepstrum, oś kwefrencji oraz zakres spodziewanej częstotliwości podstawowej (domyślnie 50-500 Hz)
 - Konwertuje zakres częstotliwości [min_f0, max_f0] na odpowiedni zakres kwefrencji [max_quefrency, min_quefrency], wykorzystując zależność $f = \frac{1}{q}$, gdzie f to częstotliwość, a q to kwefrencia
 - Znajduje indeksy w tablicy kwefrencji odpowiadające wyliczonym wartościami granicznym
 - Implementuje zabezpieczenie przed pustym zakresem indeksów
 - Wyszukuje pozycję maksymalnego piku w cepstrum w określonym zakresie kwefrencji
 - Konwertuje znalezioną kwefrencję na częstotliwość podstawową zgodnie z zależnością $F0 = \frac{1}{q_{peak}}$
 - Zwraca parę: (estymowana częstotliwość podstawowa, indeks piku)

3.6 Inne

Wcześniej powstałe moduły i klasy zostały odpowiednio zaaktualizowane i przerobione w taki sposób, aby nowe elementy bezproblemowo zostały w nie wkomponowane. Pierwotny styl aplikacji został przeniesiony na nowe komponenty, które zostały także wzbogacone o nowe, drobne aktualizacje designowe.

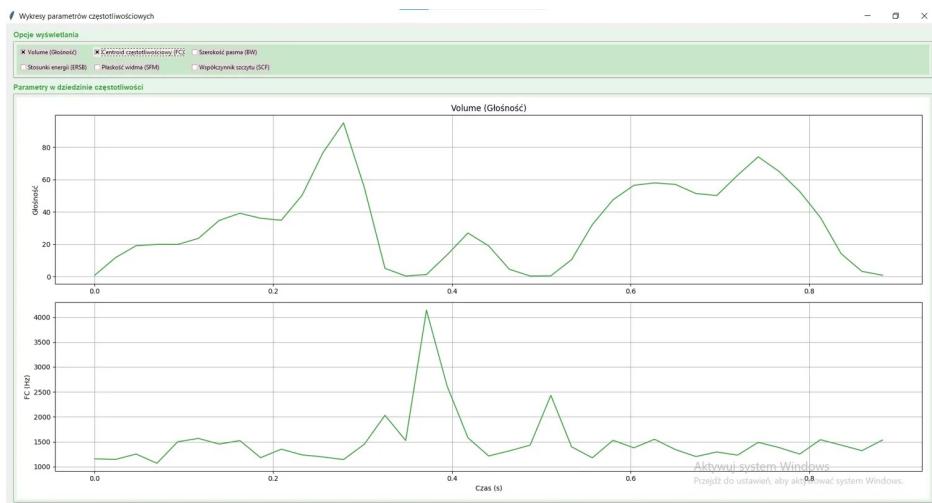
4 Eksperymenty

4.1 Analiza parametrów częstotliwościowych dla różnych głosów i głosek

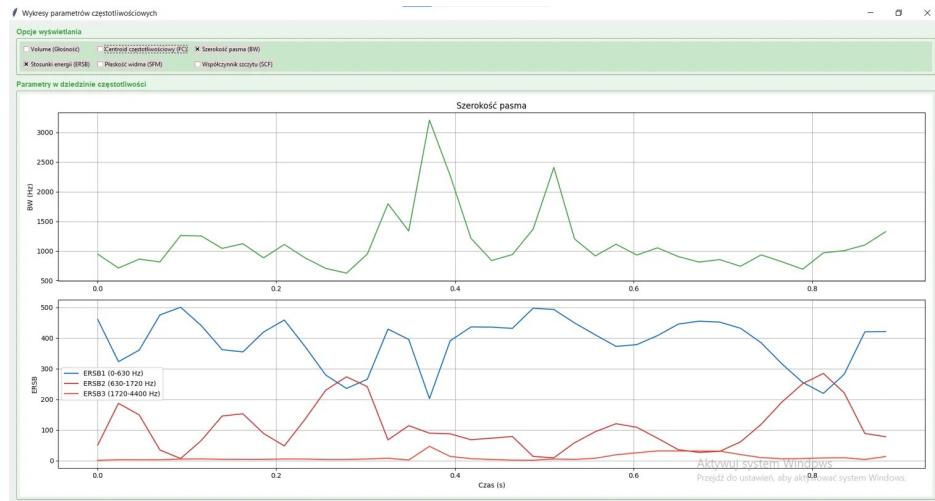
4.1.1 Materiał badawczy

Analizę przeprowadzono na podstawie nagrania słowa „onomatopeja” wymówionego przez głos męski i żeński. Wykresy parametrów częstotliwościowych zostały przedstawione na rysunkach:

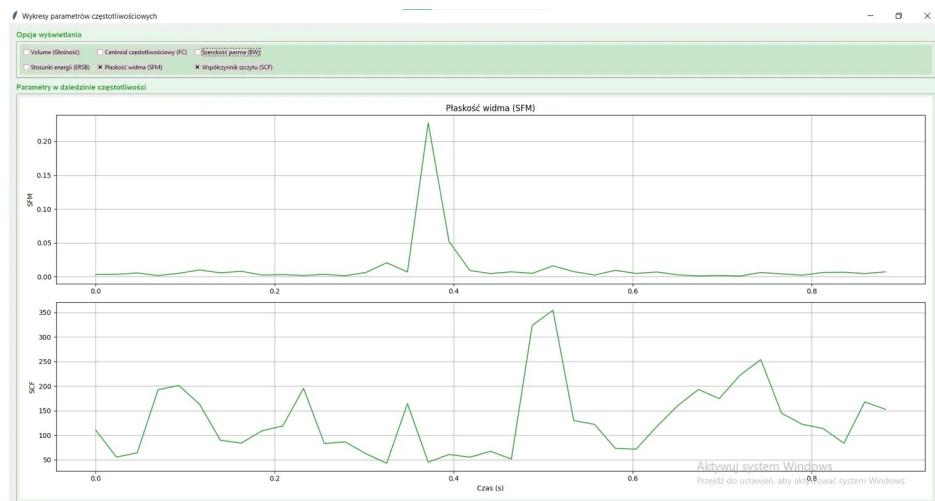
- Rys. 1-3: Głos męski (onomatopeja_man_1.jpg, onomatopeja_man_2.jpg, onomatopeja_man_3.jpg)
- Rys. 4-6: Głos żeński (onomatopeja_woman_1.jpg, onomatopeja_woman_2.jpg, onomatopeja_woman_3.jpg)



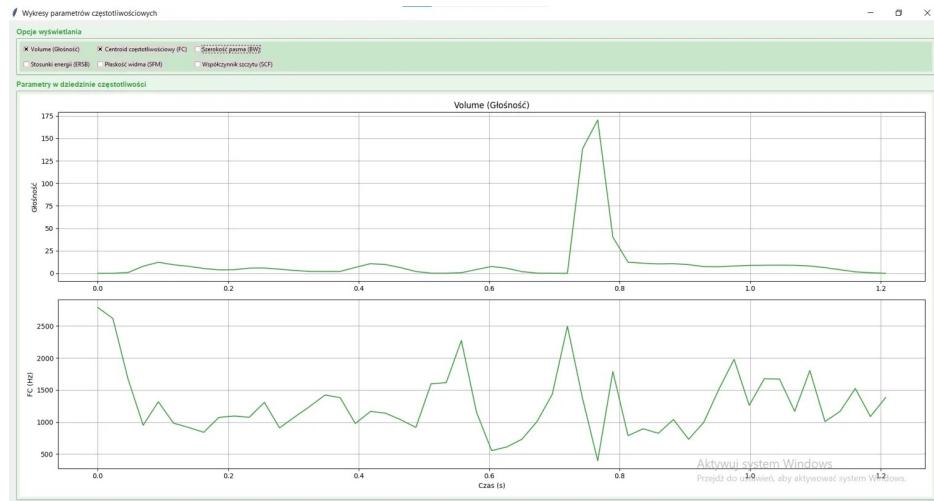
Rysunek 1: Parametry Volume i FC dla głosu męskiego



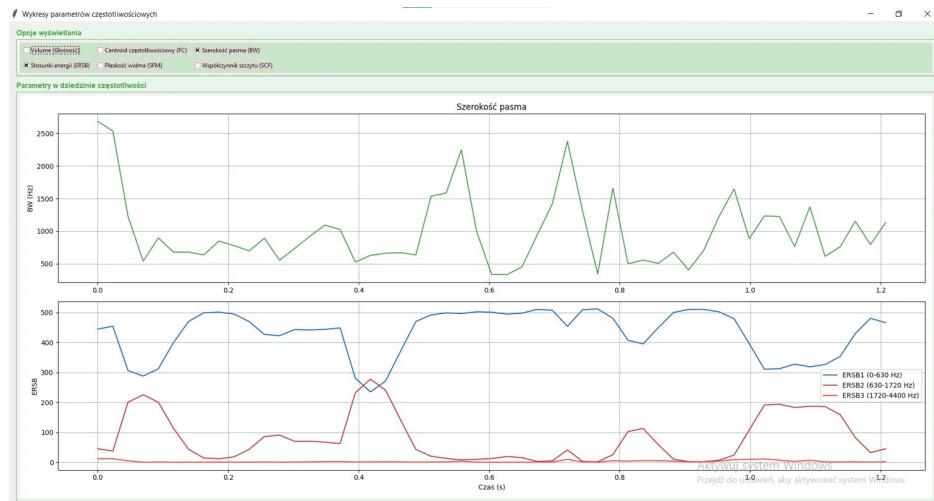
Rysunek 2: Parametry BW i ERSB dla głosu męskiego



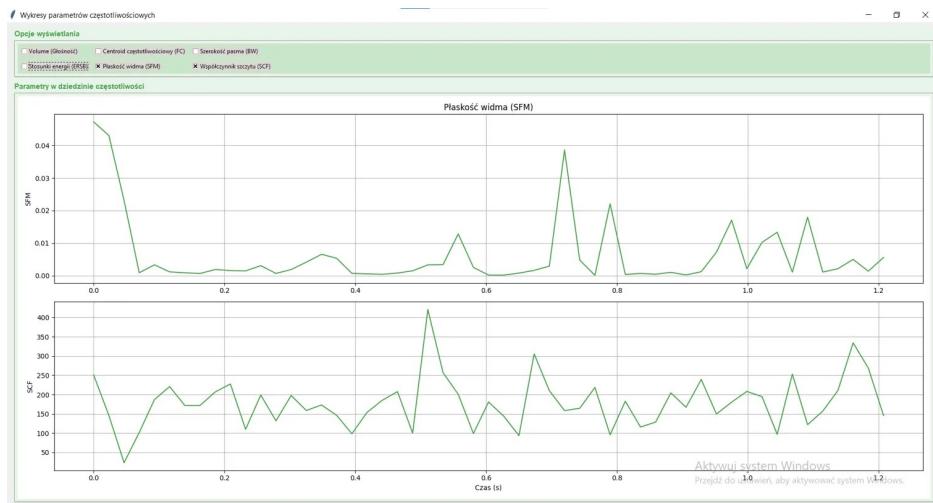
Rysunek 3: Parametry SFM i SCF dla głosu męskiego



Rysunek 4: Parametry Volume i FC dla głosu żeńskiego



Rysunek 5: Parametry BW i ERSB dla głosu żeńskiego



Rysunek 6: Parametry SFM i SCF dla głosu żeńskiego

4.1.2 Volume (Głośność) a płeć

- Głos męski: Na rys. 1 widoczna wyższa amplituda bazowa z pikiem przy 0,35s (około 90 jednostek)
- Głos żeński: Rys. 4 pokazuje mniejszą amplitudę bazową, ale wyraźniejszy pik przy 0,75s (około 170 jednostek)

4.1.3 Centroid częstotliwościowy (FC) a płeć

- Głos męski: Rys. 1 pokazuje FC w zakresie 1000-2000 Hz, z pikiem do 4100 Hz przy 0,35s
- Głos żeński: Na rys. 4 FC utrzymuje się w wyższym zakresie 1000-2500 Hz z pikami do 2500 Hz

4.1.4 Szerokość pasma (BW) a płeć

- Głos męski: Rys. 2 przedstawia szerokość pasma z nagłymi zmianami i pikiem około 3200 Hz
- Głos żeński: Rys. 5 pokazuje większą stabilność BW z wartościami 1000-2400 Hz

4.1.5 Płaskość widma (SFM) a płeć

- Głos męski: Na rys. 3 SFM osiąga szczytową wartość 0,23 przy 0,35s

- Głos żeński: Rys. 6 pokazuje niższe wartości SFM z maksimum 0,04 przy 0,75s

4.1.6 Stosunek energii (ERSB) a płeć

- Głos męski: Rys. 2 ukazuje dominację ERSB1 (około 450) z niewielkim udziałem ERSB2 (maksymalnie 270)
- Głos żeński: Na rys. 5 widoczny bardziej równomierny rozkład między ERSB1 (400-500) a ERSB2 (100-200)

4.1.7 Samogłoski

- W słowie „onomatopeja” wyróżniamy samogłoski /o/, /e/, /a/
- Na rys. 1 i 4 widoczne jako obszary podwyższonego Volume
- W głosie męskim (rys. 2): Niższe wartości FC i dominujące ERSB1
- W głosie żeńskim (rys. 5): Wyższe wartości FC i większy udział ERSB2

4.1.8 Spółgłoski nosowe (/m/, /n/)

- Widoczne na rys. 1 i 4 przy około 0,15s i 0,25s
- Charakteryzują się obniżoną głośnością i FC w porównaniu do sąsiadujących samogłosek
- Na rys. 3 i 6 widoczne jako obszary o podwyższonym SCF

4.1.9 Spółgłoski zwarte (/p/, /t/)

- Spółgłoska /p/ widoczna przy około 0,55s na rys. 1 i 4
- Spółgłoska /t/ widoczna przy około 0,45s na tych samych wykresach
- Oba głosy wykazują gwałtowne spadki Volume i wysokie piki FC przy tych głoskach
- Na rys. 3 i 6 widoczne jako wyraźne szczyty SCF

4.1.10 Spółgłoski szczelinowe (/j/)

- Głoska /j/ widoczna przy około 0,7s na rys. 1 i 4
- Charakteryzuje się podwyższonym FC i BW przy umiarkowanej głośności
- Na rys. 2 i 5 widoczny zwiększyły udział ERSB3 przy tej głosce

4.1.11 Wnioski

Analiza parametrów częstotliwościowych słowa „onomatopeja” wykazuje systematyczne różnice między głosem męskim i żeńskim. Jak widać na rysunkach 1-6, głos żeński charakteryzuje się wyższymi wartościami centroidu częstotliwościowego, bardziej zrównoważonym rozkładem energii między pasmami oraz niższą płaskością widma, wskazującą na bardziej harmoniczną strukturę spektralną. Parametry częstotliwościowe pozwalają również efektywnie różnicować poszczególne klasy głosek, niezależnie od typu głosu.

