



# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY

A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

## VÝZKUM TEMPOVÝCH PARAMETRŮ POROVNÁVAJÍCÍ ZVUKOVÉ NAHRÁVKY

RESEARCH OF TEMPO FEATURES COMPARING AUDIO RECORDS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Matěj Ištvanek

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Tomáš Kiska

BRNO 2017

# Bakalářská práce

bakalářský studijní obor **Audio inženýrství**  
Ústav telekomunikací

**Student:** Matěj Ištváněk

**ID:** 161433

**Ročník:** 3

**Akademický rok:** 2016/17

**NÁZEV TÉMATU:**

## Výzkum tempových parametrů porovnávací zvukové nahrávky

**POKYNY PRO VYPRACOVÁNÍ:**

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

**DOPORUČENÁ LITERATURA:**

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

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

**Termín zadání:** 1.2.2017

**Termín odevzdání:** 8.6.2017

**Vedoucí práce:** Ing. Tomáš Kiska

**Konzultant:**

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

**UPOZORNĚNÍ:**

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

## ABSTRAKT

Tato práce se zabývá technickými vlastnostmi audio signálů, respektive skladbami z připravené databanky, a popisuje používané parametry pro transkripci a rozbor audio signálů. Shrnuje poznatky z oblasti hudební teorie a automatické transkripce zvukových nahrávek, uvádí odborné studie zabývající se problematikou analýzy signálů a jejich výsledky. Dále uvádí atributy s největší schopností rozlišovat dané skladby podle tempa a obecně z rytmického a metrického hlediska. Práce analyzuje pomocí dvou zvolených metod jednotlivé interpretace skladby Leoše Janáčka *Smyčcový kvartet No. 1 – IV. Con moto* z definované databanky v jazyku *MATLAB* a předkládá výsledky analýzy a porovnání mezi jednotlivými metodami. Na závěr uvádí shrnutí celé práce a probírané problematiky.

## KLÍČOVÁ SLOVA

Automatická transkripce, tempové parametry, analýza signálů, elektronický notový zápis, tempo, onset.

## ABSTRACT

This thesis deals with technical properties of audio signals or more precisely of recordings from the prepared database and describes parameters which are used for music transcription and analysis of audio signals. It summarises information about music theory and automatic transcription of audio recordings, introduces specialist studies that deal with problems of signal analysis and their results. Furthermore it mentions attributes with the best ability to generally differentiate included songs from rhythmical and metrical aspects. Thesis analyses in the *MATLAB* language, from the prepared database, all interpretations of the piece *String Quartet No. 1 – IV. Con moto* from Leoš Janáček with two selected methods and shows results of the analysis and comparing of the methods. Finally the work summarizes all information and problems of the thesis.

## KEYWORDS

Automatic transcription, tempo parameters, signal analysis, electronic notation, tempo, onset.

IŠTVÁNEK, Matěj *Výzkum tempových parametrů porovnávající zvukové nahrávky*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2017. 61 s. Vedoucí práce byl Ing. Tomáš Kiska.

## PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Výzkum tempových parametrů porovnávací zvukové nahrávky“ jsem vypracoval(a) samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

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

Brno .....

.....

podpis autora(-ky)



## PODĚKOVÁNÍ

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

Brno .....

.....

podpis autora(-ky)



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

## PODĚKOVÁNÍ

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

Brno .....

.....

podpis autora(-ky)



EVROPSKÁ UNIE  
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ  
INVESTICE DO VAŠÍ BUDOUCNOSTI



# OBSAH

<b>Úvod</b>	<b>11</b>
<b>1 Úvodní teorie</b>	<b>12</b>
1.1 MIDI . . . . .	12
1.2 Automatická hudební transkripce . . . . .	12
1.3 Problém převodu audio signálu . . . . .	13
1.4 Umělá kompozice a lidský faktor . . . . .	14
<b>2 Zpracování transkripce</b>	<b>15</b>
2.1 Předzpracování signálu . . . . .	15
2.2 Rozdělení transkripce . . . . .	15
2.3 Základní parametry . . . . .	16
2.3.1 Témbr . . . . .	16
2.3.2 Dynamika . . . . .	17
2.3.3 Čas . . . . .	17
<b>3 Technické parametry pro výpočet tempa</b>	<b>21</b>
3.1 Doba počátků not . . . . .	21
3.2 Způsoby detekce doby počátků not . . . . .	22
3.2.1 Detekce na základě energetické obálky . . . . .	22
3.2.2 Detekce na základě spektrální obálky . . . . .	23
3.2.3 Další způsoby detekce . . . . .	24
3.2.4 Detekční funkce . . . . .	24
3.2.5 Přesnost detekce . . . . .	25
<b>4 Ostatní parametry</b>	<b>27</b>
4.1 Chroma vektor . . . . .	27
4.2 Jasnost zvuku . . . . .	27
4.3 Jasnost pulzů . . . . .	28
4.4 Spektrální plochost . . . . .	28
<b>5 Analýza</b>	<b>29</b>
5.1 Databanka . . . . .	29
5.2 Vyhodnocovací systém . . . . .	29
5.3 Výběr motivů a časová synchronizace . . . . .	30
5.4 Postup . . . . .	30

<b>6</b>	<b>Výsledky</b>	<b>34</b>
6.1	Výsledky metod . . . . .	36
6.2	Česká republika – Evropa – Svět . . . . .	39
6.3	Vývoj tempa během let . . . . .	40
<b>7</b>	<b>Závěr</b>	<b>42</b>
	<b>Literatura</b>	<b>43</b>
	<b>Seznam symbolů, veličin a zkratk</b>	<b>47</b>
	<b>Seznam příloh</b>	<b>48</b>
<b>A</b>	<b>Databanka</b>	<b>49</b>
<b>B</b>	<b>První motiv</b>	<b>51</b>
<b>C</b>	<b>Druhý motiv</b>	<b>53</b>
<b>D</b>	<b>Analýza hodnot segmentování</b>	<b>55</b>
<b>E</b>	<b>Notový přepis skladby</b>	<b>57</b>
<b>F</b>	<b>Obsah příloženého DVD</b>	<b>61</b>

# SEZNAM OBRÁZKŮ

2.1	Ilustrační ADSR obálka jedné noty. . . . .	20
5.1	Blokové schéma vyhodnocovacího systému. . . . .	29
5.2	Extrakce obálky pro první skladbu prvního motivu. . . . .	31
5.3	Detekce začátků not a rozdělení prvního motivu. . . . .	32
5.4	Křivka doby začátků not pomocí určení výšky tónů. . . . .	32
5.5	Výsledná křivka doby začátků not první skladby pomocí určení výšky tónů. . . . .	33
5.6	Výsledná křivka doby začátků not druhé skladby pomocí určení výšky tónů. . . . .	33
6.1	Tempový vývoj motivu č. 1 při použití spektrální dekompozice singálu. . . . .	34
6.2	Tempový vývoj motivu č. 1 při použití autokorelační funkce a spektrální dekompozice. . . . .	34
6.3	Tempový vývoj motivu č. 2 při použití spektrální dekompozice singálu. . . . .	35
6.4	Tempový vývoj motivu č. 2 při použití autokorelační funkce a spektrální dekompozice. . . . .	35
6.5	Histogram rozložení tempa všech skladeb motivu č. 1. . . . .	38
6.6	Histogram rozložení tempa všech skladeb motivu č. 2. . . . .	38
6.7	Hodnota tempa všech skladeb motivu č. 1 podle roku vzniku kvarteta. . . . .	41
6.8	Hodnota tempa všech skladeb motivu č. 2 podle roku vzniku kvarteta. . . . .	41
E.1	Notový zápis prvního motivu. . . . .	57
E.2	Notový zápis druhého motivu. . . . .	58

## SEZNAM TABULEK

6.1	Porovnání průměru tempa všech nahrávek databanky motivu č. 1. . .	37
6.2	Porovnání průměru tempa všech nahrávek databanky motivu č. 2. . .	37
6.3	Porovnání rozdílů skutečného tempa motivu č. 1 českých interpretů vůči zbytku Evropy a světa. . . . .	39
6.4	Porovnání rozdílů skutečného tempa motivu č. 2 českých interpretů vůči zbytku Evropy a světa. . . . .	40
A.1	Informace o souborech v databance. . . . .	49
B.1	Porovnání skutečného tempa prvního motivu a výpočtu tempa po- mocí dvou zvolených metod. . . . .	51
C.1	Porovnání skutečného tempa druhého motivu a výpočtu tempa po- mocí dvou zvolených metod. . . . .	53
D.1	Rozbor hodnot vektorů z dekompozice druhého motivu při použití spektrální obálky. . . . .	55

# ÚVOD

Dříve lidé zaznamenávali hudbu a zvuky jako takové pomocí specifických symbolů, které se postupně významově zlepšovaly, zpřesňovaly a unifikovaly, až dospěly do podoby, jakou známe dnes. Proces sjednocení značení trval velmi dlouho. Pokusy začaly v podobě jednoduchých značek a nákresů, poté přešly v sofistikovanější metody, až se vyvinula notová osnova, například v zápise gregoriánských chorálů. Dnes existuje a používá se více druhů notového značení, například druhy grafických zápisů, využívajících kreslené znaky a obrazce; tato práce však počítá s nejnámější, tzv. klasickou západní podobou.

Již v době vzniku prvních strojových počítačů lidé přemýšleli nad jejich využitím v oblasti hudební analýzy či dokonce interpretace, avšak tehdejší výpočetní výkon nebyl dostatečný. První zmínky o zpracování signálů můžeme najít v 17. století, kdy matematikové dali základ obecným principům, algoritmům a řešením rovnic pomocí numerické analýzy. Některé postupy, jako třeba Fourierova transformace a její obměny, jsou naprosto nezbytné nejen v programech zabývajících se zpracováním audio signálů, ale i při klasických výpočtech v oblasti hudební transkripce, a byly fundamentální složkou pro vznik vědeckého, uměleckého a hlavně technického odvětví – zpracování audio signálů (Audio Signal Processing). S nástupem dostatečně výkonných počítačů se začaly objevovat první strojové transkripce audio signálů a jak se technologie vyvíjela, přestalo být zpracování velkého množství signálů a informací v reálném čase problémem.

Cílem této práce je shrnout poznatky a studie o hudební transkripci (audio-to-score alignment). Věnuje se popisu MIDI rozhraní, jednotlivých parametrů a charakteristik hudebních signálů a jejich zpracování a využitelnosti se zaměřením na metrum a rytmus skladeb. Především rozebírá členění rytmických struktur a proces výpočtu na bázi různých dostupných parametrů tempa. Dále uvádí výsledky studií, které se soustředí na automatickou transkripci nahrávek a rozbor technických atribut signálů a jejich parametrů. Nakonec je vytvořena tempová analýza dvou charakteristických motivů skladby, respektive jejich interpretací, přiložené databanky a výsledky jsou popsány a shrnuty.

# 1 ÚVODNÍ TEORIE

## 1.1 MIDI

MIDI (Musical Instrument Digital Interface) je dnes již mezinárodně uznávaný softwarový a hardwarový standard pro komunikaci mezi rozličnými zařízeními (počítače, syntetizéry), zavedený v roce 1983 [1]. Tento elektronický protokol nahradil dřívější problémy se synchronizací syntetizérů, elektrických kláves a jiných zařízení. Základní verze poskytuje až 16 nezávislých kanálů k přenosu různých signálů. Také může emulovat vícestopé nahrávání – převážně se jedná o různé kontrolní programy, ovládání světel, změny nastavení syntetizérů pomocí počítače nebo automatizaci mixovacích procesů. MIDI rozhraní je v posílání informací mezi ovladači a komunikaci mezi přístroji hudební sféry nejrozšířenější prvek.

## 1.2 Automatická hudební transkripce

Jedná se o proces analýzy a přepisu audio signálu s informacemi o jeho parametrech – základními jsou výška a zdroj každého zvuku, délka jeho trvání, rozsah spektra a dynamika. Dnes již celosvětově rozšířený a uznávaný zápis (dle západní tradice), obsahující atributy pro následné čtení a přehrávání melodie, harmonie a not obecně, nabízí určitý druh symbolického značení, který může určovat zmíněné atributy, ale také expresivnější prvky, jako třeba legato/ligatura, staccato, portamento či decrescendo nebo slovní komentář autora či aranžéra skladby, je inspirací pro elektronický notový záznam. Vytváření takového zápisu z audio signálu poskytuje výhodu v úpravě a reprodukování původního záznamu. Je to v podstatě návod nebo druh jazyka, který muzikantovi říká, co a jak hrát na hudební nástroj. Nemusí to být jen běžně používaná úprava, ale i netradiční grafické nebo symbolické záznamy a instrukce; důležitým prvkem je, aby obsahovaly relevantní informace ohledně provedení dané skladby. Některé případy automatické hudební transkripce jsou doposud neřešitelné a vytváří tak prostor pro lepší porozumění a zlepšení použitých metod v této problematice.

Podobným zápisem, ale v elektronické formě, je zmíněný MIDI formát souborů. Často obsahuje větší množství strukturovaných dat, než jen základ pro popis hudebního signálu. Výsledky mnohých pokusů a metod od samého začátku historie automatické transkripce by se daly považovat za subjektivní a nepřesné, protože jen málo z nich staví na systematickém přístupu a dostatečném množství příkladů [2]. V dnešní době existují postupy a metody zabývající se automatickým zachycováním co největšího množství informací z nahrávky [3] – od předznamenání přes rozsah



tempa během celé skladby až po data, která vůbec nejsou součástí audio signálu, ale obsahují pouze doplňující informace (název skladby a autora, rok vydání, ad.).

Vyhodnocovací systémy se používají, kromě oblasti hudebních softwarů, studií a výzkumných programů, v internetových přehrávačích a databankách. Díky obsahové audio analýze lze podle předchozích poslechů či zadaných parametrů, jako žánr hudby, použité nástroje nebo jiné charakteristické znaky, vyhledat vhodný seznam skladeb a jednotlivá díla. Program pak doporučuje díky nashromážděným datům další skladby a může ulehčovat hledání správného materiálu pro posluchače.

