



**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**

---

**Fakulta elektrotechnická  
Katedra radioelektroniky**

**Detekce hudebních žánrů pro účely masteringu gramofonových  
desek**

**Music Genre Detection For Vinyl Mastering**

Bakalářská práce

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

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

**Jaroslav Bartoš**

---

**Praha 2016**



České vysoké učení technické v Praze  
Fakulta elektrotechnická  
katedra radioelektroniky

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student: Jaroslav Bartoš

Studijní program: Komunikace, multimédia a elektronika  
Obor: Multimediální technika

Název tématu: **Detekce hudebních žánrů pro účely masteringu gramofonových desek**

Pokyny pro vypracování:

Seznamte se s problematikou klasifikace obsahu a metodami pro automatickou klasifikaci hudebních žánrů. Připravte metodu pro automatickou klasifikaci hudebních titulů podle žánru pro účely masteringu mechanického záznamu.

Výstupem bude pravděpodobnost výskytu sledovaných žánrů v testované nahrávce.

Implementovaný systém ověřte a důkladně otestujte.

Seznam odborné literatury:

- [1] Lartillot, O.; Toivainen P., A Matlab Toolbox for Musical Feature Extraction from Audio, 10th. Int. Conf. on Digital Audio Effects, Bordeaux, 2007.
- [2] Tzanetakis, G.; Cook, P., "Musical genre classification of audio signals," in Speech and Audio Processing, IEEE Transactions on , vol.10, no.5, pp.293-302, Jul 2002 doi: 10.1109/TSA.2002.800560
- [3] Koshkina, E., Identifikace obsahu archivních zvukových záznamů, bakalářská práce, FEL ČVUT, 2015

Vedoucí: Ing. František Rund, Ph.D.

Platnost zadání: do konce letního semestru 2016/2017

L.S.

doc. Mgr. Petr Páta, Ph.D.  
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.  
děkan

V Praze dne 19. 2. 2016



## **Čestné 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 26 května 2016 v Praze

.....  
podpis studenta



## **Poděkování**

Tímto bych rád poděkoval panu Ing. Františku Rundovi, Ph.D. za odborné rady a cenné připomínky v průběhu zpracování této bakalářské práce.

Také děkuji všem členům mé rodiny za trpělivost a podporu, kterou mi věnovali při zpracování bakalářské práce a po celou dobu studia.





## Abstrakt

Tato práce se zabývá automatickou detekcí hudebních žánrů. Je zde popsána základní problematika a nejčastěji aplikované principy. Práce zkoumá vliv použití různých parametrů při použití modelu *k-nearest neighbours* (KNN). Nejlepších výsledků je dosaženo pomocí *Mel frequency cepstral coefficients* (MFCC). Zároveň je navržen alternativní způsob zvýšení úspěšnosti detekce. Na základě výsledků byl vytvořen experimentální systém v prostředí Matlab za použití toolboxu MIRtoolbox. Výstupem systému je pravděpodobnost, s jakou nahrávka patří do každého ze 7 detekovaných žánrů. Rozpoznávanými žánry jsou *Rock*, *Pop*, *Jazz*, *Folk*, *Elektronika*, *Vážná hudba* a *Hip hop*. Systém byl ověřen na 70 hudebních nahrávkách a bylo dosaženo úspěšnosti 64,3%.

**Klíčová slova:** detekce, žánr, hudba, nahrávka, klasifikace, KNN, MFCC, Matlab, MIRtoolbox

## Abstract

This thesis focuses on automatic music genre detection. It describes basic problematics and most frequently applied principles. The study concentrates on influence of different parameters when using *k-nearest neighbours* model (KNN). The best solutions are achieved by *Mel frequency cepstral coefficients* (MFCC). Alternative method of increasing success rate of detection is designed too. Based on solutions, an experimental system in Matlab with toolbox MIRtoolbox was created. Output of this system is probability with which the music record belongs to each of 7 detected genres. Distinguished genres are *Rock*, *Pop*, *Electro*, *Folk*, *Jazz*, *Classical music* and *Hip hop*. The system was verified on 70 music records and reached 64.3 % success rate.

**Key words:** detection, genre, music, record, classification, KNN, MFCC, Matlab MIRtoolbox



## Seznam obrázků

Obrázek 1 Obecné schéma procesu automatické detekce hudebních žánrů .....	19
Obrázek 2 Ilustrace KNN pro 2 parametry a $k = 3$ .....	22
Obrázek 3 Zjednodušené fáze algoritmu Decision tree.....	23
Obrázek 4 Znázornění výpočtu parametru LE.....	25
Obrázek 5 Brightness signálu [27] .....	26
Obrázek 6 Roll off signálu [27] .....	26
Obrázek 7 Závislost Mel scale na frekvenci .....	27
Obrázek 8 Příklad Mel filtrů pro frekvence v rozsahu 0 – 8 kHz [33] .....	28
Obrázek 9 Časový průběh skladby po kompresi a před kompresí [34].....	30
Obrázek 10 Pravděpodobnost výskytů sledovaných žánrů v nahrávce.....	34
Obrázek 11 Soubor matic vypočítaných parametrů pro úseky nahrávek. ....	35
Obrázek 12 Vzdálenosti úseků TR od dílů TS .....	37
Obrázek 13 Ilustrační schéma hledání optimálního řešení systému .....	37
Obrázek 14 Dvanáct variant úspěšností pro různá $k$ .....	39
Obrázek 15 Ilustrační schéma zvýšení citlivosti detekce .....	40
Obrázek 16 Úspěšnost klasifikace pro různá $k$ .....	41
Obrázek 17 Schéma detekce s klíčovým parametrem.....	43
Obrázek 18 Hodnoty parametru Brightness pro jednotlivé žánry.....	43
Obrázek 19 RMS skladeb v různém období .....	45
Obrázek 20 Koefficienty změny pro vybrané nahrávky.....	45
Obrázek 21 Průměrná vzdálenost dílů TS od TR2 .....	46
Obrázek 22 Popis rozhraní GENDET .....	57
Obrázek 23 Výstup GENDET v Microsoft Excel.....	57

## Seznam tabulek

Tabulka 1 Úprava rozřazení nahrávek GZ media.....	32
Tabulka 2 Nastavení TR1 a TR2 .....	38
Tabulka 3 Nastavení TS.....	38
Tabulka 4 Výsledky varianty 12.....	39
Tabulka 5 Hledání optimálního nastavení systému .....	40
Tabulka 6 Nastavení systému dle výsledků z fáze 1 .....	42
Tabulka 7 Orientační přehled vydaných alb dle žánrů v období 2010 - 2015 .....	44
Tabulka 8 Úspěšnosti jednotlivých variant v 1. části .....	51
Tabulka 9 Výsledky úspěšností při hledání nejlepšího nastavení systému .....	52
Tabulka 10 Nahrávky použité v TR2 .....	53
Tabulka 11 Nahrávky použité v TS.....	54

# Seznam použitých zkratek

ANN - Artificial neural networks

GMM - The Gaussian mixture model

HMM – Hidden Markov model

KNN – K-nearest neighbours

LE – Low energy rate

MFCC – Mel frequency cepstral coefficients

MIR – Music information retrieval

RMS – Root mean square energy

TR - trénovací množina nahrávek

TR1, TR2 – specifické trénovací množiny nahrávek

TS1, TS2 – testované skupiny nahrávek

ZRC – Zero crossing rate

# Obsah

1	Úvod .....	15
2	Detekce hudebních žánrů.....	16
2.1	Hudební žánr .....	16
2.2	Music Information Retrieval (MIR).....	17
2.3	Automatická detekce hudebních žánrů.....	17
2.3.1	Historie a aktuální situace .....	17
2.3.2	Základní předpoklady pro detekci .....	18
2.3.3	Metody detekce.....	19
2.3.4	Klasifikační modely: .....	19
2.3.5	Parametry pro detekci žánrů.....	23
2.4	Ovlivnění detekce .....	28
2.4.1	Trénovací množina .....	29
2.4.2	Kvalita nahrávky.....	29
2.4.3	Loudness war .....	29
3	Realizace systému pro detekci hudebních žánrů .....	31
3.1	Základní informace o návrhu realizace systému .....	31
3.2	Vytvoření systému automatické detekce hudebních žánrů.....	32
3.2.1	O programu .....	32
3.2.2	Jednotlivé části programu.....	32
3.2.3	Algoritmus programu .....	34
3.2.4	Výběr skupiny parametrů a TR.....	37
3.2.5	Další upřesnění detekce.....	40
3.2.6	Shrnutí .....	41
3.3	Vylepšení systému .....	42
3.3.1	Klíčový parametr .....	42
3.3.2	Koeficient změny.....	44
3.3.3	Klasifikace nehudebních nahrávek .....	46
4	Závěr .....	47
	Citovaná literatura .....	48
	Přílohy.....	51
	Příloha A .....	51
	Příloha B.....	51
	Příloha C.....	55
	Příloha D .....	57



# 1 Úvod

Každý hudební počín má svůj specifický a jedinečný charakter. Hudební vkus se liší jak u producentů, tak i u posluchačů hudby. Mezi některými skladbami je možné najít určité souvislosti a podobnosti. Pomocí těchto paralel lze skladby třídit do skupin označovaných jako žánry. V současné době hudba již navíc neslouží pouze k zábavě a potěšení. Užívá se v mnoha oblastech díky jejímu sociálnímu a psychologickému efektu. S neustále se zvyšujícím množstvím produkované hudby roste potřeba danou hudbu třídit.

Z tohoto důvodu vznikl tento projekt, který se zabývá automatickou detekcí hudebních žánrů.

Projekt vznikl původně pouze pro potřeby vinylového průmyslu. Rostoucí poptávka po vinylových deskách vyžaduje zvýšení výrobní kapacity. S tím jsou spojeny i přípravné práce před lisováním desky, zejména speciální úpravy finálního mixu nahrávky, které je potřeba provést z důvodu fyzikálních vlastností vinylu. Je třeba aplikovat odlišné postupy pro různé žánry. Je důležité u nahrávky zjistit hudební žánr ještě před touto úpravou.

Hlavním úkolem této práce je vytvořit systém, který bude detekovat žánr nahrávky v digitální podobě ještě před finálními úpravami nahrávky. Systém může být využit nejen ve vinylovém průmyslu, ale i pro detekci žánru u jakékoliv nahrávky v digitální podobě. Systém nezařadí skladbu pouze do jednoho žánru, jeho výstupem je pravděpodobnost, s jakou nahrávka patří do definovaných žánrů. Systém nalezne využití pro třídění velkého objemu hudby, ale i pro studium jednotlivých nahrávek.

Primárně se práce zabývá automatickou detekcí hudebních žánrů (dále jen detekce) do jednoho ze sedmi žánrů pomocí programu Matlab, zejména jeho knihovny MIRtoolbox. Její výhodou je možnost přizpůsobení se konkrétním požadavkům, co se výběru žánru týče.

Tato práce vznikla ve spolupráci se společností GZ Media, a.s., která poskytla archiv audio stop a určila hudební žánry.

Nejprve se rozebere, co se rozumí pod pojmem hudební žánr, popíší se obecně nejobvyklejší metody pro detekci a některé klasifikační modely, které těchto metod využívají. Dále jsou přiblíženy vybrané parametry pro detekci, a to hlavně ty, které bude tato práce používat. Kapitola 3 se zaměří na konkrétní řešení systému detekce. Je rozdělena na 2 části. V první části se vytvoří systém pro detekci hudebních žánrů. Zjistí se, jaké je jeho nejefektivnější nastavení, jak vstupní data ovlivňují data výstupní a porovnává jednotlivé výsledky. V druhé části jsou nabídnuty alternativní postupy, kterých by mohlo být využito pro upřesnění detekce. Jedná se pouze o návrhy vylepšení, které nebyly podrobené důkladné analýze.

Vnímání hudby je velice subjektivní. Každý má své osobité cítění pro harmonii, rytmus a celkový dojem, který na něj hudební skladba udělá. Je proto důležité si uvědomit, že i přes všechny vypočítané parametry, zůstává člověk stále hlavním rozhodujícím činitelem, co se vnímání hudby týče.

