

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Sašo Brus

**Prepoznavanje akordov s skritim
markovskim modelom**

DIPLOMSKO DELO

UNIVERZITETNI ŠTUDIJSKI PROGRAM PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Matija Marolt

Ljubljana, 2013

Rezultati diplomskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavlanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

Besedilo je oblikovano z urejevalnikom besedil \LaTeX .



Št. naloge: 00099/2013

Datum: 10.04.2013

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **SAŠO BRUS**

Naslov: **PREPOZNAVANJE AKORDOV S SKRITIM MARKOVSKIM MODELOM
CHORD RECOGNITION WITH A HIDDEN MARKOV MODEL**

Vrsta naloge: Diplomsko delo univerzitetnega študija prve stopnje

Tematika naloge:

V diplomski nalogi raziščite področje prepoznavanja akordov v glasbenih signalih. Osredotočite se na postopke, ki za prepoznavanje uporabljajo skrite markovske modele in implementirajte tovrsten algoritem. Algoritem preizkusite na anotirani zbirki posnetkov.

Mentor:

doc. dr. Matija Marolt

Dekan:

prof. dr. Nikolaj Zimic



IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Sašo Brus, z vpisno številko **63080055**, sem avtor diplomskega dela z naslovom:

Prepoznavanje akordov s skritim markovskim modelom

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom doc. dr. Matije Marolta,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki "Dela FRI".

V Ljubljani, dne 9. septembra 2013

Podpis avtorja:

Hvala.

Kazalo

Povzetek

Abstract

1	Uvod	1
1.1	Pregled področja in potek dela	4
2	Glasbena teorija	9
2.1	Elementi glasbe	9
2.2	Uglasitev	11
2.3	Intervali	11
2.4	Lestvice	12
2.5	Akordi	15
3	Digitalizacija in zaznavanje zvoka	19
3.1	Človeški sluh in zaznavanje zvoka	20
3.2	Diskretna Fourierova transformacija	21
4	HMM - Skriti markovski model	27
4.1	Definicija	27
4.2	Gradniki modela	29
4.3	Enačba modela HMM	32
4.4	Algoritem Viterbi	32

KAZALO

5 Implementacija	35
5.1 Predpostavke in omejitve	35
5.2 Vhodni podatki modela	37
5.3 Kromagram	38
5.4 Učenje	41
5.5 Evalvacija	49
5.6 Analiza rezultatov	53
6 Zaključek	59

Povzetek

V diplomski nalogi je predstavljen sistem za samodejno prepoznavo akordov podane skladbe. Sistem temelji na Skritem markovskem modelu – HMM. Elementi modela HMM so vizualno predstavljeni. Za vrednotenje stanj je uporabiljena metrika kromagram. Predstavljena sta postopka učenja in evalvacije. Naš sistem je naučen in testiran na bazi glasbenih notacij Isophonics. Z uporabo 10-kratnega prečnega preverjanja dosega naš sistem 62% klasifikacijsko točnost. Uporabljen abeceda akordov vsebuje 25 stanj akordov. Opisani so razlogi za dosežene rezultate. Opravljena je podrobna analiza prepoznavne. Naš pristop združuje tudi znanja iz področij glasbene teorije in psihoakustike. Vse uporabljene metode so argumentirane in primerjane z modernimi sistemi. Predstavljene so možnosti za izboljšanje klasifikacijske natančnosti.

Ključne besede: akord, hmm, mirex, fft

Abstract

In this paper a system for automatic chord estimation of an input song is presented. Our system is based on a Hidden Markov model – HMM. Visual representation of HMM elements is offered. Metric called Chromagram is used for evaluation of system states. Learn and evaluation processes are presented. Our system learns rules and performs evaluation on Isophonics musical database. Our system achieves 62% classification accuracy using 10-fold validation. Chord alphabet, used in our model, contains 25 chord states. We present reasons for achieved results and perform detailed estimation analysis. Our approach contains knowledge of music theory and psychoacoustics. All methods, used in our system are argued and compared with modern systems. Further, some options for improving classification accuracy are presented.

Keywords: chord, hmm, mirex, fft

Poglavje 1

Uvod

*Thank you for the music, the songs I'm singing
Thanks for all the joy they're bringing
Who can live without it, I ask in all honesty
What would life be?
Without a song or a dance what are we?
So I say thank you for the music
For giving it to me.*

Abba, 1977

“Kaj bi bilo življenje brez glasbe?”, se sprašuje skupina Abba v uspešnici Thank You for the Music iz leta 1977. Kaj takšnega si danes težko predstavljamo, saj nas glasba spremlja na vsakem koraku. Oglaševalci glasbo s pridom uporabljajo za pospeševanje prodaje, otroci si s pomočjo enostavnih kompozicij lažje zapomnijo abecedo, mnoge glasba sprošča, jim izboljšuje razpoloženje. To je pogled na glasbo skozi oči modernega človeka. Glasba je seveda veliko starejša od oglasov in abecede, vendar njen vpliv na človeka in družbo ostaja nespremenjen.

Začetki glasbenega ustvarjanja segajo okoli 50.000 let v preteklost. V to obdobje spada slavna piščal iz jame Divje Babe, ki naj bi veljala za najstarejše glasbilo na svetu [1]. Številne raziskave so pokazale, da razmerja med

luknjami ustrezajo razmerjem med nekaterimi toni molove lestvice. Replike Turka in Dimkaroskega so omogočale razpon 2 oktav [2]. Najstarejši zapisi glasbe segajo v obdobje okoli 2000 let pred našim štetjem. Zapis na glineni plošči, najdeni na ozemlju današnjega Iraka, vsebuje navodila za izvajanje glasbe. Raziskovalci so ugotovili, da vsebuje fragmente diatonične lestvice in trozvokov. Začetki uporabe notnega črtovja, kot ga poznamo danes, segajo v zgodnji srednji vek. Uporabljalo se je 4 vrstično črtovje, še vedno pa niso imeli načina za zapis ritma. Moderni 5 vrstični sistem (Slika 1.1) je v uporabi od 16. stoletja naprej.

Prelude
Op. 28, No. 7

Frederic Chopin

The image shows a musical score for Chopin's Prelude Op. 28, No. 7. It is written for piano and consists of two systems of staves. The first system includes the treble and bass clefs, with a tempo marking of 'Andantino' and a dynamic of 'p dolce'. The second system continues the piece, ending with a 'rit. e dim.' marking and a final 'pp' dynamic. The score is in G major and 3/4 time.

Slika 1.1: Notno črtovje; Uvod v Prelude.

Leta 1977 je NASA v vesolje poslala sondi Voyager 1 in 2. Na zunanjo stran trupa so pritrdili zlato gramofonsko ploščo, ki naj bi morebitni inteligentni entiteti predstavila človeško raso. Na ploščo so shranili fotografije, pozdrave v mnogih jezikih, razne posnetke in navodila za predvajanje. Plošča vsebuje 90 minut glasbe različnih kultur. Najbolj zastopan avtor je Johann Sebastian Bach. Njegove kompozicije spadajo med najbolj matematično definirane, kar je po mnenju nekaterih močno vplivalo na repertoar plošče [3]. V dobi računalništva je matematično ozadje glasbe dobilo poseben pomen. Tehnologija omogoča analizo ogromnih količin podatkov. Iz rezultatov analiz lahko npr. iščemo podobnosti med skladbami, kar je posebej zanimivo pri zaščiti avtorskih pravic. Baze se uporabljajo za učenje inteligentnih siste-

mov, ki znanje nato uporabijo pri odločanju, analizi, klasifikaciji. Primera takšnih sistemov sta Shazam in Songsmith.

Shazam

Shazam je storitev za mobilne naprave, ki uporabniku omogoča identifikacijo naslova in avtorja skladbe. Aplikacija z uporabo mikrofona zajame del skladbe, iz katerega napravi spektrogram. Z uporabo zgoščevalne funkcije izlušči značilnice posnetka, ki se na strežniku primerjajo z značilnicami v podatkovni bazi. Če pride do ujemanja, je uporabnik obveščen o naslovu skladbe.

Microsoft Songsmith

Microsoft Research je pripravil orodje, ki na podlagi posnete melodije samodejno zgradi spremljavo – izbere ustrezne akorde in ritem. Jedro programa je statistični model, ki so ga zgradili iz analize približno 300 skladb. Model se je naučil medsebojne korelacije med akordi, prav tako pa tudi povezav med akordi in melodijo. Program je primeren tako za začetnike, kot za umetnike, ki nočejo, da se jim izmuzne dobra ideja.

MIREX

MIREX (Music Information Retrieval Evaluation eXchange) je mednarodna konferenca, ki poteka znotraj dogodka ISMIR (International Conference on Music Information Retrieval). MIREX ponuja možnost evalvacije in primerjave algoritmov za pridobivanje informacij iz glasbe. Tipične pridobljene informacije so klasifikacija glasbe, prepoznavna skladbe, iskanje podobnosti med skladbami, sledenje ritma, iskanje skladbe z mrmranjem in klasifikacija akordov. Letošnje tekmovanje MIREX je 8. po vrsti.

1.1 Pregled področja in potek dela

Zaradi vedno večje procesorske moči, pa tudi zaradi vse večje prisotnosti računalniške tehnologije v vsakdanjem življenju, postaja pridobivanje informacij iz glasbe vedno bolj aktualno. Pridobljene informacije so lahko popolnoma tehničnega značaja – analiza zvočnega posnetka omogoča učinkovito kompresijo, lahko pa tudi bolj abstraktne – iskanje plagiatov, podobnosti med skladbami. Glasba ponuja mnogo informacij, zato je ustrezen pristop do problema ključnega pomena. Za izgradnjo sistema, kot je Shazam, potrebujemo učinkovit način pridobitve značilnic posnetka in možnost hitrega iskanja po podatkovni bazi. Pri implementaciji sistema, kot je Songsmith, je potrebno na podlagi učnih podatkov zgraditi bazo znanja, s pomočjo katere ustrezno obravnavamo neznane vhodne podatke. Za klasifikacijo žanra glasbe moramo iz učne množice izluščiti tiste značilnice, ki so skupne določenemu žanru. Med improvizatorji in izvajalci glasbe, kot je jazz, je pogosto edini zapis oziroma način deljenja glasbe t.i. Lead Sheet. Zapis vsebuje zaporedje akordov in osnovno melodijo, ostale podrobnosti pa so prepuščene interpretaciji izvajalca. Samodejna segmentacija glasbe bi lahko olajšala izdelavo takšnih notacij.

V zadnjem desetletju se je na področju prepoznave akordov razvilo mnogo različnih pristopov. Fujishima je v članku [20] iz leta 1999 prvi opisal metriko kromagram. Metriki se bomo posvetili v nadaljnjih poglavjih, omenimo pa, da je kromagram skupni imenovalec večine modelov za klasifikacijo akordov. Predstavljamo si ga lahko kot meritev intenzitete posameznih tonov v določenem časovnem odseku. Osnovni gradniki sistema za klasifikacijo akordov so naslednji:

- Gradnja baze znanja.
- Gradnja kromagram-a ali podobne metrike.
- Uporaba statističnega modela za izbiro najprimernejšega kandidata za podan kromagram (metriko).

Fujishima v članku [20] predlaga uporabo STFT (kratkotrajne Fourierve transformacije) za gradnjo kromagramov. Avtorji članka [10] so pokazali, da lahko CQT (konstantna Q transformacija) prinese primerljive oziroma celo boljše rezultate, kot STFT. Razlog tiči v boljši ločljivosti CQT pri nižjih frekvencah.

Pogost problem gradnje kromagramov je določitev dolžine kromagramov in s tem količine le-teh. Veliko sistemov uporablja fiksno določene dolžine kromagramov, ki so dovolj kratke, da prehodi med akordi ne povzročajo prevelikih napak meritev. Dolžine navadno znašajo od nekaj 10 do nekaj 100 ms. Takšni sistemi so občutljivi na lokalni šum in variacije akordov [16]. Obstaja več načinov za rešitev omenjenega problema; avtorji članka [15] navajajo 2 možnosti. V sistem se uvede zaznava ritma. Menjava akordov se večinoma izvede ob dobi, zato je sklep, da je med dvema dobama prisoten samo 1 akord, smiseln. Kromagrame med 2 dobama lahko torej obravnavamo kot celoto. Če generiramo kromagrame fiksnih dolžin, je potrebno ugotoviti, kateri spadajo skupaj in izračunati povprečje. Druga možnost je izdelava kromagrama, ki obsega celoten čas med dvema dobama.

Avtorji članka [17] predlagajo uporabo alternativne metrike. Metrika, ki jo imenujejo tonski centroid, je projekcija kromagrama na 6-dimezijski objekt. Terčna in kvintna razmerja se v objektu odražajo kot majhne evklidske razdalje, kar je osnova za detekcijo sprememb akordov. Uporaba 2 ločenih kromagramov za bas in višje tone se je izkazala za uspešno alternativo enemu samemu kromagramu [10]. Model z dvema ločenima metrikama namreč bolje obravnava obrate akordov in spremembe v frekvenčnem območju basa. Zaznava ključa – tonalitete, iz katere izhaja skladba se je izkazala za uporabno metodo [10]. Razloge za to bomo opisali v poglavju o glasbeni teoriji. Pri modelu, ki poleg akordov zaznava tudi tonaliteto, se prehodi med akordi obravnava v odvisnosti od tonalitete, kar posledično zmanjša nabor vseh možnih kandidatov za posamezni kromagram.

Problem učenja modela iz omejenega nabora skladb je v tem, da takšen model dobro klasificira samo podatke, ki so na nek način podobni tistim, iz

katerih se je učil. Avtorji člankov [18] in [19] predlagajo uporabo modela, ki temelji na bazi ekspertnega znanja. Takšen sistem je nepristranski, saj ne temelji na omejeni množici učnih podatkov. Model, ki temelji na učni množici lahko vsebuje kvečjemu toliko znanja, kolikor ga vsebuje učna množica. Pri evalvaciji lahko pride do problema, ko sistem napačno obravnava vhodne podatke, ki niso bili vsebovani v učni množici. Sistemi, ki se učijo na učni množici MIREX, imajo zato pogosto omejen nabor akordov, ki jih zaznavajo. Abeceda akordov pri evalvaciji MIREX sestoji iz 12 durovih in 12 molovih akordov. Model iz članka [18] vsebuje, zaradi ekspertnega značaja, večji nabor akordov. Poleg durovih in molovih vključuje še 12 zvečanih in 12 zmanjšanih trozvokov.

V poglavju o človeškem sluhu in zaznavanju zvoka bomo omenili posebnosti človeškega sluha. Avtorji članka [15] predlagajo uporabo t.i. Loudness kromagrama, ki izkorišča lastnosti človeškega slušnega organa. Predlagajo uporabo psihoakustičnega filtra, ki signal v frekvenčni domeni ustrezno normalizira z ozirom na človeški sluh.

Cilj diplomske naloge je opis in izdelava modela za samodejno klasifikacijo akordov podane skladbe. V 2. poglavju bomo predstavili vso potrebno glasbeno teorijo, na kateri temeljijo predpostavke in zmožnosti našega modela. Najprej bomo predstavili osnovne glasbene izraze, uglasitev in intervale. Nekaj besed bomo posvetili lestvicam in razložili, kako iz njih izvirajo akordi. Predstavili bomo osnovne tipe akordov, njihove notacije in razmerja med njimi.

V 3. poglavju se bomo posvetili digitalizaciji. Navedli bomo osnovne digitalne formate, postopek digitalizacije in nekaj besed namenili psihoakustiki. Predstavili bomo osnovno orodje našega dela, to je Fouriereva transformacija. Razložili bomo matematično ozadje in algoritem za hitri izračun transformacije – FFT.

V 4. poglavju bomo predstavili teoretično ozadje statističnega modela HMM – skritega markovskega modela. Razložili bomo osnovno idejo in de-

lovanje predstavili na preprostem primeru. Predstavili bomo vse gradnike modela, njihove posebnosti in mesto pri učenju in evalvaciji. Zapisali bomo enačbo HMM in predstavili algoritem Viterbi, s pomočjo katerega bo naš sistem izdelal zaporedje akordov podane vhodne skladbe.

V 5. poglavju se bomo osredotočili na praktično implementacijo modela. Predstavili bomo lastnosti in omejitve modela, delovno okolje in uporabljena orodja. Navedli bomo lastnosti vhodnih podatkov in predstavili metriko kromagram. Pokazali bomo izvedbo psihoakustičnega filtra v praksi. Navedli bomo diagrama poteka procesov učenja in evalvacije. Vsak korak algoritma bomo posebej razložili in potegnili vzporednice z glasbeno teorijo. Razložili bomo delovanje okenskih funkcij in predstavili gradnike HMM tudi v praktičnem smislu. Bazo znanja bomo predstavili v grafični obliki in razložili posebnosti in odstopanja. Opravili bomo evalvacijo modela z uporabo 10-kratnega prečnega preverjanja. Navedli bomo postopek ocenjevanja in dosežene rezultate. Z ustreznimi postopki bomo analizirali rezultate, tako na nivoju celotnega modela, kot na nivoju posameznih skladb. V luči doseženih rezultatov bomo navedli in opisali možne izboljšave modela.

