Prodor globokih arhitektur na področje pridobivanja informacij iz glasbe

Matevž Pesek
Univerza v Ljubljani, Fakulteta za računalništvo in informatiko

Izvleček

S povečano popularnostjo globokih arhitektur, ki temeljijo na nevronskih mrežah, so se v zadnjem času bistveno izboljšali rezultati pri reševanju problemov na več področjih. Zaradi popularnosti in uspešnosti teh globokih pristopov, temelječih na nevronskih mrežah, so bili drugi simbolni in kompozicionalni pristopi odmaknjeni od središča pozornosti raziskav. V članku bomo obširneje pregledali delovanje globokih in hierarhičnih pristopov na področju pridobivanja informacij iz glasbe. Omenjamo nekatere od bolj znanih problemov na področju in izpostavimo probleme, pri katerih globoki pristopi, temelječi na nevronskih mrežah, še niso bili uspešno aplicirani. Kot alternativo takšnim pristopom predstavljamo kompozicionalni hierarhični model in opisujemo uporabnost modela in rezultate na predstavljenih problemih. Pregled bomo sklenili z diskusijo o prihodnosti uporabe globokih modelov glede na druge pristope.
Ključne besede: pridobivanje informacij iz glasbe, globoke arhitekture, kompozicionalni hierarhični modeli.

Abstract

With the increasing popularity of deep neural-based architectures, the results of deep architectures have been significantly improved recently in several areas. Due to the popularity and success of these deep approaches based on neural networks, other symbolic and hierarchical approaches are no longer the focus of researchers. In this article, we review the recent progress of deep and compositional approaches in the field of music information retrieval. Furthermore, we deliberate on the most notorious issues in the field and highlight problems where deep approaches based on neural networks have not yet been successfully applied. As an alternative to such approaches, we provide an overview of hierarchical models and describe the compositional hierarchical model as an alternative deep architecture. The latter shows great usability with the presented problems. We conclude this review with a discussion of the future of deep models compared to other approaches.
Keywords: Music information retrieval, deep learning architectures, compositional hierarchical models.

1 UVOD

Področje računalništva in informatike vključuje veliko kombinacij interdisciplinarnih pristopov, katerih namen je avtomatizirati obstoječe procese ali izumiti nove, ki pomagajo našim potrebam v vsakdanjem življenju. Računalniki so v večini naših dejavnosti že v celoti vključeni v izdelke končnih uporabnikov, kot so pametne televizije, kuhinjski aparati ali avtomobili, ki jih uporabljamo za shranjevanje, upravljanje, organiziranje in razvrščanje naraščajočih količin podatkov, ki jih zbiramo. Tako se je računalništvo dotaknilo tudi na videz nasprotnih področij, kot sta umetnost in glasba. Na glasbenem področju se je vez z informatiko stkala na več ravneh: v ustvarjanju glasbe, glasbeni organizaciji in glasbeni analizi.

Kot poskus analiziranja, pridobivanja in organiziranja glasbe se je v zadnjih dveh desetletjih utrdilo področje pridobivanja informacij iz glasbe (angl. music information retrieval – MIR) (Downie 2010). Od zgodnjih začetkov se je področje signifikantno razširilo in zajema številne teme od glasbene percepcije in pridobivanja informacij do razumevanja in analize glasbe. Področje MIR meji na več dobro uveljavljenih področij, kot so psihologija (npr. Gelfand, 2004; Ti-

rovolas, 2011), muzikologija (npr. Lerdahl, 1983; McDermott, 2008) in računalništvo (npr. de Cheveigne, 2002; Mauch, 2010; Tolonen 2000).

Del raziskav MIR obravnava pridobivanje semantičnih opisov glasbe v različnih oblikah. Tako kot na mnogih sorodnih področjih sta se v zadnjih letih znatno povečali natančnost in učinkovitost najboljših algoritmov, kot so ocenjevanje melodije (npr. Ryynanen, 2008; Bittner, 2015), ocenjevanje akordov (npr. Harte, 2005; Papadopoulos, 2007; Sigtia 2015; Korzeniowski, 2017), sledenje ritmu (npr. Holzapfel, 2012; Durand 2015), ocenjevanje razpoloženja (npr. Laurier, 2009; Pesek, 2017), priporočilni sistemi v glasbi (npr. Tkalčič, 2017), klasifikacija žanrov (npr. Lee, 2018) in analiza vzorcev v glasbi (npr. Conklin, 2010; Meredith, 2002; Ren, 2017). V mnogih primerih je povečano natančnost mogoče pripisati uvedbi globokega učenja na področju (Humphrey, 2012). Za številne probleme je bilo predlaganih več globokih pristopov, vključno s transkripcijo melodije (npr. Rigaud, 2016), klasifikacijo žanrov (npr. Jeong, 2016) in ocenjevanjem akordov (npr. Deng, 2016). V svoji široki definiciji globoki učni algoritem konstruira več stopenj predstavitve podatkov (hierarhijo značilk) z namenom modeliranja visokonivojskih struktur, prisotnih v opazovanih podatkih (Bengio, 2013).

Večina globokih učnih pristopov temelji na nevronskih mrežah. Med samim učenjem nevronskih mrež so visokonivojske predstavitve podatkov kodirane v večplastni hierarhiji, vendar je kodirano znanje implicitno in ga je težko razložiti na pregleden način kot model bele škatle. V zadnjih letih je bilo sicer razvitih več pristopov za vizualizacijo naučenih konceptov v nevronskih mrežah (npr. Bilal, 2018), vendar so še vedno daleč od popolnoma razvidnega kodiranega znanja. En od takšnih pristopov je na primer zakrivanje na slikah (Zeiler, 2014), ki poskuša prepoznati regije znotraj slike, ki sprožijo opazovani odziv v nevronski mreži.

Globoke nevronske mreže imajo običajno veliko parametrov, ki so potrebni za pokritje celotne ciljne domene, kar zahteva velike zbirke podatkov za učenje. Takšne velike zbirke podatkov je težko pridobiti zaradi pomanjkanja ustreznih podatkov, morebitnih težav z avtorskimi pravicami (slike, glasba) ali zaradi same majhnosti zbirk. Nadalje je treba takšne zbirke opremiti z anotacijami. Te pa so a) najpogosteje subjektivne (npr. žanrska razvrstitev v glasbi, sledenje objektov v računalniškem vidu) in zato zahtevajo več anotatorjev, da zajamemo približno človeško percepcijo problema, b) pogosto zahtevajo strokovnjaka (npr. glasbena transkripcija) in c) zahtevajo veliko časa in delovne sile.

