

UNIVERZITET U BOGRADU
ELEKTROTEHNIČKI FAKULTET

Željko B. Nedeljković

OPTIMIZACIJA SISTEMA ZA AUTOMATSKO
PREPOZNAVANJE GOVORNIH EMOCIJA

doktorska disertacija

Beograd, 2020.

UNIVERSITY OF BELGRADE
SCHOOL OF ELECTRICAL ENGINEERING

Željko B. Nedeljković

OPTIMIZATION OF AUTOMATIC SPEECH
EMOTION RECOGNITION SYSTEMS

Doctoral Dissertation

Belgrade, 2020.

Mentor:

dr Željko Đurović, redovni profesor,
Univerzitet u Beogradu – Elektrotehnički fakultet

Članovi komisije:

dr Aleksandra Marjanović, docent,
Univerzitet u Beogradu – Elektrotehnički fakultet

dr Zoran Perić, redovni profesor,
Univerzitet u Nišu – Elektronski fakultet

dr Dragana Šumarac Pavlović, redovni profesor,
Univerzitet u Beogradu – Elektrotehnički fakultet

Datum odbrane: _____

Zahvalnica

Želeo bih da iskažem iskrenu zahvalnost svom mentoru prof. dr. Željku Đuroviću na velikoj podršci, mudrom usmeravanju i dragocenim savetima tokom istraživanja i pisanja doktorske disertacije.

Zahvaljujem se kolegici Milani Milošević na uspešno ostvarenoj saradnji tokom trajanja doktorskih studija.

Srdačno se zahvaljujem i komisiji koja je pregledala ovaj rad i svojim korisnim sugestijama uticala na konačnu verziju disertacije.

Zahvalnost na nesebičnoj podršci i razumevanju dugujem i svojoj izabranici Mirjani, roditeljima Miladini i Branku, kao i bratu Darku.

Posebnu zahvalnost na čvrstoj veri i ohrabrenju povodom ovog naučno-istraživačkog poduhvata dugujem blagopočivšim baki Perki i dedi Cakanu.

Naslov disertacije: Optimizacija sistema za automatsko prepoznavanje govornih emocija.

Sažetak: Osnov za uspešnu integraciju emocionalne inteligencije u sofisticirane sisteme veštačke inteligencije jeste pouzdano prepoznavanje emocionalnih stanja, pri čemu se paralingvistički sadržaj govora izdvaja kao posebno značajan nosilac informacija o emocionalnom stanju govornika. U ovom radu je sprovedena komparativna analiza obeležja govornog signala i klasifikatorskih metoda najčešće korišćenih u rešavanju zadatka automatskog prepoznavanja emocionalnih stanja govornika, a zatim su razmotrene mogućnosti popravke performansi sistema za automatsko prepoznavanje govornih emocija. Izvršeno je unapređenje diskretnih skrivenih Markovljevih modela upotrebom QQ krive za potrebe određivanja etalona vektorske kvantizacije, a razmotrena su i dodatna unapređenja modela. Ispitane su mogućnosti vernije reprezentacije govornog signala, pri čemu je analiza proširena na veliki broj obeležja iz različitih grupa. Formiranje velikih skupova obeležja nameće potrebu za redukcijom dimenzija, gde je pored poznatih metoda analizirana i alternativna metoda zasnovana na Fibonačijevom nizu brojeva. Na kraju su razmotrene mogućnosti integracije prednosti različitih pristupa u jedinstven sistem za automatsko prepoznavanje govornih emocija, tako da je predložena paralelna multiklasifikatorska struktura sa kombinatornim pravilom koje pored rezultata klasifikacije pojedinačnih klasifikatora ansambla koristi i informacije o karakteristikama klasifikatora. Takođe, dat je predlog automatskog formiranja ansambla klasifikatora proizvoljne veličine upotrebom redukcije dimenzija zasnovane na Fibonačijevom nizu brojeva.

Ključne reči: obrada govora, prepoznavanje emocija, klasifikatorski algoritmi, skriveni Markovljevi modeli, redukcija dimenzija, obeležja govornog signala.

Naučna oblast: Elektrotehnika i računarstvo.

Uža naučna oblast: Upravljanje sistemima i obrada signala.

Dissertation title: Optimization of automatic speech emotion recognition systems.

Abstract: The basis for the successful integration of emotional intelligence into sophisticated systems of artificial intelligence is the reliable recognition of emotional states, with the paralinguistic content of speech standing out as a particularly significant carrier of information regarding the emotional state of the speaker. In this paper, a comparative analysis of speech signal features and classification methods most often used for solving the task of automatic recognition of speakers' emotional states is performed, after which the possibilities for improving the performances of the systems for automatic recognition of speech emotions are considered. Discrete hidden Markov models were improved using the QQ plot for the purpose of determining the codevectors for vector quantization, and additional models improvements were also considered. The possibilities for a more faithful representation of the speech signal were examined, whereby the analysis was extended to a large number of features from different groups. The formation of big sets of features imposes the need for dimensionality reduction, where an alternative method based on the Fibonacci sequence of numbers was analyzed, alongside known methods. Finally, the possibilities for integrating the advantages of different approaches into a single system for automatic recognition of speech emotions are considered, so that a parallel multiclassifier structure is proposed with a combinatorial rule, which, in addition to the classification results of individual ensemble classifiers, uses information about classifiers' characteristics. A proposal is also given for the automatic formation of an ensemble of classifiers of arbitrary size by using dimensionality reduction based on the Fibonacci sequence of numbers.

Key words: speech processing, emotion recognition, classification algorithms, hidden Markov models, dimensionality reduction, speech features.

Scientific field: Electrical and Computer Engineering.

Scientific subfield: System Control and Signal Processing.

Sadržaj

| | | |
|-------|---|----|
| 1 | Uvod..... | 1 |
| 1.1 | Motivacija rada i predmet istraživanja..... | 2 |
| 1.2 | Struktura rada | 3 |
| 2 | Govor i emocije | 5 |
| 2.1 | Govor | 5 |
| 2.2 | Emocije | 6 |
| 2.3 | Emocije u govoru..... | 7 |
| 2.4 | Modeli emocija | 8 |
| 2.4.1 | Kategorički modeli | 8 |
| 2.4.2 | Dimenzionalni modeli..... | 9 |
| 3 | Obeležja govornog signala | 11 |
| 3.1 | Spektralna obeležja | 15 |
| 3.1.1 | LPCC..... | 15 |
| 3.1.2 | LFPC | 17 |
| 3.1.3 | MFCC..... | 19 |
| 3.2 | Kontinualna obeležja..... | 20 |
| 3.2.1 | Fundamentalna frekvencija | 21 |
| 3.2.2 | Energija | 21 |
| 3.2.3 | Formanti | 22 |
| 3.2.4 | Obeležja trajanja | 23 |
| 3.3 | Obeležja kvaliteta glasa..... | 23 |
| 3.3.1 | Džiter..... | 24 |
| 3.3.2 | Šimer | 24 |
| 3.3.3 | HNR | 24 |
| 3.3.4 | ZCR..... | 25 |
| 3.3.5 | Verovatnoća zvučnog govora..... | 25 |
| 3.4 | Nelinearna obeležja..... | 26 |
| 3.5 | Delta koeficijenti..... | 26 |
| 3.6 | Statistike kratkovremenih obeležja | 27 |

| | | |
|-------|--|----|
| 3.7 | Redukcija dimenzija..... | 29 |
| 4 | Tehnike klasifikacije | 31 |
| 4.1 | Skriveni Markovljevi modeli..... | 31 |
| 4.1.1 | Rešenje problema 1 | 33 |
| 4.1.2 | Rešenje problema 2..... | 35 |
| 4.1.3 | Rešenje problema 3..... | 36 |
| 4.1.4 | Vektorska kvantizacija..... | 41 |
| 4.2 | Metoda potpornih vektora | 42 |
| 4.2.1 | Linearni SVM..... | 42 |
| 4.2.2 | Linearni SVM sa mekanom marginom..... | 44 |
| 4.2.3 | Nelinearni SVM..... | 45 |
| 4.2.4 | Višeklasni SVM..... | 46 |
| 4.3 | Duboke neuralne mreže..... | 46 |
| 4.3.1 | Veštačke neuralne mreže..... | 46 |
| 4.3.2 | Konstrukcija DNN | 51 |
| 4.4 | Multiklasifikatorski sistemi | 52 |
| 4.4.1 | Ansambl klasifikatora | 52 |
| 4.4.2 | Multiklasifikatorske topologije | 53 |
| 4.4.3 | Kombinatorna pravila | 54 |
| 4.5 | Mere uspešnosti klasifikatorskih sistema | 55 |
| 5 | Analiza sistema za prepoznavanje govornih emocija | 56 |
| 5.1 | Baze emocionalnog govora | 58 |
| 5.1.1 | Baza srpskog emocionalnog govora | 58 |
| 5.1.2 | Baza poljskog emocionalnog govora..... | 59 |
| 5.1.3 | Baza ruskog emocionalnog govora..... | 59 |
| 5.1.4 | Baza nemačkog emocionalnog govora | 59 |
| 5.2 | Postavka eksperimenata | 60 |
| 5.2.1 | Koeficijenti uspešnosti..... | 61 |
| 5.3 | Rezultati testiranja | 62 |
| 5.3.1 | Uspešnost obeležja..... | 63 |
| 5.3.2 | Uspešnost klasifikatora | 64 |
| 5.3.3 | Uspešnost parova obeležje-klasifikator | 66 |
| 5.4 | Rezime analize..... | 67 |
| 6 | Optimizacija sistema za prepoznavanje govornih emocija | 68 |
| 6.1 | Unapređenje HMM klasifikatora..... | 68 |
| 6.1.1 | Vektorska kvantizacija upotrebom QQ krive | 69 |

| | | |
|-------|---|----|
| 6.1.2 | Dodatna unapređenja HMM..... | 74 |
| 6.1.3 | Analiza unapređenog HMM..... | 75 |
| 6.2 | Proširenje skupa obeležja i redukcija dimenzija..... | 77 |
| 6.2.1 | Nespektralna obeležja | 77 |
| 6.2.2 | Prošireni skup obeležja | 78 |
| 6.2.3 | Redukcija dimenzija..... | 80 |
| 6.3 | Konstrukcija multiklasifikator sistema | 83 |
| 6.3.1 | Ansambli preuzetih klasifikatorskih sistema..... | 86 |
| 6.3.2 | Ansambli klasifikatorskih sistema upotrebom Fibonači redukcije..... | 87 |
| 7 | Zaključak | 90 |
| 8 | Literatura | 93 |

Lista slika

| | |
|---|----|
| Sl. 2.1 – Delovi govornog aparata | 6 |
| Sl. 2.2 – Plutčikov „točak emocija” [29]..... | 9 |
| Sl. 2.3 – Raselov kružni model [31] | 10 |
| Sl. 3.1 – Frejmovanje govornog signala..... | 12 |
| Sl. 3.2 – <i>Hamming</i> prozor | 13 |
| Sl. 3.3 – Tretiranje frejma <i>Hamming</i> prozorom | 13 |
| Sl. 3.4 – Izdvajanje spektralnih obeležja..... | 14 |
| Sl. 3.5 – Banka pravougaonih filtera..... | 18 |
| Sl. 3.6 – Banka trougaonih filtera | 20 |
| Sl. 4.1 – Promena stanja HMM | 32 |
| Sl. 4.2 – Parametri SVM modela | 43 |
| Sl. 4.3 – Matematički model neurona | 47 |
| Sl. 4.4 – Konstrukcija DNN | 51 |
| Sl. 5.1 – Koeficijenti uspešnosti obeležja u SD postavci..... | 64 |
| Sl. 5.2 – Koeficijenti uspešnosti obeležja u SI postavci | 64 |
| Sl. 5.3 – Koeficijenti uspešnosti klasifikatora u SD postavci..... | 65 |
| Sl. 5.4 – Koeficijenti uspešnosti klasifikatora u SI postavci | 65 |
| Sl. 5.5 – Koeficijenti uspešnosti parova obeležje-klasifikator u SD postavci..... | 66 |
| Sl. 5.6 – Koeficijenti uspešnosti parova obeležje-klasifikator u SI postavci | 66 |
| Sl. 6.1 – Deo po deo linearna aproksimacija QQ krive (pristup 1)..... | 73 |
| Sl. 6.2 – Deo po deo linearna aproksimacija QQ krive (pristup 2)..... | 73 |
| Sl. 6.3 – Deo po deo linearna aproksimacija QQ krive (pristup 3)..... | 74 |
| Sl. 6.4 – Koeficijenti uspešnosti klasifikatora (proširenje: unapređeni HMM)..... | 76 |
| Sl. 6.5 – Koeficijenti uspešnosti parova obeležje-klasifikator (proširenje: unapređeni HMM)..... | 76 |
| Sl. 6.6 – Koeficijenti uspešnosti obeležja (proširenje: nespektralna obeležja) | 78 |
| Sl. 6.7 – Koeficijenti uspešnosti parova obeležje-klasifikator (proširenje: nespektralna obeležja) .. | 78 |
| Sl. 6.8 – Koeficijenti uspešnosti obeležja (proširenje: prošireni skup obeležja) | 79 |
| Sl. 6.9 – Koeficijenti uspešnosti parova obeležje-klasifikator (proširenje: prošireni skup obeležja) | 80 |
| Sl. 6.10 – Koeficijenti uspešnosti obeležja (proširenje: redukcija dimenzija) | 82 |
| Sl. 6.11 – Koeficijenti uspešnosti parova obeležje-klasifikator (proširenje: redukcija dimenzija) ... | 82 |

Lista tabela

| | |
|---|----|
| Tab. 3.1 – Statistike kratkovremenih obeležja | 28 |
| Tab. 5.1 – Sistemi za avtomatsko prepoznavanje govornih emocij testirani sa slovenskim bazama | 57 |
| Tab. 5.2 – Rezultati testiranja | 63 |
| Tab. 6.1 – Prosečna rasipanja za različito formirane kvantizacione etalone | 72 |
| Tab. 6.2 – Uspešnost klasifikacije – unapređeni HMM | 75 |
| Tab. 6.3 – Uspešnost klasifikacije – nespektralna obeležja | 77 |
| Tab. 6.4 – Uspešnost klasifikacije – prošireni skup obeležja | 79 |
| Tab. 6.5 – Uspešnost klasifikacije – redukcija dimenzija | 81 |
| Tab. 6.6 – Matrica konfuzije HMM klasifikatora, GEES baza | 84 |
| Tab. 6.7 – Matrica konfuzije SVM klasifikatora, GEES baza | 85 |
| Tab. 6.8 – Matrica konfuzije DNN klasifikatora, GEES baza | 85 |
| Tab. 6.9 – Uspešnost klasifikacije – multiklasifikator | 86 |
| Tab. 6.10 – Mere različitosti klasifikatora | 87 |
| Tab. 6.11 – Uspešnost klasifikacije – multiklasifikator (Fibonači redukcija) | 88 |
| Tab. 6.12 – Mere različitosti klasifikatora (Fibonači redukcija) | 89 |

1 Uvod

Svet u kakvom danas živimo je proizvod inteligencije. Zahvaljujući inteligenciji su stvorena neverovatna tehnološka dostignuća. Sledeće veliko tehnološko dostignuće kojem se teži jeste upravo stvaranje inteligencije. Već dugi niz godina intenzivno se radi na razvoju veštačke inteligencije i na taj način stvara potencijal koji može da pomogne civilizaciji kao nikada do sada.

Cilj istraživanja veštačke inteligencije jeste razvijanje algoritama koji mašinama omogućavaju da funkcionišu na način koji se može okarakterisati inteligentnim. Kada je reč o ljudskoj inteligenciji, generalni stav psihologa jeste da inteligencija predstavlja sposobnost prilagođenja okruženju, što uključuje veći broj kognitivnih procesa, kao što su percepcija, učenje, pamćenje, razmišljanje i rešavanje problema. Inteligencija kod mašina treba da omogući obavljanje zadataka i reagovanje poput ljudi, s tim u vezi, ulažu se naponi kako bi se i veštačkoj inteligenciji omogućilo da percipira, uči, planira, rešava probleme.

Već postoje mnogi zadaci u kojima se veštačka inteligencija pokazala uspešnijom od čoveka i to u oblastima koje deluju suviše složene, kao što je medicina, podučavanje, savetovanje. Međutim, dosadašnji uspesi sistema veštačke inteligencije se dominantno zasnivaju na iscrpnom pretraživanju ogromnog broja mogućnosti i uglavnom su vezani za uske oblasti primene. S obzirom na želju da veštačka inteligencija što vernije oslikava čovekovu inteligenciju, neophodno je uvođenje promena u pristupu kreiranja veštačke inteligencije.

Dosadašnja praksa uvođenja novih tehnoloških dostignuća u obavljanju različitih zadataka je podrazumevala da se ljudi obuče, tj. da se prilagode obavljanju zadataka upotrebom novih tehnologija. Sa stvaranjem sofisticiranih, naprednih tehnoloških dostignuća, kojima se sve više poveravaju delikatni aspekti ljudskih života, mora da se uvede adaptacija i u suprotnom smeru, tj. da se omogući prilagođenje mašina promenljivom ljudskom okruženju. Već je mnogo urađeno na polju interakcije čovek-mašina, gde su već prisutni uspešni sistemi za automatsko prepoznavanje prirodnog govora kao i za sintezu govora [1]. Međutim, u svakoj komunikaciji su vrednije informacije sadržane i izvan eksplicitnog saopštenja, što dovodi do pojma emocionalne inteligencije kao suštine mašina koje su mnogo opštijeg karaktera, robusnije, efikasnije, ali i usklađenije sa humanim vrednostima [2]. Sama činjenica da stepen emocionalne inteligencije često odlučuje o uspešnosti ljudi u obavljanju različitih poslova, kao i tvrdnje mnogih psihologa da emocije dominantno upravljaju većinom bitnih odluka u životu [3], [4], [5], su dovoljan razlog za neminovnost uključenja emocionalne inteligencije u sofisticirane sisteme veštačke inteligencije, kako bi im bila data šansa da obavljaju poverene zadatke jednako dobro kao ljudi. Emocionalna inteligencija podrazumeva sposobnost razumevanja osećanja i raspoloženja, kao i upravljanje ponašanjem u skladu sa njima, što u računarskim okvirima podrazumeva postojanje algoritama za prepoznavanje emocionalnih stanja, algoritama za razmatranje emocionalnih stanja u donošenju odluka, kao i algoritama za simulaciju emocionalnih stanja.

Kreiranje algoritama za simulaciju emocionalnog doživljaja nalik humanom nije jednostavan zadatak, ali je izvesno da svaki pomak na polju merenja, obrade i oponašanja emocionalnih znakova može pomoći veštačkoj inteligenciji da komfornije funkcioniše zajedno sa ljudima. S obzirom da je pouzdano prepoznavanje emocija osnov za pravilan rad sistema koji se bave ostalim aspektima emocionalne inteligencije, suštinski je bitno da prepoznavanje bude izvršeno sa jako visokim

stepenom pouzdanosti. Iako je zadatak prepoznavanja emocija relativno lak za čoveka, situacija nije tako jednostavna kada se zadatak postavi pred algoritam veštačke inteligencije. Posebno, postoje situacije u kojima ni ljudi ne interpretiraju iste nadražaje sa jednakim zaključkom, a često čak ni sam čovek ne može sa sigurnošću tvrditi o emocijama koje proživljava. Takođe, emocionalna stanja su uglavnom pomešana, što sve dodatno otežava zadatak automatskog prepoznavanja emocija.

Različiti su načini na koje se izražavaju emocije, pa samim tim i senzorski ulazi u sisteme za automatsko prepoznavanje emocionalnih stanja. Jedna od osnovnih manifestacija emocionalnih stanja jeste putem govora, pri čemu se razlikuju dva dela govora koji nose informacije o emocionalnim stanjima. Prvi je lingvistički, koji se odnosi na značenje osnovne poruke, a drugi je paralingvistički, koji podrazumeva implicitne poruke u govoru koje odražavaju emocionalna stanja govornika. Izrazi lica su, takođe, dominantni nosioci informacija o emocionalnim stanjima, kao i pokreti tela. Informacije o emocionalnim stanjima su sadržane i u fiziološkim znacima, kao što je puls, električna provodljivost kože, aktivnost mišića lica, glotalni impulsi. U pogledu dominantnosti pojedinačnih izvora, paralingvistički sadržaj govora se može smatrati najznačajnijim nosiocem emocionalnih informacija, s obzirom na zastupljenost verbalne komunikacije, stepen pouzdanosti, kao i mogućnost prikupljanja.

Od sredine devedesetih godina dvadesetog veka je emocionalna inteligencija počela da okupira pažnju mnogih istraživača u oblasti prepoznavanja oblika i mašinskog učenja. Okvire modernih istraživanja u oblasti emocionalne inteligencije je postavila Rozalinda Pikard 1995. godine [6]. Ubrzo zatim je Delart, zajedno sa saradnicima, objavio rad koji se smatra prvim na temu prepoznavanja emocija iz govora [7], iako je ideja automatskog prepoznavanja govornih emocija postojala i dosta ranije, sa prvim patentima objavljenim 1970-tih [8].

Napravljeni su veliki pomaci na polju automatskog prepoznavanja emocija iz govora od prvih eksperimenata do danas [9], ali svakako postoji prostor za dalja unapređenja. Ostvareni uspesi su već našli brojne primene u praktičnim aplikacijama. Jedna primena jeste u automatskom menadžmentu poziva u korisničkim servisima, gde se vrši efikasnije preusmeravanje poziva ukoliko se razmatraju emocionalna stanja korisnika [10]. Sledeća uspešna primena jeste u automatizovanim edukativnim sistemima [11]. Takođe, postoje primene u tretmanu poremećaja govora [12], za bezbedniju vožnju [13], [14], kao i u video igrama [15]. Prepoznavanje emocija može naći ulogu i u sistemima kojima detekcija emocionalnih stanje nije primaran cilj. U primenama kao što je prepoznavanje govora [16] ili govornika [17] prepoznato emocionalno stanje služi u svrhe adaptacije prilikom rešavanja primarnog zadatka.

1.1 Motivacija rada i predmet istraživanja

Stvaranje veštačke inteligencije otvara velike mogućnosti, gde se kao osnovno obeležje najvišeg nivoa veštačke inteligencije izdvaja emocionalna inteligencija. Sa druge strane, osnov za uspešnu realizaciju emocionalne inteligencije jeste pouzdana percepcija emocionalnih stanja. Motiv ovog rada jeste upravo optimizacija sistema za automatsko prepoznavanje govornih emocija.

U ranijim istraživanjima sistema za automatsko prepoznavanje govornih emocija su korišćeni različiti pristupi, gde se u svakom može razlikovati deo koji se bavi izdvajanjem obeležja govornog

signala i deo koji se bavi klasifikacijom izdvojenih obeležja [18]. Pri tome je neosporna sličnost sistema za automatsko prepoznavanje govornih emocija sa dosta duže istraživanim sistemima za automatsko prepoznavanje govora (izgovorenih reči). Mnoga obeležja, osmišljena za potrebe prepoznavanja govora, su preuzeta i korišćena u prepoznavanju govornih emocija. Takođe, mnoge metode, koje su korišćene za klasifikaciju govora, su u upotrebi prilikom prepoznavanja emocija. Iako je u pitanju isti izvor informacija, prepoznavanje govora i prepoznavanje govornih emocija se bave sasvim različitim aspektima govora, dok se jedno bavi izgovorenim sadržajem drugo se bavi načinom izgovora, što je za jedan zadatak korisna informacija za drugi je poremećaj i obrnuto. Zato se nameće potreba za ispitivanjem obeležja i klasifikatora, koji nisu specijalno dizajnirani za potrebe prepoznavanja govornih emocija, u pogledu podobnosti za upotrebu u sistemima automatskog prepoznavanja govornih emocija. Zarad izvođenja robusnijih zaključaka, korisno je testiranja izvršiti sa većim brojem baza emocionalnog govora, pri čemu je zgodno iskoristiti priliku za promovisanje slovenskih baza emocionalnog govora, kao manje istraživanih, a ujedno i poređenje sa dominantno korišćenim bazama. Rezultati analize mogu biti korisni kod izbora najprikladnije konfiguracije klasifikatorskog sistema, a takođe, mogu poslužiti kao osnova za razmatranje popravke analiziranih algoritama.

U sklopu istraživanja će biti sagledane performanse klasifikatorskih metoda zasnovanih na skrivenim Markovljevim modelima, metodi potpornih vektora i dubokim neuralnim mrežama, pri čemu će dodatan akcenat biti stavljen na popravku performansi skrivenih Markovljevih modela. Od obeležja govornog signala biće pre svega ispitivana spektralna obeležja, a potom će analiza biti proširena na kontinualna obeležja, obeležja kvaliteta glasa i nelinearna obeležja. Biće analizirane i metode za redukciju dimenzija vektora obeležja, gde će pored poznatih metoda biti ispitani i alternativni pristupi. Na kraju, biće razmotrene multiklasifikatorske strukture i mogućnosti ujedinjenja različitih pristupa.

1.2 Struktura rada

Ostatak rada je organizovan kroz nekoliko celina. Prvenstveno je dat uvod u problem automatskog prepoznavanja emocija iz govora, potom je izvršena analiza različitih elemenata sistema za automatsko prepoznavanje govornih emocija, a zatim su ispitane mogućnosti unapređenja.

U drugoj glavi je opisan proces nastajanja govora i predstavljen pojam emocija, a potom je opisana manifestacija emocija putem govora. U drugoj glavi su predstavljeni i osnovni modeli emocija.

U trećoj glavi je dat pregled obeležja koja mogu biti upotrebljena za reprezentaciju govornog signala, kao i metoda za određivanje statistika i redukciju dimenzija formiranog skupa obeležja.

Četvrta glava se bavi klasifikacionim tehnikama korišćenim u rešavanju zadatka automatskog prepoznavanja govornih emocija. Takođe, predstavljene su osnove multiklasifikatorskih sistema, kao i mere za određivanje kvaliteta klasifikatorskih sistema.

Peta glava je posvećena analizi kvaliteta obeležja i klasifikatora označenim kao dobar izbor za potrebe automatskog prepoznavanja govornih emocija. Ujedno je izvršena analiza uticaja korišćenih baza emocionalnog govora na rezultate testiranja.

Šesta glava se bavi optimizacijom sistema za automatsko prepoznavanje govornih emocija, sa ciljem postizanja što bolje performanse. Prvenstveno su analizirane mogućnosti unapređenja klasifikacione tehnike na bazi diskretnih skrivenih Markovljevih modela, pri čemu je predložena nova metoda vektorske kvantizacije, kao i nekoliko dodatnih unapređenja. U šestoj glavi su analizirane i mogućnosti proširenja skupa obeležja, kao i mogućnosti redukcije dimenzija proširenog skupa obeležja. Na kraju glave je razmotrena agregacija različitih pristupa u jedinstven multiklasifikatorski sistem.

U sedmoj glavi su sumirani zaključci rada i date preporuke za buduća istraživanja.

2 Govor i emocije

Odrastao čovek provede prosečno 70% vremena u komunikaciji, od čega 75% predstavlja verbalna komunikacija [19]. Komunikacija se može definisati kao razmena poruka između odašiljaoca i primaoca, pri čemu u verbalnoj komunikaciji, pored samih reči, učestvuju još dve komponente: promene glasa (ton glasa, boja glasa, brzina govora, izgovor, pauze u govoru) i pokreti tela (mimika i gestikulacija).

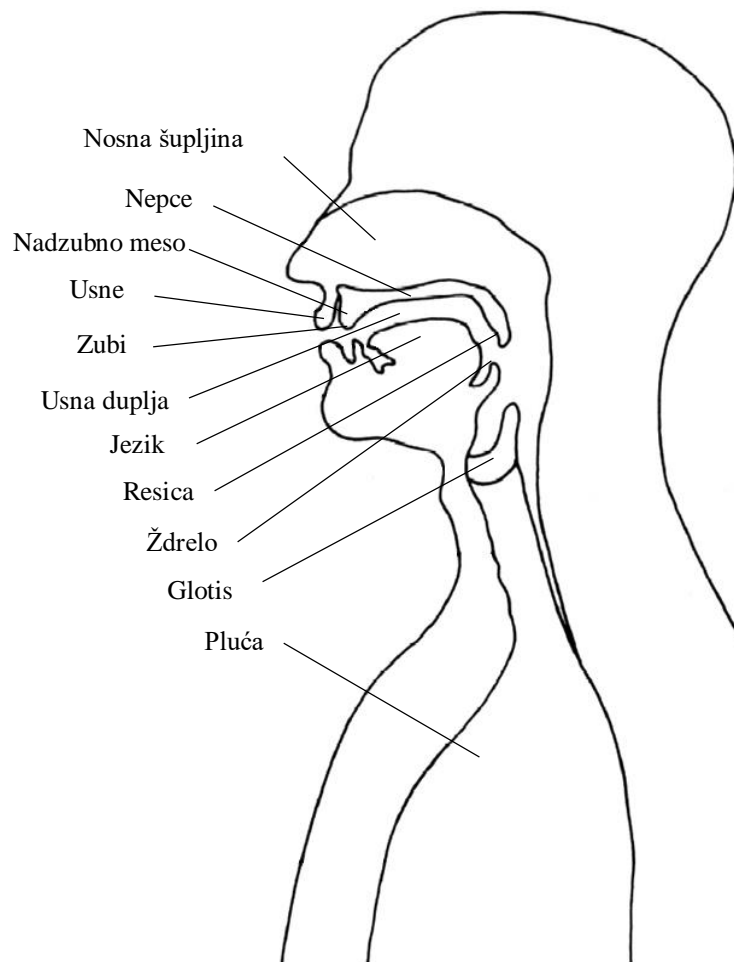
2.1 Govor

Prilikom verbalne komunikacije se u mozgu govornika formira apstraktna poruka, a zatim se generiše skup neuronskih signala koji upravljaju postupkom artikulacije. Govor nastaje kada se vazduh izdahne iz pluća uz istovremeno pomeranje artikulacionog aparata. Čovekov govorni aparat je sposoban da proizvede veoma složene i raznovrsne zvuke zahvaljujući sinhronizovanom delovanju većeg broja organa, od kojih su neki izvor zvuka, dok su drugi zaduženi za njegovo oblikovanje. Izvor zvuka su pluća i glotis, pri čemu glotis čine glasne žice (glasnice) i prostor između njih. Za oblikovanje zvuka su zaduženi organi vokalnog trakta: ždrelo, usna duplja, jezik, zadnje (meko) nepce, srednje nepce, prednje (tvrdo) nepce, nadzubno meso, zubi, usne, resica i nosna šupljina sa nosnicama (Sl. 2.1) [20].

Organi govornog aparata imaju ulogu i u mehanizmima van produkcije govora, kao što je disanje, ishrana i čulna uloga, pa je neophodna sinhronizacija funkcija. Energija potrebna za stvaranje zvuka dolazi iz pluća kao sastavni deo mehanizma disanja, tako što se izdahnuti vazduh koristi za govor. Neophodno je usklađenje disanja sa tipom govora koji se stvara, pri čemu se razlikuje zvučni i bezvučni govor. Prilikom produkcije zvučnog govora se vrši spajanje glasnih žica, privremeno zaustavljajući protok vazduha iz pluća, što dovodi do povećanja pritiska, a kada pritisak premaši otpornost glasnica, one se otvaraju dozvoljavajući vazduhu da prođe dalje. Glasnice se zatim brzo zatvaraju zahvaljujući uticaju više faktora, kao što je njihova elastičnost, napetost mišića i Bernulijev efekat. Ako se nastavi ravnomerno dopremanje vazduha pod pritiskom, glasnice će nastaviti da se pseudo periodično otvaraju i zatvaraju, stvarajući povorku vazдушnih impulsa u vokalnom traktu. Frekvencija vazдушnih impulsa određuje fundamentalnu frekvenciju (pič frekvenciju) stvorenog zvuka. Ukoliko se ne obezbeđuje konstantan dotok vazduha pod pritiskom, stvara se stohastičan vazdušni talas, što karakteriše bezvučan govor.

Vokalni trakt provodi vazdušni talas od glotisa do spoljašnje sredine, vršeći pri tome spektralno bojenje. Vokalni trakt se ponaša kao rezonantna cev, gde njegovi različiti delovi deluju u sinhronizmu kako bi proizveli odgovarajući zvuk, pri tome potiskujući određene frekvencije, a

ističući druge. Frekvencije koje vokalni trakt najviše ističe predstavljaju frekvencije formanata, pri čemu su centralne frekvencije i opsezi formanata određeni dužinom i zapreminom vokalnog trakta.



Sl. 2.1 – Delovi govornog aparata

2.2 Emocije

Emocije su dugo bile predmet inspiracije filozofa, pisaca, slikara, glumaca, dok sa psihološke strane nisu posebno izučavane, što se može objasniti širinom, raznovrsnošću i neodređenošću pojma emocija. S obzirom da su emocije u osnovi svakog ljudskog iskustva i adaptacije, neminovno je bilo pozabaviti se emocijama i sa naučne strane. Darwin je sproveo prve naučne studije emocija [21], a ubrzo zatim je Viljem Džejms objavio teoriju prema kojoj se emocije javljaju kao osećaji telesnih promena koje su direkta posledica emocionalnih stimulusa [22]. Prema drugim teorijama se kognicija stavlja na prvo mesto, gde se emocije razmatraju kao posledica misaonih procesa [23]. Prema sledećoj

grupi teorija, emocije su oblikovane socijalnom notom i zasnivaju se na naučenim društvenim pravilima [24].

Generalno prihvaćena definicija jeste da su emocije epizode koordiniranih promena u više aspekata kao odgovor na unutrašnje i spoljašnje događaje od većeg značaja za jedinku [25]. Manifestacija emocionalnih promena se može podeliti u tri komponente: fiziološka, izražajna i subjektivna. Fiziološka komponenta podrazumeva promene koje se odnose na ubrzan rad srca, ubrzano disanje, pojačano lučenje hormona, napetost mišića, širenje zenica. Izražajna komponenta se odnosi na spolja vidljive manifestacije u vidu promena u ponašanju, stavu tela, gestovima, mimici, boji glasa. Subjektivna komponenta se odnosi na lični doživljaj osećaja, koji je uvek usmeren ka nekom objektu. Što se tiče emocionalnih stimulusa, spoljni stimulusi mogu biti promena u situaciji, odgovor na drugu osobu i slično, dok unutrašnji stimulusi mogu biti u obliku misli, sećanja i osećaja.

Emocije pripadaju opštijem pojmu afektivnih stanja. U širem smislu, afekat pokriva širok raspon osećaja različitog intenziteta i trajanja, kao što su emocije, raspoloženja, stavovi, karakteristike ličnosti. Afekat u užem smislu je emocionalni doživljaj velikog intenziteta, veoma burnog toka, kratkog trajanja, sa izraženim telesnim i psihičkim promenama. Emocije predstavljaju osećaje ograničenog trajanja, koja su uvek usmerena ka nekome ili nečemu, za razliku od raspoloženja, koja ne moraju imati očigledan uzrok i uglavnom su slabijeg intenziteta i dužeg trajanja. Stavovi su relativno dugotrajna emocionalno obojena uverenja prema nekom predmetu ili osobi. Karakteristike ličnosti su emocionalno obojene karakteristike i obrasci ponašanja svojstveni za neku osobu.

Pored adaptivne uloge, prilikom čega emocionalni doživljaji pripremaju telo na ono što sledi, emocije imaju i bitnu komunikacionu ulogu. Emocija besa šalje poruku okruženju da odvijanje situacije nije u skladu sa željama jedinke. Prepoznavanje straha saopštava informaciju o ugroženosti jedinke. Ekspresija tuge šalje poziv za utehu i pomoć. Iskazivanje gađenja upozorava da je opaženo nešto neprijatno.

2.3 Emocije u govoru

Govor ima vrlo bitnu ulogu u procesu komunikacije među ljudima, pri čemu se razlikuju dve komponente govora: lingvistička i nelingvistička. Lingvistički sadržaj se odnosi na eksplicitnu (semantičku) poruku, dok se nelingvistička komponenta odnosi na implicitne (dugoročno i kratkoročno prisutne) informacije u govoru, kao što je pol, starost i emocionalno stanje govornika. Eksplicitni sadržaj poruke može, ali ne mora, sadržati informacije o emocionalnom stanju govornika, ili može sadržati informacije koje nisu tačne, zato se implicitna poruka smatra vrednijom sa aspekta prepoznavanja emocionalnih stanja govornika.

Tokom emocionalnog doživljaja dolazi do niza promena u organizmu jedinke, što za posledicu ima odstupanje određenih funkcija u odnosu na neutralno stanje. Pod uticajem emocionalnih stanja sreće, besa i straha dolazi do uzbuđivanja nervnog sistema, praćenim povišenim pulsom, povišenim krvnim pritiskom, ubrzanim tempom disanja, suvoćom usta i grčevima mišića, što rezultira glasnim, brzim govorom sa velikim rasponom pič frekvencije. Sa druge strane, pod uticajem emocionalnog stanja tuge dolazi do smanjenja pulsa i krvnog pritiska, dok se lučenje pljuvačke

povećava, što uzrokuje da je govor spor, sa niskom pič frekvencijom i malom energijom. Zaključak bi bio da se mogu prepoznati obrasci promena akustičkih parametara specifični za različita emocionalna stanja, a koji se javljaju kao posledica uticaja fizioloških promena na disanje, izgovor i artikulaciju.

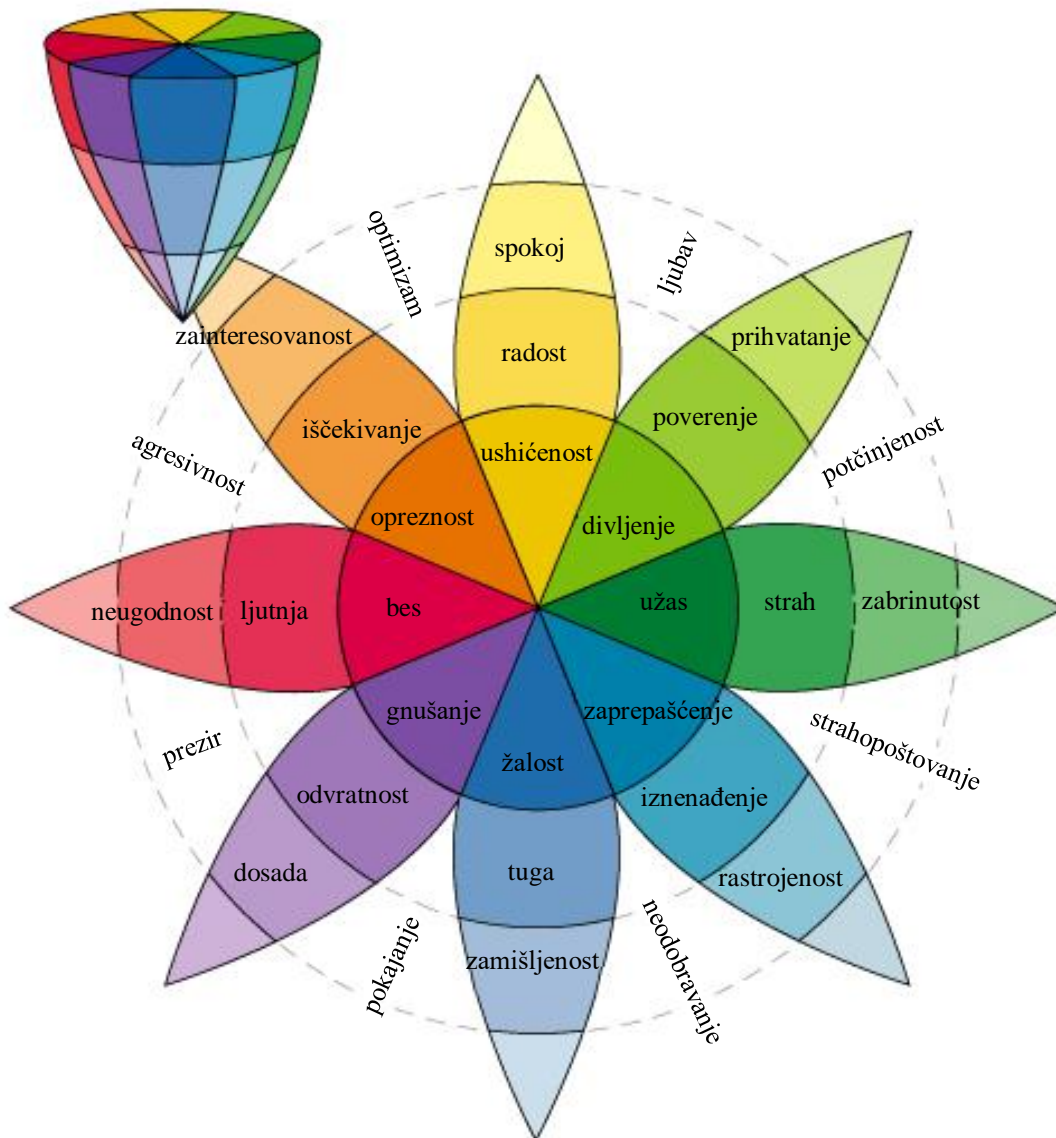
Sprovedeno je više studija koje su se bavile uticajem emocionalnih stanja na promene parametara govora, kao što su osnovna frekvencija, formanti, energija i parametri trajanja [26], [27]. Na osnovu sprovedenih istraživanja se došlo do načelnih pravila o promenama parametara govornog signala u zavisnosti od emocionalnog stanja, što je dovoljan pokazatelj da je moguće vršiti automatsko prepoznavanje emocionalnih stanja na osnovu govornog signala.

2.4 Modeli emocija

Definisanje broja emocionalnih stanja i granice između njih je vrlo važno sa aspekta njihove klasifikacije. Bilo je različitih pokušaja definisanja modela emocija sa oprečnim pristupima, među kojima su najznačajniji kategorički i dimenzionalni modeli.

2.4.1 Kategorički modeli

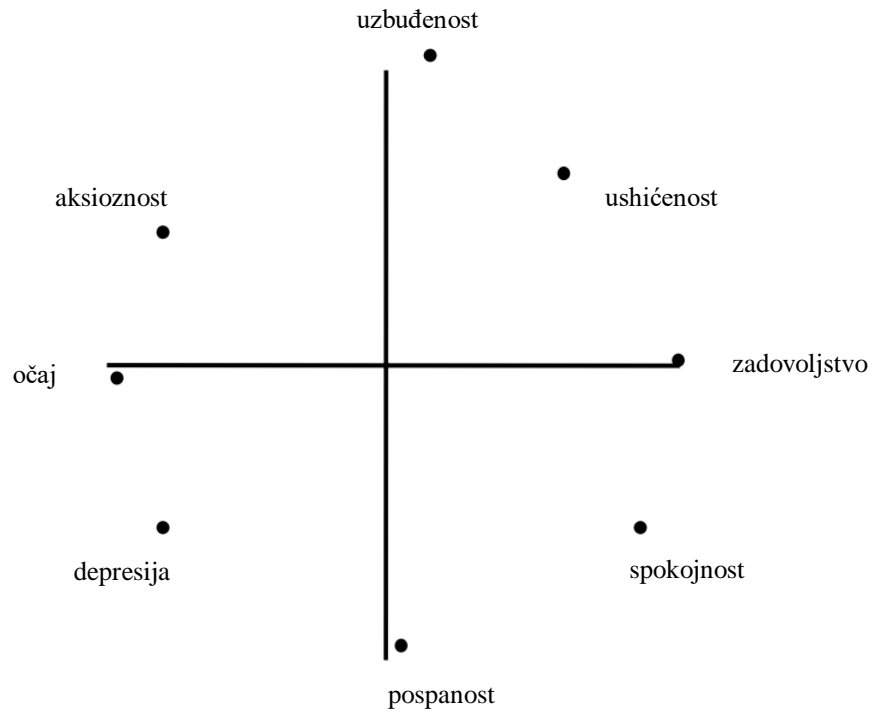
Kategorički modeli podrazumevaju diskretan skup jasno definisanih bazičnih emocionalnih stanja. Model koji je predložio Šerer [28] implicira paletu emocija, pri čemu se svaka emocija može dobiti kombinovanjem osnovnih emocija. Sličan model je ponudio i Plutčik [29], smatrajući da postoji osam osnovnih emocija, kao biološki primarnih, a da se ostale emocije izvode njihovim mešanjem. Dat je i predlog "točka emocija" koji oslikava bliskost različitih emocionalnih stanja (Sl. 2.2). Najčešće korišćen kategorički model je predložio Ekman [30] na osnovu istraživanja facijalnog izražavanja emocionalnih stanja u različitim kulturama, pri čemu se došlo do predloga šest osnovnih emocija: strah, bes, tuga, sreća, gađenje i iznenađenje.