Poglavje 2

Glasbena teorija

2.1 Elementi glasbe

Organizirano, urejeno in oblikovano vrsto tonov, zvenov, šumov, ropota in vseh drugih umetno povzročenih zvočnih pojavov imenujemo glasba [9]. Zvok nastane s prenosom nihanja s prožne snovi na zrak. Prožna snov – glasbilo je lahko struna, koža, opna, plošča, zračni steber. Preprosti (sinusni) ton povzročajo glasbene vilice. Ton je vsekakor najpomembnejši element glasbe. Ker je tudi osnova za naše nadaljnjo delo, si ga pogledjimo podrobneje.

2.1.1 Ton - zven

Ker preprosti (sinusni) ton nima barve, je na tem mestu bolj ustrezno poimenovanje zven. Zven je obarvani (preprosti) ton. Je periodično nihanje, ki vsebuje več sinusnih krivulj. Ton – zven opredelimo z naslednjimi lastnostmi:

- Višina
- Moč
- Barva
- Trajanje

2.1.2 Višina

Višina tona je določena s številom nihajev na sekundo – frekvenco. Enota za frekvenco je Hertz (Hz). Višja frekvenca pomeni višji ton, nižja nižji ton. Nadzorovano spreminjanje višine tona je osnovna naloga večine glasbenih inštrumentov.

2.1.3 Moč

Moč tona je v veliki meri odvisna od inštrumenta. Inštrumenti z ojačevalniki lahko dosežejo bistveno višje amplitude, kot tisti brez njih. Prav tako je potrebno upoštevati karakteristike človeškega sluha. Za nekatere tone smo zaradi evolucije bolj dovzetni, kot za druge. Izmerjena amplituda takšnega tona je lahko enaka ali celo manjša od “nedovzetnega” tona, vendar “dovzetnega” dojamemo kot glasnejšega.

2.1.4 Barva

Barva tona je lastnost, zaradi katere ločimo isti ton zaigran na dveh različnih vrstah inštrumentov. Frekvenca 440Hz, ki ustreza tonu A4, bo npr. slišati drugače na klavirju, kot na kitari. Do razlik pride zaradi različnih tipov zvočil. Zvočila so npr. strune, cevi, napete kože,... Tudi način igranja vpliva na barvo. Brenkanje po struni povzroči drugačno barvo, kot udarjanje po njej. Barva je posledica višjih harmonskih komponent osnovnega tona, ki jih vsak inštrument oddaja drugače. Ker so te amplitude teh komponent precej manjše, slišimo samo osnovni ton, vplivajo pa na barvo celotne percepcije.

2.1.5 Trajanje

Trajanje tona je lastnost, ki pove kako dolgo bo ton zvenel. Na kvaliteto zvena nima posebnega vpliva. Pri notaciji se za definiranje trajanja uporablja različne notne simbole.

2.2 Uglasitev

Za absolutno uglasitev inštrumentov moramo določiti izhodišče. Izhodiščni ton imenujemo komorni ton in je splošno znan. Danes se najpogosteje uporablja ton A4, ki ustreza frekvenci 440 Hz. Komorni ton je bil v zgodovini že večkrat spremenjen. Nekateri orkestri, pa tudi drugi izvajalci se poslužujejo drugih izhodiščnih tonov. Nestandardna absolutna uglasitev zahteva pri implementaciji algoritma za izločanje značilnic posebno pozornost. Več o uglasitvi in implementaciji bomo navedli v 5. poglavju.

Poleg absolutne uglasitve poznamo tudi relativno. Le-ta določa razmerja med posameznimi toni. Poznamo naravno in temperirano uglasitev. Pri naravni uglasitvi so toni v kvintnem oziroma kvartnem razmerju, pri temperirani uglasitvi pa je vsaka oktava razdeljena na 12 poltonov. Razmerja med poltoni so v vsaki oktavi enaka. Temperirana uglasitev povzroča izenačenje eharmoničnih tonov (cis = des), kar pri naravni uglasitvi ne velja. Temperirana uglasitev je osnova za diatonično funkcijo, iz katere izhajata najpomembnejši lestvici – dur in mol.

2.3 Intervali

Razmerje med dvema tonoma, ki zvenita zaporedoma ali hkrati, imenujemo interval. Razmerje med zaporednima tonoma imenujemo melodični interval, razmerje hkrati zvenceh tonov pa imenujemo harmonični interval. Intervale znotraj ene oktave poimenujemo glede na število stopenj med tonoma:

- prima: 1 stopnja (c – c)
- sekunda: 2 stopnji (c – d)
- terca: 3 stopnje (c – e)
- kvartet: 4 stopnje (c – f)
- kvinta: 5 stopenj (c – g)
- seksta: 6 stopenj (c – a)

- septima: 7 stopenj (c – h)
- oktava: 8 stopenj (c – c2)

2.4 Lestvice

Lestvica je množica različnih tonov, ki največkrat obsega eno oktavo. Toni, ki spadajo v lestvico so določeni s pravilom. Večinoma se pravilo – nabor intervalov ponovi v višjih in nižjih oktavah. Lestvico določata nabor intervalov in izhodiščni ton – tonika. V osnovi se lestvice delijo takole:

- tonalne lestvice (dur in mol)
- modalne lestvice
- ljudske lestvice
- umetne lestvice

2.4.1 Kromatična lestvica in poltoni

Kromatična lestvica je sestavljena iz vseh tonov znotraj oktave. Stopnje so konstantne in znašajo pol tona. Kromatična lestvica se vedno uporablja kot del tonalne lestvice, zato imajo stopnje tonalne lestvice znotraj kromatične lestvice posebno vlogo. Kromatično lestvico (z višaji) lahko zapišemo takole:

c cis d dis e fis g gis a b h

2.4.2 Tetrakord

Tetrakord je gradnik tonalnih lestvic. Sestavljajo ga 4 toni. Razdalje med njimi so lahko polton, ton ter ton in pol. V posamezni tonalni lestvici nastopata 2 tetrakorda. Z določitvijo intervalov znotraj tetrakorda in razdalje med obema tetrakordoma določimo pravilo za tonalno lestvico.

2.4.3 Durova lestvica

Durova lestvica je diatonična lestvica, ki jo sestavljata 2 enaka tetrakorda. Osmi ton je vedno oktavna ponovitev prvega. Razdalja med tetrakordoma je celotonska. Pravilo za gradnjo durove lestvice lahko podamo na naslednji način:

$$C C P C C C P \text{ (} C - \text{ cel ton, } P - \text{ polton)}$$

Pravilo pomeni razdalje med posameznimi toni lestvice. Za durovo lestvico sta značilni poltonski razdalji med 3. in 4. stopnjo, ter med 7. in 8. stopnjo. Sedma stopnja (tudi vodilni ton) zaradi poltonske razdalje predvideva oz. vodi v osmi ton. Pri poslušanju durovske skladbe zato “začutimo” kdaj se skladba zaključi. C-dur lestvico po zgornjem pravilu sestavimo takole:

$$c d e f g a h c$$

Vsako durovo lestvico lahko zgradimo po enakem postopku. Vsako naslednjo lestvico pa lahko zgradimo tudi z uporabo tetrakordov. Iz C-dur zgradimo G-dur tako, da ohranimo desni tetrakord, levega pa prestavimo nad desnega. Pri tem postopku je potrebno zvišati 7. stopnjo za pol tona. Nastane durova lestvica z enim višajem, poznana kot G-dur. Če postopek ponovimo, dobimo lestvico z 2 višaji – D-dur. Komplementaren postopek uporabimo za lestvice z nižaji. Število višajev / nižajev enolično določa lestvico in je v notnem črtovju navedeno ob violinskem ključu (Slika 1.1).

2.4.4 Molova lestvica

Za vsako modalno lestvico obstaja durov oziroma molov komplement. Pravimo, da sta lestvici vzporedni. Vsak molov komplement teče za malo terco pod durovo lestvico. Prav tako vsak durov komplement teče za malo terco nad molovo lestvico. Lestvico, ki jo dobimo kot rezultat zamika durove lestvice za 3 stopnje (malo terco), imenujemo naravna molova lestvica. Vsebuje iste tone, kot vzporedna durova lestvica. Molovo lestvico sestavljata 2

različna tetrakorda. Drugi (desni) tetrakord nastopa v 3 različicah, zato poznamo 3 molove lestvice: naravno, harmonično in melodično. Prvi tetrakord je konstanten in ga imenujemo molov tetrakord. Naravno molovo lestvico zgradimo po naslednjem pravilu:

$$C P C C P C C \text{ (} C - \text{ cel ton, } P - \text{ polton)}$$

Če po zgoraj opisanem pravilu zgradimo lestvico začenši s tonom A, dobimo naravno lestvico A-mol:

$$a h c d e f g a$$

Opazimo, da je molova lestvica ciklični zamik durove za 3 stopnje v desno, kar ustreza zgornjemu razmisleku. Vzporednost dura in mola je podlaga za razumevanje in uporabo kvintnega kroga.

Naravna molova lestvica zaradi celotonske razdalje med 7. in 8. stopnjo nima vodilnega tona. Zato je naravni molov sklep nekoliko nedefiniran – ne vodi jasno v toniko. To pomanjkljivost so poskušali odpraviti z zvišanjem 7. stopnje za pol tona. S tem dosežemo poltonsko razdaljo med 7. in 8. stopnjo – vodilni ton. Tako spremenjeno lestvico imenujemo harmonična molova lestvica. Lestvica se imenuje harmonična zaradi pogoste uporabe pri gradnji akordov – harmonij v molu. Zvišanje 7. stopnje za pol tona povzroči povečanje razdalje med 6. in 7. stopnjo. Popravljen interval znaša 3 poltone in zveni nenavadno, zato se zelo redko uporablja.

Nenavadnem intervalu med 6. in 7. stopnjo harmonične lestvice se skuša izogniti melodična molova lestvica. Melodično lestvico dobimo, če harmonični lestvici zvišamo še 6. stopnjo. S tem se izognemo zvišani sekundi med 6. in 7. stopnjo, saj je razdalja sedaj celotonska. Tako predelan drugi tetrakord je enak durovemu. V praksi se melodična lestvica večinoma uporablja v smeri navzgor, v smeri navzdol pa se nadomesti z naravno molovo lestvico.

2.4.5 Pentatonika

Pentatonika (penta = pet) je lestvica, sestavljena iz 5 tonov in oktavne ponovitve prvega. Pentatonična lestvica ne vsebuje poltonskih intervalov, zato nima vodilnega tona, osnovni ton pa ni jasno izražen. Glede na to, iz katere tonalne lestvice izhaja, imenujemo pentatonično lestvico durovo ali molovo. Pentatonična lestvica C-dur zгледа takole:

c d e g a c

Iz C-dur lestvice smo izpustili 4. in 7. stopnjo, ki sta osnova tonalne lestvice. Iz molove pentatonike izhajajo t.i. bluesovske lestvice, ki vsebujejo izrazit ton, imenovan ton blues. Pentatonična lestvica je pogosto uporabljena v zahodnjaški glasbi; molova pentatonika je gradnik prenekaterih rock balad.

2.5 Akordi

Akord je sozvočje treh ali več tonov. Najpogostejša oblika akorda je trozvok, ki ga sestavljajo 3 toni. Pogosti so tudi akordi z dodano 7. stopnjo. Pri našem delu se bomo omejili na trozvoke tipa dur in mol, treba pa se je zavedati, da je svet akordov mnogo širši.

Okvirni interval trozvoka je kvinta. Znotraj trozvoka je interval terca. Iz navedenih omejitev lahko sestavimo 4 tipe trozvokov:

- Velika terca + mala terca: DUR
- Mala terca + velika terca: MOL
- Mala terca + mala terca: zmanjšani trozvok (DIM)
- Velika terca + velika terca: zvečani trozvok (AUG)

V vsaki izbrani tonaliteti nastopa toliko akordov, kolikor je tonov lestvice. Vsak izmed tonov lestvice nastopa v vsaj 3 akordih te lestvice. V lestvici C-dur nastopajo naslednje stopnje akordov:

C D E F G A H

Vsak akord ima glede na razmerje terc znotraj akorda, ki so posledica intervalov izbrane lestvice, določeno tonaliteto. V durovi in naravni molovi lestvici nastopajo naslednje tonalitete (tudi kvalitete):

- 3 durovi akordi
- 3 molovi akordi
- 1 zmanjšan trozvok

Ker so kvalitete in razmerja za vse (transponirane) lestvice enake, jih pogosto označujemo z rimskimi številkami, kjer velike številke predstavljajo dur, male pa mol:

1. Tonika: I – dur, i – mol, i^o - dim
2. Subdominantna paralela: II – dur, ii – mol, ii^o - dim
3. Dominantna paralela: III – dur, iii – mol, iii^o - dim
4. Subdominanta: IV – dur, iv – mol, iv^o - dim
5. Dominanta: V – dur, v – mol, v^o - dim
6. Tonična paralela: VI – dur, vi – mol, vi^o - dim
7. Vodilni ton: VII – dur, vii – mol, vii^o - dim

Durovo lestvico lahko z rimskimi številkami opišemo na naslednji način:

I	ii	iii	IV	V	vi	vii^o
C	d	e	F	G	a	h^o

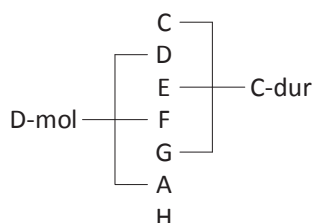
Na Sliki 2.1 je prikazan izvor prvih dveh akordov lestvice C-dur. Če akord izhaja iz neke lestvice, mora vsebovati tone, ki so del te lestvice. Če to združimo s pravili za gradnjo trozvokov, lahko iz vsake trojice tonov, z medsebojnimi intervali terce, sestavimo ustrezne akorde. Akord C-dur sestoji iz tonov C, E in G, pri čemer je razdalja med C in E velika terca (4 poltoni),

med E in G pa mala terca (3 poltoni). Razdalje ustrezajo pravilu za akord dur, akord, po osnovnem tonu, imenujemo C-dur. Akord D-mol sestoji iz tonov D, F in A, pri čemer je razdalja med D in F mala terca (3 poltoni), med F in A pa velika terca (4 poltoni). Razdalje ustrezajo pravilu za akord mol, akord, po osnovnem tonu, imenujemo D-mol.

Če zgornji postopek nadaljujemo, bomo analizirali vse akorde lestvice C-dur. Če postopek ponovimo še na vseh ostalih lestvicah, bomo navedli:

- 12 akordov dur
- 12 akordov mol
- 12 zmanjšanih trozvokov
- 12 zvečanih trozvokov

12 durovih in 12 molovih akordov je osnova za naše delo. Vsi ostali akordi so, zaradi poenostavitve modela, izpuščeni – klasificirani kot 25. razred (ni akord). Vsak trozvok predpostavlja tudi 2 obrata akorda. Obrat akorda je akord, ki vsebuje iste tone, kot osnovni akord, vendar najnižji ton ni osnovni ton, ampak terčni ali kvintni ton.



Slika 2.1: Toni, prisotni v akordih C-dur in D-mol.

2.5.1 Stopnje akordov

Znotraj vsake durove in (naravne) molove lestvice nastopa 7 akordov. Od teh so 3 tipa dur, 3 tipa mol in 1 zmanjšan trozvok. V durovi lestvici so durovi akordi tonika, subdominanta in dominantna. Vzporedno so v molovi

lestvici, molovi akordi tonika, subdominanta in dominanta. Te 3 stopnje lestvice imenujemo glavne stopnje. Akordi glavnih stopenj vsebujejo vse tone lestvice, zato pravimo, da lahko harmonizirajo celotno lestvico.

Z uporabo 7 stopenj sestavimo harmonsko zaporedje kompozicije, ki obstaja znotraj izbrane lestvice. Uporabo akordov, ki spadajo v drugo lestvico, imenujemo modulacija. Vsako zaporedje akordov ima končen cilj, ki vodi v, oziroma stran od tonike. Z uporabo modulacije ima skladatelj na voljo širok nabor akordov za uporabo v kompoziciji, vendar imajo določena zaporedja pomembnejšo vlogo od drugih.

Osnovo našega dela lahko povzamemo v 2 točkah:

1. Določeni akordi pogosto nastopajo skupaj.
2. Verjetnost pojavitve naslednjega akorda v skladbi je v veliki meri odvisna od njegovih predhodnikov; verjetnost nekaterih akordov je bistveno večja od drugih.

V popularni (zahodnjaški) glasbi, je vzorec ponovljenih zaporedij zelo opazen. Mnoge skladbe si delijo (uporabljajo) ista ali podobna zaporedja akordov. Ponovna uporaba zaporedja akordov seveda ni opredeljena kot kršenje avtorskih pravic, nam pa omogoča doseganje boljših rezultatov prepoznave. Če skladbe ne bi uporabljale takšnih vzorcev, bi bil naš model precej neuporaben.

Poglavje 3

Digitalizacija in zaznavanje zvoka

Način shranjevanja zvočnih posnetkov se od leta 1877 do osemdesetih let prejšnjega stoletja, ni dosti spremenil. Osnovna ideja je bila prenesti zvočno valovanje na nek medij in ga nato, z inverznim postopkom, ponovno predvajati. Medij se je skozi leta spreminjal. Sprva je bil to železni boben, kasneje gramofonska in vinilna plošča. Magnetni trak je svoj razcvet doživel v šestdesetih letih prejšnjega stoletja, predvsem na račun avtomobilske industrije in možnosti presnemavanja.

