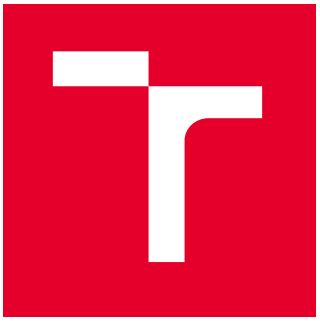
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

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 — **Bc. Viktor Slezák**

AUTHOR

VEDOUCÍ PRÁCE — **Ing. Matěj Ištvánek**

SUPERVISOR

BRNO 2022

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

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 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. 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 světel.

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, Jérome, 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 Ištvánek

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

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, 34 s. Semestrální práce. Vedoucí práce: Ing. Matěj Ištvánek

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 Ištvánek za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Seznam obrázků

Seznam tabulek

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 strukuturu 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 zaobývá vytvořením funkčního systému pro získávání parametrů z hudební nahrávky. Důraz je kladen na využití dostupných moderních metod analýzy hudebních signálu.

1 Teorie

Semestrální práce se zabývá zejména problematikou MIR. Popsánou v kapitole1.1. 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 táto části jsou popsány následující segmenty: Teorie zpracování hudební nahrávky pomocí známých algoritmů. Například důležitým algoritmem je FT[1] popsaná více v bodě 1.2.3, její variantý pak v bodech 1.2.4 a 1.2.5. Nabízené moderní metody strojového učení s využitgím 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 sostředící se na získávání infromací 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 alanýze tempe 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 [13]. 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 tabularueou. 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 sginálů a s tím související možnosti analýzy hudebních nahrávek pomocí počítačů. V těchto letech se poprvé začaly objevoat nyní známé termíny jako „computational musicology“ a „ music information retrieval“ První oblast výzkumu se soustředila na analazýzu tempa skladby. Z důvodu nízké popularity se však výzkum zpomalil. Tento útlum

[1]Fourier transform - Fourierova transformace.

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 jednodužší na realizaci [13].

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 ISMIR[2]. Každoročním vyvrcholením ISMIR je právě váše zmíněná konference, na které vědci z celého světa prezentují pokroky v oblasti výzkumu MIR. Zanedlouho naté v roce 2005 byl v rámic této konference představen model MIREX[3] sloužící jako správa zásad pro hodnocení pokroků ve výzkumu MIR[5].

1.1.2 Řetězec zpracování - pipeline

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

Vstupními date se rozumí zejména hudební informace v digitální potobě. Tyto vstupní data se rozlišují do více typů. Mohou to být obrázky představující digitální formu zápisu hudby pomocí symbolů „not“[13]. Například digitalizovaná partitura. Dalším možným typem je „digitální hudba“. Jedná se o hudbu čistě v „digitálních notách“ představujících sadu příkazů. Například zápis v MIDI[4]. 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 popsaáno 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 převládatkombinace hlubokých neuronových sítí. Tyto kombinace umožňují

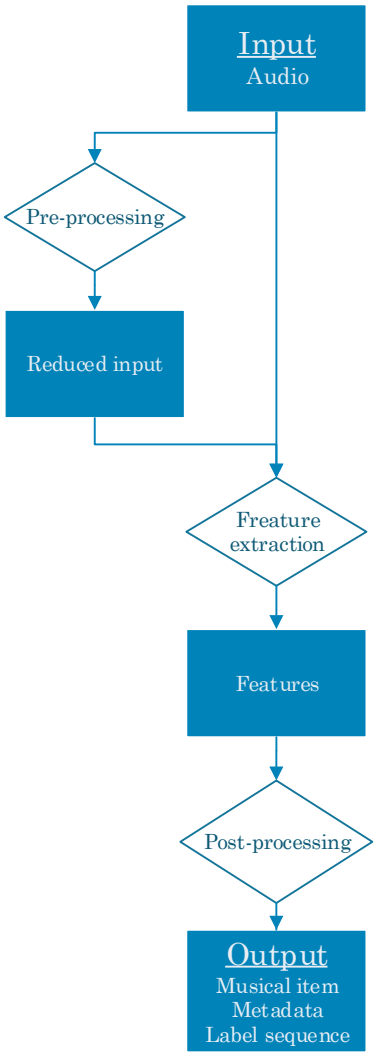
[2]International Society of Music Information Retrieval - Mezinárodní sdružení pro MIR

[3]The Music Information Retrieval Evaluation eXchange - komunitní rámec pro hodnocení pokroků výzkumu v oblasti MIR. Obhospodařovaný laboratočí International Music Information Retrieval Systems Evaluation Laboratory sídlící na University of Illinois. [5].

