

**SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
STAVEBNÁ FAKULTA**

POČÍTAČOVÉ SPRACOVANIE HUDBY: BEAT TRACKING

Bakalárska práca

SvF-5342-98404

**SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
STAVEBNÁ FAKULTA**

POČÍTAČOVÉ SPRACOVANIE HUDBY: BEAT TRACKING

Bakalárska práca

SvF-5342-98404

Študijný program: Matematicko-počítačové modelovanie
Študijný odbor: 9.1.9. aplikovaná matematika
Školiace pracovisko: Katedra matematiky a deskriptívnej geometrie
Vedúci záverečnej práce: Ing. Mgr. Lukáš Tomek, PhD.

Bratislava, 2021

Eva Bencová



ZADANIE BAKALÁRSKEJ PRÁCE

Študentka: **Eva Bencová**
ID študenta: 98404
Študijný program: matematicko-počítačové modelovanie
Študijný odbor: matematika
Vedúci práce: Ing. Mgr. Lukáš Tomek, PhD.

Názov práce: **Počítačové spracovanie hudby: Beat tracking**

Jazyk, v ktorom sa práca vypracuje: slovenský jazyk

Špecifikácia zadania:

Keď človek počuje dobrú hudbu, často si podvedome kýve hlavou alebo dupe nohou do rytmu. Ako toto naučiť počítač?

V bakalárskej práci je úlohou najprv detailne pochopiť, čo je to zvuk a ako je zaznamenaný v digitálnych zvukových nahrávkach. Následne si treba naštudovať základné algoritmy, ktoré sa na beat tracking používajú, a potom ich implementovať v softvéri Mathematica.

Algoritmy sa najskôr použijú na nájdenie beatov v rockových a popových skladbách, v ktorých je to obvykle jednoduchšie: tempo skladby je pravidelné a každý beat (hudobná doba) začína v tom časovom okamihu, keď je najväčšia intenzita zvuku (napríklad úder do bicích, brnknutie na basgitaru). Ďalej sa bude skúmať funkčnosť algoritmov na náročnejších skladbách, napríklad z folklórnej alebo klasickej hudby. Úloha je tu náročnejšia: tempo skladby sa často mení (spojito alebo aj náhle), rytmus udávajú sláčikové alebo dychové nástroje, pri ktorých je priebeh intenzity zvuku komplikovanejší (najväčšia intenzita zvuku nemusí byť na začiatku tónu).

Algoritmy beat trackingu majú mnoho aplikácií, napríklad synchronizácia svetelných efektov s hudbou, mixovanie hudby, synchronizácia skladieb pri DJ-ingu, klasifikácia hudby, či automatické odporúčanie hudby (Spotify).

Riešenie zadania práce od: 08. 01. 2021

Dátum odovzdania práce: 06. 05. 2021

Eva Bencová
študentka

Ing. Marek Macák, PhD.
vedúci pracoviska

prof. RNDr. Karol Mikula, DrSc.
garant študijného programu

Čestné prehlásenie

Vyhlasujem, že predloženú záverečnú prácu som vypracovala samostatne pod vedením vedúceho záverečnej práce, s použitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú citované v práci a uvedené v zozname použitej literatúry. Ako autor záverečnej práce ďalej prehlasujem, že som v súvislosti s jej vytvorením neporušil autorské práva tretích osôb.

Bratislava, 6. mája 2021

Názov práce: Počítačové spracovanie hudby: Beat tracking

Kľúčové slová: zvuk, zvuková nahrávka, frekvencia, Fourierova transformácia, tempo

Abstrakt: Človek pri počúvaní hudby podvedome dupe nohou do rytmu piesne, čím dokáže automaticky určiť jej tempo. V bakalárskej práci sa bude zaoberať tým, ako počítačovým spracovaním zvuku dosiahnuť rovnaký výsledok, a teda nájsť údery (beaty) v danej piesni. Začneme s podrobným vysvetlením hlavných pojmov a algoritmov, ktoré sa na beat tracking používajú a následne ich implementujeme v softvéri Wolfram Mathematica, kde sonifikujeme údery pre piesne rôzneho žánru a tempa.

Title: Music processing: Beat tracking

Keywords: sound, sound recording, frequency, Fourier transform, tempo

Abstract: When listening to music, one subconsciously stomps his feet to the beat of a song, thus being able to automatically determine its tempo. The bachelor's thesis will deal with how to achieve the same result, i.e. to find beats in a given song, using music processing. We begin with a detailed explanation of the main concepts and algorithms that are used for beat tracking and then implement them in Wolfram Mathematica, where we sonify beats in songs of various genres and tempos.

Predhovor

Táto bakalárska práca sa zaoberá algoritmami, ktoré sa používajú na beat tracking piesní. Cieľom práce je detailne popísať kroky algoritmov a vysvetliť základné pojmy potrebné pre ich porozumenie. Algoritmy beat trackingu implementujeme v softvéri Wolfram Mathematica, kde vypracujeme program, ktorý použijeme na niekoľko zvukových nahrávok. K výsledným nahrávkam sme vytvorili [zoznam piesní na YouTube](#). Odkazy na piesne sa nachádzajú aj priamo v texte, pričom sú vždy zvýraznené touto farbou. Hlavným zdrojom bakalárskej práce je kniha [2], ktorej hlavné myšlienky v tejto práci reprodukuje.

Kľúčovým pojmom bakalárskej práce je zvuk a s ním súvisiace pojmy ako zvuková vlna, frekvencia a fáza, ktoré si objasníme v kapitole 1. Taktiež sa tu zaoberáme vznikom digitálnej zvukovej nahrávky.

V kapitole 2 si predstavíme definíciu Fourierovej transformácie, založenú na výpočte podobnosti. Následne definíciu upravíme pomocou použitia komplexných čísel a pre počítačové spracovanie si zavedieme aj definíciu diskretnej Fourierovej transformácie. V poslednej sekcii tejto kapitoly si popíšeme výpočet diskretnej krátkodobej Fourierovej transformácie, čím dostaneme informáciu o prítomných frekvenciách v čase.

V kapitole 3 sa budeme zaoberať prvým krokom algoritmu beat trackingu, ktorým je detekcia nástupov tónov. Bližšie si popíšeme niekoľko pístupov detekcií. Začneme s detekciou nástupov založenou na energii zvuku, a budeme pokračovať s detekciou založenou na spektre zvuku, na fáze a nakoniec na kombinácii spektra a fázy.

Dalším krokom algoritmu, je analýza tempa. V kapitole 4 najprv určíme tempo piesne pomocou výpočtu diskretnej Fourierovej transformácie novinkovej funkcie a neskôr pomocou diskretnej krátkodobej Fourierovej transformácie novinkovej funkcie, čím získame informáciu o prítomných tempách piesne aj v prípade, že sa tempo mení v čase.

V poslednej kapitole 5 si predstavíme posledný krok algoritmu beat trackingu, ktorým je nájdenie PLP funkcie, čím zachytíme lokálne tempo zvukovej nahrávky.

Na záver by som sa chcela poďakovať vedúcemu bakalárskej práce Ing. Mgr. Lukášovi Tomekovi, PhD. za ústretový prístup, rady a pomoc pri jej vypracovaní.

Bratislava, 6. mája 2021

Eva Bencová

Obsah

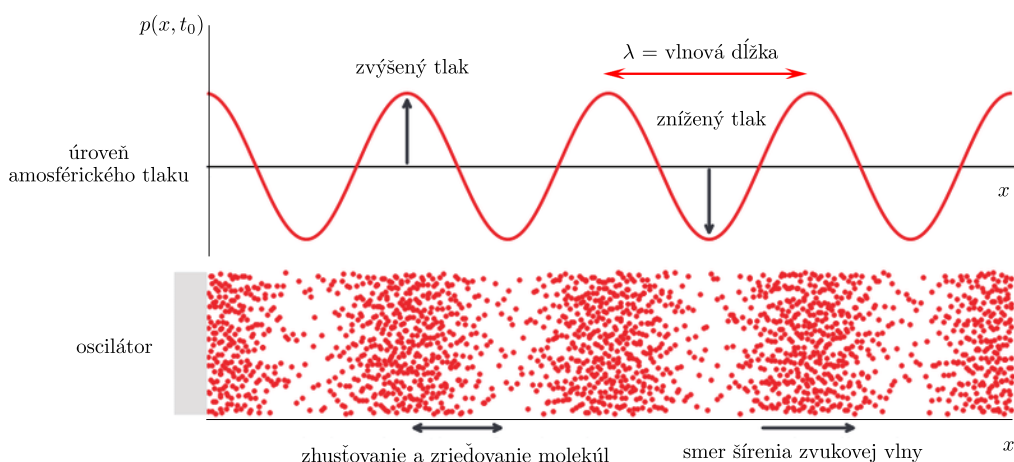
1	Úvod – Zvuk a zvukové nahrávky	1
1.1	Zvukové vlny	1
1.2	Zvukové nahrávky	2
1.3	Frekvencie	4
1.4	Fáza	5
2	Fourierova transformácia	6
2.1	Výpočet podobnosti funkcií pomocou integrálov	6
2.2	Definícia Fourierovej transformácie	7
2.3	Komplexné čísla	8
2.4	Definícia Fourierovej transformácie pomocou komplexných čísel	9
2.5	Diskrétna Fourierova transformácia	10
2.5.1	Odvodenie diskkrétnej Fourierovej transformácie	10
2.6	Krátkodobá Fourierova transformácia	13
2.6.1	Hannovo okno	14
2.6.2	Definícia diskkrétnej krátkodobej Fourierovej transformácie	15
3	Detekcia nástupu tónov	17
3.1	Detekcia založená na energii zvuku	17
3.1.1	Výpočet energie	17
3.2	Detekcia založená na spektre zvuku	20
3.3	Detekcia založená na fáze	21
3.4	Detekcia založená na spektre a fáze	23
4	Analýza tempa	25
4.1	Tempové spektrum	25
4.2	Tempogram z krátkodobej Fourierovej transformácie	27
5	Beat tracking	29
5.1	Beat tracking pomocou tempového spektra	29
5.2	PLP funkcia	30
5.3	Zhrnutie algoritmu beat trackingu	33
6	Záver	35
	Literatúra	36

1 Úvod – Zvuk a zvukové nahrávky

V prvej kapitole si predstavíme základné pojmy potrebné pre spracovanie zvuku. Zvuk definujeme ako *vlnenie*, ktoré vznikne vibrovaním nejakého objektu, príkladom môže byť struna na husliach alebo naše hlasivky. Vibrovanie spôsobí kmitanie molekúl vzduchu, ktoré si tento vzruch ďalej odovzdávajú, a teda dochádza k šíreniu vibrácií, vlneniu. Toto vlnenie je pozdĺžne, čo znamená, že častice kmitajú v smere vlnenia [2].

1.1 Zvukové vlny

Na šírenie vibrácií je potrebné médium [1]. V tejto práci sa budeme zaoberať šírením vo vzduchu. Vibrácie oscilátora spôsobia zmenu polohy molekúl vzduchu, tie narazia do ďalších molekúl, ktoré opäť narazia na iné. Ich zhutňovanie a následné zriedenie, spôsobuje zmenu atmosférického tlaku. Tento proces odovzdávania si energie nazývame *zvuková vlna*. Najjednoduchším príkladom zvukovej vlny je pre nás sínusoida, tvorená harmonickým kmitaním oscilátora, zobrazená aj na obrázku 1.1. Vo všeobecnosti však oscilátor nemusí kmitať pravidelne a zvuková vlna môže mať oveľa zložitejší priebeh. Odchýlkam tlaku, inak povedané pretlaku alebo podtlaku voči atmosférickému tlaku, sa hovorí akustický tlak. Akustický tlak $p(x, t)$ je funkciou závislou od dvoch premenných, od priestorovej súradnice x a času t .¹ Na obrázku 1.1 vidíme graf meniaceho sa akustického tlaku v priestore pre fixovanú hodnotu času $t = t_0$.



Obr. 1.1: Vizualizácia molekúl vzduchu a k tomu prislúchajúci graf akustického tlaku [4].

¹Vo všeobecnosti je p funkciou až 4 premenných: priestorových súradníc x, y, z a času t . Tu pre ilustráciu uvažujeme tzv. rovinnú vlnu, v ktorej p nezávisí od y, z .

Zvukové vlny sú riešeniami vlnovej rovnice, ktorá má v 1D tvar

$$\partial_t^2 p(x, t) = c^2 \partial_x^2 p(x, t), \quad (1.1)$$

kde c vyjadruje rýchlosť zvuku. Jedným z mnohých riešení vlnovej rovnice je aj spomínaná sínusoida. Predpis tejto funkcie je

$$p(x, t) = A \cos(kx - 2\pi ft), \quad (1.2)$$

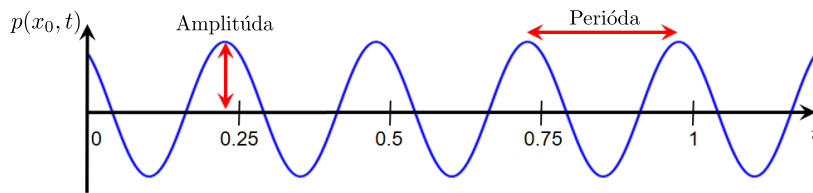
kde A je *amplitúda* akustického tlaku, ktorá vyjadruje veľkosť najväčšej zmeny tlaku a súvisí s energiou zvukovej vlny a s hlasitosťou zvuku. Číslo k sa nazýva vlnové číslo a je dané vzťahom $k = \frac{2\pi}{\lambda}$, kde λ je vlnová dĺžka (obr. 1.1). Vlnová dĺžka sa rovná vzdialenosti dvoch po sebe nasledujúcich vrcholov vlny (bodov s najväčšou hodnotou akustického tlaku). Ďalej f označuje frekvenciu, ktorej sa viac budeme venovať v sekcii 1.3. Keďže sínusoida (1.2) je riešením vlnovej rovnice (1.1), tak pre rýchlosť zvuku platí

$$c = \lambda f, \quad (1.3)$$

čo sa dá ľahko vidieť dosadením predpisu (1.2) do vlnovej rovnice (1.1).

1.2 Zvukové nahrávky

Pre počítačové spracovanie zvuku potrebujeme vytvoriť zvukovú nahrávku, a teda zaznamenať analógový signál a previesť ho na digitálny. Analógový signál zaznamenáme pomocou mikrofónu. Funkciu analógového signálu, ktorá je spojitá v čase označíme ako $s(t)$. Zvukovej vlne do cesty postavíme mikrofón v mieste $x = x_0$, čím získame závislosť akustického tlaku v mieste x_0 od času t . Pre vlnu (1.2) dostaneme funkciu $p(x_0, t) = A \cos(kx_0 - 2\pi ft)$. Mikrofón v skutočnosti zaznamená signál $s(t)$ úmerný akustickému tlaku, $s(t) \sim p(x_0, t)$.



Obr. 1.2: Akustický tlak v bode x_0 závislý od času [2].

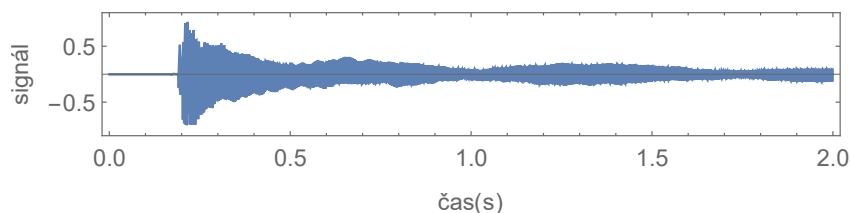
Prevod analógového signálu na digitálny sa nazýva digitalizácia. Spočíva v diskretizácii analógového signálu, čím dostaneme hodnoty, ktorým hovoríme vzorky. Diskretizujeme analógový signál, čo znamená funkciu $s(t)$, pre diskkrétne hodnoty času

$$t_n = nT_s, \quad n \in [0, N - 1], \quad (1.4)$$

kde n je časový index a T_s je vzorkovacia periódá. Počet vzoriek označíme ako N . Vzorkovacia periódá je daná vzťahom

$$T_s = \frac{1}{F_s}, \quad (1.5)$$

kde F_s predstavuje vzorkovaciu frekvenciu zvukovej nahrávky, ktorá sa udáva v Hz. Definujeme ju ako počet vzoriek analógového signálu za sekundu. Pre MP3 nahrávky sa najčastejšie používa $F_s = 44\,100 \text{ Hz} = 44.1 \text{ kHz}$. Vytvorená zvuková nahrávka je teda funkciou, závislou od času a pozostáva zo vzoriek merania. Príklad vidíme na obrázku 1.3, kde graf signálu predstavuje **zvukovú nahrávku tónu C4, zahraného na klavíri**.

