

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky  
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2020

Bc. Jana Jelínková



# 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

# ROZPOZNÁNÍ HUDEBNÍHO SLOHU Z ORCHESTRÁLNÍ NAHRÁVKY ZA POMOCI TECHNIK MUSIC INFORMATION RETRIEVAL

RECOGNITION OF MUSIC STYLE FROM ORCHESTRAL RECORDING USING MUSIC INFORMATION  
RETRIEVAL TECHNIQUES

DIPLOMOVÁ PRÁCE  
MASTER'S THESIS

AUTOR PRÁCE  
AUTHOR

Bc. Jana Jelínková

VEDOUCÍ PRÁCE  
SUPERVISOR

Ing. Tomáš Kiska

BRNO 2020



# Diplomová práce

magisterský navazující studijní obor **Audio inženýrství**

Ústav telekomunikací

**Studentka:** Bc. Jana Jelínková

**ID:** 174319

**Ročník:** 2

**Akademický rok:** 2019/20

**NÁZEV TÉMATU:**

## **Rozpoznání hudebního slohu z orchestrální nahrávky za pomoci technik Music Information Retrieval**

**POKyny PRO VYPRACOVÁNÍ:**

V rámci této práce budou shrnuty dosavadní poznatky v oblasti zvané Music Information Retrieval. Konkrétně poznatky věnující se rozpoznávání hudebního slohu. Bude sestavena databáze nahrávek, na které bude toto rozpoznávání testováno. Výstupem diplomové práce bude vyhodnocovací systém využívající strojového učení. Dále budou analyzovány hudební nahrávky z hlediska barvy zvuku, rytmiky a dynamiky a budou vybrány takové hudební parametry, které jsou schopny identifikovat/diferencovat jednotlivé hudební slohy. Úspěšnost systému bude hodnocena pomocí několika metrik.

**DOPORUČENÁ LITERATURA:**

[1] Music similarity and retrieval: an introduction to audio- and web-based strategies. New York, NY: Springer Berlin Heidelberg, 2016. ISBN 9783662497203.

[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:** 3.2.2020

**Termín odevzdání:** 1.6.2020

**Vedoucí práce:** Ing. Tomáš Kiska

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

**UPOZORNĚNÍ:**

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

## **ABSTRAKT**

Pojem klasická hudba obsahuje, stejně jako žánry populární hudby, mnoho žánrových variant. Cílem práce je pomocí strojového učení rozlišit tyto žánrové varianty z orchestrálních nahrávek. Tato práce se zaměřuje na skladby z období od konce 16. století do začátku 20. století, jedná se tedy o baroko, klasicismus a romantismus.

K identifikaci těchto tří hudebních slohů je využita metoda Music Information Retrieval, která spočívá v získávání parametrů z nahrávek a jejich evaluaci. Evaluované parametry jsou využity jako vstupní data pro klasifikátory. Klasifikace probíhá pomocí metod strojového učení, konkrétně se jedná o algoritmy kNN (K-Nearest Neighbor, v češtině k-nejbližší soused), LDA (Linear Discriminant Analysis, lineárně diskriminační analýza), GMM (Gaussian Mixture Models, Gaussovy smíšené modely) a SVM (Support Vector Machines, podpůrné vektory).

V závěru práce je shrnuta dosažená úspěšnost jednotlivých klasifikačních algoritmů. Výsledky ukázaly, že značná odlišnost barokního slohu od zbylých dvou slohů, umožňuje jeho velmi přesnou klasifikaci. Naproti tomu klasicismus vykazuje velkou podobnost s romantismem a jeho identifikace proto dosahuje nižší úspěšnosti. Toto zjištění je ve shodě s předpokladem, který vychází z hudebně-teoretické charakteristiky slohů, která je popsána v první kapitole této práce.

## **KLÍČOVÁ SLOVA**

hudební slohy, baroko, klasicismus, romantismus, orchestrální nahrávky, Music Information Retrieval, extrakce parametrů, minimální redundance, maximální relevance, strojové učení, k-nejbližší soused, lineárně diskriminační analýza, Gaussovy smíšené modely, podpůrné vektory, Support Vector Machines, Gaussian Mixture Models, Linear Discriminant Analysis, K-Nearest Neighbor, Wavelet Scattering, Error-Correcting Output Codes

## ABSTRACT

As all genres of popular music, classical music consists of many different subgenres. The aim of this work is to recognize those subgenres from orchestral recordings. It is focused on the time period from the very end of 16th century to the beginning of 20th century, which means that Baroque era, Classical era and Romantic era are researched. The Music Information Retrieval (MIR) method was used to classify chosen subgenres.

In the first phase of MIR method, parameters were extracted from musical recordings and were evaluated. Only the best parameters were used as input data for machine learning classifiers, to be specific: kNN (K-Nearest Neighbor), LDA (Linear Discriminant Analysis), GMM (Gaussian Mixture Models) and SVM (Support Vector Machines).

In the final chapter, all the best results are summarized. According to the results, there is significant difference between the Baroque era and the other researched eras. This significant difference led to better identification of the Baroque era recordings. On the contrary, Classical era ended up to be relatively similar to Romantic era and therefore all classifiers had less success in identification of recordings from this era. The results are in line with music theory and characteristics of chosen musical eras.

## KEYWORDS

historical periods of music, Baroque era, Classical era, Romantic era, orchestral recordings, Music Information Retrieval method, parameters extraction, minimum Redundancy Maximum Relevance method, machine learning, K-Nearest Neighbor, Linear Discriminant Analysis, Gaussian Mixture Models, Support Vector Machine, Wavelet Scattering, Error-Correcting Output Codes

JELÍNKOVÁ, Jana. *Rozpoznání hudebního slohu z orchestrální nahrávky za pomoci technik Music Information Retrieval*. Brno, 2020, 79 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 „Rozpoznání hudebního slohu z orchestrální nahrávky za pomoci technik Music Information Retrieval“ jsem vypracovala 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 autorka uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma 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 autorky

## PODĚKOVÁNÍ

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

Brno .....

.....

podpis autorky



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 popsáný 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 autorky



EVROPSKÁ UNIE  
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ  
INVESTICE DO VAŠÍ BUDOUCNOSTI



OP Výzkum a vývoj  
pro inovace



# Obsah

Úvod	12
<b>1 Slohy klasické hudby</b>	<b>13</b>
1.1 Identifikace slohů posluchačem . . . . .	13
1.2 Instrumentální hudba před příchodem baroka . . . . .	14
1.3 Baroko . . . . .	15
1.4 Klasicismus . . . . .	19
1.5 Romantismus . . . . .	21
<b>2 Dataset</b>	<b>24</b>
2.1 Datasets klasické hudby . . . . .	24
2.2 Dataset orchestrálních nahrávek . . . . .	26
<b>3 Music Information Retrieval</b>	<b>30</b>
3.1 Klasifikátory . . . . .	31
3.1.1 kNN . . . . .	31
3.1.2 Gaussovy smíšené modely . . . . .	32
3.1.3 Lineární diskriminační analýza . . . . .	33
3.1.4 Support Vector Machines . . . . .	33
3.1.5 Umělé neuronové sítě . . . . .	34
<b>4 Parametry</b>	<b>36</b>
4.1 Parametry nízké úrovně . . . . .	37
4.2 Parametry střední úrovně . . . . .	40
<b>5 Přehled dosavadních prací souvisejících s rozpoznáváním hudebních slohů</b>	<b>42</b>
5.1 Rozpoznávání hudebních žánrů . . . . .	42
<b>6 Implementace</b>	<b>47</b>
6.1 Návrh řešení . . . . .	47
6.2 Výpočet parametrů . . . . .	48
6.3 Výběr parametrů metodou mRMR . . . . .	49
6.4 Optimalizace hyperparametrů . . . . .	52
<b>7 Úspěšnost klasifikátorů</b>	<b>54</b>
7.1 kNN . . . . .	54
7.2 LDA . . . . .	58
7.3 GMM . . . . .	59
7.4 SVM . . . . .	60

7.4.1	Error-Correcting Output Codes a SVM . . . . .	61
7.4.2	Wavelet Scattering . . . . .	62
<b>8</b>	<b>Závěr</b>	<b>65</b>
	<b>Literatura</b>	<b>67</b>
	<b>Seznam symbolů, veličin a zkratk</b>	<b>73</b>
	<b>Seznam příloh</b>	<b>74</b>
<b>A</b>	<b>Seznam parametrů</b>	<b>75</b>
<b>B</b>	<b>Úspěšnost klasifikátorů</b>	<b>77</b>
<b>C</b>	<b>Obsah přiloženého CD</b>	<b>79</b>

# Seznam obrázků

1.1	MP3 tagy: žánry . . . . .	14
1.2	Vývoj hudebního umění . . . . .	15
1.3	Vivaldi: Čtvero ročních období - Léto [42] . . . . .	18
1.4	Bach: Braniborský koncert č. 4, BWV 1049 [4] . . . . .	18
1.5	Mozart: Symfonie č.40 g-moll, K.550 [29] . . . . .	20
1.6	Beethoven: Symfonie č.3 „Eroica“ Es dur, Op. 55 [5] . . . . .	21
1.7	Dvořák: Symfonie č.9, Op. 95 [9] . . . . .	23
2.1	Cross-Era dataset . . . . .	25
2.2	Cross-Composer dataset . . . . .	25
2.3	MusicNet dataset . . . . .	26
2.4	Vlastní dataset orchestrálních nahrávek . . . . .	27
2.5	Dataset: baroko . . . . .	27
2.6	Dataset: klasicismus . . . . .	28
2.7	Dataset: romantismus . . . . .	28
3.1	Metoda MIR . . . . .	30
3.2	Hranice třídy pro různá k . . . . .	32
3.3	Rozdělení prostoru nadrovinou na dvě třídy . . . . .	34
3.4	Schéma neuronu . . . . .	35
4.1	Proces získání parametrů nízké úrovně . . . . .	38
6.1	Návrh implementace . . . . .	48
6.2	Výpovědní hodnota všech parametrů . . . . .	50
6.3	Závislost přesnosti klasifikátoru SVM na počtu parametrů . . . . .	51
6.4	Závislost přesnosti klasifikátoru kNN na počtu parametrů . . . . .	51
6.5	Závislost přesnosti klasifikátoru kNN na počtu parametrů . . . . .	52
6.6	Příklad výsledků optimalizace hyperparametrů pro klasifikátor LDA . . . . .	53
7.1	Závislost přesnosti klasifikátoru kNN na vzdálenostní funkci . . . . .	54
7.2	Závislost přesnosti klasifikátoru kNN na vzdálenostní funkci . . . . .	55
7.3	Závislost přesnosti klasifikátoru kNN na počtu sousedů . . . . .	55
7.4	Úspěšnost klasifikátoru kNN při řešení binárního problému . . . . .	57
7.5	Úspěšnost klasifikátoru kNN při řešení binárního problému . . . . .	57
7.6	Celková úspěšnost klasifikátoru kNN . . . . .	58
7.7	Úspěšnost klasifikátoru GMM v závislosti na počtu parametrů . . . . .	59
7.8	Závislost úspěšnosti klasifikátoru ECOC SVM na počtu parametrů . . . . .	62
7.9	Klasifikace s využitím metody Wavelet Scattering . . . . .	63
7.10	Získávání parametrů metodou Wavelet Scattering . . . . .	63
7.11	Výsledky klasifikátoru SVM/Wavelet Scattering - binární problém . . . . .	64

# Seznam tabulek

5.1	Výzkumy na poli klasické hudby . . . . .	46
7.1	Úspěšnost klasifikátoru kNN při určení jednoho slohu z celého datasetu . .	56
7.2	Úspěšnost klasifikátoru kNN při řešení sloh X nebo Y . . . . .	56
7.3	Úspěšnost klasifikátoru LDA v závislosti na počtu parametrů . . . . .	59
7.4	Úspěšnost klasifikátoru GMM v závislosti na počtu parametrů . . . . .	60
7.5	Úspěšnost klasifikátoru SVM v závislosti na počtu parametrů . . . . .	60
7.6	Úspěšnost klasifikátoru SVM v závislosti na počtu parametrů . . . . .	61
A.1	Seznam parametrů vybraných metodou mRMR . . . . .	75
B.1	Celkové výsledky . . . . .	77

# Úvod

Tato práce se zabývá analýzou parametrů orchestrálních nahrávek za účelem rozpoznání hudebního slohu nebo-li historické epochy, ve které interpretovaná skladba vznikla. K tomuto účelu je využita metoda Music Information Retrieval (dále MIR), která zahrnuje algoritmy strojového učení, popř. umělou neuronovou síť. Z historického hlediska se práce zabývá hudebním uměním v časovém rozsahu od konce 16. století do začátku 20. století a zahrnuje tři hlavní hudební slohy, a to baroko, klasicismus a romantismus. Jak již název práce napovídá, jedná se o analýzu orchestrálních nahrávek. Všechny analyzované nahrávky jsou tedy orchestrálního charakteru, v případě baroka se jedná o hudební tělesa s větším obsazením. Nejedná se tedy o sólové koncerty s orchestrem v doprovodné funkci ani o vokálně-instrumentální díla.

V první fázi řešení problému je nutné vytvořit databázi vhodných nahrávek, tzv. dataset. Jelikož metoda zahrnuje algoritmy strojového učení, které velmi závisí na kvalitě dat, od datasetu se značně odvíjí kvalita výsledku.

V dalším kroku jsou ze všech nahrávek získány parametry pomocí metody mRMR. Tyto parametry jsou následně využity jako vstupní data pro algoritmy strojového učení. Tato práce se zabývá klasifikací hudebních slohů pomocí algoritmů kNN, LDA, GMM a SVM. Přesnost modelů vytvořených těmito klasifikátory se testuje na testovací části datasetu. Výsledky testů všech klasifikátorů jsou uvedeny v tabulce B.1.

# 1 Slohy klasické hudby

Během dějin prošlo hudební umění dlouhým vývojem a mnoha změnami. Tyto změny jsou úzce navázány na změny ve společnosti. Hudební umění tedy odpovídá momentálnímu stavu společnosti v dané epoše dějin. Je tedy ovlivňováno vlivy kulturními, náboženskými, politickými a vývojem vědy a techniky. Tato práce se zabývá třemi významnými obdobími hudebních dějin, a to barokem, klasicismem a romantismem. Cílem je rozpoznat tyto hudební epochy z orchestrálních nahrávek na základě zvukových odlišností, kterými je každá z těchto epoch typická. Zvukové odlišnosti mezi epochami vznikaly z několika důvodů. Pro každé období byly typické skladatelské postupy a formy výsledných skladeb. V závislosti na formách a určení skladby se lišilo obsazení orchestru a tudíž i zvuk, který orchestr produkoval. Během doby rovněž docházelo k vývoji nástrojů a způsobů (neboli technik) hry. O rozdílech mezi jednotlivými epochami bude detailněji pojednávat tato kapitola.

## 1.1 Identifikace slohů posluchačem

Školený posluchač dokáže při poslechu nahrávky určit sloh, ve kterém byla skladba napsána i přes to, že nezná jejího autora. Při identifikaci slohu vychází posluchači z několika vlastností nahrávky. Každý sloh má charakteristické obsazení orchestru a používá charakteristické nástroje. Pro baroko je to například cembalo – slyší-li tedy posluchač v nahrávce cembalo, pravděpodobně zařadí nahrávku do baroka či raného klasicismu. Pro romantismus jsou typická velká obsazení orchestru. Pokud tedy posluchač slyší orchestr s velkým obsazením (melodické perkusní nástroje, harfa apod.), patrně bude nahrávku zařazovat do období romantismu. Krom obsazení orchestru se bude posluchač orientovat pomocí toho, jak skladatel pracuje s tématy a motivy, což může také úzce souviset s celkovou formou skladby. Pokud se téma opakuje ve všech hlasech, posluchač určí formu skladby jako fugu a dílo tak zařadí do baroka. K identifikaci žánru pomáhá také poslech práce s dynamikou, četnost a styl ornamentiky (nátryl, obal atd.), práce s tóninami a modulacemi, hudební interpretace apod. K přesnému určení slohu tedy potřebuje posluchač vědomosti a ideálně také poslechovou zkušenost. Zatímco pro školeného posluchače není určení slohu z nahrávky složitým úkolem, posluchač bez vědomostí a zkušeností má zpravidla s určením problém. Jelikož se dá široká veřejnost považovat za neškolené posluchače má bližší kategorizace klasické hudby význam.

Například, běžná nahrávka formátu MP3 nese digitální informaci neboli tag. V tagu je krom autora, interpreta a roku vydání často obsažen i žánr. Žánry populární hudby se zde dělí poměrně detailně. Pokud se například podíváme na tagy uváděné u metalových nahrávek, nalezneme několik běžně užívaných subžánrů (black, death, heavy, progressive,

power, symphonic, thrash atd.). U nahrávek klasické hudby je ve většině případů uváděn pouze tag CLASSICAL. Jednotlivá díla klasické hudby jsou přitom od sebe často vzdálena po všech stránkách mnohem více, než například jednotlivé subžánry metalu. Tagy se k nahrávkám přidávají především proto, aby bylo usnadněno jejich řazení v digitálních zařízeních (PC, MP3 přehrávače apod.), a tím se zjednodušilo jejich vyhledávání. Takže zatímco potenciální neškolený koncový uživatel bude schopen bez problému vyhledat nahrávky subžánru ACID ELECTRO, nahrávky spadající do klasicistního slohu bude hledat mnohem déle a mnohem obtížněji. Částečnou pomocí jsou v tomto streamovací služby typu IDAGIO, kde je možné nahrávky klasické hudby vyhledávat i v detailních kategoriích (např. barokní fuga).

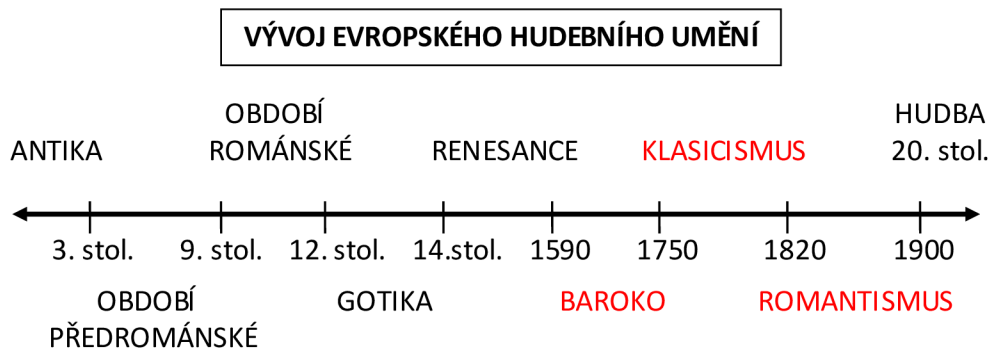


Obr. 1.1: MP3 tagy: žánry

## 1.2 Instrumentální hudba před příchodem baroka

Do starších dějin instrumentální hudby tato práce nezasahuje hned z několika důvodů. Jedním z nich je nedostatek materiálu pro tvorbu vhodného datasetu, a to kvůli relativně malému počtu dochovaných skladeb. Nejvíce dochovaných skladeb je duchovního a liturgického charakteru (písňe, hry, chorály), přičemž se jedná o formy zpívané. Formy instru-

mentální, tedy převážně světská hudba určená např. k tanci, vznikly často improvizací a nejsou tedy zaznamenány. Dalším problémem je nejednotné uskupení větších ansámbků. O prvních uskupeních orchestrálního typu začínáme hovořit až v průběhu 17. stol. Do té doby se jednalo o ansámby hudebníků při hradech mocných velmožů (od 14. století). Tyto ansámby měly typicky různý počet hráčů a různě pestré obsazení. V případě renesance a starších epoch tedy nelze hovořit o orchestrálních uskupeních [38].



Obr. 1.2: Vývoj hudebního umění

### 1.3 Baroko

Země západní Evropy bohatnou díky nově získaným koloniím. Existuje tedy velký počet patronů, kteří jsou schopni na svém dvoře vydržovat ansámby o velkém počtu muzikantů. Začínáme tedy mluvit o prvních orchestrech. V závislosti na tomto faktu, začínají vznikat také skladby pro větší počet muzikantů, ve většině případů na objednávku. Období baroka je i počátkem opery.

Z hlediska tonality je baroko prvním slohem, ve kterém se prosadilo dur-mollové cítění. Do té doby využívala hudba především církevních tónin neboli modů. S užíváním modů úzce souvisí polymelodické smýšlení, v jehož duchu se nesla tvorba předchozích epoch. Přejchod na dur-mollovou tonalitu tedy předestřel nástup harmonického smýšlení. Dalším stupněm ve vývoji tonality byl rozvoj melodicko-harmonického smýšlení, které vychází z principu doprovázeného jednohlasu. Zde je nutné podotknout, že nástup nového smýšlení neznamenal okamžité vymizení prvků, které v hudbě panovaly v předchozích érách. Typicky se jedná především o vícehlas neboli polyfonii. Barokní polyfonie je však na rozdíl od té renesanční harmonicky podmíněná. V baroku tedy roste význam vztahů jednotlivých tónů k tónu centrálnímu a význam harmonických funkcí (tónika, subdominanta, dominanta). Díky střídání harmonických funkcí lze nově ve skladbách docílit střídání „klidu“ a „napětí“, a to i v instrumentálních skladbách bez zpívaného partu. Pociť napětí se typicky vyjadřuje čtyřzvuky (např. dominantní nebo zmenšené septakordy).



Nově se začínají objevovat i modulační postupy, které svou oblíbeností kladou větší nároky na ladění, a tím i na technické zdokonalování nástrojů. Baroko je rovněž obdobím, kdy se velmi ctí umění hráčské improvizace, a to mnohdy více než vlastní skladatelova práce. Běžně tedy docházelo k tomu, že hráči během hry dotvářeli skladbu bohatou ornamentikou. Tato praxe časem zanikla, jelikož přehnaně bohatá ornamentika narušovala charakter díla a měnila záměr autora. Pro baroko je rovněž typické instrumentační umění doprovodu – doprovodný part byl vypsán pouze basovými tóny a hráči na základě těchto tónů sami dotvořili finální podobu doprovodu. Pro potřeby novodobé interpretace bývá doprovodný part vypsán, k improvizaci tedy ve většině případů nedochází. Barokní doprovod se v praxi nazývá basso continuo nebo generálbas. Ve větších ansámblech tvořily tuto složku většinou varhany, cembalo, loutna nebo theorba či různé kombinace těchto nástrojů. Melodickou linku doprovodu potom podtrhoval fagot, violoncello nebo viola da gamba. Oproti doprovodnému bassu continuu stojí v baroku pouze melodie ve vyšším hlasu. V homofonních skladbách jsou tedy střední hlasy v roli harmonické výplně a jsou oproti doprovodu a melodii značně oslabeny. Kostrou barokní skladby je pouze basso continuo a melodie. Výjimkou je specifická barokní forma – fuga.

Další novinkou, kterou s sebou barokní sloh přinesl a zůstala zachována dodnes, je užití taktů obsahujících přízvučné a nepřízvučné doby. Doprovodný hlas tedy krom melodické a harmonické funkce nese mnohdy ještě metrický pulz skladby. Popřípadě se tento metrický pulz promítá i do jiných hlasů, tím je zachován typický motorický chod skladeb. Zde stojí za zmínku, že pulz barokních skladeb bývá narušován agogikou spíše zřídka.

Baroko se rovněž od dalších zkoumaných slohů liší vnímáním a užíváním tempa a dynamiky. Tempové označení skladeb se začalo užívat od začátku 17. století, ale nesloužilo k označení tempa tak, jak jej známe dnes. Tempovým označením byl dán celkový charakter skladby. Dynamika byla ve skladbách užívána na velkých plochách. Tento způsob dynamických proměn vycházel z instrumentace. Skladatelé jednotlivé skupiny nástrojů neslučovali, naopak je často využívali zvlášť. Část skladby hraná pouze smyčcovými nástroji tedy přirozeně vykazuje jinou dynamiku než další část hraná nástroji žesťovými. Dynamické změny typu crescendo nebo diminuendo užívají barokní skladatelé pouze v rámci jednoho, déle drženého tónu [39].

## Barokní formy

Jak již bylo řečeno, forma skladby může napovědět posluchači, o jaký hudební sloh se jedná. Pro digitální analýzu forma prakticky ztrácí význam, jelikož z každé skladby analyzujeme pouze úsek o délce třiceti sekund. Pro ucelení přehledu jednotlivých slohů tedy uvádím pouze v krátkosti.

Velký rozvoj zaznamenaly vokální či vokálně-instrumentální formy. Do období baroka se například datuje vznik opery a ustálení jejích částí, a to árie, recitativu a instrumentální

operní přede hry. Z instrumentální tvorby to byla již zmiňovaná fuga, passacaglia, ciacona, sólový koncert, concerto grosso a jiné drobnější útvary, například preludia a toccaty [49].