### 1.3 Problém převodu audio signálu

U automatického převodu záznamu na elektronický zápis not se potýkáme s mnoha problémy – například pokud vícero muzikantů hraje rozdílné noty různé intonace a dynamiky odlišných nástrojů dohromady ve stejný čas [4], musíme určit fundamentální frekvenci každého tónu i přesto, že se překrývají s vyššími harmonickými ostatních hudebních nástrojů. Lze použít databanku předdefinovaných zvuků, které program vyhodnocuje a porovnává dle spektrální analýzy se zadanou nahrávkou. Pokud bychom věděli, že bude přehrán třeba smyčcový kvintet, tato knihovna by mohla obsahovat více podobných samplů často používaných nástrojů – několik houslí, viol a violoncell, mnohokrát zaznamenaných, aby se zvýšila pravděpodobnost teoretické shody s předloženým materiálem. Následná analýza by vyřešila identifikování nástrojů v dané skladbě – algoritmus by porovnával poměr energií a amplitud signálu v určitých frekvenčních pásmech, kde se vyskytují charakteristické vyšší harmonické složky, formanty, instrumentů (ty se nachází přibližně v násobcích základní frekvence, podle harmonické řady a podle nástrojové konstrukce ad.). Hudební nástroje bohatší na barvu, a tedy i na spektrum signálu a energii ve vyšších harmonických složkách, dosahují méně přesných výsledků při identifikování nástrojů v nahrávce – barevnější a spektrálně rozsáhlejší piano mělo u frekvenčního měření podle práce Š. Albrechta [5] úspěšnost 60-70 %, při použití pouze vyšší dynamiky (rozdíl mezi *mezzoforte* a *forte* stejného nástroje), namísto presence více vyšších harmonických složek, úspěšnost přes 90 %. Simultánní hra více instrumentů je často spektrálně nahodilá; formanty se různě sčítají a odečítají, vytváří nové rezonance a místa s větší či menší energií, a to v každé kombinaci výšky vícero tónů rozdílně.

## 1.4 Umělá kompozice a lidský faktor

Z estetické a umělecké stránky můžeme považovat převod audia na notovou kompozici za určitého umělého skladatele, dirigenta či umělce obecně. Díky nepřesnostem ve vyhodnocovacích algoritmech tak vzniká originální kompozice, která se jen těžko předpovídá, dokud neproběhne. Následné přehrání signálu z předchozích záznamů může přinést různý výsledek, protože mezi konverzí na symbolický jazyk a následným přehráním zvuků (pomocí rozličných databank s různými zvuky a nástroji) se některé parametry původní skladby změní. Doposud nejsou používané algoritmy natolik technicky propracované, aby dokázaly napodobit jakýkoliv výraz muzikanta. Výsledný zvuk a hudební vyjádření zkušeného hudebníka se dá považovat za kvalitnější a umělecky hodnotnější, než u umělé reprodukce, kde drobné nepřesnosti mizí, a tím vytváří pocit nepřirozené dokonalosti, na kterou je lidské ucho citlivé. Na druhou stranu třeba právě díky výpočetním schopnostem strojů bude možné dosáhnout dříve nepředstavitelných hudebních výkonů, které by byly umělecky všeobecně akceptovány.

Lidské sluchové ústrojí je výkonným systémem pro hudební analýzu. Proto je důležité znát jeho funkce a principy. Existují dva typy získávání znalostí o jeho stavbě a fungování – studie na subjektech s poškozenými mozky (a sluchovými ústrojími) a neurologické experimenty na zdravých lidech [6]. Provádí se různé pokusy a testy, vyhodnocují se schopnosti lidí s rozličnými vadami rozeznávat výšky a další atributy hudebních signálů a reakce jejich obou mozkových hemisfér [7, 8]. Výsledky pomáhají při pochopení lidského vnímání zvuku a tím ovlivňují vývoj a pokrok techniky obecně. Mozek jako takový je koncovým „posluchačem“ v analýze interpretovaných hudebních signálů, a je tedy vhodné si uvědomit, nakolik v této věci zasahuje lidská anatomie i individuální subjektivita.

## 2 ZPRACOVÁNÍ TRANSKRIPCE

### 2.1 Předzpracování signálu

Hudební signály se ještě před samotnou analýzou často předzpracovávají. Většina hudebních signálů totiž obsahuje informace, které při následném určování parametrů nehrají žádnou roli a jsou tudíž pouze nadbytečným objemem dat nebo naopak vnášejí do výpočtů chyby, které zvětšují výsledné nepřesnosti. Mezi procesy předzpracování hudebních signálů patří například normalizace, vzorkování signálu, kvantizace nebo odstranění stejnosměrné složky signálu (Direct Current offset) a šumu. U analýzy se pochopitelně požaduje co nejkvalitnější záznam a komprese na jiný zvukový formát by často mohla nepříznivě ovlivnit výsledek. Komprimace signálu, využívající poznatky z vědního oboru *psychoakustiky* – maskovací jevy, tzv. místa ticha aj., však s sebou nese řadu výhod – především menší objem dat, které zabírají méně fyzického místa na disku. Tento proces se začal hojně využívat s nástupem internetu, nicméně pro zachování původních informací se používají především nekomprimované soubory nebo soubory s bezztrátovou kompresí.

Podle studie G. Peeterse [9] předzpracování signálu zahrnuje

- odhad energetické obálky signálu (výpočet globálních dočasných parametrů, popisujících vlastnosti signálu)
- krátkodobou Fourierovu transformaci (Short Time Fourier Transform – STFT)
- sinusové harmonické modelování signálu (výpočet vrcholů STFT v každém použitém časovém rámci a výběr těch, které jsou přibližně v násobcích fundamentálních frekvencí)
- sled procesů, snažící se napodobit lidské vnímání zvuku (vytvoření percepčního modelu)

Některé metody synchronizace polyfonních notových zápisů s hudebními záznamy používají tzv. off-line mód [10] (zpracování neprobíhá v reálném čase zároveň s přehráváním hudební nahrávky) – všechny data jsou již k dispozici. Ostatní algoritmy pracují v reálném čase, a tudíž v daný čas dokážeme určit jen momentální a minulé vzorky a jejich vlastnosti [3]. Práce se zabývá výhradně off-line zpracováním.

### 2.2 Rozdělení transkripce

Rozdělení a určení vlastností zvukového signálu je v transkripčních metodách velmi důležité. Podle A. Klapuriho [6] se základní algoritmy v automatické transkripci hudebního signálu rozdělují na výpočet více fundamentálních frekvencí zároveň (Mul-

tiple Fundamental Frequency Estimation) a na výpočet hudebního tempa a rytmu (Musical Meter Estimation). Existují samozřejmě další postupy, které mohou například zjišťovat a porovnávat barevnost signálů, energetickou rozmanitost spektra, dynamiku nahrávky ad. Určování parametrů a vlastností nahrávek je tedy klíčovým faktorem pro volbu algoritmu a snahu o dosažení požadovaného výsledku.

Pro větší přesnost a úspěšnost metod i zjednodušení jejich postupů se transkripční postupy v hudební oblasti často dále dělí na ty, které počítají jen s harmonickými zvuky, jen s bicími nástroji a na ty, které analyzují zpěv [11]. Využívají psychoakustiky, zaměřují se na rytmus, melodické a basové linky i harmonii. V práci M. Rynänenena a A. Klapuriho [11] je vytvořen algoritmus pro převedení nahrávky zpěváka na symbolický záznam. Jedna část procesu je výpočet relativní výšky záznamu během celé doby jeho trvání a druhá rozčlenění jednotlivých tónů a jejich následné kvantování pro určení intonace.

## 2.3 Základní parametry

### 2.3.1 Témbr

Každý nástroj má svoji jedinečnou barvu (témbr) [12], která se odvíjí od jeho parametrů, proporcí a charakteristik – způsob tvorby tónů (proud vzduchu přes dřevěný plátek/plátka, kmitající struny ad.), použitý materiál (dřevo, plech, plast, kov ad.) a samozřejmě originální lidská interpretace a jedinečnost techniky, zkušenost nebo styl dýchání a motoriky při hře každého jedince. Stejně tak záleží na snímací technice, použitých mikrofonech a na prostředí při nahrávání. Každá místnost či objekt má jiné akustické vlastnosti a existuje mnoho parametrů, jež mohou měnit barvu zvuku, jako třeba rozměry stěn, stropu a podlahy – módy a rezonance, pohltivost použitých materiálů a osob v prostoru, teplota nebo vlhkost vzduchu, která při vysokém zastoupení utlumuje vyšší frekvence a tím nepříznivě ovlivňuje věrnost snímání. Navíc každý nástroj je zaznamenán s jinou intenzitou a fluktuace výšky tónu (vibrato), pravidelná rychlá změna dynamiky (tremolo) a další výrazové prvky ovlivňují přesnost určení aspektů hudební nahrávky. Dále musíme počítat s případným dozvukem, ať už přirozeným či případně umělým, který vyhodnocování ještě ztíží a rozmaže frekvenční spektrum. Postprodukční úpravy a použití různých efektů, jako například kompresor a jeho variace, jsou pro následné zpracování nevýhodou, jelikož také zkreslují a mění původní signál. Dosavadní metody převodu audia do notového záznamu s těmito úpravami většinou nepočítají, protože velké množství proměnlivých parametrů a jejich kombinací by nebylo možné ve výpočtech jednoduše obsáhnout a navíc nebývají natolik výrazné, aby radikálně měnily výsledek, takže se v algoritmech nezahrnují.

### 2.3.2 Dynamika

Dynamika signálu je velmi často používaný parametr při popisu hudebních signálů a jejich transkripcích. Důkazem toho je grafické prostředí softwaru pro editaci hudebních nahrávek, kde je již v náhledu k dispozici dynamická obálka. V práci A. Lercha [3] se v úvodu objevuje definice termínu hlasitost (Loudness) a intenzita (Intensity). Intenzita je fyzikálně měřitelná veličina, kdežto hlasitost je subjektivní lidské posouzení zvuku. Z důvodu individuálního uzpůsobení sluchového ústrojí, vnímají lidé hlasitost přicházejících zvuků při lineárním zvětšování energie logaritmicky. Proto pro výpočet hladiny akustického tlaku používáme logaritmickou rovnici, jejíž výsledkem je logaritmický poměr dvou veličin (v tomto případě výkonů) s jednotkou decibel (dB) [13].

$$L_p = 20 \cdot \log \frac{P}{P_0}. \quad (2.1)$$

Kvůli praktickým účelům se používají i odvozené jednotky od decibelu – dBm, dBV, dBFS ad. Další často používaný parametr je kvadratický průměr (Root Mean Square – RMS) [14] a z audio vzorků se spočítá následovně:

$$v_{RMS}(n) = \sqrt{\frac{1}{K} \sum_{i=i_s(n)}^{i_e(n)} x(i)^2}. \quad (2.2)$$

Z uměleckého hlediska se reprezentace dynamiky signálu v klasickém notovém zápisu značí většinou zkratkami italského názvosloví – *p* (piano – potichu), *mp* (mezzopiano – středně potichu), *f* (forte – nahlas) a další. Vývoj, zvětšování či zmenšování intenzity zvuku, se také označuje slovy – *decrescendo/diminuendo* (postupně zeslabovat) nebo *crescendo* (postupně zesilovat) či jiné. V notovém zápise se ale značí převážně znaky – pro zesilování otevřená šipka, zvětšující se s nárůstem intenzity zvukového signálu, tvořená dvěma mimoběžnými úsečkami, a pro zeslabování stejná šipka, ale s opačným průběhem. Dalším dynamickým značením může být například *tremolo* (modulace hlasitosti v čase), které se v zápise značí stejně jako opakování stejného tónu, přeškrtnutou nožičkou, ale několikrát. K notě bývá někdy připsána slovní zkratka *trem.*

### 2.3.3 Čas

Čas, jakožto muzikální aspekt, je velmi důležitý parametr hudby jako takové. I pokud se jedná o tradiční kmenovou expresi z Afriky, americký jazz či italskou operu, vždy je čas, a tudíž i rytmus a metrum, fundamentálním rysem při poslechu, tvorbě

a provedení jakékoliv skladby. Pro popis časových a tempových vlastností nahrávek v transkripčních programech se používají různé parametry, mající svůj základ většinou v matematických postupech a také v algoritmech založených na autokorelačních funkcích a výpočtech pravděpodobností. Jednou ze základních složek času v hudbě je doba začátku zvuků – může se jednat o začátek hudebního tónu v díle, o úder perkusního nástroje nebo výjev zpěváka či umělce, zasazený do vhodného místa v daném provedení. Právě doba počátku hudebního jevu se obecně považuje za důležitější, než moment jeho konce, protože posluchač vnímá tuto událost výrazněji [15].

Dalším velmi významným parametrem je tempo. Jedná se o závislost posloupnosti hudebních not na čase, tudíž rychlost hry. Většinou se označuje slovem, slovy, číslem nebo jejich kombinací v notovém zápise. Slovní definice má základ v italském názvosloví a je mezinárodně uznávaná – od *Lentissima* (co nejpomaleji) přes *Con moto* (hybně, s pohybem) až po *Prestissimo* (co nejrychleji); někdy se zadává rozmezím (dvěma hodnotami) nebo pouze jedním číselným údajem. Je také nutno podotknout, že rozmezí hodnot různých označení se v průběhu historie měnilo.

Za číselnou hodnotou se většinou uvádí jednotka. Standardně se u nás používá jednotka M.M. (Mälzelův metronom), podle vynálezce metronomu. V anglické terminologii se však ustálilo označení BPM (z anglického Beats Per Minute – úderů za minutu) a značně převažuje, co se týče frekvence užívání ve světové literatuře, nad starším metronomickým značením. Ve skladbách se také obvykle značí, jaké notě odpovídá předložené tempo – zda půlové, čtvrtové s tečkou, čtvrtové apod., změna tempa v určité části díla nebo následný návrat na tempo původní – vrácení se na původní tempo se v zápise označuje jako *Tempo primo* nebo *Tempo I* a pokud došlo ke zrychlení či zpomalení, poté výrazem *a tempo*. Běžně se tempo  $\mathfrak{S}$  vypočítá, pokud je neměnné, jako počet úderů za minutu vydělené délkou daného časového úseku:

$$\mathfrak{S} = \frac{B \cdot 60}{\Delta t_s}, \quad (2.3)$$

kde  $B$  je počet úderů a  $\Delta t_s$  je zvolený časový úsek měření. Jednotkou je BPM. Pokud se jedná o dynamické, proměnné tempo, lze ho spočítat pomocí zjištění doby času každého úderu a výpočtu doby mezi dvěma sousedními údery [3]:

$$\mathfrak{S}_{local}(j) = \frac{60}{t_b(j+1) - t_b(j)}, \quad (2.4)$$

kde  $t_b$  je doba každého úderu a jednotkou je BPM.