Leta 1982 sta Sony in Philips predstavila nov format shranjevanja glasbenih posnetkov. CD, oziroma Compact Disk je ponujal velike kapacitete, višjo kakovost zvoka ter daljšo življenjsko dobo v primerjavi z vinilnimi ploščami [4]. Zvok ni več zapisan kot zvezno valovanje, ampak je rezultat postopkov kvantizacije in diskretizacije. Digitalni zapis zvoka nam omogoča analizo značilnic, s pomočjo katerih iščemo zakonitosti v glasbi. Danes se uporablja mnogo zvočnih formatov; vsak izmed njih ima svoje posebnosti in namen.

LPCM (Linear Pulse Code Modulation)

Standardni zapis zvoka na CD in pri ostalih nekompresiranih zvočnih formatih (npr. WAVE). Kakovost zvoka je odvisna od naslednjih parametrov: frekvence vzorčenja, števila bitov na vzorec in števila zvočnih kanalov. Najpogostejša oblika je 44.1 kHz, 16 bit, stereo.

Stisnjeni zvočni formati (MP3, Vorbis, AAC, WMA)

Stisnjeni formati za svoje delovanje izkoriščajo 2 dejavnika: strukturo glasbenega posnetka in posebnosti človeškega sluha. Neizgubni formati kompresijo dosežejo z neenakomerno količino bitov na vzorec – tišina v posnetku npr. zavzame zelo malo bitov. Izgubni formati iz posnetka izločijo nepomembne frekvence (tiste, ki so izven slušnega spektra, so pretihe, ali pa so, zaradi prisotnosti sosednjih frekvenc, človeku neslišne).

MIDI (Musical Instrument Digital Interface)

Standard MIDI je bil razvit kot skupen jezik digitalnih glasbenih sintetizatorjev. Podpira 16 komunikacijskih kanalov, preko katerih se prenaša identifikacija zaigranega tona. V poli foničnem načinu omogoča tudi prenos harmonij. Na ta način lahko z eno MIDI napravo upravljamo več inštrumentov [5]. Datoteka MIDI definira standarden zapis glasbe, orodja za notacijo pa omogočajo enostavno izdelavo in urejanje datotek MIDI.

3.1 Človeški sluh in zaznavanje zvoka

V procesu glasbenega izražanja sodelujeta 2 entiteti: izvajalec in poslušalec. Vsak izmed njiju je nepogrešljiv, zato je pomembno, da si pogledamo njune lastnosti. Naloga izvajalca je jasna: igrati mora pravilno in všečno. Medtem, ko je pravilnost dobro definirana v glasbeni teoriji, katere smo se dotaknili v prejšnjem poglavju, je všečnost nekoliko bolj ohlapna. Vsekakor je odvisna

od poslušalca in njegovega slušnega sistema. Poglejmo si nekaj posebnosti človeškega sluha, ki bodo osnova za kasnejše delo analize zvoka.

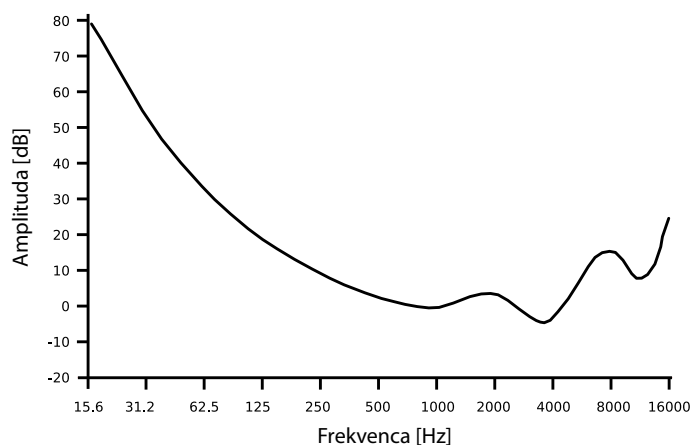
Človeško uho delimo v 3 dele: zunanje, srednje in notranje uho. Zunanje uho skrbi za prenos zvočnega valovanja do bobniča. Zaradi resonančne frekvence kanala, se frekvence od 2 kHz do 5 kHz lahko ojačajo do 10 krat [4]. Preko koščic se vibracije prenesejo do polža. Odprtina v polžu meri približno 4 mm^2 , medtem ko je površina bobniča okoli 60 mm^2 . Zaradi spremembe površine, je po Bernoullijevi enačbi faktor ojačitve približno 15 [4]. Polž vsebuje okoli 12.000 slušnih celic. Celice so na različne frekvence različno občutljive. Polž deluje kot nekakšen spektralni analizator. Pri visokih frekvencah se bolj odzivajo celice pri vhodu, notranje celice pa so bolj občutljive na nizke frekvence. Tekočina v polžu deluje kot kompresor. Večina energije se odbije – v polž potuje le majhen del, kar omogoča nemoteno zaznavo zelo glasnih in zelo tihih zvokov.

Pri razvoju sistemov za zvočno obdelavo je potrebno paziti, da ustrezno upoštevamo posebne lastnosti človeškega sluha. Frekvenčni razpon sluha je od 20 Hz do 20 kHz. Območje med 500 Hz in 5 kHz je dosti bolj občutljivo, predvsem zaradi resonančne frekvence slušnega kanala. V to območje spada človeški govor. V standardu ISO 226 je definirana krivulja za enakomerno zvočno zaznavanje (Slika 3.1). Iz slike so razvidna področja višje občutljivosti človeškega sluha. Več o praktični aplikaciji psihoakustičnega filtra sledi v poglavju 5.

3.2 Diskretna Fourierova transformacija

Fourierova transformacija je postopek, s katerim preslikamo signal iz časovne v frekvenčno domeno. Ker je naš signal diskreten (LPCM), govorimo o diskretni Fourierovi transformaciji. Postopek je leta 1807 predstavil francoski matematik Joseph Fourier. Osnovna ideja transformacije je naslednja:

“Vsak kompleksen signal lahko zapišemo kot vsoto sinusoid z različnimi frekvencami.”



Slika 3.1: Občutljivost človeškega ušesa na različne frekvence. Graf predstavlja amplitude, ki so potrebne za enakomerno zaznavo zvoka po spektru.

Akademija znanosti je novelo istega leta zavrnila, predvsem zaradi njene ohlapnosti. Danes se zavedamo, da je Fourierjeva transformacija temeljni kamen prenekaterega sistema. Aplikacij, kot so procesiranje signalov, obdelava slik, podatkovna kompresija, telekomunikacije, si brez Fourierjeve transformacije ne moremo predstavljati.

Matematična definicija transformacije je navedena v Enačbi (3.1). Postopek transformacije deluje v obe smeri – pretvorbo iz frekvenčne v časovno domeno predstavlja Enačba (3.2).

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \quad (3.1)$$

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{i2\pi kn/N} \quad (3.2)$$

Pri našem delu bomo eksplicitno uporabljali samo 1. del transformacije, saj signal zgolj analiziramo, potrebno pa se je zavedati, da na zelo podoben način izdelamo inverzni transform. Enačba (3.1) določa postopek za pridobitev posameznega elementa transformacije, to je X_k . To storimo tako, da

signal pomnožimo s sinusoido. Ker gre za diskretno transformacijo, seštejemo zmnožke vseh elementov signala in sinusoid. Sinusoida je v enačbi navedena kot kompleksni eksponent:

$$e^{-i2\pi kn/N}$$

Kompleksno sinusoido lahko zapišemo kot vsoto funkcij sinus in kosinus. Tako dobimo zapis, ki je naveden v enačbi (3.3).

$$X_k = \sum_{n=0}^{N-1} x_n \cos \frac{2\pi kn}{N} - i \sum_{n=0}^{N-1} x_n \sin \frac{2\pi kn}{N} \quad (3.3)$$

Postopek množenja signala s sinusoido imenujemo korelacija. Korelacijo si lahko predstavljamo kot mero prisotnosti sinusoid v podanem signalu. Rezultat X_k je torej mera prisotnosti sinusoid s frekvencami, ki so k-kratniki osnovne frekvence.

Delovanje algoritma si pogledjmo na primeru. Vhodni signal je realiziran kot vsota 2 sinusoid in je definiran z naslednjo enačbo:

$$V(n) = 4 \sin \frac{2\pi(1)n}{N} + 2 \cos \frac{2\pi(3)n}{N} \quad (3.4)$$

Na Sliki 3.2 je prikazano delovanje algoritma za $k = 0, 1, 2, 3, 4$. Vhodni signal se na vsakem koraku algoritma pomnoži z ustrezno bazno funkcijo (b1...b9). Bazne funkcije so trigonometrične funkcije sinus in kosinus. Bazna funkcija na posameznem koraku algoritma ima en cikel več, kot njena predhodnica. Na desni strani Slike 3.2 vidimo rezultat produkta vhodnega vektorja in posamezne bazne funkcije. Enačba (3.1) narekuje, da je potrebno elemente rezultata sešteti. Rezultati seštevanja so prikazani na skrajni desni strani Slike 3.2 pod kategorijo SUM.

Opazimo, da se elementi vektorja rezultatov pri večini primerov med seboj izničijo. Ker so bazne funkcije ciklične, se, ob odsotnosti le-teh v signalu, vektorji rezultatov seštevajo v 0. Pri baznih funkcijah, ki so prisotne v signalu, normaliziran seštevek vektorja rezultatov pomeni prispevek posamezne bazne

funkcije k signalu. Normaliziran rezultat naše Fouriereve transformacije, za $k = 0, 1, 2, 3, 4$ je naslednji:

$$X = [0, 4, 0, 2, 0]$$

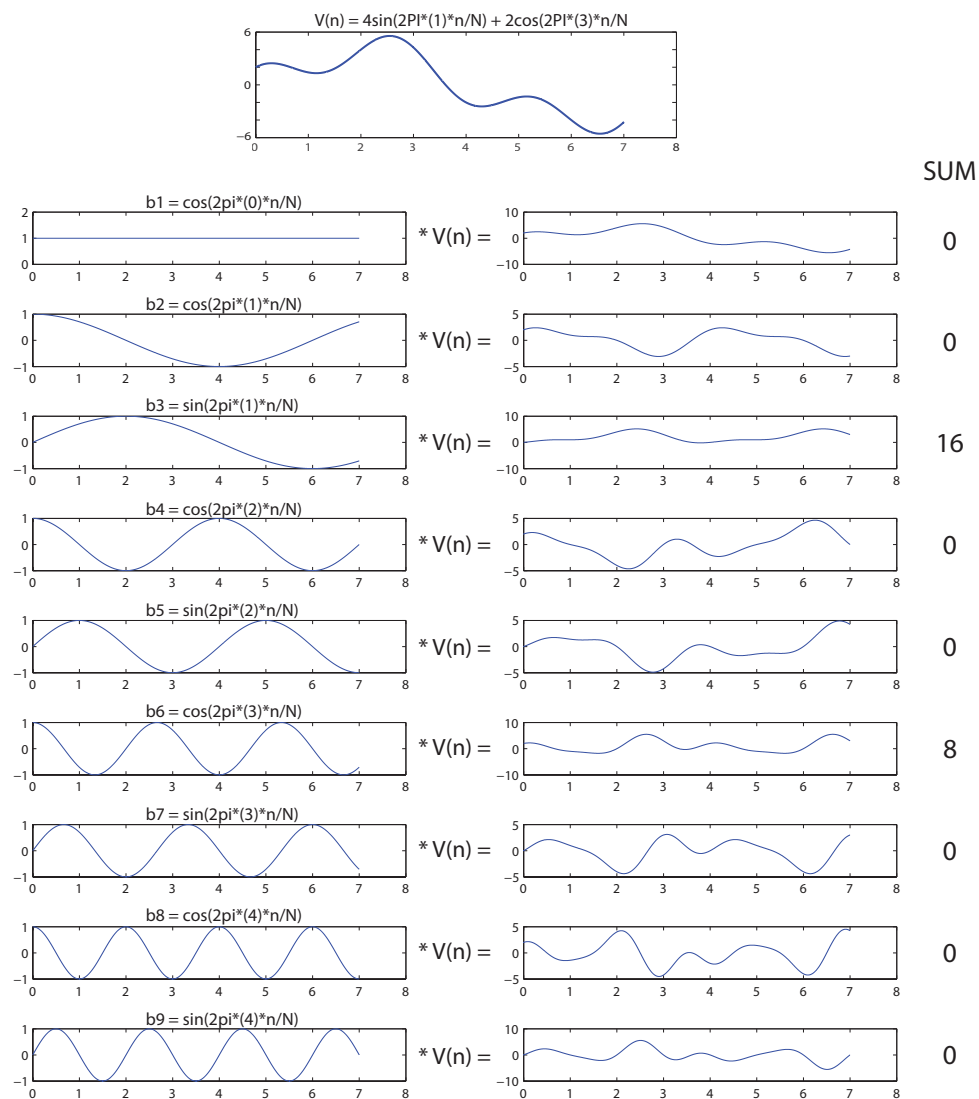
Če primerjamo rezultat z Enačbo vhoda (3.4), vidimo, da algoritem deluje pravilno. Vektor X navaja, da signal vsebuje sinusoido z 1 ciklom in amplitudo 4 ter sinusoido s 3 cikli in amplitudo 2.

V primeru na Sliki 3.2 smo analizirali enostaven signal brez faznega zamika. Prav tako je bilo število ciklov vhodnega signala vedno celo število. V realnosti seveda ni tako. Fazni zamik lahko obravnavamo kot vsoto ustrezno obteženih sinusoid, število ciklov signala pa predstavlja večji problem. Če z zgornjimi baznimi funkcijami analiziramo signal, ki nima celega števila ciklov, bodo seštevki vektorjev rezultatov neničelni tudi pri baznih funkcijah, ki niso vsebovane v signalu. Ta pojav imenujemo spektralno puščanje. Rešitev problema spektralnega puščanja ponujajo okenske funkcije, ki jih bomo omenili v 5. poglavju.

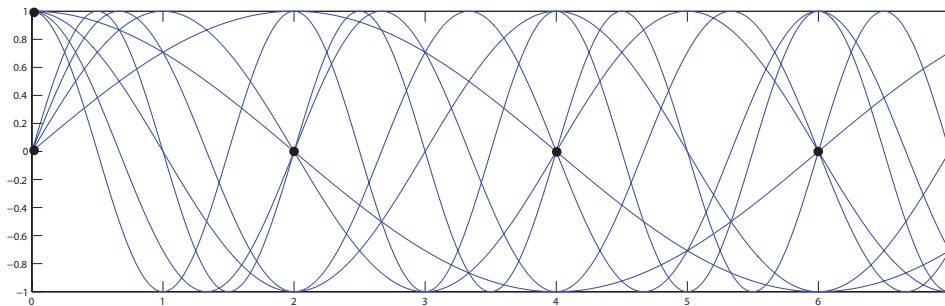
3.2.1 Hitra Fouriereva transformacija – FFT

Algoritem, ki smo ga prikazali v prejšnjem poglavju izvira iz definicije. Težava algoritma DFT je veliko število računskih operacij, ki so potrebne za izračun transformacije. Kompleksnost znaša $O(N^2)$, kjer je N število elementov transformacije.

Algoritem FFT – hitra Fouriereva transformacija deluje po principu “deli in vladaj”. Če bi vse bazne funkcije iz zgornjega primera izrisali na skupen graf, bi opazili, da grafi vsebujejo veliko skupnih točk (Slika 3.3). FFT izkorišča to dejstvo z uporabo rekurzije. Za točke, kjer se stika več baznih funkcij, se izračun opravi samo enkrat, rezultat pa se rekurzivno ekstrapolira na vsa ustrezna mesta v transformu. Kompleksnost algoritma FFT je $O(N \log N)$, kjer je N število elementov transformacije. Bistveno manjša kompleksnost je omogočila uporabo algoritma FFT v aplikacijah, ki zahtevajo realno časovni odziv.



Slika 3.2: Prikaz delovanja diskretne Fouriereve transformacije. Zgoraj je prikazan vhodni signal, na levi strani so navedene bazne funkcije, na desni strani je rezultat produkta posamezne bazne funkcije z vhodnim signalom. Stolpec SUM predstavlja seštevek elementov posameznega vektorja rezultatov.



Slika 3.3: Skupni izris baznih funkcij pokaže skupne točke grafov funkcij (npr. pri 0, 2, 4 in 6).

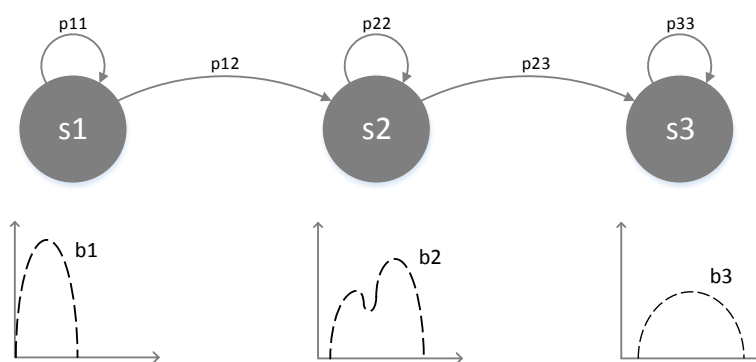
Poglavje 4

HMM - Skriti markovski model

4.1 Definicija

Skriti markovski model je statistični markovski proces prvega reda s skritimi – neopazovanimi stanji. Markovski proces prvega reda pomeni, da je trenutno stanje odvisno le od neposrednega predhodnika:

$$p(x_i|x_{i-1}, x_{i-2}, \dots x_1) = p(x_i|x_{i-1}) \quad (4.1)$$



Slika 4.1: Shema delovanja skritega markovskega modela. s1, s2, s3 - skrita stanja, b1, b2, b3 - opazovana stanja.

