

Dokumentacja

Analiza i przetwarzanie dźwięku

Projekt 2 - cechy sygnału audio w dziedzinie częstotliwości

Bogumiła Okrojek

9 maja 2025

Spis treści

1	Wstęp	2
2	Opis aplikacji	3
2.1	Wykorzystane technologie i struktura	3
2.2	Interfejs użytkownika	4
3	Opis metod	5
3.1	Transformacja na dziedzinę częstotliwości	5
3.1.1	Wykres widma częstotliwościowego	5
3.1.2	Wykres sygnału w dziedzinie czasu i częstotliwości po funkcji okna	6
3.2	Parametry z dziedziny częstotliwości	8
3.2.1	Głośność (Volume)	8
3.2.2	Frequency Centroid (FC)	10
3.2.3	Effective Bandwidth (BW)	11
3.2.4	Band Energy (BE)	13
3.2.5	Band Energy Ratio (BER or ERSB)	14
3.2.6	Spectral Flatness Measure (SFM)	16
3.2.7	Spectral Crest Factor (SCF)	17
4	Wnioski	19
4.1	Parametry w dziedzinie częstotliwości dla różnych głosek	19
4.2	Wnioski na temat wpływu funkcji okienkowych na widmo sygnału	22
5	Podsumowanie	26

1 Wstęp

Projekt stanowi aplikację graficzną (GUI) stworzoną w języku **Python** z wykorzystaniem biblioteki **Tkinter**. Narzędzie umożliwia wczytywanie oraz analizę plików audio typu **.wav**. System pozwala na wizualizację parametrów dźwięku zarówno w **diedzinie czasu**, jak i **diedzinie częstotliwości**.

Cechy sygnału w dziedzinie czasu zostały zaimplementowane i szczegółowo opisane w **Projekcie 1**. Dokumentacja tego projektu obejmuje m.in.:

- Cechy sygnału audio na poziomie ramki (Frame-Level):
 - Volume – głośność
 - Short Time Energy (STE)
 - Zero Crossing Rate (ZCR)
 - Silent Ratio (SR)
 - Częstotliwość tonu podstawowego F0: z autokorelacji oraz z funkcji AMDF
- Cechy sygnału audio na poziomie klipu (Clip-Level):
 - VSTD, VDR, VU (bazujące na głośności)
 - LSTER, Energy Entropy (bazujące na energii)
 - ZSTD, HZCRR (bazujące na ZCR)
- Analizy dodatkowe:
 - Detekcja ciszy
 - Określenie fragmentów dźwięcznych i bezdźwięcznych
 - Klasyfikacja fragmentów jako muzyka lub mowa
 - Odtwarzanie i wczytywanie plików audio

Uwaga: W niniejszym raporcie opisane zostaną jedynie **dodane funkcjonalności** w zakresie zakładek:

- Transformacja na dziedzinę częstotliwości
- Parametry z dziedziny częstotliwości

oraz zastosowane GUI dotyczące tej części. Pozostałe elementy zostały omówione w poprzedniej dokumentacji.

2 Opis aplikacji

Aplikacja została wykonana w języku programowania **Python**, który jest powszechnie stosowany w analizie sygnałów, przetwarzaniu dźwięku oraz tworzeniu graficznych interfejsów użytkownika. Jego elastyczność i bogactwo bibliotek sprawiają, że idealnie nadaje się do szybkiego prototypowania oraz tworzenia aplikacji edukacyjnych i badawczych.

Do realizacji interfejsu graficznego (GUI) wykorzystano bibliotekę **Tkinter**, będącą standardowym narzędziem GUI w Pythonie. Umożliwia ona tworzenie okien, przycisków, suwaków, menu i innych komponentów, co pozwoliło na zbudowanie intuicyjnego i przejrzystego systemu do analizy plików audio.

2.1 Wykorzystane technologie i struktura

W projekcie zastosowano również następujące biblioteki:

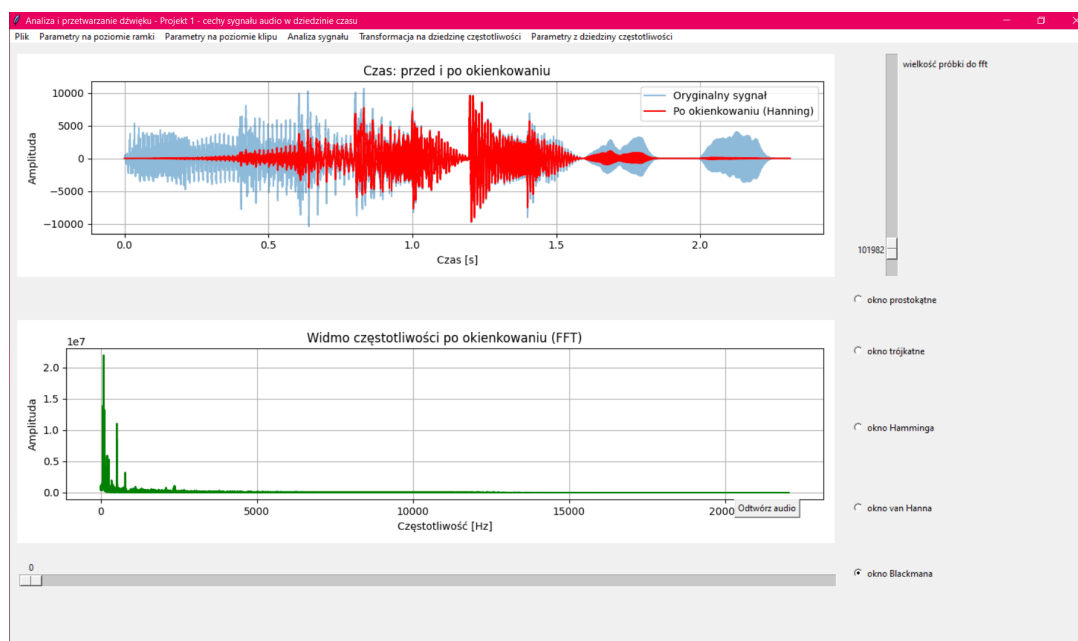
- **NumPy** – do obliczeń numerycznych, takich jak przekształcenia Fouriera, filtracja czy obliczenia statystyczne.
- **SciPy** – do wczytywania i zapisu plików dźwiękowych w formacie `.wav`.
- **Matplotlib** – do generowania wykresów i wizualizacji parametrów sygnału w dziedzinie czasu i częstotliwości.
- **sounddevice** – do odtwarzania dźwięku bezpośrednio w aplikacji.

Decyzja o wyborze tych bibliotek wynikała z ich popularności, szerokiego wsparcia społeczności oraz dobrej dokumentacji. Dzięki temu aplikacja jest nie tylko funkcjonalna, ale także łatwa do rozbudowy i modyfikacji.

Główne elementy interfejsu widoczne na Rysunek 1 to:

- **Główne okno interfejsu** – zbudowane na bazie klasy `Tk()` z elementami, takimi jak paski menu, suwaki i przyciski do nawigacji,
- **Moduł funkcji analizy sygnału** – zawiera funkcje do analizy sygnałów, takie jak **Short Time Energy (STE)**, **Zero Crossing Rate (ZCR)**, czy wykrywanie tonu podstawowego za pomocą algorytmów autokorelacji,
- **Moduł przetwarzania w dziedzinie częstotliwości** – obejmuje analizę widmową sygnału (FFT), funkcję częstotliwości centroidu, pasmo efektywne, energię w paśmie, współczynnik spłaszczenia widma oraz inne cechy widmowe,

- **Obsługa funkcji okna** – użytkownik może wybrać jeden z pięciu typów funkcji okna (prostokątne, trójkątne, Hamminga, Hann’a, Blackmana), które wpływają na wyniki analizy widmowej. Wybór okna realizowany jest poprzez przyciski radiowe w interfejsie,
- **Moduł wizualizacji** – umożliwia wyświetlanie wykresów wyników analizy przy użyciu biblioteki `matplotlib`, oddzielnie dla sygnału w dziedzinie czasu oraz częstotliwości.