Dle studie vypracované M. Kunou a M. Bláhou [16] 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

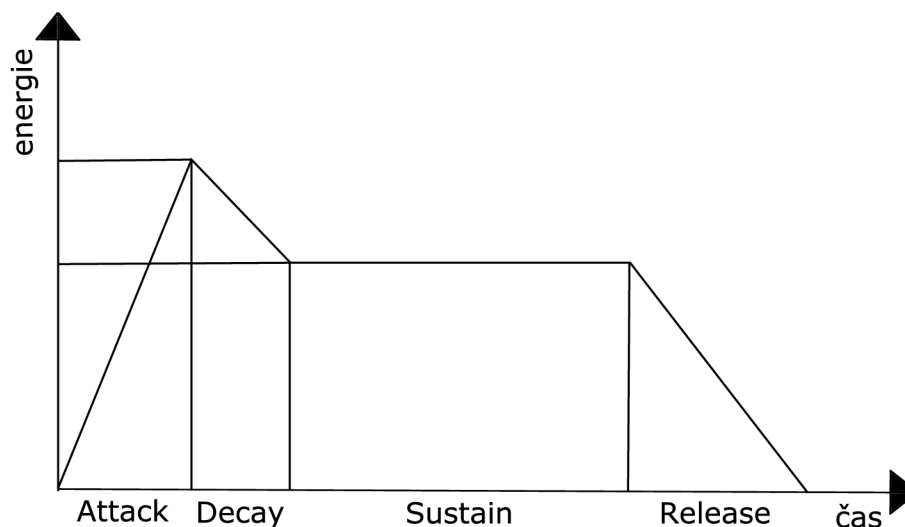
Nejčastěji používané metrizedované tempo se dále dělí na:

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)

V práci A. Gabrielsona [17] bylo rozlišeno ještě *střední tempo* (mean tempo), které neuvažuje různé změny tempa na začátku a na konci skladby (například ritardando na konci s korunou), a tím zpřesňuje přehled rychlosti skladby při počítání pouze jediné hodnoty na celou skladbu. V další práci zabývající se výpočtem tempa [18] je uvedeno *mode tempo*, vypočítané na základě mezi-úderových intervalů.

Další charakteristikou hudebního signálu může být tzv. ADSR obálka. Používá

se například při generování vlastností zvuků syntetizérů. Jedná se o obálku energetického vývoje, popsanou několika parametry. U zmíněného syntetizéru je to – náběh (Attack), který určuje, za jakou dobu od zprávy o zahájení tónu (stisk klávesy) se signál dostane na maximální úroveň; útlum (Decay) udává dobu, za jakou se signál z maxima dostane na hodnotu „sustain“; podržení (Sustain) symbolizuje neměnnou velikost signálu při stisknuté klávese; uvolnění (Release) značí dobu poklesu energie signálu po uvolnění klávesy.



Obr. 2.1: Ilustrační ADSR obálka jedné noty.

Algoritmy pro výpočet tempa jsou podle práce M. Alonsa et al. [19] rozřazeny do tří kategorií:

- detekce začátku not – měřené pomocí spektrálního toku energie
- výpočet periodicity – zde jsou využívány techniky pro určení výšky tónů (opakovatelnost úderů a důrazů)
- výpočet lokalizace úderů – pomocí křížové korelační funkce mezi nahrávkou a umělými pulzy



## 3 TECHNICKÉ PARAMETRY PRO VÝPOČET TEMPA

Uvedené parametry jsou popsány v pracích A. Lercha [3], G. Peeterse [9], O. Lartillota a P. Toiviainen [14], O. Lartillota et al. [21] a M. Müllera [23].

Výpočet tempa z hudebního signálu není jednoduchá záležitost. Hudebních žánrů je totiž mnoho a každý obsahuje specifikace, které ztěžují nebo naopak zjednodušují výpočet. U populární hudby nebo rocku lze často snadno detekovat silné krátkodobé přechodové jevy (Transient) nebo znatelné, opakující se impulzy, které vytváří bicí nástroje – především basový buben, činely ad. Právě díky opakovatelnosti a relativní neměnnosti tempa je snadnější určit rytmické a metrické struktury. V některých dílech klasické hudby (či vážné hudby) je naopak velmi obtížné zjistit tempo, a to z mnoha důvodů. Klasická hudba je většinou interpretována větším počtem hráčů, a pokud se jedná o smyčcový kvartet, jsou zde 4 zdroje zvuků dohromady. V uskupení není žádný bicí nástroj a na rozdíl od perkusí nebo třeba klavíru nedisponují housle, violy a violoncello silným přechodovým dějem, který umožňuje detekci typické rytmické konstrukce. Navíc skladba v přiložené databance může být tempově velice nečitelná (z poslechu) i pro zkušeného hudebníka, jelikož obsahuje proměnlivé důrazy a neustále se měnící tempa s různými vývoji. Různé hudební nástroje mívají odlišné průběhy vzniků not. Perkuse nebo klavír mají na začátku noty výrazný transient, housle nebo viola však nikoliv – zde je energetický vývoj většinou pozvolnější, méně průrazný – i proto je složitější určit u smyčcových nástrojů dobu začátku jednotlivých not.

### 3.1 Doba počátků not

Existuje více možností a metod, jak určit tempo nahrávky. Naprostá většina používá určení doby začátků not (Onset). Jedná se o místo na časové ose, kde s určitou pravděpodobností vznikla nová nota. Každá taková nota má svoji dobu náběhu (Attack Time) a dobu uvolnění (Release Time), viz obr. č. 2.1. Vrcholy křivky doby počátku zvukové události (Onset Time Curve) ukazují místa největší energie – předpokládané časové místo noty [14]. Jako doba náběhu se považuje čas, ve kterém se zvýší energie instrumentu až do okamžiku, kdy dojde k vytvoření noty. Náběh může být prudký (Steep Attack), podle rychlé změny energie, ale krátké doby trvání, nebo pozvolný (Gradual Attack). Každý nástroj se liší rychlostí vzniku noty – nejmenší hodnotu dosahují za určitých podmínek zpravidla bicí nástroje (od 5 ms) a největší dřevěné nástroje (až 200 ms). Doba uvolnění je pak pokles energie po existenci noty. Transient, jak bylo zmíněno výše, je přechodový jev. Může být popsán jako časově

krátký ruchový prvek s velkým spektrálním zastoupením, který se objevuje především na začátcích zvukových jevů – drnknutí do struny kytary, klavíru, vyslovení neznělé souhlásky při zpěvu. Objevuje se ale po dobu určitého času a vyvíjí se nepředvídatelně, podobá se spíše šumu (například u klavíru je transient vytvořen nejen úderem kladívka na struny, ale i zmáčknutím klávesy, zvednutím kladívka, prvotním rozvibrováním a rezonancí strun a ozvučné desky apod.) Definice termínu „doba počátku not“<sup>1</sup> se v mnoha pracích a studiích liší, a proto dochází k nepřesnostem při výkladu závěrů. Přechodový jev a detekci přechodů v audio signálu, což je další metoda pro určení doby začátku not, zmiňuje studie A. Røbela [24], kde je uveden matematický postup výpočtu:

$$t_{cg}(t_m) = \frac{\int_{\omega_l}^{\omega_h} -\frac{\partial \Phi(\omega, t_m)}{\partial \omega} |X(\omega, t_m)|^2 d\omega}{\int_{\omega_l}^{\omega_h} |X(\omega, t_m)|^2 d\omega}. \quad (3.1)$$

V práci J. W. Gordona [21] se doba začátku zvukové události dělí na několik typů:

- doba počátku noty (Note Onset Time – NOT) je definována jako čas, ve kterém byl nástroj excitován, aby zahrál zvuk
- akustická doba počátku (Acoustic Onset Time – AOT/Physical Onset Time – POT) je čas, ve kterém je děj nebo signál poprvé teoreticky měřitelný
- percepční doba počátku (Perceptual Onset Time – POT) – doba, ve které může poprvé posluchač vnímat daný zvukový jev nebo událost)

## 3.2 Způsoby detekce doby počátků not

Dobu začátků not lze vypočítat na základě energetické obálky, spektrální obálky, rozdílů fáze a na základě komplexní oblasti [23, 3], nicméně Toolbox, ze kterého tato práce při výpočtech vychází, umožňuje výpočet pouze spektrální a energetické obálky a jejich kombinací.

### 3.2.1 Detekce na základě energetické obálky

Účelem této metody je změřit lokální energetické výkyvy. Například při rozvibrování struny (strunné nástroje) nebo prvotního nárazu vzduchu o hranu (dechové nástroje) vznikne energetické maximum (dle ADSR obálky), které značí začátek

---

<sup>1</sup>V anglicky psané literatuře se používá parametr *Onset*, který vyjadřuje začátek zvukové události a často se zaměňuje právě za *Onset Time*, což by měla být pouze jedna z vlastností parametru *Onset* [3].

nové noty. Tyto nárůsty energie v signálu lze zachytit a následně pomocí detekční funkce spočítat.

V práci M. Müllera [23] je uveden následující postup výpočtu. Lokální energie diskrétního signálu  $x$  je popsána jako:

$$E_w^x(n) = \sum_{m=-M}^M |x(n+m)w(m)|^2 = \sum_{m \in \mathbb{Z}} |x(m)w(m-n)|^2, \quad (3.2)$$

pro  $n \in \mathbb{Z}$  a kde  $w$  je diskrétní funkce zvonovitého tvaru (křivky)  $w : \mathbb{Z} \rightarrow \mathbb{R}$ , posunuta vůči signálu  $x$ . Její střed se nachází v nule. Protože chceme změřit pouze nárůsty energie, nikoliv poklesy, používá se pro odfiltrování záporných hodnot jednocestný (půlvlnný) usměrňovač (jako u spektrální obálky), který se matematickou funkcí simuluje. Nakonec dostaneme detekční funkci  $\Delta_{\text{Energy}} : \mathbb{Z} \rightarrow \mathbb{R}$ :

$$\Delta_{\text{Energy}}(n) := |E_w^x(n+1) - E_w^x(n)|_{\geq 0} \quad (3.3)$$

pro  $n \in \mathbb{Z}$ .

Výhoda metody spočívá v poměrně přesném určení energeticky výrazných začátků not. Každá nota však nezačíná impulzivně a pro méně výrazné nástroje s méně transientním zvukem se tato metoda nehodí. V hudebním signálu i malé hodnoty energie, které by neprošly nastaveným energetickým prahem, mohou být důležité. Proto se používá převedení měřítka hodnot energie na logaritmické (nebo se zavádí logaritmická komprese) [23]:

$$\Delta_{\text{Energy}}^{\text{Log}}(n) := |\log(E_w^x(n+1)) - \log(E_w^x(n))|_{\geq 0} = \left| \log\left(\frac{E_w^x(n+1)}{E_w^x(n)}\right) \right|_{\geq 0} \quad (3.4)$$

Nevýhoda však spočívá v zesílení šumu a ruchových složek, které mohou vnášet do výsledků chyby. Na druhou stranu lze docílit chtěného zvýraznění některých začátků not, a tím je výrazněji oddělit od ostatních složek signálu. Většinou se používá pro dosažení přesnějších výsledků kombinace vícera metod detekování začátků not zároveň.

### 3.2.2 Detekce na základě spektrální obálky

Pro určování začátku not v hudebním signálu se jako jedna z metod používá kalkulace spektrálního toku (Spectral Flux) [10]. Jedná se o určení pravděpodobnosti začátku hrané noty pomocí spektrálních změn ve frekvenčních pásmech signálu (dle různého rozdělení). Hodnota změny spektrálního toku v signálu se nejprve znormalizuje, spočítá se prahová hodnota pomocí speciálního filtru na výstupu, který se dále odečte od zmíněné normalizované hodnoty a nakonec je použita logistická funkce. V podstatě se jedná o matematický proces, který v konečném důsledku určí,

zda s určitou pravděpodobností došlo k takovým frekvenčním změnám, aby vznikla nová nota. Spektrální tok je popsán vzorcem:

$$SF(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n, k)| - |X(n-1, k)|), \quad (3.5)$$

kde  $H(x) = \frac{x+|x|}{2}$  je „půlvlnný“ usměrňovač. Existuje více možností výpočtu detekční funkce na základě spektrální analýzy a následném vyhodnocení míst s největší hodnotou energie.

V práci M. Müllera [23] se nejprve pro získání zkomprimovaného spektra použije funkce  $\Gamma_\gamma$  na amplitudový spektrogram  $|\chi|$ :

$$y = \Gamma_\gamma(|\chi|) = \log(1 + \gamma \cdot |\chi|), \quad (3.6)$$

kde se dosadí vhodná konstanta  $\gamma \geq 1$ . Pokud by nedošlo ke zkomprimování původního spektrogramu, nebyly by pozice úderů a začátků not výrazné. Při  $\gamma = 1$  jsou maxima již dobře znatelná; se zvyšováním hodnoty  $\gamma$  se i slabé vrcholy ve spektru projeví, ale zesilují se i šumové složky záznamu. Poté se vypočítá diskrétní časová derivace zkomprimovaného signálu  $y$ . Ve výpočtu se zahrnují pouze kladné rozdíly (záporné se vyloučí), protože je použit usměrňovač. Tato detekční funkce  $\Delta_{\text{Spectral}} : \mathbb{Z} \rightarrow \mathbb{R}$  je poté vypočítána jako:

$$\Delta_{\text{Spectral}}(n) = \sum_{k=0}^K |y(n+1, k) - y(n, k)|_{\geq 0}, \quad (3.7)$$

pro  $n \in \mathbb{Z}$ .

### 3.2.3 Další způsoby detekce

Pro určení doby začátků not a pro tempovou analýzu obecně se používají i další metody a parametry, jako třeba PLP (Predominant Local Pulse) funkce, sledování úderů (Beat Tracking), dynamické programování (s využitím dynamického borcení časové osy – DTW (Dynamic Time Warping), adaptivní funkce či různé druhy tempogramů (Fourierův, cyklický, autokorelační) [3, 23]. V práci však tyto metody z různých důvodů nejsou aplikovány – algoritmy a parametry nemusí být vhodné pro analýzu nahrávky ze zadané databanky nebo přesahují rozsah této práce.

### 3.2.4 Detekční funkce

Výpočet parametru *Onset* využívá detekční funkce (Detection Function/Novelty Function). Jedná se o funkci, která zaznamenává míru změn hudebního signálu [22]. Nejprve proběhne segmentace signálu na bloky, pak se spočítá množství nových

informací v jednotlivých sekcích, a poté se vyhodnotí pozice maxim, značící dobu náběhu noty.

