

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 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, 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, ?? 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 Ištvanék 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!** (**MIR!**). 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 zaobývá 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???. 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 **FT!**<sup>1</sup> popsána více v bodě ??, její varianty pak v bodech ?? a ??. Nabízené moderní metody strojového učení s využitím hlubokých neuronových sítí při detekci tempa skladby ?? a určení žánru ??. Struktura a možnosti systému Spectoda pro generování interaktivních světelných animací je podrobně popsána v bodě ??.

## 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 [?]. **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 laboratorů 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

---

<sup>1</sup>**FT!**.

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 [?].

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!**<sup>2</sup>. 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 **MIREX!**<sup>3</sup> sloužící jako správa zásad pro hodnocení pokroků ve výzkumu **MIR!**[?].

### 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“ [?]. Například digitalizovaná partitura. Dalším možným typem je „digitální hudba“. Jedná se o hudbu čistě v „digitálních notách“ představujících sadu příkazů. Například zápis v **MIDI!**<sup>4</sup>. 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 ??.

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

---

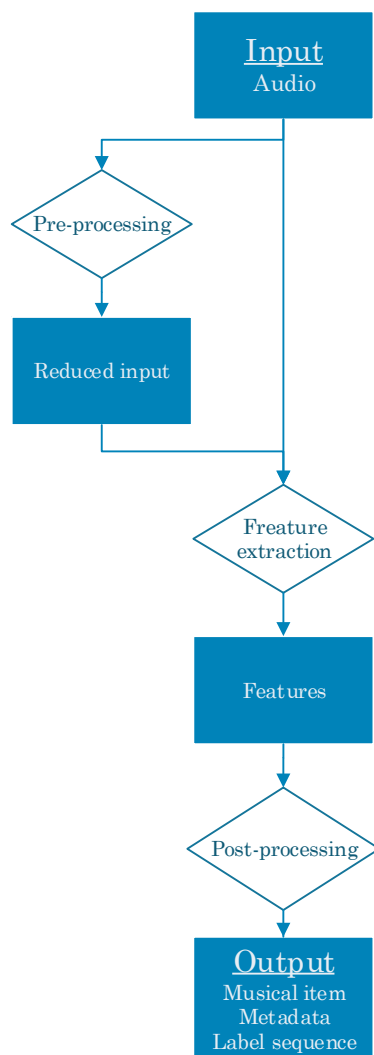
<sup>2</sup>International Society of Music Information Retrieval - Mezinárodní sdružení pro **MIR!**

<sup>3</sup>The 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. [?].

<sup>4</sup>Musical Instrument Digital Interface - Volně dostupný hudební standart specifikující hardwarové a softwarové požadavky pro digitální přenos hudební notace a komunikace mezi nástroji.[?]

přesnější parametrizaci a menší chybovost.

**Post-processing - konečné zpracování** Posledním blokem v řetězci je tzv. „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 [?]

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 [?].

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[?]. 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. [?]

### 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 ?? a ?? 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 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 [?].

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.

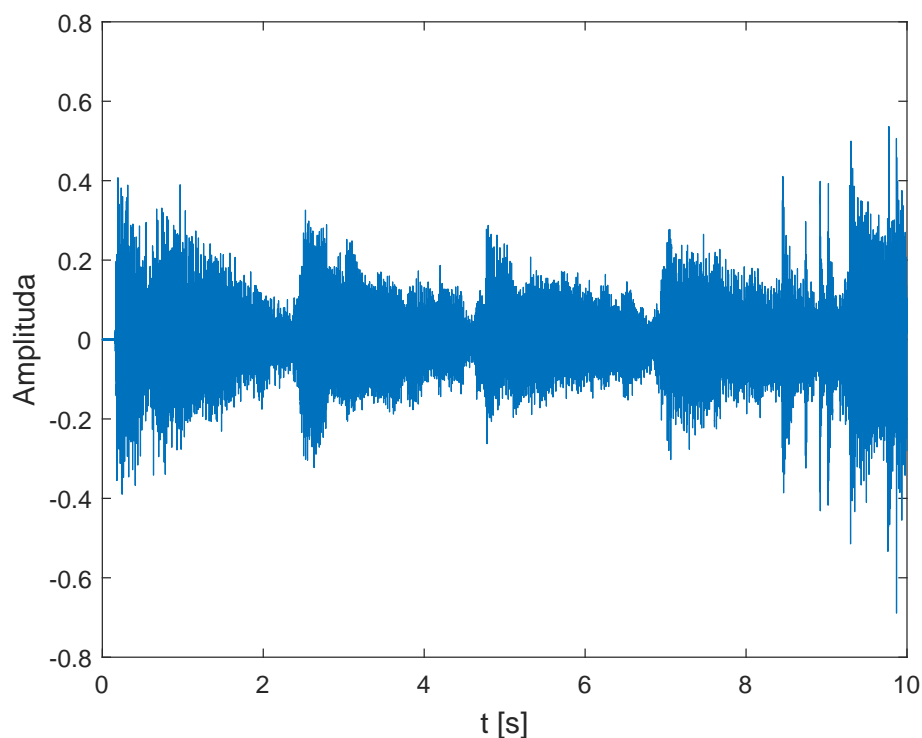
Jednotlivé částice kmitají pouze kolem své rovnovážné polohy. Nedomáhají se tak k přenosu látky ale pouze energie a hybnosti[?]. Popsané kmitání jsme schopni zachytit pomocí akustických měničů. Je získán analogový signál šířící se zvukovými vlnami nazývaný jako audio signál. Pojem audio je označován 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 [?].