4.2 Porównanie parametrów w dziedzinie częstotliwości dla różnych głosów i głosek

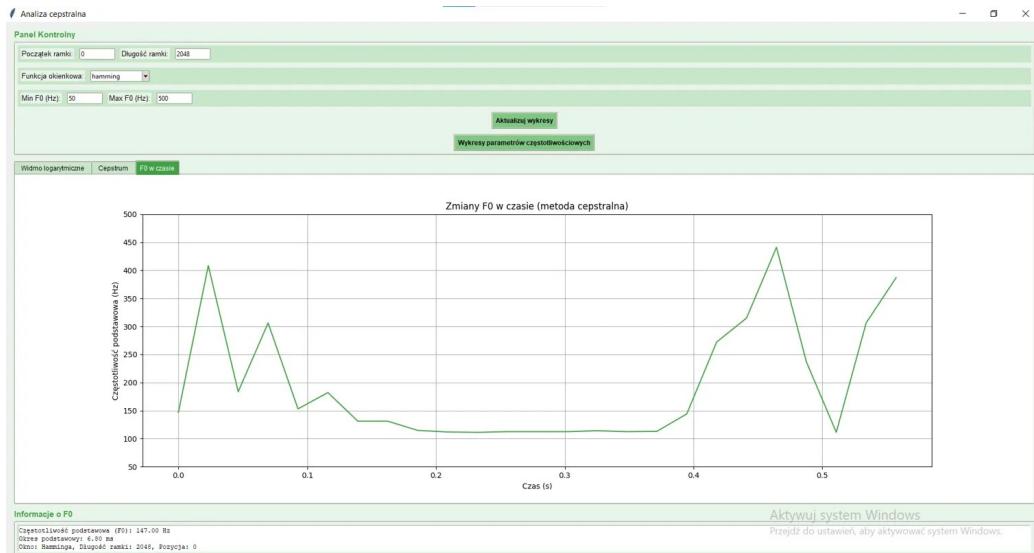
4.3 Analiza tonu podstawowego w głosie męskim

Przeprowadzono analizę cepstralną nagrań trzech polskich słów: *chrząszcz*, *jeden* i *zero*, wypowiadanych przez mężczyznę. Wykorzystano okno Hamminga o długości ramki 2048 próbek. Zakres analizy częstotliwości podstawowej (F0) ustalono między 50 Hz a 500 Hz.

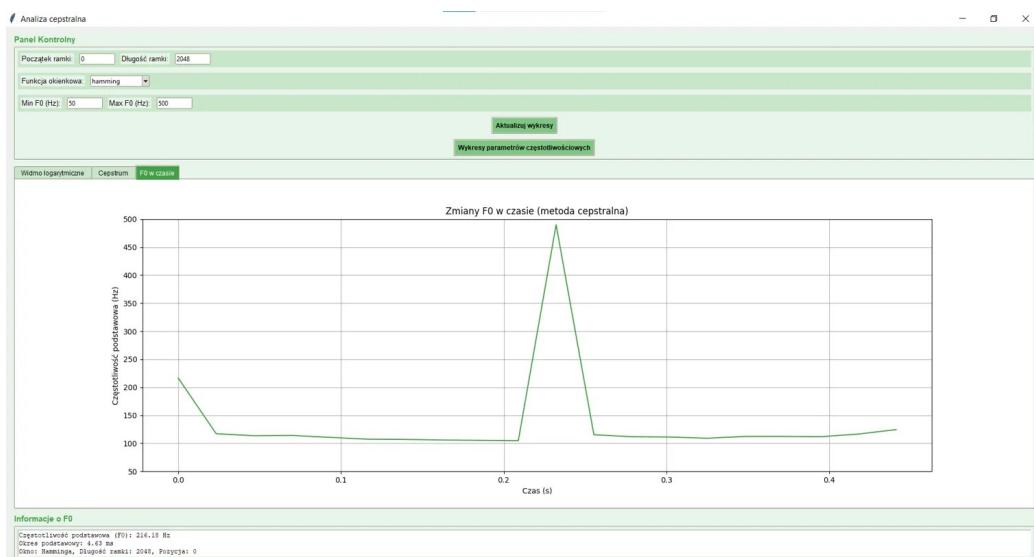
4.3.1 Częstotliwość podstawowa

W badanych nagraniach zaobserwowano następujące wartości F0:

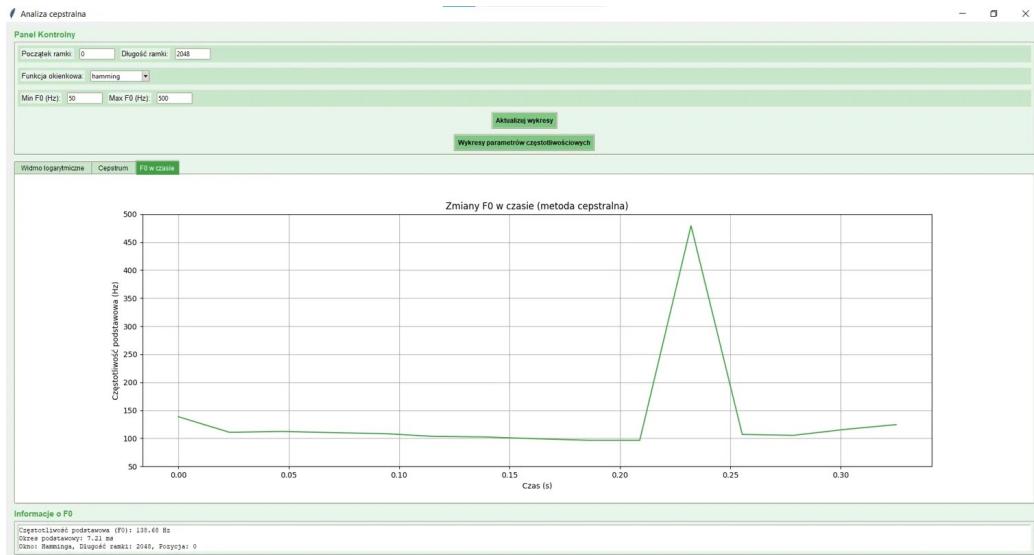
- Pierwsze nagranie: F0 = 147,00 Hz (okres: 6,80 ms)
- Drugie nagranie: F0 = 216,18 Hz (okres: 4,63 ms)
- Trzecie nagranie: F0 = 138,68 Hz (okres: 7,21 ms)



Rysunek 7: Zmiany F0 w czasie dla wyrazu "chrząszcz" (głos męski)



Rysunek 8: Zmiany F0 w czasie dla wyrazu "jeden" (głos męski)



Rysunek 9: Zmiany F0 w czasie dla wyrazu "zero" (głos męski)

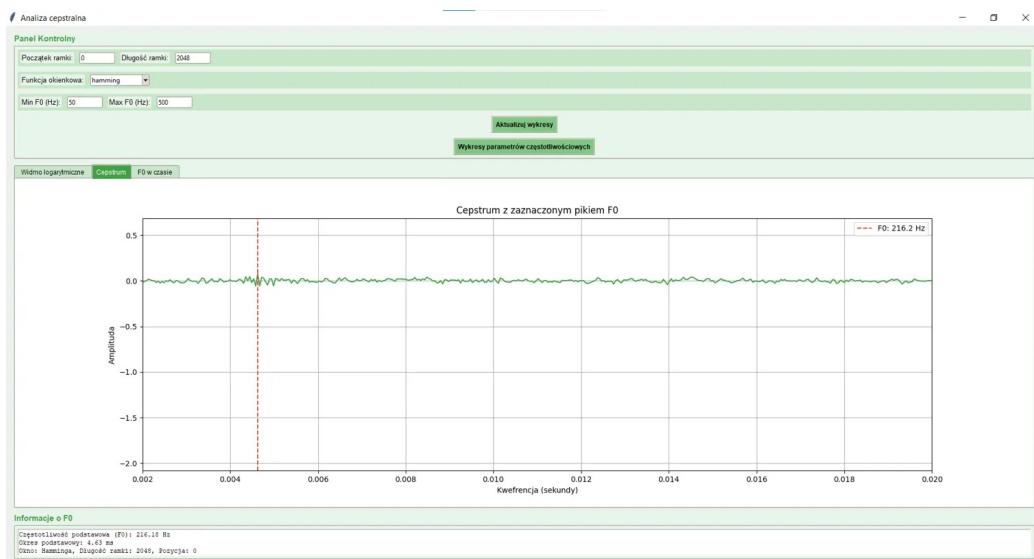
4.3.2 Kontury F0

Analiza zmian F0 w czasie wykazała charakterystyczne wzorce dla poszczególnych wyrazów:

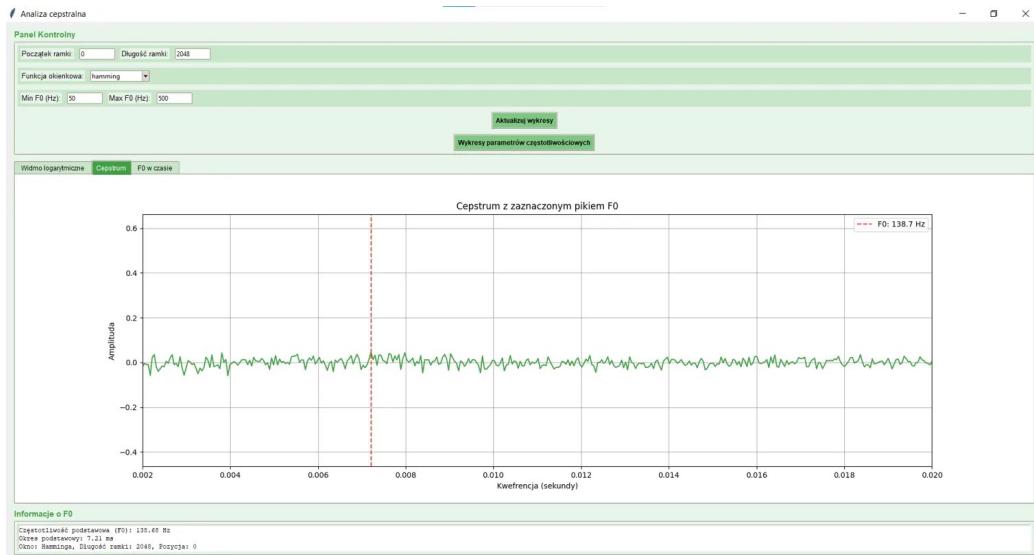
- *Chrząszcz*: kontur F0 wykazuje liczne wahania między około 150 a 400 Hz, co odzwierciedla złożoną strukturę fonetyczną tego wyrazu z licznymi zbitkami spółgłoskowymi.
- *Jeden*: widoczny jest wyraźny szczyt F0 przy około 0,22 sekundy, osiągający prawie 500 Hz, po którym następuje gwałtowny spadek. Szczyt ten występuje podczas wymowy sylaby "de-", co jest interesującym odkryciem, ponieważ w standardowym polskim akcent powinien padać na sylabę "je-".
- *Zero*: kontur F0 wykazuje znaczący szczyt około 0,22 sekundy (około 480 Hz), co odpowiada wymowie fragmentu "e-ro", a nie pierwszej sylabie "ze-".



Rysunek 10: Cepstrum z zaznaczonym pikiem F0 dla wyrazu "chrząszcz" (głos męski)



Rysunek 11: Cepstrum z zaznaczonym pikiem F0 dla wyrazu "jeden" (głos męski)

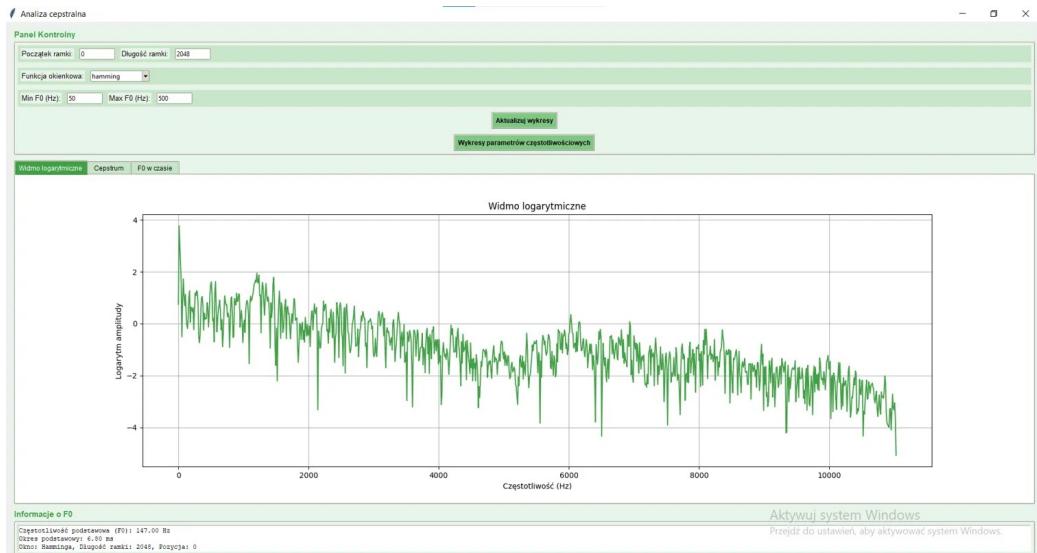


Rysunek 12: Cepstrum z zaznaczonym pikiem F0 dla wyrazu "zero" (głos męski)

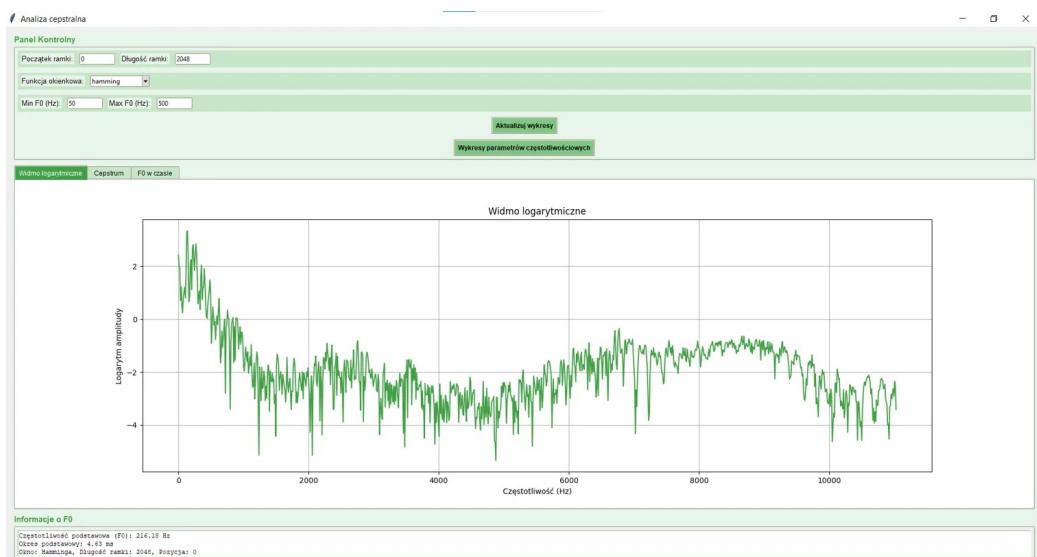
4.3.3 Charakterystyka widmowa

Widma logarytmiczne ujawniły odmienne struktury dla każdego wyrazu:

