

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

ROZPOZNÁVÁNÍ AKORDŮ

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

ONDŘEJ SOUDEK

BRNO 2015



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

ROZPOZNÁVÁNÍ AKORDŮ

CHORD RECOGNITION

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

ONDŘEJ SOUDEK

VEDOUcí PRÁCE

SUPERVISOR

Doc. Dr. Ing. JAN ČERNOCKÝ

BRNO 2015

Abstrakt

Cílem této bakalářské práce je vytvořit aplikaci, která bude v reálném čase analyzovat signál, pocházející z elektrické nebo akustické kytary, a bude rozpoznávat, jaký akord nebo tón uživatel hraje, případně akord tomu nejbližší. Práce obsahuje základy hudební teorie, způsob řešení zadaného problému a výsledky testování.

Abstract

The goal of this bachelor thesis is to create an application, that will perform a real-time analysis of a signal from electric or acoustic guitar and will recognize a chord or tone, played by user. This document contains basics of music theory and the solution of the problem, as well as the results of testing.

Klíčová slova

Audio, analýza, rozpoznávání, signál, kytara, akordy, vzorky, vzorkování, Steinberg, ASIO, reálný čas, FFT, Fourierova transformace, frekvence

Keywords

Audio, analysis, recognition, signal, guitar, chords, sample, sampling, Steinberg, ASIO, real-time, FFT, Fourier transform, frequency

Citace

Ondřej Soudek: Rozpoznávání akordů, bakalářská práce, Brno, FIT VUT v Brně, 2015.

Rozpoznávání akordů

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Doc. Dr. Ing. Jana Černockého.

.....
Ondřej Soudek
20. května 2015

Poděkování

Rád bych poděkoval vedoucímu mé bakalářské práce, panu Doc. Dr. Ing. Janu Černockému, za vstřícnost, ochotu, věnovaný čas a poskytnutí odborné pomoci. Dále také všem, kteří mi pomohli aplikaci otestovat.

© Ondřej Soudek, 2015.

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

1 Úvod	3
1.1 Existující aplikace	3
2 Kytara a hudební teorie	4
2.1 Tón a akord	4
2.2 Stupnice a stavba akordů	4
2.3 Kytara a hmatník	7
3 Kytarový signál a jeho zpracování	9
3.1 Fourierova transformace	9
3.2 Frekvence jednotlivých tónů	9
3.3 Fyzika strun	10
3.4 Skládání tónů v akord	11
4 Proces rozpoznávání	13
4.1 Obecné schéma rozpoznávače	13
4.2 Klasifikace jednoho akordu	13
4.3 Referenční akordy	14
4.3.1 Obdélníkové referenční spektrum	15
4.3.2 Obdélníky se snižovanou výškou	15
4.3.3 Obdélníky se stejným obsahem	16
4.3.4 Trojúhelníkové referenční spektrum	16
4.4 Dekódování sledu akordů	17
5 Softwarový návrh a implementace	18
5.1 Technologie a knihovny	18
5.2 Návrh a implementace aplikace	18
5.2.1 Základní informace	18
5.2.2 Blokové schéma rozpoznávače	19
5.2.3 Popis tříd	19
5.3 Uživatelské rozhraní	20
6 Testování	22
6.1 Testování detekce	22
6.1.1 Metriky	22
6.1.2 Výsledky	23
6.2 Funkční testování a ladění aplikace	25
6.3 Uživatelské testování	25

7 Závěr	27
7.1 Shrnutí	27
7.2 Další vývoj	27
7.2.1 Blízká budoucnost	27
7.2.2 Složitější cíle	27
A Návod k použití	30
B Obsah CD	32

Kapitola 1

Úvod

Tato práce byla vytvořena na Fakultě informačních technologií na Vysokém učení technickém v Brně. Práce se zabývá problematikou rozpoznávání akordů. Aplikací s podobným zaměřením na trhu není mnoho. Z toho důvodu není moc možností, kde hledat inspiraci pro řešení tohoto problému. Nicméně i tak existuje více způsobů, jak zpracovávat a analyzovat digitální audio signál.

Hlavním úkolem této práce bylo navrhnout a vytvořit algoritmus a aplikaci, která bude v reálném čase analyzovat signál, pocházející z elektrické nebo akustické kytary a bude rozpoznávat a zobrazovat akordy, které uživatel hraje. Pro implementaci aplikace byl zvolen jazyk Java. Na úvod této práce si řekneme něco o základech hudební teorie, potřebných pro pochopení terminologie, použité v práci, a pro zasazení práce do kontextu reálného světa. Druhá část se věnuje řešení zadaného problému. V této části se dozvíme něco o zvoleném řešení a způsobu tvorby referenčních dat. Nakonec se podíváme na samotnou implementaci zvoleného algoritmu a projdeme si několik způsobů jeho testování.

1.1 Existující aplikace

Aplikací, provádějících rozpoznávání akordů v reálném čase není mnoho. I proto se tato práce zabývá právě tímto tématem.

Rocksmith 2014

Rocksmith 2014 je počítačová hra / naučný software, vyvíjený studiem Ubisoft San Francisco. Hra umožňuje hráči zapojit elektrickou kytaru nebo baskytaru a hrát akordy nebo tóny, které hra zobrazuje, přičemž v reálném čase kontroluje, zda jsou doopravdy hrány[8].

Narozdíl od této práce se Rocksmith 2014 v kontextu rozpoznávání akordů nestará o rozpoznání a zobrazení hraného akordu, ale o kontrolu, zda uživatel zahrál zobrazený akord. Aplikace již předem přesně ví, jaké akordy nebo tóny v daný čas očekávat a nemusí tedy kontrolovat ostatní.

Kapitola 2

Kytara a hudební teorie

2.1 Tón a akord

Než probereme samotný proces analýzy a zpracování signálu, musíme si nejprve říct něco o hudební teorii. Hudba se skládá z jednotlivých tónů, které mají různé vlastnosti. Pro tuto práci je nejdůležitější výška tónu, která silně souvisí s frekvencí daného tónu. O frekvencích se dozvíme více v sekci 3.2.

V hudbě používáme a rozlišujeme 12 základních půltónů[10]:

| C | C#/Db | D | D#/Eb | E | F | F#/Gb | G | G#/Ab | A | A#/Hb | H |

Dva půltóny tvoří celý tón. V základní notové řadě máme přirozené půltóny, které jsou mezi tóny E - F, H - C. Základním intervalem v hudbě je oktáva. Oktáva je interval mezi dvěma tóny s poměrem frekvencí 2:1. Mějme například tón T1 s frekvencí 20 Hz. Tón o oktávu výše, T2, by tedy měl frekvenci 40 Hz a tón o oktávu níže, T0, by měl frekvenci 10 Hz.

Jako akord můžeme obecně označit souzvuk tří a více tónů, uspořádaných do určitého systému. O tomto systému pojednává následující sekce.

2.2 Stupnice a stavba akordů

Stupnice je množina tónů, s definovanými intervaly vzhledem k základnímu tónu stupnice. Existuje mnoho druhů stupnic s různým využitím v hudbě. Stupnice se mohou dělit například podle počtu obsažených tónů na pentatonické (5 tónů), diatonické (7 tónů) nebo chromatické (12 tónů).

Asi nejznámější stupnicí je diatonická durová stupnice, která obsahuje v rámci jedné oktávy 7 tónů a tón o oktávu výše, než základní tón stupnice. Tyto tóny se označují jako prima, sekunda, tercie, kvarta, kvinta, sexta, septima, oktáva. Například pro stupnici C_{dur} jsou to tyto tóny:

Tabulka 2.1: Stupnice C_{dur}

prima	sekunda	tercie	kvarta	kvinta	sexta	septima	oktáva
C1	D1	E1	F1	G1	A1	H1	C2

Stejně jako existuje mnoho druhů stupnic, existuje i mnoho druhů akordů. Každý typ akordu využívá některé z tónů určité stupnice. Pro lepší představu si tento systém předvedeme na několika typech akordů se základem v tónu **A** a pomocí tabulatur si ukážeme rozdíly mezi danými akordy, a jakým způsobem lze akordy zahrát na kytaru [2]:

Durové akordy

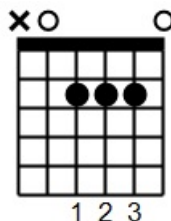
Durové akordy jsou jedny z nejpoužívanějších akordů v hudbě všech žánrů. Jsou definovány jako **prima + tercie + kvinta durové stupnice** daného tónu.

Tabulka 2.2: Jónská (durová) stupnice tóniny A

prima	sekunda	tercie	kvarta	kvinta	sexta	septima	oktáva
A2	H2	C#3	D3	E3	F3	G3	A3

Akord A (A_{dur}) je tedy tvořen tóny A2, C#3 a E3, případně jinou oktávou těchto tónů. Na kytaru lze zahrát například následujícím způsobem:

Obrázek 2.1: Tabulatura akordu A_{dur}



Mollové akordy

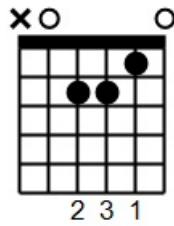
Mollové akordy jsou stejně jako ty durové jedny z nejpoužívanějších nezávisle na žánru. Skládají se z **primy + tercie + kvinty mollové stupnice** základního tónu.

Tabulka 2.3: Aiolská (mollová) stupnice tóniny A

prima	sekunda	tercie	kvarta	kvinta	sexta	septima	oktáva
A2	H2	C3	D3	E3	F3	G3	A3

Akord Am (A_{moll} / A_{mi}) je tedy složen z tónů A2, C3 a E3, nebo jinou oktávou těchto tónů.

Obrázek 2.2: Tabulatura akordu A_{moll}



Dominantní septakordy

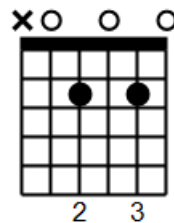
Dominantní septakordy, označované číslem 7 (A_7), nejsou tak běžné, jako durové a mollové, nicméně žánry jako blues a jazz by se bez nich kvůli jejich specifickému zvuku neobešly. Stejně jako akordy durové obsahují **primu** + **tercii** + **kvintu durové stupnice**, ke kterým ještě přibyla **malá septima**.

Tabulka 2.4: Jónská (durová) stupnice tóniny A

prima	sekunda	tercie	kvarta	kvinta	sexta	septima	oktáva
A2	H2	C#3	D3	E3	F3	G3	A3

Akord A_7 je tedy tvořen tóny A2, C#3, E3 a G3.

Obrázek 2.3: Tabulatura akordu A_7



„Silové akordy“

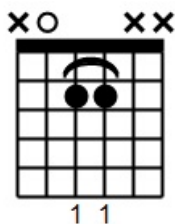
Tzv. silové akordy (anglicky power chords) jsou speciální typy akordů, vyskytující se převážně v rockové, metalové a punkové hudbě. Jsou označovány číslem 5 (A_5). Jelikož obsahují pouze dva tóny, nemusí být v tradičním smyslu slova považovány za akord. Z těchto akordů je vyloučena tercie a není je tedy možné považovat ani za durové, ani za mollové. Zjednodušeně by se dalo říct, že využívají durovou stupnici a obsahují pouze **primu** a **čistou kvintu**.

