

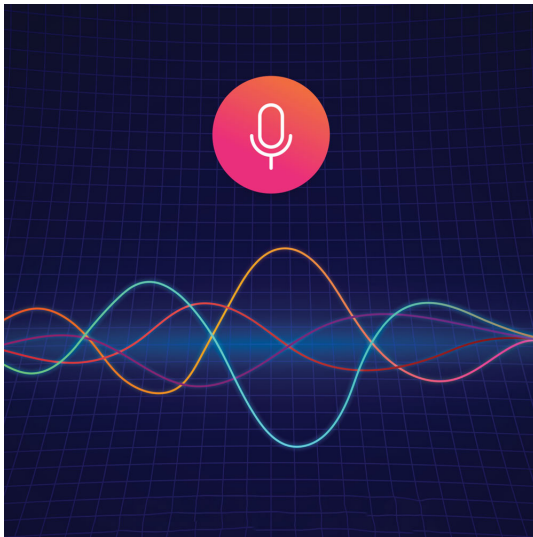
Speech emotion recognition

Računarska inteligencija

Jelena Keljać, Aleksandra Pešić

Matematički fakultet

Septembar 2022



- ▶ "Problem"
- ▶ MFCC
- ▶ Chroma
- ▶ Zero-crossing rate

- ▶ Predstavlja skup vrednosti, tj. koeficijenata koje sažeto opisuju oblik spektralnog sloja i često se može koristiti za opisivanje boje tona.
- ▶ Mel-skala je skala odnosa percipirane frekvencije od strane ljudi sa zapravo izmerenom frekvencijom zvuka.

Chroma and Zero cross rate

- ▶ Chroma se odnosi na 12 različitih klasa visine tonova i predstavlja intezitet zvuka za svaku od 12 različitih klasa koje predstavljaju 12 različitih polutonova muzičke oktave.
- ▶ Stopa prolaska kroz nulu meri koliko puta u određenom vremenskom intervalu je amplituda zvučnog signala prošla kroz vrednost nula.

Ulazni podaci

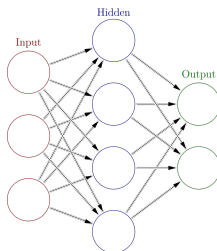
- ▶ Za pravljenje našeg modela korišćena su tri seta podataka:
 1. Ryerson Audio-Visual Database of Emotional Speech and Song (Ravdess)
 2. Toronto emotional speech set (Tess)



- ▶ `librosa.feature.mfcc`

Model

- ▶ Veštačke neuronske mreže oponašaju biološke neuronske mreže.
- ▶ Svaki sloj se sastoji od jedinica, neurona, povezanih sa neuronima iz susednih slojeva težinskim koeficijentima.



- ▶ `sklearn.neural_network.MLPClassifier`
- ▶ `tf.keras.Sequential`

Tačnost modela je oko 84%.

Confussion matrix for model 1

	neutral	happy	sad	angry	fear	disgust	surprise
neutral	170	9	6	2	1	0	1
happy	3	160	3	7	4	1	11
sad	3	4	158	2	1	5	3
angry	6	2	6	129	0	0	7
fear	1	5	1	2	207	4	0
disgust	1	1	4	3	9	159	2
surprise	1	5	1	5	2	3	152

Confussion matrix for model 2

	neutral	happy	sad	angry	fear	disgust	surprise
neutral	172	9	5	1	0	1	1
happy	0	155	7	3	4	9	11
sad	2	5	148	3	6	9	3
angry	7	2	8	122	2	3	6
fear	0	7	3	2	199	7	2
disgust	0	4	5	10	15	141	4
surprise	3	5	3	7	4	6	141



The End