Rysunek 1: GUI aplikacji

2.2 Interfejs użytkownika

Interfejs graficzny został zaprojektowany w sposób intuicyjny i funkcjonalny, oferując:

- Menu podzielone na sekcje, takie jak Plik, Parametry na poziomie klipu, Parametry na poziomie ramki, Analiza sygnału, Transformacja na dziedzinę częstotliwości oraz Parametry z dziedziny częstotliwości, umożliwiające użytkownikom dostęp do szerokiego zakresu funkcji analitycznych,
- Suwaki umożliwiające interaktywną regulację parametrów, takich jak poziom RMS, poziom ZCR czy wielkość ramki analizy,
- Przyciskowe przełączniki (radio buttons) pozwalające na wybór rodzaju funkcji okna, co daje użytkownikowi kontrolę nad kształtem okna używanego przy analizie widmowej,
- Panel wizualizacji, który prezentuje wyniki analizy w formie wykresów – zarówno dla danych surowych, jak i po przekształceniu Fouriera. Dwa oddzielne obszary wykresów (dla

dziedziny czasu i częstotliwości) zapewniają przejrzystość oraz możliwość bezpośredniego porównania wyników.

3 Opis metod

3.1 Transformacja na dziedzinę częstotliwości

Analiza sygnału audio w dziedzinie częstotliwości pozwala na uzyskanie informacji o składnikach częstotliwościowych obecnych w danym sygnale. W przeciwieństwie do analizy w dziedzinie czasu, która koncentruje się na zmianach amplitudy sygnału w czasie, transformacja do dziedziny częstotliwości umożliwia identyfikację dominujących częstotliwości, pasma energetycznego oraz cech widmowych. Proces ten jest szczególnie istotny w zadaniach takich jak rozpoznawanie mowy, klasyfikacja dźwięków, detekcja tonów oraz analiza muzyki.

3.1.1 Wykres widma częstotliwościowego

Wykres widma częstotliwościowego przedstawia rozkład amplitud w funkcji częstotliwości dla analizowanego fragmentu sygnału audio. Jest to jedno z najczęściej wykorzystywanych narzędzi w analizie sygnałów dźwiękowych, pozwalające na uzyskanie informacji o składnikach częstotliwościowych sygnału. W aplikacji realizowana jest ta operacja za pomocą szybkiej transformaty Fouriera (FFT), która przekształca sygnał z dziedziny czasu do dziedziny częstotliwości, umożliwiając wizualizację jego składników częstotliwościowych.

Idea i metoda

W aplikacji analiza widma częstotliwościowego odbywa się na wybranym fragmencie sygnału audio. Proces polega na następujących krokach:

- **Podział sygnału** – sygnał audio jest dzielony na małe fragmenty, tzw. ramki, które poddawane są dalszej analizie. Użytkownik może wybrać wielkość ramki oraz funkcję okna, co ma na celu ograniczenie efektu *spectral leakage*, czyli przecieku widmowego.
- **Zastosowanie funkcji okna** – przed przeprowadzeniem transformacji Fouriera, sygnał jest przetwarzany przez funkcję okna, która zmniejsza amplitudę próbek na brzegach ramki. Dzięki temu unikamy zniekształceń widma, które mogą wystąpić, gdy fragment sygnału jest „obcięty” na brzegach.
- **Transformacja Fouriera (FFT)** – na każdym z fragmentów sygnału wykonuje się szybką transformatę Fouriera, co umożliwia przekształcenie sygnału z dziedziny czasu do dziedziny

częstotliwości. W wyniku tej operacji otrzymujemy widmo amplitudowe, które przedstawia zależność między częstotliwościami a ich amplitudą w analizowanym fragmencie.

- **Wykres widma** – wynik FFT jest następnie prezentowany na wykresie, na którym oś X reprezentuje częstotliwości (w Hz), a oś Y – amplitudy tych częstotliwości. Tylko pierwsza połowa widma jest wyświetlana, ponieważ drugi fragment jest lustrzanym odbiciem pierwszego.
- **Wykrywanie pików** – dla lepszej analizy, na wykresie wyświetlane są największe piki w widmie. Użytkownik może zobaczyć wartości częstotliwości, które dominują w analizowanym sygnale.

Zastosowana metoda pozwala użytkownikowi na dokładną analizę składników częstotliwościowych sygnału audio i identyfikację jego dominujących częstotliwości. Ponadto, użytkownik może manipulować parametrami analizy, takimi jak funkcja okna, aby zobaczyć, jak wpływają one na uzyskane wyniki.

Wady i ograniczenia

Chociaż metoda FFT jest bardzo popularna i skuteczna w analizie częstotliwościowej, ma pewne ograniczenia:

- **Wielkość ramki** – wyniki analizy FFT zależą od wielkości ramki, która jest stosowana do analizy. Zbyt małe ramki mogą prowadzić do niewielkiej rozdzielczości częstotliwościowej, co utrudnia identyfikację dokładnych składników częstotliwościowych. Zbyt duże ramki mogą z kolei utrudnić uchwycenie szybkich zmian w sygnale.
- **Spectral leakage** – mimo użycia funkcji okna, wciąż może występować efekt przecieku widmowego, szczególnie jeśli sygnał jest nieciągły w obrębie ramki. Efekt ten może prowadzić do zniekształceń wyników analizy.
- **Brak rozdzielczości czasowej** – FFT pozwala na doskonałą analizę częstotliwości, ale nie dostarcza informacji o czasie, w którym konkretne składniki częstotliwościowe występują. Dlatego do analizy zmian w czasie często stosuje się podejście oparte na oknach przesuwanych, takie jak STFT (Short-Time Fourier Transform).

3.1.2 Wykres sygnału w dziedzinie czasu i częstotliwości po funkcji okna

Ta funkcjonalność umożliwia użytkownikowi obserwację wpływu zastosowania funkcji okna na sygnał audio – zarówno w dziedzinie czasu, jak i częstotliwości. Operacja ta pozwala porównać

oryginalny przebieg sygnału z jego wersją po zastosowaniu funkcji okna, a następnie przeanalizować, jak modyfikacja ta wpływa na wynik transformacji Fouriera. Dzięki temu użytkownik może lepiej zrozumieć znaczenie okienkowania sygnału w kontekście redukcji zniekształceń widmowych, takich jak *spectral leakage*.

Idea i metoda

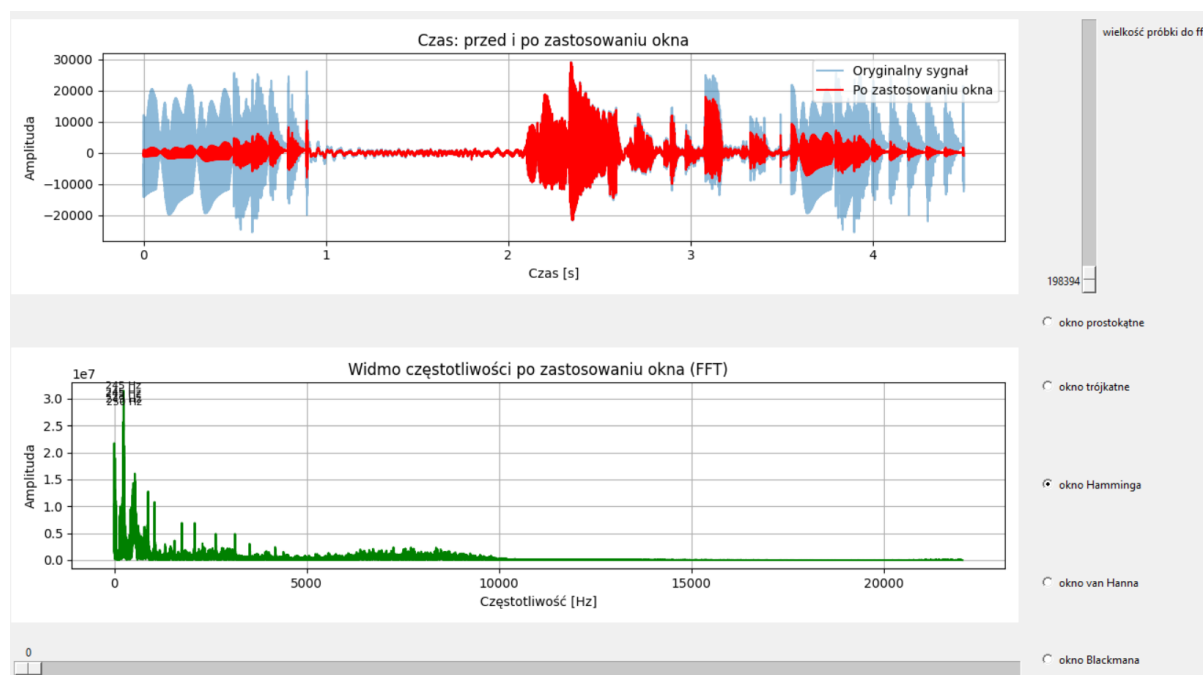
Metoda analizy sygnału przed i po zastosowaniu funkcji okna przebiega według następujących etapów:

