

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2024

Bc. Viktor Slezák



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

SVĚTELNÉ ANIMACE PRO SYSTÉM SPECTODA NA ZÁKLADĚ ANALÝZY PARAMETRŮ Z HUDEBNÍCH NAHRÁVEK

LIGHT ANIMATIONS FOR THE SPECTODA SYSTEM BASED ON THE ANALYSIS OF PARAMETERS FROM
MUSIC RECORDINGS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Viktor Slezák

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Matěj Ištvanek, Ph.D.

BRNO 2024



Diplomová práce

magisterský navazující studijní program **Audio inženýrství**
specializace Zvuková produkce a nahrávání
Ústav telekomunikací

Student: Bc. Viktor Slezák

ID: 203745

Ročník: 2

Akademický rok: 2023/24

NÁZEV TÉMATU:

Světelné animace pro systém Spectoda na základě analýzy parametrů z hudebních nahrávek

POKYNY PRO VYPRACOVÁNÍ:

Vytvořte systém pro výpočet parametrů z hudební nahrávky s důrazem na dynamickou, rytmickou a akordickou strukturu. Otestujte výhody a nevýhody nejnovějších přístupů založených na metodách strojového učení pro extrakci relevantních parametrů. Získaná data analyzujte a na jejich základě navrhněte a naprogramujte algoritmus generující specifický kód „SpectodaCode“ pro následné vytváření světelných animací. Výstupem práce bude jednoduché webové rozhraní, které po nahrání hudební skladby vygeneruje unikátní světelné animace.

DOPORUČENÁ LITERATURA:

- [1] MÜLLER, Meinard. Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications. Cham: Springer International Publishing, 2015. ISBN 978-3-319-21945-5.
- [2] CARSAULT, Tristan, NIKA, Jérôme, ESLING, Philippe a ASSAYAG, Gérard. 2021. Combining Real-Time Extraction and Prediction of Musical Chord Progressions for Creative Applications. Electronics, vol. 10, no. 21: 2634. DOI <https://doi.org/10.3390/electronics10212634>.

Termín zadání: 5.2.2024

Termín odevzdání: 21.5.2024

Vedoucí práce: Ing. Matěj Ištvanek, Ph.D.

doc. Ing. Jiří Schimmel, Ph.D.

předseda rady studijního programu

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

V práci je prozkoumána problematika oboru Music information retrieval. Na základě získaných znalostí je navržena struktura systému pro generování animací z parametrů hudební nahrávky. Dále jsou popsány a porovnány možnosti jak tyto parametry extra-hovat.

KLÍČOVÁ SLOVA

MIR, extrakce parametrů, audio analýza, generátor animací

ABSTRACT

In this paper, is explored the field of Music Information Retrieval. Based on the knowledge acquired, a system structure for generating animations from parameters of a music recording is proposed. Furthermore, the possibilities for extracting these parameters are described and compared.

KEYWORDS

MIR, feature extraction, audio analysis, animation generator

SLEZÁK, Viktor. *Světelné animace pro systém Spectoda na základě analýzy parametrů z hudebních nahrávek*. Online: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2024, 50 s. Diplomová práce. Vedoucí práce: Ing. Matěj Ištvánek

Obsah

Úvod	10
1 Teorie	11
1.1 MIR - Music information retrieval	11
1.1.1 Historie	11
1.1.2 Řetězec zpracování – pipeline	12
1.2 Parametrizace hudebních nahrávek	14
1.2.1 Reprezentace audio signálů	14
1.2.2 Časová oblast	14
1.2.3 Frekvenční oblast	15
1.2.4 DFT – Diskrétní Fourierova transformace	17
1.2.5 STFT – Short-time Fourier transform	19
1.2.6 Dynamika hlasitost a intenzita	21
1.2.7 Barva	22
1.3 Detekce nástupů a analýza tempa skladby	23
1.3.1 Využití energie signálu	24
1.3.2 Využití spektra signálu	26
1.3.3 Detekce periodicity	28
1.4 Analýza chroma vlastností	29
1.5 Segmentace	29
1.6 Klasifikace žánrů a nálady	29
1.7 Dostupná řešení	29
1.7.1 Librosa	29
1.7.2 Madmom	31
1.7.3 Aubio	31
1.7.4 Hodnocení extrakce informací	31
1.8 Systém Spectoda	31
1.8.1 Tvorba animace	32
2 Výsledky studentské práce	33
2.1 Návrh výsledného systému	33
2.1.1 Uživatelské rozhraní	33
2.1.2 Parametry hudební nahrávky	33
2.1.3 Systém pro generování animací	35
2.1.4 Databáze bloků animací	38
2.2 Výběr vhodných metod pro extrakci parametrů z hudební nahrávky .	38
2.2.1 Detekce dob a tempa	38

2.2.2	Analýza chromavektorů	40
2.2.3	Efektivní hodnota signálu	42
2.3	Realzace	43
2.3.1	Uživatelské rozhraní	43
2.3.2	Extrakce parametrů	43
2.3.3	Přiřazení barev	44
	Závěr	45
	Literatura	46
	Seznam symbolů a zkratek	50

Seznam obrázků

1.1	Řetězec procesů MIR [22]	13
1.2	Zobrazení časového průběhu signálu	15
1.3	Reprezentace tónu E zahráneného na basovou kytaru. a) Časová oblast b) Frekvenční spektrum	16
1.4	Časově spojitý signál a diskrétní signál	18
1.5	Signál o délce 1 s s počáteční frekvencí 10 Hz a koncovou frekvencí 30 Hz a) Původní signál b) Signál s okénkem od 0,2 s do 0,5 s c) Signál s okénkem od 0,35 s do 0,65 s d) Signál s okénkem od 0,5 s do 0,8 s [18]	20
1.6	Nahrávka piana a) Okamžitý výchylka nahrávky b) Frekvenční spek- trum nahrávky zobrazené pomocí spektrogramu	21
1.7	Tón A5 zahráný na klavír a) Okamžitá výchylka tónu b) ADSR obálka tónu	23
1.8	Detekce nástupů perikusivního zvuku a) Okamžitá výchylka nahrávky b) Lokální energie signálu $E_{xw}(n)$ c) Derivace lokální enegrie signálu s půlvlnným usměrněním $\Delta_E(n)$	25
1.9	Výpočet spektrálního toku pro nahrávku piana a) Okamžitá výchylka nahrávky b) Spektrogram nahrávky c) Spektrální tok bez komprese d) Spektrální tok s kompresí spektra $\gamma = 1$	27
1.10	Výpočet spektrálního toku z mel spektrogramu pro nahrávku piana a) Okamžitá výchylka nahrávky b) Mel spektrogram nahrávky c) Spektrální tok	28
2.1	Blokové schéma postupu uživatele webovou stránkou	33
2.2	Blokový diagram výběru datasetu.	35
2.3	Blokový diagram struktury rozhodovacího procesu	37
2.4	Blokové diagramy tříd <i>Dataset</i> , <i>AnimationBlock</i>	38
2.5	Porovnání metod detekce dob na úryvku skladby Oh-Darling!. a) Mel spektrogram b) Detekce dob pomocí Librosa c) Detekce dob pomocí Madmom d) Detekce dob pomocí Aubio	39
2.6	Porovnání přesnosti a času metod detekce dob na skladbách Oh- Darling!, Come Together a Here Comes The Sun. a) Čas výpočtu b) Cemgil skóré	40
2.7	Porovnání výpočtu chromavektorů na skladbě Oh-Darling! v časovém rozmezí 0 - 60 s.	41
2.8	Porovnání délky výpočtu chromavektorů	42
2.9	Zobrazení efektivní hodnoty skladby Belly dancer	43
2.10	Struktura třídy <i>ChromaFeatures</i>	43

Úvod

Práce se zabývá přípravou teoretických podkladů a průzkumem vhodných metod pro úspěšnou realizaci navazující diplomové práce v oblasti tvorby světelných animací na základě parametrisace hudební nahrávky. Získané znalosti jsou v semestrální práci využity pro výběr vhodných parametrů a návrh struktury výsledného systému. Tento systém bude generovat kód jež je následně převáděn na světelnou animaci pomocí řešení od společnosti Spectoda. Práce je rozdělena na 3 nasebe navazující cíle.

Prvním cílem je teoretický průzkum a vybudování dostatečné báze znalostí. Problematikou získávání informací z hudebních nahrávek se celosvětově zabývá vědní obor Music information retrieval zkráceně MIR. V této oblasti je zapotřebí prozkoumat existující vhodná řešení a zjistit jakými postupy jsou realizovány.

Druhým cílem je ze získaných znalostí navrhnut strukturu výsledného programu pro generování SpectodaCode, který je následně převáděn na světelné animace. Jsou prozkoumány a vybrány parametry hudebních děl potřebné pro realizaci systému. Při návrhu systému je nutné dbát aby byl realizovatelný, pro uživatele jednoduchý na použití a výpočetně nenáročný. Zároveň je kladen důraz na kvalitu generovaných animací.

Třetím cílem je praktická realizace získávání parametrů. Je nutné vyzkoušet dostupné metody pro analýzu vybraných parametrů, porovnat je mezi sebou na základě stanovených kritérií a určit jejich vhodnost pro použití ve struktuře výsledného systému. Z průzkumu vzniknou návrhy pro zlepšení a úpravu vybraných parametrů.

1 Teorie

V této kapitole je shrnuta teorie a sepsány potřebné informace pro realizaci a pochopení výstupů práce. Jsou popisovány zejména problematiky z oboru MIR popsaném v bodě 1.1. Metody pro extrakci parametrů z hudebních nahrávek a principy pro pochopení jak tyto metody fungují. Poslední bod 1.8 této kapitoly je zaměřen na systém Spectoda jeho možnosti tvorby interaktivních světelných animací a ovládání světelných zdrojů.

1.1 MIR - Music information retrieval

Music information retrieval je interdisciplinární vědní obor soustředící se na získávání informací z hudebních nahrávek. Jsou zde kombinovány znalosti mnoha oborů jako jsou muzikologie, psychoakustika, strojové učení, zpracování signálů a další.

Výstupy jeho výzkumu jsou využívány v populárních technologických aplikacích. Jednou z aplikací je personalizované doporučování hudebních skladeb, které se nachází v moderních streamovacích platformách. Další využití je v programech pro mixování hudby používaných diskžokeji k plynulejší práci díky analýze tempa a klíčových částí skladby. Tyto technologie se nachází v mnoha dalších aplikacích a s rozširováním digitálního audia jejich důležitost stále roste.

1.1.1 Historie

V tomto bodě je napsán souhrn historie MIR z knihy A New Companion to Digital Humanities [22]. Výzkumy v oblasti MIR se začínají objevovat na přelomu devatenáctého a dvacátého století s příchodem moderních statistických metod. Objevují se první pokusy o aplikování statistických metod na hudební partitura. Protože ještě nebyly natolik dostupné počítače jednalo se spíše o ruční práci s partiturami a tabulaturou. Z grafických notací se analyzovaly rysy skladeb a specifikovaly charakteristiky hudebního díla. S příchodem počítačů do výzkumných laboratoří se v letech 1960 až 1970 začalo více rozvíjet zpracování signálů a s tím související možnosti analýzy hudebních nahrávek pomocí počítačů. V těchto letech se poprvé začaly objevovat nyní známé termíny jako „Computational musicology“¹ a „Music information retrieval“. První oblast výzkumů se soustředila na analýzu tempa skladby. Z důvodu malé využitelnosti v komerčním sektoru nebyly výzkumu přiděleny patřičné priority. Tento útlum pokračoval až do roku 1990 kdy výzkumům v oblasti MIR pomohly dvě změny. První důležitou změnou byly rostoucí databáze digitální hudby, která

¹Computational musicology je vědní obor zahrnující veškeré výzkumy využívající výpočetní techniku pro práci s hudbou.

se stala lehce dostupná pro výzkumné týmy. Druhým bodem který přispěl k vývoji MIR byl nárůst výpočetního výkonu počítačů a nižší náklady s nimi spojené. Díky těmto změnám se stal výzkum dostupnější a jednodužší na realizaci [22].

V říjnu roku 2000 bylo uspořádáno první mezinárodní symposium zabývající se vývojem v oblasti MIR. Z této mezinárodní konference se stala tradice a vybudovala se kolem ní velká komunita nazývaná ISMIR². Každoročním vyvrcholením ISMIR je právě výše zmíněná konference, na které vědci z celého světa prezentují pokroky v oblasti výzkumu MIR. Zanedlouho naté v roce 2005 byl v rámci této konference představen model MIREX³ sloužící jako správa zásad pro hodnocení pokroků ve výzkumu MIR [6].

1.1.2 Řetězec zpracování – pipeline

V tomto bodě je popsán postup zpracování dat v aplikaci MIR. Jedná se o systém, kterým jsou data zpracovávána a určuje standardně využívaný řetězec jak při tvorbě algoritmů postupovat.

Vstupními daty se rozumí zejména hudební informace v digitální podobě. Tyto vstupní data se rozlišují do více typů. Mohou to být obrázky představující digitální formu zápisu hudby pomocí symbolů „not“ [22]. Například digitalizovaná partitura. Dalsím možným typem je „digitální hudba“. Jedná se o hudbu čistě v „digitálních notách“ představujících sadu příkazů. Například zápis v MIDI⁴. Nejrozšířenější formou vstupních dat jsou digitální hudební nahrávky představující audio signály.

Pre-processing – předzpracování signálu: Na začátku řetězce je zařazen blok předzpracování vstupních signálů. Tento blok se postará o připravení dat do podoby vhodné pro extrakci vlastností. Jedná se například o komprimaci komplexních vstupních signálů popsaných níže. Nebo je signál převáděn z časové do frekvenční oblasti. Více o technikách předzpracování je popsaano v bodě 1.2.

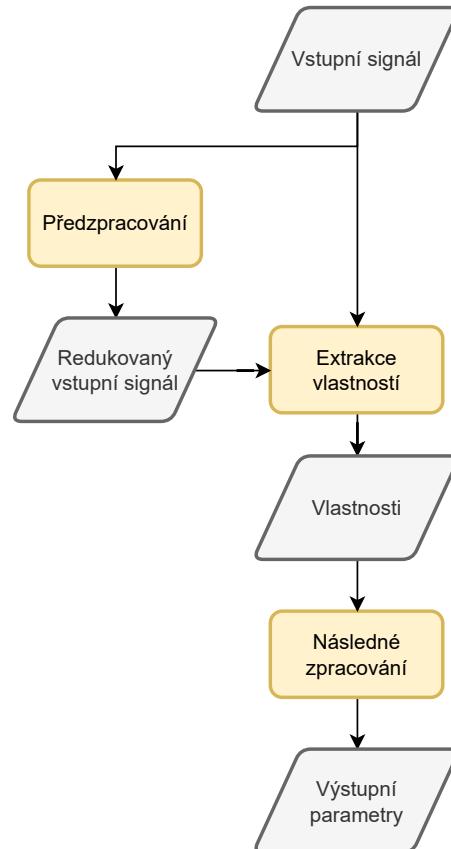
Feature extraction – extrakce vlastností signálu: Podle požadovaných vlastností pro extrakci je využíváno modelů popsaných v bodech 1.3, 1.6 a 1.4. S rostoucí popularitou strojového učení začaly při extrakci vlastností hudební nahrávky převládat kombinace hlubokých neuronových sítí. Tyto kombinace umožňují přesnější parametrizaci a menší chybovost.

