

Određivanje sličnosti audio numera aproksimacijom karakteristika audio signala



Stefan Isidorović
Matematički fakultet
Univerzitet u Beogradu

Master rad

Beograd 2018

Univerzitet u Beogradu, Matematički fakultet, Master rad

Autor: Stefan Isidorović

Naslov: Određivanje sličnosti audio numera
aproximacijom karakteristika audio signala

Mentor: prof. dr Vladimir Filipović,
Matematički fakultet, Beograd

Članovi komisije: dr Aleksandar Kartelj,
Matematički fakultet, Beograd,
dr Dragan Matić,
Prirodno-matematički fakultet, Banjaluka

Datum: x.x.2018.

Rezime

Ovaj rad pripada oblasti pretraživanja muzičkih informacija. Oblast pretraživanja muzičkih informacija se bavi procesiranjem i obradom audio signala kako bi se odredile karakteristike zvuka. Cilj ovog rada je da se razmotre i iskoriste različiti načini aproksimiranja karakteristika audio signala, prvenstveno radi kreiranja algoritma koji bi pametno birao pesme za reprodukciju. Značaj ovog rada, pored komercijalnih potencijala, leži u analizi samog audio signala i izvlačenju informacija iz njega. Mere sličnosti, koje će biti opisane u okviru ovog rada, mogu se iskoristiti za dalji razvoj i istraživanje oblasti gde postoje audio podaci. Aproksimacija sličnosti dve muzičke numere predstavlja samo jednu od mogućih primena opisanih odlika audio signala. Nakon upoznavanja sa problemom, u radu se razvijaju i primenjuju tehnike istraživanja i izvlačenja karakteristika audio signala. Vršiti se analiza dobijenih karakteristika. Razvijena metodologija je primenjena na konkretne slučajeve upotrebe dobijenih karakteristika audio signala.

Sadržaj

1	Uvod	1
1.1	Struktura rada	2
1.2	Sličnost i nalaženje muzičkih numera	2
1.3	Istraživanje podataka u muzici	3
1.4	Sličnost muzike	5
1.4.1	Uticaoj emocija na muzičku sličnost	6
1.4.2	Računski faktori muzičke sličnosti	7
1.4.3	Semantički raskorak	8
2	Osnovne metode obrade audio signala	10
2.1	Prozorske funkcije i okviri	11
2.1.1	Implementacioni detalji	13
2.2	Furijeova transformacija	15
2.3	Implementacioni detalji	16
2.3.1	Biblioteke	17
3	Atributi	19
3.1	Atributi niskog nivoa apstrakcije	19
3.1.1	Atributi iz vremenskog domena	20
3.1.1.1	Broj promena znaka amplitude	20
3.1.1.2	Autokorelacija prvog reda	21
3.1.1.3	Maksimalna vrednost amplitude	21
3.1.1.4	Srednjekvadratna energija	22
3.1.2	Atributi iz frekvencijskog domena	22
3.1.2.1	Spektralni centroid	22
3.1.2.2	Spektralni raspon	23
3.1.2.3	Glatkoća zvuka	24
3.1.2.4	Spektralna asimetrija	24
3.1.3	Implementacioni detalji	25

3.1.4	Atributi srednjeg nivoa	25
4	Mera sličnosti	28
4.1	Statistički parametri	29
4.2	Mera sličnosti	30
4.3	Test skup	31
4.3.1	Predstavljanje rezultata	32
4.4	Normalizacija statističkih ocena atributa	34
4.5	Ocena mere sličnosti	37
4.6	Implementacioni detalji	38
5	Zaključak	40
A	Prateće informacije	41

Poglavlje 1

Uvod

Muzika prati ljude kroz celokupnu njihovu istoriju. Većini ljudi muzika predstavlja više od čistog akustičnog fenomena. Danas je to kulturni fenomen koji je istrajao u ljudskoj istoriji. U životu većine ljudi muzika zauzima bitno mesto, bilo da su to producenti i ljudi koji učestvuju u samom stvaranju muzike, ili samo obični slušaoci. S obzirom na svoju kulturnu važnost, muzika je prvi tip medija koji je doživeo takozvanu digitalnu revoluciju. Zahvaljujući tehnološkom napretku u kompresiji i enkodiranju audio signala način konzumiranja muzike se znatno promenio tokom istorije. Zajedno sa razvojem Interneta i razvoju portabilnih uređaja za reprodukciju zvuka, digitalna muzika nije samo uzburkala svet informacionih tehnologija već je uzrokovala znatnu promenu u načinu kako (i koliko) ljudi slušaju muziku. Uprkos pomenutim promenama, način skladištenja i reprodukcije audio materijala je ostao, u velikoj meri, isti. I što je najbitnije, način na koji čovek pretražuje i dolazi do željene muzičke numere je ostao, takođe, isti. U nekoj meri je poboljšano pretraživanje filtriranjem rezultata na osnovu algoritama računarske inteligencije, ali ovaj pristup je u potpunosti zasnovan na akcijama korisnika sistema (Youtube, Deezer, iTunes i drugo) i ne vrše analizu prirodnih karakteristika zvuka. Ovakvim pristupom se donekle ograničava pretraživanje i vrši diskriminacija numera koje nisu toliko zastupljene u javnosti. Omogućen je pristup velikom broju muzičkih numera, preko raznih servisa, i dalje je odabir muzičkih numera, u velikoj meri, ostao manuelan, odnosno reprodukcione liste se generišu ručno od strane korisnika, što predstavlja veoma naporan i mukotrpan posao. Treba dodati da zbog prirode muzike, reprodukcione liste kreirane ručno, vremenom mogu postati neodgovarajuće, što iziskuje ponovno pravljenje istih. Ovo povećava potrebu za sistemima za pronalaženje pesama koji su svesni konteksta pesme, a koji takođe mogu biti personalizovani tako da mogu da pruže zadovoljavajuće rezultate. Aplikacije poput Shazam-a (identifikacija muzičke numere), Pandore (automatsko generisanje presonalizovane radio stanice), Spotify/Deezer(strimovanje muzike) ili Last.fm (sistem za preporuku pesama i društvena mreža) se smatraju neophodnim servisima za veliki broj korisnika danas. Sve veći broj korisnika pomenutih servisa dovodi do povećanja značaja razvoja pomenutih sistema, čime se postiže mnogo veći uticaj na živote ogromnog broja ljudi.

1.1 Struktura rada

U ovom radu fokus će biti stavljen na osnovne probleme i tehnike za aproksimaciju sličnosti audio numera baziranih na prirodnim karakteristikama zvuka. Proći će se svi aspekti koji se odnose na dizajniranje takvog sistema, počevši od prirode problema, ograničenja i mogućnosti. Prvo će biti opisana oblast kojoj ovaj problem pripada, njen istorijat, razvoj i trenutno stanje. Biće formulisana postavka problema odnosno istraživački i korisnički zahtevi sistema, kao i cilj kome se teži. Ukratko će biti opisane tehnologije koje se koriste za izradu rezultujućeg sistema, izabranih biblioteka i slično, kao i kratak pregled dizajna samog sistema, čime će uvod biti zaokružen. Nakon toga biće navedeni najvažniji aspekti teorije obrade signala, koji se odnose na audio signal. Biće spomenuti osnovni problemi na koje se nailazi, kao i par osnovnih načina za rešavanje tih problema. Nakon toga će biti formulisana pojam sličnosti dve audio numere. Biće definisana osnovna terminologija vezana za prirodne karakteristike zvuka, kao i njihova povezanost sa odabranim skupom atributa. Svaki odabrani atribut biće ponaosob posmatran i objašnjen, kao razlozi za njegov odabir. U ovom radu fokus će biti na primitivnim atributima, odnosno atributima koji pružaju dovoljno dobru aproksimaciju tražene osobine, na uštrb složenosti izračunavanja. Zatim dolazi diskusija o načinu izračunavanja same sličnosti, da li će biti korišćeno izračunavanje mere sličnosti na osnovu redosleda vrednosti ili ne. Biće uveden i bazični sistem za personalizaciju dobijenih predloga, na osnovu kalibracije korisnika pa i kasnijih akcija samog korisnika. Na samom kraju biće prikazani rezultati testiranja, kao i njihova analiza. Takođe, razmatraće se i mogućnosti daljih unapređenja samog sistema.

Kroz svaki pređeni aspekt biće paralelno obrađena odgovarajuća komponenta sistema, njena implementacija i motivacija za svako upotrebljeno rešenje u nadi da će ovakav pristup omogućiti lakše razumevanje, kasniju modifikaciju i istraživanje baziranog na pruženom sistemu i informacijama.

1.2 Sličnost i nalaženje muzičkih numera

Tradicionalno, pretraživanje muzike sačuvane u digitalnom zapisu, bilo da je kolekcije od par stotina ili hiljada numera (privatne kolekcije) ili par miliona pesama (veći striming ili komercijalni servisi) je pretraživanje baze podataka na osnovu meta podataka. Meta podaci, u ovom slučaju, predstavljaju skup podataka kojima se opisuje ili identifikuje pesma. Primer takvih podataka su ime pesme, ime izvođača, godina izdavanja, žanr i stil. Indeksiranje u malopre pomenutim kolekcijama se uglavnom vrši preko žanra ili stila, u kombinaciji sa podacima kao što su izvođač, album i slično. Ovakav pristup je donekle prihvatljiv kod malih kolekcija pesama (do par stotina pesama), gde korisnik u značajnoj meri poznaje sadržaj kolekcije, međutim kod većih kolekcija ovakav pristup nije dovoljan. Kada većina pesama nije prethodno poznata korisniku, veliki broj pesama se zanemaruje i često

omete nalaženje željene pesme. Problem nastaje u veoma malom skupu dostupnih podataka, pogotovu ako taj skup nije kompletan, što je često slučaj kod privatnih kolekcija. Ispravnost i postojanje meta podataka zavisi od dosta faktora, da li je uz samu numeru isporučen i odgovarajući skup meta podataka, da li format u kom se nalazi podržava svaki podatak, da li su softveri za kompresiju zadržali meta podatke i slično. Iako je muzičkom produkcijom došlo do pojave ogromnog broja muzičkih numera, zbog nedovoljno razvijenih servisa za distribuciju i pretraživanje, značajan broj njih ostaju neprikazane krajnjim korisnicima. Ovim se dolazi do zaključka da je zasnivanje pretrage na meta podacima podložno greškama i u većini slučajeva nedovoljno. Veliki prostor pretrage omogućuje da se iz aspekta pretraživanja klasičnih simboličkih podataka (meta podaci) pređe u pretraživanje sub-simboličkih ili prirodnih karakteristika zvuka - na primer, više u fokusu neće biti samo izvođač pesme, već ritam pesme ili tempo ili dominacija visokih tonova. Sistemi za pronalaženje koji ne obraćaju pažnju na prirodne karakteristike zvuka odnosno pesme su mnogo drugačiji i inferiorniji od načina kako ljudi organizuju i koriste muzičke kolekcije - ili u terminima istraživanja podataka, takvi sistemi ne obraćaju pažnju za "muzičke informacije" koje su potrebne njihovim korisnicima. Danas se teži ka sistemima za pronalaženje koji će obraćati pažnju na prirodne karakteristike i tako poboljšati svoju efikasnost. Ovaj aspekt je izuzetno bitan kada se u obzir uzme da je muzika veoma isprepletana sa pop kulturom i da u većini slučajeva, prosečan korisnik koji pristupa jednom takvom sistemu, pristupa primarno radi čiste zabave i hedonističkih i rekreativnih potreba, pa je nalaženje optimalnog rezultata izuzetno bitna karakteristika, koja često dominantno odlučuje o tome da li će sistem opstati [20].

1.3 Istraživanje podataka u muzici

Istraživanje podataka u muzici, u daljem tekstu MIR (eng. *MIR - Music Information Retrieval*) je disciplina računarskih nauka koja se bavi istraživanjem zvuka (odnosno muzike) kao podatka. Kao odgovor na izazove i specifične potrebe nalaženja podataka u muzičkom domenu, oblast istraživanja podataka poznatija kao istraživanje podataka u muzici je počela svoju evoluciju tokom devedesetih godina XX veka, a postala je priznata kao oblast serijom konferencija organizacije ISMIR [2]¹. Između ostalog MIR se bavi istraživanjem i razvojem inteligentnih metoda koje imaju za cilj da odrede muzički značajne attribute, direktno iz audio signala ili drugih izvora. Ovi atributi se kasnije mogu upotrebiti kao poboljšani interfejsi prema muzičkim kolekcijama. U literaturi se pominju definicije poput:

"MIR je multidisciplinarna oblast istraživanja koja obuhvata računarske nauke istraživanje podataka, muzikologiju, muzičku teoriju, audio inženjering, obradu digitalnog signala, kognitivne

¹ISMIR - International Symposium for Music Information Retrieval, od 2009. godine poznata kao International Society for Music Information Retrieval

nauke, bibliotečke veštine, izdavanje i pravna pitanja. Za cilj ima razvoj različitih načina za upravljanje kolekcijama muzičkog materijala u svrhe čuvanja, pristupa, istraživanja i drugih upotreba [7] ”

Definiciju su formulisali Futrelle i Downie [9] koji naglašavaju da je oblast multidisciplinarna i vuče korene iz istraživanja u okviru digitalnih biblioteka. Kasnije Downie više favorizuje analizu sadržaja, odnosno automatsko izdvajanje muzičkih atributa iz audio signala [6]. Na kraju se iskristalisala sledeća defnacija: ”MIR se bavi sa ekstrakcijom, analizom i upotrebom informacija o bilo kakvoj vrsti muzičkog entiteta na bilo kom nivou njegove reprezentacije.”

Pod muzičkim entitetom možemo podrazumevati i druge stvari pored same pesme, kao na primer konkretnog izvođača ili skup pesama (npr. žanr), dok pod nivoom reprezentacije se misli na sveobuhvatne informacije u muzičkoj numeru gde spadaju i simbolički i sub-simbolički podaci. Pored navedenih defnicija u literaturi se mogu naći još mnoštvo drugih defnicija same oblasti. Različitost ovih različitih defnicija demonstrira širinu koju ova oblast poseduje.

Neki od problema kojima se MIR bavi a tiču se samo automatske analize kolekcija muzičkih numera (katalogizacije, arhiviranje i sličnog što je je fokus ovog rada) su:

- Prepoznavanje izvođača
- Audio otisak prsta
- Prepoznavanje akorda
- Detektovanje muškog/ženkog vokala
- Klasifikacija žanra
- Identifikovanje instrumenta
- Identifikovanje raspoloženja
- Sličnost ritma
- Preporučivanje muzičkih numera
- Aproksimacija sličnosti pesama
- Estimacija ritma
- Transkripcija
- ...

U okviru ovog rada fokus će biti na aspektima obrade signala, preporuke muzičkih numera i aproksimacije sličnosti audio numera. Ostali aspekti će biti obrađeni više ili manje u zavisnosti od potrebe.

Slično kao i kod ostalih domena ovde se identifikuju tri glavne paradigme što se tiče pristupa informacijama o muzici. Među autorima iz ove oblasti najšire se prihvata sledeća klasifikacija:

Pronalaženje Korisnik ima određene informacije o pesmi koju želi da nađe i ove informacije aktivno koristi u upitu. Upit može biti tradicionalni tekstualni ili simbolička reprezentacija dela pesme ili audio snimak dela pesma.

Pretraga Korisnik ima okvirne informacije o pesmi/pesmama koje traži i želi da istraži dostupnu kolekciju. Pretraga je interaktivni i iterativni proces koji se bazira na intuitivnim i efektivnim korisničkim interfejsima.

Preporuka Sistem filtrira kolekciju za potencijalno relevantnim unosima. Relevantnost se određuje na osnovu akcija i preferencijama korisnika. Ove preferencije su eksplicitno date ili zaključene na osnovu ponašanja korisnika. U ovom slučaju od korisnika se ne zahteva da eksplicitno pretražuje sistem.

Primena istraživanja iz ove oblasti je prilično velika. Od sistema za identifikaciju pesme koja se trenutno reprodukuje (Shazam, SoundHound i drugi), nalaženje pesama na osnovu zvuka koji pruža sam korisnik do detekcije instrumenata ili izdvajanja izvora zvuka i mnogih drugih tehnika [17]. Fokus će biti na sličnosti dve audio numere, kao glavne teme ovog rada.

