Počítačové spracovanie hudby: Beat tracking

Eva Bencová

Zvukové nahrávky

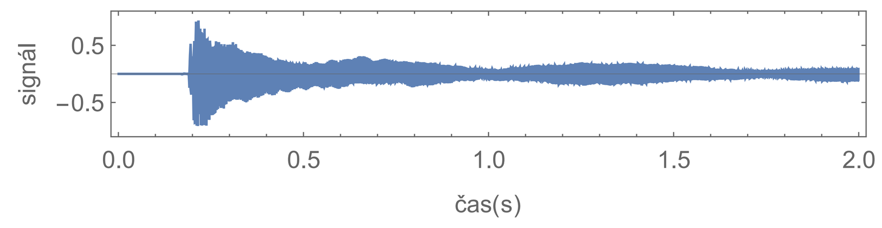
Pre počítačové spracovanie zvuku potrebujeme vytvoriť zvukovú nahrávku, a teda zaznamenať **analógový signál** a previesť ho na **digitálny**. Diskretizáciou analógového signálu $s(t)$ dostaneme hodnoty, ktorým hovoríme **vzorky**. Signál diskretizujeme pre diskrétne hodnoty času

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

kde n je časový index, N počet vzoriek a T_s je vzorkovacia perióda daná vzťahom

$$T_s = \frac{1}{F_s},$$

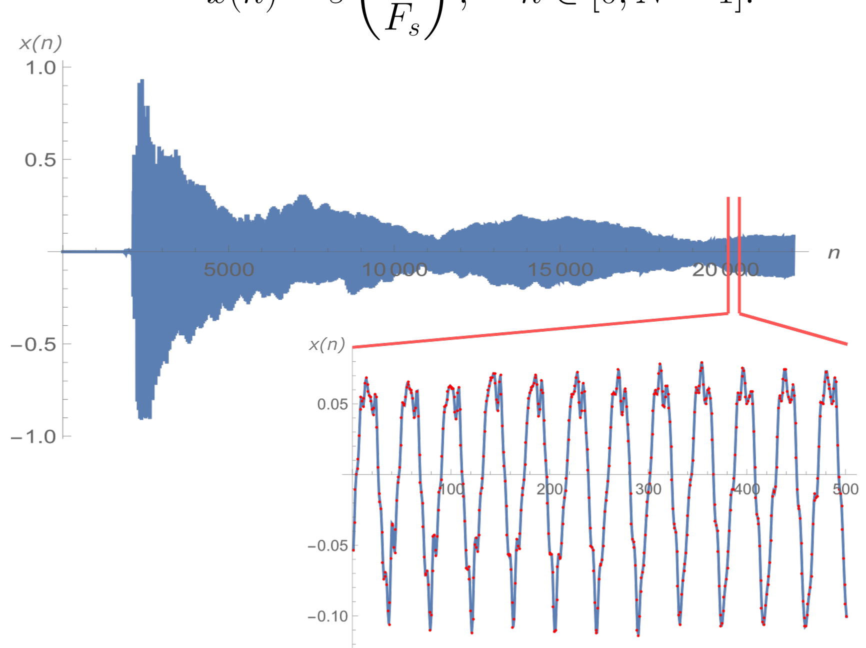
kde F_s predstavuje **vzorkovaciu frekvenciu** zvukovej nahrávky. Definujeme ju ako počet vzoriek analógového signálu za sekundu. Pre MP3 nahrávky sa najčastejšie používa $F_s = 44100$ Hz.



Graf signálu zvukovej nahrávky tónu C4, zahraného na klavíri.

Hodnoty vzoriek zvukovej nahrávky, ktoré nadobúda signál diskretizovaný v čase, zapíšeme do poľa

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



Pole hodnôt vzoriek zvukovej nahrávky tónu C4, zahraného na klavíri.