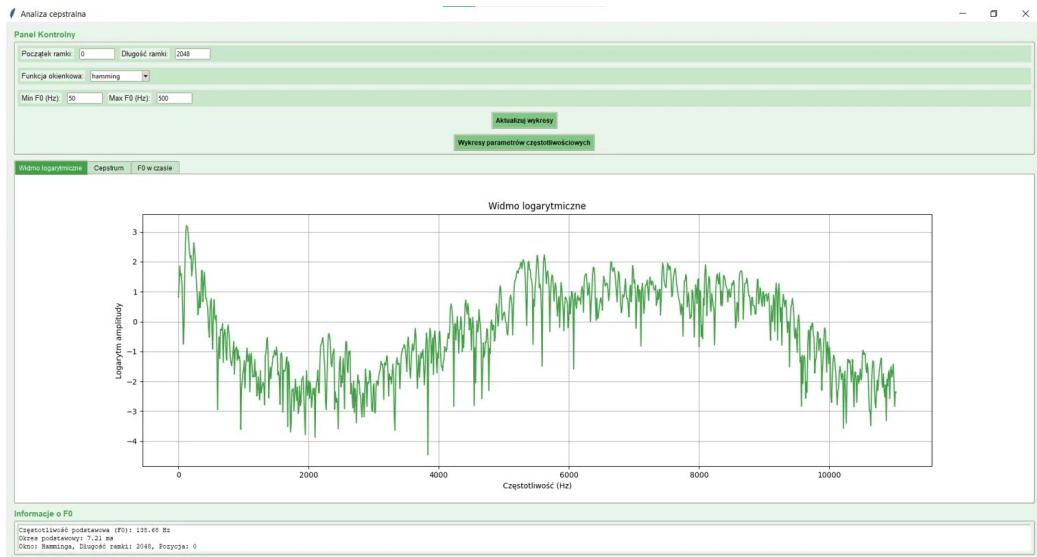
- *Chrząszcz*: energia skupiona głównie w niższych częstotliwościach.
- *Jeden*: energia skoncentrowana w kilku wyraźnych pasmach częstotliwości, szczególnie w zakresie 5000-9000 Hz.
- *Zero*: bardziej równomierne rozłożenie energii z istotnymi składowymi zarówno w niskich (0-2000 Hz), jak i średnich częstotliwościach.



Rysunek 13: Widmo logarytmiczne dla wyrazu "chrząszcz" (głos męski)



Rysunek 14: Widmo logarytmiczne dla wyrazu "jeden" (głos męski)



Rysunek 15: Widmo logarytmiczne dla wyrazu "zero" (głos męski)

4.3.4 Wnioski

Analiza wykazała interesujące odkrycia dotyczące rozkładu częstotliwości podstawowej:

1. Obserwowane szczyty F0 nie pokrywają się z oczekiwany wzorcem akcentowym języka polskiego. W wyrazie *jeden* szczyt F0 (ok. 500 Hz) pojawia się podczas wymowy sylaby "de-", a nie akcentowanej sylaby "je-". Podobnie w wyrazie *zero*, szczyt F0 (ok. 480 Hz) występuje podczas wymowy fragmentu "e-ro". Być może dobrano zbyt dużą wielkość ramki, przez co wyniki nie są satysfakcjonujące.
2. Złożone zbitki spółgłoskowe w wyrazie *chrząszcz* generują bardziej nie-regularne wzorce F0 w porównaniu do regularniejszych konturów obserwowanych w wyrazach *jeden* i *zero*, co potwierdza wpływ struktury fonetycznej na realizację F0.

4.4 Analiza tonu podstawowego w głosie żeńskim

W tym wypadku także przeprowadzono analizę cepstralną nagrani trzech słów: *chrząszcz, jeden i zero*, tym razem wypowiadanych przez kobietę. Parametry analizy były takie same jak w przypadku męskim.

4.4.1 Częstotliwość podstawowa

W badanych nagraniach zaobserwowano następujące wartości F0:

- Pierwsze nagranie (*chrząszcz*): $F_0 = 386,84 \text{ Hz}$ (okres: 2,59 ms)
- Drugie nagranie (*jeden*): $F_0 = 103,04 \text{ Hz}$ (okres: 9,71 ms)
- Trzecie nagranie (*zero*): $F_0 = 344,53 \text{ Hz}$ (okres: 2,90 ms)



Rysunek 16: Zmiany F0 w czasie dla wyrazu "chrząszcz" (głos żeński)



Rysunek 17: Zmiany F0 w czasie dla wyrazu "jeden" (głos żeński)

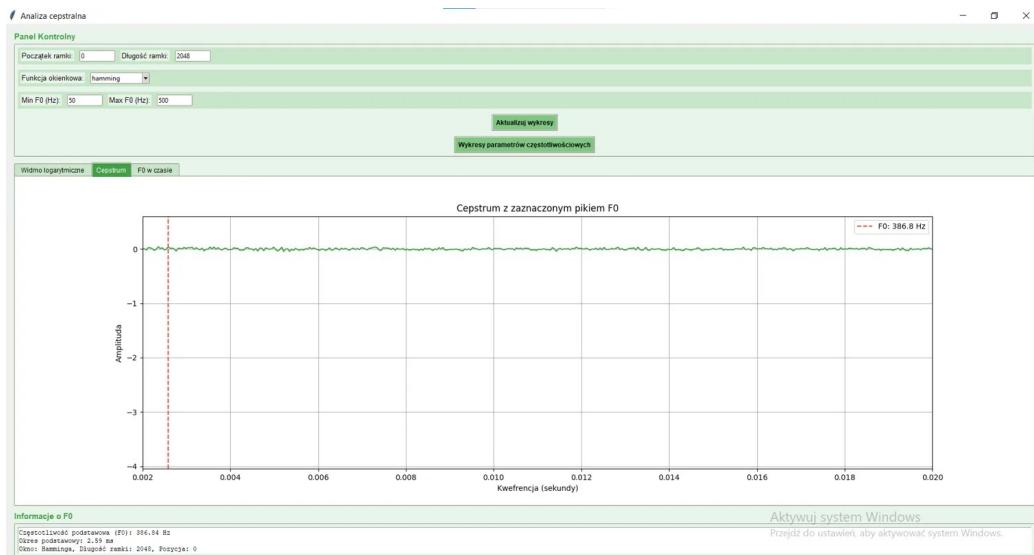


Rysunek 18: Zmiany F0 w czasie dla wyrazu "zero" (głos żeński)

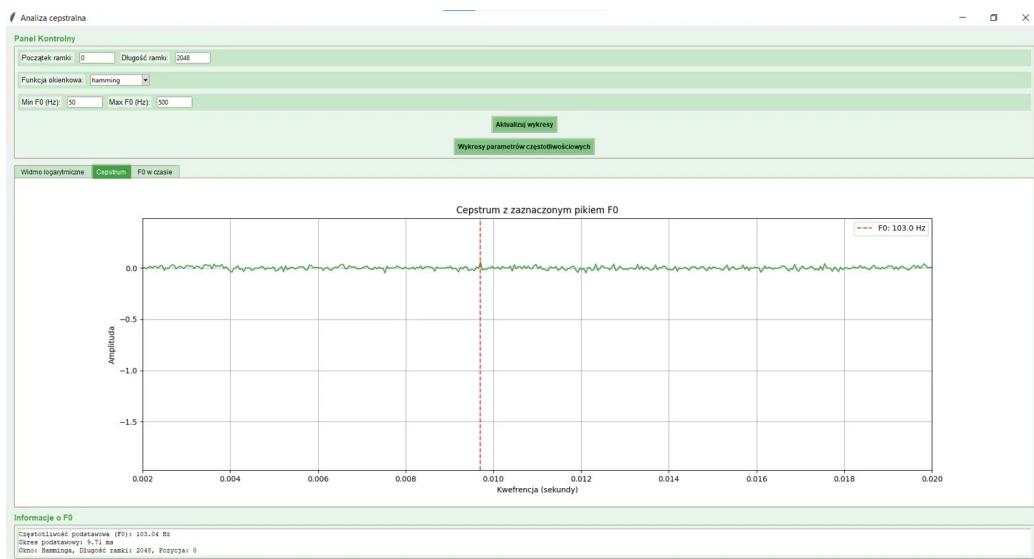
4.4.2 Kontury F0

Analiza zmian F0 w czasie wykazała następujące charakterystyczne wzorce:

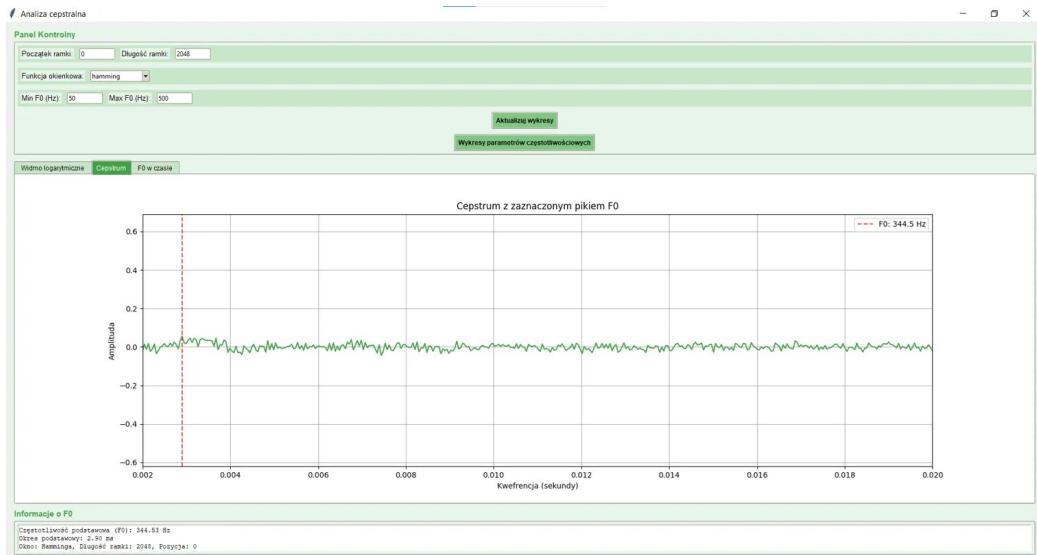
- *Chrząszcz*: kontur F0 wykazuje złożoną strukturę z wieloma wahaniemami między 150 a 480 Hz. Szczególnie widoczne są dwa wyraźne szczyty - jeden przy około 0,05 s (450 Hz) i drugi, największy przy około 0,4 s (prawie 500 Hz). Ta nieregularność odzwierciedla złożoność artykulacyjną słowa.
- *Jeden*: interesująco, w przeciwieństwie do typowego głosu żeńskiego, to nagranie charakteryzuje się znacznie niższą częstotliwością podstawową (około 103 Hz), co jest wartością typową dla głosu męskiego. Widoczny jest wyraźny szczyt przy około 0,07 s sięgający 380 Hz, po którym następuje stabilizacja w zakresie 250-270 Hz. Był może popełniono tu jakieś błędy, chociaż w innych przypadkach wyniki są w miarę oczekiwane. Możliwe, że po prostu metodologia cepstrum nie sprawdza się w tym konkretnym przypadku.
- *Zero*: kontur F0 wykazuje względnie stabilny przebieg w zakresie 260-300 Hz z wyraźnym wzrostem do około 370 Hz w końcowej fazie wyrazu (około 0,35 s). Charakterystyczny jest gwałtowny spadek F0 pod koniec nagrania.



Rysunek 19: Cepstrum z zaznaczonym pikiem F0 dla wyrazu "chrząszcz" (głos żeński)



Rysunek 20: Cepstrum z zaznaczonym pikiem F0 dla wyrazu "jeden" (głos żeński)

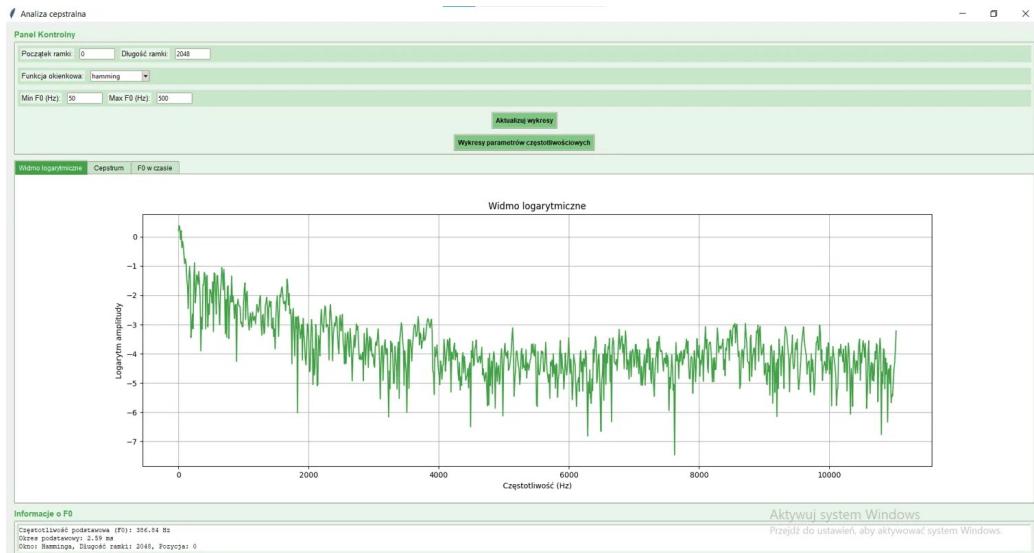


Rysunek 21: Cepstrum z zaznaczonym pikiem F0 dla wyrazu "zero" (głos żeński)

