

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2019

Bc. Šárka Zemánková



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

ROZPOZNÁVÁNÍ HUDEBNÍHO ŽÁNRU ZA POMOCI TECHNIK MUSIC INFORMATION RETRIEVAL

MUSIC GENRE RECOGNITION USING MUSIC INFORMATION RETRIEVAL TECHNIQUES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Šárka Zemánková

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Tomáš Kiska

BRNO 2019

Diplomová práce

magisterský navazující studijní obor **Audio inženýrství**

Ústav telekomunikací

Studentka: Bc. Šárka Zemánková

ID: 174479

Ročník: 2

Akademický rok: 2018/19

NÁZEV TÉMATU:

Rozpoznávání hudebního žánru za pomoci technik Music Information Retrieval

POKyny PRO VYPRACOVÁNÍ:

V rámci této práce budou shrnuty dosavadní poznatky z oblasti zvané Music Information Retrieval. Konkrétně pak poznatky věnující se rozpoznávání hudebního žánru. Bude sestavena databáze nahrávek, na které bude toto rozpoznávání testováno. Dále pak budou analyzovány hudební nahrávky z hlediska barvy zvuku, rytmiky a dynamiky a vybrány takové hudební parametry, které mají největší diskriminační sílu rozpoznat jednotlivé hudební žánry. V neposlední řadě bude vyhodnocena klasifikační úspěšnost u jednotlivých druhů těchto hudebních žánrů.

DOPORUČENÁ LITERATURA:

[1] TZANETAKIS, G. a P. COOK, 2002. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing[online]. 10(5), 293-302 [cit. 2017-11-02]. DOI: 10.1109/TSA.2002.800560. ISSN 1063-6676.

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

Termín zadání: 1.2.2019

Termín odevzdání: 16.5.2019

Vedoucí práce: Ing. Tomáš Kiska

Konzultant:

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

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

Tato diplomová práce se zabývá rozpoznáváním hudebního žánru za pomoci technik Music Information Retrieval. Je zde stručně popsána podstata této oblasti výzkumu i její podobor zvaný Music Genre Recognition. Následující kapitola obsahuje výběr nejvhodnějších parametrů pro určení hudebního žánru. Dále jsou v této práci popsány metody strojového učení, využívané v této oblasti. Další kapitola se věnuje popisu databází nahrávek vytvořených pro výzkumy žánrové klasifikace. Následuje návrh a implementace vyhodnocovacího systému pro rozpoznávání hudebního žánru. V poslední části práce jsou popsány výsledky analýzy dílčích parametrů, závislost přesnosti žánrové klasifikace na množství využitých parametrů a diskutovány příčiny úspěšnosti zařazení jednotlivých žánrů.

KLÍČOVÁ SLOVA

Music Information Retrieval, Music Genre Recognition, zvukový signál, zvukové parametry, hudební žánr, strojové učení

ABSTRACT

This diploma work deals with music genre recognition using the techniques of Music Information Retrieval. It contains a brief description of the principle of this research area and its subfield called Music Genre Recognition. The following chapter includes selection of the most suitable parameters for describing music genres. This work further characterizes machine learning methods used in this field of research. The next chapter deals with the descriptions of music datasets created for genre classification studies. Subsequently, there is a draft and evaluation of the system for music genre recognition. The last part of this work describes the results of partial parameter analysis, dependence of genre classification accuracy on the amount of parameters and contains a discussion on the causes of classification accuracy for the individual genres.

KEYWORDS

Music Information Retrieval, Music Genre Recognition, audio signal, audio features, music genre, machine learning

ZEMÁNKOVÁ, Šárka. *Rozpoznávání hudebního žánru za pomoci technik Music Information Retrieval*. Brno, 2019, 72 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: Ing. Tomáš Kiska

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Rozpoznávání hudebního žánru 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 popsaný v této diplomové práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

Brno

.....

podpis autorky



EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI



Obsah

Úvod	11
1 Music Information Retrieval	12
1.1 Výška tónu	12
1.2 Časový aspekt	12
1.3 Harmonie	12
1.4 Témbr	13
1.5 Vydavatelský aspekt	13
1.6 Textový aspekt	13
1.7 Bibliografický aspekt	13
2 Rozpoznávání hudebního žánru	15
2.1 Žánr	15
2.2 Výzkumy zabývající se rozpoznáváním hudebního žánru	16
3 Parametrizace hudebního signálu	20
3.1 Parametry využívané pro určení žánru	20
3.1.1 Parametry založené na dynamice	21
3.1.2 Parametry založené na rytmu	21
3.1.3 Parametry založené na témbru	22
4 Metody strojového učení	25
4.1 AdaBoost	25
4.2 Gaussův smíšený model	26
4.3 k-NN	27
4.4 Lineární diskriminační analýza	27
4.5 Náhodné lesy	28
4.6 Naivní Bayesův klasifikátor	28
4.7 Skrytý Markovův model	29
4.8 Support Vector Machines	29
4.9 Umělé neuronové sítě	30
5 Datasetsy	31
5.1 GTZAN	31
5.2 ISMIR2004	32
5.3 Free Music Archive	33
5.4 Million Song Dataset	33
5.5 Unique Dataset	34

6	Návrh vyhodnocovacího systému	35
7	Evaluace systému	37
7.1	Volba parametrů	37
7.2	Výběr metody klasifikace	38
8	Testování a vyhodnocení	39
8.1	Databáze ISMIR2004	39
8.1.1	Individuální analýza parametrů	39
8.1.2	Přesnost klasifikace v závislosti na počtu parametrů	39
8.1.3	Další klasifikační metody	42
8.2	Unique dataset	43
8.2.1	Individuální analýza parametrů	43
8.2.2	Přesnost klasifikace v závislosti na počtu parametrů	43
8.2.3	Další klasifikační metody	46
8.3	Free Music Archive	46
8.3.1	Individuální analýza parametrů	46
8.3.2	Přesnost klasifikace v závislosti na počtu parametrů	47
8.3.3	Další klasifikační metody	49
8.4	Shrnutí výsledků	49
9	Závěr	51
	Literatura	53
	Seznam symbolů, veličin a zkratk	60
	Seznam příloh	61
A	Obsah přiloženého CD	62
B	Výzkumy zabývající se rozpoznáváním žánru	63
C	Seznam vypočítaných parametrů	64
D	Individuální analýza parametrů	65
D.1	ISMIR2004	65
D.2	Unique	67
D.3	Free Music Archive	70

Seznam obrázků

4.1	Metoda 3-NN	27
5.1	Žánrové složení databáze GTZAN	31
5.2	Žánrové složení databáze ISMIR2004	32
5.3	Žánrové složení databáze Unique dataset	34
6.1	Blokové schéma vyhodnocovacího systému	35
8.1	Úspěšnost klasifikace jednotlivých žánrů databáze ISMIR2004	40
8.2	Celková úspěšnost klasifikace databáze ISMIR2004	41
8.3	Úspěšnost klasifikace jednotlivých žánrů databáze Unique	44
8.4	Celková úspěšnost klasifikace databáze Unique	45
8.5	Žánrová úspěšnost klasifikace databáze Free Music Archive	47
8.6	Celková úspěšnost klasifikace databáze Free Music Archive	48

Seznam tabulek

1.1	Příklady využití MIR	14
7.1	Seznam statistických parametrů	37
7.2	Předpokládaná úspěšnost metod strojového učení	38
8.1	Přesnost rozřazení datasetu ISMIR2004 udaná v procentech	42
8.2	Klasifikace databáze ISMIR2004 dalšími metodami strojového učení	42
8.3	Přesnost rozřazení datasetu Unique udaná v procentech	45
8.4	Klasifikace databáze Unique dalšími metodami strojového učení	46
8.5	Přesnost rozřazení datasetu Free Music Archive udaná v procentech	49
8.6	Klasifikace databáze Free Music Archive dalšími metodami strojového učení	49
B.1	Výzkumy zabývající se rozpoznáváním žánru	63
C.1	Seznam vypočítaných parametrů	64
D.1	Individuální analýza parametrů databáze ISMIR2004	65
D.2	Individuální analýza parametrů databáze Unique	67
D.3	Individuální analýza parametrů databáze Free Music Archive	70

Úvod

Tato práce pojednává o rozpoznávání hudebního žánru za pomoci technik Music Information Retrieval (MIR).

Populární hudba se začala rozvíjet na počátku 20. století. V průběhu let docházelo k proměně jejích formálních i obsahových charakteristik. Z tohoto důvodu bylo nutné nalézt způsob třídění a uspořádání hudebních děl. Nejběžnějším způsobem klasifikace populární hudby je rozdělení do hudebních žánrů. Skladby, náležící jednomu žánru, obvykle spojuje podobná instrumentace, typický rytmus, obsah textu i emoce, které u posluchače vyvolávají. Určení hudebního žánru je však subjektivní, každý jedinec jej může vnímat jinak. Mnoho skladeb se také nachází na pomezí více žánrů.

Zejména v důsledku rozšíření internetu narůstá v současné době využívání digitálních zvukových souborů a s nimi i potřeba jejich kategorizace. K tomu lze využít automatické rozpoznávání hudebního žánru (MGR), tedy odvětví MIR. Tato oblast výzkumu slouží ke snadnějšímu vyhledávání požadovaného typu hudby, sdružování nahrávek se stejnými vlastnostmi do skupin nebo návrhy a doporučení podobných skladeb.

Cílem této práce je shrnutí dosavadních poznatků z oblasti Music Information Retrieval, zejména jejího odvětví, které se týká rozpoznávání hudebního žánru. Dále budou popsány parametry, stěžejní pro určení žánru. Následuje charakteristika metod strojového učení běžně využívaných v této oblasti. Budou zde také uvedeny databáze nahrávek, na kterých probíhá testování zvolené metody rozpoznávání žánru. Poté dojde k vytvoření návrhu vyhodnocovacího systému a jeho realizaci. Závěr práce se věnuje vyhodnocení výsledků žánrové klasifikace a diskuzi ohledně přesnosti zařazení jednotlivých žánrů.

1 Music Information Retrieval

Music Information Retrieval (MIR) je oblast výzkumu, zabývající se získáváním, analýzou a zpracováním informací hudby, reprezentované různými způsoby (zvukový signál, MIDI, jméno autora apod.) [45]. Tato oblast se začala rozvíjet během 90. let 20. století. K výraznému pokroku došlo na začátku nového tisíciletí v důsledku založení organizace ISMIR (The International Society for Music Information Retrieval), která každoročně pořádá konference, týkající se zpracování a organizování hudebních dat [20].

Podle S. J. Downieho [13] sestávají hudební informace ze sedmi aspektů. Těmi jsou výška tónu, harmonie, tónbr, časové, textové, vydavatelské a bibliografické aspekty. Tato hlediska se v mnoha ohledech vzájemně prolínají, nevylučují se. Jako příklad lze uvést harmonii, která je souhrou časových aspektů a výšky tónu.

1.1 Výška tónu

Výšku tónu lze definovat jako vnímanou vlastnost zvuku, která je funkcí fundamentální frekvence zvuku, určené počtem kmitů za sekundu [43]. Výšku tónu je možné zadat pomocí grafické reprezentace, názvu noty, čísla stupnice apod. Rozdíl mezi dvěma výškami tónu se nazývá interval. Tuto vzdálenost lze označit počtem půltónů, nacházejících se mezi danými tóny, nebo za pomoci názvosloví, vycházejícího z tradice hudební teorie (např. velká tercie, čistá kvarta) [13].

1.2 Časový aspekt

Časový aspekt popisuje rytmickou složku hudebního díla, kterou tvoří metrum, tempo, délka trvání tónu, délka trvání harmonie a akcenty. Vyjádření časových informací může být absolutní (např. indikací metronomu $MM = 80$), obecné (*adagio*, *presto*) a relativní (*rychleji*, *pomaleji*) [13].

1.3 Harmonie

Jako harmonie je označována situace, kdy současně znějí dva nebo více tónů různých výšek. Tento stav je také znám jako polyfonie. V rámci hudební teorie byly za pomoci základní frekvence a intervalů mezi ostatními znějícími tóny definovány harmonické události. Nejběžnější z nich je *akord* [13].

1.4 Témbr

Témbrální aspekt zahrnuje informace o barvě zvuku. Díky tomu je ucho schopno vzájemně rozlišit jednotlivé hudební nástroje. Z toho důvodu jsou pro tento aspekt důležité informace o instrumentaci, ačkoliv bývají tyto údaje často součástí bibliografického aspektu. Na témbr má vliv také způsob interpretace hudebního díla, neboť se barva zvuku liší při různé dynamice provedení i při využití dusítek nebo rozmanitých stylů hry smyčcem. Zde se témbrální aspekt prolíná s aspektem vydavatelským [13].

1.5 Vydavatelský aspekt

Pokyny, týkající se způsobu provedení hudebního díla, jsou hlavním obsahem vydavatelského aspektu. Tyto instrukce zahrnují prstoklad, hudební ozdoby, dynamické značky, artikulaci, způsoby hry smyčcem a mnohé další. Vydavatelský aspekt může být zadán symbolicky nebo slovně, případně oběma způsoby současně. Často tyto údaje naopak nejsou přítomny vůbec. Pro systémy, zabývající se MIR, jsou však nerosvnalosti ve vydavatelských aspektech různých verzí jednoho díla velmi problematické [13].

1.6 Textový aspekt

Textový aspekt hudebního díla je tvořen texty písní, áriemi, chorály i librety. Tyto informace jsou ovšem téměř nezávislé na aranžmá nahrávky a získat pomocí systémů MIR melodii skladby lze na jejich základě jen velice obtížně. Tato situace je o to složitější při překladech textů do jiných jazyků. Důležité je také nezapomenout na množství děl, které neobsahují žádný text [13].

1.7 Bibliografický aspekt

Údaje o názvu díla, hudebním skladateli, textaři, editorovi, vydavateli, čísle katalogu, datu vydání – to vše a mnohé další je součástí bibliografického aspektu. Jsou to informace, které nelze čerpat z obsahu díla, nýbrž informace popisující hudební dílo. Označují se také jako hudební metadata [13].

Vstupy a výstupy MIR mohou být textové informace (metadata), notové zápisy, zvukové nahrávky nebo jejich úseky, či hudební parametry. Výsledek je buď přesný (např. získání hudby o určitém obsahu) nebo přibližný (např. získání hudby v závislosti na podobnosti) [8]. Systém lze také definovat na základě specifičnosti.

Systémy, které identifikují přesný obsah nahrávek, se nazývají systémy s vysokou specifičností. Ty, které ke svému popisu využívají vysokoúrovňové parametry, jsou označovány jako systémy se střední specifičností. Naopak systémy, využívající nízkoúrovňové parametry a všeobecný popis hudby, mají nízkou specifičnost. Do této kategorie patří i systémy, rozpoznávající hudební žánr [8].

Několik příkladů oblastí, v nichž se MIR využívá, a jejich specifičnosti jsou uvedeny v tabulce 1.1.

Tab. 1.1: Příklady využití MIR

Detekce	Specifičnost	Popis
Hudební dílo	Vysoká	Získ metadat neznámé skladby, mobilní MIR (např. <i>shazam.com</i>)
Plagiátorství	Vysoká	Zjištění zneužití duševního vlastnictví, týkající se hudebního díla
Verze	Vysoká, střední	Detekce remixů, coververzí, studiových vs. živých nahrávek
Melodie	Vysoká, střední	Nalezení díla, obsahujícího určitý melodický úsek
Název	Střední	Nalezení děl o stejném názvu
Umělec, skladatel	Střední	Nalezení děl daného umělce nebo skladatele
Podobnost	Střední	Nalezení hudby o podobném zvuku
Doporučení	Střední, nízká	Nalezení hudby, odpovídající preferencím uživatele
Nálada	Nízká	Nalezení hudby, odpovídající daným emocionálním konceptům
Instrumentace	Nízká	Nalezení hudby se stejnými hudebními nástroji
Žánr	Nízká	Nalezení hudby, odpovídající danému žánru

2 Rozpoznávání hudebního žánru

Rozpoznávání hudebního žánru – Music Genre Recognition (MGR) je jednou z podoblastí Získávání informací z hudby – Music Information Retrieval (MIR). Jedná se o klasický příklad strojového učení, kde jsou ze zvukového signálu extrahovány vhodné parametry, pomocí nichž je následně systém trénován za účelem klasifikace [26].

2.1 Žánr

Termín *žánr* pochází z latinského slova *genus*, tedy druh nebo třída. Lze jej popsat jako kategorii, definovanou strukturálními, tematickými nebo funkčními znaky. Hudební žánr proto můžeme považovat za specifický typ hudby se souborem společných vlastností, které při běžném poslechu odlišují danou skladbu od skladeb odpovídajících jiné kategorii [23]. Hlavními faktory definujícími hudební žánr jsou instrumentace, tempo, rytmické struktury a průběh dynamiky.

Určení hudebního žánru je sémantický problém, tedy otázka klasifikace. Lidé hudbu rozdělují do jednotlivých kategorií na základě individuálního vnímání zvukového signálu. Z toho důvodu je určení hudebního žánru subjektivní a nejednoznačné. Navíc je obvykle možné skladbu přiřadit k více než jednomu žánru, případně do mnoha různých žánrových podkategorií [52]. Ty mohou souviset s obdobím vzniku nahrávky (např. *šedesátkový rock*), místem původu (např. *Chicago blues*), obsahem (např. *zamilované písně*) a dalšími faktory [22].

Subjektivitou lidské klasifikace žánrů se zabývaly mnohé výzkumy. Ve studii D. Perrota and R. Gjerdingena [39] figurovali vysokoškolští studenti, jejichž úkolem bylo rozřadit vyslechnuté nahrávky o délce 300 ms do deseti žánrů. Výsledkem byla 70 % přesnost klasifikace u lidí s malým až středním hudebním vzděláním. Tím byla potvrzena schopnost identifikovat i velice krátký časový úsek. To také znamená, že nejdůležitějším prvkem pro rozpoznání žánru je barva zvuku. Tempo a dynamika, které se v takto malém časovém rozpětí nemohou projevit, jsou tedy až druhořadými ukazateli. V další studii [32] byly experimenty provedeny na 27 posluchačích, kteří rozdělovali nahrávky o délce 30 s do šesti žánrů. Zde bylo dosaženo vzájemné shody pouze 76 %. Tyto výsledky potvrzují výraznou subjektivitu v charakterizaci žánru posluchači.

2.2 Výzkumy zabývající se rozpoznáváním hudebního žánru

První publikace, týkající se automatického určení hudebního žánru, se objevily v devadesátých letech minulého století. Zpočátku byly výzkumy zaměřeny zejména na rozpoznávání mezi řečovým a hudebním signálem, ačkoliv se objevily i algoritmy, určující více kategorií [26]. Jedním z nich je práce J. Foota [15], který za pomoci strojového učení s učitelem (viz 1) rozlišuje signál hudební, řečový (mužský a ženský) a nehudební (zejména perkusivní zvuky). Využívá k tomu výpočetně nenáročnou stromovou kvantifikaci. Jako parametry byly zvoleny ZCR, MFCC a hodnoty, popisující dynamiku zvuku.

Jedním z prvních, kdo se zabýval žánrovou klasifikací, byl kolektiv H. Soltaua [47] v roce 1998. Skupina pracovala s databází o 360 skladbách, rozdělených do čtyř žánrů: rock, pop, techno a klasická hudba. Pro každou skladbu byly zjištěny témbrální parametry, popisující barvu zvuku. Analyzovány byly také časové struktury, tedy četnost výskytu jednotlivých událostí v sekvenci. Pro klasifikaci byly zvoleny dvě metody strojového učení, a to skrytý Markovův model (HMM), se kterým bylo dosaženo úspěšnosti 79%, a explicitní modelování času s neuronovou sítí (ETM-NN), s úspěšností 86%.