[4]Musical Instrument Digital Interface - Volně dostupný hudební standart specifikující hardwerové a softwarové požadavky pro digitální přenos hudební notace a komunikace mezi nástroji.[18]

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 ovlivní přesnost zpracování.



Obr. 1.1: Řetězec procesů MIR [13]

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 [13].

Z důvodu velké komplexsnosti vstupních dat se využívá několik technik komprimaca signálů. Slučování vícekanálových nahrávek do mono sginálu. 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	Indetifikace 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. [13]

1.1.3 Současné problémy

1.2 Parametrizace hudebních nahrávek

V této kapitole je popsán audio signál. Jak vzniká, jeho reprezentace v číslicovám 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 interptretaci. 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 [10].

Při pohledu z fyzikálního hlediska důsledkem interpretace díla vznikají zvukové vlny šířící se vzduchem. Tyto vlny jsou reprezentovány kmitáním 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 částit a prostředím se začne šířít vlna.

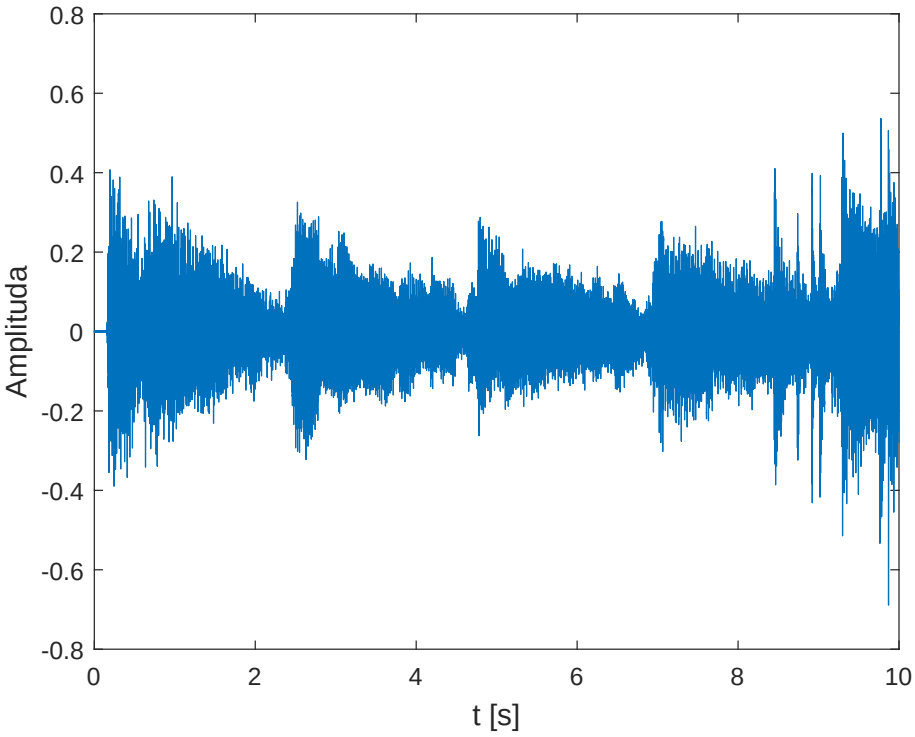
Jednotlivé částice kmitají pouze kolem své rovnovážné polohy. Nedochází tak k přenosu látky ale pouze energie a hybnosti[4]. Popsané kmitání jsme schopní zachytit pomocí akustických měničů. Je získán analogový signál šířících se zvukových vln nazývaný jako audio signál. Pojmem audio je označován řetězec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

$$f_{vz} > 2f_{max} \tag{1.1}$$

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.



