Analiza i Przetwarzanie Dźwięku — Projekt 2

Anna Hoang
305922

1 Opis aplikacji

Niniejszy projekt został napisany w Pythonie w formie notatnika Jupyter. Wykorzystano bogate narzędzia do analizy sygnałów audio, którymi dysponuje pakiet `librosa` oraz biblioteki `scipy` i `numpy`, zaś do prezentacji wykresów użyto `matplotlib.pyplot`.

2 Wymagania

- Rysowanie przebiegu czasowego
- Wyznaczenie i rysowanie wykresów parametrów dźwięku z dziedziny częstotliwości.
  - Volume
  - Frequency/Spectrum Centroid (**SC**)
  - (Effective) Bandwidth (**BW**)
  - Band Energy Ratio (**BER**) Band Energy
  - *Spectral Flatness Measure
  - *Spectral Crest Factor
- Rysowanie wykresu widma częstotliwościowego poprzez zastosowanie FFT dla całego sygnału lub dla konkretnej ramki sygnału o dowolnej długości, niekoniecznie ramki początkowej.
- Zastosowanie na takiej ramce/sygnale funkcji okienkowej i narysowanie wykresu w dziedzinie czasu, oraz w dziedzinie częstotliwości po zastosowaniu okna.
- Wybór funkcji okienkowych:
  - prostokątne
  - Hamminga
  - von Hamma

- Rysowanie spectogramu z możliwością wyboru:
  - okna
  - długości ramki
  - overlap
- Rysowanie wykresu częstotliwości krtaniowej, tym razem obliczanej za pomocą cepstrum

3 Opis metod

3.1 Widmo sygnału

Analizując parametry sygnału w dziedzinie czasu, operowaliśmy bezpośrednio na próbkach. W tym projekcie interesujące nas parametry możemy wyróżnić dopiero po przejściu z dziedziny czasu do dziedziny częstotliwości. W tym celu stosuje się **transformację Fouriera**, by uzyskać jego tzw. **widmo**, czyli reprezentację częstotliwościową. Do jego obliczenia (dokładniej jego dyskretnej reprezentacji, czyli DFT) posługujemy się algorytmem szybkiego przekształcenia Fouriera (FFT).

Z matematycznego punktu widzenia gotowy wzór na DFT dla N próbek sygnału (za N również przyjmuje się liczbę rozpatrywanych częstotliwości) to

$$\hat{x}(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-i2\pi n \frac{k}{N}},$$

gdzie x to amplituda próbki, $k \in [0, N-1]$ to pasma częstotliwościowe

$$F(k) = \frac{k}{NT} = \frac{k \cdot f_s}{N}$$

gdzie $f_s = \frac{1}{T}$ to częstotliwość próbkowania (sampling rate).

W praktyce interesuje nas zakres częstotliwości nieprzekraczający $\frac{f_s}{2}$ zwanego częstotliwością Nyquista. Prawa połowa stanowi bowiem odbicie lustrzane, tego co jest po lewej na wykresie widma.

3.2 Funkcje okienkowe

W teorii mowa jest o tym, że to **okresowy** sygnał można przedstawić jako szereg Fouriera i poddać go przekształceniu z dziedziny czasu do częstotliwości. W sytuacji gdy poddawane FFT fragmenty (ramki) nie stanowią okresu całego sygnału, dochodzi do tzw. przecieku widma, który zniekształca amplitudy ramek. Może dojść do tego, że końcowe fragmenty sygnału będą przekształcone na wysokie częstotliwości, których nie ma wprawdzie w oryginalnym sygnale po

zaaplikowaniu FFT.

Można zminimalizować efekty powyższego zjawiska, korzystając z funkcji okienkowych. Idea polega na zaaplikowaniu takiej funkcji na każdej z ramek przed wyznaczeniem ich transformaty Fouriera. Tym sposobem eliminujemy wartości na krańcach ramek, otrzymując sygnał okresowy. Dobór ramek ma również duże znaczenie i żeby w wyniku okienkowania nie zatracić informacji o oryginalnym przebiegu czasowym na końcach, umożliwia się ich zakładkowanie/nakładanie (np. o stałą liczbę próbek).

