

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2019

Bc. Matěj Ištváněk



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

ANALÝZA INTERPRETACE HUDBY METODAMI ČÍSLICOVÉHO ZPRACOVÁNÍ SIGNÁLU

ANALYSIS OF EXPRESSIVE MUSIC PERFORMANCE USING DIGITAL SIGNAL PROCESSING METHODS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Matěj Ištvanek

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. Zdeněk Smékal, CSc.

BRNO 2019

Diplomová práce

magisterský navazující studijní obor **Audio inženýrství**

Ústav telekomunikací

Student: Bc. Matěj Ištvanek

ID: 161433

Ročník: 2

Akademický rok: 2018/19

NÁZEV TÉMATU:

Analýza interpretace hudby metodami číslicového zpracování signálu

POKyny PRO VYPRACOVÁNÍ:

Agogika v hudbě vyjadřuje změnu tempa v průběhu hudebního díla. Výrazně tím ovlivňuje celkovou interpretaci a umělecký přednes. Stejnou skladbu hraje každý orchestr či hudebník rozdílně a i malé interpretační a agogické rozdíly lze určit a následně porovnat. Cílem diplomové práce je provedení analýzy z hlediska agogiky a porovnání interpretace stejného díla zahraničního různými interprety. Analýza hudebních nahrávek s využitím metod číslicového zpracování signálů bude provedena pomocí vhodného programovacího prostředí. V rámci diplomové práce ověřte vhodně zvolenou metodu určování tempa a vyhodnoťte srovnání stejně vybraných úseků nahrávek různými interprety.

DOPORUČENÁ LITERATURA:

[1] COOK, N.: Beyond the Score: Music as Performance. Oxford University Press 2013. ISBN 978-0-19-935743-7

[2] KUNA, M., BLÁHA, M.: Čas a hudba: K dramaturgii časových prostředků v hudebně interpretačním výkonu. Academia, Praha 1982.

Termín zadání: 1.2.2019

Termín odevzdání: 16.5.2019

Vedoucí práce: prof. Ing. Zdeněk Smékal, CSc.

Konzultant:

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

UPOZORNĚNÍ:

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

ABSTRAKT

Tato diplomová práce se zabývá metodami detekce not a tempa v hudebních signálech pomocí specifických metod číslicového zpracování. Rozebírá a popisuje danou problematiku jak z hudební, tak z technické stránky. Nejprve je testováno několik implementací využívajících různá programovací prostředí, poté je vybrán fungující systém s nejvyšší přesností detekce a s nastavitelnými parametry, který je následně použit pro testování funkčnosti na zvolené referenční databázi. Je vytvořena také nástavba uvedeného algoritmu pomocí aplikace Teagerova-Kaiserova energetického operátoru ve fázi předzpracování signálu. Rozdíl v přesnosti obou systémů je porovnán – operátor v průměru zvýšil přesnost detekce globálního tempa i časových pozic jednotlivých dob. Nakonec je analyzován druhý dataset obsahující 33 různých interpretací první věty skladby Bedřicha Smetany *Smyčcový kvartet No. 1 e moll „Z mého života“*. Výsledky ukazují, že průměrné tempo celé první věty skladby se v závislosti na pozdějším roku vzniku nahrávky mírně zmenšuje.

KLÍČOVÁ SLOVA

Analýza audio signálů, detekce tónů, interpretace, onset, systém sledování rytmu, Teagerův-Kaiserův energetický operátor, tempo, zpracování digitálních signálů.

ABSTRACT

This diploma thesis deals with methods of the onset and tempo detection in audio signals using specific techniques of digital processing. It analyzes and describes the issue from both the musical and the technical side. First, several implementations using different programming environments are tested. The system with the highest detection accuracy and adjustable parameters is selected, which is then used to test functionality on the reference database. Then, an extension of the algorithm based on the Teager-Kaiser energy operator in the preprocessing stage is created. The difference in accuracy of both systems is compared – the operator has on average increased the accuracy of detection of a global tempo and inter-beat intervals. Finally, a second dataset containing 33 different interpretations of the first movement of Bedřich Smetana's composition, String Quartet No. 1 in E minor "From My Life". The results show that the average tempo of the entire first movement of the song slightly decreases depending on the later year of the recording.

KEYWORDS

Audio signal analysis, beat tracking system, digital signal processing, interpretation, onset, Teager-Kaiser energy operator, tempo, tone detection.

IŠTVÁNEK, Matěj. *Analýza interpretace hudby metodami číslicového zpracování signálu*. Brno, 2019, 72 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: prof. Ing. Zdeněk Smékal, CSc.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Analýza interpretace hudby metodami číslicového zpracování signálu“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

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

Brno

.....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu prof. Ing. Zdeňku Smékalovi CSc. za odborné vedení, konzultace a za upřímný a vstřícný přístup.

Brno

.....

podpis autora



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

PODĚKOVÁNÍ

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

Brno

.....

podpis autora



EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI



Obsah

Úvod	11
1 Hudební teorie	12
1.1 Hudba	12
1.2 Časová reprezentace	13
1.3 Rytmus	13
1.4 Metrum	13
1.5 Tempo	14
2 Metody zpracování hudebního signálu	16
2.1 Parametry	16
2.2 Onset	16
2.3 Struktura detekčních metod	17
2.4 Volba reprezentace	18
3 Metody detekce onsetů	20
3.1 Metody na bázi energie	20
3.2 Metody na bázi fáze signálu	21
3.3 Metody na bázi spektrálních změn	22
3.4 Další metody	23
4 Dostupné zdroje a knihovny	24
5 Návrh detekce tónů	25
5.1 Test implementace	25
5.2 Implementace	30
6 Detekce tempa	32
6.1 Datasetsy	33
6.2 Návrh vylepšení metody	33
7 Testování metod	36
7.1 Testování výpočtu pozic dob	39
7.2 Analýza datasetu	42
8 Závěr	47
Literatura	48

Seznam symbolů, veličin a zkratk	53
Seznam příloh	54
A Dataset	55
B Časová délka motivů datasetu	56
C Odhadnuté tempo motivů datasetu	57
D Notový zápis	58
E Obsah přiloženého DVD	72

Seznam obrázků

5.1	Spektrogram první testované nahrávky	26
5.2	Spektrální tok první testované nahrávky	26
5.3	Detekční funkce a pozice skutečných onsetů pro první testovanou nahrávku	27
5.4	Hainsworthova detekční funkce pro první testovanou nahrávku [HA-03]	28
5.5	Larocheho detekční funkce pro první testovanou nahrávku [LA-03] . .	29
5.6	Vyhlazená Larocheho detekční funkce pro první testovanou nahrávku	29
5.7	Spektrogram a detekční funkce první testované nahrávky	31
5.8	Spektrogram a detekční funkce pro Dvořákovu Mazurku op. 56 Allegro moderato	31
6.1	Původní vstupní signál a TKEO nahrávky klarinetového sóla	34
6.2	Spektrogram a detekční funkce klarinetového sóla, ODF	34
6.3	Spektrogram a detekční funkce klarinetového sóla, TKEO	34
6.4	Spektrogram a detekční funkce houslového sóla, ODF	35
6.5	Spektrogram a detekční funkce houslového sóla, TKEO	35
7.1	Medián časových rozdílů mezi vypočítanými pozicemi dob pro systém TKEO vůči referenci	41
7.2	Medián časových rozdílů mezi vypočítanými pozicemi dob pro systém ODF vůči referenci	41
7.3	Odhadnuté tempo vůči roku vzniku nahrávky – Expozice, hlavní téma (Zacatek)	43
7.4	Odhadnuté tempo vůči roku vzniku nahrávky – Expozice, vedlejší téma (A)	43
7.5	Odhadnuté tempo vůči roku vzniku nahrávky – malá coda (B)	44
7.6	Odhadnuté tempo vůči roku vzniku nahrávky – Provedení (C)	44
7.7	Odhadnuté tempo vůči roku vzniku nahrávky – Repríza (D)	45
7.8	Odhadnuté tempo vůči roku vzniku nahrávky – závěrečná coda (E)	45
7.9	Závislost časové délky každé interpretace první věty skladby a roku vzniku dané nahrávky	46
D.1	Notový zápis první věty analyzované skladby	71

Seznam tabulek

7.1	Testování výpočtu GT na referenční <i>SMC_MIREX</i> databázi	36
7.2	Výsledky testování GT referenční databáze	37
7.3	Testování rozdílů pozic dob nahrávek referenčního datasetu	39
7.4	Výsledky testování rozdílů pozic dob u nahrávek s mediánem menším než 1 s u systému TKEO a ODF	39
7.5	Výsledky testování rozdílů pozic dob u nahrávek s průměrem menším než 1 s u systému TKEO a ODF	40
7.6	Určení počtu taktů a dob jednotlivých úseků	42
A.1	Informace o databázi	55
B.1	Časová délka všech nahrávek a motivů datasetu smyčcových kvartetů	56
C.1	Odhadnuté tempo všech nahrávek a motivů datasetu smyčcových kvartetů	57

Úvod

Většina dnes běžně využívaných elektronických zařízení využívá a zpracovává nějaké formy signálů, avšak nemusí to být pouze člověkem vyrobené přístroje. Příjímačem je i lidské ucho, potažmo mozek, který může zpracovávat a vyhodnocovat signály. Díky tomuto procesu vnímáme i hudbu a její vlastnosti. Číslíkové zpracování hudebních signálů nám může pomoci s tím, co již naše subjektivní kapacita mozku nedokáže. Řeč není jen o teoreticky časově neomezeném záznamu hudby a jejím opětovné reprodukci či úpravě, ale i o možné hudebně-vědecké analýze. Aspekty interpretace lze vnímat, zaznamenávat a popisovat, nicméně lidský faktor, který do vyjadřování vstupuje, je vždy spjatý jak s daným interpretem, tak s osobou, která danou věc či problematiku popisuje.

Muzikologická věda se snaží o popsání jevů souvisejících s hudbou. Cílem muzikologického výzkumu v kombinaci s technickou stránkou věci může být například nalezení objektivních parametrů, které mají schopnost určitým způsobem popisovat a porovnávat hudební díla. Dnešní výpočetní výkon počítačů nám umožňuje zpracovávat hudbu způsobem, který byl v dřívějších dobách nemyslitelný. Odpadla závislost živého poslechu – vznikly záznamy skladeb s možností neustálého přehrávání a zpětné reprodukce. Nahrávky však vyřešily jen část problému, časové úseky a motivy se odečítaly buďto pomocí ručních stopek, kde vznikaly chyby především kvůli nedostatečné reakci měřící osoby, nebo později pomocí magnetofonů, které také nenabízely dostatečnou přesnost. Postupem času se výkon zařízení zvětšil a podrobnější analýza, například v rámci časových a tempových vlastností skladeb, začala být realitou.

Přesto je porovnávání interpretací prováděno především subjektivně, a to uznávanými kritiky a je poplatné individuálnímu pohledu hodnotitele. Lze ale zachovat určitou objektivitu, která je tvořena právě technicky přesným zpracováním časové reprezentace hudebních motivů nebo jednotlivého zastoupení hudebních elementů. Je zde vyvíjena snaha objektivizovat hodnocení interpretace daného díla různými interprety, přestože konečný umělecký pohled a následné závěry na danou problematiku budou vždy subjektivní.

Cílem této práce je shrnutí algoritmů a metod pro detekci not a tempa v hudebním signálu pomocí číslíkového zpracování. Předmětem je především detekce relevantních informací z hudebního signálu neboli mezioborová vědní disciplína Music Information Retrieval (MIR). Nejprve práce uvádí teoretickou, a poté technickou stránku problematiky včetně implementace detekčního algoritmu a následného vylepšení. Nakonec je vytvořena analýza různých interpretací první věty skladby Bedřicha Smetany *Smyčcový kvartet No. 1 e moll „Z mého života“*.

1 Hudební teorie

1.1 Hudba

Na hudbu můžeme nahlížet z různých pohledů. Pro muzikanta může hudba představovat symbolický zápis – reprezentuje instrukce, jak má předepsaný materiál zahrát. Ze zápisu vyčte rytmus, metrum, melodickou strukturu i chtěný výraz (velkou roli hraje i zkušenost nebo supervize dirigenta či hudebního režiséra). Podle jeho předchozích zkušeností a schopnosti číst tyto instrukce provede transformaci symbolů na hudbu – z notového zápisu díky jeho hudebnímu nástroji vytvoří uměleckou činnost. Každá nuance od předepsaného nebo od výkonu jiných interpretů vytváří nové dílo, originál. Svou roli hrají jemné odchylky od ladění, individuální umělecké cítění, relativní nedokonalosti vůči tempovému předpisu a celkové fyzické i psychické rozpoložení interpreta. Muzikanti nehrají mechanicky, nýbrž mírně zpomalují a naopak zrychlují v různých úsecích dané skladby. Stejně je to i s dynamikou a zvýrazňováním určitých not v motivu.

Z jiného pohledu můžeme hudbu brát jako mechanické vlnění, které se šíří látkovým prostředím. Zvukové vlny, které mají schopnost vyvolat u člověka sluchový vjem, mají frekvenci vlnění přibližně v rozmezí 16–20000 Hz². Audio signál jako číselnou reprezentaci signálu pro naši potřebu pak předpokládáme a definujeme právě v tomto rozmezí. Na rozdíl od symbolického zápisu, zde máme všechny potřebné informace jasně zaznamenané. Ve většině případů musíme ale použít nějaký druh transformace nebo jiné reprezentace, abychom určili požadované parametry – délka trvání dob, tempo, barva, dynamika, výška not ad. Číselné hodnoty, které nám poskytne samotný záznam bez další úpravy, totiž ani zdaleka nestačí. Zde nastává úskalí analýzy. Velmi záleží na druhu zpracování signálu a na metodách, které se používají. Je obtížné určit většinu parametrů efektivně a účinně, pokud je analyzovaným signálem například záznam polyfonní hudby, kde hraje více nástrojů zároveň. Podle M. Müllera se nástroje spektrálně překrývají, hráči nemusí hrát synchronně a důležitou roli hraje také efekt maskování [MÜ-15]. V neposlední řadě je důležité vzpomenout i subjektivní vnímání hudby – při analýze je vhodné myslet také na lidské vnímání a s tím spojenou psychoakustiku a její zákonitosti.

²Tato čísla se mohou z různých důvodů (predispozice, fyzické poškození nebo třeba aspekt věku) u každého jedince lišit.

1.2 Časová reprezentace

Čas a časové parametry hudebních záznamů jsou jedny z nejdůležitějších prostředků pro vyjádření interpretačních rozdílů mezi různými nahrávkami stejného díla. Předmětem zkoumání často bývá porovnání časové reprezentace určitého hudebního motivu, potažmo celé skladby nebo její části vůči ostatním interpretacím. Změření odchylky v absolutní časové oblasti však nemá dostatečně vypovídající hodnotu. Kratší čas motivu ještě nemusí vždy znamenat, že muzikanti zahráli úsek rychleji. V práci se budou vyskytovat některé pojmy z hudební terminologie, proto je vhodné je nejprve definovat.

1.3 Rytmus

Především s nástupem a rozvojem polyfonie byl v Evropě vytvořen a postupně rozvíjen zápis respektující rytmické členění skladeb. Předtím přesná časová struktura nebyla zaznamenána, jednalo se spíše o zápis melodických posunů, které určovaly, jak se skladba či píseň v čase vyvíjí a kam směřuje. Rytmus je v hudbě ve své zjednodušené podobě struktura not různých délek – střídání přízvučných a nepřízvučných dob (tzv. těžké a lehké) a jejich pravidelné opakování. Dnešní klasická západní notace respektuje značení rytmu a taktů, a tím poskytuje možnost vyjádření tempového členění skladeb. Takty byly v notaci vytvořeny mj. kvůli (a díky) zápisu pravidelnosti rytmické struktury.

1.4 Metrum

Metrum úzce souvisí s rytmem i s takty. Může být definováno například jako pravidelná, většinou opakující se struktura v hudebním přednesu – akcenty, takty, rytmus..., někdy se hovoří o tzv. pulzech nebo pulzaci. Bývá reprezentováno zápisem na začátku notové osnovy, během skladby se však může měnit. Typická rytmická změna ve skladbě může být znázorněna třeba přesunutím přízvuku z přízvučných dob na nepřízvučné a metrická změna zase střídáním taktů s odlišným metrem ($\frac{4}{4}$, $\frac{2}{4}$, $\frac{3}{4}$, $\frac{6}{8}$ a mnohé další). Příklad jiného značení $\frac{4}{4}$ taktu je uveden například na začátku notového zápisu analyzované skladby (obr. č. D.1)³.

³<https://imslp.org/wiki/Special:ImagefromIndex/216217/pm53>

1.5 Tempo

Pro určení rychlosti přednesu se používá termín tempo. V podstatě se jedná o závislost posloupnosti hudebních not na čase, neboli jak rychle se skladba hraje. Většinou bývá označeno v notovém zápise (dle představy skladatele) vlevo nahoře nad notovou osnovou na začátku díla. Může mít italskou slovní podobu (Lentissimo, Con moto, Prestissimo, ...) nebo může být vyjádřeno číselnou reprezentací. Někdy lze vidět použití obojího značení zároveň. Jednotkou je M.M. (Mälzelův metronom), podle vynálezce metronomu. V anglické literatuře (a dnes často i u nás) převažuje jednotka BPM (Beats Per Minute – úderů za minutu) a ve většině případů bývá číselná hodnota přiřazena čtvrté době. Značení závisí na stylu skladby a na zvoleném rytmu a metru. Pokud bychom tempo T definovali jako neměnnou, přiřazenou hodnotu rychlosti hry, můžeme ho určit jako počet úderů za minutu vydělený délkou časového úseku:

$$T = \frac{B \cdot 60}{\Delta t_s} \quad (1.1)$$

kde B je počet úderů a Δt_s je zvolený časový úsek měření. Jednotkou je BPM. Neměnné tempo je však v klasické interpretaci v souhře pouze hráčů na nástroje mezi sebou v podstatě nereálné. Záleží na mnoha faktorech, jakými jsou individuální zkušenost a hudebnost jednotlivých muzikantů, souhra celého tělesa a také vlivy prostředí a výukového stylu, ve kterých byli muzikanti vycvičeni a ve kterém danou skladbu realizují. Dané odchylky od konstantního tempa lze určovat a porovnávat v rámci vícera interpretací.

Podle studie vypracované M. Kunou a M. Bláhou se rozlišuje více druhů a typů tempa:

1. tzv. tempo metrizedované
 - tempo s přísnou vazbou hudebního metra
 - obsahuje teze a arze, přízvučné a nepřízvučné doby a metrické struktury
2. tempo s libovolnou vazbou hudebního metra
 - mezistupeň tempa metrizedovaného a nemetrizedovaného
 - libovolná artikulace (dle hudebníka)
3. tzv. tempo nemetrizedované
 - tempo bez vazby na hudební metrum
 - postaveno na základě sekundového plynutí času
 - různá znázornění – například grafické zpodobnění měřítka, kde 1 cm délky notové osnovy představuje 1 s hudebního času

Nejčastěji používané metrizované tempo se může dále dělit na:

1. základní tempo

- je nadřazeno všem ostatním typům
- tempová rychlost, která vyjadřuje abstraktní představu rychlosti a určuje průměrnou rychlost hudby během celého daného časového úseku bez ohledu na agogické individuální nepřesnosti muzikantů
- určeno především pro prvotní představu tempa začátku díla a pro zásadní změny tempa v průběhu skladby

2. pevné tempo

- charakterizováno časovým pohybem hudby, ve kterém nedochází k úmyslným ani náhodným odchylkám – strojová přesnost
- používáno v transkripčních programech nebo v syntetické hudbě, kde se tempo během určitého časového úseku v absolutní míře nemění

3. oscilující tempo

- podoba nahodilých odchylek od průměrné doby taktu
- není pevně dáno, záleží na každé interpretaci

4. proměnné tempo

- vytváření záměrných odchylek
- narušení pravidelného vývoje tempa – vědomý odklon (fermata, ritardando, rubato)

Předkládaná práce se zabývá metrizovaným tempem, které vychází z podstaty vlastností skladeb databanky, se zaměřením na základní, oscilující a proměnné tempo a na jejich odchylky vůči jednotlivým interpretacím. Nejčastěji pracuje s pojmem globální tempo (GT), což je průměrné tempo celého analyzovaného úseku nebo hudebního motivu.

