SADRŽAJ:

**Napomena: Uz rad je priložen CD.

1. UVOD

Stvaranje ogromnih digitalnih muzičkih audio baza podataka dolazi usljed digitalizacije postojećih analognih arhiva i potrebe arhiviranja novih sadržaja, što zahtijeva sve pouzdaniji i brži alat za analizu i deskripciju sadržaja, koji će omogućiti pretraživanje, ispitivanje sadržaja i interaktivni pristup. U tom kontekstu, muzički žanrovi su krucijalne kategorije jer su godinama široko korišteni za organizovanje muzičkih kataloga, biblioteka i muzičkih kolekcija i radnji. Uprkos njihovoj primjeni, muzički žanrovi nisu dovoljno definisani, što problem automatske klasifikacije muzičkih zapisa čini netrivijalnim.

Muzički žanrovi su takođe najvažniji deskriptori korišteni i od strane muzičkih dilera i kolekcionara za organizaciju njihovih muzičkih kolekcija. Oni predstavljaju pojednostavljen umjetnički muzički izraz i interesantni su kao skup zajedničkih karakteristika muzičkih zapisa.

Sa početkom elektronske distribucije muzike, muzički katalozi teže da postanu ogromni (najveći online servisi nude 1 milion zapisa). U tom kontekstu, udruživanje žanra sa muzičkim zapisom je veoma bitno kao pomoć korisniku da nađe ono što traži. Zapravo, količina digitalne muzike podstiče na razmišljanje o efikasnijem načinu otvaranja, orgnizovanja i dinamičkog obnavljanja muzičkih kolekcija. To definitivno zahtijeva novi pristup u konstrukciji muzičkog sadržaja koji će biti automatizovan. U slučaju klasifikacije muzičkog žanra, Weare[1] izvještava da je za manuelno klasifikovanje 100,000 kompozicija za Microsoft's MSN muzički pretraživač potrebno oko 30 muzikologa godišnje.

U isto vrijeme, termini kao što su jazz, rock ili pop, iako široko korišteni, često bivaju slobodno definisani, tako da problem automatske klasifikacije žanra postaje netrivijalan problem.

Muzički žanr

Muzički žanrovi su kategorije, kreirane od strane čovjeka, koje su nastale kroz kompleksno međusobno djelovanje kulture, umjetnosti i marketinga da bi se okarakterisale sličnosti između muzičara ili kompozicija, kao i da bi se organizovale muzičke kolekcije. Muzički žanr nema striktne definicije ni granice. Kako god, jasno je da muzički žanrovi međusobno dijele određene karakteristike koje su tipične za instrumentaciju, ritmičku strukturu i tonski sadržaj muzike.

Problem automatske klasifikacije muzičkih audio zapisa zahtijeva eksplicitno definisanje žanrova, tj. hijerarhijski set kategorija koji će biti preslikan na muzičku kolekciju. U nekim istraživanjima [2] o broju žanrovskih kategorija koje se koriste u muzičkoj industriji pokazano je da nije jasno izgrađena takva hijerarhija žanrova.

Izvođač, album ili kompozicije?

Jedno osnovno pitanje na koje treba dati odgovor glasi: Na koje dijelove muzike treba primjeniti žanr klasifikaciju: na kompozicije, album ili izvođača? Ako pretpostavimo da jednu pjesmu možemo klasifikovati u samo jedan žanr, to više nije tako jednostavno za jedan album, jer on može biti višežanrovski materijal. Isto važi i za izvođača. Neki izvođači pokrivaju širok spektar žanrova tokom karijere i nema smisla svrstavati ih u jednu specifičnu klasu.

Neslaganje oko taksonomije (definisanja žanrova)

Pachet and Cazaly [3] u svojim istraživanjima kažu da generalno ne postoji sporazum o taksonomijama žanra koji bi se poštovao u praksi. Uzimajući kao primjer dobro poznate Web sajtove, kao što su Allmusic (http://www.allmusic.com – sadrži 531 žanr), Amazon (http://www.amazon.com – 719 žanrova), Mp3 (http://mp3.com - 430 žanrova), oni pronalaze samo 70 termina koji su zajednički za sve tri taksonomije. Oni takođe primjećuju da široko korišteni termini kao što su *rock* i *pop* označavaju različit skup kompozicija i te hijerarhije žanrova su različito organizovane od jedne do druge taksonomije.

Loše definicije žanrova

Ako se bliže pogledaju neki specifični i široko korišteni muzički žanrovi, može se vidjeti koliko je različit kriterij definisanja specifičnosti žanra. Kao na primjer:

- Indijska muzika (*Indian music*) ja geografski definisana,
- Barokna muzika (*Baroque music*) je povezana sa jednim istorijskim razdobljem koje uključuje različite stilove i širok geografski region,
- Post-rok (*Post-rock*) je termin izmišljen od strane kritičara *Simon Reynolds*a.

Pachet i *Cazaly* uvjeravaju da ova semantička zabuna između pojedinih taksonomija može da vodi u redundantnost, koja možda neće biti smetnja za ljudski faktor, ali će automatskim sistemima biti veoma otežavajuća okolnost. Pored toga, taksonomija može da zavisi i od kulturnih vrijednosti. Na primjer, pjesme francuskog pjevača *Charles Aznavour*a[1] mogu se različito tumačiti u Francuskoj, ali će u Velikoj Britaniji biti svrstane u *world music*.

Skalabilnost taksonomije žanrova

Hijerarhije žanrova takođe trebaju razmatrati mogućnosti dodavanja novih žanrova na račun muzičke evolucije. Novi žanrovi se učestalo pojavljuju i oni su rezultat djelimične ili potpune integracije različitih žanrova (npr. *psychobily* je mješavina *rockabily*a i *punk*a) ili posljedica razdvajanja žanrova na podžanrove (npr. *hip-hop* se dijeli na *gangsta rap*, *turntablism* i *conscious rap*). Ovo je važan rezultat za automatske sisteme. Dodavanje novih žanrova i podžanrova u taksonomiju je jednostavno, ali zahtijeva automatski sistem sa nadgledanim treningom sposoban da se sam adaptira na novonastale promjene.

Sve prethodno rečeno ukazuje na to, koliko je problem automatske klasifikacije muzičkih audio zapisa zaista problem, odnosno, na koja sve pitanja treba odgovoriti, u smislu preprocesiranja, da bi se automatska klasifikacija omogućila. Takođe, da bi se omogućila klasifikacija potrebno je audio signal predstaviti na način pogodan za obučavanje i korištenje klasifikatora. Originalni audio signal nije pogodan u ovu svrhu zbog velike redundanse koju sadrži, a koja rezultuje i velikim memorijskim zahtjevima, odnosno velikim zahtjevima za procesnu moć računara. Dakle, neophodno je signal predstaviti pomoću određenih obilježja, veličina koje oslikavaju određene karakteristike signala bilo u vremenskom, bilo u transformacionom domenu. Izdvojena obilježja se zatim koriste za obučavanje klasifikatora. Za klasifikaciju se uglavnom koriste statistički klasifikatori, a rjeđe klasifikatori zasnovani na pravilima (ekspertni sistemi).

Metode automatske klasifikacije audio zapisa vode porijeklo iz prepoznavanja govora i imaju dužu istoriju. Mel cepstralni koeficijenti (MFCC[2]) su skup perceptualno motivisanih obilježja koja su često korištena u prepoznavanju govora. U radovima [4,5] predstavljeno je klasifikovanje audio zapisa na govorne i muzičke. Takođe, u radovima [4,6], audio signal je segmentiran i klasifikovan na muziku, govor, smijeh i bezgovorne signale korištenjem cepstralnih koeficijenata i skrivenih Markovljevih modela (HMMs[3]). U radu [7] prikazano je klasifikovanje muzike primjenom vještačkih neuronskih mreža (ANNs[4]).

Automatska klasifikacija audio zapisa, kao dio sistema za semantičko pretraživanje multimedije, svoju popularnost i značaj doživila je zahvaljujući velikoj ekspanziji audio zapisa na Webu (uskoro će svi muzički audio zapisi u ljudskoj istoriji biti dostupni na Webu [4]), a sve to kroz potrebu velikih kompanija da ostvare profit, za uzvrat pružajući odgovarajuće usluge korisnicima.

---

[1] Kod nas bi smo mogli uzeti primjer benda Sanja Ilić & Balkanika ili Mostar Sevdah Reunion
[2] MFCC-Mel Frequency Cepstral Coeficients
[3] HMMs-Hiden Markov Models
[4] ANNs-Artifical Neural Networks

Sve ovo, kao i činjenica da će u dogledno vrijeme pojavom mreža nove generacije (NGNs[5]) doći do globalnog umrežavanja (u isto vrijeme i kontrole), čini problem automatskog prepoznavanja i klasifikovanja značajnim.

U ovom radu se razmatra mogućnost klasifikacije muzičkih audio zapisa na žanrove, korištenjem niza obilježja izdvojenih kako iz reprezentacije signala u vremenskom, tako i u frekvencijskom domenu. Dat je teorijski pregled pojedinih korištenih obilježja u drugoj glavi, zatim teorijski opis popularnih metoda klasifikacije uzoraka u trećoj glavi, te u eksperimentalnom dijelu, koji je opisan u četvrtoj glavi, detalji implementacije i rezultati klasifikacije. Peta glava sadrži zaključak, a predposljednja, šesta glava, listu korištenih skraćenica. U posljednjoj glavi rada dat je spisak korištene literature, a uz rad je priložen i CD na kojem se nalazi rad u elektronskom obliku, kao i sva prateća dokumentacija.

[5] NGNs-New Generation Networks

2. PREGLED OBILJEŽJA

U ovoj glavi dat je teorijski pregled odabranih obilježja za klasifikaciju muzičkih audio zapisa. Da bi se mogla izvršiti klasifikacija audio zapisa potrebno ga je predstaviti na način koji je pogodan za obučavanje i korištenje klasifikatora. Originalni audio signal nije pogodan u ovu svrhu zbog velike redundanse koju sadrži, a koja rezultuje i velikim memorijskim zahtjevima za smještanje odmjeraka signala, odnosno velikim zahtjevima za procesnu moć računara. Dakle, neophodno je signal predstaviti pomoću određenih obilježja, veličina koje oslikavaju određene karakteristike signala bilo u vremenskom, bilo u transformacionom, npr. frekvencijskom, domenu. Postupak izdvajanja obilježja iz signala naziva se indeksiranje signala. Izdvojena obilježja se sada koriste za obučavanje klasifikatora, a klasifikacija novih signala se vrši na osnovu njihovih obilježja izdvojenih korištenjem iste procedure.

2.1 Obilježja teksture (*Timbral Texture Features*)

Zvučni signali spadaju u grupu nestacionarnih signala, tj. njihove spektralne karakteristike se mijenjaju u vremenu. Zbog toga se analiziraju na kratkim vremenskim intervalima (*frame*-ovima). Ukoliko je interval analize dovoljno kratak može se smatrati da je signal u njemu stacionaran i parametri signala su konstantni na tom intervalu. Ovaj vremenski interval naziva se *prozor analize.* Za zvučne signale kao što su govor i muzika obično se uzima da je trajanje prozora analize dvadesetak milisekundi. Kada se zvučnom signalu intervali sa različitim spektralnim karakteristikama izmjenjuju sa određenom pravilnošću, možemo govoriti o zvučnoj teksturi.

Da bi se ova pojava kvantitativno ispitala neophodno je signal posmatrati na većem intervalu koji se naziva *prozor teksture*. Prozor teksture se sastoji od više prozora analize i njegovo trajanje je oko jedne sekunde. Istraživanja na ljudskim subjektima su pokazala da je čovjeku za prepoznavanje muzičkog žanra potrebno svega tri sekunde muzičkog zapisa [4]. Iz ovoga se dolazi do zaključka da čovjek za prepoznavanje muzičkog žanra koristi, pored drugih karakteristika audio signala, i upravo opisanu muzičku teksturu. Da bi se muzička tekstura kvantitativno opisala koriste se sljedeća obilježja, zasnovana na spektralnim karakteristikama signala:

- **Spektralni centroid** se izračunava za svaki prozor analize (*frame*) i predstavlja centar mase amplitudnog spektra tog prozora određenog pomoću kratkotrajne Furijeove transformacije (STFT-Short Time Fourier Transform). Matematički ovo se može iskazati kao:

$$C_t = \frac{\sum_{k=1}^{N} k \cdot M_t(k)}{\sum_{k=1}^{N} M_t(k)}\,, \tag{2.1}$$

  gdje indeks t označava prozor analize, a $M_t(k)$ je vrijednost amlitudnog spektra prozora t za k-tu diskretnu frekvenciju. U daljem tekstu ćemo pod pojmom diskretna frekvencija smatrati indeks diskretne Furijeove transformacije. Veća vrijednost ovog obilježja ukazuje na veći udio visokih frekvencija u spektru signala u prozoru analize. Prozori muzičkog signala imaju veću vrijednost spektralnog centroida od prozora govornog signala zato što muzički instrumenti proizvode tonove viših frekvencija od ljudskog glasa. Takođe, vrijednosti spektralnog centroida su različite za zvučni i bezvučni govor.

- **Spekralni rolloff** predstavlja diskretnu frekvenciju R_t ispod koje se nalazi 85% raspodjele magnituda signala, tj.

$$\sum_{k=1}^{R_t} M_t(k) \approx 0.85 \cdot \sum_{k=1}^{N} M_t(k)\,. \tag{2.2}$$

  Vrijednost ovog obilježja je veća ukoliko je više energije signala sadržano u visokim frekvencijama. Veći dio energije bezvučnog govora i muzike sadržan je na nižim frekvencijama, dok je kod zvučnog govora više energije sadržano na višim frekvencijama pa zvučni govor ima veću vrijednost ovog obilježja od bezvučnog govora i muzike.

- **Spektralni fluks** odražava promjenu spektra između dva susjedna prozora analize. Izračunava se kao suma kvadrata razlika normalizovanih magnituda signala u dva susjedna prozora:

$$F_t = \sum_{k=1}^{N} \left(N_t(k) - N_{t-1}(k)\right)^2 , \tag{2.3}$$

gdje je $N_t(k)$ normalizovana magnituda signala u prozoru *t*, a $N_{t-1}(k)$ normalizovana magnituda signala u prethodnom prozoru *t-1*. Magnitude u svakom prozoru se normalizuju zbirom magnituda signala na svim frekvencijama za dati prozor. Ovo obilježje označava dinamiku promjene spektra signala. Naravno, muzički signal se brže mijenja od govornog i ima veću vrijednost ovog obilježja.

- **Broj prolazaka kroz nulu** je obilježje koje se izračunava u vremenskom domenu. Njegova vrijednost je broj prolazaka signala kroz nulu na datom prozoru. Matematički,

$$Z_t = \frac{1}{2} \sum_{m=1}^{M} \left|\mathrm{sgn}(x(m)) - \mathrm{sgn}(x(m-1))\right|, \tag{2.4}$$

gdje je $x(n)$ signal u prozoru *t,* a *M* dužina tog prozora. Bezvučni govor ima višu vrijednost ovog obilježja od zvučnog govora. Pošto se u govornom signalu smjenjuju intervali zvučnog i bezvučnog govora to znači da se smjenjuju i intervali sa velikom i malom vrijednošću ovog obilježja. Sa druge strane broj prolazaka kroz nulu na jednom prozoru je kod muzičkog signala relativno konstantan.

- **Prozori sa niskom energijom** su prozori analize čija je RMS energija manja od prosječne RMS energije u jednom prozoru teksture. Ukoliko signal ima veći broj " tihih " prozora analize vrijednost ovog obilježja će biti veća. Veći broj " tihih " prozora analize karakterističan je za govorni signal. Kao obilježje se uzima procentualno učešće ovih prozora u ukupnom broju prozora analize signala.

- **Mel-skalirani cepstralni koeficijenti (MFCC)** su obilježja motivisana ljudskom percepcijom audio signala i često se koriste za modeliranje u sistemima za prepoznavanje govora. Da bi se odredili MFCC, signal se propušta kroz banka filtar čije su centralne frekvencije uniformno raspoređene na logaritamski transformisanoj frekvencijskoj osi. Razlog za ovo su eksperimenti na ljudskim subjektima koji su pokazali da uho frekvenciju zvučnih signala opaža na logaritamskoj skali. Takođe je pokazano da postoje određeni opsezi frekvencija, *kritični opsezi*, unutar kojih nije moguće razlikovati frekvencije zvukova.

U radu je iskorišten ISP (Intelligent sound implementation) model realizacije MFCC-a [8]. Najprije se signal podijeli na kratkotrajne prozore dužine N na kojima se izračunava diskretna Furijeova transformacija (DFT) po jednačini:

$$X(k) = \sum_{n=0}^{N-1} w(n)x(n)\exp(-j2\pi kn/N), \tag{2.5}$$

za *k=0,1,...,N-1*, gdje k odgovara frekvenciji $f(k) = k \cdot f_S / N\,[Hz]$, pri čemu je f_S frekvencija odmjeravanja u *Hz* a *w(n)* posmatrani prozor. Za vremenski prozor odabran je Hammingov[6] prozor zbog njegovih optimalnih karakteristika pri izračunavanju STFT-a.