Pokud by funkce měla úspěšnost 100 %, zachycovala by každý kladný lokální extrém. V praxi však tato metoda vytváří mnoho chyb – chybných pozitiv (False Positives), což jsou špatně detekované doby not, které mají za následek zvětšenou nepřesnost výsledků [2]. Proto se používá speciální filtr s nastavitelným prahem  $G$  (Threshold) podobný šumové bráně (Noise Gate) v audio technice, kde všechna maxima nad nastavenou hodnotou jsou detekována jako začátky not. Pokud by prahová hodnota byla konstantní, pak:

$$G_{d,c} = \lambda_1. \quad (3.8)$$

Pokud není konstantní, nazývá se přizpůsobivý nebo také MA (Moving Average) filtr, jehož práh se vypočítá jako:

$$G_{d,ma} = \lambda_2 + \sum_{j=0}^{O-1} b(j) \cdot d(i-j), \quad (3.9)$$

kde  $b(j)$  je nastavitelná funkce okna (Window Function) a  $\lambda_2$  značí váhu posouvání hodnoty prahu, respektive citlivosti filtru.

### 3.2.5 Přesnost detekce

Rozlišují se dvě základní detekce – přesnost detekce (Detection Accuracy) a detekční výkonnost (Detection Performance) [3]. První systém hodnotí časovou přesnost algoritmu a druhý vyhodnocuje počet správných a chybných detekcí s poměrně malou tolerancí. Určené časové odchylky slouží jako vstupní data pro vyhodnocovací algoritmus. Aritmetický průměr těchto hodnot ukazuje, zda vzorky vzniklých not zpožďují nebo naopak předbíhají požadovaný signál:

$$d_{\text{mean}} = \sum_{\forall j} \Delta t_{R,D}(j), \quad (3.10)$$

kde  $\Delta t_{R,D}$  značí časový rozdíl mezi detekovanými časy začátků not a referenčními hodnotami. Absolutní průměr značí průměrnou hodnotu časové vzdálenosti mezi detekovanou dobou začátku not a referenční hodnotou:

$$d_{\text{abs}} = \sum_{\forall j} |\Delta t_{R,D}(j)|. \quad (3.11)$$

Dalším používaným parametrem je standardní odchylka:

$$\sigma_d = \sqrt{\frac{1}{O_{\text{DP}}} \sum_{\forall j} (\Delta t_{R,D}(j) - d_{\text{mean}})^2} \quad (3.12)$$

a absolutní maximální hodnota odchylky:

$$d_{\max} = \max|\Delta t_{R,D}(j)|. \quad (3.13)$$

Nastavování prahové hodnoty filtru pro odstranění nepřesností při detekci může být kontraproduktivní. V energetické obálce spektra například nahrávky smyčcových nástrojů může často dojít k situaci, že vibrato silně zahrané noty energeticky přesahuje hodnotu slaběji zahrané noty, a tudíž odstraní i ty detekce, které ve skutečnosti značí reálný začátek notové události.

## 4 OSTATNÍ PARAMETRY

Kromě tempových vlastností a struktur porovnávají různé algoritmy další aspekty zvukových nahrávek, a to především dynamiku a barvu (spektrální rozmanitost) audio signálů. Uvedeno je pouze několik ukázek z mnoha rozličných parametrů.

### 4.1 Chroma vektor

Chroma vektory poskytují reprezentaci výšek tónů hudebního signálu pro převod zvuku do symbolického zápisu. Je to vektor o dvanácti rozměrech (dvanáct půltónů v rovnoměrném temperovaném ladění), kde každá z těchto dimenzí vyjadřuje rozdělení energie ve frekvenčních pásmech dvanácti půltónů za sebou, podle chromatické stupnice [10]. Následná extrakce probíhá většinou na základě součtu amplitud každé části STFT. Některé studie používají chroma vektory k automatickému rozpoznávání akordů ve skladbách. Ch. Harte a M. Sandler ve své práci [25] použili k identifikování akordů kvantované chroma vektory a rozčlenění harmonických změn a dosáhli až 62,5% úspěšnost v případě, kde byly zahrnuty durové a mollové varianty. Vektor lze počítat jako:

$$C_f(b) = \sum_{z=0}^{Z-1} |X_{if}(b + z\beta)|, \quad (4.1)$$

kde  $X_{if}$  je logaritmické spektrum,  $Z$  – počet oktáv;  $z$  – celočíselný oktávový index  $\in [0, Z - 1]$ ;  $b$  – celočíselný výškový index  $\in [0, \beta - 1]$ ;  $\beta$  – počet „zásobníků“ na oktávu [26].

### 4.2 Jasnost zvuku

Měření tzv. jasnosti zvuku (Brightness, Clarity) se používá především pro charakterizování odlišností barev hudebních nástrojů. Její křivka znázorňuje vývoj jasnosti během časového intervalu. Vyšší hodnoty znamenají, že většina energie se nachází ve vyšších frekvencích a menší hodnoty indikují většinové zastoupení zvukové energie v basech až středních basech (nižší rejstříky). Jejím opakem může být temnost či tmavost – i přesto, že je tato veličina fyzikálně měřitelná, její vnímání a popis závisí na lidské subjektivitě. Podobnou veličinou je **spektrální těžiště (Spectral Centroid)**, které určuje, kde (na jakých frekvencích) se nachází největší množství energie celého spektra. Jasnost zvuku a spektrální těžiště se často používají ve vyhodnocovacích algoritmech pro automatické určování hudebních žánrů, stylů a barev.

Vypočítá se jako:

$$SC = \frac{\sum_{k=1}^N kF[k]}{\sum_{k=1}^N F[k]}, \quad (4.2)$$

kde  $F[k]$  je amplituda odpovídající  $k$ -položce v DFT spektru [9, 27].

### 4.3 Jasnost pulzů

Parametr popisující intervaly ve zvukovém signálu, které určují základní tempo a sílu rytmických úderů, se nazývá jasnost pulzů (Pulse Clarity). Výpočet probíhá podobně jako u spektrálního toku. Používá se stejná diferenciatní metoda na spektrální obálku, ale bez rozřazení frekvenčních pásem na skupiny [20]. V podstatě se jedná o obdobu detekční funkce ve frekvenčním měřítku pro zjištění času začátku not.

### 4.4 Spektrální plochost

Spektrální plochost (Spectral Flatness) je měření míry „šumovosti“ (Noisiness) a „sinusovosti“ (Sinusoidality) neboli také tonality signálu. Jedná se o poměr geometrické a aritmetické střední hodnoty energie spektra. Pokud vyjde výpočet malé číslo, je spektrum spíše tonálního charakteru, pokud velké číslo, je spektrum více ploché nebo se jedná spíše o šum (pro tonální signál by měl hodnotu přibližně 0 a pro šumivé 1). Spektrální plochost lze vypočítat jako:

$$SF(x) = \frac{(\prod_{k \in x} a(k))^{\frac{1}{\kappa}}}{\frac{1}{\kappa} \sum_{k \in x} a(k)}, \quad (4.3)$$

kde  $a(k)$  je amplituda ve frekvenčním pásmu číslo  $k$  [9].



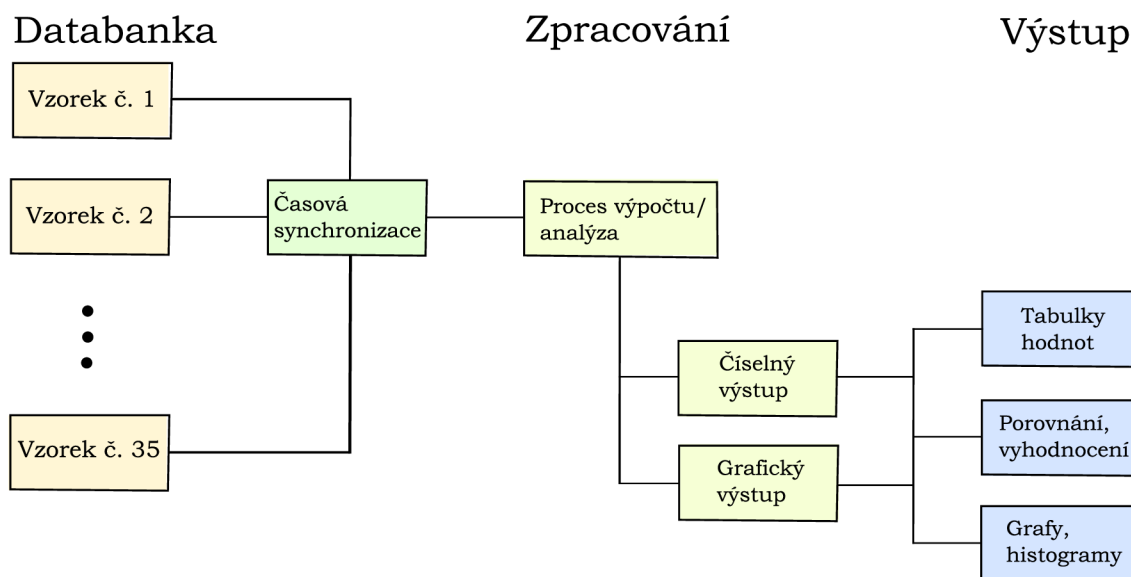
## 5 ANALÝZA

### 5.1 Databanka

Databanka, která byla zadána a následně použita pro analýzu a vyhodnocovací systém, obsahuje 35 interpretací skladby *Smyčkový kvartet No. 1 – IV. Con moto* od Leoše Janáčka. V příložené tabulce A.1 je uvedeno označení čísla skladby, zvukový formát, interpret, země původu a rok vzniku uskupení. Všechny nahrávky jsou bezztrátově zkomprimovány ve formátu WAV (Waveform) se vzorkovací frekvencí 44,1 kHz a rozlišením 16 bitů. Většina pochází ze zdrojového formátu ALAC, některé však byly zpracovány z CD. Dodatečně bylo zjištěno, z jakého státu jednotlivé smyčkové kvartety pochází a v jakém roce byly založeny. Tyto data slouží jako vstupní parametry pro následnou analýzu.

### 5.2 Vyhodnocovací systém

Systém vyhodnocení je vypracován v prostředí *MATLAB* za použití speciálních funkcí balíčku *MIRtoolbox* s nastavitelnými atributy. Některé z nahrávek databanky obsahují ruchy a zvuky, které jsou při výpočtu parametrů nechtěné, jako například potlesk diváků či doba před nástupem první noty instrumentu. Tyto okolnosti by mohly znatelně ovlivňovat výsledek výpočtu. Proto byla nejprve provedena časová synchronizace všech nahrávek. Grafické znázornění vyhodnocovacího systému je na příloženém obrázku 5.1.



Obr. 5.1: Blokové schéma vyhodnocovacího systému.

## 5.3 Výběr motivů a časová synchronizace

Pro analýzu byly vybrány dva motivy (charakteristické úseky) z celé skladby. Oba motivy byly přepsány do elektronické verze pomocí notačního programu *Sibelius* (viz příloha E), nicméně následující údaje o členění taktů jsou vytvořeny podle původního zápisu not, a tudíž jejich číslování s elektronickým zápisem v příloze nebude shodné (původní zápis se nachází v příloženém DVD a byl pořízen na webové stránce *imslp.org*). První motiv začíná v notovém zápise třetí dobou v taktu č. 30 a končí koncem taktu č. 36. Jedná se o sólový úsek prvních houslí. Během těchto sedmi taktů se různě mění dynamika i tempo jednotlivých provedení – dynamika začíná *mezzopianem*, následuje *crescendo* a rychlejší *decrescendo*, na konci fráze na tónu heses *diminuendo* s následným *crescendem* do *piana*. Celý úsek je zapsán v tříosminovém taktu a tempovém označení ( $\text{♩} = \text{cca } 84$ ). Tento motiv byl vybrán právě kvůli možné proměnlivosti interpretací, relativně krátkému trvání (10–20 s) a přítomnosti pouze jediného nástroje z celého kvartetu.

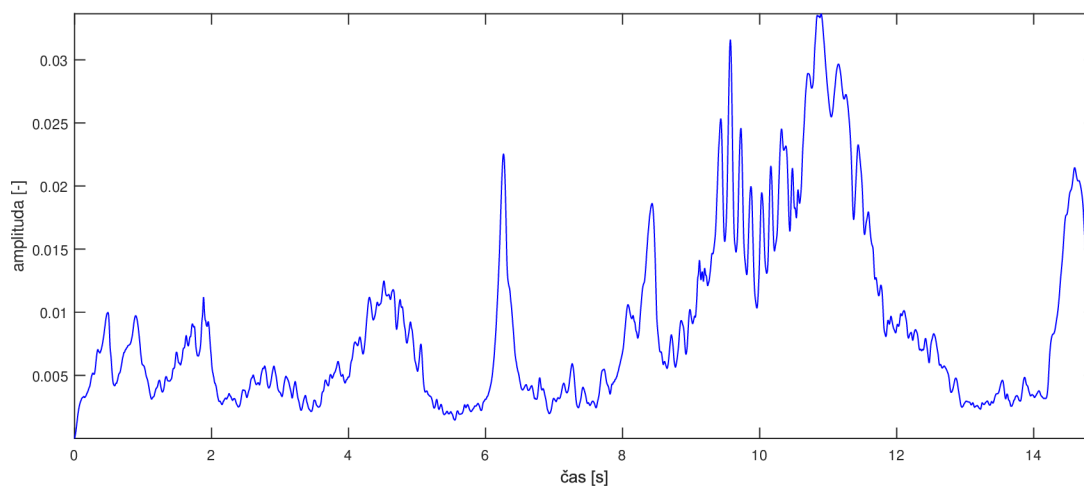
Druhý motiv hned navazuje v notovém zápise na první motiv – začíná jednu dobu před taktom č. 37 a končí až na začátku druhé doby taktu č. 60. Poté skladba moduluje, mění se motiv a také takt na šestišestnáctinový. Tempo je na rozdíl od předchozího motivu již na začátku úseku velmi rychlé ( $\text{♩} = 138$ ) a postupně se ještě zvyšuje pomocí zápisu *poco a poco accelerando*, který se zde opakuje několikrát. Na koncích taktů č. 43 a č. 51 jsou navíc zapsány přesné hodnoty tempa ( $\text{♩} = 144$  a  $\text{♩} = 152$ ) postupného zrychlování. V celém motivu hrají 3 nebo 4 nástroje dohromady, přičemž druhé housle hrají po celý čas ve dvaatřicetinových notách. Úsek byl vybrán z důvodu tempové proměnlivosti jednotlivých interpretací a kvůli jeho kontrastním vlastnostem a charakteristikám vůči prvnímu motivu.

Časová synchronizace a oddělení vybraných motivů od zbytků skladeb bylo provedeno v DAW (Digital Audio Workstation). „Fade-in“ a „fade-out“ byl u každé stopy nastaven na velmi malou hodnotu tak, aby co nejméně ovlivňoval parametry nahrávek. Motivy byly nakonec zvlášť vyexportovány a tím připraveny pro analýzu.