Sl. 2.2 – Plučikov „točak emocija” [29]

2.4.2 Dimenzionalni modeli

Dimenzionalni modeli se baziraju na određivanju dimenzija kao pojednostavljenih svojstava emocionalnih stanja i definisanju položaja emocija u tako određenom prostoru. Najčešće se predlaže upotreba dve do tri dimenzije, opisane osama valentnosti, aktivacije i snage. Valentna osa daje informaciju o tome koliko je pozitivna ili negativna neka emocija, odnosno, koliko je emocionalno stanje prijatno ili neprijatno. Aktivaciona osa daje informaciju o intenzitetu emocije, od pospanosti do izrazitog uzbuđenja. Osa snage odnosi se na mogućnost kontrole emocija i omogućava razdvajanje emocija koje potiču od same jedinice i onih koje su posledica spoljašnjih faktora. Često se treća osa izostavlja, tako da je Rasel [31] predložio dvodimenzionalni model sa kružnom rasporedom emocija u prostoru valentnosti i aktivacije (Sl. 2.3).



Sl. 2.3 – Raselov kružni model [31]

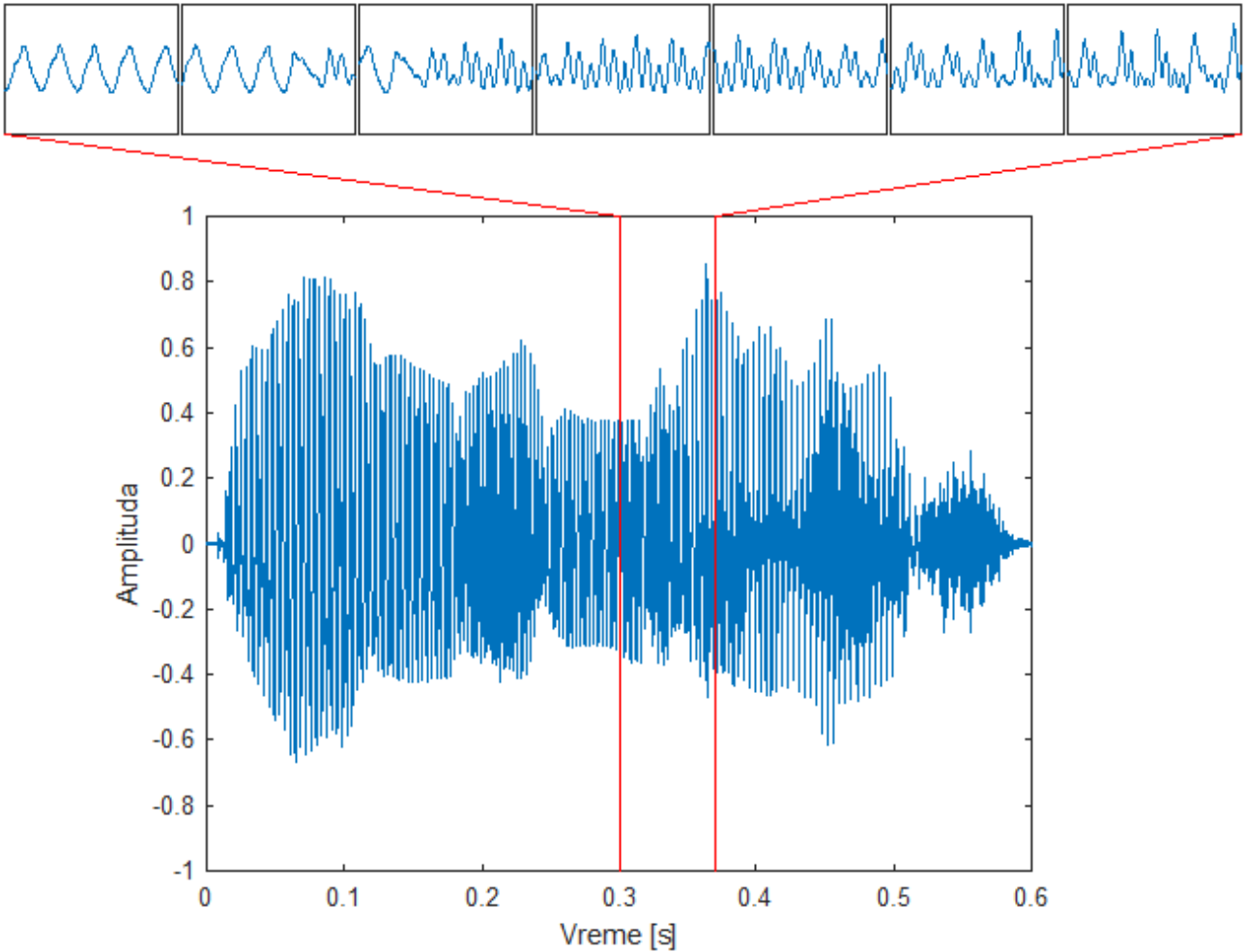
Svaki od modela ima svoje prednosti i mane. Kategorički modeli su lako interpretabilni, dok dimenzionalni omogućavaju predstavljanje proizvoljno mnogo emocionalnih stanja. U istraživanjima je najčešće u upotrebi Ekmanov kategorički model sa šest osnovnih emocija, a u zavisnosti od primene se razmatra manje ili više bazičnih emocija.

3 Obeležja govornog signala

Neosporno je da su u snimljenom govornom signalu sadržane informacije o emocionalnom stanju govorniku, pri čemu je neizostavni deo i lingvistička poruka, kao i informacije svojstvene za govornika, poput boje glasa, pola, starosti, ali i karakteristike prenosnog kanala. Klasifikacija emocionalnih stanja upotrebom neizmenjenog govornog signala ne bi bila efikasna, stoga je potrebno izvršiti izdvajanje obeležja (karakteristika) govornog signala koja što vernije oslikavaju emocionalna stanja uz potiskivanje ostalih sadržaja. Postavljeni problem nije trivijalan, sa poznatim i opšteprihvaćenim rešenjem, a sa druge strane, uspešnost klasifikacije emocionalnih stanja direktno zavisi od kvaliteta obeležja, tako da obezbeđenje adekvatnih obeležja govornog signala zahteva posebnu pažnju. Istraživanja u oblasti psihologije i psihoakustike ističu povezanost emocionalnih stanja sa promenama u fundamentalnoj frekvenciji govora, energiji i različitim obeležjima trajanja, ali dobijeni rezultati nisu dovoljni za uspešno rešavanje problema automatskog prepoznavanja emocionalnih stanja, a takođe, zaključci različitih istraživanja nisu uvek istovetni [32], [26], [27].

Uspešno izdvajanje obeležja govornog signala podrazumeva pronalaženje odgovora na niz otvorenih pitanja. Prvo pitanje jeste odabir govorne jedinice na nivou koje će se vršiti razmatranje obeležja, gde su dominantna dva pristupa. Prema prvom pristupu se podrazumeva analiza nad malim jedinicama od nekoliko desetina milisekundi – frejmovima, dok je prema drugom pristupu analizu potrebno vršiti na nivou govorne sekvence. Obeležja govornog signala izdvojena na nivou frejma se nazivaju lokalnim ili kratkovremenim, dok se obeležja izdvojena na nivou govorne sekvence nazivaju globalnim. Iako je govorni signal brzopromenljiv i nestacionaran, ukoliko se podela govorne sekvence vrši na frejmove dovoljno kratkog trajanja, može se pretpostaviti aproksimativna stacionarnost govornog signala unutar takve govorne jedinice, pa je moguće primeniti različite tehnike obrade stacionarnih signala. Globalna obeležja se formiraju direktnom analizom kompletne govorne sekvence ili određivanjem statistika kratkovremenih obeležja. Upotrebom globalnih obeležja je cela govorna sekvenca opisana jednom vrednošću, pri čemu se pretpostavlja trajanje emocionalnog naboja jednakog intenziteta tokom cele sekvence, odnosno, gubi se informacija o vremenskom poretku, ali bez obzira na navedeno, mnoga istraživanja podržavaju tvrdnju da globalna obeležja obezbeđuju dobre rezultate [33], [34]. Izdvajanje obeležja govornog signala je moguće vršiti i na nivou segmenata (sačinjenih od nekoliko susednih frejmova), slogova, reči ili zvučnih segmenata.

Sva kratkovremena obeležja u ovom radu su izdvajana na nivou frejma, pri čemu je potrebno definisanje vremenskog trajanja jednog frejma. Ukoliko je trajanje prekratko, u razmatranje se uzima premalo odbiraka govornog signala, što se odražava na kvalitet procene obeležja, sa druge strane, ukoliko je trajanje predugo, prestaje da važi pretpostavka o statističkoj stacionarnosti signala. Praksa je usvajanja družine frejma za problem prepoznavanja emocija od 15 do 25 milisekundi, a u ovom radu je usvojeno trajanje jednog frejma od 16 milisekundi sa preklapanjem susednih frejmova od devet milisekundi. Postupak formiranja frejmova je ilustrovan na Sl. 3.1, gde je prikazan tipičan talasni oblik govornog signala i niz frejmova koji odgovara jednom delu signala.

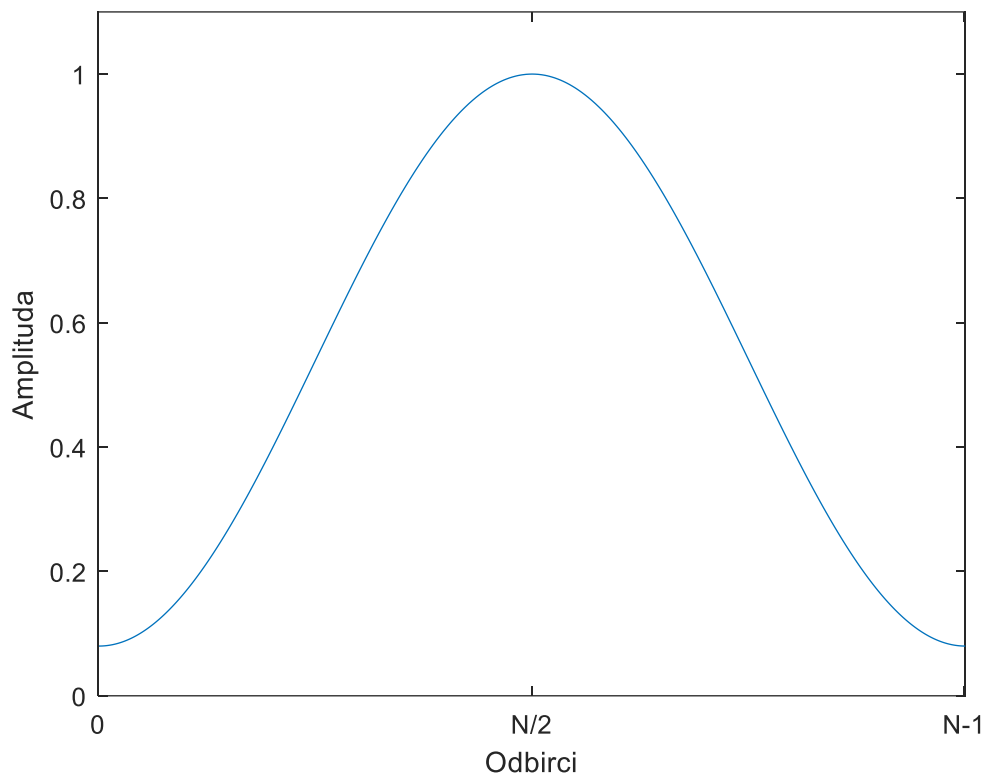


Sl. 3.1 – Frejmovanje govornog signala

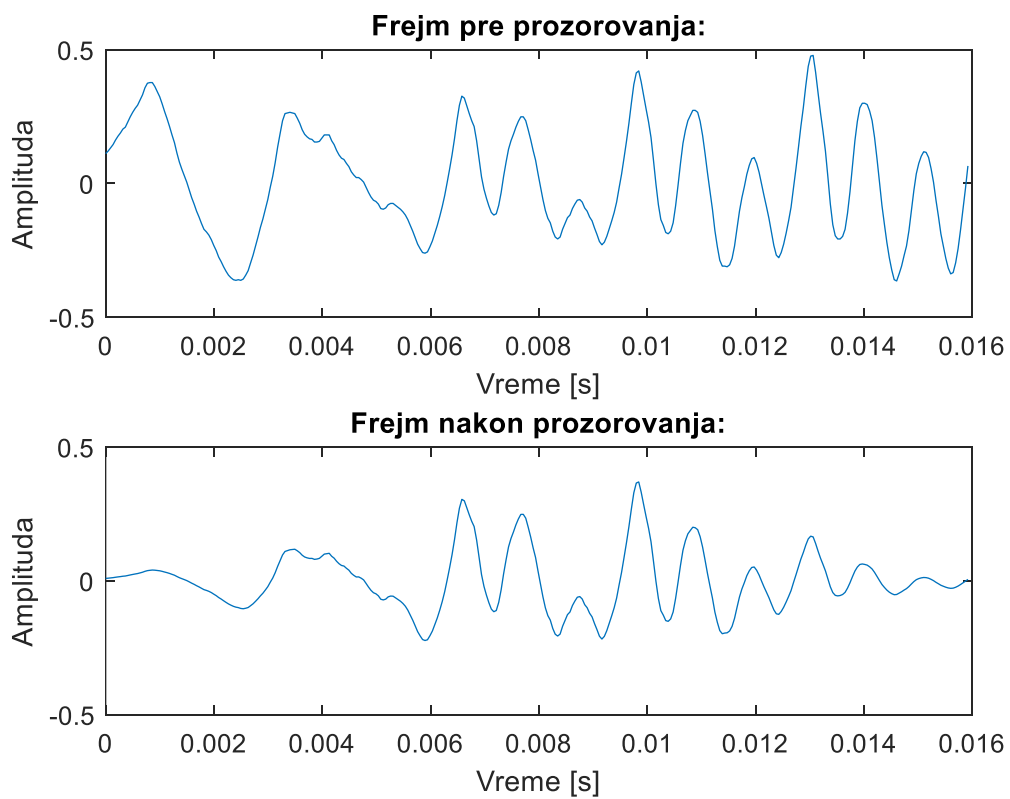
Kako bi se izbegli nepoželjni efekti diskontinuiteta na početku i kraju frejma potrebno je istaći centralne odbirke, tako da se svaki frejm tretira *Hamming* prozorom, definisanim izrazom:

$$w(n) = 0.54 - 0.46 \cdot \cos\left(2\pi \frac{n}{N-1}\right), \quad n = 0, \dots, N-1, \quad (3.1)$$

gde je N broj odbiraka u frejmu. Talasni oblik *Hamming* prozora je dat na Sl. 3.2, a tretiranje frejma prozorskom funkcijom je ilustrovano na Sl. 3.3.



Sl. 3.2 – Hamming prozor



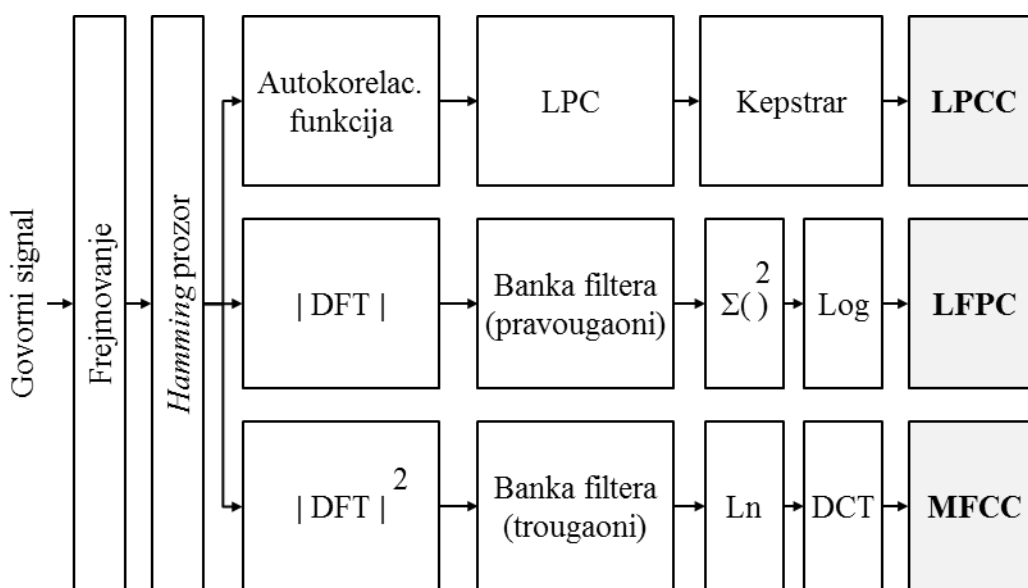
Sl. 3.3 – Tretiranje frejma Hamming prozorom

Izuzetno je važan odabir obeležja govornog signala koja će efikasno odražavati emocionalno stanje, a potiskivati irelevantne informacije. Mnoštvo obeležja je izučavano, ali se nije došlo do univerzalnog rešenja, pri tome je akcenat ispitivanja prevashodno bio na akustičkim obeležjima, a u nekim studijama su kao dopuna korišćena i lingvistička obeležja [35].

U literaturi se mogu naći različite kategorizacije akustičkih obeležja govornog signala, koje često nisu konzistentne, pri čemu grupisanje zavisi od razmatranih osobina, a granice između kategorija nisu striktno, ali to nije suštinski bitno kod projektovanja sistema za automatsko prepoznavanje emocionalnih stanja. Jedna moguća kategorizacija obeležja govornog signala jeste u sledeće četiri grupe [36]:

1. Spektralna obeležja: LPCC (eng. *Linear Prediction Cepstral Coefficients*), MFCC (eng. *Mel Frequency Cepstral Coefficients*), LFPC (eng. *Log Frequency Power Coefficients*);
2. Kontinualna obeležja: fundamentalna frekvencija, energija, formanti, obeležja trajanja;
3. Obeležja kvaliteta glasa: džiter (eng. *jitter*), šimer (eng. *shimmer*), HNR (eng. *Harmonic to Noise Ratio*), ZCR (eng. *Zero Crossing Rate*), verovatnoća zvučnog govora;
4. Nelinearna obeležja: Tajger (eng. *Teager*) energija.

Nakon izdvajanja obeležja govornog signala je moguće vršiti normalizaciju, što umanjuje uticaj prenosnog kanala i obezbeđuje ravnomeran uticaj obeležja sa različitim opsezima promene. Moguće je vršiti normalizaciju u odnosu na srednju vrednost i varijansu, uz napomenu da i varijansa može sadržati vredne informacije. Drugi način za sprovođenje normalizacije jeste svodenjem svih obeležja na predefinisani opseg promene.



Sl. 3.4 – Izdvajanje spektralnih obeležja

3.1 Spektralna obeležja

Spektralna obeležja se zasnivaju na kratkovremenoj spektralnoj analizi govornog signala. U nastavku će biti opisane implementacije spektralnih obeležja korišćenih u ovom radu: LPCC, LFPC i MFCC. Blok dijagram izdvajanja spektralnih obeležja je dat na Sl. 3.4.

3.1.1 LPCC

LPCC [37], [20] se formiraju u dva koraka, gde se u prvom formiraju LPC (eng. *Linear Prediction Coefficients*) na osnovu odbiraka govornog signala, a u drugom koraku se LPC koriste za računanje LPCC.

Osnovna ideja na kojoj je zasnovan LPC model jeste da se odbirak $x(n)$ u datom trenutku n može aproksimirati linearnom kombinacijom prethodnih p odbiraka:

$$x(n) \approx \hat{x}(n) = a_1x(n-1) + a_2x(n-2) + \dots + a_px(n-p), \quad (3.2)$$

gde je $\hat{x}(n)$ predikcija odbirka $x(n)$, a LPC a_1, a_2, \dots, a_p se smatraju konstantnim za dati frejm. Greška predikcije se može odrediti kao:

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{i=1}^p a_ix(n-i), \quad (3.3)$$

a transformacijom u z -domen se dobija:

$$E(z) = X(z) - \sum_{i=1}^p a_iX(z)z^{-i} = X(z) \left[1 - \sum_{i=1}^p a_iz^{-i} \right] = X(z)A(z), \quad (3.4)$$

tako da je:

$$A(z) = 1 - \sum_{i=1}^p a_iz^{-i}. \quad (3.5)$$

Na osnovu jednakosti (3.4) se može zapisati:

$$X(z) = E(z) \frac{1}{A(z)}, \quad (3.6)$$

tj. govorni signal se može posmatrati kao izlaz IIR (eng. *Infinite Impulse Response*) filtera sa funkcijom prenosa:

$$\frac{1}{A(z)}, \quad (3.7)$$

na čiji ulaz se dovodi signal greške aproksimacije $e(n)$.

Potrebno je odrediti koeficijente a_1, a_2, \dots, a_p tako da se što vernije modeluje govorni signal, odnosno, tako da se minimizuje greška. Optimizacija se može sprovesti primenom metode najmanjih kvadrata, pri čemu se ukupna greška može izraziti kao:

$$E = \sum_{n=n_0}^{n_1} e^2(n), \quad (3.8)$$

gde su n_0 i n_1 granice u kojima se vrši optimizacija. Ukoliko se greška $e(n)$ izrazi preko signala $x(n)$, kao u izrazu (3.3), može se zapisati:

$$E = \sum_{n=n_0}^{n_1} \left[x(n) - \sum_{i=1}^p a_i x(n-i) \right]^2. \quad (3.9)$$

Kako bi se izvršila minimizacija E u odnosu na LPC, potrebo je odrediti parcijalne izvode u odnosu na a_k , $k = 1, \dots, p$, i izjednačiti ih sa nulom. Optimalni koeficijenti se dobijaju kao rešenja sistema jednačina:

$$\frac{\partial E}{\partial a_k} = 0, \quad k = 1, \dots, p, \quad (3.10)$$

odnosno,

$$\sum_{n=n_0}^{n_1} x(n-k)x(n) = \sum_{i=1}^p a_i \sum_{n=n_0}^{n_1} x(n-k)x(n-i), \quad k = 1, \dots, p. \quad (3.11)$$

U prethodnoj jednačini je moguće uočiti postojanje korelacionog člana:

$$R(k, i) = \sum_{n=n_0}^{n_1} x(n-k)x(n-i). \quad (3.12)$$

Tako da se sistema jednačina (3.11) može zapisati kao:

$$R(k, 0) = \sum_{i=1}^p a_i R(k, i), \quad k = 1, \dots, p, \quad (3.13)$$

što predstavlja sistem od p jednačina sa p nepoznatih. Kako bi se došlo do rešenja prethodnog sistema, neophodno je prvenstveno odrediti $R(k, i)$ za $k = 1, \dots, p$ i $i = 0, \dots, p$. U zavisnosti od načina određivanja korelacije, razlikuju se autokorelaciona i kovarijantna metoda, pri čemu druga metoda obezbeđuje bolje iskorišćenje dostupnih odbiraka. U ovom radu će biti korišćena modifikovana kovarijantna metoda, tako da se stepen korelacije određuje na osnovu izraza:

$$R(k, i) = \frac{1}{2(N-p)} \left(\sum_{n=p}^{N-1} x(n-k)x(n-i) + \sum_{n=0}^{N-1-p} x(n+k)x(n+i) \right), \quad 0 \leq k, i \leq p, \quad (3.14)$$

gde je N broj odbiraka u frejmu, a zamenom u (3.13) i rešavanjem sistema jednačina se dolazi do vrednosti LPC za dati frejm. Ovako određeni LPC se mogu direktno koristiti u analizi govornog signala, ali se izvođenje LPCC pokazalo kao efikasnije rešenje.

Kepstralna analiza je proistekla iz ideje da se govorni signal može posmatrati kao pobudni signal glotisa oblikovan filterom vokalnog trakta, pri čemu je uloga kepstralne analize da omogući razdvajanje karakteristika pobude i filtera. Procedura određivanja kepstralnih koeficijenata se ostvaruje prevođenjem signala u z -domen, primenom logaritamske funkcije, a potom primene inverzne z -transformacije. Ukoliko je z -transformacija pobudnog signala $U(z)$, funkcija prenosa filtera $F(z)$ i z -transformacija izlaznog signala $X(z)$, tada važi:

$$X(z) = U(z)F(z), \quad \log(X(z)) = \log(U(z)) + \log(F(z)), \quad (3.15)$$

$$Z^{-1}(X(z)) = Z^{-1}(U(z)F(z)), \quad Z^{-1}(\log(X(z))) = Z^{-1}(\log(U(z))) + Z^{-1}(\log(F(z))), \quad (3.16)$$

gde je Z^{-1} inverzna z -transformacija. Može se zaključiti da primena kepstralne analize omogućava izražajnije isticanje karakteristika pobude i filtera.

Umesto direktne primene nad govornim signalom, kepstralna analiza se može sprovesti upotrebom LPC reprezentacije govornog signala, pri čemu je moguće izbeći prevođenje u z -domen upotrebom rekurzivnih formula:

$$c_0 = R_{0,0}, \quad (3.17)$$

$$c_m = a_m + \sum_{k=1}^{m-1} \binom{k}{m} c_k a_{m-k}, \quad 1 \leq m \leq p, \quad (3.18)$$

$$c_m = \sum_{k=m-p}^{m-1} \binom{k}{m} c_k a_{m-k}, \quad p < m \leq M - 1, \quad (3.19)$$

gde je p broj LPC, a M željeni broj LPCC. U ovom radu će biti određivano $p = 8$ LPC, a potom će se izvoditi $M = 13$ LPCC, da bi se u daljoj analizi koristili LPCC od c_1 do c_{12} .

3.1.2 LFPC

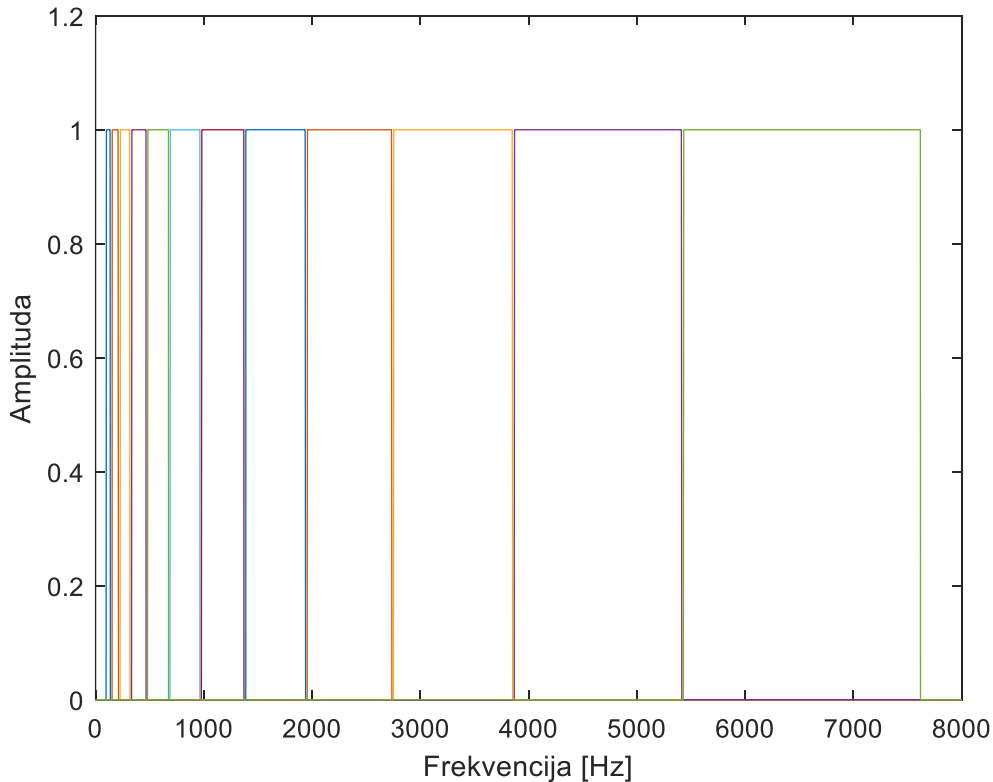
LFPC [38] prate ideju da ljudski slušni aparat nije jednako osetljiv na svim učestanostima, gde je osetljivost veća na nižim učestanostima, a sa porastom učestanosti osetljivost opada. LFPC treba da pruže informaciju o spektralnoj raspodeli energije prema opsezima koji odgovaraju kritičnim frekvencijskim opsezima ljudskog slušnog aparata.

Izdvađa se $M = 12$ frekvencijskih opsega upotrebom banke pravougaonih filtera, čije su centralne frekvencije i propusni opsezi određeni formulama:

$$b_m = a b_{m-1}, \quad 2 \leq m \leq M, \quad (3.20)$$

$$f_m = f_1 + \sum_{j=1}^{m-1} b_j + \frac{b_m - b_1}{2}, \quad (3.21)$$

pri čemu je f_m centralna frekvencija, a b_m propusni opseg m -tog filtera. Za centralnu frekvenciju i propusni opseg prvog filtera usvaja se $f_1 = 127$ Hz, $b_1 = 54$ Hz. Konstanta α definiše stepen logaritamskog porasta, a ujedno određuje i frekvencijski opseg govornog signala koji se razmatra. S obzirom da su relevantne informacije za emocionalno stanje prisutne i na višim učestanostima, usvojena je vrednost $\alpha = 1.4$, tako da je gornja granica analiziranog frekvencijskog sadržaja 7.6 kHz. Amplitudsko-frekvencijska karakteristika formirane banke filtera je data na Sl. 3.5.



Sl. 3.5 – Banka pravougaonih filtera

Postupak određivanja koeficijenata započinje prevođenjem signala $x(n)$, $n = 0, \dots, N - 1$ u frekvencijski domen, tj. određivanjem diskretne Furijerove transformacije (DFT):

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-\frac{i2\pi kn}{N}}, \quad k = 0, \dots, N - 1. \quad (3.22)$$

Filteri su opisani izrazom:

$$W_m(f) = \begin{cases} 1, & f_m - \frac{b_m}{2} \leq f \leq f_m + \frac{b_m}{2} \\ 0, & f < f_m - \frac{b_m}{2} \vee f > f_m + \frac{b_m}{2} \end{cases}, \quad (3.23)$$

pri čemu je $m = 1, \dots, M$, $f \in \{kF_s/N, k = 0, \dots, N/2\}$ i F_s učestanost odabiranja.

Energija koja odgovara m -tom frekvencijskom opsegu je određena izrazom:

$$S(m) = \sum_{f=f_m-\frac{b_m}{2}}^{f_m+\frac{b_m}{2}} (|X(f)|W_m(f))^2, \quad m = 1, \dots, M, \quad (3.24)$$

gde je $X(f)$ DFT spektralna komponenta koja odgovara učestanosti f .

Konačna mera energije m -tog frekvencijskog opsega je data sa:

$$SE(m) = \frac{10 \log_{10}(S(m))}{N_m}, \quad (3.25)$$

pri čemu je N_m broj spektralnih komponenti obuhvaćenih m -tim filterom.

3.1.3 MFCC

MFCC [39] su najčešće korišćena obeležja u obradi govornog signala, a predstavljaju proširenje koncepta kepra, sa težnjom bolje aproksimacije osetljivosti ljudskog slušnog aparata. S obzirom da osetljivost slušnog aparat nije ista na svim učestanostima linearne frekvencijske skale, u upotrebi je mel-frekvencijska skala na kojoj je osetljivost slušnog aparata jednaka duž cele skale. Mel-frekvencijska skala u odnosu na linearnu frekvencijsku skalu ima približno linearnu zavisnost na učestanostima do 1000 Hz, a logaritamsku na višim učestanostima, što modeluje veliku osetljivost slušnog aparata na nižim učestanostima, a slabiju sa porastom učestanosti.

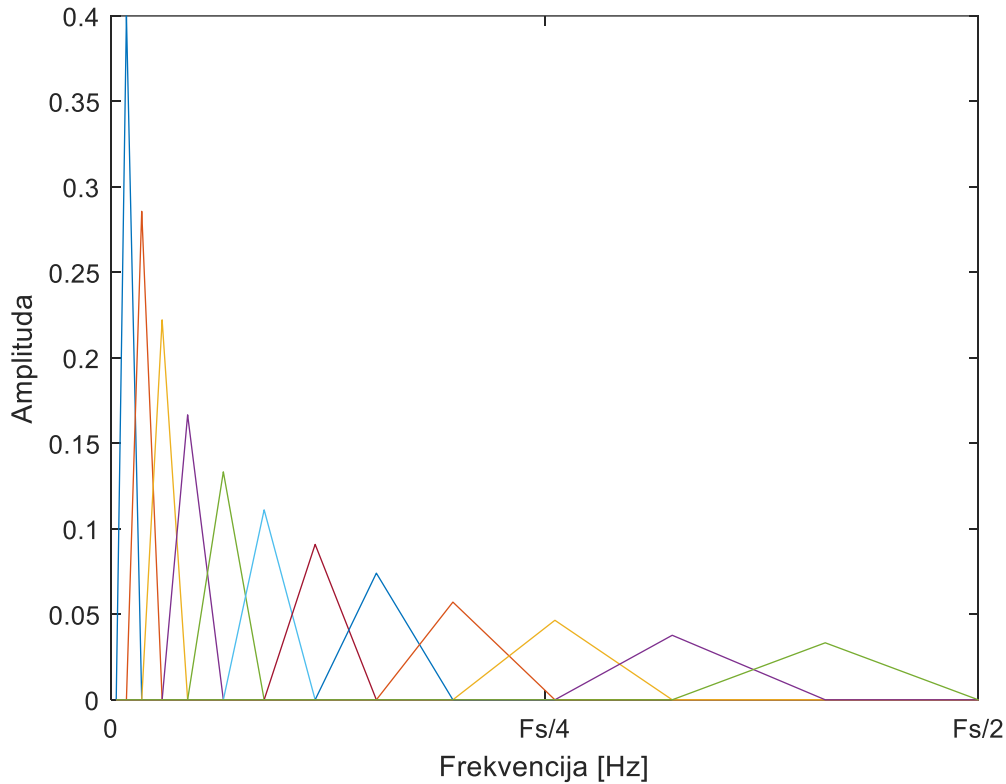
Kako bi se primenila prethodna ideja, formira se banka preklapajućih filtera jednako udaljenih na mel skali, gde je m -ti filter opisan jednakošću:

$$H_m(\phi) = \begin{cases} \frac{\phi - \phi_{b_{m-1}}}{\phi_{b_m} - \phi_{b_{m-1}}}, & \phi_{b_{m-1}} \leq \phi \leq \phi_{b_m} \\ \frac{\phi_{b_{m+1}} - \phi}{\phi_{b_{m+1}} - \phi_{b_m}}, & \phi_{b_m} \leq \phi \leq \phi_{b_{m+1}} \\ 0, & \phi < \phi_{b_{m-1}} \vee \phi > \phi_{b_{m+1}} \end{cases}, \quad m = 1, \dots, M, \quad (3.26)$$

U prethodnoj formuli je $M = 12$ ukupan broj filtera, a ϕ predstavlja diskretnu frekvenciju na mel skali. Granične frekvencije $\phi_{b_0}, \dots, \phi_{b_{M+1}}$ dele mel skalu na $M + 1$ jednakih frekvencijskih opsega, pri čemu maksimalna frekvencija na mel skali odgovara frekvenciji $F_s/2$ na linearnoj skali, gde je F_s učestanost odabiranja. Ovako formirana banka filtera se preslikava na linearnu skalu prema relaciji:

$$\phi = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right), \quad f \in \{kF_s/N, k = 0, \dots, N/2\}. \quad (3.27)$$

Potom se vrši normalizacija tako da se ostvari jedinična suma odbiraka svakog filtera i time banka filtera dobija konačnu formu $\bar{H}_m(k)$, $m = 1, \dots, M$ (Sl. 3.6).



Sl. 3.6 – Banka trougaonih filtera

Određivanje koeficijenata započinje prevođenjem govornog signala u frekvencijski domen upotrebom DFT transformacije, a potom se određuje spektar snage kvadriranjem amplitudskog spektra. Primenom prethodno formirane banke filtera na spektar snage se dobija mera raspodele energije na mel skali:

$$MS(m) = \sum_{k=0}^{N/2} |X(k)|^2 \bar{H}_m(k), \quad m = 1, \dots, M. \quad (3.28)$$

Na kraju, MFCC se dobijaju primenom logaritma, a potom i diskretne kosinusne transformacije:

$$MFCC(m) = \sum_{k=1}^M \ln(MS(k)) \cos(m(k - 1/2)\pi/M), \quad m = 1, \dots, M. \quad (3.29)$$

3.2 Kontinualna obeležja

Promene kontinualnih obeležja govora usled različitih emocionalnih stanja su izučavane u mnogim istraživanjima, sa posebnim akcentom na kontinualnim prozodijskim obeležjima poput

fundamentalne frekvencije, energije i parametara trajanja [26], [27]. Iako zaključci različitih studija nisu jednoznačni, neosporno je da su kontinualna obeležja značajni pokazatelji emocionalnih stanja govornika.

3.2.1 Fundamentalna frekvencija

Fundamentalna frekvencija (osnovna učestanost ili pič frekvencija) predstavlja frekvenciju vibracije glasnica tokom nastajanja zvučnog govora. Iako je fundamentalna frekvencija svojstvenost samog govornika, pod uticajem različitih emocionalnih stanja menja se zategnutost glasnica i subglotalni vazdušni pritisak, što se odražava na promene u fundamentalnoj frekvenciji. Razvijeno je više metoda za određivanje fundamentalne frekvencije, kako u vremenskom, tako i frekvencijskom domenu, a u ovom radu je korišćena kepralna metoda [40].

Određivanjem kepra se razdvajaju informacije koje potiču od vokalnog trakta i onih koje potiču od pobude, pri tome se karakteristike vokalnog trakta preslikavaju na niže delove kepra, dok se karakteristike pobude odražavaju na više delove kepra.

Za dati signal $x(n)$ određuje se kepr star snage prema izrazu:

$$C(\tau) = IDFT(\ln(|DFT(x(n))|^2)), \quad (3.30)$$

gde su DFT i $IDFT$ diskretna Furijerova transformacija i inverzna diskretna Furijerova transformacija. Estimacija fundamentalne frekvencije se određuje traženjem pikova u višim delovima kepra:

$$\hat{f}_0 = \frac{1}{\tau_{max}}, \quad C(\tau_{max}) = \max_{\tau} C(\tau), \quad \tau > \tau_0. \quad (3.31)$$

3.2.2 Energija

Kratkovremena energija govornog signala pruža informaciju o intenzitetu zvuka i direktno je povezana sa nivoom emocionalnog naboja u govoru. Kratkovremena energija govornog signala $x(n)$, $n = 0, \dots, N - 1$, je određena izrazom:

$$E = \frac{1}{N} \sum_{n=0}^{N-1} x(n)^2. \quad (3.32)$$

Vredan pokazatelj emocionalnog stanja može biti informacija o prisustvu energije u različitim frekvencijskim opsezima. Signal se prvo tretira filterom propusnikom opsega, a potom se vrši računanje energije filtriranog signala prema prethodnoj formuli. U ovom radu su razmatrani

frekvencijski opsezi (u hercima): [20,100], [100,260], [260,580], [580,1220], [1220,2500], [2500,5060].

Mogu se određivati logaritamske vrednosti energije nefiltriranog i filtriranog signala, što simulira nelinearnu osetljivost ljudskog slušnog aparata na različite intenzitete zvuka.

3.2.3 Formanti

Formanti se odnose na rezonantne učestanosti vokalnog trakta, tako da imaju blisku vezu sa njegovom geometrijom, a sa druge strane, geometrija vokalnog trakta se menja i pod uticajem emocionalnog stanja govornika. U ranijim istraživanjima su korišćene različite metode za određivanje pozicije formanata zasnovane na Furijerovoj transformaciji, kepstru i LPC analizi [41], pri čemu se pristup na bazi LPC analize pokazao superiornim. Kod LPC analize se pretpostavlja autoregresivni model vokalnog trakta, odnosno, odbirak $x(n)$ govornog signala u trenutku n se predstavlja kao linearna kombinacija prethodnih p odbiraka i greške predikcije $e(n)$:

$$x(n) = \sum_{k=1}^p a_k x(n-k) + e(n), \quad (3.33)$$

gde su a_k , $k = 1, \dots, p$, LPC koje je potrebno odrediti, a p predstavlja red modela. Opšte je pravilo da se za red modela bira vrednost duplo veća od broja očekivanih formanata uvećana za dva. U ovom radu će biti određivana prva tri formanta, tako da je za red modela usvojeno $p = 8$.

Slično kao i ranije, ako je z -transformacija govornog signala $X(z)$ i z -transformacija signala greške $E(z)$, tada se može zapisati:

$$X(z) = E(z) \frac{1}{A(z)}, \quad (3.34)$$

pri čemu je funkcija prenosa vokalnog trakta:

$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} = \frac{1}{\prod_{k=1}^p (1 - c_k z^{-1})}. \quad (3.35)$$

Pikove funkcije prenosa vokalnog trakta je moguće odrediti na osnovu kompleksnih polova $c_k = r_k e^{j\omega_k}$. Odnosno, pozicija formanta k , f_k , se može odrediti na osnovu kompleksnog pola $c_k = r_k e^{j\omega_k}$ i frekvencije odabiranja F_s :

$$f_k = \frac{F_s}{2\pi} \omega_k, \quad k = 1, \dots, p. \quad (3.36)$$

Konačan rezultat predstavljaju tri najniže uzastopne vrednosti f_k dobijene prethodnom analizom.

3.2.4 Obeležja trajanja

Obeležja trajanja se bave pokazateljima trajanja koja se mogu dovesti u vezu sa različitim emocionalnim stanjima govornika. Obeležja trajanja su globalna obeležja, tj. izdvajaju se na nivou cele govorne sekvence. Pod obeležjima trajanja se mogu smatrati različite statistike trajanja slogova, reči, rečenica i pauza u govoru, ali i statistike kratkovremenih obeležja, kao što su relativni položaj maksimuma, minimuma i slično. Statistike kratkovremenih obeležja će biti razmatrane u poglavljima koja slede, dok će na ovom mestu fokus biti na različitim parametrima trajanja govora i pauza. U ovom radu će biti izdvajana sledeća obeležja trajanja:

- odnos trajanja govora i pauza;
- odnos trajanja govora i ukupnog trajanja sekvence;
- odnos trajanja pauza i ukupnog trajanja sekvence;
- prosečno trajanje govora;
- prosečno trajanje pauze;
- najduže trajanje govora;
- najduže trajanje pauze.

Da bi se omogućilo izdvajanje prethodnih parametara neophodno je izvršiti klasifikaciju frejmova na one koji sadrže govor i one koji sadrže pauze. Prvo se za svaki frejm određuje kratkovremena energija E (kao u 3.2.2) i broj prolazaka kroz nulu ZCR (kao što će biti opisano u 3.3.4), a potom se određuje prosečna vrednost energije E_{mean} i prosečna vrednost broja prolazaka kroz nulu ZCR_{mean} na nivou cele govorne sekvence. Frejm n sa energijom E_n i brojem prolazaka kroz nulu ZCR_n se proglašava govornim ako su zadovoljeni uslovi da je $E_n > 0.1E_{mean}$ i $ZCR_n < 0.9ZCR_{mean}$, a u suprotnom se proglašava pauzom.

3.3 Obeležja kvaliteta glasa

Pod obeležjima kvaliteta glasa se podrazumevaju šumnost, hrapavost, promuklost, napetost, šapat i druge promene u glasu koje se javljaju kao posledica promena u pobudnom signalu. Prema načinu izdvajanja, obeležja kvaliteta glasa se mogu podeliti u ona koja zahtevaju estimaciju pobudnog signala i ona koja se izdvajaju direktno iz govornog signala. Na ovom mestu će akcenat biti na obeležjima kvaliteta glasa koja se mogu izdvojiti direktno iz govornog signala: džiter, šimer i odnos intenziteta harmonika i šuma (HNR) [42]. U širem smislu, pod kvalitetom glasa se mogu podrazumevati i obeležja koja karakterišu različite tipove govornog signala, kao što su zvučni govor, bezvučni govor i tišina. S tim u vezi, biće razmatrani broj prolazaka kroz nulu (ZCR) i verovatnoća zvučnog govora.

