



**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**

---

**Fakulta elektrotechnická  
Katedra radioelektroniky**

**Klasifikace hudebních žánrů pomocí MFCC**

**MFCC Based Music Genre Classification**

Bakalářská práce

Studijní program: Komunikace, Multimédia a Elektronika  
Studijní obor: Multimediální technika

Vedoucí práce: Ing. František Rund, Ph.D.

**Jan Melechovský**

---

**Praha 2017**



## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Melechovský** Jméno: **Jan** Osobní číslo: **435031**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávající katedra/ústav: **Katedra radioelektroniky**  
Studijní program: **Komunikace, multimédia a elektronika**  
Studijní obor: **Multimediální technika**

## II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

**Klasifikace hudebních žánrů pomocí MFCC**

Název bakalářské práce anglicky:

**MFCC Based Music Genre Classification**

Pokyny pro vypracování:

Seznamte se s problematikou klasifikace obsahu a metodami pro automatickou klasifikaci hudebních žánrů. V prostředí Matlab implementujte metodu pro automatickou klasifikaci hudebních titulů podle žánru se zaměřením na parametr MFCC (Mel-Frequency Cepstral Coefficients). Ověřte úspěšnost klasifikace na dodaných vzorcích a výsledky porovnejte s výsledky práce [1]. Navrhněte možná vylepšení implementované metody.

Seznam doporučené literatury:

- [1] Bartoš, J., Detekce hudebních žánrů pro účely masteringu gramofonových desek, FEL ČVUT v Praze, 2016, Bakalářská práce  
[2] Lartillot, O., Toivainen, P., A Matlab Toolbox for Musical Feature Extraction From Audio, International Conference on Digital Audio Effects, Bordeaux, 2007

Jméno a pracoviště vedoucí(ho) bakalářské práce:

**Ing. František Rund Ph.D., katedra radioelektroniky FEL**

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **13.02.2017** Termín odevzdání bakalářské práce: **26.05.2017**

Platnost zadání bakalářské práce: **31.08.2018**

Podpis vedoucí(ho) práce

Podpis vedoucí(ho) ústavu/katedry

Podpis děkana(ky)

## III. PŘEVZETÍ ZADÁNÍ

Student bere na vědomí, že je povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

Datum převzetí zadání

Podpis studenta



## **Prohlášení**

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Dne 24. května 2017 v Praze

.....

## **Poděkování**

Děkuji svému vedoucímu práce panu Ing. Františku Rundovi, Ph.D. za cenné rady a přátelský přístup během vypracovávání této bakalářské práce.

Taktéž děkuji panu Georgi Tzanetakisovi za to, že mi umožnil použít jeho databázi hudebních žánrů v této práci. Děkuji i GZ Media, a.s. za poskytnutí bohaté sbírky hudebních ukázek, jíž bylo rovněž využito v této práci.

Na závěr děkuji svým blízkým za podporu během vypracovávání této práce i během celého studia.

## Abstrakt

Tato bakalářská práce se zabývá problematikou automatické klasifikace hudebních žánrů. Cílem je vytvořit algoritmus schopný třídit hudbu dle hudebních žánrů. Audio signál se nejdříve dělí na kratší časové úseky zvané segmenty a dále na kratší minisegmenty, ve kterých je spočteno několik koeficientů MFCC. Koeficienty jsou zpracovány pomocí statistických veličin aritmetického průměru a směrodatné odchylky. Klasifikaci realizuje klasifikátor k-NN. Algoritmus rozpoznává 8 žánrů: Classical, Electro, Folk, Hip Hop, Jazz, Metal, Pop, Reggae. Bylo dosaženo úspěšnosti 68,8 % v pomalé verzi a 67,2 % v rychlé verzi algoritmu.

**Klíčová slova:** klasifikace, hudba, žánr, parametrizace, segmentace, MFCC, k-NN, MATLAB, MIR Toolbox

## Abstract

This bachelor's thesis deals with automatic music genre classification. The goal is to create an algorithm able to distinguish music by music genres. At first, the audio signal is divided into smaller frames called segments and further into even smaller frames called minisegments. The MFCCs are taken from each minisegment. Using mean and standard deviation, the coefficients are integrated in a segment. We further use the k-NN algorithm to classify music genres. The following music genres are distinguished: Classical, Electro, Folk, Hip Hop, Jazz, Metal, Pop, Reggae. A total of 68,8 % success rate has been reached for the slower version and 67,2 % for the faster version of the algorithm.

**Key words:** classification, music, genre, feature extraction, segmentation, MFCC, k-NN, MATLAB, MIR Toolbox

# Obsah

Seznam tabulek .....	10
Seznam obrázků .....	10
Seznam příloh.....	10
Seznam použitých zkratek.....	11
Úvod.....	13
1 Hudební žánry a jejich klasifikace .....	15
2 Parametrizace .....	17
2.1 Segmentace.....	17
2.2 MFCC.....	18
2.3 Další spektrální parametry.....	19
2.3.1 Spectral centroid.....	19
2.3.2 Spectral Rolloff .....	20
2.3.3 Spectral Flatness.....	20
2.4 Statistické veličiny .....	20
2.4.1 Aritmetický průměr .....	20
2.4.2 Směrodatná odchylka .....	21
2.4.3 Koeficient šikmosti.....	21
2.4.4 Kovariance.....	21
3 Metody klasifikace .....	23
3.1 k-NN (k-Nearest Neighbors) .....	23
3.2 Distance-weighted k-NN (DWKNN).....	25
3.3 k-Means .....	26
3.4 SVM .....	26
4 Realizace .....	27
4.1 Návrh realizace.....	27
4.2 Algoritmus programu .....	28
4.2.1 Fáze trénování .....	28
4.2.2 Fáze testování .....	29
4.3 Test 1 - kombinace parametrů, vliv normování parametrů, 4 žánry.....	30
4.4 Test 2 - databáze GTZAN, 10 žánrů, vliv převzorkování .....	32
4.5 Test 3 - změna počtu filtrů $b$ .....	33
4.6 Test 4 - změna počtu použitých koeficientů.....	34
4.7 Test 5 - změna délky segmentu a minisegmentu.....	35
4.8 Test 6 - vlastní databáze, 8 žánrů, vliv trénovací množiny .....	36
4.9 Test 7 - změna počtu nejbližších sousedů $k$ .....	38



4.10	Test 8 - zrychlení algoritmu .....	40
5	Zhodnocení výsledků .....	43
5.1	Srovnání systémů .....	43
5.2	Návrh možných zlepšení, budoucí práce.....	44
5.3	Rozlišení nehudebních a hudebních zvuků .....	44
	Závěr .....	47
	Seznam použité literatury.....	48
	Přílohy.....	51
	Příloha 1 - CD .....	51
	Příloha 2 - Tabulky vlastní databáze, příkladové grafy .....	51

## Seznam tabulek

Tab. 1 Nastavení systému pro test 1 .....	31
Tab. 2 Test kombinace parametrů bez použití normování parametrů (15).....	31
Tab. 3 Test parametrů s použitím normování parametrů (15).....	31
Tab. 4 Nastavení systému pro test 2.....	32
Tab. 5 Úspěšnost klasifikace testu 2 - kombinace parametrů a vzorkovací kmitočet .....	33
Tab. 6 Nastavení systému pro test 3.....	34
Tab. 7 Úspěšnost klasifikace v závislosti na počtu použitých filtrů b pro $r=13$ .....	34
Tab. 8 Nastavení systému pro test 4.....	34
Tab. 9 Úspěšnost klasifikace v % pro 4 různá b v závislosti na r .....	35
Tab. 10 Nastavení systému pro test 5.....	35
Tab. 11 Úspěšnost klasifikace v % pro různé délky segmentů a minisegmentů .....	35
Tab. 12 Srovnání žánrů vlastní databáze s žánry použitými v [1].....	36
Tab. 13 Nastavení systému pro test 6.....	36
Tab. 14 Úspěšnost klasifikace, podtrénovaný systém .....	37
Tab. 15 Úspěšnost klasifikace po dotrénování vybranými žánry s nízkou úspěšností .....	37
Tab. 16 Úspěšnost klasifikace - konečná verze.....	38
Tab. 17 Nastavení pro test 7.....	39
Tab. 18 Úspěšnost klasifikace v závislosti na k .....	39
Tab. 19 Úspěšnost klasifikace pro $k=7$ .....	39
Tab. 20 Úspěšnost klasifikace pro $k=11$ .....	40
Tab. 21 Úspěšnost klasifikace rychlého systému .....	40
Tab. 22 Úspěšnost klasifikace rychlého systému pro $k=11$ .....	41
Tab. 23 Úspěšnost klasifikace systému [1] .....	43

## Seznam obrázků

Obr. 1 Segmentace nahrávky.....	17
Obr. 2 Spektrum signálu do 8 kHz rozdělené bankou 10 Mel-frekvenčních filtrů (převzato z [23]) .....	19
Obr. 3 Algoritmus k-NN .....	24
Obr. 4 Rozhodování DAG SVM (převzato z [21]) .....	26
Obr. 5 Algoritmus programu .....	28
Obr. 6 Grafický výstup algoritmu .....	30

## Seznam příloh

Tab. přílohy 1 Nahrávky použité pro dotrénování systému .....	51
Tab. přílohy 2 Testovací množina - Classical, Electro.....	52
Tab. přílohy 3 Testovací množina - Folk, Hip Hop .....	52
Tab. přílohy 4 Testovací množina - Jazz, Metal .....	53
Tab. přílohy 5 Testovací množina - Pop, Reggae .....	54
Obr. přílohy 1 Grafické vyjádření klasifikace nahrávky žánru Metal - jednoznačná klasifikace	55
Obr. přílohy 2 Grafické vyjádření klasifikace nahrávky žánru Pop - nejednoznačná klasifikace	55
Obr. přílohy 3 Grafické vyjádření klasifikace nahrávky žánru Blues - neznámý žánr.....	56

## **Seznam použitých zkratek**

MFCC - Mel-Frequency Cepstral Coefficients

k-NN - k-Nearest Neighbors

ANN - Artificial Neural Networks

SVM - Support Vector Machine

ELM - Extreme Learning Machine

DWKNN - Distance-weighted k-Nearest Neighbors

M - aritmetický průměr (Mean)

D - směrodatná odchylka (Standard Deviation)

S - koeficient šikmost (Skewness)

FFT - Fast Fourier Transform

DCT - Discrete Cosine Transform

MIR - Music Information Retrieval

CD - Compact Disc

GUI - Graphical User Interface

.wav - Waveform audio file format

.mp3 - MPEG-2 Audio Layer III

.au - Audio file format



## Úvod

Cílem této práce je vytvořit algoritmus schopný klasifikace hudebních žánrů založený na použití Mel-Frequency Cepstral Coefficients (MFCC) jako parametru.

S neustále se rozmáhajícími možnostmi využití internetu roste i počet různých databází hudby. Zároveň se objevuje stále více interpretů a populární je dnes i hudbu kupovat online jako soubory mp3. Ne všechny skladby musí být nutně označeny svým žánrem. V takovýchto případech by nebylo od věci mít program, který dokáže jednotlivé hudební žánry roztrždit, aby se uživatel nebo zákazník lépe zorientoval.

Algoritmus pro klasifikaci hudebních žánrů může být využit např. ve výrobě vinylových desek, kde je velmi důležité provést správný mastering, aby mohla být deska vyřezána. Obsahuje-li hudba nepřírozeň skoky, může dojít k poškození desky, v horším případě k poškození řezacího zařízení. Algoritmus může pomoci při třídění nahrávek od zákazníků mezi žánrově specializované odborníky masteringu. Touto problematikou se zabývá práce [1], s jejíž výsledky dle zadání porovnáme výsledky naše.

Zároveň může být algoritmus použit pro účely výstupní kontroly, kde každý žánr může mít nastaven specifické limity výskytu různých chyb. Např. v žánru metal nebude vadit větší počet lupanců nebo větší intenzita lupanců tak, jako by vadila třeba v žánru vážné hudby nebo country.

Pro jakoukoliv klasifikaci audio signálů, ať už se jedná o detekci nástrojů, zpěvných ptáků nebo hudebních žánrů, je obecný postup stejný. Nejprve se signály dělí na kratší časové úseky zvané segmenty. Ty je možno dělit na ještě kratší úseky, tzv. minisegmenty. Proces segmentace napomáhá důkladnějšímu popisu signálu, tedy parametrizaci. Při procesu parametrizace se minisegmenty nebo segmenty popisují jedním nebo více vhodnými parametry. Třetím krokem je již parametrizované segmenty klasifikovat, tj. řadit do vybraných tříd (žánry, nástroje apod.). K tomu slouží řada algoritmů strojového učení.

Nejprve se rozebere, co se rozumí pod pojmem hudební žánr. V druhé kapitole se zaměříme na parametrizaci a MFCC. Třetí kapitola se týká metod klasifikace, kde zmíníme metodu k-NN a jiné. Následně se budeme věnovat realizaci algoritmu pro klasifikaci hudebních žánrů a testům spojeným s realizací. V poslední kapitole zhodnotíme výsledky a srovnáme s výsledky práce [1].



# 1 Hudební žánry a jejich klasifikace

S historickým vývojem lidstva se vyvíjí i hudba a s vývojem hudby se vyvíjí, rozšiřují, objevují nové hudební směry. Pod pojmem hudební žánry rozumíme definovatelné rozdělení hudby podle jejich formálních (struktura, použité nástroje) a obsahových charakteristik [15]. V [20] uvádí definovatelnost žánrů podle hudebních technik, stylu, kulturního kontextu a zároveň jsou zmíněny tři proměnné pro dělení hudby: arousal, valence, depth. Arousal (vzrušení) reprezentuje energii v hudbě, valence (mocenství) představuje rozpětí mezi smutnými a šťastnými emocemi a depth (hloubka) představuje hloubku těchto emocí.

Základních hudebních žánrů je podle [15] sedm, jsou to: lidová hudba, vážná hudba, jazz, rock, pop, metal, electronica. Dělení na žánry je však nejednotné a liší si literaturu od literatury, občas dokonce i akademii od akademie. V [18] zmiňují možnost doplnění sedmi zmíněných o swing, rhythm and blues, country, folk, šanson a reggae. V [1, 16, 19] je žánr metalu začleněn do žánru rock a v [16, 19] přidávají žánr hip hop/rap.

Dále existuje pojem hudebních podžánrů a hudebních stylů [15, 16, 20]. Dělení dle hudebních stylů je rozdílné, zpravidla jsou styly podřazené žánrům. Zdroj [20] uvádí, že v současnosti existuje více jak 1200 hudebních stylů. Tématem této práce však není vyřešit problematiku hudebních žánrů a stylů, nadále budeme používat termínu "hudební žánr".

V této práci použijeme tři různých databází a pro každou bude výběr žánrů trochu odlišný. V prvním kroku to bude čtveřice: Metal, Reggae, Dance a Classical. Při práci s databází GTZAN [27] použijeme všech 10 žánrů, a to: Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae, Rock. Nakonec sestavíme vlastní databázi, abychom výsledky mohli porovnat s výsledky práce [1]. Bude obsahovat těchto 8 žánrů: Classical, Electro, Folk, Hip Hop, Jazz, Metal, Pop, Reggae. Pro přiblížení si uveďme základní rysy některých žánrů.

- Blues vznikl v afro-americké komunitě. Obsahem je většinou smutný, depresivní [34]. Často poznáme blues typickými bluesovými stupnicemi nebo "bluesovou dvanáctkou" (předpis opakujících se taktů a tóniny).
- Classical označuje vážnou hudbu. Ta je typická svým dynamickým rozsahem a množstvím použitých nástrojů.
- Dance označuje taneční hudbu z žánru electronica.
- Electro je zkrácený název pro electronica. Jde především o uměle (počítačově) vytvořenou hudbu.
- Folk může označovat lidovou hudbu a písničkáře. Pro lidovou hudbu se u nás více vžilo označení folklór [35]. Folkem se spíše rozumí moderní folk, a podobně to bude i v této práci (databáze Folku použita v této práci obsahuje pouze 1 lidovou píseň). Především se jedná o klidnější projev bez bicích (moderní folk může bicí obsahovat), s kytarovým doprovodem.
- Hip hop je typický použitím rapu a samplů, což jsou smyčky melodií, často i známých, třeba z vážné hudby.
- Jazz vznikl z Blues míšením s dalšími žánry a typické je použití dechových nástrojů nebo piana.
- Metalem jsou míněny všechny metalové podžánry: heavy, thrash, doom, death, metalcore, black atd. Metal je typický zkreslenými kytarami, důraznými bicími a u některých podžánrů i technikami zpěvu, jako jsou growl nebo scream.
- Pop je velice rozmanitý žánr, jelikož vlastně označuje, co je v danou dobu "populární". Může být míchaný s žánrem Electro i s žánrem Folk, či Rock.

