

ISS Projekt 2019 / 20

Michal Koval - xkoval17

xkoval17@stud.fit.vut.cz

16.12.2019

Úloha 1:

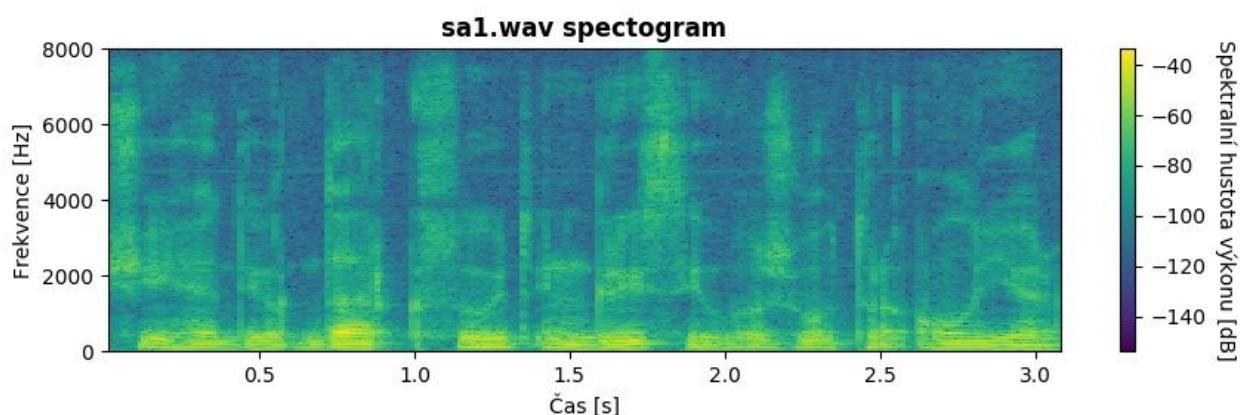
Názov	Dĺžka vo vzorkách	Dĺžka v sekundách	Použitie
sa1.wav	81200	5.08	c
sa2.wav	70800	4.42	c
si938.wav	106000	6.62	c
si1568.wav	52400	3.27	c
si2198.wav	96800	6.05	c
sx38.wav	62000	3.88	c
sx128.wav	69200	4.33	c
sx218.wav	78400	4.90	c
sx308.wav	67200	4.20	c
sx398.wav	72400	4.53	c

Úloha 2:

Názov	Dĺžka vo vzorkách	Dĺžka v sekundách	Použitie
q1.wav	19625	1.23	c
q2.wav	7716	0.48	c

Na zaznamenania nahrávok som použil môj mobilný telefón značky OnePlus. Default recorder v telefóne neposkytoval možnosť zmeny formátu výstupného súboru, tak som stiahol aplikáciu Easy Voice Recorder, kde som frekvenciu vzorkovania na 16kHz a .wav kódovanie. Správnosť parametrov nahrávok som následne ešte skontroloval s programom soxi. Ticho som odstránil v programe Audacity.