- **Wybór fragmentu sygnału** – użytkownik wybiera początek analizowanego fragmentu sygnału oraz jego długość przy pomocy suwaków interfejsu. Dzięki temu może elastycznie analizować dowolne partie nagrania.
- **Zastosowanie funkcji okna** – na wybranym fragmencie stosowana jest wybrana przez użytkownika funkcja okna, np. Hanninga, Hamminga, Blackmana itp. Funkcja okna tłumi wartości sygnału przy jego brzegach, co zmniejsza ryzyko wystąpienia nieciągłości sygnału na końcach analizowanego fragmentu.
- **Wizualizacja sygnału w dziedzinie czasu** – na pierwszym wykresie prezentowany jest oryginalny sygnał oraz jego wersja po zastosowaniu funkcji okna. Pozwala to zaobserwować efekt wygładzenia krawędzi oraz zmianę kształtu przebiegu czasowego.
- **Transformacja Fouriera i analiza widmowa** – na tym samym sygnale (po okienkowaniu) wykonywana jest szybka transformata Fouriera (FFT), która pozwala na uzyskanie widma amplitudowego. Widmo to przedstawia rozkład częstotliwościowy sygnału, już zredukowany pod względem efektu przecieku widmowego.
- **Wizualizacja w dziedzinie częstotliwości** – drugi wykres przedstawia wynik transformacji FFT. Pokazuje on, jak zastosowanie okna wpływa na kształt i dokładność widma sygnału.

Na wykresie przedstawionym na Rysunek 2 widać, że po zastosowaniu wybranego okna (np. Hanninga) na sygnał czasowy, jego widmo częstotliwościowe ulega znacznemu wygładzeniu. Dodatkowo, na wykresie spektrum (prawej stronie Rysunek 2) zaznaczono pięć najwyższych wartości amplitudy widma, których częstotliwości podano bezpośrednio nad odpowiadającymi im szczytami. Można zauważyć, że dla tego sygnału najwyższe amplitudy są dla wartości $\approx 250\text{Hz}$.

Wady i ograniczenia

- **Wpływ wyboru funkcji okna** – różne funkcje okna mają różne właściwości: np. lepsze



Rysunek 2: Wykres dźwięku w dziedzinie czas i w dziedzinie częstotliwości

tłumienie bocznych pasm, ale gorszą rozdzielczość częstotliwościową. Użytkownik musi świadomie dobrać funkcję do rodzaju analizowanego sygnału.

- **Wpływ długości sygnału** – długość analizowanego fragmentu (próbki) wpływa na rozdzielczość częstotliwościową FFT. Zbyt krótki fragment może nie pozwolić na dokładne rozróżnienie sąsiadujących częstotliwości.
- **Sygnały nieliniowe lub zmienne w czasie** – analiza pojedynczego, nieruchomego okna czasowego może nie oddać pełnej dynamiki sygnałów zmiennych w czasie. W takich przypadkach lepszym rozwiązaniem może być analiza krótkoczasowa (STFT).

3.2 Parametry z dziedziny częstotliwości

Parametry wyznaczone w dziedzinie częstotliwości umożliwiają dokładniejszą analizę właściwości sygnału, które nie są bezpośrednio widoczne w dziedzinie czasu. W niniejszej sekcji omówiono metody obliczania oraz interpretację wybranych cech spektralnych, takich jak: głośność (Volume), środek ciężkości widma (Frequency Centroid), czy szerokość pasma (Bandwidth). Parametry te stanowią podstawę do dalszej klasyfikacji i identyfikacji sygnałów akustycznych.

3.2.1 Głośność (Volume)

W tej funkcjonalności przedstawiona jest analiza głośności sygnału audio, lecz w odróżnieniu od klasycznej metody w dziedzinie czasu (bazującej na RMS), tutaj głośność estymowana jest na

podstawie mocy widma (ang. *spectral power*) w dziedzinie częstotliwości. Jest to alternatywna forma oceny energii sygnału, wykorzystująca przekształcenie Fouriera dla poszczególnych ramek.

Idea i metoda

Celem metody jest wyznaczenie średniej energii spektralnej sygnału audio w kolejnych ramach czasowych. Operacja przebiega według następujących kroków:

- **Podział sygnału na ramki** – analizowany fragment sygnału jest dzielony na ramki o ustalonej długości (np. 512 próbek). Niepełne ramki na końcu są pomijane.
- **Zastosowanie funkcji okna** – na każdą ramkę nakładana jest wybrana funkcja okna (np. Hanninga, Hamminga), co redukuje efekty przecieku widmowego i poprawia dokładność analizy częstotliwościowej.
- **Transformacja FFT i obliczenie mocy widma** – dla każdej okienkowanej ramki wykonywana jest szybka transformata Fouriera (FFT), a następnie wyznaczana jest moc widma jako suma kwadratów modułów wartości zespolonych (czyli $|FFT|^2$):

$$P = \frac{1}{N} \sum_{i=1}^N |X_i|^2 \quad (1)$$

gdzie X_i to kolejne wartości transformaty Fouriera dla danej ramki, a N to liczba próbek w ramce.

- **Wizualizacja** – wartości głośności są wykreślane w funkcji czasu. Oś X przedstawia czas w sekundach (liczony od początku ramki), a oś Y prezentuje wartość średniej energii widmowej.

Zalety i zastosowania

- Analiza głośności w dziedzinie częstotliwości pozwala lepiej ocenić, jak zmienia się całkowita energia sygnału niezależnie od jego fazy i przebiegu czasowego.
- Jest to przydatne m.in. w zadaniach klasyfikacji dźwięków, wykrywaniu zmian głośności oraz estymacji intensywności w muzyce i mowie.

Wady i ograniczenia

- Wynik zależy od wyboru funkcji okna oraz długości ramki – różne ustawienia mogą wpływać na dokładność estymacji energii.

- Energia całkowita w dziedzinie częstotliwości może być wrażliwa na zakłócenia lub obecność hałasu, dlatego przy interpretacji wyników należy uwzględniać także charakter sygnału źródłowego.
- Metoda ta nie uwzględnia subiektywnej percepcji głośności przez człowieka, która zależy również od charakterystyki częstotliwościowej ucha.

Analiza głośności w dziedzinie częstotliwości stanowi cenne rozszerzenie klasycznych metod opartych na dziedzinie czasu, pozwalając na głębszą eksplorację zawartości energetycznej sygnału audio.

3.2.2 Frequency Centroid (FC)

Frequency Centroid, czyli środek ciężkości widma, jest jednym z najważniejszych parametrów widmowych stosowanych w analizie sygnałów audio. Parametr ten wskazuje, wokół jakiej częstotliwości skupiona jest energia spektralna danej ramki sygnału. W praktyce Frequency Centroid jest często interpretowany jako miernik „jasności” dźwięku – wyższe wartości sugerują obecność bardziej energetycznych składowych wysokoczęstotliwościowych.

Idea i metoda

Analiza przebiega w następujących krokach:

- **Podział sygnału na ramki** – fragment sygnału jest dzielony na niepokrywające się ramki o ustalonym rozmiarze, np. 512 próbek.
- **Zastosowanie funkcji okna** – na każdą ramkę nakładana jest funkcja okna (np. Hanninga), co redukuje przeciek widmowy i poprawia precyzję analizy.
- **Transformacja Fouriera** – wykonywana jest szybka transformata Fouriera (FFT) każdej ramki po okienkowaniu, a następnie obliczany jest moduł wartości zespolonych (amplituda widma).
- **Obliczenie FC** – wyznaczenie środka ciężkości widma następuje na podstawie wzoru:

$$FC = \frac{\sum_{i=1}^N f_i \cdot |X_i|}{\sum_{i=1}^N |X_i|} \quad (2)$$

gdzie:

- f_i to częstotliwość odpowiadająca i -tej próbce widma,
- $|X_i|$ to amplituda spektrum dla danej częstotliwości f_i ,

- N to liczba próbek w analizowanej połowie widma (ponieważ widmo sygnału rzeczywistego jest symetryczne).
- **Wizualizacja** – wyznaczone wartości FC są nanoszone na wykres w funkcji czasu (oś X). Oś Y reprezentuje wartość środka ciężkości widma wyrażoną w Hz.

Zastosowania

Frequency Centroid znajduje szerokie zastosowanie w:

- analizie i rozpoznawaniu dźwięków,
- analizie muzycznej (różnicowanie instrumentów, gatunków),
- estymacji barwy dźwięku oraz jego „jasności”,
- wykrywaniu zmian tonalnych i segmentacji mowy lub muzyki.

