

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 374

**GRADNJA I EVALUACIJA KORPUSA
HRVATSKOG EMOCIONALNOG GOVORA**

Emilija Damić

Mentor: prof.dr.sc. Davor Petrinović

Zagreb, lipanj 2012.

Zahvaljujem mentoru prof.dr.sc. Davoru Petrinoviću.

Zahvaljujem Branimiru Dropuljiću, dipl.ing., na izradi aplikacije za eksperiment te na beskonačnom strpljenju.

Zahvaljujem dr.sc. Dragutinu Ivanecu i dr.sc. Mirjani Tonković s Katedre za eksperimentalnu psihologiju te dr.sc. Aniti Lauri Korajlija s Katedre za zdravstvenu i kliničku psihologiju pri Filozofskom fakultetu u Zagrebu.

Zahvaljujem svim studentima s Odsjeka za psihologiju i Odsjeka za fonetiku pri Filozofskom fakultetu u Zagrebu koji su sudjelovali u eksperimentu.

Zahvaljujem Vedranu koji me uvjerio da ako želim, onda i mogu.

Zahvaljujem svojoj sestri Heleni za svu podršku.

Zahvaljujem svojoj mami bez čijih bi savjeta i potpore fakultetske obveze bile daleko teže.

Naposlijetku, zahvaljujem svojem dragom tati koji me čuva s neba, a njemu i posvećujem ovaj Diplomski rad.

Sadržaj

1. Uvod.....	1
2. Emocije	2
2.1. Vrste emocija	3
2.2. Reprerzentacija emocija.....	3
2.2.1. Diskretni prikaz emocija.....	3
2.2.2. Kontinuirani prikaz emocija.....	4
3. Prikupljanje audio-snimaka	6
4. Eksperiment.....	8
5. Modeliranje emocija Gausovim modelima.....	11
5.1. Modeliranje histograma za svaku emociju	12
5.2. Aproksimacija histograma Gausovom funkcijom dviju varijabli	15
5.3. Trajektorije koje opisuju pomak Gausovih modela	18
5.4. Treniranje Gausovih modela	22
5.5. Testiranje Gausovih modela.....	27
6. Modeliranje emocija GMM modelima.....	29
6.1. Treniranje GMM modela	30
6.2. Testiranje GMM modela.....	41
7. Modeliranje emocija GMM modelima s preferencijama	43
7.1. Treniranje GMM modela s preferencijama	43
7.2. Testiranje GMM modela s preferencijama	51
8. Usporedba uspješnosti modela.....	53
9. Zaključak.....	54
10. Literatura	55
11. Sažetak	56
12. Abstract.....	57

1. Uvod

Emocije su oduvijek izazivale pažnju, kako znanstvenika i filozofa, tako i ljudi u svakodnevnim situacijama. Emocije se mogu iskazati na razne načine, a u ovome Diplomskom radu bavit ćemo se svjesnim i nesvjesnim ispoljavanjem emocija govorom.

Razvojem tehnologije pokazalo se zanimljivim izazovom prepoznavanje emocija iz glasa govornika, a ovaj rad služi kao svojevrsni uvod u tu problematiku. U predstojećim poglavljima bit će dana definicija emocija te opisani načini reprezentacije emocija, koje čine temelj statističke analize provedene kroz ovaj rad. Također ćemo se upoznati s vrstom eksperimenta koja je potrebna kako bi se prikupili uzorci za statističku analizu, a naposljetku će biti izvršena sama analiza.

Konačni cilj rada jest pokazati uspješnost različitih načina modeliranja uzoraka koji reprezentiraju procjene emocionalnih stanja govornika od strane različitih anotatora kroz testiranje samih modela.

2. Emocije

Ljudi nisu isključivo racionalna bića koja svoje postupke i htijenja poduzimaju samo na osnovu umnog ramišljanja i zaključivanja. Mnogi naši postupci izazvani su i drugim činiteljima, a jedan od tih činitelja su upravo emocije. Emocije ili osjećaji definiraju se kao doživljaj našeg subjektivnog odnosa prema stvarima, ljudima, događajima i samima sebi. Emocije su oduvijek bile zanimljiva tema širokog spektra znanstvenih disciplina. Proučavali su ih filozofi, biolozi, psiholozi. Svaka znanstvena grana pokušava što preciznije opisati tu izrazito složenu pojavu, iz njoj zanimljive perspektive promatranja.

Velik je broj različitih emocija koje se koriste, a neke od njih su sreća, tuga, ljubav, simpatija, ponos, oholost, sram, zahvalnost, briga, strepnja, nada itd. Emocije se na čovjeku prepoznaju kroz unutrašnje i vanjske indikatore. Unutrašnji su indikatori promjena rada unutarnjih organa, poput bržeg kucanja srca prilikom osjeta emocije straha. Vanjski indikatori su, primjerice, izrazi lica, tjelesna napetost, promjene u glasu. Promjena u glasu zbog trenutne emocije čini okosnicu ovog rada.

Uz emocije, postoje još neki slični procesi koji se od emocija razlikuju po duljini trajanja te uzroku nastajanja.

- emocija – intenzivna, relativno kratkog trajanja, usmjerena prema vanjskom objektu ili internim mislima
- raspoloženje – manje intenzivno od emocije, duljeg trajanja (moguće i danima), nije usmjereno prema konkretnom objektu
- sentiment – općeniti odnos prema nekom objektu
- crta ličnosti – dugoročna karakterna osobina

2.1. Vrste emocija

Prema podjeli emocija navedenoj u [1], emocije se dijele na dvije skupine:

- osnovne ili primarne – sreća, tuga, strah, ljutnja, iznenađenje, gađenje
- složene ili sekundarne – ljubav, ljubomora, zavist, ponos, krivnja...

Primarne su emocije u svim kulturama jednake. Sekundarnih emocija ima vrlo mnogo, a one su kombinacija dviju ili više primarnih emocija. Tako je, primjerice, ljubomora sastavljena od primarnih emocija straha, ljutnje i tuge.

Prema drugom kriteriju, emocije se mogu podijeliti na:

- pozitivne
- negativne

2.2. Rerezentacija emocija

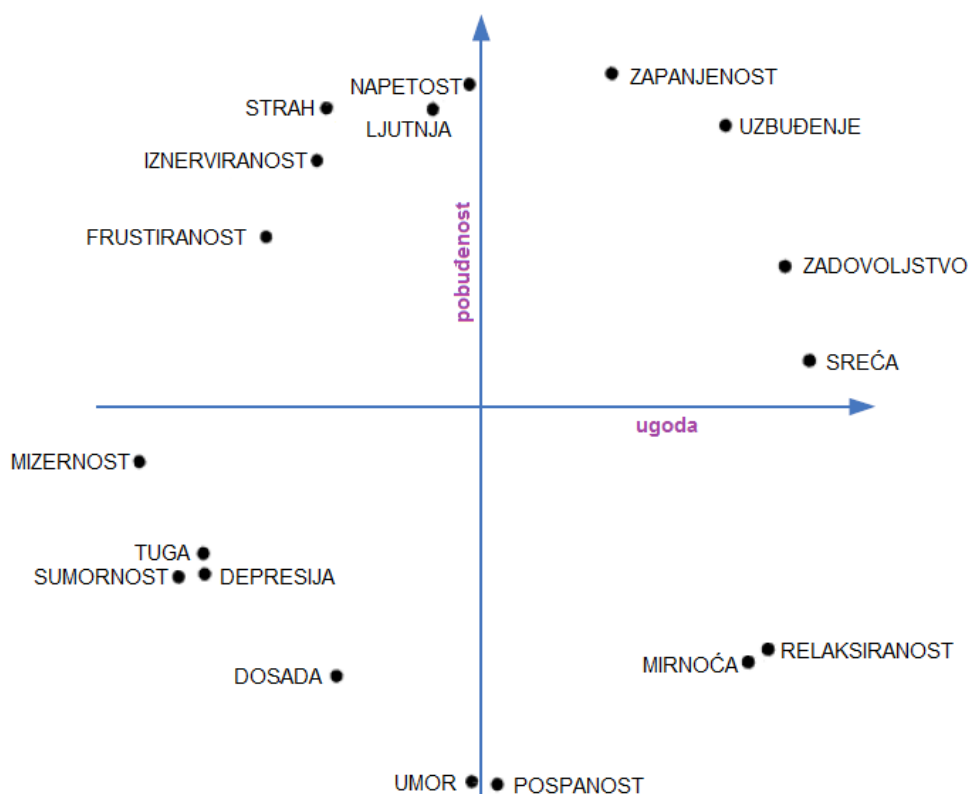
Znanstvenici su se složili oko dvije vrste vizualnog prikaza emocija. Prva pretpostavlja ograničen skup diskretnih emocija, dok drugi opisuje emocije kao kontinuiranu funkciju u dvije ili tri dimenzije. Diskretne emocije su upravo one primarne emocije navedene u poglavlju 2.1., iako oko broja diskretnih emocija postoji razilaženje u mišljenju unutar znanstvenih krugova.

2.2.1. Diskretni prikaz emocija

Emocija se u ovoj vrsti prikaza karakterizira kao jedna od diskretnih emocija, a dodatno se može navesti i intenzitet navedene emocije na nekoj zadanoj skali (primjerice, od 0 do 10).

2.2.2. Kontinuirani prikaz emocija

Ovakav prikaz podrazumijeva bliskost različitih emocija, zbog čega je emociju potrebno opisati kroz nekoliko različitih vrijednosti. James Russel 1970. predlaže dvije dimenzije za opis emocija: ugodu (eng. *valence*) i pobuđenost (eng. *arousal*). Prikaz emocija na predloženom dimenzionalnom prostoru može se vidjeti na Slici 1.



Slika 1 Russelov dimenzionalni model emocija [2]

Uočeno je da znanje vrijednosti ugone i pobuđenosti nije dovoljno za potpuno precizno razlikovanje svih diskretnih emocija, s obzirom da se neke emocije nalaze vrlo blizu jedna drugoj, iako se u stvarnom životu vrlo dobro razlikuju. Primjer takvog para emocija su strah i ljutnja. Zbog svega navedenoga, predloženo je uvođenje treće dimenzije koja bi se nazivala kontrola, a povezuje se s tendencijom neke akcije, poput sukobljavanja ili

bježanja. U okviru ovog diplomskog rada ova se dodatna dimenzija neće razmatrati.

3. Prikupljanje audio-snimaka

Kako bismo imali dovoljno veliki korpus hrvatskog emotivnog govora za provođenje statističke analize, bilo je potrebno prikupiti dovoljno audio-snimaka. Uvjeti za prikupljanje bili su:

- Snimka je na hrvatskom jeziku (može biti na narječju, ali nije poželjno)
- Snimka je dulja od 1 sekunde, a kraća od 20 sekundi
- Na snimci se čuje samo jedan govornik
- Poželjno je da snimkom dominira jedna emocija (sreća, tuga, strah, ljutnja, neutralno stanje)

Snimke je bilo potrebno pretvoriti u format *.wav* te u jednokanalni način (eng. *mono*). Za obradu snimaka korišten je program Audacity.

Korpus je prije početka prikupljanja u sklopu ovog diplomskog rada sadržavao 1037 snimaka. Te snimke prikupljene su u sklopu diplomskog rada Milosza Chmure [3] te seminarskih radova u sklopu predmeta Interaktivni simulacijski sustavi Emilije Damić i Tee Tomić [4] te Berislava Marszaleka. Kako je cilj bio da korpus sadrži podjednak broj snimaka za svaku od zadanih emocija te podjednak broj snimaka za svaki od spolova (muški i ženski), prvi zadatak u sklopu ovog diplomskog rada bio je nadopuniti korpus. U tu svrhu prikupljeno je dodatnih 109 snimki te ih sada ukupno ima 1146. Kako bi se vodila evidencija o prikupljenim snimkama, izrađena je tablica u kojoj se nalaze podaci o svakoj snimci. Dio tablice prikazan je u Tablici 1.

ID snimke	spol	dob	ID govornika	ekspr	transkripcija	inicijalna emocija
775	M	O	248	G	EJ EVO ZNAŠ ONU KLINKU KOJU TRAŽE SALE JU JE PUSTIO HTJELI SMO JU VRATIT ALI VODE	ST
776	M	O	248	G	OVO JE SLJEPA ULICA	ST
777	M	O	251	G	ZAŠTO TO RADITE	ST
778	Z	O	252	G	E A ONDA MOJ MUŽ ODE DOLE I IZVUČE OSIGURAČ IDH ZA CIJELU KUĆU	SR
779	Z	O	253	G	HA NE ZNAM OBIČNO OD MENE TRAŽE ONO ŠTO NE MOGU DOMA DOBIT	SR
780	Z	O	253	G	HIHI HIHI UDH TO SAM I MISLILA	SR
781	Z	O	252	G	OOO LJEPOTICA NAŠA IDE KAKO SMO EEE TAKO I TREBA TAKO I TREBA SRCE MOJE LJJEPO	SR
782	M	O	254	G	IDEM EVO ME EVO ME OOO GOSPA BARO OPET VI	SR
783	Z	O	253	G	E SUPER JA SAM ISTO UČILA PRIJE A ONDA SAM SI REKLA STARA MOJA NEMA TU PARA STA	SR
784	M	O	254	G	OOO SUSEDA SUSEDICA MOJA UDH ZNAL SAM JA DA JE PITANJE VURE KAD BUŠ DOŠLA K M	SR
785	M	O	254	G	OOO VELEČASNE OPET VI ZASTAJK SAMO NAPRED NIŠ SE NEBOJTE ZASTAJK AJDE AJDE IDI	SR
786	Z	O	253	G	MAMIN IMPOTENTNI KRETEN IDH HAHA HAHA UDH UDH	SR
787	M	O	255	G	ŽIVIO SUDAC ŽIVIO	SR
788	Z	O	256	G	HIHI MOCE HAHA UDH PA MOCE TO LUDONJO UPLAŠI ME UDH UDH HIHI UDH IDH	SR

Tablica 1 Pohranjeni podaci o snimkama

Podaci koji se pohranjuju za svaku snimku su:

- ID snimke
- Spol govornika
- Dob govornika
- ID govornika
- Ekspresija (G = glumljena, S = stvarna)
- Transkripcija
- Inicijalna emocija (emocija govornika procijenjena od strane prikupljača snimaka)

U Tablici 2 možemo vidjeti broj prikupljenih snimaka prema inicijalnoj emociji, spolu govornika te ekspresiji.

EMOCIJA	broj snimki	M			Z		
		G	S	ukupno	G	S	ukupno
SR	250	113	24	137	56	57	113
LJ	287	89	68	157	53	77	130
NE	201	21	107	128	16	57	73
TU	207	46	24	70	68	69	137
ST	200	114	0	114	73	13	86

SPOL	broj snimki	EKSPRESIJA	broj snimki
M	607	G	650
Z	539	S	496

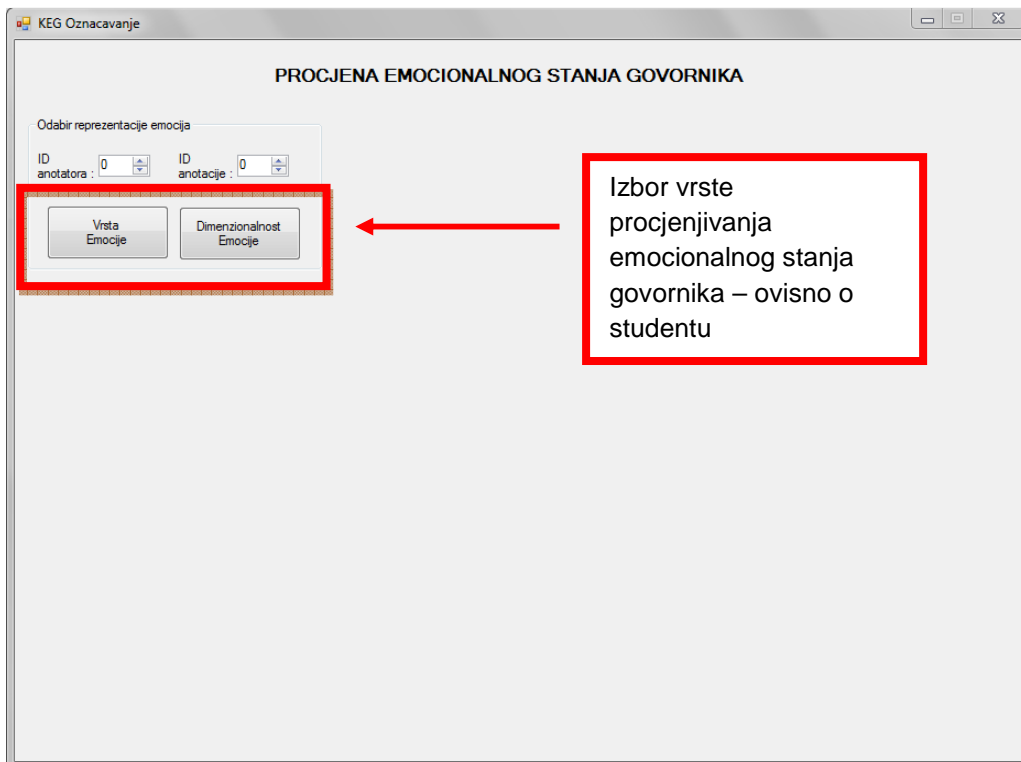
Tablica 2 Broj prikupljenih snimaka

4. Eksperiment

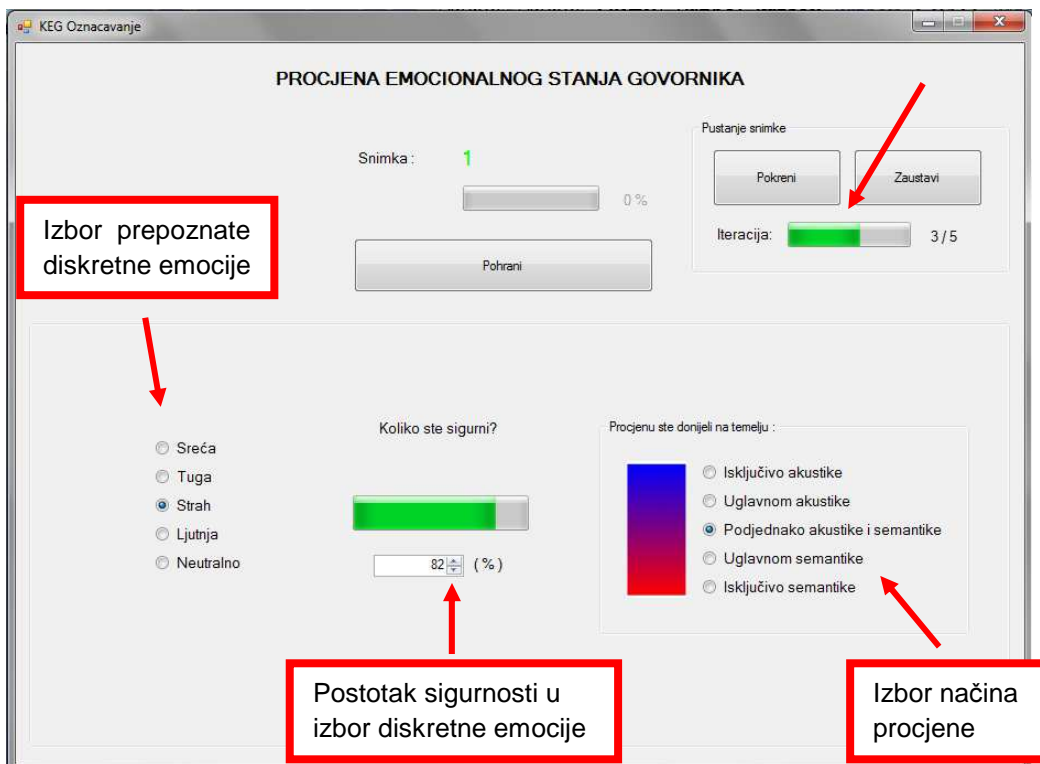
Kako bismo prikupili dovoljno anotacija snimki, u suradnji s Odsjekom za psihologiju pri Filozofskom fakultetu u prostoru Fakulteta elektrotehnike i računarstva, organizirali smo eksperiment. Eksperimentu se pridružilo 115 različitih anotatora, studenata na Odsjecima za psihologiju te fonetiku Filozofskog fakulteta, a cjelokupni je eksperiment trajao 3 tjedna. Svaki je student u tri dolaska u trajanju od jednog sata imao zadatak procijeniti emocionalno stanje govornika u ukupno 228 snimaka (svaki dolazak bio je predviđen za ocjenjivanje njih 76). Nakon ocijenjenih 46 snimaka, a prije preostalih 30, studenti su imali pauzu od nekoliko minuta kako umor ne bi utjecao na rezultate istraživanja. Neki studenti bili su zaduženi za procjenjivanje emocionalnog stanja govornika diskretnom emocijom, a neki su pak morali procijeniti razinu ugone i pobuđenosti emocije, dakle dimenzionalnu emociju. Za vrijeme pet godina studija, ovi studenti trebaju prikupiti 50 eksperimentalnih sati. Sudjelovanjem u eksperimentu za svaki su dolazak dobili po dva eksperimentalna sata.

U svrhu automatizacije preslušavanja snimaka te pohranjivanja procjena studenata, izrađena je računalna aplikacija u programskom jeziku C#. Pri procjeni diskretnih emocija, studenti su uz procjenu konkretnog emocionalnog stanja trebali ocijeniti i razinu svoje sigurnosti u procjenu, kao i označiti je li pri njihovoj procjeni prevagnula akustika ili semantika snimke. Kako mnogo snimki ipak uz jednu prevladavajuću emociju sadrži i dodatne emocije, studenti su imali priliku navesti intenzitete svih emocija koje prepoznaju u snimci. Studenti koji su procjenjivali dimenzionalne emocije, utvrđivali su razinu ugone i pobuđenosti govornika, vlastitu sigurnost u izbor navedenih razina te su također naznačili na temelju čega su donijeli zaključke: je li prevagnula akustika ili semantika snimke.

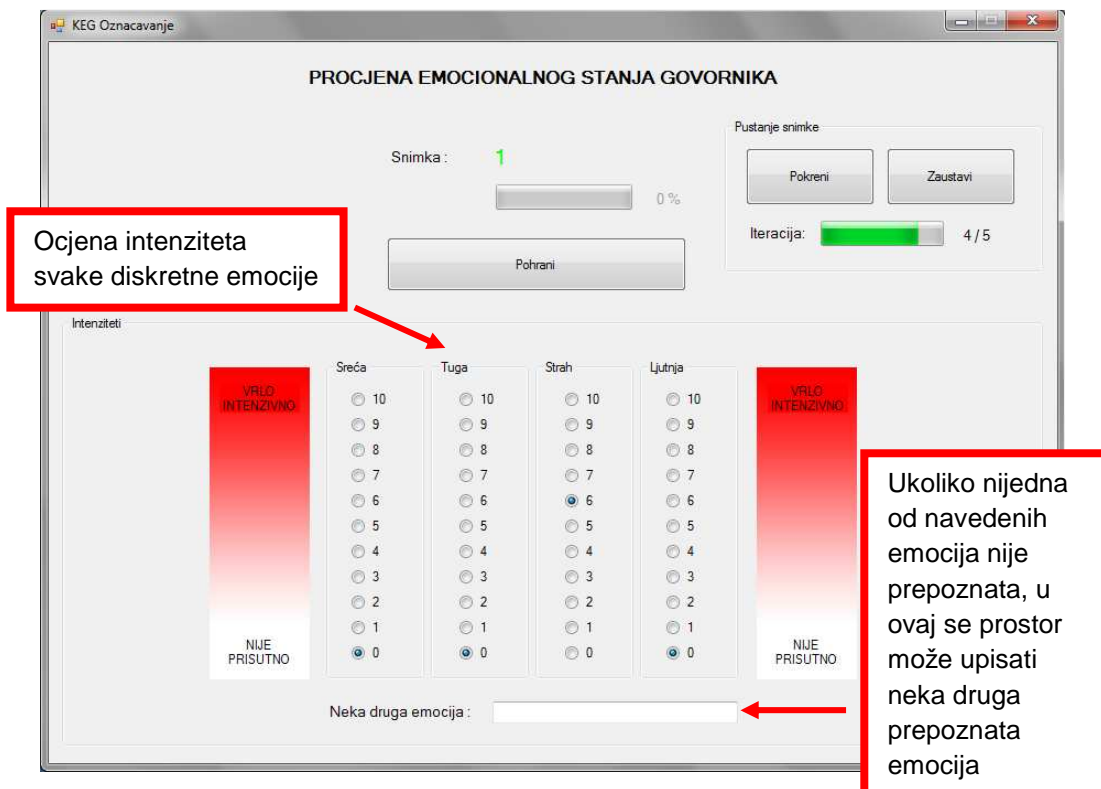
Izgled i opis pojedinog dijela aplikacije možemo vidjeti na sljedećim slikama.