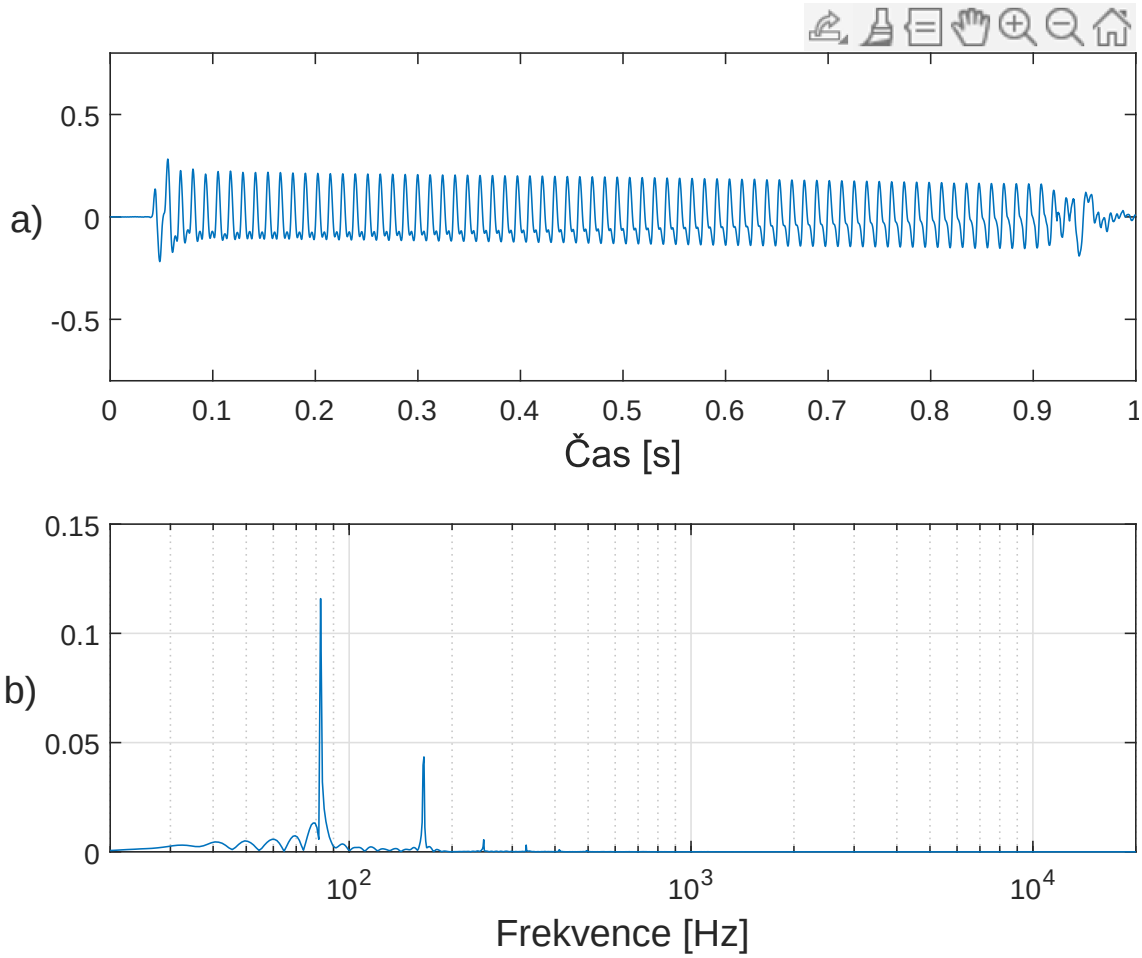
Obr. 1.2: Zobrazení časového průběhu signálu

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

1.2.3 Frekvenční oblast

Pro získání dalalších 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 magnitudu „velikosti“ funkce sinus. Imaginární složka komplexního čísla pak udává počáteční fázi. V grafu jsou poté zobrazy frekvenční složky se kterých se signál skládá viz obr. 1.3.



Obr. 1.3: Reprezentace tónu E zahraného 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 složýitý proces. Díky transformaci do frekveční oblasti je patrná fundementální 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 vyžívá zejména Diskrétní Fourierovy transformace popsané v bodě 1.2.4.

