VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2019

Bc. Pavel Smělý

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ
DEPARTMENT OF TELECOMMUNICATIONS

ROZPOZNÁVÁNÍ HUDEBNÍ NÁLADY A EMOCÍ ZA POMOCI TECHNIK MUSIC INFORMATION RETRIEVAL
MUSIC MOOD AND EMOTION RECOGNITION USING MUSIC INFORMATION RETRIEVAL TECHNIQUES

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — **Bc. Pavel Smělý**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. Tomáš Kiska**
SUPERVISOR

BRNO 2019

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Diplomová práce

magisterský navazující studijní obor **Audio inženýrství**
Ústav telekomunikací

Student: Bc. Pavel Smělý ***ID:*** 164617
Ročník: 2 ***Akademický rok:*** 2018/19

NÁZEV TÉMATU:

Rozpoznávání hudební nálady a emocí za pomoci technik Music Information Retrieval

POKYNY PRO VYPRACOVÁNÍ:

V rámci této práce budou shrnuty dosavadní poznatky z oblasti zvané Music Information Retrieval. Konkrétně pak poznatky věnující se rozpoznávání hudebních emocí a nálady. Bude sestavena databáze nahrávek, na které bude toto rozpoznávání testováno. Dále budou analyzovány hudební nahrávky z hlediska barvy zvuku, rytmiky a dynamiky a vybrány takové hudební parametry, které mají největší sílu rozpoznat jednotlivé druhy hudebních emocí a nálady. V neposlední řadě bude vyhodnocena klasifikační úspešnost u jednotlivých druhů těchto hudebních emocí a nálad. Důraz testování bude kladen na co možno nejpestřejší nástrojové složení a žánrové zastoupení.

DOPORUČENÁ LITERATURA:

[1] From sounds to music and emotions. 9th International Symposium, CMMR 2012, London, UK, June 19-22, 2012, Revised Selected Papers. New York: Springer, 2013. ISBN 978-3-642-41247-9.

[2] MÜLLER, M. Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications [online]. Springer International Publishing Switzerland, 2015, 483 s. ISBN 978-3-319-21945-5.

Termín zadání: 1.2.2019 ***Termín odevzdání:*** 16.5.2019

Vedoucí práce: Ing. Tomáš Kiska
Konzultant:

prof. Ing. Jiří Mišurec, CSc.
předseda oborové rady

ABSTRAKT

Tato práce se zabývá oblastí Music Information Retrieval, přesněji její podoblastí zaměřující se na rozpoznávání hudebních emocí s názvem Music Emotion Recognition. Počáteční kapitoly práce se věnují obecnému přehledu a definici MER, kategorizaci jednotlivých metod a nabízejí tak komplexní pohled na tuto vědní disciplínu. Práce se dále zabývá výběrem a popisem vhodných parametrů pro rozpoznávání emocí, k čemuž využívá nástroje openSMILE a MIRtoolbox. K získání databáze nahrávek a jejich subjektivních emočních popisů byla použita volně dostupná databáze DEAM. Praktická část práce se již plně zabývá návrhem statického dimenzionálního regresního vyhodnocovacího systému pro číselnou predikci hudebních emocí u hudebních nahrávek, přesněji jejich polohy v AV emočním prostoru. Práce publikuje a komentuje přehled dosažených výsledků jak pro individuální analýzy významnosti jednotlivých parametrů pro úspěšnost predikce, tak celkové analýzy úspěšnosti predikce navrženého modelu.

KLÍČOVÁ SLOVA

získávání informací z hudby, rozpoznávání hudebních emocí, ReliefF, metoda podpůrných vektorů, regresní, číselná predikce, anotace, Gauss

ABSTRACT

This work focuses on scientific area called Music Information Retrieval, more precisely it's subdivision focusing on the recognition of emotions in music called Music Emotion Recognition. The beginning of the work deals with general overview and definition of MER, categorization of individual methods and offers a comprehensive view of this discipline. The thesis also concentrates on the selection and description of suitable parameters for the recognition of emotions, using tools openSMILE and MIRtoolbox. A freely available DEAM database was used to obtain the set of music recordings and their subjective emotional annotations. The practical part deals with the design of a static dimensional regression evaluation system for numerical prediction of musical emotions in music recordings, more precisely their position in the AV emotional space. The thesis publishes and comments on the results obtained by individual analysis of the significance of individual parameters and for the overall analysis of the prediction of the proposed model.

KEYWORDS

music information retrieval, music emotion recognition, MIR, MER, support vector regression, SVR, SVM, numerical prediction, subjective annotations, arousal, valence, GPR

SMĚLÝ, Pavel. *Rozpoznávání hudební nálady a emocí za pomoci technik Music Information Retrieval*. Brno, 2019, 85 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: Ing. Tomáš Kiska

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Rozpoznávání hudební nálady a emocí za pomoci technik Music Information Retrieval“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno ....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu Ing. Tomáši Kiskovi za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno ....

podpis autora

Faculty of Electrical Engineering
and Communication
Brno University of Technology
Purkynova 118, CZ-61200 Brno
Czech Republic
http://www.six.feec.vutbr.cz

PODĚKOVÁNÍ

Výzkum popsaný v této diplomové práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

Brno ....

podpis autora

EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI

OP Výzkum a vývoj
pro inovace

Obsah

Seznam obrázků

Seznam tabulek

Úvod

Tato práce se zabývá zkoumáním vědecké oblasti nazývané *Music information retrieval*, přesněji její podoblastí soustřeďující se na rozpoznávání hudebních nálad a emocí *Music emotion recognition*. Cílem diplomové práce je prozkoumat a shrnout znalosti této poměrně mladé a dynamické vědní disciplíny, navrhnout vhodný vyhodnocovací systém, který dokáže hudební emoce predikovat, a sestavit databázi, na které bude rozpoznávání testováno.

Hned v úvodu práce se nachází část motivace a zasazení problematiky do obecného kontextu a praktického využití. Tato část pomáhá také objasnit rozdíl mezi tzv. *hudební emocí* a *náladou* a naznačuje směrování této práce.

První kapitola se již plně zabývá definicí a popisem MER. Důraz je kladen především na podrobné zmapování celé této progresivně se vyvíjející oblasti, kategorizaci a popisu používaných metod a modelů. Dále rozebírá tzv. *modely emocí*. Tato kapitola tedy obsahuje jakýsi komplexní přehled, který umožňuje potřebnou orientaci v problematice, a tak následně pomáhá při návrhu vhodného vyhodnocovacího systému.

Ve druhé kapitole se nachází jak obecný přehled hudebních parametrů používaných pro rozpoznávání hudebních emocí, tak podrobný popis jednotlivých parametrů, které jsou v této práci prakticky využity. Kapitola je členěna na jednotlivé části podle typu nástroje, který byl použit k extrakci daných parametrů. V neposlední řadě se v této kapitole nachází také část o algoritmech pro výběr a předzpracování hudebních parametrů.

Potřebné hudební nahrávky a anotace emocí obsahuje databáze *DEAM*, která je popsána ve třetí kapitole. Od povahy a typu informací databáze vstupních dat se posléze z velké části odvíjí typ a metoda navrhovaného vyhodnocovacího systému.

Čtvrtá kapitola nabízí informace o použitých metodách strojového učení, především použité *metodě podpůrných vektorů*, její jádrových funkcích a použité implementaci. Na ni navazuje krátká pátá kapitola, který popisuje použité metody statistické analýzy.

Šestá kapitola se již plně zabývá praktickou částí práce, tedy návrhem systému na rozpoznávání emocí. Na prvních stranách kapitoly se nachází popis a návrh statického dimenzionálního regresního vyhodnocovacího systému pro číselnou predikci hudebních emocí u hudebních nahrávek, přesněji jejich polohy v tzv. *AV emočním prostoru*. Jako první se tato část zabývá výběrem ground-truth dat, poté navazuje hudebními parametry a jejich předzpracováním. Nakonec práce zdůvodňuje samotný výběr metod strojového učení.

V sedmé kapitole se nachází jak individuální hodnocení významnosti použitých parametrů, tak analýza použitých sad parametrů a celkové vyhodnocení navrženého

dimenzionálního vyhodnocovacího systému pro predikci hudebních emocí.

Osmá kapitola se zabývá shrnutím nabytých poznatků, diskuzí výsledků a srovnáním s podobnými realizovanými pracemi.

Závěr práce obsahuje shrnutí dosažených poznatků a zhodnocení splněných úkolů zadání.

Motivace

Hudba má bezesporu velkou a jen těžko odmyslitelnou roli v životě člověka. Již od pradávna lidstvo provází a vyvíjí se stejně tak, jako lidská kultura sama. Zdá se, že je hudba s emocemi velmi blízce propojena jako určitý zprostředkovatel prožitků. Právě jedním z kritérií kvality hudby může být míra schopnosti předat posluchači zamýšlenou emoci a náladu. Lze si totiž položit otázku, jakou hodnotu má hudba, která není schopna navodit v příjemci jakoukoliv emocionální odezvu.

Již od počátků prvních empirických prací až po moderní rozsáhlé výzkumné studie se ukazuje mnoho silných důkazů o tom, že v závislosti na kontextu, může hudba reálně navozovat a zprostředkovávat různé typy emocí. [13, 27] Jak zdůraznila Krumhanslová v [21], je důležitý rozdíl mezi *cítěním* emocí (ang. *feel*) a *vnímáním* emocí (*perceive*). Zatímco v prvním případě jde o posluchačovu emocionální odezvu v jeho vlastní mysli, druhý případ je spojen spíše s faktem, že hudba může obecně zprostředkovávat určité vlastnosti – kvality spojené s emocemi. V oblasti rozeznávání hudebních emocí a nálad bývá zjišťováno právě toto vnímání hudebních emocí. [4]

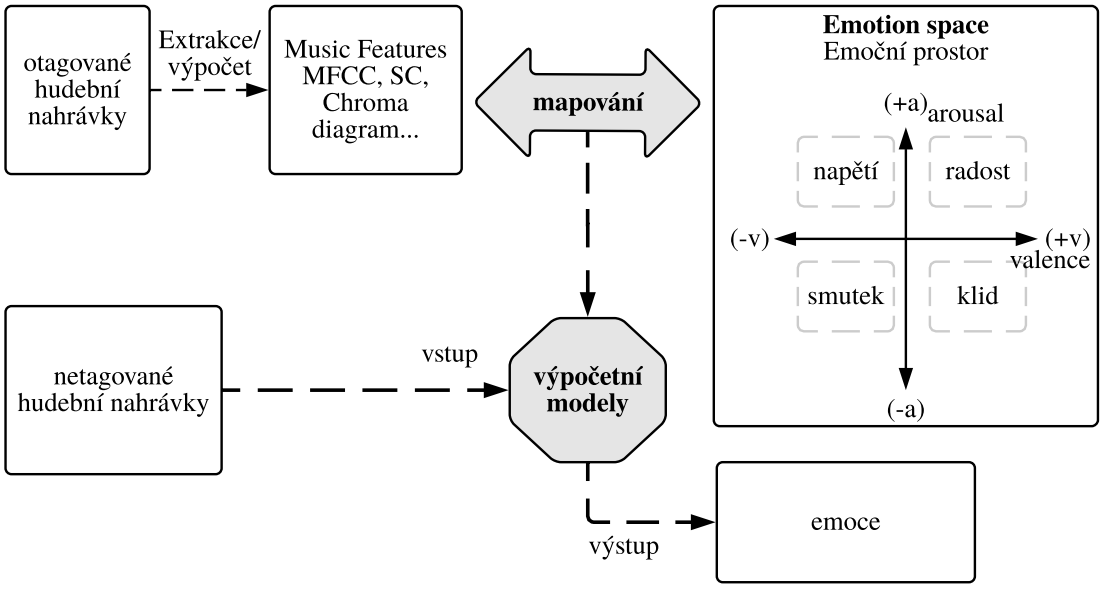
Ve výzkumu, při kterém bylo testováno 427 respondentů na míru zájmu vyhledávat či procházet hudební nahrávky podle třídění na základě emocí, se ukázalo, že jedno ze tří „vyhledání“ by využilo právě tuto metodu. Důležitost a perspektivu označování hudebních emocí podtrhuje výzkum, který ukázal, že 15 % všech hledání v internetové hudební službě *Last.fm* bylo právě podle tzv. *mood tags* (anotace, označení podle nálad). Tyto tendence a preference posluchačů zapříčinily rozmach a rozsáhlejší kategorizaci hudby na základě emocí. V současnosti vzniká veliká poptávka po systémech automatického doporučování a organizace hudby podle hudebních emocí a tedy i po nových studiích v oblasti, zabývající se výpočetními modely a detekováním emocí v hudbě, nazývanou souhrnně **Music Emotion Recognition**, zkráceně MER [4].

0.0.1 Hudební emoce vs. nálada

Porozumění rozdílu *mezi hudební emocí* a *hudební náladou* se může zdát pro pochopení problematiky zásadní. Pravdou však je, že neexistuje žádná zavedená definice, či jasné porozumění, co to vlastně *emoce* či *nálada* je. Neplatí to pouze v kontextu hudby, ale i v obecném měřítku. Existuje velké množství věděckých definic nálady a emocí, ale nikdy nebylo možné dojít ke konsenzu napříč věděckou obcí. Jedna z lepších definicí (Weld) popisuje emoce jako něco krátkého a vyprchávajícího, kdežto náladu jako déle trvalejší a stabilnější stav. Tyto dva pojmy bývají však velmi často zaměňovány [23].

1 Rozpoznávání hudebních emocí – Music Emotion Recognition

Oblast MER se snaží o sestavení co nejdokonalejšího procesu pro extrakci a analýzu hudebních parametrů (hudebních prvků, příznaků, *music features*) s využitím výpočetních technologií, dále se zabývá mapováním vztahů mezi těmito hudebními parametry a tzv. *emočním prostorem* (*emotion space*), a v neposlední řadě rozeznáváním emocí, které daná hudba vyjadřuje. Na základě těchto výpočtů mohou být poté sestaveny hudební databáze, které jsou tříděny podle hudebních emocí. Model typického procesu rozeznávání hudebních emocí zobrazuje schéma 1.1. MER modely mohou být použity také v oblastech doporučování hudby, hudební terapie či v jiných oborech MIR. Od roku 2007 se začaly objevovat v každoročním hodnocení *MIREX (Music Information Retrieval Evaluation eXchange)* vyhodnocující algoritmy MIR, extrémně úspěšné studie o klasifikaci hudby. V tom samém čase začalo mnoho internetových portálů, služeb a sociálních sítí jako např. *Stereomood* a *Sensbeat* používat emoce jako klíč k doporučování hudby či k jiným interakcím s posluchači. MER se začalo aplikovat také do různých zařízení, přehrávačů a aplikací v mobilních telefonech a dokonce do systémů tzv. „chytrých domácností“ (*Smart home systems*) [39].



Obr. 1.1: Model typického procesu MER [39].

1.1 Obecný model MER

Typická MER metoda se skládá ze tří kroků. Prvním krokem, který definuje i předurčuje povahu výsledků, je výběr různých typů a vlastností hudebních nahrávek a modelu reprezentací emocí – *emočních modelů*. Druhým krokem je extrakce hudebních parametrů z hudebních nahrávek, přičemž v MER se obecně používají dva typy: již zmíněné hudební parametry a základová data zvaná výstižněji *ground-truth data*. Zjednodušeně popsáno, tato data jsou povětšinou získána označováním jednotlivých emocí na základě daného emočního modelu prostřednictvím subjektů – respondentů. V třetím a posledním kroku bývá v trénovacím modelu využito metod *strojového učení* k určení vztahu a spojitosti extrahovaných hudebních parametrů a emocí [39].

Kromě predikce emočního označení či hodnot, které jednoduše vyjadřují emoce daných skladeb v reprezentativním úseku, můžeme do metod MER zavést také rozměr času. Této oblasti se říká *Music emotion Variation Detection*, zkráceně MEVD a zabývá se dynamickým procesem hudebních emocí a jejich predikci pro každý krátký časový úsek vedoucí k výsledné řadě hodnot emocionálních predikcí. Přestože MEVD i dimenzionální MER pohlíží na emoce z dimenzionální perspektivy, jsou rozdílné ve výpočetním modelu těchto emocí. Zatímco MEVD počítá hodnoty *VA systému* (*valence-arousal*) v každém krátkém časovém úseku a vyjadřuje skladbu řadou hodnot, dimenzionální MER počítá hodnoty VA pro reprezentativní úsek (často 30 s) skladby. Celý hudební úsek je tedy vyjádřen jedním bodem ve VA prostoru [41]. Podle výše uvedených vlastností můžeme rozdělit modely MER na:

- **statické modely**
- **dynamické modely** (které pracují s časovou informací), např. MEVD

1.2 Ground-truth data a modely emocí

Ve studiích zabývajících se lidskými emocemi psychologové často využívají lidských verbálních popisů emocionálních odezev [32]. Například známý Hevnerův článek z roku 1935 popisuje vztah mezi hudbou a emocemi pomocí experimentu, při kterém se dotazoval subjektů na slovní popis emocí, které jim přijdou první na mysl při poslechu určitých pasáží vybraných skladeb. Od těchto dob bylo předkládáno mnoho modelů, pokoušejících se o empirický popis emocí, z nichž většina spadá do dvou typů modelů; kategorických a dimenzionálních. Pro správnou volbu daného typu modelu emocí pro použití v MER neexistuje žádný konsenzus. Nicméně platí, že při volbě číselných vstupních dat se vyplatí používat dimenzionální modely, zatímco při slovním hodnocení se přirozeně vybírají kategorické modely. Záleží především

na výběru vstupních, **ground-truth** dat použitých při výzkumu. Ty můžeme podle [41] rozdělit na tři typy:

- **Metoda slovních anotací AA (adjectives annotation)**
  Tato metoda bývá povětšinou realizovaná pomocí dotazníků, kdy dotazovaný vybírá vnímané emoce z kategorického modelu emocí. Výsledkem této metody je tzv. *popisný typ (label-type)* ground-truth dat. Tato metoda jednoduše přijímá emoce jako diskrétní označení s jasným emočním významem, snadno se aplikuje a je v souladu se subjektivním pocitem, což je příznivější pro návrh personalizovaných MER systémů. Má také nižší požadavky na profesionalitu respondentů ve smyslu vnímání hudby - stačí, když vyberou správný popis – adjektivum. Nicméně kvůli nejednoznačnosti definice daných popisů – adjektiv – v kategorických modelech (viz kap. 1.2.1) může vést k silné subjektivitě [39].
- **Metoda anotací hudebních parametrů MFA (music feaures annotation)**
  Psychologové uvádějí, že zatímco má ve VA modelu emocí rozměr *arousal* (energie) spojitost s tempem, výškou, hlasitostí a barvou zvuku, rozměr *valence* (pozitivní–negativní, viz. 1.2.2) je spojován s libozvučností (dur/moll) a mírou disonance. Pro využití této metody je potřeba, aby respondenti při poslechu hudebních nahrávek hodnotili jednotlivé hudební parametry. Z těchto informací se na základě předpokladu o spojitosti hudebních parametrů s emocemi určí daná ground-truth data. Tato metoda není často využívána a také neexistuje veřejně dostupná databáze tohoto typu anotace z důvodu neefektivity, velké subjektivity a posléze nepřesným výsledkům [39].
- **Metoda anotací dimenzionálních modelů DMA (Dimensional models annotation method)**
  Metoda DMA vyžaduje přímé manuální hodnocení emocí v rozměrech daného dimenzionálního modelu emocí (typicky VA viz. 1.2.2). Emoce je vyjádřená jako bod v n-dimenzionálním emočním prostoru. Jedná se tedy o číselný typ ground-truth dat. Tento typ anotace však vyžaduje u respondentů vysokou míru profesionality kvůli vysokým nárokům na schopnosti vnímání hudby. Příkladem využívající tento typ anotace je databáze DEAM, popsaná v kap. 3 nebo databáze AMG1608 či DEAP120.

1.2.1 Kategorické modely emocí

Podle tohoto přístupu lidé zažívají emoce, které lze rozčlenit do vzájemně rozdílných kategorií. Pro tento přístup je zásadní tzv. koncept *základních emocí (basic emotions)*, tedy předpoklad, že existuje určitý počet takzvaných *primárních emocí* (*primary emotions*), jako radost, štěstí, smutek, zlost, strach, překvapení, ze kterých

mohou být poté odvozeny tzv. *sekundární třídy emocí*. Tyto základní emoce mohou být nalezeny ve všech kulturách a bývají spojovány s odlišnými vzory fyziologických změn nebo emočních výrazů. Zde je vhodné zmínit kategorický model *FACS* (*Facial Action Coding System*) vyvinutý Eckmanem, který rozlišuje emoce právě podle těchto fyziologických změn a typů mimických pohybů obličeje. Pojem základních emocí byl však hojně kritizován z řady důvodů, zejména proto, že různí badatelé přicházeli s odlišnými sadami základních emocí [32, 41].

Výzkum emocí v hudbě se velmi často provádí ve spojení s analýzou hudebního výrazu. Za účelem rozdělení rozmanité škály emocí co nejobjektivnější a nejjednodušší cestou a zároveň zachování svobody vyjadřování respondentů, odvodila K. Hevnerová seznam přídomků – adjektiv souvisejících s emocemi, uspořádaných v 8 skupinách, jak ukazuje obrázek 1.2. V tomto modelu jsou adjektiva uspořádána do jednotlivých skupin podle významové podobnosti a zároveň jsou jednotlivé skupiny uspořádány v kruhu tak, aby opačné strany kruhu vyjadřovaly i pokud možno opačný význam. Tento model navržený již v roce 1935 byl později přeskupen a doplněn do 10 skupin adjektiv Farnsworthem (1954) a poté v roce 2010 do 9 skupin E. Schubertem, nazvaný jako *UHM* (*Updated Hevner model*) [4].

Některé kategorické modely, jako například AVQ4, vznikly z dimenzionálního Thayer-Russelova (AV) modelu (1.2.2). Využívají rozdělení podle osy A-V již zmíněného modelu na 4 kvadranty jako 4 klasifikační skupiny emocí:

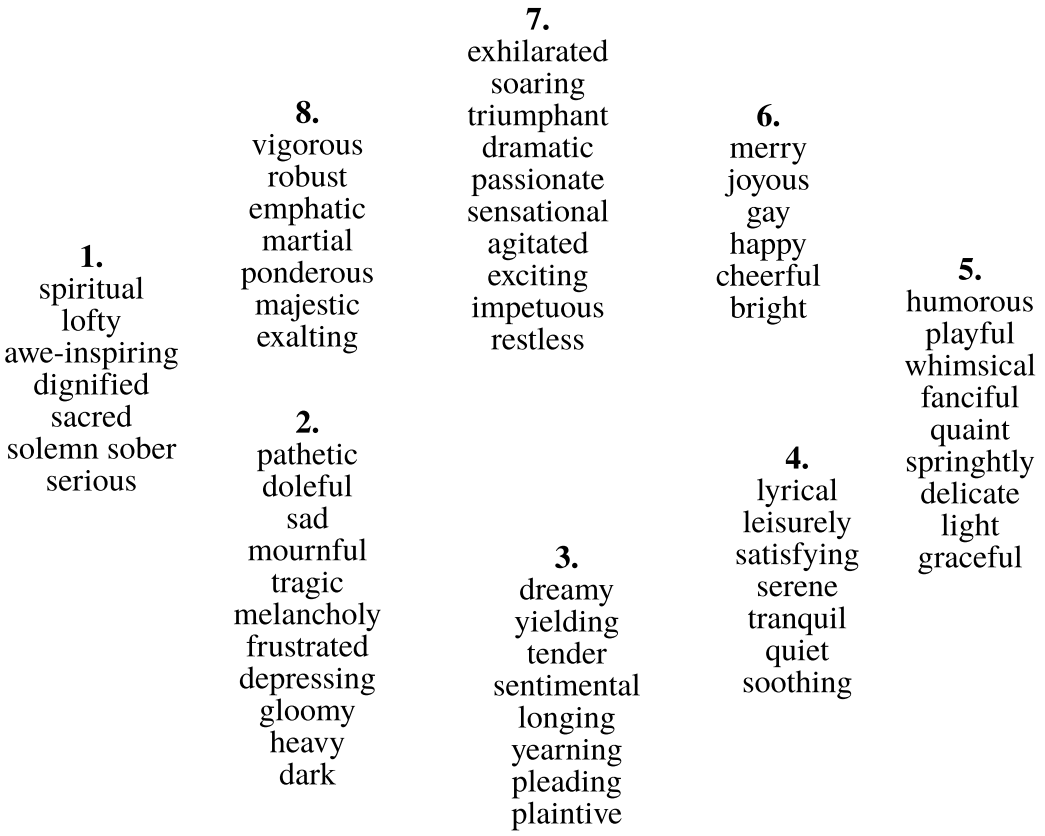
- Q1 – vysoká energie (high energy) a vysoké napětí (high stress)
- Q2 – vysoká energie a nízké napětí
- Q3 – nízká energie a vysoké napětí
- Q4 – nízká energie a nízké napětí

Tento model můžeme spolu s Thayer-Russellovým vidět na obrázku 1.3.

