VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

Signály a systémy

2022/2023

Projekt - Syntetický klavír

Onegen Something

(xonege99)

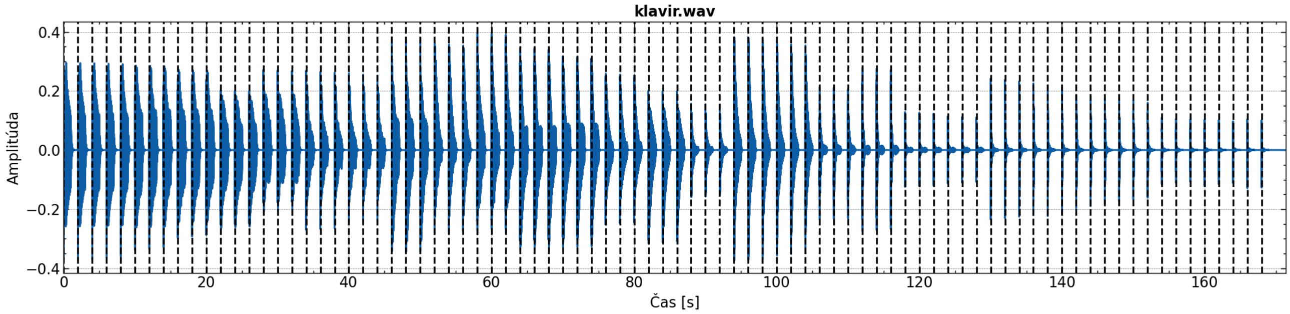
18. 12. 2022

Brno

Tento projekt, zameriavajúci sa na spektrálnu analýzu, je písaný v programovacom jayzku Python (Jupyter Notebook) — zdrojový súbor `projekt.ipynb` sa nachádza v priečinku `src`. Projekt využíva knižnice `numpy` a `scipy` na výpočty, knižnicu `soundfile` na načítanie zvukového súboru, a knižnice `matplotlib`, `IPython` a `scienceplots` na vykreslovanie grafov. Konečným cieľom projektu bolo pomocou analýzy súboru `klavir.wav` vytvoriť syntetický klavír. Pri jednotlivých krokoch som mal zadané tri konkrétne tóny:

`xonege99.txt`	MIDI 36	MIDI 43	MIDI 83
	65,41 Hz	98,00 Hz	987,77 Hz

1 Základy



klavir.wav	
Počet tónov	85
Dĺžka tónu	2 s
Dĺžka celkom	170 s
Vzorkovacia frekvencia	48 kHz

Po načítaní je zvukový súbor rozdelený na jednotlivé tóny s dĺžkou 0,5 sekundy.

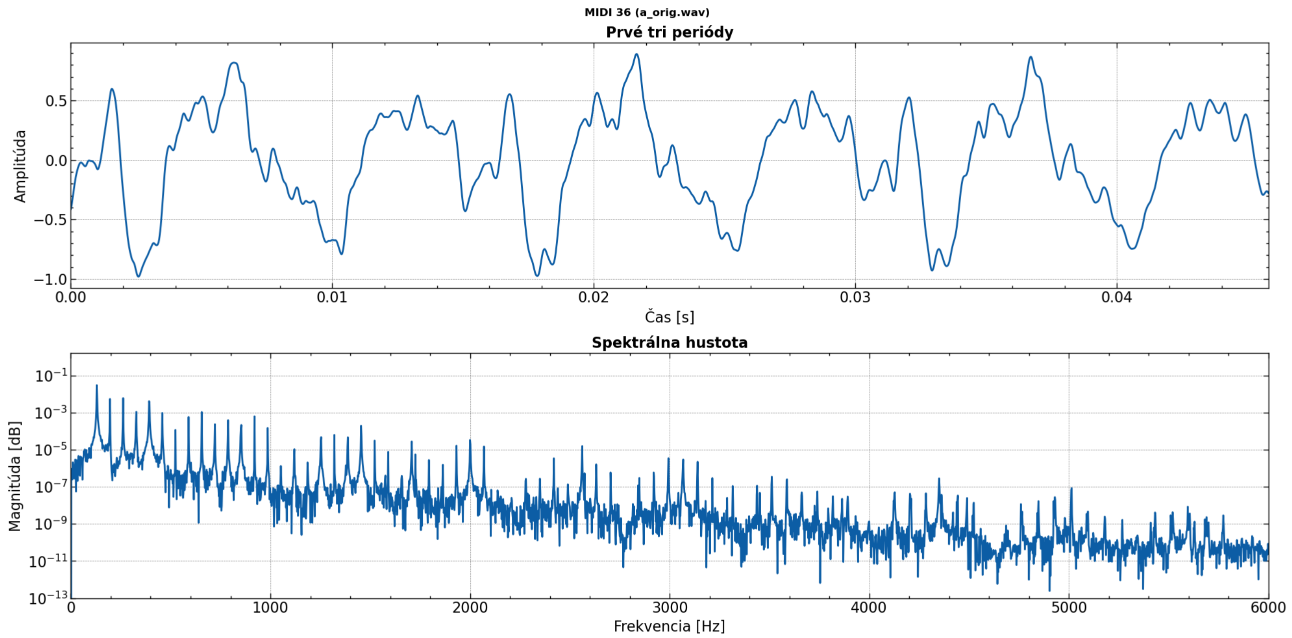
Z celkovej vzorkovacej frekvencie f_0 je možné vypočítať čas medzi vzorkami — **vzorkovaciu periódu** T:

