

UNIVERZA V LJUBLJANI  
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Primož Godec

# **Evalvacija čustev iz glasbe**

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 mentorja<sup>1</sup>

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



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:

*Evalvacija čustev iz glasbe*

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 11. januarja 2011

Podpis avtorja:





*Želim se zahvaliti mentorju Matiji Maroltu, za pomoč in spodbujanje pri raziskovanju in izdelavi diplomskega dela. 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**

<b>1</b>	<b>Uvod</b>	<b>1</b>
<b>2</b>	<b>Pregled področja</b>	<b>3</b>
2.1	Pridobivanje informacij iz glasbe . . . . .	4
2.2	Evalvacija čustev iz glasbe . . . . .	5
2.3	Pregled osnovnih elementov . . . . .	10
<b>3</b>	<b>Naš dataset</b>	<b>11</b>
3.1	Zbiranje dataseta . . . . .	11
3.2	Sestava datasta . . . . .	16
3.3	Analiza podatkov . . . . .	18
<b>4</b>	<b>Algoritmi za evalvacijo čustev iz glasbe</b>	<b>27</b>
4.1	Regresijski algoritem . . . . .	27
4.2	Essentia Gaiatransofrm algoritem . . . . .	30



# Seznam uporabljenih kratic

<b>SVM</b>	support vector machine	metoda podpornih vektorjev
...	...	...





# Povzetek

V vzorcu je predstavljen postopek priprave diplomskega dela z uporabo okolja L<sup>A</sup>T<sub>E</sub>X. Vaš povzetek mora sicer vsebovati približno 100 besed, ta tukaj je odločno prekratek.

**Ključne besede:** glasba, razpoloženje, evalvacija.



# Abstract

This sample document presents an approach to typesetting your BSc thesis using L<sup>A</sup>T<sub>E</sub>X. A proper abstract should contain around 100 words which makes this one way too short.

**Keywords:** music, mood, evaluation.



# Poglavje 1

## Uvod



## Poglavje 2

### Pregled področja

Lahko bi rekli, da je glasba ena najstarejših in zelo pomembnih aktivnosti na svetu. Razširjena je praktično po celem svetu in poznajo jo še tako izlirana in od ostalega sveta odmaknjena plemena. Znano je, da glasba na svetu obstaja že vsaj 50 000 let. Prva glasba naj bi takrat izvajali na Afriških tleh. Nato se je skozi čas razvijala in postala eden najpomembnejših sestavnih delov človekovega življenja. [23] [11] 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 ko smo v trgovini. Ponekod z pravo iziro glabe vplivajo na človekove odločitve. Na primer v raznih trgovinah in lokalih z glasbo privlačijo kupce. Vse to priča o pomembnem vplivu glasbe pri človeku.

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 zbudi potlačena čustva. Na vzdušje vpliva razmerje med toni. Mol pričara bolj melanholično vzdušje. Dur pa ba bolj veselo. [14]

V nadaljevanju poglavja bom povedal še nekaj o področju imenovanem pridobivanje informacij iz glasbe (Music information retrieval - MIR). To področje je pomembno, saj je osnova za temo moje diplomske naloge. Povedal bom nekaj o povezavi med razpoloženjem in glasbo. Pregledal bom datase, ki obstajajo trenutno na področju razpoloženja in glasbe. Za konec pa bom predstavil še algoritme za evalvacijo čustev iz glasbe.

## 2.1 Pridobivanje informacij iz glasbe

Pridobivanje informacij iz glasbe (Music information retrieval - MIR) je interdisciplinarna znanost, ki povezuje predvsem muzikologijo in računalništvo. [16] Vključuje pa tudi vede kot so psihologija, akademski študij glasbe, procesiranje signalov, strojno učenje in mnoge ostale. To področje je dokaj novo in se trenutno res hitro širi. Kljub temu, da je področje novo, je postalo že zelo široko. Pojavljajo se velike razlike v načinu obdelave in uporabe podatkov. Prav tako so cilji raziskovalcev zelo različni.

Ključni namen področja je pridobiti informacije iz glasbe in še bolj pomembno te informacije potem uporabiti v različnih sistemih. Trenutno se največ dela na naslednjih sistemih.

**Sistem za predlaganje glasbe (Music Recommendation System)** V osnovi taki sistemi predlagajo glasbo na podlagi zgodovine poslušane glasbe s strani uporabnika. Trenutno obstaja veliko takih sistemov kot sta Pandora in Last.fm, ampak le redki uporabljajo informacije pridobljene z MIR za svoje delovanje. Namesto tega veliko sistemov uporablja informacije na podlagi primerjave med uporabniki. Naprimer sistem predlaga glasbo, ki so jo poslušali uporabniki s podobno zgodovino poslušanja. Drugi spetu uporabljajo oznake k določeni glasbi ali pa kakšen druge informacije, ki niso del MIR-a. Te oznake lahko dodajo ali uporabniki ali pa tudi experti. Pri sistemu Pandora glasbo označujejo strokovnjaki pri sistemu Last.fm pa uporabljajo oznake dodane s strani uporabnikov. Oznake, ki se uporabljajo so lahko različne od tega, da označijo zvrst glabe, lahko oznake opisujejo tudi razploženje, ali je pesem instrumental ali ali ima voka, je vokal ženski, moški in podobno. Od kar pa se je področje MIR zelo razširilo vedno več sistemov uporablja tudi te informacije za predlaganje glasbe.

**Ločevanje pesmi na več pasov in prepoznavanje instrumentov** Sistem za ločevanje pesmi zna pesem v kateri nastopa več instrumentov in vokali, razstaviti tako, da imamo posamezne pasove v pesmi ločene. Na primer lahko loči po instrumentih. Da lahko loči posamezne pasove potrebuje sistem za prepoznavanje instrumentov. Zaradi tega sta ta sistema tako tesno povezana med seboj.



Na tem področju obstaja že kar nekaj sistemov. Ti sistemi se veliko uporabljajo pri ustvarjanju podlag za karaoke. Tam je naprimer potrebno pesmi odvzeti vokal. Uporabljajo se tudi za izključevanje posameznih inštrumentov iz podlage. Ti sistemi pa seveda še niso popolni. Veliko je še prostora za izboljšave. Velik problem je v tem kako najbolje ločiti vokal od ostalih delov pesmi, ker si vokal z določenimi inštrumenti deli frekvence.

**Avtomatično prepisovanje glasbe** Ti sistemi delujejo tako, da glasbo iz posnetka pretvarjajo v simbolični zapis. Največ se to uporablja za prepisovanje glasbe v zapis MIDI. Taki sistemi vključujejo kar nekaj podsistemov: detekcijo pojavov (onset detection), ocenjevanje trajanja, identifikacijo instrumentov, prepoznavanje ritma in ostale. Sistemi, ki trenutno obstajajo še niso popolni. Predvsem nastane problem, ko je v pesmi veliko inštrumentov in več polifoničnih nivojev.

**Avtomatična kategorizacija glasbe** To so sistemi, ki znajo razvrstiti glasbo v več prej definiranih skupin. Na področju MIR se raziskovalci trenutno ukvarjajo predvsem z kategorizacijo po žanrih in glede na čustva v glasbi. Za obe dve temi MIREX (Music Information Retrieval Evaluation eXchange) organizira tekmovanja v kategorizaciji glasbe. Raziskovalci lahko oddajo svoj algoritem, ki ga potem pežnejo na MIREX-ovem datasetu. Za kategorizacijo se uporabljajo tehnike za strojno učenje, kot so SVM [3], regresija, različna drevesa in še ostale.

**Generiranje glasbe** Eden od ciljev raziskovalcev v MIR pa je tudi narediti dober sistem za avtomatično generiranje glasbe. Zaenkrat še ne obstaja sistem, ki bi samostojno zgeneriral dobro pesem. V trenutnih sistemih je potrebno predvsem veliko ročnega prilagajanja.

## 2.2 Evalvacija čustev iz glasbe

Eno od pomembnih področij v MIR je zagotovo evalvacija čustev iz glasbe s pomočjo računalniških algoritmov. Ti algoritmi v osnovi delujejo tako, da najprej iz zvočnega zapisa izračunajo določene opisnike. V naslednjem koraku je potrebno na podlagi že obstoječe podatkovne zbirke natrenirati algoritem. Nato pa lahko izvedemo klasifikacijo na večji zbirki zvočnih posnetkov. Za delovanje teh

algoritmo potrebujemo podatkovno zbirko z že obstoječimi podatki. Zato bom v naslednjem podpoglavju opisal kaj na tem področju že obstaja. Poleg tega bom naredil še pregled algoritmov, ki že obstajajo.