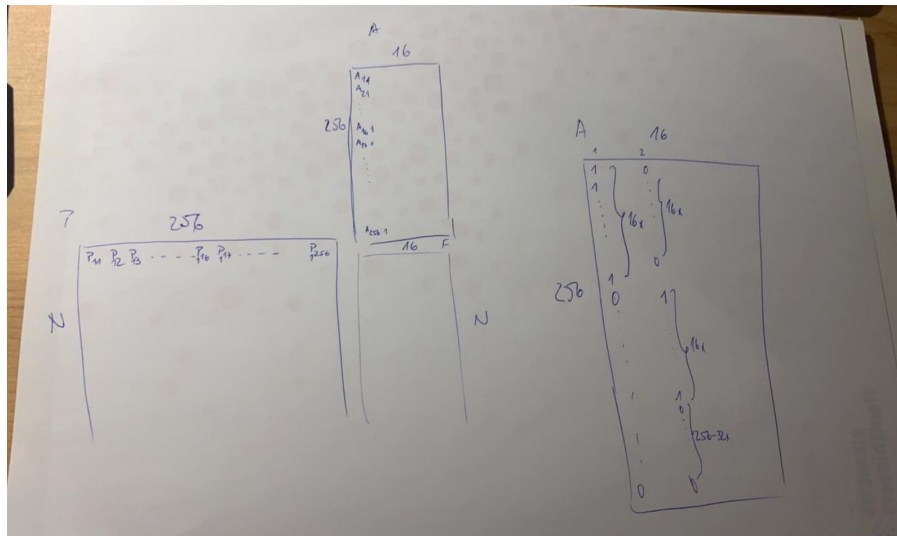
Úloha: 3



Úloha 4:

Parametre počítam použitím funkcie spectrogram. Z nej dostanem maticu $256 \times N$, kde N je počet segmentov a závisí na dĺžke nahrávky. Teda ku každému segmentu pripadá 256 hodnôt. Ďalej spravím lineárnu banku filtrov, teda každý segment zrátam vektor o 16 parametroch. Jednotlivé parametre sú súčet 16 hodnôt zo stĺpca spektra. Pri rátaní ďalšieho parametra sa posunieme o 16 indexov, posledný parameter je teda súčet posledných 16 indexov. Toto sa vykoná pre každý segment.

Matica A:



Úloha 5:

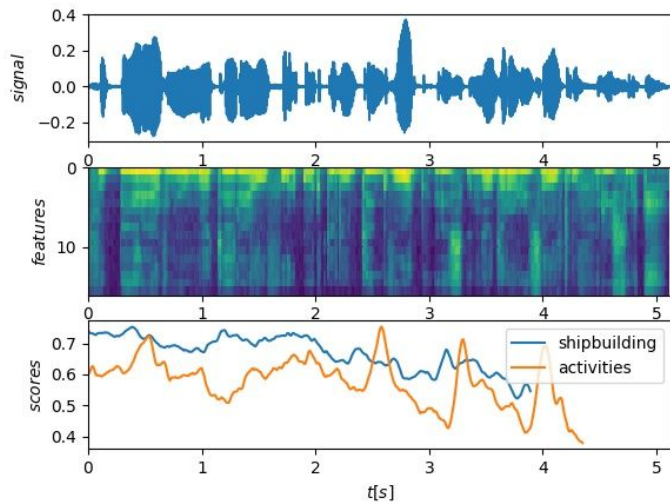
```
def get_pearson_score(q_f, s_f):
    score = []
    q_f = np.transpose(q_f)    # transponujem maticu parametrov query a aj sentence
    s_f = np.transpose(s_f)    # aby obe boli rozmeru Nx16 a aby som ich mohol jednoducho
                                # vložiť do pearsonr()
    score_len = s_f.shape[0] - q_f.shape[0]    # max počet cyklov bude počet
                                                # segmentov vety - počet segmentov slova,
                                                # pretože na konci vety by nebolo dost
                                                # zvyšných segmentov na porovnanie so
                                                # segmentmi slova

    sum = 0
    n_sum = 0
    for pp in range(score_len):                # prechod všetkých potencionálnych pozícií
        for q in range(q_f.shape[0]):          # prechod cez všetky segmenty queue
            r, p_value = pearsonr(q_f[q], s_f[pp + q])    # pearsonova funkcia
            if not np.isnan(r):                # ak nie je variácia v parametroch vracia nan
                sum += r    # priratanie výsledku porovnania aktualneho segmentu
                n_sum += 1

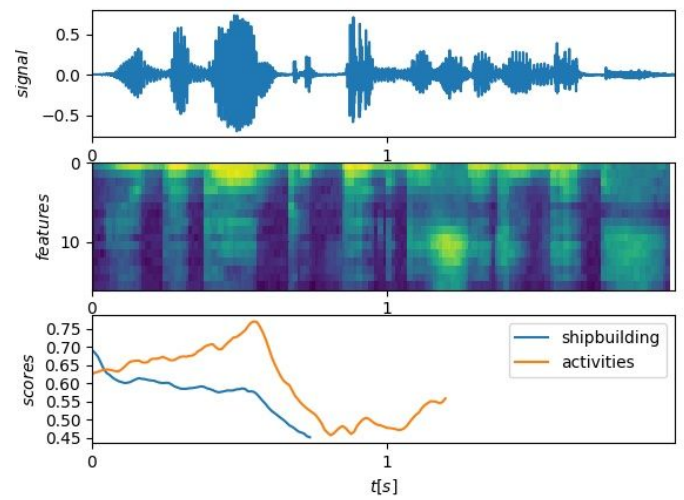
        sum = sum / n_sum    # predelenie počtom prirataných výsledkov
        score.append(sum)   # vloženie do poľa, toto je koniec výpočtu score pre jednu pp
        sum = 0
        n_sum = 0
    return score
```

