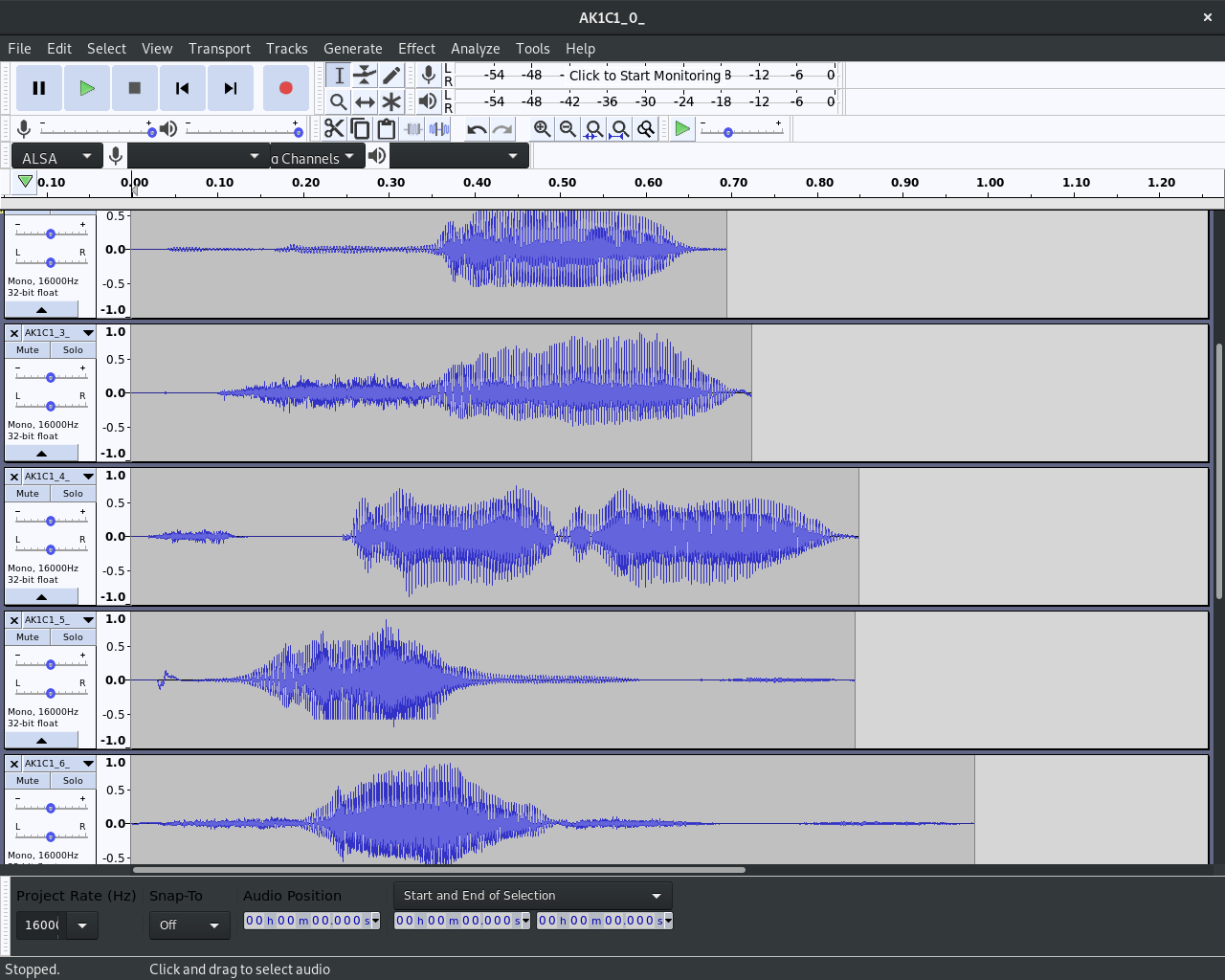
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DOKUMENTACJA PROJEKTU** | | | | Uczelnia **PP**Rok akademicki**2018/19** |
| Przedmiot: **przetwarzanie i rozpoznawanie dźwięku** | | | |
| Temat projektu: **Rozpoznawanie wypowiedzi z wykorzystaniem DTW** | | | |
| Specjalność:**ISWD** | Tryb / semestr:STAC. II ST / II | Grupa:środowa, 16:50 | Autorzy:Damian Jurga Marcin Drzewiecki | |
| Termin konsultacji:14 listopada | | | Termin oddania sprawozdania:21 listopada 2018 | |