V tem članku bomo predstavili obstoječe globoke pristope na področju pridobivanja informacij iz glasbe in najbolj popularne probleme, ki jih rešujejo ti pristopi. Predstavili bomo tudi probleme, pri katerih so bili globoki pristopi, temelječi na nevronskih mrežah, še uspešno aplicirani. Predvsem pri problemih nenadzorovanega odkrivanja vzorcev v podatkih na področju pridobivanja informacij iz glasbe še vedno uporabljamo hierarhične in druge pristope. Med drugim bomo omenili kompozicionalni hierarhični model kot alternativno globoko arhitekturo, ki presega nekatere omejitve drugih globokih pristopov. Pregled bomo sklenili z opisom nadaljnjih korakov pri razvoju pristopov v današnjem času, v katerem dominirajo globoki pristopi z nevronskimi mrežami.

2 PREGLED PODROČJA

V tem poglavju je predstavljen pregled hierarhičnega modeliranja, ki se nadaljuje s pregledom hierarhičnih in drugih globokih pristopov v področju MIR. Predstavljamo tudi več popularnih problemov na področju. Za nekatere probleme, npr. ocenjevanje osnovnih frekvenc, je uspešnost globokih pristopov signifikantno presegla uspešnost drugih pristopov. Pri drugih, kot je npr. odkrivanje vzorcev, pa so drugi, predvsem hierarhični pristopi, še vedno dominantni.

3 HIERARHIČNI MODELI NA PODROČJU MIR

Hierarhični modeli v glasbi dobro sovpadajo s samo domeno – glasba je hierarhična v času (zaporedja) in prostoru (harmonije). Številni pristopi za hierarhično modeliranje glasbe izhajajo s področja teorije glasbe, ki ponuja dobro uveljavljene hierarhične sisteme za glasbeno analizo.

Generativna teorija tonske glasbe (angl. Generative Theory of Tonal Music – GTTM) Lerdahla in Jackendoffa (1983) ponuja pristop hierarhičnega modeliranja glasbe v muzikologiji, ki je v sodobni glasbeni teoriji zelo znan. GTTM poskuša formalizirati sistem, ki odraža poslušalčevo razumevanje glasbe. GTTM predlaga štiri hierarhične vidike: skupinske in metrične strukture, redukcijo časovnega razpona in strukture za podaljševanje. Drug znani hierarhični pristop je predlagal Heinrich Schenker (1980). Schenkerjeva analiza, poimenovana po avtorju, poskuša v

glasbi razkriti temeljno osnovno strukturo (nem. Ursatz).

Čeprav sta GTTM in Schenkerjeva analiza večinoma odvisni od ekspertnih pravil, je koncept hierarhičnega strukturiranja v glasbi intuitiven način, saj temelji na vzorcih človeškega zaznavanja in razmišljanja. Ker pravila v takšnih pristopih niso zelo strogo opredeljena, je avtomatizacija procesa analize netrivialna. V preteklosti je bilo predstavljenih več sistemov za GTTM ali Schenkerjevo analizo (npr. Hamanaka, 2006; Hirata, 2007; Marsden, 2010). Marsden (2010) je predlagal sistem za samodejno pridobivanje Schenkerjeve redukcije. Pokazal je, da je mogoče takšno analizo opraviti samodejno, vendar je z implementacijo sistema odkril tudi nekatere pomanjkljivosti avtomatizacije. Predlagani postopek zahteva precejšnjo računsko moč. Poleg tega sistem ustvari veliko možnih analiz, ki se razlikujejo po kakovosti. Avtor je ugotovil, da je sicer implementacija sistema uspešna, a so potrebni dodatni koraki, ki bi sistem izboljšali do primernosti za vsakdanjo uporabo.

Človeško zaznavanje pogosto modeliramo z enim ali več hierarhičnimi sistemi, saj je takšno modeliranje intuitivno podobno našemu razumevanju zaznavanja. Farbood idr. (2010) so raziskali medsebojno povezavo med omejitvami delovnega spomina in hierarhičnimi strukturami v glasbi. Poročali so, da razlike v optimalnem časovnem usklajevanju tonskih harmonij v primerjavi z ritmom in konturo pomenijo različno obdelavo za vsako od teh modalnosti v glasbi. Poskusi, kot sta na primer Sapp (2005) in Woolhouse (2006), so tudi empirično pokazali prisotnost takšnih hierarhičnih predstavitev, ki jih povzročajo človeški kognitivni procesi. Conklin in Anagnostopoulou (2001) sta predlagala algoritem odkrivanja vzorcev z več vidikov (angl. viewpoint), ki temelji na priponskem drevesu. Za izbrani vidik (preoblikovanje glasbenega dogodka v abstraktno funkcijo) algoritem gradi priponsko drevo. Po izbiri vzorcev, ki ustrezajo določenim pogostejšim vrednostim in pragom pomembnosti, liste drevesa obravnavata kot najdaljše pomembne vzorce v korpusu. Conklin in Bergeron (2008) sta kasneje predstavila dva algoritma, ki temeljita na Conklinovemu modelu. Prvi algoritem najde vse »maksimalne pogoste vzorce«, drugi pa je optimizacijski algoritem z uporabo hitrejšega hevrističnega pristopa, pri katerem najdeni vzorci morda niso vedno največji in najbolj pogosti vzorci. Največji pogost vzorec je vzorec, katerega sestava funkcij komponent ni mogoče nadalje specializirati, ne da bi vzorec postal redek.

Wiggins in Forth (2015) sta opisala kognitivno arhitekturo »informacijska dinamika razmišljanja« (angl. information dynamics of thinking – IDyOT). Arhitektura je korak k modelu, ki vključuje vidike človeške ustvarjalnosti in druge oblike kognitivne obdelave v smislu predzavedne napovedne zanke. Predlagani model je hierarhična arhitektura, ki na prvem sloju vključuje številne generatorje, uporabljene za vzorčenje vnosa. Vsak generator proizvaja distribucijo izhodov glede na vhodno zaporedje. Arhitektura poskuša modelirati kognitivni cikel, ki temelji na statističnih opazovanjih vhodnih zaporedij. Vhodna zaporedja predstavljajo različne vidike glasbe, kot so tonska višina, barva, amplituda in čas. Napovedi, ki se ujemajo s perceptualnim vhodom, so razvrščene v zaporedje. Ta dinamični vidik ima za posledico proces inkrementalnega učenja.

4 GLOBOKI NEVRONSKI PRISTOPI V MIR

Pridobivanje informacij iz glasbe vključuje širok nabor problemov, ki obsegajo ustvarjalne, analitične in priklicne vidike dela z glasbo v različnih zapisih in oblikah.

Mnogi od teh problemov so formalizirani v okviru pobude MIREX (angl. eXchange Evaluation Evaluation), ki si prizadeva za vzpostavitev okvirov za vrednotenje in primerjavo različnih pristopov v MIR (Downie, 2008). Pobuda MIREX je zdaj dobro uveljavljena v skupnosti MIR, rezultate evalvacij pa so predstavili na konferenci ISMIR. Izmed množice obstoječih problemov smo izbrali nekatere, pri katerih so pogosteje uporabljeni globoki ali hierarhični pristopi. Prav te bomo v nadaljevanju obravnavali podrobneje.

