

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Primož Godec

**Zbiranje baze in evalvacija algoritmov
za ocenjevanje razpoložanja v glasbi**

DIPLOMSKO DELO
UNIVERZITETNI ŠTUDIJSKI PROGRAM PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Matija Marolt

Ljubljana 2014

Rezultati diplomskega dela so intelektualna lastnina avtorja. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja¹

Besedilo je oblikovano z urejevalnikom besedil \LaTeX .

Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Tematika naloge:

Besedilo teme diplomskega dela študent prepíše iz študijskega informacijskega sistema, kamor ga je vnesel mentor. V nekaj stavkih bo opisal, kaj pričakuje od kandidatovega diplomskega dela. Kaj so cilji, kakšne metode uporabiti, morda bo zapisal tudi ključno literaturo.

IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Primož Godec, z vpisno številko **63110452**, sem avtor diplomskega dela z naslovom:

Zbiranje baze in evalvacija algoritmov za ocenjevanje razpoložanja v glasbi

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom doc. dr. Matije Marolta,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela,
- soglašam z javno objavo elektronske oblike diplomskega dela na svetovnem spletu preko univerzitetnega spletnega arhiva.

V Ljubljani, dne 30. avgust 2014

Podpis avtorja:

Želim se zahvaliti mentorju Matiji Maroltu, za pomoč in spodbujanje pri raziskovanju in izdelavi diplomske naloge. Prav tako se želim zahvaliti Matevžu Pesku, ki si je vedno vzel čas, ko sem ga potreboval in s spodbujanjem poskrbel, da je bilo diplomsko delo napisano hitreje, kot bi bilo drugače. Zahvalil bi se tudi ekipi s katero smo sodelovali na projektu raziskovanja razpoloženja in glasbe. Ekipo sestavljajo Matevž Pesek, Matija Marolt, Mojca Poredoš, Jože Guna, Gregor Strle, Emilija Stojmenova in Matevž Pogačnik. Nazadnje bi se zahvalil še družini in prijateljem, ki mi vedno stojijo ob strani.

Moji družini in prijateljem.

Kazalo

Povzetek

Abstract

1	Uvod	1
2	Pregled področja	3
2.1	Pridobivanje informacij iz glasbe	4
2.2	Ocenjevanje razpoložanja v glasbi	6
3	Moodo podatkovna zbirka	13
3.1	Zbiranje podatkovne zbirke	13
3.2	Sestava podatkovne zbirke	19
3.3	Analiza podatkov	24
4	Algoritmi za ocenjevanje razpoložanja v glasbi	37
4.1	Regresijski algoritem	37
4.2	Essentia Gaiatransofrm algoritem	40

Seznam uporabljenih kratic

SVM	support vector machine	metoda podpornih vektorjev
MIR	music information retrieval	pridobivanje informacij iz glasbe
MIREX	Music Information Retrieval Evaluation eXchange	
VA	Valence-arousal	prijetnost-aktivnost
MIDI	Musical instrument digital interface	
MFCC	Mel-frequency cepstral coefficients	

Povzetek

V vzorcu je predstavljen postopek priprave diplomskega dela z uporabo okolja L^AT_EX. Vaš povzetek mora sicer vsebovati približno 100 besed, ta tukaj je odločno prekratek.

Ključne besede: glasba, razpoloženje, čustva, algoritem za ocenjevanje razpoloženja, .

Abstract

This thesis present an evaluation of mood estimation algorithms.

This thesis presents a new dataset gathered containing perceived and induced emotions for 200 audio clips. The gathered dataset also provides users' association of color for each clip, along with users' demographic and personal data, such as users' emotion state, preferred genres, music experience and daily inference, and others. With an online survey we collected more than 7000 responses for a dataset of 200 audio excerpts, thus providing about 37 user responses per clip. We introduced a new methodology for gathering user perception of emotions in a form of two new interfaces - the MoodGraph and MoodStripe. We present a preliminary evaluation of classifying the present emotions with a regression algorithm on the gathered dataset, and perform a comparison towards other datasets.

Keywords: music, mood, emotions, mood clasification algorithm.

Poglavje 1

Uvod

Na svetovnem spletu se velikokrat srečamo s priporočilnimi sistemi. Ko poslušamo neko pesem, si ogledamo video ali preberemo članek, nam ti sistemi sami predlagajo vsebine, ki nas bi še lahko zanimale. S takimi sistemi se srečujemo na vsakem koraku, ne zavedamo pa se, kako pravzaprav delujejo.

Na področju glasbe od priporočilnega sistema pričakujemo, da nam predlaga skladbe, ki bi jih želeli poslušati v danem trenutku. Priporočilni sistemi, ki so trenutno zelo razširjeni so YouTube, Pandora, Last.fm in ostali. Najboljši sistemi so tisti, ki izbor čim bolj prilagodijo našim željam in potrebam. Večino takih sistemov na področju glasbe deluje glede na podobnost uporabnikove zgodovine poslušanja, uporabljajo različne uporabniške oznake ali delujejo na drug način, ki ne vključuje elementov iz področja pridobivanja informacij iz glasbe (MIR - Music information retrieval). Drugi sistemi uporabljajo podatke izračunane iz glasbe. Glede na naš pregled področja noben javno dostopen algoritem za priporočilni sistem ne uporablja podatkov o razpoloženju v glasbi.

Razpoloženje je eden pomembnih dejavnikov, ki vplivajo na to, kakšno glasbo bi uporabnik želel poslušati v določenem trenutku. Obenem pa tudi izbor glasbe lahko spremeni razpoloženje uporabnika. Zaradi tega se nam zdi pomembno, da bi priporočilni sistemi vključevali tudi podatke o razpoloženju. Odločili smo se, da naredimo tak sistem. Preden pa lahko ta sistem razvijemo, bo potrebno še nekaj raziskovanja.

Osnova za raziskavo je dobra podatkovna zbirka. Na področju razpoloženja trenutno obstaja nekaj podatkovnih zbirk [10, 30, 34, 31, 25], ampak nobena v

celoti ne vključuje podatkov, ki bi zadostili našim potrebam. Zato smo se odločili, da zberemo svojo podatkovno zbirko. To smo izvedli z anketo. V glavnem delu ankete smo spraševali po razpoloženju v glasbi in o razpoloženju, ki ga glasba vzbudi pri uporabniku. Poleg tega nas je še zanimalo, s kakšno barvo bi uporabnik opisal določeno pesem. Zbirali smo še podatke o uporabnikovi percepciji razpoloženja in barve, o njegovem trenutnem razpoloženju in nekatere demografske podatke. Glede na naše podatke je ta zbirka največja tovrstna zbirka na področju razpoloženja.

Drugi pomemben element v priporočilnem sistemu na podlagi razpoloženja je algoritem, ki bo znal določiti razpoloženje v glasbi čim bolj natančno. Zaenkrat algoritmi še niso zelo zanesljivi, saj je njihova natančnost ocenjevanja še ni tako dobra. Se pa ta iz leta v leto izboljšuje.

Algoritmi v osnovi delujejo tako, da na podlagi značilnic iz glasbe in podatkov o razpoloženju iz podatkovne zbirke izračunajo določene parametre, na podlagi katerih potem izvajajo ocenjevanje. V ta namen so uporabljeni različni algoritmi kot so metoda podpornih vektorjev (SVM - Support Vector Machine), regresija, različna drevesa, naivni bayes in podobni.

V tej nalogi bomo preizkusil več takih sistemov na naši podatkovni zbirki. Pokazali bomo kakšni so rezultati, če uporabimo regresijski algoritem, ki določa prijetnost (valence) in aktivnost (arousal) v glasbi. Prav tako bomo preizkusili algoritem, ki za svojo osnovo uporablja SVM in na podlagi tega določi enega od petih razpoloženskih gruč, ki so uporabljene tudi na MIREX (Music Information Retrieval eXchange) opravilu. To je algoritem iz knjižnice Essentia.

V naslednjem poglavju bomo pregledali področje, ki ga obsega ta diplomska naloga. V poglavju 3 bomo bolj natančno predstavili zbiranje naše podatkovne zbirke in naredil krajšo analizo podatkov v njej. V četrtem poglavju bomo predstavili, kako delujejo različni algoritmi za določanje razpoloženja v glasbi in predstavil rezultate, ki jih dobimo, če te algoritme uporabimo na naši podatkovni zbirki. V zadnjem poglavju bomo predstavili lasten algoritem, njegovo delovanje in rezultate. [todo: spremeni]

Poglavje 2

Pregled področja

Lahko bi rekli, da je glasba ena najstarejših in zelo pomembnih aktivnosti na svetu. Razširjena je po celem svetu, poznajo jo tudi izolirana in od ostalega sveta odmaknjena plemena. Znano je, da glasba na svetu obstaja že vsaj 50 000 let [18]. Prvo glasbo naj bi takrat izvajali na Afriških tleh. Nato se je skozi čas razvijala in postala ena najpomembnejših sestavnih delov človekovega življenja. O pomembnosti glasbe priča dejstvo, da jo lahko slišimo praktično na vsakem koraku. Poslušamo jo doma, na poti, ko nam je dolgčas ali ko se želimo razvedriti, poslušamo jo, ko smo na kavi ali v trgovini. Ponekod s pravo izbiro glasbe vplivajo na človekove odločitve. Na primer, v trgovinah in lokalih z glasbo privlačijo kupce. Vse to priča o pomembnem vplivu glasbe na človeka.

Glasba pomembno vpliva na človekova čustva in razpoloženje. To moč ima predvsem zaradi tega, ker ima neposredno pot do čustev. Glasbo namreč doživljamo z notranjimi čuti, zato ni potrebne predhodne interpretacije, kot je potrebna pri razumevanju tiskane besede. Znano je, da različna glasba vzbudi različna čustva in ima moč, da vzbudi potlačena čustva. Na vzdušje vpliva razmerje med toni. Mol pričara bolj melanholično vzdušje, dur pa bolj veselo [22].

V nadaljevanju poglavja bomo povedali še nekaj o področju, imenovanem pridobivanje informacij iz glasbe. To področje je pomembno, saj je osnova za temo diplomske naloge. Povedali bomo nekaj o povezavi med razpoloženjem in glasbo. Pregledali bomo podatkovne zbirke, ki obstajajo trenutno na področju razpoloženja in glasbe. Za konec pa bomo predstavili še algoritme za ocenjevanje razpoloženja iz glasbe.

2.1 Pridobivanje informacij iz glasbe

Pridobivanje informacij iz glasbe (MIR) je interdisciplinarna znanost, ki povezuje predvsem muzikologijo in računalništvo[26]. Vključuje pa tudi področja kot so psihologija, procesiranje signalov, strojno učenje in ostale.

To področje je dokaj novo in se trenutno hitro širi. Področje se je začelo razvijati v 80ih letih prejšnjega stoletja in je močno povezano z razvojem računalništva. Šele v današnjem času pa je, predvsem zaradi povečanih računskih zmognosti računalnikov, doživelo razcvet. Pojavljajo se velike razlike v načinu obdelave in uporabe podatkov. Prav tako so cilji raziskovalcev zelo različni. Pomembni cilji v MIR so: narediti dober algoritem za ocenjevanje razpoloženja in žanrov iz glasbe, dobro ocenjevati akorde, izboljšati algoritme za transkripcijo glasbe v simbolično obliko in ostali.

Ključni namen področja je pridobiti informacije iz glasbe in jih uporabiti. Poznamo več vrst informacij. Lahko so pridobljene direktno iz zapisa glasbe ali simboličnega zapisa, lahko pa so metapodatki, ki so pridobljeni na drug način. Rezultati služijo za različna opravila. V nadaljevanju bomo predstavili opravila, na katerih se trenutno največ dela.

Priporočilni sistemi za glasbo (Music Recommendation Systems)

Trenutno obstaja veliko takih sistemov kot so YouTube, Pandora in Last.fm, ampak samo nekateri uporabljajo informacije pridobljene z MIR za svoje delovanje. Namesto tega veliko sistemov uporablja informacije na podlagi primerjave med zgodovinami poslušanja uporabnikov. Na primer, sistem predlaga glasbo, ki so jo poslušali uporabniki s podobno zgodovino poslušanja. Drugi spet uporabljajo oznake k določeni glasbi ali druge informacije, ki niso del MIR-a. Te oznake lahko dodajo uporabniki ali strokovnjaki. Pri sistemu Pandora glasbo označujejo strokovnjaki, pri sistemu Last.fm pa uporabljajo oznake dodane s strani uporabnikov. Oznake, ki se uporabljajo, so lahko različne. Lahko označijo zvrst glasbe (npr. rock, pop, ...), opisujejo razpoloženje (vesel, žalosten, sproščen, ...), povejo ali je pesem instrumentalno izvedena ali vsebuje vokal, je ta vokal ženski ali moški in podobno. Od kar se je področje MIR zelo razširilo vedno več priporočilnih sistemov uporablja informacije pridobljene iz glasbe.

Ločevanje pesmi na več izvorov in prepoznavanje inštrumentov Sistem za ločevanje pesmi poizkuša pesem, v kateri nastopa več instrumentov in vokali, razstaviti tako, da imamo posamezne izvore v pesmi ločene. Na primer lahko loči glede na instrumente, za kar potrebuje sistem za prepoznavanje instrumentov. Zaradi tega sta ta sistema tako tesno povezana med seboj.

Na tem področju obstaja že kar nekaj sistemov [12, 23]. Sistemi se uporabljajo za ločevanje posameznega inštrumenta iz glasbe, kar pride prav pri izključevanju posameznih instrumentov iz podlage. Na tak način lahko izključimo tudi vokal iz glasbe, kar je uporabljeno pri sistemih za pripravljanje pesmi za karaoke. Pri sistemih za izključevanje vokala iz glasbe je še veliko možnosti za izboljšave. Problem je v tem, da v vokalu nastopajo določene frekvence, ki se pojavijo tudi pri določenih inštrumentih.

Avtomatična transkripcija glasbe Ti sistemi delujejo tako, da glasbo iz posnetka pretvarjajo v simbolični zapis [11, 16, 17]. Najpogosteje je to uporabljeno pri prepisovanju glasbe v zapis MIDI [32]. Rešitve za transkripcijo vključujejo številne podsisteme: pojavitev glasbenih dogodkov (onset detection), ocenjevanje trajanja, identifikacijo instrumentov, prepoznavanje ritma in ostale. Rešitve, ki trenutno obstajajo še niso popolne. Predvsem nastane problem, ko je v pesmi veliko inštrumentov in več stopenj polifonije (nastopa več sočasnih osnovnih frekvenc).