## 2 Detekce hudebních žánrů

### 2.1 Hudební žánr

Definice hudebních žánrů se s vývojem hudby mění. Hudba provází člověka od počátku jeho existence. V každé kultuře se vyvíjela jinak. Tato práce se bude věnovat hlavně hudbě, jež se vyvinula z evropských uměleckých slohů nazývané souhrnně jako vážná hudba. V průběhu 20. a 21. století se postupně od vážné hudby začaly oddělovat další skupiny žánrů.

Hudebním žánrem se rozumí definovaný okruh skladeb s podobnými charakteristickými prvky, díky kterým dokáže posluchač jednotlivé skladby rozlišit. Mezi takové patří např. dynamický rozsah skladby, trvání skladby, tempo, frekvenční spektrum, použité hudební nástroje a způsob, jakým se na ně hraje, kompoziční a harmonické postupy a v neposlední řadě textový obsah [1].

Neexistuje žádné oficiální rozdělení hudby. Uvádí se různé rozdělení, viz např. v [1], [2], [3], [4].

Zajímavé rozřazení nabízí i literatura [5], která kromě dalších rozlišuje žánry pomocí činností a příležitostí, ke kterým se daná hudba hodí – např. relaxační, dovolenková či taneční, atd.

Na základě uvedených rozdělení se určily v práci za hlavní tyto žánry: Rock, Metal, Pop, Folk, Elektronika, Vážná hudba a Jazz.

Na základě doporučení GZ Media, a.s. byl žánr Metal zařazen jako součást Rocku a navíc se přidává žánr Hip hop, který má v moderní hudbě silné postavení. V roce 2014 bylo například v USA 14% prodaných alb Hip hopových [6]. Rozlišované žánry v této práci jsou tedy Rock, Pop, Jazz, Folk, Elektronika, Vážná hudba a Hip hop. Do vážné hudby se řadí též hudba filmová. Od 21. století s vývojem výpočetních zařízení zažívá velký boom žánr Elektronika, od kterého všechny nově vznikající skladby často přejímají spoustu prvků. Všeobecně nelze určit přesnou hranici jednotlivých žánrů. Předpokládá se, že k největším problémům při klasifikaci bude docházet mezi žánry Pop a Elektronika, kde většina novodobých popových skladeb obsahuje prvky Elektroniky, což je způsobeno hlavně počítačovou syntézou hudebních nástrojů. Žánr Hip hop zase využívá tzv. „samplů“, což jsou ve smyčce opakované hudební segmenty, které jsou vyjmuté z nějaké skladby jiného žánru. Často tak v Hip hopové skladbě můžeme zaslechnout například úryvek od Beethovena či motiv nějaké slavné skladby. Charakteristickým vokálním projevem v Hip hopu je rap. Tento projev může být často velice monotónní, proto se v řadě dnešních nahrávek kombinuje Hip hop a Pop. Mění se tak celková forma skladby a to také zhoršuje přesnost klasifikace.

Žánry mohou být příliš obecný pojem. Proto se občas uvádí ještě Hudební styl [7]. Žánry a styly spolu často kolidují. Tato problematika je velice složitá. Různě jsou uváděny



styly jako žánry a obráceně. Neexistuje přesná definice. Všeobecně platí, že styl je podmnožinou žánru. Zabíhat do této problematiky ale není smyslem práce. Proto se bude dále pracovat jen s žánry, které jsou definovány výše.

## 2.2 Music Information Retrieval (MIR)

Music Information Retrieval neboli MIR je vědní obor zabývající se získáváním informací z hudby. Vznikl z výpočetní muzikologie v 60. letech 20. století. Každoročně se pořádá konference ISMIR, kde se prezentují novinky z tohoto oboru. Patří sem informace ze zpracování signálů, statistiky, muzikologie a psychologie. Hlavní využití těchto informací je pro obchodní, ale i akademické účely. Jako příklad lze uvést různé hudební kolekce, systémy reagující na hudbu, převod hudebních formátů či samotná klasifikace. [8] Z MIR vychází například aplikace SoundHound, či Shazam, které dokáží identifikovat skladby na základě „odposlechu“ z okolí. Na webové stránce Youtube.com se zase MIR využívá pro navržení doporučených videí. Podobné interprety hledá pomocí MIR software s obrovskou hudební databází Spotify. Součástí MIR je i detekce hudebních žánrů.

## 2.3 Automatická detekce hudebních žánrů

Hudba je všeobecně velice komplikovaná a v některých ohledech jako například harmonické změny či tempo nepředvídatelná. Automatická detekce hudebních žánrů (dále jen detekce) je poměrně nová problematika. Začala se vyvíjet v 90. letech 20. století za využití různých postupů a vychází zejména z automatického rozpoznávání řeči. [8]

### 2.3.1 Historie a aktuální situace

V roce 1997 představuje J.T. Foote systém, který dokáže najít podobnosti mezi jednotlivými audio signály. Základem práce je využití MFCC (kap. 2.3.5). Informace z jednotlivých audio stop se shlukují do skupin. Různé nahrávky tak tvoří různé shluky. Jsou úspěšně rozpoznány různé druhy zvuků s rozdílnými úspěšnostmi (11 % - 100 %) za použití několika systémů. Zároveň je možné v reálném čase sledovat, jak se od sebe liší dva právě přehrávané zvuky [9]. V roce 1999 na práci navazují Zang a Kuo, kteří pomocí parametrů RMS, ZRC, Fundamental frequency, Timbre a rytmus vytvořili systém rozlišující 10 druhů zvuků za použití metody HMM s úspěšností 80 % [10]. Práce [11] rozlišuje v reálném čase za použití metody KNN hudební nástroje používané ve vážné hudbě. V roce 2001 pomocí GMM v práci [4] byla dosažena úspěšnost detekce 10 hudebních žánrů 62%, což je pro detekci takového počtu žánrů vysoká úspěšnost. Využita byla Autokorelační funkce, Centroid, Flux, MFCC, Roll Off, ZRC a LE. Poslední čtyři jmenované jsou použity i v této práci a jsou popsány v kap. 2.3.5.

Práce [12] kombinuje v různých variantách koeficienty MFCC, Centroid, Roll Off, Flux, ZRC a LE. Využívá hlavně modely KNN a GMM. Trénovací množina obsahuje

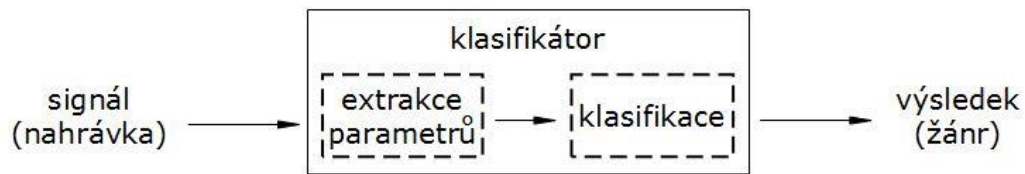
1000 nahrávek, 100 pro každý žánr a dosahuje úspěšnosti až 80 %. To je pro takto velkou trénovací množinu vysoké číslo při porovnání s [4]. Existuje celá řada dalších prací. Obecně se dosahuje maximální úspěšnosti okolo 80 %. Záleží na velikosti trénovací množiny a počtu rozlišovaných žánrů.

Nejrozšířenější metoda klasifikace hudebních žánrů je v současné době pomocí ID3 tag, či CUE sheet což jsou metadata kontejnery, které se v případě ID3 tagů používají převážně u formátu MP3 a v případě CUE sheetu u formátu wav. V těchto kontejnerech lze kromě názvu skladby, či délky trvání uložit i informaci o žánru dané skladby. Tyto informace jsou souhrnně nazývány jako tagy [13]. Tagů využívá k detekci např. software Jaikoz od firmy JThink Ltd. Výhodou je rychlost detekce. Pokud ovšem daná skladba sebou tuto informaci nenese, software je nepoužitelný. Nevýhodou je, že data musí být vkládána manuálně, což je velice časově zdlouhavé. Též se musí počítat s chybou zapříčiněnou lidskou subjektivitou. Nejedná se tedy o typickou automatickou detekci. Musí se počítat s chybou zapříčiněnou lidskou subjektivitou.

## 2.3.2 Základní předpoklady pro detekci

Cílem je vyvinout systém, kde na vstupu bude neznámá nahrávka a na výstupu se určí, do jakého žánru nahrávka náleží. Jedná se tedy o Klasifikátor hudebních žánrů. Tvorbu klasifikátoru lze rozdělit do 3 dílčích kroků - určení metody, modelu a posledním krokem je určení parametrů. Je třeba rozhodnout, zda bude detekce vycházet z předem daných předpokladů, či zda detekce proběhne přímo na základě sledování souvislostí mezi detekovanými skladbami. Tedy, zda systém již bude mít schopnost rozpoznávat nahrávky, či zda si tuto schopnost osvojí až v průběhu detekování nějaké množiny nahrávek. Obecně se pro detekci využívají tři metody. Jsou popsány v kap. 2.3.3. Tato práce bude využívat metodu učení s učitelem, principy ostatních jsou pouze zmíněné, protože nejsou pro tuto práci důležité, ale lze je použít také. Dále je třeba určit, pomocí jakého modelu bude metoda skladby zařazovat. Bude tedy třeba určit, jaký základní algoritmus systém využije. V této práci byl zvolen model *k-nearest neighbours (KNN)*. Popis tohoto a dalších nejpoužívanějších modelů je v kap. 2.3.4. Třetím krokem je volba parametrů, pomocí kterých se budou žánry vyhodnocovat. Vybrané parametry jsou popsány v kap. 2.3.5.

Klasifikátor funguje obecně tak, že nejprve analyzuje vstupní nahrávku a na základě výsledků analýzy určí pomocí algoritmu žánr nahrávky. Toho je docíleno ve dvou etapách. V první etapě se extrahují parametry z neznámé nahrávky. Pro klasifikaci hudby není jednoduché parametry vybrat. Výsledná skupina parametrů by měla spolehlivě obsáhnout ty vlastnosti, pomocí kterých se poté provede druhá etapa a tou je klasifikace. Základní schéma je na obr. 1.



Obrázek 1 Obecné schéma procesu automatické detekce hudebních žánrů

### 2.3.3 Metody detekce

**Učení s učitelem** - princip této metody spočívá v použití předem známé databáze prvků, které jsou již rozděleny do kategorií. Právě kategorie budou později rozlišované. U těchto prvků jsou známé hodnoty jednotlivých parametrů. Neznámý prvek, jenž je snaha kategorizovat, podstoupí výpočet totožných parametrů jako všechny prvky databáze. Neznámý prvek je poté za pomoci specifického modelu (kap. 2.3.4) zařazen. Výhodou tohoto postupu je možnost úpravy databáze. Nevýhodou je větší náročnost na početní výkon [14].

**Metoda učení bez učitele** - tato metoda je vhodná, pokud nejsou předem dané kategorie, ale jenom jejich počet. Z neznámé skupiny prvků jsou vypočítány určité parametry a poté se  $n$  dimenzionální prostor rozdělí na úseky, jejichž množství je dáno počtem kategorií a všechny prvky, které spadají do daného úseku jsou prohlášeny za prvky dané kategorie [15].

**Metoda posíleného učení** - metoda je postavena na pozorování. Systém učiní rozhodnutí na základě pozorování svého okolí. Pokud je ve výsledku pozorování prohlášeno jako chybné, je systém upraven váhováním pozorovaných jevů, aby bylo možné provést správné rozhodnutí příště [16].

### 2.3.4 Klasifikační modely:

Zde uvedené klasifikační modely se úspěšně používají v audio oblasti zejména při detekci řeči. Pro detekci hudebních žánrů však existují omezené zdroje, proto jejich implementace do systému detekce hudebních žánrů může být v některých případech velice složitá. V této části jsou uvedeny jejich obecné principy.

#### The Gaussian mixture model (GMM)

The Gaussian mixture model GMM neboli Model směsí Gaussovských křivek je stochastickým modelem. Je používán zejména v systémech rozpoznávání řeči a je založen na metodě učení bez učitele. Využívá Gaussovy funkce ve významu hustoty pravděpodobnosti.

Každá kategorie je zastoupena více Gaussovskými křivkami, tedy výsledná hustota pravděpodobnosti je směsí těchto křivek. Směs je poté vážená suma definovaných parametrů.

