



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

**POROVNÁVÁNÍ ZVUKOVÝCH NAHRÁVEK ZA POMOCI
PARAMETRŮ POPISUJÍCÍ BARVU ZVUKU**

SOUND RECORDS COMPARISON USING TIMBRE FEATURES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Štěpán Miklánek

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Tomáš Kiska

BRNO 2017

Bakalářská práce

bakalářský studijní obor **Audio inženýrství**
Ústav telekomunikací

Student: Štěpán Miklánek

ID: 174465

Ročník: 3

Akademický rok: 2016/17

NÁZEV TÉMATU:

Porovnávání zvukových nahrávek za pomoci parametrů popisující barvu zvuku

POKYNY PRO VYPRACOVÁNÍ:

V rámci této práce budou shrnuty dosavadní poznatky z oblasti srovnávání zvukových nahrávek. Bude navržen a implementován vyhodnocovací systém v prostředí MATLAB. Dále budou analyzovány interpretace z hlediska barvy zvuku. Rovněž budou za pomoci statistické analýzy vybrány ty parametry, které z hlediska barvy zvuku mají největší schopnost diferencovat jednotlivé interpretace.

DOPORUČENÁ LITERATURA:

[1] MAEZAWA, A.; OKUNO, H. G. Bayesian Audio-to-Score Alignment Based on Joint Inference of Timbre, Volume, Tempo, and Note Onset Timings. *Computer Music Journal*, roč. 39, č. 1, 2015: s. 74–87.

[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: 1.2.2017

Termín odevzdání: 8.6.2017

Vedoucí práce: Ing. Tomáš Kiska

Konzultant:

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

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské 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 práce se zabývá výzkumem parametrů popisující nahrávky na základě barvy zvuku. Nejprve je popsán historický vývoj a novodobý přístup v oblasti Music Information Retrieval (MIR), poté je popsán postup při zpracovávání hudebního signálu a daná problematika je nastíněna jak z pohledu hudební teorie, tak z pohledu číslicového zpracování signálu. Následuje popis předzpracování signálu, tato část je důležitá z hlediska parametrizace hudebního signálu. V kapitole parametrizace jsou shrnuty poznatky o parametrech, které jsou běžně využívány při získávání informací z nahrávek, přičemž je kladen důraz zejména na parametry týkající se barvy zvuku. Je také představena databáze nahrávek k analýze a návrh vyhodnocovacího systému, který bude nahrávky analyzovat. Nakonec je představena individuální analýza parametrů, kterými jsou popisovány nahrávky na základě barvy zvuku.

KLÍČOVÁ SLOVA

barva zvuku, parametry hudební nahrávky, statistická analýza, klasifikace, získávání informací z hudby, podobnost

ABSTRACT

This thesis deals with research of musical features, which are describing music recordings relating to timbre. First chapter deals with historical development and modern approach in a discipline called Music Information Retrieval (MIR), further there is a description of music processing from the perspective of music theory and digital signal processing. Then followed by a description of signal pre-processing. This part is very important when retrieving features from music recordings. In chapter concerned about retrieving features there are summarized all common features used when retrieving information from musical recordings with main concern to timbral features. A database of music recordings and a feature retrieving system is introduced. The last chapter deals with individual analysis of timbral features.

KEYWORDS

timbre, musical features, statistical analysis, classification, music information retrieval, similarity

MIKLÁNEK, Štěpán *Porovnávání zvukových nahrávek za pomoci parametrů popisující barvu zvuku*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2017. 72 s. Vedoucí práce byl Ing. Tomáš Kiska

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Porovnávání zvukových nahrávek za pomoci parametrů popisující barvu zvuku“ jsem vypracoval(a) samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

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

Brno

.....

podpis autora(-ky)

PODĚKOVÁNÍ

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

Brno

.....

podpis autora(-ky)



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

Brno

.....

podpis autora(-ky)



EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI



OBSAH

1	Úvod	12
2	Zpracování hudebního signálu	13
2.1	Historický vývoj	13
2.2	Novodobý přístup	13
3	Srovnání z pohledu hudební teorie a číslicového zpracování signálu	15
3.1	Intenzita a hlasitost	15
3.2	Tónový obsah	15
3.3	Časové hledisko, rytmus a tempo	17
3.3.1	Počátek zvukové události	17
3.3.2	Tempo	17
3.3.3	Časové hledisko v hudební teorii	17
4	Předzpracování hudebního signálu	19
4.1	Převod z analogové do číslicové oblasti	19
4.2	Smíchávání kanálů	19
4.3	Odstranění stejnosměrné složky	20
4.4	Normalizace	20
4.5	Další možnosti předzpracování signálu	21
5	Parametrizace hudebního signálu	22
5.1	Nízkoúrovňové parametry	22
5.1.1	Časová oblast	23
5.1.2	Frekvenční oblast	24
5.2	Rytmické parametry	26
5.2.1	Fluktuace	26
5.2.2	Spektrum úderů	26
5.2.3	Počátek zvukové události	27
5.2.4	Hustota	28
5.2.5	Tempo	28
5.2.6	Metrum	28
5.2.7	Metrický centroid a metrická síla	28
5.2.8	Čitelnost pulzu	29
5.3	Dynamické parametry	29
5.3.1	Efektivní hodnota signálu	29
5.3.2	Segmentace	29
5.3.3	Nízká energie	29

5.4	Parametry popisující barvu zvuku	30
5.4.1	Doba náběhu	30
5.4.2	Sklon náběhu	30
5.4.3	Intenzita náběhu	30
5.4.4	Sklon poklesu	30
5.4.5	Doba trvání	31
5.4.6	Počet průchodů nulovou úrovní	31
5.4.7	Pokles spektrální energie	31
5.4.8	Světlost spektra	31
5.4.9	Melovské keprální koeficienty (MFCC)	31
5.4.10	Drsnost	33
5.4.11	Parametry spektrálního rozložení	33
6	Výsledky dosažené v zahraničí	34
7	Synchronizace interpretací	35
7.1	Parametry popisující tónový obsah	35
7.1.1	Reprezentace výšky tónu	36
7.1.2	Problematika ladění	36
7.1.3	Parametr Chroma Pitch	37
7.1.4	Normalizace chroma parametru	37
7.1.5	Parametr Chroma-Log-Pitch	37
7.1.6	Parametr CENS	37
7.2	Dynamické borcení časové osy	38
7.2.1	DTW Algoritmus Itakura	38
7.3	Postup při výběru notového úseku	39
8	Návrh vyhodnocovacího systému	42
9	Evaluace systému	43
9.1	Klasifikační stromy	43
9.2	Křížová validace	43
9.3	Vyhodnocení algoritmů	43
9.3.1	Přesnost	44
9.3.2	Spearmanův koeficient pořadové korelace	44
9.3.3	Vzájemná informace	44
9.3.4	Mann-Whitney U test	45

10 Testování a vyhodnocení	46
10.1 Databáze nahrávek k analýze	46
10.2 Vypočítané parametry	46
10.3 Individuální analýza parametrů	47
11 Závěr	51
Literatura	53
Seznam symbolů, veličin a zkratk	58
Seznam příloh	60
A Obsah přiloženého DVD	61
B Partitura notového úseku	62
C Databáze nahrávek k analýze	63
D Seznam vypočítaných parametrů	65

SEZNAM OBRÁZKŮ

5.1	Ukázka amplitudové obálky hudebního signálu.	23
5.2	Ukázka efektivní energie hudebního signálu.	23
5.3	Počet průchodů nulovou úrovní.	24
5.4	Ukázka spektrálního centroidu hudebního signálu.	25
5.5	Ukázka spektrální proměnlivosti hudebního signálu	26
5.6	Blokové schéma výpočtu křivky počátků zvukových událostí	27
5.7	Blokové schéma výpočtu Melovských keprálních koeficientů	32
7.1	Postup synchronizace nahrávek	40
8.1	Návrh vyhodnocovacího systému	42
9.1	Grafické znázornění vzájemné informace.	45

SEZNAM TABULEK

3.1	Tónové třídy	16
3.2	Diatonická stupnice	16
5.1	Přehled nízkoúrovňových parametrů	22
5.2	Termíny používané při definicích nízkoúrovňových parametrů	22
10.1	Přehled vysokoúrovňových parametrů	47
10.2	Výsledky individuální analýzy Melovských kepstrálních koeficientů.	48
10.3	Výsledky individuální analýzy parametrů popisujících rozložení spektrální energie.	48
10.4	Výsledky individuální analýzy parametrů popisujících spektrální rozložení.	49
10.5	Výsledky individuální analýzy parametrů popisujících křivku počátků zvukových událostí.	49
10.6	Výsledky individuální analýzy parametrů drsnost a počet průchodů nulovou úrovní.	50
C.1	Databáze nahrávek k analýze	63
D.1	Seznam vypočítaných parametrů	65

1 ÚVOD

Tato práce je věnována zejména oblasti s názvem Music Information Retrieval (MIR), v češtině by se dal tento název chápat jako získávání informací z hudby [12]. Jedná se o interdisciplinární oblast výzkumu zahrnující počítačovou vědu, získávání informací, muzikologii, hudební teorii, audio inženýrství, číslicové zpracování signálu a kognitivní vědu. Jejím cílem je vyvinout postupy jak spravovat databáze hudebního materiálu pro zachování, přístup, výzkum a další využití.

V oblasti výzkumu barvy zvuku je kladen důraz zejména na klasifikaci jednotlivých hudebních nástrojů, kde jednotlivé výpočetní systémy reprezentují audio signál pomocí dané sady parametrů a je využíváno této reprezentace pro přiřazení kategorií ke zvukům, které jsou vydávány hudebními nástroji [34].

Při získávání informací z nahrávek je využíváno všech druhů a žánrů hudby a jsou popisovány všechny různé kvality jako například rytmus, barva zvuku, melodie nebo harmonie. Při výpočetních postupech je typicky využíváno parametrů a jsou vytvářeny modely pro popis hudby jednou nebo více kategoriemi vnímání hudby. Hudebním obsahem jsou myšleny aspekty, které jsou zakódovány v audio signálu, kdežto hudební kontext zahrnuje faktory, které nemohou být extrahovány přímo z audio nahrávky, ale i přesto souvisejí s hudbou nebo umělcem [32].

Parametry týkající se rytmické struktury, melodie nebo barvy zvuku patří do kategorie hudebního obsahu, kdežto informace o daném umělci nebo žánrové zařazení patří do kategorie hudebního kontextu.

V ideálním případě by měly postupy získávání informací z hudby překonávat „sémantickou mezeru“, kterou je myšlena neshoda mezi strojově získávanými hudebními parametry a popisy, které mají význam pro vnímání hudby posluchačem.

2 ZPRACOVÁNÍ HUDEBNÍHO SIGNÁLU

2.1 Historický vývoj

V této oblasti došlo k významnému pokroku na konci 90. let 20. století, kdy tato oblast postupně získávala svoji nynější podobu. Ačkoli se MIR technologie zabývají širokým spektrem vlastností hudby, nejsou dodnes až tak široce využívány [19].

Vědecký pohled na hudbu je také součástí důležité tradice počítačové muzikologie, která vznikla již v počátcích výpočetních technologií. To je také doba, kdy se poprvé objevil termín MIR, tedy kolem roku 1965. Po ustálení podoby této disciplíny na konci 90. let, hraje od roku 2000 klíčovou roli každoroční konference pořádaná organizací The International Society for Music Information Retrieval (ISMIR), která poskytuje platformu pro setkávání představitelů zabývajících se touto problematikou [12]. Počínajíc rokem 2005 se také každoročně koná kampaň s názvem the Music Information Retrieval Evaluation eXchange (MIREX), kde se vyhodnocují algoritmy zabývající se získáváním informací z hudebního obsahu.

2.2 Novodobý přístup

Hlavním úkolem v obsahově založeném MIR je extrahovat hudební vlastnosti a rysy, které jsou zakódovány v digitálním hudebním souboru. Termín hudební vlastnost se může vztahovat k číselnému vyjádření (skalár, vektor nebo matice), anebo k výsledku procesu extrakce. Extrakce dat spočívá v přetvoření surových dat reprezentujících hudební objekt ve více popisnou reprezentaci [19]. Nejlépe popisující hudební aspekty tak, jak je vnímají posluchači. Tedy například instrumentace, rytmická struktura, melodie nebo harmonie.

Výpočetní vlastnosti, které míří na modelování aspektů lidského vnímání hudby, mohou být roztrženy do několika úrovní. Například úroveň abstrakce, čas působnosti a signálovou doménu ve které jsou počítány.

Úroveň abstrakce hudebního parametru je typicky uváděna ve třech úrovních: nízká úroveň (např. obálka amplitudy, energie, těžiště spektra, spektrální proměnlivost, počet průchodů nulovou úrovní), střední úroveň (např. parametry vztahující se k výšce tónu a rytmu - počátky not, fluktuace tempa) a vysoká úroveň (např. instrumentace, tónina, akordy, melodie, rytmus, tempo, text, žánr nebo nálada). Zatímco nízkourovňové parametry jsou typicky vypočítávány přímo ze surové zvukové křivky a jsou to často jednoduchá statistická shrnutí, parametry střední úrovně mají již více hudebního významu. Například parametry vztahující se k notám a rytmu

a využívají i vlastností psychoakustického modelu. Vysokoúrovňovými parametry jsou popisovány vlastnosti hudby, tak jak ji vnímají běžní posluchači.

Z pohledu času působnosti jsou parametry rozlišeny na okamžité, úsekové a globální. Okamžité parametry jsou vypočítávány pro určitý bod v čase. Pokud je brána v potaz rozlišovací schopnost lidského ucha tj. přibližně 10 ms, pokrývají tyto parametry typicky časovou oblast několika desítek milisekund. Úsekové parametry jsou počítány pomocí časového okna o dané velikosti (např. 5 s). Může se také jednat o významovou část hudební nahrávky, například hudební fráze nebo refrén skladby. Globálními parametry je popisován celý hudební objekt zájmu, kupříkladu skladba, věta nebo úryvek audio záznamu.

Další kategorizace hudebních parametrů je prováděna na základě oblasti ve které jsou vypočítávány. V základu může být audio signál reprezentován v časové nebo frekvenční oblasti. Reprezentace v časové oblasti indikuje amplitudu signálu pro každý bod v čase, přesněji pro každý bod, kdy byl zachycen vzorek. Frekvenční oblast signálu je typicky výsledek Fourierovy transformace, která nám umožňuje rozdělit daný signál na jeho jednotlivé harmonické složky. Tímto procesem je popsán signál dle jeho amplitudy (modulu) na jednotlivých frekvencích.

Proces výpočtu parametru z hudebního audio signálu může být popsán tak, že vstupem tohoto systému je zvukový generátor, například libovolný hudební nástroj. Nejdříve je provedeno nahrání a konverze analogového signálu, vyprodukovaného hudebním nástrojem, do digitální reprezentace, kterou je možno snadno zpracovávat pomocí počítače. Jednotlivé vzorky vzniklé převodem z analogové oblasti jsou reprezentace hodnoty amplitudy signálu v čase.

3 SROVNÁNÍ Z POHLEDU HUDEBNÍ TEORIE A ČÍSLICOVÉHO ZPRACOVÁNÍ SIGNÁLU

3.1 Intenzita a hlasitost

Intenzitou je označována fyzikální, měřitelná veličina, například amplituda zvuku. Hlasitostí je označována veličina, kterou je možné měřit jedině na základě vnímání lidského ucha posluchače. Lidské vnímání intenzity je spojeno s amplitudou audio signálu tak, že pokud zvětšíme amplitudu, vnímaná hlasitost se zvětší. Tento vztah je nelineární. Lineární nárůst amplitudy signálu neznamena lineární nárůst vnímané hlasitosti. Je zaváděna jednotka *decibel* (dB), která je vypočítána jako logaritmus intenzity parametru $v(n)$ (vypočítáno z několika vzorků)

$$v_{\text{dB}}(n) = 20 \cdot \log_{10} \left(\frac{v(n)}{v_0} \right), \quad (3.1)$$

kde v_0 reprezentuje referenční hodnotu. V číslicové oblasti je počítáno s amplitudami v intervalu $[-1;1]$, referenční hodnota $v_0 = 1$. Výsledná jednotka úrovně je označována jako dBFS (dB *full scale*).[17]

V tradičním notovém zápisu jsou instrukce o dynamice značně neurčité. Typicky je používáno pět až osm úrovní, kterými je popisována hudební dynamika (**pp**: *pianissimo*, **p**: *piano*, **mf**: *mezzoforte*, **f**: *forte*, **ff**: *fortissimo* pro úroveň dynamiky od „velmi slabé“ po „velmi silnou“). Tyto úrovně jsou doprovázeny značením o plynulé změně hlasitosti (například *crescendo* a *decrescendo* pro zvýšení a snížení hlasitosti) a dynamické akcenty (například **sf**: *sforzando*). Těmito instrukcemi není přímo hlasitost určena, protože je kladen důraz také na faktory jako instrumentace, barva zvuku, počet hlasů, provedení a hudební kontext. Měřením intenzity nebo hlasitosti tedy může být do jisté míry zjištěna informace o hudební dynamice.

