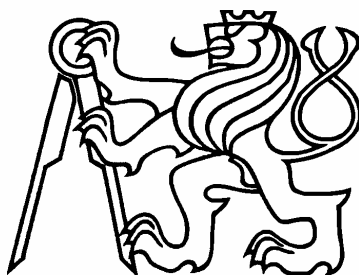


ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ

Fakulta elektrotechnická

Katedra teorie obvodů



Diplomová práce na téma

Kytarový MIDI převodník

2003

vypracoval: Hynek Bořil

vedoucí diplomové práce: Ing. Václav Hanžl, CSc.

## Prohlášení

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a veškerá literatura a prameny, ze kterých jsem čerpal, jsou uvedeny na konci této práce.

Nemám námitky proti užití tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne 24. 1. 2003

Hynek Bořil

## Poděkování

Děkuji Ing. Petru Pollákovi, CSc., Ing. Hynku Kocourkovi, Ing. Vítu Kavanovi, CSc., Ing. Pavlu Hrzinovi a Tomáši Bořilovi za cenné rady a připomínky při tvorbě a implementaci této práce a Ing. Václavu Hanžlovi, CSc., který se krom rad a připomínek uvolil tuto práci i vést. Také děkuji rodině a přátelům za odhalení značného množství gramatických chyb a překlepů.

|         |   |    |
|---------|---|----|
| 1       | Úvod  | 5  |
| 2       | Syntetický hudební nástroj MIDI                 | 7  |
| 2.1     | MIDI kontroler                                  | 7  |
| 2.2     | MIDI převodník                                  | 8  |
| 2.3     | Zvukový modul MIDI                              | 8  |
| 2.4     | FM syntéza                                      | 8  |
| 2.5     | Wave Table syntéza                              | 11 |
| 2.5.1   | Sample Looping                                  | 11 |
| 2.5.2   | Pitch Shifting                                  | 12 |
| 2.5.2.1 | Decimace signálu                                | 13 |
| 2.5.2.2 | Interpolace signálu                             | 13 |
| 2.6     | Generátor obálky tónu                           | 17 |
| 3       | Syntetická kytara – kontroler                   | 18 |
| 3.1     | Selektivní snímač                               | 18 |
| 3.2     | A/D převodník                                   | 19 |
| 3.3     | Detektor dynamiky                               | 19 |
| 3.4     | Pitch detektor                                  | 19 |
| 3.5     | Generátor MIDI událostí                         | 19 |
| 4       | Vybrané metody určení základní frekvence tónu   | 20 |
| 4.1     | Diskrétní Fourierova transformace (DFT)         | 21 |
| 4.1.1   | Amplitudové spektrum                            | 22 |
| 4.1.2   | Polyfonní signál                                | 23 |
| 4.1.3   | Implementace DFT detekce v Matlabu              | 24 |
| 4.1.4   | Vztah mezi časovým a frekvenčním rozlišením DFT | 25 |
| 4.2     | Lineární predikce                               | 26 |
| 4.2.1   | Autokorelační metoda                            | 27 |
| 4.2.2   | Kovarianční metoda                              | 33 |
| 4.3     | Tónová banka filtrů                             | 34 |
| 4.3.1   | Realizace banky filtrů půlením intervalu        | 34 |
| 4.3.2   | Ukázka realizace 1. kroku půlení intervalu      | 36 |
| 4.4     | Autokorelace                                    | 39 |
| 4.4.1   | Detekce periody                                 | 39 |
| 4.4.2   | Zvýšení frekvenčního rozlišení autokorelace     | 46 |
| 4.4.3   | Implementace autokorelace v Matlabu             | 48 |
| 4.5     | Přímé měření délky periody                      | 48 |
| 5       | Implementace syntetické kytary                  | 51 |
| 5.1     | DP - dolní propust                              | 53 |
| 5.1.1   | Výběr aproximace                                | 53 |
| 5.1.2   | Toleranční schéma DP                            | 53 |
| 5.2     | Obálka $P(n)$                                   | 55 |
| 5.2.1   | Blokový odhad výkonu                            | 56 |
| 5.2.2   | Průběžný odhad výkonu                           | 58 |
| 5.3     | Funkce $f_{detect}()$                           | 58 |
| 5.3.1   | Detekce lokálních extrémů                       | 59 |
| 5.3.2   | Detekce význačných extrémů                      | 59 |
| 5.3.3   | Měření délky periody a výpočet frekvence        | 61 |
| 5.4     | Funkce $true\_note()$                           | 61 |
| 5.4.1   | Energetické kritérium                           | 61 |
| 5.4.2   | Frekvenční kritérium                            | 62 |
| 5.4.3   | Majoritní kritérium                             | 62 |
| 5.5     | Tvorba událostí                                 | 63 |
| 5.6     | Funkce $gensound()$ – jednoduchý syntezátor     | 64 |
| 5.7     | Tvorba MIDI souboru                             | 66 |
| 5.7.1   | Meta-události                                   | 67 |
| 5.7.2   | MIDI události                                   | 67 |
| 5.7.3   | SYSEX události                                  | 67 |
| 5.8     | Práce se syntetickou kytarou                    | 67 |
| 6       | Závěr   | 69 |
| 7       | Abstract  | 70 |
|         | Příloha   | 71 |
|         | Literatura                                      | 72 |

# 1 Úvod

Bez nadsázky lze říci, že jedním z nástrojů, které prošly za posledních sto let nejdramatičtějším vývojem, je právě kytara. Nástroj byl svou dynamikou původně určen spíše ke komorní produkci, neboť nemohl konkurovat svým hlasitějším kolegům. Na počátku 20. století však začala popularita kytary stoupat, na jedné straně díky nové generaci hráčů, jimž virtuózní technika umožnila opustit barokní a renesanční etudy a orientovat se na repertoár, který osloví současné posluchače, na straně druhé díky pronikání černošské lidové hudby – blues, jazzu a soulu do podvědomí široké veřejnosti. Kytara se tak počala pomalu přesouvat z úzkých komunit do koncertních a tanečních sálů. Odtud byl už jen krůček k její amplifikaci. Zatímco snímání zvuku mikrofonom nevyžadovalo žádné zvláštní úpravy nástroje, ve 20. letech minulého století se objevily první úspěšné pokusy nahradit mikrofón magnetodynamickými snímači. Tyto snímače byly nejprve instalovány na klasické akustické kytary s dutým korpusem, posléze se ukázalo vhodnější použít korpus plný, který nemá takové sklony k uzavření kladné akustické zpětné vazby při větším zesílení. Nástup zesilovačů krom zesílení reprodukce přinesl i škálu nových barev kytarového zvuku díky elektronkovému zkreslení v zesilovacím řetězci. Toto zkreslení umožnilo zdůraznit smutek blues a později dravost rock'n'rollu. S rozvojem a miniaturizací elektroniky jdoucím ruku v ruce s vývojem vkusu se objevují stále nové cesty, jak kytarový zvuk v tomto směru obohatit.

Chuť experimentovat se zvukem se přenesla postupně i na další nástroje. Jedním z prvních následníků byly elektrické varhany. Zvuk je zde tvořen dvanácti harmonickými generátory příslušejícími jednotlivým půltónům v oktávě. Pro tvorbu barvy tónu je možno signály od jednotlivých generátorů a jejich harmonické kombinovat – podobně jako u klasických varhan tvoříme rejstřík tónu kombinací několika píšťal – a rozličnými způsoby modifikovat obálku. Zatímco první elektrické varhany měly snahu přiblížit se zvuku píšťalových varhan, později se začaly objevovat i zvuky avantgardní, ryze syntetické. Jiný typ syntezátorů využíval tónů přirozených nástrojů zaznamenaných na magnetofonových nekonečných páskách (Melotron) a umožňoval tak ovládat z kláves rejstřík zvuků smyčců apod. S nástupem číslicového zpracování signálu mohly být v 70. letech magnetické pásky nahrazeny polovodičovými paměťmi a klávesové syntezátory dospěly do stádia, v jakém je známe prakticky i nyní.

Vývoj klávesových syntezátorů brzy vnuknul některým výrobcům myšlenku konstrukce syntezátoru kytarového. První pokusy se objevily v polovině sedmdesátých let, přinášely však

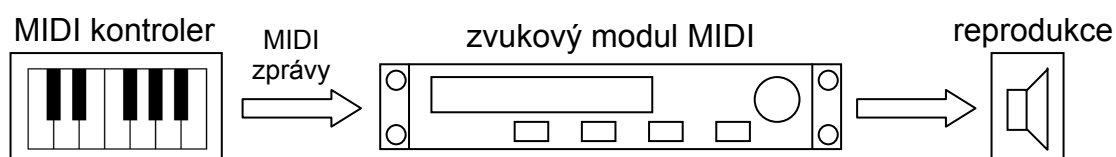
kromě příslibu nových zvuků především necitlivé zásahy do konstrukce nástroje, nutící ke změně techniky hry, značnou nespolehlivost, a v neposlední řadě byly dosti nákladné. Proto se v hráčské veřejnosti setkávaly až do konce 80. let spíše s odmítnutím. V posledních letech byly speciální syntezátorové kytary nahrazeny převodníky, které stačí nainstalovat na běžnou kytaru, a tudíž hráče příliš neomezují. Tyto převodníky jsou v závislosti na výrobci různě rychlé a spolehlivé, společným jmenovatelem je však jejich ještě stále vysoká cena a fakt, že výrobci z pochopitelných důvodů neuvádějí použité metody rozpoznávání.

Mým cílem v této práci bylo navrhnout metodu analýzy kytarového signálu, která bude dostatečně spolehlivá a rychlá, tudíž použitelná pro hraní v reálném čase, a zkonstruovat na základě této metody kytarový MIDI převodník.

## 2 Syntetický hudební nástroj MIDI

Syntetické nástroje vystřídal během svého vývoje řadu architektur v závislosti na době vzniku a značce výrobce. První pokusy o zavedení komunikačního protokolu, jímž by bylo možno přenášet informace o hře, se objevily již u analogových syntezátorů. Výška hraného tónu se zde transformovala přímo na hodnoty napětí. S nástupem číslicové architektury potřeba možnosti komunikace narůstala. Tentokrát již společné úsilí nejvýznamnějších výrobců vyústilo v roce 1982 resp. 1983 v definici protokolu MIDI (Musical Instrument Digital Interface), určeného pro záznam událostí vznikajících při hře na hudební nástroj a jejich případnou reprodukci. Protokol je založen na posílání MIDI zpráv, popisujících události v čase. Základními událostmi jsou vznik a zánik tónu, hlasitost, plynulá změna výšky tónu (pitch bend), výběr nástroje atd. Protokol podporuje polyfonii (vícehlas). Podrobněji bude zmíněna struktura MIDI zpráv v oddíle o tvorbě MIDI souboru.

Syntetický nástroj založený na MIDI architektuře je tvořen dvěma základními částmi – kontrolerem a zvukovým modulem – *obr. 2.1*.



*Obr. 2.1: Základní architektura MIDI*

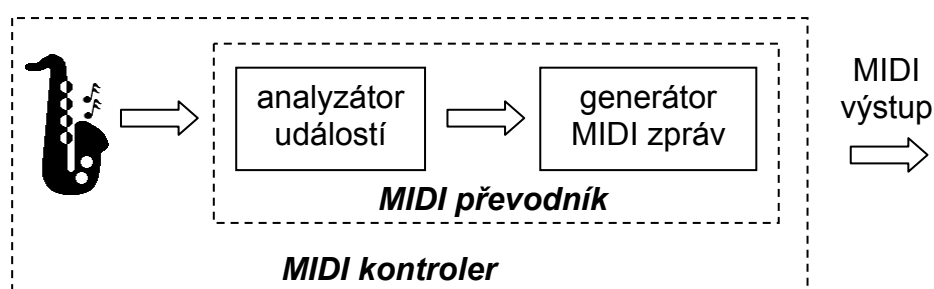
### 2.1 MIDI kontroler

MIDI kontroler obsluhovaný hráčem vytváří odpovídající události a zasílá je zvukovému modulu v podobě zpráv. Zvukový modul událostem přiřazuje vybrané syntetické nástroje, obálky a efekty, a generuje reprodukovatelný zvukový signál.

Pokud jsou kontrolerem klávesy, je úloha převodu hry na MIDI události relativně jednoduchá. Každé klávese jsou přiřazeny dva snímače pro polohy stisknuto a uvolněno, dynamiku (razanci úderu) pak určíme z rychlosti přechodu mezi těmito polohami – odtud plyne poněkud překvapivé označení pro hlasitost při nástupu tónu v MIDI - *velocity*, tj. rychlost změny stavu. Pokud chceme dosáhnout přesnější odezvy a jemnějšího rozlišení polohy klávesy, je možno přiřadit klávese více snímačů.

## 2.2 MIDI převodník

Nový typ úlohy se objeví, pokud chceme jako MIDI kontroler použít přirozený hudební nástroj, u něhož pro popis hry nestačí snímat stav mechaniky. Možným řešením je snímání zvuku nástroje a následný převod na události. V tomto případě analyzující část kontroleru zakončenou generátorem MIDI zpráv nazýváme MIDI převodník – *obr. 2.2*.



*Obr. 2.2: Struktura kontroleru s MIDI převodníkem*

## 2.3 Zvukový modul MIDI

Jak již bylo zmíněno, zvukový modul má za úkol provádět na základě příchozích MIDI zpráv syntézu zvuku, jedná se tedy o syntezátor\*.

Mezi nejpoužívanější principy číslicové syntézy zvuku nástrojů patří FM a Wave Table syntéza.

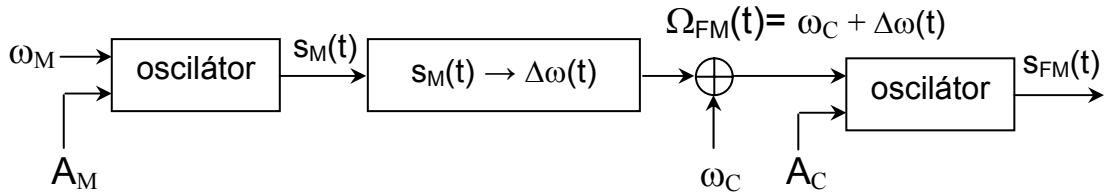
## 2.4 FM syntéza

FM syntéza, jak již z názvu vyplývá, je založena na principu frekvenční modulace. V nejjednodušším případě jsou jejím vstupem dva harmonické signály – nosný a modulační, *obr. 2.3*.

---

\* Často je nazýván syntezátorem celý syntetický hudební nástroj (viz Úvod), tedy spojení kontroleru a zvukového modulu. Pro přehlednost se proto budu od tohoto momentu snažit tomuto termínu ve spojení s celým nástrojem vyhýbat, a používat jej striktně pro označení zvukového modulu.





Obr. 2.3: FM syntéza ze dvou kmitočtů

Pro modulační signál  $s_M(t)$  s amplitudou  $A_M$  a kmitočtem  $\omega_M$  a nosný signál  $s_C(t)$  s amplitudou  $A_C$  a kmitočtem  $\omega_C$  můžeme psát:

$$s_M(t) = A_M \cos(\omega_M t), \quad (2.1)$$

$$s_C(t) = A_C \cos(\omega_C t). \quad (2.2)$$

Výstupem FM syntézy je signál

$$s_{FM} = A_C \cos(\Omega_{FM}(t) \cdot t). \quad (2.3)$$

Jak je ukázáno na obr. 2.3, okamžité hodnoty modulačního signálu  $s_M(t)$  jsou převáděny na změny kmitočtu  $\Delta\omega(t)$ . Součet konstantního kmitočtu nosné  $\omega_C$  a proměnného kmitočtu  $\Delta\omega(t)$  pak tvoří proměnný kmitočet  $\Omega_{FM}(t)$  výstupního FM signálu:

$$\Omega_{FM}(t) = \omega_C + \Delta\omega(t) = \omega_C + \Delta\Omega \cos(\omega_M t), \quad (2.4)$$

kde  $\Delta\Omega$  je kmitočtový zdvih přímo úměrný amplitudě modulačního signálu. Modulační signál  $s_M(t)$  tedy moduluje kmitočet nosného signálu, nikoliv jeho amplitudu, jak je tomu u amplitudové modulace. Jelikož je představa proměnného kmitočtu náročná, je vhodnější přejít k fázovým vztahům.

Pro okamžitý kmitočet platí

$$\omega(t) = \frac{d\varphi(t)}{dt}. \quad (2.5)$$

Aplikací vztahu (2.5) na (2.4) dostáváme pro okamžitou fázi FM signálu

$$\varphi_{FM}(t) = \int_0^t \Omega_{FM}(\tau) d\tau = \int_0^t (\omega_c + \Delta\Omega \cos(\omega_M \tau)) d\tau = \varphi_c + \frac{\Delta\Omega}{\omega_M} \sin(\omega_M t). \quad (2.6)$$

Podíl na pravé straně rovnice (2.6) se nazývá modulační index,

$$M = \frac{\Delta\Omega}{\omega_M} = \frac{\Delta f_{FM}}{f_M}. \quad (2.7)$$

Modulační index udává hloubku modulace pro daný modulační signál a je přímo úměrný amplitudě a nepřímo úměrný kmitočtu modulačního signálu. Pokud by byl modulační signál proměnlivý co do amplitudy či kmitočtu, bude se měnit i modulační index.

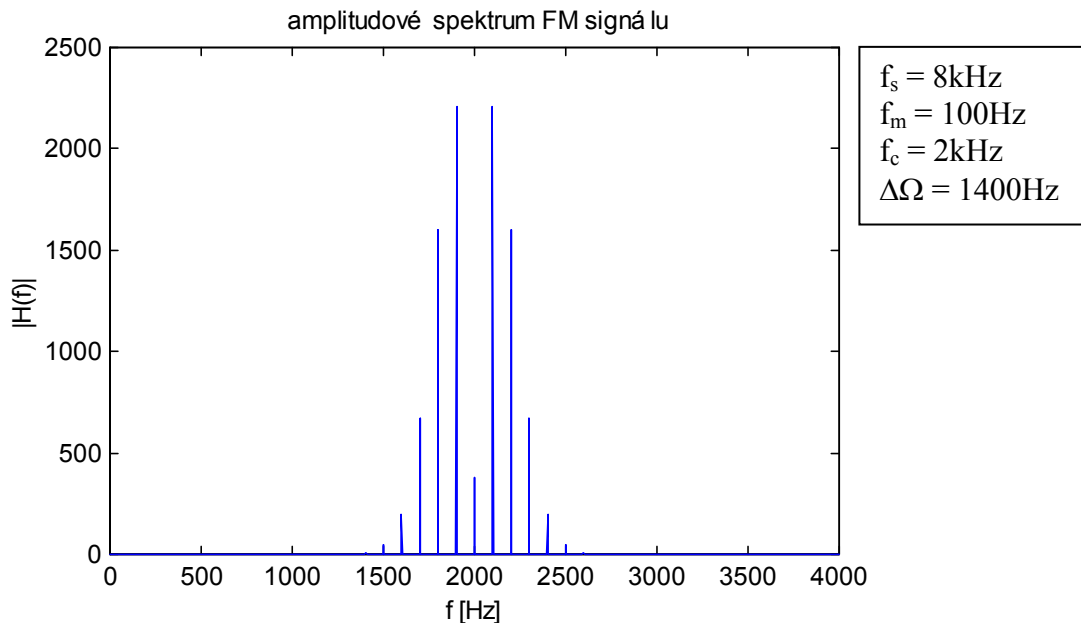
Pro FM signál můžeme tedy psát

$$s_{FM}(t) = \sin(\varphi_c + M \sin(\omega_M t)). \quad (2.8)$$

Modulaci nosné harmonickým signálem odpovídá spektrum postranních kmitočtů rozložených s konstantním krokem  $\omega_M$  symetricky vůči nenulové  $\omega_c$ . Polohy i amplitudy jednotlivých spektrálních čar jsou závislé na modulačním indexu  $M$ , tedy na amplitudě a kmitočtu modulačního signálu, jejich velikost lze určit pomocí Besselových funkcí. Při modulaci nosné neharmonickým signálem přecházejí postranní kmitočty do postranních pásem.

Předností FM syntézy je fakt, že již pouhé dva harmonické signály (tj. dva operátory) umožňují vytváření signálů s komplikovanými spektry za minimálních nároků na výpočet. V praxi se můžeme setkat i s víceoperátorovou syntézou, která poskytuje další rozšíření možností tvorby zvuku.

Pokud známe spektrum tónu reálného nástroje, můžeme se mu FM syntézou snažit přiblížit, nebo vytvářet zcela nové barvy zvuku. Na *obr. 2.4* je ukázáno spektrum signálu zvuku kuličkové píšťalky vytvořené dvouoperátorovou FM syntézou.



Obr. 2.4: Spektrum FM signálu – zvuk kuličkové píšťalky

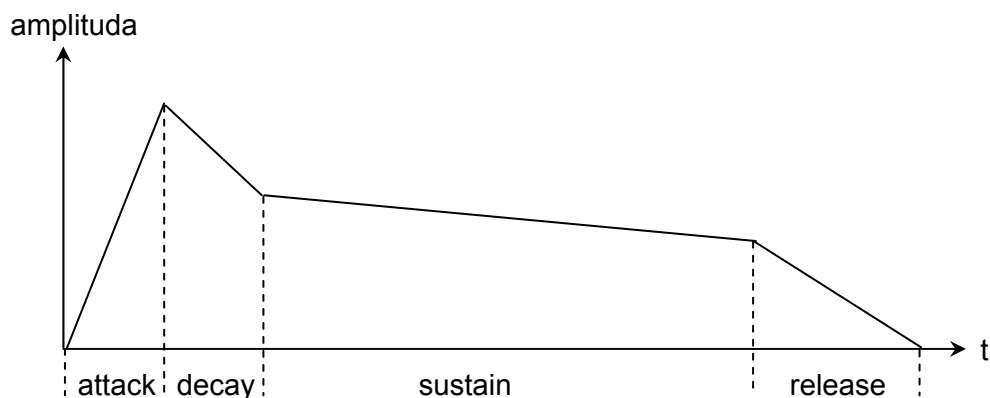
Daní za výpočetní nenáročnost této metody je fakt, že přiblížit se požadovanému průběhu spektra může být značně náročné. Proto se FM syntéza hodí spíše k modelování „jednoduchých“ nástrojů či tvorbě zcela nových zvuků. Krom harmonických modulačních signálů se používají i trojúhelníkové, pilové, obdélníkové a jiné průběhy. FM signálu je na konci řetězce přidána obálka, která má zásadní vliv na konečnou podobu zvuku. Obálce se budu podrobněji věnovat v následujících oddílech 2.5 a 2.6.