Tabulka 2.5: Jónská (durová) stupnice tóniny A

prima	sekunda	tercie	kvarta	kvinta	sexta	septima	oktáva
A2	H2	C#3	D3	E3	F3	G3	A3

Akord A_5 je tvořen tóny A2 a E3.

Obrázek 2.4: Tabulatura akordu A_5



2.3 Kytara a hmatník

Vzhledem k tomu, že se tato práce soustředí na signál elektrické a akustické kytary, řekneme si základní informace o kytáře. Standardní kytara má 6 strun a obvykle alespoň 21 pražců. Existují ovšem i kytary, které nemají strun 6, ale například 7, 8 nebo 12. Struny jsou číslovány odspodu nahoru, přičemž první struna má nejvyšší tón, šestá struna naopak nejnižší. Jednotlivých tónů je docíleno stisknutím struny na konkrétním pražci a zahráním dané struny. Výška tónu se zvyšuje se zkracováním hrané části struny a to vždy o jeden půltón na každý pražec[2]. Nejpoužívanější ladění kytary je následující:

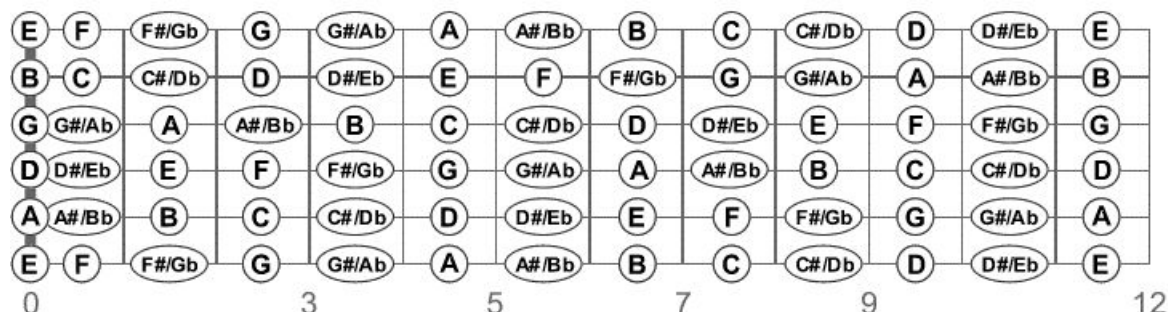
Tabulka 2.6: Základní ladění kytary

Struna	6	5	4	3	2	1
Tón	E2	A2	D3	G3	H3	E4

Není to ovšem ani zdaleka jediný používaný způsob ladění. Jiné způsoby mohou například zahrnovat snížení všech tónů o jeden až několik půltónů („D” ladění), snížení tónu šesté struny o celý tón („Drop D” ladění), případně tak, aby zahráním všech strun bez stisku hmatníku zněl akord („Otevřené E” ladění). Jejich použití závisí na žánru a konkrétním hudebníkovi.

Kytara je specifická tím, že existuje mnoho různých způsobů, jak zahrát jeden tón (například tón A_3 lze při standardním ladění zahrát na čtyřech místech: šestá struna - 17. pražec, pátá struna - 12. pražec, čtvrtá struna - 7. pražec a třetí struna - 2. pražec).

Obrázek 2.5: Část hmatníku kytary při použití standardního ladění. V obrázku je tón H označen jako B. Toto označení se používá v anglické literatuře. Převzato z <https://shekhassan.wordpress.com/2013/11/22/guitar-chord-mastery-studiopigeon-method-ch-4/>.



Na obrázku 2.5 můžeme vidět rozmístění jednotlivých tónů na hmatníku kytary. Tóny jsou s každou další strunou (s výjimkou druhé struny) posunuty o 5 půltónů (5 prážců). Například tón C3 zle tedy zahrát na šesté struně - 8. prážci, ale také na páté struně - 3. prážci. Na obrázku je zobrazeno pouze 12 prážců. Protože existuje jenom 12 půltónů, od dvanáctého prážce se tóny opakují, pouze jsou položeny o oktávu výše.

Kapitola 3

Kytarový signál a jeho zpracování

3.1 Fourierova transformace

Fourierova transformace (pojmenovaná po Josephu Fourierovi) je proces rozkladu časově závislého signálu na součet harmonických funkcí sinus a cosinus. Dochází při něm k převodu signálu z časové oblasti do oblasti frekvenční. Signál může být ve spojitém i diskrétním čase. Pro tuto práci je důležitý diskrétní čas. Vstupní signál je na začátku digitalizován a zaznamenán jako množina N hodnot. Hodnoty jsou zaznamenány s pravidelným rozestupem. Diskrétní Fourierova transformace využívá následující vztah^[4]:

$$D(k) = \sum_{n=0}^{N-1} d(n)e^{-\frac{i2\pi}{N}nk} \quad k = 0, \dots, N - 1 \quad (3.1)$$

Podle tohoto vztahu vyžaduje výpočet diskrétní Fourierovy transformace N^2 komplexních součtů a součinů. Takové množství se příliš nehodí pro zpracování v reálném čase. V této práci tedy byla použita efektivnější metoda diskrétní Fourierovy transformace, tzv. rychlá Fourierova transformace (anglicky Fast Fourier Transform¹). Tento algoritmus popsali v roce 1965 J.W. Cooley a J.W. Tukey. Metoda pro výpočet potřebuje jenom $N \log_2 N$ komplexních součtů a $\frac{N}{2} \log_2 N$ komplexních součinů. Vzhledem k tomu, že rozlišení spektra zvukového signálu je konstantní (dáno podílem vzorkovací frekvence a počtu vzorků - $\frac{f_s}{N}$), není tento způsob analýzy příliš vhodný pro zpracovávání nízkých frekvencí. Rozlišení hudebních tónů totiž není konstantní, ale geometrické. Rozdíl frekvencí dvou nízkých tónů je podstatně menší, než rozdíl dvou vysokých tónů (viz tabulka 3.1). Tato práce se však nesoustředí na přesnou detekci nízkých tónů, ale na detekci celých akordů. Metoda rychlé Fourierovy transformace je tedy pro práci dostatečně přesná.

Výsledkem FFT je $\frac{N}{2}$ komplexních čísel, které se rovnicí $X = \sqrt{R^2 + I^2}$, kde R je reálná část komplexního čísla a I je imaginární část komplexního čísla, převedou na absolutní hodnoty, se kterými se dále pracuje.

3.2 Frekvence jednotlivých tónů

Jak již bylo zmíněno v sekci 2.1, výška tónu silně souvisí s frekvencí tónu. Na kytáře je tónu dosaženo pomocí vibrací struny a frekvence těchto vibrací potom udává tón. Frek-

¹Běžně zkracováno jako FFT.

vence všech jednotlivých tónů se dají vypočítat z frekvence jakéhokoliv jednoho tónu[1]. Vzhledem k tomu, že máme 12 půltónů a tón o oktávu výše má dvojnásobnou frekvenci, můžeme odvodit následující vzorec:

$$f_{ref+i} = f_{ref} \times \sqrt[12]{2^i}, \quad (3.2)$$

kde f_{ref+i} je frekvence vzniklého tónu, f_{ref} je frekvence původního tónu a i je počet půltónů mezi původním a novým tónem. Například tón E2 má frekvenci 82,406 Hz. Pokud z tohoto tónu budeme chtít vypočítat například tón G2, tedy tón o 3 půltóny výše, uděláme to následujícím způsobem:

$$f_{G2} = 82,406 \times \sqrt[12]{2^3} \text{ Hz}$$

$$f_{G2} = 97,998 \text{ Hz}$$

Nyní tedy potřebujeme pouze jeden základní tón a přesně definovanou frekvenci. Základním tónem byl mezinárodní dohodou hudebníků ve Vídni stanoven tón **A4** (označovaný také jako **komorní A**) s frekvencí **440 Hz**. Není to ale pravidlem. Existuje mnoho hudebníků, kteří, obzvláště na kytáře, používají jinou základní frekvenci. Ta se může lišit o jednotky až malé desítky Hz.

Tabulka 3.1: Tabulka frekvencí tónů při základní frekvenci 440 Hz

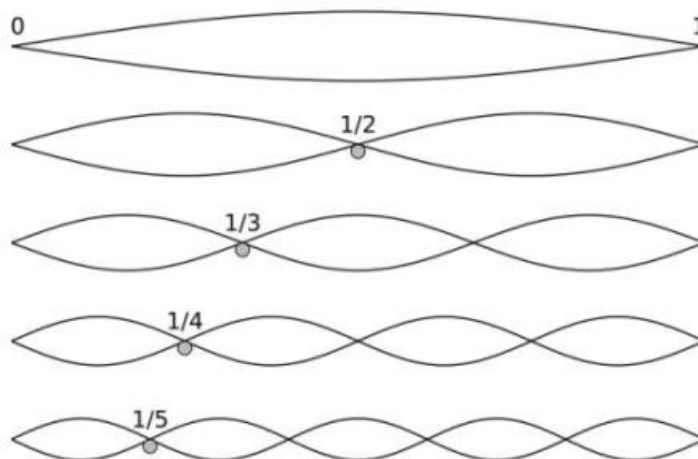
tón	0	1	2	3	4	5	6
C	16,351	32,702	65,404	130,808	261,616	523,232	1046,464
C#	17,323	34,646	69,292	138,584	277,168	554,336	1108,672
D	18,354	36,708	73,416	146,832	293,664	587,328	1174,656
D#	19,445	38,890	77,780	155,560	311,120	622,240	1244,480
E	20,621	41,242	82,484	164,968	329,936	659,872	1319,744
F	21,826	43,652	87,304	174,608	349,216	698,432	1396,864
F#	23,124	46,248	92,496	184,992	369,984	739,968	1479,936
G	24,499	48,998	97,996	195,992	391,984	783,968	1567,936
G#	25,956	51,912	103,824	207,648	415,296	830,592	1661,184
A	27,500	55,000	110,000	220,000	440,000	880,000	1760,000
A#	29,135	58,270	116,540	233,080	466,160	932,320	1864,640
H	30,867	61,734	123,468	246,936	493,872	987,744	1975,488

3.3 Fyzika strun

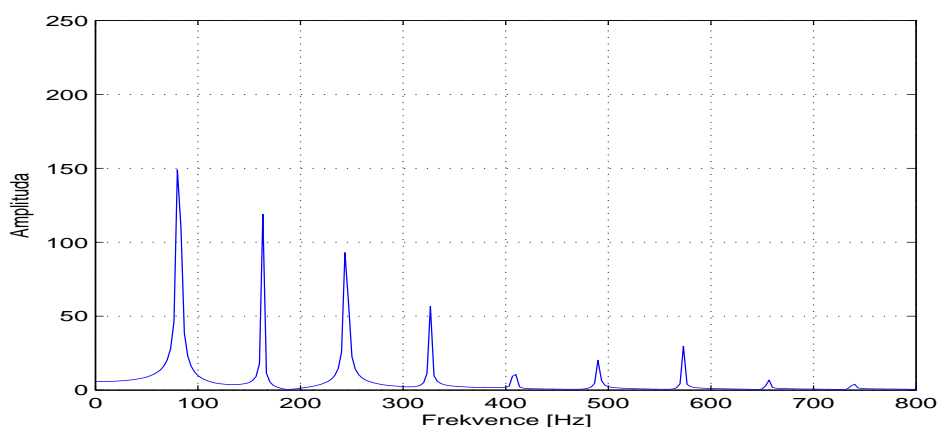
Když jsme se v předchozí sekci bavili o frekvencích tónů, měli jsme se spíše bavit o **základních frekvencích tónů**. V reálném světě nejsou tóny tvořeny pouze z jejich základních frekvencí, ale také z tzv. harmonických frekvencí. Na vibrující struně totiž vzniká tzv. stojaté vlnění. Struna vibruje v celém harmonickém spektru. Základní frekvence je nejhlubší a nejhlasitější. S touto frekvencí zní další harmonické tóny, které jsou tišší a vyšší. Harmonické frekvence jsou celočíselné násobky základní frekvence [3]. Tento fakt značně ztěžuje analýzu signálu, protože harmonické frekvence jednoho tónu jsou také základní frekvence

tónu jiného.

