

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Matevž Pesek

**Prepoznavanje akordov s hierarhičnim
kompozicionalnim modelom**

DIPLOMSKO DELO
NA UNIVERZITETNEM ŠTUDIJU

Mentor: doc. dr. Matija Marolt

Somentor: prof. dr. Aleš Leonardis

Ljubljana, 2012

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

Besedilo je oblikovano z urejevalnikom besedil $\text{\LaTeX}$.



Št. naloge: 01789/2011

Datum: 01.12.2011

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

Kandidat: **MATEVŽ PESEK**

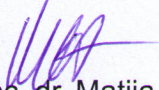
Naslov: **PREPOZNAVANJE AKORDOV S HIERARHIČNIM
KOMPOZICIONALNIM MODELOM
HIERARCHICAL COMPOSITIONAL MODEL FOR CHORD ESTIMATION**

Vrsta naloge: Diplomsko delo univerzitetnega študija

Tematika naloge:

V nalogi preučite kako bi lahko kompozicionalni hierarhični model, ki se uspešno uporablja za prepoznavanje vizualnih kategorij, uporabili na glasbeni domeni za prepoznavanje akordov v glasbenih posnetkih. Izdelajte postopek, ki bo uporabljal psihoakustični model zaznavanje glasbe za preslikavo glasbenega signala v značilnice na prvem nivoju hierarhičnega modela in ki bo zgradil večnivojsko hierarhijo, ki bo posnemala človeško dožemanje harmonije. Implementirajte tudi vizualizacijo dobljenih hierarhij in jih preizkusite na problemu prepoznavanja akordov na standardnih testnih množicah.

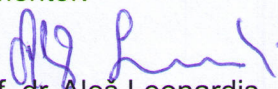
Mentor:


doc. dr. Matija Marolt

Dekan:


prof. dr. Nikolaj Zimic

Somentor:


prof. dr. Ales Leonardis



IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani/-a Matevž Pesek,

z vpisno številko 63070192,

sem avtor/-ica diplomskega dela z naslovom:

Glasbeni kompozicionalni hierarhični model

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal/-a samostojno pod mentorstvom
doc. dr. Matije Marolta
in somentorstvom
prof. dr. Aleša Leonardisa
- 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 15.1.2012

Podpis avtorja/-ice:

Zahvala

Rad bi se zahvalil tudi svoji družini za pomoč pri uspešnem zaključku študija. Hvala vam za moralno podporo pri polaganju izpitov in pisanju diplome. Brez vas mi ne bi uspelo.

Nato naj se zahvalim svojim mentorjema, doc. dr. Matiji Maroltu in prof. dr. Alešu Leonardisu. Skupaj sta že pred to diplomsko nalogo pričela z razvojem glasbenega kompozicionalnega hierarhičnega modela, v času pisanja diplomskega dela pa posvetila nešteto ur sestankov in razmišljanj ter kilobajtov elektronske pošte temu problemu.

Zahvalil bi se rad tudi članom Laboratorija za računalniško grafiko in multimedije, ki so mi pomagali z nasveti in znanjem v zadnjem letu študija: prof. Divjak, Marko, Alenka, Jernej, Ciril in Robert - hvala vam.

Hvala Ani in Vidi, ki sta me neutrudno poskušali naučiti $\text{\LaTeX}$.

Posebej še enkrat hvala Lei, ki mi je pomagala s slovničnimi popravki in predlogi.

Moji Družini

Kazalo

Povzetek	15
Abstract	17
1 Uvod	1
2 Pregled področja	5
2.1 Glasba in zgodovina notnega zapisa	5
2.2 Problemi relacije notnega zapisa in posnetka glasbe	7
2.2.1 Problem določljivosti tonske višine	7
2.2.2 Problem relativnosti časa v zapisu	7
2.2.3 Problematika percepcije harmonije	8
2.3 Pridobivanje informacij iz glasbe	10
2.3.1 Značilnice	11
2.3.2 Pristopi, ki temeljijo na pravilih	12
2.3.3 Statistični pristopi	16
2.3.4 Vizualni hierarhični kompozicionalni model	17
3 Glasbeni kompozicionalni hierarhični model	21
3.1 Uvod	21
3.1.1 Motivacija	21
3.1.2 Pomen kompozicionalnosti in hierarhičnosti	21
3.1.3 Delovanje modela	24
3.2 Priprava glasbenega posnetka	24
3.2.1 Konstantna Q transformacija	24
3.2.2 Psihoakustika	25
3.2.3 Implementacija psihoakustičnega filtra	27
3.3 Gradnja hierarhičnega kompozicionalnega modela na značilkah .	28
3.3.1 Ničti nivo hierarhije	28
3.3.2 Prvi nivo - učenje in izbira kompozicionalnih delov . . .	30

3.3.3	Aktivacija dela	30
3.3.4	Izbira kandidatov za naslednji nivo	31
3.3.5	Pokritje izbrane podmnožice	32
3.3.6	Inferenca posnetka s kompozicionalnim modelom	33
3.3.7	Inhibicija	33
3.4	Dodatne omejitve	34
3.4.1	Harmonična omejitev	34
3.4.2	Območja delov	34
3.4.3	Obtežitvena funkcija delnih odmikov	35
3.4.4	Omejevanje epsilon okolice združevanja parov	35
3.5	Višji nivoji - učenje in izbira	36
3.5.1	Doprinos združenega para	37
3.5.2	Izbira kandidatov	37
3.5.3	Nenadzorovano učenje pri gradnji modela	37
4	Interpretacija hierarhičnega kompozicionalnega modela	39
4.1	Model logistične regresije	41
4.2	Detekcija tonske višine s hierarhičnim kompozicionalnim modelom	41
5	Vizualizacija	43
5.1	Uporabljene rešitve	43
5.2	Izris hierarhičnega modela	44
5.3	Interakcija hierarhičnega modela in posnetka	46
6	Rezultati	49
6.1	Uporabljene množice avdio posnetkov	49
6.2	Testiranje stabilnosti hierarhičnega kompozicionalnega modela .	50
6.3	Testiranje modela na klavirskih akordih	51
6.4	Testiranje modela skladbah skupine <i>The Beatles</i>	54
6.5	Primerjava s sodobnimi rešitvami	58
7	Zaključek	59
7.1	Prednosti in omejitve modela	59
7.2	Nadaljnje delo	60
	Seznam slik	61
	Seznam tabel	65
	Literatura	67

Seznam uporabljenih kratic in simbolov

MIR - *Music Information Retrieval*

MIDI - *Musical Instrument Digital Interface*

TSR - *Time-Span Reduction*

PR - *Prolongational Reduction*

WPF - *Windows Presentation Foundation*

MP3 - *MPEG-1 Audio Layer III*

MIREX - *Music Information Retrieval Evaluation eXchange*

ISMIR - *The International Society for Music Information Retrieval*

Povzetek

V diplomski nalogi implementiramo nov pristop avtomatiziranega modeliranja in pridobivanja informacij iz glasbe. Model temelji na vizualnem kompozicionalnem hierarhičnem modelu, ki daje v praksi na svoji domeni izjemno dobre rezultate pri kategorizaciji objektov na slikah. Pristop prevedemo na glasbeno področje in ga uspešno implementiramo ter vizualiziramo. Pristop vsebuje uporabljeno znanje s področja psihoakustike za pripravo zvočnega posnetka in izračun značilnic. Izdelamo tudi vizualizacijo modela, ki služi kot pomoč pri interpretaciji interakcije modela z zvočnim posnetkom.

Model preizkusimo na množici štirih osnovnih vrst klavirskih akordov, zaigranih preko osmih oktav. Dosežemo 83% klasifikacijo v tipa dur in mol ter 78% klasifikacijo v tipe dur, mol, zvečan in zmanjšan akord. Model preizkusimo tudi na posnetkih glasbene skupine *The Beatles*. Uspešno klasificiramo 60% akordov v tipa dur in mol.

Prevedbo koncepta kompozicionalnega modela na glasbeno področje označujemo kot uspešnega in vrednega nadaljnjega razvoja.

Ključne besede:

kompozicionalni hierarhični model, klasifikacija akordov, pridobivanje informacij iz glasbe

Abstract

The thesis offers an implementation of a new approach for automatic modelling and music information retrieval. The model is based on a visual compositional hierarchical model, which provides good results in the area of categorization of visual objects in pictures. We translate the approach to the music domain by implementing and visualizing the model. The approach uses knowledge of psychoacoustic music modelling for the preparation of input audio clips.

The model is tested on a set of four basic piano chords, played through all eight octaves. We achieve 83% classification accuracy by classifying into major and minor chords and 78% classification accuracy by classifying into major, minor, diminished and augmented chords. The model is also tested on a set of songs composed by *The Beatles*. We successfully classify 60% of chords into major and minor chord classes.

We find the translation of the compositional hierarchical model to the music domain successful and worth of further development.

Key words: compositional hierarchical model, chord classification, music information retrieval

Poglavje 1

Uvod

Računalniška znanost se v zadnjih letih ne omejuje več le na strogo matematične probleme, kar je pripeljalo do tako imenovane informatizacije v celotnem družbenem procesu. V aplikativni znanosti najdemo najrazličnejša področja našega življenja, kjer se uporabljajo bolj ali manj kompleksni podprti sistemi, ki podpirajo in povečujejo storilnost, kapaciteto in hitrost obdelave najrazličnejših podatkov. Zato je na mestu pripomba, da mora biti inženir računalništva ekspert na dveh področjih: na svojem lastnem ter na področju, v katerega spada naloga, ki mu je bila zastavljena. V praksi se večinoma srečamo z različnimi informacijskimi sistemi, ki pogosto ne zadevajo akademskih ekspertnih področij. A se ravno na teh ekspertnih področjih odvija večji del raziskav v današnjem času - ne le iz vidika računske zahtevnosti in ostalih strogo računalniških ved, temveč tudi aplikativne uporabe najnovejših dovršenih pristopov, ki sprva služijo v podporo le ožjim ekspertnim skupinam ljudi, nato pa njihova uporaba, sicer počasi, preide v vsakdanjo rabo kot skrito gonilo splošno namenskih aplikacij.

Tudi na področju glasbe se je v zadnjem času odvilo nekaj takšnih preobrazb rešitev specifičnih problemov v osnovo za splošno namenske aplikacije. Kot primer omenimo znan program *Shazam* (<http://www.shazam.com>), s pomočjo katerega uporabniki mobilnih telefonov enostavno pridobijo informacijo o naslovu, avtorju in izvajalcu predvajane pesmi. Pridobitev informacije poteka s pomočjo mikrofona v telefonu, ki del pesmi zajame in preko značilnosti, ki to pesem definirajo, v bazi pesmi poišče pravo pesem. Uporabnik tako pridobi informacijo o naslovu pesmi in lahko slednjo uporabi pri nakupu pesmi v spletni trgovini. Ideja se je seveda rodila v strogo akademskih vodah, kjer so raziskovalci poskušali določiti nov pristop, ki bi pripomogel k enostavni klasifikaciji pesmi in iskanju podobnosti. Zaradi enostavnega pristopa so inženirji v

zabavni industriji idejo uporabili v aplikacijah na mobilnih telefonih in drugih spletnih aplikacijah.

Glasba je seveda mnogo več, kot le sistem enačb tonskih višin in relacij slednjih. A se tudi skladatelji in izvajalci mnogokrat znajdejo pred problemom iskanja notnih zapisov željenega posnetka in, ko iskanje ne prinese rezultatov, transkripcije glasbe. Tudi za nas je bilo to eno izmed vodil, na podlagi katerega smo poskušali ubrati nov pristop pri zaobseganju tega problema. Izgradnja modela, ki bi znal pojasniti relacije in podobnost, predvideti stil, opisati tonsko višino, harmonije in še mnogo drugega, za nas pomeni deterministično odkrivanje tistega, kar je v glasbi mnogim zakrito. Odpirajo se nova področja percepcije, ki lahko spodbudijo drugačne načine učenja glasbe, hkrati pa olajšajo delo - od transkripcije in opisa harmonij, pa vse do iskanja plagiatov.

Predstavljen model temelji na predpostavkah nekaterih nevrologov ([6]), da si človek sveta ne zapomni v razmerju ena proti ena - torej za vsako slišano melodijo ne obstaja direktna preslikava v nek spomin, vektor ali skupek nevronov - temveč si predstav o slišani melodiji gradimo hierarhično. Včasih je morda že nek zvok - žvižg, vzklik ali škripanje vrat - dovolj, da obudimo spomin o že slišani melodiji, ki se nam v trenutku odvrti v glavi. Po drugi strani pa si tudi bolj glasbeno nadarjeni med nami ob poslušanju svoje najljubše skladbe ne predstavljamo vsakega tona ali akorda posebej - a bomo v trenutku vedeli, da je nekaj narobe, če bo skladba le malo drugačna od tiste v našem spominu. Kako si torej zapomnimo toliko skladb, melodij, zvokov?

Naša hipoteza, ki bo v tej diplomski nalogi predstavljena, temelji na glasbenem hierarhičnem kompozicionalnem modelu. Vemo, da si vsak zvok in melodijo lahko predstavljamo v osnovi specifično definirano, hkrati pa zelo abstraktno, neodvisno od hitrosti izvajanja in tonske višine. S hierarhičnega kompozicionalnega modela bomo poskušali doseči približek modeliranja naše zavesti ob poslušanju glasbe. Predpostavljamo, da bomo na takšen način lažje modelirali glasbo in razbrali ključne informacije za natančnejšo opredelitev, ki bo podobna opredelitvi nas samih ob poslušanju glasbe.

Hierarhični kompozicionalni model glasbe je prevedba pristopa hierarhičnega kompozicionalnega modela za kategorizacijo slik ([7],[8]), ki se trenutno vedno bolj uveljavlja na področju vizualnega kognitivnega zaznavanja. Sama zamisel o hierarhičnem modelu se v muzikologiji pojavlja že dlje časa, uporabljajo pa se različni (predvsem) analitični pristopi, ki ne temeljijo toliko na determinizmu temveč na svobodnejši interpretaciji sistemov pravil oziroma preferenc (ohlapnejših pravil). Slednje bomo podrobneje spoznali v 2. poglavju. 3. poglavje se navezuje na izgradnjo modela od analize glasbenega posnetka

in izbire značilnic do same gradnje, dejavnikov vpliva in uporabljenih orodij. Predstavili bomo tudi morebitne druge interpretacije modela in vizualizacijo v 4. in 5. poglavju. Na koncu bomo v 6. poglavju predstavili rezultate testiranja modela in v 7. poglavju zaključili s prednostmi in omejitvami modela ter opisali možnosti za nadaljnji razvoj.

Poglavje 2

Pregled področja

Sprva se bomo v prvem podpoglavju spoznali s kratkim razvojem glasbe in osnovnimi gradniki notnega zapisa, ki so ključno sredstvo za predstavitev glasbe z izjemo posnetkov na medijih. Predstavili bomo problematike zapisa in izvedbe harmonij, s katerimi se srečujemo pri izvajanju zapisane in transkripciji posnete glasbe ter se nato osredotočili na trenutni razvoj analitičnih in drugih pristopov od začetka razvoja računalniško podprte analize, do trenutnih raziskovalnih smernic in rezultatov. Opisali bomo tudi že izdelane hierarhične modele za predstavitev glasbe. Poglavje bomo sklenili z opisom vizualnega hierarhičnega kompozicionalnega modela, ki je načrtal smernice našega pristopa na področju kategorizacije objektov v slikah, teoretični pristop kompozicionalnosti in hierarhičnosti pa se pomembno odraža tudi v naši implementaciji.

2.1 Glasba in zgodovina notnega zapisa

Zapis glasbe kot en izmed konstanto spremljajočih socioloških pojavov se po zgodovinskih virih pojavi v starem veku okoli 2000 pr. n. št. v obliki piktogramov (slikovna pisava). Vztrajen napredek privede v stari Grčiji do zapisa dolžine in višine tona (z določenimi osnovami harmonije). Šele v 7. stoletju se je pojavil zapis z nevmami, pri čemer se je nad besedilo zapisovalo poudarke (na nek način predhodnik ostrivca, strešice in krativca) in dolžine. Kmalu za tem se je pojavila bolj napredna oblika zapisa z nevmami, poznana predvsem po religioznih zapisih gregorjanskih koralov, ki se v zahodnem bogoslužju še uporablja.

Kasneje se je izoblikoval način zapisa dolžine not in višine, ki je bila normirana z označenim prvim tonom s črko. V sodobnem notnem zapisu sedaj to označujejo violinski, violski, altovski in basovski ključ. Iz črkovnih označb

izhaja tudi ime ključa - G ključ oziroma violinski ključ.

Za zapis poltonov na notnem črtovju sta v začetku služila dva znaka, iz katerih so se razvili zdaj imenovani višaj, nižaj in razvezaj/vračaj. Prav tako se je pričela razvijati tehnika zapisa za dinamiko in tempo (slednjega sedaj velikokrat nadomešča zapis za osnovno dobo in število udarcev osnovne dobe na minuto).

Z željo po zapisu glasbe je nastal notni zapis, ki objektivno povzema večino informacije v glasbi, del pa je še vedno prepuščen interpretaciji.

Oglejmo si osnovne glasbene gradnike notnega zapisa:

- nota - osnova zapisa je definirana kot simbol za zapis tonske višine in dolžine trajanja, slednje relativno glede na tempo in takt. Zapis tonske višine je v notnem črtovju relativen glede na ključ, ki označuje osnovno tonaliteto.
- Notno črtovje - baza notnega zapisa je definirana s pet-črtnim notnim črtovjem, ki služi za lažji zapis višin not. Note lahko pišemo bodisi na črte ali v prostor, ki ga omejujeta dve notni črti. Zapis lahko tudi poljubno (čeprav zaradi težjega branja ni zaželeno) raztegnemo izven notnega črtovja z dodajanjem krajših črt za vsako notno črto bodisi nad ali pod osnovnim notnim črtovjem. Te črte ponazarjajo dodatne vzporednice notnega črtovja in služijo lažjemu določanju tonske višine pri branju.



Slika 2.1: Sosledno zapisane note v območju dveh oktav. Note, ki so po odmiku nižje od območja notnega črtovja, vsebujejo dodatne črte, ki omogočajo lažje branje

- takt je v notnem zapisu označen na začetku zapisa s številom dob na takt, poševnico (zapis v obliki ulomka), in definicijo osnovne dobe.
- tempo - tempo je v sodobnih notnih zapisih definiran kot število osnovnih dob na minuto. Obstaja tudi bolj klasičen in starejši opisni zapis tempa, ki grobo definira hitrost izvajanja. Primer tega je *adagio* - počasi, *moderato* - zmerno, *allegro* - hitro, *veselo* in drugi. Občasno je opisnemu zapisu tempa

dodan tudi izrazni način, na primer: *veselo, funky, otožno*, ki se nanaša tudi na interpretacijo igranja notnega zapisa.

2.2 Problemi relacije notnega zapisa in posnetka glasbe

2.2.1 Problem določljivosti tonske višine

Tonska višina skozi posnetke ni vedno določljiva. Kot primer vzemimo petje melodije. Pri izgovorjavi besed, ki vsebujejo nezveneče nezvočnike za kratek čas tonska višina nedoločljiva. Prav tako se to pojavlja pri določenih tolkalih. Tonska višina ni popolnoma določljiva, lahko pa določamo relativno višino med posameznimi tolkali - percipiramo jih kot višje in nižje tonske višine, ne moremo pa jih povezati z nekim specifičnim tonom drugega inštrumenta. V notnem črtovju zapisujemo tonske višine s pomočjo not na različnih lokacijah v vertikalni smeri.

2.2.2 Problem relativnosti časa v zapisu

Čas je v notnem zapisu predstavljen kot množica sosledij relativnih dolžin tonskih višin glede na tempo. To privede do neenoličnosti zapisa iste melodije. Z manipulacijo tempa in takta lahko dosežemo na videz dugačen notni zapis, katerega izvedba se ne razlikuje. Če pri neki skladbi uporabimo na primer dvakrat daljše dobe, ki jih izvajamo dvakrat hitreje (v absolutnem času), bomo tako skladbo dojemali popolnoma enako kot osnovno. Zapis pa se bo bistveno razlikoval. Vseeno pri tem relativne razlike pri soslednih tonskih višinah ostajajo enake (slika 2.2).