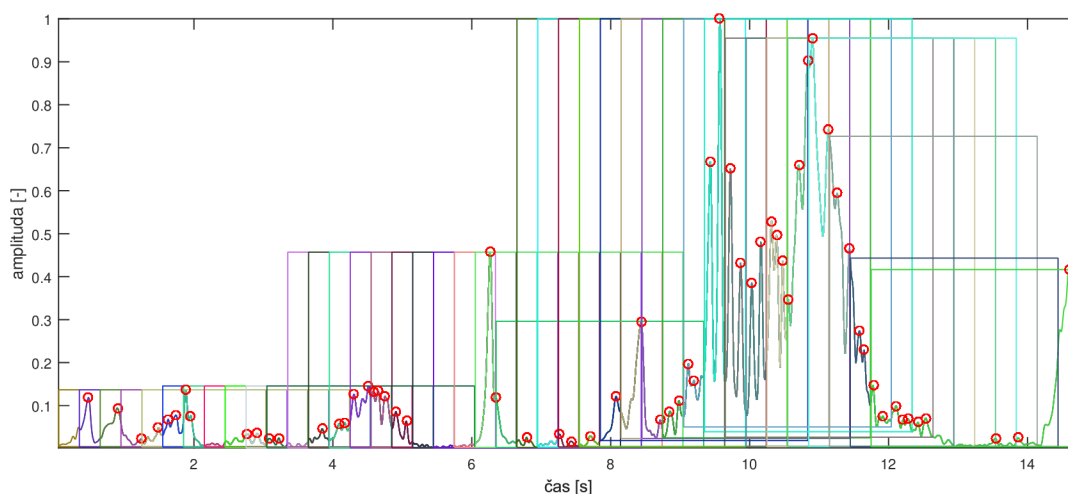
## 5.4 Postup

Podle teorie zpracování hudebních signálů a možností implementace parametrů *MIR-Toolboxu* byly zvoleny metody výpočtů tempa a rytmických struktur nahrávek databanky. Zvolené metody se skládají z následujících procesů: nejprve se extrahuje spektrální obálka (kalkulace spektrogramu), dále se použije usměrňovač, který všechny záporné hodnoty změní na kladné a poté se použije filtr typu dolní propust, který odstraní všechny rychlé oscilace signálu (pomocí autoregresivního filtru s nekonečnou impulzní odezvou (IIR) [14], viz obrázek 5.2). Časová konstanta filtru (na výstupu

je 63 % hodnoty vstupu) je nastavena na 0,02 s. Poté následuje určení začátků not – signál se rozdělí na menší části s délkou okna 3 s a hop faktorem 10 % (0,3 s), což je jedna z dalších nastavitelných metod v toolboxu. Impulzivní shluky energie (lokální extrémy) se označí jako začátky not, jak lze vidět na dalším obrázku 5.3. Dalším krokem je detekční funkce, která následně určí výsledné tempo v čase – průměrné celého úseku nebo dílčích segmentů. První metoda tedy používá spektrální dekompozici křivky doby počátků not, druhá metoda aplikuje na křivku doby počátků not autokorelační funkci (porovnání posunutého signálu s jeho originálem), která je posléze přeložena do frekvenční domény (periody jsou vyjádřeny v Hz místo sekund) a porovnána se spektrální křivkou a následně jsou obě křivky vynásobeny [14]. Parametr jasnost pulzů nakonec nebyl zahrnut do analýzy, protože je obdobou použitých metod a vykazoval velmi podobné výsledky.

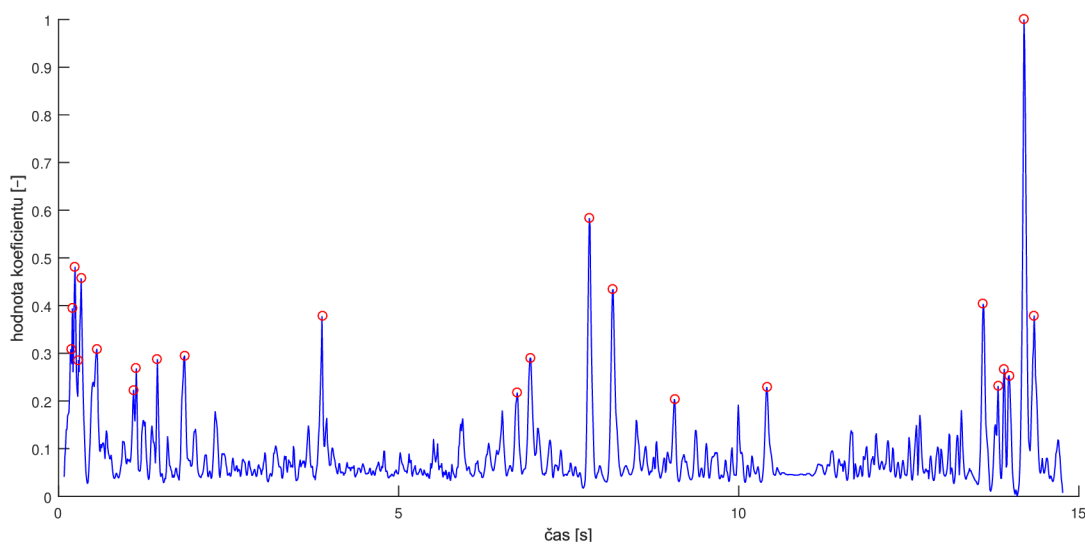


Obr. 5.2: Extrakce obálky pro první skladbu prvního motivu.



Obr. 5.3: Detekce začátků not a rozdělení prvního motivu.

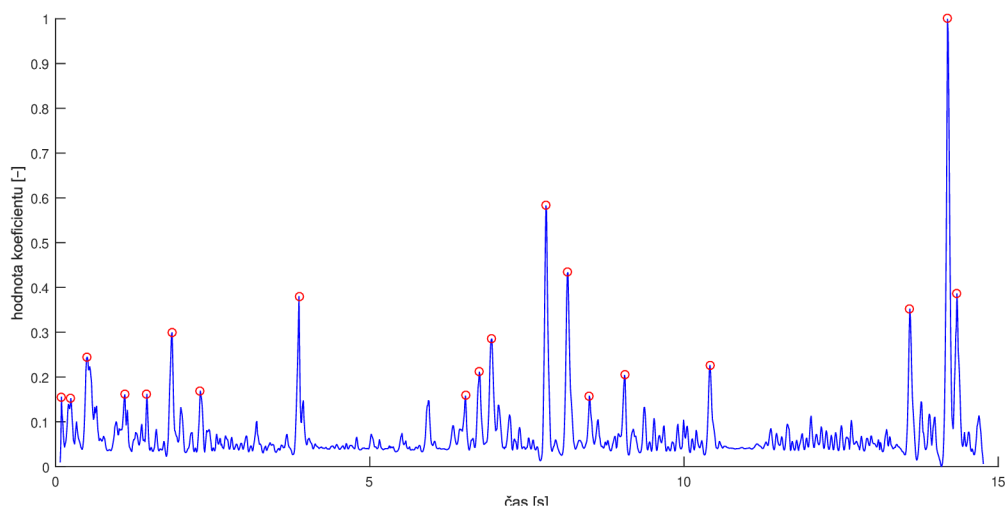
Jedna z testovaných metod, která se nakonec ukázala jako nevhodná, je detekce na základě výpočtu výšky tónu. Výsledek lze vidět na obrázku 5.4, kde je zachycena křivka (rozložená na snímky) určená autokorelační funkcí FFT spektra. Rozsah frekvencí byl ponechán v základním nastavení od 30 do 1000 Hz. Pro odstranění výchylek byl nastaven prahový filtr na 0,2 normalizované hodnoty.



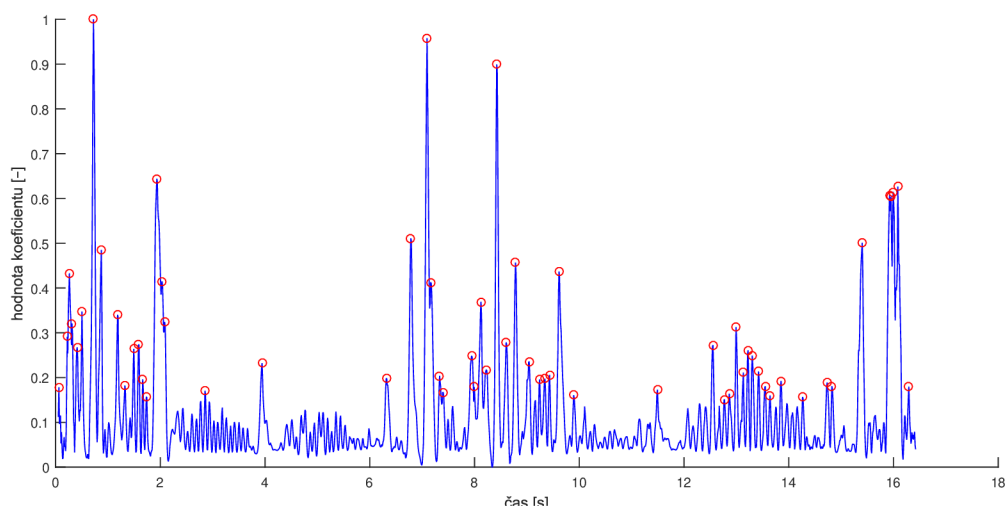
Obr. 5.4: Křivka doby začátků not pomocí určení výšky tónů.

Toto nastavení parametrů je ale nedostačující, protože rozsah zkoumaných frekvencí je příliš velký, což vnáší nepřesnosti do výsledku. Nejnižší tón úseku je des1 (cca 270 Hz) a nejvyšší es2 (cca 620 Hz). Proto byl upraven rozsah frekvencí – od 250

do 650 Hz s prahovým filtrem 0,15 normalizované hodnoty. Výsledek ukazuje obrázek 5.5. Pomocí tohoto postupu bylo detekováno 19 začátků not (v notovém zápise je 15 začátků not) v první nahrávce, což je vůči ostatním metodám dobrá úspěšnost, nicméně pro většinu ostatních nahrávek je tato metoda i přes různá nastavení parametrů v podstatě nepoužitelná. Graf detekce začátků not pro druhou skladbu prvního motivu je znázorněn na obrázku 5.6.



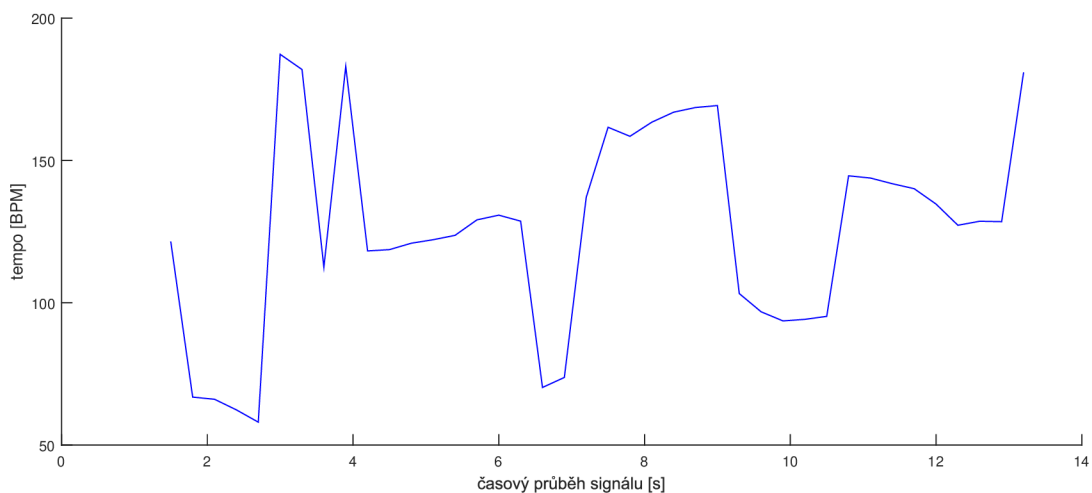
Obr. 5.5: Výsledná křivka doby začátků not první skladby pomocí určení výšky tónů.



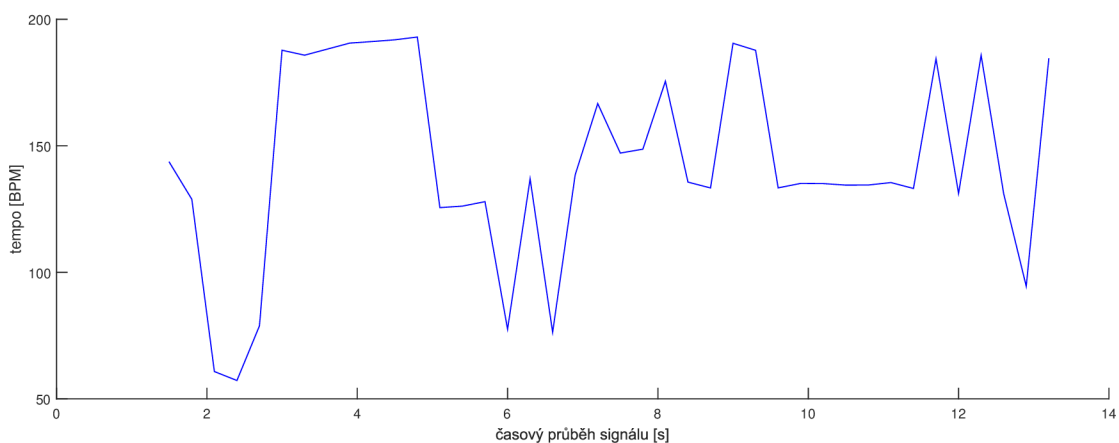
Obr. 5.6: Výsledná křivka doby začátků not druhé skladby pomocí určení výšky tónů.

## 6 VÝSLEDKY

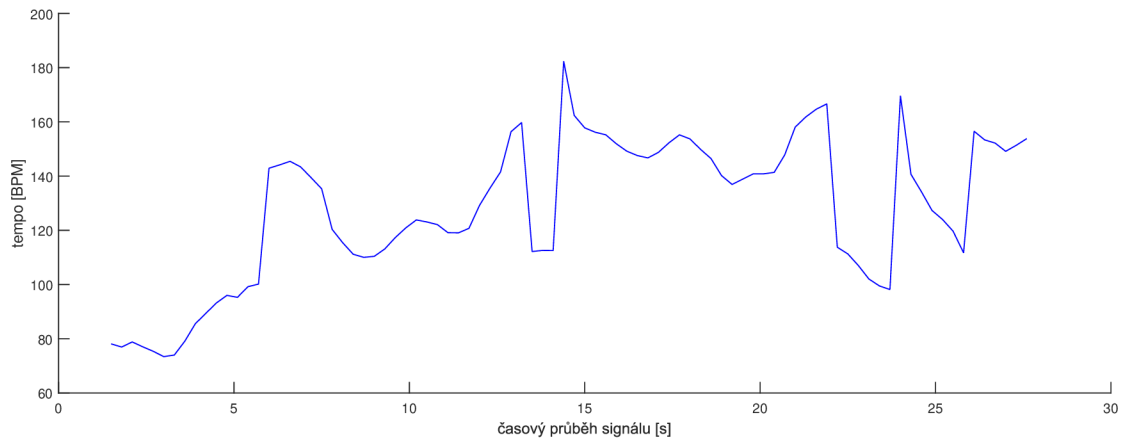
Pomocí dvou metod, zmíněných na začátku předchozí kapitoly 5.4, bylo vypočteno tempo jednotlivých nahrávek obou motivů. Na obrázcích 6.1 a 6.2 je ukázán vývoj tempa nahrávky č. 1 prvního motivu při použití první i druhé metody a na obrázcích 6.3 a 6.4 průběh tempa druhého motivu.



