

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE

Brno, 2017

Šárka Zemánková



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY

A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

VÝZKUM DYNAMICKÝCH PARAMETRŮ POROVNÁVAJÍCÍ ZVUKOVÉ NAHRÁVKY

RESEARCH OF DYNAMICS FEATURES COMPARING AUDIO RECORDS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Šárka Zemánková

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Tomáš Kiska

BRNO 2017



Bakalářská práce

bakalářský studijní obor **Audio inženýrství**
Ústav telekomunikací

Studentka: Šárka Zemánková

ID: 174479

Ročník: 3

Akademický rok: 2016/17

NÁZEV TÉMATU:

Výzkum dynamických parametrů porovnávací zvukové nahrávky

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 dynamiky. Rovněž budou za pomoci statistické analýzy vybrány ty parametry, které z hlediska dynamiky 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 analýzou parametrů, souvisejících s dynamikou zvukové nahrávky. Je zde stručně popsána historie zpracování nahrávek v analogové a digitální podobě i průběh zpracování zvukového signálu v současnosti. Následující kapitola obsahuje výběr nejvhodnějších parametrů pro popis zvukové nahrávky, především těch, které popisují dynamiku. Tato práce dále charakterizuje metody, využívané při podobných výzkumech ve světě. Také je zde navržen systém pro výpočet 43 dynamických parametrů a nastíněny možnosti jejich analýzy. Porovnávalo bylo 35 různých interpretací jednoho hudebního díla. Nakonec došlo k zaznamenání vypočítaných parametrů do bodových diagramů a k jejich vyhodnocení pomocí vizuální shlukové analýzy.

KLÍČOVÁ SLOVA

digitální zpracování signálu, zvukový signál, zvukové parametry, dynamika, shluková analýza

ABSTRACT

This work deals with the analysis of parameters related to the dynamics of sound recordings. It contains a brief description of the history of sound processing in analogue and digital form and the process of audio signal processing nowadays. The following chapter includes selection of the most suitable parameters for describing an audio recording, especially those describing the dynamics. This work further characterizes the methods used in similar researches in the world. There is also a system designed to calculate 43 dynamic parameters and the possibilities of their analysis are outlined as well. 35 different interpretations of one musical work were compared. Finally, the calculated parameters were drawn into scatter plots and evaluated using visual cluster analysis.

KEYWORDS

digital signal processing, audio signal, audio features, dynamics, cluster analysis

ZEMÁNKOVÁ, Šárka *Výzkum dynamických parametrů porovávající zvukové nahrávky*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2017. 49 s. Vedoucí práce byl Ing. Tomáš Kiska

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Výzkum dynamických parametrů porovnávací zvukové nahrávky“ 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

Ráda bych poděkovala vedoucímu bakalářské práce panu Ing. Tomáši 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



OP Výzkum a vývoj
pro inovace

OBSAH

Úvod	11
1 Úvod do zpracování hudebního signálu	12
1.1 Historický vývoj	12
1.1.1 Analogový signál	12
1.1.2 Digitální signál	12
1.2 Hudební teorie vs. číslicové zpracování signálu	13
1.3 Metody hodnocení hudebních nahrávek	14
2 Předzpracování hudebního signálu	18
3 Parametrizace hudebního signálu	19
3.1 Všeobecné parametry	19
3.2 Parametry založené na dynamice	19
4 Statistické zpracování parametrů	23
5 Nové metody v oblasti analýzy hudebních nahrávek	24
5.1 Charakteristika nejvyužívanějších metod	24
5.2 Nejvýznamnější výzkumy	25
6 Návrh vyhodnocovacího systému	27
7 Evaluace systému	29
7.1 Volba parametrů	29
7.2 Analýza vypočítaných parametrů	30
8 Vyhodnocení	32
9 Závěr	37
Literatura	39
Seznam symbolů, veličin a zkratk	43
Seznam příloh	44
A Obsah přiloženého CD	45
B Seznam nahrávek	46

SEZNAM OBRÁZKŮ

1.1	Fonograf [16]	13
1.2	Průběhy signálů	15
6.1	Vyhodnocovací systém	27
8.1	Závislost parametru E (rms) na roce vzniku	32
8.2	Závislost parametru Y (mean) na roce vzniku	33
8.3	Závislost parametru Harmonics (mean) na roce vzniku	34
8.4	Závislost parametru TKEO (mean) na roce vzniku	34
8.5	Závislost parametru shimmer (local) na roce vzniku	35

SEZNAM TABULEK

B.1 Seznam nahrávek	46
C.1 Seznam parametrů	48

ÚVOD

Tato práce se zabývá analýzou dynamických parametrů hudebních nahrávek a jejich vzájemným porovnáním. Díky tomu je následně možné u různých interpretací jednoho díla hledat souvislosti a zařadit danou verzi do časové oblasti vzniku nebo k místu původu.

Sledování dynamiky hudební nahrávky je také součástí *Audio-to-Score Alignmentu*, neboli výzkumu, který se věnuje přiřazování hudebního signálu k notovému zápisu. Při něm je, mimo dynamiky, sledována též výška a délka tónu, tempo skladby a další parametry.

Audio-to-Score Alignment pak spadá do ještě širší vědecké oblasti, kterou je *Music Information Retrieval*, tj. získávání informací z hudby. Tyto informace je posléze možno využít pro různé aplikace, analyzující hudební skladby. Jsou to např. programy sledující průběh živě hrané hudby v reálném čase, které dokáží ve správný časový okamžik spustit reprodukováný doprovod, automaticky otáčet stránky, případně také identifikovat chyby, kterých se hudebník dopustil. Dále se Music Information Retrieval využívá ve v dnešní době populárních mobilních aplikacích, které jsou schopny analyzovat právě hraný hudební signál a identifikovat název skladby a autora, případně vyhledat také text písně a posouvat jej podle průběhu skladby.

Cílem této práce je shrnout zpracování hudebních signálů z hlediska historického i novodobého a srovnat fyzikální a umělecký přístup k hudbě. Následuje sumarizace parametrů pro popis zvukových nahrávek, zejména z hlediska dynamiky. Dalším úkolem je charakterizace metod, využívaných k získávání informací z hudby ve světě. Následně byl v rámci této práce vytvořen návrh a implementace vyhodnocovacího systému pro výpočet parametrů hodnotících dynamiku nahrávek. Finálně došlo k jejich porovnání a výběru parametrů vhodných pro diferenciaci interpretací.

1 ÚVOD DO ZPRACOVÁNÍ HUDEBNÍHO SIGNÁLU

1.1 Historický vývoj

1.1.1 Analogový signál

Už od druhé poloviny 19. století se lidé pokoušeli o zaznamenávání a reprodukci zvuku. Mezi první zařízení, které sloužilo k tomuto účelu, patřil *fonograf* sestavený vynálezcem Thomasem Edisonem a patentovaný v roce 1878 [24]. Zvukový signál byl zaznamenáván přes vibrace kovové roury na voskové, později celuloidové válečky. Tato technika byla o několik let později nahrazena záznamem na gramofonovou desku. Nahrávání probíhalo čistě mechanicky a zvuk proto nemohl být nijak dodatečně upravován. Hlavní nevýhodou této techniky byla omezenost frekvenčního pásma nahrávaného signálu, zejména absence nižších frekvencí.

Již na začátku 20. století bylo mechanické zaznamenávání nahrazeno záznamem magnetickým, ke kterému se již využíval elektrický proud [30]. Zvukové vibrace dopadající na membránu mikrofону byly přeměňovány na střídavý elektrický proud, který působil na zmagnetizovaný pásek a vytvářel tak na něm záznam zvukového signálu. Tato metoda má oproti mechanickému záznamu výraznou výhodu – záznam lze z pásky po nahrání i vymazat a opakovaně nahrát.

Od 50. let 20. století už bylo při nahrávání a reprodukci možné využít zařízení pro úpravu hudby - zvukový signál mohl být zpracováván korekčními zesilovači, které zdůrazňují/potlačují určitá frekvenční pásma, tzn. hloubky, středy a výšky, nebo zařízeními pro úpravu dynamiky (kompresory, limity atd.). Byly také vynalezeny první přístroje pro vytváření zkreslení, echa, modulací (vibrato, tremolo) a dalších efektů.

1.1.2 Digitální signál

V průběhu 70. let 20. století došlo k zásadní inovaci – převedení klasického analogového záznamu do digitálního. Během tohoto procesu je vstupní analogový signál vzorkován (dochází k odebrání vzorků signálu po časovém úseku o délce vzorkovací periody) a následně kvantován (hodnoty vzorků jsou zaokrouhleny na vzorkovací hladiny). Výsledkem je signál číslicový (digitální). Ten pak lze oproti původnímu analogovému signálu zpracovávat mnohem jednodušeji využitím nepřeberného množství naprogramovaných digitálních efektů, jako jsou reverb, echo, delay, kompresor, gate, distortion a další. Jeho výhody tkví také ve větším odstupu signálu od šumu, omezení nežádoucího zkreslení signálu, snadnějším přenosu signálu, možnosti dosažení



Obr. 1.1: Fonograf [16]

většího dynamického rozsahu, a daleko širších možnostech při upravování a modifikování signálu - snadnější stříhání, kopírování, mazání jednotlivých úseků a podobně.

Otázka kvality analogového a digitálního audio signálu je poměrně diskutabilní. Nedokonalost digitálního signálu spočívá v zaokrouhlování jednotlivých hodnot na kvantovací hladiny, čímž je hudební nahrávka znehodnocena. Bylo však zjištěno, že pokud tento proces probíhá na převodníku o dostatečném počtu bitů, jsou změny téměř nepostřehnutelné pro lidské ucho. Za dostačující se běžně považuje 16 bitů. S rostoucím počtem bitů se zvyšuje dynamický rozsah a tím pádem také odstup signálu od šumu. Výhoda digitálního signálu naopak tkví v lepší odolnosti vůči šumu, která opět závisí na počtu bitů signálu [32].