1.4 Sličnost muzike

Pojam sličnosti muzike je centralni koncept u većini sistema za pronalaženje muzičkih numera, posebno za sisteme koji koriste nekakav upit za pretragu. Da bi se izvršila pretraga, potrebno je definisati relaciju između dve muzičke numere, odnosno način na koji se odlučuje koja muzička numera, od dve ponuđene, ima prednost. U okviru sistema za identifikaciju puštene muzičke numere, dovoljna je relacija jednakosti. Kod generisanja reproduccionih lista, relacija jednakosti ne može biti dovoljna. Potrebno je definisati relaciju poretka. Relacija poretka može biti realizovana i računanjem distance između dve muzičke numere, odnosno njihove sličnosti, ovim se zadovoljava i uslov relacije jednakosti (u slučaju kada distanca teži 0). Zbog same prirode muzike vrlo je nejasno šta pojam sličnosti predstavlja. Intuitivno korisnik muziku može opisati korišćenjem nekih od sledećih koncepata [9]

- Žanr (npr. zvuči kao pop muzika)
- Instrumentacija (npr. učtvuju samo harmonika i bubnjevi)
- Karakteristike vokala (npr. pevač zvuči skoro kao žena)
- Melodija (npr. pesma ide nešto kao "pam pam param pam tšš")

- Tempo (npr. brz)
- Ritam (npr. dvojka)

Međutim ako bi trebalo da se opišu i kvantifikuju svi muzički aspekti i njihovo menjanje tokom vremena, ovakav pristup baziran na signalu, ne bi uspeo da obuhvati širok fenomen spoznate muzičke sličnosti [27].

1.4.1 Uticaj emocija na muzičku sličnost

Pitanje i potraga za muzičkom sličnošću može ići i preko akustičnih svojstava muzičke numere. Primer gde korisnik ne želi da sluša muzičku numeru iz razloga koji su strogo subjektivni, u slučaju gde korisnik ne podržava stavove izvođača ili zbog privrženosti drugom žanru, čak i u slučaju da akustične osobine pesme odgovaraju zvuku koji inače sluša. U ovom slučaju prirodna sličnost muzičkih numera nema nikakvog značaja. Kao drugi primer se navodi situacija gde određena osoba tokom vožnje preferira isključivo pesme sa ženskim vokalima i brzim tempom, ali dok odmara više preferira pesme sa slabijim ritmom, nezavisno od vokala. Na pomenutim primerima se vidi da situacija u kojoj korisnik odlučuje da koristi sistem ima ogroman uticaj na to kako se manifestuje sličnost muzike i preferencije korisnika za sledeću pesmu. Ove lične preferencije i shvatanje muzike su vezani za emocije koje prenosi muzička numera. Pojedini autori smatraju da se konkretno shvatanje muzike (sa emocionalne strane) može povezati sa različitim vrstama kodiranja, odnosno specifičnih načina na koji muzika prenosi emociju. Džoslin [10] razlikuje tri vrste kodiranja:

- **Ikonično kodiranje** (eng. *Iconic coding*) koje povezuje formalnu sličnost između signala, odnosno mapira osnovne emocije kroz akustične osobine poput tempa ili vokalnih izraza.
- **Unutrašnje kodiranje** (eng. *Intrinsic coding*) se odnosi na "unutrašnje, sintatičke veze u okviru same muzike" to jest kompleksnije, strukturalnije osobine.
- **Asocijativno kodiranje** (eng. *Associative coding*) se odnosi na konotacije između muzike i drugih proizvoljnih događaja ili objekata, na primer na romantičnu poruku neke balade.

Iz navedenog se vidi da asocijativno kodiranje u velikoj meri može da utiče na razumevanje i doživljaj pesme. Aspekti od kojih zavisi asocijativno kodiranje se ne mogu izvesti iz samog audio signala i individualni su. Pošto se sve navedene osobine ne mogu tačno odrediti, potreban je način da se iz sadržaja muzičke numere, računskim pristupom, aproksimiraju i izvuku informacije koje su nam dostupne [9].

1.4.2 Računski faktori muzičke sličnosti

Umesto da računski bude modelovan proces razumevanja muzike individualnog korisnika, biće razvrstane sve dostupne informacije povezane sa slušanjem muzike, u cilju postavljanja matematički definisanih mera za određivanje sličnosti dve muzičke numere. Među postojećim klasifikacijama u ovom radu će detaljno biti opisana klasifikacija koju su formulisali Knees i Schedl [20]. Mogu se odrediti četiri kategorije koje utiču na opažanje sličnosti muzike:



Slika 1.1: Četiri različite kategorije koje utiču na shvatanje muzike

- **Muzički sadržaj** se odnosi na sve što je sadržano i može se izdvojiti iz samog signala muzičke numere. Ovde se prvenstveno misli na ritam, boju zvuka, melodiju, harmoniju i slično. Ovi faktori odgovaraju ikoničnom i na višem nivou unutrašnjem kodiranju.
- **Muzički kontekst** su informacije koje se ne nalaze i ne mogu se izračunati na osnovu audio signala muzičke numere. Ovo su podaci koji spadaju pod kategoriju meta-podataka prošireno sa dodatnim informacijama. To su marketing, zemlja porekla, način oblačenja izvođača i slično. Karakterišu kao društveni meta-podaci ili kulturološke osobine. Ovi faktori odgovaraju asocijativnom kodiranju.
- **Osobine korisnika** su lične odlike koje karakterišu korisnika, kao što su preferencije, ukus kao i muzičko znanje i iskustvo. Ovi faktori se odnose na unutrašnje i asocijativno kodiranje.
- **Kontekst korisnika** odnosi se na trenutnu situaciju u kojoj korisnik sluša muziku. Uticaj na kontekst korisnika imaju lokacija, vreme, aktivnost ili raspoloženje. Ovi faktori utiču na sve tipove kodiranja.

U cilju stvaranja univerzalne i sveobuhvatne mere sličnosti, u teoriji, trebalo bi da budu izračunati svi faktori, koji bi zatim bili kombinovani na odgovarajući način, što je netrivialan zadatak. Ono što se zapaža je da faktori nisu disjunktni, odnosno postoji prilično velika sprega među njima, što direktno povlači tvrdnju da nisu svi potrebni da bi dovoljno dobro aproksimirali sličnost.

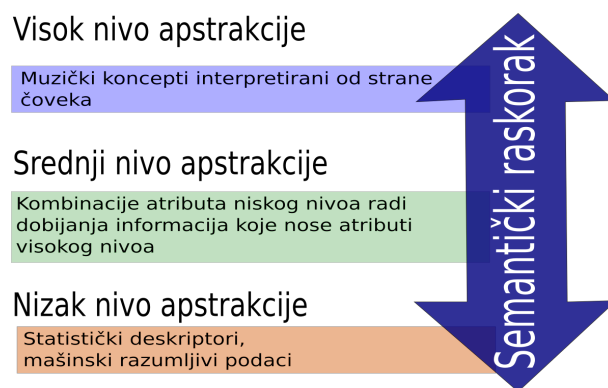
Da bi se aproksimirali prethodno opisani faktori potrebno je identifikovati signale koje nose upotrebljivu informaciju o jednom ili više faktora.

- Za muzički sadržaj, signal koji se traži je digitalna reprezentacije muzičke numere.
- Muzički kontekst se može odrediti iz više izvora, iz dostupnih meta-podataka ili pretragom informacija sa drugih izvora takvih informacija (npr. Internet).
- Osobine korisnika se mogu aproksimirati putem tehnika preuzetih iz psihologije, poput testova i upitnika. Upitnici su veoma korisni za određivanje preferencija ili muzičkog poznavanja korisnika.
- Kontekst korisnika se može odrediti pomoću raznih senzora na samom uređaju, međutim ovakve informacije su podložne greškama usled veoma velikog šuma koji se stvara. Na primer, koristeći pedometar možemo odrediti da li korisnik trči u datom trenutku, međutim ukoliko korisnik se vozi autobusom, vibracije prouzrokovane pomeranjem ruke ili vožnje po neravnom terenu, mogu uzrokovati da pedometar da vrednosti slične kao pri trčanju. Postojanje šuma koji se pojavljuje u određenim situacijama kod ovakvih senzora dovodi do pogrešnog tumačenja rezultata.

U drugom koraku, trebalo bi izvući neki smisao iz dobijenih podataka, odnosno izdvojiti skup atributa. Ovo je ključni korak i ima najveći uticaj na finalni proizvod, odnosno na rezultate koje će sistem proizvesti. Na osnovu dostupnih tipova podataka mogu se odrediti atributi višeg ili nižeg nivoa apstrakcije. Na primer, iz muzičkog sadržaja se mogu dobiti atributi niskog nivoa apstrakcije, koji se relativno teško interpretiraju od strane čoveka, dok tokom analiziranja tekstova sa interneta se mogu izvući atributi višeg nivoa apstrakcije koje čovek može veoma lako da razume i interpretira. Atributi niskog nivoa apstrakcije predstavljaju numeričke vrednosti dobijene pomoću formule koja se koristi za izračunavanje datog atributa. Zbog fundamentalne razlike između atributa niskog nivoa i atributa visokog nivoa koje čovek koristi za opis pesme, ne postoji očigledna korelacija pa zbog toga dobijene numeričke vrednosti ne mogu biti efikasno interpretirane od strane čoveka.

1.4.3 Semantički raskorak

Audio zapis je predstavljen kao niz numeričkih vrednosti, čovek koristi opisne attribute da bi predstavio muzičku numeru. Niz numeričkih vrednosti upotrebljen za čuvanje i zapis u memoriji, kao i vrednosti dobijene njihovom obradom nazivaju se atributi niskog nivoa. Opisni atributi koje korisnik koristi se nazivaju atributima visokog nivoa. Nesaglasnost atributa niskog nivoa i visokog nivoa apstrakcije naziva se "semantički raskorak" [4]. Semantički raskorak je prisutan u svim medijskim domenima [13] i dok ne opravdava u potpunosti opis "semantički" u praksi može biti definisan kao



Slika 1.2: Različiti nivoi apstrakcije atributa i semantički raskorak između njih

krovni termin za ograničenja trenutnih atributa inteligentnih sistema u kontekstu nivoa satisfakcije korisnika. Iako je semantički raskorak svuda pristuan, zbog veoma subjektivnog shvatanja od strane korisnika, najviše je prisutan u domenu muzike [21] [20]. Ublažavanje problema semantičkog raskoraka se rešava pomoću uvođenja atributa srednjeg nivoa apstrakcije. Atributi srednjeg nivoa apstrakcije nose informacije bliže čoveku, a dobijaju se kombinacijom atributa niskog nivoa. Ovakav pristup nije karakterističan samo za oblast muzike, već i za ostale tipove podataka. Atributi srednjeg nivoa apstrakcije, takođe, pružaju mnogo relevantnije informacije za pristupe mašinskog učenja.

Poglavlje 2

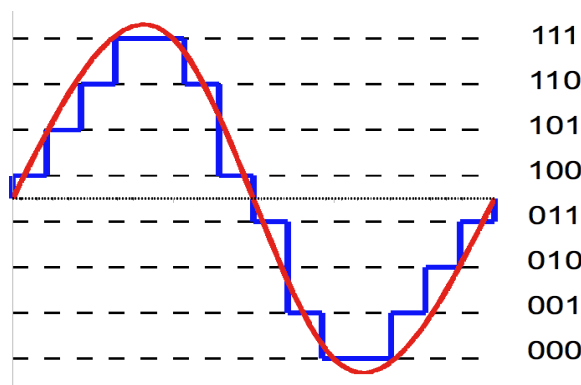
Osnovne metode obrade audio signala

Aproksimacija sličnosti dve muzičke numere na osnovu prirodnih karakteristika zvuka se primarno odnosi na izvlačenje skupa atributa iz muzičkog sadržaja. Muzički sadržaj se reprezentuje datotekama koje sadrže digitalni zapis zvuka. Izvođenje atributa je glavni i centralni proces u okviru ovog rada. Proces izvođenja atributa, kojim se bave komponente poznatije kao "ekstraktori atributa", ima za cilj da na osnovu audio signala formira skup atributa. Pre svega toga, sirov audio signal mora da prođe kroz niz operacija i obrada da bi se dobili podaci optimizovani i spremni za nalaženje atributa. U okviru ovog poglavlja će biti razmatrani osnovni aspekti i elementarne tehnike obrade audio signala. Pored teorijske osnove, biće i reči o implementacionim detaljima u okviru testnog sistema koji će biti razvijen radi demonstracije prikazanih stvari u ovom radu.

Zvuk se definiše kao mehanički talas frekvencija između 16 i 20000 Herca, odnosno zvuk koji čovek može da čuje. Postoje i infrazvuk i ultrazvuk, odnosno zvuk ispod granice od 16Hz i iznad granice od 20KHz respektivno. Akustični muzički instrumenti podstiču vazduh da oscilira na različitim frekvencijama čime proizvode zvuk. Zvuk proizveden od strane akustičnih muzičkih instrumenata je analogne prirode i može se predstaviti netrivialnom matematičkom funkcijom, koja za određen trenutak u vremenu rezultuje frekvencijskom amplitudom. Uređaji za snimanje imaju ograničenu vremensku rezoluciju, te stoga ne mogu u snimiti kontinualni signal. Problem se prevazi-
lazi uzimanjem uzoraka signala na određenom vremenskom intervalu. Uzimanje uzoraka je proces merenja amplitude frekvencije u datom trenutku. Merenje se vrši pomoću specijalizovane membrane koja detektuje frekvenciju oscilacije (mikrofona¹). Nakon merenja frekvencije, dobijena vrednost se normalizuje na unapred zadat interval. U ovom slučaju normalizacija podrazumeva dva koraka, prvi korak je ignorisanje vrednosti van intervala zvuka [16Hz,20000Hz], odnosno njihovo zaokruživanje na najbližu granicu intervala. Drugi korak predstavlja preslikavanje vrednosti iz intervala zvuka u interval zadat prilikom digitalizacije (npr. [-1,1]). Opisan proces uzimanja uzoraka se zove i kvanti-

¹Uređaji kojim se vrši ovo merenje su uglavnom mikrofoni. Princip po kome funkcioniše mikroskop je veoma sličan principu reprodukcije zvuka preko zvučnika, zapravo mikroskop predstavlja mali zvučnik i sposoban je da reprodukuje zvuk dosta manje jačine nego zvučnik. Posledica toga je manja membrana, zbog osetljivosti na spoljašnje vazdušne vibracije.

zacija signala. Proces kvantizacije stvara i grešku kvantizacije. Greška kvantizacije nastaje kada se vrednost kontinualnog signala i mapirana diskretna vrednost razlikuju.



Slika 2.1: Greška kvantizacije. Analogni signal je predstavljen crvenom bojom, dok je plavom predstavljen odgovarajući digitalni signal.

Na slici se lako uočava greška kvantizacije, razlika u vrednostima između plave i crvene linije. Iz čega zaključujemo da je rezolucija, ili diskretni interval iz kog uzimamo vrednosti prilikom analogno-digitalne konverzije, važan za kvalitet rezultata. Često se rezolucija poistovećuje sa memorijom dostupnom za zapis vrednosti, odnosno brojem bitova u koji možemo da zapišemo jedan uzorak. Kod analogno-digitalne konverzije pored rezolucije, frekvencija uzimanja uzoraka predstavlja još jedan izuzetno bitan faktor. Nyquist-Shannon-ova teorema kaže da frekvencija uzimanja uzoraka signala mora da bude najmanje $2 \cdot f$ da bi se zapisale frekvencije do f Herca [23]. Ako je frekvencija suviše mala postaje nemoguće zabeležiti pun spektar frekvencija koje ljudsko uho može da prepozna.