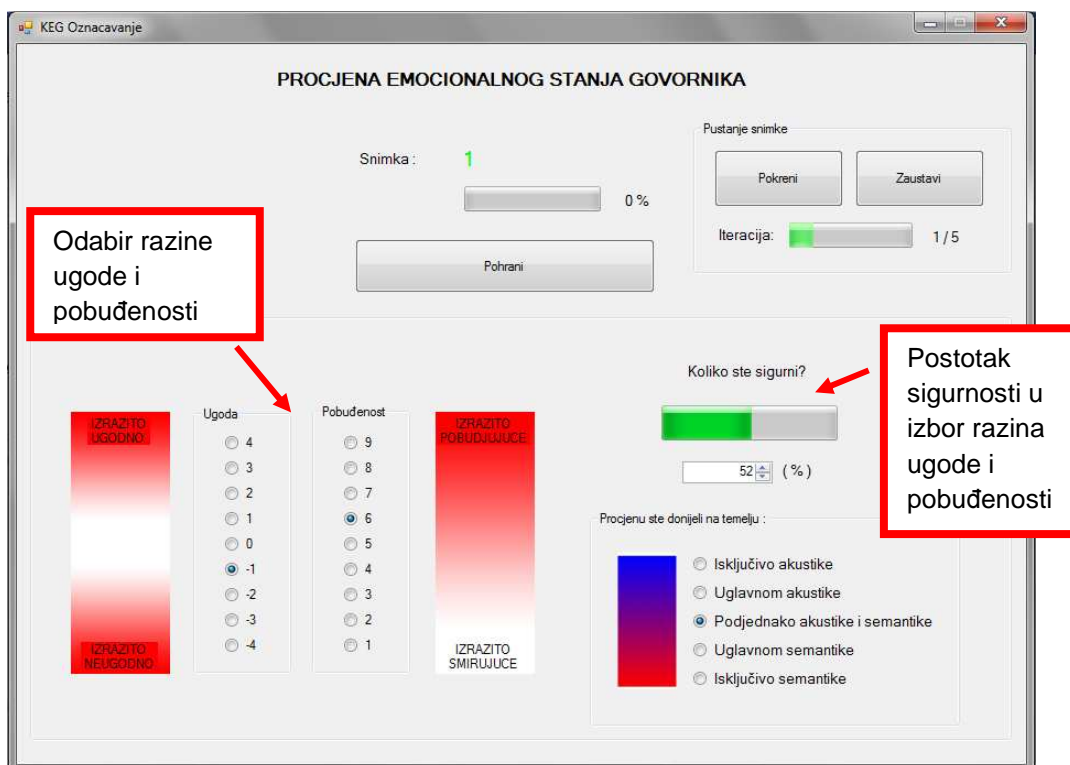
Slika 2 Početni prozor aplikacije



Slika 3 Procjena emocionalnog stanja govornika diskretnim emocijama - prozor 1



Slika 4 Procjena emocionalnog stanja govornika diskretnim emocijama - prozor 2



Slika 5 Procjena emocionalnog stanja govornika dimenzionalnim emocijama

5. Modeliranje emocija Gaussovima modelima

Za potrebe ovog diplomskog rada korišteni su podaci iz starog korpusa koji sadrži 674 snimke ocijenjene od strane pet anotatora. Svaki anotator ocjenjivao je i dimenzionalne i diskretne emocije za svaku snimku koju je preslušao.

Tablica u kojoj su pohranjeni podaci o procjenama emocionalnog stanja govornika prikazana je u Tablici 3.

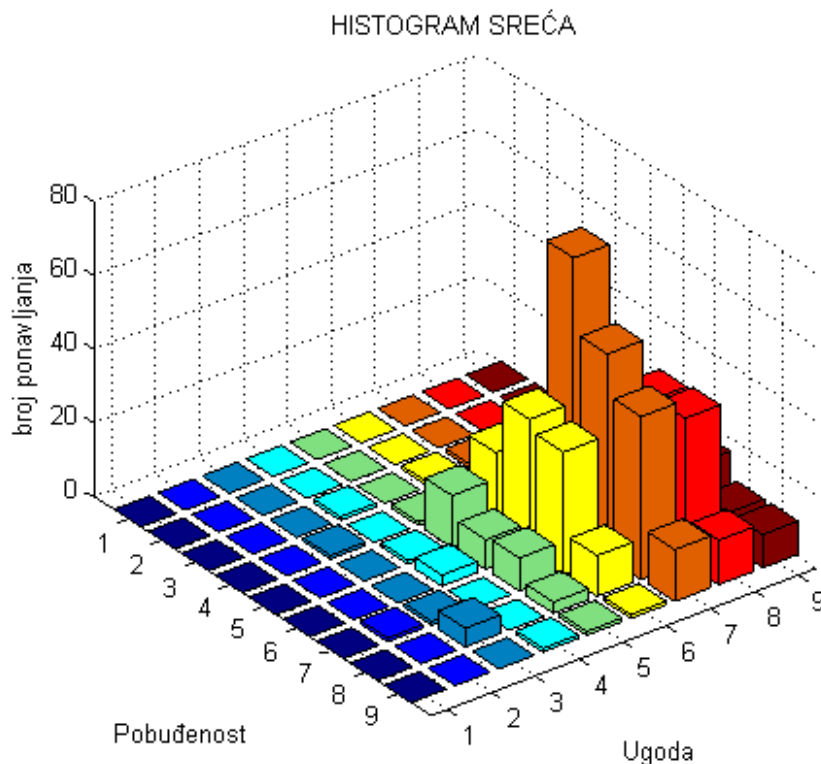
r1 - SR	r1 - TU	r1 - ST	r1 - U	r1 - V	r1 - A
0	0	0	8	3	8
0	0	0	8	2	8
0	0	0	8	2	9
0	0	0	8	2	9
0	0	0	7	3	8
0	0	0	8	3	8
0	0	0	7	3	8
0	0	0	8	2	8

Tablica 3 Ocjene jednog anotatora

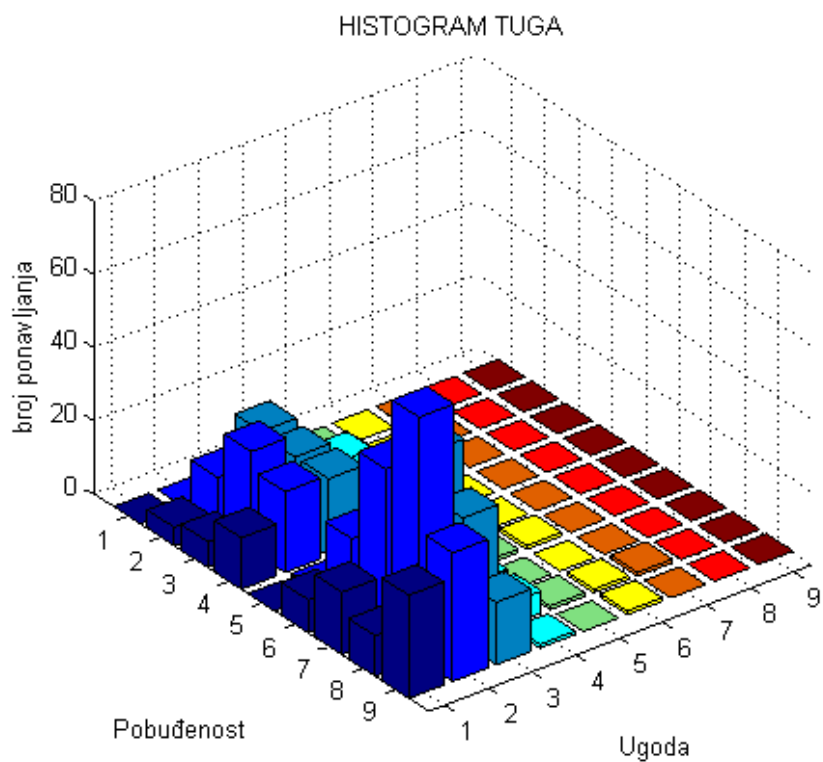
U gornjoj tablici anotator je za prvu snimku intenzitete svake emocije ocijenio s 0, osim intenziteta emocije ljutnja kojemu je pridodao vrijednost 8 na skali od 0 do 10. Dimenzionalnu emociju prve snimke ocijenio je davši vrijednosti 3 za ugodu, odnosno 8 za pobuđenost govornikove emocije.

5.1. Modeliranje histograma za svaku emociju

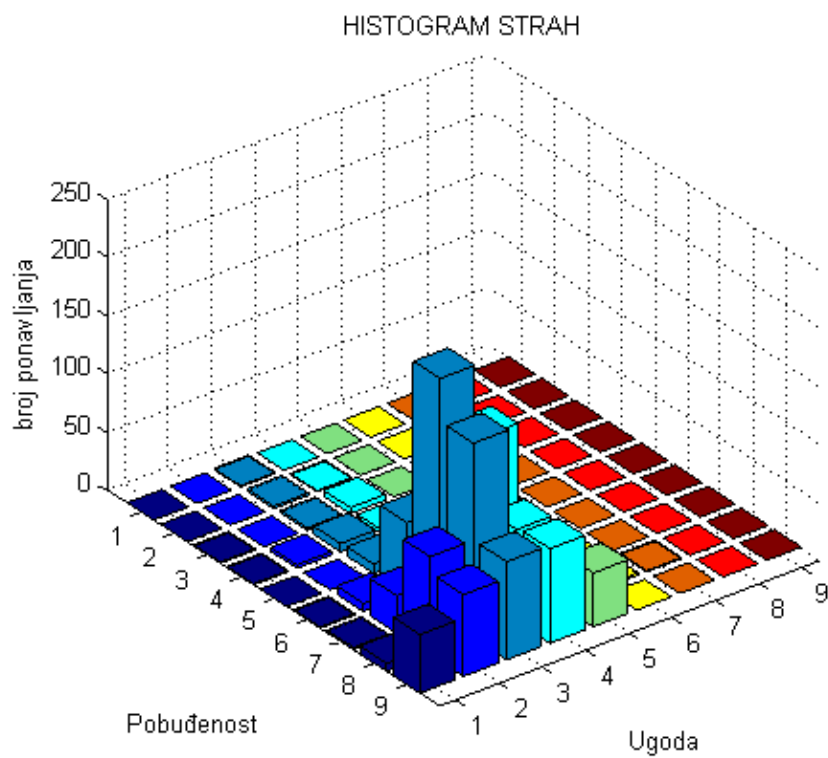
Histogram je prikaz broja ponavljanja neke emocije u ovisnosti o ugodi i pobuđenosti. Ukoliko je anotator za neku snimku označio sve emocije intenzitetom manjim od 3, smatra se da je emocija govornika s te snimke za tog anotatora neutralna te se histogram neutralnog stanja povećava za 1 za varijable ugode i pobuđenosti koje su označene za istu tu snimku. Ukoliko je intenzitet emocije ljutnja za neku snimku veći od 3, histogram emocije ljutnja povećava se za jedan za varijable ugode i pobuđenosti koje odgovaraju istoj snimci. Intenziteti od 4 do 10 uključivo tretiraju se jednako, dakle intenzitet ne predstavlja zasebnu varijablu. Vratimo li se ponovno na Tablicu 3 i proanaliziramo li prvu navedenu snimku, prema navedenim pravilima o gradnji histograma, histogram emocije ljutnja će se povećati za 1 za varijablu ugode 3 i pobuđenosti 8. Isti postupak ponavlja se kroz sve 674 snimke. Rezultat su histogrami za sve četiri emocije te neutralno stanje. Histogrami su prikazani na sljedećim slikama.



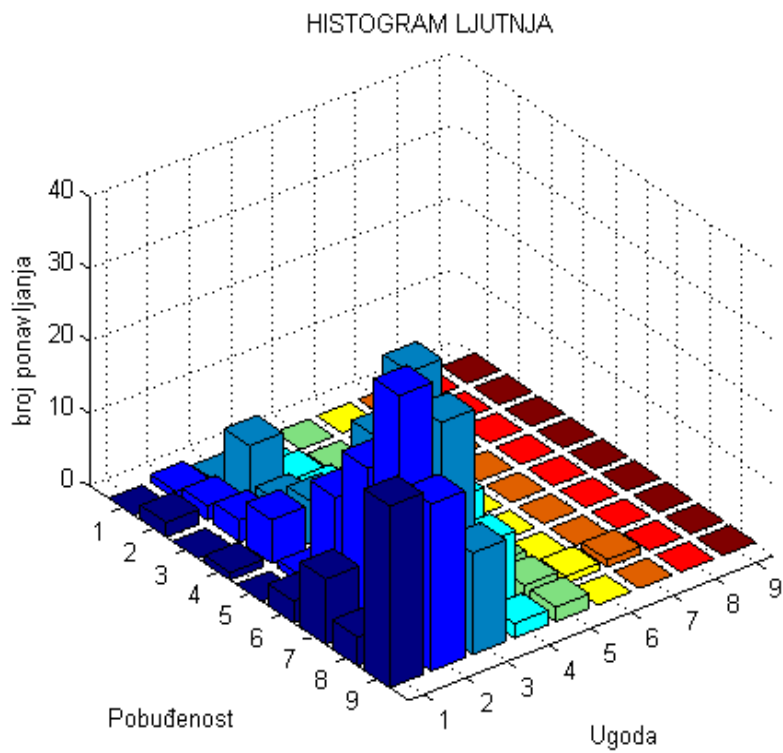
Slika 6 Histogram emocije sreća



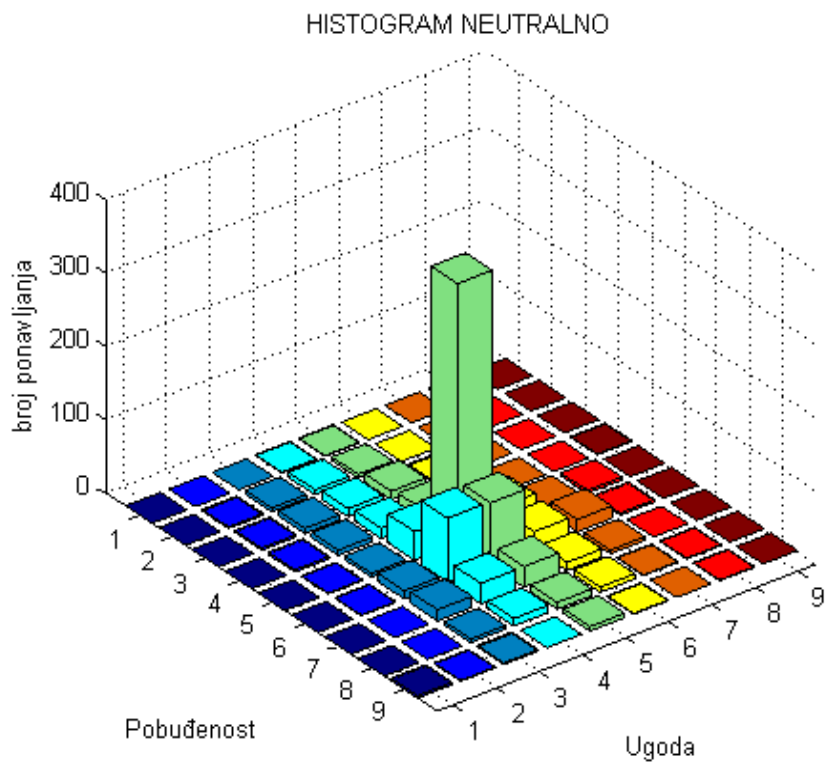
Slika 7 Histogram emocije *tuga*



Slika 8 Histogram emocije *strah*



Slika 9 Histogram emocije *ljutnja*



Slika 10 Histogram *neutralnog stanja*

5.2. Aproximacija histograma Gaussovom funkcijom dviju varijabli

Cilj je ovog dijela rada aproksimirati svih pet ranije prikazanih histograma Gaussovima funkcijama gustoće vjerojatnosti. Za aproksimaciju je korištena gotova funkcija [8]. Funkcija je obavila aproksimaciju histograma Gaussovima plaštom, a kao dodatak učinjeno je podešavanje plašta kako bi integral dobivene funkcije, tj. volumen prostora ispod grafa bio jednak 1.

Funkcija gustoće vjerojatnosti normalne distribucije s dvije varijable:

$$PDF = \frac{1}{2 \cdot \pi \cdot \sqrt{\det(C)}} \cdot e^{-\frac{1}{2} \cdot (x-\mu)^T \cdot C^{-1} \cdot (x-\mu)}$$

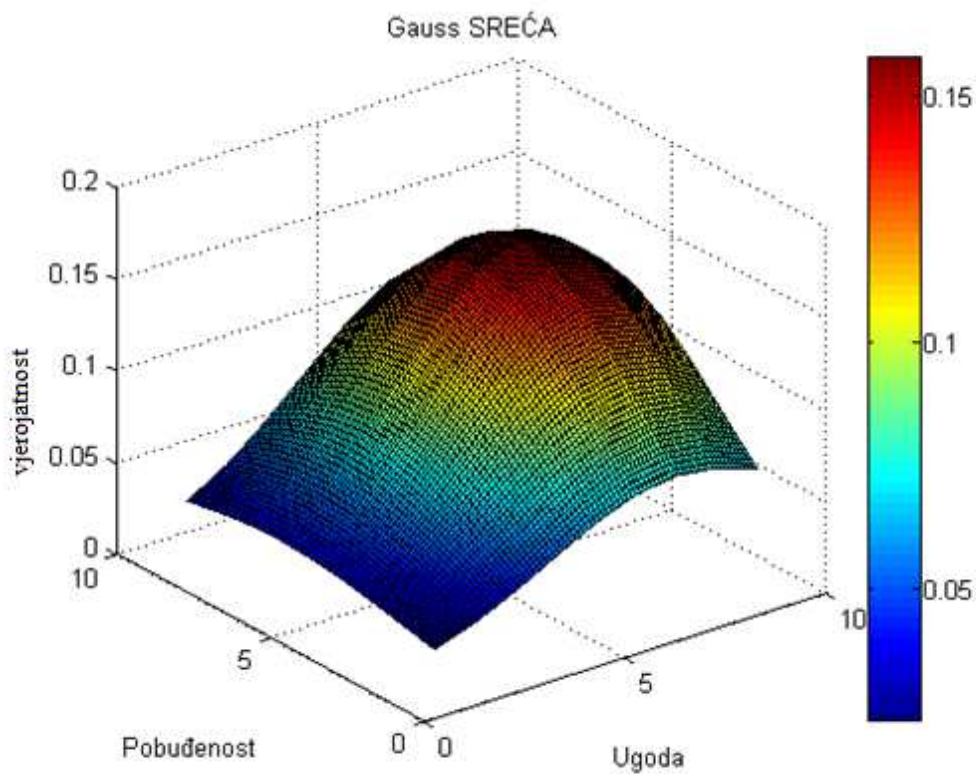
gdje su:

$$\mathbf{x} = \begin{bmatrix} V \\ A \end{bmatrix}, \quad V=\text{valence (ugoda)}, A=\text{arousal (pobuđenost)}$$

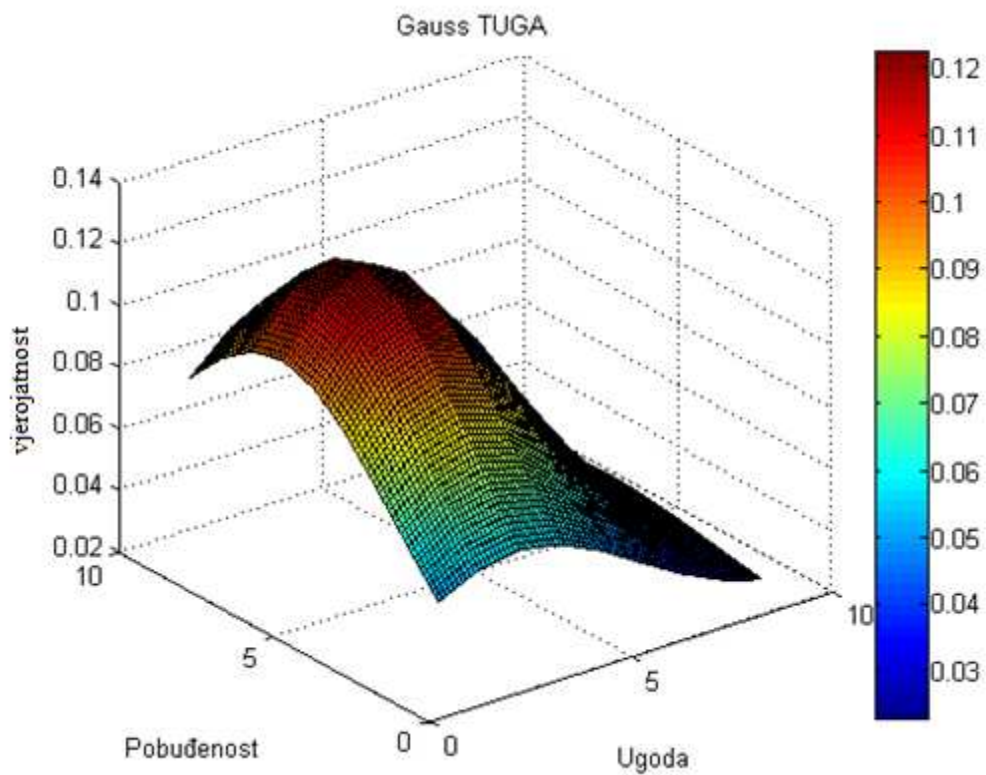
$$\boldsymbol{\mu} = \begin{bmatrix} \mu_V \\ \mu_A \end{bmatrix}, \quad \text{matrica sa srednjim vrijednostima po obje osi}$$

$$C = \begin{bmatrix} \text{var}(V, V) & \text{cov}(V, A) \\ \text{cov}(A, V) & \text{var}(A, A) \end{bmatrix}, \quad \text{matrica kovarijance}$$

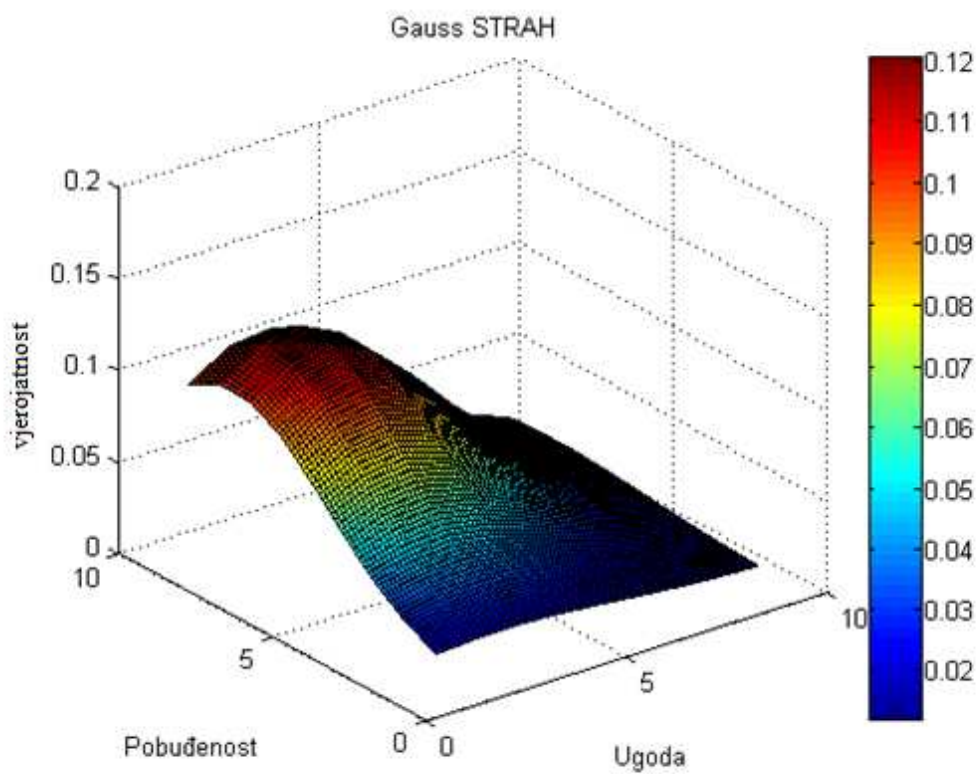
Na sljedećim slikama možemo vidjeti Gaussove modele dobivene aproksimacijom histograma za pojedinu emociju te neutralno stanje. Prilikom prikaza, početni je koordinatni sustav po x i y-osima interpolacijom namješten na razlučivost 0.1 kako bi funkcija bila što glađa.