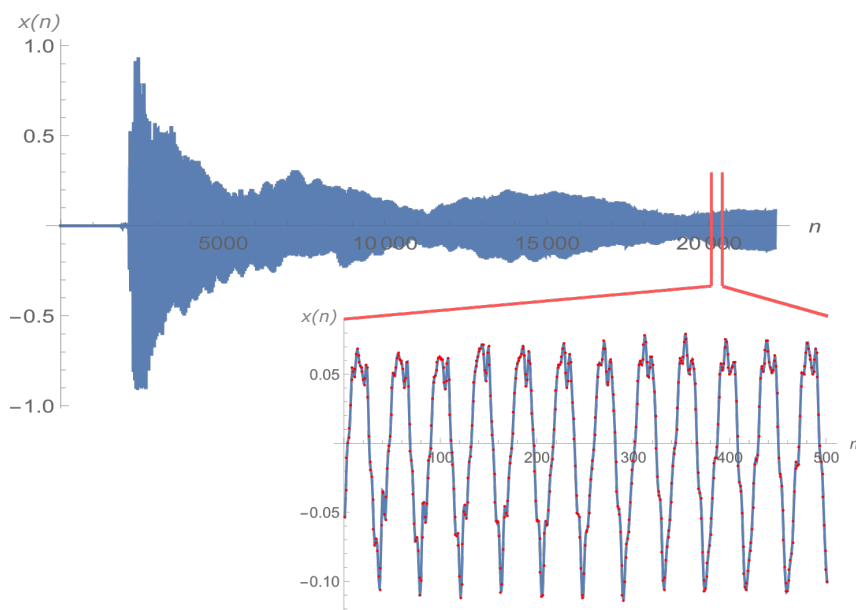


Obr. 1.3: Graf signálu diskrétného v čase.

Hodnoty vzoriek zvukovej nahrávky, ktoré nadobúda signál diskretizovaný v čase, zapíšeme do poľa. Toto pole dĺžky N označíme ako x a dostaneme ho vzťahom

$$x(n) = s\left(\frac{n}{F_s}\right), \quad n \in [0, N-1], \quad (1.6)$$

kde $x(n)$ predstavuje hodnotu analógového signálu v čase $t_n = \frac{n}{F_s}$. Takéto pole vidíme na obr. 1.4, kde na vodorovnej osi už nie je čas, ale časový index n a na zvislej osi hodnota $x(n)$.



Obr. 1.4: Pole hodnôt vzoriek **zvukovej nahrávky tónu C4, zahraného na klavíri**.

1.3 Frekvencie

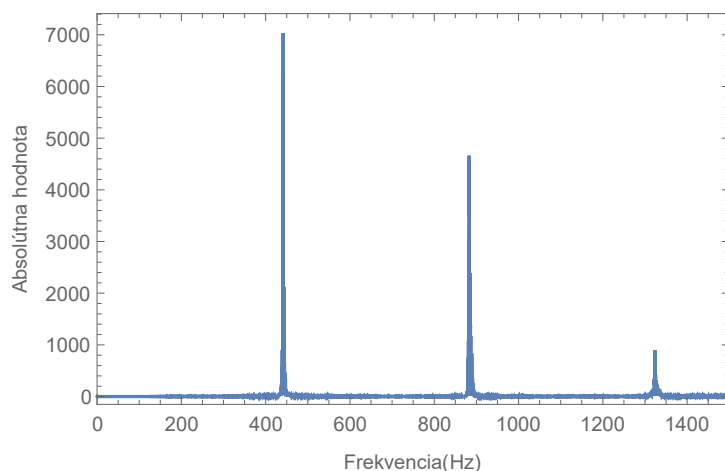
Ak sa zvuková vlna správa ako periodický jav, množstvo periód za jednotku času udáva *frekvencia* f zvukovej vlny, ktorá je definovaná vzťahom

$$f = \frac{1}{T}, \quad (1.7)$$

ktorý vyjadruje, že frekvencia je obrátenou hodnotou periódy. Perióda T predstavuje čas, za ktorý sa oscilátor pri kmitaní vráti na pôvodné miesto. Jednotkou frekvencie je Hertz (Hz). Na obrázku 1.2 vidíme zvukovú vlnu s frekvenciou 4 Hz (medzi časom 0 s a 1 s vidíme 4 periódy). Čím kratší je čas medzi osciláciami objektu, t.j. čím viac periód za sekundu, tým je frekvencia vyššia [3]. Pomocou rýchlosti zvukovej vlny (1.3) si vieme frekvenciu vyjadriť aj ako

$$f = \frac{c}{\lambda}. \quad (1.8)$$

Hodnotu frekvencie vnímame ako výšku tónu. Čím je väčšia hodnota frekvencie, tým vyšší tón počujeme. Ľudské ucho dokáže vnímať frekvencie v rozmedzí od 20 Hz do 20000 Hz [2]. Čistý tón zodpovedá zvukovej vlny, ktorá má sínusový priebeh s danou frekvenciou f , napr. sínusoida s frekvenciou $f = 440$ Hz, predstavuje čistý tón A4, ktorý sa používa pri ladení hudobných nástrojov. Žiaden hudobný nástroj však nevie zahrať čistý tón a ako sme videli na obrázku 1.4 priebeh zvukovej vlny je preto vo všeobecnosti zložitejší, než sínusoida. Pri zahrnutí jedného tónu na hudobnom nástroji je zvuk zmesou mnohých sinusoíd s rôznymi frekvenciami a amplitúdami. Vtedy je tón tvorený základnou frekvenciou a vyššími harmonickými frekvenciami, ktoré sú celočíselnými násobkami základnej. Za základnú frekvenciu považujeme najnižšiu prítomnú frekvenciu.² Pre vizualizáciu frekvencií, ktoré zvuková nahrávka obsahuje, sa používa spektrum, kde x-ová os reprezentuje frekvencie a y-ová ich intenzitu. Na obrázku 1.5 vidíme spektrum zvukovej nahrávky pri zahrnutí tónu A4 na priechnej flaute. V spektre si môžeme všimnúť základnú frekvenciu $f = 440$ Hz, a taktiež vyššie harmonické frekvencie, $2f = 880$ Hz, $3f = 1320$ Hz.



Obr. 1.5: Spektrum pre tón A4, zahrnutý na priechnej flaute.

²V skutočnosti je to však zložitejšie, pretože základná frekvencia nemusí byť prítomná, ale človek ju napriek tomu počuje. Tento fenomén sa nazýva chýbajúca základná frekvencia.

1.4 Fáza

Ďalšou charakteristikou zvukovej vlny, ktorá bude taktiež zohrávať dôležitú úlohu pri spracovaní signálu je *fáza*. Na zistenie fázy sa vždy pýtame, aká je miera posunutia sínusoidy od počiatku súradnicovej sústavy. Pre danú funkciu sínusoidy

$$s(t) = A \cos(2\pi(ft - \varphi)), \quad (1.9)$$

parameter φ označujeme ako fázu. Meria sa v normalizovaných radiánoch, teda v intervale $\varphi \in [0, 1)$, kedy hodnota fázy vyjadruje o akú časť periódy je $\cos$ posunutý ³.

³Obvykle sa fáza definuje inak, meria sa v radiánoch $\varphi \in [0, 2\pi)$ a namiesto (1.9) vtedy máme funkciu sínusoidy $s(t) = A \cos(2\pi ft - \varphi)$

2 Fourierova transformácia

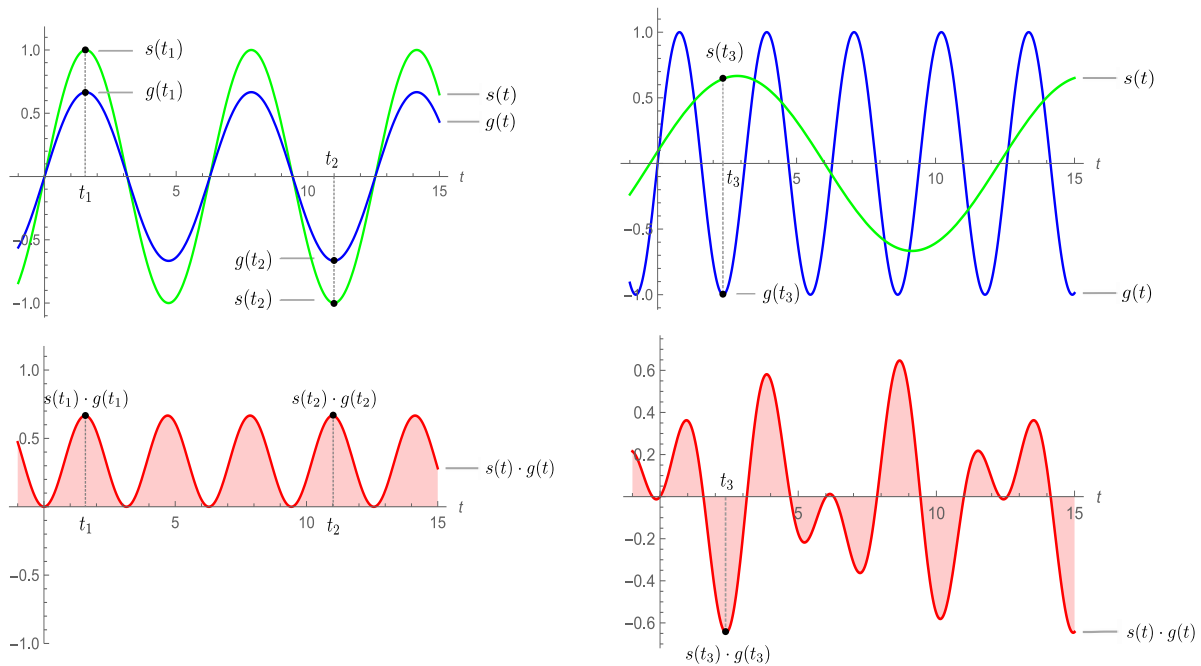
Po vytvorení zvukovej nahrávky je pre jej ďalšie spracovanie potrebné získať o nej bližšie informácie. V sekcii 1.3 sme hovorili o tom, že frekvencia zodpovedá výške tónu. Pomocou Fourierovej transformácie získame informáciu o prítomných frekvenciách v zvukovej nahrávke, a teda vieme určiť aj výšku tónov, ktoré v zvukovej nahrávke zazneli. Základnou myšlienkou Fourierovej transformácie je rozklad signálu na frekvenčné spektrum porovnaním signálu so sínusoidami s rôznymi frekvenciami. Pre jednoduchosť budeme najprv pracovať s analógovým signálom, pričom neskôr vysvetlíme Fourierovu transformáciu aj pre signál diskretizovaný v čase.

2.1 Výpočet podobnosti funkcií pomocou integrálov

Fourierovou transformáciou dostaneme pre danú frekvenciu koeficient podobnosti signálu so sínusoidou s danou frekvenciou. Čím väčšia hodnota koeficientu, tým viac sa signál podobá na danú sínusoidu. Pre získanie týchto koeficientov je teda potrebné signál porovnať s rôznymi sínusoidami. Majme dve funkcie, funkciu analógového signálu $s(t)$, a funkciu $g(t)$, s ktorou ju chceme porovnať. Pomocou výpočtu ich skalárneho súčinu, čo znamená integrálu zo súčinu funkcií, zistíme ich podobnosť. Výpočet podobnosti funkcií je:

$$\int_{t \in \mathbb{R}} s(t) \cdot g(t) dt. \quad (2.1)$$

Súčinom hodnôt $s(t)$ a $g(t)$ pre pevne zvolené t zistíme, či sú hodnoty funkcií v bode t zhodné ich znamienkom. Ak majú obe funkcie v danom čase t kladnú hodnotu (Obr. 2.1, vľavo, $t = t_1$) alebo obe zápornú hodnotu (Obr. 2.1, vľavo, $t = t_2$), znamienkom sa zhodujú, a teda sú si podobné, z ich súčinu dostaneme kladnú hodnotu. Ak však má v danom čase t jedna funkcia kladnú a druhá zápornú hodnotu (Obr. 2.1, vpravo, $t = t_3$), nie sú si podobné a zo súčinu dostaneme zápornú hodnotu. Integrál definujeme ako plochu pod krivkou, pričom plochu pod x-ovou osou berieme so záporným znamienkom, odpočítame ju. Na obrázku. 2.1, vľavo, dostávame súčin funkcií $s(t) \cdot g(t) \geq 0$, pre $t \in \mathbb{R}$, a preto integrálom z tohto súčinu získame väčšiu hodnotu než v prípade, ktorý vidíme na obrázku. 2.1, vpravo, kde sme zo súčinu pre niektoré t dostali aj záporné hodnoty, a tak bude hodnota integrálu bližšie k nule. Skalárnym súčinom podobných funkcií teda dostaneme väčšiu hodnotu ako skalárnym súčinom funkcií, ktoré sú si menej podobné.



Obr. 2.1: Skalárny súčin funkcií ako plocha pod grafom ich súčinu. Vľavo: Súčin podobných funkcií. Vpravo: Súčin veľmi rozdielnych funkcií.

2.2 Definícia Fourierovej transformácie

Ako už bolo napísané, základnou myšlienkou Fourierovej transformácie je porovnanie signálu $s(t)$ so sínusoidami s rôznymi frekvenciami a fázami, čím signál pretransformujeme na frekvenčné spektrum. Funkcia sínusoidy má tvar

$$\cos_{f,\varphi}(t) := A \cos(2\pi(ft - \varphi)), \quad (2.2)$$

kde zvolíme $A = \sqrt{2}$. Skalárnym súčinom funkcií vypočítame ich podobnosť, preto pre fixovanú hodnotu frekvencie f určíme

$$d_f := \max_{\varphi \in [0,1)} \int_{t \in \mathbb{R}} s(t) \cdot \cos_{f,\varphi}(t) dt, \quad (2.3)$$

$$\varphi_f := \operatorname{argmax}_{\varphi \in [0,1)} \int_{t \in \mathbb{R}} s(t) \cdot \cos_{f,\varphi}(t) dt, \quad (2.4)$$

kde d_f predstavuje koeficient podobnosti a φ_f koeficient fázy. Koeficient podobnosti vypočítame ako maximum zo skalárneho súčinu signálu a sínusoid s danou frekvenciou, ale rôznymi fázami, $\varphi \in [0, 1)$. Argumentom tohoto maxima je koeficient fázy, ktorý vyjadruje optimálnu fázu, pre ktorú je skalárny súčin maximálny. Tak určíme o akú časť periódy musí byť sínusoida s danou frekvenciou posunutá, aby bola najviac podobná signálu.

Takýto výpočet je však pomerne komplikovaný, pretože zahŕňa aj optimalizačný krok, maximalizáciu. Pre odstránenie tohoto problému použijeme komplexné čísla, čím dostaneme jednoduchšiu definíciu Fourierovej transformácie.