2.1 Prozorske funkcije i okviri

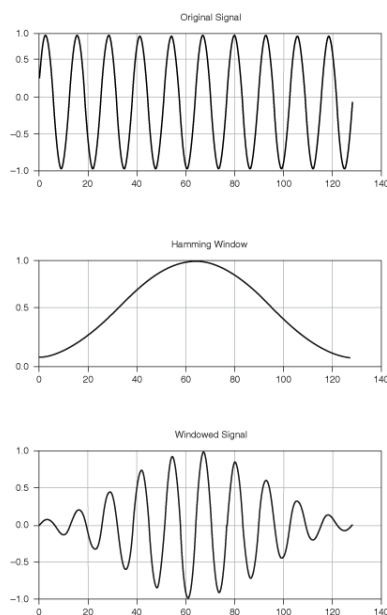
Posmatrajući digitalni signal kao niz uzoraka, svaki uzorak ponaosob predstavlja izuzetno mali deo vremena. Kod najrasprostranjenije frekvencije uzimanja uzoraka od 44KHz, uzorak pokriva oko 23 mikrosekunde. Ovaj interval je suviše kratak da bi ga ljudsko uho registrovalo kao zvuk. Ljudsko uho može da prepozna dužinu zvuka od najmanje 10 milisekundi. Izračunavanje atributa se vrši na skupu uzoraka, koji kad se reprodukuje proizvodi dovoljno dug zvuk da ga ljudsko uho registruje. Izračunavanje atributa na celoj muzičkoj numeri u celosti, iako moguća tehnika, neće pružiti mnogo informacija. Osnovni razlog za ovo je što se na taj način ne dobija nikakva informacija o raspodeli atributa u toku trajanja pesme, već samo vrednost atributa za celu pesmu. Stoga se vrši izračunavanje atributa nad manjim skupovima uzoraka, koji se nazivaju okviri (eng. *frame*). Dužina okvira se definiše kao stepen dvojke u intervalu 2^8 (256) - 2^{13} (8192). Moguće je uzimanje dužine i van pomenutog intervala, ali u literaturi i današnjim sistemima je to vrlo retko. Dužina okvira mora zadovoljiti uslov da je jednaka nekom stepenu dvojke zbog kasnije primene brze Furijeove transformacije. Iako stepen dvojke nije obavezan uslov za Furijeovu transformaciju, dosta doprinosi

performansama izračunavanja Furijeove transformacije i često je obavezan uslov za korišćenje već implementiranih transformacija u okviru dostupnih biblioteka.

Okviri predstavljaju ulaz za komponente koje vrše izračunavanje atributa koji rade sa signalom u vremenskom domenu. Za attribute iz frekvencijskog domena, primenjuje se Furijeova transformacija na okvir pre izračunavanja. Pre primene Furijeove transformacije nad okvirom obavezno je uraditi neki vid normalizacije nad okvirima. Normalizacija se vrši primenom unapred određene prozorske funkcije nad okvirom. Uobičajen izbor je takozvana Hann-ova prozorska funkcija, nazvana po austrijskom meteorologičaru Julius-u von Hann-u. Prozorska funkcija je definisana u jednačini 2.1, gde je N dužina okvira i $n = 1 \dots N$.

$$w(n) = 0.5 \cdot \left(1 - \cos \left(\frac{2 \cdot \pi \cdot n}{N - 1} \right) \right) \quad (2.1)$$

Množenjem svakog uzorka u okviru sa odgovarajućom vrednošću Hannove funkcije dobijamo periodični signal. Ovo je izuzetno bitno zato što Furijeova transformacija stvara artefakte u rezultujućem spektrumu za funkcije koje nisu periodične. Ovaj problem se naziva *spektralno curenje*. Proces normalizacije okvira sa prozorskom funkcijom je prikazan i grafički na slici 2.2.



Slika 2.2: Prikaz primene Hannove prozorske funkcije na signal

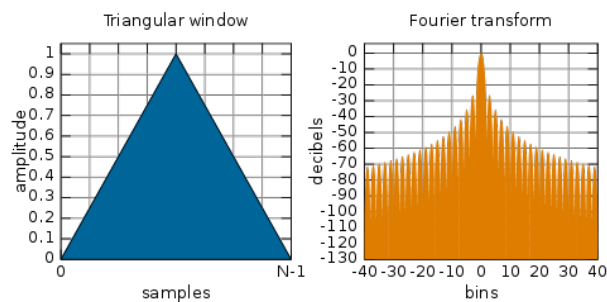
Gornji dijagram prikazuje ulazni signal, odnosno okvir od 128 uzoraka. Srednji dijagram prikazuje izgled Hannove funkcije nad istim brojem uzoraka. Najniži dijagram prikazuje izgled izlaznog signala u okviru posle primene normalizacije prozorskom funkcijom. Kao što se vidi na slici, primena normalizacije prozorskom funkcijom potiskuje uzorke sa oba kraja okvira. Da bi se izbeglo gubljenje informacija zbog pomenutog potiskivanja, obično se uzastopni okviri preklapaju. Dužina preklapanje se naziva i hop dužina (eng. *hop size*). Ilustrovano na primeru, uzećemo dužinu okvira 256 i

hop dužinu od 128 uzoraka, prvi okvir uzima vrednosti uzoraka na pozicijama 1-256, drugi okvir uzima uzorke na pozicijama 129-385, treći okvir uzorke na pozicijama 257-513 i sve tako redom. Na prikazanom primeru uzastopni uzorci se preklapaju u proporciji od 50% što je čest izbor u procesiranju muzičkog signala [20].

2.1.1 Implementacioni detalji

U okviru testnog sistema koji će biti razvijen, pružene su opcije biranja prozorske funkcije za normalizaciju. Kako je sistem predviđen da demonstrira navedene koncepte u ovom radu, pored toga dizajniran je sa namerom da čitaocu pruži mogućnost daljeg eksperimentisanja sa pomenutim konceptima. Korisniku je omogućeno da učitava sopstvenu prozorsku funkciju za normalizaciju kao i sledeće podrazumevane prozorske funkcije, odakle korisnik može da izabere koju želi da primeni. Svi koncepti se mogu konfigurisati u okviru `config.json` datoteke. Odabir prozorske funkcije za normalizaciju se vrši setovanjem parametra `NormFunc` kome se prosleđuje ili putanja do datoteke ili ime neke od već implementiranih prozorskih funkcija za normalizaciju. Već implementirane prozorske funkcije za normalizaciju, pored Hann-ove su:

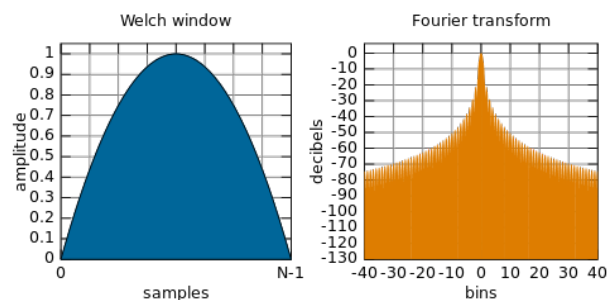
- Bartletova prozorska funkcija



Slika 2.3: Bartletova ili trougaona prozorska funkcija [28]

$$w(n) = 1 - \left| \frac{n - \frac{N-1}{2}}{\frac{N}{2}} \right| \quad (2.2)$$

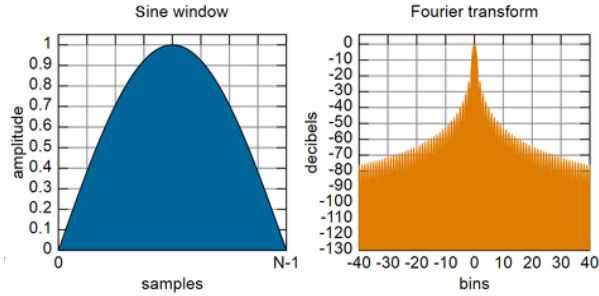
- Welch-ova prozorska funkcija



Slika 2.4: Welchova prozorska funkcija [28]

$$w(n) = 1 - \left(\frac{n - \frac{N-1}{2}}{\frac{N-1}{2}} \right)^2 \quad (2.3)$$

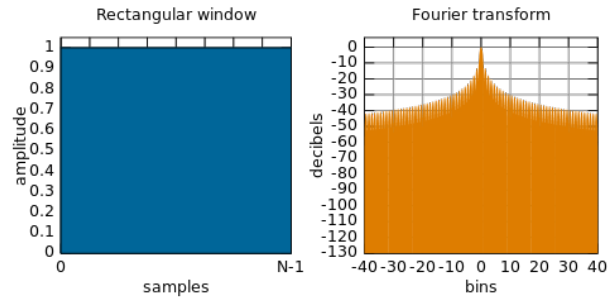
- Sinusna prozorska funkcija



Slika 2.5: sinusna prozorska funkcija [28]

$$w(n) = \sin \left(\frac{\pi \cdot n}{N - 1} \right) \quad (2.4)$$

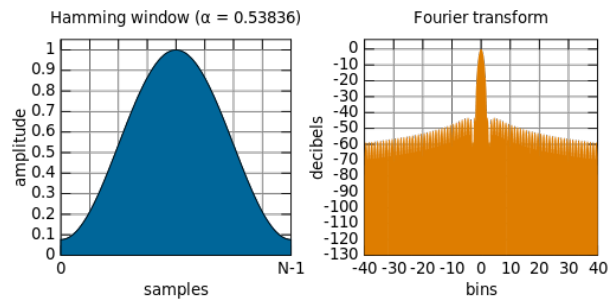
- Pravougaona prozorska funkcija



Slika 2.6: Pravougaona prozorska funkcija [28]

$$w(n) = 1 \quad (2.5)$$

- Hammingova prozorska funkcija



Slika 2.7: Hammingova prozorska funkcija [28]

$$w(n) = \alpha - \beta \cdot \cos \left(\frac{2 \cdot \pi \cdot n}{N - 1} \right) \quad (2.6)$$

sa $\alpha = 0.54$, $\beta = 1 - \alpha = 0.46$

Uz svaku prozorsku funkciju je data formula koja je određuje, kao i slika koja predstavlja prozorsku funkciju i rezultujuć izlaz Furijeove transformacije za konstantan ulazni signal.

2.2 Furijeova transformacija

Furijeova transformacija je nazvana prema francuskom matematičaru i fizičaru Jean Baptiste Joseph Fourier. Predstavlja transformaciju signala iz vremenskog u frekvencijski domen. Zasniva se na Furijeovoj teoremi koja kaže: "Svaka kontinualna i periodična funkcija može biti predstavljena sumom sinusnih i kosinusnih talasa koji osciliraju na različitim frekvencijama." Furijeova transformacija rezultuje diskretnim skupom kompleksnih vrednosti, koje opisuju frekvencijski spektar ulaznog signala. Pošto Furijeova transformacija predstavlja fundamentalni alat u obradi signala, razvijeno je nekoliko različitih pristupa i algoritama za njeno izračunavanje, kratko će biti opisani pristupi koji se tiču obrade audio signala. Računski najmanje zahtevna varijanta Furijeove transformacije je tzv. brza Furijeova transformacija (eng. *Fast Fourier Transformation - FFT*, koja može biti korišćena ako je ulazni signal diskretan, što jeste slučaj kod digitalnih audio signala. Stoga FFT je diskretna Furijeova transformacija (DFT) definisana jednačinom 2.7, gde je k celobrojna vrednost na intervalu od 0 do $N - 1$, N je dužina okvira odnosno broj uzoraka u okviru i x_n je n^{ti} uzorak iz okvira, odnosno odgovarajuća frekvencijska amplituda.

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{2\pi \cdot i}{N} \cdot n \cdot k} \quad (2.7)$$

Svaka izlazna vrednost X_k diskretne furijeove transformacije je kompleksan broj, čiji realan deo $Re(X_m)$ predstavlja značaj kosinusnih talasa i imaginarni deo $Im(X_m)$ predstavlja značaj sinusnih talasa. Odakle možemo da dobijemo informacije o vrednosti i fazi date frekvencije.

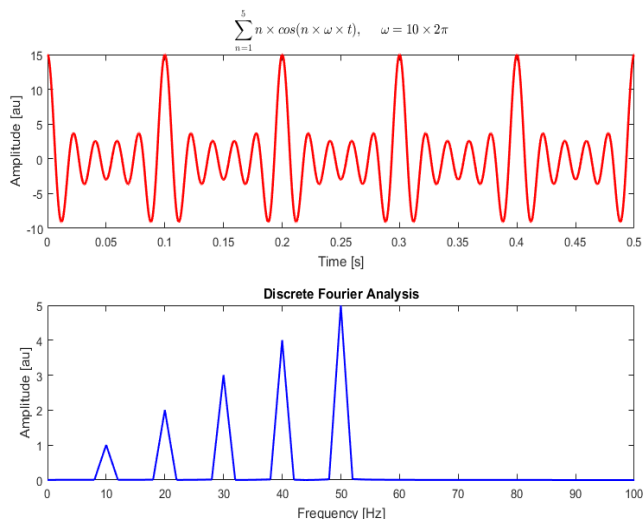
$$|X_n| = \sqrt{Re(X_n)^2 + Im(X_n)^2} \quad (2.8)$$

Jednačina 2.8 nam daje način za određivanje amplitude kombinovanih sinusnih i kosinusnih talasa.

$$\phi(X_n) = \text{tg}^{-1} \frac{Im(X_n)}{Re(X_n)} \quad (2.9)$$

Jednačinom 2.9 se definiše faza tj. relativna proporcija sinusa i kosinusa. Izračunavanjem X_n za različite frekvencije iz vremenskog domena ulaznog signala x_n direktno iz jednačine 2.7, dobija se vremenska složenost $O(n^2)$, odnosno za svaki od N izlaza sumira se N elemenata. Da bi se značajno smanjila vremenska složenost, Cooley i Tukey 1965. godine predlažu algoritam koji se izvršava u $O(N \cdot \log N)$ složenosti [5]. Kao početni zahtev je postavljen uslov da ulazni signal pripada diskretnom skupu kao i to da njihov algoritam zahteva da je N , odnosno dužina okvira,

stepen dvojke. Pomenuti uslovi, obično nisu ograničavajući faktori u svakodnevnoj upotrebi pa je ova verzija algoritma (odnosno njene sitne modifikacije) danas i najrasprostranjenija u industriji.

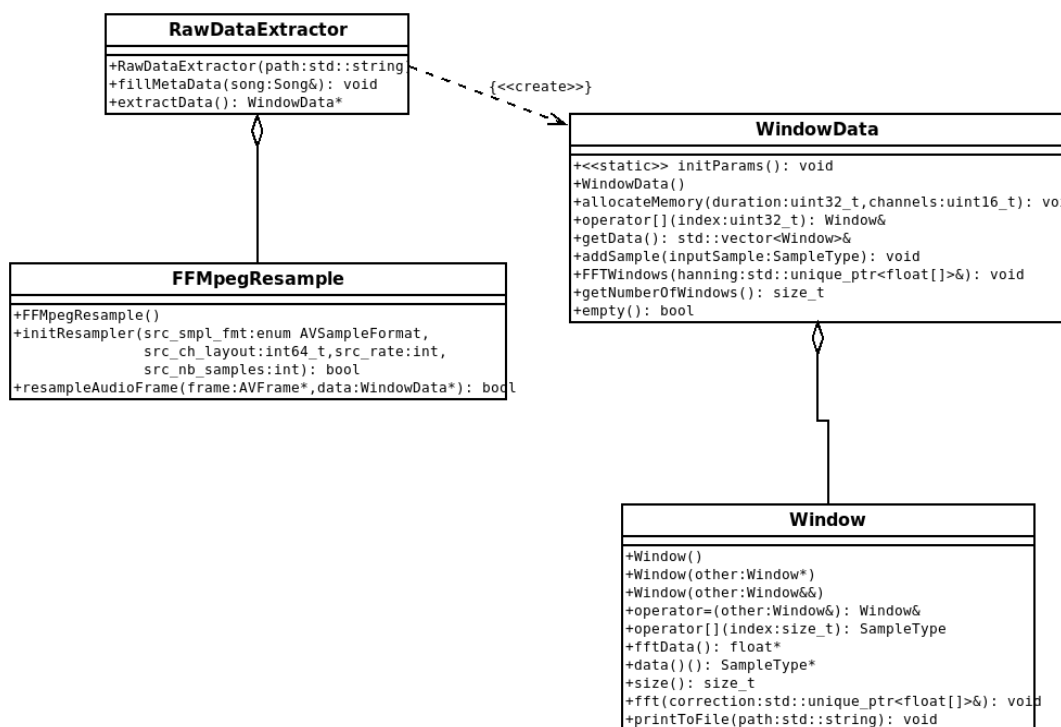


Slika 2.8: Prikaz ulaznog signala u vremenskom domenu i prikaz rezultata DFTa (frekvencijski domen)