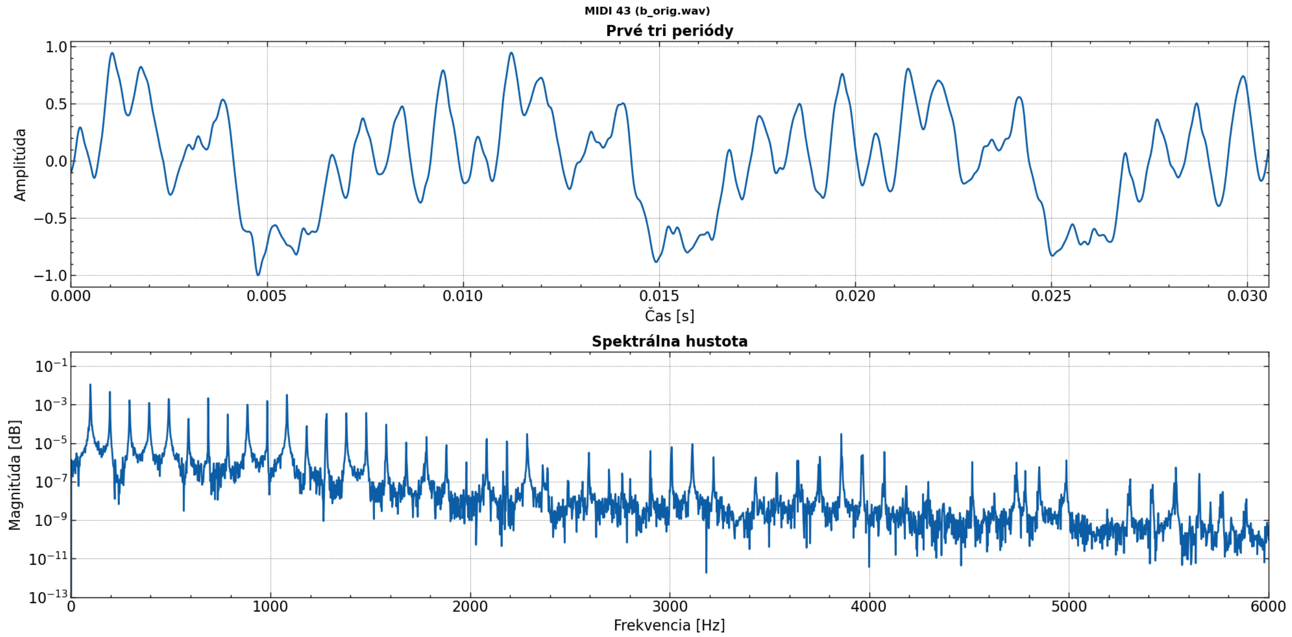
$$T = \frac{1}{f_s} = \frac{1}{48000} = 2,08\overline{3} \cdot 10^{-5} \text{ s}$$