2 Metody zpracování hudebního signálu

Existuje mnoho výzkumů na poli číslicového zpracování hudebních signálů. Mnohé se zabývají transkripcí hudby na symbolický zápis, automatickou detekcí hudby a jejich parametrů nebo segmentováním nástrojů z vícestopé polyfonní hudby. Základním problémem, který je potřeba v takových případech vyřešit, je mimo jiné časová reprezentace a s ní spojená detekce jednotlivých not v hudebním signálu. Právě pozice not je důležitým aspektem při porovnávání interpretací nahrávek mezi sebou. Vychází z ní tempové, rytmické a metrické struktury skladby. Metrickou strukturu může zkušený hudebník často poměrně přesně určit na základě subjektivního vnímání a poslechu dle jeho předchozích zkušeností a znalostí. Efektivní a úspěšná detekce not v komplexnější hudební struktuře, která navíc obsahuje prvky způsobené vlastností digitálního záznamu (šumy, ruchy, vlastnosti vlnění jako takového či nepřesnosti v záznamu), bývá v mnoha případech obtížným úkolem.

Algoritmy pro detekci interpretačních odchylek změn se neustále vyvíjejí a jejich efektivita se zvyšuje. Metody se vylepšují a dochází k jejich různým kombinacím, které posléze dosahují lepších výsledků. Nynější „state-of-the-art“ využívá k dosažení optimalizace nastavení parametrů jednotlivých metod především strojového učení a neuronových sítí. K používání zcela nových technik pro detekování základních časových vlastností signálu však nedochází tak často, jak tomu bývalo předtím – spíše se metody kombinují a vylepšují. Práce se zabývá především upravenými základními algoritmy bez použití strojového učení a pravděpodobnostních modelů, které i tak mohou vykazovat, v některých případech, poměrně vysokou přesnost.

2.1 Parametry

Obecně můžeme zkoumané parametry signálu rozdělit na nízkoúrovňové a vysokoúrovňové. Nízkoúrovňové parametry, někdy také lokální, se počítají přímo ze zdrojového signálu – může se jednat o detekci začátků tónů, úroveň přechodů nulou, výpočet krátkodobé energie ad. Vysokoúrovňové parametry pak vycházejí z těchto lokálních – využívá se statistika (transformace z vektorů nebo matic na skalární hodnoty, které se posléze analyzují).

2.2 Onset

Termín „Onset“ se běžně vyskytuje v odborné literatuře a znamená určitou významnou událost v akustickém signálu. Často se zaměňuje za výraz „Onset Time“, který značí začátek takové události. V rámci hudby a této práce bude brán v potaz především „Note Onset Time“ (dále jen NOT), který značí časovou informaci

začátku vzniku nového tónu nebo výraznější události v signálu – obecně se může jednat o změnu energie spektra, změnu výšky tónu nebo rychlou změnu fáze. Přestože se jedná o anglické výrazy, budou se v práci v rámci unifikace objevovat ve svém původním znění, bez překladu.

Představme si nahraný akustický signál, ve kterém je zaznamenán jeden delší úhoz kladívka na strunu klavíru bez použití pedálů (mírné podržení klávesy). Pokud bychom vynesli na osu y energii signálu, můžeme vidět, jak se tato energie noty v čase t mění. Nejprve nastává tzv. náběh (Attack), který značí, za jakou dobu se od stisku klávesy na klavíru dostane signál na maximální úroveň; dále útlum (Decay) určuje dobu, za jakou se signál z maxima dostane na hodnotu „sustain“; podržení (Sustain) udává relativně neměnnou velikost signálu během stisknuté klávesy; uvolnění (Release) značí dobu poklesu energie signálu po následném uvolnění klávesy klavíru [IŠ-17, s. 20]. Všem částem dohromady se říká ADSR obálka. NOT je tedy okamžik, kdy začíná náběh tónové obálky noty. Analogicky můžeme pozorovat vývoj energie not i u dalších nástrojů. Různé instrumenty se ale mohou lišit svým průběhem, energií i zastoupením vyšších harmonických složek v signálu. I proto se detekční algoritmy často snaží ve své implementaci a struktuře přizpůsobit vlastnostem nástrojů, pro které jsou navrženy.

2.3 Struktura detekčních metod

Celou metodu detekce lze ve většině případů dělit na několik fází:

- **Předzpracování signálu** – jedná se o převedení původního signálu na jinou reprezentaci, vytvoření mono záznamu z vícekanálové nahrávky nebo vypuštění nepotřebných informací a dat. Následující kapitola 2.4 pojednává o předzpracování více.
- **Redukce** – v této fázi dojde k výrazné redukci dat v rámci použitého algoritmu – například po aplikování výpočtu spektrálních změn může být vzorkovací frekvence detekční funkce okolo 100 Hz vůči původní vzorkovací frekvenci 44,1 kHz. Obecně se tato fáze dělí na dvě odvětví podle toho, jakým způsobem redukuje data, buďto podle spektrálních parametrů, nebo podle pravděpodobnostních modelů [BÖ-10, s. 5].
- **Detekce lokálních maxim** – úkolem poslední fáze je zpracovat výstupní signál a určit časovou pozici onsetů. K tomu se využívá vyhlazování výsledné detekční funkce, fixní nebo adaptivní prahování (thresholding) a závěrečná detekce lokálních maxim – tzv. peaků.

2.4 Volba reprezentace

Jak již bylo zmíněno, většina metod detekce NOT využívá nějakého předzpracování pro vhodnější reprezentaci signálu. Kromě sloučení stereo stop do jedné monofonní stopy a následné normalizaci se nejčastěji využívá tzv. krátkodobá Fourierova transformace diskrétního signálu (anglicky Short Time Fourier Transform – STFT). Ta vytvoří reprezentaci závislosti času a frekvence, tzv. časově-frekvenční analýzu. Jejím základem je Fourierova transformace [SM-13, s. 124]:

$$S(e^{j\omega}) = \sum_{n=-\infty}^{\infty} s[n] e^{-j\omega n} \quad (2.1)$$

$$s[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(e^{j\omega}) e^{-j\omega n} d\omega \quad (2.2)$$

Transformace se aplikuje na analyzovaný signál po krátkých úsecích, které se určují podle symetrického okna $w[n]$, které se postupně posouvá buďto o jeden vzorek (klouzavé spektrum), nebo o skupinu vzorků (skokové spektrum). Každý časový rámec je před výpočtem rychlé Fourierovy transformace (FFT) násoben oknem. Můžeme tudíž psát [SM-13, s. 125]:

$$S_{\text{STFT}}(e^{j\omega}, m) = \sum_{n=-\infty}^{\infty} s[n] w[n - m] e^{-j\omega_k n} \quad (2.3)$$

kde $\omega_k = 2\pi f_k$ a $k = 0, 1, 2, \dots, N - 1$. Vybírá se N časových vzorků, takže počítáme pouze N složek spektra. Pro efektivitu výpočtu FFT se rámce doplňují nulovými hodnotami do počtu vzorků odpovídající mocnině o základu dva. Většinou se ve zmiňovaných aplikacích používá Hammingovo nebo Hannovo okno a hodnota jeho posunutí se značí m .

V dnešních systémech pro určení onsetů na bázi spektrálních změn se často používá i jiná reprezentace – tzv. mel-frekvenční spektrální (nebo keprální) analýza, kterou využívá například práce S. Böcka [BÖ-13]. Před použitím STFT se signál filtruje pomocí banky filtrů, které jsou odvozeny podle poznatků z psychoakustiky a vychází z modelů lidského slyšení (například větší počet filtrů na nižších frekvencích, zlogaritmování spektra, ...). Tato reprezentace často dosahuje při implementaci stejné detekční funkce lepších výsledků.

Pro popis náhodných diskrétních signálů ve spektrální oblasti můžeme uvažovat například výkonovou spektrální hustotu (PSD), která popisuje rozložení výkonu daného procesu na kmitočtové ose. K tomu lze využít výpočet diskrétní Fourierovy transformace (DFT) pomocí algoritmu FFT. Provede se výpočet pro konečný počet bodů na kmitočtové ose, to znamená pro $\omega_k = 2\pi \frac{k}{N}$, $k = 0, 1, 2, \dots, N - 1$, a tím vznikne z odhadu výkonové spektrální hustoty pomocí odhadu autokorelační

posloupnosti $P(e^{j\omega})$ (rovnice č. 2.4) nový tvar periodogramu $P[k]$ (rovnice č. 2.6) [SM-13, s. 220–221].

$$P(e^{j\omega}) = \sum_{m=-(N-1)}^{N-1} r[m]e^{-j\omega m} = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n]e^{-j\omega n} \right|^2 = \frac{1}{N} |X(e^{j\omega})|^2 \quad (2.4)$$

kde $X(e^{j\omega})$ je Fourierova transformace diskrétního signálu $x[n]$.

Odhad autokorelační posloupnosti $r[m]$ se určí následovně:

$$r[m] = \frac{1}{N} \sum_{n=0}^{N-m-1} x^*[n]x[n+m], m \in \langle 0, N-1 \rangle, r \in \mathbb{R}, m \in \mathbb{N} \quad (2.5)$$

$$P[k] = P(e^{j\omega_k}) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n]e^{-jk\frac{2\pi}{N}n} \right|^2, k = 0, 1, 2, \dots, N-1 \quad (2.6)$$

Znak * značí komplexní sdružení. Při výpočtu a popisu spektrální výkonové hustoty se často používá zlogaritmovaná hodnota a jednotka vztažená k 1 W – dBW/Hz.

Kvůli použití funkce Spectral Flux (SF), neboli spektrálního toku, (kapitola č. 3.3) je vhodné ještě definovat l_p -normu. Pokud vektorem x rozumíme matici typu $(n,1)$ – neboli vektor $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$, pak můžeme obecně psát pro $p \geq 1$ [CHL-18]:

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (2.7)$$

Speciálním případem je pak oktaedrická forma pro $p = 1$ a euklidovská norma pro $p = 2$.

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad (2.8)$$

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad (2.9)$$

3 Metody detekce onsetů

3.1 Metody na bázi energie

Pokud bychom chtěli detekovat začátky not nástroje, který má při začátku každé noty velmi silný transientní charakter (rychlá změna stavu signálu způsobuje výraznou krátkodobou změnu energie), například bicí nástroje, můžeme použít relativně jednoduchou metodu. Stačí vzít hodnoty signálu v časovém měřítku, signál usměrnit pomocí výpočtu absolutní hodnoty každého vzorku a použít filtr s klouzavým průměrem pro vytvoření detekční funkce [HE-11, s. 5–7]. Pokud by hodnota takto upraveného signálu přesáhla určitou mez, filtr by zaznamenal místo překročení prahové hodnoty a označil časový okamžik (respektive vzorek signálu) jako NOT.

Další možností je rozdělit signál na menší části a vypočítat průměrnou energii každého takto vytvořeného úseku, kde N je velikost okna, které vytváří jednotlivé úseky a $x[n]$ je vstupní signál:

$$E[n] = \frac{1}{N} \sum_{-\frac{N}{2}}^{\frac{N}{2}-1} x^2[n] \quad (3.1)$$

Dále může být detekční funkce vylepšena měřením změny energie pomocí výpočtu její derivace. Například v práci M. Alonsa [AL-04] je pro tento účel použit FIR filtr a Remezův algoritmus. Jinou možností je odvození detekční funkce z lokální energie signálu v krátkých úsecích (dle délky k volitelného okna) [ZH-10, s. 297–316]:

$$E[n] = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} s^2[n+k] w[k] \quad (3.2)$$

Když dojde k rychlému nárůstu energie, často je takový okamžik vznikem nového tónu (proběhne excitace oscilátoru, který za pomoci rezonátoru vytvoří tón), nemusí to být ale výhradně pravda. Zaleží také na dynamice přednesu a povaze instrumentu jako takového. I proto je zmíněná metoda vhodná pouze pro perkusní nástroje s výraznými změnami energie a nehodí se pro složitější analýzu. Přesto byly tyto metody upraveny a vylepšeny k dosažení lepších výsledků – často se jedná o rozdělení signálu do několika frekvenčních pásem, neboť jednotlivá pásma nesou rozdílnou informaci z hlediska určení vzniku nové noty. Například A. Klapuri [KL-99] využil ve své práci rozdělení signálu do 21 frekvenčních pásem pomocí banky filtrů, a poté vypočítal modulovou obálku pro zjištění NOT – použil zde mimo jiné poznatky z psychoakustiky. M. Goto [GO-01] ve své práci rozdělil signál tentokrát do sedmi různých frekvenčních pásem a určoval náhlé změny energie pro detekci NOT a vytvoření tempové a rytmické struktury nahrávky.

U některých metod je důležité si uvědomit, že místo nejvyšší lokální energie nemusí znamenat časovou pozici NOT, ale začátek fáze „sustain“ v rámci ADSR obálky. Proto se pro lokalizaci začátku onsetů v konečné fázi algoritmu používá funkce často označovaná jako „backtrack“, která posune výsledné časové okamžiky o takovou hodnotu, aby odpovídala skutečné hodnotě NOT.

3.2 Metody na bázi fáze signálu

Energetické metody (myšleno ke vztahu k energii, nikoliv energetiky) poskytují poměrně nepřesnou detekci tzv. soft onsetů, neboli jemných, velmi slabých onsetů. Ty se totiž poměrně jednoduše mohou zakrývat a maskovat dalšími jevy v signálu. Obvyklé problematické nástroje s nevýraznými onsety jsou například strunné nástroje – typicky housle, viola nebo violoncello. Vznik nové noty nemusí být vždy doprovázen výraznější změnou energie. Navíc výrazové prvky, jako třeba vibrato, mohou detekci ještě ztížit. I proto vzniklo odvětví metod zaměřující se na určení NOT pomocí měření změny fáze.

Tato metoda je mimo jiné více popsána například v práci A. Hesse [HE-11], M. Müllera [MÜ-15, s. 313–315] nebo R. Zhoua [ZH-10]. Základem jsou následující předpoklady. Nechť je $X[n, k] \in \mathbb{C}$ komplexní Fourierův koeficient pro jednotlivé frekvence $k \in [0 : K]$ v diskretním čase $n \in \mathbb{Z}$ [MÜ-15, s. 313].

$$X[n, k] = |X[n, k]|e^{2\pi j\varphi[n, k]} \quad (3.3)$$

Pokud předpokládáme, že signál x má vysokou korelaci se sinusoidou určité frekvence $F_{koeff}[k] = fs \cdot k/N$, která je posunuta podle fáze $\varphi[n, k]$ tak, aby co nejvíce korelovala se signálem po aplikaci funkce okna n -tého segmentu (takže $|X[n, k]|$ má velkou hodnotu), a že vykazuje relativně neměnné hodnoty pro po sobě jdoucí segmenty $(\dots, n-2, n-1, n, n+1, \dots)$, pak fáze $(\dots, \varphi[n-2, k], \varphi[n-1, k], \varphi[n, k], \varphi[n+1, k], \dots)$ se postupně zvyšuje lineárně podle překryvu okna v STFT [MÜ-15, s. 313]. Poté by se fáze neměla v takovém úseku příliš měnit a skokové změny by měly být způsobeny změnou zahraničního tónu – a ty lze detekovat.

Metoda pro detekci tedy funguje na předpokladu, že během útlumu a podržení noty (viz kap. 2.1) se fáze signálu téměř nemění. Naopak při vzniku nové noty nebo obecně při změně kmitočtů začne tato změna fáze nabývat vyšší hodnoty (pokud frekvence koeficientu $X[n, k]$ není konstantní).

Metody na bázi fáze signálu vykazují většinou lepší výsledky pro detekci jemných onsetů než obdobné metody s využitím energetické nebo modulové obálky. Často dokáží zachytit i ty, které neprovází výraznější změna energie při vytvoření nového tónu. Přesto jsou samostatně málo využívané, a to hlavně kvůli možnému fázovému

zkreslení a fázovému šumu nízkoenergetických komponentů v signálu, které vytvářejí poměrně výrazné nepřesnosti při detekci. Práce tuto metodu nevyužívá, proto není dále rozebírána.

3.3 Metody na bázi spektrálních změn

Někdy je výhodné se zaměřit na hodnoty změny spektra v jednotlivých úsecích vytvořených váhovacím oknem. Změna nebo nové nasazení tónu je často doprovázeno výraznější změnou ve spektru signálu. Někdy můžeme čerpat informace obsažené právě ve vyšších frekvencích signálu – ty mohou být lineárně váhované díky výpočtu energie spektra [HE-11, s. 6–7]:

$$HFC[n] = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |W_k X_k[n]|^2 \quad (3.4)$$

Toho lze dále využít při zjišťování spektrálních změn, které mohou vypovídat o změně či vytvoření nových hudebních tónů. Metoda počítá změnu spektra mezi dvěma následujícími úseky (vytvořené pomocí STFT) a poté sečte všechny kladné změny pro vytvoření detekční funkce. Používají se dva podobné výpočty – jeden založen na l_2 -normě a druhý na l_1 -normě. Tyto parametry se označují jako Spectral Difference (SD) [HE-11, s. 7] a Spectral Flux (SF) [BÖ-10, s. 6]:

$$SD[m] = \frac{1}{K} \sum_{k=0}^{K-1} \{H(|X[m+1, k]| - |X[m, k]|)\}^2 \quad (3.5)$$

$$SF[m] = \frac{1}{K} \sum_{k=0}^{K-1} H(|X[m+1, k]| - |X[m, k]|) \quad (3.6)$$

pro $m = 0, 1, 2, \dots, M-2$, kde M je počet segmentů, K je polovina frekvenčních pásem STFT nebo počet melovských pásem a $H[x]$ se chová jako tzv. jednocestný (půlvlnný) usměrňovač – $H[x] = \frac{(x+|x|)}{2}$. Pokud je hodnota argumentu záporná, pak $H[x] = 0$ – zajímá nás pouze kladná změna energie. Někdy se signál nejprve upraví logaritmickou kompresí spektrálních koeficientů, aby se slabé spektrální komponenty zvýraznily [MŮ-15, s. 309–311]. Až poté je zavedena tzv. spectral-based novelty function neboli samotný výpočet SF.

Na určení informací ze spektrálních složek signálu je v dnešní době založeno velké množství algoritmů pro detekci not v hudebním signálu. Jedná se totiž o relativně jednoduchou a výpočetně nenáročnou metodu pro zjištění NOT (vůči například

neuronovým sítím), která v rámci svých různých variací a implementací zároveň poskytuje poměrně dobrou úspěšnost při analýze nahrávek. Důkazem toho je i relativní četnost takových metod a algoritmů na konferenci ISMIR v dřívějších ročnících⁴.

3.4 Další metody

Jinou možností jak určit výskyt nové noty v signálu je zaměřit se na výšku tónu (respektive tzv. fundament v kmitočtové oblasti). Částečně tuto specifikaci pokrývají metody na bázi změny spektra signálu, nicméně některé algoritmy mohou používat detekce založené výhradně na změně výšky tónů. Pro tento účel se používají například chroma vektory [MÜ-15, s. 118–128], které mají své využití i v pokročilejších transkripčních systémech. Problém však nastává při analýze polyfonie a komplexnější metrické struktury skladeb. Další metody detekce vycházejí například z pravděpodobnostních modelů (HMM – Hidden Markov Models, neboli skryté Markovovy modely) nebo z tzv. komplexních domén a jejich kombinací [MÜ-15, s. 315–316]. Bývají však náchylné na šum a jiné rušivé složky v nahrávkách a použití u starších nahrávek bývá problémem. Nynější „state-of-art“ je využití strojového učení a neuronových sítí [SCH-14], [EY-10]. Nutnost velkého množství často nedostupných trénovacích dat pro správnou funkci algoritmu je zřejmou limitací těchto metod – přesto tyto systémy dosahují v podstatě v každém odvětví MIR nejvyšší úspěšnosti.