Pri markovskem procesu opazujemo stanja in prehode med njimi. Skriti markovski proces vsebuje 2 verigi: opazovano in skrito. Opazovana stanja so

tista, ki jih lahko na nek način izmerimo. Skrita stanja poskušamo ugotoviti na podlagi opazovanih stanj (Slika 4.1).

$$\begin{aligned} s(t) &\in \{s_1, s_2, s_3, \dots, s_n\}; && \text{skrita stanja} \\ b(t) &\in \{b_1, b_2, b_3, \dots, b_n\}; && \text{opazovana stanja} \end{aligned}$$

Skriti markovski model razložimo na naslednjem primeru: Zaprti smo v sobo brez oken in vrat. Vreme zunaj se spreminja; predpostavimo, da je jutrišnje vreme odvisno le od tega, kakšno vreme je danes – markovski proces prvega reda. Matriko prehodov med stanji predstavlja Tabela 4.1

	Sončno	Deževno	Megleno
Sončno	0.8	0.1	0.1
Deževno	0.2	0.6	0.2
Megleno	0.2	0.3	0.5

Tabela 4.1: Verjetnosti prehajanja stanj modela vremena. Vrstice - trenutno stanje, stolpci - možni prehodi

Ugotoviti želimo, kakšno je vreme zunaj. Enkrat na dan nam skrbnik prinese živež. Vsa informacija, ki jo dobimo o zunanjem svetu je ta, ali skrbnik prinese s seboj dežnik, ali ne. Verjetnosti, da skrbnik nosi dežnik glede na vreme, so podane v Tabeli 4.2

	Sončno	Deževno	Megleno
Prisotnost dežnika	0.1	0.9	0.3

Tabela 4.2: Verjetnosti za prisotnost dežnika v odvisnosti od vremena.

Predpostavimo, da je bilo na dan, ko so nas zaprli, zunaj sončno. S postopkom skritega markovskega modela lahko odgovorimo na naslednja vprašanja:

1. Kolikšna je verjetnost, da je drugi dan deževalo, če je skrbnik s seboj imel dežnik?

2. Kolikšna je verjetnost, da je tretji dan megla, če je skrbnik s seboj imel dežnik drugi dan, tretji dan pa ne?
3. Kakšna je najverjetnejše zaporedje stanj vremena za 1 teden, če smo si vsak dan beležili prisotnost / odsotnost dežnika?

4.2 Gradniki modela

V primeru smo navedli elemente, ki jih potrebujemo za gradnjo skritega markovskega modela. Navedimo jih v splošni obliki:

1. Množica skritih stanj S
2. Množica opazovanih stanj O
3. Matrika verjetnosti prehodov med stanji T
4. Matrika emisij E
5. Začetna razporeditev verjetnosti stanj π

4.2.1 Množica skritih stanj S

Množica skritih stanj vsebuje elemente, ki jih želimo s pomočjo modela oceniti. Skrita stanja so lahko diskretne vrednosti, oznake, skratka kakršna koli veličina, ki jo lahko segmentiramo. Množica skritih stanj vsebuje n elementov. Spodnja notacija predstavlja skrito stanje sistema v času t :

$$s(t) \in \{s_1, s_2, s_3, \dots, s_n\}$$

Elementi množice skritih stanj v zgornjem primeru so: “Sončno”, “Deževno” in “Megleno”. Matrika verjetnosti prehodov med stanji T glasi na skrita stanja, torej elemente množice S .

4.2.2 Množica opazovanih stanj O

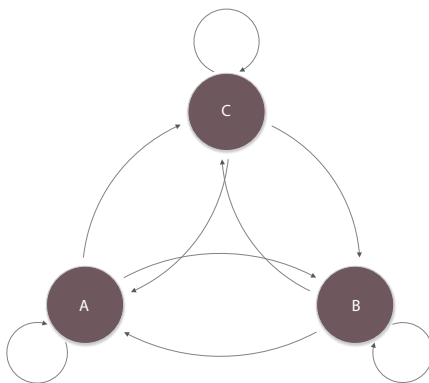
Množica opazovanih stanj O vsebuje meritve oziroma rezultate opazovanj. Za vsako skrito stanje sistema opravimo meritev, zato je velikost množice O enaka velikosti množice S .

$$o(t) \in \{o_1, o_2, o_3, \dots, o_n\}$$

Meritve lahko zavzamejo številsko ali opisno vrednost. Pri opisnih vrednostih je pomembno, da določimo nabor možnih vrednosti. Številske vrednosti lahko obstajajo tudi kot večdimenzionalni vektorji, ki bodo osnova našega nadaljnjega dela.

4.2.3 Matrika verjetnosti prehodov med stanji **T**

Z matriko verjetnosti prehodov med stanji določimo markovski proces prve stopnje (Slika 4.2). Matrika je dimenzij $n \times n$, pri čemer je n število stanj sistema. Matrika določa, kolikšne so verjetnosti prehodov iz podanega stanja v vsa ostala stanja (vključno s preходом sam vase).



Slika 4.2: Markovski proces 1. stopnje.

Matrika verjetnosti prehodov vsebuje naslednje elemente:

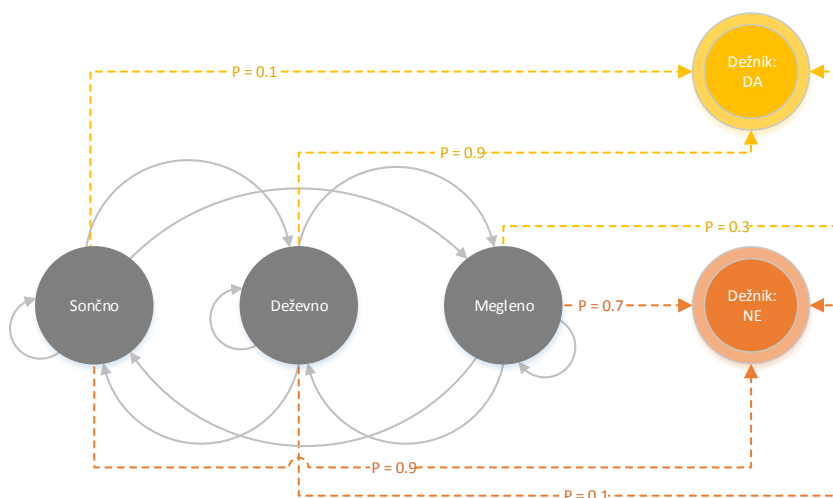
$$T_{i,j} = P(s_j(t)|s_i(t-1)); 1 \leq i, j \leq N$$

4.2.4 Matrika emisij **E**

Matrika emisij se nanaša na skriti del markovskega modela. Ker nimamo dostopa do skritih stanj modela (če bi jih imeli, bi bil celoten postopek trivialen), se moramo zadovoljiti z opazovanimi stanji. Opazovana stanja so naše

meritve opazovanega procesa. Pomembno je, da izberemo takšne metrike, ki so čim bolj konsistentne s skritimi stanji. V našem primeru je prisotnost dežnika dokaj ustrezna metrika za ocenjevanje vremena. Ne bi pa nam dosti pomagala npr. pri ocenjevanju gostote prometa na bližnji cesti.

$$E_i(s) = P(S_k = s | O_k = i)$$



Slika 4.3: Shematični prikaz modela za primer z vremenom. Naša meritev (prisotnost dežnika) je predstavljena v rumeni / oranžni barvi. Verjetnosti na črtah predstavljajo matriko emisij.

Matrika emisij določa mero statistične povezanosti metrike s skritim stanjem. Na Sliki 4.3 je shematična predstavitev matrike emisij in markovskega modela 1. stopnje. Izbira metrike in posledično gradnja matrike emisij je osnovna naloga pri implementaciji modela HMM. Potrebno je dobro poznavanje problema, saj z napačno oz. slabo izbiro metrike ne bomo dosegli željenih rezultatov. Ker se morajo metrike čim bolj prilagajati skritim stanjem sistema, je lahko implementacija dobre metrike precej težavna. V enostavnih modelih je metrika lahko kar numerična – verjetnost. Poseben problem so modeli s področja multimedije, kjer so ustrezne metrike pogosto kompozitumi večdimenzionalnih vektorjev. Ker naš model spada v področje multimedije,

bomo izbiro in implementacijo metrike podrobno razložili v naslednjem poglavju. Matriko emisij za problem z vremenom predstavlja Tabela 4.2.

4.2.5 Začetna porazdelitev verjetnosti stanj π

Sistem za delovanje potrebuje izhodišče – začetno porazdelitev. Matrika je dimenzij $1 \times n$, kjer je n število vseh skritih stanj. Za vsako možno stanje moramo določiti, kakšna je verjetnost, da se sistem v začetku nahaja v njem. V našem primeru z vremenom začetni porazdelitvi ustreza informacija, da je bilo na dan, ko so nas zaprli, zunaj sončno. Izbira začetnih porazdelitev je, tako kot matrika emisij, odvisna od problema. Pri vremenu smo uporabili znanje, ki smo ga pridobili na podlagi 1 vzorca. Lahko bi uporabili tudi enakomerno porazdelitev z verjetnostjo $1/n$. Slednja vsa stanja obravnava enakovredno – uporabili jo bomo tudi pri implementaciji modela prepoznave akordov.

4.3 Enačba modela HMM

Z uporabo gradnikov iz prejšnjega poglavja lahko model HMM zapišemo z enačbo:

$$p(x_1, x_2, \dots, x_n) = p(\pi_1)p(o_1|s_1) \prod_{k=2}^n p(s_k|s_{k-1})p(o_k|s_k) \quad (4.2)$$

$p(\pi_1)$...začetna porazdelitev

$p(o_k|s_k)$...matrika emisij

$p(s_k|s_{k-1})$...matrika prehodnih verjetnosti

4.4 Algoritem Viterbi

Algoritem Viterbi je algoritem dinamičnega programiranja, ki ga je prvi opisal Andrew Viterbi leta 1967. Uporablja se na mnogih področjih, kot so dekodiranje GSM signalov, sinteza govora, razpoznavanje govora, bioinformatika. Z uporabo algoritma Viterbi določimo najverjetnejšo sekvenco skritih

stanj S , ki je povzročila sekvenco opazovanih stanj O . Osnovni razmislek algoritma je naslednji:

“Če gre najverjetnejša pot skozi sekvenco stanj, ki se zaključi v s_n , skozi s_{n-1} , potem ta pot sovпада z najverjetnejšo potjo, ki se zaključi v s_{n-1} .”

Zgornji razmislek je temelj za implementacijo rekurzije v algoritmu Viterbi. Za delovanje algoritma potrebujemo vse elemente modela HMM. Rekurzijo navedimo takole:

$$w(s_1) = p(s_1)p(o_1|s_1) \quad (4.3)$$

$$w(s_n) = p(o_n|s_n) \max_{s_{n-1}} \{w(s_{n-1})p(s_n|s_{n-1})\} \quad (4.4)$$

$w(s_1)$...baza rekurzije

$w(s_n)$...splošna enačba rekurzije

$p(s_1)$...začetna porazdelitev

$p(o_n|s_n)$...matrika emisij

$p(s_n|s_{n-1})$...matrika preh.verjetosti

Z uporabo zgornje rekurzije dobimo verjetnost najbolj verjetne sekvence stanj. Ker je naš cilj najti sekvenco in ne njene verjetnosti, je potrebno na vsakem koraku rekurzije zabeležiti najbolj verjetno pot.

Poglavje 5

Implementacija

V prejšnjih poglavjih smo si pogledali teorijo, potrebno za razumevanje in implementacijo praktičnega modela prepoznave akordov. Kot bomo videli v nadaljevanju, je lahko gradnja modela poljubno komplicirana. V začetku je potrebno opredeliti predpostavke in omejitve, na podlagi katerih se določi nivo podrobnosti modela. Nekatero omejitev spadajo v glasbeno teorijo in smo jih že navedli v pripadajočem poglavju, druge pa so praktičnega značaja in jih bomo omenili pri posameznih komponentah. Razvojno okolje našega modela je Mathworks Matlab. Zaradi matrične usmerjenosti in mnogih vgrajenih funkcij za procesiranje signalov, je razvoj v okolju Matlab hiter in učinkovit.

5.1 Predpostavke in omejitve

Model razpoznave akordov predpostavlja 2 množici podatkov: učno in testno. Za vse podatke imamo na voljo temeljno resnico – dejanski akord s podanim časovnim zapisom. Iz učne množice podatkov se model nauči parametre, potrebne za nadaljnjo razpoznavo. Prepoznane sekvence testnih podatkov nato primerjamo s temeljno resnico testnih podatkov. Odstotek ujemanja nam pove natančnost modela.

5.1.1 Omejitev abecede akordov

Kot smo že spoznali v poglavju o glasbeni teoriji, je število različnih akordov zelo veliko. Če želimo, da bo naš model rezultate podal v doglednem času, moramo množico nekako omejiti. Prav tako se pri veliki količini akordov pojavlja problem ločljivosti, saj se število različnih tonov, ter razdalje med toni v akordu, manjšajo. Že v začetku smo se omejili na trozvoke in tonalne lestvice. V okviru tonalnih lestvic nastopa 12 durovih akordov, 12 molovih akordov, 12 zvečanih trozvokov in 12 zmanjšanih trozvokov. Ker se zvečani in zmanjšani trozvoki v primerjavi z durovimi in molovimi trozvoki pojavljajo precej manj pogosto, jih pri klasifikaciji izpustimo. Naša abeceda sestoji iz akordov, ki jih navaja Tabela 5.1

C	C#	D	D#	E	F	F#	G	G#	A	A#	B
Cm	C#m	Dm	D#m	Em	Fm	F#m	Gm	G#m	Am	A#m	Bm

Tabela 5.1: Abeceda akordov, uporabljenih v našem modelu.

Akord (stanje) NC je univerzalni razred, v katerega klasificiramo vse neprepoznane akorde, prav tako tudi zvečane in zmanjšane trozvoke. V razred NC (no – chord) spadajo tudi tišina in tonsko neopredeljivi elementi glasbe. Nekateri sistemi za prepoznavo akordov [10], ločijo med obrati akordov, prav tako pa tudi ločeno analizirajo basovske vzorce segmentov. Zaradi kompleksnosti implementacije, naš sistem takšnih podrobnosti ne vsebuje.

5.1.2 HMM Toolbox

Implementacija skritega markovskega modela je, kljub na videz enostavnemu algoritmu, precej obsežna. Ker za okolje Matlab obstaja mnogo implementacij, t.i. toolbox-ov, smo se pri naši implementaciji poslužili obstoječe implementacije za skriti markovski model [11]. Toolbox avtorja Kevina Murphy-ja vsebuje vse potrebne funkcije za uporabo skritega markovskega modela. Več o posameznih funkcijah bomo navedli pri elementih modela, v splošnem pa

toolbox poskrbi za naslednja opravila:

1. Učenje: Iz podanih podatkov generira matriko prehodov, matriko emisij in vektor začetnih stanj.
2. Evalvacija: Iz zaporedja opazovanih stanj zgradi vektor verjetnosti skritih stanj, katerih posledica so opazovana stanja; vektor se uporabi v algoritmu Viterbi, ki izdelava najverjetnejšo sekvenco stanj.

Za metrike in vse vhodne podatke algoritmov moramo poskrbeti sami, saj so specifični za vsak problem. Iz naslova toolbox-a izhaja tudi omejitev mešanja Gaussovih krivulj, ki jo bomo omenili v poglavju o možnih izboljšavah modela.

5.2 Vhodni podatki modela

Vhodni podatki modela so 2 vrst:

- Zvočni posnetki skladb v obliki wave.
- Notacije s časovno oznako in imenom akorda.

Transkripcije skladb so prenesene iz spletne strani Isophonics [12]. V zbirki so vključeni albumi skupin Queen, The Beatles in Carole King. Ker so posnetki skladb na različnih albumih različni, je pomembno, da za posamezno transkripcijo pridobimo ustrezen posnetek. Odstopanja se lahko pojavijo tako v dolžini, kot tudi v samem tempu posnetka.

Podatki v zbirki Isophonics so v obliki tekstovnih datotek. Vsaka datoteka vsebuje časovne žige in oznake akordov:

52.762 56.448 A

Notacija izraža, da je v času od 52.762s do 56.448s v posnetku prisoten akord A. Notacije v zbirki na žalost niso popolnoma konsistentne, kar predstavlja problem pri obravnavi neznanih akordov. Potrebno je poskrbeti, da se vse različne notacije istega akorda obravnavajo enako. V tabeli 5.2 so navedene nekatere različne oznake, ki se pojavijo pri akordih C-dur in C-mol.