Prav tako lahko z zapisom načina izvajanja manipuliramo zapis do te mere, da se neupoštevanje načina izvedbe izraža v popolnoma drugačni skladbi. Primer tega je način igranja v *swing*-u napram igranju v *straight* načinu. Pri prvem zapišemo dve zaporedni enako dolgi tonski višini (prvi takt na sliki 2.3), igramo pa jih tako, kot bi v *straight* načinu igranja igrali triolsko obliko z vezanima prvo in drugo ter tretjo komponento triole (drugi takt na sliki 2.3). Pri tem je prva komponenta podaljšana za dvokratnik dobe (vezaj med prvo in drugo noto šteje kot ena nota), nato pa ji sledi tretja komponenta. Štejemo triolsko, torej v časovnem intervalu dveh dob preštejemo tri dobe.



Slika 2.2: Primer iste melodije, zapisane v štiri četrtinskem taktovskem načinu (zgoraj) in nato v dvo polovinskem taktovskem načinu. Glede na podan tempo (zapis dolžine osnovne dobe na minuto) na začetku notnega črtovja, sta zaigrani melodiji identični.

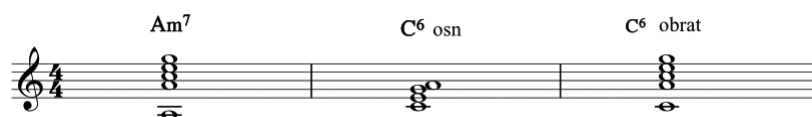


Slika 2.3: Prvi takt prikazuje *swing* zapis, ki se metrično izvaja kot je zapisano v drugem taktu v *straight* zapisu

2.2.3 Problematika percepcije harmonije

Obstaja tudi problematika pri zapisu določenih harmonij. Tipičen primer tega je zapis akorda C-E-G-A, kot je predstavljeno na 2.4. Slednje - če v spodnji legi inštrumenta, kot je klavir - nastopa ton C (kar lahko igra tudi drug inštrument, na primer kontrabas), imenujemo C6, oziroma C dur z dodano seksto - A. Na tem mestu je potrebna pripomba, da obstaja več obratov, inverzov akorda. Slednjega lahko igramo tudi kot tone A-C-E-G. To pa lahko interpretiramo tudi kot A mol z dodanim 7 tonom - G. Problem je torej tudi v sami interpretaciji. Večina teh primerov v je klasični in tudi deloma moderni glasbi sorazmeroma enostavno rešena s pomočjo pregleda celotnega vzorca z dodanimi glasbenimi pravili, ki določajo sorodne harmonije, ki nastopajo v specifičnem zapisu. Pri moderni glasbi (jazz) je problem toliko večji, kar nam izpostavlja že hiter pregled notnih zapisov, prosto dostopnih na spletu. Različne spletne strani z dostopnimi zapisi okvirnih harmonij skladb vsebujejo primerljive harmonije, zapisane v obliki drugega akorda. Občasno se pri transkripcijah glasbenih posnetkov pojavljajo tudi zapisi harmonij v oklepajih. To nakazuje, da lahko namesto ene harmonije zaigramo tudi drugo in je ta harmonija prejšnji na nek način ekvivalentna.

Drug primer nakazuje na človeško halucinacijo (domišljanje) prisotnosti



Slika 2.4: V prvem taktu nastopa akord A mol z dodano malo septimo. V drugem taktu je zapisan akord C dur z dodano seksto v osnovni legi, napram slednjemu pa je v tretjem akordu zapisan možen obrat akorda. Tretji takt odraža podobnost harmonije, opisane v prvem taktu, saj so vse tonske višine, z izjemo najnižje, enake. Takšen zapis akorda v tretjem taktu bi lahko interpretirali tudi kot A mol z dodano malo septimo in tonom C v basu oziroma krajše: Am7/C

tonskih višin v glasbi. V splošnem v audio posnetku določene tonske višine, ki se pojavljajo znotraj nekega časovnega intervala skupaj s preostalimi tonskimi višinami, določajo tip harmonije oz. akorda, ki je prisoten. Tako lahko poenostavljeno predstavimo dva tipa, ki ju uspešno razloči večina poslušalcev - dur in mol. Na sliki 2.5 sta predstavljena oba tipa v osnovni legi v C-dur tonaliteti.



Slika 2.5: Predstavitev C dur akorda (levo), C mol akorda (sredina) in dvoumnega akorda (desno).

Kot je iz slike 2.5 razvidno, se edina razlika nahaja v drugi noti (od spodaj navzgor), kjer je na eni strani zapisana tonska višina E, na drugi pa E♭ (es). Torej lahko dur in mol predstavimo v obliki intervala glede na osnovno tonsko višino v obliki velike (dur) oziroma male (mol) terce. Naslednji interval (med drugo in tretjo noto) je prejšnjemu obraten - če se med prvo in drugo noto nahaja interval velike terce, se med drugo in tretjo nahaja mala terca, in obratno pri molu. Na vrhu je označena pojavitev tonske višine C oktavo višje.

Problematika nastane pri odsotnosti tega tona v harmoniji. Vzemimo za primer le osnovni C dur akord, enak prej opisanima, ki mu odvzamemo noto E. Tako dobimo intervala kvinte in oktave, ki sama po sebi ne določata harmonije v pomenu dura in mola. Z vidika poslušalca pa je, glede na kontekst, vseeno možno razbrati, da slednji dojema interval kot akord in sicer - v kolikor posluša

skladbo, napisano v molu, bo zelo verjetno tudi interval kvinte dojemal kot mol. V kolikor je glasba napisana v duru, pa bo interval dojemal kot dur. Torej je kontekst za poslušalca (in ne za zapis) v določenih primerih edini faktor, ki opisuje njegovo dožemanje takšnih specifičnih primerov.

Tako neka bijektivna preslikava med glasbenim zapisom in dejansko izvedbo ali posnetkom ne obstaja. Že pri ročni analizi oz. transkripciji skladb pridemo do reda težavnosti, ki je v najboljšem primeru za večino večjih glasbenikov zamuden premo sorazmerno s težavnostjo skladbe.

Ne glede na to, da obstaja več interpretacij notnega zapisa in je percepcija glasbe nedvomno podvržena znanju in posledično subjektivni analizi posameznika, problem vsekakor ni nerazrešljiv vsaj do določene mere. Nasprotno se na tem področju že kaže napredek pri računalniškem zaznavanju in klasifikaciji glasbenih posnetkov v različnih oblikah.

Doprinos sistema, ki bi (do neke mere služil) kot pomoč pri klasifikaciji elementov v nekem glasbenem posnetku, je v praksi velik.

2.3 Pridobivanje informacij iz glasbe

Kot preplet muzikologije in računalništva se je razvilo področje pridobivanja informacij iz glasbe (MIR - Music Information Retrieval). Področje je glede na svojo kratko zgodovino precej obširno, kar se odraža v popolnoma različnih pristopih raziskovalcev. Predvsem gre za bistveno razliko med obdelavo in pridobivanjem informacij iz avdio posnetka in pridobivanjem informacij iz delno anotirane glasbe (MIDI zapis in drugi). Prav tako se pristopi razlikujejo tudi v ciljih, po katerih segajo.

Splošen pregled ([16],[15]) MIR sistemov, ki delujejo predvsem na primerjavi vhodnih (anotiranih ali avdio) posnetkov z neko bazo podatkov ali znanjem, pridobljenim iz baze, nakazuje na naslednje smernice:

- Povpraševanje z mrmranjem (Query by humming): Povpraševalec poskuša z mrmranjem imitirati glavno melodijo skladbe. Pridobiti želi informacijo o skladbi iz baze, ki vsebuje vzorec, ki je zamrmrani melodiji najbolj podoben.
- Primerjava notnih zapisov glasbe: Avtomatizacija medsebojnih vplivov različnih skladateljev je močan dejavnik tega področja. Slednji je zanimiv predvsem akademskim ekspertom s področja muzikologije, saj sta postopek in analiza vplivov s pomočjo prebiranja notnih zapisov dolgotrajna. Računalniško

podprt sistem sicer ni primerljiv z ekspertom do mere zamenljivosti, a v izjemno krajšem času nakaže osnovne vplive zunanjih socioloških dejavnikov na avtorja analizirane skladbe.

- Iskanje plagiatov: S pomočjo avtomatizacije iskanja plagiatov lahko po eni strani avtomatsko ocenjujemo pesmi kot avtentične ali ponarejene, hkrati pa se lahko tudi glasbeniki zavarujejo pred nezavednim prepisom in vključitvijo dela drugega avtorskega dela.

Glavni namen pristopov je torej pridobivanje različnih informacij iz glasbe. Prvi muzikološki pristopi so bili opisani že v zgodnjih 70ih letih, računalniško podprt razvoj tega področja pa se je pričel v zgodnjih 80ih letih in v današnjem času skupaj z razvojem tehnologije na glasbenem področju in posledično povečano doseglivostjo snemalne opreme doživlja razcvet. Ob povečani računski moči današnjih računalnikov se je tudi število statistično orientiranih pristopov povečalo in ravno slednji kažejo pomemben napredek pri klasifikaciji tipov harmonij, ki je tudi naše glavno področje dela v tej diplomski nalogi.

Sprva bomo predstavili osnovne značilnice, ki so v veliki meri temelj vseh sistemov. Ker v diplomski nalogi združujemo dva različna pristopa k modeliranju glasbe, bomo opisali trenutno uporabljene pristope, ki temeljijo na pravilih in gramatiki, in vnašajo hierarhični vidik posnetka. V nasprotju s slednjimi bomo predstavili tudi drugačne, statistične pristope, ki so trenutno najpogostejše uporabljeni za klasifikacijo tipov harmonij in detekcije tonskih višin. Na drugi strani pa želimo to klasifikacijo dobiti kot izhod iz hierarhičnega pristopa k modeliranju glasbe. Nato bomo predstavili modele, ki so trenutno v uporabi in uspešno ekstrahirajo informacije iz takšne ali drugačne oblike zvočnega posnetka (avdio signal, MIDI), nato pa bomo predstavili obstoječe hierarhične modele, uporabljene pri analizi glasbe, ki ne spadajo popolnoma v kategorijo klasifikacije eksplicitnih vsebin posnetka, temveč se večinoma bolj ukvarjajo z glasbo kot celoto. Ti pristopi so na tem mestu zanimivi zaradi vpeljave hierarhičnega modela v analizo glasbe.

2.3.1 Značilnice

Skupni temelj pristopov na tem raziskovalnem področju je razbiranje glasbe s pomočjo iz glasbe pridobljenih in izračunanih značilnic ([15]). Značilnice so definirane na več načinov, ki označujejo različne abstraktne elemente glasbe.

Tako v grobem ločimo med naslednjimi deli, ki jih zaobsegamo z značilnicami([16]):

- Tonska višina - za nek monofoničen vhodni posnetek poskušamo natančneje zaobseči tonsko višino predvajane melodije. Tonska višina je v nekem trenutku definirana kot percepcija osnovne frekvence neke igrane note, povezana s spremembo glede na prejšnjo igrano noto ([12]).
- Intenziteta - glede na energijo v signalu ugotavljamo glasnost predvajane melodije v nekem trenutku.
- Barva - Opis barve nekega signala v splošnem poslušalcu omogoča razlikovanje med različnimi izvori zvoka. Človek ob poslušanju orkestralne glasbe v nekem segmentu razloči godala od pihal, ne glede na isto tonsko višino in intenziteto. Slednje v splošnem težji problem od določanja prejšnjih dveh vrst značilnic.

V abstraktnejšem pogledu želimo določiti tudi kompleksnejše značilnice:

- Harmonija - Večji del glasbe predstavlja izvajanje več kot le enega instrumenta. V splošnem nastopa v posnetku več različnih tonskih višin, ki se v medsebojnih odmikih ujema in tvorijo harmonijo. Harmonije zavestno ob poslušanju glasbe ne dojemamo ekzaktno temveč nastopa v ozadju kot podpora vodilni melodiji.
- Ritem - ne glede na prisotnost ritmične baze (rhythm base - tolkala in percussion) nam predstavitev vsake melodije tvori neke enakomerno ponavljajoče abstraktne segmente. Percepcija ritma ni vezana na melodijo, barvo ali tonsko višino, temveč na hitrost spreminjanja melodije oziroma harmonije, ki nastopa v posnetku.

2.3.2 Pristopi, ki temeljijo na pravilih

V začetku osemdesetih let prejšnjega stoletja se je izoblikovala ideja o gramatičnem zapisu glasbe. Gramatičen zapis omogoča zajemanje preprostejših entitet in povezovanje v splošnejše vzorce, ki s poenostavitvijo in združevanjem nižjih nivojev, ki predstavljajo osnovne premike v harmoniji, na najvišjih povezovalnih nivojih zaobsegajo gradnike glasbe v obliki motivov in tem. Pristop

zaradi svoje gramatične narave temelji na opisovanju pravil, ki omogočajo gramatičen (in posledično hierarhičen) zapis. Na podobnem pristopu temelji tudi kasneje opisana Schenkerjeva analiza. Pripomniti pa je potrebno, da so takšni opisani pristopi iz vidika determinizma algoritmov zelo ohlapni, pravila pa prepuščena lastni interpretaciji.

Generative theory of tonal music

Gre za enega izmed osnovnih širše uporabljenih pristopov. Pristop temelji na delno subjektivni analizi, pri kateri uporabimo serijo hierarhičnih poenostavitev glasbenega posnetka kot osnovo za človekovo dožemanje glasbe. Vsak element se lahko povezuje v večje strukture in je hkrati sestavljen iz manjših struktur. Osnovni pristop je povezovanje glasbenih elementov. Časovno gledano se po tem pristopu uporabita dva principa poenostavitve:

- Poenostavitev osnovnih časovnih gradnikov (Time-span reduction) - S pomočjo ritmičnih in povezovalnih struktur poskušamo izluščiti informacijo. Na ta način pridobimo hierarhične organizacije časovnih intervalov, ki tvorijo drevo s pomočjo združevanja intervalov na vseh časovnih nivojih nekega zvočnega posnetka. Začenši pri najnižjih nivojih, sprva metrična struktura zaobsega razdelitev posnetka na ritmične enote enake dolžine (natančneje v "točke napada", ločene z enakomernimi časovnimi intervali), se analiza nadaljuje na višjih s pristopom razdeljevanja posnetka na motive, fraze, tematske in druge večje skupine.
- Poenostavitev psiholoških časovnih gradnikov (Prolongational reduction) - V tej fazi analiziramo psihološki učinek tenzij in sprostitve v vzorcih glasbe z natančno določenimi pogoji. Pri poenostavitvi osnovnih časovnih gradnikov je hierarhija vzpostavljena z bolj ali manj pomembnimi dogodki glede na ritmično stabilnost. pri Prolongational reduction daje hierarhija poudarek predvsem na relativno stabilnost skozi kontinuiteto in napredovanje (progression), premiki med tenzijami in sprostitvami. Uporaba PR analize je obrazložena skozi dve omejitvi TSR. Prva omejitev - TSR analiza ne omogoča izražanja občutka kontinuitete kot posledice harmoničnega ritma (harmonic rhythm). Druga - čeprav PR analiza pokaže, da so določene pojavitve tonskih višin zaznane glede na specifičen ritem (beat) znotraj določene skupine - ne zmore zaobseči informacije o tem, kako glasba napreduje skozi te segmente preko posameznega elementa.

- Pravila - Model je po definiciji predstavljen kot množica pravil, ki sooblikujejo različne korake analize in bolj ali manj restriktivno omejujejo obliko dreves, nastalih kot produkt analiz. Delimo jih na

- 1 Strukturna povezovalna pravila (Grouping Structure Rules) - Pravila oblikujejo pogoje katerih izpolnitev tvori odgovor o smiselnosti povezovanja elementov v skupino.

- 2 Preferenčna povezovalna pravila (Grouping Preference Rules) - Pravila govori o osnovnih vzorcih in o preferencah povezovanja opazovanih elementov.

- 3 Pravila o metrumskih strukturah - Pravila označujejo preference za označevanje takta v posnetku.

- 4 Pravila o poenostavitvi časovnih gradnikov - Pravila določajo osnovne elemente časovnih gradnikov, njihovo nastopanje znotraj skupine v času.

- 5 Pravila o poenostavitvi psiholoških časovnih gradnikov - Pravila določajo preference o pojavitvah psiholoških dejavnikov in položajev elementov znotraj teh časovnih skupin.

Problem pristopa nastopa predvsem v pomanjkanju determinizma pri določanju in upoštevanju pravil, ki jim modeli zadoščajo. Zaradi ohlapnosti pravil je problematična že računalniška predstavitev modela. Razvoj takšnih avtomatiziranih sistemov zahteva tudi veliko ekspertnega znanja. Čeprav je pristop med najstarejšimi na tem področju, se avtomatizirani sistemi razvijajo tudi v sedanjem času (5). Primer hierarhičnega združevanja je prikazan na sliki 2.6.

Schenkerjeva analiza (Schenkerian analysis)

V analizi glasbe se med drugim širše uporablja tudi Schenkerjeva analiza (Schenkerian analysis). Slednja temelji na interpretaciji glasbenega dela s pomočjo hierarhičnih razmerij med različnimi tonskimi višinami ter na tej podlagi potegne zaključke o strukturi glasbenega posnetka. Analiza torej temelji na iskanju vzorcev v glasbi na različnih nivojih in povezovanju vzorcev v večje skupine, kar nakazuje na hierarhični pristop. Vzorci se glede na nivo povezujejo v

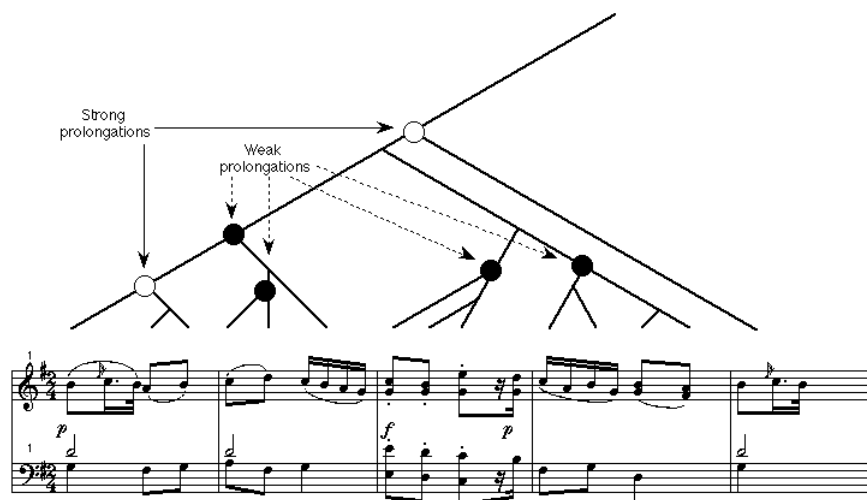


Figure 1(b)

Slika 2.6: Primer drevesa. (Vir: <http://www.mus.cam.ac.uk/~ic108/MusicAnalysis/>)

okviru ritma, melodičnega sosledja in na koncu tudi motivov. Takšna analiza je v praksi težko izvedljiva in je v večini primerov uporabljena s strani eksperta muzikologa brez računalniške podpore. Iskanje takšnih vzorcev ni (in ne more biti) deterministično. V zadnjem času se pojavljajo poskusi izgradnje avtomatiziranih modelov, ki vsebujejo tak pristop k analizi glasbe ([13], [9]). Veliko poskusov se orientira le na del analize, kot na primer iskanje variacij melodij ([13]). Glede na obsežnost analize so takšni specifično omejeni modeli popolnoma razumljivi, saj je takšen način analize računsko požrešen.