## 2.5 Wave Table syntéza

Wave Table syntéza pracuje s navzorkovanými zvuky nástrojů uloženými v paměti syntezátoru. Umožňuje tak lépe se přiblížit zvuku přirozených nástrojů. Jelikož kvalitní záznam všech tónů hracelných nástrojů by zabíral příliš paměťového prostoru, používají se různé techniky pro redukci potřebného množství navzorkovaného materiálu pro syntézu. Dvě základní metody úspory paměťového prostoru jsou Sample Looping a Pitch Shifting.

### 2.5.1 Sample Looping

Tato metoda využívá znalostí vývoje tónu nástroje. Velká skupina nástrojů má podobný průběh obálky tónu, jako ukazuje obr. 2.5. Jedná se především o nástroje, u nichž dochází k perkusivní tvorbě tónu.



Obr. 2.5: Typický průběh ADSR obálky

Zkratka ADSR je tvořena počátečními písmeny pojmenování jednotlivých fází vývoje obálky. První fází je attack - nasazení tónu provázené prudkým nárůstem amplitudy. Následně dochází k odeznění přechodového děje, provázené relativně rychlým poklesem amplitudy – decay. V další fázi, nazvané sustain, je tón prakticky v ustáleném stavu a obálka pozvolně klesá. Poslední fází je zánik tónu - release, kdy dojde k prudkému poklesu amplitudy na nulovou hodnotu.

ADSR obálka má nejen přímou spojitost s hlasitostí tónu, ale i s jeho spektrální charakteristikou, neboli barvou. Při attacku se obvykle krátkodobé spektrum rychle mění a obsahuje velký počet vyšších harmonických. Po odeznění ve fázi decay se naopak stabilizuje a po relativně dlouhou dobu nemění příliš svůj charakter, pouze dochází k mírnému ubývání vyšších harmonických. Ve fázi release dochází opět k prudké změně barvy, způsobené zatlumením tónu.

Pokud má nástroj, jehož syntézu provádíme, takovýto průběh obálky, stačí mít uložen v paměti úsek odpovídající fázím attack a decay a krátký segment z fáze sustain, který můžeme ve smyčce opakovaně přehrávat, zatímco generátor obálky bude snižovat amplitudu signálu. Attack a decay trvají obvykle v poměru k fázi sustain nepatrně dlouho, dojde tedy k velké úspoře paměťového prostoru.

## 2.5.2 Pitch Shifting

Dalším používaným způsobem pro úsporu paměťového prostoru je Pitch Shifting – transpozice výšky tónu. Máme-li v paměti uložen tón určité výšky, můžeme jej transponovat změnou rychlosti pohybu v paměti vzorků, tj. změnou

kroku adresového ukazatele. Abychom zabránili výraznému zkreslení signálu, je výhodné posupovat následujícím způsobem.

### 2.5.2.1 Decimace signálu

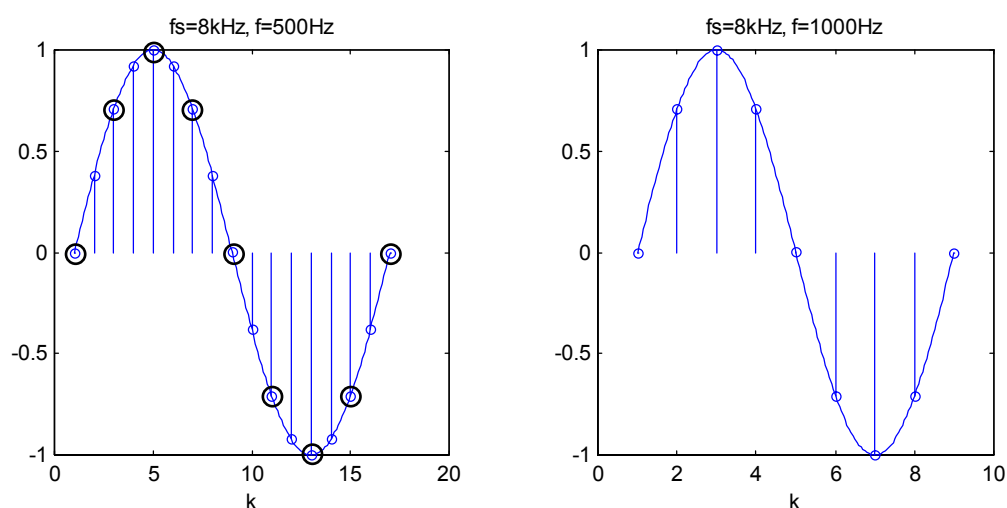
O decimaci signálu hovoříme v souvislosti se zvyšováním poměru frekvence signálu ke vzorkovací frekvenci, resp. zvyšováním výšky tónu. Pro signál o frekvenci  $f$  vzorkovaný  $f_s$  můžeme psát

$$f = \frac{f_s}{N}, \quad (2.9)$$

kde  $N$  je počet vzorků připadajících na jednu periodu. Změnou délky periody měníme frekvenci signálu a tedy vnímanou výšku tónu. Pro změnu frekvence signálu z  $f_1$  na  $f_2$  je třeba změnit počet vzorků periody následovně:

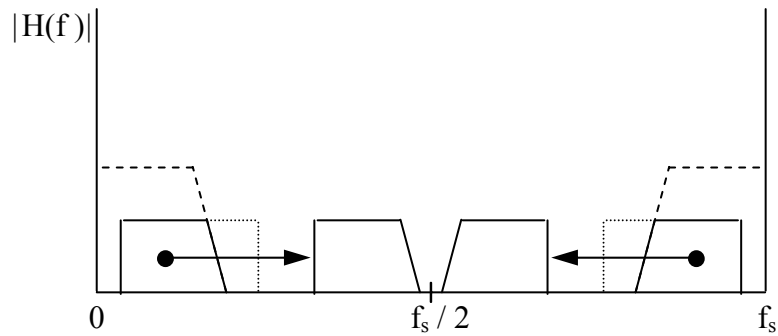
$$\frac{f_2}{f_1} = \frac{N_1}{N_2} = D, \quad (2.10)$$

kde  $D$  je faktor udávající decimaci – míru vypouštění vzorků.



Obr. 2.6: Decimace signálu,  $D = 2$

Na obr. 2.6 je příklad zvýšení frekvence na dvojnásobek – transpozice o oktávu. Signál vpravo tvoříme decimačním výběrem vzorků původního signálu. Ve frekvenční rovině se decimace projevuje posunem spektra signálu směrem k  $f_s/2$ .



Obr. 2.7: Vliv decimace na spektrum obecného reálného signálu

Abychom zabránili spektrálnímu aliasingu, je obecně třeba před decimací signál filtrovat dolní propustí, jak ukazuje obr. 2.7. Amplitudové spektrum reálného signálu je zde znázorněno obdélníky symetricky umístěnými podle  $f_s/2$ , přenos dolní propusti čárkovanou čarou. Výsledné spektrum po decimaci je tvořeno dvěma lichoběžníky opět v symetrii podle  $f_s/2$ . Z obrázku je patrné, že dolní propust zde zabraňuje prolnutí spektrálních pásem.

### 2.5.2.2 Interpolace signálu

Transponujeme-li výšku tónu směrem dolů, je třeba zvýšit počet vzorků na periodu signálu, čehož dosáhneme interpolací vzorků, které máme k dispozici.

Jednou z nejjednodušších metod interpolace je proklad vzorku původního signálu nulovými vzorky. Počet vkládaných nul opět vyplývá z poměru výchozí a cílové frekvence. Jestliže mezi každé dva vzorky původního signálu vložíme  $p$  nulových vzorků, spektrum původního signálu se nám  $p$ -krát periodicky zopakuje ve spektru nového signálu. Tuto periodicitu lze snadno ověřit.

Interpolací signálu  $s_1(n)$  sériemi  $p$  nulových vzorků získáme signál

$$s_2(n) = s_1\left(\frac{n}{p}\right), \quad n = 0, p, 2p, 3p, \dots \quad (2.11)$$

$$s_2(n) = 0, \quad \text{jinde.}$$

Diskrétní Fourierova transformace je definována [1]

$$X(k) = \sum_{n=0}^{N-1} s(n) e^{-j2\pi nk/N}, \quad k = 0, 1, \dots, N-1. \quad (2.12)$$

$N$  je počet prvků diskrétní posloupnosti  $s(n)$ ,  $X(k)$  je  $k$ -tá čára komplexního spektra signálu. Pro jednoduchost budeme předpokládat, že délka  $N$  signálu  $s_2(n)$  je celým násobkem  $p$ . Z (2.12) plyne, že Fourierovou transformací signálu délky  $N$  získáme spektrum tvořené  $N$  spektrálními čarami. Pokud má být tvořeno  $p$ -krát se periodicky opakující složkou, délka periody bude

$$N_{per} = \frac{N}{p}. \quad (2.13)$$

Vzhledem k tomu, že jeden nenulový vzorek  $s_2(n)$  je následován  $p$  nulami, můžeme pro spektrum psát

$$X(k) = \sum_{n=0}^{N-1} s(n)e^{-j2\pi nk/N} = \sum_{n=0}^{N/p-1} s(np)e^{-j2\pi npk/N}. \quad (2.14)$$

Platí-li (2.13), musí také platit

$$X(k) = X\left(k + \frac{N}{p}\right), \quad k = 0, 1, \dots, N - \frac{N}{p} - 1. \quad (2.15)$$

Dosazením (2.14) do (2.15) dostaneme

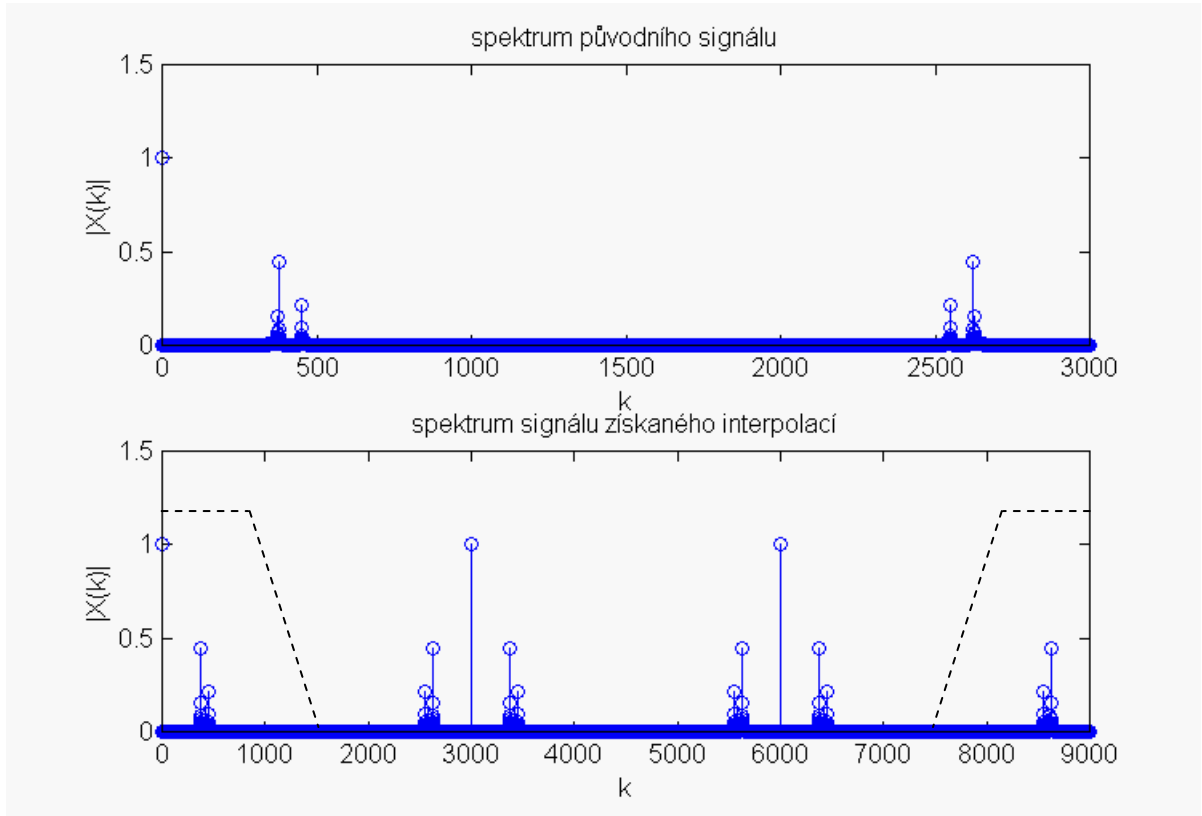
$$\sum_{n=0}^{N/p-1} s(np)e^{-j2\pi npk/N} = \sum_{n=0}^{N/p-1} s(np)e^{-j2\pi np(k + \frac{N}{p})/N}. \quad (2.16)$$

Úpravou exponenciály z pravé strany rovnice dostaneme

$$e^{-j2\pi np(k + \frac{N}{p})/N} = e^{-j2\pi np(\frac{k}{N} + \frac{1}{p})} = e^{-j2\pi npk/N} \cdot e^{-j2\pi n} = e^{-j2\pi npk/N}. \quad (2.17)$$

Rovnice (2.16) a tedy i (2.15) platí, tím je tedy potvrzena periodicitu spektra signálu získaného interpolací nulovými vzorky.

Dále víme, že amplitudové spektrum reálného signálu je až na stejnosměrnou složku symetrické podle  $f_s/2$  [1], proto jednotlivé periodicky se opakující části spektra signálu získaného interpolací budou symetrické podle svého středu, s výjimkou vždy prvního prvku. Pokud  $N$  nebude celým násobem  $p$ , dojde v porovnání s předchozími úvahami k „rozostření“ spektrálních čar, nicméně periodický charakter spektra zůstane zachován.



Obr. 2.8: Vliv interpolace nulovými vzorky na spektrum signálu;  $p=3$

Na obr. 2.8 je ukázán případ, kdy signál složený ze dvou sinusovek relativně blízké frekvence byl interpolován nulovými vzorky, vždy třemi na jeden původní vzorek. Podle předpokladu je spektrum tvořeno třemi po sobě jdoucími vnitřně symetrickými identickými částmi.

Abychom získali spektrum odpovídající výškově transponovanému tónu, je tedy nutno odstranit nežádoucí složky spektra dolní propustí, její přenos je na obr. 2.8 naznačen čárkovaně. Je potřebné, aby dolní mez nepropustného pásma začínala nejpozději v bodě

$$k_{stop} = \frac{N}{2p} - 1. \quad (2.18)$$



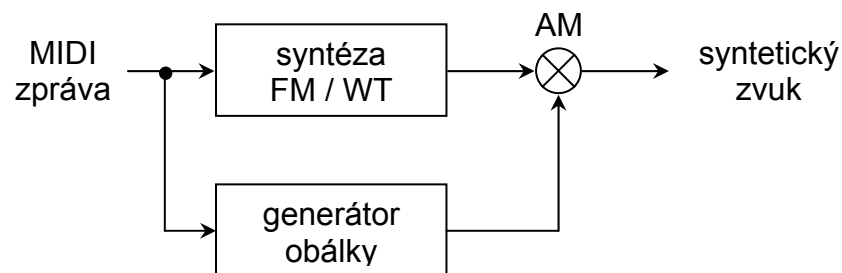
Je-li signál vzorkován  $f_s$ , dolní mezi nepropustného pásma  $k_{stop}$  odpovídá frekvence

$$f_{stop} = \frac{f_s}{N} \cdot k_{stop}. \quad (2.19)$$

## 2.6 Generátor obálky tónu

V oddílu 2.5.1 jsme popsali průběh ADSR. Tento tvar obálky však není společným jmenovatelem pro všechny nástroje. Kupříkladu dechové a smyčcové nástroje umožňují dobře kontrolovat průběh obálky při attacku i po něm, v kontrastu s perkusivními nástroji může být attack velmi pozvolný. Průběh amplitudové obálky má velký vliv na vnímání charakteru nástroje posluchačem, může převážet i nad barvou tónu – příkladem je efekt hry smyčců na elektrické kytáře docílený postupným otevíráním hlasitostní clony až po odeznění attacku.

Proto jak u FM syntézy tak Wave Table syntézy hraje velkou roli generátor obálky, který amplitudově moduluje syntetický signál a dotváří tak výsledný zvuk – obr. 2.9.



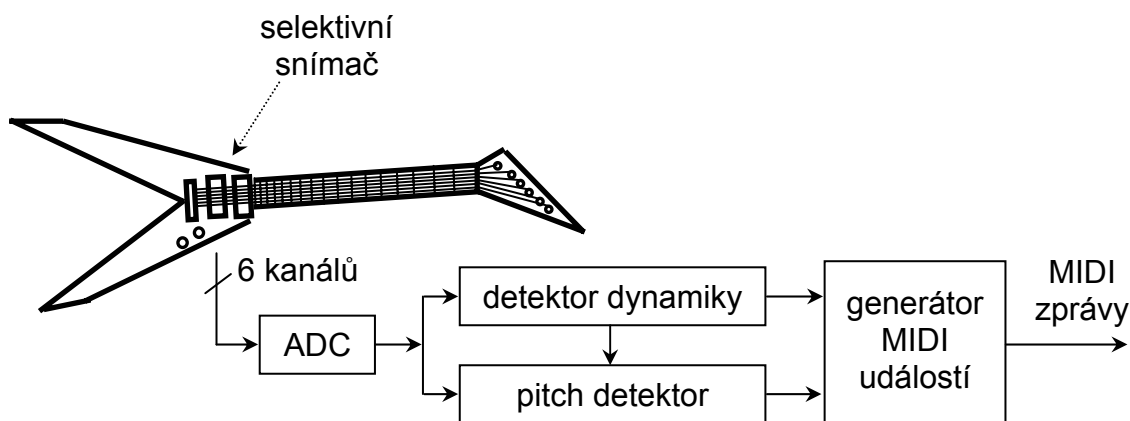
Obr. 2.9: Koncepce syntezátoru

Pokud jsou změny v průběhu pozvolné, vnímáme je jako změnu hlasitosti hraného tónu, pokud se frekvence změn blíží nebo zcela spadá do pásma slyšitelných kmitočtů, dochází k změně barvy tónu. Generátor obálky tak může vytvářet krom průběhů odpovídajícím přirozenému vývoji hlasitosti tónu také průběhy sinusové, pilové, obdélníkové (tremolo efekt), apod.

### 3 Syntetická kytara

Metod, jak snímat události vznikající při hře na kytaru, byla vyzkoušena celá řada, zdaleka ne všechny se však setkaly s pochopením hudebníků. Například firma Vox vyrobila v 60. letech nástroj nazvaný Guitar Organ, u kterého byl každý pražec rozdělen na 6 izolovaných částí příslušejících jednotlivým strunám a generování varhanního zvuku bylo spuštěno dotekem struny a pražce [2]. Tento způsob snímání nutí hráče ke značné modifikaci techniky hry, což je nevýhodné.

V současnosti používaná koncepce kytarového MIDI převodníku je založena na použití selektivního snímače a prakticky nevyžaduje žádnou zásadní změnu v přístupu ke hře. MIDI kontroler uvažovaný v této práci, založený na selektivním snímači, má podobu znázorněnou na *obr. 3.1*.



*Obr. 3.1: Architektura kytarového MIDI převodníku*

#### 3.1 Selektivní snímač

Selektivní snímač, který jsem použil při návrhu MIDI převodníku, je tvořen šesti oddělenými magnetodynamickými snímači přiřazenými jednotlivým strunám. Proto je také nazýván Hex snímač. MIDI převodníkem podle *obr. 3.1* vyhodnocujeme signály od jednotlivých strun odděleně, čímž se vyhýbáme potřebě analyzovat jeden polyfonní signál. Důvody této koncepce budou vysvětleny v kapitole 4. Na snímač je proto kladen nárok minimálních přeslechů od sousedních strun. Parametry snímače jsou uvedeny v příloze.

## 3.2 A/D převodník

A/D převodník obsahuje šest vstupních kanálů pro zpracování signálů od jednotlivých strun. Každý kanál je před převodem filtrován antialiasingovým filtrem, jehož parametry jsou dány vzorkovací frekvencí převodníku. Požadavky na vzorkovací frekvenci budou zmíněny v kapitole 4. Detaily o převodníku použitém v této práci jsou uvedeny v příloze.

## 3.3 Detektor dynamiky

Detektor dynamiky počítá energetické obálky signálů příslušejících jednotlivým strunám. Hodnota energetické obálky je použita pro určení hlasitosti tónu, a také slouží jako kritérium pro pozitivní detekci frekvence tónu – viz kapitola 5.

## 3.4 Pitch detektor

Pitch detektor rozpoznává výšku hraného tónu. Jsou na něj kladeny nároky na rychlost a spolehlivost pozitivní detekce. Je úhelným kamenem celého MIDI převodníku, a proto je jeho návrhu věnováno v této práci nejvíce pozornosti.

Pitch detektor je tvořen dvěma částmi. První část na základě vybrané metody určuje frekvence z průběhu signálu. Výběru vhodné metody je věnována kapitola 4. Druhá část pak z detekovaných frekvencí na základě určitých kritérií vybírá pouze platné hodnoty. Návrh těchto kritérií je proveden v kapitole 5.

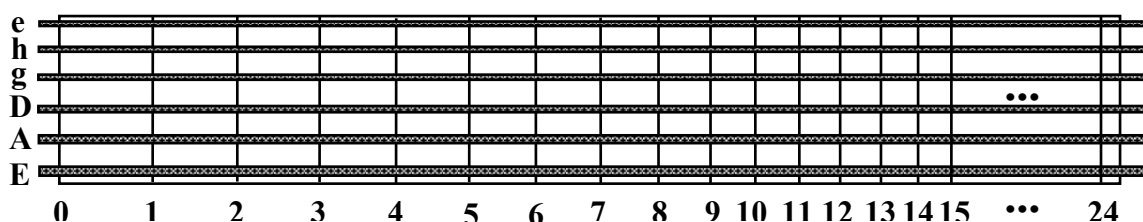
## 3.5 Generátor MIDI událostí

Generátor MIDI událostí informacím přicházejícím z Pitch detektoru a detektoru dynamiky přiřazuje MIDI události a z nich vytváří MIDI zprávy. Množství událostí, se kterými generátor pracuje, určuje míru využití možností poskytovaných MIDI protokolem.