⁴viz <https://dblp.uni-trier.de/db/conf/ismir/index.html>

4 Dostupné zdroje a knihovny

Existuje mnoho toolboxů, knihoven a přídatných modulů vytvořených především MIR komunitou a výzkumníky z celého světa. Většinou pracují buď jako samostatné aplikace, pluginy pro další aplikace, nebo jako softwarové knihovny. Umožnily mimo jiné seskupit databáze pro výzkumné a vývojové skupiny (například *The Million Song Dataset* [BE-11]).

Mezi nejznámější a nepoužívanější patří mj. podle D. Moffata [MO-15]:

- **Aubio** – Knihovna pro extrakci⁵ vysokoúrovňových parametrů jako jsou onset detekce, sledování tempa a melodie⁶.
- **Essentia** – Prostředí pro zpracování nízkoúrovňových i vysokoúrovňových parametrů a zpracování vstupních a výstupních signálů včetně statistické analýzy. Napsáno v C++ v propojení s Pythonem [BO-13].
- **jAudio** – Samostatná aplikace pro jazyk Java s vlastním GUI a CLI [Mc-05].
- **LibROSA** – API pro extrakci parametrů a zpracování dat v Pythonu⁷.
- **LibXtract** – Nástroj pro extrakci nízkoúrovňových parametrů se zaměřením na zpracování v reálném čase. Podpora v C, Max-MSP, Pure Data, Super Collider a Vamp [BU-07].
- **Marsyas** – Samostatná aplikace pro zpracování signálu v reálném čase s GUI a CLI. Možnost extrakce nízkoúrovňových parametrů pomocí strojového učení [TZ-99].
- **Meyda** – API s možností reálné extrakce nízkoúrovňových parametrů napsané v Javascriptu a vytvořené pro webový prohlížeč [RA-15].
- **MIR Toolbox** – API pro offline extrakci nízkoúrovňových i vysokoúrovňových parametrů v Matlabu [LA-07].
- **Timbre Toolbox** – Nástroj obdobný MIR Toolboxu, optimalizovaný především na analýzu tónu [PE-11].
- **YAAFE** – Knihovna pro extrakci nízkoúrovňových parametrů s požadavkem na výpočetní efektivitu. Napsána v C++ s CLI a možností použití i v Matlabu a Pythonu [MA-10].
- **Sonic Visualiser** – Samostatný program s možností využití přídatných modulů a pluginů pro detekci onsetů, tempa a zobrazení vícera spektrogramů současně⁸. Příkladem vytvořených pluginů je projekt Mazurka⁹.

⁵Extrakcí je míněno „vytažení“ informací z dat pomocí různých metod číslicového zpracování hudebních signálů.

⁶<https://aubio.org/>

⁷<https://librosa.github.io/librosa/>

⁸<https://www.sonicvisualiser.org/>

⁹<http://www.mazurka.org.uk>

5 Návrh detekce tónů

Každá zmiňovaná metoda má své výhody i nevýhody. Co se týče návrhu, implementace a výkonové náročnosti v porovnání s efektivitou a úspěšností detekce se jeví metoda založená na detekci onsetů pomocí spektrálních změn jako nejvhodnější.

5.1 Test implementace

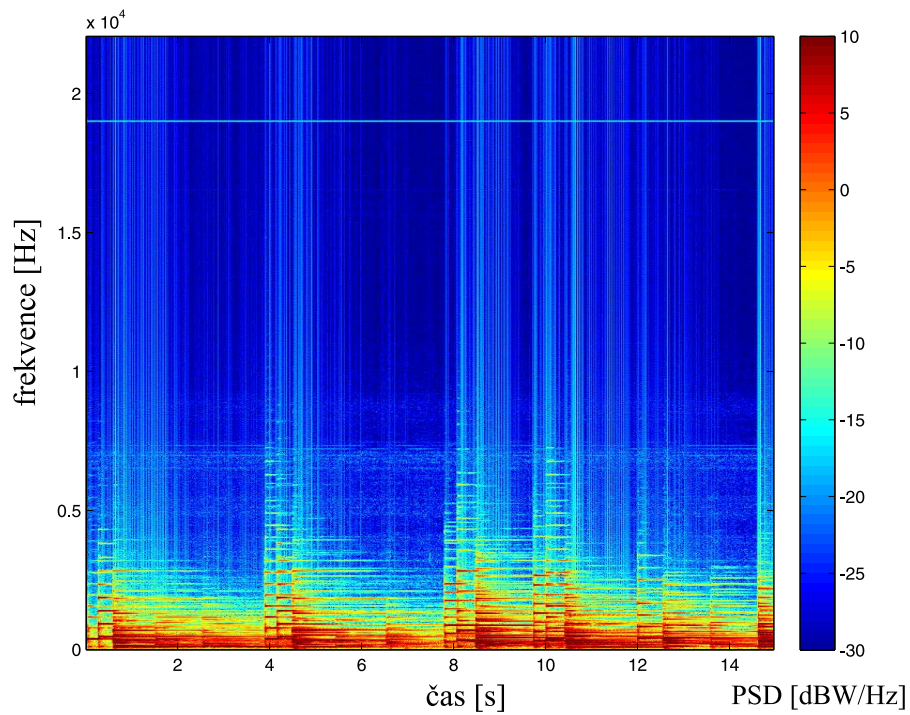
Pro tento účel byla nejprve vytvořena jednoduchá detekční funkce v programovacím jazyku Matlab. Skládá se z následujících bloků:

- načtení zvukového souboru a zjištění vzorkovací frekvence; načtení doplňkových informací k dané nahrávce (skutečné pozice onsetů, které jsou dostupné v testovací databázi),
- předzpracování signálu – pokud je nahrávka ve stereo formátu, je zkonvertována do mono formátu,
- zvolení parametrů – počet bodů algoritmu FFT, délka a typ okna a velikost překryvu oken v rámci SFFT,
- výpočet SFFT a vykreslení spektrogramu,
- implementace spektrálního toku – výpočet změn mezi dvěma sousedními segmenty spektrogramu a jejich usměrnění pro určení kladné změny spektra a následná normalizace,
- vyhledání takových časových pozic v signálu, kde se mění derivace funkce z kladné na zápornou (lokální maxima),
- nastavení prahové hranice pro zaznamenání onsetů – částečná eliminace *vibrata* a šumových složek.

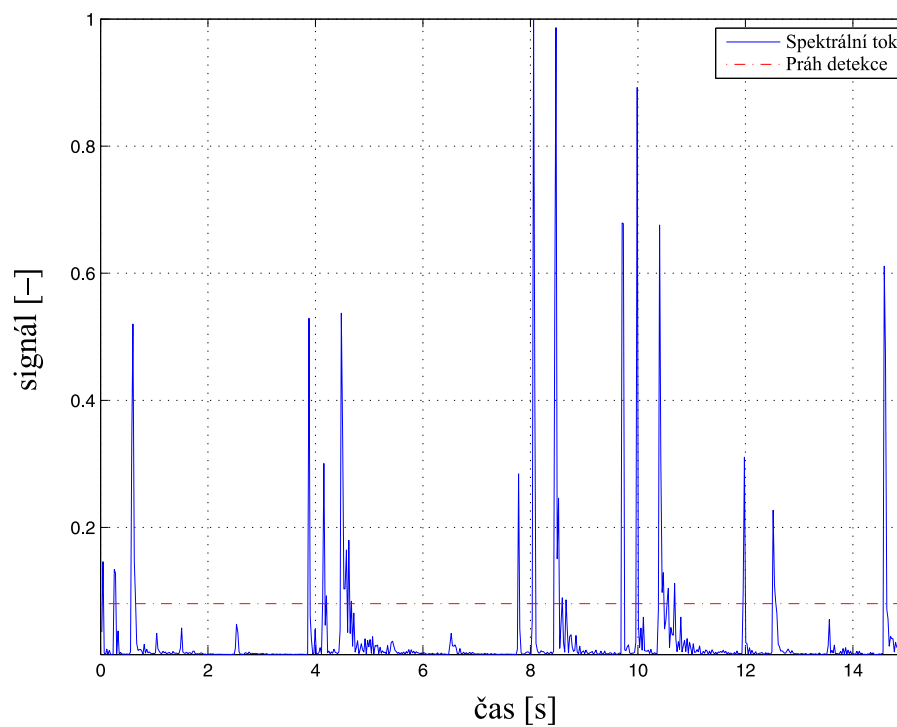
Na následujících grafech lze vidět výsledek jednotlivých sekcí při vytvoření detekční funkce – spektrogram, poté hodnoty spektrálního toku s přednastaveným prahem a nakonec výsledná detekční funkce s označením detekovaných onsetů a časové pozice skutečných onsetů v nahrávce. Zadané parametry jsou:

- délka FFT = 2048, délka okna = 2048, Hammingovo okno, 50% překryv (1024 vzorků); všechny nahrávky mají vzorkovací kmitočet $f_{vz} = 44100$ Hz.

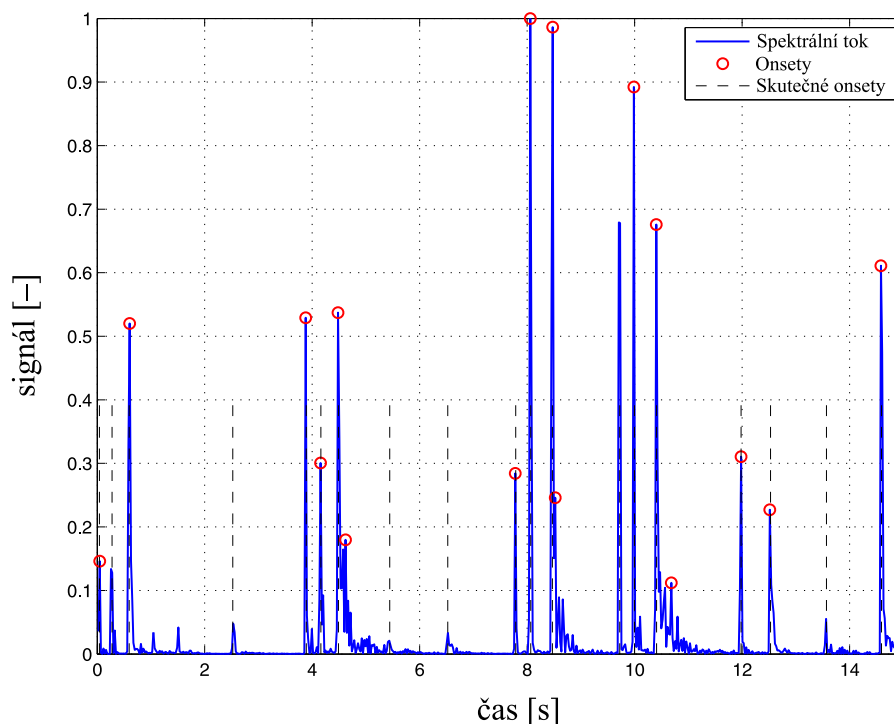
První analyzovaná nahrávka je klavírní sólový 15sekundový úsek v poměrně pomalém tempu. Pár tónů melodie pravé ruky většinou doprovází změna drženého akordu v levé ruce s očividným použitím pedálu.



Obr. 5.1: Spektrogram první testované nahrávky



Obr. 5.2: Spektrální tok první testované nahrávky

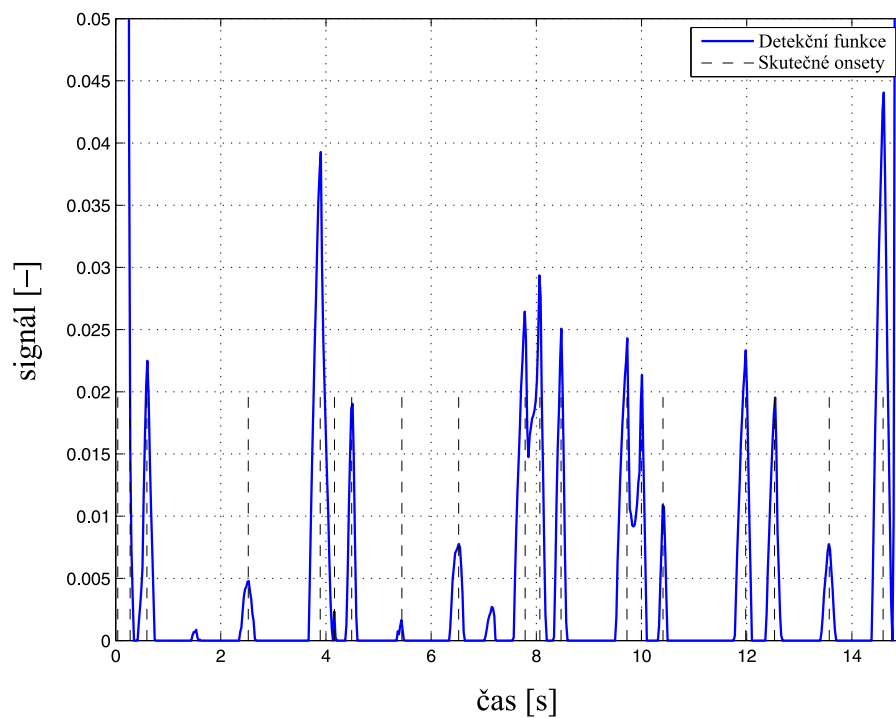


Obr. 5.3: Detekční funkce a pozice skutečných onsetů pro první testovanou nahrávku

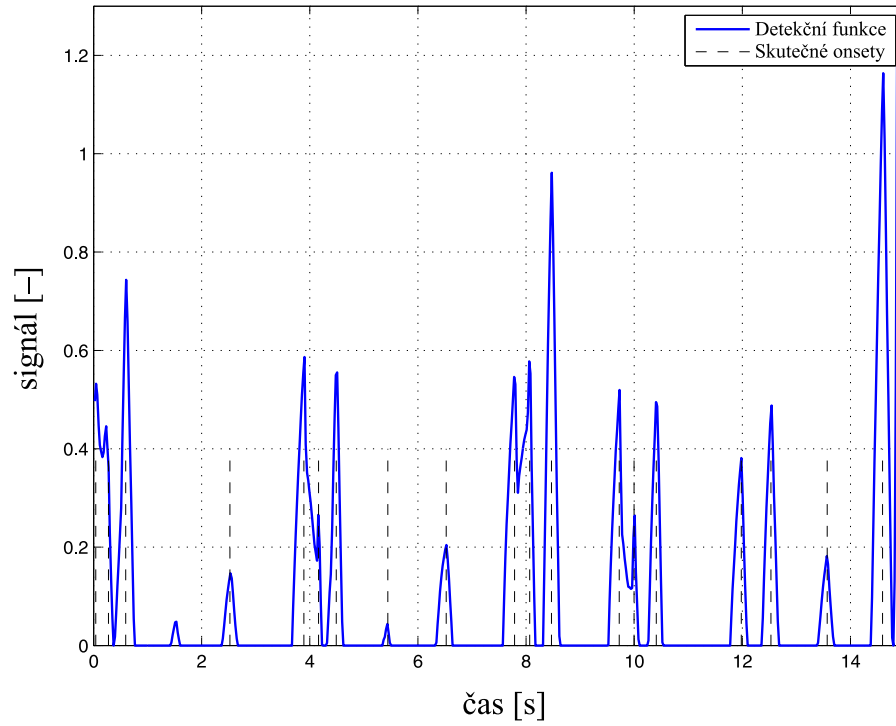
Jak lze v posledním grafu vidět, u analyzované nahrávky je celkem 19 onsetů, detekční funkce však zaznamenala pouze 16 onsetů (při počítání absolutní hodnoty výskytu). I díky typu nástroje s poměrně solidním perkusivním charakterem je možné detekovat výraznější změny spektra (například při změně akordu v levé ruce) nebo výraznější melodickou linku v pravé ruce. Některé změny, kdy jedna ruka drží akord a druhá zahraje v *piano* pouze jediný tón, nebyly touto metodou detekovány, protože nevykazují dostatečně velkou změnu v energii spektra. Jak lze na obrázku č. 5.3 vidět, 2. a 13. skutečný onset nebyl zaznamenán, přestože funkce v tomto okamžiku vykazuje větší hodnotu než jak je nastavena prahová hodnota pro detekci. Naopak 6. a 13. detekovaný onset je sice určen, ale chybně, skutečný onset se zde nenachází.

Z tohoto důvodu je aplikováno vyhlazení detekční funkce. Záleží na parametrech filtru – je vhodné zvolit takové koeficienty přenosové funkce, aby rychlé změny byly vyhlazeny, ale zároveň nedošlo ke spojení a nepřehlednosti onsetů, které se nachází blízko sebe. Často se používá adaptivní detekce špiček pro zjištění i méně výrazných lokálních maxim. Dalším důležitým aspektem implementace je její optimalizace – mimo jiné i těmito otázkami se zabírají vědecké skupiny z celého světa, které se snaží navrhnout co nejpřesnější detekční funkci v rámci daného žánru hudby nebo v obecných měřítkách (vytvoření univerzálního systému, který by se dal aplikovat na jakýkoliv styl hudby nebo hudební nástroj, bývá velmi problematické).

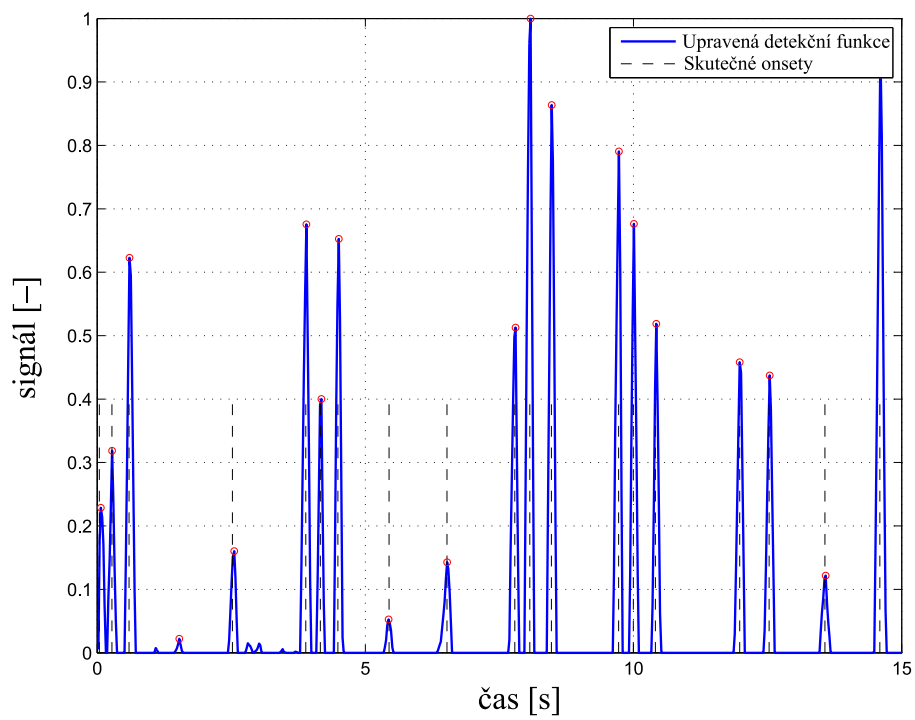
V rámci porovnání jsou zde zobrazeny další implementace detekčních metod, které jsou založeny na podobném principu, využívají ale jiný výpočet změn spektra (obr. 5.4 a obr. 5.5). Například Larocheho detekční funkce [LA-03] (obr.,5.5) lépe zachycuje méně výrazné onsety a při správném nastavení mění se prahové hodnoty správně určí naprostou většinou onsetů v předložené nahrávce klavíru. Hainsworthova funkce [HA-03] vykazovala na testované nahrávce také dobré výsledky, měla však problém s chybnými hodnotami detekční funkce pro úplný začátek nahrávky. Můžeme však vidět, že filtr je příliš pomalý a měl by být nastaven jinak, aby lépe reagoval na rychlejší změny ve spektru. Jako poslední příklad je uvedena správně vyhlazená detekční funkce s adaptabilní prahovou hodnotou pro určení onsetů – obr. č. 5.6.