Problematika te metode je kot pri predhodno omenjeni v subjektivnosti pristopa, saj je interpretacija odvisna od analitika. Avtomatizacija in vpeljava računalniške podpore v analizo sicer bistveno pripomore k hitrosti, a je po drugi strani cena izračuna analize takšnega sistema previsoka. Po drugi strani se iz našega vidika problem skriva tudi v obliki vhodnih podatkov, saj se za takšno analizo uporablja predvsem MIDI zapis glasbe, med tem, ko je naš namen izvirati iz posnetka zaigrane glasbe.

2.3.3 Statistični pristopi

Statistični pristopi temeljijo na upoštevanju frekvenc pojavitev značilnic in njihovih sosledij. Predstavili bomo široko uporabljen skriti Markovski model. Implementacije skritega Markovskega modela v glasbi občasno vsebujejo tudi delno hierarhično podobo na podlagi združevanja več različnih hierarhičnih modelov, ki ločeno obdelujejo različne vrste glasbenih značilnic istega vhoda, v celoto.

Skriti Markovski model

Skriti Markovski model temelji na statističnem modelu, s pomočjo katerega poskušamo opisati vhod, pri čemer nismo seznanjeni z notranjim stanjem sistema. Pridevnik skriti se torej nanaša na notranja stanja, pri čemer je preostali del sistema jasno viden ([3]). Skozi proces učenja poskušamo pridobiti skriti Markovski model, za katerega smo izvedeli njegova stanja in prehode med njimi. skriti Markovski model se v praksi izkaže kot zelo učinkovito orodje za modeliranje glasbe skozi čas, kot tudi za samo statistično napovedovanje nadaljevanja prekinjenega glasbenega posnetka.

Skriti markovski model pogosto temeljijo na značilnicah, poimenovanih *kromagrami*. Kromagram je 12-dimenzionalni realni vektor, kjer vsak element vektorja prikazuje energijo posamezne tonske višine v posnetku ob nekem času t_n ([14]). Kromagrame pridobimo z *diskretno Fourierjevo transformacijo* audio signala. V frekvenčnem prostoru si za vsak ton zapomnimo vsoto magnitud v spektru. Vrhove v spektru preslikamo v območje ene oktave tako, da seštejemo vse tonske višine, ki predstavljajo isti ton. Tako dobimo informacijo o magnitudi pojavitev posameznih tonov v spektru. To se izkaže kot dobra predstavitev, saj omogoča relativno lahko klasifikacijo akordov [11].

Za analizo se običajno uporablja dva kromagrama, pri čemer en kromagram pokriva nižji frekvenčni interval (običajno med 50 Hz in nekaj 100 Hz), drugi pa preostali del območja od zgornje meje kromograma nižjega frekvenčnega intervala do zgornje meje zajetega frekvenčnega območja.

Z uporabo skritega Markovskega modela se raziskovalci trenutno uspešno spopadajo s prepoznavanjem akordov v zahodni moderni rock glasbi ([10], [14]), pri čemer se klasifikacija nanaša na prepoznavanje dveh do petih osnovnih tipov akordov. Viden napredek trenutno vsebuje klasifikacijo v več različnih razredov, kar je netrivialen problem, katerega del je opisan tudi v poglavju 2.2.3 Problematika percepcije harmonije.

2.3.4 Vizualni hierarhični kompozicionalni model

Na področju vizualizacije se že desetletja aktivno raziskuje klasifikacija oz. kategorizacija objektov. Slednje je bilo problematično v preteklosti predvsem iz vidika velikostnega razreda podatkov za obdelavo glede na razpoložljivo računalniško infrastrukturo, kar se je v zadnjih letih spremenilo. Ne glede na to je problem še vedno brez splošno uporabljene učinkovite rešitve, ki bi predstavljala *de facto* pristop k problemu. Z aplikativnega vidika je problem več kot le zanimiv, saj bi učinkovit pristop nadgradil zaznavanje v množici problemskih tematik od industrije (proizvodne linije) in izboljšanja standarda (varnost v prometu) do kognitivnih spoznavnih sistemov za domačo uporabo.

Predlagani pristop, ki ga uporabljamo prenesenega na drugo področje - glasbo, obeta zadovoljivo rešitev.

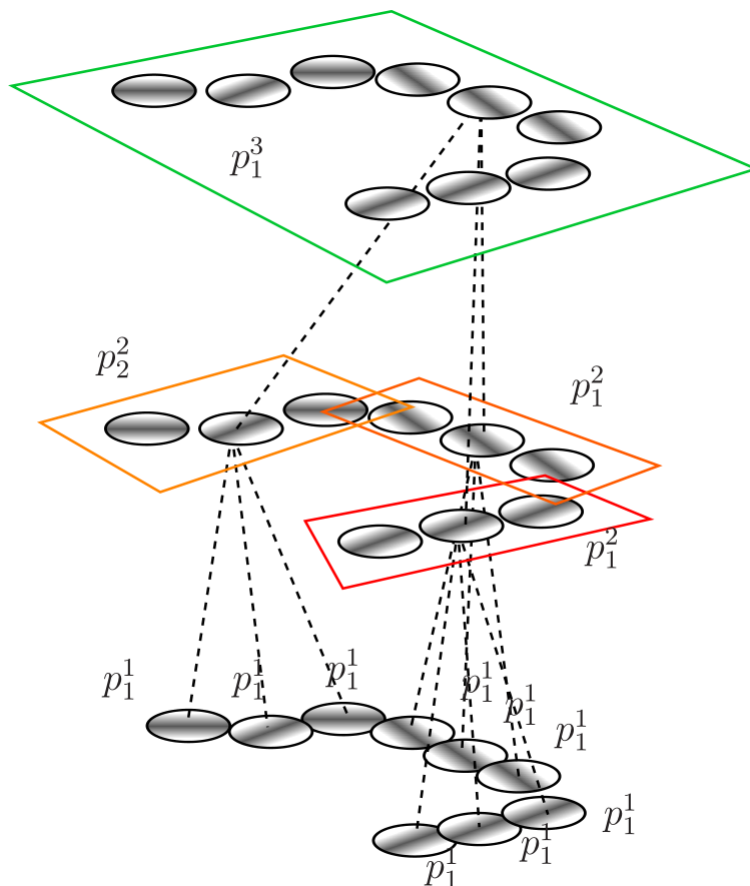
Osnovna predpostavka sistema izhaja tako iz mentalnega vidika človekove percepcije, kjer dojemamo stvari kategorizirano (jablana je drevo, drevesa si predstavljamo kot nekaj abstraktnega, a hkrati zelo specifičnega), kot tudi z vidika delovanja živčnega sistema, ki nakazuje na hierarhično delovanje na celičnem nivoju organizmov.

Problem direktnega pristopa kategorizacije tiči še vedno v velikostnem redu pomnjenja vseh možnih kombinacij predstavitve enega predmeta in semantičnega povezovanja v kategorije. Kombinatorično gledano je še vedno nemogoče zavzeti nek reprezentativen vzorec brez dodatnega učenja do te mere, da bi v splošnem dosegal neko zadovoljivo človeku primerljivo klasifikacijo. Glede na samo analizo slik pridemo do ugotovitve, da so pojavitve določenih segmentov slik skupne več različnim predmetom in celo kategorijam (štirinožne živali so med seboj zelo različne, a hkrati glede na trup zelo podobne).

S tem znanjem lahko predpostavimo, da je učenje in povezovanje objektov po gradnikih v prvi vrsti prostorsko manj zahtevno. Po drugi strani pa že direktno nakazuje na idejo hierarhičnega združevanja segmentov v višje, bolj specifične skupine, ki eventuelno z dodajanjem bolj specifičnih, statistično manj pomembnih segmentov določajo specifične objekte.

Hierarhija je torej sestavljena iz gradnikov, ki rekurzivno tvorijo kompozicije po nivojih le-te. Enostaven primer kompozicionalnega hierarhičnega modela je prikazan na sliki 2.7.

Ta hierarhičen pristop ni edini predstavljen na tem področju, bistveno pa se razlikuje zaradi same avtomatizacije kompozicioniranja gradnikov, ki temelji na statistični obdelavi vhodnih podatkov s pomočjo nenadzorovanega učenja po principu od spodaj - navzgor.



Slika 2.7: Prikaz vizualnega hierarhičnega modela, sestavljenega iz treh nivojev. Na prvem nivoju so vidni posamezni deli, ki predstavljajo ravne črte. Na drugem nivoju so opazne združitve v tri nove dele, tretji nivo pa vsebuje en sam del, ki pokriva celotno sliko. Vir: [4]

Učenje in gradnja hierarhije

Učenje hierarhije poteka na množici slik, ki so podlaga za izgradnjo prvega nivoja hierarhije. To je edini nivo, katerega rezultat se direktno navezuje na slike kot vhod pri operaciji učenja hierarhije [7].

Prvi nivo hierarhije se lahko izračuna s pomočjo množice filtrov, zaradi eksperimentalne potrditve hipoteze, da so prvi nivoji na različnih reprezentativnih učnih množicah stabilni, se v praksi zaradi znižanja računske kompleksnosti uporablja že naučen prvi nivo na katerem se gradi hierarhijo.

Naslednji nivoji se gradijo s pomočjo kompozicij delov na nižjih nivojih.

Povezovanje delov v nove dele na višjih nivojih je statistično upravičeno glede na učno množico.

Na višjih nivojih hierarhičnega modela se preneha z uporabo nenadzorovanega učenja. Raziskovalci so namreč prišli do ugotovitve, da se kompozicije na višjih nivojih preveč razlikujejo, da bi jih bilo možno dobro zaobseči s pomočjo statistike. Zato se uporablja nadzorovano učenje po kategorijah, ki v splošnem daje dobre rezultate, ki so primerljivi s trenutno najboljšimi vizualnimi spoznavnimi sistemi.

Poglavje 3

Glasbeni kompozicionalni hierarhični model

3.1 Uvod

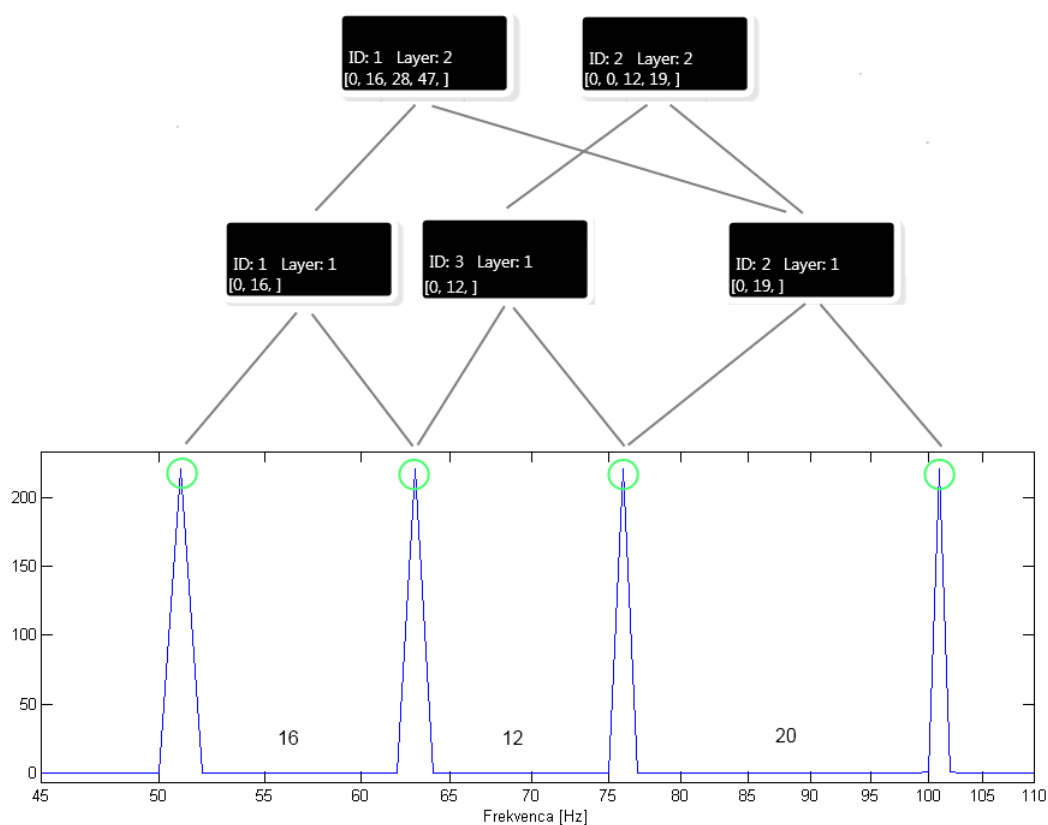
3.1.1 Motivacija

Po temeljitem pregledu področja smo prišli do ugotovitve, da problemu razbiiranja informacij iz glasbe še zdaleč ni zadoščeno v celoti. V želji po raziskovanju tega področja smo v kompozicionalnem hierarhičnem modelu videli dobro priložnost drugačnega načina modeliranja glasbe. V nadaljevanju si bomo ogledali pripravo osnovnih glasbenih značilnic posnetka, na katerih pričnemo z gradnjo hierarhičnega modela. Pri izboru značilnic smo upoštevali splošno uporabljeno znanje s področja psihoakustike ([5]). Nadaljevali bomo z izgradnjo hierarhičnega modela, z opisom dodatnih omejitev in razširitev modela, ki smo jih izoblikovali tekom razvoja pristopa. Podrobneje bomo opisali tudi nekatere ključne algoritme, uporabljene tako v vizualnem ([7] [8]) kot v glasbenem hierarhičnem modelu, skupaj z našimi implementacijami.

3.1.2 Pomen kompozicionalnosti in hierarhičnosti

Model, ki ga predstavljamo, vsebuje množico delov, ki so med seboj hierarhično razporejeni. Hierarhija je ureditev delov glede na lestvico kompleksnosti, pri čemer je lahko del *nad*, *pod* ali na *istem nivoju* glede drugega. Hierarhija nakazuje tudi na združevanje v kompleksnejše oziroma vsebinsko širše dele. Zaradi združevanja delov v kompleksnejše dele, je hierarhija tudi kompozicionalna, saj je vsak del na nekem nivoju (z izjemo najnižjega) *kompozicija*

delov na nižjem nivoju.



Slika 3.1: Enostavni kompozicionalni hierarhični model, sestavljen iz dveh nivojev. Spodaj se nahaja frekvenčni spekter, ki prikazuje ničti nivo hierarhije. Na ničtem nivoju nastopajo štirje absolutni deli, ki so označeni z zeleno barvo. Na prvem nivoju nastopajo trije deli, na drugem nivoju pa dva dela, ki sta sestavljena iz posameznih dveh delov prvega nivoja.

Model torej sestavlja manjše dele v večje in tako tvori hierarhijo. Na najvišjem nivoju hierarhije želimo, da vsak del pojasnjuje neko določeno vrsto dogodka (okvirja), na primer harmonijo določenega tipa. Deli na nižjih nivojih niso nujno značilni le za en del specifičnega tipa harmonije, temveč se pojavljajo na različnih mestih v različnih okvirjih. Ob pogledu *od spodaj navzgor* želimo na najnižjem nivoju takšne dele, ki so popolnoma neodvisni od kakršnih koli višjih struktur in tipov. Kompozicije teh delov pa so na višjih nivojih specifično

izoblikovane in predstavljajo različne tipe harmonij.

Na sliki 3.1 je predstavljen poenostavljen hierarhični model. Na prvem nivoju hierarhije so predstavljeni trije deli. Dva dela, ki sta narisana nad prvim nivojem predstavljata drugi nivo hierarhije. Puščice, ki kažejo iz delov prvega nivoja v dele na drugem nivoju, predstavljajo povezavo posameznih dveh gradnikov na prvem nivoju v *kompozicijo* - torej del na drugem nivoju. V prihodnje bomo označevali posamezen del z zaporednim indeksom dela (od leve proti desni), poševnico in zaporedno številko nivoja.

Prvi del na prvem nivoju $d_{1/1}$ predstavlja sopojavitev dveh frekvenc v signalu, ki sta med seboj odmaknjeni 16 frekvenčnih kanalov - torej štiri poltone. To je razvidno iz atributa *partial offsets*, ki nastopa kot zadnji v oknu posameznega dela. Drugi del $d_{2/1}$ tako predstavlja sopojavitev harmonikov z odmikom 19 frekvenčnih kanalov in tretji $d_{3/1}$ odmik 12 frekvenčnih kanalov. Ti deli so po svoji naravi tako enostavni, da nastopajo v večini tipov harmonij zahodne glasbe. Pozorni moramo biti na naravo višjih harmonikov posameznega tona. Zmotno bi lahko namreč pričakovali, da se prvi del na prvem nivoju vzbudi le, ko na vhodu zaigramo posnetek intervala velike terce. V takšnem primeru se ta del aktivira, a pride do aktivacije tudi pri igranju poljubnega klaviškega tona, saj lahko odmik predstavlja razliko med četrtim in petim višjim harmonikom. Oba sta namreč prisotna v zvoku, ki ga proizvede struna.

Takšni pari harmonikov se pojavljajo na različnih mestih v posnetku in niso zadostni za razbiranje informacije iz posnetka. Zato naredimo statističen pregled sopojavitev takšnih delov in zgradimo naslednji nivo hierarhičnega modela. Ugotovimo torej, kateri deli na prvem nivoju so se hkrati pojavili v posnetku največkrat in takšne dele poskusimo združiti. Pri združevanju smo seveda pozorni na oddaljenost posameznih dveh delov, ki se hkrati pojavita. Prvi del drugega nivoja $d_{1/2}$ na sliki 3.1 prikazuje primer, ko se je drugi del na prvem nivoju $d_{2/1}$ pojavil 28 frekvenčnih kanalov višje od dela $d_{1/1}$. Primer dela $d_{2/2}$ pa nakazuje, da sta se oba dela $d_{2/1}$ in $d_{3/1}$ pojavila na istem mestu, zato se število 0 v delnih odmikih dela $d_{2/2}$ pojavi dvakrat.

Oglejmo si še ostale attribute posameznega dela.

- Atribut *CP range* označuje zaporedno številko območja, v katerega del spada. Območja so definirana v poglavju 3.4.2.
- Atribut *FrameNum* označuje zaporedno številko trenutno predvajanega

okvirja.

- Atributi *ProbFrequ1*, *ProbFrequ2* in *ProbFrequ3* izpisujejo prve tri najmočnejše aktivacije (med 0 in 1) dela in točne lokacije teh aktivacij (v Hz).

3.1.3 Delovanje modela

Model kot vhod jemlje posnetek, ki ga razdeli na okvirje dolžine 50 milisekund. Modeliranje poteka na okvirjih, ki jih jemljemo kot medsebojno neodvisne. Vsak okvir torej modeliramo kot samostojno entiteto. Pri ugotavljanju frekvence sopojavitev harmonikov na ničtem nivoju ali delov na višjih nivojih hierarhičnega dela preko posnetka tako pregledujemo posamezne okvirje, posnetek pa je iz vidika modela množica okvirjev, pri čemer zaporednost elementov množice nima vpliva.

Pri vizualni animaciji aktivacij modela ob igranju posnetka spremljamo aktivacije posameznih okvirjev. Okvirji nosijo informacijo o natančni lokaciji pojavitve glede na čas posnetka. Tako se prvi okvir pojavi ob času $t_0 = 0s$, drugi okvir ob času $t_1 = 0.05s$, in tako dalje.

3.2 Priprava glasbenega posnetka

Kot vhod smo uporabili standardne WAV datoteke pri 44.1kHz frekvenci vzorčenja z 16 bitnim zapisom. Prebrani zvočni posnetek smo filtrirali s pomočjo psihoakustičnega filtra, ki simulira delovanje zunanjega in srednjega ušesa ([5], <http://is.rice.edu/~welsh/elec431/psychoAcoustic.html>). Tako zmanjšamo vpliv močnih harmonikov. Oglejmo si transformacijo posnetka v frekvenčni prostor kot predpripravo za uporabo psihoakustičnega filtra.