Obrázek 3.1: Ukázka vibrace struny. Převzato z <http://www.ethanhein.com/wp/2009/tuning-the-quantum-guitar/>.



Obrázek 3.2: Tón E2 ve frekvenční doméně

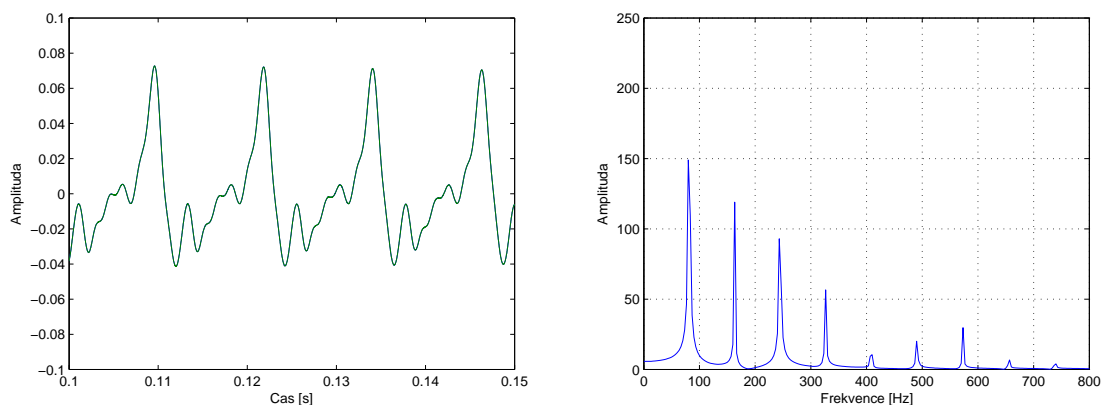


Na obrázku 3.2 můžeme vidět frekvenční spektrum tónu E2. Tón E2 je složen ze základní frekvence (82, 484 Hz) a jejích harmonických frekvencí.

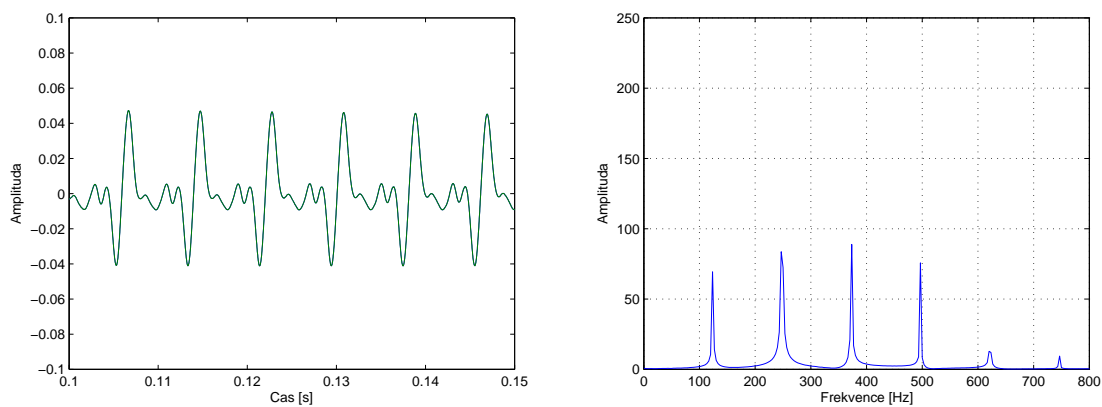
3.4 Skládání tónů v akord

Jak jsou tedy v kontextu signálu a frekvencí tvořeny akordy? Na reálné kytáře se vibrace více strun ovlivňují skrze tělo kytary, nicméně tato skutečnost je velmi neurčitá, protože tento efekt se s časem mění a navíc různé kombinace tónů se ovlivňují jiným způsobem. Pro zjednodušení tedy musíme tuto skutečnost zanedbat. Zjednodušeně se dá říct, že se tóny, hrané na různých strunách současně, sečtou a vznikne akord. Toto si předvedeme na následujících obrázcích. Akord E5 je tvořen z tónů E2 a H2. Na obrázcích 3.3 a 3.4 můžeme vidět signál těchto tónů v čase a frekvenční spektrum těchto signálů a na obrázku 3.5 kombinaci těchto tónů, akord E5.

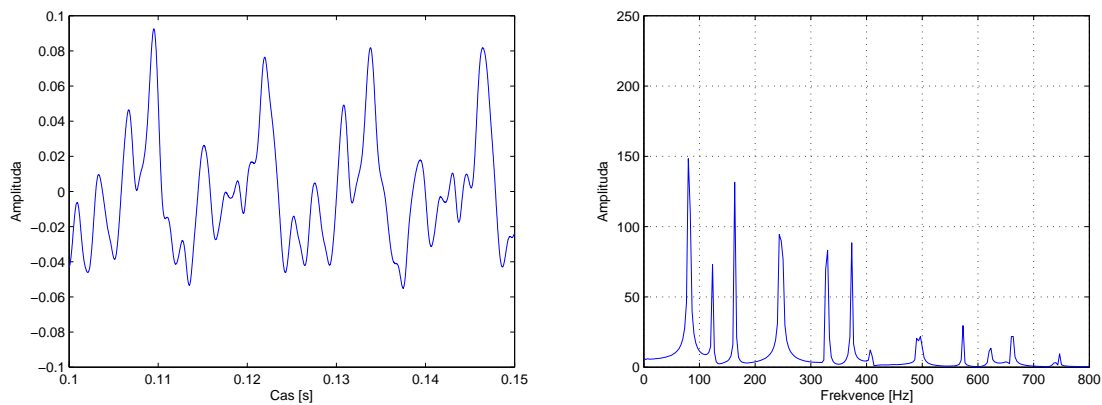
Obrázek 3.3: Tón E2 v časové a frekvenční doměně



Obrázek 3.4: Tón H2 v časové a frekvenční doměně



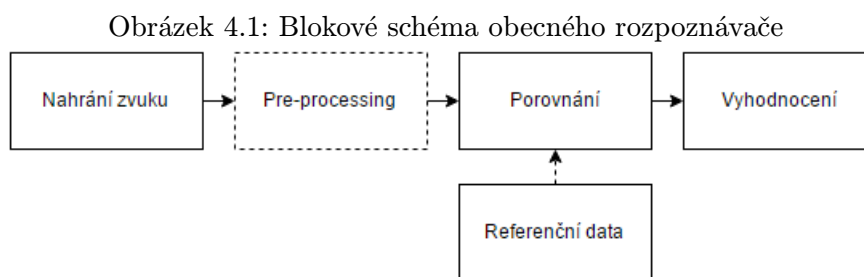
Obrázek 3.5: Akord E5 v časové a frekvenční doměně



Kapitola 4

Proces rozpoznávání

4.1 Obecné schéma rozpoznávače



Všechny jednoduché rozpoznávače mají několik společných částí. Na začátku potřebují nahrát data, se kterými budou pracovat (v našem případě zvuk) a tato data upravit tak, aby je bylo možno porovnávat. V další fázi následuje samotné porovnání s referenčními daty a nakonec interpretace výsledků porovnání.

4.2 Klasifikace jednoho akordu

Pro rozpoznání akordu potřebujeme způsob, jakým nahraná data porovnat s referenčními daty. O tvorbě těchto referenčních dat pojednává následující sekce. Pro účel porovnávání byl zvolen skalární součin. Jednoduše se tedy vezmou frekvenční data nahraného zvuku a vzorky se postupně násobí s odpovídajícími vzorky referenčních akordů. V referenčních akordech jsou vysoké hodnoty pouze v rozsahu očekávaných frekvencí, ve zbytku je 0. Pokud tedy frekvence nahraného zvuku odpovídají frekvencím v referenčním akordu, vynásobí se dvě vysoké hodnoty a vznikne opět vysoká hodnota. Naopak pokud frekvence neodpovídají, vynásobí se vysoká hodnota frekvence z nahraného zvuku s číslem 0, vznikne tedy 0. Výsledné násobky se sečtou a vznikne nám jedno číslo. Toto číslo by se dalo označit jako hodnota příslušnosti k danému akordu, ale řekněme mu jednoduše skóre. Rovnice tohoto systému tedy vypadá následovně:

$$S = \sum_{n=0}^{\frac{N}{2}} f_n^{ref} \times |D(k)|_n, \quad (4.1)$$

kde S je výsledné skóre, N je počet vzorků FFT, f^{ref} je frekvenční spektrum referenčního akordu a $|D(k)|$ je frekvenční spektrum reálných nahraných dat. Čím větší je toto skóre, tím větší je pravděpodobnost, že nahraný zvuk je akord s tímto hodnocením.

To ovšem není všechno. Akordy, které jsou složeny z více tónů, mají větší potenciál získat vyšší skóre, neboť referenční akordy, těmto akordům odpovídající, mají více nenulových koeficientů. Například akord E5 je složen z tónů E2, H2 a případně E3, tedy 2-3 tóny, E_{dur} obsahuje tónů 6 - E2, H2, E3, G#3, H3 a E4. Musíme tedy výsledné skóre každého referenčního akordu dělit konstantou, která se zvolí podle počtu tónů v referenčním akordu. Tyto konstanty byly stanoveny experimentálně a jsou uvedeny v následující tabulce.