Obr. 5.4: Hainsworthova detekční funkce pro první testovanou nahrávku [HA-03]



Obr. 5.5: Larocheho detekční funkce pro první testovanou nahrávku [LA-03]



Obr. 5.6: Vyhlazená Larocheho detekční funkce pro první testovanou nahrávku

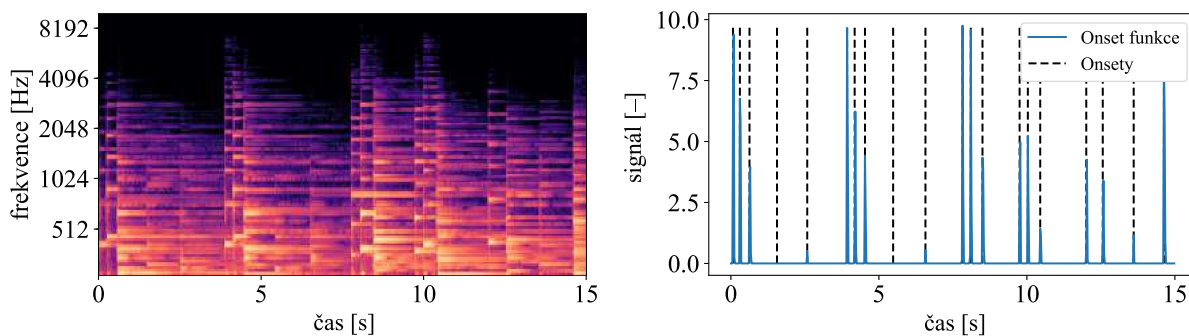
Na základě testování těchto algoritmů na některých nahrávkách MIREX databáze¹⁰ bylo zjištěno, že tyto metody jsou nedostatečné. Některé nahrávky (např. sólový part klavíru), respektive začátky tónů, byly určeny s dobrou přesností a některé (rychlé sólo houslí s výrazným dozvukem) nikoliv. Systém vykazoval výraznou chybovost (často došlo k určení onsetů na místech, kde se ve skutečnosti nevyskytovaly) v detekci onsetů u mnoha nahrávek databáze.

Proto byly testovány další dvě metody, které jsou již odladěné a funkční. Vykazují solidní úspěšnost onset detekce a zároveň jsou jejich zdrojové kódy dostupné. První využívá spektrálních změn s logaritmickou kompresí spektra a je dostupná v rámci Tempogram Toolboxu (kapitola č. 6). Druhý systém využívá psychoakustický model a speciální banku filtrů pro úpravu signálu před provedením STFT a následného spektrálního toku. Je inspirována prací S. Böcka [BÖ-13], metodou zvanou „SuperFlux“, a je dostupná v knihovně LibROSA pro programovací jazyk Python.

5.2 Implementace

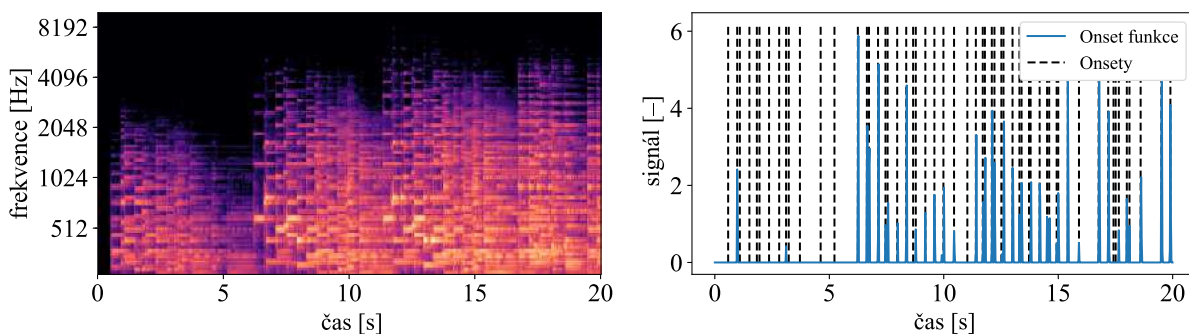
Nejlepší přesnost detekce z testovaných systémů vykazovala poslední zmíněná metoda založená na spektrálním toku a tzv. mel-scaled spektrogramech (viz kapitola č. 2.4), a proto byla vybrána pro další použití. Vstupní signál se nejprve podvzoruje na $f_{vz} = 22050$ Hz a převede do mono formátu. Poté se upraví bankou filtrů a vypočítá se STFT a spektrální tok, na který je aplikována funkce hledání lokálních maxim s adaptabilní prahovou hodnotou. Pro vyhlazení byl použit mediánový filtr. Často se také používá filtr typu horní propust s mezním kmitočtem okolo 0,4 Hz a následné vyhlazení konvolucí – Gaussovská funkce (obálka) s šířkou okolo 20 ms [EL-07]. Parametry výpočtu mel-spektra a STFT byly nastaveny následovně: minimální frekvence 50 Hz, maximální frekvence 10 kHz, 138 melovských pásem, délka FFT 2048 vzorků a překryv 512 vzorků za použití Hannova okna. Výhodou této implementace je především její potlačení rušivých vlivů ve spektru a zároveň dobrá citlivost na vznik poměrně nevýrazných tónů.

¹⁰(https://www.music-ir.org/mirex/wiki/2017:Audio_Onset_Detection)



Obr. 5.7: Spektrogram a detekční funkce první testované nahrávky

Na předchozím obrázku (obr. 5.7) můžeme vidět detekční funkci stejné testované nahrávky klavíru jako v předchozích příkladech a zmiňovaný spektrogram. Pro klavírní skladby tato metoda funguje velmi dobře, na dalším obrázku můžeme vidět 20sekundový úsek Dvořákovy Mazurky op. 56 Allegro moderato.



Obr. 5.8: Spektrogram a detekční funkce pro Dvořákovu Mazurku op. 56 Allegro moderato

Zobrazení detekční funkce není příliš přehledné, ale naprostá většina onsetů byla určena správně. K diplomové práci je přiloženo DVD, které obsahuje mj. zvukové výstupy každé zde uvedené testovací nahrávky po úpravě vybraným systémem. Zvuková stopa vždy obsahuje původní nahrávku a v místech, kde byl detekován onset, je přehráán zvuk kliku pro větší názornost. Testovány byly také nahrávky perkusivních nástrojů, houslí, kytar nebo klasické, popové, rockové a jazzové hudby.

6 Detekce tempa

Pro vytvoření reprezentace tempa, respektive pro zjištění časových okamžiků, kde se nacházejí jednotlivé doby rytmické struktury, je zapotřebí určit, jak často se onsets opakují. Většina systémů využívá detekci onsetů, i proto je jejich správné určení důležité. Ke zjištění tempa skladby lze využít několik metod, od výpočtu autokorelační funkce a hledání pravidelnosti onset detekční funkce [GR-11], [EL-07], [LA-07], přes algoritmy využívající pravděpodobnostní modely [SR-16], až po kombinaci jednotlivých metod s využitím neuronových sítí [BÖ-14].

Existuje systém s nízkou výpočetní náročností v kombinaci s poměrně dobrou přesností při detekci GT a jednoduchou implementací bez nutnosti trénovacích dat – systém D. Ellise, tzv. Beat Tracking by Dynamic Programming [EL-07]. Tento systém vybírá onset kandidáty v rámci předem vypočítané detekční funkce na základě stanoveného časového intervalu (default tempo – DT), které je možné manuálně zadat, nebo je vypočítáno automaticky na základě autokorelační funkce (ostatní onsets, které nevyhovují hranicím nastaveného tempa, se zanedbají). Nevýhodou je, že pro správné fungování je potřeba dopředu zjistit referenční GT (respektive DT), podle kterého bude algoritmus vyhledávat časové pozice dob. Pro analýzu předem neznámých skladeb se takový systém příliš nehodí, pokud však máme k dispozici notový zápis nebo víme, o jakou skladbu se jedná, může tato metoda fungovat dobře. Systém navíc poskytuje pouze omezenou možnost reagovat na rychlé změny tempa, což nemusí být problém, pokud v rámci předzpracování datasetu vybereme pouze určité charakteristické úseky, které se výrazně tempově nemění.

Další možností je využití tzv. Tempogram Toolboxu [GR-11], který je volně dostupný v programovacím prostředí Matlab. Poskytuje nejen dobrou onset detekci na základě spektrálních změn s logaritmickou kompresí spektra, ale také metodu PLP (Predominant Local Pulse). Tu lze využít k vytvoření rytmické struktury s možností zaznamenání větších rytmických změn. Algoritmus se také pohybuje v rámci nastavených hranic předem určeného tempa, testy na složitějších polyfonických nahrávkách však ukazují, že větší rytmická volnost při určování klíčových onsetů může znamenat větší chybovost detekce časových pozic dob.

Pro testování a analýzu byla nakonec zvolena metoda D. Ellise – Beat Tracking by Dynamic Programming [EL-07]. Systém využívá výpočtu periodicity onsetů v rámci předem definovaného tempa nebo pomocí autokorelační funkce. Hledá vhodné onset kandidáty s určitou volností k zadanému tempu a doplňuje pravidelnou rytmickou strukturu i tam, kde žádné onsets nebyly detekční funkcí zaznamenány. Ostatní onsets potlačuje, a tím vytváří rytmickou reprezentaci s určením časových pozic jednotlivých dob – nerozlišuje však metrickou strukturu (nerozděluje takty, ani nedetekuje důrazy na specifické doby v rámci taktů). Může dobře určit GT celého

hudebního úseku nebo motivu, pro automatickou hudební transkripci se však nehodí. V původní práci D. Ellise je použit odlišný detektor onsetů, než navrhuje tato práce. Rozdíly jsou především v menším podvzorkování (původně 8 kHz), jiné volbě parametrů melovské reprezentace, STFT a způsobu vyhlazení detekční funkce.

6.1 Datasetsy

Pro testování funkčnosti systému byla použita *SMC_MIREX* databáze [HO-12], která obsahuje 217 nahrávek různých hudebních žánrů a nástrojů. Nahrávky mají vzorkovací frekvenci 44,1 kHz a každá obsahuje metadata s ručně upravenými časovými pozicemi dob. Tento referenční dataset tedy může sloužit k zjištění efektivity a přesnosti navrženého algoritmu.

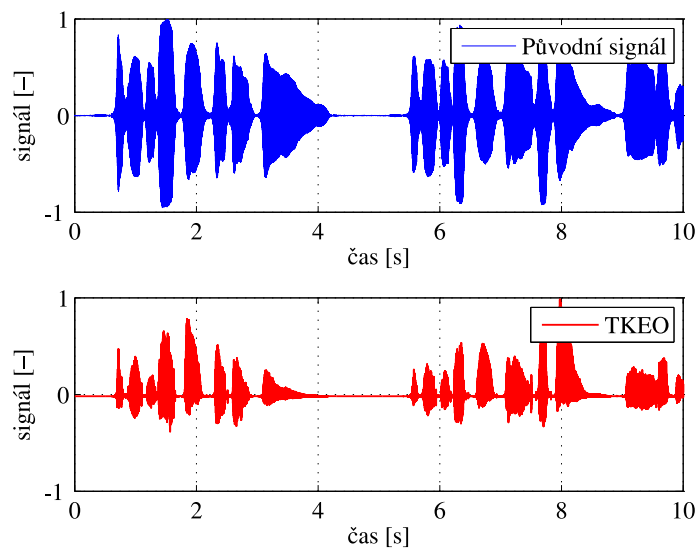
Druhý dataset je složen z 33 různých interpretací první věty skladby Bedřicha Smetany *Smyčcový kvartet No. 1 e moll „Z mého života“* včetně dvou orchestrálních verzí (CD28 a CD29). Skladba má sonátovou formu, proto byla rozdělena do 6 hudebních úseků – Expozice (hlavní a vedlejší téma), malá coda před Provedením, Provedení, Repríza a závěrečná coda. Všechny nahrávky byly manuálně rozděleny na jednotlivé motivy v DAW (Digital Audio Workstation) a každý úsek byl zvlášť vyexportován ve formátu *wav* (vzorkovací frekvence 44,1 kHz, 16bitové rozlišení).

6.2 Návrh vylepšení metody

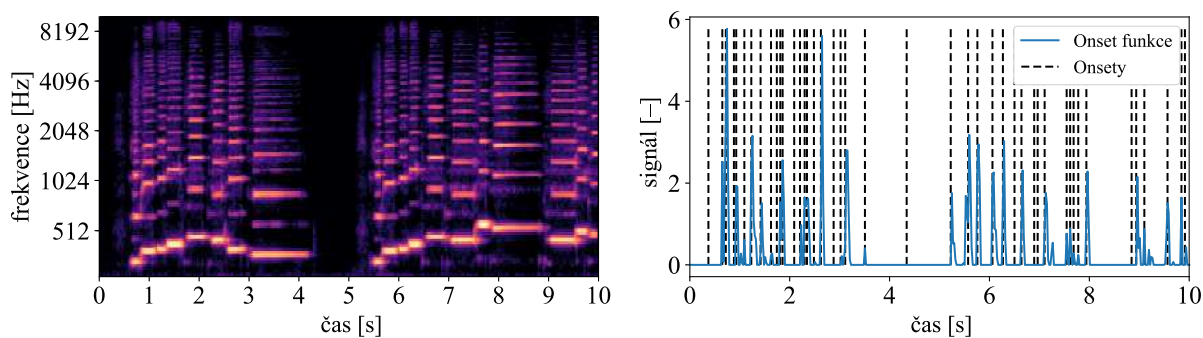
Ve fázi přezpracování signálu před vytvořením časově-frekvenční reprezentace se používají různé metody pro redukci redundantní informace nebo vylepšení výsledné detekční funkce (kap. 2.3). V rámci testování onset detektoru byl vyzkoušen také Teagerův-Kaiserův energetický operátor (TKEO) – $\Psi(x[n])$. Jeho definice je následující [SM-16]:

$$\Psi(x[n]) = x^2[n] - x[n-1] \cdot x[n+1] \quad (6.1)$$

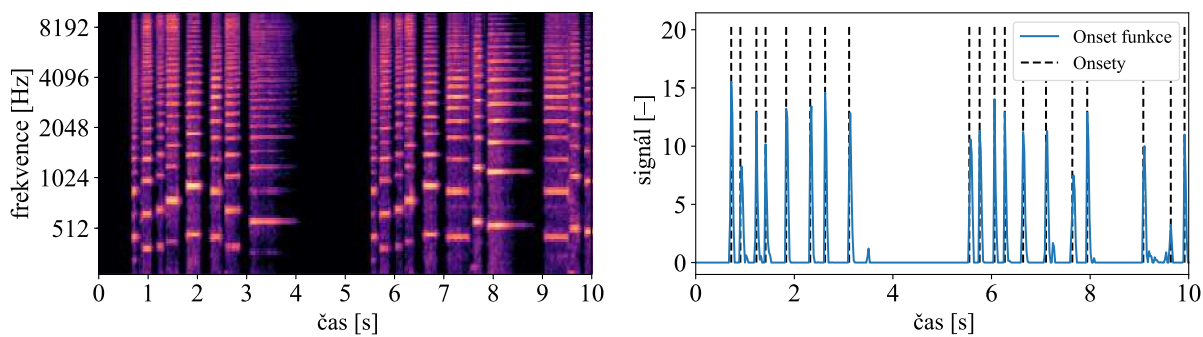
Je to nelineární energetický operátor, který dokáže vytvořit energetickou obálku daného signálu. Může nabývat i záporných hodnot a bere v potaz také frekvenci signálu [DI-09]. Používá se například pro vylepšení onset detekce v EMG signálech [SO-08], ale pro onset detekci tónů v hudebních signálech nikoliv. Po implementaci TKEO do systému došlo u některých nahrávek ke zpřesnění detekce, u některých naopak ke zhoršení přesnosti. Celkové výsledky jsou sporné. Na následujících grafech je vyzobrazen nejprve původní signál v porovnání s TKEO, a poté detekční funkce systému bez TKEO i s TKEO pro nahrávku klarinetového sóla. Pro větší přehlednost bude systém bez použití TKEO pojmenován ODF (Onset Detection Function).



Obr. 6.1: Původní vstupní signál a TKEO nahrávky klarinetového sóla

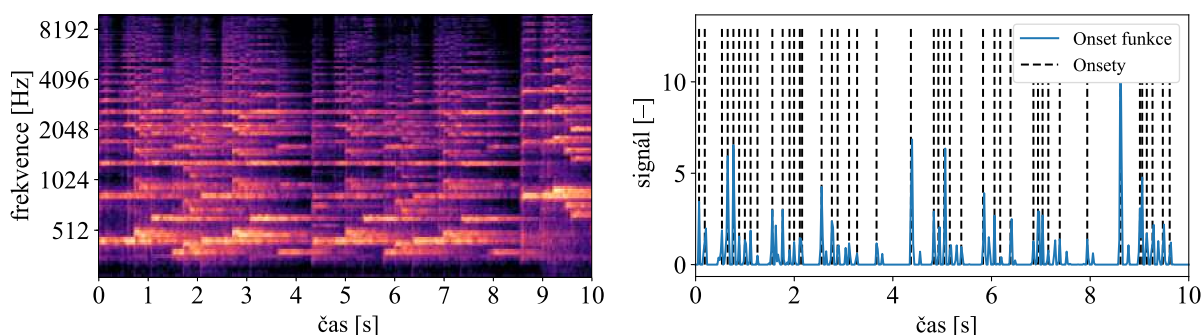


Obr. 6.2: Spektrogram a detekční funkce klarinetového sóla, ODF

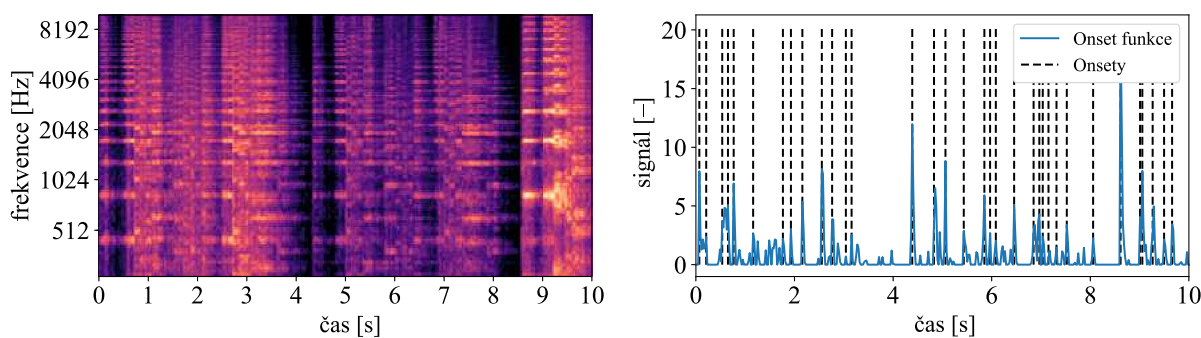


Obr. 6.3: Spektrogram a detekční funkce klarinetového sóla, TKEO

Jak lze vidět, ODF vykazoval vysokou chybovost v detekci klarinetu. Stejný systém, jen s přidáním TKEO, určil skoro všechny onsety správně. Na dalších obrázcích je znázorněn opačný případ, kdy ODF určil pozici onsetů lépe než TKEO. Samozřejmě záleží na nastavení parametrů například váhovacího okna, zdá se však, že TKEO nedokáže při stejném nastavení dostatečně reagovat na rychlé nebo energeticky příliš slabé změny. Houslové sólo v testované nahrávce je zahráno ve velmi rychlém tempu a vyskytují se zde i dvaatřicetinové tóny. Onset funkce systému ODF je v tomto případě přesnější. Testování onset detektoru nebylo uvedeno v číselných hodnotách, protože je důležitější závěr z testování detektoru tempa.