→ Reggae je typické houpavým rytmem a projevem. Žánr pochází z Jamajky, tudíž můžeme i často slyšet ve skladbách typickou jamajskou angličtinu.

Hudební žánry se však často míchají dohromady za účelem vzniku nového žánru nebo zkrátka interpret tvrdí, že hraje ten i ten žánr. Nemožnost správně definovat hranice jednotlivých žánrů potvrzuje i experiment zmíněný v literatuře [29]. Vysokoškolští studenti měli poznávat jednotlivé hudební žánry z výběru 10 žánrů. Nejprve jim bylo pouštěno několik ukázek každého žánru a poté měli hádat pro úseky dlouhé 250 ms a poté pro úseky dlouhé 3 s, o jaký žánr se jedná. Pro 250 ms dosáhli úspěšnosti pouze 53 % a pro úseky dlouhé 3 s dosáhli 70 %. Při delších úsecích hudby už úspěšnost nestoupala. Úspěšnosti automatické klasifikace se od tohoto čísla tolik neliší. Fakt, že ani lidé nedokáží správně zařadit jednotlivé žánry, poukazuje na skutečnost míšení žánrů a jejich nejednoznačnost, neomezenost.

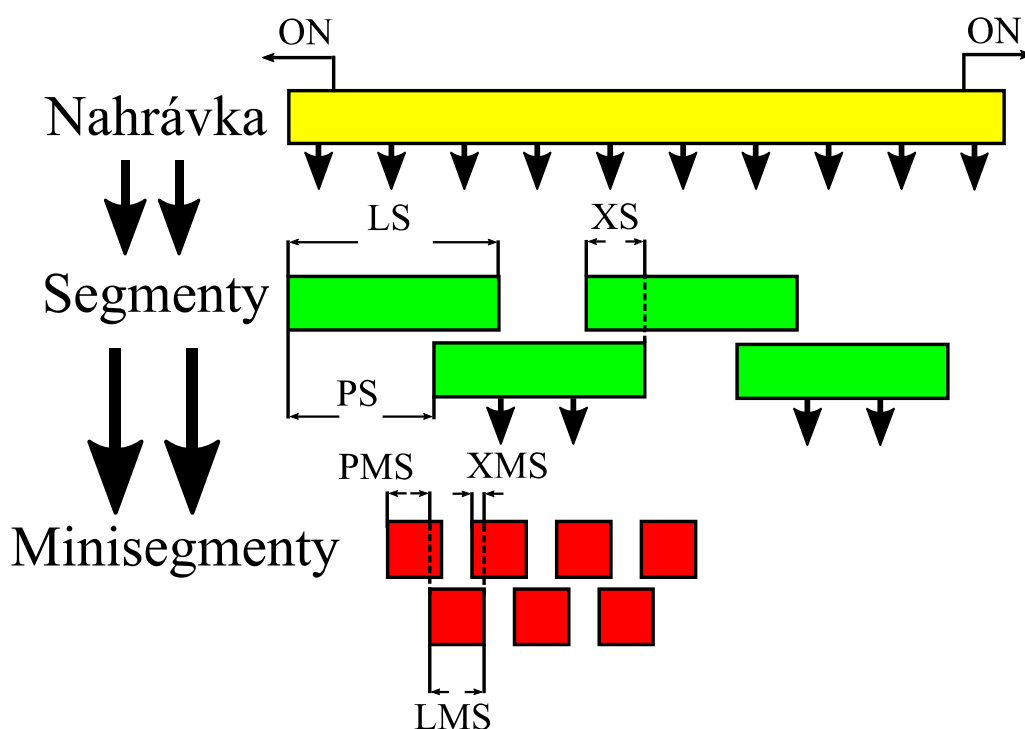


## 2 Parametrizace

Parametrizace je proces, který popíše signály z hlediska vybraných parametrů. Tento krok je pro veškeré rozpoznávací algoritmy z oblasti zpracování signálů zásadní. Parametrů existují desítky, ale ne všechny se hodí pro popis hudby. V této kapitole si nejdříve představíme proces segmentace, dále se podíváme na několik vybraných parametrů. Zaměříme se na parametr MFCC, který je v oblasti MIR (Music Information Retrieval) poměrně rozšířený, přestože hlavní využití má spíše v oblasti zpracování řeči.

### 2.1 Segmentace

Pro lepší parametrizaci je vhodné signál nejdříve rozdělit na kratší časové úseky, ve kterých se parametry vyhodnocují. Tomuto procesu říkáme segmentace a jednotlivé časové úseky označíme jako segmenty. Jejich délka se může pohybovat v jednotkách vteřin. V [13] zmiňují, že pro jednoduché parametry není vhodné dělat segmenty kratší než 2,4 s. Délka segmentů by však neměla být příliš velká, protože segment má zachytit nějakou část signálu s pokud možno neměnnými či málo měnícími se parametry. Pro ještě důkladnější popis můžeme segmenty dělit na ještě menší úseky - minisegmenty [1, 2, 14, 17, 21] s typickou délkou v rozmezí 10 až 100 ms. K znázornění segmentace poslouží Obr. 1.



Obr. 1 Segmentace nahrávky

Na Obr. 1 vidíme ořezání nahrávky ON o část začátku a konce, kde předpokládáme výskyt hluchých míst, netypických inter či outer. Nahrávka se dále dělí na segmenty o délce LS, které se překrývají o XS, z čehož plyne posun segmentu PS. Segmenty se dělí na minisegmenty o délce LMS, překryvu XMS a posuvu PMS.

V literatuře [17, 21, 29] zmiňují, že nejlépe pro detekci hudebních žánrů vychází spektrální parametry, za nimi zaostávají parametry harmonické. Parametry časové (rytmické a dynamické) nedosahují zdaleka tak dobrých výsledků. Proto bude v této práci využito spektrálního parametru MFCC.

## 2.2 MFCC

Mel-Frequency Cepstral Coefficients (dále MFCC) popisují spektrum signálu za použití Mel-frekvenční škály. Tento parametr je velmi často používán v oblastech rozpoznávání řeči. Výpočet [1, 5] probíhá v následujících krocích:

1. Dělení signálu na krátké úseky (20 až 40 ms, standard 25 ms) [23], ořezání krajů okénkem (Hamming)
2. Fourierova transformace signálu
3. Rozdělení spektra bankou 40 trojúhelníkových filtrů, přepočítání na Mel-frekvenci (Obr. 1)
4. Absolutní hodnota, logaritmus
5. Diskrétní kosinová transformace

Mel-frekvence je škála frekvence, která lépe popisuje skutečné lidské vnímání výšky tónů než normální škála frekvence. Zhruba do 1 kHz je závislost lineární, nad 1 kHz logaritmická [4]. Většinou se v 2. kroku použije filtrů, které jsou již přepočtené na Mel-frekvenci, tedy samotné spektrum se nepřepočítává. Na Obr. 2 je vidět, že šířka filtrů (ačkoliv jich je znázorněno pouze 10) je do 1 kHz zhruba stejná, ale od vyšší frekvence šířka roste, což odpovídá popisu Mel-frekvence. Používají se následující dva přepočtení vztahy [23]:

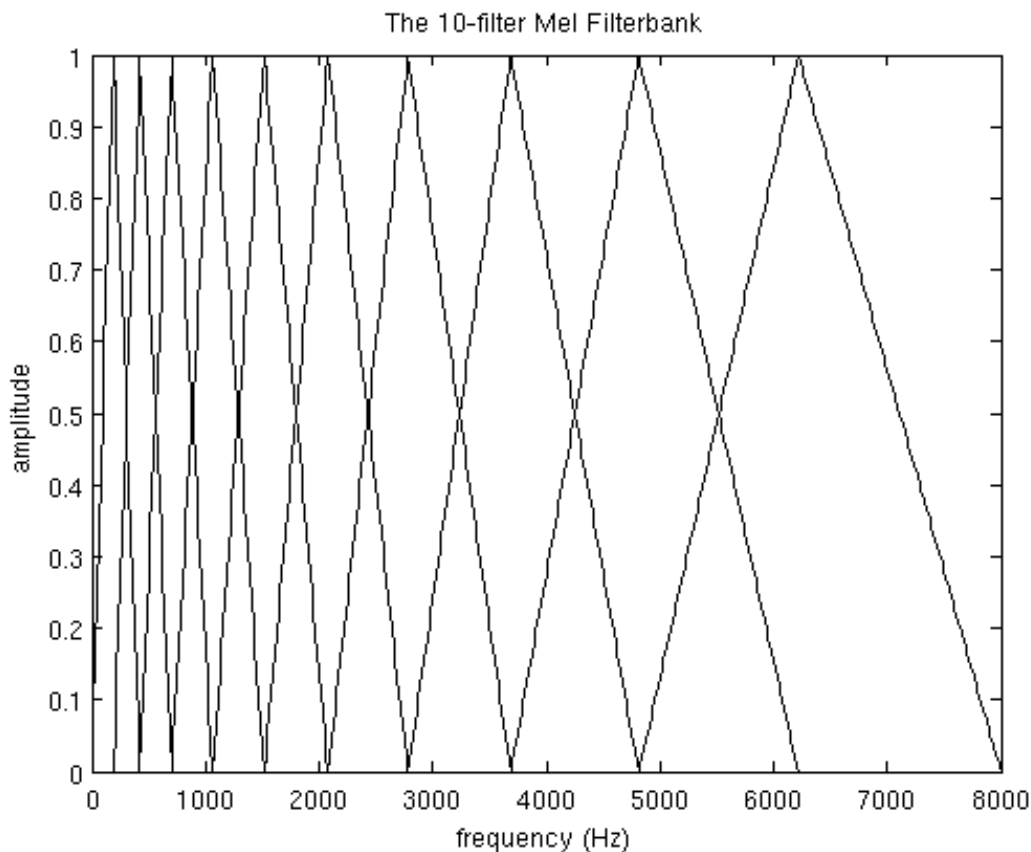
$$Mel(f) = 2595 \cdot \lg\left(1 + \frac{f}{700}\right), \quad (1)$$

$$Mel(f) = 1125 \cdot \ln\left(1 + \frac{f}{700}\right). \quad (2)$$

Vztahy jsou téměř ekvivalentní, dávají velmi podobné hodnoty. Pro zpětný přepočítání z Mel-frekvence na frekvenci můžeme použít vzorců [23]:

$$Freq(m) = 700 \cdot 10^{\frac{m}{2595}-1}, \quad (3)$$

$$Freq(m) = 700 \cdot e^{\frac{m}{1125}-1}. \quad (4)$$



Obr. 2 Spektrum signálu do 8 kHz rozdělené bankou 10 Mel-frekvenčních filtrů (převzato z [23])

V článku [6] se zaměřují na použitelnost MFCC pro rozpoznávání hudby v porovnání s rozpoznáváním řeči, zda-li je Mel-frekvenční škála vhodná i pro hudbu a zda-li by pro hudbu nevyhovovalo namísto DCT transformace použít KL transformaci. Zmíněné se v [6] neprokázalo, MFCC se dle [6, 21] hodí i pro hudbu.

## 2.3 Další spektrální parametry

Přestože v této práci využijeme MFCC, uvedeme si i tři jiné spektrální parametry, kterých využívají v práci [17].

### 2.3.1 Spectral centroid

Spektrální centroid je definován jako těžiště spektra. Je to aritmetický průměr frekvencí ve spektru násobených jejich příslušnou amplitudou [29]. K výpočtu používáme vzorec

$$C_t = \frac{\sum_{n=1}^N M_t[n] \cdot n}{\sum_{n=1}^N M_t[n]}, \quad (5)$$

kde  $C_t$  značí hodnotu spektrálního centroidu v čase  $t$ ,  $M_t[n]$  je amplituda ve frekvenčním intervalu  $n$ ,  $N$  je počet vzorků FFT.

### 2.3.2 Spectral Rolloff

Spectral rolloff značí hodnotu frekvence, pod kterou se nachází určité množství frekvence. V [29] používají 85 %. K výpočtu pro 85 % slouží vzorec

$$\sum_{n=1}^{R_t} M_t[n] = 0,85 \cdot \sum_{n=1}^N M_t[n], \quad (6)$$

kde  $R_t$  značí hodnotu spectral rolloff,  $M_t[n]$  je amplituda ve frekvenčním intervalu  $n$ ,  $N$  je počet vzorků FFT.

### 2.3.3 Spectral Flatness

Parametr spectral flatness popisuje tvar spektra. Spočte se jako geometrický průměr spektra dělený aritmetickým průměrem. Spectral flatness spočteme vzorcem

$$F_t = \frac{\sqrt[N]{\prod_{n=1}^N M_t[n]}}{\frac{\sum_{n=1}^N M_t[n]}{N}}, \quad (7)$$

kde  $F_t$  značí hodnotu spectral flatness,  $M_t[n]$  je amplituda ve frekvenčním intervalu  $n$ ,  $N$  je počet vzorků FFT.

Rovné spektrum (bílý šum) by mělo mít hodnotu spectral flatness blízké 1 zdola. Naopak nerovné až "špičaté" spektrum může dosahovat hodnoty spectral flatness blízké 0 shora. Pro výpočet spectral flatness je potřeba, aby byly všechny hodnoty nenulové.

## 2.4 Statistické veličiny

Popis pouze pomocí MFCC (nebo jiných parametrů) vypočtených v krátkých časových úsecích (20 až 40 ms), odpovídajících délkou námi označeným minisegmentům by mohl být nedostačující. Určitě by se vyskytla řada minisegmentů, která by klasifikaci zhoršovala, např. krátké ticho, pazvuk, lupanec na vinylové desce apod. I proto je vhodné signál dělit na delší segmenty složené z několika minisegmentů. K popisu celých segmentů můžeme využít dále zmíněných statistických veličin. V literatuře [17] používají k parametrizaci aritmetického průměru, směrodatné odchylky, koeficientu šikmosti a koeficient špičatosti, všechny aplikované na spektrální parametry. Při použití průměru a odchylky dosahují úspěšnosti 79,7 %, s přidáním koeficientu šikmosti pak 83,3 % a nakonec po přidání koeficientu špičatosti dosáhnou úspěšnosti 85,6 %.

### 2.4.1 Aritmetický průměr

Velmi oblíbený a užívaný aritmetický průměr je definován jako suma  $N$  hodnot  $X_n$ , dělená počtem hodnot  $N$ . Dle [17] jsou aritmetický průměr a směrodatná odchylka nejpoužívanějšími integračními metodami pro kratší segmenty (10 až 100 ms). Samotný aritmetický průměr může být velmi zkreslující veličinou. Nedává žádnou informaci o proměnlivosti průměrované náhodné veličiny. Aritmetický průměr spočteme známým vztahem

$$\mu = \frac{1}{N} \sum_{n=1}^N X_n. \quad (8)$$

#### 2.4.2 Směrodatná odchylka

Směrodatná odchylka je kvadratickým průměrem odchylek hodnot znaku od jejich aritmetického průměru [8]. Narozdíl od aritmetického průměru nám tato statistická veličina popisuje proměnlivost náhodné veličiny. V našem případě se jedná o proměnlivost v čase. Větší odchylku MFCC v segmentu budeme očekávat u žánrů rychlejších, např. metal nebo electro, naopak malé odchylky mohou mít žánry klidnější, např. classical nebo folk. Směrodatnou odchylku spočteme vztahem

$$\delta = \sqrt{\frac{1}{N} \sum_{n=1}^N (X_n - \mu)^2}. \quad (9)$$

#### 2.4.3 Koefficient šikmosti

Koefficient šikmosti je charakteristika rozdělení náhodné veličiny, která popisuje jeho nesymetrii [9]. Nulová hodnota šikmosti znamená, že jsou hodnoty náhodné veličiny rovnoměrně rozděleny vlevo a vpravo od střední hodnoty. Záporná hodnota šikmosti značí, že se odlehlejší hodnoty nachází vlevo od střední hodnoty a většina hodnot bude blízko střední hodnoty zprava. Kladná hodnota značí opak. Koefficient šikmosti spočteme vztahem

$$\gamma = \frac{\sum_{n=1}^N (X_n - \mu)^3}{(N - 1)\delta^3}. \quad (10)$$

#### 2.4.4 Kovariance

Další možností zpracování koeficientů je popis pomocí kovariance, která představuje míru spolupráce dvou náhodných veličin. V literatuře [4] používají pro popis signálu společně s průměrem koeficientů kovarianční matici. Výpočetní vztah je

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])], \quad (11)$$

kde  $E$  představuje střední hodnotu,  $X$  a  $Y$  jsou porovnávané hodnoty.



