

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

SEMESTRÁLNÍ PRÁCE

Brno, 2022

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

SEMESTRÁLNÍ PRÁCE

SEMESTRAL THESIS

AUTOR PRÁCE

AUTHOR

Bc. Viktor Slezák

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Matěj Ištváněk

BRNO 2022

Semestrální práce

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

Student: Bc. Viktor Slezák

ID: 203745

Ročník: 2

Akademický rok: 2022/23

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. Využijte nejnovější přístupy založené na metodách strojového učení pro extrakci relevantních parametrů. Získaná data analyzujte a na jejich základě navrhnete a naprogramujete 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. Cílem semestrálního projektu je popis parametrů a informací, které lze smysluplně použít pro generování světelných animací. Semestrální práce bude obsahovat implementaci skriptů pro výpočet parametrů a návrh struktury výsledného systému. V budoucí diplomové práci budou parametry využity a optimalizovány pro generování kódu, který bude data převádět na sekvence animací pro ovládání světla.

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: 1.10.2022

Termín odevzdání: 12.12.2022

Vedoucí práce: Ing. Matěj Ištváněk

doc. Ing. Jiří Schimmel, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor semestrální práce nesmí při vytváření semestrální 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.

SLEZÁK, Viktor. *Světelné animace pro systém Spectoda na základě analýzy parametrů z hudebních nahrávek*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2022, 31 s. Semestrální práce. Vedoucí práce: Ing. Matěj Ištváněk

Prohlášení autora o původnosti díla

Jméno a příjmení autora:	Bc. Viktor Slezák
VUT ID autora:	203745
Typ práce:	Semestrální práce
Akademický rok:	2022/23
Téma závěrečné práce:	Světelné animace pro systém Spectoda na základě analýzy parametrů z hudebních nahrávek

Prohlašuji, že svou závěrečnou práci jsem vypracoval samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

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

Brno
podpis autora*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu Ing. Matěj Ištváněk za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Úvod	11
1 Teorie	12
1.1 MIR - Music information retrieval	12
1.1.1 Historie	12
1.1.2 Řetězec zpracování - pipeline	13
1.1.3 Současné problémy	15
1.2 Parametrizace hudebních nahrávek	15
1.2.1 Reprezentace audio signálů	15
1.2.2 Časová oblast	16
1.2.3 Frekvenční oblast	17
1.2.4 DFT - Diskrétní Fourierova transformace	19
1.2.5 STFT - Short-time Fourier transform	20
1.2.6 Dynamika hlasitosti a intenzita	22
1.2.7 Barva	22
1.3 Detekce dob a analýza tempa skladby	25
1.3.1 Využití energie signálu	25
1.3.2 Využití spektra signálu	25
1.3.3 Detekce periodicity	25
1.3.4 Využití neuronových sítí	25
1.3.5 Více vrstvé perceptronové sítě	25
1.3.6 Konvoluční neuronové sítě	25
1.3.7 Rekurentní neuronové sítě	25
1.3.8 Hybridní architektury	25
1.4 Klasifikace žánrů a nálady	25
1.5 "Získání" chromavektorů	25
1.6 Systém Spectoda	25
1.7 Hudební signál jako animace	25
2 Výsledky studentské práce	26
2.1 Návrh struktury výsledného algoritmu	26
Závěr	27
Literatura	28
Seznam symbolů a zkratk	30

Seznam obrázků

1.1	Řetězec procesů MIR [11]	14
1.2	Zobrazení časového průběhu signálu	17
1.3	Reprezentace tónu E zahraného na basovou kytaru.	18
1.4	Časově spojitý signál a diskrétní signál	19
1.5	Signál o délce 1s s počáteční frekvencí 10Hz a koncovou frekvencí 30Hz a) Původní signál b) Signál s okénkem od 0,2s do 0,5s c) Signál s okénkem od 0,35s do 0,65s d) Signál s okénkem od 0,5s do 0,8s	21
1.6	Tón A5 zahraný na klavír a jeho ADSR obálka	24

Seznam tabulek

1.1	Typické procesy na základně vstupních a výstupních dat. [11]	15
-----	--	---------	----

Úvod

V rámci semestrální práce jsou popsány možnosti pro dolování parametrů z hudebních nahrávek a jejich analýzu. Tyto techniky jsou využity pro získání potřebných informací o skladbě. Například data o tempu a rozmístění dob, žánr a tónové či spektrální rozložení skladby. Dále je navržena struktura algoritmu sloužícího pro převod získaných parametrů na sekvence animací kompatibilních se systémem Spectoda.