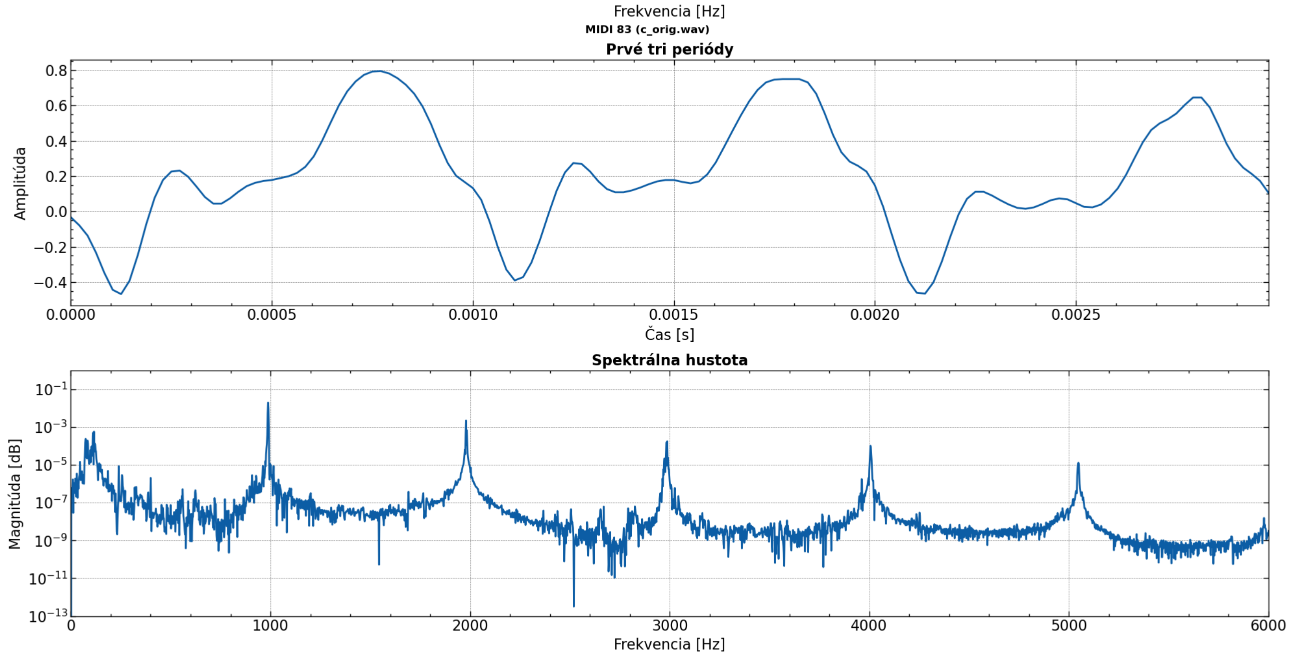
Pri analýze zvukového signálu je základom tento signál frekvenčne analyzovať, teda previesť na spektrum zobrazujúce aké frekvencie sú v signály obsiahnuté a do akej miery. Toto spektrum sa zobrazuje ako logaritmická **spektrálna hustota výkonu** $G[k]$:

$$G[k] = 10 \log_{10} \frac{|X[k]|^2}{N}$$

Na tento výpočet sa používa **rýchla Fourierova transformácia** (FFT), čo je zrýchlený algoritmus výpočtu diskrétnej Fourierovej transformácie (DFT).


MIDI 36 (a_orig.wav)
Prvé tri periódy
Amplitúda
Čas [s]
Spektrálna hustota
Magnitúda [dB]
Frekvencia [Hz]


MIDI 43 (b_orig.wav)
Prvé tri periódy
Amplitúda
Čas [s]
Spektrálna hustota
Magnitúda [dB]
Frekvencia [Hz]


MIDI 83 (c_orig.wav)
Prvé tri periódy
Amplitúda
Čas [s]
Spektrálna hustota
Magnitúda [dB]
Frekvencia [Hz]

2 Určenie základnej frekvencie

Základná frekvencia f_0 je frekvencia tvorená periodickým signálom, ktorá sa v spektre tohoto signálu vyskytuje najvýraznejšie. V prípade zvukového tónu ide o frekvenciu, ktorú pri počutí zvukovej vlny vnímame ako hlavnú. Na jej výpočet sú možné dva spôsoby — *diskrétna Fourierova transformácia* a *autokorelácia* — ktoré majú na istý rozsah tónov odlyšnú presnosť.