Avtomatična kategorizacija glasbe To so sistemi, ki znajo razvrstiti glasbo v več vnaprej definiranih skupin. Na področju MIR se raziskovalci trenutno ukvarjajo predvsem s kategorizacijo po žanrih in razpoloženju v glasbi. Za obe temi MIREX organizira opravilo v kategorizaciji glasbe. Raziskovalci lahko oddajo svoj algoritem, ki ga potem poženejo na MIREX-ovi podatkovni zbirki. Za kategorizacijo se uporabljajo tehnike strojnega učenja, kot so SVM [3], regresija, različna drevesa in druge.

Generiranje glasbe Eden od ciljev raziskovalcev v MIR je tudi izdelati dober sistem za avtomatično generiranje glasbe. V trenutnih sistemih je potrebno predvsem veliko ročnega prilagajanja in nastavljanja različnih parametrov. Cilj je

narediti sistem, ki bi generiral glasbo popolnoma samostojno in bo slednja primerljiva človeškim izdelkom.

2.2 Ocenjevanje razpoloženja v glasbi

Eno pomembnih področij v MIR je ocenjevanje razpoloženja v glasbi s pomočjo računalniških algoritmov. Ti algoritmi v osnovi delujejo tako, da najprej iz zvočnega zapisa izračunajo določene značilnice. V naslednjem koraku se na podlagi že obstoječe podatkovne zbirke nauči algoritem. Izračunajo se določeni parametri, na podlagi katerih potem poteka razvrščanje. Nato se izvede klasifikacija na večji zbirki zvočnih posnetkov. Za delovanje teh algoritmov potrebujemo podatkovno zbirko z že obstoječimi podatki o razpoloženju, zato bomo v naslednjem podpoglavju opisali, kaj na tem področju že obstaja. Poleg tega bomo naredili še pregled algoritmov.

2.2.1 Podatkovne zbirke

Na področju razpoloženja v MIR obstaja nekaj podatkovnih zbirk, ki vsebujejo različne oznake za glasbo.

Podatkovna zbirka s filmsko glasbo Eerola et. al [10] so zbrali podatkovno zbirko, ki vključuje filmsko glasbo iz različnih znanih filmov. Ta podatkovna zbirka je razdeljena na dva dela. Prvi del vsebuje 361 glasbenih odlomkov. Za vsak odlomek imajo povprečno vrednost za prijetnost (ang. valence), aktivnost (ang. arousal) in napetost (ang. tension). Poleg tega so za vsako pesem zbrali tudi oznake za razpoloženje (vesel, žalosten, nežen in podobne). K vsakemu odlomku je dodana tudi povprečna številčna vrednost o prisotnosti posameznega razpoloženja izmed nabora razpoloženj: jeza, strah, veselje, žalost in nežnost. Drugi del zbirke vsebuje 110 glasbenih odlomkov, ki so podmnožica odlomkov iz prvega dela. Poleg vseh oznak, ki jih ima prvi del ima ta del dodane številčne vrednosti, ki označujejo prisotnost za občutij lepota (beauty) in naklonjenost (liking). Ta podatkovna zbirka poleg podatkov vsebuje tudi glasbene odseke, ki so dolgi 15 sekund. Kar je dobro za evalvacijo podatkovne zbirke na različnih algoritmihi. Podatkovna zbirka

vsebuje še podatke o naslovu filma za posamezen odlomek in podatek iz katerega dela pesmi je bil odrezan.

Mood Swing Turk Dataset Podatkovna zbirka Mood Swing Turk Dataset [30] je bila zbrana za 240 glasbenih odlomkov popularne glasbe. Vsak odlomek vsebuje povprečno 17 vrednosti, ki opisujejo prijetnost (valence) in aktivnost (arousal). Dodani so še podatki o glasbi (naslov, avtor, album) in tudi podatki o poteku ocenjevanja (krog ocenjevanja, v katerem so bile zbrane ocene in identifikacijska številka uporabnika, ki je podal posamezno oceno).

Sama podatkovna zbirka ne vsebuje uporabljenih glasbenih odsekov, so pa zaradi tega objavljene že izračunane značilnice za posamezno pesem. Podatkovna zbirka vsebuje naslednje značilnice: Mel-frequency cepstral coefficients (MFCCs), Octave-Based Spectral Contrast, Statistical Spectrum Descriptors (SSDs), Chromagram in EchoNest Audio Features.

Cal500 Podatkovna zbirka Cal500 [34] ima zbrane podatke o razpoloženju za 500 pesmi. Pesmi so zbrane med zahodno popularno glasbo. Za vsako pesem vsebuje oznako za razpoloženje (vesela, žalostna, jezna in podobne). V naboru je na voljo 18 možnih oznak. Podatkovna zbirka je bila označena ročno in prikazuje 3 oznake na pesem. Podatkovna zbirka vsebuje tudi datoteke z glasbo v MP3 obliki.

MTV Music Dataset MTV Music Dataset [31] vsebuje podatke za 192 pesmi izbrane iz MTV Europe Most Wanted lestvic med leti 1981 in 2000. Celotna podatkovna zbirka je bila označena s strani petih ocenjevalcev. Vsaka ocena vsebuje vrednost za prijetnost (valence) in vrednost za aktivnost (arousal). Posamezna pesem je bila označena s strani treh ocenjevalcev.

LAMP Podatkovna zbirka LAMP [7] vsebuje podatke za 492 popularnih pesmi izdanih med 2002 in 2008. Ta zbirka je bila označena s strani 400 anketirancev v treh korakih. V prvem je anketiranec dobil samo besedilo in na podlagi tega določil prijetnost (valence) in aktivnost (arousal). V drugem koraku je dobil samo zvočni posnetek ter določal aktivnost in prijetnost. V tretjem koraku je postopek ponovil z zvočnim posnetkom in besedilom.

Multi Modal Zadnja zbirka, ki jo bom opisal je tako imenovana Multi Modal [25] podatkovna zbirka. Vsebuje 903 glasbene odlomke, ki so dodani podatkovni zbirki. Vključena je večinoma popularna zahodna glasba. Za vsako pesem je določena oznaka z razpoloženjem iz nabora: boisterous, confident, passionate, rowdy, amiable-good natured, cheerful, fun, rollicking, sweet, autumnal, bittersweet, brooding, literate, poignant, wistful, campy, humorous, silly, whimsical, witty, wry, aggressive, fiery, intense, tense - anxious, visceral. Poleg tega pa je tudi vsaka razvrščena v enega od petih gruč definiranih s strani MIREX-a (gruče so predstavljene v poglavju 2.2.2).

2.2.2 Algoritmi za ocenjevanje razpoloženja v glasbi

Vsi algoritmi za ocenjevanje razpoloženja iz glasbe potrebujejo za svojo delovanje podatkovno zbirko, zato smo se v prejšnjem poglavju posvetili pregledu področja obstoječih zbirk. Poleg podatkov moramo zgenerirati tudi različne značilnice, na podlagi katerih se potem izvedemo klasifikacijo v VA prostor ali katero drugo kategorijo.

Delovanje algoritmov za ocenjevanje razpoloženja v glasbi delimo na dva dela. Prvi del je treniranje algoritma. V drugem delu z algoritmom izvedemo klasifikacijo. Treniranje poteka tako, da uporabimo eno izmed metod strojnega učenja. Na podlagi značilnic in podatkov o razpoloženju iz podatkovne zbirke algoritem izračuna pravilo za klasifikacijo. Ta postopek se izvaja na delu podatkovne zbirke namenjene treniranju algoritma. Z naučenim algoritmom izvedemo klasifikacijo na preostanku podatkovne zbirke. Rezultat, ki ga vrne algoritem, je lahko skupina po razpoloženju, v katero uvrstimo določeno pesem ali pa številčna vrednost, ki določa le prijetnost in aktivnost. Ocenjevanje se izvede na drugem delu zbirke pesmi.

To je splošen postopek klasifikacije. Vsak algoritem ima svoje posebnosti. Obstoječi algoritmi so opisani v nadaljevanju.

Schmidt et. al [29] so uporabili regresijski algoritem (metoda najmanjših kvadratov) za klasifikacijo. Podatkovno zbirko so razdelili na dva dela tako, da so 70% podatkovne zbirke uporabili za treniranje algoritma in 30% za testiranje. Za značilnice so uporabili mel-frequency cepstrum in kromatski vektor. Algoritem napoveduje prijetnost (valence) in aktivnost (arousal) pesmi. Parametra algoritem

napoveduje ločeno.

Panda et al. [25] so uporabili več algoritmov za napovedovanje in primerjali točnost napovedovanja. Uporabili so: support vector machines, k-nearest neighbours, C4.5 in naive bayes. Algoritem izvaja klasifikacijo na podlagi 19 značilnic izračunanih iz glasbenega posnetka. Tem značilnicam so dodali tudi značilnice pridobljene iz MIDI signala in značilnice iz besedil. Izvajali so klasifikacijo v pet gruč, ki so definirane za MIREX opravilo (tabela 2.1). Kot najboljši algoritem se je izkazal SVM. Na podlagi značilnic iz zvočnega zapisa je dosegel natančnost 64%. Ko so bile dodane še značilnice iz MIDI zapisa in besedil, pa se je natančnost zmanjšala na 61.1%. Algoritem so preizkusili na podatkovni zbirki z 903 pesmimi.

Laurier et. al [20] so za ocenjevanje uporabili support vector machine [3]. Algoritem za delovanje uporablja 133 značilnic. Tudi ta algoritem deluje tako, da kot rezultat za vsako pesem vrne eno od petih gruč opisanih v tabeli 2.1. Preizkusili so več različnih SVM metod in ugotovili, da najboljše rezultate vrača metoda C-SVC z RBF (radial basis function) jedrom iz SVMlib knjižnice.

Diane Watson [35] je uporabila klasifikator z Bayesovimi mrežami in Markove ocene. Kot rezultat je algoritem napovedoval prijetnost (valence) in aktivnost (arousal). Za razliko od ostalih je uporabila podatke, ki niso bili zajeti v laboratorijskem okolju temveč v vsakdanjem življenju. Uporabniki so skladbe ocenjevali s pomočjo pametnih telefonov kjerkoli so se nahajali v trenutku, ko jih je aplikacija prosila za oceno. Poleg značilnic iz glasbe je algoritem upošteval tudi podatke o tem kako prijetno se uporabnik počuti v času ocenjevanja in podatke o njegovi aktivnosti ter še nekatere druge podatke o stanju in okolju med tem ko je ocenjeval. Algoritem je dosegel natančnost 67% za aktivnost (arousal) in 75% za prijetnost (valence).

Saari et. al [28] so uporabili semantic layer projection (SLP) metodo za klasifikacijo. Za klasifikacijo so uporabili značilnice, končni rezultat pa so oznake z razpoloženji. Za razliko od prej opisanih metod, ki klasifikacijo izvedejo v enem koraku, je tu drugače. Tukaj najprej preslikajo značilnice v tri dimenzionalni prostor z uporabo metode delnih najmanjših kvadratov (partial least squares - PLS). V drugem koraku na podlagi teh vrednosti določijo razpoloženje. Algoritem so poskusili tudi v kombinaciji z tekstovnimi oznakami, vendar je deloval slabše.

Tabela 2.1: Gruče razpoloženj, ki se uporabljajo v MIREX mood opravilu

Gruče	Razpoloženja
Gruča 1	strastno, vznemirjeno, razposajeno
Gruča 2	veselo, zabavno, prijetno, ljubeznivo
Gruča 3	ganljivo, bridko, potrto, otožno, grenko-sladko, malodušno
Gruča 4	šaljivo, duhovito, smešno, neumno, čudaško, muhasto
Gruča 5	agresivno, nasilno, ognjevito, razdražljivo, napeto, intenzivno, nestanovitno, spreminjajoče

2.2.3 Značilnice

Značilnice razdelimo v skupine po podobnosti. Eno od takih razporeditev, povzeto po [20, 4], bomo predstavili v nadaljevanju :

- **Spektralne značilnice** se nanašajo na obliko spektra v glasbi. Izmed teh značilnic se na MIR področju veliko uporabljajo MFCC, HFC (high frequency content), kromatski vektor, spektralni močnostni vrh (spectral strong peak), GFCC (gammatone feature cepstrum coefficients) in podobne. MFCC (mel-frequency cepstral coefficients) [15] je koeficient kratkotrajnega močnostnega spektra zvoka. Temelji na linearni kosinusni transformaciji logaritmičnega močnostnega spektra na nelinearni mel frekvenčni skali. Spectral strong peak [14] nam pove ali spekter vključuje izrazit vrh. Vrednost je večja, če je vrh višji in tanjši. HFC je izračunan na podlagi spektra in označuje količino visokih frekvenc v signalu [6].
- **Časovne značilnice** se nanašajo na časovne spremembe v signalu in trajanje nekega stanja v signalu. Efektivno trajanje (effective duration) nam pove koliko časa je bil signal nad določeno energijo. ZCR (zero-crossing rate) [13] je vrednost, ki nam pove kolikokrat je signal prešel iz pozitivne vrednosti v negativno in obratno. Leq, LARM, ludness, loudness so različne mere za glasnost.
- Med **tonskimi značilnicami** je pomembna tonaliteta [37], ki je skupina vseh karakteristik, ki povezujejo niz tonov ali akordov neke tonalne kompozicije okrog tonike, t.j. okrog središča tonalitete. Kosonanca (consonance)

[33] je definirana s frekvenčnimi razmerji. Nanaša se na akorde, harmonijo ali intervale, ki jih obravnavamo kot stabilne, ter nam izzovejo občutje ugodja. Ravno nasprotje kosonanci je disonanca (dissonance). Ta na poslušalca deluje neugodno in se odraža z negativno psihofizično reakcijo. Harmonični ritem (harmonic rhythm) [19] je razmerje, ki pove kako se akordi spreminjajo v sestavi glasbe. Poleg teh se uporablja tudi histogram akordov (chord histogram) in ostali.