Zadanie polega na przygotowaniu programu (w dowolnym języku) do rozpoznawania pojedynczych słów z danej grupy (cyfry, komendy, imiona…) z wykorzystaniem metody *dynamicznego zakrzywiania czasu* DTW (*Dynamic Time Warping)* oraz zbadaniu jakości rozpoznawania. Wypowiedzi z bazy CORPORA (częst. próbkowania 16kHz) są dostępne pod adresem: www.cs.put.poznan.pl/elukasik/PIRD18.

Etapy projektu:

1. Dokonać obserwacji przebiegów czasowych, widm i spektrogramów wypowiedzi.

Dokonano obserwacji przebiegów czasowych, widm i spektrogramów wypowiedzi za pomocą programu Aucacity. Na ich podstawie wstępnie ustalono odpowiednią długość ramki.

1. Dobrać odpowiednią długość ramki

Dobrano długość ramki: 0.0125s, oraz długość okna: 0.005s

1. Dokonać parametryzacji wypowiedzi z wykorzystaniem CC lub MFCC lub LPC (mogą to być gotowe procedury)

Dokonano parametryzacji za pomocą procedury MFCC z biblioteki python\_speech\_features.

1. Przygotować bazę wzorców z dostępnych wypowiedzi

Skorzystano z gotowej bazy wzorców dostępnej na wskazanej stronie.

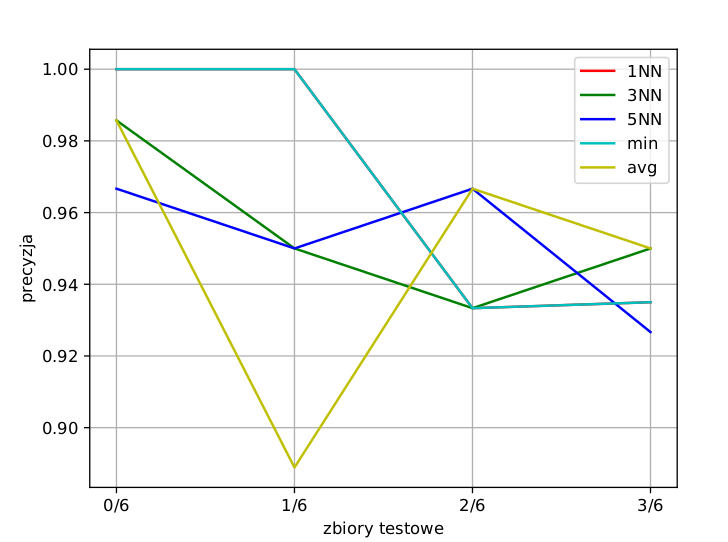
1. Napisać lub zaadaptować procedurę DTW

Napisano procedurę DTW, która okazała się działać zbyt wolno. Zaadaptowano procedurę DTW z pakietu tslearn.

1. Wykonać badania poprawności rozpoznawania dostępnych wypowiedzi.

Wykonano badania poprawności rozpoznawania dostępnych wypowiedzi. W tabeli przedstawiono jakość klasyfikacji w zależności od długości ramki oraz liczby współczynników cepstralnych.



Wykonano również badanie skuteczności różnych algorytmów. Na wykresie przedstawiono jakość klasyfikacji w zależności od tego, jaka część zbioru stanowiła zbiór testowy.