Obr. 6.4: Spektrogram a detekční funkce houslového sóla, ODF



Obr. 6.5: Spektrogram a detekční funkce houslového sóla, TKEO

Pro detekci tempa ale nemusí být nutně potřebné zachytit každý onset správně. Je důležité detekovat významné onsety, které mají vlastnost určit rytmickou strukturu. Proto i přes nedostatečné vylepšení onset detekční funkce na větším množství nahrávek je možné, že systém s TKEO bude vykazovat lepší výsledky v rámci detekce tempa u komplexnějších polyfonních nahrávek. Z tohoto důvodu byl na referenční databázi testován systém ODF i TKEO.

7 Testování metod

Pro testování byla vybrána *SMC_MIREX* databáze [HO-12] (viz kap. 6.1). Výsledek obou systémů je vždy porovnán s referenční hodnotou v rámci anotace datasetu. Některé nahrávky v databázi chybí (například č. 20, 25, 29 a 31), a proto je jejich číslování zanedbáno – nahrávka č. 20 ve skutečnosti odpovídá nahrávce s popisem *SMC_021*. Veškeré hodnoty jsou vloženy v excel sešitu přiloženého DVD.

Tab. 7.1: Testování výpočtu GT na referenční *SMC_MIREX* databázi

Index	Ref (BPM)	TKEO (BPM)	ODF (BPM)	Δ TKEO (%)	Δ ODF (%)	Δ TKEO (BPM)	Δ ODF (BPM)
1	48,12	47,85	47,85	0,57	0,57	0,27	0,27
2	66,10	73,83	73,83	10,47	10,47	7,73	7,73
3	65,56	95,70	95,70	31,50	31,50	30,14	30,14
4	55,64	48,75	68,00	14,14	18,17	6,89	12,36
5	40,04	42,36	42,36	5,48	5,48	2,32	2,32
6	59,40	47,85	47,85	24,14	24,14	11,55	11,55
7	52,25	54,98	54,98	4,96	4,96	2,73	2,73
8	135,08	136,00	143,55	0,67	5,90	0,92	8,47
9	55,79	56,17	56,17	0,67	0,67	0,38	0,38
10	76,01	80,75	78,30	5,87	2,92	4,74	2,29
11	89,26	95,70	95,70	6,73	6,73	6,44	6,44
12	86,85	86,13	184,57	0,83	52,95	0,72	97,72
13	95,67	99,38	95,70	3,73	0,03	3,71	0,03
14	77,61	86,13	89,10	9,89	12,89	8,52	11,49
15	35,18	42,36	42,36	16,94	16,94	7,18	7,18
16	70,09	66,26	66,26	5,77	5,77	3,83	3,83
17	71,93	73,83	73,83	2,57	2,57	1,90	1,90
18	83,50	89,10	117,45	6,29	28,91	5,60	33,95
19	40,78	46,98	42,36	13,19	3,72	6,20	1,58
20	88,59	99,38	123,05	10,86	28,01	10,79	34,46
.
.
.
30	62,81	63,02	63,02	0,33	0,33	0,21	0,21
Průměr	76,20	83,75	88,97	11,22	16,11	9,75	16,52

Bylo testováno 30 nahrávek různých hudebních žánrů a instrumentů. Pro systém ODF i systém s TKEO byly použity stejné parametry – DT každého záznamu bylo dopředu stanoveno a nastaveno dle referenčního GT z anotace databáze. Tab. 7.1 ukazuje referenční GT, výpočet GT pomocí systému s TKEO a ODF a rozdíl výpočtu od referenčního tempa v procentech i BPM. Systém s TKEO vykazuje průměrnou odchylku od reference 9,75 BPM (11,22 %) a ODF 16,52 BPM (16,11 %). Je zde tedy patrné obecné vylepšení systému pro detekci GT hudební skladby. V některých případech (například nahrávka č. 12 v tabulce) je odchylka systému ODF velmi vysoká, ale systém s TKEO určil GT přesně. Jedná se právě o vlastnost TKEO potlačit některé méně významné onsety při stejném nastavení jako u metody ODF. Systém ODF by pravděpodobně určil více onsetů v rámci dané skladby, ale pro určení tempa to nemusí být dostatečně relevantní. Pokud by se nastavil systém ODF s menším DT, pravděpodobně by došlo k výpočtu pozic dob, které by více odpovídaly referenčnímu tempu skladby.

Tab. 7.2: Výsledky testování GT referenční databáze

typ	medián	\bar{x}	sd	rsd	var
Ref	68,09	76,20	33,36	0,44	1113,16
TKEO	82,05	83,75	37,30	0,45	1391,05
ODF	76,07	88,97	41,05	0,46	1685,16
Δ TKEO	6,32	9,75	11,29	1,16	127,52
Δ ODF	9,98	16,52	20,05	1,21	402,15

V příložené tabulce Tab. 7.2 je uveden medián, aritmetický průměr (\bar{x}), směrodatná odchylka (sd), variační koeficient (rsd) a rozptyl (var) referenčního GT datasetu obou testovaných systémů a absolutních odchylek výpočtu systémů od referenčního GT v jednotkách BPM. I zde lze vidět, že systém s TKEO vykazuje lepší výsledky určení GT.

U detekce GT vykazovaly obecně nejvyšší chybovost nahrávky expresivních sólových partů akustické kytary (SMC_3, 6, 15, 22, 24, 28 a 33), zpěvu s doprovodem (SMC_27), orchestrálních verzí proměnlivé filmové hudby (SMC_32) a tempově se měnícího sóla klavíru (SMC_23), které dosahovalo vůbec největší odchylky detekce systému TKEO v rámci testovaného datasetu. U systému ODF byla určena nejvyšší odchylka u skladby zpěvu s doprovodem (SMC_12). Zde byl také největší rozdíl v určení tempa oběma systémy (TKEO – 0,83 %, 0,72 BPM a ODF – 52,95 %, 97,72 BPM). ODF určil velkou hustotu onsetů v nahrávce a i přes zadané referenční tempo, vypočítal GT nepřiměřeně vysoké (184,57 BPM ku referenčním 86,85 BPM).

Systém TKEO naopak určil GT velmi přesně – s odchylkou 0,72 BPM. Je zajímavé, že i u další nahrávky sólového zpěvu s doprovodem (SMC_30) byl TKEO systém velmi přesný – odchylka zde dosahovala hodnoty 0,03 % (0,06 BPM), což je vůbec nejpřesnější výsledek z celé databáze. ODF určil tempo nahrávky s relativní odchylkou 7,66 % (14,14 BPM). Třetí skladba zpěvu s doprovodem (SMC_9) byla vyhodnocena oběma systémy stejně (0,67 %, 0,38 BPM).

Systém ODF přesněji detekoval GT v rámci testovaného datasetu pouze u 4 nahrávek (SMC_10, 13, 19, 22). Až na poslední nahrávku (sólo akustické kytary) se jedná o různé skladby s instrumentací připomínající menší orchestr (sekce houslí, varhany, trubky). Nebyla zde pozorována žádná souvislost mezi přesnější detekcí průměrného tempa systémem ODF a druhem hudby, hudebním nástrojem či interpretací.

Pro systém TKEO i ODF byla v případě nahrávek sóla akustické kytary naměřena odchylka 11,60–31,50 %. V absolutním měřítku pak 7,18–43,41 BPM. Pro systém ODF 16,15–31,50 % (7,18–43,41 BPM). Nejnižší nebo nejvyšší hodnota relativní odchylky při uvedeném rozsahu nemusí vždy odpovídat té absolutní. U některých nahrávek byla relativní odchylka vyšší, skladba však byla v pomalém tempu a v absolutním měřítku dosahovala nízkých hodnot.

Limitace tohoto testování spočívá ve skutečnosti, že dochází k výpočtu průměrného tempa měřeného úseku, ale ne k určení přesnosti detekce jednotlivých časových pozic dob. Aby systém fungoval správně pro analýzu interpretačního výkonu, je zapotřebí identifikace jemných změn v agogice přednesu, a to kalkulace GT nezajistí. Následující porovnání je zaměřeno na odchylky mezi referenčními hodnotami časových pozic dob nahrávek datasetu a výstupem předložených systémů.

7.1 Testování výpočtu pozic dob

Referenční dataset obsahuje přesné časové pozice jednotlivých dob. Ty byly porovnány s výstupy systému ODF i TKEO. U absolutních odchylek pozic dob v sekundách byl vypočítán průměr a medián v rámci každé nahrávky a následně byl určen průměr a medián těchto hodnot celého testovaného datasetu dohromady. V následujících tabulkách jsou vyneseny výsledky v sekundách.

Tab. 7.3: Testování rozdílů pozic dob nahrávek referenčního datasetu

	TKEO	ODF
Průměr	2,30	2,84
Medián	1,81	2,57
sd průměru	1,90	2,17
sd mediánu	2,14	2,31

Průměrná odchylka mezi vypočítanou a referenční hodnotou je 2,30 s u systému s TKEO a 2,84 s pro ODF. To je velmi vysoké číslo a je zkresleno především nahrávkami, které podle systémů vykazují velmi odlišné tempo od reference. Proto byly následně vybrány pro další analýzu ty nahrávky, které v podstatě kopírují stanovené tempo – s mediánem nebo průměrnou odchylkou menší než 1 s. Nejprve byly zvoleny nahrávky splňující tuto podmínku (soubory, které splňují podmínku u systému s TKEO jsou jiné než u systému ODF), poté byl určen aritmetický průměr a také medián (Me) průměru časových rozdílů dob mezi systémy a referencí.

Tab. 7.4: Výsledky testování rozdílů pozic dob u nahrávek s mediánem menším než 1 s u systému TKEO a ODF

	Pro rozdíl < 1 s u mediánu TKEO		< 1 s u mediánu ODF	
	TKEO	ODF	TKEO	ODF
	\bar{x}	Me	\bar{x}	Me
Průměr	0,67	0,52	1,22	0,71

Tab. 7.4 vyjadřuje průměr hodnot aritmetického průměru a mediánu časových rozdílů dob u nahrávek, které dosahovaly mediánu rozdílu menšího než 1 s. Nahrávky byly vybrány nejprve podle hodnot systému TKEO (14 nahrávek – 47 % testované databáze), a poté podle ODF (pouze 9 z 30 nahrávek splňovalo tuto podmínku

– 30 % testované databáze). Směrodatná odchylka průměru časových rozdílů dob vybraných nahrávek je v návaznosti na tabulku postupně 0,63 a 0,33 pro TKEO, 1,38 a 0,32 pro ODF.

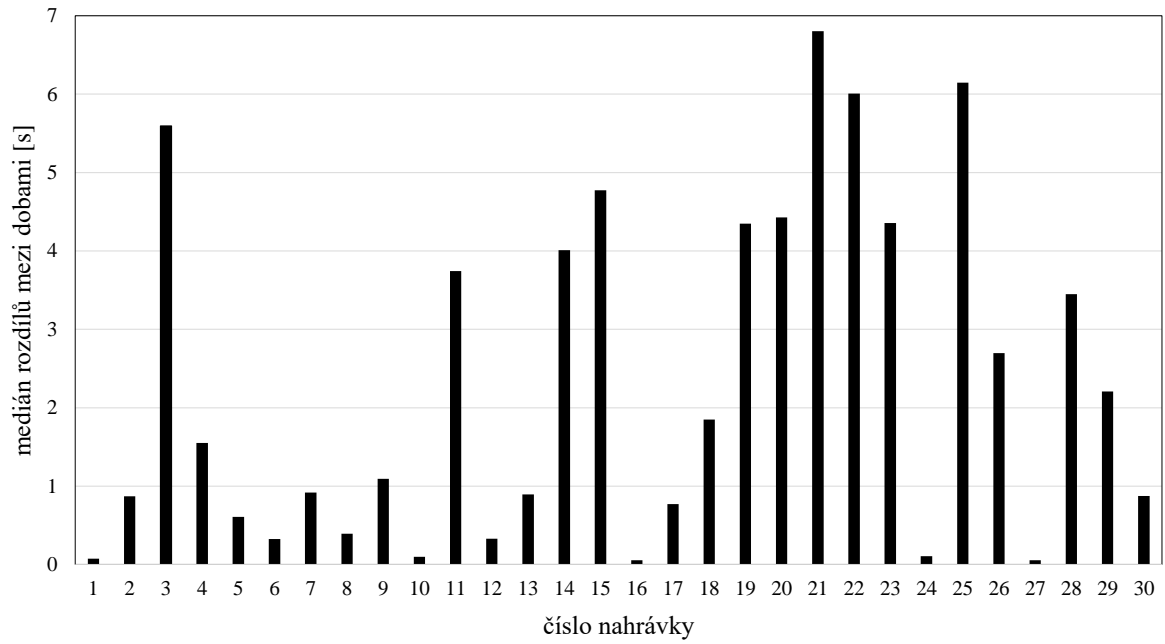
Tab. 7.5: Výsledky testování rozdílů pozic dob u nahrávek s průměrem menším než 1 s u systému TKEO a ODF

	Pro rozdíl < 1 s u průměru TKEO				< 1 s u průměru ODF			
	TKEO		ODF		TKEO		ODF	
	\bar{x}	Me	\bar{x}	Me	\bar{x}	Me	\bar{x}	Me
Průměr	0,39	0,38	0,95	0,70	0,29	0,12	0,36	0,22

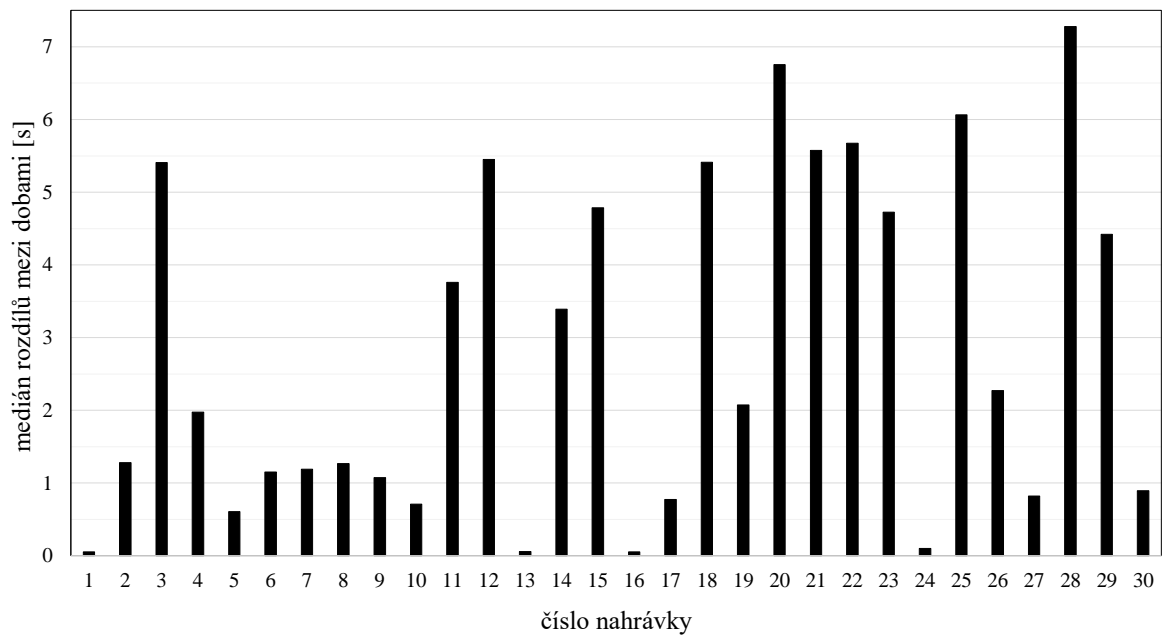
Tab. 7.5 vyjadřuje průměr hodnot aritmetického průměru a mediánu časových rozdílů dob u nahrávek, které dosahovaly průměrného rozdílu menšího než 1 s. Hodnoty systému TKEO – 11 nahrávek (37 % testované databáze) a ODF – 9 nahrávek – (30 % testované databáze). Směrodatná odchylka průměru časových rozdílů dob vybraných nahrávek je v návaznosti na tabulku postupně 0,30 a 0,31 pro TKEO, 1,37 a 0,30 pro ODF.

Je zapotřebí si uvědomit, že pokud se výpočet pozic dob u nějaké skladby začal odchylovat (určené referenční tempo se lišilo), pak začal časový rozdíl výrazně narůstat. Nastaly tedy situace, kde se každá z posledních 40 hodnot lišila o více jak 3 s – průměr a medián časových odchylek u některých nahrávek je vysoký, přestože GT bylo vypočítáno podobně. I proto byly odděleny a samostatně vypočítány hodnoty pro skladby, které se drží referenčního tempa (odchylka rozdílů mezi dobami menší než 1 s). Na obrázcích obr. 7.1 a obr. 7.2 je zobrazen medián odchylek výpočtů obou systémů vůči referenčním časovým hodnotám.

Výsledky ukazují, že použitý systém využívající TKEO ve fázi předzpracování signálu dokáže obecně přesněji určit nejen GT, ale i časové pozice jednotlivých dob v rámci skladby. V určení GT byla průměrná odchylka systému s TKEO 9,75 BPM a u systému ODF 16,52 BPM od referenčního tempa. U porovnání výpočtu časových pozic jednotlivých dob TKEO systém vykazoval v průměru o 0,54 s přesnější detekci časové pozice k dané referenční poloze každé doby. Systém s použitím TKEO dosahuje obecně nižších průměrných odchylek od referenčních hodnot a má vyšší procentuální zastoupení nahrávek v rámci předložené databáze, u kterých byl detekován průměr nebo medián rozdílů časových pozic dob menší než 1 s, než systém ODF. Zároveň mnohem lépe detekuje GT u nahrávek zpěvu s doprovodem – je však zapotřebí dalšího testování na větším počtu nahrávek, aby byl tento závěr v obecném měřítku potvrzen.