### 3 Metody klasifikace

Nezbytným krokem pro klasifikaci hudebních žánrů bude zajisté samotná klasifikace. Jedná se o proces, který porovnává nasbíraná data mezi sebou za účelem rozdělení dat do vybraných tříd. Třídy mohou být předem definované, ale nemusí (hledáme jen souvislosti - viz Metody bez učitele). V našem případě budou třídy představovat vybrané hudební žánry. Klasifikace dat souvisí se strojovým učením. V této kapitole si uvedeme několik klasifikačních metod.

Metody strojového učení dělíme především na [14, 17, 32]:

#### Metody s učitelem (supervised learning)

Algoritmu jsou poskytnuta vzorová data jednotlivých tříd. Tato fáze se nazývá trénování a vzorová data nazýváme jako **trénovací množinu**. Vstupem je vždy sada parametrů (vektor příznaků), kterou se snažíme přiřadit ke správnému výstupu reprezentovanému nějakou třídou. Typickým představitelem je algoritmus k-NN.

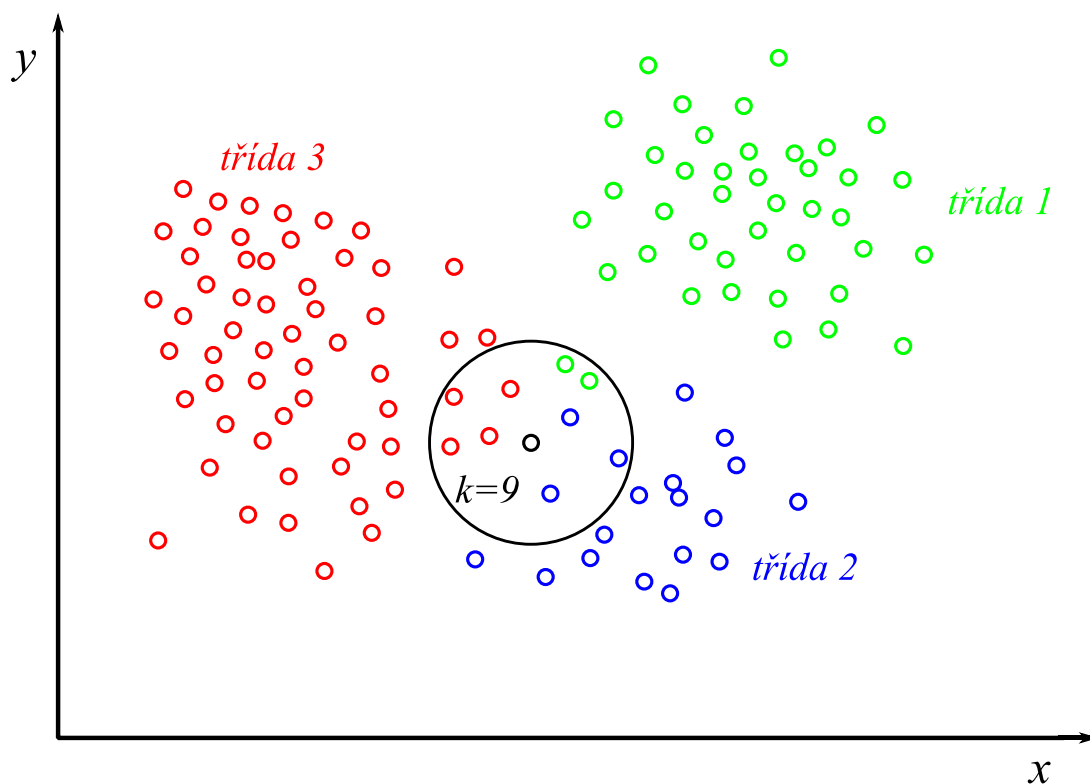
#### Metody bez učitele (unsupervised learning)

Tato metoda si vytváří třídy sama na základě podobnosti vstupních dat. Cílem nemusí být přiřazování vstupů k výstupům, ale často jde o poznání skrytých souvislostí mezi daty. Typickým představitelem je k-Means.

#### 3.1 k-NN (k-Nearest Neighbors)

Metoda k-NN je klasifikační metodou s učitelem, jenž spočívá v hledání  $k$  nejbližších bodů v  $n$ -dimenzionálním prostoru, kde  $n$  je počet posuzovaných parametrů [12]. Nejprve se nechá parametrizovat tzv. trénovací množina, což je množina dat, o kterých poskytneme klasifikátoru informaci, do které třídy patří, tj. v našem případě jakého žánru jsou. V  $n$ -dimenzionálním prostoru se vytvoří body příslušící parametrizovaným úsekům trénovací množiny, v našem případě jednotlivým parametrizovaným segmentům nahrávek. Testovací množina obsahuje data, která klasifikujeme. Jednotlivé parametrizované úseky testovací množiny se vnesou do  $n$ -dimenzionálního prostoru s body trénovací množiny a počítá se vzdálenost bodu testovací množiny od všech segmentů trénovací množiny. Ze získaných rozdílů pro každý jednotlivý bod testovací množiny se poté vybere  $k$  nejmenších rozdílů spolu s informací, od jakých bodů trénovací množiny tyto rozdíly vznikly. Jelikož je pro každý bod trénovací množiny známa jeho třída, výsledkem bude počet nejbližších sousedů z  $k$  ke každé třídě.

Klasifikace je znázorněna na Obr. 3. V prostoru k-NN se nachází body trénovací množiny, která má v tomto případě pouze 3 třídy. Testovaný bod (černý) je zanesen do prostoru a do klasifikace je zahrnuto  $k=9$  nejbližších bodů, tj. 4 červené, 3 modré a 2 zelené. V tomto případě bude testovaný bod zařazen do červené třídy. Na obrázku je vidět i nedostatek algoritmu k-NN. Zelené body, které ovlivňují klasifikaci, jsou od ostatních bodů své třídy poněkud vzdálené. V případě hudebních žánrů by mohly vzniknout parametrizováním netypických úseků pro daný žánr. Problém těchto "náhodných sousedů" by jiné, rafinovanější klasifikační algoritmy, které pracují "shlukově", mohly potlačit, např. k-Means nebo SVM.



Obr. 3 Algoritmus  $k$ -NN

K hledání nejbližších sousedů lze použít [10] několik metrik: Euklidovskou metriku

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (12)$$

Manhattanovu metriku

$$d = \sum_{i=1}^n |x_i - y_i|, \quad (13)$$

nebo Minkowskiho metriku

$$d = \left( \sum_{i=1}^n |x_i - y_i|^q \right)^{1/q}, \quad (14)$$

kde  $d$  je vzdálenost,  $n$  je počet parametrů,  $x_i$  jsou parametry trénovací množiny a  $y_i$  jsou parametry testovací množiny a  $q$  je konstanta, jejíž hodnota se volí mezi 1 a 2. Vztah (14) je v případě  $q=1$  roven vztahu (13) a v případě  $q=2$  je roven vztahu (12). Klasifikátor  $k$ -NN použitý v této práci používá metriku Euklidovskou (12).



V [7] se autor zmiňuje o problému Euklidovské metriky k-NN při použití dvou a více parametrů. Hodnoty jednotlivých parametrů se mohou číselně velmi lišit (např. parametr energie a parametr počet průchodů nulou) a je potřeba je normovat vztahem

$$\hat{X}_j = \frac{X_j - \mu}{\delta}, \quad (15)$$

kde  $X_j$  je hodnota parametru v j-tém bodě (segmentu),  $\mu$  je aritmetický průměr (8) parametru získaný ze všech bodů (segmentů) a  $\delta$  je směrodatná odchylka (9).

Při použití k-NN je možno vynechat aritmetický průměr  $\mu$ , protože se při počítání rozdílu mezi sousedy odečte, jak je vidět ve vztahu

$$d_{ji} = \hat{X}_{jtrain} - \hat{X}_{itest} = \frac{X_{jtrain} - \mu}{\delta} - \frac{X_{itest} - \mu}{\delta}, \quad (16)$$

kde  $d_{ji}$  je rozdíl j-tého trénovacího od i-tého testovacího bodu. Dostáváme vztah pro výsledný rozdíl

$$d_{ji} = \frac{X_{jtrain} - X_{itest}}{\delta}. \quad (17)$$

Lze tedy pouze dělit rozdíly od ostatních bodů v prostoru směrodatnou odchylkou  $\delta$ .

### 3.2 Distance-weighted k-NN (DWKNN)

Jak již název napovídá, DWKNN je variantou k-NN, která se liší tím, že nejbližším sousedům při klasifikaci dává váhu dle jejich vzdálenosti od testovaného bodu [24]. Váhu vypočteme vztahem

$$w_j = \frac{d_k - d_j}{d_k - d_1}, \quad (18)$$

kde  $w_j$  značí váhu j-tého bodu,  $k$  značí počet sousedů,  $d_k$  je vzdálenost posledního ze sousedů (je maximální),  $d_j$  je vzdálenost j-tého bodu.

Výhoda DWKNN je ta, že nezáleží až tolik na počtu sousedů, často se ke klasifikaci použije celé trénovací množiny, tj.  $k$  je maximální možné [26]. V takovém případě je ale více než vhodné mít pro každou třídu stejný počet bodů.

Při použití podobného počtu sousedů jako v k-NN by mohlo DWKNN dosahovat lepších výsledků např. v případech, kdy většina sousedů při klasifikaci je z nesprávné třídy a menšina ze správné, avšak vzdálenost správných bodů bude mnohem menší než těch nesprávných. Přidáním váhy dle vzdálenosti je pravděpodobné, že dojde ke klasifikaci do správné třídy. Váhování může mít ale i negativní vliv v případě "náhodných" bodů nesprávné třídy, tj. v trénovací množině je nahrávka nebo část nahrávky, která není čistě typická pro daný žánr. V [25] dosáhli v úloze klasifikace hudby podle emocí nejvyšší úspěšnosti použitím DWKNN (96,7 %), následované SVM (90 %) a k-NN (88%).

### 3.3 k-Means

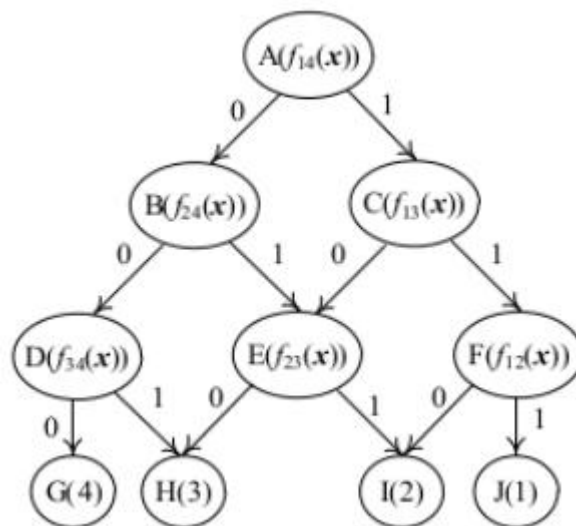
K-Means je metodou strojového učení bez učitele. Algoritmus pracuje na základě shluků dat. Volbou  $k$  se určí počet shluků (clusterů) a v počáteční fázi se (většinou jako náhodné body) vyberou centroidy jednotlivých shluků. Algoritmus dále iterativně přepočítává, zda je daný centroid těžištěm blízkého shluku a na základě výpočtu upraví pozici centroidu. To se opakuje, dokud se pozice centroidů neustálí. Výstupem algoritmu je jak pozice jednotlivých centroidů, tak i informace o příslušnosti všech ostatních bodů jednotlivým shlukům. Většinou se volí  $k$  rovno počtu rozlišovaných tříd, ale je možno zvolit  $k$  větší a mít více shluků pro méně tříd. [30, 31]

### 3.4 SVM

Support Vector Machine [11] je metodou strojového učení pro klasifikaci. Mezi daty se hledá nadrovina, která data rozdělí na dvě skupiny, přičemž nadrovina musí mít co největší vzdálenost od obou stran, tedy nejbližší bod z jedné či druhé skupiny by měl být co nejdále.

Z definice plyne, že SVM je metoda klasifikace pouze dvou kategorií, lze ji však využít i pro více kategorií [3, 21]. K tomu se dá použít např. DAG (directed acyclic graph) [21], viz. Obr. 3.2. Obvykle dosahuje pro klasifikaci žánrů lepších výsledků než k-NN [3, 4, 21].

V [21] porovnávají metody k-NN, k-Means, SVM a ANN. První dvě dosáhnou úspěšnosti 80 %, DAG SVM 87 % a ANN 96 %. Literatura [17] uvádí, že SVM dosahovalo o 0,5 % větší úspěšnosti než ELM (Extreme learning machine), což je typ feedforward ANN [22]. S použitím statistických veličin (viz kapitola 2.4) a spektrálních parametrů dosáhli s klasifikátorem SVM úspěšnosti 85,6 %. Literatura [3] uvádí, že SVM je přesný klasifikátor, avšak s pomalejším trénováním.



Obr. 4 Rozhodování DAG SVM (převzato z [21])

## 4 Realizace

V této kapitole si nejdříve představíme návrh realizace algoritmu pro klasifikaci hudebních žánrů. Dále si přiblížíme funkci algoritmu, který má trénovací a testovací fázi. Následuje řada testů, která vedla ke konečnému nastavení algoritmu. Změnou nastavení (počet koeficientů MFCC, počet filtrů v MFCC, počet sousedů v k-NN atd.) ladíme algoritmus, aby dosáhl co největší úspěšnost klasifikace.

### 4.1 Návrh realizace

K realizaci algoritmu pro klasifikaci hudebních žánrů použijeme prostředí MATLAB R2014b obohacené o Music Information Retrieval Toolbox 1.6.1 (MIR Toolbox) [2], který nabízí velké množství funkcí pro zpracování signálů. Signály budeme parametrizovat v segmentech. Ty rozdělíme na minisegmenty, pro které se spočte několik MFCC. Koeficienty dále zpracujeme pomocí statistických veličin (viz kap. 2.4). Ke klasifikaci použijeme klasifikátoru k-NN především kvůli jeho jednoduchosti implementace a zřetelnější funkčnosti. Práci si ulehčíme tím, že použijeme algoritmus z práce [14], který přetvoříme pro účely klasifikace hudebních žánrů.

Pro vytvoření programu pro klasifikaci hudebních žánrů a všechny testy s tím spojené použijeme 3 různé databáze hudby. Každá je něčím specifická. První je databáze od GZ Media, která sestává ze 188 stop o průměrné délce 15 minut reprezentujících okolo 20 žánrů. Několik vybraných žánrů bude manuálně sestříháno a použito pro první test. Druhou databází je GTZAN MARSYAS music genre database [27]. Ta se skládá z deseti žánrů, každý reprezentovaný 100 úryvky. Výhoda je ve větší diverzitě interpretů. Pro porovnání s prací [1] ale nakonec použijeme třetí databázi, a to vlastní, skládající se z 8 žánrů. Tu pro trénování doplníme několika úryvky z databáze první. Podrobněji se zmíníme o databázích až při použití pro jednotlivé testy.