Fourierovy transformace zkráceně definována jako transformace převádějící signál mezi časovou a fekvenční oblastní poomcí harmonických signálů jež popisují

funkce sinus a cosinus.[1] Funkce sinus a cosinus představují komplexní exponenciály.

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

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

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

kde $\omega = 2\pi f$ a udává uhlovou frekvenci. Magnituda $|X(f)|$ je potom funkcí sudou [14].

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

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

kde $f_k = k \times f_0$ a $k \in (\mathbb{Z}; 0, \pm1, \pm2, \dots)$. Vypočítané spektrum je diskrétní a neperiodické [14].

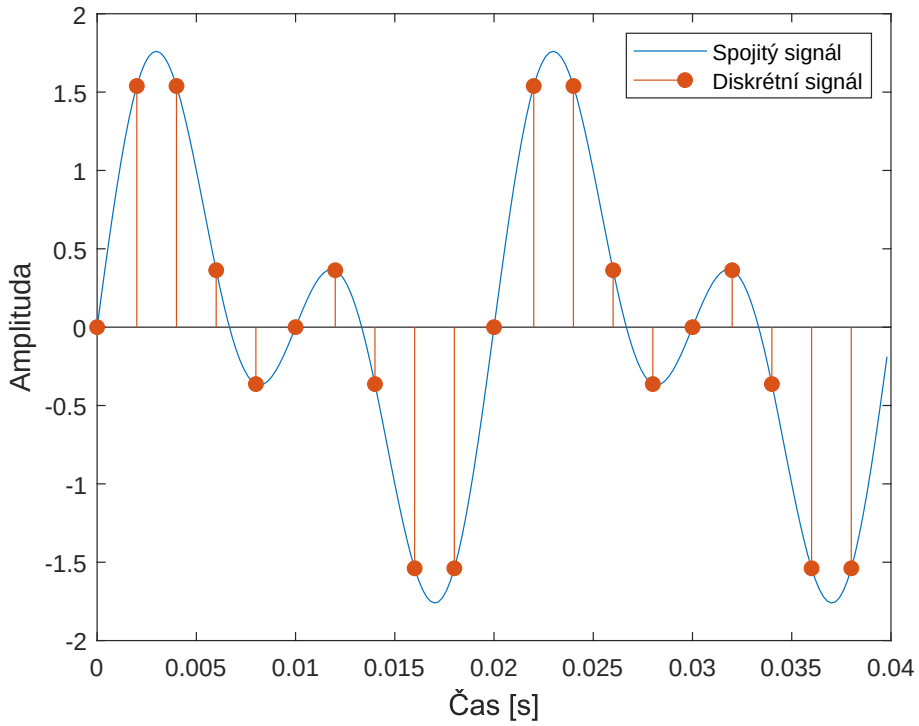
Pokud je vstupní signál diskrétní hovoříme o Diskrétní Fourierově transformaci která je více popsána v následujícím bodě 1.2.4.

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

1.2.4 DFT - Diskrétní Fourierova transformace

[15] [1]

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 pospsány algoritmus DFT přizpůsobené právě pro zpracování diskrétních signály nespojité v čase.