5 AVTOMATSKO OCENJEVANJE AKORDOV

Zaporedja akordov in melodija sta dva najbolj prepoznavna gradnika zahodne glasbe – po navadi si z njima zapomnimo posamezno pesem. Samodejno ocenjevanje akordov lahko zato uporabimo za transkripcijo (Mauch, 2008; Mauch, 2007, Papadopoulos, 2011; Smith, 2011) in klasifikacijo glasbe (Ni, 2012)

ter druge naloge. Ocenjevanje akordov lahko uporabljamo tudi za zbiranje informacij ali izluščevanje metapodatkov (Sheh, 2003; Papadopoulos, 2007) in analizo vzorcev (Scholz, 2009).

Algoritmi za ocenjevanje akordov so najpogosteje sestavljeni iz dveh modelov: akustičnega, ki preoblikuje avdio signal v značilke, in jezikovnega, ki modelira časovna razmerja med akordi.

Pri tradicionalnih pristopih so najpogosteje uporabljene kromatske značilke (Mauch, 2008; Muller, 2011) ali razredi tonskih višin (Gomez, 2004). Te značilke zagotavljajo vmesno predstavitev zvočnega signala in običajno vsebujejo dvanajst dimenzij, od katerih vsaka predstavlja moč oktavno invariantne tonske višine v signalu. Vsako komponento kromatskega vektorja izračunamo kot vsoto ustreznih frekvenc. Ker kromatski vektorji ohranijo informacije o tonski višini, jih je mogoče uporabiti za ocenjevanje akordov s standardnimi algoritmi za strojno učenje, kot je na primer metoda podpornih vektorjev. Vendar takšna klasifikacija ne upošteva časovne odvisnosti akordov, saj so vektorji obravnavani neodvisno. Za časovno odvisnost pogosto uporabljajo prikrite modele Markova (angl. hidden Markov model – HMM) (npr. Bello, 2005; Noland, 2006; Papadopoulos, 2007).

V zadnjem času za ocenjevanje akordov pogosto uporabljajo globoko učenje. Boulanger-Lewandowski in drugi (2013) so predlagali model RNN. Z učenjem modela na celotnem naboru podatkov poročajo o 93,5-odstotni povprečni natančnosti na posamezni razred akorda. Vendar pa tudi podrobneje razložijo rezultate, pri čemer navajajo, da je »ta scenarij močno nagnjen k prekomerni ureditvi: z vidika strojnega učenja je trivialno oblikovati neparametrični model, ki deluje pri 100-odstotni natančnosti« (str. 5).

Sigtia in drugi (2015) so predlagali hibridno rekurenčno nevronsko mrežo za prepoznavanje akordov. Najprej so uporabili prtslojno globoko nevronsko mrežo, pri kateri so za vhod uporabili avdio signal, transformiran z metodo Constant-Q. DNN je bil uporabljen kot akustični model, ki odpravlja potrebo po kromatskih vektorjih ali podobnih značilkah. Za jezikovni model so uvedli hibridno rekurenčno nevronsko mrežo (RNN), ki modelira razmerja med izhodi. Ta model učinkovito nadomešča HMM, ki ga uporabljajo pri tradicionalnih pristopih. Hibridni model je bil preizkušen na podatkovnem nizu MIREX 2014 s štirikratno navzkrižno validacijo, pri čemer je bila učna množica nadalje razdeljena na 80 odstotkov za usposabljanje in 20 odstotkov za validacijo. Rezultati kažejo večjo učinkovitost kot pri akustičnih modelih (približno 3 % pri natančnosti na ravni okvira).

Deng in Kwok (2016) sta predlagala pristop hibridnega Gaussovega-HMM-globokega učenja. Model Gauss-HMM se uporablja za segmentacijo kromagramov in jih posreduje globokemu modelu za klasifikacijo akordov. Avtorja predlagata dva globoka učna modela, model globokega zaupanja in LSTM RNN (angl. long-short-term-memory – LSTM). Avtorja pri evalvaciji pokažeta, da njun model doseže boljše rezultate kot obstoječi sistem Chordino na podatkovnih nizih, ki so anotirani z večjim številom akordov, vendar doseže Chordino boljše rezultate na podatkovnih zbirkah z manjšim številom akordov.

Korzeniowski in Widmer (2016) sta predlagala sistem globokega učenja s kromatskimi vektorji. Model sta evalvirala na petih razpoložljivih podatkovnih zbirkah – albumih skupin Beatles, Queen in Zweieck, naboru podatkov RWC pop in diskografiji pevca Robbieja Williamsa – in dosegla do sedaj najboljše rezultate.

6 OCENJEVANJE OSNOVNIH TONSKIH VIŠIN

Cilj glasbene transkripcije je pretvoriti avdio zapis v notni zapis. Osnovni del transkripcije je ocena osnovnih frekvenc, ki so prisotne v signalu (angl. multiple fundamental frequency estimation – MFFE), pri čemer je cilj oceniti vse osnovne frekvence (ki ustrezajo tonskim višinam) v posameznih časovnih okvirih glasbenega signala. Kot pomemben cilj MIR je bil problem transkripcije raziskovan že od zgodnjih sedemdesetih let (Benetos, 2015; Gerhard, 2003; Klapuri, 2004 in 2006). Nekateri pristopi se problema lotevajo skozi analizo spektra signala (npr. Roebel, 2010; Pertusa, 2012), medtem ko drugi pristopi signal modelirajo kot kompozicijo različnih virov (npr. Dessein, 2010; Grindlay, 2011; Smaragdis, 2003). Veliko pristopov je specializiranih za posamezne inštrumente (npr. Marolt, 2004; Weninger, 2013; Boulanger-Lewandowski, 2012; Vincent, 2010) ali se fokusirajo na transkripcijo simboličnih podatkov, značilnih za posamezne inštrumente (npr. Barbancho, 2012).

Za ocenjevanje osnovnih frekvenc v signalu je bilo predstavljenih tudi več globokih pristopov, ki temeljijo na nevronskih mrežah (npr. Bock, 2012; Nam, 2011; Rigaud, 2016). Bock in Schedl (2012) sta uporabila rekurenčni model nevronske mreže za

transkripcijo klavirja, Nam idr. (2011) pa so združili mreže globokega zaupanja z metodo podpornih vektorjev in prikritim modelom Markova za isto nalogo. Rigaud in Radenen (2016) sta predlagala kombinacijo dveh globokih nevronskih mrež za transkripcijo pevskega glasu.