Wady i ograniczenia

- FC może być zakłócany przez szum i obecność wielu składowych o zbliżonej energii, co może prowadzić do nieintuicyjnych wyników.
- Wynik analizy zależy od długości ramki oraz typu zastosowanego okna – inne ustawienia mogą znacząco wpłynąć na interpretację wartości FC.
- Nie daje pełnej informacji o charakterze sygnału – często analizowany łącznie z innymi parametrami widmowymi (np. Spectral Spread czy Spectral Flatness).

Frequency Centroid stanowi efektywną miarę lokalizacji energii w widmie i może być wykorzystywany jako kluczowy element w klasyfikacji dźwięków oraz ekstrakcji cech sygnału.

3.2.3 Effective Bandwidth (BW)

Effective Bandwidth, czyli efektywna szerokość pasma, opisuje rozrzut energii sygnału w dziedzinie częstotliwości wokół jego środka ciężkości (ang. Frequency Centroid). Parametr ten odpowiada odchyleniu standardowemu rozkładu mocy widma i informuje o tym, jak szeroko rozciągnięte są składowe częstotliwościowe danego fragmentu sygnału. W kontekście sygnałów muzycznych czy mowy, BW pozwala lepiej scharakteryzować barwę dźwięku i stopień jego „rozmycia” spektralnego.

Idea i metoda

Proces obliczania Effective Bandwidth obejmuje następujące kroki:

- **Podział sygnału na ramki** – analizowany fragment sygnału dzielony jest na ramki o zadanym rozmiarze.
- **Zastosowanie funkcji okna** – każda ramka poddawana jest okienkowaniu (np. oknem Hanninga), co ogranicza efekty przecieku widmowego i poprawia jakość transformacji FFT.
- **Transformacja Fouriera** – przeprowadzana jest szybka transformata Fouriera (FFT), z której analizowana jest tylko pierwsza połowa (symetria widma).
- **Obliczenie środka ciężkości (Centroid)** – punkt odniesienia dla wyznaczenia rozrzutu, obliczany ze wzoru:

$$FC = \frac{\sum_{i=1}^N f_i \cdot |X_i|}{\sum_{i=1}^N |X_i|} \quad (3)$$

- **Wyznaczenie szerokości pasma (BW)** – jako pierwiastek z wariancji rozkładu amplitud względem centroidu:

$$BW = \sqrt{\frac{\sum_{i=1}^N (f_i - FC)^2 \cdot |X_i|}{\sum_{i=1}^N |X_i|}} \quad (4)$$

gdzie:

- f_i – częstotliwość przypisana do i -tej próbki widma,
 - $|X_i|$ – amplituda FFT dla tej częstotliwości,
 - FC – Frequency Centroid dla danej ramki.
- **Wizualizacja** – wartości BW są nanoszone na wykres względem czasu rozpoczęcia każdej ramki.

Zastosowania

Efektywna szerokość pasma znajduje zastosowanie m.in. w:

- analizie barwy dźwięku i klasyfikacji instrumentów,
- wykrywaniu różnic tonalnych i jakości nagrań,
- segmentacji i klasyfikacji fragmentów mowy lub muzyki.

Wady i ograniczenia

- BW zależy bezpośrednio od poprawności obliczenia Frequency Centroid – błędy w centroidzie wpływają na wynik BW,

- nie odróżnia lokalnych pików w widmie – opisuje jedynie ogólny rozrzut,
- w przypadku sygnałów o niskiej energii w wysokich częstotliwościach, BW może być zaniżone i nieadekwatne.

Effective Bandwidth to cenne narzędzie do oceny szerokości widma energetycznego i charakterystyki tonalnej dźwięku, szczególnie gdy analizowany sygnał zawiera skomplikowane struktury widmowe.

3.2.4 Band Energy (BE)

Band Energy (energia pasma) to parametr opisujący ilość energii zawartą w określonym zakresie częstotliwości ($[f_{\text{low}}, f_{\text{high}}]$) w sygnale audio. Pozwala on na ocenę, jak silnie reprezentowane są konkretne zakresy pasma w analizowanym dźwięku, co ma zastosowanie m.in. w klasyfikacji typów dźwięków, analizie muzyki, detekcji mowy czy identyfikacji szumów.

Idea i metoda

Obliczanie Band Energy w aplikacji przebiega według następującego schematu:

- **Ramkowanie sygnału** – sygnał dzielony jest na niepokrywające się ramki o zadanym rozmiarze (np. 1024 próbek).
- **Okienkowanie sygnału** – na każdą ramkę nakładana jest funkcja okna (np. Hanninga), co redukuje przeciek widmowy i poprawia analizę FFT.
- **Transformacja FFT i widmo mocy** – dla każdej ramki obliczana jest szybka transformata Fouriera (FFT), po czym z amplitudy widma tworzone jest widmo mocy poprzez podniesienie wartości bezwzględnych do kwadratu:

$$P(f) = |X(f)|^2 \quad (5)$$

- **Filtracja pasma** – spośród wszystkich częstotliwości wybierane są tylko te, które mieszczą się w zadanym zakresie $[f_{\text{low}}, f_{\text{high}}]$.
- **Suma energii w paśmie** – energia sygnału w wybranym paśmie obliczana jest jako suma wartości widma mocy:

$$BE = \frac{1}{N} \sum_{f=f_{\text{low}}}^{f_{\text{high}}} P(f) \quad (6)$$

gdzie N to liczba próbek w ramce, a $P(f)$ to moc widmowa dla częstotliwości f .

- **Wizualizacja** – wartości Band Energy są prezentowane na wykresie względem czasu rozpoczęcia każdej ramki.

Zastosowania

Band Energy jest przydatne w wielu kontekstach:

- analiza mowy – np. różnicowanie samogłosek i spółgłosek,
- analiza muzyczna – np. identyfikacja partii basowych lub wysokotonowych,
- detekcja hałasów lub tonów o określonej charakterystyce,
- analiza emocji w głosie – różne emocje mogą powodować zmiany energii w różnych pasmach.

Wady i ograniczenia

- Wymaga dobrania pasma $[f_{\text{low}}, f_{\text{high}}]$ odpowiedniego do danego zastosowania – nie ma jednego uniwersalnego zakresu,
- Wysoka zależność od jakości FFT i parametrów ramki (rozmiar, okno),
- Może nie być wystarczająco informatywne w przypadku złożonych sygnałów z rozproszoną energią.

Parametr *Band Energy* stanowi prosty, lecz skuteczny wskaźnik do oceny koncentracji energii w określonych zakresach częstotliwości, szczególnie przydatny w zadaniach klasyfikacyjnych i diagnostycznych.

3.2.5 Band Energy Ratio (BER or ERSB)

Band Energy Ratio (BER) lub *Energy Ratio Subbands* (ERSB) to wskaźnik opisujący stosunek energii w wybranych pasmach częstotliwości do całkowitej energii sygnału. Używa się go do analizy, jak energia sygnału jest rozłożona w różnych zakresach częstotliwości, co jest przydatne w rozpoznawaniu cech sygnałów, klasyfikacji dźwięków czy w analizie mowy.

Idea i metoda

Obliczanie wskaźnika BER (ERSB) w aplikacji przebiega według następującego schematu:

- **Ramkowanie sygnału** – sygnał dzielony jest na ramki o zadanym rozmiarze. Dla każdej ramki obliczane są wskaźniki stosunku energii dla różnych pasm częstotliwości.

- **Okienkowanie sygnału** – na każdą ramkę nakładane jest okno Hanninga, aby zmniejszyć wpływ przecieku widmowego na transformację FFT.
- **Transformacja FFT i widmo mocy** – dla każdej ramki sygnału obliczany jest widmo mocy poprzez wykonanie transformacji Fouriera (FFT) i obliczenie kwadratu amplitudy dla każdej częstotliwości.
- **Podział na pasma** – częstotliwości sygnału są dzielone na subbandy. Zwykle stosowane pasma to:
 - 0 – 630 Hz (wersja ERSB1),
 - 630 – 1720 Hz (wersja ERSB2),
 - 1720 – 4400 Hz (wersja ERSB3),
 - Dodatkowo, dla wyższych częstotliwości, może zostać zastosowane pasmo $4400 - 0.5 \times f_{\text{sample rate}}$ (wersja ERSB4) w przypadku próbkowania powyżej 8800 Hz.
- **Obliczenie stosunku energii** – dla każdego pasma obliczany jest stosunek energii w tym paśmie do całkowitej energii sygnału w ramce:

$$\text{ERSB} = \frac{\sum \text{energia w paśmie}}{\text{energia całkowita}} \quad (7)$$

Wartości ERSB dla każdego pasma są obliczane osobno i zapisywane dla każdej ramki.

- **Wizualizacja** – dla każdej ramki obliczane są wartości ERSB, które następnie są wykreslane na wykresie w zależności od czasu.