Obr. 1.4: Časově spojitý signál a diskrétní signál

Opět jsou dány dolišné 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

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

kde

$$\Omega = 2\pi(f/f_s) \tag{1.6}$$

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

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

kde

$$\Omega_k = 2\pi\frac{f_k}{f_s} \tag{1.8}$$

$$f_k = \frac{kf_s}{N_0} \tag{1.9}$$

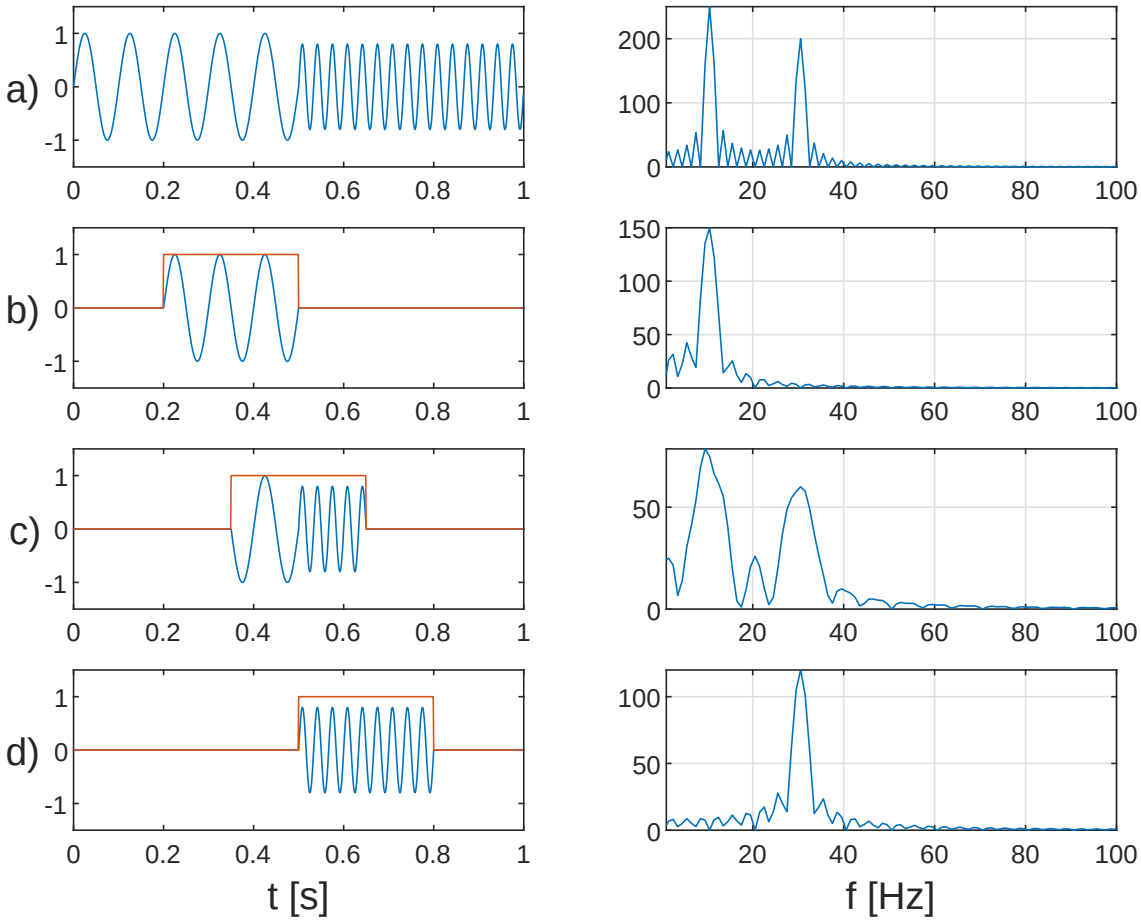
a $k \in (\mathbb{Z}; 0, N_0 - 1)$. N_0 Počet vrzorků v jedné periodě signálu. Hustota spektra K v takovém případě odpovídá $K = N_0$.

Ze strany výpočetní náročnosti je takto definovyný algoritmus neefektivní a výpočetně náročný. Pro výpočet DFT je zapotřebí velkého množství operací složitost algoritmu je pak zapsána jako $O(N^2)$. 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. Jeho složitost pro výpočet byla snížena na $O(N \log_2 N)$ [10]. 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. 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 [10].

1.2.5 STFT - Short-time Fourier transform

V roce 1946 Dennis Gabor představil STFT jako možnost zařazení frekvenčních složek do konkrétnímo času signálu [15]. Fourierova transformace umožňovala převod signálu z časové oblasti do frekvenční ale nebylo zřejme v jakém časovém úseku signálu se získané fekvenční složky nachází. Hlavní myšlenkou STFT je, že namísto analyzování celého signálu je analyzována pouze jeho malá část. Za tímto účelem je definována tzv. okénokvá 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 $1s$ 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 $0,8s$ [10]

Na obr. 1.5 je graficky znázorněna myšlenka STFT, která ukazuje princip určování frekvenčních složek v čase a jejich výhzody. 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 poomcí 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 a Haming a další [3].