- **Ritmične značilnice** Zelo pomembna značilnica v tem sklopu je število ritmičnih enot na minuto (BMP - beats per minute). To je značilnica, ki nam opisuje tempo v glasbi. Ker se v večini pesmi skozi čas BMP spreminja, je pomembna enota tudi BMP histogram. Uporablja se tudi BeatLoudnes, ki predstavlja glasnost signala v oknu okoli udarca.
- **Visoko nivojske značilnice v glasbi** so bolj abstraktne in med seboj zelo različne. To so značilnice, ki jih v večini ocenjujemo na podlagi ostalih značilnic opisanih zgoraj. Gre za značilnice kot so razpoloženje, žanri, barva glasu in ostale.

Tabela 2.2: Pogosto uporabljene značilnice za ocenjevanje čustev iz glasbe

Skupina	Značilnice
Spektralne značilnice	MFCC, HFC, kromatski vektor, spektralni močnostni vrh, GFCC
Časovne značilnice	efektivno trajanje, ZRC, leq, LARM, ludnessm, loudness
Tonske značilnice	tonaliteta, kosonanca, disonanca, harmonični ritem, histogram akordov
Ritmične značilnice	BPM, BPM histogram, beat loudness
Visoko nivojske značilnice	žanri, barva glasu, razpoloženje

Poglavje 3

Moodo podatkovna zbirka

Tema moje diplomske naloge je ocenjevanje razpoloženja v glasbi s pomočjo računalniških algoritmov. Dobrega algoritma za ocenjevanje ni možno narediti brez dobre podatkovne zbirke. Na področju razpoloženja obstaja že nekaj podatkovnih zbirk, vendar glede na naše informacije nobena ne vključuje dodatnih povezav, kot je povezava med glasbo in barvo ali podatkov o percepciji razpoloženja. Zaradi tega smo se odločili, da zberemo svojo podatkovno zbirko, ki bo osnova za raziskovanje povezave med razpoloženjem in glasbo. Vključevala bo tudi podatke o barvah, ki po mnenju anketirancev najbolje opisujejo posamezno pesem, uporabnikovo percepcijo povezave med glasbo, barvami in razpoloženjem ter nekatere demografske podatke.

V tem poglavju bomo opisali, kako smo zbirali podatke, namenil nekaj besed preliminarne analizi, ki smo jo opravili na krajši anketi. Opisali bomo glavno anketo, s katero smo zbirali podatke, ki sestavljajo podatkovno zbirko. Za konec bomo še opisali, kaj sestavlja podatkovno zbirko in podatke analizirali.

3.1 Zbiranje podatkovne zbirke

Podatkovno zbirko smo zbirali s spletno anketo, ki smo jo implementirali sami. Pred oblikovanjem glavne ankete, smo morali sprejeti še nekaj odločitev o tem, kako zgraditi anketo, da bo dala dobre in smiselne rezultate. Prva stvar, kjer je bil potreben premislek, je izbor razpoloženskih oznak. Ugotovili smo, da obstajajo nekatere osnovne oznake za razpoloženja, na primer [9]. Zaradi subjektivne narave

področja ni standardnega seta oznak, ki bi se uporabljale na področju povezanem z razpoloženjem in glasbo. Nekateri avtorji so izbrali zbirko oznak čisto intuitivno, na primer Wu et al. [36]. V izogib subjektivnosti smo se odločili, da naredimo preliminarno raziskavo v obliki ankete, v kateri smo preverjali smiselnost oznak.

3.1.1 Preliminarna analiza

Kot omenjeno, smo preliminarno analizo izvajali s pomočjo ankete, ki smo jo izvedli v elektronski obliki. Želeli smo preveriti smiselnost oznak, primernost elementov ankete in ugotoviti, katere oznake za razpoloženja so tista, ki bodo uporabljena v glavni anketi. Razpoloženja smo izbirali tako, da je moral uporabnik za 46 razpoloženskih oznak označiti, v kolikšni meri je neko razpoloženje pri njem prisotno v tistem trenutku. Uporabnik je to označil na skali od 1 do 7, prikazani na sliki 3.1. Iz tega seznama smo v nadaljevanju izbrali 17 osnovnih oznak uporabljenih v glavni anketi: aktivno, budno, dremavo, neaktivno, nesrečno, nezadovoljno, razočarano, sproščeno, srečno, utrujeno, vedro, veselo, zadovoljno, zaspano, žalostno, mirno in ježno.



Slika 3.1: Del sedem stopenjske lestvice, s pomočjo katere so anketiranci ocenjevali trenutno razpoloženje. Oznaka 1 pomeni, da se tisto razpoloženje pri anketirancu sploh ne pojavi, 7 pa da je zelo izrazito.

Zanimala nas je tudi struktura vprašalnika, ki je ostala približno enaka v glavnem vprašalniku, s to razliko, da smo tam dodali del z glasbenimi odlomki. Več o

tem v poglavju 3.1.2.

Poleg tega smo testirali tudi elemente uporabljene v anketi (sedem stopenjska skala, neskončni barvni krog, izbira z radio gumbi in tekstovnimi polji). Ugotovili smo, da moramo nekatere elemente spremeniti. Spremenili smo barvno skalo, saj smo jo omejili na barvni krog z 49 možnostmi izbire. To je bilo potrebno, ker je imel uporabnik na neskončni skali preveliko možnost izbiranja, obenem pa je bil sistem tri dimenzionalen, kar večina uporabnikov ni opazila in so nastavljali samo odtenek barve, na svetlost in nasičenost pa so pozabili. Skala z 49 možnostmi (prikazana na sliki 3.2) se je izkazala kot boljša alternativa, saj še vedno ponuja veliko možnosti izbire barv, je prijazna uporabniku in pridobljeni podatki so boljši.

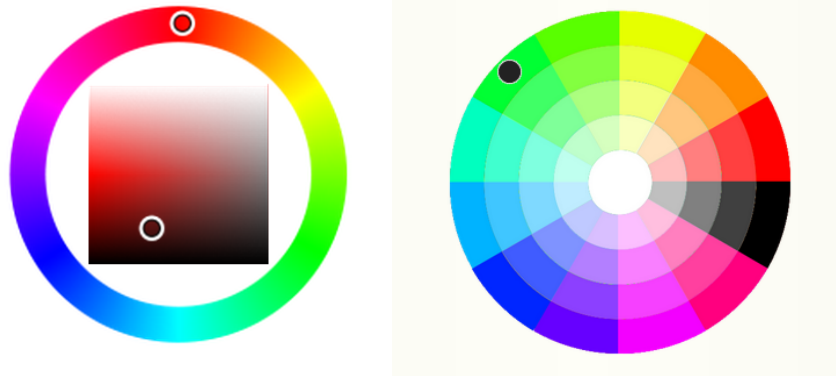
Poleg zamenjave barvnega kroga smo zamenjali tudi nekaj ostalih elementov. V delu, kjer uporabnik označi tri svoje najljubše žanre, smo se odločili namesto vpisnih polj ponuditi uporabniku seznam 20 žanrov, iz katerega je izbral in potegnil svoje tri izbire na nov seznam. Za to smo se odločili, ker so uporabniki v preliminarni analizi v prostor vpisovali tudi žanre, ki niso osnovni in oznake, ki sploh niso žanri. Prav tako nam je preliminarni vprašalnik pomagal izbrati, kateri žanri so tisti, ki jih bomo uporabniku ponudili. Zamenjali smo tudi način, kako uporabnik vnese svoje trenutno razpoloženje in za ta namen uporabili nov element MoodGraph, ki ga bom opisal v poglavju 3.1.2.

3.1.2 Glavna anketa

Glavni vprašalnik, s katerim smo zbrali našo podatkovno zbirko, je bil implementacija preliminarne analize, saj smo v skladu z zbranimi podatki prilagodili oznake in strukturo vprašalnika.

Glavni vprašalnik je bil sestavljen iz treh delov. V prvem delu smo spraševali o uporabnikovih navadah, o poslušanju glasbe in njegovi glasbeni izobrazbi. V drugem delu nas je zanimala predvsem uporabnikova percepcija razpoloženja, glasbe in barv. V tretjem delu so morali uporabniki označiti razpoloženje in barve v odlomku glasbe.

Prvi del V tem delu smo spraševali o udeleženci starosti, spolu in o tem, ali živi na podeželju, ali v mestu. Uporabnika smo vprašali tudi, če je pod vplivom drog ali drugih substanc. Zanimali so nas podatki o glasbeni izobrazbi in o tem,



Slika 3.2: Barvni krog na levi smo uporabili v preliminarni anketi. V zunanjem krogu je bilo možno izbrati odtenek (hue) v notranjem pa je bilo možno izbirati nasičenost in svetlost. Desni krog smo uporabili v glavni anketi. Krog je omogoča izbiro med 49 barvami. Rezina predstavlja en barvni odtenek, proti notranjosti pa se spreminja svetlost in intenzivnost.

če udeleženec igra inštrument ali poje, kljub temu da ne obiskuje več glasbene šole ali pa je ni nikoli obiskoval. Poleg tega nas je zanimalo še, koliko časa dnevno uporabnik posluša glasbo in kateri so njegovi najljubši žanri. Izbrati je moral tri svoje najljubše žanre in jih razporediti po priljubljenosti. To smo zajemali s pomočjo elementa prikazanega na sliki 3.3, kjer je uporabnik izmed nabora 20 žanrov izbral najljubše in jih povlekel v stolpec desno ter razporedil glede na priljubljenost.

Drugi del Drugi del je namenjen ocenjevanju anketirančevega trenutnega razpoloženja, njegove percepcije oznak za razpoloženje in barve za razpoloženje. Te podatke zajemamo zato, ker nam omogočajo, da ugotovimo povezave med različnimi ocenami razpoloženja v tretjem delu, ki vključuje glasbene odlomke.

Na začetku smo udeleženca vprašali, kako bi svoje razpoloženje označil s točko v prostoru, ki ga označujeta prijetnost in aktivnost (valence-arousal space) [8]. To je dvodimenzionalen prostor, kjer na vodoravni osi od leve proti desni narašča prijetnost in od spodaj navzgor aktivnost. Svoje razpoloženje je moral udeleženec opisati tudi z izbiro barve s pomočjo barvnega kroga predstavljenega v poglavju 3.1.1.



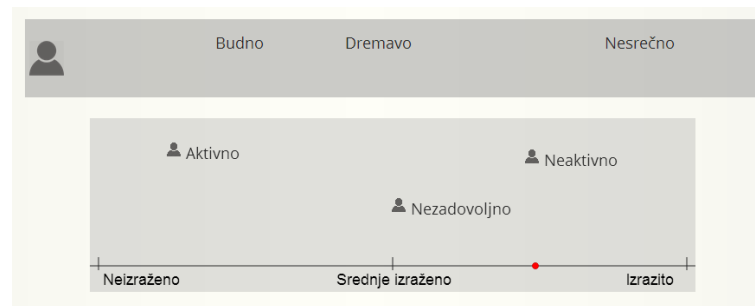
Slika 3.3: Element v anketi uporabljen za izbiro najljubših žanrov

Udeleženec je moral svoje razpoloženje opisati tudi s tem, da je označil, kako je posamezno razpoloženje pri njem izraženo v trenutku reševanja ankete. To smo zajemali s pomočjo novega elementa imenovanega MoodStripe. Potrebno je bilo potegniti posamezne oznake razpoloženj v prostor, kjer od leve proti desni narašča prisotnost posameznega razpoloženja. Če je anketiranec postavil razpoloženje skrajno levo to pomeni, da to razpoloženje pri njem ni prisotno, če pa ga je postavil skrajno desno to pomeni, da je razpoloženje zelo prisotno.

V drugem delu smo uporabnike povprašali tudi o njihovi percepciji posameznega razpoloženja. Udeleženec je za vsako razpoložensko oznako v elementu MoodGraph (slika 3.5) označil, kako prijetno in aktivno je posamezno razpoloženje. To je dvodimenzionalen prostor, kjer je na vodoravni osi prijetnost in na navpični osi aktivnost. Uporabnik je oznake razpoloženj prikazane nad grafom potegnil v to ravnino na, mesto za katerega je mislil, da najbolje opisuje razpoloženje.

Poleg tega, kako si uporabnik predstavlja posamezno razpoloženje v VA prostoru, nas je zanimalo tudi, kako bi uporabnik opisal ista razpoloženja z barvo. To smo izvedli s pomočjo barvnega kroga (slika 3.2), ki sem ga že opisal zgoraj.

Tretji del V tretjem delu smo želeli, da udeleženci za 10 glasbenih odlomkov dolgih 15 sekund označijo razpoloženje, ki ga zaznavajo pri glasbi (perceived) in tistega, ki občutijo (induced) ter razpoloženje opišejo z barvo. Odlomki so bili

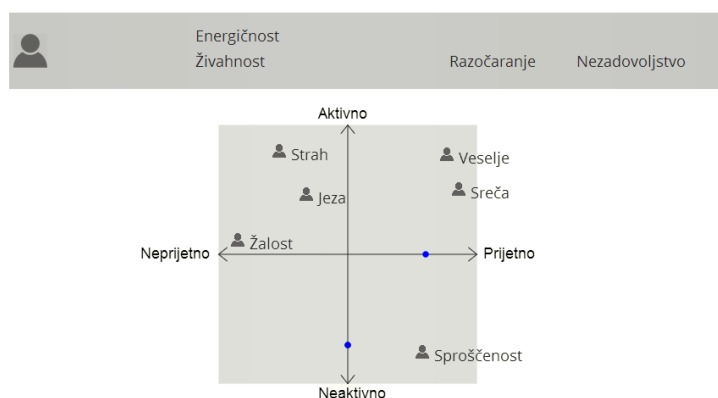


Slika 3.4: Element za zajemanje pristnosti posameznega razpoloženja imenovan moodstripe. Anketiranec je iz zgornjega pravokotnika povlekel razpoloženja in jih vstavil na ustrezno mesto na traku glede na prisotnost posameznega razpoloženja. Prisotnost narašča iz leve (razpoloženje ni izrazito) proti desni (razpoloženje je zelo izrazito).

naključno izbrani iz nabora 200 odlomkov. Ker smo se želeli izogniti pristranskosti uporabnika zaradi poznavanja določene glasbe ali dogodkov, ki so se mu zgodili ob poslušanju določene pesmi, smo odlomke izbrali iz zbirk, ki vsebujejo uporabnikom nepoznano glasbo (Jamendo, zbirka s filmsko glasbo, zbirka slovenskih ljudskih pesmi in zbirka moderne elektro-akustične glasbe).