Frekvencie

Množstvo periód za jednotku času udáva **frekvencia** f zvukovej vlny, ktorá je definovaná vzťahom

$$f = \frac{1}{T}.$$

Hodnotu frekvencie vnímame ako výšku tónu. Pri zahraní jedného tónu na hudobnom nástroji je zvuk zmesou mnohých sinusoíd s rôznymi frekvenciami a amplitúdami. Vtedy je tón tvorený **základnou frekvenciou** (najnižšou prítomnou) a **vyššími harmonickými frekvenciami**.

Fourierova transformácia

Fourierovou transformáciou získame informáciu o prítomných frekvenciách zvukovej nahrávky. Základnou myšlienkou je rozklad signálu na **frekvenčné spektrum** porovnaním signálu so sínusoidami s rôznymi frekvenciami. **Funkcia sínusoidy** má tvar

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

Skalárnym súčinom funkcií vypočítame ich podobnosť, preto pre fixovanú hodnotu frekvencie f určíme

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

kde d_f predstavuje **koeficient podobnosti** a φ_f **koeficient fázy**. Fourierovu transformáciu môžeme definovať aj pomocou **komplexných čísel**. Jedným z hlavných výsledkov Fourierovej teórie je, že hodnotu $\hat{s}(f)$ vypočítame ako

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

Funkcia $\hat{s}(f)$ sa označuje ako Fourierova transformácia signálu $s(t)$. **Diskrétnu Fourierovu transformáciu** odvodeníme z analytického tvaru a definujeme ako

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

Pomocou **diskrétnej krátkodobej Fourierovej transformácie** vieme zistiť, v akom čase boli prítomné jednotlivé frekvencie. Základnou myšlienkou je výpočet Fourierovej transformácie iba v časových výsekoch zvukovej nahrávky. Tie vytvoríme vynásobením signálu $x(n)$ a funkcie Hannovho okna $w(n)$. Predstavíme si nový parameter, ktorým je veľkosť skoku H, $H \in \mathbb{N}$. Pre spracovávanie signálu budeme používať veľkosť skoku rovnú tretine veľkosti okna, čiže $H = N_w/3$. Tak dostaneme definíciu diskrétnej krátkodobej Fourierovej transformácie

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

kde $k \in [0, N_w/2]$ a m predstavuje index časového výseku. Spektrogram označujeme ako $\mathcal{Y}$ a definujeme

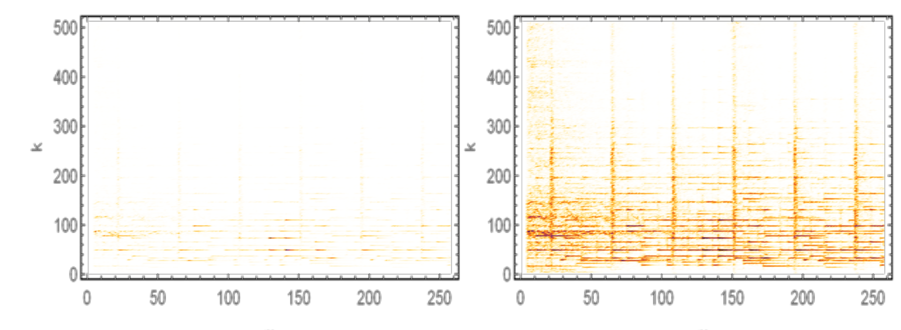
$$\mathcal{Y}(m,k) := |\mathcal{X}(m,k)|^2.$$

Detekcia založená na spektre zvuku

Prvým krokom beat trackingu je detekcia nástupov výraznej zmeny, nejakej novinky v signále zvukovej nahrávky. V tomto prístupe detekcie budeme hľadať časy, v ktorých sa náhle zmení frekvenčné spektrum signálu, čím určíme jednotlivé nástupy. Ako prvé vypočítame diskrétnu krátkodobú Fourierovu transformáciu. Ďalej použijeme **logaritmickú kompresiu**

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

kde budeme dosádzať $\gamma = 1$ a za $|\mathcal{X}|$ absolútnu hodnotu komplexného čísla $\mathcal{X}(m,k)$.



Spektrogram pred a po použití logaritmickej kompresie.