Obr. 7.1: Medián časových rozdílů mezi vypočítanými pozicemi dob pro systém TKEO vůči referenci



Obr. 7.2: Medián časových rozdílů mezi vypočítanými pozicemi dob pro systém ODF vůči referenci

7.2 Analýza datasetu

Ve spolupráci s Ústavem hudební vědy Masarykovy univerzity v Brně byla vytvořena databáze 33 různých interpretací převážně smyčcových kvartetů první věty skladby Bedřicha Smetany *Smyčcový kvartet No. 1 e moll „Z mého života“* (kap. 6.1). Tab. A.1 obsahuje informace o interpretech a roku vzniku nahrávek. Každá nahrávka byla posléze rozdělena do 6 úseků dle sonátové formy. Motivy byly nejprve změřeny a bylo vypočítáno tzv. odhadnuté tempo. Výpočet vyšel ze znalosti notového zápisu (obr. č. D.1), respektive počtu dob v jednotlivých úsecích, a změřené délky trvání hudebních motivů. Odhadnuté tempo t_o lze určit jako:

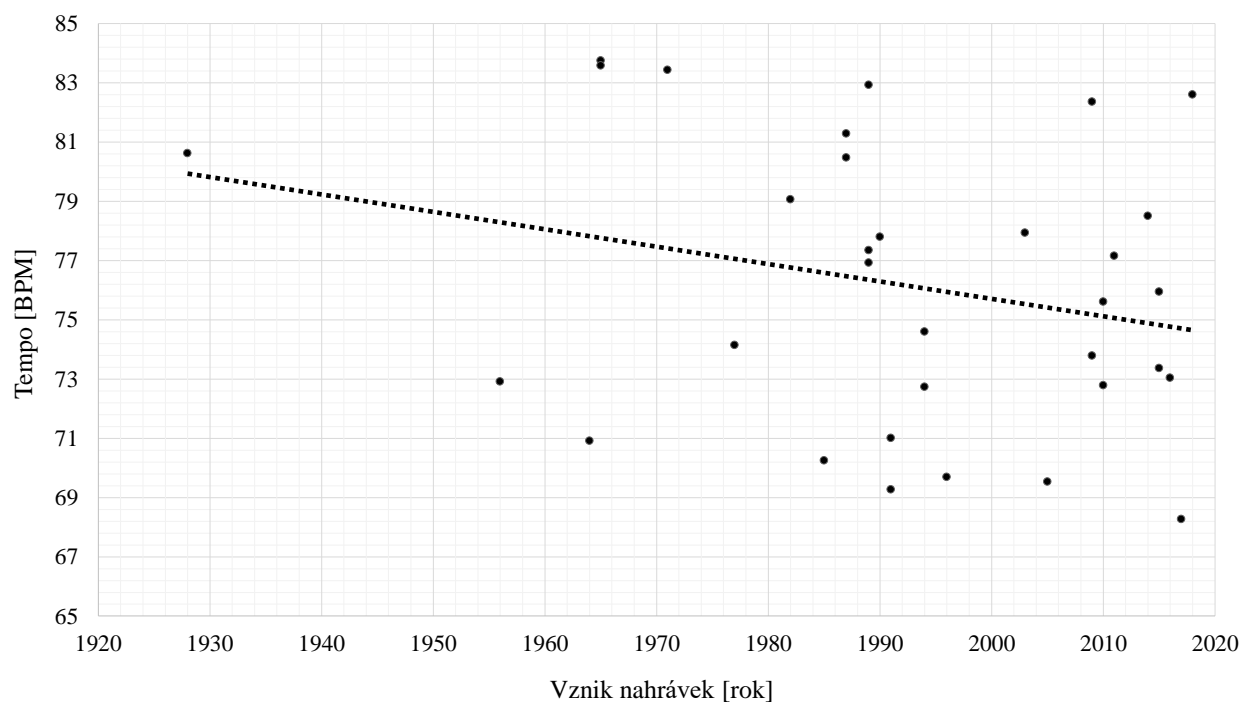
$$t_o = \frac{\frac{d}{2} \cdot 60}{t} \quad (7.1)$$

kde d je počet dob v měřeném motivu a t značí časovou délku úseku. Ve vzorci je d poděleno dvěma, protože celá první věta skladby je chápána jako tzv. *tactus alla breve* – počet dob byl určen čtvrtovou notou (v tomto případě jsou čtyři čtvrtové noty v jednom taktu), ale jeden takt je ve skutečnosti na dvě doby (dvoupůlový). Tab. 7.6 znázorňuje počet taktů a dob všech segmentů. V tabulkách je označení hlavního tématu Expozice – Zacatek, vedlejšího tématu Expozice – A, malé cody – B, Provedení – C, Reprízy – D a závěrečné cody – E.

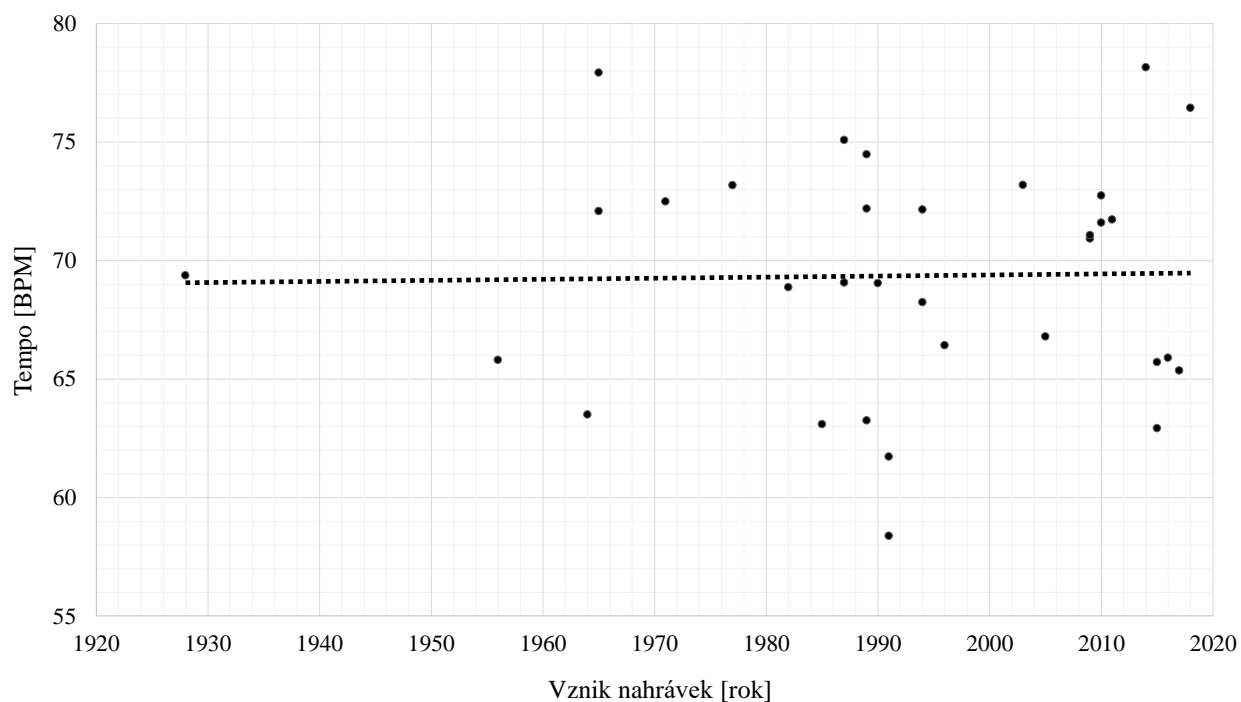
Tab. 7.6: Určení počtu taktů a dob jednotlivých úseků

Motiv	Zacatek	A	B	C	D	E
Takty	1–70	71–110	111–118	119–164	165–225	226–262
Čtvrtkové noty	280	160	32	184	244	148

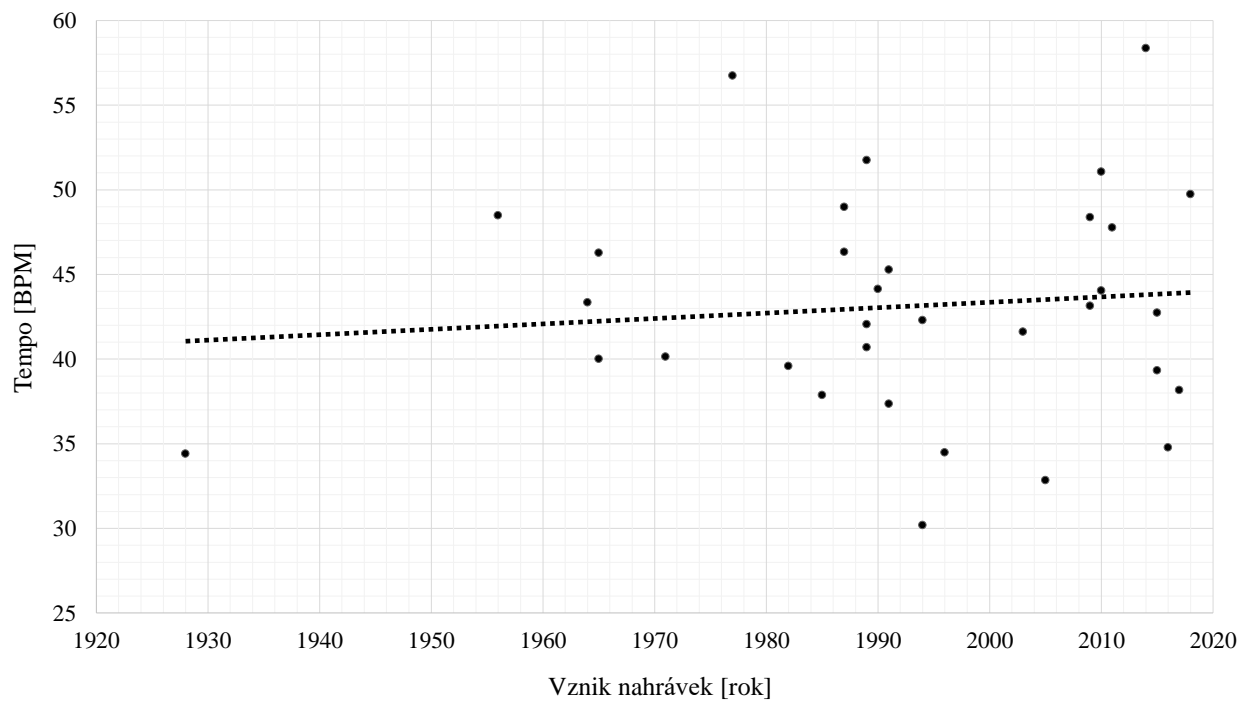
Zde se naskytá otázka – proč není pro analýzu datasetu použit implementovaný algoritmus pro detekci tempa? Je zde potřeba dosáhnout co možná nejvyšší přesnosti detekce délky a tempa úseků. Předkládaná „manuální“ metoda nevypočítá přesné pozice dob, ani nezaznamená agogický vývoj hudebního úseku. Dokáže však teoreticky určit průměrné tempo motivu, do výpočtu se nezanáší chyby vzniklé při výpočtu změn spektra nebo při výběru vhodných onsetů. Dokud nebude automatický systém vykazovat extrémní přesnost detekce nebo neproběhne manuální korekce výstupu systému u každé nahrávky, muzikologové sáhnou po přesnějších datech, protože ty mohou lépe odhalit různé trendy, o kterých lze diskutovat a vyvozovat z nich závěry. Tab. C.1 předkládá odhad tempa pro každý motiv všech nahrávek databáze. Obrázky č. 7.3–7.8 znázorňují závislost odhadnutého tempa na roku vzniku nahrávek včetně uvedené spojnice trendu.



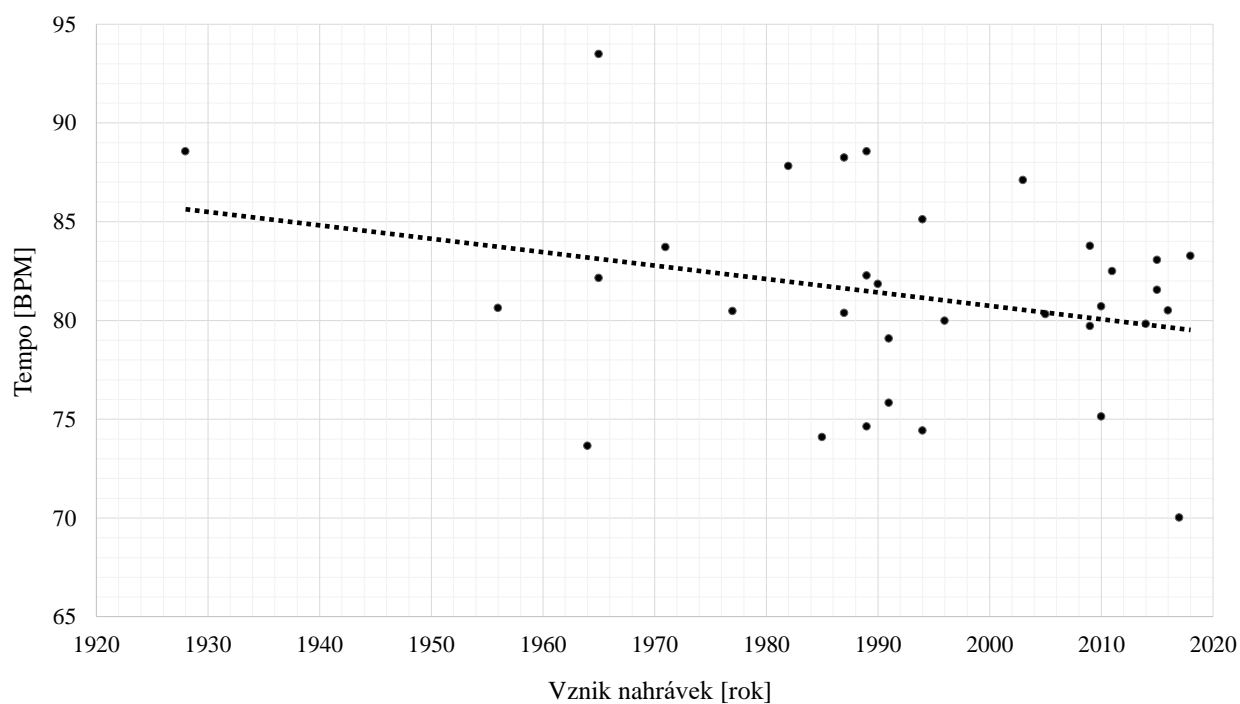
Obr. 7.3: Odhadnuté tempo vůči roku vzniku nahrávky – Expozice, hlavní téma (Zacatek)



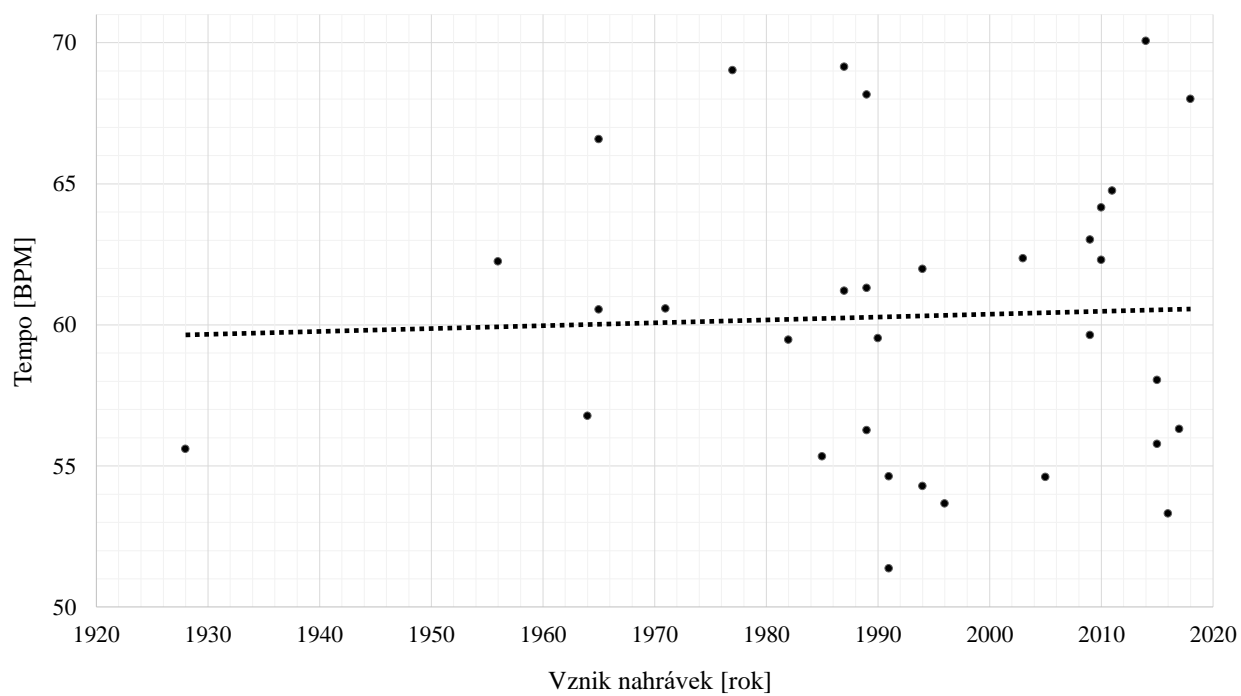
Obr. 7.4: Odhadnuté tempo vůči roku vzniku nahrávky – Expozice, vedlejší téma (A)



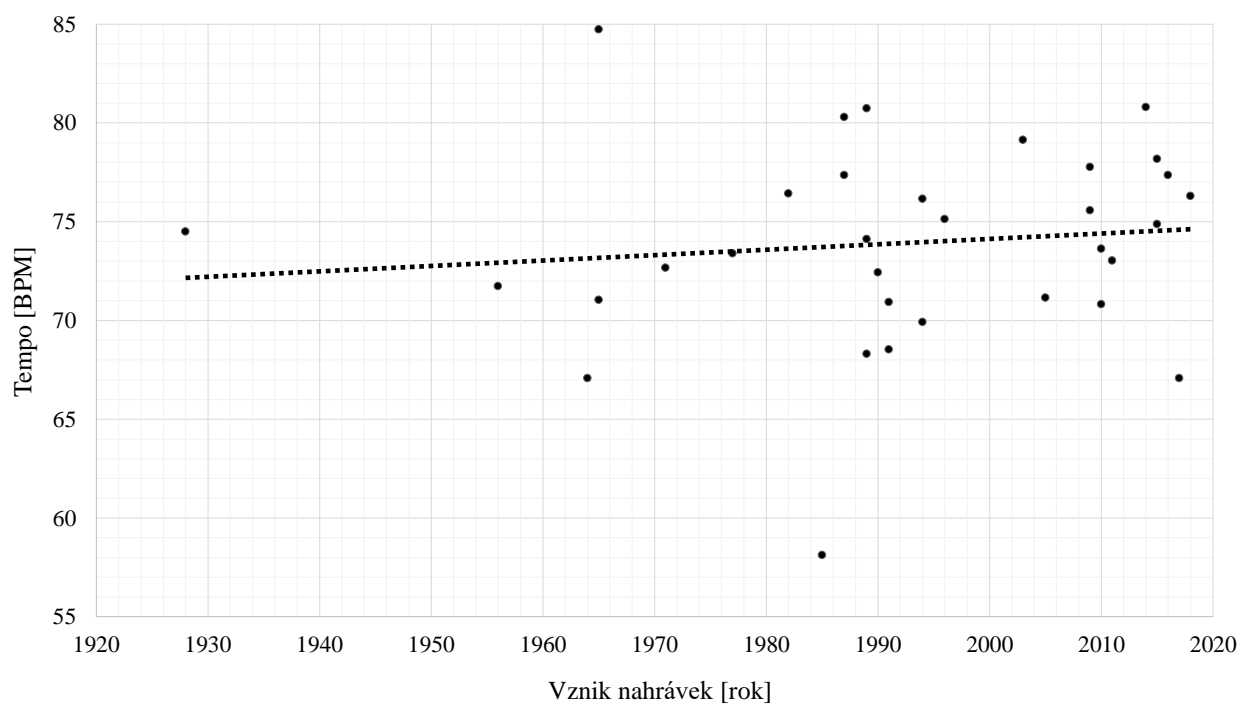
Obr. 7.5: Odhadnuté tempo vůči roku vzniku nahrávky – malá coda (B)



Obr. 7.6: Odhadnuté tempo vůči roku vzniku nahrávky – Provedení (C)



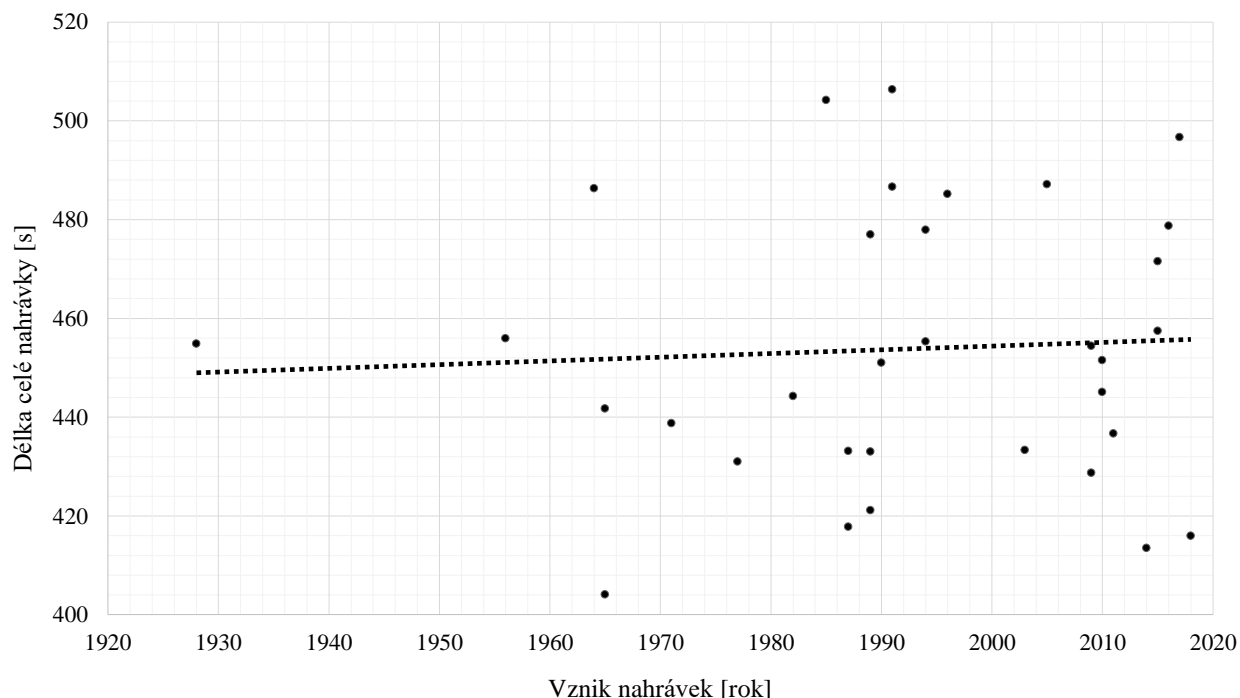
Obr. 7.7: Odhadnuté tempo vůči roku vzniku nahrávky – Repríza (D)



Obr. 7.8: Odhadnuté tempo vůči roku vzniku nahrávky – závěrečná coda (E)

Dalším důvodem, proč nepoužívat odhadnuté tempo, je jeho možná nepřesnost. Odhadnuté tempo totiž nezahrnuje expresivní výrazové prostředky jako hudební ko-runy nebo cézury. Takový výpočet tedy nelze považovat za spolehlivou referenci pro systém a výsledek automatické metody by musel být porovnán s manuálně uprave-nými hodnotami, které prozatím nejsou k dispozici. Přesto je použití vypočítaného odhadu v tomto případě výhodnější.