## 4 Vybrané metody určení základní frekvence tónu

V následujících úvahách budeme v souladu s koncepcí MIDI převodníku zmíněnou v kapitole 3 předpokládat, že signál z elektrické kytary bude snímán pro každou strunu zvlášť. Vyhneme se tak potřebě analyzovat polyfonii a získáme možnost přistupovat k jednotlivým signálům podle znalostí vlastností jednotlivých strun (frekvenční rozsah tónů, spektrální zastoupení vyšších harmonických).

Běžné šestistrunné elektrické kytary mají 22 nebo 24 pražců, každý pražec představuje pro danou strunu zvýšení tónu o jeden půltón. V jedné oktávě se nachází 12 půltónů [3], pokud tedy hmatník obsahuje 24 pražců, má každá struna rozsah 2 oktávy. Nejobvyklejší ladění strun kytary je zobrazeno na *obr. 4.1*.



*Obr. 4.1: Obvyklé ladění strun kytary*

Struny *E*, *A*, *D* se často označují jako basové a struny *g*, *h*, *e* jako melodické. Tuto terminologii budu v textu v případě potřeby používat.

Vytahováním struny můžeme měnit výšku tónu (pitch bending) plynule i v rozsahu mezi sousedními půltóny. Proto bude jedním z požadavků na zkoumané metody frekvenční rozlišení odpovídající alespoň jednomu čtvrttónu a dále minimální zpoždění detekce – pro použitelnost v reálném čase. Nejprve se budeme zabývat možnostmi detekce ve frekvenční oblasti (DFT, LPC), poté v časové oblasti (banka filtrů, autokorelace, přímé měření periody).

Popsané metody jsem implementoval a testoval v programovém prostředí Matlab. V některých případech zmíním pro názornost specifika této implementace.

## 4.1 Diskrétní Fourierova transformace (DFT)

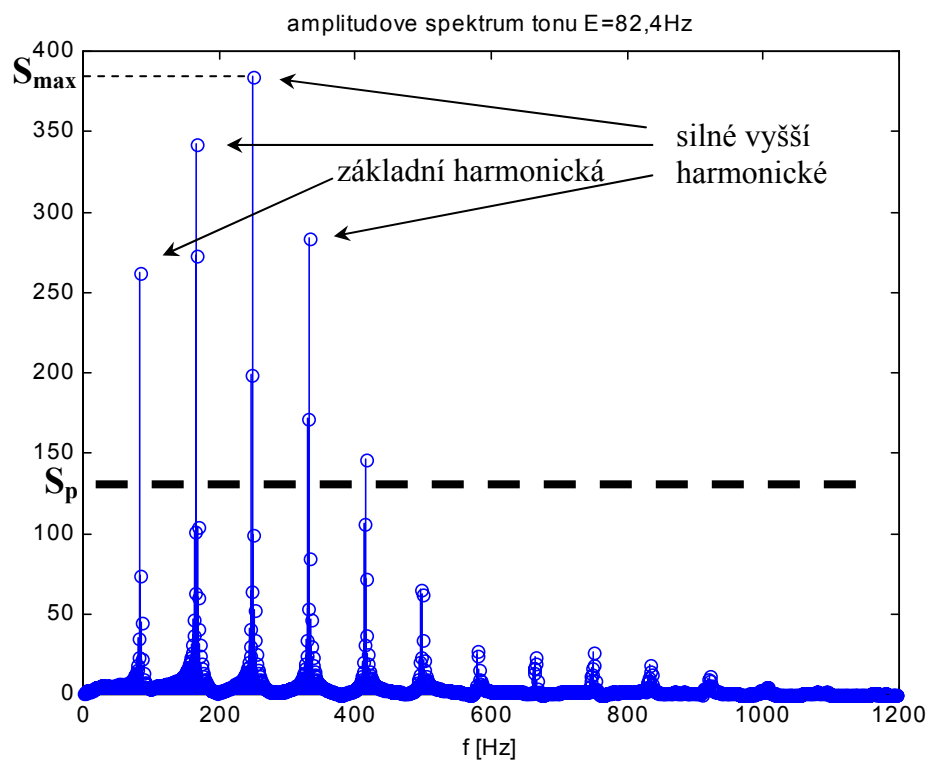
Diskrétní Fourierova transformace (DFT), zmíněná již v kapitole 2, je definována vztahem

$$S(k) = \sum_{n=0}^{N-1} s(n) e^{-j2\pi kn/N}, \quad k = 0, 1, \dots, N-1, \quad (4.1)$$

který váže diskrétní posloupnost  $s(n)$  a diskrétní periodické komplexní spektrum  $S(k)$  signálu  $s(n)$  [4]. DFT nám tedy umožňuje analyzovat signál ve frekvenční oblasti.

### 4.1.1 Amplitudové spektrum

Typické spektrum kytarového tónu je na *obr. 4.2*.



*Obr. 4.2: Amplitudové spektrum tónu E = 82,4Hz*

Jak je z *obr. 4.2* patrné, u kytarového tónu není základní harmonická, tedy frekvence odpovídající výšce tónu, vždy nejsilnější.

Abychom získali komplexnější představu o možných realizacích amplitudového spektra kytarového tónu, provedl jsem analýzu signálu v závislosti na modifikovatelných vlastnostech

nástroje či způsobu hry na něj. Poměr amplitud základní harmonické a několika dalších silných harmonických závisí na typu a stáří strun – obecně vykazuje největší obsah vyšších harmonických nová ocelová prázdná struna (není přitisklá na žádný pražec). Dále poměr závisí na délce kmitající části struny – se hrou ve vyšších polohách zastoupení vyšších harmonických klesá, stejně tak s narůstající dobou znění tónu od jeho rozeznění - attacku. Další vliv na barvu tónu má poloha snímačů. Čím blíže je snímač umístěn ke kobylce, tím obsah a amplitudy vyšších harmonických rostou – tón má ostřejší barvu, totéž platí o místě úderu do struny při attacku. Tyto poznatky jsou zcela v souladu s fyzikálními předpoklady o vlastnostech kmitající struny.

Analýza získaných amplitudových spekter ukázala, že v extrémním případě z hlediska množství zastoupených vyšších harmonických, kdy udeříme do struny v přímé blízkosti kobylky a strunu v tomtéž místě i snímáme, je nejsilnější harmonická s dvojnásobnou či trojnásobnou frekvencí vůči základní harmonické a má maximálně 2,5 až 3-krát vyšší amplitudu.

Základní harmonickou, a tedy i aktuální výšku hraného tónu, můžeme pak určit na základě vhodného prahování amplitudového spektra, jak je naznačeno na *obr. 4.2*. Prahovou hodnotu  $S_p$  můžeme volit z amplitudy nejsilnější harmonické  $S_{max}$ :

$$S_p = S_{max} \cdot th, \quad (4.2)$$

kde konstanta  $th$  představuje poměrný prahový koeficient.

Prahový koeficient jsem vzhledem k výše uvedeným poznatkům o spektru s dobrými výsledky volil  $th = 0,3$ . Základní harmonickou nalezneme coby nejnižší harmonickou, která je vyšší než stanovený práh.

#### 4.1.2 Polyfonní signál

Amplitudové spektrum polyfonního signálu vzniká součtem amplitudových spekter signálů od jednotlivých strun kytary. Jak jsme již zdůraznili, obvykle není hlavní harmonická jednotlivých tónů nejsilnější. Po sečtení jejich spekter již nemůžeme použít pro detekci jednoduché prahování, ztrácíme informaci, na jaké harmonické se soustředit. Navíc u kytary můžeme zahrát souzvuk i několika tónů stejné výšky, neboť rozsahy strun se značně překrývají. V tomto případě je třeba volit algoritmus analýzy, který vede k rozkladu

polyfonního spektra na složky odpovídající jednotlivým tónům. Řešením může být výpočet vzájemné energie známého spektra hledaného tónu se spektrem polyfonním. Pokud najdeme typické spektrum tónu pro určitý tónový rozsah, můžeme výpočet provádět pomocí korelace typického spektra s polyfonním, posuv v korelaci odpovídá změně výšky tónu. S posuvem musíme vždy přepočítat polohy složek typického spektra, abychom zachovali jejich frekvenční poměry. Tato metoda je výpočetně náročná a nevede zřejmě k spolehlivým výsledkům. Zároveň vykazuje i nevýhody uvedené v odstavci 4.1.4, platné pro detekci v jednohlasém signálu.

### 4.1.3 Implementace DFT detekce v Matlabu

Pro hledání lokálních maxim ve spektru (peaků), resp. v jakémkoliv segmentu dat, jsem napsal v Matlabu funkci *peaks()*. Vstupem je analyzovaný segment dat a výstupem je vektor s hodnotami nalezených lokálních maxim v prvním řádku a příslušnými indexy jejich pozic v druhém řádku. Lokální maximum nalezneme podle změny znaménka difference mezi sousedními prvky

$$d(n) = s(n) - s(n-1), \quad n = 1 \dots N. \quad (4.3)$$

Indexování zde provádíme od jedné, v souladu s konvencí Matlabu.

Pokud

$$(sign(d(n)) = -1) \& (sign(d(n-1)) = 1), \quad (4.4a)$$

kde znak “&” značí logickou konjunktci a pro *sign()* platí

$$\begin{aligned} sign(x) &= 1, & x > 0 \\ sign(x) &= 0, & x = 0 \\ sign(x) &= -1, & x < 0, \end{aligned} \quad (4.4b)$$

pak  $s(n-1)$  je lokální maximum. Pokud po sobě následuje několik prvků stejné hodnoty (ploška), pak je třeba si zapamatovat, zda funkce před touto ploškou naposledy stoupala

či klesala a podle toho opět nalezneme maximum, pokud

$$(\text{sign}(d(n)) = -1) \ \& \ (\text{sign}(d(n-k)) = 1), \quad (4.5)$$

kde diference  $d(n-k)$  je poslední nenulová diference před začátkem plošky. Pokud je maximum tvořeno takovouto ploškou, funkce  $peaks()$  vrátí polohu lokálního maxima coby středu plošky.

Frekvence tónové řady leží na exponenciále (pokud zvolíme za první tón např. subkontra  $A_2 = 22,5\text{Hz}$ ), pak

$$f_i(m) = A_2 \cdot 2^{\frac{m}{12}}, \quad m = 0, 1, 2, \dots, \quad (4.6)$$

kde  $f_i(m)$  je frekvence  $m$ -tého půltónu tónové řady [3].

Pro určení jména tónu z detekované frekvence základní harmonické jsem na základě vztahu (4.6) napsal funkci  $note()$ . Jako vstup má detekovanou frekvenci a jako výstup jméno tónu, číslo oktávy, ve které se nachází (0 pro subkontra oktávu), správnou výšku tónu (dle ladění komorního  $a = 440\text{Hz}$ ) a odchylku v %.

Pro detekci tónu v hudebním signálu jsem napsal funkci  $tuner()$ . Vstupem je analyzovaný signál, vzorkovací frekvence, výstup je identický s výstupem funkce  $note()$ , navíc je přidána hodnota detekované frekvence. Funkce  $tuner()$  funguje dobře pro dlouhé vstupní signály, pro krátké přesnost detekované frekvence klesá. Důvody uvedeme v následujícím odstavci.

#### 4.1.4 Vztah mezi časovým a frekvenčním rozlišením DFT

Z důvodu potřeby rychlé detekce tónu máme požadavek na velké časové rozlišení analýzy. DFT aplikovaná na krátké segmenty signálu, často nazývaná STDFT (Short Time DFT), má předpis shodný s DFT, navíc zde nabývá na důležitosti aplikace váhovacího okna  $w(n)$  na segment signálu:

$$S(k) = \sum_{n=0}^{N-1} w(n)s(n) e^{-j\frac{2\pi kn}{N}}, \quad k = 0, 1, \dots, N-1. \quad (4.7)$$



Váhovací okno má za úkol zabránit prosakování spektra, způsobenému segmentací signálu [4]. Obvykle při výpočtech uvažujeme, že signál mimo segment je nulový, a tím se vystavujeme nebezpečí, že na přechodu mezi doplněnými nulami a začátkem segmentu vznikne váhovaný jednotkový skok, který v původním signálu nebyl. Váhovací okno může zajistit pozvolný náběh amplitudy v segmentu a tudíž skok odstranit. Dále se váhovací okna používají pro dělení signálu na segmenty s překryvem, pak má okno za úkol také zajistit, aby součet překryvů rekonstruoval věrně původní signál – metoda OLA (Overlap And Add). Jak je však v [4] ukázáno, váhovací okno může způsobit určité nežádoucí rozostření spektrálního odhadu. Z tohoto důvodu jsem použil obdélníkové váhovací okno, které nemění amplitudové poměry v signálu a pouze vybírá jednotlivé segmenty.

Ze (4.7) je patrné, že pro segment signálu délky  $N$  získáme spektrum o  $N$  spektrálních čarách. Frekvenční rozlišení DFT je tedy

$$\Delta f = \frac{f_s}{N}. \quad (4.8)$$

Toto rozlišení se také zřejmě nezvýší se zvýšením  $f_s$ , jen dojde k rozšíření pásma analyzovaných frekvencí. Frekvenční rozlišení můžeme zvýšit jedinečně prodloužením segmentu – na úkor časového rozlišení. Další nevýhodou je, že toto frekvenční rozlišení je lineární pro celý rozsah frekvencí, zatímco frekvence tónové řady leží na exponenciále podle (4.6). Pro nízko položené tóny tak určujeme frekvenci s velkou relativní chybou, zatímco pro vyšší tóny chyba klesá. Čtvrttón od nejhlubšího tónu na kytáře  $E = 82,4\text{Hz}$  je vzdálen

$$\Delta f = 82,4 - 82,4 \cdot 2^{\frac{0,5}{12}} = 2,4\text{Hz}.$$

Abychom tedy dosáhli frekvenčního rozlišení na čtvrttóny v celém rozsahu kytary, potřebujeme analyzovat segment délky

$$\frac{1}{\Delta f} = \frac{1}{2,4} = 0,42\text{s}.$$

I kdybychom mohli zanedbat čas potřebný pro výpočet DFT, výšku tónu bychom mohli takto určit se zpožděním rovným minimálně délce segmentu – tedy  $0,42\text{s}$ . To je ovšem

nepoužitelné pro analýzu signálu v reálném čase. Navíc při analýze hudebního signálu o takovéto délce nemůžeme předpokládat, že byl hrán po celou dobu pouze jeden tón. V amplitudovém spektru bychom pak měli promítnuty všechny tóny zahrané v daném segmentu. Z tohoto důvodu je STDFT nepoužitelná pro rychlou detekci výšky tónu.

## 4.2 Lineární predikce

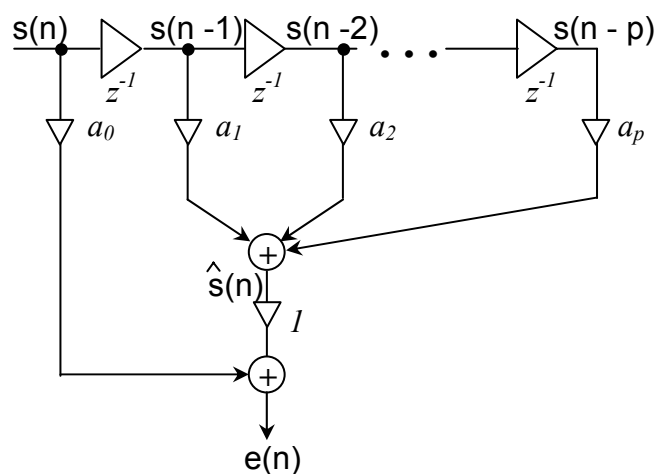
Metoda LPC (Linear Predictive Coding) vychází z předpokladu, že aktuální vzorek signálu můžeme predikovat z předchozích hodnot. Popíšeme ji v souladu s [4]. Prediktor  $p$ -tého řádu odhaduje  $n$ -tý prvek analyzovaného signálu na základě předchozích hodnot

$$\hat{s}(n) = -\sum_{k=1}^p a_k s(n-k), \quad (4.9)$$

kde  $a_k$  je  $k$ -tý koeficient prediktoru. Naším úkolem je nalezení takových koeficientů  $a_k$ , aby prediktor odhadoval co nejlépe hodnoty analyzovaného signálu  $s(n)$ . Zavedeme proto chybový signál

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^p a_k s(n-k) = \sum_{k=0}^p a_k s(n-k). \quad (4.10)$$

Ze (4.10) je zřejmé, že struktura výpočtu chybového  $e(n)$  je identická se strukturou FIR filtru (Finite Impulse Response), tedy číslicového filtru s konečnou impulsní odezvou, jehož koeficienty jsou rovny koeficientům prediktoru – obr. 4.3. Jedná se o filtr analyzující.



Obr. 4.3: Výpočet chybového signálu – analyzující filtr

Minimalizací výkonu chybového signálu  $e(n)$  získáme tzv. normální rovnice, ze kterých je možno vyjádřit velikosti  $a_k$ . Minimalizaci chyby  $e(n)$  budeme provádět metodou nejmenších čtverců, zavedeme proto kriteriální funkci

$$J = \sum e^2(n). \quad (4.11)$$

Meze v (4.11) jsou dány použitou metodou určení koeficientů  $a_k$ . Normální rovnice získáme hledáním minima kriteriální funkce (4.11), musí platit

$$\frac{\partial J}{\partial a_k} = 0, \quad k = 1, 2, \dots, p. \quad (4.12)$$

#### 4.2.1 Autokorelační metoda

Jedním z nejpoužívanějších postupů výpočtu koeficientů  $a_k$  je autokorelační metoda. Vycházíme z předpokladu, že analyzovaný segment signálu má konečnou délku  $N$  a vzorky signálu vně segmentu jsou nulové. Kriteriální funkci pak volíme

$$J = \sum_{n=0}^{N+p-1} e^2(n). \quad (4.13)$$

Minimalizací kritéria (4.13) získáme soustavu normálních rovnic. Postup jejich odvození je podrobně uveden v [4]. Výsledkem je soustava Yule-Walkerových rovnic:

$$\sum_{i=1}^p R(|i-j|) \cdot a_i = -R(j), \quad j = 1, \dots, p, \quad (4.14)$$

autokorelační koeficienty  $R(k)$  spočteme

$$R(k) = \frac{1}{N} \sum_{n=0}^{N-k-1} s(n)s(n+k), \quad k = 1, \dots, p-1. \quad (4.15)$$

Maticová podoba (4.15) je následující:

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(p-1) \\ R(1) & R(0) & \cdots & R(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & \cdots & R(0) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix}. \quad (4.16)$$

Je zřejmé, že autokorelační matice na levé straně (4.16) je symetrická a ekvidiagonální. Dále je pozitivně-definitní, všechna její vlastní čísla jsou kladná.

Jak již bylo řečeno, výpočet chybového signálu odpovídá realizaci analyzujícího FIR filtru. Transformací (4.10) do  $z$ -roviny získáme přenosovou funkci FIRu

$$A(z) = \sum_{k=0}^p a_k z^{-k}. \quad (4.17)$$

Chybový signál  $e(n)$  na výstupu FIRu je pro daný řád  $p$  nekorelovaný a má pro náhodné signály charakter bílého šumu [4].



Obr. 4.4: Výpočet chybového signálu

Jak je znázorněno na obr. 4.4, inverzní filtrací syntetizujícím filtrem můžeme získat zpět analyzovaný signál. Vzhledem k vlastnostem  $e(n)$  můžeme analyzovaný signál získat buzením syntetizujícího filtru přímo bílým šumem. Pro zachování výkonových poměrů je třeba bílý šum zesílit

$$G = \sqrt{\frac{\sigma_e^2}{\sigma_u^2}}, \quad (4.18)$$

čitatel pod odmocninou je rozptyl chybového signálu a jmenovatel rozptyl bílého šumu.

Výsledná přenosová funkce syntetizujícího filtru buzeného bílým šumem je rovna

$$H(z) = \frac{G}{A(z)}. \quad (4.19)$$

Jelikož čtvercová matice v (4.16) je pozitivně-definitní, poskytuje vždy stabilní model syntetizujícího filtru [4]. Pro modelování výkonového spektra signálu můžeme psát

$$S_s(z) = |H(z)|^2 \cdot S_u(z), \quad (4.20)$$

kde  $S_u(z)$  je výkonové spektrum bílého šumu, pro které platí

$$S_u(z) = \sigma_u^2, \quad (4.21)$$

tedy je konstantní a rovno rozptylu bílého šumu [5]. Dosadíme-li (4.18), (4.19) a (4.21) do (4.20), získáme vztah

$$S_s(z) = \frac{\sigma_e^2}{|A(z)|^2}, \quad (4.22)$$

což můžeme po substituci  $e^{j\Theta}$  za  $z$  přepsat

$$S_s(e^{j\Theta}) = \frac{\sigma_e^2}{|A(e^{j\Theta})|^2}, \quad (4.23)$$

kde  $\Theta$  je normovaná frekvence resp. fáze definovaná

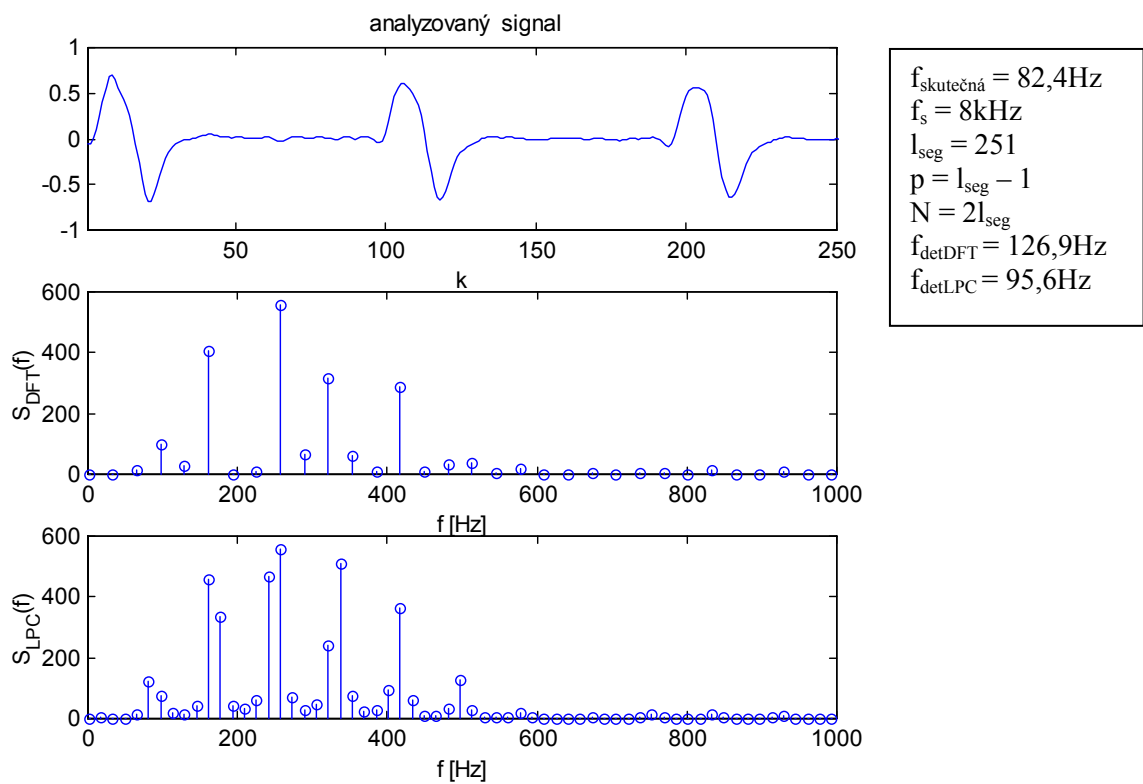
$$\Theta = 2\pi \frac{f}{f_s} \quad [\text{rad}]. \quad (4.24)$$

Výkonové spektrum modelovaného signálu je tedy rovno kvadrátu amplitudové charakteristiky syntetizujícího filtru.