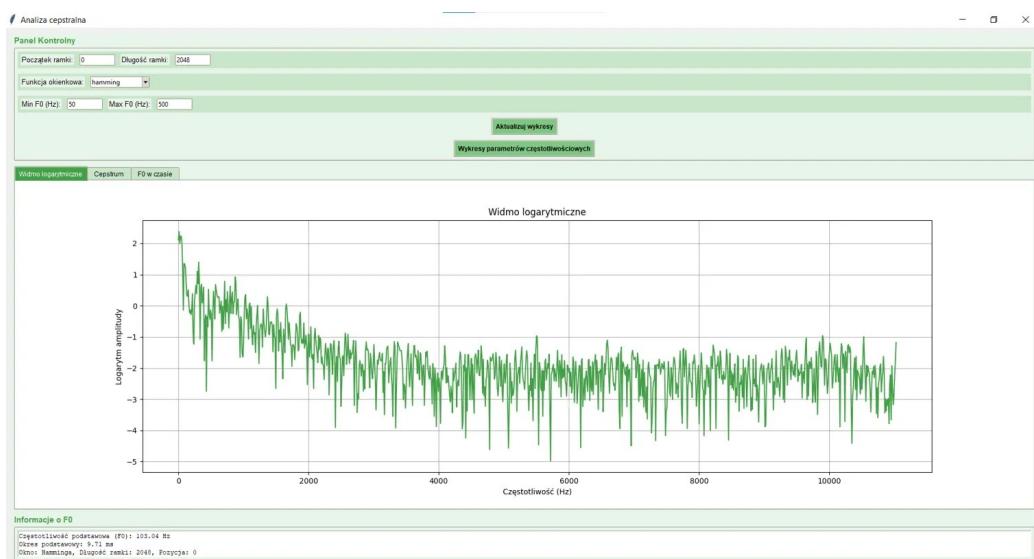
4.4.3 Charakterystyka widmowa

Widma logarytmiczne ujawniły odmienne struktury formantowe dla każdego wyrazu:

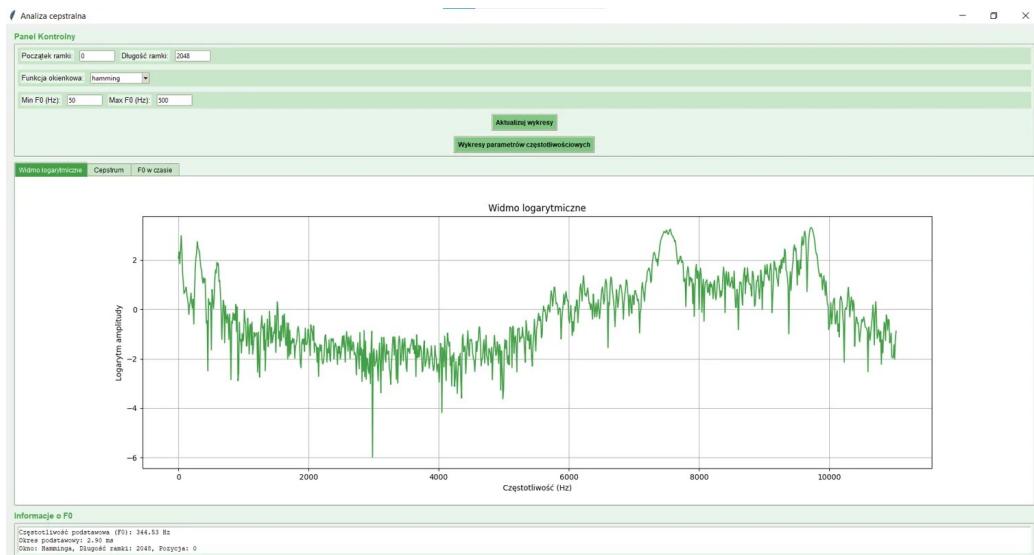
- *Chrząszcz*: widmo logarytmiczne wskazuje na koncentrację energii w zakresie niskich częstotliwości (poniżej 4000 Hz) z wyraźnym tłumieniem powyżej tego zakresu.
- *Jeden*: widmo wykazuje bardziej równomierny rozkład energii z wyraźnymi fragmentami w środkowym zakresie częstotliwości (2000-5000 Hz). W górnym zakresie (powyżej 6000 Hz) widoczne jest stopniowe tłumienie.
- *Zero*: najbardziej charakterystyczną cechą tego widma jest podwyższona energia w zakresie 7000-9000 Hz, tworząca wyraźne ”wzniesienie” w środkowej części widma.



Rysunek 22: Widmo logarytmiczne dla wyrazu "chrząszcz" (głos żeński)



Rysunek 23: Widmo logarytmiczne dla wyrazu "jeden" (głos żeński)



Rysunek 24: Widmo logarytmiczne dla wyrazu "zero" (głos żeński)

4.4.4 Wnioski

Analiza nagrań głosu żeńskiego doprowadziła do następujących wniosków:

1. Zaobserwowano znaczną różnicę w wartościach F0 między poszczególnymi nagraniami (103-387 Hz). Skrajny przypadek nie powinien być jednak wyznacznikiem tendencji w częstotliwości podstawowej głosu żeńskiego. Otrzymany wynik może być wynikiem niedokładności algorytmu analizy cepstralnej dla tego konkretnego nagrania.
2. W przeciwieństwie do głosu męskiego, częstotliwość podstawowa w głosie żeńskim wykazuje większą zmienność i szerszy zakres wartości, co jest zgodne z obserwacjami w literaturze fonetycznej.
3. W przypadku wyrazu "zero", ciekawym zjawiskiem jest podwyższona energia w paśmie 7000-9000 Hz, co może wskazywać na specyfczną realizację spółgłoski "z" lub indywidualną charakterystykę głosu mówcy w tym konkretnym kontekście fonetycznym.

Powysza analiza pokazuje, że nawet w obrębie mowy jednego mówcy (w tym przypadku kobiety) mogą występować znaczące różnice w parametrach akustycznych między poszczególnymi wypowiedziami, co potwierdza złożoność i dynamiczną naturę ludzkiej mowy.

4.5 Wnioski dotyczące częstotliwości podstawowej u różnych mówców

Przeprowadzone analizy cepstralne nagrań głosów męskich i żeńskich pozwalały sformułować następujące wnioski:

1. Zakres wartości F0:

- Mężczyzna: 138-216 Hz
- Kobieta: 103-387 Hz (z anomalią 103 Hz, prawdopodobnie błąd pomiaru)
- Głos żeński wykazuje większą zmienność F0

2. Wpływ struktury fonetycznej:

- “Chrząszcz” generuje nieregularne wzorce F0 w obu głosach ze względu na złożone zbitki spółgłoskowe
- Szczyty F0 nie pokrywają się z oczekiwany akcentem w języku polskim (np. w “jeden” szczyt na “-de-” a nie “je-”)

3. Obserwacje metodologiczne:

- Metoda cepstralna ma ograniczenia w niektórych kontekstach fonetycznych
- Te same parametry analizy mogą nie być optymalne dla różnych głosów

4. Końcowe wnioski:

- Istnieją systematyczne różnice między płciami, ale widoczna jest też znacząca zmienność indywidualna
- Kontekst fonetyczny silnie wpływa na realizację F0, co powinno być uwzględniane w systemach przetwarzania mowy

4.6 Formanty w samogłoskach - analiza spektrograficzna

4.6.1 Definicja formantów

Formanty to wzmocnione pasma częstotliwości w widmie akustycznym dźwięków mowy, szczególnie samogłosek, powstające w wyniku rezonansów w torze głosowym. Odpowiadają naturalnym częstotliwościom rezonansowym traktu

wokalnego, które są modyfikowane przez ułożenie artykulatorów (języka, warg, żuchwy). Formanty stanowią kluczowy element akustyczny pozwalający na identyfikację i rozróżnianie samogłosek.

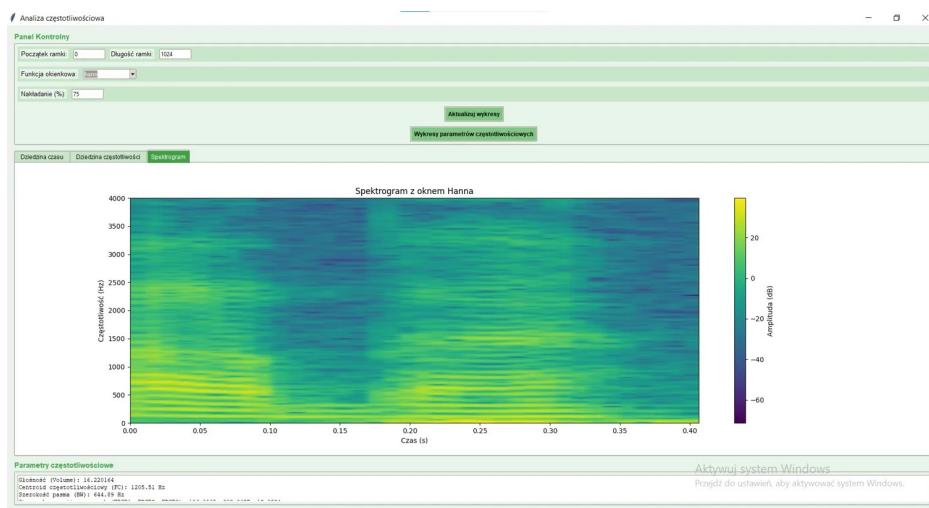
Najważniejsze formanty w analizie fonetycznej to:

- **F1 (pierwszy formant)** – związany głównie z pionowym położeniem języka i otwartością jamy ustnej; wyższe wartości F1 odpowiadają bardziej otwartej artykulacji
- **F2 (drugi formant)** – związany z poziomym położeniem języka (przód-tyl); wyższe wartości F2 odpowiadają bardziej przedniej artykulacji
- **F3 (trzeci formant)** – związany głównie z zaokrągleniem warg oraz innymi niuansami artykulacyjnymi

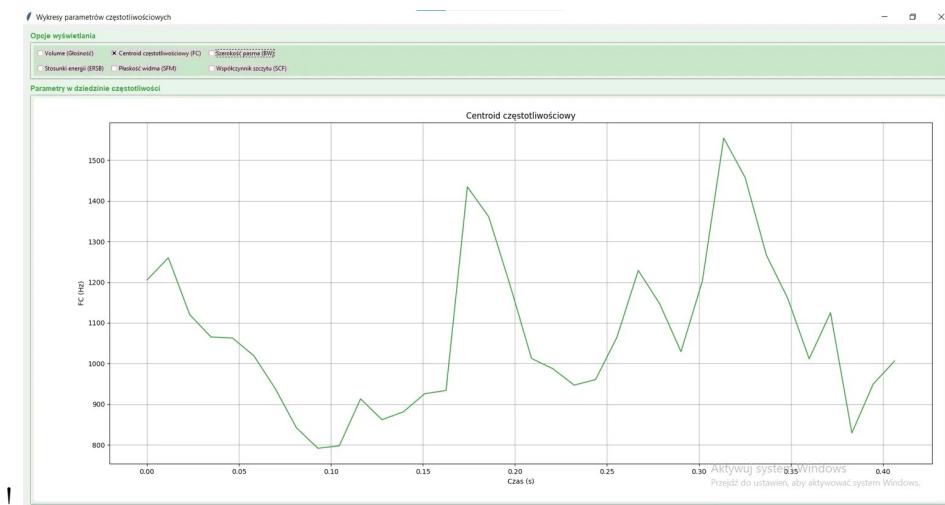
4.6.2 Analiza spektrogramów

Eksperymenty przeprowadzono używając okna Hanna i stosując nakładanie na poziomie 75%. Długość ramki wynosiła 1024. Przeprowadzona analiza spektrogramów słów zawierających różne samogłoski jednoznacznie potwierdza możliwość zaobserwowania struktury formantowej. Na spektrogramach formanty są widoczne jako poziome pasma o podwyższonej energii (jaśniejsze obszary). Częstotliwości formantów można odczytać z osi pionowej spektrogramu.

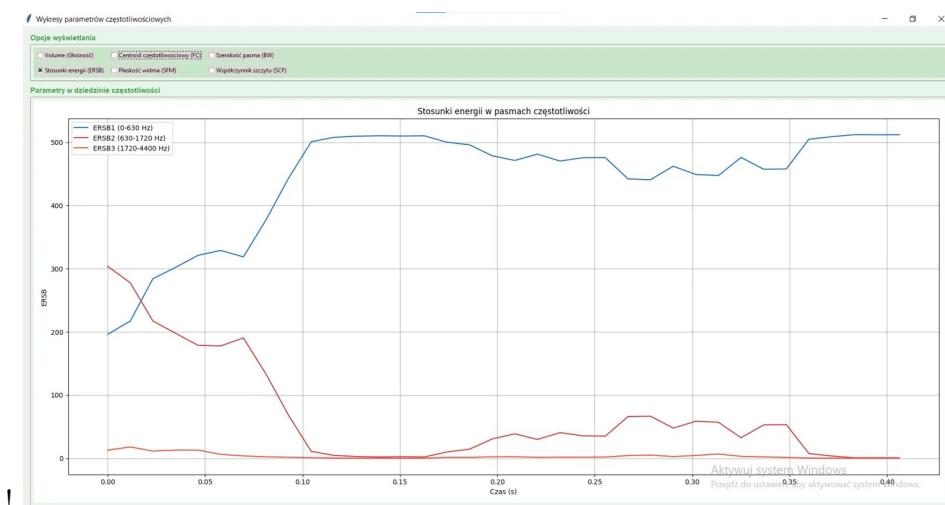
Poniżej przedstawiono wykresy dla słowa "abe":



Rysunek 25: Spektrogram słowa "abe" z widocznymi formantami samogłosek [a] (0.00-0.05s) i [e] (0.15-0.20s)

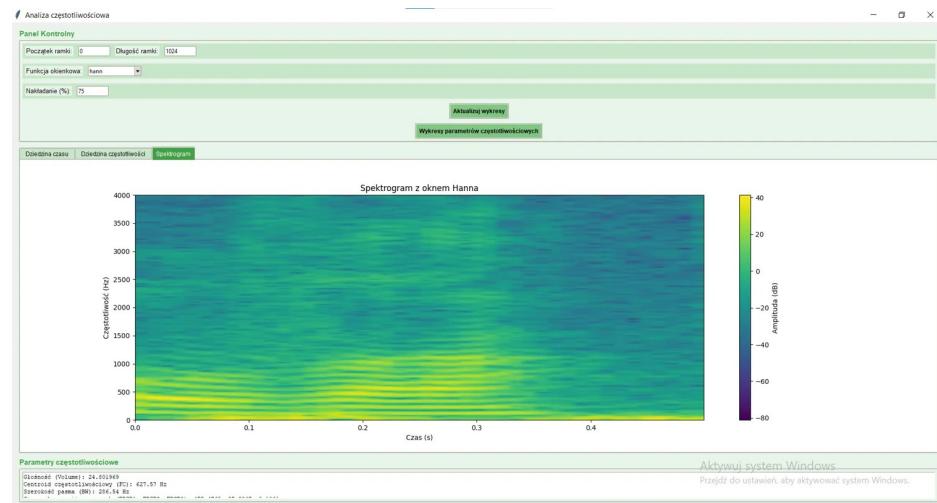


Rysunek 26: Wykres centroidu częstotliwościowego dla słowa ”abe”