2.3 Komplexné čísla

Rozšírením množiny reálnych čísel $\mathbb{R}$ dostaneme množinu komplexných čísel, ktorú označujeme ako $\mathbb{C}$. Z toho vyplýva, že $\mathbb{R} \subset \mathbb{C}$. Nech i je imaginárna jednotka, pre ktorú platí vzťah $i = \sqrt{-1}$. Komplexné číslo je usporiadaná dvojica reálnych čísel (a, b) . Algebraický tvar komplexného čísla je

$$c = a + bi, \quad a, b \in \mathbb{R}, \quad (2.5)$$

kde a je reálna časť a b je imaginárna časť komplexného čísla c , čo zapíšeme ako

$$a = \operatorname{Re}(c), \quad (2.6)$$

$$b = \operatorname{Im}(c). \quad (2.7)$$

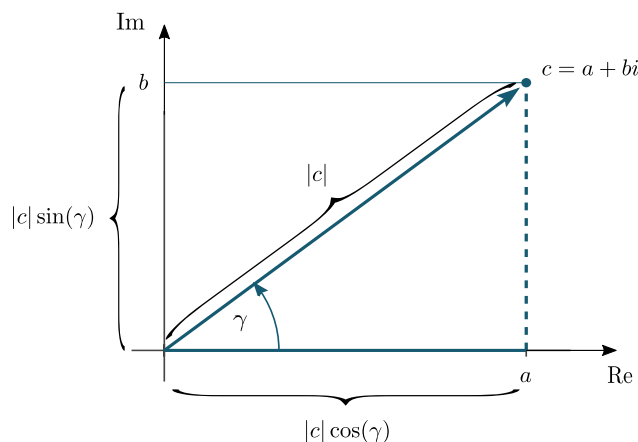
Platí, že $c \in \mathbb{R}$ vtedy a len vtedy, keď $b = 0$. Množinu komplexných čísel môžeme reprezentovať aj ako 2D rovinu, ktorú nazývame komplexná rovina, kde x-ová os predstavuje reálnu časť a y-ová imaginárnu časť. Komplexné číslo vieme teda zakresliť ako bod v komplexnej rovine. Vzdialenosť tohoto bodu od počiatku súradnicovej sústavy sa rovná hodnote

$$|c| = \sqrt{a^2 + b^2}, \quad (2.8)$$

ktorá sa označuje ako absolútna hodnota alebo veľkosť komplexného čísla c . Úsečka, ktorá spája bod, predstavujúci komplexné číslo, a počiatok súradnicovej sústavy zvierá s kladnou poloosou osi x uhol $\gamma \in [0, 2\pi)$ (Obr. 2.2). Okrem algebraického tvaru vieme komplexné číslo vyjadriť aj v goniometrickom tvare s použitím polárnych súradníc, pričom využijeme veľkosť komplexného čísla a uhol γ . Komplexné číslo c vyjadríme ako

$$c = |c|(\cos(\gamma) + i \sin(\gamma)), \quad (2.9)$$

kde sme reálnu časť vyjadrili ako $a = |c| \cos(\gamma)$, imaginárnu časť ako $b = |c| \sin(\gamma)$ a následne sme veľkosť komplexného čísla vyňali pred zátvorku.

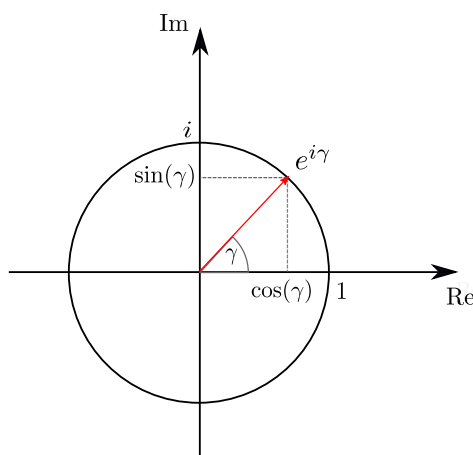


Obr. 2.2: Reprezentácia komplexného čísla $c = a + bi$ pomocou polárnych súradníc.

Goniometrický tvar komplexného čísla ďalej vieme upraviť použitím *Eulerovej identity*, ktorá vraví, že

$$e^{i\gamma} = \cos(\gamma) + i \sin(\gamma). \quad (2.10)$$

Znázornenie goniometrického tvaru vidíme aj na obrázku 2.3.



Obr. 2.3: Eulerova identita.

Dosadením vzorca (2.10) do (2.9) dostaneme komplexné číslo c v tvare

$$c = |c|e^{i\gamma}. \quad (2.11)$$

2.4 Definícia Fourierovej transformácie pomocou komplexných čísel

Fourierovu transformáciu si teraz zdefinujeme pomocou komplexných čísel, vďaka čomu sa vyhneme optimalizačnému kroku v (2.3) a (2.4), a tak zjednodušíme výpočet. Základnou myšlienkou je zahrnúť obe čísla, koeficient podobnosti d_f a koeficient fázy φ_f do komplexného čísla c_f . Použijeme polárnu reprezentáciu komplexného čísla (2.11), kde koeficient podobnosti využijeme pre výpočet veľkosti komplexného čísla a koeficient fázy zakódujeme do uhla γ . Komplexné číslo zapíšeme v tvare

$$c_f := \frac{d_f}{\sqrt{2}} e^{2\pi i(-\varphi_f)}, \quad (2.12)$$

kde veľkosť komplexného čísla sa rovná $|c_f| = \frac{d_f}{\sqrt{2}}$ a uhol $\gamma_f = -2\pi\varphi_f$. Rovnako, ako sme koeficienty podobnosti a fázy určovali pre sínusoidy s rôznymi frekvenciami, aj komplexné číslo c_f chceme vypočítať pre rôzne frekvencie. Vytvoríme teda funkciu $\hat{s}(f)$, ktorá danej frekvencii priradí komplexné číslo c_f ,

$$\hat{s}(f) = c_f. \quad (2.13)$$

Jedným z hlavných výsledkov Fourierovej teórie je, že hodnotu $\hat{s}(f)$ vypočítame ako

$$\hat{s}(f) = \int_{t \in \mathbb{R}} s(t) e^{-2\pi i f t} dt. \quad (2.14)$$

Namiesto výpočtu dvoch reálnych čísel, ktorý sme definovali v (2.3) a (2.4), máme teraz výpočet jedného komplexného čísla. Funkcia $\hat{s}(f)$ sa označuje ako Fourierova transformácia signálu $s(t)$. Hodnotami funkcie sú komplexné čísla a nazývajú sa Fourierove koeficienty. Pomocou Fourierovho koeficientu, do ktorého sme zakódovali informáciu o koeficiente podobnosti d_f a fázy φ_f , vieme spätne dostať tieto hodnoty výpočtom

$$d_f = \sqrt{2} |\hat{s}(f)|, \quad (2.15)$$

$$\varphi_f = -\frac{\gamma}{2\pi}. \quad (2.16)$$

2.5 Diskrétna Fourierova transformácia

V predchádzajúcej sekcii sme uviedli definíciu Fourierovej transformácie pre analógový signál, teda funkciu $s(t)$, spojitú v čase. V našom prípade pre spracovanie zvukovej nahrávky však potrebujeme odvodiť túto transformáciu pre signál, ktorý je diskretný v čase. Pracovať budeme s poľom hodnôt vzoriek $x(n)$, $n \in [0, N-1]$, ktoré sme zdefinovali vo vzorci (1.6).

2.5.1 Odvodenie diskkrétnej Fourierovej transformácie

Diskrétnu Fourierovu transformáciu odvodíme z analytického tvaru (2.14) diskretizáciou pre hodnoty času $t_n = nT_s$. Integrál nahradíme jeho aproximáciou, Riemannovou sumou, cez všetky vzorky signálu.¹ Hodnota dt bude teraz predstavovať vzorkovaciu periódu $T_s = \frac{1}{F_s}$, ktorá vyjadruje čas, ktorý uplynie medzi dvoma vzorkami. Následne pomocou diskretizácie definovanej v (1.6) upravíme Fourierovu transformáciu na tvar

$$\hat{s}(f) = \int_{t \in \mathbb{R}} s(t) e^{-2\pi i f t} dt \approx \sum_{n=0}^{N-1} s(nT_s) e^{-2\pi i f n T_s} T_s = \sum_{n=0}^{N-1} x(n) e^{-2\pi i f n T_s} T_s. \quad (2.17)$$

Ďalej zavedieme novú premennú

$$\bar{f} = f T_s, \quad (2.18)$$

pomocou ktorej vyjadríme f ako $f = \frac{\bar{f}}{T_s}$ a dosadením dostaneme

$$\hat{s}\left(\frac{\bar{f}}{T_s}\right) \approx \sum_{n=0}^{N-1} x(n) e^{-2\pi i \bar{f} n T_s} \quad (2.19)$$

$$\frac{1}{T_s} \hat{s}\left(\frac{\bar{f}}{T_s}\right) \approx \sum_{n=0}^{N-1} x(n) e^{-2\pi i \bar{f} n}. \quad (2.20)$$

¹Signál $s(t)$ je mimo zvukovej nahrávky nulový, preto v skutočnosti integrujeme iba cez interval $[0, \text{dĺžka nahrávky}]$.

Funkcia $\hat{s}(f)$ predstavuje Fourierovu transformáciu pre analógový signál $s(t)$. Pre diskretný signál $x(n)$ zavedieme funkciu $\hat{x}(\bar{f})$, ktorá predstavuje odvodenú aproximáciu

$$\hat{x}(\bar{f}) \approx \frac{1}{T_s} \hat{s}\left(\frac{\bar{f}}{T_s}\right), \quad (2.21)$$

$$\hat{x}(\bar{f}) = \sum_{n=0}^{N-1} x(n) e^{-2\pi i \bar{f} n}. \quad (2.22)$$

Diskrétné hodnoty chceme dosádzať nielen za čas, ale aj za $\bar{f}$. Pre tento krok potrebujeme najprv zadať pojem *Nyquistova frekvencia*². Predstavuje maximálnu frekvenciu, ktorú ešte má zmysel hľadať v diskretnom signáli, a rovná sa hodnote

$$f = \frac{1}{2} F_s. \quad (2.23)$$

Do vzorca (2.18) dosadíme hodnotu Nyquistovej frekvencie, vďaka čomu zistíme, aká je horná hranica intervalu, z ktorého budeme voliť diskretné hodnoty pre $\bar{f}$

$$\bar{f} = \frac{1}{2} F_s T_s, \quad (2.24)$$

kde $T_s = \frac{1}{F_s}$, a preto horná hranica sa rovná hodnote

$$\bar{f} = \frac{1}{2} F_s \frac{1}{F_s} = \frac{1}{2}. \quad (2.25)$$

Interval, z ktorého budeme voliť diskretné hodnoty je $\bar{f} \in [0, \frac{1}{2}]$. Volíme ich ako

$$\bar{f} = \frac{k}{M}, \quad k \in [0, M-1], \quad (2.26)$$

pričom hodnotu M často volíme ako $M = N$. Dosadením (2.26) do (2.21) dostaneme funkciu, ktorú označíme $X(k)$, a tak zadefinujeme diskretnú Fourierovu transformáciu predpisom

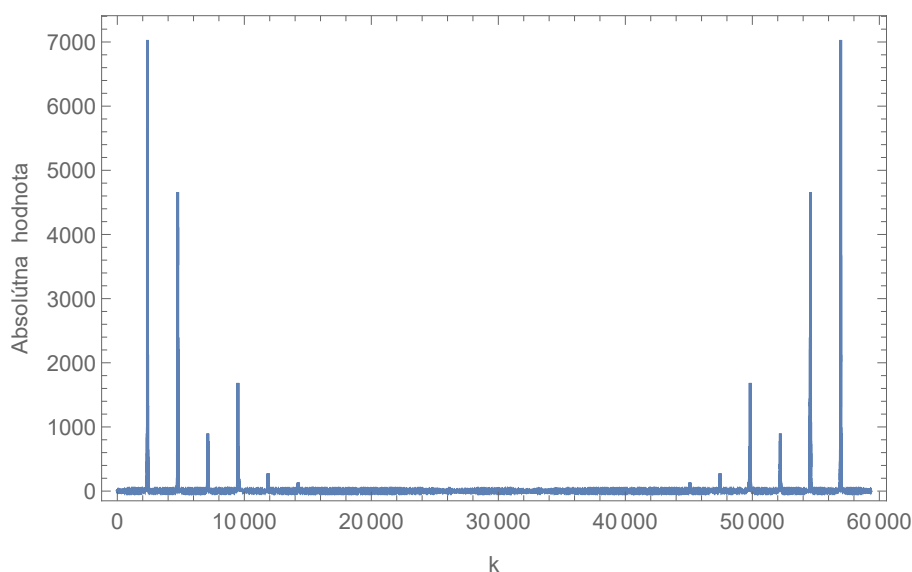
$$X(k) := \hat{x}(k/N) = \sum_{n=0}^{N-1} x(n) e^{\frac{-2\pi i k n}{N}}, \quad k \in [0, N-1]. \quad (2.27)$$

Pre každú hodnotu k tak dostaneme hodnotu diskretnéj Fourierovej transformácie. Príslúchajúcu frekvenciu $f(k)$ vieme dopočítať odvodením pomocou (2.18) ako

$$f(k) = \frac{\bar{f}}{T_s} = \frac{k}{N T_s} = \frac{k F_s}{N} \quad (2.28)$$

Intenzitu danej frekvencie, inak povedané to, ako veľmi je frekvencia f v zvukovej nahrávke zastúpená, vypočítame ako absolútnu hodnotu komplexného čísla, ktoré je výsledkom diskretnéj Fourierovej transformácie, a teda hodnoty $X(k)$.

²Táto frekvencia je pomenovaná po matematikovi Harrym Nyquistovi.



Obr. 2.4: Graf diskkrétnej Fourierovej transformácie zvukovej nahrávky tónu A4. Na x-ovej osi sú indexy $k \in [0, N - 1]$, prislúchajúce frekvenciám, a na y-ovej hodnoty $|X(k)|$.

Po vykreslení si môžeme všimnúť, že diskrétna Fourierova transformácia má vlastnosť symetrie, presnejšie $|X(n - k)| = |X(k)|$.³ Hodnoty $X(k)$ ako aj k nim prislúchajúce frekvencie stačí vypočítať iba pre interval $k \in [0, N/2]$, pretože po dosadení hornej hranice $k = N/2$ do (2.28) dostaneme hodnotu rovnú Nyquistovej frekvencii. Symetria je dôsledkom diskretizácie. Analytická Fourierova transformácia túto vlastnosť nemá.

Teraz už máme všetky potrebné informácie pre vykreslenie frekvenčného spektra, kde sú na vodorovnej osi frekvencie a na zvislej je ich intenzita. V nasledujúcom príklade sme aplikovali diskrétnu Fourierovu transformáciu na zvukovú nahrávku úvodných tónov piesne Greensleeves (Obr. 2.5), ktorú si môžeme vypočítať [tu](#). Výsledkom je frekvenčné spektrum (Obr. 2.6), v ktorom vidíme základné frekvencie tónov zapísané v tabuľke 2.1. Okrem základných frekvencií vidíme v spektre aj prvé harmonické frekvencie zodpovedajúce dvojnásobku základných frekvencií. Ich hodnoty sme zapísali do tabuľky 2.2.



Obr. 2.5: Notový zápis začiatku piesne [Greensleeves](#).

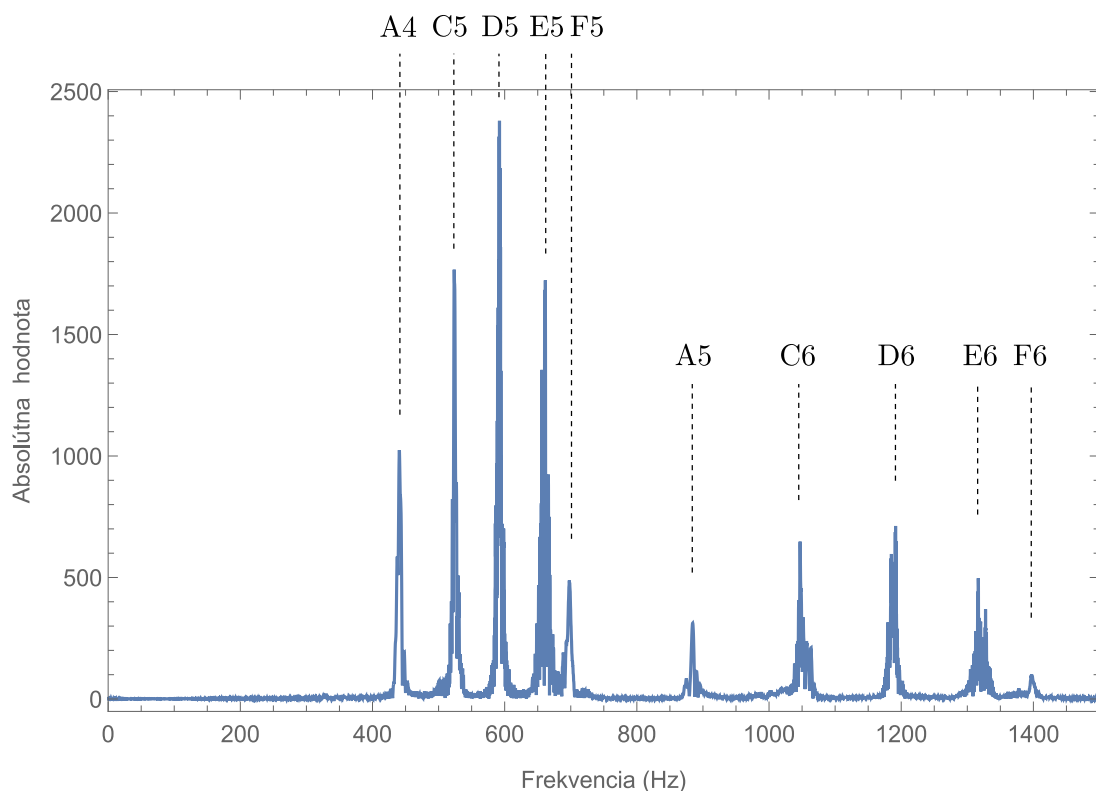
Tón	A4	C5	D5	E5	F5
Frekvencia (Hz)	440	523.25	587.33	659.25	698.46

Tabuľka 2.1: Tóny, ktoré zaznejú v zvukovej nahrávke a k nim prislúchajúce hodnoty frekvencií.