3.2 Tónový obsah

Lidské vnímání výšky tónu je přímo je přímo související s frekvencí signálu. Čím vyšší je frekvence, tím se zdá být i vnímaný tón vyšší. Jestliže je signál složen z více harmonických složek s frekvencemi $f_0, 2f_0, 3f_0, \dots$, potom *základní harmonická složka - fundament* f_0 převažuje při vnímání výšky tónu.

Vztah mezi základní frekvencí a vnímanou výškou tónu není lineární. Když jsou položeny stejně vzdálené tóny z hlediska výšky tónu v nízkých a vysokých frekvencích, nebude jejich frekvenční rozdíl stejný. Tato nelinearita je způsobena frekvenč-

ním rozlišením lidského ucha, zejména hlemýždě. Existují různé způsoby jak měřit frekvenční mapu hlemýždě. Jeden z běžných modelů je *mel stupnice* [17].

Termín *mel* byl poprvé představen v roce 1937 jako název subjektivní jednotky výšky tónu. Jsou vypočítávány Melovské keprální koeficienty (MFCC), které našly svůj původ v zpracovávání řeči. Jsou získávány výpočtem logaritmu energie krátkodobé fourierovy transformace před aplikováním diskretní kosinové transformace, což má za následek reprezentaci spektrálního tvaru. Koeficienty nižšího řádu je reprezentována hrubá spektrální proměnlivost, zatímco koeficienty vyššího řádu je reprezentován jemnější detail spektra [34].

Běžně je používán Fantův

$$m_F(f) = 1000 \cdot \log_2 \left(1 + \frac{f}{1000 \text{ Hz}} \right) \quad (3.2)$$

a O'Shaughnessyho model

$$m_S(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700 \text{ Hz}} \right). \quad (3.3)$$

Pozdější model je používán ve formě

$$m_S(f) = 1127 \cdot \log \left(1 + \frac{f}{700 \text{ Hz}} \right). \quad (3.4)$$

Z hlediska hudební teorie jsou vlastnosti spojené s výškou tónu dále rozváděny i ve spojitosti vzájemné interakce ve formě akordů, melodie, harmonie a tónin.

Tóny tvoří oktávy, kde každá oktáva má počáteční a konečný tón. Frekvenční poměr těchto tónů je 1:2. V jedné oktávě je obsaženo celkem 12 stejně velkých částí - tónů.

Tab. 3.1: Tónové třídy

0	1	2	3	4	5	6	7	8	9	10	11
<i>C</i>	<i>C♯/D♭</i>	<i>D</i>	<i>D♯/E♭</i>	<i>E</i>	<i>F</i>	<i>F♯/G♭</i>	<i>G</i>	<i>G♯/A♭</i>	<i>A</i>	<i>A♯/B♭</i>	<i>B</i>

Sedmi z těchto tónových tříd je tvořena takzvaná diatonická stupnice. V tomto případě stupnice C dur.

Tab. 3.2: Diatonická stupnice

0	2	4	5	7	9	11
<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>A</i>	<i>B</i>

Na základě tabulky je možno pozorovat, že vzdálenost mezi dvěma výškami tónů je ve většině případů rovna dvěma půltónům, ve dvou případech je tato vzdálenost jen půltón. Jakákoliv výška tónu mezi těmito diatonickými stupni může být vytvořena zvýšením nebo snížením jednotlivých stupňů. Z hlediska výšky mohou být například tóny $C\sharp$ a $D\flat$ považovány za stejné - platí enharmonická záměna.

3.3 Časové hledisko, rytmus a tempo

3.3.1 Počátek zvukové události

Je zaváděn pojem *onset*, což je počátek zvukové události [17]. V příkladě reálné nahrávky, začátek hudebního zvuku není určen jako přesný bod v čase, ale je popsán časovým průběhem - *attack time* neboli *rise time*. Tento časový úsek může být definován jako doba od první oscilace vytvořené hudebním nástrojem do dosažení maximální amplitudy. Doba *attack time* se značně liší s druhem hudebního nástroje a může se pohybovat v rozmezí 5 ms u perkusivních nástrojů až do 200 ms u nástrojů dechových. Dobu *onset* nelze jednoduše popsat a jsou zaváděny jednotlivé definice [30].

1. *Note Onset Time* - bod v čase, kdy je hudebním nástrojem vytvořen zvuk.
2. *Acoustic Onset Time* - první bod v čase, kdy je signál nebo zvuk měřitelný.
3. *Perceptual Onset Time* - bod v čase kdy je poprvé zvuk vnímán posluchačem.

3.3.2 Tempo

Tempo je definováno jako míra vnímaných pulsů se stejnou dobou trvání [9]. Typická hodnota přirozeného tempa je uváděna jako 100 Beats per Minute (BPM) - úderů za minutu [17].

Pro hudební úseky s konstantním tempem může být tempo T v BPM pro časový úsek δt_s a počet úderů B vypočítáno jako

$$T = \frac{B \cdot 60}{\delta t_s} [\text{BPM}]. \quad (3.5)$$

3.3.3 Časové hledisko v hudební teorii

Celkové tempo hudebního díla je většinou voleno umělci, přestože skladatel uvádí preferované tempo [17]. Mnoho hudebních děl z období baroka není doprovázeno instrukcemi o tempu. V pozdějších obdobích se začíná používat značení tempa italskými výrazy jako *Largo* (velmi pomalu), *Adagio* (pomalu), *Andante* (volně), *Moderato* (mírně rychle), *Allegro* (rychle) a *Presto* (velmi rychle). V posledních staletích se stalo běžnějším specifikovat tempo v jednotkách BPM.

Lokální tempo se mění v průběhu hudebního díla téměř ve všech žánrech. Tyto instrukce jsou v notovém zápisu zapisovány jako *ritardando* (zpomalování) a *accelerando* (zrychlování).

Notový zápis je rozdělen na začátku každého taktu vertikální čarou. První doba taktu je vnímána nejzřetelněji a je označována jako těžká.

4 PŘEDZPRACOVÁNÍ HUDEBNÍHO SIGNÁLU

4.1 Převod z analogové do číslicové oblasti

Původně spojitý analogový signál je potřeba vzorkovat ve stejně velkých časových intervalech a také kvantovat, tzn. kódovat s určitým počtem bitů. Tento proces je nazýván jako analogově-digitální konverze. Vzorkováním je popisováno měření amplitudy signálu, typicky použitím mikrofону. Získaná hodnota je stále analogová. Pro další zpracování musí být daný signál kvantován, znamená to, že je signál transformován do digitální reprezentace s použitím konečného počtu hodnot. Tento proces je obvykle prováděn zaokrouhlením na určitou jednotku přesnosti. Například při použití 4 bitů pro kódování signálu je dosaženo 16 diskrétních úrovní na které se původně spojitý signál může namapovat. Počtem bitů je udávána bitová hloubka analogově-digitální konverze. Tímto převodem je přivedena nežádoucí chyba kvantizační zkreslení, jedná se o rozdíl mezi hodnotami spojitého signálu a diskrétní hodnotou, ke které byla hodnota spojitého signálu přiřazena. Počet bitů použitý pro kódování jednotlivých vzorků je důležitý pro výsledek konverze. Běžná nahrávka na kompaktním disku je kódována s bitovou hloubkou 16 bitů na jeden audio kanál. Kromě bitové hloubky je důležitým parametrem také vzorkovací frekvence [12]. Podle Nyquistova-Shannonova vzorkovacího teorému je vzorkovací frekvence udávána jako dvojnásobek nejvyšší frekvenční složky daného signálu. Při volbě příliš nízké vzorkovací frekvence není možné zachytit celé frekvenční spektrum tak, jak je zaznamenáváno lidským uchem (přibližně 20 Hz až 20 kHz) [17].

Surová audio data jsou často předzpracovávána než je proveden výpočet okamžitých parametrů. Důvodem procesu předzpracování je redukovat objem audio dat, které budou analyzovány. To je prováděno vynecháním nedůležitých informací, aby bylo zajištěno minimalizování dopadu nechtěných informací na vypočítávané parametry. Standardní postupy se liší zejména v oblasti ve kterých navrhovaný algoritmus pracuje. Tyto oblasti jsou rozděleny do dvou kategorií: reálný čas - známe pouze nynější a minulé vzorky a offline - známe i všechny nadcházející vzorky audio souboru.

4.2 Smíchávání kanálů

Ve většině analyzovaných problémů mohou být informace reprezentovány jak jedním tak i několika vstupními kanály. Pro příklad, tempo nebo informace o hudebním žánru může být zjištěna jak z mono, tak i ze stereo nebo vícekanálové nahrávky.

Smíchávání kanálu je prováděno jednoduchým výpočtem aritmetického průměru všech vstupních kanálů. Aritmetický průměr je definován všemi vstupními kanály $x_c(i)$ ze vzorků i . Počet vstupních kanálů je C

$$x(i) = \frac{1}{C} \sum_{c=0}^{C-1} x_c(i). \quad (4.1)$$

Je také možné kanály smíchat s různou vahou. Surroundové kanály mohou mít například menší váhu, než levý a pravý kanál. Alternativou ke smíchávání kanálů je použití pouze jednoho kanálu. To ale může mít za následek ztrátu informace, jestliže byly kanály míchány za účelem, aby bylo docíleno širokého prostorového dojmu. Někdy je také prováděno fázové zpoždění některého z kanálů o 90° před smícháváním stereo signálu na mono. Důvodem je potlačení zvýšení úrovně částí signálu, které jsou přítomny ve všech audio kanálech (mono složka).

4.3 Odstranění stejnosměrné složky

Stejnosemřná složka (tzn. pokud je aritmetický průměr signálu různý od nuly) neposkytuje žádné užitečné informace a může mít nežádoucí dopad na výsledek parametru.

Při výpočtu offline máme všechna audio data $x_{DC}(i)$ o délce L dostupná a je možné jednoduše vypočítat aritmetický průměr a odečíst ho od každého vzorku

$$x(i) = x_{DC}(i) - \frac{1}{L} \sum_{i=0}^{L-1} x_{DC}(i). \quad (4.2)$$

V reálném čase vzniká možnost použití filtru typu horní propusti pro odstranění stejnosměrné složky. Jednou z metod jak provádět tento proces je aplikace derivačního članku:

$$x(i) = x_{DC}(i) - x_{DC}(i-1). \quad (4.3)$$

Derivační článek ovšem způsobuje silné potlačení vysokých frekvencí signálu. Tento efekt může být potlačen použitím dolní propusti. Výsledný filtr stejnosměrné složky bude

$$x(i) = (1 - \alpha) \cdot (x_{DC}(i) - x_{DC}(i-1)) + \alpha \cdot x(i-1). \quad (4.4)$$

4.4 Normalizace

Ve snaze zjištění parametrů nezávisle na amplitudě vstupního signálu, může být signál normalizován na předdefinovanou (maximální) úroveň nebo výkon.

Jednoduchá a často používaná metoda při normalizaci audio signálu je detekce celkového maxima jeho absolutní hodnoty vzorku a škálovat signál tak, že absolutní hodnota maxima je rovna jedné:

$$x(i) = \frac{x_s(i)}{\max_{\forall i}(|x_s(i)|)} \quad (4.5)$$

Tímto přístupem je zajištěn normalizovaný modul signálu, ale není zaručena stejná hlasitost různých vstupních signálů.

Normalizace signálu v reálném čase je obtížná. Může být prováděna algoritmy pro automatické řízení zesílení nebo kompresory a limitery, které monitorují okamžitou vstupní charakteristiku signálu, jsou řízeny například střední hodnotou RMS nebo okamžitou maximální hodnotou (peak level) a mají za cíl nastavovat časově proměnnou hodnotu zesílení.

4.5 Další možnosti předzpracování signálu

Mezi další možnosti předzpracování signálu patří například decimace. Při decimaci je používán algoritmus konverze vzorkovací frekvence, který konvertuje vstupní vzorkovací frekvenci f_s na nižší vzorkovací frekvenci f_a . Decimace redukuje jak počet vzorků audio signálu, tak šířku jeho výsledného spektra. Tento proces nebude používán, protože by došlo ke ztrátě informací na vyšších kmitočtech. Tak jako výběr vhodných parametrů, možnosti předzpracování signálu budou upravovány podle aplikace. Každé předzpracování, které má za důsledek vylepšení přesnosti, stabilitu a snižuje výpočetní nároky algoritmu je prospěšné. Další z metod předzpracování signálu může být například potlačení nechtěné frekvenční oblasti pomocí signálového filtru.

5 PARAMETRIZACE HUDEBNÍHO SIGNÁLU

5.1 Nízkoúrovňové parametry

Nízkoúrovňové parametry jsou přímo vypočítávány z reprezentací audio signálu jak v časové, tak ve frekvenční oblasti. Každý z těchto parametrů je počítán v rámcové úrovni a z časového hlediska je okamžitý [12].

Tab. 5.1: Přehled nízkoúrovňových parametrů

Parametr	Zkratka	Oblast
Amplitudová obálka signálu	AE	Časová
Efektivní hodnota signálu	RMS	Časová
Počet průchodů nulovou úrovní	ZCR	Časová
Poměr energií pásem	BER	Frekvenční
Spektrální centroid	SC	Frekvenční
Šířka pásma	BW	Frekvenční
Spektrální proměnlivost	SF	Frekvenční

Přestože každou hodnotou parametru je popisován pouze jeden rámeček, je možné shrnout všechny hodnoty parametru příslušející ke stejné nahrávce. To může být prováděno pomocí výpočtu statistických funkcí jako například průměr, medián, minimum nebo maximum. Mezi další metody mimo jiné patří modelování rozdělení jednotlivých parametrů vztahujících se ke stejné nahrávce pomocí funkce Gaussian mixture model (GMM) - parametrická funkce rozložení pravděpodobnosti. Tyto funkce jsou běžně používány jako modely pravděpodobnostního rozložení parametrů daného systému [31].

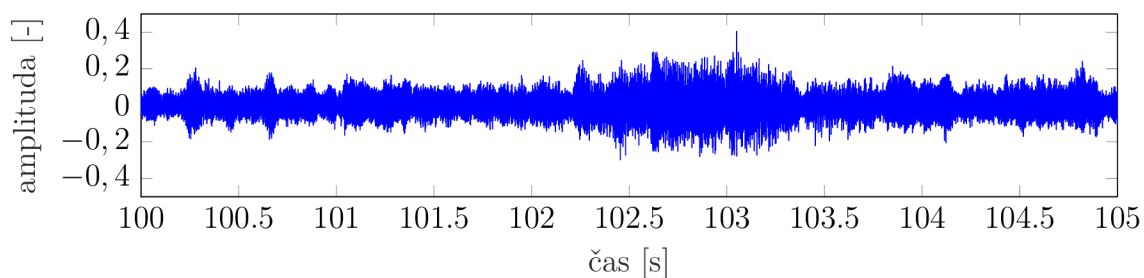
Tab. 5.2: Termíny používané při definicích nízkoúrovňových parametrů

Pojem	Popis
$s(k)$	Amplituda k-tého vzorku
$m_t(n)$	Modul signálu ve frekvenční oblasti v rámci t ve frekvenčním pásmu n
K	Velikost rámce, počet vzorků v jednom rámci
N	Číslo nejvyššího frekvenčního pásma

5.1.1 Časová oblast

Amplitudová obálka

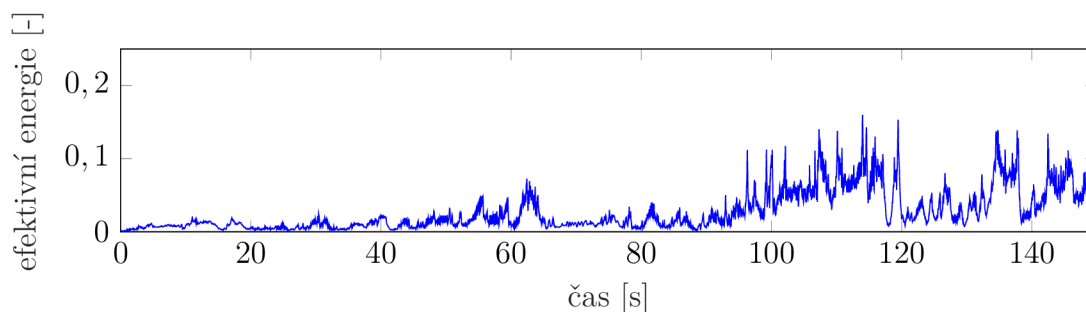
Amplitude Envelope neboli *amplitudová obálka* je maximální amplituda všech vzorků daného rámce t . Amplitudou obálky je znázorňován vnější tvar signálu. Tato reprezentace je přínosná při sledování dlouhodobého vývoje signálu.



Obr. 5.1: Ukázka amplitudové obálky hudebního signálu.