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**

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

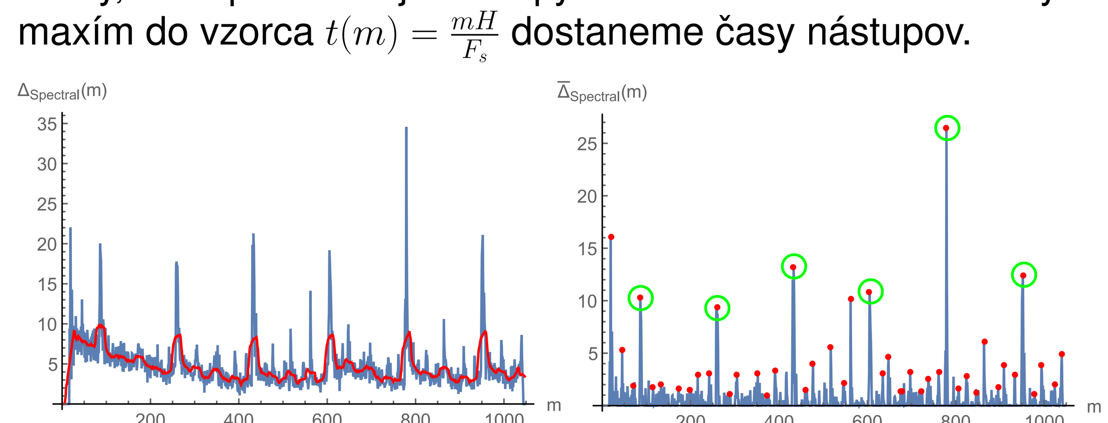
Následne môžeme novinkovú funkciu upraviť pre lepšiu a jednoduchšiu detekciu nástupov. Vypočítame jej kĺzavý priemer,

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

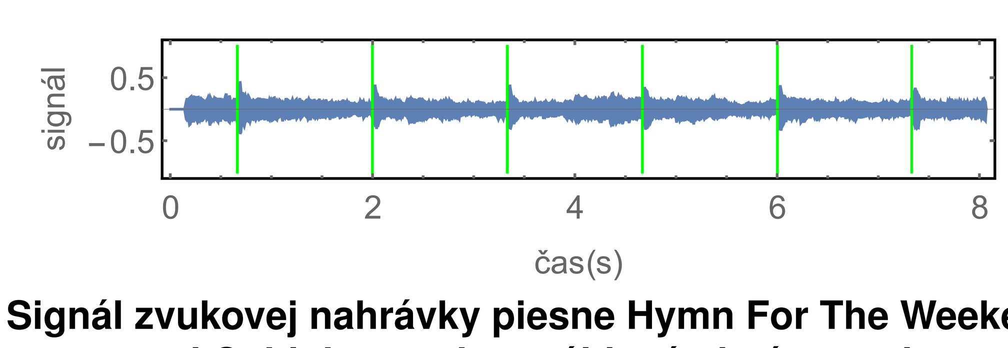
kde $2M_\mu + 1$ predstavuje veľkosť intervalu, na ktorom priemer počítame. Odpočítaním kĺzavého priemeru $\mu(m)$ od novinkovej funkcie $\Delta_{Spectral}(m)$ dostaneme **vylepšenú novinkovú funkciu**,

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

Posledným krokom je nájdenie lokálnych maxím novinkovej krivky, ktoré predstavujú nástupy. Dosadením indexov lokálnych maxím do vzorca $t(m) = \frac{mH}{F_s}$ dostaneme časy nástupov.



Novinkové funkcie k zvukovej nahrávke piesne Hymn For The Weekend od Coldplay.