Zastosowania

Band Energy Ratio (ERSB) jest użyteczny w kontekście:

- *Analiza mowy* – rozkład energii w różnych pasmach częstotliwości jest pomocny w rozróżnianiu głosów, intonacji i innych cech mowy.
- *Rozpoznawanie muzyki* – stosunek energii w różnych pasmach pozwala na identyfikację instrumentów lub tonów w utworach muzycznych.
- *Detekcja hałasu* – analiza rozkładu energii w pasmach pomaga w identyfikacji szumów o określonym widmie.

Wady i ograniczenia

- *Dobór pasm* – wyniki analizy zależą od wyboru pasm częstotliwości, co może być różne w zależności od zastosowania.
- *Złożoność obliczeniowa* – dla sygnałów o dużych próbkach, obliczenia FFT i stosunki energii mogą być zasobożerne.
- *Zależność od jakości próbkowania* – dla niskiej jakości próbkowania (np. zbyt mała częstotliwość próbkowania) analiza pasm wyższych częstotliwości może być mniej dokładna.

Band Energy Ratio (ERSB) stanowi skuteczną metodę do analizy rozkładu energii w sygnale w różnych pasmach częstotliwości, przydatną szczególnie w zadaniach związanych z rozpoznawaniem i klasyfikowaniem dźwięków oraz analizy mowy.

3.2.6 Spectral Flatness Measure (SFM)

Miara *Spectral Flatness* (SFM) jest używana do oceny "płaskości" widma sygnału. SFM mierzy, jak bardzo widmo sygnału przypomina szum (płaskie widmo) w porównaniu do dźwięku tonalnego (o wyraźnych, wąskich pikach). Wartości SFM są użyteczne do rozróżniania między dźwiękami tonalnymi (np. muzyka, głos) a niskoczęstotliwościowymi szumami (np. szumy tła, dźwięki losowe).

Idea i metoda

Obliczenie miary SFM odbywa się poprzez następujące kroki:

- **Ramkowanie sygnału** – sygnał jest dzielony na ramki o zadanym rozmiarze, co pozwala na lokalną analizę w dziedzinie czasu.
- **Okienkowanie sygnału** – dla każdej ramki nakłada się okno Hanninga w celu zmniejszenia wpływu przecieku widmowego podczas obliczeń FFT.
- **Transformacja FFT i widmo mocy** – dla każdej ramki obliczane jest widmo sygnału przy użyciu transformacji Fouriera (FFT), a następnie kwadratu amplitudy (mocy) każdej częstotliwości. Dodanie małej wartości (1e-6) do widma mocy zapobiega dzieleniu przez zero w dalszych obliczeniach.
- **Obliczanie miary SFM** – miara SFM jest obliczana jako stosunek średniej geometrycznej do średniej arytmetycznej widma mocy:

$$\text{SFM} = \frac{\exp\left(\frac{1}{N} \sum_{i=1}^N \log(x_i)\right)}{\frac{1}{N} \sum_{i=1}^N x_i}$$

gdzie x_i to wartości mocy w poszczególnych pasmach częstotliwości, a N to liczba punktów w widmie. Średnia geometryczna mierzy rozkład mocy w sposób bardziej odpowiedni do analizowania szeroko rozłożonych wartości (szum), podczas gdy średnia arytmetyczna jest wrażliwa na wysokie wartości (tonalne).

Zastosowanie

Miara SFM jest szeroko stosowana w analizie sygnałów audio w następujących dziedzinach:

- *Rozpoznawanie mowy i muzyki* – SFM pomaga w identyfikacji, czy sygnał pochodzi od źródła tonalnego (np. mowa lub muzyka) czy od źródła szumowego (np. hałas tła).
- *Separacja dźwięku* – w systemach separacji dźwięków SFM może być używana do oddzielania dźwięków tonalnych od szumów.
- *Analiza audio* – w obróbce dźwięku SFM może być używana do analizy charakterystyki dźwięku, wykrywania obecności szumów lub tonów.

Zalety

- *Prosta implementacja* – metoda SFM jest stosunkowo łatwa do zaimplementowania, wymagając jedynie obliczenia FFT i kilku operacji arytmetycznych.
- *Szerokie zastosowanie* – może być używana w różnych dziedzinach, takich jak rozpoznawanie mowy, klasyfikacja dźwięków czy analiza jakości nagrań.

Wady

- *Wrażliwość na szum* – SFM może być mniej skuteczna w przypadku dźwięków o bardzo niskiej lub zmiennej intensywności, gdzie rozkład energii w widmie jest mniej wyraźny.
- *Zależność od parametrów sygnału* – skuteczność miary zależy od parametrów sygnału wejściowego, takich jak wielkość ramki czy częstotliwość próbkowania.

Miara SFM jest cenionym narzędziem w analizie sygnałów dźwiękowych, szczególnie przy rozróżnianiu pomiędzy sygnałami tonalnymi a szumowymi, co czyni ją użyteczną w wielu aplikacjach związanych z przetwarzaniem dźwięku i analityką audio.

3.2.7 Spectral Crest Factor (SCF)

Spectral Crest Factor (SCF) to miara stosunku maksymalnej wartości widma mocy (PSD - Power Spectral Density) do średniej wartości tego widma. Wysoki SCF wskazuje na obecność

wyraźnych, ostrych pików w widmie (np. w sygnałach tonalnych lub impulsowych), podczas gdy niski SCF sugeruje bardziej rozproszone widmo, charakterystyczne dla szumów.

Definicja i obliczenia

Miara SCF jest obliczana na podstawie następujących kroków:

- **Ramkowanie sygnału** – sygnał dzielony jest na małe ramki o stałej długości, które następnie są analizowane w celu uzyskania reprezentacji czasowej i częstotliwościowej sygnału.
- **Okienkowanie sygnału** – na każdą ramkę nakładane jest okno (np. okno Hanninga), aby zminimalizować efekty przecieku widmowego i poprawić jakość obliczeń FFT.
- **Obliczanie widma mocy** – dla każdej ramki obliczane jest widmo mocy za pomocą transformacji Fouriera (FFT), przy czym analiza ograniczona jest do dodatnich częstotliwości. Następnie obliczana jest moc widma, która jest równa kwadratowi amplitudy każdego punktu widma.
- **Obliczanie SCF** – miara SCF jest obliczana jako stosunek maksymalnej wartości mocy widma do średniej wartości mocy widma w danej ramce:

$$\text{SCF} = \frac{\max(\text{Power Spectrum})}{\text{mean}(\text{Power Spectrum})}$$

Wysoki SCF wskazuje na dominację wąskich, wyraźnych pików w widmie, co jest charakterystyczne dla sygnałów tonalnych lub impulsowych, natomiast niski SCF wskazuje na bardziej równomierne, "szumowe" widmo.

Zastosowanie

Miara SCF znajduje szerokie zastosowanie w analizie sygnałów audio:

- *Rozpoznawanie tonów i impulsów* – SCF jest przydatny do rozróżniania między sygnałami tonalnymi, które charakteryzują się ostrymi pikami w widmie, a szumami, które mają bardziej rozproszone widmo.
- *Analiza mowy* – w systemach rozpoznawania mowy SCF może pomóc w wykrywaniu fragmentów zawierających wyraźne tonalne elementy, takie jak słowa i sylaby.
- *Przetwarzanie dźwięku* – SCF jest używany do analizy charakterystyki sygnałów audio, szczególnie przy próbach oddzielania dźwięków o różnych cechach (np. tonalnych i szumowych).

Zalety

- *Łatwość interpretacji* – SCF jest prostą miarą, która dostarcza jednoznacznych wyników, rozróżniając sygnały tonalne od szumów.
- *Skuteczność w rozpoznawaniu impulsów i tonów* – SCF jest szczególnie skuteczny w identyfikacji sygnałów o wyraźnych szczytach w widmie.

Wady

- *Wrażliwość na zmienne pasma* – w przypadku sygnałów złożonych, w których różne pasma częstotliwości mają różne cechy, SCF może dawać trudności w dokładnej interpretacji.
- *Ograniczenia w analizie szumów* – w przypadku szumów o bardzo rozproszonym widmie, SCF może być mniej przydatny, ponieważ nie uwzględnia lokalnych charakterystyk szumów.

Miara SCF jest użytecznym narzędziem w analizie dźwięków o wyraźnych tonach, szczególnie w kontekście rozpoznawania mowy, muzyki czy analizy impulsów, pozwalając na łatwe rozróżnienie sygnałów tonalnych od szumowych.