2.3 Implementacioni detalji

Kao što je bilo prethodno reći o tome, u okviru ovog rada je razvijen sistem koji će demonstrirati sve pomenute koncepte. Koristimo reč sistem jer svaka komponenta u okviru aplikacije razvijana je sa mogućnošću da se koristi samostalno. U ovoj sekciji će biti opisane komponente i organizacija dela za obradu signala i pripremu za računanje atributa. Problem postojanja velikog broja formata audio zapisa unosi veliki nivo kompleksnosti, potrebno je podržati barem najčešće formate. Svaki format sa sobom nosi različite načine zapisa uzorka, nivoa kompresije, broja kanala i sličnog. Da bi bilo moguće raditi smislenu analizu i poređenje, prvi korak je svođenje na istu kategoriju, odnosno prevođenje u isti format sa istim brojem uzoraka, kanala i tipom uzorka. Komponente koje to rade su `RawDataExtractor` i `FFmpegResampler`. Komponentne u "ispod haube" koriste biblioteku `FFMPEG` prilikom rada sa formatima. Njihova primarna uloga je da za datu putanju audio datoteke pročitaju uzorke i formiraju okvire za kasniju obradu. Iako `FFMPEG` biblioteka podržava rad sa formatima, potrebno je prilagoditi ih za izvlačenje audio signala u odgovarajućem obliku, stoga su ove komponente razvijene. Prilikom čitanja uzoraka bilo je neophodno izvršiti njihovu modifikaciju i ponovno uzorkovanje radi "svođenja na isto" zbog kasnije ravnopravne analize i poređenja. `FFmpegResampler` služi da svaki fajl svede na predefinisano, istu za sve, frekvenciju uzimanja uzoraka. Pored toga se radi i pretvaranje višekanalnog zapisa u jednokanalni ili mono zapis. Postojanje više kanala, u zavisnosti od sadržaja, može da oteža ili olakša posao, međutim u ovom slučaju to znatno proširuje prostor pretrage i prostor parametara, što nije fokus ovog rada. Stoga se radi prebacivanje u jedan

kanal koji će sadržati sve informacije sa svih postojećih kanala. Ovakav pristup predstavlja kompromis između kompleksnosti rezultujućeg sistema i izračunavanja. Za aproksimaciju mere sličnosti, prema ljudskom shvatanju muzike, broj kanala ne predstavlja toliko bitan uslov. `RawDataExtractor` za prosleđenu putanju obezbeđuje formiranu i napunjenu komponentnu `WindowData` koja sadrži formirane okvire prema konfiguracionim parametrima. U trenutku pisanja komponente terminologija je bila premutovana, pa se zapravo reč `Window` odnosi na okvir. `WindowData` je zapravo omotač oko vektora `Window` komponente. Obe komponente služe kao kontejner za čuvanje podataka u željenom i odgovarajućem obliku. Konfiguracioni parametri vezani za dužinu okvira, frekvenciju uzimanja uzoraka i sličnog se nalaze u `config.json` datoteci. Na slici 2.9 se nalazi dijagram klasa pomenutih komponenti, sa akcentom na javni interfejs i relacije između komponenti.



Slika 2.9: Klasni dijagram za komponente vezane za obradu audio signala i čuvanje relevantnih podataka.

2.3.1 Biblioteke

Postoji veliki broj formata audio zapisa. Svaki podrazumeva različit nivo kompresije i različite načine zapisa podataka, bilo da li je do tipa zapisa uzorka (`uint16_t`, `float`, `uint8_t`), broj kanala i slično. Kako to nije glavna tema ovog rada, za korišćenje, čitanje i neznatnu obradu ulaznih audio datoteka korišćenja je biblioteka FFmpeg (www.ffmpeg.org). FFmpeg je biblioteka otvorenog kôda za rad sa multimedijalnim datotekama. Sastoji se iz nekoliko manjih biblioteka za obradu multimedijalnog signala. Podržava rad sa audio i video materijalom i orijentisana je da objedini sve

potrebne formate i kodeke u okviru jedne biblioteke. Primarni fokus biblioteke je okrenut ka konverziji i čitanju formata radi kasnije reprodukcije. Takođe je podržano osnovno modifikovanje multimedijalnog materijala. FFmpeg predstavlja industrijski standard kada je obrada multimedijalnog signala i danas je veoma rasprostranjena i stoji iza većine aplikacija koje se bave ovom tematikom. Zbog svoje stabilnosti i širine predstavlja logičan izbor za realizaciju ovog rada, osim toga pisana je u C-u. C interfejs se bez mnogo muke integriše i koristi u C++ kôdu, što je još jedan razlog za korišćenje biblioteke. Takođe, zbog svoje popularnosti postoji verzija biblioteke za većinu dostupnih platformi i operativnih sistema. U okviru ovog rada biblioteka je korišćena za čitanje i dekompresiju ulaznih audio datoteka i fft transformaciju ulaznog signala. Implementirana je i mogućnost čitanja dostupnih meta podataka, ali akcenat nije na korišćenju istih, već kao način za dobavljanje podataka za kasniji bolji estetski prikaz.

Poglavlje 3

Atributi

Sličnost muzike je, kao što je ranije pomenuto, procena sličnosti muzičkih numera često zasnovana na subjektivnom doživljaju. Atributi kojima se opisuje muzička numera se mogu podeliti u tri osnovne kategorije:

- Muzički atributi niskog nivoa apstrakcije
- Muzički atributi srednjeg nivoa apstrakcije
- Muzički atributi visokog nivoa apstrakcije

Atributi srednjeg nivoa apstrakcije su uvedeni da bi premostili semantički raskorak između atributa niskog i visokog nivoa. Fokus ovog rada je na atributima niskog nivoa i njihovo korišćenje za formiranje mere sličnosti. Uprkos tome formulisaće se i osnova za attribute srednjeg nivoa apstrakcije.

3.1 Atributi niskog nivoa apstrakcije

Atributi niskog nivoa se direktno računaju na osnovu audio signala u vremenskom i/ili frekvencijskog domena. Obradiće se najkorišćeniji atributi niskog nivoa apstrakcije. Svi atributi se računaju na nivou okvira stoga su svi trenutni, to jest svaki zauzima najmanji mogući vremenski interval koji čovek može da prepozna. Iako se svi atributi računaju na nivou okvira, moguće je sumirati sve vrednosti određenog atributa na nivou jedne muzičke numere. Ovo se može uraditi računajući statističke ocene kao što su medijana, srednja vrednost, varijansa ili maksimum. Kasnije će biti reči o korišćenju statističkih ocena. Postoje i drugi načini sumiranja vrednosti, jedan od njih je modelovanje distribucije pojedinačnog atributa pomoću Gausovog modela mešavine (eng. *GMM - Gaussian mixture model*)[29]. Prilikom diskusije muzičkih atributa niskog nivoa koristićemo terminologiju navedenu u tabeli 3.1.

Muzički atributi niskog nivoa apstrakcije se izračunavaju direktno iz sadržaja muzičke datoteke. Glavna podela ovakvih atributa je na attribute iz vremenskog domena i na attribute iz frekvencijskog

Termin	Opis termina
$s(n)$	Amplituda k-tog uzorka
$m_t(k)$	Vrednost signala u frekvencijskom domenu okvira t u frekvencijskom opsegu k
N	Dužina okvira to jest, broj uzoraka u pojedinačnom okviru
K	Broj frekvencijskih kanti tj. broj najvećeg frekvencijskog opsega

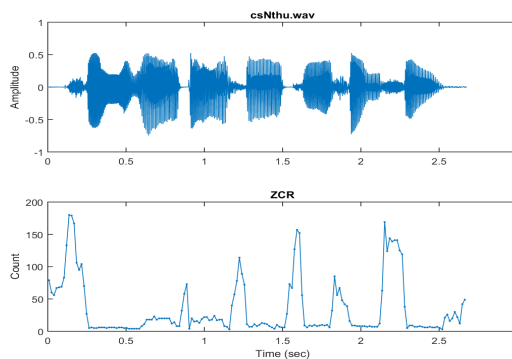
Tabela 3.1: Tabela termina korišćena prilikom diskusiji o atributima niskog nivoa apstrakcije

domena. Atributi iz vremenskog domena se računaju na osnovu uzoraka tj. frekvencijskih amplituda, dok se atributi frekvencijskog domena računaju na osnovu spektrograma [20].

3.1.1 Atributi iz vremenskog domena

3.1.1.1 Broj promena znaka amplitude

Broj promena znaka amplitude (eng. *ZCR - Zero Crossing Rate*) predstavlja broj puta koliko signal pređe iz negativne u pozitivnu vrednost frekvencije i obratno. Može se koristiti kao primitivna mera detekcije visine tona. Vrlo uspešno se koristi u klasifikaciji perkusivnih zvukova i u detekciji govora.



Slika 3.1: Prikaz dijagrama za frekvenciju prelaska nule. Prvi dijagram je audio signal, drugi dijagram predstavlja dijagram frekvencije prelaska nule za dat audio signal.

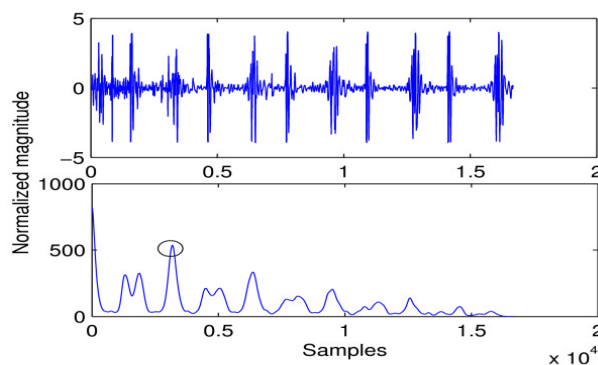
Audio signal koji ne sadrži govor (instrumentalna muzika ili periferni zvuk) broj promena znaka amplitude teži ka visokim vrednostima, dok kod audio signala gde je prisutan govor, obično ima dosta manju vrednost. Zbog ove osobine se koristi kao primitivna mera za određivanje delova signala gde ima, odnosno nema, govora. Iako postoji mnogo stabilniji i tačniji algoritmi [14] [18] za određivanje postojanja govora ili detekcije visine zvuka, zbog relativno lakog izračunavanja a dovoljno dobre aproksimacije pomenutih osobina koristićemo broj promena znaka amplitude. Broj promena znaka amplitude može da služi i kao primitivni indikator koliko smetnji postoji u analiziranom signalu. Visoka vrednost obično znači da su smetnje prisutne u velikoj meri u analiziranom signalu. Izračunavanje broja promena znaka amplitude predstavljeno je u jednačini 3.1. Prilikom formiranja podataka treba obratiti pažnju na tip uzorka i skup odnosno interval vrednosti pomoću kojih se kodira digitalni signal koji analiziramo. Konkretni problem sa tipom uzoraka je što uzorak

može da bude neoznačenog tipa, to jest može da uzima samo pozitivne vrednosti. U ovom slučaju se umesto nule bira prag, obično na sredini intervala.

$$ZCR = \frac{1}{2} \cdot \sum_{n=0}^{N-1} |sgn(s(n)) - sgn(s(k+1))| \quad (3.1)$$

3.1.1.2 Autokorelacija prvog reda

Autokorelacija prvog reda (eng. *First Order Autocorrelation*) je mera koliko se signal poklapa sa pomerenom verzijom samog sebe u vremenu, odnosno prosečna vrednost odstupanja svaka dva susedna uzorka u signalu. Kao i frekvencija prelaska nule, može da se iskoristi da se identifikuju nagle promene intenziteta muzike tj. delove muzičke numere gde je pojačan intezitet muzičkih instrumenata i/ili vokala, osim toga veoma je korisna kao primitivna mera grubosti boje zvuka ¹.



Slika 3.2: Prikaz dijagrama autokorelacije prvog reda na primeru snimka otkucaja srca

Autokorelacija se često koristi prilikom određivanja mere sličnosti dva audio signala, pogotovo zbog osobine da može da aproksimira razliku postojećeg sa istim signalom ali različitom fazom. Autokorelaciju prvog reda izračunavamo pomoću formule prikazane jednačinom 3.2.

$$FOA = \frac{1}{N} \cdot \sum_{n=1}^{N-1} s(n) \cdot s(n-1) \quad (3.2)$$

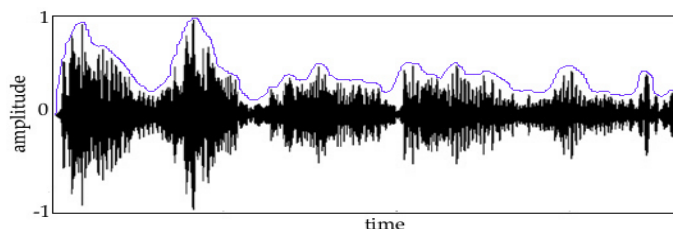
3.1.1.3 Maksimalna vrednost amplitude

Maksimalna vrednost amplitude (eng. *Amplitude Envelope - AE*), kao što joj ime kaže, je maksimalna vrednost uzorka u okviru. Stoga se, prirodno, računa u vremenskom domenu. Formalna definicija je data u formuli (3.3)

$$AE = \max \{s(n) : n \in 0, 1, \dots, N-1\} \quad (3.3)$$

¹<http://www.musanim.com/wavalign/foote.pdf>

Maskimalna vrednost amplitude je veoma prost atribut koji se vezuje za muzički otkucaj (eng. *Music Beat*). Muzički otkucaj je osnovni deo takta, odnosno svaki takt se sastoji iz određenog broja otkucaja.



Slika 3.3: Uprošćeni prikaz maksimalne vrednosti amplitude u odnosu na audio signal.

Zbog svoje prirode je veoma podložan uticaju ekstremnih vrednosti. Ekstremna vrednost je vrednost signala koja u velikoj meri odudara od prosečne vrednosti, na primer, ako nam je od 100 uzoraka, 98 uzoraka u intervalu $[a, b]$ dok su poslednja dva uzorka iz intervala $[2 \cdot b, 6 \cdot b]$ poslednja dva uzorka se mogu okarakterisati kao ekstremne vrednosti.

3.1.1.4 Srednjekvadratna energija

Srednjekvadratna energija (eng. *Root-Mean-Square Energy* ili *RMS energy*) je još jedan atribut iz domena vremena. U literaturi se često naziva i RMS energija kao i RMS nivo ili RMS jačina. RMS potiče kao skraćenica engleskog naziva ovog atributa. Vezuje se za intezitet prepoznatog zvuka. Računa se prema formuli (3.4)

$$RMSEnergy = \sqrt{\frac{1}{N} \cdot \sum_{n=0}^{N-1} s(n)^2} \quad (3.4)$$

Može se koristiti za određivanje nivoa glasnoće muzičke numere i kao indikator novih događaja u audio segmentu [3]. Srednjekvadratna energija, kao atribut, je veoma slična maksimalnoj vrednosti amplitude, ali je manje osetljiva na ekstremne vrednosti, koje se često tumače kao šum.

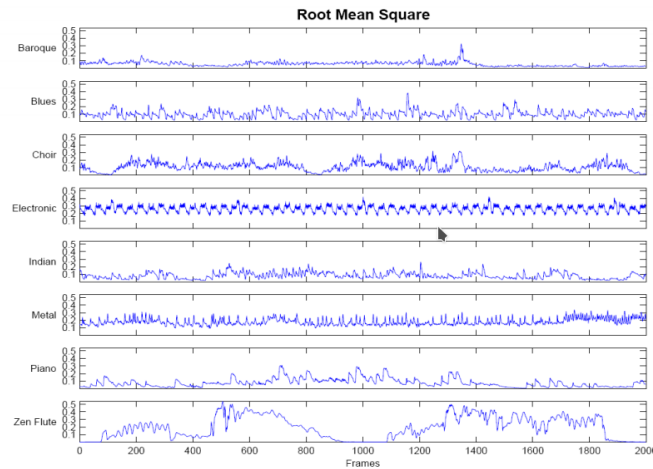
Srednjekvadratna energija se može koristiti i za određivanje osobina muzičke numere vezanih za muzički otkucaj. Zbog svoje povezanosti sa muzičkim otkucajem, ritmom i glasnoćom muzičke numere, neretko se ovaj atribut koristi i u aproksimaciji emocije koju muzička numera sa sobom nosi.