### 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 ??,

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

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



Obr. 1.2: Zobrazení časového průběhu signálu

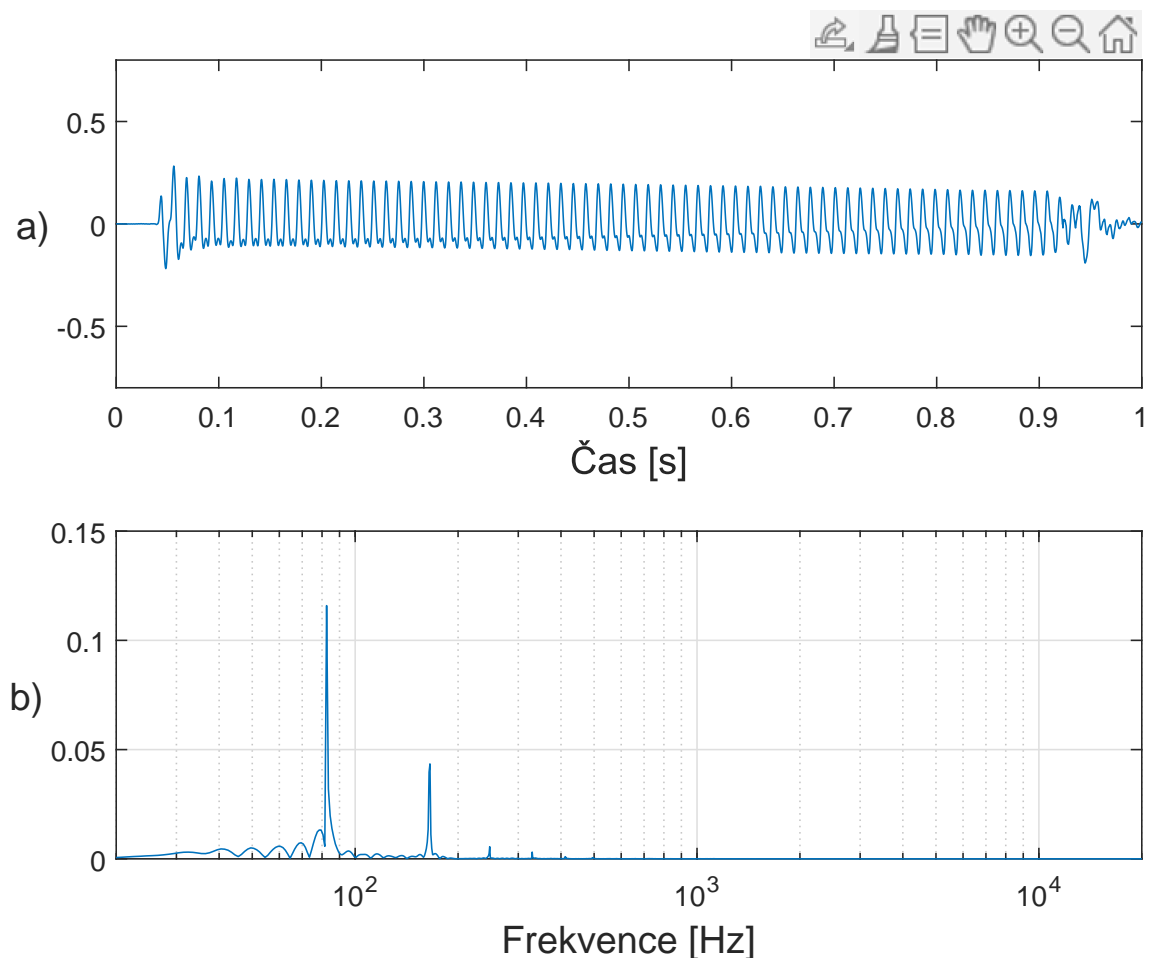
Tato reprezentace audio signálu poskytuje informace průběhu amplitudy signálu. Využívá například pro výpočet energie signálu popsany v bodě ??.