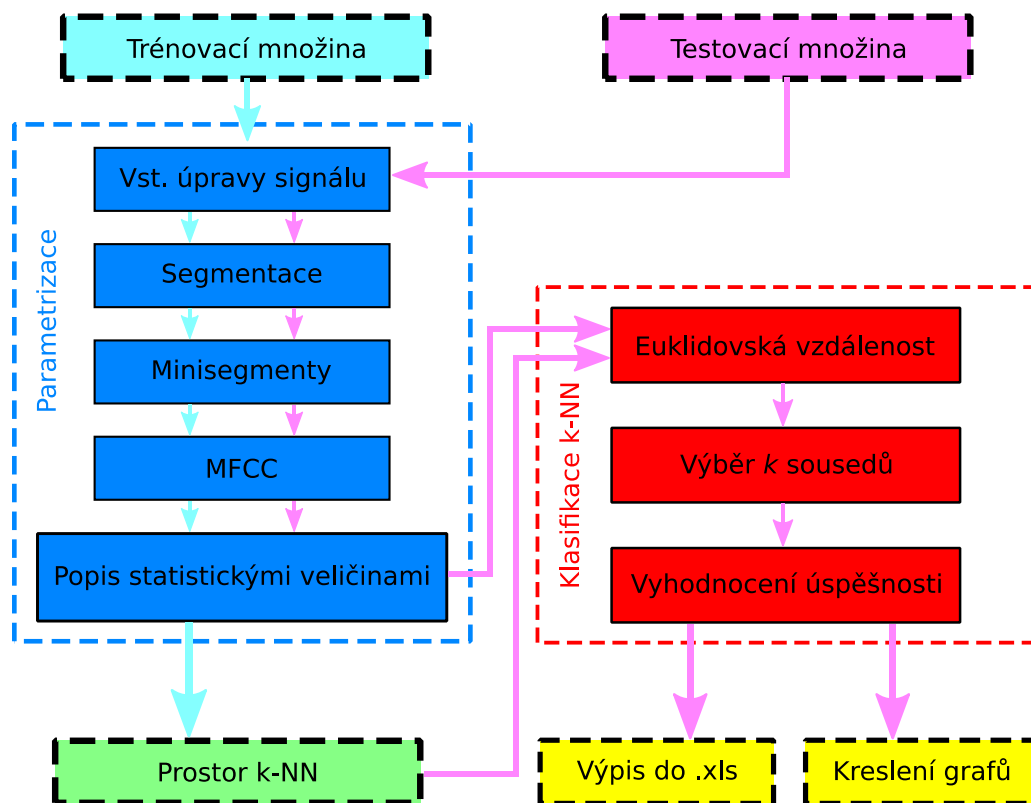
Výsledky reprezentujeme úspěšností klasifikace, což je poměr správně poznanych skladeb vůči celkovému počtu skladeb v rámci jednoho žánru. Označme je dílčími úspěšnostmi. Celkovou úspěšnost klasifikace programu pak můžeme spočítat dvěma způsoby. První způsob je spočítat sumu dílčích úspěšností a podělit ji počtem žánrů. Tak získáme celkovou **úspěšnost žánrovou**. Druhý způsob je spočítat sumu správně poznanych skladeb všech žánrů a dělit ji celkovým počtem skladeb všech žánrů. Označme ji jako **úspěšnost skladbovou**.

V případě, že testovací množina sestává ze stejného počtu skladeb pro každý žánr, bude úspěšnost žánrová rovna úspěšnosti skladbové. Představme si ale extrémní případ rozpoznávání dvou žánrů, kde jeden je v testovací množině zastoupen 100 skladbami a druhý 2 skladbami. Poznáme-li z prvního žánru všech 100 skladeb a z druhého jen 1, bude žánrová úspěšnost 75 % (průměr 100 % a 50 %), ale skladbová přes 99 %. Poznáme-li z prvního žánru skladeb 50 a z druhého skladby 2, pak je žánrová úspěšnost také 75 %, ale skladbová bude 51 %. Je-li žánr, který rozpoznáme bez problémů, zastoupen více skladbami, bude s velkou pravděpodobností skladbová úspěšnost větší než žánrová. Problém žánrové a skladbové úspěšnosti lze však jednoduše vyřešit stejným počtem skladeb pro všechny žánry. Pokud bychom v praxi očekávali, že na vstupu programu se častěji objeví specifický žánr, pak můžeme chtít testovat více skladeb tohoto žánru, abychom důkladněji prověřili úspěšnost. V takovém případě se skladbová úspěšnost k popisu může hodit.

V této práci uvádíme úspěšnost žánrovou, nebude-li uvedeno jinak.

## 4.2 Algoritmus programu

Jelikož používáme klasifikátoru k-NN, je potřeba jej nejdřív natrénovat (kap. 4.2.1). Algoritmus programu znázorňuje Obr. 5, podrobnější popis následuje v nadcházejících podkapitolách.



Obr. 5 Algoritmus programu

### 4.2.1 Fáze trénování

Pro trénování je potřeba mít připravenou množinu typických nahrávek daných žánrů. Programu musíme nějakým způsobem také předat informaci o tom, jakého žánru jednotlivé nahrávky trénovací množiny jsou. To lze (a také je tomu tak v našem případě) udělat náležitým pojmenováním jednotlivých nahrávek, např. přidat počáteční písmeno, které značí daný žánr.

Samotný algoritmus nejprve převede vstupní nahrávky formátu .wav na mono. Dále následuje případné zkrácení nahrávky o 30 sekund od začátku a 20 sekund od konce z důvodu hluchých míst nebo netypických inter a outer (záleží na tom, zda trénujeme již typickými krátkými úseky, nebo celými nesestříhanými nahrávkami). Poté následuje převzorkování signálu na požadovaný vzorkovací kmitočet. Konečná verze používá standardní CD vzorkovací kmitočet 44,1 kHz. Následuje normování amplitudy signálu tak, aby se amplituda pohybovala v intervalu  $< -1; 1 >$ .

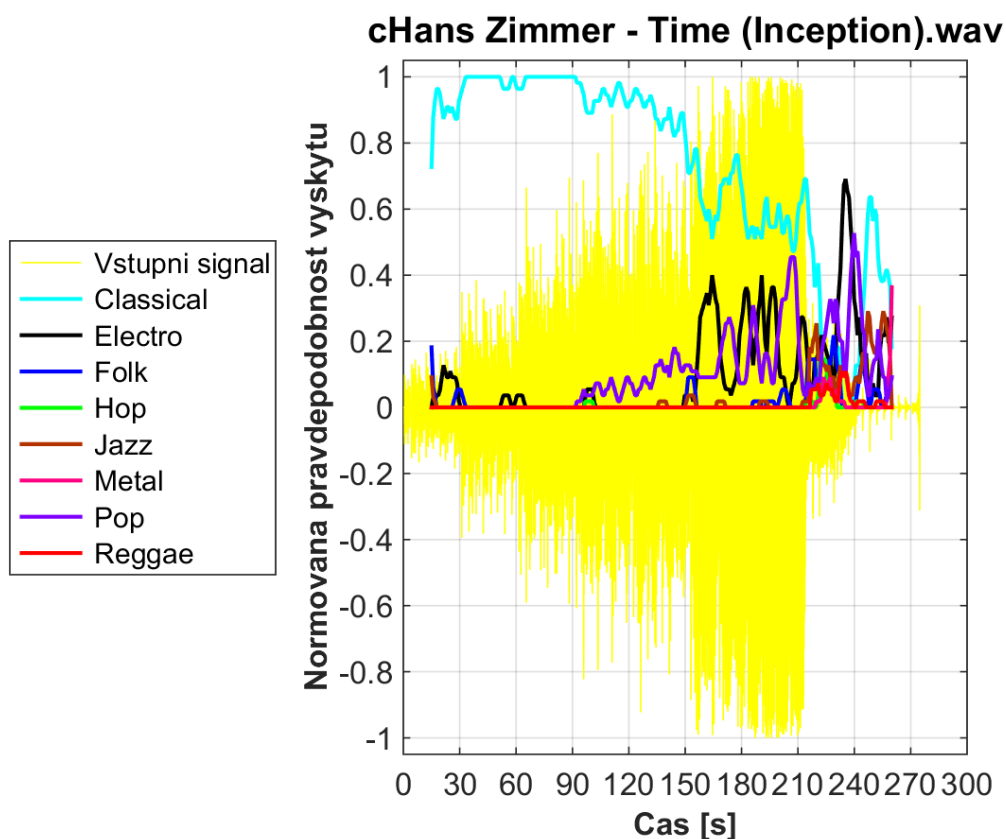
Program dále dělí signál na segmenty a jednotlivé segmenty procházejí procesem parametrizace. Ve funkci *MFCC.m* je nejprve segment pomocí funkce *mirframe.m* [2] rozdělen na několik minisegmentů. K ořezání okrajů se ve funkci použije Hammingova okna. Vstupními argumenty lze modifikovat délku minisegmentů a jejich překryv. Pro jednotlivé minisegmenty je pak funkcí *mirmfcc.m* [2] spočítáno  $r$  koeficientů MFCC. Vstupními argumenty lze měnit počet koeficientů  $r$  (Rank) a počet použitých filtrů ve filtrové bance  $b$  (Bands). Výstupem funkce *MFCC.m* je matice o rozměrech  $r \times m$ , kde  $r$  je počet koeficientů a  $m$  je počet minisegmentů v jednom segmentu. Ta je dále upravena pomocí statistických veličin na vektor o délce  $n$ , kde  $n$

se spočte jako součin  $r$  a počtu použitých statistických veličin. Tento vektor pak zaujme pozici ve větším vektoru, který bude obsahovat všechny vektory jednotlivých segmentů seřazených za sebou. Současně s tímto procesem je přečten první znak názvu nahrávky, z které byl segment vyňat, a podle toho, o jaký znak jde, se do vektoru uloží string příslušící danému žánru, tzv. label.

Když jsou všechny nahrávky trénovací množiny parametrizované, spočte se směrodatná odchylka každého z parametrů, která poslouží k normování jednotlivých parametrů.

#### **4.2.2 Fáze testování**

Vstupní nahrávka je upravena stejně jako ve fázi trénování, s výjimkou rozdílného zkrácení, které je 15 sekund od začátku i od konce. Opět parametrizujeme po jednotlivých segmentech. Každý jednotlivý segment vstoupí do klasifikátoru  $k$ -NN, který pracuje s  $n$ -rozměrným prostorem. Spočtou se rozdíly testovaného segmentu od všech segmentů trénovací množiny a na každý rozdíl je aplikována Euklidovská metrika (12). Vzniklý vektor s již jednočíselnými rozdíly se seřadí dle velikosti, přičemž se zachová informace o původních indexech daných rozdíly. Dále se vezme  $k$  nejmenších rozdíly a podle uložených indexů se nahlédne do vektoru labelů, který předá informaci o žánru těchto  $k$  vybraných segmentů z trénovací množiny. Výstupem jsou procenta zastoupení jednotlivých žánrů v tomto testovaném segmentu. Po klasifikování všech segmentů dané nahrávky se zastoupení žánrů v jednotlivých segmentech nahrávky zprůměrují a program vypíše výsledná procenta zastoupení jednotlivých žánrů v testované nahrávce. Program také umí vykreslit grafy (Obr. 6) znázorňující klasifikaci nahrávky v čase, z kterých lze v některých případech vyčíst, proč byla klasifikace neúspěšná apod., a to např. pozorováním klasifikace v časovém úseku, kde skladba přejde v netypický break, bridge nebo mezihru. Nakonec se vypíše průměrná procenta poznání jednotlivých žánrů za celý testovaný soubor nahrávek a v neposlední řadě počty skladeb predikované k jednotlivým žánrům. Výpisy se také uloží do souboru Microsoft Office Excel (.xls).



Obr. 6 Grafický výstup algoritmu

### 4.3 Test 1 - kombinace parametrů, vliv normování parametrů, 4 žánry

Nejdříve se omezíme na 4 žánry, a to na Classical, Dance, Metal, Reggae, a budeme zkoumat, které kombinace parametrů vykazují největší úspěšnost klasifikace. Tyto žánry by měly být velmi odlišné (viz kap. 1), a tudíž jejich rozlišení by nemělo být moc složité. Zároveň otestujeme, zda normování vzdáleností v k-NN dle vztahu (11) zlepší klasifikaci oproti žádnému normování. K tomuto účelu využijeme databáze od GZ Media. Stopy vybraných žánrů byly ručně sestříhány na úseky od sebe odlišné, tj. z jedné stopy, která může představovat jednu vinylovou desku, bylo získáno až 20 různých úseků. Odlišnost úseků byla posuzována subjektivně. Většinou úseky reprezentují samostatnou sloku, samostatný refrén nebo samostatné solo.

V trénovací množině je každý z žánrů zastoupen asi 15 úseky tak, aby měly úseky všech žánrů zhruba stejnou délku, která činí v průměru 13 minut. Zbytek úseků je použit jako testovací množina, dostáváme tedy počty úseků v testovací množině následující: Classical - 106, Dance - 72, Metal - 138, Reggae - 40.

Testované parametry jsou veškeré kombinace těchto tří statistických veličin: aritmetický průměr - **M** (Mean), směrodatná odchylka - **D** (Standard Deviation), koeficient šikmosti - **S** (Skewness). Kombinaci parametrů budeme značit jako **M+D**, **M+D+S** atd. Všechny tyto parametry jsou aplikovány na jednotlivé koeficienty MFCC získané z minisegmentů v rámci jednoho segmentu. Pokud např. počítáme  $r=13$  koeficientů MFCC v minisegmentech, pak jeden segment při použití kombinace parametrů M+D+S popisuje 39 číselných hodnot.

Pro všechny testy byly zvoleny pevné parametry dle (Tab. 1).

Tab. 1 Nastavení systému pro test 1

Nastavení systému	
Délka segmentu [s]	2,4
Překryv segmentu [s]	0,9
Délka minisegmentu [ms]	25
Překryv minisegmentu [ms]	6,25
Počet filtrů $b$ [-]	40
Počet koeficientů $r$ [-]	13
Počet sousedů $k$ [-]	7
Parametry statistické	Kombinace M, D a S
Vzorkovací kmitočet [kHz]	44,1

Tab. 2 Test kombinace parametrů bez použití normování parametrů (15)

	M	D	S	M+D	M+S	S+D	M+D+S
Metal	96,4 %	90,6 %	88,4 %	97,8 %	86,8%	16,0%	87,7%
Classical	84,0 %	72,6 %	13,2 %	86,8 %	63,9%	51,4%	66,7%
Dance	75,0 %	62,5 %	40,3 %	66,7 %	97,8%	94,2%	97,8%
Reggae	62,5 %	75,0 %	2,5 %	77,5 %	25,0%	27,5%	47,5%
<b>Celkem</b>	<b>79,5%</b>	<b>75,2%</b>	<b>36,1%</b>	<b>82,2%</b>	<b>68,4%</b>	<b>47,3%</b>	<b>74,9%</b>

Tab. 3 Test parametrů s použitím normování parametrů (15)

	M	D	S	M+D	M+S	S+D	M+D+S
Metal	94,2 %	94,9 %	90,6 %	97,8 %	97,8 %	98,6 %	98,6 %
Classical	83,0 %	69,8 %	12,3 %	90,6 %	84,9 %	45,3 %	88,7 %
Dance	73,6 %	61,1 %	31,9 %	68,1 %	68,1 %	50,0 %	65,3 %
Reggae	65,0 %	72,5 %	5,0 %	85,0 %	25 %	65,0 %	82,5 %
<b>Celkem</b>	<b>79,0%</b>	<b>74,6%</b>	<b>35,0%</b>	<b>85,4%</b>	<b>69,0%</b>	<b>64,7%</b>	<b>83,8%</b>

Z výsledků (Tab. 2), (Tab. 3) je vidět, že při kombinaci více parametrů normování jednotlivých parametrů úspěšnost klasifikace zlepšila. Pro parametr jeden jsou výsledky však velmi podobné, přestože jeden parametr sám tvoří více dimenzí pro k-NN rovné počtu použitých koeficientů MFCC. Nejlépe vychází použití kombinace parametrů aritmetického průměru a směrodatné odchylky, která dosáhla v testu úspěšnosti 85,4%.

