

Uporabniški vmesniki in metodologija pridobivanja večmodalnih podatkov o glasbi

Matevž Pesek¹, Gregor Strle², Matija Marolt¹

¹Univerza v Ljubljani, Fakulteta za računalništvo in informatiko,

²Znanstvenoraziskovalni center slovenske akademije znanosti in umetnosti

¹ E-mail: matevz.pesek@fri.uni-lj.si

Povzetek. Raziskave, ki se ukvarjajo s priporočanjem glasbe in glasbeno vizualizacijo, pogosto črpajo podatke iz podatkovnih zbirk, ki vsebujejo podatke, pridobljene z anketiranjem uporabnikov. Kljub pomembnosti postopka zbiranja uporabniških podatkov so omenjene raziskave redkeje osredinjene na metodologijo zbiranja podatkov in uporabniške vmesnike. Zadnji igrajo ključno vlogo pri graditvi glasbenih podatkovnih zbirk in evalvaciji algoritmov za pridobivanje informacij iz glasbe. Članek predstavlja temeljne elemente postopka graditve glasbene podatkovne zbirke *Moodo*, ki poleg demografskih podatkov, uporabnikovega počutja in splošnega zaznavanja čustev vsebuje tudi uporabnikove čustvene in vizualne odzive na glasbo. V ta namen sta bila razvita nova uporabniška vmesnika *MoodStripe* in *MoodGraph*, ki presemeta omejitve klasičnih modelov. Rezultati raziskave kažejo na izboljšano uporabniško izkušnjo, tako glede intuitivnosti kot funkcionalnosti, predstavljena vmesnika pa sta aplikativna širše, predvsem na področjih, ki pri zbiranju podatkov merijo uporabnikov odziv.

Ključne besede: komunikacija človek—računalnik, pridobivanje informacij iz glasbe, zbiranje uporabniških podatkov, uporabniški vmesniki

User interfaces and methodology for gathering multimodal data about music

Several studies dealing with music recommendation and visualization base their approaches on datasets gathered with user surveys. However, the gathering procedure is seldom the focus of music research, even though the user interfaces and methodology are an important part of gathering the music data and evaluation of the music information retrieval algorithms. The paper presents the main elements of gathering the *Moodo* dataset that combines the demographic data, the users' mood and perception of emotions with the users' emotional and color responses to music. For this purpose, two novel user interfaces were developed, i.e. the *MoodStripe* and *MoodGraph*, which have several advantages over the existing classical models, both in terms of intuitiveness and functionality. The proposed interfaces are also applicable to other domains dealing with the user data.

1 UVOD

Raziskave človekovega zaznavanja glasbe so interdisciplinaren izziv, ki povezuje raziskovalna področja psihologije, kognitivne znanosti, komunikacije človek-računalnik, strojnega učenja in pridobivanja informacij iz glasbe (angl. music information retrieval - MIR). Opirajo tako teoretična kot praktična vprašanja o zaznavanju glasbe [1], z vidika MIR pa so aktualne predvsem zaradi možnosti nadgradnje obstoječih priporočilnih sistemov v smeri uporabniku prilagojenih poizvedb in

interakcij [2], [3]. Poslušanje glasbe je lahko izjemno individualna in čustveno navdana izkušnja, odvisna od osebnostnih lastnosti posameznika, glasbenih preferenc, trenutnega počutja, kot tudi bolj splošnega konteksta (starost, spol, izobrazba, socialni in kulturni vpliv, itn.), zato je način, kako merimo uporabnikovo zaznavanje, še posebno pomemben. Trenutne MIR raziskave so osredinjene na integracijo omenjenih parametrov, njihov namen pa je razvoj algoritmov za računalniško analizo glasbe in njihovo implementacijo v priporočilne sisteme. Pri tem sta pomembni predvsem učinkovita in intuitivna predstavitev in interakcija z vsebinami. Eden glavnih izzivov pri tem pa je zaobseganje kompleksnosti interakcij za graditev sistema, ki bo prilagojen preferencam in trenutnemu počutju posameznega uporabnika.

Članek predstavlja uporabniške vmesnike in metodologijo zbiranja podatkov o uporabnikovih čustvenih in vizualnih odzivih na glasbo, ki smo jih uporabili pri graditvi glasbene podatkovne zbirke *Moodo*. Drugo poglavje predstavlja ozadje in trenutne trende raziskav na področju pridobivanja informacij iz glasbe, glasbene vizualizacije in komunikacije človek-računalnik. Tretje poglavje opisuje metodologijo zbiranja podatkov in zasnovo uporabniških vmesnikov za modeliranje večmodalnih interakcij za potrebe glasbene zbirke *Moodo*. Predstavljena sta postopka preliminarne analize nabora čustvenih oznak in raziskave zaznavanja glasbe. Četrto poglavje obsega analizo zbranih podatkov, raziskavo uporabniške izkušnje in rezultate evalvacije novih

uporabniških vmesnikov *MoodStripe* in *MoodGraph* ter primerjavo s klasično metodo zbiranja podatkov. V zaključnem, petem poglavju avtor razpravlja o možnostih nadaljne uporabe novih vmesnikov.

Članek prinaša:

- celovit pristop k načrtovanju metodologije za zbiranje podatkov o uporabnikih;
- nova grafična uporabniška vmesnika *MoodStripe* in *MoodGraph*, ki po intuitivnosti in uporabniški izkušnji presegata obstoječe vmesnike za zbiranje podatkov o uporabnikih.

2 PREGLED PODROČJA

Pregled področja zajema problematiko pridobivanja informacij iz glasbe s poudarkom na glasbenih priporočilnih sistemih in glasbeni vizualizaciji ter splošnih izzivih s področja komunikacije človek—računalnik.

