

1 Uvod

Emocija je kompleksno psihofiziološko iskustvo stanja uma pojedinog individualca koje je podložno interakcijama s biokemijskim (unutarnjim) i okolinskim (vanjskim) utjecajima.

Postoji više klasifikacija emocija te više desetaka diskretnih emocionalnih stanja. Tema emocija je veoma kompleksna i usko povezana s psihologijom.

Područje automatskog prepoznavanja emocija na temelju govora prvi je ozbiljno počeo istraživati Klaus Scherer.

Ne postoji automatski klasifikator emocija za hrvatski jezik.



2 Opis problema

Cilj diplomskog rada je da se istraži te izgradi automatski sustav za klasifikaciju diskretnih emocionalnih stanja u hrvatskom jeziku na osnovu akustičkih značajki govora. Emocionalna stanja koja se klasificiraju su:

- ljutnja
- strah
- tuga
- sreća
- neutralno

Statistika korpusa emocija nakon ocjenjivanja

Emocije	ljutnja	strah	tuga	sreća	neutralno	Ukupno
Broj snimaka	231	33	83	77	72	496
Trajanje [mm:ss]	13:27	2:26	8:57	5:27	10:59	41:16

Tri glavna problema kod izgradnje sustava su:

1. teško je skupiti odgovarajući korpus sa snimkama stvarnih emocija i visoke kvalitete
2. često nije moguće primijeniti druge korpuse zbog međukulturalnih razlika u ekspresiji pojedine emocije
3. nije jednostavno pronaći odgovarajuće značajke

3 Metoda

Istražuju se dvije metode klasifikacije:

$$\hat{P}(O|M) = \max_X \left\{ a_{x(0)x(1)} \prod_{t=1}^T b_{x(t)}(o_t) a_{x(t)x(t+1)} \right\}$$

Vjerodostojnost za HMM

1. skriveni Markovljevi modeli (eng. *Hidden Markov Models – HMM*)
2. mješoviti Gaussovi modeli (eng. *Gaussian Mixture Models – GMM*)

$$P(o|\lambda) = \sum_{j=1}^M w_j g_j(o)$$

Vjerodostojnost za GMM

Razlika: HMM sadrži vremensku dinamiku, dok je GMM statičan.

$$b_j(o_t) = g_j(o_t) = \frac{1}{\sqrt{(2\pi)^p |\Sigma_j|}} e^{-\frac{1}{2}(o_t - \mu_j)^T \Sigma_j^{-1} (o_t - \mu_j)}$$

Gaussova komponenta gustoće vjerojatnosti

Markovljev model je automat s konačnim brojem stanja gdje se prijelazi iz stanja i u stanje j događaju u vremenskim koracima t s vjerojatnošću a_{ij} . Svaki prijelaz popraćen je opservacijom o_t nastalom iz gustoće vjerojatnosti $b_j(o_t)$. Prijelaz stanja nije poznat, prema tome radi se o skrivenom Markovljevom modelu.

Zadatak treniranja HMM-a je za zadani skup sekvenci opservacija pronaći najbolji skup prijelaza $\{a_{ij}\}$ i gustoća vjerojatnosti $\{b_j(o_t)\}$.