4 Wnioski

Na podstawie wyników i implementacji można wyciągnąć następujące wnioski:

4.1 Parametry w dziedzinie częstotliwości dla różnych głosek

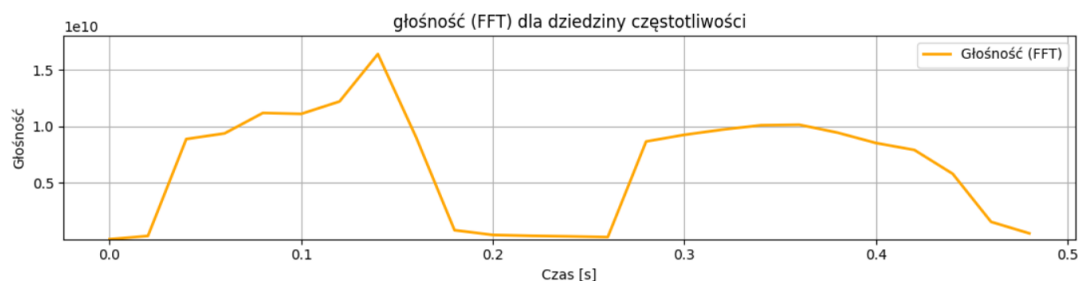
Na podstawie analizy parametrów sygnału w dziedzinie częstotliwości, zmiany w charakterystyce nagrania słowa "abe", które widać na Rysunek 3, można opisać następująco:



Rysunek 3: Wykres nagrania słowa "abe" w dziedzinie czasu z oknem Hamminga

- **Głośność (Volume):**

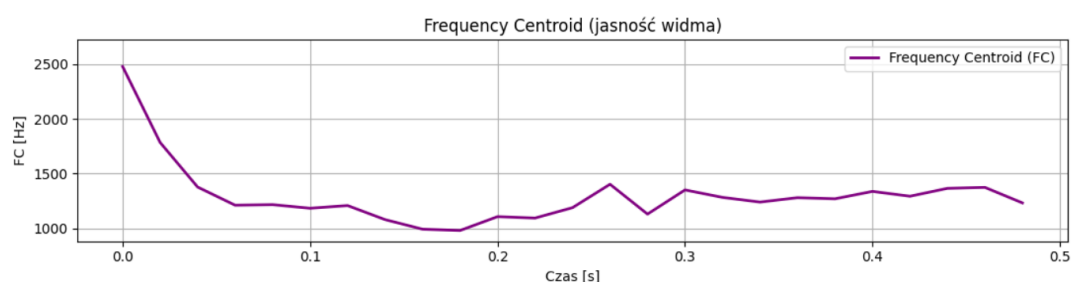
- Głośność sygnału zmieniała się w zależności od siły wymowy samogłoskowych dźwięków. Samogłoska "a" miała wyższą głośność, ponieważ jest bardziej otwartą samogłoską, podczas gdy "e" miało nieco niższą głośność. Zmiana ta jest widoczna na wykresie Rysunek 4.



Rysunek 4: Wykres głośności dla nagrania

- **Frequency Centroid (FC):**

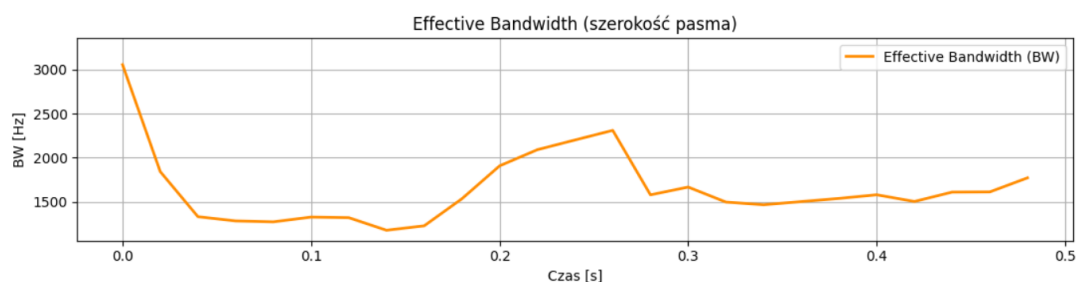
- Środek ciężkości widma (Frequency Centroid) na początku nagrania miał wysoką wartość, ponieważ dźwięk "a" ma tendencję do generowania bardziej niskoczęstotliwościowych formantów. W miarę jak nagranie przechodzi do samogłoski "e", FC przesuwa się w kierunku niższych częstotliwości, ponieważ "e" ma wyższe formanty, które wpływają na obniżenie wartości FC. Zmiany te są przedstawione na wykresie Rysunek 5.



Rysunek 5: Wykres FC dla nagrania

- **Bandwidth (BW):**

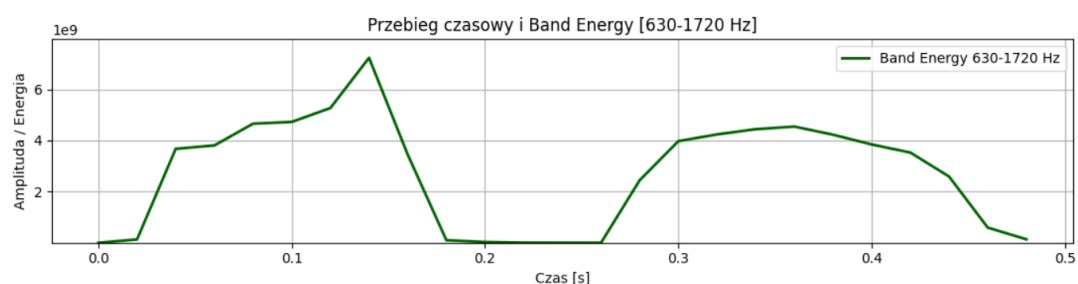
- Szerokość pasma była największa dla samogłoski "a", która ma szersze pasmo formantów. Po tej części następuje też zwiększenie szerokości pasma w czasie wymowy "b", ponieważ spółgłoska "b" wprowadza dodatkową szerokość pasma z powodu charakterystycznych dla niej cech akustycznych. Zmiany te są przedstawione na wykresie Rysunek 6.



Rysunek 6: Wykres Bandwidth dla nagrania

- **Band Energy (BE):**

- Energia pasma była najwyższa dla samogłoski "a", która ma więcej energii skoncentrowanej w niższych częstotliwościach. Dla "e" energia pasma była nieco niższa, ponieważ "e" koncentruje swoją energię w wyższych częstotliwościach. Dla spółgłoski "b" energia pasma była prawie zerowa, ponieważ dźwięk "b" ma bardzo mało energii w paśmie, co jest charakterystyczne dla spółgłoskowych dźwięków w tej pozycji. Zmiany te ilustruje wykres Rysunek 7.



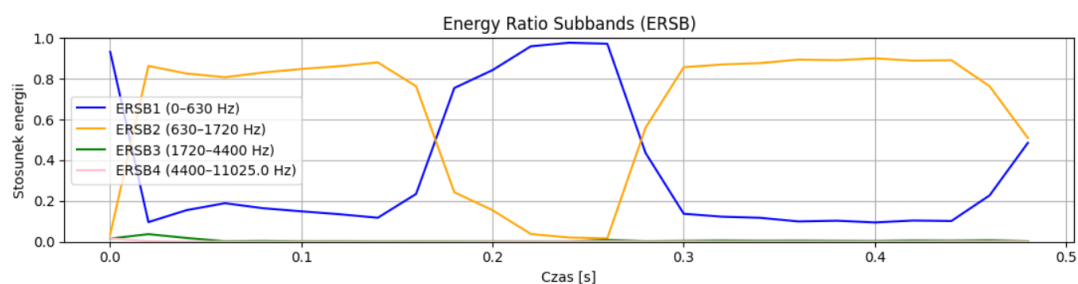
Rysunek 7: Wykres Band Energy dla nagrania

- **Effective Rectangular Bandwidth (ERSB):**

- Na Rysunek 8 widać, że wartości **ERSB1** i **ERSB2** są symetryczne względem $x = 0.5$, podczas gdy **ERSB3** i **ERSB4** przyjmują bardzo małe wartości przez całe nagranie, co wskazuje na minimalną efektywną szerokość pasma w wyższych zakresach częstotliwości.

- **Spectral Flatness Measure (SFM):**

- Miara płaskości widma była stosunkowo niska dla obu samogłoskowych dźwięków, ale "a" miało nieco niższą wartość SFM ze względu na wyraźniejsze formanty w niskich częstotliwościach. Zmiany SFM są widoczne na wykresie Rysunek 9.