2.1 Pridobivanje informacij iz glasbe

Raziskave pri pridobivanju informacij iz glasbe (MIR) se primarno ukvarjajo z algoritmi za analizo podatkov iz glasbe. Med drugimi je eden od ciljev področja tudi graditev glasbenih priporočilnih sistemov. Ti lahko delujejo na podlagi različnih glasbenih parametrov. Sistem lahko glasbo priporoča glede na podobnost trenutno predvajanega posnetka z drugimi, izračunano na podlagi značilnic izločenih iz posnetkov. Posnetki se lahko primerjajo tudi na ravni metapodatkov, kjer sistem priporoča skladbe, ki so medsebojno podobne glede na izvajalca, obdobje ali zvrst. Čedalje pogostejši so tudi sistemi, ki primerjajo družabne oznake (angl. social tags), ki poleg pogosto vključenih oznak o žanru vsebujejo tudi podrobnosti o glasbenih entitetah, kot so: način izvedbe skladbe, tip vokala, uporabljeni inštrumenti in situacijske opise, npr. glasba za ozadje, telesne vaje ali delo. Prav tako so v oznakah pogosto prisotne čustvene oznake, ki se navezujejo na zaznana čustva v glasbi (npr. melanholično), kot tudi čustva, vzbujena v uporabniku (npr. vznemirjenje, energično).

Prav čustva imajo pri izboru priporočil pomembno vlogo, zato je na področju MIR veliko aktivnih raziskav, ki se ukvarjajo z razvojem algoritmov za razpoznavo čustev v glasbi. Pri tem je potreba najprej definirati, kako lahko čustva sploh opišemo. Uporabnikovo počutje in trenutna čustva, kakor tudi čustveno označevanje glasbe, tipično zbiramo na dva načina: diskretno, kjer so čustva definirana kot diskretne kategorije, in zvezno, na podlagi prepletanja več dimenzij. Diskretni čustveni model meri intenziteto oz. prisotnost vsakega čustva posebej, ločeno od drugih čustev, tipično z večstopenjsko lestvico (npr. prisotno—neprisotno). Zvezni čustveni model pri merjenju čustev uporablja več dimenzij. Najbolj znan in pogosto uporabljen zvezni čustveni model je Russellov krožni model afekta (angl. Circumplex model of affect) [4]. V tem modelu so čustva predstavljena v dvodimenzionalnem koordinatnem prostoru prijetnosti (angl.

valence) in intenzitete (angl. arousal) — poimenovanem tudi "valence-arousal (VA) space" (v nadaljnjem besedilu VA- prostor). Izбира čustvenega modela ni trivialna in vpliva na rezultate modeliranja [5]. Za potrebe raziskav glasbene kognicije je bilo razvitih več različic Russellovega modela, na primer [6], [7], [8], [9], ki je v omenjenih raziskavah tudi najpogosteje uporabljen.

Evalvacija MIR-algoritmov navadno poteka na anotiranih zbirkah podatkov, ki so zbrane z anketiranjem uporabnikov. V preteklih letih je bilo zgrajenih več javno dostopnih baz. Eerola idr. [5] so pri graditvi podatkovne zbirke, ki vsebuje 360 anotiranih posnetkov filmske glasbe, uporabili tridimenzionalni zvezni model (dimenzije prijetnosti, intenzitete in napetosti). Podatkovna zbirka Mood Swings Turk vsebuje povprečno 17 anotacij za 240 posnetkov popularne glasbe; tu je bil uporabljen klasičen VA-model [10]. Prav tako so dvodimenzionalen VA model uporabili pri graditvi podatkovne zbirke MTV, ki vsebuje pet bipolarnih anotacij za 192 popularnih pesmi [11] s seznamom predvajanj na kanalu MTV. Glavna omejitev obstoječih glasbenih zbirk za potrebe MIR je predvsem pomanjkanje zadostnega števila uporabniških podatkov, ki se navezujejo na zaznavanje glasbe.

Poznamo tudi diskretne modele, ki v določenih vidikih presegajo različice Russellovega modela, na primer model GEMS [12] in zbirka čustev AllMusicGuide [13]. Zadnji naj bi bili primernejši za zajem vzbujenih čustev in so uporabljeni v več raziskavah [3], [14].

2.2 Glasbena vizualizacija

Raziskav na področju glasbene vizualizacije, ki poudarjajo uporabniku prilagojen pristop, je razmeroma malo [15]. Eden večjih izzivov je povezovanje avdio—vizualnih modalitet uporabniškega vmesnika. Čeprav glasbene vizualizacije uporabljajo barvne sheme za upodobitev razmerij med entitetami, trenutni sistemi večinoma ne uporabljajo barvnih kombinacij, ki bi temeljile na analizi človekovega zaznavanja barv in glasbe, marveč so povezave med modalitetama izbrane naključno ali pa na podlagi nekih splošnih razmerij. Posledično je še danes glavni poudarek na vizualizacijah raznovrstnih glasbenih in zvočnih značilnk, od nizkonivojskih spektralnih značilnk do visokonivojskih glasbenih vzorcev in glasbenih metapodatkov [16]. Glasbene vizualizacije te vrste lahko na grobo razdelimo v dve kategoriji: vizualizacije glasbenih parametrov (npr. harmonij, časovnih vzorcev in drugih glasbenih entitet) [17], [18], [19], [20] in vizualizacije virtualnih prostorov za pojasnjevanje razmerij med različnimi glasbenimi posnetki. Zadnji so primernesši za priporočilne sisteme in sisteme za analizo glasbe. Namen vizualizacij prostorov variira od aplikacij za virtualne glasbene knjižnice [21], [22], prilagoditve za mobilne naprave [23] do izdelave predogledov za predstavitev glasbenih datotek [24]. Nedavni izzivi, kot je npr. "The Grand challenge at

Mirex evaluation exchange initiative”* pa kljub vsemu kažejo na to, da se zanimanje na področju uporabniško usmerjene interakcije in glasbene vizualizacije v zadnjem času povečuje.

2.3 Uporabniški vidik: komunikacija človek—računalnik

Področje komunikacije človek—računalnik (angl. human computer interaction - HCI) se osredinja na funkcionalnost in uporabnost uporabniških vmesnikov ter zajema celotno uporabnikovo izkušnjo. Zahteva integracijo treh osnovnih vidikov: človeški vidik (uporabnik), tehnološki vidik (sistem) in interakcijo med obema (komunikacija). Področje se v zadnjem času razvija v smer večmodalnih arhitektur [25] in povezovanje različnih sorodnih panog, kot so računalniški vid in procesiranje signalov z raziskavami v psihologiji, kognitivni znanosti in umetni inteligenci.