Sada se magnituda *X(k)* skalira po frekvenciji i po amplitudi. Po frekvenciji se logaritamski skalira korištenjem tzv. Mel banka filtra *H(k,m)* , a zatim se skalira po amplitudi prirodnim logaritmiranjem.

$$X'(m) = \ln\left(\sum_{k=0}^{N-1} |X(k)| \cdot H(k,m)\right), \tag{2.6}$$

za *m=1,2,...M,* gdje je M broj filtara u filtar banci i $M<<N$. Mel banka filtar je kolekcija filtara čije su amplitudne karakteristike trougaonog oblika sa centralnim frekvencijama $f_c(m)$, date sa:

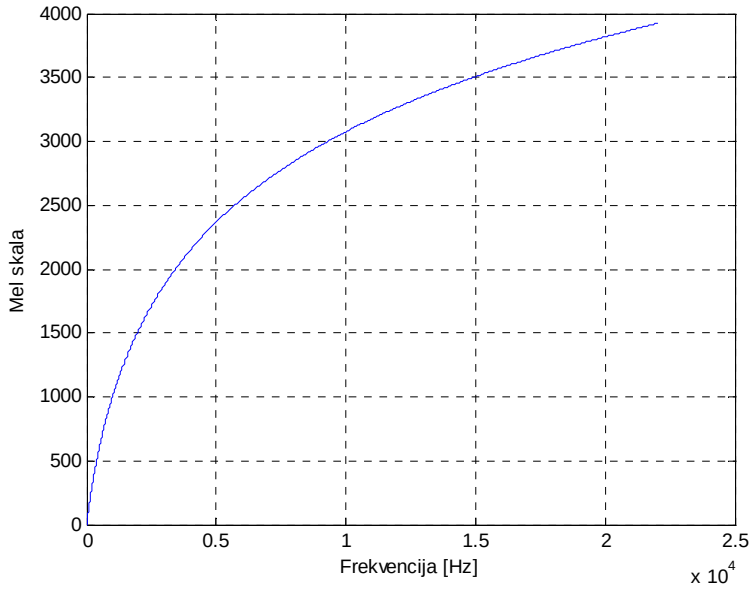
$$H(k,m) = \begin{cases} 0 & \text{za } f(k) < f_c(m-1) \\ \dfrac{f(k) - f_c(m-1)}{f_c(m) - f_c(m-1)} & \text{za } f_c(m-1) \le f(k) < f_c(m) \\ \dfrac{f(k) - f_c(m+1)}{f_c(m) - f_c(m+1)} & \text{za } f_c(m) \le f(k) < f_c(m+1) \\ 0 & \text{za } f(k) \ge f_c(m+1) \end{cases}. \tag{2.7}$$

Za logaritamsku transformaciju frekvencijske ose koristi se Mel frekvencija koja je sa frekvencijom u hercima povezana jednačinom:

$$\phi = 2595 \log_{10}\left(1 + \frac{f[Hz]}{700}\right). \tag{2.8}$$

[6] Slabljenje glavnog luka Hammingovog prozora iznosi -46dB, a slabljenje bočnih lukova opada sa -6dB/oktavi

Na Slici 2.1 prikazana je zavisnost Mel frekvencije od frekvencije u Hz po prethodnoj jednačini:



***Slika 2.1**-Zavisnost Mel frekvencije od frekvencije u Hz*

Frekvencijska rezolucija u Mel skali data je sa:

$$\Delta\phi = \left(\phi_{\max} - \phi_{\min}\right)/(M+1)\ , \tag{2.9}$$

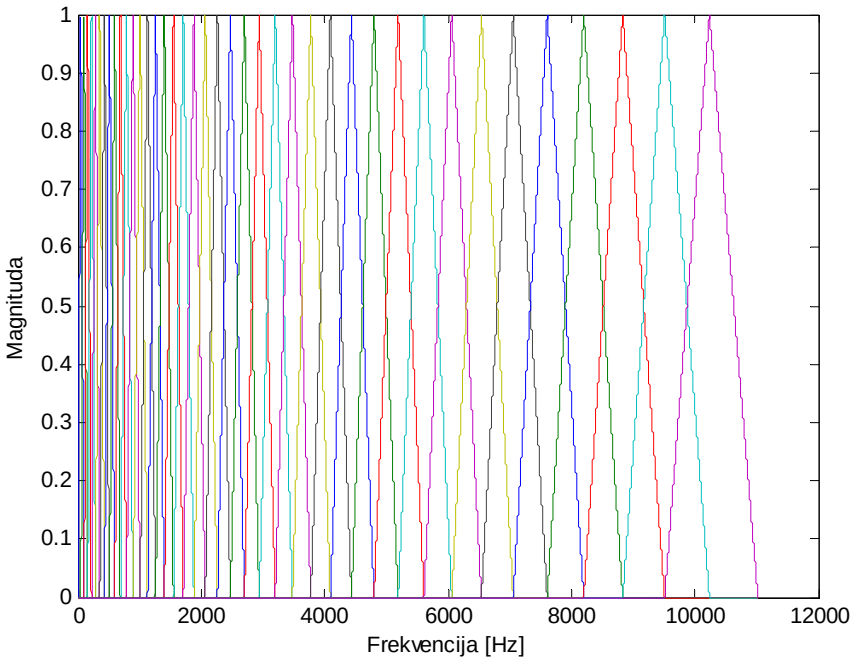
gdje je $\phi_{\max}$ maksimalna frekvencija filtar banke u Mel skali, izračunata iz $f_{\max}$ pomoću jednačine (2.8) , a $\phi_{\min}$ minimalna frekvencija filtar banke u Mel skali, izračunata iz $f_{\min}$ pomoću jednačine (2.8). $f_{\min}$ i $f_{\max}$ su minimalna i maksimalna frekvencija date u *Hz*. U radu je uzeto $f_{\min} = 0$ i $f_{\max} = f_s = 22050\ [Hz]$. Centralne frekvencije u Mel skali date su sa $\phi_c(m) = m \cdot \Delta\phi$ za $m=1,...,M$. Da se dobiju centralne frekvencije u *Hz* primjeni se inverzno jednačina (2.8), $f_c(m) = 700 \cdot \left(10^{\phi_c(m)/2595} - 1\right)$, i uvrsti u izraz (2.7) da se dobije Mel banka filtar.

Na kraju, da bi se smanjila dimenzionalnost i korelacija između obilježja izračunava se diskretna kosinusna transformacija (DCT-*Discrete Cosine Transform*) od *X'(m)* :

$$c(l) = \sum_{m=1}^{M} X'(m) \cdot \cos\left(l \cdot \frac{\pi}{M}(m - 0.5)\right), \tag{2.10}$$

za $l=1,...,M$, gdje *c(l)* predstavlja *l*-ti MFCC koeficijent.

Na Slici 2.2 je prikazana frekvencijska karakteristika jedanog banka filtra koji se sastoji od 40 propusnika opsega čije su centralne frekvencije linearno raspoređene na mel-trasformisanoj frekvencijskoj osi.



***Slika 2.2**-Frekvencijska karakteristika banka filtra*

Većina opisanih obilježja su vremenski promjenjiva tj. njihova vrijednost se razlikuje u pojedinim prozorima analize u kojima se smatra da je zvučni signal stacionaran. Spektralni centroid, spektralni rolloff, spektralni fluks, broj prolazaka kroz nulu i MFCC se računaju za svaki prozor analize signala. Ova obilježja izračunata za svaki prozor analize mogu poslužiti kao osnova za klasifikator koji bi radio u realnom vremenu. Međutim, ako se projektuje klasifikator koji koristi čitav raspoloživi signal potrebno je izračunati globalna obilježja koja predstavljaju čitav signal. Da bi se ovo postiglo koriste se srednja vrijednost i varijansa obilježja na prozoru teksture. Na ovaj način modelirane su prosječna vrijednost obilježja i mjera odstupanja stvarnih vrijednosti od te prosječne vrijednosti. Ovako dobijena obilježja vrijede na pojedinim prozorima teksture. Sa druge strane, procenat prozora sa niskom energijom se izračunava na prozoru teksture i vrijednost ovog obilježja se dodaje u vektor obilježja za pojedini prozor teksture. Čitav signal se sada opisuje jedinstvenim vektorom obilježja koji predstavlja srednju vrijednost opisanih vektora obilježja za prozore teksture.

2.2 Obilježja ritma (*Rhythmic Content Features*)

Ritam je, uopšte, najsamostalnija – tačnije, jedina samostalna kategorija u muzici. Drugi njeni činioci – melodija, harmonija, oblik, tonalitet – nužno podrazumijevaju i podlogu nekakvog ritma, samim tim što se ispoljavaju u vremenu, a ritam je taj u kome se iskazuje muzičko vrijeme! On može da se ispoljava i da dejstvuje, u muzičkom smislu, sam za sebe-na primjer, u zvuku neke udaraljke neodređene zvučne visine (razumije se, uz izvjesnu jačinu i boju tog zvuka, jer su te osobine neodvojive od same pojave zvuka, kao takvog). Po nekim teorijama o porijeklu muzike, ritam je i njeno izvorište, prvobitni polazni činilac (" Na početku biješe ritam", Hans von Bulow, 1830-1894 [9]) – ako je suditi upravo po njegovoj veoma čestoj i bogatoj samostalnoj primjeni u muzici najprimitivnijih naroda. Elementarnost ritma ispoljava se i u njegovom najneposrednijem i najdubljem dejstvu na čovjeka svakog vremena i podneblja – o čemu naročito vidljivo svjedoči uticaj savremene popularne muzike na masovno slušalište. U svakom slučaju, ritam čini samu osnovu svakog muzičkog zbivanja, svojevrsnu "kičmu" i njeno vremensko "bilo", bez kojeg ona ne može ni da postoji[9]!

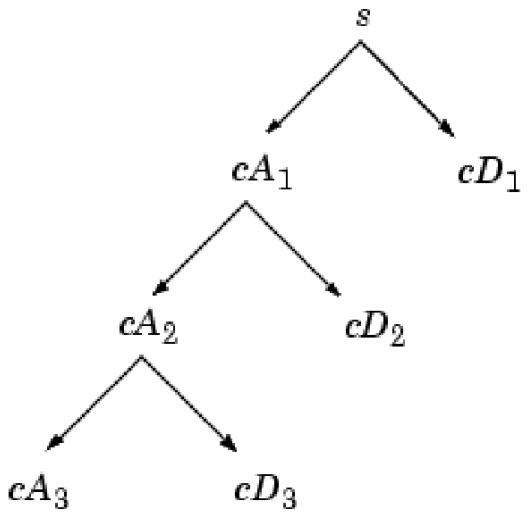
Iako je ritam kao muzički pojam jednostavan za shvatiti, nije ga jednostavno definisati. Ljudska percepcija ritma je subjektivan doživljaj (postoje i ljudi bez naročito izraženog osjećaja za ritam), ali u osnovi je ritam uvijek opisivan kao ponavljanje naglašenih elemenata ili cijelih segmenata unutar kompozicije. Pravilnost ritma, veza između osnovnog, tj. glavnog bita (eng. *beat*) i sporednih bita, tj. harmonika i relativna jačina sporednih bita i glavnog bita su karakteristike, odnosno, obilježja koja želimo predstaviti u karakterističnom vektoru obilježja. Da bi se došlo do vektora obilježja, potrebno je pethodno izvršiti detekciju bita, kao i konstruisati Beat Histogram (BH). Bit predstavlja pojam koji je usko vezan za ritam i tempo. Dok ritam uključuje kompleksnu strukturu, a tempo s druge strane krajnje jednostavnu, bit čini vezu između njih. Dakle, bit predstavlja skup pojedinačnih tempa cjelokupne instrumentacije kompozicije koji, konačno, tvore ritam.

2.2.1 Algoritam za detekciju bita

Jedan od uobičajenih algoritama za automatsku detekciju bita sastoji se iz dekompozicije signala pomoću banke filtara na podopsege (oktave), koja je praćen izdvajanjem anvelope signala podopsega i algoritmom koji se koristi za detekciju vremenskih perioda u kojima je anvelopa signala najsličnija samoj sebi. Izdvajanje obilježja za reprezentaciju ritmičkog sadržaja iz audio signala bazirana je na Wavelet transformaciji (*Wavelet Transform*-WT) koja predstavlja tehniku za analizu signala koja je razvijena kao alternativa kratkotrajnoj Furijeovoj transformaciji (STFT) da se prevaziđu problemi sa rezolucijom. Tačnije, kratkotrajna Furijeova transformacija omogućuje jednaku rezoluciju u vremenu za sve frekvencije, tj. prozor kroz koji se posmatra signal u vremenu je iste dužine za sve frekvencije, dok je kod Wavelet transformacije prozor promjenjive dužine, odnosno, visoka rezolucija u vremenu a niska u frekvenciji za visoke frekvencije (prozor male dužine) i niska rezolucija u vremenu i visoka u frekvenciji za niske frekvencije (prozor velike dužine).

Diskretna Wavelet transformacija (*Discrete Wavelet Transform*-DWT) je specijalan slučaj Wavelet transformacije koja daje kompaktnu reprezentaciju signala u vremenu i frekvenciji, i koja može biti uspješno izračunata korištenjem brzog piramidalnog algoritma dekompozicije pomoću banke filtara. Više o WT i DWT može se pronaći u [10] .

DWT je u ovom radu upotrebljena kao tehnika dekompozicije signala na oktave u frekvencijskom domenu. Tačnije, centralne frekvencije propusnih opsega banke filtara se razlikuju za jednu oktavu (pomnožene faktorom 2). U piramidalnom algoritmu dekompozicije, signal je analiziran u različitim frekvencijskim opsezima sa različitom rezolucijom za svaki opseg. Ovo je postignuto dekompozicijom signala na grubu aproksimaciju (cA_k) i detalje (cD_k), zatim se ponovo vrši dekompozicija aproksimacije na novu aproksimaciju i detalje u sljedećem nivou, itd. Postupak je prikazan na Slici 2.3.



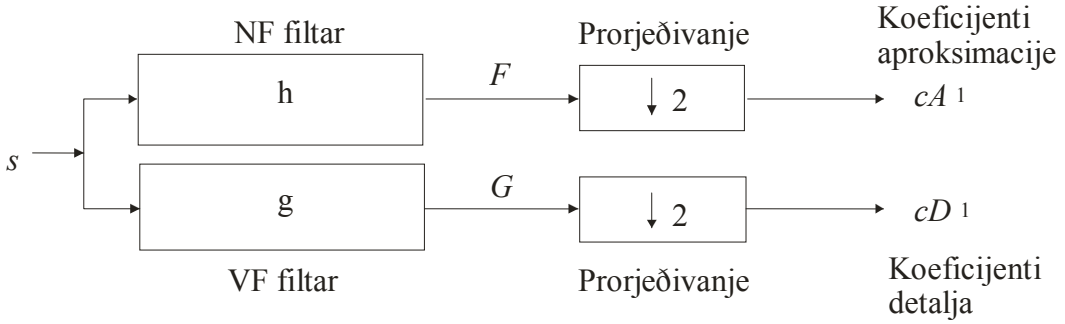
***Slika 2.3**-Dekompozicija signala pomoću DWT [10]*

Jedan nivo dekompozicije signala vrši se filtriranjem signala visokopropusnim i niskopropusnim filtrima u vremenskom domenu kako je definisano sljedećim jednačinama:

$$y_{high}[k] = \sum_n x[n] \cdot g[2k - n], \tag{2.11}$$

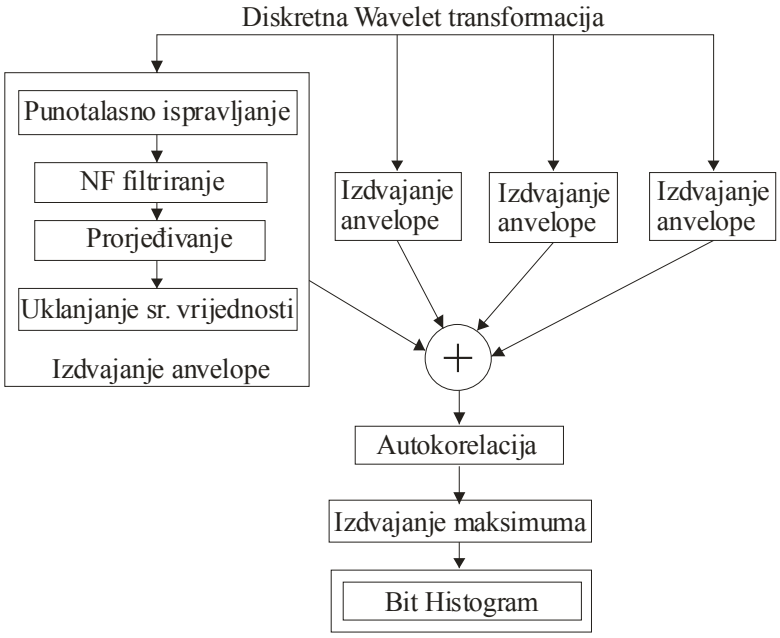
$$y_{low}[k] = \sum_n x[n] \cdot h[2k - n], \tag{2.12}$$

gdje su : $y_{high}[k]$ i $y_{low}[k]$ izlazi visokopropusnog (g) i niskopropusnog (h) filtra, respektivno nakon prorjeđivanja sa faktorom 2. Postupak je dat na dijagramu sa Slike 2.4



***Slika 2.4**-Jedan nivo dekompozicije primjenom DWT*

U radu je iskorišten *Daubechies* filtar četvrtog reda, "db4". Na Slici 2.5 prikazan je dijagram koji opisuje algoritam za izdvajanje bita (dobijanje bit histograma):



***Slika 2.5**-Izračunavanje bit histograma[4]*

Nad signalom se prvo izvrši dekompozicija na četiri nivoa (oktave) primjenom DWT-a. Nakon dekompozicije, izdvaja se anvelopa signala u vremenskom domenu za svaki opseg posebno. Ovo je postignuto primjenom tehnika punotalasnog ispravljanja, niskopropusnog filtriranja, prorjeđivanja u vremenu i uklanjanja srednje vrijednosti (istosmjerne komponente) za svaku oktavu. Nakon toga, anvelope svakog opsega su sumirane i izračunata je autokorelacija tako dobijenog signala. Dominantni pikovi autokorelacione funkcije odgovaraju različitim periodičnostima anvelope signala, tj. bitu koji je sadržan u datom audio signalu.