Proces učení probíhá tak, že algoritmus hledá parametry modelu, dokud nedosáhne definované prahové hodnoty [17]. Využívá se k tomu algoritmus *Expectation Maximization*, který postupně přehodnocuje parametry, zmenšuje prostor, který je definuje, až do finálního vytvoření kategorií [18].

Výhodou GMM je to, že model má malý počet parametrů a je tedy rychlý. Nevýhodou je to, že algoritmus může přestat pracovat při větším počtu dimenzí signálu [19].

## **Artificial neural networks (ANN)**

Tato metoda je analogie k lidské nervové soustavě, kde jsou informace šířeny a předávány přes neurony. R. Hecht –Nielsen definoval nervovou síť jako „Výpočetní systém, tvořený množstvím elementárních vysoce propojených elementů, které zpracovávají informace jejich dynamickým vztahem k externím vstupům.“ Tato metoda, konkrétně typ FeedBack je použita v [20].

ANN jsou tvořeny jednotlivými uzly, které imitují neurony lidského mozku. Uzel může přijmout vstupní data a provést elementární funkci. Výsledek funkce je distribuován do dalších uzlů.

Jednotlivé buňky vykonávají elementární funkce. Uzly jsou propojeny cestami, kdy každá cesta má svoji určitou váhu.

Existují 2 typy ANN - FeedForward a FeedBack.

U FeedForward jsou distribuována jen jedním směrem. Jsou zde neměnné vstupy a výstupy. Ve FeedBack se používají zpětné smyčky, které mohou upřesnit výsledky [16].

ANN je schopno učení, ale je třeba před detekcí trénovat. Dle [16] je možné využít všechny 3 metody učení (viz. kap. 2.3.3).

Výhodou je, že algoritmus umožňuje efektivně tvořit nelineární závislosti, které přesně popisují množinu dat, je velice rychlý, je odolný vůči šumům vstupních signálů – systém určí chybovost signálu a přiřadí mu příslušnou váhu, či v případě naprosté chyby signál vyřadí úplně.

Hlavní nevýhodou je složitá implementace, kdy je třeba popsat všechny možnosti, jakých mohou nabývat jednotlivé uzly. Na každý problém musí být vytvořena speciální architektura [17].

## Hidden Markov model (HMM)

Jedná se o zvláštní případ stochastických konečných automatů. Využívá tzv. Markovské řetězce, což jsou posloupnosti náhodných proměnných, kdy následující hodnota proměnné je určena pouze hodnotou předchozí proměnné a nezávislá na předchozích stavech. Tento model je často využíván např. při rozpoznávání řeči [21].

Princip HMM lze zobecnit na problému urn. V prostoru, do kterého pozorovatel nevidí, jsou urny  $X_1$ ,  $X_2$  a  $X_3$ . V každé urně je známý počet předmětů  $y_k$ , které jsou označeny  $y_1, y_2, y_3, y_4$ . Je náhodně vybrána jedna z urn a z ní vybrán předmět, jenž je exportován ven z prostoru, do kterého pozorovatel nemá přístup. Tímto způsobem se postupuje dále. Pozorovatel tedy nevidí posloupnost urn, z jakých byl předmět vybrán, vidí pouze posloupnost vybraných předmětů. Výběr urny pro vytažení  $k$ -tého předmětu závisí pouze na náhodném čísle a na výběru urny pro vytažení  $(k-1)$ -ho předmětu. Výběr předmětu není závislý na pořadí předchozích urn, ale pouze na urně, ze které byl vytažen  $(k-1)$ -tý předmět.

Markovův proces tedy nemůže být pozorován – proto se nazývá Skrytý Markovův proces. Z každého stavu může být vytažen pouze jeden předmět. Ačkoliv pozorovatel zná rozmístění urn a posloupnost tažených předmětů, neví, ze které urny byl vytažen poslední předmět. Lze určit pouze pravděpodobnost, s jakou byl vytažen z jednotlivých urn [22].

Výhodou HMM je to, že není nutné pochopení vnitřního procesu dynamických změn v systému. Umožňuje simulace dlouhých časových závislostí.

Největší nevýhodou při práci s informací o audio signálu je problém zpracování velkého množství dat, je třeba velký objem paměti pro uchování parametrů modelu a trénovací množinu [17].

## K-nearest neighbours (KNN)

KNN neboli  $k$ -nejbližších sousedů je metoda klasifikace pro učení s učitelem. Vzhledem ke svým vlastnostem je vhodná pro použití v této práci, proto je rozebrána důkladněji. Neznámé prvky se klasifikují do 2 a více kategorií z předem definované trénovací množiny, která obsahuje prvky již zařazené do kategorií. Prvky jsou definované vektory [23].

### Postup klasifikace:

Je vytvořena trénovací množina tvořená z  $P$  prvků. Každý prvek zastupuje jednu z  $N$  kategorií. Každému prvku je přiřazen vektor  $v = \{v_1, v_2, \dots, v_m\}$ , který definuje hodnoty jednotlivých parametrů. Výsledkem je  $m$ -dimensionální prostor.

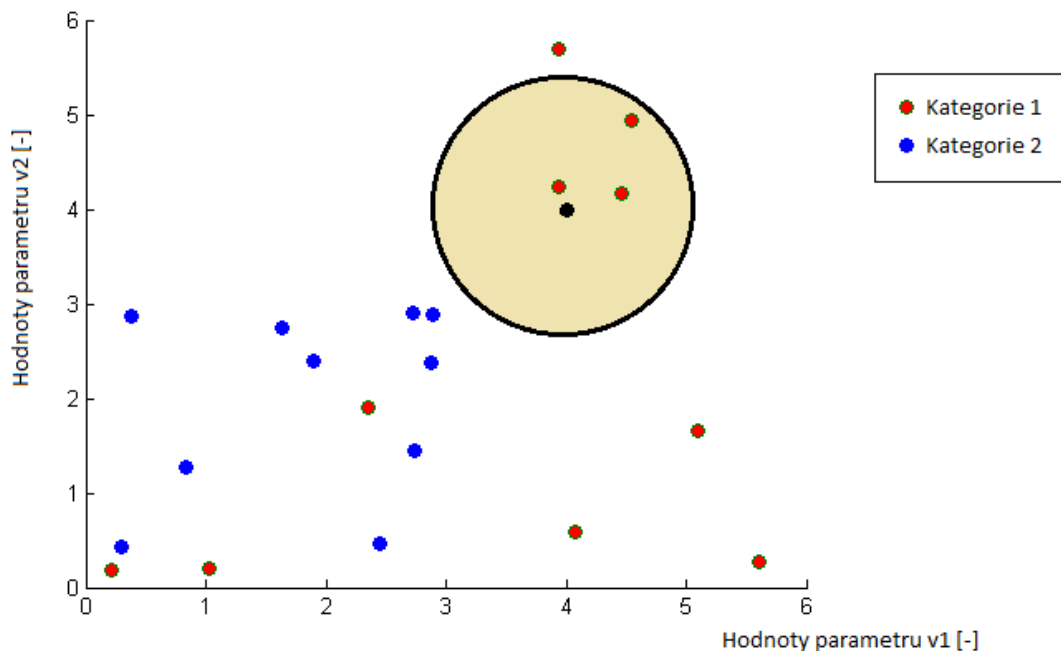
Neznámému prvku  $P_x$  je vytvořen vektor  $x = \{x_1, x_2, \dots, x_m\}$

Prvek je umístěn do stejného prostoru jako trénovací množina a klasifikace probíhá metodou nalezení  $k$ -nejbližších sousedů, kde  $k$  je právě počet nejbližších sousedů. Prvek je zařazen do té kategorie, s nejvíce sousedy.

Pro hledání se nejčastěji používá Euklidovská metrika, která je daná vztahem

$$D(x, v) = \sqrt{\sum_{i=1}^m (x_i - v_i)^2} \quad (1)$$

kde  $x = \{x_1, x_2, \dots, x_m\}$  a  $v = \{v_1, v_2, \dots, v_m\}$  jsou vektory popisující parametry kategorií. [24]



Obrázek 2 Ilustrace KNN pro 2 parametry a  $k = 3$

Na obr. je 2-dimenzionální prostor. Modré body jsou prvky kategorie 1 a červené jsou kategorie 2. Černý bod je neklasifikovaný. Při zvolení hodnoty  $k = 3$  (pro lepší znázornění jsou v ohraničeném kruhu) se naleznou pomocí vztahu(1) nejbližší sousedi. V tomto případě jsou to všechno prvky kategorie 1. Neznámý bod je tedy klasifikován jako bod kategorie 1.

Výhodou této metody je konkrétní implementace, malý počet vstupních parametrů a možnost jednoduché změny  $k$ . Nevýhodou je velká výpočetní náročnost v případě obsáhlé trénovací množiny.

## Decision Tree

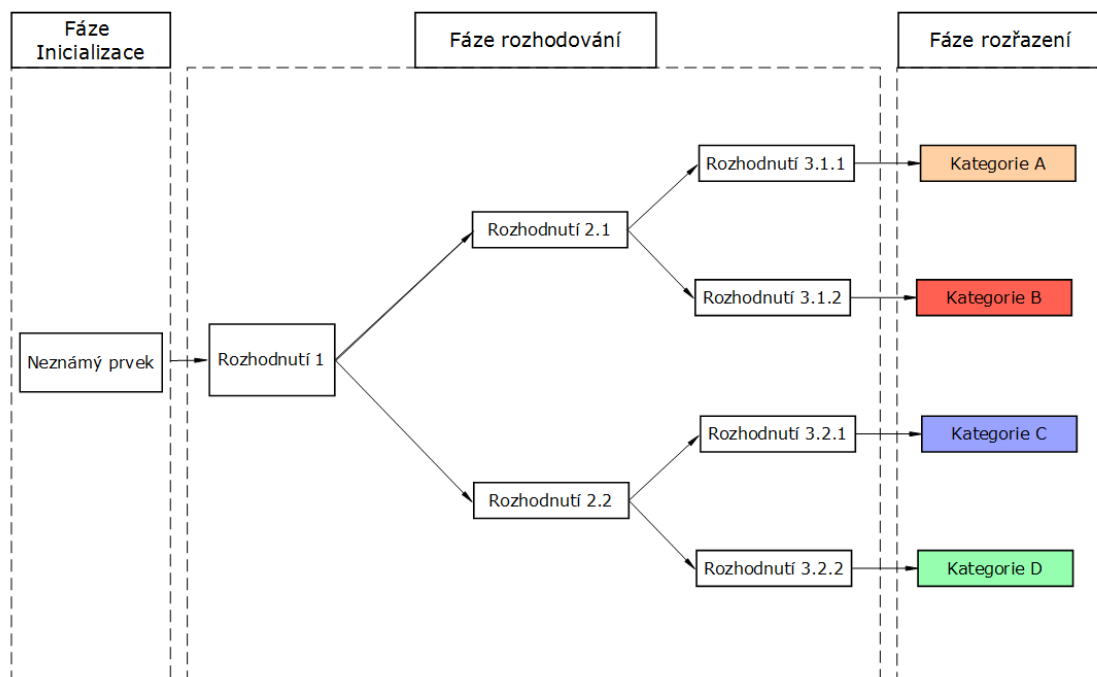
Tento model funguje na principu postupného rozhodování. Problém se řeší deduktivně od primárního rozdělení za předem definovaného výběru možností až ke konečné větvi. V případě klasifikace je při posledním rozhodnutí určena kategorie. [25] Pro klasifikaci hudby je využíván např. v [26].

Lze rozlišit 3 základní fáze algoritmu:

Fáze inicializace – je vybrán neznámý prvek

Fáze rozhodování – prvek postupně prochází řadou deterministických rozhodnutí

Fáze zařazení – prvek je zařazen do kategorie



Obrázek 3 Zjednodušené fáze algoritmu Decision tree

Výhodou této metody je jednoduchost a přehlednost rozložení faktů daného problému. Umožňuje analyzovat všechny možnosti.

Nevýhodou je potřeba znát celkovou problematiku. Při řešení složitého problému tak může docházet k nepřehlednostem [25].

### 2.3.5 Parametry pro detekci žánrů

Parametrů, které lze využít k detekci hudebních žánrů existuje nespočetné množství, jejich výčet a podrobný popis by zabral celou knihu. Tato práce se bude zabývat jen někte-

rými vybranými, které budou implementovány v prostředí Matlab pomocí toolboxu MIR-toolbox.