1.2 Hudební teorie vs. číslicové zpracování signálu

Hudební zvuk, neboli tón, je ve skutečnosti mechanické vlnění s periodickým průběhem. Jednoduché tóny mají harmonický průběh, většina hudebních nástrojů však produkuje složené tóny, které obsahují i vyšší harmonické složky.

Podle hudební teorie posuzujeme tóny podle čtyř základních parametrů: hlasitost, výška, barva a délka.

Mezi hudební teorií a digitálním signálem můžeme nalézt určitou analogii.

- Hlasitost tónu je dána intenzitou signálu.
- Výška tónu je reprezentována pomocí základní frekvence signálu.
- Barva poté závisí na počtu a amplitudách vyšších harmonických složek, tzv. *aliquót*.

Frekvence vyšších harmonických složek však nejsou náhodné. Existují dva základní přístupy, kterými se tyto frekvence určují. Podle pythagorejského ladění jsou to vždy celočíselné násobky fundamentu (první harmonická složka). Tato metoda je však poměrně nepřesná, jednotlivé půltóny podle ní nejsou stejně velké a na 7 oktávách vzniká rozdíl přibližně o čtvrtinu tónu. Modernějším přístupem je temperované ladění, podle něhož je oktáva rozdělena na 12 stejných půltónů, jejichž frekvence je vždy $\sqrt{2}$ krát vyšší než frekvence předcházejícího půltónu [12].

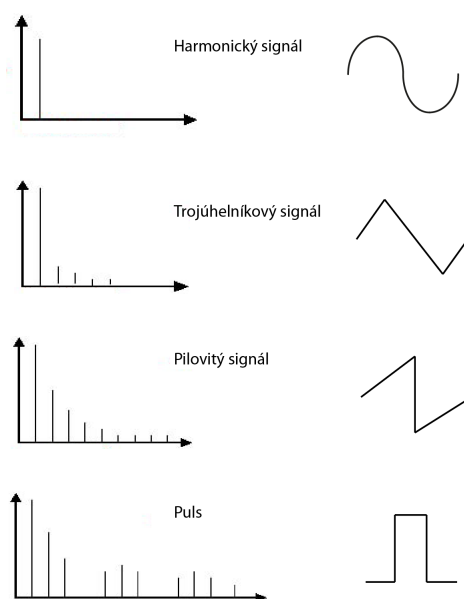
Různé hudební nástroje obsahují různý počet vyšších harmonických složek a tím se také liší průběh jejich signálu a spektrum frekvencí:

- K harmonickému signálu má nejbližší např. flétna, jelikož obsahuje pouze slabé aliquóty.
- Trojúhelníkový signál obsahuje slabé liché harmonické složky, hodnoty sudých jsou nulové. Vytváří jej např. klarinet nebo saxofon.
- Obdélníkový signál je tvořen také pouze lichými složkami, které jsou však silnější, a vzniká při hře na hoboj nebo anglický roh.
- Pilovitý signál, obsahující sudé i liché harmonické složky, generují především žestové nástroje.

1.3 Metody hodnocení hudebních nahrávek

Z fyzikálního hlediska se při hodnocení hudebních nahrávek posuzuje především způsob jejich zpracování. Dřívější techniky zpracování digitálního signálu byly založeny na operacích v časové doméně, jako je konvoluce, korelace, skalární součin a průměrování signálu [34]. Nedokonalost těchto technik tkví v jejich neschopnosti vyjádřit zastoupení jednotlivých frekvenčních složek signálu. Pomocí Fourierovy analýzy [29] lze naopak zjistit celkový frekvenční obsah signálu, ne však jeho proměnlivost v čase. Z toho důvodu byla zavedena časově-frekvenční analýza signálu, která je vhodná především ke zpracování nestacionárních signálů (vč. zvukových).

U nepravidelných signálů se také využívají pravděpodobnostní statistiky, které mají za cíl předvídat hodnoty signálu v jednotlivých časových okamžicích.



Obr. 1.2: Průběhy signálů

V rámci hudební teorie se u nahrávek hodnotí estetické parametry, jako je tempo, metrum a dynamika.

Tempo, neboli rychlost, kterou po sobě následují základní doby skladby, může být interpretovi zadáno počtem úderů za minutu (bpm - beats per minute) nebo tempovým označením (adagio, andante, moderato, allegro, presto). Jednotlivá tempová označení mají dānu svou hodnotu bpm, hudebník ji však nemusí striktně dodržovat. Celková délka skladby je proto relativní a u různých interpretů se může značně lišit. Rozmezí minimální a maximální délky skladby, které na posluchače působí kladným dojmem, se nazývá rozmezí estetické rezonance. Celkový průběh skladby je z hlediska tempa ve většině případů nepravidelný. V rámci skladby často dochází ke zpomalování či zrychlování určitých úseků, nebo pouze k jednomu z těchto jevů. Toto tempové kolísání bývá doprovázeno změnou dynamiky. Krátkodobé odchylky od původního tempa dělíme na dva druhy - záměrné a nahodilé. Záměrné odchylky mohou být zadány skladatelem a do partitury jsou zapsány pomocnými označeními (ritardando, accelerando, rubato), případně je podle vlastního citu vědomě vytváří sám interpret. Pokud je změna tempa způsobena technickou nedokonalostí interpreta a nevznikla záměrem skladatele ani hudebníka, označuje se jako nahodilá odchylka a rozpětí, v němž se tyto odchylky pohybují je tzv. pásmo nahodilých odchylek. Každá hudební interpretace tedy podléhá určitým tempovým výkyvům. K jejich zhodnocení se používá míra rovnoměrnosti tempa. Ta se stanovuje jako podíl ko-

eficientu rovnoměrnosti (průměrná odchylka od průměrné délky taktu) a průměrné délky taktu a udává se v procentech.

Dalším parametrem, který se u hudebních nahrávek využívá je tzv. hybnost hudebního proudu. Podle [13] je to „konkrétní pohyb zvukového proudu se všemi podrobnostmi, týkajícími se absolutního trvání zvuků a jejich vzájemných časových vztahů“. Průměrná hybnost se vyjadřuje podíl počtu zvukových impulsů a reálného času hudebního úseku.

Přechodem mezi tempem a dynamikou je metrum. Hudební metrum znamená střídání těžkých (přízvučných) a lehkých (nepřízvučných) dob taktu v určitém pořadí. Je úzce spjato s tempem a na jeho základě rozlišujeme tři druhy tempa: tempo metrizované, tempo s libovolnou vazbou hudebního metra a tempo nemetrizované. Metrizované tempo se dále dělí na čtyři typy:

- základní tempo - charakterizováno průměrnou rychlostí tempa, není brán zřetel na záměrné ani nahodilé tempové odchylky
- pevné tempo - vyjádřeno časovým pohybem skladby, ve kterém se nevyskytují žádné tempové odchylky
- oscilující tempo - tempo, ve kterém dochází k nahodilým odchylkám pohybu jícím se kolem základního tempa
- proměnné tempo - tempo, ve kterém se vyskytují záměrné odchylky

Pokud je skladbě zadáno tempo s libovolnou vazbou hudebního metra, znamená to, že není metrum zadáno skladatelem, ale je ponecháno na vůli hudebníka. Partitura však může obsahovat příležitostné akcenty.

Nemetrizované tempo, neboli tempo bez vazby na hudební metrum, je pro hudebníky i posluchače nepřírozené a v praxi se příliš často nevyskytuje.

Hudební dynamika udává sílu vloženou do zpěvu nebo hry na hudební nástroj. V partituře je zadána pomocí dynamických značek (pianissimo, piano, mezzo-forte, forte, fortissimo) pro celou skladbu, příp. úsek skladby, krátkodobých značek (crescendo, decrescendo) pro jeden nebo několik taktů nebo akcentů na jednotlivých tónech. Tyto údaje však nejsou přesné, ale jen velmi přibližné. Klasická notace ponechává prostor pro umělecký cit hudebníka, v praxi proto můžeme najít značné rozdíly mezi různými interpretacemi stejného díla. Při jejich posuzování hraje důležitou roli psychoakustika a rozlišovací schopnosti lidského ucha. Podle výzkumu [26] dokáže většina posluchačů identifikovat crescendo (zesilování hlasitosti) pokud při něm dojde k zesílení o více než 2 dB. Decrescendo (zeslabování hlasitosti) nebylo ve většině případů identifikováno vůbec, jelikož odeznívání tónu působí na posluchače jako přirozený jev. Rozlišovací schopnost lidského ucha závisí z velké části na věku posluchače, neboť s narůstajícím věkem dochází ke snížení citlivosti sluchu a ztrátě slyšení na vysokých frekvencích (nad 16 kHz) [25].

S dynamikou skladby také úzce souvisí barva zvuku, tzn. počet vyšších harmonických složek signálu. Většina nástrojů hrajících forte generuje mnohem více alikvót než při hře piano. Tento jev souvisí s rezonancí hudebního nástroje.

Na základě studií [26] lze také tvrdit, že existuje spojitost mezi vnímáním hudební dynamiky, tempa, délky a výšky tónů. Postupné narůstání výšky tónů u více než poloviny posluchačů vyvolává dojem nárůstu dynamiky, naopak klesání výšky působí jako snižování dynamiky. Podobně je tomu i u změn tempa - pokud dochází ke zrychlování tempa, působí tato gradace na posluchače také jako nárůst dynamiky, zpomalování vytváří naopak dojem zeslabení dynamiky. Také délka tónu ovlivňuje lidské vnímání, pokud jsou za sebou zahrány tóny o stejné intenzitě, lidské ucho vnímá delší tón jako hlasitější [25]. K podobnému závěru dospěli také britští vědci v rámci psychologických výzkumů [33].