²International Society of Music Information Retrieval - Mezinárodní sdružení pro MIR

³The Music Information Retrieval Evaluation eXchange - Komunitní rámec pro hodnocení pokroků výzkumu v oblasti MIR. Obhospodařovaný laboratoří International Music Information Retrieval Systems Evaluation Laboratory sídlící na University of Illinois [6].

⁴Musical Instrument Digital Interface - Volně dostupný hudební standart specifikující hardwerové a softwarové požadavky pro digitální přenos notace a komunikace mezi nástroji [30].

Post-processing – konečné zpracování: Posledním blokem v řetězci je konečné zpracování dat zajišťující jejich optimalizaci a převedení do požadované formy. V některých případech může ovlivnit přesnost zpracování.



Obr. 1.1: Řetězec procesů MIR [22]

Digitální hudební nahrávka se jako forma vstupních dat stala hlavním trendem výzkumu MIR. Je to způsobeno zejména dostupností velkých databází nahrávek ke kterým mají vědecké instituce přístup a nepotýkají se s problémy souvisejícími s autorskými právy [22].

Z důvodu velké komplexnosti vstupních dat se využívá několik technik komprimace signálů. Slučování vícekanálových nahrávek do mono signálu. Převzorkování signálu na nižší vzorkovací kmitočty, a rozložení na krátké překrývající se úseky, ze kterých mohou být nezávisle extrahovány jejich vlastnosti [12]. Výsledkem je kolekce paralelně složených sekvencí hodnot jednotlivých vlastností, které se následně zpracují na požadovaná výstupní data.

1.2 Parametrizace hudebních nahrávek

V této kapitole je popsán audio signál. Jak je reprezentován v oboru číslicového zpracování a základní principy práce s audiosignálem. V bodech 1.2.6 a 1.2.7 jsou popsány parametry získávané z audio signálu. Získané parametry slouží pro přesnější popis skladby.

1.2.1 Reprezentace audio signálů

Hudební dílo může být reprezentováno více formami. Jako tradiční médium pro její ukládání ještě před vznikem záznamu sloužily noty a další typy zápisů pomocí symbolů. Výsledné hudební dílo ale představuje mnohem více než počáteční notový zápis. Každý hudebník a hudební nástroj do skladby předává svou unikátnost. Při hře se noty začnou proměňovat v harmonické zvuky, hladké melodie a nástroje vzájemně rezonují. Každý z hudebníků do skladby přináší svou interpretaci. Jinak reagují na tempo, zvýrazňují odlišné noty a liší se jejich artikulace. Všechny tyto proměnné ve výsledku způsobují, že dílo není jen mechanické přehrání napsané partitury. Jeho součástí se stává unikátní přednes umělce a hudebního nástroje [18].

Z fyzikálního hlediska při interpretaci díla vznikají zvukové vlny šířící se vzduchem. Tyto vlny jsou reprezentovány kmítáním částic v pružném prostředí. V takovém prostředí jsou částice na sebe vázány a vytvářejí soustavu oscilátorů. Pokud dojde k vychýlení jedné částice ze své rovnovážné polohy, vlivem okolních částic dochází k působení pružných sil a vzniká její kmítání. Zároveň dochází k vzájemnému rozkmitání okolních částí a prostředím se začne šířit vlna. Jednotlivé částice kmitají pouze kolem své rovnovážné polohy. Nedochází tak k přenosu látky ale pouze energie a hybnosti [5]. Popsané kmítání je možné zachytit pomocí akustických měničů. Je získán analogový signál šířících se zvukových vln nazývaný jako audio signál. Pojmem audio je označován řetězec sloužící k záznamu, přenosu a reprodukci zvuků v mezích lidského slyšení. Avšak v audio signálu se už nenachází přesná reprezentace not a jejich paramterů jako jsou čas nástupu, tón, délka trvání, dynamika. Díky tomu je analýza hudebních signálů obtížným úkolem a je ovlivněna reprezentací interpreta, stavbou nástroje, akustikou prostoru a vnímáním posluchače. Zmíněnými problémy se zabývá samostatný vědní obor s názvem psychoakustika. Nejdůležitějšími parametry audio signálu které jsou podrobně popsány níže definujeme: frekvence, výška tónu, dynamika, intenzita, hlasitost a barva [18].

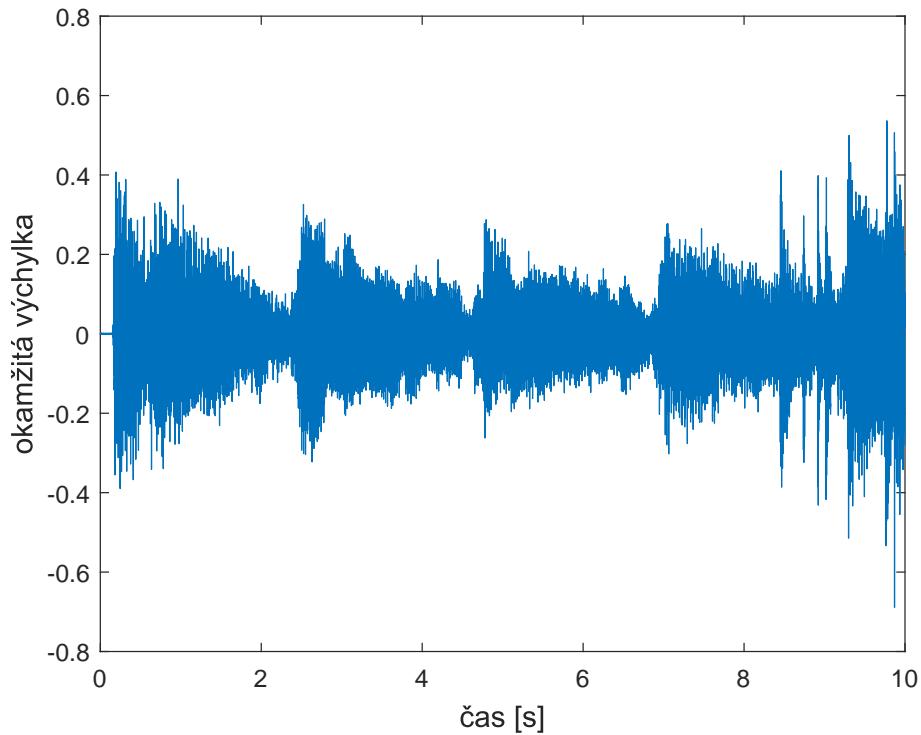
1.2.2 Časová oblast

Základní reprezentací audio signálu je tzv. zobrazení v **časové oblasti**. V časové oblasti představují číslicový signál vzorky. Jednotlivé vzorky udávají hodnotu signálu v

daném čase. Počet vzorků vztažených na jednotku času určuje vzorkovací frekvence signálu. Důležitým pravidlem pro vzorkování signálu je Shannonův-Nyquistův vzorkovací teorém

$$f_{vz} > 2f_{max} \quad (1.1)$$

kde f_{vz} je vzorkovací frekvence a f_{max} je maximální frekvence v audio signálu [2]. Pokud jednotlivé vzorky zobrazíme graficky získáme průběh signálu v čase viz obr.1.2. Tato reprezentace audio signálu poskytuje informace o průběhu amplitudy signálu. Využívá se například pro výpočet energie signálu popsaný v bodě 1.3.1.

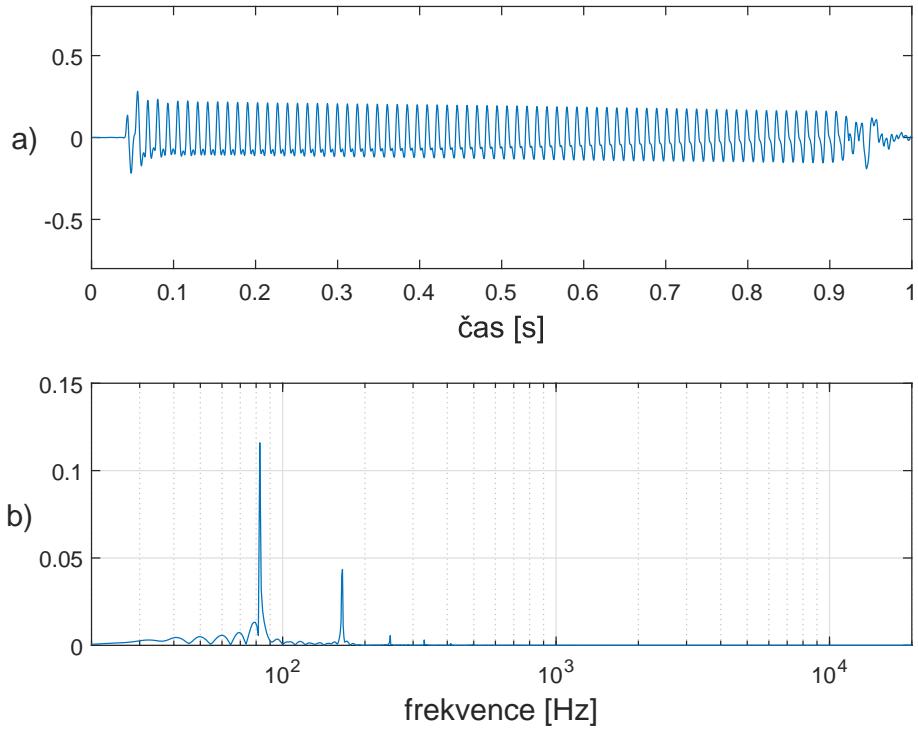


Obr. 1.2: Zobrazení časového průběhu signálu

1.2.3 Frekvenční oblast

Pro získání dalších informací o hudebním díle se využívá transformace signálu do frekvenční oblasti umožňující odlišné znázornění struktury signálu.

Ve frekvenční oblasti je signál reprezentován jeho frekvenčními složkami popsanými v komplexním tvaru. Spektrum představuje rozložení původní části signálu na jednotlivé frekvenční složky popsané funkcí sinus. Kde reálná složka obsahuje informaci o magnitudě „velikosti“ funkce sinus. Imaginární složka komplexního čísla pak udává počáteční fázi. V grafu jsou poté zobrazeny frekvenční složky se kterých se signál skládá viz obr. 1.3.



Obr. 1.3: Reprezentace tónu E zahráného na basovou kytaru. **a)** Časová oblast
b) Frekvenční spektrum

Jako názorný důvod proč je transformace do frekvenční oblasti přínosná je dán příklad. Na nástroj je zahrán tón, který je zaznamenán. V časové oblasti je možné určit délku tónu a jeho průběh podle ADSR obálky popsané v bodě 1.2.7. Pokud je ale potřeba zjistit výšku tónu a určit notu, tak se jedná o složitý proces. Díky transformaci do frekveční oblasti je patrná fundamentální frekvence tónu. Označována také první harmonická. Tato frekvence udává výšku tónu a je tak možné stanovit notu která byla zahrána.

Pro získání frekvenčního spektra signálu je třeba transformovat signál s časové oblasti. K tomu slouží několik úprav Fourierovy transformace podle vlastností vstupního signálu. Tyto metody jsou dále nazývány jako Fourierovy řady, Diskrétní časová Fourierova transformace a Diskrétní Fourierova transformace. V případě audio signálu se využívá zejména Diskrétní Fourierovy transformace popsané v bodě 1.2.4.

Fourierovy transformace zkráceně definována jako transformace převádějící signál mezi časovou a frekvenční oblastní pomocí harmonických signálů jež popisují funkce sinus a cosinus [2]. Funkce sinus a cosinus představují komplexní exponenciály.

$$e^{i\alpha t} = \cos(\alpha t) + i \sin(\alpha t) \quad (1.2)$$

Fourierova transformace pro **spojitý neperiodický signál** je pak zapsána jako

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt \quad (1.3)$$

kde $\omega = 2\pi f$ a udává uhlovou frekvenci. Magnituda $|X(f)|$ je potom funkcí sudou [23].

Pro signál který je **spojitý a periodický** se definují Fourierovy řady a integrální funkce je počítaná pouze pro jednu periodu signálu

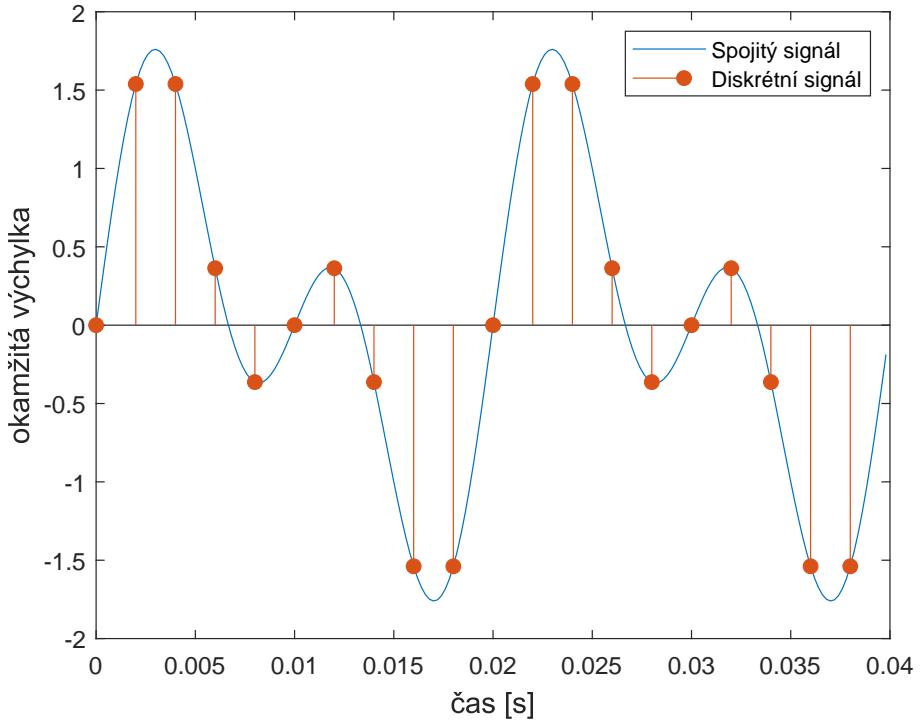
$$c[f_k] = \frac{1}{T_0} \int_0^{T_0} x(t) e^{-ik\omega_0 t} dt \quad (1.4)$$

kde $f_k = k \times f_0$ a $k \in (\mathbb{Z}; 0, \pm 1, \pm 2, \dots)$. Vypočítané spektrum je diskrétní a ne-periodické [23]. Pokud je vstupní signál diskrétní hovoříme o Diskrétní Fourierově transformaci která je více popsána v následujícím bodě 1.2.4.

Po transformaci signálu do frekvenčního spektra jsou data signálu v komplexním tvaru a jejich magnituda $|X(f)|$ je funkce sudá, tím pádem symetrická kolem nuly a fáze $\varphi_x(f)$ je funkci lichou čili je středově symetrická. Pro analýzu audio signálů se využívá kladná část spektra.

1.2.4 DFT – Diskrétní Fourierova transformace

Pokud jsou signály zpracovávány pomocí výpočetních procesorů, tak může být uložen pouze omezený počet hodnot signálu. To znamená, že analogový signál spojitý v čase musí být převeden na signál digitální tvz. signál diskrétní, který je není spojitý v čase. Diskrétní signál je potom vhodný pro číslicové zpracování. Proto jsou pospané algoritmy DFT přizpůsobené právě pro zpracování diskrétních signálů nespojitéch v čase.