Ukazuje se však, že hlavní nevýhodou kategorického přístupu je příliš malý počet tříd emocí v porovnání s rozmanitostí hudebních emocí, které může člověk při poslechu hudby vnímat. Řešení, které se na druhou stranu nabízí v podobě jemnější granularity, nemusí nutně řešit tento problém, protože jazyk pro popis hudebních emocí je ve své podstatě nejednoznačný a použití daných adjektiv se liší od člověka k člověku. Navíc může používání velkého počtu emočních tříd ještě více zmást a zatížit testovaného respondenta, což má na výsledek dané studie negativní dopad [32].

1.2.2 Dimenzionální modely emocí

Na rozdíl od kategorického přístupu, dimenzionální modely se zaměřují na identifikaci vnímaných emocí na základě jejich pozic v několika daných rozměrech.

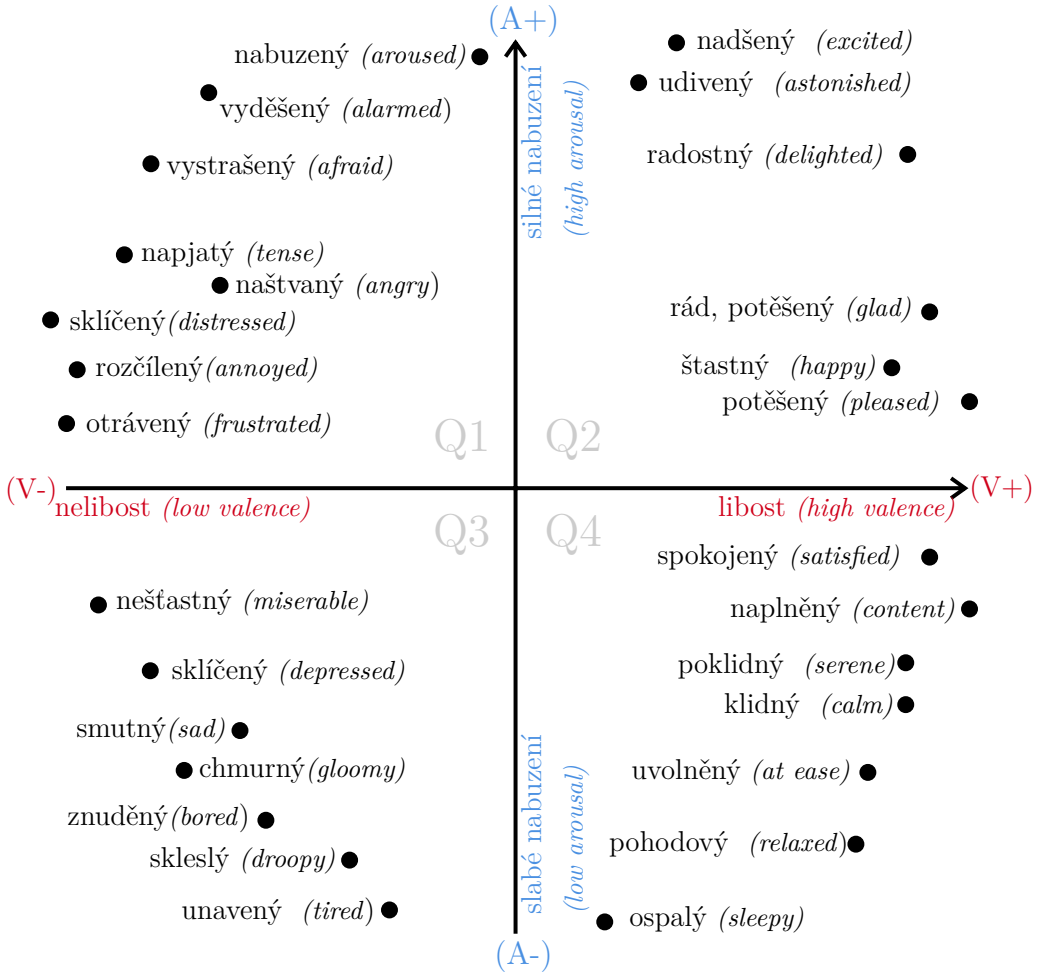


Obr. 1.2: Kruhový model adjektiv podle Hevnerové z roku 1935 [41].

Práce, které měly největší vliv na výběr emočních modelů pro MER jsou výzkumy Russella a Thayera. Russell odvodil tzv. *Kruhový model afektů (Circumplex Model Of Affect* [30]), který sestává z 2-dimenzionální kruhové struktury s rozměry původně anglicky označenými jako ***valence*** (nebo *pleasantness*) a ***arousal*** (také označováno jako *activation*). Thayersovy poznatky zase ověřily spojitost tohoto arousal-valence modelu v hudební doméně, kde může být rozměr arousal definován jako míra energičnosti – jak vzrušující nebo uklidňující emoce hudba sděluje. Valence (také ozn. jako *stress*) zde vyjadřuje jak pozitivně a příjemně, či negativně a nepříjemně hudba působí. Tento systém se souhrnně označuje jako *2D Thayer-Russellův prostor* a můžeme jej vidět znázorněn na obrázku 1.3 [4, 41].

Přestože se AV prostor oproti jiným modelům vyznačuje velkou jednoduchostí a robustností, ukazuje se, že v určitých případech je vhodné model rozšířit o další dimenze jako například *potency* nebo *dominance* pro rozlišení emoce strachu a zlosti, které jsou obě v AV prostoru na stejném místě. Je třeba také zmínit, že díky di-

menzionálnímu přístupu je možné v nahrávkách efektivně zkoumat a zaznamenávat i časový vývoj vnímaných emocí, což vede k hlubším možnostem analýzy MER [4].



Obr. 1.3: Russellův AV model, vícerozměrové váhování, převzato a přeloženo z [30]. Šedou barvou je zde znázorněn také kategorický model AVQ4.

1.3 Metody MER

Metody MER můžeme dělit podle povahy modelu vstupních dat popisujících emoce. V případě tohoto přístupu rozlišujeme metody, které využívají kategorické modely emocí, a metody pracující s dimenzionálními modely a jejich číselnými vstupními daty, což můžeme přehledně vidět v tabulce 1.1, která ukazuje typické varianty existujících MER metod.

Tab. 1.1: Srovnání nejpoužívanějších metod automatického MER [39].

Předpo. výsledek	Typ metody	Typ ground-truth	Model emocí
slovní popis (adjektivum)	Single–label klasifikace	popisný (tagy)	kategorický
více slovních popisů (adjektiva)	Multi–label klasifikace	popisný (tagy)	kategorický
číselné hodnoty	predikce číselné hodnoty	číselný	dimenzionální
spojité rozdělení pravděpodobnosti	predikce spojitého rozdělení pravděpodobnosti	číselný	dimenzionální
MEVD	predikce klasifi. a regrese	popisný (tagy) a číselný	kategorický a dim. model

1.3.1 Rozdělení podle vstupních dat

Pro vstupní data MER predikčních systémů se používají, jak již bylo zmíněno, 2 typy dat: základová ground-truth data (nejčastěji v podobě subjektivních anotací) a extrahované hudební parametry z nahrávek. Ground-truth data se podle kapitoly 1.2 dělí na číselná a popisná. Podle kombinací těchto skupin rozlišujeme MER metody na tyto typy:

- Metody využívající **pouze ground-truth** (základová data)
- Metody využívající **hudební parametry**
- Metody využ. kombinaci **hudebních param. a ground-truth dat** (viz. 1.4)
  - Metody využívající **popisná** (label-type) ground-truth data
  - Metody využívající **číselná** ground-truth data

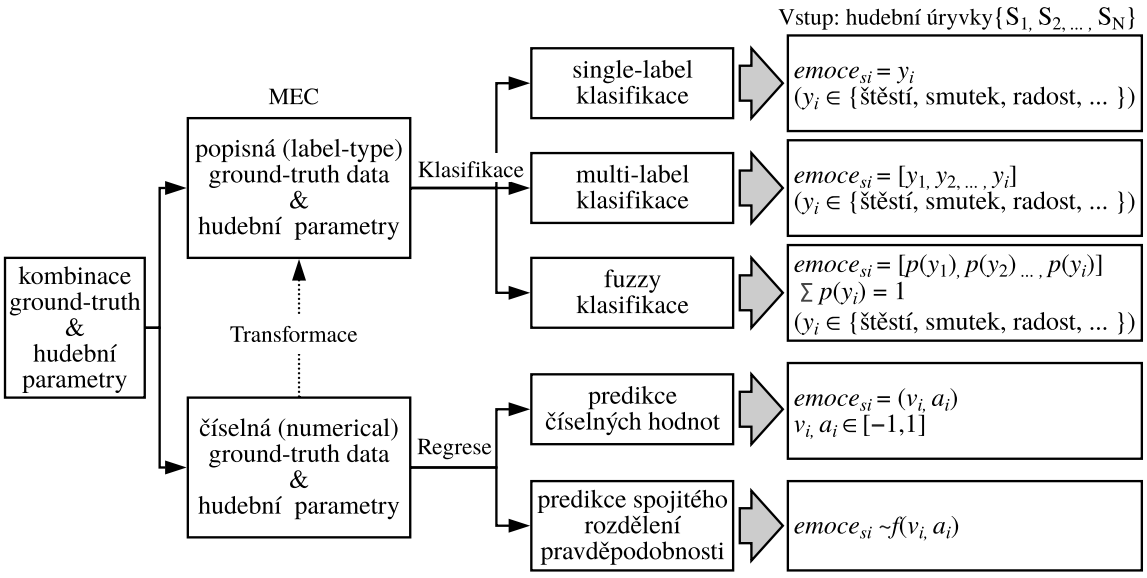
Výsledky metod, které využívají pouze ground-truth data bez hudebních parametrů, bývají přirozeně velmi subjektivní díky náhodnosti, nejasnosti a nepřesnosti lidské anotace, způsobené problematickým procesem získávání těchto dat, jak popisuje kapitola 1.2. Existuje tedy pouze pár metod tohoto typu. Kim et al. [16] rozdělili pomocí algoritmu k–means získané VA hodnoty do 8 oblastí a vyjádřili každou oblast rozložením pravděpodobností adjektiv pomocí statistických metod. Tak mohou být slovní emocionální označení (label-type) predikována pomocí jejich VA hodnot [39].

Metody využívající pouze hudebních parametrů nejsou také optimální, ale od výše zmíněné metody mají výsledky o poznání lepší. Dle analýzy hudebních parametrů vyplývá, že existuje přímý vzájemný vztah mezi vnímanými emocemi a hudebními parametry. Například „vysoko“ znějící tón může evokovat emoci „vzrušení (excitement)“, zatímco „nízký“ tón „smutek“. Bartoszewski et al. [5] vyvinuli v roce

2008 klasifikační systém strojového učení bez učitele pouze s využitím hudebních parametrů s tím, že přesnost predikce byla poté vypočítána s využitím ground-truth anotací. Metody založené na korelaci hudebních prvků a emocí jsou tedy proveditelné. Problémem však zůstává skutečnost, že hudební úryvky odpovídající různým emocím mohou mít stejné určité hudební parametry jako například tempo či efektivní hodnotu (RMS). Navíc bývají vzájemně zaměňovány při procesu shlukování, což ovlivňuje přesnost klasifikace. Obecně nejpoužívanějšími metodami jsou ty, které využívají jak hudební parametry, tak ground-truth data. V dnešní době se jedná téměř o součást definice samotné MER. Rozdělení a vlastnosti této skupiny metod jsou popsány v následující části [39].

1.3.2 Metody využívající kombinaci hudebních parametrů a ground-truth dat)

Ve starších vědeckých studiích byla nejčastěji využívanou metodou, díky omezenosti tehdejších anotačních metod, kterými byly především metody AA (kap. 1.2), a obecnou univerzalitou, klasifikační metoda. Výstupem této metody je získání jednoho nebo více emocionálních označení (labels), korespondujících s daným hudebním úryvkem. V závislosti na různém počtu anotací jednoho hudebního úryvku, rozlišujeme Single-Label klasifikaci, Mutli-Label klasifikaci a Fuzzy klasifikaci, jak přehledně zobrazuje obrázek 1.4.



Obr. 1.4: Klasifikační metody MER v závislosti na různých typech ground-truth vstupních dat [39].

Single–label klasifikace vyjadřuje hudební emoci prostřednictvím jednoho určitého emočního označení daného hudebního úryvku *si*: $emoce_{si} = y_i$ kde y_i je adjektivum – emocionální označení. Tato metoda je jednoduchá, intuitivní a výpočetně relativně nenáročná. Na druhou stranu je zde zanedbána subjektivita a dynamický rozsah lidského vnímání emocí a přesnost klasifikace je nepřímo úměrná počtu kategorií emocí. To je důvod, proč se tato metoda používá čím dál méně [39].

Multi-label klasifikace již bere v úvahu složitost a nepřesnost lidské anotace a klasifikuje emoce hudebního úryvku do většího počtu emočních kategorií: $emoce_{si} = [y_1, y_2, \ldots, y_i]$.

Fuzzy klasifikace vyjadřuje emoci hudebního úryvku jako nespojité (diskrétní) rozdělení pravděpodobnosti určitého počtu emocionálních kategorií: $emoce_{si} = [p(y_1), p(y_2), \ldots, p(y_i)]$. Přestože jsou tyto metody v porovnání se Single–label klasifikaci znatelně komplexnější, vynikají vyšší přesnosti a nižším vlivem subjektivity při hodnocení výsledků experimentů. Tato práce se však bude blíže věnovat metodám, které kombinují hudební parametry s číselným typem ground-truth dat [39].

Kombinace číselných typů anotací a hudebních parametrů

Stejně, jako předchozí typy ground-truth dat, mohou být i číselné typy těchto dat transformovány pomocí souřadnic a poté použity pro proces **klasifikace**. Podle literatury [39], která dokládá příklady reálných studií, však výsledky těchto postupů vykazují malou výkonnost a velkou nepřesnost. Hudební segmenty s poměrně shodnými hodnotami v AV prostoru byly rozděleny do naprosto odlišných emočních kategorií. Bylo tedy usouzeno, že pro číselný typ numerických ground-truth dat je vhodnější použití **regresních metod učení**. Regresní metody strojového učení vyžadují právě tuto kombinaci číselných typů anotací a hudebních parametrů a umožňují mapování jejich vzájemného vztahu. Posun od předešlých popisných metod k číselným, a tedy přechod od kategorických modelů k dimenzionálním, ukazuje jasný vývoj MER. Regresní metody využívající číselný typ ground-truth dat mohou být podle obr. 1.4 rozděleny na dva typy.

Predikce číselných hodnot vyjadřuje hudební emoci jako bod, který určuje přesně specifikovanou pozici emoce v dimenzionálním prostoru. V prostoru VA je emoce hudebního úryvku *si* definována: $emoce_{si} = (v_i, a_i)$, kde $v_i, a_i \in [-1; 1]$.

Predikce spojitého rozdělení pravděpodobnosti je metoda, která vyjadřuje hudební emoci jako spojité rozdělení pravděpodobnosti v dimenzionálním prostoru: $emoce_{si} = f(v_i, a_i)$, kde $f(\cdot)$ je dvourozměrná funkce *Gaussova rozdělení pravděpodobnosti*. Díky tomuto lze reflektovat rozdílné emocionální prožitky různých subjektů poslouchajících tutéž skladbu, zmírnit vliv subjektivity označování emocí a pomoci

vytvářet výzkumníkům lepší a více personalizované MER systémy.

Existuje mnoho metod, jak vyjádřit hudební emoci jako jeden diskrétní bod v emočním prostoru. Mezi standardní a efektivní metody patří *metoda vícenásobné lineární regrese* se zkratkou MLR (Multiple Linear Regression). Metoda *SVR (Support Vector Regression)* je základní nelineární regresní algoritmus, který mapuje nelineárně vstupní vektor do mnoho dimenzionálního parametrického prostoru a vytváří mnoho dimenzionální lineární rozhodovací funkci k realizaci nelineární regrese. Hojně používané jsou také další metody jako *regrese gaussovskými procesy* – GPR (Gaussian Process Regression) či algoritmus *k–nejbližších sousedů KNN* (K-Nearest Neighbours).

2 Hudební parametry

Cílem extrakce datových parametrů je redukce informace skladeb do deskriptorů, které je mohou plně popsat. Datové parametry se v oblasti MER dělí na 2 skupiny. První skupinou jsou hudební parametry. To jsou akustické veličiny a hudební text, které bývají z nahrávek získávány pomocí odpovídajícího výpočetního procesu. Odrážejí daný hudební styl, strukturu, hudební emoce a jsou přímo spjaty s formátem hudební nahrávky. Druhou skupinou jsou ground-truth data, tedy označení (onálepkování) daných skladeb příslušnou emocí, kterou v testovacím subjektu, posluchači, hudební nahrávka vyvolává. Častým bývá i rozlišování tzv. *nízkoúrovňových (low-level)* parametrů podle toho, zda jsou extrahovány přímo z reprezentace signálu.

Ve většině MER studiích se vyskytuje stejné pozorování, tedy že modelování emocí spojených s rozměrem valence (jako radost či smutek) je vždy obtížnější než pro emoce spojené s rozměrem arousal. Valence je totiž spojena s harmonickým a melodickým obsahem nahrávek a je nemožné tyto emoce předpovídat pouze pomocí nízkoúrovňových spektrálních parametrů, jako u arousal. Intenzita je považována za základní parametr a často silně koreluje s vnímáním rozměru arousal a bývá použitá ke klasifikaci tohoto rozměru. [39].

Energie písně je blízce spjata s vnímáním v rozměru arousal. Rytmus je často popisován pomocí tempa nebo frázování. U skladeb s rychlým tempem je často vnímána vysoká úroveň hodnoty arousal. Zatímco plynulý, měnící se rytmus je spojován s pozitivní hodnotou valence, pevný rytmus s valenci negativní [41]. V tabulce 2.1 můžeme vidět pět běžně používaných parametrů, které Yang ve svém obecném shrnutí oblastí MER [39] z roku 2016 vyzdvihuje.

Tab. 2.1: Přehled běžně užívaných deskriptorů – parametrů pro MER [39].

	Parametr	**Zkratka**	**Rozměr**
1	Melovské kepstrální koeficienty	MFCC	20-D
2	Oktávový spektrální konstrast (Octave-based Spectral Contrast)	OSC	14-D
3	Statistical Spectrum Descriptors	SSD	4-D
4	Chromagram	Chroma	12-D
5	Daubechies Wavelet Coefficient Histrograms	DWCH	21-D

Protože se bude regresní vyhodnocovací systém této práce inspirovat cestou současných studií MER, bude používáno a extrahováno velké množství z celé škály hudebních parametrů (podrobněji viz. praktická část této práce). Tyto parametry budou téměř všechny popsány v následující podkapitole.

2.1 Extrakce parametrů

Pro potřeby této práce byly vybrány dva nástroje pro extrakci parametrů jak s poměrně rozdílným přístupem jednotlivých nástrojů, tak i rozdílnými extrahovanými parametry. Jedná se o nástroje běžně používané v oblastech MIR a MER. Takto bude možné otestovat oba soubory parametrů a provést jejich srovnání.

2.1.1 MIRtoolbox

Jedná se o volně dostupný externí toolbox v prostředí *MATLAB*, který umožňuje extrakci hudebních parametrů, které jsou obecně považovány jako vhodné pro oblast získávání informací z hudby (MIR). Práce s tímto sofistikovaným toolboxem není složitá a má velmi dobře zpracovanou dokumentaci [22].

2.1.2 openSMILE

Pro druhou extrakci parametrů ze zvukových nahrávek je použit nástroj openSMILE[1] [10] vyvinutý společností audEERING, který nabízí možnost výpočtu velkého množství parametrů ze zvukových vzorků. Je psán v jazyce C++ a nabízí rychlé a efektivní zpracování dat, podporu multi-threading pro paralelní procesy extrakce a vysokou modularitu. Tento nástroj byl vybrán především ze dvou hlavních důvodů. Prvním důvodem je, že byl používán spolu s databází DEAM pro soutěž iniciativy MediaEval. Druhým důvodem může být zajímavé srovnání úspěšnosti MIR hudebních parametrů ve srovnání parametrů extrahovaných z openSMILE, které lze zařadit spíše mezi řečové parametry (příznaky), přestože byly používány pro rozpoznávání hudebních emocí. Soubor parametrů, který bude použit v této práci je velmi podobný tomu, který byl právě použit pro potřeby této soutěže.

2.2 Parametry MIRtoolboxu

V tabulce 2.2 můžeme vidět výčet 23 parametrů, které byly extrahovány ze souboru nahrávek pomocí nástroje OPENSmile.

V tabulce 2.3 se nachází výčet statistických parametrů, které MIRtoolbox nabízí, a které byly vypočteny pro každý extrahovaný parametr pro každou skladbu. Obsahuje jak standardní střední hodnotu parametru, směrodatnou odchylku, tak například parametr **slope**, který značí lineární sklon trendu napříč jednotlivými okny (hodnotami parametru v čase). Jinými slovy se jedná o derivaci přímky, která

[1] www.audeering.com/technology/opensmile/

Tab. 2.2: Přehled parametrů vyextrahovaných nástrojem MIRtoolbox. V kap. 2.2.2 se nachází popis těchto parametrů [35].

Rozměr	**Parametr (LLD)**	**Zkratka**
1-D	efektivní hodnota signálu	m_rms
1-D	počet přechodů nulovou úrovní	m_zerocross
1-D	spektrální fluktuace	m_spectralflux
10-D	fluktuace v okt. pásmech	m_sflux_oct
1-D	pokles spektrální energie 95%	m_roloff95
1-D	pokles spektrální energie 85%	m_roloff85
1-D	bělost spektra	m_brightness
1-D	spektrální centroid	m_centroid
1-D	spektrální entropie	m_specentropy
1-D	spektrální plochost	m_flatness
1-D	spektrální špičatost	m_kurtosis
1-D	spektrální šikmost	m_skewness
1-D	spektrální rozptyl	m_spread
1-D	drsnost spektra	m_roughness
1-D	nepravidelnost spektra	m_irregularity
1-D	doba náběhu	m_attack
1-D	tempo	m_tempo
1-D	fluktuace tempa	m_fluct
12-D	chromagram	m_chroma
13-D	Melovské kepstrální koef.	m_mfcc
13-D	Δ Melovské kepstrální koef.	m_dmfcc
13-D	$\Delta\Delta$ Melovské kepstrální koef.	m_ddmfcc
1-D	parametr nízké energie	m_lowenergy

nejlépe odpovídá sledované závislosti. Mezi další parametry patří **perFreq** – kmitočet maximální periodicity detekovaný v daném rozsahu hodnot vyjádřený výpočtem autokorelační posloupnosti. **perAmp** je normalizovaná amplituda periodicity a **perEntropy** značí Shannonovu entropii autokorelační posloupnosti [35].

2.2.1 Parametry časové oblasti

Efektivní hodnota signálu

Efektivní hodnota signálu (root mean square) může určitým způsobem popisovat dynamiku hudebního signálu. Při malém dynamickém rozsahu signálu se její hodnota

Tab. 2.3: Přehled statistických parametrů MIRtoolboxu [35].

Parametr	Zkratka
střední hodnota	mean
směrodatná odchylka	std
lineární sklon trendu	slope
kmitočet maximální zjištěné periodicity	perFreq
normalizovaná amplituda zjištěné periodicity	perAmp
Shannonova entropie autokorelační posloupnosti	perEntropy

bude blížit špičkové hodnotě. Efektivní hodnota je definována vztahem:

$$RMS_t = \sqrt{\frac{1}{K} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1)} s(k)^2}, \tag{2.1}$$

kde K udává velikost rámce vzorků (počet vzorků v každém rámci) a $s(k)^2$ je druhá mocnina amplitudy k-tého vzorku [18].

Počet přechodů nulovou úrovní

Tento parametr, označován také jako (ZRC - *Zero Crossing Rate*), udává, kolikrát se změní znaménko hodnoty amplitudy. Využívá se mimo jiné pro detekci perkusivních zvuků. Je definován vztahem:

$$ZRC_t = \frac{1}{2} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1)} \left| sgn(s(k)) - sgn(s(k+1)) \right|, \tag{2.2}$$

kde K je velikost rámce vzorků a s(k) je amlituda k-tého vzorku [18].