2 PŘEDZPRACOVÁNÍ HUDEBNÍHO SIGNÁLU

Před samotným zpracováním hudebního signálu musí být provedeno několik úprav.

Na začátku je třeba odstranit rušivé části signálu, které vznikají při jeho záznamu, přenosu, převodu do digitální podoby a zpracování. Patří sem zejména šum, v případě živých vystoupení to mohou být i nehudební zvuky (kašel, šramot apod.).

Při odstraňování šumu je nejprve třeba zjistit šumový profil, neboli hladinu hlasitosti jednotlivých frekvencí, obsažených v šumu. Tyto frekvence jsou následně v určité míře vyfiltrovány. Čím je tato míra vyšší, tím více šumu je odstraněno, může to však mít negativní vliv i na samotnou hudební stopu.

Při analýze nepravidelných signálů, měnících se v čase (tzn. i hudebních signálů) se používá krátkodobá Fourierova transformace [6, 31]. Pomocí ní lze dlouhý signál rozdělit na krátké úseky tzv. váhovacími okny, a tyto úseky poté zpracovávat odděleně. Váhovací okna nabývají nulové hodnoty mimo určitou oblast, takže signál vynásobený tímto oknem nabývá mimo tuto oblast také nulové hodnoty [25]. V praxi existuje mnoho druhů váhovacích oken, která se liší svými vlastnostmi - tvarem, spektrem, šířkou hlavního laloku nebo ekvivalentní šumovou šířkou. Je to např. Dirichletovo okno (pravoúhlé), Hanningovo okno (speciální druh sinového okna), Barletovo okno (trojúhelníkové) nebo Blackmanovo okno. Při násobení signálu váhovacími okny někdy může docházet k tzv. prosakování spektra (spectral leakage). Tento nežádoucí jev je způsoben velkými postranními laloky ve spektrech oken, jejichž důsledkem po vynásobení oknem vznikají ve výsledném spektru nežádoucí frekvenční složky. Míra prosakování je závislá na typu použitého okna. Jelikož má každý typ okna také jinou šířku, závisí na něm i rozlišovací schopnost spektrální analýzy. Každý typ okna má navíc svůj vlastní širokopásmový šum, který je přímo úměrný jeho ekvivalentní šumové šířce [6].

3 PARAMETRIZACE HUDEBNÍHO SIGNÁLU

V rámci dalšího zpracování je signál parametrizován, čímž jsou vyhledávána nejdůležitější kritéria pro posuzování hudební nahrávky.

3.1 Všeobecné parametry

- Tempo: počet úderů za minutu (bpm), míra rovnoměrnosti tempa, hybnost hudebního proudu
- Rytmika: počet dob v taktu, hudební metrum
- Tonalita: výška tónu (frekvence), barva tónu (počet vyšších harmonických složek), identifikace hudebního nástroje
- Harmonie: tónina, hudební interval (poměr frekvencí)
- Textura: vztah melodie/melodií a doprovodu

3.2 Parametry založené na dynamice

Při hodnocení dynamických vlastností nahrávek je potřeba sledovat zejména průběh energie hudebního signálu v čase, tedy především jeho minimální a maximální hodnotu, variabilitu a gradaci.

- Obálka ADSR - charakterizuje průběh hlasitosti tónu v čase
 1. attack - doba náběhu tónu na maximum
 2. decay - doba úpadku na stabilní hlasitost tónu
 3. sustain - doba trvání konstantní hlasitosti tónu
 4. release - doba doznívání
- Minimální hodnota energie hudebního signálu $E(\min)$
- Maximální hodnota energie hudebního signálu $E(\max)$
- Dynamický rozsah - rozdíl mezi maximální a minimální energií signálu
- Střední hodnota energie signálu \bar{E}

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

- Střední hodnota hlasitosti jednoho vzorku signálu

$$\bar{E} = \frac{1}{N} \sum_{n=0}^{N-1} E(n). \quad (3.2)$$

- Směrodatná odchylka σ - míra proměnlivosti energie kolem své střední hodnoty

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (E_i - \bar{E})^2}. \quad (3.3)$$

- Efektivní hodnota energie (RMS)

$$\text{RMS}_t = \sqrt{\frac{1}{K} \sum_{k=t*K}^{(t+1)*K-1} s(k)^2}, \quad (3.4)$$

kde $s(k)$ označuje amplitudu k -tého vzorku a K označuje velikost rámce t (počet vzorků v každém rámci) [17].

- Počet průchodů nulovou hodnotou – Zero Crossing Rate (ZCR) - měří, kolikrát hodnota amplitudy změni znaménko během rámce t . Při hodnocení zvukových nahrávek se používá např. k detekci perkusivních zvuků [17]

$$\text{ZCR}_t = \frac{1}{2} \sum_{k=t*K}^{(t+1)*K-1} |\text{sgn}(s(k)) - \text{sgn}(s(k+1))|. \quad (3.5)$$

- Rozdíl energií ΔE - vyjadřuje maximální a minimální rozdíl hodnoty energie
- Modus energie - nejčtenější hodnota energie
- Medián energie - hodnota, která dělí vzestupně seřazené hodnoty energií na dvě stejné poloviny
- Modulační energie - rozložení energie ve výkonovém spektru
- Shimmer - lokální kolísání amplitudy signálu. Existuje několik druhů shimmeru [8]:
 - shimmer (dB) - vyjadřuje se jako proměnlivost špičkové hodnoty amplitudy v decibelech

$$\text{Shimmer (dB)} = \frac{1}{N-1} \sum_{i=1}^{N-1} |20 \log(A_{i+1}/A_i)|, \quad (3.6)$$

kde A_{i+1} a A_i označují špičkové amplitudy signálu v průběhu po sobě jdoucích period, N udává počet extrahovaných period základní frekvence [8]

- APQ3 (Three-Point Amplitude Perturbation Quotient) - průměrný absolutní rozdíl mezi amplitudou periody a průměrem amplitud sousedních period, podělený průměrnou hodnotou amplitudy.
- APQ5 (Five-Point Amplitude Perturbation Quotient) - průměrný absolutní rozdíl mezi amplitudou periody a průměrem amplitud čtyř sousedních period, podělený průměrnou hodnotou amplitudy.
- APQ11 (Eleven-Point Amplitude Perturbation Quotient) - průměrný absolutní rozdíl mezi amplitudou periody a průměrem amplitud deseti sousedních period, podělený průměrnou hodnotou amplitudy.
- shimmer (ddp) - průměrný absolutní rozdíl mezi následujícími rozdíly mezi amplitudami následujících period.

- Odstup signálu od šumu – Signal to Noise Ratio (SNR)
- Teager-Kaiserův energetický operátor (TKEO) - charakterizuje současně energii i frekvenci signálu, jelikož např. při hlasitější hře na hudební nástroj vzniká více vyšších harmonických složek
- Medián výkonové spektrální hustoty – Median of Power Spectral Density (MPSD) - definuje rozložení energie ve výkonovém spektru
- Low Short-Time Energy Ratio (LSTER) - umožňuje diferenciaci hudebního signálu, který vykazuje výrazné odchylky v energii na krátkých časových úsecích
- Spektrální tok – Spectral Flux (SF) - parametr ve frekvenční doméně, popisuje změnu ve výkonovém spektru mezi po sobě jdoucími rámci [17]

$$SF_t = \sum_{n=1}^N (D_t(n) - D_{t-1}(n))^2, \quad (3.7)$$

kde D_t označuje normalizované frekvenční rozdělení v rámci t [17].

- Poměr energetických pásem – Band Energy Ratio (BER) - udává poměr energií pásem o nízké a vysoké frekvenci, respektive udává, zda v signálu převažují nízké nebo vysoké kmitočty. [17] Tento parametr lze využít i při určování dynamiky, neboť hudební úseky o výraznější dynamice obsahují více vyšších harmonických složek, tzn. čím větší je hodnota BER, tím méně vyšších harmonických složek signál obsahuje

$$BER_t = \frac{\sum_{n=1}^{F-1} m_t(n)^2}{\sum_{n=F}^N m_t(n)^2}, \quad (3.8)$$

kde F označuje mezní frekvenci a $m_t(n)$ označuje rozsah frekvenční domény signálu v rámci t a frekvenčním pásmu n [17].

- Spektrální těžiště – Spectral Centroid (SC) - charakterizuje frekvenční oblast, ve které se soustřeďuje největší množství energie. Tento parametr se často nazývá „jas“ zvuku [17]. Zároveň je to první moment a geometrický střed frekvenčního rozdělení [18]

$$SC_t = \frac{\sum_{n=1}^N m_t(n)}{\sum_{n=1}^N m_t(n)}. \quad (3.9)$$

Podle [18] je lze také popsat vztahem

$$\mu_1 = \int x f(x) dx. \quad (3.10)$$

- Spectral spread - frekvenční rozptyl - popisuje druhý centrální moment frekvenčního rozdělení a tím i míru disperze tohoto rozdělení [18]

$$\sigma^2 = \mu_2 = \int (x - \mu_1)^2 f(x) dx. \quad (3.11)$$

- Skewness - šikmost - udává míru symetrie frekvenčního rozdělení a je jeho třetím centrálním momentem. Šikmost zcela symetrického rozdělení je rovna nule [18]

$$\mu_3 = \int (x - \mu_1)^3 f(x) dx. \quad (3.12)$$

- Kurtosis - špičatost - označuje čtvrtý centrální moment frekvenčního rozdělení [18].
- Flatness - plochost - je poměrem mezi geometrickou a aritmetickou střední hodnotou signálu a hodnotí množství špiček signálu [18].
- Entropy - entropie - hodnotí zda v signálu převládají špičky [18].

4 STATISTICKÉ ZPRACOVÁNÍ PARAMETRŮ

Po průchodu signálu systémem, který určuje hledané parametry hudebního signálu, jsou tyto parametry rozříděny a porovnávány. Výsledkem tohoto procesu mohou být tabulky a grafy, na základě kterých lze vzájemně srovnávat jednotlivé hudební nahrávky.