5.3 Kromagram

V poglavju o skritem markovskem modelu smo omenili, da je izbira ustrezne metrike fundamentalnega pomena za delovanje modela. Izbrati moramo takšno metriko, ki se čim bolj prilega skritim stanjem sistema. To pomeni, da mora biti ob pojavitvi istega skritega stanja opazovano stanje čim bolj konsistentno – opazovano stanje, ki je posledica skritega stanja moramo znati čim bolje klasificirati. V našem modelu smo za metriko uporabili kromagram. V poglavju o glasbeni teoriji smo spoznali tonalne lestvice, ki so osnova našega dela. Tonalne lestvice so sestavljene iz 12 različnih tonov in oktavnih ponovitev le-teh. Vsi akordi našega modela izhajajo iz tonov tonalnih lestvic. Okravne ponovitve tonov lahko združimo tako, da njihove amplitude seštejemo. V splošnem se amplitude, ki so v frekvenčnem spektru dovolj narazen, seštevajo [13].

V Tabeli 5.3 so navedeni vsi toni s pripadajočimi frekvencami, ki jih uporabljamo v našem modelu. Če seštejemo vse amplitude po navedenih frekvencah in vrednosti normaliziramo, dobimo kromagram. Kromagram izraža odstotek prisotnosti posameznih tonov znotraj izbranega frekvenčnega območja. Kromagram je primerna metrika zato, ker ustreza pravilom iz glasbene teorije. Na Sliki 5.1 zgoraj je akord C-dur. V poglavju o glasbeni teoriji smo navedli, da je akord C-dur sestavljen iz tonov C, E in G. Prav tako smo navedli, da je interval med C in E velika terca, med E in G pa mala terca. Iz slike je razvidno, da so odstotki tonov C, E in G bistveno večji od ostalih tonov. Izrazite so tudi razdalje velike (4) in male (3) terce. Na Sliki 5.1 spodaj je prikazan akord A-mol. Tudi ta kromagram je v skladu z

Akord	Različne oznake				
C-dur	C	Cmaj	C:maj	C:	
C-mol	c	Cmin	C:min	Cm	C:m

Tabela 5.2: Različne notacije istega akorda v zbirki Isophonics.

Ton	Frekvence po oktavah						
C	32.70	65.41	130.8	261.6	523.3	1047	2093
C#	34.65	69.30	138.6	277.2	554.4	1109	2217
D	36.71	73.42	146.8	293.7	587.3	1175	2349
D#	38.89	77.78	155.6	311.1	622.3	1245	2489
E	41.20	82.41	164.8	329.6	659.3	1319	2637
F	43.65	87.31	174.6	349.2	698.5	1397	2794
F#	46.25	92.50	185.0	370.0	740.0	1480	2960
G	49.00	98.00	196.0	392.0	784.0	1568	3136
G#	51.91	103.8	207.7	415.3	830.6	1661	3322
A	55.00	110.0	220.0	440.0	880.0	1760	3520
A#	58.27	116.5	233.1	466.2	932.3	1865	3729
B	61.74	123.5	246.9	493.9	987.8	1976	3951

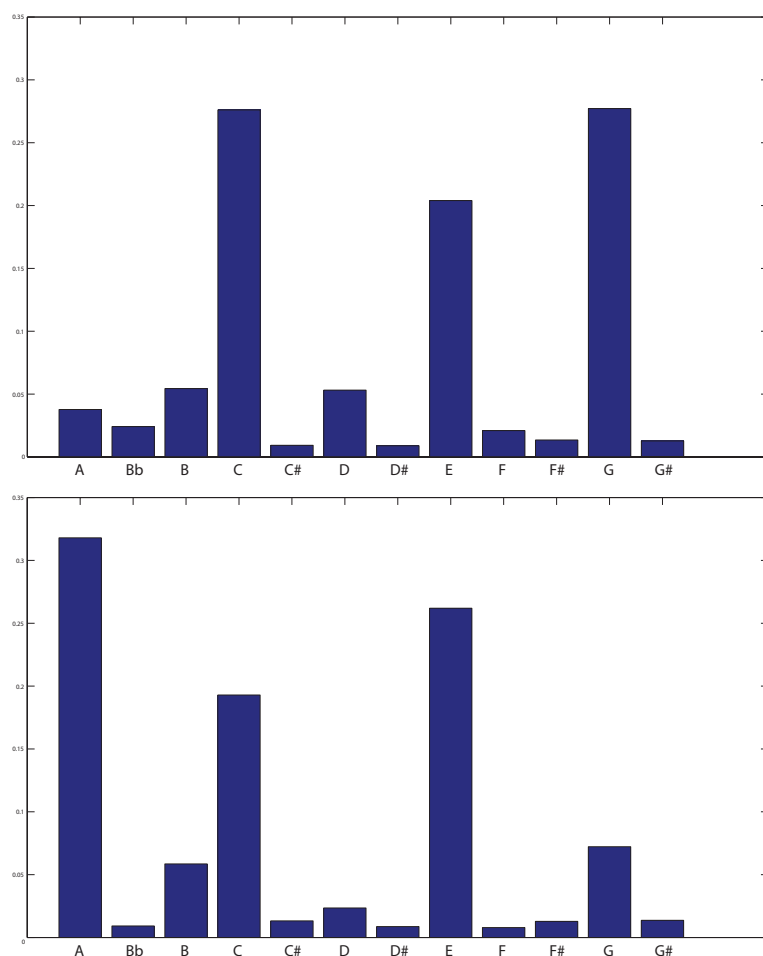
Tabela 5.3: Frekvence tonov in oktavnih ponovitev v modelu.

glasbeno teorijo. Kromagrama na Sliki 5.1 sta del matrike emisij, ki jo bomo omenili v nadaljevanju. Matrika emisij vsebuje 25 kromagramov – za vsako stanje svojega.

5.3.1 Uglasitev

Poseben problem izdelave kromagrama je uglasitev skladbe. Uglasitev zgodnjih skladb skupine The Beatles ni sledila komornemu tonu 440 Hz ampak 435 Hz. Pri novejših skladbah so prevzeli komorno uglasitev. Poleg različnih izvedenk istih skladb se pojavlja tudi problem raztegovanja in krčenja magnetnih trakov, kar lahko vpliva na višino celotnega posnetka.

Pri generiranju kromagrama je potrebno upoštevati različne uglasitve. Pri našem modelu smo problem rešili z uporabo intervala okoli frekvenc iz Tabele 5.3. Na intervalu poiščemo najvišjo amplitudo – lokalni maksimum, ki ga uporabimo pri gradnji kromagrama.



Slika 5.1: Kromagrama akordov C-dur (zgoraj) in A-mol (spodaj).

5.3.2 Psihoakustično filtriranje

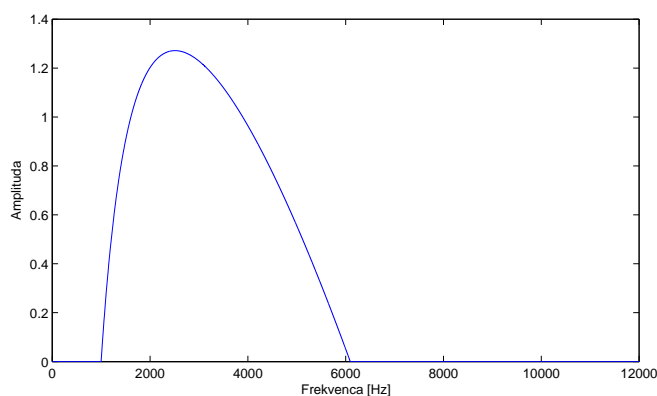
Človeško uho je zaradi evolucije bolj občutljivo na določene frekvence. Pri gradnji kromagrama je potrebno to dejstvo upoštevati. Ker se glasba, kot harmonija tonov, ustvarja z ozirom na človeški sluh in dožemanje zvoka, je smiselno, da pri gradnji modela upoštevamo te specifikke.

Človeško uho je najbolj dovzetno za frekvence med 2kHz in 5kHz. To pomeni, da tone v tem območju, ki so sicer tišji, slišimo enako glasno oz. glasneje kot ostale tone. Fourierova transformacija nam takšnih informacij seveda ne more priskrbeti. Amplitude tonov iz tega frekvenčnega območja

bodo zato po transformaciji premajhne, kljub temu, da so natančna preslikava v frekvenčni prostor. Enačbi (5.1) in (5.2) prikazujeta postopek normalizacije vhodne frekvence. Frekvenca F_2 predstavlja normalizirano frekvenco. Na Sliki 5.2 je prikazan psihoakustični filter, ki ustrezno upošteva lastnosti človeškega sluha [10]. Z uporabo filtra normaliziramo amplitude, kar posledično pomeni bolj natančne kromagrame.

$$R_a(f) = \frac{12200^2 f^4}{(f^2 + 20.6^2) \sqrt{(f^2 + 107.7^2)(f^2 + 737.9^2)(f^2 + 12200^2)}} \quad (5.1)$$

$$F_2(f) = 2.0 + 20 \log_{10}(R_a(f)) \quad (5.2)$$



Slika 5.2: Frekvenčni odziv psihoakustičnega filtra iz Enačbe (5.1).

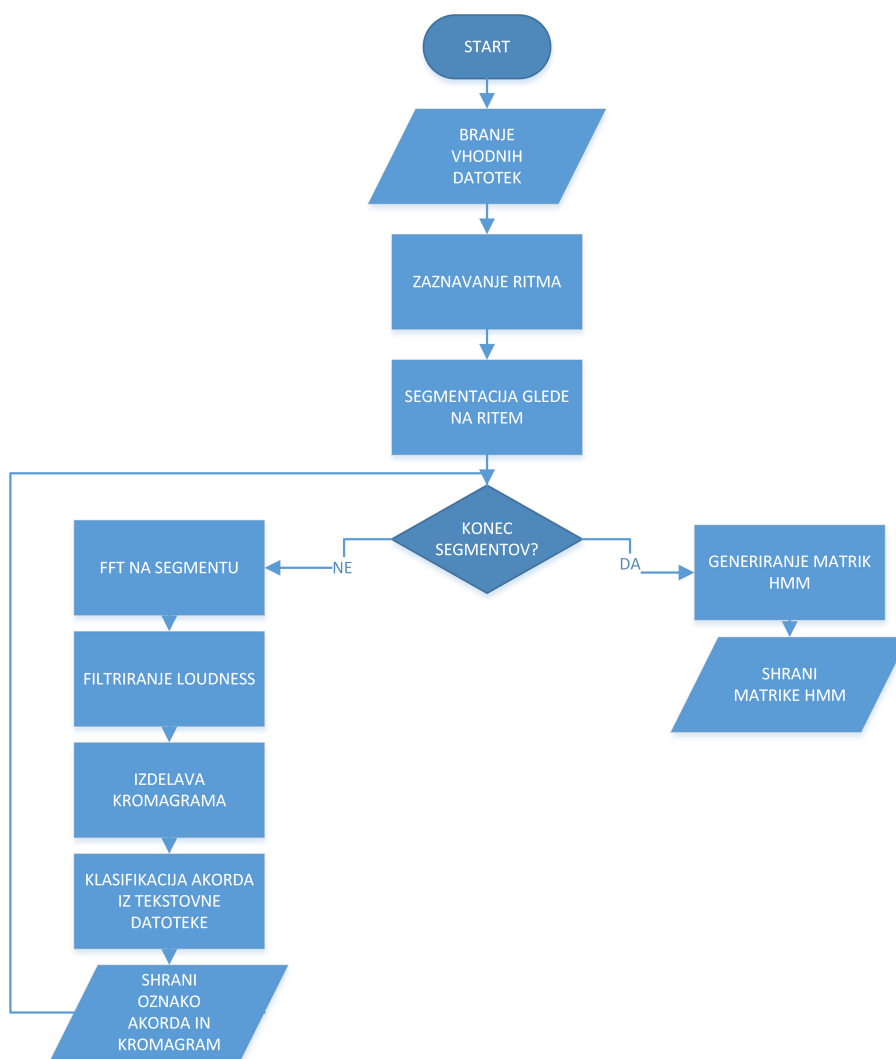
5.4 Učenje

Postopek učenja je proces, pri katerem iz vhodnih podatkov izluščimo znanje, s pomočjo katerega naš model prepoznava akorde. Za postopek učenja potrebujemo tako podatke za gradnjo kromagramov – zvočne datoteke, kot tudi podatke o notaciji – tekstovne datoteke. Na Sliki 5.3 je prikazan diagram delovanja učnega procesa.

5.4.1 Zaznavanje tempa in segmentacija

Vhodni učni podatki – notacije posamezne skladbe so navedeni v vektorju s pripadajočimi časovnimi oznakami. Če želimo podatke koristno uporabiti, moramo tudi zvočne posnetke na nek način segmentirati. V splošnem imamo za segmentacijo na voljo 3 možnosti:

1. Za segmentacijo uporabimo časovne oznake notacij: glasbeni posne-



Slika 5.3: Diagram delovanja učnega procesa modela.

tek razdelimo na segmente, ki so istih dolžin, kot trajanje posameznih akordov. Dolžine so variabilne.

2. Posnetek segmentiramo na podlagi tempa, ki ga predhodno izračunamo. Segmenti so fiksnih dolžin, njihove dolžine so nekajkrat krajše od časovnih oznak notacij akordov.
3. Posnetek segmentiramo na odseke fiksnih dolžin brez zaznavanja tempa. Potrebno je paziti, da so dolžine dovolj majhne, da odseki ne segajo v območje sosednjih akordov. Število odsekov je lahko zelo veliko.

V našem modelu smo uporabili metodo z zaznavanjem tempa (2. točka). Ritem zaznavamo s pomočjo funkcije *beat2*, ki je del projekta LabROSA [14]. Funkcija generira vektor časovnih oznak, na podlagi katerih izdelamo segmente posnetka. Vsak posamezen segment se analizira neodvisno od ostalih.

5.4.2 Fourierova transformacija

Nad vsakim segmentom posnetka se opravi Fourierova transformacija. V okolju Matlab je za naše potrebe najprimernejša metoda Spectrogram. Metoda izvede STFT (kratkotrajno Fourierovo transformacijo). Vzorčna frekvenca naših segmentov je 44.1kHz. Nyquist-ova frekvenca je naslednja:

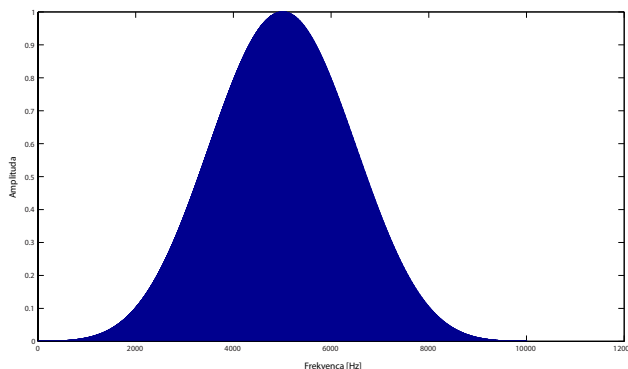
$$F_{nq} = \frac{f_s}{2} = 22050Hz, pri f_s = 44100Hz \quad (5.3)$$

Najvišja frekvenca v Tabeli 5.3 je približno 4kHz, zato funkciji Spectrogram podamo najvišjo frekvenco 5kHz. Funkcija Spectrogram predvideva tudi vnos okenske funkcije.

Okenske funkcije

Diskretna Fourierova transformacija predvideva sinusni signal, ki je vzorčen tako, da je v vzorcu zajeta celotna perioda. V realnosti bo naš signal prisekan

– nezvezen. Rezultat Fouriereve transformacije nad takšnim signalom bo vseboval veliko šuma – frekvenc, ki sicer niso prisotne v signalu. Drug problem predstavlja samo delovanje DFT. Med frekvenčnimi razdelki je precej prekrivanja, zaradi česar rezultat transformacije nima ostro definiranih vrhov. Okenske funkcije ponujajo rešitev za zgoraj opisana problema. Poznamo jih več vrst, njihova uporaba pa je odvisna od posameznega problema. V našem modelu smo uporabili Blackman-Harrisovo okno, ki je prikazano na Sliki 5.4. Z uporabo okenske funkcije dosežemo bistveno boljše rezultate transforma-

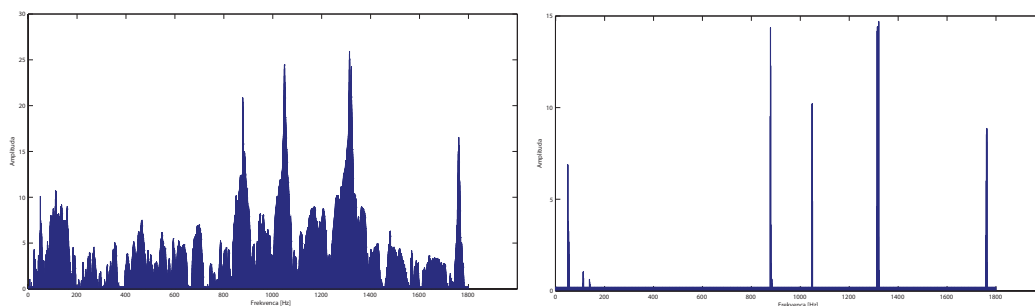


Slika 5.4: Blackman-Harris-ovo okno.

cije, kot brez nje. Na Sliki 5.5 je prikazana primerjava med transformacijo z in brez okenske funkcije. Transformirali smo studijski posnetek akorda Amol, zaigranega na klavir. Prikazan je izsek na območju do 1800Hz. Okenska funkcija nekoliko zniža amplitudo signala, zato je pomembno, da imamo dovolj močan signal.