Efektivní hodnota signálu

Root-Mean-Square Energy neboli *efektivní hodnota signálu*, také označována jako hladina RMS je jeden z dalších časově závislých parametrů [2]



Obr. 5.2: Ukázka efektivní energie hudebního signálu.

$$\text{RMS}_t = \sqrt{\frac{1}{K} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1} s(k)^2}. \quad (5.1)$$

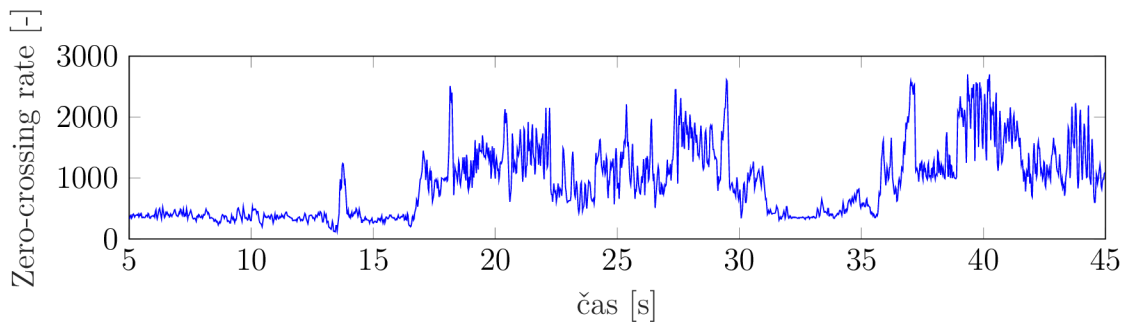
Efektivní hodnota je jeden z běžných parametrů popisující intenzitu a může být použit jako ukazatel hlasitosti nebo jako indikace změny v audio nahrávce.

Počet průchodů nulovou úrovní

Je používán název *zero-crossing rate*. Tímto parametrem je udáván počet, kolikrát se změnilo znaménko hodnoty amplitudy na daném rámcu t podle rovnice

$$\text{ZCR}_t = \frac{1}{2} \cdot \sum_{k=t \cdot K}^{(t+1) \cdot K - 1} |\text{sgn}(s(k)) - \text{sgn}(s(k+1))|. \quad (5.2)$$

Přestože je tento parametr používán zejména v oblasti rozpoznávání řeči, v hudební nahrávce mohou být detekovány perkusivní zvuky nebo hluk. V oblasti klasické hudby je tento parametr obvykle nižší a více stabilní [8].



Obr. 5.3: Počet průchodů nulovou úrovní.

5.1.2 Frekvenční oblast

Poměr energií pásem

Tento parametr původně *band energy ratio* je používán pro rozlišování mluveného slova, hudby a mimo jiné i pro klasifikaci hudebního žánru [12]. Pomocí tohoto parametru jsou porovnávány energie v nižších pásmech spektra s vyššími pásmy, čímž je měřeno jak nízké frekvence převažují [20]

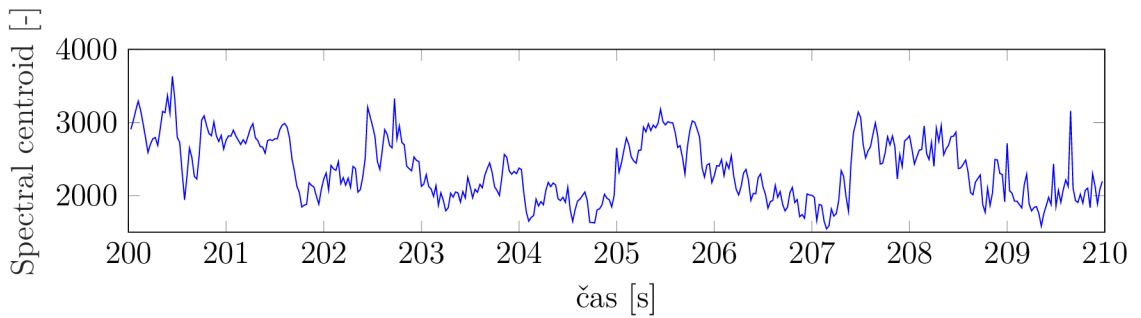
$$\text{BER}_t = \frac{\sum_{n=1}^{F-1} m_t(n)^2}{\sum_{n=F}^N m_t(n)^2}. \quad (5.3)$$

F značí frekvenční pásmo, které rozděluje pásma nízkých a vysokých frekvencí. Zvolením pásma F je ovlivněn výsledný rozsah hodnot. Protože energie obsažená v nízkých pásmech je často mnohem vyšší než ve vysokých pásmech, je používán ořez nízkých frekvencí.

Spektrální centroid

Parametr, kterým je určováno centrum modulu frekvenčního spektra, neboli frekvenčního pásma, ve kterém je zastoupeno nejvíce energie. Je používán k měření „jasu“ zvuku a vztahuje se přímo k barvě zvuku [12]. Tento parametr je velkou měrou ovlivňován při filtrování signálu filtrem typu dolní propust, protože je při výpočtu kladen důraz spíše na vysoké frekvence. Tato vlastnost se může projevit jako problematická zejména při decimaci signálu (viz. kapitola 4.5), protože dojde k ořezu nebo zkreslení frekvencí vyšších jak polovina vzorkovací frekvence a je tím ovlivněna výsledná hodnota parametru. U klasické hudby jsou pozorovány obvykle nízké změny hodnoty tohoto parametru [33].

$$SC_t = \frac{\sum_{n=1}^N m_t(n) \cdot n}{\sum_{n=1}^N m_t(n)}. \quad (5.4)$$



Obr. 5.4: Ukázka spektrálního centroidu hudebního signálu.

Šířka pásma

Označení toho parametru bývá také uváděno jako *spectral spread* a je odvozeno z předchozího parametru spektrální centroid. Je jím označována oblast spektra, ve které se nacházejí důležité části spektra (okolí pásma s nejvyšší energií) a může být vyjádřen jako frekvenční průměr daného audio signálu [12]. Průměrná šířka spektra může být použita k popisu barvy zvuku nahrávky [17]

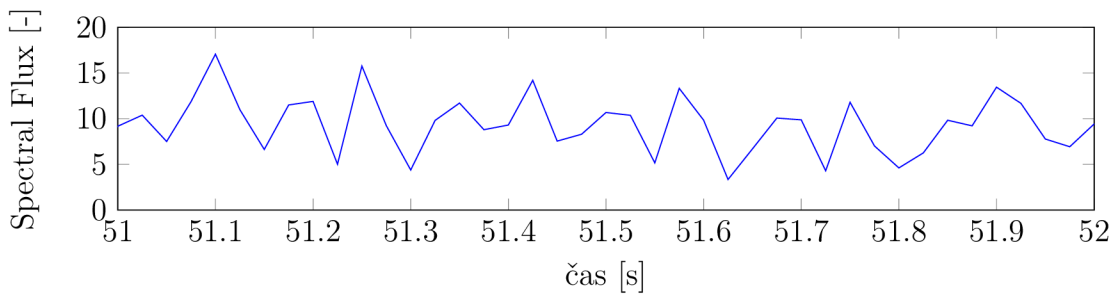
$$BW_t = \frac{\sum_{n=1}^N |n - SC_t| \cdot m_t(n)}{\sum_{n=1}^N m_t(n)}. \quad (5.5)$$

Spektrální proměnlivost

Parametr spektrální proměnlivosti neboli *spectral flux* může být vypočítán jako vzdálenost mezi spektry po sobě jdoucích rámců [12]

$$SF_t = \sum_{n=1}^N (D_t(n) - D_{t-1}(n))^2. \quad (5.6)$$

Tento parametr může být použit při provádění detekce řeči a také pro popis barvy zvuku audio signálu a detekci počátků not (onset detection) [4].



Obr. 5.5: Ukázka spektrální proměnlivosti hudebního signálu

5.2 Rytmické parametry

5.2.1 Fluktuace

Jedná se o rytmickou periodicitu podél sluchových kanálů *fluctuation*. Tento parametr je pokládán za jednu z možností jak odhadovat rytmus. Výpočet je založen na spektrogramu, který je transformován sluchovým modelem a poté je vypočítán odhad spektra v každém pásmu [25]. Nejdříve je vypočítán výkonnostní spektrogram, přičemž jsou použity rámce o velikosti 23 ms. Poté je provedena vícepásmová redistribuce energie za použití *bark* nebo *mel* pásem. Jsou odhadnuty maskovací efekty a nakonec jsou reprezentovány amplitudy pomocí dB stupnice. Dalším krokem je provedení rychlé Fourierovy transformace (FFT) na každém pásmu, čímž je dosaženo výpočtu matice rytmické periodicity pro každé pásmo *bark* či *mel*.

5.2.2 Spektrum úderů

Původním názvem je parametr označován jako *beat spectrum*. Rytmus je považován za důležitý prvek skladby. Dvě skladby se stejnou melodií mohou být zahrány s jiným tempem a rytmem a to má za následek i změnu nálady. Spektrum úderů je parametr, kterým je možno automaticky charakterizovat rytmus a tempo hudby.

Maxima tohoto parametru jsou obvykle pozorována u skladeb se složitějším rytmickým členěním, protože je tímto parametrem popisováno jak tempo, které se v čase mění, tak relativní síla jednotlivých úderů. Spektrum úderů nezávisí na energii nebo frekvenci, je počítáno v časové oblasti a může být použito pro všechny hudební žánry [38]. Výpočet je prováděn parametrizací při použití spektrální nebo jiné reprezentace.

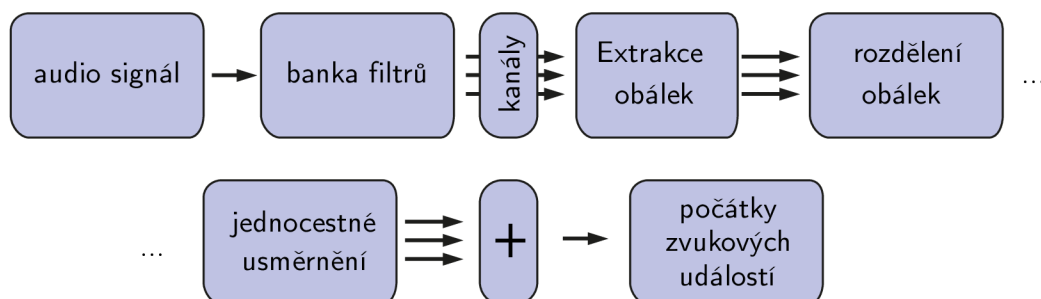
5.2.3 Počátek zvukové události

Tento parametr byl již zmíněn v kapitole 3.3.1. Často je prováděna detekce tohoto parametru v reálném čase za použití zejména krátkodobé Fourierovy transformace (STFT) [37].

Obecně je detekce prováděna ve třech krocích. Prvním krokem je předzpracování vstupního audio signálu, kdy je prováděna normalizace, poté je naprogramována detekční funkce a nakonec je proveden výběr maximálních hodnot v dané nahrávce (tzv. *peak picking*). Funkce pro detekci tohoto parametru může být implementována více způsoby. Jedním ze způsobů je výpočet pomocí parametru fluktuace spektra. Při této metodě je využito rozdílu v po sobě jdoucích frekvenčních pásech ve výkonovém spektru.

Při výpočtu křivky počátků zvukových událostí je nejdříve audio signál rozdělen do sluchových kanálů za použití banky filtrů. Existuje více typů těchto bank a počet kanálů může být měněn. Dále je extrahována obálka každého z kanálů [15]. Jelikož je pulzace obecně vztažena k nárůstu energie, tak jednotlivé obálky jsou rozděleny, jednocestně usměrněny a poté opět sečteny. Touto cestou je dosaženo přesného popisu změn energie, které jsou způsobeny každou zvukovou událostí (zahráním noty) z různých sluchových kanálů.

Výpočet počátků zvukových událostí je znázorněn na níže uvedeném blokovém diagramu 5.6.



Obr. 5.6: Blokové schéma výpočtu křivky počátků zvukových událostí

5.2.4 Hustota

Výpočtem tohoto parametru dostaneme hodnotu kolik počátků zvukových událostí proběhne za jednu sekundu. Jinými slovy se jedná o průměrnou frekvenci událostí. Parametr je nazýván jako *density*.

Výpočet je prováděn dekompozicí na jednotlivé, nepřekrývající se rámce o délce 10 s.

5.2.5 Tempo

Jak již bylo zmíněno v kapitole 3.3.2, tento parametr je uváděn v jednotkách BPM. Odhad je prováděn detekcí periodicit pomocí křivky detekce počátků hudebních událostí.

Obecné schéma pro odhadování tempa je založeno na detekci periodicit v rozsahu hodnot v jednotkách BPM a poté ve výběru maximální hodnoty periodicity v každém rámci zvlášť.

Často je také uváděna hodnota změny tempa jako rozdíl mezi po sobě jdoucími hodnotami tempové křivky.

5.2.6 Metrum

Metoda používána při výpočtu tohoto parametru vychází zejména z křivky počátků zvukových událostí, která je dále analyzována. Dále je prováděn odhad periodicity. Křivka parametru zvukových událostí může být vypočítána jak z amplitudové obálky, tak za použití parametru fluktuace spektra.

Obě metody poskytují relevantní výsledky zejména v nahrávkách, kde jsou nástroje od sebe izolovány nebo jsou jejich party zdůrazněny.

Parametr metra je důležitý zejména pro vizualizace rytmické složitosti a rytmického členění. Úrovně metra jsou určovány na základě výpočtu autokorelačního diagramu, metrická struktura je vysledována pomocí algoritmu.

5.2.7 Metrický centroid a metrická síla

Jsou používány názvy *metric centroid* a *metric strenght*. Metrický centroid je považován za odhad metrické aktivity, který je postaven na porovnávání centroidu s vybranými metrickými úrovněmi. Výsledkem je křivka metrického centroidu, která může být použita jako vývoj tempa v čase. Tyto hodnoty mohou být porovnávány s již dříve vypočtenými hodnotami tempa. Vysokými hodnotami tohoto parametru jsou obvykle znázorňovány více elementární rytmická členění dané skladby. Nízké hodnoty naopak ukazují, že je rytmické členění složitější.

Dynamickou metrickou silou je znázorňováno, jestli se jedná o zřetelnou pulzaci a pravidelnou metrickou hierarchii nebo naopak jestli je pulzace nějakým způsobem skrytá nebo nepravidelná, či se jedná o komplexní mixturu pulzací.

5.2.8 Čitelnost pulzu

Jedná se o parametr, kterým je popisováno, jaká je míra srozumitelnosti vnímání rytmické nebo metrické pulzace během dané skladby nebo její části. Původní název je *pulse clarity*

Díky charakterizaci, která je poskytována tímto parametrem je možné ho použít při rozpoznávání hudebního žánru, protože je jím umožněno podrobněji třídít a nalézat odlišnosti mezi jednotlivými žánry. Jedná se zejména o žánry, které mají sice podobné průměrné tempo, ale jsou odlišné v míře výraznosti pulzace.

Pro výpočet tohoto parametru je nejdříve vypočítána křivka změny tempa v čase, která vychází z počátků hudebních událostí. Pro popis čitelnosti pulzu je tato křivka dále charakterizována.

5.3 Dynamické parametry

5.3.1 Efektivní hodnota signálu

Za typický parametr popisující dynamiku audio signálu je možné považovat efektivní hodnotu signálu. Tento parametr byl již zmíněn v kapitole 5.1.1.

5.3.2 Segmentace

Tímto parametrem je vypočítávána segmentace v pozicích, kdy se v audio signálu vyskytuje ticho. Původní název parametru je *segmentation*. Segmenty jsou vybrány z časových pozic, kdy se efektivní hodnota zvýší nad prahovou úroveň, do úrovně kdy efektivní hodnota klesne pod prahovou úroveň.

5.3.3 Nízká energie

Křivka popisující energii *low energy*, může být použita k odhadu rozložení energie v čase. Tímto způsobem je zjištěno jestli je energie v čase konstantní nebo jsou-li naopak některé rámce více kontrastní než ostatní. Jeden ze způsobů odhadu je výpočet míry nízké energie, tedy počet rámců, které mají energii menší než je průměrná energie [35].

5.4 Parametry popisující barvu zvuku

5.4.1 Doba náběhu

Původní název je *attack time*. Jedna z možností jak popsat tento parametr je jeho zjištění z křivky počátků hudebních událostí. Pomocí tohoto parametru popisována doba od první zvukové vlny vydané nástrojem, do nabytí maximální velikosti amplitudy. Mnoho z algoritmů je naprogramováno s použitím pevné prahové úrovně. Kupříkladu, jestliže energie audio signálu přesáhne tuto úroveň o 10 %, jedná se o počátek náběhu. Konec doby náběhu je určen již zmíněnou prahovou úrovní. Tento algoritmus byl dále vylepšen o automaticky nastavovanou prahovou úroveň odhadovanou na základě charakteru signálu při samotné fázi náběhu [26].

5.4.2 Sklon náběhu

Attack slope neboli sklon náběhu může být popsán na základě průměrného sklonu samotné křivky náběhu. Jsou vypočítány lokální nárůsty energie a následně je spočítán vážený průměr těchto nárůstů. Vážený průměr je volen z důvodu zdůraznění hodnot nárůstů v polovině fáze náběhu. Při výpočtu je používána Gaussova funkce centrovaná okolo prahové úrovně 50 % se standardní odchylkou 0,5 [27].

5.4.3 Intenzita náběhu

Jedná se o parametr popisující rozdíl amplitud na počátku a na konci fáze náběhu, původní název je uváděn jako *attack intensity*. Hodnoty tohoto parametru jsou vyjádřeny ve stejné stupnici jako byl původní audio signál.