Obr. 1.4: Časově spojitý signál a diskrétní signál

Opět jsou dány odlišné definice pro signál diskrétní neperiodický a diskrétní periodický. Protože se jedná o signál diskrétní, tak zde odpadají integrální funkce. Pokud se jedná o signál **diskrétní neperiodický** jeho výsledné spektrum bude spojité a hovoříme o Fourierově transformaci diskrétní v čase. Matematicky je zapsána v následujícím tvaru

$$X(f) = \sum_{n=-\infty}^{\infty} x[n]e^{-i\Omega n} \quad (1.5)$$

kde

$$\Omega = 2\pi(f/f_s) \quad (1.6)$$

a f_s je vzorkovací frekvence signálu.

Protože v praxi signál není nikdy nekonečně dlouhý, tak je možné jej poskládat za sebe a vytvořit tak signál periodický. Pro periodické signály je výpočet DFT zapsán ve tvaru

$$X[f_k] = \sum_{n=1}^{N_0} x[n]e^{-i\Omega_k n} \quad (1.7)$$

kde

$$\Omega_k = 2\pi \frac{f_k}{f_s} \quad (1.8)$$

$$f_k = \frac{k f_s}{N_0} \quad (1.9)$$

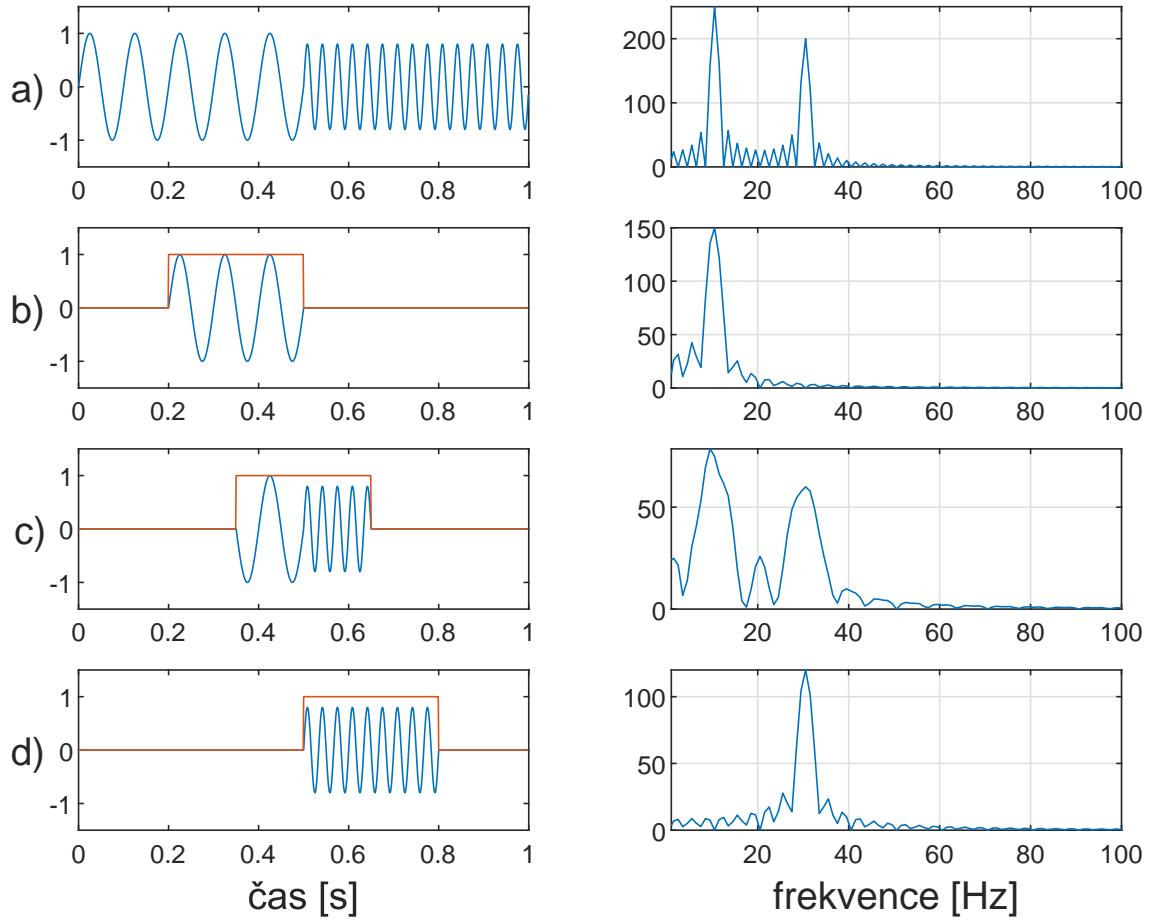
a $k \in (\mathbb{Z}; 0, N_0 - 1)$. N_0 udává počet vzorků v jedné periodě signálu. Hustota spektra K v takovém případě odpovídá $K = N_0$.

Ze strany výpočetní náročnosti je takto definováný algoritmus neefektivní a výpočetně náročný. Pro výpočet DFT je zapotřebí velkého množství operací. Složitost algoritmu je pak zapsána jako $\mathcal{O}(N^2)$. Proto pokud počet vzorků N dosahuje většího množství je ve spoustě případů tento algoritmus příliš pomalý a neefektivní pro praktické využití.

Počet potřebných operací může být výrazně redukován. Na vývoj efektivního řešení výpočtu DFT se zasloužil Carl Friedrich Gauss a Joseph Fourier zhruba před dvěma sty lety. Tento algoritmus nazýváme Rychlou Fourierovou transformací zkráceně FFT. Počet operací pro výpočet takového algoritmu byl snížen na $\mathcal{O}(N \log_2 N)$ [18]. Například při použití vzorků $N = 2^{10} = 1024$. FFT vyžaduje $N \log_2 N = 10240$ operací namísto $N^2 = 1048576$ operací při použití DFT. Jak je vidět snížení výpočetní náročnosti je velké a exponenciálně roste s větším počtem vzorků N . Vynález FFT změnil odvětví zpracování signálů a je dnes využíván v miliardách telekomunikačních zařízeních. Stejně tak i ve zpracování a analýze zvukových signálů hraje důležitou roli [18].

1.2.5 STFT – Short-time Fourier transform

V roce 1946 Dennis Gabor představil STFT jako možnost zařazení frekvenčních složek ke konkrétnímu času signálu [26]. Fourierova transformace umožňovala převod signálu z časové oblasti do frekvenční ale nebylo zřejmé v jakém časovém úseku signálu se získané frekvenční složky nachází. Hlavní myšlenkou STFT je, že namísto analyzování celého signálu je analyzována pouze jeho malá část. Za tímto účelem je definovaná tzv. okénková funkce, která je nenulové pouze v malé části signálu. Analyzovaný signál je následně vynásoben vzniklou okénkovou funkcí a díky tomu vzniká malá nenulová část signálu dle okénkové funkce viz obr. 1.5. Chceme-li analyzovat signál v různých časech je tato funkce po signálu posouvána a následně se počítá DFT pro každý výsledný okénkový signál.



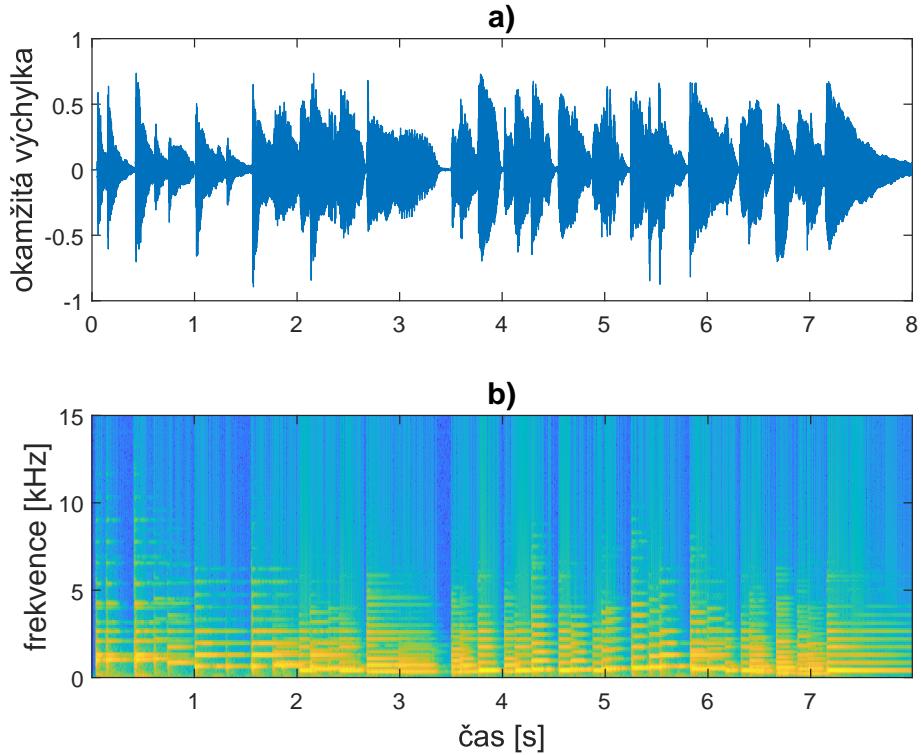
Obr. 1.5: Signál o délce 1 s s počáteční frekvencí 10 Hz a koncovou frekvencí 30 Hz
a) Původní signál **b)** Signál s okénkem od 0,2 s do 0,5 s **c)** Signál s okénkem od 0,35 s do 0,65 s **d)** Signál s okénkem od 0,5 s do 0,8 s [18]

Na obr. 1.5 je graficky znázorněna myšlenka STFT, která ukazuje princip určování frekvenčních složek v čase a jejich výhody. Signál je násoben obdélníkovou okénkovou funkcí ve třech místech. Tyto tři vzniklé signály jsou následně na sebe nezávisle transformovány do frekvenční oblasti. Z výsledků Fourierovy transformace lze vidět, že každá z těchto částí má jiné frekvenční spektrum. Pokud by bylo zapotřebí například určit přesný přechod mezi dvěma frekvencemi nacházejícími se v signálu. Lze zpřesnit časové měřítko analýzy pomocí délky okénka. Tím ale dochází ke zmenšení přesnosti ve frekvenční oblasti.

Na výsledku přesnosti analýzy pomocí STFT závisí také tvar použité okénkové funkce. V obr. 1.5 je použito obdélníkového okénka které díky svým ostrým hranám zkresluje výsledek o nechtěné frekvenční složky. Existuje více tvarů okénkových funkcí pro odstranění nežádoucích složek. Například to jsou Kaise, Chebyshev, Hann, Haming a další [4].

Pokud jsou analyzované data skrze okénka funkce STFT poskládány zpět za

sebe, tak představují matici průběhu frekvenčního spektra signálu v čase. Takové zobrazení se nazývá **spektrogram** a v případě 2D zobrazení se skládá z časové a frekvenční osy. Magnituda frekvencí je pak zobrazena barevnou škálou viz obrázek 1.6.



Obr. 1.6: Nahrávka piana **a)** Okamžitý výchylka nahrávky **b)** Frekvenční spektrum nahrávky zobrazené pomocí spektrogramu

1.2.6 Dynamika hlasitost a intenzita

V češtině se pojem hlasitost využívá pro reprezentaci subjektivního vnímání akustického tlaku definovanou například jednotkou phon⁵. Stejně tak je hlasitost využívána, hovoří li se o měřené hlasitosti vyjádřené například hladinou intenzity zvuku popsanou níže nebo efektivní hodnotou signálu. Z důvodu lepší srozumitelnosti jsou dále využívána anglické pojmy „volume“ a „loudness“.

Dynamika popisuje průběh hlasitosti „volume“ interpretovaného hudebního díla. Udává jeden z faktorů jak lze například odlišit stejnou skladbu zahranou různými muzikanty. Interpretací skladby umělec vytváří dynamiku přednášeného díla [18]. V notovém zápisu je dynamika neboli hlasitost přednesu popsána symboly jako jsou například pianissimo „pp“, piano „p“, forte „f“ a další.

⁵Phon - logaritmická jednotka vyjadřující individuální vnímání hlasitosti. Vnímání hlasitosti lidského ucha je závislé na křivce prahu slyšení a může se lišit pro každý tón [29].

V audio signálu je dynamika brána jako hlasitost „loudness“ Jedná se o změny apmlitudy signálu nebo jeho efektivní hodnoty RMS⁶ v čase.

Při měření hlasitosti „loudness“ v akustickém prostoru je pak využíváno pojmu **intenzita** zvuku a **akustický výkon**. Akustický výkon je definován jako množství energie vyzářené akustickým vysílačem ve vzduchu za jednotku času [8]. Jednotkou je W . A intenzita zvuku je pak definována jako množství energie, které projde jednotkovou plochou kolmou na směr šíření na jednotku času. Jednotkou je Wm^{-2} [31].

Z pohledu vnímání hlasitosti lidským uchem je rozsah vnímané intenzity zvuku v řádech bilionů. Práh slyšení činí $10^{-12} Wm^{-2}$ a práh bolesti je $10 Wm^{-2}$. Pro zmenšení tak velkého rádu je definována hladina intenzity zvuku v decibelech dB . Kde vztažnou hodnotou je práh slyšení $I_0 = 10^{-12} Wm^{-2}$. Hladina intenzity se vypočítá dle rovnice 1.10.

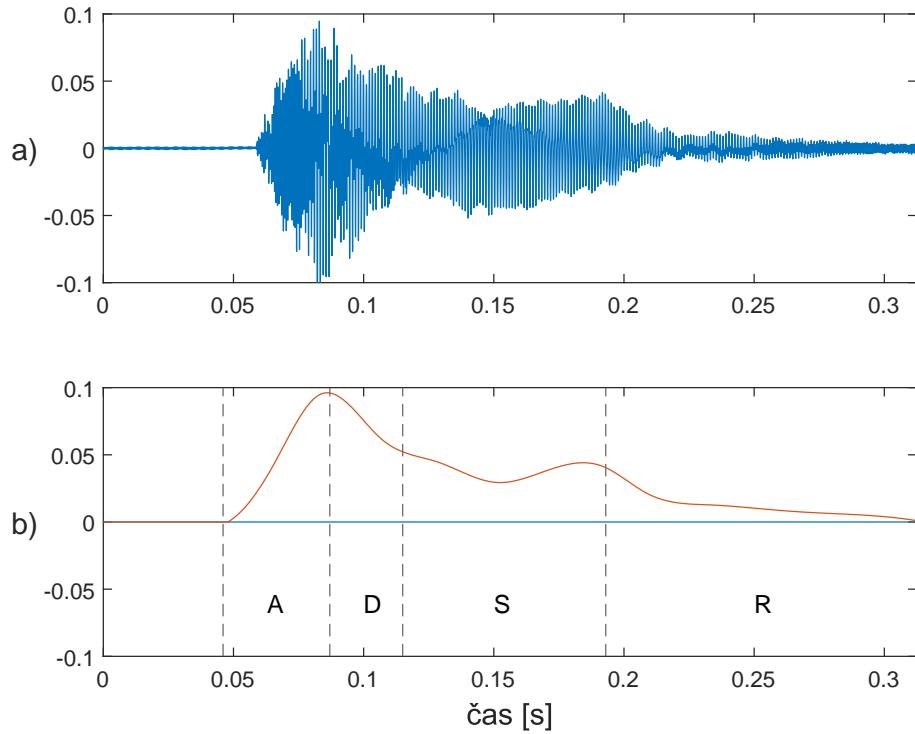
$$L_I = 10 \log\left(\frac{I}{I_0}\right) \quad (1.10)$$

1.2.7 Barva

V hudebním vyjádření se za slovem barva skrývá komplexní sdružení atributů. Jedná se jak o psychologický, tak hudební problém, který je vnímán individuálně[15]. Zjednodušeně se barva definuje jako vlastnosti, díky kterým je možné rozpoznat tón o stejné výšce a hlasitosti zahrnutý na dva různé nástroje[19]. Díky barvě je posluchač schopen rozpoznávat odlišné zvuky nástrojů a typů interpretace. Z důvodu špatné kategorizace barvy fyzikálními veličinami je většinou interpretována přídavnými jmény. Například je barva popisována jako jasná, temná, ostrá, čistá, teplá, pestrá, nazální, tonální a další.