V práci G. Tzanetakise z roku 2001 [51] je problém automatického určení žánru řešen za pomoci Gaussova rozdělení. Databáze pro tento výzkum byla získána z rádia, CD a internetu a skládá se ze 750 nahrávek o délce 30 s, rovnoměrně rozdělených do patnácti žánrů. Parametry pro popis těchto skladeb byly spektrální těžiště, spektrální tok, rolloff, low energy a ZCR. Dále byl s využitím diskretní vlnkové transformace (DWT) vypočítán soubor rytmických parametrů, jako např. perioda nebo amplituda několika prvních špiček signálu.

Specifickou žánrovou oblastí se zabýval S. Dixon [12]. Ten v roce 2003 analyzoval taneční hudbu na základě jejích rytmických vzorců. Dataset pro tuto práci byl sestaven ze standardní a latinskoamerické taneční hudby, pro kterou je rytmus stěžejním elementem. K analýze bylo využito zejména určení počátků not (onset time), pomocí nichž lze odhadnout tempo a metrum skladby. Určení periodicity proběhlo s využitím autokorelace. Získané hodnoty byly poté porovnávány s údaji na přebalu CD, obsahujícího zpracovávané nahrávky. Úspěšnost správného zařazení v této práci dosáhla 80 %. Nevýhodou určení periodicity je skutečnost, že tento parametr poskytuje informace o metrických strukturách, nikoliv však o rytmice, která by byla pro identifikaci žánru vhodnější [12].

Autory dalšího z mnoha výzkumů byli v roce 2003 S. Z. Li a G. Guo [28]. Ti pro klasifikaci využili metodu, zvanou Support Vector Machines (viz 4.8). Jako dataset sloužil soubor 409 zvuků, manuálně rozdělených do šestnácti kategorií, který byl

využit již v dřívější práci E. Wolda [53]. U každé z nahrávek byla určena celková energie, energie v pásmech, brightness (neboli jas zvuku) a dvanáct koeficientů MFCC. S touto metodou bylo dosaženo úspěšnosti 78 % při využití MFCC a 85 % při klasifikaci pomocí zbývajících parametrů.

Identifikací žánru se zabývala také organizace ISMIR, která již byla zmíněna v kapitole 1. Významná byla pro tuto oblast její pátá mezinárodní konference v roce 2004. Zde došlo k uvedení datasetu ISMIR2004 (viz 5.2), který byl vytvořen pro jednu z pěti soutěží, jejíž smysl spočíval právě v určení hudebního žánru [20]. Tato databáze byla později využívána i v mnoha dalších výzkumech.

Ve stejném roce proběhl výzkum J. J. Burreda a A. Lerche [7]. Zde byl využit dataset rozdělený do 17 tříd o 50 nahrávkách, z nichž tři třídy obsahovaly řečový projev, jedna třída šumové signály a zbývajících třináct tvořilo jednotlivé hudební žánry. Délka trvání každé nahrávky činila 30 s. U každé skladby byly určeny témralní parametry (ZCR, MFCC, rolloff, spektrální tok), rytmické parametry (tempo, pravidelnost rytmu) a mnohé další. Klasifikace probíhala za pomoci Gaussova smíšeného modelu (viz 4.2) a metody k -NN (viz 4.3). Při diferenciaci základních tříd (řeč, šum a hudba) bylo dosaženo úspěšnosti 95 %, při určování jednotlivých hudebních subžánrů však klesla úspěšnost na 59 %.

V roce 2009 se rozpoznáváním hudebního žánru zabýval Ch. Lee et al. [27]. Analyzovány byly dva nejznámější datasety, a to GTZAN (viz 5.1) a ISMIR2004 (viz 5.2). Méně typické však byly parametry použité pro popis signálu. Mimo koeficientů MFCC byl určen oktávový spektrální kontrast (OSC) a normalizovaná zvuková spektrální obálka (NASE). OSC reprezentuje spektrální charakteristiku signálu tak, že popisuje spektrální špičku a propad nezávisle v každém frekvenčním pásmu. Parametr NASE poskytuje informaci o rozdělení výkonového spektra v každém rámci. Žánrová klasifikace poté proběhla za pomoci lineární diskriminační analýzy (viz 4.4). Na datasetu GTZAN bylo dosaženo úspěšnosti téměř 91 %, s datasetem ISMIR2004 necelých 87 %.

Určením žánru při využití kompresivního odběru vzorků se v roce 2010 zabýval tým K. K. Changa [19]. Analýzou databáze GTZAN byly získány parametry MFCC, OSC, low energy a spektrální parametry (ZCR, spektrální tok, rolloff). Metoda kompresivního samplování (CS) byla zvolena z důvodu své efektivity v oblastech, jako je zobrazování, rozpoznávání obličejů nebo odhad pohybu. Tím bylo dosaženo úspěšnosti 82 %.

Přístup odlišný ode všech dříve zmíněných zvolili G. Marques et al. [33]. V jejich práci nebyly nahrávky reprezentovány statistickým zpracováním vektorů parametrů, nýbrž za pomoci tzv. *kódového seznamu*. Tento přístup umožňuje explicitní charakterizaci jednotlivých částí prostoru parametrů a tím je možné sledovat umístění parametrů daných žánrů v tomto prostoru a jejich vliv na klasifikaci. Žánrové zařazení

proběhlo s využitím následujících tří metod: 5-NN (viz 4.3), Markovových modelů (viz 4.7) a Support Vector Machines (viz 4.8). Marques zvolil jako analyzované parametry ZCR, spektrální tok a těžiště, rolloff a 13 koeficientů MFCC. Systém byl testován na datasetu ISMIR2004 (viz 5.2) a Latin Music Datasetu, který obsahuje 900 skladeb latinskoamerické taneční hudby. Úspěšnost tohoto přístupu byla 75 % pro 5-NN a SVM, 83 % pro HMM.

Existují také výzkumy, pojednávající o specifické žánrové oblasti a rozřazení do jejích podkategorií. Mezi ně patří také vědecká práce V. Tsatsishviliho z roku 2011 [50] nebo D. G. J. Muldera z roku 2014 [36]. Obě zmíněné se zabývají rozborem metalové hudby. Pro každý byl sestaven vlastní dataset, obsahující 17 subžánrů. Tsatsishvili u nahrávek určoval běžně používané parametry, jakými jsou ZCR, MFCC a hodnoty, popisující spektrum signálu. Pro klasifikaci byly využity metody k -NN (viz 4.3) a AdaBoost (viz 4.1), s každou z nich bylo dosaženo úspěšnosti pouze 45 %. Oproti tomu Mulder pro popis signálu využíval horizontální intervaly, tedy intervaly mezi po sobě jdoucími výškami tónu, a vertikální intervaly, neboli intervaly mezi současně znějícími výškami tónu. Tyto parametry byly zpracovány metodou k -NN a pomocí Mahalanobisovy vzdálenosti. Tímto postupem bylo dosaženo úspěšnosti 70 %.

Dalším možným přístupem ke klasifikaci je využití umělých neuronových sítí (viz 4.9). Ty ve své práci využívali B. Vogler a A. Othman v roce 2016 [52]. Také pro tuto metodu je důležitá volba parametrů. Jako ve většině prací, byly i zde vypočítány koeficienty MFCC. Druhým určovaným parametrem se stal oktávový spektrální kontrast (OSC). V důsledku častých kritik databáze GTZAN (viz 5.1) došlo k vytvoření vlastního datasetu, obsahujícího sto hudebních děl od každého ze tří žánrů. Všechny skladby byly rozděleny na úseky o délce 30 s, a ty pak byly analyzovány jednotlivě. Nejčastěji predikovaný žánr byl určen jako finální žánr nahrávky. S touto metodou bylo dosaženo úspěšnosti 83 %.

Jedním z nejnovějších výzkumů zabývajících se rozpoznáváním hudebního žánru je práce H. Pálmasona et al. [38]. Pro klasifikaci zde byla zvolena metoda 3-NN (viz 4.3). Jako porovnávané parametry sloužily koeficienty MFCC a míra spektrální rovinnosti (SFM). SFM slouží k popisu „tónovosti“ a „šumovosti“ zvukového signálu. Tato vlastnost je určena na základě množství špiček ve výkonovém spektru signálu. Pokud je spektrum rovné, signál je označen jako šumový. Při větším množství špiček existuje větší pravděpodobnost výskytu sinusových vln a tím je signál označen za tónový [14]. Analýza byla provedena na datasetu GTZAN. Dále je zde také rozvedena diskuse, týkající se chyb v této databázi a jejich vliv na klasifikaci. Více je uvedeno v kapitole 5.1. I přes úpravy štítků žánrů v tomto datasetu však klasifikace dosáhla úspěšnosti necelých 81 %.

Přehled výzkumů, týkajících se rozpoznávání hudebního žánru za pomoci strojového učení, je uveden v přílohách, v tabulce B.1.

3 Parametrizace hudebního signálu

Jedním z hlavních úkolů MIR je parametrizace, tedy získání parametrů analýzou zvukového signálu. Pojem „parametr“ může znamenat numerický popis signálu (ve formě skalární hodnoty, vektoru nebo matice), ale také jeho nominální popis (často ve formě tzv. štítků). Účelem získání těchto parametrů je charakteristika zvukového signálu z hlediska, jakým je vnímán posluchači, tedy zejména určení hudebních aspektů jako je instrumentace, rytmika, melodie a harmonie [22].

Zvukové parametry lze rozdělit na základě úrovně abstrakce do následujících tří skupin:

1. **Nízkoúrovňové parametry** jsou obvykle vypočítány přímo ze zvukové vlny a bývají prostým statistickým popisem signálu (např. energie signálu, spektrální tok, spektrální těžiště, počet průchodů nulovou hodnotou).
2. **Středoúrovňové parametry** charakterizují základní zvukové aspekty, jež mohou být spojené mimo jiné s výškou tónu nebo rytmem (např. MFCC, onsets, fluktuace). Tyto parametry jsou získány kombinací nízkoúrovňových parametrů, případně aplikací psychoakustického modelu.
3. **Vysokoúrovňové parametry** popisují zvuk z hlediska lidského vnímání. Jsou také nejsrozumitelnějším popisem hudby pro běžného posluchače. Spadá sem určení instrumentace, tóniny, žánru nebo hudební nálady.

Dalším způsobem třídění parametrů je dělení z hlediska hudebního aspektu, který charakterizují [26]:

- **Temporální (časové) vlastnosti** - tempo, taktové předznamenání, rytmické struktury
- **Dynamické vlastnosti** - rozsah hlasitostí, změna hlasitosti v čase, akcenty
- **Tonální vlastnosti** - melodické vlastnosti, harmonie, netypické výšky tónu
- **Témbrální vlastnosti** - počet a typ hudebních nástrojů, jejich vzájemné vztahy, rozdělení do stera

3.1 Parametry využívané pro určení žánru

Rané systémy, zabývající se rozpoznáváním hudebního žánru, používaly pouze omezený počet zvukových parametrů. Byly jimi především počet průchodů nulovou hodnotou (ZCR) a parametry popisující intenzitu zvuku. Brzy nato však byly připojeny také parametry založené na témbru. Určení výšky tónu se taktéž objevilo v několika z těchto prací, ale dnes se již obvykle nevyužívá [26].

Během let množství analyzovaných parametrů narůstalo. V současnosti jsou těmi nejpoužívanějšími hodnoty založené na dynamice a témbru, byly však přidány i pa-

rametry, týkající se rytmu a výšky tónu [26]. Jejich charakteristiky jsou uvedeny v následujících pododdílech.

3.1.1 Parametry založené na dynamice

- Pro popis intenzity zvuku lze použít parametry, související s energií signálu. Jsou jimi *Střední hodnota energie signálu* \bar{E} a *Efektivní hodnota energie – Root mean square (RMS)* [22]:

$$\bar{E} = \frac{1}{\tau} \int_0^{\tau} E(t) dt, \quad (3.1)$$

$$\text{RMS} = \sqrt{\frac{1}{\tau} \int_0^{\tau} E^2(t) dt}, \quad (3.2)$$

kde $E(t)$ označuje průběh energie v čase a τ označuje periodu signálu.

- *Low energy* je parametr, udávající procento rámců, v nichž je hodnota energie nižší než průměrná hodnota energie [51]. Tím je definována stejnoměrnost časového rozložení energie v rámci signálu. Při kategorizaci hudebního žánru lze tento parametr využít pro zjištění dynamiky nahrávky, která je obvykle rovnoměrná v rockové hudbě, avšak velice nestabilní v hudbě vážné.
- *Spektrální tok – Spectral Flux (SF)* je parametr ve frekvenční doméně, popisuje změnu ve výkonovém spektru mezi po sobě jdoucími rámci [22]

$$\text{SF}_t = \sum_{n=1}^N (D_t(n) - D_{t-1}(n))^2, \quad (3.3)$$

kde D_t označuje normalizované frekvenční rozdělení v rámci t [22].

- *Teager-Kaiserův energetický operátor (TKEO)* charakterizuje současně energii i frekvenci signálu [11]. Toho lze využít pro rozlišení dynamiky hudebníkovy hry, neboť při hlasitější hře na hudební nástroj vzniká více vyšších harmonických složek. Zároveň se změna toho parametru může projevit při využití kytarových efektů (zkreslení), produkujících vyšší harmonické složky. Ty jsou typické pro rockovou a metalovou hudbu.
- *Medián výkonové spektrální hustoty – Median of Power Spectral Density (MPSD)* definuje rozložení energie ve výkonovém spektru [17].

3.1.2 Parametry založené na rytmu

- Nejpoužívanějším parametrem pro určení rytmu skladby je *onset time*, neboli doba počátku noty. Ta je určena z průběhu signálu v časové doméně, resp. ze špiček jeho křivky [34]. Na základě nalezených pravidelností ve výskytu těchto špiček je odvozen rytmus a tempo nahrávky.

- Dalším parametrem pro odhad rytmických struktur je *fluktuace*. Ta je založena na výpočtu spektrogramu, transformovaného sluchovým modelem, a následném odhadu spektra v každém z pásem [34].
- Tempové vlastnosti nahrávky lze charakterizovat také pomocí parametru *event density*, tedy hustoty událostí. Tato proměnná odhaduje průměrnou frekvenci výskytu událostí, jinými slovy množství počátků not za sekundu [34].
- Pro zkoumání metrických struktur byl zaveden také parametr *metroid*. Je založen na stanovení těžiště vybraných metrických úrovní (*metrical centroid*) a určení tzv. metrické síly. Výsledná křivka metrického těžiště udává časový vývoj metrické aktivity, vyjádřené v úderech za minutu (BPM). Vysoké hodnoty BPM v tomto případě odpovídají základním metrickým úrovním, tj. rychlým rytům. Naopak nízké hodnoty BPM znamenají, že ve skladbě převažují vyšší metrické úrovně, tedy celé noty ve formě pomalých pulzací [34]. Odlišností od běžného určení metra je zde zahrnutí výpočtu metrické síly. Ta udává, nakolik jsou pulzace v signálu výrazné a jednoznačné nebo naopak nevýrazné až skryté. Může také nastat situace, kdy signál obsahuje směs pulzací s různým tempem. Metrická síla je určena sumou autokorelací jednotlivých metrických úrovní [34].

3.1.3 Parametry založené na témbu

- *Počet průchodů nulovou hodnotou – Zero Crossing Rate (ZCR)* měří, kolikrát hodnota amplitudy změní znaménko během rámece t . Při hodnocení zvukových nahrávek se používá např. k detekci perkusivních zvuků [22]. Tato vlastnost souvisí také s určením žánru, respektive s výskytem bicích nástrojů v dané nahrávce. Jako příklad lze uvést rockovou hudbu, pro níž jsou perkuse téměř podmínkou, naopak v klasické hudbě se vyskytují spíše výjimečně.

$$ZCR_t = \frac{1}{2} \sum_{k=t*K}^{(t+1)*K-1} |sgn(s(k)) - sgn(s(k+1))|, \quad (3.4)$$

kde $s(k)$ znamená k -tý vzorek signálu, K označuje počet vzorků v mikrosegmentu a t udává rámeček.

- *Melovské keprální koeficienty – Mel-Frequency Cepstral Coefficients (MFCC)* jsou koeficienty, popisující spektrální tvar zvuku [34]. Po určení hodnot frekvencí v každém rámci je měřítko v hertzech převedeno na měřítko ve stupnici Mel, kde jsou frekvenční pásma rozložena logaritmicky, čímž je simulováno vnímání lidského sluchového ústrojí. Dále je určena logaritmická hodnota v těchto pásmech a provedena diskrétní cosinova transformace.

Vektor MFCC popisuje periodicitu nalezenou v hodnotách velikosti frekvenčních pásem v daném rámci. Výjimkou je první koeficient, který odpovídá průměrné energii signálu v daném rámci [22, 34].

Tyto koeficienty lze využít při určování žánru, neboť nízké hodnoty MFCC (a tím i nízká periodicitu) jsou příznačné pro klasickou hudbu, naopak vysoké hodnoty MFCC odpovídají rocku nebo metalu [22].

- Parametr *Brightness*, neboli jas zvuku, udává množství energie, vyskytující se nad mezní frekvencí. Typicky je hodnota mezní frekvence volena v rozmezí 1000 Hz [25] až 3000 Hz [21]. Výsledná hodnota je udávána v procentech [21, 34]. Tímto parametrem je možné odhadnout množství vyšších harmonických složek signálu.
- Další parametr, který hodnotí množství energie vyšších harmonických složek signálu, je *Rolloff*. Ten udává frekvenci, pod kterou se nachází určitá část celkové energie signálu. Tzanetakis a Cook [51] uvádějí tuto část jako 85 % celkové energie, oproti tomu podle autorů Pohleho, Pampalka a Widmera [40] je to 95 %.
- *Spektrální těžiště – Spectral Centroid (SC)* charakterizuje frekvenční oblast, ve které se soustřeďuje největší množství energie. Tento parametr se často označuje jako „jas“ zvuku [22]. Zároveň je to první moment a geometrický střed frekvenčního rozdělení [22]

$$SC_t = \frac{\sum_{n=1}^N m_t(n) \cdot n}{\sum_{n=1}^N m_t(n)}. \quad (3.5)$$

kde $m_t(n)$ označuje váženou frekvenční hodnotu histogramu o střední frekvenci n .

Podle [34] jej lze také popsat vztahem

$$\mu_1 = \int x f(x) dx. \quad (3.6)$$

kde $f(x)$ popisuje hustotu rozdělení frekvencí.

Při analýze různých žánrů bylo zjištěno, že klasická hudba dosahuje nejnižších a velice stabilních hodnot spektrálního těžiště [22].

- *Spectral spread – frekvenční rozptyl* popisuje druhý centrální moment frekvenčního rozdělení a tím i míru disperze tohoto rozdělení [34]

$$\sigma^2 = \mu_2 = \int (x - \mu_1)^2 f(x) dx. \quad (3.7)$$

Průměrná hodnota frekvenčního rozptylu v rámci hudebního díla může sloužit k popisu vnímané barvy zvuku [26]. Podle Kneese a Schedla [22] se tento parametr u různých žánrů liší - klasická hudba má ve většině případů menší rozptyl než metalová nebo elektronická hudba.

- *Spectral skewness* – *spektrální šikmost* udává míru symetrie frekvenčního rozdělení a je jeho třetím centrálním momentem. Šikmost zcela symetrického rozdělení je rovna nule [34]

$$\mu_3 = \int (x - \mu_1)^3 f(x) dx. \quad (3.8)$$

- *Spectral kurtosis* – *spektrální špičatost* označuje čtvrtý centrální moment frekvenčního rozdělení [34].
- *Spectral flatness* – *spektrální plochost* je poměrem mezi geometrickou a aritmetickou střední hodnotou signálu a hodnotí množství špiček signálu [34].

4 Metody strojového učení

Strojové učení je podoborem umělé inteligence, spočívající ve schopnosti stroje automaticky zpracovávat a analyzovat data. Pojem „učení“ vyjadřuje postupné zlepšování provedení určitého úkolu, aniž by byla tato operace explicitně naprogramována [24].