Pro porovnávání hudební nahrávky z hlediska dynamiky je nejpodstatnější sledování průběhu hodnot energie signálu v čase. Různé interpretace jedné hudební skladby se v tomto ohledu mohou značně lišit. Projevuje se to zejména na celkové hlasitosti celého hudebního díla, kterou lze nejlépe charakterizovat pomocí střední hodnoty signálu. Dále pak na maximální hlasitosti v rámci skladby, která je úzce spjatá s maximální hodnotou energie a dynamickým rozsahem. Podstatným parametrem pro rozpoznávání dané interpretace je také shimmer, který popisuje rychlost změny dynamiky (crescendo, decrescendo) v různých oblastech nahrávky. Pro různá hudební tělesa je totiž typické, že mají odlišný cit pro dramatičnost hudebního proudu a mění se u nich proto rychlost a míra gradace.

5 NOVÉ METODY V OBLASTI ANALÝZY HUDEBNÍCH NAHRÁVEK

Ve světě se problematikou sledování dynamiky signálu zabývají vědci v rámci rozsáhlého výzkumu, označovaného jako audio-to-score alignment. Cílem tohoto výzkumu je synchronizace hudební nahrávky s jejím notovým zápisem. Rozděluje se na dva typy - offline a online alignment. Offline přístup využívá hudební nahrávku jako celek a může tak detailně sledovat i následující stavy. Online verze pracuje v reálném čase, což je vhodné například pro aplikace zajišťující automatické otáčení stránek notového zápisu nebo aplikace ovládající reprodukováný doprovod sólisty při živém vystoupení [4].

Existují dva typy metod, které se využívají k analýze audio signálu. První z nich jsou statistické přístupy, vhodné spíše pro online alignment (Skrytý Markovův model, Bayesovská statistika), tou druhou tzv. Dynamické borcení časové osy – Dynamic Time Warping (DTW), typické pro offline alignment [20].

5.1 Charakteristika nejvyužívanějších metod

- Markovův Model – Markov Model (MM) - statistický model, který předpokládá, že budoucí stavy systému závisí pouze na aktuálním stavu, ne však na již známých předchozích stavech [9].
- Skrytý Markovův Model – Hidden Markov Model (HMM) - statistický model, jenž je používán, pokud jsou předchozí stavy neznámé, skryté. HMM je považován za nejjednodušší formu DBN. Zásadní nedokonalostí této metody je její náročnost na paměť (hodnoty parametrů narůstají exponenciálně) a čas [10, 15].
- Bayesovská statistika – Bayesian Statistics (BS) - podobně jako skrytý Markovův model se používá, pokud nejsou známy předešlé stavy systému. Je založena na tzv. „stupních víry“ (degrees of belief) [2].
- Dynamické Bayesovské sítě – Dynamic Bayesian Networks (DBN) - grafický model pro reprezentaci podmíněné nezávislosti mezi náhodnými veličinami v různých časových úsecích za předpokladu, že událost může mít za následek budoucí událost, ale nikoli naopak [10].
- DTW - algoritmus pro sjednocení dvou časových sekvencí s ohledem na časový posun a změny v rychlosti. Sekvence jsou vzájemně porovnány a je určena jejich míra podobnosti [23]. Tato metoda byla poprvé využita v práci Nicolý Oriá [28].

Výhody DTW tkví v dobré práci s paměťovými prostředky a práci s komplexními hudebními prvky, jako jsou vibrata, trioly, rychlé hudební sekvence a polyfonie. Problémem tohoto systému je však jeho neschopnost pracovat bez znalosti delších časových úseků a je proto vhodný pouze jako offline metoda [23].

Pro zajištění synchronizace hudebního signálu s notovým zápisem je potřeba vyhodnocovat množství parametrů signálu. Seznam těchto parametrů je uveden v kapitole 3.

5.2 Nejvýznamnější výzkumy

Výzkum španělských univerzit [4] je založen na využití DTW a rozkladu na nezáporné matice. Spektrální rozklad se využívá pro určení vzorků spektra každé kombinace souběžně znějících tónů MIDI partitury. DTW se poté používá k udání vztahu mezi zvukovým signálem a partiturou. Je zde využita online i offline metoda zpracování těchto hodnot. Oba tyto přístupy byly testovány na různých stupních polyfonie a porovnávány s dalšími referenčními metodami. Tento výzkum dosahuje vysoké míry přesnosti rozlišení souběžně znějících melodií.

Práce japonských vědců [21] se zabývá srovnáním hudebního zápisu a interpretace za pomoci Bayesovské metody. Hudební signál prezentuje jako Skrytý Markovův model. Vyhodnocováno bylo šedesát děl klasické hudby, s instrumentací zahrnující jak sólový klavír, tak celý orchestr. Po odstranění dozvuku ze signálu je opakovaně identifikován tón nástroje a vyrovnanost hlasitosti v každém bodě partitury. Tato metoda, na rozdíl od metod využívajících DTW, nevyžaduje pečlivé přizpůsobování modelových parametrů. Má však nedostatky v určování tónové délky.

Na základě výzkumu pařížského institutu TELECOM [15] lze k offline synchronizaci hudebního signálu a notového zápisu využít statistické modely. Skrytý stavový model je vytvořen z partitury, na základě čehož lze následně použít dvě různé funkce - chroma vektory a funkce detekce počátku. Tato metoda je vhodná pro určování hudebních skladeb a následné přiřazení k partituře. K experimentu byla využita databáze 94 populárních skladeb, které byly hodnoceny na základě výšek tónů a dynamiky. Informace o rytmicích nebyly pro přiřazování využity.

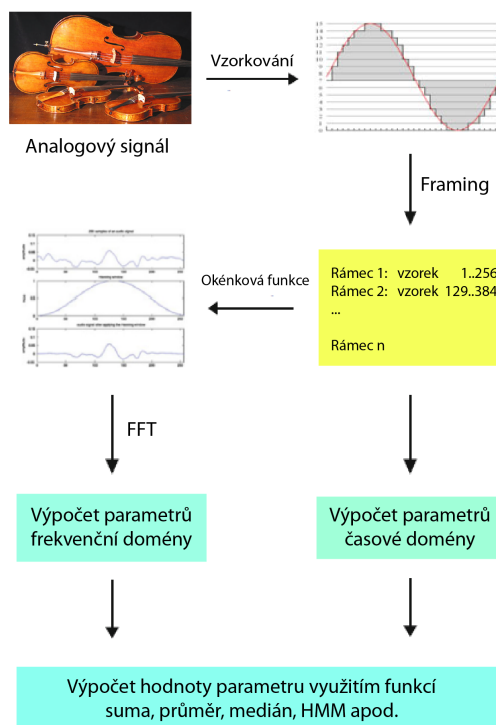
Výzkum Roberta Macrae a Simona Dixona [22] se zabývá WTW (Windowed Time Warping), metodou, jež je na rozdíl od DTW vhodná k online alignmentu. WTW rozděluje signály na posloupnost oken, která lze zpracovávat s vyšší přesností a v reálném čase. Výhodou oproti DTW je také nižší doba zpracování. Tento způsob zpracování je vhodný pro využití v zařízeních s méně výkonnými procesory (mobilní telefony) nebo pro aplikace zpracovávající velké množství dat (např. při identifikaci

cover verzí písní).

Práce rakouských vědců Bernharda Niedermayera a Gerharda Widmera [27] pojednává o odlišném přístupu k práci s partiturou. Namísto zkoumání jednotlivých rámců, během kterých se může vyskytovat více hudebních not, se zabývá zkoumáním odhadu počáteční doby samostatných not. K tomu využívá maticový rozklad. Nedostatkem této metody jsou chyby v časové linii, dochází k dislokaci not v rámci notového zápisu.

6 NÁVRH VYHODNOCOVACÍHO SYSTÉMU

Vstupem systému pro hodnocení dynamiky zvukových nahrávek je digitální signál, který byl vytvořen navzorkováním analogového signálu, tj. každému vzorku v pevně daném časovém intervalu byla přiřazena momentální amplituda signálu. Jelikož tyto vzorky trvají velice krátký časový úsek, je dalším krokem tzv. *framing* (vytváření rámců). Při něm je několik po sobě jdoucích vzorků spojeno a označeno jako rámec. Z těchto rámců již lze určit parametry v časové doméně. K určení parametrů ve frekvenční doméně je potřeba rámce dále zpracovat, obvykle to znamená použití okénkových funkcí (*windowing*). Tyto parametry však popisují pouze jednotlivé rámce, k charakterizování celého hudebního segmentu nebo skladby je nutné provést ještě dodatečný výpočet [17].



Obr. 6.1: Vyhodnocovací systém

Nízkoúrovňové parametry, ať už v časové nebo frekvenční doméně, jsou z reprezentace hudebního signálu určeny přímo. Mezi tyto parametry patří např. efektivní hodnota energie (RMS), amplitudová obálka, počet průchodů nulovou hodnotou (ZCR), spektrální těžiště nebo spektrální tok. Tyto parametry se vztahují k jednotlivým rámcům, získáváme je tedy ve formě vektorů. K získání jediné hodnoty

definující celý signál se provádí statistické sumarizace, jako je střední hodnota nebo medián.

Vysokoúrovňové parametry, mezi něž patří např. určení výšky tónu, tempa, instrumentace, akordů, melodie a rytmu, jsou pro výpočet mnohem komplikovanější. Některé z nich jsou již mnohem více abstraktní a k jejich určení je potřeba znalost lidského sluchu a způsobu vnímání zvuku.

7 EVALUACE SYSTÉMU

V této práci byly porovnávány nahrávky různých interpretací IV. věty *Con moto* Smyčcového kvarteta č. 1 od Leoše Janáčka. Seznam nahrávek je možné vidět v příloze v tabulce B.1. Z důvodu získání co nejpřesnějších parametrů byl zpracováván pouze jeden z hudebních motivů této věty (takty 37-50).