### 1.2.3 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 magnitudu „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. ??.





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 ???. 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 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 ??.

**Fourierovy transformace** zkráceně definována jako transformace převádějící signál mezi časovou a frekvenční oblastí pomocí harmonických signálů jež popisují

funkce sinus a cosinus.[?] 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 [?].

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 neperiodické [?].

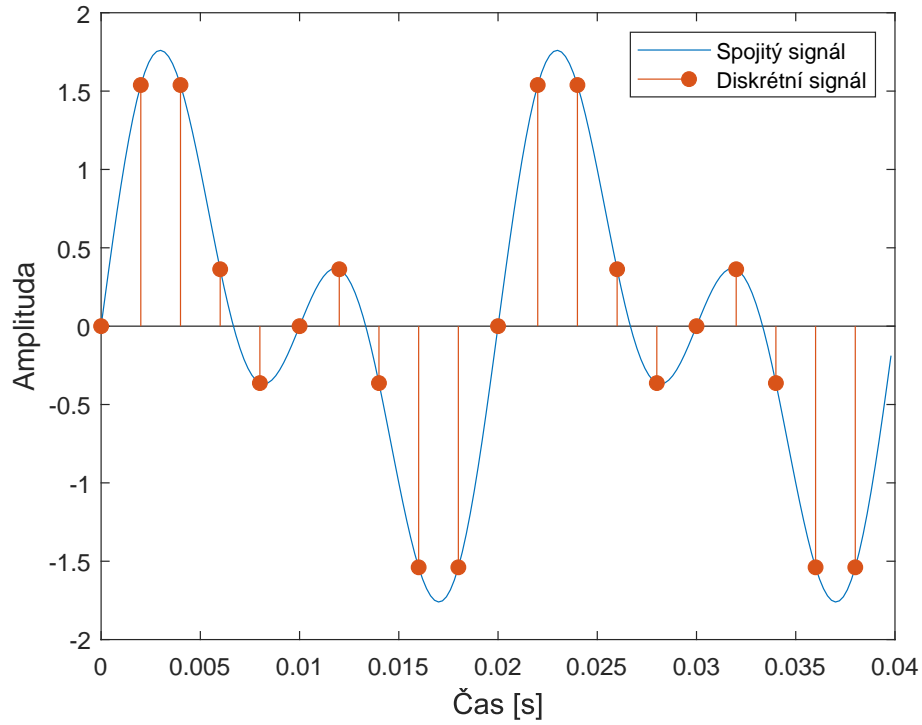
Pokud je vstupní signál diskrétní hovoříme o Diskrétní Fourierově transformaci která je více popsána v následujícím bodě ??.

Po transformaci signálu do frekvenčního spektra jsou data signálu v komplexním tvaru 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

[?] [?]

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 tzv. signál diskrétní, který je není spojitý v čase. Diskrétní signál je potom vhodný pro číslicové zpracování. Proto jsou pospšány algoritmus **DFT!** přizpůsobené právě pro zpracování diskrétních signálů nespojitě 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