Strojové učení se v dnešní době využívá v mnoha oblastech. Jednou z nich je např. hodnocení internetových stránek z hlediska relevance při vyhledávání webovými prohlížeči. Podobným problémem je i tzv. kolaborativní filtrování, jež spočívá v doporučení podobného obsahu uživateli na základě jeho předchozí činnosti (např. doporučení zboží na základě dřívějších nákupů). Další významnou oblastí, v níž se strojové učení využívá, je automatický překlad dokumentů. Zde je stěžejní gramatická správnost, kterou si stroj osvojí na základě překladů podobných dokumentů. Širokou oblastí strojového učení je též rozpoznávání. Do této kategorie spadá především rozpoznávání řeči, rukopisu, tváří apod. [3, 46].

Jedním z hlavních využití strojového učení je klasifikace. Tímto pojmem rozumíme třídění vstupních dat do několika kategorií, případně rozhodnutí, že vstupní informace do žádné ze známých kategorií nenáleží [46]. Do této oblasti spadá také rozpoznávání hudebního žánru.

V praxi rozlišujeme dva základní typy strojového učení:

1. **Učení s učitelem** je metoda, při níž jsou nejprve zpracovány parametry trénovacích dat, jejichž kategorie jsou známé. Následně jsou analyzovány parametry testovacích dat a na základě podobnosti s dříve získanými hodnotami jsou testovací data rozdělena do kategorií [35].
2. **Učení bez učitele** analyzuje data, která nebyla rozdělena do kategorií. Pracuje s podobnými znaky zpracovávaných dat a jejich rozdělením do shluků (clusterů), které se zakládá na této podobnosti [35].

4.1 AdaBoost

Adaptive Boosting, zkráceně AdaBoost, je metoda strojového učení, využívající iterativní algoritmus. V každém opakování tohoto algoritmu je volána sada předdefinovaných jednoduchých, tzv. „slabých“, klasifikátorů (např. binární prahové hodnoty jednoho z parametrů [22]). Přizpůsobováním váhy slabých klasifikátorů na základě jejich výkonu na trénovacích datech je vytvořen tzv. silný klasifikátor, který následně třídí testovací data [16, 22].

Všechny slabé klasifikátory jsou váhovány po každém opakování a to tak, že váhy nesprávně určených případů zůstávají nezměněny, zatímco při správném určení do-

cháží ke změně váhy pomocí klasifikační chyby podle vzorce

$$W_k = W_{k-1} \frac{e}{1 - e}, \quad (4.1)$$

kde W_k označuje váhu, přiřazenou klasifikátoru v k -tém opakování a e označuje klasifikační chybu [50].

Pokud klasifikační chyba dosáhne hodnoty $e = 0$ (tedy dokonalá klasifikace), nabudou všechny váhy nulových hodnot a proces skončí. V případě, že k tomu nedojde, je algoritmus ukončen po počtu opakování, zadaném uživatelem. Výstup procesu sestává z klasifikátorů a jejich vah. Výsledná třída je stanovena na základě součtu vah klasifikátorů určujících jednotlivé třídy, resp. podle nejvyšší hodnoty součtu [50].

4.2 Gaussův smíšený model

Smíšený model je pravděpodobnostní model, reprezentující přítomnost tříd v celkovém prostoru dat. Klasifikátor Gaussova smíšeného modelu – Gaussian Mixture Model (GMM) modeluje každou třídu jako lineární kombinaci Gaussových hustot. Každá třída k je tedy reprezentována vícerozměrnou podmíněnou hustotou

$$p(x|\omega_k) = \sum_{m=1}^M w_{km} p_{km}(x), \quad (4.2)$$

kde x označuje vektor funkcí, ω_k událost náležící třídě k , w_{km} váhu směsi, M celkový počet hustot ve směsi a p_{km} normální hustotu [7].

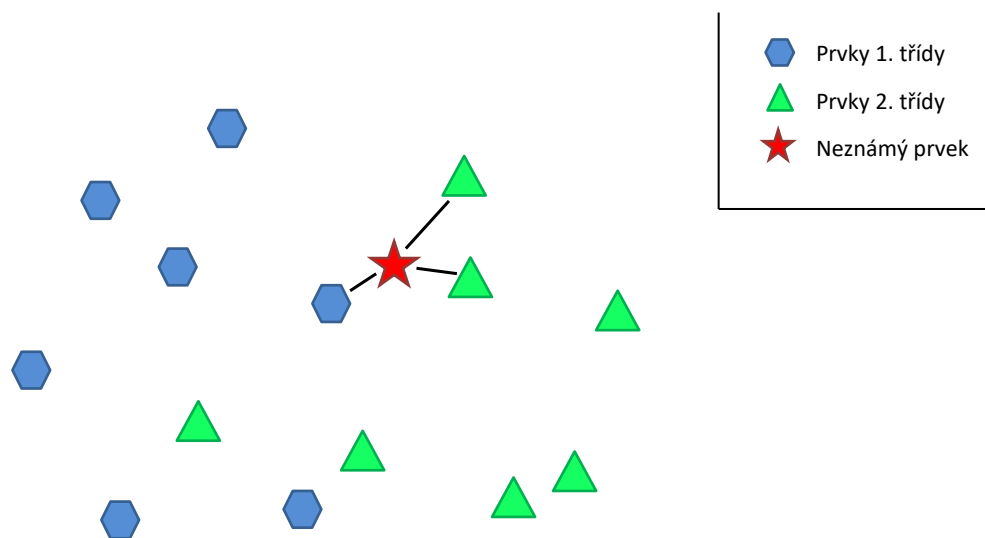
Podmíněná hustota se také nazývá pravděpodobnost třídy k vzhledem k vektoru parametrů x . V případě, že celkový počet hustot ve směsi $M = 1$ je každá třída modelována normálním rozdělením a klasifikátorem je jednoduše Gaussův klasifikátor [7, 22].

Při využití Gaussova smíšeného modelu v oblasti rozpoznávání hudebního žánru jsou nejprve zpracována trénovací data. Parametry každé kategorie, tedy váhy směsi, střední vektory a kovarianční matice každé složky, jsou odhadnuty využitím algoritmu, zvaného maximalizace očekávání. Pro určení vstupního vektoru je za pomoci odhadnutých parametrů vypočítána podmíněná hustota v každé ze tříd. Kategorii, jejíž hodnota hustoty je největší, je daný vektor přiřazen. Tento postup se nazývá kritérium maximální pravděpodobnosti [7].

Výhodou tohoto postupu oproti jiným metodám, jako jsou např. SVM nebo k -NN, je menší množství dat potřebných pro klasifikaci, neboť nejsou ukládány všechny parametry trénovacích dat, nýbrž pouze set hodnot, typických pro každou třídu.

4.3 k-NN

Metoda *k* nejblížeších sousedů – *k* nearest neighbor (*k*-NN) bývá označována jako tzv. „líné učení“, neboť trénovací fáze spočívá pouze v uložení vektoru parametrů trénovacích dat se správným označením žánru [22]. Tím je vytvořen „slovník“, v němž je každý žánr charakterizován několika vektory hodnot. Testovací fáze probíhá porovnáním vektoru hodnot, získaných z testovacího signálu, s vektory hodnot ve slovníku. Toto porovnání spočívá v hledání *k* nejblížeších vektorů využitím určité metriky, nejčastěji Euklidovské vzdálenosti. Jako výsledný žánr je určen ten, který se mezi *k* nejblížešími vektory vyskytuje nejčastěji [1, 22, 30].



Obr. 4.1: Metoda 3-NN

Důležitá je volba parametru *k*. Pro nízké hodnoty *k* je tato metoda příliš citlivá na šum, což znamená, že dochází k nepřesnosti klasifikace. Naopak při vysokých hodnotách *k* dochází ke zpomalení algoritmu v důsledku zpracování velkého množství dat, čímž je současně zvýšena paměťová náročnost [30].

4.4 Lineární diskriminační analýza

Další metodou, hojně využívanou v pracích, zabývajících se rozpoznáváním žánru hudebního díla, je *Lineární diskriminační analýza* – *Linear Discriminant Analysis* (LDA). Její základní myšlenkou je nalézt lineární transformaci, která nejlépe rozlišuje mezi jednotlivými třídami. Záměrem nalezení této transformace je snížení

korelace mezi parametry, čímž je snížena také nadbytečnost [30]. Následně je provedena klasifikace prostoru, transformovaného na základě určité metriky, jakou je např. Euklidovská vzdálenost [29]. Tím je určeno, která ze dvou kategorií je pravděpodobnější.

Při rozdělování dat do více než dvou kategorií je obvykle postupováno tak, že jedna kategorie náleží do první skupiny a zbylé kategorie do druhé. Poté je provedena LDA, jejímž výsledkem je stejný počet klasifikátorů, jaký je počet kategorií. Nakonec jsou získané hodnoty klasifikátorů zkombinovány a tím je určena výsledná kategorie [29, 30].

4.5 Náhodné lesy

Strojové učení využívající tzv. náhodné lesy (random forests) je kombinovaná metoda, sloužící současně ke klasifikaci i regresi. Základem jsou klasifikační stromy, které lze obecně za účelem klasifikace zdokonalit technikou *bagging* nebo *boosting*. Technika boosting (posilování) spočívá ve změně vah prediktorů v závislosti na správnosti predikce předchozích stromů. Oproti tomu v technice bagging (akronym pojmu *bootstrap aggregating*) klasifikační stromy nezávisějí na předchozích stromech a jsou vytvořeny pomocí bootstrap vzorku trénovacího datasetu [31].

Náhodné lesy jsou vytvořeny způsobem, který dodává technice bagging další úroveň náhodnosti. Dochází totiž ke změně konstrukce jednotlivých stromů. Ve standardních stromech je každý uzel rozdělen nejlepším rozdělením mezi všemi parametry. V náhodných lesích je však rozdělení určeno z podmnožiny prediktorů, náhodně zvolených v daném uzlu [31].

Tato strategie se překvapivě ukázala být vysoce efektivní v porovnání s jinými metodami, jako jsou Umělé neuronové sítě 4.9 nebo Support Vector Machines 4.8 [5].

4.6 Naivní Bayesův klasifikátor

Ve strojovém učení se využívá také naivní Bayesův klasifikátor, což je jednoduchý pravděpodobnostní klasifikátor, založený na Bayesově větě. Naivní Bayesův klasifikátor provádí statistickou analýzu trénovacích dat, vytváří odhad maximální pravděpodobnosti a využívá podmíněné pravděpodobnosti hodnot zkoumaných parametrů jako rozhodovací kritéria. To znamená, že je pomocí trénovacích dat určena pravděpodobná hodnota všech parametrů pro každý žánr. Testovací data jsou nakonec přiřazena příslušným žánrům na základě maximální pravděpodobnosti hodnot svých parametrů [2, 6].

4.7 Skrytý Markovův model

Skrytý Markovův model – Hidden Markov Model (HMM) je dvojnásobný náhodný proces se stavy, jež nejsou pozorovatelné. Viditelný je však výstup systému, který je na skrytých stavech závislý. Každý skrytý stav má své vlastní rozdělení pravděpodobnosti, model má tudíž možnost měnit toto rozdělení v průběhu zpracovávání [9, 42]. Skrytý Markovův model lze specifikovat pomocí pěti parametrů. Jsou jimi počet stavů modelu, počet pozorovaných symbolů v každém stavu, distribuce pravděpodobnosti přechodu stavu, distribuce pravděpodobnosti výskytu pozorovaného symbolu a distribuce počátečního stavu [18, 42].

Skrytý Markovův model často nachází své uplatnění v oblastech rozpoznávání řeči i analýze hudebního signálu [54]. Pro klasifikaci hudebního žánru je nejprve třeba zvolit vhodné parametry modelu. Následně je s pomocí „algoritmu dopředu-zpět“ vypočtena pravděpodobnost pozorovaných vlastností. Dále je využit Baum-Welchův algoritmus pro zvýšení této pravděpodobnosti a nakonec Viterbiho algoritmus, sloužící pro výpočet nejpravděpodobnější posloupnosti stavů modelu [9].

4.8 Support Vector Machines

Další metoda strojového učení s učitelem se nazývá *Support Vector Machines (SVM)*. Tato klasifikační metoda spočívá v převodu parametrů získaných z trénovacích dat do dimenzionálního prostoru. Následně nachází nadrovinu (separátor), jež optimálně rozděluje prostor tak, aby se na každé straně nadroviny nacházely prvky jedné kategorie. Nadrovina je zvolena tím způsobem, aby se od ní nejbližší prvky nacházely co nejdále. Pomocí nejbližších bodů, tzv. podpůrných vektorů (support vectors), je pak nadrovina definována. Tyto body mají nejvýraznější vliv na klasifikaci [22, 26, 28].

V základní představě je separátor lineární, tvoří tedy přímku. Ve většině případů, zejména při rozdělení dat do více než dvou tříd, je však tento způsob klasifikace neefektivní. Proto je třeba zvolit vhodný typ jádra (např. polynomiální nebo Gaussovské), pomocí něhož je v trojrozměrném prostoru zvolen tvar dělicí nadroviny tak, aby byla klasifikace co nejpřesnější [26].

Tato metoda má oproti tradičním technikám klasifikace nízké riziko nesprávného rozřazení testovacích dat, neboť je zde pevně stanovené, ačkoliv neznámé, rozdělení pravděpodobnosti dat. Přesnost se obecně snižuje s rostoucím počtem analyzovaných tříd [28].

4.9 Umělé neuronové sítě

Artificial Neural Networks – Umělé neuronové sítě (ANN) tvoří strukturu, složenou z několika vzájemně propojených výpočetních jednotek, tzv. umělých neuronů. Každá jednotka přijímá několik vstupů, na něž používá aktivační funkce. V závislosti na těchto funkcích je poté určen výstup jednotky. Každý vstup má své odpovídající váhy, které si lze představit jako sílu vazeb mezi neurony. Neurony jsou rozděleny do vrstev: jedné vstupní vrstvy, několika skrytých vrstev a jedné výstupní vrstvy. Pokud má tato struktura více než dvě skryté vrstvy, je označována jako hluboká neuronová síť [30, 52].

Ve fázi trénování si ANN přizpůsobuje sílu vazeb mezi neurony v závislosti na učícím algoritmu. Tento algoritmus řídí přenos vstupní hodnoty ze vstupní vrstvy do výstupní vrstvy. Poté jsou srovnány výstupní a skutečné hodnoty, jejich rozdíl je označován jako chyba. Pro snížení chyby jsou pomocí optimalizačních technik přizpůsobeny váhy neuronů [30, 52].

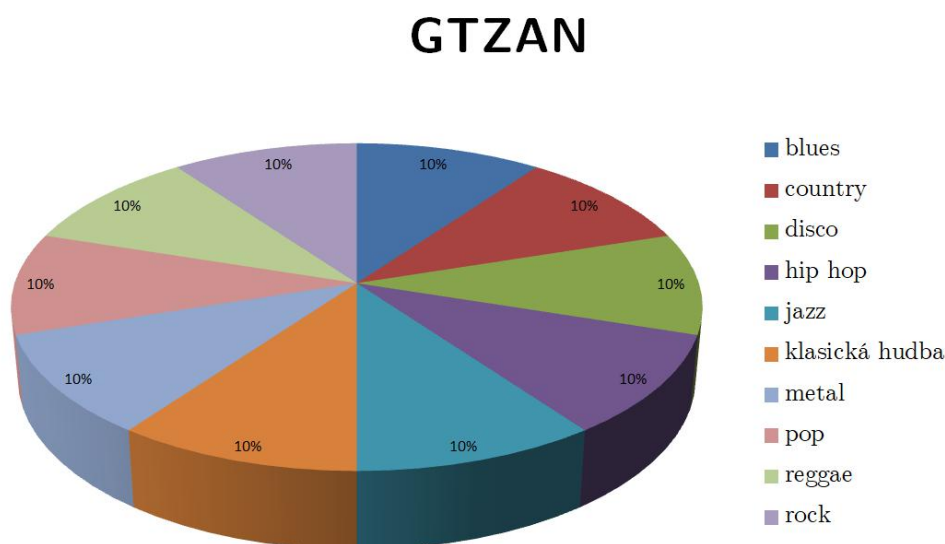
Výhodou ANN je vysoká odolnost vůči šumu, nevýhody však spočívají ve výpočetní náročnosti, týkající se zejména rychlosti a výkonnosti [52].

5 Datasets

Důležitým prvkem pro rozpoznávání hudebního žánru je volba hudebního datasetu, který bude zvolenou metodou zpracováván. Na jeho typu a kvalitě závisí celková přesnost analýzy. Existuje velké množství veřejných i komerčních souborů dat, určených pro rozpoznávání žánru. Většina těchto datasetů je autory předem rozdělena do jednotlivých žánrových kategorií.

5.1 GTZAN

Jedním z nejčastěji využívaných datasetů v oblasti rozpoznávání hudebního žánru je veřejný soubor dat GTZAN. Jeho autory byly v roce 2002 G. Tzanetakis a P. Cook [51]. Dataset obsahuje celkem 1000 skladeb o délce 30s, stejnoměrně rozdělených do deseti žánrů: blues, country, disco, hip hop, jazz, klasická hudba, metal, pop, reggae a rock.



Obr. 5.1: Žánrové složení databáze GTZAN

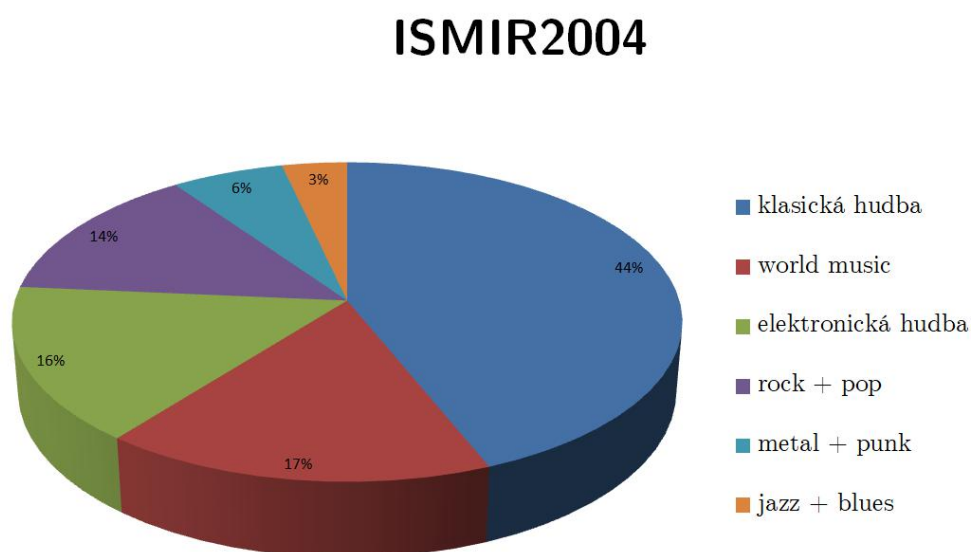
Ačkoliv existuje velké množství výzkumů, které v minulosti GTZAN využívaly, byly na tento dataset vzneseny kritiky. Jedním z hlavních problémů je jeho nedostatečná obsáhlost, tedy pouze sto skladeb na žánr. Z důvodu malého množství trénovacích dat je tedy snížena přesnost klasifikace. Další z chyb datasetu je způsobena zařazením jedné skladby (Queen - Tie your mother down) ke dvěma žánrům (rock a metal). Pokud je tedy v trénovacích datech skladba označena jako rock,

při testovací fázi bude nesprávně identifikována jako metal, a naopak. Další problematickou skladbou je jeden z titulů, označený jako reggae, který obsahuje pouze 6 sekund hudby a zbylých 24 sekund hlasitého šumu [38, 49].

Jedním z nedostatků datasetu GTZAN může být také nesprávné určení žánrů skladeb samotným autorem G. Tzanetakisem. Tímto problémem se ve své práci zabýval H. Pálmason et al. [38]. Pro tuto studii zvolil dvacet účastníků, aktivních v hudebním průmyslu. Ti na základě poslechu označili každou v minulosti nesprávně klasifikovanou skladbu žánrem, který ji podle jejich názoru nejlépe definoval. Výsledkem studie byla častá shoda volby žánru jednotlivými účastníky, která však byla v rozporu s označením žánru autorem G. Tzanetakisem. Závěrem je značná subjektivita v definici žánru uživateli.