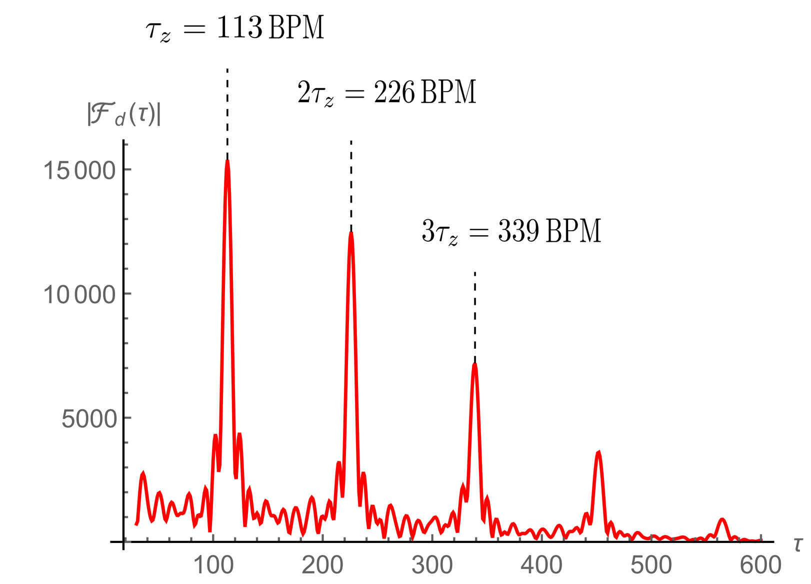
Signál zvukovej nahrávky piesne Hymn For The Weekend od Coldplay spolu s nájdenými nástupmi.

Analýza tempa

V ďalšom kroku sa budeme zaoberať určením tempa zvukovej nahrávky. Tempo τ sa udáva v počte úderov za minútu, čo značíme ako BPM. V zvukovej nahrávke sú okrem základného tempa τ_z aj jeho celočíselné násobky, čo znamená hodnoty $2\tau_z, 3\tau_z$... Pre analýzu tempa použijeme najprv **diskrétnu Fourierovu transformáciu novinkovej funkcie** $\Delta(n)$. Počet hodnôt novinkovej funkcie označíme ako N_Δ.

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

Vďaka tomuto výpočtu zanalyzuje lokálne periodické správanie sa novinkovej funkcie. **Tempové spektrum** predstavuje graf, kde na vodorovnej osi sú hodnoty τ a na zvislej osi sú intenzity prislúchajúce daným tempám, ktoré vypočítame ako $|\mathcal{F}_d(\tau)|$.



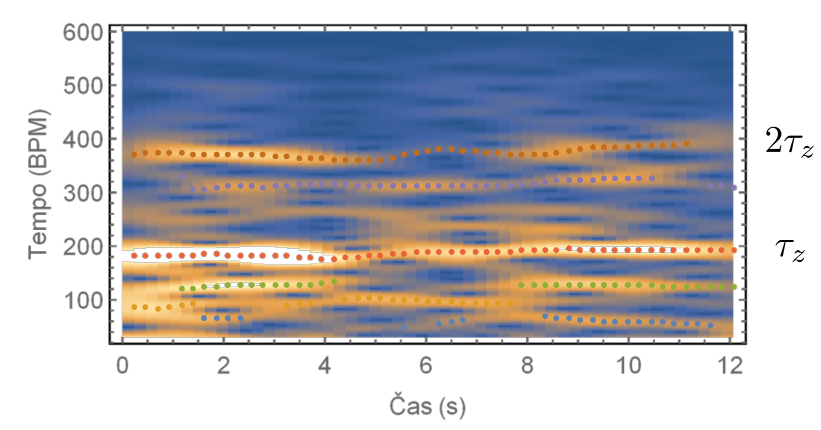
Tempové spektrum pre zvukovú nahrávku piesne Break My Heart od Dua Lipy.

Krátkodobú Fourierovu transformáciu novinkovej funckie definujeme ako

$$\mathcal{F}\left(l, \frac{\tau H}{60F_s}\right) = \sum_{m=-M_\tau}^{M_\tau} \Delta(lH_\tau + m)w(m)e^{\frac{-2\pi i(lH_\tau+m)\tau H}{60F_s}},$$

kde M_τ je veľkosť okna a H_τ veľkosť skoku. Ďalej definujeme **Fourierov tempogram**, ktorého hodnoty vieme zobraziť grafom, kde na vodorovnej osi je čas a na zvislej osi sú tempá.