Tabulka 4.1: Tabulka konstant

Počet tónů	1	2	3	4	5	6
Konstanta	9,4	9,8	11,4	11,5	11,9	12

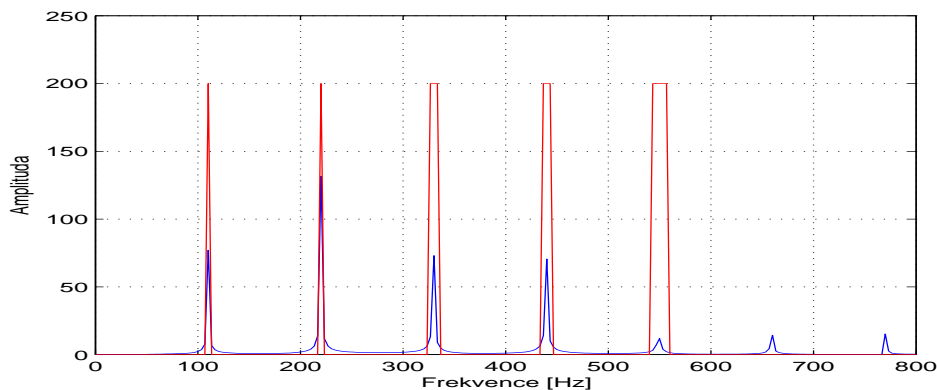
4.3 Referenční akordy

Důležitou částí této práce je vytvoření dat, se kterými se nahraná data budou porovnávat. Říkejme těmto datům referenční akordy. K vytvoření těchto referenčních akordů využijeme znalosti, které jsme získali v předchozích kapitolách. Jelikož budeme porovnávat pouze frekvence hraných akordů, časovou doménou se nemusíme zabývat. Nejdříve si tedy musíme vytvořit základní jednotky akordů - tóny. Rovnicí 3.2 si vypočítáme základní frekvence všech tónů. Z těchto frekvencí si vypočítáme nejsilnější (experimentálně stanoveno na 4) harmonické frekvence jednotlivých tónů (viz sekce 3.3). Když máme vypočítány základní i harmonické frekvence, vytvoříme si referenční spektra tónů. Způsobů vytváření referenčního spektra bylo více. Řekněme si tedy něco k těmto způsobům.

4.3.1 Obdélníkové referenční spektrum

V prvním pokusu byly kolem vypočítaných frekvencí vytvořeny obdélníky o šířce $\pm \frac{1}{4}$ dané frekvence a konstantní výšky. Předvedeme si to na následujícím obrázku tónu A2. Modrou barvou je označeno frekvenční spektrum reálného tónu A2, červeně pak referenční tón A2. Tento způsob však nebyl příliš úspěšný.

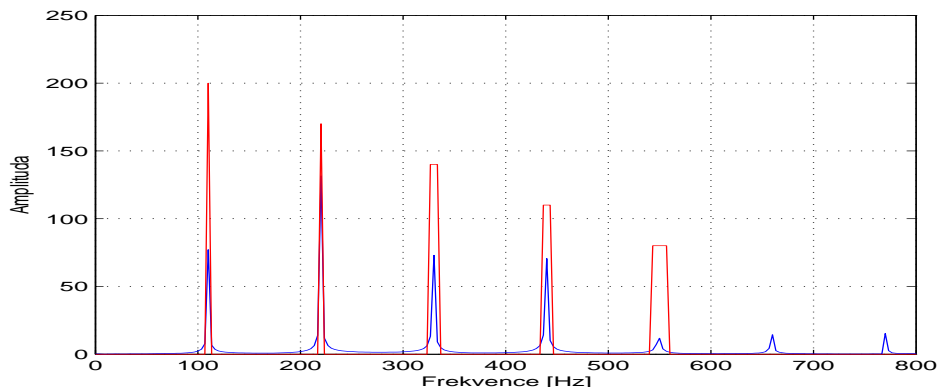
Obrázek 4.2: Tón A2 ve frekvenční doméně a referenční spektrum tónu A2



4.3.2 Obdélníky se snižovanou výškou

První věc, která byla změněna, je snižování výšky obdélníků pro harmonické frekvence, neboť síla těchto frekvencí se u reálného tónu snižuje. Výška obdélníků pro harmonické frekvence byla vypočítána rovnicí $v = max - (n \times max \times 0,15)$, kde max je výška obdélníku základní frekvence tónu a n je pořadí harmonické frekvence. Tento způsob prokazoval značně lepší výsledky, než předchozí, nicméně stále ne dostatečné.

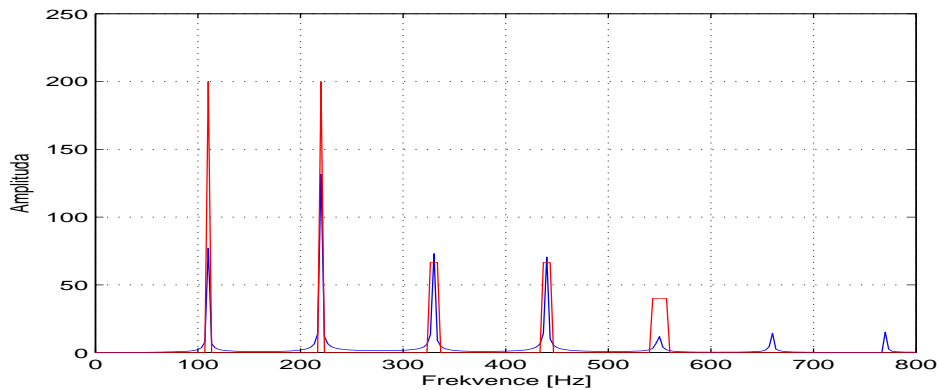
Obrázek 4.3: Tón A2 ve frekvenční doméně a referenční spektrum tónu A2



4.3.3 Obdélníky se stejným obsahem

V další pokus je podobný tomu prvnímu. Změněno bylo to, že není konstantní výška obdélníku, ale jeho obsah. Výška se tedy mění s počtem vzorků v daném rozsahu. Předvedme si to opět na tónu A2. Tato metoda však přinesla zdaleka nejhorší výsledky a byla tedy zavržena.

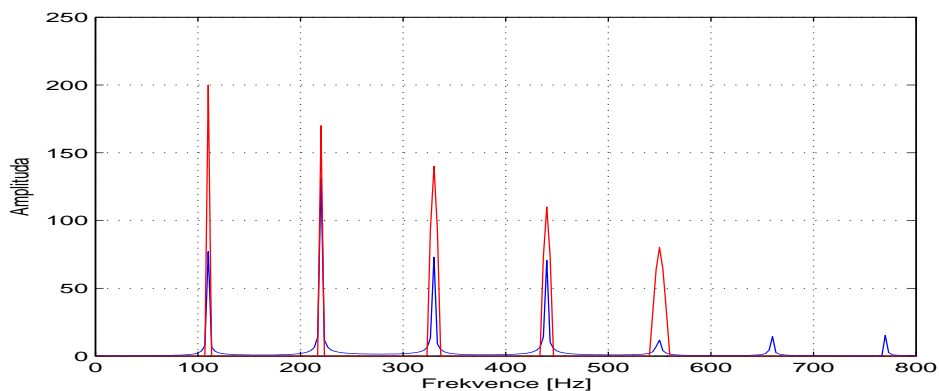
Obrázek 4.4: Tón A2 ve frekvenční doméně a referenční spektrum tónu A2



4.3.4 Trojúhelníkové referenční spektrum

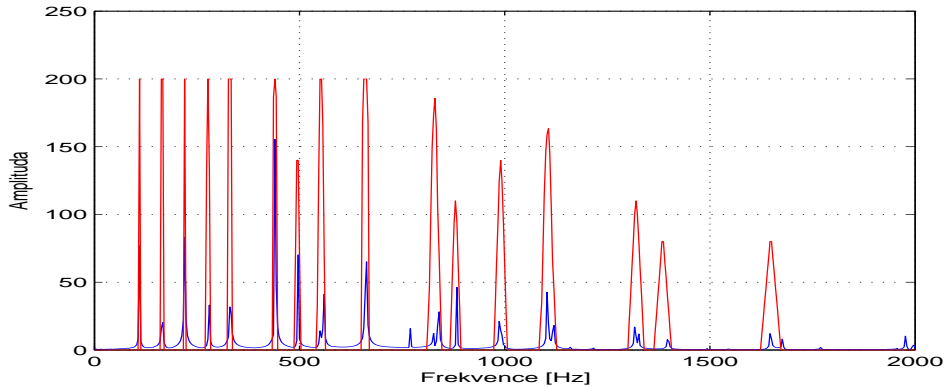
Poslední pokus zahrnoval zohlednění síly harmonických frekvencí, ale obdélníky byly nahrazeny trojúhelníky s vrcholem, ležícím na dané frekvenci. Předvedme si to opět na tónu A2. Tento způsob tvorby referenčního spektra vykazoval nejlepší výsledky, které byly shledány za dostatečné a model byl tedy použit pro výslednou aplikaci.

Obrázek 4.5: Tón A2 ve frekvenční doméně a referenční spektrum tónu A2



Pokud z těchto referenčních tónů chceme vytvořit akordy, jednoduše je spojíme dohromady sečtením jednotlivých hodnot. Jak tedy vypadá referenční spektrum akordu? Ukážeme si na následujícím obrázku.

Obrázek 4.6: Akord A ve frekvenční doméně a referenční spektrum akordu A



Na obrázku 4.6 můžeme vidět akord A (A_{dur}), který je složen z tónů A2, E2, A3, C#4 a E4 a referenční akord A, který vznikl spojením vyjmenovaných referenčních tónů.

4.4 Dekódování sledu akordů

Kytarový signál a jeho frekvenční spektrum není v čase konstantní, i když se akord nemění. Rozdíly mezi některými akordy jsou tak minimální, že i takové malé změny mohou způsobit, že podobný akord najednou dostane vyšší skóre, než hraný akord. Navíc pravděpodobnost přechodu na jiný akord mezi dvěma přibližně 42,6 ms okny je mnohem menší, než možnost, že uživatel hraje stejný akord, jako před 42,6 ms. Proto byl implementován jednoduchý vyhlazovací model. Model funguje následovně. Po zjištění skóre jednotlivých referenčních akordů způsobem, popsaným z předchozí sekce, se toto skóre použije v následující rovnici.

$$V = (1 - \lambda) \times (p \times S) + (\lambda \times S), \quad (4.2)$$

kde V je výsledné skóre referenčního akordu, λ je váha nového skóre, S je skóre, vypočítané způsobem, popsaným v předchozí sekci a p je pravděpodobnost přechodu z předchozího hraného akordu na aktuálně zkoumaný akord. Hodnota konstanty λ byla experimentálně stanovena na 0,9. Jako pravděpodobnost přechodu na stejný akord byla zvolena hodnota 0,01 a pravděpodobnost přechodu na jakýkoliv jiný akord je vypočítána rovnicí $\frac{1-0,01}{N}$, kde N je počet definovaných akordů. V našem případě je počet definovaných akordů 258, tedy výsledná pravděpodobnost přechodu na jakýkoliv jiný akord je přibližně 0,0038. To je asi 2,6× menší pravděpodobnost, než přechod na stejný akord.

Po aplikaci vyhlazovacího modelu a úpravě skóre referenčních akordů je algoritmem vybrán referenční akord s nejvyšším skóre a ten je následně zobrazen jako rozpoznáný akord. Výsledný akord k zobrazení (A^*) se tedy určí následující rovnicí:

$$A^* = \arg \max_{\forall A^{ref}} V(A^{ref}) \quad (4.3)$$

Kapitola 5

Softwarový návrh a implementace

5.1 Technologie a knihovny

Práce je napsána v jazyce Java [6]. Java byla zvolena mimo jiné pro jednoduchou portovatelnost na platformu Android. Uživatelské rozhraní je vytvořeno na platformě JavaFX, která v roce 2014 oficiálně nahradila Swing, zastaralé prostředí pro tvorbu uživatelských rozhraní pro jazyk Java. Program byl vyvíjen pod operačním systémem Windows 8, verze 64-bit.