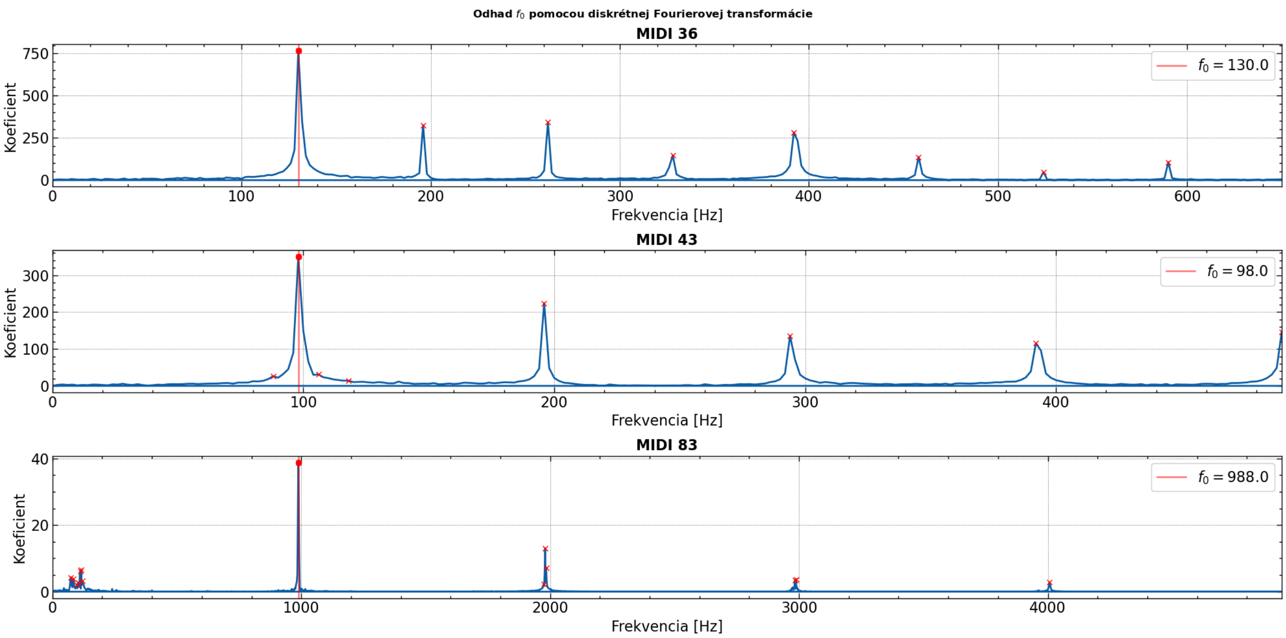
2,1 Diskrétna Fourierova transformácia

Prvý spôsob určenia f_0 je použitie **diskrétnej Fourierovej transformácie** (DFT):

$$c[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-\jmath \frac{2\pi}{N} kn}$$

Použitím DFT na zvukový signál tónu získavame frekvenčnú doménu signálu, na ktorej je vidno zložky tvoriace tento signál. Najväčšia z týchto frekvenčných zložiek — *dominantná frekvencia* — je *zvyčajne* aj základnou frekvenciou.

$$f_0 = f_{max} = \max \left| \sum_{n=0}^{N-1} x[n] \cdot e^{-\jmath \frac{2\pi}{N} kn} \right|$$



MIDI 36	MIDI 43	MIDI 83
130,0	98,0	988,0

Po použití tejto metódy na všetky tóny je možné určiť presnosť tejto metódy porovnaním s hodnotami v súbory `midi.txt`:

$$E = |f_{0_{calculated}} - f_{0_{true}}|$$

Interval [MIDI]	**Chyba [Hz]**
24–40	33 až 166
41–52	0 až 3
56–108	
53–55	176 až 200

Obecne sa ukázalo, že táto metóda je nepresná pre tóny z nižšou frekvenciou, kde odchýľka môže dosiahnuť skoro až 200 Hz, zatiaľ čo na vyšších tónoch je odchýlka značne menšia (do 2 Hz, pravdepodobne malé rozladenie klavíru). Túto vysokú odchýľku spôsobuje jav zvaný **efekt chýbajúceho základného tónu** - čo je jav, kde podtóny *naznačujú* existenciu základnej frekvencie f_0, ktorá sa však v zvukovom signály nenachádza. Ku príkladu, v mojom tóne MIDI 36 (nota C2, správna f_0 je 65,41 Hz) bola detekovaná f_0 130 Hz, čo však korešponduje s tónom o oktávu vyššie (dvojnásobok; MIDI 48, nota C3). Následujúce vrcholy viditeľné v grafe sú taktiež posunuté o oktávu (teda sú násobky správnej f_0) — 192 Hz, 262 Hz, 325 Hz a tak ďalej. Toto naznačuje, že správna f_0 je 65 Hz aj keď sa v signály nenachádza — napriek tomu túto neprítomnú frekvenciu ľudské ucho ju vníma ako základnú (tzv. *psychoakustický jav*).