3.3.1 Džiter

Džiter predstavlja perturbacije fundamentalne frekvencije u dve uzastopne periode pobudnog signala, a posledica je poremećaja kontrole vibracija glasnih žica, čiji uzrok može biti emocionalno stanje govornika. Džiter se može određivati kao globalno obeležje na nivou govorne sekvence upotrebom formule:

$$jitter = \frac{1}{L-1} \sum_{k=2}^L |f_0(k) - f_0(k-1)|, \quad (3.37)$$

gde L predstavlja broj frejmova, a $f_0(k)$ je fundamentalna frekvencija pridružena frejmu k .

3.3.2 Šimer

Šimer predstavlja amplitudsku perturbaciju pobudnog signala u dve uzastopne periode, a posledica je izmenjene otpornosti glotisa i promene u masi glasnih žica. Šimer se može određivati kao globalno obeležje prema formuli:

$$shimmer = \frac{1}{L-1} \sum_{k=2}^L |E(k) - E(k-1)|, \quad (3.38)$$

pri čemu je L broj frejmova, a $E(k)$ je kratkovremena energija frejma k .

3.3.3 HNR

HNR predstavlja meru periodičnosti govornog signala, tj. odnos periodične i aperiodične komponente govora, pri čemu prva komponenta potiče od vibracije glasnica, dok druga potiče od glotalnog šuma. Odnos periodične i aperiodične komponente oslikava efikasnost govora, odnosno intenzitet vibracija koju izdahnuti vazduh proizvodi na glasnim žicama, tako da je zvučni govor okarakterisan višim vrednostima HNR.

HNR se može proceniti na osnovu kratkovremene estimacije autokorelacione funkcije [43], određivanjem relativnog odnosa pikova. Autokorelaciona funkcija signala $x(n)$, $n = 0, \dots, N-1$, se može estimirati na osnovu formule:

$$r(\tau) = \frac{1}{N} \sum_{n=0}^{N-\tau-1} x(n)x(n+\tau), \quad \tau = 0, \dots, N-1. \quad (3.39)$$

Globalni maksimum funkcije $r(\tau)$ odgovara indeksu $\tau = 0$, dok je sledeći maksimum određen sa:

$$r_{max} = \max_{\tau} r(\tau) , \quad \tau > 0 . \quad (3.40)$$

HNR datog frejma se određuje prema formuli:

$$HNR = \frac{r_{max}}{r(0) - r_{max}} . \quad (3.41)$$

3.3.4 ZCR

ZCR daje informaciju o broju promena znaka govornog signala za vreme trajanja jednog frejma. ZCR predstavlja grubu procenu učestanosti zastupljenih u govornom signalu, tako da se može zaključiti da li je u pitanju zvučni govor, bezvučni govor ili pauza. Očekuju se relativno visoke vrednosti ZCR za bezvučni govor u odnosu na zvučni, dok bi u toku pauze vrednosti trebale da budu bliske nuli (iako ambijentalni šum može da prouzrokuje jako visoke vrednosti ZCR). Za datu sekvencu odbiraka govornog signala $x(0), \dots, x(N-1)$, ZCR je određen formulom:

$$ZCR = \frac{1}{2N} \sum_{n=1}^{N-1} |\text{sign}(x(n)) - \text{sign}(x(n-1))| , \quad (3.42)$$

pri čemu je:

$$\text{sign}(x(n)) = \begin{cases} 1, & x(n) > 0 \\ -1, & x(n) \leq 0 \end{cases} . \quad (3.43)$$

3.3.5 Verovatnoća zvučnog govora

Još jedna mera čijom se interpretacijom može zaključivati o svojstvu govora sadržanom u datom frejmu jeste verovatnoća zvučnog govora. Polazeći od ideje da zvučni govor karakteriše periodičnost koja potiče od vibracije glasnih žica, analizom stepena periodičnosti se može zaključivati o verovatnoći zvučnog govora. Na osnovu kratkovremene estimacije autokorelacione funkcije $r(\tau)$, kao u (3.39), i maksimuma r_{max} , kao u (3.40), moguće je odrediti verovatnoću zvučnog govora za dati frejm prema izrazu:

$$voicingP = \frac{r_{max}}{r(0)} . \quad (3.44)$$

3.4 Nelinearna obeležja

Većina obeležja se zasniva na pretpostavci linearnog modela produkcije govornog signala. S obzirom da govorni signal nastaje kao posledica nelinearnog protoka vazduha kroz govorni aparat, nameće se potreba za opisivanjem govornog signala upotrebom nelinearnih obeležja. Pod nelinearnim obeležjima u prvom planu se podrazumeva Tajger energija. Ideju ove nelinearne mere je prvi izneo Tajger [44], a algoritam za računanje je prvi definisao Kajzer [45]. Tajger energija predstavlja meru energije potrebne za generisanje signala, a ne energije sadržane u samom signalu. Tajger energija je lokalni operator koji se odnosi na pojedinačne odbirke, a ne na frejm u celini. Za n -ti odbirak signala $x(n)$, $n = 0, \dots, N - 1$, Tajger energija je data formulom:

$$TE(x(n)) = x(n)^2 - x(n-1)x(n+1). \quad (3.45)$$

Na nivou frejma se može odrediti ukupna Tajger energija:

$$TE_{sum} = \sum_{n=1}^{N-2} TE(x(n)). \quad (3.46)$$

Posebno vredan pokazatelj emocionalnih stanja može biti Tajger energija računata za različite frekvencijske opsege govornog signala. Govorni signal se prvo tretira filterom propusnikom opsega, a potom se vrši računanje Tajger energije filtriranog signala prema prethodnim izrazima. U ovom radu će biti razmatrani sledeći frekvencijski opsezi (u hercima): [20, 100], [100, 260], [260, 580], [580, 1220], [1220, 2500], [2500, 5060].

Moguće je vršiti dodatno uobličenje informacija o nelinearnim interakcijama logaritmovanjem vrednosti Tajger energije.

3.5 Delta koeficijenti

Izdvajanjem kratkovremenih obeležja se svakom frejmu pridružuje mera u odnosu na dato obeležje, pri čemu bi određivanje i informacije o vremenskoj promeni obeležja kroz frejmove moglo biti veoma korisno. U tu svrhu moguće je odrediti aproksimacije prvog i drugog izvoda uvođenjem Δ i $\Delta\Delta$ koeficijenata [46]. Za datu sekvencu kratkovremenih obeležja o_1, \dots, o_L , gde je L broj frejmova govorne sekvence, prvi izvod se može aproksimirati:

$$\Delta(o_l) = \frac{\sum_{k=1}^K k(o_{l+k} - o_{l-k})}{2 \sum_{k=1}^K k^2}, \quad l = 1, \dots, L, \quad (3.47)$$

pri čemu se pretpostavlja da je $o_l = 0$ za $l < 1$ i $l > L$, a $K = 2$ predstavlja susedstvo koje se uzima u obzir prilikom određivanja koeficijenata.

Drugi izvod se može aproksimirati:

$$\Delta\Delta(o_l) = \frac{\sum_{k=1}^K k(\Delta(o_{l+k}) - \Delta(o_{l-k}))}{2 \sum_{k=1}^K k^2}, \quad l = 1, \dots, L, \quad (3.48)$$

uz pretpostavku da je $\Delta(o_l) = 0$ za $l < 1$ i $l > L$.

Informacije o promeni obeležja su često vrednije od informacija sadržanih u samim obeležjima, zato će u ovom radu biti određivane aproksimacije prvog i drugog izvoda za svako kratkovremeno obeležje, za svaki frejm.

3.6 Statistike kratkovremenih obeležja

Većina prethodno opisanih obeležja su kratkovremena, tj. izdvajaju se na nivou frejma. Iako je klasifikaciju moguće vršiti na osnovu kratkovremenih obeležja, mnoge klasifikatorske metode postižu bolje rezultate ukoliko se koriste globalna obeležja na nivou govorne sekvence. Primenom statističkih pokazatelja je moguće predstaviti proizvoljan niz kratkovremenih vrednosti jednim globalnim reprezentom. U upotrebi su najrazličitije statistike, a u ovom radu će biti korišćena dva skupa: prvi skup se sastoji od pet osnovnih statistika, dok drugi (prošireni) skup pored statistika iz prvog skupa sadrži i niz dodanih statistika.

Za govornu sekvencu predstavljenu nizom od L kratkovremenih obeležja o_1, \dots, o_L je moguće izdvojiti statistike kao u Tab. 3.1. Statistike od 1 do 5 u Tab. 3.1 čine osnovni skup statistika, dok svih 17 statistika datih u tabeli čine prošireni skup. Statistika pod rednim brojem 17 određuje stepen perturbacije kratkovremenog obeležja na nivou govorne sekvence, što u slučaju primene nad fundamentalnom frekvencijom i kratkovremenom energijom daje ranije interpretirana obeležja kvaliteta glasa, džiter (odjeljak 3.3.1) i šimer (odjeljak 3.3.2).

Tab. 3.1 – Statistike kratkovremenih obeležja

| Redni broj | Naziv statistike | Formula |
|------------|-----------------------------------|---|
| 1 | Srednja vrednost | $mean = \frac{1}{L} \sum_{l=1}^L o_l$ |
| 2 | Standardna devijacija | $sd = \left(\frac{1}{L} \sum_{l=1}^L (o_l - mean)^2 \right)^{\frac{1}{2}}$ |
| 3 | Minimum | $min = \text{minimum}(o_1, \dots, o_L)$ |
| 4 | Maksimum | $max = \text{maximum}(o_1, \dots, o_L)$ |
| 5 | Opseg promene | $range = max - min$ |
| 6 | Medijan | $median = (\text{sort}(o_1, \dots, o_L))_{\frac{L}{2}}$ |
| 7 | Geometrijska srednja vrednost | $Gmean = \left(\prod_{l=1}^L o_l \right)^{\frac{1}{L}}$ |
| 8 | Srednjekvadratna vrednost | $Qmean = \left(\frac{1}{L} \sum_{l=1}^L o_l^2 \right)^{\frac{1}{2}}$ |
| 9 | Pozicija minimuma | $pos_min = \frac{l_{min}}{L}, \quad o_{l_{min}} = min$ |
| 10 | Pozicija maksimuma | $pos_max = \frac{l_{max}}{L}, \quad o_{l_{max}} = max$ |
| 11 | Drugi centralni moment | $M2 = \frac{1}{L} \sum_{l=1}^L (o_l - mean)^2$ |
| 12 | Treći centralni moment | $M3 = \frac{1}{L} \sum_{l=1}^L (o_l - mean)^3$ |
| 13 | Četvrti centralni moment | $M4 = \frac{1}{L} \sum_{l=1}^L (o_l - mean)^4$ |
| 14 | Koeficijent asimetrije | $skewness = \frac{M3}{M2^{\frac{3}{2}}}$ |
| 15 | Koeficijent spljoštenosti | $kurtosis = \frac{M4}{M2^2}$ |
| 16 | Broj presecanja srednje vrednosti | $MCR = \frac{1}{2L} \sum_{l=1}^{L-1} \text{sign}(o_l - mean) - \text{sign}(o_{l+1} - mean) ,$ $\text{sign}(x) = \begin{cases} 1, & x > 0 \\ -1, & x \leq 0 \end{cases}$ |
| 17 | Perturbacija | $pertur = \frac{1}{L-1} \sum_{l=2}^L o_l - o_{l-1} $ |

3.7 Redukcija dimenzija

U prethodnim poglavljima je predstavljeno mnoštvo obeležja govornog signala, ali se ne može sa sigurnošću tvrditi o stepenu podobnosti u najrazličitijim primerima klasifikacije emocionalnih stanja. Jedno rešenje problema je optimizacija izbora obeležja prema konkretnoj primeni, dok bi univerzalnije rešenje bilo formiranje velikog skupa obeležja u cilju uključenja i onih obeležja koja nose relevantne informacije. Sa druge strane, povećanjem broja obeležja povećava se i računaska složenost, a moguća je i degradacija uspešnosti klasifikacije usled razuđenja prostora obeležja. Redukcija dimenzija se primenjuje sa ciljem smanjenja skupa obeležja bez velikog gubitka značajnih informacija.

U upotrebi su različite tehnike za redukciju dimenzija skupa obeležja, pri čemu se može izvršiti podela na metode koje vrše selekciju i metode koje vrše transformaciju obeležja. Pod selekcijom se podrazumeva odabir podskupa obeležja iz polaznog skupa tako da se sa što manjim brojem uključenih obeležja sačuva što više relevantnih informacija. Odabir skupa obeležja testiranjem svih mogućih kombinacija je nepraktično, umesto toga se odabir vrši prema usvojenoj strategiji pretrage, pri čemu se razlikuju: sekvencijalne (sekvencijalno dodavanje ili uklanjanje obeležja), slučajne (postoji slučajnost u strategiji pretrage) i eksponencijalne (broj kombinacija raste eksponencijalno sa brojem obeležja). Nakon što se odabere skup obeležja potrebno je izvršiti njegovu evaluaciju prema nekom od dva pristupa: na bazi filtera ili na bazi omotača. Evaluacija na bazi filtera podrazumeva ocenu izabranog skupa obeležja na osnovu informacionog sadržaja. Kod omotača se efikasnost izabranog skupa obeležja procenjuje na osnovu rezultata kompletnog klasifikatorskog sistema. Neke od često korišćenih tehnika za selekciju obeležja jesu: sekvencijalna selekcija unapred, sekvencijalna selekcija unazad, bidirekciona selekcija, metoda grana i grančica, metoda na bazi genetskih algoritama.

Drugi pristup redukcije dimenzija podrazumeva transformaciju datog prostora obeležja u prostor manjih dimenzija. Transformacija u opštem slučaju može biti nelinearna, ali je u praksi najčešće linearna, tako da se problem svodi na pronalaženje linearne transformacije A koja dati vektor obeležja X preslikava u novi vektor obeležja $Y = A^T X$. Pri projektovanju linearne transformacije se za kriterijum postavlja ili što tačnije očuvanje originalnih uzoraka u novom prostoru obeležja ili očuvanje što većeg međuklasnog rastojanja. Najpoznatije tehnike redukcije dimenzija na bazi transformacija jesu: analiza glavnih komponenti (PCA – eng. *Principal Component Analysis*) i linearna diskriminantna analiza (LDA – eng. *Linear Discriminant Analysis*). U ovom radu će biti u upotrebi linearna diskriminantna analiza, tako da će u nastavku biti detaljnije opisana.

Linearna diskriminantna analiza [47] sprovodi redukciju dimenzija vodeći računa o separabilnosti klasa. Pretpostavka je da je na raspolaganju obučavajući skup od N vektora X_i , $i = 1, \dots, N$, dimenzije n , $X_i = [x_{i1} \ x_{i2} \ \dots \ x_{in}]^T$, i da je moguće dati skup vektora podeliti u c klasa $\{X_i^{(k)}, i = 1, \dots, N_k, k = 1, \dots, c\}$. Potrebno je odrediti transformacionu matricu A dimenzije $n \times m$ koja će izvršiti linearnu transformaciju svakog od vektora iz obučavajućeg skupa X_i , $i = 1, \dots, N$ u odgovarajući m -dimenzionalni vektor Y_i , $i = 1, \dots, N$, tako da se i dalje može izvršiti razdvajanje u c klasa $\{Y_i^{(k)} = A^T X_i^{(k)}, i = 1, \dots, N_k, k = 1, \dots, c\}$. Svakoju od klasa se može pridružiti srednja vrednost i kovarijaciona matrica:

$$M_k = E\{Y^{(k)}\}, \quad \Sigma_k = E\{(Y^{(k)} - M_k)(Y^{(k)} - M_k)^T\}, \quad (3.49)$$

gde je $E\{\cdot\}$ operator matematičkog očekivanja. Kako u praksi najčešće nisu poznate funkcije gustine verovatnoće, određuju se aproksimacije:

$$M_k \approx \frac{1}{N_k} \sum_{j=1}^{N_k} Y_j^{(k)}, \quad \Sigma_k \approx \frac{1}{N_k} \sum_{j=1}^{N_k} (Y_j^{(k)} - M_k) (Y_j^{(k)} - M_k)^T. \quad (3.50)$$

Moguće je odrediti i procenu apriorne verovatnoće pojave određene klase u obučavajućem skupu:

$$P_k \approx \frac{N_k}{N}, \quad k = 1, \dots, c. \quad (3.51)$$

Definišu se matrica unutarklasnog rasejanja, S_w , i matrica međuklasnog rasejanja, S_b :

$$S_w = \sum_{k=1}^c P_k \Sigma_k, \quad S_b = \sum_{k=1}^c P_k (M_k - M_0) (M_k - M_0)^T, \quad M_0 = \sum_{k=1}^c P_k M_k, \quad (3.52)$$

koje će biti iskorišćene za postavljanje kriterijuma separabilnosti klasa. Matrica unutarklasnog rasejanja opisuje rasipanje uzoraka oko centara klasa, dok matrica međuklasnog rasejanja opisuje rasipanje oko združene srednje vrednosti. Kako bi razdvojenost klasa bila što izraženija, neophodno je da elementi matrice unutarklasnog rasejanja budu što manji, dok elementi matrice međuklasnog rasejanja treba da budu što veći. Moguće je postavljanje kriterijuma na više načina, a na ovom mestu će biti razmatrana postavka:

$$J_1 = \text{tr}(S_2^{-1} S_1), \quad (3.53)$$

pri čemu je $\text{tr}(\cdot)$ trag kvadratne matrice. U prethodnom kriterijum se za S_1 najčešće usvaja S_b , dok se za S_2 usvaja S_w ili $S_w + S_b$. Postavljeni kriterijum je potrebno maksimizovati u odnosu na nepoznatu transformacionu matricu A , što odgovara maksimizaciji S_1 i minimizaciji S_2 . Maksimum kriterijuma se ostvaruje ako se matrica A formira kao:

$$A = [\Psi_1 \ \Psi_2 \ \dots \ \Psi_m], \quad (3.54)$$

gde su Ψ_1, \dots, Ψ_m sopstveni vektori koji odgovaraju najvećim sopstvenim vrednostima matrice $S_2^{-1} S_1$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \lambda_{m+1} \geq \dots \geq \lambda_n$:

$$(S_2^{-1} S_1) \Psi_i = \lambda_i \Psi_i, \quad i = 1, \dots, m, \quad (3.55)$$

pri čemu su u prethodnom izrazu S_1 i S_2 matrice međuklasnog i unutarklasnog rasejanja originalnog skupa vektora X_i , $i = 1, \dots, N$.

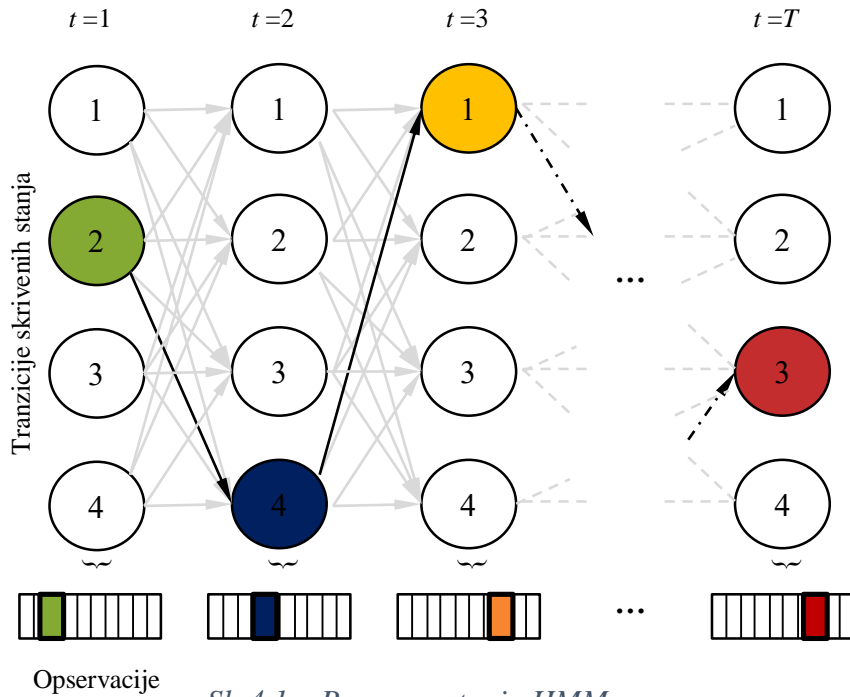
4 Tehnike klasifikacije

Postupak prepoznavanja emocionalnog stanja govornika sadržanog u datoj govornoj sekvenci se sastoji iz dva koraka. Prvi korak podrazumeva izdvajanje obeležja govornog signala prema nekoj od metoda iz prethodne glave, dok je u drugom koraku potrebno doneti odluku o sadržanom emocionalnom stanju na osnovu izdvojenih obeležja. Različite tehnike su korišćene u klasifikaciji emocionalnih stanja, a najčešće su to: HMM (skriveni Markovljevi modeli, eng. *Hidden Markov Models*), GMM (model Gausovih mešavina, eng. *Gaussian Mixture Models*), SVM (metoda potpornih vektora, eng. *Support Vector Machines*), ANN (veštačke neuralne mreže, eng. *Artificial Neural Networks*) i kNN (metoda k najbližih suseda, eng. *k Nearest Neighbours*). U nastavku će biti nešto više reči o klasifikatorskim metodama korišćenim u ovom radu, pri čemu su metode od izbora: HMM kao tradicionalan metod u obradi govora, SVM koji se pokazao superiornim u mnogim zadacima prepoznavanja oblika i DNN (duboke neuralne mreže, eng. *Deep Neural Networks*) kao reprezent dubokog učenja, pristupa koji je u povelju u obradi govora.

4.1 Skriveni Markovljevi modeli

Struktura skrivenih Markovljevih modela (HMM) [37], [48] je u velikoj meri istraživana u zadacima prepoznavanja govora. Iako se ne može napraviti striktna interpretacija modela, HMM je često u ulozi klasifikatora i u istraživanjima prepoznavanja emocija [36], [49], [38], [33]. Glavna karakteristika HMM je mogućnost modelovanja dinamike promene govornih obeležja, što može biti korisno prilikom klasifikacije emocionalnih stanja.

HMM je dvostruko stohastički proces. Prvi (skriveni) stohastički proces opisuje tranzicije između skrivenih stanja, gde naredno stanje zavisi samo od trenutnog stanja, a nezavisno je od prethodnih promena. Prvi proces se može sagledati samo kroz drugi stohastički proces, koji generiše različite opservacije zavisno od trenutnog stanja prvog procesa, a nezavisno od prethodnih stanja i opservacija. Na Sl. 4.1 je ilustrovan princip promene stanja i generisanja opservacija.



HMM je okarakterisan brojem skrivenih stanja N , brojem mogućih diskretnih opservacionih simbola M , matricom verovatnoće tranzicija između stanja A , matricom verovatnoće opservacija B i vektorom verovatnoće početnog stanja π . HMM se može kompaktno zapisati:

$$\lambda = (A, B, \pi). \quad (4.1)$$

Stanja u kojima se model može naći su iz skupa S , dok su moguće opservacije iz skupa V :

$$S = (s_1, s_2, \dots, s_N), \quad (4.2)$$

$$V = (v_1, v_2, \dots, v_M). \quad (4.3)$$

Za datu sekvencu stanja Q dužine T , O je odgovarajuća sekvenca opservacija:

$$Q = (q_1, q_2, \dots, q_T), \quad (4.4)$$

$$O = (o_1, o_2, \dots, o_T). \quad (4.5)$$

Matrica tranzicija, A , definiše verovatnoću prelaska iz stanja i u stanje j , nezavisno od vremena t :

$$A = [a_{ij}], \quad a_{ij} = P(q_t = j | q_{t-1} = i), \quad 1 \leq i, j \leq N. \quad (4.6)$$

Matrica verovatnoće opservacija, B , definiše verovatnoću generisanja opservacije k iz stanja i , nezavisno od vremena t :

$$B = [b_i(k)], \quad b_i(k) = P(o_t = k | q_t = i), \quad 1 \leq i \leq N, \quad 1 \leq k \leq M. \quad (4.7)$$

Vektor verovatnoće početnog stanja, π , definiše verovatnoću da je stanje i početno stanje:

$$\pi = [\pi_i], \quad \pi_i = P(q_1 = i), \quad 1 \leq i \leq N. \quad (4.8)$$

HMM karakterišu dva ograničenja. Prvo ograničenje je da trenutno stanje zavisi samo od prethodnog stanja, što predstavlja memoriju modela:

$$P(q_t|q_1, \dots, q_{t-1}) = P(q_t|q_{t-1}). \quad (4.9)$$

Drugo ograničenje jeste da opservacija u trenutku t zavisi samo do trenutnog stanja i da je nezavisna od prethodnih stanja i opservacija:

$$P(o_t|o_1, \dots, o_{t-1}, q_1, \dots, q_t) = P(o_t|q_t). \quad (4.10)$$

Za HMM se vezuju tri problema [37]:

Problem 1: Za datu sekvencu opservacija $O = (o_1, o_2, \dots, o_T)$ i dati model $\lambda = (A, B, \pi)$ je potrebno odrediti verovatnoću sa kojom je data sekvenca opservacija generisana od strane datog modela, $P(O|\lambda)$.

Problem 2: Za datu sekvencu opservacija $O = (o_1, o_2, \dots, o_T)$ i model $\lambda = (A, B, \pi)$ je potrebno odrediti odgovarajuću sekvencu stanja $Q = (q_1, q_2, \dots, q_T)$.

Problem 3: Za datu sekvencu opservacija $O = (o_1, o_2, \dots, o_T)$ je potrebno odrediti model $\lambda = (A, B, \pi)$, tako da se maksimizuje $P(O|\lambda)$.

4.1.1 Rešenje problema 1

Rešenje problema 1 podrazumeva određivanje verovatnoće sa kojom je data sekvenca opservacija generisana od strane datog modela, tj. određivanje u kojoj meri model odgovara opservacijama. Direktan pristup određivanja $P(O|\lambda)$ podrazumeva ispitivanje svih mogućih sekvenci stanja dužine T . Verovatnoća sekvence opservacija O za sekvencu stanja Q je određena izrazom:

$$P(O|Q, \lambda) = \prod_{t=1}^T P(o_t|q_t, \lambda) = b_{q_1}(o_1) \cdot b_{q_2}(o_2) \cdot \dots \cdot b_{q_T}(o_T), \quad (4.11)$$

pri čemu se pretpostavlja statistička nezavisnost opservacija. Verovatnoća sekvence stanja se može zapisati:

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T}. \quad (4.12)$$

Združena verovatnoća O i Q , tj. verovatnoća istovremene pojave O i Q je proizvod prethodno definisanih verovatnoća:

$$P(O, Q|\lambda) = P(O|Q, \lambda)P(Q|\lambda). \quad (4.13)$$

Verovatnoća sekvence O za dati model se dobija sumiranjem prethodno definisane združene verovatnoće za sve moguće sekvence stanja Q :

$$P(O|\lambda) = \sum_{all\ Q} P(O|Q, \lambda)P(Q|\lambda) = \sum_{all\ Q} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \dots a_{q_{T-1} q_T} b_{q_T}(o_T). \quad (4.14)$$

Izvedena formula ima jasnu interpretaciju, međutim, zahteva $(2T - 1)N^T$ množenja i $N^T - 1$

sabiranja, što je čini neprimjenjivom na realne probleme, tako da su razvijene efikasnije metode.

4.1.1.1 Procedura računanja unapred

Definiše se varijabla $\alpha_t(i)$ koja predstavlja verovatnoću sekvence opservacija do trenutka t , o_1, o_2, \dots, o_t , i stanja i u trenutku t , za dati model λ :

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = i | \lambda). \quad (4.15)$$

Vrednosti $\alpha_t(i)$, $1 \leq i \leq N$, $1 \leq t \leq T$, se mogu odrediti induktivnom metodom prema narednim koracima [37]:

1. Inicijalizacija:

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N. \quad (4.16)$$

2. Indukcija:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N. \quad (4.17)$$

3. Terminacija:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i). \quad (4.18)$$

4.1.1.2 Procedura računanja unazad

Definiše se varijabla $\beta_t(i)$ koja predstavlja verovatnoću sekvence opservacija od trenutka $t+1$ do trenutka T , $o_{t+1}, o_{t+2}, \dots, o_T$, za dato stanje i u trenutku t i dati model λ :

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = i, \lambda). \quad (4.19)$$

Vrednosti $\beta_t(i)$, $1 \leq i \leq N$, $1 \leq t \leq T$, se mogu odrediti induktivnom metodom prema narednim koracima [37]:

1. Inicijalizacija:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N. \quad (4.20)$$

2. Indukcija:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1, \quad 1 \leq i \leq N. \quad (4.21)$$

4.1.2 Rešenje problema 2

Rešenje problema 2 podrazumeva određivanje sekvence skrivenih stanja datog modela koja je generisala datu sekvencu opservacija, pri čemu ne postoji tačna sekvenca stanja, već se traži optimalna sekvenca stanja u odnosu na usvojeni kriterijum. Jedan mogući kriterijum jeste odabir najverovatnijeg stanja q_t za svaki individualan trenutak t , čime se maksimizuje verovatnoća korektnog odabira individualnih stanja.

Definiše se varijabla $\gamma_t(i)$ koja predstavlja verovatnoću stanja i u trenutku t za datu sekvencu opservacija O i model λ :

$$\gamma_t(i) = P(q_t = i | O, \lambda). \quad (4.22)$$

Dalje se može izvesti:

$$\gamma_t(i) = \frac{P(O, q_t = i | \lambda)}{P(O | \lambda)} = \frac{P(O, q_t = i | \lambda)}{\sum_{j=1}^N P(O, q_t = j | \lambda)}. \quad (4.23)$$

S obzirom da je $P(O, q_t = i | \lambda) = \alpha_t(i)\beta_t(i)$, pri čemu se $\alpha_t(i)$ odnosi na deo sekvence do trenutka t , o_1, o_2, \dots, o_t , i stanje i u trenutku t , dok se $\beta_t(i)$ odnosi na ostatak sekvence, $o_{t+1}, o_{t+2}, \dots, o_T$, može se zapisati:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}. \quad (4.24)$$

Na osnovu $\gamma_t(i)$ se određuje individualno najverovatnije stanje q_t^* u trenutku t :

$$q_t^* = \arg \max_{1 \leq i \leq N} [\gamma_t(i)], \quad 1 \leq t \leq T. \quad (4.25)$$

Prethodnim kriterijumom se maksimizuje broj korektnih stanja, ali se ne vodi računa o sekvenci stanja, tako da je moguće doći i do sekvence koja nije validna. Moguće je proširiti kriterijum tako da se maksimizacija vrši u odnosu na parove ili triplete stanja, ali se ipak češće kriterijum postavlja tako da se traži jedinstvena najverovatnija sekvenca stanja, tj. maksimizuje se $P(Q | O, \lambda)$. Metoda koja određuje najverovatniju sekvencu stanja jeste Viterbijev algoritam [50].

4.1.2.1 Viterbijev algoritam

Viterbijev algoritam se primenjuje sa ciljem određivanja jedinstvene najverovatnije sekvence stanja $Q = (q_1, q_2, \dots, q_T)$ za datu sekvencu opservacija $O = (o_1, o_2, \dots, o_T)$ i dati model $\lambda = (A, B, \pi)$. Za datu sekvencu opservacija do trenutka t , o_1, o_2, \dots, o_t , i najverovatniju sekvencu stanja do trenutka t koja se završava u stanju i , $q_1, q_2, \dots, q_{t-1}, q_t = i$, definiše se:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_{t-1}, q_t = i, o_1, o_2, \dots, o_t | \lambda). \quad (4.26)$$

Indukcijom se dobija:

$$\delta_{t+1}(j) = \max_i [\delta_t(i) a_{ij}] b_j(o_{t+1}). \quad (4.27)$$

Da bi se omogućilo praćenje sekvence stanja, neophodno je praćenje argumenata koji maksimizuju prethodni izraz za svako t i j , a u tu svrhu se uvodi $\psi_t(j)$. U nastavku su dati koraci za određivanje najverovatnije sekvence stanja [37]:

1. Inicijalizacija:

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1) \\ \psi_1(i) &= 0 \end{aligned}, \quad 1 \leq i \leq N. \quad (4.28)$$

2. Rekurzija:

$$\begin{aligned} \delta_t(j) &= \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t) & 2 \leq t \leq T \\ \psi_t(j) &= \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] & 1 \leq j \leq N \end{aligned} \quad (4.29)$$

3. Terminacija

$$\begin{aligned} P^* &= \max_{1 \leq i \leq N} [\delta_T(i)] \\ q_T^* &= \arg \max_{1 \leq i \leq N} [\delta_T(i)] \end{aligned} \quad (4.30)$$

4. Izdvajanje sekvence:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1. \quad (4.31)$$

4.1.3 Rešenje problema 3

Treći problem se odnosi na podešavanje parametara modela (A, B, π) u odnosu na datu sekvencu opservacija, tako da se zadovolji postavljeni optimizacioni kriterijum. Nije poznato analitičko rešenje problema, ali je moguće odrediti model $\lambda = (A, B, \pi)$ koji lokalno maksimizuje $P(O|\lambda)$. Lokalno optimalno rešenje je moguće odrediti upotrebom Baum-Velč (eng. *Baum-Welch*) metode [51] ili upotrebom gradijentnih algoritama. Na ovom mestu biće više reči o Baum-Velč metodi, poznatoj i kao metoda maksimalnog očekivanja (EM, eng. *Expectation Maximization*).

4.1.3.1 Baum-Velč algoritam

Za dati model λ i sekvencu opservacija O se definiše verovatnoća stanja i u trenutku t i stanja j u trenutku $t+1$:

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda). \quad (4.32)$$

Na osnovu definicija varijabli $\alpha_t(i)$ i $\beta_{t+1}(j)$ datih u tačkama 4.1.1.1 i 4.1.1.2, $\xi_t(i, j)$ se može zapisati u obliku [37]:

$$\xi_t(i, j) = \frac{P(q_t = i, q_{t+1} = j, O | \lambda)}{P(O | \lambda)} = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i_1=1}^N \sum_{j_1=1}^N \alpha_t(i_1) a_{i_1 j_1} b_{j_1}(o_{t+1}) \beta_{t+1}(j_1)}, \quad (4.33)$$

za $1 \leq i \leq N, 1 \leq j \leq N, 1 \leq t \leq T-1$.

Ranije je $\gamma_t(i)$ (odjeljak 4.1.2) definisano kao verovatnoća stanja i u trenutku t za dati model i sekvencu opservacija, tako da se može zapisati:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j). \quad (4.34)$$

Sumiranjem $\gamma_t(i)$ po t se dobija očekivani broj ponavljanja stanja i . Sumiranjem $\xi_t(i, j)$ po t se dobija očekivani broj tranzicija iz stanja i u stanje j . Na osnovu uočenog je moguće izvesti reestimacione formule parametara modela A, B i π :

$$\bar{\pi}_i = \gamma_1(i), \quad (4.35)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad (4.36)$$

$$\bar{b}_j(k) = \frac{\sum_{t=1}^{T-1} \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(j)} \cdot \quad (4.37)$$

Parametri aktuelnog modela $\lambda = (A, B, \pi)$ se koriste u reestimacionim formulama i time se dolazi do reestimiranog modela $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$. Ukoliko se u prethodnim formulama iterativno koristi $\bar{\lambda}$ umesto λ , popravljaje se verovatnoća $P(O|\lambda)$ sve dok se ne dostigne lokalni maksimum. Primenom reestimacionih formula se čuvaju stohastička ograničenja HMM parametara: $\sum_{i=1}^N \bar{\pi}_i = 1$, $\sum_{j=1}^N \bar{a}_{ij} = 1$, za $1 \leq i \leq N$ i $\sum_{k=1}^M \bar{b}_j(k) = 1$, za $1 \leq j \leq N$.

U dosadašnjem izlaganju je podrazumevan potpuno povezan (eng. *ergodic*) HMM, kod koga je moguć prelazak između bilo koja dva stanja u jednom koraku, a moguće su i drugačije strukture, kao što je struktura sa leva na desno ili neka druga. Različite strukture modela se postižu izjednačavanjem odgovarajućih elemenata matrica π i A sa nulom, pri čemu primena reestimacionih formula ne utiče na promenu vrednosti nultih elemenata.

4.1.3.2 Skaliranje

Prilikom implementacije HMM postaje značajna činjenica da su članovi a_{ij} i $b_j(k)$ uvek manji od jedan, a često i mnogo manji, pri čemu prethodno izvedene formule zahtevaju množenje niza takvih vrednosti, što u slučaju sekvenci opservacija velikih dužina može rezultovati suviše malim vrednostima i narušenom preciznošću. Rešenje problema se može naći u uvođenju skaliranja u prethodno opisani algoritam.

Ideja skaliranja se zasniva na množenju varijabli $\alpha_t(i)$ i $\beta_t(i)$ skalirajućim koeficijentima za svako t nezavisno od i , održavajući time preciznost, bez narušavanja rezultata procedure [37]. Za prikaz ideje biće iskorišćena reestimaciona formula (4.36) koeficijentata a_{ij} , koja sa uključenim varijablama $\alpha_t(i)$ i $\beta_t(i)$ ima oblik:

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j_1=1}^N \alpha_t(i) a_{ij_1} b_{j_1}(o_{t+1}) \beta_{t+1}(j_1)}. \quad (4.38)$$

U nastavku će $\alpha_t(i)$ označavati neskaliranu vrednost, $\hat{\alpha}_t(i)$ će označavati skaliranu vrednost, a $\hat{\hat{\alpha}}_t(i)$ će označavati lokalnu verziju pre skaliranja. Za $t = 1$, određuje se $\alpha_1(i)$ kao u formuli (4.16), i vrši se postavljanje:

$$\hat{\alpha}_1(i) = \alpha_1(i), \quad c_1 = \frac{1}{\sum_{i=1}^N \alpha_1(i)}, \quad \hat{\alpha}_1(i) = c_1 \alpha_1(i). \quad (4.39)$$

Za svako t , $2 \leq t \leq T$, prvo se računa $\hat{\alpha}_t(i)$ u odnosu na skalirane vrednosti $\hat{\alpha}_{t-1}(j)$ prema formuli (4.17), tako da se dobija:

$$\hat{\alpha}_t(i) = \sum_{j=1}^N \hat{\alpha}_{t-1}(j) a_{ji} b_i(o_t). \quad (4.40)$$

Skalirajući koeficijent c_t se određuje na osnovu $\hat{\alpha}_t(i)$:

$$c_t = \frac{1}{\sum_{i=1}^N \hat{\alpha}_t(i)}, \quad (4.41)$$

što daje:

$$\hat{\alpha}_t(i) = c_t \hat{\alpha}_t(i), \quad (4.42)$$

tako da se $\hat{\alpha}_t(i)$ može zapisati:

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^N \hat{\alpha}_{t-1}(j) a_{ji} b_i(o_t)}{\sum_{i_1=1}^N \sum_{j=1}^N \hat{\alpha}_{t-1}(j) a_{ji_1} b_{i_1}(o_t)}. \quad (4.43)$$

Indukcijom se može doći do:

$$\hat{\alpha}_{t-1}(j) = \left(\prod_{\tau=1}^{t-1} c_\tau \right) \alpha_{t-1}(j), \quad (4.44)$$

odnosno, $\hat{\alpha}_t(i)$ se dalje može zapisati kao:

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^N \alpha_{t-1}(j) (\prod_{\tau=1}^{t-1} c_\tau) a_{ji} b_i(o_t)}{\sum_{i_1=1}^N \sum_{j=1}^N \alpha_{t-1}(j) (\prod_{\tau=1}^{t-1} c_\tau) a_{ji_1} b_{i_1}(o_t)} = \frac{\alpha_t(i)}{\sum_{i_1=1}^N \alpha_t(i_1)}, \quad (4.45)$$

tako da se svaki član $\alpha_t(i)$ efektivno skalira sumom $\alpha_t(i)$ po svim stanjima.

Skaliranje $\beta_t(i)$ se sprovodi na sličan način, pri čemu se skaliranje vrši koeficijentima određenim za $\alpha_t(i)$, tako da je skalirana forma $\hat{\beta}_t(i)$:

$$\hat{\beta}_t(i) = c_t \hat{\beta}_t(i) = c_t \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \hat{\beta}_{t+1}(j). \quad (4.46)$$

Ako se u reestimacionoj formuli (4.38) koriste skalirane vrednosti, dobija se:

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) a_{ij} b_j(o_{t+1}) \hat{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j_1=1}^N \hat{\alpha}_t(i) a_{ij_1} b_{j_1}(o_{t+1}) \hat{\beta}_{t+1}(j_1)}. \quad (4.47)$$

Svaki član $\hat{\alpha}_t(i)$ može biti zapisan kao:

$$\hat{\alpha}_t(i) = \left(\prod_{\tau=1}^t c_\tau \right) \alpha_t(i) = C_t \alpha_t(i). \quad (4.48)$$

Svaki član $\hat{\beta}_{t+1}(j)$ se može zapisati kao:

$$\hat{\beta}_{t+1}(j) = \left(\prod_{\tau=t+1}^T c_\tau \right) \beta_{t+1}(j) = D_{t+1} \beta_{t+1}(j). \quad (4.49)$$

Zamenom (4.48) i (4.49) u izraz (4.47) se dobija:

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} C_t \alpha_t(i) a_{ij} b_j(o_{t+1}) D_{t+1} \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j_1=1}^N C_t \alpha_t(i) a_{ij_1} b_{j_1}(o_{t+1}) D_{t+1} \beta_{t+1}(j_1)}. \quad (4.50)$$

Proizvod $C_t D_{t+1}$ se može zapisati kao:

$$C_t D_{t+1} = \prod_{\tau=1}^t c_\tau \prod_{\tau=t+1}^T c_\tau = \prod_{\tau=1}^T c_\tau = C_T, \quad (4.51)$$

odnosno, član $C_t D_{t+1}$ je nezavistan od t , što omogućavanja njegovo skraćivanje u brojiocu i imeniocu izraza (4.50), a time se dolazi do ranije izvedene reestimacione formule za a_{ij} , tako da je pokazano da upotreba skalirajućih koeficijenata ne menja rezultat reestimacione formule.

Procedura skaliranja je primenjiva i na reestimacione formule za određivanje π i B . Sa uvođenjem skaliranja je neophodno izvršiti samo promenu procedure za računanje $P(O|\lambda)$. Na osnovu svojstva:

$$\prod_{t=1}^T c_t \sum_{i=1}^N \alpha_T(i) = C_T \sum_{i=1}^N \alpha_T(i) = 1, \quad (4.52)$$

važi:

$$\prod_{t=1}^T c_t P(O|\lambda) = 1, \quad (4.53)$$

odakle sledi:

$$P(O|\lambda) = \frac{1}{\prod_{t=1}^T c_t}, \quad (4.54)$$

odnosno:

$$\log(P(O|\lambda)) = - \sum_{t=1}^T \log c_t. \quad (4.55)$$

4.1.3.3 Višestruke sekvence opservacija

Upotrebom većeg broja sekvenci opservacija se može postići pouzdanija estimacija parametara modela. Neka je dat skup od K uzajamno nezavisnih sekvenci opservacija:

$$O = [O^{(1)}, O^{(2)}, \dots, O^{(K)}], \quad (4.56)$$

gde je $O^{(k)} = (o_1^{(k)}, o_2^{(k)}, \dots, o_{T_k}^{(k)})$ k -ta sekvenca opservacija. Potrebno je odrediti parametre modela λ , tako da se maksimizuje kriterijum:

$$P(O|\lambda) = \prod_{k=1}^K P(O^{(k)}|\lambda) = \prod_{k=1}^K P_k. \quad (4.57)$$

Reestimacione formule prate učestanost određenih događaja, pa je odgovarajućim izmenama ranije izvedenih formula moguće omogućiti razmatranje većeg broja sekvenci opservacija. Uvođenje skaliranja ne menja rezultat izraza, tako da reestimacione formule za \bar{a}_{ij} i $\bar{b}_i(l)$ na bazi višestrukih opservacija i uključenim skaliranjem imaju oblik [37]:

$$\bar{a}_{ij} = \frac{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^k(i) a_{ij} b_j(o_{t+1}^{(k)}) \hat{\beta}_{t+1}^k(j)}{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^k(i) \hat{\beta}_t^k(i)}, \quad (4.58)$$

$$\bar{b}_i(l) = \frac{\sum_{k=1}^K \frac{1}{P_k} \sum_{\substack{t=1 \\ o_t=l}}^{T_k-1} \hat{\alpha}_t^k(i) \hat{\beta}_t^k(i)}{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^k(i) \hat{\beta}_t^k(i)}. \quad (4.59)$$

4.1.3.4 Inicijalizacija

Primenom reestimacionih formula se vrši korekcija parametara HMM u pravcu lokalnog maksimuma optimizacionog kriterijuma. Ostaje otvoreno pitanje izbora inicijalnih parametara tako da ostvareno lokalno optimalno rešenje bude jednako, ili što bliže, globalno optimalnom rešenju. Na ovo pitanje nema jasnog odgovora, tako da se u praksi najčešće koristi inicijalizacija slučajnim vrednostima ili uniformna inicijalizacija. S obzirom da je matrica verovatnoće opservacija, B , osetljivija na odabir početnih vrednosti, moguće je izvršiti ručnu segmentaciju sekvenci opservacija prema skrivenim stanjima, a potom proceniti inicijalne parametre modela, ili primeniti neki napredniji pristup.