Práce je rozložena do tří na sebe navazujících cílů. Prvním z cílů je průzkum vědních oborů soustředících se na danou problematiku. Například MIR (Music information retrieval - Obor zabývající se vyhledávání informací v hudebních dílech). Z existujících výzkumů jsou vybrány postupy analýzy hudebních signálů vyhovující pro použití v rámci výsledného algoritmu.

Druhým cílem práce je navrhnout vnitřní strukturu výsledného algoritmu převádějícího získané parametry na sekvence animací pro systém Spectoda. Důležitým úkolem je vymyslet jak bude docházet k takovému přenosu a co dané parametry ovlivní v rámci generování unikátních sekvencí animace.

Poslední třetí cíl se zabývá vytvořením funkčního systému pro získávání parametrů z hudební nahrávky. Důraz je kladen na využití dostupných moderních metod analýzy hudebních signálů.

1 Teorie

Semestrální práce se zabývá zejména problematikou MIR. Popsanou v kapitole 1.1. Nabízejí se otázky jak by měla daná animace reagovat na konkrétní děj skladby. Jakým způsobem navrhnout strukturu algoritmů a co by měly získané parametry ovlivňovat při vytváření animací.

V této části jsou popsány následující segmenty: Teorie zpracování hudební nahrávky pomocí známých algoritmů. Například důležitým algoritmem je FFT (Fast Fourier transform - Rychlá Fourierova transformace) popsaná více v bodě 1.2.3, jeho varianty pak v bodech 1.2.4 a 1.2.5. Nabízené moderní metody strojového učení s využitím hlubokých neuronových sítí při detekci tempa skladby 1.3 a určení žánru 1.4. Struktura a možnosti systému Spectoda pro generování interaktivních světelných animací je podrobně popsána v bodě 1.6.

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.

Výstup jeho výzkumu je využíván populárními technologiemi. Jednou z aplikací je personalizované doporučování hudebních skladeb, která 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 šířením se digitálního audia jejich důležitost stále poroste.

1.1.1 Historie

V tomto bodě je napsán souhrn historie MIR z knihy [11]. MIR se začíná objevovat na přelomu devatenáctého a dvacátého století s příchodem moderních statistických metod. Začínají se objevovat pokusy o aplikování statistických metod na hudební partitury. Protože ještě nebyly natolik dostupné počítače jednalo se spíše o ruční práci s partiturami a tabulacemi. Z grafických notací se analyzovaly jejich rysy a specifikovaly charakteristiky hudebního díla. S příchodem počítačů do výzkumných laboratoří v letech 1960 až 1970 se 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“ a „music information retrieval“. První oblast výzkumu se soustředila na analýzu tempa skladby. Z důvodu nízké popularity se však výzkum zpomalil. Tento útlum

pokračoval až do roku 1990 kdy výzkumu MIR pomohly dvě změny. První důležitou změnou byly rostoucí databáze digitální hudby, která se staly lehce dostupné 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 jednodušší na realizaci [11].

Poté v říjnu roku 2000 bylo uspořádáno první mezinárodní symposium soustředící se na MIR. Z této mezinárodní konference se stala tradice a vybudovala se kolem ní velká komunita nazývaná ISMIR¹. Každoročním vyvrcholením ISMIR je právě vaše 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 MIREX² sloužící jako správa zásad pro hodnocení pokroků ve výzkumu MIR[5].

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“ [11]. Například digitalizovaná partitura. Dalším možným typem je „digitální hudba“. Jedná se o hudbu čistě v „digitálních notách“ představujících sadu příkazů. Například zápis v MIDI³. 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 popsáno v bodě 1.2.

Feature extraction - extrakce vlastností signálu Podle požadovaných vlastností pro extrakci je využito různých modelů popsaných v bodech 1.3, 1.4 a 1.5. S rostoucí popularitou strojového učení začaly při extrakci vlastností hudební nahrávky převládat kombinace hlubokých neuronových sítí. Tyto kombinace umožňují