Zaradi pomanjkanja anotiranih podatkovnih baz mnogi globoki mrežni pristopi za MFFE (npr. Bock, 2012; Nam, 2011; Kelz, 2016; Rigaud, 2016) uporabljajo velik del anotiranih podatkov za učenje modelov. Podatkovna zbirka MAPS (Emiya, 2010) je ena izmed najpogosteje uporabljenih zbirk za učenje in vrednotenje algoritmov MFFE. Sestavljena je iz 30 skladb, predvajanih z Yamaha Disklavierjem in sintetiziranih s 7 vzorci klavirja (približno milijon tonskih višin). Bock in Schedl (2012) sta svoj model rekurenčne nevronske mreže učila na štirih različnih zbirkah, vključno z zbirko MAPS. Poročala sta o visoki F_1 natančnosti (do 93,5 %) pri zaznavanju začetka tonske višine na zbirki MAPS; vendar pa sta pri učenju uporabila tudi znatno količino zbirke za učenje in validacijo (približno 75 odstotkov za učenje in 9,4 odstotka za validacijo). Nam in drugi (2011) so poročali o rezultatih za 30-sekundne odlomke iz zbirke MAPS (74,4-odstotna natančnost F_1), ob uporabi približno 60 odstotkov zbirke za učenje in 25 odstotkov za validacijo. Ker ti pristopi uporabljajo znaten del nizov podatkov za učenje in testiranje, so lahko rezultati preveč optimistični v primerjavi z njihovim delovanjem v realnem svetu.

Bittner in drugi (2017) so predlagali model za ocenjevanje osnovnih frekvenc, ki temelji na polno povezani konvolucijski mreži. Dosegli so najboljše rezultate na dveh od treh podatkovnih baz vrednotenja MFFE in presegli najsodobnejše pristope pri izluščevanju melodije.

Med najnovejšimi so Hawthorne in drugi (2017) predstavili kombinacijo konvolucijskih nevronskih omrežij in LSTM mrež. Na ravni okvirja dosežejo 78,30-odstotni F_1 rezultat, medtem ko na ravni notnega zapisa presegajo rezultate drugih pristopov za približno 30 odstotkov in dosegajo 82,29 odstotka. Avtorji rezultate opisujejo kot dokaz koncepta za svoje delo in poudarjajo vprašanja učenja in validacije na tako majhnem naboru podatkov.

Zaradi prej omenjenih omejitev evalvacije je težko uporabiti dosežene rezultate v scenarijih »realnega sveta«, ki lahko vključujejo posnetke, ki morda niso bili posneti v idealnih studijskih okoljih ali s profesionalnimi izvajalci. Pri takšnih pogojih pristope redko ocenjujejo predvsem zaradi pomanjkanja različnih označenih podatkovnih nizov – večina podatkovnih baz je sestavljena predvsem iz sintetiziranih posnetkov, ki jih je mogoče preprosto dobiti in vsebujejo le majhno število anotiranih dejanskih posnetkov. Posledično lahko trpi robustnost algoritmov, saj se lahko ti dobro prilegajo majhnim podatkovnim zbirkam, kar vodi k slabemu delovanju na različnih materialih in ob prisotnosti hrupa.

7 ODKRIVANJE VZORCEV IN TRENUTNI PRISTOPI

Odkrivanje ponavljajočih vzorcev je znan problem na različnih področjih, vključno z računalniškim vidom (npr. Campilho, 2012), bioinformatiko (npr. Coward, 1998) in pridobivanjem informacij iz glasbe (MIR). Čeprav je problem razširjen na več področij, se njegova definicija kot tudi algoritmi za odkrivanje vzorcev med področji močno razlikujejo. V glasbi so o pomembnosti ponavljanja razpravljali številni glasbeni teoretiki in nedavno tudi raziskovalci, ki so razvili algoritme za polavtomatsko analizo glasbe, kot na primer Marsden (2010). V ogrodju MIREX so raziskovalci izpostavili več nalog, ki se ukvarjajo z vzorci in strukturami v glasbi, vključno s strukturno segmentacijo, melodično podobnostjo v simbolnih podatkih in ujemanjem vzorcev ter odkrivanjem vzorcev.

Namen opravila »odkrivanje ponavljajočih tem in odsekov« je najti ponovitve, ki so eden najpomembnejših vidikov glasbenega dela (Meredith, 2002). Definicija MIREX določa, da »algoritmi vzamejo glasbo kot vhod in izpisujejo seznam vzorcev, ki se ponovijo znotraj tega dela« (Collins, 2015). Opravilo se lahko zdi podobno tudi sicer znani nalogi ujemanja vzorcev (Collins, 2010). Cilj algoritma za ujemanje vzorcev je najti mesto iskanega vzorca znotraj nabora podatkov, medtem ko algoritem za odkrivanje vzorcev najde lokacije podobnih zaporedij podatkov, ki se večkrat ponovijo v neki podatkovni zbirki, brez kakršnih koli informacij o iskanem vzorcu. Kot je zapisal Wang (2015), se opravilo odkrivanja vzorcev razlikuje od opravila strukturne segmentacije, pri katerem segmenti pokrivajo celotno glasbeno delo in predstavljajo disjunktne segmente v glasbi. V opravilu odkrivanja vzorcev se lahko vzorci delno prekrivajo ali so podmnožice drugega vzorca.

Za odkrivanje vzorcev v glasbi so predlagani različni pristopi. Večina pristopov ne temelji na globokih,

temveč na kompozicionalnih modelih. Hsu in drugi (2001) so poskušali odkrivati netrivialne vzorce s korelacijskimi matrikami in odkrivanjem daljših vzorcev z večkratnim povezovanjem krajših ponavljajočih vzorcev. Knopke in drugi (2009) so analizirali 101 delo Giovannija Pierluigija da Palestrine in poskušali odkrivati vzorce s priponskimi matričnimi strukturami.

Cambouropoulos in drugi (2005) so uporabili pristop priponskega drevesa z dovoljenim delnim prekrivanjem vzorcev. Da bi presegli omejitve priponskega drevesa, so se osredotočili tudi na problem približnega ujemanja odkritih vzorcev.

Za to opravilo je bilo predstavljenih še več drugih nehierarhičnih in neglobokih pristopov (npr. Meredith, 2013; Velarde, 2014; Lartillot, 2014), katerih obširnejše povzetke del v tem pregledu izpuščamo, a jih je vredno omeniti zaradi njihovih rezultatov. V našem pregledu za to opravilo nismo zasledili globokih pristopov, ki temeljijo na nevronskih mrežah. Globoki pristopi sicer dosegajo dobre rezultate pri opravilih razvrščanja, a so takšni modeli črne škatle, pri katerih si težko razlagamo naučene strukture, zaradi česar jih posledično težko uporabljamo za opravila odkrivanja zakonitosti.

8 ALTERNATIVNI GLOBOKI MODELI

Čeprav globoki modeli, ki temeljijo na nevronskih mrežah, dosegajo izvrstne rezultate, so kompozicionalni in hierarhični modeli še vedno prisotni pri problemih, ki vsebujejo koncept odkrivanja vzorcev v podatkih. Takšni sistemi omogočajo vpogled v naučene abstrakcije, kar je še vedno netrivialen postopek pri modelih, ki temeljijo na nevronskih mrežah.