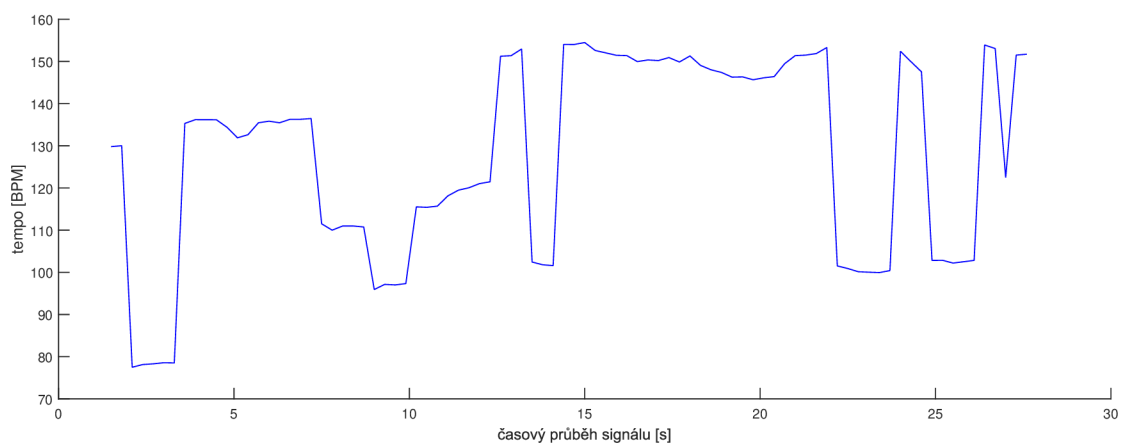
Obr. 6.1: Tempový vývoj motivu č. 1 při použití spektrální dekompozice singálu.



Obr. 6.2: Tempový vývoj motivu č. 1 při použití autokorelační funkce a spektrální dekompozice.



Obr. 6.3: Tempový vývoj motivu č. 2 při použití spektrální dekompozice singálu.



Obr. 6.4: Tempový vývoj motivu č. 2 při použití autokorelační funkce a spektrální dekompozice.

## 6.1 Výsledky metod

Jak lze vidět v tabulce B.1, hodnoty výpočtu pro první motiv se velmi liší od skutečného tempa nahrávek. Skutečné tempo každého záznamu bylo vypočítáno pomocí určení počtu dob v celém motivu vynásobeného číslem 60 (kapitola 2.3.3) a následného vydělení délkou časového úseku motivu. Počet dob z notového zápisu byl určen jako 19 osminových not pro první motiv (6 taktů a 1 doba) a 68 osminových not pro druhý motiv (22 taktů a 2 doby navíc). Rozdíl hodnot je dán především nepřesností, kterou vytváří vlastnosti vstupního materiálu, a také způsob výpočtu – zvolené metody. Metod bylo testováno několik, ale žádná neměla v prvním motivu slibné výsledky. Úsek tvoří expresivní sólo houslí s rychle se měnící dynamikou a velmi výrazným vibratem. Algoritmy často detekovaly oscilace vibrata, které nebylo možné efektivně potlačit nastavením prahu filtru – v druhé polovině motivu byly energeticky výraznější, než všechny skutečné tóny první poloviny motivu, a i proto výpočet ukazoval vyšší hodnoty tempa. Navíc se v daném sólu v podstatě nevyskytuje pravidelná metrická struktura, takže i kdyby program zachytil všechny začátky not a žádná chybná pozitiva, tempo by stále neodpovídalo notovému zápisu. Použitý *MIRToolbox* v základu neobsahuje výpočty na bázi změn fáze či složitější kombinace různých metod – větší úspěšnost by pravděpodobně byla dosažena u jiných žánrů hudby, jiných hudebních nástrojů nebo jiného motivu skladby.

U druhého motivu bylo určení tempa podle výsledků mnohem přesnější, a to i přesto, že zde hraje vícero smyčcových nástrojů zároveň. Výhodou oproti prvnímu motivu je rytmická pravidelnost – úsek totiž obsahuje charakteristickou metrickou strukturu, která se opakuje a je dostatečně výrazná – na každou dobu je zde důraz druhých houslí a ostatní nástroje častěji kopírují důraz na první době taktu. Minimální možná rychlost tempa v jednom snímku je nastavená na 40 BPM a maximální 200 BPM – pouze některé vektorové hodnoty (v řádu jednotek) vykazovaly při jejich exportu *NaN* (Not a Number), což značí výchylku od nastaveného rozsahu hodnot. Tabulka C.1 obsahuje porovnání vypočítaných hodnot se skutečnou hodnotou tempa daného úseku skladby.

Průměr tempa motivu č. 1 všech 35 nahrávek, které bylo vypočítáno první metodou, je 122,28 BPM a druhou metodou 122,92 BPM. Průměr skutečného tempa je však 81,73 BPM. U druhého motivu bylo vypočítáno první metodou 133,61 BPM a druhou metodou 128,80 BPM. Průměr skutečného tempa druhého motivu odpovídá hodnotě 135,10 BPM – rozdíl je tedy výrazně menší. První metoda se u druhého motivu liší v průměru všech skladeb pouze o 1,49 BPM a jeví se tedy jako přesnější, nicméně odchylka u obou metod je velmi malá.

V příloze je dále tabulka D.1, která ukazuje průměrnou hodnotu vektorů (vzniklých rozdělením úseku na tzv. snímky) první nahrávky motivu č. 2, jejich rozptyl,



medián, 1. percentil, 99. percentil a mezipercentilový rozsah. Z hodnot lze vyčíst, že vektory vykazují velké výchyly a jednotlivé snímky se tempově velmi liší, což naznačuje nestabilní vývoj tempa. Ve skutečnosti by tempový vývoj měl být stoupajícího charakteru, ale jak ukazuje graf výpočtu 6.3 a 6.4, je tomu tak pouze částečně. Místa poklesu tempa jsou způsobeny lokální nepřítomností prvních houslí (nejvýraznějším instrumentem motivu), které hrají hlavní melodickou linku, což lze vyčíst i z notového zápisu. V tu chvíli jsou začátky not odečítání pouze z doprovodných nástrojů, které nejsou energeticky tak výrazné a mírně mění charakter výpočtu (tento fakt přináší otázku, nakolik je metoda opravdu přesná a zda přesnost není určena jinými faktory). V následujících tabulkách 6.1 a 6.2 jsou uvedeny průměrné hodnoty tempa prvního i druhého motivu.

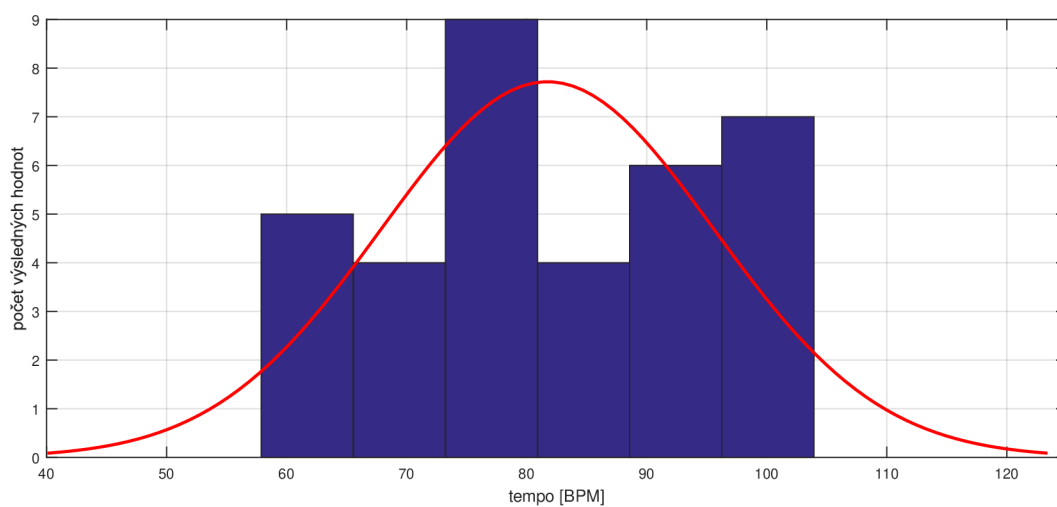
Tab. 6.1: Porovnání průměru tempa všech nahrávek databanky motivu č. 1.

<b>Skutečné tempo</b> [BPM]	<b>První metoda</b> [BPM]	<b>Druhá metoda</b> [BPM]
81,73	122,28	122,92

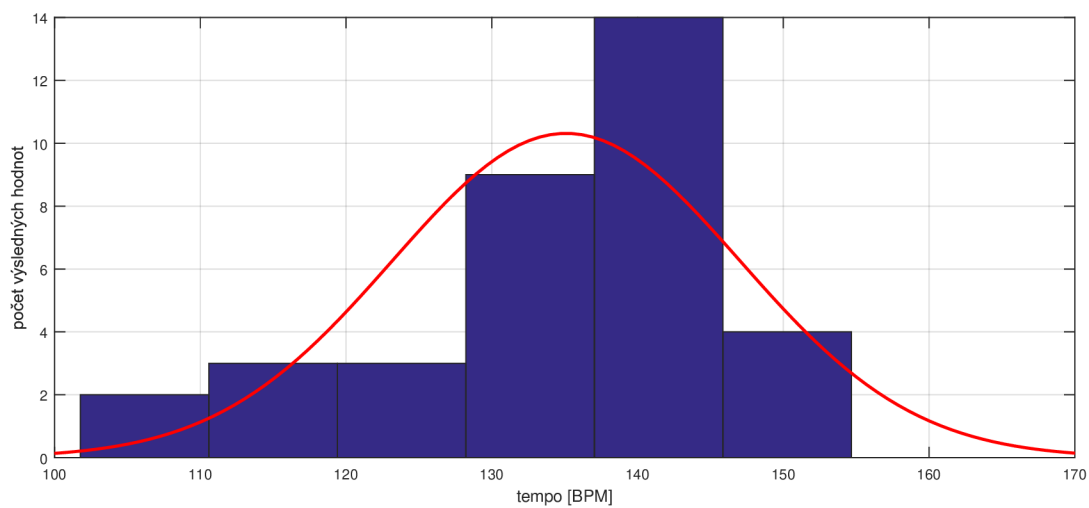
Tab. 6.2: Porovnání průměru tempa všech nahrávek databanky motivu č. 2.

<b>Skutečné tempo</b> [BPM]	<b>První metoda</b> [BPM]	<b>Druhá metoda</b> [BPM]
135,10	133,61	128,79

V grafech 6.5 a 6.6 jsou uvedeny histogramy rozložení četnosti tempa obou motivů. Nejvíce interpretů hrálo první motiv v tempu 73–81 BPM a druhý v tempu 138–146 BPM. Kvarteta tedy docela dodržují předepsané tempo – cca 84 BPM pro první motiv a 138–152 BPM pro druhý motiv, nicméně tempo je v průměru mírně pomalejší.



Obr. 6.5: Histogram rozložení tempa všech skladeb motivu č. 1.



Obr. 6.6: Histogram rozložení tempa všech skladeb motivu č. 2.

## 6.2 Česká republika – Evropa – Svět

V následujících tabulkách 6.3 a 6.4 je uveden rozdíl mezi průměrným tempem jednotlivých českých nahrávek a průměrným tempem všech nahrávek z Evropy (mimo Česko) a zbytku světa – na kolik se každá interpretace liší od průměrného tempa Evropy a světa (6 záznamů pochází z USA, 1 z Izraele a 1 z Austrálie). Kladná hodnota znamená vyšší tempo, záporná nižší tempo české nahrávky, než byl průměr v Evropě či ve zbytku světa. Průměrné tempo Evropy (mimo Česko) je u prvního motivu 80,80 BPM a u druhého motivu 130,32 BPM. Průměrné tempo zbytku světa je u prvního motivu 82,99 BPM a u druhého motivu 141,81 BPM.

Tab. 6.3: Porovnání rozdílů skutečného tempa motivu č. 1 českých interpretů vůči zbytku Evropy a světa.

Nahrávka	Interpret	Evropa [BPM]	zbytek světa [BPM]
010.wav	Janáček String Quartet	−6,96	−9,15
014.wav	Martinu Quartet	−15,22	−17,41
018.wav	Panocha Quartet	−18,61	−20,79
019.wav	Pavel Haas Quartet	−18,78	−20,97
023.wav	Smetana Quartet	−1,59	−3,78
024.wav	Stamitz Quartet	21,50	19,31
025.wav	Škampa Quartet	23,13	20,94
026.wav	Talich Quartet	12,41	10,23
029.wav	Wihan Quartet	19,23	17,05

Průměrné tempo motivu č. 1 nahrávek českých interpretů je 82,48 BPM, takže o 1,68 BPM vyšší, než průměrné tempo Evropy, ale o 0,51 BPM menší, než průměrné tempo zbytku světa. Rozdíly mezi jednotlivými skladbami jsou výrazné (až 23,13 BPM), přesto je průměr tempa všech skladeb dohromady velmi podobný a odchylka malá.

Tab. 6.4: Porovnání rozdílů skutečného tempa motivu č. 2 českých interpretů vůči zbytku Evropy a světa.