Parametry lze zařadit do několika oblastí:

**Spektrální-** MFCC, **Brightness**, Spectral flux, **Roll off**, Centroid a další...

**Časová** – Tempo, Dynamický průběh, **ZRC**, **RMS**, **LE** a další...

**Harmonická** – Harmonické změny, Modalita, Pitch, Harmonicity, Tonalcentroid a další...

Všechny zde vybrané parametry nabízí MIRtoolbox, Jejich popis je dohledatelný v manuálu MIRtoolbox. [27]

Pozn.: Tučně zvýrazněné parametry budou použity pro detekci v této práci.

## Zero crossing rate (ZRC)

Zero crossing rate je parametr, který popisuje, kolikrát signál  $s(k)$  v počtu vzorků projde nulou, tedy změni znaménko. Pro klasifikaci hudby se používá např zde: [4] Můžeme ho definovat jako

$$ZCR = \frac{1}{N-1} \sum_{k=1}^{N-1} f\{s(k) \cdot s(k-1) < 0\} \quad (2)$$

kde  $N$  je počet vzorků a  $f$  je logická funkce která při splnění dané podmínky vrací 1 a při jejím nesplnění vrací 0.

Zero crossing rate se využívá převážně při detekci řeči, jeho využití je ovšem i v ostatních odvětvích zpracování signálu, zejména pro jeho jednoduchou implementaci. Důležité je zvolit dostatečnou vzorkovací frekvenci, aby nedocházelo k chybám při průchodu nulou [28].

## RMS energy (RMS)

RMS neboli Root Mean Square energy vyjadřuje energii signálu. Jedná se o druhou odmocninu aritmetického průměru druhých mocnin daných hodnot signálu v určitém časovém úseku. Pro signál  $s$  délky  $N$  vzorků vyjádříme RMS:

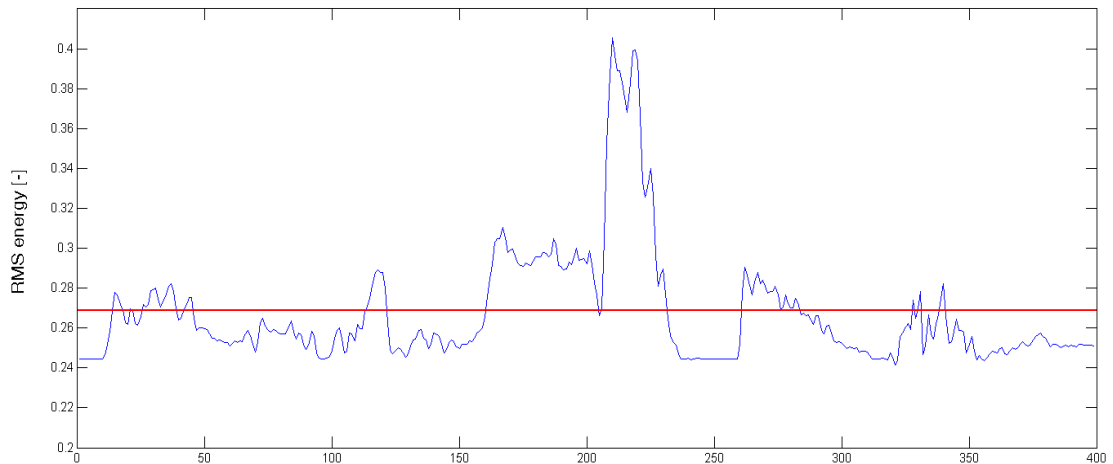
$$RMS = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N s(i)^2} \quad (3)$$

Pomocí nalezení RMS v krátkých časových úsecích lze získat časový rozvoj energie signálu. Tento průběh se nazývá obálka signálu [29].



## Low energy rate (LE)

Parametr Low energy rate udává, kolik procent signálu má energii nižší než je průměrná energie. Pro ilustraci je uvedený náhodný signál o délce 10 s, který byl rozdělen do 400 úseků a z každého úseku bylo vypočítáno RMS. Byla tedy získána obálka signálu. Červená čára značí průměrnou hodnotu RMS celého signálu, která je rovna 0,269 [-].



Obrázek 4 Znáznornění výpočtu parametru LE

Z grafu je patrné, že větší část signálu leží pod průměrnou hodnotou RMS. Výpočet LE je následující:

$$LE = \frac{\text{počet úseků RMS} < \text{průměrná hodnota RMS}}{\text{počet úseků RMS}} \quad (4)$$

Po dosazení

$$LE = \frac{270}{400} = 0,675 [-]$$

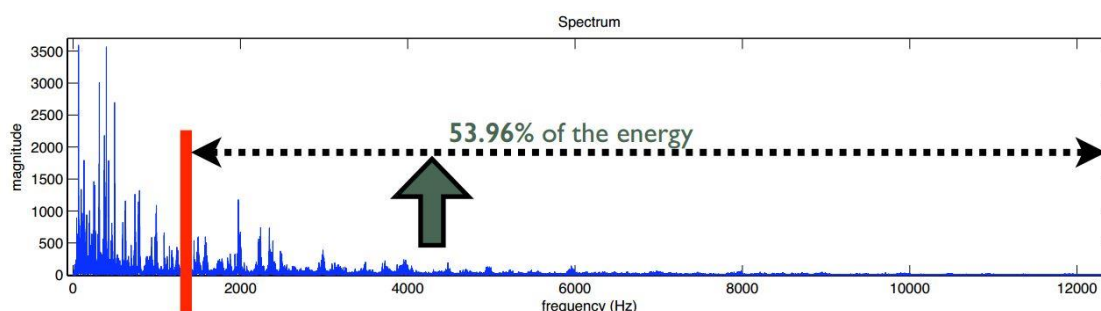
Tedy 67,5 % signálu leží pod průměrnou energií signálu [27]. Používá se například v [4], [12].

## Brightness

Parametr Brightness, definuje barvu zvuku. Zvuky s vyšší hodnotou energie na vyšších frekvencích mají také vyšší hodnotu Brightness.

Brightness informuje o tom, kolik energie se nachází nad určitou frekvencí  $f_c$ . Hraniče se nastavuje obvykle na 1500 Hz. V některých případech je možno setkat se též s hodnotami 1000 Hz či 3000 Hz [27].

Na obr. 5 je znázorněno spektrum náhodného signálu. Za kmitočtem 1500 Hz se nachází 53,96 % energie, což je právě hodnota Brightness.

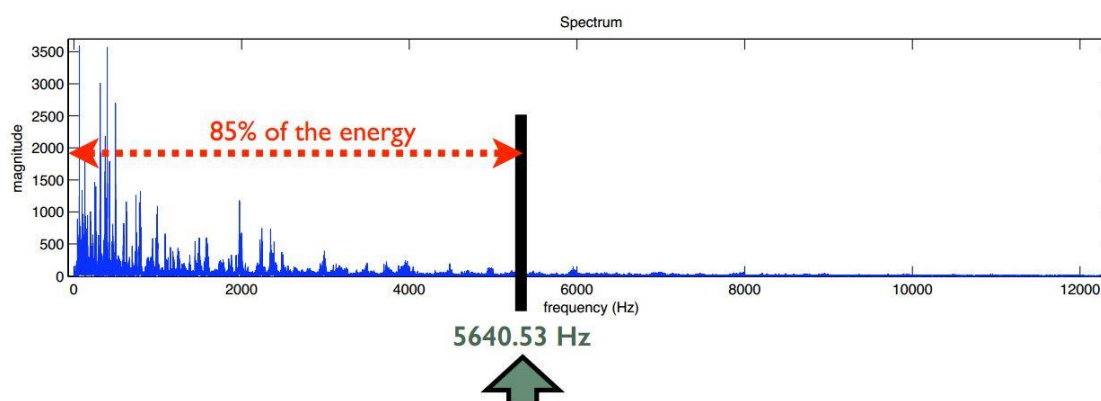


Obrázek 5 Brightness signálu [27]

## Roll off

Parametr Roll off je podobný parametru Brightness. Zjišťuje mezní kmitočet  $f_r$ , kdy je dosaženo 85 % energie signálu, ale používá se i hodnota 95% [4]. Zvuk s vyšší energií na nižších kmitočtech bude mít tedy mezní kmitočet  $f_r$  nižší než zvuk s vyšším energetickým zastoupením vyšších frekvencí.

Na obr. 6 je stejný signál jako pro parametr Brightness. Hraničním kmitočtem  $f_r$  pro 85 % energie a tedy Roll off je kmitočet 5640,53 Hz.



Obrázek 6 Roll off signálu [27]

Je použit např. v pracích [4], [12].

## MFCC

Mel frequency cepstral koeficienty (MFCC) jsou velice často a běžně používány v automatické detekci řeči (ASR) [30] a také hudby [12]. Byly představeny v 80. letech 20. století Davisem a Mermelsteinem.

Audio signál se v průběhu času neustále mění, proto je třeba jej rozdělit na velice krátké *úseky* za předpokladu, že tyto *úseky* se statisticky nebudou tolik měnit. Audio signál se tedy rozdělí na 20 - 50 ms *úseky*. Pokud je *úsek* menší, není k dispozici dostatek vzorků k získání spolehlivé spektrální charakteristiky, pokud je delší, mění se signál v daném *úseku* až příliš [4]. Literatura [31] ovšem jako klíčovou délku uvádí 15 ms a literatura [32] uvádí hodnotu 20 ms.

Dále se provede výpočet výkonového spektra každého *úseku*, čímž je zjištěno, jaké frekvence jsou zastoupeny v jednotlivých úsecích.

Na spektrum aplikuje série tzv. Mel filtrů. Obvykle se používá 26 – 40 filtrů. První filtr je velice úzký a dává informace o tom, kolik energie se nachází v okolí 0 Hz, postupně se filtry s rostoucí frekvencí rozšiřují. Jak široké a jak vzdálené filtry mají být říká Mel scale.

### Mel scale

Mel scale je percepční stupnice výšek tónů, které mají mezi sebou stejnou vzdálenost, což je založeno na úsudku posluchačů. Referenční bod mezi Mel scale a normálním měřením frekvence je definován přiřazením percepční výšky 1000 mels k 1000 Hz.

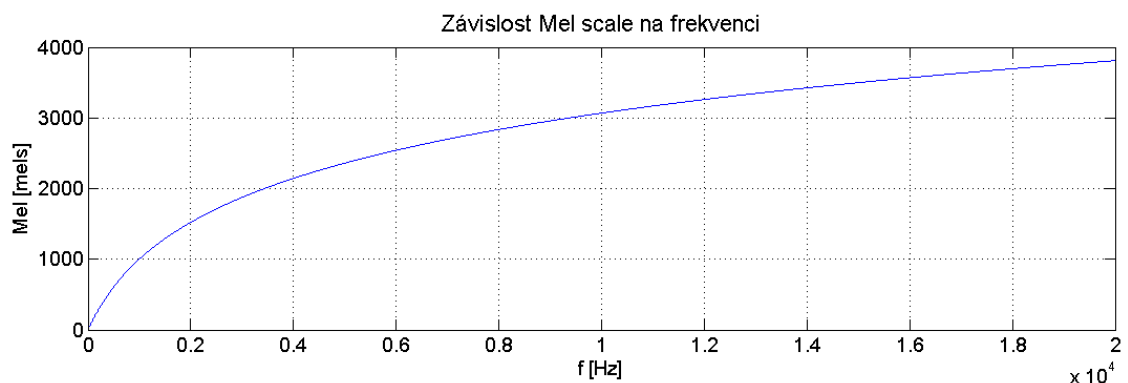
Člověk je více všímavý, co se týče malých změn výšky tónu u nízkých frekvencí, než na vyšších frekvencích. Toto měřítko se tedy přibližuje k lidskému sluchu.