Glasba je bila izbrana iz štirih različnih virov. Iz odprte podatkovne zbirke Jamendo smo izbrali 80 pesmi, ki so bile iz različnih žanrov in dokaj vsakdanje. Naslednjih 80 pesmi smo vzeli iz zbirke filmske glasbe opisane v [10]. Dodali smo še 20 slovenskih ljudskih pesmi in 20 pesmi iz nabora moderne elektro-akustične glasbe. Uporabnikov nismo želeli preveč obremeniti, zato je bilo vsakemu uporabniku naključno dodeljenih zgolj 10 odlomkov.

Pri vsakem odlomku je udeleženec označil, katera razpoloženja so izražena v odlomku, katera razpoloženja posamezna glasba vzbudi v njem in s katero barvo iz barvnega kroga (slika 3.2) bi posamezno pesem opisal. Razpoloženja izražena v odlomku je izbral iz nabora 14 oznak, razpoloženja, ki jih je glasba pri njem vzbudila, pa iz nabora 10 oznak. Udeleženec je lahko za posamezen odlomek izbral več oznak iz obeh naborov. Poleg tega, da je izbral posamezno oznako jo je moral še uvrstiti v VA prostor. S tem smo zajeli tudi podatek kako si predstavlja določeno oznako pri posamezni pesmi. Na primer, pri neki pesmi je lahko veselje zelo aktivno, pri drugi pa manj. Poleg tega nam ta podatek omogoča, da razi-



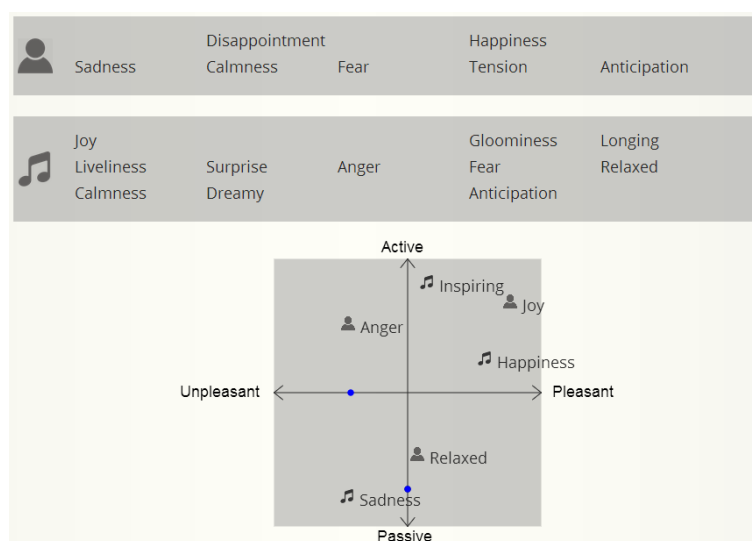
Slika 3.5: Enonivojski MoodGraph, element za zajemanje precepcije razpoloženja v VA prostoru. Anketiranec je razpoloženijske oznake uvrstil v graf glede na prijetnost in aktivnost. Prijetnost narašča iz leve proti desni. Aktivnost narašča od spodaj navzgor.

skujemo tudi, kako si uporabnik v VA prostoru predstavlja posamezno pesem. Za zajemanje tega podatka smo uporabili dvokategorni MoodGraph prikazan na sliki 3.6.

3.2 Sestava podatkovne zbirke

Struktura podatkovne zbirke sledi strukturi uporabljene ankete, zato je razdeljena v tri dele. V prvem delu smo zbrali 1423 odgovorov in s tem tudi toliko vpisov v našo zbirko. V drugem delu je vpisov 1090. V tretjem delu, kjer je moral anketiranec oceniti 10 pesmi, smo zbrali 7187 odgovorov. Ta del je izpolnjevalo 741 anketirancev.

V prvem delu so podatki, ki opisujejo udeležence (podrobnosti so predstavljene v tabeli 3.1). Za vsakega udeleženca zbirka vsebuje podatek o starosti na leto natančno, o spolu in o tem ali živi na podeželju ali v mestu. Poleg tega smo zbrali podatke o tem, koliko časa se ukvarja z glasbo, koliko časa je hodil v glasbeno šolo do leta natančno ter koliko časa na dan posluša glasbo. Tukaj je uporabnik poslušanje glasbe uvrstil v eno od kategorij: do 1 ure, od 1 do 2 uri, od 2 do 3 ure in več kot 3 ure. Imamo tudi podatek o najljubših glasbenih žanrih udeleženca.



Slika 3.6: Dvonivojski MoodGraph, element za zajemanje razpoloženja v glasbi in njegovo umestitev v VA prostor. Nota označuje nabor razpoloženj, ki jih udeleženec zazna v glasbi (perceived). Figura človeka označuje nabor razpoloženj, ki jih udeleženec občuti (induced). Udeleženec je iz vsake skupine razpoloženj izbral tista, ki po njegovem izražene ali vzbujene in jih umestil na v VA prostor glede na prijetnost in aktivnost.

Udeleženec je podal najmanj en in največ tri žanre. Zanimalo nas je še psihofizično stanje anketiranca med reševanjem. Torej imamo podatek ali jemlje zdravila, ki vplivajo na razpoloženje ter, če je bil v trenutku reševanja pod vplivom drog ali alkohola.

Drugi del podatkovne zbirke vsebuje podatke o udeležencem razpoloženju v trenutku, ko je izpolnjeval anketo in o tem kako si anketiranec predstavlja posamezna razpoloženja. Uporabnikovo razpoloženje opisujejo trije podatki. Prvi je točka v VA prostoru (x in y koordinata). Drugi je barva v barvnem krogu (tukaj hranimo podatek o tem, katero barvo je udeleženec izbral). Tretji pa je vrednost, kako močno je posamezno razpoloženje iz nabora 17 razpoloženj, izraženo pri udeležencu v tistem trenutku. Hranimo vrednost med 0 in 1 za vsako razpoloženje. Kot smo že omenili, drugi del vsebuje tudi podatek o tem, kako si udeleženec predstavlja razpoloženja. Za nabor 10 razpoloženj imamo podatek, kam v VA prostoru spada to razpoloženje po mnenju udeleženca in barvo, s katero ga je anketiranec označil.

Tabela 3.1: Vprašanja prvega dela ankete. Tabela poleg vprašanj vsebuje še podatke o možnih odgovorih in komentar. V prvem delu ankete smo pridobivali demografske podatke in podatke o uporabnikovih izkušnjah z glasbo.

Vprašanje	Odgovori	Komentar
Starost	[5, 99]	izražena v letih
Spol	{Moški, Ženski}	
Področje prebivanja	{mesto, podeželje}	
Leta glasbenega izobraževanja	[0, 20]	v letih, 0 - pomeni, da se ni glasbeno izobraževal
Igranje instrumenta ali petje	[0, 20]	v letih, 0 - pomeni, da se nikoli ni ukvarjal z glasbo
Uživanje drog	{da, ne}	
Vpliv drog	{da, ne}	Je bil anketiranec pod vplivom drog med reševanjem ankete?
Najljubši žanri	{classical, opera, country, folk, latin, dance/disco, electronic, RnB/soul, hip hop/rap, reggae, pop, rock, alternative, metal, blues, jazz, vocal, easy listening, new age, punk}	Anketiranec je moral izbrati do tri žanre (najmanj enega) in jih razvrstiti od najljubšega do najmanj priljubljenega.
Čas poslušanja glasbe	{manj kot 1, 1-2, 2-3, več kot 3}	v urah na dan

Tretji del podatkovne zbirke je malo drugačen. Tu nimamo le enega odgovora na od posameznega udeleženca, ampak do 10 odgovorov, za vsako pesem en odgovor. Imamo podatek o tem, katera razpoloženja so izražena v glasbi in katera razpoloženja glasba vzbudi pri anketirancu. Izbrana razpoloženja je udeleženec umestil v VA prostor, zato imamo tudi podatek, kam v tem prostoru se umeščajo. Za vsako pesem imamo tudi podatek o barvi, s katero je udeleženec označil glasbeni odlomek.

Tabela 3.2: Vprašanja drugega dela ankete. Tabela vsebuje vprašanja, možne odgovore in komentar. Vprašanja sprašujejo po udeležencevem razpoloženju, percepciji razpoloženja in povezavi med barvami in razpoloženjem.

Vprašanje	Odgovori	Komentar
Trenutno razpoloženje v VA prostoru	VA prostor	Uporabnik je označil razpoloženje v VA prostoru, predstavljenem v poglavju 3.1.2.
Barva razpoloženja	Barvni krog	Uporabnik je izbral barvo trenutnega razpoloženja z barvnim krogom predstavljeni v poglavju 3.1.1.
Percepcija razpoloženja	{strah, energičnost, jeza, sproščenost, sreča, žalost, živahnost, veselje, razočaranje, nezadovoljstvo}	Uporabniki so umestili razpoloženja v VA prostor z uporabo Mood-Grapha predstavljenega v poglavju 3.1.2
Trenutno razpoloženje z oznakami	{aktivno, budno, dremano, neaktivno, nesrečno, nezadovoljno, razočarano, sproščeno, srečno, utrujeno, vedro, veselo, zadovoljno, žalostno, mirno, jezno}	Uporabnik je razvrstil vse oznake na Mood-Stripe (predstavljen v poglavju 3.1.2), glede na prisotnost od neizraženega do zelo izraženega.
Barva razpoloženja	{sreča, žalost, nezadovoljstvo, energičnost, razočaranje, veselje, sproščenost, živahnost, strah, jeza}	Za vsako razpoloženje je uporabnik izbral barvo, ki ga najbolj opisuje z uporabo barvnega kroga.

3.3 Analiza podatkov

3.3.1 Demografska analiza

Povprečna starost anketirancev je 26,5 let. Najmanjši anketiranec je bil star 15 let, najstarejši pa 64 let. Največ udeležencev je bilo starih 20 let. Anketo je rešilo več udeleženk ženskega spola, kar 65%. Večina udeležencev prihaja iz urbanega okolja, 66% anketirancev živi v mestu in 34% na podeželju.

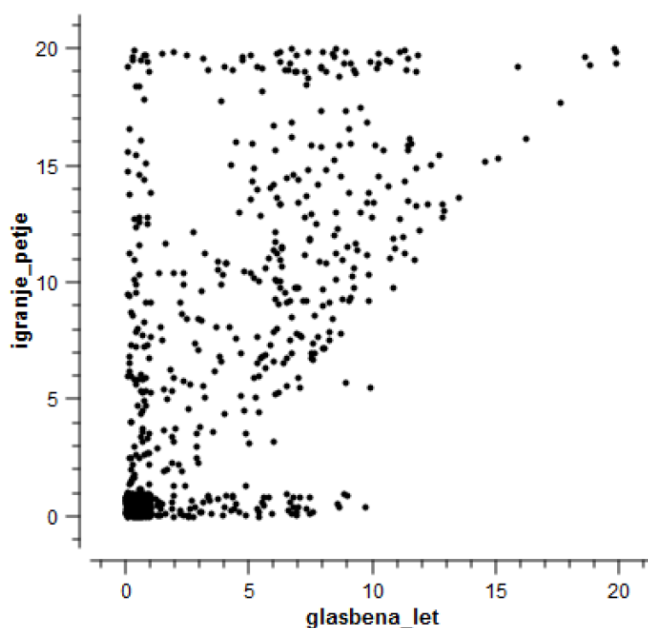
Ukvarjanje z glasbo Zanimivo je, da ima kar polovica vseh anketirancev glasbeno izobrazbo, z glasbo pa se ukvarja 53% anketirancev (igrajo inštrument ali pojejo). Udeleženci, ki so obiskovali glasbeno šolo, so jo v povprečju obiskovali 6 let. Tisti, ki se z glasbo ukvarjajo neformalno se z njo ukvarjajo povprečno 11 let. Pričakovano igranje inštrumenta in petje ter leta glasbene šole zelo visoko korelirata ($r=0,653$). Iz slike 3.7 poleg korelacije razberemo tudi, da se večina ljudi z glasbo ukvarja tudi po dokončani formalni izobrazbi, kar sklepamo iz podatka, da se večina ljudi z glasbo ukvarja dlje, kot traja zgolj formalna izobrazba.

Največ anketirancev posluša glasbo 1 do 2 uri na dan, najmanj pa 2 do 3 ure na dan. Količina poslušanja glasbe ne korelira z nobeno od ostalih demografskih spremenljivk.

Zdravila in droge Zdravila, ki vplivajo na razpoloženje, uživa 4,2% udeležencev. Kar 3% udeležencev je poročalo, da so bili med reševanjem ankete pod vplivom drog ali alkohola. Tudi ta dva parametra visoko korelirata med seboj ($r = 0,519$).

Žanri Kot je razvidno iz grafa na sliki 3.8, je večina udeležencev kot najljubši žanr označila rock (31,6% anketirancev). Na drugem mestu med najljubšimi žanri je pop z 16,6% in na tretjem alternativna glasba. Na zadnjih dveh mestih sta opera in reggae, ki sta najljubši žanr manj kot odstotku vprašanih. Razporeditev žanrov, ki so jih uporabniki uvrstili na prvo mesto, ni enakomerna, saj kot največkrat izbrana žanra zelo izstopata rock in pop.