V zadnjem času je čedalje več zanimanja za uporabniško usmerjen pristop, ki ne temelji zgolj na logični interakciji človek—računalnik, marveč poudarja subtilen (pogosto izmuzljiv in nelogičen), čustven del človeške komunikacije. To imenujemo ”čustveno računalništvo” [26] in pomeni aktualno temo uporabniško usmerjenih računalniških raziskav. Cilj čustvenega računalništva je simulacija vedenjskih vzorcev, osebnostnih in čustvenih značilnosti človeka z računalniškimi sredstvi: z opazovanjem, tolmačenjem in ustvarjanjem značilk čustvenega vpliva, vezanih na obdelavo informacij o uporabniku [27]. Danes je čedalje več primerov praktičnih aplikacij na različnih področjih, kot so navidezna resničnost, pametni nadzor, večmodalni vmesniki, nosljivi senzorji itn., s splošnim ciljem, da se izboljša interakcija človek—računalnik v bolj intuitivno, prilagojeno in prijetno uporabniško izkušnjo.

3 GRADITEV PODATKOVNE ZBIRKE *Moodoo*

V nadaljevanju predstavljamo metodologijo graditve podatkovne zbirke *Moodoo*, ki združuje podatke o uporabnikih in njihovih čustvenih in vizualnih odzivih na glasbo. Naš cilj je bil zbrati čim več čim bolj relevantnih podatkov, brez prevelikega bremena za anketirance, ki bi negativno vplivalo na kakovost zbranih podatkov.

Zbiranje podatkov smo začeli s preliminarno analizo, s katero smo definirali nabor čustvenih oznak, uporabljenih za označevanje čustev v zbirki. V nadaljevanju opisujemo osrednjo raziskavo, pri kateri predstavljamo tudi dva nova grafična uporabniška vmesnika za zajem podatkov: *MoodStripe* in *MoodGraph* in strukturo celotnega vprašalnika. Postopek zbiranja podatkov, vključno s preliminarno analizo, glavno raziskavo in naknadno evalvacijo vmesnikov, smo izvedli v slovenskem jeziku.

3.1 Preliminarna analiza: izbor čustvenih oznak in barv

Namen preliminarne analize je bil izbor relevantnih čustvenih oznak za glavno raziskavo o čustvenem in barvnem zaznavanju glasbe. Pregled literature in raziskav s področja glasbenega zaznavanja, predvsem v psihologiji in MIR, namreč ni dal konkretnih rešitev, saj do danes nista bila splošno sprejeta model in nabor čustev, primernih za označevanje glasbe - obstoječe zbirke čustvenih oznak so večinoma izbrane intuitivno, brez dodatne razlage. Težava je tudi v tem, da niso vsa čustva primerna za opisovanje glasbe (npr. gnev). Posledično smo lasten nabor čustvenih oznak črpali iz več dobro definiranih dimenzionalnih in diskretnih modelov s področja raziskav glasbe [28], [12], [13]. Ker smo raziskavo izvedli v slovenskem jeziku in za slovensko govoreče uporabnike, smo v izhodišču zbrali širši nabor slovenskih čustvenih oznak, ki so bile kot podlaga za preliminarno analizo.

V preliminarni analizi smo zbrali podatke 63 uporabnikov. Vprašalnik je bil sestavljen iz 48 čustvenih oznak, s katerimi so uporabniki ocenjevali prisotnost čustev na sedemstopenjski (Likertovi) lestvici, od popolnoma neprisotno/neaktivno (1), do močno izraženo/aktivno (7). Z metodo glavnih komponent (angl. principal component analysis - PCA) smo identificirali prve tri komponente prostora oznak, ki so pojasnjevale več kot 64 odstotkov variance v podatkih. V končno zbirko smo vključili 17 čustvenih oznak, ki so močno korelirale s prvimi tremi komponentami.

Preizkusili smo tudi najprimernejši način za izbor barve, ki opisuje posamezen glasbeni izsek, pri čemer smo uporabnikom ponudili možnost, da barvo izberejo s pomočjo zveznega barvnega kroga. Pozneje se je pri analizi pisnih opažanj uporabnikov (ob koncu ankete so uporabniki imeli možnost izražanja dodatnih opažanj) izkazalo, da je zvezni barvni krog preobsežen in za uporabnika prezahteven. Posledično smo razvili diskretni barvni krog z izborom 49 barv, kar je kompromis med kompleksnostjo in možnostjo izbire barve. Primernost izbranega diskretnega barvnega kroga smo v naknadnih raziskavah tudi evalvirali.

Pomemben rezultat preliminarne analize je bilo opažanje uporabnikov o samem vmesniku za opisovanje čustev. Standardna Likertova lestvica je za uporabnika premalo intuitiven in preveč naporen vmesnik, saj je s to metodo prisiljen izbrati intenzivnost posameznega čustva na ločeni lestvici, za vsako čustvo posebej. Zaradi preglednosti, poenostavitve in zmanjšanja nepotrebnega napora pri označevanju smo posledično razvili dva nova grafična vmesnika za čustveno označevanje: *MoodStripe* in *MoodGraph*, ki ju predstavljamo v naslednjem razdelku.

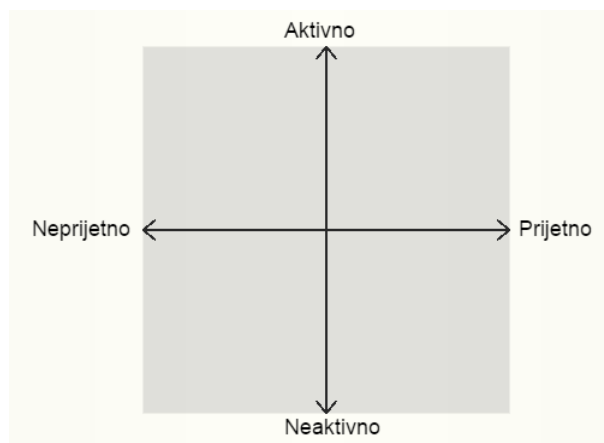
*<http://www.music-ir.org/mirex/wiki/2014:GC14UX>

3.2 Glavna raziskava: pridobivanje uporabnikovega zaznavanja čustev, barv in glasbe

Glavno raziskavo smo načrtovali ob podpori rezultatov preliminarne analize. Potekala je v obliki daljšega vprašalnika, razdeljenega na tri dele. Prvi del vsebuje osnovna demografska vprašanja, ki vključujejo vprašanja o uporabnikovem glasbenem predznanju in izkušnjah. Drugi del se nanaša na uporabnikovo občutje ter zaznavanje čustev in barv, tretji del pa obravnava čustveno in barvno zaznavanje glasbe.