5.4.3 Branje in urejanje akordov

Model vzporedno uporablja oba vektorja vhodnih podatkov. Ko se izvrši transformacija zvočnega segmenta, je potrebno prebrati notacijo akorda, ki ustreza trenutnemu segmentu. Ker akordi navadno trajajo več dob, je zvočnih segmentov več kot notiranih akordov. To pomeni, da več zvočnih segmentov predstavlja isti akord. S tem ni nič narobe, saj tako dobimo več



Slika 5.5: Primerjava transformacije z in brez okenske funkcije: rezultat funkcije FFT brez okna (levo) in funkcije SPECTROGRAM (z Blackman-Harris-ovim oknom (desno).

vzorcev za opazovana stanja, kar nam lahko prinese boljše srednje vrednosti v matriki emisij. Omenili smo, da je potrebno poskrbeti za transparentno branje različnih notacij, ki predstavljajo isti akord. Pri branju notacij pa je potrebno poskrbeti še na problem četverozvokov posebnih akordov. Četverozvoki so trozvoki z dodatnim tonom. Pogosto je ta ton 6. ali 7. stopnje. Akord A-mol z dodano 7. stopnjo npr. označimo z Am7. Tudi pri četverozvokih prihaja pri notacijah do nedoslednosti. Kljub temu, da smo se pri našem delu omejili na trozvoke, moramo pri učenju ustrezno upoštevati vse prebrane akorde. Poglejmo si naslednje zaporedje akordov:

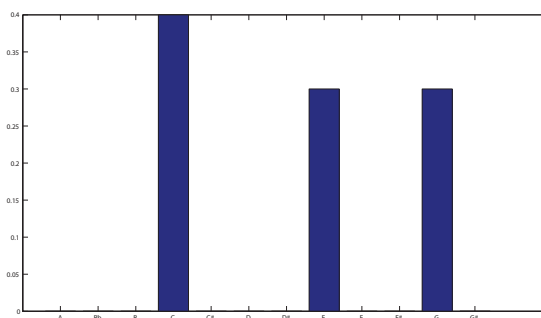
$$C \ Dm \ G7 \ C$$

Na 3. mestu je četverozvok G7, ki je za naš sistem posebnost. Potrebno je določiti, kako naš sistem reagira na tak akord in sicer v dveh ozirih:

1. Kako izdelati ustrezen kromagram?
2. Kaj storiti z zaporedjem akordov (glej poglavje o matriki prehodov)?

Odgovor na prvo vprašanje je lahko enostaven: kromagrama za takšen akord ne izdelamo. Vendar lahko tako nastane problem zaradi majhne učne množice. Recimo, da je G7 edini (ali zelo redek) zapis akorda G (tudi brez dodane 7. stopnje) v naši učni množici. Če ga izpustimo, naš sistem ne bo imel informacije o kromagramu akorda G. Zato je bolje, da kromagram takšnega

akorda nekako ustvarimo. Z ozirom na glasbeno teorijo, bi lahko problem rešili na naslednji način: Ustvarimo kromagram akorda G7, mu ustrezno zmanjšamo amplitudo 7. stopnje in ga shranimo kot kromagram akorda G. Vendar četvero-zvok z dodano 7. stopnjo še zdaleč ni edini “poseben” akord v našem modelu. Takšnih akordov je veliko. Zato smo se problema lotili nekoliko drugače. Ustvarili smo matriko t.i. umetnih kromagramov. To so kromagrami, ki izhajajo iz idealnega sveta glasbene teorije. V takšnih kromagramih vedno nastopajo samo tisti toni, ki so del akorda, vsi ostali imajo amplitudo enako 0. Primer takšnega kromagrama je na Sliki 5.6. V primeru,



Slika 5.6: Umetni kromagram akorda C-dur.

da v notaciji naletimo na “poseben” akord, ga ustrezno klasificiramo in namesto klasičnega kromagrama uporabimo umetni kromagram. V množico “posebnih” akordov našega modela spadajo vsi četvero-zvoki z dodano 6. in 7. stopnjo, prav tako tudi tisti z dodano zvišano 7. stopnjo (maj7), sus, sus2, sus4 in nekateri intervali.

Drugo vprašanje “posebnih” akordov se nanaša na matriko prehodov. Kot bomo v nadaljevanju omenili, je postopek za gradnjo matrike prehodov štetje pojavitev posameznih parov akordov. V zgornjem primeru imamo 3 pare: C – Dm, Dm – G7, G7 – C. Če akord G7 izpustimo, izgubimo 2 para, kar lahko močno vpliva na pravilnost delovanja. Tudi tukaj lahko uporabimo podoben razmislek, kot pri kromagramih. Če takšne akorde obravnavamo kot trozvoke, bomo v matriko prehodov vnesli bistveno več informacij, le-te pa

zaradi specifik našega modela, ne bodo napačne. Potrebno je še omeniti, da je na tem mestu izpuščanje akorda iz vektorja nesprejemljivo. Če iz zgornjega zaporedja preprosto izpustimo G7, dobimo prehod $Dm - C$, ki lahko ni v skladu z glasbeno teorijo in skladbo, ter nam lahko v matriko prehodov vnese napačne informacije.

V notacijah nastopajo tudi akordi, ki jih ne moremo obravnavati tako kot v zgornjem primeru. To so večinoma zvišani in znižani trozvoki. Ker naš sistem takšnih akordov ne generira, jih moramo pri učenju obravnavati posebej. Zvišani in znižani trozvoki niso podobni nobenemu izmed stanj našega sistema, zato jih moramo izpustiti. Vse neprepoznane akorde in izjeme klasificiramo v 25. stanje – NC.

5.4.4 Priprava podatkovnih množic in izgradnja matrik

Postopek branja in urejanja akordov opravimo nad vsemi segmenti skladb učne množice. Rezultat procesiranja posamezne skladbe zapišemo v naslednji obliki:

- `obsData[]` vektor dimenzij $12 \times n$; vsebuje n kromagramov
- `hiddenData[]` vektor dimenzij $1 \times n$; vsebuje številsko oznako akorda $[1 \dots 25]$

Vektorje vseh skladb združimo v množico na način, ki ga pričakuje HMM Toolbox. Preostane nam le še klic ustrezne funkcije, ki nam zgradi vse potrebne matrike:

- Vektor začetnih verjetnosti
- Matrika prehodnih verjetnosti
- Matrika emisij

Namesto naučenega vektorja začetnih verjetnosti pri našem modelu uporabljamo enakomerno porazdelitev – vsa začetna stanja so enako verjetna. S

tem se izognemo težavam iz naslova velikosti učne množice. Začetno stanje, ki ne bi bilo prisotno v učni množici bi bilo v takšnem modelu nemogoč dogodek.

5.4.5 Matrika prehodnih verjetnosti

Vrednosti v matriki prehodnih verjetnosti so normalizirane vsote vseh parov akordov v vseh skladbah učne množice. Funkcija pregleda vse pare akordov in na ustrezno mesto v matriki zapisuje število pojavitev. Matematično je matrika opredeljena na naslednji način:

$$T = [p_{i,j}]; \quad p_{i,j} = P(X_{n+1} = j \mid X_n = i) \quad (5.4)$$

$$0 \leq p_{i,j} \leq 1 \quad (5.5)$$

$$\sum_j p_{i,j} = 1 \quad (5.6)$$

Matrika prehodnih verjetnosti je stohastična. Za izdelavo ustrezne matrike prehodnih verjetnosti je potrebna zadostna količina učnih podatkov. V primeru majhnih učnih množic se lahko zgodi, da se določeni akordi nikoli ne pojavijo. Za takšen akord ne pridobimo niti kromagrama, niti zapisa v matriki prehodnih verjetnosti. Pri našem modelu je učna množica dovolj velika, da ustvarimo vseh 25 kromagramov.

Matrika prehodnih verjetnosti beleži prehode med akordi. Če se določen par akordov v učni množici ne pojavi, bo verjetnost prehoda enaka 0. Prehod v takšno stanje bo nemogoč dogodek. Ker naša učna množica ni zelo velika, obstaja verjetnost, da na ta način onemogočimo prehod, ki bi bil glede na glasbeno teorijo povsem legalen. Ničelne vrednosti matrike prehodov pri našem modelu nadomestimo z vrednostmi, ki so nekajkrat manjše od najmanjših neničelnih vrednosti.

V Prilogi 1 je grafični prikaz matrike prehodnih verjetnosti. Podroben pregled matrike pokaže skladnost z glasbeno teorijo. Matrika je zaradi boljšega kontrasta prikazana z zmanjšanimi vrednostmi na diagonali. Diagonala predstavlja verjetnost prehoda akorda samega vase. V poglavju o

zaznavanju tempa in segmentaciji smo omenili, da skladbe segmentiramo glede na zaznan tempo. Ker posamezen akord traja več dob tempa, je verjetnost prehoda samega vase tolikokrat večja. Ker isti postopek uporabljamo tudi pri evalvaciji modela, so vrednosti na diagonali ustrezne, čeprav zaradi visoke verjetnosti nekoliko neprimerne za prikaz.

5.4.6 Matrika emisij

Matrika emisij vsebuje informacije o izračunanih kromagramih. Funkcijsko sestoji iz dveh sklopov:

- Matrika srednjih vrednosti μ
- Kovariančna matrika Σ

Funkciji za izdelavo matrik smo podali zaporedje kromagramov s pripadajočimi stanji – oznakami akordov. Funkcija za vsako izmed stanj izračuna Gaussovo krivuljo - srednjo vrednost in kovariančno matriko. Srednja vrednost je kromagram, povprečen preko vseh kromagramov posameznega stanja. Velikost matrike μ je torej 12 x 25. Vsaki srednji vrednosti pripada kovariančna matrika, ki predstavlja odvisnost med toni posameznega kromagrama (Slika 5.7). Matrika emisij vsebuje 25 kromagramov, primera za C-dur in A-mol sta na Sliki 5.1

5.5 Evalvacija

Postopek evalvacije, gradnje najprimernejšega zaporedja akordov, je podoben postopku učenja. Vsako skladbo najprej segmentiramo glede na tempo. Iz posameznih segmentov zgradimo kromagrame. Pri postopku evalvacije nimamo na voljo notacij, saj je naš cilj njihova izdelava. Originalne notacije uporabimo v zadnji fazi, to je izračun pravilnosti generiranih akordov. Na Sliki 5.8 je prikazan diagram poteka za postopek evalvacije.

	A	Bb	B	C	C#	D	D#	E	F	F#	G	G#
A	0.04	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
Bb	-0.00	0.01	-0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.00
B	-0.00	-0.00	0.02	-0.00	0.00	-0.01	0.00	0.00	0.00	-0.00	0.00	0.00
C	-0.00	0.00	-0.00	0.02	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00
C#	-0.00	0.00	0.00	-0.00	0.01	-0.00	0.00	0.00	0.00	-0.00	0.00	0.00
D	0.00	-0.00	-0.01	-0.00	-0.00	0.04	-0.00	-0.01	-0.00	0.00	-0.00	-0.00
D#	-0.00	0.00	0.00	0.00	0.00	-0.00	0.01	-0.00	0.00	-0.00	0.00	0.00
E	-0.00	-0.00	0.00	-0.00	0.00	-0.01	-0.00	0.02	-0.00	-0.01	0.00	0.00
F	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.01	-0.00	0.00	0.00
F#	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.01	-0.00	0.04	-0.00	-0.00
G	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.02	-0.00
G#	-0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.01

Slika 5.7: Kovariančna matrika akorda D-dur. Opazimo višje vrednosti po diagonali, posebno pri tonih D, F# in A, ki sestavljajo akord D-dur.

5.5.1 Viterbi

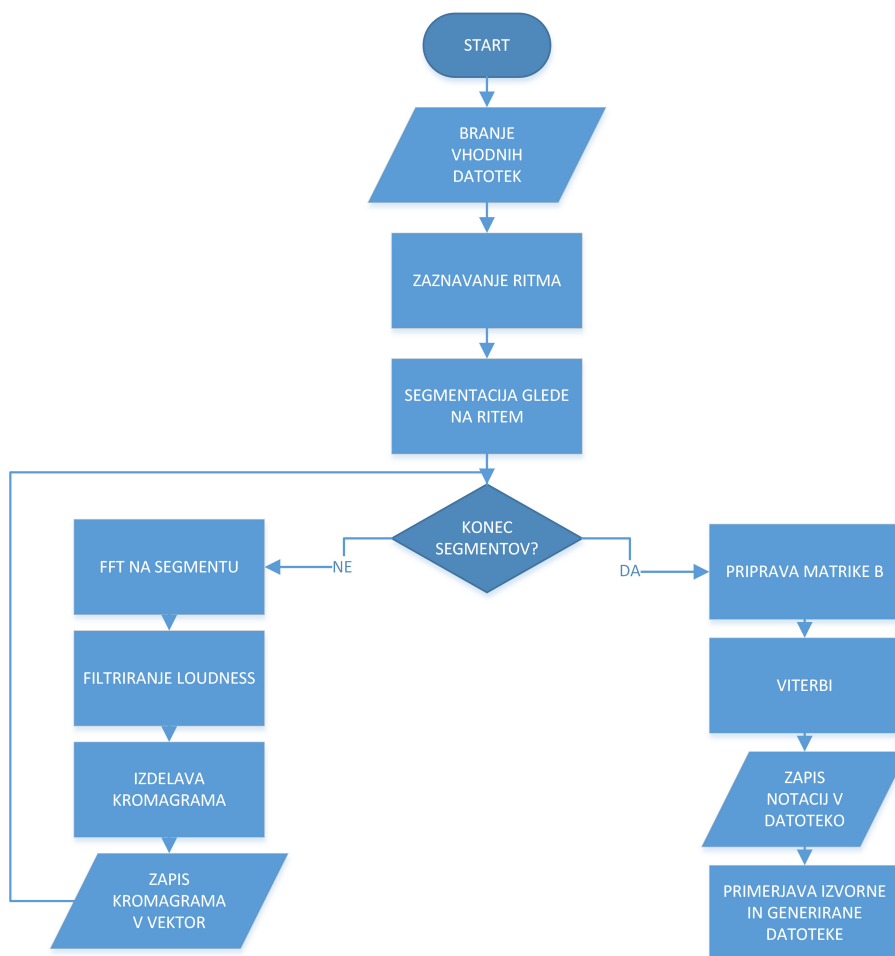
Pri postopku evalvacije za vsak segment skladbe izdelamo kromagram. Ker nimamo podatka o stanju, ki je povzročilo frekvenčni odziv v določenem segmentu, so veljavni vsi kromagrami. Kromagrame zapišemo v vektor; dolžina vektorja je enaka številu dob posamezne skladbe. Algoritem Viterbi za svoje delovanje potrebuje naslednje podatke:

- Vektor začetnih verjetnosti
- Matriko prehodnih verjetnosti
- Matriko B

Vektor začetnih verjetnosti je enakomerno porazdeljen. Matriko prehodnih verjetnosti se je naš model naučil v fazi učenja. Matriko B je potrebno izdelati s pomočjo matrike emisij in vektorja kromagramov posamezne skladbe. Elementi matrike B so naslednji:

$$B(i, t) = P(y_t \mid q_t=i), \text{ za vse } t, i$$

y...vektor opazovanih stanj
q...vektor skritih stanj



Slika 5.8: Diagram poteka postopka evalvacije.

Matrika B za vsak kromagram posamezne skladbe izraža verjetnost, da je le-ta posledica skritega stanja i . Matrika je dimenzij $25 \times t$, kjer je t število kromagramov posamezne skladbe. Algoritem Viterbi iz vhodnih podatkov izdelava vektor numeričnih stanj, ki ustrezajo notacijam akordov. Vektor predstavlja najverjetnejše zaporedje akordov, pri podanih vhodnih matrikah.

5.5.2 Zapis v datoteko

Izdelan vektor zaporedja akordov je potrebno pretvoriti v obliko, ki jo bomo lahko enostavno primerjali z originalnimi notacijami. Numerična stanja akordov model pretvori v standardno notacijo akordov. Z uporabo informacij o tempu, se vektor notacij zapiše v datoteko s pripadajočimi časovnimi oznakami. Ker vektor notacij vsebuje 1 zapis za vsako dobo, je potrebno zapis nekoliko poenostaviti. Akordi večinoma trajajo več dob zaznanega tempa, zato je v vektorju mnogo zaporednih notacij istega akorda. Takšne notacije je potrebno združiti in ustrezno popraviti časovne oznake. Urejen zapis olajša pregledovanje datoteke.