³Dá sa to ľahko overiť úpravou zo vzorca (2.27) napísaného pre $\overline{X(n - k)}$. Vyjde $\overline{X(n - k)} = X(k)$, z čoho plynie symetria $|X(n - k)| = |X(k)|$.

Tón	A5	C6	D6	E6	F6
Frekvencia (Hz)	880	1046.50	1174.66	1318.50	1396.92

Tabuľka 2.2: Hodnoty prvých harmonických frekvencií, ktoré sú taktiež zaznamenané pomocou diskretnéj Fourierovej transformácie.



Obr. 2.6: Frekvenčné spektrum, v ktorom vidíme základné frekvencie tónov A4, C5, D5, E5, F5, ktoré zazneli v zvukovej nahrávke piesne **Greensleeves**, ale aj prvé harmonické frekvencie, ktoré prislúchajú tónom A5, C6, D6, E6, F6.

2.6 Krátkodobá Fourierova transformácia

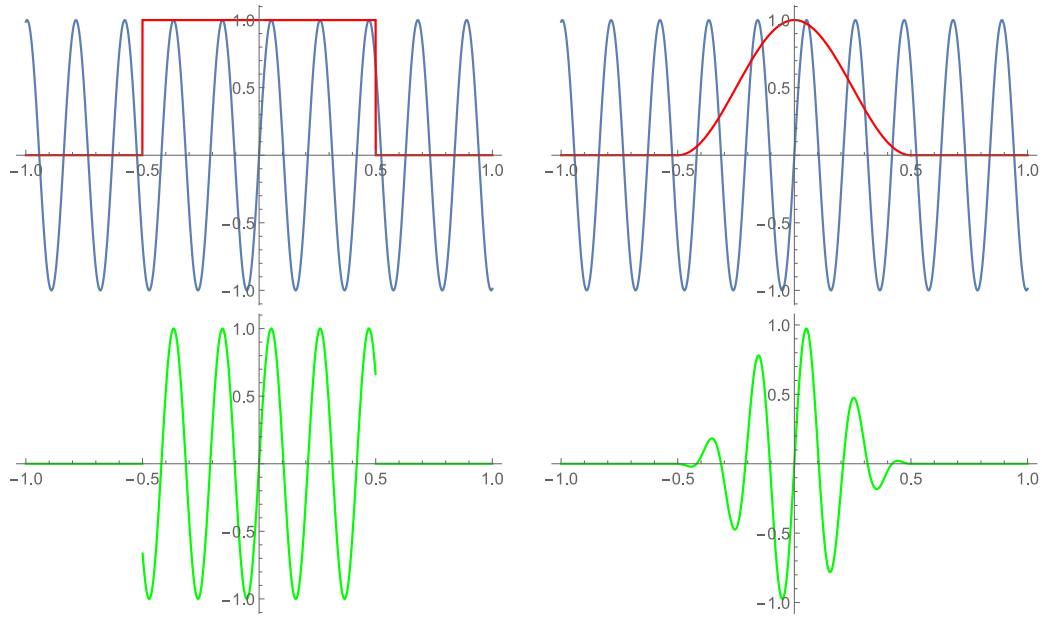
Fourierovu transformáciu sme si už zadefinovali pre analógový signál a aj pre signál diskretný v čase, čoho výsledkom bolo frekvenčné spektrum. Ako však vidíme aj na obrázku 2.6, po transformácii signálu sme síce získali informáciu o prítomných frekvenciách, ale stratili sme informáciu o plynúcom čase v zvukovej nahrávke. Úlohou krátkodobej Fourierovej transformácie je teda zistiť, v akom čase boli prítomné jednotlivé frekvencie. Základnou myšlienkou je výpočet Fourierovej transformácie iba v úzkych časových výsekoch zvukovej nahrávky, vďaka čomu si ponecháme aj informáciu o čase. Najprv si potrebujeme zadefinovať funkciu s názvom *Hannovo okno*, ktorú budeme neskôr vo výpočte krátkodobej Fourierovej transformácie používať.

2.6.1 Hannovo okno

Okno je funkcia, ktorá má nulovú hodnotu mimo vybraného intervalu a je symetrická okolo stredu tohto intervalu. Takáto funkcia je často používaná na výpočet váženého klzavého priemeru. Základným a najjednoduchším oknom je *obdĺžnikové okno*, ktoré hodnoty na vybranom intervale vezme s rovnakou váhou. Pre spracovanie signálu sa však najčastejšie používa Hannovo okno⁴, ktoré vezme hodnoty okolo bodu symetrie s najväčšou váhou a smerom ku okrajom okna sa váhy hodnôt postupne znižujú. Využíva trigonometrickú funkciu kosínus a definícia tejto funkcie je

$$w(u) := \begin{cases} \frac{1}{2} (1 + \cos(2\pi u)), & \text{ak } -0.5 \leq u \leq 0.5 \\ 0, & \text{inak} \end{cases} \quad (2.29)$$

Jedno z ďalších použití okna je také, že danú funkciu, v našom prípade signál, vynásobíme funkciou okna, čím dostaneme časový výsek signálu. Na obrázku 2.7 vidíme rozdiel medzi použitím obdĺžnikového a Hannovho okna.



Obr. 2.7: Vľavo hore: Obdĺžnikové okno a funkcia, na ktorú ho budeme aplikovať. Vpravo hore: Hannovo okno a daná funkcia. Dole: Výsledok súčinu funkcií.

Kedže my pracujeme s digitálnou zvukovou nakrávkou, teda funkciou diskretnou v čase, tak aj funkciu Hannovho okna potrebujeme diskretizovať. Zavedieme si dve diskretizácie Hannovho okna, s ktorými budeme neskôr pracovať. Prvou je diskretizácia

$$w(n) := \frac{1}{2} \left(1 + \cos \left(2\pi \left(\frac{n}{N_w - 1} - \frac{1}{2} \right) \right) \right), \quad n \in [0, N_w - 1], \quad (2.30)$$

kde N_w predstavuje veľkosť okna.

⁴Pomenované po matematikovi Juliusovi von Hannovi.

Druhá diskretizácia je centrovaná okolo nuly a veľkosť okna sa v tomto prípade rovná hodnote $2M + 1$. Funkciu w preto diskretizujeme ako⁵

$$w(n) := \frac{1}{2} \left(1 + \cos \left(\frac{2\pi n}{2M} \right) \right), \quad n \in [-M, M]. \quad (2.31)$$

2.6.2 Definícia diskkrétnej krátkodobej Fourierovej transformácie

Opäť budeme pracovať s poľom hodnôt vzoriek signálu, ktoré sme zadefinovali ako x v (1.6). Ako bolo spomenuté v úvode tejto sekcie, Fourierovu transformáciu budeme počítať pre časové úseky zvukovej nahrávky. Tie vytvoríme vynásobením signálu a Hannovho okna, ako sme videli na obrázku 2.7. Predstavíme si nový parameter, ktorým je veľkosť skoku H , $H \in \mathbb{N}$. Veľkosť skoku určuje krok, ktorým sa budeme posúvať po indexoch poľa x . Najmenšia veľkosť skoku je $H = 1$, kedy sa vždy posunieme o jeden index poľa x a pre vybraný interval (časový úsek) vypočítame Fourierovu transformáciu. Vybraný interval predstavuje veľkosť Hannovho okna, ktoré sme určili pre diskkrétne hodnoty v (2.30). Pre spracovávanie signálu budeme používať veľkosť skoku rovnú tretine veľkosti okna, čiže $H = N_w/3$. Definíciu diskkrétnej krátkodobej Fourierovej transformácie odvodíme pomocou diskkrétnej Fourierovej transformácie (2.27). V tomto prípade však sumu nebudeme počítať cez celú nahrávku, ako to bolo v diskkrétnej Fourierovej transformácii, ale iba pre vzorky z okna, keďže hodnoty mimo okna sú kvôli použitiu Hannovho okna nulové. Tak dostaneme definíciu diskkrétnej krátkodobej Fourierovej transformácie

$$\mathcal{X}(m, k) := \sum_{n=0}^{N_w-1} x(n + mH)w(n)e^{\frac{-2\pi i k n}{N_w}}, \quad (2.32)$$

kde $k \in [0, N_w/2]$ a m predstavuje index časového výseku. Pre každú hodnotu m , teda pre každý časový výsek, dostaneme vektor Fourierových koeficientov (komplexných čísel), dĺžky $N_w/2 + 1$. Prislúchajúce frekvencie $f(k)$ vieme dopočítať rovnako, ako sme odvodili v (2.28). Časovému indexu m zodpovedá čas

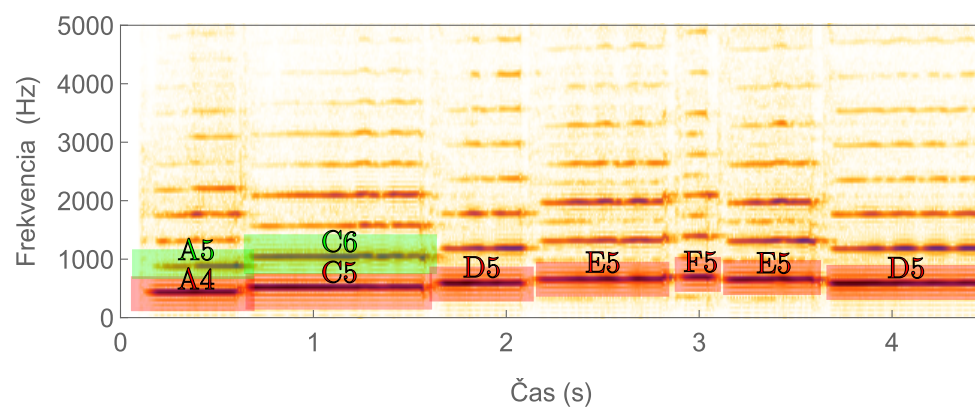
$$t(m) := \frac{mH}{F_s} \quad (2.33)$$

To, ako veľmi je zastúpená daná frekvencia v čase, zistíme pomocou funkcie, ktorú nazývame spektrogram. Spektrogram je 2D reprezentácia kvadrátu absolútnej hodnoty diskkrétnej krátkodobej Fourierovej transformácie $\mathcal{X}(m, k)$. Označujeme ho ako $\mathcal{Y}$ a definujeme

$$\mathcal{Y}(m, k) := |\mathcal{X}(m, k)|^2. \quad (2.34)$$

Spektrogram vieme zobrazit' grafom, kde vodorovná os predstavuje indexy m a zvislá index k . Pomocou výpočtu $t(m)$ podľa (2.33) vodorovnú os upravíme na časovú os a výpočtom $f(k)$ podľa (2.28) bude zvislá os zodpovedať frekvenciám. Intenzita frekvencie v zvukovej nahrávke zodpovedá intenzite farby v grafe, ako vidíme aj na obrázku 2.8, kde sme vykreslili zastúpenie frekvencií v čase pre zvukovú nahrávku piesne Greensleeves. Správnosť výsledku môžeme overiť pomocou notového zápisu (obr. 2.5) a tabuľky prislúchajúcich frekvencií (tab. 2.1). Na obrázku vidíme základnú frekvenciu každého tónu vykreslenú intenzívnejšou, tmavou farbou a zastúpenie vyšších harmonických frekvencií je taktiež viditeľné a vykreslené bledšou farbou.

⁵Pre spracovávanie signálu sa pri vzorkovacej frekvencii $F_s = 44\,100$ Hz obvykle nastavuje $M = 1024$.

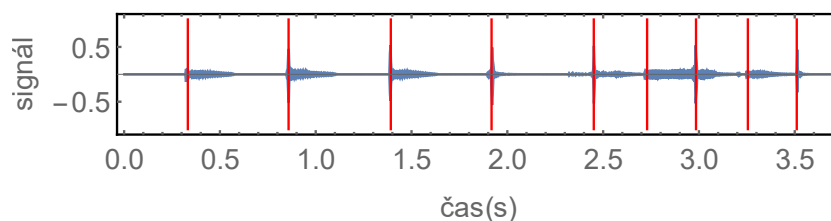


Obr. 2.8: Krátkodobá Fourierova transformácia piesne **Greensleeves**, kde na vodorovnej osi je čas a na zvislej frekvencie.

3 Detekcia nástupu tónov

V tejto kapitole sa budeme zaoberať rôznymi prístupmi detekcie nástupu výraznej zmeny signálu v zvukovej nahrávke. Výraznú zmenu, ktorú hľadáme, budeme nazývať novinka. Príkladom novinky v zvukovej nahrávke môže byť začiatok nového tónu alebo úder na bicie (obr. 3.1).

Nájdenie nástupov je prvým krokom k analýze tempa a beat trekingu. Cieľom je nájsť vhodný algoritmus, ktorý správne deteguje aj menej zjavné nástupy, napríklad v polyfónnej hudbe.¹ Výsledkom všetkých prístupov bude novinková funkcia, ktorej grafom je *novinková krivka*. Lokálne maximá tejto krivky predstavujú jednotlivé nástupy.



Obr. 3.1: Signál zvukovej nahrávky popovej piesne s názvom *Break My Heart* od Dua Lipy a zaznamenané nástupy.

3.1 Detekcia založená na energii zvuku

Prvým a najjednoduchším prístupom, je energeticky orientovaná detekcia nástupov. V popových a rockových piesňach, kde hlavné tempo udávajú bicie alebo gitara, sú začiatky tónov spojené s náhlym zvýšením amplitúdy signálu zvukovej nahrávky [2]. Ako sme hovorili v časti 1.1, amplitúda (hlasitosť) súvisí s energiou zvukovej vlny a preto sa detekcii založenej na amplitúde hovorí energeticky orientovaná detekcia. Zo signálu x teda potrebujeme nejakým spôsobom odvodiť novinkovú funkciu, ktorá bude mať lokálne maximá v časoch, kde má x prudký nárast amplitúdy.