3.2.1 Konstantna Q transformacija

Na zvoku izvedemo Konstantno Q transformacijo, ki je sorodna *Fourierjevi* transformaciji, saj transformira iz časovne v frekvenčno domeno. Konstantna Q transformacija poda izhod v frekvenčnem prostoru kot amplituda v odvisnosti od logaritemske frekvenčne lestvice, kar daje transformaciji pomembno prednost z vidika smotrnega pokritja frekvenčnega intervala s frekvenčnimi kanali. Nižje frekvence imajo tako zadovoljivo frekvenčno ločljivost, med tem, ko se časovna ločljivost izboljša pri višjih frekvencah. S tem zadoščamo želji po posnemanju delovanja človeškega ušesa. V vhodnem zvoku upoštevamo

le frekvence med 55 Hz in 8000 Hz. Slednje razporedimo v 345 frekvenčnih kanalov, to je 48 frekvenčnih kanalov na oktavo. Zvok vzorčimo v 50 milisekundnih korakih z največjo širino okna 100 milisekund. Na posameznem okvirju, ki ga pridobimo s transformacijo, uporabimo psihoakustični filter. Sprva pa si oglejmo kratko definicijo psihoakustike.

3.2.2 Psihoakustika

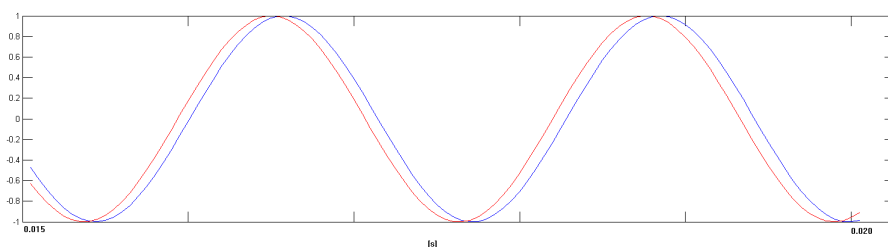
Psihoakustika se ukvarja tako s psihološkim, kot s fiziološkim vidikom človeške percepcije zvoka. Po eni strani na podlagi modeliranja ušesnega organa, po drugi strani pa s pomočjo serije psihoanalitičnih testov poskušamo doseči transformacijo predvajanega zvoka v zaznan zvok. Slednji se lahko bistveno razlikuje od zajema zvoka z avdio opremo. To dognanje sprejememo kot dejstvo, saj nam pomaga pri nadaljnji obdelavi zvoka in pridobivanju značilk. Sledijo opisi ključnih dejavnikov pri percepciji zvoka.

Zaznavanje tonske višine in razlike v tonskih višinah

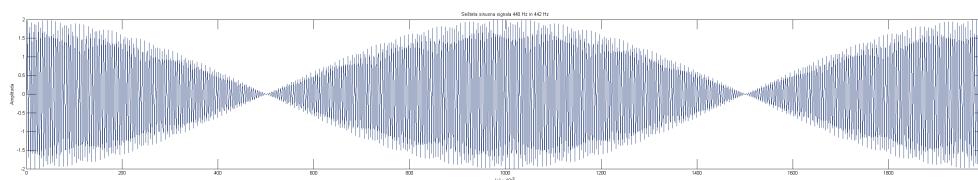
Človeško uho v splošnem zaznava zvoke med 20Hz in 20.000Hz. Znotraj tega območja obstaja ožje območje med 2 in 4 kHz, pri katerem je človek bolj dovzeten za zaznavo oziroma te frekvence sliši kot glasnejše, glede na ostale izven ožjega območja. Resolucija percepcije tonske višine dveh tonov je prav tako odvisna od lege frekvence znotraj intervala zaznavanja, hkrati pa je povezana tudi s pojavitvijo dveh tonov. V idealnih pogojih je človeško uho sposobno razlikovati med dvema bližnjima frekvencama (slika 3.2) na podlagi zaznavanja utripanja signala, ki je posledica vsote dveh bližnjih signalov (slika 3.3). Primer tega je lahko predstavljen tudi vizualno, ko gledanju dveh stroboskopskih luči. Svoje opažanje o frekvenci obračanja luči izpostavljammo na podlagi skupnih ojačitev in ošibitev svetlobnega toka v smer gledalca in ne s štetjem obratov vsake posamezne luči.

Maskiranje (masking)

Potrebno je omeniti tudi značilnost maskiranja. Pri izpostavljenosti neki tonski višini znotraj slušnega območja ter drugi, šibkejši tonski višini, ki je zelo blizu prej omenjene, opazovana oseba zaznava le močnejšo tonsko višino. To je posledica maskiranja tišje frekvence zaradi močnejše. Primer s sveta telekomunikacij nastopa pri brezžičnih omrežjih tipa g, kjer je na voljo 12 kanalov na frekvenci 2.4GHz z majhnimi odmiki. V praksi lahko uporabimo le 3 kanale (prvega, šestega in dvanajstega) tako, da izkoristimo maksimalno možno



Slika 3.2: rahel zamik sinusnih signalov pri frekvencah 440 Hz (označen z modro barvo) in 442 Hz (označen z rdečo barvo) na intervalu med 0,015s in 0,02s.



Slika 3.3: Prikazana sta šesteta signala iz slike 3.2 na intervalu med 0 in 1 sekundo. Na sliki so jasno vidne amplitudne ojačitve pri 0, 0,5 in 1 sekundi ter ošibitvi signala pri 0,25 in 0,75 sekundi.

pasovno širino signala. Sicer pride do t.i. motenj signala, ker je digitalni filter za sprejem podatkov na točno določeni frekvenci kanala premalo natančen, da bi lahko uspešno izoliral le posamezen kanal, hkrati pa lahko druga frekvenca zaradi moči povzroča interferenco.

Pri človeškem sluhu pride do podobnega problema. Ob pojavitvi močnega signala pri neki frekvenci, ki je nesorazmerno večja od ostalih prisotnih signalov pri drugih frekvencah, postane človek gluha za ostale signale, ki so najmočnejšemu signalu blizu in v tem območju dojema le najmočnejšega.

Halucinacija

Barvo zvoka določa nastop različnih harmonikov v frekvenčnem spektru. Človeška percepcija posameznih harmonikov zavestno večinoma ne zazna zaradi nizke magnitude, skupaj z osnovno tonsko višino pa se harmoniki odražajo v barvi zvoka. Tako lahko iz posnetka identificiramo dva poljubna različna inštrumenta, seveda glede na naše znanje o inštrumentih. Pri psihoanalitičnih poskusih pa so raziskovalci prišli do dognanja, da tudi če osnovne frekvence ni v signalu, lahko zaznamo ton na višini, ki je enaka višini manjkajoče frekvence.

V praksi je to jasno razvidno iz področja analogne telefonije, kjer se je zaradi potrebe po zmanjšanju prenešene količine podatkov pri telefonskem pogovoru tipično uporabljalo filtre, ki so omejili prenešen frekvenčni spekter med 300 Hz in 3000 Hz ([1]). Ne glede na to je lahko pogovor potekal nemoteno, poslušalec pa je lahko kakovost zvoka ocenjeval kot nizko, saj so pri pogovoru poleg osnovne tonske višine, ki je pri ljudeh tipično med 85 in 255 Hz - je pod spodnjo mejo zajetega frekvenčnega intervala - bili odrezani tudi višji harmoniki. Na podlagi tega je jasno razvidno, da lahko poslušalec razbere informacijo iz posnetka kljub odstranjenemu delu signala v enaki meri, kot bi to storil pri polnem signalu.

Uporaba psihoakustičnih filtrov

Psihoakustični filtri so pogosti v vsakdanji uporabi, najbolj znan primer uporabe je prisoten pri kompresiji zvoka v avdio predvajalnikih. MPEG-1 Audio Layer III (MP3) je splošno sprejet standard kodiranja glasbe z izgubami, ki s pomočjo psihoakustičnega filtra iz signala odstrani frekvence, ki jih človeško uho ne zaznava in s tem pripomore k manjši velikosti zapisa posnetka na trdem disku ali drugem mediju. Zapis se zmanjša na približno eno desetino velikosti originalnega posnetka (odvisno od kakovosti kodiranja), kar v praksi pomeni, da lahko na CD zgoščenko zapišemo približno 800 minut posnetkov (v primerjavi z 80 minutami nekompresiranih posnetkov).

3.2.3 Implementacija psihoakustičnega filtra

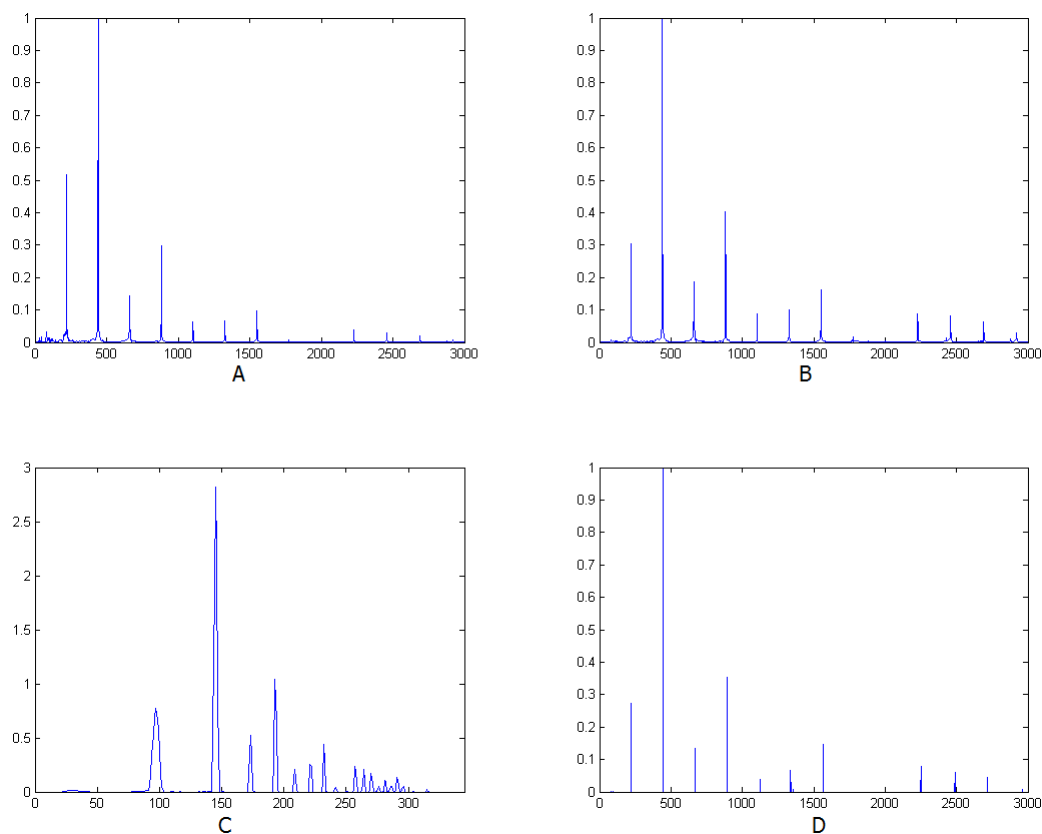
Posnetek torej želimo prilagoditi na način, kot ga naše uho dojema. Na sliki 3.4 je prikazan postopek filtriranja posnetka tona A1 (440 Hz), zaigranega na klavir. Sprva zvočni posnetek preberemo, kot je prikazano na podsliki A.

Nato na posameznem okvirju simuliramo človeško uho tako, da skladno z višino frekvence povečujemo magnitudo frekvence, kar je vidno v razliki višin vrhov med podslikama A in B pri frekveni 220 Hz, 880 Hz in 1760 Hz.

V tretjem koraku izračunamo pokritja posameznih prisotnih harmonikov v okvirju, kot je izrisano na sliki C okoli frekvenčnega kanala 96.

V četrtem koraku izvedemo maskiranje glede na funkcijo pokritja iz tretjega koraka. Maskiramo bližnje harmonike, prav tako pa zanemarimo dele signala, ki so več kot -35 dB tišji od najglasnejše frekvence v okvirju in globalno pod -70 dB (noise floor).

Na tem nivoju izvedemo lokalno inhibicijo harmonikov znotraj posameznega okvirja. Pri inhibiciji znotraj posameznega okvirja želimo doseči filtracijo



Slika 3.4: Slika prikazuje štiri korake obdelave psihoakustičnega filtra na posnetku tona A (440 Hz), zaigranega na klavirju. Pod slike A, B in D prikazujejo graf amplitude v odvisnosti od frekvence, pod slika C pa vsebuje graf maskiranja posameznih harmonikov glede na frekvenčne kanale.

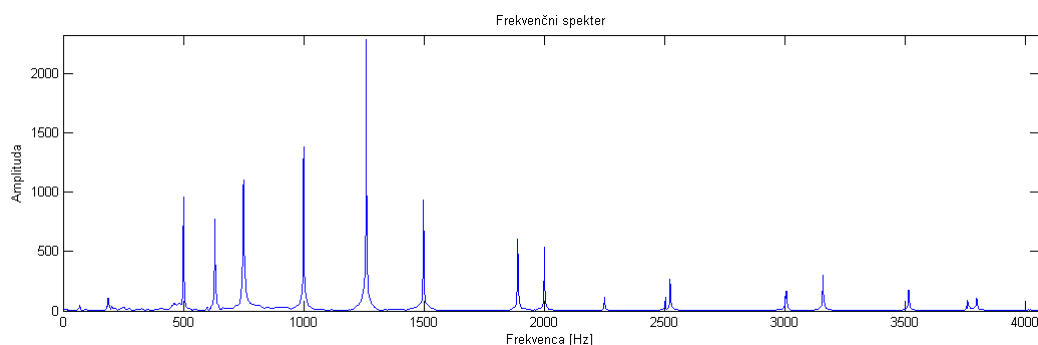
šibkejših harmonikov, ki jih po psihoakustičnih dognanjih ne glede na prisotnost v spektru zavestno ne dojemamo ob prisotnosti močnejšega harmonika.

3.3 Gradnja hierarhičnega kompozicionalnega modela na značilkah

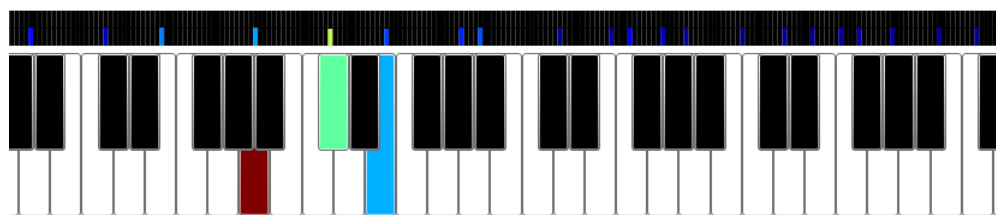
3.3.1 Nični nivo hierarhije

Hierarhični model obravnava vhodni posnetek v obliki okvirjev, ki so bili filtrirani s psihoakustičnim filtrom. Predstavljen izračun modela se procesira na

vsakih zaporednih 50 milisekundah vhodnega posnetka z največjo širino okna 100 milisekund. Osnovni gradniki hierarhije so frekvenčni kanali med 55 Hz in 8000 Hz. Ti gradniki predstavljajo absolutni nivo hierarhije - frekvenčni spekter ob določenem trenutku t_n . Te gradnike označujemo tudi kot ničti nivo hierarhije, saj predstavljajo posamezne osnovne dele, ki se pojavijo v glasbenem zapisu. Pri tem nivoju gre za reprezentacijo osnovnih značilk glasbe, iz katerih nato učimo hierarhični model. Na ničtem nivoju tako dobimo 345 delov, ki sestavljajo naše frekvenčno območje. Ta nivo je fiksni in v praksi predstavlja izhodišče za izgradnjo kompozicionalnega modela. Ta nivo je hkrati tudi edina povezava med glasbenim posnetkom in hierarhičnim modelom, saj so za gradnjo višjih nivojev potrebne izključno aktivacije nižjih nivojev. Primer izrisa frekvenčnega prostora je viden na sliki 3.5, aktivacija ničtega nivoja istega dela ob istem trenutku pa je prikazana na sliki 3.6.



Slika 3.5: Prikazan je frekvenčni spekter posnetka akorda A dur (A4, C#5, E5). Vrhovi predstavljajo potencialne aktivacije ničtega nivoja.



Slika 3.6: Prikazan je ničti nivo hierarhije ob predvajanju posnetka akorda A dur (A4, C#5, E5). Aktivni deli so obarvani v *Jet* barvni shemi. Neaktivirani deli so obarvani črno.

3.3.2 Prvi nivo - učenje in izbira kompozicionalnih delov

Učenje prvega nivoja hierarhije se prične z generiranjem možnih parov harmonikov (oziroma parov delov na ničtem nivoju), ki hkrati nastopajo v določenem izračunanem okvirju. Za posamezen okvir poteka iskanje parov na sledeč način:

- Na ničtem nivoju - predstavitvi frekvenčnega spektra po frekvenčnih kanalih - izberemo poljuben del, katerega aktivacija (v frekvenčnem spektru amplituda harmonika) je dovolj močna (višja od pragovne vrednosti).
- V kombinaciji s tem delom, ki ga imenujemo centralni del, povežemo pojavitev nekega drugega dela, ki nastopa višje v frekvenčnem spektru, *sekundarnega dela*, (harmonika) znotraj okvirja
- par shranimo v zbirko kandidatov za dele na tem nivoju. Število frekvenčnih kanalov med obema deloma, ki tvorita par, shranimo kot relativen odmik sekundarnega dela od centralnega dela, v kolikor takšna kombinacija istih dveh delov v tem razmiku še ne obstaja.

Ob tem je potrebno podrobneje pojasniti definicijo aktivacije določenega dela.

3.3.3 Aktivacija dela

Aktivacija nekega dela $d_{i/1}$ na prvem nivoju neposredno pogojena aktivacijo vsaj enega dela $d_{j/0}$ na ničtem nivoju, ki ta del na prvem nivoju sestavlja. To je potreben pogoj za aktivacijo dela $d_{i/1}$. V kolikor je ta pogoj izpolnjen, se v naslednjem koraku izračuna pokritje delnih odmkov dela $d_{i/1}$. Če je pokritje večje ali enako 50% (torej če se pokrije vsaj en harmonik na ničtem nivoju), se del aktivira glede na vsoto magnitud aktivacij na prejšnjem nivoju, znormirano s številom vseh možnih dvojic harmonikov, ki zadoščajo delnim odkom dela $d_{i/1}$ znotraj enega časovnega okvirja.

Posledično se lahko posamezen del na prvem nivoju aktivira v istem časovnem okvirju na večih mestih (frekvencah centralnega dela) znotraj spektra hkrati.

$$Act_{d_{i/1}} = \frac{Act_{d_{j/0}} + Act_{d_{k/0}}}{\sum (Diff_{d_{i/1}} == diff(d_{a/0}, d_{b/0}))}, \forall a, b : d_{a/0}, d_{b/0} \in Layer(0)$$

Slika 3.7: Formula za izračun magnitude aktivacije dela $d_{i/1}$. Aktivacija dela $d_{i/1}$ je vsota aktivacij delov, $d_{j/0}$ in $d_{k/0}$, deljena z vsoto števila vseh takšnih delov ničtega nivoja, katerih razlika v lokaciji pojavitve je enaka delnemu odmiku dela $d_{i/1}$.

3.3.4 Izbira kandidatov za naslednji nivo

Tako pridobimo zbirko vseh kandidatov za dele na prvem nivoju. Teh kandidatov je mnogo več (velikostni razred 100) kot je po končanem učenju delov, ki postanejo gradniki hierarhije (velikostni razred 10). Izbira med kandidati za prvi nivo hierarhije temelji na statistiki in pokritju značilnic v učni množici. Pomembna sta predvsem dva (delno izključujoča faktorja):

- pokritje učne množice naj bo čim večje in
- pokritja posameznih delov naj bodo medsebojno čim bolj disjunktna.

