



przedmiot
Analiza i Przetwarzanie Dźwięku, Projekt 2.



Aplikacja do analizy częstotliwościowej nagrań audio

Antoni Kingston

Numer albumu

327284

prowadzący

dr. inż. Janusz Rafałko

WARSZAWA kwiecień 2025

Spis treści

1. Opis ogólny	3
1.1. Użyte technologie	3
1.2. Interfejs	3
2. Wyznaczane parametry	5
2.1. Volume	5
2.2. Frequency centroid	5
2.3. Effective bandwidth	5
2.4. Band energy ratio	5
2.5. Spectral flatness measure	6
2.6. Spectral crest factor	6
2.7. FFT (czytać tylko w przypadku nadmiaru czasu a najlepiej po jego końcu)	6
3. Przykłady działania	7
3.1. abeuwo	7
3.2. alesza	9
3.3. Czyste i nieczyste A4(440Hz)	10
3.4. ludzki głos	12
4. Uwagi końcowe	13

1. Opis ogólny

1.1. Użyte technologie

Aplikacja została napisana w języku python, jej interfejs graficzny jest oparty na bibliotece PyQt5, wykresy na bibliotekach matplotlib oraz seaborn. Większość obliczeń korzysta z biblioteki numpy. Szczegółowe wymagania są zawarte w pliku 'requirements.txt'.

1.2. Interfejs

Aplikacja posiada toolbar z możliwością wczytania pliku audio (W formacie .wav). Pod paskiem znajduje się 5 zakładek z różnymi (czasem interaktywnymi) wizualizacjami:

1. –

- wykres przebiegu czasowego wczytanego pliku, z możliwością wyboru fragmentu poprzez przeciągnięcie myszą
- wykres przebiegu czasowego wybranego fragmentu

2. –

- wykres głośności
- wykres frequency centroid
- wykres effective bandwidth
- wykres band energy ratio
- wykres spectral flatness measure
- wykres spectral crest factor trzy ostatnie wykresy wizualizują dane z 3 pasm zgodnych z filtrami cochleara, ponadto można wybrać długość i stopień nakładania się ramek na podstawie, których wykresy są tworzone

3. –

- wykres przebiegu czasowego wczytanego pliku, z możliwością wyboru fragmentu poprzez przeciągnięcie myszą
- wykres przebiegu czasowego wybranego fragmentu
- wykres w dziedzinie częstotliwości wybranego fragmentu wizualizacja fragmentu odbywa się po zastosowaniu funkcji okienkowej, dostępne są funkcje:
 - prostokątna (identyczność)
 - trójkątna
 - hamminga
 - hanninga
 - blackmana

1. Opis ogólny

4. spektrogram z możliwością wyboru długości ramki, nakładania się ramek, zakresu częstotliwości oraz funkcji okienkowej
5. wykres częstotliwości krtaniowej wyznaczanej przy pomocy cepstrum

2. Wyznaczane parametry

2.1. Volume

Nic odkrywczego, uśredniamy kwadraty amplitud po analizowanych częstotliwościach

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

2.2. Frequency centroid

Średnia częstotliwości ważona po amplitudzie

$$FC(n) = \frac{\sum_{k=0}^{N-1} f(k) S_n(k)}{\sum_{k=0}^{N-1} S_n(k)}$$

2.3. Effective bandwidth

Wyliczana przy użyciu frequency centroid, można określić jako miara rozstrzału występujących częstotliwości

$$BW(n) = \sqrt{\frac{\sum_{k=0}^{N-1} (f(k) - FC(n))^2 S_n(k)^2}{\sum_{k=0}^{N-1} S_n(k)^2}}$$

2.4. Band energy ratio

Częstotliwości w zakresie słyszalnym można podzielić ze względu na różnice w ludzkiej percepcji na 4 pasma:

1. 0 – 630 Hz
2. 630 – 1720 Hz
3. 1720 – 4400 Hz
4. 4400 – ... Hz

dla każdego z past można wyznaczyć jego udział w całkowitej energii (głośności) sygnału:

$$BER_{[f_0, f_1]}(t) = \frac{BE_{f_0, f_1}(t)}{Vol(t)}$$

gdzie:

$$BE_{[f_0, f_1]}(t) = \frac{\sum_{i=i_{f_0}}^{i_{f_1}} S_t(i)^2}{\sum_{k=0}^{N-1} w(k)}$$