## Obsazení orchestru

Obsazení orchestru má přímý vliv na frekvenční složení nahrávky. V barokním období se obsazení často liší skladba od skladby, a to mnohem více než u zbylých dvou slohů. Všeobecně ale platí, že v nahrávkách identifikujeme nástroj nebo skupinu nástrojů v úloze bassa continua a nástroje či skupiny nástrojů střídající se ve hře melodie. Obecně má také barokní orchestr menší počet hráčů než klasicistní nebo romantický. Jako příklad uvádím obsazení Bachova Braniborského koncertu č. 4 a Vivaldiho Čtvero ročních období, část Léto.

V případě Vivaldiho úplně chybí sekce dechových nástrojů, a to jak dřevěných, tak i žesťových. V případě Bacha jsou sice dřevěné dechové nástroje zastoupeny, ale pouze flétnami. Dechové nástroje jiných barev či nižších frekvenčních poloh opět chybí. Základem zvuku barokního orchestru jsou obecně smyčce. Počet hráčů některých smyčcových partů se může lišit, ale ve standardním obsazení hraje obě uvedené skladby pouze 15 až 20 hráčů, někdy i méně. V obou případech je také v partituře uveden part continua, který v této podobě z partitury v následujících érách zmizel. Jak již bylo zmíněno, byl hrán především nástroji schopnými vícehlasé hry (tedy varhany, cembalo, popřípadě drnkací nástroje) [39].

## Hudební nástroje

Hudební nástroje užívané v baroku jsou tak odlišné od moderních nástrojů, že dodnes tvoří zvláštní skupinu. Během barokního období prodělaly například zásadní změnu smyčce. Postupně se přestaly používat smyčce violového typu a byly nahrazeny smyčcovými nástroji houslového typu s jasnějším a pevnějším tónem a většími dynamickými schopnostmi. Barokní hudba rovněž užívala specifických dechových, drnkacích i bicích nástrojů. V dnešní praxi je možné se setkat s interpretací barokní hudby na moderní nástroje a zároveň také s interpretací na nástroje barokní. Pro posluchače to znamená především rozdíl v barvě celkového zvuku [39].

Durata: min.10

# CONCERTO in Sol minore

1

per Violino, Archi e Organo (o Cembalo)

## L'Estate

Da "Il cimento dell'armonia e dell'invenzione,,

a cura di  
Gian Francesco Malipiero

F. I n° 23

Antonio Vivaldi  
(1678-1741)

LANGUIDEZZA PER IL CALDO

Sotto dura stagion dal sole accesa Langue l'huom, langue'l gregge, ed arde il Pino;\*)

Allegro non molto

Violino principale  
pp

Violini I.  
pp

Violini II.  
pp

Viola  
pp

Violoncelli  
pp

Contrabbassi

Organo (o Cembalo)  
pp

Obr. 1.3: Vivaldi: Čtvero ročních období - Léto [42]

## CONCERTO IV.

J. S. Bach.

Allegro.

Violino principale.

Flauto Iº

Flauto IIº

Violino Iº di ripieno.

Violino IIº di ripieno.

Viola di ripieno.

Violoncello.

Violone.

Continuo.

Obr. 1.4: Bach: Braniborský koncert č. 4, BWV 1049 [4]

## 1.4 Klasicismus

Klasicismus se postupně vyvinul z barokního slohu. Polyfonie pozvolna ustupuje, faktura skladeb je průzračnější, což vede k menší vypjatosti; užívá se především durových tónin. Nejvíce se užívají tóniny s nejvýše dvěma křížky nebo dvěma bé. Skladby mají homofonní charakter, tedy obsahují jednu vedoucí melodickou linku a té jsou ostatní hlasy podřízeny a plní doprovodnou funkci. Melodie je v klasicismu nadřazena nad ostatní složky. Základní myšlenkou je princip periodicity a symetričnosti, který se projevuje i v nejmenších stavebních jednotkách skladby, tedy v motivech. V praxi to znamená, že vnitřní stavba vět je symetrická. Základem je osmitaktová perioda dělená na dvě poloviny. Tento základ však není užíván naprosto striktně.

Posun nastal i v oblasti rytmu. Nově se užívají synkopy, trioly, tečkovaný rytmus či lombardský rytmus. Rytmičky nejpestřejší je hlas nesoucí melodii, doprovodné hlasy jsou, co do pestrosti rytmů, střídmější. Tempových výrazů se užívá široká škála a v oblibě jsou především tempa rychlejší.

Díky novému posunu ve vývoji orchestru i technickému zdokonalování nástrojů začínají terasovitou barokní dynamiku nahrazovat pozvolné změny typu crescendo a diminuendo. Harmonie se zjednodušuje, často jsou užívány jen základní funkce (tónika, subdominanta, dominantanta). Největšího užití zaznamenal vztah tónika - dominantanta či vztah tónika - paralelní durová tónina. K oddělení jednotlivých vět se využívá tzv. autentická kadence, tj. dominantanta - tónika. Uvnitř vět se používá tonální kontrast, nicméně věta vždy začíná a končí v hlavní tónině.

Vrcholná fáze klasicismu je nejlépe reprezentovaná tvorbou Josepha Haydna, Wolfganda A. Mozarta a Ludwiga van Beethoven. Přičemž v tvorbě posledního jmenovaného se projevuje tzv. expresivní styl. Jedná se o tedy o dramatičtější hudbu s větším zastoupením mollových tónin [39].

### Klasicistní formy

Mnohé z barokních forem v klasicismu zanikají. Z tanečních forem přežívá menuet, který se stává součástí sonátového cyklu. V orchestrální tvorbě jednoznačně převládají symfonie dělené na věty, a to buďto v italském stylu (rychlá, pomalá, rychlá) nebo ve stylu francouzském (pomalá, rychlá, pomalá). První věta často obsahuje sonátovou formu, druhá věta prostou dvoudílnou či třídílnou formu a třetí věta menuet. Symfonie o čtyřech větách byly spíše výjimkou. Další orchestrální formou byla koncertantní symfonie, tj. útvar pro dva a více sólových nástrojů s orchestrem. Nově vzniklé klasicistní druhy jsou například divertimento, kasace či serenáda. Skládá se také ve formě rondové a variační.[49, 39]

## Klasicistní orchestr

Během období klasicismu se vyhraňuje nový ideál orchestrálního zvuku. Střední hlasy se svou důležitostí dorovnávají k basu a všechny se zároveň přizpůsobují melodické linii ve vrchním hlase. Základem orchestru je smyčcová skupina, která se ustálila na čtyřech hlosech, a to 1. housle, 2. housle, violy a bas. Bas tvořily kontrabasy a violoncella hrající totožný part v oktávách. Ke smyčcovému základu se přidávají dechové nástroje, zpravidla v párech. Nejprve lesní rohy a hoboje, v polovině 18. století střídá zobcovou flétnu flétna příčná. Začíná se používat také nový nástroj - klarinet. V pozdním klasicismu se stávají stabilními členy orchestru také trubky. Naopak mizí loutny a o něco pomaleji také cembala.

Orchestry se velmi liší počtem členů, v zásadě mají ale více členů než orchestry barokní. Velké orchestry mohly mít až 30 houslí, 20 ostatních smyčcových nástrojů, až 16 dechových nástrojů a tympanů.[39]

Mozart  
Symphony No. 40  
in G minor  
K. 550

*Allegro molto.*

The image shows a page of a musical score for Mozart's Symphony No. 40 in G minor, K. 550. The score is for the woodwind and string sections. The instruments listed are Oboi, Clarinetti in B., Flauto, Oboi, Fagotti, Corno in Balto., Corno in G., Violino I., Violino II., Viola, and Violoncello e Basso. The tempo is marked 'Allegro molto.' The key signature is G minor (two flats) and the time signature is common time (C). The score shows the first few measures of the piece, with the woodwinds and strings playing a rhythmic pattern. The strings are marked with a piano (p) dynamic.

Obr. 1.5: Mozart: Symfonie č.40 g-moll, K.550 [29]

# Symphony No. 3

Eroica

in E $\flat$  Major

Op. 55

*Allegro con brio, o-so.*

Flauti.

Oboi.

Clarinetti in B.

Fagotti.

*Allegro con brio, o-so.*

Corni in Es.

Corno 3<sup>o</sup> in Es.

Trombe in Es.

Timpani in Es.

*Allegro con brio, o-so.*

Violino I.

Violino II.

Viola.

Violoncello e Basso.

Obr. 1.6: Beethoven: Symfonie č.3 „Eroica“ Es dur, Op. 55 [5]

## 1.5 Romantismus

Romantismus se vyznačuje snahou o návrat k přírodě a vyjádření hlubokých emocí. V praxi to znamenalo vzrůstající oblibu fantazijních či tajemných nálad nebo hledání inspirace v lidové tvorbě. Melodika skladeb se stává složitější a pracuje s větším tónovým rozsahem. V rytmu se kromě četných rytmických ozdob užívá i nepravidelný pohyb. Vznikají tak rytmicky komplikované útvary i na úrovni taktů. Harmonie více pracuje s akordy (mimotonální dominanty, alterované akordy), častěji využívá modulace i stupňových chromatických postupů. Rychlost harmonického pohybu se mění v závislosti na

vypětí skladby. Disonantní souzvuky se nově rozvádí do konsonantních souzvuků oddáleně a nebo se nerozvádí vůbec. Postupně se skladatelé dostávají na hranice tonální harmonie (R. Wagner). Tvorba romantická tak přechází na jedné straně do tvorby atonální a na druhé straně do tvorby impresionistické. Instrumentace bohatě pracuje s barevnými možnostmi nástroje.

Na romantická díla byl obecně požadavek, aby nesla ideovou výpověď. Vznikaly proto nové typy skladeb, například symfonické básně. Jednalo se o jednovětou formu s mimohudebním námětem. Mnoho romantických autorů tíhlo k monumentálnosti a okázalosti, tomu odpovídalo i složení orchestru [38, 39].

## Formy a žánry

Období romantismu je charakteristické rozmachem programní hudby, tj. hudby, která má mimohudební podtext. Vedle nové symfonické básně se užívá forma čtyřvěté symfonie. Na poli instrumentálně-vokální tvorby zůstává vedoucí formou opera. Provádí se také menší vokální formy, např. operety a v oblasti duchovní hudby oratoria [49].

## Romantický orchestr

Skladatelé mají v oblibě zvuk dechových nástrojů, a to jak dřevěných, tak žestových. Zejména dřevěné dechové nástroje dostávají v orchestru stěžejní úlohu. Nové, rozšířené využití mají i bicí nástroje a harfa. Žestová sekce bývá doplňována novými nástroji, např. Wagnerovými tubami. Ve smyčcových nástrojích je upřednostňován temnější zvuk viol a violoncell [39].

Adagio. M.M. ♩ = 126.

Flauti.

Oboi.

Clarineti in A.

Fagotti.

Corni I, II. in E.

Corni III, IV. in C.

Trombe in E.

Tromboni I, II.

Trombone basso.

Tympani A, B, H.

Adagio. M.M. ♩ = 126.

Violino I.

Violino II.

Viola.

Violoncello.

Contrabasso.

The image displays a page of a musical score for Dvořák's Symphony No. 9, Op. 95. The tempo is marked 'Adagio' with a metronome marking of 126. The score is divided into two systems. The first system includes woodwind instruments: Flauti (Flutes), Oboi (Oboes), Clarineti in A (Clarinets in A), Fagotti (Bassoons), Corni I, II. in E (Trumpets in E), Corni III, IV. in C (Trumpets in C), Trombe in E (Trumpets in E), Tromboni I, II. (Trumpets in C), Trombone basso (Trombone), and Tympani A, B, H (Timpani). The second system includes string instruments: Violino I (Violin I), Violino II (Violin II), Viola (Viola), Violoncello (Cello), and Contrabasso (Double Bass). The woodwinds and strings are playing a melodic line with dynamic markings such as *p*, *fz*, *dim. p*, and *pp*. The strings are playing a rhythmic accompaniment.

Obr. 1.7: Dvořák: Symfonie č.9, Op. 95 [9]



## 2 Dataset

Jelikož metoda MIR využívá algoritmů strojového učení či neuronových sítí, jsou kvalitní data nutným základem. Dataset se obvykle dělí na dvě množiny:

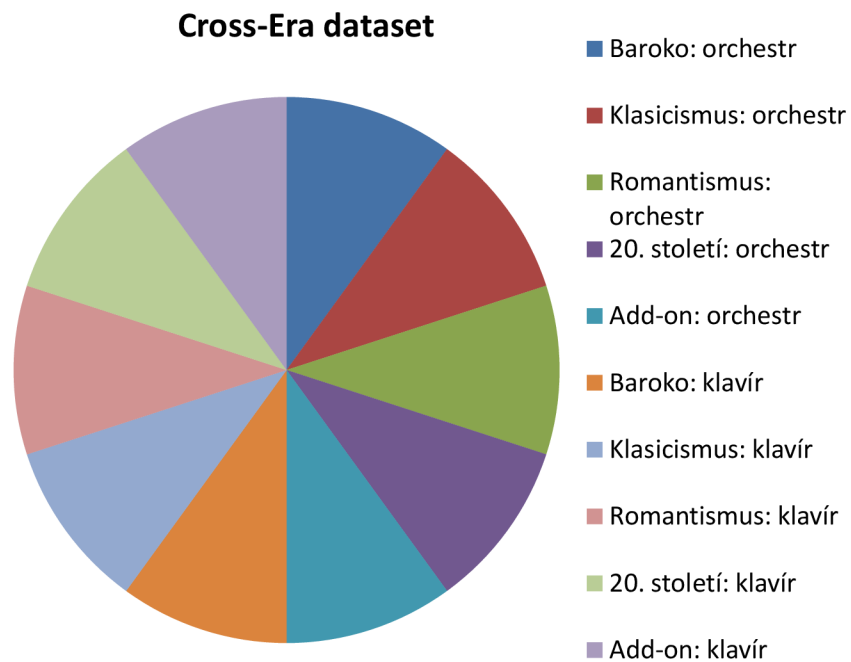
- trénovací data
- testovací data.

Trénovací data tvoří 60 % až 80 % z celého datasetu. Výstup této množiny je předem známý a zvolený model se tzv. „učí“ vztah mezi daty a výstupem [15]. Testovací data tvoří zbytek datasetu. Tato data jsou během trénovací fáze skrytá a mají neznámý výstup. Na této množině se daný model otestuje a pokud určí výstup správně, je připraven k použití.

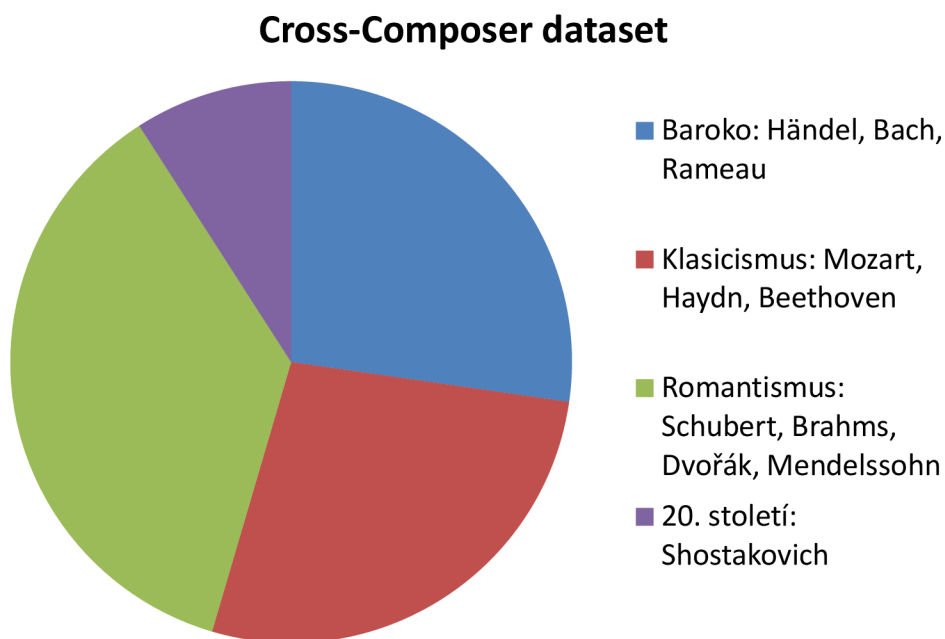
### 2.1 Datasetsy klasické hudby

Podobných výzkumů na poli klasické hudby zatím neproběhlo mnoho. Existující datasety jsou často přizpůsobeny typu výzkumu a zároveň, z důvodu možného porušení autorských práv, nejsou veřejně přístupné. Ismir.net uvádí databázi několika set existujících datasetů vhodných pro analýzu pomocí metody MIR. Z oblasti klasické hudby je zde například veřejně přístupný dataset Cross-Composer, který obsahuje 1100 děl od 11 autorů. Aby bylo možné jej distribuovat veřejně, neobsahuje dataset žádné audio nahrávky, nýbrž chromagramy vytvořené pomocí algoritmu NNLS chroma dostupného volně jako součást pluginu VAMP. Jak už napovídá název, je dataset určen především pro identifikaci autora z audio nahrávky. Dataset obsahuje nenormalizované nahrávky orchestrální, vokální i sólové [2, 15].

Další podobný dataset je Cross-Era obsahující 2000 děl ze 4 různých období hudebních dějin, konkrétně je zastoupeno baroko, klasicismus, romantismus a hudba 20. století. Aby autoři datasetu předešli porušení autorských práv, jsou všechna data opět ve formě chromagramů. Každá epocha je zastoupena 200 orchestrálními nahrávkami a 200 klavírními. Další nahrávky z dané epochy obsahují také data nazvaná „add-on“. Jedná se o díla autorů ze všech epoch, která mají zacelit mezery mezi jednotlivými epochami. V každé epoše je vždy minimálně 5 autorů ze 3 různých zemí. Autoři datasetu upozorňují na fakt, že pokud se jedná o orchestrální nahrávku, je vždy instrumentální a zároveň se nejedná o koncerty sólového nástroje, kde by byl orchestr v doprovodné funkci. Za zmínku také stojí, že klavírní skladby zařazené do barokní éry (a ještě také část klasicistních klavírních skladeb) jsou interpretovány na moderní klavír, nikoli na cembalo. Přestože to z historického hlediska není ideální, autoři datasetu tak zajistili, aby se model „neučil“ typickou barvu cembala a nevznikala tak chyba. Z datasetu byly také vyřazeny skladby Beethovena a Schuberta, jelikož se, dle autorů datasetu, nacházejí na rozhraní dvou epoch a jejich zařazení je tak nejasné [3].



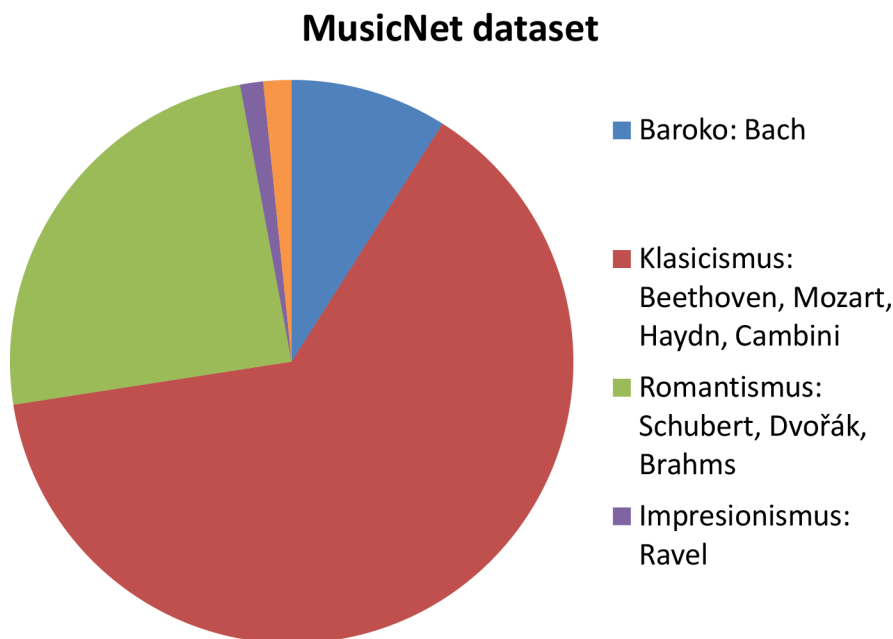
Obr. 2.1: Cross-Era dataset



Obr. 2.2: Cross-Composer dataset

Volně šiřitelné audio nahrávky obsahuje dataset MusicNet. Jedná se o 330 skladeb deseti autorů. Skladby jsou netradičně celé, ne jen formou vzorků. Dataset tedy obsahuje 2 048 minut hudby. Především se jedná o skladby pro sólové nástroje nebo pro menší

uskupení do velikosti kvintetu. Dataset je tedy ideální například na detekci nástroje nebo skladatele. Na detekci hudebního slohu se příliš nehodí, jelikož jednotlivé slohy nejsou zastoupeny stejně početně a v případě skladeb Faurého je sporné i zařazení do určitého slohu [32].

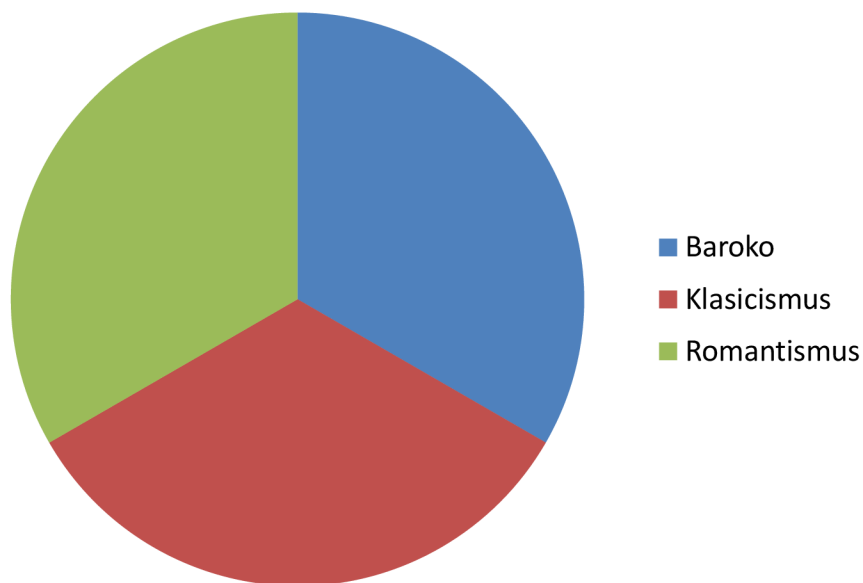


Obr. 2.3: MusicNet dataset

## 2.2 Dataset orchestrálních nahrávek

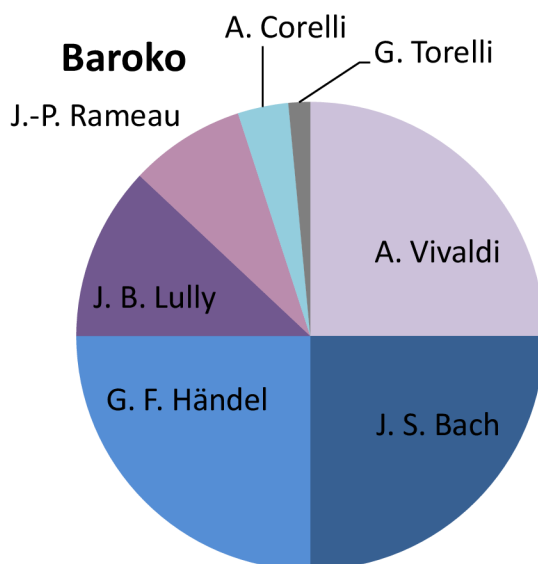
Protože žádný z již existujících datasetů není vhodný pro řešení problému, kterým se zabývá tato práce, byl pro účely rozpoznání slohů z orchestrálních nahrávek vytvořen nový dataset. Všechny nahrávky jsou ve formátu MP3 a byly pomocí batch processingu zkráceny na vzorky o délce 30 s a normalizovány na hladinu -1,5 dB. Dataset obsahuje celkem 600 orchestrálních nahrávek a zastoupeny jsou v něm stejným dílem tři různé slohy.