Pri čim večjem pokritju učne množice želimo doseči statistično sprejemljivo bazo delov, ki glede na pojavitve dosegajo visok delež parov aktivacij v posnetku. Pri slednjem pa smo pozorni, da ne dodajamo novih delov za vsako ceno: s pokrivanjem prenehamo, ko nam dodatno dodajanje delov na prvi nivo doprinaša manj pokritja, kot je določen minimum (prag 0,5% dodanega pokritja učne množice na dodan del). S tem se poskušamo izogniti dodajanju velike podmnožice delov, ki sicer pokrijejo različne robne primere v posnetku (ki so lahko tudi posledica šuma), ki imajo posledično nek doprinos (drugi deli ne pokrivajo tega dela učne množice), a lahko prekomerno prilagodijo model učni množici.

Pri medsebojni disjunktnosti pokritij kandidatov smo pozorni na prekrivanje pokritij delov, ki so že v izbrani množici, s preostalimi kandidati. Preostali kandidati, ki imajo lahko relativno veliko pokritje učne množice, a se slednje v veliki večini prekriva s pokritjem že izbranih kandidatov, modelu posledično ne doprinesejo dodatne informacije in so zato zavrženi.

Na podlagi zgoraj opisanih pogojev in vedenja o problematiki smo zgradili požrešen pristop za izbiro kandidatov:

```

while(pokritje(izbraniKandidati) < pragPokritja)
{
    //glede na ze izbrane kandidate izracunamo
    //relativne doprinose pokritij preostalih kandidatov
    pokritjaPreostalihKandidatov =
        izracunajPokritjaPreostalihKandidatov
        (izbraniKandidati, preostaliKandidati);

    //izberemo kandidata, ki relativno najvec
    //pokrije glede na ze izbrane kandidate
    izbraniKandidat =
        izberiOptimalnegaKandidata(pokritjaPreostalihKandidatov);

    //pazimo na preveliko prileganje
    if(pokritje(izbraniKandidat) < minimalniDoprinos)
    {
        //doprinos kandidatov je
        //manjsi od minimuma zakljucimo
        break;
    }
    //dodamo izbranega kandidata
    izbraniKandidati.Add(izbraniKandidat);

    //ponovimo postopek
}

```

Požrešnost se v algoritmu odraža v izbiri najboljšega kandidata - vedno bomo izbrali tistega, ki relativno pokrije največji del še ne pokrite učne množice. Ta pristop ni vedno najbolj optimalen, saj lahko obstaja določena podmnožica kandidatov, ki skupaj bolj optimalno pokrijejo učno množico, pri posamičnem dodajanju pa se to ne izrazi in se zato tvori drugačna, suboptimalna podmnožica.

3.3.5 Pokritje izbrane podmnožice

Pri požrešnem pristopu izbora kandidatov je potrebno prikazati še način računanja doprinosa (psevdokoda - `izracunajPokritjaPreostalihKandidatov()`). Za vsak posamezen že izbrani del (izbrani kandidati) moramo namreč pregledati vse aktivacije in med tistimi, ki jih je izbrani del že pokril, izločiti slednje, ko za novega kandidata računamo relativno pokritje učne množice. Na podlagi tega pridobimo ravno to "relativnost pokritja", ki nam omogoča učinkovit požrešen

izračun. Aktivacije, ki so že bile pokrite s strani že izbranih kandidatov, se ne upoštevajo pri izračunu pokritja še neizbranih kandidatov. Tako vsakega preostalega kandidata hipotetično priključimo že izbrani množici in pogledamo celotno pokritje hipotetične množice. Izberemo tistega kandidata, ki skupaj z množico izbranih kandidatov pokrije največ.

Pri pokritju smo uvedli tudi faktor bližine harmonika glede na centralni del. Tako v praksi v večji meri upoštevamo pokritje harmonikov, ki so blizu centralnega dela, v manjši meri pa pokritje harmonikov, ki so bolj oddaljeni. S tem smo želeli doseči pokritje slednjih, bolj oddaljenih harmonikov, z nekim drugim delom, ki bo pokril te harmonike s centralnim delom, ki se aktivira na enem od teh harmonikov, ostali pa so relativno bližji temu delu, kot preostalim. Definicija obtežitvene funkcije harmonikov je natančneje opisana v poglavju 3.4.3.

3.3.6 Inferenca posnetka s kompozicionalnim modelom

Kot smo za ničti nivo že omenili, je magnituda harmonika enaka aktivaciji dela na tem nivoju. Ker se na višjih nivojih deli ne navezujejo na posnetek, se na tem mestu poslužujemo izračuna aktivacij delov, glede na aktivacijo delov prejšnjega nivoja, ki sestavljajo ta del.

Posamezna aktivacija na novem nivoju je torej posledica aktivacije (vsaj enega) dela na prejšnjem nivoju, pri čemer je moč aktivacije izračunana iz moči aktivacij delov na prejšnjem nivoju in relativnega odmika med lokacijami aktivacij. Pri kasneje omenjeni dodatni omejitvi območij, smo pozorni tudi na lokacijo aktivacije glede na območje, kateremu posamezen del pripada. V kolikor bi lahko prišlo do morebitne aktivacije dela in je lokacija aktivacije izven območja, se takšna aktivacija ne upošteva.

Aktivacije na koncu tudi normaliziramo in tiste, ki so najšibkejše odstranimo, saj predstavljajo šum v vhodnih podatkih. Čeprav vpliv takšnih aktivacij statistično ni relevanten, takšne aktivacije pobrišemo, saj v nasprotnem primeru zasedejo dobršen del procesorske moči pri izračunu prvega nivoja.

3.3.7 Inhibicija

Inhibicija je postopek zmanjševanja vpliva podobnih delov na istem nivoju. Nek del $d_{i/\alpha}$ pokriva del frekvenčnega spektra v času t_n . V primeru, da obstaja nek drug del $d_{j/\alpha}$, katerega pokritje signala v času t_n je vsaj delno enako pokritju dela d_i , delu d_j del aktivacij, ki jih je pokril že del d_i ne upoštevamo.

Na tak način želimo z uporabo požrešnega pristopa odvzeti del moči aktivacij preostalim delom, ki so bili kasneje dodani v množico kandidatov.

V praksi želimo imeti čim bolj disjunktne dele. V primeru odštevanja aktivacij takšen proces vpliva le na tiste dele, ki pokrivajo nek delež istih aktivacij. Na druge dele, ki teh aktivacij ne pokrivajo, to direktno ne vpliva, kvečjemu se jim statistično gledano dvigne verjetnost, da bodo ti deli izbrani, saj smo prej omenjenim delom zmanjšali pokritje.

3.4 Dodatne omejitve

Z razvojem glasbenega kompozicionalnega modela smo uvedli nekatere dodatne omejitve povezovanja delov. Nekatere omejitve sovpadajo z omejitvami vizualnega kompozicionalnega modela, druge pa smo razvili sami z ozirom na glasbeno področje.

3.4.1 Harmonična omejitev

Pri obravnavanju posnetkov, v katerih je prisotna le ena tonska višina v vsakem trenutku, smo poskušali učenje hierarhije nižjih nivojev omejiti s harmonično omejitvijo pri povezovanju delov nižjega nivoja v nov del na naslednjem nivoju. Tako smo v hierarhičnem modelu ob vsaki sopojava tvitvi harmonikov preverili, če njihovi relativni delni odmiki predstavljajo harmonično vrsto. Sicer je bilo povezovanje neuspešno. Na višjih nivojih je takšna omejitev nesmiselna, saj harmonije ni mogoče modelirati zaradi pojavitve več vzporednih harmoničnih vrst, ki tvorijo akord.

3.4.2 Območja delov

Skozi proces gradnje kompozicionalnega hierarhičnega modela smo uvedli določene glasbi specifične izboljšave. Dejstvo je, da se znotraj našega opazovanega območja pri igranju nižjih tonov vzbudi veliko več harmonikov kot pri višjih tonih. S tem dognanjem smo pričeli naše opazovano območje opazovati na podlagi delitve območja v intervale.

Posamezen del na nekem nivoju je definiran s pripadnostjo podintervalu tako, da je njegov centralni del (oziroma na ničtem nivoju frekvenčni kanal) znotraj določenega podintervala. Prav tako se del aktivira le, če se kombinacija delnih odmikov, ki jih ta del predstavlja, pojavi na lokaciji znotraj podintervala, ki mu del pripada. Ta izboljšava po našem mnenju pri nenadzorovanemu učenju pripomore k različnemu obravnavanju harmonij v različnih legah. Tako

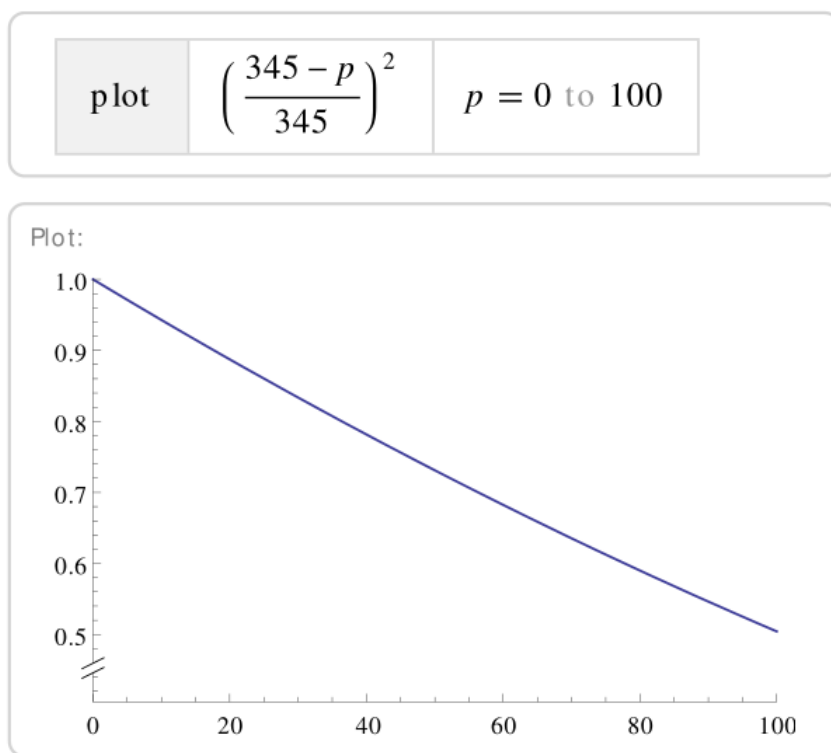
poskušamo usmeriti učenje hierarhije na višjih nivojih v izgradnjo delov, ki pretežno pokrivajo različne tipe harmonij v spodnjem nivoju območja in prav tako drugih delov, ki pokrivajo iste tipe harmonij v višjih nivojih. Načeloma nas dodatna količina delov ne moti, saj je obvladljiva z določitvijo števila podintervalov.

3.4.3 Obtežitvena funkcija delnih odmikov

V želji, da pridobimo dele, ki na višjih nivojih predstavljajo intervale in akorde, smo se poslužili funkcije obtežitve delnih odmikov. Funkcija modelira uteži, pri izboru primernosti novega para kot kandidata za nov nivo hierarhičnega modela. Obtežujemo vse delne odmike glede na njihovo oddaljenost od osnovnega odmika. Z oddaljenostjo se pomembnost odmika zmanjšuje. Tako želimo preprečiti velika pokritja zelo oddaljenih harmonikov, saj obstaja majhna verjetnost, da se ti harmoniki pojavljajo v harmoniji na takšen način, da je harmonija neposredno odvisna od teh harmonikov. Seveda se višji harmoniki pojavljajo na velikih odmikih že pri posameznem tonu, a se slednji pojavijo z relativno nižjo amplitudo. Funkcija je prikazana na sliki 3.8, ki izrisuje utež posameznega delnega odmika na intervalu med 0 in 100.

3.4.4 Omejevanje epsilon okolice združevanja parov

Kot paralelo epsilon okolice posameznih delov na različnih nivojih vizualnega hierarhičnega kompozicionalnega modela smo uvedli epsilon okolico delnih odmikov za posamezen del, glede na nivo, kateremu pripada. Tako se v nižjih nivojih omejimo na združevanje delov, ki so v frekvenčnem spektru relativno blizu, na višjih nivojih pa kombinacij ne omejujemo. Na takšen način želimo pridobiti gradnike, ki že v osnovi predstavljajo pokritje kombinacij več različnih harmoničnih serij (intervalov, akordov). Problem izhaja iz statistične pojavitve ene harmonične serije, ki modelira en sam ton. Takšen del lahko pokrije večino signala z aktivacijami na različnih lokacijah ob istem času, saj se glede na frekvenco aktivacije pojavljajo tako na posameznih tonih kot na sozvočjih. Takšen del je, glede na našo željo po klasifikaciji harmonij, nezanimiv, po drugi strani pa lahko dobro pojasnjuje definicijo tona, ki je zaigran s specifičnim inštrumentom. Takšni deli so torej po eni strani z vidika določanja tonskih višin zaželeni, z vidika klasifikacije harmonij pa želimo pridobiti tudi dele, ki so statistično manj pogosti, a pojasnjujejo korelacijo med tonskimi višinami.



Slika 3.8: Prikaz obtežitvene funkcije delnih odmikov glede na centralni del na intervalu med 0 in 100. Glede na frekvenčne kanale je odmik ene oktave razdalja 48 frekvenčnih kanalov. Na sliki je ordinatna os zaradi preglednosti primerno skrajšana.

3.5 Višji nivoji - učenje in izbira

Učenje višjih nivojev hierarhije poteka po enakem pristopu kot učenje prvega nivoja. Razlika v nivojih se odraža v delovnih podatkih učenja nivoja: na prvem nivoju delamo na samih značilnicah, kar je neposreden dostop do podatkov glasbenih posnetkov. Na višjih nivojih se pri izgradnji ne zanašamo več na značilnice, pridobljene iz glasbenih posnetkov, temveč na nižje že zgrajene nivoje hierarhije. Pri izračunu statistike značilnice glasbenih posnetkov še vedno uporabljamo, a se pri izgradnji parov ne zanašamo več na sopojavaivte frekvenčnih kanalov znotraj okvirja, temveč na sopojavaivte aktivacij delov na nižjem nivoju, ki posledično tvorijo kandidata višjega nivoja.

Generiranje množice kandidatov torej poteka s sprehodom čez vse aktivacije parov na nižjem nivoju in - glede na njihovo sopojavaitev - tvorbo novih

kandidatov, ki so sestavljeni iz parov tega nižjega nivoja.

3.5.1 Doprinosa združenega para

Pri izbiri parov, ki kasneje tvorijo potencialne kandidate za nove dele na višjem nivoju v izgradnji, smo pozorni tudi na doprinosa para kot skupka glede na število delnih odmkov, ki jih par pokriva. Običajno si obseg delnih odmkov sledi po harmonični vrsti 2 na prvem nivoju, 4 na drugem, 8 na tretjem in tako dalje. Ker pa pride do prekrivanja vsaj enega delnega odmika na združevanje, se ta številka zmanjša vsaj za ena na operacijo združitve. Pri tem smo bili pozorni, da ne pride do združitve delov, ki že sami po sebi predstavljajo večji del istih odmkov. V praksi nam torej preveliko ujemanje delnih odmkov ne ustreza saj slednje posledično prinaša nevarnost, da samo združevanje ne predstavlja neke diskriminacije dela oz. implicitne specializacije za določen tip harmonije, ki je nastopala v učni množici, temveč umetno napihovanje delov na nižjih nivojih, ki pokrivajo iste delne odmike.

Po drugi strani pa želimo pri izračunu doprinosa tudi obtežiti pokrivanje harmonikov, ki so bližje lokaciji centralnega dela, saj lahko večja količina višjih harmonikov nakazuje na nek doprinosa, ki v praksi pri človeškem poslušalcu ne predstavlja bistvenega vpliva.

3.5.2 Izbira kandidatov

Izbira med kandidati poteka s pomočjo istega požrešnega pristopa, opisanega v algoritmu 1. Pred izbiro izvedemo tudi postopek inference, ki generira aktivacije novonastalih kandidatov. To upoštevamo pri računanju doprinosa kandidatov.

3.5.3 Nenadzorovano učenje pri gradnji modela

Glede na vizualni kompozicionalni hierarhični model trenutno uporabljamo drugačen pristop h kategorizaciji. Osnovna implementacija vizualnega modela temelji na nekaj nižjih nivojih hierarhije, ki so skupni večini objektov in se zanje uporablja nenadzorovano učenje. Za višje nivoje hierarhije pa se smatra, da sama statistika ne zadošča za doseg cilja kategorizacije predmetov. Zato se za višje nivoje vizualnega hierarhičnega modela uporablja nadzorovano učenje po kategorijah.

V nasprotju s tem v našem primeru je gradnja glasbenega hierarhičnega

kompozicionalnega modela podvržena izključno učni množici in nenadzorovanemu učenju na podlagi statistike. Trenutno kompozicionalni model deluje na časovno neodvisnem modeliranju glasbe. Tudi število delov na posameznih nivojih glasbenega modela je za nekaj velikostnih razredov manjše glede na vizualni model. Na podlagi tega smo sklepali, da nadzorovana gradnja modela za potrebe klasifikacije tipov harmonij (slednji so le štirje napram nekaj sto kategorijam predmetov pri vizualnem modelu) ni potrebna, dovolj bo le enakomerno porazdeljena učna množica.

Kot je kasneje opisano, pri klasifikaciji akordov uporabljamo nadzorovano učenje v okviru modela logistične regresije na značilnicah, ki so izračunane s pomočjo nenadzorovano zgrajenega hierarhičnega modela.

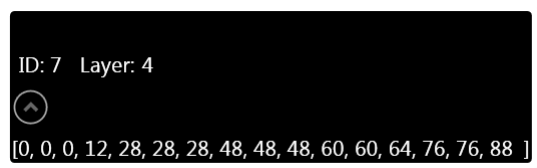
V primerjavi z vizualnim hierarhičnim modelom se na tem mestu razlikujemo, saj je za učinkovito klasifikacijo predmetov pri hierarhičnem modelu potrebna nadzorovana gradnja hierarhičnega modela.

Poglavje 4

Interpretacija hierarhičnega kompozicionalnega modela

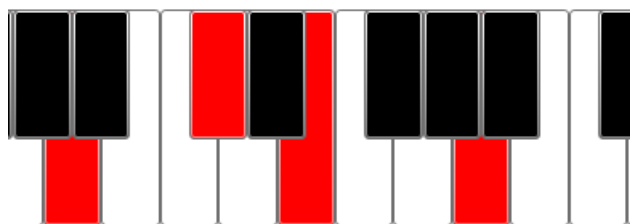
Hierarhični model želimo uporabiti za prepoznavanje harmonij v avdio posnetku. V ta namen smo se odločili z nadzorovano učenje modela logistične regresije na podlagi aktivacij zadnjih dveh nivojev modela. Zgrajeni hierarhični kompozicionalni model predstavlja na nižjih nivojih modeliranje osnovnih gradnikov avdio signala. S povezovanjem osnovnih gradnikov ob statistični podpori pojavitev gradnikov v učni množici poskušamo zaobseči abstraktnější pojem harmonije. Pričakovani rezultat višjih nivojev je torej množica delov na najvišjem nivoju, ki se z nenadzorovanim učenjem izoblikujejo v gradnike različnih tipov harmonij. Ob predpostavki, da nek del ali podmnožica delov pokriva vse harmonije enega tipa, pričakujemo nekaj delov za vsakega od štirih tipov harmonij, v katere želimo klasificirati. Tako nastane ob neki primerljivi frekvenci pojavitev različnih tipov harmonij v učni množici več delov, ki predstavljajo različne harmonije v učni množici. Za te dele sklepamo, da so (delno) disjunktni glede na tip harmonij. Glede na to, da v učni množici eksplicitno ne omejujemo tudi drugih kompleksnejših tipov harmonij, lahko pričakujemo tudi določene dele na višjem nivoju, ki pojasnjujejo drugačne tipe harmonij od tistih, v katere želimo klasificirati. Vseeno pa je takšna možnost relativno majhna, saj so ostali tipi harmonij v moderni (rock) glasbi precej manj zastopani ([2]).

Na sliki 4.1 je viden primer dela na četrtem nivoju, kjer delni odmiki nakazujejo na mol harmonijo. Lepo je vidna mala (molova) terca, kvinta, oktava (12, 28, 48) ter ponovitve istega vzorca oktavo višje (48, 60, 76). Odmika, ki nakazujeta veliko terco in malo septimo oktavo višje, sta verjetna posledica višjih harmonikov osnovnega tona.