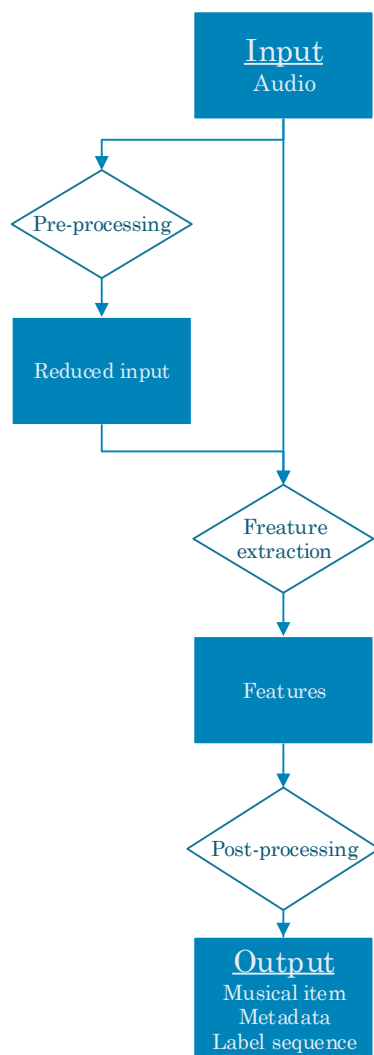
¹International Society of Music Information Retrieval - Mezinárodní sdružení pro MIR

²The 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. [5].

³Musical Instrument Digital Interface - Volně dostupný hudební standart specifikující hardwarové a softwarové požadavky pro digitální přenos hudební notace a komunikace mezi nástroji.[14]

přesnější parametrizaci a menší chybovost.

Post-processing - konečné zpracování Posledním blokem v řetězci je tvz. „post-procesing“ zajišťující zpracování a optimalizaci získaných dat. Post-procesing zpracuje data do požadované formy. V některých případech také dokáže ovlivnit přesnost zpracování.



Obr. 1.1: Řetězec procesů MIR [11]

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 [11].

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[7]. Výsledkem je kolekce paralelně složených sekvencí hodnot vlastností, které se následně zpracují na požadovaná výstupní data.

Data	Vyhledávání informací	Klasifikace a odhad	Sekvenční značení
Audio	Identifikace kopi „coverů“, Řazení skladeb, Měření podobnosti, Získání otisku, Generování seznamu skladeb	Identifikace umělce a skladatele, Žánr a nálada, Určení tempa	Extrakce melodie, Odhad akordů, Detekce nástupů, Segmentace

Tab. 1.1: Typické procesy na základně vstupních a výstupních dat. [11]

1.1.3 Současné problémy

1.2 Parametrizace hudebních nahrávek

V této kapitole je popsán audio signál. Jak vzniká, jeho reprezentace v číslicovém 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

Hudba může být reprezentována spoustou forem. Jako tradiční médium pro její ukládání ještě před vznikem záznamu sloužily vždy 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 dodává 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 zvyrazňují 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 [10].

Při pohledu z fyzikálního hlediska důsledkem interpretace 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 kmitání. Zároveň dochází k vzájemnému rozkmitání okolních částic a prostředím se začne šířit vlna.

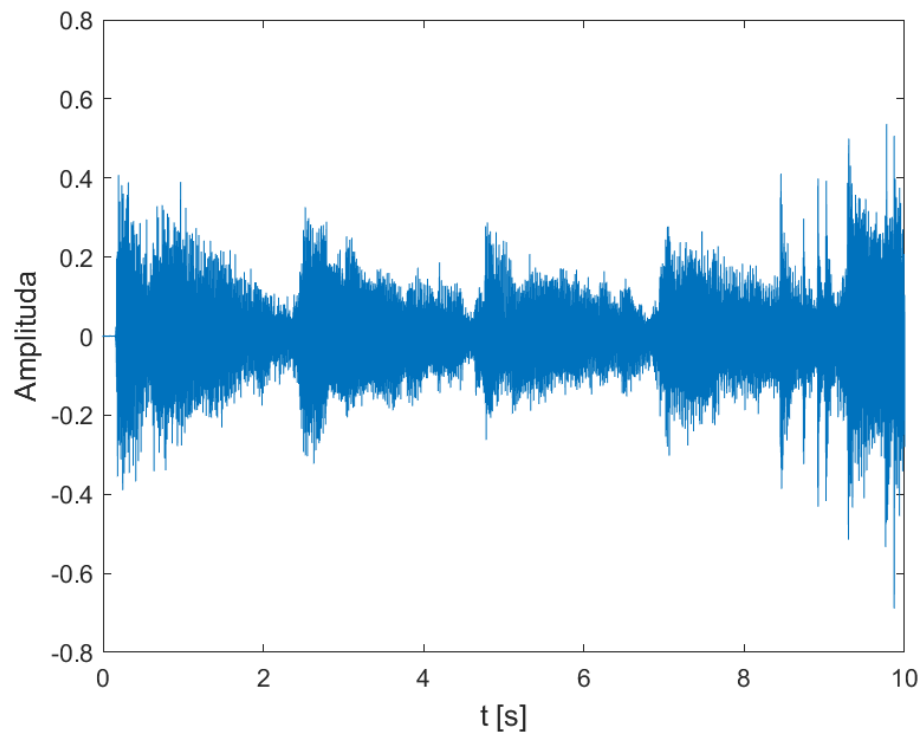
Jednotlivé částice kmitají pouze kolem své rovnovážné polohy. Nedochází tak k přenosu látky ale pouze energie a hybnosti[4]. Popsané kmitání jsme schopni zachytit pomocí akustických měničů. Je získán analogový signál šířících se zvukových vln nazývaný jako audio signál. Pojmem audio je označován řetězec 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 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 také barva [10].