5.4.4 Sklon poklesu

Tímto parametrem je vyjadřován popis části *release* audio signálu, původní název je uváděn jako *decrease slope*. Parametr je velice příbuzný parametru sklonu náběhu. Výpočtem tohoto parametru mohou být odlišeny zvuky z hlediska jejich délky. Příkladem krátkého zvuku může být například perkusivní zvuk anebo hráčská technika *pizzicato* při hře na housle. Výpočet je založen na poklesu exponenciálního modelu obálky energie, pokles začíná v bodě kdy je energie největší t_{\max} . Výpočet je dán rovnicí

$$\hat{e}(t_n) = Ae^{-\alpha(t_n - t_{\max})}t_n \max, \quad (5.7)$$

kde α je odhadnuto lineární regresí logaritmu obálky energie [27].

5.4.5 Doba trvání

Parametr, jímž je popisována doba trvání noty v sekundách, tedy od náběhu do fáze *release*. Je jím dále měřena vnímaná délka signálu. Tímto parametrem lze také odlišovat perkusivní zvuky od zvuků s delší dobou trvání. Výpočet je prováděn amproximací a jedná se o čas kdy obálka energie přesáhne danou prahovou úroveň. Jednou z používaných prahových úrovní je 40 % [27]. Původní název je *duration*.

5.4.6 Počet průchodů nulovou úrovní

Tento parametr byl již zmíněn v kapitole 5.1.1. Patří mezi parametry popisující i barvu zvuku.

5.4.7 Pokles spektrální energie

Původní název je uváděn jako *spectral rolloff*. Jedná se o jeden ze způsobů jak odhadnout míru zastoupení vysokých frekvencí v signálu. To je provedeno nalezením dané frekvence, pod kterou je obsažen daný zlomek celkové energie audio signálu. Procentuálně je množství této energie vyjádřeno jako 85 % nebo 95 % [35].

5.4.8 Světlost spektra

Světlost spektra neboli *spectral brightness* je parametr používaný při výpočtu procentuálního zastoupení množství energie nad zvolenou frekvencí. Tato frekvence má hodnoty 1 kHz, 1,5 kHz nebo 3 kHz [14].

5.4.9 Melovské keprální koeficienty (MFCC)

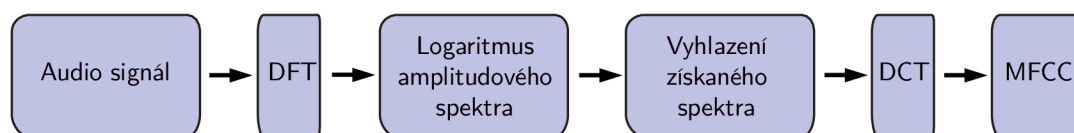
Melovské keprální koeficienty jsou krátkodobé spektrální parametry hojně používané zejména v oblasti rozpoznávání lidské řeči. V prvním kroku výpočtu těchto parametrů je audio signál rozdělen do rámců pomocí okna v pevně daných intervalech. Cílem tohoto kroku je rozdělit signál na krátké sekce (řádově desítky milisekund). Typicky je tento krok prováděn pomocí Hammingova okna, které má za účel odstranit přechodové jevy.

Dalším krokem je výpočet diskrétní Fourierovy transformace (DFT) na každém rámci audio signálu. Po provedení DFT je vypočten logaritmus amplitudového spektra signálu. Informace o fázi signálu není pro další výpočty důležitá [18]. S logaritmem amplitudového spektra je počítáno, protože vnímaná hlasitost signálu má přibližně logaritmický průběh.

Následně je nutné provést vyhlazení spektra s důrazem na frekvence, které jsou vnímány jako důležité. Tento úkon je prováděn shromážděním spektrálních komponentů do menšího počtu frekvenčních oblastí. Rozmístění těchto oblastí je dáno tzv. Mel stupnicí, která byla již zmíněna v kapitole 3.2.

Stupnice Mel je založena na závislosti mezi skutečnou frekvencí a vnímanou výškou tónu. Tato závislost je přibližně lineární do frekvence 1 kHz a logaritmická nad frekvencí 1 kHz.

Složky Melovských spektrálních vektorů vypočítané pro každý rámeček signálu jsou vysoce korelované. Posledním krokem výpočtu Melovských keprálních koeficientů je aplikování transformace na Melovské spektrální vektory, což má za důsledek redukce autokorelace jejich složek. Pro tento krok je používána diskrétní Kosinová transformace (DCT). V konečném důsledku jsou získány Melovské keprální koeficienty (MFCC) pro každý rámeček audio signálu. Celý postup výpočtu Melovských keprálních koeficientů je znázorněn na blokovém schématu 5.7.



Obr. 5.7: Blokové schéma výpočtu Melovských keprálních koeficientů

Použití MFCC při analýze hudebního signálu

Přestože Melovské keprální koeficienty nacházejí využití zejména v rozpoznávání a analýze lidské řeči, existuje již řada výzkumů, při kterých bylo využito těchto koeficientů v analýze hudebního signálu. V počátcích nebylo zcela jisté zda-li budou Melovské keprální koeficienty pro získávání informací z hudebního signálu vhodné, viz publikace [18]. Následovaly ale výzkumy, ve kterých bylo dosaženo dobrých výsledků.

Příkladem může být například výzkum který provedl v roce 2009 Chang-Hsing Lee et al., přičemž byla na základě spektrálních a keprálních parametrů provedena klasifikace hudebních žánrů s přesností 86,63 % [10]. V roce 2015 provedl N. J. Nalini et al. úspěšný výzkum, kde bylo využito Melovských keprálních koeficientů a parametrů reziduální fáze pro rozpoznávání hudebních emocí. Zejména se jednalo o emoce jako vztek, strach, štěstí nebo smutek. Pomocí kombinace těchto parametrů a následného statistického vyhodnocení různými metodami bylo dosaženo přesnosti 95 % – 99 % [24].

5.4.10 Drsnost

Tento parametr, původním názvem *roughness*, reprezentuje míru vnímané disonance (nelibozvuku) dané zvukové nahrávkou. Jev disonance vzniká, pokud se jednotlivé harmonické složky hudebního zvuku nacházejí blízko sebe ve frekvenční oblasti. Výsledkem tohoto parametru je odhad drsnosti zvuku na základě poměru každého páru sinusoid, které jsou obsaženy ve frekvenčním spektru daného zvuku. Čím více složek s dostatečně velkou amplitudou se nachází v těsné blízkosti, tím je zároveň větší celková drsnost zvuku [14].

5.4.11 Parametry spektrálního rozložení

Spektrální centroid a šířka pásma

Tyto parametry byly již zmíněny v kapitole 5.1.2.

Šikmost spektra

Tímto parametrem je udávána míra asymetričnosti spektra okolo jeho průměrné hodnoty, původní název je *spectral skewness*. Jestliže je hodnota rovna nule, jedná se o symetrické rozložení. V případě, že je hodnota menší než nula, je zastoupeno více energie v nízkých frekvencích. Při hodnotě větší než nula je více energie zastoupeno ve vysokých frekvencích.

Špičatost spektra

Jedná se o parametr, jímž je udávána míra špičatosti frekvenčního spektra kolem jeho průměrné hodnoty, původní název je *spectral kurtosis*. Jestliže je tato hodnota rovna třem, jedná se o normální (Gaussovo) rozložení, hodnota menší než tři znamená plošší rozložení, hodnota větší než tři potom strmější („špičatější“) rozložení.

Plochosť spektra

Parametr, který je vypočítán porovnáním geometrického průměru a aritmetického průměru frekvenčního spektra. Původní název tohoto parametru je *spectral flatness*. V kontextu charakterizace barvy zvuku je používáno jedno frekvenční pásmo pokrývající celý frekvenční rozsah. Při tonální analýze je hodnota blízká nule, spektrum obsahuje hodně vrcholů. Naopak při analýze šumu je hodnota bližší jedné, spektrum je na celém frekvenčním rozsahu vyrovnané.

6 VÝSLEDKY DOSAŽENÉ V ZAHRANIČÍ

Parametry popisující barvu zvuku byly doposud používány zejména ke klasifikaci (rozpoznávání) hudebních nástrojů a také k rozpoznávání hudebních žánrů. Běžně používané algoritmy jsou tvořeny dvěma částmi. První z nich se zabývá reprezentací audio signálu pro kterou jsou zvoleny nebo nově vytvořeny parametry. Druhá část spočívá ve zvolení klasifikačního modelu. Mezi často používané modely patří metoda nejbližšího souseda (k-NN), Gaussian mixture models (GMM) nebo skrytý Markovův model (HMM) [34].

V roce 2010 provedl Mathieu Barthe et al. [3] výzkum, který ukazuje rozdíl ve vněm barvy zvuku klarinetů za použití syntetického modelu, který pracuje v trojrozměrném prostoru. Vnímaná barva zvuku klarinetů byla nejlépe popsána parametry charakterizující náběh jednotlivých tónů (*attack time*) a spektrální bohatost (spektrální centroid). Tyto parametry byly již zmíněny v kapitole 5.4. Bylo prokázáno, že barva zvuku klarinetu je determinována zejména již zmíněným parametrem spektrální centroid.

V roce 2011 byl objasněn vzájemný některých používaných parametrů. Byla použita obsáhlá databáze hudebních nástrojů (více jak 6000 vzorků), bylo zjištěno, že původních 164 parametrů použitých v Timbre Toolboxu pro MATLAB spadá do 10 nezávislých kategorií, přičemž parametry v každé kategorii byly vysoce kolineární. Tato skutečnost naznačuje, že v mnoha situacích je možné využívat mnohem menšího počtu proměnných, než vykazují surová data parametrů. Na vývoji zmíněného toolboxu do software MATLAB pracoval Geoffroy Peeters et al. [27].

George Tzanetakis et al. [36] provedl výzkum klasifikace nástrojů pomocí jednotlivých harmonických složek. Při této metodě byl použit postup porovnávání jednotlivých párů audio vzorků. Touto metodou bylo zjištěno, které hudební nástroje korespondují s danou harmonickou složkou. Výsledky přinesly konečný odhad nástrojů přítomných v audio signálu. Byly použity dvě databáze, přičemž se ukázalo, že použití této metody může být v některých případech přesnější než klasifikace lidským posluchačem.

Claudia Fritz et al. [5] zkoumala anglické slovní popisy používané houslisty k hodnocení kvality a barvy zvuku jednotlivých houslí. Bylo nasbíráno 61 běžně používaných popisů, které byly poskládány houslisty na mapu tak, že slova s podobným významem ležela blíže sobě, kdežto slova s jinými významy ležela od sebe vzdáleně. Dále byly vybrány slovní popisy, které jsou používány pro popis podobné kvality zvuku. Tyto termíny a jejich vztahy byly použity pro zjištění vnímaného efektu akustické modifikace zvuků houslí v pěti oktávových frekvenčních pásmech.

7 SYNCHRONIZACE INTERPRETACÍ

Jedním z aspektů pro úspěšnou statistickou analýzu interpretací je krok synchronizace interpretací a následný výběr stejného notového úseku ze všech nahrávek databáze zmíněné v kapitole 10.1.

Výběr stejného notového úseku by mohl být proveden metodou poslechu a následného stříhu všech nahrávek, což by ovšem bylo velmi časově náročné.

Alternativní a v současnosti hojně používanou metodou je synchronizace všech interpretací pomocí dynamického borcení časové osy (DTW) za využití parametrů, kterými jsou nahrávky charakterizovány v čase na základě jejich tónového obsahu. Všechny dostupné interpretace jsou dynamicky a zejména tempově odlišné. Tato metoda poskytuje informaci o časovém umístění vybraného notového úseku ve všech interpretacích. V kapitole 7.2 je tato metoda dále popsána.

7.1 Parametry popisující tónový obsah

Lidské vnímání výšky tónu by se dalo popsat jako periodické, v tom smyslu, že dvě výšky tónu jsou vnímány jako barevně podobné, jestliže se liší o oktávu. Na základě této skutečnosti, může být výška tónu rozdělena do dvou součástí, jimiž jsou frekvence a tónová třída.

Tímto postupem se dá získat takzvaný *Chroma* parametr. Tento parametr je reprezentován vektorem o 12 dimenzích, kde každá dimenze reprezentuje jeden tón z tónových tříd, viz kapitola 3.2.

Ve fázi extrakce parametru je daný hudební signál převeden na sekvence chroma parametrů, kdy každá sekvence reprezentuje kolik krátkodobé energie signálu je rozprostřeno do 12 chroma pásem.

Při identifikaci výšek tónů, které se od sebe o oktávu liší, vykazují chroma parametry vysokou míru odolnosti vůči změnám v barvě zvuku a blízce korelují s hudebním aspektem harmonie. Tato vlastnost je také důvodem, proč jsou chroma parametry, často nazývané jako profily tónových tříd, zavedenými nástroji pro zpracování a analýzu hudebních dat [7].

Příkladem může být obecně postup při rozpoznávání akordů, přičemž se Chroma parametry také staly standardem v úkolech jako je hudební synchronizace a zarovnání [11].

Existuje více než jedna metoda pro výpočet chroma parametrů. Jednou z metod je například převod hudebního signálu na chroma reprezentaci (Chromagram) za využití krátkodobé Fourierovy transformace nebo za použití vhodných bank filtrů [23]. Dále můžou být vlastnosti chroma parametrů měněny zavedením vhodných

metod předzpracování a zpracování, které modifikují spektrální, časové a dynamické aspekty. Tímto je dosaženo velké řady variant chroma parametrů, které mohou vykazovat různé vlastnosti v kontextu dané analýzy hudebního signálu.

7.1.1 Reprezentace výšky tónu

Prvním krokem při extrakci chroma parametrů je rozložení hudebního signálu do 88 frekvenčních pásem, přičemž střední frekvence těchto pásem korespondují s výškami tónů od $A0$ do $C8$ (v oblasti MIDI jsou to výšky tónů $p = 21$ až $p = 108$). Pro dosažení dostatečného spektrálního rozlišení pro nižší je zapotřebí nízká vzorkovací frekvence nebo velké časové okno. V Chroma Toolboxu pro MATLAB je používána banka filtrů s konstantní jakostí a je využíváno vzorkovací frekvence 22050 Hz pro vysoké tóny, 4410 Hz pro středně vysoké tóny a 882 Hz pro nízké tóny [23]. Použité filtry mají relativně velkou šířku propustného pásma, ale zároveň správně rozdělují po sobě jdoucí noty díky strmému poklesu v přechodových pásmech. Filtry výšky tónů jsou odolné vůči odchylkám až do ± 25 centů oproti tónu s výškou střední frekvence daného filtru. Aby nedocházelo k většímu fázovému zkreslení, je využíváno metody *forward-backward filtering*. Výsledný výstupní signál má nulové fázové zkreslení a jeho modul je upraven druhou mocninou modulové charakteristiky daného filtru [29].

V dalším kroku je vypočítán pro každou z 88 výšek tónů krátkodobý výkon (vzorky každé výšky tónu jsou umocněny na druhou) za použití okna s pevně danou délkou a překryvem 50 %. Příkladem může být použití okna o délce 200 ms, to vede k četnosti parametru 10 Hz, tedy 10 parametrů za sekundu. Výslednými parametry, které jsou pojmenovány jako *Pitch* (výška tónu), je udávána informace o obsahu energie v každém pásmu výšky tónu.

7.1.2 Problematika ladění

Jelikož nemusí být nástroje v každé nahrávce naladěny na stejné základní frekvence, musí být počítáno s posunem každé střední frekvenci pásmových filtrů banky. Toto je provedeno výpočtem průměru vektoru spektrogramu a odtud je odvozena odchylka ladění simulací posunů banky filtrů za použití váženého rozdělení viz [7]. V Chroma Toolboxu pro MATLAB je vypočítáno 6 různých bank filtrů, které korespondují na odchylku ladění o $0, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}$ a $\frac{3}{4}$ půltónu [21]. Z těchto bank filtrů je vybrána jedna, která odpovídá odhadnuté odchylce v ladění.

7.1.3 Parametr Chroma Pitch

Z reprezentace výšky tónu zmíněné v předchozí kapitole je možné vypočítat chroma reprezentaci jednoduše součtem hodnot, které náleží stejné chroma skupině. Například pro výpočet hodnoty korespondující chroma skupině C , musí být proveden součet hodnot výšek tónů $C1, C2, \dots, C8$. Pro každé vypočítané okno je vypočítán vektor o 12 dimenzích $x = (x(1), x(2), \dots, x(12))^T$, kde $x(1)$ odpovídá chroma skupině C , $x(2)$ odpovídá chroma skupině $C\sharp$ atd.

7.1.4 Normalizace chroma parametru

Pro dosažení neměnnosti v dynamice nahrávky může být provedena normalizace parametrů odpovídající vhodné normě. Je brána v potaz norma ℓ^p -norma definovaná jako $\|x\|_p := (\sum_{i=1}^{12} |x(i)|^p)^{1/p}$ pro daný chroma vektor $x = (x(1), x(2), \dots, x(12))^T$ a zvolené přirozené číslo $p \in \mathbb{N}$ [21].

Aby nedošlo k náhodnému rozložení energií v pasážích velmi nízké energie (například pasáže ticha před začátkem nahrávky nebo během dlouhé pauzy), je chroma vektor x nahrazen vektorem první normy, jestliže $\|x\|_p$ bude menší než daná prahová úroveň. Výchozím nastavením při výpočtu je $p = 2$ a jedná se tedy o Eukleidovskou normu.