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

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

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

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

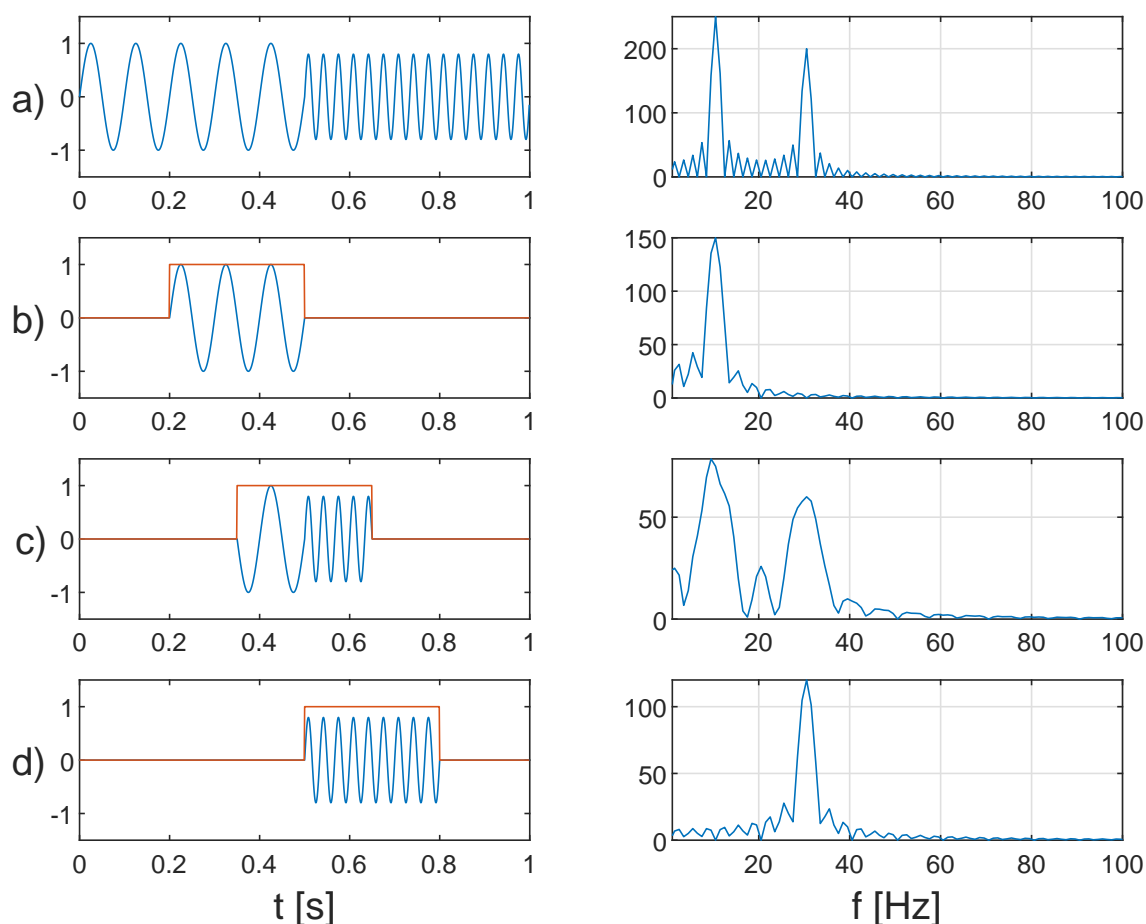
a  $k \in (\mathbb{Z}; 0, N_0 - 1)$ .  $N_0$  Počet vzorků v jedné periodě signálu. Hustota spektra  $K$  v takovém případě odpovídá  $K = N_0$ .

Ze strany výpočetní náročnosti je takto definovaný 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  $O(N \log_2 N)$  [?]. 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 [?].

### 1.2.5 STFT - Short-time Fourier transform

V roce 1946 Dennis Gabor představil **STFT!** jako možnost 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 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 **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 [?]

Na obr. ?? je graficky znázorněna myšlenka **STFT**!, která ukazuje princip určování frekvenčních složek v čase a jejich výhody. 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. ?? 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 [?].

## 1.2.6 Dynamika hlasitost a intenzita

V češtině se pojem hlasitost využívá pro reprezentaci subjektivního vnímání akustického tlaku definovanou například jednotkou phon<sup>5</sup>. 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 [?]. 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 **RMS**<sup>6</sup> 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. [?]. 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  $Wm^{-2}$  [?]

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 ??.