V rámci předzpracování zvukového signálu byla provedena jeho normalizace, resp. zesílení signálu na co nejvyšší hodnotu tak, aby nedocházelo k jeho zkreslení. To se provádí vynásobením všech vzorků signálu konstantou, závisící na úrovni hlasitosti, na kterou má být signál zesílen. Následně došlo k převodu stereofonního signálu na monofonní. Důvodem bylo získání jednotných parametrů pro skladbu jako celek, nikoliv pro jednotlivé kanály. Finálně byl ze všech nahrávek vybrán již zmíněný motiv. Tento úsek byl vybrán z toho důvodu, že se jeho dynamika u jednotlivých nahrávek nejvíce lišila. V jeho průběhu také dochází ke změnám v síle hry na hudební nástroj, jelikož notový zápis obsahuje označení *crescenda* (postupného zesilování) i *forte* („silně“).

Samotné zpracování hudebních nahrávek bylo prováděno v programu MATLAB, s využitím toolboxu MIRtoolbox 1.6.1., který obsahuje funkce pro výpočet parametrů k identifikaci zvukového signálu (Music Information Retrieval) [19]. V prostředí MATLAB byl také využit program Praat, pomocí kterého byly vypočítány komplikovanější parametry, jako je lokální kolísání amplitudy (*shimmer*), autokorelace nebo poměr signál/šum [3].

7.1 Volba parametrů

Pro popis nahrávek byly vybrány jak parametry hodnotící dynamiku (amplituda a energie signálu), tak i parametry, které s dynamikou souvisí nepřímou (zejména parametry charakterizující frekvenční spektrum). Rezonátory hudebních nástrojů se totiž vyznačují tím, že při vyšší dynamice produkují mnohem větší počet vyšších harmonických složek, což se znatelně projevuje ve vnímané barvě zvuku i spektru kmitočtů.

Celkově bylo pro každou z nahrávek vypočítáno 43 parametrů. Seznam těchto parametrů s jejich krátkým popisem se nachází v příloze v tabulce C.1. Pro parametry vyjádřené vektorem byla určena jejich střední hodnota, medián, rozptyl a směrodatná odchylka, případně maximální hodnota. Výjimkou bylo frekvenční spektrum, kde tento údaj postrádá smysl z hlediska vnímání lidského sluchu, který je schopen vnímat pouze frekvence do 20 kHz. Tomuto faktu jsou přizpůsobena zařízení pro digitální zpracování zvuku, která signály nad 20 kHz propouští jen v omezené míře.

Alikvotní tóny většiny hudebních nástrojů však mají i mnohem vyšší frekvence. Maximální hodnota frekvence tudíž závisí na AD převodu signálu, nikoliv na dynamice interpretace.

Vypočítané parametry byly uloženy do tabulky pro další možné zpracování, např. tvorbu grafů.

7.2 Analýza vypočítaných parametrů

Ke zjištění významu vypočítaných parametrů pro porovnávání zvukových nahrávek je nutné provést určitý typ statistického zpracování. Pomocí něj je pak možné určit vztahy mezi jednotlivými parametry, popisujícími jednu zvukovou nahrávku, podobnost jednoho parametru v rámci různých interpretací hudebního díla i závislost parametrů na časovém období vzniku nahrávky nebo oblasti původu hudebního souboru.

Jednou z metod, kterou lze pro klasifikaci parametrů využít je tzv. *shluková analýza (cluster analysis)*. Ta spočívá v rozdělení objektů (parametrů) do shluků tak, aby se v těchto uskupeních vyskytovaly objekty o podobných vlastnostech, respektive parametry o podobných hodnotách. U shluků jsou následně sledovány parametry jako jsou vzdálenosti mezi objekty (tedy jejich podobnost), hustota objektů uvnitř shluku i celkový počet shluků, tvořených všemi analyzovanými daty [1, 7].

Shlukovou analýzu lze rozdělit na několik typů, např. *hard clustering* - objekt do určitého shluku buď patří nebo nepatří, a *soft clustering* - objekt patří do určitého shluku do jisté míry. Dalším členěním této analýzy je hierarchické a nehierarchické shlukování. Hierarchická metoda tkví v členění shluků na podmnožiny, dochází tedy k větvení objektů. Při nehierarchickém shlukování jsou jednotlivé shluky oddělenými množinami. Vzdálenost mezi shluky je možné definovat různými způsoby, např. vzdáleností dvou nejbližších objektů různých shluků, vzdáleností dvou nejvzdálenějších objektů různých shluků, vzdáleností center dvou různých shluků, nebo také Wardovou metodou, jež vychází z analýzy rozptylu [1, 7].

Pro shlukování existuje více než sto různých algoritmů, přičemž nelze jeden z nich stanovit jako objektivně správný. Volba algoritmu totiž závisí na typu vstupních dat a účelu požadovaného výstupu. Graficky lze výsledky analýzy znázornit pomocí histogramu, kde výška sloupce udává četnost výskytu daného objektu nebo s využitím bodového diagramu, v němž lze vizuálně sledovat shluky bodů, udávajících oblast výskytu objektů [7].

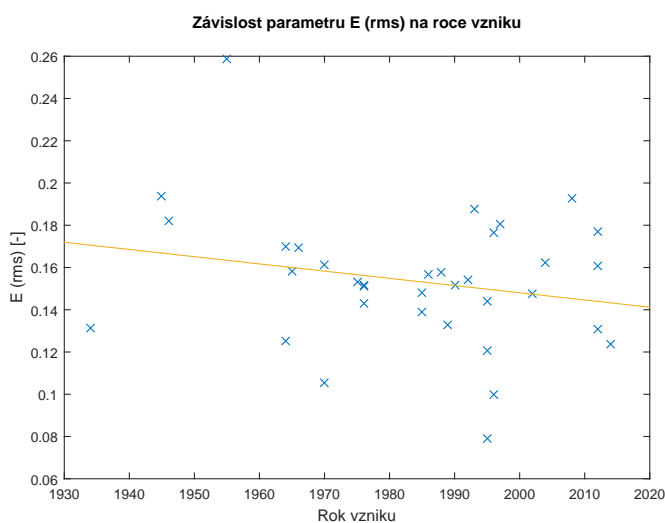
Další možností statistické analýzy parametrů je tzv. *binomické rozdělení*, označované též jako *Bernoulliho schéma*. To spočívá v diskrétním rozdělení pravděpodobnosti výskytu náhodného jevu. Této metody lze využít i v rámci hodnocení parametrů různých zvukových nahrávek. Binomické rozdělení obecně udává počet úspěchů v několika nezávislých pokusech, přičemž pravděpodobnost úspěchu nabývá hodnot od 0 do 1 [11]. Při vzájemném porovnávání parametrů různých interpretací hudebních děl můžeme tento úspěch považovat za podobnost daných parametrů. Čím je pravděpodobnost větší, tím se rozdíly mezi hodnotami parametrů snižují.

8 VYHODNOCENÍ

Tabulka obsahující vypočítané hodnoty již zmíněných 43 parametrů se nachází v příloze na CD. Sloupce tabulky náležejí jednotlivým nahrávkám, v řádcích jsou vyneseny zmíněné parametry. V rámci této práce byl pro každý parametr vytvořen graf, do něhož je vynesena závislost hodnoty daného parametru na roce vzniku souborů, interpretujících porovnávané hudební dílo. Grafem je tedy bodový diagram, na němž je možné vizuálně pozorovat vznik shluků, případně rozptyl hodnot parametrů. Graf je následně proložen regresní přímkou, která znázorňuje aproximaci těchto hodnot. Pomocí ní můžeme určit, zda mají parametry konstantní, klesající nebo naopak stoupající charakter.

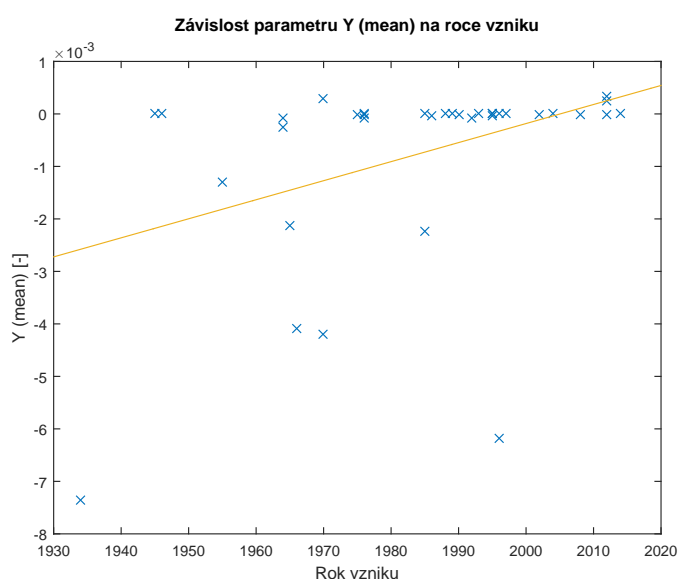
Pro porovnávání byly vybrány nejpodstatnější parametry popisující dynamiku i parametry, které v grafech vykazovaly nejzajímavější průběh. Globálně má podle regresní přímkou většina grafů klesající charakter, což znamená, že se dynamika interpretací v průběhu času snižovala. Současně však můžeme pozorovat, že u nahrávek o starších datech vznikají shluky o poměrně malém průměru, zatímco později se body mnohem více vzdalují od regresní přímkou. Z toho vyplývá, že se hodnoty parametrů v dřívější době příliš nelišily a interpretace měly tudíž podobnou dynamiku, v současnosti se však dynamika různých provedení značně liší.