Slika 4.1: Primer dela na četrtem nivoju. Delni odmiki nakazujejo na molov akord. Pri interpretaciji upoštevamo, da vsak polton modeliramo s štirimi frekvenčnimi kanali. Del tako predstavlja delne odmike treh, sedmih, dvanajs-tih poltonov glede na osnovni ton.

Primer dela, izrisanega na klavirski tipkovnici, je prikazan na sliki 4.2. Delni odmiki so izrisani v obliki obarvanih klavirskih tipk glede na tipko A.



Slika 4.2: Primer dela na drugem nivoju, ki združuje delne odmike v obliki klavirskih tipk.

Osredotočimo se na četrti (najvišji) nivo hierarhičnega modela. Na tem nivoju so deli natančno izoblikovani tako, da predstavljajo harmonije. Modela od tega nivoja dalje ne gradimo več, temveč se poglobimo v aktivacije delov na tem nivoju.

Sprva določimo novo učno in testno množico. Pri obeh množicah potrebujemo dodatno informacijo o anotaciji. Za obe množici izračunamo inferenco s hierarhičnim modelom tako, da za vsak posamezen okvir zapišemo vse magnitude aktivacij zadnjega nivoja hierarhičnega modela. Model nam torej služi za izračun zveznih atributov (ki predstavljajo dele na zadnjem nivoju) posameznega okvirja.

Od te točke dalje modela ne uporabimo več. Zapisa učne in testne množice sedaj obravnavamo kot matrična zapisa vektorjev atributov in razreda. Razred pridobimo iz anotacij posnetkov učne in testne množice. Učno množico uporabimo za izgradnjo modela logistične regresije.

4.1 Model logistične regresije

Za klasifikacijo akordov smo uporabili model logistične regresije pri čemer smo magnitude aktivacij delov na zadnjem nivoju uporabili kot značilnice okvirja. Značilnice smo modelirali z modelom logistične regresije v programskem okolju Weka 3 (<http://www.cs.waikato.ac.nz/ml/weka/>).

Model logistične regresije temelji na logistični funkciji, predstavljeni s formulo 4.3. Logistična funkcija je definirana na intervalu med 0 in 1, z zalogo vrednosti na intervalu $[-\infty, \infty]$. Funkcija je izrisana na sliki 4.4.

$$f(z) = \frac{1}{1 + e^{-z}}$$

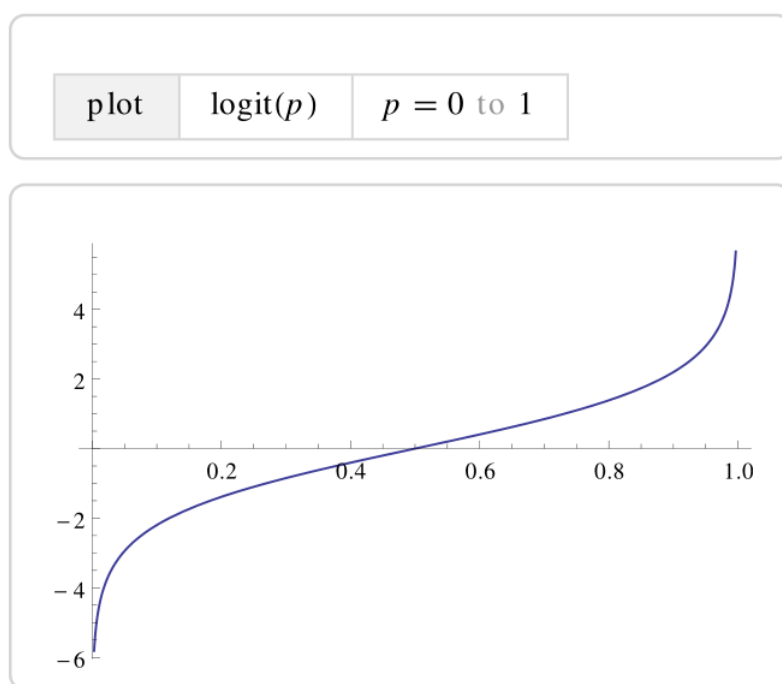
Slika 4.3: Formula logistične funkcije.

Spremenljivka z je definirana kot $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$, kjer $\beta_0 \dots \beta_k$ predstavljajo koeficiente posameznih atributov $x_1 \dots x_k$. Regresijski koeficienti torej prikazujejo vpliv posameznega atributa na ciljni razred.

4.2 Detekcija tonske višine s hierarhičnim kompozicionalnim modelom

Kompozicionalni model lahko uporabimo tudi za določanje tonske višine v signalu, ki vsebuje več tonskih višin (multi-pitch detection). Glede na aktivacije delov na višjih nivojih lahko z dodatnim algoritmom za detekcijo tonske višine sklepamo na osnovne tonske višine v signalu. Slednje ostaja zaenkrat še odprto področje našega modela, saj zahteva dodatno intenzivno poglobitev v gradnjo učinkovitega algoritma.

Implementirali smo osnovni algoritem za detekcijo tonskih višin, kar se odraža na prikazu tonskih višin na tipkovnici. Ker razvoj algoritma odpira popolnoma novo problematiko, smo nastavili osnovo za nadaljnji razvoj in te možnosti uporabe modela nismo dodatno testirali. Vsekakor pa tudi takšna uporaba izkazuje potencial, ki ga nameravamo v prihodnje izkoristiti.



Slika 4.4: Izris logistične funkcije na intervalu med 0 in 1.

Poglavje 5

Vizualizacija

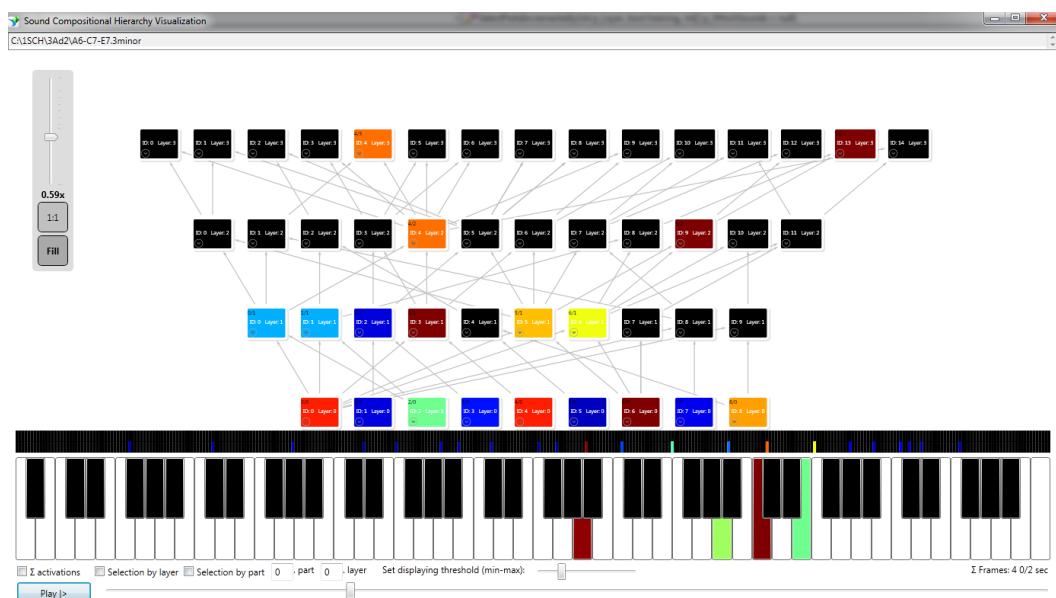
Za namen boljše predstavitve smo v sklopu diplomske naloge izdelali programsko orodje VisualSCH za interakcijo in vizualizacijo hierarhičnih modelov. VisualSCH vsebuje možnost vnosa vhodnih zbirk zvokov za izgradnjo hierarhičnega modela, nalaganje in shranjevanje že obstoječih hierarhičnih modelov ter vizualizacijo poljubnega hierarhičnega modela skupaj z realno časovnim izrisom interakcije poljubnega posnetka s hierarhijo. Motivacija za izdelavo programskega orodja temelji predvsem na težki predstavljenosti velike količine časovno odvisnih tekstovnih podatkov. Vizualen pregled aktivacij je veliko bolj intuitiven, hkrati pa nam omogoča lažjo detekcijo in odpravljanje programskih semantičnih napak. Zaradi vpeljave novega pristopa se ni bilo mogoče zanesti na že uveljavljene rešitve, kar je bil dodaten vzrok za izdelavo vizualizacije, saj smo lahko s pomočjo opazovanja obnašanja modela natančneje dopolnjevali naše vedenje o interakciji, prav tako pa je orodje služilo za empirično potrjevanje naših hipotez o odzivih hierarhičnega modela skozi čas tega pristopa.

5.1 Uporabljene rešitve

Vizualizacija je napisana v programskem jeziku Microsoft C# s podporo Windows Presentation Foundation (WPF). Pri izdelavi vizualizacije smo uporabili knjižnico Graph# (GraphSharp) (<http://graphsharp.codeplex.com>), ki omogoča enostavno vizualno manipulacijo grafov, kot je pri nas uporabljeno povečevanje in pomanjševanje pogleda in premiki vozlišč v grafu.

5.2 Izris hierarhičnega modela

Naložen hierarhičen model se prikaže v srednjem delu okna, prikazanega na sliki 5.1. Vozlišča, ki predstavljajo dele na posameznih nivojih, so po zaporednih nivojih naložena od nižjih nivojev proti višjim. Povezave med vozlišči označujejo povezovanje delov na nekem nivoju v del, ki nastopa na enem nivoju višje. Posamezen del je predstavljen z zaporedno številko (za lažjo identifikacijo), delnimi odmiki harmonikov, ki jih predstavlja, prvimi tremi frekvencami (lokacijami), na katerih se ta del aktivira in območje, v katerega centralni del tega dela spada.



Slika 5.1: Delovno okno vizualizacijskega programskega orodja, izdelanega kot pripomoček pri razvoju kompozicionalnega hierarhičnega modela.

Pod hierarhičnim grafom modela izrisujemo ničti absolutni nivo modela (slika 5.2). Ta nivo je predstavljen ločeno, brez povezav na prvi nivo modela zaradi jasnosti in berljivosti modela. Takšno povezovanje bi namreč doprineslo k velikemu številu povezav, ki zaradi količine delov na ničtem nivoju ne bi bile razberljive.

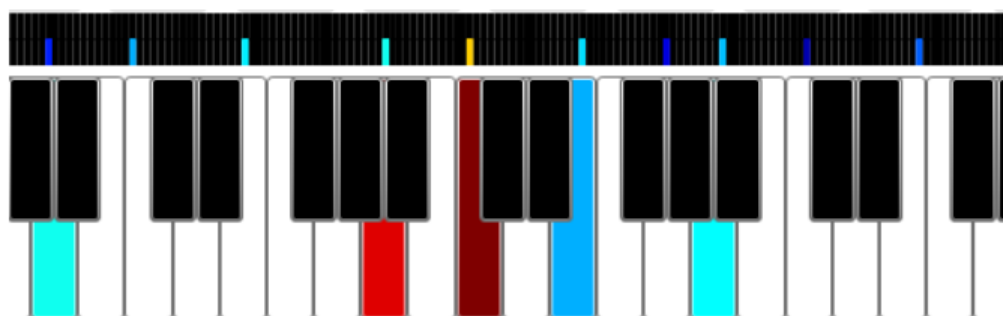
Ničti nivo je horizontalno razdeljen na dva dela: spodnji del prikazuje aktivacije delov na ničtem nivoju oziroma psihoakustično filtriran frekvenčni spekter predvajanega posnetka. Zgornji del ničtega nivoja pa je rezerviran za izris pokritij posameznih aktiviranih delov na poljubnem nivoju s projekcijo delnih

odmikov na ničti nivo.



Slika 5.2: Prikaz ničtega nivoja skupaj s pokritjem nekega dela na 3. nivoju. Za magnitudo aktivacije je uporabljena barvna shema *jet*, kot je definirana v programskem orodju *Matlab*. Zgoraj so z zeleno barvo označena pokritja izbranega dela v modelu.

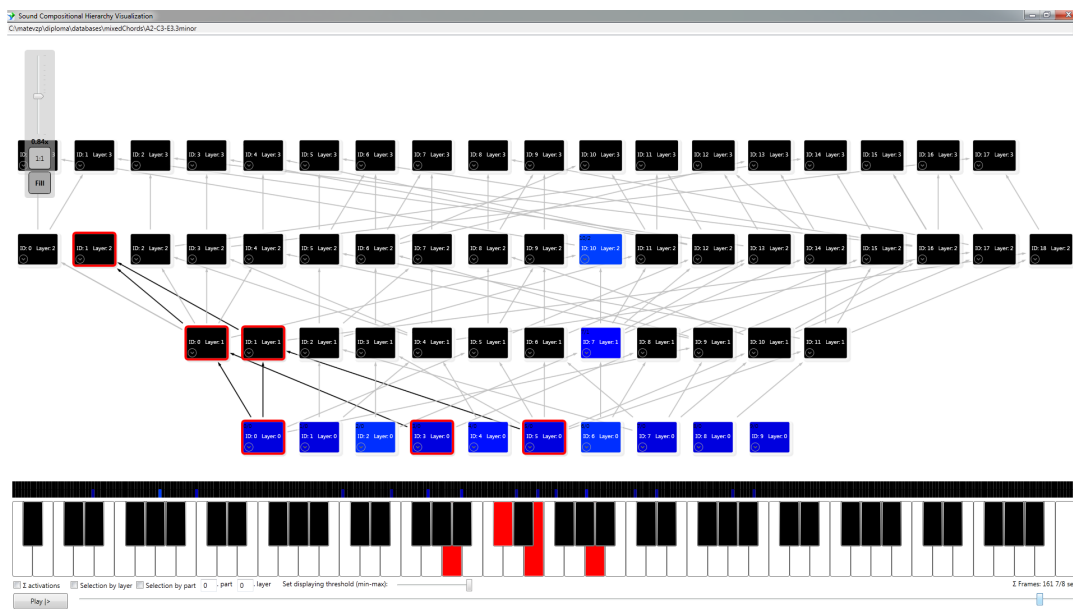
Pod ničtim nivojem modela je prikazan izris klavirske tipkovnice z 88 tipkami, kot je prikazano na sliki 5.3. Tipkovnica je delno prirejena, saj je premaknjena za dvanajst poltonov višje od standardne klavirske tipkovnice na inštrumentu. Najnižja tonska višina je torej A1 (55Hz), najvišja pa C9 (8000Hz). Na tipkovnici se izrisujejo predvidene osnovne frekvence aktiviranih delov v določenem predvajanjem časovnem okvirju. Ničti nivo hierarhičnega modela in tipkovnica sta vizualno poravnana, torej lahko frekvenčne kanale po potrebi primerjamo s klavirskimi tipkami.



Slika 5.3: Del tipkovnice ob predvajanju akorda A4 C5 E5. Na tem mestu pride do pojavitve tonske višine A3 in A5 (skrajna leva in desna obarvana tipka). Za magnitudo normiranega seštevek aktivacij je uporabljena barvna shema *jet*, kot je definirana v programskem orodju *Matlab*.

Izris klavirske tipkovnice nam lahko prav tako služi za boljšo predstavbo posameznega dela na nekem nivoju. V kolikor se slednji v določenem časovnem okvirju ne aktivira (saj bi slednjemu sledil izris frekvenc na absolutnem nivoju kar bi posledično omogočalo sklepanje na klavirske tipke glede na frekvenco),

lahko delne odmike določenega izbranega neaktivnega dela izrišemo tudi neposredno na klavirski tipkovnici, kot je prikazano na sliki 5.4.

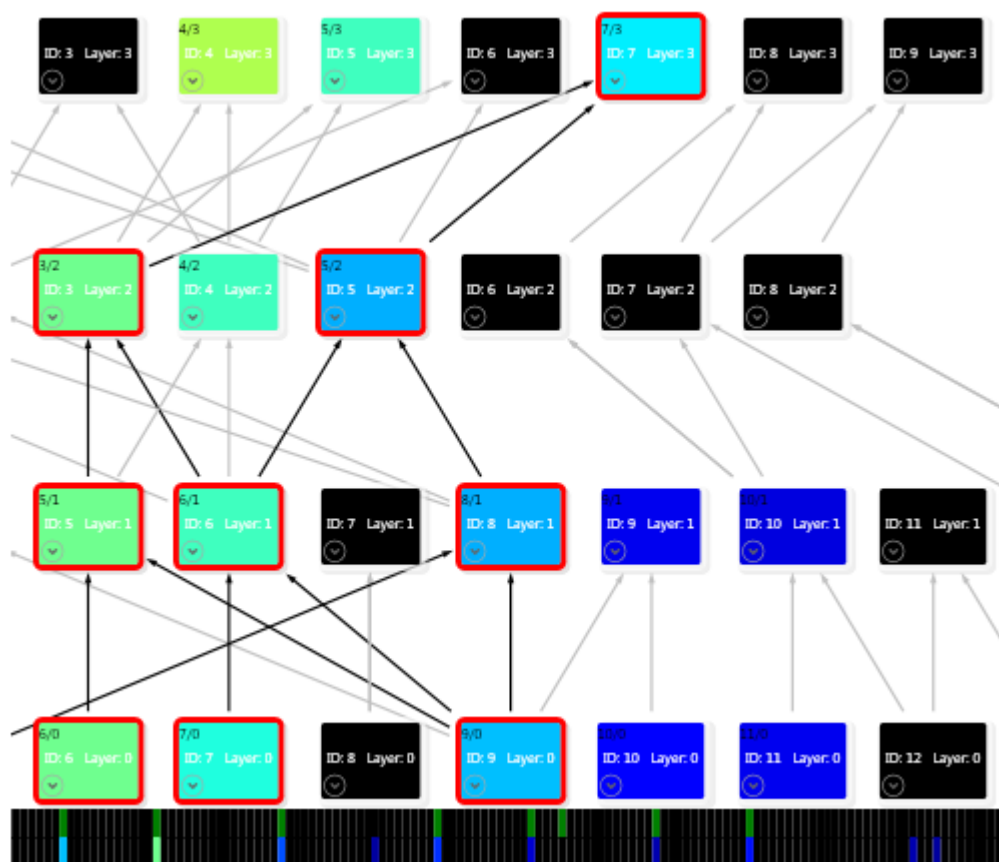


Slika 5.4: Izrisan neaktiviran del na klavirski tipkovnici. Delni odmiki so risani od A4 - 440Hz dalje za lažjo predstavbo tonskih višin, ki jih del s svojimi delnimi odmiki pokriva.

Za lažji pregled dela hierarhije in poddrevesa dela, ki nas zanima, lahko s pomočjo prehoda miške preko dela slednjega označimo. Rob označenega dela se obarva rdeče, prav tako pa se obarvajo vsi deli na nižjih nivojih, ki sestavljajo izbrani del. Povezave med izbranimi deli na nižjih nivojih se obarvajo črno. V primeru, ko ta del ni na najvišjem nivoju, se črno obarvajo tudi povezave, ki kažejo v dele na višjem nivoju. Uporabnik lahko tako enostavno razbere dele, ki izbrani del vsebujejo v svoji kompoziciji. Označevanje dela in poddrevesa, ki tvori označen del je predstavljeno na sliki 5.5.

5.3 Interakcija hierarhičnega modela in posnetka

V programsko okolje lahko uporabnik uvede nove vhodne posnetke in opazuje interakcijo hierarhičnega modela s posnetki. Naložen posnetek se ob pritisku



Slika 5.5: Primer delno označenega hierarhičnega modela glede na izbrani del (id 7 na četrtem nivoju). Na vsakem posameznem nivoju so označeni deli, ki so del kompozicije na višjem nivoju. Povezave med označenimi deli so obarvane črno.

na gumb prične predvajati. Uporabnik lahko ob poslušanju predvajanega posnetka opazuje aktivacije hierarhičnega modela na ekranu. V primeru, ko je določen del posnetka za uporabnika posebej zanimiv, lahko posnetek ustavi in se z drsnikom premakne na poljubno mesto v posnetku, pri čemer med drsenjem model prikazuje interakcije delov s posnetkom.