Aplikace využívá několik knihoven a technologií:

- Audio Stream Input/Output (ASIO) [9], ovladač zvukových karet specifikovaný společností Steinberg. Ovladač poskytuje rozhraní mezi softwarem a zvukovou kartou s nízkou latencí, na rozdíl od technologie DirectSound od Microsoftu.
- JAsioHost [5], rozhraní pro jazyk Java, které zprostředkovává komunikaci s ovladačem ASIO a nahrazuje tedy zastaralé a pomalé rozhraní Java Sound API.
- JTransforms [7], open-source knihovnu, napsanou v jazyce Java, poskytující rychlou implementaci rychlé Fourierovy transformace (FFT) a jiných transformačních metod.

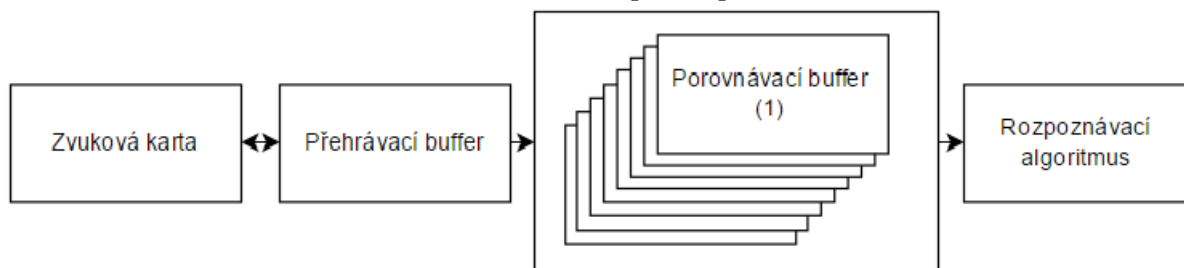
5.2 Návrh a implementace aplikace

5.2.1 Základní informace

Pokud to zařízení podporuje, aplikace používá vzorkovací frekvenci 48000 Hz a dva typy vyrovnávací paměti. Přehrávací buffer pro záznam hraného signálu a jeho reprodukci a větší, porovnávací buffer, pro analýzu nahraného signálu. Porovnávací buffer není jen jeden. Pro zrychlení zobrazování výsledků analýzy aplikace používá 8 porovnávacích bufferů, které jsou plněny prokládaně. To znamená, každá vyrovnávací paměť nahraná v čase t , obsahuje část dat nahraných v čase $t - 1$. Přehrávací buffer má implicitně velikost 512 B, ale tato velikost se dá změnit v nastavení použitého ovladače. Velikost 512 B znamená, že nahraná data jsou zpětně přehrána po asi $10,6 \text{ ms}$ ($\frac{512}{48000}$) + režie, závisující na výkonu počítače. Porovnávací buffery mají velikost 16384 B. K zobrazení tedy dojde po asi $42,6 \text{ ms}$ ($\frac{16384}{8 \times 48000}$) + doba porovnávání. Přehrávací buffer je plněn daty ze zvukové karty. Tato data následně posílá zpět zvukové kartě k přehrávání a kopíruje je do všech porovnávacích pamětí na určené místo s posunem vždy o 2048 vzorků ($\frac{16384}{8}$) s každou další pamětí. Jakmile

je jedna z porovnávacích pamětí naplněna, je na ní aplikován rozpoznávací algoritmus a po jeho dokončení je paměť plněna od začátku.

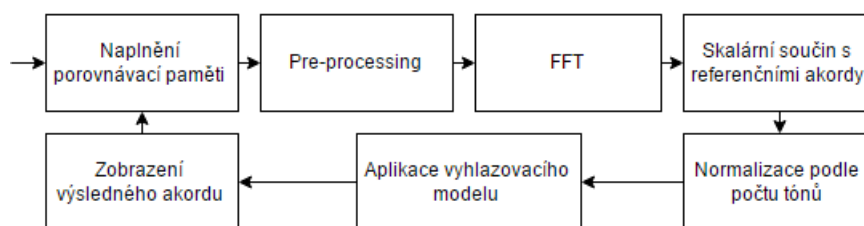
Obrázek 5.1: Schéma plnění pamětí



5.2.2 Blokové schéma rozpoznávače

Zopakujme si celý proces rozpoznávání na jeho blokovém schématu.

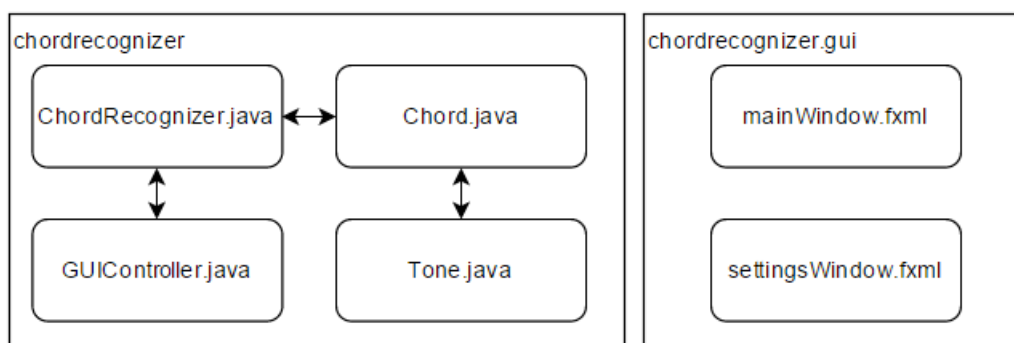
Obrázek 5.2: Blokové schéma rozpoznávače



5.2.3 Popis tříd

Pro aplikaci bylo vytvořeno celkem 6 tříd, z toho 2 třídy pro uživatelské rozhraní.

Obrázek 5.3: Diagram tříd



Popišme si nyní, co která třída dělá:

Třída **Tones** je předpis pro referenční tóny. Uchovává si u každého tónu jeho název a

frekvenční spektrum. Obsahuje také statickou metodu pro vypočítání frekvenčního spektra všech použitých tónů, která závisí na zadané referenční frekvenci tónu A4, použité vzorkovací frekvenci a velikosti porovnávacího bufferu.

Třída **Chords** je šablona pro referenční akordy. Každý referenční akord nese informaci o jménu akordu, jeho frekvenční spektrum, počet tónů v daném akordu a několik dalších informací, důležitých pro běh programu. Třída sama o sobě uchovává také všechny vypočítané referenční akordy a provádí samotný algoritmus pro porovnávání akordů. Obsahuje také metodu pro vytvoření referenčního akordu ze zadaných tónů.

Třída **ChordRecognizer** se stará o režii a komunikaci se zvukovou kartou a obsahuje paměť pro přehrávání a 8 porovnávacích pamětí. Nahrává data z ovladače zvukové karty, přehrává je a naplňuje paměť pro porovnávání. S tou poté dále pracuje. Provádí jednoduchý pre-processing a rychlou Fourierovu transformaci. Zpracovaná data poté posílá do porovnávacího algoritmu a zpracovává výsledek porovnávání.

Třída **GUIController**, jak název napovídá, zprostředkovává komunikaci s grafickým uživatelským rozhraním. Zobrazuje výsledky porovnávání a umožňuje jednoduchou změnu nastavení programu.

Třídy **mainWindow** a **settingsWindow** jsou popisy dvou obrazovek grafického uživatelského rozhraní, napsané v jazyce FXML.

Veškeré zdrojové kódy je možno nalézt na přiloženém CD.

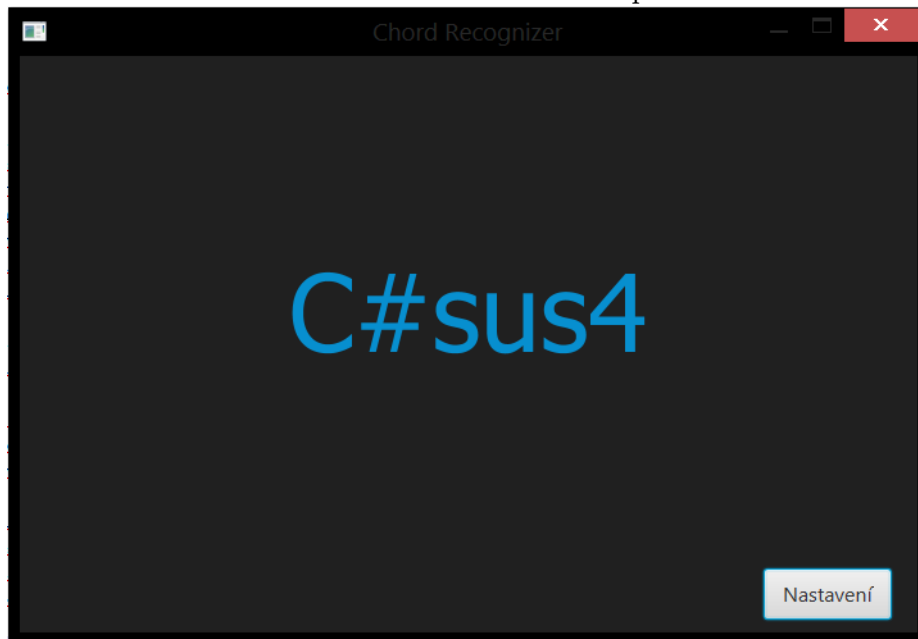
5.3 Uživatelské rozhraní

Při tvorbě grafického uživatelského rozhraní bylo vsazeno na jednoduchost. Rozhraní se skládá ze dvou obrazovek, jedné pro zobrazování rozpoznávaných akordů, druhé pro změnu nastavení programu.

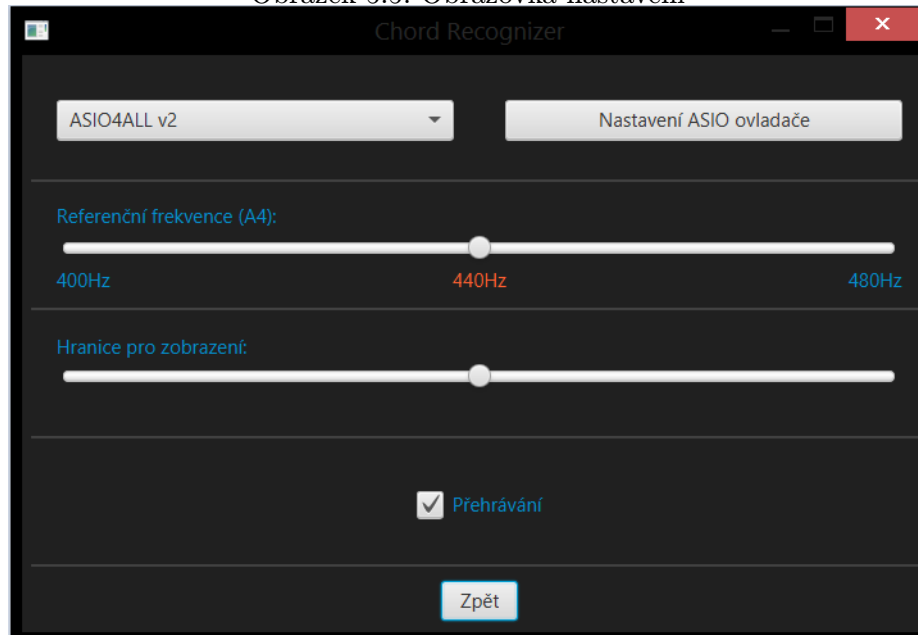
První obrazovka obsahuje dva prvky. V pravém rohu tlačítko pro přechod do obrazovky nastavení a dominantní prvek, text, který vždy zobrazuje jméno hraného akordu. Akordy a samotné tóny jsou odlišeny barvou textu. Tóny se zobrazují červenou barvou, akordy světle modrou. Hlavní obrazovku můžete vidět na obrázku 5.4.