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

## 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 [?].

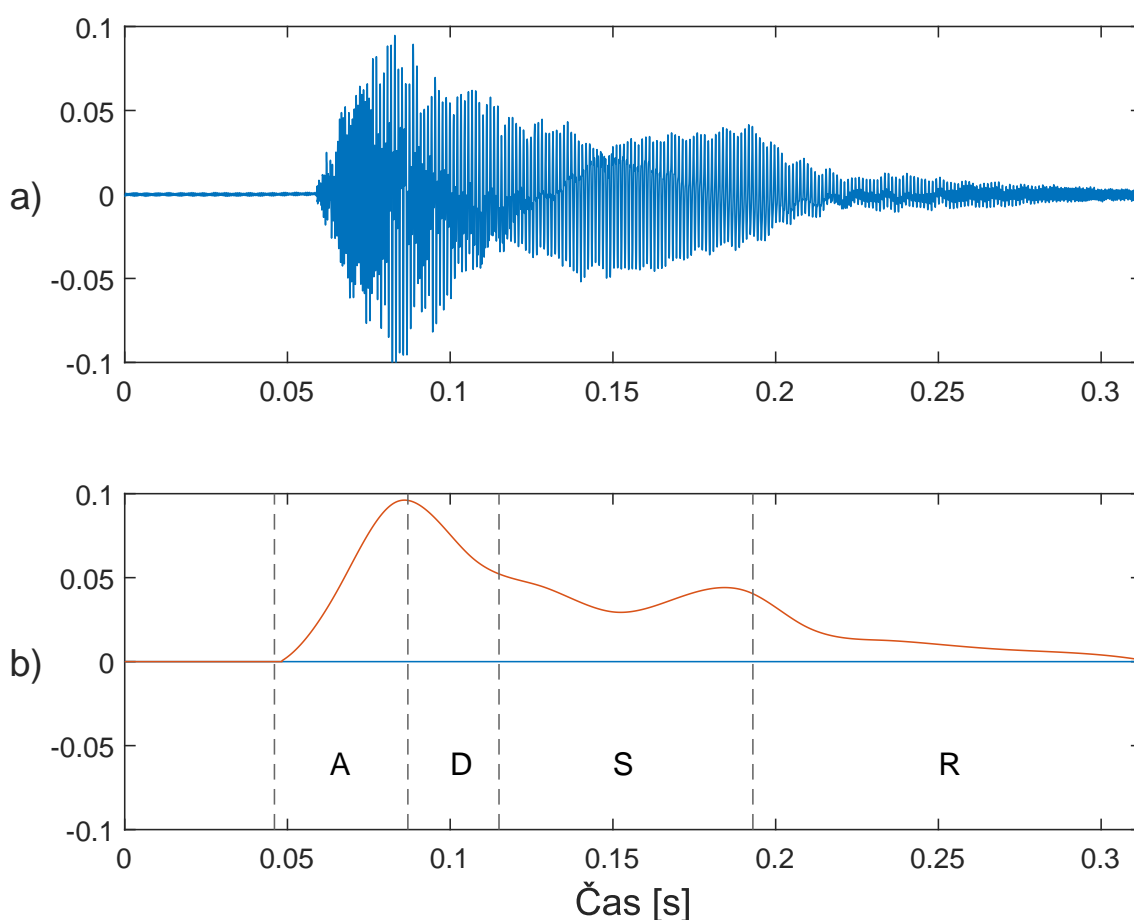
Zdjednodušeně se barva definuje jako vlastnosti, díky kterým je možné rozeznat tón o stejné výšce a hlasitosti zahráný na dva různé nástroje [?]. Díky barvě je posluchač schopen rozeznávat odlišné zvuky nástrojů a typů interpretace.

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

<sup>6</sup>RMS - udává statistickou hodnotu z měření velikosti veličin. Je využívána u periodických veličin [?].

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

Jedním z možných nástrojů pro analýzu barvy tónu je tzv. obálka tónu/signálu. Obálku určuje amplituda signálu v čase viz obr. ?? Je rozdělena na 4 fáze popsané z knihy Fundamentals of Music Processing [?]. **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.



Obr. 1.6: Tón A5 zahráný na klavír **a)** Amplituda 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 hlubokých neuronových sítí.

Detekce nástupů lze chápat jako detekci začátků not či dalších hudebních událostí, které se vyskytnout v průběhu skladby. Výskyt takových událostí je provázen zvýšením energie signálu zaznamenané ve fázi nástupu dle ADSR obálky popsané v bodě ???. 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, technice hry [?]. 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 [?].