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



Tempogram skladby Waltz No. 2 od Dmitrija Shostakovicha spolu s vyznačené bodmi, ktoré predstavujú nájdené tempá τ_p v jednotlivých časových krokoch.

Beat Tracking

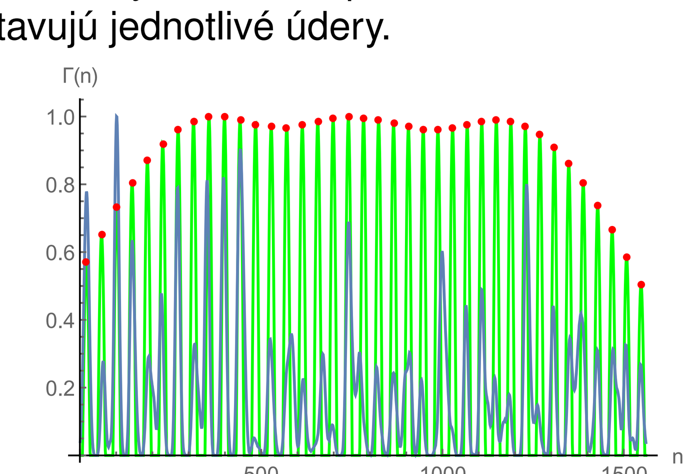
Ďalej skonštrujeme tzv. **PLP funkciu**, ktorá reprezentuje periodickosť novinkovej krivky, a indexy lokálnych maxím tejto funkcie predstavujú konečný výsledok, ktorým sú časy úderov. Pre piesne, v ktorých sa tempo v čase mení, je vhodnejšie vypočítať krátkodobú Fourierovu transformáciu novinkovej funkcie, a tak získať informáciu o prítomných tempách v čase. Prvým krokom je nájdenie lokálnych maxím funkcie $\mathcal{T}(l,\tau)$, ktorých indexy predstavujú hodnoty τ_p pre každý časový výsek. Ďalej každému tempu τ_p určíme prislúchajúcu funkciu sínusoidy pre $n \in [0, N_\Delta - 1]$

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

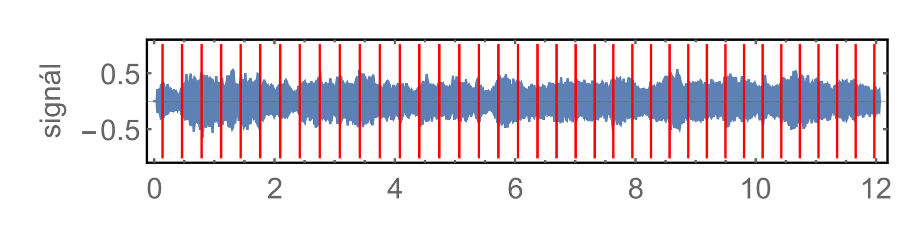
PLP funkciu dostaneme následným sčítaním funkcií sínusoíd cez všetky časové kroky, ktorých počet označíme ako l_{max}. Do úvahy vezmeme iba jej kladné hodnoty

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

Posledným krokom je nájdenie lokálnych maxím PLP funkcie. Indexy lokálnych maxím po dosadení do vzorca $t(m) = \frac{mH}{F_s}$ predstavujú jednotlivé údery.



PLP funkcia (zelená) pre základné tempo spolu s lokálnymi maximami (červená) a novinkovou funkciu (modrá) skladby Waltz No. 2 od Dmitrija Shostakovicha.



Signál spolu s nájdenými údermi (beatmi) skladby Waltz No. 2 od Dmitrija Shostakovicha.

K výsledným nahrávkam sme vytvorili zoznam piesní na YouTube, ktorý otvoríte načítaním QR kódu.

SCAN ME

Pedagóg: Ing. Mgr. Lukáš Tomek, PhD.
Poslucháč: Eva Bencová
Akademický rok: 2020/2021
Študijný program: 9.1.9 aplikovaná matematika
BAKALÁRSKA PRÁCA
Počítačové spracovanie hudby: Beat tracking