5.5.3 Analiza pravilnosti prepoznave

Za vsako skladbo pripravimo datoteko s prepoznanimi akordi. Datoteko nato primerjamo z originalnimi notacijami in izračunamo odstotek podobnosti. Postopek računanja podobnosti je dokaj enostaven. Akorde iz obeh datotek preberemo v vektor. Akorde originalnih notacij je potrebno ustrezno klasificirati zaradi posebnih akordov in nekonsistenčnih notacij (glej poglavje o branju in urejanju akordov pri postopku učenja). Oba vektorja razdelimo na dovolj majhne segmente. S tem poskrbimo, da je odstotek pravilnosti ustrezen tudi v primeru, ko je trajanje akorda v vektorjih različno. V nasprotnem primeru bi bila kazen za napačno prepoznani akord in / ali trajanje, prevelika. Potrebno je omeniti, da je ocena, pridobljena na ta način, ustrezna in nepristranska. Za vsak segment primerjamo enakost akordov v obeh vektorjih.

$$\text{Pravilnost prepoznave} = \frac{\# \text{ ujemanj}}{\# \text{ vseh segmentov}} \quad (5.7)$$

5.6 Analiza rezultatov

Postopek evalvacije temelji na principu 10-kratnega prečnega preverjanja. To pomeni, da vhodne podatke razdelimo v učno in testno množico v razmerju 90:10. Naš sistem naučimo z 90% skladb, preostalih 10% skladb pa uporabimo za testiranje. Postopek ponovimo 10 krat, vsakič izberemo drugih 10% testnih podatkov. Tako dosežemo dobro statistično oceno natančnosti klasifikacije. Rezultati validacije skladb so navedeni v Prilogi 2. Velikost celotne množice vhodnih podatkov je 200 skladb.

Skupna povprečna natančnost klasifikacije našega modela znaša 0,62. Natančnost Monte Carlo sistema, ki bi vsak akord izbral naključno, bi pri skladbi z 80 akordi znašala približno $1.46 * 10^{-112}$. Natančnosti sistemov, ki sodelujejo na tekmovanju MIREX, segajo čez 0,75. V nadaljevanju bomo omenili dodatne možnosti, s katerimi bi lahko izboljšali natančnost klasifikacije našega sistema.

V Prilogi 3 je prikazana matrika napak. Vrstice predstavljajo akorde, ki jih je klasificiral naš sistem, stolci pa akorde, ki so navedeni v notacijah. Vrednost polja predstavlja odstotek napak – kolikokrat je sistem dejanski akord (stolpec) napačno klasificiral kot akord v vrstici. Vrednosti po diagonali so nične, saj predstavljajo pravilno prepoznane akorde.

Z analizo matrike napak lahko poiščemo vzroke za napačno klasifikacijo našega sistema. Polje C#m - C navaja 53-odstotno napačno klasifikacijo akorda C-dur v C#m. Akord C#m7b5, ki ga naš sistem obravnava kot C#m ima npr. 2 skupna tona z akordom C-dur (to sta tona E in G). Akord C-dur bo, ob neizrazitem tonu C, napačno klasificiran kot C#m. Podobno lahko analiziramo tudi napako med A-dur in D-dur. Akord Asus4, ki ga naš sistem obravnava kot A-dur, ima z akordom D-dur 2 skupna tona, kar lahko povzroči napačno klasifikacijo. V nadaljevanju bomo predlagali možne izboljšave sistema, ki bi lahko odstranile mnogo napačnih klasifikacij.

5.6.1 Analiza klasifikacije na nivoju skladb

Rezultati, navedeni v prejšnjem poglavju, nam ustvarijo predstavo o delovanju sistema nad množico skladb. Potrebno se je zavedati, da so odkloni pravilnosti prepoznavne zelo veliki, zato velja analizo opraviti tudi na nivoju posameznih skladb. V analizi bomo uporabili 2 skladbi in skušali poiskati razloge za podano pravilnost klasifikacije. Na ta način bomo spoznali specifične našega modela. Opisali bomo situacije, pri katerih se naš sistem odreže dobro, pa tudi tiste, kjer je odstotek nepravilnih prepoznav velik. Dobro razumevanje slednjih, nam omogoča izboljšave in nadgradnje modela.

The Beatles – You’ve Really Got a Hold on Me

Skladba You’ve Really Got a Hold on Me skupine The Beatles je pri analizi dosegla najvišji odstotek pravilnosti prepoznavne in sicer 0.92. Primerjava prepoznane sekvence akordov z referenčnimi akordi je prikazana v Tabeli 5.4.

Iz Tabele 5.4 razberemo, da je ujemanje zadovoljivo. Opazimo napačno klasifikacijo v C#min pri 9. sekundi. Akord C#min sicer spada v lestvico Adur, v kateri je posneta skladba, tako, da je napaka v resnici manjša. Ostale napake lahko pripišemo posnetku – opazimo, da se časi začetkov in koncev akordov ne ujemajo z referenčnimi. To je posledica različnih posnetkov in medijev. Opazimo tudi, da je prepoznanih akordov več kot referenčnih. To je posledica strukture in normalizacije matrike prehodov modela in je parameter, ki ga je potrebno nastaviti empirično. Kot je opisano v poglavju o analizi pravilnosti prepoznavne, je število akordov ustrezno normalizirano in nima neposrednega vpliva na oceno pravilnosti prepoznavne. Razlogi za uspešno klasifikacijo akordov skladbe se skrivajo v skladbi sami. Le-ta vsebuje stalno kitarsko spremljavo, iz katere izhajajo dobri kromagrami. Skladba ne vsebuje preveč tišine in neharmoničnih elementov, ki bi v rezultate vnesli nepravilne klasifikacije.

Referenčni akordi			Prepoznani akordi		
0.000000	0.440395	N	0.000000	0.472000	N
0.440395	2.054309	E	0.472000	1.192000	E
2.054309	5.038074	A	1.192000	4.812000	A
5.038074	8.010228	F#:min	4.812000	7.784000	F#:min
8.010228	10.982382	A	7.784000	9.044000	A
10.982382	13.977756	F#:min	9.044000	10.020000	C#:min
13.977756	20.096214	A	10.020000	11.032000	A
20.096214	26.226282	F#:min	11.032000	13.760000	F#:min
26.226282	29.291316	A	13.760000	19.896000	A
29.291316	32.367960	D	13.760000	19.896000	A
32.367960	33.865647	B:7	19.896000	26.032000	F#:min
33.865647	35.421384	E	26.032000	29.364000	A
35.421384	41.621112	A	29.364000	31.924000	D
41.621112	47.867280	F#:min	31.924000	32.428000	A
47.867280	54.148278	A	32.428000	33.948000	B
54.148278	60.394445	F#:min	33.948000	35.484000	E
60.394445	63.540749	A	35.484000	41.440000	A
63.540749	66.640613	D	41.440000	47.956000	F#:min
66.640613	68.219570	B:7	47.956000	54.188000	A
68.219570	69.786917	E	54.188000	60.156000	F#:min
69.786917	76.079525	A	60.156000	63.564000	A
			63.564000	66.160000	D
			66.160000	66.688000	A
			66.688000	68.240000	B
			68.240000	69.800000	E
			69.800000	75.816000	A

Tabela 5.4: Primerjava referenčnih in prepoznanih akordov za uvodnih 75 sekund skladbe You've Really Got a Hold on Me

Queen – Another One Bites The Dust

Pravilnost klasifikacije skladbe Another One Bites The Dust znaša 0.05, kar skladbo uvršča med najslabše klasificirane. Tabele primerjav na tem mestu ne bomo navajali, saj vsebuje večinoma napačne klasifikacije. Razloge za slab

rezultat je potrebno poiskati v posnetku. Pri poslušanju skladbe takoj opazimo, da skoraj ne vsebuje harmonične vsebine. Skladba temelji na izrazitem ritmu, ki ga ustvarjajo bobni in bas. Tudi vokal večinoma sledi basovski liniji – prisotnost harmoničnih intervalov je nizka. Prav tako je v skladbi prisotno mnogo neharmonične vsebine – umetni zvoki, ropot, šumenje, tišina. Ti elementi v prepoznavo vnašajo veliko šuma in povzročajo nejasne kromagrame. Skladba vsebuje nekaj sekund harmonične spremljave, vendar ne dovolj, za uspešno delovanje HMM. Prepoznani akordi zato nimajo nikakršne povezave z referenčnimi.

5.6.2 Izboljšave sistema

Kot smo navedli v prejšnjem poglavju, bi lahko naš sistem nadgradili z dodatnimi funkcijami, ki bi izboljšale natančnost klasifikacije. Nekatere možnosti smo pri opisu delovanja že omenili.

Velikost učne množice

Velikost učne množice je klasična problematika strojnega učenja. V našem primeru je velikost učnih podatkov pomembna predvsem za generiranje matrike prehodov. Pri manjšem številu vhodnih podatkov se sistem neobstojećih kombinacij ne nauči, kljub temu, da so lahko popolnoma legalne. Zato je poleg velikosti pomembna tudi heterogenost učne množice. Če bi za učne podatke uporabili samo množico skladb enega samega izvajalca, bi to lahko povzročilo, da bi naš sistem dobro klasificiral samo skladbe, ki bi bile na nek način podobne tistim iz učne množice. Če bi želeli izdelati sistem, ki bi dovolj dobro klasificiral poljubno zvrst glasbe, bi bilo potrebno uporabiti temu primerno učno množico. Sistemi z ekspertnimi bazami znanja so pri evalvaciji poljubnih skladb bolj nepristranski [18], [19].

Izdelava ločenih kromagramov za bas

Ker basovski ton ne sledi vedno tonom v akordu, je smiselno izdelati ločen kromagram za frekvenčno območje basa [10]. Tako dosežemo neodvisnost vzorcev, potrebno pa je izdelati več kromagramov. Z ločenimi kromagrami bi lahko pri skladbi *Another One Bites The Dust* dosegli boljše rezultate, saj celotna skladba temelji na basovski liniji.

Uporaba omejitev lestvic in tonalitete

Prehodi med akordi so specifični za vsako izbrano lestvico - tonaliteto. Če je skladba napisana v določeni tonaliteti, bo uporabljala akorde in prehode med akordi, ki spadajo v to domeno. Če matriko prehodov zgradimo tako, da zaznamo in upoštevamo tonaliteto skladbe, bomo dosegli manj šuma – matrika prehodov bo vsebovala samo tiste prehode, ki so del zaznane tonalitete [10]. V naši implementaciji npr. obstaja prehod iz C-dur v F-dur, ki je v nekaterih lestvicah legalen, v drugih pa ne. Pri implementaciji zaznave tonalitete je potrebno upoštevati tudi menjavo ključa znotraj skladbe. Tudi omejitev lestvic in tonalitete bi se lahko izkazala za uporabno pri skladbi, kot je *Another One Bites The Dust*, kjer je prisotnih malo harmoničnih informacij in je potrebno čim bolj zmanjšati nabor možnih kandidatov za akorde.

Mešanje Gaussovih krivulj

V poglavju o branju in urejanju akordov smo navedli možno rešitev za problem “posebnih” akordov. Bistveno boljše rešitev problema ponuja postopek mešanja Gaussovih krivulj. Postopek je dokaj enostaven in smiseln, vendar zahteva podporo znotraj Toolbox-a HMM. Toolbox, ki ga uporabljamo pri našem modelu, mešanja Gaussovih krivulj žal ne omogoča, vseeno pa si pogledajmo osnovno idejo in prednosti, ki jih postopek omogoča. Predpostavimo, da imamo v naši učni množici naslednje akorde, ki temeljijo na trozvoku C-dur:

- C

- C7
- Cmaj7
- Csus4

V našem trenutnem postopku bomo za prvi akord izdelali kromagram, za vse ostale pa uporabili umetni kromagram. Ideja mešanja krivulj je v tem, da za vse 4 akorde izdelamo kromagrame. Vsak dodaten kromagram predstavlja dodatno dimenzijo. Določiti je potrebno, koliko dimenzij bo imel vsak akord. Matriko emisij izdelamo za vsako dimenzijo posebej. Akord C-dur je v takšnem sistemu predstavljen kot ena izmed 4 dimenzij – kromagramov. Z uporabo mešanja Gaussovih krivulj lahko dosežemo bistveno boljše rezultate, saj nam “posebni” akordi ne predstavljajo več šuma, temveč posebno dimenzijo znotraj istega stanja. Če naš sistem pri evalvaciji naleti na kromagram, ki ustreza kateri koli izmed štirih dimenzij akorda C-dur, bo stanje klasificiral kot C-dur.

Poglavje 6

Zaključek

V diplomski nalogi smo predstavili tehnično in teoretično ozadje sistema za samodejno klasifikacijo akordov. Sistem smo implementirali v razvojnem okolju Matlab. Vsi gradniki sistema so podrobno razloženi. Naš sistem na glasbeni zbiriki Isophonics dosega povprečno klasifikacijsko točnost 62%. Pri nekaterih skladbah dosega sistem več kot 90% klasifikacijsko točnost.

Gradnike modela HMM smo predstavili v grafični obliki. Vizualizacija nam omogoča lažje razumevanje specifik modela in iskanje razlogov za napačne klasifikacije. Prav tako lahko naučeno bazo znanja preverimo z ozirom na glasbeno teorijo, kar nam omogoča implementacijo dodatnih funkcionalnosti in izboljšav sistema. Za vrednotenje posameznih segmentov uporabljamo metriko kromagram z dodatnim predprocesiranjem - uporaba psihoakustičnega filtra in okenske funkcije.

Pri delu na diplomski nalogi smo pridobili mnogo koristnega znanja s področij glasbene teorije in statističnih modelov. Spoznali smo različne pristope k reševanju problema prepoznavne akordov. Za implementacijo smo izbrali moderen model, ki se danes uporablja v večini podobnih aplikacij. Le-te dosega bistveno boljše rezultate, kot naša implementacija, vendar klasifikacijska točnost našega sistema še zdaleč ni zanemarljiva. Z uporabo metod, predlaganih v prejšnjem poglavju, bi naš sistem lahko približali modernim sistemom.

Cilj našega dela je bila izdelava delujočega sistema za klasifikacijo akordov, zato je bilo delo razporejeno horizontalno. Določene podrobnosti so izpuščene oz. poenostavljene. Pri nadaljnjem delu bi se bilo potrebno osredotočiti na implementacijo oz. dodelavo takšnih podrobnosti. S tem bi dosegli boljšo natančnost sistema in odprli vrata mnogim aplikacijam. Možnost uporabe sistema za izdelavo t.i. “Lead Sheet” notacij smo že omenili. Prav tako bi takšen sistem lahko uporabili kot pripomoček pri učenju prenekateriga instrumenta. Z vedno večjo procesorsko močjo se odpirajo tudi vrata analiziranja v realnem času. Takšen sistem bi s pridom uporabljali improvizatorji in ostali ustvarjalci.

Literatura

- [1] Janka Istenič in drugi. “Stopinje v preteklost. Narodni muzej Slovenije”, 2006.
- [2] Turk Matija, Dimkaroski Ljuban. “Neandertalska piščal iz Divjih bab I: stara in nova spoznanja”. Založba ZRC. 2011
- [3] Rahul Siddharthan. “Music, Mathematics and Bach”. 1999
- [4] Peek Hans B. “The Emergence of the Compact Disc”. IEEE Communications Magazine 48. 2010.
- [5] Huber David Miles. “The MIDI Manual”. 1991.
- [6] Nave C.R. “Sensitivity of Human Ear”. Department of Physics and Astronomy Georgia State University
- [7] Steven W. Smith, Ph.D. “The Scientist and Engineer’s Guide to Digital Signal Processing”. 2002.
- [8] Ross, H.E. “E.H.Weber on the tactile senses”. 1996.
- [9] Pavle Mihelčič. “Osnove glasbene teorije”. Založba DZS. 2006
- [10] Yizhao Ni, Matt Mcvigar, Raul Santos-Rodriguez and Tijl De Bie. “Harmony progression analyzer for MIREX 2011”.
- [11] Kevin Murphy. “Hidden Markov Model (HMM) Toolbox for Matlab”. 1998.

- [12] Podatkovna zbirka Isophonics. Dostopno na:
<http://isophonics.net/content/reference-annotations>
- [13] T. D. Rossing. “The science of sound (second edition)”. Addison-Wesley, 1990.
- [14] Daniel P.W. Ellis, LabROSA. “Beat Tracking by Dynamic Programming”. 2007.
- [15] Yizhao Ni, Matt McVicar, Raúl Santos-Rodríguez and Tijl De Bie. “An end-to-end machine learning system for harmonic analysis of music”. 2011.
- [16] T. Cho, R. J. Weiss, and J. P. Bello, “Exploring common variations in state of the art chord recognition systems”. 2010.
- [17] C. Harte, M. Sandler, and M. Gasser. “Detecting harmonic change in musical audio”. 2006.
- [18] Johan Pauwels, Jean-Pierre Martens, Geoffroy Peeters. “The ircam-keychord submission for MIREX 2012”. 2012.
- [19] W. Bas de Haas, José Pedro Magalhães, Frans Wiering. “Improving audio chord transcription by exploiting harmonic and metric knowledge”. 2012.
- [20] T. Fujishima, “Real time chord recognition of musical sound: a system using common lisp music”. 1999.

Priloge

Priloga 1: Matrika prehodnih verjetnosti

Vrstice - izvorni akord, stolpci - ponorni akord. Vrednosti na diagonali so zmanjšane zaradi boljsega kontrasta.