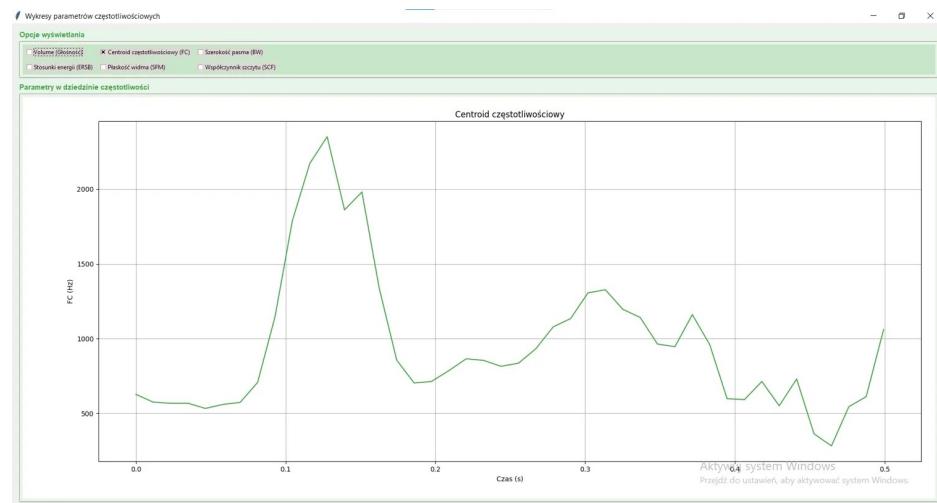


Rysunek 27: Stosunki energii w pasmach częstotliwości (ERSB) dla słowa ”abe”

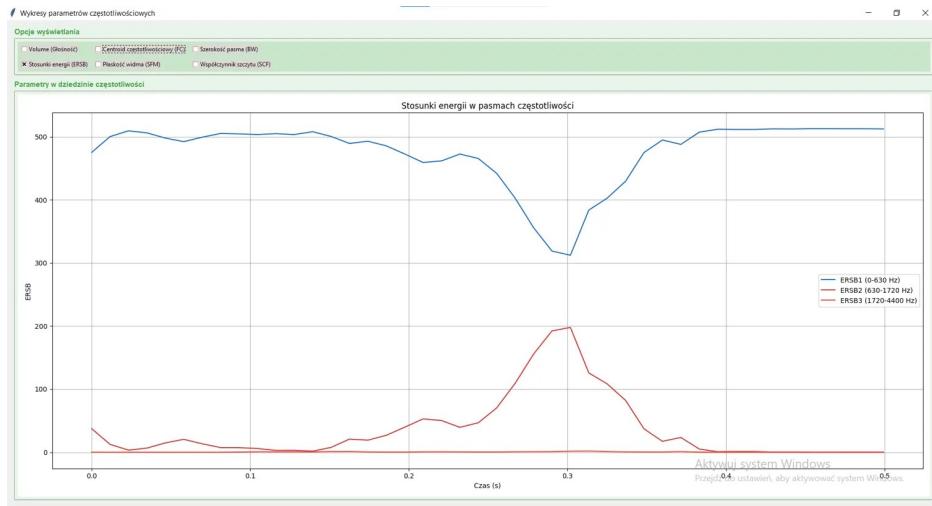
Poniżej znajdują się wykresy dla słowa ”uwo”:



Rysunek 28: Spektrogram słowa "uwo" z widocznymi formantami samogłosek [u] (0.00-0.10s) i [o] (0.20-0.40s)



Rysunek 29: Wykres centroidu częstotliwościowego dla słowa "uwo"



Rysunek 30: Stosunki energii w pasmach częstotliwości (ERSB) dla słowa "uwo"

4.6.3 Wartości formantów dla poszczególnych samogłosek

Na podstawie analizy spektrogramów możemy określić typowe wartości formantów dla badanych samogłosek:

!

Tabela 2: Wartości formantów dla analizowanych samogłosek

Samogłoska	F1 [Hz]	F2 [Hz]	Charakterystyka artykulacyjna
[a]	700–800	1200–1300	samogłoska niska, centralna/tylna
[e]	600–700	1800–2000	samogłoska średnio-niska, przednia
[u]	300–400	600–700	samogłoska wysoka, tylna, zaokrąglona
[o]	500–600	900–1100	samogłoska średnio-wysoka, tylna, zaokrąglona

4.6.4 Zależności między wartościami formantów a cechami artykulacyjnymi

Wartości formantów pozostają w ścisłym związku z cechami artykulacyjnymi samogłosek:

1. Otwartość jamy ustnej (wysokość języka):

- Samogłoski otwarte ([a]) mają wysokie wartości F1 (700–800 Hz)
- Samogłoski zamknięte ([u]) mają niskie wartości F1 (300–400 Hz)

2. Przedniość-tylność (poziome położenie języka):

- Samogłoski przednie ([e]) mają wysokie wartości F2 (1800–2000 Hz)
- Samogłoski tylne ([u], [o]) mają niskie wartości F2 (600–1100 Hz)

3. Zaokrąglenie warg:

- Zaokrąglenie warg obniża wartości formantów, szczególnie F2 i F3
- Samogłoski zaokrąglone ([u], [o]) mają niższe wartości F2 niż ich niezaokrąglone odpowiedniki o podobnej wysokości języka

4.6.5 Dynamika formantów

Analiza spektrogramów pokazuje również dynamiczną naturę formantów – zmieniają się one w czasie zarówno w obrębie jednej samogłoski (szczególnie na początku i końcu, gdy wpływ mają sąsiednie spółgłoski), jak i przy przejściach między głośkami.

4.6.6 Wnioski

Przeprowadzona analiza jednoznacznie potwierdza, że w samogłoskach można zaobserwować strukturę formantową. Wartości formantów dla poszczególnych samogłosek są względnie stabilne i charakterystyczne, co pozwala na ich identyfikację oraz klasyfikację. Różnice w wartościach formantów bezpośrednio odzwierciedlają różnice artykulacyjne między samogłoskami.

4.7 Analiza zmienności formantów w różnych realizacjach tego samego fonemu samogłoskowego

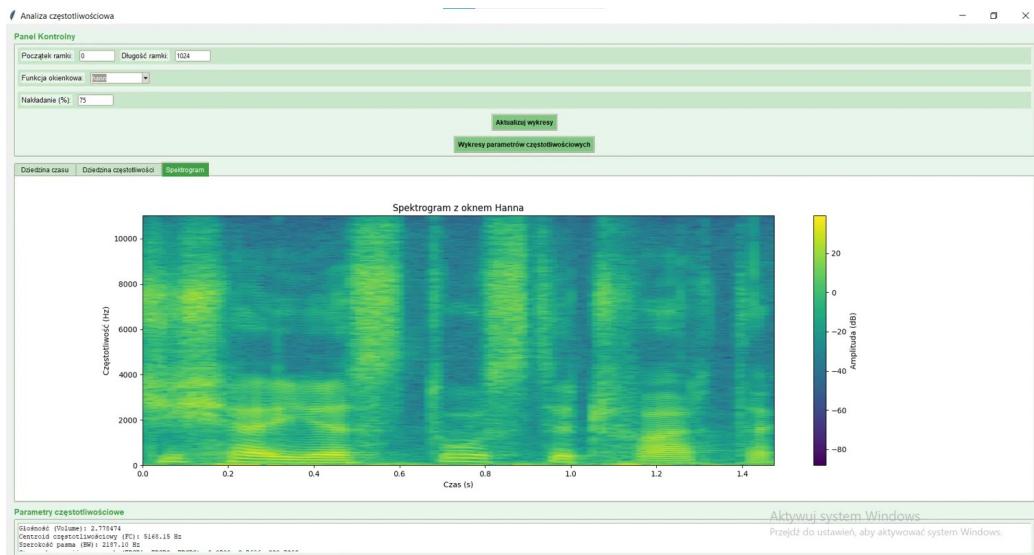
Przeprowadzono analizę zmienności formantów tego samego fonemu samogłoskowego u jednego mówcy (mężczyzna) w dwóch różnych wypowiedziach: „dzisiaj jest tłusty czwartek” oraz „czy przyniósł ktoś dzisiaj pączka”.

4.7.1 Metodologia

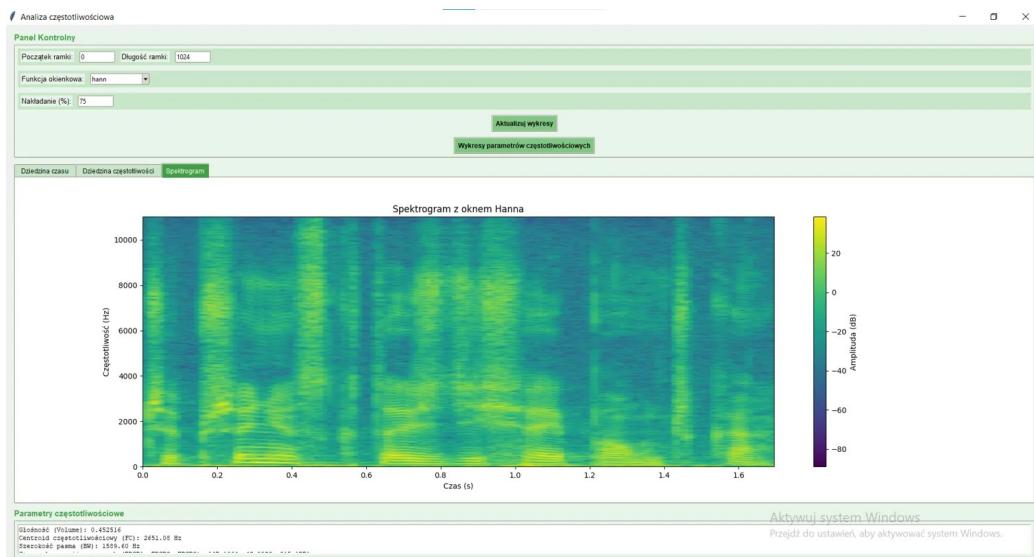
Analiza została przeprowadzona przy użyciu następujących narzędzi:

- Spektrogramów z wykorzystaniem okna Hanna (ramka 1024 próbki, nakładanie 75%)
- Centroidu częstotliwościowego (FC)
- Stosunków energii w pasmach częstotliwości (ERSB)

4.7.2 Wyniki analizy



Rysunek 31: Spektrogram wypowiedzi „dzisiaj jest tłusty czwartek” z oknem Hanna

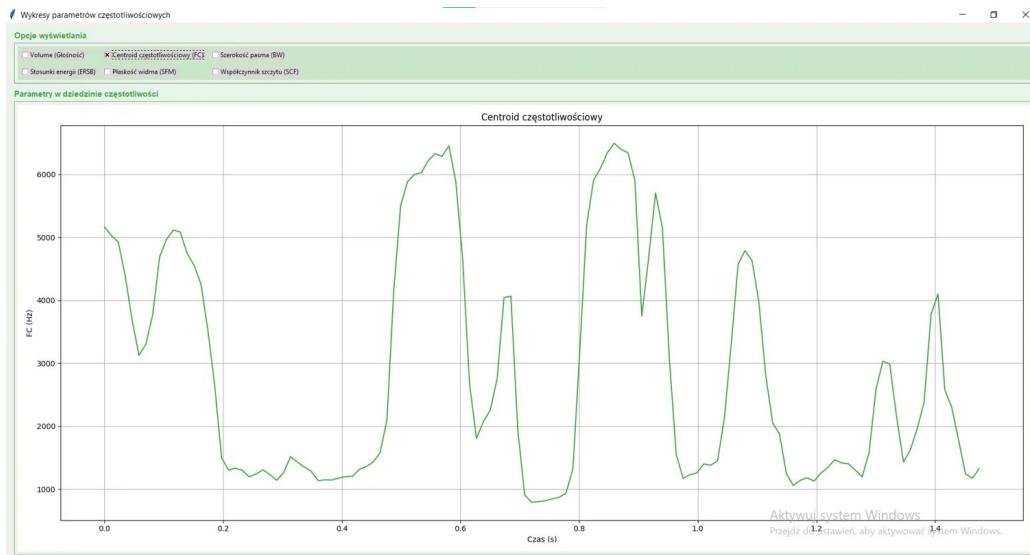


Rysunek 32: Spektrogram wypowiedzi „czy przyniósł ktoś dzisiaj pączka” z oknem Hanna

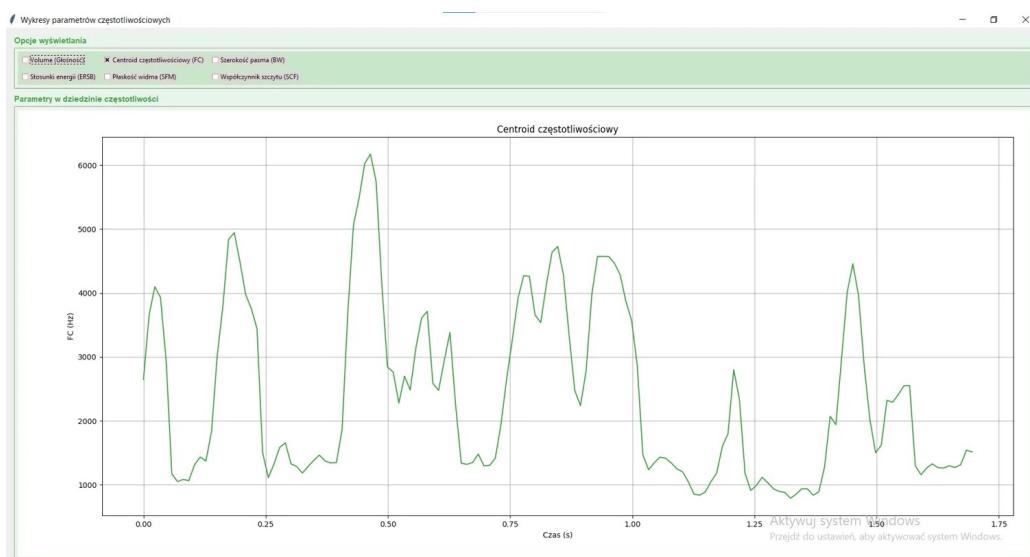
Realizacja samogłoski ”i” w wyrazie ”dzisiaj”: Na podstawie spektrogramów (rys. 31 i 32) wyodrębniono charakterystykę formantową dla samogłoski ”i” w

wyrazie "dzisiaj":