Tento fakt však Fourierova transformácia neberie do úvahy — vrcholy sú zobrazené len pri frekvenciách, ktoré sa v signály skutočne nachádzajú. Na nižšie tóny je teda vhodnejšia iná metóda, ktorá berie do úvahy akékoľvek harmonické frekvencie — autokorelácia.

2,2 Autokorelácia

Autokorelácia (alebo *autokorelačná funkcia* — ACF) je korelácia (závislosť, podobnosť) signálu s opozdenou kópiou samého seba — teda určuje ako sa signál podobá sám sebe v rôznych časových bodoch.

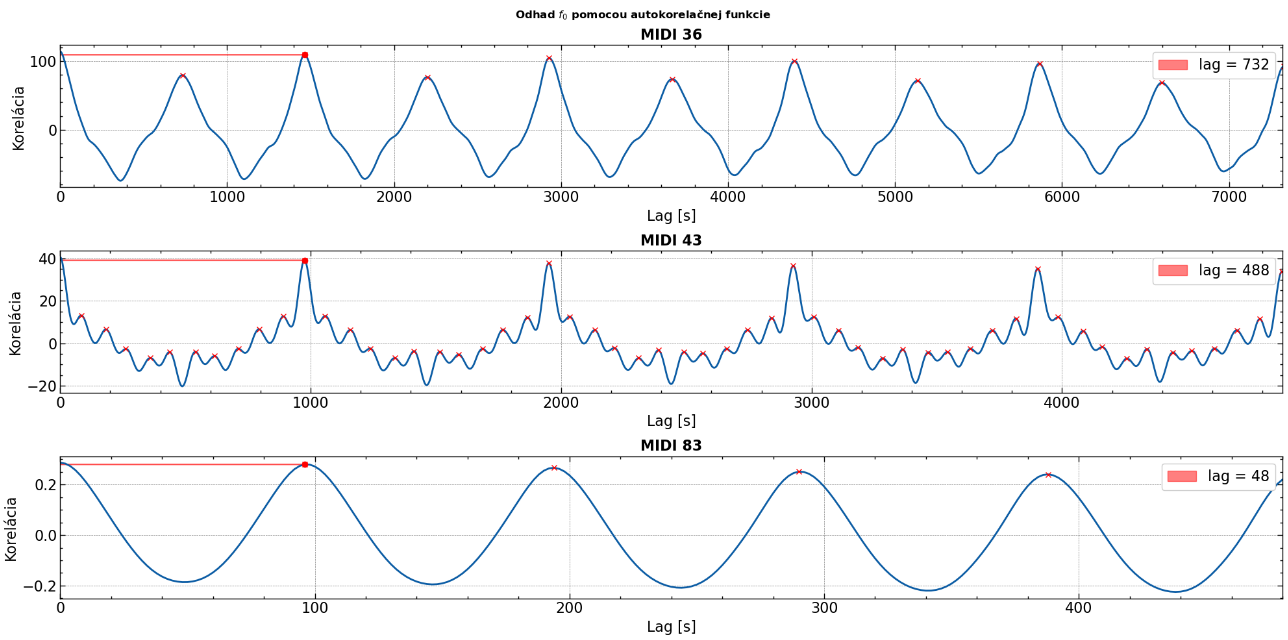
$$R(m) = \sum_{n=0}^{N-1-m} s(n) \cdot s(n+m)$$

$$R(m) = \sum_{n=m}^{N-1} s(n) \cdot s(n-m)$$

Z výslednej autokorelačnej funkcie $R(m)$ sa potom určuje hodnota *lag* hľadaním indexu jeho maxima:

$$lag = \arg\max R(m)$$

$$f_0 = \frac{1}{lag} \cdot F_s$$



MIDI 36	MIDI 43	MIDI 83
65,57	98,36	1000

Rovnako ako pri predošlej metóde som porovnal získanú hodnotu f_0 zo správnou hodnotou z `midi.txt`:

Interval [MIDI]	Chyba [Hz]
24–65	0 až 1,1
66–80	0 až 3
81–101	3 až 40
102–108	37 až 2100

Ukázalo sa, že táto metóda je vhodnejšia pre tóny z nižšou frekvenciou — presne opačne od DFT. Autokorelácia zdanlivo ignoruje problém chýbajúcej základej frekvencie, a to preto, lebo analýzou všetkých harmonických frekvencií *detekuje* základnú frekvenciu v zložkách (v násobkoch f_0 v následujúcich oktávach). Autokorelácia má však svoje problémy — napríklad detekcia nesprávnych vrcholov (lag).