(górną całką jest po częstotliwościach, dolną po czasie, gdzie w to stosowana funkcja okna)

2.5. Spectral flatness measure

Jest to iloraz średniej geometrycznej i arytmetycznej kwadratu spektrum w zadanym zakresie, ta pierwsza jest mniej podatna na wartości odstające co skutkuje pewną adekwatnością nazwy

$$SFM_{[f_0, f_1]}(n) = \frac{\sqrt[i_{f_1} - i_{f_0} + 1]{\prod_{k=i_{f_0}}^{i_{f_1}} S_n(k)^2}}{\frac{1}{i_{f_1} - i_{f_0} + 1} \sum_{k=i_{f_0}}^{i_{f_1}} S_n(k)^2}$$

2.6. Spectral crest factor

Podobna do poprzedniej wartość, w liczniku zamiast średniej geometrycznej jest stosowane maksimum

$$SCF_{[f_0, f_1]}(n) = \frac{\max_i (S_n(i)^2)}{\frac{1}{i_{f_1} - i_{f_0} + 1} \sum_{k=i_{f_0}}^{i_{f_1}} S_n(k)^2}$$

2.7. FFT (czytać tylko w przypadku nadmiaru czasu a najlepiej po jego końcu)

Szybka transformacja Fouriera to oczywiście nie parametr lecz algorytm, niezwykle jednak przy wyznaczaniu powyższych przydatna. Wielkości $S_n(k)$ to widma częstotliwościowe uzyskiwane z danych w dziedzinie czasu poprzez zastosowanie dyskretnej transformacji Fouriera (DTFT). Domyślny algorytm ich wyznaczania jest złożoności $O(n^2)$ co nie jest wynikiem umożliwiającym sensowne obliczanie wielu parametrów na nawet kilkusekundowych danych. Ludzkość dysponuje jednak FFT - algorytmem o złożoności $O(n \log n)$. Poniższy strumień świadomości autora prędkiej niż wyjaśnić osobie nierozumiejącej działanie algorytmu spowoduje, że ta, która się do niego przyzwyczaiła się odzwyczai, trudno poza tym suchym raportowym tekstem objaśniać matematyczne złożoności a wstawianie tu wykładu mija się z celem, lekkomyślne zobowiązanie się w raporcie za projekt 1. musi mieć jednak swoje konsekwencje, poniżej one.

Całość tłumaczenia będzie opierała się na pewnym izomorfizmie, którego istnienie nie zostanie wykazane da się jednak w nie uwierzyć. Mianowicie mamy odpowiedniość między dyskretnym sygnałem w dziedzinie czasu wraz jego transformatą fouriera, oraz wielomianem w postaci listy współczynników wraz z nim samym postaci listy różnych punktów jego wykresu (listy wartości dla pewnych sprytnie dobranych punktów). Innymi słowy jeśli znaleźć by dla wielomianów metodę przechodzenia między postacią współczynnikową a punktową to jest ona również metodą przechodzenia z dziedziny czasu w dziedzinę częstotliwości, i vice versa - odwrotnej transformacji można dokonać dochodząc do postaci współczynnikowej z punktowej. Jak zatem obliczyć n wartości wielomianu (n-1)-tego stopnia (z n wyrazami) w czasie $O(n \log n)$ znając jego współczynniki? Złożoność sugeruje pewną rekurencję i rzeczywiście trzeba znaleźć sposób rozbicia dużego problemu na

mniejsze, celem tego rozważmy ogólny wielomian n -tego stopnia (tutaj założymy, że n jest liczbą parzystą, nawet potęgą 2):

$$w(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

można go zapisać w nieco innej postaci

$$\begin{aligned} w(x) &= (a_0 + a_2x^2 + \dots + a_nx^n) + (a_1x + a_3x^3 + \dots + a_{n-1}x^{n-1}) = (a_0 + \dots + a_nx^n) + x(a_1 + a_3x^2 + \dots + a_{n-1}x^{n-2}) = \\ &= e(x^2) + xo(x^2) \end{aligned}$$