Izdvajaju se tri dominantna pika i dodaju u bit histogram. Svaki pik bit histograma odgovara periodu bita u bpm (*beats-per-minute*). Na ovaj način, tamo gdje je signal sebi najsličniji, pik u bit histogramu će biti najveći. Za izdvajanje obilježja korištene su sljedeće metode obrade signala.

- Punovalno ispravljanje:

$$y[n] = |x[n]|, \tag{2.13}$$

  je primjenjeno da bi se tačnije i lakše izdvojila privremena anvelopa signala .

- Niskopropusno filtriranje:

$$y[n] = (1-\alpha)\cdot x[n] + \alpha \cdot y[n-1], \tag{2.14}$$

  pomoću filtra sa jednim polom, sa $\alpha = 0.99$, koristi se za glačanje anvelope i uz punovalno ispravljanje predstavlja standardnu tehniku izdvajanja anvelope.

- Prorjeđivanje:

$$y[n] = x[kn], \tag{2.15}$$

  gdje je uzeto $k = 16$, koristi se za smanjenje broja odmjeraka signala radi smanjenja vremena izračunavanja autokorelacije bez efekta na performanse algoritma.

- Uklanjanje istosmjerne komponente:

$$y[n] = x[n] - E[x[n]], \tag{2.16}$$

  se vrši da bi se signal centrirao na nulu za izračunavanje autokorelacije.

- Autokorelacija:

$$y[n] = \frac{1}{N}\sum_{n} x[n]\cdot x[n-k], \tag{2.17}$$

  je metoda kojom se vrši prepoznavanje periodičnosti (sličnosti) u signalu, tj. tempa (u našem slučaju).

U paragrafu 2.3 detaljno je objašnjen pojam poboljšane autokorelacione funkcije (*Enhanced Summary AutoCorrelation Function* – ESACF), koja se dobija tako što se suma anvelopa prvo pozitivno klipuje, zatim vremenski proširi sa faktorom 2 i oduzme od originalne klipovane funkcije. Isti proces se može ponoviti sa drugim cjelobrojnim faktorima kako bi se otklonili harmonici osnovnog pika.

Dominantna tri pika (lokalna maksimuma) poboljšane autokorelacione funkcije koji su u rangu za bitsku detekciju, izdvojena su i dodana u bit histogram. Svaki bin histograma odgovara bitu po minuti "bpm", od 60 do 220 bpm. Bit histogram daje detaljne informacije o ritmičkom sadržaju, koje mogu biti upotrebljene za klasifikaciju muzičkog žanra. Vektor obilježja, baziranih na bit histogramu, izračunat je da reprezentuje ritmički sadržaj i služi za automatsku klasifikaciju muzičkih audio zapisa.

2.3 Obilježja tonaliteta (*Pitch Content Features*)

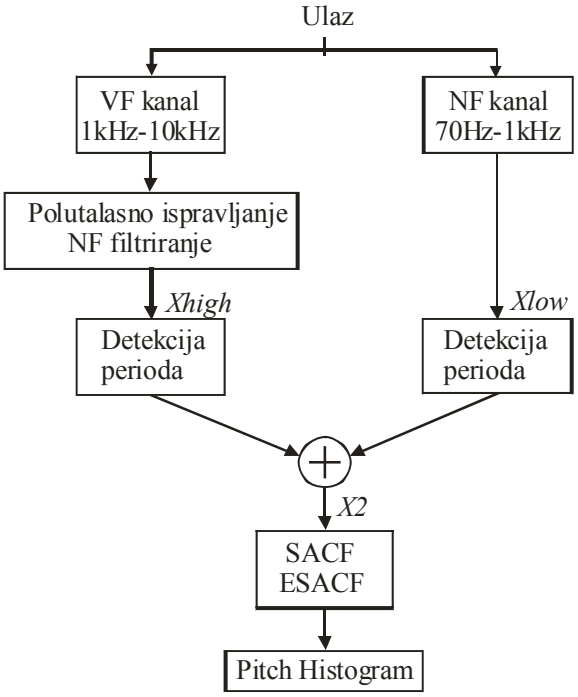
Već najprimitivnije čovjekovo muzičko izražavanje ili ono primarno – na primjer u pjevanju djeteta – pokazuje potrebu i težnju da ima neki intonacioni oslonac, tj. neku tonsku visinu koja se među ostalim ističe češćom pojavom, višestrukim ponavljanjem, i naročito time što se melodijsko kretanje na njoj zaustavlja i/ili završava, dakle "smiruje" uz izvjesno psihološko opuštanje. Ta pojava vezanosti muzičkog toka za jedno tonsko središte naziva se tonalitet, i zasniva se na svakako prirodnoj, psihološkoj potrebi čovjeka, jednako kao izvođača i kao slušaoca muzike [9].

U sistemima za audio analizu osobine tonaliteta najčešće se izražavaju uz pomoć Pitch Histograma (PH). PH predstavlja statističku reprezentaciju tonskog sadržaja muzičkog audio zapisa. Karakteristike tonaliteta izdvojene iz PH formiraju set obilježja tonaliteta. Obilježja izračunata iz PH mogu zajedno sa obilježjima teksture i ritma biti iskorištena za automatsku klasifikaciju muzičkih zapisa, što je i pokazano u ovom radu.

PH se definiše kao dijagram koji prikazuje zavisnost broja pojavljivanja svake note (tona) u muzičkom audio zapisu od cjelobrojnih vrijednosti (binova) indeksiranih MIDI (*Musical Instruments Digital Interface*-MIDI) brojevima. Pitch Histogram, u suštini, treba da prikaže tonski sadržaj, odnosno, strukturu muzičkog audio zapisa koja bi trebala da karakteriše određeni žanr. Žanrovi sa složenijom tonskom strukturom (kao što su klasika ili džez) imaju raznovrsniji spektar tonova i manje izražene pikove u svojim histogramima nego žanrovi sa "jednostavnijom akordskom progresijom" kao što su rok, pop ili hiphop.

Algoritam za izračunavanje PH poznat je pod nazivom *Multiple Pitch Detection Algorithm* [13]. Ovaj algoritam bazira se na modelu dvokanalne pič (eng. p*itch*) analize. Blok dijagram ovog modela prikazan je na Slici 2.6. Signal se razdvaja na dva kanala, ispod i iznad 1kHz, pomoću filtara propusnika opsega. Niskopropusni kanal je dobijen filtrom čiji propusni opseg iznosi od 70Hz do 1KHz, a visokopropusni kanal filtrom čiji je propusni opseg od 1KHz do 10KHz. Za razdvajanje kanala iskorišteni su filtri sa slabljenjem 12dB/oktavi[7] u nepropusnom opsegu.

[7] 12dB/oktavi=40dB/dekadi (Batterworth-ovi filtri drugod reda)



***Slika 2.6**-Multiple Pitch Detection Algorithm*

Visokopropusni kanal je još polutalasno ispravljen i "niskopropusno" filtriran filtrom propusnikom opsega koji se koristio pri odvajanju niskopropusnog kanala. Detekcija periodičnosti (*periodicity detection*) bazira se na autokorelacionoj funkciji, tj. izračunava se diskretna Furijeova transformacija (DFT-*Discrete Fourier Transform*), vrši se kompresija magnitude parametrom k , zatim primjenjuje inverzna diskretna Furijeova transformacija (IDFT-*Inverse Discrete Fourier Transform*). Signal x_2 sa Slike 2.6 dat je izrazom:

$$x_2 = IDFT\left(\left|DFT(x_{low})\right|^k + \left|DFT(x_{high})\right|^k \right), \tag{2.18}$$

gdje su x_{low} i x_{high} signali prije detekcije periodičnosti u niskopropusnom i visokopropusnom kanalu respektivno.

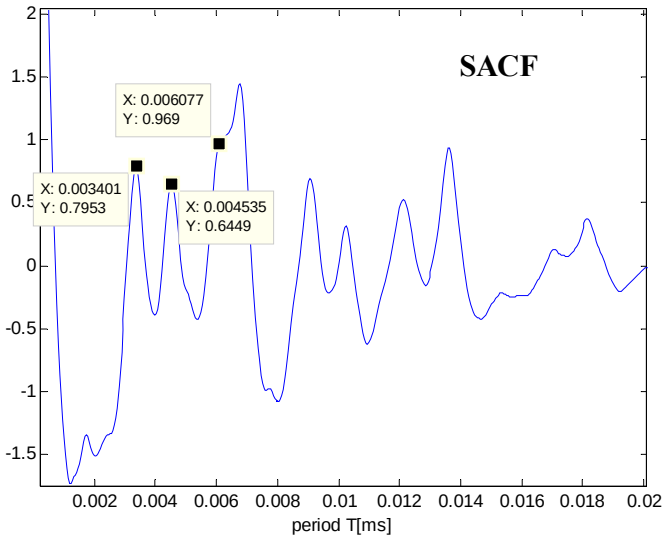
Parametar k definiše kompresiju signala u frekventnom domenu (za standardnu korelaciju je $k=2$, optimalno $k=0.67$[8]). FFT[9] algoritam se koristi za brže izračunavanje.

Pikovi u sumiranoj autokorelacionoj funkciji (SACF-Summary AutoCorrelation Function) su relativno dobri indikatori potencijalnih pič perioda u analiziranom signalu. Da bi se isključili cjelobrojni umnošci osnovnog perioda izračunava se poboljšana sumirana autokorelaciona funkcija (ESACF). SACF sadrži redundantne i lažne informacije koje otežavaju utvrđivanje koji pikovi su stvarni pič pikovi.

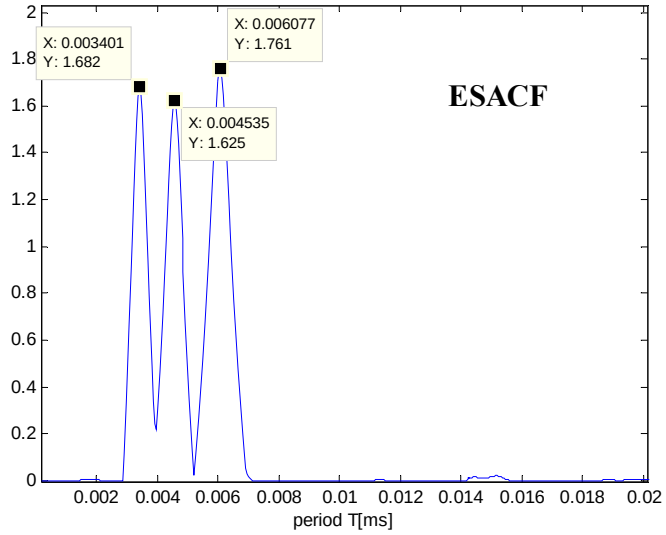
[8] Eksperimentalno je pokazano u radu [11] da optimalna vrijednost koeficijenta kompresije spektra signala iznosi $k=0.67$.
[9] FFT-Fast Fourier Transform

Tehnika izračunavanja ESACF je sljedeća: Originalna kriva SACF se prvo klipuje tako što se odsijeku negativne vrijednosti, a zatim se proširi u vremenu sa faktorom 2 i oduzme od originalne klipovane SACF funkcije. Tako dobijena funkcija se ponovo klipuje da se dobiju pozitivne vrijednosti. Ovaj postupak će otkloniti ponovljene pikove sa dvostruko većim periodom gdje je osnovni pik veći od ponovljenog. Ako je osnovni pik manji, ukloniće ga djelimično. Takođe ukloniće i dio autokorelacije koji je blizak nuli, tj. DC komponentu. Ovaj postupak može se ponavljati sa većim faktorom proširenja (3,4,5,...) radi uklanjanja pikova sa većim periodom.

Na Slikama 2.7.a i 2.7.b dat je primjer SACF i ESACF tri tona gitare (E3, A3 i D4) sa fundamentalnim frekvencijama na 165Hz (*T=6.07ms*), 220Hz (*T=4.54ms*) i 294Hz (*T=3.4ms*).



***Slika 2.7.a**-SACF*



***Slika 2.7.b**-ESACF*

Kada je dobijena ESACF uzimaju se tri dominantna pika iz svakog prozora analize i stavljaju u histogram. Tamo gdje se pikovi budu najviše poklapali amplituda u histogramu će biti najveća. Frekvencije koje odgovaraju svakom piku histograma su konvertovane u muzički ton, tako što svaki bin PH odgovara muzičkoj noti odgovarajuće frekencije (na primjer A4=440Hz). Muzičke note su definisane MIDI notnim sistemom. Konverzija frekvencije u MIDI notni broj izvršena je jednačinom:

$$n = 12\log_2\left(\frac{f}{440}\right) + 69, \tag{2.19}$$

gdje je f frekvncija u Hz, a n histogram bin (MIDI notni broj). Postoje dvije verzije PH: *folded* (FPH) i *unfolded* histogram (UPH). UPH je kreiran prema jednačini (2.19).

U slučaju FPH, sve note su mapirane u jednu oktavu pomoću jednačine:

$$c = n \cdot \mathrm{mod}12, \tag{2.20}$$

gdje je c FPH bin (tonovi jedne oktave), a n MIDI notni broj. Zatim se FPH mapira u kvintne krugove, tj. tako da se susjedni tonovi razlikuju za kvintu unaprijed, a kvartu unazad. Mapiranje je implementirano formulom:

$$c' = (7 \cdot c)\,\mathrm{mod}12, \tag{2.21}$$

gdje su c' novi histogram binovi nakon mapiranja. Broj 7 potiče od broja polutonova u okviru kvintnog intervala. Na ovaj način se dobija bolja slika odnosa između tonova, tj. dobija se tonika i dominanta [9], a i obilježja izabrana na ovaj način daju veću tačnost klasifikacije.

3. KLASIFIKACIJA UZORAKA

Algoritmima za klasifikaciju pokušava se izvršiti kategorizacija uzoraka u odgovarajuće klase ili grupe uzoraka prema klasifikacijskoj šemi. Uzorak je sačinjen od jednog ili više obilježja (deskriptora). Klase uzoraka su skupovi (familije) uzoraka koji dijele neke zajedničke osobine. Tačnost klasifikacije uzoraka bitno zavisi od izbora odgovarajućih obilježja koja će omogućiti separaciju klasa. Šema klasifikacije obično je bazirana na skupu uzoraka koji je već ranije klasifikovan ili prepoznat, tj. zna se kojoj klasi pripada. To su tzv. nadgledane metode klasifikacije (*supervised approach*). Ovaj skup uzoraka naziva se trening skup (*training set*) ili skup za obučavanje, a sam proces se naziva *obučavanje* (*učenje*). Postoji i tzv. nenadgledana (*unsupervised*) šema klasifikacije. Ovaj pristup koristi objektivnu mjeru sličnosti između podataka za klasifikaciju bez unaprijed poznatih klasa.

Šema klasifikacije obično koristi jedan od sljedećih pristupa: statistički, strukturni ili neuronski. Statistička klasifikacija uzoraka je bazirana na statističkim osobinama uzoraka pod pretpostavkom da su uzorci generisani probabilističkom metodom (funkcijom raspodjele vjerovatnoće). Strukturno prepoznavanje uzoraka je bazirano na strukturnim međuodnosima deskriptora. Neuronsko prepoznavanje uzoraka bazira se na radu procesirajućih elemenata (neurona) međusobno povezanih u jednu cjelinu koja se naziva neuronska mreža.

U literaturi postoji mnogo različitih klasifikatora i ne može se tvrditi koji je bolji, jer se međusobno razlikuju u mnogim aspektima kao što su: brzina učenja, količina podataka za obuku, brzina klasifikacije, robusnost, itd.

Glavna ideja koja se krije iza većine ovih šema klasifikacije je "učenje" geometrijskih struktura trening podataka u prostoru obilježja i njihovo korištenje za klasifikaciju novih uzoraka.