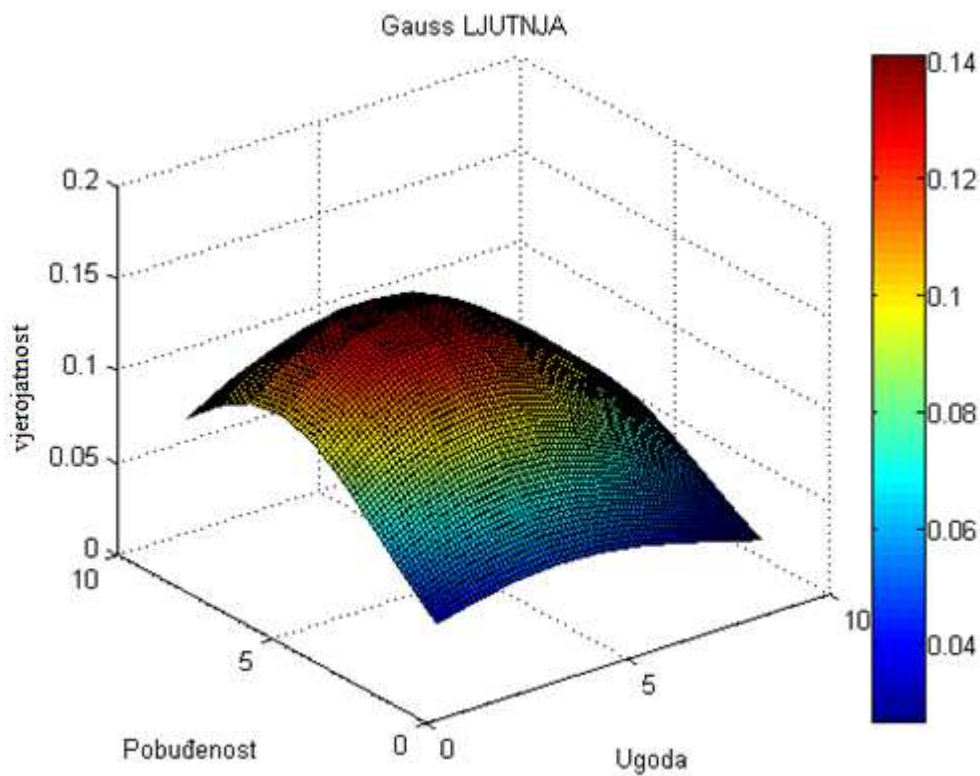
Slika 11 Gaussov model za emociju *sreća*



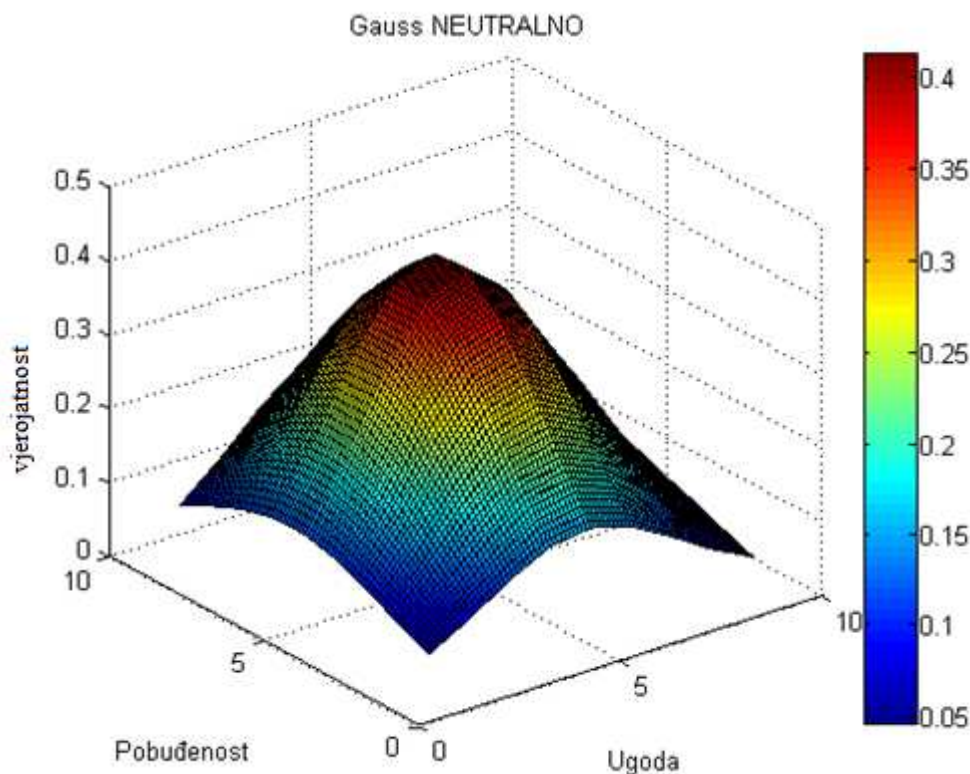
Slika 12 Gaussov model za emociju *tuga*



Slika 13 Gaussov model za emociju *strah*



Slika 14 Gaussov model za emociju *ljutnja*



Slika 15 Gaussov model za *neutralno stanje*

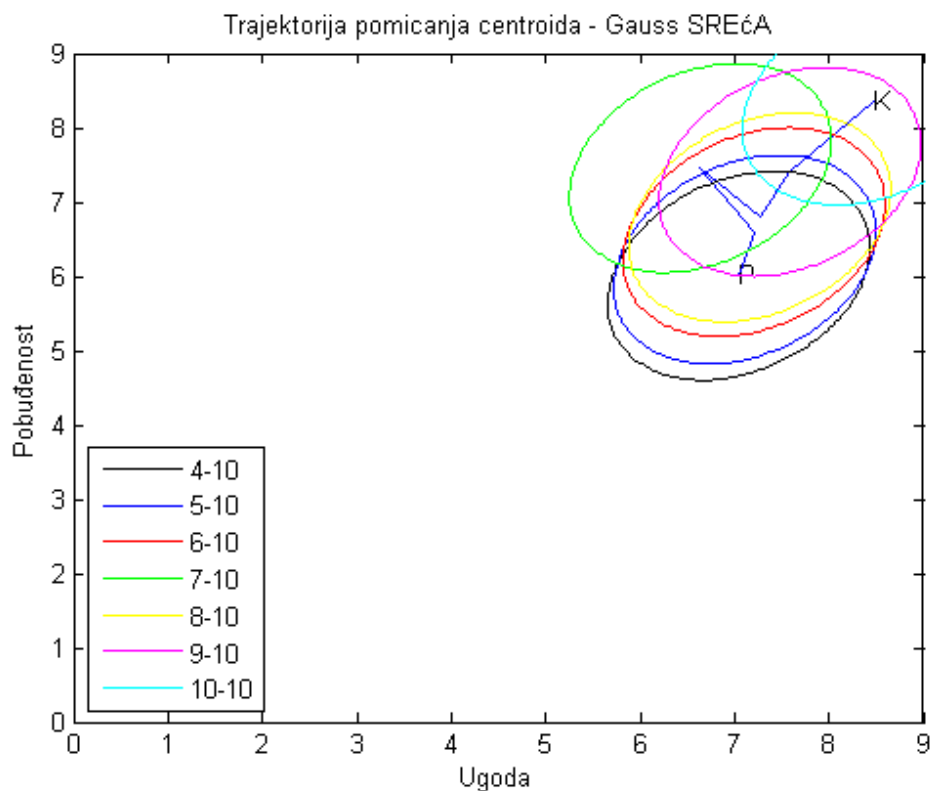
5.3. Trajektorije koje opisuju pomak Gaussovih modela

Svaki Gaussov model možemo predočiti u dvodimenzionalnoj ravni (na osima su ugoda, odnosno pobuđenost) pomoću elipsi koje su zapravo presjek Gaussovog modela na mjestu koje volumen prostora ispod Gaussovog plašta siječe na polovicu. Ukoliko bismo se vratili u postupak gradnje histograma opisan u Poglavlju 5.1. te postupak gradnje Gaussovih modela opisan u Poglavlju 5.2., možemo postupke ponoviti za različite raspone intenziteta emocija. U navedenim poglavljima za gradnju histograma i modela četiriju emocija sreće, tuge, straha i ljutnje koristio se raspon intenziteta tih emocija od 4 do 10. Neutralno nam stanje u ovom dijelu rada nije posebice zanimljivo jer za njega koncept intenziteta nema bitnu ulogu (sjetimo se, neutralnim je stanjem proglašena svaka snimka kojoj niti jedan intenzitet četiriju emocija ne prelazi razinu 3). Ovo poglavlje pruža nam uvid u

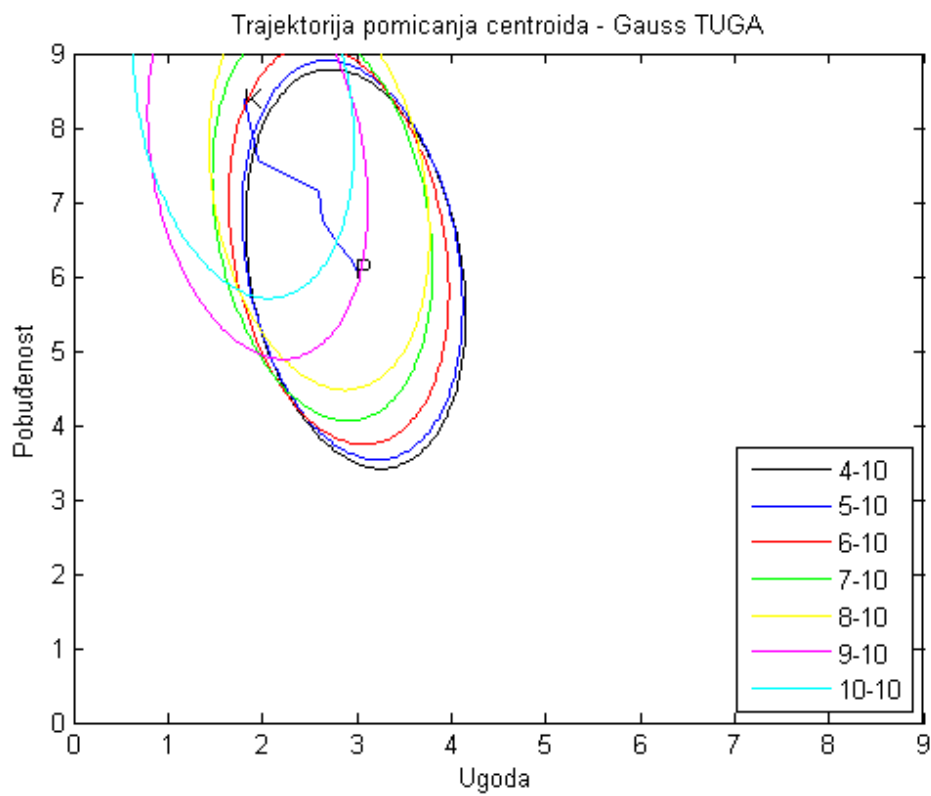
kretanje ranije spomenutih Gaussovih elipsi s obzirom na zadani raspon intenziteta. Korišteni rasponi intenziteta za svaku emociju su:

- 4–10
- 5–10
- 6–10
- 7–10
- 8–10
- 9–10
- 10–10

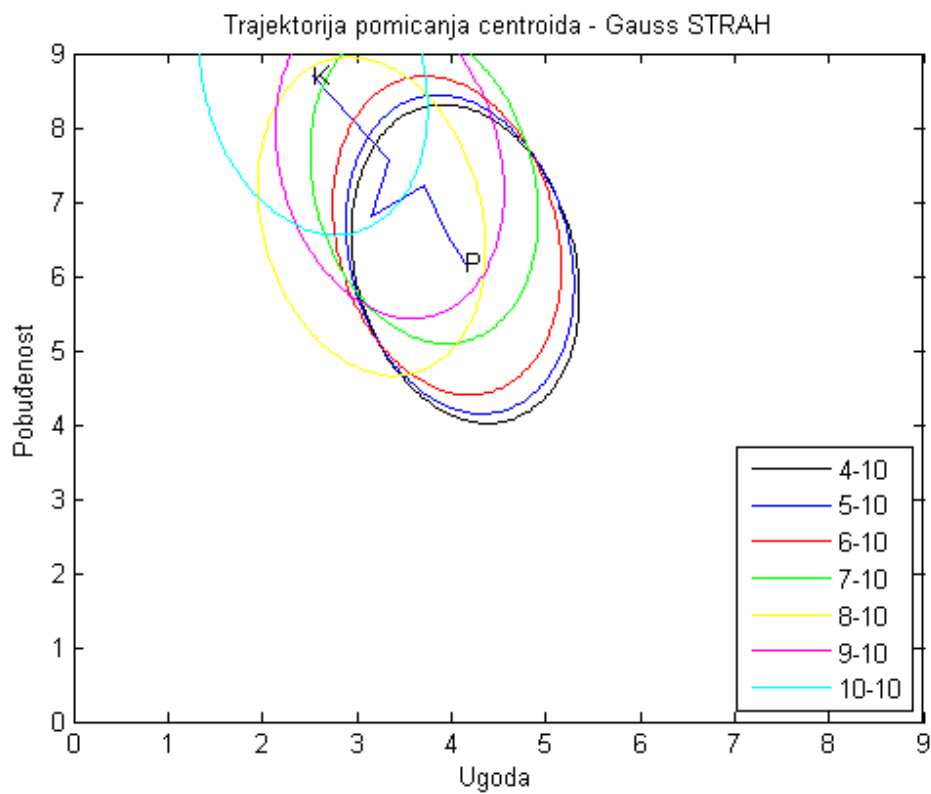
Svaka elipsa ima svoj centroid koji je zadan srednjim vrijednostima varijabli ugoda i pobuđenost za odgovarajući Gaussov model. Promotrimo li kretanje centroida te spojimo li centroide linijama, dobit ćemo trajektoriju pomicanja Gaussovih elipsi za svaku emociju zasebno. Elipse i trajektorije prikazane su na sljedećim slikama, a dobivene su pomoću funkcije [9]. Slovo „P“ označava centroid početne elipse, a slovo „K“ krajnje.



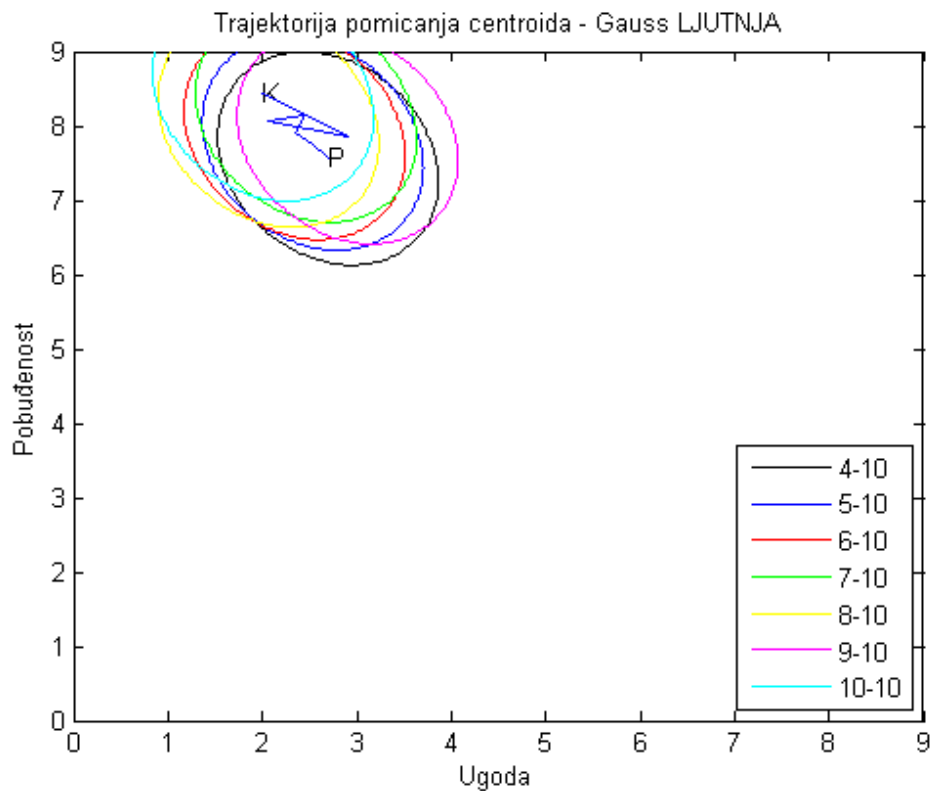
Slika 16 Trajektorija pomicanja centroida elipsi presjeka Gaussovog modela emocije sreća



Slika 17 Trajektorija pomicanja centroida elipsi presjeka Gaussovog modela emocije tuga



Slika 18 Trajektorija pomicanja centroida elipsi presjeka Gaussovog modela emocije strah



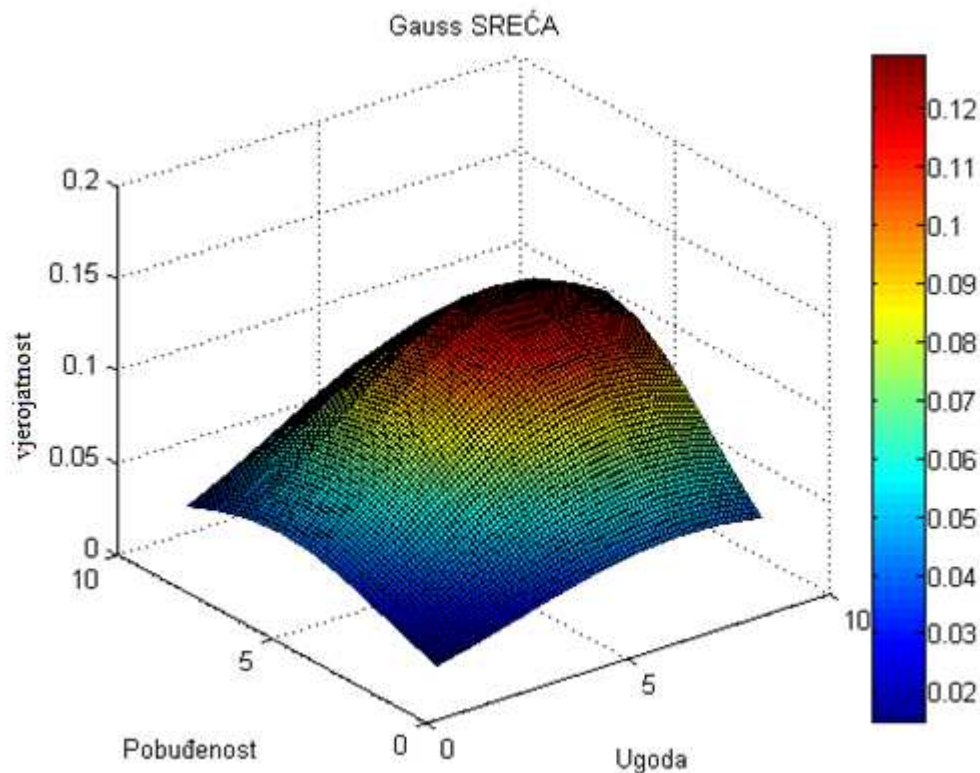
Slika 19 Trajektorija pomicanja centroida elipsi presjeka Gaussovog modela emocije ljutnja

Možemo zamijetiti da se elipse modela emocije sreća s promjenom raspona od intenziteta 4 do intenziteta 10 prema rasponu od intenziteta 10 do intenziteta 10, kako je i navedeno ranije u tekstu, pomiču prema višoj razini ugode te višoj razini pobuđenosti, što je logično jer intenziteti koji se promatraju rastu, a veći intenzitet sreće po definiciji znači i višu razinu ugode te višu razinu pobuđenosti. Elipse modela emocije tuga pomiču se prema nižoj razini ugode i višoj razini pobuđenosti, što vrijedi i za elipse modela emocije strah. Naposljetku, elipse modela emocije ljutnja također se pomiču prema nižoj razini ugode i višoj razini pobuđenosti. Ove pojave su očekivane, što možemo ustvrditi i vratimo li se još jednom na Sliku 1 gdje, prema definiciji, vidimo lokaciju za svaku od navedenih emocija u koordinatnom sustavu na čijim su osima varijable ugode i pobuđenosti.

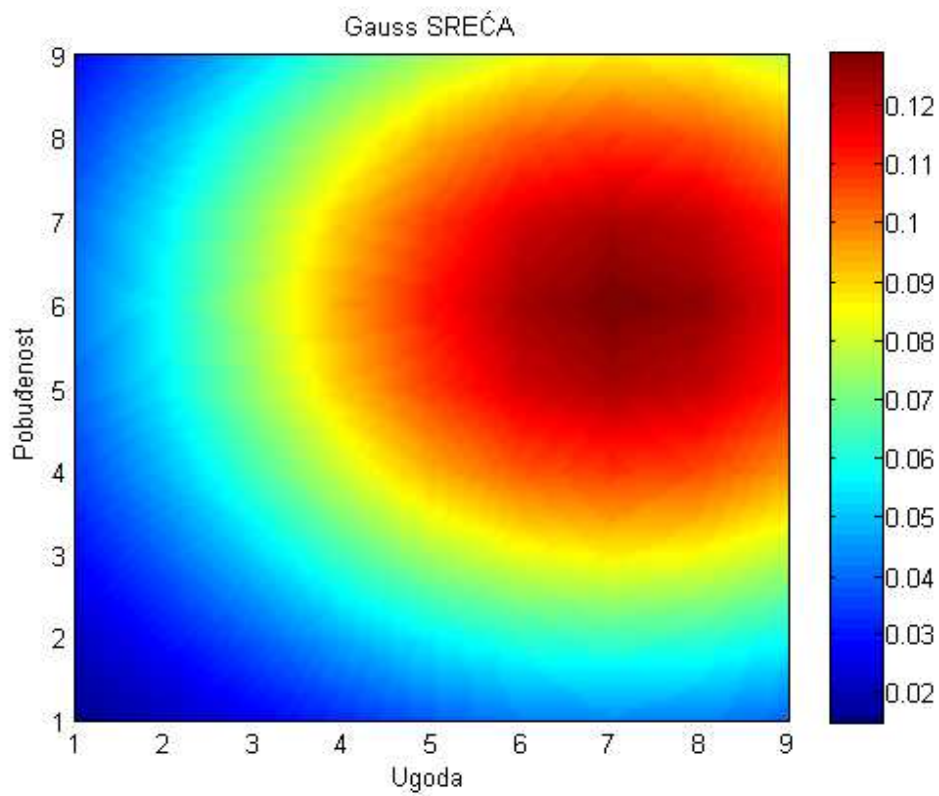
5.4. Treniranje Gaussovih modela

U Poglavlju 5.2. opisan je postupak izgrađivanja modela na temelju cijelog korpusa koji ukupno broji 674 snimke. Ono što bismo mi željeli je izgraditi modele svake emocije te neutralnog stanja na temelju 90% korpusa, a na preostalih 10% korpusa testirati dobivene modele te procijeniti njihovu vjerodostojnost.