2,3 Výsledok

MIDI 36	MIDI 43	MIDI 83
65,57	98,00	980,00

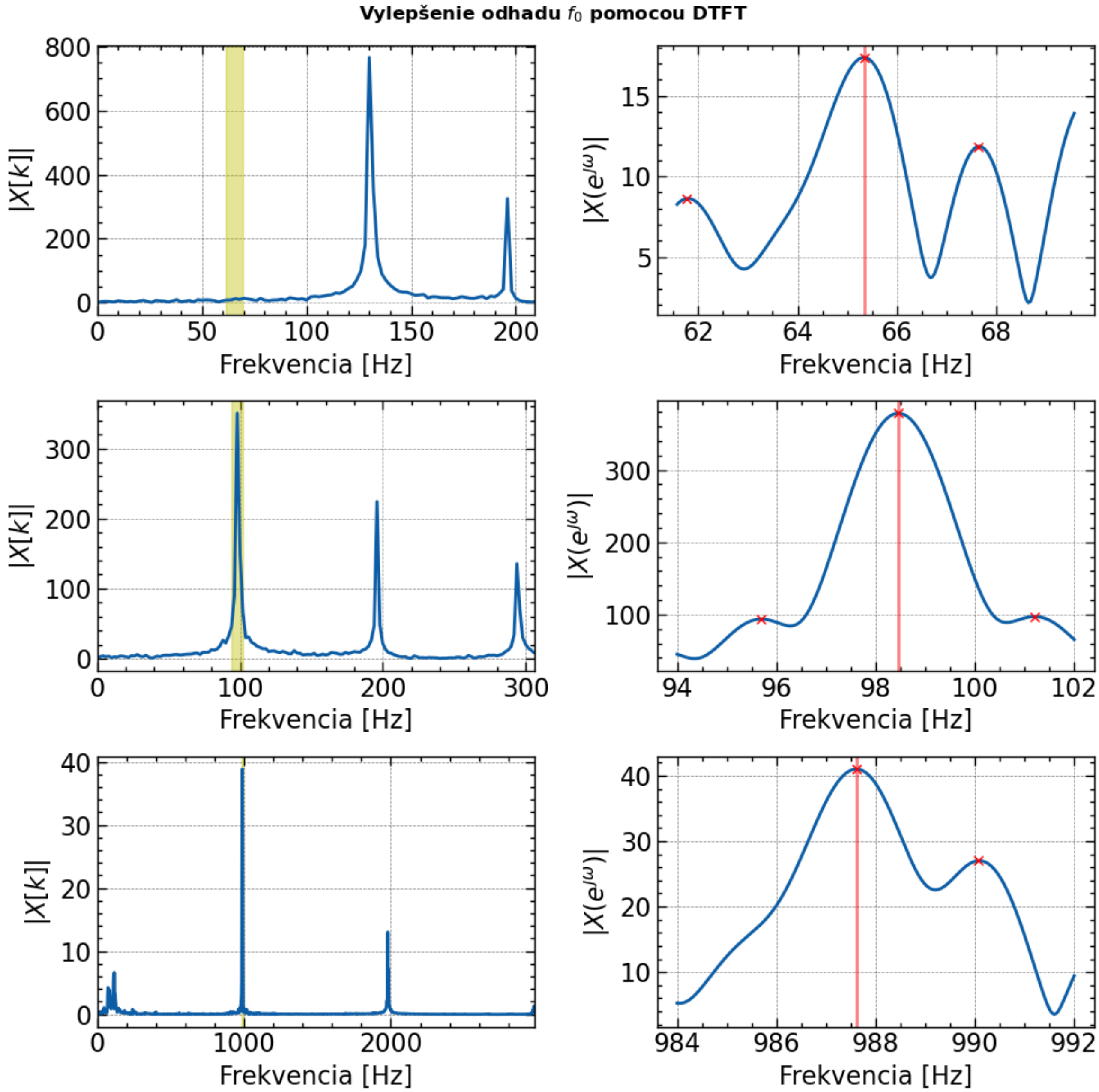
Podľa výsledkov presnosti sa ukázalo, že metóda diskrétnej Fourierovej transformácie je presnejšia na nižších tónoch (MIDI >40) a metóda autokorelačnej funkcie je presnejšia na vyšších tónoch (MIDI <81).