Parametarski klasifikatori temelje se na pretpostavci da je pripadnost uzoraka određena funkcijom raspodjele vjerovatnoće obilježja, dok neparametarski direktno koriste trening skup za klasifikaciju.

U ovoj glavi dat je sažet opis tri metoda klasifikacije (klasifikatora): Gaussian Mixture Model (GMM), vještačke neuronske mreže (Artifical Neural Networks-ANNs) i metod *k* najbližih susjeda (*k*-Nearest Neighbor - *k*-NN).

3.1 Gaussian Mixture Model (GMM)

GMM klasifikator spada u grupu statističkih parametarskih klasifikatora. Ideja statističkih klasifikatora je da se vektor obilježja interpretira kao stohastička varijabla čija raspodjela zavisi od klase uzoraka. Kao veoma važan alat koristi se Bayesova formula:

$$P(\omega_i \mid x) = \frac{P(x \mid \omega_i) \cdot P(\omega_i)}{P(x)}, \tag{3.1}$$

gdje $P(\omega_i \mid x)$ označava uslovnu vjerovatnoću klase ω_i, uslovljenu vektorom obilježja x.

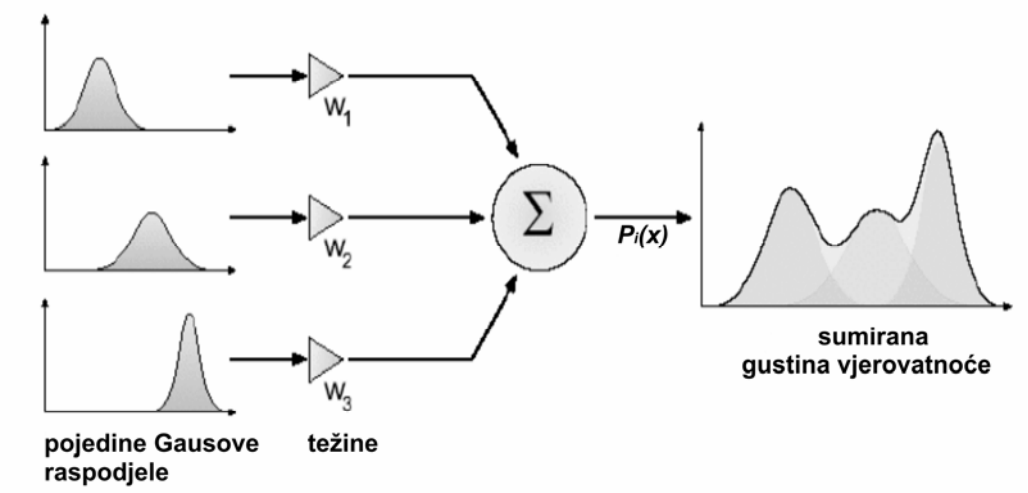
Mora se primjetiti da je $P(\omega_i \mid x)$ nepoznato, jer je nepoznata vjerovatnoća obilježja x koja je uslovljena klasom ω_i, $P(x \mid \omega_i)$, ali se ona može estimirati pomoću trening skupa. Potrebno je, dakle, da se parametriziraju i nauče vjerovatnoće $P(x \mid \omega_i)$ i definišu ili nauče vjerovatnoće klasa $P(\omega_i)$.

GMM je veoma pogodan za reprezentaciju uslovne vjerovatnoće $P(x \mid \omega_i)$, tj. za reprezentaciju višedimenzionalne raspodjele vektora obilježja za i-tu klasu ω_i. Težinska suma višedimenzionalnih Gausovih raspodjela data je sa:

$$P(x \mid \omega_i) = \sum_{q=1}^{Q} w_{i,q} \cdot N(x; \mu_{i,q}, \Sigma_{i,q}), \tag{3.2}$$

gdje su $w_{i,q} \leq 0$ težine, pri čemu važi $\sum_{q=1}^{Q} w_{i,q} = 1$, a $N(x, \mu, \Sigma)$ Gausove raspodjele .

Na Slici 3.1 prikazan je GMM jednodimenzionalni model.



Slika 3.1*-Jednodimenzionalni GMM (Q=3)*

Gausova (normalna) raspodjela data je poznatim izrazom:

$$N(x;\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)^{D}|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)^{T}\Sigma^{-1}(x-\mu)\right), \tag{3.3}$$

gdje su μ - D dimenzionalni vektor srednjih vrijednosti od x i Σ - $D\times D$ kovarijansna matrica od x. Zadovoljavajuća generalizacija modela zahtijeva konačan broj Q Gausovih raspodjela.

GMM clasifikator je i parametarski klasifikator. Parametri su: težine $w_{i,q}$, srednja vrijednost $\mu_{i,q}$ i kovarijansna matrica $\Sigma_{i,q}$. Ako su obilježja dekorelisana može se upotrebiti dijagonalna kovarijansna matrica, jar sadrži manje parametara.

Dakle, pomoću GMMa vrši se estimacija parametara modela za svaku klasu i uzimaju se maksimumi vjerovatnoće $P(x\,|\,\omega_i)$, tj. vjerovatnoće trening skupa za svaku klasu ω_i. Sa ovako izdvojenim maksimumima pomoću Bayesove formule dolazi se do konačne vjerovatnoće klase ω_i kojoj pripada vektor obilježja x.

3.2 Artifical Neural Networks (ANNs)

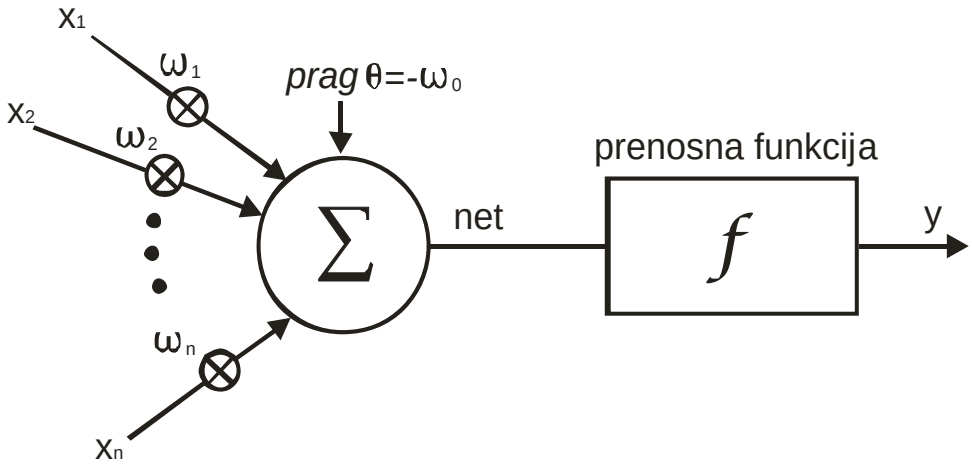
Neuronske mreže simuliraju način rada ljudskog mozga pri obavljanju datog zadatka ili neke funkcije. Neuronska mreža je masovno paralelizovan distribuirani procesor sa prirodnom sposobnošću memorisanja iskustvenog znanja i obezbjeđivanja njegovog korištenja. Vještačke neuronske mreže podsjećaju na ljudski mozak u dva pogleda:

- neuronska mreža prikuplja znanje kroz proces obučavanja,
- težine između neurona mreže (jačina sinaptičkih veza) služe za memorisanje znanja.

Procedura kojom se obavlja obučavanje je algoritam obučavanja. Kroz ovu proceduru se na algoritamski način mjenjaju sinaptičke težine u cilju dostizanja željenih performansi mreže.

Osnovnu računarsku snagu neuronskih mreža čini masovni paralelizam, sposobnost obučavanja i generalizacija. Generalizacija predstavlja sposobnost produkovanja zadovoljavajućeg izlaza neuronske mreže i za ulaze koji nisu prisutni u toku obučavanja.

Osnovni elemenat neuronske mreže je neuron. On izgleda kao na sledećoj slici.



***Slika 3.2**-Model vještačkog neurona*

Ulazne signale, njih ukupno n, označavamo sa $x_1, x_2, \ldots x_n$. Težine označavamo sa $w_1, w_2, \ldots w_n$. Ulazni signali, uopšte, su realni brojevi u intervalu [-1,1], [0,1] ili samo elementi iz {0,1}, kada govorimo o Booleovom ulazu. Težinska suma *net* data je sa:

$$net = \omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_n x_n - \theta\,, \tag{3.4}$$

ali se zbog kompaktnosti često dogovorno uzima da je vrijednost praga $\theta = -\omega_0$, te se dodaje ulazni signal x_0 sa fiksiranom vrijednošću 1.

Sada imamo:

$$net = \omega_0 x_0 + \omega_1 x_1 + \omega_2 x_2 + ... + \omega_n x_n = \sum_{i=0}^{n} \omega_i x_i \,, \tag{3.5}$$

dok je izlaz y rezultat prenosne funkcije primjenjene na izraz (3.5):

$$y = f(\sum_{i=0}^{n} \omega_i x_i) = f(net) \,. \tag{3.6}$$

Neuronske mreže rješavaju probleme *klasifikacije* i *predikcije*, odnosno uopšte sve probleme kod kojih postoji odnos između prediktorskih (ulaznih) i zavisnih (izlaznih) varijabli, bez obzira na visoku složenost te veze (nelinearnost). Široko korištena mreža za prepoznavanje uzoraka je višeslojni perceptron (*multilayer perceptron-MLP*). To je generalizovana mreža koja u principu može da aproksimira bilo koju nelinearnu funkciju. Neuronske mreže, kao i ostale opisane arhitekture, mogu da riješavaju jedino zadatke čija se obilježja ne mijenjaju u vremenu. Ova slabost se djelimično može prevazići korištenjem tzv. *feedforward* (kontrola procesa korištenjem očekivanih rezultata) mreža.

Neuronske mreže nisu do sada često korištene u klasifikaciji muzičkih audio zapisa, jer generalno neuronske mreže su neistraženo područije i njihovo vrijeme tek dolazi. U [12] je predstavljen jedan originalan metod za klasifikaciju muzičkog žanra u realnom vremenu pomoću neuronske mreže (*Explicite Time Modeling, ETM-NN*).

U ovom radu je implementiran klasifikator na bazi neuronske mreže kao što je opisano u četvrtoj glavi u poglavlju 4.1.4.

3.3 *k*-Nearest Neighbor (*k*-NN)

Za razliku od parametarskih klasifikatora, *k*-NN klasifikator direktno koristi trening skup za klasifikaciju, bez korištenja ikakve matematičke forme za funkcije gustine vjerovatnoće osnovnih klasa. Kod NN (*Nearest Neighbor*) klasifikatora svaki uzorak se klasifikuje prema klasi svog najbližeg susjeda iz trening skupa. Kod *k*-NN klasifikatora, pronalazi se *k* najbližih susjeda uzorka koji se klasifikuje i glasanjem se utvrđuje kojoj klasi uzorak pripada. Detaljnije, za svaki ulazni vektor obilježja koji se klasifikuje (pripada test skupu), pronalaze se klase *k* najbližih vektora susjeda iz trening skupa, a zatim se ulazni vektor svrstava u onu klasu koja ima najviše članova. Kao mjera klase najbližih susjeda koristi se metrika. Najčešće je to Euklidova ili Mahalanobisova distanca. Za k=1 dobija se najjednostavniji slučaj (NN klasifikator), gdje se klasa ulaznog vektora obilježja određuje prema klasi najbližeg (po izabranoj metrici) vektora iz trening skupa.

Euklidova distanca između vektora obilježja $X = [x_1, x_2, ..., x_n]^T$ i $Y = [y_1, y_2, ...y_3]^T$ je data jednačinom:

$$D_E(X,Y) = \sqrt{(X-Y)^T \cdot (X-Y)} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}\,, \tag{3.7}$$

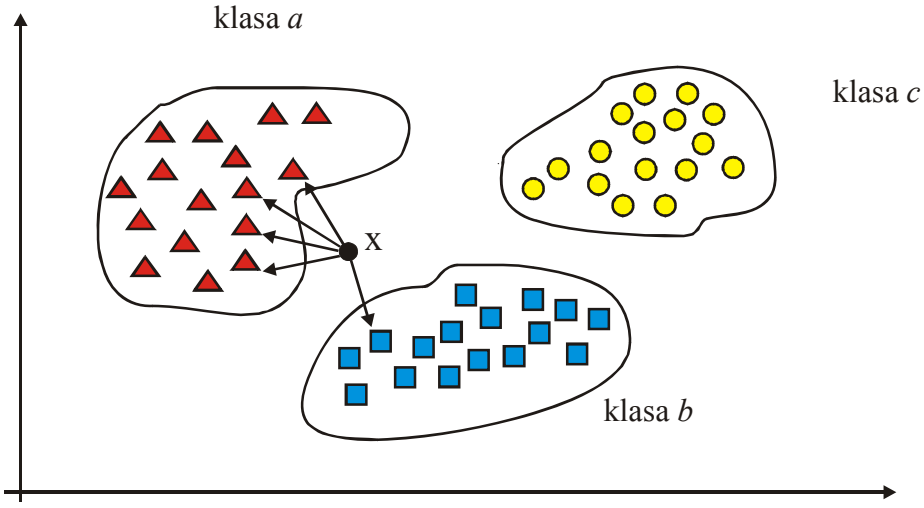
dok je Mahalanobisova distanca:

$$D_M(X,Y) = (X-Y)^T \cdot C^{-1} \cdot (X-Y)\,, \tag{3.8}$$

gdje je C kovarijansna matrica trening skupa. Mahalanobisova distanca se češće koristi prilikom klasifikacije obilježja koja su medjusobno korelisana.

U slučaju da obilježja imaju različite dinamičke opsege, što znači da bi onda velika udaljenost po nekom obilježju mogla da dominira distancom, narušile bi se performanse klasifikatora. Moguće rješenje ovog problema je da se sva obilježja normalizuju u opseg [0, 1], tako što će se naći maksimalne vrijednosti za svako obilježje i podjeliti ulazni vektor tim vrijednostima.

Na Slici 3.3 prikazan je primjer klasifikacije triju klasa korištenjem *k*-NN klasifikatora.



***Slika 3.3**-k-NN klasifikator (k=5)*

U ovom primjeru *k*-NN klasifikator ima za zadatak da pronađe nepoznatu klasu vektora X. Kao što se vidi, od pet najbližih susjednih vektora četiri pripadaju klasi *a*, a jedan klasi *b*, pa prema tome, vektor X pripada klasi *a*.

Nedostaci *k*-NN klasifikatora su:

- zahtijeva cijeli trening skup kada je potrebno klasifikovati novi vektor obilježja, što zahtijeva veliki memorijski kapacitet;
- vrijeme klasifikacije je duže u poređenju sa drugim klasifikatorima.

Prednosti *k*-NN klasifikatora su:

- ne zahtijeva obučavanje (trening) što je naročito od pomoći kada se klasifikuju novi uzorci
- koristi lokalne informacije tako da može obrađivati kompleksne funkcije koje nisu eksplicitno zadate

3.3.1 Unakrsna provjera (Cross - Validation)

Prilikom evaluacije nekog od metoda automatske klasifikacije tipično je da se set podataka koji se koristi za trening koristi i za testiranje. Prilikom provjere (validacije) set raspoloživih podataka se dijeli na dva skupa. Jedan je tradicionalno trening skup, koji se koristi za podešavanje parametara modela klasifikatora, a drugi, tzv. validacioni skup (validation set), se koristi za estimaciju greške koja se koristi pri poređenju različitih parametara klasifikatora ili klasifikatora uopšte. Očigledno, ova greška zavisi od dijeljenja podataka na trening i test skup, a postoji i opasnost od tzv. *overfittinga*.

Da bi se riješio ovaj problem, jednostavnom generalizacijom prethodnog metoda, koristi se tzv. *m-fold cross-validation* metod. Sada se raspoloživi skup podataka slučajno podijeli na *m* dijelova jednakih dužina *n/m*, gdje je *n* ukupan broj raspoloživih podataka. Klasifikator se trenira *m* puta, svaki put sa drugim skupom koji predstavlja validacioni, odnosno, test skup. Ukupna greška je srednja vrijednost grešaka po iteraciji. Kada se podaci slučajnim odabirom podijele na *m* jednakih dijelova, na taj način da svaki dio dostojno reprezentuje određenu klasu, dobija se tzv. *stratified m-fold cross-validation* metod.

Pošto različiti *m-fold cross-validation* eksperimenti sa istim setom podataka i šemom učenja ponekad daju različite rezultate, kao i iz razloga smanjivanja uticaja određene podjele raspoloživog skupa podataka (pristrasnosti), česo se koristi određeni broj ponavljanja *m-fold cross-validation* algoritma. Ukupna greška jednaka je srednjoj vrijednosti grešaka po iteraciji.

U ovom radu primjenjen je *10-fold cross-validation* algoritam koji je ponovljen na raspoloživom skupu podataka 100 puta.

4. EKSPERIMENTALNI REZULTATI

Klasifikacija muzičkih audio zapisa korištenjem opisanih obilježja i klasifikatora konstruisanog metodom *k*-Nearest Neighbor (*k*-NN) izvršena je implementacijom programa za izdvajanje obilježja i obučavanje klasifikatora u MATLAB-u. U ovom poglavlju dat je pregled detalja implementacije programa, kao i statistika, odnosno rezultati koji potvrđuju teorijska razmatranja opisana u prethodnim poglavljima.