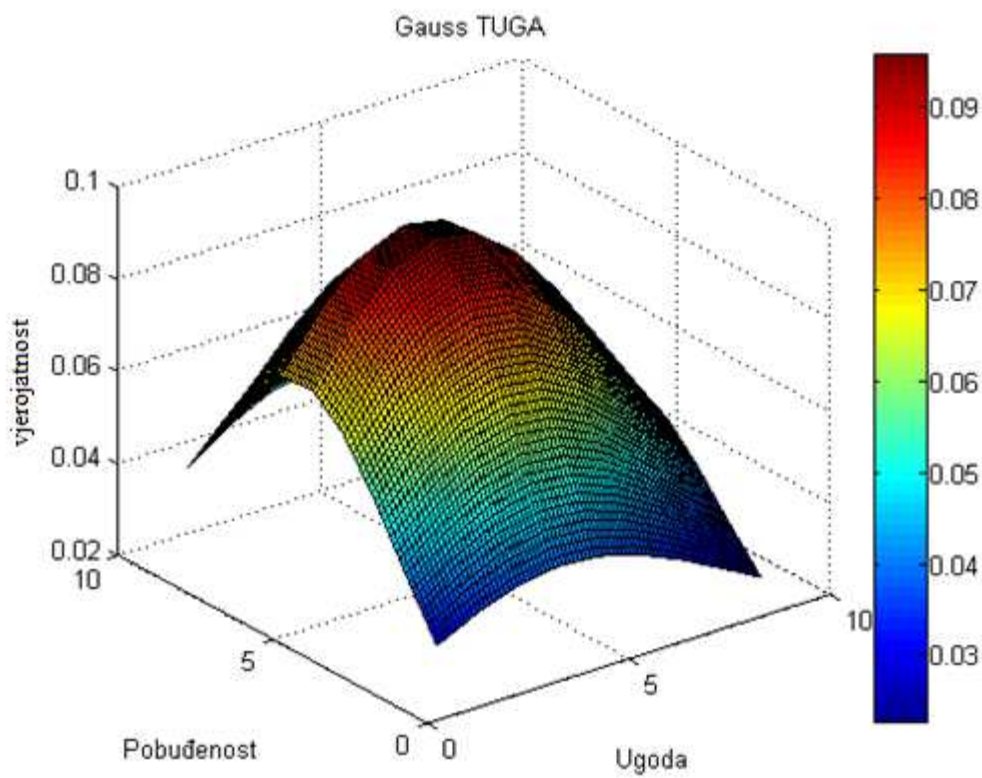
Postupak treniranja modela identičan je postupku iz Poglavlja 5.3., osim postavljenog uvjeta u petlji koja učitava podatke o snimkama koji propušta u obradu one snimke čiji redni broj nije djeljiv s 10. Razlog ovome uvjetu jest činjenica da za treniranje želimo koristiti samo 90% korpusa. Dvodimenzionalni i trodimenzionalni prikazi dobivenih Gaussovih modela vidljivi su na sljedećim slikama, a modeli su interpolacijom na razlučivost 0.1 prikazani preciznije.



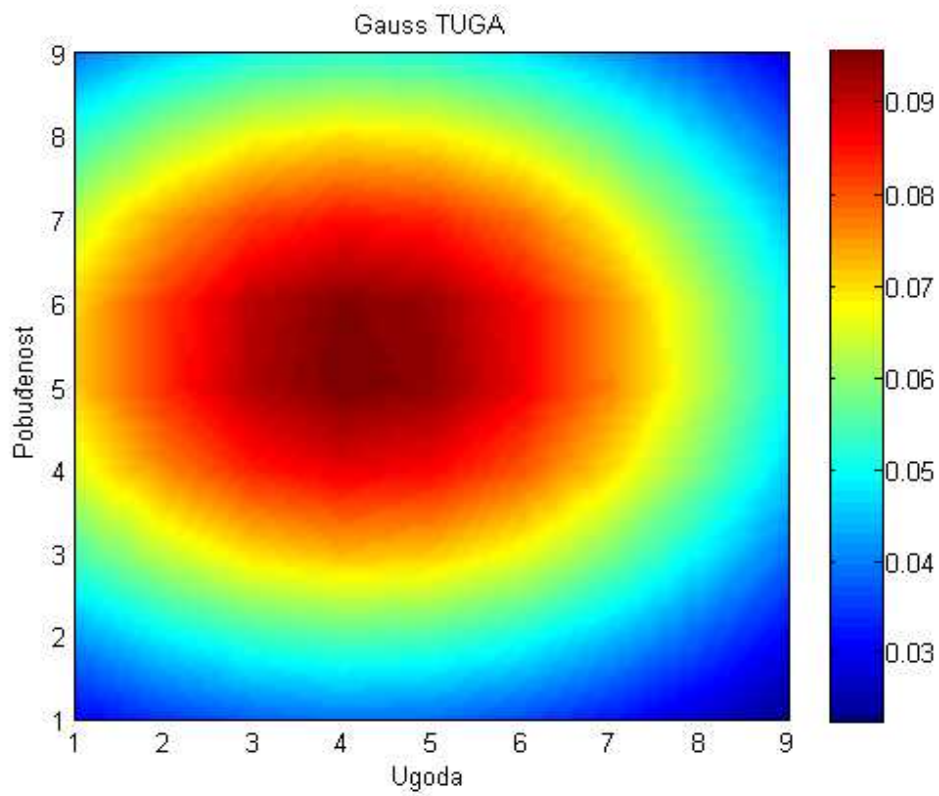
Slika 20 Gaussov model za emociju sreća



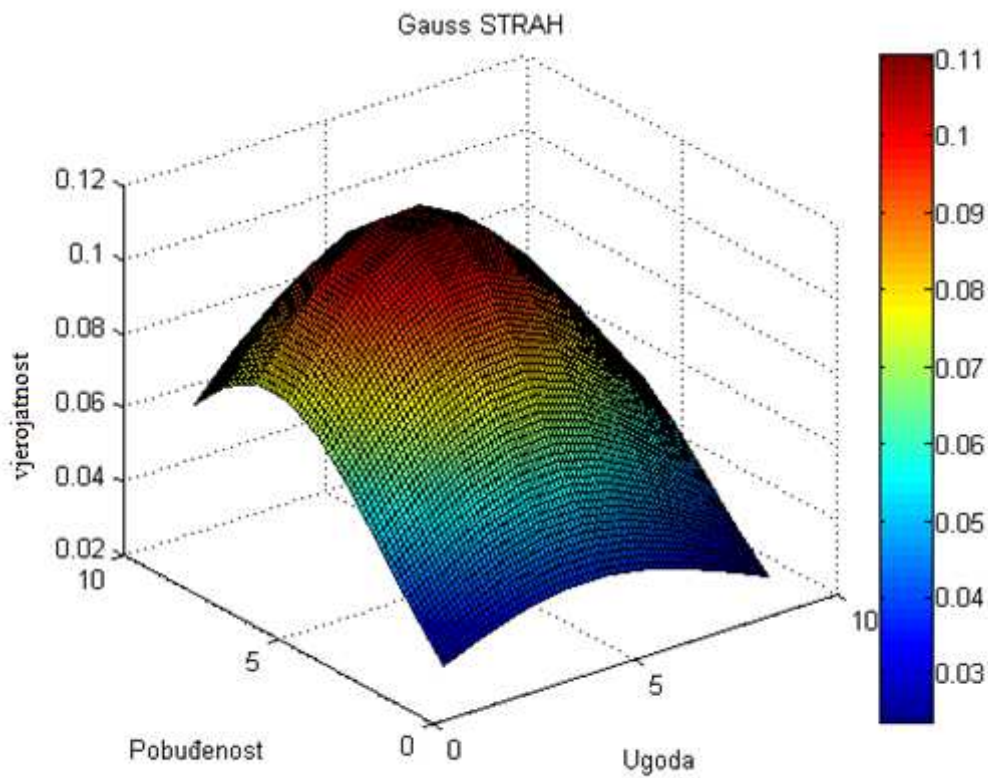
Slika 21 2D prikaz Gaussovog modela za emociju sreća



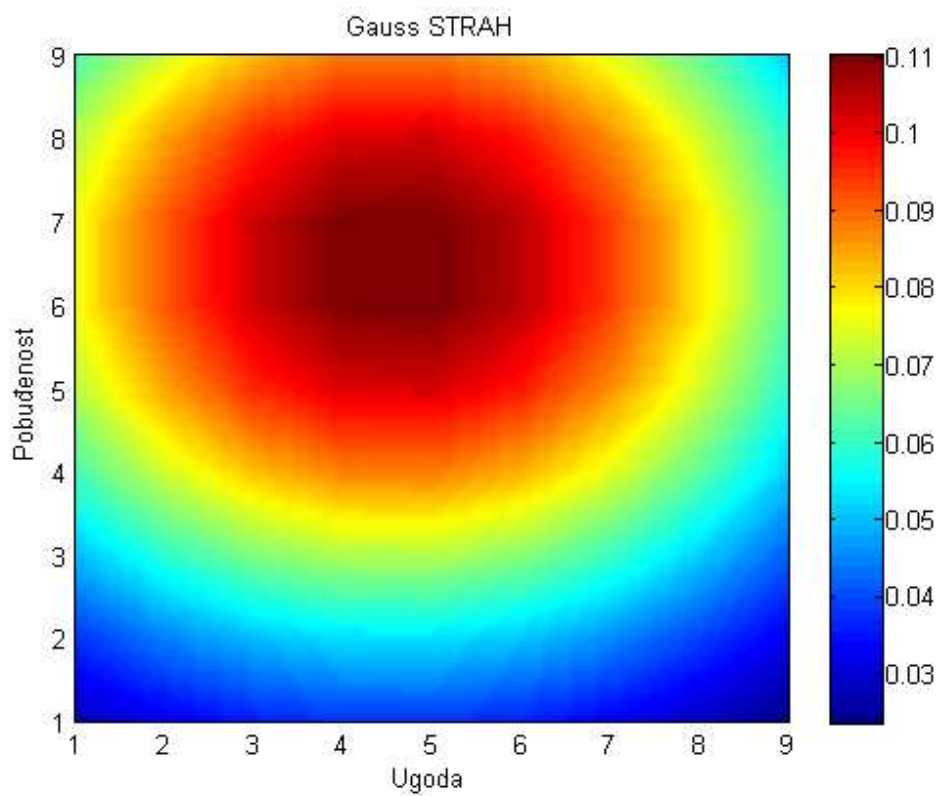
Slika 22 Gaussov model za emociju tuga



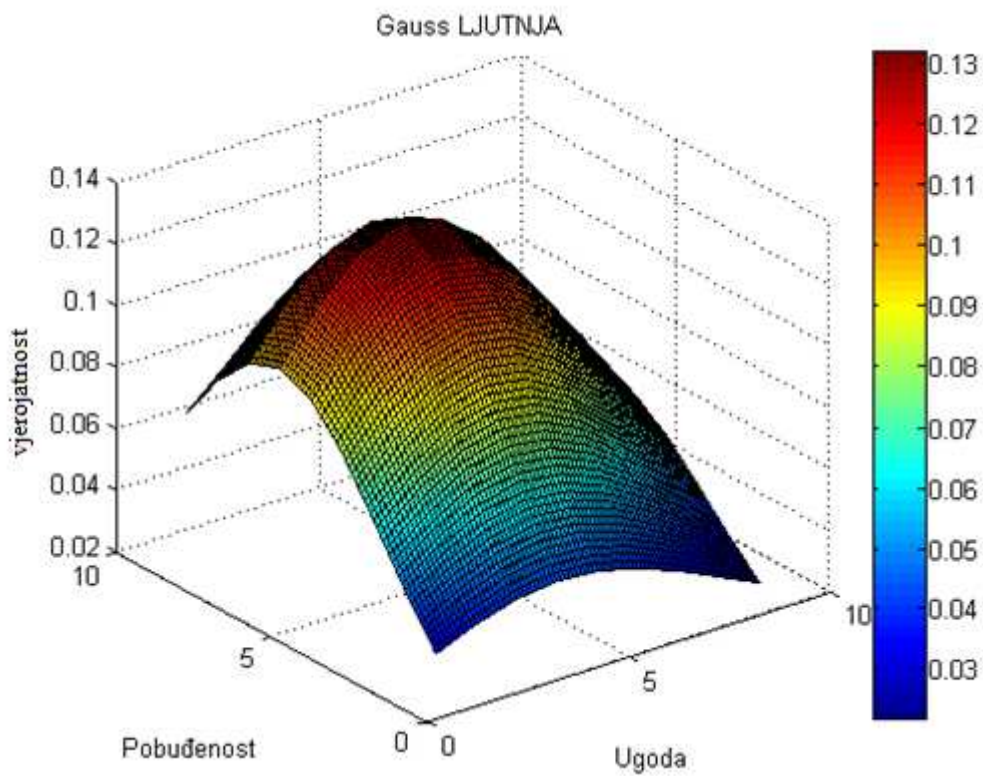
Slika 23 2D prikaz Gaussovog modela za emociju tuga



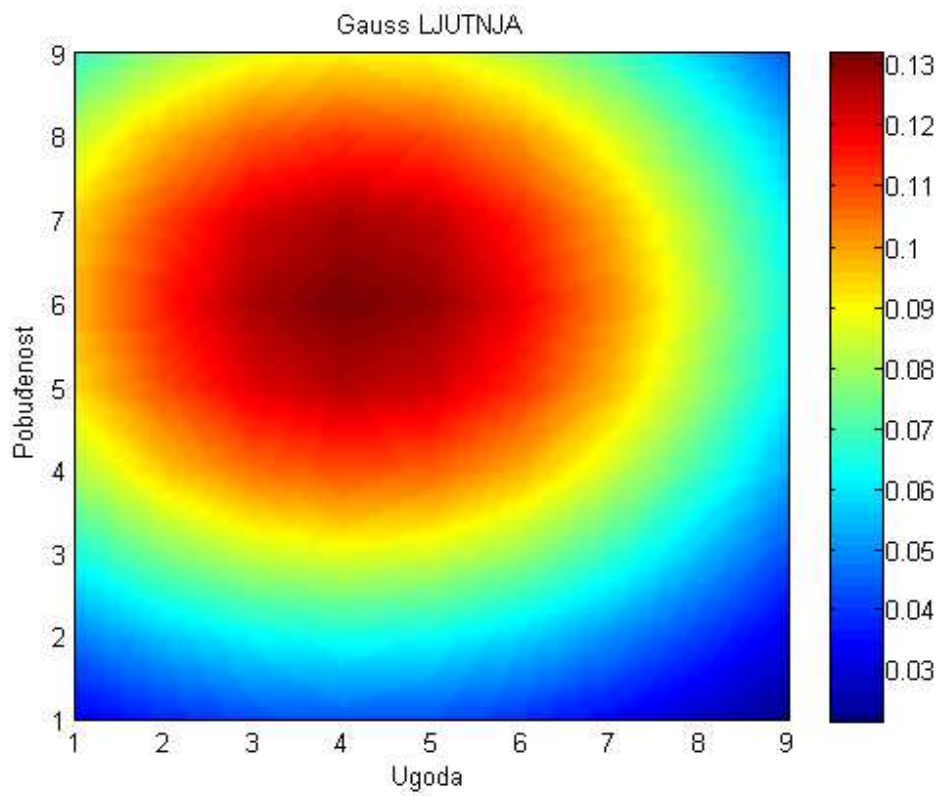
Slika 24 Gaussov model za emociju strah



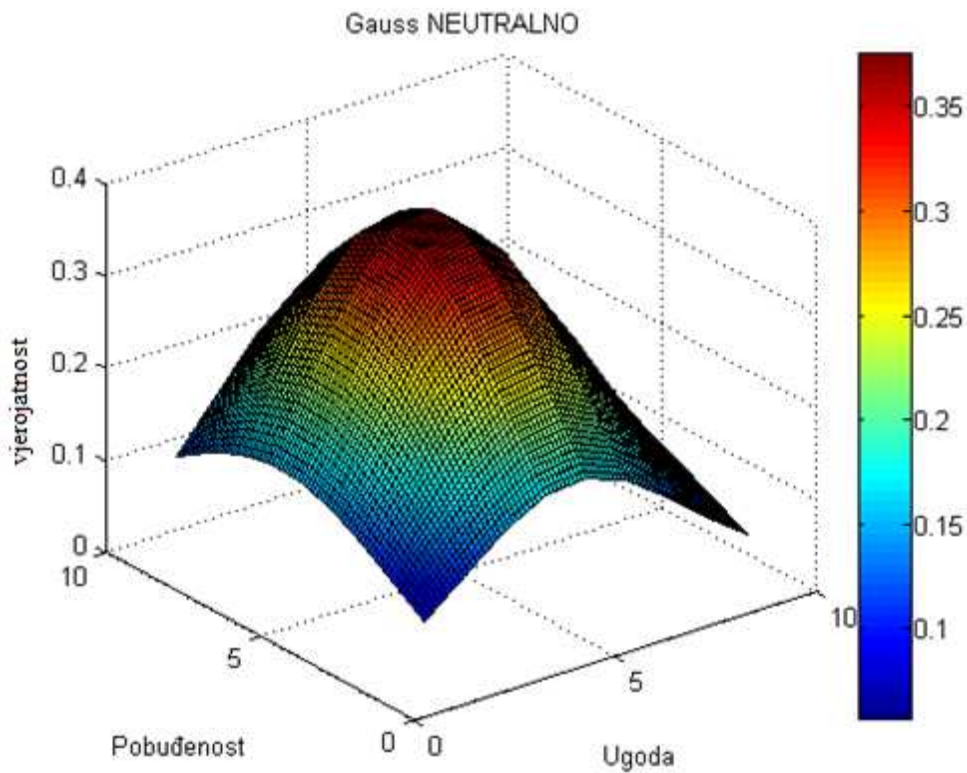
Slika 25 2D prikaz Gaussovog modela za emociju *strah*



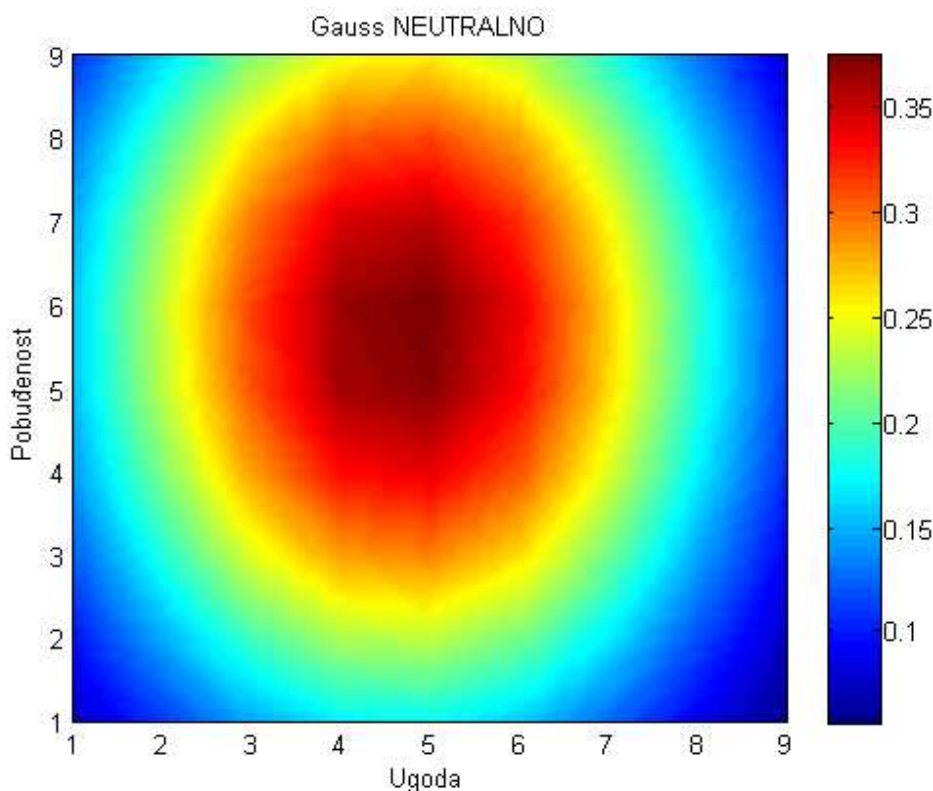
Slika 26 Gaussov model za emociju *ljutnja*



Slika 27 2D prikaz Gaussovog modela za emociju *ljutnja*



Slika 28 Gaussov model za neutralno stanje



Slika 29 2D prikaz Gaussovog modela za neutralno stanje

5.5. Testiranje Gaussovih modela

Testiranje modela dobivenih u Poglavlju 5.4. provodi se na 10% ukupnog korpusa. Taj uvjet osiguran je tako što su u petlju za testiranje ušle samo one snimke čiji je redni broj djeljiv s 10. Testiranje je postupak traženja onog Gaussovog modela koji za razinu ugode i pobuđenosti testne snimke ima najveću vjerojatnost. Nakon što je pronađen odgovarajući Gaussov model, valja ga usporediti s emocijom kojoj je pridijeljen maksimalni intenzitet za istu snimku. Ukoliko se ove dvije emocije (Gaussov model odnosi se na jednu emociju, a druga emocija je ona s maksimalnim intenzitetom) poklapaju, broj pogođenih emocija raste za 1. Ako se dogodi da neka snimka od strane nekog anotatora za više emocija ima pridijeljen jednak i ujedno maksimalan intenzitet ili je vjerojatnost Gaussovog modela za više emocija jednaka i ujedno maksimalna, postotak pogođenosti raste ukoliko se barem

jedna emocija označena maksimalnim intenzitetom poklapa s nekim od maksimalno vjerojatnih Gaussovih modela. Kada se obrade sve testne snimke po svim anotatorima, dobivamo postotak pogođenosti po svim anotatorima. Postotci pogođenosti po anotatorima prikazani su u Tablici 4.

	Anotator 1	Anotator 2	Anotator 3	Anotator 4	Anotator 5
Postotak pogođenosti (%)	40,3	31,3	35,8	34,3	22,4

Tablica 4 Tablica postotaka pogođenosti po anotatorima

6. Modeliranje emocija GMM modelima

Pogledamo li histograme u Poglavlju 5.1., ne možemo reći da su svi unimodalni, tj. da imaju jedan izražen maksimum. Gaussova dvodimenzionalna funkcija može dobro aproksimirati unimodalne histograme, međutim ako želimo preciznu aproksimaciju unimodalnih ili višemodalnih histograma, radije koristimo Gaussove mješavine (GMM – eng. *Gaussian Mixture Models*). Gaussove mješavine su dvodimenzionalne funkcije gustoće vjerojatnosti koje su zapravo težinski zbroj više Gaussovih dvodimenzionalnih funkcija gustoće vjerojatnosti. Kako mi raspolažemo dvjema varijablama (ugoda i pobuđenost), formule koje predstavljaju jedan GMM model su sljedeće:

$$GMM(o; w_i, \mu_i, C_i) = \sum_{i=1}^M w_i \cdot gauss(o; \mu_i, C_i)$$
$$gauss(o; \mu, C) = \frac{1}{2 \cdot \pi \cdot \sqrt{\det(C)}} \cdot e^{\frac{-1}{2} \cdot (x-\mu)^T \cdot C^{-1} \cdot (x-\mu)}$$

gdje su:

o = vektor observacije, tj. vektor čiji su elementi razina ugode i pobuđenosti

w = težinski koeficijent za pojedinu Gaussovu komponentu

μ = vektor srednjih vrijednosti po obje osi za svaku Gaussovu komponentu

C = matrica kovarijance za pojedinu Gaussovu komponentu

Matrice kovarijanci mogu biti dijagonalne ili punog ranga, a u ovome radu koriste se matrice kovarijance punog ranga kako ne bismo ograničavali Gaussove modele na poravnanje njihovih elipsoidnih presjeka sa koordinatnim osima. Također koristimo continuous model parametara, što znači da se za svaku emociju zasebno generiraju GMM modeli te svaki model ima zasebne parametre.