Nahrávka	Interpret	Evropa [BPM]	zbytek světa [BPM]
010.wav	Janáček String Quartet	5,86	-5,63
014.wav	Martinu Quartet	-13,16	1,67
018.wav	Panocha Quartet	9,78	-1,71
019.wav	Pavel Haas Quartet	13,11	1,61
023.wav	Smetana Quartet	-12,85	-24,34
024.wav	Stamitz Quartet	10,36	-1,14
025.wav	Škampa Quartet	18,52	7,03
026.wav	Talich Quartet	10,16	-1,33
029.wav	Wihan Quartet	7,34	-4,16

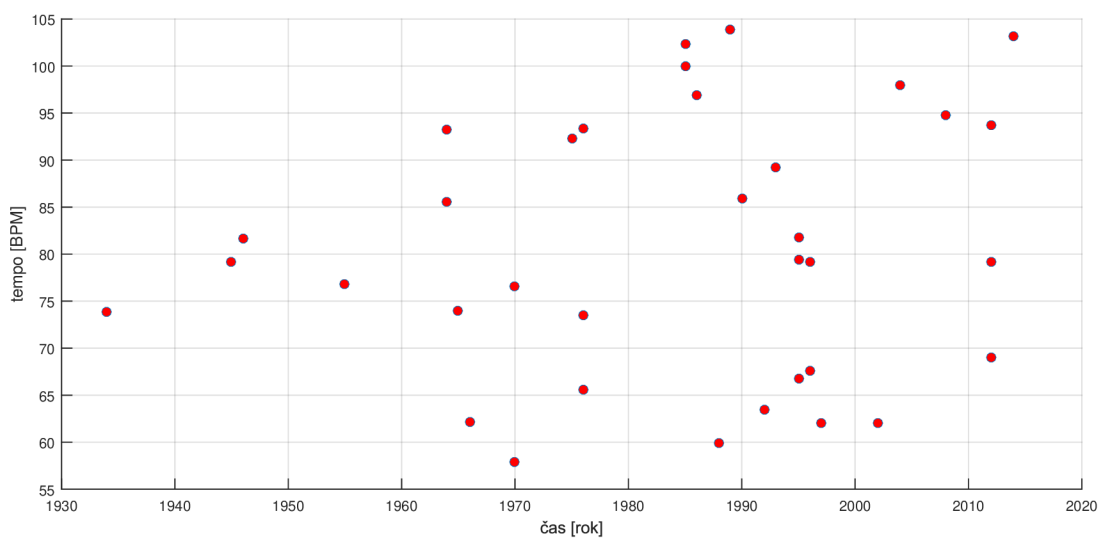
Rozdíly mezi českými a evropskými nahrávkami jsou u motivu č. 2 větší, než mezi českými a světovými. Průměrné tempo českých nahrávek je 138,70 BPM – o 8,38 BPM větší, než Evropy a o 3,11 BPM menší, než zbytku světa. Pouze dvě české interpretace jsou v pomalejším tempu k průměru Evropy. Vůči zbytku světa jsou nahrávky naopak většinou pomalejší.

### 6.3 Vývoj tempa během let

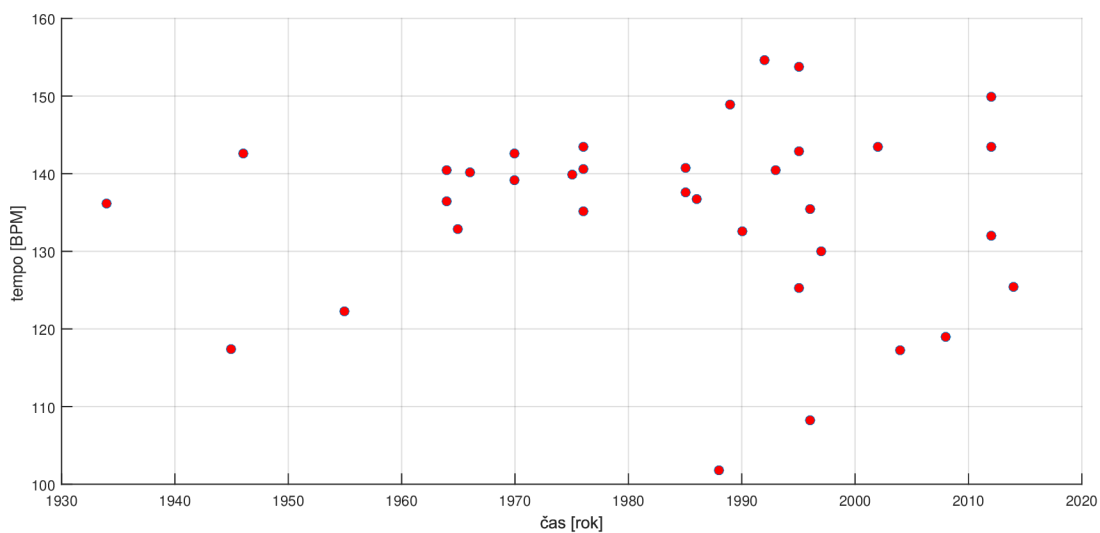
V databance A.1 je mimo jiné uveden i rok vzniku kvartetu. Pomocí těchto údajů byla vytvořena závislost mezi tempem nahrávky a rokem vzniku uskupení. Členové kvartetů se však obměňují a není zde dále dohledáno, jestli skladbu hraje původní sestava. Navíc rok založení souboru nemusí korespondovat s rokem vytvoření nahrávky (což je obtížně dohledatelné), nicméně určitou vypovídající hodnotu tato analýza poskytuje.

Lze tedy říct, že první motiv je tempově rozmanitější – jsou zde větší rozdíly jednotlivých nahrávek od průměrné hodnoty celého úseku. Tento fakt odpovídá expresivní povaze motivu – sólový part houslí, u kterého je ponechána volnost interpretace. Pozdější provedení vykazují obecně větší tempové výkyvy. Druhý motiv je mnohem vyrovnanější – odchylky jednotlivých nahrávek od průměrného tempa jsou menší, než u prvního motivu. U nahrávek druhého motivu souborů, které vznikly mezi lety 1960 a 1980, lze pozorovat velmi podobné hodnoty, které se vychylují od předepsaného tempa jen velmi málo.

V grafech 6.7 a 6.8 jsou uvedeny závislosti hodnot tempa podle roku vzniku kvartetu.



Obr. 6.7: Hodnota tempa všech skladeb motivu č. 1 podle roku vzniku kvarteta.



Obr. 6.8: Hodnota tempa všech skladeb motivu č. 2 podle roku vzniku kvarteta.

## 7 ZÁVĚR

V této bakalářské práci je uvedena historie a teorie základních faktorů a vlastností transkripce audio signálů na notový záznam. Další podkapitoly zmiňují umělecké, ale i (a to převážně) technické problémy analýzy hudebních signálů a určování základních vlastností záznamů. Dále je v textu popsáno zpracování a rozdělení transkripce jako takové a je uvedeno základní rozdělení parametrů zvukových nahrávek na tři hlavní oblasti – tónbr, dynamika a čas. Nejvíce je rozebrán časový aspekt, tempo a jejich rozdělení a definice. Během celé práce jsou postupně zmíněny odborné studie, které pracují s danou problematikou.

Ve třetí kapitole jsou vysvětleny technické parametry a jejich použití. Je zde popsána doba začátků not, různé způsoby jejího zjišťování, a s tím související metody detekční funkce. Práce popisuje také postup při zjišťování přesnosti detekce not v audio signálu.

V neposlední řadě práce okrajově uvádí ostatní parametry, které se využívají v transkripčních algoritmech pro zjištění ostatních vlastností zvukových záznamů. Jedná se o chroma vektor, jasnost zvuku, jasnost pulzů a spektrální plochost.

V páté kapitole jsou uvedeny informace o souborech v databance a je zde předložen vyhodnocovací systém. Následuje popis předzpracování nahrávek a výběru motivů, které tvoří vstupní data analýzy, a vysvětlení postupu výpočtu dvou použitých metod pro zjištění tempa hudebního signálu.

Poslední kapitola shrnuje dosažené výsledky výpočtů jednotlivých metod a vysvětluje proces určení skutečného tempa nahrávek. Analýza ukázala, že určení tempa je pro danou skladbu velmi obtížné. Celá skladba, ale hlavně první vybraný motiv, vykazuje velmi nepravidelné metrum a silné nežádoucí jevy pro většinu výpočtů. Přesto byla dosažena vysoká přesnost při určení tempa druhého motivu. Práce také porovnává nahrávky z České republiky, Evropy a zbytku světa. Podle výsledků čeští hudebníci hrají většinou oba motivy rychleji než zbytek Evropy, ale zároveň pomaleji, než smyčcové soubory ze zbytku světa.

Nakonec práce uvádí rozbor tempa nahrávek v závislosti na roku vzniku daného interpreta. Výrazný tempový rozdíl nahrávek v závislosti na roku vzniku uskupení pozorován není, pouze se obecně zvětšují tempové výchylky mezi jednotlivými interpretacemi s pozdější dobou vzniku kvarteta.

## LITERATURA

- [1] ROTHSTEIN, Joseph. *MIDI: a comprehensive introduction* 2nd ed. Madison, Wis.: A-R Editions, c1995. ISBN 0895793091.
- [2] CONT, Arshia, Diemo SCHWARZ, Norbert SCHNELL a Christopher RAPHAEEL. *Evaluation of Real-Time Audio-to-Score Alignment* [online]. International Symposium on Music Information Retrieval (ISMIR), 2007, Vienna, Austria. [cit. 2016-12-06]. Dostupné z URL: <<https://hal.inria.fr/hal-00839068>>.
- [3] LERCH, Alexander. *An Introduction to Audio Content Analysis Applications in Signal Processing and Music Informatics* [online]. Ausg. New York: Wiley, 2012. ISBN 9781118393505. Dostupné z URL: <<http://eu.wiley.com/WileyCDA/WileyTitle/productCd-111826682X.html>>.
- [4] OTSUKA, Takuma, Kazuhiro NAKADAI, Tetsuya OGATA a Hiroshi OKUNO. *Incremental Bayesian Audio-To-Score Alignment with Flexible Harmonic Structure Models* [online]. ISMIR 2011: proceedings of the 12th International Society for Music Information Retrieval Conference, Miami: University of Miami, 2011. s. 525-530, ISBN 9780615548654. Dostupné z URL: <[http://ismir2011.ismir.net/ISMIR2011\\_complete\\_proceedings.pdf](http://ismir2011.ismir.net/ISMIR2011_complete_proceedings.pdf)>.
- [5] ALBRECHT, Štěpán. *Model-based Approaches for Automatic Transcription of Music* [online]. Plzeň, 2013 [cit. 2016-12-07]. Disertace. University of West Bohemia, Department of Computer Science and Engineering. Vedoucí práce Prof. Ing. Václav Matoušek, CSc. Dostupné z URL: <[https://otik.uk.zcu.cz/bitstream/handle/11025/12980/d05\\_06\\_final.pdf?sequence=1](https://otik.uk.zcu.cz/bitstream/handle/11025/12980/d05_06_final.pdf?sequence=1)>.
- [6] KLAPURI, Anssi. *Signal processing methods for the automatic transcription of music* [online]. Tampere: Tampere University of Technology, 2004. ISBN 9521511478. Dostupné z URL: <[https://www.cs.tut.fi/sgn/arg/klap/phd/klap\\_phd.pdf](https://www.cs.tut.fi/sgn/arg/klap/phd/klap_phd.pdf)>.
- [7] PERETZ, Isabelle. *Music perception and recognition* [online]. In The Handbook of Cognitive Neuropsychology, B. Rapp (Ed.), Hove: Psychology Press, pp. 519–540, 2001 [cit. 2016-12-07]. Dostupné z URL: <<http://psycnet.apa.org/psycinfo/2001-16360-021>>.
- [8] ZATORRE, Robert, Pascal BELIN a Virginia PENHUNE. *Structure and function of auditory cortex: music and speech* [online]. TRENDS in Cognitive Sciences, 6(1), 2002. s. 37–46, [cit. 2016-12-07]. Dostupné z URL: <<http://psych.colorado.edu/~kimlab/zatorre.etal.tics2002.pdf>>.

- [9] PEETERS, Geoffroy. *A large set of audio features for sound description (similarity and classification) in the CUIDADO project* [online]. CUIDADO I.S.T. Project Report, 2004 [cit. 2016-12-07]. Dostupné z URL: <[http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters\\_2003\\_cuidadoaudiofeatures.pdf](http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf)>.
- [10] CYRIL, Joder, Essid SLIM a Richard GAEL. *An Improved Hierarchical Approach for Music-to-Symbolic Score Alignment* [online]. Institut TELECOM, TELECOM ParisTech, CNRS LTCI. ISMIR, 2010 [cit. 2016-12-07]. Dostupné z URL: <<http://biblio.telecom-paristech.fr/cgi-bin/download.cgi?id=10383>>.
- [11] RYYNÄNEN, Matti a Anssi KLAPURI. *Transcription of the Singing Melody in Polyphonic Music* [online]. 7th International Conference on Music Information Retrieval, Victoria, Canada, 2006 [cit. 2016-12-07]. Dostupné z URL: <[http://www.cs.tut.fi/sgn/arg/matti/ryynanen\\_ismir06.pdf](http://www.cs.tut.fi/sgn/arg/matti/ryynanen_ismir06.pdf)>.
- [12] KLAPURI, Anssi a Manuel DAVY. *Signal Processing Methods for Music Transcription* [online]. New York, 2006 [cit. 2016-12-07]. Dostupné z URL: <<http://www.springer.com/la/book/9780387306674>>.
- [13] TAYLOR, B. N. *Guide for the use of the International System of Units (SI)* [online]. 1995 ed. Washington, D.C.: For sale by the Supt. of Docs., U.S. G.P.O., 1995 [cit. 2016-12-07]. Dostupné z URL: <<http://physics.nist.gov/cuu/pdf/sp811.pdf>>.
- [14] 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. 2016-12-07]. Dostupné z URL: <<http://dafx.labri.fr/main/papers/p237.pdf>>.
- [15] WRIGHT, Matthew. *The Shape of an Instant: Measuring and Modeling Perceptual Attack Time with Probability Density Functions* [online]. Disertace. Stanford University, Stanford, 2008 [cit. 2016-12-07]. Dostupné z URL: <<https://ccrma.stanford.edu/~matt/diss/Matthew-Wright-Dissertation.pdf>>.
- [16] 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.