### Vlastní dataset



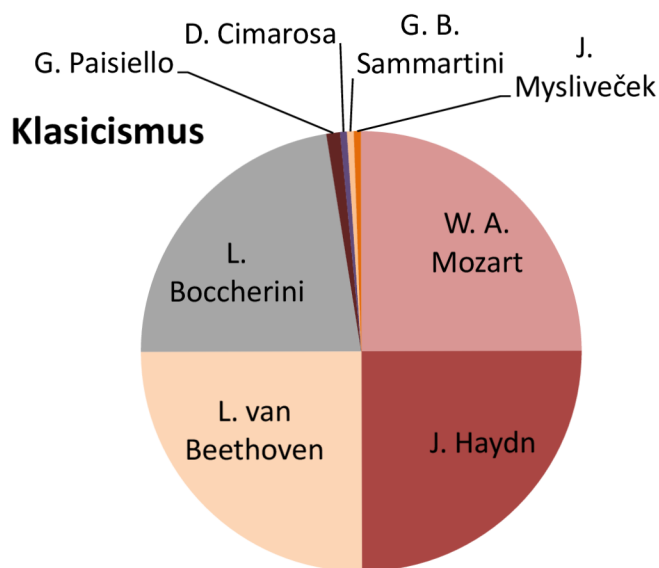
Obr. 2.4: Vlastní dataset orchestrálních nahrávek

Barokní část datasetu tvoří z poloviny zástupci německého baroka G. F. Händel a J. S. Bach, italské baroko zastupuje A. Vivaldi, A. Corelli a G. Torelli, francouzské baroko potom J. B. Lully a J.-P. Rameau. Vesměs se jedná o concerta, concerta grossa, orchestrální suity, přede hry a instrumentální části z oper či oratorií.

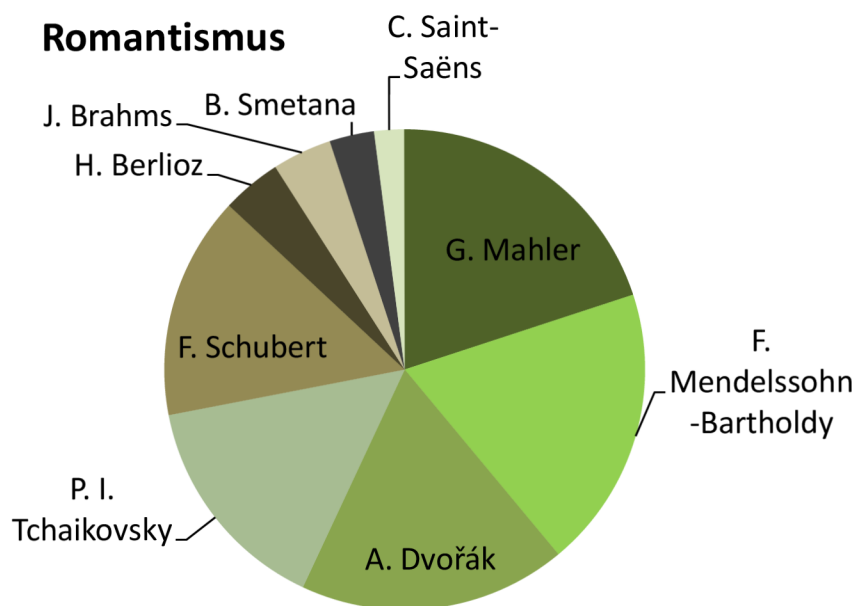


Obr. 2.5: Dataset: baroko

Klasicistní část obsahuje ze tří čtvrtin skladby autorů tzv. první vídeňské školy, to jest W. A. Mozarta, J. Haydna a L. van Beethovena. Styl posledního jmenovaného stále spadá do klasicismu, nicméně je silně individuální a dle hudebních historiků „otvírá dveře romantismu“. Proto bývá někdy pro dosažení lepších výsledků systému z datasetů vyřazován. Formálně se jedná o symfonie a přehry oper.



Obr. 2.6: Dataset: klasicismus



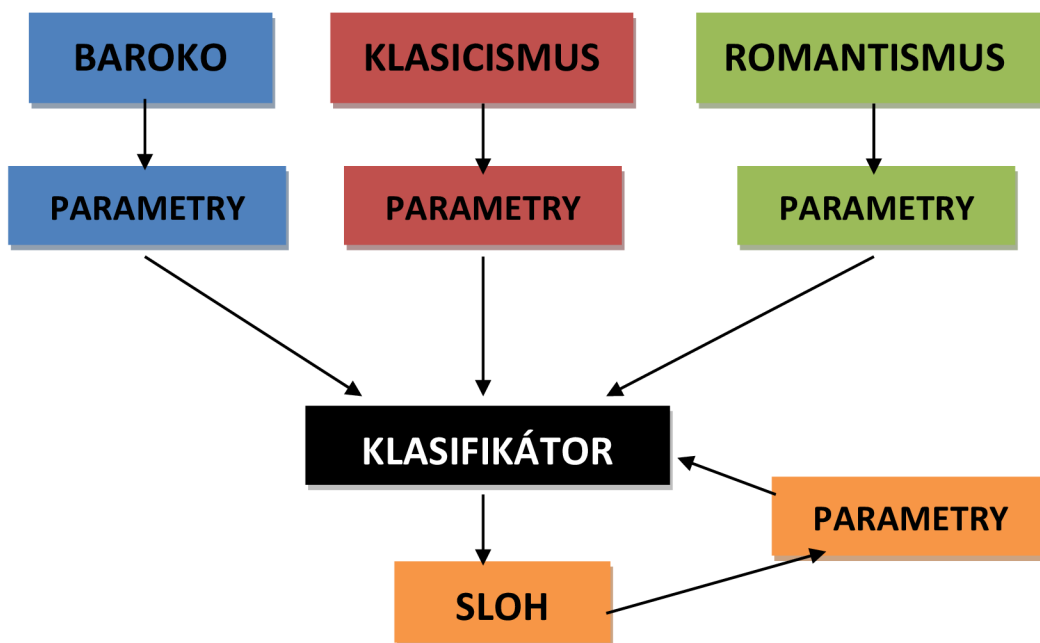
Obr. 2.7: Dataset: romantismus

Romantismus sám o sobě obsahuje několik vývojových etap a na rozdíl od předešlých epoch se na svém konci rozvětňuje do několika naprosto rozdílných stylů. Raná fáze romantismu je zastoupena F. Schubertem a F. Mendelssohnem, novoromantismus zastupuje H. Berlioz, vrcholný romantismus A. Dvořák, B. Smetana, J. Brahms, P. I. Tchaikovsky, C. Saint-Saëns a pozdní romantismus je zastoupen G. Mahlerem. Z formálního hlediska se jedná o symfonické básně, symfonie a přehry z oper (tzv. ouvertury).

### 3 Music Information Retrieval

Metoda Music Information Retrieval (dále MIR) se zabývá získáváním informací z hudebních dat. Jedná se o komplexní metodu zahrnující analýzu dat, extrakci vybraných parametrů a jejich porovnání pomocí algoritmů strojového učení, popřípadě pomocí umělé neuronové sítě. Metoda MIR staví na faktu, že zpracovávaná hudební data vykazují určitou míru podobnosti, která definuje jejich vztah. Každý vzorek je reprezentován hodnotou parametrů. Míru podobnosti zjišťujeme jejich porovnáváním a na základě výsledků potom daný vzorek připadne k určité skupině dat.

Metodu MIR lze využít k řešení široké škály problémů. Typově podobná problematika jako rozpoznání hudebních slohů, je například určení hudebního žánru z nahrávky. V praxi se metoda MIR používá například pro usnadnění vyhledávání v osobních zařízeních určených k přehrávání hudby (iPod, MP3 přehrávače). Streamovací služby využívají tuto metodu k vyhledávání skladeb podobných těm, které daný uživatel poslouchá (Spotify Radio, Pandora, iTunes Radio). V neposlední řadě lze tuto metodu využít v aplikacích identifikujících skladby pomocí nahrávání vzorků či broukání melodie (Shazam, SoundHound)[37].



Obr. 3.1: Metoda MIR

Podle toho jaká data obsahuje zvolený dataset rozlišujeme parametry na obsahové a kontextové. V případě řešení problému pomocí kontextu skladby využíváme data patřící ke skladbě, nikoli skladbu samotnou. Může se jednat o obal CD na kterém skladba

vyšla nebo o hodnocení a recenze na webu. Pokud využíváme k řešení problému obsah skladby, můžeme využít obsah symbolický nebo vlastní nahrávku. Za symbolický obsah bývá považována reprezentace skladby v MIDI, popřípadě partitura. Při extrakci parametrů z vlastní nahrávky pracujeme buď v časové nebo frekvenční doméně. V případě této práce se parametry extrahují z audio dat, pracuje se tedy s obsahem hudební nahrávky.

Jak jde vidět z obrázku, v prvním kroku je nutné každý hudební sloh definovat pomocí typických nahrávek. Každé nahrávce z trénovací části datasetu je přiřazen štítek a jsou z ní extrahovány parametry. Štítky a parametry jsou využívány k natrénování vybraného klasifikátoru. Natrénovaný klasifikátor se otestuje na testovací části datasetu tak, že předpoví slohy obsažených nahrávek [17].

## 3.1 Klasifikátory

Klasifikátor je centrální částí metody MIR. Tato část metody MIR spadá do oblasti strojového učení (Machine Learning), popřípadě do oblasti práce s umělými neuronovými sítěmi (Artificial Neural Network). Nicméně v obou případech jde o užití umělé inteligence (Artificial Intelligence, AI). Úkolem klasifikátoru je na základě daných informací (parametrů, štítků) rozlišit do jaké skupiny bude analyzovaná nahrávka zařazena. V závislosti na tom, zda jsou klasifikátoru poskytnuta specifikovaná data v podobě trénovací množiny, rozlišujeme proces učení na „učení s učitelem“ a „učení bez učitele“. Jak již bylo zmíněno, v případě učení s učitelem je klasifikátoru poskytnuta trénovací množina nahrávek se štítky a známými parametry. Takto natrénovaný klasifikátor se dále zkouší na testovací množině. V případě učení bez učitele trénovací množina poskytnuta není a klasifikátor parametry zjišťuje sám. Na základě odlišnosti či podobnosti hodnot parametrů rozděluje dataset do tříd.

### 3.1.1 kNN

*K-Nearest-Neighbor*, v češtině *k-nejbližší-soused*, je základní klasifikátor pro klasifikaci dat a detekci struktur. Klasifikátor počítá vzdálenost parametrů umístěných v rovině a zakládá se na předpokladu, že parametry instancí stejné třídy budou k sobě blíže. Ke každému bodu v rovině je možné najít  $k$ -nejbližší bod. Tímto způsobem jsou v případě kNN definovány třídy [1, 20]

Vzdálenost mezi vektory parametrů lze spočítat jako *Euklidovskou vzdálenost*:

$$d(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (3.1)$$

kde  $n$  je rozměr vektoru parametrů.

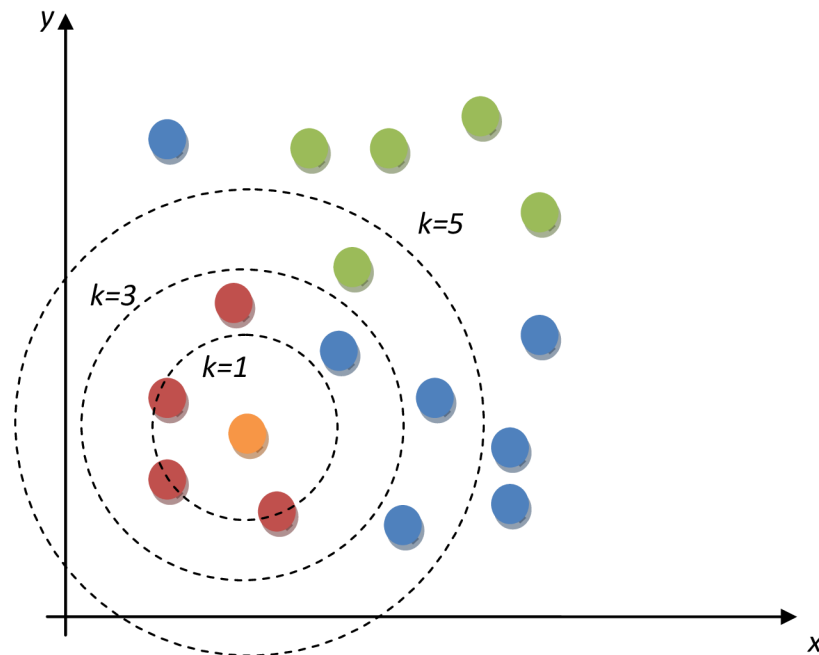
V datasetu s body  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , kde  $x_i$  reprezentuje vektor parametrů



a  $y_i$  reprezentuje cílovou třídu. Pro nový bod  $x$  bude třída určena:

$$nn(x, 1) = y_p, p = \operatorname{argmin}_i \|x - x_i\|^2. \quad (3.2)$$

Vektory parametrů se seřadí podle vypočtené vzdálenosti a poté platí, že  $k = 1$  odpovídá nejbližším sousedním vektorům, a ty tím pádem spadají do stejné třídy jako nový bod  $x$ . Se zvětšujícím se  $k$  se roztahují pomyslné hranice třídy a spadá do ní tedy více vektorů[17, 20].



Obr. 3.2: Hranice třídy pro různá  $k$

### 3.1.2 Gaussovy smíšené modely

*Gaussovy smíšené modely* (Gaussian Mixture Models, GMM) jsou klasifikační metodou, která pracuje s hustotou pravděpodobnosti. Model třídy je odvozen prostřednictvím hustot pravděpodobnosti z vektorů parametrů, které jsou s danou třídou asociovány. GMM jsou charakterizovány počtem  $K$  Gaussových rozdělení. Tato rozdělení jsou definována středními hodnotami  $\mu_k$  a kovariancemi  $\Sigma_k$  (míra lineární závislosti veličin). Každému rozdělení je přidělena váha  $w_k$ , která reprezentuje počet hodnot (vektorů parametrů), které rozdělení obsahuje. Součet vah všech rozdělení musí odpovídat hodnotě 1. Váhy slouží jako indikace pravděpodobnosti, že nový vektor parametrů spadá do jedné ze tříd. Hustota pravděpodobnosti jednotlivých tříd je pak definována takto:

$$p(x|w_k) = \sum_{m=1}^M w_{km} p_{km}(x), \quad (3.3)$$

kde  $p_{km}$  vyjadřuje normální hustotu [6, 17].

Vlastnosti GMM bývají obvykle upravovány metodou maximalizace očekávání (Expectation Maximization, EM). Tato metoda se skládá ze dvou kroků, a to:

- zjištění kvality modelu GMM a posouzení, zda-li je model vhodný pro zpracovávaná data (expectation),
- adaptace modelu na zpracovávaná data, úprava jeho vlastností (maximization).

Po natrénování každý GMM určí, s jakou pravděpodobností spadá nový prvek do jeho třídy. Nový prvek je následně přiřazen do třídy, u které je určena největší pravděpodobnost [17].

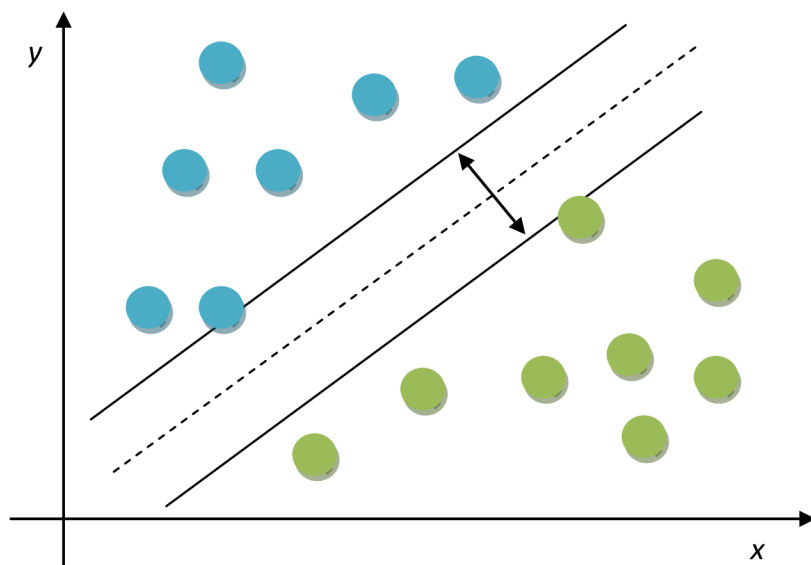
### 3.1.3 Lineární diskriminační analýza

Tato metoda je často používána při řešení rozpoznávání hudebních žánrů. Využívá se také při analýze klasické hudby. Podstatou *lineární diskriminační analýzy* (Linear Discriminant Analysis, LDA) je projekce všech dat do prostoru tak, aby byla snížena korelace mezi parametry. V tomto modifikovaném prostoru se pomocí výpočtů Euklidovských vzdáleností definují třídy. LDA má blízko k jiným lineárním klasifikátorům jako je analýza hlavních komponent, která se využívá k redukci dimenzionality nebo logistická regrese. Na rozdíl od logistické regrese však podporuje jak řešení binárních problémů, tak řešení problémů s více třídami a zůstává stabilní i v případech, kdy jsou jednotlivé třídy jednoznačně odděleny. Počet klasifikátorů odpovídá počtu tříd [20, 46].

### 3.1.4 Support Vector Machines

V češtině se někdy tato metoda označuje jako *podpůrné vektory*, zkratkou SVM. Základním krokem je nalezení nadroviny, která nejlépe rozdělí pomyslný multidimenzionální prostor s hodnotami parametrů do dvou tříd. Nadrovina se umísťuje mezi hodnoty, které jsou od sebe v prostoru nejvíce vzdálené. Pokud nelze tuto nadrovinu najít přímo, používá se heuristické řešení a navýšení počtu dimenzí tak, aby byly jednotlivé třídy lineárně oddělitelné. Při řešení nedělitelnosti tříd se pro převod hodnot do vyšší dimenze používá jádrová transformace (kernel function)[22].

Nevýhodou SVM je, že dokáže hodnoty rozdělit pouze do dvou tříd, tudíž musí být celý proces klasifikace upraven. Nejběžnějším postupem je zredukovat každý krok tak, aby jej šlo řešit rozdělením na pouhé dvě třídy. Výhodou naopak je, že můžeme kontrolovat správnost procesu učení nezávisle na dimenzi řešeného problému [20, 22].



Obr. 3.3: Rozdělení prostoru nadrovinou na dvě třídy

### 3.1.5 Umělé neuronové sítě

Vzorem *umělých neuronových sítí* (Artificial Neural Network, ANN) je chování biologické neuronové sítě. Jedná se tedy o paralelní zpracování dat. Nejmenší jednotkou neuronových sítí je umělý neuron. Všechny neurony jsou propojeny do sítě. Každý neuron má specifické chování (local computation). Běžně se používají různé varianty, některé mají pouze nelineární přenosovou funkci, jiné simulují biologický neuron. Chování neuronu je tedy závislé na jeho naprogramované funkci a na neuronech se kterými je propojen [44]. Neurony jsou rozděleny do vrstev:

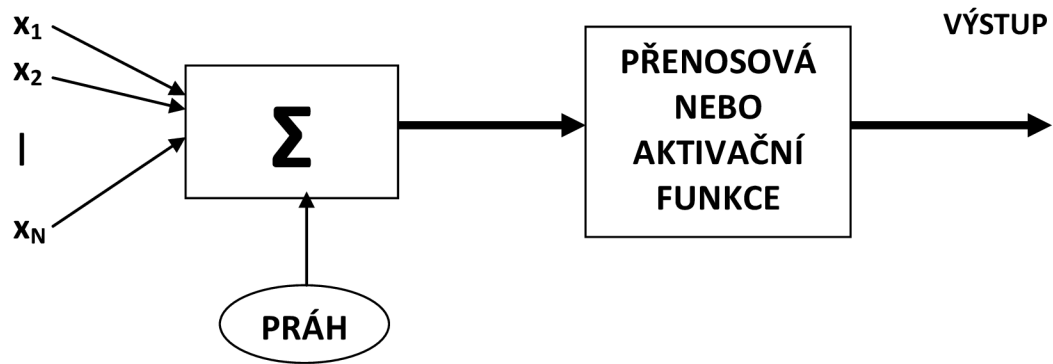
- vstupní vrstva,
- $n$  skrytých vrstev,
- výstupní vrstva.

Počet skrytých vrstev závisí na implementaci. Sítě s menším počtem se dají rychleji natrénovat a jsou ekonomičtější na složitost výpočtu[20, 44].

Každý neuron má tři části:

- synapse – propojení s dalšími neurony, každé propojení je váhováno,
- sumační bod – kombinuje váhované vstupní signály,
- aktivační funkce – vnáší do procesu nelinearity, kontrola výstupu.

V trénovací fázi jsou nastaveny synaptické váhy tak, aby vektory parametrů spadaly do „svých“ tříd. Každá třída má odpovídající neuron, ten je aktivován, pokud je rozhodnuto, že vektor parametrů do dané třídy spadá. Existuje několik variant umělých neuronových sítí. Na poli analýzy hudebního signálu se nejvíce používá vícevrstvý perceptron (Multilayer Perceptron, MLP). Neurony mají více vstupů, ale pouze jeden výstup[20, 45].



Obr. 3.4: Schéma neuronu

## 4 Parametry

Metoda MIR se snaží přiblížit rozhodovacímu procesu reálného posluchače. V ideálním případě by tedy analýza i rozhodovací proces měli nějakým způsobem, krom zřejmých parametrů, obsáhnout také osobu posluchače. Jedná se tedy o velmi náročný proces, jelikož na rozhodnutí posluchače má vliv několik faktorů:

- samotný hudební obsah – rytmus, melodie, harmonie, hudební barva, hlasitost. . .
- hudební kontext – videoklip, grafická úprava alba, umělec. . .
- stav posluchače – momentální nálada, sociální kontext, psychologické aspekty. . .
- vlastnosti posluchače – poslechová zkušenost, znalosti z oblasti hudebního umění, hudební preference, demografie. . .

Jelikož jsou některé z výše zmíněných faktorů do určité míry abstraktní, jsou i parametry rozděleny podle míry abstrakce do tří skupin:

- parametry nízké úrovně (vlastnosti audio signálu, např. energie, počet průchodů nulou, . . .)
- parametry střední úrovně (kombinace více parametrů nízké úrovně s poznatky z vysoké úrovně, popisují základní aspekty zvuku, vyplňují „významovou mezeru“ mezi parametry nízké a vysoké úrovně, např. MFCC, FP, . . .)
- parametry vysoké úrovně (přibližují se vnímání posluchače, rozpoznání motivů, témat, melodie, . . .).

Z hudebního hlediska parametry popisují následující vlastnosti skladby.

- **Dynamika** je intenzita zvuku, která se v čase mění. Skladatelem požadované změny dynamiky jsou zapsané v notovém zápise a interpret je uplatňuje dle dobových tradic a svého estetického cítění. Dále je dynamika dána především hudebním nástrojem, množstvím hrajících nástrojů a zpracováním nahrávky.
- **Výška** vnímaného zvuku je závislá především na frekvenci, ale také na intenzitě a délce. Proto bývá rozlišena tzv. absolutní výška tónu, která je dána frekvencí čistého tónu. Relativní výška, která je dána intervalem a subjektivní výška, která byla vytvořena empiricky a zohledňuje vliv intenzity a délky [41]. Subjektivní výška má jednotku 1 mel.
- **Barva** (témbr) je určena spektrálním složením zvuku, tedy množstvím a silou vyšších harmonických složek a výskytem formantových oblastí. Je závislá rovněž na výšce i dynamice, jelikož se změnou těchto vlastností dochází k přesunu formantových oblastí ve frekvenčním spektru [41]. Každý hudební nástroj je charakteristický svojí barvou, která většinou obsahuje rejstříky, tj. mírné změny barvy v závislosti na výšce hraného tónu.
- **Harmonie** je souzvuk více tónů. V závislosti na frekvenčním poměru znějících tónů vnímá lidské ucho souzvuk jako konsonantní (libozvučný) nebo disonantní (nelibozvučný). Akord je souzvuk tří a více tónů. Shluk více tónů, které spolu těsně

sousedí se nazývá klastr.

- **Rytmus** je dán střídáním delších a kratších not během taktu.
- **Metrum** je střídání přízvučných a nepřízvučných dob během taktu. Stálou součástí skladeb se stalo ke konci baroka.
- **Tempo** je rychlost, kterou je skladba hrána. Často se během skladby mění a je tak závislé na čase. V případě historických hudebních slohů tempo buď není zadáno vůbec nebo je zadáno slovně. Závisí tedy především na interpretech, v případě orchestrů na dirigentovi. V současné hudbě je tempo značeno i v BPM (beats per minute) nebo indikací metronomu (např. čtvrtová nota odpovídá tempu 80).
- **Melodie** je rytmický sled tónů vyjadřující hudební myšlenku. V hudbě identifikujeme motiv – krátký úryvek reprezentující ucelenou myšlenku a téma – delší, výrazná hudební myšlenka. Motiv i téma nebývají výrazné jen melodicky, ale také rytmicky.