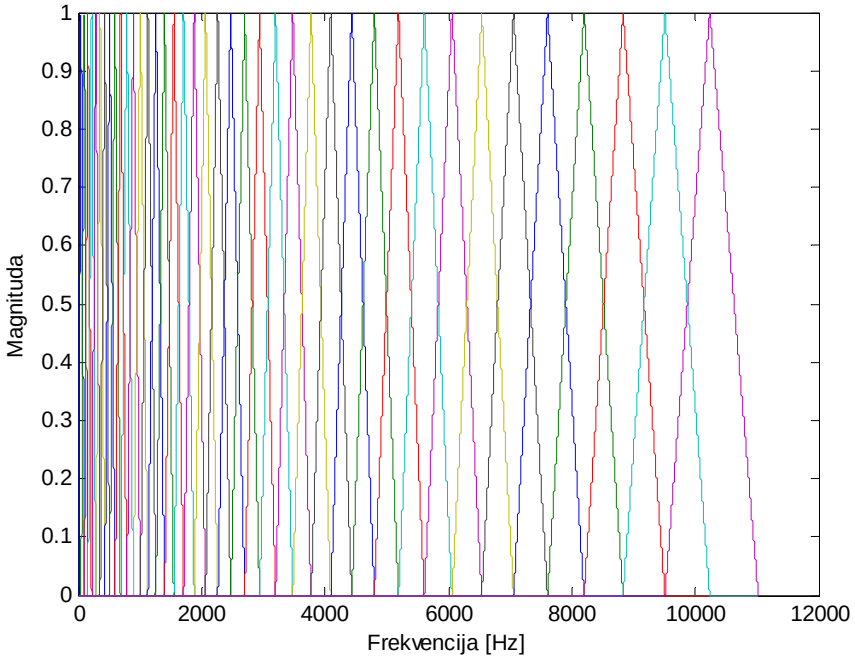
4.1 Detalji implementacije

Implementirani program se sastoji iz dva dijela. Prvi dio izdvaja obilježja opisana u Poglavlju 2 i smješta ih u MATLAB-ove strukture podataka. Drugi dio koristi izračunata obilježja za obučavanje i testiranje klasifikatora. Izdavajanje vektora obilježja implementirano je funkcijom *features.m* u okviru koje se pozivaju podfunkcije koje izračunavaju karakteristične vektore obilježja teksture, ritma i tonaliteta, a nazvane su *texture.m*, *rhythm.m*, *pitch.m*, respektivno. Program za obučavanje i testiranje klasifikatora sastoji se iz dvije funkcije. Jedna je *genres.m* koja učitava vektore obilježja svih žanrova, a druga je *KNN.m* koja predstavlja klasifikator. Treba napomenuti da izdvojena obilježja zahtijevaju znatno manje memorije od ''sirovih'' audio signala tako da je, ukoliko je potrebno ponoviti obučavanje klasifikatora, znatno pogodnije čuvati samo obilježja.

4.1.1 Implementacija obilježja teksture

U kratkotrajnoj audio analizi postoji potreba da se signal u vremenu rastavi na manje segmente koji se mogu i preklapati i da se onda svaki segment posmatra odvojeno. Ovi segmenti se nazivaju *prozori analize (frame)* i trebaju biti toliko mali da se signal unutar njih može smatrati stacionarnim, odnosno da je spektar signala u okviru prozora vremenski nepromjenljiv.

Termin *prozor teksture* se koristi da opiše jedan veći prozor koji odgovara minimalno potrebnoj količini vremena koja je dovoljna da se identifikuje određeni zvuk ili muzička tekstura. Heuristički je pokazano da prozor analize traje 23ms (512 odmjeraka pri frekvenciji odmjeravanja od 22050Hz), a prozor teksture 1s (43 prozora analize) [4]. Da bi se izdvojile karakteristične osobine teksture kroz cijeli audio zapis vrši se izračunavanje srednje vrijednosti i varijanse obilježja preko svih prozora analize u okviru svakog prozora teksture, a zatim se vrši usrednjavanje tako dobijenih vrijednosti preko svih prozora teksture.
Implementacija *spektralnog centroida* izvršena je po formuli (2.1) duž svakog prozora analize i data je funkcijom *spectralcentroid.m*. *Spektralni rolloff*, *spektralni fluks* i *broj prolazaka kroz nulu* su takođe implementirani duž svakog prozora analize po relacijama (2.2-2.4) i predstavljeni su funkcijama *spectralrolloff.m*, *spectralflux.m*, *zerocross.m* respektivno. U implementaciji izdvajanja obilježja *Mel cepstralnih koeficijenata* (MFCC) bilo je potrebno projektovati banka filtar. To je predstavljeno funkcijom *mel.m*. Banka filtar je trebalo konstruisati tako da njegove centralne frekvencije budu logaritamski raspoređene na frekvencijskoj osi, a propusni opsezi odgovaraju kritičnim opsezima. U radu je iskorišten ISP (*Intelligent Sound Implementation* [8]) model izračunavanja MFCC koji je opisan u 2. poglavlju. Za izračunavanje MFCC koriste se filtri čija je karakteristika trougaonog oblika, a granične frekvencije na mel-transformisanoj frekvencijskoj osi se nalaze na polovini udaljenosti između centralnih frekvencija susjednih filtara. Na Slici 4.1 prikazan je jedan banka filtar koji se sastoji od 40 propusnika opsega.



***Slika 4.1**-Banka filtar za izračunavanje MFCC*

Posljednje obilježje teksture je *prozori sa niskom energijom* koje je implementirano u samoj glavnoj funkciji koja kao izlaz daje vektor obilježja texture, *texture.m* .

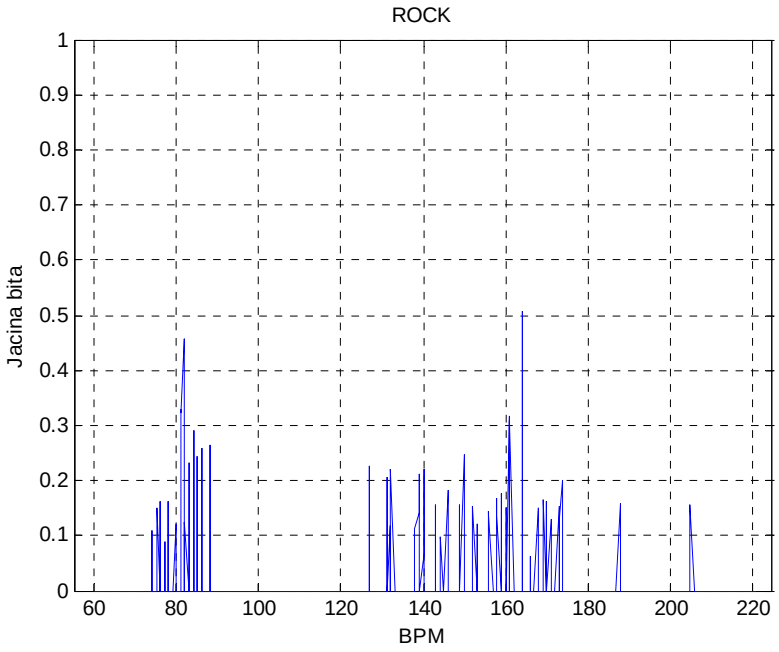
Implementirani vektor za opis obilježja teksture sastoji se od sljedećih obilježja: srednje vrijednosti i varijanse *spektralnog centroida, rolloffa, fluksa, broja prolazaka kroz nulu* za pojedini prozor teksture, koje su usrednjene duž svih prozora teksture (što daje 8 obilježja), zatim broja prozora sa niskom energijom za pojedini prozor teksture, koji su usrednjeni duž svih prozora teksture (1 obilježje), i srednje vrijednosti i varijanse prvih 5 MFCC koeficijenata za pojedine prozore teksture, koje su takodje usrednjene duž svih prozora teksture (10 obilježja). Konačni vektor obilježja teksture predstavlja 19-dimenzionalni vektor.

4.1.2 Implementacija obilježja ritma

Da bi se došlo do vektora obilježja ritma potrebno je izračunati bit histogram. Bit histogram daje sliku o zavisnosti jačine bita od perioda bita izraženog u bita-po-minuti (bpm). Prije toga, potrebno je izračunati poboljšanu autokorelacionu funkciju, ESACF, iz sumirane autokorelacione funkcije. Teorijski opis dobijanja poboljšanja dat je u paragrafu 2.3. Praktična implementacija zahtijevala je nešto drugačiji pristup. Naime, prilikom proširivanja u vremenu sa faktorom 2 (3,4,5 ili višim) u diskretnom domenu dolazi do pojavljivanja nula na svakom drugom odmjerku (ili većem) i tako dobijena proširena funkcija kada se oduzme od polazne autokorelacione funkcije (SACF) daje nove nepoželjne odmjerke. Ti odmjerci kvare sliku ESACF i praktično se ne poboljšava SACF. Da bi se to izbjeglo potrebno je izvršiti niskopropusno filtriranje proširene autokorelacije.

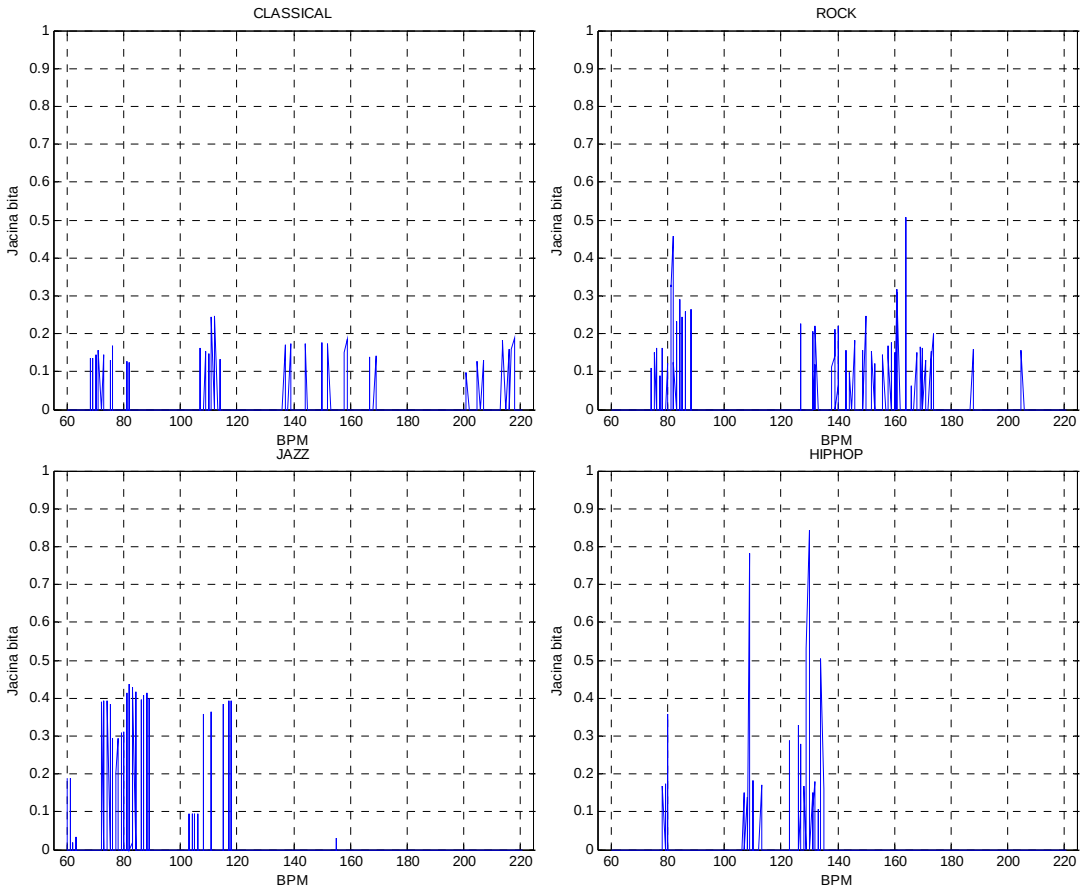
Još bolje rezultate dalo je ubacivanje srednje vrijednosti odmjeraka koji se nalaze ispred i iza nultih odmjerka, na njihovo mjesto. Ovo je također filtriranje niskopropusnim filtrom, samo što je filtar nekauzalan. Kada se ovako dobijena proširena funkcija oduzme od polazne SACF u potpunosti se eliminišu harmonici osnovne frekvencije koji odgovaraju faktoru proširenja. Nakon toga uzimaju se tri pika ESACF koji su u odgovarajućem opsegu za bitsku detekciju i stavljaju se u bit histogram. Svaki bin histograma odgovara bitu u opsegu od 60bpm do 220bpm. Na taj način pikovi koji imaju veću amplitudu (tamo gdje je signal najsličniji samom sebi) potiskuju pikove sa manjom amplitudom i bivaju izražajniji u histogramu.

Na Slici 4.2 prikazan je bit histogram isječka trajanja 30s audio zapisa pjesme "Come Together" rok benda The Beatles.



***Slika 4.2**-Primjer bit histograma kompozicije* "Come Together", The Beatles

Dva najveća pika histograma odgovaraju glavnom bitu, tj. tempu zapisa, na 80bpm i njegovom prvom harmoniku, dva puta većeg tempa, na 160bpm. Heuristički je pokazano da tempo pjesme najčešće odgovara prvom ili drugom piku histograma [4]. Na Slici 4.3 prikazana su četiri bit histograma kompozicija iz različitih muzičkih žanrova.



***Slika 4.3**-Primjeri bit histograma*

U gornjem lijevom uglu prikazan je bit histogram klasike. To je histogram Mozartove četrdesete simfonije. Primjećuje se da kompozicija nema izraženih pikova u histogramu, kao i da je jačina postojećih pikova veoma mala. Ova pojava je karakteristična za klasični žanr i dešava se zbog kompleksnosti i višestrukosti instrumenata u orkestru kao i zbog činjenice da u klasičnoj muzici nije naglašena ritam sekcija. Malo jači pikovi se mogu vidjeti u donjem lijevom uglu gdje je predstavljen histogram za pjesmu I Can't Stop Loving You koju izvodi Ray Charles. U pitanju je džez. I ovdje su binovi histograma podjednake snage. Ističu se pikovi oko 80bpm i 120bpm. U gornjem desnom uglu dat je histogram rok pjesme Come Together, The Beatles. Pikovi su više izraženiji jer rok žanr ima snažniji bit. Najveći pikovi u donjem desnom uglu prikazuju snažnu ritmičku strukturu hip-hop pjesme Candy Shop izvođača 50Cent. Sa slike 4.3 se još može uočiti da se muzički žanrovi mogu i vizuelno razlikovati.

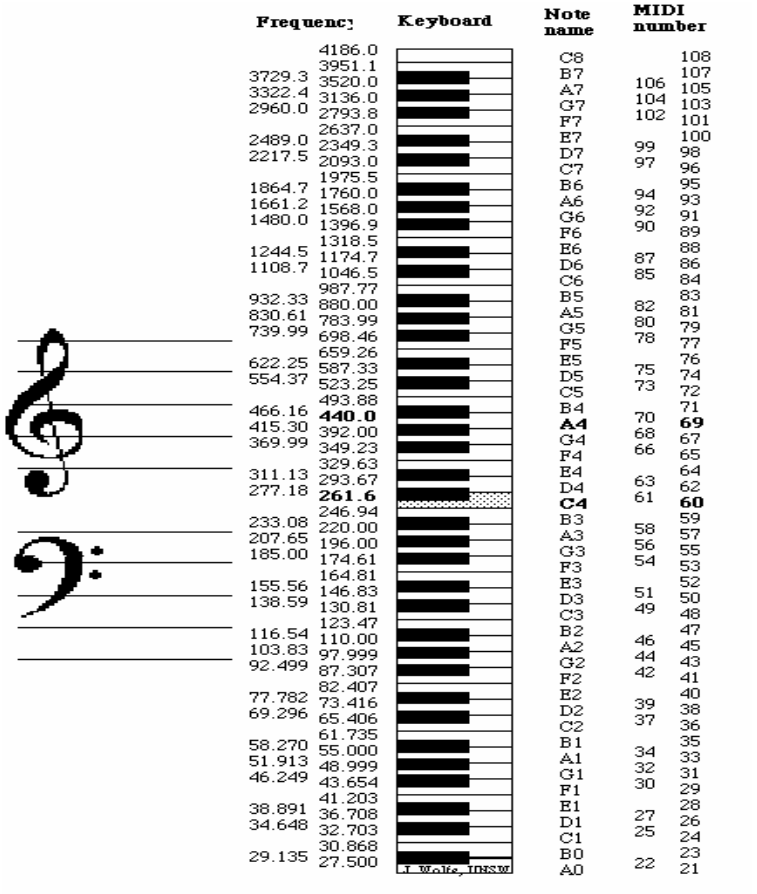
Set obilježja baziran na bit histogramu je izračunat tako da predstavlja ritmički sadržaj u svrhu automatske klasifikacije muzičkih zapisa po žanru. U ritmička obilježja spadaju:

- **A0**,**A1**: Relativne amplitude (podijeljene sumom amplituda) prvog i drugog pika histograma;
- **RA**: Odnos amplituda drugog i prvog pika histograma;
- **P1**,**P2**: Period prvog i drugog pika izražen u bpm;
- **SUM**: Suma duž cijelog histograma.