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 popsán rovnicí 1.1,

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

kde f_{vz} je vzorkovací frekvence a f_{max} je maximální frekvence v audio signálu [1]. Pokud jednotlivé vzorky zobrazíme graficky získáme průběh signálu v čase viz obr.1.2.



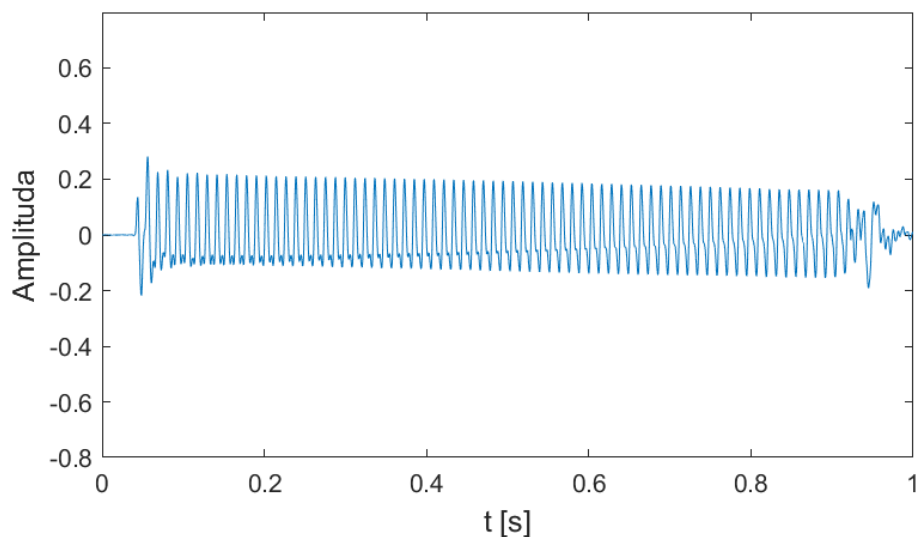
Obr. 1.2: Zobrazení časového průběhu signálu

Tato reprezentace audio signálu poskytuje informace průběhu amplitudy signálu. Využívá například pro výpočet energie signálu popsany v bodě 1.3.1.

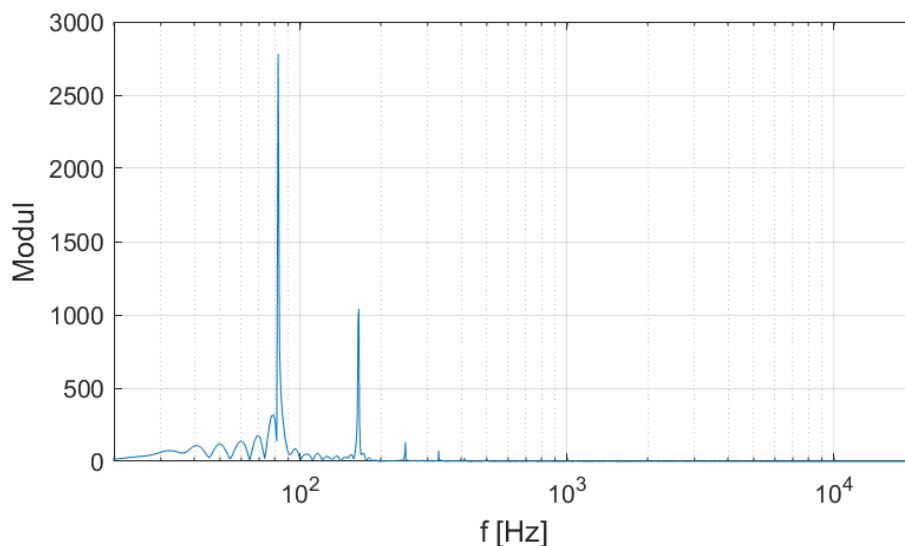
1.2.3 Frekvenční oblast

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 popsány v komplexním tvaru. Spektrum představuje rozložení původní části signálu na jednotlivé frekvenční složky popsané funkcí sinus. Kde reálná složka obsahuje informaci o modulu „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.