## 4.1 Parametry nízké úrovně

Parametry nízké úrovně popisují měřitelné vlastnosti nahrávky a jsou extrahovány přímo z časové nebo frekvenční domény. Daný parametr je počítán pro jeden rámeček. Počet výsledných hodnot tedy odpovídá počtu rámečků. Na konci extrakce dochází ještě ke zpracování získaných hodnot, a to buď statistickou sumarizací (průměr, maximum, medián), kdy dochází k datové ztrátě nebo modelováním distribuce hodnot (Gaussův smíšený model, skrytý Markovův model), kdy k datové ztrátě nedochází.

### Parametry časové domény

*Amplitudová obálka* (Amplitude Envelope, AE) je maximální hodnota amplitudy všech vzorků v jednom rámečku. Jedná se tedy o parametr citlivý na veškeré odchylky hodnot amplitudy. Dalšími parametry jsou *střední hodnota energie signálu* ( $\bar{E}$ ) a *efektivní hodnota energie* (Root-Mean-Square Energy, RMS) signálu popisují intenzitu nahrávky. Spolu s amplitudovou obálkou se tedy vztahují k dynamice. Lze je získat pomocí následujících vzorců:

$$\bar{E}_t = \frac{1}{\tau} \int_0^{\tau} E(t) dt, \quad (4.1)$$

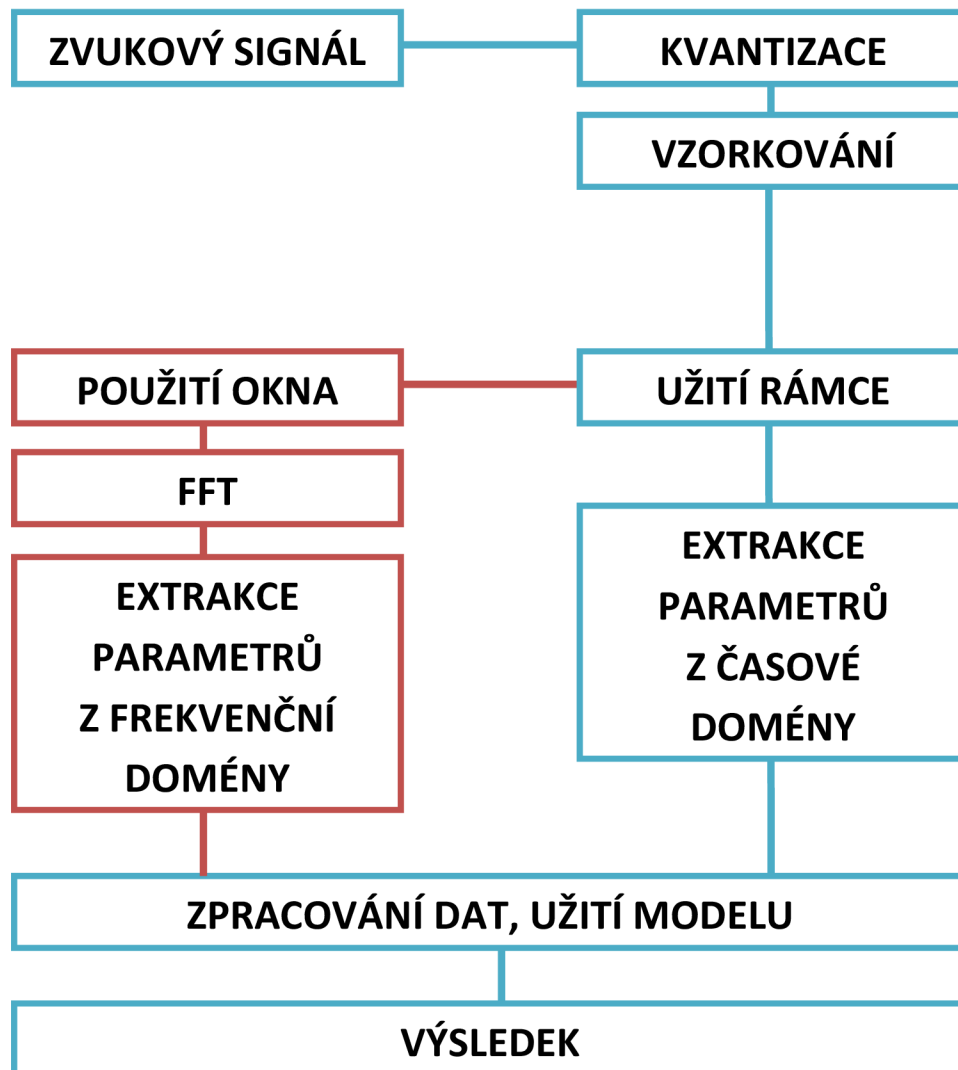
$$RMS_t = \sqrt{\frac{1}{\tau} \int_0^{\tau} E^2(t) dt}. \quad (4.2)$$

Efektivní hodnotu energie signálu lze krom zjištění intenzity využít také k detekci nových událostí v průběhu nahrávky [17, 19].

Dalším často používaným parametrem je *počet průchodů nulou* (Zero-Crossing Rate, ZCR). Prakticky se jedná o počet změn znaménka amplitudy signálu. Jedná se o často

využívaný parametr v analýze mluveného slova, v hudbě se užívá na detekci perkusivních zvuků a hluků. Vyšší hodnoty parametru ZCR obvykle znamenají detekci vyšší frekvence. Pro skladby s absencí perkusivních zvuků dosahují parametry ZCR nižších hodnot a jsou také stabilnější. To může být případ klasické hudby [17]. Výpočet parametru ZCR je následující:

$$ZCR_t = \frac{1}{2} \cdot \sum_{K=t \cdot K}^{(t+1) \cdot K - 1} |sgn(s(k)) - sgn(s(k + 1))|. \quad (4.3)$$



Obr. 4.1: Proces získání parametrů nízké úrovně

## Parametry frekvenční domény

*Poměr energie v pásmech* (Band Energy Ratio, BER) je parametr hojně využívaný při problému rozlišení mluveného slova a hudby. Dále se využívá k určení hudebního žánru.

Energie ve vyšších pásmech se dávají do poměru k energiím nízkých pásem. Tímto způsobem lze získat informaci o tom, jak moc dominantní roli hrají nízké frekvence. Využívá se také inverzní parametr k BER, který popisuje, jak moc jsou vyšší frekvence slabší vůči nižším. Parametr se počítá pomocí následujícího vzorce:

$$BER_t = \frac{\sum_{n=1}^{F-1} m_t(n)^2}{\sum_{n=F}^N m_t(n)^2}. \quad (4.4)$$

$F$  je hraniční frekvence rozdělující pásma na vyšší a nižší. Volba hraniční frekvence může zásadně ovlivnit výsledné hodnoty parametru [17].

*Spektrální těžiště* (Spectral Centroid, SC) určuje frekvenční pásmo s největší koncentrací energie, popisuje barvu zvuku (míru ostrosti a světlosti). Nižší a stabilnější hodnoty ukazují na zvukově světlejší, popř. ostřejší nahrávku. Parametr SC je velmi citlivý na radiální podvzorkování nahrávek (např. ze 44,1 kHz na 11,025 kHz) a filtraci dolní propustí. Oba tyto postupy totiž značně mění původní frekvenční spektrum [17]. Pro výpočet parametru ze spektra existuje vzorec:

$$SC_t = \frac{\sum_{n=1}^N m_t(n) \cdot n}{\sum_{n=1}^N m_t(n)}. \quad (4.5)$$

*Šířka pásma* (Bandwidth, BW) a *spektrální rozsah* (Spectral Spread, SS) popisují frekvenční rozsah v okolí spektrálního těžiště. U klasické hudby se dají očekávat spíše nižší hodnoty, s rostoucím počtem nástrojů by se ale měly mírně zvyšovat [17]. Šířku pásma lze spočítat:

$$BW_t = \frac{\sum_{n=1}^N |n - SC_t| \cdot m_t(n)}{\sum_{n=1}^N m_t(n)}. \quad (4.6)$$

Parametrem popisujícím spektrum je také *spektrální šikmost* (Spectral Skewness, SKW). V podstatě se jedná o míru asymetrie rozložení spektra kolem jeho průměrné hodnoty. Pokud se parametr  $SKW = 0$ , spektrum je rozloženo symetricky. Pokud je  $SKW < 0$ , více energie je detekováno vpravo a pokud  $SKW > 0$ , je více energie vlevo [17].

Parametr *tloušťka spektra* (Spectral Kurtosis, SK) vyjadřuje, jak ploché či špičaté je spektrum v okolí průměrné hodnoty. Pokud je  $SK > 3$  je spektrum spíše špičaté, pokud je  $SK < 3$  je spektrum spíše ploché. Normální distribuce spektra odpovídá hodnotě 3 [17].

Dalším parametrem může být například *spektrální tok* (Spectral Flux, SF), který popisuje změnu v energii po sobě následujících rámcích. Detekuje tedy změny na lokální úrovni. Parametr lze získat pomocí následujícího vzorce:

$$SF_t = \sum_{n=1}^N (D_t(n) - D_{t-1}(n))^2. \quad (4.7)$$

Parametr SF je tedy suma rozdílů energií v sousedících rámcích ve všech frekvenčních pásmech [17].

*Spektrální roll-off* (Spectral Roll-Off) je bod ve spektru, pod kterým se nachází 95 %, popř. 85 % energie signálu [17, 19].



## 4.2 Parametry střední úrovně

V této skupině parametrů jsou zakomponovány psychoakustické modely a proto se více přibližují vnímání posluchače. Jelikož lidské ucho nevnímá absolutní výšku, ale *subjektivní výšku* tónů, jsou některé parametry týkající se výšky uváděny se stupnicí v melech. Stejně subjektivní je i vjem hlasitosti. Lidské ucho se vyznačuje rozdílnou citlivostí v různých frekvenčních pásmech. Proto byla zavedena veličina *hladina hlasitosti* jejíž jednotkou je fón. Vnímání hlasitosti v závislosti na intenzitě a frekvenci udávají tzv. křivky stejné hlasitosti (Fletcher-Munsonovy křivky, Kingsburyho křivky). Nutno podotknout, že křivky platí za určitých podmínek a pouze pro sinusové signály. Hladina hlasitosti neodpovídá subjektivnímu vjemu úplně. Dvojnásobek hlasitosti ve fónech se posluchači nejvíce jeví jako dvojnásobně hlasitý vjem. Proto byla zavedena veličina *hlasitost*, která má stupnici v sonech, přičemž 1 son odpovídá hladině hlasitosti 40 fónů pro tón s frekvencí 1 kHz [41]. Parametry střední úrovně můžeme dále rozdělit podle toho, zda-li je parametr počítán z jednoho rámce nebo z většího bloku nahrávky. Parametry počítané z většího bloku nahrávky (v řádu sekund) jsou schopny popsat tempo [17].

### Mel-frekvenční keprální koeficienty

*Mel-frekvenční keprální koeficienty* (Mel Frequency Cepstral Coefficients, MFCC) jsou získávány ze spektrogramu. Spadají tedy do skupiny parametrů extrahovaných z frekvenční domény. V prvním kroku se zjistí hodnoty frekvencí pro každý rámec. Tyto hodnoty v hertzech jsou následně převedeny na mely. Z logaritmu hodnot v melech je pomocí diskrétní kosinové transformace získáno spektrum frekvencí v melech, které přímo reprezentuje parametry MFCC. Tímto postupem se získá keprum ze všech rámců. Výsledkem jsou vektory parametrů MFCC. Jeden vektor tedy popisuje periodicity nalezené v hodnotách frekvencí obsažených v jednom rámci.

Běžně se počítá 13 až 25 parametrů pro jeden rámec. První parametr MFCC jako jediný nepopisuje frekvence v daném rámci, ale celkovou energii rámce. U dalších parametrů v pořadí platí, že s narůstajícím číslem parametru roste i periodičita.

Výpočet MFCC parametrů je náročný na velikost výsledných dat. Často se tedy jako poslední krok zařazuje statistická sumarizace, kvantizace vektorů nebo se aplikuje pravděpodobnostní model. Tím je možné získat MFCC parametry popisující celou nahrávku [17, 19].

### Fluktuace

*Fluktuace* (Fluctuation Pattern, FP) popisuje, které frekvence se opakovaně objevují, v jakých intervalech, a to v blocích o délce 6 s.

Poté, co je nahrávka rozdělena do bloků, je z každého z nich spočítáno výkonové spektrum pomocí rychlé fourierovy transformace (FFT). V následujících krocích jsou aplikovány psychoakustické modely (barková stupnice, frekvenční maskovací křivky...). Na takto zpracované spektrum, které je rozděleno na kritická pásma, je aplikována další FFT. Tím je možné získat frekvenční struktury v jednotlivých pásmech a ty jsou reprezentovány, co by závislost periodicity na frekvenci. Víme tedy, jaké frekvenční struktury (tj. fluktuace) se objevují opakovaně. Jelikož je vnímání těchto struktur závislé na jejich periodicitě, aplikuje psychoakustický model síly fluktuace. Tím lze detekovat rytmické struktury. Při zjišťování míry podobnosti se porovnávají vektory hodnot globálně reprezentující rytmické struktury obou nahrávek[17].

Parametr fluktuace má v praxi několik variant. Běžně používanou variantou je *logaritmická fluktuace* (Logarithmic Fluctuation Pattern). Na rozdíl od klasické fluktuace je místo lineárního měřítka použito logaritmické. To umožňuje zobrazení periodicit nezávisle na tempu[19].

## Spektrální struktura

*Spektrální struktura* (Spectral Pattern, SP) slouží k detekci tónu a extrahuje se ze spektrogramu každého bloku. Blok je rozdělen na 10 rámců a spektrogram je rozdělen na 98 pásem. Nejedná se tedy o kritická pásma – frekvenční osa je v centech. Výsledkem je matice o rozměrech 98x10.

Stejně jako v případě fluktuace se i tento parametr používá v různých variantách k zjištění specifických vlastností spektra. Například *Delta Spectral Pattern* slouží k detekci počátku jednotlivých tónů [17].

## Korelace

*Korelace* (Correlation Pattern, CP) popisuje harmonické vztahy mezi frekvenčními pásmy. V prvním kroku je spektrogram rozdělen na 52 pásem. Jedná se o doporučený kompromis mezi dostačujícím frekvenčním rozlišením a náročností na výpočet. Jednotlivé bloky nahrávky jsou dále děleny do 256 rámců. Korelace je potom počítána pro všechny dvojice frekvenčních pásem. Vznikají tedy matice o rozměrech 52x52. Všechny matice se dále zpracují statistickou sumarizací (výpočet mediánu) [17, 19].

## 5 Přehled dosavadních prací souvisejících s rozpoznáváním hudebních slohů

Rozpoznávání slohů klasické hudby je problém, který se na poli zpracování audio nahrávek metodou MIR, řeší velmi vzácně. Frekventovaněji se vyskytují témata rozpoznávání hudebních žánrů a identifikace autora skladeb. Obě úzce souvisí s rozpoznáváním hudebního slohu a postup práce je prakticky stejný. Co se týče dat, vychází výzkumy buď ze symbolických dat (MIDI, partitury) nebo z audio nahrávek.

### 5.1 Rozpoznávání hudebních žánrů

Yusuf Yaslan a Zehra Cataltepe ve své práci z roku 2006 hledají nejúspěšnější kombinaci parametrů a metody jejich vyhodnocení. Využívají Tzanetakisův dataset (GTZAN), což je databáze obsahující 10 žánrů. Ke každému z nich 100 vzorků, které jsou dlouhé 30 s. K získání parametrů z audio nahrávek využívají software MARSYAS. Zkoumány jsou konkrétně sety Beat (6 parametrů), STFT (Short-Time Fourier Transform, 9 parametrů), MFCC (Mel-Frequency Cepstral Coefficients, 10 parametrů), MPITCH (5 parametrů) a set All, který zahrnuje všechny dříve zmíněné sety. Zkoumané metody jsou potom tyto: Fisherův klasifikátor, LDC (lineární klasifikátor předpokládající normální rozložení hustot), UDC (kvadratický klasifikátor předpokládající normální nekorelované hustoty), QDC (kvadratický klasifikátor předpokládající normální hustoty), NBC (naivní Bayesův klasifikátor), PDC (metoda Parzenových okének), kNN (metoda nejbližšího souseda), kNN1 (s 1 nejbližším prvkem), kNN3 (se 3 nejbližšími prvky) a kNN5 (s 5 nejbližšími prvky). Trénovací množina byla volena jako 90 % datasetu, testovací data obsahovala zbylých 10 % [48]. Úspěšnost kombinací:

- Fisher s parametry MFCC 79 %,
- různé verze kNN s MFCC 59 % až 79 %,
- kNN a set MPITCH 70 %,
- LDC a set All 70 %,
- QDC při kombinaci 10 parametrů 61 %,
- UDC nejlepší při 25 parametrech,
- Fisher a PDC nejlepší výkon se setem All,
- NBC nejlepší při užití 10 parametrů 78 %.

Z této práce rovněž vyplývá, že větší přesnosti lze dosáhnout kombinací klasifikátorů, např. PNU (kombinace PDC, NBC a UDC) dosahoval přesnosti 80 %. Při užití transformace PCA (analýza hlavních komponent) klesla dimenzionalita setu All a tím stoupla přesnost u metod kNN a QCD na 80 % [48].

Se stejným datasetem pracovali také P. Kozakowski a B. Michalak. Jejich cílem bylo rovněž rozlišení hudebních žánrů z dané databáze. Dataset byl využit v poměru 7:3 na trénování a testování. Ze vzorků byly extrahovány především spektrální parametry (ze spektrogramů v melech). Autoři zvolili metodu konvolučních neuronových sítí a dospěli k celkové úspěšnosti 67 % [18].

Podobně, avšak s jiným datasetem zpracovávali tuto úlohu M. Haggblade, Y. Hong a K. Kao. V práci uveřejněné v roce 2011 používali k řešení opět software MARSYAS. Zvolili set parametrů MFCC a na tomto setu porovnávali metody kNN, k-means, SVM a řešení pomocí neuronových sítí. Zvolený dataset byl GTZAN Genre Collection, ze kterého extrahovali 4 žánry (classical, jazz, metal, pop). Ve výsledku se tedy jednalo o dataset se 400 vzorky (22,05 kHz, 16 bit, AU formát, mono). Trénovací množinu tvořilo 70 % celkových dat [11]. Výsledky:

- metoda KNN a k-means, obě cca 80 %,
- SVM 87 %,
- neuronová síť 96 %.

Výborných výsledků dosáhli F. Medhat, D. Chesmore a J. Robinson při rozpoznávání hudebních žánrů pomocí neuronových sítí (konkrétně metodou MCLNN: Masked Conditional Neural Networks). Jako dataset volili vzorky hudby pro společenské tance ze sbírky Ballroom. Cílem bylo odlišit 8 subžánrů taneční hudby ze 698 vzorků (30 s, 22,05 kHz, 16 bit, mono, WAV formát). Parametry nahrávek byly v tomto případě odečítány z spektrogramů v melech. Úspěšnost tohoto řešení dosahovala cca 92 % [28]. V článku autoři poukazují také na výsledky předchozích studií, které se stejným datasetem, ale jinou metodou dosáhly ještě lepších výsledků. Např. Ballroom dataset zkoumaný metodou SVM s 28 tempovými parametry vykazuje úspěšnost 96 % [34]. Druhou nejúspěšnější kombinací (93 %) je potom metoda kNN s využitím spektra modulace (Modulation Scale Spectrum), které zobrazuje zastoupení frekvencí v čase a tím poskytuje také informaci o rytmu. Ten zásadně odlišuje subžánry taneční hudby [23].

Některé studie využívají k analýze zvukové vzorky ve formátu MIDI. Například práce Ponce de Leona zkoumala úspěšnost algoritmů na vlastním MIDI datasetu, a to se žánry jazz a klasická hudba. Parametry datasetu: 110 vzorků, z toho 45 vzorků klasické hudby a 65 jazzových. U vzorků jsou zjištěny tempové, rytmické, frekvenční a tonální parametry, v celkovém počtu 28 parametrů. Další řešení se pak ubírá dvěma směry, a to využitím metody bez učitele a naopak dvou metod s učitelem. Nejúspěšnější byla metoda kNN s učitelem a využitím všech 28 parametrů – 96,4 %. Úspěšností se jí blížila i naivní Bayesova metoda s učitelem a využitím 10 parametrů – 95,5 %. Nejnižší úspěšnost měla metoda bez učitele, a to konkrétně neuronová síť typu SOM (Self-Organizing Map) s využitím 6 parametrů – 90,7 % [36]. Tento výzkum ve srovnání s výše zmíněnými výzkumy podává velmi dobré výsledky, nicméně v praxi je využívání formátu MIDI často omezeno pouze na určité aplikace. Velké databáze hudebních děl, které je nejčastěji nutné nějakým

způsobem třídit užívají zvukové formáty (WAV, MP3).

Pro velké databáze hudebních děl je poměrně praktické tzv. multi-label rozpoznávání žánrů. V tomto případě není rozhodující pouze analýza audia, ale také analýza textu, obalu alba apod. Tímto typem analýzy se zabýval například S. Oramas a spol. K analýze je v tomto případě nutné využít dataset obsahující celá alba, texty a přebaly CD. Proto byl využit dataset MUMU obsahující 250 žánrů (přes 31 000 alb) skládající se z části datasetu MSD (Million Song Dataset) a z části datasetu Amazon Reviews. Autoři pracují se spektrálními parametry a metodou konvolučních neuronových sítí se čtyřmi konvolučními vrstvami. Data byla rozdělena na množiny – 80 % trénovací, 10 % validační a 10 % testovací. Celková úspěšnost dosáhla 88,8 % [33].

Frekventovaně se na poli MIR řeší také odlišení jednotlivých skladatelů klasické hudby. Ch. Weiss, M. Mauch a S. Dixon pracují ve své studii s datasetem o 1600 vzorcích ve formátu MP3. Mezi vzorky jsou různé hudební formy - klavírní sonáty, orchestrální předehry k operám apod. Autoři článek rovněž dělí skladatele podle období do čtyřech skupin, a to baroko, klasicismus, romantismus a moderna. Každé období je zastoupeno 400 různorodými vzorky (200 orchestrálních, 200 klavírních) od různého počtu skladatelů [47]. Zde je nutné zmínit, že autoři záměrně používají barokní nahrávky hrané na moderní klavír, ne na cembalo. Z historického pohledu je to samozřejmě chyba, jelikož moderní podoba klavíru se ustálila až v průběhu 19. století. Nicméně cílem jejich práce bylo k identifikaci skladatele využít především parametry nesouvisející s barvou zvuku a tónbr cembala je tak výrazně odlišný od klavíru, že bylo nutné z datasetu tento nástroj vyřadit. K identifikaci byly využívány chromatické a tonální parametry z Chroma Toolboxu (pro Matlab)[31], ty byly dále zpracovávány pomocí metody SVM, implementované v knihovně LIBSVM. Dle výsledků dosahuje metoda přesnosti 82,5 % při kombinaci nejúspěšnějších parametrů [47].

Ještě větší úspěšnosti dosáhla studie F. Clara, která porovnává přesnost metod SVM a Random Forest. Použitý dataset obsahoval téměř 10 000 skladeb od celkem 106 autorů. Z parametrů byly voleny především ty, popisující tóninu a chromatiku skladeb (odečet hodnot z chromagramů apod.). Jako úspěšnější se prokázala metoda Random Forest, která s 91 % úspěšností určila skladby spadající do klasicistní éry [7].

Jiných výsledků dosáhl ve své studii G. Ruiz. Jeho cílem bylo určit autora klasické hudby, přičemž dataset obsahoval 800 vzorků (30 s, 44,1 kHz, formát WAV) od 2 barokních autorů, 3 klasicistních a 3 autorů spadajících do éry romantismu. Ze vzorků extrahoval parametry popisující především spektrum a k jejich zpracování použil metodu SVM implementovanou ve WEKA Toolbox (Waikato Environment for Knowledge Analysis). Tímto postupem dosáhl autor pouze 55 % úspěšnosti [24].