3.1.2 Atributi iz frekvencijskog domena

3.1.2.1 Spektralni centroid

Spektralni centroid (eng. *Spectral Centroid*) predstavlja epicentar magnitude spektra, odnosno frekvenciju gde je energija najviše koncentrisana. Ovaj atribut se najviše koristi kao mera svetline muzike, stoga je direktno vezana sa bojom zvuka [8][22]. Boja zvuka je karakteristična za svaku vrstu

RMS Energy: Illustration



Slika 3.4: Prikaz srednjekvadratne energije, po okvirima, za različite muzičke žanrove

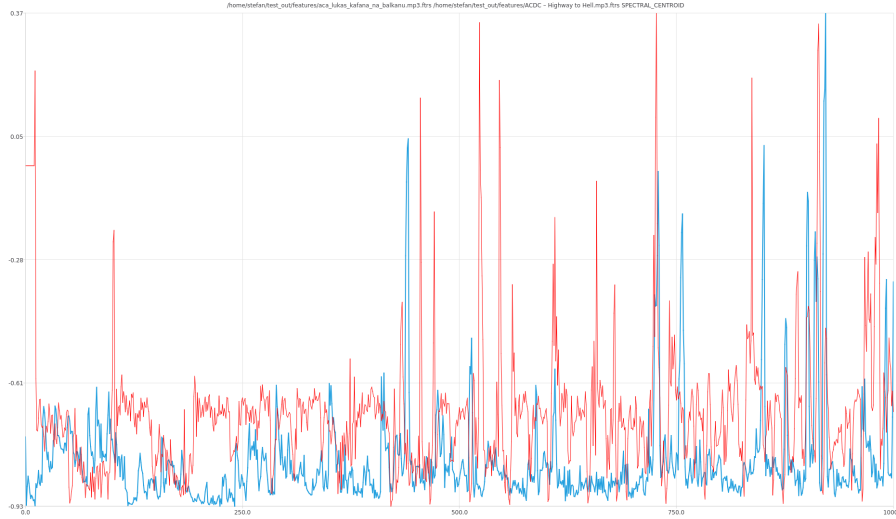
zvuka. Isti ton može zvučati drugačije u zavisnosti od izvora zvuka. Boju određuje broj harmonika, odnosno njihovih amplituda i frekvencija. Parni harmonici daju toplinu i mekoću zvuku, dok neparni daju hladnoću i oštrinu [16]. Spektralni centroid je veoma osetljiv na frekvencijske filtere, pogotovo na one koji "razblažuju" visoke frekvencije. Razlog ovome je što visoke frekvencije nose veću težinu prilikom računanja vrednosti spektralnog centroida. Ovaj problem posebno dolazi do izražaja kada je frekvencija uzimanja uzoraka mala. Posmatrajmo slučaj kada se ista numera posmatra u dve različite frekvencije uzimanja uzoraka, 44KHz i 11KHz. Na osnovu Nyquist-Shannon-ove teoreme, smanjenje frekvencije uzoraka će odseći ili proizvesti distorziju visokih frekvencije viših od polovine frekvencije uzimanja uzoraka. Prema tome, vrednost spektralnog centroida će se promeniti shodno izmeni i uticaju na amplitude frekvencije uzoraka. Formula za dobijanje vrednosti spektralnog centroida:

$$SC = \frac{\sum_{k=1}^K m_t(k)^2 \cdot k}{\sum_{k=1}^K m_t(k)^2} \quad (3.5)$$

Na slici 3.5 prikazane su vrednosti spektralnog centroida za prvih desetak sekundi pesama koje pripadaju potpuno različitim žanrovima. Da bi primer bio što ilustrativniji, korišćena je jedna pesma koja pripada tzv. novokomponovanom stilu (Aca Lukas - Kafana na Balkanu) i jedna pesma "hard rok" stila (ACDC - Highway to hell). Podaci su normalizovani u odnosu na sebe na interval [-1,1] radi lepšeg prikaza.

3.1.2.2 Spektralni raspon

Spektralni raspon (eng. Spectral Spread), takođe poznat u literaturi i kao Spektralni protok (eng. *Spectral Bandwidth*) je atribut izveden iz spektralnog centroida. Spektralni protok predstavlja raspon spektra interesantnih delova signala, odnosno delova oko centroida. Može se interpretirati i kao



Slika 3.5: Vrednosti spektralnog centroida Kafana na Balkanu - Aca Lukas (plavo) i Highway to hell - ACDC (crveno)

varijansa srednje frekvencije signala. Definicija je data formulom 3.6. Prosečna vrednost spektralnog raspona muzičke numere se može koristiti kao opisna ocena zapažene boje zvuka [12].

$$SS = \frac{\sum_{k=1}^K |k - SC| \cdot m_t(k)}{\sum_{k=1}^K m_t(k)} \quad (3.6)$$

3.1.2.3 Glatkoća zvuka

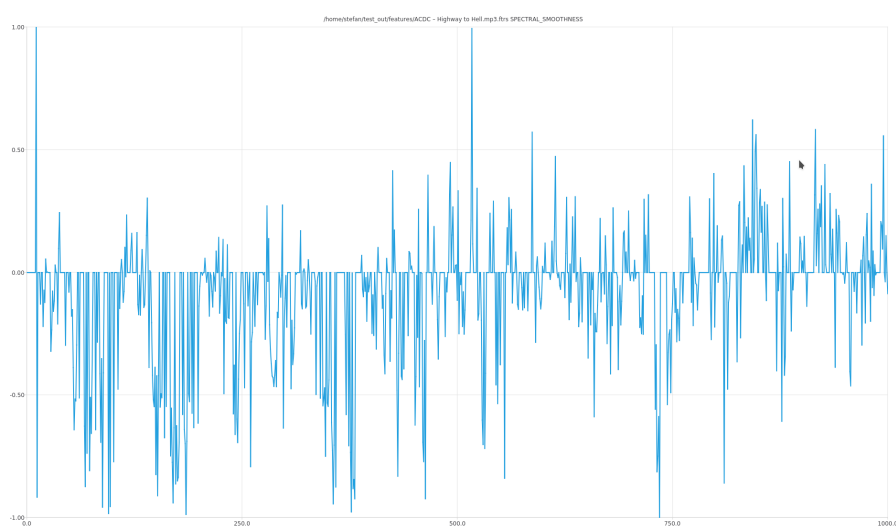
Glatkoća zvuka (eng. *Smoothness*) predstavlja ocenu koliko je zvuk "ravan" ili preciznije predstavlja ocenu meru promene zvuka u spektrumu [11]. To je mera koliko je signal rasprostranjen po spektrumu. Beli šum (eng. *white noise*), signal koji ima istu snagu na svim frekvencijama, će imati glatkoću oko 1, dok sinusoidni talas će imati glatkoću 0, pošto ima jedan vrh u spektrumu [19].

Koristeći se ovim osobinama glatkoća zvuka nam daje dobru aproksimaciju buke (šuma) kao i ocenu harmoničnosti signala. Glatkoća se računa pomoću formule 3.7. Osim pomenutih primena, glatkoća zvuka se takođe koristi dosta i u prepoznavanju instrumenata. Zbog svoje karakterističnosti da reaguje na vrhove signala, može se modelovati sistem koji na osnovu ovog parametra može prepoznati da li određeni instrument učestvuje u numeri. [26]

$$SpectralSmoothness = \sum_{k=1}^K 20 \cdot \log_e m(k) - \frac{20 \cdot \log_e m(k-1) + 20 \cdot \log_e m(k) + 20 \cdot \log_e m(k+1)}{3} \quad (3.7)$$

3.1.2.4 Spektralna asimetrija

Spektralna asimetrija (eng. *Spectral Dissymetry*) ili iskrivljenost signala je mera koliko je spektrum iskrivljen u odnosu na spektralni centroid i shodno tome, koliko teži niskim odnosno visokim



Slika 3.6: Glatkoća zvuka u prvih 10ak sekundi pesme Highway to hell. Podaci su normalizovani u odnosu na sebe na interval [-1,1]

frekvencijama. Zajedno sa glatkoćom zvuka i rasponom predstavlja još jedan atribut koji nam pruža informaciju o obliku spektra. U ovom slučaju, mera je bazirana na harmoničnosti zvuka oko spektralnog centroida. Spektralna asimetrija se dobija korišćenjem formula 3.8.

$$SpectralAsymetry = \sqrt{\frac{\sum_{k=1}^K m(k) \cdot (\frac{D_{rate}}{2} - SC)}{\sum_{k=1}^K m(k)}} \quad (3.8)$$

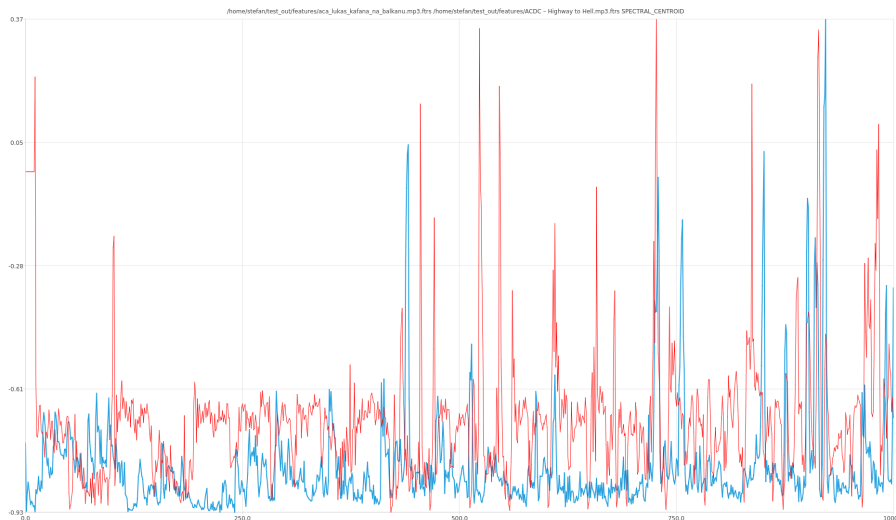
D_{rate} predstavlja frekvenciju uzimanja uzoraka referentnog signala, u ovom slučaju, kao što je već napomenuto, vrednost frekvencije uzimanja uzoraka je 44100Hz, što direktno znači da je $D_{rate} = 22050$. Na slici 3.7, kao kod slučaja spektralnog centroida, su prikazane normalizovane vrednosti spektralne asimetrije za dve muzičke numere. Jedna pripada novokomponovanom žanru, dok druga pripada hard rok žanru.

3.1.3 Implementacioni detalji

Računanje pomenutog skupa atributa se dešava u komponenti `WindowFeatureExtractor`, čija se implementacija nalazi u okviru datoteka `WindowFeatureExtractor.hh/.cc`. Računanje skupa atributa je za sada implementirano preko skupa metoda pomenute komponente, odnosno klase. U daljem radu i proširenju sistema treba uzeti u obzir i moguć način spoljašnjeg definisanja računanja atributa, u vidu dodataka (eng. *plug-in*), ili pomoću nekog definisanog pseudo-računarskog jezika koji će se izvršavati dinamički.

3.1.4 Atributi srednjeg nivoa

U prethodnom tekstu su definisani nivoi atributa koji se koriste za aproksimaciju muzičkih numera. Fokus ovog rada je na atributima niskog nivoa, ali ćemo zbog kompletnosti, ukratko objasniti i



Slika 3.7: Vrednosti spektralne asimetrije Kafana na Balkanu - Aca Lukas (plavo) i Highway to hell - ACDC (crveno).

atribute srednjeg nivoa. Atributi niskog nivoa su prvenstveno statističke ocene nad signalom i njegovom frekvencijskom i/ili vremenskom obliku. Sami za sebe ne nose nikakvu značajnu informaciju koju bi čovek mogao da interpretira i na osnovu njih izvuče smislene zaključke. Na primer, na osnovu samih brojeva ne može se zaključiti da li je pesma bržeg ili sporijeg tempa, da li je reč o baladi ili pesmi veselije melodije itd. Takvi opisi se nazivaju atributi visokog nivoa, odnosno atributi koje čovek koristi da bi opisao pesmu. Da bi se premostio jaz između atributa niskog nivoa i visokog nivoa uvode se atributi srednjeg nivoa koji kombinuju vrednosti jednog ili više atributa niskog nivoa da bi se bliže definisali atributi visokog nivoa.

U okviru atributa srednjeg nivoa se često koriste Cepstralni koeficijenti Melove frekvencije. Cepstrum predstavlja rezultat dobijen primenom inverzne furijeove transformacije nad logaritmom spektruma. Ime Cepstrum potiče od permutacije prva četiri slova engleske verzije termina spektrum. Cepstrum se dosta koristi u analizi i prepoznavanju govora, a i u oblasti istraživanja muzičkih podataka [1]. Mel je izvedena jedinica za frekvenciju koja se definiše preko formule

$$mel = 1127 \cdot \log\left(1 + \frac{f}{700}\right) \quad (3.9)$$

gde je f predstavlja vrednost frekvencije u hercima.

Takođe se, za svrhe definisanja atributa srednjeg nivoa se izvodi još jedna jedinica za frekvenciju, Bark.

$$bark = \left(\frac{26.81 \cdot f}{1960 + f}\right) - 0.53 \quad (3.10)$$

gde f takođe predstavlja vrednost frekvencije u hercima [20].

Izvedene jedinice frekvencije se koriste radi bližeg određivanja vrednosti frekvencije ljudskom shvatanju muzike. Pomenutim formulama dobijamo preslikavanje frekvencija tako da se više obraća

WindowFeatureExtractor
<pre> -m_win_size: uint32_t -m_win_per_sec: uint32_t -m_fft_bins: uint32_t -m_dest_samplerate: uint32_t -m_Sx: int32_t -m_Sxx: int32_t -m_total_amp: float </pre>
<pre> +getFeatures(): vector<vector<float>> -f_zeroCrossingRate(SampleType*): float -f_firstOrderCorrelation(data:SampleType*): float -f_rms_energy(data:SampleType*): float -f_max_amp(data:SampleType*): float -f_spectralCentroid_fftf(data:float*): float -f_linearRegression_fftf(data:float*): float -f_spectralSmoothness_fftf(data:float*): float -f_bandEnergyRatio_fftf(data:float*): float -f_spectralDissymetry_fftf(float*,SpectralCentroid:float): float -f_spectralSpread_fftf(data:float*,SpectralCentroid:float): float -initNormalizationWindow(): void -initParams(): void -initValues(): void -windowFeatureCalc(win:Window&): vector<float> </pre>

Slika 3.8: Klasni dijagram implementacionih detalja za računanje atributa. Samo jedna komponenta je odgovorna za ovo stoga samo jedna klasa

pažnja na intervale odnosno vrednosti frekvencija na koje čovek više reaguje. Razlog izvođenja ovih jedinica leži u psihoakustici, oblasti koja se bavi ljudskom percepcijom zvuka, odnosno načinu na koji čovek interpretira neke od osnovnih zvučnih karakteristika, glasnoću, frekvenciju, intezitet i slično.

Osim pomenutih izvedenih jedinica, definisanje atributa srednjih nivoa se odnosi i na kombinaciju dobijenih atributa. Ovde se prvenstveno misli na računanje atributa po okviru, pa se rezultati kasnije kombinuju pomoću Gausovih modela i vektorizacije radi daljeg računanja. Vektorizacija podrazumeva grupisanje podataka u vektore. Korišćenjem vektora umesto pojedinačnih promenljivih dosta uprošćava potrebne korake pri računu svodeći na vektorski račun. Još jedna bitna tehnika je formiranje takozvanih blok-okvira. Blok-okvir je grupisanje određenog broja okvira u posebnu grupu koja će kasnije tokom izračunavanja sličnosti biti uzeta u obzir prilikom računanja, to direktno znači da će svaki blok učestovati u izračunavanju aproksimacije sličnosti. Ovaka pristup je vrlo efikasan, iako računski prilično zahtevan. Efikasnost se prvenstveno ogleda u sposobnosti da se računa sličnost za pojedinačne fragmente muzičke numere a ne cele muzičke numere u potpunosti. Ovde se koriste tehnike logaritmski fluktuirajućih šablona, spektralnih šablona i slično. Detaljno objašnjenje tehnika i pojmova atributa srednjeg nivoa izlazi iz obima ovog rada. Više informacija o ovim atributima može se pronaći u [20].

Poglavlje 4

Mera sličnosti

U ovom poglavlju biće objašnjen rad sa atributima niskog nivoa i njihovom osnovnom korišćenju radi trivijalne aproksimacije sličnosti muzičkih numera na osnovu prirodnih atributa zvuka. Meru sličnosti ćemo definisati kao rastojanje između dve tačke u višedimenzionalnom prostoru. Ose ovog prostora će predstavljati statistički parametri koje ćemo izračunavati nad atributima. Svaki atribut iz skupa atributa niskog nivoa se računa na nivou okvira. Na osnovu toga koristiće se statistički parametri da bi sumirali ‘prosečne’ vrednosti na nivou cele muzičke numere, odnosno malim skupom vrednosti aproksimirali promenu tog atributa kroz sve okvire u muzičkoj numeri.