Druhá obrazovka obsahuje ovládací prvky programu. Umožňuje například změnu použitého ASIO ovladače a tlačítko pro vyvolání nastavení samotného vybraného ovladače. Dále obsahuje možnost zvolit si referenční frekvenci tónu A4. S touto změnou dojde znovu k vytvoření všech referenčních tónů a referenčních akordů. Umožňuje také měnit hranici skóre pro zobrazení akordů. Tato hranice je důležitá pro eliminaci šumu a různých rušení. Hodnotu této hranice je možno měnit z důvodu, že například akustická kytara, nahrávaná ne příliš kvalitním mikrofonom, může dosahovat menšího skóre, než elektrická kytara, spojená přímo se zvukovou kartou. Jako poslední možnost nastavení je možnost volby, zda nahraný zvuk přehrávat zvukovou kartou. Přehrávání se hodí při použití elektrické kytary, při použití akustické kytary ale může být nežádoucí. Poslední prvek obrazovky nastavení je tlačítko zpět, kterým se uživatel vrátí na hlavní obrazovku. Ukázkou této obrazovky můžete vidět na obrázku 5.5.

Obrázek 5.4: Hlavní obrazovka aplikace



Obrázek 5.5: Obrazovka nastavení



Kapitola 6

Testování

6.1 Testování detekce

První fáze testování probíhala na statických datech. Předem bylo nahráno 110 akordů různých druhů (durové, mollové, dominantní septakordy, suspendované akordy, durové a mollové septakordy a jiné). Všechny akordy byly nahrány s čistým zvukem bez efektu. Byl použit jeden zvukový kanál a vzorkovací frekvence 48000 Hz . Vždy bylo nahráno několik sekund těchto akordů a ty byly poté uloženy do formátu *.wav*. Z prostřední části těchto akordů bylo poté extrahováno 300 *ms* zvuku, se kterým se dále pracovalo. Veškeré akordy byly nahrány na elektrickou kytaru za použití kvalitní externí zvukové karty a specializovaného kabelu. Cílem tohoto testování bylo porovnat výsledky jednotlivých modelů ze sekce 4.3, udělat si hrubý odhad jejich přesnosti a zvolit ten nejpřesnější model, který bude později vylepšován.

6.1.1 Metriky

Jako faktor pro rozhodování byly zvoleny dva způsoby zobrazení výsledků. Jednočíselnou hodnotu, udávající procentuální úspěšnost rozpoznávacího algoritmu (správně rozpoznáno / celkový počet testovaných akordů) a modifikovanou verzi tzv. konfúzní matice (matice záměn, anglicky confusion matrix) pro detailnější zkoumání jednotlivých akordů. Konfúzní matice je speciální tabulka, která dovoluje zobrazení přesnosti algoritmu. Pro tuto práci vypadá konfúzní matice následovně. Na prvním řádku jsou umístěny všechny referenční akordy, v prvním sloupci zase všechny nahrané akordy. Hodnota na průniku nahraného a referenčního akordu udává skóre, které při porovnávání nahraného akordu referenční akord získal. Předvedme si tento systém na jednoduchém příkladu.

Tabulka 6.1: Příklad konfúzní matice

	A	E	C	D
A	15,4	8,6	9,7	13,6
E	6,5	16,8	12,3	8,1
C	14,7	16,1	15,9	9,6
D	11,6	10,7	9,7	15,0

V matici můžeme vidět, že při zkoumání nahraného akordu A získaly referenční akordy A - 15,4, E - 8,6, C - 9,7, a D - 13,6 bodů. Referenční akord A získal nejvyšší bodové hodnocení a byl by tedy zobrazen. Naopak při porovnávání nahraného akordu C byl nesprávně

vyhodnocen jako nejpravděpodobnější referenční akord E, který by se také zobrazil. Kromě samotných nejlépe ohodnocených referenčních akordů ale můžeme vidět i hodnocení ostatních akordů a můžeme tedy případně pozměnit a zpřesnit algoritmus podle těchto výsledků. Vzhledem k tomu, že N-tý nahraný akord odpovídá N-tému referenčnímu akordu, v tomto případě například 3. nahraný akord (C) odpovídá 3. referenčnímu akordu (C), v ideálním případě by se měly maximální hodnoty řádku objevovat na hlavní diagonále matice.

6.1.2 Výsledky

V této sekci si projdeme výsledky jednotlivých modelů a popíšeme si jejich problémy. U každé varianty můžeme vidět výřez konfúzní matice pro daný model. Plné verze konfúzních matic všech variant referenčního spektra lze nalézt na přiloženém CD ve formátech .csv a .xlsx.

Obdélníkové referenční spektrum

Rozpoznávací algoritmus, používající obdélníkové referenční spektrum (sekce 4.3.1), dokázal rozpoznat 39 ze 110 akordů. Jeho procentuální úspěšnost byla tedy přibližně **35,5%**.

Hlavním problémem obdélníkového referenčního spektra byla pravděpodobně skutečnost, že do okrajů těchto obdélníků mohou zasahovat části frekvencí vedlejších púltónů při nedokonalém naladění kytary nebo při kolísání frekvencí, způsobeném vzájemným ovlivňováním strun. Přesnost 35,5% nebyla shledána za dostatečnou a model nebyl použit.

Obrázek 6.1: Část konfúzní matice pro obdélníkové referenční spektrum

	A#	A#5	A#6	A#7	A#m	A#m7	A#maj7	A#sus4	A	A5	A6	A7	Am	Am7
A#	783613,6	946253,9	776829,6	755732,6	755573,7	705428,6	755754,1	755271,8	17135,19	19148,58	18478,15	49861,56	75432,27	108147,5
A#5	704544,7	889323,5	703727,7	676654,3	704518,3	661859,7	676684,4	704318,5	3083,372	3394,251	3136,081	11798,43	59133,96	67842,33
A#6	794348,3	964200,8	800683,2	766455,3	763783,2	700377,9	766478,3	765653,9	15576,12	19114	16942,59	53000,9	75310,93	109036,7
A#7	547458,2	643849,7	537268,3	613393,3	530687,1	578607,9	547468,1	542678,3	28195,2	31814,97	29631,8	61208,04	88892,93	121877,6
A#m	709882,6	888688,5	703231,5	689447,8	741025,4	698438,5	682119	709662,4	35660	9710,315	35713,13	44360,18	67908,02	76509,8
A#m7	709882,6	888688,5	703231,5	689447,8	741025,4	698438,5	682119	709662,4	35660	9710,315	35713,13	44360,18	67908,02	76509,8
A#maj7	534146,8	636417,7	527307,3	534153,9	523796,2	516720,5	696681,8	510319,1	171617,3	213723,2	172713,9	213131,2	229929,9	258193,8
A#sus4	746521,5	934772,9	740329,8	698364,5	746439,3	683533,6	681768,2	775223,2	4558,701	3300,66	4657	13293,13	62886,74	71613,17
A	15294,97	16173,5	15760,05	23584,81	61728,69	61831,42	322780	13536,17	437812,7	509290,7	429808,7	424148,9	410903,5	381434,6
A5	14712,71	16237,15	15132,74	15505,18	30472,33	30587,11	302816	12962,55	371478	468282,5	370445,3	357726,8	371512,3	342030,6
A6	17367,17	16347,92	17121,87	25657,92	69603,55	69734,89	310057	13723,59	410178,2	479416,1	416552,8	396415,3	380320,7	345944,8
A7	79337,99	98953,09	132064	87724,9	113075,5	108335,7	198478,1	79145,35	253961,3	254092,2	229132,6	394586	268232	366244,4
Am	34303,33	33259,05	40692,8	35061,16	50115,56	50194,21	342047,4	32602,69	405202,6	504233	398916,5	402663,8	442057,1	412510,7
Am7	34303,33	33259,05	40692,8	35061,16	50115,56	50194,21	342047,4	32602,69	405202,6	504233	398916,5	402663,8	442057,1	412510,7

Obdélníky se snižovanou výškou

Při použití těchto obdélníků (sekce 4.3.2) algoritmus rozpoznal 60 akordů ze 110. Jeho procentuální úspěšnost tedy byla přibližně **54,5%**.

Toto zlepšení je způsobeno právě tím, že model reflektuje snižování energie harmonických frekvencí reálných tónů. Problém s použitím obdélníkového referenčního spektra, popsáný na předchozím případě však stále trvá. Ani přesnost 54,5% nebyla považována za dostatečnou a tento model také nebyl použit.

Obrázek 6.2: Část konfúzní matice pro obdélníky se snižovanou výškou

A#	A#5	A#6	A#7	A#m	A#m7	A#maj7	A#sus4	A	A5	A6	A7	Am	Am7	
A#	745733,9	742953,9	744023,2	654920,3	680710	589879,2	655559,8	692350,2	9503,557	9341,263	10294,59	34554,68	58519,32	83568,67
A#5	673955,1	723472,2	674226,3	585011,8	658545,5	569586,1	585011,7	670900,8	2013,511	1836,98	2093,136	8374,384	49577,51	55937,1
A#6	756852,5	749851,3	766149,8	659564,3	682566	585261,9	661310,5	695698,5	8872,804	9333,039	9687,245	36804,62	59025,68	86032,88
A#7	521838,2	503260,8	516977,1	537729,1	472891,1	486042,2	486518,6	483890,2	20181,33	13743,35	21031,02	45455,64	64458,88	89727,92
A#m	677193,6	722953,4	673752,3	593480,9	687681,6	602093,4	588323,5	674143,4	28707,65	4028,923	30416,6	35029,27	53627,35	59945,64
A#m7	677193,6	722953,4	673752,3	593480,9	687681,6	602093,4	588323,5	674143,4	28707,65	4028,923	30416,6	35029,27	53627,35	59945,64
A#maj7	510748,2	499775,1	508239	475128,5	467026,8	431391,3	602312	458813,6	166897,3	174185,5	168238,3	151468	208926,4	193490,7
A#sus4	697170,3	744794,8	693986,7	600757,7	682342,9	585914,4	588029,3	734568,6	2570,545	1746,627	2677,444	8948,953	51606,62	57983,09
A	14425,03	5849,856	14707,41	18987,61	40932,94	43456,19	263542,4	5744,725	412845,6	411707,4	410522	303512,9	380283	270946,5
A5	14133,57	5904,911	14501,52	14579,35	15527,15	15811	249682,1	5466,04	354904,2	378895,9	355060,6	244763,1	345366	235222,6
A6	15353,11	5953,98	15347,79	19921,91	45249,57	47786,43	252589,1	5885,019	386477,6	382524	395374,9	274507	348666,6	236692,3
A7	78590,06	34555	124980,5	81292,64	59184,8	59833,9	187974,8	31910,8	230243,5	206550,8	215324,4	306875,5	235116,3	301150,3
Am	27811,25	16257,55	33791,34	29957,08	29257,48	31241,09	277170,2	19165,52	386458,8	410126,9	383482,1	283121,6	405904,7	299755,1
Am7	27811,25	16257,55	33791,34	29957,08	29257,48	31241,09	277170,2	19165,52	386458,8	410126,9	383482,1	283121,6	405904,7	299755,1

Obdélníky se stejným obsahem

Algoritmus, používající tyto obdélníky (sekce 4.3.3) dokázal rozpoznat 36 ze 110 akordů. S procentuální úspěšností přibližně **32,7%** byl tedy nejméně úspěšný.