Pro zvolený řád  $p$  můžeme určit odhad vyhlazeného výkonového spektra analyzovaného signálu. Počet maxim a míra vyhlazenosti spektra je přímo dána řádem syntetizujícího filtru, dvojici komplexně sdružených pólů přenosové funkce odpovídají dvě špičky symetrické podle  $f_s/2$ , reálnému pólu jedna špička na nulové frekvenci nebo na  $f_s/2$  [4]. Polynom ve jmenovateli (4.23b) je funkcí spojité fáze  $\Theta$ , při numerickém výpočtu výkonového spektra pro konečný počet spektrálních čar ji nahradíme diskrétní fází

$$\Theta_a(k) = \text{round}\left(\frac{2\pi}{N} \cdot k\right), \quad k = 0, 1, \dots, N-1, \quad (4.25)$$

kde  $N$  je požadovaný počet čar spektra a funkce  $\text{round}()$  zaokrouhluje na celá čísla. Míra přiblížení poloh maxim vyhlazeného spektra skutečným frekvencím význačných harmonických analyzovaného hudebního signálu závisí na vzorkovací frekvenci  $f_s$ , řádu LPC, délce analyzovaného segmentu a velikosti  $N$  ve smyslu (4.25). Volba  $f_s$  ovlivňuje autokorelační koeficienty (podrobněji se vlivu velikosti  $f_s$  budeme věnovat v oddílu věnovanému analýze v časové oblasti). Vliv výpočtu LPC spektra v konečném počtu bodů na přesnost detekce základní harmonické je patrný z následujícího příkladu – obr. 4.5, 4.6.



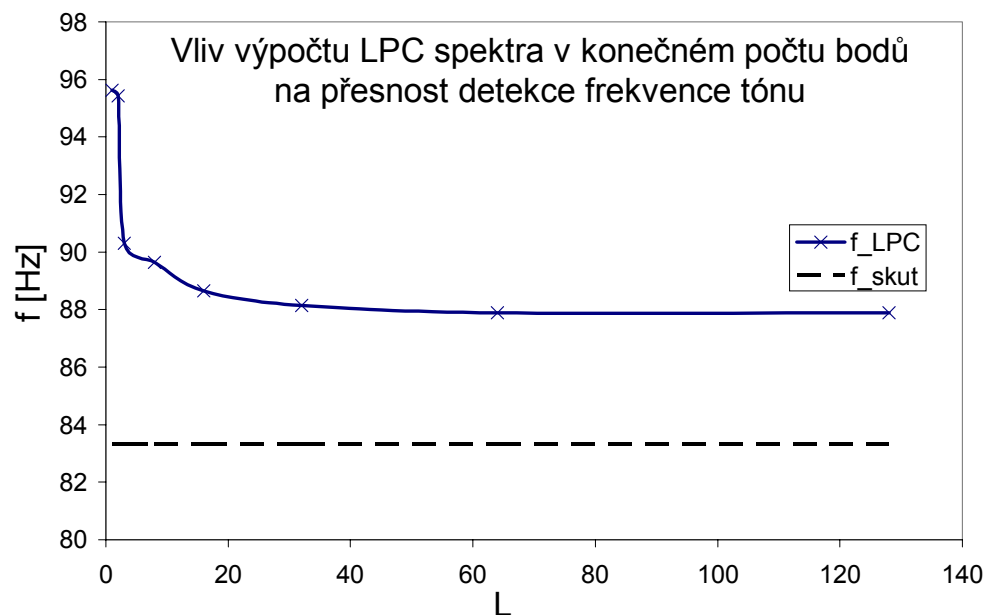
Obr. 4.5: Detekce  $f$  tónu z DFT a LPC výkonového spektra

Na obr. 4.4 je zobrazen segment analyzovaného signálu délky  $l_{seg}$  odpovídajícího tónu  $E = 82,4\text{Hz}$  a příslušná výkonová spektra určená z DFT a LPC řádu  $l_{seg} - 1$ . V tomto případě byl počet bodů  $N$  pro výpočet LPC spektra roven  $2l_{seg}$ , získáváme tak dvojnásobek počtu spektrálních čar oproti DFT, kde je jich jen  $l_{seg}$ . Ve spektru LPC můžeme pozorovat oproti DFT určité vyhlazení přechodů mezi maximy a minimy. Jak je patrné z tabulky v obr. 4.5, základní frekvence detekovaná z LPC spektra způsobem popsáním v 4.1.3 se blíží více skutečné frekvenci tónu nežli frekvence detekovaná z DFT.

Na obr. 4.6 je ukázán vliv konečného počtu bodů, v nichž LPC spektrum počítáme, na přesnost detekce základní frekvence tónu ze segmentu konstantní délky. Frekvence přiřazené proměnné  $L$  byly detekovány z LPC spektra napočítaného v  $N$  bodech:

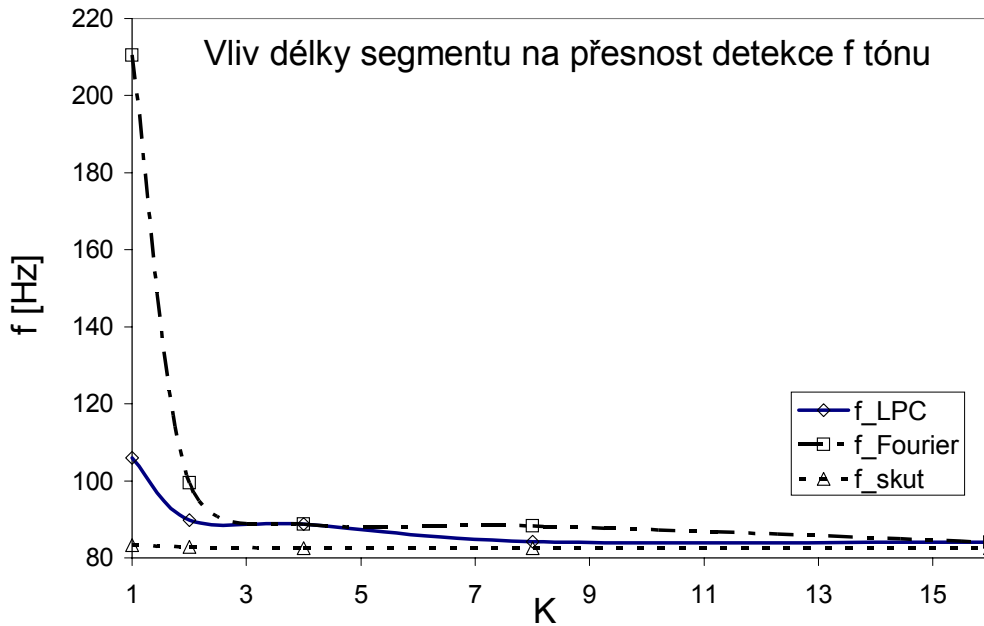
$$N = L \cdot l_{seg}. \quad (4.26)$$

Jak je z grafu patrné, se zvyšujícím se  $N$  přesnost detekce do určité doby roste, ale je omezena délkou analyzovaného segmentu signálu  $l_{seg}$ . Čárkovaně je v grafu vyznačena skutečná výška tónu analyzovaného signálu.



Obr. 4.6: Detekce  $f$  tónu z LPC spektra – vliv diskretizace spektra

Na obr. 4.7 je ukázán vliv délky segmentu analyzovaného signálu na přesnost detekce výšky tónu pomocí DFT a LPC.



Obr. 4.7: Vliv délky segmentu na přesnost detekce  $f$  tónu pro DFT a LPC

Délka segmentu zde byla

$$l_{seg} = 251 \cdot K. \quad (4.27)$$

Řád LPC zvolíme

$$\begin{aligned} p &= l_{seg} - 1, & K &= 1,3,5,\dots \\ p &= l_{seg} - 2, & K &= 2,4,6,\dots \end{aligned} \quad (4.28)$$

Důvodem odlišného výpočtu řádu LPC pro lichá a sudá  $K$  je potřeba, aby řád LPC byl vždy sudý. Získáme tak vždy jen komplexně sdružené póly syntetizujícího filtru. Jak již bylo zmíněno, reálné póly vznikající při lichém řádu generují špičku na nulovém kmitočtu nebo na  $f_s/2$ , což je v našem případě nežádoucí, neboť signál byl v daném segmentu vždy vystředěn a neobsahoval stejnosměrnou složku. Jak je z grafu patrné, pro krátké segmenty přináší detekce z LPC spektra přesnější výsledky než z DFT, pro delší segmenty je přesnost obou metod srovnatelná. Jak je z obr. 4.7 také vidět, detekce z LPC i DFT začíná vykazovat použitelnou přesnost rozlišení základní frekvence pro srovnatelnou délku segmentu, která odpovídá časové délce



$$t_{del} = \frac{l_{seg}}{f_s} = \frac{16 \cdot 251}{8000} = 0,5s. \quad (4.29)$$

Je zřejmé, že autokorelační LPC není použitelná pro tónovou detekci ze stejných důvodů jako DFT.

## 4.2.2 Kovarianční metoda

Kovarianční metoda výpočtu LPC se liší od autokorelační volbou kritériální funkce

$$J = \frac{1}{N-p} \sum_{n=p}^{N-1} e^2(n). \quad (4.30)$$

Z mezí sumy (4.30) vyplývá, že v tomto případě pracujeme pouze se vzorky segmentu, není zde tedy třeba trvat na nulovosti vzorků mimo segment. Postup odvození normálních rovnic je obdobný jako u autokorelační metody.

Normální rovnice mají následující podobu [4]

$$\sum_{i=1}^p \Phi(j, i) \cdot a_i = -\Phi(0, j), \quad j = 1, \dots, p. \quad (4.31)$$

Korelační koeficienty z (4.31) jsou definovány

$$\Phi(i, j) = \frac{1}{N-p} \sum_{n=p}^{N-1} s(n-i)s(n-j). \quad (4.32)$$

Soustavu rovnic (4.31) můžeme opět zapsat maticově

$$\begin{bmatrix} \Phi(1,1) & \Phi(1,2) & \cdots & \Phi(1,p) \\ \Phi(2,1) & \Phi(2,2) & \cdots & \Phi(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi(p,1) & \Phi(p,2) & \cdots & \Phi(p,p) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} \Phi(0,1) \\ \Phi(0,2) \\ \vdots \\ \Phi(0,p) \end{bmatrix}. \quad (4.33)$$

Matice soustavy je zde opět symetrická, ale není ekvidiagonální. Na rozdíl od autokorelační metody zde při výpočtu normálních rovnic závisí také na absolutním posunutí signálu, kovarianční matice proto přináší větší informaci o signálu. Z tohoto důvodu můžeme použitím kovarianční metody dosáhnout většího spektrálního rozlišení pro krátké segmenty signálu. Nevýhodou je možná nestabilita syntetizujícího filtru, kterou je třeba řešit stabilizací pólů jejich zrcadlením do jednotkové kružnice.

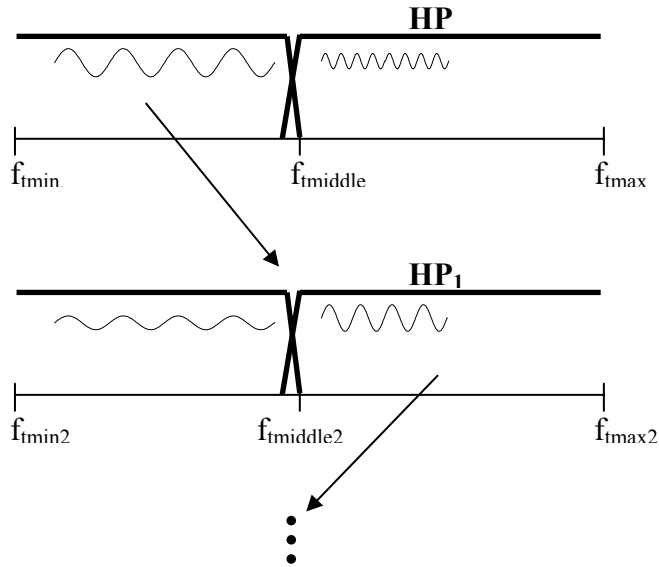
Tato metoda se ukázala jako nepoužitelná pro rychlou analýzu hudebního signálu, zvláště vzhledem ke své nestabilitě, a nebudeme se jí zde dále zabývat.

### 4.3 Tónová banka filtrů

Jednu z možností zkoumání vlastností signálu v časové oblasti přináší tónová banka filtrů. Pomocí vhodně zvolené banky můžeme provést dekompozici signálu na jednotlivá frekvenční pásma a porovnat jejich energii. Jak je ukázáno v [4], takovouto bankou filtrů je i DFT, kde každá spektrální čára energetického spektra reprezentuje energii pásma harmonických o šířce dané zpracovávanou délkou segmentu signálu. Jak již bylo zmíněno, nevýhodou je lineární frekvenční rozlišení DFT.

#### 4.3.1 Realizace banky filtrů půlením intervalu

Pro rozlišení jednoho čtvrttónu je třeba signál každé struny filtrovat bankou 48 pásmových propustí, neboť rozsah dvou oktáv připadající každé struně je tvořen 48 čtvrttóny. Požadavek na minimální přeslechy mezi jednotlivými pásmy vede na velmi strmé pásmové propusti vysokých řádů. Jinou možností je použít metodu půlení intervalu pomocí dolních (DP) a horních (HP) propustí, u nichž již požadavky na strmost nejsou tak velké a nevedou na tak vysoké řády – *obr. 4.8*.



Obr. 4.8: Banka filtrů realizovaná půlením intervalu

Frekvence  $f_{tmin}$ ,  $f_{tmiddle}$  a  $f_{tmax}$  vyznačují frekvenční pásma, v nichž určujeme energii signálu. Frekvence  $f_{tmiddle}$  definuje mez propustnostního pásma DP a HP. Filtrací signálu  $s$  dolní propustí získáme signál  $s_1$ , horní propustí  $s_2$ . Pro jejich energie pak platí

$$E_1 = \sum_{k=1}^N s_1^2(k), \quad E_2 = \sum_{k=1}^N s_2^2(k), \quad (4.34)$$

kde  $N$  je délka segmentu signálu  $s$ , ze které vyhodnocujeme příslušné energie. Pokud bude např. jako na obrázku  $E_{11} > E_{12}$ , signál  $s_2$  opět filtrujeme jak DP tak HP, tentokrát s mezí propustného pásma zvolenou v rozsahu předchozích  $f_{tmin}$  a  $f_{tmiddle}$ . Proces se opakuje až do okamžiku, kdy rozsah  $f_{tmin_k} - f_{tmax_k}$  odpovídá požadovanému frekvenčnímu rozlišení.

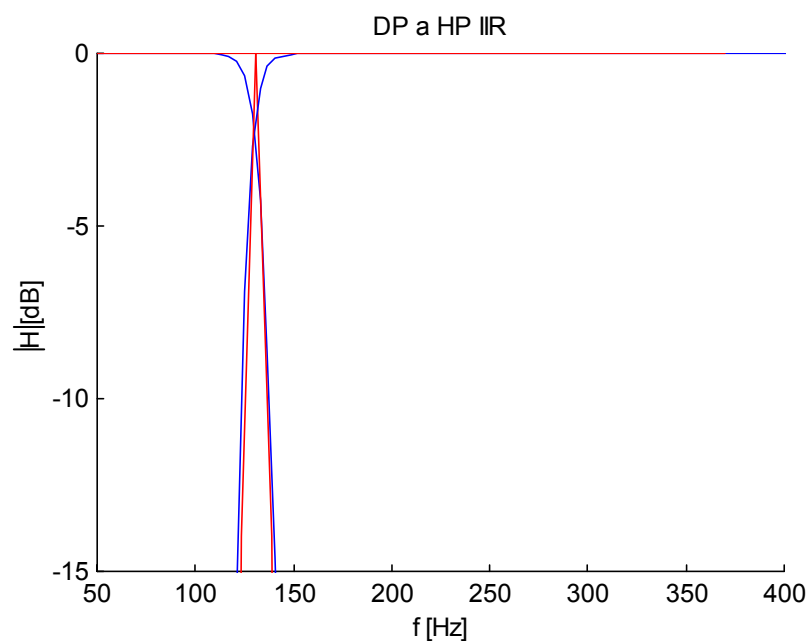
Při filtraci signálu dochází k jeho zpoždění danému řádem filtru a zkreslení vlivem přechodových dějů fázové charakteristiky filtru. Pro dosažení čtvrttónového rozlišení tato metoda předpokládá průchod signálu několika filtry po sobě. Zkreslení a zpoždění signálu tak narůstá.

Dalším problémem je interpretace významu získaných energií. Jak jsme ukázali v oddíle 4.1.1, základní harmonická odpovídající frekvenční výšce tónu nemá u kytary obvykle nejvyšší energii. Narozdíl od amplitudového spektra, kde stačí vybrat nejnižší harmonickou, která amplitudou překračuje vhodně zvolenou prahovou hodnotu, u banky filtrů získáváme součet energií širokého pásma harmonických, a nemůžeme tudíž snadno najít obdobu prahování jako u spektrálních čar. Řešením může být filtrace analyzovaného signálu dolní propustí zařazenou před banku filtrů, která omezí obsah vyšších harmonických takovým

způsobem, že součet jejich energií bude menší nežli energie základní harmonické. Tím ovšem zařazujeme do řetězce další zpoždění a zkreslení signálu.

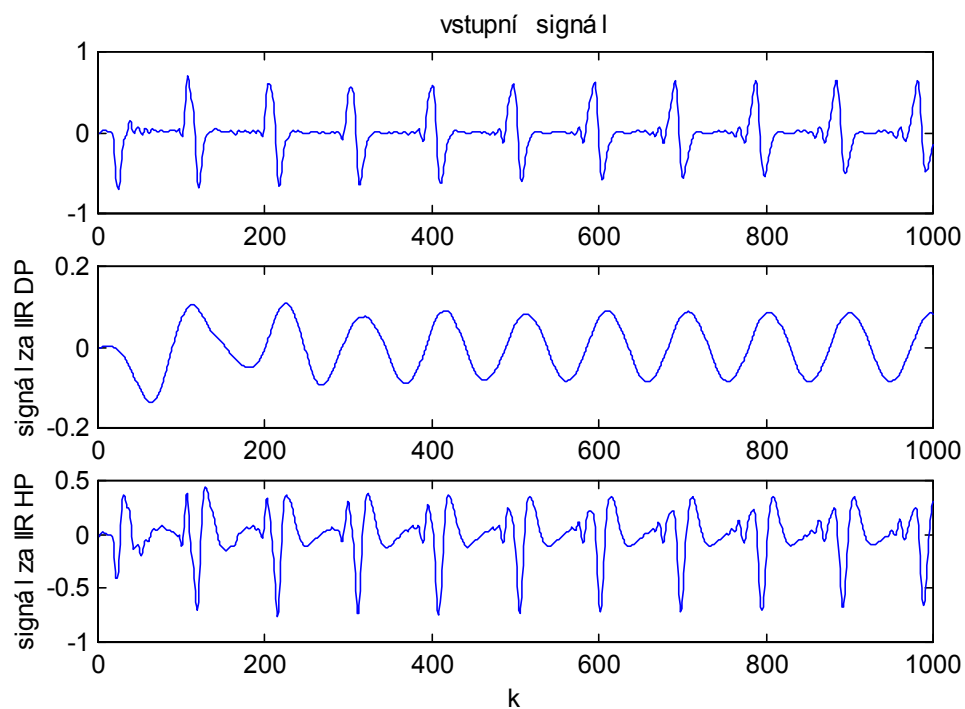
### 4.3.2 Ukázka realizace 1. kroku půlení intervalu

V Matlabu jsem navrhl DP a HP pro basovou strunu  $E$  odpovídající prvnímu článku filtrace bankou filtrů. Pro  $f_{middle}$  jsem zvolil frekvenci odpovídající tónu na 8. pražci – tedy  $c = 130,8Hz$ . Jako DP a HP IIR (Infinite Impulse Response) jsem použil inverzní Čebyševovy filtry řádu 10 a 8 – obr. 4.9.



Obr. 4.9: Modulová charakteristika DP a HP pro basovou strunu  $E$

Na obr. 4.10 je ukázána dekompozice signálu odpovídajícího tónu  $E = 82,4Hz$  na dvě složky reprezentující dolní a horní kmitočtové pásmo ve smyslu obr. 4.8.