3 Spresnenie odhadu základnej frekvencie

Získaná základná frekvencia f_0 je však stále iba odhadom Fourierovej transformácie alebo autokorelácie — obe tieto metódy majú svoje obmedzenia a výsledky nemusia byť úplne presné. Tento odhad však môže byť zlepšení použitím **Fourierovej transformácie v diskrétnom čase** (DTFT).

$$X(e^{\jmath\omega}) = \sum_{n=-\infty}^{\infty} x[n] \cdot e^{-\jmath\omega n}$$

DTFT som aplikoval na časť tónu v blízkosti aktuálneho odhadu. V predošlej časti bola zistená najhoršia odchýľka 2,93 Hz, čo som zaokrúhlil na 3 Hz a pre istotu pridal 1 Hz, teda na časť tónu v rozpätí ± 4 Hz od aktuálneho odhadu.

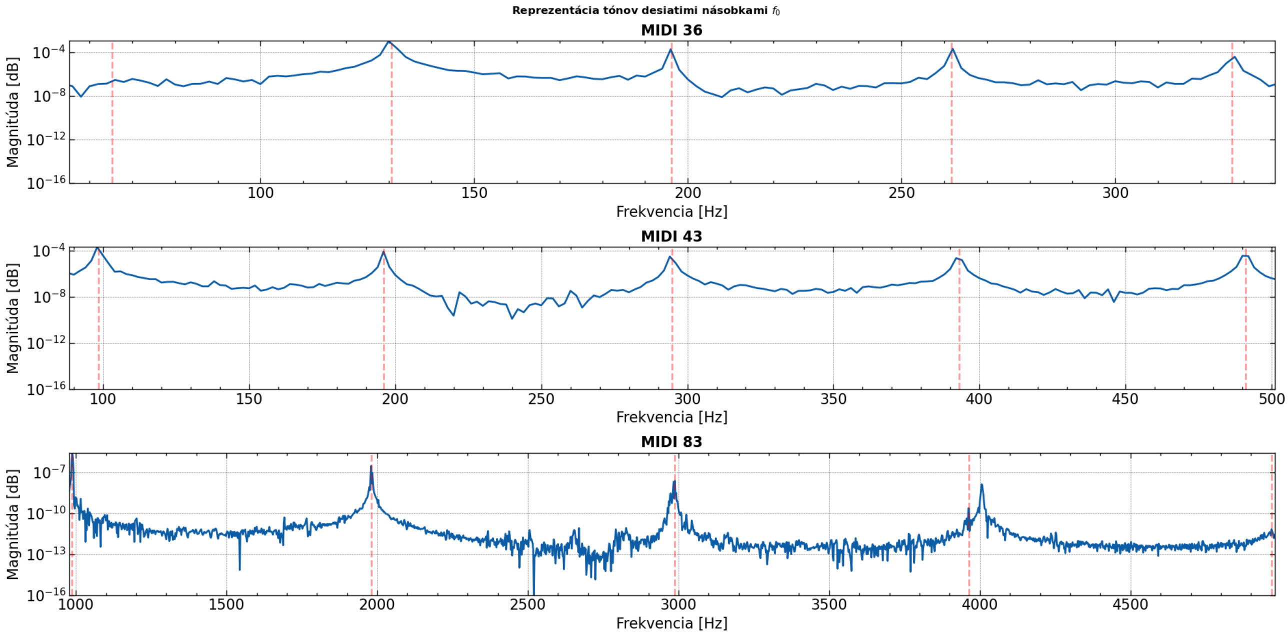


MIDI 36	MIDI 43	MIDI 83
65,35	98,46	987,62

Výsledky DTFT boli zmiešané. Väčšie zmeny nad 0,5 Hz boli skôr k lepšiemu, zatiaľ čo malé zmeny pod 0,5 Hz boli skôr negatívne. Zmeny nepresahovali najväčšiu odchýlku odhadu (2,93 Hz). Výsledok taktiež záležal na výške tónu - nižšie tóny mali prevažne negatíny výsledok, zatiaľ čo vyššie tóny mali f_0 odhad zväčša zlepšený.