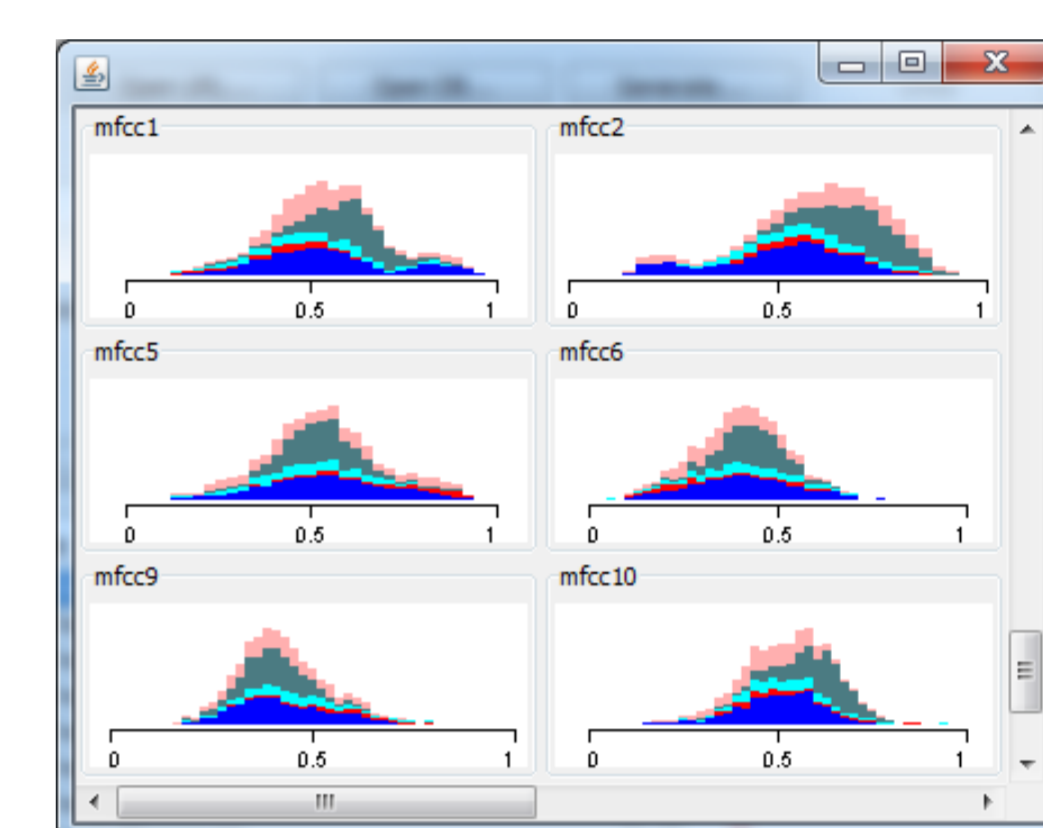
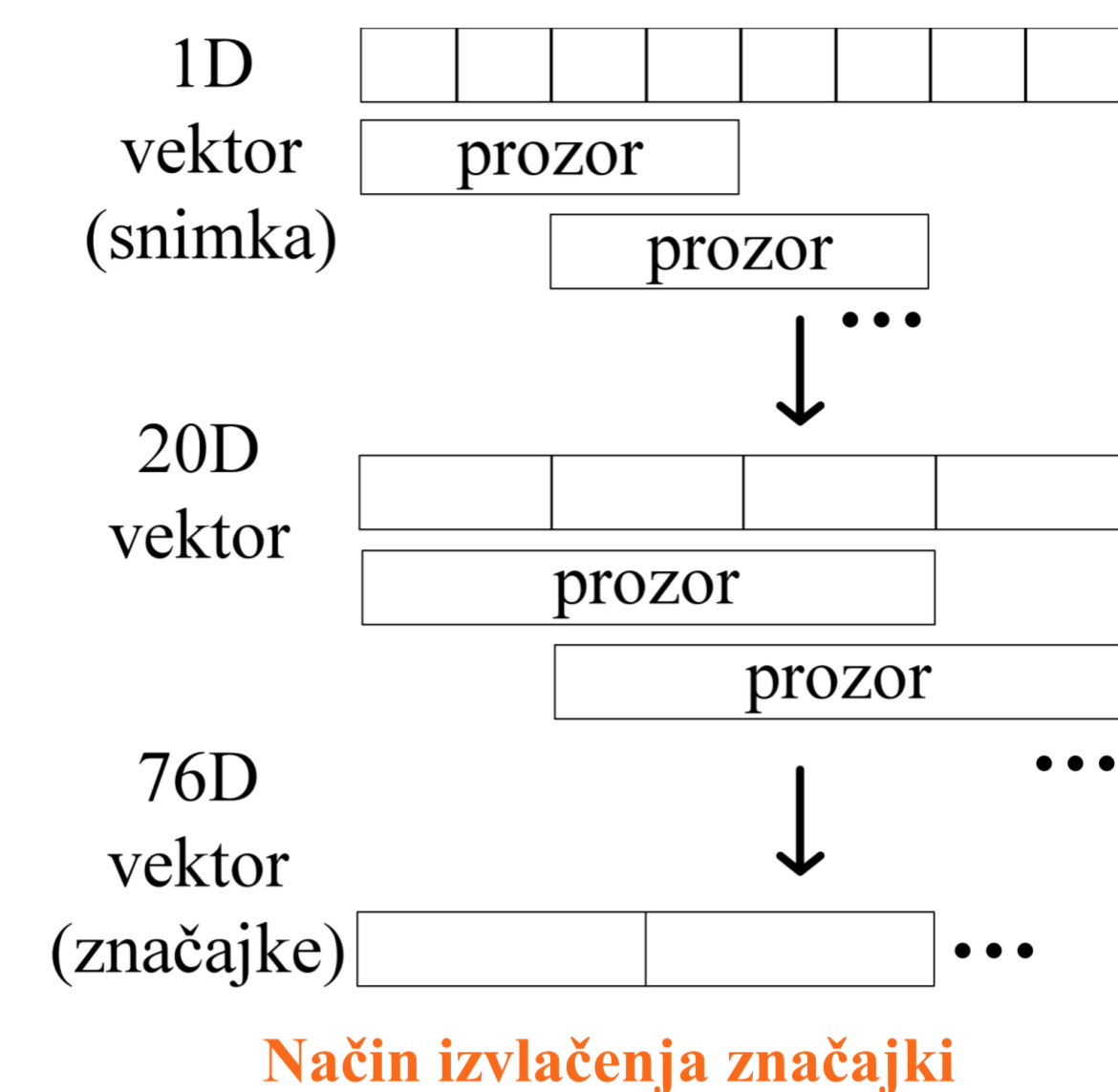
Zadatak prepoznavanja je odabir modela koji daje najveću vjerodostojnost za dani niz opservacija.