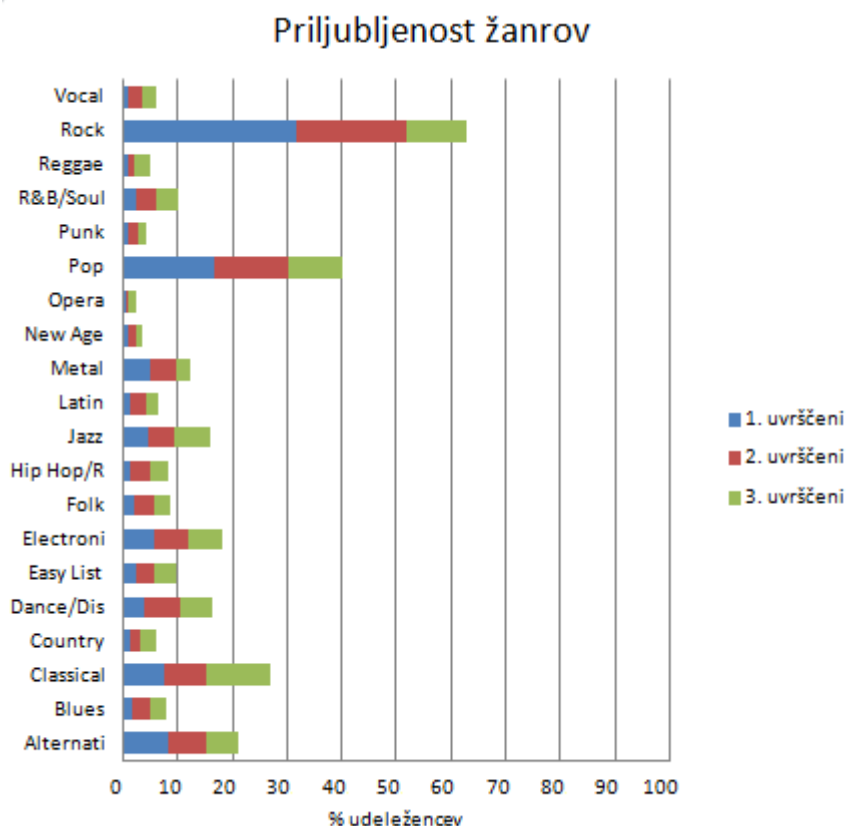
Na drugo mesto je spet največ vprašanih uvrstilo rock (20% vprašanih), drugi največkrat uvrščen žanr na drugo mesto je pop z 13,9%. Najmanjkrat izbran žanr je opera. Na drugem mestu so izbire že bolj raznolike kot pri najljubšem žanru, vendar vrh še vedno izstopa.



Slika 3.7: Vsaka točka na sliki prikazuje povezavo med trajanjem formalnega glasbenega izobraževanja na vodoravni osi in leta ukvarjanja z glasbo (igranje inštrumenta ali petje) na navpični osi. Iz slike je razvidno, da se večina anketirancev tudi po končani formalni izobrazbi ukvarja z glasbo.

Zanimivo je, da je največ udeležencev na tretje mesto uvrstilo klasično glasbo. Za to izbiro se je odločilo 11,9% anketirancev. Klasiki sledi rock (11,3%). Najredkeje so udeleženci na tretje mesto uvrstili new age in punk. Za izbire na tretjem mestu je značilno, da so bolj raznolike, saj noben žanr ne izstopa.

Izmed vseh žanrov smo potem izbrali najbolj priljubljene: rock, pop, country, alternativna glasba in klasična glasba. Naredili smo primerjavo med najljubšimi žanri po starostnih skupinah (slika 3.9). Anketirance smo razdelili v 6 starostnih skupin. Vsaka od skupin zajema 10 let. Pri vseh starostnih skupinah je popularna rock glasba. Pri najmlajših štirih starostnih skupinah je na prvem mestu. Tudi pop glasba se pojavi pri vseh skupinah razen najstarejših dveh. Razlike se pojavijo v ostalih glasbenih zvrsteh. Pri mlajših so popularne še zvrsti kot so metal, alternative in elektronska glasba. Pri starejših je zelo popularna klasična glasba. V najstarejših dveh skupinah se pojavi na prvem mestu. V teh skupinah so popularne še folk, jazz in country. Srednji dve starostni skupini pa sta kombinacija



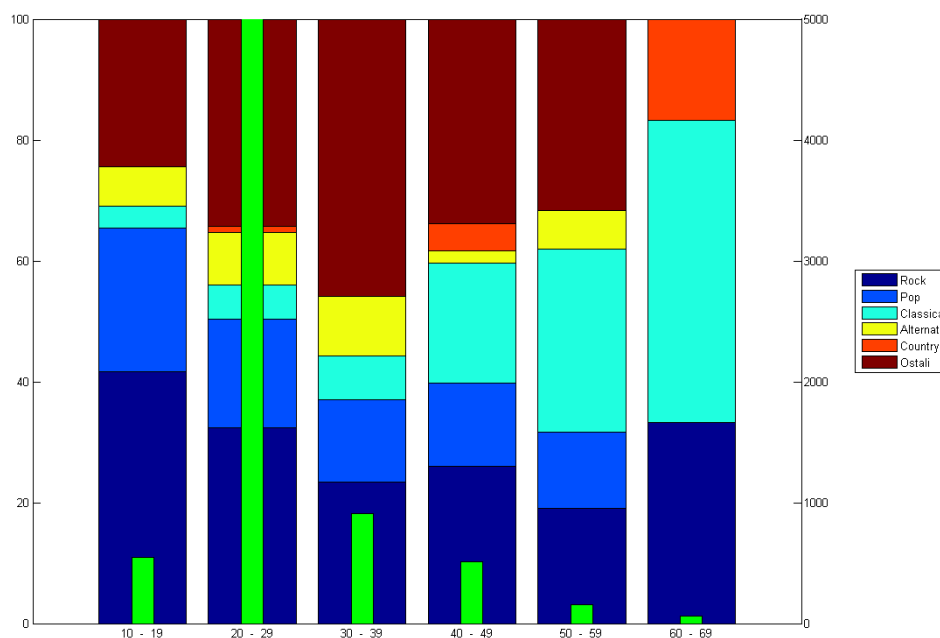
Slika 3.8: Priljubljenost žanrov ter uvrstitev teh na prvo, drugo ali tretje mesto. Razvidno je, da je najbolj priljubljen žanr rock, sledi mu pop in klasična glasba. Najmanj popularen žanr je opera.

oboje. V njih se pojavi še alternativna glasba in tudi jazz ter klasična glasba.

Naredili smo tudi analizo priljubljenosti glede na spol. Ugotovili smo, da je pri obeh spolih enako zastopana rock glasba. S tem, da pri udeležencih prevladuje še elektronska glasba in metal. Pri udeleženkah pa je na prvem mestu pop glasba.

Za konec še nekaj zanimivosti:

- Podatki iz ankete kažejo, da osebe, ki se z glasbo ukvarjajo več kot 15 let, najraje poslušajo klasično glasbo.
- Osebe, ki jemljejo zdravila ne poslušajo klasične glasbe.
- Po podatkih iz ankete je med ljudmi, ki uživajo droge bolj popularna elek-



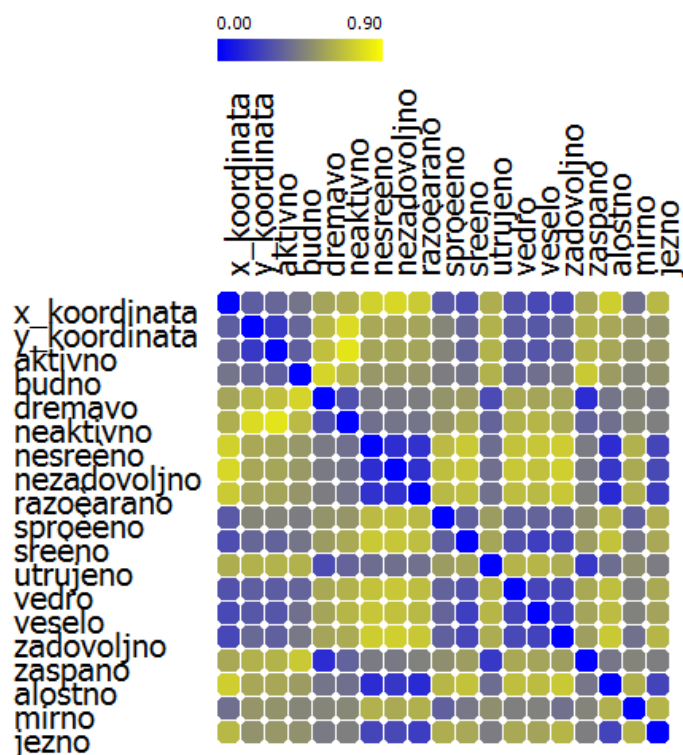
Slika 3.9: Priljubljenost posameznih glasbenih zvrsti po starostnih skupinah. Na vodoravni osi so označene starostne skupine. Širši stolpci predstavljajo razporeditev posameznih žanrov v odstotkih, zeleni pa število anketirancev v posamezni starostni skupini.

tronska glasba, kot pri ostali populaciji.

3.3.2 Razpoloženje, čustva in barva

Za ta del analize smo iz podatkovne zbirke odstranili vse, ki so bili pod vplivom drog ali zdravil, ki vplivajo na razpoloženje. V anketi smo anketirančev trenutno razpoloženje zajemali na dva načina. Najprej je anketiranec moral označiti svoje razpoloženje na VA prostoru. Torej je označeval prijetnost in aktivnost. Sledila so tudi vprašanja kjer je anketiranec trenutno razpoloženje ocenjeval z tem, da je povedal kako zelo prisotno je posamezno razpoloženje pri njem. Zanimala nas je predvsem korelacija med prisotnostjo posameznih razpoloženj in med koordinatama v VA prostoru. Kot je razvidno iz slike 3.10 je prijetnost (koordinata

x) najmanj oddaljena od razpoloženj: sproščenost, sreče, zadovoljstvo, vedrost in veselje. Koordinata y pa po pričakovanjih najbolj korelira z razpoloženjem aktivnost. Z koordinata x se najmanj ujema s razpoloženjem nesrečno, nezadovoljno in razočarano, kar je bilo pričakovati, saj je prijetnost (predstavlja jo koordinata x) ravno nasprotna s temi razpoloženji. Z koordinato y se pričakovano najmanj ujema razpoloženjem neaktivno zaradi enakega razloga kot pri koordinati x. Med razpoloženji lahko opazimo še dobro ujemanje med razpoloženji nezadovoljno in razočarano ter med razpoloženji veselo in zadovoljno.

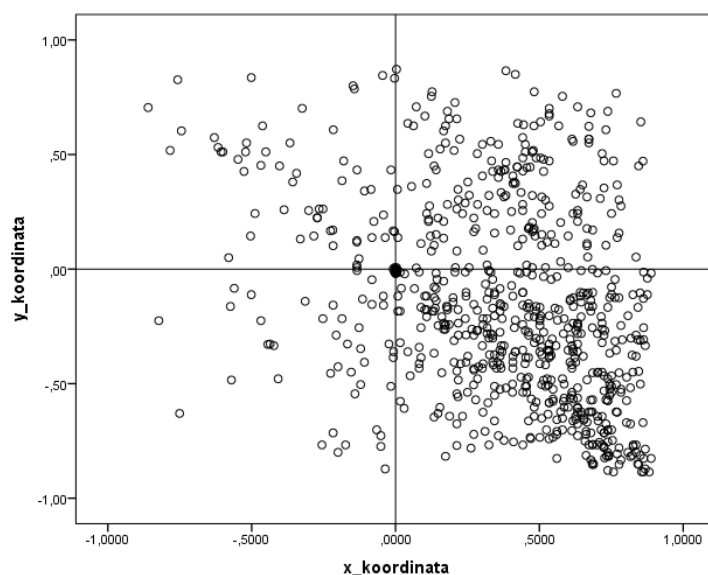


Slika 3.10: Povprečna evklidska razdalja med posamezni atributi. Iz slike je razvidno, da prijetnost najbolj korelira z razpoloženji: sproščenost, sreča, zadovoljstvo, vedrost in veselje. Aktivnost (koordinata y) se najbolj ujema z razpoloženji aktivno, vedro, veselo in zadovoljno. Opazimo lahko tudi ujemanje med razpoloženji nezadovoljno in razočarano ter veselo in zadovoljno.

Slika 3.11 prikazuje kako so ljudje uvrstili svoje razpoloženje v VA prostor.

Opazno je, da so se anketiranci v večjem številu uvrstili v desno polovico koordinatnega sistema, kar pomeni, da so svoje razpoloženje označili kot bolj prijetno. Več anketirancev občuti bolj neaktivno razpoloženje, kar se odraža v tem, da je več udeležencev svoje razpoloženje uvrstilo v spodnji del koordinatnega sistema. Največja gostota anketirancev se nahaja v spodnjem desnem kvadrantu, kar pomeni, da se počutijo bolj prijetno in so bolj neaktivni.

Zelo zanimivo je, da s koordinato y korelira starost in tudi poslušanje glasbe. Tisti, ki so starejši, so označili, da se počutijo manj aktivne in obenem manj poslušajo glasbo.



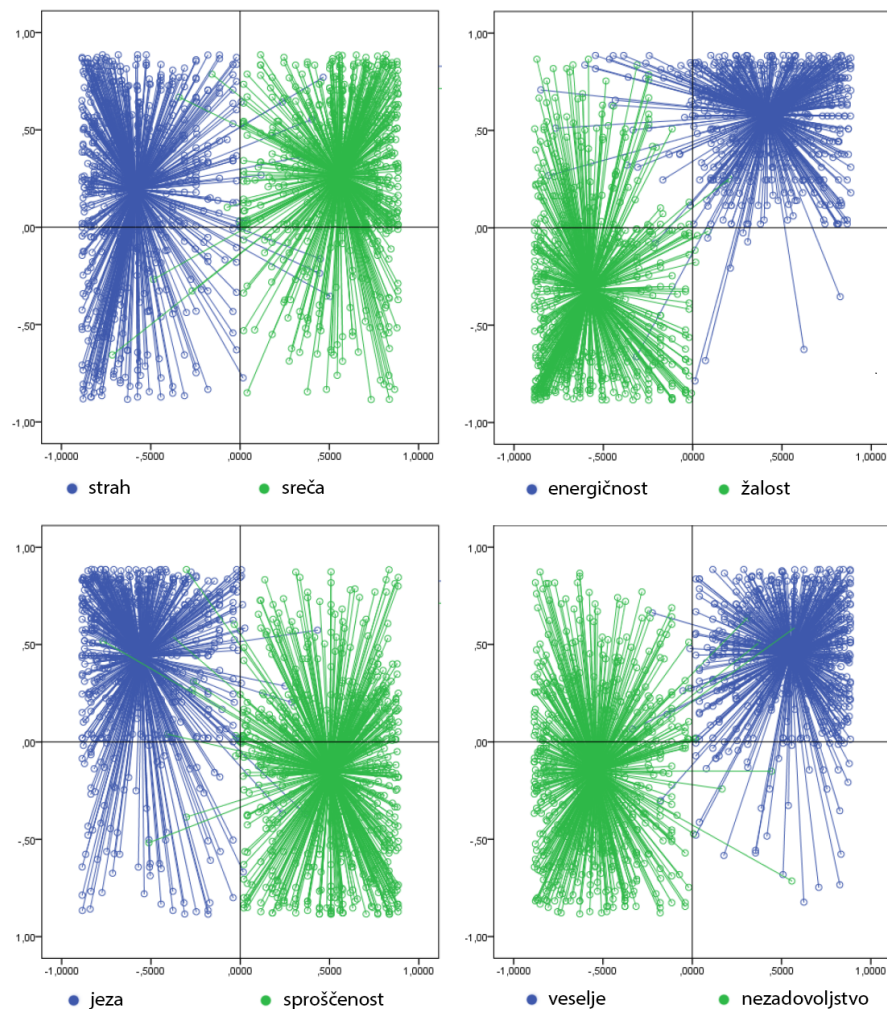
Slika 3.11: Slika prikazuje odgovore vseh uporabnikov o razpoloženju v VA prostoru. Vidi se, da je bilo največ anketirancev v času reševanja ankete prijetnega in bolj neaktivnega razpoloženja.