3.2.1 Prvi del: Demografski podatki: Z namenom evalvacije vpliva posameznikovega ozadja na zaznavo čustev, barv in glasbe smo v prvem delu zbirali osnovne demografske podatke. Ta del (tabela 1) vsebuje tri demografska vprašanja, vprašanja o glasbenih izkušnjah posameznika in glasbenih preferencah in dve vprašanji o prisotnosti zdravil in drog, ki bi lahko vplivala na zaznavanje. Ker smo pri izpolnjevanju celotnega vprašalnika želeli doseči povprečen čas reševanja 15 minut, smo se posledično namenoma izognili večjemu naboru demografskih vprašanj, saj je zaradi vključevanja večih modalitet struktura raziskave že dovolj kompleksna.

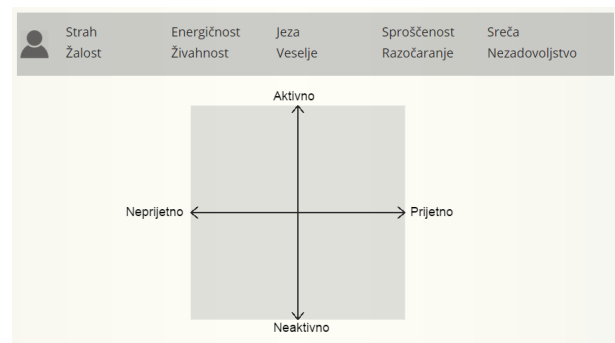
3.2.2 Drugi del: Podatki o trenutnem počutju in zaznavanju čustev in barv: Drugi del vprašalnika je osredotočen na pridobivanje informacij o uporabnikovem trenutnem počutju, njegovi zaznavi čustev in asociaciji čustev z barvami. Čustveno stanje je uporabnik sprva označil v dvodimenzionalnem VA—prostoru, glede na trenutno prijetnost in intenziteto čustvenega stanja (slika 1). V naslednjem koraku je uporabnik iz predstavljenega nabora čustvenih oznak vmesnika *MoodGraph* vsako posamezno čustvo označil v VA-prostoru, glede na njegovo dožemanje posameznega čustva (slika 3). Na koncu smo uporabnika povprašali še o trenutnem čustvenem stanju, tako da je z oznakami, ki jih je postavljajal v enodimenzionalen grafični vmesnik *MoodStripe* (slika 4), izrazil prisotnost posameznega čustva. Struktura vprašanj in reference na posamezen grafični vmesnik, uporabljen pri posameznem vprašanju, so predstavljeni v tabeli 2.



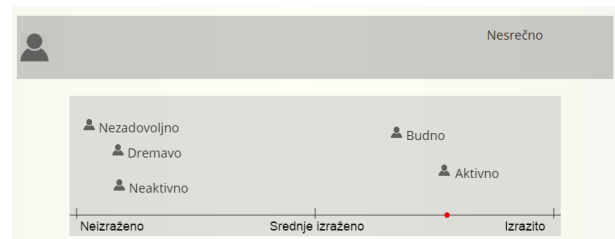
Slika 1: VA-prostor. Osi grafa označujejo prijetnost (abscisa) in intenziteto (ordinata) čustev.



Slika 2: Diskretni barvni krog z 49 barvami (črna pika označuje izbrano barvo). Vprašanja 2 in 5 v drugem delu vprašalnika se nanašata na izbor barve v vmesniku.



Slika 3: Grafični vmesnik *MoodGraph*: oznake čustev lahko po principu povleci in spusti (angl. drag-and-drop) uporabnik postavi v dvodimenzionalni VA-prostor.



Slika 4: Grafični vmesnik *MoodStripe*: zaradi večje zbirke oznak smo v vprašalniku uporabili tri instance tega uporabniškega vmesnika za izvedbo vprašanja 4. Dimenzija se razteza od odsotnosti čustva (primerljiva večstopenjska lestvica pri označbi 1) do izrazito izražena (lestvica pri najvišji stopnji) od leve proti desni.

Oba grafična uporabniška vmesnika: *MoodGraph* in *MoodStripe* nadomeščata zbirko standardnih vmesnikov, t. j. Likertovih ordinalnih večstopenjskih lestvic, implementiranih v obliki sistema gumbov (radio button), ki bi jih bilo treba implementirati za vsako čustveno oznako posebej. Glede na rezultate evalvacije vmesnikov (4.

Tabela 1: Demografska vprašanja. Posamezno vprašanje (prvi stolpec) vsebuje nabor mogočih odgovorov in dodatnih komentarjev (drugi stolpec). Odgovori na vprašanja so namenjeni odkrivanju potencialnih vplivov uporabnikovega ozadja na njegovo zaznavo.

| Vprašanje | Mogoči odgovori in komentarji |
|-------------------------------|--|
| Starost | <i>v letih</i> |
| Spol | {moški, ženska} |
| Kraj bivanja | {mesto, podeželje} |
| Obiskovanje glasbene šole | <i>v letih,</i> <i>0 - ni obiskoval/a</i> |
| Igranje inštrumenta ali petje | <i>v letih,</i> <i>0 - ni igral/a</i> |
| Uporaba zdravil in drog | <i>ali uporabnik uživa zdravila ali droge</i> |
| Vpliv zdravil in drog | <i>ali je uporabnik trenutno pod vplivom zdravil ali drog, ki bi lahko vplivali na njegovo razpoloženje</i> |
| Preferenca glasbenih žanrov | {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} - <i>uporabnik je lahko zbral do tri (a najmanj enega) najljubše žanre. Žanri so bili predstavljeni v angleškem jeziku zaradi splošne uporabe tujih oznak v slovenščini</i> |
| Dnevno poslušanje glasbe | {manj kot 1, 1-2, 2-3, več kot 3} - <i>v urah na dan</i> |

Tabela 2: Drugi del: Zbiranje podatkov o trenutnem počutju ter zaznavanju posameznih čustev in barv. Pri četrtem vprašanju smo zbirko čustvenih oznak zaradi obsežnosti razdelili na tri podvprašanja. Pri petem vprašanju je uporabnik dojemanje posamezne čustvene oznake določal v VA-prostoru (slika 1) kot tudi z barvo, ki po njegovem mnenju ponazarja čustvo (slika 2).