Tento model by v jistém rozsahu frekvencí mohl fungovat úspěšně, nicméně problémem je, že pod tímto rozsahem jsou obdélníky zbytečně a neúměrně vysoké, neboť čím nižší frekvence tohoto obdélníku je, tím méně vzorků se v obdélníku nachází, naopak nad tímto rozsahem jsou obdélníky velmi nízké z obdobného důvodu. Jelikož předchozí modely pracovaly úspěšněji, ani tento model nebyl použit.

Obrázek 6.3: Část konfúzní matice pro obdélníky se stejným obsahem

A#	A#5	A#6	A#7	A#m	A#m7	A#maj7	A#sus4	A	A5	A6	A7	Am	Am7	
A#	491439,7	409532,7	411181,9	384476,9	474936,7	367974	386048,1	477974,3	2507,7	2324,873	2564,433	7850,523	13131,09	18473,92
A#5	458727,8	393780,1	382045,8	347968,9	455238,7	344479,9	348268,2	458433,7	785,0771	648,2533	730,8659	2152,321	11140,74	12507,98
A#6	482328,5	403028,8	408600	369698,9	464255,5	351625,9	371646	467568	2387,715	2326,731	2451,49	8448,226	13314,87	19375,38
A#7	363835,7	289841,3	293858,7	332376,2	349988,5	318529	318235,7	352789,2	5107	3723,11	5180,98	10521,39	14827,11	20241,5
A#m	470949	401973,3	389892,3	365455,2	473375	367881,2	364408,9	470653	7596,213	1587,314	7924,042	8957,015	12411,45	13772,25
A#m7	470949	401973,3	389892,3	365455,2	473375	367881,2	364408,9	470653	7596,213	1587,314	7924,042	8957,015	12411,45	13772,25
A#maj7	360786,7	288259,5	291340,4	313838,5	347907,9	300959,6	367317,2	346496,1	98859,11	92366,3	91429,09	69685,73	108323,1	79149,68
A#sus4	474887,7	406041,8	393845,8	367460	471509	364081,3	364310	485561,6	841,8411	598,0231	814,8393	2218,369	11469,29	12845,82
A	3584,908	1942,519	3612,453	4643,995	9384,041	10443,13	104011,4	1949,198	242700,2	228583,7	228422	164648	235495,7	157443,4
A5	3578,826	1973,371	3606,072	3647,198	3681,955	3750,327	100389,8	1957,727	216748,4	214620,4	208982,5	141981,8	215043	140276,5
A6	3798,093	1993,356	3772,903	4856,93	10146,17	11205	101000,3	2031,071	224164,6	215703,7	218587,2	148977,2	216065,7	140878,3
A7	28112,96	14696,14	36409,97	28774,91	20290,77	20952,72	111098,2	14174,62	152999,9	144416	144886,2	168803,4	153606,1	169409,5
Am	6645,873	4539,562	7586,195	7299,138	6759,635	7412,9	107160,3	5019,92	236207,5	227912,5	221722,1	159536,2	242150	165478,7
Am7	6645,873	4539,562	7586,195	7299,138	6759,635	7412,9	107160,3	5019,92	236207,5	227912,5	221722,1	159536,2	242150	165478,7

Trojúhelníkové referenční spektrum

Algoritmus, používající trojúhelníkové referenční spektrum (sekce 4.3.4) dokázal rozpoznat 77 ze 110 akordů. Byl tedy s procentuální úspěšností **70%** nejvyšší.

S použitím trojúhelníků odpadl problém s obdélníky. Navíc je přesnost zvýšena tím, že kombinuje i snižování energie harmonických frekvencí. Jelikož model dosahoval nejlepších výsledků, byl vybrán pro použití v práci a k dalšímu vylepšování.

Obrázek 6.4: Část konfúzní matice pro trojúhelníkové referenční spektrum

A#	A#5	A#6	A#7	A#m	A#m7	A#maj7	A#sus4	A	A5	A6	A7	Am	Am7	
A#	689333,8	651714	687752,5	605388,5	629227,7	545266,5	605979,7	639987,6	8784,8	8194,09	9516,009	31941,3	54093,49	77248,35
A#5	622983,7	634624,8	623234,4	540767,2	608739,5	526508,2	540767,1	620160,4	1861,229	1611,386	1934,832	7741,028	45827,95	51706,57
A#6	699611,5	657764,3	708205,7	609681,3	630943,3	540998,4	611295,4	643082,6	8201,752	8186,877	8954,597	34021,08	54561,55	79526,19
A#7	482371,4	441456,9	477878	497060,5	437126,2	449282,7	449723,1	447293,4	18655,01	12055,57	19440,44	42017,82	59583,84	82941,77
A#m	625977,3	634169,6	622796,2	548595,8	635672,1	556556,9	543828,4	623157,7	26536,48	3534,143	28116,19	32380	49571,5	55411,94
A#m7	625977,3	634169,6	622796,2	548595,8	635672,1	556556,9	543828,4	623157,7	26536,48	3534,143	28116,19	32380	49571,5	55411,94
A#maj7	472120,2	438399,2	469800,8	439194,4	431705,4	398765,1	556759	424113,4	154274,8	152794,3	155514,4	140012,5	193125,2	178857
A#sus4	644443,1	653328,8	641500,3	555322,2	630737,1	541601,5	543556,5	679013	2376,134	1532,129	2474,948	8272,141	47703,6	53597,82
A	13334,06	5131,453	13595,08	17551,57	37837,17	40169,59	243610,6	5310,25	381621,9	361146,9	379474,2	280558,1	351522,1	250454,8
A5	13064,65	5179,747	13404,77	13476,71	14352,83	14615,21	230798,6	5052,642	328062,7	332364,8	328207,3	226251,6	319245,9	217432,7
A6	14191,95	5222,79	14187,03	18415,21	41827,33	44172,33	233485,7	5439,933	357248,2	335547,4	365472,6	253746	322296,8	218791,2
A7	72646,27	30311,4	115528,2	75144,46	54708,64	55308,65	173758,2	29497,38	212830,1	181184,9	199039,3	283666,4	217334,4	278374,2
Am	25707,87	14261,01	31235,69	27691,42	27044,73	28878,32	256207,7	17716,02	357230,8	359760,4	354479,3	261709,1	375206	277084,6
Am7	25707,87	14261,01	31235,69	27691,42	27044,73	28878,32	256207,7	17716,02	357230,8	359760,4	354479,3	261709,1	375206	277084,6

6.2 Funkční testování a ladění aplikace

V této fázi probíhalo testování aplikace, analyzující data, hraná v reálném čase. Toto testování není podloženo žádnými výstupními daty. Jedná se o alfa testování a ladění algoritmu dle citění autora při hře na kytaru. Podle přesnosti rozpoznávání bylo manipulováno s konstantami pro normalizaci skóre (tabulka 4.1) tak, aby bylo dosaženo co možná největší rovnováhy skóre akordů s různým počtem tónů. Dále byly prováděny změny poměrů výšky trojúhelníků u harmonických frekvencí a základní frekvence u referenčních tónů (viz sekce 4.3.2). Přidávaly se nové akordy a měnila se váha skóre (λ) ve vyhlazovacím modelu (viz rovnice 4.3) pro hladší přechody mezi akordy.

K testování byly použity dva typy nástroje - elektrická a akustická kytara. Elektrická kytara byla zapojena dvěma způsoby. Kvalitní externí zvukovou kartou a kvalitním kabelem a méně kvalitní interní zvukovou kartou s méně kvalitním kabelem. Akustická kytara byla testována se dvěma typy mikrofonů. Nekvalitním interním mikrofonem notebooku a kvalitnějším mikrofonem headsetu. Obecně se dá tvrdit, že čím kvalitnější technologie byla použita, tím přesnější bylo rozpoznávání. Elektrická kytara dosahovala vyšší přesnosti než akustická. Mimo to, že kabelem zapojená elektrická kytara má kvalitnější signál, při hře na akustickou kytaru a použití mikrofonu také dochází k rušení okolním hlukem. I přesto se algoritmus při použití akustické kytary jevil jako dostatečně přesný.

6.3 Uživatelské testování

Po dostatečném otestování aplikace autorem byla aplikace poskytnuta nezúčastněným uživatelům. Jako cílová skupina testu byli vybráni relativně zkušenější hráči na kytaru. Nejméně zkušený uživatel dle dotazníku hraje na kytaru 5 let. Uživatelé si mohli zvolit, zda použijí akustickou kytaru a mikrofon, nebo elektrickou kytaru a kytarový kabel. Všem byl poskytnut návod, jak aplikaci zprovoznit na počítači, kde aplikace a ovladač, nutný k běhu aplikace, nikdy nainstalovány nebyly. Nikdo z testerů neměl problém program dostat do chodu. Součástí návodu bylo také to, jak ovládat aplikaci a ASIO ovladač, nutný k běhu programu. Poté uživatel hrál na zvolenou kytaru a porovnával akordy, které hraje se zobrazovanými akordy. Nakonec testeré vyplnili dotazník, obsahující například otázky na posouzení přesnosti programu, hodnocení rozpoznávacího algoritmu, vzhled nebo obecně celkový dojem z aplikace. Dotazník je k dispozici na příloženém CD.

Na otázku „Setkal(a) jste se dříve s podobným typem programem? Jestli ano, jaký program to

byl?" všichni uživatelé, kromě jednoho, odpověděli „ne”, jediný odpověděl „ano, už nevím”.

Podle dotazníku aplikace neměla problém fungovat v reálném čase ani na průměrném procesoru. Ani jeden uživatel neměl problém s tím, že by program nedokázal rozpoznat nějaký běžně používaný akord, a to jak při použití elektrické kytary, tak při použití akustické kytary.

Všichni uživatelé byli spokojeni s kvalitou grafického uživatelského prostředí, jenom jeden na otázku „Je něco, co byste změnil(a) na uživatelském prostředí?” odpověděl „Pouze po zapnutí není jasné jak aplikace pracuje nebo co zobrazuje.”. Takový problém by se dal vyřešit například tím, že aplikace při prvním spuštění zobrazí jednoduchou nápovědu ke svým ovládacím a zobrazovacím prvkům.

