

Opis schematu przetwarzania

Do znalezienia początku fragmentów, w których mówiłem ja albo Maciej wykorzystałem podejście zaprezentowane na zajęciach do szukania różnych tonów. Do znajdowania miejsc zmiany mówiącego wykorzystałem funkcję `findpeaks()` z trzema parametrami pomocniczymi:

- `MinPeakHeight` – parametr mówiący jaką minimalną wysokość musi mieć szczyt żeby uznać za maksimum lokalne. W tym przypadku uznałem 0.15 za odpowiednią wielkość.
- `MinPeakDistance` – parametr mówiący jak daleko od siebie mają być znalezione szczyty. W tym przypadku wiem, że mówca zmieniał się mniej więcej co 1/6 trwania nagrania. Żeby wziąć pod uwagę pewne pomyłki programu postanowiłem, że odpowiednia wielkość to będzie 1/7 z ilości wszystkich próbek nagrania.
- `NPeaks` – parametr mówiący ile szczytów ma znaleźć funkcja.

Następnie po znalezieniu fragmentu znajduje jego maksymalną częstotliwość. Częstotliwości szukam z pomocą funkcji `fft()`. Na podstawie maksymalnej wartości modułu z szybkiej transformaty Fouriera. Jestem w stanie zdeterminować, czy fragment jest teraz badany. Wartość progowa jest równa 240.

Następnie program zlicza ilość słów w każdym z moich trzech segmentów. Do zliczania słów wykorzystałem prostą dylację i filtr progowy. Dylacja ma na celu zgeneralizowanie sygnału i pozbycie się skoków w amplitudzie, które mogłyby być błędnie zinterpretowane przez program jako początki lub końce słów. Maskę do dylacji miała długość 5000 próbek. Następnie tak uproszczony sygnał zamieniam na sygnał binarny, na podstawie amplitudy sygnału w danym miejscu. Wartością progową było 0.01. Potem korzystając z funkcji `strfind()` szukam początków i końców słów.

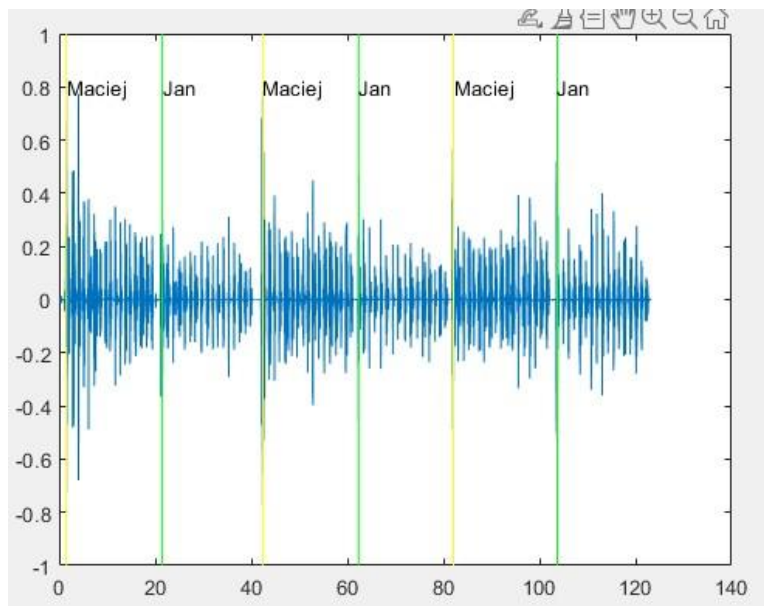
Mimo tego że metoda jest dosyć prosta to daje świetne wyniki. Największym kłopotem zaprezentowanego rozwiązania jest błąd w określaniu granic fragmentów. Ten błąd powoduje, że część słów z innego fragmentu może być podana na koniec badanego dźwięku lub początkowe słowo może zostać ucięte.