2.2.2 Parametry spektrálního rozložení

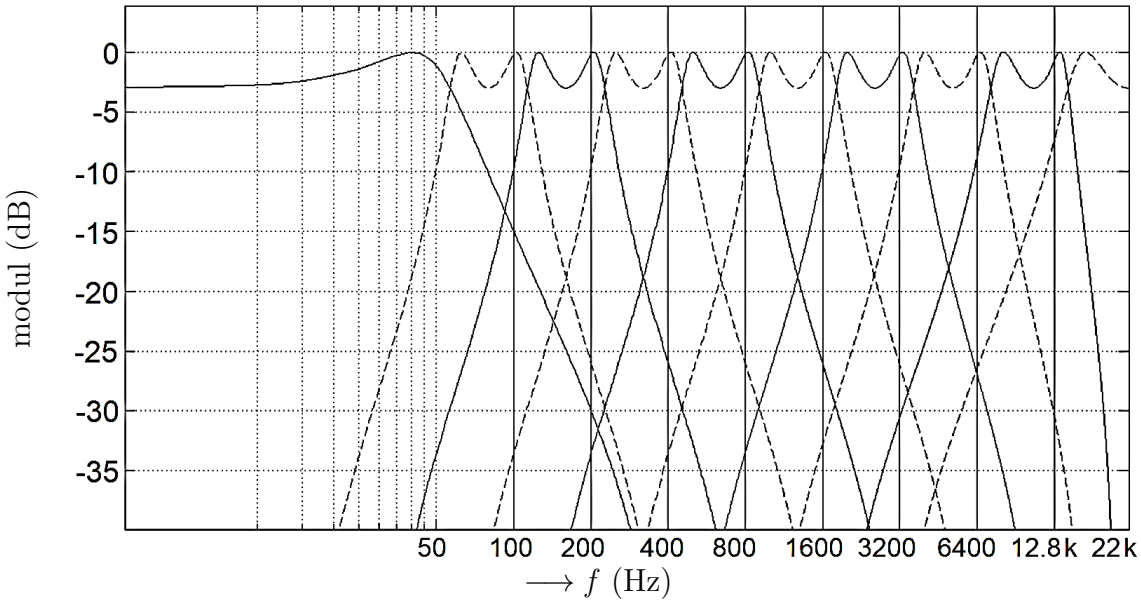
Spektrální fluktuace

Tento parametr, anglicky označován jako SF - *Spectral Flux*, udává změny ve výkonovém spektru mezi po sobě jdoucími rámci. Je definován vztahem:

$$SF_t = \sum_{n=1}^{N} (D_t(n) - D_{t-1}(n))^2, \tag{2.3}$$

kde D_t značí rámec po rámci normalizovanou frekvenční distribuci za čas t. N je číslo nejvyššího kmitočtu [22].

Parametr spektrální fluktuace je v této práci využit ve dvou variantách. Parametr **m_spectralflux** udává hodnotu fluktuace přes celé spektrum, zatímco **m_flux_oct**



Obr. 2.1: Rozdělení spektra pomocí eliptických oktávových filtrů druhého řádu pro parametr spektrální fluktuace [22].

je 10 rozměrný parametr, který udává fluktuaci v 10 frekvenčních pásmech daných oktávovými eliptickými filtry druhého řádu s použitým filtrem dolní propusti na 50 Hz. Na obrázku 2.1 můžeme vidět zobrazení těchto filtrů [22].

Pokles spektrální energie

Pokles spektrální energie (*spectral rolloff*) vyjadřuje hodnotu frekvence, pod kterou se nachází nachází určené množství energie. Ve většině případů se udává hodnota energie 85 % či 95 %. Tento parametr často slouží k odhadnutí množství vysokých frekvencí v nahrávce [22].

Bělost spektra

Tento parametr označovaný v anglické literatuře jako *spectral brightness* vyjadřuje množství energie, která se vyskytuje nad určitou frekvencí. Běžně se volí hodnota frekvence 1 kHz, 1,5 kHz nebo 3 kHz. Tato práce bude počítat s hodnotou 1,5 kHz.

Spektrální centroid

Pomocí parametru (angl. spectral centroid) je analyzována tzv. „jasnost“ barvy zvuku a udává frekvenční pásmo, ve kterém je koncentrováno nejvíce energie analy-

zovaného signálu. Rovnice je dána:

$$SC_t = \frac{\sum_{n=1}^{N} m_t(n) \cdot n}{\sum_{n=1}^{N} m_t(n)}, \tag{2.4}$$

kde N udává číslo nejvyššího kmitočtu a $m_t(n)$ je modul signálu ve frekvenční oblasti za čas t ve frekvenčním pásmu n [18].

Spektrální entropie

Míra entropie velmi zjednodušeně vyjadřuje míru neurčitosti. Parametr spektrální entropie je vyjádřen pomocí tzv. Shannonovy entropie [22].

Spektrální plochost

Parametr plochosti je vypočítán jako poměr geometrického průměru spektra a jeho aritmetického průměru a popisuje tvar spektra. Čím více je spektrum signálu členité, tím bude hodnota spektrální plochosti blíže nule. Naopak bílý šum, který má rovnoměrné rozložení výkonové spektrální hustoty, bude mít hodnotu tohoto parametru blížící se jedné zdola [18].

Spektrální špičatost

Spektrální špičatost (angll. *spectral kurtosis*) udává míru „špičatosti“ spektra v okolí její střední hodnoty. Hodnota tohoto parametru rovná nule značí normální rozdělení. Kladná hodnota tohoto parametru značí vysokou míru špičatosti spektra, což znamená, že většina spektrálních složek leží blízko střední hodnoty. Pokud je hodnota špičatosti naopak záporná, spektrum je plošší a rozložení spektrálních složek je rovnoměrnější [18].

Spektrální šikmost

Spektrální špičatost (*spectral kurtosis*) udává míru „šikmosti“ spektra v okolí jeho střední hodnoty. V případě, že je hodnota spektrální šikmosti rovna nule, jde o symetrické rozložení spektrálních složek kolem střední hodnoty spektra. Pokud je hodnota tohoto parametru větší než nula, je více spektrálních složek zastoupeno ve vysokých kmitočtech a naopak, pokud je hodnota koeficintu menší než nula, je více spektrálních složek obsaženo v nízkých kmitočtech [18].

Drsnost spektra

Plomp a Levelt navrhli v roce 1996 metodu pokoušející se o stanovení míry nelibozvuku dané zvukové nahrávky, která se označuje anglicky *roughness*. Jev nelibozvučnosti (disonance) vzniká, nacházejí-li se jednotlivé harmonické složky zvuku

velmi blízko sebe ve frekvenční oblasti. V praxi se tento parametr počítá jako poměr amplitud všech možných sousedících maxim (špiček) získaných ze spektra analyzovaného signálu [22].

Nepravidelnost spektra

Nepravidelnost spektra (angl. *spectral peaks variability*) udává míru variability po sobě následujících vrcholů spektra [22].

Doba náběhu

Parametr doby náběhu, označovaný také jako *attack time*, je definován jako čas, za který se hodnota energie zvukové události změní z 10 % na 90 % celkové energie dané události [20].

Parametr nízké energie

Parametr nízké energie (*low energy*) může být použit k získání informace o časovém rozložení energie. Lze pomoci něj zjistit, zda zůstává energie v celém pásmu konstantní, nebo jsou některé úseky více kontrastní, než jiné [22]. Jeden z možných způsobu výpočtu hodnoty nízké energie je výpočet procentuálního zastoupení těch analyzovaných oken, které mají energii nižší, než je průměrná energie zkoumaného úseku [37].

2.2.3 Parametry popisující tempo zvukové nahrávky

Tempo

Tempo hudební nahrávky je běžně uváděno v počtu úderů za minutu (*beats ber minute, BPM*). Jedná se o hudební parametr, který informuje o tom, jak rychle se daná hudební skladba hraje. Parametr tempa extrahovaný pomoci MIRtoolboxu je určen na základě detekce periodicity z křivky počátků zvukových událostí [22].

Fluktuace tempa

Parametr fluktuace tempa (*tempo fluctuation*) je určován ze spektrogramu zkoumaného signálu. Ten je rozdělen na segmenty s délkou okna 23 ms a poté je frekvenční osa přepočítána na melovskou či barkovou škálu. Následně jsou odhadnuty maskovací efekty lidského ucha a je provedena rychlá Fourierova transformace na každém pásmu. Následně je sečteno výsledné spektrum napříč pásmy, což vede k celkovému shrnutí spektra, které ukazuje celkové rozdělení rytmických periodicit. Výstupem výpočtu je matice rytmické periodicity [22].

2.2.4 Chromagram

Chromagram (také ozn. jako *Harmonic Pitch Class Profile*) vyjádřený pomocí tzv. *chroma parametru* vyjadřuje distribuci spektrální energie v jednotlivých tónových výškách („*chromas*“) či tónových třídách. Jedná se o vektor o 12 dimenzích, kde jedna dimenze zastupuje jeden tón tónových tříd. Tento typ parametru se stal velmi důležitým při úkolech synchronizace hudebních nahrávek nebo při rozpoznávaní akordů. Pro výpočet tohoto parametru existuje více metod, přičemz MIRtoolbox využívá pro výpočet právě chromagram. V prvním kroku je pomocí FFT vypočítáno spektrum v logaritmickém měřítku s výběrem např. pouze horních 20 dB – a to pouze pro určité užitečné frekvenčním pásmo. Poté je vypočítán chromagram pro jednotlivé tónové třídy [22].

2.3 Parametry OPENSmile

V tabulce 2.4 můžeme přehledně vidět jednotlivé parametry extrahované pomocí programu OPENSmile.

Tab. 2.4: Přehled parametrů vyextrahovaných nástrojem OPENSmile [10].

Rozměr	**Parametr (LLD)**	**Zkratka**
1-D	základní frekvence (viz kap.2.3.1)	F0final
1-D	obálka základní frekvence	F0fin_Env
1-D	param. znělosti F_0, z autokor. fce (viz. 2.3.2)	voicingFinalUnclipped
8-D	Line Spectral Pairs (viz. 2.3.5)	lspFreq
15-D	Melovské kepstrální koeficienty (viz. 2.3.3)	mfcc
8-D	Melovské spektrální koeficienty (viz. 2.3.4)	MelFreqBand
1-D	normalizovaná intenzita umocněna 0,3	loudness
1-D	jitter (viz. 2.3.6)	jitterLocal
1-D	DDP jitter (viz. 2.3.6)	jitterDDP
1-D	shimmer (viz. 2.3.6)	shimmerLocal

Všechny tyto parametry byly vyexportovány v jejich základní podobě a navíc také v podobě (delta koeficientu) diferenciálu prvního řádu. Ve vyhodnocení a kódu jsou tyto parametry označeny příponou _de. Dále byly pro každý parametr vypočítány statistické parametry, které již slouží pro regresní učení samotné. Výčet a krátký popis se nachází v tabulce 2.5.

Tab. 2.5: Přehled statistických parametrů OPENSmile [10].

Parametr	**Zkratka**
pozice max. hodnoty pro daný úryvek (norm. do 0 – 1)	maxPos
pozice min. hodnoty pro daný úryvek (norm. do 0 – 1)	minPos
střední hodnota	mean
sklon (k) směrnice přímky lineární regrese	linregc1
svislý posun (q) přímky lineární regrese	linregc2
absolutní chyba (rozdíl střední hodnoty a lineární aprox.)	linregerrA
kvadratická chyba (rozdíl střední hodnoty a lineární aprox.)	linregerrQ
směrodatná odchylka	stddev
šikmost	skewness
špičatost	kurtosis
1. kvartil	quartile1
2. kvartil	quartile2
3. kvartil	quartile3
mezikvartilové rozpětí (1. – 2. kvartil)	iqr1_2
mezikvartilové rozpětí (2. – 3. kvartil)	iqr2_3
mezikvartilové rozpětí (1. – 3. kvartil)	iqr1_3
99. percentil	percentile990
1. percentil	percentile10
mezipercentilové rozpětí	pctlrange0_1
čas, kdy je parametr nad 90 % rozsahu + min (%)	uplvlt90
čas, kdy je parametr nad 75 % rozsahu + min (%)	uplvlt75

2.3.1 Základní frekvence

Základní (fundamentální) frekvence signálu patří jak mezi základní parametry řeči, tak i hudebního signálu. Pro její výpočet je nezbytné na začátku provést segmentaci signálu, tedy rozdělení vstupního signálu na menší úseky, ze kterých se poté počítají parametry základního tónu. Většina metod pro určení základního tónu (f_0) používá při výpočtu rychlou Fourierovu transformaci (FFT). Mezi základní metody určení parametrů základního tónu patří tyto metody [33]:

1. detekce základního tónu v časové oblasti
2. detekce základního tónu v kmitočtové oblasti
3. detekce základního tónu v reálném kepstru

Nástroj OpenSMILE, který byl v této práci využit pro extrakci parametrů, používá třetí typ z výše uvedeného seznamu - detekci pomocí kepstra [10]. Postup výpočtu této metody je následující:

- standardní segmentace signálu
- výpočet modulu spektra pro jednotlivé segmenty pomocí FFT
- logaritmování modulu s využitím přirozeného logaritmu
- provedením zpětné FFT je získáno reálné kepstrum každého segmentu
- Z kepstra bývá případě řečového signálu vyseknuta pomocí okna jeho část mezi 60 Hz a 400 Hz, kde se může frekvence f_0 nacházet. Běžný rozsah základního kmitočtu hudebních nástrojů však bývá mezi 20 a 50 Hz (pro kontrabas) a $3-5$ kHz (pro pikolu) [23]. Potom je ve vyseknutém okně nalezena maximální hodnota a pomocí ní se určí kmitočet základního tónu daného segmentu [33].

Parametr vyjadřující základní frekvenci je v praktické části práce označován jako **F0final** a obálka tohoto parametru jako **F0fin_Env**.

2.3.2 Parametr znělosti f_0

Voicing probability p_v (přeloženo jako znělost) ve vztahu k základní frekvenci indikuje, do jaké míry je signál podobný ideálnímu harmonickému signálu (vysoká znělost, high probability) či šumovému signálu (nízká znělost). Je založený na principu autokorelační metody a vyskytuje se v mnoha detektorech výšky tońu (PDA, pitch detection algorithms). Zjišťování základního tónu v segmentovaném signálu pomocí autokorelační funkce (ACF) je poměrně robustní a jednoduchá metoda. Protože je základní frekvence f_0 určována pouze pro znělý, harmonický signál, je v každém rámci provedeno rozhodnutí, zda-li se jedná o zmíněný, či šumový signál. Ve znělých hlasových segmentech jsou nalezena maxima, která mají určité vzájemné minimální a maximální vzdálenosti $T_{0,min}$ a $T_{0,max}$ [9].

Autokorelační metoda předpokládá, že základní perioda T_0 je dána polohou nejvyšší hodnoty autokorelační funkce v daném segmentu v rozmezí $T_{0,min}$ do $T_{0,max}$. Amplituda této špičkové hodnoty normalizována amplitudou nultého koeficientu ACF může vyjadřovat právě znělost (voicedness) signálu. Tento parametr znělosti, je stanoven jako:

$$p_v = \frac{ACF_{max}}{ACF_0}, \tag{2.5}$$

kde ACF_{max} je maximální hodnota v rozsahu $T_{0,min} \dots T_{0,max}$ a ACF_0 je energie rámce (nultý koeficient ACF) [9]. Tento parametr je použit v základním souboru parametrů, vyextrahovaném pomocí nástroje OPENSmile. V praktické části práce je uveden pod zkratkou **voicingFinalUnclipped**.

2.3.3 Melovské kepstrální koeficienty – MFCC

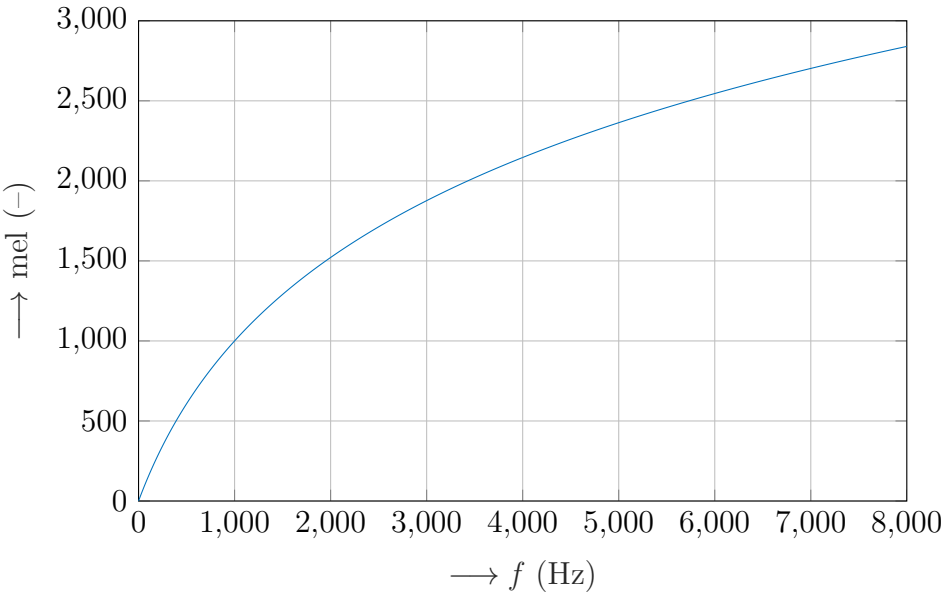
MFCC koeficienty jsou v oblasti MER vůbec nejpoužívanějšími parametry, obzvláště pro predikci dimenze *valence*. Mají svůj původ v oblasti *zpracování řeči*, ale nacházejí

své místo i všude tam, kde je potřeba reprezentovat barvu zvuku. Koeficienty MFCC byly první parametry, které braly v úvahu nelineární maskovací vlastnosti lidského slyšení. V principu je jeho získání provedeno výpočtem reálného kepstra signálu, který je poté ve spektrální oblasti podroben nelineární transformaci kmitočtové osy - převod Hz na jednotky mel [33]. Jejich výpočet se skládá z těchto kroků:

1. Dělení signálu na krátké časové rámce (20 až 40 ms), ořezání okrajů Hammingovým oknem
2. Fourierova transformace signálu
3. Rozdělení spektra pomocí banky složené z n trojúhelníkových filtrů, přepočet na Mel-frekvenci
4. Výpočet absolutní hodnoty a logaritmování
5. Diskrétní kosinová transformace [24]

Mel frekvence je škála frekvence, která co nejvěrněji popisuje nelineární vnímání výšek tónů a jejích intervalů lidského ucha. Přibližně do 1 kHz je závislost přibližně lineární, nad 1 kHz již logaritmická. Závislost přepočtu kmitočtu na melovskou stupnici můžeme vidět na obrázku 2.2. Pro přepočet se používá následující vztah:

$$Mel(f) = 1127,01048 \cdot ln\left(1 + \frac{f}{700}\right) = 2595,0375 \cdot log\left(1 + \frac{f}{700}\right). \tag{2.6}$$



Obr. 2.2: Melovská stupnice v závislosti na kmitočtu.

Koeficienty MFCC se používají také v oblasti MIR, obzvláště v oblasti rozeznávání hudebních žánrů či hudebních nástrojů. Ve většině případů počet používaných koeficientů variuje mezi 4 a 20 [18, 23].

V praktické části této práce jsou tyto koeficienty nazvány zkratkou **mfcc**.

2.3.4 Melovské spektrální koeficienty

V této práci je také využit parametr označován zkratkou **logMelFreqBand**. Podle [10] se jedná o logaritmovanou energii osmi mel-frekvenčních pásem od 0 Hz do 8 kHz. Jde o parametry česky označované jako melovské spektrální koeficienty. Tyto parametry jsou získávány stejným způsobem, jako koeficienty MFCC, akorát je při jejich výpočtu vynechán poslední krok zpětného převodu spektra do nelineární časové oblasti pomoci diskrétní kosinové transformace. Postup výpočtu těchto koeficientů je tedy výpočet spektra jednotlivých dílčích rámců pomoci FFT a poté vynásobení modulu spektra bankou melovských filtrů [33].

2.3.5 LSP - Line Spectral Pairs

Line Spectral Pairs koeficienty jsou přímou matematickou transformací souboru *lineárních prediktivních koeficientů* LPC tak, jak jsou generovány mnoha systémy jako *CELP (Codebook-Excited Linear Prediction)*. Použití LSP je velmi populární díky dobré kvantifikační charakteristice a účinnosti následné reprezentace. Jednotlivé části párů jsou často nazývány jako *Line Spectral Frequencies* (LSF nebo LSP frequencies). Tento parametr je také hojně využíván v oblasti MER a nachází se v souborech parametrů určených k predikci hudebních emocí.

Lineární prediktivní kódování

Lineární prediktivní kódování (LPC) je jedna z nejefektivnějších metod analýzy původně řečového signálu k získání spektrální obálky signálu [15]. Je založena na matematickém popisu hlasového ústrojí, tedy ploše reprezentované tubusem s proměnným průměrem. Nejdůležitější součástí LPC je *lineární prediktivní filtr*, který linearizuje hodnotu dalšího vzorku, která je určena lineární kombinací zase vzorku předchozího [36]. Od lineárního prediktivního kódování byly odvozeny tzv. LSP *koeficienty*, které jsou využity k výpočtu již zmíněných Line Spectral Frequencies. Ty jsou využity v této práci jako jeden z extrahovaných parametrů zvukových ukázek.

Princip Line Spectral Pairs

Jak již bylo zmíněno, LSP koeficienty jsou odvozeny od filtru lineárního prediktivního kódování reprezentující vokální trakt v řečovém signálu pro řád analýzy p [25]:

$$A_p(z) = 1 + a_1 z^{-1} + a_{2p} z^{-2} + ... + a_p z^{-p}. \tag{2.7}$$

Definujme dva polynomy řádu p+1 získané z prediktoru $Ap(z)$, které pojmenujme $P(z)$ a $Q(z)$. Tím získáme asymetrické a symetrické složky daného signálu, které

jsou založeny na porovnávání svých koeficientů. Výsledný polynom definuje model trubice, která reprezentuje hlasový trakt člověka:

$$A_p(z) = \frac{P(z) + Q(z)}{2} \tag{2.8}$$

Tyto dva polynomy, kde je možnost vytvořit zpětnou vazbu s jsou vytvořeny z LPC polynomu jsou definovány následujícími rovnicemi [25]:

$$\begin{aligned} P(z) =& A(z) + z^{-(p+1)}A(z^{-1}) \\ Q(z) =& A(z) - z^{-(p+1)}A(z^{-1}) \end{aligned} \tag{2.9}$$

Kořeny těchto dvou polynomů jsou dílčí složky jednotlivých lineárních spektrálních párů, tedy již zmíněné LS-frekvence. V této práci je zmíněný parametr označován jako **lspFreq** a vyjadřuje 8 LSP frekvencí vypočítaných z 8 LPC koeficientů. Parametr byl pro potřeby této práce extrahován pomocí nástroje openSMILE.

2.3.6 Jitter a shimmer

Tyto parametry patří především do skupiny řečových parametrů („příznaků“) a jsou získány z průběhu hlasivkových pulzů. Dobře detekují například stres. Jedná se o příznaky, které popisují změnu velikosti pulzů A_g a změnu jejich periody T_g. Uvažujme segment, který obsahuje N period hlasivkových pulzů. Jitter, označovaný také jako třes nebo chvění hlasivek, udává rozdíl v délce dvou sousedních period hlasivkových pulzů dělený průměrnou délkou periody [33]:

$$J_g = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} \mid T_g[i] - T_g[i-1] \mid}{\frac{1}{N}\sum_{i=0}^{N-1} T_g[i]} \tag{2.10}$$

Změna jitteru v jednotných rámcích může být nazvána také jako „jitter jitteru“. Jedná se tedy o změnu rozdílu délky dvou sousedních period hlasivkových pulzů dělený průměrnou délkou periody. Tento parametr se označuje jako DDP jitter (Difference of Differences of Periods) a je definován takto [9]:

$$J_{ddp} = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1} \mid\mid A_g[i] - A_g\,[i-1] \mid - \mid A_g[i-1] - A_g[i-2] \mid\mid}{\frac{1}{N}\sum_{i=0}^{N-1} A_g[i]} \tag{2.11}$$

Jako vibrace nebo kolísání pulzů se označuje shimmer, který je vyjádřen jako rozdíl velikosti dvou sousedních pulzů dělený jejich průměrnou velikostí [33]:

$$S_g = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} \mid A_g[i] - A_g[i-1] \mid}{\frac{1}{N}\sum_{i=0}^{N-1} A_g[i]} \tag{2.12}$$

V praktické části této práce lze tyto parametry nalézt pod zkratkami **jitterLocal**, **jitterDDP** a **shimmerLocal**.