4.1.3.5 Nedovoljan broj obučavajućih uzoraka

U praksi je broj obučavajućih uzoraka uvek limitiran, a sa druge strane, nedovoljan broj uzoraka može dovesti do loše estimiranih parametara modela. Ako se obrati pažnja na reestimacionu formulu za $b_j(k)$, ona određuje koliko puta se opservacija k javi u stanju j , tako da u slučaju nedovoljno velikog obučavajućeg skupa uzoraka može da se dobije vrednost $b_j(k) = 0$. Problem se može prevazići povećanjem obučavajućeg skupa ili redukcijom dimenzija modela, ali to često nije izvodljivo. Jednostavan način za prevazilaženje problema nedovoljnog broja obučavajućih uzoraka jeste postavljanje praga kako bi se osiguralo da parametri modela ne padnu ispod zadate vrednosti [37], što bi u slučaju $b_j(k)$ značilo uvođenje provere tako da:

$$b_j(k) = \begin{cases} b_j(k), & b_j(k) \geq \delta_b \\ \delta_b, & b_j(k) < \delta_b \end{cases} \quad (4.60)$$

4.1.3.6 Korektivna obuka

Kod formiranja modela je potrebno što vernije predstaviti obučavajući skup podataka. Nakon što se za dati obučavajući skup i dati broj klasa formiraju odgovarajući modeli, moguće je sprovesti dodatno podešavanje parametara modela primenom korektivne obuke [52]. Za sekvencu opservacija O^u iz klase u se određuje $P(O^u|\lambda_u)$, gde je λ_u model klase u , kao i $P(O^u|\lambda_w)$, gde je λ_w model klase w , $w \neq u$. Ukoliko je $\log P(O^u|\lambda_w) > \log P(O^u|\lambda_u) - \Delta$, pri čemu je Δ zadati prag, potrebno je izvršiti korekciju modela λ_u i λ_w . Korekcija parametara započinje određivanjem matrica verovatnoće opservacija \tilde{B}^u i \tilde{B}^w za korektnu klasu, u , i za pogrešnu klasu, w , upotrebom samo sekvence opservacija O^u . Potom se vrši modifikacija originalne matrice B^u iz modela λ_u tako da glasi $B^u + \gamma\tilde{B}^u$, dok se originalna matrica B^w iz modela λ_w modifikuje tako da glasi $B^w - \gamma\tilde{B}^w$, pri čemu parametar γ kontroliše intenzitet promena.

4.1.4 Vektorska kvantizacija

U dosadašnjem izlaganju su podrazumevane diskretne opservacije iz konačnog skupa simbola, međutim, čest je slučaj da realne primene ne odgovaraju datoj pretpostavci, gde su opservacije kontinualne i višedimenzionalne. Jedan način za prevazilaženje problema kontinualnih višedimenzionalnih opservacija jeste modifikacija modela tako da se umesto jednodimenzionalnih diskretnih funkcija gustine verovatnoće opservacija upotrebe višedimenzionalne kontinualne funkcije gustine verovatnoće. Uvođenje koraka vektorske kvantizacije je drugo rešenje problema kontinualnih vektorskih opservacija, što će biti primenjeno i u ovom radu. Ukoliko se pretpostavi postojanje konačnog skupa simbola sa pridruženim etalonima, vektorska kvantizacija se sprovodi tako što se proizvoljnom vektoru dodeljuje simbol etalona sa kojim ostvaruje najbolje podudaranje. Skup etalona se može formirati grupisanjem obučavajućih uzoraka, a potom određivanjem predstavnika svake od grupa, pri čemu se za potrebe grupisanja obučavajućih uzoraka najčešće koristi *K-means* klasterizacija [53].

K-means klasterizacija vrši razdvajanje datog skupa od L vektora u M klastera prema sledećoj proceduri:

1. Na slučajan način se vrši izbor M vektora koji se proglašavaju centroidima, tj. predstavnicima pojedinačnih klastera.
2. Svaki od L vektora se pridružuje klasteru najbližeg centroida.
3. Za svaki od M klastera se određuje nova pozicija centroida kao srednja vrednost vektora pridruženih klasteru.
4. Vršiti se provera konvergencije, a ukoliko postoji promena pozicije centroida ponavljaju se koraci 2 i 3.

Centroidi određeni prethodnom procedurom predstavljaju etalone grupa, a simbolička vrednost svake grupe je njen redni broj. Euklidsko rastojanje se koristi kao mera podudaranja prilikom određivanja rastojanja vektora do centroida, kao i kod postupka kvantizacije.

4.2 Metoda potpornih vektora

Metoda potpornih vektora (SVM) [54], [55], [56] je binarni klasifikator koji konstrukcijom hiper-ravni vrši razdvajanje podataka iz različitih klasa, uz istovremenu maksimizuju rastojanja od hiper-ravni do tačaka obe klase. SVM je inicijalno razvijen za potrebe binarnog razdvajanja linearno separabilnih klasa, a naknadno je ideja proširena kako bi se omogućilo razdvajanje nelinearno separabilnih i neseperabilnih klasa.

4.2.1 Linearni SVM

Pretpostavka je da su obučavajući podaci organizovani u skup uređenih parova $\{(x_1, y_1), \dots, (x_n, y_n)\}$, pri čemu za svako $i = 1, \dots, n$ važi da je $x_i \in \mathbb{R}^d$ tačka u prostoru podataka i da je $y_i \in \{-1, +1\}$ odgovarajuća oznaka klase. Potrebno je razdvojiti tačke različitih klasa određivanjem parametara hiper-ravni:

$$\langle w \cdot x \rangle + b = 0, \quad (4.61)$$

gde je w vektor normalan na hiper-ravan, b ofset, a $\langle \cdot \rangle$ skalarni proizvod. Razdvajanje tačaka linearno separabilnih klasa je moguće postići postavljanjem hiper-ravni na proizvoljno mnogo načina, ali se ostvaruje bolja generalizacija ukoliko se maksimizuje rastojanje od hiper-ravni do najbližih tačaka datih klasa [57].

Ako su sa d_+ i d_- označena najkraća rastojanja od hiper-ravni do tačaka jedne i druge klase, tada je zbirom $d_+ + d_-$ određena veličina margine hiper-ravni. Prilikom postavljanja hiper-ravni je potrebno maksimizovati marginu i pri tome nije dozvoljeno da se bilo koja tačka nađe u oblasti margine, što se može zapisati u obliku uslova:

$$\langle w \cdot x_i \rangle + b \geq +1, \quad \text{za } y_i = +1, \quad (4.62)$$

$$\langle w \cdot x_i \rangle + b \leq -1, \quad \text{za } y_i = -1. \quad (4.63)$$

Prethodna dva uslova se združeno mogu zapisati:

$$y_i(\langle w \cdot x_i \rangle + b) \geq 1, \quad i = 1, \dots, n. \quad (4.64)$$

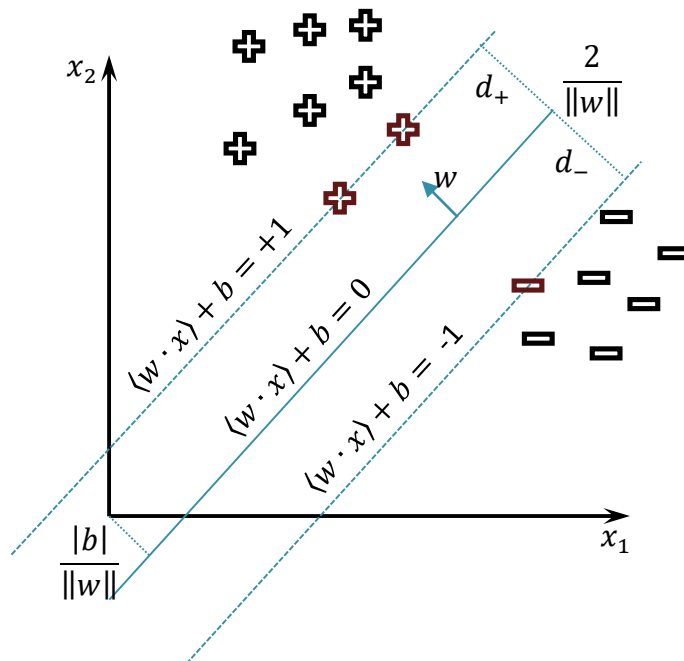
Rastojanje tačke x_i od hiper-ravni $\langle w \cdot x \rangle + b = 0$ je određeno izrazom:

$$d(w, b; x_i) = \frac{|\langle w \cdot x_i \rangle + b|}{\|w\|}, \quad (4.65)$$

gde je $\|w\| = \sqrt{w_1^2 + \dots + w_d^2}$. Na osnovu prethodnog izraza je rastojanje od hiper-ravni $\langle w \cdot x \rangle + b = 0$ do koordinatnog početka $|b|/\|w\|$. Slično je rastojanje hiper-ravni $\langle w \cdot x \rangle + b = 1$ do koordinatnog početka $|b - 1|/\|w\|$, a rastojanje hiper-ravni $\langle w \cdot x \rangle + b = -1$ do koordinatnog početka je $|b + 1|/\|w\|$. Najkraća rastojanja od hiper-ravni $\langle w \cdot x \rangle + b = 0$ do tačaka jedne i druge klase su $d_+ = 1/\|w\|$ i $d_- = 1/\|w\|$, odakle sledi da je veličina margine $d_+ + d_- = 2/\|w\|$ (Sl. 4.2).

Maksimizacija margine podrazumeva minimizaciju $\|w\|$, tako da se optimizacioni problem može zapisati:

$$\min_{w,b} \frac{1}{2} \|w\|^2, \text{ za } y_i(\langle w \cdot x_i \rangle + b) \geq 1, \quad i = 1, \dots, n. \quad (4.66)$$



Sl. 4.2 – Parametri SVM modela

Postavljeni problem je kvadratni optimizacioni problem sa linearnim ograničenijima, koji se može rešiti upotrebom matematičke tehnike Lagranžovih multiplikatora [58], [59], traženjem optimuma Lagranžijana:

$$L_P(w, b, \alpha_1, \dots, \alpha_n) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (\langle w \cdot x_i \rangle + b) - 1), \quad (4.67)$$

pri čemu su $\alpha_i, i = 1, \dots, n$, Lagranžovi multiplikatori. Za minimizaciju Lagranžijana se postavljaju jednakosti $\frac{\partial L_P}{\partial w} = 0$ i $\frac{\partial L_P}{\partial b} = 0$, odakle se dobija:

$$w = \sum_{i=1}^n \alpha_i y_i x_i, \quad 0 = \sum_{i=1}^n \alpha_i y_i. \quad (4.68)$$

Zamenom (4.68) u (4.67) se dolazi do dualnog problema:

$$L_D(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle, \quad (4.69)$$

čijom maksimizacijom se dolazi do vrednosti $\alpha_i, i = 1, \dots, n$, a nadalje se mogu odrediti parametri razdvajajuće hiper-ravani prema izrazima:

$$w^* = \sum_{i=1}^n \alpha_i y_i x_i, \quad b^* = \frac{1}{y_i} - \langle w^* \cdot x_i \rangle = y_i - \langle w^* \cdot x_i \rangle, \quad \alpha_i > 0. \quad (4.70)$$

Sve tačke x_i , $i = 1, \dots, n$, za koje važi $\alpha_i > 0$ se nazivaju potpornim vektorima.

Bilo koja nova tačka x' se klasifikuje na osnovu znaka izraza:

$$y' = \text{sign} (\langle w^* \cdot x' \rangle + b^*) = \begin{cases} +1, & \langle w^* \cdot x' \rangle + b^* > 0 \\ -1, & \langle w^* \cdot x' \rangle + b^* \leq 0 \end{cases}. \quad (4.71)$$

4.2.2 Linearni SVM sa mekanom marginom

U prethodnom izvođenju je pretpostavljena linearna separabilnost klasa, pri čemu se nije dozvoljavalo narušavanje granice margine razdvajajuće hiper-ravni, tako da se izvedena struktura može nazvati linearnim SVM sa tvrdom marginom. Ukoliko nije moguće odrediti hiper-ravan koja razdvaja date klase bez greške koristi se linearni SVM sa mekom marginom [54]. Uvođenjem koeficijenata ξ_i , $i = 1, \dots, n$, se dozvoljava postojanje tačaka koje narušavaju granice margine, tako da optimizacioni problem dobija formu:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i, \quad \text{za } y_i(\langle w \cdot x_i \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, n, \quad (4.72)$$

gde je C faktor koji određuje balans između maksimizacije margine i minimizacije greške klasifikacije. Upotrebom Lagranžovih multiplikatora se problem može predstaviti u vidu Lagranžijana:

$$\begin{aligned} L_P(w, b, \xi_1, \dots, \xi_n, \alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n) = \\ = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i(\langle w \cdot x_i \rangle + b) - 1 + \xi_i) - \sum_{i=1}^n \beta_i \xi_i. \end{aligned} \quad (4.73)$$

Prilikom minimizacije prethodnog izraza se postavljaju jednakosti $\frac{\partial L_P}{\partial w} = 0$, $\frac{\partial L_P}{\partial b} = 0$ i $\frac{\partial L_P}{\partial \xi_i} = 0$, odakle se dobija:

$$w = \sum_{i=1}^n \alpha_i y_i x_i, \quad 0 = \sum_{i=1}^n \alpha_i y_i, \quad \alpha_i + \beta_i = C. \quad (4.74)$$

Zamenom (4.74) u (4.73) se formira dualni problem:

$$L_D(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle, \quad (4.75)$$

čijom se maksimizacijom dolazi do α_i , $i = 1, \dots, n$, a parametri hiper-ravni su određeni izrazima:

$$w^* = \sum_{i=1}^n \alpha_i y_i x_i, \quad b^* = y_i - \langle w^* \cdot x_i \rangle, \quad i = \operatorname{argmax}_{j=1, \dots, n}(\alpha_j). \quad (4.76)$$

Slično kao i ranije, bilo koja nova tačka x' se klasifikuje na osnovu znaka izraza:

$$y' = \operatorname{sign}(\langle w^* \cdot x' \rangle + b^*) = \begin{cases} +1, & \langle w^* \cdot x' \rangle + b^* > 0 \\ -1, & \langle w^* \cdot x' \rangle + b^* \leq 0 \end{cases}. \quad (4.77)$$

4.2.3 Nelinearni SVM

Često je potrebno izvršiti razdvajanje klasa koje nisu linearno separabilne, pri čemu primena linearne klasifikacije nije prihvatljivo rešenje. Kako bi se omogućila primena linearnih klasifikatora za razdvajanje nelinearno separabilnih klasa, upotrebljavaju se kernel funkcije [60]. Time se originalni, nelinearno separabilni podaci transformišu u prostor veće dimenzije u kojem se može primeniti linearna klasifikacija. Za datu tačku $x_i \in \mathbb{R}^d$ se uvodi transformacija $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^D$ ($D > d$) pomoću koje se vrši preslikavanje $x_i \rightarrow \Phi(x_i)$, tako da se može postaviti optimizacioni problem:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i, \quad \text{za } y_i(\langle w \cdot \Phi(x_i) \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, n. \quad (4.78)$$

Slično kao i ranije, dolazi se do dualnog problema:

$$L_D(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \Phi(x_i) \cdot \Phi(x_j) \rangle. \quad (4.79)$$

Nakon što se odrede α_i , $i = 1, \dots, n$, parametri w i b su određeni izrazima:

$$w^* = \sum_{i=1}^n \alpha_i y_i \Phi(x_i), \quad b^* = y_i - \sum_{j=1}^n \alpha_j y_j \langle \Phi(x_j) \cdot \Phi(x_i) \rangle, \quad i = \operatorname{argmax}_{j=1, \dots, n}(\alpha_j). \quad (4.80)$$

Nova tačka x' se klasifikuje na osnovu znaka izraza:

$$y' = \operatorname{sign}(\langle w^* \cdot \Phi(x') \rangle + b^*) = \operatorname{sign}\left(\sum_{i=1}^n \alpha_i y_i \langle \Phi(x_i) \cdot \Phi(x') \rangle + b^*\right). \quad (4.81)$$

U prethodnim izrazima figuriše skalarni proizvod transformisanih vrednosti, $\langle \Phi(x_i) \cdot \Phi(x_j) \rangle$, tako da je zgodno koristiti kernel funkcije za koje važi:

$$K(x_i, x_j) = \langle \Phi(x_i) \cdot \Phi(x_j) \rangle, \quad (4.82)$$

što omogućava direktno izračunavanje skalarnog proizvoda u transformisanom prostoru i na taj način se gubi potreba za određivanjem transformisanih vrednosti tokom računanja parametra i tokom klasifikacije. Neki od često korišćenih kernela jesu: linearni, polinomijalni, Gausov RBF (eng. *Radial Basis Function*), eksponencijalni RBF, sigmoid i drugi.

4.2.4 Višeklasni SVM

SVM se u osnovnoj formi može primeniti isključivo na probleme binarne klasifikacije. Ukoliko je potrebno vršiti razdvajanje većeg broja klasa, neophodno je usvojiti strategiju za kombinovanje rezultata binarnih klasifikacija. Postoji više predloga za proširenje SVM na višeklasne probleme [61], [62], a dva osnovna pristupa su data u nastavku.

Prvi pristup za formiranje višeklasnog SVM podrazumeva konstruisanje hiper-ravani koja razdvaja elemente jedne klase od elemenata svih preostalih klasa, što zahteva konstruisanje onoliko binarnih klasifikatora koliko ima i klasa. Nova tačka se pridružuje onoj klasi koja je proglašena u nekoj od binarnih klasifikacija, a u nerešenim slučajevima se razmatraju rastojanja od razdvajajućih hiper-ravni. Dobra strana pristupa jeste potreba za onolikim brojem binarnih klasifikacija koliko ima i klasa, sa druge strane, moguća je degradacija performansi binarnih klasifikacija usled neizbalansiranosti skupova podataka.

Kod drugog pristupa za formiranje višeklasnog SVM se za svaki par klasa konstruiše po jedan binarni klasifikator, a nova tačka se pridružuje onoj klasi koja je proglašena u najvećem broju binarnih klasifikacija. Broj potrebnih binarnih klasifikatora kod ovog pristupa je $M(M - 1)/2$, gde je M broj klasa. Mana ovog pristupa jeste značajno povećanje broja binarnih klasifikacija sa povećanjem broja klasa, ali je broj podataka za binarne obuke manji, pa je obuka brža, kao i klasifikacija.

4.3 Duboke neuralne mreže

Duboke neuralne mreže (DNN) [63] su veštačke neuralne mreže (ANN) sa većim brojem skrivenih slojeva između ulaznog i izlaznog sloja. Skriveni slojevi vrše modelovanje sa različitim stepenom generalizacije što omogućava rešavanje veoma složenih klasifikacionih problema. U nastavku će biti dat uvod u ANN, a potom će biti prikazan postupak realizacije DNN.

4.3.1 Veštačke neuralne mreže

Veštačke neuralne mreže (ANN) [64], [65], ili samo neuralne mreže (NN), su paralelni distribuirani sistemi za obradu informacija koji oponašaju biološke neuralne mreže. Slično biološkim sistemima, ANN imaju sposobnost učenja, memorisanja i generalizacije, pri čemu se osposobljavanje za dati zadatak sprovodi na osnovu primera, bez postavljanja eksplicitnih pravila. ANN se formiraju povezivanjem velikog broja procesorskih elemenata, a znanje je sadržano u samim vezama.

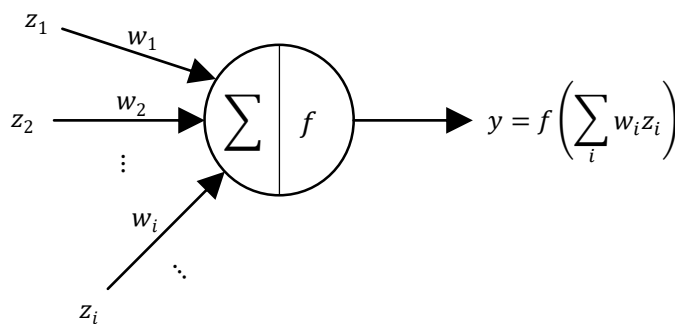
Procesorski elementi, kao osnovne gradivne komponente ANN, se po analogiji sa biološkim neuralnim sistemima nazivaju neuronima. Funkcija pojedinačnih neurona je vrlo jednostavna, gde se na osnovu niza ulaza vrši izračunavanje skalarnog izlaza. Jednovremeno funkcionisanje velikog broja

takvih jedinica omogućava izvođenje obimnih paralelnih izračunavanja. Zavisno od pozicije u ANN strukturi, razlikuje se nekoliko tipova neurona: ulazni neuroni prihvataju podatke iz spoljašnje sredine i prosleđuju ih na dalju obradu, skriveni neuroni vrše interne transformacije podataka, dok izlazni neuroni plasiraju rezultate obrade u spoljašnju sredinu.

Neuroni u ANN mogu biti povezani na proizvoljan način, pri čemu je svakoj vezi pridružen težinski koeficijent kojim se reguliše međusobni uticaj neurona. Matematički model neurona se može zapisati:

$$y = f\left(\sum_i w_i z_i\right). \quad (4.83)$$

U prethodnom izrazu je z_i jedan ulaz neurona, w_i je težinski koeficijent veze, a $\sum_i w_i z_i$ težinska suma svih ulaza neurona. Preslikavanje težinske sume u izlaz neurona, y , se vrši pomoću aktivacione funkcije, f (Sl. 4.3). U skladu sa primenom, moguće je koristiti različite aktivacione funkcije, kao što su: linearna, funkcija praga, rampa, sigmoid.



Sl. 4.3 – Matematički model neurona

Mogu se prepoznati različiti tipovi ANN u zavisnosti od usvojenog kriterijuma. Ukoliko se podela vrši na osnovu načina povezivanja neurona, moguće je razlikovati nestruktuirane, slojevite i modularne neuralne mreže. Ako se za kriterijum podele usvoji tok informacija, razlikuju se mreže kod kojih se propagacija signala vrši u jednom smeru (eng. *feedforward*) i one kod kojih je dozvoljena propagacija signala u oba smera (eng. *feedback*). U slučaju podele na osnovu načina obučavanja, može se napraviti razdvajanje na mreže koje se obučavaju sa nadgledanjem i mreže koje se obučavaju bez nadgledanja.

Obučavanje neuralnih mreža podrazumeva podešavanje težinskih koeficijenata neuronskih veza u skladu sa željenom funkcionalnošću mreže. Pored toga, moguće je sprovođenje obuke sa promenom strukture mreže, što se u praksi ređe radi. Određivanje težinskih koeficijenata nije trivijalan zadatak sa poznatim analitičkim rešenjem, ali su razvijeni algoritmi za iterativno podešavanje parametara na osnovu dostupnih primera.

4.3.1.1 Propagacija unazad

Propagacija unazad (eng. *backpropagation*) [66], [67] je najčešće korišćena metoda za obuku ANN. Postoji više varijanti algoritma propagacije unazad u zavisnosti od korišćene optimizacione tehnike, pri čemu je standardan pristup opadajućeg gradijenta [68], kod koga se težinski koeficijenti podešavaju u smeru opadanja kriterijumske funkcije. Ovakav algoritam obuke se može primeniti na višeslojne *feedforward* NN sa aktivacionom funkcijom diferencijabilnom bez prekida.

Za dati obučavajući skup sačinjen od P parova ulaza i očekivanih izlaza $\{(X^{(p)}, T^{(p)}) \mid p = 1, \dots, P\}$ procedura obuke se sprovodi u dva koraka [65]. U prvom koraku se ulazni podatak, $X^{(p)}$, dovodi na ulaz mreže i određuje se izlaz, $Y^{(p)}$, koji mreža ostvaruje za dati ulaz. U drugom koraku se greška u vidu odstupanja stvarnog izlaza, $Y^{(p)}$, od očekivanog, $T^{(p)}$, propagira unazad od izlaznog sloja ka prethodnim slojevima kako bi se izvršilo podešavanje težinskih koeficijenata. Procedura se ponavlja za svaki obučavajući par $(X^{(p)}, T^{(p)})$, $p = 1, \dots, P$, do dostizanja konvergencije.

Neka su $i = 1, \dots, I$ indeksi neurona iz prethodnog sloja, $j = 1, \dots, J$ indeksi neurona iz datog sloja, O skup izlaznih neurona i $p = 1, \dots, P$ indeksi obučavajućih uzoraka. Sa $x_j^{(p)}$, $y_j^{(p)}$ i $t_j^{(p)}$ su označeni težinska suma, ostvareni izlaz i očekivani izlaz neurona j koji odgovaraju obučavajućem uzorku p . Za aktivacionu funkciju, f , je zgodno usvojiti sigmoid, s obzirom da ima osobinu diferencijabilnosti bez prekida. Na osnovu uvedene postavke se može zapisati:

$$x_j^{(p)} = \sum_{i=1}^I w_{ji} y_i^{(p)}, \quad (4.84)$$

$$y_j^{(p)} = f(x_j^{(p)}) = \frac{1}{1 + e^{-x_j^{(p)}}}, \quad (4.85)$$

Grešku koja odgovara obučavajućem uzorku p i ukupnu izlaznu grešku za ceo skup uzoraka, $E^{(p)}$ i E , je moguće izraziti na sledeći način:

$$E^{(p)} = \frac{1}{2} \sum_{j=1}^J (y_j^{(p)} - t_j^{(p)})^2, \quad j \in O, \quad (4.86)$$

$$E = \sum_{p=1}^P E^{(p)}. \quad (4.87)$$

Procedura propagacije unazad vrši promenu težinskih koeficijenata w_{ji} srazmerno njihovom uticaju na grešku, i to tako da utiče na njeno smanjenje, pri čemu konstanta ε kontroliše brzinu promene:

$$\Delta^{(p)} w_{ji} = -\varepsilon \frac{\partial E^{(p)}}{\partial w_{ji}} = -\varepsilon \frac{\partial E^{(p)}}{\partial y_j^{(p)}} \frac{\partial y_j^{(p)}}{\partial x_j^{(p)}} \frac{\partial x_j^{(p)}}{\partial w_{ji}} = -\varepsilon \frac{\partial E^{(p)}}{\partial y_j^{(p)}} f'(x_j^{(p)}) y_i^{(p)}. \quad (4.88)$$

U prethodnom izrazu se prilikom računanja parcijalnog izvoda $\frac{\partial E^{(p)}}{\partial y_j^{(p)}}$ pravi razlika u zavisnosti da li je j izlazni neuron mreže ($j \in O$) ili se njegov izlaz koristi kao ulaz neurona $k = 1, \dots, K$ iz narednog sloja, tako da važi:

$$\begin{aligned} \frac{\partial E^{(p)}}{\partial y_j^{(p)}} &= y_j^{(p)} - t_j^{(p)}, & \text{za } j \in O, \\ \frac{\partial E^{(p)}}{\partial y_j^{(p)}} &= \sum_{k=1}^K \frac{\partial E^{(p)}}{\partial y_k^{(p)}} \frac{\partial y_k^{(p)}}{\partial x_k^{(p)}} \frac{\partial x_k^{(p)}}{\partial y_j^{(p)}} = \sum_{k=1}^K \frac{\partial E^{(p)}}{\partial y_k^{(p)}} f'(x_k^{(p)}) w_{kj}, & \text{za } j \notin O, \end{aligned} \quad (4.89)$$

Na osnovu pokazanog, $\frac{\partial E^{(p)}}{\partial y_j^{(p)}}$ je moguće direktno izračunati ukoliko neuron j pripada izlaznom sloju, u suprotnom je do rešenja moguće doći koristeći $\frac{\partial E^{(p)}}{\partial y_k^{(p)}}$ određene za neurone iz narednog sloja, $k = 1, \dots, K$. Dakle, korekciju koeficijenata je potrebno započeti od izlaznog sloja, pa propagacijom unazad vršiti korekcije redom u prethodnim slojevima. Postupak se ponavlja za sve obučavajuće uzorke, $p=1, \dots, P$, dok se ne postigne konvergencija parametara.

4.3.1.2 Obuka metodom skaliranog konjugovanog gradijenta

Prethodno opisani pristup obučavanja neuralnih mreža ima nedostatak u vidu spore konvergencije kao i zavisnosti performanse od izbora parametara, što posebno dolazi do izražaja kod ANN sa velikim brojem neurona. Razmatrane su različite mogućnosti unapređenja standardnog postupka obuke propagacijom unazad, pri čemu se izdvaja ideja upotrebe optimizacionih tehnika konjugovanih gradijenata [69]. U nastavku će biti detaljnije opisan postupak obuke ANN metodom skaliranog konjugovanog gradijenta (SCG, eng. *Scaled Conjugate Gradient*).

Obučavanje ANN se može posmatrati kao problem minimizacije globalne funkcije greške, E , u prostoru težinskih koeficijenata, $w \in \mathbb{R}^N$. U opštem slučaju je u pitanju lokalni iterativni proces minimizacije aproksimacije funkcije greške u okolini trenutne tačke prostora težinskih koeficijenata, pri čemu je aproksimacija funkcije greške određena Tejlorovim polinomom prvog ili drugog reda. Strategija optimizacije se može opisati sledećim koracima:

1. Postavlja se početni vektor težinskih koeficijenata, w_1 , i indeks, $k = 1$.
2. Određuje se pravac pretrage, p_k , i veličina koraka, α_k , tako da je $E(w_k + \alpha_k p_k) < E(w_k)$.
3. Ažurira se vektor težinskih koeficijenata, $w_{k+1} = w_k + \alpha_k p_k$.
4. Ukoliko je $E'(w_k) \neq 0$, inkrementira se indeks, $k = k + 1$, i postupak se ponavlja od koraka 2, u suprotnom je w_{k+1} traženo rešenje.

U prethodnom iterativnom postupku je potrebno definisati strategiju za određivanje sledeće tačke prostora težinskih koeficijenata, što podrazumeva određivanje pravca pretrage i veličine koraka u datom pravcu. Ukoliko se za pravac pretrage, p_k , usvoji negativni gradijent funkcije greške, $-E'(w)$, i ukoliko se za veličinu koraka, α_k , usvoji konstanta, ε , strategija optimizacije postaje prethodno opisani postupak opadajućeg gradijenta. Minimizacija zasnovana na linearnoj aproksimaciji $E(w + y) = E(w) + E'(w)^T y$ uzrokuje sporu konvergenciju, a konstantna veličina koraka utiče na smanjenje robusnosti.

Metode optimizacije na bazi konjugovanih gradijenata koriste prethodni iterativni postupak, ali sa pažljivijim određivanjem pravca pretrage i veličine koraka, pri čemu se upotrebljava aproksimacija funkcije greške drugog reda, $E(w + y) = E(w) + E'(w)^T y + \frac{1}{2} y^T E''(w) y$. Kod optimizacionih metoda konjugovanih gradijenata se vrši pretraga u konjugovanim pravcima u odnosu na prethodne pravce pretrage ($p_k^T E''(w) p_j = 0, j = 1, \dots, k - 1$) sa koracima odgovarajuće veličine, što isključuje potrebu za ponavljanjem pretrage u istim pravcima i garantuje konvergenciju u maksimalno N iteracija. Predloženo je nekoliko strategija optimizacije na bazi konjugovanih gradijenta [70], [71] koje mogu da obezbede bržu konvergenciju u odnosu na standardnu obuku propagacijom unazad, ali pokazuju slabost u vidu računске složenosti i zavisnosti performanse od usvojenih parametara, što je prevaziđeno kreiranjem algoritma skaliranog konjugovanog gradijenta (SCG) [72].

Moler je predložio sprovođenje SCG obuke prema sledećim koracima [72]:

1. Bira se početni vektor težinskih koeficijenata, w_1 , i parametri $0 < \sigma \leq 10^{-4}$, $0 < \lambda_1 \leq 10^{-6}$, $\bar{\lambda}_1 = 0$. Vršiti se postavljanje $p_1 = r_1 = -E'(w_1)$, $k = 1$ i marker *sucess* = *true*.
2. Ako važi *sucess* = *true*, potrebno je izvršiti izračunavanje:

$$\sigma_k = \sigma/|p_k|,$$

$$s_k = (E'(w_k + \sigma_k p_k) - E'(w_k))/\sigma_k,$$

$$\delta_k = p_k^T s_k.$$
3. Skaliranje δ_k :

$$\delta_k = \delta_k + (\lambda_k - \bar{\lambda}_k)|p_k|^2.$$
4. Ukoliko je $\delta_k \leq 0$, vrše se korekcija kako bi matrica $E''(w)$ bila pozitivno definitna:

$$\bar{\lambda}_k = 2(\lambda_k - \delta_k/|p_k|^2),$$

$$\delta_k = -\delta_k + \lambda_k|p_k|^2,$$

$$\lambda_k = \bar{\lambda}_k.$$
5. Računanje veličine koraka:

$$\mu_k = p_k^T r_k,$$

$$\alpha_k = \mu_k/\delta_k.$$
6. Računanje parametra za poređenje:

$$\Delta_k = 2\delta_k[E(w_k) - E(w_k + \alpha_k p_k)]/\mu_k^2.$$
7. a) Ako je $\Delta_k \geq 0$, može se uspešno izvršiti redukcija greške:

$$w_{k+1} = w_k + \alpha_k p_k,$$

$$r_{k+1} = -E'(w_{k+1}),$$

$$\bar{\lambda}_k = 0,$$
sucess = *true*.
 Ukoliko je k celobrojni umnožak od N , potrebno je restartovati algoritam:

$$p_{k+1} = r_{k+1},$$
 u suprotnom:

$$\beta_k = (|r_{k+1}|^2 - r_{k+1}^T r_k)/\mu_k,$$

$$p_{k+1} = r_{k+1} + \beta_k p_k.$$
 Ukoliko je $\Delta_k \geq 0,75$, potrebno je smanjiti skalirajući parametar:

$$\lambda_k = \frac{1}{4}\lambda_k.$$
- b) Ukoliko je $\Delta_k < 0$:

$$\bar{\lambda}_k = \lambda_k,$$
sucess = *false*.
8. Ukoliko je $\Delta_k < 0,25$, potrebno je povećati skalirajući parametar:

$$\lambda_k = \lambda_k + (\delta_k(1 - \Delta_k)/|p_k|^2).$$
9. Ako je $r_k \neq 0$, postavlja se $k = k + 1$ i algoritam se ponavlja od koraka 2, u suprotnom je w_{k+1} traženi minimum.

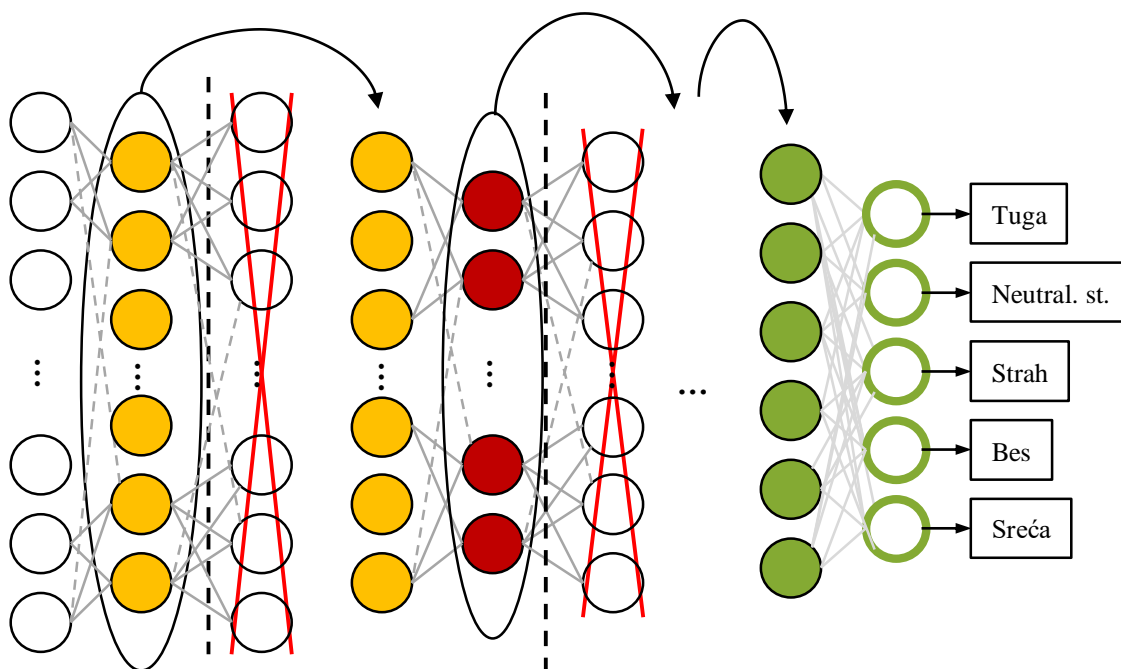
Usvajanjem malih vrednosti σ u prethodnoj proceduri se minimizuje uticaj parametra na performanse algoritma, tako da se ne zahteva zadavanje ni jednog parametra koji je ključan za uspešnost algoritma. Upotrebom prikazane optimizacione strategije se povećava računaska složenost jedne iteracije u odnosu na metod opadajućeg gradijenta, ali je konvergencija brža, tako da se SCG pokazuje kao efikasniji metod obuke ANN.

Bez obzira na odabir optimizacione tehnike, obuka ANN procedurom propagacije unazad podrazumeva lokalnu minimizaciju funkcije greške. Za mali broj slojeva procedura propagacije unazad postiže dobre rezultate, međutim, sa povećanjem broja slojeva postaje izraženije ograničenje lokalne minimizacije, što se odražava na loše performanse konstruisane ANN.

4.3.2 Konstrukcija DNN

Povećanje broja slojeva ANN omogućava značajnu popravku efikasnosti u pogledu broja potrebnih procesorskih elemenata za predstavljanje date funkcije, pri čemu se većim brojem nivoa nelinearnih preslikavanja može postići kompaktna reprezentacija izrazito nelinearnih funkcija sa velikim brojem varijacija. Međutim, povećanje broja slojeva dovodi do problema neadekvatne obuke klasičnim metodama, što je razlog pribegavanju ANN sa manjim brojem slojeva. Poslednjih godina je veća pažnja posvećena ispitivanju mogućnosti uspešnog projektovanja neuralnih mreža sa velikim brojem slojeva – dubokih neuralnih mreža (DNN). Jedan način formiranja DNN biće opisan u nastavku [73].

DNN mogu biti efikasno obučene ukoliko se vrši nezavisna obuka svakog sloja mreže. Do prvog sloja mreže se dolazi konstruisanjem *feedforward* NN sa ulaznim, izlaznim i jednim skrivenim slojem. Broj neurona u ulaznom i izlaznom sloju je jednak dimenziji ulaznih podataka, dok je broj neurona u skrivenom sloju nešto manji. Za aktivacionu funkciju je zgodno izabrati sigmoid. Prilikom obuke se za ulaze i ciljane izlaze koriste parovi istih vrednosti, a obuka se sprovodi upotrebom procedure propagacije unazad sa SCG optimizacijom. Skriveni sloj tako obučene mreže sadrži model ulaznih podataka u prostoru manjih dimenzija, tako da se uklanjanjem izlaznog sloja dolazi do strukture koja vrši kompresiju ulaznih podataka. Na sličan način se formiraju sledeći slojevi, sa upotrebom izlaza prethodnih slojeva kao ulaza. Procedura se ponavlja do željenog broja skrivenih slojeva, pri čemu se sa svakim novim slojem povećava stepen generalizacije ulaznih podataka.



Sl. 4.4 – Konstrukcija DNN

Konačnu odluku sprovodi izlazni sloj koji na osnovu rezultata poslednjeg skrivenog sloja treba da odredi klasnu pripadnost ulaznih podataka. Obuka se sprovodi upotrebom SCG metode propagacije unazad, a za aktivacionu funkciju je zgodno odabrati *softmax*, s obzirom da izlazi dobijeni pomoću *softmax* aktivacione funkcije daju meru sigurnosti date klasifikacije. Analitički oblik *softmax* funkcije za c klasa je:

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^c e^{x_j}}. \quad (4.90)$$

Nakon što se izvrši obučavanje svih sloja mreže ponaosob, vrši se spajanje slojeva i finalno podešavanje težinskih koeficijenata primenom SCG metode propagacije unazad na celu mrežu. Ilustracija formiranja DNN je data na Sl. 4.4.

4.4 Multiklasifikatorski sistemi

Od mnoštva dostupnih klasifikatorskih metoda prepoznaju se prednosti različitih pristupa u zavisnosti od problema koji se rešava. Standardan postupak odabira odgovarajućeg klasifikatora za dati zadatak podrazumeva projektovanje nekoliko različitih klasifikatorskih struktura, određivanje performanse upotrebom skupa dostupnih uzoraka i odabir strukture koja ostvaruje najbolje rezultate. Navedeni pristup može da dovede do dobrih rezultata ukoliko skup dostupnih uzoraka dovoljno dobro oslikava rešavani problem, ali se u praksi lako može desiti da to nije slučaj. Nedovoljno reprezentativni uzorci dovode do loše procene performansi poređenih klasifikatorskih struktura, što u najgorem slučaju može prouzrokovati odabir najlošije strukture. Jednostavnim usrednjavanjem izlaza pojedinačnih klasifikatora se može osigurati postizanje rezultata koji su bolji ili jednaki rezultatu najlošijeg pojedinačnog klasifikatora [74]. Dodatno, adekvatnim kombinovanjem rezultata pojedinačnih klasifikatora je moguće postići performanse koje su bolje od najboljeg pojedinačnog klasifikatora, pri čemu se zahteva da pojedinačni klasifikatori ostvaruju rezultate koji su bolji od slučajnog izbora i da greške koje prave pojedinačni klasifikatori nisu korelisane. U opštem slučaju, stepen uspešnosti formirane multiklasifikatorske strukture zavisi od odabira ansambla pojedinačnih klasifikatora, topologije u koju su organizovani pojedinačni klasifikatori, kao i metode za kombinovanje rezultata pojedinačnih klasifikacija.

4.4.1 Ansambl klasifikatora

Formiranje ansambla klasifikatora je vrlo bitan korak u realizaciji multiklasifikatorske strukture, pri čemu uspešnost pojedinačnih klasifikatora nije od presudnog značaja. Kombinovanjem klasifikatora visokih performansi koji čine iste greške se ne može postići popravka kvaliteta. Kod formiranja ansambla klasifikatora je od suštinske važnosti kombinovanje klasifikatorskih sistema različitih karakteristika, što se postiže varijacijom klasifikatorske metode, skupa obučavajućih uzoraka, obeležja, parametara algoritama, arhitekture, inicijalizacije.

U literaturi su predloženi različiti pokazatelji za poređenje karakteristika klasifikatora [75], a u nastavku će akcenat biti na grupi mera koje vrše analizu parova klasifikatora prebrojavanjem tačnih klasifikacija. Ako je $Z = \{z_1, \dots, z_N\}$ skup uzoraka rešavanog klasifikacionog problema, tada se na osnovu rezultata klasifikatora D_i , $i = 1, \dots, L$, može formirati N -dimenzionalni vektor binarnih

vrednosti $y_i = [y_{1,i}, \dots, y_{N,i}]^T$, tako da je $y_{j,i} = 1$ ako D_i korektno klasifikuje z_j , $j = 1, \dots, N$, a u suprotnom je $y_{j,i} = 0$. Na osnovu vektora y_i i y_k pridruženih klasifikatorima D_i i D_k se može odrediti: broj uzoraka korektno klasifikovanih od strane oba klasifikatora, N^{11} , broj uzoraka pogrešno klasifikovanih od strane oba klasifikatora, N^{00} , broj uzoraka korektno klasifikovanih od strane samo prvog klasifikatora, N^{10} , i broj uzoraka korektno klasifikovanih od strane samo drugog klasifikatora, N^{01} . Prethodno definisane vrednosti omogućavaju poređenje karakteristika datog para klasifikatora prema različitim statističkim pokazateljima.