V anketi smo povpraševali tudi o tem kako uporabniki dojemajo posamezno razpoloženje. Razpoloženje so morali opisati tako, da so posamezno oznako postavili v VA prostor. Na sliki 3.12 je prikazana umestitev 8 razpoloženj v VA prostor.

Slika levo zgoraj prikazuje primerjavo med razpoloženjema strah in sreča. Strah je zelo neprijetno razpoloženje, sreča pa prijetno. Oba sta razporejena praktično

preko celotne y osi, kar si lahko razlagamo s tem, da sta to razpoloženji, za katera je težko določiti ali sta aktivna ali ne.

Iz slike desno zgoraj je po pričakovanjih razvidno, da je energičnost zelo prijetno in aktivno razpoloženje, obenem pa je žalost neprijetna in neaktivna. Zanimivo je, da je kar nekaj uporabnikov označilo, da je žalost aktivno razpoloženje.



Slika 3.12: V zgornjih valence-arousal prostorih je prikazana razporeditev 8 razpoloženj, glede na to kako so jih v ta prostor uvrstili anketiranci. V splošnem je opazno, da so se anketiranci glede prijetnosti odločili dokaj enotno, medtem ko so večje razlike v aktivnosti.

Na tretjem grafu (levo spodaj) je narejena primerjava med jezo in sproščenostjo. Jeza je po pričakovanjih zelo aktivno in negativno razpoloženje, čeprav je zanimivo, da nekaj anketirancev meni, da je neaktivna. Med tem je sproščenost prijetna in po aktivnosti razporejena preko celotnega prostora, čeprav prevladuje mnenje, da je bolj neaktivna. Tukaj je tudi veliko odgovorov, ki so nevtralni.

Iz četrte slike se vidi, da je veselje zelo prijetno in aktivno, medtem ko je nezadovoljstvo neprijetno. Po aktivnosti je nezadovoljstvo malo bolj razpršeno, kar govori o tem, da je ta parameter pri tem razpoloženju težje določiti.

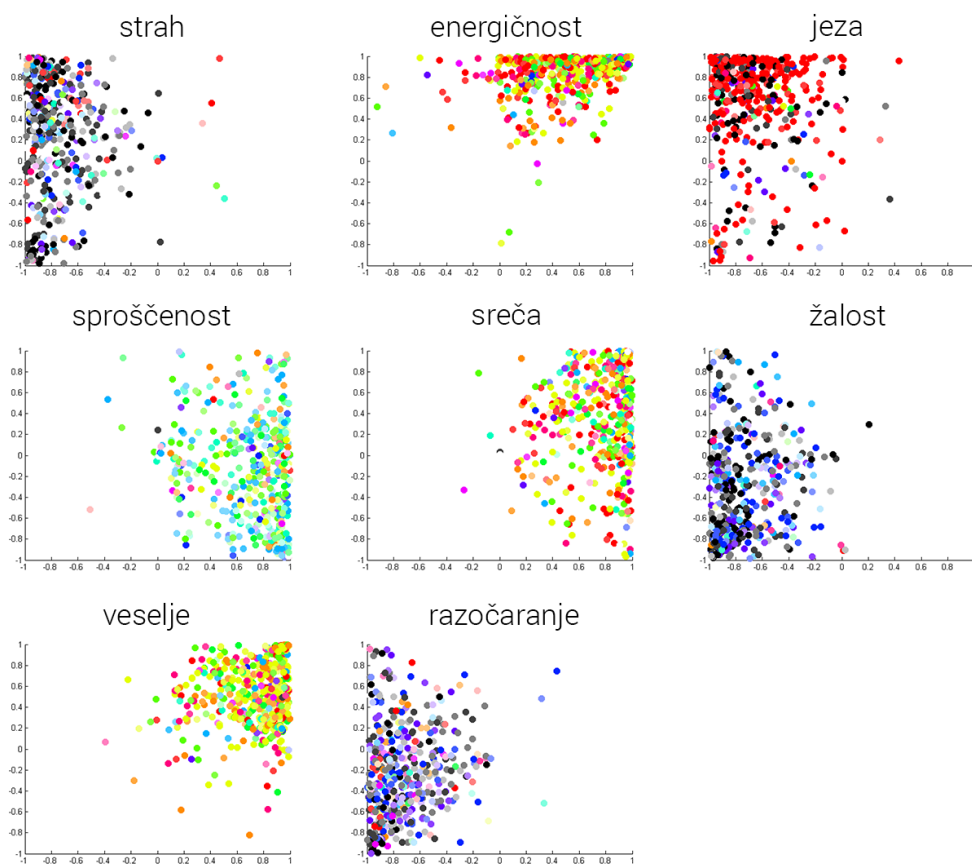
Barva Ugotovili smo, da obstaja zelo dobra korelacija med razpoloženjem in pozicijo razpoloženja v VA prostoru, kar je možno opaziti na sliki 3.13. Pri prijetnih razpoloženjih se barva v večini primerih giblje med rdečo in nežno modro. Pri neprijetnih razpoloženjih so odtenki večinoma modri, črni, rdeči in vijolični. Barve so tu zelo intenzivne.

Iz slike 3.13 je razvidno da so bili pri strahu pogosto izbrani bolj sivi in črni odtenki. Veselje je bilo v večini primerov označeno z bolj intenzivnimi toni. Odtenki pa so bili v tem primeru bolj rdeči, oranžni, rumeni ter svetlo zeleni. Pri energičnosti se je večina uporabnikov odločila za rdečo, zeleno in rumeno barvo. Za sproščenost je zanimivo, da je dokaj enakomerno razporejena preko celotnega barvnega spektra. Torej si to razpoloženje vsak udeleženec predstavlja drugače. Intenzivnost barv je pri sproščenosti manjša. Pri jezi je bila pogosto izbrana barva rdeča, modra ali vijolična. Izbirali so praktično med vsemi intenzivnostmi.

3.3.3 Glasba, razpoloženje in barva

Za konec smo izbrali še 4 glasbene odlomke in v sliki 3.14 prikazal rezultate. Predstavili smo VA vrednosti, kamor so anketiranci postavili razpoloženja, ki nastopajo v posamezni pesmi. Z krogci so označena mesta kamor so postavili razpoloženja, ki jih po njihovem mnenju glasba vzbudi pri človeku. Z kvadratki so predstavljena razpoloženja, ki jih pesem izraža. Obenem sem vizualiziral tudi barve, ki po mnenju uporabnikov opisujejo posamezno pesem.

Pri prvi pesmi nastopajo razpoloženja, ki so bolj pozitivna in aktivna (večina oznak se nahaja desno zgoraj). Tudi barve so glede na analizo barv zgoraj temu primerne, saj so svetle in niso tako intenzivne, kar se ujema z bolj aktivnimi



Slika 3.13: Izbira barv za posamezna razpoloženja in umestitev razpoloženj v VA prostor. Iz slike je razvidno, da so pri prijetnih razpoloženjih izbrane barve manj intenzivne, odtenki so v tem primeru večinoma rdeči, rumeni in zeleni. Pri neprijetnih razpoloženjih so izbrani bolj intenzivni in temnejši odtenki.

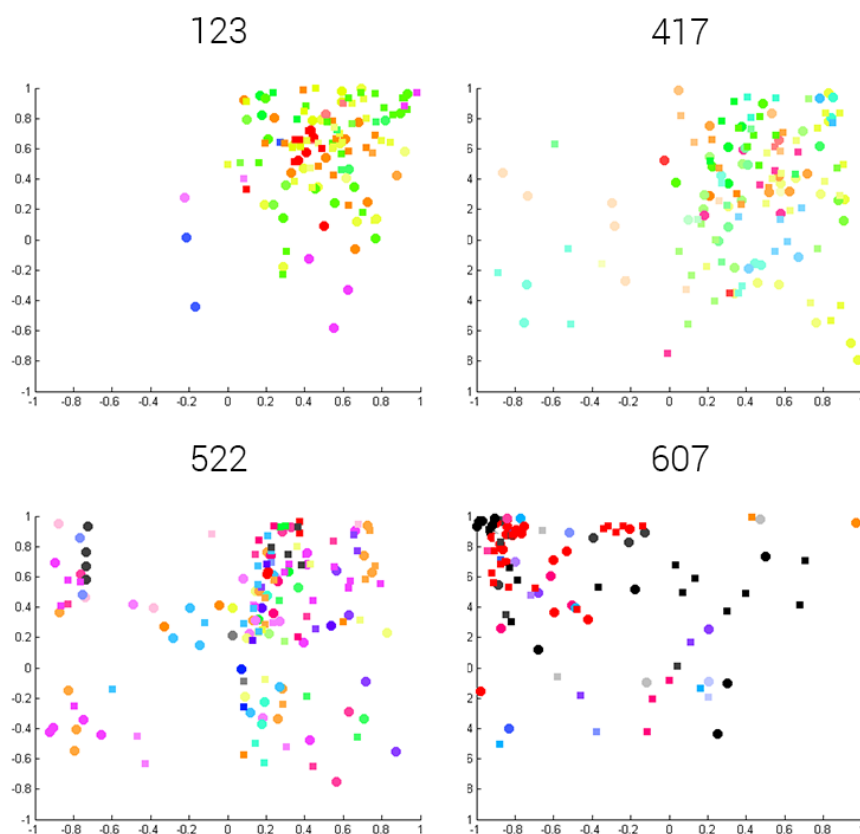
razpoloženji. Iz grafa se opazi, da so razpoloženja, ki jih pesem vzbudi malo bolj razpršena, kot tista, ki jih pesem izraža.

Pri drugi pesmi (desno zgoraj) so razpoloženja še vedno bolj prijetna in tudi malo bolj aktivna kot neaktivna. Tudi barve so enako kot pri prejšnji sliki temu primerne. V primerjavi s prejšnjo pesmijo pa so anketiranci veliko bolj neodločeni glede aktivnosti v pesmi. Tukaj so izražena razpoloženja manj razpršena in se v večini nahajajo v zgornjem desnem delu grafa za razliko od tistih, ki jih pesem

vzbuja.

Pri pesmi na grafu levo spodaj odločitve niso tako enotne. V večji meri prevladujejo prijetna in obenem aktivna razpoloženja. Je pa zanimivo, da se pojavijo 4 gruče odgovorov. Prva je skrajno levo zgoraj, druga levo spodaj, tretja največja je desno zgoraj in četrta desno spodaj. Zanimivo je, da so odločitve o izraženih razpoloženjih z nekaj izjemami bolj aktivne, tiste o vzbujenih pa se v večji meri pojavijo tudi v ostalih delih prostora.

Graf za četrto pesem (desno spodaj) pa je pravo nasprotje vsem ostalim. Notri nastopajo zelo neprijetna razpoloženja, ki pa so v večini primerov aktivna. Značilno je, da je večina ljudi prijetnost in tudi aktivnost označila na skrajnem robu prostora. V tem primeru so izražena razpoloženja bolj razpršena kot tista, ki jih glasba vzbudi. Tudi barve pri tem grafu s primerne izbranim razpoloženjem, saj so bolj intenzivne in predvsem v rdeče črnih odtenkih.



Slika 3.14: Prikaz razpoloženj v VA prostoru skupaj z barvami, s katerimi so razpoloženja označili anketiranci. S krogi so označena razpoloženja, ki jih po njihovem mnenju glasba vzbudi pri poslušalcu, z kvadrati pa razpoloženja, ki so izražena v glasbi. Narejen je prikaz za 4 pesmi. Nad slikami so označene identifikacijske številke pesmi. Prva pesem je iz zbirke filmske glasbe (identifikacijska številka se začne z 1), druga je iz zbirke etnografskih pesmi (identifikacijska številka se začne z 4), tretja iz zbirke Jamendo (identifikacijska številka se začne z 5) in četrta iz nabora moderne elektro-akustične glasbe (identifikacijska številka se začne z 6).

Poglavje 4

Algoritmi za ocenjevanje razpoloženja v glasbi

Zadali smo si, da preizkusimo delovanje nekaterih algoritmov za ocenjevanje razpoloženj v glasbi. V tem poglavju bomo predstavili delovanje in rezultate za regresijski algoritem in algoritem Gaiatransform iz knjižnice Essentia.

4.1 Regresijski algoritem

Prvi algoritem, ki smo ga preizkusil je algoritem, ki so ga implementirali Schmidt et al. [29]. Ni nam uspelo dobiti izvirnega algoritma avtorja. Po informacijah, ki smo jih pridobil ob natančnem pregledu avtorjevega članka in nekaj izmenjanih e-poštah z avtorjem, smo sami implementiral algoritem, po navodilih avtorja. Algoritem iz glasbe izračuna značilnice. Na podlagi teh pa potem napove prijetnost (valence) in aktivnost (arousal) v skladbi.