Ke spočtení směrodatných odchylek sloužících k normování jednotlivých parametrů (nikoliv směrodatných odchylek používaných jako parametr samotný) je vždy použit průměr daného parametru za celou trénovací množinu. Pro zřetelnost uveďme rozvíjející příklad. Všechny segmenty trénovací množiny se parametrizují a tak při počtu koeficientů  $r=13$  dostáváme pro každý segment 13 aritmetických průměrů (parametrů M), které byly získány jako průměry jednotlivých koeficientů v minisegmentech příslušících danému segmentu. Z těchto průměrů (parametrů) spočítáme jejich aritmetický průměr. Ten poté dosadíme do vzorce (9) a spočteme směrodatnou odchylku tohoto parametru pro normování. Stejný postup platí i pro počítání normovací směrodatné odchylky jiných parametrů.

#### 4.4 Test 2 - databáze GTZAN, 10 žánrů, vliv převzorkování

Dále se budeme držet stejného nastavení a vyzkoušíme klasifikovat více žánrů. K tomuto účelu však použijeme jiné databáze, a to MARSYAS GTZAN databáze hudebních žánrů [27], jejíž použití v této práci mi bylo panem Georgem Tzanetakisem [21, 29] schváleno. Výhoda této databáze je její větší rozmanitost z hlediska interpretů. Ačkoliv není přiložen žádný seznam s tituly skladeb, při poslechu zřídka narazíme na dvě skladby stejného interpreta. Databáze obsahuje 10 žánrů, každý po 100 stopách o délce 30 sekund, Mono 16-bit, vzorkováno 22050 Hz ve formátu .au. Obsažené žánry jsou následující: Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae, Rock.

Databáze GTZAN je velmi často používaná různými experimentátory pro testování klasifikace hudebních žánrů [3, 4, 17, 28], ale ne vždy používají všechny žánry. My použijeme všech 10 žánrů. Databázi náhodně rozdělíme tak, aby 50 stop od každého žánru tvořilo trénovací množinu a 50 zbylých od každého žánru testovací množinu.

Nejprve porovnáme použití dvou nejúspěšnějších kombinací parametrů z předchozího testu, tedy M+D a M+D+S. Jelikož je trénovací množina pro každý žánr zhruba dvojnásobně dlouhá oproti předchozímu testu a zároveň obsahuje více žánrů, bylo zvoleno  $k=15$ .

Prvně byl vzorkovací kmitočet pro parametrizaci zvolen 22,05 kHz, tedy souhlasný se vzorkovacím kmitočtem databáze a podruhé 44,1 kHz, tedy standardní vzorkovací kmitočet pro audio záznam na CD, protože předpokládáme, že většina audio záznamů používá tento kmitočet, a systém chceme vyladit při použití tohoto kmitočtu. Převzorkování na jiný kmitočet realizuje MATLAB funkce resample používající k interpolaci implementace polyphase filterů.

Rozdíly v dílčích i celkových úspěšnostech klasifikace mezi zmíněnými vzorkovacími kmitočty nás vedl k vyzkoušení třetího vzorkovacího kmitočtu, a to 88,2 kHz.

Tab. 4 Nastavení systému pro test 2

Nastavení systému	
Délka segmentu [s]	2,4
Překryv segmentu [s]	0,9
Délka minisegmentu [ms]	25
Překryv minisegmentu [ms]	6,25
Počet filtrů $b$ [-]	40
Počet koeficientů $r$ [-]	13
Počet sousedů $k$ [-]	15
Parametry statistické	$M+D$ , $M+D+S$
Vzorkovací kmitočet [kHz]	22,05; 44,1; 88,2



Tab. 5 Úspěšnost klasifikace testu 2 - kombinace parametrů a vzorkovací kmitočet

Žánr	Úspěšnost M+D			Úspěšnost M+D+S		
	$f_s=22,05$ kHz	$f_s=44,1$ kHz	$f_s=88,2$ kHz	$f_s=22,05$ kHz	$f_s=44,1$ kHz	$f_s=88,2$ kHz
Blues	52,0%	46,0%	60,0%	24,0%	20,0%	56,0%
Classical	90,0%	100,0%	100,0%	80,0%	100,0%	100,0%
Country	26,0%	46,0%	46,0%	22,0%	38,0%	32,0%
Disco	68,0%	60,0%	40,0%	74,0%	72,0%	44,0%
Hip Hop	60,0%	76,0%	56,0%	72,0%	80,0%	58,0%
Jazz	66,0%	76,0%	66,0%	52,0%	62,0%	46,0%
Metal	94,0%	92,0%	86,0%	92,0%	92,0%	92,0%
Pop	62,0%	56,0%	94,0%	54,0%	54,0%	92,0%
Reggae	48,0%	40,0%	68,0%	36,0%	36,0%	72,0%
Rock	18,0%	26,0%	36,0%	30,0%	32,0%	38,0%
<b>Celkem</b>	<b>58,4%</b>	<b>61,8%</b>	<b>65,2%</b>	<b>53,6%</b>	<b>58,6%</b>	<b>63,0%</b>

Z výsledků opět pozorujeme, že kombinace parametrů M+D dosáhla větší úspěšnosti než M+D+S, a proto se nadále budeme držet pouze kombinace M+D. Zvýšení úspěšnosti klasifikace při použití vyššího vzorkovacího kmitočtu může souviset s těmito dvěma tvrzeními:

**1. Rozšíření spočítatelného frekvenčního spektra (Tvrzení 1)** - FFT počítá spektrum do poloviny vzorkovacího kmitočtu. Jelikož MFCC používá filtrů lineárně rozložených na Mel-frekvenční škále až do maximálního kmitočtu, zvýšením vzorkovacího kmitočtu se rozšíří spočítatelné spektrum a s ním i frekvenční šířka jednotlivých filtrů. Teoreticky by tedy šlo při použití nižšího vzorkovacího kmitočtu dosáhnout podobného výsledku pokud zmenšíme počet použitých filtrů  $b$ , čímž je také frekvenčně rozšíříme.

**2. Nárůst vzorků v minisegmentech (Tvrzení 2)** - více vzorků v minisegmentu znamená i více vzorků pro FFT aplikovanou na daný minisegment, což zlepšuje frekvenční rozlišení. Teoreticky bychom mohli podobného výsledku dosáhnout při použití delších minisegmentů, abychom dostali více vzorků, ale narozdíl od případu 1, kde můžeme dosáhnout ekvivalentní šířky filtrů, zde nebude ekvivalentní délka minisegmentu a tím pádem parametrizujeme jiné úseky.

#### 4.5 Test 3 - změna počtu filtrů $b$

Tvrzení 1 nás navedlo na otestování vlivu použitých filtrů  $b$  v MFCC na úspěšnost klasifikace. Pro konstantní parametry (Tab. 6) budeme jemně měnit počet filtrů  $b$ . Dle vztahu (3) bylo spočteno, že dvojnásobnému vzorkovacímu kmitočtu s  $b=40$  odpovídá při normálním vzorkovacím kmitočtu  $b=34$ . Výsledek tohoto testu může potvrdit nebo vyvrátit tvrzení 1 o převzorkování na vyšší vzorkovací kmitočet.

Tab. 6 Nastavení systému pro test 3

Nastavení systému	
Délka segmentu [s]	2,4
Překryv segmentu [s]	0,9
Délka minisegmentu [ms]	25
Překryv minisegmentu [ms]	6,25
Počet filtrů $b$ [-]	26 - 40
Počet koeficientů $r$ [-]	13
Počet sousedů $k$ [-]	15
Parametry statistické	M+D
Vzorkovací kmitočet [kHz]	44,1

Tab. 7 Úspěšnost klasifikace v závislosti na počtu použitých filtrů  $b$  pro  $r=13$

Počet filtrů $b$	Úspěšnost klasifikace [%]	Počet filtrů $b$	Úspěšnost klasifikace [%]
26	55,0	34	61,4
27	55,4	35	61,4
28	55,8	36	63,0
29	56,6	37	62,8
30	57,2	38	63,0
31	57,4	39	62,6
32	57,0	40	61,8
33	59,6		

Nejllepších výsledků dosahuje pro 13 koeficientů počet filtrů 36 nebo 38. Tvrzení 1 o převzorkování na vyšší kmitočet se nepotvrdilo. Úspěšnost klasifikace pro 34 filtrů je srovnatelná s úspěšností pro 40 filtrů.

#### 4.6 Test 4 - změna počtu použitých koeficientů

V následujícím testu budeme měnit počet použitých koeficientů pro 4 různé počty filtrů  $b$  (Tab. 8). Počet použitých koeficientů představuje část spektra, kterou použijeme. Zkoumáme, zda není vhodnější použít menší či větší část spektra a také zkoumáme, zda menší počet filtrů při menším počtu použitých koeficientů (nějak široká část spektra) bude dosahovat podobných výsledků jako vyšší počet filtrů s vyšším počtem koeficientů (podobně široká část spektra).

Tab. 8 Nastavení systému pro test 4

Nastavení systému	
Délka segmentu [s]	2,4
Překryv segmentu [s]	0,9
Délka minisegmentu [ms]	25
Překryv minisegmentu [ms]	6,25
Počet filtrů $b$ [-]	26; 31; 36; 40
Počet koeficientů $r$ [-]	5 - 35
Počet sousedů $k$ [-]	15
Parametry statistické	M+D
Vzorkovací kmitočet [kHz]	44,1

Tab. 9 Úspěšnost klasifikace v % pro 4 různá  $b$  v závislosti na  $r$

$r$	5	7	9	11	13	15	17	20	23	26	30	35
$b=26$	52,6	54,6	55,2	56	55	51,8	52,8	50,2	-	-	-	-
$b=31$	54,2	55	55,4	57,6	57,4	55,8	56	56,6	55,4	52,8	-	-
$b=36$	57,6	60,4	59,4	61	<b>63</b>	61,6	61,8	60,4	59	59	56	-
$b=40$	56	57,8	61,2	60,4	61,8	61,6	61,6	59,8	59,6	59,2	58,2	55,4

Z výsledků je vidět, že při použití menšího počtu filtrů nedochází ke znatelnému zlepšení při změně počtu koeficientů. Maximum úspěšnosti se však u  $b=26$  kolem  $r=11$  a u  $b=31$  mezi  $r=11$  a  $r=13$ , kdežto u  $b=40$  se maximum pohybuje někde mezi  $r=13$  a  $r=17$ . Z toho vyvozujeme závěr, že maximum úspěšnosti klasifikace je svázáno s použitou šířkou pásma a je tohoto maxima dosáhneme při menším počtu filtrů s menším počtem koeficientů a při větším počtu filtrů s větším počtem koeficientů. Ovšem menší počet filtrů s menším počtem koeficientů poskytne menší rozlišení, méně dimenzí a proto je vhodnější použít filtrů více. Nejvyšší úspěšnosti (63 %) dosahuje nastavení  $b=36$  a  $r=13$ , které nadále budeme používat.

#### 4.7 Test 5 - změna délky segmentu a minisegmentu

Nyní otestujeme vliv délky segmentu a minisegmentu na úspěšnost klasifikace. Použití delších segmentů může posloužit k lepšímu popisu pomalé hudby, naopak kratší segmenty mohou detailněji popsat hudbu více dynamickou. U minisegmentů sledujeme, zda delší minisegmenty povedou k vyšší úspěšnosti dle tvrzení 2 o převzorkování na vyšší kmitočty.

Tab. 10 Nastavení systému pro test 5

Nastavení systému	
Délka segmentu [s]	1; 2,4; 4
Překryv segmentu [s]	0,5; 0,9; 2
Délka minisegmentu [ms]	10; 25; 40
Překryv minisegmentu [ms]	2,5; 6,25; 10
Počet filtrů $b$ [-]	36
Počet koeficientů $r$ [-]	13
Počet sousedů $k$ [-]	15
Parametry statistické	M+D
Vzorkovací kmitočet [kHz]	44,1

Tab. 11 Úspěšnost klasifikace v % pro různé délky segmentů a minisegmentů

Délka segmentu; překryv segmentu	Délka minisegmentu; překryv minisegmentu		
	10 ms; 2,5 ms	25 ms; 6,25 ms	40 ms; 10 ms
1 s; 0,5 s	62,2	62,6	63,8
2,4 s; 0,9 s	59,6	63	62,2
4 s; 2 s	61,4	62,6	63

Rozdíl mezi testovanými délkami segmentů a minisegmentů není moc zásadní. Ačkoliv nejlepšího výsledků dosáhlo nastavení {1 s; 0.5 s; 40 ms; 10 ms} budeme se nadále držet

nastavení {2,4 s; 0,9 s; 25 ms; 6,25 ms}, a to především kvůli rychlejšímu výpočtu. První možnost má třetinový posun oproti druhé možnosti a výpočet zpravidla trvá třikrát déle. Také nastavení {2,4 s; 0,9 s; 25 ms; 6,25 ms} upřednostníme nastavení {4 s; 2 s; 40 ms; 10 ms}, a to kvůli větší úspěšnosti v málo úspěšných žánrech Rocku, Country a Disco.

Tvrzení 2 o převzorkování na vyšší kmitočty se nepotvrdilo ani nevyvrátilo. Delší minisegmenty při použití nastavení {2,4 s; 0,9 s} nevedou k zlepšení, ale v ostatních dvou případech ano, ne však dostatečně dramaticky. Vliv převzorkování na úspěšnost klasifikace by jistě stálo za to prozkoumat, ale na to není v rámci této práce prostor.

#### 4.8 Test 6 - vlastní databáze, 8 žánrů, vliv trénovací množiny

Abychom mohli porovnat naše výsledky s výsledky práce [1], omezíme se na 8 žánrů pro které utvoříme databázi obsahující celé skladby, stejně jako tomu je v [1], avšak s jedním žánrem k dobru. Porovnání žánrů a jejich názvů také s přiřazením demonstruje tabulka. Žánry Metal a Rock mohou být odlišné, ovšem složení použitých nahrávek v naší databázi a v databázi [1] se výrazně neliší, tudíž prohlašujeme tyto žánry v tomto případě za identické. Předpokládáme, že proběhlé testy na předchozí databázi by vykazaly podobné nebo stejné výsledky i pro tuto databázi, protože se jedná stále o stejnou problematiku a přístup, výběr žánrů se také výrazně nezměnil.

Tab. 12 Srovnání žánrů vlastní databáze s žánry použitými v [1]

Naše žánry	Classical	Electro	Folk	Hip Hop	Jazz	Metal	Pop	Reggae
Žánry v [1]	Vážná hudba	Elektronika	Folk	Hip Hop	Jazz	Rock	Pop	-

Jako trénovací množina bude použito sestříhaných úseků z databáze GZ Media, kterých jsme použili pro testování vhodné kombinace parametrů. Průměrná délka každého žánru v této trénovací množině je 11 minut. Trénovací množinu bude potřeba obohatit některými z nahrávek naší databáze (např. žánr Electro je v sestříhaných úsecích spíše ve formě Dance). Testovací množinu zatím tvoří všechny nahrávky naší databáze. Při obohacování trénovací množiny ale budeme nahrávky přesouvat z testovací množiny za účelem zvýšení úspěšnosti klasifikace zbylých nahrávek v testovací množině a tím se sníží počet testovaných nahrávek.

Nejprve otestujeme úspěšnost při použití trénovací množiny pouze ze sestříhaných úseků.