### 2.2.1 Podatkovne zbirke

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

**Podatkovna zbirka s filmsko glasbo** Eerola et. al [8] so zbrali podatkovno zbirko, ki vključuje filmsko glasbo iz različnih znanih filmov. Ta podatkovna zbirka je razdeljena na dva dela. Prvi del vsebuje 361 glasbeni odlomkov. Za vsako 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 čustva (vesel, žalosten, nežen in podbne). K vsaki odlomku je dodana tudi povprečna številčna vrednost o prisotnosti posameznega čustva izmed nabora čustev: 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 še številčne vrednosti, ki označujejo prisotnost za čustvi lepota (beauty) in naklonjenost (liking). Ta podatkovna zbirka poleg podatkov vsebuje tudi glasbene odseke, ki so dlogi 15 sekund. Kar je dobro za evalvacijo podatkovne zbirke na različnih algoritmih. Poleg tega 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 [20] je bila zbrana za 240 glasbenih odlomkov popularne glasbe. Za 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 id 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 imenovana Cal500 [22] ima zbrane podatke o razpoloženju za 500 pesmi. Pesmi so zbrane izmed zahodne popularne glasbe. Za vsako pesem ima oznako za razpoloženje (vesela, žalostna, jezna in podobne). V naboru imajo 18 možnih oznak. Podatkovna zbirka je bila anotirana ročno in ima 3 oznake na pesem. Poleg ročne anotacije so kaseneje anotirali tudi s pomočjo algoritma. Podatkovna zbirka vsebuje tudi glasbene datoteke.

**MTV Music Dataset** MTV Music Dataset [21] vsebuje podatke za 192 pesmi izbrane iz MTV Europe Most Wanted lestvic med leti 1981 in 2000. Celotna podatkovna zbirka je bila anotirana s strani 5 ocenjevalcev. Vsaka ocena vsebuje vrednost za prijetnost (valence) in vrednost za aktivnost (arousal). Vsaka pesem je bila ozančena s strani 3h ocenejvalcev.

**LAMP** Podatkovna zbirka LAMP [6] vsebuje podatke za 492 popularnih pesmi izdanih med 2002 in 2008. Podatkovna zbirka je bila labelirana 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. Isto je moral narediti v tretjem koraku na podlagi besedila in zvočnega posnetka skupaj.

**Multi Modal** Zadnja zbirka, ki jo bom opisal je tako imenovana Multi Modal [15] podatkovna zbirka. Vsebuje 903 posnetke, ki so tudi dodani podatkovni zbirki. Glasba je večinoma popularna zahodna. Za vsako pesem je določena oznaka s čustvom. Poleg tega pa je tudi vsaka pesem razvrščena v enega od 5 razredov po razpoloženju.

### 2.2.2 Algoritmi za evalvacijo čustev iz glasbe

Vsi algoritmi za evalvacijo razpoloženja iz glasbe potrebujejo za svojo delovanje podatkovno zbirko, zato sem se v prejšnjem poglavju posvetil ravno tej temi. Poleg podatkov moramo poskrbeti tudi za različne značilke, na podlagi katerih se potem izvede klasifikacija.

V osnovi lahko delovanje algoritmov za evalvacijo čustev iz glasbe razdelimo na dva dela. Prvi del je treniranje algoritma. V drugem delu pa algoritem dejansko izvede klasifikacijo. Prvi del (treniranje poteka tako), da uporabimo eno izmed mnogih metod (Support Vector Machine, regresijo, dreves ali kakšno drugo). Na podlagi značilnic in podatkov iz podatkovne zbirke (že določene končne vrednosti za glasbo) algoritem izračuna pravilo za preslikavo. Ta postopek se izvaja na delu podatkovne zbirke namenjene treniranju algoritma. Ko algoritem ve pravilo lahko izvedemo drugi del klasifikacije. Na podlagi značilnic in pravila za preslikavo sedaj algoritem določi vrednosti. Vrednost je lahko skupina po razpoloženju v katero uvrstimo določeno pesem ali pa številčna vrednost, ko določamo le prijetnost in aktivnost. Drugi del se izvede na drugem delu zbirke pesmi.

To je opis splošnega delovanja algoritmov, ki je skupno vsem. Vsak od algoritmov pa ima svoje posebnosti. Zbral sem nekaj že obstoječih algoritmov, ki jih bom sedaj opisal.

Schmidt et. al [19] 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 značilke so uporabili Mel-frequency cepstrum in Chromo. Algoritem napoveduje prijetnost (valence) in aktivnost (arousal) pesmi. Parametra algoritem napoveduje ločeno.

Panda et al. [15] so uporabili več algoritmov za napovedovanje in primerjali točnost napovedovanja. Uporabili so: Support Vector Machines, K-nearest Neighbours, C4.5 in Naive Bayes. Algoritem je izvaja klasifikacijo na podlagi 19 značilnic izračunanih iz glasbenega posnetka. Tem značilnicam so dodali tudi značilnice pridobljene iz MIDI signala in značilnice iz besedil. Izvajal je klasifikacijo v 5 gruč, ki so definirane za MIREX tekmovanje (tabela 2.1). Kot najboljši algoritem se je izkazal SVM. Samo na podlagi značilnic iz zvočnega zapisa je dosegel natančnost 64%. Ko so dodali 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 [12] so za evalvacijo uporabili Support Vector Machine (SVM) [3]. Algoritem za delovanje uporablja 133 značilk. Tudi ta algoritem deluje tako, da kot rezultat za vsako pesem vrne eno od petih gurč 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 [24] je uporabila klasifikator z Bayesovimi mrežami in Markove ocene. Kot rezultat je algoritem napovedoval prijetnost (valence) in aktivnost (arousal). Za razliko od ostlih je ona uporabila podatke, ki niso bili zajeti v laboratorijskem okolju ampak 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).

Sari et. al [18] so uporabili Semantic Layer Projection (SLP) metodo za klasifikacijo. Za klasifikacijo so uporabili značilnice, končni rezultat pa so oznake s čustvi. Za razliko od prej opisnih metod, ki preslikavo 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) in potem v drugem koraku na podlagi teh vrednosti določijo razporeditev. Naj omenim, da so poskusili tudi algoritem v kombinaciji z tekstovnimi oznakami, vendar je deloval slabše.

Tabela 2.1: Gruče z čustvi, ki se uporabljajo v MIREX mood tekmovanju

Gruče	Čustva
Gruča 1	passionate, rousing, confident, boisterous, rowdy
Gruča 2	rollicking, cheerful, fun, sweet, amiable/good natured
Gruča 3	literate, poignant, wistful, bittersweet, autumnal, brooding
Gruča 4	humorous, silly, campy, quirky, whimsical, witty, wry
Gruča 5	aggressive, fiery, tense/anxious, intense, volatile, visceral

## 2.3 Pregled osnovnih elementov

### 2.3.1 Značilnice

Kot ste v opisih delovanja algoritmov opazili praktično vsi za svoje delvanje uporabljajo tako imenovane značilnice. To so različne vrednosti, ki jih lahko izračunamo iz glasbenega posnetka in so definirane na različne načine.

Razdelimo jih lahko v skupine po podobnosti. Eno od takih razporeditev bom predstavil v nadaljevanju. Tako so značilnice razoredili v [12]:

- **Spektralne značilnice** so značilnice, ki opisujejo obliko spektra v glasbi. Izmed teh značilnic se na MIR področju veliko uporablja MFCC, HFC (High Frequency Content), Chroma, Spectral Strong peak in podobni. MFCC [10] (Mel-frequency cepstral coefficients) je koeficient kratkotrajnega močnostnega spektra zvoka. Temelji na linearni kosinusni transformaciji logaritmičnega močnostnega spektra na nelinearni mel frekvenčni skali. Spectral strong peak [9] nam pove ali spekter vključuje izrazit vrh. Vrednost je večja, če je vrh višji in tanjši.
- **Značilnice, ki opisujejo glasnost** Med temi značilnicami sem v MIR zasledil uporabo naslednjih: Larm, Leq, RMS in podobnih. RMS (Root mean square) [2] je kvadratni koren povprečnih vrednosti kvadratov valovanja glasbe.
- **Visoko nivojski opisniki v glasbi** Višjenivojski deskriptorji so bolj abstraktni in med seboj zelo različni. Med njimi lahko najdemo opisnike kot so
- **Časovni opisniki**
- **Statistične značilnice**

## Poglavje 3

### Naš dataset

Tema moje diplomske naloge je vsekakor evalvacija čustev iz glasbe s pomočjo računalniških algoritmov. Dobre evalvacije pa ni možno narediti brez dobrega dataseta, zato smo se odločili, da zgradimo svoj dataset, ki bo osnova za raziskovanje povezave med čustvi in glasbo. Poleg tega smo se odločili, da dataset nadgradimo s podatki o barvah, ki po mnenju anketirancev najbolj odražajo posamezno skladbo. Tako smo s tem pridoli še povzavo med čustvi, glasbo in barvami.