U některých nástrojů je však energie přenášena 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 například o polyfoní skladbu. Polyfoní skladbou rozumíme skladbu tvořenou více hlasy. Zároveň nemají určenou roli vedoucího a doprovodného hlasu [?]. 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 [?]. Například pohled na krátkodobé změny ve spektru signálu pomocí využití **STFT**! popsané v bodě ??.

### 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. Na 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. Necht  $x(n)$  je diskretní signál. Stejně jako u **STFT**! popsané v bodě ?? je definována okénková funkce  $w(n)$  ve tvaru zvonu „bell-shaped function“ která je posouvána 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

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

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

Pro názornější zobrazení se následně vypočítá derivace funkce lokální energie. V případě diskretního signálu se derivace realizuje jako rozdíl dvou po sobě jdoucích vzorků. Protože pro detekci nástupů je důležitá zejména pozitivní změna energie nikoli 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“. Vzorec 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  $n \in \mathbb{Z}$ .

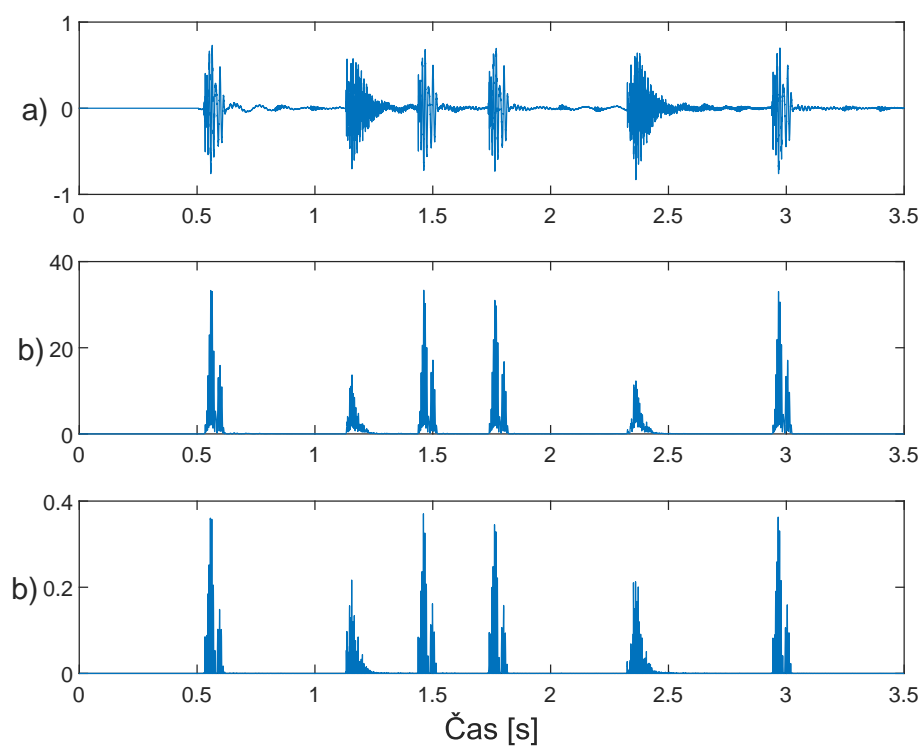
### 1.3.2 Využití spektra signálu

[?]

### 1.3.3 Detekce periodicity

### 1.3.4 Využití neuronových sítí

Deep neural networks(DNN)



Obr. 1.7: Detkce nástupů perkusivního zvuku **a)** Amplituda nahrávky **b)** Lokální energie signálu  $E_{xw}(n)$  **c)** Derivace lokální energie signálu s půlvlnným usměrněním  $\Delta_E(n)$

### **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.

## Seznam symbolů a zkratek

<b>MIR</b>	Music information retrieval - Obor zabývající se vyhledávání informací v hudebních dílech
<b>MIDI</b>	Musical Instrument Digital Interface - Digitální rozhraní hudebních nástrojů
<b>ISMIR</b>	International Society of Music Information Retrieval - Mezinárodní združení pro <b>MIR</b> !
<b>MIREX</b>	The Music Information Retrieval Evaluation eXchange
<b>FT</b>	Fourier transform - Fourierova transformace
<b>FFT</b>	Fast Fourier transform - Rychlá Fourierova transformace
<b>DFT</b>	Discrete Fourier transform - diskrétní Fourierova transformace
<b>STFT</b>	Short-time Fourier transform - krátkodobá Fourierova transformace
<b>RMS</b>	Root mean square - efektivní hodnota

## Seznam příloh