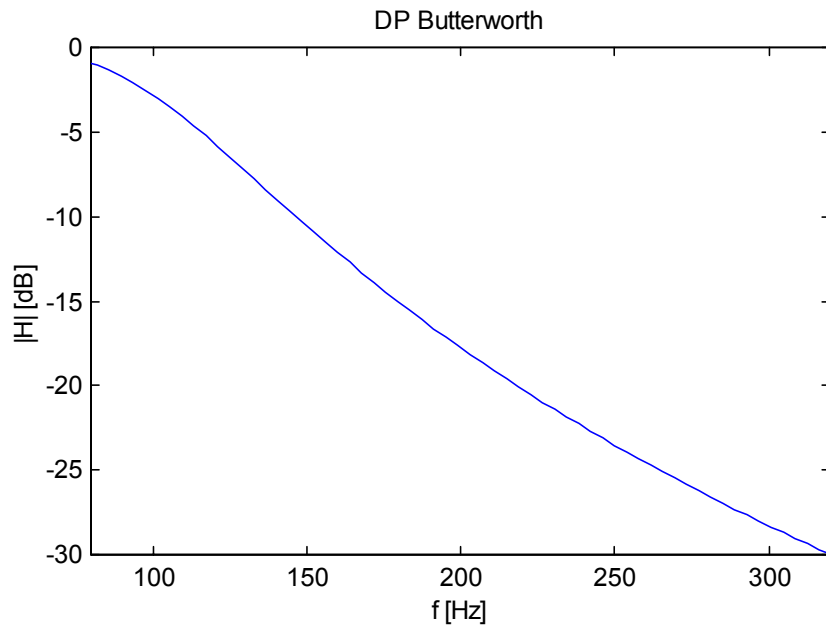
Obr. 4.10: Signál za dolní a horní propustí IIR

Energie segmentu signálu za dolní propustí  $E_{11}$  a za horní propustí  $E_{12}$  v tomto případě je

$$E_{11} = \sum_{k=1}^N s_1^2(k) = 3,95, \quad E_{12} = \sum_{k=1}^N s_2^2(k) = 47,62. \quad (4.35)$$

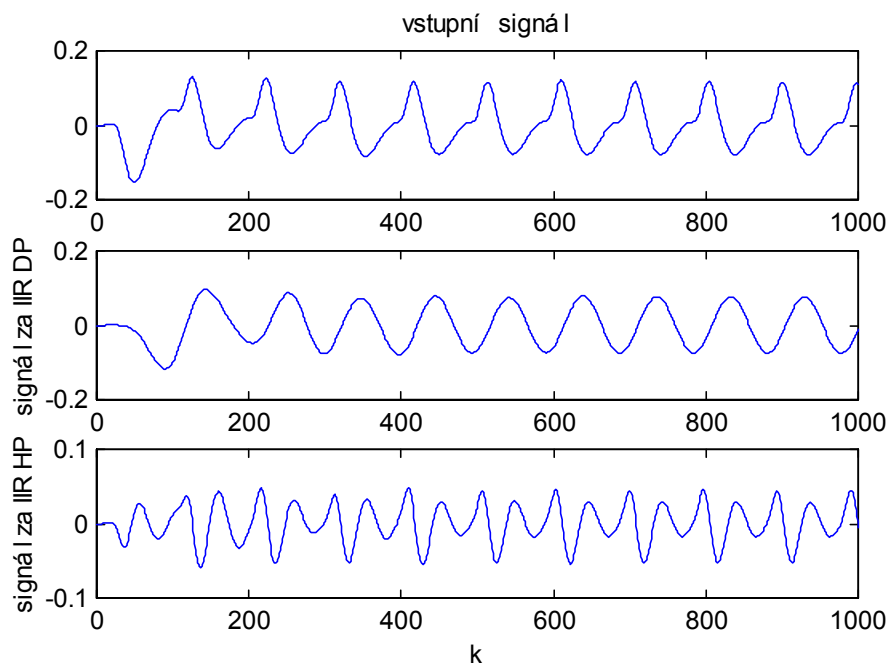
Z toho vyplývá, že v daném signálu mají více energie vyšší harmonické oproti základní harmonické, což je ve shodě s poznatky z oddílu 4.1.1.

Nyní před banku filtrů zařadíme DP pro omezení vyšších harmonických v signálu struny  $E$  s přenosem ukázaným na obr. 4.11.



Obr. 4.11: Modulová charakteristika Butterworthovi DP pro basovou strunu E

Dekompozice signálu v tomto případě je na obr. 4.12.



Obr. 4.12: Signál s omezenými vyššími harmonickými za dolní a horní propustí IIR

Pro energie dolního a horního frekvenčního pásma nyní dostáváme

$$E_{11} = \sum_{k=1}^N s_1^2(k) = 3,04, \quad E_{12} = \sum_{k=1}^N s_2^2(k) = 0,67. \quad (4.36)$$

Na základě energií pásem signálu jsme tedy správně určili, že frekvence hraného tónu je nižší nežli  $c = 130,8\text{Hz}$ .

Je tedy zřejmé, že po potlačení vyšších harmonických je v principu možno postupnou filtrací bankou filtrů odhadnout výšku hraného tónu. Tato metoda je ale zbytečně náročná oproti metodám ukázaným v dalších oddílech.

## 4.4 Autokorelace

Autokorelační funkce je definována jako [5]

$$R(\tau) = \sum_{-\infty}^{\infty} s(k + \tau) s^*(k), \quad \tau \in Z, \quad (4.37)$$

kde  $\tau$  představuje časové posunutí signálu vůči sobě samému a  $s^*(k)$  komplexně sdruženou hodnotu signálu. V našem případě budeme počítat autokorelaci pro reálný signál, kde

$$s(k) = 0, \quad k \leq 0. \quad (4.38)$$

### 4.4.1 Detekce periody

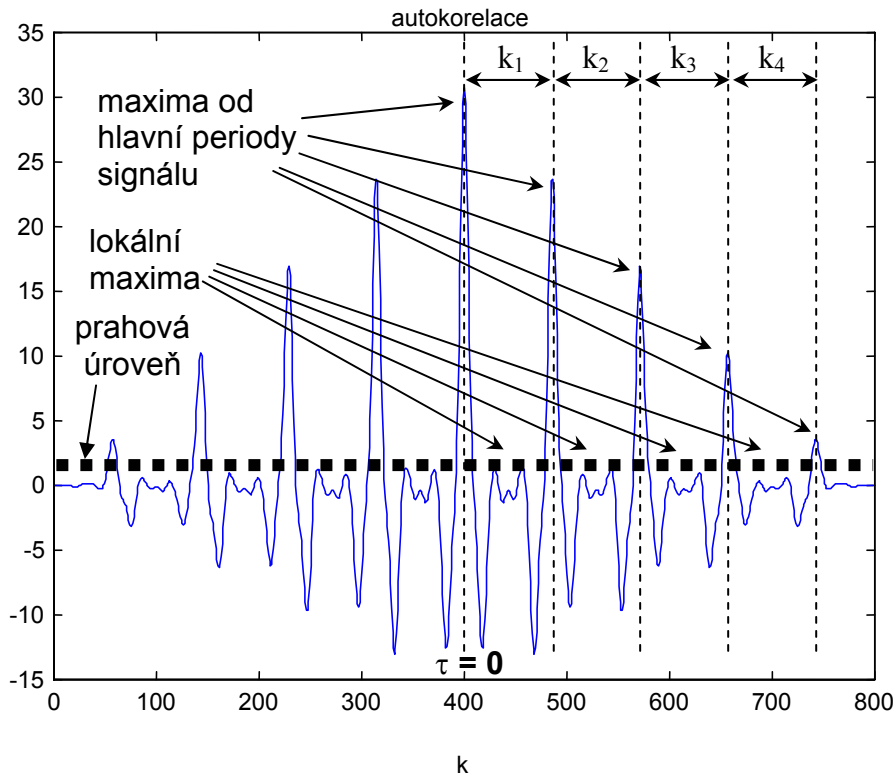
Analyzovaný signál rozdělíme na segmenty odpovídající délce periody nehlubšího detekovatelného tónu příslušné struny. Tím máme zaručeno, že v každém segmentu se bude vyskytovat minimálně 1 celá perioda hledaného tónu. Mimo segment doplníme signál nulami. Pro výpočet autokorelační funkce segmentu délky  $N$  předpokládáme:

$$s(k) = 0, \quad k < 1 \text{ nebo } k > N. \quad (4.39)$$

Autokorelační funkci pro nové meze pak můžeme přepsat následovně:

$$R(\tau) = \sum_{k=1}^N s(k + \tau) s(k). \quad (4.40)$$

Tato funkce je sudá – symetrická podle  $\tau = 0$ , kde má také maximum rovné energii signálu v segmentu. Pomocí autokorelace budeme vyšetřovat periodicitu signálu. Periodicitou v tomto kontextu rozumíme význačnou podobnost, nikoliv identitu úseků signálu. Periodicita signálu se v autokorelační funkci zobrazí v podobě lokálních maxim. Naším úkolem je vybrat ta maxima, která odpovídají periodě základního tónu, a ignorovat maxima od vyšších harmonických. Toho dosáhneme vhodným prahováním.



Obr. 4.13: Autokorelace signálu odpovídajícího tónu  $E = 82,4\text{Hz}$ ,  $f_s = 8\text{kHz}$

Na obr. 4.13 je ukázán princip měření periody signálu z autokorelace. V momentech, kdy platí pro posuv v autokorelaci (4.40)

$$\tau \rightarrow kT, \quad k \in Z, \tag{4.41}$$

kde  $T$  je délka hlavní periody signálu, v autokorelaci se objevuje význačné lokální maximum. Změřením vzdálenosti  $k_n$  sousedních význačných maxim můžeme určit odpovídající frekvenci

$$f_n = \frac{f_s}{k_n}. \tag{4.42}$$

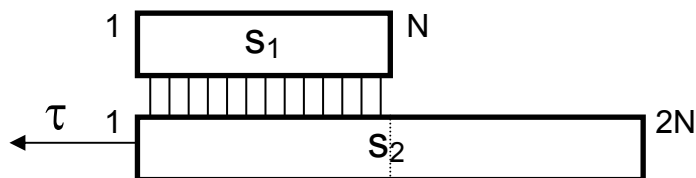


Vzhledem k sudosti autokorelační funkce stačí provádět výpočet pro  $\tau \geq 0$  (dále jednostranná autokorelační resp. korelační funkce). Jak je patrné z *obr. 4.13*, amplitudy maxim značících periodicitu signálu s  $\tau$  klesají, díky doplnění signálu mimo segment nulami. Hrozí reálné nebezpečí, že lokální maxima pro nízká  $\tau$ , odpovídající zastoupeným vyšším harmonickým v signálu, budou srovnatelně velká nebo i vyšší nežli maxima odpovídající hlavní periodě signálu (dále jen význačná maxima) pro vysoké hodnoty  $\tau$  v autokorelační funkci. V tomto případě nemůžeme použít konstantní práh v celém segmentu pro určení význačných maxim, ale musíme jeho velikost s rostoucím  $\tau$  snižovat. Další nevýhodou je fakt, že poměr výšky význačného maxima a blízkého vedlejšího poměrně rychle klesá, což dále zhoršuje možnost detekce.

Tyto nepříznivé jevy lze eliminovat následovně. Nuly, kterými jsme doplňovali daný segment nahradíme dalším segmentem signálu. Autokorelaci tedy nahradíme korelací

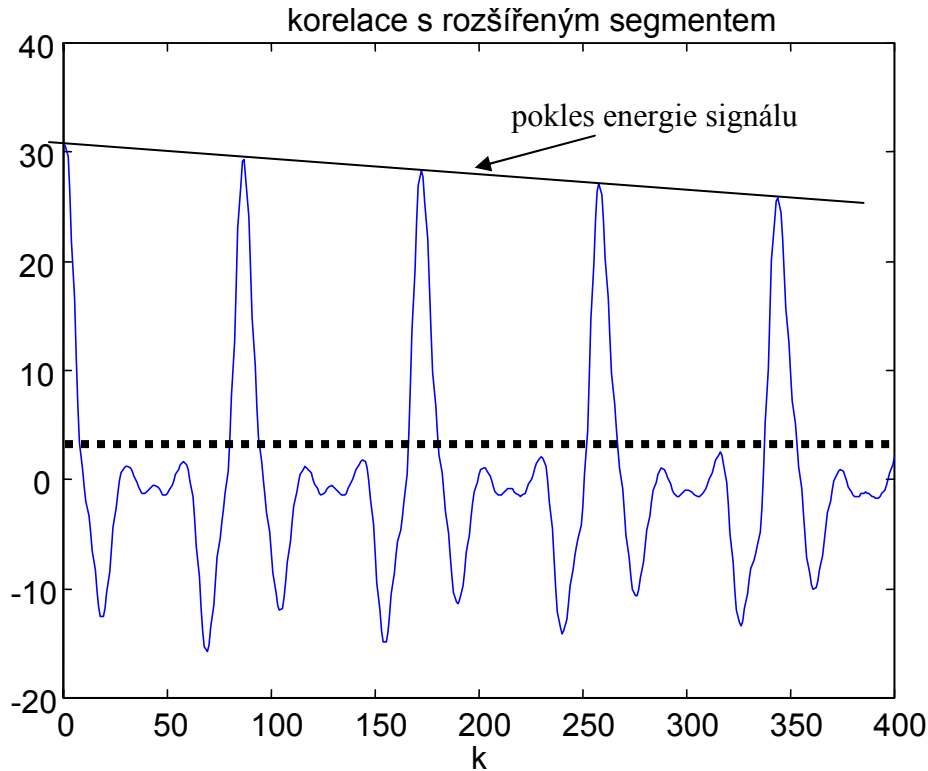
$$R_{21}(\tau) = \sum_{k=1}^N s_2(k+\tau) s_1(k), \quad (4.43)$$

kde  $s_1$  je signál tvořený původně uvažovaným segmentem a  $s_2$  signál vzniklý připojením následujícího segmentu k  $s_1$  – *obr. 4.14*. Svislé čáry zde symbolizují násobení jednotlivých vzorků segmentů.



*Obr. 4.14: Způsob výpočtu korelace*

Tímto způsobem docílíme, že pokles význačných maxim získaných korelační funkcí bude dán pouze poklesem amplitudy (přesněji energie) analyzovaného signálu. Vzhledem k malé délce segmentů vůči době dozvuku tónu je tento pokles zanedbatelný – *obr. 4.15*. Dále již neplatí, že pro  $\tau = 0$  musí nabývat korelační funkce maximální hodnoty. Nicméně hodnota  $R_{12}(\tau = 0)$  odpovídá význačnému maximu a lze z ní stanovit práh pro určování maxim následujících.



Obr. 4.15: Korelace s rozšířeným segmentem

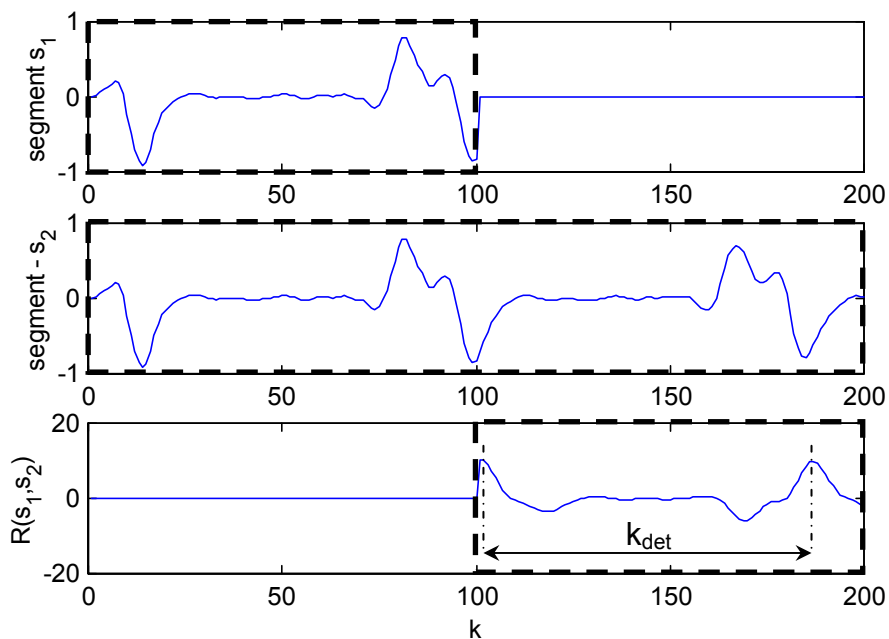
Pro výpočet „jednostranné“ korelační funkce jsem napsal v Matlabu funkci  $cor(s_2, s_1, \tau)$ , kde musí být splněno

$$length(s_2) \geq length(s_1) + \tau. \quad (4.44)$$

Funkce  $length()$  vrací délku segmentu  $s$ . Optimální délka segmentu  $s_1$  by měla odpovídat vlnové délce nejhlubšího tónu, který chceme detekovat, resp. měla by být nepatrně delší. Počet vzorků segmentu je pak

$$N \geq round\left(\frac{f_s}{f_{low}}\right), \quad (4.45)$$

kde  $f_s$  je vzorkovací frekvence,  $f_{low}$  nejnižší detekovatelná frekvence a  $round()$  je zaokrouhlení na celá čísla. Frekvence  $f_{low}$  je pro každou strunu jiná a je rovna frekvenci nejnižšího na ní hratelného tónu.



Obr. 4.16: Zpoždění detekce z korelace

Jak je patrné z obr. 4.16, v nejmíň příznivém případě je zpoždění detekce rovno délce 2 segmentů – při detekci  $f_{low}$ . Pro nejhlubší tón na basové struně  $E = 82,4\text{Hz}$  bude zpoždění detekce

$$t_{delay} \approx \frac{2}{f_{low}} = \frac{2}{82,4} \cong 0,024\text{s}. \quad (4.46)$$

Uvedený vztah je přibližný, neboť neuvažuje vliv diskretizace času. Ale pro  $f_s \geq 8\text{kHz}$  je uvedené zpoždění platné na všechny řády. Pro vyšší tóny hrané na dané struně zpoždění klesá, neboť klesá  $k_{det}$ .

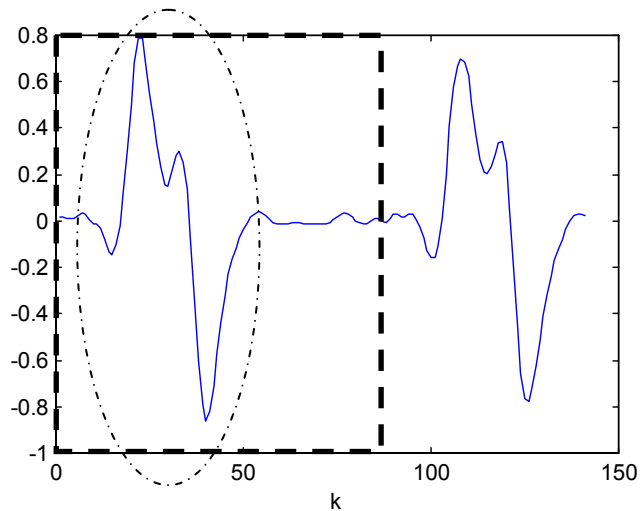
Zpoždění pro detekci tónu můžeme vyjádřit

$$t_{delay} = \frac{N + k_{det}}{f_s}, \quad (4.47)$$

kde  $N$  je délka  $s_1$ .

Metoda korelace umožňuje snížit zpoždění detekce periodicity zkrácením segmentu  $s_1$ . Tento segment totiž nemusí mít nutně délku odpovídající  $f_s / f_{low}$ . Vhodným zkrácením  $s_1$ ,

tj. takovým, že  $s_1$  stále obsahuje alespoň jeden hlavní zákmit signálu (viz obr. 4.17), můžeme korelaci s  $s_2$  stále detekovat periodicitu  $s_2$ .



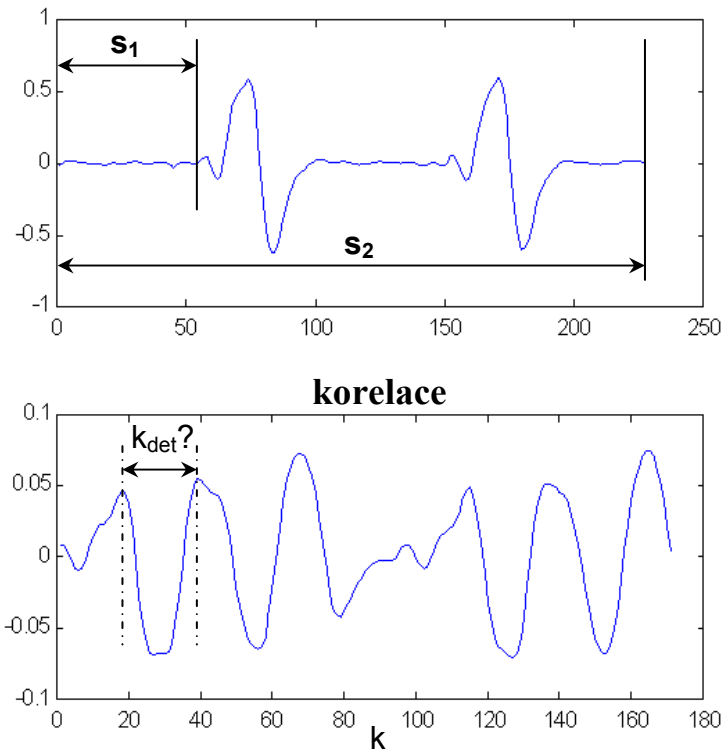
Obr. 4.17: Možnosti zkracování segmentu  $s_1$

Čárkovaně je na obr. 4.17 vyznačen původní  $s_1$ , čerchovaná elipsa vyznačuje postačující část segmentu pro korelaci s  $s_2$ .

Možnost zkracování délky segmentu jsem experimentálně ověřil v Matlabu. Pro  $E = 82,4\text{Hz}$  stačila spolehlivě délka segmentu

$$N = \frac{f_s}{130\text{Hz}} < \frac{f_s}{82,4\text{Hz}}. \quad (4.48)$$

Je třeba dbát na správný výběr zkráceného segmentu. Vybereme-li úsek s malou energií vůči zbytku segmentu, získané hodnoty budou zavádějící – obr. 4.18.