V naslednjih delih bom predstavil, kako smo zbirali podatke iz našega dataseta. Kasneje pa bom še analiziral nekatere podatke iz našega dataseta.

#### 3.1 Zbiranje dataseta

Dataset smo zbirali s spletno anketo, ki smo jo sami implementirali. Še preden pa smo lahko naredili glavno anketo, smo morali sprejeti še nekaj odločitev, kako zgraditi anketo, da bo dala dobre in predvsem uporabne rezultate. Prva stvar, kjer je bil potreben premislek, je izbor oznak za emocije. Ugotovili smo, da obstajajo nekatere osnovne oznake za čustva, ki si jih je možno pogledati v [7], ni pa standardnega seta oznak, ki bi se uporabljal na področju povezanem z razpoloženjem in glasbo. Nekatri avtorji so izbrali zbirko oznak čisto intuitivno naprimer [25]. Zaradi tega smo se odločili, da naredimo prliminarno raziskavo v obliki ankete.

### 3.1.1 Preliminarna analiza

Kot omenjeno smo preliminarno analizo izvajali s pomočjo ankete. To smo izvedli v elektronski obliki. V njej smo želeli preveriti osnovno strukturo glavne ankete, primernost elementov ankete in najbolj pomembno ugotoviti, katere oznake za čustva, so tista, ki bojo uporabljena v glavni anketi. Čustva smo izbirali tako, da je moral uporabnik za 46 oznak povedati, kako je neko čustvo prisotno pri njemu v tistem trenutku. Uporabnik je to označil na skali od 1 do 7. Iz tega seznama je bilo potem izbranih 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.

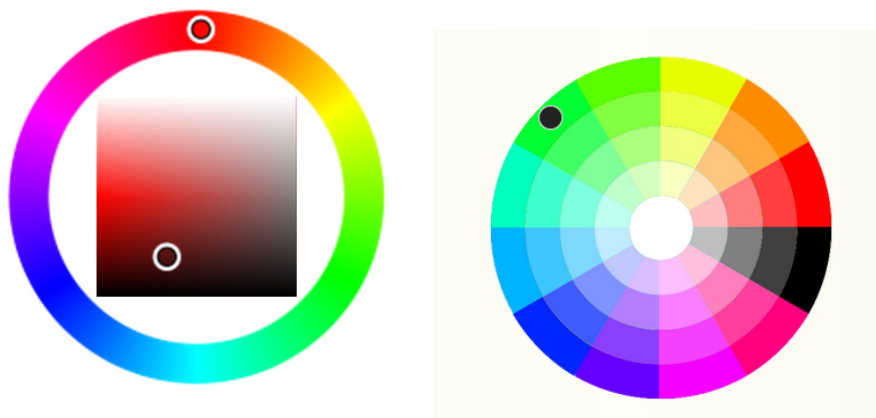
slika skale

Kot sem že omenil nas je zanimala tudi struktura vprašalnika, ki je ostala približno enaka s to razliko, da smo v novem vprašalniku dodali del z glasbnimi odseki ampak več o tem v [ref na chapter].

Poleg tega pa smo testirali tudi elemente uporabljene v anketi (7 stopenjska skala, neskončni barvni krog, izbira z radio gumbi in tekstovnimi polji). Ugotovili smo, da moramo nekatere elemente spremeniti. Najbolje smo spremenili barvno skalo, saj smo jo omejili na barvni krog z 49 možnostmi izbire. To je bilo potrebno, ker je imel na neskončni skali uporabniki preveliko možnost izbiranja, obenem pa je bil sistem trodimenzionalen, kar večina uporabnikov sploh ni opazila in so nastavljali samo odtenek na svetlost pa so pozabili. Skala z 49 možnostmi (prikazana na sliki 3.1) se je izkazala kot odlična 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 ozanči tri svoje najljubše žanre smo se odločili, da namesto vpisnih polj uporabniku ponudimo, da iz seznama izbere in na drug seznam potegne svojo izbiro. Za to smo se odločili, ker so uporabniki tja vpisovali tudi žanre, ki niso osnovni in tisti, ki sploh niso žanri. Prav tako nam je ta preliminarni vprašalnik pomagal izbrati kateri žanri so tisti, ki jih bomo uporabniku ponudili. Zamenjali smo še način, kako uporabnik vnese svoje trenutno razpoloženje in za ta namen uporabili nov element MoodGraph, ki ga bom opisal kasneje.





Slika 3.1: Prvi barni krog smo v preliminarni anketi. V zunanjem krogu je bilo možno izbrati odtenek (hue) v notranjem pa je bilo možno izbirati nasičenost in svetlost. Drugi krog smo uporabili v glavni anketi. Imel je možnost izbire med 49 barvami. Rezina predstavljajo en barvni odtenek, proti notranjosti pa se spreminja svetlost in intenzivnost.

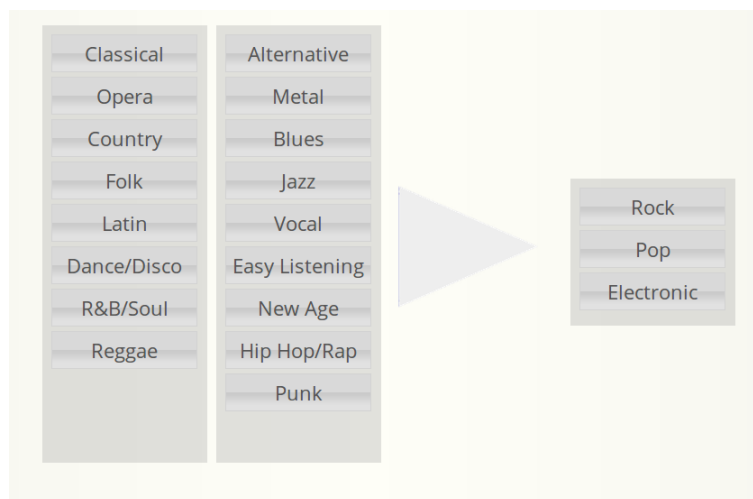
### 3.1.2 Glavna anketa

Ko so bile izbrane oznake in struktura v grobem določena smo se lotili implementacije druge verzije vprašalnika, ki predstavlja glavni vprašalnik za zbiranje naše podatkovne zbirke.

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

**Prvi del** V tem delu smo spraševali o anketirančevi starosti, spolu in o tem ali živi na podeželju ali v mestu. Uporabnika smo vprašali tudi o tem ali je pod vplivom drog ali substanc. Zanimali so nas tudi podatki o glasbeni izobrazbi in o tem ali udeleženec igra inštrument ali poje. Poleg tega nas je še koliko časa dnevno uporabnik posluša glasbo. Anketiraneec je moral tudi izbrati do tri svoje najljubše žanre in jih razporediti po priljubljenosti. To smo zajemal s pomočjo elementa

prikazanega na sliki 3.2, kjer je uporabnik izmed seta 20 žanrov izbral najljubše in jih potegnil potegnil v stolpec desno ter razporedil po priljubljenosti.

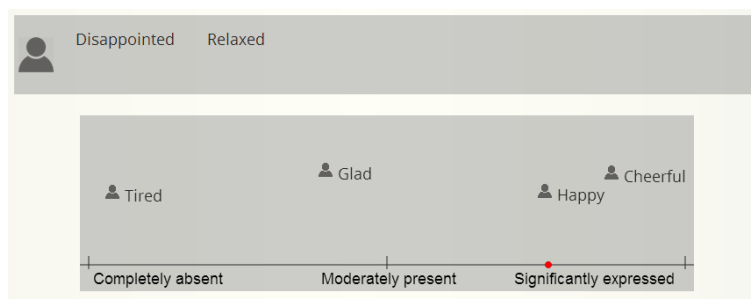


Slika 3.2: Element v anketi uporabljen za izbiro najljubših žanrov

**Drugi del** Drugi del je namenjen zaznavanju anketirancovega trenutnega razpoloženja in njegove percepcije oznak za razpoloženje in barve za razpoloženje. To podatke zajemamo zato, ker na tak način lahko ugotovimo vzrok v različnih ocenah razpoloženja pri delu z glasbo.