7.1.5 Parametr Chroma-Log-Pitch

Vnímaná hlasitost má přibližně logaritmický průběh, a tak je při výpočtu tohoto parametru použita logaritmická komprese amplitudy. Každá hodnota energie e reprezentace tónového obsahu je nahrazena hodnotou $\log(\eta \cdot e + 1)$, kde η je vhodná kladná konstanta [21]. Poté jsou vypočítány chroma hodnoty jak již bylo zmíněno v kapitole 7.1.3. Výsledné parametry, které závisí na kompresním parametru η , jsou nazývány jako *Chroma-Log-Pitch* a je označován jako CLP [η].

7.1.6 Parametr CENS

V další úrovni chroma parametru je využito krátkodobé statistiky pro rozložení energií v chroma skupinách. Takto upravený chroma parametr se nazývá CENS (*Chroma Energy Normalized Statistics*). Jedná se o skupinu škálovatelných a robustních audio parametrů. Bylo dokázáno, že tyto parametry jsou velice užitečné v oblasti MIR [13].

Při výpočtu CENS parametrů je každá chroma skupina nejprve normalizována vzhledem k normě ℓ^1 , tudíž je každým vektorem vyjadřováno relativní rozložení energie. Následně je provedena kvantizace s vhodně zvolenou prahovou úrovní. V dalším

kroku je poté parametr vyhlazen oknem o délce $w \in \mathbb{N}$ a podvzorkován koeficientem d . Výsledné parametry jsou normalizovány podle normy ℓ^2 a jsou značeny jako CENS_d^w .

7.2 Dynamické borcení časové osy

Operace roztahování a zkracování částí audio signálu v čase jsou nazývány jako borcení časové osy *time warping*. Díky tomu, že je borcení časové osy ovlivněno dynamicky během procesu porovnávání dvou sekvencí parametrů, je používán termín *Dynamic Time Warping (DTW)* pro pojmenování zarovnání dvou sekvencí parametrů, které reprezentují v čase dané audio signály. Jedním z těchto parametrů může být i CENS, který byl již zmíněn v kapitole 7.1.6.

Nechť jsou dány dvě sekvence parametrů $\mathbf{X} = \underline{x}(i); i = 1, \dots, T_x$ a $\mathbf{R} = \underline{r}(j); j = 1, \dots, T_r$, které mají být zarovnány. Obecně platí, že $T_x \neq T_r$ a zároveň $\underline{x}(i)$ a $\underline{r}(j)$ jsou vektory o K dimenzích [6].

Dále, necht' je dána funkce rozdílnosti $d(\underline{a}, \underline{b})$, například Eukleidovská vzdálenostní funkce, která je definována jako

$$d(\underline{a}, \underline{b}) = \sqrt{\sum_{l=1}^K (\underline{a}_l - \underline{b}_l)^2}. \quad (7.1)$$

Prvním krokem při výpočtu DTW algoritmu je vytvoření tzv. matice ceny (cost matrix), C . Sekvence parametrů \mathbf{R} a \mathbf{X} jsou vyneseny na horizontální a vertikální osu, respektive souřadnice uzlu (i, j) jsou i -tý element sekvence \mathbf{X} a j -tý element sekvence \mathbf{R} . Zároveň je použit algoritmus, který využívá dynamické programování. Algoritmus začíná od uzlu $(1, 1)$ a uzly matice jsou zkoumány po řádcích zleva doprava. V každém uzlu je vypočítána akumulovaná cena, za kterou se do daného uzlu algoritmus dostal. Na konci tohoto procesu reprezentuje cena v uzlu (T_x, T_r) celkovou cenu synchronizace mezi dvěma sekvencemi parametrů. Čím vyšší je celková cena, tím více jsou dané dvě sekvence rozdílné, jinými slovy, jsou si méně podobné [6].

V této práci bylo použito tzv. *Itakura* schéma pro výpočet DTW a je popsáno v následující kapitole 7.2.1.

7.2.1 DTW Algoritmus Itakura

V případě této práce se interpretace z databáze radikálně neliší. Jedná se zejména o změny v lokálním tempu a dynamice. Nástrojové obsazení je stejné a změny v barvě zvuku jsou v tomto případě z hlediska synchronizace nahrávek zanedbatelné. V

takovém případě se jako vhodný jeví algoritmus *Itakura*, který využívá následující podmínky lokální cesty [6]:

$$\begin{aligned}
 C(i, j) &= \min C(i-1, j-1), C(i-2, j-1), C(i, j-1) + d(\underline{x}(i), \underline{r}(j)), \\
 &\quad \text{jestliže } p_B(i, j-1) \neq (i, j-2) \\
 &\quad \text{a} \\
 C(i, j) &= \min C(i-1, j-1), C(i-2, j-1) + d(\underline{x}(i), \underline{r}(j)), \\
 &\quad \text{jestliže } p_B(i, j-1) = (i, j-2).
 \end{aligned} \tag{7.2}$$

Z definice vyplývá, že vertikální posun mezi uzly je zcela zakázán, je možné vynechat ze sekvence vektory parametrů, které byly umístěny na vertikální ose, díky faktu že uzel $(i-1, j-1)$ je přípustný předchůdce. Dále je nutno poznamenat, že dva horizontální po sobě jdoucí posuny mezi uzly nejsou povoleny, díky podmínce, která vyžaduje aby $p_B(i, j-1) \neq (i, j-2)$ [6]. Díky těmto vlastnostem je zajištěno, že borcení časové osy nemůže být libovolně intenzivním jevem.

Implementace použitá při synchronizaci nahrávek z databáze pochází z Audio Analysis Library toolboxu pro MATLAB [6]. Funkce pro výpočet Itakura DTW algoritmu má název `dynamicTimeWarpingItakura()`.

7.3 Postup při výběru notového úseku

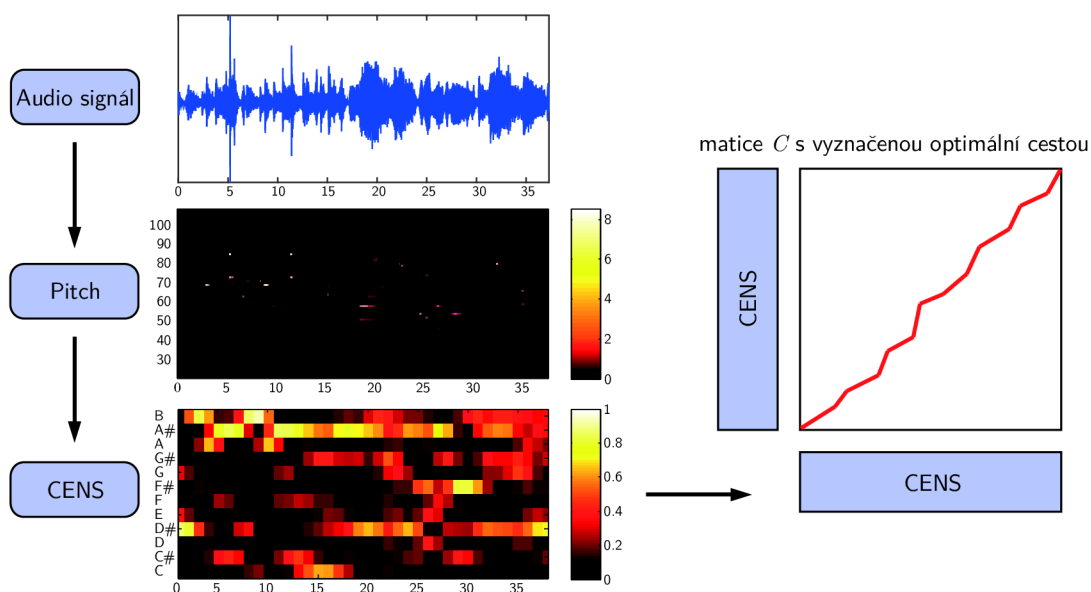
Prvním krokem při výběru stejného notového úseku bylo podvzorkování všech nahrávek na vzorkovací frekvenci 22050 Hz a následný převod ze stereofonních nahrávek na jejich mono verzi. Tento krok je prováděn zejména kvůli redukci dat a také se jedná o počáteční krok při výpočtu parametrů tónového obsahu pomocí Chroma Toolboxu pro MATLAB. Zmíněný převod je zajištěn funkcí `wav_to_audio()`, která je obsažena ve výše uvedeném toolboxu. Všechny další uvedené funkce mimo výpočet dynamického borcení časové osy pocházejí z tohoto toolboxu.

Z takto předzpracovaných nahrávek je v dalším kroku vypočítán parametr s názvem *Pitch*, který reprezentuje rozložení jednotlivých tónů v čase, viz kapitola 7.1.1. Toto je prováděno funkcí `audio_to_pitch_via_FB()`. Tato funkce má parametr `parameter.winLenSTSMP`, kterým je ovlivněna délka okna. Při výchozím nastavení je hodnota tohoto parametru 4410. To při vzorkovací frekvenci 22050 Hz znamená, že je toto okno 200 ms dlouhé. Jednotlivá okna se z 50 % překrývají, to znamená, že ve výsledku je parametr *Pitch* vypočítán 10 krát za sekundu. V rámci této práce byla zvolena hodnota parametru `parameter.winLenSTSMP` 441, tedy parametr *Pitch* byl vypočítán 100 krát za sekundu. Zavoláním funkce `audio_to_pitch_via_FB()`

je získána matice o rozměrech $120 \times N$, kde N je počet časových rámců a její první dimenze odpovídá výškám tónů podle MIDI.

V následujícím kroku je ze získaného parametru *Pitch* vypočítán parametr CENS zavoláním funkce `pitch_to_CENS()`. Tato funkce má parametry w a d , které byly popsány v kapitole 7.1.6. Při výchozím nastavení jsou parametry této funkce nastaveny na hodnoty `winLenSmooth = 41` a `windowSampSmooth = 10`. Při tomto nastavení je vypočítán parametr CENS z výše uvedeného parametru *Pitch* 10 krát za sekundu. Získaný vektor má 12 dimenzí, přičemž každá z nich reprezentuje jednu chroma skupinu.

Nakonec bylo nutné zvolit vhodný notový úsek, který bude podroben analýze. Tento úsek byl zvolen poslechem a za pomoci DTW algoritmu, který byl zmíněn v kapitole 7.2.1, byl vybrán ze všech interpretací databáze. Partitura vybraného notového úseku je obsažena v příloze B. Důvodem pro výběr tohoto úseku byla skutečnost, že se v něm objevuje část kdy hraje pouze jeden nástroj (viola) a také tento úsek obsahuje repetici. Tyto dvě skutečnosti mohou zlepšit rozlišení daných interpretací.



Obr. 7.1: Postup synchronizace nahrávek

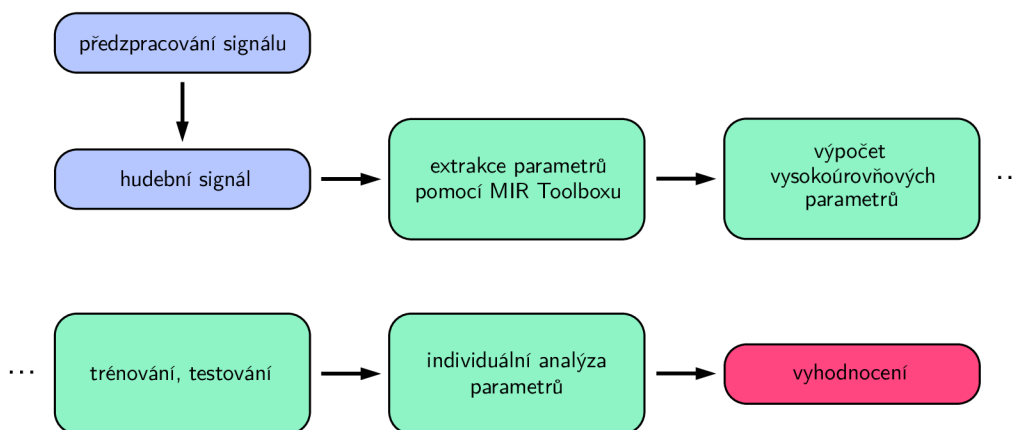
Na výše uvedeném obrázku 7.1 je naznačen postup při synchronizaci interpretací. Ze všech nahrávek byly vypočítány sekvence parametrů CENS, kterými je popsán tónový obsah nahrávek v čase. Druhým krokem bylo porovnávání těchto sekvencí pomocí algoritmu DTW zavoláním funkce `dynamicTimeWarpingItakura()`. Výstupem této funkce je optimální cesta mezi dvěma nahrávkami a cena za kterou se algoritmus dostane z počátečního uzlu matice C do uzlu konečného. Jak již bylo

zmíněno, čím je hodnota ceny nižší, tím jsou si dané dvě nahrávky více podobné. Všechny nahrávky byly porovnávány vzhledem k první nahrávce databáze. Poté byl poslechově vybrán úsek z 12. nahrávky a díky vypočítanému DTW je již získána informace, kde se tento notový úsek nachází ve všech ostatních nahrávkách. Nahrávka číslo 12 byla zvolena za výchozí, protože se jedná o interpretaci zahranou Janáčkovým kvartetem, které bylo založeno v roce 1947 studenty brněnské konzervatoře.

Synchronizace interpretací byla provedena s přesností na 0,1 s. Vyšší přesnost není proveditelná pomocí klasického DTW algoritmu zejména kvůli narůstajícím nárokům na paměť RAM. Pro účely větší přesnosti synchronizace jsou používány algoritmy MsDTW[23] a MrMsDTW[28]. Větší přesnost není v případě této práce nutným předpokladem zejména díky tomu, že pro analýzu nahrávek budou použity vysokoúrovňové statistické parametry.

8 NÁVRH VYHODNOCOVACÍHO SYSTÉMU

Vyhodnocování bude prováděno pomocí systému, který bude diferencovat jednotlivé interpretace databáze nahrávek, která je zmíněna kapitole 10.1. Pomocí tohoto systému bude ve výsledku možné diferencovat a zhodnotit jednotlivé interpretace analyzované skladby na základě parametrů popisujících barvu zvuku.



Obr. 8.1: Návrh vyhodnocovacího systému

Na obrázku 8.1 je naznačen diagram návrhu funkce vyhodnocovacího systému. Prvním krokem bude předzpracování audio signálu, přičemž bude z každé interpretace skladby volen stejný notový úsek, aby bylo možné co nejlépe nahrávky diferencovat. Volba stejného notového úseku byla již popsána v kapitole 7.3.

Dalším krokem bude provedení extrakce parametrů popisujících audio signál. V případě této práce, bude kladen důraz zejména na parametry popisující barvu zvuku, které byly zmíněny v kapitole 5.4. Výpočty budou prováděny pomocí prostředí MATLAB za použití MIR Toolboxu [14]. Všechna data získaná výpočtem parametrů jsou uložena ve formě vektorů.

Hned v dalším kroku bude proveden výpočet vysokoúrovňových statistických parametrů z výše zmíněných parametrů extrahovaných pomocí MIR Toolboxu.

Dalším krokem bude individuální vyhodnocení všech vypočtených vysokoúrovňových pomocí statistické analýzy. Pro statistickou analýzu budou použity klasifikační stromy a křížová validace.

V posledním kroku budou vybrány ty parametry, které mají nejlepší schopnost diferencovat jednotlivé interpretace.

9 EVALUACE SYSTÉMU

V následujících kapitolách jsou nastíněny metody, pomocí kterých je možné provést individuální klasifikaci jednotlivých parametrů

9.1 Klasifikační stromy

Klasifikační stromy nacházejí své využití v mnoha oborech. Jsou používány v disciplínách jako diagnóza, kognitivní věda, umělá inteligence nebo data mining. Hlavním úkolem klasifikačních stromů je vyprodukovat přesný klasifikátor a pochopení předpovědní struktury daného problému.

Klasifikační strom klasifikuje instance pomocí třídění do stromu od jeho kořene až do jednotlivých listů, které poskytují klasifikaci instance. Každý uzel stromu specifikuje a testuje danou vlastnost instance a každá větev vycházející z daného uzlu koresponduje s možnou hodnotou této vlastnosti.

Klasifikace instance začíná v kořenovém uzlu stromu testováním vlastnosti, která je specifikována tímto uzlem. Poté se pokračuje směrem dolů po větvích stromu, kde větve odpovídají vlastnostem daného příkladu.

9.2 Křížová validace

Křížová validace je statistická metoda vyhodnocování a porovnávání učících se algoritmů rozdělením dat na dvě části. Jedna část je použita pro trénování a druhá pro testování daného modelu. Tato metoda je používána zejména v případě, kdy je potřeba vyhodnotit soubor, který nedisponuje velkým množstvím prvků.

Základní formou křížové validace je takzvaná *k-fold* křížová validace. Ostatní formy křížové validace jsou specifické případy *k-fold* křížové validace nebo vyžadují několik opakování.

Jedním z těchto případů může být metoda *leave-one-out*. V tomto případě *k-fold* křížové validace je *k* rovno počtu prvků v analyzovaném souboru. Jinými slovy v každé iteraci jsou téměř všechna data použita pro trénování a model je testován pomocí jednoho prvku.

9.3 Vyhodnocení algoritmů

V této kapitole jsou představeny analytické nástroje, pomocí kterých je možné zhodnotit učící se algoritmy a porovnat jejich klasifikační schopnosti.

9.3.1 Přesnost

Pomocí přesnosti (*accuracy*) může být hodnocena klasifikace provedená pomocí binárního klasifikátoru. Tato klasifikace produkuje 4 možné výsledky.