3.1.1 Výpočet energie

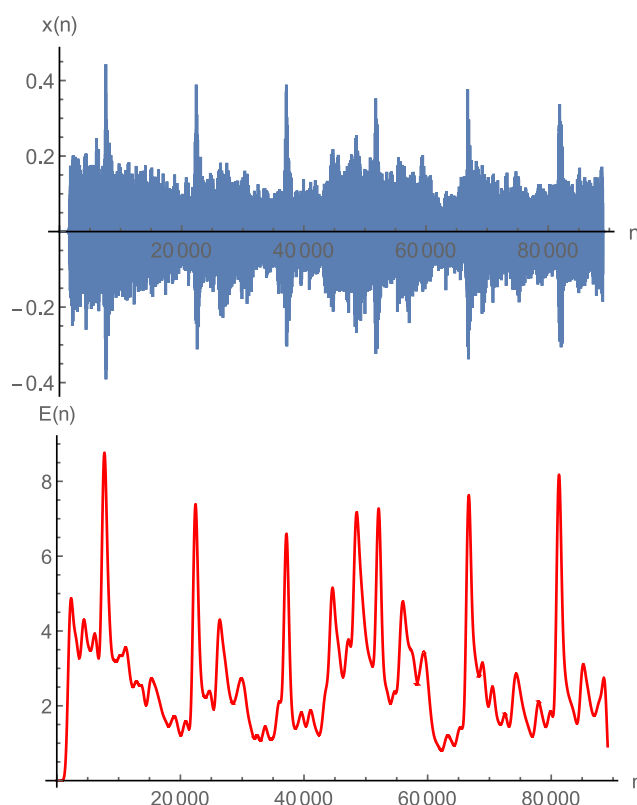
Nech x je pole hodnôt vzoriek dĺžky N , ktoré signál diskretizovaný v čase nadobúda a nech w je funkcia Hannovo okno, centrovaná okolo nuly, ktorú sme zadefinovali a diskretizovali v (2.31). Keďže nás zaujíma veľkosť amplitúdy bez ohľadu na jej znamienko,

¹Polyfónia znamená viachlas. V polyfónnej hudbe sa prekrývajú viaceré nástroje a hlasy.

prvým krokom je výpočet kvadrátu x . Ako sme videli na obrázku 1.4, hodnoty signálu sa veľmi rýchlo menia a kolíšu, preto signál zhladieme pomocou Hannovho okna a to funkciou na výpočet lokálnej energie, ktorá je definovaná ako

$$E(n) := \sum_{m=-M}^M |x(n+m)w(m)|^2, \quad n \in [0, N-1]. \quad (3.1)$$

Funkcia lokálnej energie $E(n)$ je (až na násobok normalizačnou konštantou) vážený kľzavý priemer funkcie $x^2(n)$, pričom váhy sú $w^2(m)$. Tento výpočet sa nazýva aj výpočtom obálky energie.



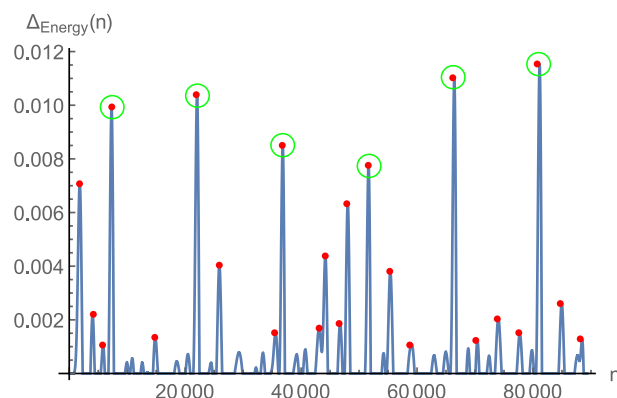
Obr. 3.2: Hore: Pole hodnôt vzoriek signálu zvukovej nahrávky piesne **Hymn For The Weekend** od **Coldplay**. Dole: Výpočet obálky energie.

Hľadáme časy nástupov, kde je prudký nárast (zmena) energie. Zmenu energie, teda funkcie $E(n)$, vyjadruje jej derivácia. V našom prípade spravíme konečnú diferenciu dvoch po sebe nasledujúcich hodnôt a vezmeme do úvahy iba kladné hodnoty, pretože tie hovoria o náraste energie, pričom záporné hovoria o jej poklese. Výsledkom je novinková funkcia energie

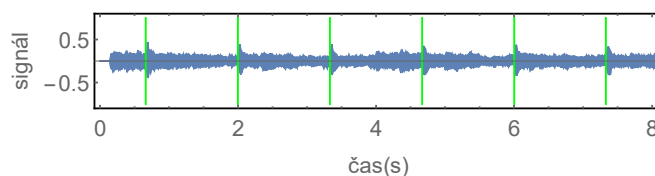
$$\Delta_{Energy}(n) := |E(n+1) - E(n)|_{\geq 0}. \quad (3.2)$$

Tak dostaneme pole dĺžky N , ktorého grafom je novinková krivka. Ďalším krokom je nájdenie lokálnych maxím, ktoré zodpovedajú začiatkom noviniek (obr. 3.3).

Dôležité sú pre nás indexy, v ktorých sme našli lokálne maximá. Tie po delení vzorkovacou frekvenciou reprezentujú čas zaznamenaných nástupov, pozri (1.4).

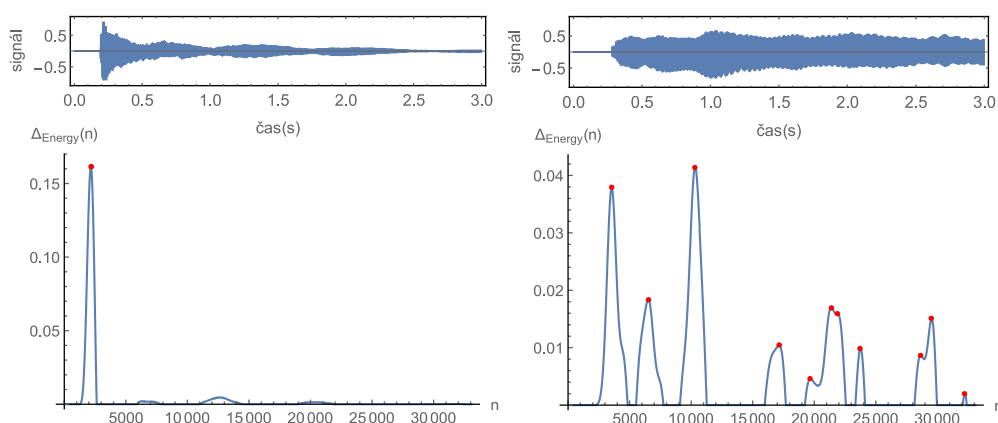


Obr. 3.3: Novinková krivka (modrá) piesne Hymn For The Weekend od Coldplay a jej lokálne maximá (červená). Najvýraznejšie lokálne maximá vyznačené zelenou farbou pri slúchajú nástupom noviniek, ako vidíme na obrázku 3.4.



Obr. 3.4: Signál zvukovej nahrávky piesne Hymn For The Weekend od Coldplay a zaznamenané nástupy (zelená), ktoré sme určili vďaka idexom lokálnych maxim novinkovej krivky. Sonifikáciu nástupov si môžeme vypočítať [tu](#).

Problém pri detekcii založenej na energii signálu nastáva vtedy, ak začiatok tónu neznamená náhle zvýšenie energie alebo energia počas trvania tónu kolíše. Napríklad pre strunové nástroje sa po nasadení tónu energia pozvoľne zvyšuje a neskôr osciluje, čo spôsobí nájdenie mnohých nástupov (obr. 3.5, vpravo). Metóda nezaznamená nástupy s nízkou amplitúdou, pretože hodnota energie bude príliš nízka.



Obr. 3.5: Porovnanie signálu a novinkových kriviek pri zaznení tónu na [klavíri](#) (vľavo) a [husliach](#) (vpravo).

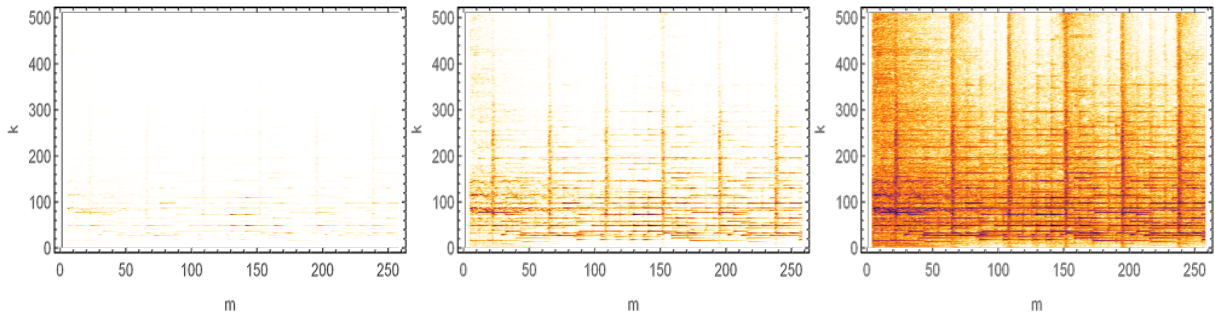
3.2 Detekcia založená na spektre zvuku

Ako už bolo spomenuté, detekcia založená na energii je vhodná hlavne pre popové a rockové piesne. Pre zložitejšie zvukové nahrávky, napríklad klasickej hudby, však detekcia nástupov môže byť náročnejšia, a preto si predstavíme ďalší prístup detekcie, ktorý je založený na spektre zvuku. Budeme hľadať časy, v ktorých sa náhle zmení frekvenčné spektrum signálu, a tak určíme jednotlivé nástupy.

Prvým krokom je výpočet diskkrétnej krátkodobej Fourierovej transformácie (2.32), pomocou čoho dostaneme hlavnú informáciu, ktorou je hodnota $\mathcal{X}(m, k)$. Ďalší krok je dôležitý napríklad pre piesne, v ktorých tempo udávajú sláčikové nástroje a ich hlasitosť je veľmi nízka. Vtedy má intenzita zodpovedajúcich frekvencií nízku hodnotu, a preto by ich nástupy nemuseli byť zaznamenané. Tento problém vyriešime pomocou operácie s názvom *logaritmická kompresia*. Použitím tejto operácie vieme prítomným frekvenciám s menšou intenzitou zvýšiť ich zastúpenie a naopak frekvenciám, ktoré sú výrazne zastúpené, ich zastúpenie zmenšiť. Logaritmickú kompresiu definujeme ako

$$\mathcal{L} := \Gamma_\gamma(|\mathcal{X}|) = \log(1 + \gamma|\mathcal{X}|), \quad (3.3)$$

kde za hodnotu $|\mathcal{X}|$ budeme dosádzať absolútnu hodnotu komplexného čísla $\mathcal{X}(m, k)$. Táto hodnota vyjadruje, ako veľmi je zastúpená frekvencia v zvukovej nahrávke prislúchajúca danému k . Čím väčšia bude hodnota γ , tým viac frekvencie zrovnomeníme. Pre príliš vysokú hodnotu γ by sa mohlo stať, že zosilníme aj také frekvencie, ktoré predstavujú nerelevantné komponenty (alebo aj šum), a preto je rozumné zvoliť $\gamma = 1$.



Obr. 3.6: Použitie logaritmickkej kompresie viditeľné na spektrograme piesne **Hymn For The Weekend** od **Coldplay**. Vľavo: Spektrogram vytvorený z krátkodobej Fourierovej transformácie. V strede: Použitie logaritmickkej kompresie s parametrom $\gamma = 1$. Na pravo: Použitie logaritmickkej kompresie s parametrom $\gamma = 1000$.

Nasledovný krok je podobný kroku, ktorý sme použili v energeticky orientovanej detekcii nástupov. Hľadáme časy nástupov, kde došlo k zmene frekvencie. Pre zistenie zmeny vypočítame diferenciu, pričom nás zaujímajú iba kladné hodnoty, ktoré predstavujú nástup nových frekvencií, ktorých intenzita vzrastie. Spravíme sumu cez všetky k , ktoré prislúchajú frekvenciám, a tak dostaneme novinkovú funkciu

$$\Delta_{Spectral}(m) := \sum_{k=0}^{N_w/2} |\mathcal{L}(m+1, k) - \mathcal{L}(m, k)|_{\geq 0}. \quad (3.4)$$

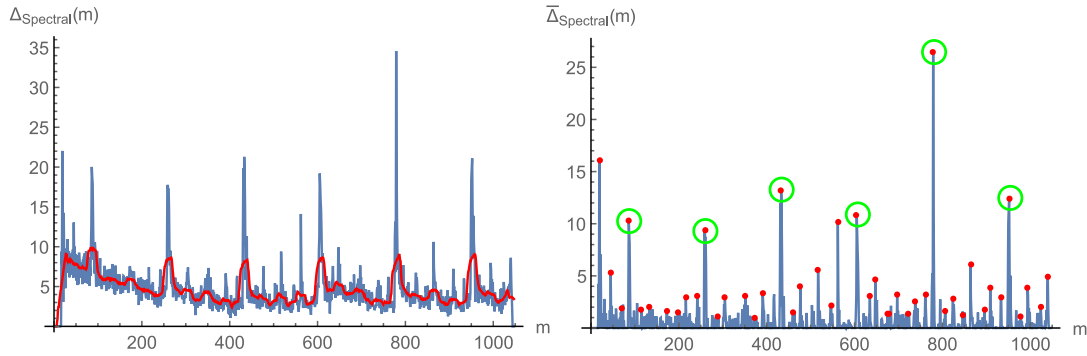
Následne môžeme novinkovú funkciu upraviť, čím dosiahneme lepšiu a jednoduchšiu detekciu nástupov. Vypočítame jej kľzavý priemer,

$$\mu(m) := \frac{1}{2M_\mu + 1} \sum_{m_\mu=-M_\mu}^{M_\mu} \Delta_{Spectral}(m + m_\mu), \quad (3.5)$$

kde $2M_\mu + 1$ predstavuje veľkosť intervalu, na ktorom priemer počítame. Odpočítaním kľzavého priemeru $\mu(m)$ od novinkovej funkcie $\Delta_{Spectral}(m)$ dostaneme vylepšenú novinkovú funkciu, ktorú označíme ako $\bar{\Delta}_{Spectral}(m)$. Z rovnakých dôvodov ako aj pri pôvodnom výpočte novinkovej funkcie, vezmeme do úvahy iba kladné hodnoty tejto funkcie

$$\bar{\Delta}_{Spectral}(m) := |\Delta_{Spectral}(m) - \mu(m)|_{\geq 0}. \quad (3.6)$$

Výsledkom je novinková krivka, ktorú vidíme na obrázku 3.7, vpravo.



Obr. 3.7: Vľavo: Novinková funkcia (modrá), ktorej hodnoty sme vypočítali pomocou vzorca (3.4) spolu s vypočítaným kľzavým priemerom (červená). Vpravo: Vylepšená novinková funkcia (modrá), ktorú sme dostali po odpočítaní kľzavého priemeru a jej lokálne maximá (červená). Novinkové funkcie sme vytvorili k zvukovej nahrávke piesne **Hymn For The Weekend** od **Coldplay**. Najvýraznejšie lokálne maximá, vyznačené zelenou farbou prislúchajú nástupom noviniiek, ktoré zodpovedajú tým na obrázku 3.4.

Rovnako ako aj v detekcii založenej na energii, posledným krokom je nájdenie lokálnych maxím novinkovej krivky, ktoré predstavujú nástupy. Čas týchto nástupov vypočítame pomocou vzorca (2.33), kam namiesto indexu m dosadíme indexy novinkovej krivky, v ktorých sme našli lokálne maximá. Ako vidíme aj na obrázku 3.7, z veľkého počtu lokálnych maxím sme vybrali iba tie najvýraznejšie. Pri výbere môžeme použiť prahovú hodnotu, pričom však vo všeobecnosti môže dôjsť k problému s jej správnou voľbou a následne so správnym výberom výrazných maxím. V ďalších kapitolách si preto ukážeme, ako z množstva lokálnych maxím efektívnejšie dostať len tie, ktoré potrebujeme.

3.3 Detekcia založená na fáze

V predchádzajúcom prístupe detekcie založenej na spektre zvuku sme použili veľkosť komplexného čísla pre získanie novinkovej funkcie a určenie nástupov. Fáza Fourierovho koeficientu je však tiež dôležitým zdrojom informácií, preto si v tejto sekcii predstavíme postup detekcie nástupov založenej na fáze.