Tab. 13 Nastavení systému pro test 6

Nastavení systému	
Délka segmentu [s]	2,4
Překryv segmentu [s]	0,9
Délka minisegmentu [ms]	25
Překryv minisegmentu [ms]	6,25
Počet filtrů $b$ [-]	36
Počet koeficientů $r$ [-]	13
Počet sousedů $k$ [-]	15
Parametry statistické	M+D
Vzorkovací kmitočty [kHz]	44,1

Tab. 14 Úspěšnost klasifikace, podtrénovaný systém

		Predikovaný žánr							Úspěšnost	
		Clas	Ele	Folk	Hop	Jazz	Met	Pop		Reg
Skutečný žánr	Classical	17	0	0	0	1	1	0	0	89,5%
	Electro	0	5	0	3	0	11	1	2	22,7%
	Folk	4	1	5	9	0	4	0	0	21,7%
	Hip Hop	0	1	0	16	0	0	0	4	76,2%
	Jazz	0	5	2	7	3	0	1	2	15,0%
	Metal	0	0	0	0	0	26	0	0	100,0%
	Pop	1	8	0	6	0	4	4	1	16,7%
	Reggae	0	0	0	15	0	0	0	2	11,8%
									<b>44,2%</b>	

Z výsledků je vidět vysoká úspěšnost žánrů Classical, Metal a Hip Hop, ale velmi nízká u ostatních žánrů. Navíc je vidět, že žánry Metal a Hip Hop ovlivňují klasifikaci v jiných žánrech. Povšimněme si klasifikace žánru Reggae, který byl v 15 ze 17 případů predikován jako Hip Hop. Nízké úspěšnosti souvisí s malou trénovací množinou. Systém označíme jako podtrénovaný. Šlo by také spekulovat o tom, zda vysoká úspěšnost zmíněných 3 žánrů souvisí s jejich "soudružností", myšleno tak, že se skladby daných žánrů neliší tolik interpret od interpreta na rozdíl od ostatních žánrů. Trénovací množinu dále dotrénujeme 5 neúspěšnými žánry zmíněným přesunem nahrávek z testovací množiny. Jelikož trénujeme celými (nesestříhanými) nahrávkami, zavedeme zkrácení nahrávky o 30 sekund na začátku a o 20 sekund na konci a to z důvodu možného intra, outra nebo hluchých míst. Zároveň omezíme maximální počet segmentů, který z jednotlivých nahrávek parametrizujeme, aby nám zbytečně nepřeplnily prostor k-NN (pravděpodobně nebudou moc různorodé). Posuv segmentů zvolíme jako délku ořezané skladby dělenou 100. V případě, že je toto číslo menší, držíme se posuvu 1,5 s (odpovídající překryvu segmentů 0,9 s).

Tab. 15 Úspěšnost klasifikace po dotrénování vybranými žánry s nízkou úspěšností

		Predikovaný žánr							Úspěšnost	
		Clas	Ele	Folk	Hop	Jazz	Met	Pop		Reg
Skutečný žánr	Classical	13	1	0	0	2	0	3	0	68,4 %
	Electro	0	12	0	0	0	0	4	2	66,7 %
	Folk	4	1	9	0	0	1	3	1	47,4 %
	Hip Hop	0	1	0	2	0	0	0	18	9,5 %
	Jazz	1	2	2	2	3	0	4	2	18,8 %
	Metal	0	1	0	0	0	24	1	0	92,3 %
	Pop	1	5	1	0	0	0	9	4	45,0 %
	Reggae	0	0	1	1	0	0	1	10	76,9 %
									<b>53,1 %</b>	

Nyní pozorujeme značné zlepšení v žánru Reggae, Electro, Folk a Pop, ale žánr Hip Hop jakoby si s žánrem Reggae vyměnily místo, většina žánru Hip Hop je klasifikována jako Reggae. Doplněním trénovací množiny o další nahrávky Hip Hopu bychom měli úspěšnost zlepšit.

Po několika úpravách dostáváme výsledky (Tab. 16).

Tab. 16 Úspěšnost klasifikace - konečná verze

		Predikovaný žánr								Úspěšnost
		Clas	Ele	Folk	Hop	Jazz	Met	Pop	Reg	
Skutečný žánr	Classical	13	2	0	0	1	0	2	0	72,2 %
	Electro	0	9	0	2	0	0	4	0	60,0 %
	Folk	3	0	9	1	0	0	3	0	56,3 %
	Hip Hop	0	1	0	14	0	0	0	3	77,8 %
	Jazz	0	2	3	1	3	0	4	0	23,1 %
	Metal	0	2	1	0	0	22	0	0	88,0 %
	Pop	1	0	1	0	0	0	14	1	82,4 %
	Reggae	0	0	0	1	0	0	2	10	76,9 %
										<b>67,1 %</b>

Z tabulky (Tab. 16) je vidět vysoká úspěšnost žánrů Classical, Hip Hop, Metal, Pop a Reggae. Tyto žánry se od sebe značně liší svým charakterem. Horší úspěšnost byla dosažena pro žánry Folk a Electro. Žánr Folk se může pohybovat od klidného kytarového nebo i houslového projevu až po dynamický kytarový doprovod s výrazným zpěvem a dalšími nástroji, třeba i bicími. Tři neúspěšně klasifikované nahrávky žánru Folk byly klasifikovány jako Classical a další tři jako Pop. To potvrzuje výše popsání. Žánr Electro se v dnešní době s oblibou pojí s žánrem Pop a je těžké najít hranici mezi těmito dvěma. Často může být daná nahrávka trochu od jednoho i od druhého, tedy někde mezi žánry. To dokumentuje mírný přeslech žánru Pop do žánru Electro. Nejméně úspěšná klasifikace byla pro žánr Jazz. Bohužel se nepodařilo tento žánr vyladit na dostatečnou hladinu úspěšnosti klasifikace. Neúspěch přisuzujeme nedostatečné diverzitě trénovací množiny oproti velké diverzitě žánru Jazz v testovací množině.

#### 4.9 Test 7 - změna počtu nejbližších sousedů $k$

V poslední fázi budeme ladit počet sousedů  $k$  tak, abychom dosáhli nejvyšší možné úspěšnosti. Při použití správných parametrů, které žánry dobře oddělí, by nemělo při změně  $k$  docházet k nikterak velkým výkyvům v úspěšnosti klasifikace. Výkyvy úspěšnosti bychom také mohli očekávat, pokud by trénovací prostor  $k$ -NN obsahoval málo bodů, tedy pokud by trénovací množina byla nedostatečně velká.

Tab. 17 Nastavení pro test 7

Nastavení systému	
Délka segmentu [s]	2,4
Překryv segmentu [s]	0,9
Délka minisegmentu [ms]	25
Překryv minisegmentu [ms]	6,25
Počet filtrů $b$ [-]	36
Počet koeficientů $r$ [-]	13
Počet sousedů $k$ [-]	5 - 200
Parametry statistické	M+D
Vzorkovací kmitočet [kHz]	44,1

Tab. 18 Úspěšnost klasifikace v závislosti na  $k$

$k$	5	6	7	8	9	10	11	12	13	14	15
Úsp. žánr. [%]	67,5	<b>68,9</b>	<b>68,9</b>	68	67,1	68	<b>68,8</b>	68,3	68,3	67,4	67,4
Úsp. skl. [%]	69,6	<b>71,1</b>	<b>71,1</b>	70,4	69,6	70,4	<b>71,1</b>	70,4	70,4	69,6	69,6

17	20	25	30	40	50	100	200
66,4	66,4	66,9	66,4	66,4	65,6	64,8	65,4
68,9	68,9	69,6	68,9	68,9	68,1	68,1	68,9

V tabulce výsledků (Tab. 18) je vidět největší úspěšnost pro  $k=6$  a  $k=7$ , úspěšnost  $k=11$  zaostává o 0,1 %, ale pokud se podíváme na úspěšnost skladbovou, dostaneme stejná čísla. V (Tab. 19) a (Tab. 20) je vidět podrobněji úspěšnost variant  $k=7$  a  $k=11$ . Varianty  $k=6$  a  $k=7$  jsou totožné.

Tab. 19 Úspěšnost klasifikace pro  $k=7$

		Predikovaný žánr							Úspěšnost	
		Clas	Ele	Folk	Hop	Jazz	Met	Pop		Reg
Skutečný žánr	Classical	13	2	0	0	1	0	2	0	72,2 %
	Electro	0	10	0	1	0	0	4	0	66,7 %
	Folk	2	0	10	1	0	0	3	0	62,5 %
	Hip Hop	0	1	0	15	0	0	0	2	83,3 %
	Jazz	0	1	3	1	5	0	3	0	38,5 %
	Metal	0	2	1	0	0	22	0	0	88,0 %
	Pop	0	1	1	0	2	0	12	1	70,6 %
	Reggae	0	0	1	2	0	0	1	9	69,2 %
<b>68,9 %</b>										

Tab. 20 Úspěšnost klasifikace pro  $k=11$

		Predikovaný žánr								Úspěšnost
		Clas	Ele	Folk	Hop	Jazz	Met	Pop	Reg	
Skutečný žánr	Classical	13	2	0	0	1	0	2	0	72,2 %
	Electro	0	9	0	2	0	0	4	0	60,0 %
	Folk	2	0	10	1	0	0	3	0	62,5 %
	Hip Hop	0	1	0	14	0	0	0	3	77,8 %
	Jazz	0	2	3	1	4	0	3	0	30,8 %
	Metal	0	2	1	0	0	22	0	0	88,0 %
	Pop	0	1	1	0	0	0	14	1	82,4 %
	Reggae	0	0	1	1	0	0	1	10	76,9 %
<b>68,8 %</b>										

Varianta  $k=7$  je úspěšnější v žánrech Jazz, Electro a Hip Hop, varianta  $k=11$  je úspěšnější v žánrech Pop a Reggae. Zbývá zvolit konečnou variantu programu. Volbou  $k=7$  můžeme podpořit žánr Jazz, který dosahuje velmi malé úspěšnosti. Rozhodneme se až po testu rychlejší varianty programu (viz. kap. 4.10).

#### 4.10 Test 8 - zrychlení algoritmu

V závěrečné části se pokusíme o experimentální návrh způsobu zrychlení algoritmu, aniž by došlo k výraznému poklesu úspěšnosti. V zásadě se jedná o snížení počtu segmentů, které z testované nahrávky parametrizujeme a klasifikujeme. Výběr konkrétních segmentů ke klasifikaci by mohl realizovat např. systém detekce. Ke klasifikaci bychom mohli vybírat segmenty pouze z refrénu na základě vyšší energie nebo počtu opakování ve skladbě, nebo zkrátka jakékoliv segmenty s vyšší energií, abychom se vyhnuli segmentům ticha či nevýrazných částí skladby. Proti je však fakt, že některé žánry nejsou tolik energické a některé skladby některých žánrů nemusí nutně obsahovat refrén nebo jiné často opakující se části.

Rozhodli jsme se pro výběr segmentů v rovnoměrných intervalech. Segmenty vybereme ze skladby na 10 místech, přičemž vždy na daném místě vezmeme 3 segmenty po sobě. Dohromady tedy dostáváme 30 segmentů na skladbu, což pro délku segmentu 2,4 s a překryv 0,9 s (posun 1,5 s) představuje celkem 44 s. V případě původního (pomalého) algoritmu by pro tříminutovou skladbu po ořezání bylo vyňato přibližně 120 segmentů. V takovém případě se dostáváme na čtyřnásobnou rychlost klasifikace. V případě delší skladby, kterou je např. Klarinetový koncert A-dur od W.A. Mozarta o délce necelých 13 minut, bude zrychlení ještě výraznější.

Úspěšnost klasifikace rychlého systému může dosahovat maxima pro jiné nastavení systému než u systému pomalého. My pouze vyzkoušíme měnit počet sousedů  $k$  (Tab. 21).

Tab. 21 Úspěšnost klasifikace rychlého systému

$k$	6	7	11	15
Úspěšnost [%]	66,7	65,5	67,2	64,7



Nejúspěšnější z vyzkoušených je  $k=11$ ,

Tab. 22 Úspěšnost klasifikace rychlého systému pro  $k=11$

		Predikovaný žánr							Úspěšnost	
		Clas	Ele	Folk	Hop	Jazz	Met	Pop		Reg
Skutečný žánr	Classical	13	0	0	0	1	1	3	0	72,2 %
	Electro	0	9	0	2	0	0	4	0	60,0 %
	Folk	2	0	11	1	0	0	2	0	68,8 %
	Hip Hop	0	1	0	14	0	0	0	3	77,8 %
	Jazz	0	2	3	1	3	0	4	0	23,1 %
	Metal	0	2	1	0	0	22	0	0	88,0 %
	Pop	0	1	0	1	2	0	12	1	70,6 %
	Reggae	0	0	0	1	0	0	2	10	76,9 %
									<b>67,2 %</b>	

V (Tab. 22) si povšimněme doposud nejvyšší úspěšnosti v žánru Folk. Úspěšnost všech žánrů se drží na podobných hodnotách jako u pomalého systému. Jelikož nejlepšího výsledku u rychlého systému dosáhlo  $k=11$ , zvolíme stejné  $k$  i pro pomalý systém. Celková úspěšnost rychlého systému 67,2 % je také velmi podobná úspěšnosti 68,8 % pomalého systému.



## 5 Zhodnocení výsledků

Jedním z úkolů v zadání této bakalářské práce je srovnání výsledků s výsledky práce [1]. Porovnáme úspěšnosti a rychlosti obou systémů.

### 5.1 Srovnání systémů

Náš systém dosáhl úspěšnosti 68,8 % v pomalé verzi a 67,2 % v rychlé verzi pro 8 hudebních žánrů. V (Tab. 23) je podrobná matice klasifikace pro systém z práce [1] pro stejnou testovací množinu, která byla použita pro náš systém.

Tab. 23 Úspěšnost klasifikace systému [1]

		Predikovaný žánr							Úspěšnost
		Clas	Ele	Folk	Hop	Jazz	Met	Pop	
Skutečný žánr	Classical	14	1	1	0	0	2	0	77,8 %
	Electro	0	12	1	0	0	1	1	80,0 %
	Folk	7	0	5	1	1	0	2	31,3 %
	Hip Hop	0	4	1	6	2	3	2	33,3 %
	Jazz	2	1	2	1	6	1	0	46,2 %
	Metal	0	4	1	0	0	19	1	76,0 %
	Pop	2	5	3	0	0	1	6	35,3 %
<b>54,3 %</b>									

Pro stejnou testovací množinu dosáhne systém [1] úspěšnosti 54,3 %, pokud bychom počítali žánr Reggae jako neúspěšnou klasifikaci, pak 47,5 %. Systém [1] je úspěšnější v žánru Jazz a výrazně úspěšnější v žánru Electro, ale značně strádá v žánrech Folk, Hip Hop a Pop. Náš systém je celkově zdatně úspěšnější.

Co se rychlosti týče, náš systém je i mírně rychlejší. Rychlý systém bez vykreslování grafů otestuje žánr Metal za 112 s (25 nahrávek), žánr Jazz za 60 s (13 nahrávek), žánr Classical za 88 s (18 nahrávek). Systém [1] otestuje žánr Metal za 180 s, Jazz za 95 s a Classical za 150 s. Dosáhli jsme tedy více než 1,5 násobku rychlosti. S vykreslováním grafů se však klasifikace značně zpomalí, u Metalu třeba na 370 s. Pomalý systém je pak ještě značně pomalejší (doba klasifikace se samozřejmě bude lišit počítač od počítače, hodnoty slouží k porovnání mezi sebou).

Z důvodu značného zpomalení rychlého systému vykreslováním grafů a zároveň skutečností, že rychlý systém by měl být skutečně rychlý, je rozumné možnost kreslení grafů u rychlého systému úplně vypustit. Takový graf by ukazoval klasifikaci pouze v 10 posuzovaných úsecích (kap. 4.10), což není moc nápomocné.

Rozdíly v úspěšnosti klasifikace našeho systému a systému [1] mohou být důsledkem jiné trénovací množiny, ale pravděpodobně jde spíše o rozdílný přístup parametrizace. Systém [1] používá také MFCC, ale bez zapojení statistických veličin. Součástí systému [1] je však i jednoduché GUI, které v našem systému chybí.