V hudebních kruzích existuje domněnka, že se interpretace hudebních děl obecně během let zrychluje. Právě na základě objektivního měření délky nebo tempa inter-pretací v rámci větší sady nahrávek lze tuto domněnku vyvrátit nebo potvrdit. Přestože motivy A, B, D i E vykazují s přibývajícimi roky zrychlení, v celkovém měřítku dochází ke zpomalování celé skladby – zpomalení je výraznější u hlavního tématu Expozice a Provedení. Je zřejmé, že databáze obsahuje mezi lety 1982–2018 větší množství nahrávek než mezi lety 1928–1981 (z tohoto období je obtížné získat relevantní nahrávky, pokud tedy vůbec existují), a proto konstatování naprosto jed-noznačného trendu není úplně na místě. Přesto výsledky naznačují, že dochází spíše ke zpomalování hlavního tématu Expozice, celé Expozice dohromady a Provedení. Ostatní motivy se mírně zrychlují, ale méně výrazně – tudíž celková délka první věty skladby Bedřicha Smetany *Smyčcový kvartet No. 1 e moll „Z mého života“* se spíše zvětšuje a celkové průměrné tempo skladby klesá. Obr. 7.9 ukazuje, jak se délka celé skladby v závislosti na pozdějším roku vzniku nahrávky obecně mírně zvětšuje.



Obr. 7.9: Závislost časové délky každé interpretace první věty skladby a roku vzniku dané nahrávky

8 Závěr

V diplomové práci je nejprve uvedena a popsána hudební a technická terminologie, která se v odvětví MIR často používá. Následně jsou rozebrány metody detekce tónů v hudebním signálu – především se zaměřením na detekce na bázi spektrálních změn. V kap. č. 4 je předložen výčet dostupných zdrojů a knihoven pro extrakci nízkourovňových a vysokoúrovňových parametrů z hudebních signálů v rámci různých programovacích prostředí. Z některých práce následně čerpá a využívá jejich funkce.

V následující kapitole je navržen, implementován a testován onset detektor v prostředí Matlab. Pro některé nahrávky (především sólový klavír) testované databáze tento detektor fungoval dobře, obecně však vykazoval příliš velkou chybovost, a to především kvůli nedostatečné úpravě signálu před provedením STFT. Další uvedené a testované metody využívají filtraci pomocí banky filtrů nebo logaritmickou kompresi spektra. Dosahují vyšší úspěšnosti detekce onsetů v hudebním signálu a jsou vhodnější pro další použití. Nakonec byl systém implementován v prostředí Python kvůli možnosti využití knihoven LibROSA.

V kap. č. 6 jsou nejprve uvedeny různé metody a systémy pro detekci tempa. Pro vytvoření detektoru globálního tempa a určení časových pozic jednotlivých dob v nahrávce byla zvolena metoda D. Ellise. Následně byl do systému ve fázi předzpracování signálu implementován Teagerův-Kaiserův energetický operátor. V rámci předloženého referenčního datasetu byl testován systém s použitím operátoru (TKEO) i bez něj (ODF) a výsledky byly následně porovnány. TKEO systém dosahoval v průměru lepších výsledků při detekci GT i časových pozic dob. Prokázala se tedy domněnka, že tento operátor může pomoci systému s určením vhodnějších onset kandidátů pro zjištění odpovídající periodicity onset detekční funkce – a určením výsledného tempa. Systém TKEO vykazoval při detekci tónů u nahrávek sólových instrumentů obecně vyšší chybovost než ODF (nedetekoval méně výrazné změny spektra), přesto byla výsledná rytmická struktura, kterou systém TKEO určil, přesnější. ODF se tedy hodí více pro detekci pozic onsetů, TKEO pro určení tempa skladby.

Nakonec byl analyzován dataset obsahující 33 různých interpretací první věty skladby Bedřicha Smetany *Smyčcový kvartet No. 1 e moll „Z mého života“*. Každá nahrávka byla rozdělena na 6 logických úseků podle sonátové formy. Z měření vyplývá, že průměrné tempo Expozice a Provedení skladby obecně postupem času klesá a tempo ostatních částí stoupá. Přesto se délka celé skladby mírně zvětšuje a tempo klesá. Trend obecného zvyšování tempa interpretací stejných skladeb vůči provedením v minulosti tedy není povrzen, právě naopak. Výstup této práce a její výsledky budou dále sloužit muzikologům pro analýzu interpretačních výkonů, resp.pektive pro interpretaci analýzy změřených parametrů nahrávek.

Literatura

- [1] ALONSO, Miguel, Bertrand DAVID a Gaël RICHARD. *Tempo and beat estimation of musical signals* [online]. 46, rue Barrault, Paris: ENST-GET, Département TSI. Paris, 2004 [cit. 2018-12-06]. Dostupné z URL: <<http://www.cs.tut.fi/~digaudio/htyo/lahteet/p032-page-158-paper191.pdf>>.
- [2] BERTIN-MAHIEUX, Thierry, Daniel ELLIS, Brian WHITMAN a Paul LAMERE. *The million song dataset* [online]. Proceedings of the 12th International Society for Music Information Retrieval Conference, 2011, Miami, Florida. University of Miami, 2011, s. 591–596 [cit. 2018-12-06]. Dostupné z URL: <<http://ismir2011.ismir.net/papers/OS6-1.pdf>>.
- [3] BOGDANOV, Dmitry, Nicolas WACK a Emilia GÓMEZ. *Essentia: An audio analysis library for music information retrieval* [online]. ISMIR, 2013, s. 493–498 [cit. 2018-12-06]. Dostupné z URL: <<https://zenodo.org/record/1415016#.XAaeYWhKjcs>>.
- [4] BÖCK, Sebastian. *Onset, Beat, and Tempo Detection with Artificial Neural Networks* [online]. Diploma Thesis. Technische Universität München, Lehrstuhl für Mensch-Maschine-Kommunikation, Mnichov, 2010. Vedoucí práce Univ. Prof. Dr. Ing. habil. Gerhard Rigoll [cit. 2018-12-06]. Dostupné z URL: <<http://mir.minimoog.org/sb-diploma-thesis>>.
- [5] BÖCK, Sebastian, Florian KREBS a Gerhard WIDMER. *A Multi-Model Approach To Beat Tracking Considering Heterogeneous Music Styles* [online]. Department of Computational Perception Johannes Kepler University, Linz, Austria. Ismir, 2014, s. 603–608 [cit. 2019-03-04]. Dostupné z URL: <<https://pdfs.semanticscholar.org/0cc2/952bf70c84e0199fcf8e58a8680a7903521e.pdf>>.
- [6] BÖCK, Sebastian a Gerhard WIDMER. *Maximum filter vibrato suppression for onset detection* [online]. In Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13), 2013. Dostupné z URL: <http://www.cp.jku.at/research/papers/Boeck_Widmer_DAFx_2013.pdf>.
- [7] BULLOCK, Jamie a UCE Conservatoire, *Libxtract: A lightweight library for audio feature extraction* [online]. Proceedings of the International Computer Music Conference, Kodaň, 2007 [cit. 2018-12-06]. Dostupné z URL: <<https://quod.lib.umich.edu/cgi/p/pod/dod-idx/libxtract-a-lightweight-library-for-audio-feature-extraction.pdf?c=icmc;idno=bbp2372.2007.116;format=pdf>>.

- [8] DIMITRIADIS, Dimitrios, Alexandros POTAMIANOS a Petros MARAGOS. *A Comparison of the Squared Energy and Teager-Kaiser Operators for Short-Term Energy Estimation in Additive Noise* [online]. In IEEE Transactions on Signal Processing, vol. 57, no. 7, 2009, s. 2569–2581. DOI: 10.1109/TSP.2009.2019299 [cit. 2019-03-04]. Dostupné z URL: <<https://dl.acm.org/citation.cfm?id=1653359>>.
- [9] ELLIS, Daniel. *Beat Tracking by Dynamic Programming* [online]. J. New Music Research, Special Issue on Beat and Tempo Extraction, vol. 36 no. 1, March 2007, s. 51–60. DOI: 10.1080/09298210701653344 [cit. 2019-03-04]. Dostupné z URL: <<http://www.ee.columbia.edu/~dpwe/pubs/Ellis07-beattrack.pdf>>.
- [10] EYBEN, Florian, Sebastian BÖCK, Björn SCHULLER a Alex GRAVES, *Universal onset detection with bidirectional long-short term memory neural networks* [online]. In Proceedings of the International Society for Music Information Retrieval Conference, Utrecht, Netherlands, August, 2010 [cit. 2019-04-02]. Dostupné z URL: <https://www.researchgate.net/publication/224929665_Universal_Onset_Detection_with_Bidirectional_Long-Short_Term_Memory_Neural_Networks>.
- [11] GOTO, Matasaka. *An audio-based real-time beat tracking system for music with or without drum-sounds* [online]. Journal of New Music Research, vol. 30, no. 2, 2001 [cit. 2018-12-06]. Dostupné z URL: <<https://www.tandfonline.com/doi/abs/10.1076/jnmr.30.2.159.7114>>.
- [12] GROSHE, Peter a Meinard MÜLLER. *Tempogram Toolbox: MATLAB tempo and pulse analysis of music recordings* [online]. In 12th International Conference on Music Information Retrieval (ISMIR, late-breaking contribution), 2011 [cit. 2019-03-04]. Dostupné z URL: <<https://www.audiolabs-erlangen.de/resources/MIR/tempogramtoolbox#ref>>.
- [13] HAINSWORTH, Stephen a Malcolm MACLEOD. *Onset Detection in Musical Audio Signals* [online]. Cambridge University Engineering Department, Cambridge, CB2 1PZ, UK QinetiQ, Malvern, WR14 3PS. UK, 2003 [cit. 2018-12-06]. Dostupné z URL: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.7782&rep=rep1&type=pdf>>.
- [14] HESS, Adam. *Beat Detection for Automated Music Transcription: An exploration of Onset Detection Algorithms* [online]. Thomas J. Watson School of Engineering and Applied Science State University of New York at

- Binghamton. B.S. Electrical Engineering, B.A. Music, 2011 [cit.2018-12-06]. Dostupné z URL: <http://bingweb.binghamton.edu/~ahess2/Onset_Detection_Nov302011.pdf>.
- [15] HOLZAPFEL, André, Matthew E.P.DAVIES, José R.ZAPATA, João Lobato OLIVEIRA a Fabien GOUYON. *Selective Sampling for Beat Tracking Evaluation* [online]. IEEE Transactions on Audio, Speech and Language Processing, vol. 20, no. 9, 2012. s. 2539–2548, DOI: 10.1109/TASL.2012.2205244 [cit. 2018-04-04]. Dostupné z URL: <<https://www.diva-portal.org/smash/get/diva2:1040421/FULLTEXT01.pdf>>.
- [16] CHLEBOUN, Jan. *Matematika 4: Příručka pro přežití* [online]. České vysoké učení technické v Praze. Fakulta stavební, katedra matematiky, Praha, 2018 [cit. 2018-12-06]. Dostupné z URL: <https://mat.fsv.cvut.cz/chleboun/JCh_vyuka/101MA4/MA4_prirucka.pdf>.
- [17] IŠTVÁNEK, Matěj. *Výzkum tempových parametrů porovnávající zvukové nahrávky* [online]. Bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2017. s. 20 [cit. 2018-12-06]. Vedoucí práce byl Ing. Tomáš Kiska. Dostupné z URL: <<https://www.vutbr.cz/studenti/zav-prace/detail/101925>>.
- [18] KLAPURI, Anssi. *Sound onset detection by applying psychoacoustic knowledge* [online]. Proc. IEEE International Conf. Acoustics, Speech, and Signal Processing (ICASSP-99), 1999. s. 3089–3092 [cit. 2018-12-06]. Dostupné z URL: <<https://dl.acm.org/citation.cfm?id=1258114>>.
- [19] KUNA, Milan a Miloš BLÁHA. *Čas a hudba: K dramaturgii časových prostředků v hudebně interpretačním výkonu*. Praha: nakladatelství Československé akademie věd, 1982. ISSN 0018-7003.
- [20] LAROCHE, Jean. *Efficient Tempo and Beat Tracking in Audio Recordings* [online]. Creative Advanced Technology Center, Scotts Valley, CA 95066, USA. Vol. 51, No. 4, 2003 [cit. 2018-12-06]. Dostupné z URL: <<https://pdfs.semanticscholar.org/e2fd/f3918394adbd2eb9fb8e523895e703da2986.pdf>>.
- [21] LARTILLOT, Olivier a Petri TOIVIAINEN. *A Matlab Toolbox for Musical Feature Extraction From Audio* [online]. Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, 2007 [cit. 2018-12-06]. Dostupné z URL: <<http://dafx.labri.fr/main/papers/p237.pdf>>.

- [22] MATHIEU, Benoit, Slim ESSID, Thomas FILLON, Jacques PRADO a Gaël RICHARD. *YAAFE, an easy to use and efficient audio feature extraction software* [online]. Proc. of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010), Utrecht, Netherlands, 2010, s. 441–446 [cit. 2018-12-06]. Dostupné z URL: <<https://dspace.library.uu.nl/handle/1874/294715>>.
- [23] McENNIS, Daniel, Cory McKAY, Ichiro FUJINAGA a Philippe DEPALLE. *jAudio: A feature extraction library* [online]. Proceedings of the International Conference on Music Information Retrieval, 2005 [cit. 2018-12-06]. Dostupné z URL: <http://jmir.sourceforge.net/publications/ISMIR_2005_jAudio.pdf>.
- [24] MOFFAT, David, David RONAN a Joshua REISS. *An evaluation of audio feature extraction toolboxes* [online]. Proc. of the 18th Int. Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway, 2015, s. 2 [cit. 2018-12-06]. Dostupné z URL: <https://www.ntnu.edu/documents/1001201110/1266017954/DAFx-15_submission_43_v2.pdf>.
- [25] MÜLLER, Meinard. *Fundamentals of Music Processing* [online]. New York, NY: Springer Berlin Heidelberg, 2015. ISBN 9783319219455. DOI 10.1007 [cit. 2018-12-06]. Dostupné z URL: <<http://www.springer.com/us/book/9783319219448>>.
- [26] PEETERS, Geoffroy, Bruno GIORDANO, Patrick SUSINI, Nicolas MISDARIIS a Stephen McADAMS. *The timbre toolbox: Extracting audio descriptors from musical signals* [online]. The Journal of the Acoustical Society of America. Vol. 130, No. 5. 2011. s. 2902–2916 [cit. 2018-12-06]. Dostupné z URL: <https://www.mcgill.ca/mpcl/files/mpcl/peeters_2011_jasa.pdf>.
- [27] RAWLINSON, Hugh, Nevo SEGAL a Jakub FIALA. *Meyda: an audio feature extraction library for the web audio api* [online]. Web Audio Conference, 2015 [cit. 2018-12-06]. Dostupné z URL: <<https://zenodo.org/record/34093#.XAahJmhKjcs>>.
- [28] SCHLÜTTER, Jan a Sebastian BÖCK. *Improved musical onset detection with convolutional neural networks* [online]. In Proceedings of the International conference on Acoustics, speech and signal processing, Florence, Italy, May, 2014 [cit. 2019-04-02]. Dostupné z URL: <http://www.ofai.at/~jan.schlueter/pubs/2014_icassp.pdf>.
- [29] SMĚKAL, Zdeněk. *Systémy a signály: 1D a 2D diskrétní a číslicové zpracování*. Praha: Sdělovací technika, 2013. ISBN 978-808-6645-230.

- [30] SMÉKAL, Zdeněk, Tomáš KISKA a Jiří MEKYSKA. *Teagerův-Kaiserův energetický operátor*. Sdělovací technika č. 6, Praha, 2016. s. 10–13. ISSN 0036-9942.
- [31] SOLNIK, Stanislaw, Paul DEVITA, Patrick RIDER, Benjamin LONG a Tibor HORTOBÁGYI. *Teager-Kaiser Operator improves the accuracy of EMG onset detection independent of signal-to-noise ratio* [online]. Acta of bioengineering and biomechanics, vol. 10, no. 2, 65-8, 2008 [cit. 2019-04-02]. Dostupné z URL: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2596643/>>.
- [32] SRINIVASAMURTHY, Ajay. *A Data-driven Bayesian Approach to Automatic Rhythm Analysis of Indian Art Music* [online]. Universitat Pompeu Fabra, Department of Information and Communication Technologies, Barcelona, 2016 [cit. 2019-04-02]. Dostupné z URL: <<http://mtg.upf.edu/node/3593>>.
- [33] TZANETAKIS, George a Perry COOK. *Marsyas: A framework for audio analysis* [online]. Organised sound, Vol. 4, No. 03. 1999. s. 169–175 [cit. 2018-12-06]. Dostupné z URL: <<https://dl.acm.org/citation.cfm?id=972857>>.
- [34] ZHOU, Ruohua a Josh REISS. *Music Onset Detection* [online]. Center for Digital Music, Electronic Engineering Department Queen Mary University of London, UK, 2010 [cit. 2018-12-06]. Dostupné z URL: <<http://eecs.qmul.ac.uk/~josh/documents/2010/Zhou%20Reiss%20-%20Music%20Onset%20Detection%202010.pdf>>.