5.2 ISMIR2004

Dalším z veřejných datasetů, využívaných v oblasti rozpoznávání hudebního žánru, je ISMIR2004. Ten byl vytvořen pro soutěž v klasifikaci žánru, která se uskutečnila během páté mezinárodní konference zabývající se získáváním informací z hudby v roce 2004 [33].



Obr. 5.2: Žánrové složení databáze ISMIR2004

Dataset se skládá z 1458 skladeb v plné délce, z nichž je 729 určených pro trénovací fázi a 729 pro testování. Nahrávky jsou rozřazeny do následujících šesti žánrů:

klasická hudba (320 skladeb), elektronická hudba (115 skladeb), jazz a blues (26 skladeb), metal a punk (45 skladeb), rock a pop (101 skladeb) a world music (122 skladeb) [27].

Nevýhodou datasetu ISMIR2004 je rozdělení do pouhých šesti kategorií, přičemž v některých z nich dochází ke spojení dvou žánrů dohromady. To je z hlediska běžného posluchače často nevhodné, neboť většina z nich rozlišuje rock a pop jako zcela odlišné žánry. Problematická může být také analýza plné délky skladby, jelikož se zde mohou vyskytovat dlouhé přede hry či pasáže, netypické pro daný žánr.

5.3 Free Music Archive

Free Music Archive (FMA) je volně přístupný soubor dat, určený pro práci v oblasti MIR, a to zejména pro určení hudebního žánru. Kompletní dataset obsahuje 106 574 skladeb v plné délce, vytvořených 16 341 různými autory. Skladby jsou rozděleny do 161 žánrů, z nichž 16 jsou základní žánry a zbylých 145 subžánry. Dále poskytuje předem vypočítané parametry těchto nahrávek, metadata, štítky a další dodatečné informace [10].

Dataset FMA se vyskytuje i v následujících zredukovaných verzích:

- Large dataset – veškeré skladby jsou omezeny na délku 30 s, extrahovanou ze středu nahrávky
- Medium dataset – výběr 25 000 skladeb o délce 30 s, rozřazených do 16 různých žánrů
- Small dataset – výběr 8 000 skladeb o délce 30 s, rozdělených do 8 nejpopulárnějších žánrů. Tento dataset bývá přirovnáván k datasetu GTZAN 5.1.

5.4 Million Song Dataset

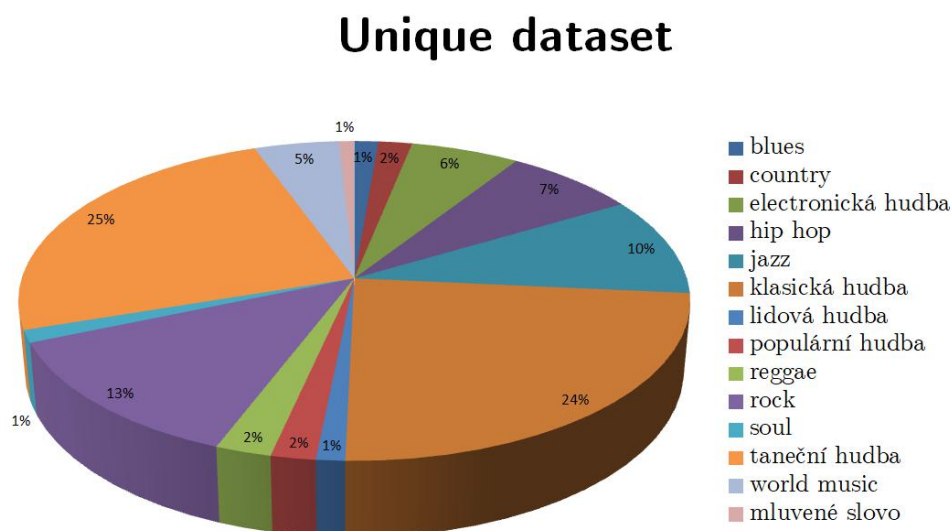
Mezi jedny z nejobsáhlejších volně dostupných datasetů patří Million Song Dataset. Jeho obsahem však nejsou přímo hudební nahrávky, ale již vypočítané zvukové parametry a metadata jednoho milionu skladeb, reprezentujících současnou komerční hudbu. Ke každé skladbě jsou také přiřazeny informace o jejím interpretovi, roku vzniku, žánru apod. [4].

Výhodou takto objemného datasetu je velké množství dat, pomáhající odhalit ojedinělé struktury, které se nevyskytují v menších souborech dat. Mohou však být stěžejní při analýze hudby, se kterou se uživatel setkává v každodenním životě. Další výhodou Million Song Datasetu je možnost odhalení podkategorií, jejich vzájemných spojitostí a parametrů, které je charakterizují [4].

5.5 Unique Dataset

Jedním z méně známých datasetů, který byl vytvořen za účelem rozpoznávání žánru, je Unique Dataset Klause Seyerlehnera [44]. Jeho obsahem je 3315 skladeb, rozdělených do následujících čtrnácti žánrů: blues, country, elektronická hudba, hip hop, jazz, klasická hudba, německá lidová hudba, německá populární hudba, reggae, rock, soul, taneční hudba, world music a mluvené slovo. Z každé nahrávky byl vybrán úsek o délce 30 s.

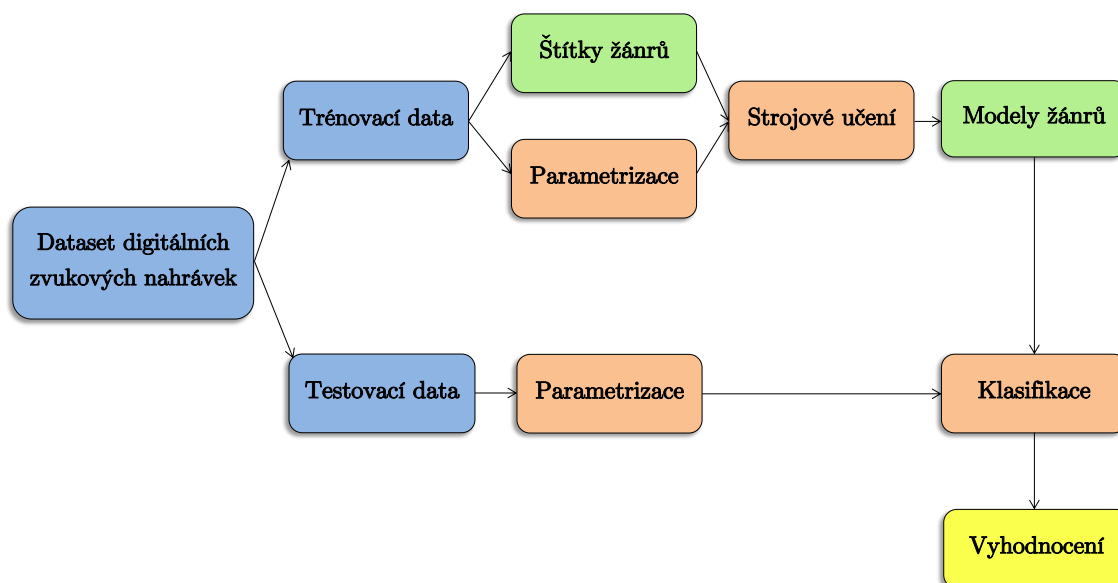
Dataset je sestaven takovým způsobem, aby každá skladba byla vytvořena jedinečným umělcem. Nejsou zde tedy žádné dvě nahrávky od totožného autora. Tím byl odstraněn vliv specifické produkce skladby na žánrovou klasifikaci. Tento vliv je běžně způsoben charakteristickým stylem umělce, který souvisí např. s instrumentací, barvou hlasu a typem hudby. V důsledku toho jsou si spektrální reprezentace písní jednoho autora vzájemně velice podobné. Nejinak je tomu při analýze nahrávek jednoho alba, kde může být podobnost způsobena společnou postprodukcí [44].



Obr. 5.3: Žánrové složení databáze Unique dataset

6 Návrh vyhodnocovacího systému

Vstupem systému určeného pro rozpoznávání hudebního žánru je databáze hudebních nahrávek. Jelikož většina přístupů využívá jednu z metod strojového učení s učitelem, je třeba dataset rozdělit na trénovací a testovací podmnožinu dat. Ke každé nahrávce trénovacích dat je posléze nutné přiřadit metadata, udávající žánr příslušné skladby.



Obr. 6.1: Blokové schéma vyhodnocovacího systému

Ve druhé fázi jsou pro každou nahrávku testovacích i trénovacích dat vypočítány její parametry. Jako nejefektivnější se v moderních výzkumech [38, 50, 52] ukázaly parametry, popisující tónbr zvuku, a to zejména koeficienty MFCC a počet průchodů nulovou hodnotou (ZCR). Skutečnost, že jsou tónbrální parametry pro určení žánru nejefektivnější, je však v rozporu literaturou, zabývající se teoretickou analýzou hudebního stylu a žánru [41, 48]. Zde jsou jako stěžejní elementy, definující styl skladby, určeny rytmické a melodické struktury a akordové postupy. Ty lze částečně popsat parametry, jako je onset time nebo fluktuace, případně průběh energie signálu, která úzce souvisí s dynamikou.

Alternativním přístupem k určení hudebního žánru je využití tzv. kulturních prvků, tj. metadat, získaných z internetových zdrojů [22]. Analyzována tedy není přímo zvuková nahrávka, ale např. informace, nacházející se na webových stránkách umělců, údaje poskytnuté vydavatelem nebo recenze hudebních kritiků.

V následujícím kroku probíhá klasifikace za pomoci zvolené metody strojového učení. Nejprve je analýzou parametrů trénovacích dat s údaji o příslušném žánru určen klasifikátor, definující předpokládané hodnoty jednotlivých parametrů u nahrávek jednoho žánru. V testovací fázi jsou poté pomocí tohoto klasifikátoru zhodnoceny parametry neznámých hudebních děl z testovací množiny dat. Jejich žánry jsou následně predikovány na základě podobnosti s parametry trénovacích dat.

Výsledná úspěšnost metody je dána porovnáním predikovaného a skutečného žánru skladby [22].

7 Evaluace systému

V rámci této práce byly zpracovány databáze ISMIR2004 (viz. 5.2), Unique dataset (viz. 5.5) a Free Music Archive (viz. 5.3).

7.1 Volba parametrů

Pro každou nahrávku trénovacích dat bylo vypočítáno 124 parametrů z oblasti barvy zvuku, dynamiky i tempa. Seznam základních parametrů je uveden v příloze v tabulce C.1. Pro parametry reprezentované vektorem, tedy pro všechny s výjimkou mediánu výkonové spektrální hustoty a Melovských keprálních koeficientů, byly na základě statistického zpracování vypočítány další parametry, které jsou uvedeny v následující tabulce 7.1.

Tab. 7.1: Seznam statistických parametrů

Označení parametru	Popis
mean	střední hodnota
median	medián
max	maximální hodnota
var	rozptyl
std	směrodatná odchylka
Q1	první kvartil
Q3	třetí kvartil
IQR	mezikvartilní rozsah
1p	první percentil
99p	devadesátý devátý percentil
IR	mezipercentilový rozsah

Výpočet veškerých parametrů hudebních nahrávek byl proveden v programu MATLAB, s využitím toolboxu MIRtoolbox 1.6.1., který obsahuje funkce pro výpočet parametrů, vhodných k identifikaci zvukového signálu (Music Information Retrieval) [34].

V další fázi byly za pomoci metody *minimum Redundancy Maximum Relevance* (mRMR) hledány parametry, které mají největší vliv na správnou klasifikaci jednotlivých žánrů. Současně také došlo k určení parametrů, majících na klasifikaci spíše negativní vliv a zkreslujících výsledek zařazení do správné třídy. Mezi parametry s nejvyšší relevancí se nejčastěji vyskytovaly některé ze statistických parametrů RMS, ZCR a také koeficienty MFCC. Naopak jako nadbytečný parametr byl určen

odhad tempa nahrávky. Důvodem je obtížnost správného určení tempa z nahrávky, v jejímž průběhu se tempo ve většině případů opakovaně mění. Větší úspěšnosti dosáhl parametr EventDensity, který popisuje hustotu událostí v nahrávce a s tempem úzce souvisí. S využitím metody mRMR byly také vyřazeny parametry dosahující často pouze nulových hodnot, jako např. rozptyl a směrodatná odchylka parametru Brightness.

7.2 Výběr metody klasifikace

Následujícím krokem je zvolení metody strojového učení, s jejíž pomocí budou zpracovány dříve vypočítané parametry. Využitím zvolené techniky dojde na základě trénovacích dat k vytvoření klasifikátorů, vhodných pro analýzu testovacích dat. Účinnost jednotlivých metod závisí na typu získaných informací, jejich variabilitě, počtu kategorií i dalších vlastnostech databází.

Klasifikace byla provedena taktéž v prostředí MATLAB, s využitím aplikace *Classification learner*. Pro stanovení nejefektivnější metody strojového učení byla analyzována trénovací část databáze ISMIR2004. Nejvyšší předpokládaná úspěšnost byla vyhodnocena u metody kvadratické Support Vector Machines (SVM), a to 78,1%. Z tohoto důvodu byly kvadratické SVM zvoleny jako stěžejní metoda klasifikace pro tuto práci.

Předpokládané úspěšnosti dalších metod jsou v sestupném pořadí uvedeny v následující tabulce 7.2.

Tab. 7.2: Předpokládaná úspěšnost metod strojového učení

Název metody	Úspěšnost
kvadratické SVM	78,6 %
lineární diskriminant	76,4 %
lineární SVM	76,3 %
střední Gaussovské SVM	75,7 %
10-NN	75,7 %
váňované k -NN	75,0 %
bagged trees	74,8 %
boosted trees	65,8 %

8 Testování a vyhodnocení

8.1 Databáze ISMIR2004

Specifika databáze ISMIR2004 jsou blíže popsána v sekci 5.2. Pro analýzu byla využita veškerá trénovací i testovací data.

8.1.1 Individuální analýza parametrů

Pro zjištění efektivnosti jednotlivých parametrů při určení žánru byla provedena jejich individuální analýza. Zpracováním pomocí strojového učení, přesněji metodou kvadratické SVM, byl každému z parametrů přidělen samostatný klasifikátor, kterým byla analyzována testovací data. Výsledná úspěšnost při zpracování jednotlivých žánrů je uvedena v příloze v tabulce D.1.

Z procentuální úspěšnosti uvedené v tabulce lze vyzorovat, že většina parametrů je vhodná pro určení jednoho až tří žánrů, pro ostatní žánry je úspěšnost nulová. Žádný z parametrů nemá takovou výpovědní hodnotu, aby jej bylo možné použít pro všechny žánry současně s alespoň přibližnou přesností jejich zařazení.

Parametry jsou v tabulce seřazeny na základě jejich významnosti, určené metodou mRMR. Z úspěšnosti klasifikace pomocí jednotlivých parametrů je zřejmé, že i tato metoda má své nedokonalosti. Problematické mohou být např. parametry třetí kvartil spektrálního toku Flux(Q3), medián Teager-Kaiserova energetického operátoru TKEO(median) nebo první kvartil efektivní hodnoty energie RMS(Q1). Tyto parametry se objevily mezi prvními deseti, vhodnými pro klasifikaci, jejich úspěšnost však není příliš vysoká. Naopak parametr střední hodnota hustoty událostí EventDensity(mean) je schopen rozřazení do tří žánrů s více než 50 % úspěšností, ačkoliv se nachází na 41. pozici v tabulce.

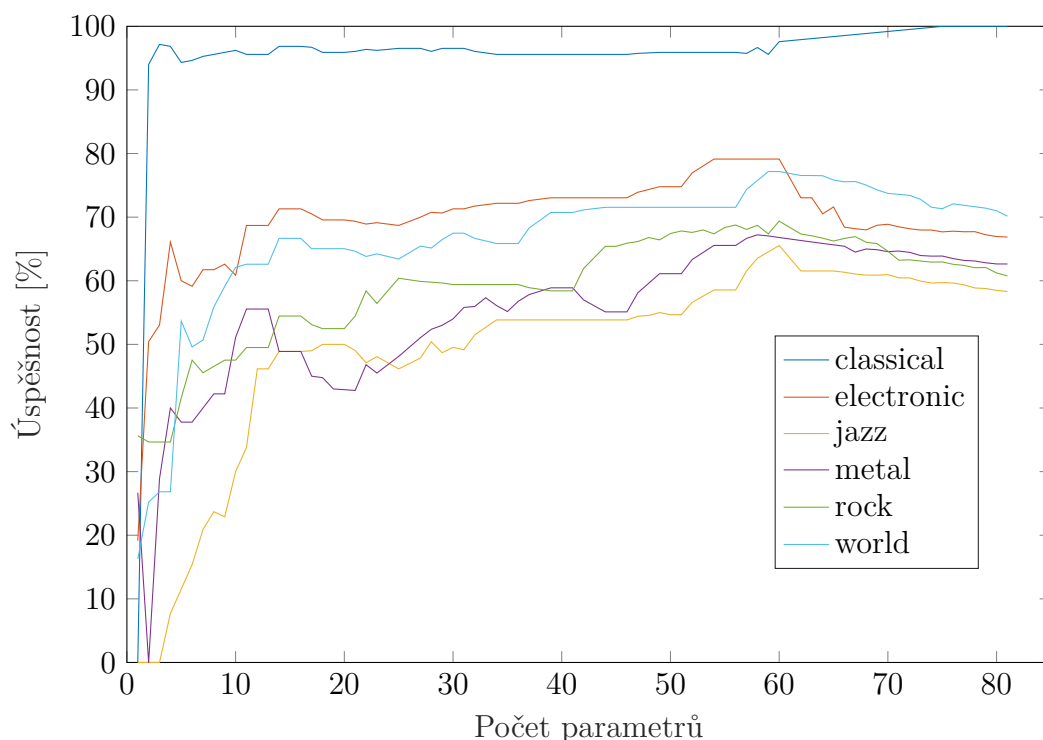
Lze si také povšimnout parametrů, nacházejících se na konci tabulky. Pro množství z nich totiž platí, že s jejich pomocí lze určit pouze jeden ze šesti žánrů, zbývajících pět bylo zařazeno ve všech případech nesprávně. Toto zjištění tedy vypovídá o skutečnosti, že jsou dané parametry nejen neefektivní pro výpočet, ale navíc snižují celkovou úspěšnost klasifikace.

Nejproblematictějším žánrem je na základě individuální analýzy jazz, který u žádného z jednotlivých parametrů nedosahuje úspěšnosti ani 40 % a ve většině případů je úspěšnost jeho zařazení dokonce nulová.