Za izračunavanje bit histograma DWT je primjenjena na prozore dužine 65536 odmjeraka sa frekvencijom odmjeravanja 22050Hz, što odgovara dužini od 3s. Prozor je pomjeran sa pomakom od 32768 odmjeraka, što odgovara dužini od 1,5s. Izračunavanje obilježja ritma implementirano je u MATLABu funkcijom *rhythm.m* .

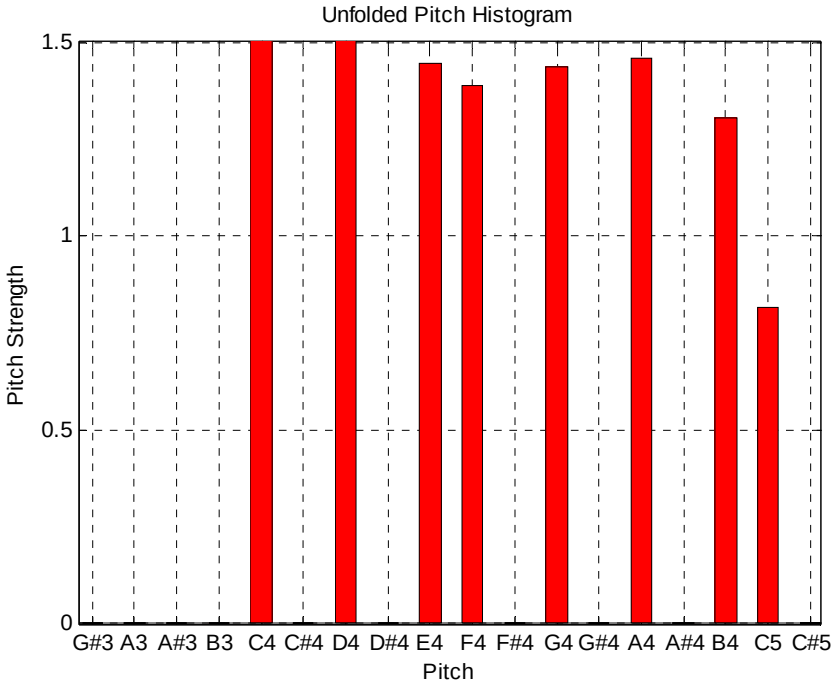
4.1.3 Implementacija obilježja tonaliteta

Slično kao kod izračunavanja obilježja ritma, da bi se došlo do obilježja tonaliteta potrebno je izračunati pič (tonski) histogram. Kao što je pomenuto u paragrafu 2.3, postoje dvije verzije pič histograma: *folded* (FPH) i *unfolded* (UPH). UPH predstavlja tonski sadržaj zapisa kroz nekoliko oktava, dok FPH predstavlja sliku svih tih tonova u jednoj oktavi (opseg od 12 tonova). Pri tome se kod FPH vrši još i mapiranje tonova u kvintne krugove, tako da se susjedni binovi histograma razlikuju za čistu kvintu (pet stupnjeva, tonika i dominanta). Za implementaciju pič histograma bitno je poznavati parametre kao što su MIDI broj, nota, fundamentalna frekvencija, kao i veze između njih. Na sljedećoj slici prikazani su međusobni odnosi između pomenutih parametara.

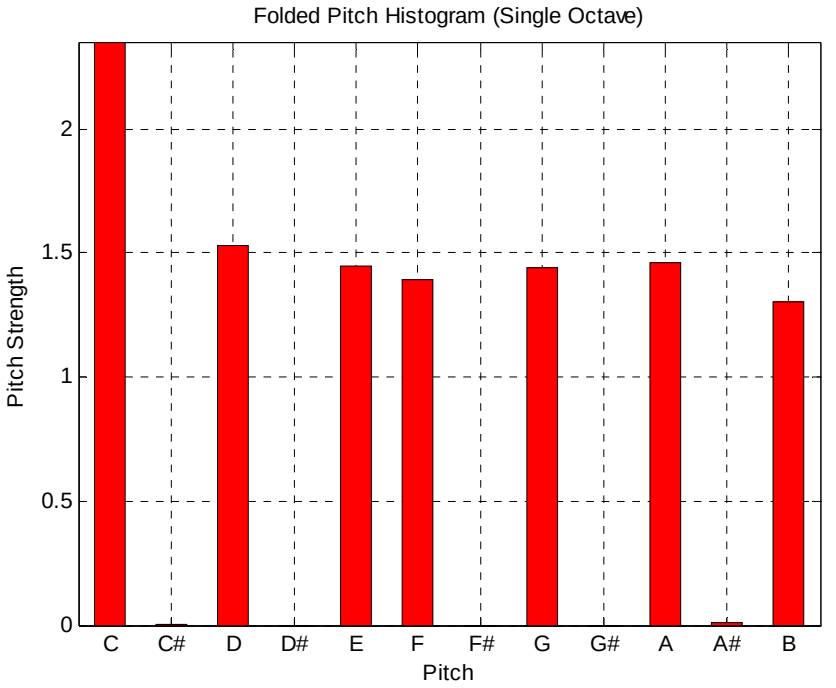


Slika 4.4*-Note, frekvencije i MIDI brojevi[17]*

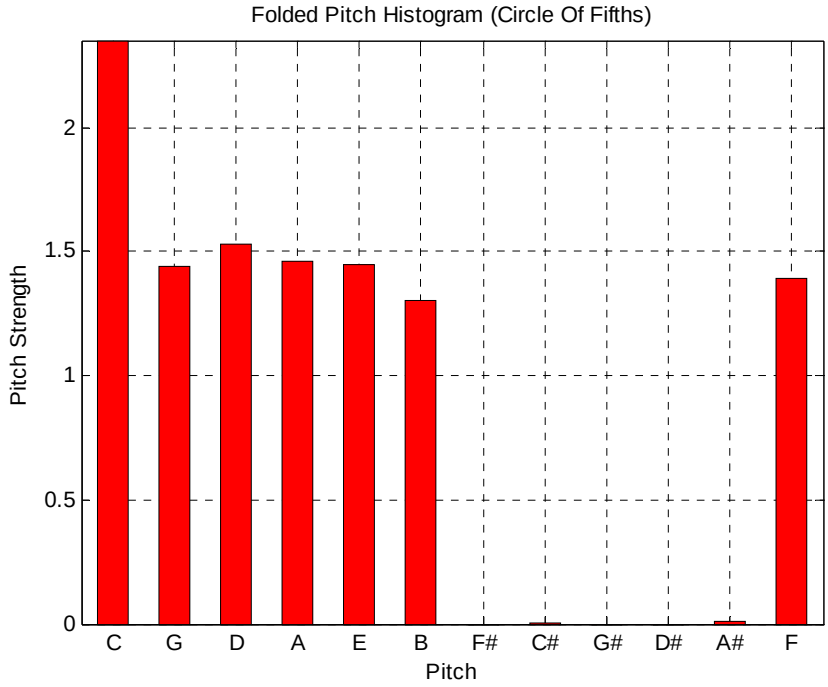
Veza između MIDI broja i fundamentalne frekvencije data je u paragrafu 2.3 relacijom (2.19). Na Slikama 4.5-4.7 prikazani su primjeri pič histograma na Cdur skali iz četvrte oktave (C4 osnovni ton, F=261,63Hz, Midi=60).



***Slika 4.5**-UPH C skale iz četvrte oktave*

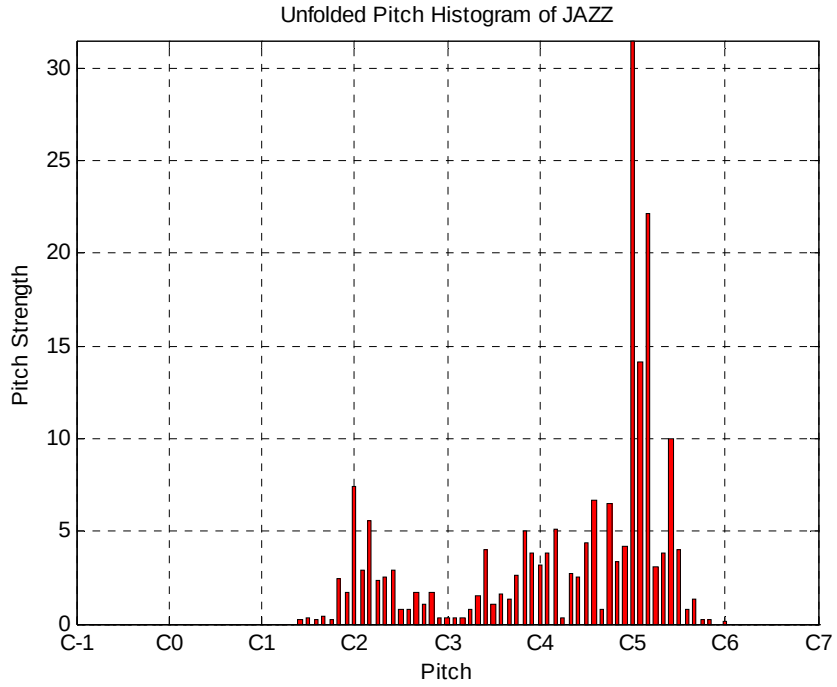


***Slika 4.6**-FPH (Single Octave) C skale iz četvrte oktave*

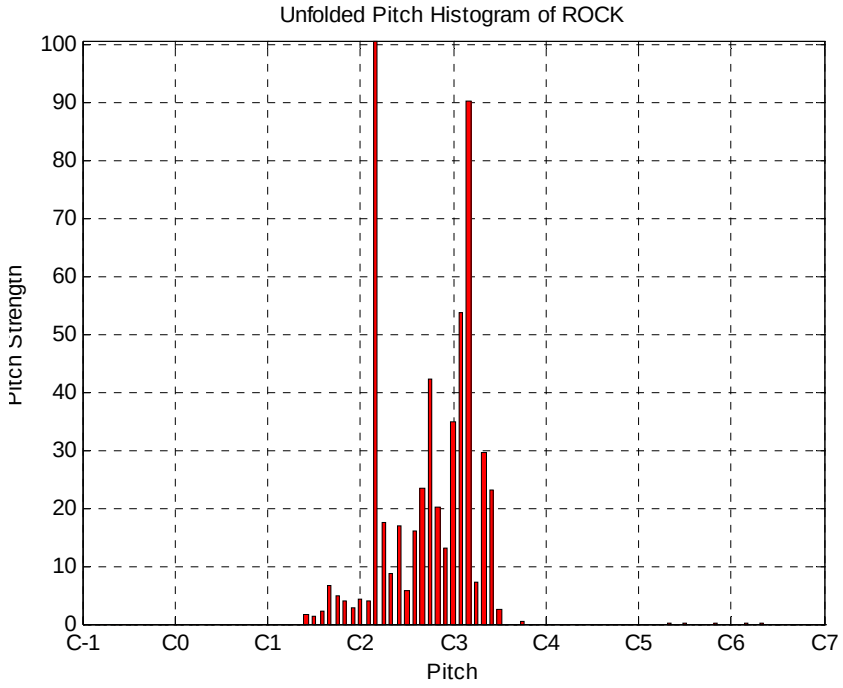


***Slika 4.7**-FPH (Circle Of Fifths) C skale iz četvrte oktave*

Mapiranje tonova iz FPH u kvintne krugove čini da se u histogramu bolje izraze odnosi tonova u zapisu, a i empirijski je dokazano da obilježja dobijena na ovaj način daju tačniju klasifikaciju [13]. Kao primjer može se pokazati da akord u C duru ima jače pikove na C i G tonovima (to su njegova tonika i dominanta) i sličniji je G duru (koji ima za toniku i dominantu G i D ton) nego C# dur akord (koji ima za toniku i dominantu C# i G#). FPH sadrži informacije vezane za tonski sadržaj muzike (tonalitet), dok UPH određuje raspon tonova. Također, na sljedećim slikama se može vidjeti da žanrovi poput džeza ili klasike imaju širi raspon tonaliteta nego žanrovi kao što su rok ili pop. Kao posljedica, dešava se da pop ili rok žanrovi imaju rjeđe i više izraženije pikove u histogramu nego džez ili klasični žanr.



***Slika 4.8**-UPH za džez*



***Slika 4.9**-UPH za rok*

Obilježja koja predstavljaju tonski sadržaj formiraju se iz UPH i FPH. To su:

- **FA0**: Amplituda maksimalnog pika FPH-a. Ovo odgovara osnovnom (glavnom) tonalitetu pjesme. Najčešće je to tonika ili dominanta. Ovaj pik će biti veći za pjesme koje nemaju mnogo harmonijskih promjena.
- **UP0**: Period maksimalnog pika UPH-a u bpm, što odgovara rasponu oktava glavnog tonaliteta pjesme.
- **FP0**: Period makrimalnog pika FPH-a u bpm, što odgovara osnovnom tonalitetu pjesme.
- **IPO1**: Interval između dva najveća pika FPH-a u bpm, što odgovara odnosu između tonskih intervala (terca, kvarta, kvinta,...). Za pjesme sa jednostavnom harmonijom ovo obilježje će imati vrijednosti 1 ili -1, što odgovara kvintnom ili kvartnom intervalu dva najveća pika.
- **OSUM**: Suma duž histograma. Ovo obilježje daje mjeru jačine pič detekcije.

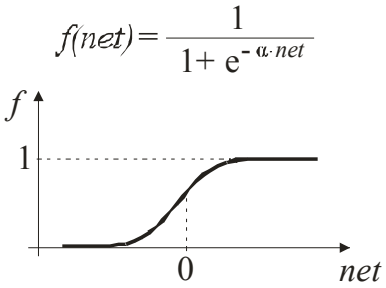
Za izračunavanje pič histograma korišteni su prozori analize dužine 512 odmjeraka pri frekvenciji odmjeravanja od 22050 Hz, što iznosi oko 23ms. Obilježja tonaliteta implementirana su u MATLAB-u funkcijom pitch.m .

4.1.4 Implementacija ANN klasifikatora

Prvobitna zamisao bila nam je da izvršimo klasifikaciju muzičnih audio zapisa primjenom vještačkih neuronskih mreža (ANNs). Implementiran je klasifikator na bazi neuronske mreže.

Učinkovita i popularna metoda učenja višeslojnih mreža jeste algoritam sa širenjem greške unazad, tzv. *backpropagation* algoritam. Ovaj algoritam je iskorišten u implementaciji klasifikatora. Realizovana mreža ima tri sloja: ulazni sloj, skrivini sloj i izlazni sloj. Učenje višeslojne mreže pomoću *backpropagation* algoritma svodi se na pretraživanje u *n*-dimenzionalnom prostoru hipoteza, gdje je *n* ukupan broj težinskih faktora u mreži. Grešku u takvom prostoru možemo vizualizovati kao hiper-površinu koja, za razliku od parabolične površine jednog procesnog elementa, može sadržavati više lokalnih minimuma. Zbog toga postupak gradijentnog spusta lako može zaglaviti u nekom lokalnom minimumu. U većini praktičnih primjena se pokazuje da algoritam i pored toga daje vrlo dobre rezultate.

U prvom sloju mreže imamo 30 neurona, a prenosna funkcija je sigmoidna funkcija koja izgleda kao na Slici 4.10.



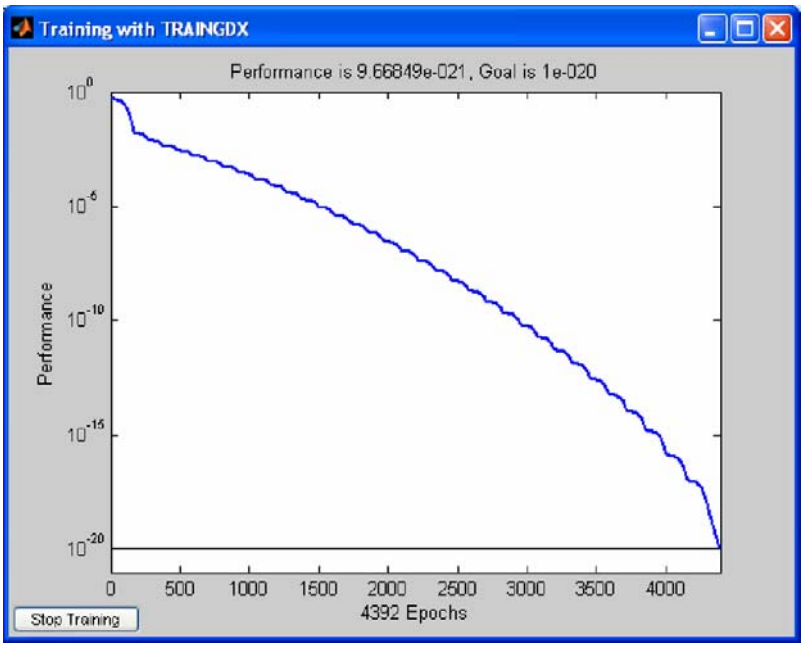
Slika 4.10 *- Sigmoidna funkcija*

Također i u drogom sloju imamo sigmoidnu prenosnu funkciju. Drugi sloj ima 15 neurona, a u trećem sloju imamo 3 neurona sa istom funkcijom prenosa kao u prethodna dva sloja. Prilikom treniranja mreže vršene su različite kombinacije raspodjele broja neurona po slojevima, ali to nije davalo značajnijih rezultata. Isto tako vršene su izmjene funkcija po slojevima mreže, pa su tako pored *logsig* funkcije korištene i *tansig* i *purelin* funkcije. Mreža je trenirana i sa različitim tipovima treninga, kao što su: *trainrp, traingd, traingdx, trainlm* [10].