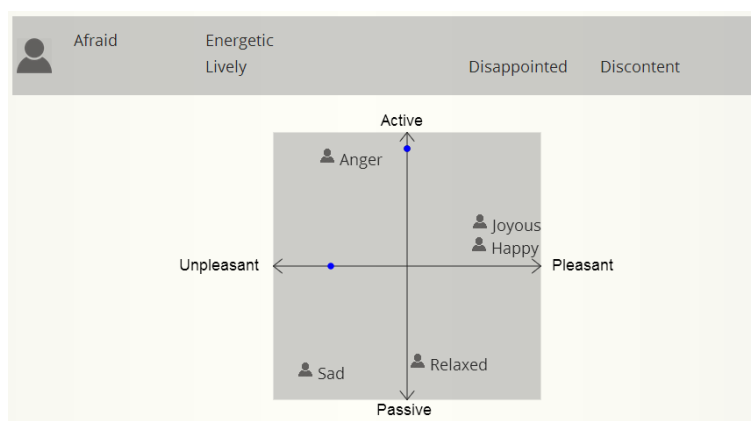
Na začetku je bil uporabnik vprašan kako bi svoje razpoloženje opisal s točko v prostoru, kjer zajemamo prijetnost in aktivnost (VA prostor) [ref]. To je dvodimenzionalen prostor, kjer na x osi od leve proti desni narašča prijetnost in od spodaj navzgor aktivnost. Svoje razpoloženje je moral opisati tudi z izbiro barve s pomočjo elementa 3.1.

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



Slika 3.3: Element za zajemanje pristnosti posameznega čustva imenovan moodstripe.

V drugem delu smo uporabnika povprašali tudi o njegovi percepciji posameznega čustva. Anketiranec je moral za vsako čustveno oznako povedati kako prijetno je to čustvo in kako aktivno je (VA vrednost). To smo zajemali s pomočjo elementa imenovanega enokategorni MoodGraph (slika 3.4). To je 2D prostor, kjer je na vodoravni osi prijetnost in na navpični osi anktivnost. Uporabnik je ozake čustev prikazane nad grafom potegnil v to ravnino na mesto za katerega misli, da ga najbolje opisuje.



Slika 3.4:

Poleg tega kako si uporabnik predstavlja posamezno čustvo v VA prostoru nas je zanimalo tudi kako bi uporabniku opisal ista čustva z barvno. To smo izvedli s pomočjo barvnega krogra (slika 3.1), ki sem ga že opisal zgoraj.

**Tretji del** V tretjem delu smo vprašali anketirance, da ozačijo 10 glasbenih odlomkov dolgih 15 sekund. Odlomki so bili izbrani iz nabora 200 odlomkov. Glasba je bila izbrana tako, da je bila uporabnikom nepoznana. S tem smo se izognili pristranskosti zaradi uporabnikovega poznavanja določene glasbe in vpliva dogodkov, ki so se mu zgodili ob poslušanju določene pesmi. Vsak uporabnik je dobil samo 10 odlomkov iz tega razloga, da uporabnikov nebi preveč obremenili in bi zaradi tega dobili slabše ocene.

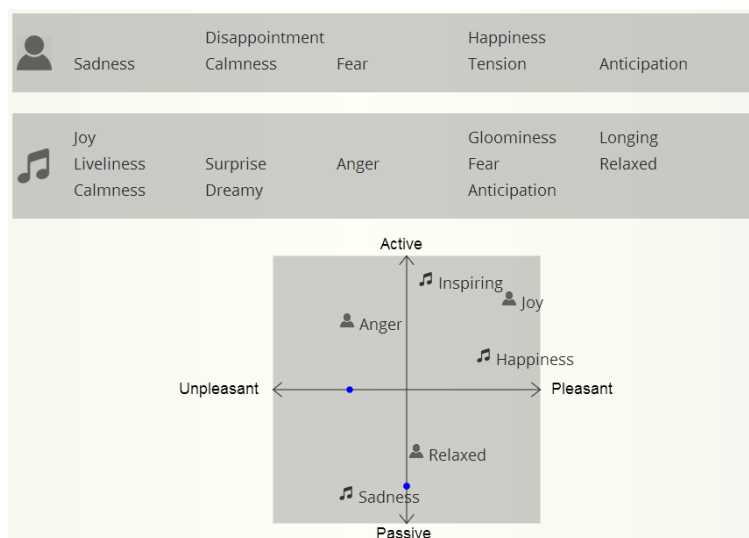
Glasba je bila izbrana iz 4 različnih virov. 80 pesmi smo izbrali iz odprte podatkovne zbirke Jamendo. Iz tega vira smo izbrali bolj vsakdanjo glasbo različnih zvrsti. Naslednjih 80 pesmi smo vzeli iz zbirke filmske glasbe opisane v [8]. Dodali smo še 20 slovenskih etno pesmi in 20 pesmi iz nabora moderne elektro-akustične glasbe.

Za vsak odlomek je moral uporabnik narediti dve stvari. Najprej je s pomočjo barvnega kroga (slika 3.1) povedal s katero barvo bi opisal posamezno pesem. Nato pa je v drugem koraku izbral katera čustva so izražena v odlomku in katera čustva posamezna glasba vzbudi pri njem. Za prvo je izbiral izmed nabora 14 oznak, za drugo pa je iz nabora 10 ozank. 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. Naprimer pri neki pesmi je lahko veselje zelo aktivno, pri drugi pa majn. Poleg tega nam ta podatek da možnost, da raziskujemo tudi samo kako si uporabnik v VA prostoru predstavlja posamezno pesem. Za zajemanje tega podatka smo uporabili dvokategorni MoodGraph prikazan na sliki 3.5.

### 3.1.3 Evaluacijska anketa

## 3.2 Sestava datasta

Sedaj sem opisal kako smo zbrali podatke za podatkovno zbirko, nisem pa še povedal kakšni so ti podatki. Kot anketa so tudi podatki razdeljeni v tri dele. V prvem delu smo zbrali več kot 1400 odgovorov in s tem tudi toliko vpisov v naši zbirki. V drugem delu je vpisov malo več kot tisoč. V tretjem delu, kjer je moral uporabnik oceniti 10 pesmi pa smo zbrali več kot 7200 vpisov v podatkovno zbirko. Torej je ta del izpolnjevalo nekaj več kot 700 anketirancev.



Slika 3.5:

V prvem delu so podatki, ki opisujejo anketirance. Za vsakega imamo 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 in koliko časa je hodil v glasbeno šolo do leta natančno ter koliko časa na dan posluša glasbo. Tukaj je uporabnik poslušanje glabe uvrstil v eno od kategorij: do 1 ure, od 1 do 2 ure, od 2 do 3 ure in več kot 3 ure. Imamo tudi podatek o največ treh najljubših glazbenih zvrsteh, ki jih anketiranec najraje posluša. Tukaj je anketiranec podal najmanj eno in največ tri zvrsti. Zanimalo nas je še psihofizično stanje anketiranca med reševanjem. Torej imamo podatek ali jemlje zdravila, ki vplivajo na razploženje ter če je bil v trenutku reševanja pod vplivom drog ali alkohola.

Drugi del podatkovne zbirke vsebuje podatke o anketirancevem razpoloženju v trenutku, ko je izpolnjeval anketo in o tem kako si anketiranec predstavlja posamezna čustva. Uporabnikovo razploženje opisujejo trije podatki. Prvi je točka v VA prostoru ( $x$  in  $y$  koordinata). Drugo je barva v barvnem krogu (tukaj hranimo podatek o tem katero barvo je anketiranec izbral). Tretji pa je vrednost kako močno je posamezno čustvo iza naobra 17 čustev, izraženo pri anketirancu v tistem trenutku. Hranimo vrednost med 0 in 1 za vsako čustvo. Kot sem že omenil drugi del vsebuje tudi podatek o tem kako si anketiranec predstavlja čustva. Za nabor 10 čustev imamo podatek kam v VA prostoru spada to čustvo po mnenju

anketiranca in barvo s katero ga je anketiranec označil.

Tretji del podatkovne zbirke je malo drugačen. Tu nimamo le enega odgovora na anketiranca ampak do 10 odgovorov. Za vsako pesem en odgovor. Imamo podatek o tem katera čustva so izražena v glasbi in katera so vzbujena pri anketirancu. Poleg tega pa imamo za vsakega od izbranih čustev še podatek kam v VA prostor ga bi anketiranec postavil. Za vsako pesem imamo še podatek o barvi s katero je anketiranec označil glasbeni odlomek.