Seznam symbolů, veličin a zkratek

ADSR	Attack, Decay, Sustain, Release – náběh, útlum, podržení, uvolnění v rámci popisu energetické obálky tónu
BPM	Beats Per Minute – údery/doby za minutu
DAW	Digital Audio Workstation – Digitální pracovní stanice, software pro úpravu zvuku
DFT	Discrete Fourier Transform – diskrétní Fourierova transformace
DT	Default Tempo – zadané referenční tempo pro systém detekování průměrného tempa skladby
EMG	Electromyography – elektromyografie
FFT	Fast Fourier Transform – rychlá Fourierova transformace
FIR	Finite Impulse Response – filtr s konečnou impulzní odezvou
FVZ	f_{vz} – vzorkovací frekvence
GT	Global tempo – průměrné tempo celého měřeného úseku
HMM	Hidden Markov Models – skryté Markovovy modely
ISMIR	The International Society for Music Information Retrieval – specializovaná organizace, která mj. pořádá konference v oboru MIR
MIR	Music Information Retrieval – vědní obor zabývající se extrakcí informací z hudebních signálů
M.M.	Mälzelův metronom
NOT	Note Onset Time – časový okamžik vzniku hudebního tónu
ODF	Onset Detection Function – detekční funkce onsetů bez použití TKEO
ONSET	Onset – začátek tónu nebo specifické hudební události
PLP	PLP – Predominant Local Pulse
PSD	Power Spectral Density – výkonová spektrální hustota
SD	Spectral Difference – spektrální rozdíl
SF	Spectral Flux – spektrální tok
STFT	Short Time Fourier Transform – krátkodobá Fourierova transformace
TKEO	Teager-Kaiser Energy Operator – Teagerův-Kaiserův energetický operátor

Seznam příloh

A Dataset	55
B Časová délka motivů datasetu	56
C Odhadnuté tempo motivů datasetu	57
D Notový zápis	58
E Obsah přiloženého DVD	72

A Dataset

Tab. A.1: Informace o databance

Index	Názvy kvartetů	Rok	Interpret
CD01	The Czech Quartet Tradition	1928	c
CD02	Pražák Quartet	1990	c
CD03	Škampa Quartet	2003	c
CD04	Medici String Quartet	1987	z
CD05	Wihan Quartet	2005	c
CD06	Talich Quartet	1994	c
CD07	Pavel Haas Quartet	2015	c
CD08	The Martinů Quartet	1996	c
CD09	Smetana Quartet	1971	c
CD10	Panocha Quartet	1989	c
CD11	Janáček Quartet	1964	c
CD12	Kocian Quartet	2009	c
CD13	Stamitz Quartet	1991	c
CD14	Trávniček Quartet	1987	c
CD15	Wihan Quartet a Josef Suk	1994	c
CD16	Melos Quartet	1982	z
CD17	Vlach Quartet Of Prague	1991	c
CD18	Dante Quartet	2010	z
CD19	Amadeus Quartet	1977	z
CD20	Alban Berg Quartet	1989	z
CD21	Seoul String Quartet	2011	z
CD22	Pavel Haas Quartet	2015	c
CD23	Zemlinsky Quartet	2016	c
CD24	Armida Quartet	2014	z
CD25	Haba Quartet	2010	z
CD26	Guarneri Quartet	1965	z
CD27	Helikon Quartet	2018	z
CD28	Cleveland Orchestra	1956	z
CD29	The London Symphony Orchestra	1985	z
CD30	Shoreline String Quartet	2017	z
CD31	VIV-Ardo String Quartet	2009	z
CD32	Smetana Quartet	1965	c
CD33	Smetana Quartet	1989	c

Rok – datum vzniku nahrávky; Interpret – český (c) a zahraniční (z)

B Časová délka motivů datasetu

Tab. B.1: Časová délka všech nahrávek a motivů datasetu smyčcových kvartetů

Motiv	Zacatek	A	B	C	D	E	Celkem
CD01	104,20	69,19	27,90	62,33	131,65	59,60	454,87
CD02	107,97	69,53	21,75	67,45	122,98	61,30	450,98
CD03	107,79	65,59	23,08	63,38	117,39	56,10	433,32
CD04	104,38	63,93	19,60	68,68	105,87	55,30	417,76
CD05	120,80	71,88	29,24	68,74	134,07	62,40	487,12
CD06	115,49	66,54	22,70	74,18	118,11	58,30	455,32
CD07	110,61	73,05	24,41	66,46	126,12	56,80	457,45
CD08	120,54	72,27	27,85	69,02	136,41	59,10	485,19
CD09	100,68	66,23	23,92	65,95	120,85	61,10	438,73
CD10	101,30	66,50	23,59	62,33	119,40	59,90	433,02
CD11	118,45	75,60	22,15	74,95	128,95	66,20	486,30
CD12	102,00	67,69	19,85	65,90	116,15	57,10	428,69
CD13	121,27	77,78	21,20	69,80	134,00	62,60	486,65
CD14	103,35	69,50	20,72	62,56	119,60	57,40	433,13
CD15	112,60	70,35	31,80	64,85	134,85	63,50	477,95
CD16	106,25	69,70	24,25	62,86	123,10	58,10	444,26
CD17	118,30	82,22	25,70	72,80	142,50	64,80	506,32
CD18	115,40	67,05	18,80	73,47	114,10	62,70	451,52
CD19	113,30	65,60	16,92	68,60	106,05	60,50	430,97
CD20	108,60	64,45	18,55	67,10	107,40	55,00	421,10
CD21	108,87	66,93	20,10	66,92	113,05	60,80	436,67
CD22	114,50	76,30	22,46	67,70	131,25	59,30	471,51
CD23	115,00	72,85	27,60	68,57	137,30	57,40	478,72
CD24	107,00	61,43	16,45	69,16	104,48	54,95	413,47
CD25	111,10	66,00	21,80	68,40	117,50	60,30	445,10
CD26	100,30	61,60	20,75	59,05	109,95	52,40	404,05
CD27	101,70	62,80	19,30	66,30	107,65	58,20	415,95
CD28	115,20	72,95	19,80	68,47	117,60	61,90	455,92
CD29	119,57	76,08	25,35	74,50	132,30	76,40	504,20
CD30	123,05	73,45	25,15	78,85	129,99	66,20	496,69
CD31	113,85	67,55	22,25	69,25	122,75	58,76	454,41
CD32	100,50	66,60	24,00	67,20	120,90	62,50	441,70
CD33	109,20	75,90	22,83	73,97	130,10	65,00	477,00

C Odhadnuté tempo motivů datasetu

Tab. C.1: Odhadnuté tempo všech nahrávek a motivů datasetu smyčcových kvartetů

Motiv	Zacatek	A	B	C	D	E
CD01	80,61	69,37	34,41	88,56	55,60	74,50
CD02	77,80	69,03	44,14	81,84	59,52	72,43
CD03	77,93	73,19	41,60	87,09	62,36	79,14
CD04	80,48	75,08	48,98	80,37	69,14	80,29
CD05	69,54	66,78	32,83	80,31	54,60	71,15
CD06	72,74	72,14	42,29	74,41	61,98	76,16
CD07	75,94	65,71	39,33	83,06	58,04	78,17
CD08	69,69	66,42	34,47	79,98	53,66	75,13
CD09	83,43	72,48	40,13	83,70	60,57	72,67
CD10	82,92	72,18	40,70	88,56	61,31	74,12
CD11	70,92	63,49	43,34	73,65	56,77	67,07
CD12	82,35	70,91	48,36	83,76	63,02	77,76
CD13	69,27	61,71	45,28	79,08	54,63	70,93
CD14	81,28	69,06	46,33	88,24	61,20	77,35
CD15	74,60	68,23	30,19	85,12	54,28	69,92
CD16	79,06	68,87	39,59	87,81	59,46	76,42
CD17	71,01	58,38	37,35	75,82	51,37	68,52
CD18	72,79	71,59	51,06	75,13	64,15	70,81
CD19	74,14	73,17	56,74	80,47	69,02	73,39
CD20	77,35	74,48	51,75	82,27	68,16	80,73
CD21	77,16	71,72	47,76	82,49	64,75	73,03
CD22	73,36	62,91	42,74	81,54	55,77	74,87
CD23	73,04	65,89	34,78	80,50	53,31	77,35
CD24	78,50	78,14	58,36	79,81	70,06	80,80
CD25	75,61	72,73	44,04	80,70	62,30	73,63
CD26	83,75	77,92	46,27	93,48	66,58	84,73
CD27	82,60	76,43	49,74	83,26	68,00	76,29
CD28	72,92	65,80	48,48	80,62	62,24	71,73
CD29	70,25	63,09	37,87	74,09	55,33	58,12
CD30	68,26	65,35	38,17	70,01	56,31	67,07
CD31	73,78	71,06	43,15	79,71	59,63	75,56
CD32	83,58	72,07	40,00	82,14	60,55	71,04
CD33	76,92	63,24	42,05	74,62	56,26	68,31
Průměr	76,17	69,35	43,10	81,28	60,30	73,91

D Notový zápis

2

Quartett.

I.

Allegro vivo appassionato.

B. Smetana.

The musical score is arranged in four systems, each with four staves. The top staff is Violino I, the second is Violino II, the third is Viola, and the bottom is Violoncello. The key signature is one sharp (F#) and the time signature is common time (C). The first system includes dynamic markings *ff* and *pp* for the violins and *ff* for the viola and cello. The Viola part has a *f sf espress.* marking. The score features various musical notations including slurs, accents, and triplets.

Edition Peters.

9055

First system of musical notation. It consists of four staves: two treble clefs and two bass clefs. The key signature has one sharp (F#). The first staff has a melodic line with slurs. The second staff has a similar melodic line. The third staff has a more rhythmic line with slurs. The fourth staff has a bass line with slurs. Dynamics include *sf* and *pp*.

Second system of musical notation, continuing the four-staff structure. The melodic lines in the upper staves continue with slurs. The bass line in the lower staves has more rhythmic activity. Dynamics include *sf*.

Third system of musical notation. The upper staves continue with melodic lines. The lower staves feature more complex rhythmic patterns, including sixteenth notes and slurs. Dynamics include *sf* and *rfz*.

Fourth system of musical notation, starting with a section marked 'A'. It features a dense texture with many slurs and accents. Dynamics include *sf* and *rfz*. The system concludes with a trill in the upper staff and a final note in the bass staff.

First system of musical notation, featuring four staves (two treble and two bass clefs). The music includes various notes, rests, and dynamic markings such as *sf* and *ff*.

Second system of musical notation, continuing the piece with four staves and dynamic markings like *sf*.

Third system of musical notation, including a section marked 'B' and dynamic markings such as *sf* and *ff*.

Fourth system of musical notation, featuring dynamic markings such as *dim.*, *p*, and *sf*.

pp dolce pp dolce pp f sf pp dim. dim.

pp

pp

This system contains four measures of music. The first measure has a piano (*pp*) dynamic. The second measure is marked *dolce*. The third measure has a piano (*pp*) dynamic. The fourth measure is marked *dim.*. The bass line includes dynamics *pp*, *f*, *sf*, and *pp*.

cresc. f < sf f < sf sf

This system contains four measures of music. The first measure has a piano (*pp*) dynamic. The second measure is marked *cresc.*. The third measure has a piano (*pp*) dynamic. The fourth measure is marked *dim.*. The bass line includes dynamics *f*, *sf*, *f*, *sf*, and *sf*.

C
espress. p dolciss. ma espress. p espress. p

This system contains four measures of music. The first measure has a piano (*pp*) dynamic. The second measure is marked *dolciss.*. The third measure has a piano (*pp*) dynamic. The fourth measure is marked *dim.*. The bass line includes dynamics *pp*, *f*, *sf*, and *pp*.

This system contains four measures of music. The first measure has a piano (*pp*) dynamic. The second measure is marked *dolciss.*. The third measure has a piano (*pp*) dynamic. The fourth measure is marked *dim.*. The bass line includes dynamics *pp*, *f*, *sf*, and *pp*.

First system of musical notation, featuring four staves. The top two staves are in treble clef, and the bottom two are in bass clef. The music includes various note values, rests, and dynamic markings such as *p* (piano) and *f* (forte).

Second system of musical notation, featuring four staves. This system includes dynamic markings such as *cresc.* (crescendo), *pp* (pianissimo), and *f* (forte). A large letter 'D' is written above the top staff in the final measure.

Third system of musical notation, featuring four staves. The music continues with various rhythmic patterns and dynamic markings.

Fourth system of musical notation, featuring four staves. This system includes dynamic markings such as *f* (forte), *cresc.* (crescendo), and *ff* (fortissimo).

First system of musical notation, featuring a treble and bass staff with various notes and rests.

Second system of musical notation, including the tempo marking *rit. E Meno Allegro.* and dynamic markings *p* and *pp*.

Third system of musical notation, including the tempo marking *rall.* and dynamic markings *piu p*.

Fourth system of musical notation, including the tempo marking *a tempo* and dynamic markings *pp* and *sf*.

First system of musical notation, consisting of four staves (treble and bass clefs). The music features a melodic line in the upper staves and a more rhythmic accompaniment in the lower staves. A dynamic marking of *sf* is present at the end of the system.

Second system of musical notation, consisting of four staves. It includes dynamic markings such as *cresc. sf* and *sf*. The music continues with similar melodic and rhythmic patterns.

Third system of musical notation, consisting of four staves. It features a dynamic marking of *piu f < sf* and a fermata over a note in the upper staff. The music continues with similar melodic and rhythmic patterns.

Fourth system of musical notation, consisting of four staves. It features a dynamic marking of *sf* and continues with similar melodic and rhythmic patterns.

First system of musical notation, consisting of four staves (treble, two alto, and bass clefs). The music features a complex rhythmic pattern with many sixteenth and thirty-second notes, and various accidentals.

Second system of musical notation, consisting of four staves. It begins with a *ff* dynamic marking. The notation continues with intricate rhythmic figures and accidentals.

Third system of musical notation, consisting of four staves. It features a *sf* dynamic marking and a key signature change to G^b major. The music is highly rhythmic and detailed.

Fourth system of musical notation, consisting of four staves. It begins with a *sf* dynamic marking and continues with complex rhythmic patterns and accidentals.

The musical score is presented in four systems, each containing three staves. The top staff is in treble clef, the middle staff is in alto clef, and the bottom staff is in bass clef. The key signature is G major (one sharp) for the first two systems, and B-flat major (two flats) for the third and fourth systems. The time signature is 3/4. The music is characterized by a dense, polyphonic texture with many sixteenth and thirty-second notes. Dynamics include *sf* (sforzando) and *sfz* (sforzando *z*). The score includes various musical notations such as slurs, accents, and ties.

H

sf dim. p pp
sf dim. p pp
sf dim. p pp f
dim. p pp

This system contains four staves of music. The first staff has a treble clef and a key signature of one flat. It begins with a forte dynamic (sf) and a decrescendo (dim.) leading to piano (p) and then pianissimo (pp). The second staff has a treble clef and a key signature of one flat, featuring triplet markings (3) and dynamics of sf dim., p, and pp. The third staff has an alto clef and a key signature of one flat, with dynamics of sf dim., p, pp, and f. The fourth staff has a bass clef and a key signature of one flat, with dynamics of dim., p, and pp.

p dolce p dolce dim. al pp dim. al pp
pp pp

This system contains four staves of music. The first staff has a treble clef and a key signature of one flat, with dynamics of p dolce and dim. al pp. The second staff has a treble clef and a key signature of one flat, with dynamics of p dolce and dim. al pp. The third staff has an alto clef and a key signature of one flat, with dynamics of sf and pp. The fourth staff has a bass clef and a key signature of one flat, with a dynamic of pp.

espress. dim. p espress.

This system contains four staves of music. The first staff has a treble clef and a key signature of one flat, with dynamics of sf and espress. The second staff has a treble clef and a key signature of one flat, with dynamics of sf and dim. The third staff has an alto clef and a key signature of one flat, with dynamics of f, sf, and p espress. The fourth staff has a bass clef and a key signature of one flat, with dynamics of sf and p.

p

This system contains four staves of music. The first staff has a treble clef and a key signature of two sharps, with a dynamic of p. The second staff has a treble clef and a key signature of two sharps. The third staff has an alto clef and a key signature of two sharps. The fourth staff has a bass clef and a key signature of two sharps, with a dynamic of p.

First system of musical notation, featuring four staves (treble, alto, tenor, and bass clefs) in a key signature of three sharps (F#, C#, G#). The music includes various rhythmic values and dynamic markings such as *f* and *sf*.

Second system of musical notation, continuing the piece with four staves. It includes dynamic markings like *sf* and *f*.

Third system of musical notation, featuring four staves. It includes dynamic markings such as *p* and *p tranquillo*.

Fourth system of musical notation, featuring four staves. It includes dynamic markings such as *triquillo*, *f*, and *cresc.*

First system of musical notation, featuring four staves (treble and bass clefs). The music is in a key with three sharps (F#, C#, G#) and a 2/4 time signature. The first two staves are marked with a fortissimo (*ff*) dynamic. The bottom two staves also feature rhythmic patterns with *ff* markings.

Second system of musical notation, continuing the four-staff arrangement. The top two staves are marked with a crescendo (*cresc.*). The bottom two staves continue with rhythmic accompaniment.

Third system of musical notation, starting with a ritardando (*rit.*) and a tempo change to **Meno allegro.** The top two staves show a decrescendo (*dim.*) and a piano (*p*) dynamic. The bottom two staves also feature *dim.* and *p* markings. A *trm* (trill) is indicated in the second staff.

Fourth system of musical notation, beginning with a rallentando (*rall.*). The top two staves are marked with a piano (*p*) dynamic, and the bottom two staves also feature *p* markings. The music concludes with a double bar line.

a tempo

The musical score is arranged in four systems, each with four staves (two treble and two bass clefs). The first system begins with a tempo marking. Dynamics include *pp*, *p*, *f*, *sf*, and *cresc.*. The second system features *sf* and *sff*. The third system includes *p* and *espress.*. The fourth system includes *p*, *rall.*, and *a tempo*. The score concludes with a final *p* dynamic marking.

The musical score is presented in four systems, each containing three staves (treble, alto, and bass clefs). The key signature is one sharp (F#) and the time signature is 3/4. The first system begins with a forte (*f*) dynamic. The second system features a *dim.* (diminuendo) marking in the second measure and a *piu p* (pianissimo) marking in the fourth measure. The third system includes *pp* (pianissimo) markings in the first and third measures, and *ppp* (pianississimo) markings in the fourth and fifth measures. The fourth system starts with a *pizz.* (pizzicato) marking in the first measure and another *pizz.* marking in the third measure of the bass staff.

Edition Peters.

9055

Obr. D.1: Notový zápis první věty analyzované skladby

E Obsah přiloženého DVD

Přiložené DVD obsahuje několik dokumentů týkajících se práce. Nejprve jsou zde dvě složky – v adresáři *Onset_detekce* jsou vybrané výstupy testovaných nahrávek na onset detekci, ve složce *Tempo_detekce* pak několik nahrávek referenční databáze a první tři interpretace druhého datasetu analyzované oběma systémy. Hlavním dokumentem je elektronická verze práce s názvem *DiplomovaPrace_Istvanek*. Následují dva soubory *MirexDataset_tempo* a *Rozdeleni - B. Smetana* obsahující výsledky jednotlivých měření, mezivýpočty, grafické výstupy a další vypočítané parametry, které v práci nebyly použity.

```
DVD
├── Onset_detekce
│   ├── Onset_Test_clarinet_ODF.wav
│   ├── Onset_Test_clarinet_TKEO.wav
│   ├── Onset_Test_classic.wav
│   ├── Onset_Test_classic2.wav
│   ├── Onset_Test_distguit.wav
│   ├── Onset_Test_drums.wav
│   ├── Onset_Test_Dvorak.wav
│   ├── Onset_Test_guit.wav
│   ├── Onset_Test_Chip.wav
│   ├── Onset_Test_jazz.wav
│   ├── Onset_Test_piano.wav
│   ├── Onset_Test_pop.wav
│   ├── Onset_Test_rock.wav
│   ├── Onset_Test_rock2.wav
│   ├── Onset_Test techno.wav
│   ├── Onset_Test_violin_ODF.wav
│   └── Onset_Test_violin_TKEO.wav
├── Tempo_detekce
│   ├── SMC1_TKEO.wav
│   ├── SMC2_TKEO.wav
│   ├── SMC3_TKEO.wav
│   ├── SMC12_ODF.wav
│   ├── SMC12_TKEO.wav
│   ├── Tempo_CD1 Zacatek_ODF.wav
│   ├── Tempo_CD1 Zacatek_TKEO.wav
│   ├── Tempo_CD2 Zacatek_ODF.wav
│   ├── Tempo_CD2 Zacatek_TKEO.wav
│   ├── Tempo_CD3 Zacatek_ODF.wav
│   └── Tempo_CD3 Zacatek_TKEO.wav
├── DiplomovaPrace_Istvanek.pdf
├── MirexDataset_tempo.xlsx
└── Rozdeleni - B. Smetana.xlsx
```