Vzorec pro konverzi frekvence na Mel scale:

$$M(f) = 1125 \ln \left( 1 + \frac{f}{700} \right) \quad (5)$$

Vzorec pro konverzi Mel scale na frekvenci:

$$M^{-1}(m) = 700 \left( e^{\frac{m}{1125}} - 1 \right) \quad (6)$$



Obrázek 7 Závislost Mel scale na frekvenci

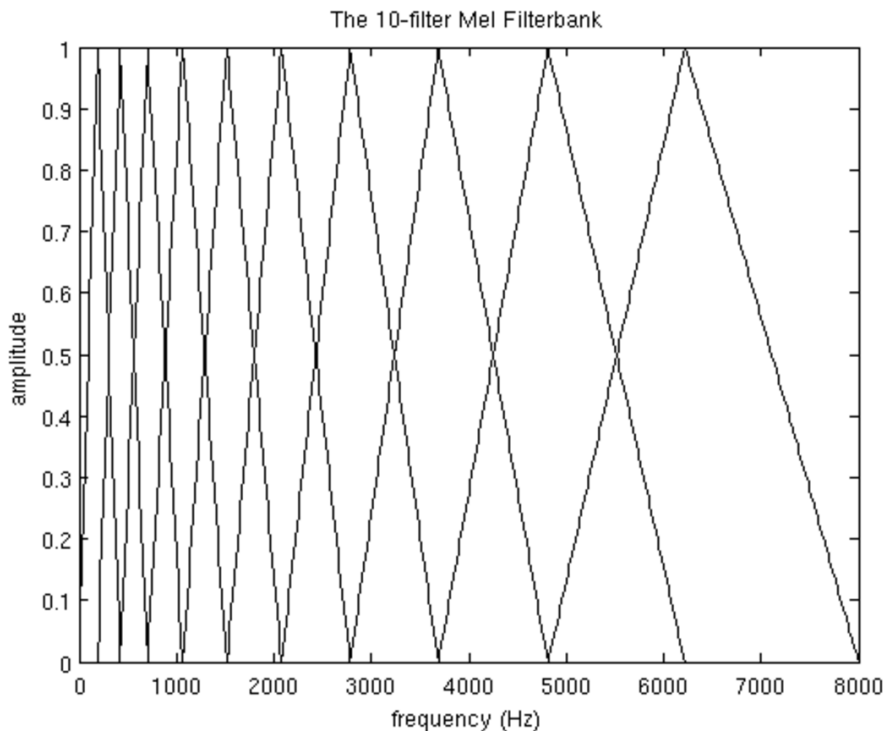
Frekvenční rozsah hudby v rozmezí 20 – 20 000 Hz odpovídá rozsahu Mel scale 31,7 – 3810,2 mels. Při použití  $n$  počtu filtrů je třeba mezi rozmezí Mel scale vložit lineárně celkem  $n$  hodnot. Celkem je tedy  $n+2$  hodnot v Mel scale. Mel filtry jsou trojúhelníkové filtry rozděleny podle Mel scale tak, že  $n$ -tý filtr začíná v  $n$ -té hodnotě Mel scale, jeho vrchol je v  $(n+1)$ . hodnotě a končí v  $(n+2)$ . hodnotě. Další filtr začíná v  $(n+1)$ . hodnotě, atd. [33]

### Koeficienty

Následně se vypočítá energie spektra v každém z filtrů. Dále se vypočítá logaritmus těchto energií z toho důvodu, že lidské ucho nevnímá úroveň hlasitosti lineárně.

Posledním krokem je výpočet DCT z logaritmu filtrovaných energií a to hlavně z důvodu překrývání filtrů. Energie filtrů jsou mezi sebou částečně korelované.

MFCC koeficienty jsou amplitudy výsledného spektra [33].



Obrázek 8 Příklad Mel filtrů pro frekvence v rozsahu 0 – 8 kHz [33]

## 2.4 Ovlivnění detekce

Detekce může být ovlivněna mnohými vlivy. Zde jsou uvedené pouze ty, které jsou pro tuto práci považovány za nejdůležitější.

## 2.4.1 Trénovací množina

Vzhledem k rozsáhlosti hudby je téměř nemožné zvolit trénovací množinu tak, aby bezpečně obsáhla všechny žánry a styly (problematika hudebních žánrů a stylů je rozebrána v kap. 2.1). Vždy bude existovat nějaká nahrávka, kterou systém na základě trénovací množiny nebude schopen korektně detekovat. Pokud bude trénovací množina příliš obsáhlá, zvýší se čas, po který bude detekce probíhat. Je tedy důležité zvolit trénovací množinu, aby obsáhla co nejširší záběr nahrávek, ale zároveň neobsahovala nahrávek příliš. Je vhodné volit nahrávky tak, aby byli zastoupeni rovnoměrně zástupci všech žánrů, popřípadě stylů.

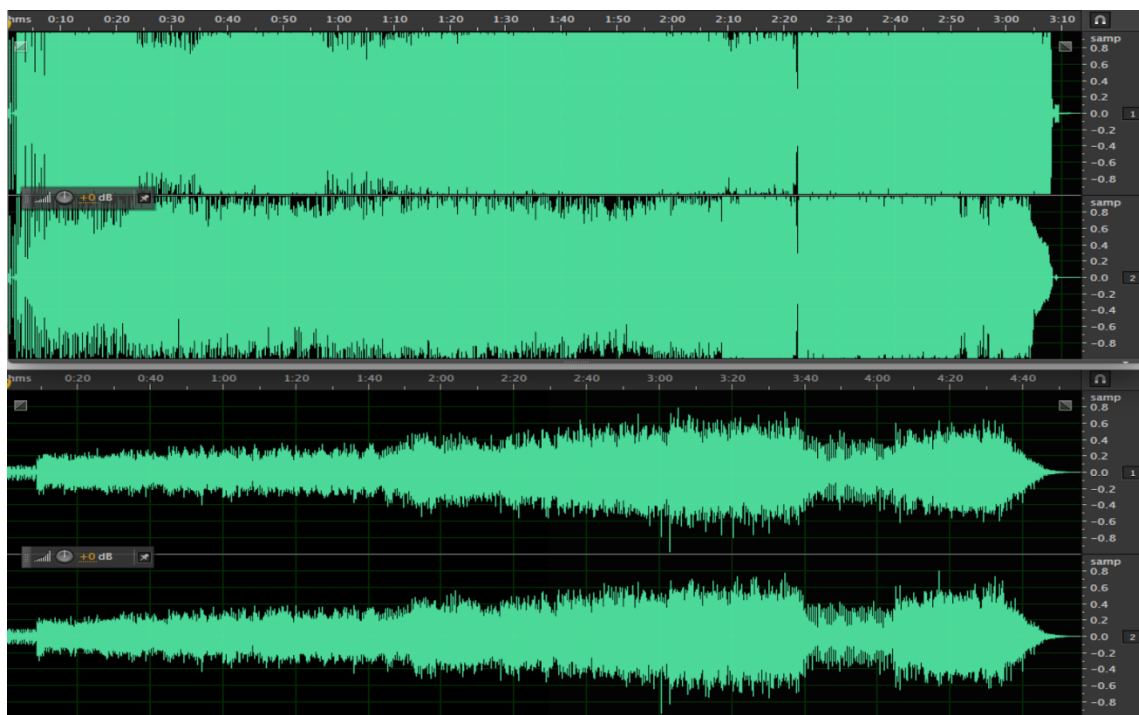
## 2.4.2 Kvalita nahrávky

V průběhu let se měnil způsob, jakým byly nahrávky vytvořeny. Zlepšovala se technika a tedy i kvalita nahrávek. To ovlivní detekci, protože nahrávky pořízené v ranějších dobách zní odlišně od novějších nahrávek. Rané nahrávky obsahují také větší podíl nežádoucích signálů. Tyto signály jsou např. šum, lupání či praskot. Před nástupem digitální techniky byly nahrávky uchovávané zejména na mechanických nosičích, např. na vinylových deskách, ve kterých se tyto vady častěji vyskytují a mění celkový charakter nahrávky.

Nahrávky mohou být profesionální a neprofesionální. Profesionální nahrávkou se rozumí nahrávka vytvořená v profesionálních hudebních studiích. Neprofesionální nahrávka byla vytvořena laikem, její kvalita je nízká a nesplňuje obecné zásady využívané v profesionální sféře. Vliv rušivých elementů v neprofesionálních záznamech může převládat nad charakteristickými znaky daných žánrů, což může vést k chybné detekci. Dále lze rozlišovat „živé“ a studiové nahrávky. „Živou nahrávkou“ se rozumí záznam z koncertu. Studiová nahrávka vznikla ve studiu. V této práci se předpokládá, že detekované nahrávky budou nahrávky vytvořené v profesionálních hudebních studiích. Toto je třeba zohlednit v trénovací množině (viz 2.4.1).

## 2.4.3 Loudness war

Od počátku 90. let 20. století po rozmachu komerčních rádií bylo zjištěno, že posluchače více zaujme nahrávka, která mu zní hlasitěji. Proto spolu jednotlivá vydavatelství a radia začala vést boj označovaný jako „loudness war“ za účelem vytvořit co nejhlasitější nahrávku. Hudební přehrávače ovšem dokáží přehrát jen signál do určité intenzity, proto ji nelze zvyšovat donekonečna. Po překročení prahu maximální amplitudy, kterou je zařízení schopné přehrát, dochází ke zkreslení výsledného zvuku, což degraduje jeho kvalitu. Nahrávky se proto začaly upravovat tak, aby se v každém okamžiku co nejvíce blížil signál maximu. Nejhlasitější místa jsou upravována jen minimálně, zato místa s nejmenší amplitudou jsou zesílena na maximální úroveň. Tomuto jevu se říká dynamická komprese.



*Obrázek 9 Časový průběh skladby po kompresi a před kompresí [34]*

Na obr. 9 je dole nahrávka před dynamickou kompresí a nahoře po kompresi. Kompresí vytvořila z časového průběhu „obdélník“, což způsobí, že skladba bude působit hlasitěji a ve všech místech bude mít stejnou úroveň. To ovšem zničí její dynamický rozsah. Od roku 1982 se RMS produkovaných nahrávek průběžně zvyšuje [35]. Parametry pro detekci hudebních žánrů, jako dynamický rozsah či RMS nemají velký vliv na rozpoznání žánrů nahrávek, které prodělaly dynamickou kompresi, protože v těchto případech budou jejich hodnoty podobné. Jejich použití je přesto vhodné, protože nesou informaci o tom, zda byla komprese provedena. U ranějších nahrávek se tyto parametry liší.

Pozn.: Kapitoly 2.4.1 a 2.4.2 vznikly na základě zkušeností autora této práce.

# 3 Realizace systému pro detekci hudebních žánrů

## 3.1 Základní informace o návrhu realizace systému

K detekci hudebních žánrů je třeba využít předem známou skupinu skladeb, bude tedy nejlepší použít metodu učení s učitelem. Pro tuto metodu je vhodné použít například model KNN. Protože je tento model jednoduchý, snadno upravitelný a s malým počtem vstupních parametrů, bude použit i v této práci.

System detekce byl v této práci vytvořen v prostředí Matlab (verze 2015b) s využitím toolboxu MIRtoolbox (verze 1.6). V této práci bude testována efektivita systému detekce hudebních žánrů při kombinaci různých parametrů a nastavení. Bude operováno s malým počtem dat. Výsledky se tedy mohou lišit, pokud se použijí objemnější data. Cílem práce je najít postup pro co nejefektivnější klasifikaci dané množiny dat. Tento postup bude poté aplikovatelný i na jiné množiny. Jednotlivé detaily nastavení se však mohou změnit.

Problém byl rozdělen na 2 části. V první části je popsán základní princip programu, který řeší navržený systém, dále se určí, jaké parametry je vhodnější použít a jaké volit optimální nastavení programu. Výsledkem této fáze je funkční systém, který bude schopen detekce. Ve druhé části se naváže na výsledky první části, bude se experimentovat se zvýšením úspěšnosti detekce.

### Definice pojmů

Jelikož se v tomto oboru nevyskytuje přesně daná terminologie, bylo třeba pro účely práce definovat některé pojmy.

**stopa, nahrávka** – signál, který reprezentuje hudbu či ostatní zvuky v systému automatické detekce hudebních žánrů

**fragment** – část stopy

**úsek** – část *fragmentu*, ze které se počítají parametry trénovací množiny

**díl** - část *fragmentu*, ze které se počítají parametry testovací množiny

**hluché místo** – nežádoucí část stopy, nevhodná pro detekci žánru – např. ticho, potlesk, smích...

## 3.2 Vytvoření systému automatické detekce hudebních žánrů

V první části je popsán algoritmus programu, vytvoření základního systému pro detekci hudebních žánrů a ověřena jeho funkčnost.