2.4 Algoritmy pro výběr a předzpracování parametrů

2.4.1 PCA – Analýza hlavních komponent

Metoda PCA (*Principal Component Analysis*) je často používaná metoda ke snížení dimenze dat s co nejmenší ztrátou informace. Bývá označována také jako *Karhunen-Loèveho transformace*. Tato metoda mapuje vstupní proměnné (povětšinou vektor prvků $\boldsymbol{v}$) do nového souřadného systému lineární kombinací jednotlivých parametrů:

$$\mathbf{u}(n) = \mathbf{T}^{\mathrm{T}} \cdot \mathbf{v}(n). \tag{2.13}$$

Výsledný vektor $\mathbf{u}(n)$ jsou data v novém souřadnicovém systému pro pozorování n a $\mathbf{T}^{\mathrm{T}}$ je transformační matice obsahující různé lineární kombinace pro vstupní vektor parametrů $\mathbf{v}(n)$. Počet parametrů ve vektoru bude označováno jako p. Formulace uvedené rovnice 2.13 neplatí pro pouze jedno pozorování, ale pro množství vektorů parametrů $\mathbf{V}$:

$$\mathbf{U} = \mathbf{T}^{\mathrm{T}} \cdot \mathbf{V}. \tag{2.14}$$

Transformační matice je čtvercová matice s rozměry $p \times p$. Je složená z vektorů definujících lineární kombinaci vstupních parametrů:

$$\mathbf{T} = \begin{bmatrix} \mathbf{c}_0 & \mathbf{c}_1 & \dots & \mathbf{c}_{p-1} \end{bmatrix}. \tag{2.15}$$

Transformační matice $\mathbf{T}$ má následující hlavní vlastnosti:

- vektory $\mathbf{c}_i$ jsou ve směru největší odchylky v datech a odchylka je koncentrovaná do co nejnižšího počtu výstupních komponent.
- vektory $\mathbf{c}_i$ jsou vzájemně ortogonální [23]:

$$\mathbf{c}_i^{\mathrm{T}} \cdot \mathbf{c}_j = 0 \quad \forall\, i \neq j. \tag{2.16}$$

- transformace je invertovatelná:

$$\mathbf{v}(n) = \mathbf{T}^{\mathrm{T}} \cdot \mathbf{u}(n). \tag{2.17}$$

2.4.2 Standardizace dat

Standardizace je proces, při kterém se proměnné souboru dat převádějí na stejné měřítko a tedy přestává záležet na skutečném rozměru příslušných proměnných. Jak bude uvedeno dále v práci, má aplikace tohoto procesu velmi pozitivní vliv na výsledky predikce regresní metody podpůrných vektorů. Protože existují různé metody

standardizace dat, bude blíže rozebrána metoda, která je využívána ve funkci **fitrsvm**, která trénuje regresní predikční model SVM a je používána v této práci. Zmíněná metoda funguje na principu centrování a úpravy rozpětí pomocí váženého průměru a vážené směrodatné odchylky. Standardizace prediktoru $j(x_j)$ je tedy dána:

$$x_j^* = \frac{x_j - \bar{x}_j}{\sigma_j}, \tag{2.18}$$

kde $\bar{x}_j$ značí vážený průměr ve sloupcích (jednotlivých parametrů), σ_j je vážená směrodatná odchylka a j označuje jednotlivá pozorování. [35].

2.4.3 Metoda RReliefF

Metoda RReliefF (*Regressional ReliefF*) je pro regresi upravená metoda ze skupiny metod *Relief*, které jsou schopné detekovat podmínkové závislosti mezi atributy a tím umožnit lepší pohled na určování parametrů při klasifikaci a regresi. Zatímco dříve byly tyto metody vnímány hlavně jako nástroj aplikovaný pro předzpracování parametrů před samotným procesem strojového učení za účelem výběru podmnožin parametrů, dnes se využívá i ve více oblastech [14].

Klíčová myšlenka původního Relief algoritmu bylo vyhodnotit kvalitu parametrů v závislosti na tom, jak moc jsou jejich hodnoty rozdílné mezi případy ležícími blízko sebe. Algoritmus funguje na principu penalizace prediktorů, které dávají rozdílné hodnoty sousedům se stejnými hodnotami, zatímco přidává skóre prediktorům, které dávají rozdílné hodnoty sousedům s rozdílnou hodnotou. Poté jsou vypočítané váhy jednotlivých prediktorů. Celý algoritmus, jak původního Relief tak upravené metody *ReliefF*, lze nalézt v původním článku [17]. Tato modifikace ReliefF už není limitována na dvě třídy problému a je více robustní. Popis a algoritmus této pro regresi sofistikovaně upravené metody je prezentován v dostupném článku [19]. Metoda typu Relief je implementována do prostředí Matlab pod názvem `relieff` [19].

3 Databáze DEAM

MediaEval Database of for Emotional Analysis in Music je souhrnná databáze pro potřeby MER skládající se z menších celků, které vznikaly pro potřeby účastníků jednotlivých ročníků soutěže iniciativy MediaEval[1]. Tato iniciativa se zabývá měřením, analýzou a vyhodnocováním nových algoritmů pro zpracování a získávání informací multimediálního charakteru. Hlavním záměrem této iniciativy v oblasti MER bylo vyhodnocování emocí, které jsou při poslechu hudby proměnné v čase tak, jak se mění samotný obsah hudebních skladeb. Tyto algoritmy dynamického MER se označují jako *MEVD (music emotion variation detection)*.

Mezi léty 2013 a 2016 bylo jedním z témat soutěže MediaEval Challenge právě vyhodnocování MER, pro jejíž potřeby vznikala tato databáze obsahující jak zvukové vzorky a emoční anotace, tak již extrahované hudební parametry skladeb. Tyto tři součásti budou níže popsány.

3.1 Hudební nahrávky

- 58 nahrávek plné délky a 1740 úryvků délky 45 s
- freemusicarchive.org, jamendo.com a medleyDB dataset [7]
- bezplatné licence (royalty-free),
- různorodé žánrové zastoupení - pop, rock, electronic, country, jazz, atd...
- formát MPEG layer 3 (MP3), f_{vz}=44,1 kHz
- v případě 45 s úryvků byl začátek určen náhodně (rovnoměrné rozdělení pravděpodobnosti)

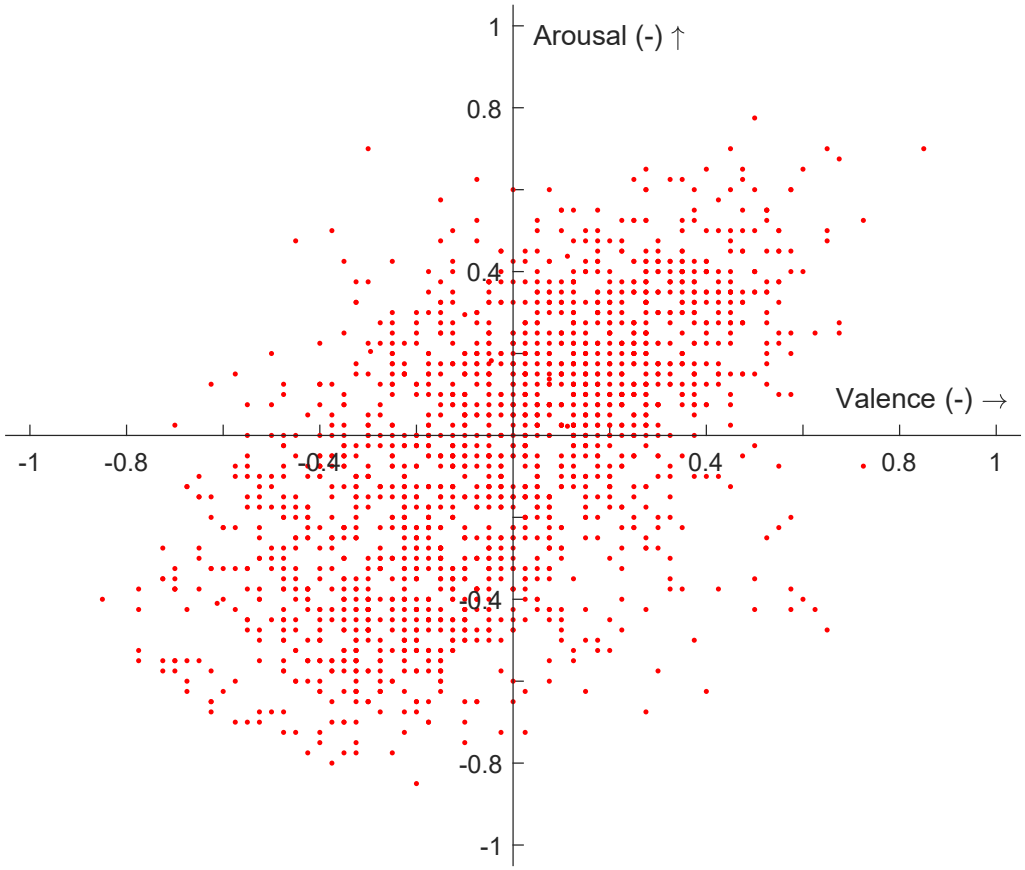
3.2 Anotace emocí

Důležitou součástí databáze jsou také informace o emocích nahrávek, které byly získávány jak označováním samotnými výzkumníky, tak hlavně s využitím tzv. crowdsourcingové platformy *Amazon Mechanical Turk*[2]. Každá zvuková ukázka byla hodnocena 7 až 23 subjekty – posluchači. Tento numerický typ označování byl zaznamenáván v čase s frekvencí záznamu hodnot 2 Hz a v klasickém Thayer-Russelově AV systému (viz. 1.2.2), přičemž rozsah hodnot pro toto časově proměnné hodnocení variuje v hodnotách mezi -1 a 1. Zaznamenána byla také statická hodnota vyjadřující celkové emoce skladby, tedy neproměnné v čase. Toto hodnocení bylo provedeno v rozsahu hodnot 1 až 9. Je třeba zdůraznit, že u všech záznamů bylo ignorováno prvních 15 s [2, 1].

[1] MediaEval Benchmarking Initiative for Multimedia Evaluation, www.multimediaeval.org

[2] www.mturk.com

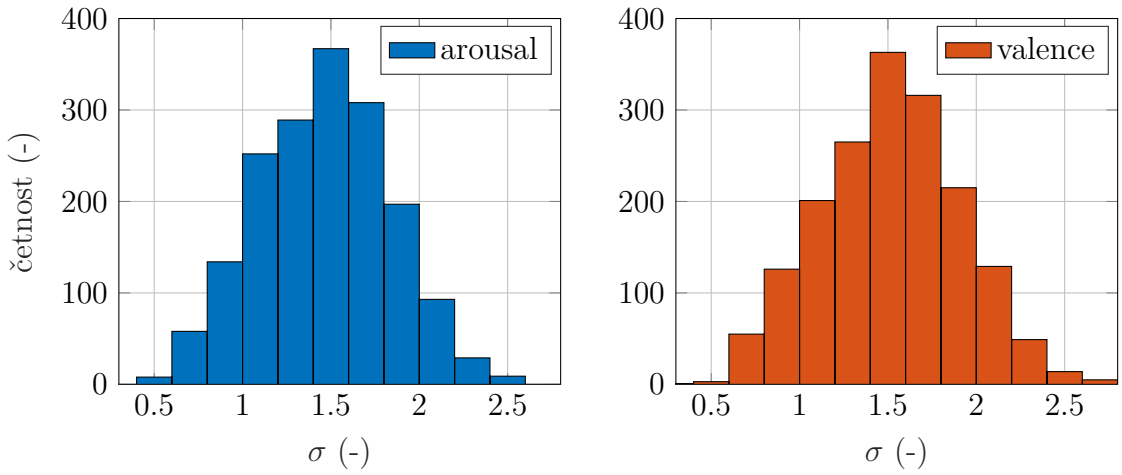
Zastoupení statických hodnocení emocí, která budou využita v této práci, můžeme vidět na obr. 3.1.



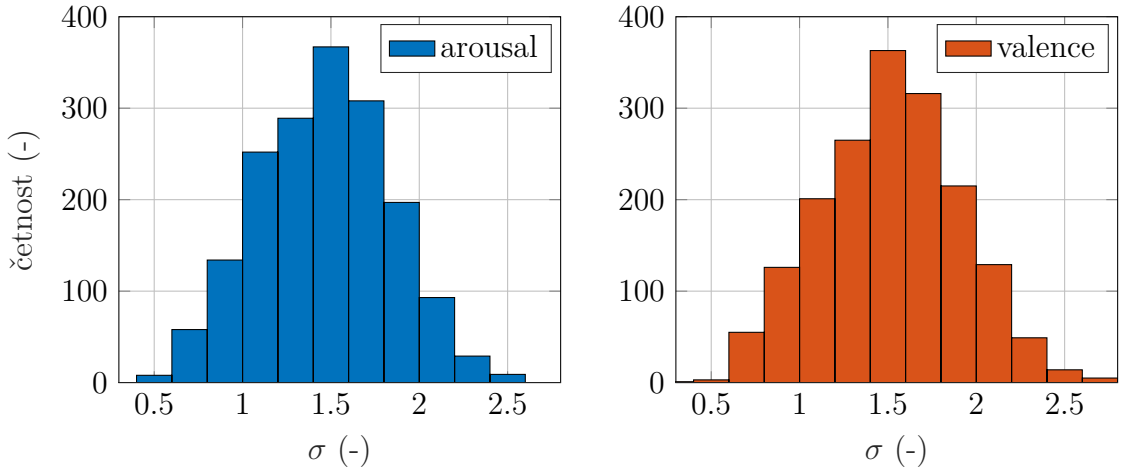
Obr. 3.1: Zastoupení emocionálního hodnocení skladeb databáze DEAM v Russell-Thayerově prostoru. Původní hodnoty jsou normalizované do rozsahu [-1,1].

Jak již bylo zmíněno, každá skladba byla hodnocena větším počtem hodnotitelů, přičemž pro tuto práci bude použita střední hodnota – aritmetický průměr vypočítaný ze všech hodnocení dané skladby. Zastoupení těchto hodnot pro dané skladby lze vidět v histogramu středních hodnot na obr. 3.2. Nejčastější hodnota v obou dimenzích se přirozeně pohybovala ve středu rozsahu, což zhruba odpovídá předpokládanému normálnímu rozdělení náhodné veličiny.

Zastoupení hodnot směrodatných odchylek v množině anotací všech skladeb přibližně odpovídá normálnímu rozdělení a můžeme jej vidět na obr. 3.3. Průměrná směrodatná odchylka anotace pro rozměr arousal je $\sigma_a = 1{,}46$ a pro valence $\sigma_a = 1{,}52$, což odpovídá v celkovém rozsahu hodnot $[1, 9]$ 18,5 % pro rozměr arousal a 19 % pro valence. Nejnižší průměrná hodnota anotace je 1,6 a nejvyšší 8,1.



Obr. 3.2: Histogram středních hodnot anotací jednotlivých skladeb na celkovém hodnocení v rozsahu 1 až 9.



Obr. 3.3: Histogram směrodatných odchylek pro anotace jednotlivých skladeb na celkovém hodnocení v rozsahu 1 až 9.

4 Strojové učení

4.1 Metoda podpůrných vektorů (SVM)

SVR (Support Vector Machines) je algoritmus sloužící pro klasifikaci a regresní analýzu, kterou představil Vladimir Vapnik a jeho kolegové v 1995, ačkoliv první zmínka o tomto algoritmu pochází již z roku 1979.

Princip této metody strojového učení s učitelem je založen na rozdělení dat, reprezentovaných vektory v mnoha-dimenzionálním příznakovém prostoru, do dvou skupin (tříd) s pomocí lineárního klasifikátoru – roviny. Jinými slovy se tato metoda snaží nalézt nadrovinu, která rozděluje prostor příznaků na dva poloprostory tak, že data náležející odlišným třídám leží v opačných poloprostorech. Cílem je oddělit tyto odlišné třídy dat na co největší vzdálenost. Protože se nestává často, že by vstupní data byla lineárně separovatelná, SVM využívá tzv. *jádrové transformace* (angl. *kernel transformation*), která umožňuje převést původně lineárně neseparovatelnou úlohu na úlohu lineárně separovatelnou, na kterou lze dále aplikovat optimalizační algoritmus pro nalezení rozdělující roviny. [8]

Řešení tohoto problému se provádí převodem daného vstupního prostoru do jiného více-dimenzionálního prostoru, kde se již i jinak lineárně nerozdělitelné třídy dají separovat lineárně. Jinými slovy probíhá zobrazení trénovacích dat z původního prostoru do jiného euklidovského prostoru $\Phi : \mathbf{R}^d \mapsto \mathcal{H}$

4.2 SVM regrese

Regresní typ metody podpůrných vektorů Support Vector Regression (SVR) je velmi používaná metoda skupiny SVM. Nejde v principu o klasifikaci vstupních dat do určitých kategorií, ale o predikci reálných číselných hodnot. Mezi výhody regresní SVM oproti jiným metoda patří jak odolnost vůči vzdáleným pozorováním, tak poměrně rychlý výpočet předpovědí.

4.2.1 Princip SVR

Mějme množinu trénovacích dat v podobě dvojic:

$$\{(\vec{x}_i, y_i)...(\vec{x}_l, y_l)\} \subset \mathcal{H} \times \mathbf{R}^d, \tag{4.1}$$

kde $\mathcal{H}$ značí prostor vstupních dat (například $\mathcal{H} = \mathbf{R}^d$), $\vec{x}_i$ reprezentuje vstupní vektor hodnot a y_i vyjadřuje informaci od učitele. V případě regrese se jedná o reálnou

číselnou hodnotu, zatímco v případě klasifikace o hodnoty +1 nebo -1. Nadrovina rozdělující body do dvou tříd lze popsat rovnicí [34]:

$$\vec{\omega} \cdot \vec{x} + b = 0, \tag{4.2}$$

kde $\vec{\omega}$ je normála nadroviny. V regresní SVR je tedy potřeba najít funkci $f(x)$, která má největší odchylku ε ze získaných hodnot y_i pro všechna trénovací data a zároveň byla co nejvíce plochá. Pro případ lineární funkce f platí:

$$f(x) = \langle \omega, x \rangle + b \qquad \omega \in \mathcal{H}, b \in \mathrm{R}, \tag{4.3}$$

kde $\langle ., . \rangle$ znamená skalární součin v $\mathcal{H}$. Plochost v rovnici 4.3 značí malá hodnota ω. Proto je potřeba minimalizovat její Euklidovskou normu $||\omega||^2$. Pro tento optimalizační problém se zavádí se také pomocné proměnné ξ_i a ξ_i^*. Nakonec lze tedy formulovat konvexní optimalizační problém:

$$\text{minimalizovat} \quad C\frac{1}{2}\|\omega\|^2 + \sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{4.4}$$

$$\text{za podmínek} \begin{cases} y_i - \langle \omega, x_i \rangle - b & \leq \varepsilon + \xi_i \\ \langle \omega, x_i \rangle + b - y_i & \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0 \end{cases}$$

Konstanta $C > 0$ vyjadřuje poměr – kompromis (angl. trade-off) mezi plochostí funkce f a maximální hodnotou, do které jsou odchylky větší než ε tolerovány. Regresní SVR využívá tzv. ε-necitlivou ztrátovou funkci $|\xi|_\varepsilon$, která je popsána takto: [31]

$$|\xi|_\varepsilon := \begin{cases} 0 & \text{pokud } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{v ost. případech} \end{cases} \tag{4.5}$$

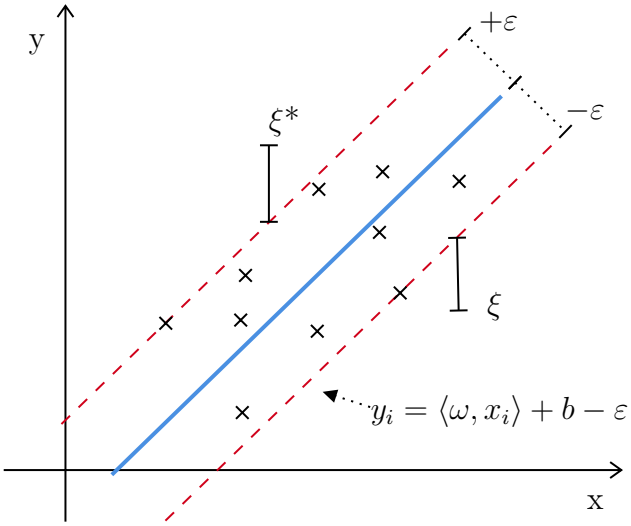
Graficky je tento popis znázorněn na obr. 4.1. Pokud je rozdíl mezi očekávanou a skutečnou hodnotou predikce menší, než je hodnota ε, není to považováno za chybu. Chceme tedy, aby:

$$-\varepsilon \leq \omega \cdot x_i - b - y_i \leq \varepsilon. \tag{4.6}$$

4.2.2 Jádrové funkce

Mezi často využívané jádrové funkce patří:

- lineární $K(x, y) = (x, y)$
- polynomiální s polynomem stupně p: $K(x, y) = (x \cdot y + 1)^p$
- radiální bázové funkce: $K(x, y) = e^{-\|x-y\|^2}$



Obr. 4.1: Rozdělující nadrovina a hraniční pásmo u lineární SVM. Body na okrajích pásma jsou podpůrné vektory [6].

4.2.3 Implementace

V této práci bude vypočítáváno a vyhodnocování více druhů modelů SVR s různými jádrovými funkcemi pro zajištění nejvhodnější konfigurace a pro zjištění nejvhodnějšího modelu pro daný typ tréningových dat. Trénovaní je provedeno v prostředí *Matlab* za využiti *Statistics and Machine Learning Toolbox*, který lze použít jak v grafickém uživatelském rozhraní, tak pomocí funkcí, např. **fitrsvm**, která slouží pro trénování regresního modelu. Pro potřeby této práce, tedy vypočítávání velkého počtu modelů, nebylo toto grafické rozhraní použito a všechny iterované výpočty byly provedeny pomoci vytvořených skriptů. Typy použitých regresních SVM metod jsou uvedeny v tabulce 7.2. V této tabulce můžeme vidět také flexibilitu jednotlivých metod, která je ovládána parametrem S_K - *kernel scale* (jádrové měřítko). Čím nižší S_K, tím je model flexibilnější. Možnost automatického zvolení tohoto parametrů znamená výběr S_K podle heuristického postupu.

Tab. 4.1: Přehled typů použitých SVR metod a jejich flexibility [35].

Typ SVM modelu	**Zkratka**	**Interpretace**	**Flexibilita a S_K**
lineární SVR	SVR_{lin}	snadná	nízká flexibilita
kvadratická SVM	SVM_{quad}	obtížná	střední flexibilita
kubická SVR	SVM_{cub}	obtížná	střední flexibilita
jemná Gaussova SVR *(Fine Gaussian SVR)*	SVR_{GaussF}	obtížná	vysoká flexibilita $S_K = \sqrt{4}/4$
střední Gaussova SVR *(Medium Gaussian SVR)*	SVR_{GaussM}	obtížná	střední flexibilita $S_K = \sqrt{4}$
hrubá Gaussova SVR *(Coarse Gaussian SVR)*	SVR_{GaussC}	obtížná	nízká flexibilita $S_K = 4\sqrt{4}$

S_K – jádrové měřítko (kernel scale)

5 Statistická analýza

Pro vyhodnocení úspěšnosti predikce regresního číselného predikčního systému je potřeba využít metod statistické analýzy, které přinesou přehledné relevantní výsledky. Nejběžněji používanou metodou, která je použita i v této práci při všech srovnáních, je statistika R^2 [39]. Přestože je v práci při všech výpočtech vyhodnocována větší skupina parametrů (uvedeny níže), v textu práce se objevuje především již zmíněná statistika R^2.

5.1 Koeficient determinace R^2

Standardní metrikou pro vyhodnocování regrese je statistika R^2, neboli *Coefficient of determination*, nazývaný též „R kvadrát". Jedná se o míru kvality regresního systému a v základní podobě vyjadřuje, jaký podíl rozptylu v pozorování závislé proměnné se podařilo regresi vysvětlit (vyšší hodnoty variability znamenají větší úspěšnost regrese). Hodnoty tohoto koeficientu nabývají hodnot mezi 0 a 1, s tím, že pokud $R^2 = 1$, jedná se o dokonalou predikci. Naopak hodnota $R^2 = 0$ znamená, že model nepřináší pro poznání závislé proměnné žádnou informaci a je tedy zcela neužitečný. Běžnou praxí je udávat tento koeficient v procentech, tedy $0 - 100\,\%$ [41].

Koeficient determinace bývá většinou definován jako jedna mínus podíl rozptylu chyb a rozptylu nezávislé proměnné. Jedná se tedy o poměr vysvětlené variability k celkové variabilitě proměnné Y. Definiční rovnice je dána takto [12]:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y}_i)^2} \tag{5.1}$$

kde SS_{res} je suma čtverců chyb, SS_{tot} suma kvadratických odchylek závislé proměnné y od její střední hodnoty $\bar{y}$ a $\hat{y}_i$ je regresní odhad i-tého pozorování. Pro potřeby této práce byl výpočet implementován v prostředí Matlab podle [26].

5.2 Střední kvadratická chyba – MSE

Střední kvadratická chyba (anglicky *mean squared error*, *MSE*) vyjadřuje přesnost odhadů pomoci střední hodnoty druhých mocnin rozdílu mezi měřenými a skutečnými hodnotami. V predikčním systému tedy vyjadřuje průměr kvadratické chyby mezi reálnou hodnotou a hodnotou predikovanou. Čím nižší je hodnota MSE, tím je předpověď hodnot přesnější. Platí tedy, že:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 , \tag{5.2}$$

kde n je počet pozorování, y_i je hodnota závislé proměnné i-tého pozorování a $\hat{y}_i$ je její odhad.