## 3.3 Analiza podatkov

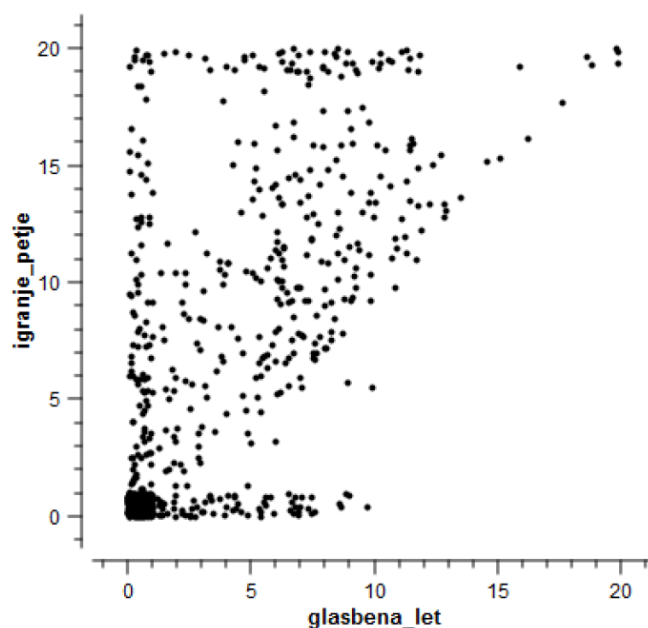
### 3.3.1 Demografska analiza

Povprečna starost anketirancev je 26,5 let. Najmanjši anketiranec je bil star 15, najstarejši pa 64. Največ udeležencev je bilo starih 20 let. Anketo je rešilo več anketirancev ženskega spola, kar 65%. Večina udeležencev prihaja iz urbanega okolja, kar 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 dobro korelirata ( $r=0,653$ ). Iz slike 3.6 je poleg te korelacije razvidno, da se večina ljudi z glasbo ukvarja tudi po dokončani formalni izobrazbi, kar sklepam iz podatka, da se večina ljudi več let neformalno ukvarja z glasbo, kot pa traja formalna izobrazba.

Največ anketirancev posluša glasbo 1 do 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** 4.2% ljudi pravi, da jemljejo zdravila, ki vplivajo na razpoloženje. Kar 3% anketirancev pa je priznalo, da so bili med reševanjem ankete pod vplivom drog. Tudi ta dva parametra dobro korelirata med sabo ( $r = 0,519$ ).



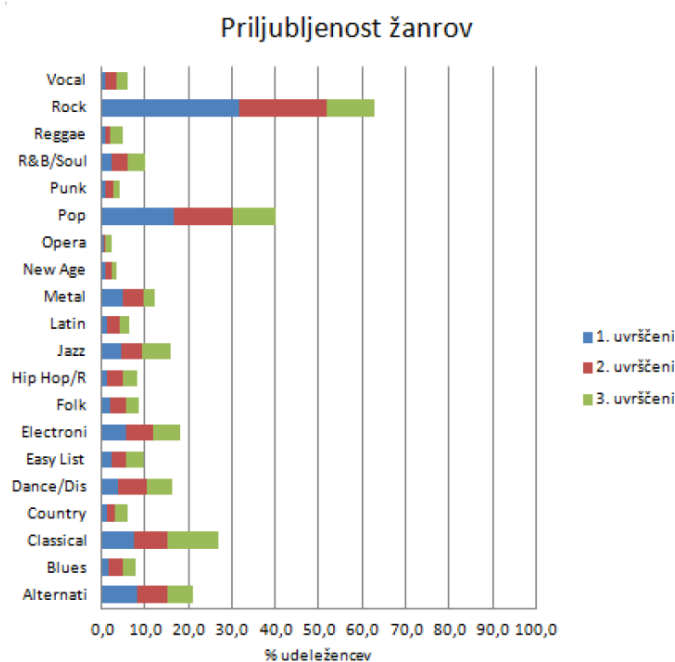
Slika 3.6: Vsaka točka na sliki prikazuje povezavo med trajanjem formalnega glasbenega izobraževanja (na vodoravni osi) in leta ukvarjanja z glasbo (igranje instrumenta ali petje) na navpični osi.

**Žanri** Kot je razvidno iz grafa na sliki [?] je večina označila kot najljubši žanr rock (31,6% anketiranih). Na drugem mestu med najljubšimi žari je pop z 16,6% in na tretjem alternativna glasba. Na zadnjih dveh mestih sta opera, ki je najljubši žanr manj kot odstotku vprašanih. Enako velja za reggae. Žanri na prvo mesto niso izbrani enakomerno, 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 uvrščen žanr na drugo mesto je Opera. Na drugem mestu so izbire že bolj raznolike kot pri najljubšem žanru, vendar vrh še vedno izstopa.

Zanimivo je, da je največ udeležencev na tretje mesto uvrstilo klasično glasbo. Za to izbiro se je odločilo kar 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 preveč.

Izmed vseh žanrov smo potem izbrali najbolj priljubljene: rock, pop, country,



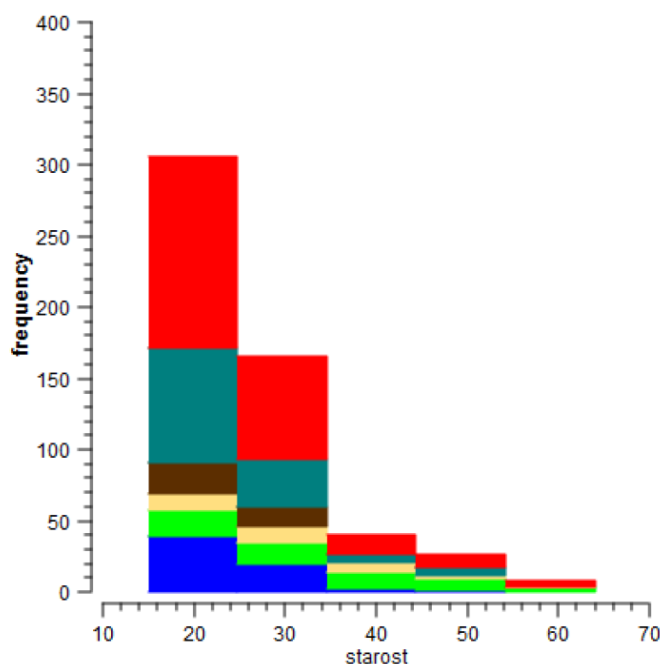
Slika 3.7: Priljubljenost žanrov ter uvrstitev teh na prvo, drugo ali tretje mesto.

alternativna glasb in klasična glasba. Naredili smo primerjavo med najljubšimi žanri po starostnih skupinah (slika 3.8). Anketirance smo razdelili v 6 starostnih skupin. Vsaka od skupin zajema 10 let. Pri vseh starostnih skupinah je popularna rock glasba. Pri prvih štirih je celo na prvem mestu. Tudi pop glasba se pojavi pri vseh skupinah razen zadnjih 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 zadnjih dveh skupinah se pojavi na prvem mestu. V teh skupinah so popularne še folk, jazz in country. Srednji dve skupini pa sta mešanici obojega. 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 pa postrežem še nekaj zanimivosti. Ugotovili smo, da osebe, ki se z glasbo ukvarjajo več kot 15 let najraje poslušajo klasično glasbo. Osebe, ki





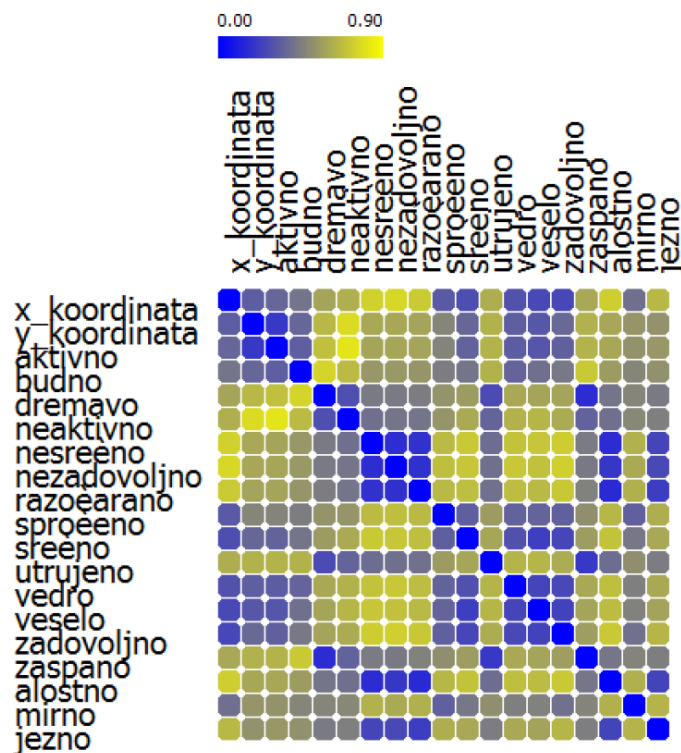
Slika 3.8: Priljubljenost posameznih glasbenih zvrsti po starostnih skupinah.

jemljejo zdravil sploh ne poslušajo klasične glasbe. Zanimivo je tudi, da pri ljudeh, ki uživajo droge bolj popularna elektronska glasba, kot pri ostali populaciji.