### 3.2.1 O programu

Program se skládá z několika částí. Obsahuje 2 databáze stop, z kterých se vytvoří trénovací množina *TR1* a *TR2*. Z těchto množin se počítají skupiny parametrů *A* a *B*. Dále obsahuje testovací množinu *TS*, ve které se také počítají skupiny parametrů *A* a *B*. Zahrnuje také vyhodnocovací část, kde pomocí metody *KNN* dojde k zařazení neznámých skladeb do žánrů.

### 3.2.2 Jednotlivé části programu

#### Trénovací množina TR1

Tato trénovací množina je tvořena archivem audio stop poskytnutých od GZ Media, a. s. Je zde zahrnuto celkem 188 stop o celkové délce 47 hodin rozdělených do 7 žánrů. Průměrná délka stopy je 15 minut. Znatelně se liší celková délka jednotlivých žánrů. Největším žánrem je Rock s celkovou délkou stop 14,8 hodiny a nejkratší je Hip Hop o celkové délce 3,5 hodiny. Každá stopa představuje soubor nahrávek určených pro 1 stranu gramofonové desky. Poskytnutý archiv do skupin rozdělených nahrávek byl rozdělen do definovaných žánrů dle následující tabulky.

Tabulka 1 Úprava rozřazení nahrávek GZ Media, a. s.

Zařazeno do žánru	Rock	Pop	Jazz	Folk	Elektronika	Vážná hudba	Hip Hop
Styly v archivu GZ media	Punk	R&B	Swing	Country	Dance	Soundtrack	Rap
	Metal	Funk	Latin		Electro		
	Ska	Disco	Blues		Techno		
	Reggae				Chiptunes		
	Post Rock						
	Indie rock						
	Hardcore						
	Rock'N'Roll						
	Industrial						



## Trénovací množina TR2

TR2 obsahuje 142 skladeb o celkové délce 11 hodin. Celková délka každého žánru je větší než 1,5 hodiny. Stopy reprezentují skladby, které byly zvoleny jako charakterističtí zástupci jednotlivých žánrů. Jména vybraných stop a interpretů jsou uvedena v příloze B v tabulce 10.

## Testovací množina TS

Testovací množina TS je tvořena celkem 70 stopami, v zastoupení po 10 pro každý žánr. Byla vybírána náhodně ze všeobecně známých skladeb tak, aby každá nahrávka byla reprezentativní vzorek žánru. Jména jednotlivých skladeb a interpretů jsou uvedena v příloze B v tabulce 11.

## Skupiny parametrů A a B

Obě skupiny parametrů byly počítány pomocí volně dostupné knihovny matlabu MIRtoolbox 1.6. [36]. Počítají se jak pro TR1, TR2, tak i pro TS. Parametry byly voleny hlavně na základě literatury [4], [12], [24], [27], [28] a [32].

## Parametry A

Parametry ve skupině A byly voleny na základě počtu jejich výstupů. Každý parametr má pouze jeden výstup, což umožňuje snadnou implementaci. Dalším důvodem volby skupiny A je ten, že každý parametr je možný aplikovat na každý *úsek*. Z každého *úseku* se tedy vypočte vždy jedna hodnota pro každý parametr. Je celkem 5 parametrů, vznikne tak 5dimensionální prostor, s kterým bude pracovat *KNN*.

Do skupiny A patří tyto parametry: Zero crossing rate (ZRC), Root mean square energy (RMS), Low energy rate (LE), Brightness a Roll off.

### Konkrétní informace o nastavení parametrů

**ZRC** hodnota se z *úseku* přepočítává vždy vzhledem k 1 vteřině.

**Brightness** hodnota frekvence Cut off byla nastavena na 1500 Hz.

**RMS** se vypočítá tak, že se nejprve provede rozdělení na 50 ms dílce, z kterých se vypočítají hodnoty RMS. Tyto hodnoty se poté sečtou.

U **LE** je každý *úsek* nejprve rozdělen na 50 ms dílce, z kterých je poté pomocí vztahu (4) určena hodnota LE.

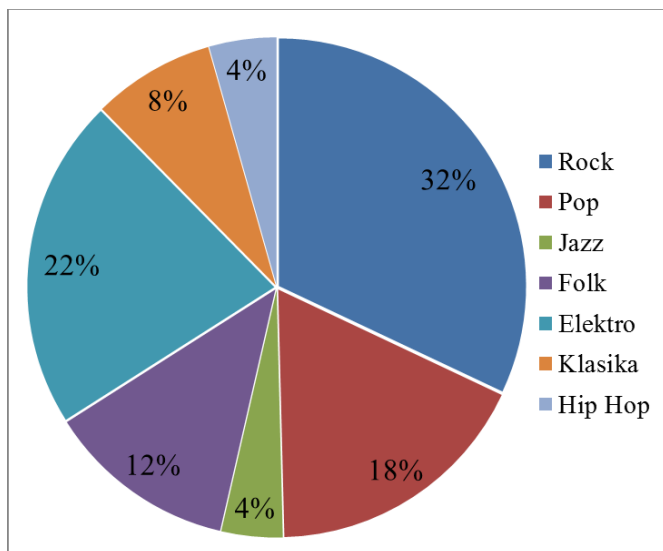
Hodnotě **Roll Off** odpovídá mezní kmitočet  $f_r$  pro 85 % energie signálu.

## Parametry B

Ve skupině parametrů B se počítají Mel Frequency Cepstral coefficients (MFCC). Literatura [33] uvádí 12 koeficientů, literatura [32] uvádí 15 koeficientů. Jako kompromis bude tato práce počítat s 13 koeficienty MFCC a to také z důvodu velké výpočetní náročnosti pro větší počet koeficientů. Vzniká tak 13dimensionální prostor parametrů, s kterým bude pracovat *KNN*.

## Výstup programu

Pomocí metody *KNN* se vyhodnotí vzdálenosti množiny parametrů TR od TS a najde se *k* nejbližších sousedů. Výsledkem je podílové rozdělení členů TS do žánrů tak, že člen s největším podílem určitého žánru bude s největší pravděpodobností do daného žánru patřit. Ilustrační znázornění je na Obrázek 10.



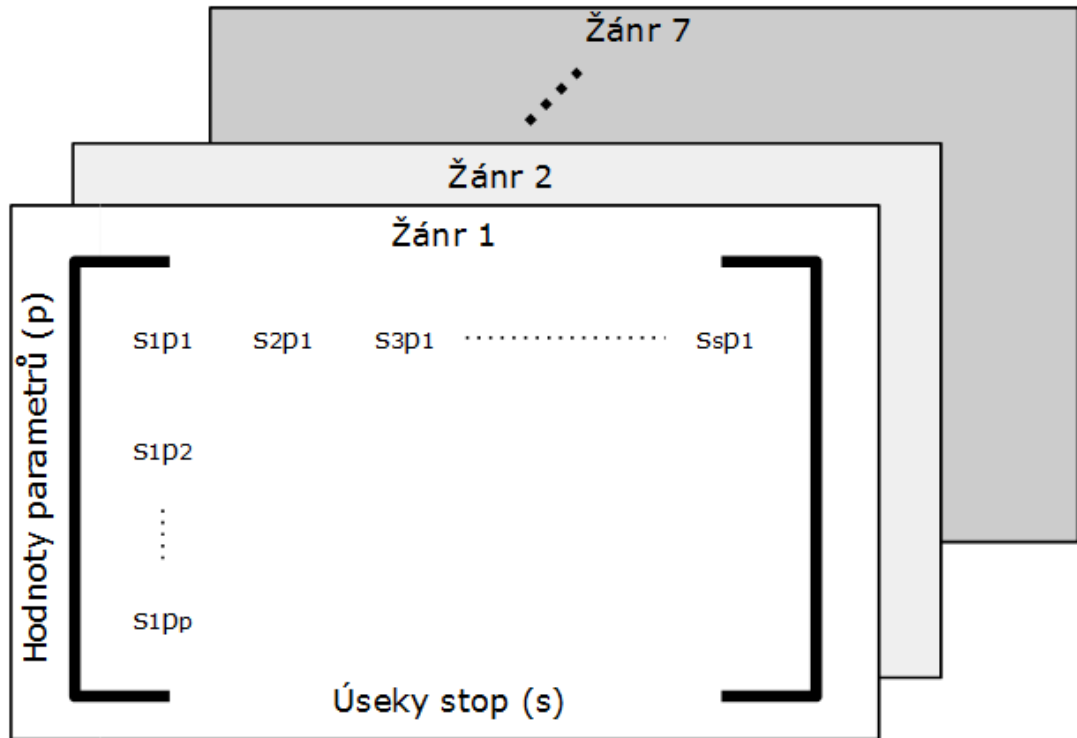
Obrázek 10 Pravděpodobnost výskytů sledovaných žánrů v nahrávce

Na obr. 10 má největší zastoupení v nahrávce žánr Rock, proto bude nahrávka zařazena do tohoto žánru.

### 3.2.3 Algoritmus programu

Nadefinuje se, jaké nahrávky budou použity pro TR1 a TR2. Tyto nahrávky jsou následně převedeny do mono, normalizovány tak, že jejich maximální hodnota amplitudy je prohlášena za 1 a minimální za -1. Následně proběhne převzorkování s vzorkovacím kmitočtem 44100 Hz. Postupně je potom z každé nahrávky odstraněn definovaný počet vteřin na začátku a na konci z důvodu eliminace *hluchých míst* a nahrávka je rozdělena do *fragmentů* o definované délce a z každého *fragmentu* je posléze vyňat *úsek*, z kterého

jsou vypočítány parametry A nebo B. Zbytek *fragmentu* se nevyužije. Délka *úseku* závisí na tom, zda se počítá A nebo B. Pro první část experimentu je totiž nutné, aby množiny obsahovaly stejný počet *úseků*. Výsledkem je soubor matic, kde v každé je uložena hodnota všech *úseků* všech parametrů pro daný žánr. Pro 7 žánrů tedy vznikne 7 matic.



Obrázek 11 Soubor matic vypočítaných parametrů pro úseky nahrávek.

Je třeba dosáhnout stejných počtů parametrů pro všechny žánry. To vyžaduje stejný počet *úseků* pro všechny nahrávky. Nejprve program pracoval tak, že se všechny nahrávky celé TR rozdělily do stejně dlouhých *fragmentů* a následně se vypočítaly parametry ze všech *úseků* všech nahrávek. Poté se zjistilo, který žánr obsahuje nejméně *úseků* a ostatní žánry byly tomuto počtu přizpůsobeny tak, že se vypočítal podíl počtu *úseků* většího žánru ku nejmenšímu  $d$  a následně se z většího žánru vzal každý  $d$ -tý *úsek*. Zbylé *úseky* se nepoužily.

To bylo ale velice pomalé a neefektivní. Např. parametry A se počítaly z TR1 přes 8 hodin. Proto byl algoritmus upraven tak, že se nejdříve zjistí délky jednotlivých nahrávek. Proveďte se prealokace [37], která ještě urychlí průběh programu. Poté se sečte délka nahrávek pro každý žánr a zjistí se, který žánr bude obsahovat nejméně *fragmentů*. To bude maximální počet *fragmentů* TR. Program poté vypočítá rovnoměrné rozdělení počtu *fragmentů*, které se vezmou z nahrávek ostatních žánrů tak, aby se z každého žánru vzal stejný počet *fragmentů*, jako je v nejmenším žánru. Rozdělení *fragmentů* každé skladby je provedeno pomocí vztahu (7).

$$N_f = \frac{n_n}{n_z} \cdot n_{max} \quad (7)$$

kde  $N_f$  je počet nově získaných *fragmentů*,  $n_n$  je celkový počet *fragmentů*, které lze z dané nahrávky získat,  $n_z$  je maximální počet *fragmentů*, které lze získat z daného žánru a  $n_{max}$  je maximální počet *fragmentů*, které lze získat z nejmenšího žánru. *Fragmenty* jsou tak rovnoměrně rozděleny mezi celý žánr. Ze vztahu (7) vyplývá, že nejvíce *fragmentů* je vyňato z nejdelších nahrávek a nejméně z nejkratších.