gdzie wielomiany $e(x^2)$ oraz $o(x^2)$ są takimi aby się zgadzało, istotne jest tu to, że są wielomianami stopnia $n/2$ (czyli wielomianami) wystarczy teraz, że obliczymy wartości wielomianów $e(x^2)$ oraz $o(x^2)$ w $n/2$ punktach, koszt takiej operacji będzie 2 razy mniejszy niż pierwotnej (2 wielomiany, ale 2 razy mniejsze i 2 razy mniej punktów), dlaczego jednak wystarczy? Otóż znając wartości $e(x^2)$ oraz $o(x^2)$ możemy wyznaczyć 2 wartości pierwotnego wielomianu $w(x)$:

$$w(x) = e(x^2) + xo(x^2)$$

$$w(-x) = e(x^2) - xo(x^2)$$

Wszystko „ładnie i cacy”, rekurencja aby jednak rzeczywiście zadziałała musi wystartować z pewnych specyficznych punktów, mianowicie pierwiastków n -tego stopnia z 1, celem zachowania odpowiedniego stężenia raportu w raporcie zostawię to jako urwany wątek.

3. Przykłady działania

3.1. abeuwo

Te dwie monosylaby zdają się być dobrymi przykładami do porównywania samogłosek (mamy aż 4) oraz spółgłosek (b i w niewiele się od siebie w wymowie różnią w języku hiszpańskim np. dystynkcja między nimi jest dość niepewna)