Ačkoliv úspěšnost klasifikace 68,8 % je blízká úspěšnosti 70 %, které dosáhli studenti v testu (kap. 1), ve srovnání s jinými systémy není moc velká. V [17] dosáhli úspěšnosti 85,6 % pro 10 žánrů s použitím databáze GTZAN MARSYAS [27]. Pro parametrizaci použili mnoho spektrálních parametrů, např. spectral centroid, brightness, spectral entropy, flatness. Pro lepší popis použili také statistických veličin, a to aritmetického průměru, směrodatné odchylky, koeficientu šikmosti a koeficientu špičatosti. Pro klasifikaci použili klasifikátor ELM (Extreme Learning Machine).

## 5.2 Návrh možných zlepšení, budoucí práce

Existuje řada možností, jak systém dále zlepšovat. První je volba jiných nebo dalších parametrů. Mohli bychom se inspirovat spektrálními parametry použitými v [17], taktéž zkusit využít koeficientu šikmosti a koeficientu špičatosti. Změnou klasifikátoru bychom také mohli dosáhnout lepší úspěšnosti. Klasifikátory pracující se shluky, jako např. k-Means, by mohly vykázat větší úspěšnost lepším potlačením "náhodných sousedů" nesprávných žánrů při klasifikaci, jejichž vliv je v k-NN zřetelný (viz kap. 3.1).

Systém by mohl využívat stromové struktury, tj. nejdříve by se rozhodlo na základě nějakého obecnějšího parametru, zda nahrávka patří do některé z nadskupin, např. klidná hudba a energická hudba, nebo hudba se zpěvem a bez zpěvu. Dále by na základě jiného parametru systém rozhodl, zda patří nahrávka do nějaké ze žánrových skupin, např. skupina Electro + Metal nebo skupina Folk+Country. A v poslední fázi by se rozhodlo na základě zase jiných, nebo stejných, ale upravených parametrů, o jaký konkrétní hudební žánr se jedná. Podobného přístupu využívají v [33], kde nejdříve dělí na instrumentální hudbu nebo píseň. Dále rozlišují, jaké nástroje se v nahrávce vyskytují. Podle nástrojů bychom také mohli rozhodovat o příslušnosti jednotlivým žánrům.

Nejvlivnějším faktorem je však dle mého názoru volba trénovací množiny. Pokud trénovací množinou obsáhneme téměř všechny existující varianty prezentace jednotlivých žánrů, pak by měla klasifikace být značně jednodušší. Problémem však stále bude skutečnost, že ani lidé sami nejsou schopni vždy správně označit jednotlivé hudební žánry a také fakt, že v dnešní době se velmi často vyskytují interpreti hrající na pomezí dvou či více žánrů, tedy dochází k míšení žánrů.

Zajímavý je také vliv převzorkování nahrávky na úspěšnost klasifikace (kap. 4.4). Rozhodně by stálo za to tento jev prozkoumat do větší hloubky. Pokud by to zlepšilo klasifikaci, pak by převzorkování mohlo být využito. Zásadní nevýhoda je však v tom, že převzorkování výrazně zpomaluje algoritmus. Bylo by více než na místě převzorkovávat pak pouze úseky, které použijeme pro klasifikace (např. vybrané segmenty), a nikoliv celou nahrávku.

Pokud bychom dosáhli větší úspěšnosti při stejném počtu žánrů, pak by stálo za zvážení rozšířit systém na více žánrů. Dříve to však není rozumné, neboť úspěšnost klasifikace značně klesá s počtem žánrů [4].

## 5.3 Rozlišení nehudbních a hudbních zvuků

Kompletní systém klasifikace hudebních žánrů určený pro použití v praxi by také měl být schopný rozlišit, zda je testovaná nahrávka skutečně hudební, nebo jde o nějaký pazvuk či řeč. Náš algoritmus není bohužel schopný rozlišit nehudbní zvuky od hudebních. Jednoduchá realizace takové funkce je implementována v práci [1]. Spočítáním průměrné vzdálenosti všech

segmentů testovací množiny od všech segmentů trénovací množiny byla vybrána hranice vzdálenosti, při které algoritmus varuje, že se pravděpodobně nejedná o hudbu.

Detekce nehudebních zvuků v [1] však byla vyzkoušena pouze pro mluvené slovo. Nehudebních zvuků je však ohromné množství. Problém by mohl nastat s hlasitějšími nehudebními zvuky, např. hluk na staveništi, ve výrobě, na ulici a další. Takové nehudební zvuky by se mohly podobat některým zvukům používaným v elektronické hudbě. Zpěv ptáků by se třeba mohl podobat hře na píšťalu v lidové hudbě.

Důkladně vyladit systém detekce nehudebních zvuků jistě nebude jednoduchý úkol a přístup použitý v [1] pravděpodobně nebude dostačující.

Prozatím si musíme vystačit bez možnosti rozlišení nehudebních zvuků a předpokládat, že na vstup algoritmu pro klasifikaci hudebních žánrů budou vpuštěny pouze hudební nahrávky.



## Závěr

Cílem této práce bylo vytvořit algoritmus schopný klasifikace hudebních žánrů s využitím spektrálního parametru MFCC. Tento úkol byl splněn. Algoritmus má dvě varianty, pomalou a rychlou. Obě používají 13 koeficientů MFCC, které jsou získány v minisegmentech o délce 25 ms s překryvem 6,25 ms. Koeficienty jednotlivých minisegmentů jsou zpracovány v rámci segmentů dlouhých 2,4 s s překryvem 0,9 s pomocí statistických veličin aritmetického průměru a směrodatné odchylky. Ke klasifikaci je použito klasifikátoru k-NN s počtem sousedů  $k=11$ . Výstupem programu je procentuální zastoupení žánrů Classical, Electro, Folk, Hip Hop, Jazz, Metal, Pop a Reggae v nahrávce. Pomalý algoritmus klasifikuje všechny segmenty v nahrávce a má možnost grafického výstupu klasifikace jednotlivých žánrů v čase. Rychlý algoritmus klasifikuje pouze 10 rovnoměrně rozprostřených trojic segmentů, čímž dosahuje značně vyšší rychlosti klasifikace. Pomalý algoritmus dosáhl úspěšnosti 68,8 %, rychlý 67,2 %.

Porovnali jsme náš systém se systémem [1]. Ten dosáhl pro stejnou testovací množinu pouze 54,3 % na 7 žánrech. Náš rychlý systém je zároveň přibližně 1,5 násobně rychlejší než systém [1]. Nejmenší úspěšnosti klasifikace náš systém dosahuje u žánru Jazz. To je do jisté míry pravděpodobně způsobeno velkou diverzitou žánru Jazz v testovací množině. Systém [1] nedosáhl pro žánr Jazz o moc větší úspěšnosti.

Možná zlepšení systému spočívají ve výběru dalších pomocných parametrů, použití jiného klasifikátoru nebo důkladnější volbě trénovací množiny (viz kap. 5.2).

## Seznam použité literatury

- [1] Bartoš Jaroslav, „*Detekce hudebních žánrů pro účely masteringu gramofonových desek*“, FEL ČVUT v Praze, 2016, Bakalářská práce, [online], dostupné z: [https://dspace.cvut.cz/bitstream/handle/10467/65310/F3-BP-2016-Bartos-Jaroslav-detekce\\_hudebnich\\_zanru\\_pro\\_ucely\\_masteringu\\_gramofonovych\\_desek.pdf](https://dspace.cvut.cz/bitstream/handle/10467/65310/F3-BP-2016-Bartos-Jaroslav-detekce_hudebnich_zanru_pro_ucely_masteringu_gramofonovych_desek.pdf)
- [2] Olivier Lartillot, Petri Toivainen: *Matlab Toolbox for Musical Feature Extraction From Audio*, International Conference on Digital Audio Effects, Bordeaux, 2007.
- [3] Mitsunori Ogihara, Tao Li: *Music genre classification with taxonomy*, IEEE
- [4] Michael Haggblade, Yang Hong, Kenny Kao: *Music Genre Classification*, [online], dostupné z: <http://cs229.stanford.edu/proj2011/HaggbladeHongKao-MusicGenreClassification.pdf>
- [5] Wikipedia "*Mel-frequency cepstrum*", [cit. 28.1.2017], [online], dostupné z: [https://en.wikipedia.org/wiki/Mel-frequency\\_cepstrum](https://en.wikipedia.org/wiki/Mel-frequency_cepstrum)
- [6] Beth Logan: *Mel Frequency Cepstral Coefficients for Music Modeling*, [online], dostupné z: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.11.9216&rep=rep1&type=pdf>
- [7] Abdillahi Hussein Omarm: *Audio Segmentation and Classification*, [online], dostupné z: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.8354&rep=rep1&type=pdf>
- [8] Wikipedia "*Směrodatná odchylka*", [cit. 28.1.2017], [online], dostupné z: [https://cs.wikipedia.org/wiki/Směrodatná\\_odchylka](https://cs.wikipedia.org/wiki/Směrodatná_odchylka)
- [9] Wikipedia "*Koeficient šikmosti*", [cit. 28.1.2017], [online], dostupné z: [https://cs.wikipedia.org/wiki/Koeficient\\_šikmosti](https://cs.wikipedia.org/wiki/Koeficient_šikmosti)
- [10] *K Nearest Neighbors Classification*, [online], dostupné z: [http://www.saedsayad.com/k\\_nearest\\_neighbors.htm](http://www.saedsayad.com/k_nearest_neighbors.htm)
- [11] Wikipedia "*Support vector machines*", [cit. 28.1.2017], [online], dostupné z: [https://cs.wikipedia.org/wiki/Support\\_vector\\_machines](https://cs.wikipedia.org/wiki/Support_vector_machines)
- [12] Wikipedia "*k-nearest neighbors algorithm*", [cit. 28.1.2017], [online], dostupné z: [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)
- [13] Lie Lu, Hao Jiang, HongJiang Zhang: *A Robust Audio Classification and Segmentation Method*, [online], dostupné z: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-2001-79.pdf>
- [14] Koshkina Ekaterina, „*Identifikace obsahu archivních zvukových záznamů*“, FEL ČVUT v Praze, 2015, Bakalářská práce, [online], dostupné z:



[https://dspace.cvut.cz/bitstream/handle/10467/61652/F3-BP-2015-Koshkina-Ekaterina-identifikace\\_obsahu\\_archivnich\\_zvukovych\\_zaznamu.pdf](https://dspace.cvut.cz/bitstream/handle/10467/61652/F3-BP-2015-Koshkina-Ekaterina-identifikace_obsahu_archivnich_zvukovych_zaznamu.pdf)

[15] Wikipedia "*Hudební žánr*", [cit. 28.1.2017], [online], dostupné z: [https://cs.wikipedia.org/wiki/Hudební\\_žánr](https://cs.wikipedia.org/wiki/Hudební_žánr)

[16] *Music Genres List*, [cit. 4.2.2017], [online], dostupné z: <http://www.musicgenreslist.com>

[17] Baniya, B.K.; Ghimire, D; Joonwhoan Lee: *Evaluation of different audio features for musical genre classification*, IEEE

[18] Málková Tereza: „*Hudební žánry a jejich působení na krátkodobou paměť*“, PDF MU Brno, 2015, Diplomová práce, [online], dostupné z: [http://is.muni.cz/th/350643/pdf\\_m/](http://is.muni.cz/th/350643/pdf_m/)

[19] All music, *Genres* [online] [cit. 4.2.2017], <http://www.allmusic.com/genres>

[20] Wikipedia "*Music genre*", [cit. 28.1.2017], [online], dostupné z: [https://en.wikipedia.org/wiki/Music\\_genre](https://en.wikipedia.org/wiki/Music_genre)

[21] Tzanetakis George, Essl Georg, Cook Perry: *Automatic Musical Genre Classification Of Audio Signals*, [online], dostupné z: <http://ismir2001.ismir.net/pdf/tzanetakis.pdf>

[22] Wikipedia: "*Extreme learning machine*", [cit. 5.2.2017], [online], dostupné z: [https://en.wikipedia.org/wiki/Extreme\\_learning\\_machine](https://en.wikipedia.org/wiki/Extreme_learning_machine)

[23] Lyons James: *Mel Frequency Cepstral Coefficient (MFCC) tutorial*, [cit. 3.5.2017], [online], dostupné z: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>

[24] Sahibsingh A. Dudani: *The Distance-Weighted k-Nearest-Neighbor Rule*, IEEE

[25] Tsang-Long Pao, Yun-Maw Cheng, Jun-Heng Yeh, Yu-Te Chen, Chen-Yu Pai and Yao-Wei Tsai: *Comparison between Weighted D-KNN and Other Classifiers for Music Emotion Recognition*, IEEE

[26] <http://www.data-machine.com/nmtutorial/distanceweightedknnalgorithm.htm>

[27] Tzanetakis George, *MARSYAS GTZAN Music Genre dataset*, dostupné z: [http://marsyasweb.appspot.com/download/data\\_sets/](http://marsyasweb.appspot.com/download/data_sets/)

[28] John Cast, Chris Schulze, Ali Fauci: *Music Genre Classification*, 3.5.2017

[29] Tzanetakis George, Cook Perry: *Musical Genre Classification of Audio Signals*, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 5, JULY 2002

[30] Trevino Andrea: *Introduction to K-means Clustering*, [cit. 9.5.2017], [online], dostupné z: <https://www.datascience.com/blog/introduction-to-k-means-clustering-algorithm-learn-data-science-tutorials>

[31] Wikipedia "K-means", [cit. 9.5.2017], [online], dostupné z: <https://cs.wikipedia.org/wiki/K-means>

[32] Wikipedia "*Machine learning*", [cit. 9.5.2017], [online], dostupné z: [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)

[33] Ghosal Arijit, Chakraborty Rudrasis, Chandra Dhara Bibhas, Sanjoy Kumar Saha: *Music Classification based on MFCC Variants and Amplitude Variation Pattern: A Hierarchical Approach*, [online], dostupné z: [http://www.sersc.org/journals/IJSIP/vol5\\_no1/10.pdf](http://www.sersc.org/journals/IJSIP/vol5_no1/10.pdf)

[34] Wikipedia "*Blues*", [cit. 17.5.2017], [online], dostupné z: <https://cs.wikipedia.org/wiki/Blues>

[35] Wikipedia "*Folk*", [cit. 17.5.2017], [online], dostupné z: <https://cs.wikipedia.org/wiki/Folk>

## Přílohy

### Příloha 1 - CD

Na CD jsou všechny skripty použité v této práci spolu s dalšími konkrétnějšími tabulkami klasifikace a grafy.

### Příloha 2 - Tabulky vlastní databáze, příkladové grafy

Pro dotrénování systému bylo použito skladeb z Tab. přílohy 1. K povšimnutí je nerovnoměrné zastoupení jednotlivých žánrů. Některé bylo potřeba dotrénovat více, některé méně (viz kap. 4.8).

*Tab. přílohy 1 Nahrávky použité pro dotrénování systému*

Žánr	Název	Žánr	Název
Classical	Joseph Haydn - Serenade	Jazz	Alissia - Let it Out
Electro	Daft Punk - One More Time	Jazz	Amy Winehouse - You Know I'm No Good
Electro	Gorillaz - Dare	Jazz	Frank Sinatra - I Won't Dance
Electro	Nicky Romero - Toulouse	Jazz	Marcus Miller - Power
Electro	Sigala - Sweet Lovin'	Jazz	Miles Davis - Portia
Electro	Skrillex - Bangarang	Metal	Black Sabbath - Wishing Well
Electro	Swedish House Mafia - Greyhound	Pop	Beyonce - If I Were a Boy
Folk	Bob Dylan - Like a Rolling Stone	Pop	Katy Perry - Roar
Folk	Jaromír Nohavica - Hlídač Krav	Pop	Maroon 5 - Sugar
Folk	Johnny Cash - Heart of Gold	Pop	Michael Jackson - Thriller
Folk	Mumford & Sons - Babel	Pop	OneRepublic - Counting Stars
Folk	The Dubliners - Dirty Old Town	Pop	Robbie Williams - Feel
Folk	Tomáš Klus - Nina	Pop	Tatu - All The Things She Said
Folk	Tomáš Klus - Přeju Ti	Reggae	Alborosie - Herbalist
Hip Hop	Obie Trice - Adrenaline Rush	Reggae	Danakil - Champs De Roses
Hip Hop	Xzibit - Spit Shine	Reggae	Dennis Brown - Money In My Pocket
Jazz	Dean Martin - Volare	Reggae	Inner Circle - Bad Boys

V testovací množině (kap. 4.8 až 4.10) byly zastoupeny skladby uvedené v Tab. přílohy 2, Tab. přílohy 3, Tab. přílohy 4 a Tab. přílohy 5.