5.3 RMSE

Stejně tak, jako je směrodatná odchylka σ odmocninou rozptylu náhodné veličiny:

$$\sigma = \sqrt{var(X)} = \sqrt{E((X - E(X)^2)} \tag{5.3}$$

kde X je náhodná veličina, $var(X)$ rozptyl a $E(X)$ její střední hodnota, tak se analogicky používá odmocnina střední kvadratické chyby - **RMSE** (*root mean square error*). Tento typ měření se využívá velmi často při vyhodnocování rozdílu mezi hodnotami předpovězenými modelem a hodnotami reálnými. V oblasti MER se jedná o jeden z nejvíce vyskytovaných typů měření chyb predikce. Pomocí MSE se lze RMSE definovat jednoduše [3]:

$$RMSE(X) = \sqrt{MSE(X)}. \tag{5.4}$$

5.4 Střední absolutní chyba – MAE

Střední absolutní chyba (*mean absolute error, MAE*) je průměr rozdílu mezi měřenými (odhadovanými) a skutečnými hodnotami a vyjadřuje průměrnou absolutní velikost chyby odhadu:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| . \tag{5.5}$$

5.5 Spearmanův korelační koeficient pořadí

Tento koefient navrhnul Ch. E. Spearman tak, že koreloval postupem podle Pearsona pořadí jednotlivých měření obou proměnných. Tento koeficient zachycuje monotónní vztahy (ne pouze lineární, ale obecně rostoucí nebo klesající) a je rezistentní vůči odlehlým hodnotám. Jedná se o neparametrickou metodu, která je založena na výpočtu pořadí sledovaných veličin. Výhodou této metody je, že je možno ji použít pro popis jakékoliv závislosti - lineární i nelineární. Standardně se tato metoda využívá v takových případech, kdy nemůžeme očekávat linearitu očekávaného vztahu nebo normální rozdělení sledovaných proměnných. Spearmanův koeficient pořadové korelace je definován vztahem [12, 20]:

$$\rho_s = 1 - \frac{6 \sum_{i=1}^{n} (R_i - Q_i)^2}{n(n^2 - 1)}, \tag{5.6}$$

kde R_i označuje pořadí náhodné veličiny X_i a Q_i udává pořadí náhodné veličiny Y_i. Parametr n udává počet korelačních dvojic a koeficient ρ může nabývat hodnot $-1 \leq \rho \leq 1$.

Pro potřeby této práce je tento koeficient vypočítáván v prostředí Matlab pomoci již zabudované funkce **corr(X,Y,'type','Spearman')**.

5.6 Směrodatná odchylka

Směrodatná odchylka s je odmocnina z rozptylu a vrací míru rozptýlenosti do měřítka původních dat. Je definována jako:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \tag{5.7}$$

kde n je počet měření, x_i jsou naměřené hodnoty a $\bar{x}$ jejich aritmetický průměr [12].

6 Návrh systému MER

Pro správný výběr MER systému, který má vykazovat maximální možnou výkonnost, je potřeba přijmout mnoho dílčích rozhodnutí ve výběru mezi mnoha typy systémů, metod a vstupních dat. Za tímto účelem byla vypracována poměrně rozsáhlá teoretická část práce, která se pokouší tuto dynamickou vědní oblast zpřehlednit, strukturalizovat a jaksi přehledně rozčlenit. V praktické části bude postupně objasněn proces návrhu vyhodnocovacího systému.

6.1 Výběr typu systému podle ground-truth dat

Z metod MER systémů popsaných v kap. 1.3.1 vyplývá, že největší efektivitu mají systémy využívající k predikci kombinaci hudebních parametrů a ground-truth dat. Jako nejvhodnější typ reprezentace ground-truth dat na základě referencí literatury a kapitoly 1.3.2 se jeví typ číselný, který trpí nižší mírou subjektivity anotací. Jedná se tedy o využití dimenzionálního modelu emocí (který je popsán v kap. 1.2.2).

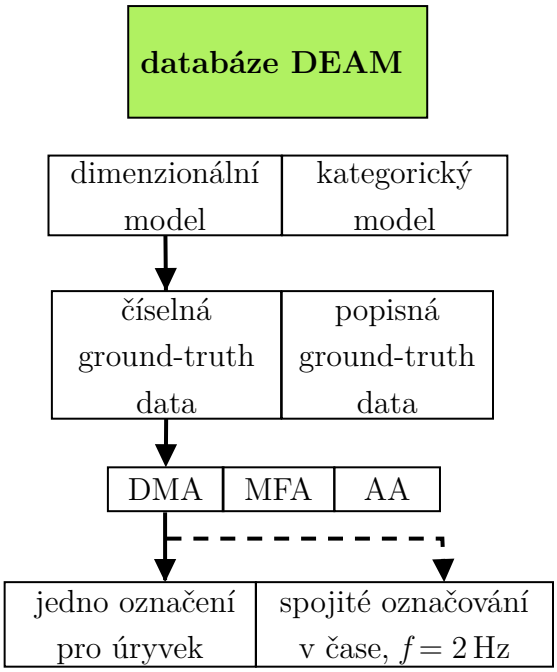
Pokud má být takový systém sestaven, je třeba vstupní ground-truth data získat. Protože není náplní této práce se zaměřovat na získávání emocionálních označení nahrávek, vytvářet pro tyto potřeby hudební databáze a anotační systémy, či dotazníky, je potřeba použít již existující vhodnou volně dostupnou databázi anotací (tzv. datasheet). Z veřejně dostupných databází, které by obsahovaly číselný typ popisu emocí a byly tedy založeny na dimenzionálním modelu emocí, je nejvhodnější databáze DEAM popsáná v kapitole 3. Tato databáze využívá metodu anotací DMA popsanou v kap. 1.2 a mezi jinými vyniká především svou obsáhlostí.

Na obrázku 6.1 můžeme vidět přehledné zařazení databáze a potažmo celého navrženého modelu MER do kategorií, které rozebírá teoretická část této práce. Jedná se tedy o databázi využívající dimenzionální model emocí s číselnými ground-truth daty v podobě statických (časově invariantních) anotací emocí.

6.2 Výběr hudebních parametrů

Proces výběru správných typů a počtu parametrů (deskriptorů) pro klasifikaci či regresi v MER bývá často těžkým úkolem. Jak je uvedeno v teoretické části 2, existují určité teoretické předpoklady o tom, které parametry mohou mít větší či menší vliv na vnímání emocí. Tyto předpoklady jsou však značně omezené a nedá se na ně při návrhu MER systému plně spolehnout [40]. Tuto skutečnost dokládá typ systémů založených pouze na hudebních parametrech popsaný v kapitole 1.3.1).

Trendem výzkumů v oblasti MER je snaha o vytvoření co možná největšího souboru dat (datasetu) všech extrahovaných hudebních parametrů, které je možno



Obr. 6.1: Systémová kategorizace databáze DEAM.

vypočítat a které by mohly mít potenciální vliv na vnímání hudebních emocí (např. [14] či většina soutěžních výzkumů MediaEval 2013 – 2015 viz. [1]). Předpoklad, že zvýšením počtu parametrů se automaticky zvýší úspěšnost systému nemusí být však vždy pravdivý. Experimenty ukázaly, že přestože se výkonnost systému může s nárůstem počtu parametrů zvýšit, použití přílišného počtu systém naopak degraduje [39]. Proto bude nutné otestovat systém pro různé počty parametrů a zjistit tak konfiguraci nabízející nejlepší výsledky predikce. Zároveň je potřeba vybrat z velkého výčtu parametrů pomocí *metod výběru parametrů* (*Feature Selection Methods*, FSA) ty parametry, které mají největší vliv na úspěšnost predikce a vyřadit parametry, které nemají na predikci vliv, či ji dokonce zhoršují [40].

6.3 Extrakce hudebních parametrů

Jak již bylo uvedeno v teoretické části práce (viz. 2.1), pro extrakci parametrů byly vybrány dva dostupné nástroje s poměrně rozdílným přístupem. Prvním nástrojem je MIRtoolbox, vytvořený v prostředí MATLAB, který byl navržen pro potřeby oblasti MIR, která se získáváním informací z hudby přímo zabývá. Tento nástroj je tedy pro rozpoznávání emocí v hudbě přímo určen. Druhým nástrojem je openSMILE společnosti audEERRING, který nabízí možnost extrakce velkého množství parametrů, které však již poněkud více spadají do oblasti analýzy řečového signálu,

tedy o řečové parametry (příznaky), viz. 2.1.2.

6.3.1 Předzpracování dat a redukce dimenzionality

Studie [1] na vlastním datasetu ukázala, že proces standardizace (*standardization*) zvyšuje efektivitu MER. Například strojové učení metodou SVR mělo se standardizací nejlepší výsledky ze všech metod. Také se ukázalo, že tento proces nemá vliv na další použité algoritmy. Autor uvádí, že je pro SVR prakticky nutností a zároveň ale negativně neovlivňuje výsledky jiných regresních metod. Standardizace tedy bude automaticky zařazena do předzpracování dat pro výpočet SVR.

Také je vhodné do výpočetního řetězce zahrnout algoritmus PCA (viz. 2.4.1). Metody, které redukují počet parametrů a sníží dimensionalitu systému a jeho výpočetní náročnost se označují souhrnně FSA (Feature selection algorithms). Tyto algoritmy jsou popsány v 2.4.

6.3.2 Metoda strojového učení

Pro tuto práci bylo navrženo použití regresního typu metody podpůrných vektorů (SVR, viz. kap. 4.2) jako hlavní metody, jejíž výsledky budou prozkoumány a vyhodnoceny. Zmíněná metoda byla vybrána nejen z důvodů vysoké úspěšnosti ve vyhodnocení posledních ročníků soutěže iniciativy MediaEval s názvem „*Emotion in Music*“ [1], tak i díky tomu, že z provedené rešerše již existujících prací vyplývá, že je tato metoda jedna z nejpoužívanějších a také přináší stabilně dobré výsledky. Porovnání výsledků soutěže MediaEval je příhodné nejen nejen kvůli jednotné metodologii vyhodnocení úspěšnosti velkého množství týmů, ale hlavně díky velmi podobnému setu vstupních dat a anotací, které byly používány (dílčí části dnešní databáze DEAM).

V návrhu vyhodnocovacího systému se počítá s výpočtem regresního modelu pomocí strojového učení jak zvlášť pro rozměr valence, tak i rozměr arousal.

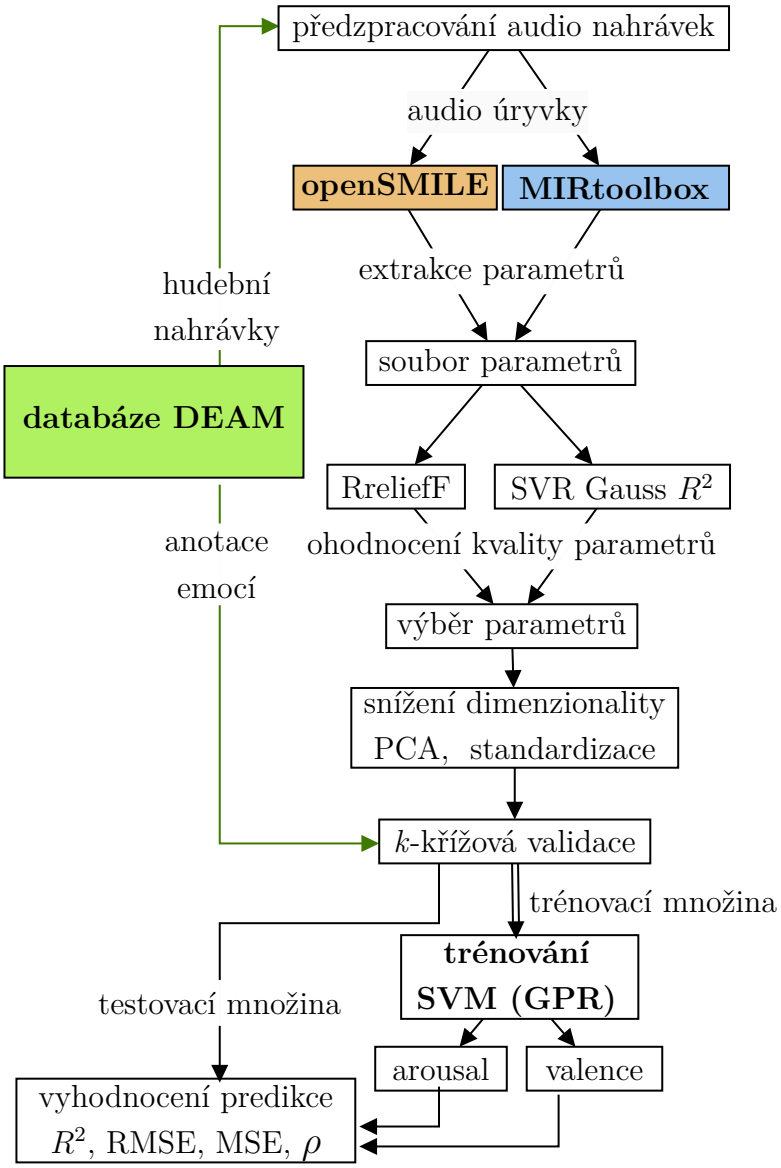
6.4 Návrh vyhodnocovacího systému MER

Na obrázku 6.2 se nachází blokové schéma celkového návrhu MER systému. Jak již bylo zmíněno, kategorické zařazení systému se převážně odvíjí od použité databáze – její kategorické zařazení je na obr. 6.1. Pro tuto práci byl vybrán nejběžnější typ MER modelu, který využívá jako vstupní data (ground-truth) spolu s extrahovanými parametry nahrávek také subjektivní anotace emocí. Potřebná data poskytuje databáze DEAM, která obsahuje 1740 hudebních úryvků délky 45 s ve formátu MPEG-3.

Ve fázi předzpracování audio nahrávek bylo potřeba nahrávky v původní formátu MPEG-3 převést na formát PCM souborů typu .wav. Pro extrakci parametrů spolu s jejich statistickými hodnotami byl použit MIRtoolbox a externí nástroj openSMILE. Analyzovány budou tedy tři soubory parametrů, jak popisuje tabulka 7.2.

První použitá metoda pro výběr parametrů je RreliefF, jejíž výstupem je číselné ohodnocení kvality každého parametru. Pokud provedeme seřazení od nejvýznamnějšího, máme k dispozici datový set počínaje nejvýznamnějším parametrem konče parametry nechtěnými. Ten je poté podroben analýze. Druhou metodou pro individuální ohodnocení kvality parametrů byla zvolena metoda podpůrných vektorů (SVR) s Gaussovou jádrovou funkci. Pro každý parametr byl tedy natrénován model SVR s 20násobnou křížovou validací, poté bylo provedeno vyhodnocení pomocí nástrojů statistické analýzy a seřazení podle významnosti jako u předchozí metody.

Vybrané datové sady budou poté rozděleny a vyhodnoceny pomocí k-násobné křížové validace. Trénování modelu proběhne jak pomocí různých typů metody SVR, tak pomocí metody GPP (*gaussian process regression*) a to vždy zvlášť pro rozměr arousal i valence.



Obr. 6.2: Blokové schéma návrhu vyhodnocovacího MER regresního systému.

7 Vyhodnocení systému MER

V první části této kapitoly se nachází analýza a vyhodnocení metody výběru hudebních parametrů RReliefF. Tato metoda slouží k ohodnocení významnosti parametrů a výstupem je vektor v významnosti jednotlivých parametrů, pomocí kterého budou parametry seřazeny od nevýznamovějšího. Další část kapitoly se zaměřuje na již zmíněné individuální hodnocení parametrů a to zvlášť pro oba soubory parametrů, které jsou v práci použity. V poslední části kapitoly je již komplexní vyhodnocení úspěšnosti navrženého MER systému.

7.1 Hodnocení metody RReliefF

7.1.1 Optimální hodnota k_R

Pro zjištění optimální hodnoty k-nejbližších sousedů metody RReliefF (viz. popis metody v kap. 2.4.3, označeno k_R) byl proveden výpočet tohoto algoritmu pro 11 hodnot k_R v rozsahu $k_{R1} = 5$ až $k_{R11} = 500$. Pro každé k_R bylo tedy vypočteno pořadí jednotlivých parametrů sestupně podle jejich kvality a tedy důležitosti pro úspěšnou predikci. Kalkulace proběhla zvlášť pro dimenzi arousal i valence a byl použit datový soubor openSMILE. Je tedy připraveno 11×2 datových sad určených ke strojovému učení a vyhodnocení.

Druhým krokem bylo trénování a vyhodnocení úspěšnosti predikce jednotlivých získaných datových sad a to s různým počtem použitých parametrů. Tím byla získána informace o úspěšnosti predikce v závislosti na hodnotě k_R a zároveň na počtu použitých parametrů vybraných podle důležitosti určené metodou RReliefF. Pro natrénovaní modelu byla v tomto případě použita jak lineární metoda SVR, tak zástupce skupiny nelineárních metod a tedy SVR s Gaussovou jádrovou funkcí. Pro vyhodnocení byla zvolena metoda křížové 20 násobné validace.

Parametry analýzy:

- výpočet RReliefF pro arousal a valence
  - pro 11 hodnot k_R v intervalu $\langle 5, 500 \rangle$
  - získáno celkem 22 datových sad v pořadí od nejdůležitějšího
- vyhodnocení úspěšnosti predikce každé datové sady
  - pro 27 hodnot počtu parametrů
    - rozsah hodnot parametrů v intervalu $\langle 1, 1579 \rangle$
  - pomocí lineární a Gaussovy SVR
    - 20 násobná k-křížová validace
    - vyhodnocení pomocí statistické analýzy, především R^2

7.1.2 Vyhodnocení

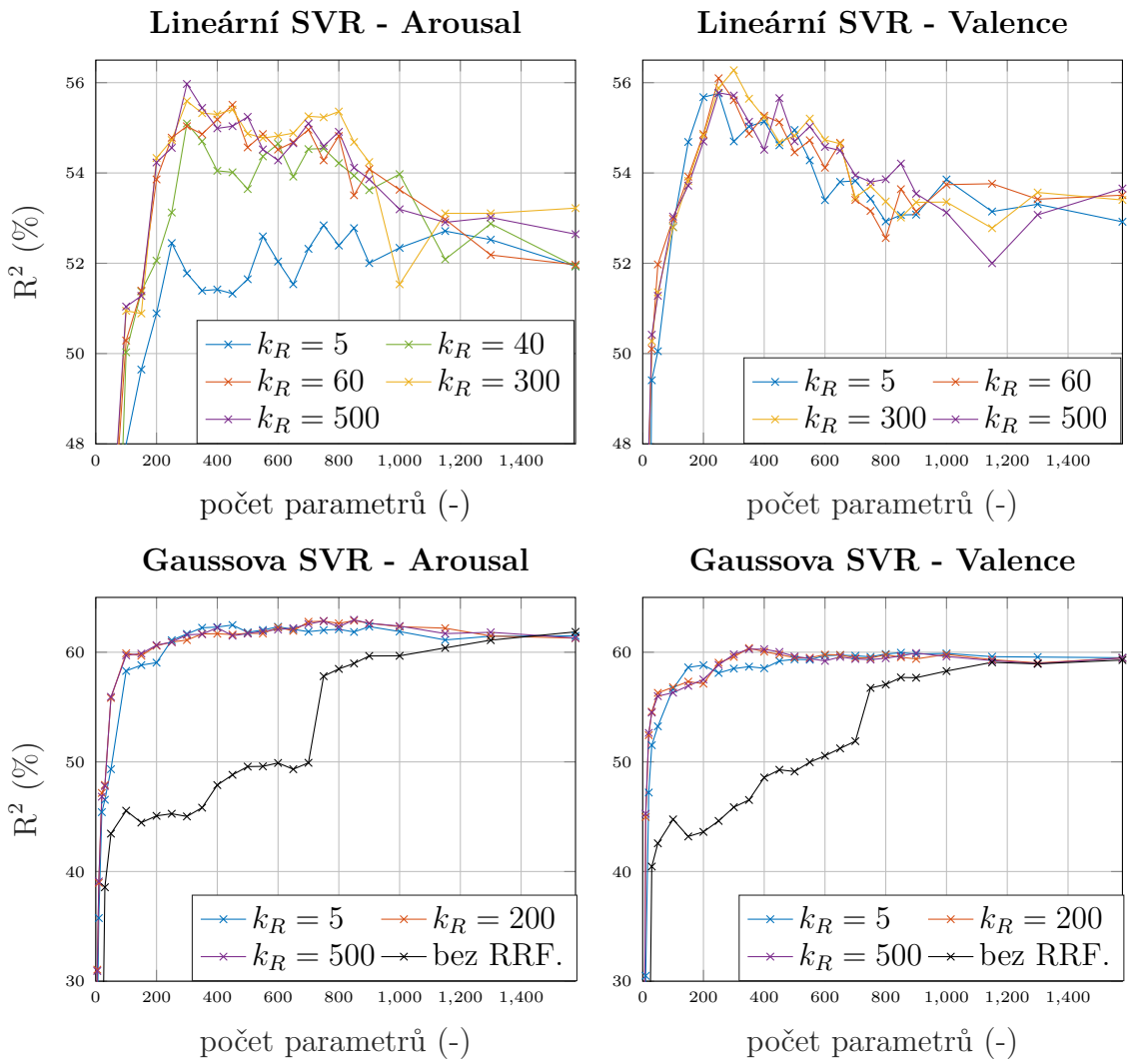
Na obrázku 7.1 můžeme vidět souhrnné grafické zobrazení provedeného vyhodnocení metody RReliefF pomocí úspěšnosti predikce hodnot arousal a valence a to pro různé hodnoty k_R v závislosti na počtu použitých parametrů pro trénování. Kvalita predikce je reprezentována hodnotami R^2. Přestože bylo testování provedeno pro mnoho hodnot k_R, pro přehlednost byl zobrazen pouze reprezentativní vzorek.

Již na první pohled je patrné, že lineární SVR nedosahuje takových výsledků jako Gaussova SVR, což není díky povaze těchto jednotlivých metod překvapivé. U lineární SVR nedochází k transformaci příznakového prostoru, jako u metod nelineárních (viz. kap. 4.2). Také je patrné, že má tato metoda větší výkyvy v úspěšnosti pro různý počet parametrů zahrnutých do trénovaní, což lze interpretovat jako menší „odolnost“ vůči neužitečným parametrům. Právě pro tuto metodu je stanovení ideálního k_R zásadní, což je cílem tohoto vyhodnocení – analyzovat vliv k_R na kvalitu predikce a stanovit její ideální hodnotu.

V tabulce 7.1 jsou uvedeny maximální a střední hodnoty úspěšnosti regrese pro jednotlivé hodnoty k_R. Lze si všimnout mírného nárůstu přesnosti predikce se stoupajícím k_R a to u obou metod v obou rozměrech. Na základě analýzy lze tedy potvrdit, že má vyšší hodnota k_R kladný vliv na úspěšnost predikce. Přestože rozdíly nejsou příliš velké, nelze říct, že by byly zanedbatelné, obzvláště v případě vyhodnocení rozměru arousal pomocí lineární SVR, jak lze pozorovat na obrázku 7.2. Zatímco optimální určení k_R znamená 1 % navíc (v R^2 statistice) u Gaussovy SVR, v případě lineární SVR se pro toto vyhodnocení jedná o přínos více než 3 %. Pro další výpočty budou tedy využity hodnoty $k_R > 150$, kdy již nedochází k významnému nárůstu úspěšnosti. Zároveň je ale potřeba poznamenat, že podle výsledků viditelných na obrázku 7.2 stále hraje roli i počet použitých parametrů a tak nelze explicitně konstatovat, která hodnota k_R je obecně nejvhodnější. Stále je třeba pokaždé experimentálně vyhodnotit a vybrat variantu s nejlepšími výsledky predikce.

7.1.3 Přínos metody

Přestože bylo toto měření provedeno se záměrem zjištění vhodného k_R, lze na výsledcích demonstrovat i přínos této metody. Přínos metody RReliefF ve smyslu zvýšení přesnosti predikce emocí díky výběru nejvhodnějších parametrů datasetu lze nejlépe demonstrovat na obrázku 7.2, kde vidíme, že při trénování celého datasetu (poslední hodnoty grafu) je úspěšnost výrazně nižší, než při výběru pouze těch nejvhodnějších parametrů (nejlepší výsledky při ≈ 300 parametrech). Přínos metody je nezanedbatelný i pro nelineární Gaussovu SVR (viz. přehled 7.1). Také lze sledovat výrazně nižší úspěšnost pro soubor dat, který nebyl vyhodnocen a seřazen pomocí zkoumané

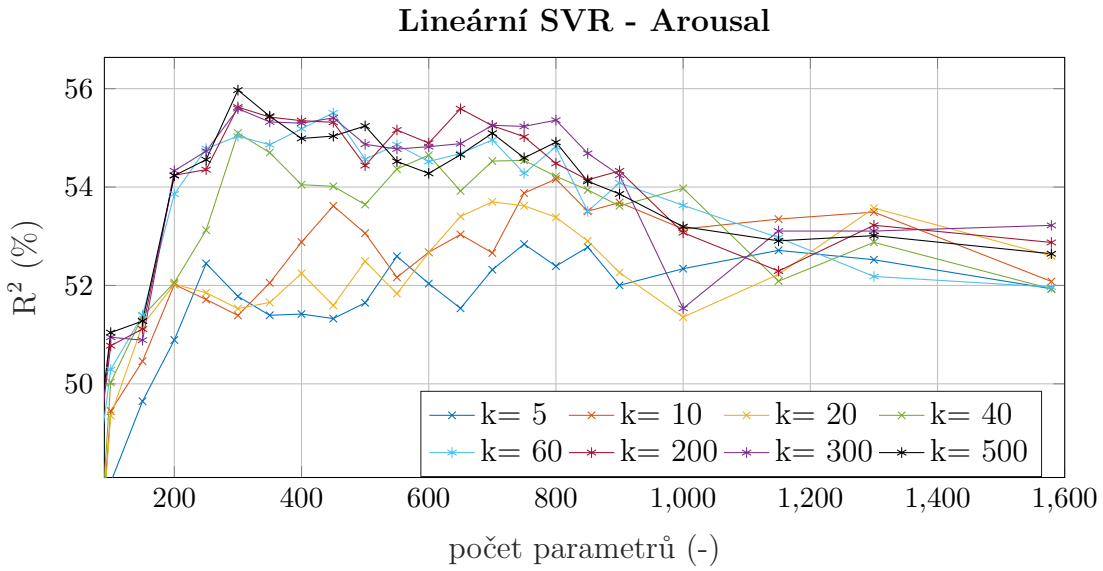


Obr. 7.1: Vyhodnocení metody RReliefF pomocí úspěšnosti predikce pro různé hodnoty k_R v závislosti na počtu použitých parametrů pro trénování.

metody. V grafech pozorujeme, že zatímco bez použití RreliefF počet kvalitních parametrů v datasetu pomalu roste, s použítím metody se kvalitní parametry nacházejí na více předních pozicích a úspěšnost je vyšší i s menším počtem parametrů.