Definisaće se pet statističkih parametara koje će se koristiti u okviru mere sličnosti. Ovo znači da će svaki izračunat atribut niskog nivoa biti predstavljen sa 5 vrednosti na nivou cele muzičke numere. Ovakav pristup je jednostavniji kada se upoređi sa kompleksnijim načinima određivanja aproksimacija muzičkih numera, kao na primer gausovih modela ili neke slične tehnike koja se zasniva na određivanju (i upoređivanju) šablona u okviru muzičke numere. Jednostavan pristup nam takođe donosi i veoma malu računsku složenost, pa samim tim je krajnji sistem dosta robusniji i sposobniji da se brže i lakše izbori sa većim količinama ulaznih podataka. Ali isto tako je podložniji greškama usled međusobno permutovanih delova muzičke numere, pošto kao što je već rečeno, ocene gledaju pesmu u celosti a ne po segmentima. U okviru ovog rada definisaćemo statističke parametre za svaki atribut pojedinačno na nivou cele muzičke numere, odnosno u obzir se uzimaju svi okviri odjednom umesto grupisanja okvira u posebne grupe koje će kasnije učestvovati u računanju mere sličnosti.

Na kraju će se na osnovu definisanog test skupa odrediti koje ocene najviše utiču na attribute visokog nivoa koji će se definisati i odrediti na osnovu subjektivnog osećaja. Iako ovaj pristup sa sobom nosi i određene nedostatke, on predstavlja dobru osnovu za dalji rad i unapređivanje mera sličnosti muzičkih numera.

Prilikom računanja atributa, kao i statističkih parametara koristićemo jedan vid normalizacije. Prilikom čitanja vrednosti uzoraka iz muzičke datoteke koriste se `int16_t` vrednosti, odnosno označene celobrojne vrednosti iz intervala $[-32768, 32767]$. U toku učitavanja ove vrednosti se normalizuju na interval $[-1, 1]$ radi, uslovno rečeno, lepših vrednosti atributa.

4.1 Statistički parametri

Statistički parametri se koriste da se brojevnom vrednošću bliže odrede vrednosti skupa podataka. U ovom slučaju je to odličan alat za okvirnu aproksimaciju sličnosti muzičkih numera. Ovde se koristi pet statističkih parametara odnosno ocena, da bi se dobila približna slika o fluktuaciji vrednosti atributa kroz okvire.

- **Srednja vrednost** je osnovna statistička ocena. Informaciju koju pruža je prosečna vrednost atributa na nivou cele muzičke numere. Srednja vrednost se može iskoristiti da grubo aproksimira vrednost atributa i poslužiti za osnovno upoređivanje atributa između dve muzičke numere

$$SrednjaVrednost = \frac{\sum_{i=0}^{N-1} s_i}{N} \quad (4.1)$$

- **Varijansa** predstavlja očekivano matematičko očekivanje odstupanja slučajne promenljive od srednje vrednosti. Varijansa nikada ne sme biti negativna vrednost, to ukazuje na grešku u računu.

$$Varijansa = \frac{\sum_{i=0}^{N-1} (s_i - SrednjaVrednost)^2}{N} \quad (4.2)$$

- **Standardna devijacija**, u statistici, predstavlja apsolutnu meru disperzije skupa, odnosno koliko u proseku elementi odstupaju od aritmetičke sredine skupa. Standardna devijacija i varijansa su u direktnom odnosu, pošto je varijansa kvadratni stepen standardne devijacije. U okviru implementiranog sistema i aproksimacije sličnosti, je zadržana zajedno sa varijansom kao statistički parametar zbog uticaja na izračunavanje ostalih statističkih parametara.

$$StdDev = \sqrt{Varijansa} \quad (4.3)$$

- **Koeficijent asimetrije (eng. *Skewness*)** je statistička mera koja pokazuje koliko se vrednosti skupa slučajne promenljive razlikuju u odnosu na normalnu raspodelu, odnosno definiše asimetriju skupa sa normalnom raspodelom.

$$Koeficijent_{asimetrije} = \frac{\sum_{i=0}^{N-1} (s_i - SrednjaVrednost)^3}{N} \cdot StdDev^3 \quad (4.4)$$

- **Koeficijent spljoštenosti (eng. *Kurtosis*)** mera koliko su vrednosti skupa spljoštene u odnosu na normalnu raspodelu. Aproksimira se vrh krive koja predstavlja vrednosti iz skupa slučajne veličine.

$$Koeficijent_{spljotenosti} = \frac{\sum_{i=0}^{N-1} (s_i - SrednjaVrednost)^4}{N} \cdot StdDev^4 \quad (4.5)$$

4.2 Mera sličnosti

Mera sličnosti će se definisati kao rastojanje u više dimenzionalnom prostoru, odnosno posmatrač će sve dobijene vrednosti kao osu u višedimenzionalnom prostoru. Svaka muzička numera će na osnovu vrednosti dobijenih izračunavanjem atributa i statističkih parametara imati mesto u tom prostoru. Meru sličnosti definišemo kao rastojanje između dve tačke u tom prostoru. U ovom slučaju će se koristiti Euklidsko rastojanje koje se definiše na sledeći način:

$$EuclideanDist = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad (4.6)$$

Gde su $A_i, B_i, i \in [0, n - 1]$ izračunate vrednosti statističkih parametara za svaki od atributa niskog nivoa za dve muzičke numere A i B . n predstavlja ukupan broj vrednosti, po muzičkoj numeri, koje učestvuju u izračunavanju mere sličnosti (*brojAtributa · brojStatistckihParametara*).

Ovakav način aproksimacije sličnosti potencijalno zanemaruje odnos intervala u muzičkim numerama, odnosno vrlo je verovatno da ista muzička numera sa izmenjenim redosledom dobije isto mesto u višedimenzionalnom prostoru u okviru kog određujemo sličnost dve muzičke numere. Može se diskutovati da li je muzička numera sa permutovanim intervalima ista muzička numera? Ovde se misli na prvenstveno drugačiji redosled istih intervala u muzičkoj numeri, na primer, okrenuti strofa i refren, ili drastičniji slučaj, kada se ispremeštaju mnogo manji vremenski intervali na slučajan način. Radi lakše implementacije i otpornosti samog pristupa na različitost u dužini trajanja muzičkih numera, u ovom radu je korišćen pristup gde je muzička numera posmatrana u celosti. I pored očiglednih nedostataka, mera sličnosti nam pruža subjektivno primetan poredak i samim tim zadovoljava početni uslov gde se redosled muzičkih numera određuje na osnovu samog audio signala i odgovara subjektivnom osećaju korisnika (čoveka) na muzičke numere.

Treba formulisati način na koji će se mera sličnosti upotrebiti u svrhu što boljeg generisanja reprodukcione liste. Prisutnost semantičkog raskoraka, dosta otežava formiranje dobrog poretka muzičkih numera. Korišćenje dobijenih vrednosti pruža objektivnu aproksimaciju sličnosti audio numera, kako je ideja da se mera sličnosti iskoristi u svrhu pravljenja odgovarajuće liste, potrebno je dodatno analizirati dobijene podatke i odrediti vezu, ukoliko postoji, između karakteristika muzičke numere koje čovek koristi da bi opisao pesmu. Ovakvim pristupom može se približno odrediti koje ocene direktno utiču na karakteristike od interesa. Mera sličnosti u osnovi predstavlja relaciju između dve muzičke numere. Pristup apsolutnog određivanja ocena na nivou cele pesme nam omogućava da dobijene ocene iskoristimo samostalnu aproksimaciju muzičke numere. Ovo efektivno znači da, ukoliko se izoluje određena ocena, možemo sortirati odnosno kreirati listu uređenu prema zadatom kriterijumu. Osim toga, pruža dobru osnovu za korišćenje raznih metoda mašinskog učenja za klasterovanje i upoređivanje i izolovanje sličnih pesama, što nije fokus ovog rada, ali predstavlja dobru osnovu za dalje istraživanje. Da bi definisali takav pristup koristićemo test skup nasumično

izabраниh muzičkih numera. Test skup je formiran tako da se biraju 3 (ili više) numera iz istog žanra, odnosno subjektivnog osećaja o sličnosti. Ovakav pristup nam pruža dobru sliku o tome koliko je mera sličnosti efikasna i koliko je sličnost dobro određena.

4.3 Test skup

Izabrani test skup čini 21 muzička numera, odabrane prema subjektivnom osećaju. Muzičke numere su preuzete sa sajta YouTube, odnosno samo audio tok podataka sa video klipova. Treba napomenuti da muzička numera generisana na ovaj način može da sadrži i neke nepotrebne elemente, kao što su pozadinski razgovor u toku spota ili tišinu tokom odjavne scene.

Tabela 4.1: Test skup

ID	Izvođač	Muzička numera	Žanr	Kategorija 1	Kategorija 2
1	Lady Gaga	Bad Romance	Techno	brza	jaka
2	Foreigner	I want to know what love is	RnR	spora	dinamična
3	Robbie Williams and Nicole Kidman	Something stupid	Pop-Bolero-Jazz	melodična	romantična
4	Luis Fonsi ft. Daddy Yankee	Despacito	Pop-regeton	brza	latino
5	Ricky Martin	Maria	Latino pop	energična	latino
6	Coldplay	Fix you	Alternativni Rock	spora	smirujuća
7	Eminem	Rap God	Hip-hop	brza	energična
8	Beogradski sindikat	Sistem te laže	Hip-hop	agresivna	energična
9	Joe Cocker	You can leave hat on	Rock	srednji tempo	vesela
10	Ceca	Znam	Narodna	ritmična	vesela
11	Orthodox Celts	Far away	Irski folk	brza	energična
12	Đorđe Balašević	Lepa protina kći	Šansona	spora	melanholična
13	Šaban Šaulić	Samo za nju	Narodna muzika	balada	srednji tempo
14	Queen	Dont stop me now	Classic RockNRoll	brza	ritmična
15	Guns N Roses	Welcome to the jungle	RockNRoll	brza	jaka
16	Zdravko Čolić	Ti si mi u krvi	Pop	balada	spora
17	Dubioza Kolektiv	Rijaliti	Rap-rock	energetična	stimulativna
18	Johannes Brahms	Hungarian dance	Klasična muzika	dinamična	jaka
19	Ludwig van Beethoven	Za Elizu	Klasična muzika	nežna	ceremonijalna
20	Rimsky-Korsakov	Bumbarov let	Klasična muzika	brza	energična
21	Miroslav Ilić	Božanstvena žena	narodna	melodična	nežna

Na sam sistem takve stvari ne bi trebalo da utiču mnogo, ali su svakako prisutne u svakodnevnoj upotrebi, stoga postojanje nesavršenosti u muzičkim numerama ne utiče značajno na dalju analizu atributa. U tabeli 4.1 je opisan test skup. Test skup sadrži 21 muzičku numeru. Svaka numera je opisana izvođačem, nazivom numere, žanrom kojem pripada, kao i sa dva atributa visokog nivoa. Inicijalno testiranje sistema nad test skupom koji sadrži muzičke numere različitog žanra i muzičke numere sa izraženo različitim prirodnim karakteristikama će dati okvirne vrednosti i intervale u kojima se atributi kreću. Ovaj korak je izuzetno bitan u analizi dobijenog sistema pošto, kao što je već rečeno, koristićemo Euklidsko rastojanje kao aproksimativnu meru sličnosti dve muzičke numere. Kako svaki atribut učestvuje u računanju sličnosti, ako se atributi kreću u različitim intervalima vrednosti, na primer $a = [0, 1]$ i $b = [1000, 5000]$ atribut b bi mnogo više uticao na rezultat rastojanja, od atributa a . Ovakva situacija je nepoželjna, bar ne u nekontrolisanim uslovima, jer atribut a možda nosi mnogo bitnije informacije od atributa b .

Za svaku muzičku numeru su dodata dva atributa visokog nivoa. Vrednosti atributa visokog nivoa su izvedena na osnovu subjektivnog osećaja na skupu ljudi. Kategorije nisu nužno disjunktne, odnosno vrednosti iz jedne i druge kategorije se mogu poklapati. Scenario gde se vrednosti dve kategorije poklapaju, to jest, nose istu informaciju, je vrlo česta pojava. Muzička numera, kao entitet, koji se vrlo subjektivno doživljava, najizraženija osobina pesme ostavlja jak utisak na slušalaca i utiče na doživljaj same muzičke numere. Ovo dovodi do toga da se slušalac opisuje pesmu imajući u vidu dominantnu karakteristiku pesme, često koristeći sinonime za istu karakteristiku, na primer, pesma može biti brza i energična, što se može vrlo lako reći za pesmu koja ima brz tempo. Takođe, to nužno ne mora biti tačno, jer na doživljaj muzičke numere može da utiče i poruka pesme. Ovde se prvenstveno misli na reči pesme i poruke koju muzička numera nosi na "pesničkom" nivou koji nema direktnu vezu sa samim karakteristikama zvuka, odnosno "pozadinske" muzike. Muzička numera može da ima spor ritam i da ima nežne tonove a da ima takav tekst da korisnika može poneti poruka koju tekst nosi i pesmu protumačiti kao energičnu.

Prilikom analize sistema i daljeg modeliranja je bitno imati ispravnu predstavu o podacima. Kako je muzika veoma subjektivan pojam i zavisi od mnogo različitih faktora, od kojih neke trenutno ne možemo izmeriti pomoću računara. Aproksimacija sličnosti će se bazirati na karakteristikama zvuka, koje ne moraju nužno biti dovoljne za tačnu aproksimaciju.

4.3.1 Predstavljanje rezultata

Kao što je već pomenuto sami rezultati nisu dovoljni, to jest, potrebno je izvršiti analizu i po potrebi normalizovanje vrednosti. Normalizacija je neophodna da bi svaki atribut, odnosno ocena, podjednako učestvovala u inicijalnoj aproksimaciji vrednosti.

Tabela 4.2: Vrednosti atributa bez normalizacije ulaznih vrednosti

Atribut	Srednja vrednost	Varijansa	Koeficijent asimetrije	Koeficijent spljoštenosti
Broj promena znaka amplitude	-0.99770	0.00442	29.437	875.606
Autokorelacija prvog reda	0.49110	0.26131	-0.08103	1.28028
Srednjekvadratna energija	0.69864	0.09966	-0.69755	4.86035
Najveća amplituda	-0.99753	0.00491	28.4429	810.131
Spektralni centroid	-0.71259	0.03957	0.06025	1.45597
Linearna regresija	-0.60128	0.1633	0.25590	2.08424
Glatkoća zvuka	-0.99917	0.00094	44.0803	2224.82
Spektralni raspon	-0.67355	0.09124	0.02554	1.0868
Spektralna asimetrija	-0.55007	0.20413	0.01306	1.01433

Tabela 4.3: Vrednosti atributa sa normalizacijom ulaznih vrednosti

Atribut	Srednja vrednost	Varijansa	Koeficijent asimetrije	Koeficijent spljoštenosti
Broj promena znaka amplitude	2.243e-06	4.221e-09	29.437	875.606
Autokorelacija prvog reda	0.74482	0.06520	-0.08102	1.28003
Srednjekvadratna energija	0.84930	0.02491	-0.69735	4.86055
Najveća amplituda	0.00123	0.00122	28.4429	810.131
Spektralni centroid	178.228	15217.2	0.05988	1.45616
Linearna regresija	-1.31383	0.10993	0.25528	2.0844
Glatkoća zvuka	0.02252	0.69573	44.078	2224.66
Spektralni raspon	571.042	279308	0.02623	1.08698
Spektralna asimetrija	1213.71	1.587e+06	0.01287	1.01432

Kao uzorak u uzeta je muzička numera *Despacito*, kao najpregledaniji video snimak na sajtu *YouTube*.

Tabele 4.1 i 4.2 sadrže vrednosti dobijene bez normalizacije ulaznih vrednosti i sa. Prva tabela predstavlja statističke ocene sa vrednostima gde su atributi normalizovani na interval $[-1, 1]$ u odnosu sami na sebe, dok druga tabela predstavlja vrednosti bez normalizacije. Lako se uočava da su vrednosti u prvoj tabeli daleko lepše i uslovno rečeno stabilnije, tako da će se zadržati ovakav vid normalizacije. Moguće je diskutovati o gubitku određenog dela informacija normalizacijom podataka na ovaj način, ali kako se podaci računaju nad okvirima određene dužine, i uzimajući u obzir njihovo vremensko trajanje, mnogo više informacija se dobija analizom fluktuacije njihovih vrednosti, pa samim tim normalizacija podataka u odnosu na same sebe, ne remeti šablon fluktuacije i ne utiče značajno na statističke ocene po atributima.