Jedním z možných nástrojů pro analýzu barvy tónu je tzv. obálka tónu „signálu“. Obálku určuje okamžitá výchylka signálu v čase viz obr. 1.7 Je rozdělena na 4 fáze popsané z knihy Fundamentals of Music Processing [18]. **Attack „náběh“** určující začátek tónu. Například úder paličkou na blánu bubnu. V této fázi se nachází více ruchových složek z daného úderu a má velkou dynamiku. Následuje fáze s názvem **Decay „útlum“**. Po hlasitém úderu okamžitá výchylka signálu klesá a začíná převládat tonální složka. Decay udává dobu za kterou se signál z jeho maxima sníží na hodnotu sustain. **Sustain „podržení“** je fáze ve které je zřetelný tón a stálá hlasitost. Rezonující blána bubnu. Poslední fází je **Release „uvolnění“** při kterém dochází k poklesu hlasitosti zdroje zvuku až k uplnému utlumení. Například přiložení tlumítka na rezonující strunu.

⁶RMS - udává statistickou hodnotu z měření velikosti veličin. Je využívána u periodických veličin[3].



Obr. 1.7: Tón A5 zahráný na klavír a) Okamžitá výchylka tónu b) ADSR obálka tónu

Další informace o barvě signálu se nacházejí v jeho frekvenčním spektru. Tón zahráný na hudební nástroj má svou fundamentální „nosnou“ frekvenci nazývanou první harmonická udávající jeho výšku. Dle konstrukce nástroje se v signálu objevují násobky nosné frekvence. Tyto násobky představují vyšší harmonické složky tónu. Počet a amplituda vyšších harmonických složek má vliv na výslednou barvu tónu a je to hlavní důvod proč je lidské ucho schopné rozeznat stejný tón znějící na různé nástroje.

1.3 Detekce nástupů a analýza tempa skladby

S vývojem nových technologií se mění i přístupy využívané v MIR pro analýzu tempa skladby. V této kapitole je popsán postupný vývoj algoritmů. Od nejjednodušších přístupů po komplexní řešení využívající moderní metody strojového učení.

Detekce nástupů lze chápat jako detekci začátků not či dalších hudebních událostí, které se vyskytnou v průběhu skladby. Výskyt takových událostí je provázen zvýšením energie signálu zaznamenané ve fázi nástupu dle ADSR obálky popsané v bodě 1.2.7. Typickým znakem pro fázi nástupu je rychlý nárůst obálky amplitudy signálu. V této fázi se při vytváření tónu vyskytují **Tranzienty**. Tranzienty je možné

pozorovat zejména u neperkusivních nástrojů například u dechových nebo smyčcových. Jsou představovány nakmitávajícími a dokmitávajícími pochody. Jedná se o zvuk netónové podoby připomínající hluk s výraznými frekvenčními změnami. U smyčcových nástrojů takový zvuk může být ze začátku způsoben drhnutím smyčce o strunu do chvíle než se ustálí její kmitání. Délka tranzientů pak může dosahovat od jednotek milisekund až po stovky milisekund v závislosti na typu nástroje a technice hry [27]. Například v případě piana tranzient odpovídá počáteční fázi kdy byla zmáčknuta klávesa. Na základě zmáčknutí klávesy dochází ke zvednutí tlumítka, kladívko udeří do struny, struna začíná vibrovat a vibrace se přenáší do těla piana. V této fázi začíná tělo rezonovat a konečně dochází k ustálení tónové složky. Při úderu kladívka dochází k velkému přenosu energie patrném na obálce tónu. Díky tomu je lehké určit začátek tónu podle nárůstu amplitudy obálky [18].

U některých nástrojů je však energie přenášená po celou dobu znění tónu konstantně a dochází k pomalému jemnému náběhu tónu například u určitých technik hry na smyčcové nástroje nebo u dechových nástrojů. Fáze náběhu je pak pomalá a dlouhá a plynule přechází do fáze podržení. Pro tyto jemné zvuky se stává obtížné určit skutečnou pozici začátku noty.

Náročnost detekce nástupů se zvyšuje v případě, že nehráje pouze jeden nástroj. Jedná se například o polyfoní skladbu. Polyfoní skladbou rozumíme skladbu tvořenou více hlasy. Zároveň nemají určenou roli vedoucího a doprovodného hlasu [20]. Takové uspořádání hlasů může vést k překrývání a nástupy mohou být maskovány. Díky jevu maskování je obtížné zaznamenat změny energie signálu. V tomto případě přichází potřeba po komplexnějších metodách detekce nástupů [18]. Například pohled na krátkodobé změny ve spektru signálu pomocí využití STFT popsané v bodě 1.2.5.

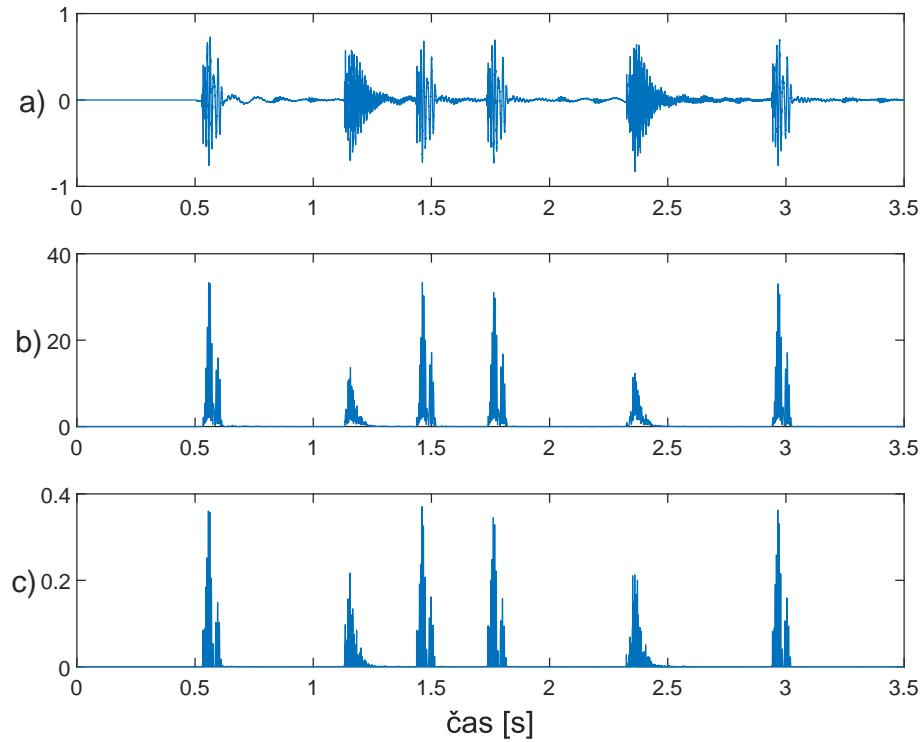
1.3.1 Využití energie signálu

Jak již bylo zmíněno při hraní dochází k přenosu energie od hráče na hudební nástroj. Tento přenos je ve značné míře provázen rychlým nárůstem amplitudy ve fázi náběhu na začátku tónu. Například při úderu kladívka piana na strunu nebo úderu paličky na blánu bubnu. Ná základě tohoto jevu je možnost detektovat nástup tónu pomocí funkce pro výpočet energie signálu v daném místě. Náhlé změny signálu v takto definované funkci ukazují potencionální místa nástupů. Matematicky pak funkci popisujeme. Stejně jako u STFT popsané v bodě 1.2.5 je definována okénková funkce $w(n)$ ve tvaru zvonu „bell-shaped function“ která je posouvaná po diskrétním signálu $x(n)$. Okénková funkce je symetrická podle počátku a potom platí že $m \in [-M : M]$

a $M \in \mathbb{N}$. Funkce lokální energie signálu je pak zapsána

$$E_{wx}(n) = \sum_{m=-M}^M |s(n+m)w(m)|^2 \quad (1.11)$$

pro $n \in \mathbb{Z}$ [18].



Obr. 1.8: Detekce nástupů perkusivního zvuku **a)** Okamžitá výchylka nahrávky **b)** Lokální energie signálu $E_{wx}(n)$ **c)** Derivace lokální energie signálu s půlvlnným usměrněním $\Delta_E(n)$

Pro názornější zobrazení se následně vypočítá derivace funkce lokální energie. V případě diskrétního signálu se derivace realizuje jako rozdíl dvou po sobě jdoucích vzorků. Protože pro detekci nástupů je důležitá zejména pozitivní změna energie nikoliv její pokles jsou ponechány pouze pozitivní rozdíly a negativní jsou zapsány jako nulové. V anglické literatuře je tento proces známý také jako půlvlnné usměrnění „half-wave rectification“. Vzorce jsou zapsány následovně

$$r = E_{wx}(n+1) - E_{wx}(n) \quad (1.12)$$

$$\Delta_E(n) = \frac{r + |r|}{2} = \begin{cases} r, & \text{if } r \geq 0 \\ 0, & \text{if } r < 0 \end{cases} \quad (1.13)$$

kde $r \in \mathbb{R}$ a $n \in \mathbb{Z}$. Na obr. 1.8 lze vidět zobrazení takto vypočítané lokální energie signálu a její derivace. Analyzovaný signál se skládá z několika perkusivních úderů. Pro výpočet bylo použito okno typu „bell-shaped function“ o velikosti 201 vzorků.

1.3.2 Využití spektra signálu

Díky rozložení signálu na jeho spektrum pomocí Fourierovy transformace popsané v bodě 1.2.3 je možné lépe rozpoznat strukturu nahrávky a zmírnit efekt maskování, který nastává při metodách založených na energii signálu popsaných v bodě 1.3.1. V polyfonní hudbě mohou interpretace o nižší hlasitosti být maskovány hlasitějšími projevy. Kdy například jeden nástroj v tónové fázi podržení ADSR obálky, popsané v bodě 1.2.7, může být hlasitější než fáza náběhu druhého nástroje. Takový nástup je pak maskován a je obtížná jeho detekce. V případě maskování některých z nástrojů v časové oblasti je možné spektrum signálu zaměřit na frekvenční oblast spektra ve které se maskovaný nástroj nachází. Díky tomu je možné nástup tónu takového nástroje detektovat pomocí spektra signálu. Každý hudební nástroj obsahuje jiné frekvenční spektrum [18]. Díky tomu se různé nástroje nachází na odlišných místech spektra. Jak spektrum daného nástroje vypadá určuje také barvu popsanou v bodě 1.2.7. S takovým jevem je důležité počítat při detekci nástupů založené na spektru signálu.

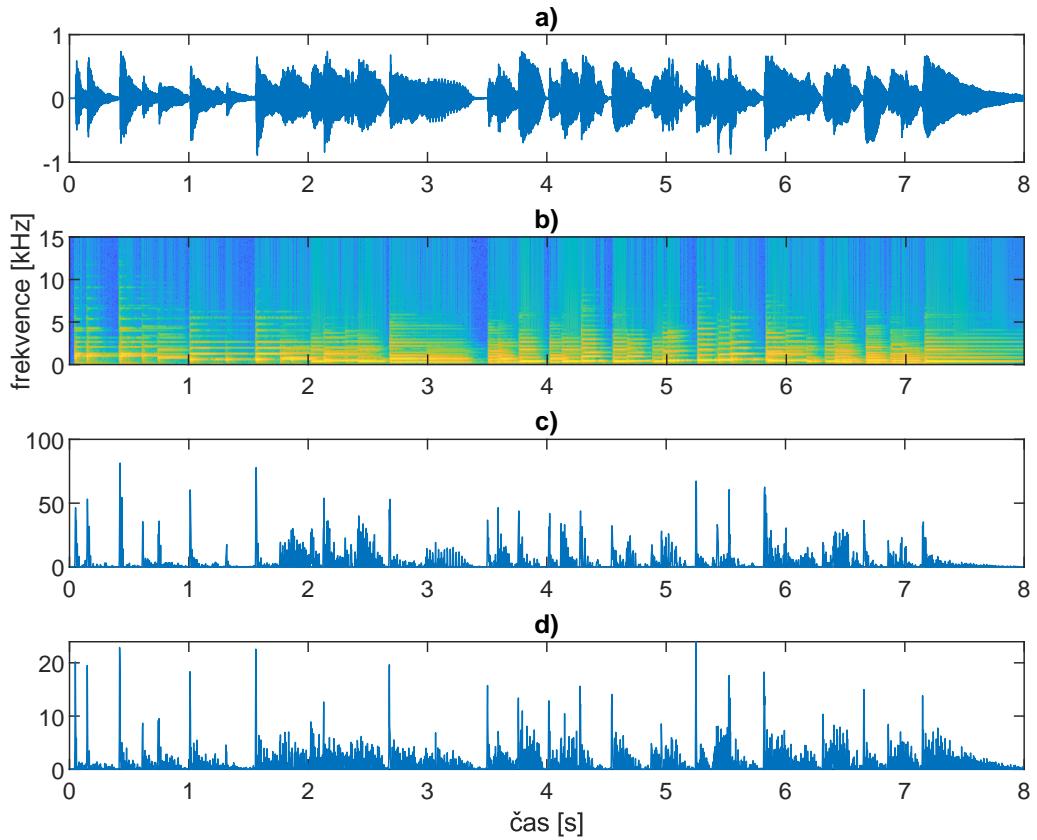
Jedním ze spůsobů analýzy nástupů pomocí spektra je detektovat změny ve spektru v průběhu času. Při zobrazení časového průběhu spektra nazývaného spektrogram popsaný v bodě 1.2.5. Jsou za sebou v čase poskládány vektory nesoucí informaci o spektru. Pokud jsou počítány rozdíly mezi dvěma po sobě jdoucimi vektory spektrogramu jsou získány informace o změně spektra v čase. Protože tranzienty se při fázi náběhu skládají z části z ruchové složky rozléhající se přes velkou část slyšitelného frekvenčního spektra je tak možné detektovat nástupy. Popsané metodě porovnávání vektorů se říká **spektrální tok** [18]. Pro výpočet spektrálního toku existuje více druhů přístupů lišících se předszpracováním dat a konečným zpracováním výsledků. Níže jsou popsány dvě metody výpočtu ze spektrogramu a následně z mel spektrogramu.