Tabela z opisem danych wejściowych

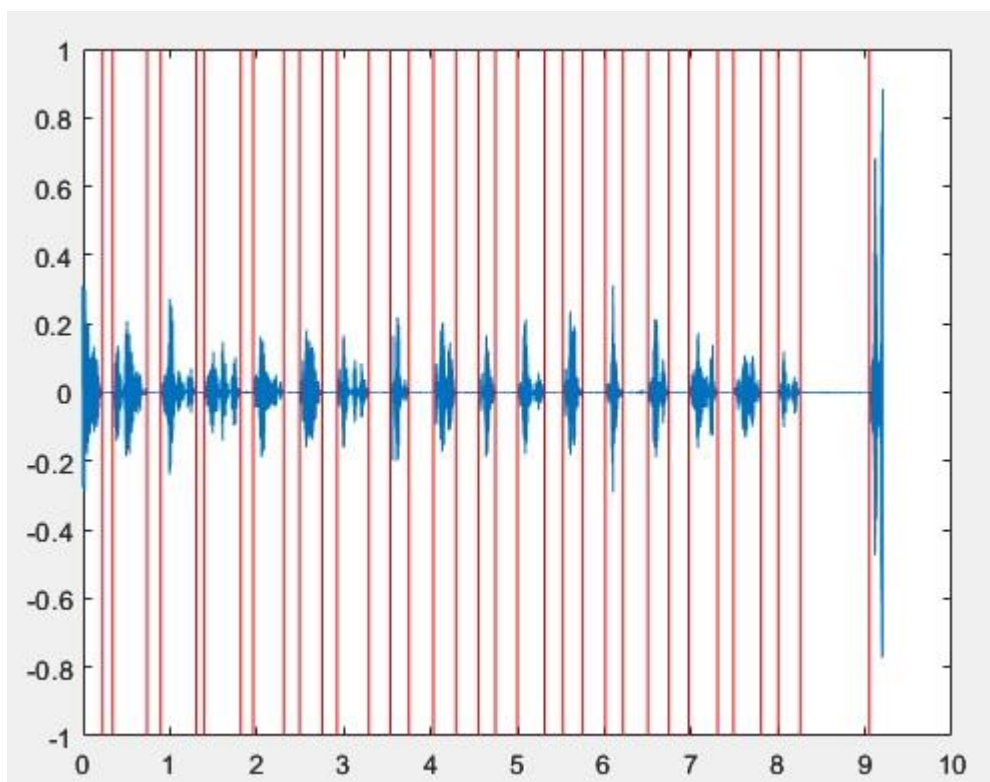
Numer fragmentu	Początek [s]	Koniec[s]	Ilość wypowiedzianych słów	Ilość słów obliczonych	Błąd[%]
2	21.4885	42.3602	17	17	0.000
4	62.3344	82.0979	16	17	6.2500
6	103.7005	123.2076	17	16	5.8824
Suma	-	-	50	50	0.000

Błąd z całości nagrania jest tutaj dosyć zaskakującą statystyką. Trochę więcej informacji niesie ze sobą średnia z błędów z każdego fragmentu, która wynosi 4.0441%.