6.1. Treniranje GMM modela

Treniranje modela provodi se na 90% korpusa. Kao pretkorak treniranju modela, najprije se izvršava zašumljivanje uzoraka. Proces zašumljivanja uzoraka provodi se pomoću funkcija:

$$nova_ugoda = ugoda - 0.5 + rand$$

$$nova_pobuđenost = pobuđenost - 0.5 + rand$$

Ovakvim zašumljivanjem uzoraka postiže se veća raspršenost uzoraka te se na taj način pospješuje precizno određivanje GMM modela, koji se grade na temelju gustoće raspršenosti odgovarajućih uzoraka. Svaki uzorak (V,A) gornjim se formulama pomakne u neku točku u području $(V \pm 0.5, A \pm 0.5)$. Dodatno valja pripaziti da ukoliko jedna od dvije nove vrijednosti postane manja od 1, odnosno veća od 9, da se zadrže njihove stare vrijednosti 1, odnosno 9 kako bi se zadržala skala ugode i pobuđenosti na vrijednostima od 1 do 9.

Treniranje modela provodi se pomoću iterativnog EM algoritma (eng. *Expectation Maximization*) koji u svakoj iteraciji reestimacijom parametara C i μ podešava GMM model. Broj iteracija zadaje se kao parametar.

EM algoritam koristi Bayesovo pravilo uvjetne vjerojatnosti:

$$P(\lambda_e|o) = \frac{P(o|\lambda_e) \cdot P(\lambda_e)}{P(o)}$$

gdje su:

$$P(\lambda_e|o) = a \text{ posteriori vjerojatnost klase } e$$

$$P(o|\lambda_e) = \text{vjerodostojnost (eng. likelihood)}$$

$$P(o) = \text{vjerojatnost observacije}$$

$P(\lambda_e)$ = a priori vjerojatnost klase e - u našem slučaju ovaj se parametar zanemaruje jer je svaki model jednako vjerojatan

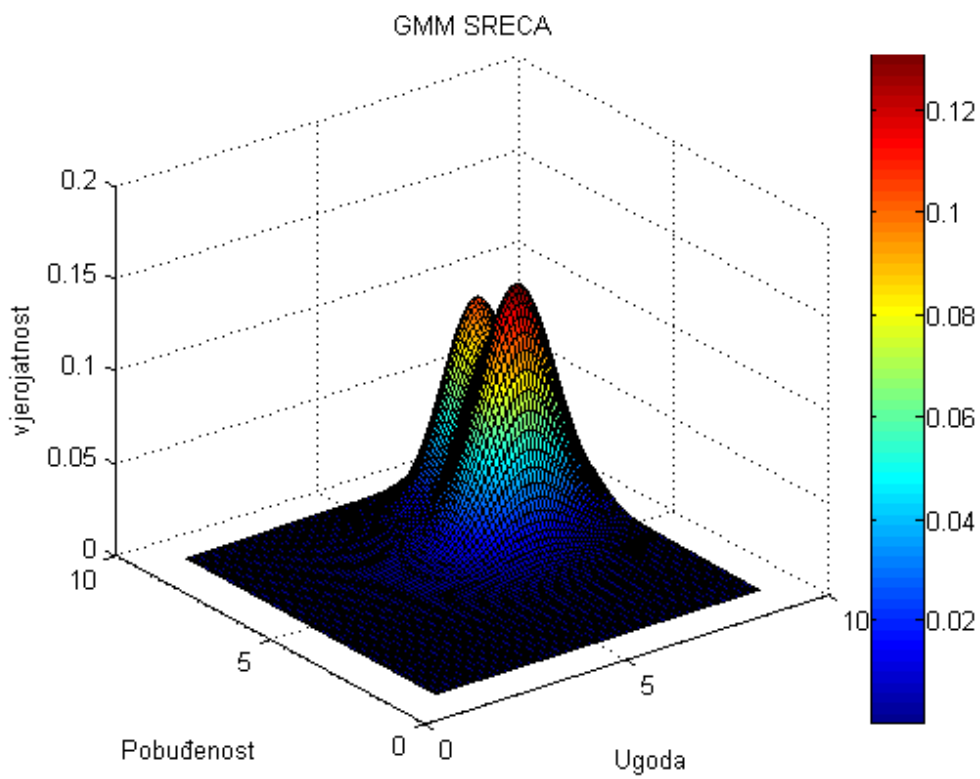
Nakon određenog broja iteracija, vjerodostojnost počinje konvergirati te više nema smisla provoditi algoritam jer su promjene nezamjetne.

Svaki GMM model sastoji se od nekoliko Gaussovih komponenata. Nakon eksperimentalne promjene parametra koji označava broj Gaussovih modela u GMM-u, odnosno broj iteracija EM algoritma, uz primjećivanje pri izvršenju koje po redu iteracije vjerodostojnost konvergira, vizualnom se ocjenom Gaussovog modela došlo se do zaključaka koliko je Gaussovih komponenata, odnosno iteracija EM algoritma potrebno za pojedini GMM model, a ti zaključci predstavljeni su u Tablici 5.

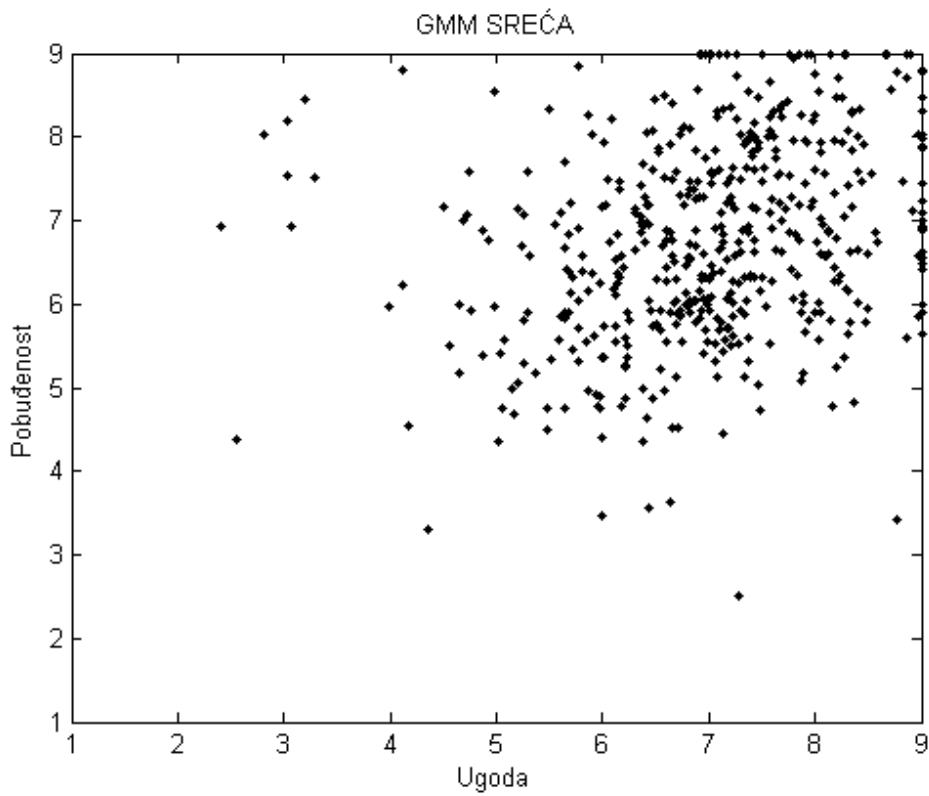
emocija	potreban broj iteracija	broj Gaussovih komponenata
SREĆA	60	3
TUGA	5	2
STRAH	15	2
LJUTNJA	20	2
NEUTRALNO	50	2

Tablica 5 Prikaz broja iteracija te broja Gaussovih komponenata za pojedini GMM model

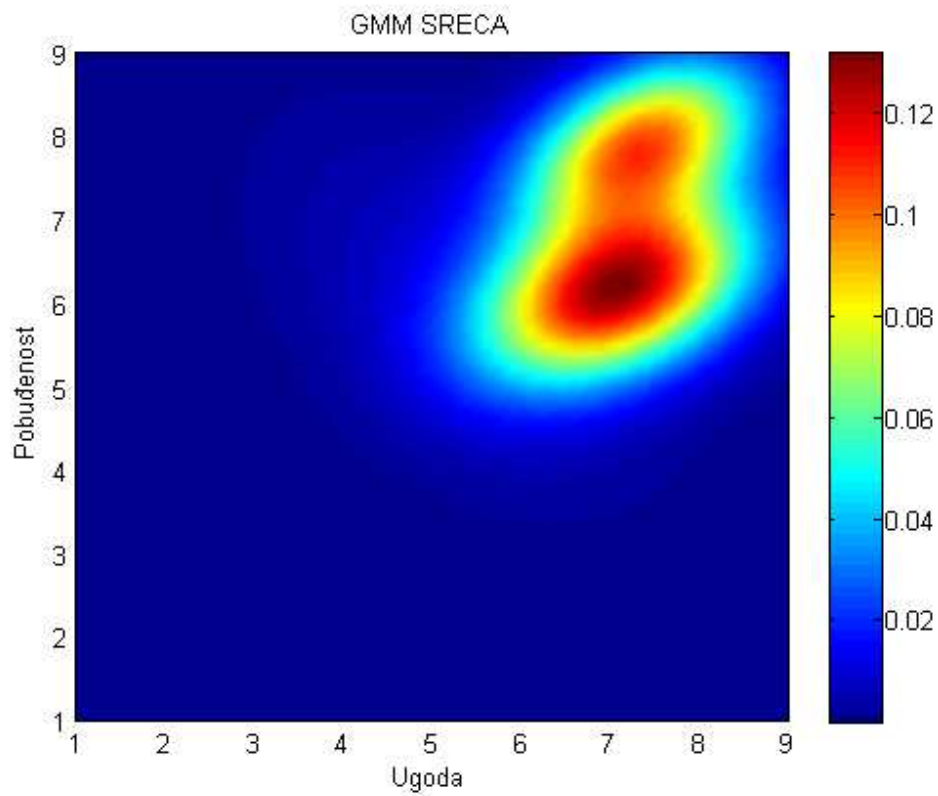
Na sljedećim slikama možemo vidjeti trodimenzionalne i dvodimenzionalne GMM modele dobivene za svaku emociju uz parametre navedene u Tablici 5.



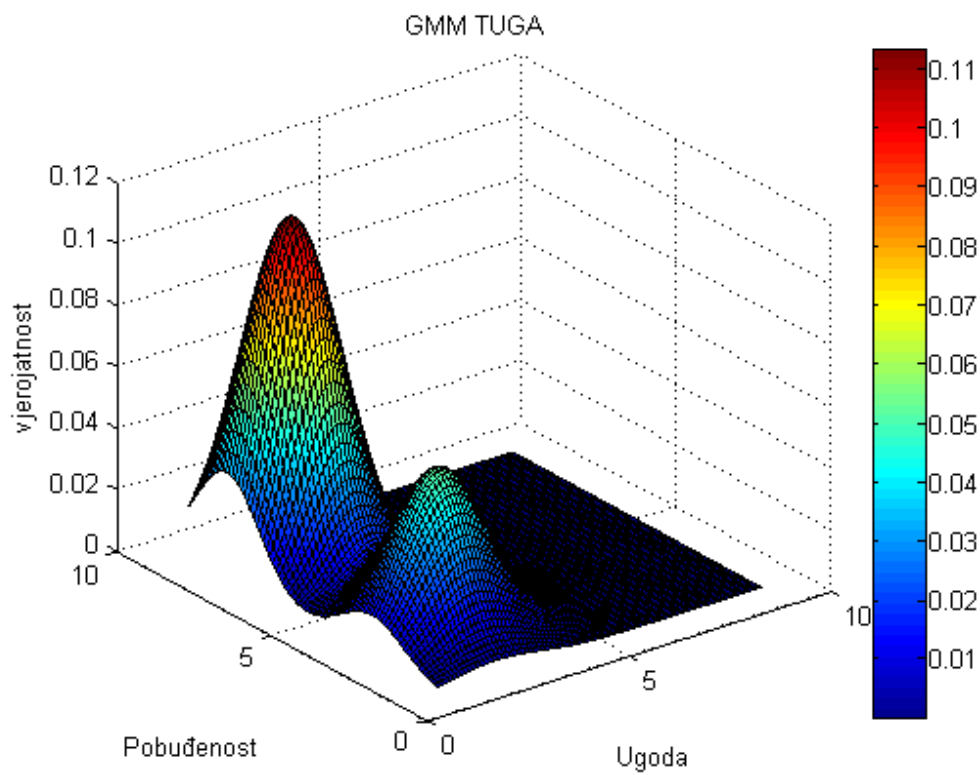
Slika 30 GMM model za emociju sreća



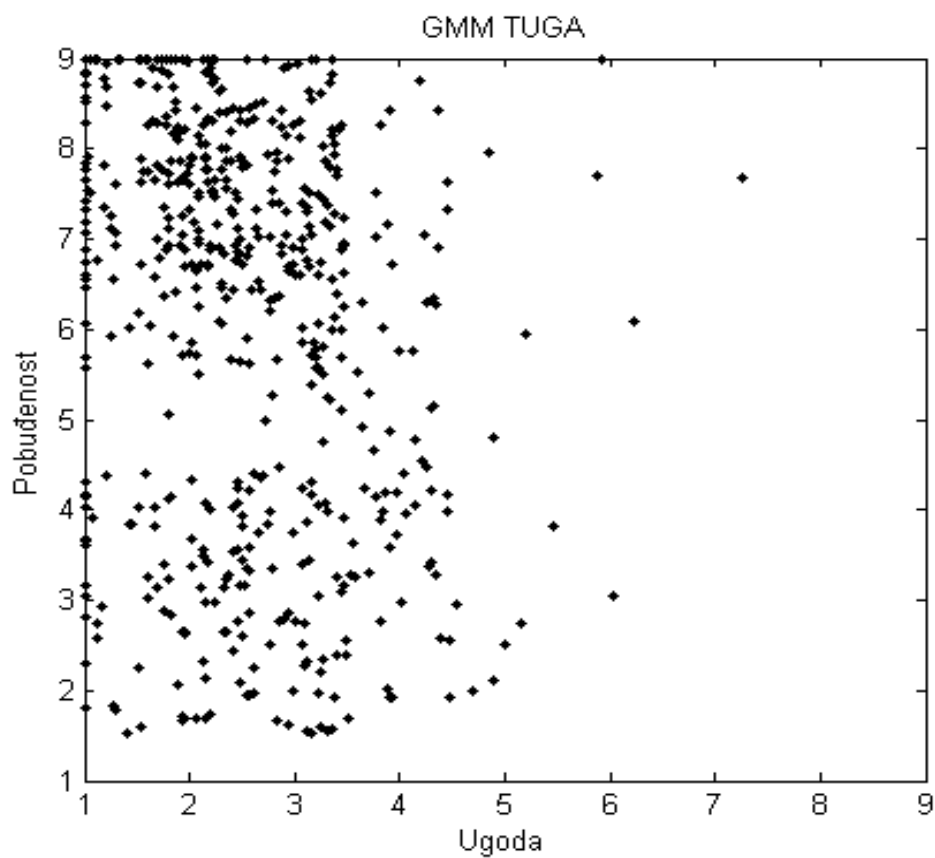
Slika 31 Raspršenost uzoraka za emociju sreća



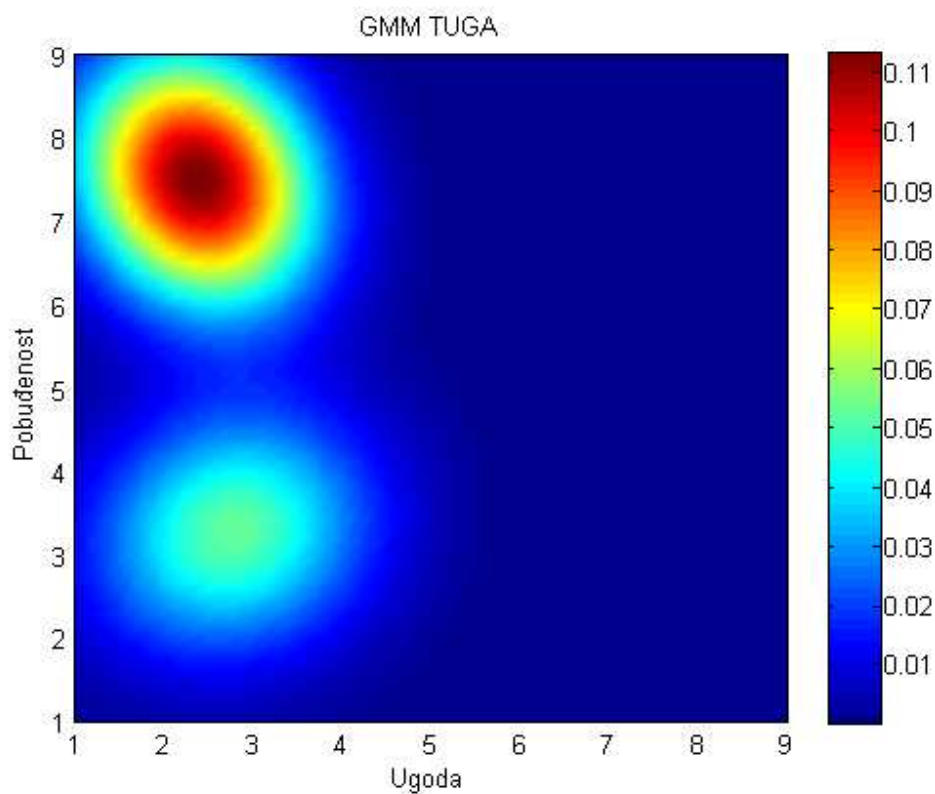
Slika 32 2D prikaz GMM modela za emociju sreća



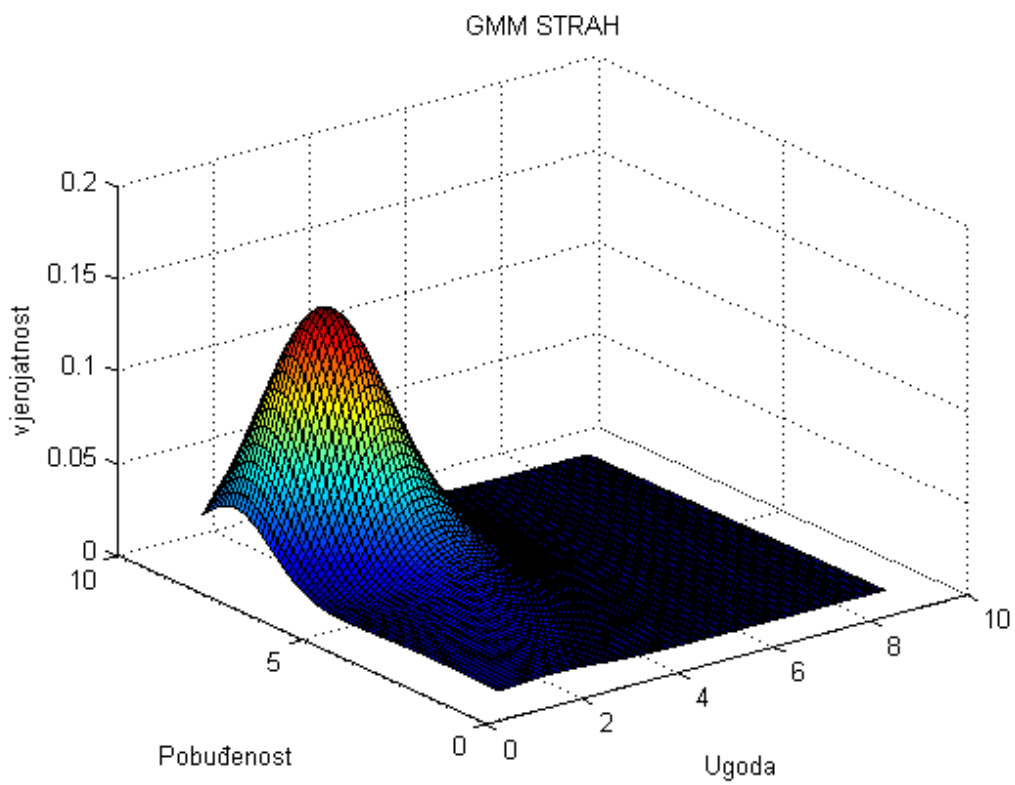
Slika 33 2D prikaz GMM modela za emociju tuga



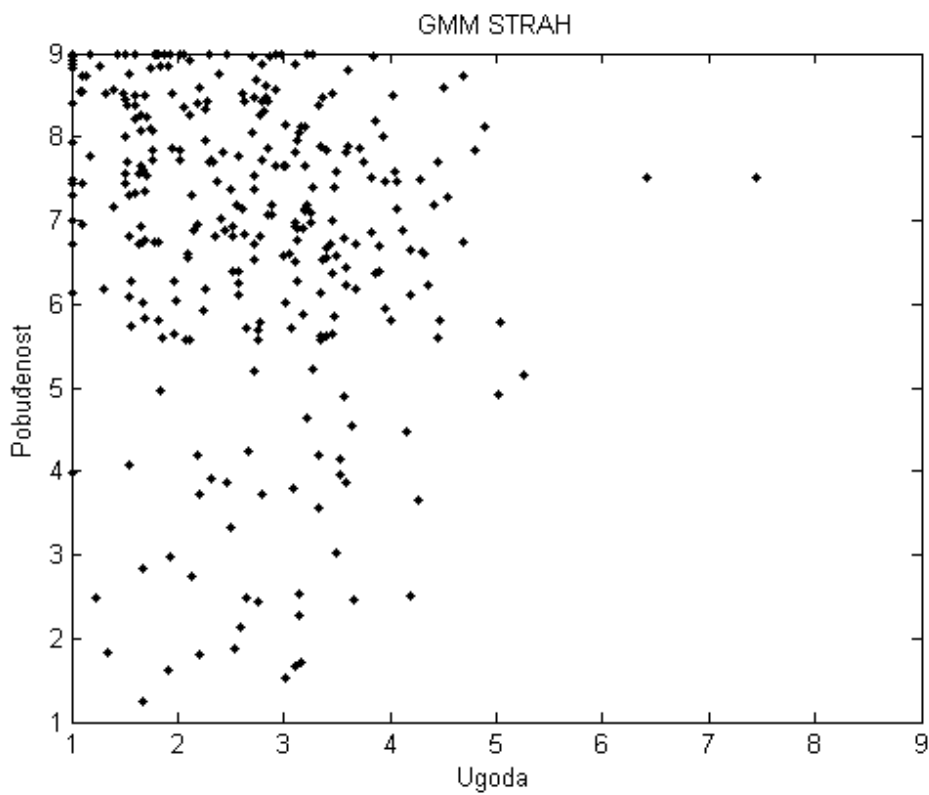
Slika 34 Raspršenost uzoraka za emociju *tuga*