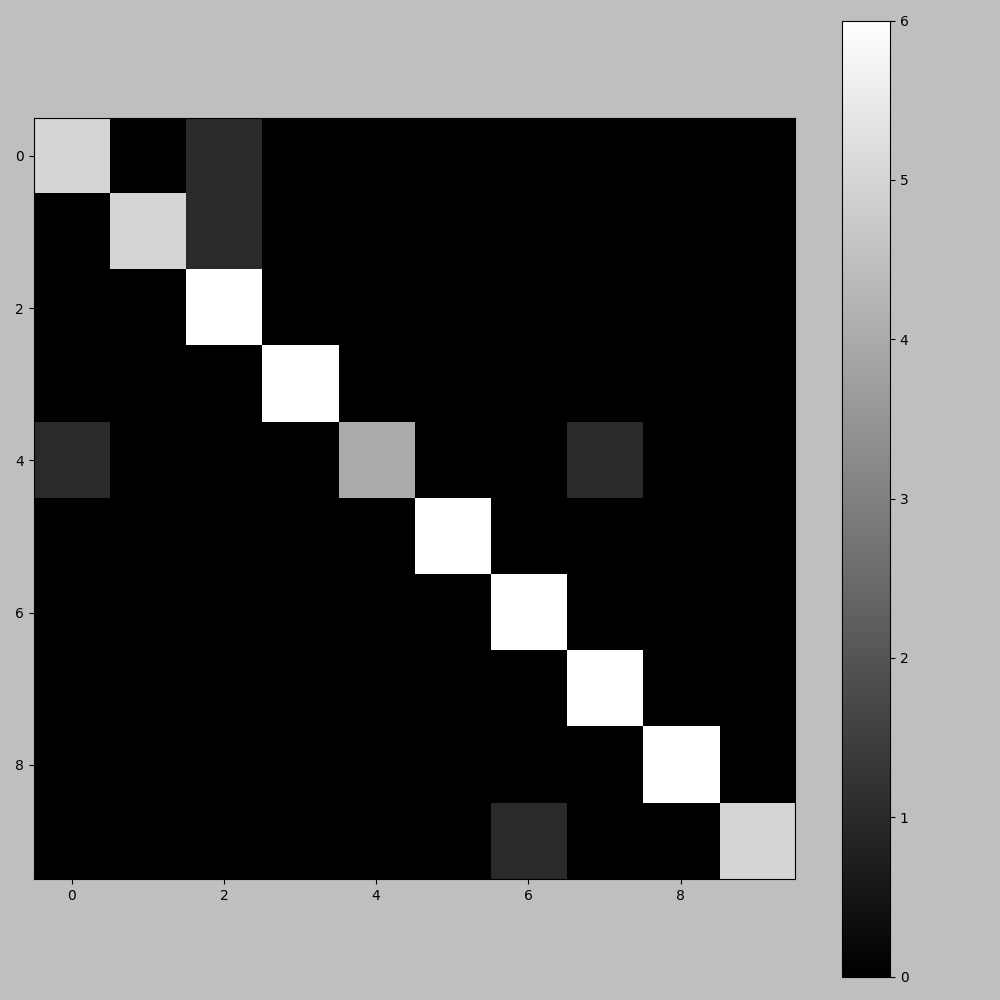
W związku z tym, że najlepsze wyniki dawał klasyfikator najbliższego sąsiada, wybrano ten algorytm jako ostateczny.

Na osi poziomej ułamek oznacza część zbioru, która stanowi zbiór testowy. Pozostała część zbioru to część treningowa.

1. Można dokonać wizualizacji mapy podobieństw ramek według wybranej miary oraz najkrótszej ścieżki dopasowania.

? ?

1. W raporcie umieścić opis poszczególnych etapów projektu, wyniki rozpoznawania i ewentualne rysunki.



Przedstawiono wynik działania dla 1NN i trzech współczynników cepstralnych. Dla każdego nagrania sprawdzano jakie jest najbliższe mu nagranie, a następno przypisywano mu klasę tego sąsiada. Dla algorytmu obliczono miarę *precision* oraz *recall*, które wynosiły odpowiednio 0.93 (55/60) oraz 0.92.