Kot alternativni globoki model, ki ne temelji na nevronskih mrežah, so Pesek in drugi (2017) razvili kompozicionalni hierarhični model za pridobivanje informacij iz glasbe. Z nenadzorovanim učenjem model zgradi hierarhično predstavitev konceptov od preprostih konceptov na najnižjem nivoju proti najkompleksnejšim konceptom na najvišjih nivojih. Ideja o takšni strukturi modela izvira iz raziskav na področju strojnega vida. Na tem področju sta Leonardis in Fidler (2007, 2009) predstavila koncept lHoP (angl. learned Hierarchy of Parts). Njun model se lahko nauči hierarhične predstavitve objektov na slikah, začenši z enostavnimi gradniki na nizkih nivojih, ki jih združuje v kompleksnejše dele objektov na višjih nivojih. Model se uči na podlagi statistike pojavitev in ga je moč uporabiti kot robusten način za kategorizacijo objektov in druge sorodne probleme na področju računalniškega vida.

Ideja modela temelji na predpostavki, da lahko kompleksne sestavljene signale razdrobimo na enostavnejše gradnike – *dele*. Deli so lahko različno kompleksni in glede na kompleksnost tvorijo različne nivoje. Posamezne dele na višjih nivojih lahko tvorimo s kombiniranjem delov na nižjih nivojih in tako tvorimo kompozicionalni model. V glasbi je takšen pristop človeku intuitiven, saj so glasbeni dogodki tvorjeni na podoben način: akord je sestavljen iz vsaj treh tonov, posamezni ton pa iz več frekvenc. Posamezni del tako opisuje posamezne frekvence na nižjem nivoju, na višjih nivojih pa njegove tvorjene kombinacije – kompozicije – tvorijo kompleksnejše dogodke. Na enak način lahko modeliramo tudi vzorce v glasbi, sosledja tonskih višin in akordov. Celotna struktura modela je transparentna, saj lahko za vsak del pregledamo in interpretiramo njegovo vlogo.

Pesek in drugi so model aplicirali na različna opravila s področja MIR, med drugim na avtomatsko ocenjevanje akordov, ocenjevanje osnovnih frekvenc in iskanje vzorcev v simbolni glasbi. Za opravilo avtomatskega ocenjevanja akordov so zgradili tronivojski model. Model so naučili na 88 klavirskih tonih in ga aplicirali na glasbeno zbirko skupine The Beatles. Dele na tretjem nivoju so uporabili kot kromatske vektorje in s pomočjo prikritega modela Markova napovedovali sosledja akordov.

Pri ocenjevanju osnovnih frekvenc so ponovno uporabili transparentno strukturo modela (Pesek, 2017a). Model so naučili na 88 klavirskih tipkah in ga aplicirali na več različnih zbirkah podatkov. Poleg javno dostopne zbirke MAPS, ki se pogosto uporablja za evalvacijo pristopov pri ocenjevanju osnovnih frekvenc, so predstavili svojo zbirko slovenske ljudske glasbe. Zbirka vsebuje 38 ljudskih pesmi, ki jih večglasno poje več amaterskih pevcev. Zbirka je posneta v vsakdanjih prostorih z osnovno produkcijsko opremo. Na tej zbirki so evalvirali tudi druge pristope in pokazali, da se predlagani kompozicionalni hierarhični model zaradi svoje robustnosti odreže bolje od drugih pristopov. Prav tako so analizirali hitrost delovanja in ugotovili, da je predlagani model hitrejši od drugih pristopov in je zato primeren za aplikacije v vgrajenih sistemih, mobilnih napravah in drugih podobnih, računsko manj zmožnih napravah.

Predlagani model so avtorji kasneje dodatno nadgradili (Pesek idr., 2017b) in prilagodili za delo

s simbolnimi glasbenimi predstavitvami z namenom razširitve nabora opravil na področju pridobivanja informacij iz glasbe. To razširitev modela so poimenovali SymCHM. Ker model vsebuje transparentno hierarhično strukturo, so model aplicirali na problem odkrivanja vzorcev v simbolnih glasbenih predstavitvah. Zaradi transparentnosti strukture je model moč uporabiti za opravila odkrivanja, kar je izredno težko doseči pri drugih strukturah, ki temeljijo na nevronskih mrežah.

9 DISKUSIJA

Na področju pridobivanja informacij iz glasbe so se v zadnjih letih izrazito izboljšali rezultati na mnogih popularnih problemih, med prej omenjenimi na problemu avtomatskega ocenjevanja akordov in glasbene transkripcije. Kljub napredku pa trenutni globoki pristopi ne prinašajo popolne rešitve. Čeprav omogočajo nenadzorovano učenje in dosegajo zadovoljive rezultate pri klasifikacijskih nalogah, jim manjka transparentnost, kar bi omogočilo vpogled v naučene koncepte. Vizualizacija naučenih konceptov trenutnih pristopov je netrivialen problem. Velikokrat zato pristope uporabijo kot črne škatle (angl. black box), ki sicer rešujejo nalogo, a jih je težko nadalje izboljšati in nadgrajevati v kontekstu inteligentnega doprinosa k reševanju širšega problema percepcije glasbe.

Prav tako so zaradi velike količine vozlišč in povezav v strukturah modelov za učenje potrebne velike podatkovne zbirke, ki jih je težko pridobiti. Med najpogostejšimi težavami pridobivanja zbirk so potencialni problemi z avtorskimi pravicami in količina potrebnega časa ter ekspertnega znanja za anotacijo zbirk. S tem je tudi povezan problem evalvacije, saj mnogo pristopov uporablja večji del podatkovne zbirke za učenje in validacijo. Avtorji nekaterih omenjenih pristopov ta problem tudi sami izpostavljajo in poudarjajo, da je treba tako pridobljene rezultate jemati z rezervo v kontekstu scenarijev v realnem svetu.

Kot je razvidno iz pregleda področja, so globoki pristopi učinkovito uporabljeni za več različnih nalog, ki se nanašajo na razlikovanje med naučenimi koncepti. Takšni sistemi rešujejo vprašanje, ali opazovani vhod pripada eni ali drugi skupini. Nasprotno je težko uporabiti takšen model za odkrivanje zakonitosti, pri čemer naj bi model izdelal lastno opažanje visokonivojskih abstraktnih pojmov, ki so prisotni na vhodu nenadzorovano. V tem kontekstu so hierarhični in drugi pristopi še vedno prisotni in dosegajo najboljše rezultate. Kot alternativa se je pojavil tudi kompozicionalni hierarhični model, ki je bil uspešno uporabljen za več različnih nalog, tako klasifikacijskih nalog kot nalog odkrivanja.