- Pierwsza realizacja (0,2-0,25 s na rys. 31): $F1 \approx 400\text{-}500$ Hz, $F2 \approx 2200\text{-}2400$ Hz
- Druga realizacja (1,0-1,05 s na rys. 32): $F1 \approx 350\text{-}450$ Hz, $F2 \approx 2100\text{-}2300$ Hz



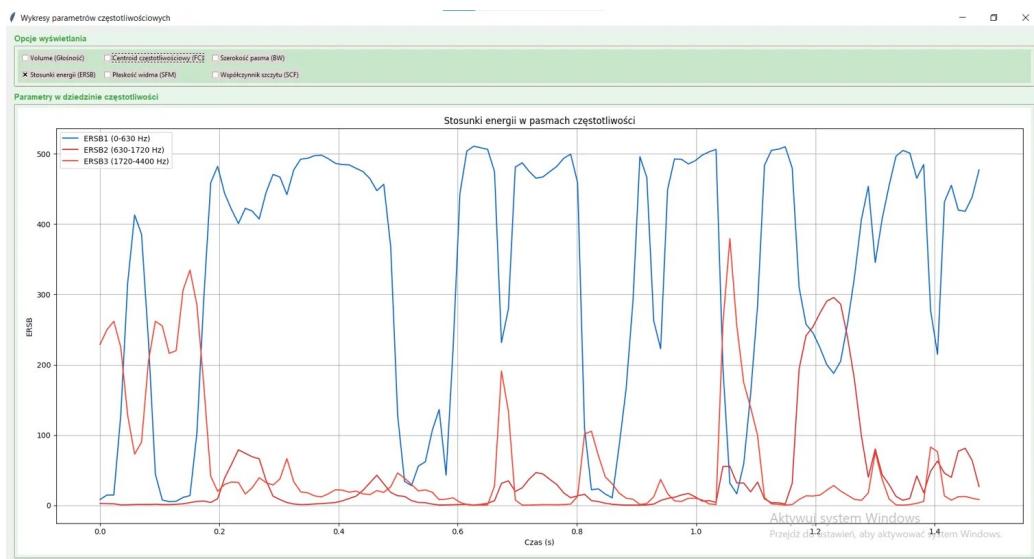
Rysunek 33: Centroid częstotliwościowy dla pierwszej wypowiedzi



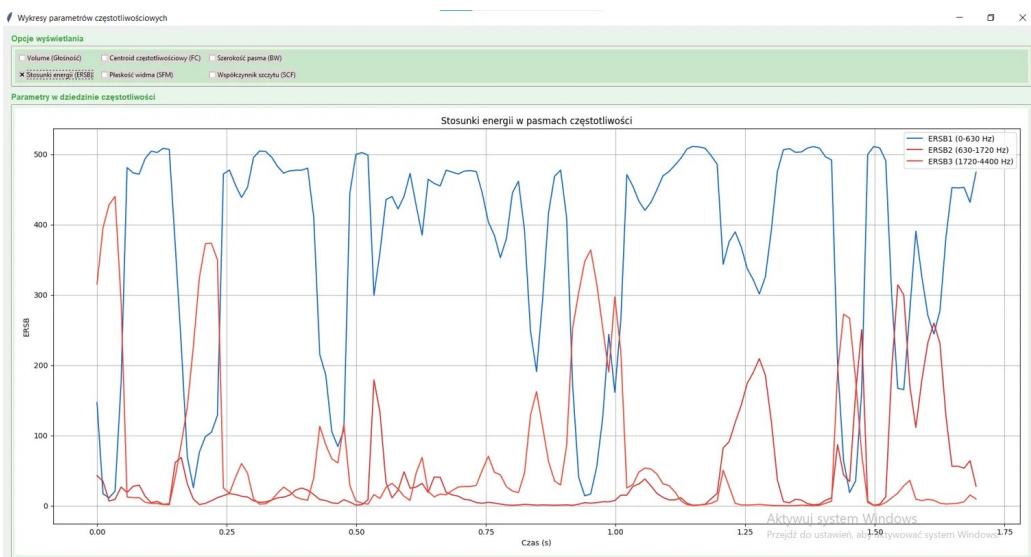
Rysunek 34: Centroid częstotliwościowy dla drugiej wypowiedzi

Analiza centroidu częstotliwościowego (rys. 33 i 34) dla samogłoski "i" w wyrazie "dzisiaj" wykazała różnicę między realizacjami sięgającą około 400 Hz:

- Pierwsza realizacja: $FC \approx 5000$ Hz
- Druga realizacja: $FC \approx 4600$ Hz



Rysunek 35: Stosunki energii w pasmach częstotliwości dla pierwszej wypowiedzi



Rysunek 36: Stosunki energii w pasmach częstotliwości dla drugiej wypowiedzi

Realizacja samogłoski "a" w wyrazie "dzisiaj": Dla samogłoski "a" w tym samym wyrazie również zaobserwowano pewne różnice:

- Pierwsza realizacja (0,3 s): $F1 \approx 700-800$ Hz, $F2 \approx 1400-1500$ Hz, $FC \approx 1500$ Hz
- Druga realizacja (1,1 s): $F1 \approx 700-800$ Hz, $F2 \approx 1500-1600$ Hz, $FC \approx 1400$ Hz

Analiza ERSB (rys. 35 i 36) wykazała, że dla samogłoski "a" w obu realizacjach dominuje ERSB1 (0-630 Hz), jednak z różną intensywnością.

4.7.3 Wnioski

1. Formanty tego samego fonemu samogłoskowego wykazują mierzalną zmienność między różnymi realizacjami, nawet gdy są wypowiadane przez tę samą osobę. Różnice te mogą wynikać z:

- Kontekstu fonetycznego (wpływ sąsiadujących głosek)
- Czynników takich jak akcent czy intonacja
- Potencjalnych niedokładności nagrań i pomiarów

2. Obserwowane prawidłowości:

- Pierwszy formant (F1) wykazuje większą stabilność między realizacjami niż drugi formant (F2)
- Centroid częstotliwościowy może różnić się dość widocznie między realizacjami tego samego fonemu
- Samogłoski wysokie (np. "i") cechują się większą zmiennością formantów niż samogłoski niskie (np. "a")

3. Mimo obserwowanych różnic, formanty zachowują swoją charakterystyczną strukturę, co pozwala na ich identyfikację jako tych samych fonemów samogłoskowych. Zauważone różnice mieszczą się w granicach naturalnej zmienności realizacyjnej.

4. Różnice w parametrach spektralnych mogą być również częściowo spowodowane niedokładnościami w procesie nagrywania lub analizy.

Uzyskane wyniki potwierdzają, że nawet u tej samej osoby realizacja fonemów samogłoskowych charakteryzuje się pewną naturalną zmiennością.

4.8 Analiza porównawcza formantów samogłosek u różnych osób

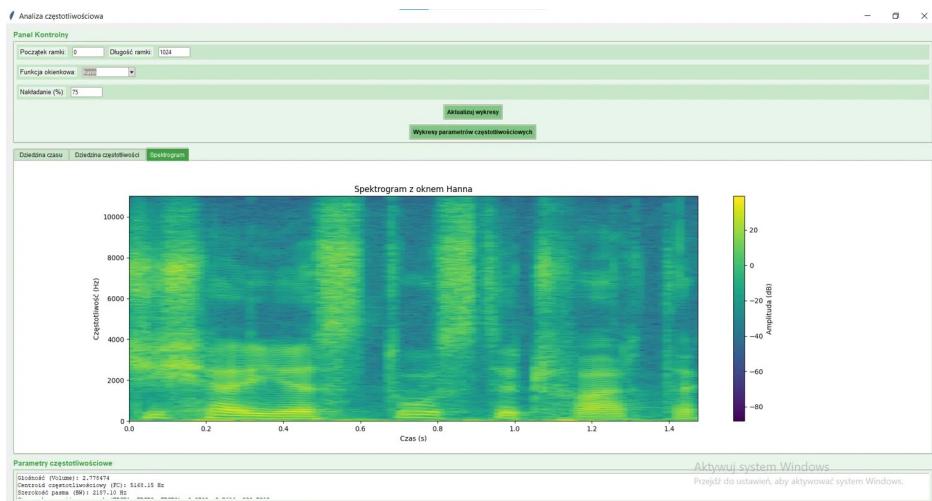
W celu porównania formantów samogłosek u różnych osób, przeprowadzono wieloaspektową analizę z wykorzystaniem różnych reprezentacji parametrów częstotliwościowych. Do testów użyto tych samych zdań, co dla mężczyzn, z uwzględnieniem tego, że tym razem mówcą była kobieta.

4.8.1 Metodologia

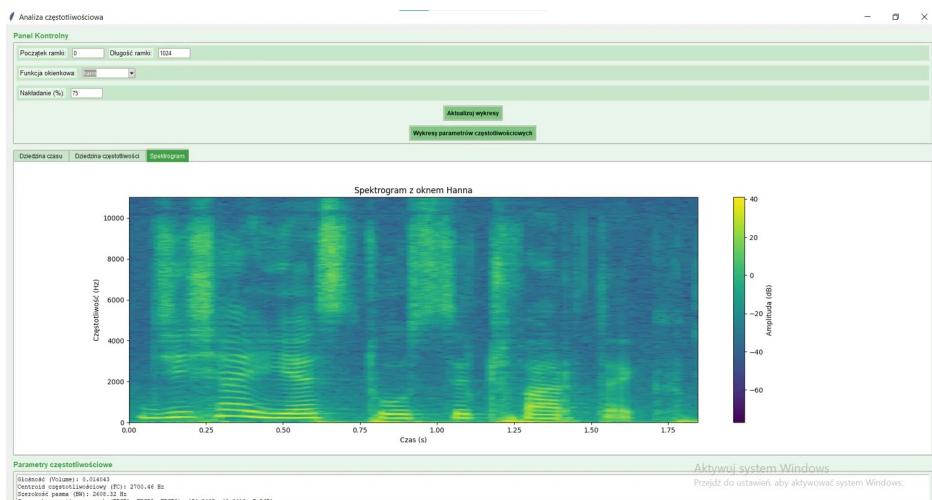
Analiza obejmowała badanie:

- Spektrogramów (okno Hanna, długość ramki 1024, nakładanie 75%)
- Centroidu częstotliwościowego (FC)
- Stosunków energii w pasmach częstotliwości (ERSB)

4.8.2 Analiza spektrogramów



Rysunek 37: Spektrogram zdania "dzisiaj jest tłusty czwartek" dla mężczyzny



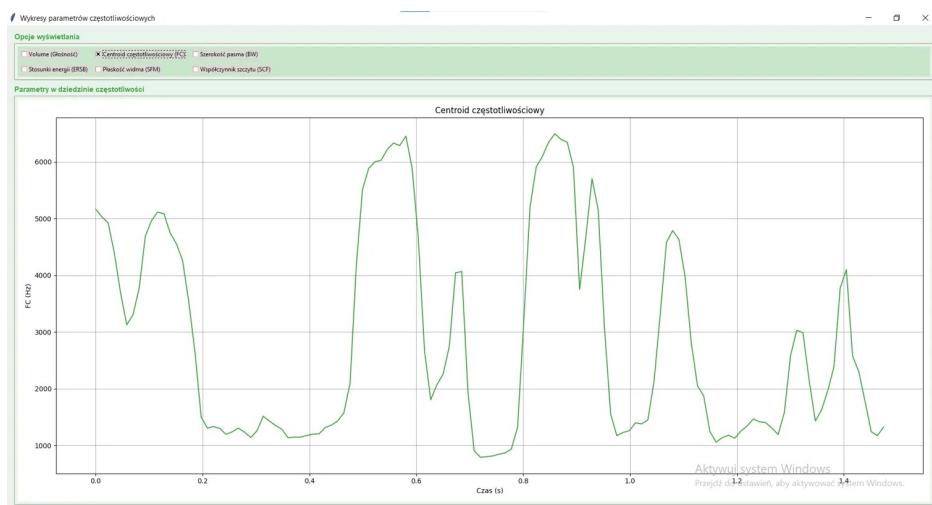
Rysunek 38: Spektrogram zdania "dzisiaj jest tłusty czwartek" dla kobiety

Porównanie spektrogramów (Rys. 37 i 38) ujawnia różnice w strukturze formantowej między różnymi mówcami:

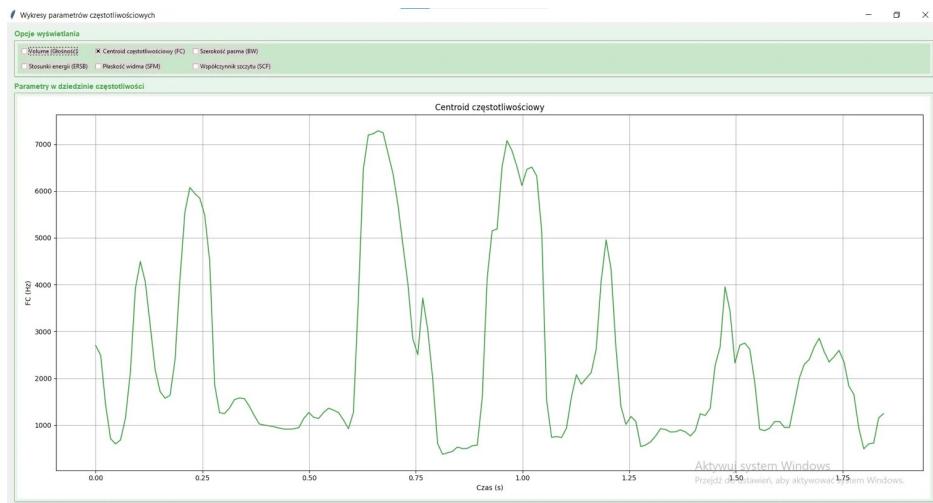
- Widoczne są różnice w położeniu i intensywności poziomych pasm formantowych

- U różnych mówców energia w tych samych zakresach częstotliwości ma różną intensywność, co manifestuje się różnym nasyceniem kolorów
- Charakterystyczne dla każdego mówcy są układy formantów szczególnie w przedziale 0-3000 Hz

4.8.3 Analiza centroidu częstotliwościowego



Rysunek 39: Centroid częstotliwościowy zdania "dzisiaj jest tłusty czwartek" dla mężczyzny