4 Reprezentácia klavíra

Ďalej je na reprezentáciu každého tónu získaných 10 čísel, ktoré budú použité pri syntéze. Ja som tieto čísla získaval na piatich násobkoch f_0 vypočítaním DTFT z magnitúdy tónu a jeho fáze.



MIDI 36	MIDI 43	MIDI 83
65,25	98,36	987,52
78,29	104,38	986,92
130,61	196,02	1980,35
124,18	215,29	1978,95
196,16	294,68	2988,04
180,71	309,73	2937,27
261,71	392,95	3963,01
246,86	389,53	3947,56
327,26	491,01	4968,09
325,66	495,22	4946,82

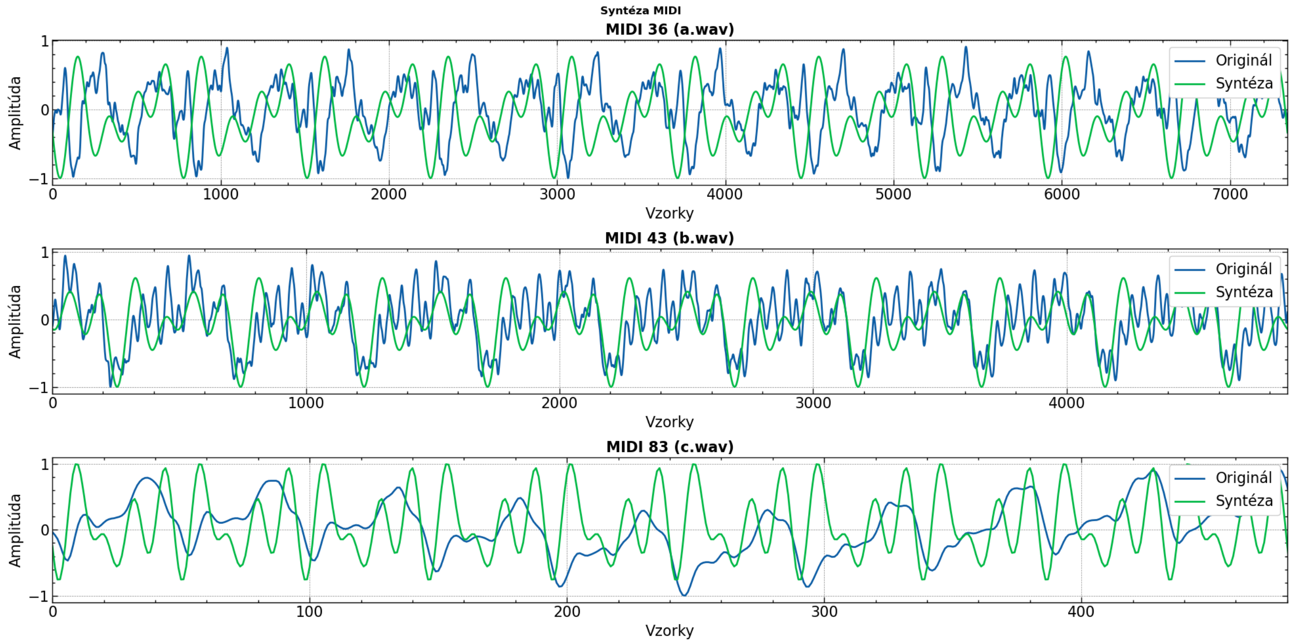
5 Syntéza

Použitím **inverznej diskrétnej Fourierovej transformácie (IDFT)** sa ďalej syntetizujú tóny z čísel získaných v predošlej časti sa ďalej syntetizujú tóny z čísel získaných v predošlej časti.

$$x[n] = \frac{1}{N} \cdot \sum_{k=0}^{N-1} y[k] \cdot e^{\jmath \frac{2\pi k n}{N}}$$

Z desiatich čísel získaných v predošlej časti je 5 modulov r (párne indexy) a 5 fází ϕ (nepárne indexy), čo tvorí 5 komplexných Fourierových koeficientov, použitých ako vstup IDFT:

$$y[k] = r \cdot e^{\jmath \phi}$$



6 Záver

Na vypracovanie som sa hlavne inšpiroval notebookmi z prednášok (DTFT a syntéza) a študíjnej etapy (spektrogram), a pri implementácií som používal dokumentáciu knižníc numpy, scipy a matplotlib na porozumenie funkciám.