- *True positive (TP)*: počet správně pozitivních nálezů
- *False positive (FP)*: počet falešně pozitivních nálezů
- *True negative (TN)*: počet správně negativních nálezů
- *False negative (FN)*: počet falešně negativních nálezů

Pokud bereme v potaz tyto výsledky, celková přesnost (ACC) je definována jako:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \cdot 100. \quad (9.1)$$

Výsledkem této rovnice je procentuální vyjádření přesnosti (ACC).

9.3.2 Spearmanův koeficient pořadové korelace

Při výpočtu korelace mezi dvěma veličinami se často používá Spearmanův koeficient (*Spearman's rank sum correlation*). Tento koeficient je obvykle značen řeckým písmem ρ . Definice tohoto koeficientu vychází z předpokladu dvou neznámých X a Y u kterých není známo jejich pravděpodobnostní rozdělení. Je provedeno uspořádání n jejich hodnot $x_{(i)}$ a $y_{(i)}$ a podle velikosti jsou těmto hodnotám přiřazena pořadová čísla $p_{(i)}$ a $q_{(i)}$.

Hodnota koeficientu je pak rovna:

$$\rho = 1 - \frac{6 \sum_i (p_{(i)} - q_{(i)})^2}{n(n^2 - 1)}. \quad (9.2)$$

Spearmanův koeficient nabývá hodnot z intervalu $\langle -1; 1 \rangle$, přičemž krajní hodnoty tohoto intervalu udávají nejvyšší korelaci.

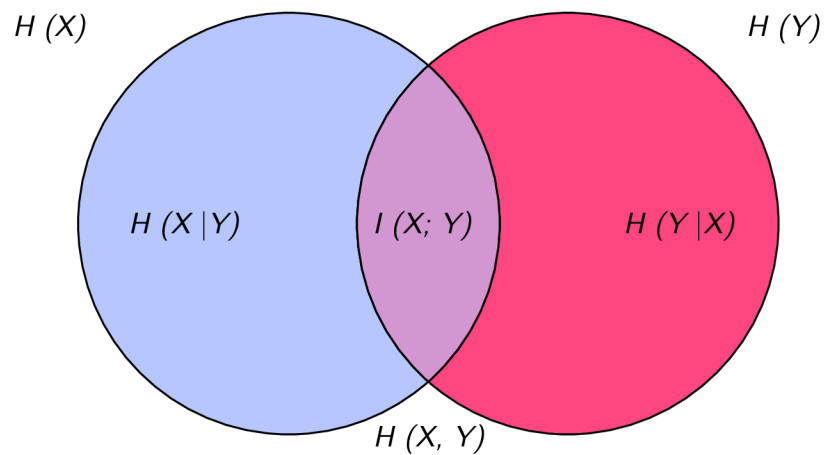
9.3.3 Vzájemná informace

Vzájemná informace je veličina, která porovnává vztah mezi dvěma náhodnými veličinami. Přesněji řečeno, touto veličinou je udávána míra vzájemné závislosti dvou proměnných. Tato veličina je značena jako MI.

Definice vzájemné informace dvou náhodných proměnných X a Y , jejichž sdružené rozložení je definováno jako $P(X, Y)$ je dána rovnicí [16]

$$I(X; Y) = \sum_{x \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right), \quad (9.3)$$

kde $P(X)$ a $P(Y)$ jsou marginální distribuční funkce proměnných X a Y . Hodnota MI se pohybuje v intervalu $\langle 0; 1 \rangle$, přičemž hodnota 1 značí maximální míru závislosti proměnných X a Y .



Obr. 9.1: Grafické znázornění vzájemné informace.

9.3.4 Mann-Whitney U test

Další možnou metodou pro vyhodnocení algoritmu je metoda Mann-Whitney U test nazývaná také jako *Wilcoxon rank-sum test* (WRS). Jedná se o neparametrický test nulové hypotézy, když dva vzorky pocházejí ze stejné populace proti alternativní hypotéze, a to zejména v případě populace, která má tendenci nabývat vyšších hodnot než druhá. Využívá se pro hodnocení nepárových pokusů porovnávající 2 různé výběrové soubory (tzv. pokusný zásah A , B). Testuje se hypotéza, že veličina X odpovídá pokusnému zásahu A a naopak veličina Y odpovídá pokusnému zásahu B mají shodné rozdělení pravděpodobnosti. Zároveň veličiny X a Y nemusí odpovídat Gaussovu normálnímu rozdělení.

10 TESTOVÁNÍ A VYHODNOCENÍ

V této kapitole je popsána datábáze nahrávek, která bude podrobena analýze. Dále jsou představeny parametry, které budou vypočítány. Nakonec je shrnuta individuální analýza parametrů a jsou vybrány ty parametry, které mají nejlepší schopnost rozlišit jednotlivé interpretace.

10.1 Databáze nahrávek k analýze

Analýze bude podrobena čtvrtá věta Houslového kvartetu č. 1 českého hudebního skladatele Leoše Janáčka. Databáze nahrávek je tvořena 35 verzemi této skladby v podání kvartet z celého světa. Jako zdroj byly použity nahrávky v bezztrátovém formátu ALAC a nahrávky z CD. V rámci základního předzpracování byly nahrávky převedeny do formátu `.wav` a sjednoceny na vzorkovací frekvenci 44,1 kHz a bitovou hloubku 16 bitů. Přehled všech nahrávek je uveden v příloze C.

Důvodem použití bezztrátových zdrojových formátů je zejména to, že při kompresi audio nahrávek dochází ke ztrátám dat, které by mohly ovlivnit výpočet parametrů. Ke ztrátám dochází zejména vlivem využití frekvenčního a časového maskování při procesu komprese.

Nahrávky byly rozděleny do dvou kategorií, podle toho, jestli pocházejí z Česka nebo se jedná o nahrávky světové. Rozdělení bylo provedeno pomocí kategorické proměnné v prostředí MATLAB, českým nahrávkám byla přidělena kategorie `'Česko'` a světovým `'svet'`. Toto rozdělení bylo použito pro individuální analýzu všech vypočítaných vysokoúrovňových parametrů, přičemž byl použit model klasifikačních stromů a následně bylo provedeno zhodnocení schopnosti parametrů správně zařadit jednotlivé interpretace do těchto kategorií.

10.2 Vypočítané parametry

V této práci bylo hlavním úkolem vypočítat co nejvíce parametrů, které popisují barvu zvuku a z těch pak vybrat ty, které mají největší schopnost diferencovat dané interpretace. Celkem bylo vypočítáno přesně 300 parametrů, přičemž 120 z nich bylo vypočítáno z Melovských kepstrálních koeficientů 2–13. Tyto parametry byly vypočítány u každé z 35 nahrávek databáze interpretací. Seznam všech těchto parametrů je uveden v příloze D.

Jak již bylo zmíněno, všechny parametry byly vypočítány pomocí MIR toolboxu pro prostředí MATLAB. Protože se jedná o data ve formě vektorů, byly z těchto dat

dále vypočítány vysokoúrovňové statistické parametry, které byly vstupní informací pro učící se algoritmy. Seznam těchto parametrů je znázorněn v tabulce 10.1.

Tab. 10.1: Přehled vysokoúrovňových parametrů

Označení parametru	Popis
mean	střední hodnota
var	rozptyl
std	směrodatná odchylka
median	medián
1p	1. percentil
99p	99. percentil
ir	interpercentilový rozsah
Q1	1. kvartil
Q3	3. kvartil
IQR	interkvartilový rozsah

Výpočet parametrů byl proveden v prostředí MATLAB. Všechny vypočítané parametry popisující barvu zvuku byly představeny v kapitole 5.4. Jednotlivé parametry byly rozděleny do 5 kategorií. Každé z těchto kategorií odpovídá vypočtená tabulka uložená do souboru `.mat`, kde každý sloupec reprezentuje jeden parametr a každý řádek odpovídá dané interpretaci. V níže uvedeném seznamu jsou uvedeny názvy kategorií, do kterých byly parametry roztrženy.

- parametry popisující tvar spektra
- parametry popisující křivku počátků zvukových událostí
- parametry popisující rozložení spektrální energie
- Melovské keprstrální koeficienty
- počet průchodů nulovou úrovní a drsnost

10.3 Individuální analýza parametrů

Z kategorií, které byly zmíněny v předchozí kapitole 10.2, bylo vybráno vždy 5 nejlepších parametrů podle vyhodnocené přesnosti. Pokud jsou brány v potaz parametry ze všech kategorií, přesnost se u nich pohybuje od 77,14 % do 97,14 %. Tyto parametry vykazují nejlepší schopnost zařadit dané interpretace do již zmíněných kategorií 'Cesko' a 'svet'. Takto vysoká míra přesnosti je ale do jisté míry dána i faktem, že databáze nahrávek není obsáhlá a čítá pouze 35 interpretací. V databázi je obsaženo celkem 9 českých a 26 světových nahrávek, což může mít za následek

zvýšení přesnosti na úkor zkreslení konečných výsledků. Jinými slovy, pokud by v hypotetické situaci učící se algoritmus zařadil všechny nahrávky do kategorie 'svet', byla by i přesto celková přesnost přibližně 74,29 %. I přes tyto skutečnosti byl získán přehled o tom, které parametry mají v tomto případě nejlepší rozlišovací schopnosti.

Jako první byla provedena individuální analýza Melovských keprálních koeficientů, viz tabulka 10.2. Je možné vyzorovat, že parametry vypočítané z 8. a 9. MFCC vykazují velmi vysokou přesnost. Konkrétně parametr mfcc9 (mean) vykazuje přesnost 97,14 % – nesprávně byla zařazena pouze jedna nahrávka. Přesnost ostatních parametrů se pohybuje v rozmezí 82,86 % do 88,57 %. Výše zmíněný parametr mfcc9 (mean) má také nejvýznamnější hodnotu vzájemné informace (MI = 0,6461) a hodnotu Spearmanova koeficientu pořadové korelace ($\rho = 0,9252$).

Tab. 10.2: Výsledky individuální analýzy Melovských keprálních koeficientů.

Parametr	ACC	MI	p	ρ
mfcc9 (mean)	97,14 %	0,6461	0,7881	0,9252
mfcc9 (median)	88,57 %	0,3111	0,5774	0,6864
mfcc9 (Q3)	88,57 %	0,3111	0,5774	0,6864
mfcc9 (Q1)	85,71 %	0,2314	0,3902	0,5997
mfcc8 (median)	82,86 %	0,1764	0,5774	0,5230

ACC – přesnost, MI – vzájemná informace, p – hladina významnosti (Mann-Whitney U test), ρ – Spearmanův koeficient pořadové korelace.

V kategorii parametrů, které popisují rozložení spektrální energie mají shodnou přesnost parametry bright1_5k (median) a bright1_5k (1p) (ACC = 82,86 %). Tyto parametry mají také nejvýznamnější hodnotu Mann-Whitney U testu ($p = 0,2387$) a Spearmanova koeficientu pořadové korelace ($\rho = 0,5071$). Nejvýznamnější hodnotu vzájemné informace (MI = 0,1669) vykazuje parametr bright1_5k (ir).

Tab. 10.3: Výsledky individuální analýzy parametrů popisujících rozložení spektrální energie.

Parametr	ACC	MI	p	ρ
bright1_5k (median)	82,86 %	0,1621	0,2387	0,5071
bright1_5k (1p)	82,86 %	0,1621	0,2387	0,5071
roll85 (var)	80,00 %	0,1374	0,7881	0,4581
roll85 (std)	80,00 %	0,1374	0,7881	0,4581
bright1_5k (ir)	80,00 %	0,1669	0,7955	0,4961

ACC – přesnost, MI – vzájemná informace, p – hladina významnosti (Mann-Whitney U test), ρ – Spearmanův koeficient pořadové korelace.

Z kategorie parametrů popisujících spektrální rozložení, měly parametry *skw* (mean) a *krt* (Q1) největší a zároveň shodnou přesnost ($ACC = 85,71\%$). Parametr *skw* (mean) vykazuje také nejvýznamnější hodnotu Spearmanova koeficientu pořadové korelace ($\rho = 0,6138$). Nejvýznamnější hodnotu Mann-Whitney U testu ($p = 0,1293$) má parametr *flt* (median).

Tab. 10.4: Výsledky individuální analýzy parametrů popisujících spektrální rozložení.

Parametr	ACC	MI	p	ρ
<i>skw</i> (mean)	85,71 %	0,2487	0,7881	0,6138
<i>krt</i> (Q1)	85,71 %	0,2314	0,3902	0,5997
<i>skw</i> (99p)	82,86 %	0,1374	1,0000	0,5513
<i>flt</i> (median)	80,00 %	0,1374	0,1293	0,4051
<i>ent</i> (1p)	80,00 %	0,1669	0,7955	0,4961

ACC – přesnost, MI – vzájemná informace, p – hladina významnosti (Mann-Whitney U test), ρ – Spearmanův koeficient pořadové korelace.

Z kategorie parametrů, které popisují křivku počátků zvukových událostí má parametr *atktime* (99p) největší přesnost ($ACC = 85,71\%$). Dále má tento parametr nejvýznamnější hodnotu vzájemné informace ($MI = 0,2314$) a Spearmanova koeficientu pořadové korelace ($\rho = 0,5997$).

Tab. 10.5: Výsledky individuální analýzy parametrů popisujících křivku počátků zvukových událostí.

Parametr	ACC	MI	p	ρ
<i>atktime</i> (99p)	85,71 %	0,2314	0,3902	0,5997
<i>duration</i> (var)	82,86 %	0,1764	0,5774	0,5230
<i>duration</i> (std)	82,86 %	0,1764	0,5774	0,5230
<i>atsklope</i> (IQR)	82,86 %	0,1764	0,5774	0,5230
<i>atktime</i> (ir)	80,00 %	0,1374	0,7881	0,4581

ACC – přesnost, MI – vzájemná informace, p – hladina významnosti (Mann-Whitney U test), ρ – Spearmanův koeficient pořadové korelace.

Do poslední kategorie parametrů jsou zařazeny parametry drsnost a počet průchodů nulovou úrovní. Největší přesnost ($ACC = 85,71\%$) má parametr *rgl* (99p). Parametry *rgl* (99p) a *rgl* (ir) mají shodně významnou hodnotu Mann-Whitney U testu ($p = 0,0000$) a také hodnotu Spearmanova koeficientu pořadové korelace ($\rho = -0,5071$).

Tab. 10.6: Výsledky individuální analýzy parametrů drsnost a počet průchodů nulovou úrovní.

Parametr	ACC	MI	p	ρ
rgl (99p)	85,71 %	0,1621	0,0000	-0,5071
rgl (ir)	82,86 %	0,1621	0,0000	-0,5071
zcr (median)	80,00 %	0,1019	0,5774	0,4051
zcr (Q1)	80,00 %	0,1019	0,5774	-0,4051
rgl (var)	77,14 %	0,0838	0,7881	-0,3595

ACC – přesnost, MI – vzájemná informace, p – hladina významnosti (Mann-Whitney U test), ρ – Spearmanův koeficient pořadové korelace.

Z výsledků individuální analýzy vyplývá, že by databáze interpretací musela být obsáhlejší, aby bylo možné snížit zkreslení výsledku vzniklé malým počtem nahrávek. I přesto se však vyhodnocení určitých parametrů dá považovat za směrodatné. Největší přesnost při klasifikaci vykázal parametr mfcc9 (mean) (ACC = 97,14 %), což se i při takto malém počtu nahrávek dá pokládat za významný výsledek. Pokud by se měla potvrdit schopnost ostatních parametrů při zařazování do výše zmíněných kategorií 'Cesko' a 'svet' bylo nutné provést analýzu s větším počtem nahrávek a zároveň by bylo vhodné, aby byl v obou kategoriích přibližně stejný počet nahrávek.

11 ZÁVĚR

První kapitola této práce se zabývá shrnutím dosavadních poznatků o parametrech popisující nahrávky na základě barvy zvuku. Je proveden náhled do metod zpracování signálu zejména z pohledu disciplíny Music Information Retrieval (MIR).

Dále je nastíněno zpracování hudebního signálu z pohledu hudební teorie a číselového zpracování signálů, přičemž jsou vysvětleny některé základní pojmy z těchto dvou oblastí.

Třetí kapitolou je předzpracování signálu, kde jsou zpracovány některé z postupů, které jsou používány před samotnou analýzou hudebních nahrávek.

Následuje kapitola, ve které jsou shrnuty běžně používané parametry, pro získávání informací z hudebních nahrávek, jedná se o parametry nízkourovňové, rytmické, dynamické a pro tuto práci stěžejní parametry popisující barvu zvuku.

Následně bylo představeno dynamické borcení časové osy (DTW), které bylo použito při synchronizaci interpretací a následném výběru stejného notového úseku ze všech nahrávek. Notový úsek byl ze všech nahrávek vybrán s přesností 0,1 s. Pokud by byla vyžadována vyšší přesnost synchronizace daných nahrávek, musely by být v dalším výzkumu implementovány již zmíněné algoritmy MsDTW[23] a MrMsDTW[28], zejména z důvodu nároků na paměť RAM, které vzniknou při použití klasického DTW algoritmu. Konkrétně v této práci byl pro synchronizaci použit algoritmus *Itakura*, který je součástí Audio Analysis Library toolboxu pro MATLAB. Ze získaných notových úseků byly vypočítány parametry, které byly představeny v kapitole 5.4, přičemž výpočet těchto parametrů byl proveden pomocí MIR Toolboxu pro MATLAB. Takto vypočtené parametry jsou reprezentovány ve formě vektorů.