Ch. Weiss ve své disertační práci pojednává o třídění klasické hudby podle slohů. Vytvořil dataset o 1600 vzorcích. Zkoumá čtyři slohy: baroko, klasicismus, romantismus a modernu, tj. hudbu 1. poloviny 20. století. Pro každé období používá tedy 400 vzorků. Dataset

zohledňuje také různou popularitu autorů tím, že od často hraných autorů obsahuje více skladeb. Co se týče parametrů, Weiss zjišťuje poměr mezi výskytem plagálních (subdominanta - tónika) a autentických (dominanta - tónika) závěrů v jednotlivých skladbách pomocí pluginu Chordino. Sleduje také, jak často se vyskytují disonance. Dále pracuje především s tonálními a chromatickými parametry (MFCCs, plugin NNLS chroma). Na parametry týkající se chromatiky je aplikována lineární diskriminační analýza. Parametry jsou zpracovávány metodou využívající Gaussovy smíšené modely (GMM - Gaussian Mixture Models). Neúspěšnější kombinace (GMM + NNLS chroma) dosahuje cca 75 % přesnosti pro celý dataset a 79 % pro orchestrální skladby [46].

Typově podobným problémem je odlišení jednotlivých orchestrálních těles z nahrávek, popř. odlišení jednotlivých dirigentů. Například C. Liem a A. Hanjalic ve své práci dospěli k závěru, že je tento problém řešitelný pomocí analýzy hlavních komponent (metoda PCA), přičemž jako parametry jsou použity spektrogramy nahrávek. Daná metoda poté odliší rozdíly mezi spektrogramy. Dataset obsahoval 24 celých nahrávek Beethovenovy symfonie č. 3 a 6 celých nahrávek Straussovy Alpské symfonie [21].

Přehled dalších výzkumů v oblasti rozlišení slohů klasické hudby či rozlišení skladatelů je uveden v následující tabulce (Tab. 5.1).

Tab. 5.1: Výzkumy na poli klasické hudby

<b>Autor</b>	<b>Třídy</b>	<b>Data</b>	<b>Parametry</b>	<b>Úspěšnost</b>
McKay & Fujinaga[26]	3 slohy	MIDI	různé	85 %
Mearns <i>et al.</i> [27]	7 autorů	partitura	intervaly, kontrapunkt	66 %
Geertzen[10]	3 autoři	partitura	melodie, rytmus	80 %
Van Kranenburg[43]	5 autorů	partitura	intervaly, kontrapunkt	80 %
Hontanilla <i>et al.</i> [14]	5 autorů	partitura	melodické $n$ -gramy	79 %
Honingh & Bod[13]	tonalita- atonalita	MIDI	intervaly	95 %
Hillewaere <i>et al.</i> [12]	Mozart- Haydn	MIDI	melodické $n$ -gramy, statistika	75 %
Dor & Reich[46]	9 autorů	partitura	melodie, oktávy, tóny	79 %
Hu <i>et al.</i> [46]	9 autorů	audio	MFCC	76 %
Izmirli[16]	tonalita- atonalita	audio	chromagramy	91 %
Hamel[46]	11 autorů	audio	MFCC	78 %

## 6 Implementace

### 6.1 Návrh řešení

Postup práce při řešení problému pomocí metody MIR byl popsán v předešlých kapitolách. Na začátku stojí vždy dataset nahrávek, který musí být dostatečně velký a kvalitní. V případě této práce byl vytvořen vlastní dataset, který obsahuje 600 nahrávek ze tří historických období vývoje hudebního umění. Jelikož se jedná o nahrávky pocházející z různých zdrojů bylo nutné je předzpracovat, tj. upravit délku, normalizovat hlasitost a sjednotit vzorkovací frekvence (viz 2.2). Bez fáze předzpracování by byly parametry ovlivněny například různou hlasitostí nahrávek, která sice může do určité míry souviset s počtem hrajících nástrojů, ale především souvisí s tím, za jakým účelem byla nahrávka zpracována, což není předmětem zkoumání.

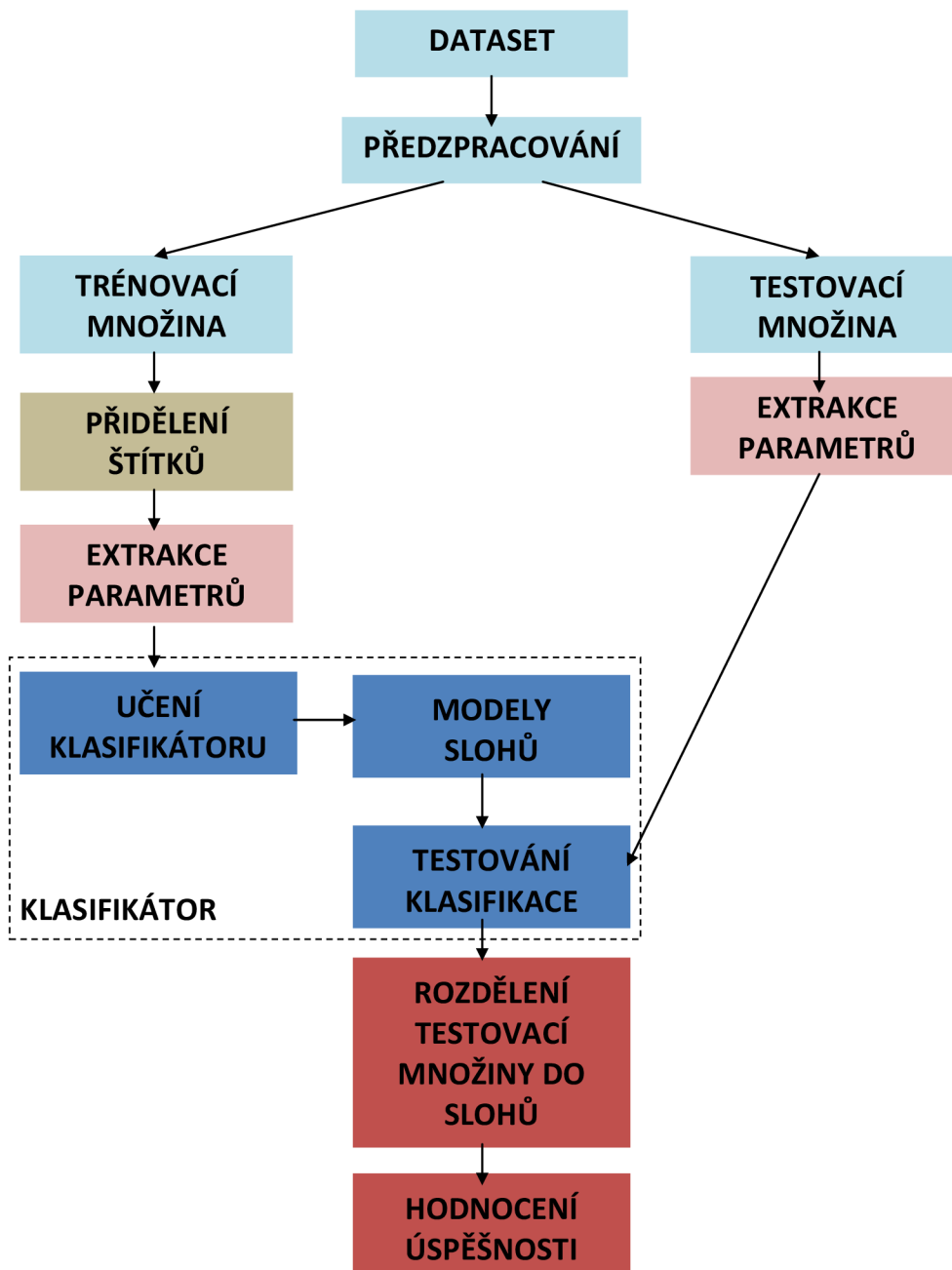
Následuje rozdělení datasetu na trénovací a testovací množinu. Trénovací množina obsahuje 480 nahrávek a testovací množina zbylých 120. Trénovací množina tedy tvoří 80 % z celkového počtu nahrávek. Nahrávkám v testovací množině jsou přiděleny štítky značící sloh, do kterého spadají. Štítky usnadňují přehlednost a orientaci v nahrávkách.

Extrakce parametrů se provádí u obou množin v prostředí Matlab, konkrétně pomocí toolboxu MIRtoolbox. Pro popis nahrávek klasické hudby jsou využívány především parametry MFCC, parametry popisující spektrum a harmonii, či tonální parametry (např. chromagramy). Méně jsou potom používány parametry popisující tempo či rytmus, a to především proto, že tyto vlastnosti jsou proměnné a nebo nemusí být v dané skladbě vůbec obsaženy, jako je tomu například u barokní fugy (viz 5.1). Získány budou parametry popisující všechny oblasti a jejich evaluace bude provedena metodou minimum Redundancy Maximum Relevance.

Po extrakci parametrů jsou již všechny nahrávky reprezentovány pomocí vektorů parametrů a v případě trénovací množiny také pomocí štítků. V dalším kroku začíná učení vybraného klasifikátoru (viz 3.1). Na poli klasické hudby jsou nejefektivnější klasifikátory LDA a SVM (viz 5.1). Neuronové sítě se využívají v kombinaci s převodem nahrávek na spektrogramy (Image-Based Approach)[21].

Po natrénování klasifikátoru jsou známé modely slohů, které se testují na parametrech nahrávek z testovací množiny. Úspěšnost klasifikátoru je dána procentem správně určených nahrávek.





Obr. 6.1: Návrh implementace

## 6.2 Výpočet parametrů

K získání parametrů z obou datasetů byly využity funkce, které jsou obsaženy v balíčku MIRtoolbox. Pomocí tohoto toolboxu je možné získat množství parametrů přímo popisujících rytmus, harmonický vývoj skladby, tónbr a dynamiku. Také je možné získat parametry statisticky hodnotící spektrum (tvar, těžiště atd.). V případě této práce bylo

celkem získáno 110 parametrů ze všech skupin. Získané parametry byly uloženy do matice tak, aby pro každý hudební vzorek existovala řada číselných hodnot. Z tohoto důvodu bylo nutné provést také redukci hodnot parametrů pomocí statistických úprav. Provedeny byly tyto statistické úpravy:

- Průměr - průměrná hodnota parametru napříč vzorky.
- Medián - rozděluje řadu hodnot parametru na dvě poloviny.
- Směrodatná odchylka - jedná se o odmocninu z rozptylu parametru. Lze ji vypočítat pomocí následujícího vzorce:

$$\sigma = \sqrt{E(X - EX)^2}. \quad (6.1)$$

Kde  $X$  je daný parametr a  $E$  je jeho střední hodnota.

- Rozptyl - vyjadřuje variabilitu rozdělení náhodných hodnot kolem střední hodnoty veličiny. Lze vypočítat pomocí vzorce:

$$\sigma^2 = \sum_{i=1}^n [x_i - E(X)]^2 p_i. \quad (6.2)$$

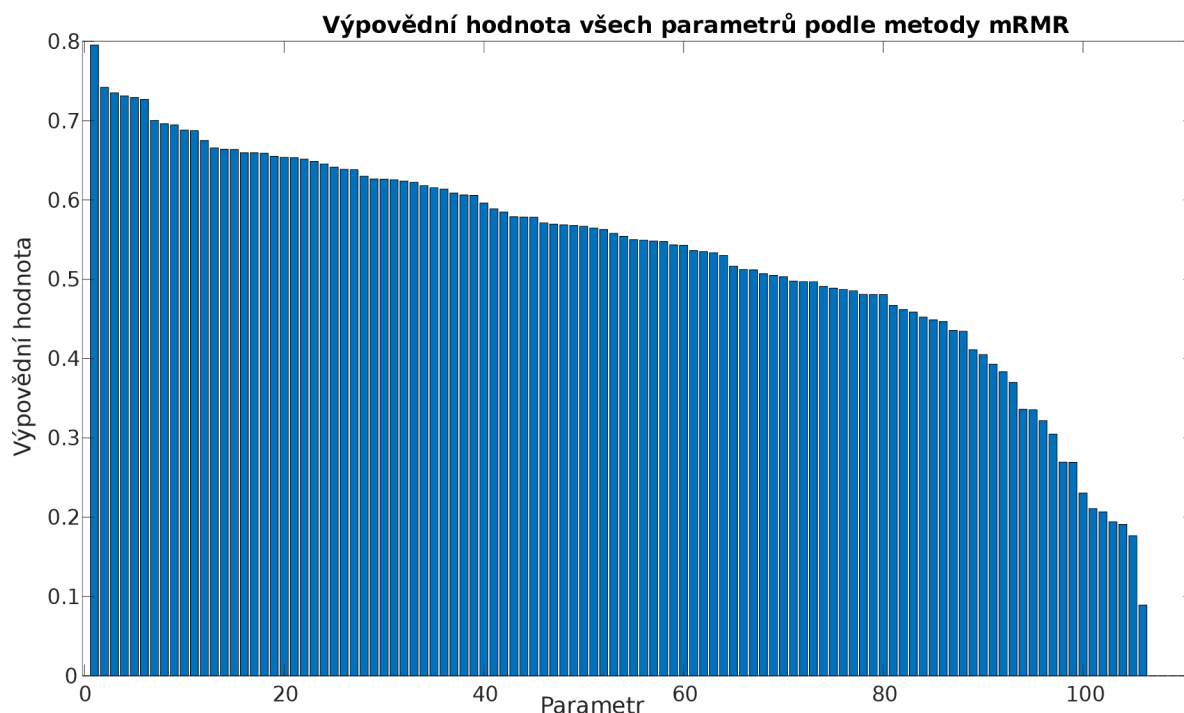
Kde  $x_i$  jsou hodnoty, kterých může parametr  $X$  nabývat s pravděpodobností  $p_i$  a  $E(X)$  je jeho střední hodnota.

### 6.3 Výběr parametrů metodou mRMR

Ke zhodnocení výpovědní hodnoty parametrů byla použita metoda minimum Redundancy Maximum Relevance (dále mRMR), a to ve verzi od čínského vědce Hanchuana Penga. Vstupní proměnné této metody jsou data ve formě matice, kde počet řádků odpovídá počtu hudebních vzorků a počet sloupců odpovídá počtu získaných parametrů, dále také vektor s informací o zařazení každého hudebního vzorku do dané třídy. Cílem metody je najít podskupinu parametrů, která optimálně charakterizuje jednotlivé třídy. Metoda mRMR nahlíží na parametry z několika hledisek. Prvním hlediskem je *maximální relevance*. Vybrány jsou tedy parametry, které danou třídu popisují nejvýznamněji. Relevance parametrů vůči třídě se dá popsat pomocí vzájemných informací (definovaných hustotou pravděpodobnosti), které charakterizují vztahy a závislosti mezi jednotlivými parametry a třídou. Z hlediska maximální relevance jsou vybrány parametry, které mají největší vzájemnou informaci s cílovou třídou, tj. mezi těmito parametry a cílovou třídou existuje silná závislost [35].

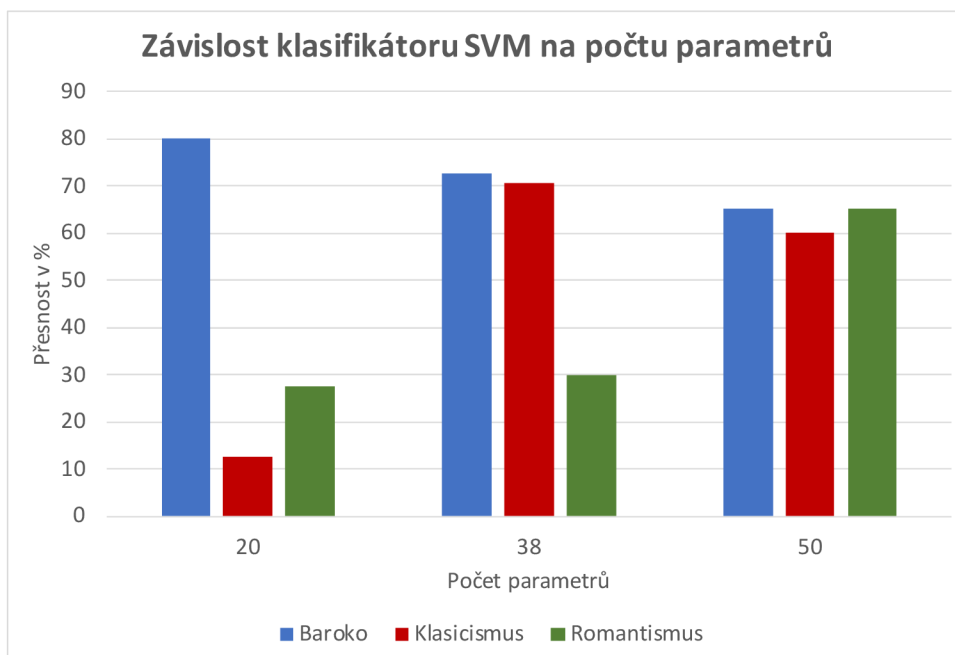
Pokud jsou parametry vybrány pouze individuálně, nemusí být výsledná skupina těchto parametrů příliš účinná. Proto probíhá také výběr parametrů z hlediska *minimální redundance*, tak aby z výsledné skupiny mohly být redundantní parametry odstraněny. Tento výběr probíhá heuristicky. Autor H. Peng ve své práci předpokládá velkou úspěšnost vybrané skupiny parametrů především s klasifikátory Naive Bayes, SVM a LDA [35].

Jednotlivé parametry byly seřazeny podle své výpovědní hodnoty, jak jde vidět z grafu 6.2. Dle autora toto seřazení nutně neznamená, že první parametr je individuálně účinnější než druhý. Seřazení popisuje jak vztah parametru k třídě, tak i vztah mezi parametry. Tedy například: kombinace prvního a druhého parametru má větší výpovědní hodnotu než kombinace druhého a třetího parametru [30].

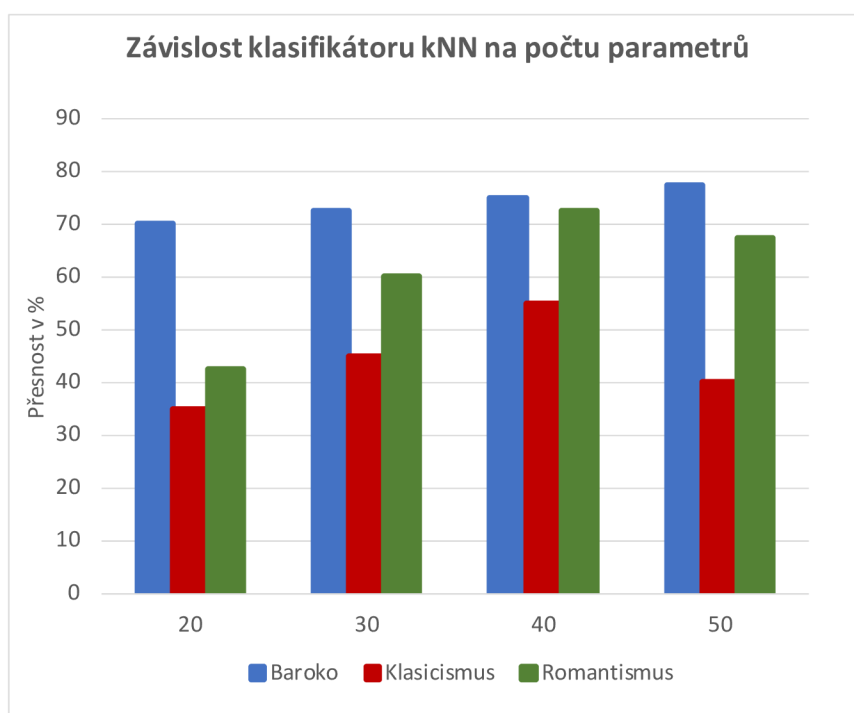


Obr. 6.2: Výpovědní hodnota všech parametrů

V případě této práce byly vybrány různě početné skupiny parametrů a byla ověřena jejich účinnost s jednotlivými klasifikátory. U klasifikátoru SVM měl počet parametrů vliv na přesnost klasifikace rozdílně pro každý hudební sloh, jak lze vidět na grafu 6.3. Podle grafu 6.5 jsou ve skupině prvních 20 parametrů obsaženy především parametry spektrální. Tím je dána vysoká úspěšnost v identifikaci barokního slohu, jelikož baroko, díky odlišnému nástrojovému složení, vykazuje výrazně jinou hudební barvu než zbylé dva slohy. Identifikace klasicismu byla nejúspěšnější se skupinou 38 parametrů. K předešlým převážně spektrálním parametrům byly přidány parametry dynamické, rytmické a tonální. Klasicismus je charakteristický užíváním převážně durových tónin a oproti baroku jsou zavedeny nové rytmické postupy a odlišná práce s dynamikou, což s malými obměnami přetrvává i během romantismu. Proto je klasifikátor úspěšný v identifikaci rozdílu mezi barokem a klasicismem, ale neúspěšný v identifikaci romantismu. Dalším zvýšením počtu parametrů, došlo k vyrovnání úspěšnosti identifikace všech slohů.



Obr. 6.3: Závislost přesnosti klasifikátoru SVM na počtu parametrů

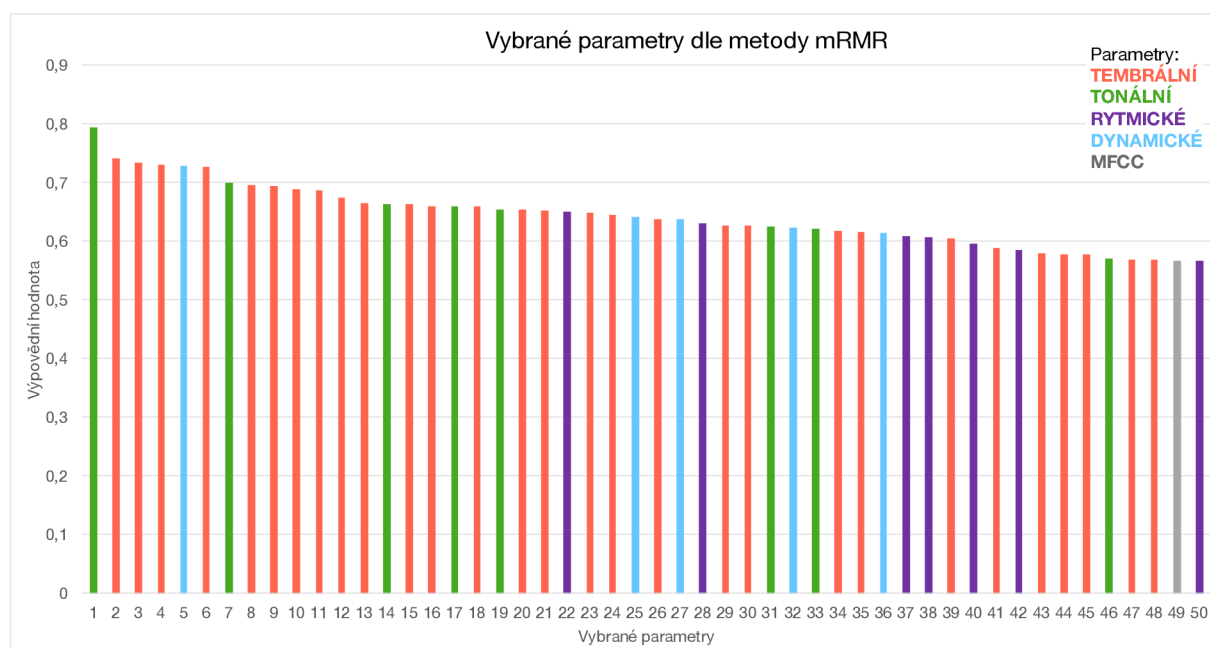


Obr. 6.4: Závislost přesnosti klasifikátoru kNN na počtu parametrů

V případě klasifikátoru kNN měla nejlepší výsledky skupina čtyřiceti parametrů. Ve většině případů měl klasifikátor kNN velkou úspěšnost v identifikaci rozdílu mezi barokem a romantismem. Menší úspěšnost měl při identifikaci klasicismu, který nejčastěji

zaměřoval s romantismem. Což je velmi pravděpodobně způsobeno expresivní tvorbou Beethovena, která se od zbylých klasicistních autorů liší a přibližuje se romantismu. Třídy klasicismus a romantismus od sebe tedy nejsou tak dobře odděleny jako baroko a klasicismus nebo baroko a romantismus. Klasifikátory LDA a GMM dosahovaly největší přesnosti se skupinou čtyřiceti nebo padesáti parametrů.

Vybrané podskupiny obsahovaly parametry popisující hudební vzorky ze všech hledisek. Nejvíce byly zastoupeny parametry obsahující informaci o spektru a hudební barvě (témbru), méně potom parametry tonální, rytmické či dynamické. Přesné zastoupení jednotlivých skupin parametrů je znázorněno v grafu 6.5. Celkový seznam parametrů je uveden v příloze, v tabulce A.1.