Tab. 7.1: Tabulka průměrů a maximálních hodnot výsledků R^2 (%) statistické úspěšnosti trénování pro různá k_R.

	Lineární SVR				Gaussova SVR			
	arousal		valence		arousal		valence	
k_R	max	mean	max	mean	max	mean	max	mean
5	52,84	46,08	55,76	49,11	63,33	53,15	60,82	53,45
10	54,16	46,68	55,95	50,03	63,86	53,48	60,56	53,93
20	53,7	46,27	56,03	49,32	63,6	53,21	60,59	53,7
30	54,7	47,14	55,67	49,36	63,81	53,53	60,65	53,61
40	55,1	47,35	56,17	49,76	63,92	53,52	60,43	53,99
60	55,51	47,73	56,1	49,86	**64,28**	53,96	60,62	53,98
100	55,64	**48,34**	56,22	49,91	63,99	54,29	**61,08**	54,1
150	55,64	48,01	56,02	**49,95**	64,12	54,36	61,1	**54,1**
200	55,63	48,23	56,27	49,88	64,01	**54,36**	60,99	54,1
300	55,59	48,29	**56,28**	49,91	64,17	54,33	61,02	54,09
500	**55,97**	48,19	55,78	49,91	64,18	54,27	60,87	54,06



Obr. 7.2: Detail vyhodnocení metody RReliefF pomocí úspěšnosti predikce pro různé hodnoty k_R v závislosti na počtu použitých parametrů pro trénování.

7.2 Individuální analýza parametrů

Bezesporu velmi hodnotnou informací pro návrh MER systému je kvalita jednotlivých parametrů, přesněji řečeno jejich vliv na úspěšnou predikci. Pro individuální analýzu všech parametrů, tedy parametrů z MIRtoolboxu i openSMILE, bude v této práci využita jak již zmíněná metoda RReliefF, tak i přímé statistické vyhodnocení modelů, které jsou vždy natrénovány pomocí jednoho daného parametru. Pro tuto metodu bude použita střední Gaussova SVM (viz. 7.2), která je výpočetně méně náročnější a vykazuje dobré výsledky. Tato metoda bude dále v práci uváděna pod zkratkou SVR_{GaussM} Pro vyhodnocení úspěšnosti této predikce bude využita standardní statistika R^2. Pro hodnocení kvality parametrů pomocí metody RReliefF je použit parametr w (weight, váha), viz kap. 2.4.3.

Tab. 7.2: Přehled testovaných sad parametrů

Zdroj sady parametrů	**Zkratka**	**Počet parametrů**
openSMILE	D_{OS}	1579
MIRtoolbox	D_M	334
openSMILE + MIRtoolbox	D_{MOS}	1913

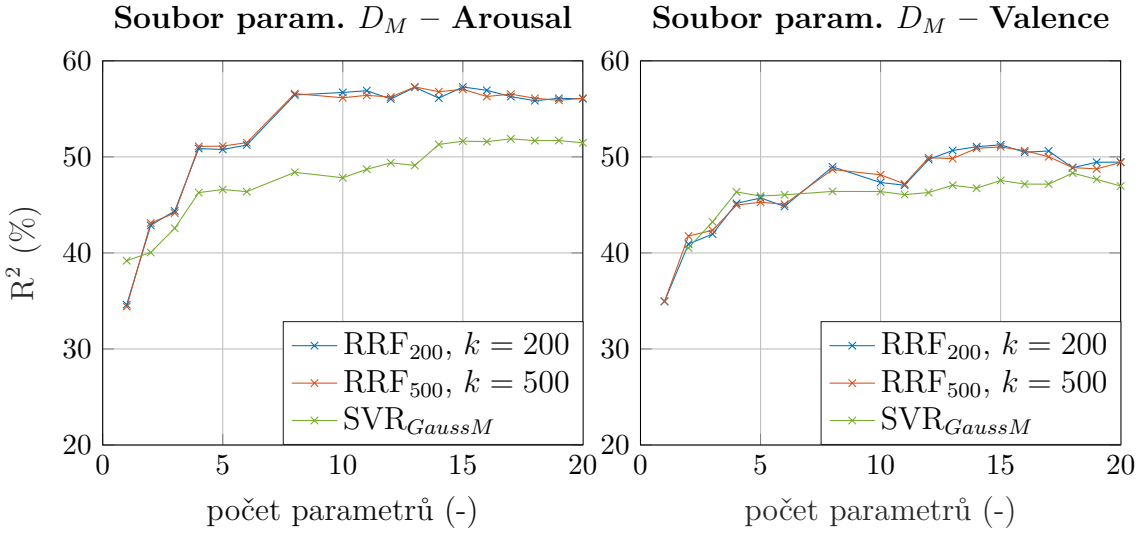
Z předchozí kapitoly o hodnocení této metody 7.1 bylo zjištěno ideální rozpětí hodnoty k_R. Pro individuální analýzu byly tedy zvoleny hodnoty $k_R = 100$, 200, 300 a 500. Protože bylo zjištěno, že jsou výsledky zvolených hodnot k_R v tomto případě prakticky totožně, budou ve srovnání zobrazeny pouze výsledky pro hodnoty $k_R = 200$. Metoda RReliefF s takto nastavenou hodnotou bude dále v práci označována jako RRF_{200}

Parametry analýzy:

- výpočet pořadí parametrů podle významnosti
  - RReliefF s $k_R = [100\ 200\ 300\ 500]$
    - pro všechny soubory parametrů a oba rozměry
  - SVR_{GaussM}
    - pro všechny soubory parametrů a oba rozměry
    - vyhodnocení pořadí pomocí R^2
    - 20 násobná k-křížová validace
- vyhodnocení úspěšnosti predikce všech kombinací souborů parametrů
  - 20 násobná k-křížová validace
  - vyhodnocení pomocí statistické analýzy, především R^2

7.2.1 Nejvýznamější parametry MIRtoolboxu

Na grafu 7.3 můžeme vidět výsledky predikce při výběru určitého počtu parametrů seřazených podle významnosti pomocí daných metod ohodnocení kvality parametrů. Z detailní analýzy dvaceti nejdůležitějších parametrů vyplývá, že pro arousal má metoda RRF_{200} lehce lepší výsledky, než vyhodnocení pomocí SVR_{GaussM}. Pro valence jsou rozdíly výsledků prakticky zanedbatelné. Pravdou však je, že obě metody vybraly povětšinou velmi podobnou množinu parametrů, pouze jejich pořadí na prvních příčkách se mění. Podrobně lze vyhodnocení kvality parametrů sledovat v tabulce A.1 v příloze této práce. V přehledové tabulce 7.3 je uvedeno pořadí parametrů vyhodnocených pomocí RRF_{200}, protože tato metoda vykazuje pro danou sadu parametrů lehce lepší výsledky, než metoda SVR_{GaussM}.



Obr. 7.3: Vyhodnocení úspěšnosti metod výběru parametrů na sadě parametrů D_M.

Jednoznačně nejvýznamnějším parametrem s nejlepšími výsledky predikce emocí u testovaných úryvků hudebních skladeb je pro oba rozměry emočního prostoru střední hodnota parametru spektrální fluktuace a to jak v podobě výpočtu pro celé spektrum, tak pro jednotlivá pásma oktávových filtrů. Jak je uvedeno v teoretické části v kap. 2.2.2, spektrální fluktuace udává změny ve výkonovém spektru mezi po sobě jdoucími rámci. Jako nejvýznamnější vychází parametr 7. oktávového filtru, který odpovídá frekvenčnímu rozsahu okolo 2 kHz – 3,5 kHz (viz obr. 2.1) a který společně s parametrem 8. filtru podle [22] nejlépe odpovídá emoci „activity“, kterou lze chápat právě jako rozměr arousal (míra nabuzení, energie). Zmíněná informace je tedy i touto prací potvrzena.

Dalším významným parametrem datové sady extrahované pomocí MIRtoolboxu se zdá být střední hodnota jasnosti spektra a spektrální entropie.

Tab. 7.3: Nejvýznamnější parametry MIRtoolboxu podle RRF_{200}.

	Parametr sady D_M – Arousal	**Parametr sady D_M – Valence**
1	spektrální fluktuace (mean)	sp. fluktuace v okt. pásmech (07, mean)
2	sp. fluktuace v okt. pásmech (07, mean)	spektrální fluktuace (mean)
3	jasnost spektra (mean)	sp. fluktuace v okt. pásmech (06, mean)
4	fluktuace tempa (mean)	sp. fluktuace v okt. pásmech (08, mean)
5	sp. fluktuace v okt. pásmech (06, mean)	spektrální entropie (mean)
6	spektrální entropie (mean)	jasnost spektra (mean)
7	doba náběhu (slope)	doba náběhu (slope)
8	spektrální šikmost (mean)	tempo (std)
9	sp. fluktuace v okt. pásmech (03, mean)	melovské kepstrální k. (de, 02, perAmp)
10	sp. fluktuace v okt. pásmech (08, mean)	melovské kepstrální k. (de, 02, perFreq)

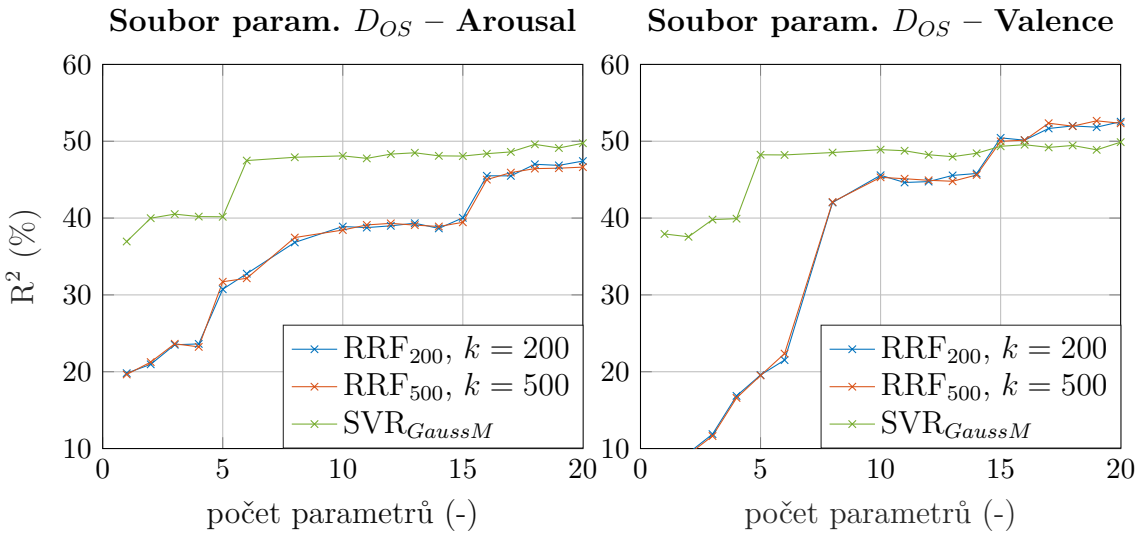
Zajímavým zjištěním také je, že čtvrtým nejvýznamnějším parametrem dimenze arousal je parametr střední hodnoty fluktuace tempa, čili rytmické periodicity, zatímco pro rozměr valence (míra libosti) hraje na předních příčkách roli směrodatná odchylka tempa. Při zkoumání definice těchto dvou parametrů lze najít určitou významovou podobnost. Významnost parametrů popisujících tempo skladeb je potvrzena i studií [41], která uvádí spojitost tempa s dimenzí arousal. Lze se také domnívat, že změny v tempu, tedy v jakési rytmické pravidelnosti hudební skladby, mají vliv i na vnímání příjemné–nepříjemné, čili dimenzi valence. Pro prokázání podrobnějších spojitostí tempa s vnímáním emocí by však bylo potřeba provést jiné a podrobnější zkoumání.

Za zmínku také stojí významnost melovských kepstrálních koeficientů, které bývají považovány za významné prediktory emocí (viz. přehled významných parametrů v tab. 2.1). V této analýze se objevují sice stále na předních místech, ne však na těch nejvyšších. Ve výčtu 10 nejvýznamnějších parametrů se vyskytují pouze u rozměru valence a to v podobě jejich delta koeficientů.

7.2.2 Nejvýznamnější parametry openSMILE

Na obr. 7.4 vidíme porovnání úspěšnosti jednotlivých metod výběru parametrů při použití parametrů nástroje openSMILE (sada parametrů D_{OS}). Zde vidíme výrazně vyšší úspěšnost predikce prvních deseti parametrů u metody SVR_{GaussM}, než u metody RRF_{200}. V přehledové tabulce nejvýznamnějších parametrů pro predikci vnímaných emocí 7.4 jsou uvedeny výsledky metody, která byla pro daný soubor parametrů úspěšnější. Opět je však nutno podotknout, že výsledky jednotlivých metod si jsou i v případe parametrů openSMILE velmi podobné, často zastávají v seznamu

dvaceti nejkvalitnějších parametrů pouze jinou pozici, jak můžeme vidět v 2. detailní přehledové tabulce A.2 v příloze této práce.



Obr. 7.4: Vyhodnocení úspěšnosti metod výběru parametrů na sadě parametrů D_{OS}.

Tab. 7.4: Nejvýznamnější parametry souboru openSMILE (D_{OS}) podle SVR_{GaussM}.

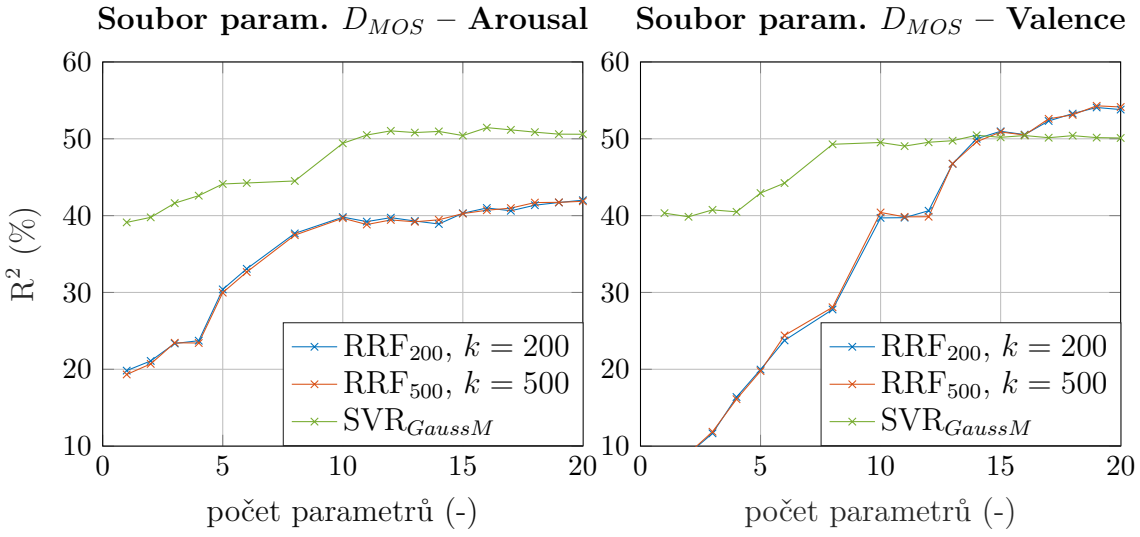
	Parametr sady D_{OS} – Arousal	**Parametr sady D_{OS} – Valence**
1	melovské spektrální koef. 07 (3. kvartil)	melovské spektrální koef. 07 (3. kvartil)
2	MFCC 00 (3. kvartil)	melovské spektrální koef. 07 (2. kvartil)
3	melovské spektrální koef. 07 (2. kvartil)	MFCC 00 (3. kvartil)
4	melovské spektrální koef. 07 (mean)	MFCC 07 (mean)
5	MFCC 00 (2. kvartil)	LS frekvence 00 (de, 2.-3. kvartil)
6	LS frekvence 00 (de, 2.-3. kvartil)	melovské spektrální koef. 06 (3. kvartil)
7	melovské spektrální koef. 06 (2. kvartil)	LS frekvence (de, 0.koef, 1.-3. kvartil)
8	LS frekvence 00 (de, abs. chyba)	LS frekvence 00 (de, abs. chyba)
9	MFCC 06 (3. kvartil)	melovské spektrální koef. 06 (2. kvartil)
10	LS frekvence 00 (de, 1.-3. kvartil)	MFCC 00 (2. kvartil)

V tabulce 7.2 můžeme jasně vidět, že nejvýznamnějšími parametry nástroje openSMILE pro predikci emocí u námi zkoumaných hudebních nahrávek jsou především melovské spektrální koeficienty, melovské kepstrální koeficienty a frekvence Line Spectral Pairs (LSP, více v kap. 2.3.5). Díky velkému množství použitých statistických parametrů se na předních příčkách objevují stejné parametry, pouze s jiným statistickým vyhodnocením. Za zmínku stojí skutečnost, že parametry vybrané pro dimenzi arousal i valence si jsou také velmi podobné.

Pokud se podíváme na pořadí parametrů podle SVR_{GaussM} dále, od 37. příčky se v seznamu pro rozměr valence začne objevovat parametr znělosti základní frekvence F_0 (kap. 2.3.2). V pořadí nejvýznamnějších parametrů pro rozměr arousal se tento parametr pohybuje až na přibližně 60. pozici.

7.2.3 Nejvýznamnější parametry souboru všech parametrů

Z obrázku 7.5 je zřejmé, že vyšší úspěšnost predikce u prvních 10 parametrů má metoda, SVR_{GaussM}. Proto budou v přehledové tabulce 7.5 zobrazeny výsledky této metody. Vůbec nejvýznamnějším parametrem pro obě dimenze je parametr spektrální fluktuace v oktávových pásmech (více v kap. 2.2.2). Tento parametr má nejvyšší potenciál jak pro dimenzi arousal, tak valence. To samé platí i pro melovské spektrální koeficienty, a to především 7. a 8. parametr. Jako další důležitý parametr ve vyhodnocení vychází spektrální entropie a to pro oba rozměry. Nelze si nevšimnout faktu, že přestože jsou oba emocionální rozměry ve své teoretické definici rozdílné, jejich nejvýznamnější parametry pro predikci jsou na předních místech značně podobné.



Obr. 7.5: Vyhodnocení úspěšnosti metod výběru parametrů na sadě parametrů D_{MOS}.

Ve zmíněné přehledové tabulce můžeme díky barevnému odlišení parametrů podle metody extrakce vidět zastoupení těchto metod. Oba nástroje MIRtoolbox i openSMILE jsou pro systém podobně přínosné.

Když se podíváme hlouběji do pořadí parametrů vyhodnocených metodou SVR, od přibližně 25. místa se objevují parametry jasnosti spektra, param. poklesu spek-

trální energie 85 % a také spektrální šikmost. To platí v různých obměnách pro arousal i valence. Tyto parametry lze tedy považovat také jako významné.

Tab. 7.5: Nejvýznamnější parametry souboru obsahující všechny extrahované parametry (D_{MOS}) podle SVR_{GaussM}. Parametry openSMILE jsou barevně odlišeny.

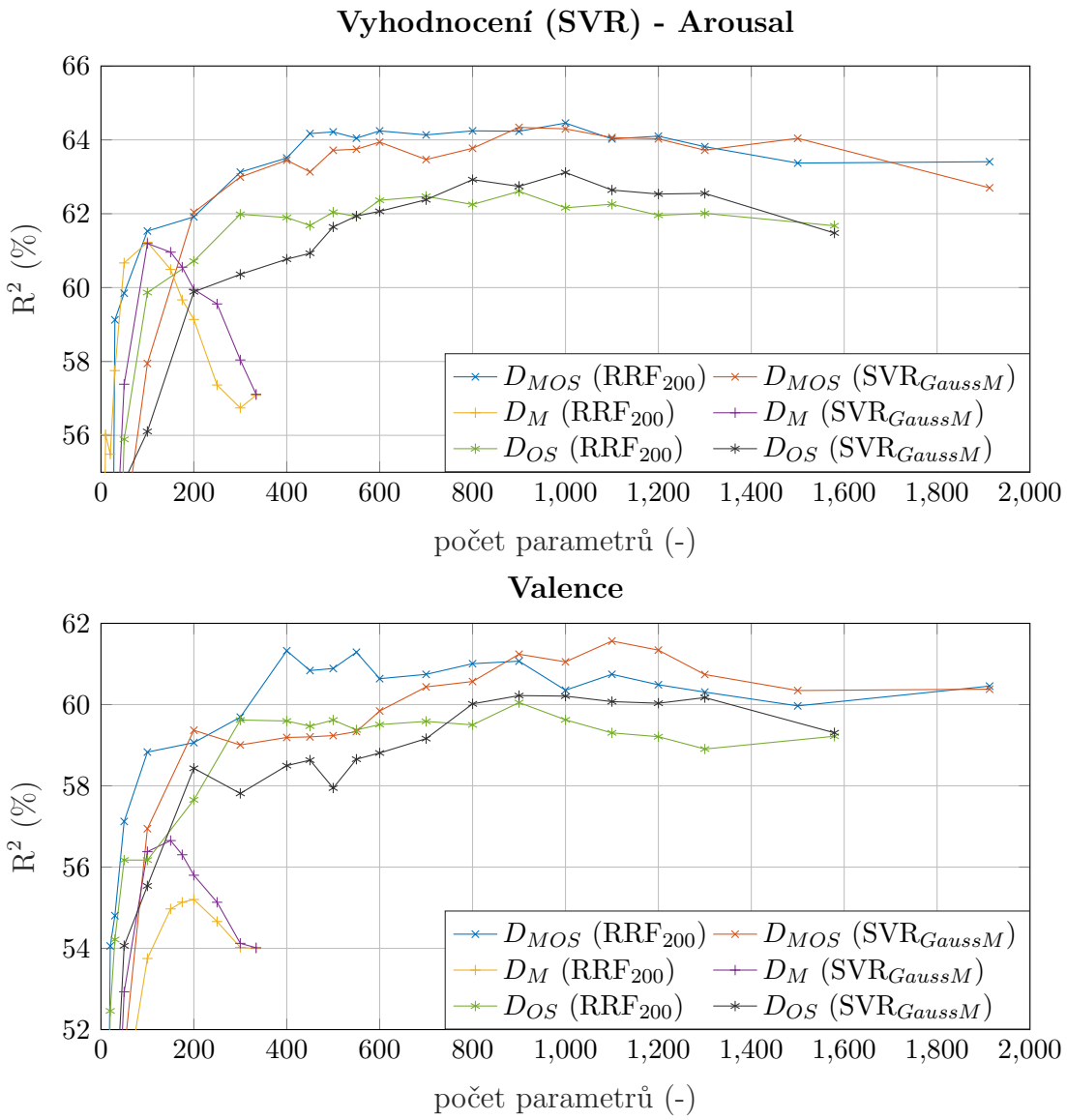
	Parametr sady D_{MOS} – Arousal	**Parametr sady D_{MOS} – Valence**
1	sp. fluktuace v okt. pásmech (08, mean)	sp. fluktuace v okt. pásmech (08, mean)
2	melovské spektrální koef. 07 (3. kvartil)	melovské spektrální koef. 07 (3. kvartil)
3	MFCC 00 (3. kvartil)	sp. fluktuace v okt. pásmech (09, mean)
4	sp. fluktuace v okt. pásmech (07, mean)	melovské spektrální koef. 07 (2. kvartil)
5	sp. fluktuace v okt. pásmech (09, mean)	spektrální entropie (mean)
6	melovské spektrální koef. 07 (2. kvartil)	MFCC 00 (3. kvartil)
7	melovské spektrální koef. 07 (mean)	melovské spektrální koef. 07 (mean)
8	MFCC 00 (2. kvartil)	LS frekvence 00 (de, 2.-3. kvartil)
9	spektrální fluktuace (mean)	spektrální fluktuace (mean)
10	spektrální entropie (mean)	melovské spektrální koef. 06 (3. kvartil)

7.2.4 Srovnání úspěšnosti datových sad

Na obrázku 7.6 můžeme vidět komplexní srovnání úspěšnosti predikce jednotlivých sad parametrů a jejich kombinace v závislosti na počtu parametrů použitých k natrénování predikčního modelu. Parametry byly seřazeny metodou RRF_{200} nebo SVR_{GaussM}. K natrénování těchto modelů byla využita opět metoda Střední SVR s Gaussovou jádrovou funkcí a 20-ti násobnou křížovou validací.