| Vprašanje | Mogoči odgovori in komentarji |
|--|--|
| Trenutno čustveno stanje v VA-prostoru | VA-prostor - <i>slika 1</i> |
| Barva trenutnega čustvenega stanja | Barvni krog - <i>slika 2</i> |
| Dojemanje čustvenih oznak | {Strah, Energičnost, Jeza, Sproščenost, Sreča, Žalost, Živahnost, Veselje, Razočaranje, Nezadovoljstvo} - <i>slika 3</i> |
| Trenutno čustveno stanje | {Aktivno, Budno, Dremavo, Neaktivno, Nesrečno, Nezadovoljno, Razočarano, Sproščeno, Srečno, Utrujeno, Vedro, Veselo, Zadovoljno, Zaspalo, Žalostno, Mirno, Jezno} - <i>slika 4</i> |
| Barve, asociirane s čustvi | {Žalost, Jeza, Sproščenost, Razočaranje, Sreča, Nezadovoljstvo, Veselje, Strah, Živahnost, Energičnost} - <i>slika 2</i> |

poglavje), sta se oba nova vmesnika iskazala za bolj intuitivna in funkcionalna.

3.2.3 *Tretji del: zbiranje podatkov o uporabnikovem dojetanju zaznanih čustev (v glasbenem posnetku) in vzbujenih čustev (v uporabniku) ter glasbeno-barvnih asociacij:* Tretji del vprašalnika se je nanašal na dva vidika dojetanja čustev v glasbi ter na barvne asociacije, ki jih v uporabniku vzbudi glasba. Uporabniku smo predvajali deset naključno izbranih 15-sekundnih posnetkov. Po poslušanju posameznega posnetka je uporabnik v barvnem krogu (vmesnik na sliki 2) izbral barvo, ki najbolje izraža konkreten glasbeni posnetek. V naslednjem koraku je uporabnik s pomočjo dvokategorne različice vmesnika *MoodGraph* (slika 5) izbiral posamezne čustvene oznake, in sicer med kategorijo

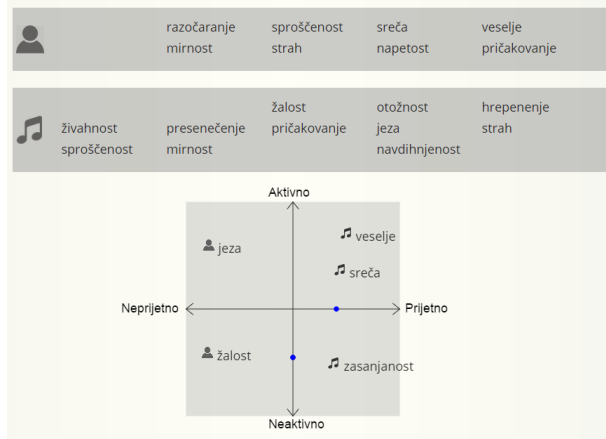
zaznanih čustev v glasbi (angl. perceived emotions; označenih z ikono glasbena nota) in kategorijo vzbujenih čustev ob poslušanju glasbe (angl. induced emotions; označenih z ikono uporabnik).

4 ANALIZA DEMOGRAFSKIH PODATKOV IN EVALVACIJA UPORABNIŠKE IZKUŠNJE

V nadaljevanju sta predstavljeni demografska analiza podatkov, zajetih v glavni raziskavi, in naknadna evalvacija uporabniške izkušnje in predlaganih uporabniških vmesnikov.

4.1 Demografski podatki

V spletni raziskavi je sodelovalo več kot 1100 udeležencev, predstavljamo pa rezultate analize 741



Slika 5: Dvokategorni grafični vmesnik *MoodGraph*: oznake čustev lahko po principu povleci in spusti (angl. drag and drop) uporabnik postavi v dvodimenzionalni VA-prostor. Oznake posamezne kategorije so označene z ikono kategorije (ikona uporabnik oz. ikona glasbena nota). Modri piki označujeta pozicijo izbranega čustva po posamezni dimenziji v VA-prostoru.

udeležencev, ki so v celoti rešili vprašalnik. Med temi je bilo 247 moških (33 %) in 494 žensk (67 %). Najmlajši je štel 15 let, najstarejši pa 64. Več kot 75 % udeležencev je bilo starih manj kot 30 let ($Q_3=28.45$ let), kar je pogojeno tudi z načinom izvedbe vprašalnika na spletu in uporabo socialnih omrežij za širjenje informacij o raziskavi.

Skoraj 60% moških, ki so sodelovali v raziskavi, ni nikoli obiskovalo glasbene šole; pri ženskah je delež manjši, le dobrih 44 odstotkov. Skoraj 12 % žensk in le 6 % moških je glasbeno šolo obiskovalo vsaj šest let. Glasbena izobrazba močno korelira z leti igranja inštrumenta in petja ($r=0.652$; $p<0.000$).

Med preferiranimi žanri je najbolj popularen rock, ki ga je na prvo mesto postavilo kar 31 % udeležencev. Sledi pop s 17 % in alternativa ter klasična glasba, vsak z dobrimi petimi odstotki. Preostali žanri so zbrali manj kot 5 % glasov. Podobno lahko opazimo pri drugem najljubšem žanru: rock je dosegel 20 %, pop pa 14 %. Varianca med žanri se pri drugem najljubšem žanru povečuje glede na prvega. Zanimivo pa je, da je klasična glasba najbolj popularna kot tretja najljubša zvrst pri 13 % udeležencev. Sledi rock (12 %) in pop (10 %). Izbira prvega najljubšega žanra korelira z leti (Spearman $\rho=-0.094$, $p=0.011$) in časom dnevnega poslušanja glasbe (Spearman $\rho=-0.111$, $p=0.002$). Celotna analiza ankete, vključno z drugim in tretjim delom, je predstavljena v Pesek idr. [29].