Obr. 6.5: Závislost přesnosti klasifikátoru kNN na počtu parametrů

## 6.4 Optimalizace hyperparametrů

Další funkcí, která byla při klasifikaci často využívána je funkce optimalizace hyperparametrů. Díky této funkci je možné optimalizovat nastavení klasifikátoru, tak aby co nejlépe reagoval na vstupní data. Každý klasifikátor je popsán jinými parametry. U klasifikátoru kNN je například nutné nastavit  $N$  počet nejbližších sousedů a také vzdálenostní funkci, kterou bude vzdálenost mezi nejbližšími sousedy počítána. U klasifikátoru LDA je naopak nutné nastavit parametr  $\Delta$ , což je rozhodovací úroveň pro koeficienty modelu a parametr  $\Gamma$ , který udává množství regularizace při určení matice kovariance vstupních parametrů. Funkce optimalizace hyperparametrů využívá metodu sekvenční optimalizace, která je založená na náhradním modelu, tzv. Bayesovská optimalizace [25].

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	Delta	Gamma
1	Best	0.67133	0.022733	0.67133	0.67133	142.28	0.30132
2	Best	0.41958	0.020827	0.41958	0.43284	3.1877e-05	0.55461
3	Best	0.3986	0.018075	0.3986	0.39866	0.00010412	0.049034
4	Accept	0.67133	0.019839	0.3986	0.39868	1.5638	0.91695
5	Accept	0.40559	0.019048	0.3986	0.39872	0.00069628	0.27818
6	Accept	0.42657	0.01855	0.3986	0.39854	0.00022921	0.66281
7	Best	0.37762	0.018657	0.37762	0.37765	1.0855e-06	0.039217
8	Accept	0.47552	0.019736	0.37762	0.37767	1.0161e-06	0.99894
9	Best	0.37063	0.019244	0.37063	0.37074	3.4659e-06	0.0013082
10	Accept	0.37063	0.019068	0.37063	0.37066	3.3716e-06	0.0029144
11	Best	0.36364	0.018094	0.36364	0.36878	2.893e-06	0.0017284
12	Accept	0.36364	0.018561	0.36364	0.36737	4.2545e-06	0.00090949
13	Best	0.35664	0.018546	0.35664	0.36476	5.4806e-06	0.00019429
14	Accept	0.35664	0.019326	0.35664	0.3617	6.7592e-06	0.00042201
15	Accept	0.35664	0.019143	0.35664	0.35865	8.7403e-06	0.00018045
16	Accept	0.37063	0.018364	0.35664	0.36246	9.572e-06	0.0037488
17	Accept	0.37063	0.018497	0.35664	0.3624	0.1162	0.0029339
18	Accept	0.36364	0.01848	0.35664	0.36256	0.017259	0.0009758
19	Accept	0.35664	0.018454	0.35664	0.3625	0.030216	0.00025446
20	Accept	0.37063	0.018234	0.35664	0.36262	0.022908	0.0012891

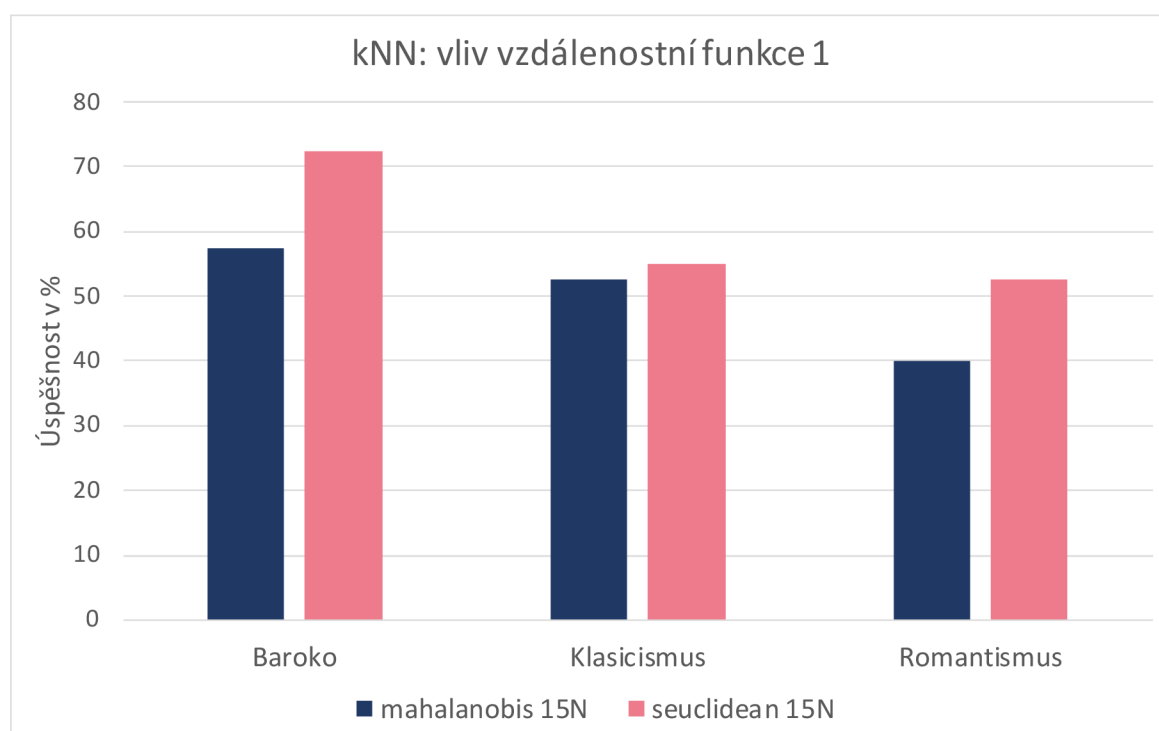
Obr. 6.6: Příklad výsledků optimalizace hyperparametrů pro klasifikátor LDA

## 7 Úspěšnost klasifikátorů

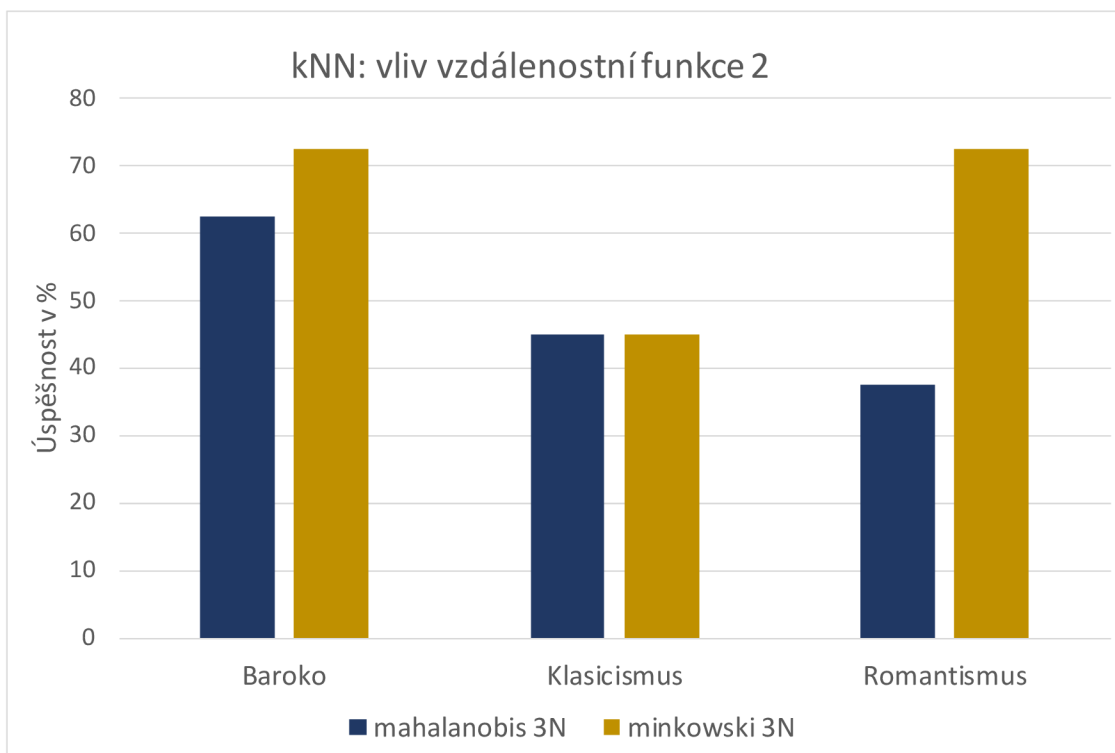
V této části práce jsou shrnuty výsledky všech zkoumaných klasifikátorů. Klasifikátory byly hodnoceny jednak na celém datasetu, kdy bylo cílem rozdělit hudební vzorky do tří tříd - baroko, klasicismus a romantismus. Dále byl tento úkol převeden do binární podoby, a to dvěma způsoby. V prvním případě byl jeden ze slohů z datasetu vyřazen a klasifikátory tak určovaly pouze dvě třídy. Ve druhém případě byl zachován celý dataset, ale klasifikátor hledal pouze hudební vzorky jednoho ze slohů. Rovněž tedy vznikly dvě třídy, přičemž jedna třída byla tvořena vybraným slohem a druhá třída byla tvořena zbylými dvěma slohy. Kompletní číselný přehled všech výsledků je v příloze B.1.

### 7.1 kNN

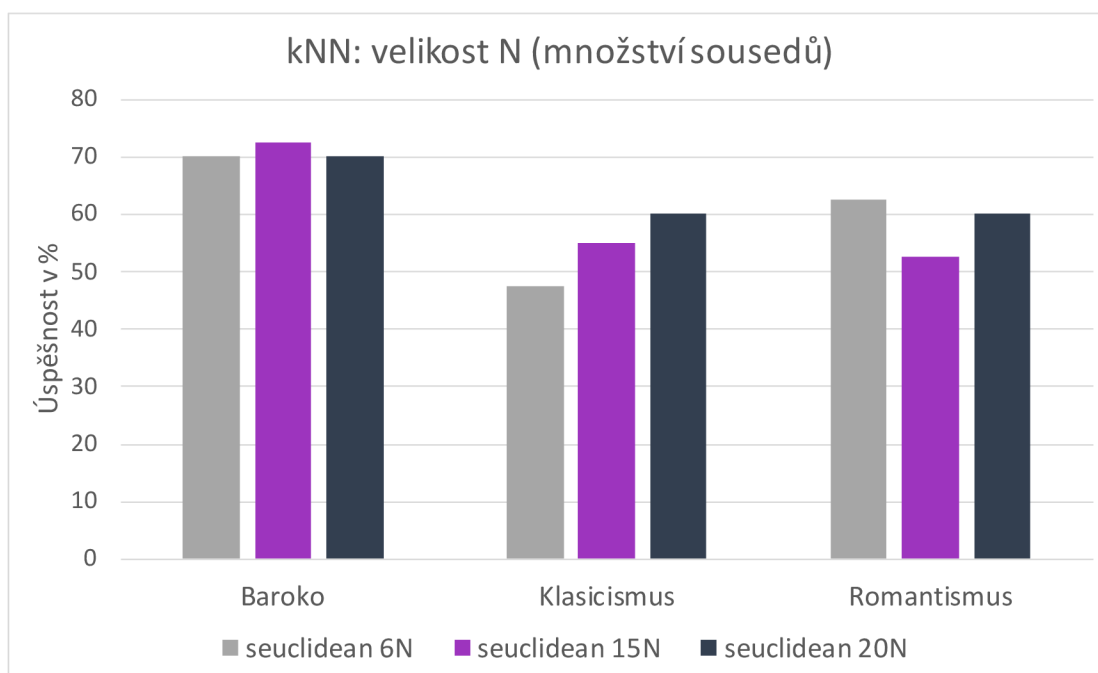
V případě metody kNN měla nejlepší výsledky skupina 40 parametrů, viz. 6.4. K optimalizaci modelu kNN byla využita funkce optimalizace hyperparametrů (parametry klasifikátoru). V případě metody kNN je nutné nastavit především správný počet „sousedů“ a také najít správnou funkci, podle které je vzdálenost mezi těmito „nejbližšími sousedy“ určena. Jak ukazují grafy 7.1, 7.2 a 7.3, různý počet N sousedů a různá vzdálenost mohou mít zásadní vliv na výsledek klasifikace.



Obr. 7.1: Závislost přesnosti klasifikátoru kNN na vzdálenostní funkci



Obr. 7.2: Závislost přesnosti klasifikátoru kNN na vzdálenostní funkci



Obr. 7.3: Závislost přesnosti klasifikátoru kNN na počtu sousedů

Z grafů 7.1 a 7.2 je patrné, že při zachování všech parametrů a změně vzdálenostní funkce může dojít ke značné změně přesnosti klasifikace. Vliv množství N sousedů na přes-



nost klasifikace je mírně menší. Jelikož jsou tyto změny rozdílné pro každou třídu, je nutné hledat kompromis. Změna hyperparametrů dobře fungující pro jednu třídu, může mít přesně opačný vliv na třídu druhou. Pomocí funkce optimalizace parametrů je možné získat ideální kompromis v nastavení vzdálenostní funkce a počtu  $N$  sousedů.

V případě celého datasetu, kdy klasifikátor rozdělval hudební vzorky do tří tříd, dosahovala nejvyšší celková úspěšnost 67,5 % (baroko: 75,0 %, klasicismus: 55,0 %, romantismus: 72,5 %). Jako vzdálenostní funkce byla přitom určena euklidovská vzdálenost a  $N = 5$  sousedů.

V druhé části testování byly v celém datasetu hledány vzorky odpovídající hudebnímu slohu  $X$ . Úspěšnost klasifikátoru se v tomto případě rapidně snížila, a to především u klasicismu a romantismu. Již dříve bylo zmíněno, že dvojice klasicismus - romantismus vykazuje teoreticky větší podobnost než dvojice baroko - klasicismus a baroko - romantismus. Proto také v tomto případě dosahuje identifikace baroka větší úspěšnosti.

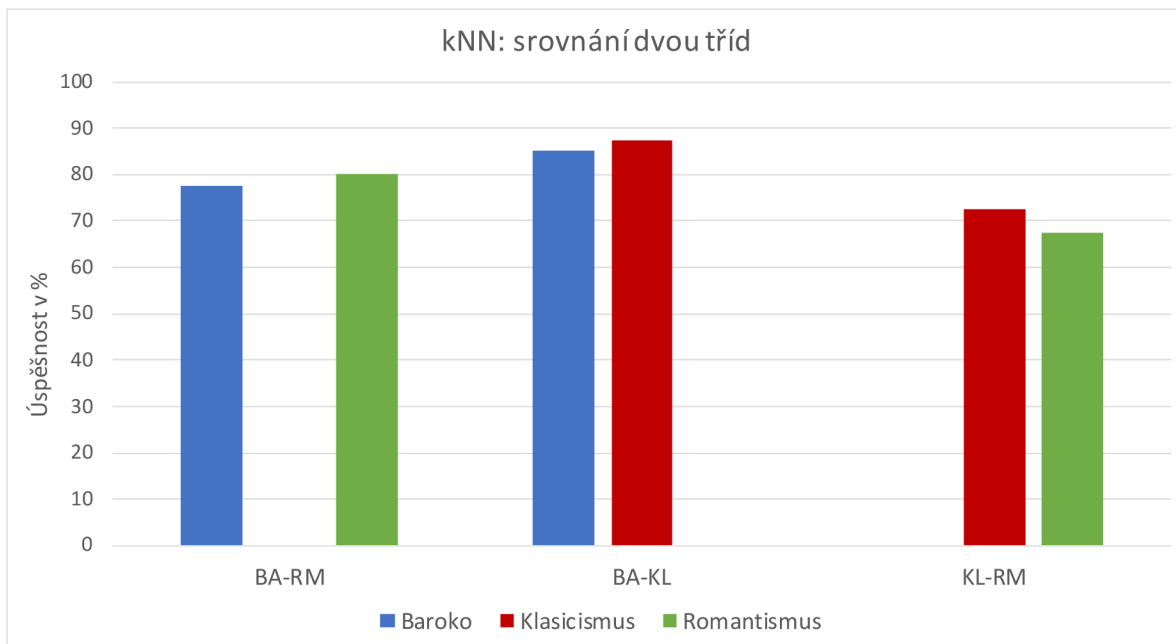
Sloh $X$	Vzdálenostní funkce	Počet sousedů	Úspěšnost
Baroko	korelační	$N = 3$	67,5 %
Klasicismus	Mahalanobis	$N = 1$	40,0 %
Romantismus	korelační	$N = 20$	42,5 %

Tab. 7.1: Úspěšnost klasifikátoru kNN při určení jednoho slohu z celého datasetu

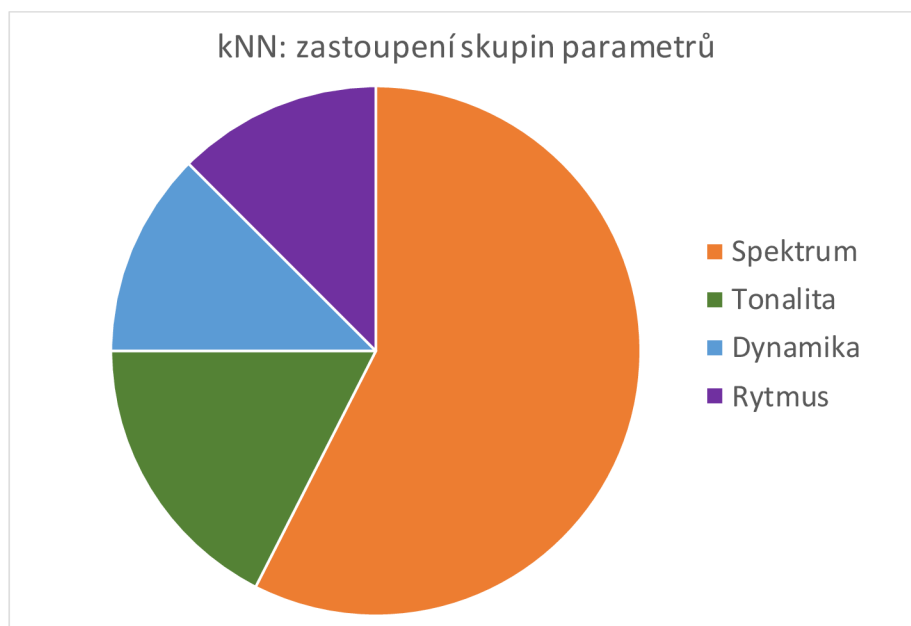
Nejvyšší úspěšnosti dosahoval klasifikátor v případě, kdy byl z datasetu vyloučen jeden ze slohů a klasifikátor tak řešil binární problém. Z grafu 7.4 je patrné, že dvojice klasicismus - romantismus je pro správnou identifikaci nejnáročnější. Nejlépe naopak funguje klasifikace v případě dvojice baroko - klasicismus. Vzhledem k tomu, že většina použitých parametrů je spektrálních, dobrá identifikace barokního ansámblu se specifickou hudební barvou je opodstatněná.

Dvojice	Sloh	Vzdálenostní funkce	Počet sousedů	Úspěšnost
BA vs. KL	Baroko	korelační	$N = 1$	85,0 %
	Klasicismus	korelační	$N = 1$	87,5 %
BA vs. RM	Baroko	korelační	$N = 1$	77,5 %
	Romantismus	korelační	$N = 1$	80,0 %
KL vs. RM	Klasicismus	Mahalanobis	$N = 11$	72,5 %
	Romantismus	Mahalanobis	$N = 11$	67,5 %

Tab. 7.2: Úspěšnost klasifikátoru kNN při řešení sloh  $X$  nebo  $Y$

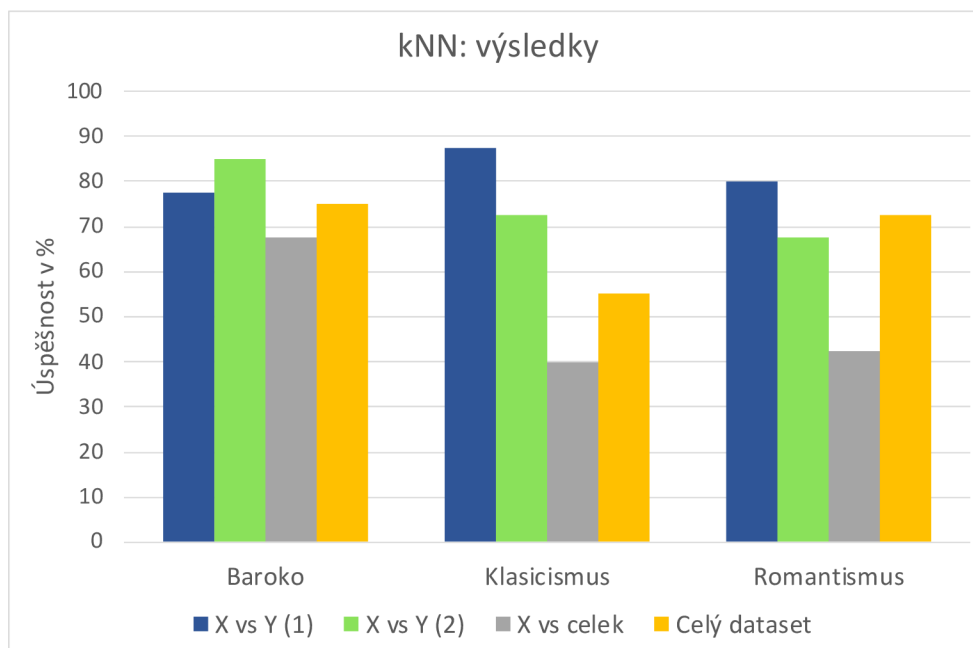


Obr. 7.4: Úspěšnost klasifikátoru kNN při řešení binárního problému



Obr. 7.5: Úspěšnost klasifikátoru kNN při řešení binárního problému

V grafu 7.6 je vyobrazeno shrnutí úspěšnosti klasifikace všech řešených úkolů. Nejlepších výsledků bylo dosaženo při odebrání jednoho slohu z datasetu. Nejméně přesná byla naopak klasifikace při rozpoznávání hudebních vzorků jednoho slohu z celého datasetu. Co se týče úspěšnosti identifikace jednotlivých slohů, baroko bylo identifikováno s nejvyšší úspěšností ve všech případech.



Obr. 7.6: Celková úspěšnost klasifikátoru kNN

## 7.2 LDA

Lineárně diskriminační analýza byla jediným klasifikátorem, který vykazoval nejnižší úspěšnost v případě barokního slohu. Klasifikátor pracoval se stejnou skupinou parametrů jako ostatní klasifikátory. Počet parametrů měl však na přesnost klasifikace minimální vliv. Nejvyšší přesnost byla dosažena jak se 40, tak s 50 parametry. V případě klasifikace všech tří tříd dosahovala přesnost těchto hodnot:

- baroko: 55,0 %,
- klasicismus: 62,5 %,
- romantismus: 72,5 %.

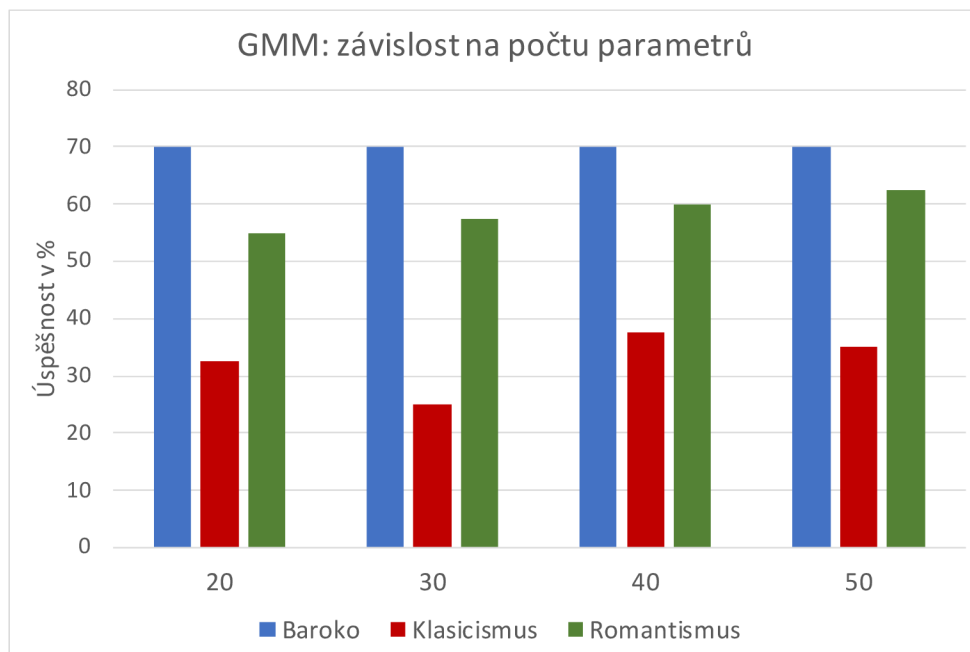
K optimalizaci klasifikátoru byla opět použita funkce optimalizace hyperparametrů. V tabulce 7.3 jsou výsledky řešení binárního problému. Z datasetu byl vždy vyloučen jeden ze slohů a zkoumány byly zbývající dva. Vyloučení jednoho slohu vedlo ke zvýšení úspěšnosti klasifikace v průměru o 10 %. Z výsledků také vyplývá, že mezi skupinou 40 a 50 parametrů je stále minimální rozdíl, tak jako tomu bylo při hodnocení všech tří tříd. I v případě binárního problému je nejnižší úspěšnost klasifikace u barokního slohu.