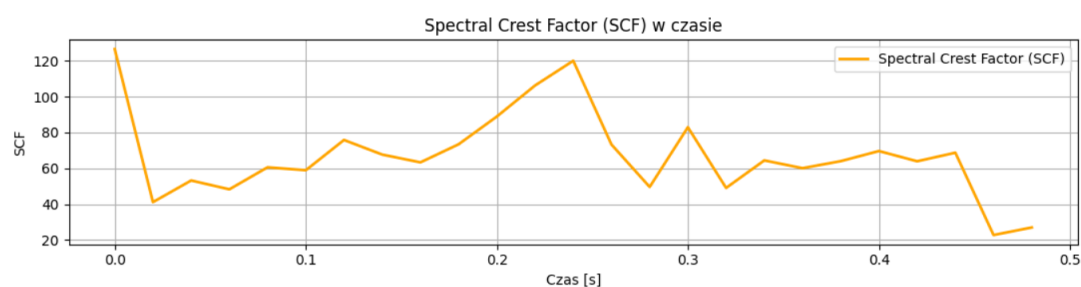
Rysunek 8: Wykres ERSB dla nagrania



Rysunek 9: Wykres SFM dla nagrania

- **Spectral Crest Factor (SCF):**

- Współczynnik szczytowości widma (SCF) był wyższy na początku nagrania oraz podczas części "b", ponieważ te fragmenty nagrania miały wyraźniejsze i bardziej skoncentrowane szczyty w widmie. Dla samogłoski "e" SCF było nieco mniejsze, co pokazuje wykres Rysunek 10.



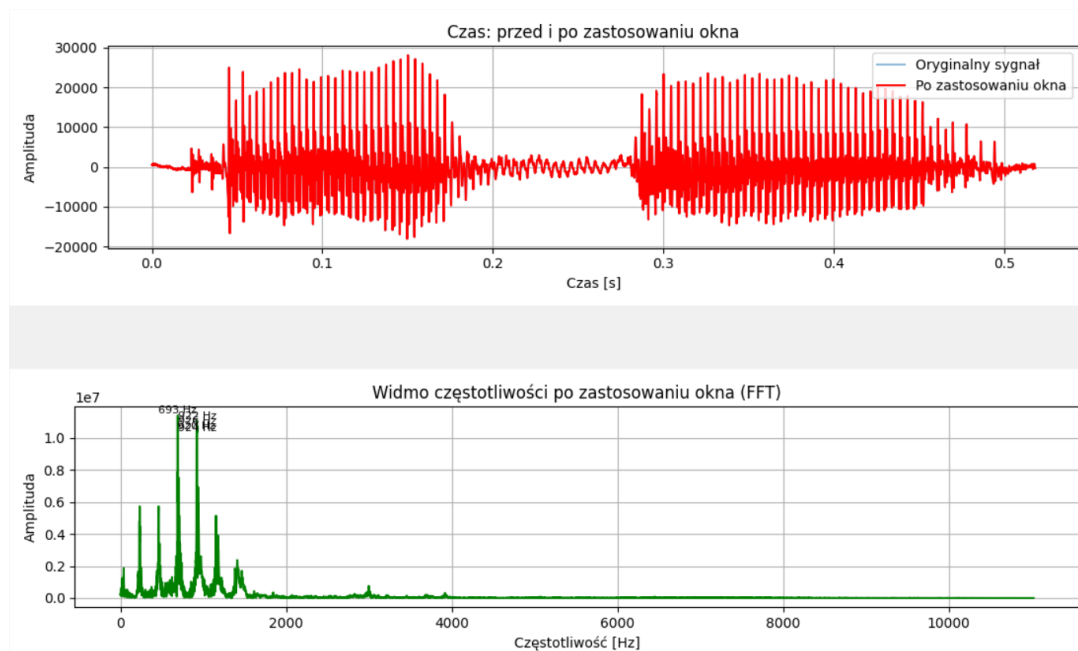
Rysunek 10: Wykres SCF dla nagrania

4.2 Wnioski na temat wpływu funkcji okienkowych na widmo sygnału

W analizie sygnałów przy użyciu transformacji Fouriera, zastosowanie różnych funkcji okienkowych ma istotny wpływ na wyniki w dziedzinie częstotliwości. Każde okno wprowadza określone zmiany w rozmyciu widma, rozdzielczości częstotliwościowej oraz poziomie bocznych harmonicznych. Poniżej przedstawiono wnioski na temat działania poszczególnych okien, ilustrowanych odpowiednimi wykresami.

- Okno prostokątne:

- Okno prostokątne, reprezentowane przez funkcję `np.ones(size)`, skutkuje najwyższym poziomem *leakage*, co powoduje szerokie piki w widmie. Rozmycie w dziedzinie częstotliwości jest znaczne, co ilustruje wykres Rysunek 11. Okno to prowadzi do rozproszenia energii na szerokie pasma częstotliwości.



Rysunek 11: Widmo po zastosowaniu okna prostokątnego

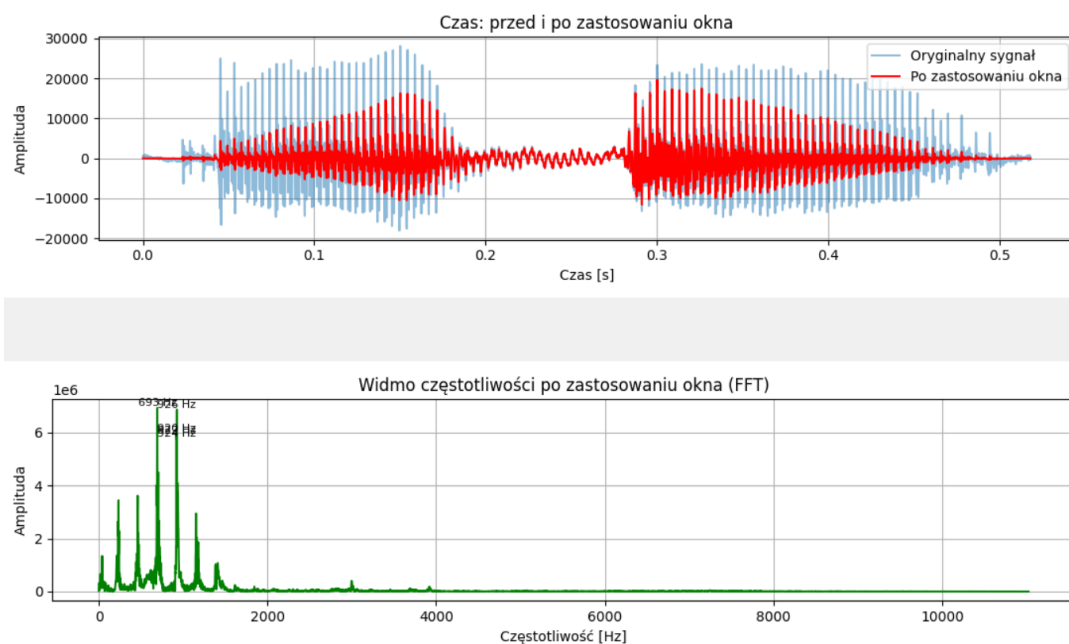
- Okno trójkątne:

- Okno trójkątne, jak pokazuje wykres Rysunek 12, zmniejsza efekty *leakage* w porównaniu do okna prostokątnego, jednak nadal występuje pewien poziom rozmycia w widmie. Zmniejsza to rozproszenie energii w bocznych pasmach, ale nadal nie jest idealnym rozwiązaniem.