Jediný uživatel na otázku „Zaznamenal(a) jste nějaké problémy s aplikací?” odpověděl „pouze počáteční nastavení ASIA, ale to občas zlobí (vlastní zkušenost)”. Vzhledem k tomu, že se tento problém týká používaného externího ovladače, nedá se s ním v této práci nic dělat.

Otázka „Jaký je Váš celkový dojem z aplikace?”, která měla za úkol zjistit obecné působení aplikace na hráče na kytaru, měla všechny odezvy kladné. Mezi odpověďmi se objevilo například „Velmi mě překvapila (pozitivně).” nebo „Především se mi líbí u mého oblíbeného bluesového akordu E7#9 (který jsem dosud nevěděl, jak že se vlastně jmenuje) motiv Jimiho Hendixe v pozadí displeje... :) V tomto smyslu je program užitečný pro začínající hráče, případně pro samouky bez větší znalosti hudební teorie, a to z toho hlediska, že jim program umožní zápis hudby. ”.

Testeři měli také za úkol ohodnotit přesnost rozpoznávání akordů známkou 1 až 5 (1 = nejlepší, 5 = nejhorší). Toto hodnocení bylo poté rozděleno podle typu použité kytary. Průměrné hodnocení rozpoznávače je uvedeno v následující tabulce:

Tabulka 6.2: Průměrná uživatelská známka rozpoznávače

Typ kytary	Elektrická	Akustická
Průměrná známka	1,0	2,0

Kapitola 7

Závěr

7.1 Shrnutí

Cílem bakalářské práce bylo vytvořit aplikace, která bude v reálném čase zpracovávat signál, vytvářený elektrickou nebo akustickou kytarou. Tento signál bude analyzovat a bude detekovat a rozpoznávat akord, hraný uživatelem aplikace. Program by implementován v jazyce Java s grafickým uživatelským rozhraním, vytvořeným technologií JavaFX. Cílů této práce se podařilo úspěšně dosáhnout.

Podle uživatelského testování byla aplikace přijata kladně a plní svůj účel.

7.2 Další vývoj

Aplikace nabízí široké spektrum možných rozšíření. Pro zjednodušení si je rozdělme do dvou kategorií:

7.2.1 Blízká budoucnost

Prvním a nejaktuálnějším cílem je přepsat aplikaci na platformu Android, aby mohla být použita na mobilních telefonech a tabletech. Jelikož ale pro běžného uživatele bohužel neexistuje mnoho způsobů, jak na mobilní zařízení připojit elektrickou kytaru, bude androidová aplikace využívat pouze vestavěný mikrofon zařízení.

Jedno z možných jednoduchých rozšíření desktopové aplikace je jednouchá vizualizace nahraných dat. Tato vizualizace by mohla být rozdělena na dvě části. Zobrazení reálného signálu v několikamilisekundovém okně a kreslení frekvenčního spektra analyzovaného zvuku. Poslední jednodušší rozšíření je možnost nahrávání a ukládání hraného zvuku do formátu *.wav*.

7.2.2 Složitější cíle

Pokud zpracováváme kytarový signál, musí být kytara správně naladěna, nabízí se tedy do programu implementovat přesnou ladičku. Složitější cíl je vytvořit systém, který bude zprostředkovávat textový přepis hraných dat. Například každých x milisekund uloží jméno aktuálně hraného akordu, aby vznikla historie hraných akordů. To se hodí například pokud si uživatel skládá vlastní píseň a nechce v budoucnu zapomenout, jak ji hrál. Na to se dá

navázat automatickou tvorbou tabulatur, nicméně k dosažení tohoto cíle by bylo potřeba zpřesnit algoritmus natolik, aby dokázal rozpoznat různé varianty téhož akordu. Ultimátní rozšíření by pak byly efekty na elektrickou kytaru.

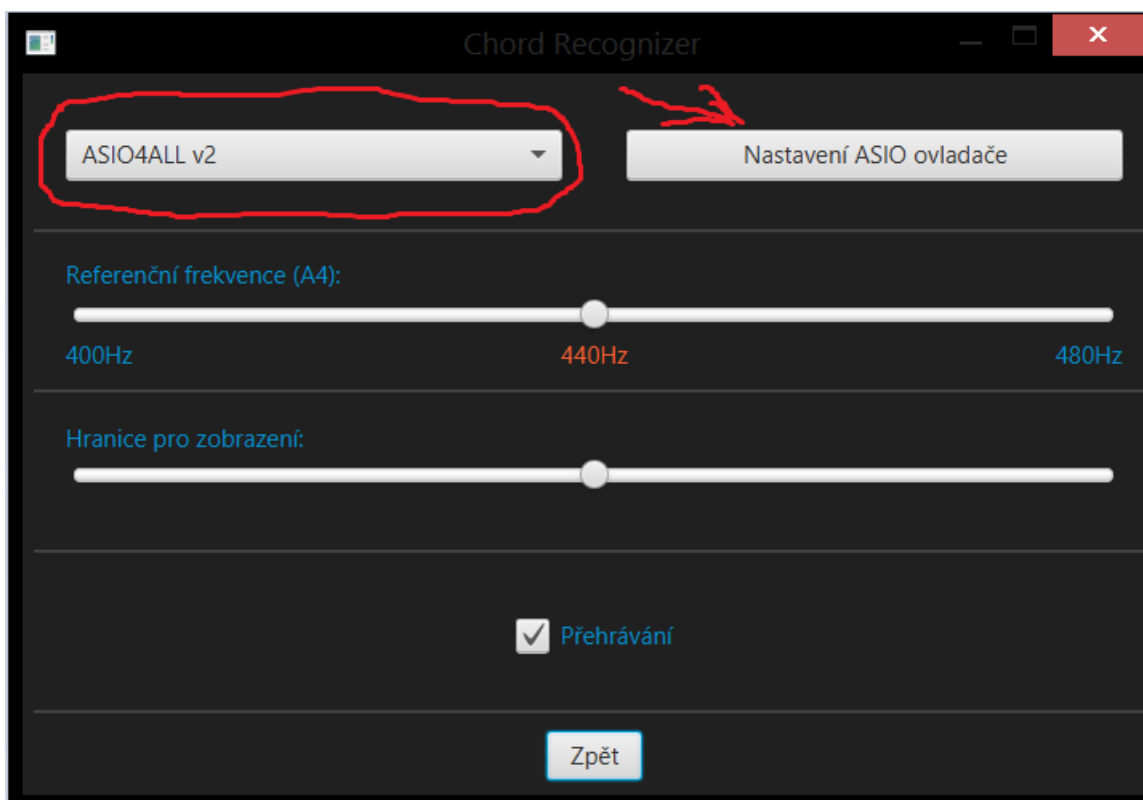
Literatura

- [1] Beauchamp, J. W.: *Analysis, synthesis, and perception of musical sounds the sound of music*. Springer, 2007, iISBN 038732576x.
- [2] Freeth, N.: *Kytara: rychle a bez problémů*. Slovart, 2009, iISBN 978-80-7391-160-7.
- [3] Rossing, T. D.: *The science of string instruments*. Springer, 2010, iISBN 1441971106.
- [4] Smith, J. O.: *Mathematics of the Discrete Fourier Transform (DFT): With Audio Applicaitons*. W3K Publishing, 2007, iISBN 978-0-9745607-4-8.
- [5] WWW stránky: JAsioHost. <https://github.com/mhroth/jasiohost>.
- [6] WWW stránky: Java. <https://www.java.com/>.
- [7] WWW stránky: JTransforms. <https://github.com/wendykierp/JTransforms>.
- [8] WWW stránky: ROCKSMITH 2014 EDITION.
<http://rocksmith.ubi.com/rocksmith/en-us/home/index.aspx>.
- [9] WWW stránky: Steinberg ASIO.
<https://www.steinberg.net/en/company/technologies.html>.
- [10] Zenkl, L.: *ABC hudební nauky*. Editio Bärenreiter Praha, 2003, iISBN 80-86385-21-3.

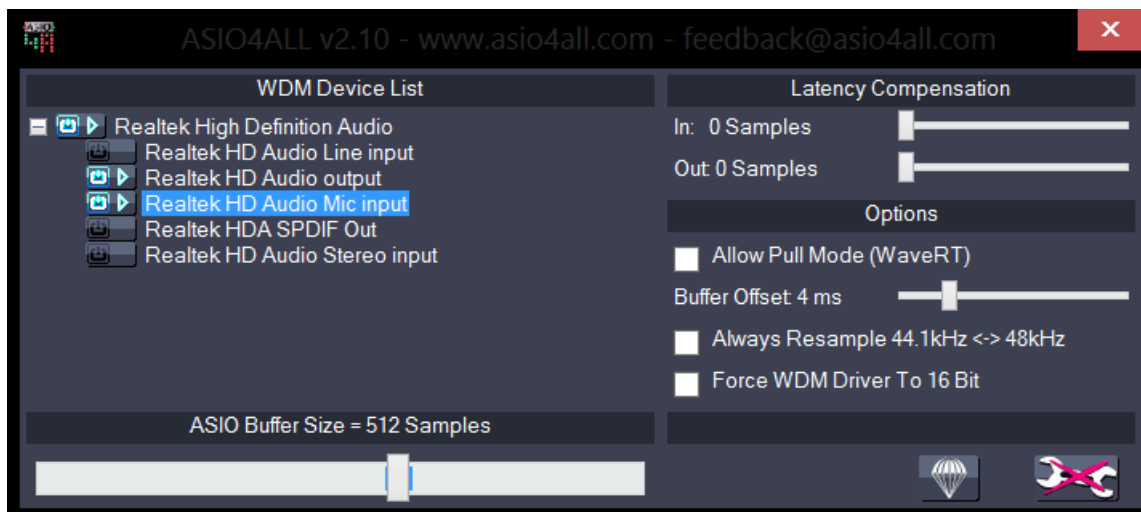
Příloha A

Návod k použití

1. Nainstalujte ASIO ovladač. (Soubor ASIO4ALL_2_12.exe)
2. Aplikace pro svůj běh potřebuje Java verze 8. Pokud nemáte, prosím nainstalujte <http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>
3. Spusťte program ChordRecognizer.jar
4. V nabídce nastavení vyberte (pokud není vybráný) ovladač ASIO4ALL v2 a klikněte na Nastavení ASIO ovladače.



5. V nastavení ovladače vyberte vstupní a výstupní zařízení (pro rozšířená nastavení klikněte na ikonu klíče) a případně restartujte aplikaci.



6. Naladte kytaru a můžete hrát.

Příloha B

Obsah CD

- *bin* - adresář s výslednou spustitelnou aplikací
- *dotaznik* - adresář s uživatelským dotzníkem
- *driver* - adresář, obsahující ASIO ovladač (Soubor ASIO4ALL_2.12.exe)
- *pdf* - adresář, obsahující tuto práci
- *src* - adresář se zdrojovými kódy aplikace
 - *netbeans* - adresář se projektem pro IDE Netbeans 8.0
- *tex* - adresář se zdrojovými kódy této práce (L^AT_EX)
- *tests*
 - *akordy* - adresář s nahranými testovacími akordy
 - *matice* - adresář s výslednými konfúzními maticemi ve formátech *.csv* a *.xlsx*
- *video* - adresář s krátkým videem, demonstrujícím práci