Toto je velká výhoda pokud TR obsahuje délkou nevyvážené žánry, protože větší žánr obsahuje větší pestrost parametrů a tedy bude schopen zařazovat úspěšněji. Toho lze využít, pokud je třeba systém využíván např. k odlišení jednoho žánru od ostatních. Další výhodou je výrazné urychlení chodu programu pro TR s rozdílně dlouhými žánry (např. v případě TR1 o desítky procent). Nevýhodou je možná chyba, pokud je poměr největšího a nejmenšího žánru příliš velký, zvyšuje se riziko, že *úseky z fragmentů* většího žánru budou obsahovat *hluchá místa*, protože se z nahrávky z objemnějších žánrů rozdělí na méně *fragmentů*. Proto byl do algoritmu zařazen překryv neboli tzv. „Hop faktor“, pomocí kterého se může na počátku stanovit, v kolika procentech předchozího *fragmentu* bude začínat další *fragment*. Tím se zvýší počet *fragmentů* na úkor překryvu jednotlivých *úseků*. Musí se proto volit obezřetně v případě, že délka *úseku* je totožná s délkou *fragmentu*.

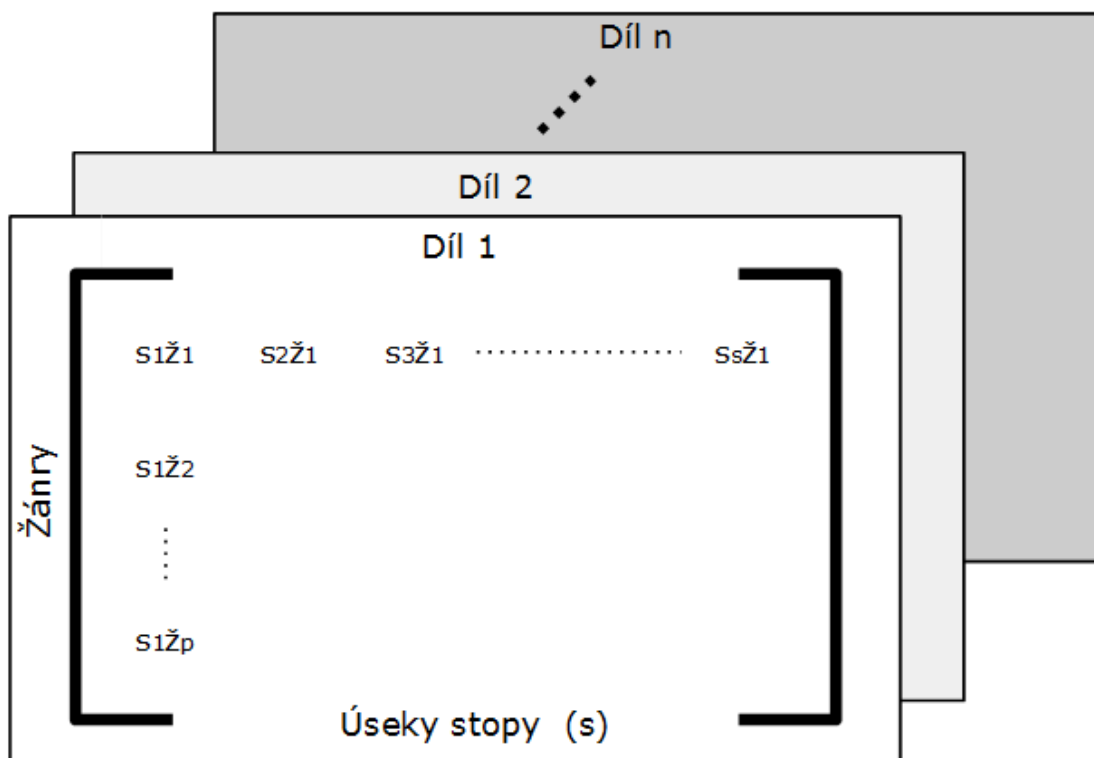
Po výpočtu TR se načte testovací množina TS. Každá nahrávka je upravena stejným způsobem jako nahrávky v TR, tedy je převedena do mono, normalizována a převzorkována. Poté se odstraní počáteční a koncová část skladby nahrávky kvůli eliminaci hluchých míst. Následně se provede opět rozdělení do *fragmentů*. Z každého *fragmentu* se opět vyjme úsek – v tomto případě bude pro rozlišení od *úseku* TR nazván *díl*. Počet *dílů* bude ovlivňovat výslednou efektivitu detekce. Předpokládá se, že čím více *dílů* ze skladby se vezme, tím úspěšnější detekce bude. Z každého *úseku* se vypočítají parametry A nebo B.

Po výpočtu parametrů u TR i TS se přejde k určení vzdáleností *úseků* od *dílů* pomocí vztahu (1) Vznikne nový soubor matic, kde v každé matici je uložena vzdálenost *úseků* všech žánrů od  $n$ -tého *dílu*. Ilustrativní znázornění je na obr. 12.

Např. v první matici je  $s_{1z_1}$  vzdálenost prvního *úseku* prvního žánru od prvního *dílu*.

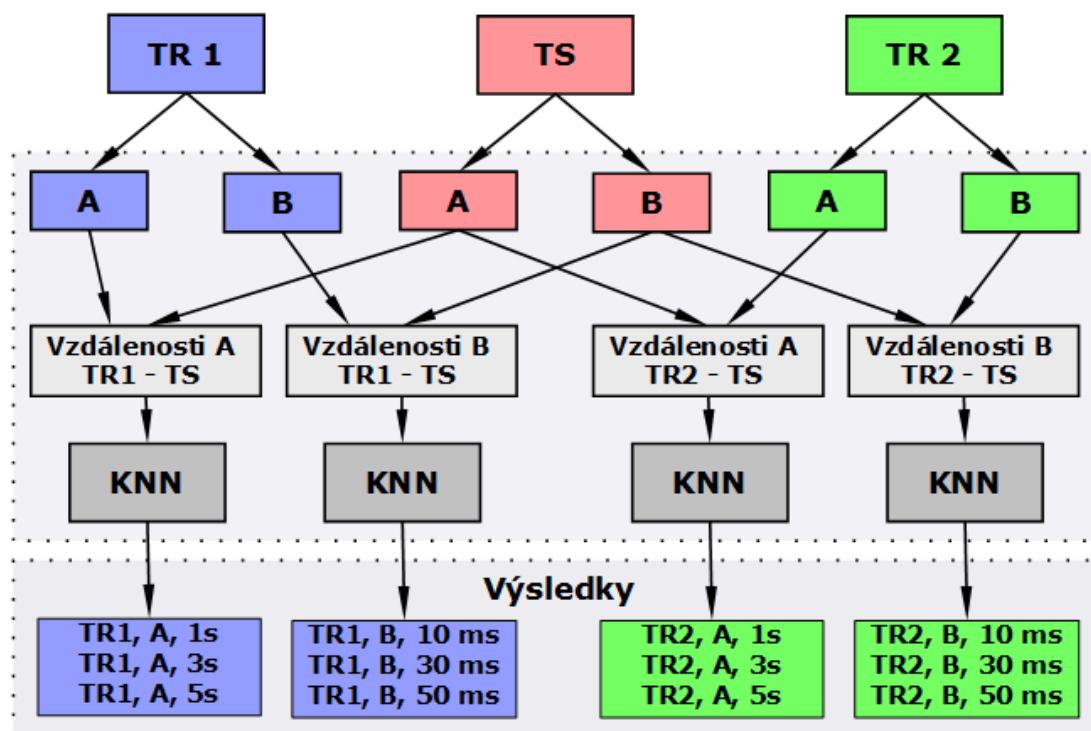
Poté se z každého *dílu* vyhledá  $k$  nejmenších vzdáleností od TR.  $K$  je parametr  $KNN$ . Zjistí se zastoupení žánrů v jednotlivých *dílech* a určí se nejvíce zastoupený žánr. viz. „Výstup programu“.

Pozn.: Při klasifikaci je vždy délka právě klasifikovaných *dílů* TS stejná jako délka *úseků* TR.



Obrázek 12 Vzdálenosti úseků TR od dílů TS

### 3.2.4 Výběr skupiny parametrů a TR



Obrázek 13 Ilustrační schéma hledání optimálního řešení systému

Na obr. 13 jsou na začátku 2 trénovací množiny (TR1 a TR2). Mezi nimi je testovací množina (TS) V každé z těchto tří množin je proveden výpočet parametrů A a B a následně jsou vždy vypočítány vzdálenosti dle vztahu (1) mezi příslušnými množinami parametrů. Poté vždy následuje blok KNN, kde se vypočítá *k- nearest neighbours*. U jednotlivých bloků se budou měnit různá nastavení a parametry a bude se zkoušet, které nastavení je nejefektivnější.

Dle [4] by délky úseků parametrů A měly být 3 vteřiny. To bylo určeno na základě experimentu se studenty, kteří detekovali žánry nahrávek dle úseků délky začínající na 250 ms. Úspěšnost klasifikace rostla do 3 vteřin a poté už se nezvyšovala. Délka *úseků* MFCC by měla být 25 ms.

Provedou se vždy tři výpočty pro parametry A i B u obou TR. Dále se provedou tři výpočty pro parametry A a B u TS. Celkem se počítá 12 variant. Pro korektnost porovnání byly nahrávky rozděleny nejprve do *fragmentů*, aby byl dosažen stejný počet *úseků* (*dílů*) pro všechny varianty, viz. tabulka 2 a tabulka 3.

Tabulka 2 Nastavení TR1 a TR2

Trénovací množina	TR1		TR2	
Skupina parametrů	A	B	A	B
Délka fragmentu [s]	5	5	5	5
Délka úseku	1 s; 3 s; 5 s	10 ms; 30 ms; 50 ms	1 s; 3 s; 5 s	10 ms; 30 ms; 50 ms
Počet úseků	2500	2500	2500	2500

Tabulka 3 Nastavení TS

Testovací skupina	TS	
Skupina parametrů	A	B
Délka fragmentu [s]	5	5
Délka dílu	1 s; 3 s; 5 s	10 ms; 30 ms; 50 ms
Počet dílů	50	50

Poté se pro každou variantu provede detekce žánrů pomocí KNN, kdy za *k* se budou dosazovat hodnoty  $k = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50\}$ . Celkem vznikne 12 variant úspěšností (viz. obr. 14).

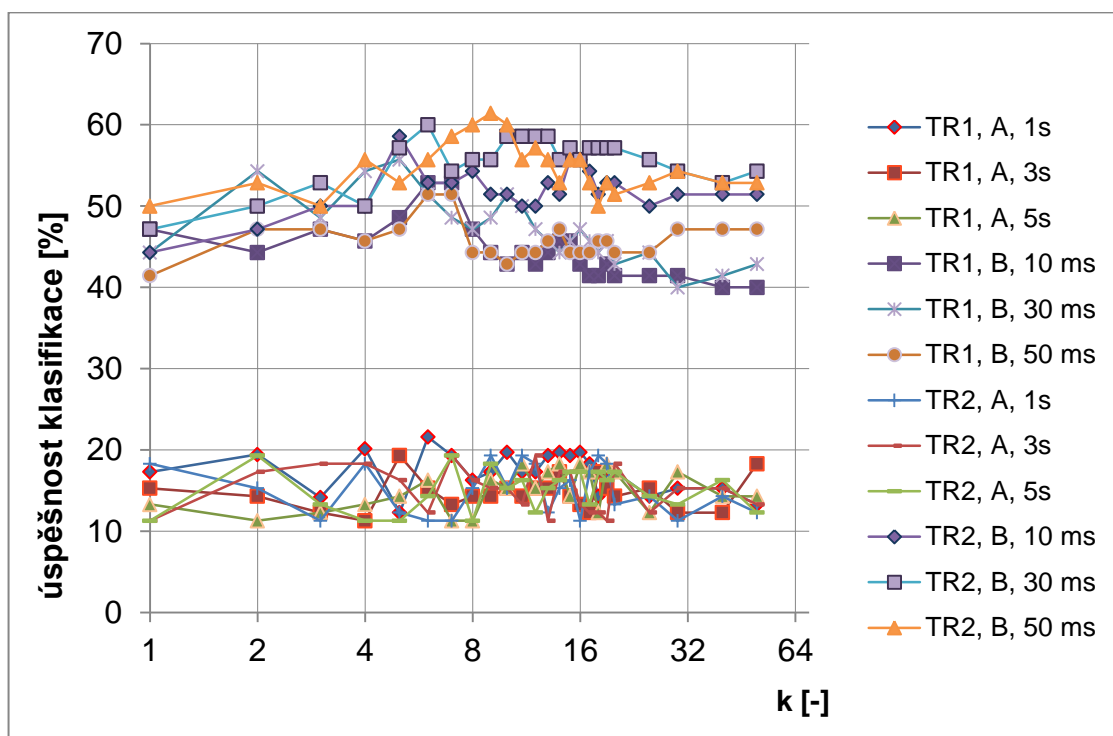
Žánry skladeb TS jsou předem známé. Klasifikátor vypočítá pravděpodobnost zastoupení žánrů v dané nahrávce. Pokud bude žánr s největší pravděpodobností stejný jako skutečný žánr nahrávky, bude klasifikace prohlášena za úspěšnou.