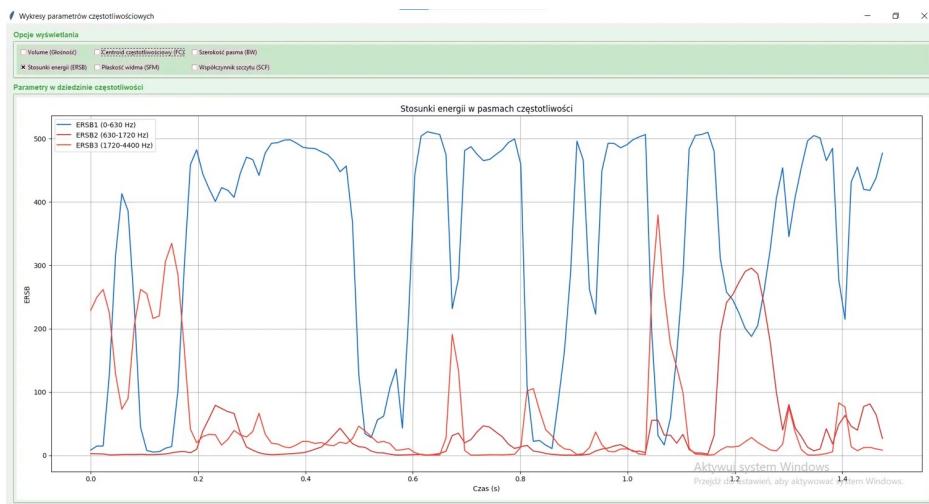


Rysunek 40: Centroid częstotliwościowy zdania ”dzisiaj jest tłusty czwartek” dla kobiety

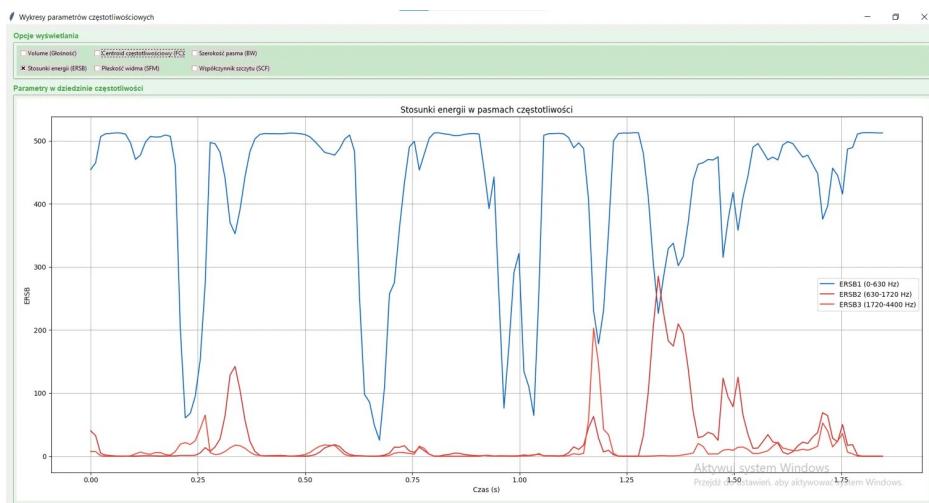
Wykresy centroidu częstotliwościowego (Rys. 39 i 40) wskazują na:

- Znaczące różnice w wartościach szczytowych FC między mówcami, sięgające 1000-2000 Hz
- Różnice w dynamice zmian FC - poszczególni mówcy wykazują charakterystyczne wzorce zmian tego parametru w czasie
- W przypadku kobiety (Rys. 40) maksimum FC osiąga około 7000 Hz, a u mężczyzn (Rys. 39) maksymalne wartości są niższe i wynoszą około 6000 Hz

4.8.4 Analiza stosunków energii w pasmach częstotliwości



Rysunek 41: Stosunki energii w pasmach częstotliwości zdania "dzisiaj jest tłusty czwartek" dla mężczyzny



Rysunek 42: Stosunki energii w pasmach częstotliwości zdania "dzisiaj jest tłusty czwartek" dla mężczyzny

Porównanie rozkładu energii w pasmach (Rys. 41 i 42) pokazuje:

- U obu mówców dominuje energia w paśmie ERSB1 (0-630 Hz), ale z różną intensywnością

- Mężczyzna (Rys. 41) wykazuje większą zmienność w stosunku energii między pasmami ERSB1 i ERSB2
- Charakterystyczne dla każdego mówcy są momenty, gdy energia w paśmie ERSB2 (630-1720 Hz) przewyższa energię w ERSB1

4.8.5 Wnioski

Analiza wykazała wyraźne różnice w charakterystyce formantowej samogłosek między różnymi osobami:

1. Formanty wykazują znaczącą zmienność międzynosobniczą, wynikającą prawdopodobnie z anatomicznych różnic
2. Różnice w wartościach centroidu częstotliwościowego między mówcami są znaczące (1000-2000 Hz) i systematyczne
3. Rozkład energii w pasmach częstotliwości jest charakterystyczny dla każdego mówcy”
4. Te same fonemy samogłoskowe realizowane przez różnych mówców wykazują różnice w strukturze formantowej, zachowując jednocześnie cechy pozwalające na ich identyfikację

Przeprowadzona analiza potwierdza, że formanty samogłosek są w znaczącym stopniu uwarunkowane indywidualnymi cechami anatomicznymi i artykulacyjnymi mówcy.

4.9 Różnice między samogłoskami a spółgłoskami w analizie widmowej

Analiza spektrogramów i transformaty Fouriera pozwala na obserwację wyraźnych różnic między samogłoskami a spółgłoskami. W dokumentacji zamieszczono szereg eksperymentów przedstawiających te różnice.

4.9.1 Charakterystyczne cechy widmowe samogłosek

Samogłoski w analizie widmowej charakteryzują się następującymi cechami:

- **Wyraźna struktura formantowa** – na spektrogramach widoczna jako poziome pasma o podwyższonej energii (jaśniejsze obszary); formanty te są kluczowe dla identyfikacji poszczególnych samogłosek

- **Stabilność czasowa** – względnie stabilna charakterystyka częstotliwościowa utrzymująca się przez dłuższy czas
- **Koncentracja energii w niższych częstotliwościach** – widoczna szczególnie w analizie stosunków energii w pasmach częstotliwości (ERSB), gdzie dominuje pasmo ERSB1 (0-630 Hz)
- **Charakterystyczne układy formantów** – umożliwiające identyfikację konkretnych samogłosek, np. dla samogłoski [a]: $F1 \approx 700-800$ Hz, $F2 \approx 1200-1300$ Hz; dla [e]: $F1 \approx 600-700$ Hz, $F2 \approx 1800-2000$ Hz

4.9.2 Charakterystyczne cechy widmowe spółgłosek

Spółgłoski wykazują znacząco odmienną charakterystykę spektralną:

- **Nieregularna struktura widmowa** – brak wyraźnych, stabilnych formantów lub ich znaczne osłabienie
- **Zróżnicowana charakterystyka** w zależności od typu spółgłoski:
 - **Zwarto-wybuchowe** (p, t, k, b, d, g) – widoczne jako momenty ciszy, zakończone krótkim wybuchem energii
 - **Szczelinowe** (f, s, sz, z, ż) – chaotyczny, szumowy rozkład energii koncentrujący się głównie w wyższych częstotliwościach
 - **Nosowe i płynne** (m, n, l, r) – słabsze pasma formantowe niż w samogłoskach, z dodatkowymi modyfikacjami.
- **Krótszy czas trwania** – szczególnie w przypadku spółgłosek zwartych
- **Energia rozłożona w szerszym zakresie częstotliwości** – widoczna w wyższych wartościach centroidu częstotliwościowego i płośkości widma
- **Charakterystyczne przejścia formantów** – na granicach między spółgłoskami a samogłoskami widoczne są charakterystyczne zmiany częstotliwości formantów

4.9.3 Przykłady w analizowanym materiale

Różnice między samogłoskami a spółgłoskami są szczególnie dobrze widoczne na następujących rysunkach:

- Spektrogramy słów „abe” (rys. 25) i „ubo” (rys. 28) – wyraźnie widoczne różnice między samogłoskami [a], [e], [u], [o] a spółgłoskami [b] i [w]

- Spektrogramy dla dłuższych wypowiedzi – „dzisiaj jest tłusty czwartek” (rys. 31) i „czy przyniósł ktoś dzisiaj pączka” (rys. 32) – ukazujące naprzemienne występowanie charakterystycznych struktur spółgłoskowych i samogłoskowych
- Analiza centroidu częstotliwościowego (rys. 33, 34) – pokazująca znaczące różnice w położeniu „środka ciężkości” widma między samogłoskami a spółgłoskami

Analiza widmowa potwierdza, że samogłoski i spółgłoski charakteryzują się fundamentalnie odmiennymi właściwościami akustycznymi, co odzwierciedla ich różnice artykulacyjne – samogłoski tworzone są przy otwartym kanale głosowym bez przeskódeł, podczas gdy spółgłoski powstają przy częściowym lub całkowitym zamknięciu kanału głosowego.

4.10 Analiza wpływu funkcji okienkowych na widmo częstotliwościowe i spektrogram

4.10.1 Wprowadzenie

Analiza porównawcza funkcji okienkowych została przeprowadzona dla słowa „brzęczyszczkiewicz” z wykorzystaniem pięciu różnych okien: prostokątnego (rectangular), trójkątnego (triangular/Bartlett), Hamminga, Hanna oraz Blackmana. Dla każdego okna analizowano zarówno widmo częstotliwościowe FFT, jak i odpowiadający mu spektrogram. Wszystkie analizy przeprowadzono przy jednakowych parametrach: długość ramki 1024 próbki oraz 90% nakładania.

4.10.2 Widmo okna prostokątnego



Rysunek 43: Widmo częstotliwościowe z oknem prostokątnym

Widmo częstotliwościowe uzyskane przy użyciu okna prostokątnego (rys. 43) charakteryzuje się:

- Wąskim listkiem głównym (wysoka rozdzielczość częstotliwościowa)
- Silnymi nieregularnościami w całym zakresie częstotliwości
- Wysokim poziomem przecieku widma, co objawia się licznymi ostrymi pikami na wykresie
- Niewielkim tłumieniem wyższych częstotliwości, co pozwala na zachowanie większej ilości szczegółów w górnym zakresie widma

4.10.3 Widmo okna trójkątnego (Bartletta)



Rysunek 44: Widmo częstotliwościowe z oknem trójkątnym (Bartletta)

Widmo z wykorzystaniem okna trójkątnego (rys. 44) wykazuje:

- Szerszy listek główny w porównaniu do okna prostokątnego (pogorszenie rozdzielczości częstotliwościowej)
- Zauważalnie lepsze tłumienie listków bocznych, co objawia się mniejszą liczbą i niższą amplitudą nieregularnych pików
- Obserwowany jest efekt wygładzania w porównaniu do okna prostokątnego
- Efektywniejszą redukcję przecieku widma, szczególnie w średnim zakresie częstotliwości (2000-6000 Hz)

4.10.4 Widmo okna Hamminga



Rysunek 45: Widmo częstotliwościowe z oknem Hamminga

Widmo z zastosowaniem okna Hamminga (rys. 45) charakteryzuje się:

- Dobrym kompromisem między rozdzielczością częstotliwościową a tłumieniem listków bocznych
- Znacznie lepszą redukcją nieregularności widma w porównaniu do okna prostokątnego i trójkątnego
- Większym wygładzeniem widma przy zachowaniu głównych cech charakterystycznych sygnału
- Porównanie z sygnałem oryginalnym pokazuje systematyczne tłumienie wysokich częstotliwości

4.10.5 Widmo okna Hanna



Rysunek 46: Widmo częstotliwościowe z oknem Hanna

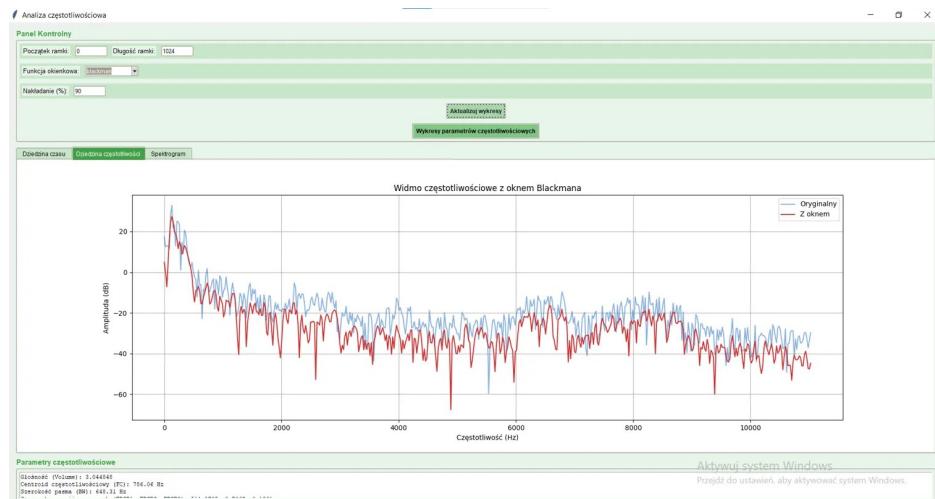
Widmo z oknem Hanna (rys. 46) wykazuje:

- Podobną charakterystykę do okna Hamminga, z nieco lepszym tłumieniem listków bocznych
- Szybszy spadek amplitudy listków bocznych
- Nieco większe wygładzenie sygnału, widoczne szczególnie w zakresie średnich częstotliwości
- Dobrą redukcję nieregularności widma przy zachowaniu istotnych cech sygnału

4.10.6 Widmo okna Blackmana

Widmo z oknem Blackmana (rys. 47) wykazuje:

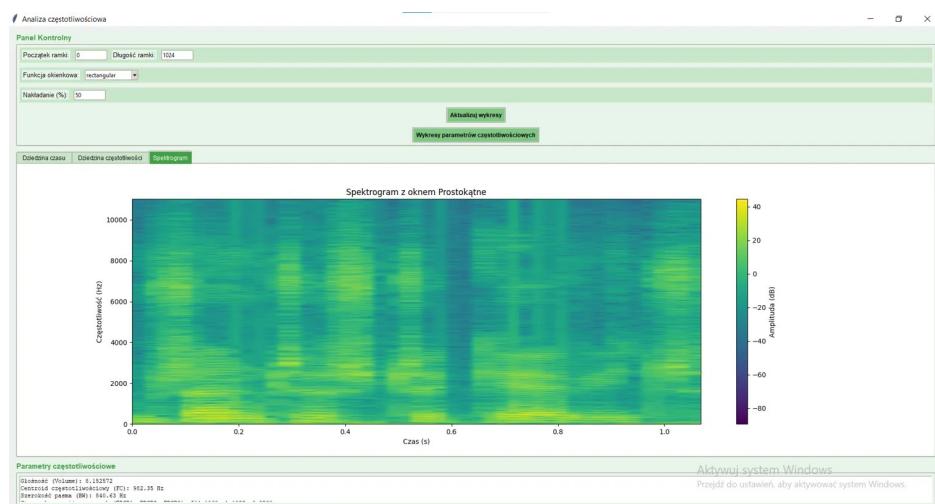
- Najszerzy listek główny spośród wszystkich badanych okien (najgorsza rozdzielcość częstotliwościowa)
- Najlepszą redukcję listków bocznych, co objawia się najbardziej wygłażonym widmem
- Najsilniejsze tłumienie przecieku widma, szczególnie w zakresach 3000-6000 Hz