Q statistika klasifikatora D_i i D_k je određena izrazom:

$$Q_{i,k} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}}, \quad (4.91)$$

sa opsegom mogućih vrednosti $[-1, 1]$. Za statistički nezavisne klasifikatore dobijaju se vrednosti bliske nuli, tendencija ka ispravnom klasifikovanju istih uzoraka utiče na ostvarenje pozitivnih vrednosti izraza, a učestale greške klasifikacije različitih uzorka uzrokuju negativan rezultat.

Koeficijent korelacije izlaza klasifikatora D_i i D_k se može odrediti na osnovu formule:

$$\rho_{i,k} = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{(N^{11} + N^{10})(N^{01} + N^{00})(N^{11} + N^{01})(N^{10} + N^{00})}}. \quad (4.92)$$

Koeficijent korelacije i Q statistika su uvek istog znaka i pri tome važi $|\rho_{i,k}| \leq |Q_{i,k}|$.

Mera neslaganja daje procenu komplementarnosti klasifikatora, a izražena je kao udeo uzoraka koji su korektno klasifikovani od strane samo jednog klasifikatora:

$$Dis_{i,k} = \frac{N^{01} + N^{10}}{N^{11} + N^{10} + N^{01} + N^{00}}. \quad (4.93)$$

Mera duple greške se može koristiti za određivanje parova klasifikatora koji su najmanje spregnuti, a definiše se kao udeo uzoraka koji su pogrešno klasifikovani od strane oba klasifikatora:

$$DF_{i,k} = \frac{N^{00}}{N^{11} + N^{10} + N^{01} + N^{00}}. \quad (4.94)$$

4.4.2 Multiklasifikatorske topologije

Uporedo sa formiranjem ansambla klasifikatora, od velikog značaja je organizacija pojedinačnih klasifikatora u odgovarajuću topologiju. Osnovna podela korišćenih topologija multiklasifikatorskih sistema jeste na: serijsku (kaskadnu), paralelnu i hijerarhijsku.

Kod serijske topologije se pojedinačni klasifikatori primenjuju sukcesivno, pri čemu svaki sledeći klasifikator sužava skup mogućih klasa. Glavna mana ovog pristupa jeste nemogućnost korekcije greške načinjene od strane klasifikatora na početku kaskade.

Paralelna topologija podrazumeva jednovremeno angažovanje svih klasifikatora ansambla, a zatim sumiranje pojedinačnih rezultata u konačnu odluku. Ovo je najviše korišćena i najviše istraživana topologija. Mana paralelne topologije je računaska zahtevnost, gde je neophodno da svi pojedinačni klasifikatori daju svoj rezultat pre donošenja konačne odluke.

Hijerarhijska topologija kombinuje paralelnu i kaskadnu strukturu u cilju postizanja optimalne performanse. Konkretna postavka ove topologije je u značajnoj meri zavisna od problema koji se rešava.

4.4.3 Kombinatorna pravila

Kombinatorna pravila treba da obezbede združenu odluku na osnovu rezultata pojedinačnih klasifikatora ansambla. Da bi kombinatorna pravila mogla da proizvedu dobar zaključak, neophodno je da klasifikatori ansambla budu visokih performansi sa različitim karakteristikama, pri čemu složenost kombinatornih pravila, neophodna za izdvajanje korisnih informacija, direktno zavisi od stepena zadovoljenja postavljenih uslova. Kombinatorna pravila se mogu formirati na proizvoljne načine, a njihovo grupisanje se može izvršiti prema različitim kriterijumima.

Kombinatorna pravila je moguće podeliti prema formi izlaza koja se očekuje od pojedinačnih klasifikatora ansambla, pri čemu se razlikuju: apstraktni izlazi, izlazi u obliku ranga i merljivi izlazi. Klasifikatori sa apstraktnim izlazom kao rezultat daju jednu klasu, a kombinatorna pravila koja koriste ovakvu formu izlaza su najgeneralnija i mogu se primeniti na ansamble proizvoljnih tipova klasifikatora. Sledeću grupu kombinatornih pravila određuju klasifikatori koji kao izlaz obezbeđuju uređenu listu klasa. Izlaz u obliku ranga ima prednost u odnosu na apstraktni izlaz, s obzirom na viši stepen informativnosti, pri čemu postaje moguće kao konačnu odluku prepoznati klasu koja se u pojedinačnim klasifikacijama često javlja blizu vrha rang liste, ali ne na samom vrhu. Treću grupu čine kombinatorna pravila koja koriste merljive izlaze pojedinačnih klasifikatora. Merljivi izlaz klasifikatora određuje stepen verovatnoće za svaku od mogućih klasa, što pruža najveću fleksibilnost kod formiranja kombinatornih pravila. Prilikom kombinovanja klasifikatora sa merljivim izlazom se može javiti problem kod normalizacije izlaza različitih opsega i tipova, pa se u takvim situacijama češće koriste kombinatorna pravila na bazi ranga.

Prema načinu formiranja se kombinatorna pravila mogu podeliti na fiksna i trenirana. Fiksna pravila sprovode konačnu odluku jednostavnim kombinovanjem rezultata pojedinačnih klasifikatora, dok se pod treniranim pravilima podrazumevaju složeni kombinatorni algoritmi koji zahtevaju obuku.

Još jedna moguća podela kombinatornih pravila jeste na integratore i selektore, zavisno od toga da li se do konačne odluke dolazi kombinovanjem rezultata pojedinačnih klasifikatora ili se odluka jednog od klasifikatora koristi kao konačna.

4.5 Mere uspešnosti klasifikatorskih sistema

Nakon što se formira klasifikatorska struktura, izuzetno je važno proceniti performanse projektovanog sistema, odnosno, odrediti meru koliko uspešno je rešen postavljeni zadatak. Performanse klasifikatorskih sistema je moguće izraziti upotrebom različitih pokazatelja: ponderisana tačnost (eng. *weighted accuracy*), neponderisana tačnost (eng. *unweighted accuracy*), preciznost (eng. *precision*), odziv (eng. *recall*), matrica konfuzije (eng. *confusion matrix*).

Najčešće korišćena mera za sagledavanje performansi klasifikatorskih sistema jeste ponderisana tačnost (prosečna uspešnost ili tačnost prepoznavanja). Ova mera govori koliko je ispravnih odluka klasifikator napravio u odnosu na ukupan broj testiranih uzoraka. Za problem sa M klasa i N testiranih uzoraka, ponderisana tačnost je određena izrazom:

$$WA = \frac{n_{11} + \dots + n_{MM}}{N}, \quad (4.94)$$

gde je n_{mm} broj tačno klasifikovanih uzoraka iz klase m , $m = 1, \dots, M$. Data mera je zavisna od stepena zastupljenosti uzoraka različitih klasa, tako da daje dobru procenu performansi za balansirane klase, ali procena postaje loša kod problema sa izraženim disbalansom klasa.

Kako bi se došlo do dobre procene performansi u slučaju nebalansiranih skupova uzoraka različitih klasa, računa se neponderisana tačnost prema formuli:

$$UA = \frac{n_{11}/N_1 + \dots + n_{MM}/N_M}{M}, \quad (4.95)$$

pri čemu je N_m broj uzoraka iz klase m , $m = 1, \dots, M$. U slučaju balansiranih klasa, ponderisana i neponderisana tačnost daju isti rezultat. U ovom radu su testiranja vršena sa dosta dobro balansiranim skupovima uzoraka, tako da je računata ponderisana tačnost bez posebnog isticanja koja mera je korišćena.

Kao mere performansi klasifikatorskih sistema se mogu koristiti preciznost i odziv, pri čemu se ove dve mere najčešće koriste u paru. Preciznost za klasu m se definiše kao udeo tačno klasifikovanih uzoraka klase m u ukupnom broju uzoraka klasifikovanih u klasu m . Odziv za klasu m se definiše kao odnos tačno klasifikovanih uzoraka klase m i ukupnog broja uzoraka klase m . Moguće je odrediti usrednjene vrednosti preciznosti i odziva, čime se dolazi do prethodno definisanih mera tačnosti.

Sveobuhvatna mera kvaliteta klasifikatorskih sistema jeste matrica konfuzije, iz koje se direktno mogu izraziti sve prethodno opisane mere. Matrica konfuzije je kvadratna matrica, dimenzija jednakih broju klasa, u kojoj su na pregledan i detaljan način prikazane ispravne i pogrešne klasifikacije:

$$CM = \begin{bmatrix} n_{11} & \dots & n_{1M} \\ \vdots & \ddots & \vdots \\ n_{M1} & \dots & n_{MM} \end{bmatrix}, \quad (4.96)$$

gde n_{ij} predstavlja broj uzoraka iz klase $i = 1, \dots, M$ klasifikovanih u klasu $j = 1, \dots, M$.

5 Analiza sistema za prepoznavanje govornih emocija

Interesovanje za automatsko prepoznavanje emocija iz govora raste sa povećanjem mogućnosti za praktičnu primenu. Informacije o emocionalnim stanjima mogu doprineti kvalitetu različitih sistema omogućavanjem adaptacije trenutnom emocionalnom stanju korisnika [32], [76], [77]. Pored toga, težnja ka ostvarenju prirodne interakcije nameće kao nužno prilagođenje u pogledu jezika i kulture. S obzirom da jezik utiče na karakteristike govora [78] i da se emocije bolje prepoznaju među pripadnicima istog govornog područja [79], važno je posvetiti pažnju načinu ispoljavanja emocija u različitim jezicima i kulturama.

Ekperimenti automatskog prepoznavanja emocija su u dosadašnjim istraživanjima retko izvođeni nad slovenskim jezicima, za šta se objašnjenje može tražiti u manjoj zastupljenosti ovih jezika, kao i nedostatku i dostupnosti istraživačkog materijala. Baze odglumljenog emocionalnog govora postoje za srpski (GEES) [80], poljski (PES) [81], [82], [83] i ruski (RUSLANA) [84] jezik, dok su baze spontanog emocionalnog govora prisutne na hrvatskom (CrES) [85], češkom (CzED) [86] i slovenačkom (EmoLUKS) [87] jeziku.

Ekperimenti klasifikacije slovenskog emocionalnog govora su najčešće vršeni sa bazom na srpskom jeziku. Primenom diskretnog HMM kao klasifikatora i skupa obeležja sačinjenog od LFPC, LPCC, energije (E), Tajger energije (TE), osnovne učestanosti (F0) i formanata (FF) je postignuto 72% tačnih klasifikacija [88]. U eksperimentima sa SVM klasifikatorom rezultati variraju od 62,78% do 91,3%, zavisno od korišćene postavke [89], [90]. Rezultati ostvareni sa bazom poljskog emocionalnog govora su raznoliki, gde se sa kNN (eng. *k Nearest Neighbours*) klasifikatorom i MFCC obeležjima došlo do 50,73% tačnih prepoznavanja [91], dok je upotrebom BDT (eng. *Binary Decision Tree*) klasifikatora i karakteristika na nivou fonema realizovano 81,9% uspešnih klasifikacija [92]. Upotrebom SVM klasifikatora i MFCC obeležja postignuta je uspešnost klasifikacije 40,5%, a sa uključenim delta koeficijentima je ostvareno 33,75% tačnih klasifikacija [93]. U testovima sa bazom hrvatskog emocionalnog govora je realizovano 65,4% tačnih klasifikacija upotrebom RF (eng. *Random Forest*) klasifikatora [94]. Ekperimenti sa bazom ruskog emocionalnog govora su se odnosili na ispitivanje akustičkih karakteristika različitih fonema u emocionalnom govoru [95], dok ekperimenti automatskog prepoznavanja emocija još uvek nisu vršeni primenom baza ruskog, češkog i slovenačkog emocionalnog govora. Detaljniji prikaz sistema za automatsko prepoznavanje govornih emocija, testiranih sa slovenskim bazama emocionalnog govora, je dat u Tab. 5.1.

Istraživanja u oblasti prepoznavanja govornih emocija su poslednjih godina uglavnom bazirana na generalnom skupu obeležja definisanom u „INTERSPEECH 2009 Challenge“ [34], gde se izdvajanje najčešće vrši pomoću „openSMILE:“ paketa [96]. Generalni je stav da su za zadatak prepoznavanja govornih emocija spektralna obeležja bolji izbor od prozodijskih [9], a do sličnih zaključaka se došlo u istraživanjima sa bazama poljskog [93] i srpskog [97] emocionalnog govora. Ako se razmatra izbor klasifikatorskih tehnika, uočavaju se prednosti i mane različitih pristupa, ali se ne mogu izvoditi zaključci o univerzalno najboljem rešenju [36].

Tab. 5.1 – Sistemi za automatsko prepoznavanje govornih emocija testirani sa slovenskim bazama

| Jezik | Skup obeležja | Klasifikator | Emocije | Referenca |
|----------|--|---|--|-------------|
| srpski | LPCC, LFPC, E, TE, F0, FF | HMM | bes, strah, sreća, tuga, neutralno stanje | [88] |
| | MFCC, E, F0, FF, harmoničnost, glasnoća, obeležja govornog izvora, obeležja trajanja | SVM | bes, strah, sreća, tuga, neutralno stanje | [98], [99] |
| | openEAR skup obeležja | SVM | bes, strah, sreća, tuga, neutralno stanje | [89] |
| | MFCC | SVM | bes, strah, sreća, tuga, neutralno stanje | [90] |
| | MFCC, vremenska diskretna kosinusna transformacija | ANN | bes, strah, sreća, tuga, pretnja | [100] |
| | MFCC, E, F0 | kNN, linearni Bajes | bes, strah, sreća, tuga, neutralno stanje | [97], [101] |
| hrvatski | MFCC, E, F0, FF, obeležja govornog izvora, lingvistička obeležja | SVM, RF | bes, strah, sreća, tuga, neutralno stanje | [102], [94] |
| poljski | E, F0, FF, ZCR, obeležja trajanja | BDT | bes, strah, sreća, tuga, dosada, neutralno stanje | [92] |
| | MFCC, HFCC (eng. <i>Human Factor Cepstral Coefficients</i>) | kNN | bes, strah, sreća, tuga, dosada, neutralno stanje | [91] |
| | E, F0, FF, spektralna perceptualna obeležja (MFCC, HFCC i druga) | kNN, SVM | bes, strah, sreća, tuga, iščekivanje, iznenađenje, gađenje, neutralno stanje | [93] |
| | LPC, E, F0, FF | SVM, kNN, BDT, LDA, ANN, linearni Bajes | bes, strah, sreća, tuga, iznenađenje, gađenje, neutralno stanje | [103] |

U ovoj glavi će biti sprovedena analiza spektralnih obeležja LPCC, LFPC i MFCC u kombinaciji sa najčešće korišćenim klasifikatorskim metodama zasnovanim na HMM, SVM i DNN [104]. Testiranja će biti vršena upotrebom slovenskih baza srpskog, poljskog i ruskog emocionalnog govora. Takođe, testovi će biti sprovedeni i sa bazom nemačkog emocionalnog govora (Berlin) [105] u cilju obezbeđenja rezultata za komparaciju. Svrha eksperimenata jeste:

1. Određivanje robusne i pouzdane ocene obeležja, klasifikatora i njihovih kombinacija.
2. Određivanje najpodobnijih konfiguracija sistema za upotrebu sa različitim jezicima.
3. Određivanje zavisnosti rezultata klasifikacije od karakteristika baze.

U nastavku će biti nešto više reči o korišćenim bazama emocionalnog govora, zatim će biti detaljnije opisane postavke eksperimenata, a na kraju će biti prikazani rezultati testiranja sa izvedenim zaključcima.

5.1 Baze emocionalnog govora

Rezultati eksperimenata, kao i kvalitet izvedenih zaključaka, u velikom meri zavise od karakteristika korišćene baze emocionalnog govora. Potrebno je da korišćeni materijal što izražajnije oslikava sadržano emocionalno stanje, a istovremeno se teži što prirodnijoj prezentaciji. Postavljene zahteve je teško jednovremeno zadovoljiti, tako da se otvara pitanje odabira odgovarajućeg tipa baze emocionalnog govora. Zavisno od načina nastanka snimljenog materijala je moguće razlikovati baze spontanog, indukovanog i odglumljenog emocionalnog govora. Baze spontanog emocionalnog govora sadrže snimke nastale u realnim situacijama, što ih čini najtežim za formiranje. Ove baze su skopčane sa problemima autorskih prava, kvaliteta snimljenog materijala, kao i nužnošću označavanja snimljenog materijala. Baze indukovanog emocionalnog govora se zasnivaju na ideji dovođenja govornika u određeno emocionalno stanje, upotrebom različitih stimulusa. Kod ovih baza je, takođe, neophodno označavanje, s obzirom na individualne razlike u percepciji stimulusa. Pored toga, snimljena emocionalna stanja su slabog intenziteta i često pomešana. Baze odglumljenog emocionalnog govora najčešće nastaju snimanjem profesionalnih glumaca koji su u stanju da verno prikažu emocionalno stanje kroz verbalno izražavanje. Ove baze su najlakše za snimanje, pri čemu se dobija čist, uređen, jasno označen materijal. Kao mana ovih baza se navodi činjenica da se vrši prototipski prikaz emocija, bez oslikavanja finih promena glasa. Takođe, kod ovih baza su emocije jasno prikazane što u realnim situacijama uglavnom nije slučaj. Na osnovu iznetih činjenica se može zaključiti da odabir tipa baze dominantno zavisi od cilja istraživanja. Ukoliko je potrebno izvršiti evaluaciju datog sistema u uslovima bliskim realnoj primeni, zgodno je odabrati bazu spontanog govora. Sa druge strane, u fazi razvoja novih algoritama, kada se sagledavaju različite zakonitosti, prikladnije su baze odglumljenog emocionalnog govora.

U nastavku će biti nešto više reči o odglumljenim bazama emocionalnog govora korišćenim u eksperimentima.

5.1.1 Baza srpskog emocionalnog govora

Baza srpskog emocionalnog govora (GEES – Govorna Ekspresija Emocija i Stavova) [80] je snimljena u akustički obrađenom studiju Fakulteta dramskih umetnosti u Beogradu. U snimanju je učestvovalo šest studenata Fakulteta dramskih umetnosti (tri devojke i tri mladića). Svaki govornik je imao zadatak da semantički neutralan tekst pročita u skladu sa pet emocionalnih stanja: bes, sreća, strah, tuga i neutralno stanje. Tekst se sastojao od 32 izolovane reči, 30 kratkih rečenica, 30 dugih rečenica i jednog pasusa. Svaki govornik je snimljen u posebnoj sesiji kako bi se izbegao međusobni uticaj na stil govora. Evaluacija snimljenog materijala je vršena od strane 30 studenata sa ukupno 95% uspešnih prepoznavanja, a za različita emocionalna stanja se uspešnost prepoznavanja kretala od 93,33% do 96,06%. Kompletan materijal je snimljen na magnetnu traku (DAT, eng. *Digital Audio Tape*) upotrebom visokokvalitetnog mikrofona sa učestanošću odabiranja 44,1 kHz, a naknadno je vršena filtracija propusnikom opsega od 70 Hz do 10 kHz i snimanje u *wav* fajlove sa učestanošću odabiranja 22,05 kHz.

5.1.2 Baza poljskog emocionalnog govora

Baza poljskog emocionalnog govora (PES – eng. *Polish Emotional Speech*) [81] je nastala u auli Poljske nacionalne škole filma, televizije i pozorišta iz Lođa, snimanjem govora četiri glumice i četiri glumca, studenata navedene škole. Svaki govornik je izgovarao pet semantički neutralnih rečenica u skladu sa šest različitih emocionalnih stanja: sreća, dosada, strah, bes, tuga i neutralno stanje. Nakon što je 50 slušalaca izvršilo evaluaciju snimljenog materijala, konstatovano je ukupno 72% tačnih prepoznavanja, pri čemu se za različite govornike uspešnost kretala od 60% do 84%. Snimci su napravljeni pomoću kondenzatorskog mikrofona, a čuvanje je izvršeno u *wav* formatu sa učestanošću odabiranja 44,1 kHz.

5.1.3 Baza ruskog emocionalnog govora

Baza ruskog emocionalnog govora (RUSLANA – eng. *RUSsian LANguage Affective speech*) [84] je kreirana u cilju obezbeđenja materijala za istraživanja lingvističke analize i obrade emocionalno obojenog govora. Baza sadrži snimke govora nativnih govornika ruskog jezika, studenata Lingvističkog fakulteta Univerziteta u Sankt Peterburgu, pri čemu je u snimanju učestvovalo 49 govornika ženskog pola i 12 govornika muškog pola. Svaki govornik je izgovarao 10 predefinisanih rečenica u skladu sa šest emocionalnih stanja: iznenađenje, sreća, bes, tuga, strah i neutralno stanje. Snimanje je vršeno u zvučno izolovanom studiju Odseka za fonetiku Univerziteta u Sankt Peterburgu pomoću „Sennheiser“ slušalica i prenosivog DAT snimača „SONY TCD-D8“ sa učestanošću odabiranja 48 kHz. Naknadno je izvršena konverzija snimaka u *pcm* format sa učestanošću odabiranja 32 kHz. Nije vršena zbirna procena kvaliteta snimljenog materijala, već su korišćena dva pristupa evaluacije pojedinačnih zapisa. Jedan pristup evaluacije zapisa se zasnivao na postavljanju svakog zapisa u prostor sa osama valentnosti i aktivacije, dok je drugi pristup evaluacije podrazumevao ocenjivanje zapisa prema vernosti prikaza emocionalnog stanja.

5.1.4 Baza nemačkog emocionalnog govora

Baza nemačkog emocionalnog govora (Berlin ili EmoDB – eng. *Emotional speech DataBase*) [105] je snimljena u anehoičnoj komori Tehničkog univerziteta u Berlinu. U snimanju je učestvovalo pet govornika ženskog pola i pet govornika muškog pola, od kojih je devet profesionalnih glumaca. Svaki govornik je 10 predefinisanih rečenica izgovorio proizvoljan broj puta, a u skladu sa sedam emocionalnih stanja: bes, sreća, strah, tuga, dosada, gađenje i neutralno stanje. Evaluaciju snimljenog materijala je vršilo 20 slušalaca sa zadatkom prepoznavanja i ocene vernosti prikaza emocionalnog stanja. Prosečna uspešnost prepoznavanja različitih emocionalnih stanja je varirala od 79,6% do 96,9%. Iz baze su izostavljeni snimci koji nisu zadovoljili uspešnost prepoznavanja iznad 80% i prirodnost izražavanja iznad 60%. Za snimanje baze je korišćen „Sennheiser MKH 40 P 48“ mikrofoni i „Tascam DA-P1“ prenosivi DAT snimač. Snimci su napravljeni sa učestanošću odabiranja 48 kHz, a kasnije je izvršena redukcija učestanosti odabiranja na 16 kHz.

Na osnovu iznetih opisa se uočavaju različite svojstvenosti korišćenih baza. Pored razlike u jeziku, prisutne su razlike u simuliranim emocionalnim stanjima, načinu snimanja, tipu validacije, broju govornika i količini snimljenog materijala. GEES bazu karakteriše mali broj govornika i veliki broj zapisa po govorniku, PES bazu odlikuje mala količina zapisa, dok RUSLANA sadrži veliki broj govornika i veliki broj snimaka po emociji. U eksperimentima sa GEES bazom su korišćeni snimci samo dugih rečenica, dok su u eksperimentima sa bazama PES, RUSLANA i Berlin korišćeni svi zapisi za emocionalna stanja sreće, besa, tuge, straha i neutralnog stanja.

5.2 Postavka eksperimenata

Zadatak sistema za automatsko prepoznavanje govornih emocija jeste da za dati ulaz u vidu govornog signala kao izlaz pruži informaciju o sadržanom emocionalnom stanju. Dizajn sistema za prepoznavanje emocija podrazumeva definisanje dva osnovna dela: metode za izdvajanje obeležja iz govornog signala i metode za klasifikaciju na osnovu izdvojenih obeležja. Prilikom izdvajanja obeležja je potrebno usvojiti strategiju predobrade govornog signala (filtracija, predisticanje, prozorovanje) i postobrade obeležja (normalizacija, izvod, statistike). U dizajn sistema se dodatno može uključiti redukcija skupa obeležja, kao i fuzija rezultata većeg broja klasifikatora.

U ovom radu je izdvajanje LPCC, LFPC i MFCC spektralnih obeležja vršeno prema procedurama opisanim u glavi 3. Na nivou svakog frejma je formiran vektor kratkovremenih obeležja, sačinjen od izdvojenih obeležja, delta i delta-delta koeficijenata. Reprezentacija govorne sekvence je vršena nizom vektora kratkovremenih obeležja ili jednim vektorom globalnih obeležja, pri čemu su globalna obeležja izvođena iz kratkovremenih primenom osnovnog skupa statistika definisanog u poglavlju 3.6 (srednja vrednost, standardna devijacija, minimum, maksimum i opseg).

U ulozi klasifikatorskih metoda su primenjivani diskretni HMM, SVM i DNN, pri čemu je, u skladu sa osobenostima klasifikatora, korišćena reprezentacija govorne sekvence u obliku niza vektora kratkovremenih obeležja (HMM) ili vektora globalnih obeležja (SVM i DNN). Teorijske osnove selektovanih klasifikatora su date u glavi 4, a u nastavku će biti date specifičnosti konfiguracija korišćenih u eksperimentima.

Struktura diskretnog HMM je određena brojem skrivenih stanja, načinom njihovog povezivanja i veličinom kodovnika. Po ugledu na ranija istraživanja [36], [38], u eksperimentima je korišćena konfiguracija HMM sa četiri potpuno povezana skrivena stanja i kodovnikom od 64 simbola, a etaloni za potrebe vektorske kvantizacije su formirani upotrebom *K-means* klasterizacije. Pre koraka vektorske kvantizacije su kratkovremena obeležja svake govorne sekvence normalizovana na opseg [0, 100]. Parametri modela su inicijalizovani slučajnim vrednostima, a zatim je sprovedeno iterativno podešavanje parametara upotrebom Baum-Velč algoritma sa uključenim skaliranjem i proširenjem za višestruke sekvence opservacija. Tokom podešavanja parametara je vršeno ograničavanje kako bi se sprečilo postojanje suviše malih verovatnoća u modelu. Za određivanje stepena slaganja modela sa datom govornom sekvencom je korišćena procedura računanja unapred.

U eksperimentima je korišćen nelinearni SVM sa mekom marginom, tako da je bilo potrebno definisati kernel funkciju, kao i balans između maksimizacije margine i minimizacije greške

modelovanja. Za kernel funkciju je odabran polinom trećeg reda $K(x_i, x_j) = (\langle x_i \cdot x_j \rangle + 1)^3$, a za balans faktor je usvojeno $C = 1$. S obzirom na potrebu za razdvajanje većeg broja klasa, organizovana su binarna razdvajanja za sve parove klasa, a konačan rezultat je određivan glasanjem.

Kod DNN klasifikatora je bilo neophodno specificirati broj slojeva i broj neurona u svakom od njih. Ulazni sloj je sačinjavalo 180 neurona u skladu sa veličinom vektora globalnih obeležja. Broj čvorova izlaznog sloja je bilo pet, što je odgovaralo broju klasa. Skriveni slojevi su organizovani sa idejom postepene generalizacije ulaznih podataka, tako da su formirana četiri skrivena sloja sa 160, 120, 80 i 40 neurona.

Analiza izabranog dizajna je sprovedena kroz korake obuke i testiranja, za šta je bilo neophodno obezbediti skupove obučavajućih i testirajućih uzoraka. Obuka je podrazumevala podešavanje parametara klasifikatora na osnovu obeležja izdvojenih iz obučavajućeg skupa, dok je testiranje podrazumevalo izdvajanje obeležja iz testirajućih uzoraka i donošenje odluke o klasnoj pripadnosti u skladu sa izlazom klasifikatora. Procena valjanosti dizajna je vršena na osnovu broja korektnih klasifikacija.

Moguće je razlikovati različite tipove testiranja u zavisnosti od načina formiranja obučavajućih i testirajućih skupova. Strukture korišćenih baza su dozvoljavale organizovanje testiranja u postavkama zavisno do govornika (SD, eng. *Speaker Dependent*) i nezavisno od govornika (SI, eng. *Speaker Independent*). Kod SD postavke je 83% uzoraka od svakog govornika korišćeno za obuku, a preostalih 17% je korišćeno za testiranje, tako da su u testiranju učestvovali poznati govornici, dok je leksički sadržaj bio nepoznat. Kod SI postavke su svi uzorci jednog govornika korišćeni za testiranje, dok su uzorci svih preostalih govornika korišćeni za obuku, i na taj način je leksički sadržaj tokom testiranja bio poznat, ali je govornik bio nepoznat. U SI postavci su obuka i testiranje sprovedeni više puta, u skladu sa brojem mogućnosti za formiranje obučavajućih i testirajućih skupova, a konačan rezultat testiranja je određivan na osnovu svih pojedinačnih testova.

Eksperimenti su realizovani kombinovanjem tri metode za izdvajanje obeležja (LPCC, LFPC, MFCC) sa tri klasifikatorske metode (HMM, SVM, DNN) i testiranjem sa četiri baze emocionalnog govora (GEES, PES, RUSLANA, Berlin) u dve postavke (SD, SI).

5.2.1 Koeficijenti uspešnosti

Sprovođenje predloženih eksperimenata uzrokuje generisanje velikog broja numeričkih rezultata, čiju analizu je moguće lakše sprovести ukoliko se uvedu mere uspešnosti različitih elemenata klasifikatorskog sistema. Ako je $R(f, c, d)$ trodimenzionalna matrica tačnosti prepoznavanja formirana upotrebom obeležja f , klasifikatora c i baze d , pri čemu je: $f \in \{LPCC, LFPC, MFCC\}$, $c \in \{HMM, SVM, DNN\}$ i $d \in \{GEES, PES, RUSLANA, Berlin\}$, tada se koeficijenti uspešnosti obeležja, klasifikatora i njihovih kombinacija mogu definisati pomoću formula datih u nastavku:

1. Koeficijent uspešnosti obeležja:

$$Q_1(f, d) = \frac{R(f, HMM, d) + R(f, SVM, d) + R(f, DNN, d)}{3}. \quad (5.1)$$

2. Koeficijent ukupne uspešnosti obeležja:

$$Q_2(f) = \frac{Q_1(f, \text{GEES}) + Q_1(f, \text{PES}) + Q_1(f, \text{RUSLANA}) + Q_1(f, \text{Berlin})}{4}. \quad (5.2)$$

3. Koeficijent uspešnosti klasifikatora:

$$Q_3(c, d) = \frac{R(\text{LPCC}, c, d) + R(\text{LFPC}, c, d) + R(\text{MFCC}, c, d)}{3}. \quad (5.3)$$

4. Koeficijent ukupne uspešnosti klasifikatora:

$$Q_4(c) = \frac{Q_3(c, \text{GEES}) + Q_3(c, \text{PES}) + Q_3(c, \text{RUSLANA}) + Q_3(c, \text{Berlin})}{4}. \quad (5.4)$$

5. Koeficijent uspešnosti para obeležje-klasifikator:

$$Q_5(f, c) = \frac{R(f, c, \text{GEES}) + R(f, c, \text{PES}) + R(f, c, \text{RUSLANA}) + R(f, c, \text{Berlin})}{4}. \quad (5.5)$$

5.3 Rezultati testiranja

Varijetet testova je realizovan kombinovanjem obeležja, klasifikatora i baza u SD i SI postavkama, a ostvarene tačnosti prepoznavanja su iskazane u procentima i sumirane u Tab. 5.2. Valjanost postignutih rezultata se može verifikovati poređenjem sa rezultatima prethodnih istraživanja u kojima su sprovedeni slični testovi. U ranijim eksperimentima sa GEES bazom u SI postavci je upotrebom MFCC obeležja i hijerarhijskog SVM klasifikatora ostvareno 63.90% uspešnih klasifikacija [98], [99], dok je na ovom mestu par MFCC-SVM zabeležio rezultat 66,78%. Testiranjem MFCC-SVM para sa PES bazom se u ranijim istraživanjima došlo do 33,75% tačnih prepoznavanja [93], dok su u ovom radu postignuti rezultati 70,00% u SD postavci i 55,50% u SI postavci. U slučaju Berlin baze se može vršiti poređenje sa ranijim rezultatima ostvarenim upotrebom neuralnih mreža kao klasifikatora, pri čemu je u prethodnim istraživanjima razmatrano sedam emocionalnih stanja. Vršena su testiranja sa skupom obeležja sačinjenim od MFCC i prozodijskih obeležja, dok su u ulozi klasifikatora korišćena dva tipa neuralnih mreža [106]. U testovima sa mrežom dubokog verovanja (eng. *deep belief network*) je postignuto 69,14% tačnih klasifikacija u SD postavci, a u SI postavci 60,32%. Upotrebom višeslojnog perceptrona (eng. *multilayer perceptron*) su ostvareni rezultati 68,10% u SD postavci i 51,65% u SI postavci. Navedeni rezultati se mogu porediti sa rezultatima para MFCC-DNN iz ovog rada, gde je postignuto 66,44% uspešnih prepoznavanja u SD postavci i 66,32% u SI postavci.

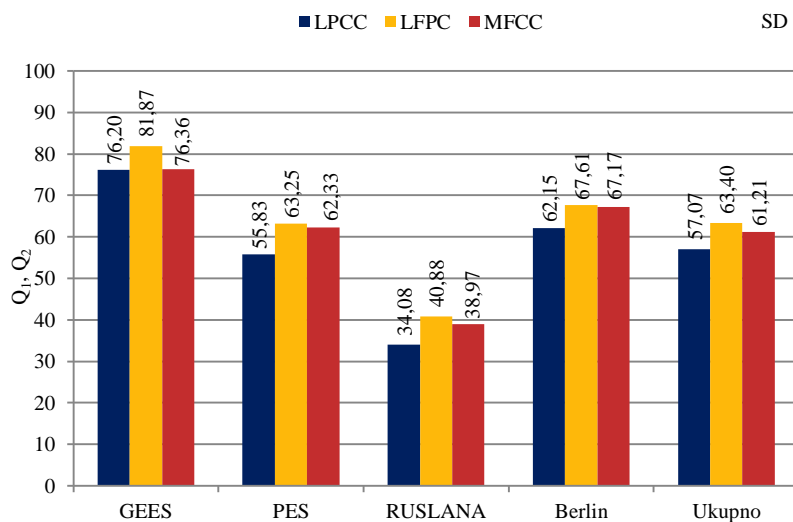
Tab. 5.2 – Rezultati testiranja

| Postavka | Klasifikator | Obeležje | Baza | | | |
|----------|--------------|----------|--------------|--------------|--------------|--------------|
| | | | GEES | PES | RUSL | Berlin |
| SD | HMM | LPCC | 72,41 | 52,75 | 32,56 | 54,41 |
| | | LFPC | 72,41 | 54,75 | 32,89 | 63,24 |
| | | MFCC | 60,69 | 52,75 | 31,40 | 72,06 |
| | SVM | LPCC | 79,33 | 57,50 | 34,43 | 65,75 |
| | | LFPC | 85,33 | 67,50 | 46,07 | 69,86 |
| | | MFCC | 83,33 | 70,00 | 43,11 | 63,01 |
| | DNN | LPCC | 76,87 | 57,25 | 35,26 | 66,30 |
| | | LFPC | 87,87 | 67,50 | 43,69 | 69,72 |
| | | MFCC | 85,07 | 64,25 | 42,41 | 66,44 |
| SI | HMM | LPCC | 57,24 | 52,50 | 30,31 | 55,22 |
| | | LFPC | 65,18 | 49,75 | 31,73 | 62,50 |
| | | MFCC | 53,45 | 44,50 | 30,27 | 59,73 |
| | SVM | LPCC | 65,11 | 55,00 | 34,00 | 59,80 |
| | | LFPC | 66,67 | 55,50 | 43,31 | 65,93 |
| | | MFCC | 66,78 | 55,50 | 42,92 | 62,25 |
| | DNN | LPCC | 63,46 | 50,50 | 30,58 | 59,83 |
| | | LFPC | 65,39 | 56,00 | 42,16 | 67,72 |
| | | MFCC | 67,85 | 54,00 | 41,23 | 66,32 |

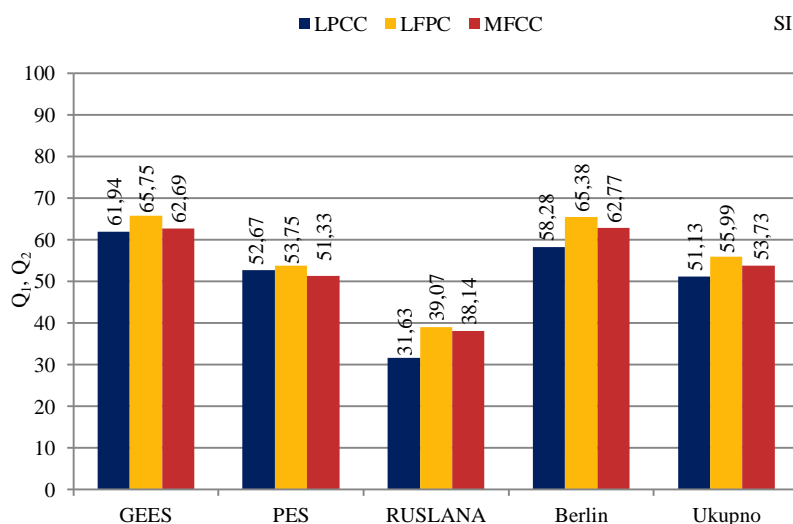
Za svaku od baza je izvršeno poređenje postignutih rezultata i u Tab. 5.2 su istaknute najviše vrednosti u SD i SI postavci, ali se ne može izvesti jasan zaključak o konfiguraciji koja ostvaruje univerzalno najbolje performanse. Ukoliko se razmatraju rezultati postignuti u SD postavci, može se konstatovati da među konfiguracijama testiranim sa Berlin bazom najbolji rezultat ostvaruje par MFCC-HMM, ali isti par beleži najlošije rezultate u slučaju slovenskih baza. Uočena nesaglasnost predstavlja indikaciju potencijalnog uticaja karakteristika baze i kulturološke pozadine ispoljavanja emocija na odabir optimalne konfiguracije sistema. Istovremeno, navedeni primer ukazuje na nužnost dodatne obrade postignutih rezultata pre donošenja opštih zaključaka. U nastavku će biti napravljena analiza rezultata iz Tab. 5.2 upotrebom koeficijenata uspešnosti definisanih u odeljku 5.2.1.

5.3.1 Uspešnost obeležja

Koeficijenti uspešnosti obeležja, Q_1 , i ukupne uspešnosti obeležja, Q_2 , su prikazani na Sl. 5.1 za SD postavku i na Sl. 5.2 za SI postavku. U SD postavci se LFPC obeležje pokazalo najefikasnijim za sve baze, iako u slučaju baza PES i Berlin samo neznatno nadmašuje MFCC (Sl. 5.1). LFPC obeležje je postiglo najviše vrednosti koeficijenata uspešnosti i u SI postavci (Sl. 5.2), tako da se može izvesti opšti zaključak da se među analiziranim obeležjima izdvaja LFPC kao najpodobnije za postavljene zadatke.



Sl. 5.1 – Koeficijenti uspešnosti obeležja u SD postavci

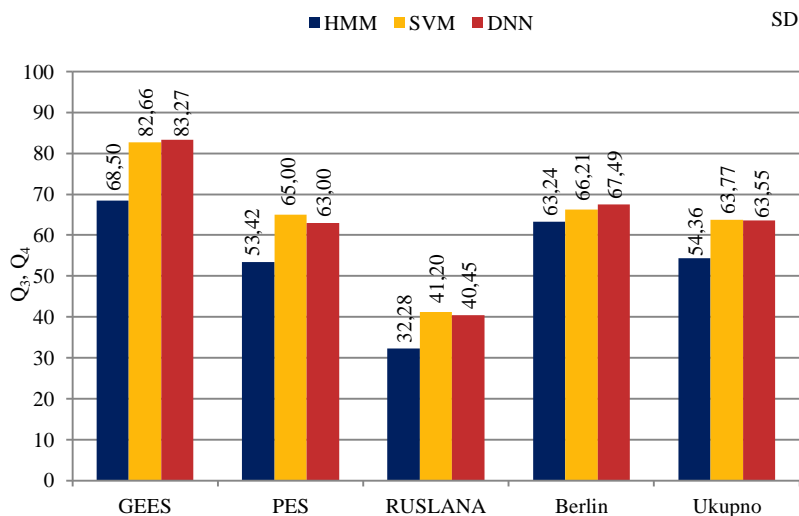


Sl. 5.2 – Koeficijenti uspešnosti obeležja u SI postavci

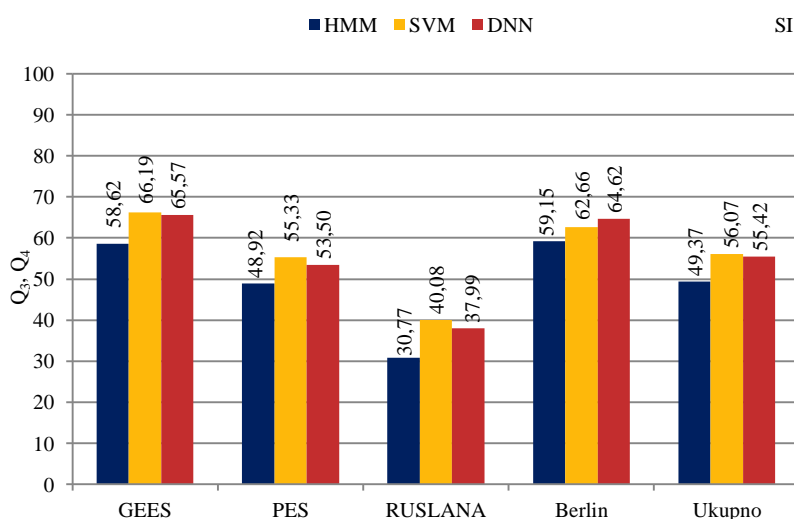
5.3.2 Uspešnost klasifikatora

Koeficijenti uspešnosti klasifikatora, Q_3 , i ukupne uspešnosti klasifikatora, Q_4 , su prezentovani na Sl. 5.3 za SD postavku i Sl. 5.4 za SI postavku. U SD postavci se SVM klasifikator izdvaja kao najpovoljniji izbor za baze PES i RUSLANA, dok je kod baza GEES i Berlin najefikasniji DNN klasifikator, pri čemu su u slučaju GEES baze uspešnosti SVM i DNN skoro izjednačene (Sl. 5.3). Do sličnih zaključaka se dolazi i kod SI postavke, gde je za baze PES i RUSLANA najbolji izbor SVM klasifikator, dok je kod Berlin baze najpodobniji DNN klasifikator (Sl. 5.4). Ponovo su u slučaju GEES baze rezultati SVM i DNN izjednačeni, ali ovog puta SVM klasifikator ostvaruje nešto višu vrednost koeficijenta uspešnosti. Ako se razmatra univerzalno najefikasnije rešenje, zaključak je da se u obe postavke (SD i SI) izdvajaju SVM i DNN sa približnim ukupnim koeficijentima uspešnosti, pri čemu je mala razlika u korist SVM (Sl. 5.3 i Sl. 5.4). Izračunati koeficijenti uspešnosti

ukazuju da HMM ostvaruje najlošije performanse među ispitivanim klasifikatorima i pored toga što je učestvovao u pojedinačnim konfiguracijama koje postižu dobre rezultate.