W obliczeniach zastosowanie funkcji okienkowej na ramce polega na wymnożeniu każdej z jej próbek przez kolejne wartości zadanego okna, czyli dla długości ramki M w próbkach wyznaczamy: $s_w(n) = s(n) \cdot w(n)$ dla $n = 0...M-1$. Poniżej przedstawione są wzory na znane okna:

- prostokątne: $w(n) = 1$ dla $n = 0...M-1$
- von Hanna: $w(n) = 0.5 - 0.5\cos(\frac{2\pi n}{M-1})$ dla $n = 0...M-1$
- Hamminga: $w(n) = 0.54 - 0.46\cos(\frac{2\pi n}{M-1})$ dla $n = 0...M-1$

3.3 Spektogram

Widmo daje nam informacje o udziale poszczególnych wartości częstotliwości występujących w sygnale, jednak jest to uogólnione i nie mamy wiedzy o tym kiedy się one pojawiają. Do analizy sygnału, którego widmo zmienia się w czasie, posługujemy się krótkookresowym przekształceniem Fouriera (STFT). Po zastosowaniu okien wyznacza się widmo dla każdego odcinka osobno i łączy się otrzymane dane. Wynik jest trójwymiarowy: czas - częstotliwość - amplituda i przedstawia się go w formie dwuwymiarowego **spektrogramu**. Jego wykres ma na osi poziomej czas, na osi pionowej - częstotliwość, natomiast amplituda widma jest pokazana za pomocą określonej mapy barw.

Spektrogram stanowi podstawę do wyznaczania poszczególnych parametrów opisujących dany sygnał w dziedzinie częstotliwości. Amplituda dla ramki m i rozpatrywanej k-tej wartości częstotliwości wynosi

$$S(k,m) = \sum_{n=0}^{M-1} x(n+mH) \cdot w(n) \cdot e^{-i2\pi n \frac{k}{N}},$$

gdzie M - długość ramki w próbkach, H - długość skoku, ile próbek jest pomiędzy sąsiednimi ramkami.

3.4 Parametry w dziedzinie częstotliwości

3.4.1 Volume

Głośność jest definiowana następująco dla każdej z ramek:

$$Vol(m) = \frac{1}{N} \sum_{k=0}^{N-1} S^2(k,m)$$

3.4.2 Frequency centroid

Ten parametr określa, w jakim pasmie częstotliwości skoncentrowana jest największa ilość energii. Można za jego pomocą zmierzyć jasność dźwięku. Ma zastosowania m.in. w klasyfikacji muzyki.

$$FC(m) = \frac{\sum_{k=0}^{N-1} S(k,m) \cdot k}{\sum_{j=0}^{N-1} S(j,m)}$$

3.4.3 Effective Bandwidth

Powiązany z wcześniej opisanym *frequency centroid (FC)* służy do określenia barwy dźwięku. Stanowi jego wariancję (średnią ważoną odległości pasm częstotliwości od FC). Ma zastosowania m.in. w klasyfikacji do gatunków muzycznych.

$$BW(m) = \frac{\sum_{k=0}^{N-1} S(k,m) \cdot |k - FC(m)|}{\sum_{j=0}^{N-1} S(j,m)}$$

3.4.4 Band Energy Ratio

Do porównania energii dla wysokich częstotliwości z niskimi częstotliwościami służy parametr BER. Znajduje zastosowanie m.in w rozróżnianiu mowy od muzyki i jest miarą tego, jak bardzo dominują niskie częstotliwości w zadanym sygnale, dokładniej definiuje się go jako stosunek mocy sygnału dla niższych przez moc dla wyższych częstotliwości

$$BER(m) = \frac{\sum_{k=0}^{K_s-1} S^2(k,m)}{\sum_{k=K_s}^{N-1} S^2(k,m)},$$

gdzie $F(K_s) = f_{split}$ (tzw. *split frequency*) je rozgranicza i zadajemy tę wartość z góry (zwykle $f_{split} = 2000$ Hz). Pasmo częstotliwości w ktorym znajduje się f_{split} opisujemy wzorem:

$$K_s = \lfloor \frac{N \cdot f_{split}}{f_s} \rfloor.$$

3.4.5 Spectral Flatness Measure

Nazywany również współczynnikiem tonalności albo entropią Wienera pozwala określić, jak bardzo przypomina czysty ton (nie stanowi szumu). Wyliczany w skali decybelowej jako iloraz średniej geometrycznej przez arytmetyczną amplitudy poszczególnych pasm częstotliwości:

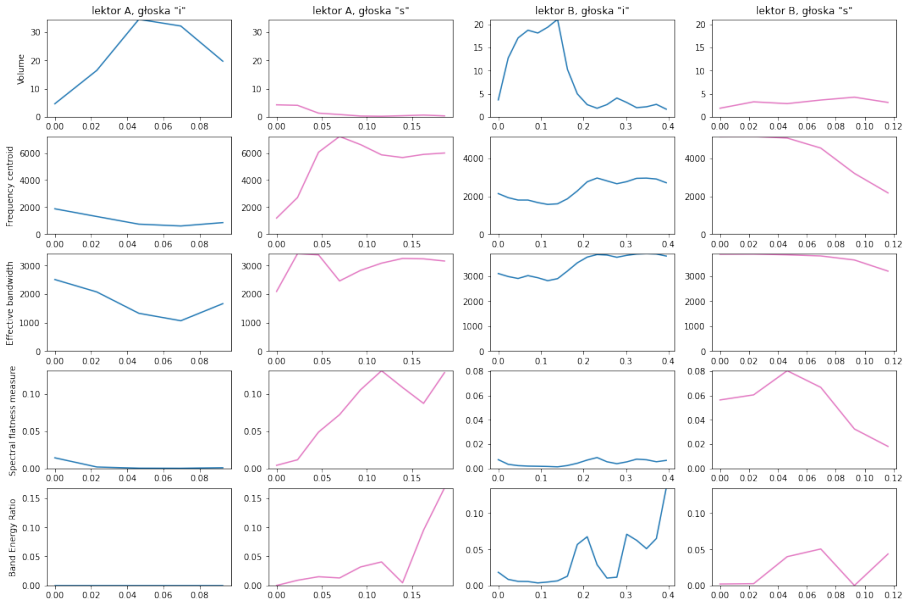
$$SFM(m) = \frac{\exp\left(\frac{1}{N}\sum_{k=0}^{N-1} \ln(S(k,m))\right)}{\frac{1}{N}\sum_{k=0}^{N-1} S(k,m)}.$$

4 Obserwacje i wnioski

Do pokazania wykresów i porównania parametrów posłużono się wyekstrahowanymi głoskami z nagrań różnych słów dwóch lektorów w programie `Audacity`. Mogło to mieć spory wpływ na wyniki, gdyż ręczne etykietowanie zawsze jest obarczone dużą niedokładnością. Niemniej jednak uzyskane wartości wydają się sensowne.

4.1 Wykresy parametrów

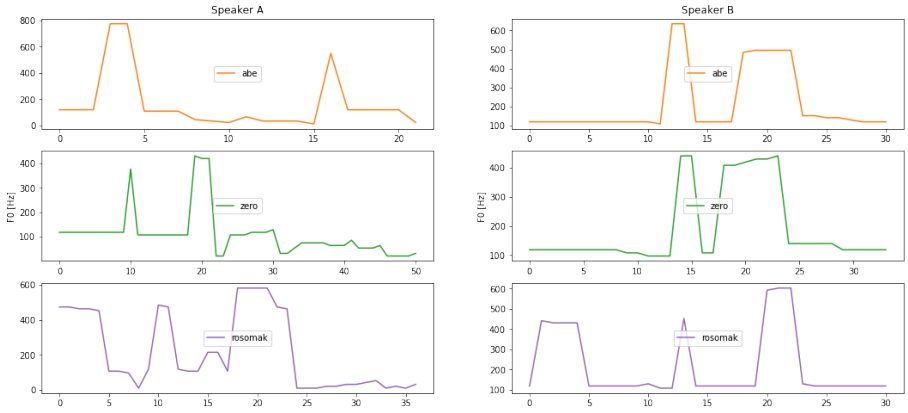
4.1.1 Dla różnych głosek i głosów



Rysunek 1: Wykresy parametrów

4.1.2 Ton podstawowy

W tej części zbadano dwa różne głosy męskie dla trzech różnych wyrazów. Dla pierwszego słowa "aba" gdzie przeważały samogłoski, zakres tonu podstawowego u pierwszego lektora wynosił 0-800 Hz, u drugiego natomiast nie przekraczał 600 Hz. Dla pozostałych wyrazów mówcy mieli podobny zakres F0. Można również zaobserwować kolejność w czasie pojawiania się szczytowych impulsów i przykładowo dla drugiego lektora przy wyrazach "abe" i "zero" wykresy mają przybliżony do siebie kształt.