Úloha 6:

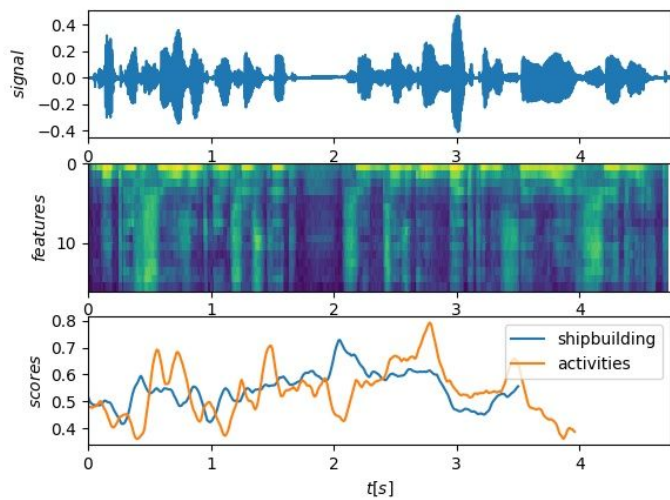
"shipbuilding" and "activities" vs ../sentences/si938.wav



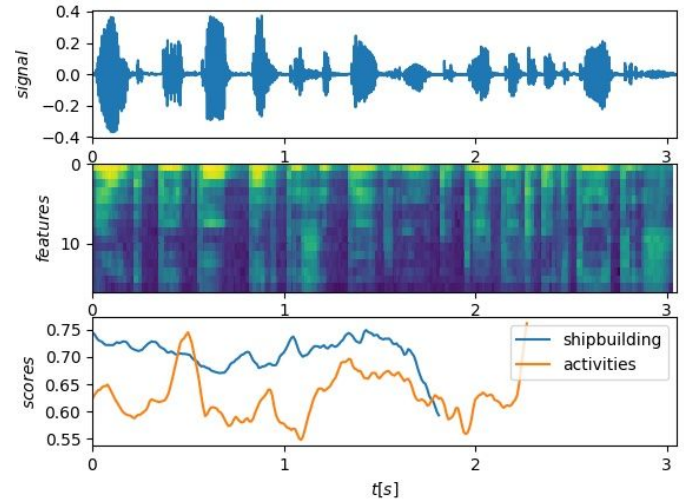
"shipbuilding" and "activities" vs ../sentences/si1568.wav



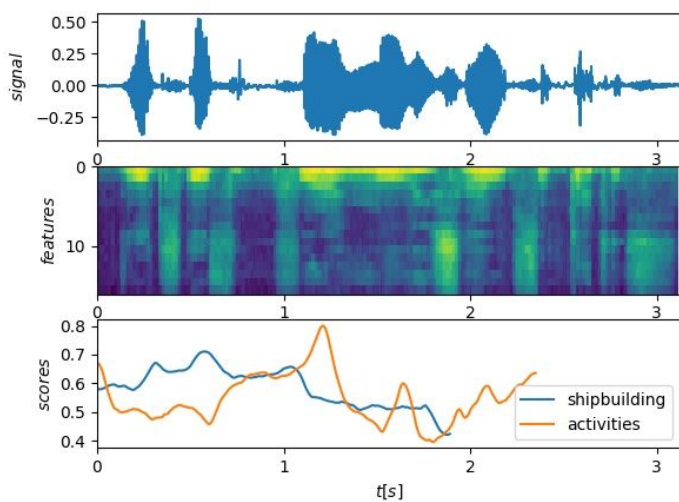
"shipbuilding" and "activities" vs ../sentences/si2198.wav