8.1.2 Přesnost klasifikace v závislosti na počtu parametrů

V následujícím kroku byl hledán počet parametrů, pro který je klasifikace nejpřesnější. Klasifikátory byly v této fázi vytvářeny postupně pro rostoucí počet paramet-

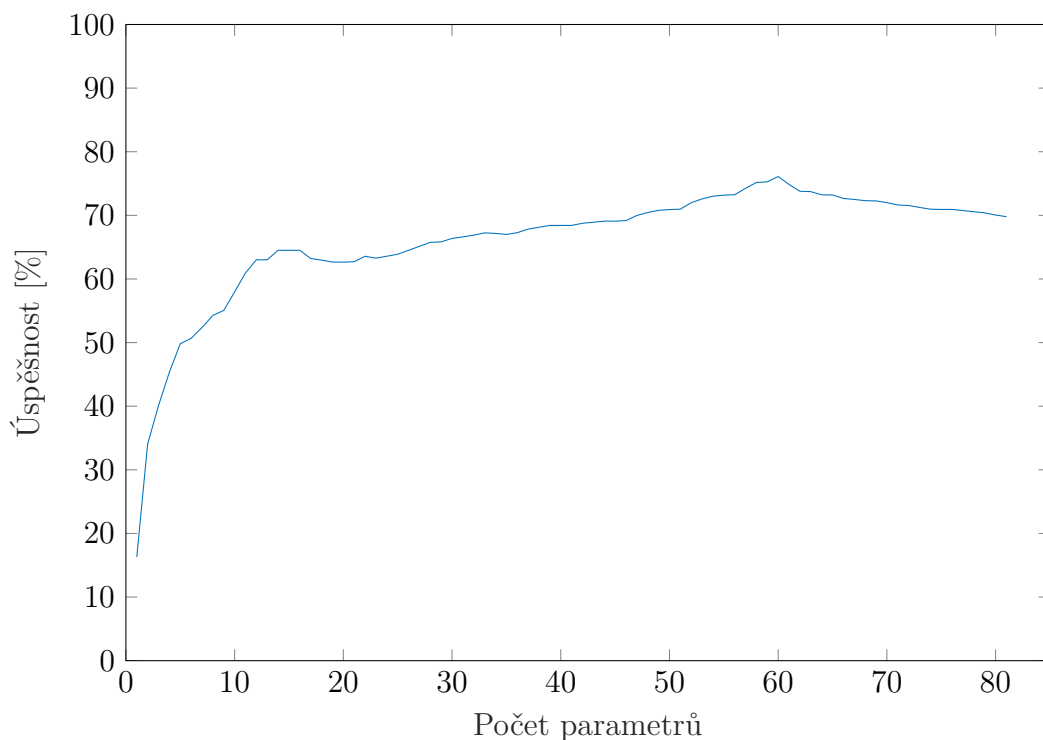
trů, seřazených podle jejich významnosti pro klasifikaci metodou mRMR. Následně byla těmito klasifikátory analyzována testovací databáze. Úspěšnost správného zařazení jednotlivých žánrů je vyobrazena v grafu 8.1.



Obr. 8.1: Úspěšnost klasifikace jednotlivých žánrů databáze ISMIR2004

Jako nejefektivnější počet parametrů se ukázala být hodnota šedesát, při které klasifikace dosahuje nejvyšší přesnosti, a to 76,1%. Při využití vyššího množství parametrů klesá úspěšnost vyhodnocení v důsledku využití parametrů o nízké vypovídací hodnotě z hlediska žánrů. Tyto parametry proto mohou nežádoucím způsobem zkreslovat klasifikaci. V grafu 8.2 je vyobrazena celková úspěšnost klasifikace v závislosti na počtu parametrů, využitých pro vytvoření modelu a analýzu.

Nejvyšší přesnosti zařazení bylo dosaženo u nahrávek klasické hudby, a to více než 95% při jakémkoli počtu využitých parametrů. Důvodem je diametrální odlišnost tohoto typu hudby od ostatních žánrů, zpracovávaných v této práci. Nejvýraznější změnu představuje barva zvuku, jež je silně ovlivněná volbou hudebních nástrojů. Klasická hudba totiž obsahuje výhradně akustické hudební nástroje, zatímco populární hudba ve velké míře využívá elektrické a elektronické hudební nástroje, jejichž odlišný charakter má výrazný vliv na frekvenční spektrum a tím i barvu zvuku. Další podstatnou vlastností klasické (artificiální) hudby je její velký dynamický rozsah, tedy rozdíl hlasitosti tichých a hlasitých pasáží, oproti nonartificiální hudbě, kde je dynamika téměř konstantní.



Obr. 8.2: Celková úspěšnost klasifikace databáze ISMIR2004

Jazz je žánr, který se v současnosti umísťuje na pomezí arificiální a nonartificiální hudby. U tohoto datasetu nebylo dosaženo vysoké úspěšnosti zařazení tohoto žánru, pouze necelých 66 %. Důvodem je malý počet trénovacích nahrávek (pouhých 26 nahrávek), a tím také nedostatečné množství informací určených pro vytvoření modelu tohoto žánru. Další příčinou může být zmíněná podobnost současně s klasickou i populární hudbou, která může být při klasifikaci zavádějící. Problematická může být také variabilita nástrojů, využívaných v jednotlivých skladbách, a stylu hry obecně.

Přesnějšího zařazení bylo dosaženo u žánrů metal a rock, konkrétně 68 % a 69 %. Tyto žánry jsou si v mnoha ohledech velice podobné, zejména z hlediska používaných hudebních nástrojů a tím i podobné barvy zvuku, stejně jako nízkého dynamického rozsahu. Z toho důvodu docházelo k chybám v zařazení zejména mezi těmito dvěma žánry navzájem, tedy vyhodnocení metalu jako rock a naopak. Tato skutečnost je patrná v tabulce 8.1. Dalším možným důvodem nepříliš vysoké úspěšnosti klasifikace je chybné zařazení nahrávek k jednotlivým žánrům samotným autorem databáze, jak již bylo diskutováno v mnohých pracích (např. [32, 33, 39]).

Výrazně vyšší úspěšnosti klasifikace bylo dosaženo u elektronické hudby (79 %). Tento žánr je autentický svou „umělou“ barvou zvuku, který nevzniká přirozeně snímáním hudebních nástrojů nebo lidského hlasu, ale je generován pomocí syn-

Tab. 8.1: Přesnost rozřazení datasetu ISMIR2004 udaná v procentech

	classical	electronic	jazz	metal	rock	world
classical	97,58	0,00	0,00	0,00	0,32	2,10
electronic	0,87	79,13	0,87	0,87	6,09	12,17
jazz	0,00	15,38	65,54	0,00	10,54	10,54
metal	1,22	3,67	0,00	67,78	27,33	0,00
rock	1,98	8,82	0,99	10,91	69,38	7,92
world	10,20	6,13	0,00	0,81	5,69	77,17

tetizérů, případně vzniká softwarově přímo v prostředí počítače. Nejčastěji došlo k jejímu zaměnění za žánr world.

S úspěšností 77 % byla klasifikována etnická hudba, označovaná jako world music. Pro ni jsou specifické různé etnické nástroje, a tím také „akustická“ barva zvuku. Z tohoto důvodu byla klasifikátorem nejčastěji nesprávně označena za klasickou hudbu.

8.1.3 Další klasifikační metody

Pro srovnání byla provedena analýza optimálního množství parametrů také dalšími metodami strojového učení. Výsledky celkové úspěšnosti této metody jsou uvedeny v tabulce 8.2. Nejvyšší úspěšnosti dosáhla podle předpokladu metoda kvadratické Support Vector Machines, následována dalšími metodami SVM.

Tab. 8.2: Klasifikace databáze ISMIR2004 dalšími metodami strojového učení

Název metody	Úspěšnost
kvadratické SVM	76,0 %
lineární diskriminant	68,4 %
lineární SVM	73,4 %
střední Gaussovské SVM	70,4 %
10-NN	65,1 %
váňované k -NN	68,4 %
bagged trees	68,6 %
boosted trees	63,2 %

8.2 Unique dataset

Databáze Unique byla blíže popsána v sekci 5.5. Z tohoto datasetu byly vybrány pouze nahrávky zařazené do stejných žánrů jako nahrávky předchozího datasetu, tedy klasická hudba, elektronická hudba, jazz, rock a world music. Důvodem bylo jejich využití k přehlednějšímu porovnání s předchozí databází. Unique však neobsahuje kategorii metal. Posléze byly jednotlivé žánrové skupiny rovnoměrně rozděleny na trénovací a testovací data.

Významnou odlišností těchto dvou databází je délka nahrávek – ISMIR2004 obsahuje skladby v plné délce, zatímco nahrávky v Unique byly zkráceny na délku 30 s. Proto nebylo možné pro obě skupiny využít stejná trénovací data.

8.2.1 Individuální analýza parametrů

Také v případě Unique datasetu byla provedena individuální analýza parametrů, získaných zpracováním této databáze. Výsledná úspěšnost správného zařazení je pro jednotlivé parametry uvedena v příloze v tabulce D.2.

Z této tabulky je patrné, že velmi dobrý vliv mají na klasifikaci parametry střední hodnota RMS a směrodatná odchylka spektrálního toku, které dosahují úspěšnosti větší než 80 % pro dva žánry současně. Mezi další významné parametry lze zařadit počet průchodů nulovou hodnotou, Melovské keprální koeficienty (zejm. první, druhý a čtvrtý) nebo první percentil Teager-Kaiserova energetického operátoru.

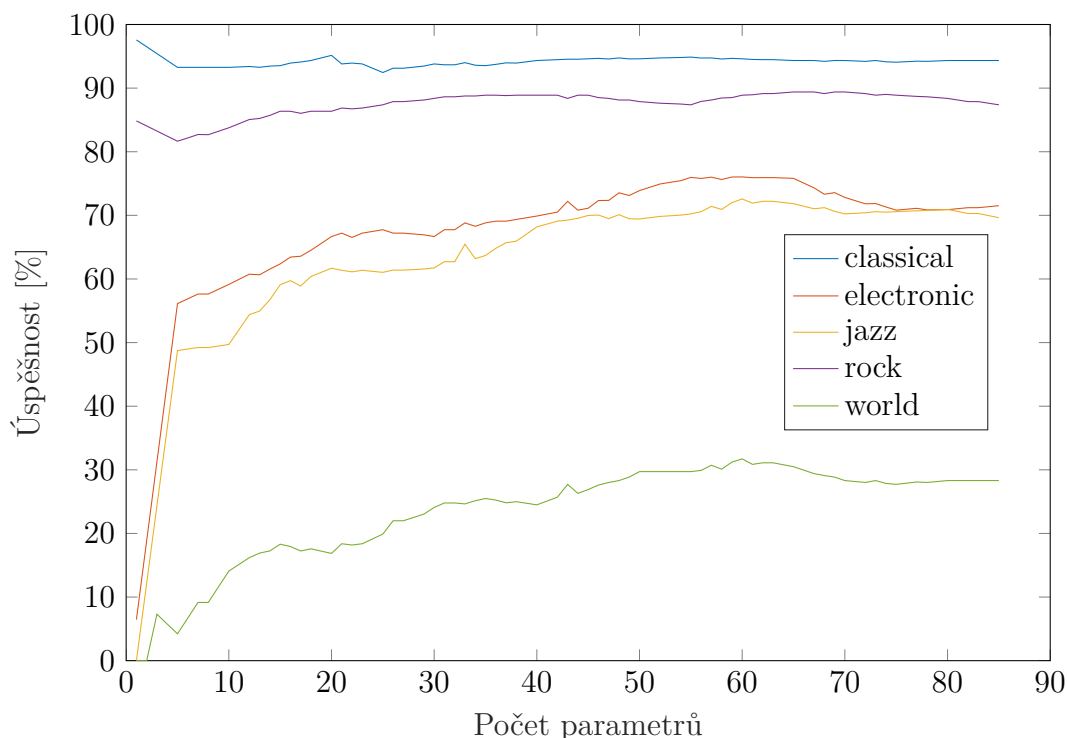
Na rozdíl od předchozího datasetu nedochází v případě Unique téměř vůbec k situacím, kdy lze pomocí jednoho parametru správně zařadit pouze jediný žánr a úspěšnost určení zbývajících je nulová. I ke konci tabulky, kde jsou umístěny parametry s nejnižší relevancí, se pravděpodobnost správné klasifikace pro jeden z žánrů pohybuje často nad hodnotou 70 %.

Nejhorších výsledků dosáhl žánr world, pro který bylo dosaženo nejvyšší hodnoty 76 %, a to v případě 99. percentilu parametru Brightness, jež se nachází až na konci tabulky. Pro téměř polovinu parametrů je úspěšnost jeho klasifikace nulová. Lze tedy předpokládat, že tento žánr nedosáhne vysoké úspěšnosti zařazení.

Naopak je tomu u klasické a rockové hudby, jež dosahují pro mnoho parametrů vysoké přesnosti zařazení, často i více než 90 %. Současně nastává nulová úspěšnost jejich zařazení spíše výjimečně.

8.2.2 Přesnost klasifikace v závislosti na počtu parametrů

Také pro databázi Unique byla vytvořena závislost přesnosti určení jednotlivých žánrů na počtu využitých parametrů. Graf této závislosti je zaznamenán na obrázku 8.3.



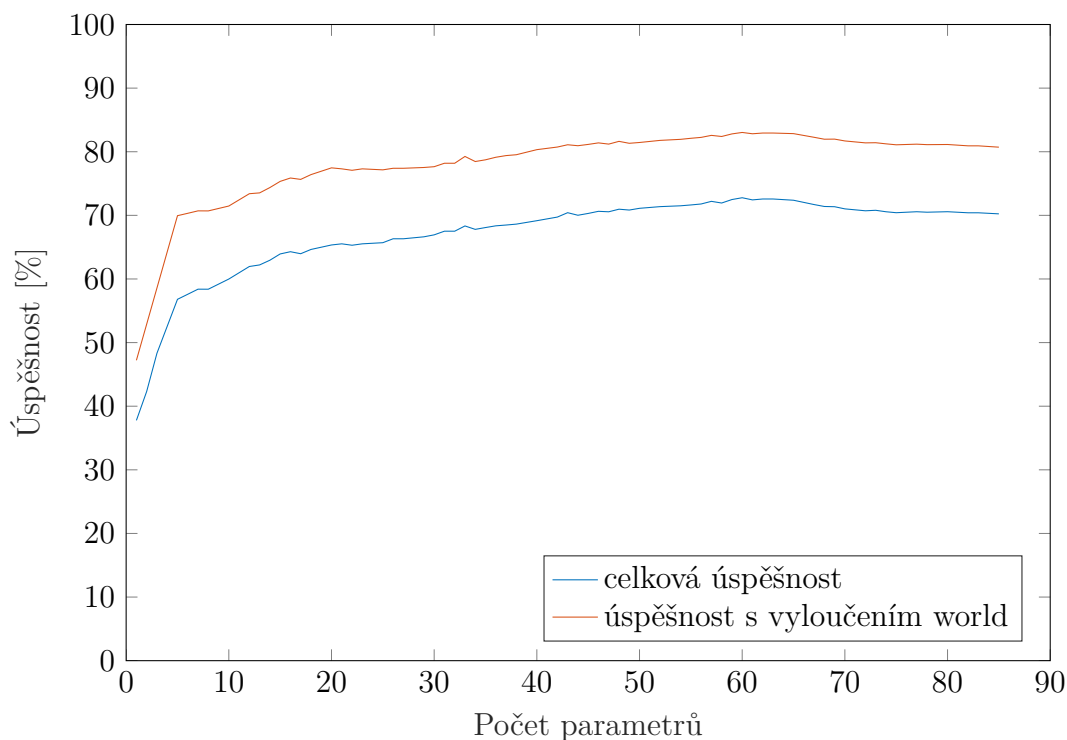
Obr. 8.3: Úspěšnost klasifikace jednotlivých žánrů databáze Unique

Přesnost klasifikace dosahuje vysokých hodnot i při malém počtu parametrů, jako neoptimálnější se však opět projevila hodnota šedesát. Při ní dosahuje klasifikace úspěšnosti 72,7 %. Výrazně negativně se zde však projevuje vliv žánru world, jehož úspěšnost dosahuje pouhých 31 %. S vyloučením této třídy dosahuje klasifikace přesnosti až 83,0 %. Rozdíl je patrný v následujícím grafu 8.4, kde modrá čára označuje celkovou úspěšnost a červená čára úspěšnost s vyloučením žánru world.

Problematika žánru world, tedy etnické hudby, spočívá s největší pravděpodobností ve volbě skladeb pro trénovací a testovací data. Mnohé z nich obsahují pouze sólové akustické nástroje (píšťaly, bubny, etnické strunné nástroje apod.), čímž se barvou zvuku blíží klasické hudbě a mohly by za ni být považovány i při subjektivním třídění nahrávek. Mnohé z nahrávek jsou také nekvalitní a zatížené šumem, což má vliv na výpočet parametrů, definujících dynamiku. Vzhledem ke skutečnosti, že byla tímto způsobem sestavena i trénovací data, byly nevhodným způsobem vytvořeny samotné klasifikátory. Z tabulky 8.3 je patrné, že pouze malé množství nahrávek kteréhokoli žánru bylo označeno za etnickou hudbu.

Nejvyšší úspěšnosti bylo opět dosaženo analýzou klasické hudby, a to téměř 95 %. Jak již bylo dříve zmíněno, je to způsobeno výraznou odlišností tohoto typu hudby, stejně jako největším množstvím trénovacích dat.

Vysoká přesnost zařazení byla získána také analýzou rockového žánru. Více



Obr. 8.4: Celková úspěšnost klasifikace databáze Unique

Tab. 8.3: Přesnost rozřazení datasetu Unique udaná v procentech

	classical	electronic	jazz	rock	world
classical	94,61	0,00	4,31	0,54	0,54
electronic	2,15	76,04	11,83	10,75	3,23
jazz	12,99	5,84	72,58	11,69	3,90
rock	1,01	2,53	6,57	88,89	1,01
world	25,94	5,63	26,03	12,68	31,72

než 90 % nahrávek této kategorie bylo zařazeno správně. V tomto datasetu je třída rock tvořena rockovými i metalovými skladbami, čímž odpadá problematika zmíněná u datasetu ISMIR2004. Výhodou je také dvojnásobné množství trénovacích nahrávek. Nejčastěji došlo k záměně s jazzem, který je nejpodobnější s ohledem na rytmické struktury i hudební nástroje.

Zpracováním elektronické hudby bylo dosaženo přesnosti zařazení 76,0 %, to znamená mírně nižší než u databáze ISMIR2004. Naopak úspěšnost klasifikace jazzu vzrostla na 72,6 %. Důvodem je v obou případech změna množství trénovacích dat. U elektronické hudby došlo k poklesu ze 115 na 97 nahrávek, naopak jazz obsahuje 158 oproti předešlým 26 nahrávkám.

8.2.3 Další klasifikační metody

Také pro databázi Unique byla provedena klasifikace dalšími metodami strojového učení. Nejvyšší celkové úspěšnosti dosáhly opět kvadratické Support Vector Machines (72,7 %), následovány lineárním diskriminantem (72,2 %). Další hodnoty jsou uvedeny v tabulce 8.4.

Tab. 8.4: Klasifikace databáze Unique dalšími metodami strojového učení

Název metody	Úspěšnost
kvadratické SVM	72,7 %
lineární diskriminant	72,2 %
lineární SVM	72,1 %
střední Gaussovské SVM	69,1 %
10-NN	66,0 %
váňované k -NN	66,8 %
bagged trees	63,1 %
boosted trees	64,0 %

8.3 Free Music Archive

Posledním zpracovaným datasetem je Free Music Archive, charakterizovaný v sekci 5.3. Z něj bylo vybráno stejných šest žánrů, které se objevují v ISMIR2004. Každá kategorie obsahuje sto trénovacích a sto testovacích úryvků nahrávek o délce 30 s.

8.3.1 Individuální analýza parametrů

Po výpočtu parametrů databáze Free Music Archive byla opět jako první krok provedena jejich individuální analýza. Získané hodnoty jsou uvedeny v příloze v tabulce D.3.

Jako nejvýhodnější parametry pro žánrovou klasifikaci byly určeny medián spektrálního toku, počet průchodu nulovou hodnotou a Melovské keprální koeficienty (zejména první, třetí a jedenáctý).