Rysunek 2: Wykresy tonu podstawowego

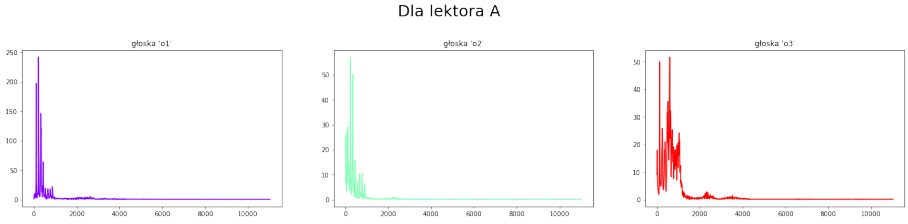
4.2 Formanty

Formant to pasmo częstotliwości w dźwięku, w którym można zaobserwować znaczne wzmocnienie tonów. Ich zbiór określa barwę dźwięku. Możliwość zlokalizowania formantów zadanej głoski może świadczyć o tym, że jest to samogłoska (badania bazowały na amerykańskich samogłoskach).

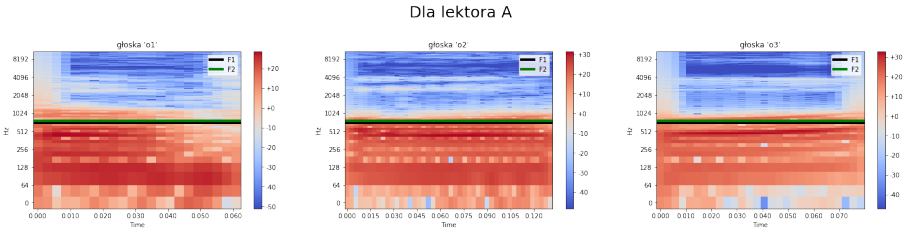
głoska (pl)	F_1 [Hz]	F_2 [Hz]
i	240	2400
u	250	595
a	850	1610
o	700	760

Tabela 1: Średnie wartości formantów w męskich głosach

4.2.1 Różne realizacje jednego fonemu

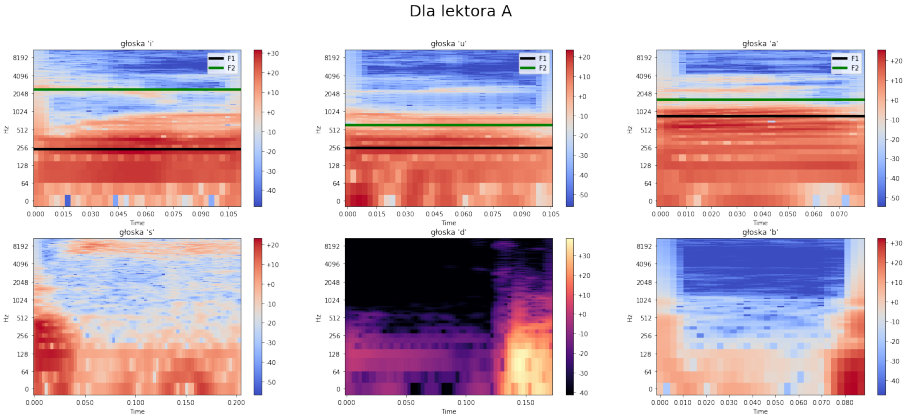


Rysunek 3: Widmo fonemu ”o” (lektor A)