X vs. Y	Sloh	40 parametrů	50 parametrů
BA vs. KL	Baroko	<b>67,5 %</b>	65,0 %
	Klasicismus	<b>87,5 %</b>	85,0 %
BA vs. RM	Baroko	62,5 %	<b>65,0 %</b>
	Romantismus	82,5 %	<b>82,5 %</b>
KL vs. RM	Klasicismus	<b>72,5 %</b>	<b>72,5 %</b>
	Romantismus	<b>77,5 %</b>	<b>77,5 %</b>

Tab. 7.3: Úspěšnost klasifikátoru LDA v závislosti na počtu parametrů

### 7.3 GMM

Maximální úspěšnosti dosahuje klasifikátor GMM při použití 40 a 50 parametrů. Na rozdíl od klasifikátoru LDA je však více závislý na počtu parametrů, a to především při klasifikaci romantismu a klasicismu. Naopak v případě barokního slohu není na počtu parametrů závislý téměř vůbec. Graf 7.7 ukazuje výsledky tohoto klasifikátoru při hodnocení všech tří tříd s různým počtem vstupních parametrů. Stejně jako v případě klasifikátorů kNN a SVM, i zde je nejméně úspěšná identifikace klasicismu.



Obr. 7.7: Úspěšnost klasifikátoru GMM v závislosti na počtu parametrů

Po vyloučení jednoho slohu z datasetu, jde vidět, že problematická je opět dvojice klasicismus - romantismus, kde přesnost klasifikace klasicismu nedosahuje ani 50 %. Klasifikace klasicismu ve dvojici s barokem přitom dosahuje téměř 80 %. Výsledky ukazují,

že zatímco v případě dvojic baroko - klasicismus a baroko - romantismus jsou vždy obě třídy dobře odděleny. V případě dvojice klasicismus - romantismus třídy dobře odděleny nejsou. Tento fakt je ve shodě s hudebně-teoretickým pohledem na oba slohy.

<b>X vs. Y</b>	<b>Sloh</b>	<b>40 parametrů</b>	<b>50 parametrů</b>
BA vs. KL	Baroko	<b>70,0 %</b>	<b>70,0 %</b>
	Klasicismus	<b>77,5 %</b>	<b>77,5 %</b>
BA vs. RM	Baroko	<b>75,0 %</b>	70,0 %
	Romantismus	<b>80,0 %</b>	82,5 %
KL vs. RM	Klasicismus	<b>45,0 %</b>	42,5 %
	Romantismus	<b>65,0 %</b>	65,0%

Tab. 7.4: Úspěšnost klasifikátoru GMM v závislosti na počtu parametrů

## 7.4 SVM

Metoda Support Vector Machines (SVM) dosahuje při klasifikaci klasické hudby zpravidla vysoké úspěšnosti. Jak již bylo zmíněno v kapitole 3.1.4, je tato metoda určena pro řešení binárních problémů. Pomocí klasifikátoru SVM tedy byla řešena identifikace slohu X z celého datasetu a zařazení hudebního vzorku do správné třídy v případě, že je dataset tvořen pouze dvěma slohy. Na klasifikaci měl značný vliv počet vstupních parametrů.

<b>X vs. zbytek</b>	<b>20 parametrů</b>	<b>38 parametrů</b>	<b>50 parametrů</b>
Baroko	80,0 %	72,5 %	65,0 %
Klasicismus	12,5 %	70,5 %	60,0 %
Romantismus	27,5%	30,0 %	65,0 %

Tab. 7.5: Úspěšnost klasifikátoru SVM v závislosti na počtu parametrů

Při řešení binárního problému, zda-li daný hudební vzorek spadá do vybraného slohu X nebo ne, pracuje klasifikátor se dvěma třídami. První třídou je vybraný sloh X a druhou třídu tvoří zbytek datasetu, tedy kombinace dvou zbývajících slohů. Z tabulky 7.5 vyplývá, že pro každý sloh je ideální jiné množství parametrů. Nejlepším kompromisem byla skupina 50 parametrů, kde úspěšnost identifikace každého slohu přesahovala nebo se rovnala hodnotě 60 %. Tento kompromis, ale v případě barokního slohu znamená snížení úspěšnosti identifikace o 20 %.

Při řešení druhého binárního problému, dojde k vyloučení jednoho slohu z datasetu. Klasifikátor tedy opět pracuje se dvěma třídami. Každá třída reprezentuje jeden sloh. Přesnost klasifikace dosahovala ve všech případech hodnot přes 70 %. Podle očekávání

byla klasifikace nejméně přesná u dvojice klasicismus - romantismus. Naopak nejvyšší přesnosti dosahovala dvojice baroko - klasicismus při použití skupiny 40 parametrů. Kompletní výsledky jsou uvedeny v tabulce 7.6.

<b>X vs. Y</b>	<b>Sloh</b>	<b>30 parametrů</b>	<b>40 parametrů</b>	<b>50 parametrů</b>
BA vs. KL	Baroko	77,5 %	<b>82,5 %</b>	77,5 %
	Klasicismus	82,5 %	<b>87,5 %</b>	90,0 %
BA vs. RM	Baroko	<b>80,0 %</b>	77,5 %	75,0 %
	Romantismus	<b>80,0 %</b>	75,0 %	80,0 %
KL vs. RM	Klasicismus	<b>72,5 %</b>	60,0 %	57,5 %
	Romantismus	<b>77,5 %</b>	77,5 %	35,0 %

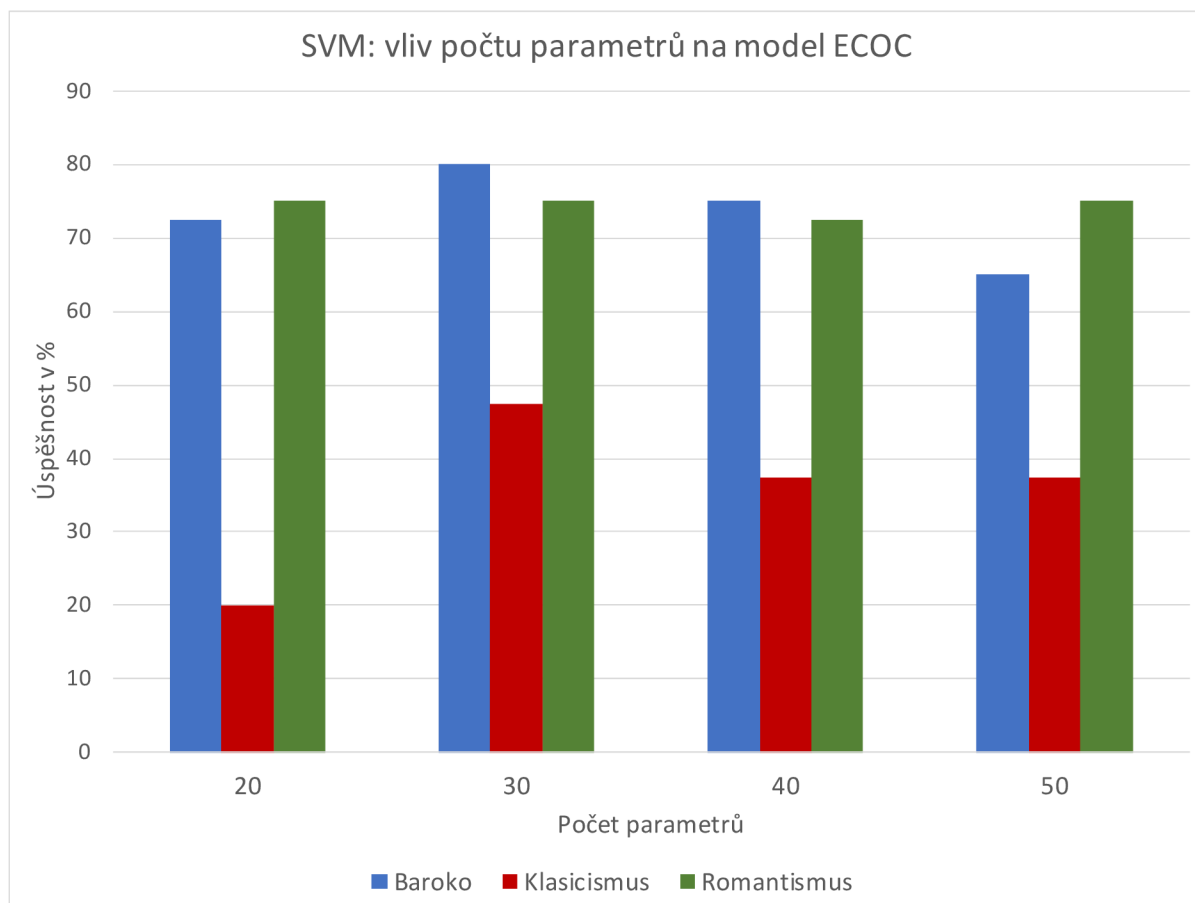
Tab. 7.6: Úspěšnost klasifikátoru SVM v závislosti na počtu parametrů

Z tabulky 7.6 rovněž vyplývá, že pro dvojice slohů, které se vyznačují velkými odlišnostmi (baroko - klasicismus a baroko - romantismus) nehraje počet parametrů příliš velkou roli. Úspěšnost klasifikace zůstává prakticky zachována. U dvojice klasicismus - romantismus, kde tak výrazný rozdíl mezi oběma slohy není, způsobuje postupné přidávání parametrů rapidní pokles úspěšnosti. Jak již bylo zmíněno v teoretické části práce, během přechodu klasicismu do romantismu zůstaly mnohé hudebně-teoretické postupy zachovány. Rapidní pokles je tedy způsoben přidáváním nerelevantních parametrů a jejich redundancí.

### 7.4.1 Error-Correcting Output Codes a SVM

Metoda Error-Correcting Output Codes (dále jen ECOC) se využívá v klasifikaci většího množství tříd při použití binárních klasifikátorů. V případě této práce, byl využit klasifikátor SVM. Metoda ECOC spočívá ve vytvoření matice, kde počet řádků odpovídá počtu tříd a počet sloupců odpovídá délce kódu. K řešení klasifikace více tříd je vytvořen kód o délce  $x$  bitů, obsahující  $x-1$  funkcí. Každá třída je tak charakterizována specifickým kódovým slovem [8].

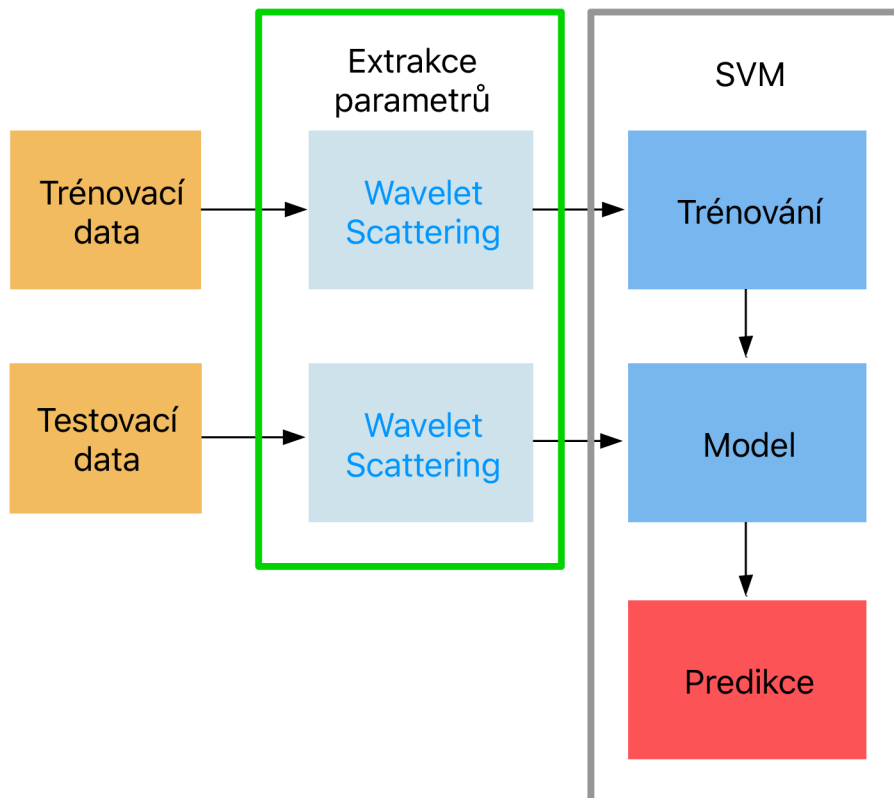
Model klasifikátoru ECOC SVM byl optimalizován pomocí funkce optimalizace hyperparametrů. Z grafu 7.8 je zřejmé, že na klasifikaci má opět značný vliv počet parametrů. Nejlepších výsledků bylo dosaženo se skupinou 30 parametrů. Největším problémem klasifikace pomocí této metody byla správná identifikace hudebních vzorků z období klasicismu. Maximální dosažená úspěšnost klasifikace klasicismu činila 47,5 %. Z povahy metody ECOC lze předpokládat, že rozšířením problému klasifikace klasické hudby o další třídy (např. moderna, impresionismus) se zlepší celková úspěšnost této metody. Pro metodu je podstatná co největší separace tříd a sloupců, která se odvíjí od počtu tříd [8].



Obr. 7.8: Závislost úspěšnosti klasifikátoru ECOC SVM na počtu parametrů

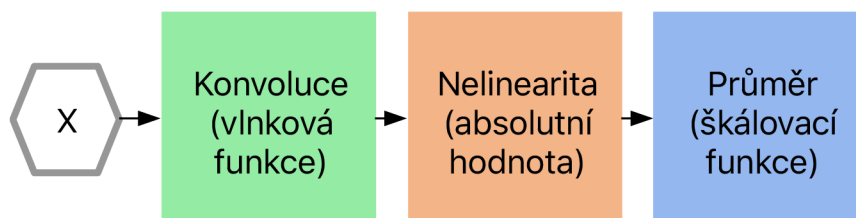
## 7.4.2 Wavelet Scattering

Transformace typu Wavelet Scattering (česky také vlnková transformace rozptylu) je způsob získávání parametrů, vhodný pro datasey s menším množstvím dat. Jedná se o kombinaci vlnkové transformace a filtrace. Díky této kombinaci minimalizuje metoda rozdíly dat uvnitř jedné třídy a maximalizuje rozdíly mezi třídami. Celý proces klasifikace zobrazuje diagram 7.9. Parametry získané z trénovacích dat pomocí metody Wavelet Scattering jsou použity ve vybraném typu klasifikátoru. V případě této práce se jedná o klasifikátor SVM. Proces klasifikace pokračuje poté stejně, jak to bylo popsáno v kapitole 3.1.4. Metoda Wavelet Scattering je použitelná jak v kombinaci se strojovým učením, tak v kombinaci s neuronovými sítěmi. Rozdíl mezi touto metodou a neuronovou sítí je v tom, že použité filtry jsou nastaveny pevně a není sledována jejich odezva.



Obr. 7.9: Klasifikace s využitím metody Wavelet Scattering

Diagram 7.10 ukazuje celý proces získávání parametrů ze vstupních dat. V první fázi projdou vstupní data konvolucí s vlnkou. Tím dojde k vytvoření koeficientů rozptylu nultého řádu, které reprezentují vstupní data. Poté jsou počítány koeficienty vyšších řádů tak, že z výsledku konvoluce je spočítán modul a ten je dále průměrován škálovací funkcí či low-pass filtrem. Platí, že koeficienty nultého řádu popisují energii signálu vstupních dat a jako reprezentace jim také nejvíce odpovídají. Energie koeficientů klesá se zvyšujícím se řádem. Vstupními daty pro výpočet koeficientů prvního řádu jsou průměrované moduly koeficientů nultého řádu. Celý proces výpočtu se tak opakuje znovu, dokud nedojde k úplné ztrátě energie. Metoda Wavelet Scattering odpovídá opakující se více-pásmové filtraci [40].

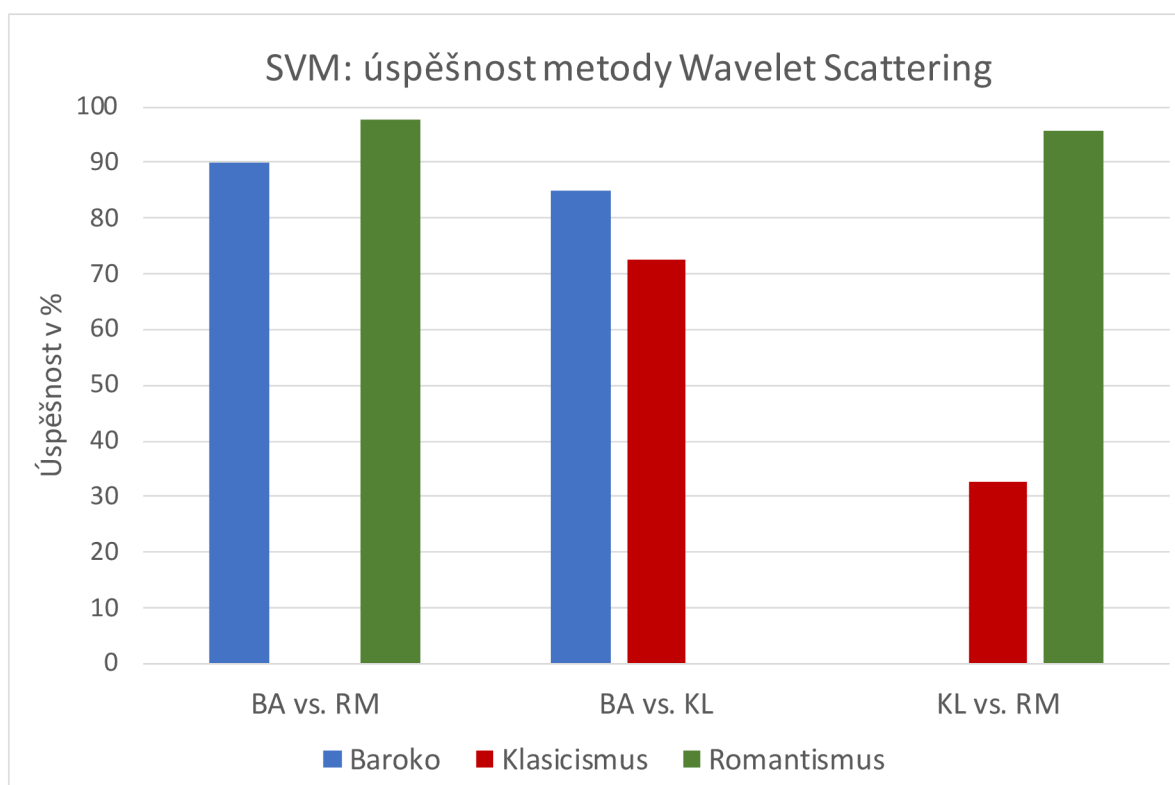


Obr. 7.10: Získávání parametrů metodou Wavelet Scattering



Klasifikátor SVM s parametry, které byly získány pomocí metody Wavelet Scattering, dosahuje velmi vysoké přesnosti v identifikaci barokního a romantického slohu. Identifikace klasicistního slohu je i v tomto případě nejméně přesná. Graf 7.11 ukazuje, že problematická je opět dvojice klasicismus - romantismus, respektive pouze klasicismus. Zatímco hudební vzorky romantického slohu byly identifikovány s přesností 95,5 %, hudební vzorky klasicistního slohu byly správně zařazeny pouze ve 32,5 % případů. Podobně tomu bylo i při hodnocení všech tří tříd s použitím modelu ECOC :

- baroko: 82,5 %,
- klasicismus: 30,0 %,
- romantismus: 92,5 %.



Obr. 7.11: Výsledky klasifikátoru SVM/Wavelet Scattering - binární problém

## 8 Závěr

Cílem této práce je zjistit, jak úspěšné jsou algoritmy umělé inteligence v rozpoznávání slohů z orchestrálních nahrávek na základě parametrů. V první kapitole je nastin vývoje hudebního umění a jsou charakterizovány jednotlivé slohy. Krátce je také pojednáno o tom, které faktory určují sloh v případě rozpoznání posluchačem.

Druhá kapitola je věnována datasetům. První část obsahuje přehled existujících datasetů klasické hudby, které jsou ovšem dostupné pouze ve formě spektrogramů. Na základě toho byl vytvořen nový dataset, který je popsán v druhé části. Úloha rozpoznávání slohů klasické hudby je nutností tvorby vlastního datasetu značně ztížena.

Ve třetí kapitole je popsána metoda Music Information Retrieval, její uplatnění v praxi a rovněž jsou zde teoreticky popsány použité klasifikátory. Ve většině případů se jedná o algoritmy strojového učení. Výběr klasifikátorů obsahuje jak základní typy (kNN, GMM), tak typy které dosahují značné úspěšnosti na poli klasické hudby (SVM, LDA).

Ve čtvrté kapitole jsou popsány parametry, jejich dělení, způsob extrakce z nahrávek a využití. Nejde o seznam parametrů, které byly z datasetu extrahovány. Počet a typ parametrů určila během počáteční analýzy jejich efektivita. V zásadě platí, že pro každý klasifikátor je vhodné jiné množství parametrů.

V páté kapitole jsou uvedeny výsledky některých dosavadních výzkumů. Jedná se o porovnání úspěšnosti klasifikátorů v závislosti na použitých parametrech a datasetech. Tabulka potom poskytuje shrnutí výsledků souvisejících s rozpoznáváním slohů či klasických autorů.

V šesté kapitole je návrh implementace celého projektu. Návrh odpovídá běžnému postupu u metody MIR. Dále šestá kapitola pojednává o získávání parametrů pomocí funkcí z toolboxu MIRtoolbox. Získané parametry bylo nutné evaluovat. Pomocí metody minimum Redundancy Maximum Relevance byly určeny parametry, které jsou nejvíce relevantní, a to jak samostatně, tak v kombinaci s ostatními vybranými parametry. O tom pojednává podkapitola 6.3. Poslední podkapitola pojednává o funkci, která byla použita k optimalizaci klasifikátorů.

V sedmé kapitole jsou uvedeny výsledky všech klasifikátorů. Nejvyšší přesnosti při hodnocení všech slohů dosáhl klasifikátor SVM v kombinaci s modelem ECOC a s parametry získanými pomocí metody Wavelet Scattering. Průměrná úspěšnost mezi všemi slohy zde dosáhla 68,33 %. Při řešení binárních problémů bylo zjištěno, že nejproblématictější dvojice tříd je klasicismus a romantismus. Tyto dvě třídy od sebe nejsou odděleny tak významně, jako dvojice tříd baroko - klasicismus a baroko - romantismus. Toto zjištění se shoduje s hudební teorií slohů uvedené v první kapitole této práce. Svou roli zde hraje také specifická tvorba L. van Beethovena, která je zařazena do klasicismu, ale od tvorby ostatních klasicistních autorů se liší značnou expresivitou a také častým užitím mollových tónin. Dvojice klasicismus - romantismus byla problematická pro většinu kla-

sifikátorů. Nejvyšší přesnosti v identifikaci těchto hudebních vzorků dosáhly klasifikátory LDA a SVM. Oba byly úspěšné na 75,00 %. V klasifikaci u dvojice baroko - klasicismus byl nejuspěšnější klasifikátor kNN s průměrnou úspěšností 86,25 %. U dvojice slohů baroko - romantismus dosáhl nejvyšší úspěšnosti klasifikátor SVM v kombinaci s modelem ECOC a s parametry získanými pomocí metody Wavelet Scattering. Průměrná úspěšnost dosahovala 93,75 %. Dle historického i hudebně-teoretického hlediska je největší rozdíl právě mezi barokem a romantismem. Jsou zde významné rozdíly ve složení orchestru, v práci s rytmikou a dynamikou, v práci s agogikou, v harmonii, ve vedení melodie atd.