Výpočet spektrálního toku se provede derivací vstupního diskrétního signálu který zde představuje vektory spektrálních složek. Výpočet probíhá pro každou spektrální složku a výsledek je sečten. Popsáno rovnici

$$r(n) = \sum_{k=0}^K |X|(n+1, k) - |X|(n, k) \quad (1.14)$$

$$\Delta S_t(n) = \frac{r(n) + |r(n)|}{2} = \begin{cases} r, & \text{if } r \geq 0 \\ 0, & \text{if } r < 0 \end{cases} \quad (1.15)$$

kde $n \in \mathbb{Z}$ a K je počet spektrálních složek v jednom vektoru.



Obr. 1.9: Výpočet spektrálního toku pro nahrávku piana **a)** Okamžitá výchylka nahrávky **b)** Spektrogram nahrávky **c)** Spektrální tok bez komprese **d)** Spektrální tok s kompresí spektra $\gamma = 1$

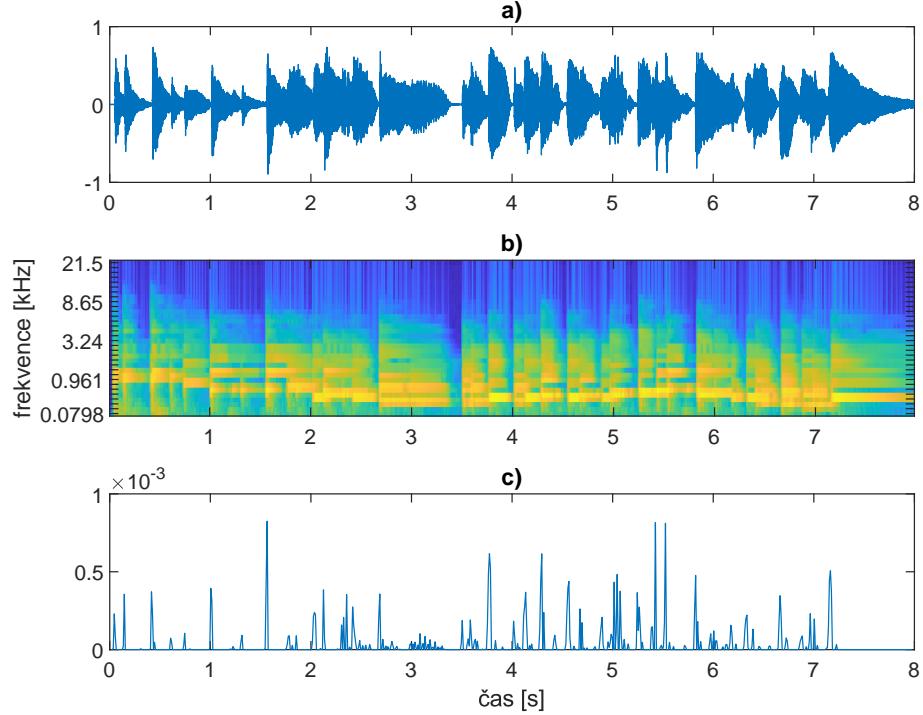
Pro dosažení přesnějších výsledků je možnost komprese spektrogramu. Funkce pro logaritmickou kompresi singálu je zapsána jako

$$X_c = \log(1 + \gamma|X|) \quad (1.16)$$

kde pro správnou kompresi je potřeba aby $\gamma \geq 1$. Míra komprese je určena velikostí γ . Nižší hodnoty jsou vhodné pro lepší detekci perkusivních zvuků. Při vyšších hodnotách γ dochází k větší komprezi spektrogramu a jsou zřetelnější zvuky o nižší intenzitě. Při velké kompresi dochází také k zvýraznění ruchových složek signálu [18].

Například v článku z ISMIR 2021 [13] je spektrální tok počítán s mel spektrogramu popsaného níže. Protože lidské ucho vnímá výšku frekvence logaritmicky je jednodužší rozoznat rozdíl mezi frekvencemi u nižší části spektra než u vyšších frekvencí. Například rozdíl mezi 500 Hz a 1 kHz je pro lidské ucho dobře detekovatelný a zřetelný. Odlišné rozlišovací schopnosti výšky tónu v různých frekvenčních pásmech jsou způsobeny biologickým rozvržením vnitřního ucha. Proto pro lepší rozlišování na základě lidského vnímání zvuku vznikla Melova stupnice. **Melova stuopnice**

je vjemová škála výšek, které posluchač posoudil jako od sebe stejně vzdálené. Referenčním bodem je pak definováno 1000 mel který odpovídá frekvenci 1000 Hz [25]. Výpočet spektrálního toku je následně stejný jak pří výpočtu ze spektrogramu. Výslednou křivku je možné vidět na obr. 1.10.



Obr. 1.10: Výpočet spektrálního toku z mel spektrogramu pro nahrávku piano **a)** Okamžitá výchylka nahrávky **b)** Mel spektrogram nahrávky **c)** Spektrální tok

1.3.3 Detekce periodicity

Detekce periodicity je nezbytnou součástí procesu při detekci dob skladby. Důležitou technikou pro detekci periodicity je autokorelační funkce. Autokorelační funkce slouží pro detekci periodických vzorů v datech. Tato metoda dokáže detektovat tempo, ale neumožňuje určit o jakou dobu se jedná. Základní použití je výpočet autokorelační funkce s energií signálu z bodu 1.3.1. Matematicky je pak popsána dle rovnice 1.17 [9].

$$r(n, i) = \sum_{u=-(T_w/2)+1}^{T_w/2} E(n+u)E(n+u-i) \quad (1.17)$$

Autokorelační funkce zde přináší několik výhod. Jednoduchost a efektivita. Je to jednoduchá metoda pro zjistění periodicity v signálech a vyžaduje menší množství paměti než jiné metody. Autokorelace může odhalit ne zcela zřejmé periodické vzory,

které nemusí být při poslechu nebo pozorování okamžitě zřejmé. Zároveň je robustní vůči šumu. Při analýze také nedochází ke ztrátě informace [11].

Autokorelace má také své nevýhody. Není vhodná pro analýzu v reláném čase protože potřebuje dostatečné množství dat a je potřebné aby v analyzovaném datovém okně byla obsažena celá perioda. V opačném případě nedojdte k jejímu nalezení. Autokorelace není schopná pracovat s fází signálu. Pro rozpoznání fáze je zapotřebí dalších metod. Nekonstantní tempo v analyzovaném signálu představuje pro autokorelační analýzu problém [11].

Při detekci periodicity je možné využít také banky rezonátorů hřebenového filtru s konstantním poločasem [21] nebo fázově blokující rezonátory [10].

1.4 Analýza chroma vlastností

1.5 Segmentace

1.6 Klasifikace žánrů a nálady

1.7 Dostupná řešení

Díky vědecké základně v oblasti music information retrieval vznikají volně šířitelné knihovny umožňující snadné použití v programovacím jazyce python. Tyto knihovny jsou publikovány za účelem usnadnění práce v oblasti MIR. Cílem některých knihoven je usnadnit přechod výzkumných týmů do programového jazyka Python a zakomponovat moderní praktiky softwarového vývoje. Díky tomu se staly tyto metody dostupné širší komunitě vědců [16]. V této kapitole je popsáno několik rozšířených knihoven poskytující metody z oblasti MIR.

1.7.1 Librosa

Knihovna Librosa vznikla v roce 2015 na základě potřeb vědecké komunity pro snadné použití metod z oblasti MIR v jazyce python. Publikována a popsána byla článkem Audio and Music Signal Analysis in Python [16] na konferenci SciPy 2015. Jedná se o otevřenou knihovnu jejíž další vývoj probíhá zapomoci vědecké komunity na GitHubu kde je kladen důraz na kompatibilitu, řádnou dokumentaci obsahující popisy funkcí s příkladovými kódy, a testování funkčnosti. Knihovna obsahuje pokročilé metody pro detekci dob.

Librosa obsahuje funkce pro extrakci audio vlastností. Pro tuto práci jsou důležitými funkcemi zejména detekce tempa a dob skladby, rozpoznání začátků tónů.

Následně jsou v knihovně obsaženy nástroje pro vizualizaci získaných vlastností. Například zobrazení spektrograma či grafické zobrazení detekovaného tempa a dob. Funkce pro detekci tempa a dob skladby pomocí dynamického programování je v knihovně zapsána:

```
tempo, beats = librosa.beat.beat_track(y=y, sr=sr)
```

Kde y je analyzovaný signál a sr je vzorkovací frekvence signálu. Detekce dob je ve funkci *beat_track()* realizována pomocí dynamického programování. Proces je rozdělen na 3 etapy popsané Danielem P.W. Ellis ve článku Beat Tracking by Dynamic Programming [7]. Jedná se o tyto kroky:

1. Výpočet obálky síly nástupů
2. Globální odhad tempa
3. Porovnání obálky síly nástupů s globálním tempem

Výpočet obálky síly nástupů: Při výpočtu obálky síly nástupů je vstupní signál převzorkován na 8 kHz. Poté je vypočítána STFT s délkou okna 32 ms a 4 ms mezerou mezi okny. Vypočítaný spektrogram je převeden na mel spektrogram o 40 mel pásmech. Mel stupnice je využívána ve snaze vyrovnat důležitost každého frekvenčního pásma v návaznosti na logaritmickém vnímání frekvencí lidským uchem. Mel spektrogram je převeden na dB a následně jsou počítány diferenční rovnice pro každé mel pásmo. Principem půlvlnného usměrnění jsou výsledné negativní hotnoty změněny na nulové hodnoty kladné jsou v daném čase sečteny napříč všemi pásmi. Výsledný signál je filtrován horní propustí s mezní frekvencí 0,4 Hz a vyhlazen pomocí konvoluce s Gaussovou křivkou s šírkou okolo 20 ms. Pomocí výše popsaného procesu je získána jednorozměrná obálka síly nástupů v závislosti na čase reprezentující proporcionální nárůst energie přibližně sečtené ve mel pásmech.

Globální odhad tempa: Globální odhad tempa je určen z obálky síly signálu $O(t)$ získané v předchozím kroku. Je počítána autokorelace mezi původní obálkou signálu $O(t)$ a jejími zpožděnými verzemi. Pro zpoždění ve kterých se potká velké množství vrcholů obálky nastává velká korelace. Tato korelace může nastat v celočíselných násobcích dané hodnoty zpoždění. Díky tomuto jevu je těžké určit který čas zpozdění představuje správné tempo skladby. Avšak lidmi vnímané tempo skladeb má sklon být kolem 120 BPM, jak bylo zjištěno ve výzkumu popsaném v článku Ambiguity in Tempo Perception: What Draws Listeners to Different Metrical Levels? [17]. Proto je na výslednou autokorelacii aplikováno váhové okno snižující hodnotu autokorelace se vzdáleností od 120 BPM. Čas ve kterém autokorelace dosahuje největší hodnoty je čas hledaného tempa skladby.

Porovnání obálky síly nástupů s globálním tempem: Matematicky je tento

krok fomulován rovnicí 1.18 kde $\{t_i\}$ je pole nalezených dob, $O(t)$ je obálka síly signálu, α je váhový koeficient určující důležitost mezi obálkou síly signálu a detekovaným tempem. Funkce $F(\Delta t, \tau_p)$ zapsaná rovnicí 1.19 měří konzistenci mezi skutečným časem od sebe vzdálených dob Δt a ideálním časem mezi dobami τ_p určeným cílovým tempem.

$$C(\{t_i\}) = \sum_{i=1}^N O(t_i) + \alpha \sum_{i=2}^N F(t_i - t_{i-1}, \tau_p) \quad (1.18)$$

$$F(\Delta t, \tau) = -(\log \frac{\Delta t}{\tau})^2 \quad (1.19)$$

1.7.2 Madmom

1.7.3 Aubio

1.7.4 Hodnocení extrakce informací

1.8 Systém Spectoda

Systém Spectoda je vytvořen a spravován společností Light Seekres s.r.o. sídlící v Praze. Spectodu představují jako nový komunikační protokol pro bezdrátové ovládání světelcích produktů [24]. Pro převod digitálních příkazů v podobě Spectoda-kódu na světelné animace je využíváno procesorů ESP32 a jejich varianty umístěné ve vlastních obvodech speciálně navržených pro různé problematiky využití. Procesory následně ovládají připojené světelné zdroje nejčastěji pak adresovatelné led pásky osazené čipy WS2812B nebo WS2815.

Spectoda kód představující animaci je zároveň uložen na plošené desce. Proto v případě ztráty signálu se zdrojem kódu, může světelné zařízení nadále pracovat. Tím se systém stává výhodnější oproti rozšířenému protokolu DMX⁷, u kterého je nutný nepřetržitý tok příkazů.

Systém spectoda využívá pro generování animací několik základních animací. Tyto animace následně mohou být mezi sebou odčítány scítány a skládány do bloků animací. Animace i bloky je možné posouvat a skládat zasebe či přes sebe v čase. Typickým příkladem jsou animace pro led pásky tvořící unikátní efekty v rámci architektonického osvětlení či přenosných světel. Animace je možné synchronizovat s hudebním podkladem, takže se hodí také pro scénické osvětlení v rámci kreativních vystoupení.

⁷Digital Multiplexer - Protokol pro digitální přenos informací a řízení scénických světel [28].

1.8.1 Tvorba animace

Pro tvorbu animací je vytvořeno Spectoda studio [24] dostupné ve webovém prohlížeči. V rámci Spectoda studia jsou dostupné nástroje pro tvorbu animací, jejich následné nahrání do fyzického kontroleru a přehrání animace. Tvorba animace probíhá pomocí blokového programování. Animace vzniká skládáním základních bloků. Bloky je možné skládat do komplexních struktur a kombinovat mezi sebou. Z bloků je generován Spectoda Code, který je kompilátorem kompilován do řad hexadecimálních čísel. Základními bloky jsou:

TODO: Místo tohoto seznamu obrázek ze studia

„solid color“

„rainbow cycle“

„transition“

„shot“

„loading“

„color cycle“

„gradient“

Práce s led pásky

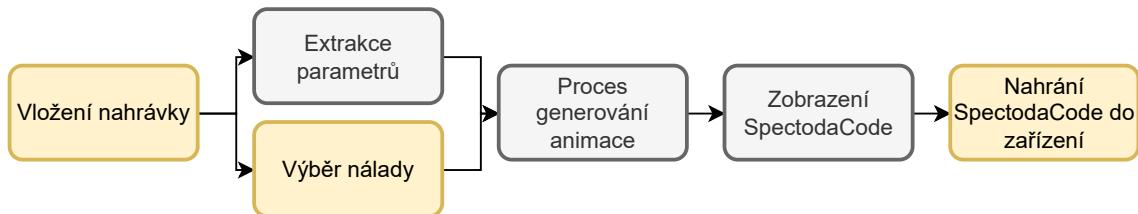
2 Výsledky studentské práce

2.1 Návrh výsledného systému

V kapitole je popsán výsledný návrh výsledného systému. Systém se skládá z několika částí: uživatelské rozhraní, algoritmů pro získání pamrametrů hudební nahrávky, algoritmu generujícího SpectodaCode na základě získaných parametrů a databáze bloků. Podkapitoly pak podrobně popisují návrh každé z částí systému.