Rysunek 47: Widmo częstotliwościowe z oknem Blackmana

- Znaczne osłabienie wysokich częstotliwości, co może prowadzić do utraty informacji o drobnych detalach w sygnale

4.10.7 Spektrogram okna prostokątnego



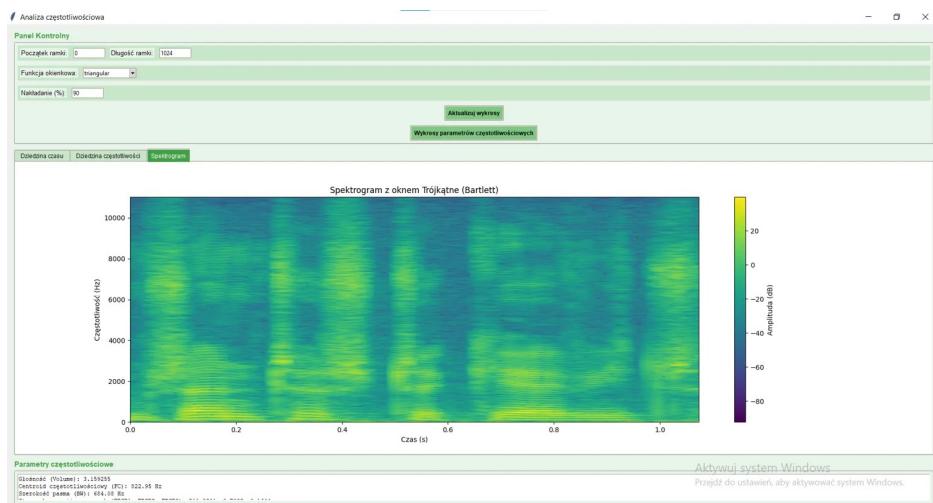
Rysunek 48: Spektrogram z oknem prostokątnym

Spektrogram z oknem prostokątnym (rys. 48) charakteryzuje się:

- Wysoką rozdzielczością częstotliwościową, co pozwala na precyzyjne rozróżnienie bliskich składowych harmonicznych

- Widocznym przeciekiem widma w postaci drobnych, pionowych artefaktów
- Dobrą widocznością struktury formantowej, szczególnie w zakresie 1000-3000 Hz
- Wyraźnie widocznymi przejściami między głoskami, co ułatwia segmentację sygnału mowy

4.10.8 Spektrogram okna trójkątnego (Bartletta)

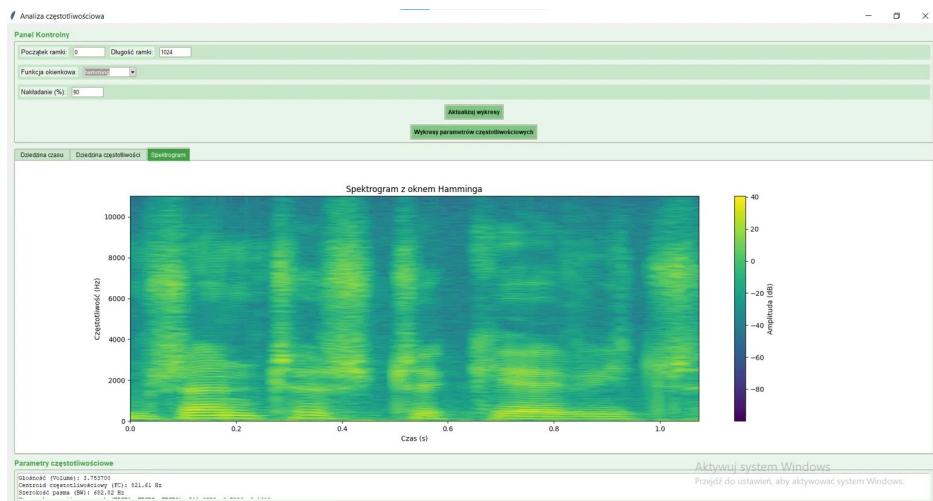


Rysunek 49: Spektrogram z oknem trójkątnym (Bartletta)

Spektrogram z oknem trójkątnym (rys. 49) wykazuje:

- Lepszą redukcję przecieku widma w porównaniu do okna prostokątnego
- Nieco rozmytą strukturę formantową, szczególnie dla wyższych formantów
- Bardziej jednorodne obszary odpowiadające samogłoskom
- Mniej zakłóceń w obszarach o niskiej energii (ciemniejsze obszary)

4.10.9 Spektrogram okna Hamminga

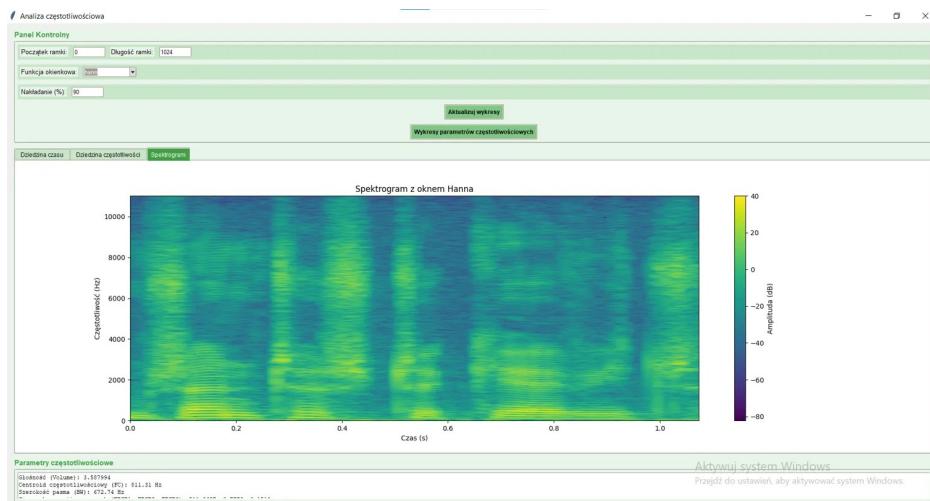


Rysunek 50: Spektrogram z oknem Hamminga

Spektrogram z oknem Hamminga (rys. 50) charakteryzuje się:

- Dobrym balansem między rozdzielczością a redukcją artefaktów
- Wyraźnymi formantami z mniejszą liczbą zakłóceń między nimi
- Lepszą wizualizacją zmian częstotliwości podstawowej (na dole spektrogramu)
- Skuteczniejszym tłumieniem składowych szumowych przy zachowaniu istotnych informacji o strukturze sygnału

4.10.10 Spektrogram okna Hanna

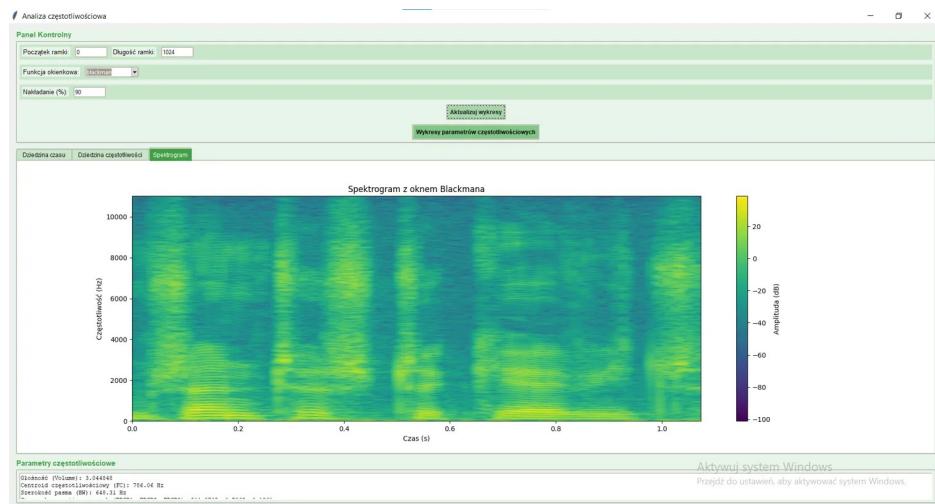


Rysunek 51: Spektrogram z oknem Hanna

Spektrogram z oknem Hanna (rys. 51) charakteryzuje się:

- Bardzo podobnymi właściwościami do okna Hamminga, z nieco lepszą redukcją przecieku widma
- Dobrze wyważonym kompromisem między rozdzielczością a wygładzeniem widma
- Wyraźnym odwzorowaniem struktur formantowych i przejść między różnymi głoskami
- Lepszą separacją składowych tonalnych od szumowych w porównaniu do okien prostokątnego i trójkątnego

4.10.11 Spektrogram okna Blackmana



Rysunek 52: Spektrogram z oknem Blackmana

Spektrogram z oknem Blackmana (rys. 52) wykazuje:

- Najlepszą redukcję przecieku widma i artefaktów
- Najwyższy kontrast między obszarami o wysokiej i niskiej energii
- Pewne rozmycie granic między formantami w wyniku zmniejszonej rozdzielczości częstotliwościowej
- Lepszą separację poszczególnych komponentów sygnału mowy, przy jednoczesnym zachowaniu ciągłości struktur formantowych

4.10.12 Wpływ funkcji okienkowych na parametry częstotliwościowe

Analiza parametrów częstotliwościowych pokazuje systematyczne zmiany wraz ze zmianą funkcji okienkowej:

Tabela 3: Porównanie parametrów częstotliwościowych dla różnych funkcji okienkowych

Parametr	Prost.	Trójk.	Hamm.	Hanna	Black.
Centroid częst. [Hz]	982,35	822,95	821,61	811,31	766,06
Szerokość pasma [Hz]	846,61	636,04	632,61	672,74	646,11
Charakt. widma	Nieregul.	Wygładz.	Bardziej wygl.	Podobne do H.	Najb. wygl.

Warto zauważyć, że wartość centroidu częstotliwościowego systematycznie maleje wraz z przejściem od okna prostokątnego do okna Blackmana. Wynika to z faktu, że okna o większym tłumieniu listków bocznych bardziej redukują energię w wyższych częstotliwościach, przesuwając tym samym „środek ciężkości” widma w kierunku niższych częstotliwości.

4.10.13 Wnioski

Przeprowadzona analiza potwierdza teoretyczne właściwości różnych funkcji okienkowych i ich wpływ na analizę częstotliwościową sygnału mowy:

- **Okno prostokątne** zapewnia najlepszą rozdzielczość częstotliwościową, ale generuje znaczny przeciek widma, co może utrudniać identyfikację słabszych składowych sygnału
- **Okno trójkątne (Bartletta)** stanowi pierwszy krok w kierunku kompromisu między rozdzielczością a przeciekiem widma, oferując wyraźną poprawę w redukcji artefaktów
- **Okno Hamminga** zapewnia dobry balans między rozdzielczością a tłumieniem listków bocznych, z dobrym tłumieniem najbliższych listków bocznych
- **Okno Hanna** oferuje podobne właściwości do okna Hamminga, ale z lepszym tłumieniem dalszych listków bocznych
- **Okno Blackmana** oferuje najlepszą redukcję przecieku widma, ale kosztem pogorszonej rozdzielczości częstotliwościowej, co może być korzystne przy analizie sygnałów o silnie zróżnicowanych amplitudach składowych

Wybór odpowiedniej funkcji okienkowej powinien być uzależniony od konkretnego zastosowania i potrzeby analitycznej

5 Podsumowanie i wnioski

5.1 Podsumowanie projektu

W ramach projektu rozszerzono aplikację do analizy plików dźwiękowych o zaawansowane funkcjonalności w dziedzinie częstotliwościowej. Zaimplementowano narzędzia do analizy widmowej sygnałów za pomocą Szybkiej Transformaty Fouriera (FFT), różne funkcje okienkowe, spektrogramy, analizę cepstralną oraz parametryzację w dziedzinie częstotliwości. Aplikacja umożliwia

kompleksową analizę dźwięku oraz wizualizację kluczowych parametrów częstotliwościowych zarówno na poziomie pojedynczych ramek, jak i w kontekście całego sygnału.

5.2 Kluczowe wnioski z przeprowadzonych eksperymentów

1. **Analiza głosów męskich i żeńskich** wykazała systematyczne różnice w parametrach częstotliwościowych. Głos żeński charakteryzuje się wyższymi wartościami centroidu częstotliwościowego, bardziej równomiernym rozkładem energii pomiędzy pasmami częstotliwości oraz niższą płaskością widma.
2. **Badanie formantów samogłosek** potwierdziło ścisły związek między wartościami formantów a cechami artykulacyjnymi. Zidentyfikowano charakterystyczne wartości F1 i F2 dla poszczególnych samogłosek, odzwierciedlające ich pozycję w systemie fonetycznym.
3. **Analiza cepstralna** umożliwiła skuteczne wyznaczanie częstotliwości podstawowej (F0) w sygnałach mowy, mimo pewnych ograniczeń metodologicznych w specyficznych kontekstach fonetycznych. Zaobserwowano, że kontekst fonetyczny silnie wpływa na realizację F0.
4. **Porównanie funkcji okienkowych** wykazało różne właściwości poszczególnych okien, wpływające na jakość analizy spektralnej. Okno prostokątne oferuje najlepszą rozdzielcość częstotliwościową kosztem wysokiego przecieku widma, podczas gdy okno Blackmana zapewnia najlepszą redukcję przecieku widma kosztem rozdzielcości. Okna Hamminga i Hanna stanowią optymalny kompromis dla większości zastosowań.
5. **Analiza różnic między głoskami** potwierdziła fundamentalnie odmienne właściwości akustyczne samogłosek i spółgłosek, co odzwierciedla ich różnice artykulacyjne. Samogłoski charakteryzują się wyraźną strukturą formantową i stabilnością czasową, podczas gdy spółgłoski wykazują nieregularną strukturę widmową i znacznie krótszy czas trwania.

5.3 Kierunki dalszego rozwoju

Zaimplementowana aplikacja stanowi solidną podstawę do dalszego rozwoju w kierunku systemów automatycznego rozpoznawania mowy, identyfikacji

mówcy, czy analizy emocji w głosie. Przyszłe rozszerzenia mogłyby obejmować:

- Implementację sieci neuronowych do klasyfikacji dźwięków
- Rozbudowę modułu analizy parametrycznej o współczynniki mel-cepstralne (MFCC)
- Dodanie funkcjonalności związanych z syntezą mowy
- Rozszerzenie możliwości wizualizacyjnych o reprezentacje trójwymiarowe

Podsumowując, projekt z powodzeniem zrealizował wszystkie założone cele, dostarczając efektywne narzędzie do analizy częstotliwościowej sygnałów dźwiękowych, co potwierdziły przeprowadzone eksperymenty praktyczne.