Glede na trenutne trende na področju je prihodnost »globoka«. S skokovitim izboljšanjem računske moči, zmožnosti oblačnega računanja in novimi ogrodji, ki omogočajo preprostejšo vzpostavitev sistemov za porazdeljeno računanje tako na splošno namenskih kot specializiranih grafičnih procesorjih, je takšne sisteme mogoče naučiti do mere, ki zadovoljivo pokriva celotno opazovano domeno. Vseeno pa se del raziskav osredotoča na razvoj simbolnih sistemov umetne inteligence (AI), ki namesto specializacije za specifičen problem ponujajo generaliziran pristop reševanja več različnih problemov (tabela 1). Takšni pristopi črpajo ideje s področij nevroznanosti, psihologije in drugih ved, ki opazujejo človeško delovanje, saj je njihov cilj približati delovanje algoritmov človeški percepciji. V dolgi zgodovini razvoja simbolnih AI pristopov je nastalo zavedanje, da ti pristopi prinašajo veliko omejitev, predvsem na nivojih vhodnih podatkov in generiranja značilk, ki jih lahko v takšnih pristopih uporabljamo. Ugibamo lahko, da je potencialna rešitev, ki bo presegala oba tipa pristopov, hibridna, pri čemer bodo združene

Tabela 1: **Pregled globokih modelov, temelječih na nevronskih mrežah (NN), in drugih simbolnih in hierarhičnih modelov na področju MIR. Uspešne aplikacije nakazujejo, da so bili pristopi uporabljeni v tem tipu problemov. Če je tip pristopov dosegel najboljše rezultate za posamezen tip problema, je to označeno v stolpcu najboljši rezultati.**

		Problemi na področju MIR			
		Klasifikacijski problemi		Problemi odkrivanja vzorcev	
		Uspešne aplikacije	Najboljši rezultati	Uspešne aplikacije	Najboljši rezultati
Tipi pristopov	Globoki NN modeli	✓	✓		
	Hierarhični in simbolni pristopi	✓		✓	✓

prednosti posameznega pristopa z uporabo globokih pristopov na vhodnih nivojih in simbolnih pristopov na višjih nivojih procesiranja.

BIBLIOGRAFIJA

[1] Barbancho, A. M., Klapuri, A., Tardon, L. J., & Barbancho, I. (2012, mar). Automatic Transcription of Guitar Chords and Fingering From Audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 20 (3), 915–921. doi: 10.1109/TASL.2011.2174227

[2] Bello, J. P., & Pickens, J. (2005). A robust mid-level representation for harmonic content in music signals. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 304–311). London.

[3] Benetos, E., & Weyde, T. (2015). Multiple-F0 estimation and note tracking for Mirex 2015 using a sound state-based spectrogram factorization model. In *11th annual music information retrieval exchange (mirex'15)* (pp. 1–2). Malaga.

[4] Bengio, Y., Courville, A., & Vincent, P. (2013, aug). Representation learning: a review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35 (8), 1798–828. doi: 10.1109/TPAMI.2013.50

[5] Bilal, A., Jourabloo, A., Ye, M., Liu, X., & Ren, L. (2018). Do Convolutional Neural Networks Learn Class Hierarchy? *IEEE Transactions on Visualization and Computer Graphics*, 24 (1), 152–165. doi: 10.1109/TVCG.2017.2744683

[6] Bittner, R. M., Justin, S., Essid, S., & Bello, J. P. (2015). Melody Extraction By Contour Classification. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 500–506). Malaga.

[7] Bittner, R. M., McFee, B., Salamon, J., Li, P., & Bello, J. P. (2017). Deep Salience Representations for F0 Estimation in Polyphonic Music. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 63–70). Suzhou, China.

[8] Bock, S., Krebs, F., & Schedl, M. (2012). *Evaluating the online capabilities of onset detection methods. In Proceedings of the international conference on music information retrieval (ismir).* Porto.

[9] Boulanger-Lewandowski, N., Bengio, Y., & Vincent, P. (2012). Discriminative Non-negative Matrix Factorization For Multiple Pitch Estimation. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 205–210). Porto, Portugal.

[10] Cambouropoulos, E., Crochemore, M., Iliopoulos, C. S., Mohamed, M., & Sagot, M.-F. (2005). A Pattern Extraction Algorithm for Abstract Melodic Representations that Allow Partial Overlapping of Intervallic Categories. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 167–174). London.

[11] Collins, T. (2016). Discovery of Repeated Themes & Sections – MIREX Wiki. Retrieved 2018-04-01, from http://www.music-ir.org/mirex/wiki/2016:Discovery_of_repeated_themes_sections

[12] Collins, T., Thurlow, J., Laney, R., Willis, A., & Garthwaite, P. H. (2010). A Comparative Evaluation of Algorithms for Discovering Translational Patterns in Baroque Keyboard Works. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 3–8). Utrecht.

[13] Conklin, D. (2008). Discovery of distinctive patterns in music. In *Proceedings of mml08: International workshop on machine learning and music* (p. 2). Helsinki. doi: 10.1.1.158.4152

[14] Conklin, D. (2010). Discovery of distinctive patterns in music. *Intelligent Data Analysis*, 14 (5), 547–554.

[15] Conklin, D., & Anagnostopoulou, C. (2001). Representation and Discovery of Multiple Viewpoint Patterns. In *Proceedings of the 2001 international computer music conference* (pp. 479–485). Cuba.

[16] Coward, E., & Drabløs, F. (1998, jan). Detecting periodic patterns in biological sequences. *Bioinformatics* (Oxford, England), 14 (6), 498–507.

[17] de Cheveigne, A. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of Acoustical Society of America*, 111 (4), 1917– 1930.

[18] Deng, J., & Kwok, Y.-K. (2016). A Hybrid Gaussian-Hmm--Deep-Learning Ap- proach for Automatic Chord Estimation with Very Large Vocabulary. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 812–818). New York.

[19] Dessein, A., Cont, A., & Lemaitre, G. (2010). Real-time polyphonic music tran- scription with non-negative matrix factorization and beta-divergence. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 489–494).

[20] Downie, J. S. (2008). The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology*, 29 (4), 247–255.

[21] Downie, J. S., Ehmann, A. F., Bay, M., & Jones, M. C. (2010). The Music Information Retrieval Evaluation eXchange: Some Observations and Insights. In W. A.A. & R. Z.W. (Eds.), *Advances in music information retrieval* (pp. 93–115). Berlin: Springer-Verlag.

[22] Durand, S., Bello, J. P., David, B., & Richard, G. (2015). Downbeat tracking with multiple features and deep neural networks. In *Acoustics, speech and signal processing (icassp)* (pp. 409–413).

[23] Emiya, V., Badeau, R., & David, B. (2010, aug). Multipitch Estimation of Piano Sounds Using a New Probabilistic Spectral Smoothness Principle. *IEEE Transactions on Audio, Speech, and Language Processing,* 18 (6), 1643–1654. doi: 10.1109/TASL.2009.2038819

[24] Farbood, M. (2010). Working memory and the perception of hierarchical tonal structures. In *Proceedings of international conference of music perception and cognition* (pp. 219–222). Seattle.