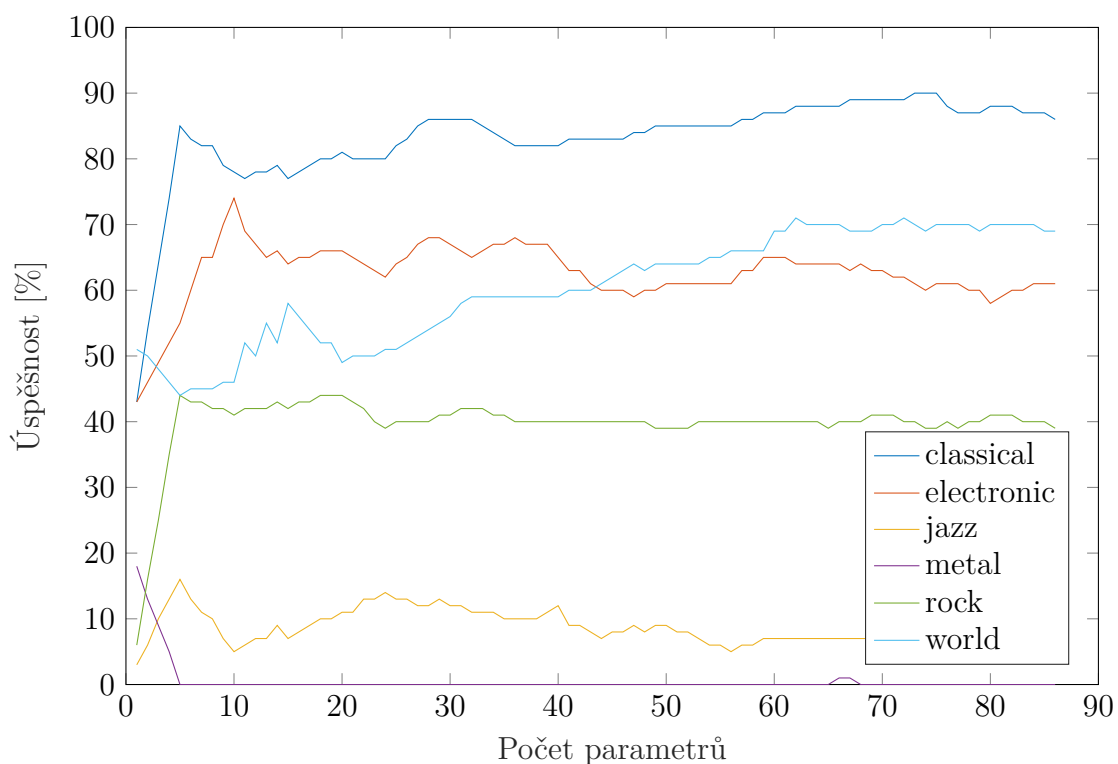
V porovnání s individuální analýzou parametrů předchozích databází je patrné, že přesnost zařazení je v případě jednotlivých parametrů Free Music Archivu nižší, což bude mít negativní vliv na celkovou úspěšnost klasifikace.

Nejvyšších hodnot dosahuje opět klasická hudba, ačkoliv míra úspěšnosti není tak vysoká jako u předchozích dvou datasetů. Naopak nejproblematictějšími žánry

jsou podle tabulky jazz a metal, jejichž úspěšnost dosahuje nejvýše 60 %, pro více než polovinu parametrů je však nulová.

8.3.2 Přesnost klasifikace v závislosti na počtu parametrů

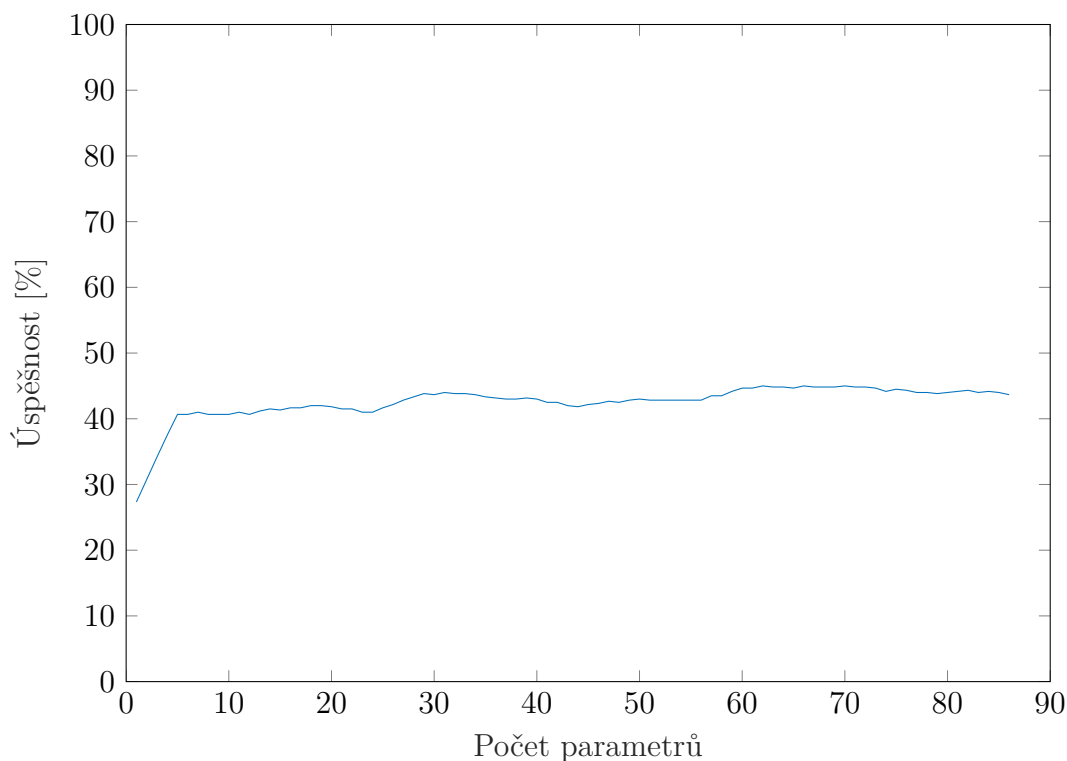
Graf závislosti úspěšnosti žánrového zařazení na množství parametrů je uveden na obrázku 8.5. Ta se pro jednotlivé žánry výrazně odlišuje. Určení klasické a world hudby dosáhlo poměrně vysoké přesnosti, naopak metal nebyl určen správně v téměř žádném z případů. Dále lze v grafu pozorovat, že křivka této závislosti má pro každý žánr jiný tvar. Úspěšnost zařazení elektronické hudby je nejvyšší pro pouhých deset parametrů a následně klesá. Naopak přesnost klasifikace etnické hudby (world) narůstá postupně a je nejvyšší pro šedesát jedna parametrů.



Obr. 8.5: Žánrová úspěšnost klasifikace databáze Free Music Archive

Graf 8.6 zobrazuje závislost celkové úspěšnosti klasifikace na počtu využitých parametrů. Ta je v celém rozsahu přibližně konstantní a pohybuje se v rozmezí 40 % až 45 %. Nejvyšší celková úspěšnost byla dosažena pro šedesát šest parametrů.

Problém nízké úspěšnosti databáze Free Music Archive spočívá s největší pravděpodobností ve způsobu zařazení trénovacích nahrávek. V této práci byla využita verze „Small“, tedy výběr základních žánrů z celkového počtu 161 subžánrů. Tyto podkategorie byly zvoleny tak, aby co nejlépe specifikovaly daný subžánr,



Obr. 8.6: Celková úspěšnost klasifikace databáze Free Music Archive

není ovšem vyloučeno, že v rámci větších skupin, tedy žánrů, nemají vhodnou výpočetní hodnotu. Dalším možným problémem je snížení počtu trénovacích dat na pouhých sto nahrávek v každé kategorii. Toto množství se však projevilo v databázi ISMIR2004 jako dostačující.

V tabulce 8.5 je uvedena přesnost rozřazení jednotlivých žánrů. Nejvyšší úspěšnosti bylo opět dosaženo u klasické hudby.

Elektronická hudba byla vyhodnocena s přesností 65 %, k nejčastější záměně došlo s klasickou hudbou a rockem. V tabulce lze však pozorovat, že má tento žánr na klasifikaci negativní vliv, neboť za něj byly často označovány ostatní kategorie.

Další hudební oblastí je jazz, který dosáhl úspěšnosti klasifikace pouhých 7 % a ve většině případů byl chybně označen právě jako zmíněná elektronická hudba.

Nejhorších výsledků v databázi Free Music Archive však docílil metal, pouhé jedno procento. Z tabulky je ale patrné, že nejčastěji došlo k jeho záměně za rock, tedy žánr velice podobný. Také lze pozorovat, že problém spočívá opět v trénovacích datech, neboť do této kategorie nebyla z ostatních žánrů zařazena žádná nahrávka.

Úspěšnost zařazení rockového žánru dosáhla 40 %. Chybně byl označen nejčastěji jako elektronická hudba, ale také jako klasická a etnická hudba.

Vyšší úspěšnost byla získána zpracováním etnické hudby, přesně 70 %. Nejčastěji byl tento žánr nesprávně klasifikován jako klasická nebo rocková hudba.

Tab. 8.5: Přesnost rozřazení datasetu Free Music Archive udaná v procentech

	classical	electronic	jazz	metal	rock	world
classical	88,00	3,00	3,00	0,00	3,00	3,00
electronic	14,00	65,00	4,00	0,00	11,00	5,00
jazz	14,00	68,00	7,00	0,00	2,00	9,00
metal	1,00	3,00	0,00	1,00	89,00	7,00
rock	18,00	24,00	0,00	0,00	40,00	18,00
world	9,00	7,00	4,00	0,00	10,00	70,00

8.3.3 Další klasifikační metody

V následující tabulce 8.6 jsou uvedeny úspěšnosti jednotlivých metod klasifikace databáze Free Music Archive pro optimální počet parametrů, tedy šedesát šest. Hodnoty jsou výrazně nižší než u předchozích databází. Nejvyšší celková úspěšnost je uvedena opět pro kvadratické Support Vector Machines (56,6%).

Tab. 8.6: Klasifikace databáze Free Music Archive dalšími metodami strojového učení

Název metody	Úspěšnost
kvadratické SVM	56,6 %
lineární diskriminant	53,1 %
lineární SVM	55,1 %
střední Gaussovské SVM	55,7 %
10-NN	48,0 %
váňované k -NN	43,9 %
bagged trees	42,9 %
boosted trees	50,0 %

8.4 Shrnutí výsledků

Při zpracování databáze ISMIR2004 bylo dosaženo celkové úspěšnosti zařazení žánrů 76,1 %, tedy o jedno procento více než v práci G. Marquese [33], který taktéž využíval tento dataset. Využita byla stejná metoda klasifikace a vypočítány přibližně podobné parametry. Výjimkou jsou TKEO a RMS, které tato diplomová práce obsahuje navíc a ukázaly se jako efektivní.

Dalším autorem, jež se zabýval analýzou ISMIR2004, je C. Lee [27]. Ten dosáhl přesnosti klasifikace dokonce 91 %. Docílil toho využitím výpočetně náročných pa-

rametrů oktávový spektrální kontrast (OSC) a normalizovaná zvuková spektrální obálka (NASE).

Analýzou databáze Unique bylo dosaženo celkové úspěšnosti 72,7 %, s vyloučením problematického žánru world až 83,0 %. Tento soubor nahrávek není příliš známý, proto jej nelze porovnat s žádnými světovými výzkumy.

Méně přesných výsledků bylo dosaženo analýzou databáze Free Music Archive, kdy se celková přesnost zařazení pohybovala pouze kolem 45 %. Příčinou byla s největší pravděpodobností nevhodně rozřazená trénovací data, neboť žánry klasická, elektronická a etnická hudba dosáhly vysokých úspěšností, ale naopak přesnost určení jazzu a metalu byla téměř nulová.

V rámci klasifikace jednotlivých žánrů dosáhla ve všech případech nejvyšší úspěšnosti klasická hudba, konkrétně 97,6 %, 94,6 % a 88,0 %. Tento žánr je specifický svou barvou zvuku i dynamickým rozsahem.

Zřejmě nejproblematictější se ukázal být žánr jazz. Důvodem může být jeho podobnost s klasickou a současně rockovou hudbou, stejně jako nevhodně zvolené nahrávky pro trénovací a testovací data.

Na přesnost zařazení mělo výrazný vliv množství nahrávek, obsažených v trénovacích databázích. Při malém počtu skladeb značně poklesla úspěšnost klasifikace. Dalším podstatným faktorem je délka skladby. Ta se projevuje zejména ve výpočtu tempa, jež se v rámci nahrávky o plné délce několikrát mění a určení jeho střední hodnoty je proto neefektivní. Problematické je také subjektivní žánrové rozřazení samotnými tvůrci databází, které se mnohdy nemusí slučovat s názorem většiny.

Jak již bylo zmíněno v sekci 2.1, označení žánru je výrazně subjektivní. Ve studii S. Lippense [32] bylo při rozdělování nahrávek jednotlivými posluchači do šesti žánrů dosaženo shody pouze 76 %, což je také klasifikační úspěšnost šestižánrové databáze ISMIR2004 v této práci.

9 Závěr

Tato práce se zabývá zpracováním zvukového signálu za pomoci technik Music Information Retrieval a následnou analýzou jeho hudebního žánru.

V první kapitole je stručně popsána oblast výzkumu, zvaná Music Information Retrieval, způsoby jejího využití a aspekty, jimiž lze definovat hudební dílo. Následující kapitola se zabývá její podoblastí Music Genre Recognition. Je zde vysvětlen pojem *žánr* a shrnuty poznatky výzkumů, zabývajících se automatickým rozpoznáváním hudebního žánru.

Třetí kapitola se věnuje parametrizaci signálu. Jsou zde uvedeny způsoby třídění parametrů a popsány nejdůležitější parametry, vhodné pro určení žánru.

Další kapitola se týká strojového učení a způsobů jeho využití. Také jsou zde stručně charakterizovány metody, nejčastěji využívané v oblasti Music Genre Recognition. Těmi jsou AdaBoost, Gaussův smíšený model, k -NN, lineární diskriminační analýza, náhodné lesy, naivní Bayesův klasifikátor, skrytý Markovův model, Support Vector Machines a umělé neuronové sítě.

V páté kapitole jsou představeny datasey nahrávek, které byly vytvořeny pro výzkumy žánrové klasifikace. U každého z nich je uveden jeho obsah, úspěšnost při rozpoznávání žánru i případné nedostatky a chyby.

Šestá kapitola této práce obsahuje návrh vyhodnocovacího systému pro určení žánru a je zde zobrazeno také blokové schéma, které bylo využito k implementaci tohoto systému.

Následuje evaluace systému, kde je blíže popsán způsob práce při zpracování hudebních nahrávek. Obsahuje volbu parametrů definujících žánr, seřazení parametrů na základě jejich relevance a následně metody, které mohou být využity k žánrové klasifikaci, včetně jejich předpokládané úspěšnosti.

Poslední kapitola se zabývá vyhodnocením získaných hodnot pro tři zvolené databáze nahrávek, kterými byly ISMIR2004, Unique a Free Music Archive. První podkapitola každého datasetu se věnuje individuální analýze jednotlivých parametrů a jejich dopadu na správné žánrové zařazení. Další sekce rozebírá vliv množství využitých parametrů na správnou klasifikaci a obsahuje vyobrazení této závislosti v grafu. Dále jsou zde diskutovány příčiny vysoké i nízké přesnosti zařazení jednotlivých žánrů. Nakonec došlo k sestavení tabulky přesnosti zařazení pro optimální počet parametrů.

Jako nejpodstatnější parametry byly určeny Melovské keprální koeficienty, spektrální tok a Teager-Kaiserův energetický operátor, který se ve světových výzkumech žánrové klasifikace (viz. tabulka B.1) obvykle nevyužívá. Za nejspěšnější metodu klasifikace byly vyhodnoceny kvadratické Support Vector Machines.

Celková úspěšnost klasifikace databáze ISMIR2004 dosáhla hodnoty 76,1 %. Vybrané žánry Unique datasetu byly zařazeny s přesností 72,7 %. Nejnižší úspěšnosti bylo dosaženo analýzou nahrávek Free Music Archivu, a to 45 %.

V budoucích výzkumech by bylo možné dosáhnout vyšších výsledků využitím většího množství parametrů, které popisují barvu zvuku. Těmi jsou např. frekvenční rozptyl, šikmost a špičatost, případně také entropie. Dále by bylo vhodné věnovat větší pozornost žánrovému zařazení nahrávek ve zpracovávaných databázích, a to volbou skladeb, které jsou pro daný žánr co nejtypičtější. Toto rozdělení lze potvrdit subjektivními testy, v nichž by nahrávky do jednotlivých žánrů zařazovala skupina jedinců s hudebním vzděláním.

Literatura

- [1] ALBERT, M., K.; AHA, D., W.: *Analyses of instance-based learning algorithms*. [online]. University of California, Irvine, 1991. [cit. 25. 10. 2018]. Dostupné z URL: <<https://www.aaai.org/Papers/AAAI/1991/AAAI91-086.pdf>>.
- [2] BASILI, R.; SERAFNI, A.; STELLATO, A.: *Classification of musical genre: A machine learning approach*. [online]. University of Rome Tor Vergata, Rome, 2004. [cit. 23. 10. 2018]. Dostupné z URL: <http://haralick.org/ML/CLASSIFICATION_OF_MUSICAL_GENRE_A_MACHINE_LEARNING_APPROACH.pdf>.
- [3] BELL, R., M.; KOREN, Y.: *Lessons from the netflix prize challenge*. [online]. SIGKDD Explorations 9, no. 2, 2007. [cit. 19. 10. 2018]. Dostupné z URL: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.401.8553&rep=rep1&type=pdf>>.
- [4] BERTIN-MAHIEUX, T.; ELLIS, D., P., W.; WHITMAN, B.; LAMERE, P.: *Million Song Dataset*. [online]. Columbia University, The Echo Nest, Somerville, 2011. [cit. 29. 10. 2018]. Dostupné z URL: <<https://www.ee.columbia.edu/~dpwe/pubs/BertEWL11-msd.pdf>>.
- [5] BREIMAN, L.: *Random Forests*. [online]. University of California, Berkeley, 2001. [cit. 15. 4. 2019]. Dostupné z URL: <<https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>>.
- [6] BOU-RABEE, A.; GO, K.; MOHAN, K.: *Classifying the Subjective: Determining Genre of Music From Lyrics*. [online]. 2012. [cit. 23. 10. 2018]. Dostupné z URL: <<http://cs229.stanford.edu/proj2012/BourabeeGoMohan-ClassifyingTheSubjectiveDeterminingGenreOfMusicFromLyrics.pdf>>.
- [7] BURRED, J., J.; LERCH, A.: *Hierarchical Automatic Audio Signal Classification*. [online]. Technical University Berlin, Berlin, 2004. [cit. 6. 10. 2018]. Dostupné z URL: <https://www.researchgate.net/publication/200054331_Hierarchical_Automatic_Audio_Signal_Classification>.
- [8] CASEY, M.; VELTKAMP, R.; aj.: *Content-Based Music Information Retrieval: Current Directions and Future Challenges*. [online]. Stanford University, Stanford, 2008. [cit. 2. 11. 2018]. Dostupné z URL: <<http://www.slaney.org/malcolm/yahoo/Casey2008-IEEEProceedingsContent-BasedMusicInformationRetrieval.pdf>>.