Z přehledu 7.6 vyplývá, že nejlepších výsledků je dosaženo při použití datasetu, který obsahuje parametry z obou nástrojů: jak MIRtoolboxu, tak openSMILE. Lze vidět, že modely natrénované pouze parametry souboru získaného MIRtoolboxem vykazují nižší úspěšnost, než modely s parametry openSMILE, či jejich kombinace. U parametrů MIRtoolboxu je zajímavé pozorovat trend rychlého růstu úspěšnosti při použití nejvýznamnějších parametrů a poté brzký a rychlý pokles díky velkému množství parametrů neužitečných. Tento jev není u openSMILE tolik znatelný. Nejlepší výsledky souboru MIRtoolboxu jsou pro arousal okolo 100 parametrů a valence 150, kde se láme křivka úspěšnosti mezi užitečnými a neužitečnými parametry. U souboru parametrů openSMILE je maximální úspěšnost také přibližně v půlce celkové počtu jeho parametrů – okolo 1000.

Nejdůležitější poznatek tohoto srovnání je již zmíněný fakt, že nejvyšší úspěšnost predikce emocí má soubor parametrů, který je kombinací všech extrahovaných parametrů obou nástrojů. Také je důležité zmínit, že není žádná metoda výběru



Obr. 7.6: Vyhodnocení závislosti úspěšnosti jednotlivých sad parametrů (při použití různých metod výběru parametrů) na počtu použitých parametru. D_{OS} je soubor parametrů openSMILE, D_M MIRtoolboxu a D_{MOS} je jejich kombinace.

parametrů, která by měla jasně lepší výsledky než druhá, a proto je potřeba zahrnout do celkové analýzy metody obě a vždy vyhodnotit, která má vyšší úspěšnost za daných podmínek.

7.3 Celkové vyhodnocení systému MER

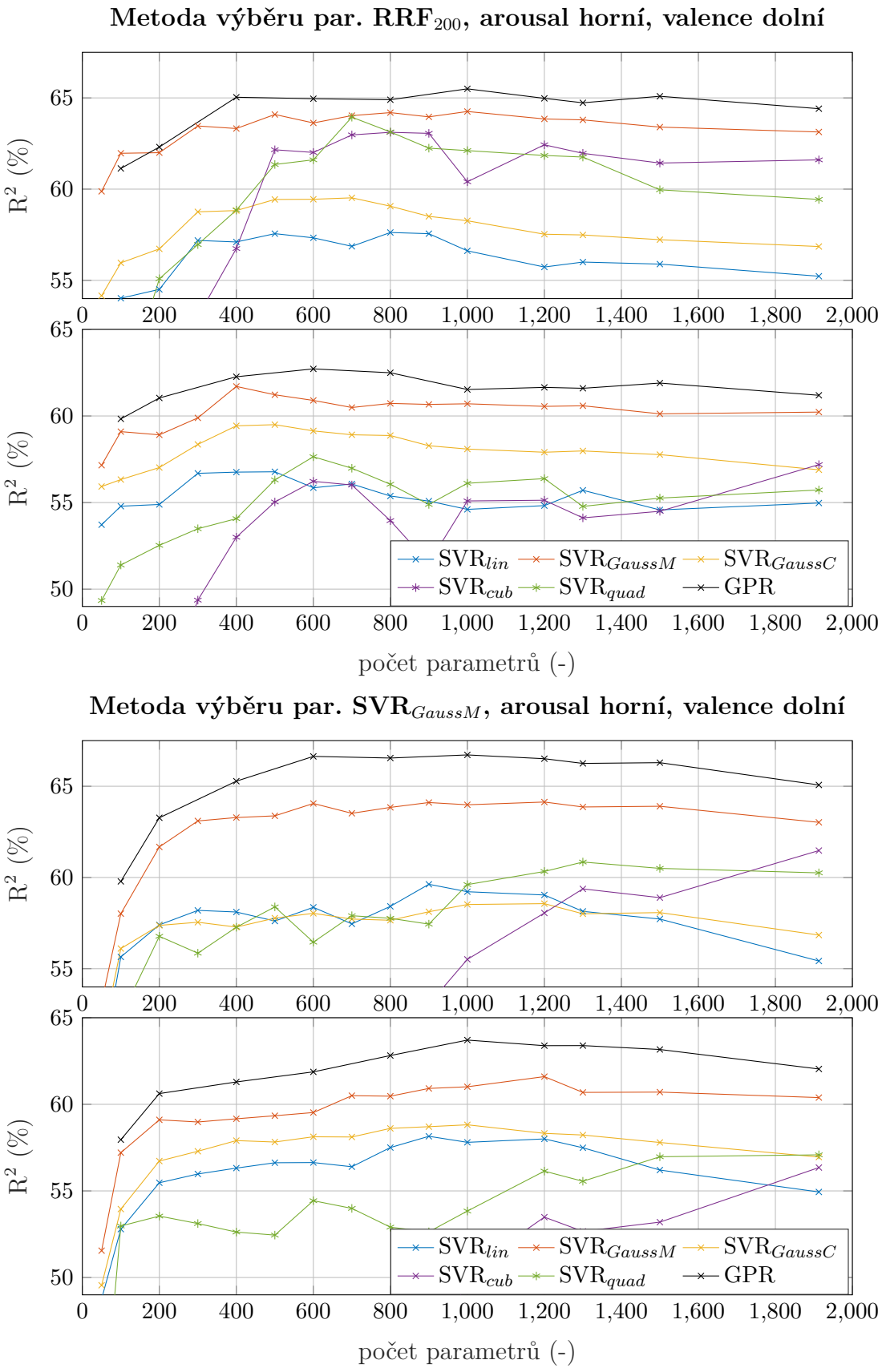
Druhou důležitou součástí vyhodnocení výsledků práce je, kromě individuální analýzy parametrů, srovnání jednotlivých metod výběru parametrů a souborů parametrů, také celková analýza úspěšnosti a uvedení nejlepších dosažených výsledků.

První se podívejme na výsledky úspěšnosti predikce emocí jednotlivých metod strojového učení. Jak bylo v předchozí kapitole ukázáno, datovým souborem s největším potenciálem pro predikci emocí je soubor obsahující jak parametry MIRtoolboxu, tak parametry openSMILE. Proto bude v této kapitole testován a vyhodnocován pouze tento set parametrů. Také bylo ukázáno, že obě metody metody výběru parametrů mají podobný potenciál, a proto budou do vyhodnocení začleněny obě. Z metod strojového učení byly vybráno 6 typů metod podpůrných vektorů (SVR) a metoda regrese pomocí Gaussovských procesů (GPR). Přehled jednotlivých metod SVR lze najít v tabulce 7.2 a jejich teoretický popis v kap. 4. Stejně jako v předchozích kapitolách je pro srovnání úspěšnosti a vynesení do grafu použita statistika R^2. Pro celkové tabulkové srovnání nejúspěšnějších konfigurací jsou zobrazeny také hodnoty RMSE, MSE, MAE a Spearmanova koeficientu pořadové korelace. Pro natrénování modelu byla použita 20 násobná křížová validace.

Na velkém přehledovém obrázku 7.7 můžeme vidět vyhodnocení úspěšnosti predikce jednotlivých vybraných typů modelů strojového učení v závislosti na počtu parametrů. Vyhodnocení bylo provedeno zvlášť pro obě představené metody výběru (ohodnocení) parametrů a také přirozeně zvlášť pro dimenzi arousal i valence. Prvním důležitým poznatkem je, že metod Gaussovských procesů má ve všech případech větší úspěšnost predikce, než jakákoliv metoda SVR. Pokud se podíváme blíže na množinu metod SVR, vidíme, že nejúspěšnější je SVR_{GaussM}, tedy střední Gaussova metoda podpůrných vektorů. Pro metodu RRF_{200} (horní grafy obrázku) má model SVR_{quad} při cca 700 použitých parametrech výsledky velmi dobré, zatímco pří hodnocení parametrů seřazených podle SVR_{GaussM} ve výsledcích propadá. Podobné chování pozorujeme i u metody SVR_{cub} (SVR s kubickou jádrovou funkcí). Lineární SVR má u obou metod stabilní výsledky, avšak s nižší úspěšností. Hrubá Gaussova SVR (SVR_{GaussC}) měla v hodnocení velmi špatné výsledky, a proto nebyla do grafů přehledu ani zanesena. Pro všechny metody platí, že predikování emocí v rozměru arousal má stabilně vyšší úspěšnost, než predikce pro rozměr valence.

7.3.1 Vyhodnocení pro arousal

V tabulce 7.6 se nachází přehled nejlepších dosažených výsledků predikce hodnot arousalu jednotlivých regresních metod. Nejlepší schopnost predikce arousalu má metoda GPR s 500 použitými parametry seřazenými metodou SVR a to 66,72 % R^2



Obr. 7.7: Vyhodnocení úspěšnosti jednotlivých typů modelů na datovém souboru RRF_{200} (nahoře) a SVR_{GaussM} (dole). Horní graf ve dvojici je pro arousal a dolní pro valence.

statistiky. Druhou nejvyšší hodnotu predikce, 64,25 % R^2 má SVR_{GaussM} s 1000 použitých parametrů seřazených pomocí RRF_{200}. Velmi překvapivého výsledku dosáhla i SVR_{quad} s výsledkem 63,94 % R^2 při 700 parametrech zařazených do trénování. Tento výsledek můžeme vidět i v horní části obrázku 7.7, kde se křivka (zelená) v okolí 700 parametrů láme a s přibývajícími parametry klesá úspěšnost. I v tomto hodnocení si lze všimnout, že i pro nejlepší predikci stačí méně než polovina z celkového počtu parametrů. Zmíněný poznatek lze ideálně demonstrovat na metodě GPR, která i s 1000 použitými parametry dosáhla úplně nejlepších výsledků predikce.

Tab. 7.6: Nejlepší výsledky vyhodnocení predikce podle metod strojového učení pro rozměr **arousal**.

metoda	**sada**	**výběr par.**	n	R^2 (%)	**RMSE**	**MSE**	**MAE**	ρ
GPR	D_{MOS}	SVR_{GaussM}	1000	**66,72**	0,093	8.6e−3	0,073	0,82
SVR_{GaussM}	D_{MOS}	RRF_{200}	1000	**64,25**	0,096	9.2e−3	0,075	0,80
SVR_{quad}	D_{MOS}	RRF_{200}	700	**63,94**	0,098	9.6e−3	0,077	0,81
SVR_{cub}	D_{MOS}	RRF_{200}	800	**63,11**	0,099	9.8e−3	0,077	0,80
SVR_{lin}	D_{MOS}	SVR_{GaussM}	900	**59,63**	0,103	1.0e−2	0,080	0,78
SVR_{GaussC}	D_{MOS}	RRF_{200}	700	**59,52**	0,103	1.0e−2	0,081	0,78

D_{MOS} – soubor parametrů MIRtoolboxu a openSMILE n – počet použitých parametrů R^2 – koeficient determinace ρ – Spearmanův koeficient pořadové korelace GPR – Gaussian Process Regression SVR – Support Vectors Regression

7.3.2 Vyhodnocení pro valence

Tab. 7.7: Nejlepší výsledky vyhodnocení predikce podle metod strojového učení pro rozměr **valence**.

metoda	**sada**	**výběr par.**	n	R^2 (%)	**RMSE**	**MSE**	**MAE**	ρ
GPR	D_{MOS}	SVR_{GaussM}	600	**63,7**1	0,088	7.8e−3	0,070	0,79
SVR_{GaussM}	D_{MOS}	RRF_{200}	400	**61,71**	0,091	8.2e−3	0,072	0,78
SVR_{GaussC}	D_{MOS}	RRF_{200}	500	**59,50**	0,094	8.7e−3	0,075	0,76
SVR_{lin}	D_{MOS}	SVR_{GaussM}	900	**58,15**	0,095	9.1e−3	0,076	0,76
SVR_{quad}	D_{MOS}	RRF_{200}	600	**57,65**	0,097	9.4e−3	0,077	0,75
SVR_{cub}	D_{MOS}	SVR_{GaussM}	1913	**56,35**	0,099	9.8e−3	0,077	0,75

D_{MOS} – soubor parametrů MIRtoolboxu a openSMILE n – počet použitých parametrů R^2 – koeficient determinace ρ – Spearmanův koeficient pořadové korelace GPR – Gaussian Process Regression SVR – Support Vectors Regression

V tabulce 7.7 můžeme vidět, že i pro dimenzi valence platí podobný vývoj výsledků jako pro arousal, popsaný v předchozí části. Rozměr valence má vždy horší výsledky predikce a výsledky této práce nejsou výjimkou. I přesto jsou dosažené výsledky pro tento rozměr velmi dobré v porovnání s výsledky jiných prací. Toto porovnání bude více rozvedeno v kapitole diskuze 8. Nejlepší dosažené výsledky má metoda Gaussovských procesů s 600 použitými parametry a hodnotou R^2 63,71 %. O přesně 2 % méně má metoda SVR_{GaussM} s 61,71 %.

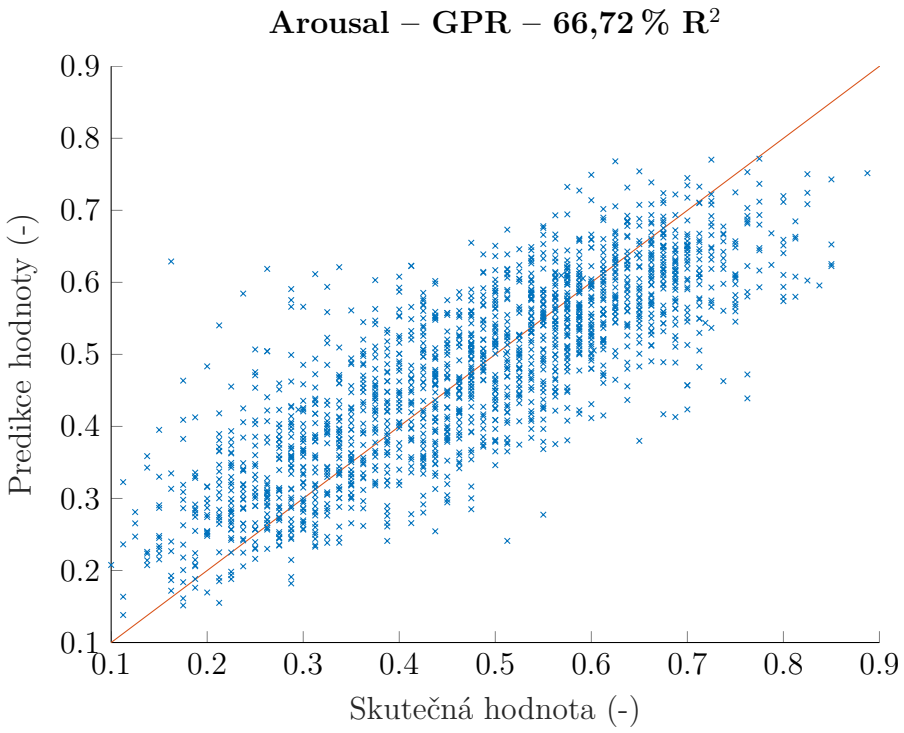
7.4 Statistické vyhodnocení nejlepších konfigurací modelu

Pro získání statisticky přesných výsledků bylo u 2 nejúspěšnějších metod provedeno 10 měření (trénování modelu). Každý tento model byl (jako všechny ostatní případy analýzy) vyhodnocen 20 násobnou křížovou validací. Z takto provedených 10 měření byla poté vypočítána průměrná hodnota R^2 a její směrodatná odchylka. Výsledky jsou prezentovány v tabulce 7.8. Na obrázku 7.8 můžeme pro ilustraci vidět zakreslené skutečné a predikované hodnoty každého pozorování pro dimenzi arousal u modelu s nejlepší úspěšností predikce. Přímka proložená grafem zobrazuje ideální predikci.

Tab. 7.8: Vyhodnocení dvou neúspěšnějších metod MER.

dimenze	**metoda**	**sada**	**výběr par.**	$\boldsymbol{n}$	$\mathbf{R^2 \pm \sigma\,(\%)}$
arousal	GPR	D_{MOS}	SVR	1000	**66,58 ± 0,36**
	SVR_{GaussM}	D_{MOS}	SVR_{GaussM}	1000	**64,40 ± 0,23**
valence	GPR	D_{MOS}	SVR	600	**61,56 ± 0,26**
	SVR_{GaussM}	D_{MOS}	SVR_{GaussM}	400	**61,44 ± 0,13**

D_{MOS} – soubor parametrů MIRtoolboxu a openSMILE n – počet použitých parametrů R^2 – koeficient determinace GPR – Gaussian Process Regression σ – směrodatná odchylka (%)



Obr. 7.8: Zobrazení skutečných a predikovaných hodnot u predikčního modelu s nejlepším výsledkem.

8 Diskuze

V této kapitole bude uvedeno porovnání dosažených výsledků s jinými dostupnými publikovanými pracemi a celkové zhodnocení úspěšnosti systému.

Jedna ze studií, která se také zabývá predikcí hudebních emocí v dimenzionálních modelech je [14]. A. Huq a kol. dosáhli v této studii pomocí SVR s implementovanou radiální bázovou funkcí výsledků 69,7 % R^2 pro arousal a 25,8 % R^2 pro valence. V dimenzi arousal jsou zmíněné výsledky lepší, než dosáhla tato práce, avšak jejich účinnost predikce pro rozměr valence je nízký. Je důležité podotknout, že je obecně poněkud problematické usuzovat závěry z takového typu srovnání výsledků mezi studiemi, protože velmi významným činitelem úspěšnosti predikce jsou vstupní data (ground-truth), která jsou povětšinou značně rozdílná. Například se lze pouze domnívat, že metoda SVR s radiální bázovou jádrovou funkcí (RBF) může mít lepší výsledky, než metody použité v této práci.

Další prací, kterou je vhodné pro srovnání uvést, je článek často citovaného autora oblasti MER X.Yanga a kol. [40]. Z této detailní studie bylo pro potřeby diplomové práce čerpáno mnoho teoretických poznatků. Studie X. Yanga a kol. zároveň dosáhla nejlepších výsledků pomocí metody výběru parametrů RReliefF spolu s metodou metodou učení SVR – 58,3 % R^2 v dimenzi arousal a 28,1 % R^2 v dimenzi valence, což jsou znatelně nižší hodnoty, než vycházejí v této práci. Přestože má tato práce stejnou metriku hodnocení, používá regresní model a také stejnou metodu výběru parametrů jako model navržený v diplomové práci, používá zmíněná studie jinou sadu parametrů (Psy15) a o mnoho menší soubor vstupních dat (25 s úryvky 195 populárních písní). V neposlední řadě také využívá jiné subjektivní označení emocí. Nižší úspěšnost predikce zmíněné studie oproti této diplomové studii byla očekávána.

Jak můžeme vidět z předchozích srovnání, v této diplomové práci se povedlo dosáhnout poměrně vysokého výsledku predikce rozměru valence oproti jiným (zmíněným) studiím. Vysvětlení může být následovné. Jedno z řešení již zmíněné soutěže v oblasti MER *MediaEval 2015*[1] pracovalo se stejnou databází ground-truth dat (databáze DEAM) jako tato práce a využívalo mimo jiné také parametry extrahované pomocí nástroje openSMILE. Přestože se práce primárně zaměřuje na predikci emocí v čase (MEVD) také používá metodu SVR. Jejich zveřejněné výsledky jsou 51,59 % R^2 pro arousal a 48,03 % R^2 pro valence. Lze se tedy domnívat, že vysoká úspěšnost predikce pro valence tkví v povaze použité databáze skladeb a emocí DEAM (viz kap. 3).

Jednoho zajímavého zjištění bylo dosaženo při porovnání jednotlivých souborů parametrů. Oba soubory parametrů openSMILE i MIRtoolbox vyšly jako významné

[1] http://www.multimediaeval.org

pro predikci. Soubor openSMILE se ukázal v určitých vyhodnoceních jako lepší, přestože obsahuje převážně typické parametry aplikované a vyvinuté pro rozpoznávání a analýzu řeči.

Ačkoliv byla metoda Gaussovských procesů zahrnuta do práce jen okrajově, vykazuje nejlepší výsledky. Tato metoda je ovšem velmi výpočetně náročná a v daných podmínkách by bylo neuskutečnitelné ji využívat při podrobné analýze, jak tomu bylo praktikováno s metodou SVR.

Oblast rozpoznávání hudebních emocí je v našem prostředí zatím poměrně neprobádaná a nelze říci, že by bylo dostatek literatury, na které by se dalo stavět a navazovat. Proto tato práce začíná na samém základu a v teoretické části popisuje všechny metody a kategorie zmíněné oblasti. V praktické části se vydává z již osvětlených důvodů směrem regresní analýzy na rozdíl od běžnější a jednodušší klasifikační analýzy.

Lepších výsledků práce by bylo dosaženo přirozeně větším souborem vstupních dat, tedy rozsáhlejší databází subjektivních anotací a hudebních nahrávek. V oblasti subjektivního hodnocení pomocí respondentů, zvláště pokud jde o vnímání hudby, má o to zásadnější roli povaha, metodika a celková kvalita sběru těchto dat, čímž se tato práce nezabývala. Dalším důležitým faktorem je pečlivý výběr parametrů pro oba emoční rozměry. Přestože tato práce provedla základní vyhodnocení významnosti parametrů, nabízí se možnost podrobného zkoumání a vyhodnocování kombinací jednotlivých parametrů. Také nebyly využity úplně všechny parametry, které byly literárními zdroji vyhodnoceny jako cenné, např. *DWCH (Daubechies Wavelet Coefficient Histograms)*. Jednalo by se však o práci značně nad rámec možností diplomové práce. To potvrzuje fakt, že během rešerše a sběru literatury nebyla nalezena jediná studentská práce, která by se tímto tématem zabývala, nýbrž práce publikované týmy vědců.

Jedním z podnětů k dalšímu výzkumu je rozhodně možnost využití neuronových sítí, které nebyly v této práci implementovány. Neuronové sítě vykazují výborné schopnosti predikce hudebních emocí a v množství studiích vycházejí ze všech metod strojového učení nejlépe. Slabina této práce je v šíři záběru, tedy snaze obecně popsat celou oblast MER a navrhnout vyhodnocovací systém. Je pravděpodobné, že kdyby bylo zadání práce úzce specializované na daný typ modelu, typu predikce a metody, mohla by optimalizace a vyhodnocení být podrobnější a tedy dosáhnout i lepších výsledků.

Také je třeba zdůraznit, že přestože je nástroj openSMILE pravděpodobně vynikající pro praktickou realizaci systému rozpoznávání emocí a extrakci parametrů v reálném čase, nehodí se příliš pro vědecký výzkum a teoretické práce z důvodu značně slabé dokumentace procesu získávání parametrů a velmi nedostatečného popisu výpočtu jednotlivých parametrů. Za zmínku také stojí to, že aplikace pro hro-

madnou extrakci parametrů z hudebních nahrávek je špatně optimalizovaná, špatně se v ní orientuje, obsahuje chyby a má obecně málo možností nastavení.

V práci bylo použito velké množství parametrů které sice byly redukovány a byly vybrány ty nejhodnotnější podle použitých metod, podrobnější analýza a detailní zkoumání významnosti jednotlivých statistických parametrů by však také mohlo zlepšit výsledky práce.

Jako poslední podnět k dalšímu zkoumání je doporučení použití tzv. **adjustovaného koeficientu determinace** místo klasického koeficientu R^2, který má tendenci lehce růst s počtem nezávisle proměnných v regresním systému, přestože přidávané proměnné nenesou žádné nové informace o závisle proměnné. Adjustovaný koeficient determinace tento efekt inflace původního koeficientu eliminuje. Zajímavé je, že i přes tento známý fakt, většina prací využívá stále koeficient původní. Srovnání výpovědní hodnoty těchto dvou koeficientů by mohlo přinést užitečnou informaci.

Využití SFFS

Pro použití v této práci byla také uvažována metoda SFFS (Sequential Forward Floating Selection). Metoda byla implementována ([11]) a pro potřeby práce testována, avšak bylo uváženo, že tento typ vyhodnocení není pro tuto práci optimální. Díky tomu, že regresní MER systém má velké množství užitečných parametrů, které jsou přínosné pro predikci a zároveň je nejlepších hodnot predikce dosahováno při vysokém počtu parametrů, musela by tato metoda (i navzdory předcházející selekci přebytečných parametrů) vyhodnocovat velké množství parametrů. Vzhledem k velké výpočetní náročnosti této metody by bylo toto vyhodnocení výpočetně enormně náročné.

9 Závěr

V úvodu této práce byla obecně představena vědní disciplína *Music Information Retrieval* a její podskupina soustřeďující se na rozpoznávání hudebních emocí a nálad, nazývaná jako *Musis emotion recognition*. V úvodu práce nalezneme obecné uvedení do kontextu problematiky a motivaci.

V první kapitole se již nachází popis, definice a rozdělení MER systémů a metod. Tato kapitola sumarizuje a kategorizuje informace z jednotlivých studií a literárních zdrojů do přehledného systému rozdělení a umožňuje vytvořit ucelený obraz obecně používaných metod a postupů v MER, což je velmi důležité pro následný návrh vyhodnocovacího systému.