Poglavje 6

Rezultati

Kompozicionalni hierarhični model ponuja obilico možnosti za raziskovanje različnih vrst informacij skritih v glasbenem posnetku. Model je v splošnem uporaben tako za detekcijo tonskih višin v glasbi, kot tudi za določanje tipov harmonij. V rezultatih bomo predstavili reševanje problematike klasifikacije tipov harmonij. Prve problematike smo se pri izgradnji modela dotaknili s pripravo osnove, implementacijo algoritma za določanje tonskih višin pa prepuščamo nadaljnjemu razvoju modela v prihodnosti.

6.1 Uporabljene množice avdio posnetkov

V razvoju smo za izgradnjo kompozicionalnega hierarhičnega modela uporabili glasbo iz zbirke slovenskih ljudskih pesmi, koncertna dela *Johanna Sebastiana Bacha* ter diskografijo skupin *The Beatles* in *The Queen*. Za dela omenjenih glasbenih skupin smo se odločili zaradi dostopnosti natančnih anotacij pesmi ter primerljivosti z drugimi raziskovalci na tem področju. Na področju pridobivanja informacij iz glasbe se ta dela namreč uporabljajo kot *de facto* množici za predstavitev rezultatov na realnih podatkih. Poleg omenjenih zbirk anotiranih glasbenih del, obstajajo še druge zbirke, ki so dostopne na spletnih straneh Centra za digitalno glasbo univerze v Londonu (<http://isophonics.net/datasets>).

Ker je naš pristop šele na začetku razvoja, smo ga testirali tudi na lastni zbirki akordov, zaigranih na klavirju čez celotno območje osmih klavirskih oktav. Tudi ti akordi so anotirani po tipu in tonskih višinah, prisotnih v posnetku. Uporabili smo štiri osnovne tipe trozvokov - dur, mol, zvečan in zmanjšan tip akorda.

6.2 Testiranje stabilnosti hierarhičnega kompozicionalnega modela

Sprva se bomo osredotočili na primerjanje strukture hierarhičnih modelov, zgrajenih na različnih po vsebini primerljivih učnih množicah. Naša domneva je, da so prvi nivoji različnih modelov relativno podobni. Ta domneva izhaja iz predpostavke, da naš um deluje na istem hierarhičnem pristopu. Predpostavko lahko v praksi pokažemo s primerom dveh poslušalcev. Poslušalca, ne glede na njuno glasbeno predznanje, v enaki meri zaznavata osnovne glasbene elemente. Razlike v percepciji nastanejo pri kompleksnejših elementih, kjer je predznanje poslušalca pomemben faktor. Kompleksnejši elementi, ki so kompozicije enostavnejših, so v našem modelu predstavljeni v višjih nivojih.

Sledi primerjava prvega nivoja hierarhičnih modelov. Za vsako učno množico smo zgradili dva modela - model z vključeno omejitvijo območja združevanja, opisano v poglavju 3.4.4 in model brez omejitve.

Za izgradnjo hierarhičnega kompozicionalnega modela smo izbrali pet učnih množic, ki so vsebovale enako količino preigravanj lestvic različnih instrumentov (saksofon, flavta, violončelo), dela iz zbirke *Slovenske Ljudske Pesmi* z enim pevcem ter dve glasbeni deli *Bach*-a. Množice načeloma ne omejujemo po velikosti, želimo pa čim bolj raznoliko kompleksnost harmonij v glasbenih posnetkih. V primeru monofonične glasbe lahko pričakujemo, da bo kompozicionalni model vseboval le dele, ki pokrivajo posamezne tonske višine in višje harmonike teh tonskih višin. Za primerjavo smo omejili velikost vsake množice na približno 40 minut. Pokazati smo želeli, da je prvi nivo teh hierarhičnih modelov podoben, podobnost pa pada s kompleksnostjo delov, ki so praviloma bolj specifični.

Tabela 6.1: Predstavitev podobnosti nižjih nivojev glede na omejitve združevanja znotraj omejenega območja in brez omejitve združevanja.

povprečje podobnosti	1. nivo	2. nivo	3. nivo	4. nivo
brez omejitve območja	85%	37%	8%	0%
z omejitvijo območja	57%	23%	1%	0%

Tabela 6.1 prikazuje povprečno podobnost med posameznimi nivoji petih testnih množic. Pričakovano je podobnost brez omejevanja območja združevanja

večja, saj takšni hierarhični modeli vsebujejo več delov, ki sestavljajo harmonične vrste posameznih tonov. Na drugi strani hierarhični modeli z omejitvijo območja združevanja vsebujejo več delov, ki na nižjih nivojih modelirajo dele harmonij.

S tem testom smo želeli pokazati v kolikšni meri je model odvisen od vhodne množice. Pri tem testu nismo upoštevali, da lahko nek del z malo drugačnim pokritjem signala pokriva isti tip harmonije ali del istega tipa harmonije, temveč smo gledali primerjavo po delnih odmikih. V kolikor se je del razlikoval v le enem delnem odmiku, smo ga obravnavali kot različen del. Model je vsekakor delno odvisen od vhodne množice, saj učenje temelji na pojavitvi določene značilnosti glasbe v učni množici. Vseeno pa pri neki normalni porazdelitvi glasbe v učni množici velike razlike med modeli niso pričakovane.

Velika podobnost nižjih nivojev se sklada z našo hipotezo, da so osnovni gradniki kompozicionalnega modela enaki ne glede na učno množico. Če potegnemo vzporednico s percepcijo novorojenčka: otrok bo percipiral osnovne gradnike harmonije na enak način tako v zahodni kot vzhodni kulturi, ne glede na kraj rojstva. Lažje pa bo kasneje dojemal kompleksnejšo glasbo svoje domače kulture, dočim se bo tuje glasbe moral priučiti (torej ustvariti nove gradnike na višjih nivojih, ki bodo pojasnjevali nove harmonije).

Stabilnost nižjih nivojev je bila opažena tudi pri vizualnem kompozicionalnem modelu, zato smatramo, da je naša implementacija skladna, prav tako pa je potrjena naša hipoteza.

6.3 Testiranje modela na klavirskih akordih

S pomočjo zgrajenega hierarhičnega kompozicionalnega modela smo izračunali aktivacije delov na tretjem in četrtem (zadnjem in predzadnjem) nivoju zbirke štirih tipov klavirskih akordov v osnovnih legah, zaigranih čez osem klavirskih oktav.

Za klasifikacijo smo uporabili model logistične regresije, kot je opisano v poglavju 4.1. Najprej smo izvedli učenje modela logistične regresije na 50% podatkov in poskusili klasificirati preostalih 50% množice. Izvedli smo dva ločena testa: v prvem testu smo poskusili klasificirati le dva najpogostejše prisotna tipa akordov - dur in mol. V drugem testu pa smo klasificirali v vse štiri tipe akordov.

V vsakem primeru so bili klasifikacijski razredi porazdeljeni enakomerno, torej je vsak razred predstavljal $\frac{1}{n}$ učne množice, kjer je n število razredov. Rezultati so predstavljeni v tabeli 6.2.

Tabela 6.2: Testiranje klasifikacije na razdeljeni učni množici.

tip klasifikacije	točnost v procentih
klasifikacija v dur/mol	83,1
klasifikacija v vse 4 tipe	78,1

Potrebna je natančnejša obrazložitev rezultatov. Klasifikacija v osnovna tipa dur in mol daje dobre rezultate. Slabši rezultati se pokažejo pri klasifikaciji v vse štiri tipe akordov, pregled matrike klasifikacij (confusion matrix) kaže na tri probleme: problem napačne klasifikacije dur akordov v mol akorde, problem navzkrižne napačne klasifikacije mol in zmanjšanih akordov in problem napačne klasifikacije zvečanih akordov v dur akorde (tabela 6.3).

Tabela 6.3: Matrika klasifikacij (confusion matrix) pri klasificiranju v vse štiri razrede.

maj	min	aug	dim	
63,9%	25%	5,55%	5,55%	maj
9,72%	77,8%	1,38%	11,1%	min
13,9%	2,8%	81,9%	1,4%	aug
1,38%	9,72%	0%	88,9%	dim

Iz tabele je razvidno, da se največji zgrešeni del zvečanih akordov klasificira v dur akorde. Pri zmanjšanih akordih se največji del zgrešenih akordov klasificira v mol akorde. Sicer to smatramo kot napako sistema, a po drugi strani vemo, da sta si dur in zvečan akord bolj sorodna kot mol in zmanjšan ter obratno. Napaka v tem pogledu ni tako moteča, a vendar prisotna.

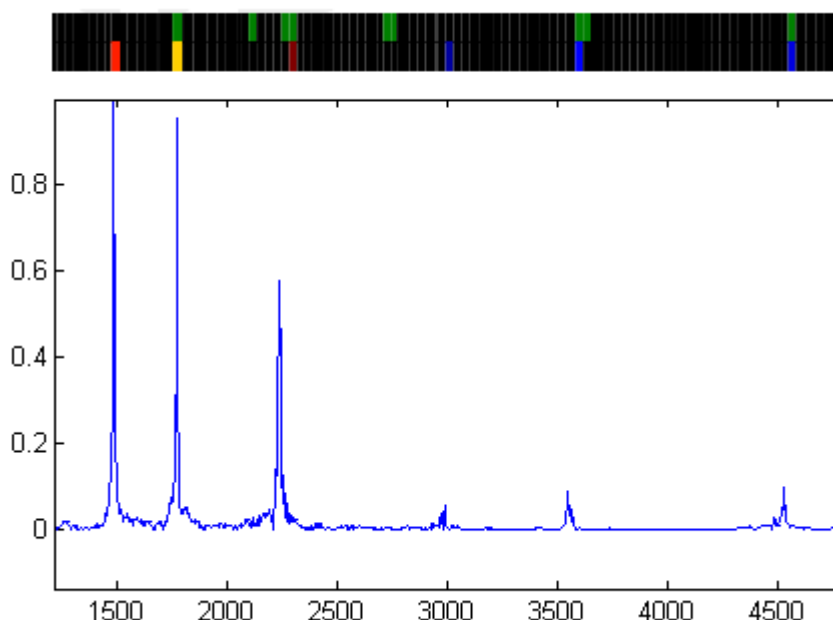
Drugačen problem se pojavi pri napačni klasifikaciji dur akordov, ki se klasificirajo večinsko v mol razred. Prav tako velja obratna relacija. Vpliv višine kvinte v akordu očitno pomembno vpliva na točnost klasifikacije, terca pa ima mnogo manjši vpliv.

Poiskali smo primer akorda, ki je napačno klasificiran. Gre za akord $F\sharp$ mol, zaigran v sedmi oktavi ($F\sharp7$, A, $C\sharp8$). Akord je napačno klasificiran v dur. Poglejmo si znormirane aktivacije posnetka. Aktivirajo se trije deli s sledečimi delnimi odmiki:

- $d_{1/4}$ [0, 9, 16, 28, 48, 64] z aktivacijo 1
- $d_{1/3}$ [0, 9, 16] z aktivacijo 1

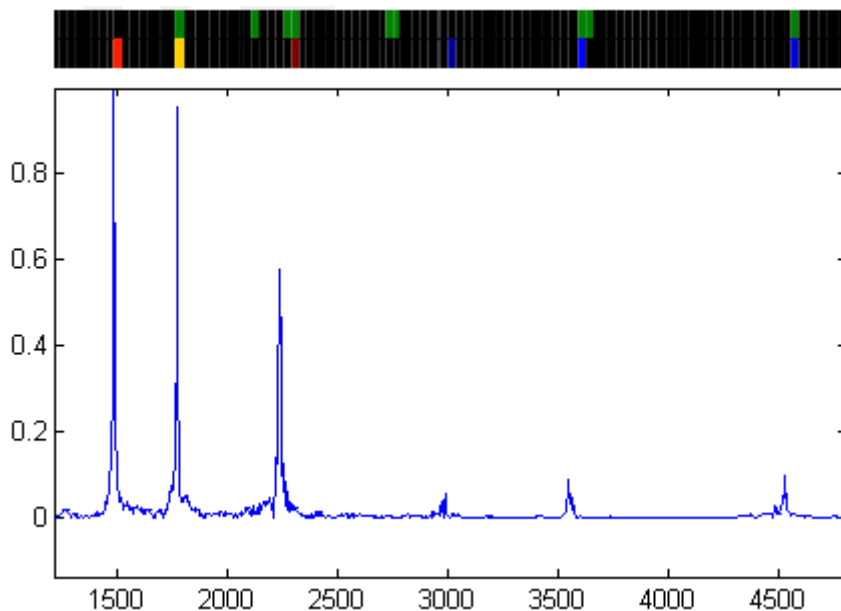
- $d_{2/3}$ $[0, 12, 28]$ z aktivacijo 0,3

Pregled lokacij aktivacij kaže, da se v območju prvih harmonikov (ki so najmočnejši) aktivirata prva dva dela, tretji del pa se aktivira v območju oktave višje in pokriva precej šibkejše harmonike. Pokritje dela $d_{1/4}$ je vidno na sliki 6.1. Opazimo lahko, da se del delno pokriva z aktivacijami na ničtem nivoju, harmonik, ki predstavlja veliko terco (dur), pa pokrije drugi interval med malo terco in kvinto. Aktivacija tega dela zaradi visokega koeficienta in močne aktivacije posledično povzroči klasifikacijo akorda v dur. Na sliki 6.2 je vidna aktivacija dela $d_{2/3}$. Ta aktivacija je v skladu z našimi željami, a na žalost ne pride do izraza zaradi aktivacije dela $d_{1/4}$.



Slika 6.1: Pokritje frekvenčnega spektra dela $d_{1/4}$. Prvi in tretji zaporedni odmik na sliki, ki predstavljata interval velike terce, pokrijeta interval med malo terco in kvinto v molovem akordu. Delni odmik, ki v delu predstavlja kvinto ne pokrije nobenega harmonika, temveč je posledica halucinacije.

Iz primera je razvidno, da se pojavijo aktivacije delov, ki jih smatramo kot zadosten pogoj za pravilno klasifikacijo. Aktivirajo pa se tudi drugi deli, ki negativno vplivajo na klasifikacijsko točnost. Na sliki 6.1 je razvidno, da del vsebuje delni odmik (četrty zaporedni odmik od leve proti desni), ki ne pokriva frekvence v signalu. Aktivacija tega dela je posledica halucinacije drugega in



Slika 6.2: Pokritje frekvenčnega spektra dela $d_{2/3}$. Del pokrije molovsko harmonijo oktavo višje. To je v skladu z našimi željami a je zaradi nizkih magnitud harmonikov posledično tudi aktivacija nizka.

četrtega delnega odmika. Takšno obnašanje modela je zaželeno, v tem primeru pa se halucinacija ne odvije v skladu z našimi pričakovanji.

Prikazani primer nakazuje na možnost, da halucinacijo preveč spodbujamo in odsotnost haluciniranih harmonikov premalo penaliziramo.

6.4 Testiranje modela skladbah skupine *The Beatles*

Model smo želeli preizkusiti tudi na bolj realni glasbi. Za primerjavo smo vzeli anotirane albume skupine *The Beatles*. Za bazo glasbenih posnetkov smo si izbrali albume glasbene skupine *The Beatles*. Baza je primerna za testiranje, ker so posnetki v polnosti anotirani, saj gre za eno izmed standardnih testnih baz pri klasifikaciji akordov. Glasba na posnetkih je harmonično enostavna, z malo dodanimi efekti, kar je dobro za testiranje modela v tej fazi.

Pri gradnji hierarhičnega modela smo uporabili del posnetkov slovenskih

ljudskih pesmi, podmnožico prej omenjene zbirke klavirskih akordov ter nekaj pesmi iz drugega albuma skupine *The Beatles*, ki ni bil vključen v testno množico. Nastavili smo tudi 16 območij delov, kar v praksi pomeni, da posamezen del pokriva pol oktave. Za takšno nastavitev smo se odločili, ker smo želeli bolj specifične dele, ki bi zajemali območja posameznih inštrumentov in glasov.

Za klasifikacijo smo uporabili dva osnovna tipa akordov - dur in mol. V posnetkih nastopa sicer še tip zmanjšanega akorda, a smo zaradi izredno nizke pojavitve v posnetkih (približno 0,3%) okvirje, ki predstavljajo ta tip, odstranili iz učne in testne množice. Za učenje modela logistične regresije smo uporabili klavirske akorde, že predhodno uporabljene v poglavju 6.3. Pri tem smo upoštevali bližino dur in zvečanih akordov ter mol in zmanjšanih akordov, ki smo jih združili v dve skupini. Pri klasifikaciji smo dobili 63,4% klasifikacijsko točnost, kot je opisano v tabeli 6.4.

Tabela 6.4: Rezultati klasifikacije v tipa dur in mol.

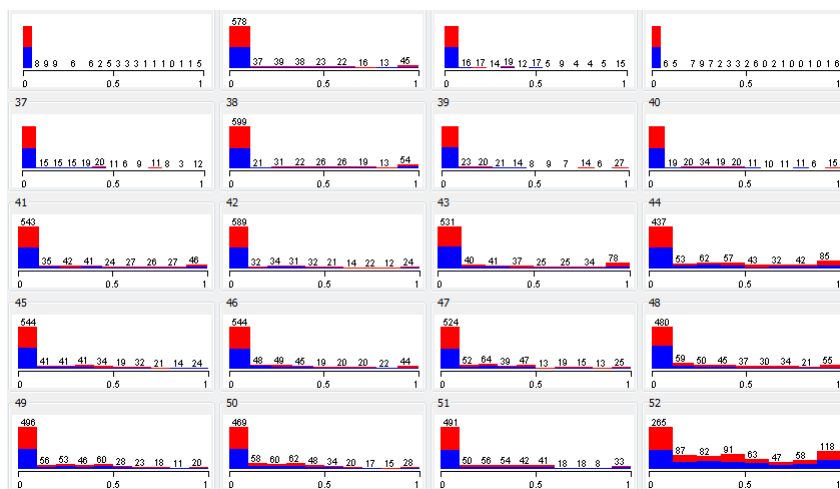
klasifikacijska točnost	mera F
63,4%	69,1%

Matrika klasifikacij prikazuje deleže klasificiranih okvirjev (skupno 27232 testnih okvirjev). Model logistične regresije smo naučili na 700 okvirjih, pri čemer smo bili pozorni na enakomerno porazdelitev dur in mol okvirjev. Matrika klasifikacij je predstavljena v tabeli 6.5.

Tabela 6.5: Matrika klasifikacij (confusion matrix) pri klasificiranju v dur in mol prvega albuma skupine *The Beatles*.

maj	min	
60,7%	39,3%	maj
41,1%	58,8%	min

Rezultat ni tako visok kot pri klasifikaciji akordov. Najprej si oglejmo porazdelitev aktivacij delov zadnjih dveh nivojev. Na sliki 6.3 je prikazan del porazdelitev aktivacij. Posamezen del je predstavljen s histogramom porazdelitev aktivacij glede na magnitudo. Modra barva predstavlja porazdelitev dur akordov, rdeča barva pa predstavlja porazdelitev mol akordov. Opazimo lahko visoko frekvenco izredno nizkih aktivacij.