	C	Cm	C#	C#m	D	Dm	D#	D#m	E	Em	F	Fm	F#	F#m	G	Gm	G#	G#m	A	Am	Bb	Bbm	B	Bm	N
C	0.37	0.00	0.00	0.00	0.09	0.02	0.00	0.00	0.02	0.02	0.11	0.02	0.00	0.01	0.21	0.01	0.00	0.00	0.01	0.03	0.01	0.00	0.04	0.00	0.02
Cm	0.00	0.39	0.00	0.00	0.02	0.01	0.01	0.00	0.01	0.00	0.17	0.08	0.00	0.00	0.05	0.06	0.00	0.01	0.03	0.00	0.01	0.01	0.10	0.00	0.04
C#	0.12	0.02	0.30	0.00	0.08	0.00	0.00	0.00	0.23	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.08	0.00	0.06	0.00	0.00	0.02	0.02	0.00	0.04
C#m	0.00	0.00	0.00	0.41	0.00	0.00	0.00	0.00	0.04	0.00	0.08	0.00	0.00	0.19	0.02	0.00	0.00	0.01	0.11	0.00	0.00	0.00	0.04	0.03	0.07
D	0.04	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.07	0.02	0.02	0.00	0.00	0.01	0.19	0.06	0.00	0.00	0.21	0.01	0.00	0.01	0.01	0.03	0.02
Dm	0.06	0.00	0.00	0.00	0.01	0.44	0.00	0.00	0.07	0.02	0.05	0.01	0.00	0.00	0.11	0.06	0.00	0.00	0.08	0.02	0.01	0.01	0.05	0.00	0.01
D#	0.00	0.02	0.06	0.00	0.01	0.00	0.26	0.00	0.05	0.00	0.02	0.08	0.00	0.00	0.00	0.02	0.07	0.00	0.08	0.00	0.10	0.00	0.07	0.00	0.15
D#m	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.07	0.07	0.00	0.00	0.28	0.14	0.00	0.00	0.03	0.00	0.00	0.03
E	0.02	0.00	0.00	0.02	0.05	0.00	0.00	0.00	0.44	0.00	0.01	0.01	0.00	0.02	0.03	0.00	0.01	0.01	0.21	0.03	0.00	0.00	0.09	0.01	0.02
Em	0.09	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.38	0.04	0.00	0.00	0.02	0.09	0.00	0.00	0.00	0.16	0.13	0.00	0.00	0.02	0.01	0.03
F	0.13	0.03	0.00	0.00	0.02	0.02	0.00	0.01	0.05	0.01	0.32	0.01	0.00	0.00	0.11	0.02	0.00	0.00	0.03	0.03	0.02	0.01	0.12	0.02	0.05
Fm	0.13	0.00	0.08	0.00	0.00	0.00	0.00	0.34	0.04	0.00	0.02	0.34	0.00	0.00	0.05	0.01	0.02	0.03	0.04	0.00	0.03	0.03	0.12	0.00	0.04
F#	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.12	0.00	0.16	0.06	0.00	0.00	0.00	0.00	0.36	0.00	0.06	0.00	0.00	0.00	0.12
F#m	0.02	0.01	0.00	0.05	0.09	0.01	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.35	0.03	0.00	0.00	0.02	0.06	0.01	0.00	0.00	0.22	0.05	0.03
G	0.18	0.00	0.00	0.01	0.16	0.00	0.00	0.00	0.02	0.02	0.04	0.00	0.00	0.01	0.34	0.02	0.00	0.00	0.08	0.04	0.00	0.00	0.02	0.02	0.03
Gm	0.13	0.08	0.00	0.00	0.15	0.03	0.00	0.00	0.02	0.00	0.07	0.01	0.00	0.04	0.01	0.29	0.02	0.01	0.05	0.01	0.00	0.01	0.03	0.01	0.01
G#	0.01	0.01	0.01	0.00	0.02	0.00	0.05	0.00	0.11	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.42	0.00	0.07	0.00	0.01	0.11	0.13	0.00	0.01
G#m	0.08	0.00	0.00	0.01	0.24	0.02	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.12	0.02	0.17	0.00	0.22	0.15	0.00	0.01	0.00	0.13	0.00	0.03
A	0.02	0.00	0.00	0.01	0.00	0.02	0.00	0.00	0.14	0.01	0.01	0.00	0.00	0.02	0.04	0.01	0.00	0.00	0.38	0.01	0.00	0.00	0.04	0.02	0.03
Am	0.10	0.00	0.00	0.00	0.09	0.05	0.00	0.00	0.03	0.07	0.05	0.01	0.01	0.00	0.09	0.02	0.00	0.01	0.01	0.40	0.00	0.00	0.02	0.02	0.03
Bb	0.27	0.01	0.01	0.00	0.01	0.02	0.12	0.00	0.02	0.00	0.07	0.02	0.00	0.00	0.01	0.04	0.01	0.00	0.02	0.03	0.33	0.00	0.01	0.00	0.00
Bbm	0.03	0.01	0.01	0.00	0.06	0.00	0.05	0.02	0.02	0.00	0.15	0.10	0.00	0.00	0.03	0.01	0.10	0.00	0.06	0.00	0.00	0.33	0.00	0.00	0.01
B	0.06	0.03	0.00	0.01	0.02	0.01	0.01	0.00	0.22	0.02	0.07	0.01	0.00	0.02	0.02	0.01	0.01	0.01	0.08	0.01	0.00	0.01	0.35	0.01	0.03
Bm	0.03	0.00	0.00	0.01	0.13	0.00	0.00	0.00	0.07	0.08	0.00	0.00	0.00	0.06	0.10	0.03	0.00	0.00	0.09	0.04	0.00	0.00	0.01	0.33	0.03
N	0.06	0.01	0.00	0.01	0.06	0.02	0.00	0.00	0.07	0.02	0.07	0.01	0.00	0.00	0.07	0.01	0.00	0.00	0.07	0.02	0.00	0.00	0.03	0.01	0.44

Priloga 2: Rezultati prepoznavne akordov posameznih skladb

Skladba	Odstotek	Skladba	Odstotek
A Kind Of Magic	0,87	Nowhere Man	0,66
Bohemian Rhapsody	0,32	Oh! Darling	0,80
I Feel The Earth Move	0,58	Rock and Roll Music	0,84
A Hard Day's Night	0,58	Beautiful	0,47
Come Together	0,51	Bicycle Race	0,39
Drive My Car	0,52	I Want To Break Free	0,85
Help!	0,70	And I Love Her	0,73
I Saw Her Standing There	0,66	Another Girl	0,54
It Won't Be Long	0,70	Boys	0,78
Magical Mystery Tour	0,90	Dig It	0,58
No Reply	0,77	Fixing A Hole	0,57
Sgt. Pepper's Lonely Hearts Club Band	0,69	Here, There And Everywhere	0,56
Taxman	0,61	I'll Follow the Sun	0,72
Two of Us	0,65	Little Child	0,85
Another One Bites The Dust	0,05	Octopus's Garden	0,84
So Far Away	0,65	Think For Yourself	0,59
All I've Got To Do	0,83	Your Mother Should Know	0,81
Dig a Pony	0,67	Way Over Yonder	0,26
Eleanor Rigby	0,85	You're My Best Friend	0,61
I'm a Loser	0,61	Ask Me Why	0,70
I Should Have Known Better	0,86	I Am The Walrus	0,76
Misery	0,81	I Want You	0,47
Norwegian Wood (This Bird Has Flown)	0,87	Let It Be	0,78
Something	0,70	Mr. Moonlight	0,59
The Fool On The Hill	0,74	She's Leaving Home	0,58
The Night Before	0,63	Tell Me Why	0,44
With A Little Help From My Friends	0,73	The Word	0,67
It's Too Late	0,65	Till There Was You	0,67
Across the Universe	0,13	Yellow Submarine	0,47
All My Loving	0,81	You're Going To Lose That Girl	0,75
Anna (Go To Him)	0,80	Don't Stop Me Now	0,54
Baby's In Black	0,79	You've Got A Friend	0,13
Flying	0,86	Being For The Benefit Of Mr. Kite!	0,45
I'm Only Sleeping	0,33	Can't Buy Me Love	0,82
If I Fell	0,74	Hello Goodbye	0,77
Lucy In The Sky With Diamonds	0,70	Here Comes The Sun	0,74
Maxwell's Silver Hammer	0,77	Kansas City- Hey, Hey, Hey, Hey	0,48
You've Got To Hide Your Love Away	0,66	Maggie Mae	0,79
You Won't See Me	0,74	Michelle	0,45
Fat Bottomed Girls	0,77	Please Mister Postman	0,89
Home Again	0,73	Please Please Me	0,83
I Want It All	0,57	She Said She Said	0,36
Blue Jay Way	0,06	Ticket To Ride	0,64
Chains	0,61	Save Me	0,70
Don't Bother Me	0,41	Act Naturally	0,92
Getting Better	0,61	Any Time At All	0,64
I'm Happy Just To Dance With You	0,35	Because	0,50
I Me Mine	0,59	Eight Days a Week	0,78
I Need You	0,57	Good Day Sunshine	0,82
Love You To	0,15	I've Got A Feeling	0,57

Skladba	Odstotek	Skladba	Odstotek
Love Me Do	0,70	Wait	0,63
Roll Over Beethoven	0,73	You Can't Do That	0,69
Strawberry Fields Forever	0,36	Play The Game	0,61
What Goes On	0,74	A Day In The Life	0,54
Within You Without You	0,00	Got To Get You Into My Life	0,56
Crazy Little Thing Called Love	0,44	I'll Be Back	0,58
Who Wants To Live Forever	0,59	If I Needed Someone	0,80
And Your Bird Can Sing	0,87	Not A Second Time	0,49
Girl	0,66	She Came In Through The Bathroom W.	0,77
Hold Me Tight	0,55	There's A Place	0,70
I'll Cry Instead	0,91	What You're Doing	0,66
It's Only Love	0,77	Yesterday	0,58
One After	0,63	Hammer To Fall	0,56
P. S. I Love You	0,81	Dizzy Miss Lizzy	0,82
Penny Lane	0,54	Everybody's Trying to Be My Baby	0,84
When I'm Sixty-Four	0,22	Golden Slumbers	0,77
Words of Love	0,81	Money	0,69
You Never Give Me Your Money	0,68	Run For Your Life	0,82
Somebody To Love	0,12	Tomorrow Never Knows	0,72
Baby It's You	0,74	Twist And Shout	0,87
Baby You're A Rich Man	0,57	Friends Will Be Friends	0,72
For No One	0,51	Seven Seas Of Rhye	0,59
Honey Don't	0,64	Carry That Weight	0,85
I'm Looking Through You	0,09	We Will Rock You	0,21
Lovely Rita	0,34	The End	0,53
Sun King	0,69	We Are The Champions	0,48
The Long and Winding Road	0,25	Her Majesty	0,45
Things We Said Today	0,80	Back in the USSR	0,79
You Like Me Too Much	0,77	Dear Prudence	0,57
You Really Got A Hold On Me	0,92	Glass Onion	0,58
All You Need Is Love	0,66	Ob-La-Di, Ob-La-Da	0,67
Do You Want To Know A Secret	0,67	Wild Honey Pie	0,02
Doctor Robert	0,49	The Continuing Story of Bungalow Bill	0,03
Every Little Thing	0,68	While My Guitar Gently Weeps	0,71
For You Blue	0,60	Happiness is a Warm Gun	0,65
Good Morning Good Morning	0,56	Martha My Dear	0,47
I Wanna Be Your Man	0,78	I'm So Tired	0,75
In My Life	0,84	Black Bird	0,42
Mean Mr Mustard	0,87	Piggies	0,20
Tell Me What You See	0,63	Rocky Raccoon	0,51
When I Get Home	0,74	Don't Pass Me By	0,22
Good Old-Fashioned Lover Boy	0,33	Why Don't We Do It In The Road	0,85
A Taste Of Honey	0,74	I Will	0,64
Devil In Her Heart	0,80	Julia	0,48
Get Back	0,63	Birthday	0,66
I've Just Seen a Face	0,82	Yer Blues	0,48
I Don't Want to Spoil the Party	0,80	Mother Nature's Son	0,64
I Want To Tell You	0,74	Everybody's Got Something To Hide	0,61
Polythene Pam	0,57	Sexy Sadie	0,63
Sgt. Pepper's Lonely Hearts Club	0,71	Helter Skelter	0,72

Priloga 3: Matrika napak

Vrstice – klasificirani akordi, stolpci – dejanski akordi

Prepoznani akordi	Dejanski akordi																									
	C	Cm	C#	C#m	D	Dm	D#	D#m	E	Em	F	Fm	F#	F#m	G	Gm	G#	G#m	A	Am	Bb	Bbm	B	Bm	N	
C	0.00	0.12	0.01	0.02	0.19	0.04	0.00	0.00	0.03	0.03	0.15	0.01	0.00	0.00	0.10	0.01	0.00	0.01	0.04	0.08	0.03	0.03	0.03	0.00	0.00	0.06
Cm	0.14	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.05	0.00	0.21	0.03	0.00	0.00	0.09	0.05	0.08	0.00	0.05	0.02	0.00	0.00	0.03	0.10	0.08	0.00
C#	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C#m	0.53	0.00	0.02	0.00	0.04	0.00	0.00	0.00	0.05	0.00	0.11	0.00	0.00	0.05	0.07	0.00	0.00	0.01	0.06	0.00	0.00	0.00	0.01	0.01	0.04	0.00
D	0.02	0.01	0.00	0.00	0.00	0.13	0.04	0.00	0.18	0.01	0.01	0.01	0.00	0.01	0.16	0.04	0.02	0.00	0.14	0.04	0.00	0.00	0.04	0.06	0.08	0.00
Dm	0.04	0.01	0.00	0.00	0.17	0.00	0.00	0.00	0.02	0.01	0.12	0.00	0.00	0.00	0.16	0.10	0.00	0.00	0.05	0.02	0.05	0.00	0.09	0.00	0.16	0.00
D#	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D#m	0.02	0.00	0.00	0.00	0.05	0.05	0.01	0.00	0.08	0.01	0.02	0.02	0.00	0.00	0.15	0.00	0.01	0.09	0.23	0.01	0.00	0.03	0.18	0.00	0.05	0.00
E	0.04	0.00	0.01	0.05	0.07	0.00	0.00	0.00	0.00	0.11	0.03	0.01	0.00	0.03	0.06	0.00	0.00	0.00	0.23	0.01	0.00	0.00	0.13	0.05	0.17	0.00
Em	0.21	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.09	0.00	0.04	0.00	0.00	0.01	0.17	0.00	0.00	0.14	0.17	0.02	0.00	0.02	0.01	0.01	0.13	0.00
F	0.19	0.02	0.00	0.00	0.04	0.11	0.00	0.00	0.02	0.07	0.00	0.13	0.01	0.02	0.08	0.01	0.00	0.02	0.02	0.04	0.03	0.02	0.07	0.00	0.11	0.00
Fm	0.06	0.03	0.05	0.00	0.09	0.01	0.02	0.00	0.03	0.23	0.04	0.00	0.00	0.00	0.08	0.01	0.04	0.08	0.08	0.05	0.00	0.02	0.08	0.00	0.09	0.00
F#	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
F#m	0.08	0.00	0.00	0.03	0.10	0.00	0.00	0.00	0.04	0.00	0.19	0.00	0.00	0.00	0.05	0.00	0.00	0.21	0.02	0.02	0.00	0.09	0.08	0.09	0.00	0.00
G	0.20	0.01	0.01	0.00	0.15	0.02	0.00	0.00	0.04	0.09	0.07	0.01	0.00	0.00	0.00	0.04	0.06	0.17	0.17	0.03	0.00	0.00	0.01	0.03	0.07	0.00
Gm	0.18	0.09	0.00	0.00	0.04	0.05	0.02	0.00	0.10	0.00	0.04	0.00	0.01	0.00	0.18	0.00	0.00	0.04	0.04	0.01	0.02	0.00	0.10	0.03	0.08	0.00
G#	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G#m	0.18	0.00	0.00	0.04	0.03	0.01	0.00	0.00	0.10	0.01	0.05	0.06	0.00	0.01	0.17	0.01	0.04	0.11	0.11	0.01	0.00	0.02	0.05	0.00	0.10	0.00
A	0.03	0.00	0.00	0.01	0.25	0.02	0.00	0.00	0.15	0.03	0.01	0.01	0.00	0.07	0.07	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.04	0.03	0.16	0.00
Am	0.13	0.00	0.00	0.00	0.12	0.05	0.00	0.00	0.06	0.09	0.19	0.00	0.00	0.01	0.07	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.01	0.01	0.08	0.00
Bb	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bbm	0.05	0.01	0.03	0.00	0.13	0.00	0.01	0.00	0.10	0.01	0.04	0.06	0.01	0.00	0.04	0.01	0.01	0.09	0.09	0.09	0.01	0.00	0.14	0.00	0.16	0.00
B	0.07	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.15	0.02	0.10	0.02	0.00	0.03	0.04	0.05	0.00	0.07	0.07	0.01	0.11	0.03	0.00	0.15	0.08	0.00
Bm	0.04	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.10	0.12	0.03	0.00	0.00	0.31	0.03	0.03	0.00	0.02	0.02	0.01	0.00	0.03	0.00	0.16	0.00	0.00
N	0.05	0.01	0.00	0.02	0.15	0.01	0.00	0.01	0.31	0.01	0.05	0.01	0.00	0.04	0.06	0.01	0.01	0.12	0.12	0.02	0.01	0.00	0.12	0.00	0.00	0.00