Sl. 5.3 – Koeficijenti uspešnosti klasifikatora u SD postavci



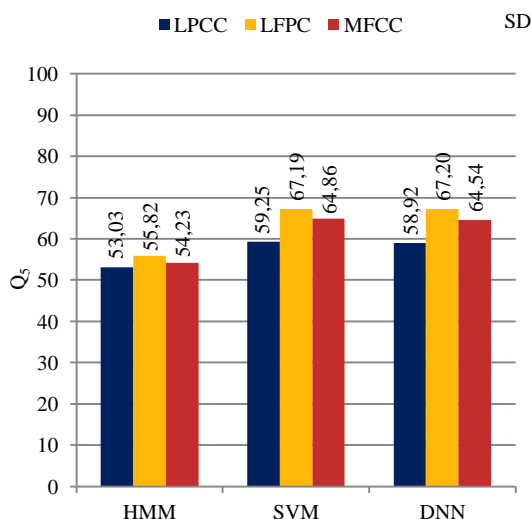
Sl. 5.4 – Koeficijenti uspešnosti klasifikatora u SI postavci

Ukoliko se analiziraju amplitude koeficijenata uspešnosti (Sl. 5.1 do Sl. 5.4), evidentno je da su sa GEES bazom ostvarene najviše vrednosti, ali i najveći pad kada se poredi SD i SI postavke. Nasuprot tome, sa RUSLANA bazom su zabeležene najniže vrednosti koeficijenata uspešnosti kao i najmanja razlika između amplituda SD i SI postavki. GEES bazu karakteriše mali broj govornika i veliki broj uzoraka po govorniku u odnosu na ostale baze, dok je za RUSLANA bazu karakterističan veliki broj govornika, tako da uočene relacije koeficijenata uspešnosti sugerišu dominantan uticaj karakteristika govornika na izražavanje emocija. To ostavlja prostor za detaljniju analizu emocionalnog izražavanja u budućim istraživanjima i razdvajanje informacija vezanih za emocionalna stanja od informacija karakterističnih za govornike.

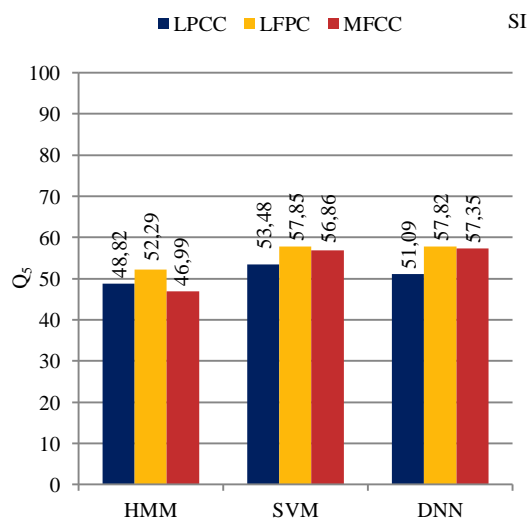
5.3.3 Uspešnost parova obeležje-klasifikator

Do sada analizirani rezultati ukazuju da obeležja i klasifikatori koji ostvaruju dobre rezultate u eksperimentima sa pojedinačnim bazama ne predstavljaju nužno najbolji izbor kada se razmatraju sve baze. Problem je očigledno veoma složen, s obzirom da uspeh sistema zavisi od izbora obeležja i klasifikatora, ali i osobenosti raspoloživih baza. U želji da se prepoznaju konfiguracije sistema koje daju dobre rezultate nezavisno od izbora baze, izračunati su koeficijenti uspešnosti svih parova obeležje-klasifikator, Q_5 , u SD i SI postavkama, a rezultati su sumirani na Sl. 5.5 i Sl. 5.6.

Na osnovu izračunatih koeficijenata uspešnosti (Sl. 5.5 i Sl. 5.6) je jasno da LFPC obeležje obezbeđuje maksimalnu performansu za dati izbor klasifikatora. Sledeći predlog je MFCC u slučaju svih klasifikatora, osim HMM u SI postavci, gde je drugi najbolji izbor LPCC. Sa druge strane, za fiksirano obeležje se ne prepoznaje značajna razlika između konfiguracija koje koriste SVM i DNN, tako da se ne može izvesti jedinstven zaključak u pogledu izbora klasifikatora. Parovi LFPC-SVM i LFPC-DNN se izdvajaju kao jednako podobni kako za SD tako i SI postavku, a u slučaju SI postavke se izbor može proširiti na parove MFCC-SVM i MFCC-DNN koji ostvaruju približno visoke performanse.



Sl. 5.5 – Koeficijenti uspešnosti parova obeležje-klasifikator u SD postavci



Sl. 5.6 – Koeficijenti uspešnosti parova obeležje-klasifikator u SI postavci

5.4 Rezime analize

Rezultati sprovedene analize se mogu sumirati u sledeće zaključke:

1. Predloženi koeficijenti uspešnosti predstavljaju robusne mere performanse obeležja i klasifikatora, koje se mogu uspešno primeniti u komparaciji kvaliteta različitih konfiguracija sistema za automatsko prepoznavanje govornih emocija. Povećanje robusnosti predloženih mera može da se postigne uključenjem dodatnih obeležja, klasifikatora i baza.

2. Na osnovu izračunatih koeficijenata uspešnosti je izdvojeno LFPC obeležje kao najpodobnije među testiranim spektralnim obeležjima. U slučaju klasifikatora nema jedinstvenog rešenja, već su prepoznati SVM i DNN kao podjednako dobri izbori. Izvedeni zaključci važe za SD i SI postavke.

3. U sprovedenim eksperimentima se uočava zavisnost rezultata od karakteristika korišćene baze, pri čemu je moguće pretpostaviti uticaj broja i vernosti zapisa, kao i broja i karakteristika govornika, dok je manje izražena mogućnost uticaja jezika i kulture.

4. Iako se HMM pokazao dobrim izborom u mnogim zadacima obrade govora, na osnovu sprovedene analize se može zaključiti da ima boljih izbora kada je u pitanju dizajn sistema za automatsko prepoznavanje govornih emocija upotrebom jednog klasifikatora.

5. Prethodna analiza ukazuje da se na osnovu rezultata postignutih u SD i SI postavkama mogu izvesti jako slični zaključci. S obzirom da SD postavka oponaša specifične primene namenjene malom broju korisnika, dok SI postavka simulira aplikacije široke upotrebe, koje su mnogo zastupljenije u praksi, potrebno je razmotriti sprovođenje budućih eksperimenata samo u SI postavci, kako bi se izbegla hiperprodukcija redundantnih rezultata.

Zaključci sprovedene analize mogu poslužiti kao osnova za buduća istraživanja u pravcu unapređenja sistema za klasifikaciju emocija. Drugi pravac budućih istraživanja može biti konstrukcija kompleksnih klasifikatorskih struktura koje uključuju širok skup obeležja i veliki broj osnovnih klasifikatora, a takve strukture bi svakako trebalo dopuniti diskriminatornim osobenostima svih obeležja i klasifikatora ispitivanih u prethodnim eksperimentima.

6 Optimizacija sistema za prepoznavanje govornih emocija

U ovoj glavi će akcenat biti na unapređenju sistema za automatsko prepoznavanje govornih emocija, oslanjajući se pri tome na rezultate komparativne analize iz prethodne glave. Jedan od zaključaka prethodne glave jeste da HMM ostvaruje lošije performanse od preostala dva ispitivana klasifikatora, ali, s obzirom da HMM karakteriše jedinstvena struktura, neće biti isključen iz dalje analize, već će u ovoj glavi biti sagledane mogućnosti njegovog unapređenja. Takođe, u prethodnoj glavi su razmatrana spektralna obeležja, u skladu sa uverenjem da su dobar izbor za zadatak automatskog prepoznavanja emocija iz govora, ali će u ovoj glavi ipak biti ispitane mogućnosti potpunije reprezentacije govornih zapisa upotrebom šireg skupa obeležja. Sa proširenjem skupa obeležja se javlja potreba za redukcijom dimenzija, gde će pored poznatih metoda biti analizirani i alternativni pristupi. Na kraju će biti razmotrena integracija prednosti različitih klasifikatorskih sistema u jedinstven multiklasifikatorski sistem.

6.1 Unapređenje HMM klasifikatora

Automatsko prepoznavanje govornih emocija je relativno mlada istraživačka tema sa mnogim elementima preuzetim iz intenzivno analiziranih sistema za automatsko prepoznavanje govora. HMM je nasleđen u ulozi klasifikatora i bio je često primenjivan u ranim istraživanjima [36]. U novim istraživanjima se HMM dosta ređe koristi, u skladu sa zaključcima da HMM generativna struktura nije najprikladnija za potrebe klasifikacije govornih emocija [107], pri čemu se u najnovijim istraživanjima dominantno koriste diskriminatorski algoritmi poput SVM [108] i DNN [109]. Međutim, može se pretpostaviti postojanje slučajeva u kojima se kvalitet modernih sistema za automatsko prepoznavanje govornih emocija može popraviti doprinosom svojstvene strukture HMM, ako ne kao individualnog klasifikatora, onda svakako u sastavu multiklasifikatorske strukture.

Prvi korak ka uspešnom uključenju HMM u moderne sisteme za prepoznavanje govornih emocija je razmatranje mogućnosti njegovog unapređenja. U proteklim godinama je bilo mnoštvo predloga za popravku kvaliteta osnovne strukture HMM [88], [110], [111], ili kreiranja hibridne strukture [112], [113], ali ni jedna nije opšteprihvaćena kao zamena za osnovnu implementaciju [37]. Mogući razlozi izostanka šireg uspeha predloženih izmena su: nedovoljna robusnost u odnosu na osnovnu implementaciju, prevelika složenost u odnosu na doprinos, preveliko udaljenje od osnovne ideje ili predložene izmene samo nisu dovoljno ubedljivo prezentovane da bi bile prihvaćene i nadalje korišćene od strane drugih istraživača. Vodeći računa o potencijalnim nedostacima ranijih predloga, u nastavku će biti razmotrene mogućnosti unapređenja osnovne implementacije diskretnog HMM, pri čemu se čini da najviše prostora za doradu ima kod koraka vektorske kvantizacije. Vektorska kvantizacija je vrlo izazovan zadatak, gde je potrebno izvršiti mapiranje vektora proizvoljne brojnosti

i dimenzija u konačan skup simbola. Po pravilu se rešenje zadatka vektorske kvantizacije oslanja na *K-means* klasterizaciju ili neku varijaciju, mada je bilo predloga i značajno drugačijih realizacija [114]. U nastavku će biti analizirane mogućnosti uključanja potpuno nove metode vektorske kvantizacije, zasnovane na upotrebi QQ krive (eng. *Quantile-Quantile plot*) [115], [116], [117]. U cilju dodatne popravke HMM, biće razmotreno uvođenje još nekoliko unapređenja, kao što su predobuka modela i višestruka obuka sa kombinovanjem modela.

6.1.1 Vektorska kvantizacija upotrebom QQ krive

Rešavanje zadatka vektorske kvantizacije se svodi na rešavanje problem klasterizacije datog skupa obučavajućih uzoraka u K klastera, gde je K željeni broj kvantizacionih simbola. Predstavnicima formiranih klastera se koriste u svojstvu etalona, a kvantizacija proizvoljnog vektora se sprovodi dodeljivanjem simbola najbližeg etalona. Na ovom mestu će biti razmotrene mogućnosti primene QQ krive za potrebe klasterizacije obučavajućih uzoraka, odnosno, određivanja etalona vektorske kvantizacije.

QQ kriva [117] je grafički metod za poređenje raspodela slučajnih promenljivih. Najčešće se koristi za određivanje stepena podudaranja raspodele datog skupa uzoraka sa nekom teorijskom raspodelom, a može se upotrebiti i za poređenje dva skupa uzoraka ili dve teorijske raspodele. QQ kriva se obrazuje uparivanjem kvantila slučajnih promenljivih, gde se pod kvantilom reda p , $p \in (0,1)$, podrazumeva vrednost koja je veća ili jednaka od tačno p realizacija slučajne promenljive. Kvantili datog skupa uzoraka se dobijaju sortiranjem, dok se kvantili teoretske raspodele dobijaju inverzijom funkcije raspodele verovatnoće. Za slučajne promenljive sa istim ili linearno zavisnim raspodelama verovatnoće se dobija približno linearna QQ kriva, a svako odstupanje od prave linije ukazuje na neslaganje raspodela, pri čemu se mogu izvoditi zaključci u skladu sa načinom devijacije.

U osnovi predloženog klasterizacionog algoritma je formiranje QQ krive za dati skup uzoraka i usvojenu teorijsku raspodelu, a zatim detektovanje lokalnih linearnih segmenata sa idejom da tačke sličnih karakteristika pripadaju istim linearnim segmentima. Istovetan rezon je primenjivan i u ranijim istraživanjima [118], gde su prepoznati linearni segmenti korišćeni kod određivanja parametara funkcije gustine verovatnoće, dok su u ovom radu tačke koje čine jedan linearni segment identifikovane kao predstavnici istog klastera.

QQ kriva se može direktno primeniti isključivo na jednodimenzionalne podatke, dok se u praksi uglavnom rešavaju višedimenzionalni problemi. Kako bi se omogućila primena QQ krive na višedimenzionalne podatke, neophodno je naći način za prevođenje vektora proizvoljnih dimenzija u skalarnu formu, pri čemu je gubitak dela informacija neminovan. Vektor se može prevesti u skalarnu predstavu jednostavnim sumiranjem njegovih elemenata, ali se na taj način vektori sa sličnim vrednostima različitih koordinata preslikavaju u približne skalarne predstave. Uvođenjem težinskih koeficijenata se dozvoljava manipulacija prioritetima koordinata i mogućnost boljeg očuvanja diskriminacionih informacija. Po uzoru na ranija istraživanja [119], za težinske koeficijente se mogu usvojiti elementi Fibonačijevog (eng. *Fibonacci*) niza brojeva. Fibonačijev niz je definisan linearnom rekursivnom formulom $F_d = F_{d-1} + F_{d-2}$, $d = 3, 4, \dots, \infty$, što će reći da je sledeći element niza određen kao suma prethodna dva elementa. Tako formirani niz ima osobinu da sa povećanjem indeksa, d , odnos susednih elemenata teži vrednosti zlatnog preseka ($\sim 1,618$). Iako su u opštem slučaju prva dva elementa Fibonačijevog niza $F_1 = 0$ i $F_2 = 1$, u ovom radu su usvojene vrednosti $F_1 = 5$ i $F_2 = 8$, čime se ostvaruje odnos svih susednih elemenata niza približan zlatnom preseku.

Za dati niz vektora V_1, V_2, \dots, V_N , sa dimenzijom pojedinačnih vektora D , formira se vektor Fibonačijevih brojeva $F = [F_1, F_2, \dots, F_D]^T$ i vektor indeksa $I = [i_1, i_2, \dots, i_D]^T$, pri čemu vektor indeksa predstavlja jednu permutaciju bez ponavljanja skupa $\{1, \dots, D\}$. Skalarne predstave zadatih vektora se određuju upotrebom elemenata vektora F kao težinskih koeficijenata, dok se vektorom I definiše prioritet koordinata:

$$S_n = \hat{V}_n(i_1) \cdot F_1 + \hat{V}_n(i_2) \cdot F_2 + \dots + \hat{V}_n(i_D) \cdot F_D, \quad n = 1, \dots, N, \quad (6.1)$$

pri čemu je $\hat{V}_{n^*}(i^*) \in [0,1]$, $n^* = 1, \dots, N$, $i^* = 1, \dots, D$, normalizovana vrednost koordinate i^* vektora V_{n^*} u odnosu na i^* koordinate svih vektora V_n , $n = 1, \dots, N$.

QQ kriva se formira za skalarne predstave S_1, S_2, \dots, S_N u odnosu na teorijsku standardnu normalnu raspodelu sa funkcijom raspodele verovatnoće $P(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$. Određivanjem inverzije funkcije raspodele verovatnoće su definisani kvantili jedne ose, $P^{-1}\left(\frac{n-0.5}{N}\right)$, $n = 1, \dots, N$, dok se sortiranjem skalarnih predstava određuju kvantili druge ose, y_1, y_2, \dots, y_N . Uparivanjem kvantila je završena konstrukcija QQ krive.

Očekuje se odstupanje formirane QQ krive od prave linije, ali se mogu prepoznati lokalni linearni segmenti koji ukazuju na grupe tačaka sličnih karakteristika. Izdvajanje linearnih segmenata se vrši određivanjem deo po deo linearne aproksimacije QQ krive, a tačke koje pripadaju istom linearnom segmentu se proglašavaju predstavnicima istog klastera. Konstrukcija deo po deo linearne aproksimacije započinje formiranjem grupa od po nekoliko susednih tačaka, tako da svaka tačka pripadne tačno jednoj grupi. Za svaku grupu tačaka se određuje linearna aproksimacija i greška aproksimacije, a potom se vrši iterativno spajanje susednih grupa održavajući ukupnu grešku aproksimacije što manjom. Postupak se ponavlja do dostizanja zadatog broja linearnih segmenata, a zatim se vrši provera postojanja grupa koje uključuju premalo tačaka u kom slučaju se vrši spajanje sa susednim. Tako formirane grupe se koriste za određivanje kvantizacionih etalona, a različiti pristupi su opisani u nastavku.

6.1.1.1 *Određivanje etalona – pristup 1*

Na osnovu deo po deo linearne aproksimacije QQ krive sa K linearnih segmenata se može formirati K klastera grupisanjem vektora čije skalarne predstave pripadaju istim linearnim segmentima, a centriodi formiranih klastera se mogu koristiti kao etaloni vektorske kvantizacije. Međutim, prilikom konstrukcije QQ krive se vektori proizvoljnih dimenzija predstavljaju skalarima (jednačina (6.1)), pri čemu se očuvanje različitih komponenti vektora ostvaruje u skladu sa vrednostima elemenata Fibonačijevog niza kojim se ponderišu. Različitim izborom vektora indeksa, I , se dolazi do drugačijih skalarnih predstava, što za posledicu ima izmenu rezultata klasterizacije i kvantizacionih etalona. Ponavljanjem procedure za različite vektore indeksa i odabirom klasterizacije koja ostvaruje najmanje ukupno rasipanje se može doći do kvalitetnijeg skupa etalona. Pri tome se rasipanje jednog klastera definiše kao kumulativno rastojanje centroida od svih tačaka klastera, a ukupno rasipanje se određuje sumiranjem rasipanja svih klastera.

6.1.1.2 *Određivanje etalona – pristup 2*

Uspešnost prethodno opisanog, neposrednog pristupa ograničava klasterizacija na osnovu nepotpunih informacija sadržanih u skalarnim predstavama vektora, pri čemu se problem ne prevazilazi ni sa ponavljanjem procedure za različito formirane skalarne predstave, s obzirom na odsustvo mehanizma za kombinovanje rezultata pojedinačnih klasterizacija. Razradom prethodnog pristupa u smeru kombinovanja rezultata pojedinačnih iteracija se dolazi do sledećeg pristupa za

formiranje etalona vektorske kvantizacije. Slično kao i ranije, za različite varijacije vektora indeksa, I , se određuju odgovarajuće skalarne predstave vektora i formira QQ kriva, a zatim se konstrukcija deo po deo linearne aproksimacije vrši za broj linearnih segmenata jednak celobrojnoj vrednosti najbližoj količniku K/C , gde je K zadati broj klastera i $C = 10$ usvojena konstanta. Sumiranje rezultata pojedinačnih iteracija se vrši pomoću kvadratne matrice T , dimenzija jednakih broju vektora u obučavajućem skupu, N , pri čemu element $T(n_1, n_2)$, $n_1 = 1, \dots, N$, $n_2 = 1, \dots, N$, predstavlja broj iteracija u kojima su skalarne predstave vektora V_{n_1} i V_{n_2} pripale istim linearnim segmentima deo po deo linearne aproksimacije. Maksimalna vrednost, $\max T$, koju mogu ostvariti elementi matrice T je jednaka broju sprovedenih iteracija, tako da se može usvojiti kriterijum $T(n_1, n_2) \geq 0.8 \cdot \max T$ kao uslov za pripadnost vektora V_{n_1} i V_{n_2} istoj grupi. Za grupe formirane prema prethodnom kriterijumu se vrši validacija u odnosu na broj sadržanih vektora, a zatim se validne grupe združuju do zadatog broja klastera, K , dok se grupe sa nedovoljnim brojem vektora pridružuju najbližim validnim grupama. Združivanje grupa se vrši po kriterijumu najmanjeg Euklidskog rastojanja centara grupa, pri čemu se centar grupe određuje kao srednja vrednost vektora koji čine grupu. Na ovaj način je izvršena klasterizacija zadatih vektora, V_1, V_2, \dots, V_N , u K klastera, a centriodi formiranih klastera predstavljaju etalone za vektorsku kvantizaciju. Mana opisanog pristupa jeste što za velike obučavajuće skupove matrica T postaje suviše velika, pa je neophodno ograničavanje broja korišćenih vektora.

6.1.1.3 *Određivanje etalona – pristup 3*

Sledeći pristup istovremeno dozvoljava upotrebu svih dostupnih uzoraka i kombinovanje rezultata dobijenih za različito formirane skalarne predstave. Kod ovog pristupa se vrši jednovremeno razdvajanje vektora u tačno dva klastera, pa je u skladu sa tim pojednostavljeno formiranje deo po deo linearne aproksimacije QQ krive. Prevođenje vektora u skalarne predstave i konstrukcija QQ krive se sprovode kao u ranijim pristupima, a formiranje deo po deo linearne aproksimacije se svodi na određivanje granice između linearnih segmenata koja obezbeđuje minimalno kvadratno rastojanje tačaka QQ krive od deo po deo linearne aproksimacije. Proverom za 10 ekvidistantnih tačaka QQ krive se pronalazi okolina tražene granice, a nadalje se metodom polovljenja dolazi do konačnog rešenja. Određivanjem granice linearnih segmenata je formirana deo po deo linearna aproksimacija QQ krive, a time je izvršeno razdvajanje datih vektora u klastere.

Za potrebe ovog pristupa se svakom klasteru pridružuju dva parametra: centroid i rasipanje. Centroid je reprezent klastera i računa se kao srednja vrednost svih vektora klastera, a rasipanje se određuje kao sumarno odstupanje pridruženih vektora od centroida klastera. Formiranje klastera se vrši prema iterativnoj proceduri, pri čemu se u svakoj iteraciji klaster sa najvećim rasipanjem deli na dva. Razdvajanje klastera se vrši konstrukcijom QQ krive i njene deo po deo linearne aproksimacije kao u prethodno opisanom postupku. S obzirom da se QQ kriva formira na osnovu skalarnih predstava koje sadrže samo deo informacija originalnih vektora, postupak razdvajanja se ponavlja za skalarne predstave formirane sa različitim prioritetima koordinata originalnih vektora, a kao konačan par novoformiranih klastera se zadržava onaj koji ostvaruje najmanje rasipanje. Iterativno razdvajanje se vrši do dostizanja zadatog broja klastera, a centriodi formiranih klastera se koriste kao etaloni za vektorsku kvantizaciju.

6.1.1.4 *Analiza različitih pristupa za određivanje etalona*

Predložene pristupe za određivanje kvantizacionih etalona karakterišu različite svojstvenosti, pri čemu prikladnost izbora zavisi od postavljenih zahteva. S obzirom da se u ovom radu vektorska kvantizacija koristi kao sastavni deo sistema za automatsko prepoznavanje govornih emocija, različiti pristupi za formiranje kvantizacionih etalona su analizirani na adekvatan način. Testiranja su

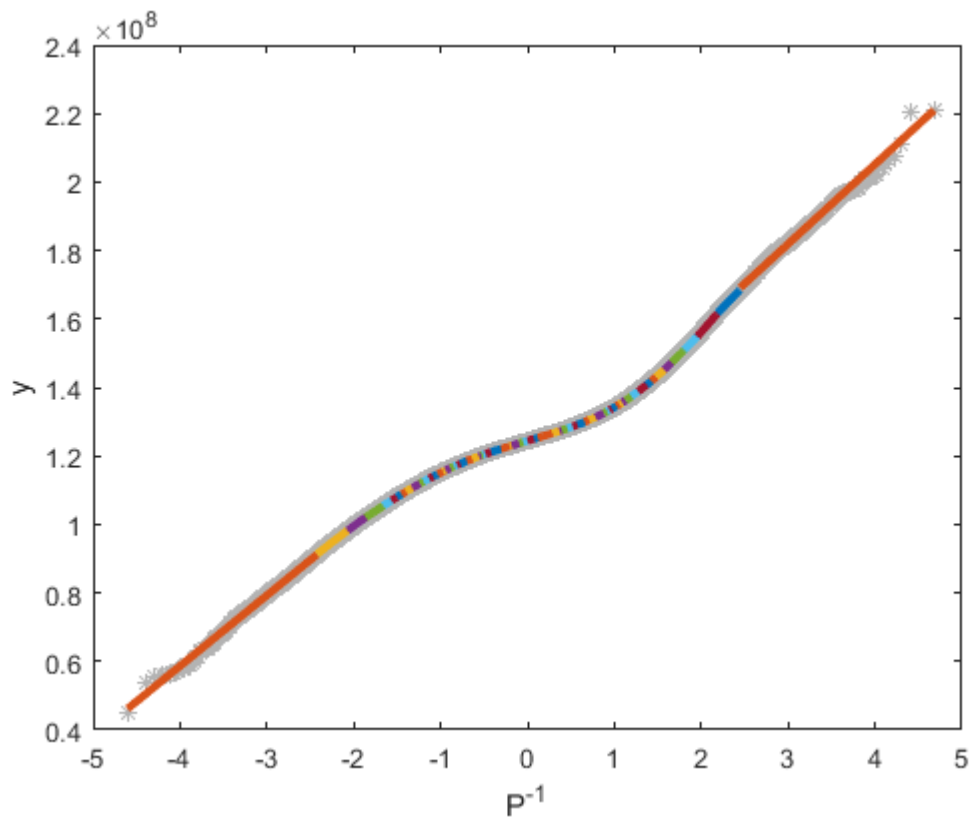
organizovana u skladu sa SI postavkom, pri čemu su korišćeni zapisi iz baza GEES, PES, RUSLANA i Berlin. Svaki zapis je predstavljen normalizovanim nizom vektora kratkovremenih obeležja sačinjenih od LFPC obeležja, delta i delta-delta koeficijenata. Etaloni su formirani na osnovu vektora kratkovremenih obeležja izdvojenih iz obučavajućeg skupa uzoraka, a za proveru valjanosti su korišćeni vektori kratkovremenih obeležja izdvojeni iz testirajućeg skupa uzoraka, pri čemu je valjanost etalona procenjivana u odnosu na grešku koja se unosi prilikom vektorske kvantizacije. U skladu sa organizacijom testiranja u SI postavci, formiranje i provera etalona su vršeni za različito realizovane skupove obučavajućih i testirajućih uzoraka, a krajnji rezultat testiranja je iskazivan kao prosečno rasipanje kvantizovanih vektora oko etalona. Analizom su obuhvaćeni svi prethodno predloženi pristupi za određivanje kvantizacionih etalona konstrukcijom QQ krive, kao i određivanje etalona upotrebom *K-means* klasterizacije. Rezultati testiranja su sumirani u Tab. 6.1, pri čemu su u testiranjima korišćene realizacije pristupa 1 sa jednom iteracijom (u tabeli označeno kao pristup 1a) i 30 iteracija (u tabeli označeno kao pristup 1b), pristupa 2 sa 1000 iteracija i ograničenjem od 5000 uzoraka, a pristupa 3 sa 30 iteracija. U svim testovima je vršeno formiranje 64 kvantizaciona etalona.

Tab. 6.1 – Prosečna rasipanja za različito formirane kvantizacione etalone

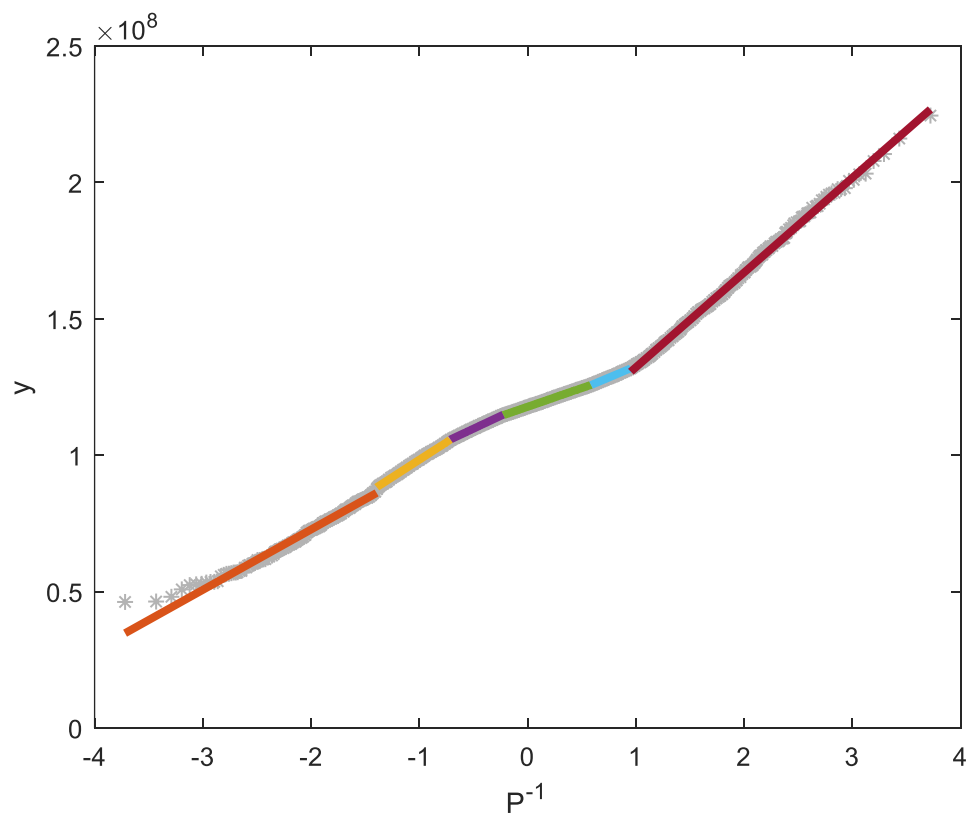
| Kvantizacija | Baza | | | |
|----------------|---------|---------|---------|---------|
| | GEES | PES | RUSL | Berlin |
| <i>K-means</i> | 70,6152 | 68,7573 | 64,3424 | 60,9774 |
| Pristup 1a | 71,3363 | 73,6103 | 67,7143 | 65,5803 |
| Pristup 1b | 71,1169 | 72,6068 | 66,8685 | 64,9859 |
| Pristup 2 | 67,3437 | 69,6927 | 63,0309 | 62,1791 |
| Pristup 3 | 66,6721 | 68,7954 | 61,9419 | 60,8027 |

Na osnovu prosečnih rasipanja datih u tabeli Tab. 6.1 se zaključuje da jednostavnost pristupa 1 za posledicu ima ostvarenje najveće greške kvantizacije, pri čemu ni uvođenje dodatnih iteracija ne obezbeđuje značajnu popravku rezultata. Pristup 2 ostvaruje bolje rezultate od pristupa 1, a za neke baze bolje i od *K-means* klasterizacije. Pristup 3 se izdvaja kao univerzalno najpodobnije rešenje, premda se uočava izraženiji doprinos za baze sa većim brojem zapisa.

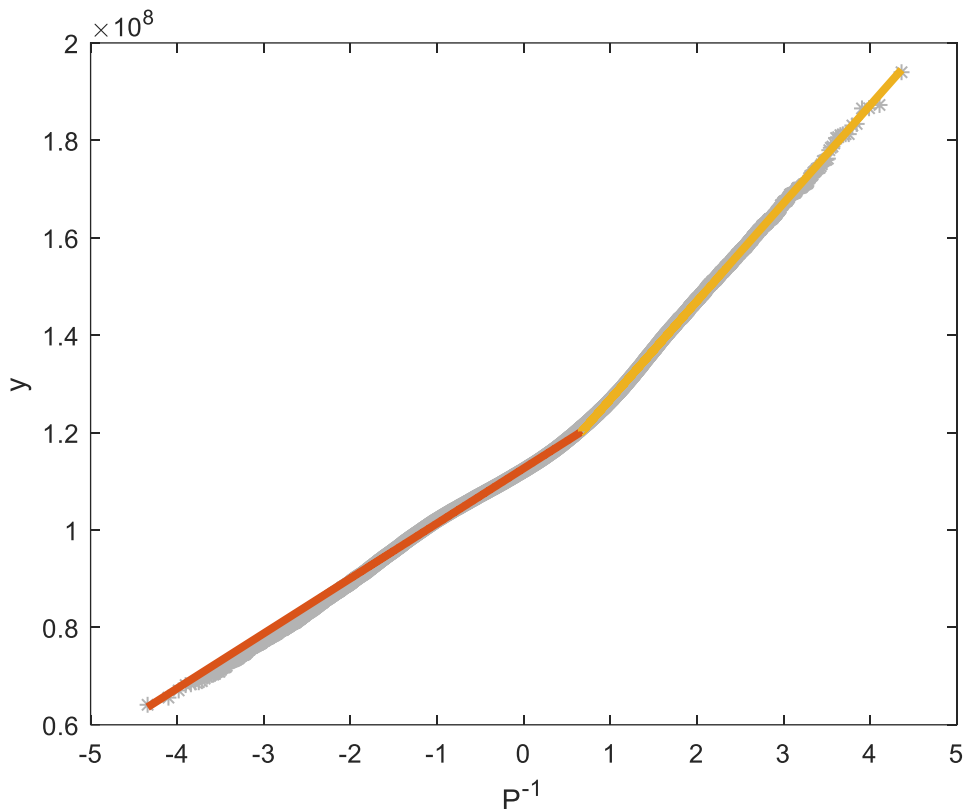
Predloženi pristupi za formiranje kvantizacionih etalona koriste iste principe prilikom prevođenja vektora u skalarne predstave i konstrukcije QQ krive, a razlike su prisutne kod određivanja deo po deo linearne aproksimacije QQ krive. Na Sl. 6.1 je ilustrovano formiranje QQ krive i njene deo po deo linearne aproksimacije u skladu sa pristupom 1, pri čemu je uočljiva veća dužina linearnih segmenata koji aproksimiraju krajeve QQ krive u odnosu na one koji odgovaraju središnjem delu, što se može objasniti velikom koncentracijom tačaka u središnjem delu QQ krive. Određivanje deo po deo linearne aproksimacije QQ krive za potrebe pristupa 2 je ilustrovano na Sl. 6.2, a na Sl. 6.3 je prikazan izgled deo po deo linearne aproksimacije QQ krive koji odgovara pristupu 3.



Sl. 6.1 – Deo po deo linearna aproksimacija QQ krive (pristup 1)



Sl. 6.2 – Deo po deo linearna aproksimacija QQ krive (pristup 2)



Sl. 6.3 – Deo po deo linearna aproksimacija QQ krive (pristup 3)

6.1.2 Dodatna unapređenja HMM

Prostor za unapređenje HMM se može tražiti i van koraka vektorske kvantizacije. Jedno moguće unapređenje jeste uvođenje predobuke, odnosno, formiranje inicijalnog modela na osnovu svih uzoraka iz obučavajućeg skupa, a parametri tako formiranog modela se mogu upotrebiti za inicijalizaciju modela pojedinačnih emocija. Baum-Velč iterativni algoritam konvergira ka lokalno optimalnom modelu koji zavisi od početnog izbora parametara, pri čemu postavljanje slučajnih vrednosti ne dozvoljava nikakvu kontrolu nad pravcem konvergencije. Uvođenjem predobuke se postiže da inicijalni parametri modela u određenoj meri odgovaraju svakom od emocionalnih stanja, a dodatna konvergencija takvih parametara, u skladu sa svojstvenostima uzoraka za konkretnu emociju, može da poveća šanse za postizanje globalno optimalnog modela. Slučajnost i dalje figuriše kroz početni odabir parametara inicijalnog modela, ali, s obzirom da se inicijalni model obučava sa mnogo više uzoraka i da se za modele pojedinačnih emocija sprovodi dodatna obuka, može se očekivati manji uticaj slučajnosti na krajnji rezultat.

Sprovođenjem višestruke obuke modela se postiže dodatno smanjenje slučajnosti u krajnjem rezultatu. Ponavljanjem obuke za različito postavljene početne parametre inicijalnog modela se može formirati proizvoljno veliki skup modela pojedinačnih emocija, a izborom kombinacije modela koja najbolje reprezentuje obučavajuće uzorke se na još jedan način obezbeđuje postizanje kvalitetnijih rezultata.

Formiranjem raznovrsnijeg skupa modela se omogućava odabir kombinacije koja adekvatnije reprezentuje obučavajuće uzorke, pri čemu je primena korektivne obuke jedan od načina za generisanje dodatnih modela. Korektivna obuka podrazumeva korekciju parametara HMM (matrice verovatnoće opservacija) sa težnjom smanjenja greške klasifikacije obučavajućeg skupa uzoraka, kao što je opisano u 4.1.3.6. Korekcija parametara se vrši u odnosu na pojedinačne uzorke, što može da naruši generalizaciona svojstva modela, zbog čega se modeli dobijeni korektivnom obukom podvrgavaju još jednoj iteraciji Baum-Velč algoritma sa svim uzorcima za obuku datog modela. Tako dobijeni modeli učestvuju u izboru najbolje kombinacije, zajedno sa modelima formiranim bez primene korektivne obuke, što isključuje mogućnost za potencijalnu degradaciju kvaliteta usled uvođenja korektivne obuke.

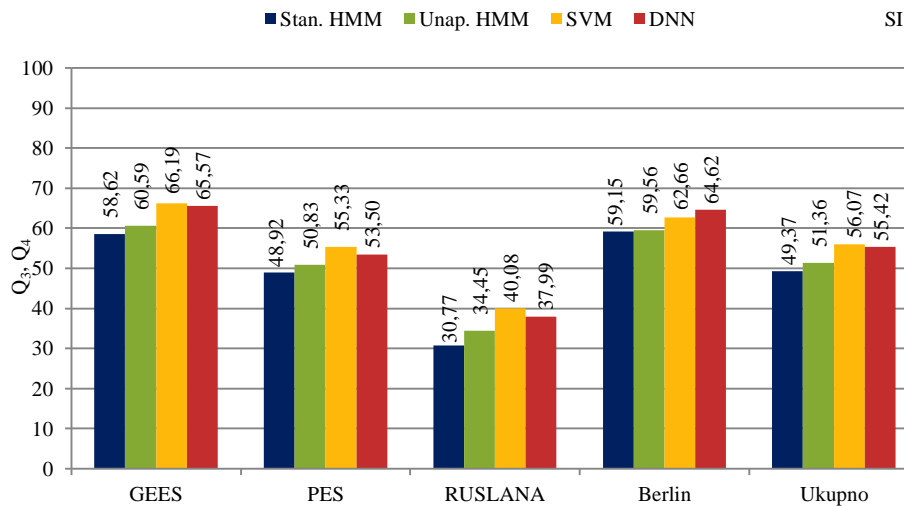
6.1.3 Analiza unapređenog HMM

Eksperimenti su organizovani u cilju verifikacije predloženih unapređenja HMM, pri čemu je testirana konfiguracija koja za određivanje kvantizacionih etalona koristi prethodno opisani pristup 3 sa 30 iteracija u pojedinačnim razdvajanjima, a kao dodatno unapređenje su uključene tri iteracije višestruke obuke sa predobukom i korektivnom obukom. Po ugledu na ranije eksperimente, u analizi su korišćena spektralna obeležja (LPCC, LFPC i MFCC) izdvojena iz tri slovenske i jedne germanske baze (GEES, PES, RUSLANA i Berlin), a u skladu sa zaključcima glave 5 su sprovedena testiranja samo u SI postavci. Takođe, za sve eksperimente koji slede će biti zajedničko izvođenje testiranja samo u SI postavci sa ponavljanjem testova za svaku od baza: GEES, PES, RUSLANA i Berlin.

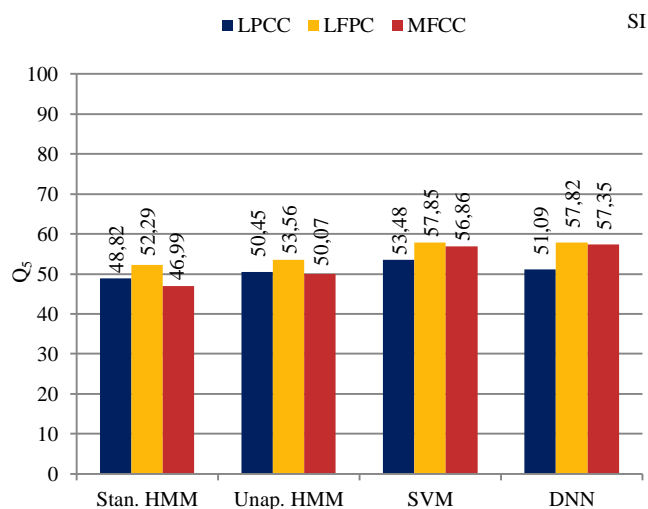
Tab. 6.2 – Uspešnost klasifikacije – unapređeni HMM

| Klasifikator | Obeležje | Baza | | | |
|----------------|----------|-------|-------|-------|--------|
| | | GEES | PES | RUSL | Berlin |
| Unapređeni HMM | LPCC | 59,22 | 53,50 | 33,67 | 55,39 |
| | LFPC | 66,22 | 51,50 | 33,28 | 63,24 |
| | MFCC | 56,33 | 47,50 | 36,39 | 60,05 |

Rezultati testiranja unapređene procedure za formiranje HMM su sumirani u Tab. 6.2, a poređenjem sa rezultatima postignutim primenom standardne procedure, Tab. 5.2, se potvrđuje da predložene izmene popravljaju kvalitet HMM bez obzira na izbor obeležja i baze. Ukoliko se poređenje vrši sa rezultatima SVM i DNN klasifikatora (Tab. 5.2), prepoznaju se pojedinačne kombinacije obeležja i baza za koje unapređeni HMM uspeva da ostvari bolje rezultate, a opšti zaključci se mogu izvesti određivanjem koeficijentima uspešnosti Q_3 , Q_4 i Q_5 za unapređeni HMM i poređenjem sa koeficijentima uspešnosti datim na Sl. 5.4 i Sl. 5.6. Na osnovu Sl. 6.4 i Sl. 6.5 je jasno da predložena unapređenja uspevaju da nadmaše rezultate standardne HMM implementacije i da se približe performansi preostala dva klasifikatora.



Sl. 6.4 – Koeficijenti uspešnosti klasifikatora (proširenje: unapređeni HMM)



Sl. 6.5 – Koeficijenti uspešnosti parova obeležje-klasifikator (proširenje: unapređeni HMM)

Predložene izmene usložnjavaju obuku HMM, pri čemu ponavljanje koraka sa slučajnim karakterom obezbeđuje verodostojnije modele, tako da je izborom broja ponavljanja moguće kontrolisati balans između računске zahtevnosti obuke i kvaliteta modela. Povećanje računске zahtevnosti obuke ne ograničava praktičnu primenljivost, s obzirom da se obuka sprovodi samo pri formiranju modela, dok je za praktičnu primenu mnogo značajniji postupak testiranja novih uzoraka, koji nije obuhvaćen izmenama. Nepromenjenost strukture modela i postupka testiranja omogućava jednostavno uključjenje unapređenog HMM u postojeće sisteme koji su dizajnirani u skladu sa standardnom HMM implementacijom.

Predloženu proceduru za formiranje HMM je moguće dodatno razrađivati, pri čemu se prostor za unapređenje prepoznaje kod preslikavanja vektora u skalarne predstave. Određivanjem relevantnosti pojedinačnih elemenata vektora obeležja bi se omogućilo formiranje vernijih skalarne predstave, što bi garantovalo adekvatnije grupisanje vektora i smanjenje, ili izostanak, potrebe za ponavljanje postupka sa različito formiranim skalarne predstavama.

U eksperimentima koji slede će biti korišćena isključivo konfiguracija unapređenog HMM iz prethodne analize, bez posebnog isticanja koja implementacija i u kojoj konfiguraciji je upotrebljena.

6.2 Proširenje skupa obeležja i redukcija dimenzija

U skladu sa opštim stavom da su spektralna obeležja dobar izbor za zadatak automatskog prepoznavanja govornih emocija [9], [93], [97], u prethodno sprovedenim eksperimentima su izdvajana LPCC, LFPC i MFCC obeležja. Međutim, može se očekivati prisustvo korisnih informacija i u drugim grupama obeležja, što će biti verifikovano sprovođenjem eksperimenata sa skupom koji izuzima spektralna obeležja. Takođe, biće razmotrena popravka informativnosti sa formiranjem proširenog skupa obeležja, sačinjenog od mnoštva predstavnika različitih grupa obeležja. Proširenje skupa obeležja nameće potrebu za uključenjem neke od tehnika za redukciju dimenzija, gde će biti angažovan LDA kao uobičajen pristup, a biće ispitane i mogućnosti redukcije dimenzija upotrebom Fibonačijevog niza brojeva.