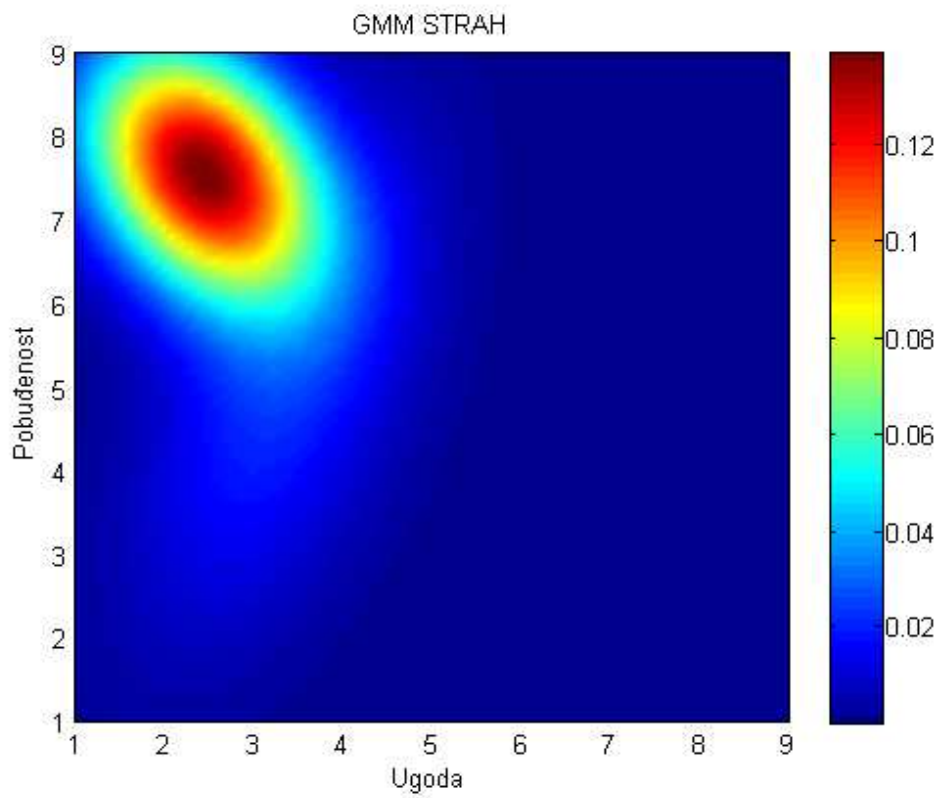
Slika 35 GMM model za emociju *tuga*



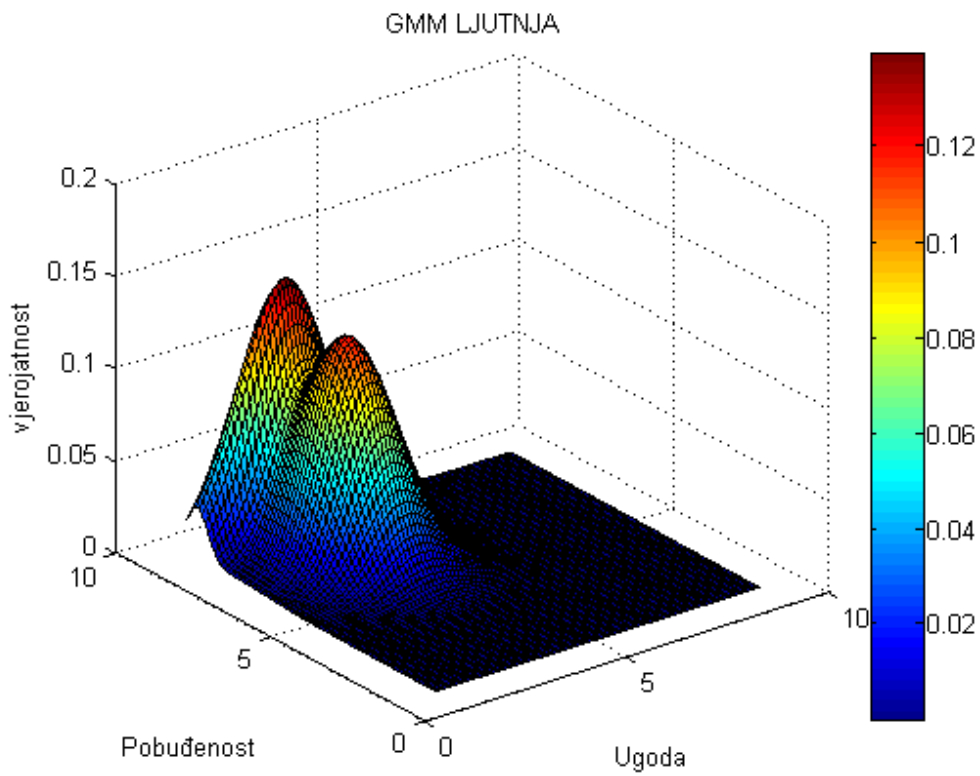
Slika 36 GMM model za emociju *strah*



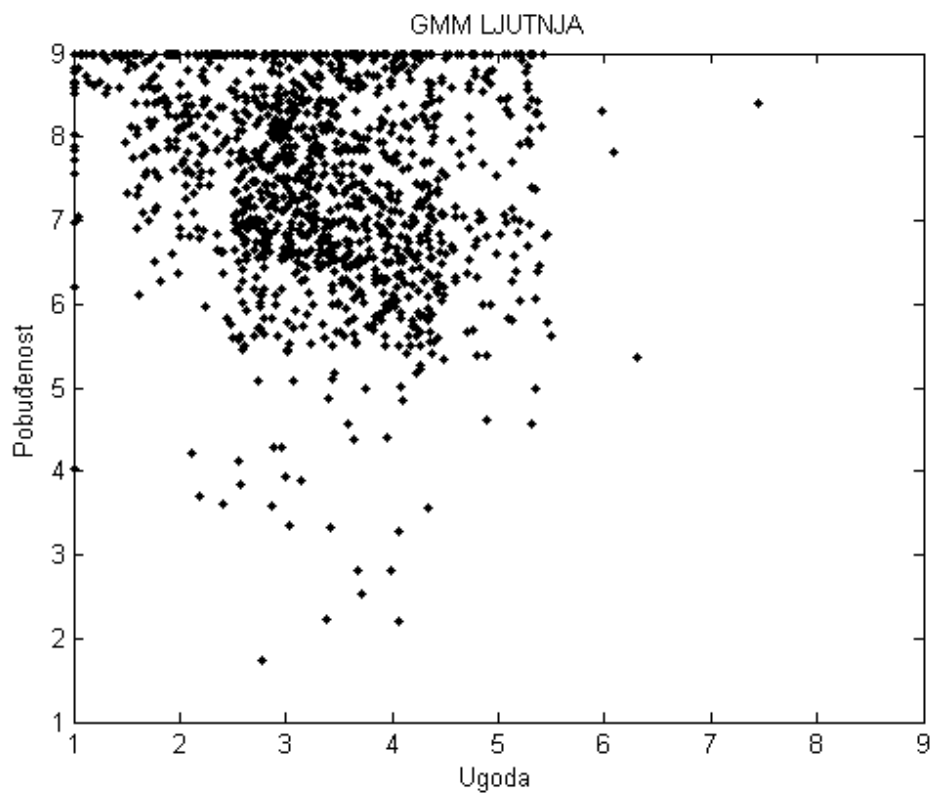
Slika 37 Raspršenost uzoraka za emociju *strah*



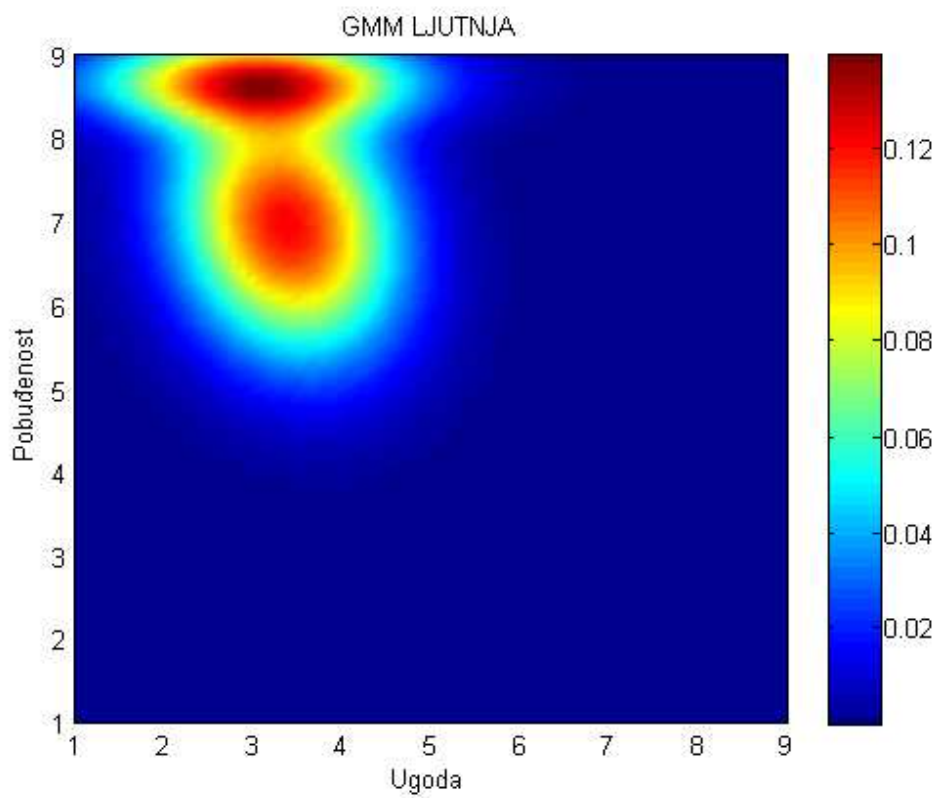
Slika 38 2D prikaz GMM modela za emociju *strah*



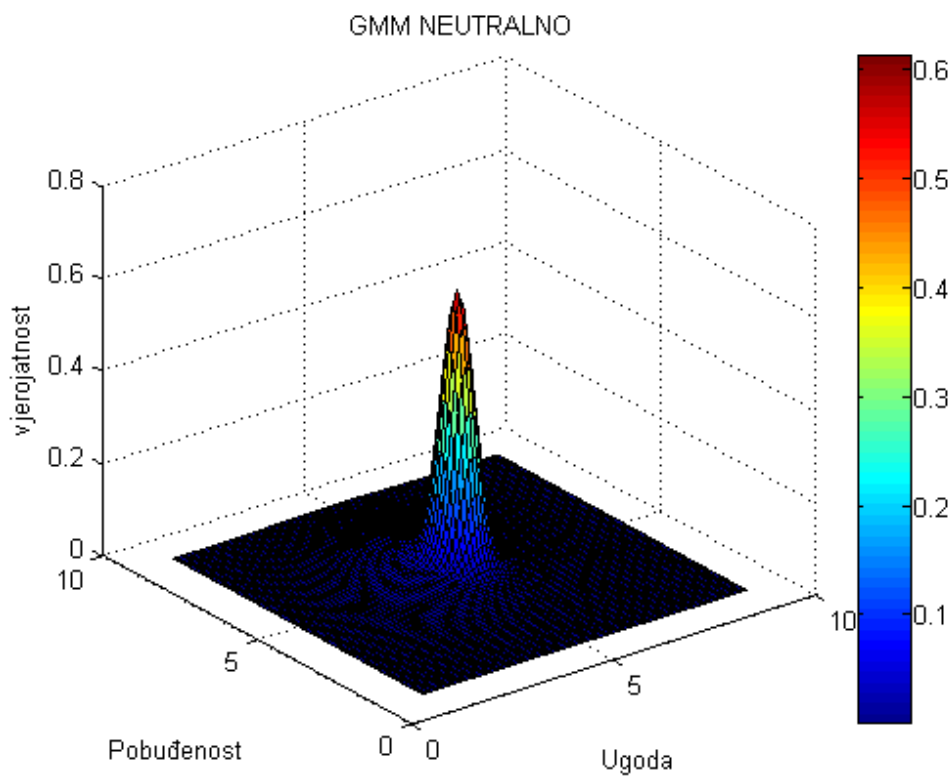
Slika 39 GMM model za emociju *ljutnja*



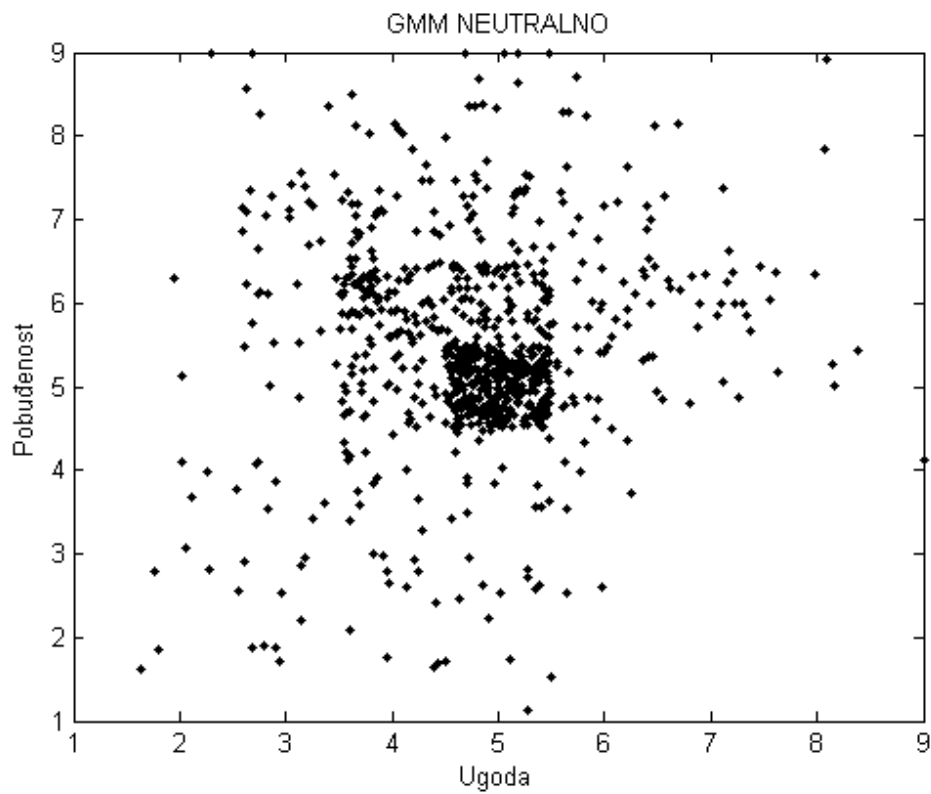
Slika 40 Raspršenost uzoraka za emociju *ljutnja*



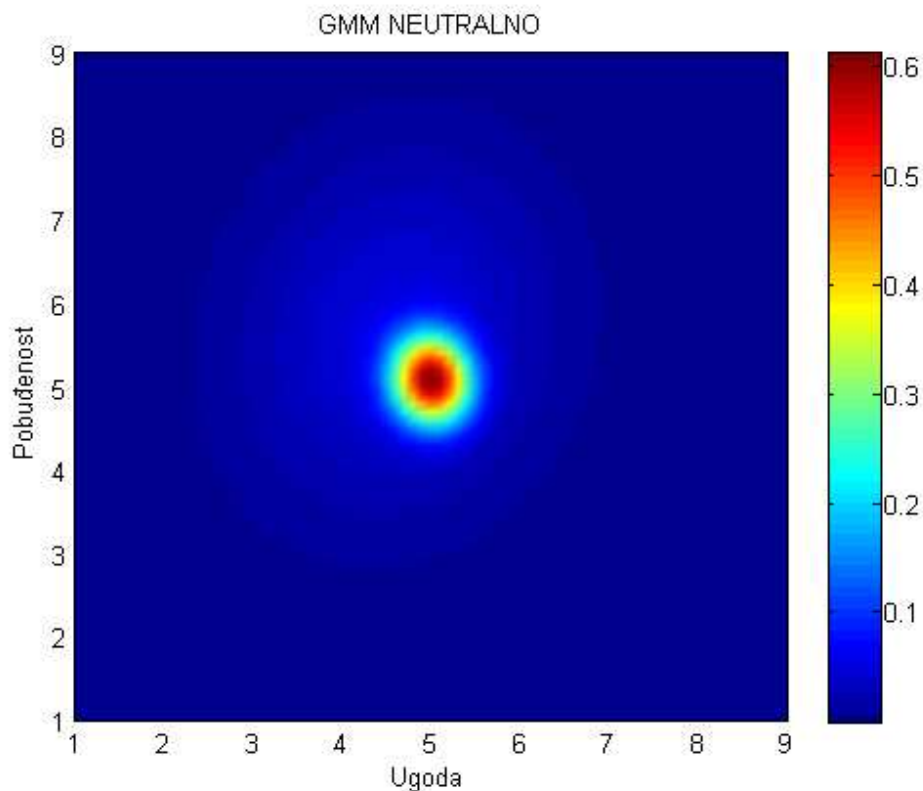
Slika 41 2D prikaz GMM modela za emociju *ljutnja*



Slika 42 GMM model za neutralno stanje



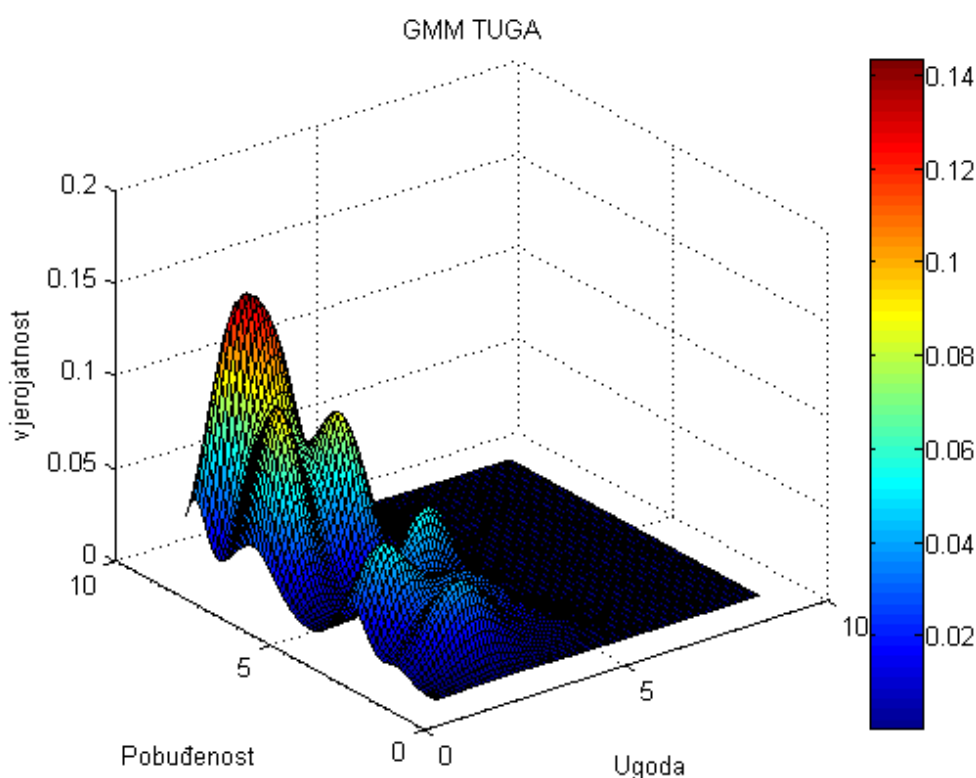
Slika 43 Raspršenost uzoraka za *neutralno stanje*



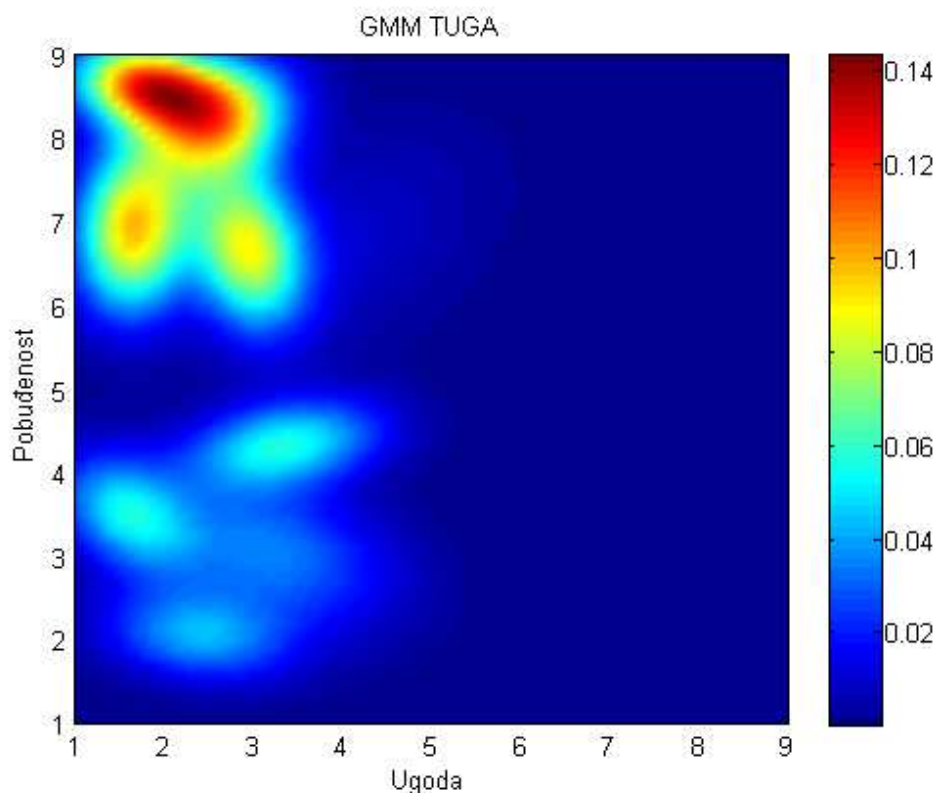
Slika 44 2D prikaz GMM modela za neutralno stanje

Ukoliko analiziramo dobivene modele, možemo primijetiti da su se uzorci emocije sreća grupirali u tri skupine, od čega su dvije dominantne, a to su sreća s visokom razinom pobuđenosti (oko 8-9) te sreća s malo nižom razinom ugode (oko 6-7). Uzorci tuge rasporedili su se također u dvije velike skupine, gdje jedna predstavlja uzorke mirnije tuge (govor kroz tiho jecanje ili tužan govor), dok druga predstavlja uzorke snažne tuge (govor kroz snažan plač i deranje). Uzorci straha koncentrirani su većinom u jednoj skupini, dok je druga neznatne visine. Ljutnja je, pak, slično kao i strah, grupirana u dvjema skupinama, od kojih jedna predstavlja snažnu ljutnju s vrlo visokom pobuđenosti (uglavnom razine oko 9), a druga nešto slabiju ljutnju s razinom pobuđenosti oko 7. Neutralno stanje koncentrirano je uglavnom u jednoj Gaussovoj komponenti, iako GMM model ima dvije komponente, a razine ugode i pobuđenosti spomenute dominantne komponente su, kako je i očekivano, oko srednjih vrijednosti ugode i pobuđenosti na skali od 1 do 9, što je razina 5 za obje varijable.

Kako bismo stekli uvid u to što se događa ukoliko neki GMM model ima previše Gaussovih komponenata, pogledajmo Slike 45 i 46 koje predstavljaju GMM model emocije *tuga* uz 100 iteracija EM algoritma te 9 Gaussovih komponenata. Ukoliko se nakratko vratimo na histogram emocije *tuga* prikazan na Slici 7, možemo zamijetiti da je GMM model modelirao skoro svaki viši stupac histograma jednom Gaussovom komponentom. Možemo zamisliti kada bismo povećali broj Gaussovih komponenata na, primjerice, nekoliko stotina, dobili bismo isključivo „šiljke“, što bi zapravo bile Diracove funkcije. Takvo rješenje za naše istraživanje ne bi bilo pogodno jer mi želimo doći do nekih općenito primjenjivih zaključaka, a ne točno modelirati procjene isključivo ovih anotatora.



Slika 45 GMM model za emociju *tuga* (broj Gaussovih komponenata = 9)



Slika 46 2D prikaz GMM modela za emociju tuga (broj Gaussovih komponenata = 9)

6.2. Testiranje GMM modela

Za testiranje GMM modela koristi se preostalih 10% korpusa, a razine ugone i pobuđenosti tih snimaka također se zašumljuju na isti način opisan u Poglavlju 6.1. Testiranje je postupak traženja onog Gaussovog modela koji za razinu ugone i pobuđenosti testne snimke ima najveću vjerojatnost. Nakon što je pronađen odgovarajući Gaussov model, valja ga usporediti s emocijom kojoj je pridijeljen maksimalni intenzitet za istu snimku. Ukoliko se ove dvije emocije (GMM model se odnosi na jednu emociju, a druga emocija je ona s maksimalnim intenzitetom) poklapaju, broj pogođenih emocija raste za 1. Ako se dogodi da neka snimka od strane nekog anotatora za više emocija ima pridijeljen jednak i ujedno maksimalan intenzitet ili je vjerojatnost GMM modela za više emocija jednaka i ujedno maksimalna, postotak pogođenosti raste ukoliko se barem jedna emocija označena maksimalnim intenzitetom poklapa s nekim od maksimalno vjerojatnih GMM modela. Kada se obrade

sve testne snimke po svim anotatorima, dobivamo postotak pogođenosti po svim anotatorima. Postotci pogođenosti po anotatorima za slučaj kada se koriste GMM modeli prikazani su u Tablici 6.