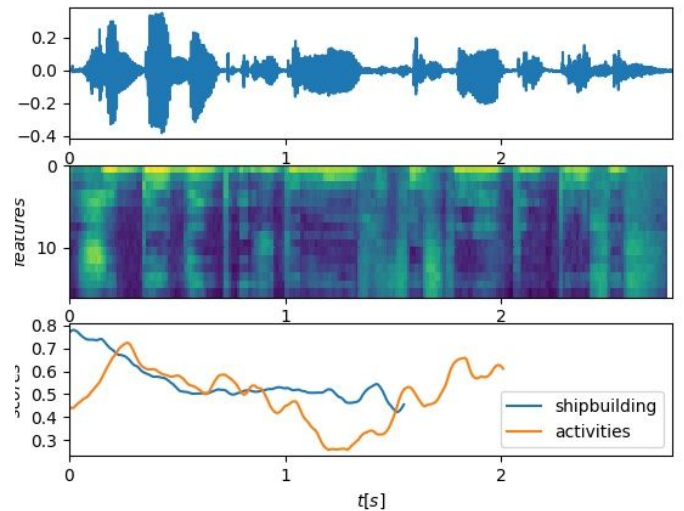
"shipbuilding" and "activities" vs ../sentences/sx38.wav



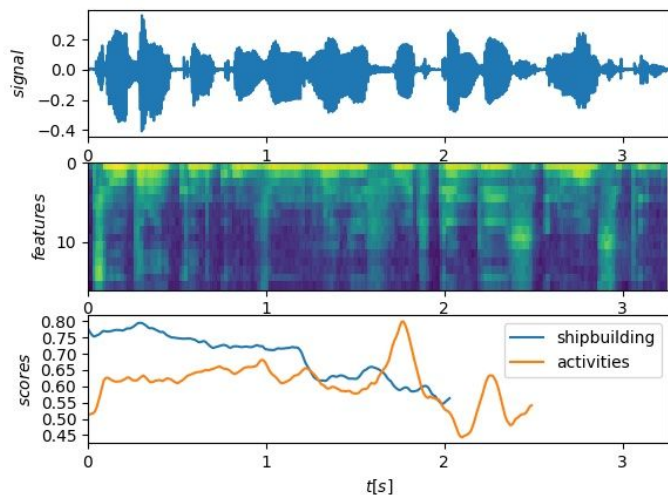
"shipbuilding" and "activities" vs ../sentences/sx128.wav



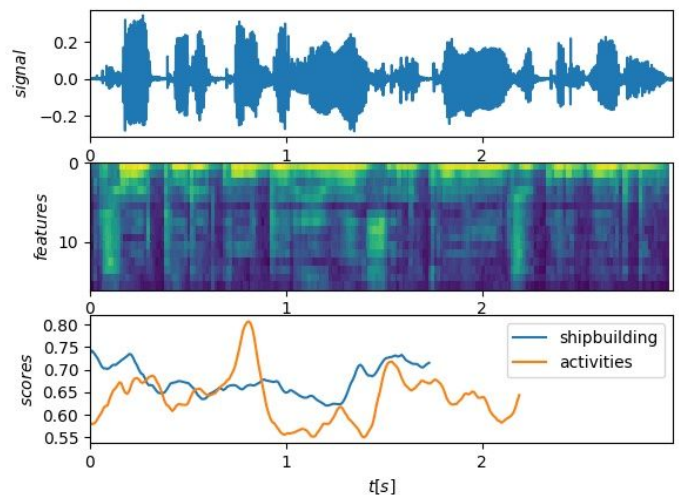
"shipbuilding" and "activities" vs ../sentences/sx308.wav