Algoritem za sovje delovanje potrebuje značilnice izračunane na podlagi glasbenih odlomkov. Uporabili smo značilnici MFCC [24] in kromatski vektor [2]. MFCC je bil izračunan s koeficientom 20, kromatski vektor pa je bil izračunan tako, da smo dobili 12 razredov tonskih višin. Za računanje teh vrednosti smo uporabili Python knjižnico LibROSA. Poleg tega potrebujemo še že znane VA vrednosti za del odlomkov, da na podlagi tega treniramo algoritem. V našem primeru smo imeli za vse odlomke znane VA vrednosti, tako smo lahko naključno določili na katerem

delu podatkovne zbirke bomo trenirali algoritem in na katerem delu testirali. Obenem smo na tak način lahko preverili natančnost naših napovedi. VA vrednosti smo smo izračunali za vsako pesem s povprečenjem vrednosti v podatkovni zbirki opisani v poglavju 3.

Algoritem deluje, tako da glasbene odlomke naključno razdeli na dva dela. 70% podatkovne odlomkov uporabi za treniranje in ostalih 30% za testiranje. Na delu zbirke za treniranje potem uporabi metodo najmanjših kvadratov (Least squares method) [1] s katero izračuna vektor b , ki ga potem uporabimo za preslikovanje iz matrike z značilnicami v VA vrednosti z enačbo 4.1. Metod najmanjših kvadratov kot vhod vzame matriko z značilnicami za posamezno pesem in vektor z VA vrednostmi. Ta postopek izvajamo ločeno za prijetnost (valence) in aktivnost (arousal).

$$y = X \cdot b \quad (4.1)$$

X v enačbi 4.1 je matrika, ki v vsaki vrstici vsebuje značilnice za posamezen glasbeni odlomek. b je vektor, ki preslikuje iz prostora, ki ga predstavljajo značilnice v VA vrednost. y je vrednost, ki jo algoritem napove. Vsaka vrednost predstavlja prijetnost ali aktivnost za en glasbeni odlomek.

4.1.1 Rezultati za regresijski algoritem

Algoritem smo kot sem že omenil trenirali z 70% glasbenih odlomkov, test pa smo izvedli na 30% odlomkov. Te napovedi, ki jih je algoritem dal na testnih odlomkih smo potem primerjali s podatki za vsako pesem iz podatkovne zbirke.

Rezultati so predstavljeni v tabeli 4.1 za oba tipa značilnic (MFCC in kromatski vektor) posebej. Za vsak primer smo izračunali povprečno razdaljo med povprečno vrednostjo iz podatkovne zbirke in algoritmično napovedano vrednostjo. Prav tako smo izračunali povprečno razdaljo med najbližjo točko v podatkovni zbirki in napovedano vrednostjo. Povprečno razdaljo do povprečne vrednosti smo izrazili tudi v večkratniku standardnega odklona v podatkih iz podatkovne zbirke. Za primerjavo smo algoritem preizkusili tudi na Mood Swing podatkovni zbirki [30].

Rezultati kažejo boljšo korelacijo med MFCC in prijetnostjo (valence) ter aktivnostjo (arousal) v primerjavi z kromatskim vektorjem. Prav tako lahko opazimo, da so podatki boljši na naši podatkovni zbirki kot na Mood Swing podatkovni

Tabela 4.1: Primerjava rezultatov dobljenih z regresijskim algoritmom na naši podatkovni zbirki in Mood Swing podatkovni zbirki z uporabo značilnic MFCC in kromatski vektor. Rezultati so predstavljeni s povprečno razdaljo med povprečno vrednostjo iz podatkovne zbirke in napovedano vrednostjo, s povprečno razdaljo do najbližje vrednosti v podatkovni zbirki in z povprečno razdaljo do povprečne vrednosti merjeno v večkratniku standardnega odklona (standardne deviacije).

Značilnica	MFCC	kromatski vektor
Naša podatkovna zbirka		
Razdalja do povprečne vrednosti	0.2060	0.2215
Razdalja do najbližje vrednosti	0.0595	0.0614
Razdalja do povprečne vrednosti v std. odkl.	0.4870	0.4993
Mood Swing podatkovna zbirka		
Razdalja do povprečne vrednosti	0.2448	0.3316
Razdalja do najbližje vrednosti	0.0641	0.1026
Razdalja do povprečne vrednosti v std. odkl.	0.6514	0.8940

zbirki. Verjetno gre to pripisati predvsem večjemu številu VA vrednosti v podatkovni zbirki.

Algoritem smo preizkusil tudi na kromatskem vektorju izračunanem s hierarhičnim kompozicionalnim modelom predstavljenim v Pesek et. al [27]. Uporaba te značilnice nam da boljše rezultate, kot jih dobimo z MFCC in kromatskim vektorjem. Na ta način dobljena povprečna razdalja do povprečne vrednosti je 0.1862, razdalja do najbližje vrednosti je 0.0719 in povprečna razdalja merjena v standardnih deviacijah je 0.4459.

V tabeli 4.2 so predstavljene povprečne razdalje izračunane ločeno za prijetnost (valence) in aktivnost (arousal). Iz teh rezultatov je razvidno, da z uporabo vseh značilnic možno bolj natančno napovedati aktivnost v primerjavi s prijetnostjo. Med tem je možno še opaziti, da kromatski vektor izračunan z Hierarhični kompozicionalnim modelom da veliko boljše rezultate pri računanju prijetnosti kot ostale

Tabela 4.2: Povprečna razdalja med povprečnimi vrednostmi in napovedmi izračunana ločeno za prijetnost in aktivnost. Razvidno je da so napovedi aktivnosti dosti boljše od napovedi prijetnosti.

	MFCC	Kromatski vektor	HKM kromatski vektor
Prijetnost (valence)	0.1734	0.1826	0.1494
Aktivnost (arousal)	0.0871	0.0940	0.0898

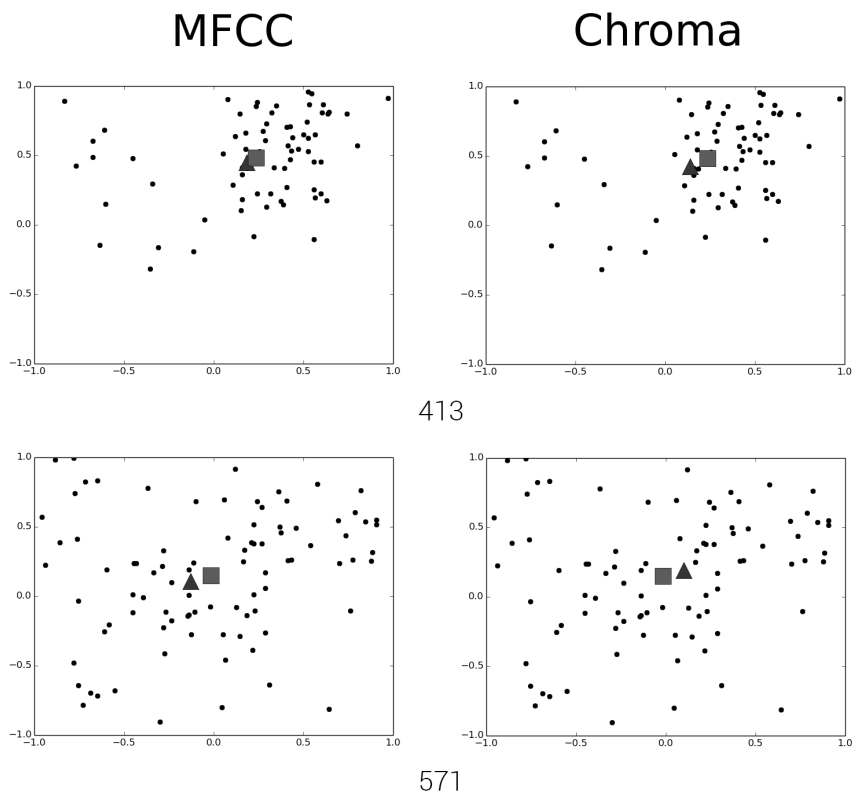
značilnice, med tem ko je aktivnost izračunana podobno natančno kot z uporabo MFCC.

4.2 Essentia Gaiatransofrm algoritem

Drug algoritem, na katerem smo preizkusili natančnost napovedovanja razpoloženja na glasbenih odlomkih iz naše podatkovne zbirke, je GaiaTransform algoritem, ki na podlagi nekaterih značilnic izračuna nekatere visoko-nivojske značilnice [4] med katerimi je tudi razpoloženje. Algoritem se nahaja v knjižnicah Essentia in Gaia [5], ki nam jo je po kar nekaj problemih uspelo naložiti na računalnik.

Essentia je knjižnica namenjena analizi zvoka in predvsem uporabi na MIR področju. Napisana je v C++ jeziku, možno pa jo je uporabiti tudi v jeziku Python. Ponuja funkcionalnost kot so branje in pisanje zvoka, standardno procesiranje signalov, statistična opredelitev podatkov, izračun velikega števila spektralnih, časovnih, tonskih in ostali značilnic. Te značilnice so osnova za izračun visoko nivojskih značilnic. Essentia skupaj s knjižnico Gaia ponuja funkcionalnosti za izračun nekaterih visoko nivojskih značilnic kot so žanr, kultura (zahodna ali nezahodna), vrsto plesa na izbrano skladbo, barvo zvoka, razpoloženja v skladbi. Ocenjuje tudi ali je pesem inštrumental ali ne. Ker v tem delu preučujemo predvsem razpoloženje v glasbi se bomo osredotočili na funkcionalnost, ki poskuša odkriti razpoloženje v glasbi.

Algoritem visoko nivojske značilnice izračuna na podlagi nizko nivojskih, ki so prav tako izračunane s knjižnico Essentia. Za določanje visoko nivojskih značilnic seveda ne uporabi vseh značilnic ampak samo podmnožico teh. To podmnožico so



Slika 4.1: Slika predstavlja 4 VA prostore za pesmi z identifikacijskima številka 413 (prva vrsta) in 571 (druga vrsta). Kvadrat predstavlja predstavlja povprečno vrednost iz podatkovne zbirke, trikotnik prikazuje napoved algoritma (z uporabo MFCC v prvem stolpcu in kromatskega vektorja v drugem). Pike prikazujejo posamezne vrednosti iz podatkovne zbirke.

za vsak razred (na primer srečo) izračunali posebej z uporabo algoritma za izbiro značilnic na podlagi korelacije (correlationbased feature selection - CFS).

Za klasifikacijo so uporabili večrazredni (muliti-class) SVM (support vector machine) algoritem z ena proti ena (one-versus-one) strategijo glasovanja (voting strategy). Uporabili so SVM algoritem iz knjižnice libSVM. Algoritem so razširili z možnostjo napovedovanja verjetnosti prisotnosti določenega razreda (na primer, da pove da je verjetnost, da v pesmi nastopa sreča 89%). SVM je treniran s 5-fold cross-validation.

Algoritem so trenirali na 20 različnih glasbenih zbirkah. Algoritem za ocenje-

vanje razpoloženja iz glasbe so trenirali na zbirki, ki so jo pripravili Laurier et. al [21]. Zbirko so potem sami še dodatno označili.

Za razpoloženje algoritem vrne verjetnost, da je v neki pesmi prisotno eno od šestih razpoloženj: srečen (happy), žalosten (sad), agresiven (aggressive), sproščen (relaxed), akustičen (acoustic), elektronski (electronic) in zabaven (pary). Poleg tega pa algoritem glasbeni odlomek glede na razpoloženje razvrsti v enega od 5 razredov (clousters), ki so uporabljeni v MIREX opravilu (MIREX task). Ti razredi so definirani v poglavju 2.2.2.

4.2.1 Natančnost razvrščanja

Zgoraj opisani algoritem po podatkih avtorjev med samostojnimi oznakami za razpoloženje (vesel, žalosten, elektronski, zabaven, sproščen in agresiven) najbolje napoveduje prisotnost agresivnosti z natančnostjo 97,5%. Potem sledi razpoloženje sproščen (92,92% natančno). Sledijo jima zabaven (88,38%), žalosten (86,96%) in elektronski (84,59%). Najslabše napovedi da algoritem za razpoloženje vesel, kjer je natančnost 82,86%.

Tabela 4.3: Natančnost napovedovanja v gruče v odstotkih glede na teste, ki so jih naredili avtorji algoritma. V navpičnem stolpcu so gruče katerim pesem pripada, v vodoravnem pa gruče kamor je algoritem uvrsti pesem.

	Gruča 1	Gruča 2	Gruča 3	Gruča 4	Gruča 5
Gruča 1	58,62	18,97	10,34	3,45	8,62
Gruča 2	29,63	48,15	16,67	5,56	0,00
Gruča 3	10,81	13,51	71,62	4,05	0,00
Gruča 4	15,62	43,75	12,50	25,00	3,12
Gruča 5	25,49	1,96	0,00	1,96	70,59

Povsem drugačno je stanje pri natančnosti napovedovanja razpoloženj v gruče. V tem primeru je natančnost le 59.14%. Iz tabele 4.3 je razvidno, da algoritem najbolj natančno napoveduje gručo 3 in gručo 5. Najslabše pa gručo 4. Iz tabele je tudi očitno, da sta gruča 2 in gruča 3 od gruče 5 tako drugačni, da ni napovedi, kjer bi pesem namesto v gručo 2 ali gručo 3 uvrstil v gručo 5. Med tem pa je veliko

napačnih preslikav med gručama 1 in 2. Veliko je tudi takšnih napačnih napovedi, ko algoritem pesmi namesto gruč 5 določi gručo 1. Obratna korelacija ni tako močna.

4.2.2 Rezultati napovedovanja z algoritmom Gaiatransform

Primerjava rezultatov pri tem algoritmu je bila malo bolj zahtevna kot primerjava z uporabo regresije, saj v naši podatkovni zbirki nimamo enakih oznak, kot jih algoritem vrača kot rezultat. Zaradi tega je bilo potrebno poenotiti oznake. Ugotovil sem, da bo primerjava najlažja, če uporabim MIREX gruč, ki jih vrača algoritem. Oznake, ki jih imamo v naši podatkovni zbirki, smo razvrstili v gruč glede na podobnost s razpoloženji, ki se nahajajo v teh gručah in so prikazane v poglavju 2.2.2. Razporeditev razpoloženj iz naše podatkovne zbirke v gruč je prikazana v tabeli 4.4.

Tabela 4.4: Razvrstitev oznak razpoloženja iz naše zbirke v MIREX gruč glede na podobnost z obstoječimi razpoloženji iz gruč. V četrti gruči ni nobenega razpoloženja, saj se nobeno razpoloženje iz naše zbirke ne ujema z razpoloženji iz te gruč.