4.2 Evalvacija uporabniške izkušnje

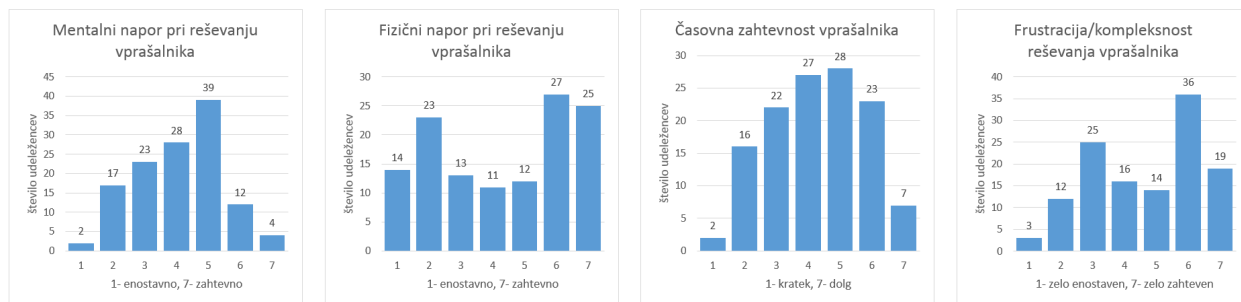
Po glavni raziskavi smo naknadno izvedli še raziskavo uporabniške izkušnje in evalvacijo uporabniških vmesnikov *MoodStripe*, *MoodGraph* in diskretnega barvnega

kroga. Uporabniki, ki so prisostvovali osnovni raziskavi, so ocenjevali več vidikov: uporabniško izkušnjo (angl. user experience - UX) [30], zahtevnost vprašalnika in različne vidike funkcionalnosti in primernosti grafičnih vmesnikov. Cilj naknadne raziskave so bile primerjava funkcionalnosti in uporabniške izkušnje novih in standardnih vmesnikov, tipično uporabljenih v raziskavah. Evalvacijo vmesnikov smo izvedli na podlagi vprašalnika NASA load task index [31] in dodatnih specifičnih vprašanj o vmesnikih, predstavljenih v poglavju 4.2.

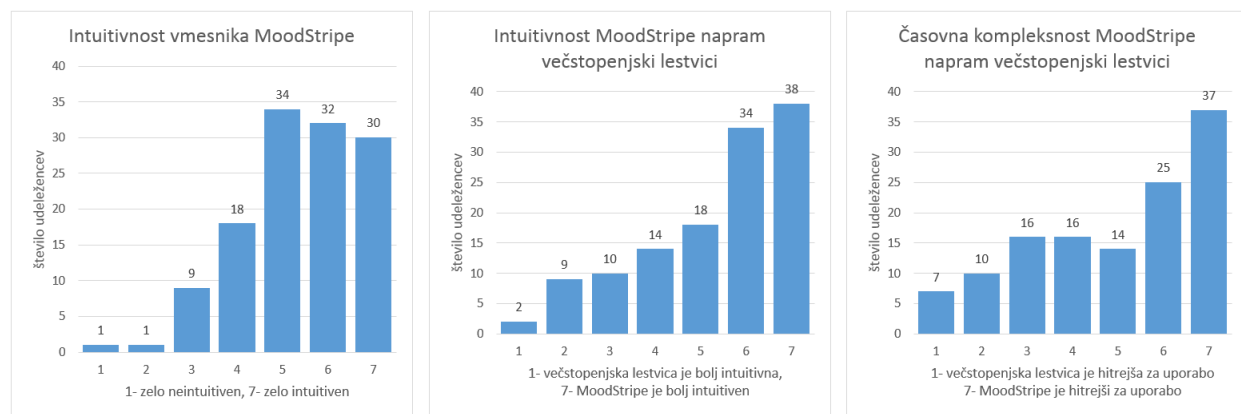
Pri evalvaciji uporabniških vmesnikov je sodelovalo 125 udeležencev. Vprašanja so skupaj z rezultati predstavljena na slikah 6, 7, 8, 9. Pri mentalni zahtevnosti vprašalnika (slika 6) rezultati nakazujejo na kompleksnost vprašalnika, kar je pričakovano glede na količino zahtevanih informacij, označevanja čustvenega zaznavanja in dolžino vprašalnika. Pri meritvi fizičnega napora se kaže morebitna nezadostna definicija vprašanja. Namen je bil zbrati informacije o fizičnem naporu (npr. premiki z miško so fizično bolj naporni kot izbira s tipkovnico) in posledično evalvirati operacijo povleci in spusti. Rezultat lahko nakazuje tudi na dve podmožici uporabnikov — tisti, ki opravijo večino dela s tipkovnico (in jim operacije povleci in spusti pomenijo fizični napor zaradi potrebe po uporabi miške), in tisti, ki preferirajo miško (in jim predstavljeni grafični vmesniki ne pomenijo dodatnega napora). Vsekakor je za utemeljen zaključek evalvacije fizičnega napora dano vprašanje problematično, saj način interakcije (miška ali tipkovnica) v vprašanju ni eksplicitno izpostavljen.

Na sliki 7 so predstavljeni rezultati meritev intuitivnosti vmesnika *MoodStripe*. Vmesnik je bil pri uporabnikih ocenjen kot izjemno intuitiven in hkrati časovno manj kompleksen kot klasična večstopenjska lestvica. Za ta vmesnik lahko z določeno mero gotovosti trdimo, da smo dosegli namen in je vmesnik primeren kot alternativa za večstopenjsko lestvico. Prav tako smo opazovali intuitivnost in časovno kompleksnost vmesnika *MoodGraph*, ki se je izkazal za intuitivnega, a pričakovano (glede na zahtevnost dvokategorijskega prepoznavanja in označevanja glasbenih posnetkov) tudi za časovno kompleksnega (slika 8).