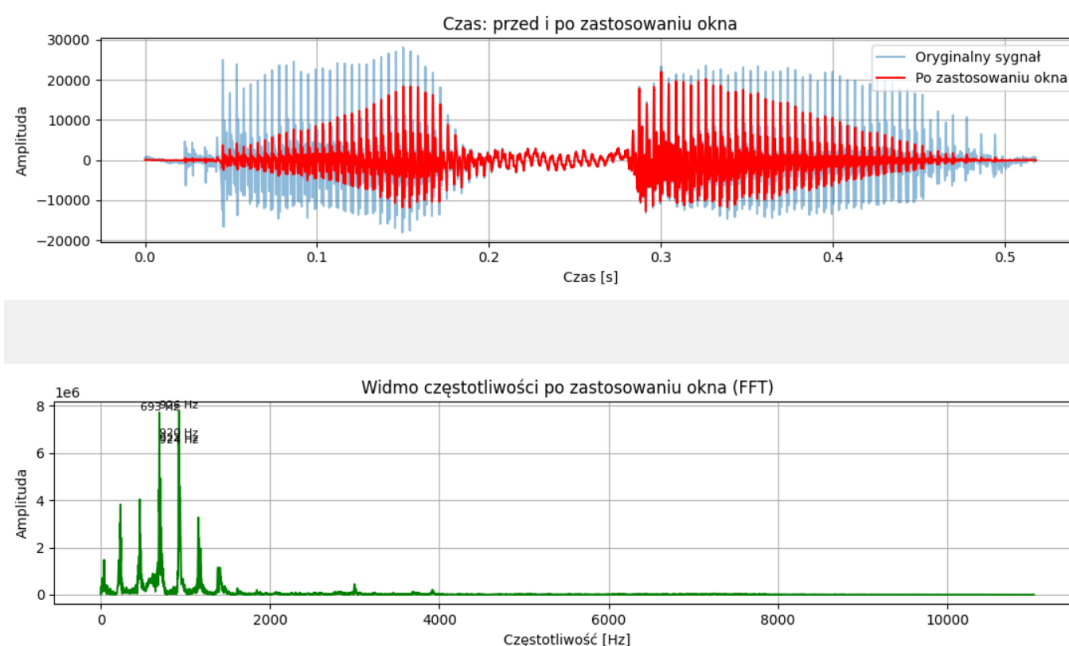
- Okno Hamming'a:

- Okno Hamming'a, przedstawione na wykresie Rysunek 13, wprowadza wyraźne wygładzenie i skutecznie redukuje *leakage*, przy zachowaniu stosunkowo dobrej rozdzielczości częstotliwościowej. Widmo sygnału jest mniej rozmyte, a boczne harmoniczne są wytłumione, co czyni je popularnym wyborem w analizach sygnałów.

- Okno Van Hana:



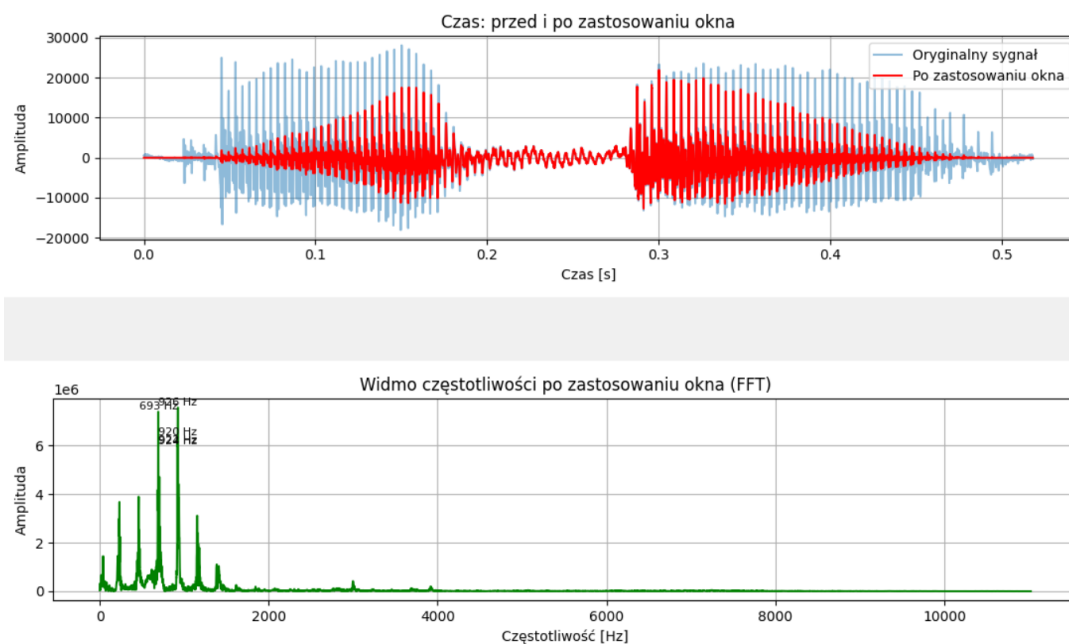
Rysunek 12: Widmo po zastosowaniu okna trójkątnego



Rysunek 13: Widmo po zastosowaniu okna Hamming'a

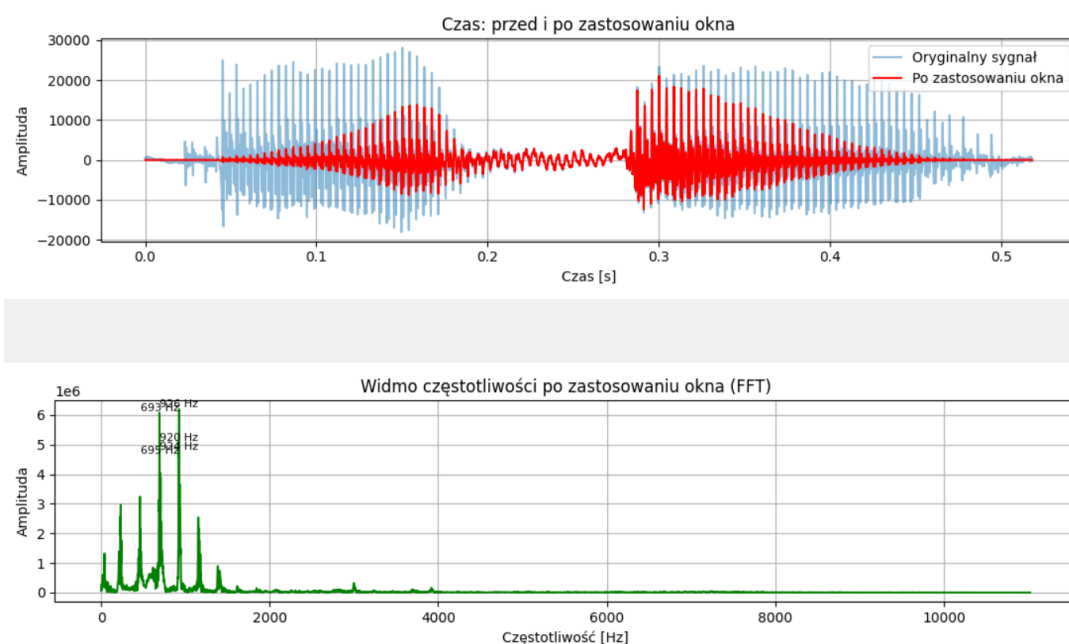
- Okno Hanning'a (Van Hann), jak pokazuje wykres Rysunek 14, również redukuje *leakage*, ale w porównaniu do okna Hamming'a ma nieco szersze pasma w widmie. Jego zastosowanie zapewnia lepsze wygładzenie i bardziej płynne przejście, ale z większym rozmyciem w porównaniu do okna Hamming'a.

- **Okno Blackmana:**



Rysunek 14: Widmo po zastosowaniu okna Van Hanna

- Okno Blackmana, zobrazowane na wykresie Rysunek 15, zapewnia najmniejsze *leakage* i skutecznie tłumi boczne harmoniczne. Widmo sygnału jest silnie wygładzone, jednak kosztem większego rozmycia w porównaniu do innych okien. Jest to najsukuteczniejsze okno w kontekście minimalizacji efektów leakage, ale z nieco większym rozmyciem w porównaniu do innych okien.



Rysunek 15: Widmo po zastosowaniu okna Blackmana

Podsumowując, wybór funkcji okienkowej zależy od specyficznych potrzeb analizy sygnału, gdzie kompromis między minimalizowaniem *leakage*, zachowaniem wyraźnych pików i rozdzielczością częstotliwościową odgrywa kluczową rolę. Wykresy przedstawione w tej sekcji ilustrują, jak różne okna wpływają na rozmycie, rozdzielczość i tłumienie harmoniczných w analizowanych sygnałach.

5 Podsumowanie

Raport przedstawia aplikację do analizy sygnałów dźwiękowych, której celem jest ekstrakcja parametrów z dziedziny częstotliwości, takich jak głośność (Volume), Frequency Centroid (FC), szerokość pasma (BW), energia pasma (BE), stosunek energii pasma (ERSB), miara płaskości widma (SFM) oraz współczynnik szczytowości widma (SCF). Aplikacja umożliwia analizę sygnałów w dziedzinie częstotliwości za pomocą transformacji Fouriera (FFT) oraz wizualizację wyników na wykresach.

Wyniki analizy wykazały wyraźne różnice w parametrach dla różnych głosek. Zastosowanie różnych funkcji okienkowych, takich jak okno prostokątne, Hamming'a i Hanning'a, miało istotny wpływ na jakość widma, poprawiając rozdzielczość częstotliwościową i minimalizując *leakage*.

Wnioski z analizy mogą znaleźć zastosowanie w dziedzinach takich jak rozpoznawanie mowy i przetwarzanie sygnałów audio, gdzie dobór odpowiednich parametrów i funkcji okienkowej jest kluczowy dla jakości wyników.