2.1.1 Uživatelské rozhraní

Jedná se o jednoduché webové rozhraní, ve kterém uživatel nahraje hudební skladbu ve formátu .wav nebo mp3. Rozhraní obsahuje pole pro vložení cesty k hudební skladbě umožňující výběr ze souborů v uživatelském úložišti. Níže je posuvník s 4 základními hodnotami pro výběr nálady „mood“. Jedná se hodnoty „chill“, „hang out“, „feeling happy“ a „dancing“, které jsou v tomto pořadí na posuvníku. Uživatel může pomocí posouvání posuvníku vybrat pro jakou náladu chce vytvořit animaci. Pod posuvníkem pro výběr nálady se nachází tlačítko spouštějící proces generování SpectodaCodu. Poslední částí webového rozhraní je textové pole, ve kterém se zobrazí vygenerovaný kód. Na blokovém schématu 2.1 je zobrazen proces postupu uživatele skrze webové rozhraní.



Obr. 2.1: Blokové schéma postupu uživatele webovou stránkou

2.1.2 Parametry hudební nahrávky

Systém pro generování SpectodaCode popsány v bodě 2.1.3 vyžaduje vstupní data o hudební nahrávce. Tyto data jsou rozdělena na 7 parametrů. Každý z parametrů představuje vlastnost analyzované nahrávky. Tyto vlastnosti jsou získány pomocí technik popsánych v bodě 2.2. Jednotlivé vlastnosti a jejich datové struktury jsou shrnutы v následujících bodech.

Detecte dob představuje pole hodnot jehož délka je závislá na době trvání nahrávky. Jednotlivé hodnoty pak udávají časy hudební nahrávky, ve kterých se doby nacházejí.

Tempo skladby je číslo typu float s jednotkou BPM vyjadřující počet úderů za minutu. Hodnota BPM se typicky vztahuje k počtu čtvrtových not za minutu [1]. Vybraný algoritmus z knihovny Librosa neumožňuje rozeznat o jaké doby se jedná. Tempo je určeno jako počet detekovaných úderů za minutu. Postup zjištění tempa je popsán v bodě 1.7.1.

Chromavektory jsou získány v podobě 2D pole. Počet řádků pole je dán 12 půltóny. Délka pole je závislá na délce nahrávky a velikosti okna při výpočtu STFT. Z chromavektorů je spočítána paleta barev dané nahrávky. Paleta barev je uložena jako pole dvanácti hodnot. Pro každý půltón jedna barva. Generování barev je popsáno více v bodech 2.3.3 a 2.3.2.

Hlasitost představuje normalizovanou vnímanou úroveň hlasitosti skladby s ohledem na vlastnosti lidského ucha popsané psychoakustikou. Je počítána buď pro celou skladbu, nebo její část/segment a je značena v jednotkách LUFS. Metoda výpočtu hlasitosti je popsána v bodě 1.2.6.

Segmentace rozděluje nahranou skladbu do 12 segmentů ve kterých se objevují stejné hudební tvary. Segmenty jsou uloženy jako pole časů hranic segmentů. Kde hranice představuje konec jednoho segmentu a začátek druhého.

Žánr je pro zvolenou nahrávku uložen v proměnné *genre* pomocí slovníku 10 žánrů, ke kterým je přiřazena jejich pravděpodobnost. Protože se v hudbě vyskytuje velké množství žánrů, které se mezi sebou prolínají nejde s přesnou jistotou zařadit každou nahrávku přesně k jednomu žánru. Proto je použita metoda pravděpodobností. Použité žánry jsou popsány v úryvku kódu 2.1.

```
# Song genre constants
BLUES      = "blues"
CLASSICAL  = "classical"
COUNTRY    = "country"
DISCO      = "disco"
HIPHOP     = "hiphop"
JAZZ       = "jazz"
METAL      = "metal"
POP        = "pop"
REGGAE     = "reggae"
ROCK       = "rock"
```

Výpis 2.1: Hodnoty proměnné *genre*

Nálada představuje proměnou *mood* typu float s přednastavenými statickými hodnotami zobrazenými v části kódu 2.2. Náladu zadává uživatel při nahrávání zvolené skladby.

```
# Song mood constants
```

CHILL	= 0
HANG_OUT	= 1
HAPPY	= 2
DANCING	= 3

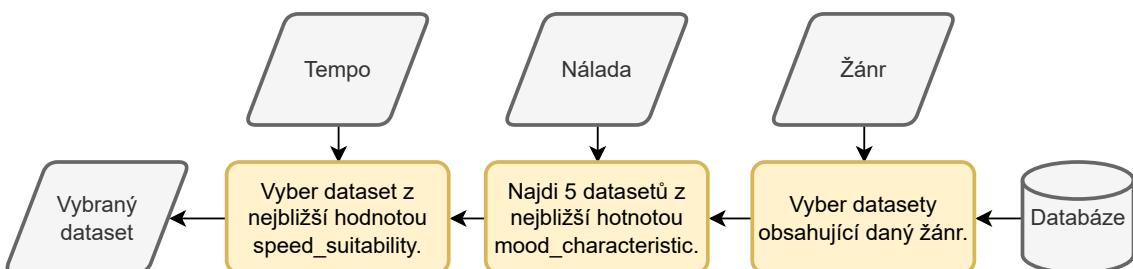
Výpis 2.2: Hodnoty proměnné *mood*

2.1.3 Systém pro generování animací

Systém pro generování animací je jádrem práce. Jeho struktura udává vizuální kvalitu animací a schopnost přizpůsobit se žánrově různorodým skladbám. V této kapitole je popsána struktura systému.

První částí systému je vstupní rozhraní, ve kterém jsou přijímány data obsahující parametry o hudební nahrávce. Struktura přijímaných dat je popsána v bodě 2.1.2. Každý ze zmíněných parametrů plní důležitou funkci v rozhodovacím procesu skládání bloků animace. Níže jsou popsány rozhodovací funkce pro jednotlivé parametry.

Žánr a nálada jsou používány pro výběr vhodného balíčku animací. Tyto balíčky jsou nazývány datasety a jejich datová struktura je popsána v bodě 2.1.4. Žánr i nálada jsou zaznamenány jako předdefinovaná celočíslená hodnota. Postup výběru datasetu je následující. Každý dataset obsahuje seznam žánrů pro který je vhodný. Na základě proměnné *genre* dochází k výběru všech datasetů vhodných pro daný žánr. Následně na základě proměnné *mood* ve které je uložena nálada je vybráno 5 datasetů s nejbližší hodnotou *mood_characteristic*.



Obr. 2.2: Blokový diagram výběru datasetu.

Tempo skladby je posledním parametrem při výběru vhodného datasetu. Hodnota proměnné *speed_suitability* každého z 5 vybraných datasetů je porovnána s tempem skladby. Finální dataset je vybrán ten jehož hodnota *speed_suitability* je nejbližší hodnotě tempa skladby. Blokový diagram procesu je znázorněn na obrázku 2.2. Tempo skladby je také použito pro výpočet délky jednotlivých animací na které je přímo úměrná rychlosť animace.

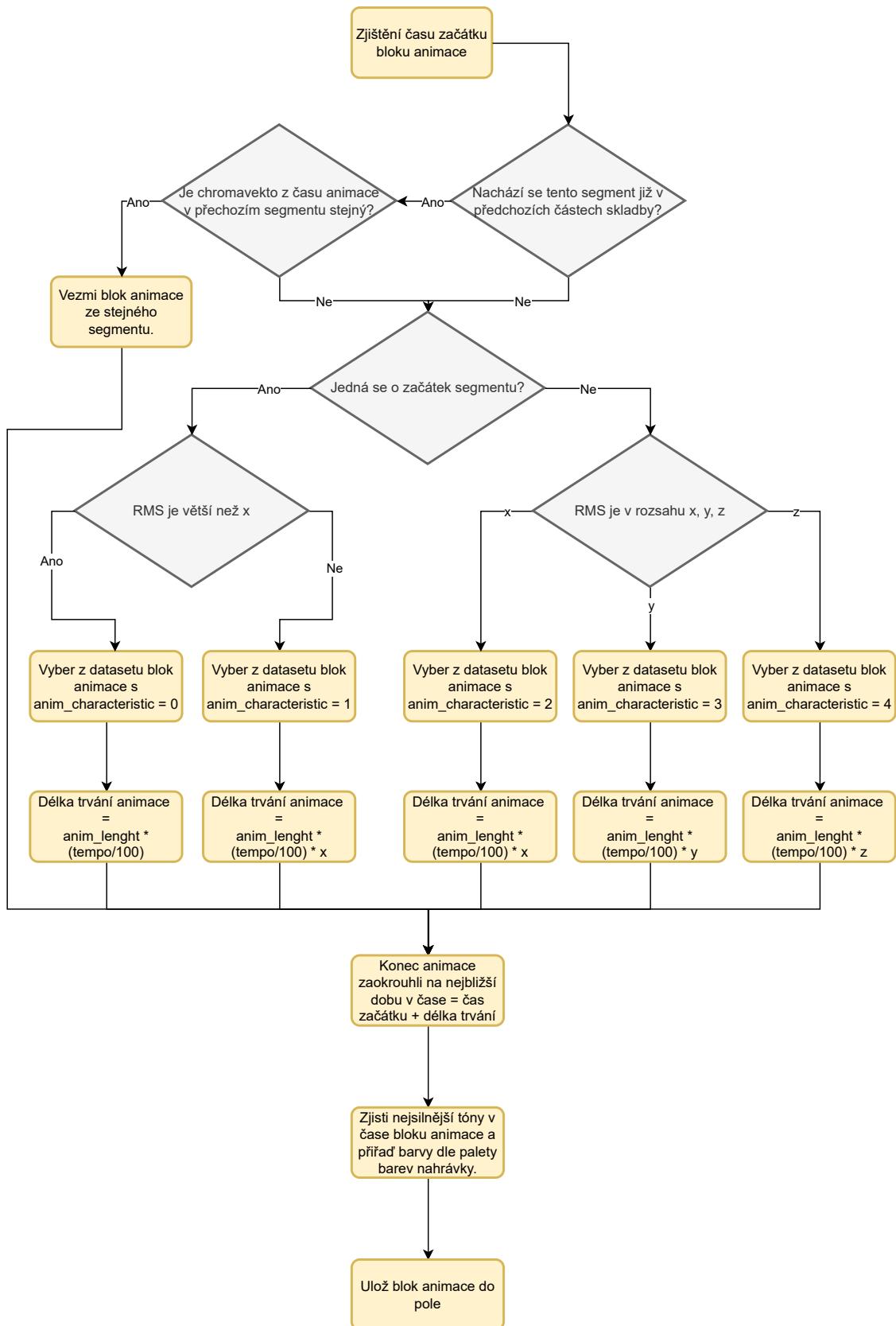
Segmentace slouží pro detekci opakujících se částí nahrávky. Například nahrávka obsahuje více než jednu sloku či refrém. Je kladen důraz aby v opakujících se segmentech nahrávky byla animace stejného typu. Z toho plyne, že v každém refrému bude podobná animace.

Detekce dob je parametrem na základě kterého se nastaví začátky a konce animací tak, aby odpovídaly rytmu nahrávky.

Chroma vektory udávají tónovou strukturu skladby v průběhu času. Tento parametr je využit pro nastavení barevné škály animací. Chromavektory se analyzují a jsou vybrány 4 hlavní tóny nejčastěji se vyskytující v nahrávce. Keždému tónu je přiřazen barevný odstín. Poté je animaci v dané části skladby přiřazena barva dle tónu určeného chromavektorem daného času nahrávky.

Efektivní hodnota signálu pomáha procesu segmentace. Refrémy a sloky se mohou lyšit hlasitostí, která je zpříjemněna s efektní hodnotou. Zároveň je použita pro výběr vhodného typu animace. Na základě velikosti efektivní hodnoty signálu v čase je vybrán typ animace. Vyšší hodnoty představují rychlejší a údernější animace. Nižší hodnoty naopak pomalejší a klidnější animace. Tento proces je zajištěn porovnáváním proměnné *anim_characteristic* obsažené v každém bloku animace s efektivní hodnotou mapovanou na určený rozsah.

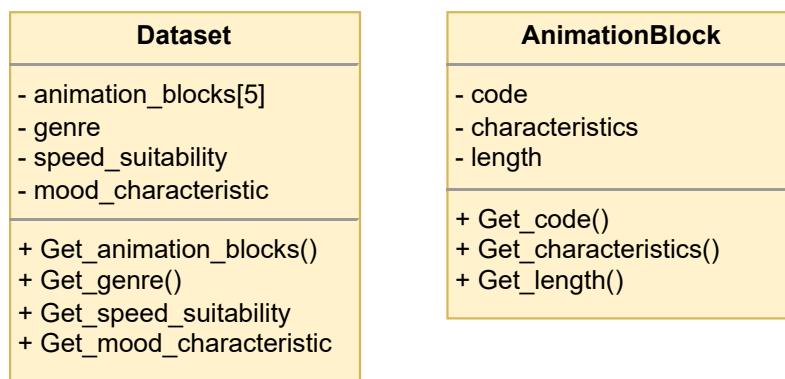
Druhá část systému tvoří samotnou logiku skládání blkoků animací. Tato rozhodovací struktura je zobrazena na blokovém diagramu 2.3.



Obr. 2.3: Blokový diagram struktury rozhodovacího procesu

2.1.4 Databáze bloků animací

Jak je popsáno v bodě 1.8 systém Spectoda obsahuje základní animace, které jsou skládány do sebe a tím vznikají komplexní animace. Navržený systém využívá již návrhářem předsložených animací, které jsou uloženy jako bloky. Těmto blokům jsou přiřazeny parametry *code*, *characteristic* a *length*. Jejich struktura je zapsaná datovou třídou *AnimationBlock*. Z objektů tipu *AnimationBlock* jsou dále tvořeny datasety zapsané datovou třídou *Dataset*. Tyto datasety obsahují 5 bloků animací a parametry *genre*, *speed_suitability* a *mood_characteristic*. UML diagramy tříd jsou zobrazeny na obrázku 2.4.