Obr. 4.18: Nevhodně zkrácený segment  $s_1$

Přesnost korelační metody závisí na volbě vzorkovací frekvence signálu. Na měřenou frekvenci  $f_i$  připadá  $N_1$  resp.  $N_2$  vzorků tak, že

$$N_1 = \text{round} \frac{f_s}{f_t}, N_2 = \text{round} \frac{f_s}{f_t} \oplus 1, f_{tdiskr1} = \frac{f_s}{N_1}, f_{tdiskr2} = \frac{f_s}{N_2} \quad (4.49)$$

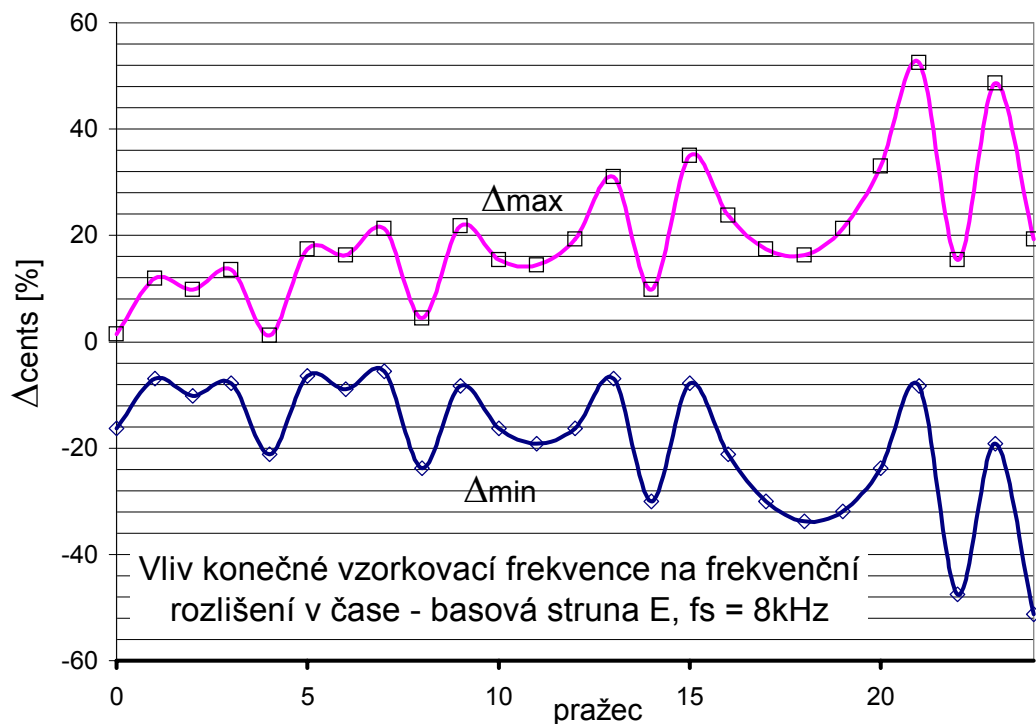
$\oplus$  zde značí +1 pokud jsme zaokrouhlovali dolů, -1 pokud nahoru.  $f_{tdiskr1}$  resp.  $f_{tdiskr2}$  je frekvence, kterou jsme schopni určit z diskrétního signálu. Vztah pro výpočet  $N_2$  můžeme přepsat

$$N_2 = \text{round} \frac{f_s}{f_t} - 2 \cdot \text{round} \left( \text{fix} \left( \frac{f_s}{f_t} \right) \right) + 1, \quad (4.50)$$

funkce  $\text{fix}()$  vrací desetinou část čísla.

Procentuální odchylku mezi skutečnou frekvencí  $f_{skut}$  a frekvencí  $f_{tdisk}$  detekovatelnou při dané  $f_s$ , respektující exponenciální rozložení frekvencí tónové řady, určíme

$$\Delta_{cents} = 1200 \cdot \log_2\left(\frac{f_{disk}}{f_{iskut}}\right) \quad [\%]. \quad (4.51)$$



Obr. 4.19: Chyba určení frekvence v důsledku diskretizace signálu

Pro hlubší tóny připadá na délku periody více vzorků, dochází zde tedy k menší chybě určení frekvence nežli u vyšších tónů. Jak je patrné z obr. 4.19, již pro basovou strunu při  $f_s = 8\text{kHz}$  nejvyšší tóny vykazují možnou odchylku vyšší než 50%, což nejen znemožňuje dosáhnout čtvrttónového rozlišení, ale i půltónového. Předpokládáme-li, že celý systém pracuje se stejnou  $f_s$ , u melodických strun již budou půltóny zcela splývat.

#### 4.4.2 Zvýšení frekvenčního rozlišení autokorelace

##### Interpolace navzorkovaného signálu

Pokud máme k dispozici dostatečnou rezervu výpočetního výkonu, můžeme navzorkovaný signál interpolovat (převzorkovat) a poté provést korelaci. Interpolaci můžeme provést způsobem uvedeným kapitole 2, nebo použít vhodnou interpolační funkci. Výhodné je interpolovat signál například polynomem třetího řádu ( v Matlabu funkce *spline()* )

$$P(x) = a_3x^3 + a_2x^2 + a_1x + a_0. \quad (4.52)$$

Převzorkováním se navyšuje množství vzorků v segmentech pro korelaci a tudíž značně narůstají výpočetní nároky.

### Interpolace korelační funkce

V tomto případě počet operací pro výpočet korelace zůstává, interpolujeme pouze peaky korelační funkce.

### Průměrování period detekovaných v segmentu

Najdeme-li v 1 segmentu  $M$  period

$$T_m = (T_1, T_2, \dots, T_M), \quad 1 \leq m \leq M, \quad (4.53)$$

průměrná délka periody pak bude

$$\bar{T} = \frac{1}{M} \sum_{m=1}^M T_m. \quad (4.54)$$

Jelikož průměr počítáme z jednoho segmentu, zpoždění je rovno délce 2 segmentů. Tímto způsobem získáme větší frekvenční rozlišení pro tóny vyšších frekvencí, čímž omezíme chybu vzniklou diskretizací času. Časové rozlišení bude pro celý rozsah struny stejné (uvažujeme konstantní délku segmentu pro jednotlivé struny).

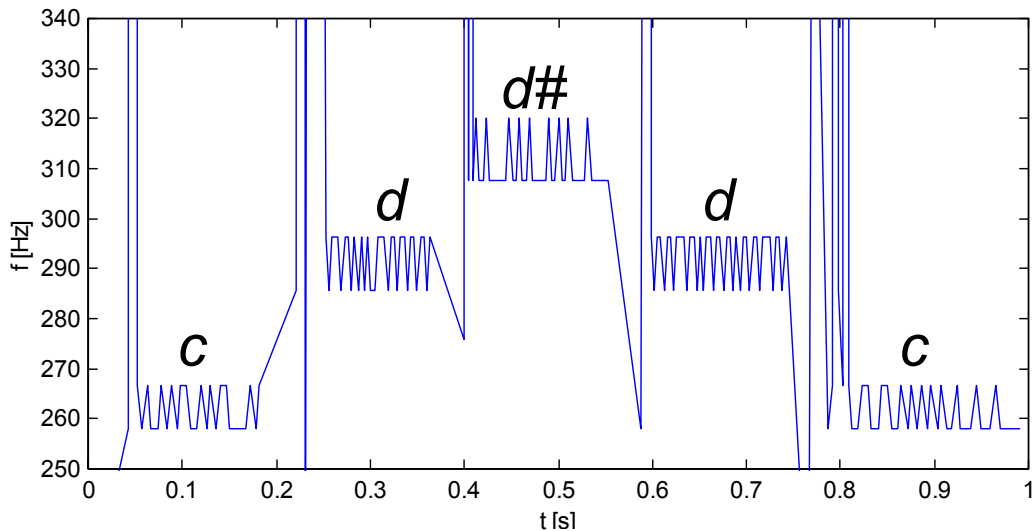
Pro vyšší struny potřebná délka segmentu rychle klesá, tudíž si můžeme dovolit průměrovat i více následujících segmentů při zachování malého zpoždění, a tím dále zvýšit frekvenční rozlišení. Průměrováním změřených hodnot můžeme chybu snížit i o několik řádů.

### Zvýšení vzorkovací frekvence

S vyšší  $f_s$  narůstá frekvenční rozlišení, pro  $f_s = 48kHz$  již zcela odpadá potřeba interpolace či průměrování.

### 4.4.3 Implementace autokorelace v Matlabu

V Matlabu jsem napsal program *autokor.m*, který rozdělí načtený signál na segmenty s volitelným překryvem, na těchto segmentech provede korelaci ve smyslu (4.43) a určí z jejích vyznačných maxim základní kmitočet tónu. Na *obr. 4.20* jsou frekvence detekované programem *autokor.m* v signálu obsahujícím krátkou melodii.



*Obr. 4.20: Frekvence detekované v melodii,  $f_s = 8\text{kHz}$*

V grafu je vidět, že frekvence příslušející danému tónu nabývají skutečně dvou hodnot, jak jsme uvažovali v (4.49). Frekvenční výstřelky mezi detekovanými tóny jsou způsobeny attackem při tvorbě tónu, v tomto případě má na nich velký podíl drhnutí trsátka o strunu. Těchto nežádoucích hodnot se zbavíme volbou vhodných kritérií, vycházejících z frekvenčního rozsahu struny a kontinuity detekovaných frekvencí.

Korelační metoda se ukázala jako velmi robustní a rychlá, tedy vhodná pro použití v reálném čase. Její jedinou nevýhodou je velké množství násobení proměnné s proměnnou připadající na jeden vzorek signálu, což vede především s vyšší  $f_s$  k velkým výpočetním nárokům.

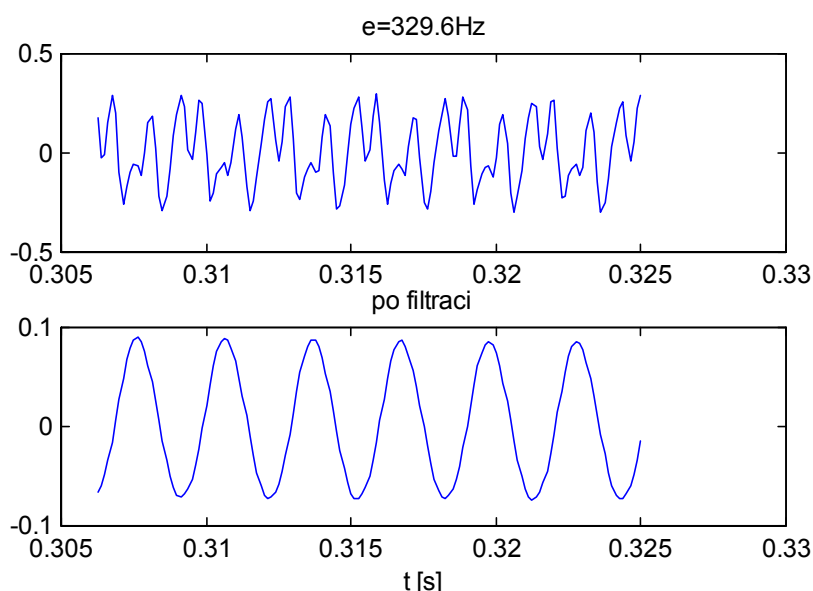
### 4.5 Přímé měření délky periody

Pro měření délky periody harmonického signálu se často používá tzv. Zero-crossing, tj. detekce průchodu signálu nulou. Z naměřených vzdáleností mezi průchody nulou  $T$  přímo plyne frekvence harmonického signálu



$$f = \frac{1}{2T}. \quad (4.55)$$

Pokud analyzujeme neharmonický periodický signál, který má v rámci periody 1 výrazné maximum (peak), pak můžeme určit periodu signálu ze vzdálenosti těchto peaků. Tomuto případu se povětšinou blíží signál z kytary. Zejména u vyšších strun, které jsou charakteristické výraznějším podílem vyšších harmonických, však vedlejší maxima dosahují prakticky stejné amplitudy jako hlavní maxima, což znemožňuje přímé měření periody. Řešením je potlačení vyšších harmonických filtrací – *obr. 4.21*.



*Obr. 4.21: Potlačení vyšších harmonických v signálu filtrací*

V tomto případě byl signál filtrován inverzní Čebyševovou DP čtvrtého řádu. Po filtraci je možné již snadno odečíst periodu a určit frekvenci tónu.

Jelikož je frekvenční rozsah jedné struny dvě oktávy, musíme filtraci provádět tak, abychom i pro nehlubší tóny dostatečně eliminovali podíl vyšších harmonických, ale zároveň abychom zcela neodfiltrovali případné hrané tóny ve vyšších polohách. Zároveň požadujeme od filtru minimální dobu odezvy na změny amplitudy signálu a minimální zpoždění signálu. Tyto vlastnosti již nevykazuje výše zmíněná Čebyševova propust, neboť požadavky na minimální přechodové zkreslení signálu při attacku a rovnoměrné potlačení vyšších harmonických v celém rozsahu struny nespĺňuje.

Pro každou strunu je výhodné navrhnout zvláštní filtr vycházející z jejího frekvenčního rozsahu a spektrálních charakteristik. Například pro basovou strunu  $E$  je vyhovující Butterworthova DP 3. řádu s přenosovou charakteristikou ukázanou na *obr. 4.11*.

Frekvenční rozlišení přímé detekce je shodné s autokorelační metodou (4.51) a můžeme jej zvyšovat obdobnými způsoby jako v 4.4.2.

Jelikož má metoda přímého měření periody minimální nároky na výpočet a vykazuje dobrou spolehlivost, rozhodl jsem se jí použít při implementaci frekvenčního detektoru. Z tohoto důvodu uvedu specifika návrhu dolních propustí jednotlivých strun a algoritmu detekce frekvence v kapitole 5 věnované implementaci syntetické kytary.

## 5 Implementace syntetické kytary

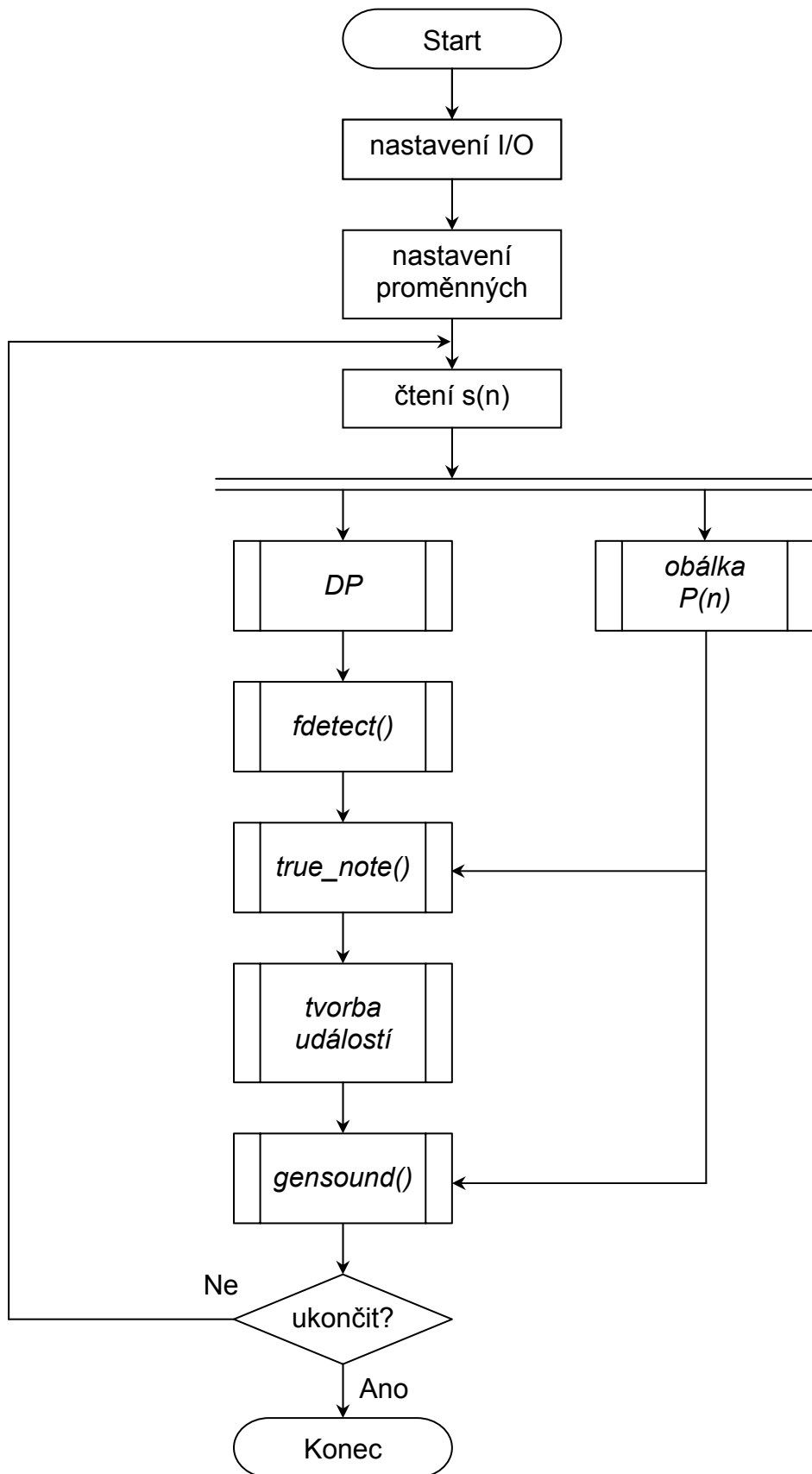
V kapitole 4 jsme porovnali vlastnosti vybraných metod analýzy výšky tónu v kytarovém signálu. Použitelnými se ukázaly autokorelace a přímá detekce periody. Pro implementaci v syntetické kytare jsme zvolili přímou detekci, jelikož má podstatně menší výpočetní nároky oproti autokorelační metodě a vykazuje srovnatelné časové a frekvenční rozlišení s autokorelací.

V této kapitole se budeme zabývat návrhem analyzující a syntetizující části syntetické kytary, tedy návrhem MIDI převodníku a jednoduchého syntezátoru. Všechny algoritmy jsem nejprve implementoval v prostředí Matlabu, po jejich odladění poté přepsal do jazyka C a kompiloval v linuxovém kompilátoru GCC.

V následujícím textu bude popsána struktura použitých algoritmů a v některých případech pro přehlednost i určitá specifika programové implementace v Matlabu. Postup přepisu do jazyka C již nebudu zmiňovat, jednalo se pouze o změny syntaxe, deklarací proměnných a způsobu práce se vstupními a výstupními zařízeními. Zdrojové soubory pro Matlab i C jsou k dispozici k nahlédnutí na přiloženém CD.

Na *obr. 5.1* je ukázáno blokové schéma řetězce syntetické kytary pro signál jedné struny. Nastavení I/O (Inputs And Outputs – vstupy a výstupy) představuje inicializaci zvukových karet pro čtení a zápis. Vzhledem k poznatkům o vlivu  $f_s$  na frekvenční rozlišení analýzy hudebního signálu jsem zvolil  $f_s = 48kHz$ . Bitová hloubka připadající na jednotlivé signály od strun je  $16b$ , což je vzhledem k dostatečnému vybuzení signálu z MIDI snímače (viz příloha) zcela vyhovující.

Nastavením proměnných definujeme počáteční podmínky pro běh systému. Následně program vstoupí do smyčky, v níž setrvá až do žádosti o ukončení procesu. Na začátku každého průchodu smyčkou dochází k načtení aktuálního vzorku signálu  $s(n)$ . Jednotlivé bloky smyčky pak na základě tohoto vzorku aktualizují své výstupy. Systém tak pracuje v reálném čase – jeho výstupy okamžitě reagují na změny vstupního signálu. Jednotlivé bloky algoritmu detailně popíšeme v následujících oddílech.



Obr. 5.1: Řetězec syntetické kytary pro signál jedné struny

## 5.1 DP - dolní propust

V kapitole 4 jsme konstatovali, že základní harmonická ve spektru kytarového tónu často nemá v poměru k vyšším harmonickým největší energii. Tento nepříznivý fakt můžeme odstranit použitím vhodné dolní propusti (DP).

### 5.1.1 Výběr aproximace

Základním požadavkem na dolní propust pro každou strunu je minimální zkreslení signálu v důsledku přechodových dějů filtru. Doba odeznění přechodového děje filtru je dána především volbou aproximace jeho přenosové charakteristiky a z toho plynoucího řádu filtru. Filtry FIR (Finite Impulse Response) mají obecně konečnou impulsní odezvu, tedy i dobu odeznění přechodového děje, ale vyžadují obvykle použití vyššího řádu aproximace nežli filtry IIR (Infinite Impulse Response), u nichž je teoreticky impulsní odezva nekonečná. U číslicové realizace IIR je skutečná délka odezvy dána kvantováním úrovní signálu – klesne-li tato úroveň pod dolní práh číslicového rozlišení, odezva končí. Filtry IIR vyžadují obvykle nižší řád aproximace oproti FIR, při vyšších řádech však mohou být nestabilní. Jelikož zpoždění signálu filtrací je přímo úměrné řádu filtru (počtu paměťových buněk jimiž signál prochází), zvolil jsem pro realizaci dolních propustí strukturu IIR.

Dalším požadavkem na filtr je minimální fázové zkreslení, tj. vzájemný posuv harmonických složek signálu vedoucí v důsledku ke zkreslení poloh maxim v průběhu, především při přechodovém ději. Výhodnou se ukazuje v tomto ohledu mnou zvolená Butterworthova maximálně plochá aproximace.

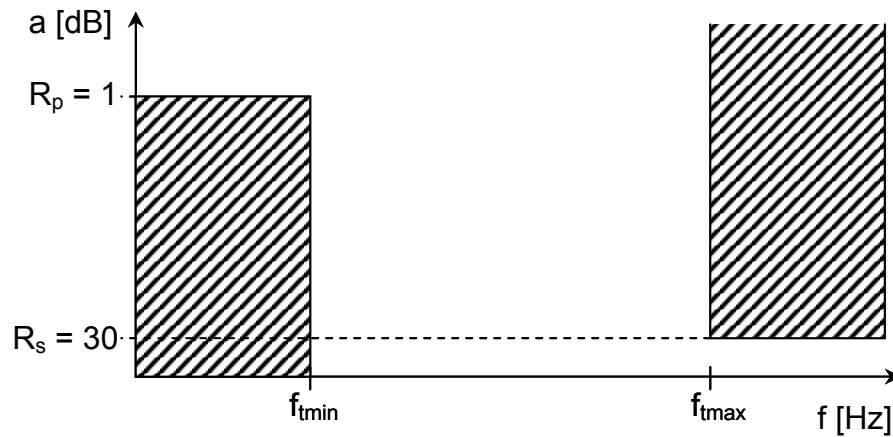
Návrh Butterworthovy DP patří mezi návrhy konvenční a v literatuře (např. [6]) dobře popsané, proto zde uvedeme pouze výchozí zadání tolerančního schématu. Pro vlastní výpočet koeficientů DP jednotlivých strun jsem použil v Matlabu funkci Signal Processing Toolboxu *buttord()*, která vypočte řád filtru a mezní kmitočet potřebný pro aproximaci zadaných mezí propustného a nepropustného pásma a útlumů jim příslušných, a funkci *butter()*, která pro získaný řád a mezní kmitočet spočte koeficienty filtru.