Tab. přílohy 2 Testovací množina - Classical, Electro

č.	Classical	č.	Electro
1	R. Brower, D. Duke & G. Stafford - Arthas, My Son	19	Avicii - Wake Me Up
2	R. Brower, D. Duke & G. Stafford - Angrathar's Shadow	20	Crystal Castles - Concrete
3	Johann Sebastian Bach - Air	21	Eric Prydz - Call On Me
4	Antonio Vivaldi - Summer	22	Goldfrapp - Ooh La La
5	Peter Gundry - The Essence	23	Jonas Blue - Fast Car ft. Dakota
6	Franz Liszt - Hungarian Rhapsody	24	Justice - D.A.N.C.E.
7	Hans Zimmer - Time (Inception)	25	Justice - Fire
8	Johann Strauss - An der schonen blauen Donau	26	M83 - Midnight City
9	W. A. Mozart - Koncert klarinetovy A-Dur	27	Martin Garrix - Animals
10	W. A. Mozart - Figarova svatba	28	MGMT - Electric Feel
11	Pirates of The Caribbean - Theme	29	Psy - Gentleman
12	Preußens Gloria	30	Rudimental - Feel The Love ft. John Newman
13	Johann Strauss Sr - Radetzky March	31	Skrillex - Rock n Roll (Will Také You to the Mountain)
14	Richard Wagner - Ride of the Valkyries	32	Swedish House Mafia - One
15	Franz Schubert - Ave Maria	33	The Knife - Heartbeats
16	Peter Gundry - Sorrows Passing		
17	The Lord of the Rings - Soundtrack		
18	Petr Iljič Čajkovskij - Waltz of the Flowers		

Tab. přílohy 3 Testovací množina - Folk, Hip Hop

č.	Folk	č.	Hip Hop
34	Willie Nelson - Always On My Mind	50	Mobb Deep - Shook Ones Pt. II
35	Bob Dylan - Blowin in the Wind	51	Notorious B.I.G. - Juicy
36	Cat Power - The Greatest	52	Mc Breed with 2Pac - Gotta Get Mine
37	Jaromír Nohavica - Pijte Vodu	53	Outkast - Player's Ball
38	Jewel - Standing Still	54	Jr. Mafia - Get Money
39	Joanna Newsom - Sprout and the Bean	55	Jay-Z feat Freeway - 8 Miles And Runnin'
40	Johnny Cash - I See A Darkness	56	Ol' Dirty Bastard - Shimmy Shimmy Ya
41	Jozka Cerny - Za tu horu	57	Wu Tang Clan - C.R.E.A.M.
42	Leonard Cohen - Suzanne	58	Rakim - R.A.K.I.M.
43	Mumford & Sons - Hopeless Wanderer	59	Gangstarr - Battle
44	Neil Young - Old Man	60	Black Eyed Peas - Like That
45	Nick Drake - Day Is Done	61	Black Eyed Peas - Bebot
46	Nick Drake - Milk & Honey	62	Black Eyed Peas - Ba Bump
47	Simon & Garfunkel - Mrs. Robinson	63	Eminem - My 1st Single
48	Sufjan Steven - Should Have Known Better	64	Eminem - Rain Man
49	Tomáš Klus - Sibyla	65	Eminem - One Shot 2 Shot (Feat. D-12)
		66	Unknown LA rapper - Something you should know
		67	Unknown LA rapper - You gotta have haters

Tab. přílohy 4 Testovací množina - Jazz, Metal

č.	Jazz	č.	Metal
68	Nat King Cole - Unforgettable	81	Satyricon - The Pentagram Burns
69	Amy Winehouse - Back to Black	82	Damageplan - Explode
70	Bud Powell - Collard Greens and Black Eyed Peas	83	Amon Amarth - For Victory or Death
71	Diana Krall - The Look of Love	84	Animals As Leaders - Tooth and Claw
72	Frank Zappa - Bobby Brown	85	Arch Enemy - Stolen Life
73	Charlie Parker - All the things you are	86	Atreyu - This Flesh A Tomb
74	Chet Baker - Born to Be Blue	87	Gojira - Backbone
75	Jaga Jazzist - Animal Chin	88	Guns N' Roses - Paradise City
76	John Coltrane - Giant Steps	89	Helloween - I'm Alive
77	Katie Melua - In My Secret Life	90	Hellyeah - Stampede
78	Louis Armstrong - La vie en rose	91	Iron Maiden - The Wicker Man
79	Nina Simone - Feeling Good	92	Iron Maiden - Can I Play With Madness
80	Victor Wooten - What You Won't Do For Love	93	Jinjer - Captain Clock
		94	Kiss - Magic Touch
		95	Kreator - Phantom Antichrist
		96	Love and Death - Whip It
		97	Motley Crue - Dr. Feelgood
		98	Feared - Your God
		99	Prong - Put Myself To Sleep
		100	Slayer - War Ensemble
		101	System of a Down - Needles
		102	Soil - Suffering
		103	Soilwork - Rejection Role
		104	Steve Harris - British Lion - This is my God
		105	Tarot - Rise!

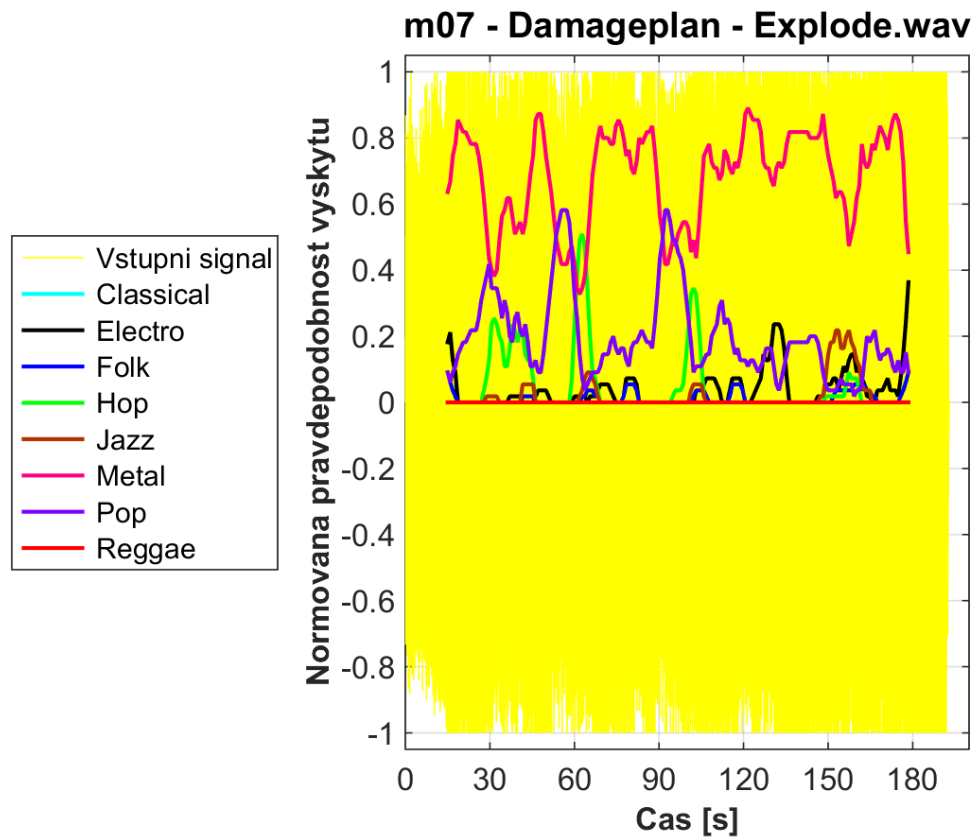
Tab. přílohy 5 Testovací množina - Pop, Reggae

č.	Pop	č.	Reggae
106	Britney Spears - Baby One More Time	123	Bob Marley - Get Up, Stand Up
107	Madonna - 4 Minutes ft. Justin Timberlake	124	Bob Marley - I Shot The Sheriff
108	Nicole Scherzinger - On the Rocks	125	Alborosie - Kingston Town
109	Justin Timberlake - Can't Stop The Feeling	126	Alpha Blondy - Brigadier Sabari
110	Nicole Scherzinger - Big Fat Lie	127	Broussai - Soif De Liberté
111	Delta Goodrem - Dear Life	128	Burning Spear - Walk
112	Adele - Someone Like You	129	Gregory Isaacs - Night Nurse
113	Anastacia - Left Outside Alone	130	Israel Vibration - My Master's Will
114	Backstreet Boys - I Want It That Way	131	Max Romeo - Chase The Devil
115	Bastille - Things We Lost In The Fire	132	Peter Tosh - Johnny B Goode
116	Bruno Mars - Just The Way You Are	133	Pierpoljak - Je sais pas jouer
117	Coldplay - The Scientist	134	Steel Pulse - Babylon Makes The Rules
118	Delta Goodrem - Believe Again	135	Tiken Jah Fakoly - Plus rien ne m'tonnes
119	George Michael - One More Try		
120	Ellie Goulding - Guns and Horses		
121	Spice Girls - Wannabe		
122	Taylor Swift - Shake It Off		

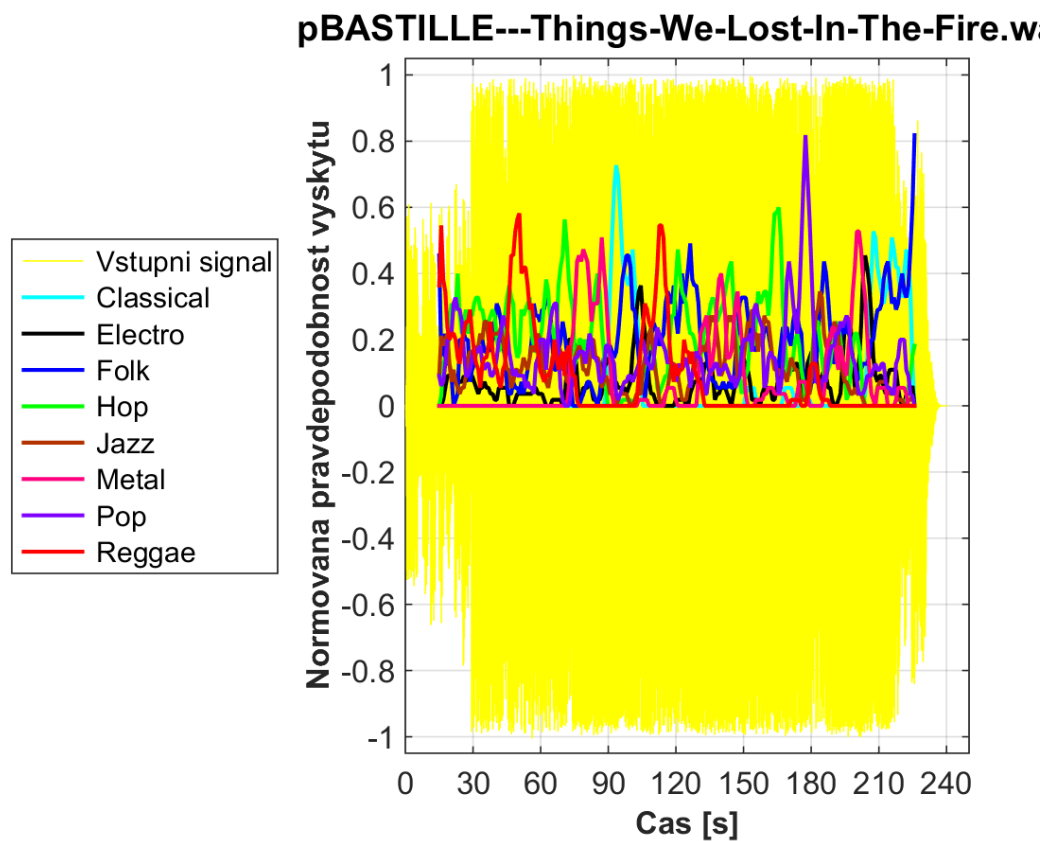
Dále se podíváme na 3 různé grafické výstupy programu. První z nich (Obr. přílohy 1) představuje poměrně jednoznačnou klasifikaci žánru Metal, který je zastoupen 68 %. V tomto případě by měla klasifikace být stejná pro rychlý i pomalý systém, protože nahrávka je téměř v každém čase klasifikována jako Metal. Pouze ve 3 krátkých úsecích je nahrávka klasifikována jinak.

V druhém případě je na vstupu nahrávka žánru Pop, která však obsahuje prvky mnoha jiných žánrů. Přesně to je také vidět v grafu (Obr. přílohy 2). Nahrávka je klasifikována jako žánr Folk s 19,95 %, těsně následuje žánr Hip Hop s 19,48 % a třetí je žánr Pop s 15,24 %.

V třetím případě je nahrávka z jiného než námi rozlišovaného žánru, a to žánru Blues (Obr. přílohy 3). Jelikož je žánr neznámý, klasifikace je také nejednoznačná, nicméně s 31 % vyhraje žánr Electro.

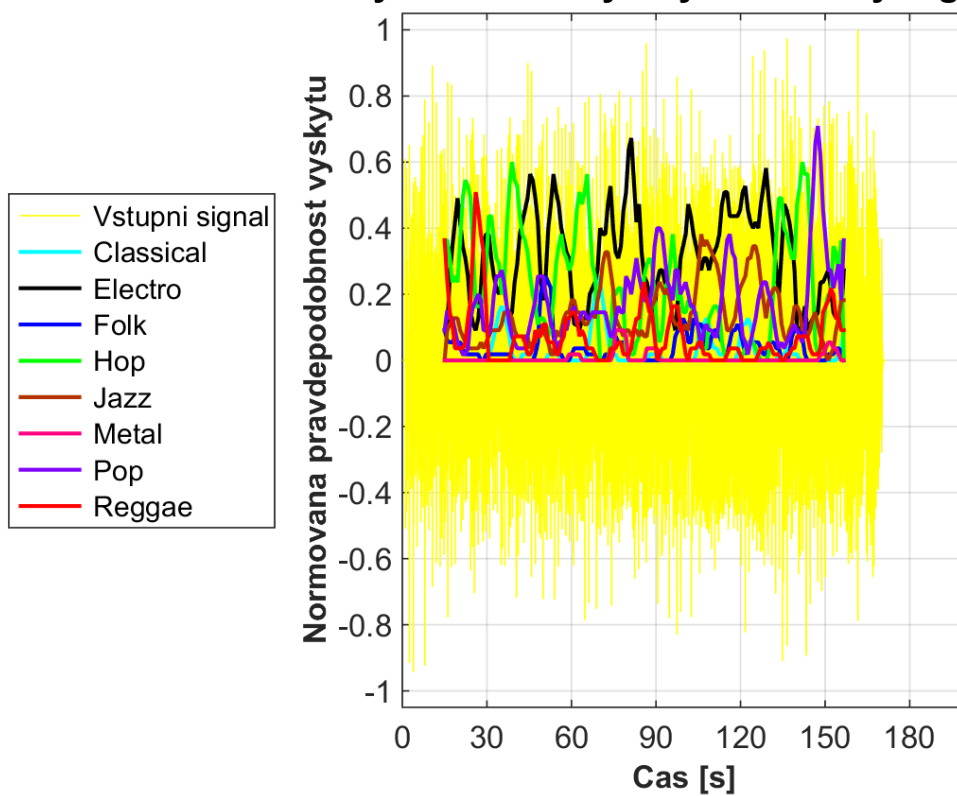


Obr. přílohy 1 Grafické vyjádření klasifikace nahrávky žánru Metal - jednoznačná klasifikace



Obr. přílohy 2 Grafické vyjádření klasifikace nahrávky žánru Pop - nejednoznačná klasifikace

### Muddy Waters Forty Days and Forty Nights.v



Obr. přílohy 3 Grafické vyjádření klasifikace nahrávky žánru Blues - neznámý žánr