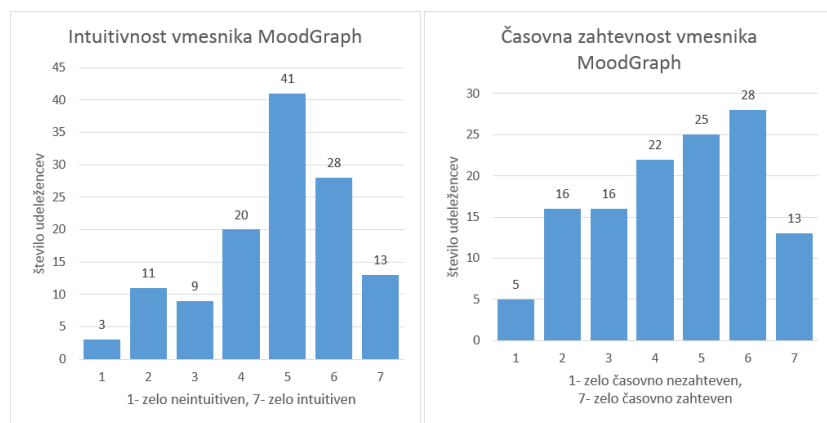
Uporabnike smo povprašali tudi po smiselnosti nabora čustvenih oznak in števila barv in za konec po oceni časovne zahtevnosti izpolnjevanja vprašalnika (slika 9). Ocena primernosti nabora čustvenih oznak nakazuje, da je čustvenih oznak morda preveč, a se rezultat nagiba k uravnoteženi porazdelitvi. Rezultati pri številu barv barvnega kroga nakazujejo na delno pomanjkanje dodatnih barv. Pri predpostavki omejitve časa reševanja vprašalnika na največ 15 minut smo se delno uštel, saj je za sorazmeren del udeležencev vprašalnik pomenil večjo časovno obremenitev.



Slika 6: Mentalna, fizična in časovna kompleksnot vprašalnika. Pričakovano je vprašalnik zahteval porazdeljen mentalni napor, fizični napor je nepričakovano bimodalno porazdeljen, časovna zahtevnost in kompleksnost reševanja pa sta rahlo nadpovprečni.



Slika 7: Intuitivnost vmesnika *MoodStripe*. Vmesnik je ocenjen kot izjemno intuitiven in časovno manj kompleksen kot klasična večstopenjska lestvica.

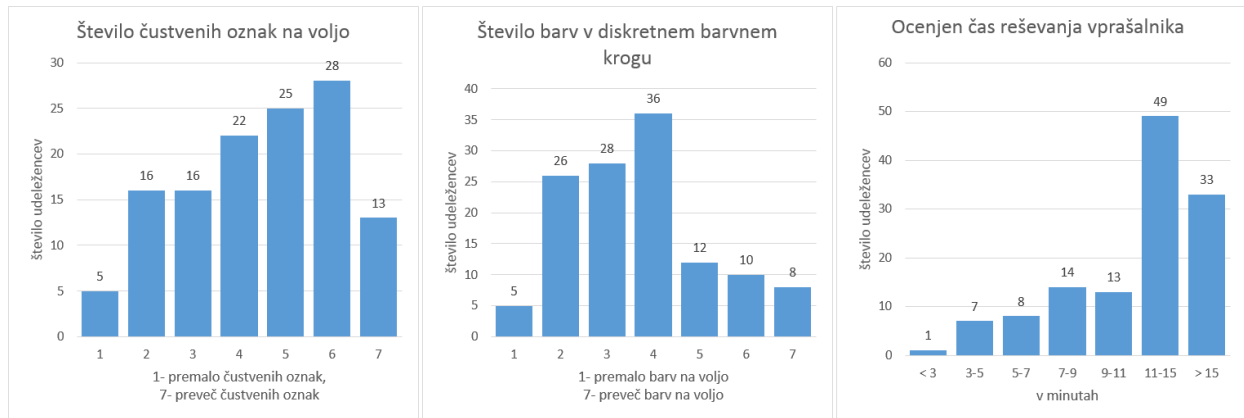


Slika 8: Ocena intuitivnosti in časovne zahtevnosti vmesnika *MoodGraph*. V tem primeru nismo izvedli primerjave s standardnimi vmesniki, saj je vmesnik kombinacija postavljanja točk v VA-prostor in oceno prisotnosti za vsako čustvo (glede na dve kategoriji), česar ni mogoče preprosto replicirati s klasičnim pristopom.

5 SKLEP

Podatkovna zbirka *Moodo* je temeljna podlaga za graditev personaliziranega priporočilnega sistema za glasbo. Z njo želimo preseči omejitve obstoječih glasbenih priporočilnih sistemov s povezovanjem avdio vizualnih modalnosti v odločitveni model in s povzemanjem trenutnega počutja posameznika, njegovih osebnostnih lastnosti in širšega konteksta.

V članku smo se osredinili na metodologijo zajemanja podatkov, ki smo jo izvedli v treh korakih, pri čemer smo za izboljšanje uporabniške izkušnje uvedli tudi dva nova grafična vmesnika: *MoodStripe* in *MoodGraph*. Z vmesniki smo izvedli zbiranje podatkov za podatkovno zbirko in jih nato evalvirali. Predstavljena metodologija je dala dobre rezultate, saj smo v glavni raziskavi zajeli več kot 1100 udeležencev, identificirali pa smo tudi



Slika 9: Ocena primernosti nabora čustvenih oznak, ki nakazuje, da je čustvenih oznak morda preveč, a se hkrati rezultat nagiba k uravnoteženi porazdelitvi. Po drugi strani bi si uporabniki želeli večji nabor barv, kot jih ponuja obstoječa različica diskretnega barvnega kroga. Meritev časovne zahtevnosti kaže, da je večina udeležencev je porabila več kot 11 minut časa za izpolnitev vprašalnika.

ključne smernice za nadaljnje izboljšave vmesnikov. Rezultati evalvacije vmesnikov *MoodStripe* in *MoodGraph* nakazujejo na večjo intuitivnost in izboljšano uporabniško izkušnjo glede na standardne vmesnike.

Predstavljeni vmesniki so uporabni širše, predvsem na področjih, ki se ukvarjajo z merjenjem uporabnikovega zaznavanja in povezovanja večih modalitet. Predvsem so to teme na področjih, kot so: psihologija, kognitivne in sociološko orientirane znanosti), kjer "temeljna resnica" ni striktno definirana, temveč je odvisna od odziva uporabnikov, ki jih zajamemo prek vprašalnikov. Evalvacija obeh uporabniških vmesnikov, *MoodStripe* in *MoodGraph*, potrjuje prednosti z vidika intuitivnosti in funkcionalnosti, ter splošne uporabniške izkušnje v primerjavi s klasičnim načinom zbiranja podatkov o uporabnikih. *MoodStripe* nam z metodo povleci in spusti omogoča preprosto označevanje več parametrov znotraj istega vmesnika, *MoodGraph* pa tudi možnost večkategoričnega označevanja v dvodimenzionalnem prostoru. S tem se močno izboljšajo sama interakcija s sistemom, preglednost označevanja in primerjanja posameznih parametrov, ki se izvaja v istem vmesniku, posledično pa tudi časovna zahtevnost celotnega procesa. Dolgoročni cilj je standardizacija vmesnikov *MoodStripe* in *MoodGraph*. V ta namen načrtujemo aplikacijo in evalvacijo uporabnosti obeh vmesnikov na drugih sorodnih področjih.