Rovnako ako predtým, prvým krokom je výpočet diskkrétnej krátkodobej Fourierovej transformácie (2.32), čoho výsledkom je hodnota $\mathcal{X}(m, k)$, kde $k \in [0, N_w/2]$ a m označuje index časového výseku. Toto komplexné číslo vieme zapísať pomocou polárnych súradníc (2.11) v tvare

$$\mathcal{X}(m, k) = |\mathcal{X}(m, k)|e^{2\pi i\varphi(m, k)}, \quad (3.7)$$

kde $|\mathcal{X}(m, k)|$ je veľkosť komplexného čísla a uhol $\gamma = 2\pi\varphi(m, k)$, $\gamma \in [0, 2\pi)$. Fázu vyjadríme ako $\varphi(m, k) = \frac{\gamma}{2\pi}$, $\varphi(m, k) \in [0, 1)$, čím zistíme, o akú časť periódy musí byť sínusoida s danou frekvenciou posunutá, aby bola najviac podobná signálu. Ďalej vypočítame spätnú diferenciu prvého rádu

$$\varphi'(m, k) := \varphi(m, k) - \varphi(m - 1, k) \quad (3.8)$$

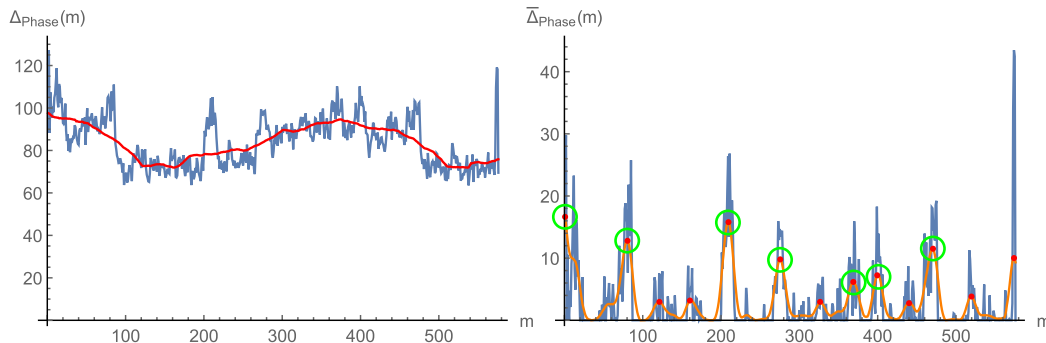
a následne spätnú diferenciu druhého rádu

$$\varphi''(m, k) := \varphi'(m, k) - \varphi'(m - 1, k). \quad (3.9)$$

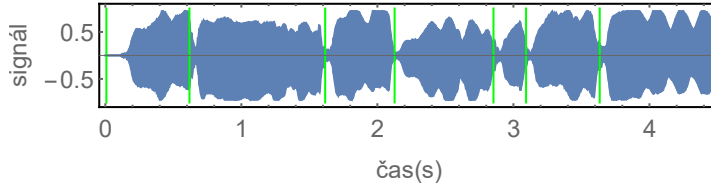
Budeme vychádzať z predpokladu, že v čase, kde sa nenachádza žiadna novinka, fáza narastá lineárne, a preto hodnoty spätnej diferencie prvého rádu majú konštantnú hodnotu a pre hodnoty spätnej diferencie druhého rádu preto platí $\varphi''(m, k) \approx 0$. V mieste novinky však dôjde k výraznej zmene hodnoty fázy a preto výrazná odchýlka v hodnotách $\varphi''(m, k)$ je vhodným indikátorom detekcie nástupu. Tak určíme novinkovú funkciu, kde vypočítame sumu cez všetky hodnoty k (prislúchajúce frekvenciám) z absolútnej hodnoty spätnej diferencie druhého rádu

$$\Delta_{Phase}(m) := \sum_{k=0}^{N_w/2} |\varphi''(m, k)|. \quad (3.10)$$

Aj túto novinkovú funkciu môžeme ďalej spracovať, pre lepšiu detekciu nástupov, napríklad výpočtom jej klzavého priemeru. Odpočítaním klzavého priemeru od novinkovej funkcie dostaneme vylepšenú novinkovú funkciu. Táto novinková funkcia však stále nie je vhodná pre detekciu kvôli veľkému množstvu lokálnych maxím (Obr. 3.8, vpravo), preto opäť vypočítame jej klzavý priemer, pre ktorý nájdeme lokálne maximá, prislúchajúce nástupom. Indexy týchto lokálnych maxím po dosadení do vzorca (2.33) zodpovedajú časom zaznamenaných nástupov.



Obr. 3.8: Vľavo: Novinková funkcia (modrá) piesne **Greensleeves**, ktorej hodnoty sme vypočítali pomocou vzorca (3.10) a jej klzavý priemer (červená). Vpravo: Vylepšená novinková funkcia (modrá) a jej klzavý priemer (oranžová), pre ktorý nájdeme lokálne maximá (červená), prislúchajúce nástupom.



Obr. 3.9: Signál zvukovej nahrávky piesne Greensleeves a zaznamenané nástupy (zelená), ktoré sme určili vďaka indexom lokálnych maxím z obrázka 3.8, vpravo. Výslednú sonifikáciu nástupov si môžeme vypočítať [tu](#).

3.4 Detekcia založená na spektre a fáze

Posledným prístupom detekcie, ktorý si predstavíme, je detekcia založená na spektre a fáze, využijeme tu teda aj informáciu o veľkosti Fourierovho koeficientu a taktiež aj informáciu o fáze. Prvý krok je preto rovnaký, ako aj v predošlých prístupoch detekcie. Našou hlavnou informáciou je opäť Fourierov koeficient $\mathcal{X}(m, k)$, ktorý získame výpočtom diskkrétnej krátkodobej Fourierovej transformácie (2.32). Predpokladom tohto prístupu detekcie je, že fázové diferencie a veľkosť Fourierovho koeficientu zostávajú v čase, kde sa nenachádza žiadna novinka, viac-menej konštantné. Hlavnou myšlienkou je preto pomocou hodnoty $\mathcal{X}(m, k)$ vypočítať odhad $\hat{\mathcal{X}}(m+1, k)$, čiže hodnoty v nasledujúcom časovom indexe a následné porovnanie odhadu so skutočným Fourierovým koeficientom. Odhad definujeme ako

$$\hat{\mathcal{X}}(m+1, k) = |\mathcal{X}(m, k)| e^{2\pi i(\varphi(m, k) + \varphi'(m, k))}, \quad (3.11)$$

kde k hodnote $\varphi(m, k)$ pripočítame hodnotu $\varphi'(m, k)$, čím v stabilných miestach získame takmer správny odhad hodnoty $\mathcal{X}(m+1, k)$. Ďalej vypočítame veľkosť rozdielu hodnoty odhadu $\hat{\mathcal{X}}(m+1, k)$ a skutočnej hodnoty $\mathcal{X}(m+1, k)$.

$$\mathcal{X}'(m+1, k) = |\hat{\mathcal{X}}(m+1, k) - \mathcal{X}(m+1, k)| \quad (3.12)$$

Prítomnosť novinky spôsobí veľkú odchýlku odhadu od skutočnej hodnoty, vďaka čomu zaznamenáme nástupy novínok signálu. Ďalej hodnotu $\mathcal{X}'(m, k)$ rozdelíme na dve zložky

$$\mathcal{X}^+(m, k) = \begin{cases} \mathcal{X}'(m, k), & \text{ak } |\mathcal{X}(m, k)| > |\mathcal{X}(m-1, k)| \\ 0, & \text{inak} \end{cases} \quad (3.13)$$

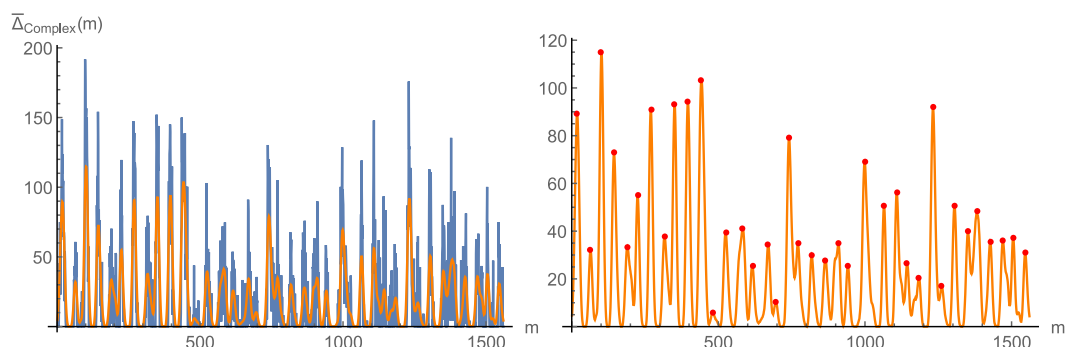
$$\mathcal{X}^-(m, k) = \begin{cases} \mathcal{X}'(m, k), & \text{ak } |\mathcal{X}(m, k)| \leq |\mathcal{X}(m-1, k)| \\ 0, & \text{inak} \end{cases} \quad (3.14)$$

Hodnota $\mathcal{X}^+(m, k)$ zodpovedá nárastu a $\mathcal{X}^-(m, k)$ poklesu hodnôt $|\mathcal{X}(m, k)|$, čiže zastúpenia frekvencie $f(k)$. Keďže nás zaujíma iba príchod nového tónu, kedy veľkosť komplexného čísla rastie, vezmeme do úvahy iba hodnoty $\mathcal{X}^+(m, k)$, ktoré použijeme pre určenie novinkovej funkcie. Novinkovú funkciu $\Delta_{Complex}(m)$ definujeme ako

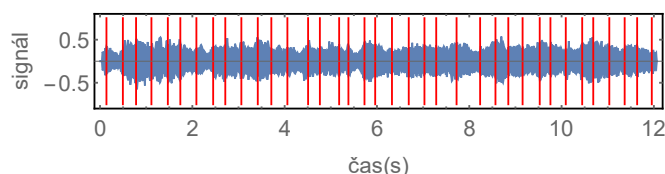
$$\Delta_{Complex}(m) := \sum_{k=0}^{N_w/2} \mathcal{X}^+(m, k). \quad (3.15)$$

Novinkovú funkciu ďalej spracujeme rovnako ako v predchádzajúcom prístupe, pre dosiahnutie lepšej detekcie nástupov. Odpočítaním kľavého priemeru od novinkovej funkcie

dostaneme vylepšenú novinkovú funkciu. Táto novinková funkcia však stále nie je vhodná pre detekciu (veľké množstvo lokálnych maxim), preto opäť vypočítame jej klzavý priemer, pre ktorý nájdeme lokálne maximá, prislúchajúce nástupom. Dosadením indexov lokálnych maxim do vzorca (2.33) dostaneme časy zaznamenaných nástupov.



Obr. 3.10: Vľavo: Vylepšená novinková funkcia (modrá), a jej klzavý priemer (oranžová). Vpravo: Klzavý priemer (oranžová), pre ktorý nájdeme lokálne maximá (červená), prislúchajúce nástupom. Novinkové funkcie sme vytvorili k zvukovej nahrávke skladby s názvom *Waltz No. 2* od Dmitrija Shostakovicha.



Obr. 3.11: Signál zvukovej nahrávky skladby *Waltz No. 2* od Dmitrija Shostakovicha a zaznamenané nástupy (červená). Sonifikáciu nástupov si môžeme vypočítať [tu](#).

Napriek tomu, že v tejto detekcii používame informáciu zo spektra aj fázy, môžeme nájsť aj také lokálne maximá, ktoré nezodpovedajú nástupom. Z tohto dôvodu s výsledkom ešte nie sme spokojní a v nasledujúcej kapitole si predstavíme ďalší krok beat trackingu, kde použitím novinkovej funkcie určíme tempo pre danú pieseň.

4 Analýza tempa

V tejto kapitole sa budeme zaoberať určením tempa zvukovej nahrávky. Tempo sa udáva v počte úderov za minútu, čo značíme ako BPM (beats per minute). Pri počúvaní hudby si človek často dupe nohou do rytmu piesne, a tak podvedome určí tempo piesne. V popových a rockových piesňach je tempo väčšinou pravidelné, avšak v klasickej alebo folklórnej hudbe sa môže tempo v priebehu piesne meniť, môže zrýchľovať alebo spomaľovať, a vtedy je analýza tempa náročnejšia. Tempo budeme označovať písmenom τ . Hodnota $\tau = 180$ BPM, teda 180 úderov za minútu zodpovedá 3 úderom za sekundu, čiže hodnote $f = 3$ Hz. Preto pomocou frekvencie úderov f vieme vypočítať hodnotu tempa ako

$$\tau = 60f. \quad (4.1)$$

Rovnako ako pri zaznení tónu, kedy boli prítomné okrem základnej frekvencie aj vyššie harmonické frekvencie, sú v zvukovej nahrávke okrem základného tempa τ_z aj jeho celočíselné násobky, čo znamená hodnoty $2\tau_z, 3\tau_z \dots$

Pre analýzu tempa budeme brať do úvahy nasledujúce predpoklady. Prvým predpokladom je, že údery sa vyskytujú v rovnakom čase ako aj nástupy, ktoré sme pomocou rôznych typov detekcie určili v predchádzajúcej kapitole. Ďalším predpokladom je pravidelnosť (periodickosť) úderov, aspoň lokálne pre určitý časový úsek zvukovej nahrávky. Za týchto predpokladov môžeme pre analýzu tempa použiť Fourierovu transformáciu novinkovej funkcie. Vďaka tomuto výpočtu zanalyzuje lokálne periodické správanie sa novinkovej funkcie a vypočítame hodnotu tempa danej zvukovej nahrávky. Tempo najprv určíme globálne pomocou diskkrétnej Fourierovej transformácie a následne aj lokálne pomocou krátkodobej Fourierovej transformácie.

4.1 Tempové spektrum

V sekcii 2.2 sme pomocou diskkrétnej Fourierovej transformácie vypočítali zastúpenie prítomných frekvencií tým, že sme porovnávali signál so sínusoidami s rôznymi frekvenciami a fázami. Rovnaký princíp použijeme aj pre určenie tempa. Namiesto signálu však budeme so sínusoidami porovnávať novinkovú funkciu, ktorú sme získali detekciou nástupov. Majme novinkovú funkciu $\Delta(n)$ a počet hodnôt novinkovej funkcie, ktorý označíme ako N_Δ . Pomocou (2.22) dostaneme diskkrétne Fourierovu transformáciu pre novinkovú funkciu

$$\hat{\Delta}(\bar{f}) = \sum_{n=0}^{N_\Delta-1} \Delta(n) e^{-2\pi i \bar{f} n}. \quad (4.2)$$

Túto definíciu chceme upraviť, aby hodnota τ bola vstupom tejto funkcie. Preto potrebujeme zistiť, aké $\bar{f}$ máme zvoliť, aby sme dostali hodnoty τ v BPM. Tempo piesne sa

pre praktické aplikácie väčšinou určuje pre celočíselné hodnoty na intervale $\tau \in [30, 600]$. Hodnotu $\bar{f}$ v tomto prípade definujeme ako

$$\bar{f} = fT_{\Delta}, \quad (4.3)$$

kde T_{Δ} predstavuje vzorkovaciu periódu medzi dvoma vzorkami novinkovej funkcie. Pre novinkovú funkciu získanú pomocou detekcie založenej na energii máme $T_{\Delta} = T_s$, pretože počet vzoriek signálu sa rovná počtu vzoriek novinkovej funkcie. V iných prístupoch detekcie, ktoré sme popísali v tretej kapitole, sa pre získanie novinkovej funkcie počítala krátkodobá Fourierova transformácia, pričom sa používal parameter H označujúci veľkosť skoku. Preto T_{Δ} vypočítame ako

$$T_{\Delta} = HT_s = \frac{H}{F_s}. \quad (4.4)$$

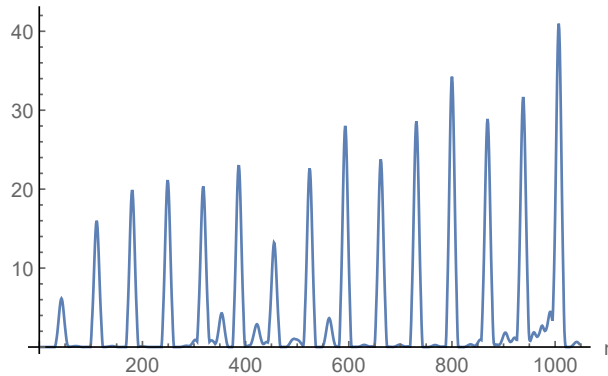
Ďalej budeme používať T_{Δ} , počítané pomocou (4.4), pričom ak pracujeme s energetickou novinkovou krivkou tak stačí vo vzorci (4.4) zobrať $H = 1$. Hodnotu f vyjadríme zo vzorca (4.1) ako $f = \frac{\tau}{60}$ a spolu s T_{Δ} dosadíme do (4.3), čím dostaneme

$$\bar{f} = \frac{\tau H}{60F_s}. \quad (4.5)$$

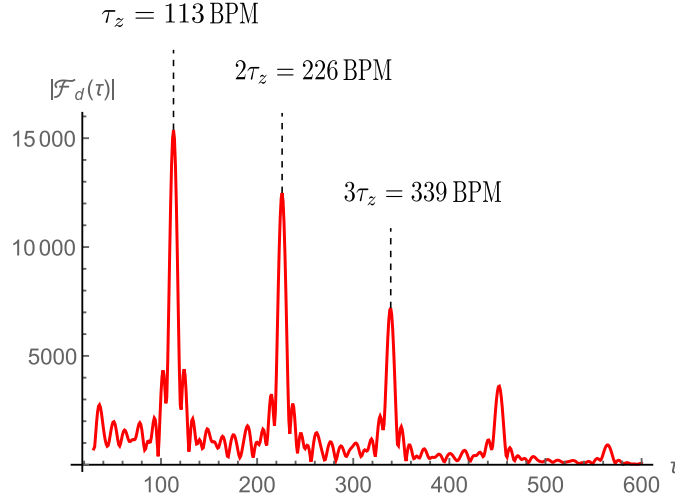
Dosadením (4.5) do (4.2) dostaneme definíciu diskkrétnej Fourierovej transformácie pre novinkovú funkciu, ktorú označíme ako $\mathcal{F}_d(\tau)$

$$\mathcal{F}_d(\tau) = \hat{\Delta} \left(\frac{\tau H}{60F_s} \right) = \sum_{n=0}^{N_{\Delta}-1} \Delta(n) e^{\frac{-2\pi i \tau H n}{60F_s}}. \quad (4.6)$$

Tempové spektrum predstavuje graf, kde na vodorovnej osi sú hodnoty τ a na zvislej osi sú intenzity prislúchajúce daným tempám, ktoré vypočítame ako $|\mathcal{F}_d(\tau)|$. Na obrázku 4.2 vidíme hlavné tempo piesne $\tau_z = 113$ BPM, ktoré je najviac zastúpené. Okrem hlavného tempa vidíme na obrázku výrazné aj jeho celočíselné násobky, teda hodnoty $2\tau_z = 226$ BPM, $3\tau_z = 339$ BPM...