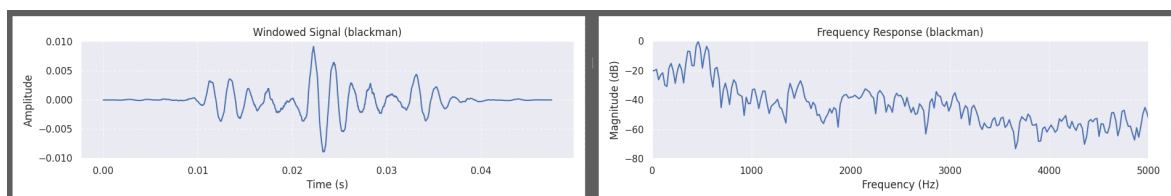
3. Przykłady działania



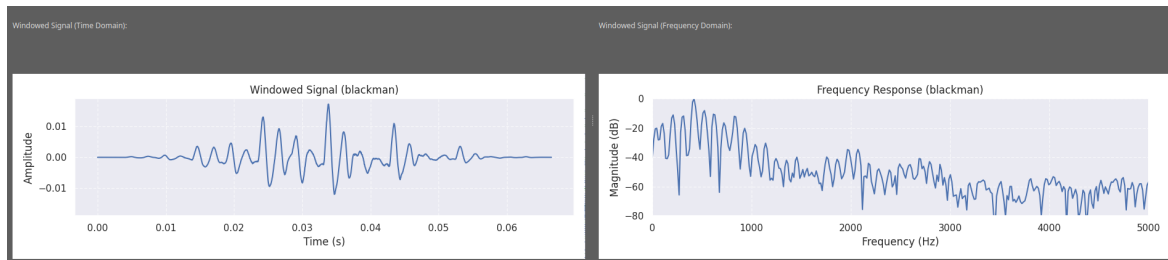
Rysunek 3.1. Próbką abe



Rysunek 3.2. Próbką uwo



Rysunek 3.3. wycięty środkowy fragment w domniemaniu z głoską b z abe



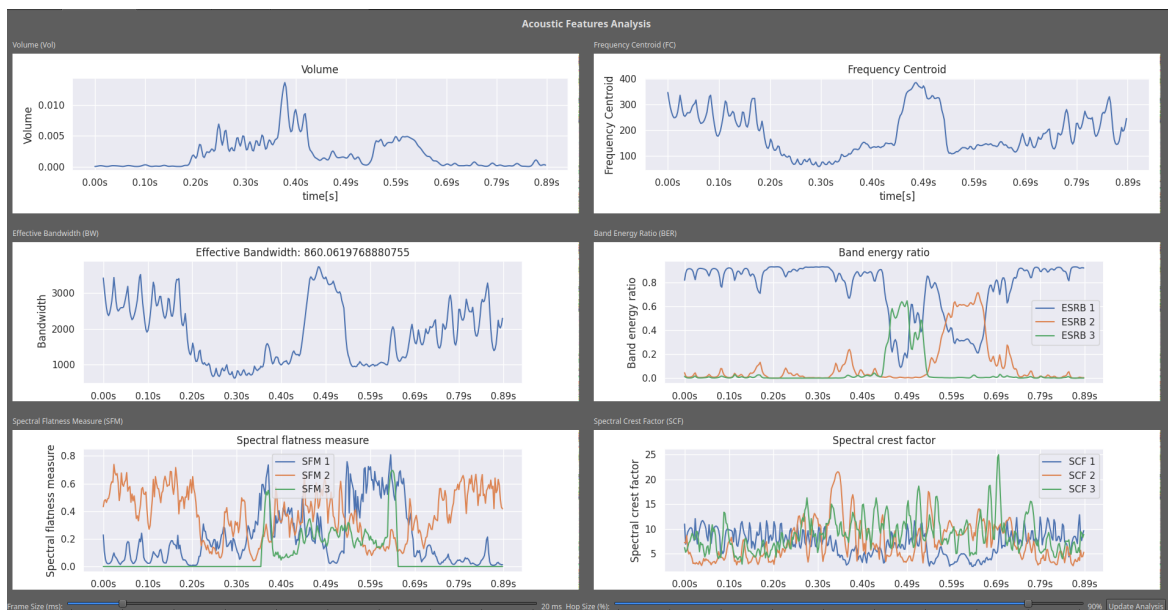
Rysunek 3.4. wycięty środkowy fragment w domniemaniu z głóską w z uwo

Obserwacje:

1. Spółgłoski są zdecydowanie cichsze niż samogłoski
2. Obydwa spółgłoski są dźwięczne (miara Frequency Centroid nie rośnie jakoś wybitnie między samogłoskami)
3. Samogłoski a i o mają równie rozdzieloną energię między pasma 1 i 2 niż samogłoski e i u (może dlatego te pierwsze uznawane są za *jaśniejsze*)?
4. spółgłoski b i w są rzeczywiście dość podobne

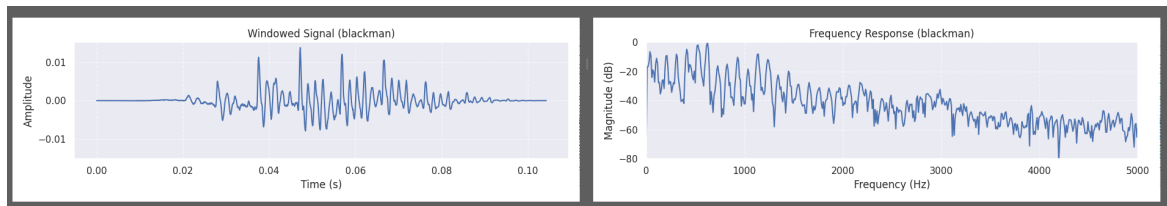
3.2. alesza

Niski poziom FC w obu poprzednich próbkach skłania do zbadania przykładu gdzie może się okazać on wyższy.