Jedním z nejvýznamnějších parametrů charakterizujících dynamiku nahrávky je efektivní hodnota energie RMS - E (rms). Jak už bylo zmíněno, hodnota energie má podobně jako většina parametrů klesající charakter. V grafu můžeme pozorovat téměř konstantní hodnotu energie v letech 1960 až 1980. V pozdějším období se její hodnota výrazně liší i v rámci jednoho roku.



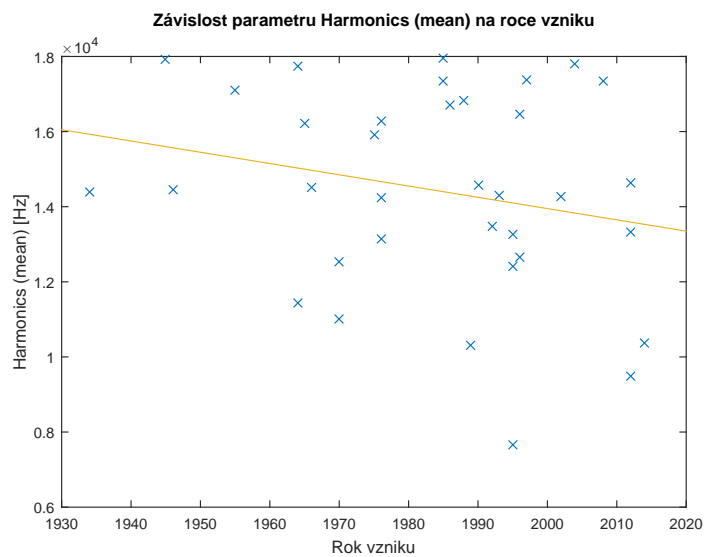
Obr. 8.1: Závislost parametru E (rms) na roce vzniku

Dalším důležitým parametrem pro popis dynamiky je amplituda signálu. Je to v podstatě vektor hodnot jednotlivých vzorků digitálního signálu. V grafu můžeme pozorovat, že střední hodnota této amplitudy, Y (*mean*), má téměř konstantní průběh. To je způsobeno faktem, že byla u signálu v rámci předzpracování provedena normalizace, čili zesílení na co nejvyšší možnou hodnotu před zkreslením signálu. Proto se většina hodnot pohybuje v okolí nuly. V grafu se však nacházejí i body o mnohem nižších hodnotách. Příčinou těchto odchylek může být skutečnost, že byla normalizace provedena pro skladbu jako celek, nikoliv pro analyzovaný motiv. Z toho vyplývá, že ostatní pasáže skladby byly interpretovány ve vyšší dynamice, než jakou měl zmíněný úsek.



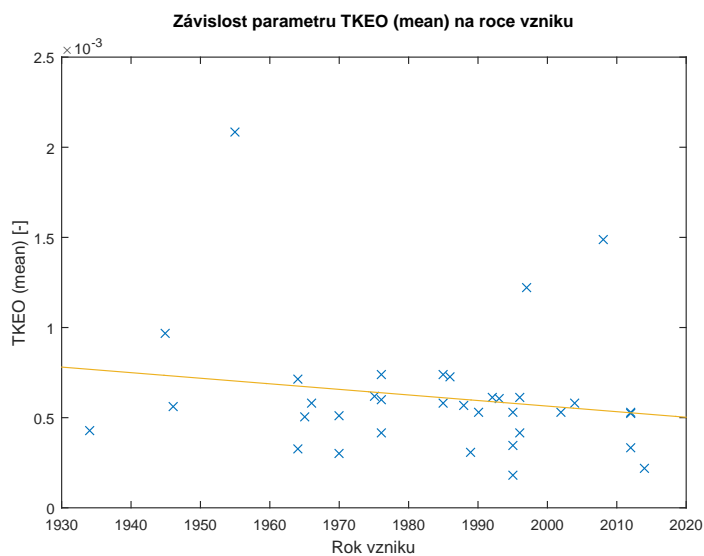
Obr. 8.2: Závislost parametru Y (*mean*) na roce vzniku

S dynamikou nahrávky souvisí také frekvenční spektrum, tedy počet vyšších harmonických složek signálu. Při hře na hudební nástroj vznikají tyto alikvotní tóny v rezonátoru nástroje a jejich počet a amplituda roste společně s dynamičností hry hudebníka. V grafu na následující straně je možno pozorovat střední hodnotu vyšších harmonických složek signálu - *Harmonics (mean)*, v tomto případě složek o frekvenci vyšší než 1500 Hz. Přímka regrese je opět klesající, což potvrzuje domněnku, že se snižující se energií klesá také množství alikvotních tónů. Při pozorování bodů uvnitř grafu si však lze všimnout, že má jejich rozložení téměř náhodný charakter a nevznikají zde žádné nápadné shluky. V průběhu času však nejnižší hodnoty v grafu stále viditelně klesají.



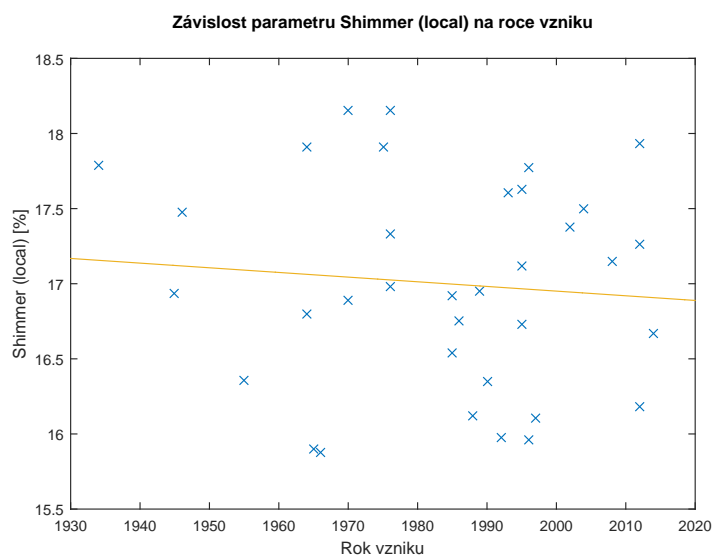
Obr. 8.3: Závislost parametru Harmonics (mean) na roce vzniku

Zajímavý průběh vykazuje parametr *TKEO (mean)*, neboli střední hodnota Teager-Kaiserova energetického operátoru. V grafu vidíme, že až na tři výjimky má tento údaj přibližně konstantní charakter, shluky bodů se tedy pohybují kolem stálé hodnoty. Současně lze i zde pozorovat, že tento parametr nabývá v průběhu od minulosti po současnost stále nižších hodnot.



Obr. 8.4: Závislost parametru TKEO (mean) na roce vzniku

Posledním porovnávaným parametrem je tzv. *Shimmer (local)*, neboli lokální kolísání amplitudy. V grafu je možné pozorovat, že má tento parametr v čase zcela náhodný charakter. Nevznikají zde žádné viditelné shluky a krajní body se nacházejí v poměrně velké vzdálenosti od regresní přímky. Pro vyjádření dynamiky nahrávky proto nemá přílišný význam. Podobně je tomu i u parametrů jako je *Centroid* (spektrální těžiště), *Entropy* (entropie) nebo *HNR* (střední hodnota poměru signál-šum), jejichž rozložení je také do značné míry náhodné a nejsou tedy vhodné pro charakterizaci dynamiky různých verzí nahrávek.



Obr. 8.5: Závislost parametru shimmer (local) na roce vzniku

Nedokonalostí této metody hodnocení nahrávek může být neznalost způsobu nahrávání a postprodukce jednotlivých interpretací, neboť při nich běžně dochází k využívání efektů typu kompresor, který záměrně snižuje dynamiku nahrávky. Proto může dojít k tomu, že není hodnocen způsob hry hudebníků, ale množství použitých efektů. Kompresor má však vliv především na energii a amplitudu signálu. Frekvenční spektrum, které je též ovlivněno dynamičností hry na hudební nástroj, zůstává nezměněno. Na množství vyšších harmonických složek má ovšem vliv další často využívaný efekt - ekvalizér. Ten se využívá nejen k úpravě barvy zvuku, ale i k potlačení nežádoucích rezonančních frekvencí, které se mnohdy vyskytují právě v oblastech o vyšším kmitočtu. Jelikož ekvalizér zesiluje nebo zeslabuje určitá frekvenční pásma, má vliv nejen na množství alikvotních tónů, ale také na hodnoty energie signálu.

Dalším drobným problémem této metody porovnávání na základě časového období může být skutečnost, že rok vzniku udává vznik hudebního souboru, nikoliv

nahrávky. Mnohdy je však efektivnější porovnávat interpretace z tohoto hlediska, neboť hudebníci s dřívějším datem narození mohou mít jiný styl hudebního provedení než mladší interpreti, a tento styl se v průběhu jejich života obvykle nemění.

9 ZÁVĚR

Tato práce se zabývá zpracováním hudebního signálu a jeho následnou analýzou z hlediska dynamiky. V první kapitole je popsána historie záznamu a zpracování analogového i digitálního zvukového signálu. Poté je zde srovnána hudba z hlediska uměleckého s přístupem fyzikálním a jsou zde tak přiblíženy vlastnosti hudebních nahrávek, které jsou důležité pro jejich popis.

V následující kapitole je popsáno předzpracování hudebního signálu, které je důležité pro následující práci se signálem. Ve třetí kapitole jsou pak uvedeny parametry, které u hudebních nahrávek hodnotíme. Zvláštní pozornost je věnována právě dynamickým parametrům, na základě nichž bude prováděno porovnávání hudebních nahrávek. V kapitole s pořadovým číslem čtyři je krátce popsáno statistické zpracování hledaných parametrů a je zde naznačen výstup systému.

Pátá kapitola se zabývá výzkumy této oblasti (Music Information Retrieval, Audio-to-Score Alignment) ve světě, metodami a způsobem práce vědců, zabývajících se těmito výzkumy a je zde také naznačeno, jakých výsledků bylo dosud dosaženo. Krátce jsou zde popsány i statistické modely, které jsou často využity pro analýzu zvukového signálu.