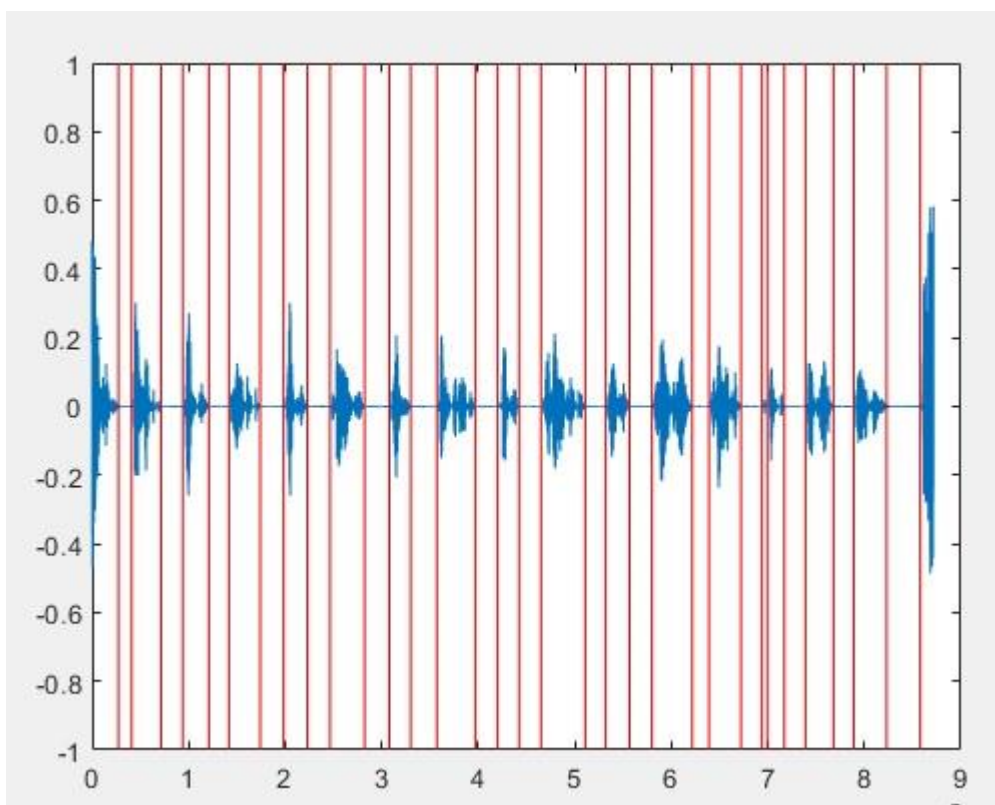
Wykresy wyników



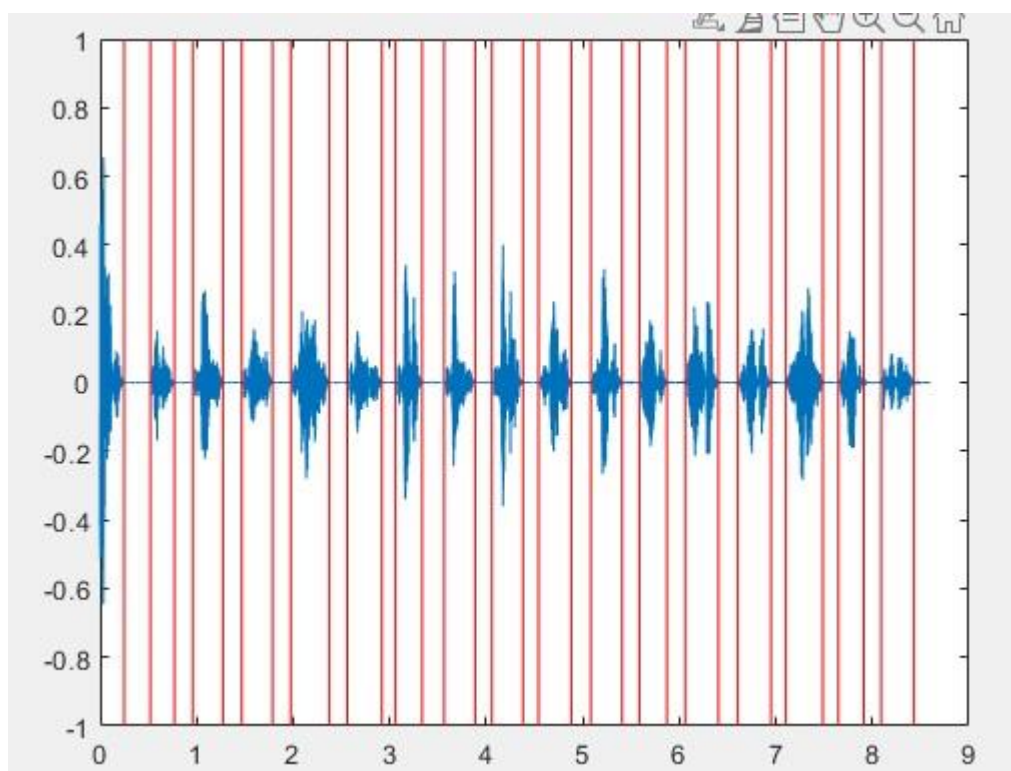
Początki poszczególnych fragmentów wypowiedzi



Początki i końce słów dla drugiego fragmentu wypowiedzi



Początki i końce słów dla czwartego fragmentu wypowiedzi



Początki i końce słów dla szóstego fragmentu wypowiedzi