1.2.6 Dynamika hlasitost a intenzita

V češtině se pojem hlasitost využívá pro reprezentaci subjektivního vnímání akustickho tlaku definovanou například jednotkou phon[5]. 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 [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 apmlitudy signálu nebo jeho efektivní hodnoty RMS[6] v čase.

Při měření hlasitosi „loudness“ v akustickém prosotru je pak využíváno pojmů **intenzita** zvuku a **akustický výkon**. Kde akustický výkon je definovánjako množství energie vyzářené akustickým vysílačem ve vzduchu za jednotku čausu.[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} [19]

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 zuvku v decibelech dB. Kde vztažnou hodnotou je práh slyšení $I_0 = 10^{-12}$ Wm^{-2}. Hladina intenzity se vypočítá dle rovnice 1.10.

$$L_I = 10log(\frac{I}{I_0}) \tag{1.10}$$

1.2.7 Barva

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

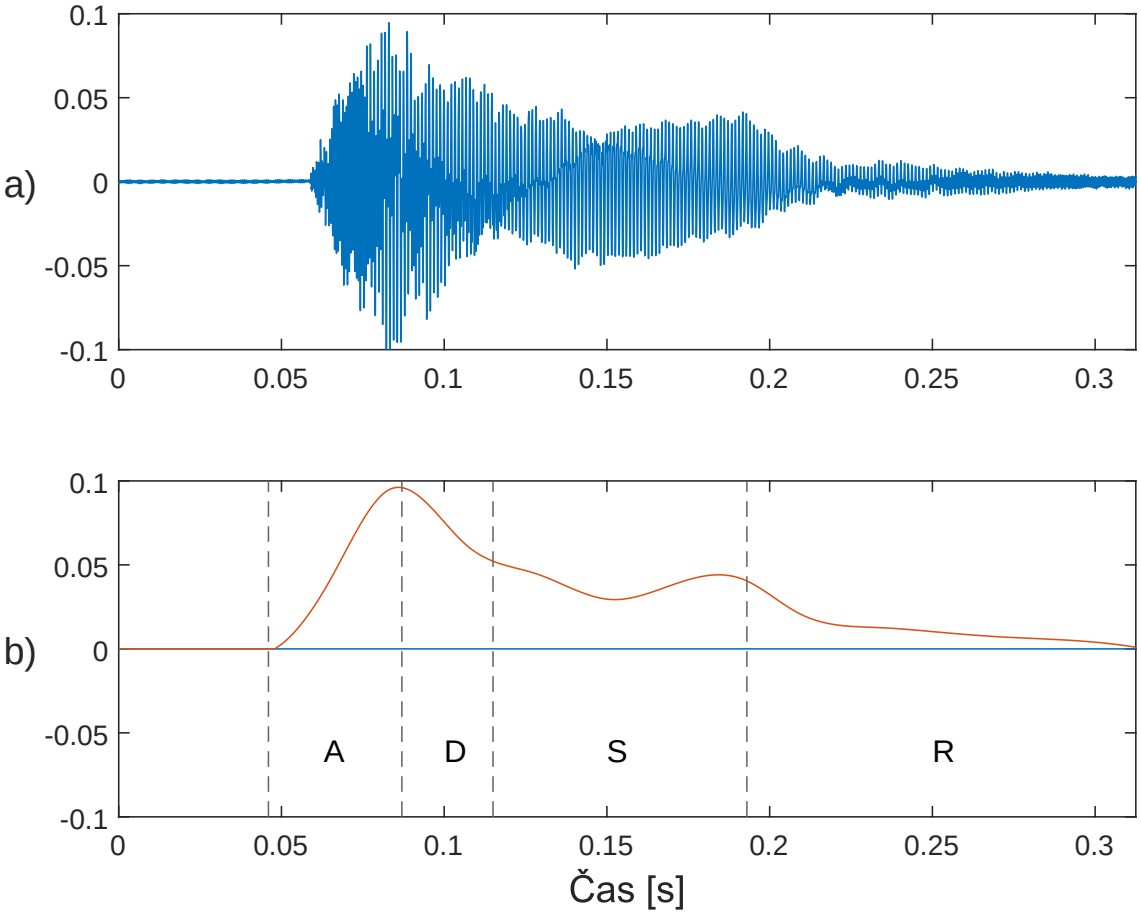
Zdjednodušeně se barva definuje jako vlastnosti, díky kterým je možné rozeznat tón o stejné výšce a hlasitosti zahraný na dva různé náastroje[11]. Díky barvě je posluchač schopen rozeznávat odlišné zvuky nástrojů a typů interpretace.