V další kapitole je přestaven návrh vyhodnocovacího systému, pomocí kterého byly nahrávky klasifikovány do dvou kategorií 'Cesko' a 'svet'. Při evaluaci tohoto systému byly použity klasifikační stromy a křížová validace. Pro vyhodnocení klasifikace je stěžejním nástrojem zejména přesnost (ACC), ale byla také využita vzájemná informace (MI), Spearmanův koeficient pořadové korelace (ρ) a Mann-Whitney U test (p).

Následuje kapitola testování a vyhodnocení, kde je představena databáze nahrávek, která byla podrobena analýze. Dále jsou zde představeny vysokoúrovňové statistické parametry, které byly podrobeny individuální analýze. Bylo vypočítáno celkem 300 vysokoúrovňových statistických parametrů z již zmíněných parametrů vypočítaných za pomoci MIR Toolboxu v prostředí MATLAB. Všechny vypočítané parametry jsou shrnuty v příloze D. Je nutné zmínit, že byla analýza prováděna na omezeném počtu nahrávek, a tak jsou výsledky individuální analýzy do jisté míry zkreslené. Za zmínku ovšem stojí vysokoúrovňové parametry vypočítané z Melov-

ských kepstrálních koeficientů, kde zejména 8. a 9. MFCC vykazují vysokou přesnost (ACC = 82,86 % – 97,14 %) při klasifikaci nahrávek.

Celkově se přesnost klasifikace všech vybraných parametrů pohybuje od 77,14 % do 97,14 %. Je nutné poznamenat, že pokud by měly mít výsledky individuální analýzy parametrů větší vypovídající hodnotu, bylo by nutné provést další výzkum na větším počtu nahrávek, kde by kategorie 'Cesko' a 'svet' byly přibližně stejně zastoupené.

LITERATURA

- [1] AMBIKAIRAJAH, E. *Auditory masking and MPEG-1 audio compression* [online]. 1997 [cit. 2016-12-11]. Dostupné z URL: <http://personal.ik.itba.edu.ar/~nbaum/Clase%20Audio%20Compression/Auditory%20masking%20and%20MPEG-1%20audio%20compression.pdf>
- [2] CAETANO, M. *Automatic Segmentation of the Temporal Evolution of Isolated Acoustic Musical Instruments Sounds Using Spectro-Temporal Cues*. In: *13th International Conference On Digital Audio Effects, DAFX 2010 Proceedings* [online]. 2010, s. 1–8 [cit. 2016-11-19]. ISBN 978-320001940-9. Dostupné z URL: <http://articles.ircam.fr/textes/Caetano10b/index.pdf>
- [3] BARTHET, M. *From Clarinet Control to Timbre Perception* [online]. 2010: s. 678–689. [cit. 2016-12-11]. Dostupné z URL: <http://www.lma.cnrs-mrs.fr/~ystad/DocsPagePerso/Articles/ActaBarthet.pdf>
- [4] DIXON, S. *Onset detection revisited*. In: *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx)* [online]. (2006) [cit. 2016-11-22]. Dostupné z URL: http://www.dafx.ca/proceedings/papers/p_133.pdf
- [5] FRITZ, C. *Exploring violin sound quality: Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties* 2012. s. 783–794. [cit. 2016-12-11]. Dostupné z URL: https://www.researchgate.net/publication/221780563_Exploring_violin_sound_quality_Investigating_English_timbre_descriptors_and_correlating_resynthesized_acoustical_modifications_with_perceptual_properties
- [6] GIANNAKOPOULOS, T. *Introduction to audio analysis: a MATLAB approach*. ISBN 9780080993881.
- [7] GOMEZ, E. *Tonal Description of Music Audio Signals* 2006. [cit. 2017-05-04] Dostupné z URL: http://mtg.upf.edu/system/files/publications/emilia-PhD-2006_0.pdf
- [8] GOUYON, F. *On The Use of Zero-crossing Rate For An Application Of Classification of Percussive Sounds* In: *Proceedings of the COST G-6 conference on the digital audio effects (DAFX-00)* [online]. 2000. [cit. 2016-11-19]. Dostupné z URL: <http://www.mtg.upf.edu/files/publications/dafx00-gouyon.pdf>
- [9] HERRE J., UHLE, C. *Estimation of Tempo, Micro Time and Time Signature from Percussive Music* In: *Proceedings of the 6th International Conference on*

- Digital Audio Effects (DAFX)* London, 2003. Dostupné z URL: <http://www.eecs.qmul.ac.uk/legacy/dafx03/proceedings/pdfs/dafx46.pdf>
- [10] CHANG-HSING, L. *Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features* [online]. 2009 [cit. 2017-05-05]. Dostupné z URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4895319>.
- [11] JODER, C. *A comparative study of tonal acoustic features for a symbolic level music-to-score alignment*. In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* [online]. IEEE, 2010, s. 409-412 [cit. 2017-05-04]. DOI: 10.1109/ICASSP.2010.5495784. ISBN 978-1-4244-4295-9. Dostupné z URL: <http://ieeexplore.ieee.org/document/5495784/>
- [12] KNEES, P., SCHEDL M. *Music similarity and retrieval: an introduction to audio- and web-based strategies*. 1. New York, NY: Springer Berlin Heidelberg, 2016. ISBN 9783662497203.
- [13] KURTH, F. a MULLER M. *Efficient Index-Based Audio Matching*. *IEEE Transactions on Audio, Speech, and Language Processing* [online]. 2008, 16(2), 382-395 [cit. 2017-05-21]. DOI: 10.1109/TASL.2007.911552. ISSN 1558-7916. Dostupné z URL: <http://ieeexplore.ieee.org/document/4432645/>
- [14] LARTILLOT, O. *MIRtoolbox 1.6.1 User's Manual* [online]. 2014 [cit. 2016-12-11]. Dostupné z URL: <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/MIRtoolbox1.6.1guide>
- [15] LARTILLOT, O. *A MATLAB toolbox for musical feature extraction from audio* [online]. 2007 [cit. 2017-04-05]. Dostupné z URL: <http://dafx.labri.fr/main/papers/p237.pdf>
- [16] LEARNED-MILLER E. *Entropy and Mutual Information* [online]. 2013 [cit. 2017-05-05]. Dostupné z URL: <https://people.cs.umass.edu/~elm/Teaching/Docs/mutInf.pdf>.
- [17] LERCH, A. *Audio content analysis: an introduction*. 1 st. Hoboken, N.J.: Wiley, 2012. ISBN 9781118266823.
- [18] LOGAN, B. *Mel Frequency cepstral Coefficients for Music Modeling* [online]. 2000 [cit. 2017-04-05]. Dostupné z URL: http://ismir2000.ismir.net/papers/logan_paper.pdf.

- [19] MARCHAND-MAILLET, S. *Adaptive multimedia retrieval: user, context, and feedback* : 4th international workshop, AMR 2006, Geneva, Switzerland, July, 27-28, 2006 : revised selected papers. 1. Berlin: Springer, 2007. ISBN 3540715444.
- [20] MCKINNEY, M. *Features for Audio and Music Classification In: Proceedings of the Fourth International Conference on Music Information Retrieval* [online]., Baltimore, Maryland, USA, 2003. s. 1–8 [cit. 2016-11-22] ISBN 2-9746194-0-1. Dostupné z: <http://ismir2003.ismir.net/papers/McKinney.pdf>
- [21] MÜLLER, M. *Chroma Toolbox: MATLAB Implementations for Extracting Variants of Chroma-based audio features* [online]. 2011. [cit. 2017-05-01] Dostupné z URL: <https://pdfs.semanticscholar.org/31cf/b91158031038ad16218997cc9724a9ab5c66.pdf>.
- [22] MÜLLER, M. *Fundamentals of music processing*. 1. New York, NY: Springer Berlin Heidelberg, 2015. ISBN 9783319219448.
- [23] MÜLLER, M. *Information retrieval for music and motion*. Online-Ausg. New York: Springer, 2007. ISBN 9783540740483.
- [24] NALINI, N. J. *Music emotion recognition: The combined evidence of MFCC and residual phase* 2015 [online]. [cit. 05-05-2017]. Dostupné z URL: http://ac.els-cdn.com/S1110866515000419/1-s2.0-S1110866515000419-main.pdf?_tid=57e426b8-317f-11e7-9b7a-00000aacb360&acdnt=1493981046_c460ac6ad452557d692ed704c525f74e.
- [25] PAMPALK, E. *Content-based Organization and Visualization of Music Archives, In: Proceedings of ACM Multimedia* Juan-les-Pins, France. 2002. [cit. 2016-12-07]. Dostupné z URL: http://www.ofai.at/~elias.pampalk/publications/pam_mm02.pdf
- [26] PEETERS, G. *A large set of audio features for sound description (similarity and classification) in the CUIDADO project* 2004. [online]. [cit. 2016-12-07]. Dostupné z URL: http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf.
- [27] PEETERS, G. *The Timbre Toolbox: Extracting audio descriptors from musical signals* 2011. [online]. [cit. 2016-12-07]. Dostupné z URL: https://www.mcgill.ca/mpcl/files/mpcl/peeters_2011_jasa.pdf
- [28] PRATZLICH, T. *Memory-restricted multiscale dynamic time warping*. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing*

- (*ICASSP*) [online]. IEEE, 2016, s. 569-573 [cit. 2017-05-21]. DOI: 10.1109/ICASSP.2016.7471739. ISBN 978-1-4799-9988-0. Dostupné z URL: <http://ieeexplore.ieee.org/document/7471739/>
- [29] PROAKIS J. a DIMITRIS M. *Digital signal processing: principles, algorithms, and applications* 3rd ed. Upper Saddle River, N.J: Prentice Hall, 1996. ISBN 9780133758337.
- [30] REPP, B. H. *Patterns of Note Onset Asynchronies in Expressive Piano Performance, Journal of the Acoustical Society of America (JASA)*, vol. 100, no. 6, 1996. s. 3917–3932, DOI: 10.1121/1.417245
- [31] REYNOLDS, D. *Gaussian Mixture Models in Encyclopedia of Biometrics* [online]. Boston, MA: Springer US, 2009, s. 659 [cit. 2016-11-18]. ISBN 9780387730028. Dostupné z URL: http://www.springerlink.com/index/10.1007/978-0-387-73003-5_196
- [32] SCHEDL, M. *Music Information Retrieval: Recent Developments and Applications* [online]. 2014 [cit. 2016-12-11]. Dostupné z URL: <http://mtg.upf.edu/system/files/publications/059-music-information-retrieval-recent-developments-applications.pdf>
- [33] SCHUBERT, E. *Spectral Centroid and Timbre in Complex, Multiple Instrumental Textures In: Music perception and cognition ICMPC8* 1st Evanston, IL, 2004, s.654–657, ISBN 1876346507.
- [34] SIEDENBURG, K. *A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology* 2016: s. 27–41. [cit. 2016-12-11]. Dostupné z URL: <http://www.tandfonline.com/doi/full/10.1080/09298215.2015.1132737>
- [35] TZANETAKIS, G. *Automatic Musical Genre Classification Of Audio Signals* [online]. s. 1–4. [cit. 2016-12-07]. Dostupné z URL: <http://ismir2001.ismir.net/pdf/tzanetakis.pdf>.
- [36] TZANETAKIS, G. *Musical Instrument Classification Using Individual Partials* [online]. 2010: s. 111–122. [cit. 2016-12-11]. Dostupné z URL: <http://ieeexplore.ieee.org.ezproxy.lib.vutbr.cz/document/5428856/>
- [37] YAMADA, M. *Development of rhythm practice supporting system with real-time onset detection* [online]. DOI: 10.1109/ICAEE.2015.7506819. 2015. s. 1–4. [cit.

- 2016-12-07]. Dostupné z URL: <http://ieeexplore.ieee.org.ezproxy.lib.vutbr.cz/stamp/stamp.jsp?arnumber=7506819>
- [38] YANG, S. *Music Emotion Detecting Based On Beat Spectrum* [online]. DOI: 10.1109/ICCIS.2010.307. 2011. s. 1–4. [cit. 2016-12-07]. Dostupné z URL: <http://ieeexplore.ieee.org.ezproxy.lib.vutbr.cz/stamp/stamp.jsp?tp=&arnumber=5709507>

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

ACC	přesnost
ALAC	Apple Lossless Audio Codec
atkleap	attack leap – velikost náběhu
atktime	attack time – doba náběhu
atkslope	attack slope – sklon náběhu
BPM	Beats Per Minute
bright1k	světlost spektra s mezní frekvencí 1 kHz
bright1_5k	světlost spektra s mezní frekvencí 1,5 kHz
bright3k	světlost spektra s mezní frekvencí 3 kHz
C	cost matrix – matice ceny
CD	Compact Disc – kompaktní disk
CENS	Chroma Energy Normalized Statistics
CLP	Chroma Log Pitch
ctr	spectral centroid – spektrální centroid
dB	decibel
dBFS	decibel full scale
dcrslope	decrease slope – sklon poklesu
DCT	Discrete Cosine Transform – diskrétní kosinová transformace
DTW	Dynamic Time Warping – dynamické borcení časové osy
duration	doba trvání
ent	spectral entropy – spektrální entropie
FFT	Fast Fourier Transform - rychlá Fourierova transformace
flt	spectral flatness – plochost spektra
GMM	Gaussian Mixture Models

HMM	Hidden Markov Models – skryté Markovovy modely
k-NN	metoda nejbližšího souseda
krt	spectral kurtosis – špičatost spektra
MFCC	Melovské keprstrální koeficienty
MI	Mutual Information – vzájemná informace
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
MsDTW	Multi-scale Dynamic Time Warping
MrMsDTW	Memory restricted Multi-scale Dynamic Time Warping
p	hladina významnosti (Mann-Whitney U test)
Pitch	reprezentace výšky tónu
rg	roughness – drsnost
roll85	parametr spectral rolloff při 85 %
roll95	parametr spectral rolloff při 95 %
ρ	Spearmanův koeficient pořadové korelace
skw	spectral skewness – šikmost spektra
sprd	spectral spread – spektrální směrodatná odchylka
STFT	Short Time Fourier Transform – krátkodobá Fourierova transformace
T	tempo
wav	Waveform audio file format
ZCR	Zero Crossing Rate – počet průchodů nulovou úrovní

SEZNAM PŘÍLOH

A	Obsah přiloženého DVD	61
B	Partitura notového úseku	62
C	Databáze nahrávek k analýze	63
D	Seznam vypočítaných parametrů	65

A OBSAH PŘILOŽENÉHO DVD

Přiložené DVD obsahuje elektronickou verzi této bakalářské práce. Hlavní dokument je nazván „Miklanek-Stepan-BP“. Ve složce „Skripty“ jsou obsaženy skripty pro synchronizaci interpretací, skript pro výpočet parametrů popisujících barvu zvuku a skript pro vyhodnocení individuální analýzy parametrů. Výpočet parametrů popisujících barvu zvuku je prováděn pomocí skriptu „VypocetParametru“ a vyhodnocení individuální analýzy parametrů je prováděno pomocí skriptu „Vyhodnoceni“. Výsledky individuální analýzy jsou uloženy ve složce „Individualni-analyza“ jako soubor typu .xlsx pod názvem „Vysledky“.

Obsah:

- Miklanek-Stepan-BP.pdf
- Skripty
- Individualni-analyza

B PARTITURA NOTOVÉHO ÚSEKU

Violin I
Violin II
Viola
Violoncello

ff
sf [fff]
sf [ff sempre]

Measures 1-4 of the score. The key signature has three flats (B-flat, E-flat, A-flat) and the time signature is 3/8. The Violoncello part starts with a forte (*ff*) dynamic. The Viola part has a *ff* dynamic and includes the marking *[rfz]*. The Violin I and II parts feature triplets and sixteenth-note patterns.

Measures 5-7 of the score. The Violoncello part has a *sf* dynamic. The Viola part has a *sf* dynamic. The Violin I and II parts continue with their melodic lines, featuring triplets and sixteenth-note patterns.

Measures 8-10 of the score. The Violoncello part has a *sf* dynamic. The Viola part has a *sf* dynamic. The Violin I and II parts continue with their melodic lines, featuring triplets and sixteenth-note patterns.