Následující kapitola této bakalářské práce popisuje návrh vyhodnocovacího systému a je zde zobrazeno také blokové schéma, které je využito k implementaci tohoto systému.

V sedmé kapitole je popsán způsob práce při zpracování hudebních nahrávek, důvod volby parametrů pro popis dynamiky a následně dvě metody, které mohou být využity k vyhodnocení a porovnání vypočítaných parametrů. Jsou jimi shluková analýza a binomické rozdělení.

Poslední kapitola se zabývá samotným vyhodnocením vypočítaných hodnot parametrů, z nichž pět je popsáno detailněji za pomoci grafů, zobrazujících závislost těchto parametrů na období vzniku uskupení, interpretujících dané hudební dílo.

Na základě provedených srovnání byly jako nejpodstatnější parametry vyhodnoceny energie E (rms) a amplituda Y (mean) signálu, střední hodnota vyšších harmonických složek signálu Harmonics (mean) a Teager-Kaiserův energetický operátor TKEO (mean).

Při sledování vývoje dynamiky v různých časových obdobích byla jako závěr ustanovena skutečnost, že se průměrná dynamika interpretací v průběhu času spíše snižovala. Současně však platí, že v minulosti se pohybovala dynamičnost skladeb kolem konstantních hodnot, zatímco v současnosti mají různé verze nahrávek naprosto odlišnou dynamiku.

Problémem této metody analýzy dynamiky nahrávek je neznalost množství efektů,

které byly využity při postprodukci nahrávky a mohou výrazně ovlivňovat její dynamický i frekvenční průběh.

Jelikož vypočítané parametry charakterizují nahrávku jako celek, bylo by vhodné v budoucnu nahrávky analyzovat souběžně a porovnávat hodnoty parametrů v jednotlivých časových okamžicích. Následně by měly být přesněji definovány vztahy mezi jednotlivými parametry.

LITERATURA

- [1] BAILEY, K., D.: *TYPOLOGIES AND TAXONOMIES: An Introduction to Classification Techniques*. [online]. University of California, Los Angeles, 1994. [cit. 20. 5. 2017]. Dostupné z URL: <<https://www.researchgate.net/file.PostFileLoader.html?id=54c946c7cf57d7772d8b46cf&assetKey=AS%3A273684877512704%401442262968252>>.
- [2] BERNARDO, J., M.: *Bayesian statistics*. [online]. Valencia: University Press, 1980. [cit. 11. 10. 2016]. Dostupné z URL: <<http://www.uv.es/bernardo/BayesStat.pdf>>.
- [3] BOERSMA, P.; WEENIK, D.: *Praat: doing phonetics by computer*. [Computer program] [online]. Version 5.3.51, 2013. [cit. 19. 4. 2017]. Dostupné z URL: <<http://www.fon.hum.uva.nl/praat>>.
- [4] CARABIAS-ORTI J. J., RODRIGUEZ-SERRANO F. J., VERA-CANDEAS P., RUIZ-REYES N., CANADAS-QUESADA F. J.: *An audio to score alignment framework using spectral factorization and dynamic time warping*. [online]. Music Technology Group (MTG), Universitat Pompeu Fabra, Spain; Polytechnical School of Linares, Universidad de Jaen, Spain [cit. 28. 10. 2016]. Dostupné z URL: <http://ismir2015.uma.es/articles/94_Paper.pdf>.
- [5] COOC, N.: *Beyond the Score: Music as Performance*. Oxford University Press, 2013. ISBN 978-0-19-935743-7.
- [6] DOLEŽAL, A.: *Programové vybavení pro frekvenční analýzu zvukového signálu*. [online]. Univerzita Tomáše Bati ve Zlíně, 2006 [cit. 21. 10. 2016]. Dostupné z URL: <http://digilib.k.utb.cz/bitstream/handle/10563/1144/dole%20C5%BEal_2006_bp.pdf?sequence=1&isAllowed=y>.
- [7] EVERITT, B.; LANDAU, S., LEESE, M.; STAHL, D.: *Cluster analysis*. [online]. King's College London, UK, 2011. [cit. 20. 5. 2017]. Dostupné z URL: <http://hbanaszak.mjr.uw.edu.pl/TempTxt/EverittEtAl_2011_ClusterAnalysis.pdf>.
- [8] FARRÚS, M.; HERNANDO, J.; EJARQUE, P.: *Jitter and Shimmer Measurements for Speaker Recognition*. [online]. Universitat Politècnica de Catalunya, Barcelona, Spain. [cit. 23. 11. 2016]. Dostupné z URL: <http://www.cs.upc.edu/~nlp/papers/far_jit_07.pdf>.
- [9] FOSLER-LUSSIÉ, E.: *Markov Models and Hidden Markov Model: A Brief Tutorial*. International Science Computer Institute, 1998 [cit. 11. 10. 2016].

- Dostupné z URL: <<http://di.ubi.pt/~jpaulo/competence/tutorials/hmm-tutorial-1.pdf>>.
- [10] GHAHRAMANI, Z.: *An Introduction to Hidden Markov Models and Bayesian Networks*. [online]. University College London, 2001 [cit. 11. 10. 2016]. Dostupné z URL: <<http://mlg.eng.cam.ac.uk/zoubin/papers/ijprai.pdf>>.
- [11] GRINSTEAD, C., M.; SNELL, J., L.: *Introduction to Probability*. [online]. Swarthmore College, Dartmouth College, 2003. [cit. 20. 5. 2017]. Dostupné z URL: <https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/amsbook.mac.pdf>.
- [12] HEJTMÁNEK, M.; HAŠEK, P.; HNILICA, P.: *Hudba očima fyzika*. [online]. Fakulta jaderná a fyzikálně inženýrská ČVUT [cit. 5. 10. 2016]. Dostupné z URL: <<http://fyzsem.fjfi.cvut.cz/2005-2006/Zima05/proc/hudba.pdf>>.
- [13] JANEČEK, K.: *Tektonika: Nauka o stavbě skladeb*. Supraphon, Praha 1968. ISBN 02-097-68.
- [14] JEHAN, T.: *Musical Signal Parameter Estimation*. [online]. Université de Rennes, 1997 [cit. 21. 10. 2016]. Dostupné z URL: <http://web.media.mit.edu/~tristan/Papers/CNMAT_Tristan.pdf>.
- [15] JODER, C., ESSID, S., GÄEL, R.: *An improved hierarchical approach for music-to-symbolic score alignment*. [online] Institut TELECOM, TELECOM ParisTech, 2010 [cit. 28. 10. 2016]. Dostupné z URL: <<http://biblio.telecom-paristech.fr/cgi-bin/download.cgi?id=10383>>.
- [16] JOSEPHSON, M.: *Edison: A Biography*. McGraw-Hill, 1959. ISBN 0471548065
- [17] KNEES, P.; SCHEDL, M.: *Music Similarity and Retrieval: An Introduction to Audio- and Web-based Strategies*. Department of Computational Perception, Johannes Kepler University, Linz. Springer, 2016. ISBN 978-3-662-49722-7
- [18] LARTILLOT, O.: *MIRtoolbox 1.6.1, User's Manual*. Aalborg University, Denmark, Department of Architecture, Design and Media Technology, 2014. [cit. 2. 5. 2017]. Dostupné z URL: <<https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/projects2/past-projects/coe/materials/mirtoolbox/manual1-6-1.pdf>>.
- [19] LARTILLOT, O.; TOIVIAINEN, P.; EEROLA, T.: *MIRtoolbox — Humanistis-yhteiskuntatieteellinen tiedekunta*. [online]. University of Jyväskylä, Humanities and Social Sciences, 2016. [cit. 19. 4. 2017]. Dostupné

- z URL: <<https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/projects2/past-projects/coe/materials/mirtoolbox>>.
- [20] MAEZAWA, A., OKUNO, H.G.: *Bayesian Audio-to-Score Alignment Based on Joint Inference of Timbre, Volume, Tempo, and Note Onset Timings*. [online] Department of Intelligence Science and Technology, Kyoto University, 2015. [cit. 8. 11. 2016]. Dostupné z URL: <http://arteca.mit.edu/journal/10.1162/COMJ_a_00286>.
- [21] MAEZAWA, A.; OKUNO, H.,G.; OGATA, T.; GOTO, M.: *Polyphonic Audio-to-Score Alignment Based on Bayesian Latent Harmonic Allocation Hidden Markov Model*. [online]. Kyoto University, 2011. [cit. 20. 11. 2016]. Dostupné z URL: <http://www.terasoft.com.tw/conf/ismir2014/proceedings/T043_178_Paper.pdf>.
- [22] MACRAE, R.; DIXON, S.: *Accurate Real-Time Windowed Time Warping*. [online]. Centre for Digital Music, 2010. [cit. 20. 11. 2016]. Dostupné z URL: <<https://www.eecs.qmul.ac.uk/~simond/pub/2010/Macrae-Dixon-ISMIR-2010-WindowedTimeWarping.pdf>>.
- [23] MORAKEAS, M.: *Audio to Score Alignment using Hidden Markov Models*. [online] Technological Educational Institute of Crete, School of Applied Science, 2014. [cit. 28. 10. 2016]. Dostupné z URL: <<http://nefeli.lib.teicrete.gr/browse/sefe/mta/2014/MorakeasMichail/attached-document-1402913270-993820-10855/MorakeasMichail2014.pdf>>.
- [24] MORTON, D.: *The history of sound recording*. [online]. 1998-2006 [cit. 1. 10. 2016]. Dostupné z URL: <http://www.recording-history.org/HTML/phono_technology1.php>.
- [25] MÜLLER, M.: *Fundamentals of Music Processing*. Springer International Publishing, Switzerland, 2015. ISBN 978-3-319-21945-5
- [26] NAKAMURA, T.: *The communication of dynamics between musicians and listeners through musical performance*. [online]. Osaka University, 1987. [cit. 1. 11. 2016]. Dostupné z URL: <<http://link.springer.com/article/10.3758/BF03210487>>.
- [27] NIEDERMAYER, B.; WIDMER, G.: *A Multi-Pass Algorithm For Accurate Audio-To-Score Alignment*. [online]. Department for Computational Perception