### 3.3.2 Razpoloženje, čustva ter 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 anketirančevo 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 čustvo pri njem. Zanimala nas je predvsem korelacija med prisotnostjo posameznih čustev in med koordinatama v VA prostoru. Kot je razvidno iz slike 3.9 je prijetnost (koordinata x) najmanj oddaljena od čustev: sproščenost, sreče, zadovoljstvo, vedrost in veselje. Koordinata y pa po pričakovanjih najboljše korelira z čustvom aktivnost. Z koordinato x se najmanj ujema s čustvi nesrečno, nezadovoljno in razočarano, kar je bilo pričakovati, saj

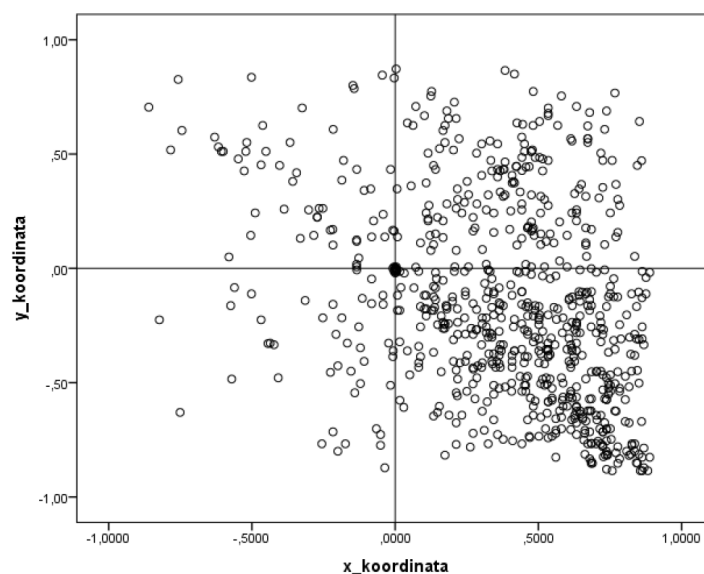
je prijetnost (predstavlja jo koordinata x) ravno nasprotna z temi čustvi. Z koordinato y se pričakovano najmanj ujema čustvo neaktivno zaradi enakega razloga kot pri koordinati x. Med čustvi lahko opazimo še dobro ujemanje med čustvoma nezadovoljno in razočarano ter med čustvoma veselo in zadovoljno.



Slika 3.9: Povprečna evklidska razdalja med posamezni atributi.

Slika 3.10 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 3.10: Razporeditev razpoloženja v VA prostoru.

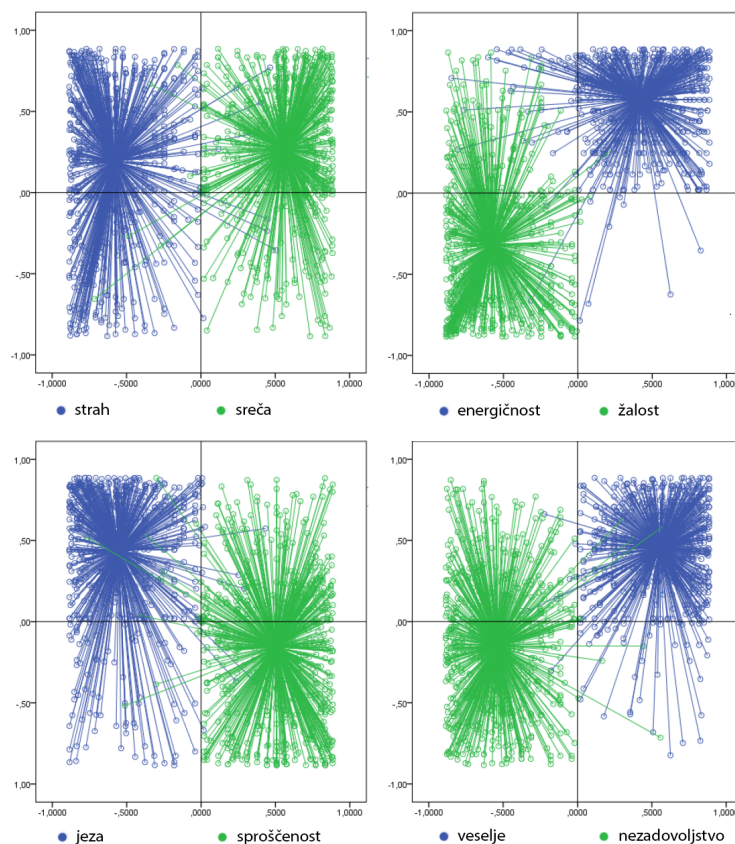
V anketi smo povpraševali tudi o tem kako uporabniki dojemajo posamezno razpoloženje. Razpoloženje so morali opisati tako, da so posamezno oznako postavili v VA prostor. Na sliki 3.11 je prikazana umestitev 8 čustev v VA prostor.

Slika levo zgoraj prikazuje primerjavo med čustvoma strah in sreča. Strah je zelo neprijetno čustvo, sreča pa prijetno. Oba sta razporejena praktično preko celotne y osi, kar si lahko razlagam s tem, da sta to čustva, 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 čustvo, obenem pa je žalost neprijetna in neaktivna. Zanimivo je, da je kar nekaj uporabnikov označilo, da je žalost aktivno čustvo.

Na tretjem grafu (levo spodaj) je narejena primerjava med jezo in sproščenostjo. Jeza je po pričakovanjih zelo aktivno in negativno čustvo, 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.

Na četrti sliki se vidi, da je veselje zelo pozitivno in aktivno čustvo, medtem ko je nezadovoljstvo neprijetno. Po aktivnosti je nezadovoljstvo malo bolj razpršeno, kar govori o tem, da je ta parameter pri tem čustvu težje določiti.



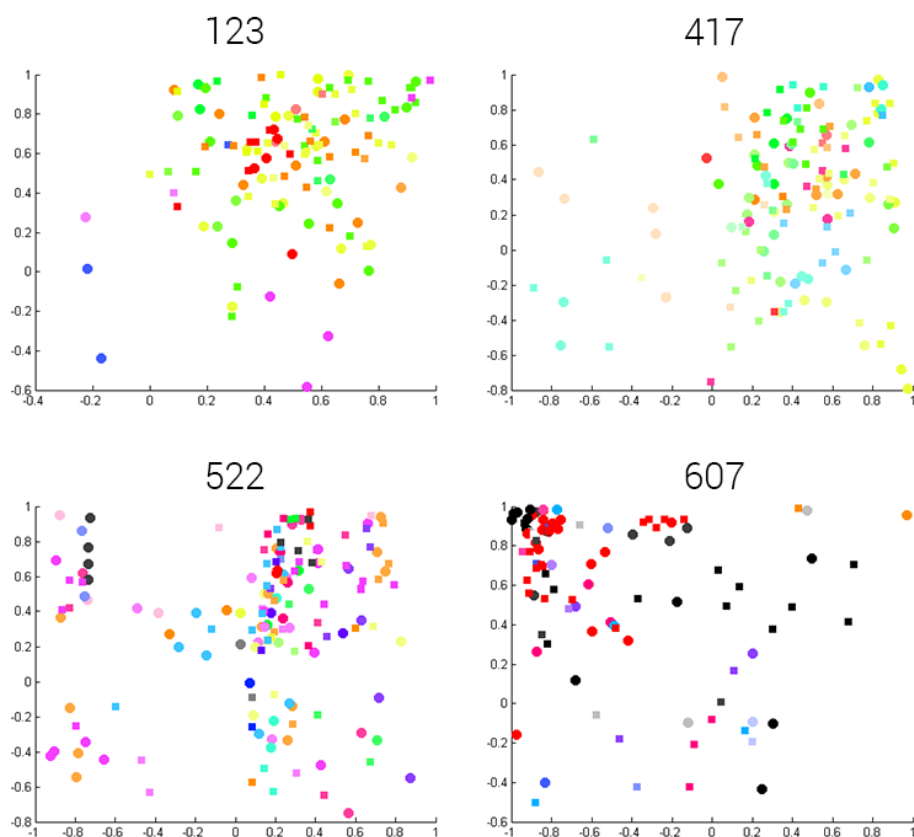
Slika 3.11: Prikaz čustev v VA prostoru.