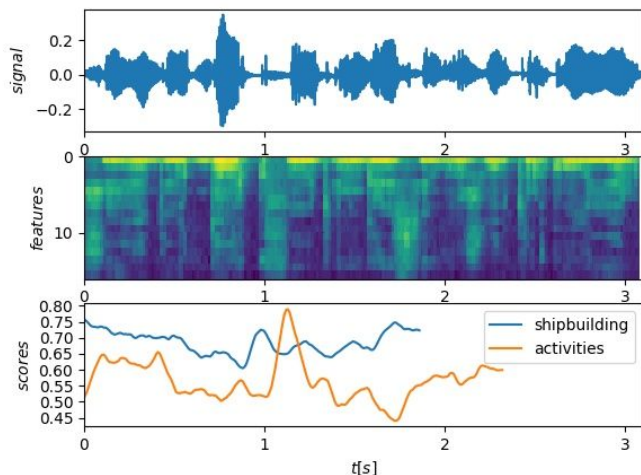
"shipbuilding" and "activities" vs ../sentences/sx398.wav



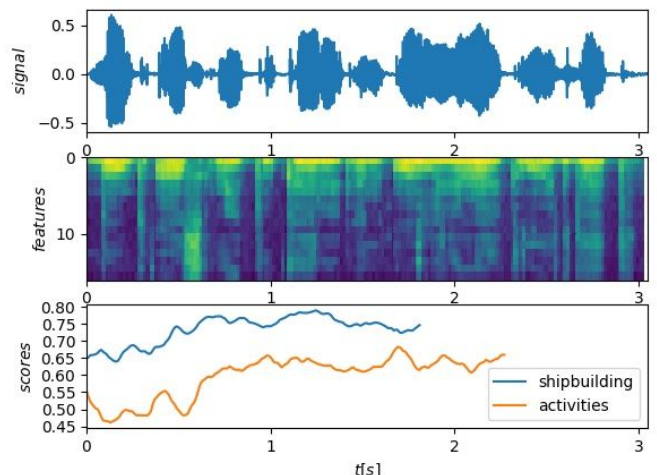
"shipbuilding" and "activities" vs ../sentences/sx218.wav



"shipbuilding" and "activities" vs ../sentences/sa1.wav



"shipbuilding" and "activities" vs ../sentences/sa2.wav



Úloha 7:

Ako queries som si zvolil slová shipbuilding a activities, lebo sú dlhé. Shipbuilding sa nachádza na začiatku vety sx308. Z grafu pre sx308 môžeme vidieť v prvej sekunde prudký nárast idúci po skóre zhruba 0.79. Na ostatných grafoch vidno že túto hodnotu nadobúda slovo shipbuilding zriedkakedy. Preto prah pre q1, teda shipbuilding, bude 0.8. Rovnakým čítaním grafov určíme prah pre q2, activities, na 0.77. Na konci nahrávky sx38 je prudký nárast ale nedosahuje úplne hodnotu 0.8. Avšak pri tomto slove je aj v ostatných grafoch veľa nárastov na túto hodnotu, takže predpokladám, že budeme mať veľa nepravdivých hits.

Úloha 8:

Hit	Výskyt	Vzorky
q1_sx308_0	áno	320-19945
q1_sx398_0	nie	
q1_sa2_0	nie	
q2_sa1_0	nie	
q2_si2198_0	nie	
q2_sx38_0	áno	36320-48677
q2_sx128_0	nie	
q2_sx398_0	nie	
q2_si1568_0	nie	
q2_sx218_0	nie	

Úloha 9:

Dostal som výsledných 10 hitov, čo samozrejme nie je správne. V každej vete jeden hit, pričom ich malo byť iba 2. Skúšal som sa hrať s prahmi ale vyššie už nemôžem ísť, pretože potom strácam aj správne hity. Je možné že je táto nepresná detekcia zapríčinená mojou nedokonalou výslovnosťou. Ďalším faktorom môže byť kvalita samotných nahrávok, boli nahrávané na obyčajnom telefóne v miestnosti kde neustále mierne hučí chladnička, aj keď tento hukot nie je na nahrávkach počuť. Samozrejme nevylučujem aj možnosť, že by moje riešenie mohlo byť vadné.