Rysunek 3.5. Próbką alesza

3. Przykłady działania



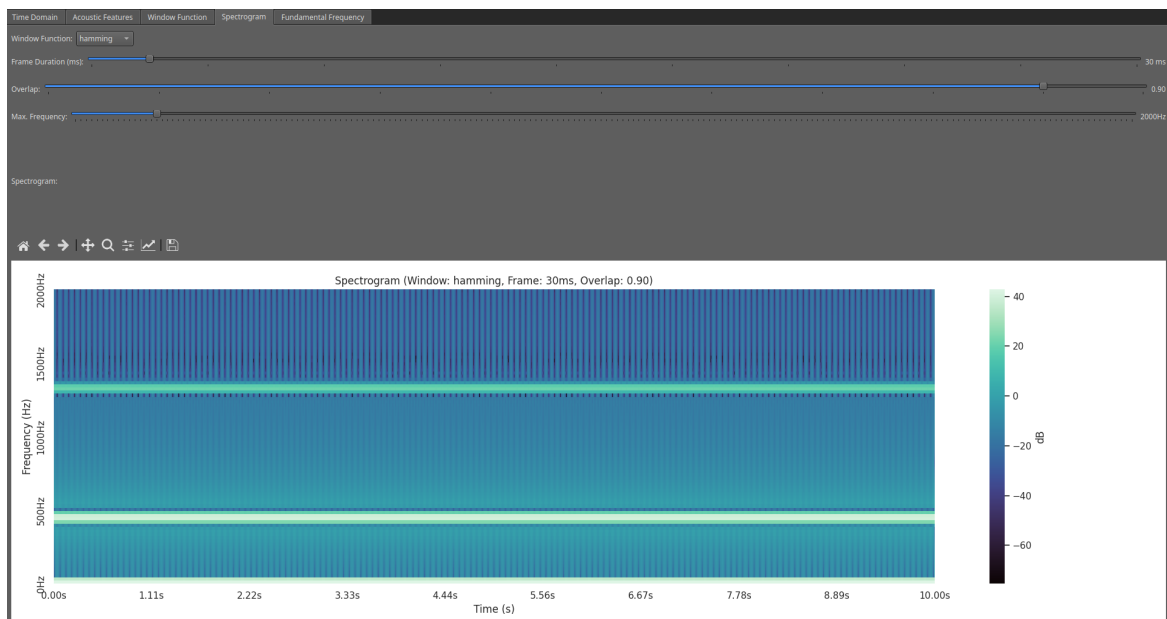
Rysunek 3.6. wycięty środkowy fragment w domniemaniu z głosek sz z alesia

Obserwacje:

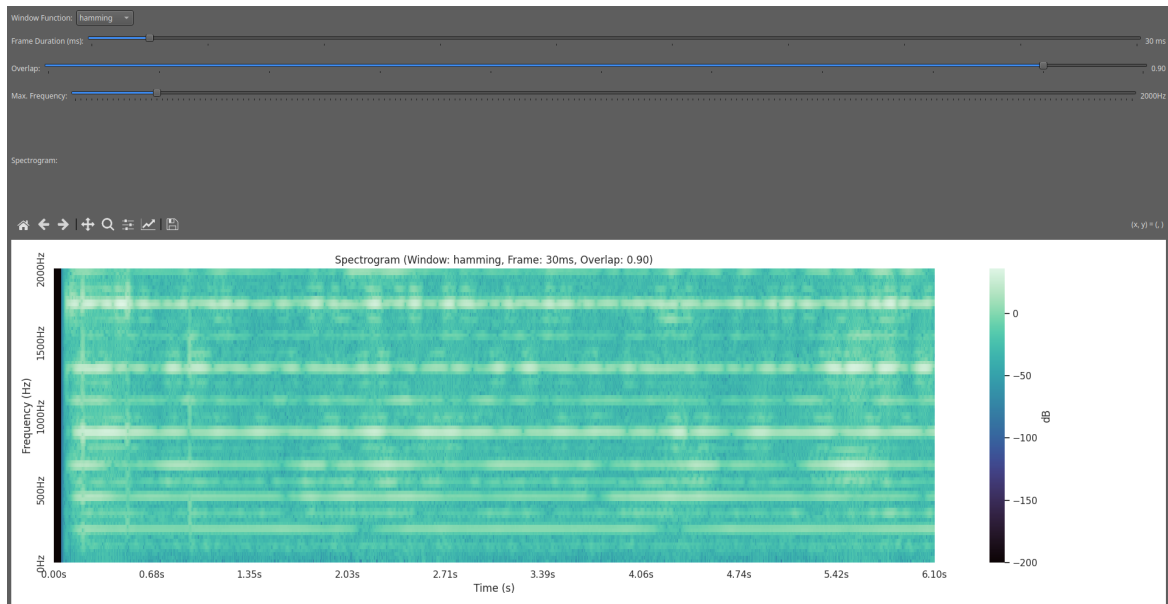
1. głoska bezdźwięczna sz rzeczywiście wybija się jeśli chodzi o FC oraz BW, są to objawy wysokiego stężenia częstotliwości z 3. pasma
2. zwiększone ZCR można zaobserwować na przebiegu czasowym
3. spektrum dalej nic nie mówi