### 5.1.2 Toleranční schéma DP

Jak již bylo v kapitole 4 zmíněno, u elektrické kytary dosahují amplitudy vyšších harmonických maximálně 3-násobku amplitudy základní harmonické. Pro dostatečné

potlačení zákmitů od vyšších harmonických v signálu se ukázal při experimentech vhodný spád  $30 \text{ dB}/2 \text{ oktávy}$ . Spád je uveden na 2 oktávy kvůli rozsahu struny.

Toleranční schéma má pak následující podobu – obr. 5.2.

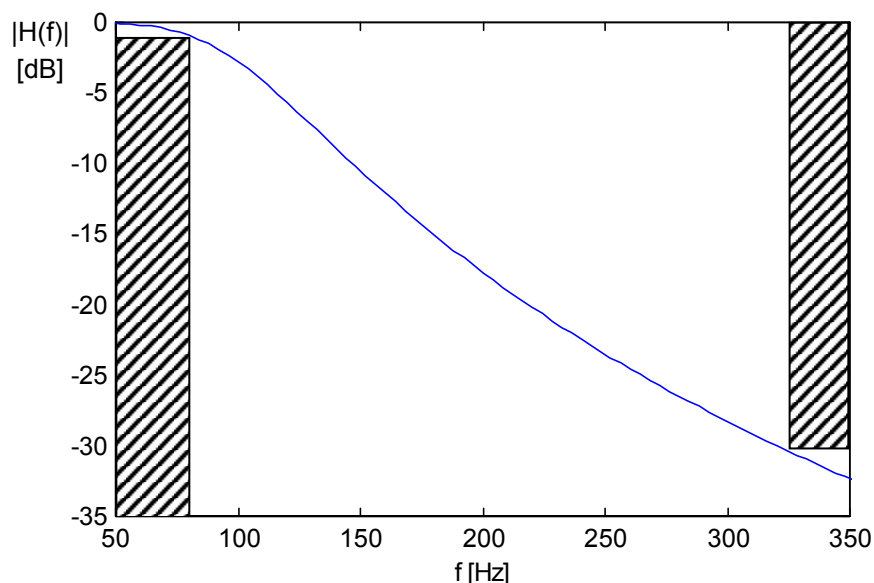


Obr. 5.2: Toleranční schéma DP pro potlačení vyšších harmonických

Frekvence  $f_{\min}$  a  $f_{\max}$  odpovídají nejnižšímu a nejvyššímu tónu hratelnému na dané struně,  $R_p$  útlum v propustném pásmu a  $R_s$  útlum v nepropustném pásmu.

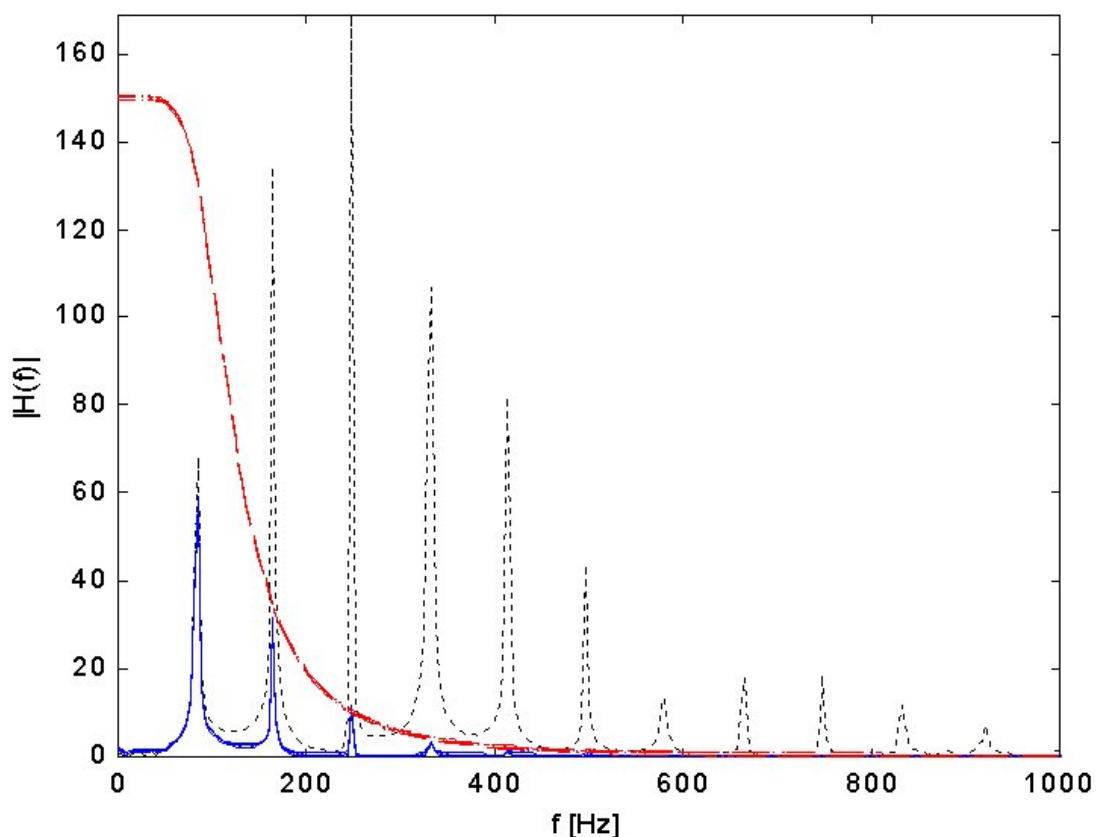
Pro basovou strunu  $E$  bude  $f_{\min} = 82,4 \approx 80 \text{ Hz}$  a  $f_{\max} = 329,6 \approx 330 \text{ Hz}$ . Velikost  $f_{\max}$  je volena vzhledem k faktu, že u tónů blížících se horní hranici rozsahu struny zcela převládá hlavní harmonická.

DP basového  $E$  má pro uvedené hodnoty tolerančního schématu následující přenos – obr. 5.3.



Obr. 5.3: Přenos DP pro basovou strunu  $E$

Na obr. 5.4 je ukázáno, jak se projeví filtrace signálu odpovídajícího tónu  $E = 82,4\text{Hz}$  dolní propustí z obr. 5.3 na jeho amplitudovém spektru.



Obr. 5.4: Vliv filtrace na amplitudové spektrum tónu  $E = 82,4\text{Hz}$

Původní amplitudové spektrum je značeno tečkovaně, spektrum po filtraci plnou čarou. Přenos dolní propustí je vynásoben 150ti, aby byl v daném měřítku vidět, a zobrazen čárkovaně. Jak je vidět, po filtraci má největší energii hlavní harmonická, což je naším cílem.

## 5.2 Obálka $P(n)$

Při analýze hudebního signálu obvykle pracujeme s amplitudovou nebo výkonovou obálkou. V našem případě budeme určovat obálku výkonovou, neboť má přímou spojitost s vnímanou hlasitostí zvukového signálu a můžeme ji tak použít pro prahování vstupního signálu a tvorbu MIDI událostí určujících hlasitost generovaného syntetického zvuku. Mezi nejpoužívanější postupy odhadu výkonové obálky patří blokový odhad a exponenciální odhad [4].

### 5.2.1 Blokový odhad výkonu

Blokový odhad výkonu  $l$ -tého segmentu délky  $N$  signálu  $s(n)$  určíme

$$P(l) = \frac{1}{N} \sum_{n=0}^{N-1} s^2(n). \quad (5.1)$$

Reakce na rychlost změn okamžité energie signálu je dána zvolenou délkou segmentu a mírou překryvů těchto segmentů. Pokud posouváme segment v signálu vždy o jeden vzorek, jedná se o tzv. klouzavé průměrování. Blokový odhad pak můžeme realizovat FIR filtrem [6] s předpisem

$$P(n) = \sum_{k=0}^{N-1} b_k s^2(n-k), \quad (5.2)$$

jehož všech  $N$  koeficientů má hodnotu

$$b_k = \frac{1}{N}, \quad k = 0, 1, \dots, N-1. \quad (5.3)$$

Na vstup FIRu v tomto případě přivádíme okamžitou hodnotu energie signálu, tedy  $s^2(n)$ .

Při velké délce segmentu  $N$  se energetické změny v signálu projevují pozvolnými změnami odhadu výkonu  $P(n)$ , naopak v extrémním případě  $N = 1$  je  $P(n)$  numericky roven okamžité energii signálu.

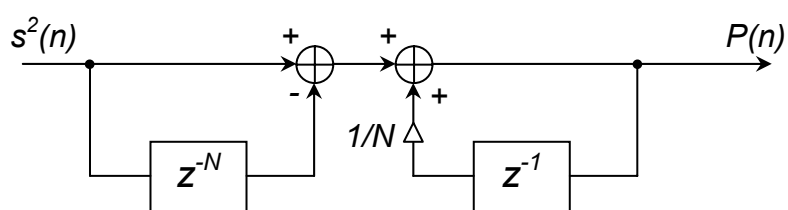
Jak bude později ukázáno, znalost odhadu  $P(n)$  nám umožní zavedení výkonového pravdivostního kritéria pro detekci význačných maxim (peaků), platných frekvencí a nástupu a zániku tónu. Proto potřebujeme, aby odhad výkonu na jednu stranu příliš nekopíroval průběh okamžité energie v rámci periody, na stranu druhou aby zachycoval s dostatečnou rychlostí (strmostí) změny energie v přechodech mezi pauzami a hrou. Pro splnění těchto požadavků se ukázala vhodná volba

$$N = 2 \cdot \frac{f_s}{f_{low}}, \quad (5.4)$$



kde  $f_{low}$  je frekvence nejnižšího tónu hratelného na struně.

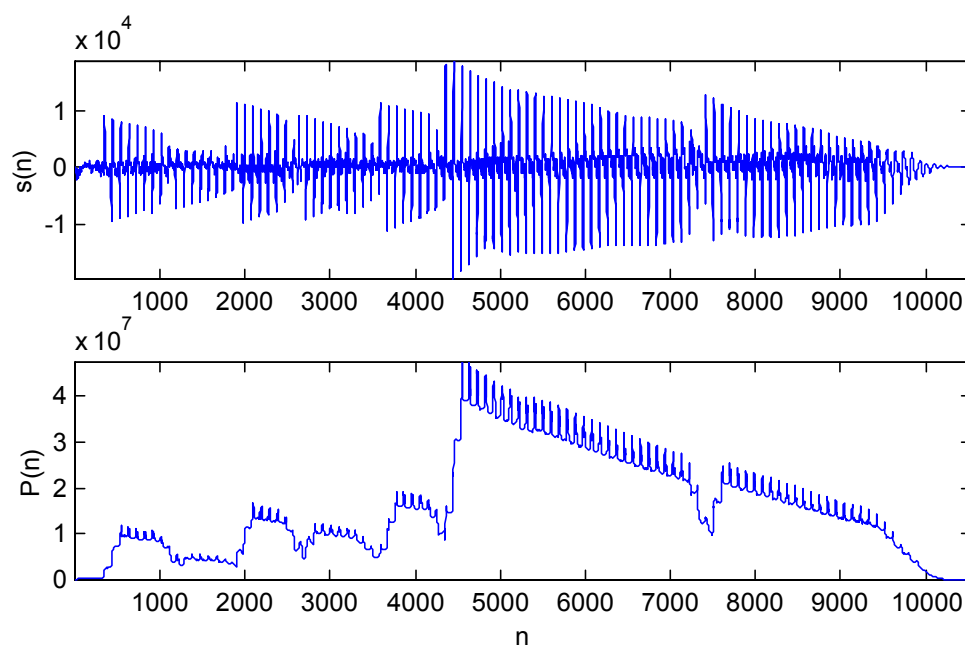
Pro výpočet je možno použít rekurentní realizaci klouzavého průměru [4] – obr. 5.5.



Obr. 5.5: Rekurentní realizace klouzavého průměru

Levá část představuje přidání nového a odstranění nejstaršího vzorku okamžité energie ze segmentu. Pravá část integruje příchozí hodnoty a počítá tak průměr přes celý segment.

Na obr. 5.6 je ukázka blokového odhadu výkonu pro signál odpovídající melodii hrané na basové struně  $E$ .



Obr. 5.6: Blokový odhad výkonu – melodie na basové struně  $E$ ,  $f_s = 8\text{kHz}$ ,  $N = 200$

V průběhu blokového odhadu jsou patrná lokální maxima odpovídající okamžité hodnotě energie v rámci jednotlivých period signálu, ale jejich amplitudy jsou oproti hlavnímu trendu výkonu malé.

## 5.2.2 Průběžný odhad výkonu

Průběžný odhad výkonu signálu můžeme definovat [4]

$$P(n) = \alpha \cdot P(n-1) + (1-\alpha) s^2(n), \quad (5.5)$$

kde  $\alpha$  představuje koeficient zapomínání a určuje míru vyhlazení odhadu. Výhodou je, že oproti blokovému odhadu, kde bylo třeba uchovávat v paměti  $N-1$  hodnot, zde uchováváme hodnotu pouze jednu. Průběžný odhad můžeme realizovat IIR integrátorem prvního řádu, pro který platí

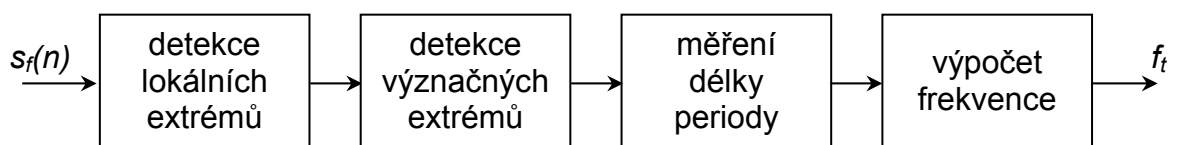
$$H(z) = \frac{1-\alpha}{1-\alpha z^{-1}}. \quad (5.6)$$

Oproti blokovému odhadu zde dochází k exponenciálnímu odeznívání vlivu předchozích hodnot, což vede k pozvolnějším reakcím na prudký pokles výkonu signálu, jak je ukázáno v [4].

Z tohoto důvodu jsem zvolil pro výpočet výkonové obálky v syntetické kytáře klouzavé průměrování.

## 5.3 Funkce *fdetect()*

Funkce *fdetect* slouží k odhadu výšky hraného tónu, tedy odhadu aktuální základní frekvence signálu. Pro odhad je třeba v signálu filtrovaném DP najít všechny lokální extrémů a z nich vybrat význačná maxima a minima. Z jejich poloh pak určíme délku základní periody signálu a odpovídající frekvenci – obr. 5.7. Signál za DP je zde značen  $s_f(n)$  a určená frekvence  $f_t$ .



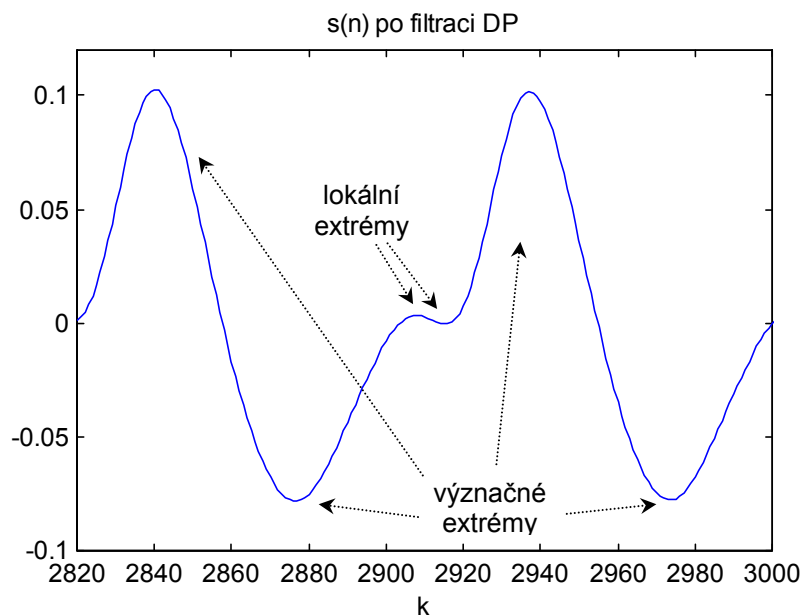
Obr. 5.7: Funkce *fdetect()*

### 5.3.1 Detekce lokálních extrémů

V oddílu 4.1.3 jsme popsali algoritmus nalezení lokálních maxim v signálu. Lokální minima určíme z lokálních maxim v signálu  $-s(n)$ .

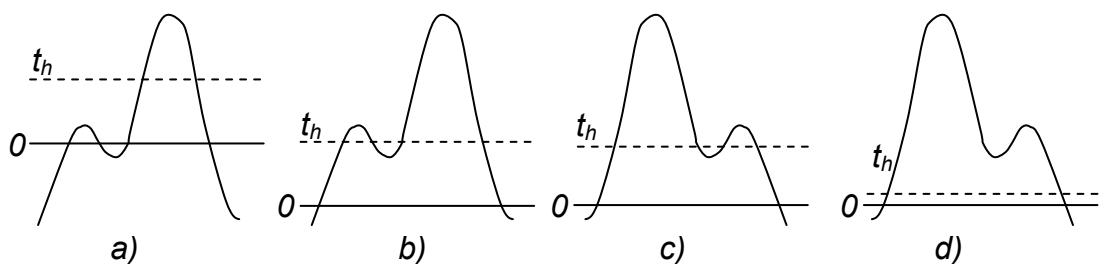
### 5.3.2 Detekce význačných extrémů

Ve spektru signálu po filtraci dolní propustí je dominantní základní harmonická. Vyšší harmonické jsou potlačeny takovým způsobem, že v naprosté většině případů nacházíme v základní periodě signálu pouze jeden zákmít odpovídající dvěma lokálním extrémům, *obr. 5.8*. Dále vždy platí, že mezi hlavními maximy dochází vždy k průchodu signálu nulou, signál má nulovou či zanedbatelnou stejnosměrnou složku.



*Obr. 5.8: Signál odpovídající tónu  $E = 82,4$  po filtraci dolní propustí*

Díky této vlastnosti filtrovaného signálu můžeme určit význačné extrémů na základě následujících kritérií.



*Obr. 5.9: Určení význačného maxima*

Protože význačná minima v  $s(n)$  určíme jako význačná maxima signálu  $-s(n)$ , budeme dále popisovat jen případ hledání význačného maxima.

Víme, že v každé hlavní periodě signálu za DP se nachází maximálně jedno vedlejší maximum. Potom vždy jedno ze dvou po sobě jdoucích maxim je význačné (*main peak*). Z toho důvodu jsou vstupem našeho algoritmu vždy dvě po sobě jdoucí nalezená maxima – označme je  $P_1$  a  $P_2$ . Maxima se zápornou hodnotou nebudeme testovat, nemohou být vzhledem minimální stejnosměrné složce obsažené v signálu *main peaky*.

Informaci, zda od posledního maxima došlo k průchodu signálu nulou (zero cross), označme  $P_{1\_zc}$  resp.  $P_{2\_zc}$ , a bude nabývat hodnot 0/1 – *nebyl/byl* průchod nulou. Pokud platí

$$P_{1\_zc} = 0, \quad (5.7)$$

pak  $P_1$  není a  $P_2$  je *main peak*. Další varianty předpokládají, že

$$P_{1\_zc} = 1. \quad (5.8)$$

Pro porovnávání  $P_1$  a  $P_2$  zavedeme práh  $th$  – obr. 5.9. Pokud

$$P_1 < P_2 \cdot th \quad \text{nebo} \quad (P_1 < P_2 \text{ a } P_{2\_zc} = 0), \quad (5.9)$$

$P_2$  je *main peak* – obr. 5.9 a), b). Pro zbylé případy je *main peak*  $P_1$  – obr. 5.9 c), d).

Pokud jsme prohlásili za *main peak*  $P_1$ , pro hledání dalšího *main peaku* přiřadíme

$$P_2 \rightarrow P_1, \quad P_{new} \rightarrow P_2, \quad (5.10)$$

kde  $P_{new}$  představuje nově nalezené maximum.

Pokud bylo *main peakem*  $P_2$ , pro hledání dalšího význačného maxima čekáme na nalezení dvou nových lokálních maxim a poté přiřadíme

$$P_{new1} \rightarrow P_1, \quad P_{new2} \rightarrow P_2, \quad (5.11)$$

kde  $P_{new1}$  a  $P_{new2}$  jsou nově nalezená maxima. Proces se pak opakuje od testu (5.7). Prahový koeficient  $t_h$  jsem z dobrými výsledky zvolil

$$t_h = 0,5. \quad (5.12)$$

### 5.3.3 Měření délky periody a výpočet frekvence

Jakmile získáme informaci o pozici dvou po sobě jdoucích význačných maxim či minim, délku periody, kterou vytyčují, určíme

$$L = P_2\_ind - P_1\_ind, \quad (5.13)$$

kde  $P_1\_ind$  resp.  $P_2\_ind$  značí diskrétní časovou pozici význačného extrému  $P_1$  resp.  $P_2$ . Odpovídající frekvence je pak

$$f_t = \frac{f_s}{L}, \quad (5.14)$$

kde  $f_s$  je vzorkovací frekvence analyzovaného signálu.

## 5.4 Funkce *true\_note()*

Při odhadu základní frekvence tónu dochází k častým chybám. Jsou způsobeny především přechodovými ději při tvorbě tónu – attacku a nežádoucí analýzou netonálních složek hudebního signálu – drhnutí o struny a podobně. Proto je třeba podrobit hodnoty vracené funkcí *fdetect()* kritériím, která eliminují chybné odhady.

### 5.4.1 Energetické kritérium

Klesne-li amplituda hudebního signálu pod určitou mez, začnou z hlediska úspěšnosti detekce převládat rušivé složky – rušivé signály naindukované do přívodního kabelu vedoucího od MIDI snímače k A/D převodníkům zvukových karet a výrazný pokles bitového rozlišení signálů ze snímače.