6.2.1 Nespektralna obeležja

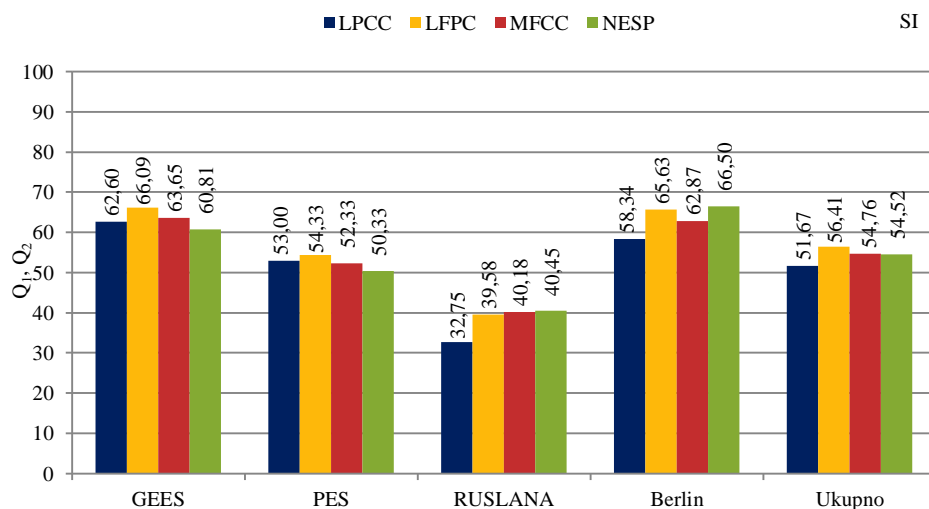
Analiza prisutnosti korisnih informacija izvan spektralnih obeležja je sprovedena sa skupom koji uključuje predstavnike kontinualnih obeležja, obeležja kvaliteta glasa i nelinearnih obeležja. Iz grupe kontinualnih obeležja su uvršteni: fundamentalna frekvencija, prva tri formanta i logaritamske vrednosti energije po frekvencijskim opsezima. Od obeležja kvaliteta glasa je uključen HNR, a kao predstavnik nelinearnih obeležja je određivan logaritam Tajger energije. Odabrana obeležja su izdvajana prema procedurama datim u glavi 3 i na taj način je svaki frejm govornog signala okarakterisan vektorom od 12 elemenata, a pridruživanjem delta i delta-delta koeficijenata su formirani konačni vektori kratkovremenih obeležja. Vektori globalnih obeležja su određivani primenom osnovnih statistika (srednja vrednost, standardna devijacija, minimum, maksimum i opseg promene), definisanih u poglavlju 3.6. Testiranje predloženih nespektralnih obeležja je vršeno u SI postavci, sa bazama i klasifikatorima kao u prethodnim eksperimentima, a rezultati testiranja su dati u Tab. 6.3.

Tab. 6.3 – Uspešnost klasifikacije – nespektralna obeležja

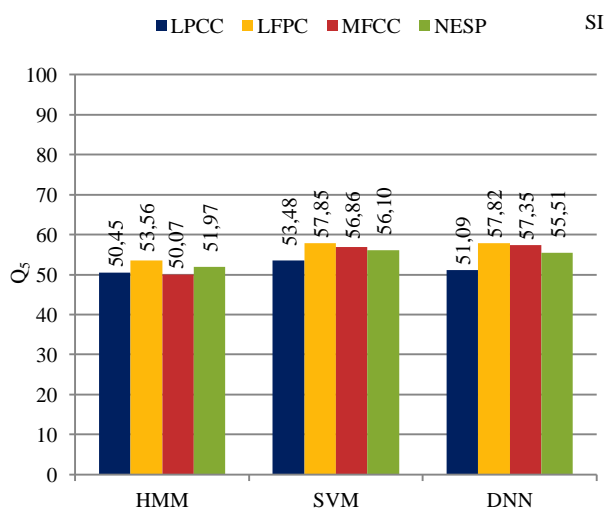
| Klasifikator | Baza | | | |
|--------------|-------|-------|-------|--------|
| | GEES | PES | RUSL | Berlin |
| HMM | 59,89 | 51,50 | 33,74 | 62,75 |
| SVM | 60,44 | 49,50 | 45,08 | 69,36 |
| DNN | 62,11 | 50,00 | 42,52 | 67,40 |

Poređenjem Tab. 6.3 sa Tab. 5.2 i Tab. 6.2 se može konstatovati bliskost rezultata ostvarenih upotrebom spektralnih i nespektralnih obeležja, a određivanjem koeficijenata uspešnosti je omogućeno izvođenje konkretnijih zaključaka. Proširenjem Sl. 5.2 koeficijentima uspešnosti nespektralnih obeležja je formirana Sl. 6.6, pri čemu je izvršena i korekcija koeficijenata uspešnosti spektralnih obeležja u skladu sa rezultatima unapređene HMM implementacije. Određivanjem koeficijenata uspešnosti parova obeležje-klasifikator za nespektralna obeležja i kombinovanjem sa koeficijentima uspešnosti sa Sl. 6.5 je formirana Sl. 6.7. Na osnovu Sl. 6.6 i Sl. 6.7 se zaključuje da nespektralna obeležja obezbeđuju performanse u opsegu ranije testiranih obeležja, pri čemu izračunati koeficijenti uspešnosti ukazuju na nespektralna obeležja kao bolji izbor za baze

RUSLANA i Berlin. Sprovedeni eksperimenti potvrđuju prisustvo vrednih informacija izvan spektralnih obeležja, što opravdava analizu šireg skupa obeležja.



Sl. 6.6 – Koeficijenti uspešnosti obeležja (proširenje: nespektralna obeležja)



Sl. 6.7 – Koeficijenti uspešnosti parova obeležje-klasifikator (proširenje: nespektralna obeležja)

6.2.2 Prošireni skup obeležja

Prethodni eksperimenti su potvrdili da se značajni pokazatelji emocionalnih stanja nalaze u različitim grupama obeležja, pri čemu se može pretpostaviti da različite grupe obeležja sadrže informacije o drugačijim manifestacijama emocionalnih stanja. Umesto determinacije najrelevantnije grupe obeležja, formiran je združeni skup, sa idejom obezbeđenja potpunijih informacija međusobnim dopunjavanjem raznorodnih obeležja. U prošireni skup su uključena sledeća obeležja:

- Spektralna: LPCC, LFPC, MFCC;

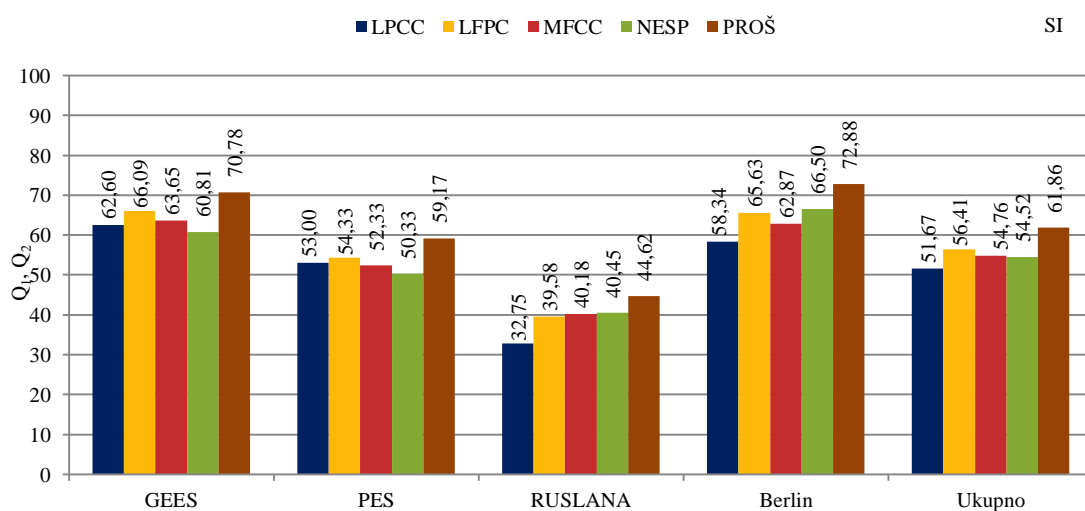
- Kontinualna: fundamentalna frekvencija, prva tri formanta, energija, logaritam energije, energija po frekvencijskim opsezima, logaritam energije po frekvencijskim opsezima;
- Kvaliteta glasa: HNR, ZCR, verovatnoća zvučnog govora;
- Nelinearna: Tajger energija, logaritam Tajger energije, Tajger energija po frekvencijskim opsezima.

Vektori kratkovremenih obeležja su formirani sprovođenjem procedura iz glave 3 i pridruživanjem delta i delta-delta koeficijenata. Vektore globalnih obeležja je sačinjavao prošireni skup statistika kratkovremenih obeležja, dat u Tab. 3.1, zajedno sa obeležjima trajanja, definisanim u odeljku 3.2.4, a džiter i šimer su uključeni kroz statistike fundamentalne frekvencije i energije. Testiranje proširenog skupa obeležja je sprovedeno sa bazama i klasifikatorima kao u prethodnim eksperimentima, a rezultati su sumirani u Tab. 6.4.

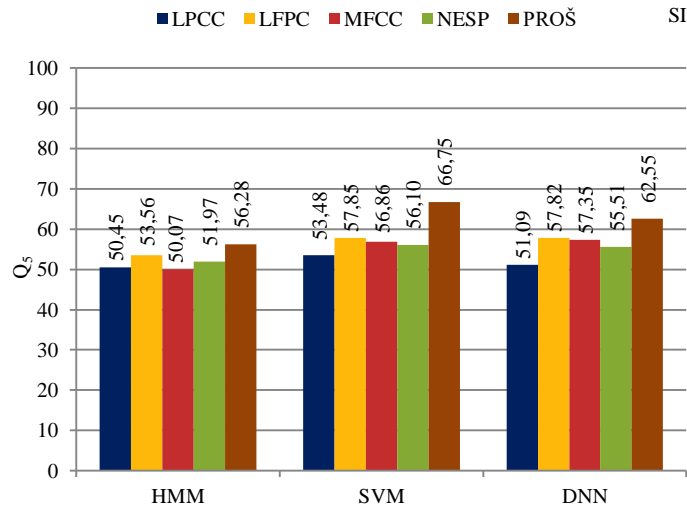
Tab. 6.4 – Uspešnost klasifikacije – prošireni skup obeležja

| Klasifikator | Baza | | | |
|--------------|-------|-------|-------|--------|
| | GEES | PES | RUSL | Berlin |
| HMM | 67,44 | 56,00 | 36,49 | 65,20 |
| SVM | 74,89 | 63,50 | 52,13 | 76,47 |
| DNN | 70,00 | 58,00 | 45,25 | 76,96 |

Poređenjem rezultata iz Tab. 6.4 sa rezultatima ranijih eksperimenata, datim u Tab. 5.2, Tab. 6.2 i Tab. 6.3, se potvrđuje ostvarenje boljih rezultata sa proširenim skupom obeležja nego sa individualnom primenom spektralnih i nespektralnih obeležja. Određivanjem koeficijenata uspešnosti obeležja i parova obeležje-klasifikator na osnovu rezultata iz Tab. 6.4 i prikazivanjem uporedo sa koeficijentima uspešnosti iz prethodnih eksperimenata, Sl. 6.8 i Sl. 6.9, se zaključuje da upotreba proširenog skupa obeležja u osetnoj meri popravlja performanse sistema za automatsko prepoznavanje govornih emocija.



Sl. 6.8 – Koeficijenti uspešnosti obeležja (proširenje: prošireni skup obeležja)



Sl. 6.9 – Koeficijenti uspešnosti parova obeležje-klasifikator (proširenje: prošireni skup obeležja)

Proizvoljno povećanje dimenzija vektora obeležja je u različitoj meri prihvatljivo za različite klasifikatorske strukture. HMM klasifikator je pod najmanjim uticajem povećanja dimenzija, s obzirom da se pri formiranju kvantizacionih etalona koriste skalarne predstave vektora, a sam model je zasnovan na simboličkom kodovniku. Povećanje dimenzija prostora obeležja doprinosi značajnom računskom opterećenju SVM klasifikatora, pri čemu se formiranje modela vrši u dodatno proširenom prostoru. DNN klasifikator može da modeluje probleme proizvoljne složenosti ukoliko se odabere adekvatna struktura modela i obezbedi dovoljno veliki broj primera. Broj neurona ulaznog sloja DNN modela mora da prati povećanje dimenzija ulaznih vektora, a srazmerno bi mogao da se poveća broj neurona i u skrivenim slojevima, ali ograničen broj uzoraka ne dozvoljava neograničeno uslozljavanje modela.

Proširenje skupa obeležja omogućava dopunjavanje informacionog sadržaja, pri čemu je za postavljanje problema u prostoru većih dimenzija potreban i veći broj uzoraka. Bez promene broja uzoraka se proređenje prostora obeležja može izbeći smanjenjem dimenzija vektora obeležja, uz čuvanje bitnih informacija.

6.2.3 Redukcija dimenzija

Izdvajanje velikog broja obeležja govornog signala uslovljava nadogradnju sistema za automatsko prepoznavanje govornih emocija korakom za redukciju dimenzija vektora obeležja. U ovom radu je primenjivana LDA redukcija dimenzija, a analizirane su i mogućnosti redukcije dimenzija upotrebom Fibonačijevog niza brojeva. U poglavlju 3.7 je opisana LDA metoda, a u nastavku će biti izneta ideja predložene metode i rezultati testiranja.

6.2.3.1 Fibonači redukcija dimenzija

Redukcija dimenzija upotrebom Fibonačijevog niza brojeva ima zajedničke elemente sa ranije predstavljenim pristupom za određivanje kvantizacionih etalona kod diskretnog HMM (odjeljak 6.1.1). Za dati niz vektora obeležja V_1, V_2, \dots, V_N , dimenzije pojedinačnih vektora D , je potrebno odrediti niz vektora redukovanih dimenzija Y_1, Y_2, \dots, Y_N , dimenzije pojedinačnih vektora K . Formira

se vektor Fibonačijevih brojeva $F = [F_1, F_2, \dots, F_D]^T$, $F_1 = 5$, $F_2 = 8$, $F_d = F_{d-1} + F_{d-2}$, $d = 3, \dots, D$, a vektori indeksa $I^{(k)} = [i_1^{(k)}, i_2^{(k)}, \dots, i_D^{(k)}]^T$, $k = 1, \dots, K$, se obrazuju kao permutacije bez ponavljanja skupa $\{1, \dots, D\}$. Normalizacijom zadatih vektora se dolazi do niza vektora $\hat{V}_1, \hat{V}_2, \dots, \hat{V}_N$, gde je $\hat{V}_n(i) \in [0,1]$, $n = 1, \dots, N$, $i = 1, \dots, D$, a zatim se postupak prevođenja vektora u skalarnu formu ponavlja K puta prema formuli:

$$S_n^{(k)} = \hat{V}_n(i_1^{(k)}) \cdot F_1 + \hat{V}_n(i_2^{(k)}) \cdot F_2 + \dots + \hat{V}_n(i_D^{(k)}) \cdot F_D, \quad n = 1, \dots, N, \quad k = 1, \dots, K. \quad (6.2)$$

Na osnovu skalarnih predstava se formiraju vektori redukovanih dimenzija $Y_n = [S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(K)}]^T$, $n = 1, \dots, N$. Predloženi pristup za redukciju dimenzija karakteriše jednostavnost implementacije, a testiranjem se može procenti uticaj na performanse klasifikatorskog sistema.

6.2.3.2 Analiza metoda za redukciju dimenzija

Uticaj metoda za redukciju dimenzija na performanse sistema za automatsko prepoznavanje govornih emocija je ispitan testiranjem sistema koji za opisivanje govorne sekvence koriste prošireni skup obeležja iz prethodne analize, a pre koraka klasifikacije sprovode redukciju dimenzija. Vektori kratkovremenih obeležja su redukovani na 36 elemenata, a vektori globalnih obeležja su redukovani na 180 elemenata, što odgovara dimenzijama vektora iz eksperimenata u kojima su analizirana individualno spektralna i nespektralna obeležja. Rezultati testiranja su sumirani u Tab. 6.5.

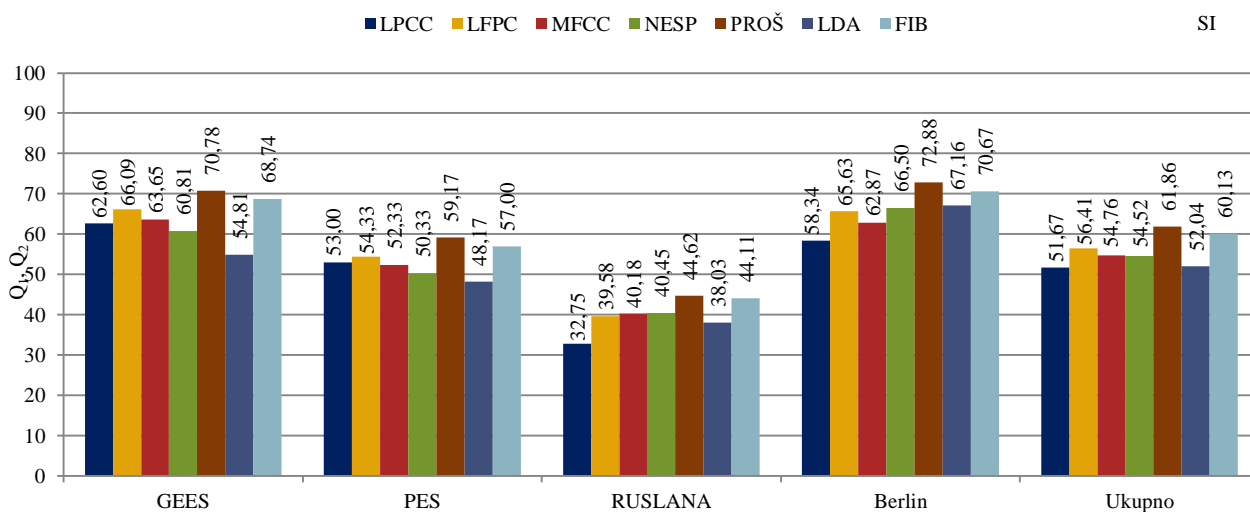
Tab. 6.5 – Uspešnost klasifikacije – redukcija dimenzija

| Klasifikator | Redukcija dimenzija | Baza | | | |
|--------------|---------------------|-------|-------|-------|--------|
| | | GEES | PES | RUSL | Berlin |
| HMM | LDA | 58,33 | 54,00 | 42,20 | 70,59 |
| | Fibonači | 63,78 | 53,50 | 41,08 | 66,42 |
| SVM | LDA | 54,11 | 48,00 | 35,31 | 63,73 |
| | Fibonači | 72,56 | 60,00 | 48,10 | 74,02 |
| DNN | LDA | 52,00 | 42,50 | 36,59 | 67,16 |
| | Fibonači | 69,89 | 57,50 | 43,15 | 71,57 |

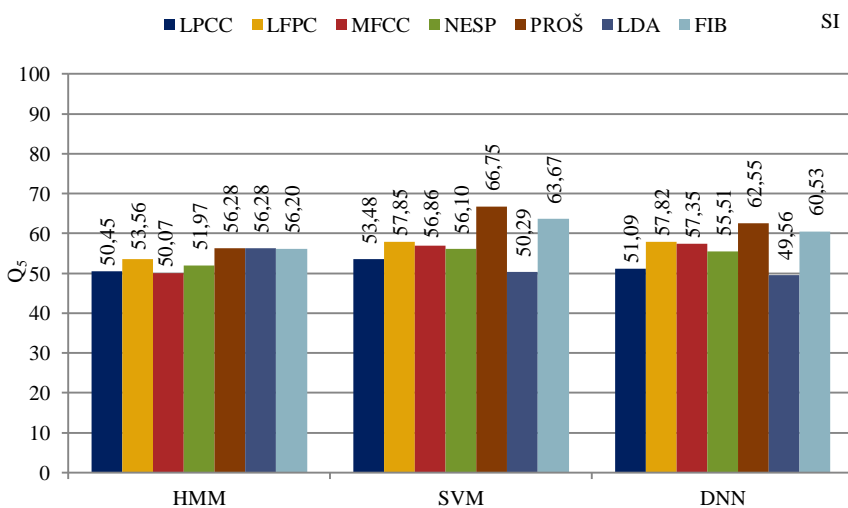
Poređenjem rezultata dobijenih upotrebom LDA redukcije i redukcije zasnovane na Fibonačijevom nizu se može zaključiti da primena Fibonačijevog niza daje iznenađujuće dobre rezultate, štaviše, za većinu izbora klasifikatora i baza se postižu čak i značajno bolji rezultati u odnosu na upotrebu LDA redukcije. Ukoliko se poredе rezultati postignuti primenom metoda za redukciju dimenzija sa rezultatima prethodnih eksperimenata, datim u Tab. 5.2, Tab. 6.2, Tab. 6.3 i Tab. 6.4, zaključuje se da za većininu kombinacija klasifikatora i baza upotreba proširenog skupa obeležja bez redukcije dimenzija obezbeđuje najbolje rezultate. Pažljivijom analizom rezultata se može prepoznati različit uticaj redukcije dimenzija na različite tipove klasifikatora, gde je uočljiva značajna degradacija performanse sa primenom LDA redukcije dimenzija kod SVM i DNN klasifikatora, dok se među konfiguracijama koje koriste HMM klasifikator beleži najbolji rezultat sa upotrebom LDA redukcije kod baza RUSLANA i Berlin. Primenom Fibonači redukcije nije ostvaren najbolji rezultat ni u jednoj kombinaciji klasifikatora i baza, ali su rezultati jako bliski najboljim ostvarenim.

Određivanjem koeficijenata uspešnosti obeležja i parova obeležje-klasifikator na osnovu rezultata ostvarenih upotrebom redukovanih vektora obeležja, Tab. 6.5, i prikazom zajedno sa

koeficijentima uspešnosti iz prethodnih eksperimenata, Sl. 6.10 i Sl. 6.11, se omogućava izvođenje opštih zaključaka o uticaju metoda za redukciju dimenzija na performanse klasifikatorskih sistema. Na osnovu Sl. 6.10 je jasno da se primenom LDA redukcije postižu niže vrednosti koeficijenata uspešnosti u odnosu na prošireni skup obeležja bez redukcije dimenzija, ali i u odnosu na većinu preostalih obeležja. Analizom koeficijenata uspešnosti parova obeležje-klasifikator, Sl. 6.11, se može konstatovati da HMM klasifikator sa uključenom LDA redukcijom ostvaruje iste performanse kao i sa proširenim skupom obeležja bez redukcije dimenzija, dok klasifikatori SVM i DNN ostvaruju najlošije performanse sa uključenom LDA redukcijom. Primenom Fibonači redukcije dimenzija su ostvareni koeficijenti uspešnosti obeležja jako bliski najboljim vrednostima za svaku od baza, kao i koeficijenti uspešnosti parova obeležje-klasifikator koji prate najbolje ostvarene vrednosti za svaki od klasifikatora.



Sl. 6.10 – Koeficijenti uspešnosti obeležja (proširenje: redukcija dimenzija)



Sl. 6.11 – Koeficijenti uspešnosti parova obeležje-klasifikator (proširenje: redukcija dimenzija)

Tokom analize metoda za redukciju dimenzija je vršeno usklađivanje dimenzija redukovanih vektora sa dimenzijama vektora pojedinačnih obeležja iz prethodnih eksperimenata, pri čemu nije isključeno da bi se drugačijim izborom dimenzija redukovanih vektora moglo uticati na popravku performansi klasifikatorskih sistema.

LDA redukcija dimenzija podrazumeva transformaciju datih vektora sa težnjom minimizacije rasipanja unutar klasa i maksimizacije rastojanja između klasa. Sprovođenje takve transformacije bi moglo da olakša zadatak klasifikacije i osigura dobre performanse sistema, međutim, uvidom u Tab. 6.5, Sl. 6.10 i Sl. 6.11 se dolazi do zaključka da jednostavna redukcija dimenzija primenom Fibonačijevog niza omogućila postizanje značajno boljih rezultata za većinu testiranih konfiguracija. S obzirom na ograničenje LDA metode u vidu podrazumevane unimodalne raspodele uzoraka svake klase, a sa druge strane, složene zavisnosti govornih obeležja u odnosu na sadržano emocionalno stanje, razlog za nezadovoljavajuće performanse se može tražiti u odstupanju vektora obeležja od unimodalne raspodele.

Sprovođenje redukcije dimenzija zasnovane na Fibonačijevom nizu brojeva je vršeno sa upotrebom ograničenog broja elemenata originalnog vektora pri formiranju jednog elementa redukovano vektora, dok je u formulaciji algoritma podrazumevano angažovanje svih elemenata originalnog vektora. Sa povećanjem indeksa, d , elementi Fibonačijevog niza, F_d , mogu da dostignu jako velike vrednosti, a sa druge strane, opseg brojnih vrednosti koji se može zabeležiti u memoriji računara je ograničen, tako da je za generisanje jednog elementa redukovano vektora korišćeno najviše 1200 elemenata originalnog vektora i isto toliko elemenata Fibonačijevog niza. Na taj način su izbegnuti preveliki elementi Fibonačijevog niza, a algoritam nije značajno narušen, s obzirom da dodani elementi svakako ne bi imali veliki uticaj. U slučaju težnje za jednovremenim angažovanjem većeg broja elemenata, bilo bi potrebno promeniti vrednosti od kojih kreće formiranje Fibonačijevog niza.

Rezultati testiranja su pokazali da Fibonači redukcija dimenzija sa nasumičnim udelima elemenata originalnog vektora obezbeđuje zadovoljavajuće performanse, ali se mogu očekivati još i bolji rezultati sa uključenjem informacija o prioritetima elemenata originalnog vektora na način da prioritetniji elementi učestvuju sa većim udelom u formiranju redukovano vektora. Drugi pravac razrade metode bi moglo da bude uvođenje optimizacionog kriterijuma i strategije pretrage, pri čemu bi se istraživanje moglo započeti usvajanjem performanse kompletnog sistema kao optimizacionog kriterijuma, a pretraga bi mogla da se sprovodi na slučajan način ili primenom genetskih algoritama. U ovom radu je potvrđeno da redukcija dimenzija zasnovana na Fibonačijevom nizu poseduje potencijal i dalje od toga neće biti ispitivana, pri čemu već osnovna forma metode može da nađe primenu u zadacima sa akcentom na računskoj efikasnosti i na mestima gde se želi izbeći formiranje glomaznih modela sa malim kompromisom u kvalitetu rezultata.

6.3 Konstrukcija multiklasifikator sistema

Na osnovu komparativne analize iz glave 5 i sprovedenih unapređenja u prethodnim poglavljima glave 6 se može zaključiti da je iznalaženje univerzalno najboljeg klasifikatorskog sistema veoma izazovan zadatak. Umesto kompromisa prilikom odabira najprihvatljivijeg rešenja, može se razmotriti integracija prednosti različitih pristupa u jedinstven sistem. Prednosti združivanja različitih pristupa su prepoznate i od strane drugih istraživača, tako da su već ispitivane mogućnosti prepoznavanja govornih emocija primenom različitih multiklasifikatorskih sistema, sa akcentom na hijerarhijskim strukturama [120], [121], [122], [123], [124], [125], [88]. Na ovom mestu je razmotrena konstrukcija paralelne multiklasifikatorske strukture sa ansamblom klasifikatorskih

sistema formiranih prilikom sprovođenja prethodnih eksperimenata. Pri tome nisu vršene nikakve izmene ranije formiranih klasifikatorskih sistema, kao ni izmene obučavajućih i testirajućih skupova uzoraka. U opštem slučaju, obezbeđenje validacionog skupa ima povoljan uticaj na formiranje multiklasifikatorskog sistema, ali je ipak za određivanje svih parametara multiklasifikatorske strukture korišćen isključivo obučavajući skup uzoraka. Zadržavanjem raspodele dostupnih uzoraka je obezbeđena potpuna uporedivost rezultata prethodnih eksperimenata sa rezultatima postignutim pomoću multiklasifikatorskog sistema.

U osnovi konstrukcije paralelne multiklasifikatorske strukture jeste formiranje kombinatornog pravila. Dizajn kombinatornog pravila je prevashodno uslovljen formom izlaza klasifikatora ansambla, pri čemu informativniji izlazi pružaju mogućnost konstrukcije efikasnijeg kombinatornog pravila. Izlazi svih klasifikatora korišćenih u ovom radu se mogu smatrati nosiocima verovatnoće pojedinačnih klasa, međutim, različite prirode klasifikatora uslovljavaju različite prirode izlaza, pa je teško odrediti preslikavanja koja obezbeđuju srazmernost izlaza različitih klasifikatora. Uspostavljanjem izlaza pojedinačnih klasifikatora ansambla u obliku rang liste se ostvaruje kompromis između srazmernosti izlaza i informacionog sadržaja. Uobičajeno se izlazu svakog klasifikatora ansambla pridružuje težinski koeficijent u skladu sa procenjenim stepenom pouzdanosti datog klasifikatora, a dodatno je moguće uključiti informacije iz matrice konfuzije kao sveobuhvatne mere kvaliteta klasifikatorskog sistema, što je razmatrano i u ranijim istraživanjima [126], [127].

Matrica konfuzije predstavlja tabelarni prikaz tačnih i pogrešnih klasifikacija, odakle se može interpretirati koliki broj od ukupno testiranih uzoraka date klase je klasifikovan u svaku od klasa. Sa druge strane, matrica konfuzije prikazuje koliki broj uzoraka iz svake od klasa je klasifikovan u datu klasu, a skaliranjem u odnosu na ukupan broj uzoraka klasifikovanih u datu klasu se dolazi do verovatnoća sa kojim klasifikovani uzorci potiču iz svake od klasa. Primeri matrica konfuzije skaliranih na predloženi način su prikazani u Tab. 6.6, Tab. 6.7 i Tab. 6.8. Formirane matrice konfuzije predstavljaju detaljniji prikaz rezultata koje je ostvario svaki od klasifikatora u testovima sa GEES bazom kada je za opisivanje govorne sekvence korišćen prošireni skup obeležja, odeljak 6.2.2. Na osnovu formiranih matrica konfuzije se može zaključiti da HMM klasifikator vrši tačno klasno pridruživanje neutralnom emocionalnom stanju sa verovatnoćom 0,67, a verovatnoća da tako klasifikovan uzorak zaista pripada klasi besa je 0,02, sreće 0,12, tuge 0,05 i straha 0,14. Slično, SVM klasifikator vrši ispravnu klasifikaciju uzorka kao neutralnog emocionalnog stanja sa verovatnoćom 0,72, a verovatnoće da uzorak zapravo pripada klasama besa, sreće, tuge i straha su redom 0,01, 0,13, 0,04 i 0,11. Na isti način je moguće tumačenje i preostalih elementa matrica konfuzije.

Tab. 6.6 – Matrica konfuzije HMM klasifikatora, GEES baza

| | | Testirane klase | | | | |
|------------------|--------|-----------------|------|-------|------|-------|
| | | Neutr. | Bes | Sreća | Tuga | Strah |
| Prepoznate klase | Neutr. | 0,67 | 0,02 | 0,12 | 0,05 | 0,14 |
| | Bes | 0,02 | 0,61 | 0,30 | 0,00 | 0,07 |
| | Sreća | 0,11 | 0,29 | 0,42 | 0,00 | 0,18 |
| | Tuga | 0,02 | 0,00 | 0,00 | 0,95 | 0,03 |
| | Strah | 0,03 | 0,04 | 0,08 | 0,00 | 0,86 |

Tab. 6.7 – Matrica konfuzije SVM klasifikatora, GEES baza

| | | Testirane klase | | | | |
|------------------|--------|-----------------|------|-------|------|-------|
| | | Neutr. | Bes | Sreća | Tuga | Strah |
| Prepoznate klase | Neutr. | 0,72 | 0,01 | 0,13 | 0,04 | 0,11 |
| | Bes | 0,02 | 0,72 | 0,17 | 0,00 | 0,09 |
| | Sreća | 0,06 | 0,21 | 0,65 | 0,00 | 0,08 |
| | Tuga | 0,04 | 0,00 | 0,00 | 0,91 | 0,05 |
| | Strah | 0,15 | 0,03 | 0,00 | 0,01 | 0,81 |

Tab. 6.8 – Matrica konfuzije DNN klasifikatora, GEES baza

| | | Testirane klase | | | | |
|------------------|--------|-----------------|------|-------|------|-------|
| | | Neutr. | Bes | Sreća | Tuga | Strah |
| Prepoznate klase | Neutr. | 0,84 | 0,01 | 0,10 | 0,00 | 0,06 |
| | Bes | 0,02 | 0,66 | 0,21 | 0,00 | 0,12 |
| | Sreća | 0,16 | 0,19 | 0,53 | 0,00 | 0,12 |
| | Tuga | 0,07 | 0,00 | 0,00 | 0,82 | 0,11 |
| | Strah | 0,18 | 0,03 | 0,02 | 0,02 | 0,75 |

Matrica konfuzije pruža detaljan uvid u performanse projektovanog sistema, pri čemu se za njeno određivanje uobičajeno koristi poseban skup uzoraka. Kako bi se omogućilo uključanje informacija o performansama pojedinačnih klasifikatora ansambla prilikom projektovanja kombinatornog pravila, nužna je procena matrica konfuzije na osnovu uzoraka koji su korišćeni i za obuku klasifikatora. Valjanost procena matrica konfuzije je uslovljena stepenom slaganja skupa korišćenog za obuku pojedinačnih klasifikatora sa skupom za testiranje kompletnog multiklasifikatorskog sistema, pri čemu povoljan uticaj na zadovoljenje uslova ima raspoloživost većih skupova uzoraka. Upotreba istih uzoraka za konstrukciju klasifikatora i određivanje performansi doprinosi izraženijim dijagonalnim elementima procenjenih matrica konfuzije, što umanjuje doprinos uključanja informacija iz matrica konfuzije, ali nema negativan uticaj na performanse kombinatornog pravila.

Ako je formiran ansambl od K klasifikatora sa izlazima pojedinačnih klasifikatora u obliku rang liste, tada se izlazu klasifikatora k , $k = 1, \dots, K$, može pridružiti vektor skorova, R_k , u skladu sa poretkom klasa u rang listi, tako da klasa sa indeksom i , $i = 1, \dots, M$, kojoj je dodeljen rang $r \in [1, \dots, M]$ ostvaruje skor $R_k(i) = p^{M-r}$, pri čemu je M broj klasa, a $p > 1$ definiše odnos skorova sukcesivno rangiranih klasa. Izuzetno, dozvoljen je izlaz proizvoljnog klasifikatora, k , sa jednakim rangovima dve ili više klasa, što iziskuje ostvarenje i istih skorova, pri čemu je potrebno da ostane zadovoljen uslov $\sum_{i=1}^M R_k(i) = \sum_{i=1}^M p^{i-1}$.

Svakom klasifikatoru ansambla k , $k = 1, \dots, K$, se pridružuju pokazatelji performansi u obliku tačnosti prepoznavanja, S_k , i matrice konfuzije, CM_k , procenjeni upotrebom obučavajućeg skupa uzoraka. Na osnovu pokazatelja performansi i skorova izlaza klasifikatora k se može definisati skor klase j , $j = 1, \dots, M$:

$$P_k(j) = S_k^q \sum_{i=1}^M R_k(i) CM_k(i, j). \quad (6.3)$$

U prethodnom izrazu, član $CM_k(i, j)$ predstavlja verovatnoću klase j u slučaju prepoznavanja klase i , dok $R_k(i)$ određuje stepen izvesnosti klase i . Sumiranjem proizvoda $R_k(i)CM_k(i, j)$ po svim

klasama, $i = 1, \dots, M$, se dolazi do skora klase j u odnosu na karakteristike i izlaz klasifikatora, a konačan skor se dobija dodatnim ponderisanjem u odnosu na pouzdanost klasifikatora, S_k^q , pri čemu je izborom eksponenta $q \in \mathbb{R}$, $q \geq 0$, moguće kontrolisati uticaj informacija o tačnosti prepoznavanja na konačnu odluku.

Ranije je napomenuto da se očekuju izraženiji dijagonalni elementi matrica konfuzije usled procene na osnovu obučavajućeg skupa uzoraka, što za posledicu ima isticanje informacija o rangu klasa prilikom računanja skora date klase, formula (6.3), pri čemu su informacije o rangu ujedno i najrelevantnije, s obzirom da se određuju na osnovu uzoraka koji su predmet klasifikacije.

Ukupan skor klase j je određen sumom skorova ostvarenih na osnovu pojedinačnih klasifikatora ansambla:

$$P(j) = \sum_{k=1}^K P_k(j). \quad (6.4)$$

Za konačnu odluku, j_{max} , se proglašava indeks klase koja ostvaruje najveći ukupan skor:

$$j_{max} = \operatorname{argmax}_{j=1, \dots, M} P(j). \quad (6.5)$$

Parametri p i q omogućavaju uobličenje uticaja angažovanih resursa pri donošenju konačne odluke. Na ovom mestu je usvojeno $p = 4$, što uzrokuje isticanje prvorangiranih klasa, a posledično i odgovarajućih elementa matrica konfuzije. Izborom parametra q je moguće isticanje odluka pouzdanijih klasifikatora, što može da bude posebno značajno kod neusaglašenih odlukama, ali se ne sme dozvoliti preveliki uticaj informacija o pouzdanosti klasifikatora procenjenih na osnovu obučavajućeg skupa, tako da je usvojeno $q = 1$.

6.3.1 Ansambl preuzetih klasifikatorskih sistema

Predloženi kriterijum za kombinovanje rezultata pojedinačnih klasifikatora ansambla paralelne multiklasifikatorske strukture je testiran sa klasifikatorskim sistemima formiranim prilikom sprovođenja prethodnih eksperimenata. Ansambl je sačinjavao 21 klasifikatorski sistem, pri čemu je varijetet konfiguracija postignut kombinovanjem tri klasifikatorske metode (HMM, SVM i DNN) i sedam metoda za određivanje obeležja govornog signala (LPCC, LFPC, MFCC, nespektralni skup obeležja, prošireni skup obeležja, prošireni skup obeležja sa LDA redukcijom dimenzija i prošireni skup obeležja sa Fibonači redukcijom dimenzija). Testovi su sprovedeni u SI postavci, sa bazama iz prethodnih eksperimenata (GEES, PES, RUSLANA i Berlin), a rezultati testiranja su dati u Tab. 6.9. Poređenjem rezultata multiklasifikatorske strukture sa rezultatima pojedinačnih klasifikatorskih sistema, Tab. 5.2, Tab. 6.2, Tab. 6.3, Tab. 6.4, Tab. 6.5, se zaključuje da formiranje multiklasifikatorske strukture obezbeđuje bolje rezultate u odnosu na pojedinačne primene.

Tab. 6.9 – Uspešnost klasifikacije – multiklasifikator

| Klasifikator | Baza | | | |
|-------------------|-------|-------|-------|--------|
| | GEES | PES | RUSL | Berlin |
| Multiklasifikator | 77,67 | 64,50 | 52,75 | 79,41 |

Prilikom konstrukcije multiklasifikatorskog sistema je posebno važno kombinovanje klasifikatorskih sistema različitih svojstava, tako da je izvršeno poređenje karakteristika parova klasifikatorskih sistema na primeru rezultata koje su ostvarili različiti klasifikatori sa upotrebom proširenog skupa obeležja za opisivanje govorne sekvence. Primenom mera predstavljenih u poglavlju 4.4, formule (4.91) do (4.94), se dolazi do rezultata sumiranih u Tab. 6.10. Izračunate mere ukazuju na veću različitost karakteristika HMM klasifikatora u odnosu na karakteristike SVM i DNN klasifikatora nego SVM i DNN klasifikatora međusobno, što opravdava ulaganje dodatnih napora ka unapređenju HMM klasifikatora. Sa druge strane, za konstrukciju uspešnog multiklasifikatorskog sistema je poželjna veća razlika karakteristika uključenih klasifikatorskih sistema, međutim, jedini izvor razlika poređenih sistema su klasifikatorske metode, takođe, mere su računane u odnosu na prve izbore klasifikatora, dok su u konstrukciji kombinatornog pravila korišćeni izlazi u formi rang liste.

Tab. 6.10 – Mere različitosti klasifikatora

| Mera različitosti | Ispitivani parovi | Baza | | | |
|-------------------|-------------------|--------|--------|--------|--------|
| | | GEES | PES | RUSL | Berlin |
| Q | HMM-SVM | 0,5622 | 0,6889 | 0,6630 | 0,6368 |
| | HMM-DNN | 0,5103 | 0,6128 | 0,4744 | 0,6876 |
| | SVM-DNN | 0,9020 | 0,8901 | 0,7156 | 0,8532 |
| ρ | HMM-SVM | 0,2719 | 0,3869 | 0,3703 | 0,3140 |
| | HMM-DNN | 0,2507 | 0,3388 | 0,2495 | 0,3494 |
| | SVM-DNN | 0,5882 | 0,5963 | 0,4166 | 0,5061 |
| Dis | HMM-SVM | 0,3022 | 0,3000 | 0,3207 | 0,2696 |
| | HMM-DNN | 0,3222 | 0,3250 | 0,3702 | 0,2549 |
| | SVM-DNN | 0,1667 | 0,1950 | 0,2948 | 0,1765 |
| DF | HMM-SVM | 0,1367 | 0,2550 | 0,3679 | 0,1299 |
| | HMM-DNN | 0,1511 | 0,2700 | 0,3767 | 0,1348 |
| | SVM-DNN | 0,1922 | 0,2950 | 0,3646 | 0,1446 |

Način formiranja i testiranja multiklasifikatorskog sistema je obezbedio potpunu uporedivost sa rezultatima prethodnih eksperimenata, pri čemu je moguće načiniti izmene koje bi omogućile ostvarenje boljih performansi multiklasifikatorskog sistema, ali bi narušile uporedivost sa prethodnim eksperimentima. Kod testirane multiklasifikatorske strukture je upitan kvalitet, kao i srazmernost, procenjenih performansi klasifikatora ansambla, obzirom na upotrebu istog skupa uzorka za obuku i testiranje, kao i različita generalizaciona svojstva klasifikatora. Određivanje matrica konfuzije i tačnosti prepoznavanja upotrebom posebnog validacionog skupa bi osiguralo valjanost i srazmernost procenjenih performansi. Dodatno, obezbeđenje validacionog skupa bi omogućilo optimizaciju parametara kombinatornog pravila, kao i konstruisanje složenijih kombinatornih pravila. Takođe, validacioni skup bi mogao biti iskorišćen za optimizaciju ansambla klasifikatora, gde je moguće usvajanje kriterijuma različitosti prema kom bi se vršilo uključivanje u ansambl samo onih klasifikatorskih sistema koji mogu da doprinesu popravci kvaliteta multiklasifikatorskog sistema.

6.3.2 Ansambl klasifikatorskih sistema upotrebom Fibonači redukcije

U odeljku 6.2.3 je ispitivana redukcija dimenzija vektora obeležja upotrebom Fibonačijevog niza brojeva, pri čemu su elementi redukovano vektora formirani nasumičnim kombinovanjem elemenata originalnog vektora. Određivanje elemenata redukovano vektora na slučajan način

uzrokuje drugačije rezultate sa svakom primenom algoritma, što je iskorišćeno za stvaranje varijacija prilikom formiranja ansambla klasifikatorskih sistema. Konstruisan je paralelan multiklasifikatorski sistem sa kombinatornim pravilom i klasifikatorima kao u prethodnom eksperimentu, a ansambl klasifikatorskih sistema je formiran tako što je za svaki klasifikator sedam puta ponavljana Fibonači redukcija dimenzija proširenog skup obeležja, kao u odeljku 6.2.3.