	Anotator 1	Anotator 2	Anotator 3	Anotator 4	Anotator 5
Postotak pogođenosti (%)	56.7	65.7	62.7	59.7	74.6

Tablica 6 Tablica postotaka pogođenosti po anotatorima

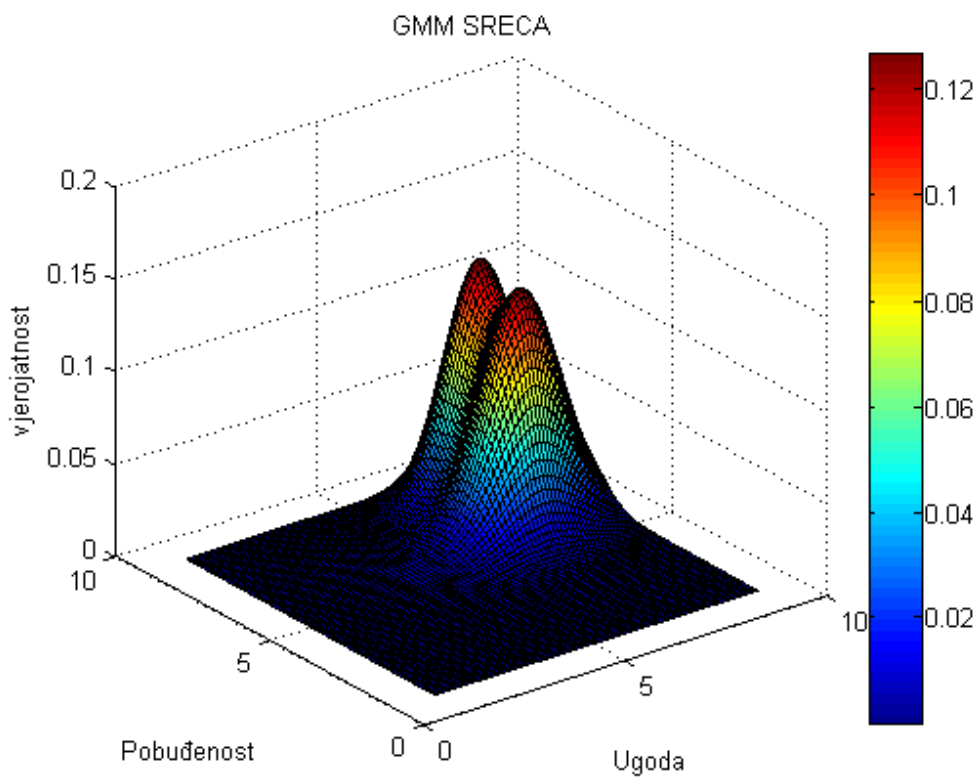
7. Modeliranje emocija GMM modelima s preferencijama

U do sada opisanim postupcima treniranja modela nismo uzimali u obzir vrijednost intenziteta emocija. Uzmimo konkretni primjer kako bismo bolje razumjeli o čemu se radi. Za neku snimku jedan anotator ocijenio je intenzitete sreće, tuge, straha i ljutnje respektivno 0,4,6,0. Prilikom gradnje histograma (Poglavlje 5.1.) broj uzoraka se nakon učitanih navedenih ocjena za emociju tuge i za emociju straha povećao za 1, bez obzira što je strahu pridodan veći intenzitet. Cilj ovog dijela rada je na neki način diferencirati ocjene prema razini intenziteta.

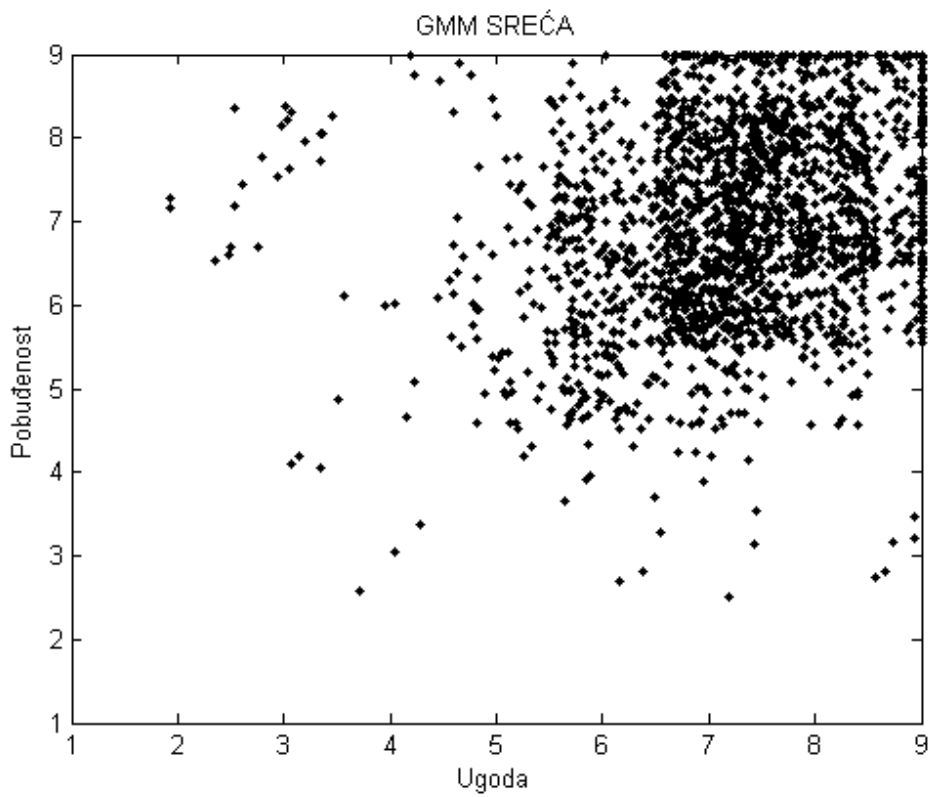
7.1. Treniranje GMM modela s preferencijama

Treniranje modela ponovno se provodilo na 90% korpusa. U ovome poglavlju ponavljamo postupke iz Poglavlja 6., ali uz preferencije uzoraka. Dakle, prilikom učitavanja ocjena snimki, ukoliko je intenzitet pojedine emocije vrijednosti x , broj uzoraka za tu emociju raste za $x-3$. Primjerice, ukoliko je intenzitet sreće za neku snimku ocijenjen intenzitetom 7, broj novih uzoraka sreće raste za 4. Svaki od tih novih uzoraka zašumljuje se zasebno, dakle uzorci će se nalaziti negdje u području $(V \pm 0.5, A \pm 0.5)$, gdje su V i A razine ugone i pobuđenosti za obrađivanu snimku. Neutralno stanje je to neutralnije što je intenzitet ostalih emocija niži. Tako ako su intenziteti sreće, tuge, straha i ljutnje ocijenjeni sa isključivo manjim intenzitetima od 4, broj uzoraka neutralnog stanja raste za y , gdje se y računa po formuli: *-maksimalni intenzitet za tu snimku + 4*. Tako, primjerice, ukoliko je za neku snimku maksimalan intenzitet 3, broj uzoraka neutralnog stanja raste za 1. Naravno, svaki novonastali uzorak zasebno se zašumljuje prema postupku iz Poglavlja 6.1. Dobiveni modeli prikazani su na sljedećim slikama.

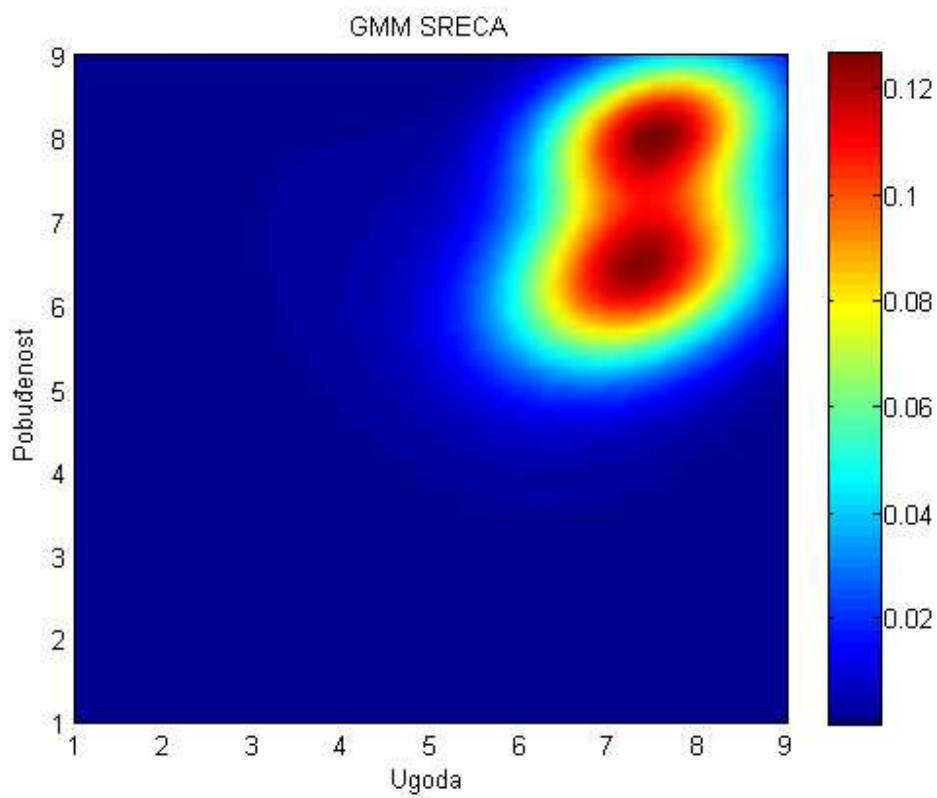
Na sljedećim slikama prikazani su GMM modeli s preferencijama.



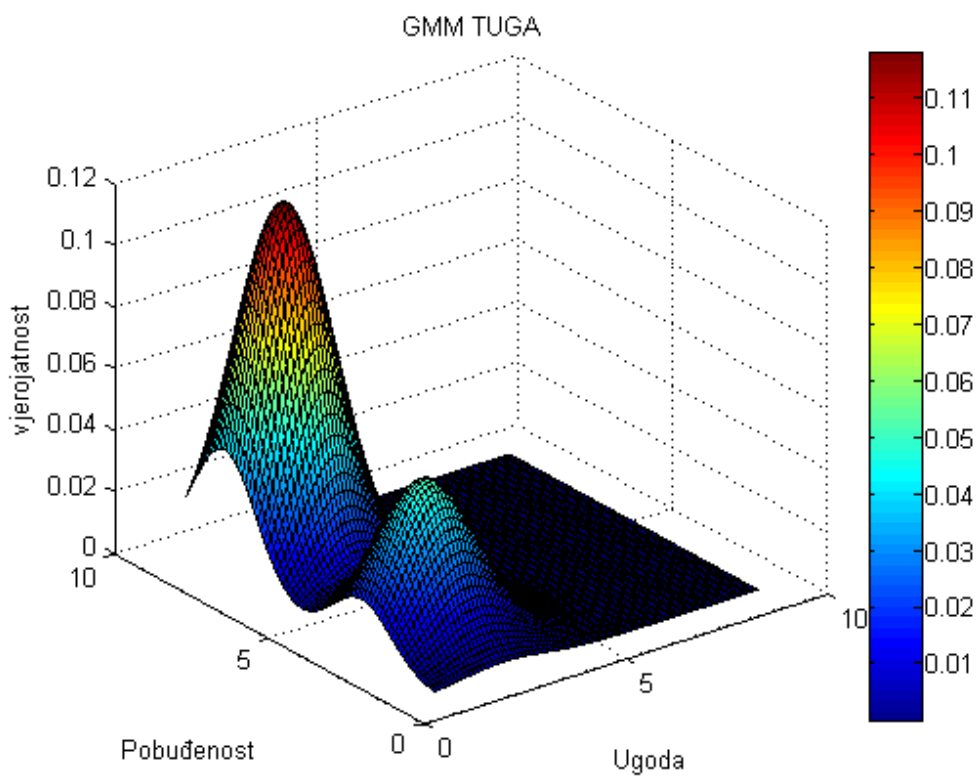
Slika 47 GMM model s preferencijama za emociju sreća



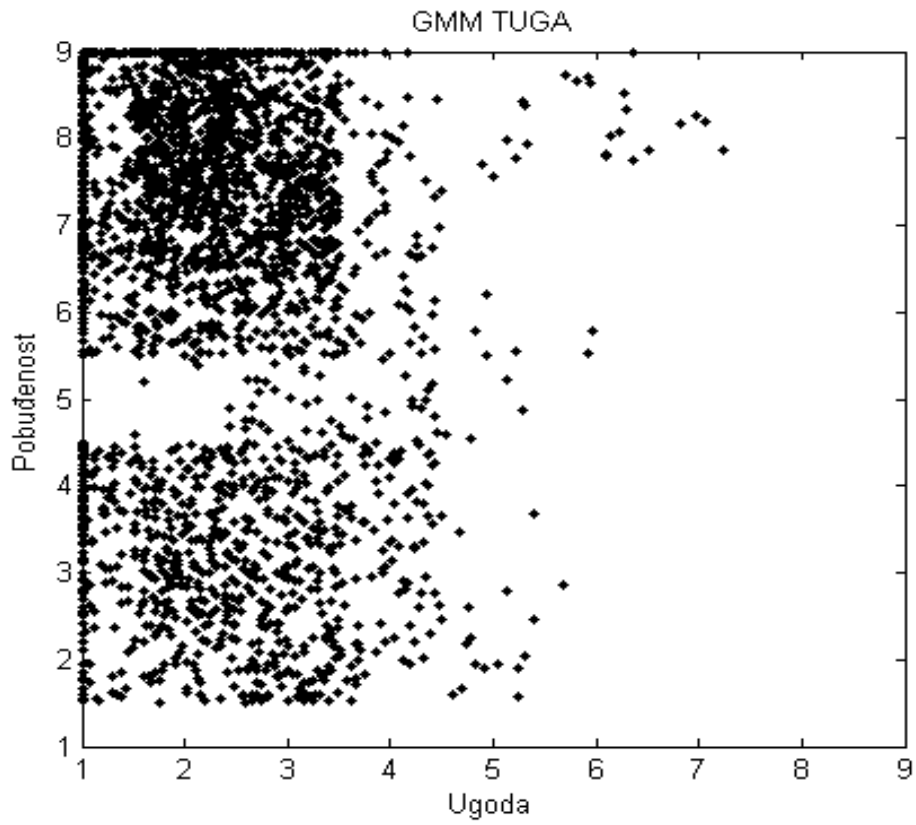
Slika 48 Raspršenost uzoraka za emociju sreća



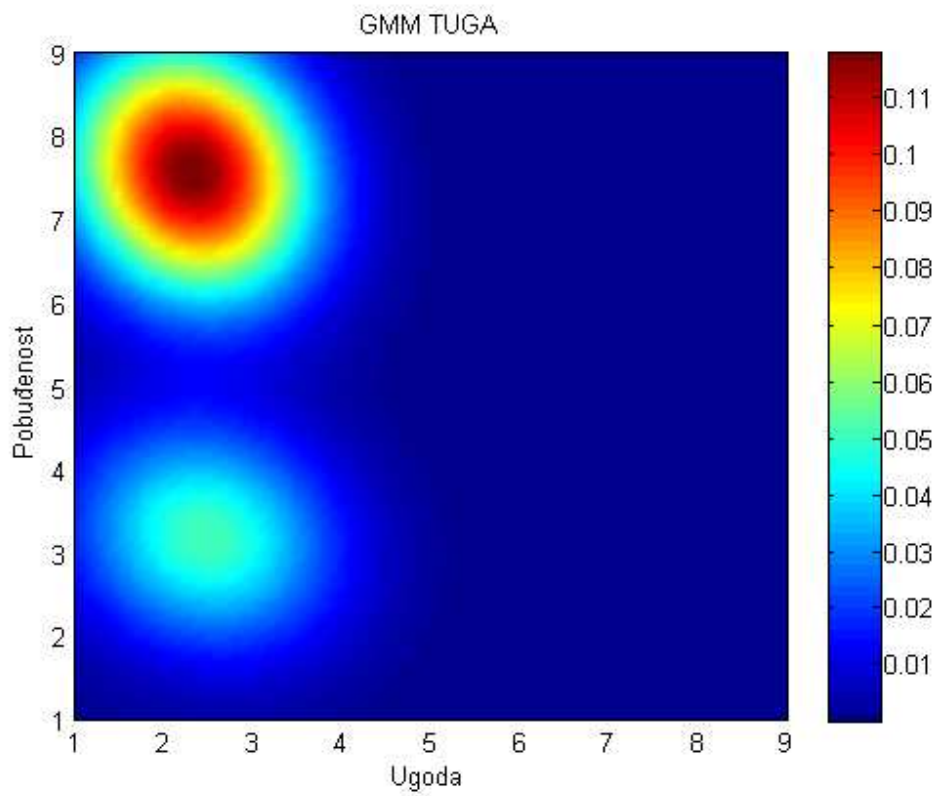
Slika 49 2D prikaz GMM modela s preferencijama za emociju sreća



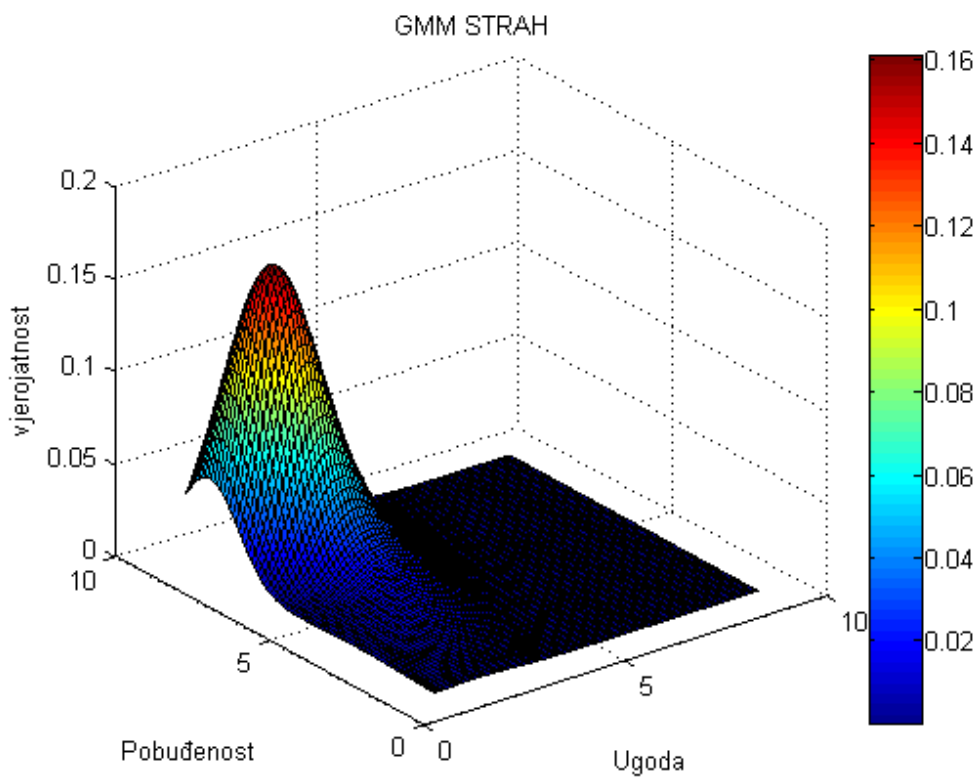
Slika 50 GMM model s preferencijama za emociju tuga



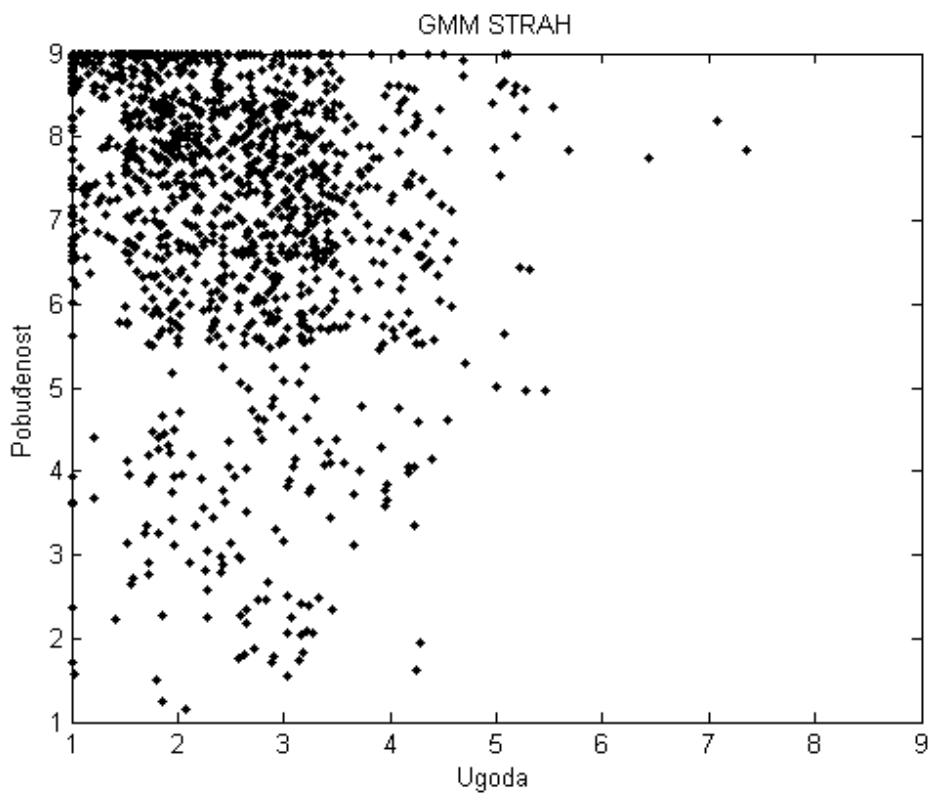
Slika 51 Raspršenost uzoraka za emociju *tuga*



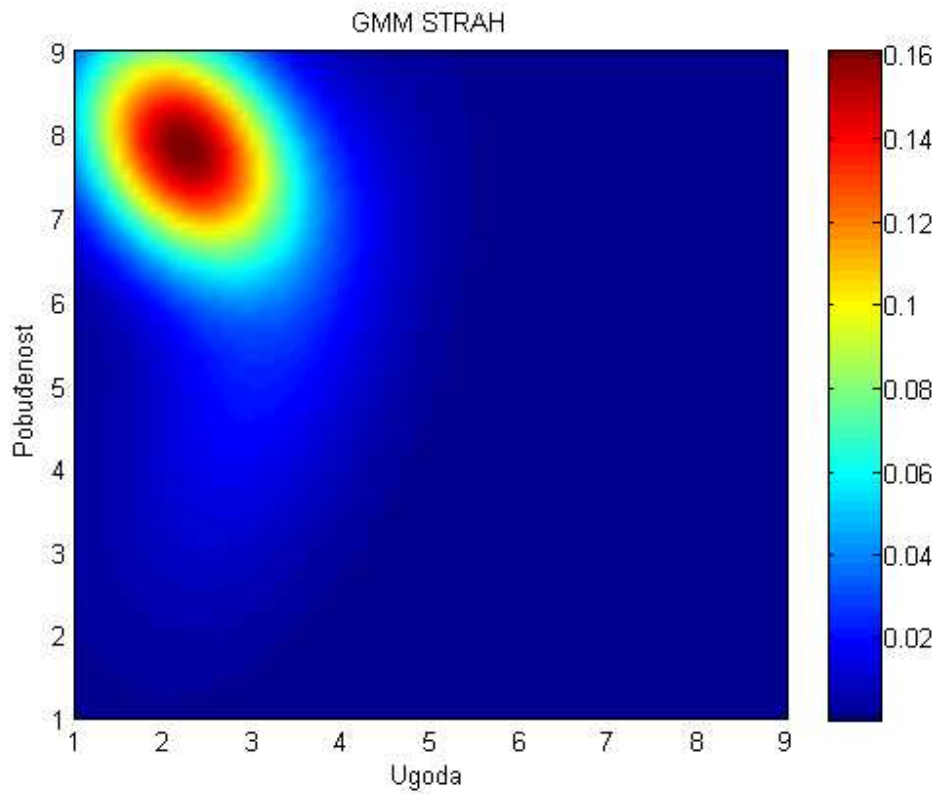
Slika 52 2D prikaz GMM modela s preferencijama za emociju *tuga*