Gruča	Razpoloženja
Gruča 1	presenečenje, navdihnjenost
Gruča 2	veselje, sreča, živahnost
Gruča 3	žalost, otožnost, hrepenenje, pričakovanje. sproščenost, mirnost, zasanjanost
Gruča 4	
Gruča 5	jeza, strah

Nato smo izvedli primerjavo tako, da smo pogledali katero je najbolj pogosto razpoloženje v podatkovni zbirki pri pri vsakem izmed 200 glasbenih odlomkov. To razpoloženje smo potem glede na preslikavo v tabeli 4.4 preslikal v gručo. Potem smo te gruč primerjali z gručami, ki jih za isti glasbeni odlomek napove algoritem.

Tabela 4.5: Rezultati napovedovanja v gruče v odstotkih. V navpičnem stolpcu so gruče katerim pesem pripada, v vodoravnem pa gruče kamor je algoritem uvrsti pesem. V četrti so ničle, ker se nobeno razpoloženje iz naše zbirke ne ujema z razpoloženji iz te gruče. Zaradi tega tudi nobena pesem ne pripada tej gruči. Najbolj natančne ocene smo dobili za gručo 3, najslabše za gručo 1.

	Gruča 1	Gruča 2	Gruča 3	Gruča 4	Gruča 5	Št. odl.
Gruča 1	14,29	14,29	57,14	0	14,29	7
Gruča 2	12,63	41,05	15,79	9,47	21,05	95
Gruča 3	8,86	25,32	58,22	1,27	6,33	79
Gruča 4	0	0	0	0	0	0
Gruča 5	26,32	5,26	15,79	5,26	47,37	19

Natančnost napovedovanja na naši podatkovni zbirki je 47,50%. Bolj natančni rezultati napovedovanja so prikazani v tabeli 4.5. Vsaka vrstica predstavlja gručo v katero spada pesem. Stolpci pa predstavljajo gruče v katere je algoritem uvrstil glasbeni odlomek. Torej na presečišču nam številka vrstice pove kam bi moral spadati določen odlomek, številka stolpca pa kam ga je uvrstil algoritem. Na presečišču je zapisan odstotek elementov uvrščenih v gručo v stolpcu glede na število elementov v dejanski gruči.

Iz teh podatkov je razvidno, da je algoritem najbolj natančno razvrstil elemente v gručo 3, med tem ko je bil najmanj natančen pri določanju gruče 1. Tukaj so rezultati slabi, saj je algoritem večino pesmi iz te gruče uvrstil v gručo 3. Mogoče je tukaj eden od problemov v tem, da je v tej gruči malo pesmi (7), zato ne moramo posplošiti teh rezultatov. V četrti vrstici so vsi rezultati nič, ker v naši podatkovni zbirki nimamo razpoloženja, za katerega bi lahko rekli, da pripada tej gruči. Temu posledično ni nobene pesmi uvrščene v to gručo. Iz rezultatov lahko opazimo, da je več pesmi, ki so v gruči 5 algoritem uvrstil v gručo 1, kar je podobno rezultatom avtorjev algoritma prikazanim v poglavju 4.2.1. Z avtorjevimi rezultati je podobno tudi to, da odlomki iz gruče 1 niso bili uvrščeni v gručo 4 in to, da je bilo malo odlomkov iz gruče 5 uvrščeno v gručo 2.

Ker se pri določenih pesmih več razpoloženj pojavi skoraj enako pogosto v naši podatkovni zbirki, sem potem primerjavo dopolnil tako, da primerja vsa razpoloženja, ki so blizu razpoloženja, ki se pojavi največkrat. Če je gruča enega od teh razpoloženj enaka gruči v katero je algoritem uvrstil glasbeni odlomek smatramo napoved kot pravilno.

Tabela 4.6: Rezultati napovedovanja v gruče v odstotkih v primeru, smo upoštevali tudi razpoloženja, ki so blizu največkrat označenem. V navpičnem stolpcu so gruče katerim pesem dejansko pripada, v vodoravnem pa gruče kamor je algoritem uvrsti pesem. V četrto gručo nismo uvrstili nobene pesmi, zaradi neujemanja vseh naših razpoloženj z razpoloženji v tej gruči. Kot pri tabeli 4.5 so tudi tukaj najboljše napovedi za gručo 3 in najslabše za gručo 1.

	Gruča 1	Gruča 2	Gruča 3	Gruča 4	Gruča 5	Št. odl.
Gruča 1	25	0	75	0	0	4
Gruča 2	12,37	44,33	15,46	9,28	18,56	97
Gruča 3	9,09	22,08	61,04	1,30	6,49	77
Gruča 4	0	0	0	0	0	0
Gruča 5	22,73	4,55	13,64	4,55	54,55	22

Pričakovano se je na tak način natančnost napovedovanja povečala in sedaj znaša 51.5%. Iz rezultatov v tabeli 4.6 je razvidno, da se je natančnost izražena v odstotkih najbolj povečala v prvi gruči, po številu pesmi pa v gruči 5. V gruči 1 so za razliko od prej samo še tiste napačne napovedi, ki uvrščajo v gručo 3. Odstotek teh se je povečal, ker se je zmanjšalo skupno število pesmi v tej gruči. Ostale napovedi se niso bistveno spremenile.

Kot smo že omenili algoritem poleg klasifikacije v MIREX gruče omogoča ocenjevanje za 6 razpoloženj. Za vsako od razpoloženj napove, če se pojavi v glasbenem odlomku. Glede na to, da se 3 od teh oznak ujemajo z oznakami v naši podatkovni zbirki, sem iz podatkovne zbirke vzel tiste pesmi, kjer je je to razpoloženje najbolj prisotno od vseh. Te pesmi sem potem primerjal z rezultati, ki jih vrne algoritem.

Tabela 4.7: Natančnost ocenjevanja razpoloženj vesel, žalosten in sproščen v odstotkih. Iz podatkov je razvidno, da algoritem na naši podatkovni zbirki zelo natančno oceni prisotnost razpoloženj žalosten in sproščen, slabši je pri ocenjevanju razpoloženja srečen.

Srečen	54,55
Žalosten	86,36
Sproščen	87,5

Za ta primer algoritem napove pravilno prisotnost razpoloženja s 79,59% natančnostjo. Kot je razvidno iz tabele 4.7 je natančnost velika pri razpoloženjih žalosten in sproščen. Dosti manjša pa je pri razpoloženju srečen. Kar kaže na to, da je algoritmično direktno iz glasbenih odlomkov dosti lažje napovedati sproščenost in žalost, medtem ko je to težje pri razpoloženju srečen.

Z uporabo algoritma za določanje MIREX gruč iz Essentie smo na naši podatkovni zbirki dobili slabše rezultate, kot so jih dobili avtorji. To gre po vsej verjetnosti pripisati temu, da naše oznake niso enake tistim, ki jih vrača algoritem, pri ročni pretvorbi pa nismo mogli zagotoviti popolnega ujemanja.

Pri napovedovanju posameznega razpoloženja pa so rezultati približno enaki tistim, ki so jih dobili avtorji za razpoloženji žalosten in sproščen. Za razpoloženje vesel so rezultati slabši. Mogoče gre za razliko v dožemanju razpoloženja srečen v slovenščini in razpoloženja "Happy" v angleščini.

Literatura

- [1] Hervé Abdi. The method of least squares. *Encyclopedia of Measurement and Statistics*. CA, USA: Thousand Oaks, 2007.
- [2] Juan P Bello and Jeremy Pickens. A robust mid-level representation for harmonic content in music signals. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 304–311, London, 2005.
- [3] Asa Ben-Hur and Jason Weston. A user’s guide to support vector machines. In *Data mining techniques for the life sciences*, pages 223–239. Springer, 2010.
- [4] D Bogdanov. *From music similarity to music recommendation: Computational approaches based on audio features and metadata*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2013.
- [5] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, Oscar Mayor, Gerard Roma, Justin Salamon, José R Zapata, and Xavier Serra. Essentia: An Audio Analysis Library for Music Information Retrieval. In *ISMIR*, pages 493–498. Citeseer, 2013.
- [6] Paul Brossier, Juan Pablo Bello, and Mark D Plumbley. Real-time temporal segmentation of note objects in music signals. In *Proceedings of the ICMC*, 2004.
- [7] Wei-rong Chu, RT-H Tsai, Ying-Sian Wu, Hui-Hsin Wu, Hung-Yi Chen, and JY-J Hsu. LAMP, a lyrics and audio mandopop dataset for music mood estimation: Dataset compilation, system construction, and testing. In *Technologies and Applications of Artificial Intelligence (TAAI), 2010 International Conference on*, pages 53–59. IEEE, 2010.

-
- [8] Tiziano Colibazzi, Jonathan Posner, Zhishun Wang, Daniel Gorman, Andrew Gerber, Shan Yu, Hongtu Zhu, Alayar Kangarlu, Yunsuo Duan, James A Russell, and Bradley S Peterson. Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion (Washington, D.C.)*, 10(3):377–389, 2010.
 - [9] Tim Dalgleish and Michael J Power. Basic Emotions. In *Handbook of cognition and emotion*. Wiley Online Library, 1999.
 - [10] T Eerola and J K Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1):18–49, 2010.
 - [11] David Gerhard. Pitch Extraction and Fundamental Frequency: History and Current Techniques. Technical report, University of Regina, Saskatchewan, Canada, Regina, 2003.
 - [12] O. Gillet and G. Richard. Transcription and Separation of Drum Signals From Polyphonic Music. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3):529–540, March 2008.
 - [13] F Gouyon, O Delerue, and F Pachet. Classifying percussive sounds: a matter of zero-crossing rate? In *Proceedings of the COST G-6 Conference on Digital Audio Effects*, 2000.
 - [14] Fabien Gouyon and Perfecto Herrera. Exploration of techniques for automatic labeling of audio drum tracks instruments. In *Proceedings of MOSART: Workshop on Current Directions in Computer Music*, 2001.
 - [15] Steinar Heimdal Gunderson. Musical descriptors: An assessment of psychoacoustical models in the presence of lossy compression. 2007.
 - [16] Anssi Klapuri and Manuel Davy, editors. *Signal Processing Methods for Music Transcription*. Springer, New York, 2006.
 - [17] Anssi P Klapuri. Automatic Music Transcription as We Know it Today. *Journal of New Music Research*, 33(3):269–282, September 2004.

-
- [18] Bernie Krause. Echonest of the past. In *The Great Animal Orchestra: Finding the Origins of Music in the World's Wild Places*, chapter Echonest o, pages 5–10. Hachette Digital, Inc., London, 2012.
- [19] Jan La Rue. Harmonic Rhythm in the Beethoven symphonies. *The Journal of Musicology*, 18(2):221–248, 2001.
- [20] Cyril Laurier, Perfecto Herrera, M Mandel, and D Ellis. Audio music mood classification using support vector machine. *Music Information Retrieval Evaluation eXchange (MIREX) extended abstract*, 2007.
- [21] Cyril Laurier, Owen Meyers, Joan Serrà, Martin Blech, Perfecto Herrera, and Xavier Serra. Indexing music by mood: design and integration of an automatic content-based annotator. *Multimedia Tools and Applications*, 48(1):161–184, 2009.
- [22] Mira Lenko, Tina Kogovšek, and Peter Stanković. *Pomen glasbe v očeh mladih: diplomsko delo*. M. Lenko, 2009.
- [23] Yipeng Li and DeLiang Wang. Separation of Singing Voice From Music Accompaniment for Monaural Recordings. *IEEE Transactions on Audio, Speech and Language Processing*, 15(4):1475–1487, May 2007.
- [24] Beth Logan and Others. Mel Frequency Cepstral Coefficients for Music Modeling. In *ISMIR*, 2000.
- [25] R Panda, R Malheiro, B Rocha, A Oliveira, and R P Paiva. Multi-Modal Music Emotion Recognition: A New Dataset, Methodology and Comparative Analysis. *Proc. CMMR*, 2013.
- [26] Matevž Pesek. *Prepoznavanje akordov s hierarhičnim kompozicionalnim modelom: diplomsko delo*. PhD thesis, M. Pesek, 2012.
- [27] Matevž Pesek and Matija Marolt. Chord estimation using compositional hierarchical model. In *6th International Workshop on Machine Learning and Music, held in conjunction with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML/PKDD 2013*, 2013.

-
- [28] Pasi Saari, Tuomas Eerola, György Fazekas, Mathieu Barthet, Olivier Lartillot, and Mark B Sandler. The Role of Audio and Tags in Music Mood Prediction: A Study Using Semantic Layer Projection. In *ISMIR*, pages 201–206, 2013.
- [29] Erik M Schmidt and Youngmoo E Kim. Projection of acoustic features to continuous valence-arousal mood labels via regression. In *10th International Society for Music Information Retrieval Conference. ISMIR*, 2009.
- [30] Erik M Schmidt and Youngmoo E Kim. Modeling Musical Emotion Dynamics with Conditional Random Fields. In *ISMIR*, pages 777–782, 2011.
- [31] Björn Schuller, Clemens Hage, Dagmar Schuller, and Gerhard Rigoll. ‘Mister DJ, Cheer Me Up!’: Musical and textual features for automatic mood classification. *Journal of New Music Research*, 39(1):13–34, 2010.
- [32] Dave Smith and Chet Wood. MIDI Musical Instrument Digital Interface Specification 1.0, 1983.
- [33] Ernst Terhardt. Pitch, consonance, and harmony. *The Journal of the Acoustical Society of America*, 55(5):1061–1069, 1974.
- [34] Douglas Turnbull, Luke Barrington, David Torres, and Gert Lanckriet. Semantic annotation and retrieval of music and sound effects. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(2):467–476, 2008.
- [35] Diane Watson and Regan L Mandryk. Modeling Musical Mood From Audio Features and Listening Context on an In-Situ Data Set. In *ISMIR*, pages 31–36, 2012.
- [36] B Wu, S Wun, C Lee, and A Horner. Spectral correlates in emotion labeling of sustained musical instrument tones. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 415–421, 2013.
- [37] Yongwei Zhu, Mohan S Kankanhalli, and Sheng Gao. Music key detection for musical audio. In *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International*, pages 30–37. IEEE, 2005.