(a) Časová oblast



(b) Frekvenční oblast

Obr. 1.3: Reprezentace tónu E zahráného na basovou kytaru.

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 frekvenční 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 se využívá **Fourierovy transformace**. Definována jako transfor-

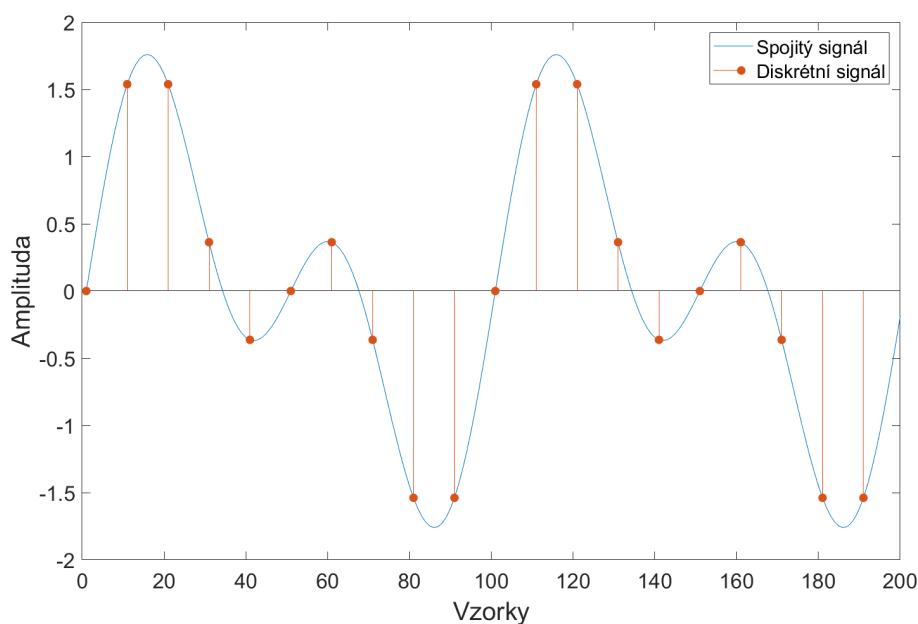
mace převádějící signál mezi časovou a frekvenční oblastí pomocí harmonických signálů „funkce sins a cosinus“.

Hlavní pilířem Fourierovy transformace je, že každý periodický signál je možné rozložit na součet nekonečně mnoha sinusových signálů s různou amplitudou a fází [1]. Toho je poté pomocí matematických postupů docílit. Analyzující signál je rozložen na jeho frekvenční složky udávané amplitudou a fází viz obr. 1.3. Jako další grafické zobrazení frekvenční oblasti hudebního signálu se pro jeho analýzu využívá spektrogramu.

1.2.4 DFT - Diskrétní Fourierova transformace

[12]

Pokud jsou signály zpracovávány pomocí výpočetních procesorů, tak může být uložen pouze omezený počet parametrů 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. Důsledkem toho bylo nutné odvodit algoritmus DFT přizpůsobený právě pro zpracování diskrétních signálů s konečným počtem hodnot.



Obr. 1.4: Časově spojitý signál a diskrétní signál

Rovnice pro DFT je potom zapsána v následujícím tvaru.

$$X(k) = \hat{x}(k/N) = \sum_{n=0}^{N-1} x(n) \exp(-2\pi i k n / N) \quad (1.2)$$

Kde $k \in [0 : M - 1] = [0 : N - 1]$ a $M \in \mathbb{N}$.