Obr. 2.4: Blokové diagramy tříd *Dataset*, *AnimationBlock*

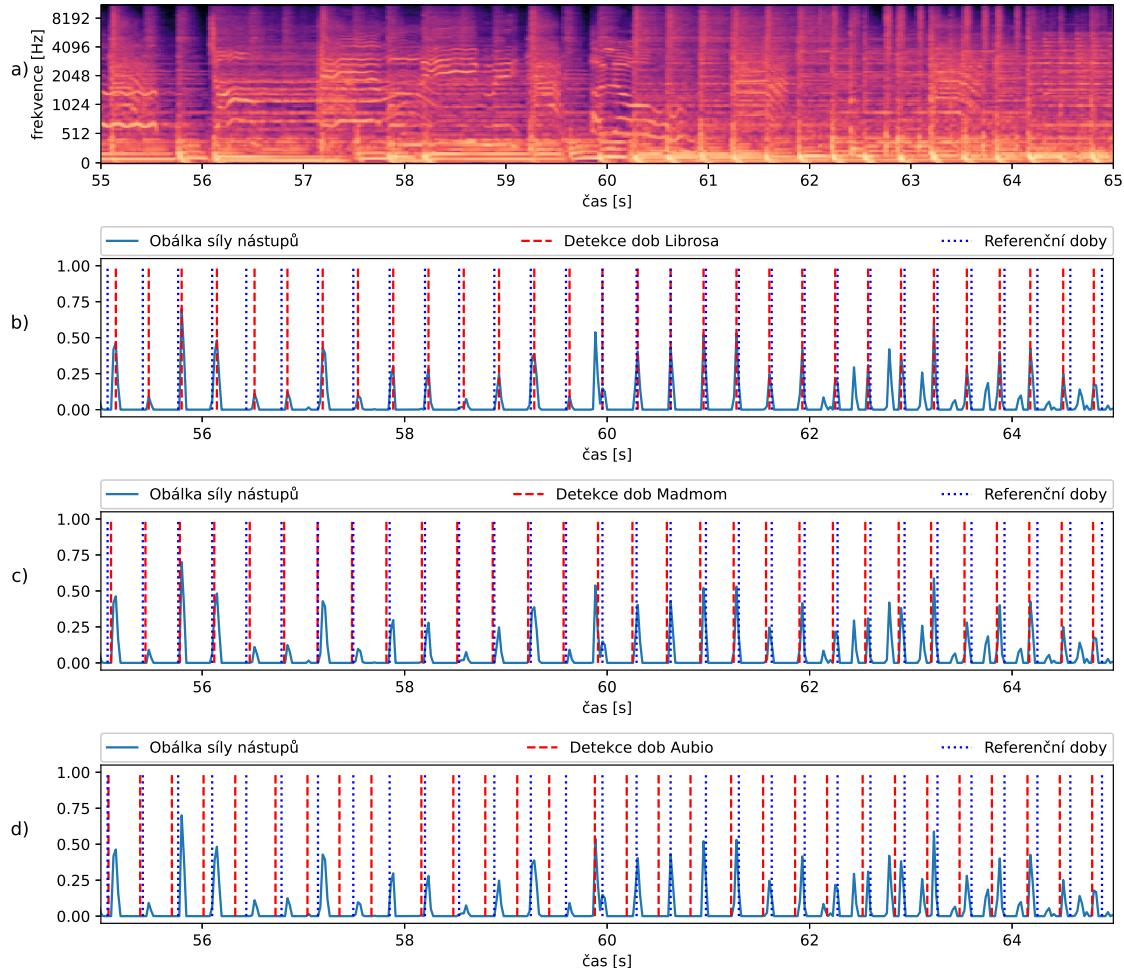
2.2 Výběr vhodných metod pro extrakci parametrů z hudební nahrávky

Díky vědecké komunitě vznikly volně dostupné knihovny obsahující techniky z oborů MIR. V této části práce jsou prozkoumány 3 knihovny zmíněné v bodě 1.7. Jsou použity jejich funkce pro získání parametrů z hudební nahrávky potřebných pro navazující diplomovou práci. Tyto funkce jsou mezi sebou porovnány z hlediska přesnosti výsledků, rychlosti výpočtů, jednoduchosti použití a možnosti využití pro komerční účely.

2.2.1 Detekce dob a tempa

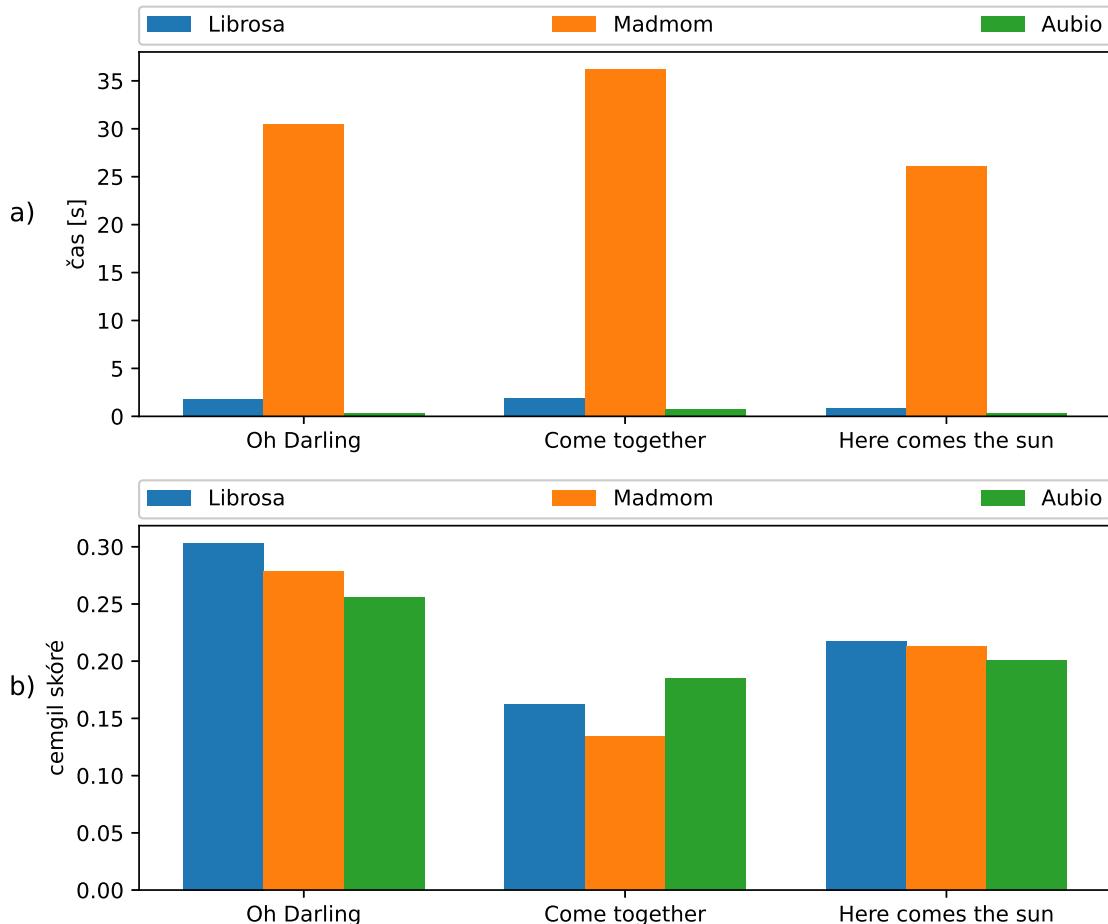
Pro porování detekce dob jsou vybrány 3 funkce. Pro každou knihovnu jedna funkce. První je z knihovny Librosa funkce *beat_track*. Z knihovny Madmom je použit *BeatTrackingProcessor* a z knihovny Aubio funkce *tempo*. Funkce jsou porovnány na třech skladbách skupiny Beatles. Níže v grafu 2.5 je vidět úryvek skladby

Oh-Darling!. Pro lepší zobrazení jsou osy grafu v rozsahu 55 - 65 s nahrávky. Na prvním z grafů je vyobrazen mels pektrogram vypočítán pomocí funkce *melspectrogram* z knihovny Librosa. Grafy b), c) a d) zobrazují v pozadí obálku síly nástupů a vertikální pruhované čáry znázorňují detekované doby dané funkce. Vertikální modré tečkované čáry jsou referenční doby zaznamenané institucí Centre for digital music na univerzitě Queen Mary, University of London [14].



Obr. 2.5: Porovnání metod detekce dob na úryvku skladby Oh-Darling!. **a)** Mel spektrogram **b)** Detekce dob pomocí Librosa **c)** Detekce dob pomocí Madmom **d)** Detekce dob pomocí Aubio

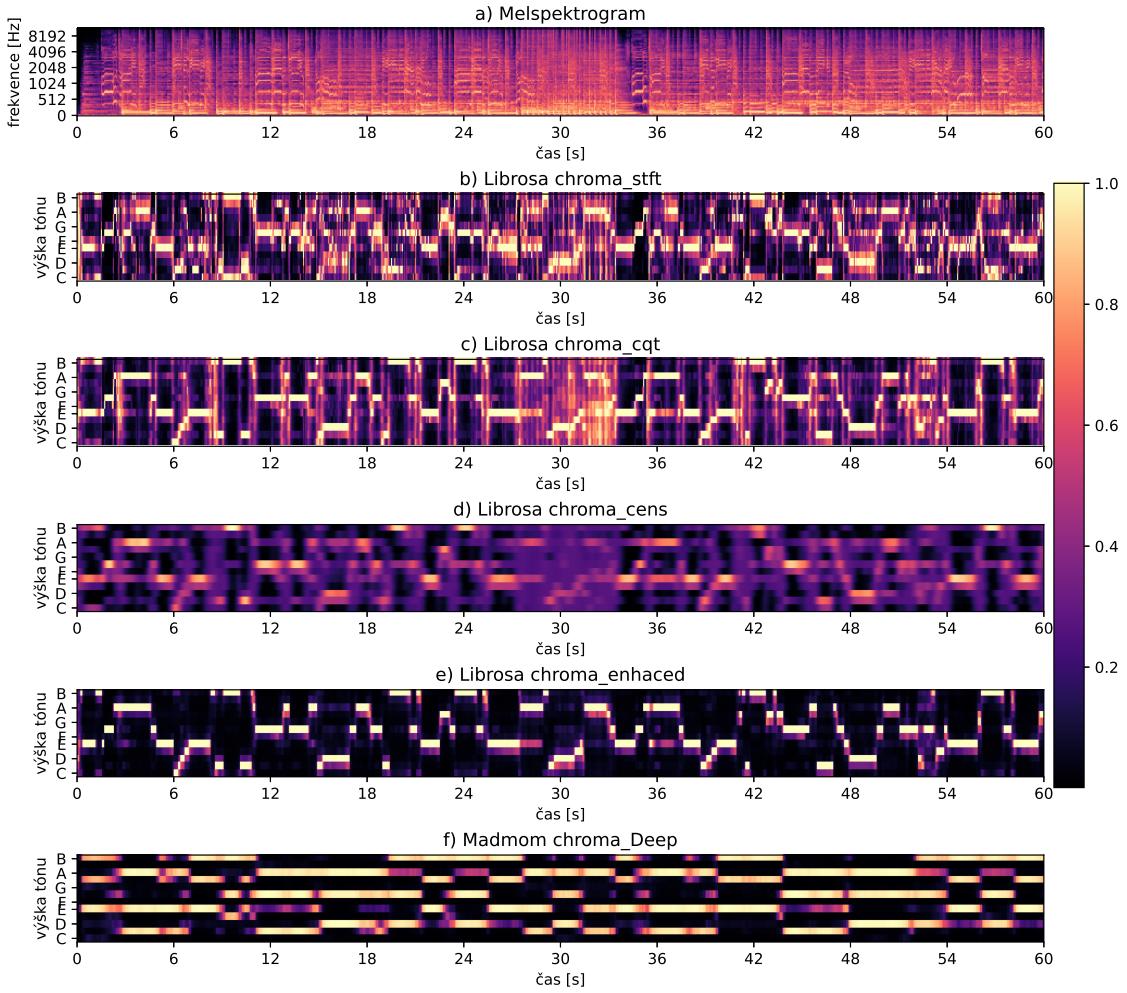
Z grafu lze vidět, že funkce z knihoven Librosa a Madmon se nejvíce přibližují referenčním dobám. Pro přesné hodnocení funkcí je využita knihovna Mir_eval poskytující funkce pro hodnocení přesnosti detekce dob. Pomocí této knihovny je počítáno Camgil skóré. Popis knihovny a výpočtu je zmíněn v bodě 1.7.4. Posledním hodnoceným parametrem je doba výpočtu. Výsledky Camgil skóré, a doby výpočtu pro jednotlivé skladby jsou zobrazeny na obrázku 2.6.



Obr. 2.6: Porovnání přesnosti a času metody detekce dob na skladech Oh-Darling!, Come Together a Here Comes The Sun. **a)** Čas výpočtu **b)** Cemgil skóre

2.2.2 Analýza chromavektorů

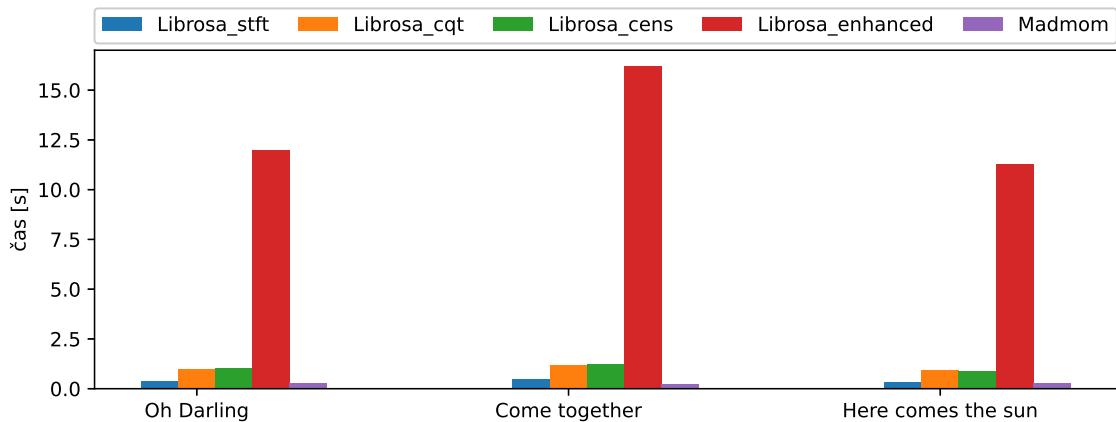
Hodnocení vhodné funkce pro výpočet chromavektorů je realizováno zejména vizuálně na zobrazených grafech referenčních skladeb. Tyto grafy lze vidět na obrázku 2.7. Porovnávány jsou 4 funkce z knihovny librosa *chroma_stft*, *chroma_cqt*, *chroma_cens* a čtvrtou je funkce *chroma_cqt* c řetězci s předzpracováním signálu a filtrací výsledných chromavektorů. Principy výpočtu jednotlivých funkcí jsou popsány v bodě 1.7.1. Z knihovny Madmom je použit *DeepChromaProcessor*.



Obr. 2.7: Porovnání výpočtu chromavektorů na skladbě Oh-Darling! v časovém rozmezí 0 - 60 s.

Z obrázku 2.7 lze vidět patrný rozdíl v přístupu výpočtu chromavektorů. U grafů *b*), *c*) a *d*) lze vidět velké množství šumu. U grafu *e*) je šum vyfiltrován přidanými metodami. Na grafu *f*) lze vidět, že knihovna Madmom a z ní použitý *DeepChromaProcessor* přistupuje k výpočtu chromavektorů odlišně. Je využito delší časové okno pro transformaci do frekvenční oblasti. Z toho je patrné, že v časovém měřítku údaj není natolik přesný jako u knihovny Librosa. Pro následné použití v algoritmu pro generování animací je však tento výsledek mnohem stálejší.

Dalším hodnoceným parametrem je délka výpočtu chromavektorů. Ta je závislá na složitosti algoritmů jež funkce využívají k výpočtu. Porovnání je zobrazeno na obrázku 2.8. Z obrázku je zřejmé, že pomocné metody pro předzpracování a filtrace chromavektorů z funkce *chroma_cqt* přidaly značné množství výpočetního času oproti samotnému výpočtu chromavektorů.