[5]Phon - logaritmická jednotka vyjadřující individuální vnímání hlasitosti. Vnímání hlasitosti lidského ucha je závislé na křivce prahu slyšení a může se lišit pro každý tón [17].

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

Z důvodu špatné kategorizace barvy fyzikálními veličinami je většinou interpretována přídavnými jmény. Například je barva popisována jako jasná, temná, ostrá, čistá, teplá, pestrá, nazální a další.

Jedním z možných nástrojů pro analýzu barvy tónu je tvz. obálka tónu/signálu. Obálku určuje amplituda signálu v čase viz obr. 1.6 Je rozdělena na 4 fáze popsané z knihy Fundamentals of Music Processing [10]. **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 kterám dochází k poklesu hlasitosti zdroje zvuku až k uplnému utlumení Například přiložení tlumítka na rezonující strunu.



Obr. 1.6: Tón A5 zahraný na klavír **a)** Amplituda tónu **b)** ADSR obálka tónu

Další informace o barvě signálu se nacházejí v jeho frekvenčním spektru. Tón zahraný 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 využiváané v MIR pro analazýzu tempa skladby. V této kapitole je popsán postupný vývoj algoritmů. Od nejjednodužších přístupů po komplexní řešení využívající moderní metody hlubokých neuronových sítí.

Detekce nástupů lze chápat jako detekci začátků not či dalších hudebních událostí, které se vyskytnout v průběhu skaldby. Výskyt takových udáslotí je provázen zvýšením energie signálu zaznamenané ve fázi nástupu dle ADSR obálky popsané v bodě 1.2.7. Typickým znakem pro fázi nátupu 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 závisloti na typu nástroje, technice hry [16]. 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 vybrovat a víbrace 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 [10].

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 nehraje pouze jeden nástroj. Jedná se napříkladod polyfoní skladbu. Polyfoní skladoubou rozumíme skladbu tvořenou více hlasy. Zároveň nemají určenou roli vedoucího a doprovodného hlasu [12]. Takové uspořádání hlasů může vést k překrývání a nástupy mohou být maskovány.

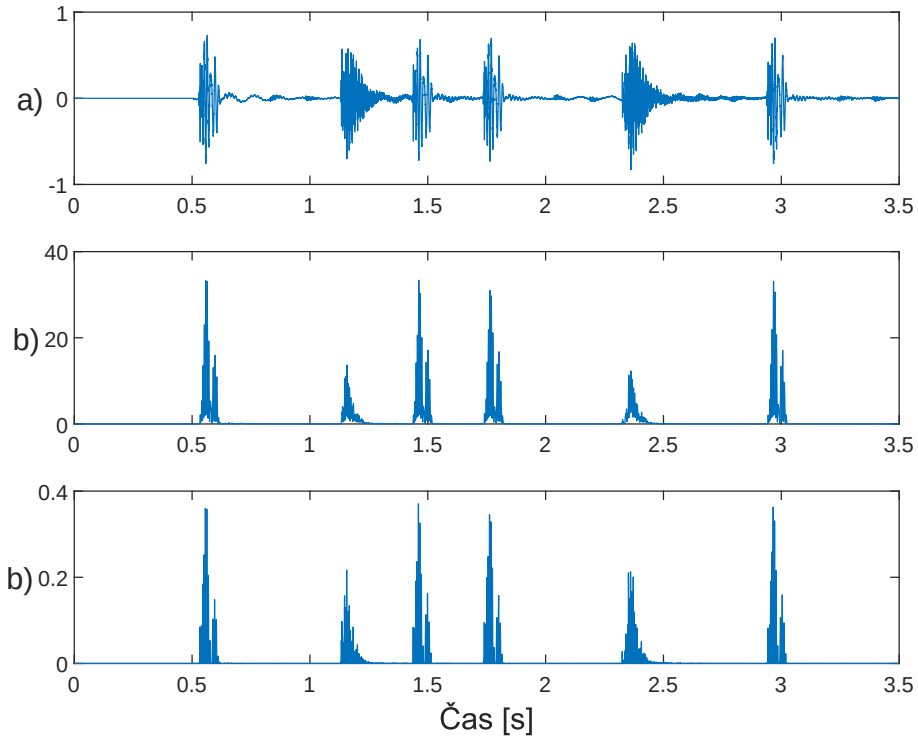
Díky jevu maskování je obtížné zaznamenat změny energie signálu. V tomto případě přichází potřeba po komplexnějších metodát detekce nástupů [10]. Například pohled na krátkodobé změny ve spektru signálu pomocí využití STFT popsané v bodě 1.2.5.