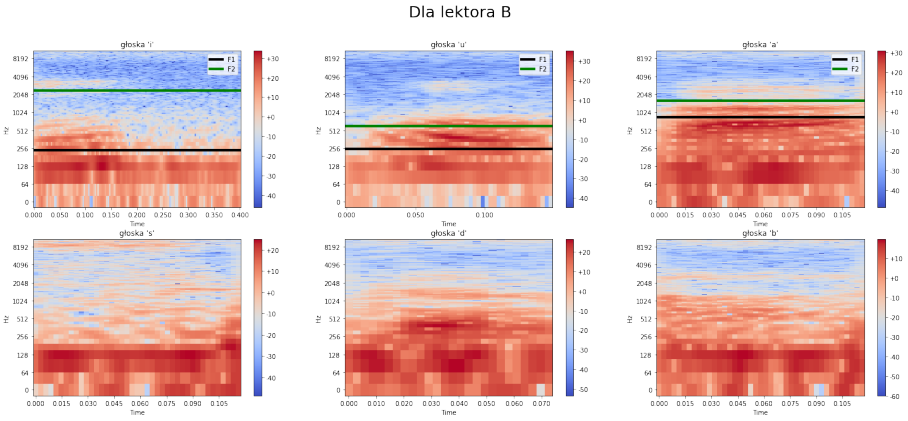


Rysunek 4: Spektogramy fonemu ”o” (lektor A)

4.2.2 Realizacje różnych fonemów dla różnych lektorów



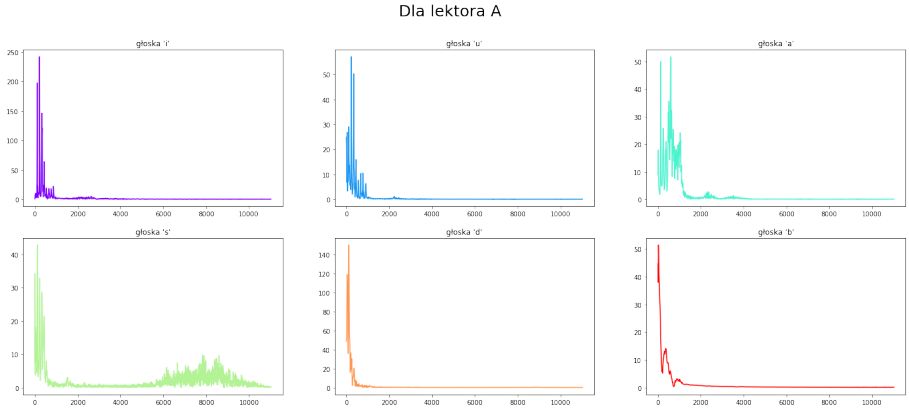
Rysunek 5: Spektogramy lektora A



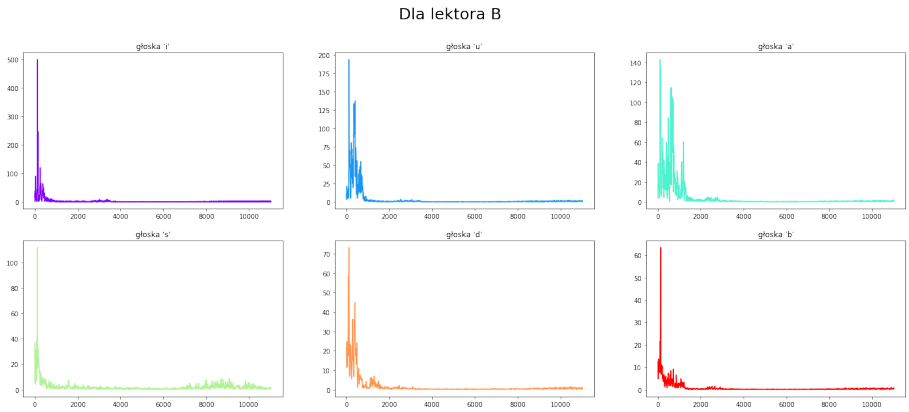
Rysunek 6: Spektogramy lektora B

4.3 Samogłoski i spółgłoski

W spektogramie można wyróżnić pasma częstotliwości z wysoką amplitudą na całej szerokości szczególnie w samogłoskach. U lektora A, który mocniej intonował poszczególne fragmenty wyrazów w nagraniach, różnice w tychże wykresach (Rys 3.) są bardzo wyraźne — w spółgłoskach "s", "d" i "b" ciężej jest zaobserwować formanty.



Rysunek 7: Widma lektora A



Rysunek 8: Widma lektora B