- Johannes Kepler University Linz; Austrian Research Institute for Artificial Intelligence Vienna, Austria, 2010. [cit. 23. 11. 2016]. Dostupné z URL: <http://www.cp.jku.at/research/papers/Niedermayer_Widmer_ISMIR_2010.pdf>.
- [28] ORIO, N.: *Alignment of Monophonic and Polyphonic Music to a Score*. [online]. Proceedings of the ICMC, 2001. [cit. 28.10.2016]. Dostupné z URL: <<http://www.influenzaarchive.org/cgi/p/pod/dod-idx?c=icmc;idno=bbp2372.2001.104>>.
- [29] OSGOOD, B.: *The Fourier Transform and its Applications*. Electrical Engineering Department, Stanford University [cit. 5.10.2016]. Dostupné z URL: <<https://see.stanford.edu/materials/lsoftaee261/book-fall-07.pdf>>.
- [30] REICHL, J.: *Encyklopedie fyziky*. [online]. 2006-2016 [cit. 1. 10. 2016]. Dostupné z URL: <<http://fyzika.jreichl.com/main.article/view/1342-porizeni-zvukoveho-zaznamu>>.
- [31] SMITH, J. O.: *Spectral Audio Signal Processing*. [online]. Center for Computer Research in Music and Acoustics (CCRMA), 2006 [cit. 21. 10. 2016]. Dostupné z URL: <<https://ccrma.stanford.edu/~jos/sasp/>>.
- [32] SMITH, S. W.: *The Scientist and Engineer's Guide to Digital Signal Processing*. [online]. 1997-2011 [cit. 2. 10. 2016]. Dostupné z URL: <<http://www.dspguide.com/ch22/3.htm>>.
- [33] TODD, N. P.: *The dynamics of dynamics: A model of musical expression*. [online]. Department of Psychology, University of Exeter, United Kingdom, 1992. [cit. 26. 11. 2016]. Dostupné z URL: <http://www.iro.umontreal.ca/~pift6080/H08/documents/papers/todd_dyn.pdf>.
- [34] UMAPATHY, K.; GHORAANI, B.; KRISHNAN, S.: *Audio Signal Processing Using Time-Frequency Approaches: Coding, Classification, Fingerprinting, and Watermarking*. [online]. EURASIP Journal on Advances in Signal Processing, 2010. [cit. 2. 10. 2016]. Dostupné z URL: <<http://asp.eurasipjournals.springeropen.com/articles/10.1155/2010/451695>>.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

BER	Poměr energetických pásem – Band Energy Ratio
BS	Bayesovská statistika – Bayesian Statistics
DBN	Dynamické Bayesovské sítě – Dynamic Bayesian Networks
DTW	Dynamické borcení časové osy – Dynamic Time Warping
HMM	Skrytý Markovův Model – Hidden Markov Model
HNR	Poměr signál/šum – Harmonics to Noise Ratio
LSTER	Low Short-Time Energy Ratio
MIR	Získávání informací z hudby – Music Information Retrieval
MM	Markovův Model – Markov Model
MPSD	Medián výkonové spektrální hustoty – Median of Power Spectral Density
SC	Spektrální těžiště – Spectral Centroid
SF	Spektrální tok – Spectral Flux
SNR	Odstup signálu od šumu – Signal to Noise Ratio
TKEO	Teager-Kaiserův energetický operátor
ZCR	Počet průchodů nulovou hodnotou – Zero Crossing Rate

SEZNAM PŘÍLOH

A	Obsah přiloženého CD	45
B	Seznam nahrávek	46
C	Seznam vypočítaných parametrů	48

A OBSAH PŘILOŽENÉHO CD

Přiložené DVD obsahuje elektronickou verzi bakalářské práce. Hlavní dokument nese název „Bakalarska prace - Sarka Zemankova.pdf“. Složka „Matlab“ obsahuje funkce pro výpočet parametrů v prostředí MATLAB. Výpočet se spouští pomocí skriptu „VypocetParametru.m“. Veškeré vypočítané parametry se nachází v tabulce pojmenované „Vysledky“. Složka s názvem „Grafy“ obsahuje závislosti všech vypočítaných parametrů na roce vzniku souborů, interpretujících danou nahrávku.

Obsah CD:

- Bakalarska prace - Sarka Zemankova.pdf
- Grafy
- Matlab
- Vysledky.xls

B SEZNAM NAHRÁVEK

Tab. B.1: Seznam nahrávek

Č.	Verze	Původ	Rok vzniku
001	Alban Berg Quartet	Rakousko	1970
002	Arcadia Quartet	Rumunsko	2012
003	Arianna Quartet	USA	1992
004	Cecilia String Quartet	USA	2012
005	Emerson Quartet	USA	1976
006	Endres Quartet	Německo	1955
007	Guarneri Quartet	USA	1964
008	Hagen Quartett	Rakousko	1970
009	Henschel Quartet	Německo	1995
010	Janáček String Quartet	Brno, Česká republika	1934
011	Jerusalem Quartet	Izrael	1993
012	Juilliard String Quartet	USA	1946
013	Mandelring Quartet	Německo	1988
014	Martinu Quartet	Praha, Česká republika	1976
015	Medici Quartet	Velká Británie	1996
016	Melos Quartet	Německo	1965
017	New Helsinki Quartet	Finsko	1995
018	Panocha Quartet	Praha, Česká republika	1966
019	Pavel Haas Quartet	Česká republika	2002
020	Quatuor Debussy	Francie	1990
021	Raphael Quartet	Nizozemí	2012
022	Schoenberg Quartet	Nizozemí	1976
023	Smetana Quartet	Praha, Česká republika	1945
024	Stamitz Quartet	Česká republika	1985
025	Škampa Quartet	Praha, Česká republika	1989
026	Talich Quartet	Praha, Česká republika	1964
027	The Fry Street Quartet	USA	1997
028	Vanbrugh Quartet	Velká Británie	1986
029	Wihan Quartet	Česká republika	1985
030	Doric String Quartet	Velká Británie	1995
031	Quartetto Energie Nove	Švýcarsko	2008
032	Wroclaw Chamber , Ernst Kovacic Orchestra Leopoldinum	Polsko	2014

Č.	Verze	Původ	Rok vzniku
033	Quatuor Diotima	Francie	1996
034	Australian Chamber Orchestra, Richard Tognetti	Austrálie	1975
035	London Sinfonietta and Wiener Philharmoniker	Velká Británie a Rakousko	2004

C SEZNAM VYPOČÍTANÝCH PARAMETRŮ

Tab. C.1: Seznam parametrů

Název parametru	Jednotka	Popis parametru
Y (mean)	[-]	střední hodnota výchylky
Y (median)	[-]	medián výchylky
Y (max)	[-]	maximální hodnota výchylky
Y (peak2rms)	[-]	rozdíl špičkové a efektivní hodnoty výchylky
Y (var)	[-]	rozptyl výchylky
Y (std)	[-]	směrodatná odchylka výchylky
E (rms)	[-]	efektivní hodnota energie
E (median)	[-]	medián energie
LowEnergy	[%]	procento rámců, které mají nižší než průměrnou energii
Harmonics (mean)	[Hz]	střední hodnota alikvót nad 1500 Hz
Harmonics (median)	[Hz]	medián alikvót nad 1500 Hz
Harmonics (var)	[Hz]	rozptyl alikvót nad 1500 Hz
Harmonics (std)	[Hz]	směrodatná odchylka alikvót nad 1500 Hz
Rolloff	[Hz]	frekvence, pod níž se nachází 85% energie signálu
Brightness	[%]	množství energie v oblasti nad 1500 Hz
Zerocross	[-]	počet průchodů nulovou hodnotou
Irregularity	[-]	míra nepravidelnosti po sobě jdoucích špiček spektra
Centroid	[Hz]	frekvenční těžiště
Spread	[Hz]	frekvenční rozptyl
Skewness	[-]	šikmost
Kurtosis	[-]	špičatost
Flatness	[-]	plochost
Entropy	[-]	entropie - míra neurčitosti náhodného procesu
Peak1	[-]	amplituda první špičky
Peak2	[-]	amplituda druhé špičky
Peak3	[-]	amplituda třetí špičky
Peak4	[-]	amplituda čtvrté špičky
BER	[-]	poměr energií pásem o nízké a vysoké frekvenci
TKEO (mean)	[-]	střední hodnota TKEO

Název parametru	Jednotka	Popis parametru
TKEO (median)	[-]	medián TKEO
TKEO (max)	[-]	maximální hodnota TKEO
TKEO (var)	[-]	rozptyl TKEO
TKEO (std)	[-]	směrodatná odchylka TKEO
MPSD	[-]	medián výkonové spektrální hustoty
Shimmer (local)	[%]	lokální kolísání amplitudy
Shimmer (dB)	[dB]	lokální kolísání amplitudy v dB
Shimmer (APQ3)	[%]	průměrný rozdíl mezi amplitudou a průměrem dvou sousedních amplitud
Shimmer (APQ5)	[%]	průměrný rozdíl mezi amplitudou a průměrem čtyř sousedních amplitud
Shimmer (APQ11)	[%]	průměrný rozdíl mezi amplitudou a průměrem deseti sousedních amplitud
Shimmer (DDP)	[%]	průměrný rozdíl mezi za sebou jdoucími rozdíly amplitud
Autocorr	[-]	střední hodnota autokorace
HNR	[-]	střední hodnota poměru signál-šum
NHR	[dB]	střední hodnota poměru signál-šum v dB