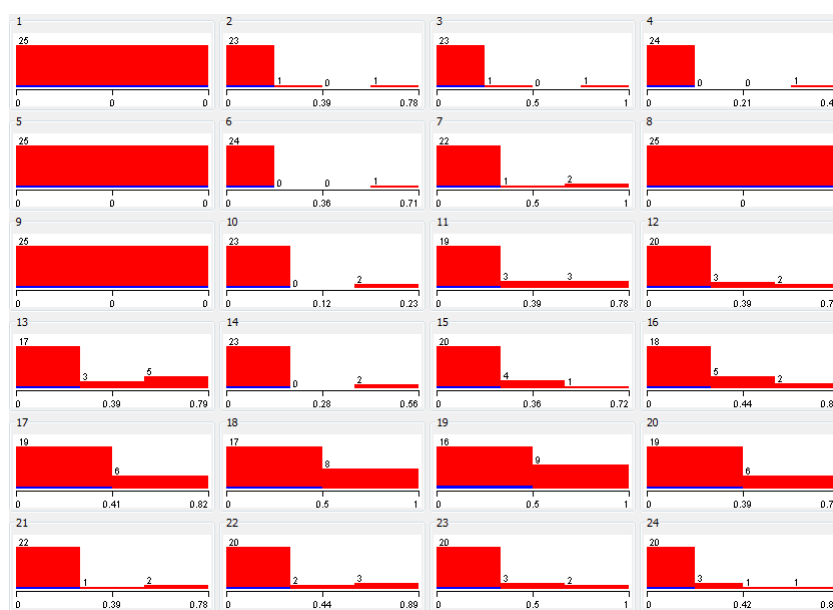
Slika 6.3: Porazdelitve aktivacij nekaterih delov kompozicionalnega modela, izrisane s programskim orodjem *Weka*. Porazdelitve nakazujejo na visoko frekvenco aktivacij z nizko magnitudo, le pri enem delu se porazdelitev pri višjih magnitudah bolj odraža (spodaj desno del 52).

Potrebno je upoštevati, da smo uporabili 16 območij, kar pomeni sorazmerno manjši delež aktivacij posameznega dela (in posledično tudi več delov), a vseeno ne daje zadostnega razloga za tako nizke aktivacije. Omejiti se moramo torej na problem nizkih magnitud aktivacij molovih akordov v pesmi.

Problema sta lahko dva - bodisi naš sistem vsebuje preveč specifičnih delov, ki so vsebovani le v učni množici, oziroma je učna množica preveč različna od testne množice. A to po našem mnenju ne more biti ključni faktor, saj učna množica vsebuje pesmi iste skupine, sicer iz drugega, a po stilu izjemno podobnega, albuma. Drug problem je morda v samem kompozicionalnem modelu pri izbiri potencialnih parov - tudi pri samem učenju hierarhičnega modela se preveč orientiramo na pare, ki v osnovi pokrivajo durove harmonije, ostale pare pa zanemarjamo. Ob pregledu zadnjega nivoja modela lahko opazimo naslednje: 50% parov pokriva specifično dur harmonije v osnovni legi. 17% pokriva specifično mol harmonije v osnovni legi. 17% delov pokriva dur akorde v prvem obratu, ostalih 16% pa navidez ne specifikira določenega tipa harmonije. Model je torej primarno osredotočen na dur akorde, kar vpliva na

klasifikacijsko točnost pri klasifikaciji mola.

Oglejmo si 14 okvirjev pesmi *I saw her standing there* iz albuma *Please please me (1963)* med 24. in 26 sekundo posnetka. Porazdelitve prvih 20 delov prikazuje slika 6.4.



Slika 6.4: Porazdelitve aktivacij 24 okvirjev mol akorda in 1 okvirja dur akorda, izrisanih s programskim orodjem *Weka*. Z modro so označene aktivacije dur okvirja, aktivacije mol okvirjev pa so označene z rdečo.

Vzeli smo zadnji okvir akorda A dur in naslednjih 24 okvirjev akorda A mol. Iz slike 6.4 so razvidne porazdelitve magnitud vžigov dur okvirja in mol okvirjev. Problematično je predvsem dejstvo, da se noben del na tem nivoju ne aktivira v polnosti tedaj, ko se v polnosti ne aktivira tudi mol. Takšni primeri delov so štirje (1, 5, 8 in 9), kjer pride do največje magnitude aktivacije tako pri duru kot pri molu. Pet delov (2, 6, 14, 22 in 23) se sicer aktivirajo z relativno veliko magnitudo, a le v enem do dveh okvirjih. Sicer so njihove aktivacije blizu nič. Ostali okvirji imajo visoko frekvenco aktivacij blizu nič in relativno majhno frekvenco aktivacij magnitud bližje 1.

Problem torej ne nastopa v modelu logistične regresije, temveč že v sami gradnji kompozicionalnega modela.

6.5 Primerjava s sodobnimi rešitvami

Sodobni algoritmi za klasifikacijo akordov, ki jih raziskovalci vsako leto testirajo na tekmovanju MIREX (*Music Information Retrieval Evaluation eXchange*), dosegajo rezultate z okoli 80% klasifikacijsko točnostjo. MIREX je primerjalno tekmovanje znotraj konference ISMIR (*The International Society for Music Information Retrieval*). Algoritmi so testirani na glasbenih posnetkih. Klasificirajo se trozvoki tipov dur, mol in nedoločljivo. Poleg tipa se klasificira tudi tonska višina akorda, česar mi ne upoštevamo.

Trenutni model klasifikacije tonske višine ne omogoča, v model pa smo vgradili osnovni algoritem, ki že omogoča določanje tonskih višin posameznih delov.

Že pri sami klasifikaciji tipov akordov ob primerjavi klasifikacijske točnosti sodobnih algoritmov še ne dosegamo zadovoljivih rezultatov.

Poglavje 7

Zaključek

Kompozicionalni hierarhični pristop modeliranja smo uspešno implementirali in pristop uporabili na področju glasbe. Pristop se je do neke mere izkazal za uspešnega, čeprav se v tej fazi razvoja še ne more meriti s trenutnimi najboljšimi rezultati drugih raziskovalcev. Povprečna klasifikacijska točnost sodobnih algoritmov je okoli 80%, pri čemer je v testiranju merjena tudi tonaliteta akorda. Za potrebe razumevanja modela smo implementirali tudi vizualizacijo, ki se je izkazala za močno orodje pri razvoju, saj je privedla do lažjega razumevanja obnašanja modela. Testiranje na množici akordov je prineslo željene rezultate (83% klasifikacijska točnost), ki so dokazali pravilnost delovanja pristopa.

7.1 Prednosti in omejitve modela

Bistvena prednost modela je po našem mnenju hkratno modeliranje posameznih tonskih višin in harmonije. Model harmonije obravnava kot kompozicije. Mnogo pristopov posnetek sprva razdeli in različne glasbene elemente obravnava ločeno ter šele v kasnejših fazah rezultate ločeno obravnavanih delov posnetka združi v kompleksnejši element. Kompozicionalni model se izkaže v avtonomni obravnavi osnovnih gradnikov in kompoziciji le-teh na podlagi statistike. Vnaprejšnja obdelava in delitev posnetka tako ni potrebna.

Ena izmed omejitev modela je časovna neodvisnost. Implementacija časovnega modeliranja glasbe bi po našem mnenju prinesla boljše rezultate, saj lahko v primeru velike negotovosti hipoteze v določenem času t_n sklepa tudi glede na najverjetnejšo hipotezo v času t_{n-1} .

Modeliranje glasbe s časovno odvisnim kompozicionalnim modelom po našem mnenju predstavlja veliko priložnost za učinkovitejšo modeliranje glasbe, saj model ni odvisen od ločenega obravnavanja različnih gradnikov glasbe, temveč lahko hkrati modelira tako posamezne tonske višine kot razvoj harmonije (chord progression).

Trenutni rezultati prav tako kažejo na trenutno omejenost modela glede na raznolikost glasbe na višjih nivojih, kar nakazuje na potrebo po vpeljavi nadzorovanega učenja v višje nivoje modela. Posledično pričakujemo tudi več delov na višjih nivojih, ki bodo pokrili raznolike harmonije bolj popolno.

7.2 Nadaljnje delo

V prihodnje želimo sprva izboljšati model in klasifikacijo akordov. Prav tako želimo prečistiti morebitne pragovne funkcije za določanje primernosti kandidatskih parov v model, saj menimo, da bi množica delov delovala bolj reprezentativno tudi na testni glasbeni množici.

Glede na problem napačne klasifikacije mola v dur, opisan v poglavju 6.3, bomo preučili možnost dodatnih omejitev pri halucinaciji.

Naslednji korak razvoja modela bo določanje višine harmonij oziroma določanje vseh tonskih višin, prisotnih v posnetku ob nekem trenutku. S tem bomo uvedli tudi klasifikacijo v specifične akorde, saj smo do sedaj določali le tipe slednjih. Razširili bomo tudi klasifikacijo v več tipov akordov, ki jih trenutno ne upoštevamo.

Nato nameravamo vpeljati časovno odvisen kompozicionalni hierarhični model. Ta model bo sposoben modeliranja fraz in motivov, uporabo pa bomo razširili tudi na iskanje podobnosti skladb. Časovno odvisni kompozicionalni model bo vseboval tudi lastnost samostojnega nadaljevanja prekinjenega vhodnega posnetka na podlagi verjetnosti.

Implementacija in vizualizacija modela je torej dokaz koncepta, ki očitno deluje tako na vizualnem kot na glasbenem področju. Pristop kompozicionalnega hierarhičnega modela v glasbi se je izkazal za uspešnega in je po našem mnenju vreden nadaljnjega razvoja.

Slike

- 2.1 Sosledno zapisane note v območju dveh oktav. Note, ki so po
odmiku nižje od območja notnega črtovja, vsebujejo dodatne
črte, ki omogočajo lažje branje 6
- 2.2 Primer iste melodije, zapisane v štiri četrtnskem taktovskem
načinu (zgoraj) in nato v dvo polovinskem taktovskem načinu.
Glede na podan tempo (zapis dolžine osnovne dobe na minuto)
na začetku notnega črtovja, sta zaigrani melodiji identični. . . . 8
- 2.3 Prvi takt prikazuje *swing* zapis, ki se metrično izvaja kot je
zapisano v drugem taktu v *straight* zapisu 8
- 2.4 V prvem taktu nastopa akord A mol z dodano malo septimo. V
drugem taktu je zapisan akord C dur z dodano seksto v osnovni
legi, napram slednjemu pa je v tretjem akordu zapisan možen
obrat akorda. Tretji takt odraža podobnost harmonije, opisane
v prvem taktu, saj so vse tonske višine, z izjemo najnižje, enake.
Takšen zapis akorda v tretjem taktu bi lahko interpretirali tudi
kot A mol z dodano malo septimo in tonom C v basu oziroma
krajše: Am7/C 9
- 2.5 Predstavitev C dur akorda (levo), C mol akorda (sredina) in
dvoumnega akorda (desno). 9
- 2.6 Primer drevesa. (Vir: [http://www.mus.cam.ac.uk/~ic108/
MusicAnalysis/](http://www.mus.cam.ac.uk/~ic108/MusicAnalysis/)) 15
- 2.7 Prikaz vizualnega hierarhičnega modela, sestavljenega iz treh
nivojev. Na prvem nivoju so vidni posamezni deli, ki predstavljajo
ravne črte. Na drugem nivoju so opazne združitve v tri
nove dele, tretji nivo pa vsebuje en sam del, ki pokriva celotno
sliko. Vir: [4] 18

3.1	Enostavni kompozicionalni hierarhični model, sestavljen iz dveh nivojev. Spodaj se nahaja frekvenčni spekter, ki prikazuje ničti nivo hierarhije. Na ničtem nivoju nastopajo štirje absolutni deli, ki so označeni z zeleno barvo. Na prvem nivoju nastopajo trije deli, na drugem nivoju pa dva dela, ki sta sestavljena iz posameznih dveh delov prvega nivoja.	22
3.2	rahel zamik sinusnih signalov pri frekvencah 440 Hz (označen z modro barvo) in 442 Hz (označen z rdečo barvo) na intervalu med 0,015s in 0,02s.	26
3.3	Prikazana sta šesteta signala iz slike 3.2 na intervalu med 0 in 1 sekundo. Na sliki so jasno vidne amplitudne ojačitve pri 0, 0,5 in 1 sekundi ter ošibitvi signala pri 0,25 in 0,75 sekundi.	26
3.4	Slika prikazuje štiri korake obdelave psihoakustičnega filtra na posnetku tona A (440 Hz), zaigranega na klavirju. Podslike A, B in D prikazujejo graf amplitude v odvisnosti od frekvence, podslika C pa vsebuje graf maskiranj posameznih harmonikov glede na frekvenčne kanale.	28
3.5	Prikazan je frekvenčni spekter posnetka akorda A dur (A4, C#5, E5). Vrhovi predstavljajo potencialne aktivacije ničtega nivoja.	29
3.6	Prikazan je ničti nivo hierarhije ob predvajanju posnetka akorda A dur (A4, C#5, E5). Aktivni deli so obarvani v <i>Jet</i> barvni shemi. Neaktivirani deli so obarvani črno.	29
3.7	Formula za izračun magnitude aktivacije dela $d_{i/1}$. Aktivacija dela $d_{i/1}$ je vsota aktivacij delov, $d_{j/0}$ in $d_{k/0}$, deljena z vsoto števila vseh takšnih delov ničtega nivoja, katerih razlika v lokaciji pojavitve je enaka delnemu odmiku dela $d_{i/1}$	31
3.8	Prikaz obtežitvene funkcije delnih odmikov glede na centralni del na intervalu med 0 in 100. Glede na frekvenčne kanale je odmik ene oktave razdalja 48 frekvenčnih kanalov. Na sliki je ordinatna os zaradi preglednosti primerno skrajšana.	36
4.1	Primer dela na četrtem nivoju. Delni odmiki nakazujejo na molov akord. Pri interpretaciji upoštevamo, da vsak polton modeliramo s štirimi frekvenčnimi kanali. Del tako predstavlja delne odmike treh, sedmih, dvanajstih poltonov glede na osnovni ton.	40
4.2	Primer dela na drugem nivoju, ki združuje delne odmike v obliki klavirskih tipk.	40
4.3	Formula logistične funkcije.	41

4.4	Izris logistične funkcije na intervalu med 0 in 1.	42
5.1	Delovno okno vizualizacijskega programskega orodja, izdelanega kot pripomoček pri razvoju kompozicionalnega hierarhičnega modela.	44
5.2	Prikaz ničtega nivoja skupaj s pokritjem nekega dela na 3. nivoju. Za magnitudo aktivacije je uporabljena barvna shema <i>jet</i> , kot je definirana v programskem orodju <i>Matlab</i> . Zgoraj so z zeleno barvo označena pokritja izbranega dela v modelu. . . .	45
5.3	Del tipkovnice ob predvajanju akorda A4 C5 E5. Na tem mestu pride do pojavitve tonske višine A3 in A5 (skrajna leva in desna obarvana tipka). Za magnitudo normiranega seštevka aktivacij je uporabljena barvna shema <i>jet</i> , kot je definirana v programskem orodju <i>Matlab</i>	45
5.4	Izrisan neaktiviran del na klavirski tipkovnici. Delni odmiki so risani od A4 - 440Hz dalje za lažjo predstavo tonskih višin, ki jih del s svojimi delnimi odmiki pokriva.	46
5.5	Primer delno označenega hierarhičnega modela glede na izbrani del (id 7 na četrtem nivoju). Na vsakem posameznem nivoju so označeni deli, ki so del kompozicije na višjem nivoju. Povezave med označenimi deli so obarvane črno.	47
6.1	Pokritje frekvenčnega spektra dela $d_{1/4}$. Prvi in tretji zaporedni odmik na sliki, ki predstavljata interval velike terce, pokrijeta interval med malo terco in kvinto v molovem akordu. Delni odmik, ki v delu predstavlja kvinto ne pokrije nobenega harmonika, temveč je posledica halucinacije.	53
6.2	Pokritje frekvenčnega spektra dela $d_{2/3}$. Del pokrije molovsko harmonijo oktavo višje. To je v skladu z našimi željami a je zaradi nizkih magnitud harmonikov posledično tudi aktivacija nizka.	54
6.3	Porazdelitve aktivacij nekaterih delov kompozicionalnega modela, izrisane s programskim orodjem <i>Weka</i> . Porazdelitve nakazujejo na visoko frekvenco aktivacij z nizko magnitudo, le pri enem delu se porazdelitev pri višjih magnitudah bolj odraža (spodaj desno del 52).	56
6.4	Porazdelitve aktivacij 24 okvirjev mol akorda in 1 okvirja dur akorda, izrisanih s programskim orodjem <i>Weka</i> . Z modro so označene aktivacije dur okvirja, aktivacije mol okvirjev pa so označene z rdečo.	57

Tabele

6.1	Predstavitev podobnosti nižjih nivojev glede na omejitve združevanja znotraj omejenega območja in brez omejitve združevanja.	50
6.2	Testiranje klasifikacije na razdeljeni učni množici.	52
6.3	Matrika klasifikacij (confusion matrix) pri klasificiranju v vse štiri razrede.	52
6.4	Rezultati klasifikacije v tipa dur in mol.	55
6.5	Matrika klasifikacij (confusion matrix) pri klasificiranju v dur in mol prvega albuma skupine <i>The Beatles</i>	55

Literatura

- [1] Wideband Audio and IP Telephony [Cisco Unified IP Phone 7900 Series]. Dostopno na http://www.cisco.com/en/US/prod/collateral/voicesw/ps6788/phones/ps379/ps8537/prod_white_paper0900aecd806fa57a.html, Januar 2012.
- [2] John Ashley Burgoyne, Jonathan Wild, in Ichiro Fujinaga. An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis. Objavljeno v *Proceedings of The International Society for Music Information Retrieval*, strani 633–638, 2011.
- [3] Wei Chai in Barry Vercoe. Folk music classification using hidden Markov models. Objavljeno v *Proceedings of International Conference on Artificial Intelligence*, 2001.
- [4] Luka Čehovin. *Kategorizacija predmetov z metodo podpornih vektorjev na osnovi hierarhičnih lokalnih značilnic*. Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, 2007.
- [5] Hugo Fastl in Eberhard Zwicker. *Psychoacoustics: facts and models*. Springer, 2007.
- [6] Daniel J. Felleman in David C. Van Essen. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1):1–47, Januar 1991.
- [7] Sanja Fidler, Marko Boben, in Aleš Leonardis. Learning Hierarchical Compositional Representations of Object Structure. Objavljeno v *Dickinson, S., Leonardis, A., Schiele, B., Tarr, M.J. (eds.) Object Categorization: Computer and Human Vision Perspectives.*, strani 196–215. Cambridge University Press, 2009.

- [8] Sanja Fidler in Aleš Leonardis. Towards scalable representations of object categories: Learning a hierarchy of parts. Objavljeno v *Computer Vision and Pattern Recognition, CVPR'07. IEEE*, stran 1–8, 2007.
- [9] Phillip B Kirlin in Paul E Utgoff. Computational Musicology - A Framework For Automated Schenkerian Analysis. Objavljeno v *Proceedings of The International Society for Music Information Retrieval*, strani 363–367, 2008.
- [10] K. Lee in M. Slaney. A Unified System For Chord Transcription And Key Extraction Using Hidden Markov Models. Objavljeno v *Proceedings of International Conference on Music Information Retrieval*, 2007.
- [11] Kyogu Lee in Malcolm Slaney. Automatic chord recognition from audio using a supervised HMM trained with audio-from-symbolic data. Objavljeno v *Proceedings of the 1st ACM workshop on Audio and music computing multimedia, AMCMM '06*, stran 11–20, New York, NY, USA, 2006. ACM.
- [12] Llewelyn Southworth Lloyd in Hugh Boyle. *Intervals, scales and temperaments*, del 1978. Macdonald and Janes, Januar 1978.
- [13] Alan Marsden. Recognition of variations using automatic Schenkerian reduction. Objavljeno v *Proceedings of The International Society for Music Information Retrieval*, strani 501–507, 2010.
- [14] Matthias Mauch in Simon Dixon. A Discrete Mixture Model for Chord Labelling. Objavljeno v *Proceedings of The International Society for Music Information Retrieval*, strani 45–50, 2008.
- [15] Nicola Orio. Music Retrieval: A Tutorial and Review. *Foundations and Trends in Information Retrieval*, 1(1):1–90, 2006.
- [16] Rainer Typke, Frans Wiering, in Remco C. Veltkamp. A Survey Of Music Information Retrieval Systems. Objavljeno v *Proceedings of The International Society for Music Information Retrieval*, strani 153–160, 2005.