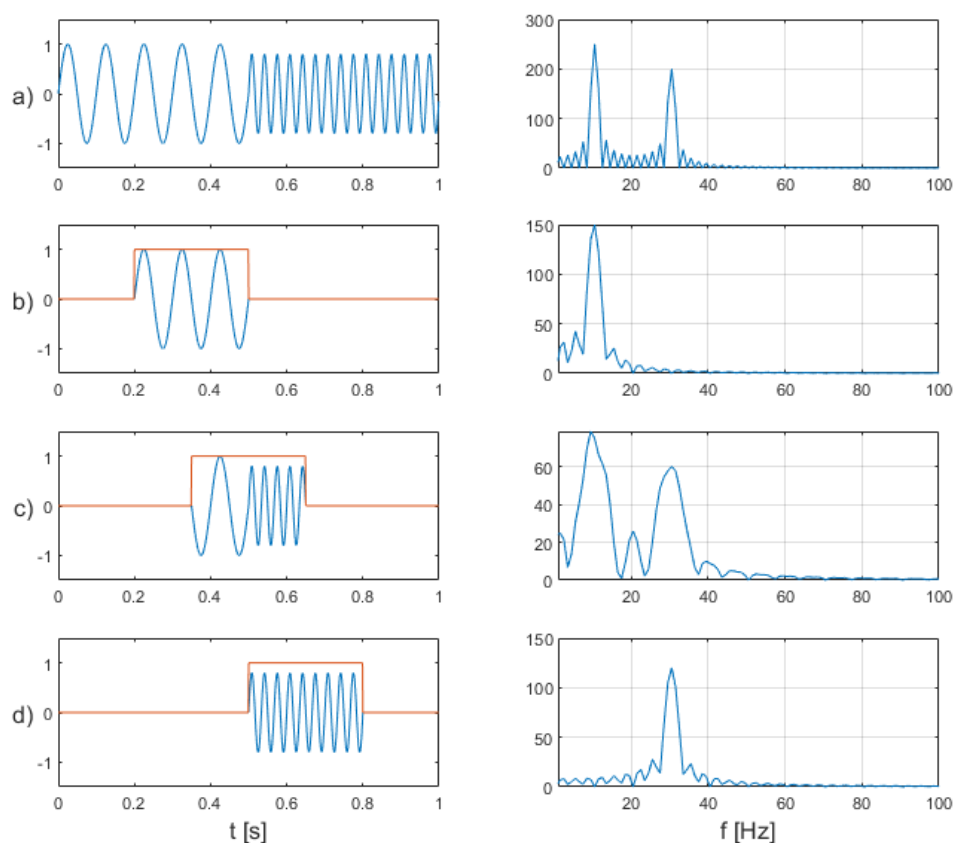
Ze strany výpočetní náročnosti je takto definovaný algoritmus neefektivní a výpočetně náročný. Při počítání Fourierova koeficientu $X(k)$ je zapotřebí velkého množství velké množství operací v řádu N^2 . Proto pokud počet vzorků N dosahuje většího množství je ve většině případů tento algoritmus příliš pomalý pro praktické využití.

Počet potřebných operací může být výrazně redukován použitím efektivního algoritmu známého jako FFT (Fast Fourier transform - Rychlá Fourierova transformace). Na vytvoření FFT se zasloužil Carl Friedrich Gauss a Joseph Fourier zhruba před dvěma sty lety. Vynález tohoto algoritmu změnil své odvětví zpracování signálů a je dnes používán v miliardách telekomunikačních zařízeních. Ačkoliv je ve velké míře využíván v telekomunikacích, tak právě i ve zpracování a analýze zvukových signálů zabírá důležitou roli.

Zjednodušeně FFT využívá redundance napříč sinusovými signály různých frekvencí ke společnému výpočtu všech Fourierových koeficientů pomocí rekurze. Díky tomu je dosaženo snížení výpočetní náročnosti počtu operací z řádu N^2 na $N \log_2 N$. Například při použití vzorků $N = 2^{10} = 1024$. FFT vzžaduje $N \log_2 N = 10240$ 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 .

1.2.5 STFT - Short-time Fourier transform

V roce 1946 Dennis Gabor představil STFT jako potřebu zařazení frekvenčních složek do konkrétního času signálu. Fourierova transformace umožňovala převod signálu z časové oblasti do frekvenční ale nebylo zřejmě v jakém časovém úseku signálu se získané frekvenční složky nachází. Hlavní myšlenkou STFT je, že namísto analyzování celého signálu je signál analyzována pouze jeho malá část. Za tímto účelem je definována tzv. okénková funkce, která je nenulová 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. ???. Chceme-li analyzovat signál v různých časech je tato funkce po signálu posouvána a následně se počítá Fourierova transformace pro každý výsledný okénkový signál.



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

;

Na obr. 1.5 je graficky znázorněna myšlenka STFT, která ukazuje výhodu přesného určení frekvenčních složek signálu v čase. Signál je násoben obdelníkovou okénkovou funkcí ve třech místech. Tyto tři vzniklé signály jsou následně na sebe nezávazně 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 spřesnit 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 obdelníkového 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 a Haming a další. [3];

1.2.6 Dynamika hlasitost a intenzita

V češtině se pojem hlasitost využívá pro reprezentaci subjektivního vnímání akustického tlaku definovanou například jednotkou phon⁴. 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 zahrnou různými muzikanty. Interpretací skladby umělec vytváří dynamiku přednášeného díla [10]. V notovém zápise je dynamika neboli hlasitost přednesu popsána symboly jako jsou například pianissimo „*pp*“, piano „*p*“, forte „*f*“ a další.