Mješoviti Gaussovi modeli su parametarske funkcije gustoće vjerojatnosti sačinjene od sume otežanih Gaussovih komponenta gustoće vjerojatnosti.

Zadatak treniranja GMM-a svodi se na pronalazak parametara w_j, η_j, Σ_j koji se na najbolji način poklapaju s distribucijom opservacija za treniranje.

Zadatak prepoznavanja je odabir modela koji daje najveću vjerodostojnost za danu opservaciju.

Jedan model predstavlja jedan razred, tj. emocionalno stanje, a opservacija predstavlja ulazni vektor sa značajkama.

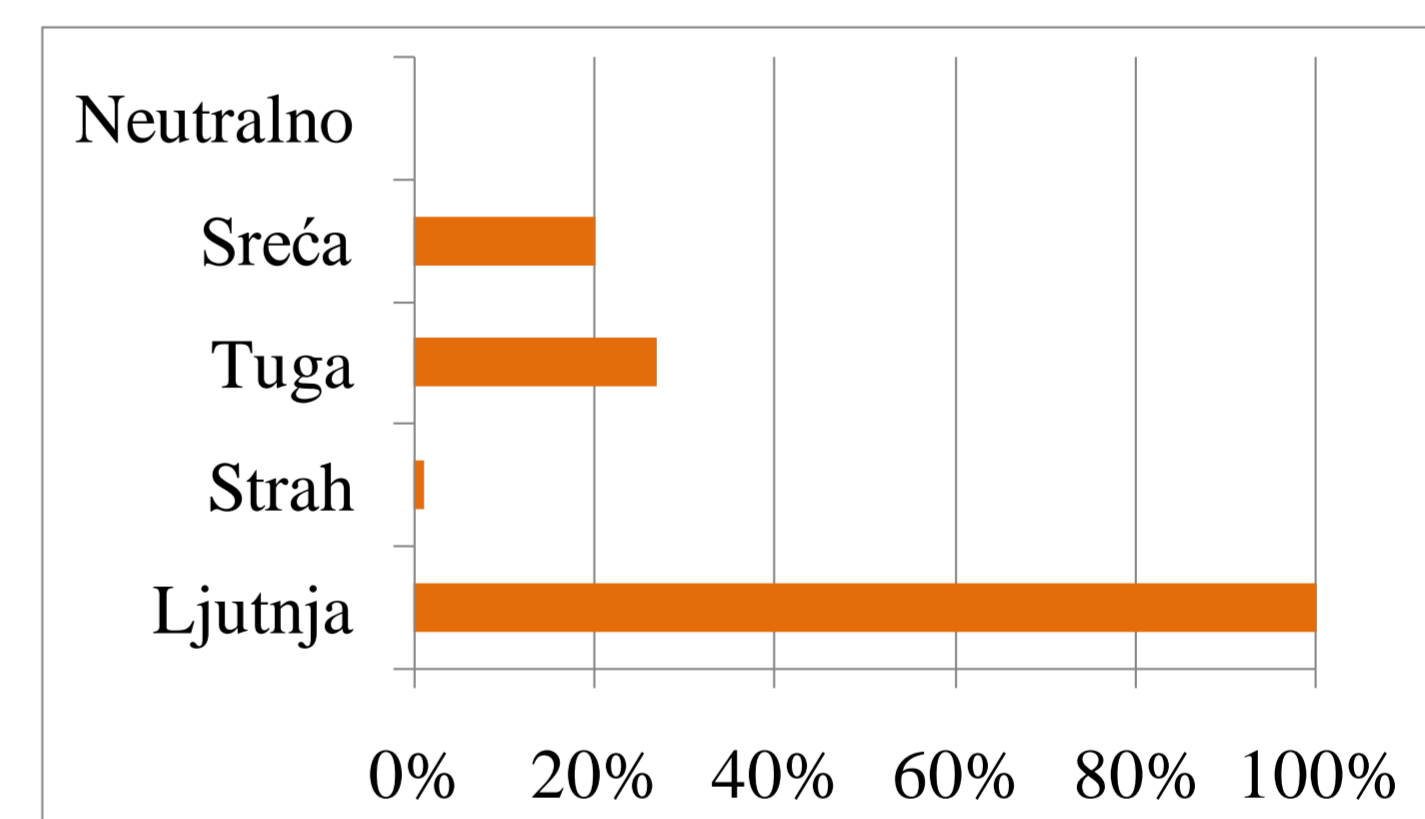


Statistički prikaz nekoliko značajki



4 Rezultati

Klasifikatori HMM i GMM testirani su na različitim izvučenim značajkama te s različitim parametrima koji podešavaju HMM klasifikator. Za HMM najbolji postotak prepoznavanja je 47.2%, a za GMM je 43.4%.



Prepoznavanje jedne snimke pomoću HMM-a

Matrice konfuzije za najbolji HMM klasifikator (47.2%) i GMM klasifikator (43.3%)

	lj.	st.	tu.	sr.	ne.		lj.	st.	tu.	sr.	ne.
ljutnja (66.7%)	16	4	2	2	0	ljutnja (83.3%)	20	0	1	3	0
strah (50.0%)	2	2	0	0	0	strah (0.0%)	3	0	0	1	0
tuga (22.2%)	1	1	2	2	3	tuga (0.0%)	2	0	0	5	2
sreća (50.0%)	3	1	0	4	0	sreća (25.0%)	4	0	2	2	0
neutralno (12.5%)	4	1	0	2	1	neutralno (12.5%)	4	1	0	2	1

5 Zaključak

Treniranje pomoću HMM-a dalo je malo bolje rezultate od GMM-a, međutim ti rezultati u oba slučaja nisu zadovoljavajući.

Ipak, rezultati pokazuju relativan uspjeh, ako se uzme u obzir činjenicu da bi za slučajni odabir postotak prepoznavanja bio 20%.

Razlozi dobivanja niskih postotaka prepoznavanja, koji istovremeno mogu poslužiti i kao polazišta za buduće unaprjeđenje sustava, su:

- korpus koji se sastoji od raznih izvora i njegova loša kvaliteta zvuka
- ipak neodgovarajuće izvučene značajke