V rámci této diplomové práce byla shrnuta hudební teorie charakterizující zkoumané slohy. Rovněž byla shrnuta teorie metody Music Information Retrieval a také teorie všech použitých klasifikátorů. Jedna kapitola byla věnována také teoretickému popisu parametrů získávaných z hudebních vzorků. Dále byly shrnuty dosavadní výsledky výzkumů na poli rozpoznání hudebních žánrů, slohů klasické hudby a identifikace klasických autorů. Poté byl proveden návrh praktického řešení a dle tohoto návrhu bylo řešení zpracováno v prostředí MATLAB. Dosažená úspěšnost klasifikace odpovídá výsledkům prací jiných autorů a výsledek se shoduje s hudebně-teoretickou charakteristikou zkoumaných slohů.

# Literatura

- [1] ALBERT, Marc, K.; David, AHA.: *Analyses of instance-based learning algorithms*. [online]. University of California, Irvine, 1991. [cit. 2019-11-25]. Dostupné z URL: <<https://www.aaai.org/Papers/AAAI/1991/AAAI91-086.pdf>>.
- [2] *Audio Labs: Cross-Composer Dataset* [online]. Erlangen: AudioLabs Erlangen, 2019 [cit. 2019-11-28]. Dostupné z: <<https://www.audiolabs-erlangen.de/resources/MIR/cross-comp>>.
- [3] *Audio Labs: Cross-Era Dataset* [online]. Erlangen: AudioLabs Erlangen, 2019 [cit. 2019-11-28]. Dostupné z: <<https://www.audiolabs-erlangen.de/resources/MIR/cross-era>>.
- [4] BACH, Johann S. a Siegfried W. DEHN.: *Brandenburg Concerto No. 4, BWV1049* [online]. Leipzig: C.F. Peters, 1851 [cit. 2019-12-20]. Dostupné z: <[http://ks.imslp.net/files/imglnks/usimg/3/36/IMSLP457279-PMLP82081-JSBach\\_Brandenburg\\_Concerto\\_No.4,\\_BWV1049\\_SWDehn\\_fullscore.pdf](http://ks.imslp.net/files/imglnks/usimg/3/36/IMSLP457279-PMLP82081-JSBach_Brandenburg_Concerto_No.4,_BWV1049_SWDehn_fullscore.pdf)>.
- [5] BEETHOVEN, Ludwig.: *Symphony No. 3 Eroica, Op. 55* [online]. 2. vydání. Mineola: Dover Publications, 1989 [cit. 2019-12-20]. Dostupné z: <<http://ks.imslp.net/files/imglnks/usimg/3/39/IMSLP504077-PMLP2581-combinepdf.pdf>>.
- [6] BURRED, Juan, José; Alexander LERCH.: *Hierarchical Automatic Audio Signal Classification*. [online]. Technical University Berlin, Berlin, 2004. [cit. 2019-11-26]. Dostupné z URL: <[https://www.researchgate.net/publication/200054331\\_Hierarchical\\_Automatic\\_Audio\\_Signal\\_Classification](https://www.researchgate.net/publication/200054331_Hierarchical_Automatic_Audio_Signal_Classification)>.
- [7] CAPÓ CLAR, Francesc.: *Western Classical Composer Identification Using Symbolic Data* [online]. Španělsko, 2015 [cit. 2019-11-03]. Dostupné z: <<http://mtg.upf.edu/node/3347>>. Diplomová práce. Universitat Pompeu Fabra Barcelona.
- [8] DIETTERICH, Thomas a Ghulum BAKIRI.: *Solving Multiclass Learning Problems via Error-correcting Output Codes*. Journal of Artificial Intelligence Research. 1995, 94(2), 263-286. ISSN 1076 - 9757 [cit. 2020-05-20].
- [9] DVOŘÁK, Antonín.: *Symphony No. 9, Op. 95* [online]. Praha: SNKLHU, 1955 [cit. 2019-12-20]. Dostupné z: <<http://imslp.simssa.ca/files/imglnks/usimg/a/a3/IMSLP547268-PMLP8710-dvoraksymphony9color.pdf>>.
- [10] GEERTZEN, Jeroen.: *Composer Classification Using Grammatical Inference*. In: Proceedings of the MML International Workshop on Machine Learning and Music. Helsinki: MML International Workshop, 2008, s. 17–18.

- [11] HAGGBLADE, Michael, Yang HONG a Kenny KAO.: *Music Genre Classification*. In: CS 229 Machine Learning Course - Final Projects [online]. Stanford University: Department of Computer Science, 2011 [cit. 2019-10-25]. Dostupné z: <<http://cs229.stanford.edu/proj2011/HaggbladeHongKao-MusicGenreClassification.pdf>>.
- [12] HILLEWAERE, Ruben, Bernard MANDERICK a Darrell CONKLIN.: *String Quartet Classification with Monophonic Models*. In: Proceedings of the 11th International Society for Music Information Retrieval Conference. Utrecht: ISMIR, 2010, s. 537–542. ISBN 978-90-393-53813.
- [13] HONINGH, Aline a Rens BOD.: *Clustering and Classification of Music by Interval Categories*. In: Proceedings of the Third International Conference on Mathematics and Computation in Music. Berlin: Springer-Verlag, 2011, s. 346–349. ISBN 978-3642215896.
- [14] HONTANILLA, Maria a Carlos PERÉZ-SANCHO.: *Modeling Musical Style with Language Models for Composer Recognition*. In: Pattern Recognition and Image Analysis. Madeira: Springer, 2013, s. 740–748. ISBN 978-3-642-38628-2.
- [15] *ISMIR* [online]. ISMIR, 2019 [cit. 2019-12-03]. Dostupné z: <<http://ismir.net/resources/datasets/>>.
- [16] IZMIRLI, Ozgur.: *Tonal-Atonal Classification of Music Audio Using Diffusion Maps*. In: Proceedings of the 10th International Society for Music Information Retrieval Conference. Kobe: ISMIR, 2009, s. 687–691. ISBN 978-0-9813537-0-8.
- [17] KNEES, Peter a Markus SCHEDL.: *Music Similarity and Retrieval*. eBook. Berlin: Springer, 2016. ISBN 978-3-662-49722-7.
- [18] KOZAKOWSKI, Piotr a Bartosz MICHALAK.: *Music Genre Recognition*. In: DeepSound [online]. Varšava, Polsko, 2016 [cit. 2019-11-03]. Dostupné z: <[http://deepsound.io/music\\_genre\\_recognition.html](http://deepsound.io/music_genre_recognition.html)>.
- [19] LARTILLOT, Olivier.: *MIRtoolbox 1.7 User's Manual* [online]. Oslo: University of Oslo, Department of Musicology, 2017 [cit. 2019-12-02]. Dostupné z: <<https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7.pdf>>.
- [20] LI, Tao, Mitsunori OGIHARA a George TZANETAKIS.: *Music Data Mining*. eBook. Boca Raton: CRC Press, 2011. ISBN 978-1-4398-3552-4.

- [21] LIEM, Cynthia C. S. a Alan HANJALIC.: *Comparative Analysis of Orchestral Performance Recordings: An Image-Based Approach*. In: Proceedings of the 16th International Society for Music Information Retrieval Conference [online]. Málaga, Španělsko: International Society for Music Information Retrieval, 2015, s. 302–308 [cit. 2019-11-03]. ISBN 987-84-606-8853-2. Dostupné z: <[http://ismir2015.uma.es/docs/ISMIR2015\\_Proceedings.pdf](http://ismir2015.uma.es/docs/ISMIR2015_Proceedings.pdf)>.
- [22] LI, Stan.; GuoDong GUO.: *Content-Based Audio Classification and Retrieval Using SVM Learning*. [online]. Microsoft Research China, Beijing, 2003. [cit. 2019-11-05]. Dostupné z URL: <<https://pdfs.semanticscholar.org/b5e8/a947b46424d54c1ea2a44abb154b4eb2857d.pdf>>.
- [23] MARCHAND, Ugo a Geoffroy PEETERS.: *The Modulation Scale Spectrum and its Application to Rhythm-Content Description*. In: 17th International Conference on Digital Audio Effects (DAFx) [online]. Erlangen, Německo, 2014 [cit. 2019-10-25]. Dostupné z: <[http://www.dafx14.fau.de/papers/dafx14\\_ugo\\_marchand\\_the\\_modulation\\_scale\\_spec.pdf](http://www.dafx14.fau.de/papers/dafx14_ugo_marchand_the_modulation_scale_spec.pdf)>.
- [24] MARCOS, Germán R.: *Audio Tag Classification: Classical Composer*. Music Information Retrieval: Student blog for the MIR course, Master in Sound and Music Computing [online]. Barcelona, Španělsko: Universitat Pompeu Fabra, 2017 [cit. 2019-11-03]. Dostupné z: <<https://musicinformationretrieval.wordpress.com/2017/03/13/classical-composer-identification-summary-and-updates/>>.
- [25] *MATLAB Documentation: Bayesian Optimization Algorithm* [online]. United States: MathWorks, 2020 [cit. 2020-05-24]. Dostupné z: <<https://www.mathworks.com/help/stats/bayesian-optimization-algorithm.html>>.
- [26] MCKAY, Cory a Ichiro FUJINAGA.: *Automatic Genre Classification Using Large High-Level Musical Feature Sets*. In: Proceedings of the 5th International Society for Music Information Retrieval Conference. Barcelona: ISMIR, 2004, s. 525–530.
- [27] MEARNS, Lesley, Dan TIDHAR a Simon DIXON.: *Characterisation of Composer Style Using High-level Musical Features*. In: Proceedings of the 8th Sound and Music Computing Conference. Padova: SMC, 2010, s. 37–40.
- [28] MEDHAT, Fady, David CHESMORE a John ROBINSON.: *Automatic Classification of Music Genre Using Masked Conditional Neural Networks*. IEEE International Conference on Data Mining (ICDM) [online]. IEEE, 2017, , s. 979–984 [cit. 2019-10-25]. DOI: 10.1109/ICDM.2017.125. ISSN 2374-8486. Dostupné z: <<https://ieeexplore.ieee.org/abstract/document/8215588>>.

- [29] MOZART, Wolfgang A.: *Symphony No. 40 in G minor K.550*. in Wolfgang Amadeus Mozarts Werke, Serie VIII: Symphonien, Bd.3, No.40 [online]. Leipzig: Breitkopf & Härtel, 1880, s. 181–229 [cit. 2019-12-20]. Dostupné z: <[http://imslp.simssa.ca/files/imglnks/usimg/f/f2/IMSLP492601-PMLP1572-IMSLP00072-Mozart\\_-\\_Symphony\\_No\\_40\\_in\\_G\\_minor,\\_K550.pdf](http://imslp.simssa.ca/files/imglnks/usimg/f/f2/IMSLP492601-PMLP1572-IMSLP00072-Mozart_-_Symphony_No_40_in_G_minor,_K550.pdf)>.
- [30] *MRMR: minimum Redundancy Maximum Relevance Feature Selection* [online]. Nanjing, China: Hanchuan Peng, 2005 [cit. 2020-05-12]. Dostupné z: <<http://home.penglab.com/proj/mRMR/>>.
- [31] MÜLLER, Meinard a Sebastian EWERT.: *Chroma Toolbox: Matlab Implementations for Extracting Variants of Chroma-Based Audio Features*. In: Proceedings of the 12th International Society for Music Information Retrieval Conference [online]. Miami, Florida, USA: University of Miami, 2011, s. 215–220 [cit. 2019-10-31]. ISBN 978-0615548654. Dostupné z: <[https://www.researchgate.net/publication/220723140\\_Chroma\\_Toolbox\\_Matlab\\_Implementations\\_for\\_Extracting\\_Variants\\_of\\_Chroma-Based\\_Audio\\_Features](https://www.researchgate.net/publication/220723140_Chroma_Toolbox_Matlab_Implementations_for_Extracting_Variants_of_Chroma-Based_Audio_Features)>.
- [32] *MusicNet: A curated collection of labeled classical music* [online]. Washington: University of Washington, 2017 [cit. 2019-11-28]. Dostupné z: <<https://homes.cs.washington.edu/~thickstn/musicnet.html>>.
- [33] ORAMAS, Sergio, Oriol NIETO, Francesco BARBIERI a Xavier SERRA.: *Multi-Label Music Genre Classification from Audio, Text and Images Using Deep Features*. In: Proceedings of the 18th International Society of Music Information Retrieval Conference [online]. Suzhou, Čína: International Society for Music Information Retrieval, 2017, s. 23–30 [cit. 2019-11-03]. ISBN 978-981-11-5179-8. Dostupné z: <[https://ismir2017.smcnus.org/wp-content/uploads/2017/10/126\\_Paper.pdf](https://ismir2017.smcnus.org/wp-content/uploads/2017/10/126_Paper.pdf)>.
- [34] PEETERS, Geoffroy.: *Spectral and Temporal Periodicity Representations of Rhythm for the Automatic Classification of Music Audio Signal*. IEEE Transactions on Audio, Speech, and Language Processing [online]. IEEE, 2011, 19(5), s. 1242–1252 [cit. 2019-10-25]. DOI: 10.1109/TASL.2010.2089452. ISSN 1558-7924. Dostupné z: <<https://ieeexplore.ieee.org/document/5607297>>.
- [35] PENG, Hanchuan, Fuhui LONG a Chris DING.: *Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy*. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. 2005, 27(08), 1226-1238. [cit. 2020-04-06] DOI: 10.1109/TPAMI.2005.159.
- [36] PONCE DE LEÓN, Pedro J. a Jose M. INESTA.: *Pattern Recognition Approach for Music Style Identification Using Shallow Statistical Descriptors*. In: IEEE Transactions on Systems Man and Cybernetics Part C [online]. IEEE Systems, 2007, 37(2),

- s. 248–257 [cit. 2019-10-29]. DOI: 10.1109/TSMCC.2006.876045. ISSN 1094-6977. Dostupné z: <[https://www.researchgate.net/publication/3421763\\_Pattern\\_Recognition\\_Approach\\_for\\_Music\\_Style\\_Identification\\_Using\\_Shallow\\_Statistical\\_Descriptors](https://www.researchgate.net/publication/3421763_Pattern_Recognition_Approach_for_Music_Style_Identification_Using_Shallow_Statistical_Descriptors)>.
- [37] SCHEDL, Markus, Emilia GÓMEZ a Julián URBANO.: *Music Information Retrieval: Recent Developments and Applications*. Foundations and Trends in Information Retrieval. Vol.8. Delft: now Publishers, 2014, s. 127–261. ISBN 978-1-60198-807-2.
- [38] SCHNIERER, Miloš.: *Dějiny hudby*. 4. vydání. Brno: Janáčkova akademie múzických umění v Brně, 2007. ISBN 80-86928-19-5.
- [39] SMOLKA, Jaroslav.: *Dějiny hudby*. 1. vydání. Brno: TOGGA agency, 2001. ISBN 80-902912-0-1.
- [40] SORO, Bedionita a Chaewoo LEE.: *A Wavelet Scattering Feature Extraction Approach for Deep Neural Network Based Indoor Fingerprinting Localization*. Sensors. 2019, 2019(19), 1790 [cit. 2019-05-21]. DOI: 10.3390/s19081790. ISSN 1424-8220.
- [41] SYROVÝ, Václav.: *Hudební akustika*. 3. doplněné vydání. Praha: Akademie múzických umění v Praze, 2013. ISBN 978-80-7331-297-8.
- [42] VIVALDI, Antonio a Gian F. MALIPIERO.: *Le quattro stagioni: Concerto in sol minore* [online]. Milano: G. Ricordi & Co., 1950 [cit. 2019-12-20]. Dostupné z: <[http://imslp.eu/files/imglnks/euimg/c/c9/IMSLP386648-PMLP126433-Vivaldi,\\_Antonio-Opere\\_Ricordi\\_F\\_I\\_No\\_23\\_scan.pdf](http://imslp.eu/files/imglnks/euimg/c/c9/IMSLP386648-PMLP126433-Vivaldi,_Antonio-Opere_Ricordi_F_I_No_23_scan.pdf)>.
- [43] VAN KRANENBURG, Peter a Eric BACKER.: *Musical Style Recognition – a Quantitative Approach*. In: Proceedings of the Conference on Interdisciplinary Musicology. Graz: CIM, 2004, s. 106–107. ISBN 3-200-00113-5.
- [44] VOGLER, Benedikt, Amir OTHMAN.: *Music Genre Recognition*. [online]. Bauhaus-Universität Weimar, Weimar, 2016. [cit. 2019-12-06]. Dostupné z URL: <<https://benediktsvogler.com/downloads/DocumentationAudiotechnikMusicGenreRecognition.pdf>>.
- [45] *What Is a Neural Network?* [online]. USA: The MathWorks, 2019 [cit. 2019-12-07]. Dostupné z: <<https://www.mathworks.com/discovery/neural-network.html>>.
- [46] WEISS, Christof.: *Computational Methods for Tonality-Based Style Analysis of Classical Music Audio Recordings* [online]. Ilmenau, 2017 [cit. 2019-11-03]. Dostupné z: <[https://www.db-thueringen.de/receive/dbt\\_mods\\_00032890](https://www.db-thueringen.de/receive/dbt_mods_00032890)>. Diplomová práce. Ilmenau University of Technology.



- [47] WEISS, Christof, Matthias MAUCH a Simon DIXON.: *Timbre-invariant Audio Features for Style Analysis of Classical Music*. In: Proceedings of the ICM-C/SMC 2014 [online]. Athens, 2014, s. 1461–1468 [cit. 2019-10-31]. ISBN 978-960-7313-28-7. Dostupné z: <<https://quod.lib.umich.edu/cgi/p/pod/dod-idx/timbre-invariant-audio-features-for-style-analysis.pdf>>.
- [48] YASLAN, Yusuf a Zehra CATALTEPE.: *Audio Music Genre Classification Using Different Classifiers and Feature Selection Method*. 18th International Conference on Pattern Recognition (ICPR'06) [online]. IEEE, 2006, 18(2), 573-576 [cit. 2019-10-25]. DOI: 10.1109/ICPR.2006.282. ISSN 1051-4651. Dostupné z: <<https://ieeexplore.ieee.org/document/1699270>>.
- [49] ZENKL, Luděk.: *ABC hudebních forem*. 4. vydání. Praha: Editio, 2014. ISBN 978-80-86385-33-4.

# Seznam symbolů, veličin a zkratk

<b>AI</b>	Artificial Intelligence (česky: umělá inteligence)
<b>ANN</b>	Artificial Neural Network (česky: umělá neuronová síť)
<b>BA</b>	baroko
<b>BPM</b>	Beats Per Minute (česky: počet úderů za minutu)
<b>ECOC</b>	Error-Correcting Output Codes model
<b>FFT</b>	Fast Fourier Transform (česky: rychlá fourierova transformace)
<b>FP</b>	Fluctuation Pattern (česky: parametr fluktuace)
<b>GMM</b>	Gaussian Mixture Models (česky: Gaussovy smíšené modely)
<b>KL</b>	klasicismus
<b>k-NN</b>	K-Nearest-Neighbor (česky: K-nejbližší sused)
<b>LDA</b>	Linear Discriminant Analysis (česky: lineárně diskriminační analýza)
<b>MFCC</b>	Mel Frequency Cepstral Coefficients (česky: mel-frekvenční keprální koeficienty)
<b>MIDI</b>	Musical Instrument Digital Interface
<b>MIR</b>	Music Information Retrieval
<b>mRMR</b>	minimum Redundancy Maximum Relevance
<b>RM</b>	romantismus
<b>RSM</b>	Root-Mean-Square (česky: efektivní hodnota)
<b>SVM</b>	Support Vector Machines (česky: podpůrné vektory)
<b>WST</b>	Wavelet Scattering Transform

# Seznam příloh

A Seznam parametrů	75
B Úspěšnost klasifikátorů	77
C Obsah přiloženého CD	79

# A Seznam parametrů

Tab. A.1: Seznam parametrů vybraných metodou mRMR

Pořadí	Název	Upřesnění	Oblast
01	Tonal Centroid	4. dimenze	Tonalita
02	Brightness	3 kHz, Global Description	Spektrum
03	Roll-Off	95 %, Global Description	Spektrum
04	Decay Leap	Median	Spektrum
05	RMS	Median	Dynamika
06	Brightness	1,5 kHz	Spektrum
07	Mode	Global Description	Tonalita
08	Irregularity	Median	Spektrum
09	Roll-Off	85 %, Median	Spektrum
10	Spectral Centroid	Global Description	Spektrum
11	Spectral Spread	Median	Spektrum
12	Skewness	Median	Spektrum
13	Irregularity	Global Description	Spektrum
14	Tonal Centroid	5. dimenze	Tonalita
15	Spectral Spread	Global Description	Spektrum
16	Roll-Off	85 %, Mean	Spektrum
17	Tonal Centroid	3. dimenze	Tonalita
18	Attack Slope	Mean	Spektrum
19	Tonal Centroid	3. dimenze	Tonalita
20	Brightness	1,5 kHz, Mean	Spektrum
21	Decay Leap	Mean	Spektrum
22	Activity (Spectral Flux)	Median	Rytmus
23	Zero Cross	Global Description	Spektrum
24	Roll-Off	85 %, Global Description	Spektrum
25	Low Energy	Average Silence Ratio	Dynamika
26	Flatness	Median	Spektrum
27	RMS	Mean	Dynamika
28	Fluctuation	Standard Deviation	Rytmus
29	Decay Slope	Standard Deviation	Spektrum
30	Kurtosis	Median	Spektrum
31	Inharmonicity	Global Description	Tonalita
32	RMS	Mean, Global Description	Dynamika

Pořadí	Název	Upřesnění	Oblast
33	Tonal Centroid	1. dimenze	Tonalita
34	Brightness	1,5 kHz, Global Description	Spektrum
35	Brightness	1 kHz, Global Description	Spektrum
36	RMS	Median, Global Description	Dynamika
37	Spectral Flux	Median	Rytmus
38	Tempo	Global Description	Rytmus
39	Attack Slope	Median	Spektrum
40	Activity (Spectral Flux)	Mean	Rytmus
41	Roll-Off	85 %, Quasi-Silent Frames, Mean	Spektrum
42	Fullness (Spectral Flux)	Mean	Rytmus
43	Flatness	Global Description	Spektrum
44	Roll-Off	85 %, Quasi-Silent Frames, Median	Spektrum
45	Irregularity	Standard Deviation	Spektrum
46	Mode	Sum	Tonalita
47	Flatness	Mean	Spektrum
48	Decay Slope	Median	Spektrum
49	MFCC	1	MFCC
50	Fullness (Spectral Flux)	Variance	Rytmus

### Legenda

- Dimenze - tonální těžiště (*Tonal Centroid*) je pro každý hudební vzorek počítáno jako šestidimenzionální veličina
- Global Description - parametr byl počítán globálně pro celý hudební vzorek [19]
- Mean, Median, Standard Deviation, Variance - statistické úpravy, viz. kapitola 6.2

## B Úspěšnost klasifikátorů

Tab. B.1: Celkové výsledky

Metoda	Baroko %	Klasicismus %	Romantismus %
kNN: 3 slohy	75,0	55,0	72,5
kNN: sloh X	67,5	40,0	42,5
kNN: BA vs. KL	85,0	87,5	-
kNN: BA vs. RM	77,5	-	80,0
kNN: KL vs. RM	-	72,5	67,5
LDA: 3 slohy	55,0	62,5	72,5
LDA: BA vs. KL	67,5	87,5	-
LDA: BA vs. RM	65,0	-	82,5
LDA: KL vs. RM	-	72,5	77,5
GMM: 3 slohy	70,0	35,0	62,5
GMM: BA vs. KL	70,0	77,5	-
GMM: BA vs. RM	75,0	-	80,0
GMM: KL vs. RM	-	45,0	65,0
SVM: sloh X	80,0	70,5	65,0
SVM: BA vs. KL	82,5	87,5	-
SVM: BA vs. RM	80,0	-	80,0
SVM: KL vs. RM	-	72,5	77,5
SVM: ECOC	80,0	47,5	75,0
SVM+WST+ECOC	82,5	30,0	92,5
SVM+WST: BA vs. KL	85,0	72,5	-
SVM+WST: BA vs. RM	90,0	-	97,5
SVM+WST: KL vs. RM	-	32,5	95,5

### Legenda zkratek

- kNN: K-Nearest-Neighbor
- BA: Baroko
- KL: Klasicismus
- RM: Romantismus
- LDA: Linear Discriminant Analysis
- GMM: Gaussian Mixture Models
- SVM: Support Vector Machines
- ECOC: Error-Correcting Output Codes

- WST: Wavelet Scattering Transform
- MFCC: Mel Frequency Cepstral Coefficients

## C Obsah přiloženého CD

- Diplomova prace Jelinkova.pdf - kompletní text diplomové práce
- Parametry.m - skript v prostředí Matlab, slouží k výpočtu parametrů
- mRMR - metoda pro evaluaci parametrů