[25] Gelfand, S. A. (2004). Hearing: An introduction to psychological and physiological acoustics. *CRC Press*.

[26] Gerhard, D. (2003). Pitch Extraction and Fundamental Frequency: History and Current Techniques (Tech. Rep.). Regina: University of Regina, Saskatchewan, Canada.

[27] Gomez, E., & Herrera, P. (2004). Estimating the Tonality of Polyphonic Audio Files: Cognitive versus Machine Learning Modelling Strategies. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 92–95). Barcelona.

[28] Grindlay, G., & Ellis, D. P. W. (2011, oct). Transcribing Multi--Instrument Polyphonic Music With Hierarchical Eigeninstruments. *IEEE Journal of Selected Topics in Signal Processing,* 5 (6), 1159–1169. doi: 10.1109/JSTSP.2011.2162395

[29] Hamanaka, M., Hirata, K., & Tojo, S. (2006, dec). Implementing "A Generative Theory of Tonal Music" *Journal of New Music Research*, 35 (4), 249–277. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/09298210701563238, doi: 10.1080/09298210701563238

[30] Harte, C., Sandler, M., Abdallah, S., & Gomez, E. (2005). Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the international conference on music information retrieval (ismir)*. London.

[31] Hawthorne, C., Elsen, E., Song, J., Roberts, A., Simon, I.,

Raffel, C., . . . Eck, D. (2017, oct). Onsets and Frames: Dual-Objective Piano Transcription. Retrieved from http://arxiv.org/abs/1710.11153

[32] Hirata, K., Tojo, S., & Hamanaka, M. (2007). Techniques for Implementing the Generative Theory of Tonal Music. In *Proceedings of the international conference on music information retrieval (ismir)*. Vienna.

[33] Holzapfel, A., Davies, M. E. P., Zapata, J. R., Oliveira, J. L., & Gouyon, F. (2012, nov). Selective Sampling for Beat Tracking Evaluation. *IEEE Transactions on Audio, Speech, and Language Processing,* 20 (9), 2539– 2548. doi: 10.1109/TASL.2012.2205244

[34] Hsu, J.-L., Liu, C.-C., & Chen, A. L. (2001). Discovering nontrivial repeating patterns in music data. IEEE Transactions on Multimedia, 3 (3), 311–325.

[35] Humphrey, E. J., Bello, J. P., & LeCun, Y. (2012). Moving beyond feature design: deep architectures and automatic feature learning in music informatics. In *Proceedings of the international conference on music information retrieval (ismir)*. Porto.

[36] Jeong, I.-Y., & Lee, K. (2016). Learning Temporal Features Using a Deep Neural Network and its Application to Music Genre Classification. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 434–440). New York.

[37] Kelz, R., Dorfer, M., Korzeniowski, F., Bock, S., Arzt, A., & Widmer, G. (2016). On the Potential of Simple Framewise Approaches to Piano Transcription. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 475–481). New York.

[38] Klapuri, A., & Davy, M. (Eds.). (2006). Signal Processing Methods for Music Transcription. New York: Springer. Retrieved from http://www.springer.com/engineering/signals/book/978-0-387-30667-4

[39] Klapuri, A. P. (2004, sep). Automatic Music Transcription as We Know it Today. *Journal of New Music Research*, 33 (3), 269–282. doi: 10.1080/0929821042000317840

[40] Knopke, I., & Ju¨rgensen, F. (2009). A system for identifying common melodic phrases in the masses of palestrina. *Journal of New Music Research*, 38 (2), 171–181.

[41] Korzeniowski, F., & Widmer, G. (2016). Feature Learning for Chord Recognition: the Deep Chroma Extractor. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 37–43). New York.

[42] Korzeniowski, F., & Widmer, G. (2017). End-to-End Musical Key Estimation Using a Convolutional Neural Network. In *Proceedings of the european signal processing conference (eusipco)* (pp. 996–1000). Kos Island, Greece.

[43] Lartillot, O. (2014). Submission to MIREX Discovery of Repeated Themes and Sections. In *10th annual music information retrieval exchange (mirex'14)* (pp. 1–3). Taipei.

[44] Laurier, C., Meyers, O., Serr`a, J., Blech, M., Herrera, P., & Serra, X. (2009, oct). Indexing music by mood: design and integration of an automatic content-based annotator. *Multimedia Tools and Applications*, 48 (1), 161– 184. doi: 10.1007/s11042-009-0360-2

[45] Lee, J., Park, J., Kim, K., & Nam, J. (2018, jan). SampleCNN: End-to-End Deep Convolutional Neural Networks Using Very Small Filters for Music Classification. *Applied Sciences*, 8 (2), 150. doi: 10.3390/app8010150

[46] Lerdahl, F., & Jackendoff, R. (1983). A generative theory of tonal music. *Cambridge: MIT Press*.

[47] Marolt, M. (2004, jun). A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music. *IEEE Transactions on Multimedia*, 6 (3), 439–449. doi: 10.1109/TMM.2004.827507

[48] Marsden, A. (2010, sep). Schenkerian Analysis by Computer: A Proof of Concept. *Journal of New Music Research*, 39 (3), 269–289. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/09298215.2010.503898 doi: 10.1080/09298215.2010.503898

[49] Mauch, M., & Dixon, S. (2008). A Discrete Mixture Model for Chord Labelling. In *Proceedings of the international conference on music information retrieval (ismir)* (Vol. 1, pp. 45–50). Philadelphia.

[50] Mauch, M., & Dixon, S. (2010). Approximate Note Transcription For The Improved Identification Of Difficult Chords. In *Proceedings of the international conference on music information retrieval (ismir)*. Utrecht.

[51] Mauch, M., Dixon, S., & Harte, C. (2007). Discovering Chord Idioms Through Beatles and Real Book Songs. In *Proceedings of the international conference on music information retrieval (ismir)*. Vienna.

[52] McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch and the auditory system. *Current Opinion in Neurobiology*, 1 (18), 452–463.

[53] Meredith, D. (2013). COSIATEC AND SIATECCOMPRESS: PATTERN DISCOVERY BY GEOMETRIC COMPRESSION. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 1–6).

[54] Meredith, D., Lemstrom, K., & Wiggins, G. A. (2002, dec). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, 31 (4), 321–345. doi: 10.1076/jnmr.31.4.321.14162

[55] Mu¨ller, M., & Ewert, S. (2011). Chroma Toolbox: MATLAB Implementations for Extracting Variants of Chroma-Based Audio Features. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 288–295). Miami.

[56] Nam, J., Ngiam, J., Lee, H., & Slaney, M. (2011). A Classification-Based Polyphonic Piano Transcription Approach Using Learned Feature Representations. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 175–180). Miami.

[57] Ni, Y., McVicar, M., Santos-Rodriguez, R., & Bie, T. D. (2012). Using Hyper-genre Training to Explore Genre Information for Automatic Chord Estimation. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 109–114). Porto.