**Barva** Ugotovili smo, da obstaja zelo dobra korelacija med razpoloženjem in izbranim pozicijo razpoloženja v VA prostoru. Za prijetnejša razpoloženja so izbrane barve bolj intenzivne. Pri prijetnih čustvih se barva v večini primerih giblje med rdečo in nežno modro. Za bolj aktivna razpoloženja so anketiranci izbrali manj intenzivne barve. Odtenki pa se v tem primeru gibljejo med modrim, zelenim in vijoličnim tonom.

Če pogledamo še izbire po čustvih so bile 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 in črno barvo. Za sproščenost je zanimivo, da je dokaj enakomerno razporejena preko celotnega barvnega spektra. Torej si to čustvo vsak udeleženec predstavlja drugače. Pri jezi

je bila pogosto izbrana barva modra ali vijolična. Izbirali so praktično med vsemi intenzivnostmi.

### 3.3.3 Glasba, razpoloženje in barva



Slika 3.12: Prikaz čustev v VA prostoru.

Za konec sem izbral še 4 glasbene odlomke in v sliki 3.12 prikazal rezultate. Predstavil sem VA vrednosti, kamor so anketiranci postavili čustva, ki nastopajo v posamezni pesmi. Z krogi so označena mesta kamor so postavili čustva, ki jih po njihovem mnenju glasba vzbudi pri človeku. Z kvadrati so predstavljena

čustva, ki jih pesem izraža. Obenem sem vizualiziral tudi barve, ki po mnenju uporabnikov izražajo v posamezni pesmi.

Pri prvi pesmi nastopajo čustva, 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, saj niso tako intenzivne, kar se ujema z bolj aktivnimi čustvi. Iz grafa se opazi, da so čustva, ki jih pesem vzbudi malo bolj razpršena, kot tista, ki jih pesem izraža.

Pri drugi pesmi (desno zgoraj) so čustva 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. Tudi tukaj so izražena čustva maj 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 čustva. Je pa zanimivo, da se pojavijo 4 gručice odgovorov. Prva je skrajno levo zgoraj, druga levo spodaj, tretja največja je desno zgoraj in četrta levo spodaj. Zanimivo je, da so odločitve o izraženih čustvih 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 čustva, 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 čustva bolj razpršena kot tista, ki jih glasba vzbudi. Tudi barve pri tem grafu s primerne izbranim čustvom, saj so bolj intenzivne in predvsem v rdeče črnih odtenkih.

## Poglavje 4

# Algoritmi za evalvacijo čustev iz glasbe

V diplomski nalogi sem si zadal, da preizkusim delovanje nekaj algoritmov za evalvacijo čustev iz glasbe. V tem poglavju bom predstavil delovanje in rezultate za nekaj algoritmov.

### 4.1 Regresijski algoritem

Prvi algoritem, ki sem ga preizkusil je algoritem predstavljen v [19]. Ni mi uspelo dobiti izvirnega algoritma avtorja. Sem pa po informacijah, ki sem jih pridobil ob natančnem pregledu avtorjevega članka in nekaj izmenjanih e-poštah z avtorjem, sam implementiral algoritem, ki je ga je avtor uporabil v svojem delu. Algoritem iz glasbe izračuna značilnice. Na podlagi teh pa potem napove prijetnost (valence) in aktivnost (arousal) v skladbi.

Na začetku bom predstavil delovanje algoritma. Sledijo še rezultati.

#### 4.1.1 Delovanje algoritma

Algoritem za sovrje delovanje potrebuje značilnice izračunane na podlagi glasbenih odlomkov. Uporabili smo značilnici MFCC [ref] in Chroma (kromogram) [ref]. MFCC je bil izračunan z koeficientom 20, Chroma pa je bila izračunana tako, da smo dobili 14 vrednosti (bins) za posamezen odlomek. Za računanje teh vrednosti

smo uporabili Python knjižnico imenovano 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 tako 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).

$$y = Xb \tag{4.1}$$

$X$  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.

Programska koda je priložena v prilogi [TODO: dodaj prilogo in referenco]

### 4.1.2 Rezultati

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 vsako značilnico (MFCC in Chroma) 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 pa smo izrazili



tudi v večkratniku standardnega odklona v podatkih iz podatkovne zbirke. Za primerjavo smo algoritem preizkusili tudi na Mood Swing podatkovni zbirki [20].

Tabela 4.1: Primerjava rezultatov dobljenih z regresijskim algoritmom na naši podatkovni zbirki in Mood Swing podatkovni zbirki z uporabo značilnic MFCC in Chroma. 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	Chroma
Naša podatkovna zbirka		
Do povprečne vrednosti	0.2060	0.2215
Do najbližje vrednosti	0.0595	0.0614
Do povprečne vrednosti v std. odkl.	0.4870	0.4993
Mood Swing podatkovna zbirka		
Do povprečne vrednosti	0.2448	0.3316
Do najbližje vrednosti	0.0641	0.1026
Do povprečne vrednosti v std. odkl.	0.6514	0.8940

Rezultati kažejo boljšo korelacijo med MFCC in prijetnostjo (valence) ter aktivnostjo (arousal) v primerjavi z Chromo. Prav tako lahko opazimo, da so podatki boljši na naši podatkovni zbirki kot na Mood Swing podatkovni zbirki. Verjetno gre to pripisati predvsem večjemu številu VA vrednosti v podatkovni zbirki.

Algoritem smo preizkusil tudi na Chroma značilnicah izračunanih s hierarhičnim kompozicionalnim modelom predstavljen v Pesek et. al [17]. Uporaba te značilnice nam da boljše rezultate, kot jih dobimo z MFCC in Chromo. 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 z prijetnost (valence) in aktivnost (arousal). Iz teh rezultatov je razvidno, da z uporabo vseh značilnic možno bolj natančno napovedati aktivnost v primerjavi s prijetnostjo.

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

	MFCC	Chroma	HKM Chroma
Prijetnost (valence)	0.1734	0.1826	0.1494
Aktivnost (arousal)	0.0871	0.0940	0.0898

Med tem je možno še opaziti, da Chroma izračunan z Hierarhični kompozicijskim modelom da veliko boljše rezultate pri računanju prijetnosti kot ostale značilnice, med tem ko je aktivnost izračunana podobno natančno kot z uporabo MFCC.

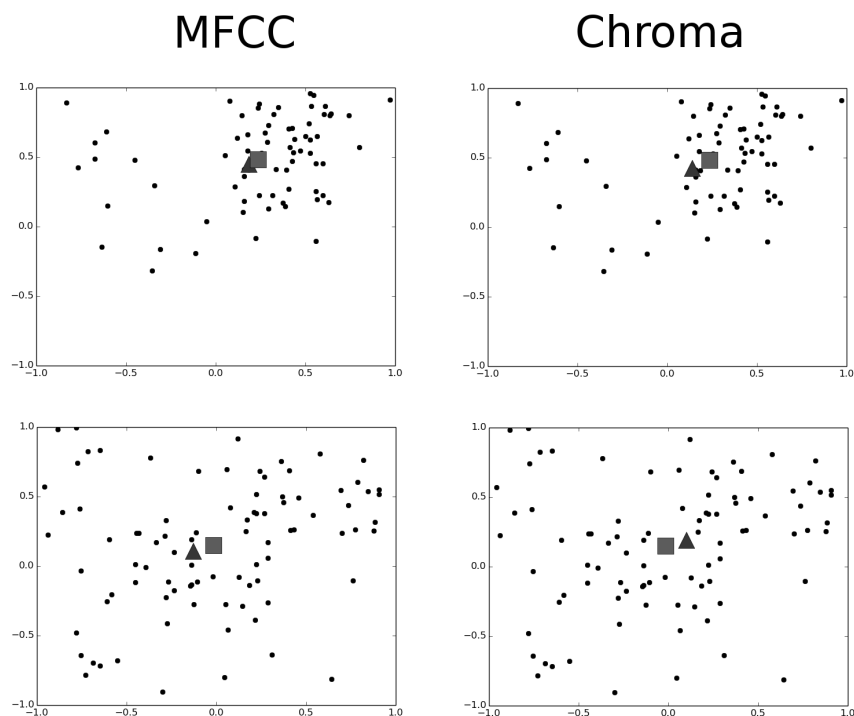
## 4.2 Essentia Gaiatransofrm algoritem

Drug algoritem na katerem sem preizkusil 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 mi jo je po kar nekaj problemih uspelo naložiti na moj računalnik.

