

Data oddania: _____

Ocena: _____

Paweł Musiał 178726
Łukasz Michalski 178724

Zadanie 4:
**Rozpoznawanie izolowanych słów w sygnale
mowy.***

Spis treści

1. Cel	2
2. Wprowadzenie	2
2.1. MCFF (<i>Mel-frequency cepstral coefficients</i>)	2
2.2. Algorytm DTW (<i>Dynamic Time Warping</i>)	2
3. Opis implementacji	2
3.1. MCFF (<i>Mel-frequency cepstral coefficients</i>)	2
3.2. Algorytm DTW (<i>Dynamic Time Warping</i>)	2
4. Wyniki	2
5. Dyskusja	2
6. Wnioski	2
Literatura	2

* SVN: https://serce.ics.p.lodz.pl/svn/labs/poid/at_sr0830/lmpm@

1. Cel

Realizacja zadania polega na stworzeniu aplikacji umożliwiającej obliczanie reprezentacji sygnału audio w postaci ciągów wektorów współczynników MFCC i porównywanie ich za pomocą algorytmu DTW. Należy stworzyć bazę zawierającą przynajmniej 10 różnych słów (przykładowo: „zero”, „jeden”, ..., „dziewięć”) i wykorzystać ją do rozpoznawania słowa wypowiedzianego przez użytkownika. W celu poprawy wyników każde słowo zawarte w bazie może być reprezentowane przez kilka wzorców (np. nagranych przez różne osoby, albo w różnych warunkach akustycznych).

Oprócz ostatecznego wyniku rozpoznania należy zaprezentować wyniki porównań dla wszystkich słów z bazy oraz tablice g (preferowana metoda – w postaci obrazu, reprezentującego wartości $g[i, j]$ za pomocą np. odcieni szarości). Należy rozważyć metodę modyfikacji algorytmu DTW pozwalającą na dopasowanie fragmentu słowa zamiast całości. Należy zaimplementować ograniczenia globalne zgodnie z przydzielonym wariantem, przy czym powinna też istnieć możliwość wyłączenia tych ograniczeń, tak aby ścieżka mogła mieć dowolny kształt.

- Ograniczenie globalne typu Sakoe and Chiba band
- **Ograniczenie globalne typu Itakura parallelogram**

2. Wprowadzenie

Rozpoznawanie mowy jest klasycznym problemem przetwarzania dźwięku, dla którego w ciągu minionych dziesięcioleci zaproponowano wiele rozwiązań. Wyróżniamy tu zasadniczo problem prostszy, polegający na rozpoznawaniu izolowanych słów oraz zadanie rozpoznawania mowy ciągłej. W obu przypadkach należy przyjąć założenia odnośnie sposobu reprezentacji i parametryzacji sygnału mowy oraz odnośnie metod dopasowania danych do wzorca w sposób niezależny od czasu trwania i zmian szybkości analizowanej wypowiedzi.

2.1. MCFF (*Mel-frequency cepstral coefficients*)

2.2. Algorytm DTW (*Dynamic Time Warping*)

3. Opis implementacji

3.1. MCFF (*Mel-frequency cepstral coefficients*)

3.2. Algorytm DTW (*Dynamic Time Warping*)

4. Wyniki

5. Dyskusja

6. Wnioski

Literatura

- [1] Fast Fourier Transform (FFT) <http://www.cmlab.csie.ntu.edu.tw/cml/dsp/training/coding/transform/fft.html>