Testiranjem predložene multiklasifikatorske strukture u SI postavci, sa bazama iz prethodnih eksperimenata, se došlo do rezultata datih u Tab. 6.11. Poređenjem sa rezultatima prethodnih eksperimenata, Tab. 5.2, Tab. 6.2, Tab. 6.3, Tab. 6.4, Tab. 6.5, Tab. 6.9, se zaključuje da analizirana multiklasifikatorska struktura nadmašuje rezultate pojedinačnih klasifikatorskih sistema, a za većinu baza i rezultate prethodno analizirane multiklasifikatorske strukture. Povoljan uticaj na performanse analizirane multiklasifikatorske strukture se može pripisati upotrebi posebnih obeležja sa svakim od klasifikatora, dok su u prethodnom eksperimentu ista obeležja korišćena više puta, a postojalo je i preklapanje korišćenih obeležja. Takođe, predložena multiklasifikatorska struktura omogućava formiranje ansambla klasifikatorskih sistema proizvoljne veličine, dok je kod prethodno analizirane multiklasifikatorske strukture bilo neophodno zasebno definisanje svih obeležja korišćenih u formiranju klasifikatorskih sistema ansambla. Navedena fleksibilnost analizirane multiklasifikatorske strukture nije iskorišćena za optimizaciju performanse, već je akcenat bio na formiranju ansambla klasifikatorskih sistema istih dimenzija kao u prethodnom eksperimentu.

Tab. 6.11 – Uspešnost klasifikacije – multiklasifikator (Fibonači redukcija)

| Klasifikator | Baza | | | |
|-------------------|-------|-------|-------|--------|
| | GEES | PES | RUSL | Berlin |
| Multiklasifikator | 78,22 | 65,50 | 53,41 | 79,17 |

S obzirom na način formiranja ansambala klasifikatorskih sistema, korisno je sagledati u kojoj meri Fibonači redukcija dimenzija doprinosi ostvarenju različitih karakteristika klasifikatorskih sistema ansambla. Određene su mere različitosti za parove klasifikatorskih sistema koji koriste HMM klasifikator i obeležja konstruisana ponovljenim primenama Fibonači redukcije dimenzija, pri čemu je vršeno poređenje obeležja konstruisanih prvom primenom Fibonači redukcije sa obeležjima konstruisanim primenom preostalih šest iteracija Fibonači redukcije za isti klasifikator. Rezultati sumirani u Tab. 6.12 potvrđuju mogućnost ostvarenja različitosti ansambla klasifikatorskih sistema primenom Fibonači redukcije. Na osnovu izračunatih mera se uočava da je kod Berlin baze postignuta najmanja različitost klasifikatorskih sistema ansambla, što može biti uzrok nešto niže tačnosti prepoznavanja u odnosu na prethodni eksperiment, a sa druge strane, ukazuje na potrebu za razmatranjem različitosti klasifikatorskih sistema prilikom formiranja ansambla.

Kao i kod prethodno analizirane multiklasifikatorske strukture, formiranje validacionog skupa uzoraka bi doprinelo fleksibilnosti prilikom konstrukcije multiklasifikatorske strukture zasnovane na Fibonači redukciji dimenzija, a dodatno bi omogućilo optimizaciju ansambla ispitivanjem proizvoljnog broja klasifikatorskih sistema konstruisanih ponovljenim primenama Fibonači redukcije dimenzija. Takođe, obezbeđenje validacionog skupa bi omogućilo razmatranje proizvoljnih načina za uvođenje različitosti u ansambl klasifikatorskih sistema, a jedna mogućnost je varijacijom dimenzija vektora obeležja formiranih primenom Fibonači redukcije dimenzija.

Tab. 6.12 – Mere različitosti klasifikatora (Fibonači redukcija)

| Mera različitosti | Ispitivani parovi | Baza | | | |
|--------------------------|-------------------|--------|--------|--------|--------|
| | | GEES | PES | RUSL | Berlin |
| <i>Q</i> | FIB1-FIB2 | 0,6912 | 0,8525 | 0,9583 | 0,8949 |
| | FIB1-FIB3 | 0,8010 | 0,7959 | 0,7337 | 0,8895 |
| | FIB1-FIB4 | 0,7505 | 0,8193 | 0,7240 | 0,9056 |
| | FIB1-FIB5 | 0,7645 | 0,8678 | 0,7247 | 0,9223 |
| | FIB1-FIB6 | 0,8231 | 0,8185 | 0,7474 | 0,8740 |
| | FIB1-FIB7 | 0,8401 | 0,8000 | 0,7457 | 0,8862 |
| <i>ρ</i> | FIB1-FIB2 | 0,3894 | 0,5598 | 0,7403 | 0,6027 |
| | FIB1-FIB3 | 0,4871 | 0,4927 | 0,4299 | 0,5934 |
| | FIB1-FIB4 | 0,4409 | 0,5198 | 0,4217 | 0,6177 |
| | FIB1-FIB5 | 0,4544 | 0,5724 | 0,4218 | 0,6412 |
| | FIB1-FIB6 | 0,5106 | 0,5198 | 0,4415 | 0,5699 |
| | FIB1-FIB7 | 0,5531 | 0,4997 | 0,4402 | 0,5887 |
| <i>Dis</i> | FIB1-FIB2 | 0,2844 | 0,2200 | 0,1252 | 0,1789 |
| | FIB1-FIB3 | 0,2378 | 0,2550 | 0,2738 | 0,1814 |
| | FIB1-FIB4 | 0,2622 | 0,2400 | 0,2784 | 0,1691 |
| | FIB1-FIB5 | 0,2567 | 0,2150 | 0,2770 | 0,1569 |
| | FIB1-FIB6 | 0,2267 | 0,2400 | 0,2672 | 0,1912 |
| | FIB1-FIB7 | 0,2178 | 0,2500 | 0,2682 | 0,1838 |
| <i>DF</i> | FIB1-FIB2 | 0,2267 | 0,3800 | 0,5331 | 0,2525 |
| | FIB1-FIB3 | 0,2444 | 0,3850 | 0,4626 | 0,2451 |
| | FIB1-FIB4 | 0,2444 | 0,3600 | 0,4577 | 0,2451 |
| | FIB1-FIB5 | 0,2500 | 0,3550 | 0,4636 | 0,2402 |
| | FIB1-FIB6 | 0,2489 | 0,3700 | 0,4702 | 0,2377 |
| | FIB1-FIB7 | 0,2611 | 0,3600 | 0,4683 | 0,2451 |

Predloženi pristup za konstrukciju multiklasifikatorskog sistema sa upotrebom Fibonači redukcije za formiranje ansambla se može posmatrati kao generičko rešenje, s obzirom da je za rešavanje zadanog klasifikacionog problema dovoljno obezbeđenje proizvoljno velikog skupa obeležja i proizvoljnog skupa klasifikatora, a konačni multiklasifikatorski sistem se može automatski formirati bez potrebe za upuštanje u osobenosti problema koji se rešava. Sa druge strane, može se vršiti adaptacija u skladu sa konkretnim klasifikacionim zadatkom, gde je moguće razdvajanje problema na više manjih zadataka, što bi u kontekstu prepoznavanja emocija moglo da znači prvenstveno razdvajanje na grupe sličnih emocionalnih stanja, a u kasnijim fazama razdvajanje individualnih emocionalnih stanja.

7 Zaključak

Kreiranje veštačke inteligencije, kao sledećeg velikog tehnološkog dostignuća, je izuzetno složen zadatak, ali se problem može rešavati kroz niz zaokruženih celina. Neizostavni deo veštačke inteligencije jeste emocionalna inteligencija, s obzirom da predstavlja suštinu najvišeg stupnja inteligencije. Emocionalna inteligencija počiva na pouzdanom prepoznavanju emocionalnih stanja, gde se kao najznačajniji izvor emocionalnih informacija izdvaja paralingvistički sadržaj govora. Svakodnevno se radi na popravci performansi sistema za automatsko prepoznavanje govornih emocija pri čemu su u upotrebi najrazličitije klasifikatorske strukture. U ovom radu je ispitana podobnost tipično korišćenih obeležja i klasifikatora, a potom su razmotrene mogućnosti unapređenja, kao i mogućnosti fuzije različitih pristupa u jedan sistem.

U prvom delu rada je dat uvod u oblast i osnove manifestacija emocija kroz govor, a potom i teorijske osnove algoritama korišćenih u konstrukciji sistema za automatsko prepoznavanje govornih emocija. Predstavljani su algoritmi za izdvajanje obeležja govornog signala, kao i metode za određivanje statistika i redukciju dimenzija vektora obeležja, a zatim su predstavljani algoritmi za klasifikaciju, multiklasifikatorske strukture i mere kvaliteta klasifikatorskih sistema.

U nastavku rada je organizovan niz testova sa ciljem analize metoda za izdvajanje obeležja i klasifikaciju. U testiranju su uključena spektralna obeležja govornog signala (LPCC, LFPC, MFCC), često korišćeni klasifikatori (HMM, SVM, DNN) i četiri baze emocionalnog govora (GEES, PES, RUSLANA, Berlin). Sprovođenje testova za sve kombinacije obeležje-klasifikator-baza u postavci nezavisno od govornika i postavci zavisno od govornika je proizvelo veliki broj rezultata. Kako bi se olakšalo izvođenje zaključaka, pridružena je mera uspešnosti svakom obeležju, svakom klasifikatoru i svakom paru obeležje-klasifikator. Na osnovu uvedenih mera uspešnosti je bilo moguće izolovanje performanse pojedinačnih elemenata sistema, kao i sagledavanje uticaja baza na performanse sistema.

Jedan od zaključaka analize poznatih metoda je bio da diskretni HMM klasifikator pokazuje nešto lošije performanse od druga dva klasifikatora. S obzirom na svojstvenu strukturu modela, nije se odustalo od primene HMM, već su uloženi napor u smeru popravke performansi. Vektorska kvantizacija se ističe kao jako izazovan zadatak kod implementacije diskretnog HMM, tako da je njoj posvećena posebna pažnja. Razmotreno je nekoliko pristupa zasnovanih na upotrebi QQ krive kod određivanja etalona za vektorsku kvantizaciju. Nakon što je usvojena konačna forma vektorske kvantizacije, razmotrene su mogućnosti dodatnog unapređenja HMM. Uveden je korak predobuke modela, kao i višestruka obuka sa kombinovanjem modela. Predložena HMM implementacija je testirana upotrebom obeležja i baza sa kojima je testirana i standardna HMM implementacija, tako da su ostvareni rezultati nedvosmisleno potvrdili da predložena unapređenja doprinose popravci performansi u odnosu na standardnu implementaciju, premda ni unapređeni HMM klasifikator nije uspeo da nadmaši performanse druga dva klasifikatora. Predloženi algoritam je moguće dalje doradivati, pri čemu se prostor za popravku uočava kod prevođenja vektora u skalarnu formu za potrebe konstrukcije QQ krive. U ovom radu je određivanje skalarnih predstava vršeno upotrebom Fibonačijevog niza brojeva kao težinskih koeficijenata i prioriteta koordinata postavljenih na slučajan način, a rezultati eksperimenata su potvrdili opravdanost dalje razrade rešenja. Predložena implementacija HMM uspeva da popravi performanse bez izmene strukture modela ili procedure

testiranja, što pojednostavljuje integraciju u postojeća rešenja, projektovana u skladu sa standardnom HMM implementacijom.

U analizama su prevashodno korišćena spektralna obeležja govornog signala, s obzirom na uverenje da su dobar izbor za zadatak automatskog prepoznavanja govornih emocija, što ne isključuje mogućnost postojanja korisnih informacija i u drugim grupama obeležja. Verifikacija postojanja korisnih informacija izvan spektralnih obeležja je izvršena testiranjem skupa obeležja sačinjenog od predstavnika preostalih grupa. S obzirom da su nespektralna obeležja postigla ujednačene performanse sa spektralnim obeležjima, i s obzirom na različite prirode obeležja, zaključeno je da se njihovim kombinovanjem mogu obezbediti potpunije informacije. Uključenjem velikog broja predstavnika različitih grupa obeležja u jedinstven prošireni skup obeležja je obezbeđena značajna popravka performansi. Sa proširenjem skupa obeležja se javlja problem manipulacije vektorima velikih dimenzija, takođe, dolazi do proređenja prostora obeležja, tako da je bilo neophodno razmatranje mogućnosti redukcije dimenzija vektora obeležja, uz što manje gubitke korisnih informacija. Eksperimenti su sprovedeni upotrebom LDA metode za redukciju dimenzija, a analizirana je i mogućnost redukcije dimenzija upotrebom Fibonačijevog niza brojeva. Primena metoda za redukciju dimenzija je uticala na degradaciju tačnosti prepoznavanja, pri čemu je redukcija dimenzija upotrebom Fibonačijevog niza izazvala mnogo manji poremećaj rezultata. Ostvarene performanse Fibonači redukcije dimenzija su ohrabrujuće u pogledu opravdanosti daljih analiza, s obzirom da se može prepoznati potreba za dodatnim testiranjem i proširenjem metode. Na ovom mestu je Fibonači redukcija dimenzija vršena sa nasumičnim udelima pojedinačnih elemenata originalnog vektora obeležja, pri čemu se može očekivati popravka performansi sa uključenjem informacija o prioritetu elemenata originalnog vektora obeležja. Takođe, računaska efikasnost metode omogućava optimizaciju proizvoljnog kriterijuma, ispitivanjem velikog broja rezultata obezbeđenih ponovljenim primenama Fibonači redukcije.

Nakon što su ispitane mogućnosti različitih popravki pojedinačnih elemenata klasifikatorskih sistema, preostala je da se razmotri integracija prednosti različitih pristupa u jedinstven sistem. Konstruisana je paralelna multiklasifikatorska struktura sa kombinatornim pravilom koje pored rezultata klasifikacije datog klasifikatorskog sistema koristi i informacije o karakteristikama klasifikatorskog sistema izraženim kroz matricu konfuzije i tačnost prepoznavanja. Predloženo kombinatorno pravilo je testirano sa klasifikatorskim sistemima formiranim prilikom sprovođenja prethodnih eksperimenata, a poređenjem rezultata multiklasifikatorske strukture sa rezultatima pojedinačnih klasifikatorskih sistema je potvrđena popravka tačnosti prepoznavanja. Takođe, razmotreno je formiranje ansambla klasifikatorskih sistema proizvoljne veličine upotrebom Fibonači redukcije dimenzija, što je za većinu baza doprinelo popravci rezultata u odnosu na prethodno ispitivanu multiklasifikatorsku strukturu, tako da se može zaključiti da su predloženi pristupi za formiranje kombinatornog pravila i ansambla klasifikatorskih sistema vredni pažnje budućih istraživanja. U ovom radu su multiklasifikatorske strukture konstruisane bez formiranja validacionog skupa uzoraka, što je omogućilo poređenje sa rezultatima prethodnih eksperimenata, ali je uticalo na lošu procenu karakteristika klasifikatorskih sistema ansambla. Uvođenje validacionog skupa bi omogućilo konstrukciju kombinatornog pravila na osnovu boljih procena karakteristika klasifikatorskih sistema, kao i razmatranje različitosti klasifikatorskih sistema prilikom formiranja ansambla.

Cilj rada je bila optimizacija sistema za automatsko prepoznavanje govornih emocija, kako u pogledu tačnosti prepoznavanja, tako i robusnosti. Predložena unapređenja su testirana sa većim brojem baza emocionalnog govora, a date su i smernice za dalje razrade predloženih izmena. Akcenat eksperimenata je bio na verifikaciji pozitivnih doprinosa predloženih ideja, pri čemu je moguće dodatno prilagođenje parametara prema konkretnim primenama. Eksperimenti su vršeni sa bazama odglumljenog emocionalnog govora kao dobrim reprezentima prototipskih emocionalnih stanja, čime su stvoreni povoljni uslovi za prepoznavanje akustičkih promena usled različitih emocionalnih stanja.

Međutim, krajnji cilj je primena projektovanih sistema u realnim uslovima, gde je često potrebno prepoznati naznake akustičkih promena koje su jasno uočljive kod odglumljenih baza. Sa druge strane, klasifikacija emocionalnih stanja u realnim uslovima obezbeđuje određene pogodnosti, kao što je mogućnost prevazilaženja ograničenog broja uzoraka za obuku, ili upotreba lingvističkog sadržaja. Primena sistema za prepoznavanje govornih emocija u realnim uslovima omogućava praćenje verovatnoće određenih reči u odnosu na emocionalna stanja, ili detekciju reči iz kojih se mogu pouzdano prepoznati razlike između emocionalnih stanja. Međusobnim dopunjavanjem različitih informacija sadržanih u govoru, kao i upotrebom drugih dostupnih izvora, je moguća realizacija sistema koji bi vršili prepoznavanje emocionalnih stanja slično ljudima i po principu i po rezultatima, pa i bolje.

8 Literatura

- [1] V. Delić, Z. Perić, M. Sečujski, N. Jakovljević, J. Nikolić, D. Mišković, N. Simić, S. Suzić i T. Delić, „Speech Technology Progress Based on New Machine Learning Paradigm,“ *Computational Intelligence and Neuroscience*, p. 2019:4368036, 2019.
- [2] D. Schuller i B. W. Schuller, „The Age of Artificial Emotional Intelligence,“ *Computer*, t. 51, br. 9, pp. 38-46, 2018.
- [3] P. Ekman, *Emotions Revealed : Recognizing Faces and Feelings to Improve Communication and Emotional Life*, New York: Henry Holt, 2007.
- [4] D. Gilbert, *Stumbling on Happiness*, New York: Knopf, 2006.
- [5] D. Keltner, K. Oatley i J. M. Jenkins, *Understanding Emotions*, New Jersey: Wiley & Sons, 2014.
- [6] R. W. Picard, „Affective Computing,“ MIT Press, Cambridge, 1995.
- [7] F. Dellaert, T. Polzin i A. Waibel, „Recognizing Emotion in Speech,“ u *Proceeding of Fourth International Conference on Spoken Language Processing*, Philadelphia, 1996.
- [8] J. Williamson, „Speech Analyzer for Analyzing Pitch or Frequency Perturbations in Individual Speech Pattern to Determine the Emotional State of the Person“. U.S. Patent 4 093 821, 1978.
- [9] B. W. Schuller, „Speech Emotion Recognition: Two Decades in a Nutshell, Benchmarks, and Ongoing Trends,“ *Communications of the ACM*, t. 61, br. 5, pp. 90-99, 2018.
- [10] C. M. Lee i S. S. Narayanan, „Toward Detecting Emotions in Spoken Dialogs,“ *IEEE Transactions on Speech and Audio Processing*, t. 13, br. 2, pp. 293-303, 2005.
- [11] H. Ai, D. J. Litman, K. Forbes-Riley, M. Rotaru, J. Tetreault i A. Purandare, „Using System and User Performance Features to Improve Emotion Detection in Spoken Tutoring Dialogs,“ u *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, Pittsburgh, 2006.
- [12] O. A. Schipor, „Improving Computer Assisted Speech Therapy Through Speech Based Emotion Recognition,“ u *The 10th International Scientific Conference eLearning and software for Education*, Bucharest, 2014.
- [13] B. Schuller, G. Rigoll, M. Grimm, K. Kroschel, T. Moosmayr i G. Ruske, „Effects of In-Car Noise-Conditions on the Recognition of Emotion within Speech,“ u *Proceedings of the 33rd Annual Conference on Acoustics, DAGA'07*, Stuttgart, 2007.

- [14] C. Jones i I.-M. Jonsson, „Using Paralinguistic Cues in Speech to Recognise Emotions in Older Car Drivers,“ u *Affect and Emotion in Human-Computer Interaction: From Theory to Applications. Lecture Notes in Computer Science*, Berlin, 2008.
- [15] T. Rudra, M. Kavakli i D. Tien, „Emotion Detection from Male Speech in Computer Games,“ u *Proceedings of IEEE Region 10 Annual International Conference, TENCON*, Taipei, 2007.
- [16] M. Bashirpour i M. Geravanchizadeh, „Robust Emotional Speech Recognition Based on Binaural Model and Emotional Auditory Mask in Noisy Environments,“ *EURASIP Journal on Audio, Speech, and Music*, p. 2018:9, 2018.
- [17] M. Milošević, Ž. Nedeljković, U. Glavitsch i Ž. Đurović, „Speaker Modeling Using Emotional Speech for More Robust Speaker Identification,“ *Journal of Communications Technology and Electronics*, t. 64, br. 11, pp. 1256-1265, 2019.
- [18] Y. B. Singh i S. Goel, „Survey on Human Emotion Recognition: Speech Database, Features and Classification,“ u *2018 International Conference on Advances in Computing, Communication Control and Networking, ICACCCN*, Greater Noida, 2018.
- [19] R. B. Adler, L. B. Rosenfeld i R. F. Proctor, *Interplay: The Process of Interpersonal Communication*, 8 ur., Fort Worth: Harcourt College Publishers, 2001.
- [20] L. Deng i D. O'Shaughnessy, *Speech Processing: A Dynamic and Optimization-Oriented Approach*, New York: Marcel Dekker Inc., 2003.
- [21] C. Darwin, *The Expression of the Emotions in Man and Animals*, London: John Murray, 1872.
- [22] W. James, „What is an Emotion?,“ *Mind*, t. 9, br. 34, pp. 188-205, 1884.
- [23] R. S. Lazarus, „Progress on a Cognitive-Motivational-Relational Theory of Emotion,“ *American Psychologist*, t. 46, br. 8, pp. 819-834, 1991.
- [24] S. S. Brehm i S. M. Kassin, *Social Psychology*, 4 ur., Boston: Houghton Mifflin Company, 1999.
- [25] K. R. Scherer, „Psychological Models of Emotion,“ *The neuropsychology of emotion*, pp. 137-166, 2000.
- [26] T. Johnstone i K. Scherer, „Vocal Communication of Emotion,“ u *Handbook of Emotions*, New York, London, 2000.
- [27] D. Ververidis i C. Kotropoulos, „Emotional Speech Recognition: Resources, Features and Methods,“ *Speech Communication*, t. 48, pp. 1162-1181, 2006.
- [28] K. R. Scherer, „On the Nature and Function of Emotion: A Component Process Approach,“ u *Approaches to emotion*, New Jersey, 1984.
- [29] R. Plutchik, „The Nature of Emotions,“ *American Scientist*, t. 89, br. 4, pp. 344-350, 2001.
- [30] P. Ekman, „Facial Expression and Emotion,“ *American Psychologist*, t. 48, br. 4, pp. 384-392, 1993.

- [31] J. A. Russell, „A Circumplex Model of Affect,“ *Journal of Personality and Social Psychology*, t. 39, br. 6, pp. 1161-1178, 1980.
- [32] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz i J. G. Taylor, „Emotion Recognition in Human-Computer Interaction,“ *IEEE Signal Processing Magazine*, t. 18, br. 1, pp. 32-80, 2001.
- [33] B. Schuller, B. V. Vlasenko, F. Eyben, G. Rigoll i A. Wendemuth, „Acoustic Emotion Recognition: A Benchmark Comparison of Performances,“ u *Proceedings of the 2009 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2009*, Merano, 2009.
- [34] B. Schuller, S. Steidl i A. Batliner, „The Interspeech 2009 Emotion Challenge,“ u *Proceedings of 10th Annual of the International Speech Communication Association, Interspeech'09*, Brighton, 2009.
- [35] B. Schuller, G. Rigoll i M. Lang, „Speech Emotion Recognition Combining Acoustic Features and Linguistic Information in a Hybrid Support Vector Machine-Belief Network Architecture,“ u *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'04*, Montreal, 2004.
- [36] M. M. E. Ayadi, M. S. Kamel i F. Karray, „Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases,“ *Pattern Recognition*, t. 44, br. 3, pp. 572-587, 2011.
- [37] L. Rabiner i B.-H. Juang, *Fundamentals of Speech Recognition*, New Jersey: Prentice Hall, 1993.
- [38] T. L. Nwe, S. W. Foo i L. C. D. Silva, „Speech Emotion Recognition Using Hidden Markov Models,“ *Speech Communication*, t. 41, br. 4, pp. 603-623, 2003.
- [39] S. B. Davis i P. Mermelstein, „Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences,“ *IEEE Transactions on Acoustics, Speech, and Signal Processing*, t. 28, br. 4, pp. 357-366, 1980.
- [40] A. M. Noll, „Cepstrum Pitch Determination,“ *The Journal of the Acoustical Society of America*, t. 41, br. 2, pp. 293-309, 1967.
- [41] R. C. Snell i F. Milinazzo, „Formant Location from LPC Analysis Data,“ *IEEE Transactions on Speech and Audio Processing*, t. 1, br. 2, pp. 129-134, 1993.
- [42] R. Fernandez i R. W. Picard, „Classical and Novel Discriminant Features for Affect Recognition from Speech,“ u *Proceedings of 9th European Conference on Speech Communication and Technology, Interspeech'05*, Lisbon, 2005.
- [43] P. Boersma, „Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonic-to-Noise Ratio of a Sample Sound,“ u *IFA Proceedings 17*, 1993.
- [44] H. M. Teager i S. M. Teager, „Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract,“ u *Speech Production and Speech Modelling. NATO ASI Series (Series D: Behavioural and Social Sciences)*, Dordrecht, 1990.

- [45] J. F. Kaiser, „On a Simple Algorithm to Calculate the 'energy' of a Signal,“ u *International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, 1990.
- [46] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, A. Ragni, V. Valtchev, P. Woodland i C. Zhang, *The HTK Book*, 3.5a ur., Cambridge: Cambridge University Engineering Department, 2015.
- [47] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Boston: Academic Press, 1990.
- [48] L. Rabiner, „A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,“ *Proceedings of the IEEE*, t. 77, br. 2, pp. 257-286, 1989.
- [49] Y.-L. Lin i G. Wei, „Speech Emotion Recognition Based on HMM and SVM,“ u *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, Guangzhou, 2005.
- [50] A. J. Viterbi, „Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm,“ *IEEE Transactions on Information Theory*, t. 13, br. 2, pp. 260-269, 1967.
- [51] L. E. Baum i T. Petrie, „Statistical Inference for Probabilistic Functions of Finite State Markov Chains,“ *Annals of Mathematical Statistics*, t. 37, br. 6, pp. 1554-1563, 1966.
- [52] L. Bahl, P. Brown, P. d. Souza i R. Mercer, „A New Algorithm for the Estimation of Hidden Markov Model Parameters,“ u *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, New York, 1988.
- [53] S. Lloyd, „Least Squares Quantization in PCM,“ *IEEE Transactions on Information Theory*, t. 28, br. 2, pp. 129-137, 1982.
- [54] C. Cortes i V. Vapnik, „Support-Vector Networks,“ *Machine Learning*, t. 20, pp. 273-297, 1995.
- [55] V. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer, 1995.
- [56] P.-N. Tan, M. Steinbach i V. Kumar, *Introduction to Data Mining*, Boston: Pearson, 2006.
- [57] V. Vapnik, *Estimation of Dependences Based on Empirical Data, Addendum 1*, New York: Springer - Verlag, 1982.
- [58] G. Arfken, „Lagrange Multipliers,“ u *Mathematical Methods for Physicists*, 3 ur., Orlando, Academic Press, 1985, pp. 945-950.
- [59] H. W. Kuhn i A. W. Tucker, „Nonlinear Programming,“ u *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, 1951.
- [60] B. E. Boser, I. M. Guyon i V. N. Vapnik, „A Training Algorithm for Optimal Margin Classifiers,“ u *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, 1992.
- [61] K.-B. Duan i S. S. Keerthi, „Which Is the Best Multiclass SVM Method? An Empirical Study,“ u *Multiple Classifier Systems. MCS 2005. Lecture Notes in Computer Science*, Heidelberg, 2005.

- [62] C.-W. Hsu i C.-J. Lin, „A Comparison of Methods for Multiclass Support Vector Machines,“ *IEEE Transactions on Neural Networks*, t. 13, br. 2, pp. 415-425, 2002.
- [63] L. Deng i D. Yu, „Deep Learning: Methods and Applications,“ *Foundations and Trends in Signal Processing*, t. 7, br. 3-4, pp. 197-387, 2013.
- [64] W. S. McCulloch i W. Pitts, „A Logical Calculus of the Ideas Immanent in Nervous Activity,“ *The Bulletin of Mathematical Biophysics*, t. 5, pp. 115-133, 1943.
- [65] C. M. Bishop, *Neural Networks for Pattern Recognition*, New York: Oxford University Press, 1995.
- [66] H. J. Kelley, „Gradient Theory of Optimal Flight Paths,“ *ARS Journal*, t. 30, br. 10, pp. 947-954, 1960.
- [67] A. E. Bryson, „A Gradient Method for Optimizing Multi-stage Allocation Processes,“ u *Proceedings of the Harvard Univ. Symposium on Digital Computers and Their Applications*, Cambridge, 1962.
- [68] H. B. Curry, „The Method of Steepest Descent for Non-linear Minimization Problems,“ *Quarterly of Applied Mathematics*, t. 2, br. 3, pp. 258-261, 1944.
- [69] R. Fletcher, *Practical Methods of Optimization*, 2 ur., New York: John Wiley & Sons, 2000.
- [70] E. M. Johansson, F. U. Dowla i D. M. Goodman, „Backpropagation Learning for Multi-Layer Feed-Forward Neural Networks Using the Conjugate Gradient Method,“ *International Journal of Neural Systems*, t. 2, br. 4, pp. 291-301, 1991.
- [71] R. Battiti i F. Masulli, „BFGS Optimization for Faster and Automated Supervised Learning,“ u *International Neural Network Conference*, Dordrecht, 1990.
- [72] M. F. Møller, „A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning,“ *Neural Networks*, t. 6, br. 4, pp. 525-533, 1993.
- [73] S. Lange i M. Riedmiller, „Deep Auto-Encoder Neural Networks in Reinforcement Learning,“ u *The 2010 International Joint Conference on Neural Networks*, Barcelona, 2010.
- [74] G. Fumera i F. Roli, „A Theoretical and Experimental Analysis of Linear Combiners for Multiple Classifier Systems,“ *IEEE Transactions on Pattern Analysis and Machine Intelligence*, t. 27, br. 6, pp. 942-956, 2005.
- [75] L. I. Kuncheva i C. J. Whitaker, „Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy,“ *Machine Learning*, t. 51, p. 181–207, 2003.
- [76] A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch i M. R. Wróbel, „Emotion Recognition and Its Applications,“ u *Human-Computer Systems Interaction: Backgrounds and Applications 3. Advances in Intelligent Systems and Computing*, Cham, 2014.
- [77] C. Vinola i K. Vimaladevi, „A Survey on Human Emotion Recognition Approaches, Databases and Applications,“ *Electronic Letters on Computer Vision and Image Analysis*, t. 14, br. 2, pp. 24-44, 2015.

- [78] M. D. Pell, S. Paulmann, C. Dara, A. Alasserri i S. A. Kotz, „Factors in the Recognition of Vocally Expressed Emotions: A Comparison of Four Languages,“ *Journal of Phonetics*, t. 37, br. 4, pp. 417-435, 2009.
- [79] M. D. Pell, L. Monetta, S. Paulmann i S. A. Kotz, „Recognizing Emotions in a Foreign Language,“ *Journal of Nonverbal Behavior*, t. 33, br. 2, pp. 107-120, 2009.
- [80] S. T. Jovičić, Z. Kašić, M. Đorđević i M. Rajković, „Serbian Emotional Speech Database: Design, Processing and Evaluation,“ u *Proceedings of the 9th International Conference Speech and Computer*, St. Petersburg, 2004.
- [81] J. Cichosz, „Database of Polish Emotional Speech,“ 2008. [Na mreži]. Available: <http://www.elel.p.lodz.pl/med/eng>. [Poslednji pristup 16 October 2015].
- [82] M. Igras i B. Ziólko, „Database of Emotional Speech Recordings,“ *Studia Informatica*, t. 34, br. 2B, pp. 67-77, 2013.
- [83] P. Staroniewicz i W. Majewski, „Polish Emotional Speech Database – Recording and Preliminary Validation,“ u *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions. Lecture Notes in Computer Science*, Berlin, 2009.
- [84] V. Makarova i V. A. Petrushin, „RUSLANA: A Database of Russian Emotional Utterances,“ u *Proceedings of the 7th International Conference on Spoken Language Processing*, Denver, 2002.
- [85] B. Dropuljić, M. T. Chmura, A. Kolak i D. Petrinovic, „Emotional Speech Corpus of Croatian Language,“ u *Proceedings of the 7th International Symposium on Image and Signal Processing and Analysis*, Dubrovnik, 2011.
- [86] D. Uhrin, P. Partila, M. Vozňák, Z. Chmelikova, M. Hložák i L. Orčík, „Design and Implementation of Czech Database of Speech Emotions,“ u *Proceedings of the 22nd Telecommunications Forum*, Belgrade, 2014.
- [87] T. Justin, V. Štruc, J. Žibert i F. Mihelič, „Development and Evaluation of the Emotional Slovenian Speech Database - EmoLUKS,“ u *Text, Speech, and Dialogue, Lecture Notes in Computer Science*, Cham, 2015.
- [88] Ž. Nedeljković i Ž. Đurović, „Automatic Emotion Recognition from Speech Using Hidden Markov Models,“ u *Proceedings of 59th Conference on Electrical, Electronic and Computing Engineering*, Silver Lake, 2015.
- [89] A. Hassan i R. I. Damper, „Multi-Class and Hierarchical SVMs for Emotion Recognition,“ u *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, Makuhari, 2010.
- [90] M. Milošević, Ž. Nedeljković i Ž. Đurović, „SVM Classifier for Emotional Speech Recognition in Software Environment SEBAS,“ u *Proceedings of 3rd International Conference on Electrical, Electronic and Computing Engineering*, Zlatibor, 2016.

- [91] D. Kamińska, T. Sapiński, D. Niewiadomy i A. Pelikant, „Comparison of Perceptual Features Efficiency for Automatic Identification of Emotional States from Speech Signal,” *Studia Informatica*, t. 34, br. 2B, pp. 59-66, 2013.
- [92] K. Ślot, L. Bronakowski, J. Cichosz i H. Kim, „Application of Poincare-Mapping of Voiced-Speech Segments for Emotion Sensing,” *Sensors*, t. 9, br. 12, pp. 9858-9872, 2009.
- [93] D. Kamińska, T. Sapiński i G. Anbarjafari, „Efficiency of Chosen Speech Descriptors in Relation to Emotion Recognition,” *EURASIP Journal on Audio, Speech, and Music Processing*, p. 2017:3, 2017.
- [94] B. Dropuljić, S. Skansi i R. Kopal, „Croatian Emotional Speech Analyses on a Basis of Acoustic and Linguistic Features,” *International Journal of Digital Technology & Economy*, t. 1, br. 2, pp. 85-96, 2016.
- [95] V. Makarova i V. A. Petrushin, „Phonetics: Tracing Emotions in Russian Vowels,” u *Russian language studies in North America: New perspectives from theoretical and applied linguistics*, London, New York, 2012.
- [96] F. Eyben i B. Schuller, „OpenSMILE:): The Munich Open-Source Large-Scale Multimedia Feature Extractor,” *ACM SIGMultimedia Records*, t. 6, br. 4, pp. 4-13, 2014.
- [97] V. Delić, M. Bojanić, M. Gnjatović, M. Sečujski i S. T. Jovičić, „Discrimination Capability of Prosodic and Spectral Features for Emotional Speech Recognition,” *Elektronika ir Elektrotehnika*, t. 18, br. 9, pp. 51-54, 2012.
- [98] A. Shaukat i K. Chen, „Emotional State Recognition from Speech via Soft-Competition on Different Acoustic Representations,” u *The 2011 International Joint Conference on Neural Networks*, San Jose, 2011.
- [99] A. Shaukat i K. Chen, „Towards Automatic Emotional State Categorization from Speech Signals,” u *Proceedings of the Annual Conference of the International Speech Communication Association*, Brisbane, 2008.
- [100] B. Popović, I. Stanković i S. Ostrogonac, „Temporal Discrete Cosine Transform for Speech Emotion Recognition,” u *Proceedings of IEEE 4th International Conference on Cognitive Infocommunications*, Budapest, 2013.
- [101] M. Bojanić, V. Delić i M. Sečujski, „Relevance of the Types and the Statistical Properties of Features in the Recognition of Basic Emotions in Speech,” *Facta Universitatis - Series: Electronics and Energetics*, t. 27, br. 3, pp. 425-433, 2014.
- [102] B. Dropuljić, S. Skansi i R. Kopal, „Analyzing Affective States Using Acoustic and Linguistic Features,” u *Proceedings of Central European Conference on Information and Intelligent Systems*, Varaždin, 2016.
- [103] P. Staroniewicz, „Automatic Recognition of Emotional State in Polish Speech,” u *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues, Lecture Notes in Computer Science*, Berlin, 2011.

- [104] Ž. Nedeljković, M. Milošević i Ž. Đurović, „Analysis of Features and Classifiers in Emotion Recognition Systems: Case Study of Slavic Languages,“ *Archives of Acoustics*, t. 45, br. 1, pp. 129-140, 2019.
- [105] F. Burkhardt, A. Paeschke, M. A. Rolfes, W. F. Sendlmeier i B. Weiss, „A Database of German Emotional Speech,“ u *Proceedings of the 9th European Conference on Speech Communication and Technology, INTERSPEECH 2005*, Lisbon, 2005.
- [106] E. M. Albornoz, M. E. Sanchez-Gutierrez, F. M. Martinez-Licon, H. L. Rufiner i J. C. Goddard, „Spoken Emotion Recognition Using Deep Learning,“ u *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Lecture Notes in Computer Science*, Charm, 2014.
- [107] R. B. Lanjewar i D. S. Chaudhari, „Comparative Analysis of Speech Emotion Recognition System Using Different Classifiers on Berlin Emotional Speech Database,“ *International Journal of Electrical and Electronics Engineering Research*, t. 3, br. 5, pp. 145-156, 2011.
- [108] U. Jain, K. Nathani, N. Ruban, A. N. J. Raj, Z. Zhuang i V. G. V. Mahesh, „Cubic SVM Classifier Based Feature Extraction and Emotion Detection from Speech Signals,“ u *2018 International Conference on Sensor Networks and Signal Processing*, Xi'an, 2018.
- [109] S. K. Pandey, H. S. Shekhawat i S. R. M. Prasanna, „Deep Learning Techniques for Speech Emotion Recognition: A Review,“ u *2019 29th International Conference Radioelektronika*, Pardubice, 2019.
- [110] H. Farsi i R. Saleh, „Implementation and Optimization of Speech Recognition System Based on Hidden Markov Model Using Genetic Algorithm,“ u *2014 Iranian Conference on Intelligent Systems*, Bam, 2014.
- [111] Z. Han, S. Lun i J. Wang, „Speech Emotion Recognition System Based on Integrating Feature and Improved HMM,“ u *Proceedings of the 2012 International Conference on Computer Application and System Modeling*, Paris, 2012.
- [112] X. Mao, L. Chen i L. Fu, „Multi-level Speech Emotion Recognition Based on HMM and ANN,“ u *2009 WRI World Congress on Computer Science and Information Engineering*, Los Angeles, 2009.
- [113] L. Li, Y. Zhao, D. Jiang, Y. Zhang, F. Wang, I. Gonzalez, E. Valentin i H. Sahli, „Hybrid Deep Neural Network--Hidden Markov Model (DNN-HMM) Based Speech Emotion Recognition,“ u *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Geneva, 2013.
- [114] J. M. Koo, H. S. Lee i C. K. Un, „An Improved VQ Codebook Design Algorithm for HMM,“ u *1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, 1992.
- [115] Ž. Nedeljković i Ž. Đurović, „QQ-plot Based Clustering,“ u *Proceedings of 6th International Conference on Electrical, Electronic and Computing Engineering*, Silver Lake, 2019.
- [116] Ž. Nedeljković, M. Milošević i Ž. Đurović, „Improved HMM for Automatic Speech Emotion Recognition,“ u *Proceedings of XXVII Pacific Voice Conference*, Krakow, 2020.

- [117] M. B. Wilk i R. Gnanadesikan, „Probability Plotting Methods for the Analysis of Data,“ *Biometrika*, t. 55, br. 1, pp. 1-17, 1968.
- [118] Ž. Đurović, B. Kovačević i V. Barroso, „QQ-Plot Based Probability Density Function Estimation,“ u *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing*, Pocono Manor, 2000.
- [119] R. Rawat, R. Nayak, Y. Li i S. Alsaleh, „Aggregate Distance Based Clustering Using Fibonacci Series-FIBCLUS,“ u *Web Technologies and Applications. APWeb 2011. Lecture Notes in Computer Science*, Berlin, 2011.
- [120] X. Ma, Z. Wu, J. Jia, M. Xu, H. M. Meng i L. Cai, „Speech Emotion Recognition with Emotion-Pair Based Framework Considering Emotion Distribution Information in Dimensional Emotion Space,“ u *Proceedings of the 18th Annual Conference of the International Speech Communication Association, INTERSPEECH 2017*, Stockholm, 2017.
- [121] T. Liogienė i G. Tamulevičius, „Comparative Study of Multi-Stage Classification Scheme for Recognition of Lithuanian Speech Emotions,“ u *2016 Federated Conference on Computer Science and Information Systems, FedCSIS*, Gdansk, 2016.
- [122] T. Liogienė i G. Tamulevičius, „Multi-Stage Recognition of Speech Emotion Using Sequential Forward Feature Selection,“ *Electrical, Control and Communication Engineering*, t. 10, br. 1, pp. 35-41, 2016.
- [123] C.-C. Lee, E. Mower, C. Busso, S. Lee i S. Narayanan, „Emotion Recognition Using a Hierarchical Binary Decision Tree Approach,“ *Speech Communication*, t. 53, br. 9-10, pp. 1162-1171, 2011.
- [124] Q. Mao i Y. Zhan, „A Novel Hierarchical Speech Emotion Recognition Method Based on Improved DDAGSVM,“ *Computer Science and Information Systems*, t. 7, br. 1, pp. 211-222, 2010.
- [125] M. Lugger, M.-E. Janoir i B. Yang, „Combining Classifiers with Diverse Feature Sets for Robust Speaker Independent Emotion Recognition,“ u *2009 17th European Signal Processing Conference*, Glasgow, 2009.
- [126] J. R. Parker, „Rank and Response Combination from Confusion Matrix Data,“ *Information Fusion*, t. 2, br. 2, pp. 113-120, 2001.
- [127] L. Xu, A. Krzyzak i C. Y. Suen, „Methods of Combining Multiple Classifiers and their Applications to Handwriting Recognition,“ *IEEE Transactions on Systems, Man, and Cybernetics*, t. 22, br. 3, pp. 418-435, 1992.

Biografski podaci o autoru disertacije

Željko B. Nedeljković je rođen 10.06.1988. godine u Užicu. Gimnaziju je završio sa odličnim uspehom. Elektrotehnički fakultet Univerziteta u Beogradu je upisao 2007. godine. Diplomirao je na modulu Signali i sistemi 2011. godine sa ukupnom prosečnom ocenom 9,5. Master studije na Elektrotehničkom fakultetu Univerziteta u Beogradu je upisao 2011. godine na modulu Signali i sistemi. Master studije je završio sa prosečnom ocenom 10,0. Doktorske studije na Elektrotehničkom fakultetu Univerziteta u Beogradu je upisao 2013. godine na smeru Upravljanje procesima i obrada signala. Položio je sve ispite sa prosečnom ocenom 10,0. U toku 2011. godine je proveo šest meseci na stručnoj praksi u firmi „Unit inženjering“ baveći se razvojem namenskih računarskih sistema. Od 2011. godine je radno angažovan u HE „Bajina Bašta“ na poslovima regulacije, upravljanja i SCADA sistema. Trenutno je na poziciji glavnog inženjera za sistem upravljanja i telekomunikacije.

Изјава о ауторству

Име и презиме аутора Жељко Недељковић

Број индекса 5003/2013

Изјављујем

да је докторска дисертација под насловом

Оптимизација система за аутоматско препознавање говорних емоција

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 01.09.2020.



**Изјава о истоветности штампане и електронске верзије докторског
рада**

Име и презиме аутора Жељко Недељковић

Број индекса 5003/2013

Студијски програм Електротехника и рачунарство

Наслов рада Оптимизација система за аутоматско препознавање говорних емоција

Ментор др Жељко Ђуровић, редовни професор

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, 01.09.2020.



Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Оптимизација система за аутоматско препознавање говорних емоција

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 01.09.2020.



1. **Ауторство.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.
2. **Ауторство – некомерцијално.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.
3. **Ауторство – некомерцијално – без прерада.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.
4. **Ауторство – некомерцијално – делити под истим условима.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.
5. **Ауторство – без прерада.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.
6. **Ауторство – делити под истим условима.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.