Obr. 4.1: Novinková krivka zvukovej nahrávky piesne **Break My Heart** od **Dua Lipy**. Po výpočte Fourierovej transformácie tejto novinkovej funkcie dostaneme hodnoty tempového spektra, ktoré vidíme na obrázku 4.2.



Obr. 4.2: Tempové spektrum pre zvukovú nahrávku piesne [Break My Heart](#) od Dua Lipy prislúchajúce novinkovej funkcii, ktorej hodnoty vidíme na obrázku 4.1.

4.2 Tempogram z krátkodobej Fourierovej transformácie

V tejto sekcii popíšeme výpočet krátkodobej Fourierovej transformácii pre novinkovú funkciu, čím získame informáciu nielen o prítomných tempách, ale ponecháme si aj informáciu o čase zvukovej nahrávky. Postup výpočtu bude rovnaký ako v predchádzajúcej sekcii. Pre diskretný signál $x(n)$ vypočítame novinkovú funkciu $\Delta(n)$. Použijeme diskretizované Hannovo okno (2.31), centrované okolo nuly, s veľkosťou okna M_τ pre vytvorenie časových úsekov novinkovej funkcie. Vo výpočte budeme používať veľkosť skoku, ktorú označíme ako H_τ . Veľkosť okna M_τ a veľkosť skoku H_τ sa nemusia zhodovať s hodnotami M a H , ktoré sme použili pre výpočet Fourierovej transformácie signálu na získanie novinkovej funkcie. Na základe (2.32) krátkodobú Fourierovu transformáciu novinkovej funkcie definujeme ako

$$\mathcal{F}(l, \bar{f}) = \sum_{m=-M_\tau}^{M_\tau} \Delta(lH_\tau + m)w(m)e^{-2\pi i \bar{f}(lH_\tau + m)}. \quad (4.7)$$

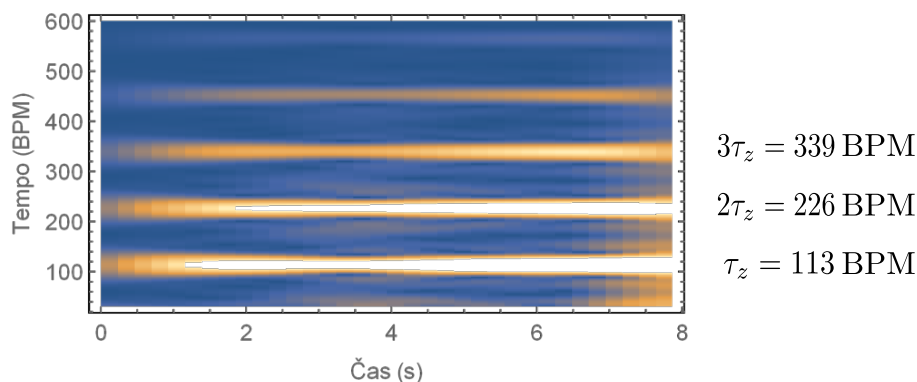
Konverziou hodnoty $\bar{f}$ na tempo τ založenou na (4.5), definujeme Fourierov tempogram

$$\mathcal{T}(l, \tau) = \left| \mathcal{F}\left(l, \frac{\tau H}{60 F_s}\right) \right|. \quad (4.8)$$

Hodnoty tempogramu vieme zobrazit grafom, kde na vodorovnej osi je čas a na zvislej osi sú tempá. Časovému indexu l zodpovedá čas

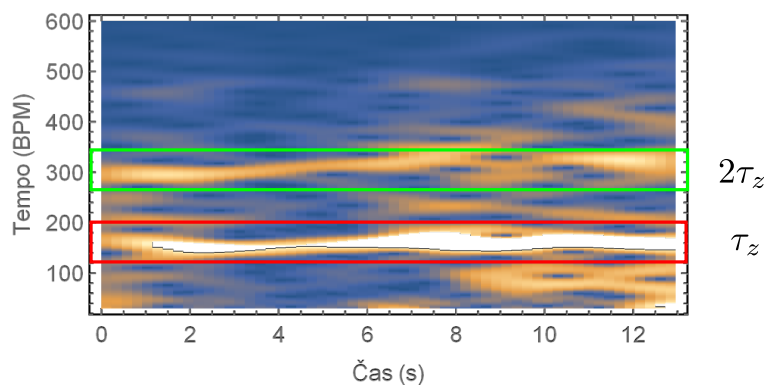
$$t(l) = nT_\Delta = lH_\tau \frac{H}{F_s}. \quad (4.9)$$

Na obrázku 4.3 vidíme tempogram prislúchajúci novinkovej funkcii, ktorej hodnoty vidíme na obrázku 4.1. Oranžové (biele) pásy na tempograme predstavujú tempá prítomné v zvukovej nahrávke, pričom čím je farba bledšia tým je tempo intenzívnejšie. Na tempograme vidíme základné tempo piesne a jeho celočíselné násobky a taktiež vidíme, že tempo sa v priebehu piesne nemení.



Obr. 4.3: Tempogram pre zvukovú nahrávku piesne **Break My Heart** od **Dua Lipy**. Hodnoty τ sú z intervalu $\tau \in [30, 600]$.

Výhodou krátkodobej Fourierovej transformácie novinkovej funkcie je, že ju počítame pre časové úseky, a teda dokážeme zaznamenať tempo aj vtedy, keď pieseň zrýchľuje alebo spomaľuje. Na obrázku 4.4 vidíme príklad meniaceho sa základného tempa, ktorého hodnoty sa pohybujú v intervale $\tau_z \in [150, 200]$.



Obr. 4.4: Tempogram pre zvukovú nahrávku skladby s názvom **In the Hall of the Mountain King** od autora **Edvarda Griega**. Na tempograme okrem základného tempa vidíme aj jeho dvojnásobok.

5 Beat tracking

V poslednej kapitole sa zameriame na posledný krok beat trackingu, pričom využijeme vedomosti získané z predchádzajúcich kapitol. Pomocou Fourierovej transformácie novinkovej funkcie, kedy sme porovnávali novinkovú funkciu s rôznymi sínusoidami, sme vypočítali hodnoty tempového spektra, prípadne tempogramu. Vieme teda určiť, aké tempo je v piesni najviac zastúpené. Túto informáciu teraz použijeme na to, aby sme skonštruovali tzv. PLP funkciu, ktorá reprezentuje periodickosť novinkovej krivky, a indexy lokálnych maxím tejto funkcie predstavujú konečný výsledok, ktorým sú časy úderov. Čas úderov najprv určíme pomocou tempa získaného diskretnou Fourierovou transformáciou a neskôr aj pomocou tempa určeného diskretnou krátkodobou Fourierovou transformáciou.

5.1 Beat tracking pomocou tempového spektra

Prvým krokom je nájdenie lokálnych maxím funkcie $|\mathcal{F}_d(\tau)|$, ktorých indexy prislúchajú najviac zastúpeným tempám v zvukovej nahrávke. Hodnotu najviac zastúpeného tempa označíme ako τ_p . K danému tempu τ_p nájdeme zodpovedajúcu funkciu sínusoidy pomocou spätnej Fourierovej transformácie. V analytickom tvare Fourierovej transformácie signálu danej frekvencii zodpovedá funkcia sínusoidy $\cos_{f,\varphi}(t)$, daná vzťahom (2.2). Rovnako aj tempu τ_p v analytickom tvare prislúcha funkcia sínusoidy v tvare

$$\kappa(t) = \cos(2\pi(ft - \varphi)). \quad (5.1)$$

Roznásobením zátvorky dostaneme

$$\kappa(t) = \cos(2\pi ft - 2\pi\varphi). \quad (5.2)$$

Funkciu upravíme pomocou vzťahu (2.16) a f vyjadríme zo vzorca (4.1) pre dané tempo $\tau = \tau_p$. Funkciu sínusoidy tak zapíšeme v tvare

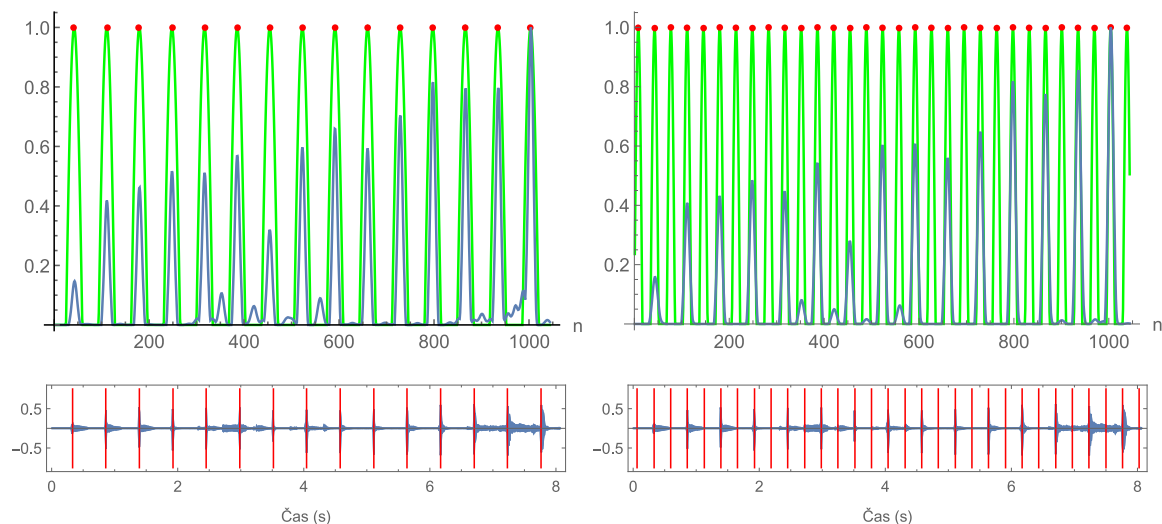
$$\kappa(t) = \cos\left(2\pi\frac{\tau_p}{60}t + \gamma\right), \quad (5.3)$$

kde hodnotu γ vypočítame ako argument komplexného čísla (2.11), ktoré prislúcha tempu τ_p a je výsledkom Fourierovej transformácie. Ďalej potrebujeme funkciu diskretizovať pre hodnoty $t = nT_\Delta$, kde $T_\Delta = \frac{H}{F_s}$. Funkciu orežeme iba pre kladné hodnoty a dosadením za t dostaneme diskkrétne hodnoty funkcie sínusoidy $\kappa(n)$, ktorá zodpovedá tempu τ_p .

$$\kappa(n) = \left| \cos\left(\frac{2\pi\tau_p n H}{60F_s} + \gamma\right) \right|_{\geq 0}, \quad n \in [0, N_\Delta - 1]. \quad (5.4)$$

Príklad funkcie $\kappa(n)$ vidíme na obrázku 5.1.

Posledným krokom je nájdenie lokálnych maxím funkcie $\kappa(n)$. Čas úderov dostaneme dosadením do vzorca (2.33), kam namiesto indexu m dosadíme indexy funkcie $\kappa(n)$, v ktorých sme našli lokálne maximá.

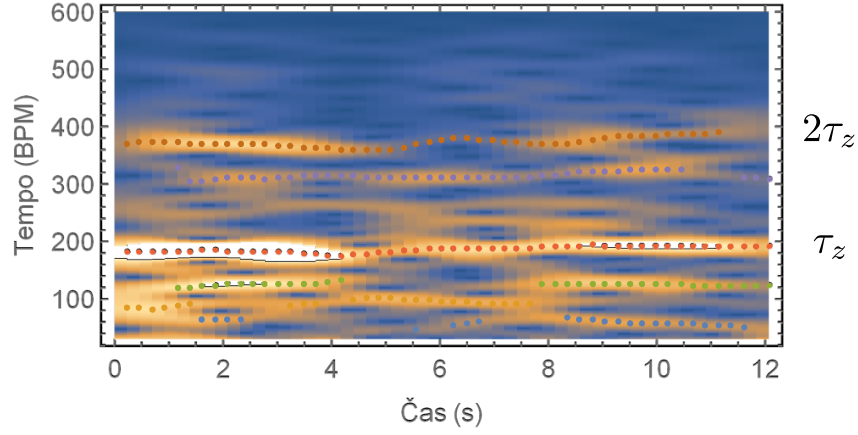


Obr. 5.1: Hore: Funkcie sínusoidy $\kappa(n)$ (zelená) prislúchajúce základnému tempu $\tau_z = 113$ BPM (vľavo) a dvojnásobku základného tempa $\tau_z = 226$ BPM (vpravo) spolu s novinkovou krivkou (modrá) a lokálnymi maximami (červená) piesne Break My Heart od Dua Lipy. Dole: Signál zvukovej nahrávky piesne spolu s nájdenými údermi. Na obrázku 4.2 si môžeme pozrieť tempové spektrum tejto piesne, pomocou ktorého sme určili základné tempo. Sonifikované údery základného tempa si môžeme vypočítať tu: [Dua Lipa-Break my Heart, 113 BPM](#) a sonifikované údery dvojnásobného tempa tu: [Dua Lipa-Break my Heart, 226 BPM](#).

Nevýhodou beat trackingu pomocou tempového spektra je, že určujeme časy úderov konkrétneho tempa pre celú zvukovú nahrávku, čím nedokážeme zachytiť zmenu tempa v priebehu piesne. Preto tento spôsob beat trackingu je vhodný iba pre piesne, v ktorých sa tempo nemení.