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 detekovat nástup tónu pomocí funkce pro výpočet energie signálu v daném místě. Náhle změny signálu v takto definované funkci ukazují potencionální místa nástupů. Matematicky pak funkci popisujeme. Nechť $x(n)$ je diskrétní signál. 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 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

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

pro $n \in \mathbb{Z}$ [10].



Obr. 1.7: Detkte nástupů perikusivního zvuku**a)** Amplituda nahrávky **b)** Lokální energie signálu $E_{xw}(n)$ **c)** Derivace lokální enegrie 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 nikooiv její pokles jsou ponechány pouze pozitivní rozdíly a negativní jsou zapsány jako nulové. V anglické literatuře je tneto proces známý také jako půlvlnné usměrnění „half-wave rectification“. Vzorece jsou zapsány následovně

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

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

kde $r \in \mathbb{R}$ a $n \in \mathbb{Z}$. Na obrázku 1.7 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 rozeznat strukturu nahrávky a zmírnit efekt maskování

který nastává při metodách založených energii signálu popsané v bodě 1.3.1. V polyfoní hudbě mohou interpretace o nižší hlasistosti bý maskovány hlasitějšími projevy. Kdy například jeden nástroj v tónové fázi podržení ADSR obálky, popsané v bodě 1.2.7, může být hlasitější než než fáza náběhu druhého nástroje. Takový nástup je pak maskován a je obtížné detekovat jeho nástup. 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 detekovat pomocí spektra signálu. Každý hudební nástroj obsazuje jiné frekvenční spektrum [10]. Díky tomu se různé nástroje nachází na odlišných místech spektra, jak spektrum daného nástroje vypadá určuje také barvu jeho barvu popsanou v bodě 1.2.7. S takovým jevem je důležité počítat při detekci nástupů založené na spektru signálu.

[8]]

TODO: popsat nějakou metodu s knížky fundamental...

1.3.3 Detekce periodicity

1.3.4 Vzužití 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 recurent 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 vyd n , 1978.

[2] Cartwright, K.: Determining the effective or RMS voltage or various waveforms without calculus. ro n k 8, 01 2007.

[3] 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.

[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, B ezen 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 B ock: *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] 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

[12] 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.

[13] Schreibman, S.; Siemens, R.; Unsworth, J. (edito i): *A new companion to Digital Humanities*. West Sussex, England: John Wiley & Sons Ltd, 2016, ISBN 9781118680599.

[14] Sneddon, I.: *Fourier Transforms*. Dover books on mathematics, Dover Publications, 1995, ISBN 9780486685229.
URL https://books.google.cz/books?id=jhpsLpRRerwC

[15] 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

[16] Syrovỳ, V.: *Hudební akustika*. Akustická knihovna Zvukového studia Hudební fakulty AMU, Akademie múzickỳch umění, 2013, ISBN 9788073312978.
URL https://books.google.cz/books?id=ikrmoAEACAAJ

[17] Tumarkin, A.: *The Decibel, The Phon and the Sone*, ro n k 64. Cambridge University Press, 1950, 178–188 s., doi:10.1017/S0022215100011919.

[18] 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

[19] 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 Frourierova transformace

RMS Root mean square - efektivní hodnota

Seznam příloh