V tem delu bom opisal delovanje algoritma, njegovo natančnost in tudi rezultate na naši podatkovni zbirki.

### 4.2.1 O essentii

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.



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

Ugotoviti poskuša celo ali gre za inštrumental ali ne.

Ker od v tem delu preučujem predvsem razpoloženje v glasbi se bom osredotočil na funkcionalnost, ki poskuša odkriti razpoloženje v glasbi. V naslednjem poglavju sem bom osredotočil na kakšen način deluje algoritem za izračun visoko nivojskih značilnic. Največjo pozornost bom dal predvsem izračunu razpoloženja.

### 4.2.2 Delovanje algoritma

Za izračun visoko nivojskih značilnic sem uporabil algoritem imenovan GaiaTransfor opisan v [4]. Ta visoko nivojske značilnice izračuna na podlagi nizko nivojskih, ki so prav tako izračunane na s knjižnico Essentia. Za določanje visoko nivojskih značilnic seveda ne uporabi vseh značilnic ampak samo podmnožico teh. To podmnožico so 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 pristnosti 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 evalvacijo razpoloženja iz glasbe so trenirali na zbirki, ki so jo pripravili Laurier et. al [13]. Zbirko so potem sami še dodatno označili.

Za razpoloženje algoritem vrne verjetnost, da je v neki pesmi prisotno eno od šestih čustev: srečen (happy), žalosten (sad), agresiven (aggressive), sproščen (relaxed), akustičen (acoustic), elektronski (electronic) in zabaven (pary). Poleg tega pa algoritem glasbeni odlomek glede na čustvo razvrsti v enega od 5 razredov (clousters), ki so uporabljeni v MIREX tekmovanju (MIREX task). Ti razredi so definirani v tabeli 2.1.

### 4.2.3 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 čustvo sproščen (92,92% natančno). Sledijo jima zabaven (88,38%), žalosten (86,96%) in elektronski (84,59%). Najslabše napovedi algoritem da za čustvo vesel, kjer je natančnost 82,86%.

Povsem drugačno je stanje pri natančnosti napovedovanja čustev v gruče. V tem primeru je natančnost le 59.14%. Iz tabele 4.3 je razvidno, da algoritem

Tabela 4.3: Natančnost napovedovanja v gruče v odstotkih. V navpičnem stolpcu so gruče katerim pesem dejansko 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

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 praktično 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 gruče 5 določi gručo 1. Obratna korelacija ni tako močna.

#### 4.2.4 Rezultati

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 imam algoritem vrača kot rezultat. Zaradi tega je bilo potrebno nekako poenotiti oznake. Ugotovil sem, da bo primerjava najlažja, če uporabim MIREX gruče, ki jih vrača algoritem. Tako sem oznake, ki jih imamo v naši podatkovni zbirki razvrstil v gruče glede na podobnost s čustvi, ki se nahajajo v teh gručah in so prikazani v tabeli 2.1. Razporeditev čustev iz naše podatkovne zbirke v gruče je prikazana v tabeli 4.4.

Primerjavo sem potem izvedel tako, da sem pogledal katero je najbolj pogosto razpoloženje v podatkovni zbirki pri pri vsakem izmed 200 glasbenih odlomkov. To razpoloženje sem potem glede na preslikavo v tabeli 4.4 preslikal v gručo. Potem sem te gruče primerjal z gručami, ki jih za isti glasbeni odlomek napove algoritem.

Natančnost napovedovanja na naši podatkovni zbirki je 47,50%. Bolj natančni

Tabela 4.4: Razvrstitev oznak razpoloženja iz naše zbirke v MIREX gruča glede na podobnost z obstoječimi čustvi iz 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

Tabela 4.5: Rezultati napovedovanja v gruča v odstotkih. V navpičnem stolpcu so gruča katerim pesem dejansko pripada, v vodoravnem pa gruča kamor je algoritem uvrsti pesem.

	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

rezultati napovedovanja so prikazani v tabeli 4.5. Vsaka vrstica predstavlja gručo v katero paše pesem. Stolpci pa predstavljajo v katero gručo 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 elementov uvrščenih v gručo v stolpcu glede na število elementov v dejanski gruč.

Iz teh podatkov je razvidno, da je algoritem najbolj natančno razvrstil elemente v gručo 3 med tem, ko je bil najmanj natančen pri določanju gruč 1. Tukaj so rezultati slabi, saj je algoritem večino pesmi iz te gruč uvrstil v gručo 3. Mogoče je tukaj eden od problemov v tem, da je v tej gruč malo pesmi, zato ne moramo

posplošiti teh rezultatov. V četrti vrstici so vsi rezultati nič, ker v naši podatkovni zbirki nimamo čustva, 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 4.2.3. 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 enako pogosto v naši podatkovni zbirki, sem potem primerjavo dopolnil tako, da primerja vsa čustva, ki so blizu razpoloženja, ki se pojavi največkrat. Če je gruča enega od teh razpoloženj enaka gruči v katero je algoritem uvrstil glasbeni odlomek smatram napoved kot pravilno.

Tabela 4.6: Rezultati napovedovanja v gruče v odstotkih v primeru, smo upoštevali tudi čustva, ki so blizu največkrat označenem. V navpičnem stolpcu so gruče katerim pesem dejansko pripada, v vodoravnem pa gruče kamor je algoritem uvrsti pesem.

	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 4.7 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 sem že omenil algoritem poleg klasifikacije v MIREX gruče omogoča evalvacijo za 6 razpoloženj. Za vsako od razpoloženj napove ali se pojavi v glasbenem

odlomku. Glede na to, da se 3 od teh oznak ujema z oznakami v naši podatkovni zbirki, sem iz podatkovne zbirke vzel tiste pesmi, kjer je to razpoloženje najbolj prisotno od vseh. Te pesmi sem potem primerjal z rezultati, ki jih vrne algoritem.

Tabela 4.7: Natančnost napovedovanja razpoloženj vesel, žalosten in sproščen v odstotkih.

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

Za ta primer algoritem napove pravilno prisotnost čustva s 79,59% natančnostjo. Natančnost je velika pri čustvih žalosten in sproščen. Dosti manjša pa je pri čustvu srečen. Kar kaže na to, da je algoritmično direktno iz glasbenih odlomkov dosti lažje napovedati sproščenost in žalost, medtem ko je to težje pri čustvu 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 čustva pa so rezultati približno enaki tisti, ki so jih dobili avtorji za čustvi žalosten in sproščen. Za čustvo 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] Rainer Bausch. Root mean square. *Journal of Chemical Education*, 73(12):A295, 1996.
- [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, 09/2013 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] 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.
- [7] Tim Dalgleish and Michael J Power. *Handbook of cognition and emotion*. Wiley Online Library, 1999.
- [8] Tuomas Eerola and Jonna K Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 2010.

- [9] 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.
- [10] Steinar Heimdal Gunderson. Musical descriptors: An assessment of psychoacoustical models in the presence of lossy compression. 2007.
- [11] Bernie Krause. *The Great Animal Orchestra: Finding the Origins of Music in the World's Wild Places*. Hachette Digital, Inc., 2012.
- [12] 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.
- [13] 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, 2010.
- [14] Mira Lenko, Tina Kogovšek, and Peter Stanković. *Pomen glasbe v očeh mladih: diplomsko delo*. M. Lenko, 2009.
- [15] R Panda, R Malheiro, B Rocha, A Oliveira, and RP Paiva. Multi-modal music emotion recognition: A new dataset, methodology and comparative analysis. *Proc. CMMR*, 2013.
- [16] Matevž Pesek. *Prepoznavanje akordov s hierarhičnim kompozicionalnim modelom: diplomsko delo*. PhD thesis, M. Pesek, 2012.
- [17] 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*, volume 2013, 2013.
- [18] 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.

- 
- [19] 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.
  - [20] Erik M Schmidt and Youngmoo E Kim. Modeling musical emotion dynamics with conditional random fields. In *ISMIR*, pages 777–782, 2011.
  - [21] 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.
  - [22] 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.
  - [23] Nils Lennart Wallin, Björn Merker, and Steven Brown. *The origins of music*. MIT press, 2001.
  - [24] 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.
  - [25] Bin Wu, Simon Wun, Chung Lee, and Andrew Horner. Spectral correlates in emotion labeling of sustained musical instrument tones. In *ISMIR*, pages 415–420, 2013.