Druhá kapitola se komplexně zabývá hudebními parametry pro rozpoznávání emocí s primárním zaměřením na popis parametrů, které jsou použity v této práci. Tato část je rozdělena na dva celky podle použitých nástrojů pro extrakci parametrů. Jedná se o nástroj MIRtoolbox, který je přímo určen pro oblast získávání informací z hudby, a nástroj openSMILE, jehož převážná část extrahovaných parametrů (příznaků) spadá spíše do oblasti rozpoznávání řeči. Je zde uveden jak tabulkový přehled všech parametrů a jejich statistických vyhodnocení, tak jejich jednotlivý popis a vysvětlení.

Obsahem třetí kapitoly je popis databáze hudebních nahrávek a anotací s názvem DEAM, která bude v práci využita. Není velké množství dostupných databází nahrávek spolu s přiloženými emocionálními popisy (anotacemi), a proto bylo potřeba vhodně vybrat. Databáze DEAM je z nich nejobsáhlejší (1740 úryvků), přičemž byl každý úryvek hodnocen mezi 7 až 23 respondenty. Jedná se tedy o statický dimenzionální model anotací s číselným typem ground-truth dat.

Čtvrtá kapitola navazuje rozborem použité regresní metody podpůrných vektorů. Následuje kapitola popisující použité metody statistické analýzy.

Šestá kapitola se zaměřuje na již samotný návrh vyhodnocovacího MER systému, který byl vhodně zvolen jak s ohledem na povahu vstupních dat, tak se zřetelem na současný vývoj a výsledky v oboru MER. Jedná se tedy o dimenzionální systém využívající extrahované hudebních parametry a číselná grount-truth data získaná metodou statického dimenzionálního modelu anotací. Tento systém používá regresní metody strojového učení, pomocí kterých predikuje polohu jednotlivých hudebních úryvků v AV emočním prostoru. Pro trénování regresního modelu byla vybrána metoda SVR pro její široké využití a také stabilně dobré výsledky predikce. Bylo navrhnuto také zapojení algoritmu standardizace pro redukci počtu parametrů pro strojové učení, a tedy snížení dimenzionality systému.

Sedmá kapitola se věnuje vyhodnocení praktické části práce, tedy navrženému predikčnímu modelu. Začíná vyhodnocením metody RReliefF, která je použita pro

ohodnocení významnosti jednotlivých parametrů. Experimentálně bylo dokázáno, že pomocí této metody lze úspěšně hodnotit významnost parametrů, tedy že je přínosná pro predikci. Byla vybrána optimální hodnota nejbližších sousedů, jako jediný vstupní parametr této metody $k_r = 200$.

Další část se zabývá individuální analýzou parametrů obou použitých datových sad a jejich kombinací. Z parametrů MIRtoolboxu je potřeba vyzdvihnout především parametr spektrální fluktuace, který má nejlepší schopnost predikce jak rozměru arousal, tak valence. Dále patří mezi významné parametry jasnost spektra, spektrální entropie a fluktuace tempa. Naopak nejvýznamnějšími parametry extrahovanými pomocí nástroje openSMILE jsou melovské spektrální koeficienty a to opět pro obě dimenze emocí. Dalšími důležitými parametry jsou melovské kepstrální koeficienty a frekvence LSP. V souhrnné analýze parametrů stojí na předních příčkách významnosti právě kombinace výše zmíněných parametrů.

Ze srovnání úspěšnosti jednotlivých sad parametrů vychází nejlépe jejich společná kombinace. Výsledkem analýzy je také zjištění, že parametry openSMILE obsahují více informací pro úspěšnou predikci emocí, než parametry MIRtoolboxu.

V poslední části kapitoly se nachází vyhodnocení jednotlivých metod strojového učení a hledání jejich nejúspěšnější konfigurace. Z metod SVR má nejúspěšnější hodnotu predikce střední Gaussova SVR (SVR_{GaussM}) s hodnotou statistiky R^2 64,40 $\pm$ 0,23 % pro arousal a 61,44 $\pm$ 0,13 % pro valence. Úplně nejvyšší úspěšnost predikce však dosáhla metoda Gaussovských procesů se statistikou R^2 66,58 $\pm$ 0,36 % pro arousal a 61,56 $\pm$ 0,26 % pro rozměr valence. Úplný závěr práce se zabývá doporučeními pro případný další výzkum, známými nedostatky práce a návrhy pro dosažení teoreticky lepších výsledků.

Literatura

[1] ALAJANKI, A., YANGYi-Hsuan, Y. a SOLEYMANI, M. Benchmarking music emotion recognition systems. *PLOS ONE*. 2016.

[2] ALJANAKI, A. *Emotion in Music: representation and computational modeling*. Utrecht University, 2016. ISBN 978-94-6328-083-9. University Utrecht.

[3] Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*. 2018, 324. DOI: 10.1088/1757-899X/324/1/012049. ISSN 1757-8981.

[4] BARTHET, M., G. FAZEKAS a M. SANDLER. *Music Emotion Recognition: From Content- to Context-Based Models*. From Sounds to Music and Emotions. CMMR: International Symposium on Computer Music Modeling and Retrieval: Springer, Berlin, Heidelberg, 2013. ISBN 978-3-642-41248-6.

[5] BARTOSZEWSKI, M., H. KWASNICKA, U. MARKOWSKA-KACZMAR a P. B. MYSZKOWSKI. *Extraction of Emotional Content from Music Data*. 2008. 293-299. DOI: 10.1109/CISIM.2008.46.

[6] BASAK, D., S. PAL a D. CH. PATRANABIS. Support Vector Regression. Indie: Electrical Laboratory, Central Institute of Mining and Fuel Research, 2007.

[7] BITTNER, R., SALAMON, J., TIERNEY, M., MAUCH, M., CANNAM, C., BELLO, J. P., M*edleyDB: A multitrack dataset for annotation- intensive mir research*, Proc. ISMIR, 2014.

[8] BOSWELL, Dustin. Introduction to Support Vector Machines. 2002.

[9] EYBEN, F. Real-time Speech and Music Classification by Large Audio Feature Space Extraction. Springer, 2016. ISBN 978-3-319-27299-3.

[10] EYBEN, F., F. WENINGER, M. WOLLMER a B. SCHULLER. Open-Source Media Interpretation by Large feature-space Extraction [online]. Verze 2.3. TU Munchen, MMK, 2016 [cit. 2019-02-28]. Dostupné z: https://www.audeering.com/opensmile/

[11] GALÁŽ, Z. Preliminary Acoustic Analysis of Noise Components in Patients In Parkinsons Disease. In Proceedings of the 21st Conference STUDENT EEICT 2015. Brno: 2015. p. 476-480. ISBN: 978-80-214-5148- 3.

[12] HENDL, Jan. *Přehled statistických metod: analýza a metaanalýza dat.* Páté, rozšířené vydání. Praha: Portál, 2015. ISBN 978-80-262-0981-2.

[13] HEVNER, K. Expression in music: a discussion of experimental studies and theories. *Psychological Review.* 1935, 42(2), 186-204.

[14] HUQ, A., J. P. BELLO a R. ROWE. Automated Music Emotion Recognition: A Systematic Evaluation. *Journal of New Music Research.* 2010, (39), 227-244.

[15] ITAKURA, F. Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals, J. Acoust. Soc. Am., Vol. 57, S35, 1975.

[16] KIM, J., S. LEE, S. KIM a W.Y. YOO. Music mood classification model based on arousal–valence values. *13th International Conference on Advanced Communication Technology (ICACT).* New York, 2011, , 292–295.

[17] KIRA, K. – RENDELL, L. A. The feature selection problem: traditional methods and a new algorithm. *Proceedings of the tenth national conference on Artificial intelligence.* San Jose, California. AAAI Press, 1992, s. 129–134. ISBN 0-262-51063-4.

[18] KNEES, P., SCHEDL, M. *Music similarity and retrieval: an introduction to audio- and web-based strategies.* Vol. 36. New York, NY: Springer Berlin Heidelberg, 2016. ISBN 9783662497203.

[19] KONONENKO I, ROBNIK-ŠIKNOJA M: Theoretical and Empirical Analysis of ReliefF and RreliefF. Machine Learning. 2003, volume 53, s. 23-69. DOI: 10.1023/A:1025667309714. Dostupné z: http://lkm.fri.uni-lj.si/rmarko/papers/robnik03-mlj.pdf

[20] KRÁL, Vítězslav. *Určování období vzniku interpretace za pomoci metod parametrizace hudebního signálu.* Brno, 2017, 72 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: Ing. Tomáš Kiska

[21] KRUMHANSLOVÁ, C.L. An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology.*1997, 51(4), 336–353

[22] LARTILLOT, O. MIRtoolbox 1.7 *User's Manual* [online]. 2017 [cit. 201711-11]. Dostupné z www.jyu.fi.

[23] LERCH, Alexander. *Audio content analysis: an introduction.* Hoboken, N.J.: Wiley, c2012. ISBN 978-1-118-26682-3.

[24] LYONS, J. *Mel Frequency Cepstral Coefficient (MFCC) tutorial* [online]. [cit. 2018-12-05]. Dostupné z: http://practicalcryptography.com/miscellaneous/-machine-learning/guidemel-frequency-cepstral-coefficients-mfccs/

[25] MCLOUGHLIN, Ian Vince. Line spectral pairs. Signal Processing. 2008, vol. 88, issue 3, s. 448-467. DOI: 10.1016/j.sigpro.2007.09.003. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/S0165168407003167

[26] MEURS, Joris. *R2 Calculator: MATLAB function* [online]. In: MathWorks File Exchange Web, 2016 [cit. 2019-04-17]. Dostupné z : wwww2.mathworks.cn

[27] MEYER, LEONARD, B., *Emotion and meaning in music*. Chicago: University of Chicago Press, 1956.

[28] MÜLLER, Meinard. *Fundamentals of music processing: audio, analysis, algorithms, applications*. Cham: Springer, 2015. ISBN 978-3-319-35765-2.

[29] PANDA, R., ROCHA, B., PAIVA R. P.(2015). *Music Emotion Recognition with Standard and Melodic Audio Features, Applied Artificial Intelligence*, 29:4, 313-334, DOI: 10.1080/08839514.2015.1016389

[30] RUSSELL, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.

[31] SCHOLKOPF, Bernhard a Alexander J. SMOLA. Learning with kernels: support vector machines, regularization, optimization, and beyond. Cambridge, Mass.: MIT Press, c2002. ISBN 0262194759.

[32] SLOBODA, J. A., JUSLIN, P. N. 2001. Psychological perspectives on music and emotion. *Music and Emotion: Theory and Research*, Oxford University Press, Oxford, UK.

[33] SMÉKAL, Z. Zpracování řeči. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2013. ISBN 978-80-214-4896-4. Elektronicky.

[34] SMOLA, A.J. a B. SCHOLKOPF. A tutorial on support vector regression. *Statistics and Computing*. Max-Planck-Institut fur biologische Kybernetik, Tubingen, Germany: Kluwer Academic Publishers. Manufactured in The Netherlands., 2002, (14), 199–222.

[35] Statistics and Machine Learning Toolbox: User's Guide. 2019. The MathWorks, Inc., Natick, MA. Dostupné také z www.mathworks.com.

[36] SVOZIL, Martin *Statistické zpracování řečových parametrů*: diplomová práce. Místo: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky, 2014. 54 s. Vedoucí práce byl Ing. Miroslav Staněk

[37] TZANETAKIS, G. a P. COOK. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing. 2002, 10(5), 293-302. DOI: 10.1109/TSA.2002.800560.

[38] WENINGER, F., EYBEN F., SCHULLER B. The TUM Approach to the MediaEval Music Emotion Task Using Generic Affective Audio Features. In: *Working Notes Proceedings of the MediaEval 2013 Workshop*; 2013.

[39] YANG, X., Y. DONG a J. LI. Review of data features-based music emotion recognition methods. *Multimedia Systems*. Berlin, Heidelberg: Springer-Verlag, 2017. ISSN 0942-4962. Dostupné také z: https://doi.org/10.1007/s00530-017-0559-4

[40] YANG, Y., H. CHEN, Y. LIN a X. SU. A Regression Approach to Music Emotion Recognition. *IEEE Transactions on Audio Speech and Language Processing*. 2008. DOI: 10.1109/TASL.2007.911513.

[41] YANG, Yi-Hsuan a Homer H. CHEN. Machine Recognition of Music Emotion. *ACM Transactions on Intelligent Systems and Technology*. 2012, (3). DOI: 10.1145/2168752.2168754. ISSN 21576904. Dostupné také z: http://dl.acm.org/citation.cfm?doid=2168752.2168754

[42] ZENTRNER, M., GRANDJEAN, D., a SCHENER, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4), 494-521.

Seznam symbolů, veličin a zkratek

DEAM	MediaEval Database of for Emotional Analysis in Music
GPR	regrese Gaussovskými procesy (Gaussian Process Regression)
LSP	lineární spektrální páry (Line Spectral Pairs)
MER	Music Emotion Recognition - rozeznávání hudebních emocí
MEVD	Music Emotion Variation Detection
MEC	Music Emotion Classification
MFCC	Melovské kepstrální koeficienty
MIR	Music Information Retrieval
MIREX	Music Information Retrieval Evaluation eX-change
PCA	Principal Component Analysis Analýza hlavních komponent
RRF	RReliefF
SFFS	Sequential Forward Floating Selection
SVM	Support Vector Machines – metoda podpůrných vektorů
SVR	Support Vector Regression – regresní typ metody podpůrných vektorů
VA	valence (libost) a arousal (energičnost)
ZRC	Zero Crossing Rate – počet průchodů nulou

Seznam příloh

A Tabulky

Tab. A.1: Individuální analýza kvality parametrů extrahovaných z MIRtoolboxu pomocí RReliefF a SVM modelu. Vysvětlení jednotlivých zkratek a popis parametrů se nachází v kap. 2.

	RreliefF $k_r = 200$, w - váha (%)				**Střední Gaussova SVM** R^2 statistika (%)			
	Arousal		**Valence**		**Arousal**		**Valence**	
	w	Parametr	w	Parametr	R^2	Parametr	R^2	Parametr
1	1,16	m_spectralflux_mean	1,15	m_sflux_oct07_mean	39,21	m_sflux_oct08_mean	40,07	m_sflux_oct08_mean
2	1,13	m_sflux_oct07_mean	1,06	m_spectralflux_mean	36,80	m_sflux_oct07_mean	37,20	m_sflux_oct09_mean
3	0,97	m_brightness_mean	0,81	m_sflux_oct06_mean	36,61	m_sflux_oct09_mean	35,88	m_spectentropy_mean
4	0,95	m_fluct_mean	0,75	m_sflux_oct08_mean	34,10	m_spectralflux_mean	34,87	m_spectralflux_mean
5	0,93	m_sflux_oct06_mean	0,69	m_spectentropy_mean	33,85	m_spectentropy_mean	34,62	m_sflux_oct07_mean
6	0,93	m_spectentropy_mean	0,67	m_brightness_mean	33,41	m_sflux_oct10_mean	33,91	m_sflux_oct10_mean
7	0,70	m_attack_slope	0,61	m_attack_slope	31,37	m_brightness_mean	31,82	m_brightness_mean
8	0,70	m_skewness_mean	0,59	m_tempo_std	31,01	m_mfcc_mean_01	31,77	m_rolloff85_mean
9	0,70	m_sflux_oct03_mean	0,57	m_dmfcc_perAmp_02	30,89	m_rolloff85_mean	31,52	m_centroid_mean
10	0,67	m_sflux_oct08_mean	0,57	m_mfcc_perFreq_11	30,50	m_centroid_mean	31,29	m_skewness_mean
11	0,65	m_sflux_oct05_mean	0,57	m_sflux_oct05_mean	29,99	m_skewness_mean	29,98	m_mfcc_mean_01
12	0,65	m_mfcc_mean_01	0,52	m_fluct_mean	27,42	m_sflux_oct06_mean	28,48	m_kurtosis_mean
13	0,61	m_rolloff85_mean	0,51	m_ddmfcc_perAmp_01	27,26	m_kurtosis_mean	27,49	m_sflux_oct06_mean
14	0,60	m_sflux_oct02_mean	0,48	m_ddmfcc_perAmp_05	26,39	m_attack_slope	26,38	m_rolloff95_mean
15	0,53	m_kurtosis_mean	0,48	m_sflux_oct03_mean	26,26	m_zerocross_mean	26,25	m_flatness_mean
16	0,53	m_ddmfcc_perAmp_04	0,47	m_sflux_oct04_mean	25,10	m_rolloff95_mean	26,16	m_zerocross_mean
17	0,53	m_kurtosis_perFreq	0,45	m_mfcc_perFreq_10	24,52	m_flatness_mean	22,48	m_spread_mean

Tab. A.2: Individuální analýza kvality parametrů extrahovaných z openSMILE pomocí RReliefF a SVM modelu.Vysvětlení jednotlivých zkratek a popis parametrů se nachází v kap. 2.

	RreliefF $k_r = 200$, w - váha (%)				**Střední Gaussova SVM** R^2 statistika (%)			
	Arousal		**Valence**		**Arousal**		**Valence**	
	w	Parametr	w	Parametr	R^2	Parametr	R^2	Parametr
1	2,08	MelFreqBand_7_upt90	1,66	MelFreqBand_7_upt90	37,09	MelFreqBand_7_quart3	37,84	MelFreqBand_7_quart3
2	1,64	MelFreqBand_6_upt90	1,58	MelFreqBand_6_upt90	36,83	mfcc_0_quart3	36,15	MelFreqBand_7_quart2
3	1,57	lspFreq_de5_upt70	1,31	MelFreqBand_5_upt90	36,54	MelFreqBand_7_quart2	35,60	mfcc_0_quart3
4	1,57	lspFreq_de4_upt70	1,18	lspFreq_de4_upt70	35,30	MelFreqBand_7_mean	35,33	MelFreqBand_7_mean
5	1,48	mfcc_de0_kurt	1,12	mfcc_0_upt90	34,56	mfcc_0_quart2	35,03	lspFreq_de0_iqr2_3
6	1,47	MelFreqBand_de7_kurt	1,04	lspFreq_de5_upt70	33,56	lspFreq_de0_iqr2_3	34,84	MelFreqBand_6_quart3
7	1,40	MelFreqBand_5_upt90	1,04	MelFreqBand_4_upt90	33,32	MelFreqBand_6_quart2	34,64	lspFreq_de0_iqr1_3
8	1,36	MelFreqBand_de0_kurt	0,98	mfcc_0_quart3	33,26	lspFreq_de0_linregerrA	34,44	lspFreq_de0_linregerrA
9	1,31	MelFreqBand_7_kurt	0,97	MelFreqBand_3_upt90	33,02	MelFreqBand_6_quart3	34,17	MelFreqBand_6_quart2
10	1,29	MelFreqBand_de7_skew	0,95	MelFreqBand_de7_kurt	32,99	lspFreq_de0_iqr1_3	34,04	mfcc_0_quart2
11	1,27	MelFreqBand_de6_kurt	0,95	mfcc_0_quart2	32,97	MelFreqBand_7_quart1	33,97	lspFreq_de0_quart1
12	1,27	mfcc_de0_skew	0,94	lspFreq_6_kurt	32,92	lspFreq_de0_quart1	33,80	lspFreq_de0_iqr1_2
13	1,25	lspFreq_6_kurt	0,94	MelFreqBand_2_upt90	32,84	mfcc_0_mean	33,59	lspFreq_de0_quart3
14	1,24	MelFreqBand_3_upt90	0,92	MelFreqBand_7_quart2	32,04	lspFreq_de0_iqr1_2	32,97	MelFreqBand_6_mean
15	1,22	MelFreqBand_1_kurt	0,90	MelFreqBand_de0_kurt	32,03	MelFreqBand_6_mean	32,35	mfcc_0_percentile990
16	1,22	mfcc_0_upt90	0,89	mfcc_0_mean	31,77	lspFreq_de0_quart3	32,28	MelFreqBand_7_quart1
17	1,22	MelFreqBand_6_kurt	0,88	mfcc_de0_kurt	31,43	mfcc_0_percentile990	32,12	mfcc_0_mean
18	1,21	MelFreqBand_de3_kurt	0,87	MelFreqBand_7_mean	31,02	mfcc_1_quart2	31,84	mfcc_1_quart2
19	1,20	MelFreqBand_3_kurt	0,87	MelFreqBand_7_quart3	30,86	MelFreqBand_5_quart2	31,14	lspFreq_de0_stddev
20	1,18	MelFreqBand_4_upt90	0,84	MelFreqBand_de1_kurt	30,59	MelFreqBand_5_quart3	31,09	mfcc_1_mean

Tab. A.3: Individuální analýza kvality všech parametrů pomocí RReliefF a SVM modelu.Vysvětlení jednotlivých zkratek a popis parametrů se nachází v kap. 2.

	RreliefF $k_r = 200$, w - váha (%)				**Střední Gaussova SVM** R^2 statistika (%)			
	Arousal		**Valence**		**Arousal**		**Valence**	
	w	Parametr	w	Parametr	R^2	Parametr	R^2	Parametr
1	2,10	MelFreqBand_7_upt90	1,58	MelFreqBand_7_upt90	39,21	m_sflux_oct08_mean	40,07	m_sflux_oct08_mean
2	1,66	lspFreq_de5_upt75	1,55	MelFreqBand_6_upt90	37,09	MelFreqBand_7_quart3	37,84	MelFreqBand_7_quart3
3	1,65	MelFreqBand_6_upt90	1,32	MelFreqBand_5_upt90	36,83	mfcc_0_quart3	37,20	m_sflux_oct09_mean
4	1,61	lspFreq_de4_upt75	1,19	lspFreq_de4_upt75	36,80	m_sflux_oct07_mean	36,15	MelFreqBand_7_quart2
5	1,47	MelFreqBand_de7_kurt	1,07	mfcc_0_upt90	36,61	m_sflux_oct09_mean	35,88	m_spectentropy_mean
6	1,46	mfcc_de0_kurt	1,07	m_flatness_perFreq	36,54	MelFreqBand_7_quart2	35,60	mfcc_0_quart3
7	1,46	MelFreqBand_5_upt90	1,06	MelFreqBand_4_upt90	35,30	MelFreqBand_7_mean	35,33	MelFreqBand_7_mean
8	1,38	MelFreqBand_de0_kurt	1,05	lspFreq_de5_upt75	34,56	mfcc_0_quart2	35,03	lspFreq_de0_iqr2_3
9	1,35	m_flatness_perFreq	1,03	m_sflux_oct07_mean	34,10	m_spectralflux_mean	34,87	m_spectralflux_mean
10	1,34	MelFreqBand_7_kurt	0,99	MelFreqBand_3_upt90	33,85	m_spectentropy_mean	34,84	MelFreqBand_6_quart3
11	1,32	lspFreq_6_kurt	0,98	MelFreqBand_2_upt90	33,56	lspFreq_de0_iqr2_3	34,64	lspFreq_de0_iqr1_3
12	1,30	MelFreqBand_de6_kurt	0,95	lspFreq_6_kurt	33,41	m_sflux_oct10	34,62	m_sflux_oct07_mean
13	1,30	MelFreqBand_de7_skew	0,93	m_spectralflux_mean	33,32	MelFreqBand_6_quart2	34,44	lspFreq_de0_linregerrA
14	1,28	MelFreqBand_de3_kurt	0,92	mfcc_0_quart3	33,26	lspFreq_de0_linregerrA	34,17	MelFreqBand_6_quart2
15	1,28	MelFreqBand_3_upt90	0,91	MelFreqBand_de7_kurt	33,02	MelFreqBand_6_quart3	34,04	mfcc_0_quart2
16	1,27	MelFreqBand_3_kurt	0,89	mfcc_0_quart2	32,99	lspFreq_de0_iqr1_3	33,97	lspFreq_de0_quart1
17	1,26	MelFreqBand_1_kurt	0,88	MelFreqBand_de0_kurt	32,97	MelFreqBand_7_quart1	33,91	m_sflux_oct10_mean
18	1,26	MelFreqBand_6_kurt	0,86	MelFreqBand_de1_kurt	32,92	lspFreq_de0_quart1	33,80	lspFreq_de0_iqr1_2
19	1,26	mfcc_de0_skew	0,84	mfcc_de0_kurt	32,84	mfcc_0_mean	33,59	lspFreq_de0_quart3
20	1,25	MelFreqBand_4_upt90	0,84	m_sflux_oct08_mean	32,04	lspFreq_de0_iqr1_2	32,97	MelFreqBand_6_mean

B Obsah přiloženého DVD

Na přiloženém DVD lze nalézt skripty a funkce prostředí MATLAB, které byly vytvořeny pro potřeby této práce. Také se zde nachází soubory s uloženými vypočtenými parametry, anotacemi a mezivýsledky včetně originálního textu práce v .pdf. DVD neobsahuje audio nahrávky databáze DEAM z důvodu jak velké datové rozměrnosti, tak možnosti volného stažení této databáze z oficiálních webových stránek databáze[1].

```
/ ..... kořenový adresář přiloženého CD
├── MATLAB ..... adresář skriptů a funkcí vytvořených pro tuto práci
│   └── variables ..... adresář s uloženými mezivýsledky a databázemi
└── DP_MER_SmelyP.pdf ..... text diplomové práce
```

[1] http://cvml.unige.ch/databases/DEAM/