ZAHVALA

Raziskavo delno financira Evropski socialni sklad v okviru programa Inovativna shema - 2012.

LITERATURA

- [1] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behavioral and brain sciences*, vol. 31, no. 5, pp. 559–575, 2008.
- [2] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," in *Proc. 9th Int. Symp. Computer Music Modelling and Retrieval (CMMR)*, London, 2012, pp. 395–410.
- [3] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "Music emotion recognition: A state of the art review," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Utrecht, 2010, pp. 255–266.
- [4] J. A. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [5] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, Aug. 2010.
- [6] M. McVicar, T. Freeman, and T. De Bie, "Mining the Correlation Between Lyrical and Audio Features and the Emergence of Mood," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Miami, 2011, pp. 783–788.
- [7] Y. Yang and X. Hu, "Cross-cultural Music Mood Classification: A Comparison on English and Chinese Songs," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Porto, 2012.
- [8] C. Laurier, M. Sordo, J. Serrà, and P. Herrera, "Music Mood Representations from Social Tags," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2009, pp. 381–386.
- [9] M. Barthet, D. Marston, C. Baume, G. Fazekas, and M. Sandler, "Design and evaluation of semantic mood models for music recommendation using editorial tags," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Curitiba, 2013.
- [10] E. M. Schmidt and Y. E. Kim, "Modeling Musical Emotion Dynamics with Conditional Random Fields," in *ISMIR*, 2011, pp. 777–782.
- [11] B. Schuller, C. Hage, D. Schuller, and G. Rigoll, "'Mister DJ, Cheer Me Up!': Musical and textual features for automatic mood classification," *Journal of New Music Research*, vol. 39, no. 1, pp. 13–34, 2010.
- [12] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evo-

- ked by the sound of music: characterization, classification, and measurement,” *Emotion*, vol. 8, no. 4, p. 494, 2008.
- [13] X. Hu and J. S. Downie, “Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, 2007.
- [14] M. Barthelet, G. Fazekas, and M. Sandler, “Multidisciplinary perspectives on music emotion recognition: Implications for content and context-based models,” in *Proc. CMMR*, London, 2012, pp. 492–507.
- [15] M. Schedl, A. Flexer, and J. Urbano, “The neglected user in music information retrieval research,” *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 523–539, Jul. 2013.
- [16] J. Donaldson and P. Lamere, “Using Visualizations for Music Discovery,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2009, p. Tutorial.
- [17] E. Isaacson, “What You See Is What You Get: On Visualizing Music,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, 2005, pp. 389–395.
- [18] A. Mardirossian and E. Chew, “Visualizing Music: Tonal Progressions and Distributions,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, 2007, pp. 189–194.
- [19] H. Grohganz, M. Clausen, N. Jiang, and M. Mueller, “on-verting path structures into block structures using eigenvalue decompositions of self-similarity matrices,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Curitiba, 2013.
- [20] N. Jiang and M. Mueller, “Automated methods for analyzing music recordings in sonata form,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Curitiba, 2013.
- [21] M. Torrens, P. Hertzog, and J. L. Arcos, “Visualizing and Exploring Personal Music Libraries,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, 2004.
- [22] R. Van Gulik, F. Vignoli, and H. Van de Wetering, “Mapping Music In The Palm Of Your Hand, Explore And Discover Your Collection,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, 2004.
- [23] C. F. Julia and S. Jorda, “SongExplorer: A Tabletop Application for Exploring Large Collections of Songs,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Kobe, 2009, pp. 675–680.
- [24] K. Yoshii and M. Goto, “Music Thumbnailer: Visualizing Musical Pieces in Thumbnail Images Based on Acoustic Features,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, 2008, pp. 211–216.
- [25] P. T. Jose, S. Miglani, and S. Yadav, “Human Computer Interaction: Analysis and Journey through Eras,” *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 4, pp. 653–650, 2014.
- [26] R. W. Picard, *Affective Computing*. MIT Press, 2000.
- [27] J. Tao and T. Tan, *Affective computing: A review*. Springer Berlin Heidelberg, 2005, vol. 1, no. 1.
- [28] P. Ekman, “An argument for basic emotions,” *Cognition and Emotion*, vol. 6, pp. 169–200, 1992.
- [29] M. Pesek, P. Godec, M. Poredos, G. Strle, J. Guna, E. Stojmenova, M. Pogacnik, and M. Marolt, “Introducing A Dataset Of Emotional And Color Responses To Music,” in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Taipei, 2014, pp. 355–360.
- [30] W. Albert and T. Tullis, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Google eBook)*. Newnes, 2013.
- [31] S. G. Hart, “Nasa-Task Load Index (NASA-TLX): 20 Years Later,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006.

Matevž Pesek je doktorski študent na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Na omenjeni fakulteti je diplomiral leta 2012 in je tam zaposlen kot asistent. Je član Laboratorija za računalniško grafiko in multimedije, kjer raziskuje biološko navdahnjene modele, globoke arhitekture in kompozicionalno modeliranje. Prav tako se ukvarja z večmodalnim zaznavanjem in komunikacijo človek—računalnik in izdelavo vizualnih orodij za analizo glasbe.

Gregor Strle je asistent z doktoratom, zaposlen na Znanstvenoraziskovalnem centru SAZU. Raziskuje na področju kognitivnih znanosti, s poudarkom na glasbeni kogniciji, modelih konceptualizacije vsebin in računalniški semantiki.

Matija Marolt je docent na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, kjer je zaposlen od leta 1995. Je vodja Laboratorija za računalniško grafiko in multimedije, kjer raziskovalno deluje na področju pridobivanja informacij iz glasbe, podrobneje s semantičnim opisom in razumevanjem avdio-signalov ter pridobivanjem in organizacijo glasbenih arhivov in komunikacijo človek—računalnik.