Nakon inicijalnog pokretanja sistema dobijaju se sledeći rezultati:

Luis Fonsi ft. Daddy Yankee – Despacito.mp3 :	0
Orthodox Celts – Far Away.mp3 :	903.523
Robbie Williams and Nicole Kidman – Something Stupid.mp3 :	2525.16
Bethoven – Fur Elise.mp3 :	5603.6
Johannes Brahms – Hungarian Dance No. 5.mp3 :	7090.75
Dubioza kolektiv – Rijaliti.mp3 :	7282.75
Djordje Balasevic – Lepa protina kci.mp3 :	9863.46
Ricky Martin – Maria.mp3 :	11089.7
Lady Gaga – Bad Romance.mp3 :	11224
Miroslav Ilic – Bozanstvena zeno.mp3 :	11737.4
Saban Saulic – Samo za nju.mp3 :	11774.5
Beogradski Sindikat – Sistem Te Laze.mp3 :	11912.2
Rimsky-Korsakov – Flight Of The Bumblebee.mp3 :	12852.5
Queen – Dont Stop Me Now.mp3 :	13372.5
Ceca – Znam.mp3 :	14424.9
Zdravko Colic – Ti si mi u krvi.mp3 :	15095.1
Coldplay – Fix You.mp3 :	16371.5
Guns N Roses – Welcome To The Jungle.mp3 :	16734.5
Foreigner – I Want To Know What Love Is.mp3 :	17909.6
Joe Cocker – You Can Leave Your Hat On.mp3 :	20539.4
Eminem – Rap God.mp3 :	22584.6

Na osnovu prethodne tabele ocena baziranih na referentnoj pesmi *Despacito* vidi velika razlika u veličinama ocena, odnosno pojedine ocene se kreću u intervalu od $[-1, 1]$ dok recimo spektralna asimetrija dostiže vrednosti preko hiljadu. Pošto se za meru rastojanja koristi Euklidsko rastojanje, veća vrednost pojedinih ocena prirodno pravi veći uticaj na samo rastojanje, pa samim tim i na dobijene rezultate. Potrebno je naglasiti da je računato rastojanje od referentne muzičke numere *Despacito* sa svakom muzičkom numerom ponaosob, odnosno broj predstavlja aproksimaciju sličnosti sa referentnom pesmom, uključujući nju samu.

4.4 Normalizacija statističkih ocena atributa

Normalizacija statističkih parametara baziranih na osnovu atributa i raspodeli njihovih vrednosti po okvirima pesme, na isti način kao što je normalizovana vrednost samih atributa nije optimalna, odnosno količina informacija koja se izgubi time nanosi više štete nego koristi. Pošto u ovom slučaju statistička ocena atributa je singularna vrednost, njena normalizacija u odnosu na samu sebe bi uvek bila maksimalna vrednost tog parametra. Normalizacija ocena pesme je izuzetno potrebna zbog rezultata predstavljenih u prethodnoj sekciji. Najmanje izmerene vrednosti ocena se nalaze u sledećoj tabeli. Svaka ćelija sadrži respektivno minimalnu i maksimalnu vrednost za zadati test skup.

Tabela 4.4: Najmanja i najveća vrednost atributa po okviru

Atribut	Srednja vrednost	Varijansa	Koeficijent asimetrije	Koeficijent spljoštenosti
Broj promena znaka amplitude	-0.9998 0.0000	0.0000 0.0132	0.0000 104.156	0.0000 10849.6
Autokorelacija prvog reda	0.4831 0.4922	0.2597 0.2717	-0.1552 - 0.0684	1.2382 1.5244
Srednjekvadratna energija	0.6895 0.6999	0.0974 0.1144	-1.1584 - 0.6067	4.4033 6.9237
Najveća amplituda	-0.9998 0.0000	0.0000 0.0132	0.0000 104.156	0.0000 10849.6
Spektralni centroid	-0.7142 0.1783	0.03954 0.6678	0.0059 0.0934	1.0081 1.7051
Linearna regresija	-0.6036 - 0.5942	0.1614 0.1764	0.2163 0.4641	1.9260 2.8799
Glatkoća zvuka	-0.9998 0.0000	0.0000 0.0016	0.0000 104.156	0.0000 10849.6
Spektralni raspon	-0.9998 0.0251	0.0002 0.9206	0.0178 126.272	1.0012 15954.7
Spektralna asimetrija	-0.5501 0.9998	0.0002 0.9988	-126.311 0.0317	1.0007 15961.5

Gde se vrednosti svi atributa kreću u okvirnom intervalu $[-126, 15000]$, što se iz može videti u izlaznim podacima aplikacije. Problem sa ovakvim intervalom je što ocene nisu podjednako dobro raspoređene po tom intervalu. Stoga će se primeniti pristup normalizacije ocena po skupu. Pošto ćemo reprodukcionu listu određivati na osnovu sličnosti svake pojedinačne muzičke numere

sa referentnom numerom, na datom skupu određuje se najmanja i najveća vrednost po atributu i normalizovati ocene svake muzičke numere ponaosob. Ovakav pristup može biti osetljiv na promene u skupu nad kojim vršimo procenu sličnosti, međutim za potrebe ovog rada pretpostavićemo da su sve muzičke numere sa kvalitetom snimka koji ne odstupa preterano od prosečnog, odnosno nema dugih i izraženih smetnji koje bi drastično uticale na rezultate.

Rezultati ocena nakon primenjene normalizacije ocena, na referentnoj pesmi su:

Tabela 4.5: Vrednosti atributa nakon normalizacije ulaznih vrednosti i normalizacije vrednosti atributa

Atribut	Srednja vrednost	Varijansa	Koeficijent asimetrije	Koeficijent spljoštenosti
Broj promena znaka amplitude	-0.99577	-0.33081	-0.43475	-0.83859
Autokorelacija prvog reda	0.73987	-0.73537	0.71063	-0.70649
Srednjekvadratna energija	0.73937	-0.73692	0.67070	-0.63734
Najveća amplituda	-0.99544	-0.25636	-0.45384	-0.8506
Spektralni centroid	-0.99635	-0.99988	0.24141	0.28506
Linearna regresija	-0.49246	-0.75229	-0.68107	-0.66838
Glatkoća zvuka	-0.99871	0.14633	-0.15357	-0.58988
Spektralni raspon	-0.3633	-0.80227	-0.99987	-0.99998
Spektralna asimetrija	-1	-0.59164	0.99970	-0.99999

Pokretanje sistema sa normalizovanim ocenama daje nešto bolje rezultate, sudeći bar po subjektivnom osećaju.

Luis Fonsi ft. Daddy Yankee – Despacito.mp3 :	0
Ricky Martin – Maria.mp3 :	3.34701
Coldplay – Fix You.mp3 :	3.83882
Dubioza kolektiv – Rijaliti.mp3 :	3.93188
Lady Gaga – Bad Romance.mp3 :	4.07851
Zdravko Colic – Ti si mi u krvi.mp3 :	4.32923
Foreigner – I Want To Know What Love Is.mp3 :	4.33964
Johannes Brahms – Hungarian Dance No. 5.mp3 :	4.64319
Guns N Roses – Welcome To The Jungle.mp3 :	4.68333
Saban Saulic – Samo za nju.mp3 :	4.79902
Eminem – Rap God.mp3 :	4.98876
Beogradski Sindikat – Sistem Te Laze.mp3 :	5.02376
Orthodox Celts – Far Away.mp3 :	5.30745
Miroslav Ilic – Bozanstvena zeno.mp3 :	5.57593
Djordje Balasevic – Lepa protina kci.mp3 :	6.14696
Queen – Dont Stop Me Now.mp3 :	6.27649
Ceca – Znam.mp3 :	6.62068
Rimsky-Korsakov – Flight Of The Bumblebee.mp3 :	6.80863
Joe Cocker – You Can Leave Your Hat On.mp3 :	6.81283
Bethoven – Fur Elise.mp3 :	7.59254
Robbie Williams and Nicole Kidman – Something Stupid.mp3 :	7.77325

Nakon pažljivog slušanja test skupa uočavaju se znatni skokovi i neskladnosti između uzastopne dve muzičke numere na dobijenoj reprodukcionalnoj listi. Razlog zbog ovoga leži u načinu određivanja sličnosti dve muzičke numere. Prva i najbliža pesma će uvek biti jedinstveno određena, ono što

predstavlja problem je taj što je lista generisana na osnovu distanci od početne referentne pesme. Ako se problem svede u jednu dimenziju i pesme posmatraju kao vrednosti na pravi, najbliža muzička numera može biti sa desne strane referentne muzičke numere dok druga najbliža može biti sa leve strane. Kako reprodukcija ide redom, već kod druge muzičke numere dobijamo nekonzistentne i nedovoljno pouzdane podatke na prvi pogled. Kako reprodukcija odmiče, problem je sve očigledniji i svodi se na slučajan izbor.

Da bi se izbegla pomenuta situacija, potrebno je izmeniti pristup formiranja reprodukcione liste. Umesto da određujemo celu listu prema sličnosti referentnoj muzičkoj numeri, svaku sledeću muzičku numeru određujemo na osnovu sličnosti trenutnoj muzičkoj numeri, izuzimajući sve već reprodukovane muzičke numere. Ovakav pristup obezbeđuje pouzdaniji izbor sledeće muzičke numere i gladak prelaz između žanrova. Sa druge strane, u nekom trenutku se može pojaviti jak skok između sličnosti dve uzastopne numere, što može predstavljati problem. Skok je neminovan ukoliko sličnost odvede na jednu stranu, kako u svakom koraku eliminišemo po jednu muzičku numeru iz skupa nad kojim se izračunava sličnost, dolazimo do momenta gde smo otišli previše u jednu stranu i prva najbliža muzička numera se nalazi na drugom kraju hiperravni.

Luis Fonsi ft. Daddy Yankee – Despacito.mp3 :	0
Ricky Martin – Maria.mp3 :	3.34701
Coldplay – Fix You.mp3 :	1.18222
Foreigner – I Want To Know What Love Is.mp3 :	1.59933
Guns N Roses – Welcome To The Jungle.mp3 :	2.22068
Zdravko Colic – Ti si mi u krvi.mp3 :	1.84687
Lady Gaga – Bad Romance.mp3 :	1.65483
Eminem – Rap God.mp3 :	2.07
Johannes Brahms – Hungarian Dance No. 5.mp3 :	4.3616
Miroslav Ilic – Bozanstvena zeno.mp3 :	2.83593
Queen – Dont Stop Me Now.mp3 :	1.33646
Ceca – Znam.mp3 :	3.216
Joe Cocker – You Can Leave Your Hat On.mp3 :	1.6826
Saban Saulic – Samo za nju.mp3 :	5.14189
Dubioza kolektiv – Rijaliti.mp3 :	1.27078
Beogradski Sindikat – Sistem Te Laze.mp3 :	1.55886
Djordje Balasevic – Lepa protina kci.mp3 :	3.61917
Rimsky-Korsakov – Flight Of The Bumblebee.mp3 :	3.61997
Bethoven – Fur Elise.mp3 :	2.1165
Robbie Williams and Nicole Kidman – Something Stupid.mp3 :	5.61598

Rezultati nakon promene pristupa prilikom generisanja reprodukcione liste pokazuju mnogo bolje rezultate nego što je to bilo sa prvobitnim računanjem rastojanja u odnosu na referentnu pesmu. Zbog test skupa, koji je prilično raznolik, postoje neznatni i приметni skokovi tokom kreiranja reprodukcione liste. Iz priloženog se vidi da se prag sličnosti nalazi negde u intervalu [2, 3.5], odnosno vrednost razdaljine koja više nego očigledno pravi razliku između dve muzičke numere. Treba napomenuti da je mera sličnosti u odnosu sa kojom se upoređuju rezultati, strogo subjektivna i zasniva se na sluhu i interpretaciji muzičkih numera od strane autora ovog rada.

4.5 Ocena mere sličnosti

Prethodni rezultati se tiču generisanja reprodukcione liste. Međutim u ovom koraku je potrebno pozabaviti se i samom ocenom mere sličnosti. Pošto je sama mera sličnosti izuzetno subjektivan pojam, odnosno dosta zavisi od slušalaca do slušalaca, mera sličnosti je posmatrana na nekoliko manjih grupa, od po pet muzičkih numera, gde su dve muzičke numere očigledno slične jedna drugoj i dosta različite u odnosu na ostale tri. U velikoj meri izražena sličnost dve pesme, odnosno različitost od ostalih, obezbeđuje nepristrasnost prilikom određivanja sličnosti. Razlog za to je već pomenuto subjektivno shvatanje muzičkih numera od strane čoveka.

Mera sličnosti je testirana na tri grupe od po pet pesama, kao što je već bilo reči. Rezultati dobijeni prilikom ovog testiranja su sledeći:

Grupa 1:

Chosen: Bryan Adams – Have You Ever Really Loved A Woman_.mp3	
Lady Gaga – Bad Romance.mp3	8.40734
Orthodox Celts – Far Away.mp3	9.64667
Djani – Ja slobodan ona neverna – (Audio 2008).mp3	9.85885
Toni Braxton – Un–Break My Heart (Video Version).mp3	6.89885
Bryan Adams – Have You Ever Really Loved A Woman_.mp3	0

Grupa 2:

Chosen: Bolero – Maurice Ravel.mp3	
Johann Strauss II – The Blue Danube Waltz.mp3	7.5346
Frank Sinatra Fly Me To The Moon.mp3	8.9684
Bolero – Maurice Ravel.mp3	0
Beogradski Sindikat – Sistem Te Laze.mp3	9.20094
Foo Fighters – The Pretender.mp3	8.30397

Grupa 3:

Chosen: Guns N Roses – Welcome To The Jungle.mp3	
Djordje Balasevic – Ne lomite mi bagrenje.mp3	7.40437
Ceca – Lepi grome moj.mp3	8.81831
Metallica – Enter Sandman.mp3	2.5572
Frank Sinatra – The Way You Look Tonight.mp3	8.6312
Guns N Roses – Welcome To The Jungle.mp3	0

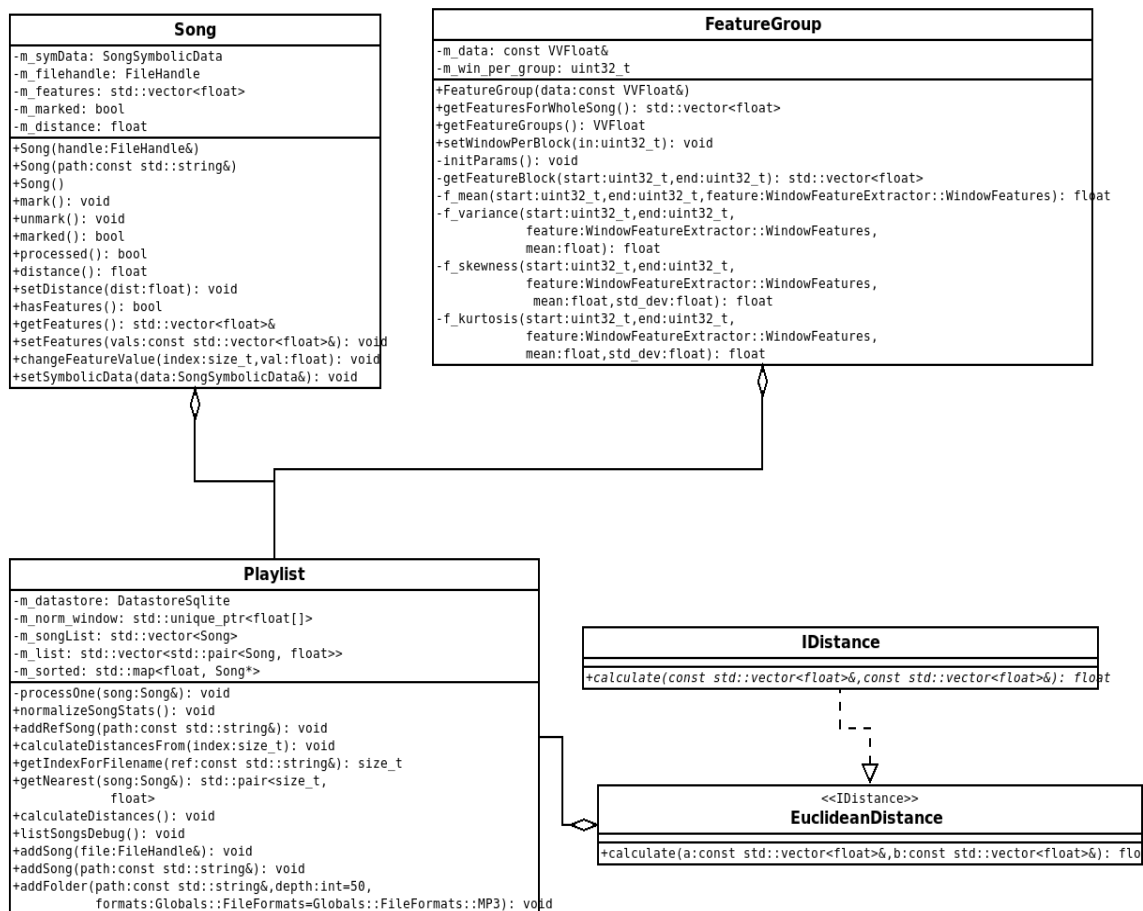
Sličnost između muzičkih numera grupisanih u istu grupu je očigledna. Na prvi pogled, mera sličnosti se pokazala odgovarajućom. Prilikom malo dublje analize rezultata, iako su rezultati zadovoljavajući, primećuje se razlika u brojevima koja nas navodi da mera sličnosti može biti nestabilna. Ovakav rezultat se dobija na manjem skupu analiziranih pesama jer je pristup normalizacije atributa takav da uzima u obzir sve vrednosti atributa iz celog skupa. Iako nije idealan, ovakav pristup će biti dovoljan za svrhu ovog rada i prikaza aproksimacije mere sličnosti u automatskom određivanju reprodukcioni lista.