Proto je vhodné zavést energetické prahování. Odhad výkonu signálu  $P(n)$  získaný klouzavým průměrováním porovnááme s nastavenou prahovou hodnotou  $P\_th$ . Pokud je

$$P(n) < P\_th, \quad (5.15)$$

příchozí frekvence z detektoru ignorujeme. Při testování systému se osvědčila normovaná prahová hodnota

$$P_{th\_norm} = 10^{-3}. \quad (5.16)$$

Tato hodnota je vztažena k signálu o jednotkové amplitudě.

#### 5.4.2 Frekvenční kritérium

Každá struna má určitý tónový rozsah, můžeme tedy ignorovat frekvence z detektoru, které v tomto rozsahu neleží. Spodní hranici je vhodné volit o něco nižší než je výška tónu prázdné struny. Jejím zavedením eliminujeme nízké frekvence, k jejichž chybné detekci došlo vlivem nenalezení jednoho či více význačných maxim. Tuto mez jsem s dobrými výsledky zvolil

$$f_{low} = \frac{f_{tlow}}{\sqrt[24]{2}}, \quad (5.17)$$

kde  $f_{tlow}$  je frekvence tónu prázdné struny. Dolní mez  $f_{low}$  zde odpovídá výšce o čtvrttón nižší než  $f_{tlow}$ .

Horní kritériální mez rozsahu jsem nezavedl, neboť k chybné detekci dvojnásobné frekvence dochází minimálně a odstraní ji Majority Buffer popsany v dalším oddílu.

#### 5.4.3 Majoritní kritérium

Předchozí kritéria neodstraní všechny chybně detekované frekvence. Například při attacku dochází vlivem přechodových dějů filtru k posuvu maxim, který má za následek chybný odhad frekvence pro prvních několik period signálu.

Z tohoto důvodu jsem zavedl takzvaný *Majority Buffer*, který pracuje s detekovanými frekvencemi zaokrouhlenými k nejbližší hodnotě půltónové řady. Toto zaokrouhlení je vhodné, neboť detekovaná frekvence tónu kolísá v určitém rozsahu z důvodu frekvenčního rozlišení daného  $f_s$  a také kvůli změnám výšky tónu způsobeným nepatrným vytahováním a změnami síly stisku struny.

Do *Majority Bufferu* délky  $M$  se ukládají zaokrouhlené detekované frekvence. S každou novou příchozí frekvencí dojde k vymazání nejstarší hodnoty uložené v bufferu. Pokud se

v Majority Bufferu nachází  $p$  stejných hodnot takových, že

$$p > \text{floor}\left(\frac{M}{2}\right), \quad M > 1, \quad (5.18)$$

kde  $\text{floor}()$  představuje zaokrouhlení desetinné části dolů, tyto frekvence splňují tzv. majoritní kritérium. Majoritní kritérium podle (5.18) značí, že v bufferu se nachází nadpoloviční počet stejných hodnot. Přísnost majoritního kritéria narůstá s délkou *Majority Bufferu*.

Toto kritérium umožňuje zredukovat počet chybně detekovaných na naprosté minimum. Experimentálně se mi osvědčila délka  $M = 5$ , musíme tedy nalézt 3 frekvence příslušející stejnému půltónu, abychom mohli potvrdit pozitivní detekci.

## 5.5 Tvorba událostí

Jak již bylo řečeno, MIDI formát používá pro popis hudební produkce události. Do syntetické kytary jsem implementoval MIDI události *Note On* a *Note Off*, které se při hře ukládají do MIDI souboru. *Note On* má za hlavní parametry jméno tónu a hlasitost (*velocity*) a značí nástup tónu. Hlavním parametrem *Note Off* je jméno noty, která má být ukončena. Tyto dvě události tedy stačí pro popis hraných tónů a jejich nástupní hlasitosti. Obálku tónu pak již generuje samotný syntezátor podle vlastností vybraného zvuku nástroje. Dále jsem zavedl interní události *Note Hold* a *Fade Out*. *Note Hold* je aktivní, pokud držíme stále týž tón, *Fade Out* se aktivuje při ukončení tónu.

Nyní si popíšeme, za jakých okolností se pro jednotlivou strunu tyto události generují.

### Note On

*Note On* generejeme v případě, že jsme na dané struně nehráli a nyní došlo k pozitivní detekci tónu funkcí *true\_note()*, nebo jsme hráli jiný tón než je nyní pozitivně detekován. Funkcí *note()* popsané v oddílu 4.1.3 určíme pro danou frekvenci odpovídající jméno noty. Zároveň odečteme aktuální odhad výkonu  $P(n)$  a provedeme výpočet parametru *velocity*. Empiricky jsem dospěl ke vztahu

$$\text{velocity} = \sqrt[4]{P(n)}, \quad (5.19)$$

který vhodně převádí dynamiku magnetodynamicky snímané struny na dynamiku MIDI syntezátoru.

Zároveň nastavujeme interní událost *Note Hold*, která zůstává aktivní až do *Note Off*. *Note Hold* má význam pro vnitřní běh programu, kde potřebujeme mít uloženou informaci o tom, že tón je stále ještě hrán. S událostí *Note Hold* pracuje i jednoduchý syntezátor, který popíšeme později.

## Note Off

*Note Off* generujeme v případě, že trvání daného tónu již není nadále potvrzeno. Stane se tak v případě poklesu hlasitosti tónu pod zvolený práh odpovídající *P\_th* nebo při přechodu na jiný tón. Při *Note Off* se také zruší událost *Note Hold* a aktivuje se *Fade Out*. *Fade Out* je událost generovaná pro jednoduchý syntezátor a bude popsána později.

## 5.6 Funkce *gensound()* - jednoduchý syntezátor

Pro syntetickou kytaru jsem napsal jednoduchý varhanní syntezátor pracující v reálném čase. Zvuk kostelních varhan je tvořen součtem zvuků jednotlivých píšťal. Jelikož průběh zvuku píšťaly se velmi blíží sinusovce, můžeme varhany snadno napodobit součtem váhovaných vyšších harmonických. Pro zjednodušení můžeme uvažovat, že poměr spektrálního zastoupení vyšších harmonických vůči základní harmonické je v tónovém rozsahu kytary konstantní.

Pro vytvoření modelu varhan jsem provedl spektrální analýzu varhanního tónu  $a = 440\text{Hz}$ , z amplitudového spektra odečetl poměr zastoupení vyšších harmonických vůči základní harmonické a z fázového spektra jejich posuvy a sestavil následující předpis:

$$\text{synth\_out}(\alpha) = \frac{\text{velocity}}{\text{norm}} \sum_{k=1}^5 A_k \sin(2^{k-1}\alpha + \alpha_k) \quad (5.20)$$

kde výstupní zvukový signál syntezátoru *synth\_out* je funkcí fáze  $\alpha$ ,  $A_k$  jsou amplitudy jednotlivých harmonických,  $\alpha_k$  jejich fázové posuvy, *norm* je konstanta pro normování signálu na jednotkovou amplitudu a *velocity* určuje požadovanou hlasitost výstupu. Pro  $\alpha$  platí:

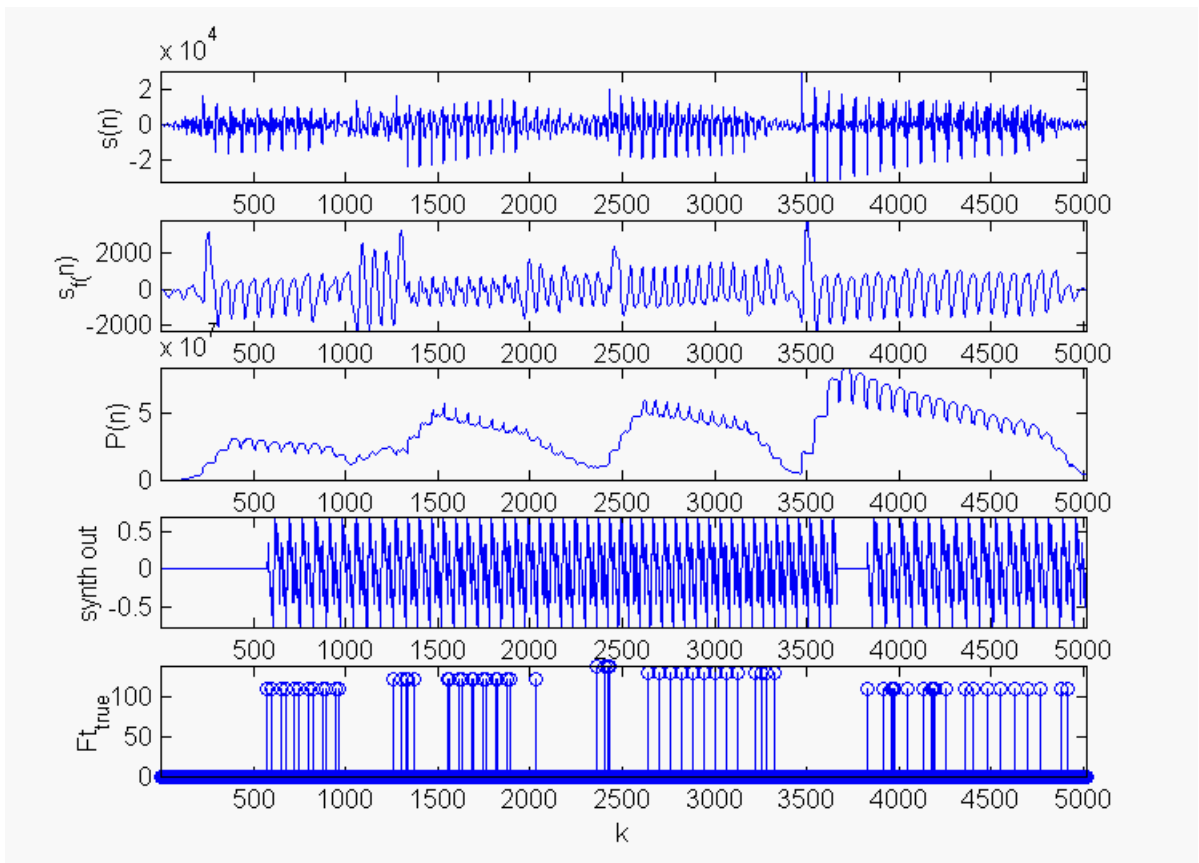
$$\alpha = 2\pi \frac{f_t}{f_s} \cdot n,$$



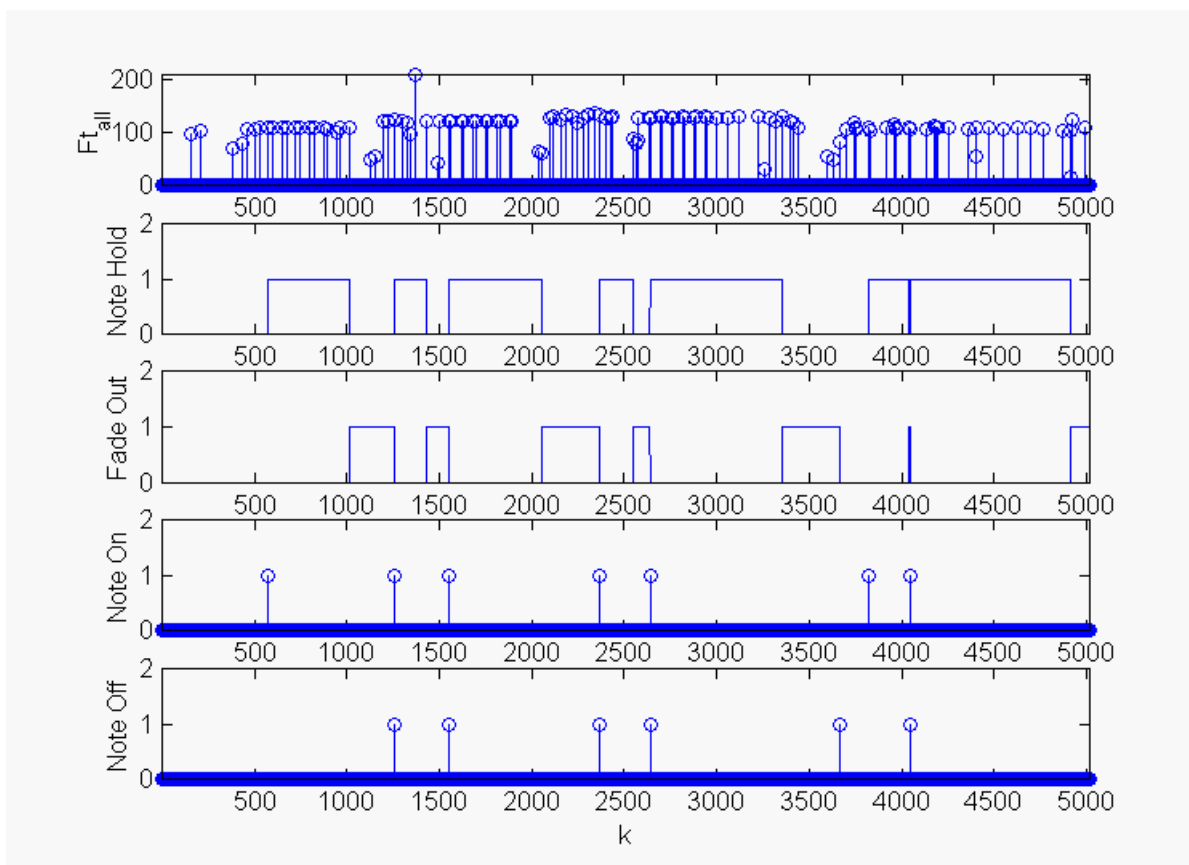
kde  $f_i$  je frekvence hraného tónu a  $n$  diskrétní čas.

Vstupem syntezátorové funkce *gensound()* jsou události *Note Hold*, *Fade Out*, frekvence tónu  $f_i$  a *velocity*. Pokud skončí *Note Hold*, aktivuje se *Fade Out*. Tato událost zajišťuje, že syntezátor pokračuje v generování zvuku až do okamžiku, kdy jeho výstup nabyde nulové hodnoty. Tímto zabráníme vzniku váhovaných jenotkových skoků vnímaných posluchačem jako praskání.

Na obr. 5.10 a 5.11 jsou zobrazeny výstupy jednotlivých částí výše popsaného řetězce syntetické kytary v následujícím pořadí: analyzovaný signál  $s(n)$ , signál za dolní propustí  $sf(n)$ , odhad výkonu signálu  $P(n)$ , výstupní signál varhanního syntezátoru *synth out*, pozitivně detekované frekvence  $F_{ttrue}$ , všechny detekované frekvence  $F_{tall}$ , události *Note Hold*, *Fade Out*, *Note On*, *Note Off*.



Obr. 5.10: Výstup syntetické kytary



Obr. 5.11: Výstup syntetické kytary

## 5.7 Tvorba MIDI souboru

Kompletní popis struktury MIDI souboru nalezneme např. v [7]. Zde zmíníme jen jeho základní rysy.

MIDI soubor je tvořen dvěma částmi. První část obsahuje hlavičku se základními informacemi, druhá část obsahuje tok MIDI dat. V hlavičkovém bloku nastavujeme, zda budeme používat jeden či více kanálů (maximálně 16). Jednotlivé kanály jsou polyfonní a umožňují přiřazení různých nástrojů. Dále definujeme, na jaké intervaly bude dělitelná čtvrtová nota. Tuto hodnotu jsem zvolil

$$tpq = 480 = 3 \cdot 5 \cdot 32. \quad (5.20)$$

To nám umožňuje hrát kromě až stovacetiosminových not také trioly a kvintoly. Jakékoliv jiné intervaly jsou zaokrouhlovány na uvedený počet tiků  $tpq$ . Dále je v hlavičkovém bloku

uvedeno tempo (počet čtvrtových not za minutu) a časové předznamenání. Výchozí tempo je 120 a časové předznamenání 4/4.

Datová část MIDI souboru je tvořena meta-událostmi, MIDI událostmi a SYSEX událostmi. Před každou událostí je uložen tzv. *delta-čas*, což je čas odvozený z tempa udávající, za jak dlouho od předchozí události se tato má provést. Pokud chceme tedy provést více událostí v jeden okamžik (např. zahrát více tónů najednou), mezi příslušné události vložíme nulové *delta-časy*.

### 5.7.1 Meta-události

Meta události umožňují zaznamenání nehuděbních informací do souboru, jako je například číslo hudební stopy, libovolný text na začátku i v průběhu souboru, jako například název nahrávky, text písně apod. Dále poskytují možnost změny tempa a časového předznamenání apod. Některé meta-události jsou povinné, například událost značící konec hudební stopy.

### 5.7.2 MIDI události

Mezi MIDI události patří již zmíněné *Note On* a *Note Off*, dále například *Key Pressure*, umožňující změnu dynamiky tónu v průběhu jeho znění, změnu nástroje reprezentuje událost *Program* a plynulou změnu výšky tónu *Pitch Wheel*.

### 5.7.3 SYSEX události

SYSEX (SYStem EXclusive) události – jedná se o rezervované systémové události, jejichž význam a obsah si definují výrobci, v této práci nejsou používány.

## 5.8 Práce se syntetickou kytarou

Softwarová část syntetické kytary je tvořena třemi programy. Program *Tuner*, představující ladičku, po spuštění zobrazuje na monitoru jméno, skutečnou výšku, správnou výšku dle komorního  $a = 440\text{Hz}$  a rozladění tónu v procentech.

Program *Calib* je určen pro kalibraci systému. Po jeho spuštění je třeba udeřit do všech strun a program si nastaví koeficienty číslicového zesílení vstupního signálu a analyzuje skutečné naladění strun. Pro toto naladění pak přepočítá relativní výšky tónu pro jednotlivé struny. Z tohoto důvodu se pak program *G\_synth* i pro zcela nenaladěnou kytaru chová, jako by naladěna byla.

Program *G\_synth* reaguje v reálném čase na kytarovou hru a generuje zvuk varhan. Zároveň ukládá na disk MIDI soubor, v němž jsou zachyceny všechny události vzniklé při hře.

## 6 Závěr

Těžištěm návrhu kytarového převodníku, potažmo celé syntetické kytary, je nalezení rychlé a spolehlivé analýzy kytarového signálu. Vzhledem k typickému spektrálnímu složení zvuku tohoto nástroje, ve kterém nejsou složky odpovídající frekvenčním výškám jednotlivých tónů obvykle dominantní, rozhodl jsem se pro oddělenou analýzu zvukových signálů jednotlivých strun. Za tímto účelem jsem pro snímání použil selektivní MIDI snímač tvořený šesti oddělenými magnetickými obvody příslušejícími jednotlivým strunám.

Pro návrh tónové detekce jsem nejprve provedl srovnání několika metod analýzy signálu ve frekvenční a časové oblasti. Frekvenční oblast byla zastoupena diskrétní Fourierovou transformací a lineární predikcí, časová oblast bankou filtrů, autokorelační metodou a přímou detekcí periody. Nejlepších výsledků dosahovala autokorelační metoda co do rychlosti i spolehlivosti. Pro její vysoké výpočetní nároky jsem se však rozhodl použít přímou detekci periody, přinášející srovnatelně dobré výsledky, zato při výpočetních nárocích v pravdě minimálních.

Pro tuto metodu jsem navrhl detekci význačných extrémů s adaptivním prahováním. Z pozic minim a maxim je odhadována délka periody a frekvence hraného tónu. Pro rozlišení chybně a správně určených frekvencí jsem definoval několik pravdivostních kritérií. Jejich aplikace má za důsledek velkou úspěšnost tónové detekce.

Na základě získaných informací o vstupním signálu jsou vytvářeny MIDI události pro ovládání MIDI syntezátoru. Dále jsem navrhl vlastní syntezátor varhanního zvuku pracující v reálném čase.

Systém popsany v této práci tak realizuje kompletní řetězec syntetické kytary od rychlého a robustního kontroleru až po syntezátor a je podle mého mínění dobře použitelný pro živou hudební produkci.

Dále mohu konstatovat, že vhodnou modifikací kritérií u přímého měření periody je analyzující část řetězce úspěšně použitelná i pro analýzu hudebních signálů jiných nástrojů a pro detekci základního řečového kmitočtu.

## 7 Abstract

This work focusses on development of a guitar-to-MIDI converter. The keystone of the converter lies in fast and reliable tone frequency recognition. That's the reason why a broad part of the work covers discussion on selecting an appropriate recognition method. I have examined various frequency (FFT, LPC) and time domain (bank of filters, autocorrelation, direct period measurement) analysis methods. The mentioned frequency methods of analysis appeared to be unusable because of their poor time/frequency resolution. The best results have been displayed by autocorrelation which brings positive detection without any additional requirements on signal preprocessing. The main disadvantage to this method is its relatively large number of variable-by-variable multiplications.

That's why I decided to use the direct period measurement which brings comparable results, while only a minimal calculation power is demanded. The analyzed signal is low-pass filtered. Then adaptive thresholding for determination of major peaks is used. Subsequently, we calculate the period length and frequency from the distance between neighboring major peaks and use several criterions to determine whether the frequency is correct or not. This seems to be a very robust technique. In the end, the MIDI messages are generated. I have also developed a simple organ synthesizer in order to make the whole system suitable for real time musical performance.

## Příloha

Kytarový MIDI snímač SH 1550

výrobce: Shadow Elektroakustik

jmenovitá impedance vinutí cívek: 1,7 k $\Omega$

počet vinutí: 6

maximální špičková výstupní napětí: cca 0,2V

přibližné rozměry snímače: 9,5 x 1 x 1 cm<sup>3</sup>

3 x zvuková karta Soundblaster 128 PCI

výrobce: Creative Labs

počet monofonních linkových vstupů: 3 x 2

vzorkovací frekvence: 48kHz

bitová hloubka: 16b

## Literatura

- [1] Kadlec, F.: Zpracování akustických signálů. ČVUT, Praha, 1996.
- [2] Bacon, T.: Všechno o kytarách. Nakladatelství Svojka a Vašut, Praha, 1995.
- [3] Zenkl, L.: ABC hudební nauky. Editio Barenreiter, Praha, 2002.
- [4] Sovka, P., Pollák, P.: Vybrané metody číslicového zpracování signálů. ČVUT, Praha, 2001.
- [5] Vejražka, F., Hrdina, Z.: Signály a soustavy. ČVUT, Praha, 2000. Skriptum ČVUT FEL.
- [6] Laipert, M., Vlček, M., Davídek, V.: Analogové a číslicové filtry. ČVUT, Praha, 2000.
- [7] Veenstra, H. J.: MIDI Documentation. [Http://www.omega-art.com/logic/index.html](http://www.omega-art.com/logic/index.html). 2002.