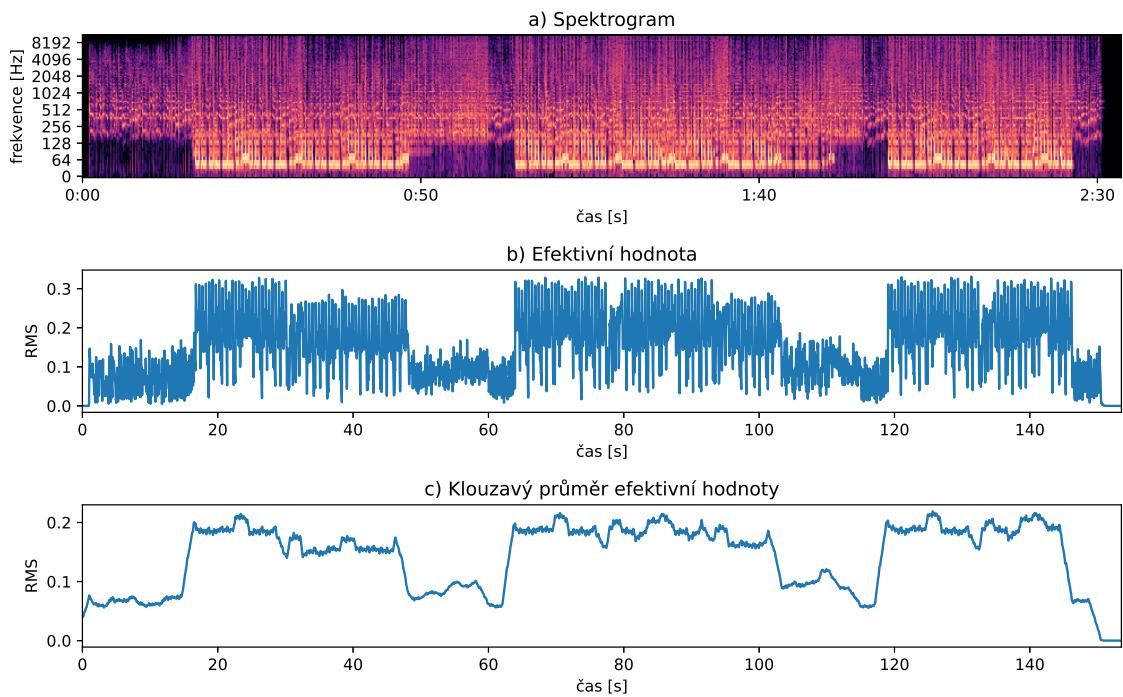


Obr. 2.8: Porovnání délky výpočtu chromavektorů

2.2.3 Efektivní hodnota signálu

TODO: Hlasitost v LUFS

Výpočet efektivní hodnoty signálu je realizován pomocí funkce z knihovny Librosa. Pro výpočet je použito okno o délce 2048 vzorků. Hodnoty jsou vycentrovány na střed rozsahu okna tedy 1024 vzorků. Získaná křivka je uhlazena pomocí výpočtu klouzavého průměru. Počet vzroků ze kterých se klouzavý průměr vypočítá odpovídá délce signálu 7,5 s a je závislý na jeho vzorkovací frekvenci. Získané informace jsou zobrazeny na obrázku 2.9.



Obr. 2.9: Zobrazení efektivní hodnoty skladby Belly dancer

2.3 Realzace

2.3.1 Uživatelské rozhraní

Uživatelské rozhraní je reprezentováno webovou stránkou a je naprogramováno pomocí značkovacího jazyka HTML spolu s formátováním v jazyce CSS. Funkčnost a propojení se systémem je řešeno pomocí frameworku Flask.

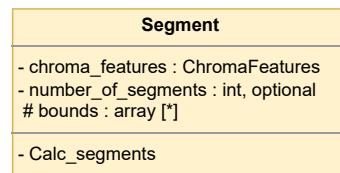
2.3.2 Extrakce parametrů

Pro získání všech potřebných parametrů z nahraného audio souboru jsou naprogramovány čtyři třídy. Tyto třídy poskytují veškeré potřebné parametry.

ChromaFeatures
- y : array [*]
- sr : float
- mood : int
chroma : array [12][*]
tones_colors : array [12]
- Calc_chroma_librosa()
- Calc_chroma_madmom()
- Calc_color_paleta()
- Generate_triadic_colors()
- Color_shift()

Obr. 2.10: Struktura třídy *ChromaFeatures*

Pro práci se segmenty je definována třída *Segmentation* a její struktura je zobrazena na obrázku 2.11.



Obr. 2.11: Struktura třídy *Segment*

2.3.3 Přiřazení barev

Závěr

V rámci semestrální práce byly splněny v úvodu stanovené cíle. Byla prozkoumána vědní oblast MIR, metody vzniklé v rámci výzkumů v této oblasti a možnost jejich využití pro navazující diplomovou práci. Komunita soustředící se kolem oblasti MIR vytvořila pro vědecké účely volně dostupné knihovny. Tyto knihovny jsou lehce implementovatelné v programovacím jazyce python a obsahují pokročilé metody pro analýzu a práci s hudebními nahrávkami v časové i frekvenční oblasti.

V části 1 je popsána teorie základních principů práce s hudební nahrávkou v digitální formě. Od metod využívajících základní principy zpracování signálů po moderní řešení s využitím strojového učení. Dále jsou v této části v bodech 1.7.1 až 1.7.4 popsány víše zmíněné dostupné knihovny jejich metody pro extrakci parametrů, potřebných pro navazující diplomovou práci, a principy jak jsou tyto metody realizovány. Zde je zapotřebí v rámci diplomové práce rádně doplnit popis knihoven Madmom, Aubio a Mir_eval.

Třetím cílem semestrální práce byl návrh systému pro generování SpectodaCodu na základě parametrů nahrávky. SpectodaCod je následně nahrán do Spectoda zařízení, které kód převádějí na světelné animace. Návrh systému je rozdělen na několik částí. Uživatelské rozhraní popsáne v bodě 2.1.1, parametry potřebné pro realizaci funkčního procesu generování animací a popis jejich datové struktury je v bodě 2.1.2. Popis jak by celý proces generování měl v principu fungovat a jeho struktura je popsána v bodě 2.1.3.

Posledním bodem práce byla samotná extrakce vybraných parametrů z hudební nahrávky. Jsou prozkoumány dostupné metody a vytvořeny Jupyter notebooky, jež jsou k nalezení v příloze práce, obsahují porovnání těchto metod na základě přesnosti a rychlosti výpočtů. Hodnocena je také vhodnost pro výsledný systému. Tato část práce je popsána v bodě 2.2.

V rámci navazující diplomové práce dojde k realizaci výsledného systému a jeho uvedení do funkčního stavu. Bude naprogramováno uživatelské rozhraní popsáne v bodě 2.1.1 a realizována rozhodovací struktura zobrazena v bodě 2.1.3. Tato struktura bude následně testována a upravena, aby bylo dosaženo funkčního generování vizuálně zajímavých animací. Na základě existence rozhodovací struktury dojde k přezkoumání vhodnosti získaných parametrů a jejich úprava pro lepší výsledky při generování animací. Budou také doplněny body, které se nezdařilo realizovat v rámci semestrální práce. Jedná se o teorii a realizaci automatické klasifikace žánrů. Popis knihoven Madmom, Aubio a Mir_eval.

Literatura

- [1] Apel, W.: *Harvard Dictionary of Music*. Harvard University Press, 1947.
URL <https://books.google.cz/books?id=Y6m0AAAAIAAJ>
- [2] Bracewell, R.: *The Fourier Transform and its Applications*. Tokyo: McGraw-Hill Kogakusha, Ltd., druh vyd n , 1978.
- [3] Cartwright, K.: Determining the effective or RMS voltage or various waveforms without calculus. ro n k 8, 01 2007.
- [4] Cohen, L.: Time-frequency distributions-a review. *Proceedings of the IEEE*, ro n k 77, . 7, 1989: s. 941–981, doi:10.1109/5.30749.
- [5] Crocker, M.: *Handbook of Acoustics*. A Wiley-Interscience Publication, Wiley, 1998, ISBN 9780471252931.
URL https://books.google.cz/books?id=1x_RvffW-hcC
- [6] Downie, J. S.; Ehmann, A. F.; Bay, M.; aj.: *The Music Information Retrieval Evaluation eXchange: Some Observations and Insights*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, ISBN 978-3-642-11674-2, s. 93–115, doi: 10.1007/978-3-642-11674-2_5.
URL https://doi.org/10.1007/978-3-642-11674-2_5
- [7] Ellis, D.: Beat Tracking by Dynamic Programming. *Journal of New Music Research*, ro n k 36, 03 2007: s. 51–60, doi:10.1080/09298210701653344.
- [8] Acoustics — Determination of sound power levels and sound energy levels of noise sources using sound pressure — Engineering methods for an essentially free field over a reflecting plane. Standard, International Organization for Standardization, B ezen 2010.
URL <https://www.iso.org/obp/ui/#iso:std:iso:3744:ed-3:v1:en>
- [9] Klapuri, A.; Davy, M.: *Signal Processing Methods for Music Transcription*. 01 2006, ISBN 978-0-387-30667-4, doi:10.1007/0-387-32845-9.
- [10] Large, E. W.; Kolen, J. F.: Resonance and the Perception of Musical Meter. *Connection Science*, ro n k 6, . 2-3, 1994: s. 177–208, doi:10.1080/09540099408915723, <https://doi.org/10.1080/09540099408915723>.
URL <https://doi.org/10.1080/09540099408915723>
- [11] Lartillot, O.; Grandjean, D.: Tempo and Metrical Analysis by Tracking Multiple Metrical Levels Using Autocorrelation. *Applied Sciences*, ro n k 9, 11 2019: str. 5121, doi:10.3390/app9235121.

- [12] Lidy, T.; Rauber, A.: Music Information Retrieval. In *Handbook of Research on Digital Libraries: Design, Development, and Impact*, IGI Global, 2009, ISBN 978-1-59904-879-6, s. 448–456.
- [13] Matthew E. P. Davies, M. F., Sebastian Bock: *Tempo, Beat and Downbeat Estimation*. <https://tempobeatdownbeat.github.io/tutorial/intro.html>, 2021.
URL <https://tempobeatdownbeat.github.io/tutorial/intro.html>
- [14] Mauch, M.: Isophonic datasets. [Online; 11. 12. 2023].
URL <http://isophonics.net/>
- [15] McAdams, S.; Giordano, B. L.: 113The Perception of Musical Timbre. In *The Oxford Handbook of Music Psychology*, Oxford University Press, 01 2016, ISBN 9780198722946, doi:10.1093/oxfordhb/9780198722946.013.12, https://academic.oup.com/book/0/chapter/292611024/chapter-ag-pdf/44515461/book_34489_section_292611024.ag.pdf.
URL <https://doi.org/10.1093/oxfordhb/9780198722946.013.12>
- [16] McFee, B.; Raffel, C.; Liang, D.; aj.: librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, ro n k 8, 2015.
- [17] McKinney, M. F.; Moelants, D.: Ambiguity in Tempo Perception: What Draws Listeners to Different Metrical Levels? *Music Perception: An Interdisciplinary Journal*, ro n k 24, . 2, 2006: s. 155–166, ISSN 07307829, 15338312.
URL <http://www.jstor.org/stable/10.1525/mp.2006.24.2.155>
- [18] Müller, M.: *Fundamentals of Music Processing*. Springer International Publishing, 2015, doi:10.1007/978-3-319-21945-5.
URL <https://doi.org/10.1007%2F978-3-319-21945-5>
- [19] Müller, M.; Klapuri, A.: Chapter 27 - Music Signal Processing. In *Academic Press Library in Signal Processing: Volume 4, Academic Press Library in Signal Processing*, ro n k 4, editace J. Trussell; A. Srivastava; A. K. Roy-Chowdhury; A. Srivastava; P. A. Naylor; R. Chellappa; S. Theodoridis, Elsevier, 2014, s. 713–756, doi:<https://doi.org/10.1016/B978-0-12-396501-1.00027-3>.
URL <https://www.sciencedirect.com/science/article/pii/B9780123965011000273>
- [20] Salamon, J.; Gomez, E.: Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, ro n k 20, . 6, 2012: s. 1759–1770, doi:10.1109/TASL.2012.2188515.

- [21] Scheirer, E. D.: Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, ro n k 103, . 1, 01 1998: s. 588–601, ISSN 0001-4966, doi:10.1121/1.421129, https://pubs.aip.org/asa/jasa/article-pdf/103/1/588/8083614/588_1_online.pdf.
URL <https://doi.org/10.1121/1.421129>
- [22] Schreibman, S.; Siemens, R.; Unsworth, J. (edito i): *A new companion to Digital Humanities*. West Sussex, England: John Wiley & Sons Ltd, 2016, ISBN 9781118680599.
- [23] Sneddon, I.: *Fourier Transforms*. Dover books on mathematics, Dover Publications, 1995, ISBN 9780486685229.
URL <https://books.google.cz/books?id=jhpsLpRERwC>
- [24] s.r.o, L. S.: Spectoda. [Online; 12. 12. 2023].
URL <https://spectoda.com/>
- [25] Stevens, S. S.; Volkmann, J.; Newman, E. B.: A Scale for the Measurement of the Psychological Magnitude: Pitch. *The Journal of the Acoustical Society of America*, ro n k 8, . 3, 1937: s. 185–190, ISSN 0001-4966.
- [26] Strichartz, R.: *A Guide To Distribution Theory And Fourier Transforms*. World Scientific Publishing Company, 2003, ISBN 9789813102293.
URL <https://books.google.cz/books?id=YfA7DQAAQBAJ>
- [27] Syrový, V.: *Hudební akustika*. Akustická knihovna Zvukového studia Hudební fakulty AMU, Akademie múzických umění, 2013, ISBN 9788073312978.
URL <https://books.google.cz/books?id=ikrmoeAEACAAJ>
- [28] United States Institute for Theatre Technology, I.: BSR E1.11, Entertainment Technology - USITT DMX512 Asynchronous Serial Digital Data Transmission Standard for Controlling Lighting Equipment and Accessories. *Entertainment Services and Technology Association*, 1999, online, 12.12.2023.
URL <http://www.jstor.org/stable/10.1525/mp.2006.24.2.155>
- [29] Tumarkin, A.: *The Decibel, The Phon and the Sone*, ro n k 64. Cambridge University Press, 1950, 178–188 s., doi:10.1017/S0022215100011919.
- [30] Wikipedie: Musical Instrument Digital Interface — Wikipedie: Otevřená encyklopédie. 2022, [Online; navštívěno 1. 12. 2022].
URL https://cs.wikipedia.org/w/index.php?title=Musical_Instrument_Digital_Interface&oldid=21081530

- [31] WikiSkripta: Vlastnosti zvuku —. 2022, [Online; navštíveno 21. 11. 2022].
URL https://www.wikiskripta.eu/index.php?title=Vlastnosti_zvuku&oldid=458442

Seznam symbolů a zkratek

MIR	Music information retrieval - Obor zabývající se vyhledávání informací v hudebních dílech
MIDI	Musical Instrument Digital Interface - Digitální rozhraní hudebních nástrojů
ISMIR	International Society of Music Information Retrieval - Mezinárodní združení pro MIR
MIREX	The Music Information Retrieval Evaluation eXchange
FT	Fourier transform - Fourierova transformace
FFT	Fast Fourier transform - Rychlá Fourierova transformace
DFT	Discrete Fourier transform - diskrétní Fourierova transformace
STFT	Short-time Fourier transform - krátkodobá Fourierova transformace
RMS	Root mean square - efektivní hodnota
BPM	Beats per minute - doby za minutu
HTML	Hypertext markup language - hypertextový značkovací jazyk
DMX	Digital Multiplexer - protokol pro digitální přenos informací
LUFS	Loudness units relative to full scale