- [9] DALIN-VOLSING, S.: *Classification Of Musical Genres Using Hidden Markov Models*. [online]. Centre for Mathematical Sciences, Lund University, 2017. [cit. 21. 10. 2018]. Dostupné z URL: <<http://www.maths.lth.se/matstat/kurser/fmsf15masc03/CoMG.pdf>>.
- [10] DEFFERRARD, M.; BENZI, K.; VANDERGHEYNST, P.; BRESSON, X.: *FMA: A Dataset For Music Analysis*. [online]. LTS2, EPFL, Switzerland; SCSE, NTU, Singapore, 2017. [cit. 29. 10. 2018]. Dostupné z URL: <https://ismir2017.smcnus.org/wp-content/uploads/2017/10/75_Paper.pdf>.
- [11] DIMITRIADIS, D.; POTAMIANOS, A.; MARAGOS, P.: *A Comparison of the Squared Energy and Teager-Kaiser Operators for Short-Term Energy Estimation in Additive Noise*. [online]. IEEE T Signal Proces, ročník 57, č. 7, 2009. [cit. 15. 10. 2018]. Dostupné z URL: <http://cvsp.cs.ntua.gr/publications/jpubl+bchap/DimitriadisPotamianosMaragos_ComparisonSquaredAmpl-TK0per-EnergyEstimation_ieeetSP2008.pdf>.
- [12] DIXON, S.; PAMPALK, E.; WIDMER, G.: *Classification of Dance Music by Periodicity Patterns*. [online]. Austrian Research Institute for AI, Vienna, 2003. [cit. 5. 10. 2018]. Dostupné z URL: <<http://www.eecs.qmul.ac.uk/~simond/pub/2003/ismir-dance.pdf>>.
- [13] DOWNIE, S., J.: *Music Information Retrieval*. [online]. Annual Review of Information Science and Technology 37, Medford, 2003. [cit. 19. 11. 2018]. Dostupné z URL: <http://www.music.mcgill.ca/~ich/classes/mumt611_06/downie_mir_arist37.pdf>.
- [14] DUBNOV, S.: *Generalization of Spectral Flatness Measure for Non-Gaussian Linear Processes*. [online]. Ben Gurion University of the Negev, Beer-Sheva, 2004. [cit. 16. 10. 2018]. Dostupné z URL: <<http://musicweb.ucsd.edu/~sdubnov/Papers/SPL.pdf>>.
- [15] FOOTE, J.: *A Similarity Measure for Automatic Audio Classification*. [online]. National University of Singapore, 1997. [cit. 10. 11. 2018]. Dostupné z URL: <<https://pdfs.semanticscholar.org/a51f/a2abe440481df5afd9a014c6ab1faa93ac05.pdf>>.
- [16] FREUND, Y.; SCHAPIRE, R.: *Experiments with a New Boosting Algorithm*. [online]. AT&T Laboratories, Murray Hill, NJ, 1996. [cit. 22. 10. 2018]. Dostupné z URL: <<https://cseweb.ucsd.edu/~yfreund/papers/boostingexperiments.pdf>>.

- [17] GONZALEZ-IZAL, M.; RODRIGUEZ-CARRENO, I.; MALANDA, A.; MALLOR-GIMENEZ, F.; NAVARRO-AMEZQUETA, I.; aj: *sEMG waveletbased indices predicts muscle power loss during dynamic contractions*. [online]. Electromyogr Kines, ročník 20, č. 6, 2010. [cit. 15. 10. 2018]. Dostupné z URL: <<http://www.sciencedirect.com/science/article/pii/S105064111000088X>>.
- [18] CHAI, W.; VERCOE, B.: *Folk Music Classification Using Hidden Markov Models*. [online]. Massachusetts Institute of Technology, Cambridge, MA, USA, 2001. [cit. 21. 10. 2018]. Dostupné z URL: <<http://compbio.fmph.uniba.sk/vyuka/gm/old/2010-02/handouts/Chai2001.pdf>>.
- [19] CHANG, K., K.; JANG, J.; ILIOPOULOS, C.: *Music Genre Classification via Compressive Sampling*. [online]. King's College London, London, 2010. [cit. 11. 10. 2018]. Dostupné z URL: <http://www.mirlab.org/conference_papers/International_Conference/ISMIR%202010/ISMIR_2010_papers/ismir2010-66.pdf>.
- [20] ISMIR, 2000-2017. [online]. [cit. 2. 11. 2018]. Dostupné z URL: <<https://www.ismir.net/>>.
- [21] JUSLIN, P., N.: *Cue utilization in communication of emotion in music performance: relating performance to perception*. [online]. Journal of Experimental Psychology: Human Perception and Performance, 26(6), Uppsala University, 2000. [cit. 14. 10. 2018]. Dostupné z URL: <<https://benediktsvogler.com/downloads/DocumentationAudiotechnikMusicGenreRecognition.pdf>>.
- [22] KNEES, P.; SCHEDL, M.: *Music Similarity and Retrieval: An Introduction to Audio- and Web-based Strategies*. Department of Computational Perception, Johannes Kepler University, Linz. Springer, 2016. ISBN 978-3-662-49722-7
- [23] KOSINA, K.: *Music Genre Recognition*. [online]. Fachhochschul Hagenberg, 2002. [cit. 4. 11. 2018]. Dostupné z URL: <http://www.music.mcgill.ca/~ich/classes/mumt611_07/similarity/KosinaMusicGenreRec.pdf>.
- [24] KOZA, J., R.; BENNETT, F., H; ANDRE, D.; KEANE, M., A.: *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. [online]. Artificial Intelligence in Design '96. Springer, Dordrecht, 1996. [cit. 19. 10. 2018]. Dostupné z URL: <https://link.springer.com/chapter/10.1007/978-94-009-0279-4_9>.

- [25] LAUKKA, P.; JUSLIN, P., N.; BRESIN, R.: *A dimensional approach to vocal expression of emotion*. [online]. Cognition and Emotion, Uppsala University, Uppsala, Sweden, 2005. [cit. 10.4.2019]. Dostupné z URL: <https://www.researchgate.net/profile/Roberto_Bresin/publication/232282425_A_dimensional_approach_to_vocal_expression_of_emotion/links/00463525dc954a44b4000000.pdf>.
- [26] LERCH, A.: *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Wiley-IEEE Press, New Jersey, 2012. 272 s. ISBN 9781439835524
- [27] LEE, C.; SHIH, J.; YU, K.; LIN, H.: *Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features*. [online]. Chung Hwa University, Hsinchu, 2009. [cit. 6.10.2018]. Dostupné z URL: <http://sjl.csie.chu.edu.tw/sjl/albums/userpics/10001/Automatic_music_genre_classification_based_on_modulation_spectral_analysis_of_spectral_and_cepstral_features.pdf>.
- [28] LI, S.; GUO, G.: *Content-Based Audio Classification and Retrieval Using SVM Learning*. [online]. Microsoft Research China, Beijing, 2003. [cit. 5.10.2018]. Dostupné z URL: <<https://pdfs.semanticscholar.org/b5e8/a947b46424d54c1ea2a44abb154b4eb2857d.pdf>>.
- [29] LI, T.; OGIHARA, M.; LI, Q.: *A Comparative Study on Content-Based Music Genre Classification*. [online]. University of Rochester, NY, University of Delaware, Newark, 2003. [cit. 23.10.2018]. Dostupné z URL: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.172.6419&rep=rep1&type=pdf>>.
- [30] LI, T.; OGIHARA, M.; TZANETAKIS, G.: *Music Data Mining*. CRC Press, Boca Raton, 2011. ISBN 978-3-662-49722-7
- [31] LIAW, A.; WIENER, M.: *Classification and Regression by Random Forest*. [online]. R news, 2002. ISSN 1609-3631 [cit. 15.4.2019]. Dostupné z URL: <https://www.researchgate.net/profile/Andy_Liaw/publication/228451484_Classification_and_Regression_by_RandomForest/links/53fb24cc0cf20a45497047ab/Classification-and-Regression-by-RandomForest.pdf>.
- [32] LIPPENS, S.; MARTENS, J. P.; DE MULDER, T.: *A comparison of human and automatic musical genre classification*. [online]. International Conference

- on Acoustics, Speech, and Signal Processing, 2004. [cit. 4. 11. 2018]. Dostupné z URL: <<https://ieeexplore.ieee.org/abstract/document/1326806>>.
- [33] MARQUES, G.; LANGLOIS, T.; GOUYON, F.; LOPES, M.; SORDO, M.: *Short-term Feature Space and Music Genre Classification*. [online]. DEETC-ISEL, Lisboa, 2011. [cit. 6. 10. 2018]. Dostupné z URL: <<http://dx.doi.org/10.1080/09298215.2011.573563>>.
- [34] LARTILLOT, O.: *MIRtoolbox 1.6.1, User's Manual*. Aalborg University, Denmark, Department of Architecture, Design and Media Technology, 2014. [cit. 14. 10. 2018]. Dostupné z URL: <<https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/projects2/past-projects/coe/materials/mirtoolbox/manual1-6-1.pdf>>.
- [35] MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A.: *Foundations of Machine Learning*. Massachusetts Institute of Technology, 2012. [cit. 19. 10. 2018]. ISBN 9780262018258. Dostupné z URL: <[https://books.google.cz/books?hl=cs&lr=&id=-ijiAgAAQBAJ&oi=fnd&pg=PR5&dq=\(2012\)+Foundations+of+Machine+Learning&ots=FUAVnRMSxI&sig=R-w3BgfF6ED7GhCftN5mshM9f11A&redir_esc=y#v=onepage&q&f=false](https://books.google.cz/books?hl=cs&lr=&id=-ijiAgAAQBAJ&oi=fnd&pg=PR5&dq=(2012)+Foundations+of+Machine+Learning&ots=FUAVnRMSxI&sig=R-w3BgfF6ED7GhCftN5mshM9f11A&redir_esc=y#v=onepage&q&f=false)>.
- [36] MULDER, D., G., J.: *Automatic Classification of Heavy Metal Music*. [online]. Korteweg-de Vries Instituut voor Wiskunde, Amsterdam, 2014. [cit. 11. 10. 2018]. Dostupné z URL: <<https://esc.fnwi.uva.nl/thesis/centraal/files/f812177081.pdf>>.
- [37] MÜLLER, M.: *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer International Publishing Switzerland, 2015, 483 s. ISBN 978-3-319-21945-5.
- [38] PÁLMASON, H.; JÓNSSON, B.; SCHEDL, M.; KNEES, P.: *Music Genre Classification Revisited: An In-Depth Examination Guided by Music Experts*. [online]. Reykjavík University, IT University of Copenhagen, Johannes Kepler University Linz, TU Wien, 2017. [cit. 6. 10. 2018]. Dostupné z URL: <http://cmmr2017.inesctec.pt/wp-content/uploads/2017/09/4_CMMR_2017_paper_71.pdf>.
- [39] PERROT, D.; GJERDIGEN, R.: *Scanning the dial: An exploration of factors in the identification of musical style*. [online]. Northwestern University; Trialgraphix-Kroll Ontrack, New York, 1999. [cit. 4. 11. 2018]. Dostupné z URL: <<http://faculty-web.at.northwestern.edu/music/gjerdigen/Papers/PubPapers/Scanning.pdf>>.

- [40] POHLE, T.; PAMPALK, E.; WIDMER, G.: *Evaluation of Frequently Used Audio Features for Classification of Music Into Perceptual Categories*. [online]. Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing, Riga, 2005. [cit. 14.10.2018]. Dostupné z URL: <http://www.cp.jku.at/research/papers/Pohle_CBMI_2005.pdf>.
- [41] PONCE DE LEÓN, P., J.; IÑESTA, J., M.: *Musical style classification from symbolic data: a two-styles case study*. [online]. Computer Music Modeling and Retrieval. Lecture Notes in Computer Science, vol. 2771. Springer, Heidelberg, 2004. [cit. 18.11.2018]. Dostupné z URL: <<https://pdfs.semanticscholar.org/3025/885a6c3afd818bc34c3b959460201f74205a.pdf>>.
- [42] RABINER, L., R.; JUANG, B., H.: *An Introduction to Hidden Markov Models*. [online]. IEEE ASSP MAGAZINE, 1986. [cit. 21.10.2018]. Dostupné z URL: <http://ai.stanford.edu/~pabbeel/depth_qual/Rabiner_Juang_hmms.pdf>.
- [43] RANDEL, D., M.: *The new Harvard dictionary of music*. Belknap Press, Cambridge, 1986.
- [44] SEYERLEHNER, K.: *Content-Based Music Recommender Systems: Beyond simple Frame-Level Audio Similarity*. [online]. Institut für Computational Perception, Linz, 2010. [cit. 4.11.2018]. Dostupné z URL: <http://www.cp.jku.at/people/seyerlehner/supervised/seyerlehner_phd.pdf>.
- [45] SCHEDL, M.: *Automatically extracting, analyzing, and visualizing information on music artists from the world wide web*. [online]. Dissertation, Johannes Kepler University Linz, 2008. [cit. 2.11.2018]. Dostupné z URL: <http://www.cp.jku.at/research/papers/schedl_phd_2008.pdf>.
- [46] SMOLA, A.; VISHWANATHAN, S., V., N.: *Introduction to Machine Learning*. [online]. Yahoo! Labs, Santa Clara; Departments of Statistics and Computer Science, Purdue University; College of Engineering and Computer Science, Australian National University, 2008. [cit. 19.10.2018]. Dostupné z URL: <<http://alex.smola.org/drafts/thebook.pdf>>.
- [47] SOLTAU, H.; SCHULTZ, T.; WESTPHAL, M.; WAIBEL, A.: *Recognition of music types*. [online]. University of Karlsruhe, 1998. [cit. 5.10.2018]. Dostupné z URL: <<http://isl.anthropomatik.kit.edu/pdf/Soltau1998a.pdf>>.
- [48] STURM, B., L.: *Classification Accuracy Is Not Enough*. [online]. Journal of Intelligent Information Systems, 41(3), 2013. [cit. 18.11.2018]. Dostupné z URL: <<http://vbn.aau.dk/files/70797941/Sturm20121030.pdf>>.

- [49] STURM, B., L.: *The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use*. [online]. CoRR, vol. 1306.1461, 2013. [cit. 28.10.2018]. Dostupné z URL: <<https://arxiv.org/pdf/1306.1461.pdf>>.
- [50] TSATSISHVILI, V.: *Automatic subgenre classification of heavy metal music*. [online]. University Of Jyväskylä, 2011. [cit. 6.10.2018]. Dostupné z URL: <<https://pdfs.semanticscholar.org/f5d6/27d830d706edb6f165cc02a1bd45c5a0e0fe.pdf>>.
- [51] TZANETAKIS, G.; ESSL, G.; COOK, P.: *Automatic musical genre classification of audio signals*. [online]. IEEE Transactions on Speech and Audio Processing, 2001. [cit. 5.10.2018]. ISSN 1063-6676. Dostupné z URL: <<http://ismir2001.ismir.net/pdf/tzanetakis.pdf>>.
- [52] VOGLER, B.; OTHMAN, A.: *Music Genre Recognition*. [online]. Bauhaus-Universität Weimar, Weimar, 2016. [cit. 6.10.2018]. Dostupné z URL: <<https://benediktsvogler.com/downloads/DocumentationAudiotechnikMusicGenreRecognition.pdf>>.
- [53] WOLD, E.; BLUM, T.; KEISLAR, D.; WHEATON, J.: *Content-based classification, search and retrieval of audio*. [online]. IEEE Multimedia Magazine, 3(3):27–36, 1996. [cit. 6.10.2018]. Dostupné z URL: <<http://musclefish.musclefish.com/ieeemm96>>.
- [54] YOUNG, S.; EVERMANN, G.; KERSHAW, D.; aj: *The HTK Book*. [online]. Cambridge University Engineering Department, 2002. [cit. 21.10.2018]. Dostupné z URL: <<http://www.dsic.upv.es/docs/posgrado/20/RES/materialesDocentes/alejandroViewgraphs/htkbook.pdf>>.

Seznam symbolů, veličin a zkratek

ANN	Artificial Neural Networks – Umělé neuronové sítě
CS	Kompresivní odběr vzorků – Compressive Sampling
ETM-NN	Explicitní modelování času s neuronovou sítí – Explicit Time Modelling with Neural Network
HMM	Skrytý Markovův model – Hidden Markov Model
GMM	Gaussův smíšený model – Gaussian Mixture Model
ISMIR	The International Society for Music Information Retrieval
k-NN	k nejbližších sousedů – k nearest neighbor
LDA	Lineární diskriminační analýza – Linear Discriminant Analysis
LSTER	Low Short-Time Energy Ratio
MFCC	Melovské keprstrální koeficienty – Mel-Frequency Cepstral Coefficients
MGR	Rozpoznávání hudebního žánru – Music Genre Recognition
MIR	Získávání informací z hudby – Music Information Retrieval
MPSD	Medián výkonové spektrální hustoty – Median of Power Spectral Density
mRMR	minimum Redundancy Maximum Relevance – minimální nadbytečnost, maximální relevance
NASE	Normalizovaná zvuková spektrální obálka – Normalized Audio Spectral Envelope
OSC	Oktávový spektrální kontrast – Octave-Based Spectral Contrast
RMS	Efektivní hodnota energie – Root mean square
SC	Spektrální těžiště – Spectral Centroid
SF	Spektrální tok – Spectral Flux
SFM	Míra spektrální rovinnosti – Spectral Flatness Measure
SVM	Support Vector Machines
TKEO	Teager-Kaiserův energetický operátor
ZCR	Počet průchodů nulovou hodnotou – Zero Crossing Rate

Seznam příloh

A	Obsah přiloženého CD	62
B	Výzkumy zabývající se rozpoznáváním žánru	63
C	Seznam vypočítaných parametrů	64
D	Individuální analýza parametrů	65
	D.1 ISMIR2004	65
	D.2 Unique	67
	D.3 Free Music Archive	70

A Obsah přiloženého CD

Přiložené CD obsahuje elektronickou verzi diplomové práce. Hlavní dokument nese název „Diplomova prace - Sarka Zemankova.pdf“. Pro výpočet parametrů v prostředí MATLAB slouží skript „Parameters.m“. Složka „mRMR“ obsahuje skripty a funkce pro výpočet metody *minimum Redundancy Maximum Relevance*. Poslední soubor se nazývá „Výsledky klasifikace.xlsx“ a obsahuje tabulky, udávající procentuální úspěšnost správného určení žánru pro rostoucí počet parametrů. Na třech kartách jsou zobrazeny tabulky pro jednotlivé datasety.