- [17] GABRIELSSON, Alf. *Music performance research at the millennium* [online]. Uppsala University, Sweden, 2003, 31(3), [cit. 2016-12-07]. Dostupné z URL: <<https://www.scribd.com/document/280528916/GABRIELSSON-Music-Performance-Research-Millennium>>.
- [18] GOEBL, Werner a Simon DIXON. *Analysis of Tempo Classes in Performances of Mozart Sonatas* [online]. Vienna, Austria: Austrian Research Institute for Artificial Intelligence, 2001 [cit. 2016-12-07]. Dostupné z URL: <[https://iwk.mdw.ac.at/goebl/papers/GoeblDixon2001\\_CogMus-Finland-tempoClasses.pdf](https://iwk.mdw.ac.at/goebl/papers/GoeblDixon2001_CogMus-Finland-tempoClasses.pdf)>.
- [19] 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. 2016-12-07]. Dostupné z URL: <<http://www.cs.tut.fi/~digaudio/htyo/lahteet/p032-page-158-paper191.pdf>>.
- [20] LARTILLOT, Olivier, Petri TOIVIAINEN a Tuomas EEROLA. *A Matlab Toolbox for Music Information Retrieval* [online]. Springer-Verlag: University of Jyväskylä, 2008 [cit. 2016-12-07]. Dostupné z URL: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.706.2450&rep=rep1&type=pdf>>.
- [21] GORDON, John William. *Perception of Attack Transients in Musical Tones* [online]. Stanford, 1984 [cit. 2016-12-07]. Disertace. Stanford University, Center for Computer Research in Music and Acoustics (CCRMA). Dostupné z URL: <<https://ccrma.stanford.edu/files/papers/stanm17.pdf>>.
- [22] FOOTE, Jonathan. *Automatic Audio Segmentation Using a Measure of Audio Novelty* [online]. In Proceedings of the International Conference on Multimedia and Expo (ICME), New York: Institute of Electrical and Electronics Engineers (IEEE), 2000, DOI: 10.1109. ICME.2000.869637. Dostupné z URL: <<https://www.fxpal.com/publications/automatic-audio-segmentation-using-a-measure-of-audio-novelty.pdf>>.
- [23] MÜLLER, Meinard. *Fundamentals of Music Processing* [online]. New York, NY: Springer Berlin Heidelberg, 2015. ISBN 9783319219455/DOI: 10.1007. [cit. 2016-12-07]. Dostupné z URL: <<http://www.springer.com/us/book/9783319219448>>.

- [24] RÖBEL, Axel. *Onset Detection in Polyphonic Signals by means of Transient Peak Classification* [online]. In Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR). Paris, 2009 [cit. 2016-12-07]. Dostupné z URL: <<http://articles.ircam.fr/textes/Roebel09a/index.pdf>>.
- [25] HARTE, Christopher a Mark SANDLER. *Automatic Chord Recognition Using Quantised Chroma and Harmonic Change Segmentation* [online]. Mile End Road, London, UK: Queen Mary, University of London, 2009 [cit. 2016-12-07]. Dostupné z URL: <[http://music-ir.org/mirex/results/2009/abs/harte\\_mirex09.pdf](http://music-ir.org/mirex/results/2009/abs/harte_mirex09.pdf)>.
- [26] BELLO, Juan PABLO. *Chroma and tonality* [online]. New York University: MPATE-GE 2623 Music Information Retrieval, 2008. [cit. 2016-12-07]. Dostupné z URL: <[http://www.nyu.edu/classes/bello/MIR\\_files/tonality.pdf](http://www.nyu.edu/classes/bello/MIR_files/tonality.pdf)>.
- [27] KNEES, Peter a Markus SCHEDL. *Music similarity and retrieval: an introduction to audio- and web-based strategies* [online]. New York, NY: Springer Berlin Heidelberg, 2016. [cit. 2016-12-08]. ISBN 9783662497207. Dostupné z URL: <<http://www.springer.com/us/book/9783662497203>>.

## SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

ADSR	Attack, Decay, Sustain, Release – náběh, útlum, podržení, uvolnění
AOT	Acoustic Onset Time – akustická doba počátku
ASP	Audio Signal Processing – zpracování audio signálů
AtSA	Audio-to-Score Alignment – převod audio signálu na notový zápis
BPM	Beats Per Minute – údery za minutu
DAW	Digital Audio Workstation – digitální zvukové pracoviště
DTW	Dynamic Time Warping – dynamické borcení časové osy
FP	False Positives – chybná pozitiva
MA	Moving Average – přizpůsobivý, pohyblivý práh
MFFE	Multiple Fundamental Frequency Estimation – výpočet více fundamentálních frekvencí zároveň
MIDI	Musical Instrument Digital Interface
MME	Musical Meter Estimation – výpočet hudebního tempa a rytmu
NaN	Not a Number – hodnota není číslo
NOT	Note Onset Time – doba počátku noty
ONSET	Onset – začátek noty
PC	Pulse Clarity – jasnost pulzů
POT	Perceptual Onset Time – percepční doba počátku
RMS	Root Mean Square – kvadratický průměr
SC	Spectral Centroid – spektrální těžiště
SF	Spectral Flatness – spektrální plochost
SFL	Spectral Flux – spektrální tok
STFT	Short Time Fourier Transform – krátkodobá Fourierova transformace

# SEZNAM PŘÍLOH

A Databanka	49
B První motiv	51
C Druhý motiv	53
D Analýza hodnot segmentování	55
E Notový přepis skladby	57
F Obsah přiloženého DVD	61

## A DATABANKA

Tab. A.1: Informace o souborech v databance.

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

033.wav	Quatuor Diotima	Francie	1996
034.wav	Australian Chamber Orchestra, R. Tognetti	Austrálie	1975
035.wav	London Sinfonietta and Wiener Phil.	V. Británie a Rakousko	2004

## B PRVNÍ MOTIV

Tab. B.1: Porovnání skutečného tempa prvního motivu a výpočtu tempa pomocí dvou zvolených metod.

Nahrávka	Skutečné tempo [BPM]	První metoda [BPM]	Druhá metoda [BPM]
01	76,55	126,35	143,60
02	69,00	115,80	99,78
03	63,43	124,82	125,13
04	79,22	126,41	110,56
05	73,51	125,24	119,37
06	76,67	123,99	133,76
07	85,50	105,50	123,67
08	57,88	122,00	128,08
09	79,36	131,57	116,04
10	73,84	128,66	149,05
11	89,25	138,68	122,42
12	81,67	122,10	130,18
13	59,88	127,03	141,14
14	65,58	115,79	110,13
15	67,61	117,30	130,39
16	73,96	110,13	106,69
17	66,72	136,47	128,54
18	62,19	118,60	131,31
19	62,02	113,86	127,30
20	85,92	138,31	125,73
21	93,75	118,72	116,59
22	93,33	105,41	108,41
23	79,21	122,58	138,64
24	102,30	122,14	118,42
25	103,93	133,80	131,51
26	93,21	130,32	106,44
27	99,04	145,19	115,99
28	96,85	126,54	135,85
29	100,03	138,84	100,85
30	81,78	104,81	108,93

31	94,82	117,15	146,23
32	103,16	91,63	117,13
33	79,17	129,37	127,21
34	92,32	118,40	115,66
35	98,00	106,23	111,35



## C DRUHÝ MOTIV

Tab. C.1: Porovnání skutečného tempa druhého motivu a výpočtu tempa pomocí dvou zvolených metod.

Nahrávka	Skutečné tempo [BPM]	První metoda [BPM]	Druhá metoda [BPM]
01	139,22	127,92	127,99
02	132,03	125,48	113,86
03	154,67	151,88	143,53
04	149,87	131,79	121,39
05	140,64	145,02	140,47
06	122,30	130,10	125,21
07	136,41	113,26	117,62
08	142,58	152,25	144,02
09	125,24	120,51	115,59
10	136,18	130,27	129,04
11	140,39	135,40	130,95
12	142,62	129,09	129,54
13	101,77	127,10	134,34
14	143,48	138,56	125,60
15	108,25	125,16	112,60
16	132,84	134,25	132,65
17	153,82	144,41	134,75
18	140,10	132,01	130,10
19	143,42	136,46	140,09
20	132,52	138,21	122,12
21	143,42	132,18	124,83
22	135,10	134,03	138,33
23	117,47	120,27	116,58
24	140,67	139,61	130,45
25	148,84	136,31	132,32
26	140,48	136,04	132,05
27	129,98	140,71	140,05
28	136,73	145,23	139,47
29	137,65	143,49	131,71
30	142,83	144,10	140,55

31	118,97	126,83	120,41
32	125,37	115,53	116,30
33	135,47	135,35	128,21
34	139,90	135,35	127,05
35	117,20	122,32	117,97

## D ANALÝZA HODNOT SEGMENTOVÁNÍ

Tab. D.1: Rozbor hodnot vektorů z dekompozice druhého motivu při použití spektrální obálky.

Číslo skladby	Prům. hodnota (mean)	Rozptyl (var)	Směr. odch. (std)	Medián	1. percentil	99. percentil	Meziperc. rozsah
1	127,92	705,87	26,72	134,75	88,24	171,28	83,04
2	125,48	458,06	21,52	126,51	96,13	163,83	67,70
3	151,88	694,16	26,52	160,88	115,42	185,17	69,75
4	131,79	1050,67	32,62	130,22	88,32	186,05	97,72
5	145,02	690,77	26,43	142,41	117,50	192,88	75,39
6	130,10	318,58	17,94	127,52	118,26	189,65	71,39
7	136,41	566,36	23,93	116,72	77,31	157,01	79,69
8	142,58	494,16	22,36	156,11	125,49	188,97	63,48
9	125,24	1367,55	37,17	117,54	73,39	187,15	113,76
10	136,18	926,68	30,61	138,12	80,72	193,11	112,38
11	140,39	388,39	19,82	142,07	103,99	167,74	63,74
12	142,62	1308,52	36,39	125,31	92,28	191,97	99,69
13	101,77	1001,15	52,05	130,44	87,92	190,30	102,39
14	143,47	842,72	29,20	133,06	104,90	187,93	83,03
15	108,25	973,75	31,34	118,51	84,74	187,07	102,33
16	132,84	835,44	29,06	132,52	101,42	192,33	90,91

17	153,82	750,27	27,57	154,38	88,34	185,36	97,03
18	140,10	566,76	23,94	136,42	112,46	163,86	51,40
19	143,42	772,09	27,95	139,66	91,78	180,72	88,95
20	132,52	835,28	29,06	129,26	107,57	190,04	82,47
21	143,42	773,48	27,98	126,81	105,22	186,82	81,61
22	135,10	248,02	15,84	139,65	103,66	152,24	48,59
23	117,47	1396,15	37,54	119,73	71,15	190,01	118,86
24	140,67	859,89	29,49	142,65	94,53	185,50	90,97
25	148,84	541,91	23,42	143,39	103,83	169,03	65,20
26	140,48	731,98	27,21	136,91	85,55	185,10	99,55
27	129,98	497,34	22,42	135,64	117,64	190,50	72,86
28	136,73	918,41	30,47	154,48	98,84	190,42	91,58
29	137,65	1076,75	33,00	143,04	84,09	191,88	107,79
30	142,83	714,52	26,89	149,57	103,28	187,97	84,69
31	118,97	788,10	30,68	124,41	94,90	190,93	96,03
32	125,37	684,15	26,29	113,22	86,81	167,90	81,09
33	135,47	784,44	31,40	141,40	101,24	191,25	90,01
34	139,90	1009,57	31,96	136,64	90,30	187,31	97,01
35	117,20	802,29	28,46	122,01	85,09	191,06	105,97

# E NOTOVÝ PŘEPIS SKLADBY

Obr. E.1: Notový zápis prvního motivu.

## Motiv 1

Leos Janacek

[♩ = cca 84]

Violin I *mp* *dim.* *rit.* *< p*

Violin II

Viola

Violoncello

Obr. E.2: Notový zápis druhého motivu.

## Motiv 2

Leos Janacek

Un poco più mosso

[Con moto ♩ = 138]

Violin I *p*

Violin II *pp senza sord.*

Viola *mf [espress.]*

Violoncello

Measures 1-3 of the score. The key signature is three flats (B-flat major/C minor). The time signature is 3/8. The first measure shows the Violin I part starting with a piano (*p*) dynamic. The Violin II part has a *pp senza sord.* marking. The Viola part has a *mf [espress.]* marking. The Cello part is silent.

Measures 4-6 of the score. The Violin I part has a fermata over the first measure. The Violin II part continues with a sixteenth-note pattern. The Viola part has a *mf* marking. The Cello part is silent.

7 [poco accel. al ♩ = 144]

*mf*

Measures 7-9 of the score. The key signature changes to two flats (B-flat major/C minor). The time signature is 3/8. The first measure of this system has a *mf* marking. The tempo marking is [poco accel. al ♩ = 144]. The Violin I part has a *mf* marking. The Violin II part continues with a sixteenth-note pattern. The Viola part has a *mf* marking. The Cello part is silent.

2

[poco a poco

10

Musical score for measures 10-13. The score is in 3/4 time with a key signature of three flats. It features a melody in the upper voice, a dense piano accompaniment in the right hand, and a bass line in the left hand. Dynamics include 'f' (forte) and 'poco a poco'.

accelerando

al

$\text{♩} = 152$

14

Musical score for measures 14-17. The score continues with the same instrumentation. It includes the instruction 'accelerando al [♩ = 152]' and 'cresc.' (crescendo). Dynamics include 'f' (forte).

18

Musical score for measures 18-21. The score continues with the same instrumentation. It features a continuation of the piano accompaniment and bass line.

22

Musical score for measures 22-24. The score is written for four staves: two treble clefs (top two) and two bass clefs (bottom two). The key signature is three flats (B-flat, E-flat, A-flat). The time signature is 6/16. Measure 22 shows a melodic line in the upper treble staff and a bass line in the lower bass staff. Measure 23 continues the melodic and bass lines. Measure 24 features a complex texture with multiple voices. Dynamic markings include *mf* (mezzo-forte) in the upper treble staff and *fp* (fortissimo) in the lower bass staff. The notation includes various note values, rests, and articulation marks.



## F OBSAH PŘILOŽENÉHO DVD

Přiložené DVD obsahuje několik dokumentů týkajících se práce. Hlavním dokumentem je elektronická verze práce s názvem „Bakalárska práce – Matej Istvanek“. DVD dále obsahuje seznam použitých příkazů v jazyku *MATLAB* a dvě složky (*Motiv1* a *Motiv2*) s časově synchronizovanými úseky dané skladby. Následuje soubor *Seznam* s informacemi o databance, elektronická verze not celé skladby (včetně 4. části) a složka *Hodnoty\_tempa*, ve které se nachází soubory typu *EXCEL* s implementovanými hodnotami vektorů několika metod obou motivů s následnou analýzou a také soubor *Skutecne\_tempo* obsahující výpočet skutečného tempa skladeb na základě jejich délky a počtu dob hudebního úseku.

Obsah:

- Bakalárska práce – Matej Istvanek.pdf
- Prikazy.m
- Seznam.xlsx
- Noty.pdf
- Motiv1
  - 001.wav – 035.wav
- Motiv2
  - 001.wav – 035.wav
- Hodnoty\_tempa
  - Skutecne\_tempo.xlsx
  - Tempo\_motiv1.xlsx
  - Tempo\_motiv2.xlsx