C DATABÁZE NAHRÁVEK K ANALÝZE

Tab. C.1: Databáze nahrávek k analýze

Označení	Verze	Zdroj
001.wav	Alban Berg Quartet	ALAC
002.wav	Arcadia Quartet	ALAC
003.wav	Arianna Quartet	ALAC
004.wav	Britten Quartet	nezískán
005.wav	Brodsky Quartet	nezískán
006.wav	Cecilia String Quartet	ALAC
007.wav	Emerson Quartet	ALAC
008.wav	Endres Quartet	ALAC
009.wav	Guarneri Quartet	ALAC
010.wav	Hagen Quartet	ALAC
011.wav	Henschel Quartet	ALAC
012.wav	Janáček String Quartet	ALAC
013.wav	Jerusalem Quartet	ALAC
014.wav	Julliard String Quartet	CD
015.wav	Kubín Quartet	nezískán
016.wav	Mandelring Quartet	ALAC
017.wav	Martinu Quartet	ALAC
018.wav	Medici Quartet	ALAC
019.wav	Melos Quartet	ALAC
020.wav	New Helsinki Quartet	ALAC
021.wav	Panocha Quartet	ALAC
022.wav	Pavel Hass Quartet	ALAC
023.wav	Pražák Quartet	nezískán
024.wav	Quatuor Debussy	ALAC
025.wav	Raphael Quartet	ALAC
026.wav	Schoenberg Quartet	ALAC
027.wav	Smetana Quartet	ALAC
028.wav	Stamitz Quartet	ALAC
029.wav	Škampa Quartet	ALAC
030.wav	Talich Quartet	CD
031.wav	The Fry Street Quartet	ALAC
032.wav	Vanbrugh Quartet	ALAC
033.wav	Vlach Quartet Prague	nezískán

034.wav	Wihan Quartet	ALAC
035.wav	Doric String Quartet	ALAC
036.wav	Quartetto Energie Nove	ALAC
037.wav	Wroclaw Chamber Orchestra Leopoldinum, Ernst Kovacic	ALAC
038.wav	M. Nostitz Quartet	nezískán
039.wav	Quatuor Diotima	ALAC
040.wav	Australian Chamber Orchestra, Richard Tognetti	ALAC
041.wav	Lindsay String Quartet	nezískán
042.wav	London Sinfonietta and Wiener Philharmoniker	ALAC

D SEZNAM VYPOČÍTANÝCH PARAMETRŮ

Tab. D.1: Seznam vypočítaných parametrů

Č.	Název	Jedn.	Popis
1	mfccn (mean)	[-]	střední hodnota n -tého MFCC
.	mfccn (std)	[-]	směrodatná odchylka n -tého MFCC
.	mfccn (var)	[-]	rozptyl n -tého MFCC
.	mfccn (median)	[-]	medián n -tého MFCC
.	mfccn (1p)	[-]	1. percentil n -tého MFCC
.	mfccn (99p)	[-]	99. percentil n -tého MFCC
.	mfccn (ir)	[-]	interpercentilový rozsah n -tého MFCC
.	mfccn (Q1)	[-]	1. kvartil n -tého MFCC
.	mfccn (Q3)	[-]	3. kvartil n -tého MFCC
120	mfccn (iq)	[-]	interkvartilový rozsah n -tého MFCC
121	bright1k (mean)	[%]	střední hodnota světlosti spektra s mezní frekvencí 1 kHz
122	bright1_5k (mean)	[%]	střední hodnota světlosti spektra s mezní frekvencí 1,5 kHz
123	bright3k (mean)	[%]	střední hodnota světlosti spektra s mezní frekvencí 3 kHz
124	roll85 (mean)	[Hz]	střední hodnota parametru rolloff při 85 %
125	roll95 (mean)	[Hz]	střední hodnota parametru rolloff při 95 %
126	bright1k (std)	[%]	směrodatná odchylka světlosti spektra s mezní frekvencí 1 kHz
127	bright1_5k (std)	[%]	směrodatná odchylka světlosti spektra s mezní frekvencí 1,5 kHz
128	bright3k (std)	[%]	směrodatná odchylka světlosti spektra s mezní frekvencí 3 kHz
129	roll85 (std)	[Hz]	směrodatná odchylka parametru rolloff při 85 %
130	roll95 (std)	[Hz]	směrodatná odchylka parametru rolloff při 95 %
131	bright1k (var)	[%]	rozptyl světlosti spektra s mezní frek- vencí 1 kHz

132	bright1_5k (var)	[%]	rozptyl světlosti spektra s mezní frekvencí 1,5 kHz
133	bright3k (var)	[%]	rozptyl světlosti spektra s mezní frekvencí 3 kHz
134	roll85 (var)	[Hz]	rozptyl parametru rolloff při 85 %
135	roll95 (var)	[%]	rozptyl parametru rolloff při 95 %
136	bright1k (median)	[%]	medián světlosti spektra s mezní frekvencí 1 kHz
137	bright1_5k (median)	[%]	medián světlosti spektra s mezní frekvencí 1,5 kHz
138	bright3k (median)	[%]	medián světlosti spektra s mezní frekvencí 3 kHz
139	roll85 (median)	[Hz]	medián parametru rolloff při 85 %
140	roll95 (median)	[Hz]	medián parametru rolloff při 95 %
141	bright1k (1p)	[%]	1. percentil světlosti spektra s mezní frekvencí 1 kHz
142	bright1_5k (1p)	[%]	1. percentil světlosti spektra s mezní frekvencí 1,5 kHz
143	bright3k (1p)	[%]	1. percentil světlosti spektra s mezní frekvencí 3 kHz
144	roll85 (1p)	[Hz]	1. percentil parametru rolloff při 85 %
145	roll95 (1p)	[Hz]	1. percentil parametru rolloff při 95 %
146	bright1k (99p)	[%]	99. percentil světlosti spektra s mezní frekvencí 1 kHz
147	bright1_5k (99p)	[%]	99. percentil světlosti spektra s mezní frekvencí 1,5 kHz
148	bright3k (99p)	[%]	99. percentil světlosti spektra s mezní frekvencí 3 kHz
149	roll85 (99p)	[Hz]	99. percentil parametru rolloff při 85 %
150	roll95 (99p)	[Hz]	99. percentil parametru rolloff při 95 %
151	bright1k (ir)	[%]	interpercentilový rozsah světlosti spektra s mezní frekvencí 1 kHz
152	bright1_5k (ir)	[%]	interpercentilový rozsah světlosti spektra s mezní frekvencí 1,5 kHz
153	bright3k (ir)	[%]	interpercentilový rozsah světlosti spektra s mezní frekvencí 3 kHz
154	roll85 (ir)	[Hz]	interpercentilový rozsah parametru rolloff při 85 %

155	roll95 (ir)	[Hz]	interpercentilový rozsah parametru rolloff při 95 %
156	bright1k (Q1)	[%]	1. kvartil světlosti spektra s mezní frekvencí 1 kHz
157	bright1_5k (Q1)	[%]	1. kvartil světlosti spektra s mezní frekvencí 1,5 kHz
158	bright3k (Q1)	[%]	1. kvartil světlosti spektra s mezní frekvencí 3 kHz
159	roll85 (Q1)	[Hz]	1. kvartil parametru rolloff při 85 %
160	roll95 (Q1)	[Hz]	1. kvartil parametru rolloff při 95 %
161	bright1k (Q3)	[%]	3. kvartil světlosti spektra s mezní frekvencí 1 kHz
162	bright1_5k (Q3)	[%]	3. kvartil světlosti spektra s mezní frekvencí 1,5 kHz
163	bright3k (Q3)	[%]	3. kvartil světlosti spektra s mezní frekvencí 3 kHz
164	roll85 (Q3)	[Hz]	3. kvartil parametru rolloff při 85 %
165	roll95 (Q3)	[Hz]	3. kvartil parametru rolloff při 95 %
166	bright1k (IQR)	[%]	interkvartilový rozsah světlosti spektra s mezní frekvencí 1 kHz
167	bright1_5k (IQR)	[%]	interkvartilový rozsah světlosti spektra s mezní frekvencí 1,5 kHz
168	bright3k (IQR)	[%]	interkvartilový rozsah světlosti spektra s mezní frekvencí 3 kHz
169	roll85 (IQR)	[Hz]	interkvartilový rozsah parametru rolloff při 85 %
170	roll95 (IQR)	[Hz]	interkvartilový rozsah parametru rolloff při 95 %
<hr/>			
171	ctr (mean)	[Hz]	střední hodnota spektrálního centroidu
172	ent (mean)	[-]	střední hodnota spektrální entropie
173	krt (mean)	[-]	střední hodnota špičatosti spektrálního rozložení
174	flt (mean)	[-]	střední hodnota plochosti spektrálního rozložení
175	skw (mean)	[-]	střední hodnota šikmosti spektrálního rozložení
176	sprd (mean)	[-]	střední hodnota spektrální směrodatné odchylky

177	ctr (std)	[Hz]	směrodatná odchylka spektrálního centroidu
178	ent (std)	[-]	směrodatná odchylka spektrální entropie
179	krt (std)	[-]	směrodatná odchylka špičatosti spektrálního rozložení
180	flt (std)	[-]	směrodatná odchylka plochosti spektrálního rozložení
181	skw (std)	[-]	směrodatná odchylka šikmosti spektrálního rozložení
182	sprd (std)	[-]	směrodatná odchylka spektrální směrodatné odchylky
183	ctr (var)	[Hz]	rozptyl spektrálního centroidu
184	ent (var)	[-]	rozptyl spektrální entropie
185	krt (var)	[-]	rozptyl špičatosti spektrálního rozložení
186	flt (var)	[-]	rozptyl plochosti spektrálního rozložení
187	skw (var)	[-]	rozptyl šikmosti spektrálního rozložení
188	sprd (var)	[-]	rozptyl spektrální směrodatné odchylky
189	ctr (median)	[Hz]	medián spektrálního centroidu
190	ent (median)	[-]	medián spektrální entropie
191	krt (median)	[-]	medián špičatosti spektrálního rozložení
192	flt (median)	[-]	medián plochosti spektrálního rozložení
193	skw (median)	[-]	medián šikmosti spektrálního rozložení
194	sprd (median)	[-]	medián spektrální směrodatné odchylky
195	ctr (1p)	[Hz]	1. percentil spektrálního centroidu
196	ent (1p)	[-]	1. percentil spektrální entropie
197	krt (1p)	[-]	1. percentil špičatosti spektrálního rozložení
198	flt (1p)	[-]	1. percentil plochosti spektrálního rozložení
199	skw (1p)	[-]	1. percentil šikmosti spektrálního rozložení
200	sprd (1p)	[-]	1. percentil spektrální směrodatné odchylky
201	ctr (99p)	[Hz]	99. percentil spektrálního centroidu
202	ent (99p)	[-]	99. percentil spektrální entropie

203	krt (99p)	[-]	99. percentil špičatosti spektrálního rozložení
204	flt (99p)	[-]	99. percentil plochosti spektrálního rozložení
205	skw (99p)	[-]	99. percentil šikmosti spektrálního rozložení
206	sprd (99p)	[-]	99. percentil spektrální směrodatné odchylky
207	ctr (ir)	[Hz]	interpercentilový rozsah spektrálního centroidu
208	ent (ir)	[-]	interpercentilový rozsah spektrální entropie
209	krt (ir)	[-]	interpercentilový rozsah špičatosti spektrálního rozložení
210	flt (ir)	[-]	interpercentilový rozsah plochosti spektrálního rozložení
211	skw (ir)	[-]	interpercentilový rozsah šikmosti spektrálního rozložení
212	sprd (ir)	[-]	interpercentilový rozsah spektrální směrodatné odchylky
213	ctr (Q1)	[Hz]	1. kvartil spektrálního centroidu
214	ent (Q1)	[-]	1. kvartil spektrální entropie
215	krt (Q1)	[-]	1. kvartil špičatosti spektrálního rozložení
216	flt (Q1)	[-]	1. kvartil plochosti spektrálního rozložení
217	skw (Q1)	[-]	1. kvartil šikmosti spektrálního rozložení
218	sprd (Q1)	[-]	1. kvartil spektrálního centroidu
219	ctr (Q3)	[Hz]	3. kvartil spektrální směrodatné odchylky
220	ent (Q3)	[-]	3. kvartil spektrální entropie
221	krt (Q3)	[-]	3. kvartil špičatosti spektrálního rozložení
222	flt (Q3)	[-]	3. kvartil plochosti spektrálního rozložení
223	skw (Q3)	[-]	3. kvartil šikmosti spektrálního rozložení

224	sprd (Q3)	[-]	3. kvartil spektrální směrodatné odchyly
225	ctr (IQR)	[Hz]	interkvartilový rozsah spektrálního centroidu
226	ent (IQR)	[-]	interkvartilový rozsah spektrální entropie
227	krt (IQR)	[-]	interkvartilový rozsah špičatosti spektrálního rozložení
228	flt (IQR)	[-]	interkvartilový rozsah plochosti spektrálního rozložení
229	skw (IQR)	[-]	interkvartilový rozsah šikmosti spektrálního rozložení
230	sprd (IQR)	[-]	interkvartilový rozsah spektrální směrodatné odchyly
<hr/>			
231	atkleap (mean)	[-]	střední hodnota velikosti náběhu
232	atkslope (mean)	[s]	střední hodnota sklonu náběhu
233	atctime (mean)	[s]	střední hodnota doby náběhu
234	dcrslope (mean)	[s]	střední hodnota sklonu poklesu
235	duration (mean)	[s]	střední hodnota doby trvání
236	atkleap (std)	[-]	směrodatná odchylyka velikosti náběhu
237	atkslope (std)	[s]	směrodatná odchylyka sklonu náběhu
238	atctime (std)	[s]	směrodatná odchylyka doby náběhu
239	dcrslope (std)	[s]	směrodatná odchylyka sklonu poklesu
240	duration (std)	[s]	směrodatná odchylyka doby trvání
241	atkleap (var)	[-]	rozptyl velikosti náběhu
242	atkslope (var)	[s]	rozptyl sklonu náběhu
243	atctime (var)	[s]	rozptyl doby náběhu
244	dcrslope (var)	[s]	rozptyl sklonu poklesu
245	duration (var)	[s]	rozptyl doby trvání
246	atkleap (median)	[-]	medián velikosti náběhu
247	atkslope (median)	[s]	medián sklonu náběhu
248	atctime (median)	[s]	medián doby náběhu
249	dcrslope (median)	[s]	medián sklonu poklesu
250	duration (median)	[s]	medián doby trvání
251	atkleap (1p)	[-]	1. percentil velikosti náběhu
252	atkslope (1p)	[s]	1. percentil sklonu náběhu
253	atctime (1p)	[s]	1. percentil doby náběhu
254	dcrslope (1p)	[s]	1. percentil sklonu poklesu

255	duration (1p)	[s]	1. percentil doby trvání
256	atkleap (99p)	[-]	99. percentil velikosti náběhu
257	atkslope (99p)	[s]	99. percentil sklonu náběhu
258	atctime (99p)	[s]	99. percentil doby náběhu
259	dcerslope (99p)	[s]	99. percentil sklonu poklesu
260	duration (99p)	[s]	99. percentil doby trvání
261	atkleap (ir)	[-]	interpercentilový rozsah velikosti náběhu
262	atkslope (ir)	[s]	interpercentilový rozsah sklonu náběhu
263	atctime (ir)	[s]	interpercentilový rozsah doby náběhu
264	dcerslope (ir)	[s]	interpercentilový rozsah sklonu poklesu
265	duration (ir)	[s]	interpercentilový rozsah doby trvání
266	atkleap (Q1)	[-]	1. kvartil velikosti náběhu
267	atkslope (Q1)	[s]	1. kvartil sklonu náběhu
268	atctime (Q1)	[s]	1. kvartil doby náběhu
269	dcerslope (Q1)	[s]	1. kvartil sklonu poklesu
270	duration (Q1)	[s]	1. kvartil doby trvání
271	atkleap (Q3)	[-]	3. kvartil velikosti náběhu
272	atkslope (Q3)	[s]	3. kvartil sklonu náběhu
273	atctime (Q3)	[s]	3. kvartil doby náběhu
274	dcerslope (Q3)	[s]	3. kvartil sklonu poklesu
275	duration (Q3)	[s]	3. kvartil doby trvání
276	atkleap (IQR)	[-]	interkvartilový rozsah velikosti náběhu
277	atkslope (IQR)	[s]	interkvartilový rozsah sklonu náběhu
278	atctime (IQR)	[s]	interkvartilový rozsah doby náběhu
279	dcerslope (IQR)	[s]	interkvartilový rozsah sklonu poklesu
280	duration (IQR)	[s]	interkvartilový rozsah doby trvání
<hr/>			
281	rgl (mean)	[-]	střední hodnota drsnosti
282	zcr (mean)	[-]	střední hodnota počtu průchodů nulovou úrovní
283	rgl (std)	[-]	směrodatná odchylka drsnosti
284	zcr (std)	[-]	směrodatná odchylka počtu průchodů nulovou úrovní
285	rgl (var)	[-]	rozptyl drsnosti
286	zcr (var)	[-]	rozptyl počtu průchodů nulovou úrovní
287	rgl (median)	[-]	medián drsnosti
288	zcr (median)	[-]	medián počtu průchodů nulovou úrovní
289	rgl (1p)	[-]	1. percentil drsnosti

290	zcr (1p)	[-]	1. percentil počtu průchodů nulovou úrovní
291	rgl (99p)	[-]	99. percentil drsnosti
292	zcr (99p)	[-]	99. percentil počtu průchodů nulovou úrovní
293	rgl (ir)	[-]	interpercentilový rozsah drsnosti
294	zcr (ir)	[-]	interpercentilový rozsah počtu průchodů nulovou úrovní
295	rgl (Q1)	[-]	1. kvartil drsnosti
296	zcr (Q1)	[-]	1. kvartil počtu průchodů nulovou úrovní
297	rgl (Q3)	[-]	3. kvartil drsnosti
298	zcr (Q3)	[-]	3. kvartil počtu průchodů nulovou úrovní
299	rgl (IQR)	[-]	interkvartilový rozsah drsnosti
300	zcr (IQR)	[-]	interkvartilový rozsah počtu průchodů nulovou úrovní