Klasifikacija je vršena na tri klase, tj. žanra: blues, classical i metal. Svaka klasa je sadržavala po 100 vektora obilježja. Pomoću konfiguracije mreže koja je data kodom:

```
net=newff(minmax(input),[30,15,3],{'logsig','tansig','logsig'},
'traingdx');
net.trainParam.epochs=10000;
net.trainParam.goal = 1e-20;
net.trainParam.min_grad=0;
train(net, input, target);
y =sim(net, input);
```

dobijena je srednjekvadratna greška od 10^{-20} što se može vidjeti na Slici 4.11.



Slika 4.11 *–Traingdx trening*

Svi prethodno opisani postupci nisu dali očekivane i željene rezultate klasifikacije, tj. neuronska mreža se ponašala kao slučajan klasifikator. Dobijali smo veoma malu grešku, ali su rezultati klasifikacije bili različiti u svakom novom pokušaju. Iz tih razloga, odlučili smo se za dalji rad sa *k*-NN (*k*-Nearest Neighbor) klasifikatorom koji je davao zadovoljavajuće rezultate.

4.1.5 Implementacija k-NN klasifikatora

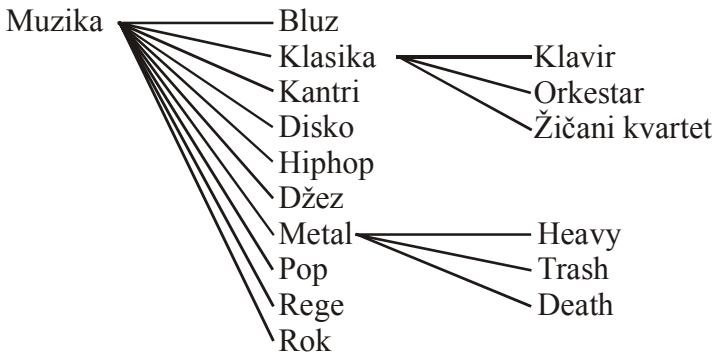
U trećoj glavi dat je opis tri vrste klasifikatora koji se mogu upotrijebiti za klasifikaciju audio zapisa. U radu je implementiran *k*-NN klasifikator. Za razliku od parametarskih klasifikatora, ovaj klasifikator je neparametarski. On direktno koristi trening skup za klasifikaciju bez pretpostavki nekih matematičkih formi funkcija gustine vjerovatnoće klase. Svaki uzorak iz testnog skupa se klasifikuje prema klasi njegovih najbližih susjeda iz trening skupa. U *k*-NN klasifikaroru, *k* susjednih vektora u prostoru obilježja učestvuje u odlučivanju. Do konačne odluke o klasi se dolazi glasanjem. U klasifikaciji korišten je metod "unakrsne provjere" (cross validation) kao što je opisano u trećoj glavi. Obilježja su podjeljena metodom slučajnog odabira u dva skupa, tj. vektora, trening i test vektor. Trening vektor sadrži 90%, a test vektor 10% ukupnih obilježja audio zapisa. Primjenjeno je 100 iteracija tzv. *ten-fold cross-validation* algoritama [14,15]. Konačni rezultati su usrednjeni. Ovo garantuje da izračunata tačnost neće imati velika odstupanja (*bias*) zbog konkretno izabranih trening i test skupova. Naime, u pojedinačnim eksperimentima može se desiti da se klasifikator testira na posebno povoljnom ili posebno nepovoljnom skupu podataka što bi rezultiralo neosnovano dobrim, odnosno, lošim performansama. Izvođenjem više eksperimenata na slučajno izabranim skupovima podataka i usrednjavanjem rezultata dobija se objektivnija slika o performansama klasifikatora. *k*-NN klasifikator je implementiran u MATLAB-u funkcijom *KNN.m* .

4.2 Statistička evaluacija

U ovom odjeljku prikazani su rezultati klasifikacije kolekcije audio zapisa, koja nam je bila na raspolaganju, metodom najbližih susjeda. Prvo je izvršena klasifikacija deset muzičkih žanrova, zatim klasifikacija dva žanra na podžanrove i na kraju je izvršena klasifikacija po važnosti pojedinih grupa obilježja. U posljednjem paragrafu prikazani su rezultati testiranja performansi *k*-NN klasifikatora.

4.2.1 Test kolekcija

Na Slici 4.10 prikazana je hijerarhija muzičkih žanrova korištenih u test kolekciji.

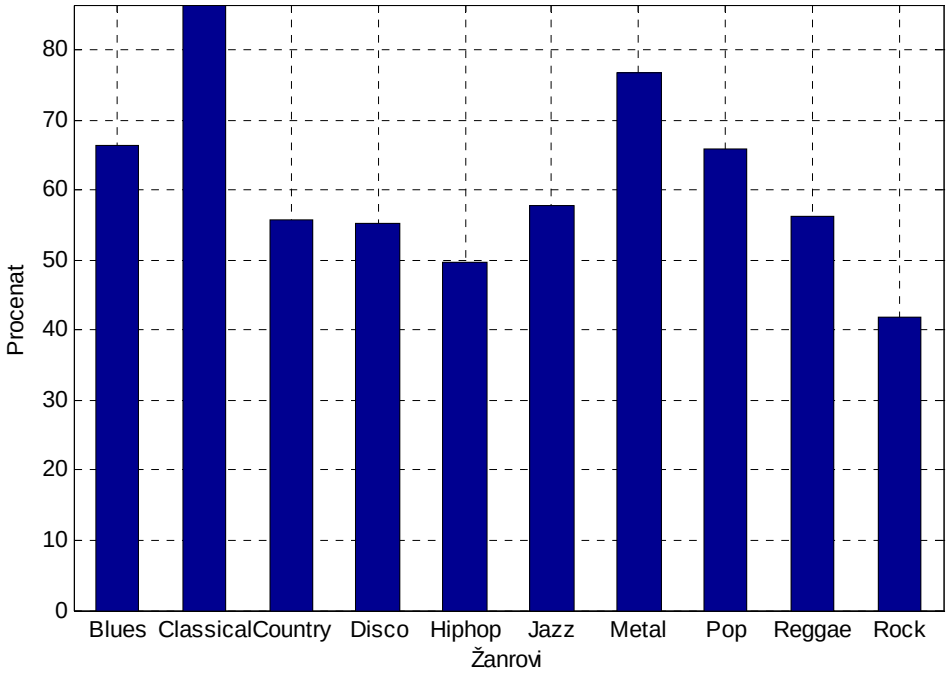


***Slika 4.10**-Hijerarhija muzičkih žanrova*

Test kolekcija se sastoji od 1000 audio zapisa sa muzikom. Svaki audio zapis je dugačak 30 sekundi i snimljen je mono, sa 16 bita i frekvencijom odmjeravanja od 22050Hz. Dakle ukupno ima $10 \cdot 100 \cdot 30s = 8.3h$ audio zapisa. Audio zapisi sadrže muziku koja obuhvata 10 različitih žanrova: bluz (*blues*), klasika (*classical*), kantri (*country*), disko (*disco*), hiphop (*hiphop*), džez (*jazz*), metal (*metal*), pop (*pop*), rege (*reggae*), rok (*rock*). U okviru klasičnog žanra postojie tri klase: klavir (*piano*), orkestar (*orchestra*), žičani kvartet (*string quartet*), kao i u okviru metal žanra: *heavy, trash, death.* Neki od muzičkih primjera su instrumentalni, a neki sadrže i vokale. Korišteni audio zapisi su različitog kvaliteta jer su sakupljani sa CD-a, radia i Web-a. Pored ovih, u radu su korišteni i muzički zapisi pojedinih muzičkih instrumenata, pojedinih tonova, skala, akorda i ritmova, odsviranih u živo ili preuzetih sa Web-a, radi provjere tačnosti implementacije pojedinih muzičkih obilježja.

4.2.2 Rezultati klasifikacije

Na Slici 4.11 prikazan je procenat tačno klasifikovanih muzičkih žanrova. Sa slike se vidi da je procenat tačno klasifikovanih žanrova dosta dobar, oko 61%, kao i to da pojedini žanrovi imaju visoku tačnost. Može se vidjeti da je klasika kao jedinstven i nezavisan žanr separabilnija od ostalih žanrova. Tačnost od oko 90% obećava. Takođe se ističe metal kao jedinstven žanr. Najmanji procenat ima rok žanr, što je logično ako se uzme u obzir njegova sveobuhvatnost.



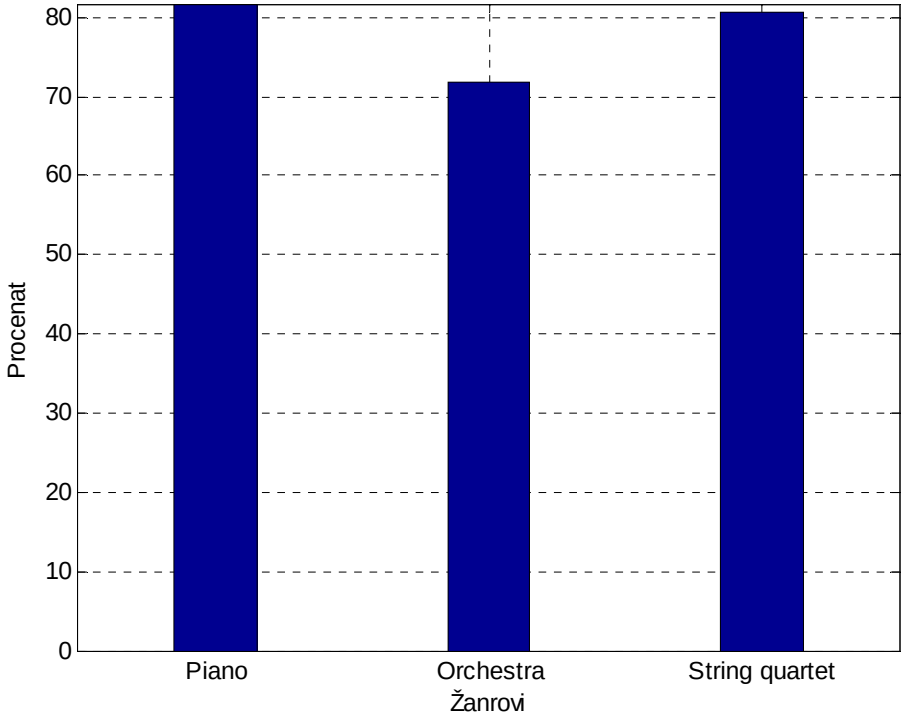
***Slika 4.11**-Klasifikacija žanra*

Tabela 4.1 daje detaljniji uvid u klasifikaciju muzičkih žanrova u formi matrice konfuzija. Kolone matrice odgovaraju aktuelnom žanru, a redovi predviđenom. Na primjer, ćelija u 6. redu i 2. koloni ima vrijednost 7, što znači da je 7% klasične muzike (kolona 2) pogrešno klasifikovano kao džez (red 6). Procenat tačno klasifikovanih žanrova nalazi se duž dijagonale matrice konfuzija. Matrica konfuzija prikazuje da je pogrešno klasifikovanje klasifikatora slično onome što bi i čovjek uradio. Na primjer, klasična muzika je klasifikovana kao džez u kompozicijama koje imaju snažan ritam, od kompozitora kao što su Leonard Bernstein i George Gershwin [4]. Bluz žanr se preklapa sa džezom, rokom i kantrijem, kantri sa džezom i rokom, rege sa hiphopom, itd. Rok žanr ima najmanju tačnost i lako se pomiješa sa ostalim žanrovima što je očekivano zbog prirode samog žanra.

***Tabela 4.1**- Matrica konfuzija žanrova*

	bl	cl	co	di	hi	ja	me	po	re	ro
bl	**67**	1	5	4	6	9	2	2	7	7
cl	0	**87**	2	1	0	12	0	0	0	2
co	8	1	**56**	7	1	13	2	6	6	17
di	3	1	5	**55**	11	1	2	7	5	7
hi	3	0	1	6	**50**	2	5	6	10	1
ja	7	7	9	1	1	**58**	0	3	1	2
me	2	1	2	3	3	0	**77**	1	0	13
po	0	0	1	11	7	1	0	**66**	9	5
re	1	0	4	5	19	0	0	3	**56**	4
ro	9	2	15	7	2	4	12	6	6	**42**

Na Slici 4.12 prikazan je procenat tačno klasifikovanih podžanrova u okviru klasične muzike, a u Tabeli 4.2 data je matrica konfuzija .



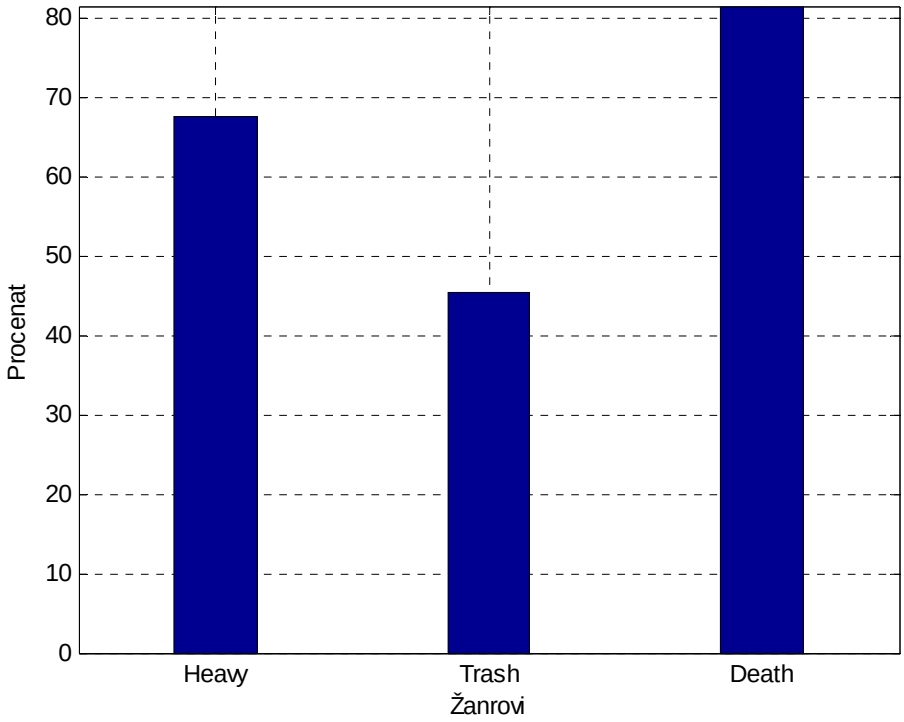
***Slika 4.12**-Klasifikacija klasičnog žanra*

Ukupna tačnost klasifikacije iznosi 78% što je odlično. Iz matrice konfuzija se može vidjeti da je *orchestralna* muzika pogrešno klasifikovana kao *string quartet* u 28% slučajeva, što je očekivano ako se uzme u obzir da se orkestri većinom sastoje od žičanih instrumenata.

***Tabela 4.2**-Matrica konfuzija clasičnog žanra*

	Piano	Orchestra	String Quartet
Piano	**82**	0	8
Orchestra	1	**72**	11
String Quartet	17	28	**81**

Na Slici 4.13 dat je prikaz klasifikacije metal žanra na predložene podžanrove. Matrica konfuzija data je Tabelom 4.3. Ukupna tačnost klasifikacije iznosi 65%.



***Slika 4.13**-Klasifikacija metal žanra*

Može se primjetiti da se izdvaja death žanr kao karakterističan. Ovaj žanr je upečatljiv po specifičnom načinu pjevanja i boji glasa vokala, kao i načinu sviranja i melodici. *Heavy* i *trash* metal se uveliko preklapaju. Može se reći da *trash* sadrži *heavy* kao i obrnuto, jer ipak *heavy metal* je korijen metal muzike.

***Tabela 4.3**-Matrica konfuzija metal žanra*

	Heavy	Trash	Death
Heavy	**68**	51	7
Trash	23	**46**	11
Death	9	2	**82**

Tabela 4.4 prikazuje procenat tačnosti klasifikacije *k*-NN klasifikatora za različite vrijednosti parametra *k* i tri seta muzičkih žanrova. Prvi set, *Genres,* sadrži svih deset žanrova, drugi set *Classical*, sadrži tri podžanra u okviru klasičnog žanra, dok treći set *Metal* sadrži tri podžanra sadržana u metal žanru. Hijerarhija je prikazana na Slici 4.10. Tačnost klasifikacije je data srednjom vrijednosti i standardnom devijacijom. Može se vidjeti da se za *k*=3 dobijaju optimalni rezultati, iako za klasični žanr se dobija nešto veća tačnost pri *k*=1. Bitno je da se za svako *k* dobija veća tačnost od random (po zakonu vjerovatnoće) koja je data u prvom redu tabele.