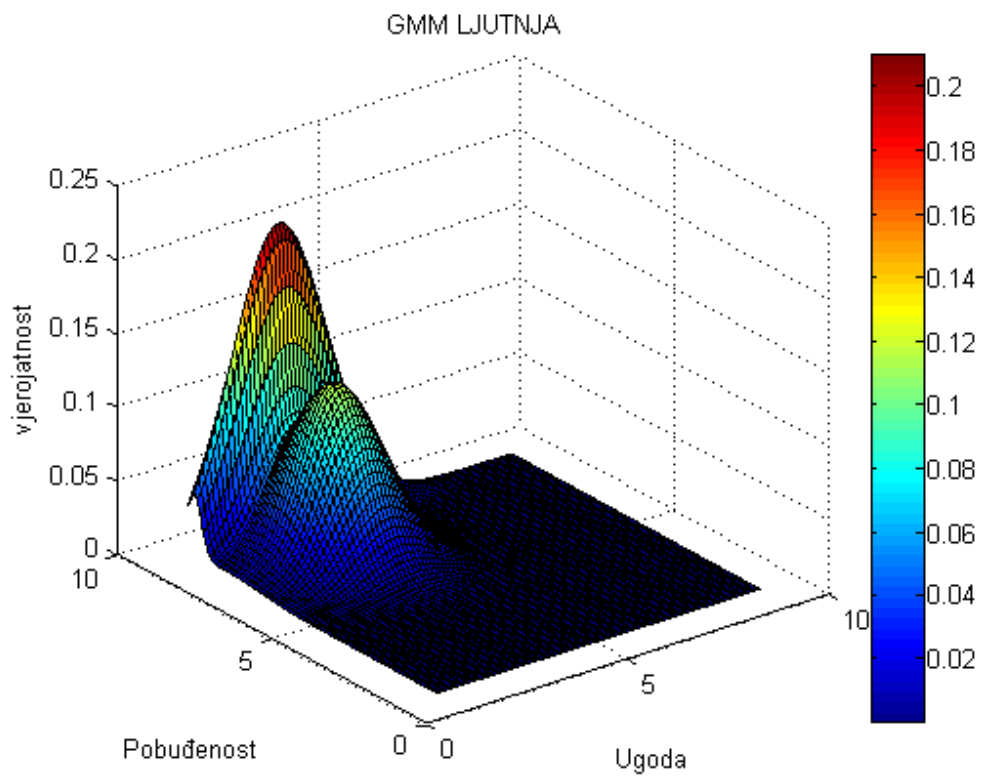
Slika 53 GMM model s preferencijama za emociju *strah*



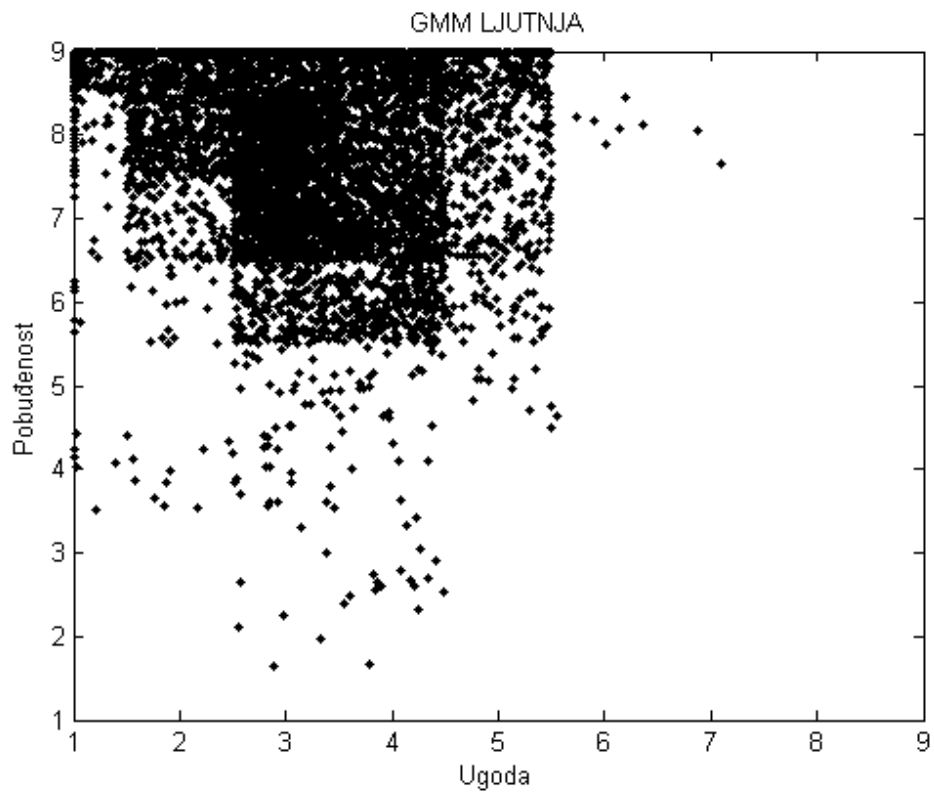
Slika 54 Raspršenost uzoraka za emociju *strah*



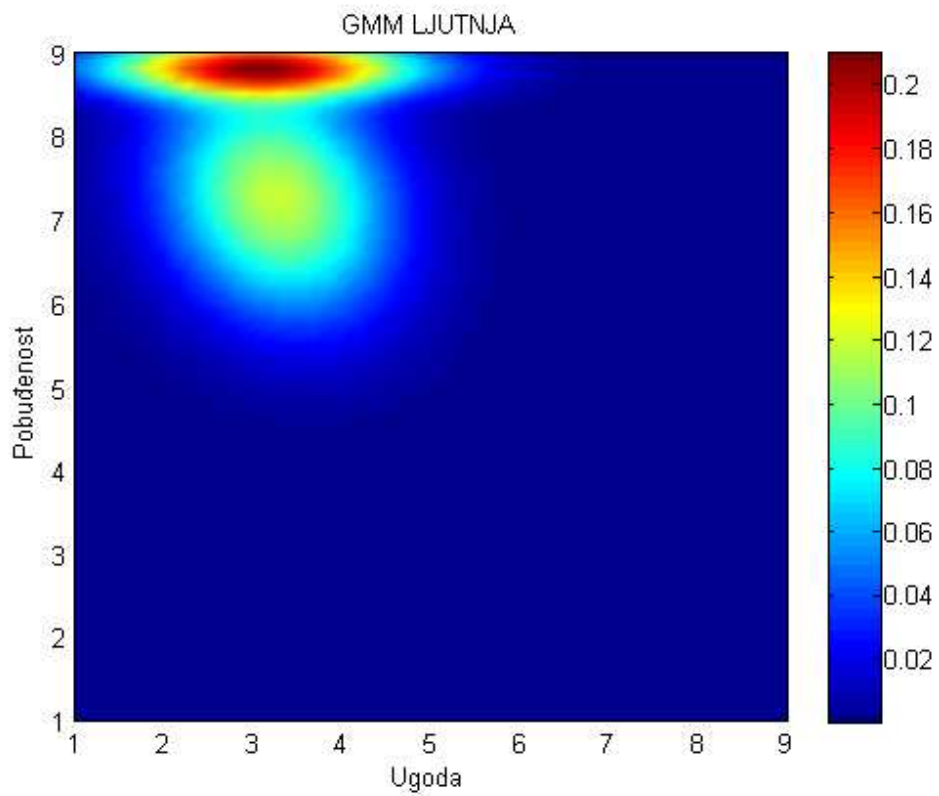
Slika 55 2D prikaz GMM modela s preferencijama za emociju *strah*



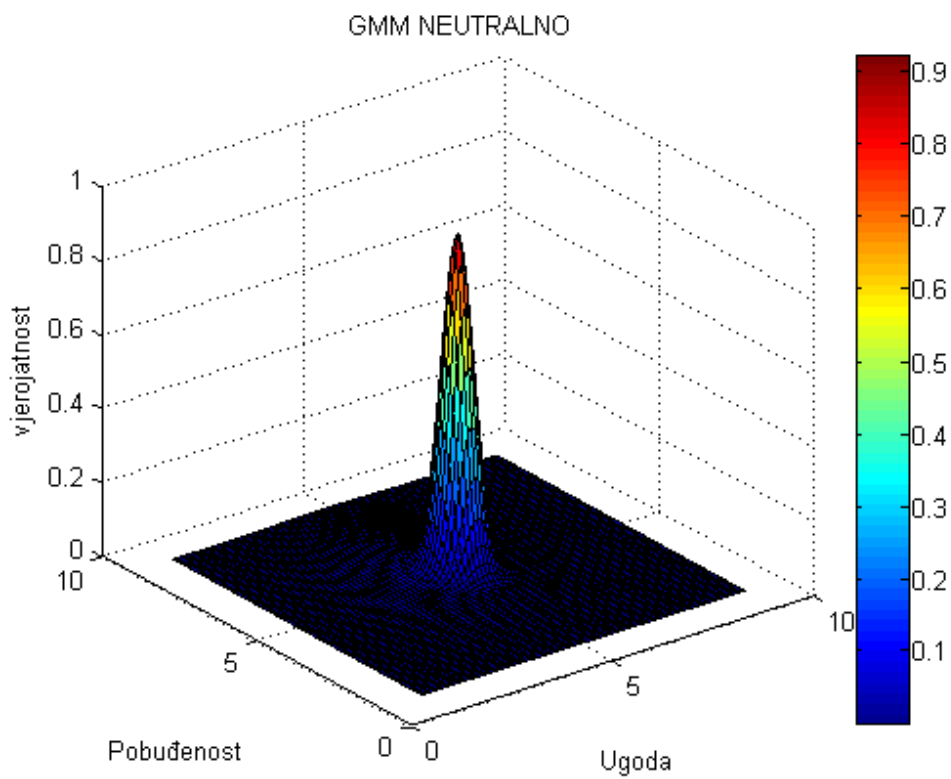
Slika 56 GMM model s preferencijama za emociju *ljutnja*



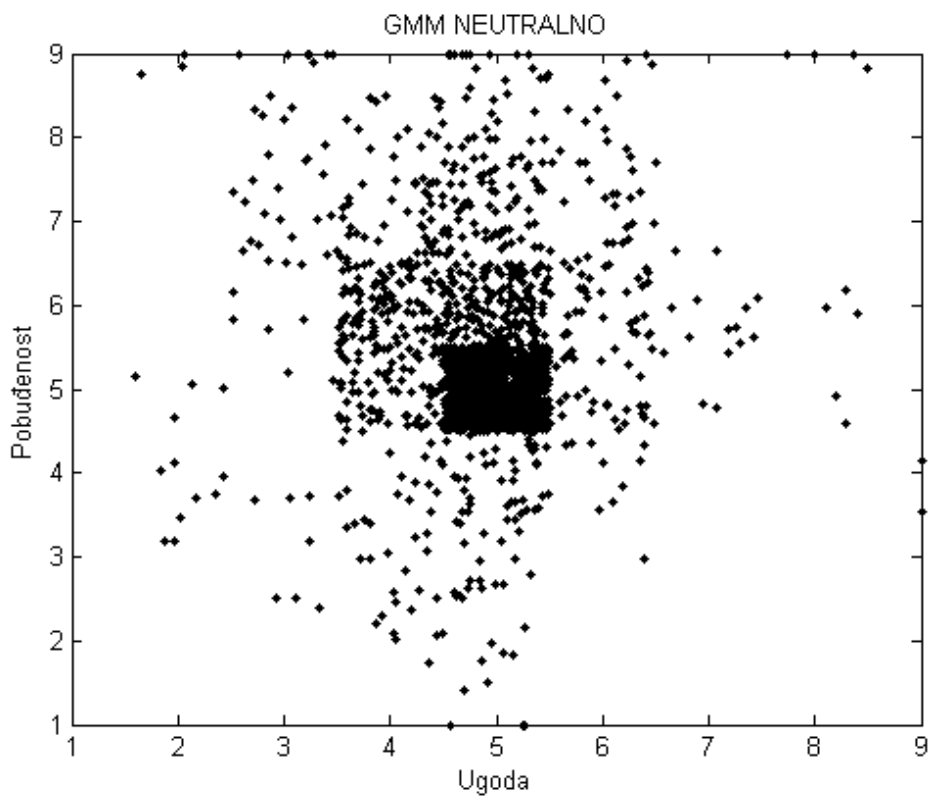
Slika 57 Raspršenost uzoraka za emociju *ljutnja*



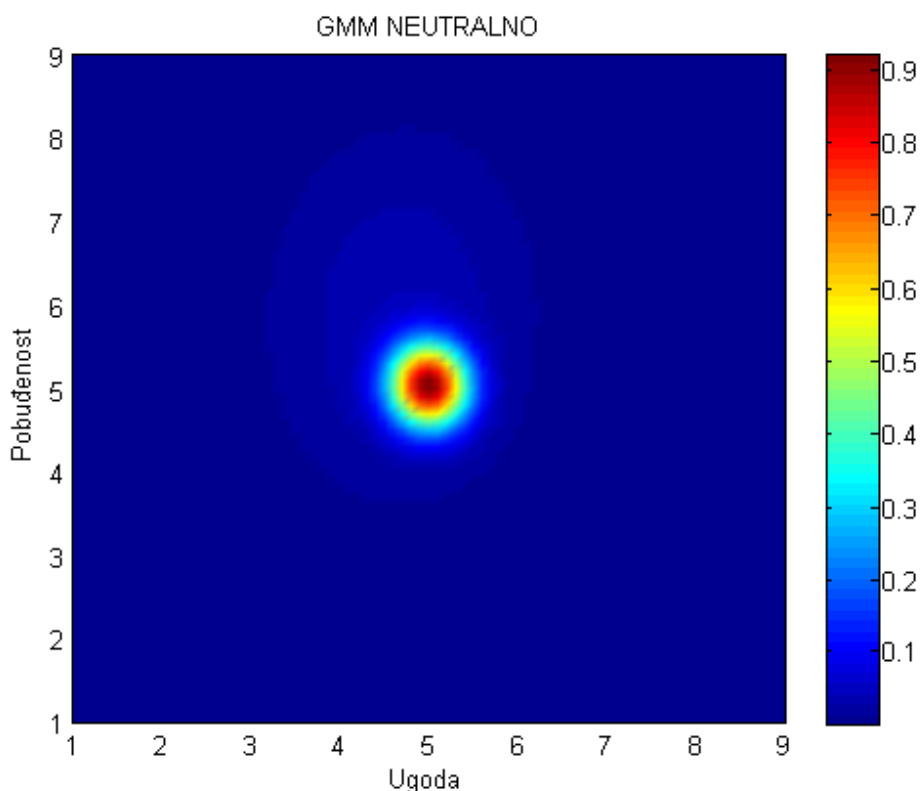
Slika 58 2D prikaz GMM modela s preferencijama za emociju *ljutnja*



Slika 59 GMM model s preferencijama za *neutralo stanje*



Slika 60 Raspršenost uzoraka za *neutraln stanje*



Slika 61 2D prikaz GMM modela s preferencijama za *neutralno stanje*

7.2. Testiranje GMM modela s preferencijama

Testiranje modela provodilo se na 10% korpusa, a razine ugode i pobuđenosti tih snimaka također se zašumljuju na isti način opisan u Poglavlju 6.1. Testiranje je postupak traženja onog Gaussovog modela koji za razinu ugode i pobuđenosti testne snimke ima najveću vjerojatnost. Nakon što je pronađen odgovarajući Gaussov model, valja ga usporediti s emocijom kojoj je pridijeljen maksimalni intenzitet za istu snimku. Ukoliko se ove dvije emocije (GMM model s preferencijama odnosi se na jednu emociju, a druga emocija je ona s maksimalnim intenzitetom) poklapaju, broj pogođenih emocija raste za 1. Ako se dogodi da neka snimka od strane nekog anotatora za više emocija ima pridijeljen jednak i ujedno maksimalan intenzitet ili je vjerojatnost GMM modela s preferencijama za više emocija jednaka i ujedno maksimalna, postotak pogođenosti raste ukoliko se barem jedna emocija

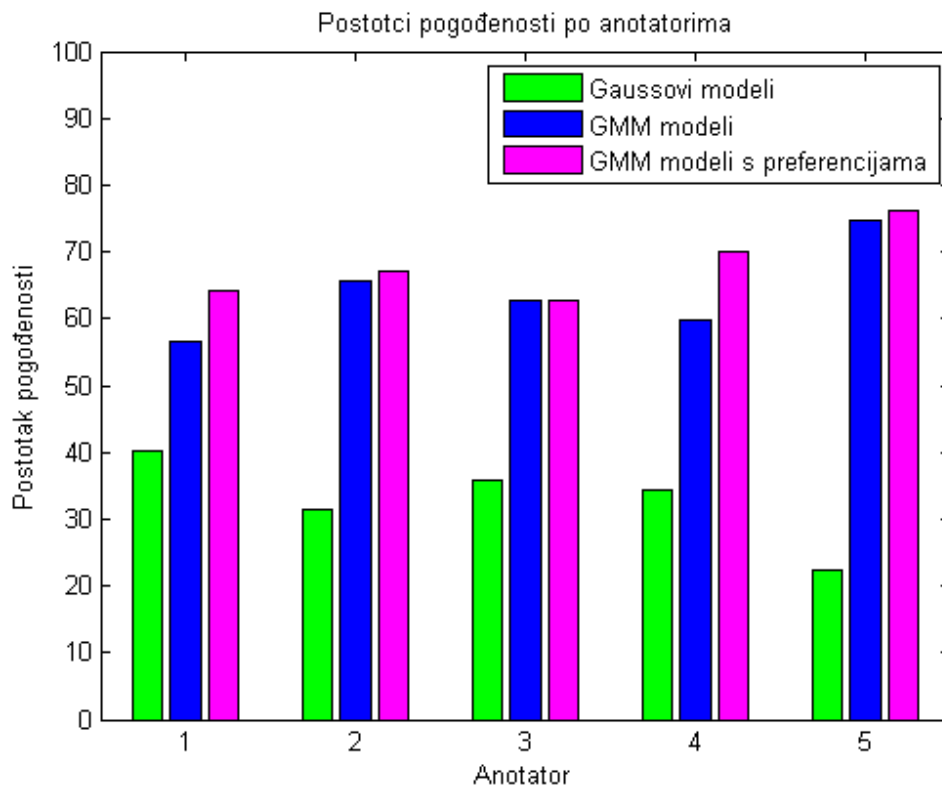
označena maksimalnim intenzitetom poklapa s nekim od maksimalno vjerojatnih Gaussovih modela. Kada se obrade sve testne snimke po svim anotatorima, dobivamo postotak pogođenosti po svim anotatorima. Postotci pogođenosti po anotatorima za slučaj kada se koriste GMM modeli s preferencijama prikazani su u Tablici 7.

	Anotator 1	Anotator 2	Anotator 3	Anotator 4	Anotator 5
Postotak pogođenosti (%)	64.2	67.2	62.7	70.1	76.1

Tablica 7 Tablica postotaka pogođenosti po anotatorima

8. Usporedba uspješnosti modela

Na sljedećem grafu možemo vidjeti usporedbu postotaka pogođenosti iz procesa testiranja u slučaju Gaussovih modela, GMM modela te GMM modela s preferencijama.



Slika 62 Usporedba postotaka pogođenosti emocija za Gaussove modele, GMM modele te GMM modele s preferencijama

Prema gornjem grafu vidljivo je da GMM modeli s preferencijama daju najbolje rezultate, tj. postotci pogođenosti emocija su najviši u konkurenciji tri načina treniranja modela. Razlog je u tome što GMM modeli s preferencijama najpreciznije opisuju raspodjelu uzoraka, dok su Gaussovi modeli najgrublja aproksimacija raspodjele. Opcija GMM modela s preferencijama najuspješnija je i zato što pri treniranju uzima u obzir ulazne intenzitete pojedine emocije te na taj način gradi modele koji ovise o navedenim intenzitetima.

9. Zaključak

Nakon provedene statističke analize zaključak je da GMM modeli s preferencijama daju najbolje rezultate jer najpreciznije modeliraju estimacije anotatora te su prilikom testiranja pomoću tih modela dobiveni najviši postotci pogodnosti emocija za svakog anotatora. Pomoću svih do sada izrađenih algoritama ostvarena je direktna veza između procjena diskretnih i procjena dimenzionalnih emocija. Konačno, nadogradnja ovog sustava tako što bi se kao ulaz promatrale i glasovne značajke poput frekvencije ili intenziteta govora, bila bi još jedan korak naprijed k raspoznavanju emocionalnog stanja iz govora govornika.

Nastavak ovog rada bio bi provedba istih algoritama na najnovijem korpusu koji se sastoji od 1146 snimaka te je ocijenjen od strane 115 anotatora te uzimanje u obzir sigurnosti koju svaki anotator izražava za pojedinu snimku, kao i izvora anotatorove procjene (akustika, semantika ili kombinacija).

10. Literatura

- [1] Šverko, Zarevski, Szabo, Kljaić, Kolega, Turudić-Čuljak, **Psihologija**, Školska knjiga, 2006.
- [2] Dropuljić, B., **Estimacija emocionalnih stanja na temelju glasovnih značajki**, 2012.
- [3] Chmura, M., **Automatska klasifikacija diskretnih emocionalnih stanja na osnovu akustičkih značajki govora**, 2011.
- [4] Damić, E., Tomac, T., **Proširenje baze audio isječaka za automatsku klasifikaciju diskretnih emocionalnih stanja na osnovu akustičkih značajki govora**, 2012.
- [5] Moore, A., **Clustering with Gaussian Mixtures**, <http://www.autonlab.org/tutorials/gmm14.pdf>, 2004., datum zadnjeg pristupa: 19. lipnja 2012.
- [6] Storkey, A., **Learning from Data: Gaussian Mixture Models**, <http://www.inf.ed.ac.uk/teaching/courses/ldf/lectures/gmm-print.pdf>, 2005., datum zadnjeg pristupa: 19. lipnja 2012.
- [7] Wikipedia, **Multivariate normal distribution**, http://en.wikipedia.org/wiki/Multivariate_normal_distribution, datum zadnjeg pristupa: 19. lipnja 2012.
- [8] **funkcija AutoGaussianSurf**, www.mathworks.com/matlabcentral/fileexchange/31485, datum zadnjeg pristupa: 18. lipnja 2012.
- [9] **funkcija error_ellipse**, www.mathworks.com/matlabcentral/fileexchange/4705, datum zadnjeg pristupa: 18. lipnja 2012.

11. Sažetak

Emocije ili osjećaji definiraju se kao doživljaj našeg subjektivnog odnosa prema stvarima, ljudima, događajima i samima sebi. One su intenzivne, relativno kratkog trajanja te usmjerene prema vanjskom objektu ili internim mislima. Reprezentiraju se kao konačan skup diskretnih emocija te kao dimenzionalne emocije s određenom razinom ugone i pobuđenosti u istoimenom koordinatnom sustavu. Najprije su pomoću ocjena od strane pet različitih anotatora formirani histogrami za pojedinu emociju, a potom su histogrami aproksimirani Gausovim dvodimenzionalnim funkcijama koji su testirani na 10% snimaka korpusa. Nadalje, uzorci su se aproksimirali GMM modelima, tj. višekomponentnim Gausovim dvodimenzionalnim funkcijama, a nakon toga je u obzir uzet i intenzitet kao zasebni parametar koji utječe na preferencije pojedinih uzoraka. Najbolje rezultate očekivano je dala zadnja opcija, gdje je postotak pogodnosti po anotatorima bio oko 70%.

12. Abstract

Emotions are defined as the subjective experience of our attitude towards things, people, events and ourselves. They are intense, of relatively short duration and focused on the external object or internal thoughts. They can be represented as a finite set of discrete emotion and dimensional emotion with a certain level of valence and arousal in the same coordinate system. First, using the annotations of five different annotators histograms for each emotion were formed, and then histograms were approximated with two-dimensional Gaussian functions that were tested at 10% of the corpus. Furthermore, the samples were approximated with GMM models, which are multicomponent two-dimensional Gaussian functions, and subsequently with GMM models but with intensity taken into account as a separate parameter which influences the preferences of individual samples. The best results were obtained using the last option, as expected. The percentage of getting the right emotion per annotator was about 70%.