[58] Noland, K., & Sandler, M. (2006). Key Estimation Using a Hidden Markov Model. In *Proceedings of the international conference on music information retrieval (ismir)*. Victoria.

[59] Papadopoulos, H., & Peeters, G. (2007). Large-case Study of Chord Estimation Algorithms Based on Chroma Representation and HMM. *Content-Based Multimedia Indexing*, 53-60 .

[60] Papadopoulos, H., & Peeters, G. (2011). Joint Estimation of Chords and Downbeats From an Audio Signal. *IEEE Transactions on Audio, Speech, and Language Processing*, 19 (1), 138–152.

[61] Pertusa, A., & In˜esta, J. M. (2012). Efficient methods for joint estimation of multiple fundamental frequencies in music signals. *EURASIP Journal on Advances in Signal Processing*, 2012 (1), 27. doi: 10.1186/1687-6180-2012- 27

[62] Pesek, M., Leonardis, A., & Marolt, M. (2017a). Robust Real-Time Music Transcription with a Compositional Hierarchical Model. *PLoS ONE*, 12 (1). doi: 10.1371/journal.pone.0169411

[63] Pesek, M., Leonardis, A., & Marolt, M. (2017b, nov). SymCHM—An Unsupervised Approach for Pattern Discovery in Symbolic Music with a Compositional Hierarchical Model. *Applied Sciences*, 7 (11), 1135. Retrieved from http://www.mdpi.com/2076-3417/7/11/1135 doi: 10.3390/app7111135

[64] Pesek, M., Strle, G., Kavčič, A., & Marolt, M. (2017). The Moodo dataset: Integrating user context with emotional and color perception of music for affective music information retrieval. *Journal of New Music Research*, 46 (3), 1–15. doi: 10.1080/09298215.2017.1333518

[65] Ren, I. Y., Koops, H. V., Volk, A., & Swierstra, W. (2017). In Search of the Consensus Among Musical Pattern Discovery Algorithms. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 671–678). Suzhou, China.

[66] Rigaud, F., & Radenen, M. (2016). Singing Voice Melody Transcription using Deep Neural Networks. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 737–743). New York.

[67] Roebel, A., & Rodet, X. (2010, aug). Multiple Fundamental Frequency Estimation and Polyphony Inference of Polyphonic Music Signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 18 (6), 1116– 1126. doi: 10.1109/TASL.2009.2030006

[68] Ryyn¨anen, M. P., & Klapuri, A. P. (2008, sep). Automatic Transcription of Melody, Bass Line, and Chords in Polyphonic Music. *Computer Music Journal*, 32 (3), 72–86. doi: 10.1162/comj.2008.32.3.72

[69] Sapp, C. S. (2005). Visual hierarchical key analysis. *Computers and Intertainment*, 3 (4), 1–19.

[70] Schenker, H. (1980). Harmony. *University of Chicago Press*.

[71] Scholz, R., Vincent, E., & Bimbot, F. (2009, apr). Robust modeling of musical chord sequences using probabilistic N--grams. In *Proceedings of international conference on acoustics, speech, and signal processing (icassp)* (pp. 53–56). IEEE. doi: 10.1109/ICASSP.2009.4959518

[72] Sheh, A., & Ellis, D. (2003). Chord segmentation and recognition using EM-trained hidden Markov models. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 1–7). Baltimore.

[73] Sigtia, S., Boulanger-Lewandowski, N., & Dixon, S. (2015). Audio Chord Recognition With A Hybrid Recurrent Neural Network. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 127– 133). Malaga.

[74] Smaragdis, P., & Brown, J. (2003). Non-negative matrix factorization for polyphonic music transcription. In *2003 ieee workshop on applications of signal processing to audio and acoustics* (pp. 177–180). IEEE. doi: 10.1109/ASPAA.2003.1285860

[75] Smith, J. B. L., Burgoyne, J. A., Fujinaga, I., De Roure, D., & Downie, J. S. (2011). Design and Creation of a Large-scale Database of Structural Annotations. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 555–560). Miami.

[76] Tirovolas, A. K., & Levitin, D. J. (2011). music perception and cognition research from 1983 to 2010: a categorical and bibliometric analysis of empirical articles in Music Perception. *Music Perception: An Interdisciplinary Journal*, 29 (1), 23–36.

[77] Tkalčič, M., Maleki, N., Pesek, M., Elahi, M., Ricci, F., & Marolt, M. (2017). A Research Tool for User Preferences Elicitation with Facial Expressions. In *Proceedings of the eleventh acm conference on recommender systems* (pp. 353–354). Como, Italy: ACM. doi: 10.1145/3109859.3109978

[78] Tolonen, T., & Karjalainen, M. (2000). A computationally Efficient Multipitch Analysis Model. *IEEE Transactions on Speech and Audio Processing*, 8 (6), 708–716.

[79] Velarde, G., & Meredith, D. (2014). Submission to MIREX Discovery of Repeated Themes and Sections. In *10th annual music information retrieval exchange (mirex'14)* (pp. 1–3). Taipei.

[80] Vincent, E., Bertin, N., & Badeau, R. (2010, mar). Adaptive Harmonic Spectral Decomposition for Multiple Pitch Estimation. *IEEE Transactions on Audio, Speech, and Language Processing*, 18 (3), 528–537. doi: 10.1109/TASL.2009.2034186

[81] Wang, C.-i., Hsu, J., & Dubnov, S. (2015). Music Pattern Discovery with Variable Markov Oracle: A Unified Approach to Symbolic and Audio Representations. In *Proceedings of the international conference on music information retrieval (ismir)* (pp. 176–182). Malaga.

[82] Weninger, F., Kirst, C., Schuller, B., & Bungartz, H.-J. (2013). A discriminative approach to polyphonic piano note transcription using supervised non-negative matrix factorization. In *Proceedings of international conference on acoustics, speech, and signal processing (icassp)* (pp. 6–10). Vancouver.

[83] Wiggins, G. A., & Forth, J. (2015). IDyOT: A Computational Theory of Creativity as Everyday Reasoning from Learned Information. In *Computational creativity research: Towards creative machines* (pp. 127–148). Atlantis Press, Paris.

[84] Woolhouse, M., Cross, I., & Horton, T. (2006). The perception of non-adjecent harmonic relations. In *Proceedings of international conference on music perception and cognition*. Bologna.

[85] Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision – eccv* 2014 se - 53 (Vol. 8689, pp. 818–833). Springer International Publishing.

▪

Matevž Pesek je asistent in raziskovalec na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, kjer je leta 2012 diplomiral in leta 2018 doktoriral iz računalništva. Od leta 2009 je član Laboratorija za računalniško grafiko in multimedije. Njegovi raziskovalni interesi so biološko navdihnjeni modeli, globoke arhitekture, kompozicionalno hierarhično modeliranje in multimodalna percepcija glasbe, vključno z interakcijo človek – računalnik in vizualizacijo za analizo zvoka in glasbe.