***Tabela 4.4**-Srednja vrijednost i devijacija tačnosti klasifikacije*

	Genres(10)	Classical(3)	Metal(3)
Random	10	33	33
KNN(1)	58 $\pm$ 1	**78** $\pm$ 5	55 $\pm$ 8
KNN(3)	**61** $\pm$ 1	72 $\pm$ 4	**65** $\pm$ 6
KNN(5)	60 $\pm$ 1	67 $\pm$ 8	54 $\pm$ 6
KNN(7)	60 $\pm$ 1	59 $\pm$ 6	52 $\pm$ 4

Tabela 4.5 prikazuje individualni značaj predloženih skupova obilježja u automatskoj klasifikaciji muzičkih žanrova. Klasifikacija je izvršena za *k*=3. Prvi red u tabeli predstavlja random klasifikaciju, dok poslednji red odgovara kompletnom setu obilježja. Broj u zagradama iza oznake obilježja predstavlja broj obilježja za taj individualni set obilježja. Kao što može da se vidi, obilježja koja nisu bazirana na teksturi, obilježja tonaliteta (Pitch Histogram Features-PHF) i obilježja ritma (Beat Histogram Features-BHF) daju lošije rezultate od obilježja zasnovanih na teksturi (STFT, MFCC) osim za slučaj metal žanra gdje su približno ista. I za džez žanr veću tačnost imaju STFT i MFCC obilježja, što se može vidjeti u [4,6]. Pošto je metal muzika veoma melodična, ritmična, harmonična i brza odatle i veća tačnost pri korištenju obilježja tonaliteta i ritma. U svim slučajevima predloženi skup obilježja daje bolje rezultate od random klasifikacije, što bi značilo da obilježja daju određene informacije o muzičkim žanrovima i muzičkom sadržaju uopšte.

***Tabela 4.5**-Značaj individualnih setova obilježja*

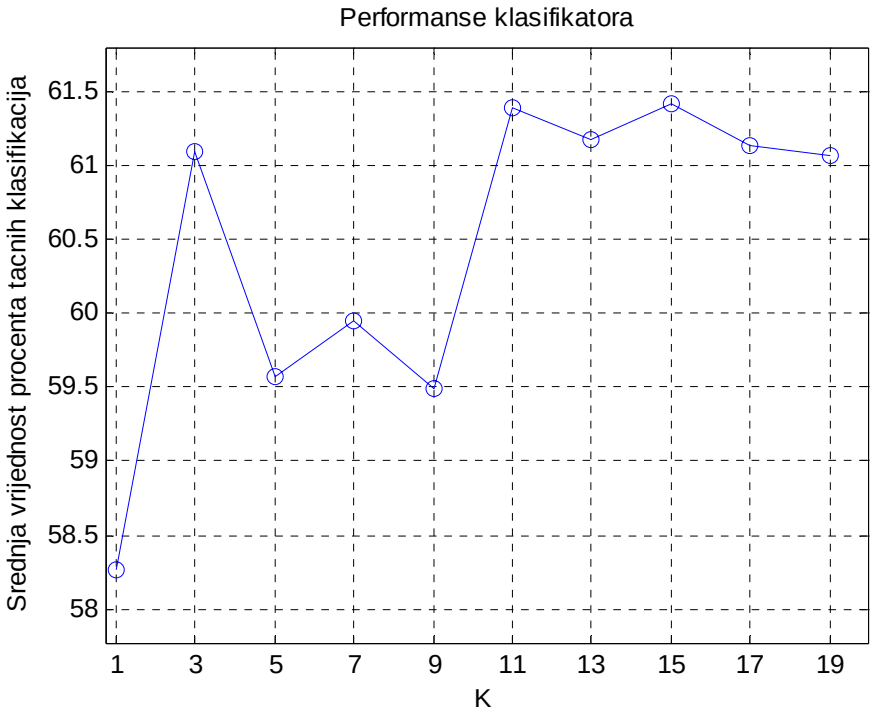
	Genres(10)	Classical(3)	Metal(3)
RND	10	33	33
PHF(5)	35	48	**48**
BHF(6)	24	46	**55**
STFT(9)	**45**	**56**	**44**
MFCC(10)	**59**	**70**	**54**
FULL(30)	61	72	65

Tačnost klasifikacije ukupnog seta obilježja (FULL(30)) u nekim slučajevima nije bitno veća od klasifikacije sa pojedinačnim setovima obilježja (što se vidi i iz tabele). Ova činjenica ne mora da znači da su obilježja međusobno korelisana ili da ne sadrže korisne informacije, jer može se desiti slučaj da se specificirani fajl korektno klasifikuje pomoću dva različita seta obilježja koji sadrže različite i nekorelisane informacije, tj. obilježja. Takođe, iako su izvjesna pojedinačna obilježja korelisana, dodavanje svakog specifičnog obilježja poboljšava tačnost klasifikacije [4,6]. Obilježja zasnovana na ritmu i tonalitetu čini se imaju veću ulogu u klasifikaciji *Classical* i *Metal* seta obilježja u poređenju sa setom *Genres*. Ovo bi moglo da znači da, ako je moguće, treba *Genres* set obilježja specificirati za detaljniju podžanrovsku klasifikaciju, odnosno, treba sve žanrove podijeliti dublje na podžanrove.

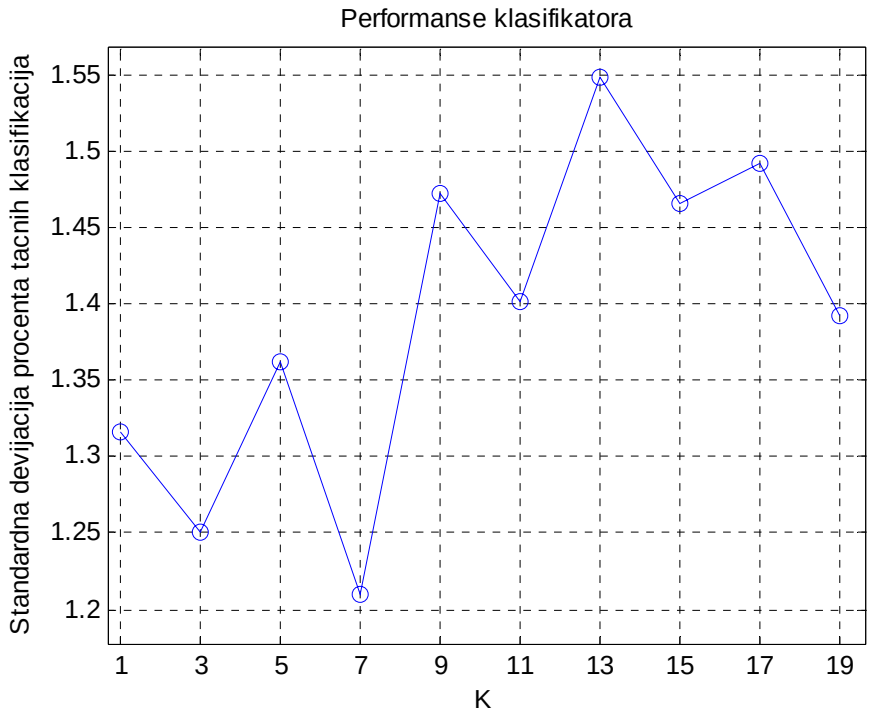
4.2.3 Performanse klasifikatora

Da bismo testirali klasifikator iz raspoloživog skupa primjera slučajno je izabrano 90% primjera koji su korišteni kao trening skup, a preostalih 10% primjera je iskorišteno za validacioni skup. Primjenjen je *10-fold cross validation* algoritam u 100 iteracija za svaku izabranu vrijednost parametra *k*. Uzete su srednja vrijednost i standardna devijacija dobijenih performansi po svakoj iteraciji. Na ovaj način nastojao se izbjeći uticaj konkretno izabranih trening i validacionih skupova na performanse klasifikatora.

Performanse *k*-NN klasifikatora zavise od parametra *k*, koji utiče na glasanje kako je to opisano u paragrafu 3.3. Na Slikama 4.14 i 4.15 prikazane su srednja vrijednost i standardna devijacija procenta tačnih klasifikacija svih žanrova u zavisnosti od vrijednosti parametra *k*. Parametar *k* je neparan broj zbog jednostavnijeg glasanja. Eksperimenti su izvedeni korištenjem svih 19 obilježja (9 obilježja zasnovanih na STFT i 10 mel-skaliranih cepstralnih koeficijenata - MFCC). Može se uočiti da klasifikator nema velika odstupanja od srednje vrijednosti, kao i da je za malo *k* (npr. *k*=1) moguć uticaj šuma na tačnost klasifikacije, dok za veće *k* postoji opasnost da distance koje su udaljenije po izabranoj metrici preglasaju kraće distance.



***Slika 4.14**-Srednja vrijednost procenta tačnih klasifikacija u zavisnosti od parametra k*



***Slika 4.15**-Standardna devijacija procenta tačnih klasifikacija u zavisnosti od parametra k*

5. ZAKLJUČAK

Uprkos nejasnoj prirodi žanrovskih granica, klasifikacija muzičkih audio zapisa po žanru može se izvršiti automatski sa rezultatima tačnosti značajnijim od slučajnih i performansama koje se mogu porediti sa subjektivnom (ljudskom) klasifikacijom. Tri seta obilježja koji predstavljaju teksturu, ritmički i tonski sadržaj muzičkog signala su izračunati i iskorišteni za klasifikaciju muzičkih audio zapisa primjenom statističkog klasifikatora (*k*-NN), koji je testiran sa velikom kolekcijom raznovrsnih audio zapisa. Korištenjem predstavljenog seta obilježja postignuta je ukupna tačnost klasifikacije od 61% na skupu muzičkih audio zapisa podijeljenih na deset žanrova, kao i 78% i 65% za klasifikaciju klasičnog i metal žanra na podžanrove. U radu je takođe prikazan i značaj pojedinačnih setova obilježja u klasifikaciji muzičkih audio zapisa. Pored toga, ispitivane su performanse *k*-NN klasifikatora, odnosno, zavisnost srednje vrijednosti i standardne devijacije procenta tačnih klasifikacija od parametra *k*, koji utiče na glasanje u *Nearest Neighbor* algoritmu. Uspjeh klasifikacije predloženim obilježjima svjedoči o njihovom potencijalu za korištenje i u drugim automatskim tehnikama, kao što su pretraživanje po sličnosti, segmentacija i audio thumbnailing-u.

Budući rad

Za dalji rad svakako bi trebalo izvršiti dodatna poboljšanja obilježja, pa čak i dodavanje novih, kao i raditi na poboljšanju algoritama za njihovo izdvajanje. Iz dosadašnje analize problema klasifikacije muzičkih audio zapisa po žanru očigledno je da je potrebno izvršiti proširenje žanr hijerarhije, kako po širini, tako i po dubini. Takođe u budućim istraživanjima treba obratiti pažnju na druge semantičke deskriptore kao što su emocije i stil pjevanja. Više istraživanja obilježja tonaliteta trebalo bi dovesti do boljih performansi.

Alternativni algoritmi za detekciju tonaliteta, na primjer algoritmi zasnovani na kohlearnim modelima, mogu biti korišteni za kreiranje pič histograma (PH).

Za izračunavanje bit histograma (BH) planira se istraživanje novih filtar banki (*front-ends*). Takođe, već postojeća istraživanja na temu klasifikacije u realnom vremenu pomoću obilježja teksture [4], mogla bi se poboljšati izračunavanjem obilježja tonaliteta i ritma u realnom vremenu. Zanimljiva mogućnost je i izdvajanje sličnih obilježja direktno iz MPEG audio kompresovanih podataka.

Posjedujući razdvojene skupove obilježja za predstavljanje teksture, ritma, tonaliteta i harmonije, moguće su implementacije različitih tipova pretraživanja po sličnosti. Dva dodatna izvora informacija o muzičkom žanru su melodija i glas izvođača. Iako je izdvajanje melodije težak zadatak koji nije riješen generalno za audio, moguće je iskoristiti neke statističke informacije čak i iz nepotpunih algoritama za izdvajanje melodije. Izdvajanje i analiza glasa izvođača tema je kojom se treba baviti u budućem radu.

Primjetno je da su istraživanja evoluirala iz isključivo prostih mašinskih izračunavanja u tehnike gdje učenje, trening skup podataka i prethodno znanje snažno utiču na performanse i rezultate. Ovo je posebno izraženo za klasifikaciju muzičkog žanra, koja je uvijek bila pod uticajem iskustva, pozadine i ponekad ličnog osjećaja. Pored muzike, i u nekoliko drugih klasifikacionih domena, povezanih sa muzikom ili ne, postoje mnogi neizvršeni zadaci gdje mašinsko učenje zajedno sa obradom signala igra glavnu ulogu.

6. PRILOG

Lista oznaka i skraćenica:

ANNs - Artifical Neural Networks
BH - Beat Histogram
BHF - Beat Histogram Features
CD – Compact Disc
DCT - Discrete Cosine Transform
DFT - Discrete Fourier Transform
DWT - Discrete Wavelet Transform
EMT - Explicite Time Modeling
ESACF - EnhancedSummary AutoCorrelation Function
FFT - Fast Fourier Transform
FPH - Folded Pitch Histogram
GMM - Gaussian Mixture Model
HMMs - Hiden Marcov Models
IDFT - Inverse Discrete Fourier Transform
ISP - Intelligent Sound Implementation
k-NN – *k*-Nearest Neighbor
MATLAB - MATrix LABoratory
MFCCs - Mel Frequency Cepstral Coeficients
MIDI - Musical Instruments Digital Interface
MLP - MultiLayer Perceptron
MPEG – ISO/IEC Moving Pictures Experts Group
NGNs - New Generation Networks
PHF - Pitch Histogram Features
PH - Pitch Histogram
RMS - Root Mean Square

RND - Random
SACF - Summary AutoCorrelation Function
STFT - Short Time Fourier Transform
UPH - Unfolded Pitch Histogram
WT - Wavelet Transform

LITERATURA

[1] R. Dannenberg, J. Foote, G. Tzanetakis and C. Weare, ''Panel: new directions in music information retrieval'', in *Proc. Int. Computer Music Conf.*, Habana, Cuba, Sept. 2001

[2] Nicolas Scaringella, Giorgio Zoila and Daniel Mlynek. Automatic Genre Classificatin of Music Content. IEEE Signal Processing Magazine,133, March 2006

[3] F. Pachet and D. Cazaly, ''A taxonomy of musical genres'', in *Proc. Content-Based Multimedia Information Access (RIAO)*, Paris, France, 2000

[4] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. IEEE Transactions on Signal Processing, Vol. 10, No.5, July 2002.

[5] Vladimir Risojević, Klasifikacija audio signala na govorne i muzičke. Seminarski rad, Elektrotehnički Fakultet, Banja Luka. Januar, 2005.

[6] George Tzanetakis. Manipulation, analysis and retrieval systems for audio signals. A Dissertation Presented to the Faculty of Princeton University in Candidacy for the Degree of Doctor of Philosophy. June 2002.

[7] Paul Scott, Music Classification using Neural Networks, EE 373B Project, Prof. Bernard Widrow, Spring 2001

[8] Sigurdur Sigurdsson, Kaare Brandt Petersen and Tue Lehn-Schiøler, Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music. Informatics and Mathematical Modelling,Technical University of Denmark,Richard Petersens Plads - Building 321, DK-2800 Kgs. Lyngby - Denmark

[9] Dejan Despić, Teorija Muzike, Zavod za udžbenike, Beograd 2007.

[10] MATLAB Help (Wavelet Toolbox, Neural Networks Toolbox)

[11] Tero Tolonen and Matti Karjalainen. A Computationally Efficient Multipitch Analysis Model. IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 6, November 2000.

[12] H. Soltau, T. Schultz, M. Westphal, and A. Waibel, "Recognition of music types," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing (ICASSP),* Seattle, WA, USA, 1998, vol. II, pp. 1137–1140.

[13] Georgio Tzanetakis, Andrey Ermolinskyi and Perry Cook, Pitch Histograms in Audio and Symbolic Music Information Retrieval, Computer Science Department 35 Olden Street Princeton NJ 08544 +1 609-258-5030

[14] Dr Chris Bryant, Cross Validation, Data Mining (CMM510), http://www.comp.rgu.ac.uk/staff/chb/teach.html

[15] Intelligent Sensor Systems, Lecture 13: Validation, Ricardo Gutierrez-Osuna,Wright State University

[16] Ana Bogdanić, Ivana Buklijaš, Krešimir Mudrovčić, Nika Parađina, Određivanje tempa i tonaliteta u glazbi. Projekt iz predmeta Slučajni procesi u sustavima. Zagreb, 2006.

[17] http://en.wikipedia.org/wiki/Tempo#Measuring_Tempo , http://en.wikipedia.org/wiki/Beats_per_minute, posjećeno: Septembar 2007.