3.3. Czyste i nieczyste A4(440Hz)

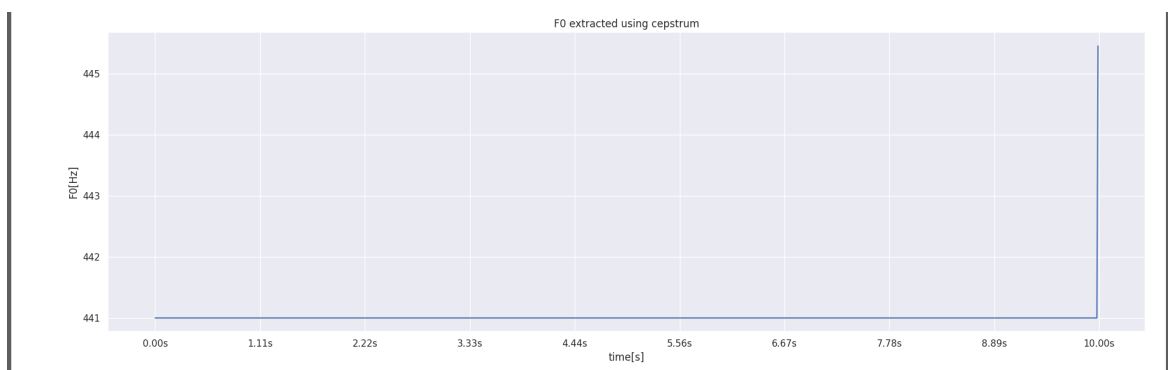
Porównano spektrogramy oraz częstotliwości podstawowe uzyskane z cepstrum dwóch nagrań tonu A4: pierwszego - czysta sinusoida (440.wav) oraz drugiego - nagranie pochodzące z aplikacji pianinowej na telefon (A4.wav).



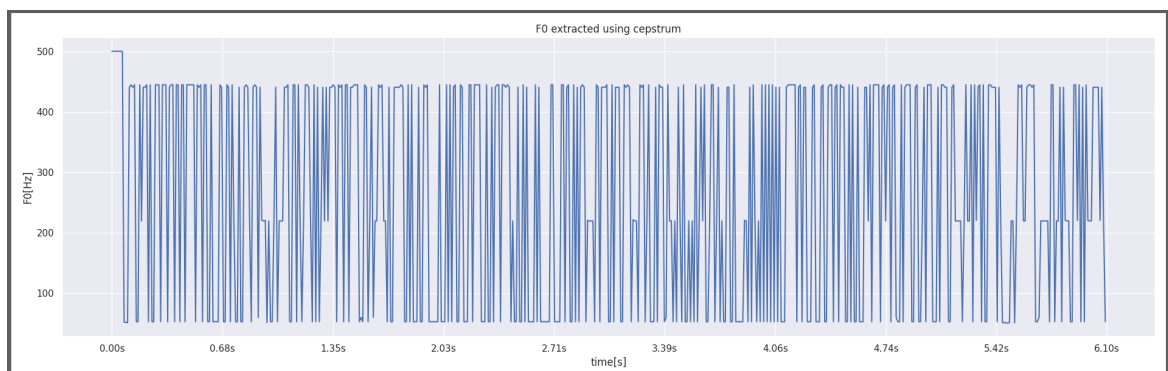
Rysunek 3.7. Próbkę czysta



Rysunek 3.8. Próbką nieczysta



Rysunek 3.9. Próbką czysta



Rysunek 3.10. Próbką nieczysta

Obserwacje:

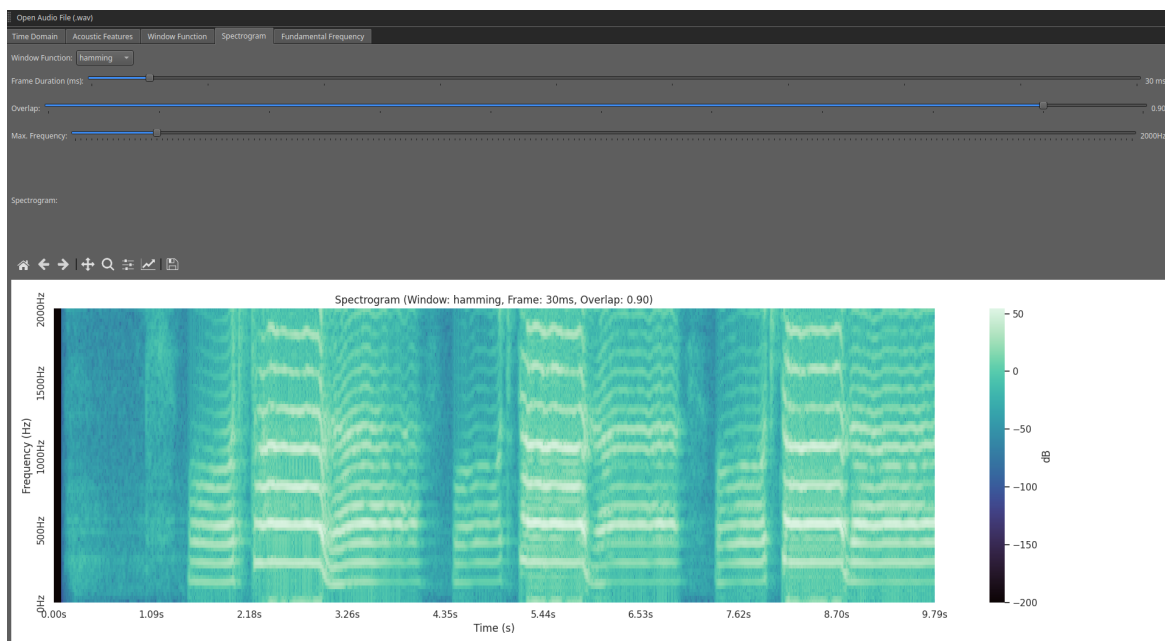
1. Na spektrogramach da się zaobserwować głośny ton w okolicach 450Hz

3. Przykłady działania

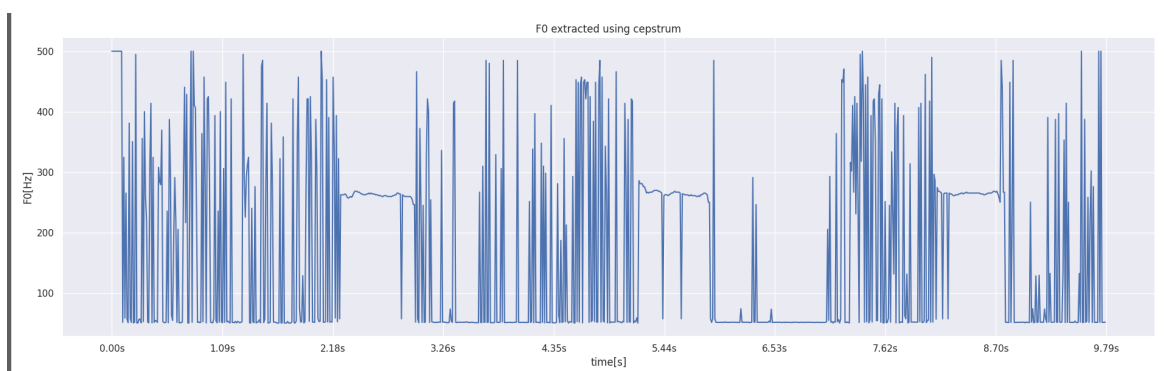
2. Na spektrogramie z próbką zanieczyszczoną poza tym, że widać mniej widać również znacznie więcej wielokrotności harmonicznych
co zdaje się odpowiadać faktowi, że jest to nagranie syntezatora organów.
3. wyznaczanie f_0 przy pomocy cepstrum dla danych czystych radzi sobie dobrze
4. dla danych nieczystych oscyluje między doskonałością a beznadzieją (czy należy przepuścić takiego studenta?)

3.4. ludzki głos

Przeanalizowano częstotliwość krtaniową dla nagrania oktawy (C3-C4-C3) (oktawka.wav).



Rysunek 3.11. spektrogram



Rysunek 3.12. krtaniowa cepstrum

Obserwacje:

1. funkcja wyznaczająca częstotliwość krtaniową zdaje się działać lepiej w bardzo głośnych fragmentach (C4, które było dość mocno wydarte objawia się dość stabilnie na ok 260Hz)
2. na spektrogramie widoczne jest dwukrotne przerzedzenie harmonii przy skoku o oktawę

4. Uwagi końcowe

Stanowczo odradza się zagłębienie do kodu źródłowego, w szczególności do funkcji main. Autor sprawozdania nie ponosi odpowiedzialności za wszelki uszczerbek materialny moralny i zdrowotny niechybnie się z takim zajrzeniem wiążący.