U okviru grupe 1, dve najbližije pesme su od izvođača Bryan Adams i Toni Braxton, u grupi 2, Johann Strauss II i Maurice Ravel, a u grupi 3, Guns N Roses i Metallica. U sva tri slučaja sistem

je pokazao zadovoljavajući rezultat. Ovo međutim, ne mora da znači da će sistem uvek biti tačan, šta više, vrlo je moguće da će za različite muzičke numere, gde postoje dve očigledno slične, dati pogrešne rezultate, usled već pomenutog, ne baš najbolje izabranog načina za normalizaciju atributa. Za potrebe ovog rada, međutim, ovakav pristup se pokazao zadovoljavajućim. Iz priloženog se vidi da se mera sličnosti ponaša dovoljno stabilno i na manjim grupama muzičkih numera. Prilikom rasta skupa, mera sličnosti će biti stabilnija jer će se normalizacija prilagođavati dominantnijem delu muzičkih numera, dok će se način izračunavanja sličnosti pobrinuti da ne bude presudan faktor. Ovo u principu znači da će na rezultate, osim korisnika, uticati i sam skup podataka u celosti. Ovakav pristup, teoretski, može doneti i još jedan nevidljiv atribut prilikom izračunavanja mere sličnosti, jer na osnovu izabranog skupa pesama od strane korisnika, mogu se donekle odrediti i preference korisnika, u ovom slučaju, u određenoj meri, izraženi prilikom normalizacije vrednosti atributa.

4.6 Implementacioni detalji

Implementacija ponašanja navedenog u okviru ovog poglavlja se nalazi u sledećim komponentama `Playlist`, `Distances` i `feature_group`. `Playlist` implementira logiku iza formiranja reprodukcioni lista. Predstavlja ulaznu tačku u sistem i kao takva rukovodi celim tokom za obradu pesme i ekstraktovanje atributa. Osnovne funkcionalnosti same komponente su implementirane u potpunosti, dok dalja istraživanja otvaraju mogućnost razvoja novih funkcionalnosti. `Distances` predstavlja klasu za izračunavanje rastojanja između dve muzičke numere. Računanje rastojanja, iako na prvi pogled trivijalno, može biti krucijalno i komplikovano u zavisnosti odabrane mere sličnosti. U okviru ovog rada je korišćeno Euklidsko rastojanje, a u budućim nadogradnjama sistema je planirana implementacija i drugih rastojanja. `feature_group` komponenta koja je zadužena za izračunavanje statističkih parametara. Osim toga, ostavljena je mogućnost i da ova komponenta vrši i podelu atributa po grupama, što će takođe biti predmet buduće nadogradnje sistema. Umesto toga fokus je stavljen na izračunavanje atributa na celoj pesmi kao celini, efektivno umesto više grupa cela pesma je posmatrana kao jedna grupa. Dodatno postoji i `Song` komponenta koja predstavlja kontejner za sve podatke koje karakterišu jednu pesmu. U okviru ove klase implementirano je i čitanje simboličkih podataka kao potencijalan izvor dodatnih atributa u daljim koracima razvoja sistema.



Slika 4.1: Klasni dijagram za Playlist, FeatureGroup, Song i Distances komponente

Poglavlje 5

Zaključak

U okviru ovog rada opisane su osnovne tehnike potrebne za analizu audio zapisa i korišćenje dobijenih podataka u svrhu aproksimacije sličnosti dve muzičke numere. Korišćenjem metoda iz obrade signala i statističkih ocena, formirani su bazični atributi niskog nivoa pomoću kojih je realizovana mera sličnosti. Rezultat ovog rada je prateći sistem u potpunosti implementiran u C++, koji demonstrira opisane koncepte. Zamisao sistema je da automatski generiše reprodukcione liste koristeći isključivo podatke dobijene analizom samog audio zapisa, međutim u toku pisanja rada i razvoja sistema, prvobitna namena je blago modifikovana: proširene su određene komponente da bi se omogućilo lakše konfigurisanje i analiziranje rezultata, kao i brza i relativno lagana promena načina generisanja atributa i statističkih ocena. Razvijeni sistem zadovoljava osnovne zahteve postavljene na početku istraživanja, ali takođe ostavlja i prostor za mnoga dodatna unapređenja. Svaki deo sistema vezan za izračunavanje atributa i statističkih momenata pisan je sa idejom da u budućnosti bude proširen i adekvatno implementiran u tok obrade podataka. Softver otvorenog kôda razvijen tokom rada na ovoj tezi sadrži tok obrade podataka, koji je otvoren za dalje dopune i nadogradnje. Ostavljeno je dosta prostora za promene u samom načinu izračunavanja atributa, što kroz sam kôd, što kroz konfiguracionu dateoteku `config.json`. Mera sličnosti opisana u ovom radu je bazirana isključivo na atributima niskog nivoa. Ovakav pristup, iako nije idealan, pruža dovoljno dobre rezultate i predstavlja jaku osnovu za dalje korišćenje u dizajniranju kompleksnijih mera sličnosti. Fokus ovog rada je da demonstrira način kako se mera sličnosti dve muzičke numere može definisati i odrediti u nekoj meri i bez simboličkih podataka vezanih za samu muzičku numeru. Simbolički podaci se, naravno, ne mogu zanemariti u potpunosti i predstavljaju drugu stranu, odnosno, dopunu teoretski kompletne mere sličnosti. Budući koraci u razvoju sistema bi sigurno predstavljali uvođenje atributa srednjeg nivoa, uvođenje grafičkog interfejsa, formiranje koeficijenata koji bi uticali na samu pretragu u zavisnosti od korisnika, odnosno njegovih preferenca u atributima i slično.

Dodatak A

Prateće informacije

Sav kôd koji je nastao tokom pisanja ovog rada dostupan je na sledećim repozitorijumima:

- Sistem(kôd) - <https://bitbucket.org/Stevos/the-tick/src/master/>
- Prateći alati - <https://bitbucket.org/Stevos/tick-tools/src/master/>
- LaTeX projekat - <https://bitbucket.org/Stevos/the-tock/src/master/>

Sistem opisan u radu je pisan ceo u C++14, na operativnom sistemu Linux (Ubuntu 15.04 - 18.04). Biblioteke od kojih zavisi su ffmpeg, sqlite3 i json++. Sve biblioteke su već sadržane u okviru projekta. ffmpeg se dinamički linkuje, dok su sqlite3 i json++ uključene pomoću amalgamacije.¹ Potrebni alati za buildovanje su cmake, git i build-essentials. Potrebna verzija g++ je 7.2.0. Sistem se može konfigurisati detaljno kroz `config.json` datoteku, da omogući izbacivanje podataka tokom svake faze obrade muzičke numere. Takođe, postoji i logovanje izvršavanja u realtimeu. Logovi se smeštaju u datoteku `/alderaan.log`².

Prateći alati, iako nisu opisani u okviru ovog rada, predstavljaju sitne alatke pisane paralelno sa sistemom radi testiranja i vizuelizacije dobijenih podataka. Za kompilaciju alata opisanih ovde potreban je Qt5 i Matlab, uz standardne alatke prisutne na Linux operativnim sistemima.

LaTeX projekat su izvorne datoteke napravljene u svrhu generisanja samog master rada.

¹Amalgamacija je postupak uključivanja izvornih datoteka eksterne biblioteke kao deo aplikacije, gde se biblioteke kompajliraju zajedno sa aplikacijom. Ovakav pristup je uglavnom karakterističan za C i C++ jezike.

²Alderaan predstavlja prvobitni naziv projekta, dobijen po planeti iz (*StarWars*TM)

Literatura

- [1] R.W. Schafer A.V. Oppenheim. “From frequency to quefrequency: a history of the cepstrum”. In: *IEEE Signal Processing Magazine (Volume 21, Issue 5)* (2004).
- [2] Fingerhut M. Byrd D. “The history of ISMIR - a short happy tale”. In: *D-Lib Mag. 8(11)* (2002).
- [3] Burred J.J. Rodet X. Caetano M. “Automatic segmentation of the temporal evolution of isolated acoustic musical instrument sounds using spectro-temporal cues.” In: *In: Proceedings of the 13th International Conference on Digital Audio Effects (DAFx), Graz* (2010).
- [4] Oscar Celma. “Foafing the Music: Bridging the semantic gap in music recommendation”. In: (2006).
- [5] Tukey J.W. Cooley J.W. “An algorithm for the machine calculation of complex Fourier series”. In: *Math. Comput. 19(90), 297–301* (1965).
- [6] J.S. Downie. “The scientific evaluation of music information retrieval systems: foundations and future.” In: *Comput. Music J. 28, 12–23* (2004).
- [7] Downie J.S. Futrelle J. “Interdisciplinary research issues in music information retrieval”. In: *ISMIR 2000–2002. J. N. Music Res. 32(2), 121–131* (2003).
- [8] Gordon J.W. Grey J.M. “Perceptual effects of spectral modifications on musical timbres.” In: *J.Acoust. Soc. Am. 63(5), 1493–1500* (1978).
- [9] J. Stephen Downie Joe Futrelle. “Interdisciplinary Communities and Research Issues in Music Information Retrieval.” In: *Conference: ISMIR 2002, 3rd International Conference on Music Information* (2002).
- [10] P.N. Juslin. “What does music express? Basic emotions and beyond”. In: (2013).
- [11] Anssi P. Klapuri. “Multiple fundamental frequency estimation based on harmonicity and spectral smoothness.” In: *IEEE Transactions on Speech and Audio Processing, volume 11, pages 804–816* (2003).
- [12] A. Lerch. “An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics.” In: *Wiley-IEEE Press, Hoboken* (2012).
- [13] Sebe N. Djeraba C.-Jain R. Lew M.S. “Content-based multimedia information retrieval: state of the art and challenges.” In: *ACM Trans. Multimedia Comput. Commun. Appl. 2(1), 1–19* (2006).
- [14] Lu Wei Li Hu Bei-Qian Dai. “A Pitch Detection Algorithm Based on AMDF and ACF”. In: *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* (2006).
- [15] A. Marsden. “Music similarity. In: Workshop on Music Similarity: Concepts, Cognition and Computation.” In: *Lorentz Center, Leiden. <http://www.lorentzcenter.nl/lc/web/2015/669/presentations/Marsden.pptx>* (2015).
- [16] Richard Kronland-Martine Solvi Ystad Mathieu Barthet Philippe Guillemain. “ON THE RELATIVE INFLUENCE OF EVEN AND ODD HARMONICS IN CLARINET TIMBRE”. In: *International Computer Music Association* (2005).

- [17] Brian McFee. “More like this: machine learning approaches to music similarity”. In: *A dissertation submitted in partial satisfaction of the requirements for the degree* (2012).
- [18] Richard Sproat Mehryar Mohri Michael Riley. “Algorithms for Speech Recognition and Language Processing”. In: *COLING96* (1996).
- [19] Geoffroy Peeters. “A large set of audio features for sound description (similarity and classification) in the CUIDADO project.” In: *Technical report, Ircam, Analysis/Synthesis Team, 1 pl. Igor Stravinsky* (2004).
- [20] Markus Schedl Peter Knees. “Music Similarity and Retrieval: An introduction to Audio- and Web-based Strategies”. In: (2016).
- [21] Slaney M. Ponceleón D. “Multimedia information retrieval.” In: *Baeza-Yates, R., Ribeiro-Neto, B. (eds.) Modern Information Retrieval – The Concepts and Technology Behind Search, Chap. 14, 2nd edn., pp. 587–639. Addison-Wesley, Pearson* (2011).
- [22] Wolfe J. Tarnopolsky A. Schubert E. “Spectral centroid and timbre in complex, multiple instrumental textures.” In: *In: Proceedings of the 8th International Conference on Music Perception and Cognition (ICMPC), Evanston* (2002).
- [23] C. Shannon. “Communication in the presence of noise.” In: (1949).
- [24] M. Slaney. “Web-scale multimedia analysis: Does content matter?” In: *IEEE MultiMedia 18(2), 12–15* (2011).
- [25] B.L. Sturm. “Classification accuracy is not enough.” In: *J. Intell. Inf. Syst. 41, 371–406* (2013).
- [26] Roland B. Valentin E. and Bertrand D. “Multipitch Estimation of Piano Sounds Using a New Probabilistic Spectral Smoothness Principle”. In: *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 18, NO. 6, (2010)*.
- [27] G.A. Wiggins. “Semantic gap?? Schemantic schmap!! Methodological considerations in the scientific study of music”. In: *Proceedings of the 11th IEEE International Symposium on Multimedia (ISM), San Diego* (2009).
- [28] *Window function*. https://en.wikipedia.org/wiki/Window_function. Accessed: 2018-09-15.
- [29] Mak M.W. Li C.K. Yiu K.K. “Gaussian mixture models and probabilistic decision-based neural networks for pattern classification: a comparative study.” In: *Neural Comput. Appl. 8, 235–245* (1999).

Lista slika

1.1 Četiri različite kategorije koje utiču na shvatanje muzike	7
1.2 Različiti nivoi apstrakcije atributa i semantički raskorak između njih	9
2.1 Greška kvantizacije. Analogni signal je predstavljen crvenom bojom, dok je plavom predstavljen odgovarajući digitalni signal.	11
2.2 Prikaz primene Hannove prozorske funkcije na signal	12
2.3 Bartletova ili trougaona prozorska funkcija [28]	13

2.4	Welchova prozorska funkcija [28]	13
2.5	sinusna prozorska funkcija [28]	14
2.6	Pravougaona prozorska funkcija [28]	14
2.7	Hammingova prozorska funkcija [28]	14
2.8	Prikaz ulaznog signala u vremenskom domenu i prikaz rezultata DFTa (frekvencijski domen)	16
2.9	Klasni dijagram za komponente vezane za obradu audio signala i čuvanje relevantnih podataka.	17
3.1	Prikaz dijagrama za frekvenciju prelaska nule. Prvi dijagram je audio signal, drugi dijagram predstavlja dijagram frekvencije prelaska nule za dat audio signal.	20
3.2	Prikaz dijagrama autokorelacije prvog reda na primeru snimka otkucaja srca	21
3.3	Uprošćeni prikaz maksimalne vrednosti amplitude u odnosu na audio signal.	22
3.4	Prikaz srednjekvadratne energije, po okvirima, za različite muzičke žanrove	23
3.5	Vrednosti spektralnog centroida Kafana na Balkanu - Aca Lukas (plavo) i Highway to hell - ACDC (crveno)	24
3.6	Glatkoća zvuka u prvih 10ak sekundi pesme Highway to hell. Podaci su normalizovani u odnosu na sebe na interval [-1,1]	25
3.7	Vrednosti spektralne asimetrije Kafana na Balkanu - Aca Lukas (plavo) i Highway to hell - ACDC (crveno).	26
3.8	Klasni dijagram implementacionih detalja za računanje atributa. Samo jedna komponenta je odgovorna za ovo stoga samo jedna klasa	27
4.1	Klasni dijagram za Playlist, FeatureGroup, Song i Distances komponente	39