V audio signálu je dynamika brána jako hlasitost „loudness“ Jedná se o změny amplitudy signálu nebo jeho efektivní hodnoty RMS⁵ v čase.

Při měření hlasitosti „loudness“ v akustickém prosotru je pak využíváno pojmů **intenzita** zvuku a **akustický výkon**. Kde akustický výkon je definován jako množství energie vyzářené akustickým vysílačem ve vzduchu za jednotku času.[6]. Jednotkou je W . A intenzita zvuku pak je definována jako množství energie, které projde jednotkovou plochou kolmou na směr šíření na jednotku času. Jednotkou pak je Wm^{-2} [15]

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 řádu 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 1.3.

$$L_I = 10 * \log\left(\frac{I}{I_0}\right) \quad (1.3)$$

1.2.7 Barva

Hudebním vyjádření se za slovem barva zkrývá velmi komplexní sdružení atributů. Jedná se jak o psychologický tak hudební problém který je vnímán individuálně. [9]

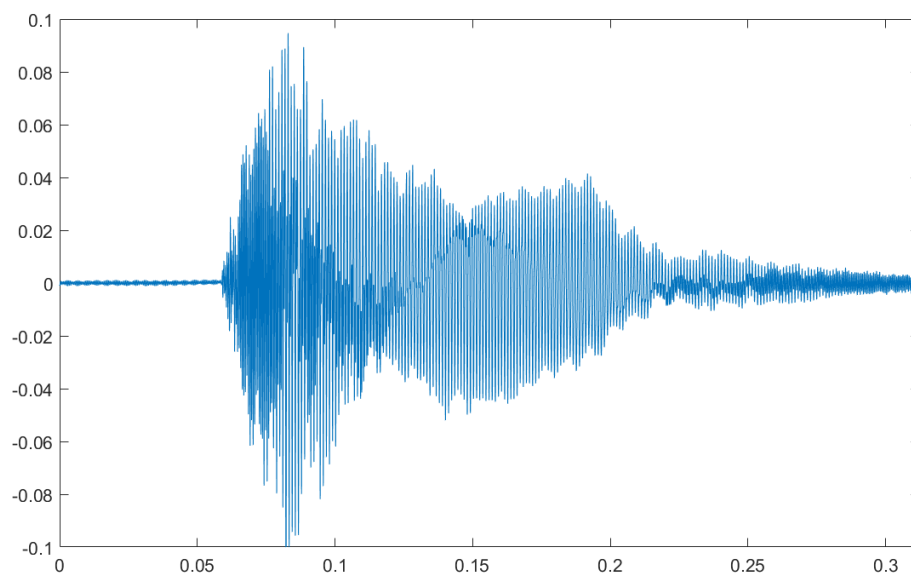
Zdjednodušeně se barva definuje jako vlastnosti díky které je možné rozeznat tón o stejné výšce a hlasitosti zahrný na dva různé náastroje. Díky barvě je posluchač schopen rozeznávat různé zvuky nástrojů a typů interpretace.

⁴Phon - logaritmická jednotka vyjadřující individuální vnímání hlasitosti. Vnímání hlasitosti lidského ucha je závislé na křivce prahu slyšení a může se lišit pro každý tón [13].

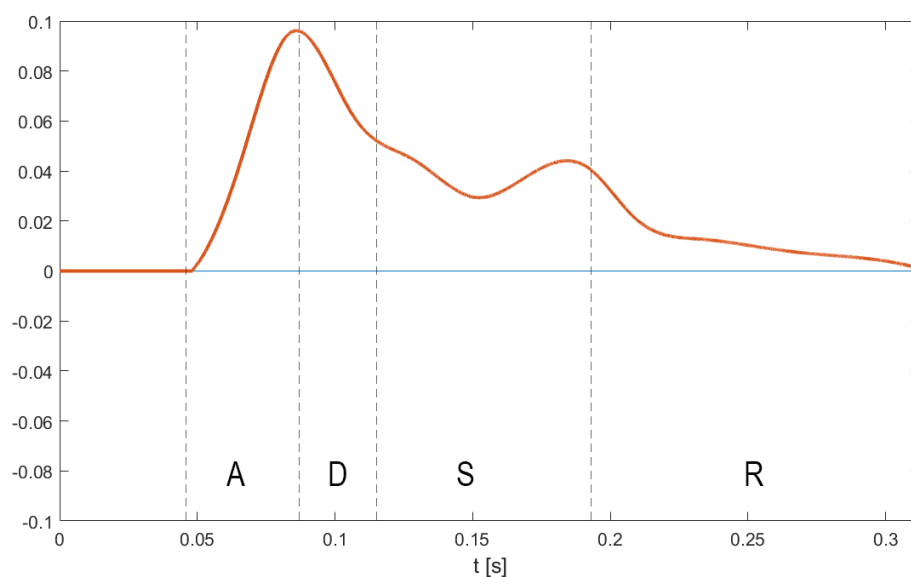
⁵RMS - udává statistickou hodnotu z měření velikosti veličin. Je využívána u periodických veličin[2].