Obsah CD:

- Diplomova prace - Sarka Zemankova.pdf
- mRMR
- Parameters.m
- Vysledky klasifikace.xlsx

B Výzkumy zabývající se rozpoznáváním žánru

Tab. B.1: Výzkumy zabývající se rozpoznáváním žánru

Autor	Rok	Metoda	Databáze	Parametry	Úspěšnost
Soltau [47]	1998	HMM, ETM-NN	360 (4 žánry)	spektrální funkce	79%, 86%
Tzanetakis [51]	2001	Gaussovo rozdělení	750 (15 žánrů)	ZCR, lowenergy, spektrální funkce	62%
Dixon [12]	2003	autokorelace	taneční hudba	onset, periodicitá	80%
Li [28]	2003	SVM	409 (16 žánrů)	MFCC, energie, brightness	70 - 90%
Burred [7]	2004	k -NN, GMM	850 (17 žánrů)	ZCR, MFCC, energie, spektrální a rytmické funkce	59%
Lee [27]	2009	LDA	GTZAN, ISMIR2004	MFCC, OSC, NASE	91%
Chang [19]	2010	kompresivní odběr vzorků	GTZAN	ZCR, MFCC, OSC, spektrální funkce	82%
Marques [33]	2011	histogram + 5-NN, SVM, HMM	ISMIR2004, Latin music set	ZCR, spektrální funkce, 13 MFCC	75%, 75%, 83%
Tsatsishvili [50]	2011	k -NN, AdaBoost	metal set (17 subžánrů)	ZCR, MFCC, chromagram, spektrální funkce	45%, 45%
Mulder [36]	2014	k -NN, Mahalanobisova vzdálenost	metal set (17 subžánrů)	horizontální a vertikální intervalové funkce	70%
Vogler [52]	2016	Neural network	300 skladeb (3 žánry)	MFCC, OSC	83%
Pálmason [38]	2017	3-NN	GTZAN	MFCC, SFM	81%

C Seznam vypočítaných parametrů

Tab. C.1: Seznam vypočítaných parametrů

Název parametru	Jednotka	Popis parametru
RMS	[-]	efektivní hodnota energie
LowEnergy	[%]	procento rámců, které mají nižší než průměrnou energii
Flux	[Hz]	spektrální tok
TKEO	[-]	Teager-Kaiserův energetický operátor
MPSD	[-]	medián výkonové spektrální hustoty
Tempo	[BPM]	odhad tempa
EventDensity	[-]	hustota událostí
ZCR	[-]	počet průchodů nulovou hodnotou
Brightness	[%]	množství energie v oblasti nad 1500 Hz
Rolloff	[Hz]	frekvence, pod níž se nachází 85% energie signálu
Centroid	[Hz]	frekvenční těžiště
MFCC(13)	[-]	Melovské spektrální koeficienty (13 hodnot)

D Individuální analýza parametrů

D.1 ISMIR2004

Tab. D.1: Individuální analýza parametrů databáze IS-MIR2004

Procentuální úspěšnost správného zařazení						
Název parametru	classical	electronic	jazz	metal	rock	world
Flux(Q3)	0,00	19,13	0,00	26,67	35,64	16,26
MFCC(4)	70,03	62,61	3,85	0,00	0,00	0,00
ZCR(max)	88,64	0,00	0,00	0,00	16,83	15,45
ZCR(p1)	93,06	0,00	0,00	15,56	0,00	5,69
TKEO(max)	0,63	41,74	0,00	0,00	46,53	4,88
EventDensity(var)	23,34	60,87	0,00	6,67	0,00	0,81
TKEO(median)	0,00	41,74	0,00	11,11	52,48	18,70
MFCC(9)	15,46	0,00	11,54	8,89	0,00	70,73
Rolloff(mean)	74,13	1,74	0,00	62,22	19,80	16,26
RMS(Q1)	43,85	22,61	0,00	20,00	18,81	41,46
MFCC(3)	63,09	21,74	23,08	0,00	25,74	9,76
Flux(std)	55,52	44,35	30,77	0,00	0,00	16,26
EventDensity(p99)	88,64	4,35	0,00	13,33	56,44	3,25
MFCC(1)	93,38	28,70	0,00	0,00	2,97	5,69
MFCC(6)	7,89	13,91	0,00	0,00	72,28	0,00
TKEO(p1)	3,47	65,22	0,00	55,56	0,00	0,00
MFCC(12)	68,77	13,91	15,38	17,78	0,00	2,44
ZCR(std)	77,92	33,91	0,00	0,00	1,98	0,00
Flux(Q1)	64,04	13,04	0,00	33,33	17,82	47,15
MFCC(7)	56,47	2,61	19,23	0,00	1,98	6,50
EventDensity(median)	19,24	0,00	0,00	42,22	58,42	40,65
MFCC(10)	51,74	0,00	0,00	0,00	2,97	43,90
RMS(Q3)	70,98	30,43	0,00	55,56	0,00	35,77
ZCR(IQR)	94,32	8,70	0,00	0,00	36,63	0,81
EventDensity(IR)	4,10	40,87	0,00	0,00	25,74	52,85
TKEO(IQR)	54,57	46,09	23,08	64,44	0,00	0,00
Flux(max)	64,98	8,70	7,69	20,00	4,95	0,00
Flux(IQR)	60,57	36,52	0,00	0,00	0,00	15,45
Rolloff(median)	94,01	0,00	30,77	8,89	5,94	0,00

Název parametru	classical	electronic	jazz	metal	rock	world
Flux(median)	35,65	52,17	0,00	64,44	0,00	16,26
ZCR(IR)	61,83	21,74	0,00	0,00	26,73	16,26
RMS(p1)	79,18	0,87	11,54	51,11	0,00	22,76
Flux(mean)	57,10	52,17	0,00	24,44	27,72	0,00
ZCR(Q1)	91,80	19,13	0,00	46,67	21,78	0,00
RMS(mean)	72,56	30,43	0,00	55,56	3,96	2,44
Flux(var)	24,92	33,04	0,00	62,22	0,99	13,01
TKEO(Q3)	83,28	15,65	38,46	88,89	0,00	19,51
EventDensity(max)	88,96	28,70	0,00	11,11	0,00	23,58
Rolloff(max)	90,85	45,22	0,00	0,00	0,00	14,63
MFCC(2)	15,77	79,13	0,00	2,22	0,00	0,00
EventDensity(mean)	64,04	0,00	3,85	57,78	66,34	3,25
ZCR(var)	51,10	11,30	15,38	0,00	0,00	8,13
RMS(median)	81,07	26,09	42,22	0,00	0,00	7,32
Flux(IR)	21,45	53,91	6,67	0,00	34,65	6,50
EventDensity(std)	74,13	0,00	8,89	0,00	3,96	24,39
Centroid(mean)	0,00	43,48	82,22	15,38	5,94	0,00
RMS(IQR)	65,93	4,35	11,11	0,00	0,00	1,63
MFCC(5)	44,48	47,83	71,11	7,69	0,00	19,51
TKEO(mean)	71,61	3,48	0,00	0,00	59,41	18,70
Rolloff(p1)	66,56	7,83	8,89	0,00	2,97	0,00
Flux(p1)	45,11	36,52	40,00	0,00	25,74	0,00
Flux(p99)	49,53	29,57	0,00	38,46	12,87	0,00
MFCC(11)	21,45	43,48	11,11	0,00	25,74	1,63
TKEO(std)	0,63	60,87	2,22	46,15	0,00	6,50
ZCR(p99)	70,03	26,96	0,00	0,00	41,58	0,00
LowEnergy(mean)	1,26	41,74	48,89	30,77	3,96	0,81
MPSD	70,35	40,00	0,00	0,00	8,91	14,63
Rolloff(p99)	7,89	31,30	0,00	0,00	64,36	0,00
Tempo(mean)	71,61	3,48	0,00	0,00	2,97	17,89
MFCC(8)	0,00	0,00	0,00	100,00	0,00	0,00
RMS(IR)	44,48	54,78	22,22	15,38	0,99	0,00
MFCC(13)	29,97	31,30	0,00	0,00	42,57	5,69
Brightness(mean)	100,00	0,00	0,00	0,00	0,00	0,00
Centroid(median)	63,72	32,17	0,00	0,00	42,57	1,63
RMS(p99)	0,00	100,00	0,00	0,00	0,00	0,00

Název parametru	classical	electronic	jazz	metal	rock	world
TKEO(Q1)	0,00	0,00	0,00	100,00	0,00	0,00
Centroid(max)	0,32	23,48	33,33	19,23	33,66	0,00
TKEO(p99)	0,00	1,74	73,33	0,00	0,00	0,00
RMS(std)	100,00	0,00	6,67	0,00	0,00	0,00
EventDensity(p1)	100,00	0,00	0,00	0,00	0,00	0,00
Brightness(median)	100,00	0,00	0,00	0,00	0,00	0,00
TKEO(IR)	36,91	50,43	2,22	0,00	0,00	22,76
Centroid(p1)	50,79	17,39	8,89	0,00	0,00	27,64
LowEnergy(median)	0,00	0,00	0,00	100,00	0,00	0,00
ZCR(mean)	0,00	100,00	0,00	0,00	0,00	0,00
RMS(var)	0,00	0,00	100,00	0,00	0,00	0,00
Centroid(p99)	0,00	0,00	0,00	100,00	0,00	0,00
RMS(max)	62,46	13,91	2,22	0,00	6,93	9,76
Brightness(max)	0,00	100,00	0,00	0,00	0,00	0,00
Brightness(p1)	11,67	15,65	0,00	0,00	47,52	28,46
LowEnergy(max)	0,00	0,00	0,00	0,00	100,00	0,00
ZCR(Q3)	100,00	0,00	0,00	0,00	0,00	0,00
Brightness(p99)	95,90	22,61	0,00	0,00	6,93	0,00
Tempo(median)	36,28	6,96	0,00	0,00	57,43	12,20
ZCR(median)	0,00	1,74	0,00	96,15	0,00	0,00

D.2 Unique

Tab. D.2: Individuální analýza parametrů databáze Unique

Procentuální úspěšnost správného zařazení					
Název parametru	classical	electronic	jazz	rock	world
RMS(mean)	97,57	6,45	0,00	84,85	0,00
ZCR(Q1)	73,85	21,51	5,19	46,46	0,00
ZCR(std)	67,39	0,00	11,69	14,65	9,86
MFCC(2)	0,00	8,28	19,48	69,70	16,90
Flux(std)	93,26	20,22	0,00	87,37	4,23
RMS(p1)	3,23	0,00	64,42	0,51	22,54
EventDensity(p1)	94,88	30,54	0,00	23,35	0,00
TKEO(max)	0,00	11,83	24,16	69,70	5,63

Název parametru	classical	electronic	jazz	rock	world
TKEO(median)	89,49	20,43	0,00	53,23	0,00
MFCC(4)	93,26	0,00	48,70	36,36	14,08
MFCC(1)	67,65	0,00	14,29	84,85	0,00
RMS(IQR)	5,12	47,31	11,83	45,45	12,49
RMS(Q1)	94,61	26,84	0,00	59,09	0,00
Flux(max)	94,88	9,68	34,16	20,20	0,00
Flux(Q1)	83,53	0,00	19,09	66,36	8,31
EventDensity(IR)	73,32	39,78	42,21	25,25	2,82
Rolloff(mean)	97,57	25,81	0,00	54,55	26,76
TKEO(p1)	70,08	0,00	0,00	97,47	0,00
Flux(p99)	5,39	8,60	12,99	4,55	45,07
MFCC(9)	19,14	0,00	12,99	89,90	0,00
EventDensity(median)	63,34	0,00	5,19	21,21	42,25
RMS(IR)	1,35	0,00	53,90	95,45	0,00
TKEO(std)	39,89	3,23	39,61	0,00	36,62
MFCC(13)	70,89	23,66	10,39	59,09	0,00
ZCR(median)	82,75	21,51	47,40	9,60	8,45
RMS(Q3)	71,16	8,60	0,00	26,26	0,00
MFCC(3)	35,31	38,71	8,44	86,36	0,00
Flux(var)	92,99	0,00	0,00	75,76	1,41
TKEO(IQR)	56,33	0,00	0,00	11,62	0,00
TKEO(Q1)	73,85	43,01	1,30	83,84	0,00
Flux(IR)	46,36	15,05	44,81	75,76	0,00
Rolloff(median)	12,67	13,98	43,51	36,36	22,54
EventDensity(mean)	95,69	8,60	0,00	66,67	0,00
RMS(median)	97,30	0,00	0,00	27,78	0,00
ZCR(p1)	85,71	31,18	21,43	20,71	0,00
Flux(IQR)	21,56	11,83	0,00	93,43	0,00
TKEO(mean)	54,72	4,30	0,00	7,07	0,00
MFCC(7)	97,84	83,87	16,88	0,00	0,00
RMS(p99)	94,61	53,76	2,60	52,53	0,00
Flux(median)	60,92	22,58	37,01	65,66	0,00
TKEO(p99)	80,05	12,90	0,00	25,76	0,00
MFCC(6)	73,32	46,24	13,64	15,66	0,00
Rolloff(max)	98,92	63,44	0,00	35,35	0,00
RMS(std)	83,56	37,63	1,30	42,93	0,00

Název parametru	classical	electronic	jazz	rock	world
EventDensity(max)	54,18	43,01	0,00	59,60	1,41
TKEO(Q3)	34,23	0,00	3,90	0,00	61,97
MFCC(11)	85,18	54,84	16,88	6,57	0,00
RMS(max)	100,00	0,00	0,00	0,00	0,00
Flux(p1)	96,23	0,00	6,49	0,00	19,72
Zerocross(p99)	18,60	8,60	16,23	83,84	0,00
Flux(mean)	62,26	62,37	0,00	27,78	1,41
TKEO(IR)	46,90	75,27	11,04	0,00	0,00
RMS(var)	73,58	0,00	5,84	78,79	7,04
Rolloff(p1)	76,82	35,48	1,30	52,02	0,00
EventDensity(p99)	0,00	33,33	7,79	29,80	19,72
ZCR(mean)	28,03	0,00	69,48	3,03	11,27
MFCC(12)	18,87	16,13	38,96	77,78	14,08
Rolloff(p99)	49,06	0,00	0,00	58,59	0,00
ZCR(max)	70,08	4,30	12,99	43,94	18,31
MFCC(8)	74,12	25,81	8,44	2,53	8,45
LowEnergy(mean)	4,58	9,68	53,25	55,05	16,90
Centroid(mean)	0,00	3,23	0,00	97,98	2,82
EventDensity(std)	63,88	16,13	37,01	0,00	12,68
ZCR(var)	39,35	2,15	57,14	50,51	0,00
Brightness(mean)	34,77	0,00	4,55	77,78	0,00
MFCC(5)	26,95	3,23	52,60	0,00	2,82
Centroid(median)	0,00	0,00	0,00	98,99	0,00
EventDensity(var)	69,00	4,30	24,03	4,55	9,86
Zerocross(Q3)	62,26	36,56	16,23	27,27	9,86
Centroid(max)	4,31	17,20	0,00	39,90	8,45
TKEO(var)	6,74	45,16	40,91	15,66	12,68
Brightness(median)	57,41	0,00	22,08	21,72	2,82
LowEnergy(median)	70,35	55,91	5,19	0,51	5,63
Centroid(p1)	14,02	8,60	0,00	76,26	0,00
ZCR(IQR)	74,12	1,08	10,39	16,67	0,00
ZCR(IR)	0,00	2,15	57,14	62,63	7,04
Centroid(p99)	61,19	0,00	21,43	29,29	11,27
Brightness(max)	74,12	2,15	9,74	27,78	2,82
LowEnergy(max)	0,00	48,39	24,03	22,22	38,03
Brightness(p1)	30,46	27,96	40,91	38,38	5,63

Název parametru	classical	electronic	jazz	rock	world
Brightness(p99)	0,00	25,81	0,00	34,85	76,06
MPSD	71,70	25,81	0,00	10,10	14,08

D.3 Free Music Archive

Tab. D.3: Individuální analýza parametrů databáze Free Music Archive

Procentuální úspěšnost správného zařazení						
Název parametru	classical	electronic	jazz	metal	rock	world
Flux(median)	43,00	43,00	3,00	18,00	6,00	51,00
ZCR(std)	19,00	69,00	0,00	0,00	14,00	0,00
ZCR(p1)	100,00	0,00	0,00	15,00	0,00	18,00
MFCC(3)	0,00	75,00	0,00	25,00	9,00	0,00
MFCC(11)	4,00	3,00	0,00	0,00	88,00	27,00
MFCC(1)	75,00	69,00	0,00	0,00	12,00	0,00
EventDensity(max)	62,00	32,00	0,00	0,00	6,00	42,00
RMS(p1)	9,00	20,00	0,00	30,00	87,00	1,00
Rolloff(mean)	68,00	0,00	62,00	0,00	45,00	12,00
RMS(IQR)	34,00	15,00	16,00	15,00	1,00	6,00
ZCR(Q1)	97,00	0,00	15,00	12,00	38,00	0,00
MFCC(6)	0,00	17,00	0,00	12,00	0,00	52,00
TKEO(max)	0,00	1,00	16,00	12,00	62,00	15,00
MFCC(8)	3,00	77,00	0,00	5,00	2,00	0,00
RMS(median)	51,00	26,00	0,00	10,00	6,00	10,00
MFCC(2)	19,00	0,00	0,00	3,00	10,00	42,00
TKEO(median)	0,00	0,00	0,00	31,00	56,00	0,00
ZCR(IQR)	0,00	76,00	7,00	5,00	4,00	0,00
EventDensity(mean)	52,00	10,00	0,00	9,00	14,00	60,00
Flux(p1)	0,00	54,00	3,00	27,00	0,00	38,00
ZCR(Q3)	20,00	41,00	0,00	0,00	36,00	16,00
Flux(std)	18,00	64,00	0,00	0,00	15,00	4,00
TKEO(mean)	0,00	63,00	8,00	15,00	33,00	0,00
MFCC(12)	6,00	14,00	0,00	0,00	62,00	36,00
Rolloff(median)	68,00	0,00	63,00	0,00	45,00	12,00
LowEnergy(mean)	0,00	7,00	0,00	0,00	83,00	1,00
Flux(Q1)	0,00	0,00	0,00	44,00	0,00	91,00

Název parametru	classical	electronic	jazz	metal	rock	world
MFCC(4)	3,00	6,00	4,00	78,00	8,00	7,00
ZCR(median)	42,00	21,00	2,00	32,00	10,00	33,00
EventDensity(median)	80,00	0,00	0,00	5,00	78,00	27,00
TKEO(p1)	92,00	47,00	0,00	35,00	1,00	0,00
ZCR(p99)	10,00	25,00	0,00	0,00	36,00	4,00
Flux(Q3)	84,00	54,00	0,00	24,00	4,00	28,00
Brightness(mean)	81,00	0,00	50,00	13,00	1,00	7,00
MFCC(7)	51,00	0,00	3,00	0,00	16,00	0,00
EventDensity(p99)	62,00	32,00	0,00	0,00	6,00	42,00
Rolloff(max)	68,00	0,00	62,00	0,00	45,00	12,00
RMS(Q1)	54,00	27,00	12,00	34,00	0,00	0,00
ZCR(mean)	22,00	17,00	12,00	3,00	55,00	10,00
RMS(var)	61,00	10,00	20,00	3,00	0,00	12,00
TKEO(Q3)	23,00	45,00	14,00	0,00	0,00	11,00
ZCR(var)	76,00	0,00	0,00	0,00	45,00	51,00
EventDensity(p1)	61,00	0,00	7,00	16,00	6,00	0,00
MFCC(9)	0,00	1,00	0,00	96,00	1,00	0,00
EventDensity(var)	48,00	7,00	4,00	0,00	62,00	0,00
Flux(mean)	68,00	0,00	63,00	0,00	45,00	12,00
Rolloff(p1)	20,00	0,00	0,00	0,00	18,00	77,00
Flux(IR)	37,00	59,00	0,00	0,00	13,00	0,00
Zerocross(IR)	81,00	0,00	50,00	13,00	1,00	8,00
Brightness(median)	0,00	0,00	28,00	81,00	2,00	0,00
TKEO(Q1)	0,00	0,00	0,00	0,00	50,00	42,00
TKEO(IQR)	68,00	0,00	63,00	0,00	45,00	12,00
Rolloff(p99)	3,00	46,00	0,00	25,00	7,00	15,00
RMS(p99)	0,00	75,00	0,00	0,00	17,00	0,00
ZCR(max)	0,00	7,00	0,00	0,00	83,00	1,00
LowEnergy(median)	5,00	7,00	10,00	6,00	32,00	0,00
TKEO(std)	81,00	0,00	50,00	13,00	1,00	7,00
Brightness(max)	12,00	75,00	0,00	22,00	4,00	0,00
MFCC(10)	60,00	0,00	25,00	29,00	19,00	1,00
RMS(mean)	0,00	87,00	0,00	0,00	23,00	6,00
MFCC(5)	30,00	32,00	0,00	14,00	68,00	6,00
Flux(var)	1,00	37,00	7,00	5,00	38,00	1,00
MPSD	81,00	0,00	50,00	13,00	1,00	7,00

Název parametru	classical	electronic	jazz	metal	rock	world
Brightness(p1)	11,00	13,00	0,00	0,00	83,00	22,00
RMS(Q3)	1,00	70,00	0,00	0,00	27,00	5,00
Tempo(mean)	81,00	0,00	50,00	13,00	1,00	7,00
Brightness(p99)	0,00	14,00	0,00	2,00	68,00	27,00
TKEO(p99)	0,00	7,00	0,00	0,00	83,00	1,00
LowEnergy(max)	82,00	4,00	16,00	0,00	8,00	0,00
RMS(std)	0,00	0,00	44,00	20,00	71,00	0,00
TKEO(IR)	39,00	61,00	8,00	0,00	0,00	1,00
Flux(IQR)	0,00	100,00	0,00	0,00	0,00	0,00
EventDensity(std)	4,00	0,00	10,00	0,00	19,00	74,00
Flux(p99)	0,00	7,00	0,00	0,00	83,00	1,00
LowEnergy(p1)	96,00	0,00	16,00	0,00	18,00	3,00
TKEO(var)	1,00	19,00	18,00	31,00	15,00	27,00
RMS(IR)	3,00	81,00	0,00	0,00	4,00	37,00
MFCC(13)	54,00	17,00	0,00	9,00	9,00	12,00
Flux(max)	0,00	7,00	0,00	0,00	83,00	1,00
LowEnergy(p99)	1,00	70,00	0,00	0,00	27,00	5,00
Tempo(median)	59,00	7,00	12,00	0,00	30,00	0,00
RMS(max)	0,00	99,00	0,00	3,00	0,00	0,00
EventDensity(IR)	1,00	61,00	0,00	0,00	37,00	5,00
Tempo(max)	1,00	70,00	0,00	0,00	27,00	5,00
Tempo(p1)	1,00	70,00	0,00	0,00	27,00	5,00