5.2 PLP funkcia

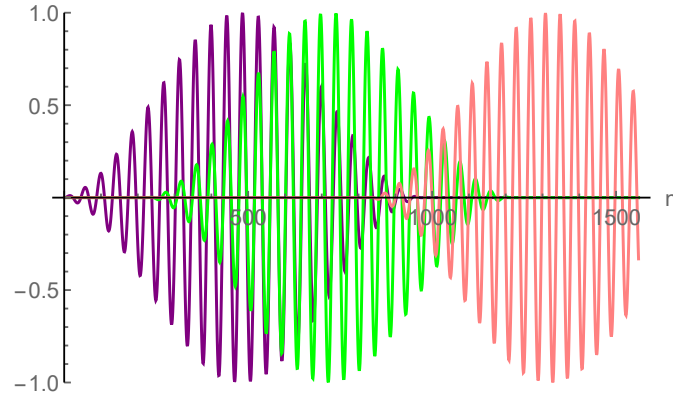
Pre piesne, v ktorých sa tempo v čase mení, je vhodnejšie vypočítať krátkodobú Fourierovu transformáciu novinkovej funkcie, a tak získať informáciu o prítomných tempách v čase. Tak ako aj v predchádzajúcej sekcii, aj teraz určíme funkciu sínusoidy pre dané tempo. V tomto prípade, ju ale určíme pre každý časový výsek použitím Hannovho okna, definovaného rovnako ako aj pre výpočet Fourierovho tempogramu s veľkosťou okna $2M_\tau + 1$. Najdôležitejšou myšlienkou je následné sčítanie týchto sínusoid, čoho výsledkom bude jedna funkcia, ktorú nazývame PLP (Predominant Local Pulse) funkcia, ktorá bude zachytávať lokálne tempo nahrávky. Prvým krokom je nájdenie lokálnych maxím funkcie $\mathcal{T}(l, \tau)$, ktorej hodnoty dostaneme pomocou vzorca (4.8). Indexy lokálnych maxím predstavujú hodnoty τ_p pre každý časový výsek, ako vidíme aj na obrázku 5.2.



Obr. 5.2: Tempogram skladby **Waltz No. 2** od **Dmitrija Shostakovicha**. Na tempograme vidíme, že dominantné tempo má hodnotu zhruba $\tau_z = 180$ BPM, ale jeho hodnota sa mení v čase. Vyznačené body predstavujú nájdené tempá τ_p v jednotlivých časových krokoch. Okrem základného tempa vidíme aj jeho dvojnásobok a pretože ide o pomerne komplikovanú polyfonickú skladbu, na tempograme vidíme aj množstvo iných temp, znázornených rôznymi farbami.

Ďalej určíme prislúchajúcu funkciu sínusoidy pre $n \in [0, N_\Delta - 1]$

$$\kappa_l(n) = \begin{cases} w(n - lH_\tau) \cos\left(\frac{2\pi(n)\tau_p H}{60F_s} + \gamma\right), & n \in [\max(0, lH_\tau - M_\tau), \min(N_\Delta - 1, lH_\tau + M_\tau)] \\ 0, & \text{inak} \end{cases} \quad (5.5)$$



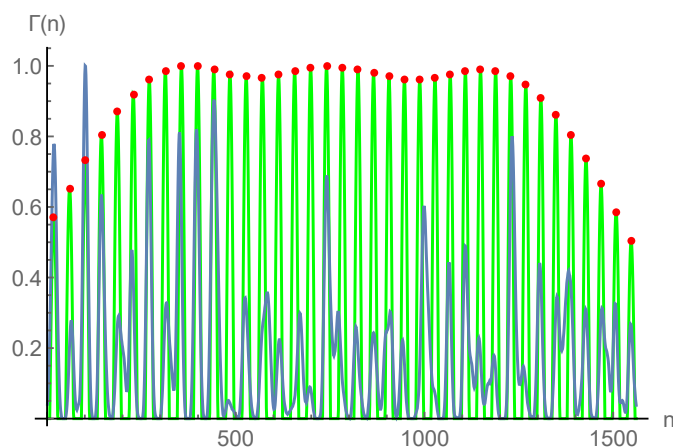
Obr. 5.3: Funkcie sínusoidy pre základné tempo skladby Waltz No. 2 od Dmitrija Shostakovicha vykreslené v troch časových krokoch, s použitím $M_\tau = 500$.

PLP funkciu dostaneme následným sčítaním funkcií sínusoid cez všetky časové kroky, ktorých počet označíme ako l_{max} . Do úvahy vezmeme iba jej kladné hodnoty

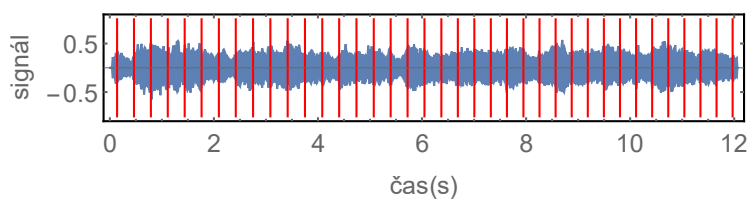
$$\Gamma(n) = \left| \sum_{l=0}^{l_{max}-1} \kappa_l(n) \right|_{\geq 0}, \quad n \in [0, N_\Delta - 1]. \quad (5.6)$$

Príklad PLP funkcie vidíme na obrázku 5.4.

Posledným krokom je nájdenie lokálnych maxím PLP funkcie. Indexy lokálnych maxím predstavujú jednotlivé údery, ktorých čas dopočítame ako v predchádzajúcej sekcii dosadením do vzorca (2.33).



Obr. 5.4: Vypočítané hodnoty PLP funkcie (zelená) pre základné tempo spolu s lokálnymi maximami PLP funkcie (červená) a novinkovou funkciou (modrá) skladby Waltz No. 2 od Dmitrija Shostakovicha.

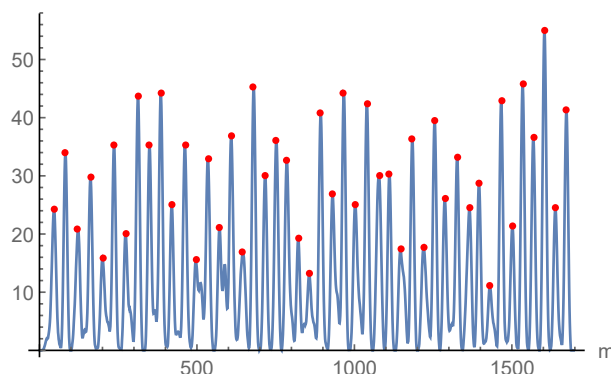


Obr. 5.5: Signál spolu s nájdenými údermi (beatmi) skladby Waltz No. 2 od Dmitrija Shostakovicha. Výslednú sonifikáciu si môžeme vypočítať [Dmitri Shostakovich-Waltz No 2](#).

5.3 Zhrnutie algoritmu beat trackingu

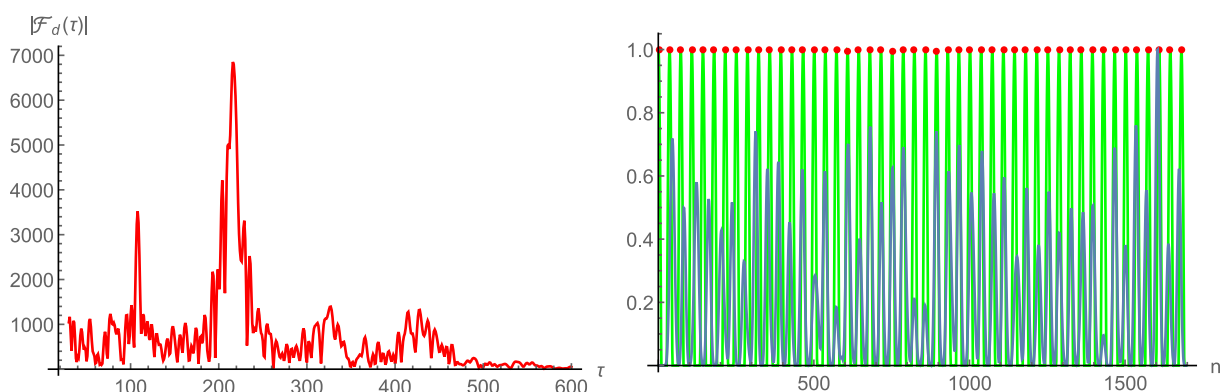
Kroky algoritmu beat trackingu si teraz postupne ukážeme pre zvukovú nahrávku folklórnej piesne s názvom **A ja taká dzivočka**.

Prvým krokom je detekcia nástupov tónov, pričom najvhodnejšie je zvoliť detekciu založenú na spektre a fáze z časti 3.4, ktorá využíva najviac informácií o danej piesni, a teda by mala byť najpresnejšou detekciou.



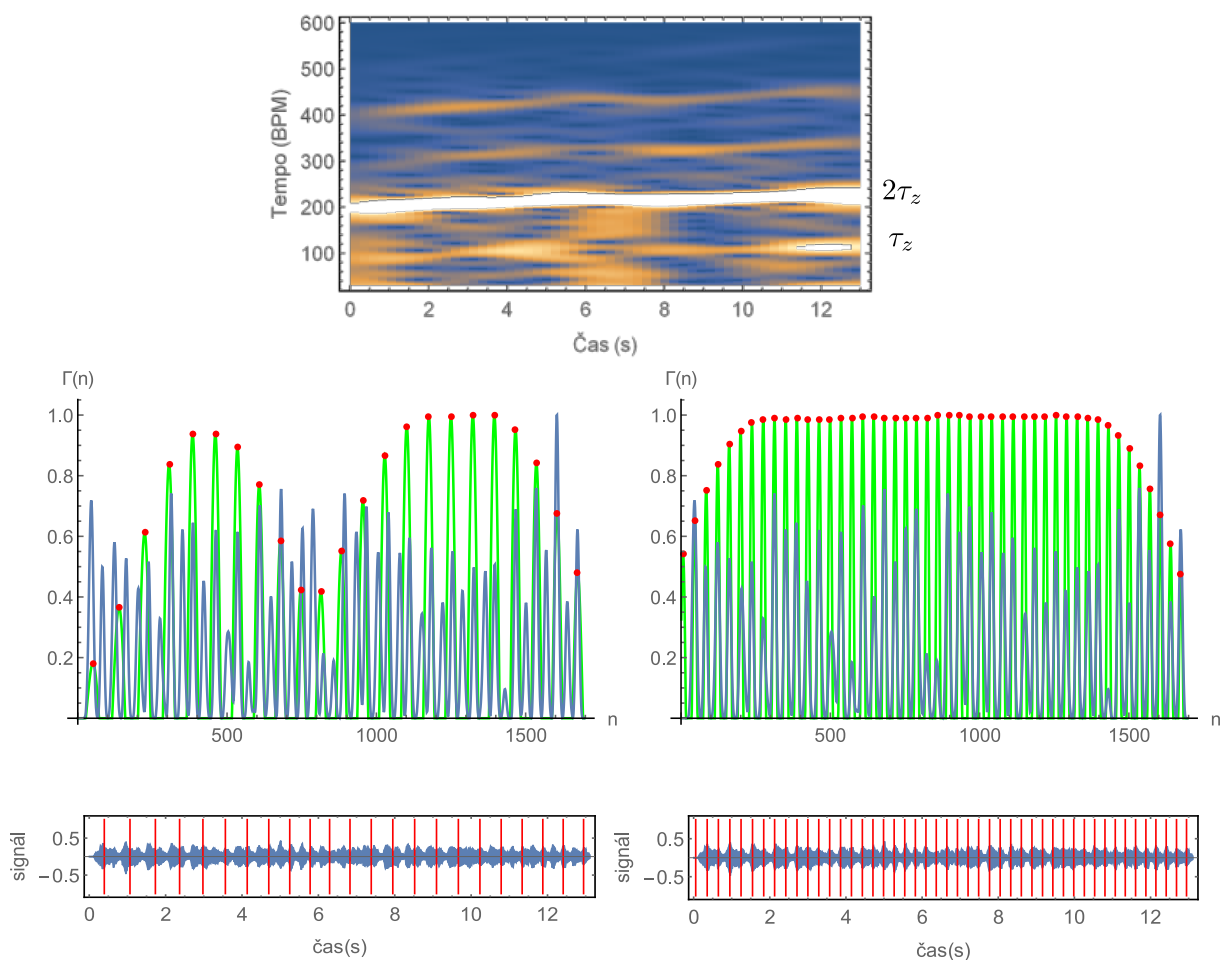
Obr. 5.6: Kľavý priemer vylepšenej novinkovej funkcie (modrá) folklórnej piesne spolu s lokálnymi maximami (červená), ktoré zodpovedajú nástupom.

Ďalším krokom je analýza tempa, kde pomocou výpočtu (4.6) dostaneme tempové spektrum, ktoré vidíme na obrázku 5.7, vľavo. Z tempového spektra vidíme, že dominantné tempo má hodnotu dvojnásobku základného tempa. Prislúchajúcou funkciou sínusoidy (5.4) nájdeme údery (beaty) piesne.



Obr. 5.7: Vľavo: Tempové spektrum. Vpravo: Funkcia sínusoidy (zelená) pre dominantné tempo piesne spolu s jej lokálnymi maximami (červená) a kľavým priemerom novinkovej funkcie (modrá). Sonifikované údery si môžeme vypočítať **tu**. V zvukovej nahrávke počujeme, že údery na začiatku predbehujú doby, v strede nahrávky sú správne a na konci zaostávajú.

Beat tracking pomocou tempového spektra pre túto zvukovú nahrávku nie je vhodný, keďže tempo sa v priebehu piesne mení, konkrétne zrýchľuje. Zmenu tempa vidíme na tempograme na obrázku 5.8, vľavo. Ďalej pomocou PLP funkcie (Obr. 5.8, vpravo), ktorá zachytáva lokálne tempo nahrávky, nájdeme údery pre danú pieseň.



Obr. 5.8: Hore: Tempogram, kde vidíme zrýchľujúce sa tempo piesne. Okrem základného tempa tu vidíme aj jeho celočíselné násobky, pričom dvojnásobné tempo je najviac zastúpené. V strede: PLP funkcie (zelená) pre základné tempo (vľavo) a dvojnásobné, dominantné tempo (vpravo) spolu s lokálnymi maximami (červená) a klzavým priemerom novinkovej funkcie (modrá). Dole: Signál zvukovej nahrávky a nájdené údery (červená), ktoré sme určili vďaka indexom lokálnych maxím PLP funkcie. Výslednú sonifikáciu úderov základného tempa si môžeme vypočítať **tu** a sonifikáciu úderov dominantného tempa **tu**. V sonifikácii úderov základného tempa počujeme prízvuk na druhú dobu, ktorý je typický pre slovenské folklórne piesne.

6 Záver

V bakalárskej práci sme popísali postup beat trackingu pre piesne rôzneho žánru a tempa. Prvým cieľom bolo vysvetlenie základných pojmov ako zvuk, zvuková nahrávka a Fourierova transformácia. Dalším cieľom práce bolo implementovať a porovnať rôzne prístupy detekcie nástupov tónov a výslednú novinkovú funkciu ďalej použiť pre analýzu tempa. K práci sme v softvéri Wolfram Mathematica vytvorili program pre detekciu nástupov a beat tracking, kde sme implementovali tieto algoritmy a použili ich na niekoľko piesní klasickej, folklórnej, popovej a rockovej hudby. Výsledné časy úderov sme sonifikovali, a tak vytvorili zvukové nahrávky piesní spolu s nájdenými údermi. Najlepšie výsledky sme dostali použitím detekcie nástupov tónov založenej na spektre a fáze, a to hlavne pre rockové a popové piesne s pravidelným tempom. Dobré výsledky sme zaznamenali aj pre prípady, kedy sa tempo nemení príliš prudko. Horšie výsledky sme dostali použitím detekcie založenej na fáze a taktiež pri analýze tempa piesní klasickej polyfónnej hudby s náhlou zmenou tempa.

Algoritmy beat trackingu, ktoré sme v tejto práci popísali, majú veľké množstvo aplikácií. V rámci počítačového spracovania hudby sa používajú aj iné algoritmy, ktorými sa dá ďalej zaoberať, ako napríklad detekcia akordov v piesni, rozpoznávanie nástrojov alebo transkripcia hudby.

Literatúra

- [1] R. Altman and P. Altman. *Sound Theory, Sound Practice*. AFI film readers. Routledge, 1992.
- [2] M. Müller. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer International Publishing, 2015.
- [3] F. Rumsey and T. McCormick. *Sound and Recording: An Introduction*. Music technology series. Focal Press, 2006.
- [4] N. Seckel and A. Singh. Physics of 3d ultrasonic sensors. 07 2019.