Jelikož je barva špatně kategorizovatelná fyzikálníma veličinama je většinou interpretována nedefinovanými slovy. Například popisujeme barvu jako jasnou, temnou, ostrou, čistou, teplou a další.

Jedním z možných nástrojů pro analýzu barvy tónu je tzv. obálka tónu/signálu. Obálku signálu určuje amplituda signálu v čase viz obr. 1.6 Je rozdělena na 4 fáze a to jsou **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 amplituda 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 úplnému utlumení. Například přiložení tlumítka na rezonující strunu.



(a) Amplituda tónu



(b) Obálka tónu

Obr. 1.6: Tón A5 zahráný na klavír a jeho ADSR obálka

Další informace o barvě signálu se nacházejí v jeho frekvenčním spektru. Tón zahráný na hudební nástroj má svou fundamentální (nosnou) frekvenci nazývanou první harmonická frekvence 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 velký 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 dob a analýza tempa skladby

V této kapitole jsou posány principy detekce tempa skladby a začátků not. S postupem času se techniky používané v MIR pro detekci dob vyvíjejí a vznikají různé přístupy. Níže jsou popsány základní principy pro detekci dob a analýzy tempa až po příchod hubokého strojového učení. Přístup k detekci dob se s vývojem hlubokých neuronových sítí značně změnil. Dnes se v této oblasti využívá zejména struktur strojového učení. [8] [10];

1.3.1 Využití energie signálu

1.3.2 Využití spektra signálu

1.3.3 Detekce periodicity

1.3.4 Využití neuronových sítí

Deep neural networks(DNN)

1.3.5 Více vrstvé perceptronové sítě

1.3.6 Konvoluční neuronové sítě

Temporal Convolutional networks

1.3.7 Rekurentní neuronové sítě

Gated recurrent units Bi-directional models

1.3.8 Hybridní architektury

1.4 Klasifikace žánrů a nálady

1.5 "Získání"chromavektorů

1.6 Systém Spectoda

1.7 Hudební signál jako animace

2 Výsledky studentské práce

Praktická část a výsledky studentské práce vhodně rozdělené do částí.

2.1 Návrh struktury výsledného algoritmu

Závěr

Shrnutí studentské práce.

Literatura

- [1] Bracewell, R.: *The Fourier Transform and its Applications*. Tokyo: McGraw-Hill Kogakusha, Ltd., druh vydání, 1978.
- [2] Cartwright, K.: Determining the effective or RMS voltage of various waveforms without calculus. ročník 8, 01 2007.
- [3] Cohen, L.: Time-frequency distributions-a review. *Proceedings of the IEEE*, ročník 77, . 7, 1989: s. 941–981, doi:10.1109/5.30749.
- [4] Crocker, M.: *Handbook of Acoustics*. A Wiley-Interscience Publication, Wiley, 1998, ISBN 9780471252931.
URL https://books.google.cz/books?id=1x_RvffW-hcC
- [5] 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
- [6] 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, Běžen 2010.
URL <https://www.iso.org/obp/ui/#iso:std:iso:3744:ed-3:v1:en>
- [7] 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.
- [8] 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>
- [9] 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>

- [10] 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>
- [11] Schreibman, S.; Siemens, R.; Unsworth, J. (edito i): *A new companion to Digital Humanities*. West Sussex, England: John Wiley & Sons Ltd, 2016, ISBN 9781118680599.
- [12] 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>
- [13] Tumarkin, A.: *The Decibel, The Phon and the Sone*, ro n k 64. Cambridge University Press, 1950, 178–188 s., doi:10.1017/S0022215100011919.
- [14] Wikipedie: Musical Instrument Digital Interface — Wikipedie: Otevřená encyklopedie. 2022, [Online; navštíveno 1. 12. 2022].
URL https://cs.wikipedia.org/w/index.php?title=Musical_Instrument_Digital_Interface&oldid=21081530
- [15] 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
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

Seznam příloh