## Výsledky

Výsledky jsou znázorněny v grafu na obr. 14. Úspěšnost klasifikace za použití parametrů A byla pro obě trénovací množiny podobná. Pohybovala v rozmezí 10 – 20 %. Průměrně 15 %. Použití množiny parametrů A tedy pro detekci není vhodné.

Úspěšnosti klasifikace za použití B parametrů byla v případě obou trénovacích množin také podobná. Průměrná hodnota úspěšnosti je v tomto případě 50 %. O trochu lépe si vedla TR2, kde při délce úseku 30 ms byla průměrná úspěšnost 55,5 %. Nejvyšší úspěšnosti klasifikace 61,4 % bylo dosaženo pro TR2 při délce úseku 50 ms, a  $k = 9$  pro B parametry. Jedná se o variantu 12.

Detailní informace o variantě 12 jsou zobrazeny v tabulce Tabulka 4.



Obrázek 14 Dvanáct variant úspěšností pro různá k

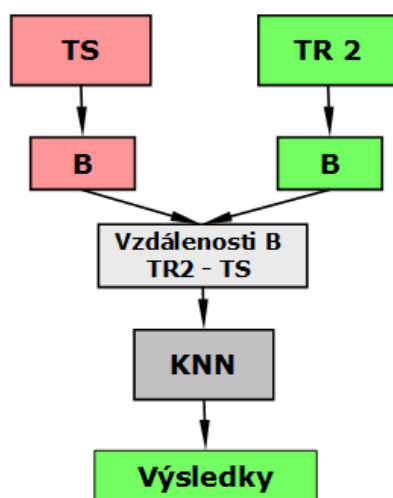
Tabulka 4 Výsledky varianty 12

		Zařazeno do žánru							Úspěšnost [%]
		ROCK	POP	JAZZ	FOLK	ELEKTR.	VÁŽNÁ H.	HIP HOP	
Žánr testované skladby	ROCK	10	0	0	0	0	0	0	100
	POP	1	4	1	0	4	0	0	40
	JAZZ	2	0	6	1	1	0	0	60
	FOLK	0	1	4	4	0	1	0	40
	ELEKTR.	2	0	1	0	6	1	0	60
	VÁŽNÁ H.	0	0	0	0	0	10	0	100
	HIP HOP	2	0	0	1	4	0	3	30

Při tomto nastavení byly rozpoznány všechny nahrávky z žánru Rock a Vážná hudba. Čtyři popové nahrávky klasifikátor zařadil do Elektroniky. Jak ale bylo zmíněno v kap. 2.1, tato chyba byla vzhledem k podobnosti žánrů očekávatelná. Překvapující není ani zařazení 4 folkových skladeb do Jazzu. Tyto žánry jsou si v mnoha ohledech podobné. Zařazení Elektroniky do Vážné hudby a Jazzu a zařazení Hip hopu do Folku jsou největší chyby. Nejmenší úspěšnost klasifikace má žánr Hip hop. Rozpoznány byly pouze 3 nahrávky. Čtyři nahrávky z tohoto žánru byly klasifikovány jako Elektronika. Stejně jako u žánrů Hip hop a Elektronika i u tohoto žánru jsou výsledky předpokládáné. Žádný jiný žánr ovšem jako Hip hop zařazen nebyl. Pokud by se tedy žánr Hip hop zcela vyřadil, vzrostla by úspěšnost až na 67 %. Klasifikátor lze prohlásit jako velice úspěšný. Dále se bude pracovat s variantou 12.

### 3.2.5 Další upřesnění detekce

Následně se zkoumalo, jak počet *dílů*, které se extrahují z neznámé nahrávky, ovlivní klasifikaci. Současně byla měněna délka jednotlivých *úseků* (*dílů*) a *k*. Pro zjednodušení zápisu se pro vyjádření sledovaných hodnot bude využívat množina  $\{dily[-], úseky [ms], k [-]\}$ . Postupovalo se dle schématu na obr.15.



Obrázek 15 Ilustrační schéma zvýšení citlivosti detekce

Koeficienty *k*, počty *dílů* a délky *úseků* jsou uvedeny v tabulce 5.

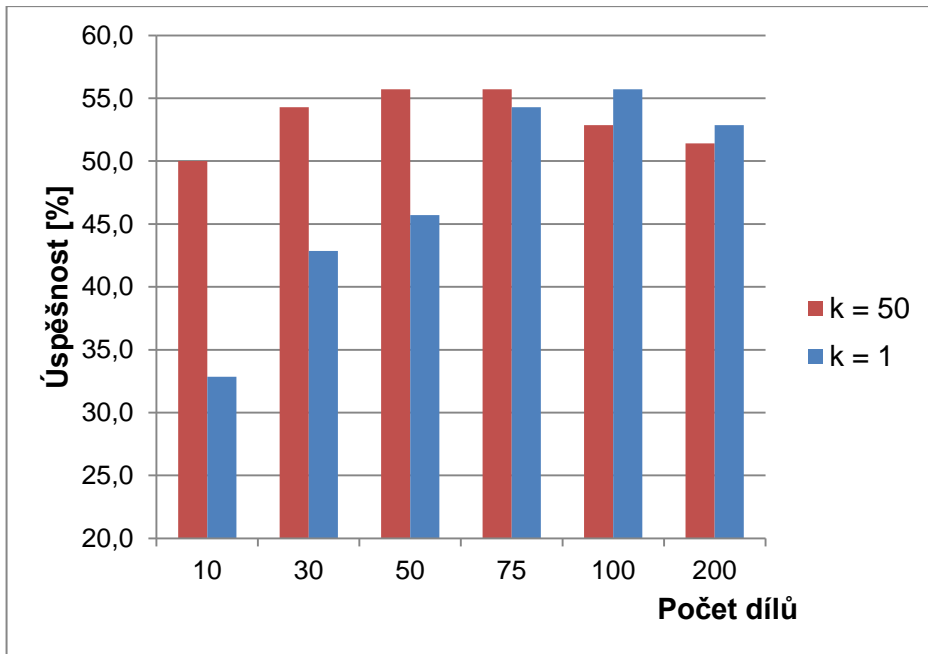
Tabulka 5 Hledání optimálního nastavení systému

<b>k</b>	1; 5; 7; 9; 11; 13; 15; 50
<b>počet dílů (úseků)</b>	10; 30; 50; 75; 100; 200
<b>délka dílu (úseku [ms])</b>	20; 30; 40; 50; 60; 100

Úspěšnost vyšla až na výjimky ve všech případech podobně. Průměrně 54,7 %. Nejmenší úspěšnost 32,9 % byla při nastavení {20, 10, 1}. Nikdy nebyla přesažena úspěšnost



64,3 %. Nejvyšší úspěšnosti bylo dosahováno při kombinaci nízké hodnoty  $k$  a velkého počtu *dílů*. Čím menší hodnota  $k$ , tím méně se bude hledat vzdáleností pro daný díl. Zvyšuje se šance, že nejmenší vzdálenosti každého *dílu* od *úseku* (klasifikace bude úspěšná) budou právě pro správný žánr. Při velmi nízkých hodnotách  $k$  ovšem hrozí, že většina *dílů* může být zařazena jako jiný žánr. Na obr. 16 jsou výsledky pro různý počet *dílů* při nastavení  $k = 1$  a  $k = 50$  při délce *úseku* 20 ms.



Obrázek 16 Úspěšnost klasifikace pro různá  $k$

Pro  $k = 50$  nejprve úspěšnost roste, ale při extrakci 50 *dílů* se začne snižovat, protože se vyhodnocují i vzdálenosti, které již mohou příslušet jiným žánrům. Pro  $k = 1$  úspěšnost roste až do extrakce 100 *dílů*, poté ale začíná také klesat, protože se zvyšuje počet *dílů*, jejichž nejkratší vzdálenost byla s nestejným žánrem. Více *dílů* navíc znamená vyšší výpočetní náročnost, protože se musí hledat více vzdáleností. Počet  $k$  by tedy neměl být příliš velký a počet *dílů* zase moc malý při zachování velké úspěšnosti. Toto splňuje nastavení {100, 60, 7}. Jeho úspěšnost je 64,3 %, tedy největší dosažená úspěšnost.

Pozn.: V některých případech (např. pro  $k = 13$ ) vycházela vysoká úspěšnost detekce pro délku *úseku* 100 ms. Podle [31] [32] by měla být optimální hodnota *úseku* pro výpočet MFCC koeficientů 15 – 50 ms. Byl tedy proveden výpočet pro {100, 200, 13} a {100, 200, 7}. V obou případech vyšla úspěšnost pod 53%. Úspěšnost pro větší délku *úseku* tedy klesla, což je v souladu s literaturou [31] [32].

Všechny výsledky jsou zobrazeny v příloze B v tabulce 9.

### 3.2.6 Shrnutí

Bylo zjištěno, že pro detekci hudebních žánrů jsou v navrženém experimentu vhodnější parametry B. *Úsek* byl nejefektivnější při délce 60 ms. Bylo rozhodnuto, že pro nej-

lepší detekci nesmí být počet  $k$  příliš velký, ale zároveň ani malý. Z experimentu byla určena hodnota  $k = 7$ . Počet 100 *dílů* byl zvolen vzhledem k výsledkům a výpočetní náročnosti. Více *dílů* by zbytečně prodlužovalo výpočet a úspěšnost by se neměnila. Byla zvolena trénovací množina TR2. Na základě těchto výsledků vznikl systém pro automatickou detekci hudebních žánrů. Systém detekce tedy bude pracovat s nastavením uvedeným v tabulce 6

Tabulka 6 Nastavení systému dle výsledků

Trénovací množina	TR2
Skupina parametrů	B
Délka úseku/dílu	60 ms
Počet <i>dílů</i> :	100
Počet <i>úseků</i> :	2546

### 3.3 Vylepšení systému

Účelem druhé části je pokusit se navrhnout vylepšení pro navržený systém. Vzhledem k rozsahu práce jsou jednotlivé složky této části navrženy za zjednodušených podmínek. Nebylo provedeno dostatečné ověření, proto tyto funkce nemusí pracovat přesně.

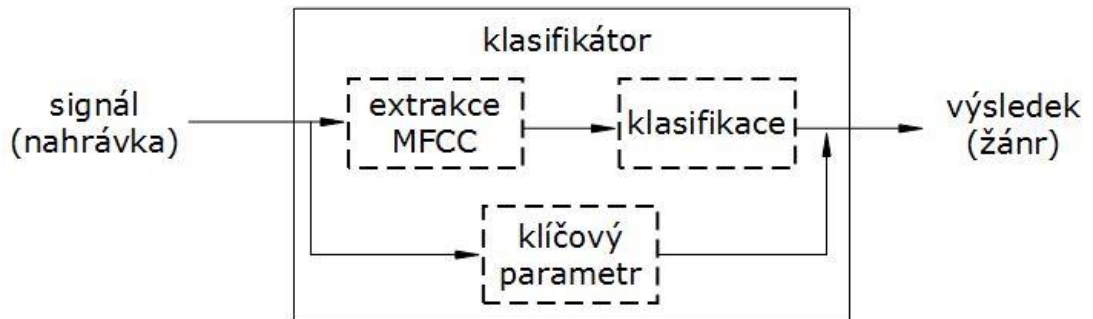
#### 3.3.1 Klíčový parametr

**Myšlenka:** Pokud by se hodnoty nějakého parametru určitého žánru zásadně lišily od ostatních žánrů, mohla by být zlepšena úspěšnost výsledné klasifikace.

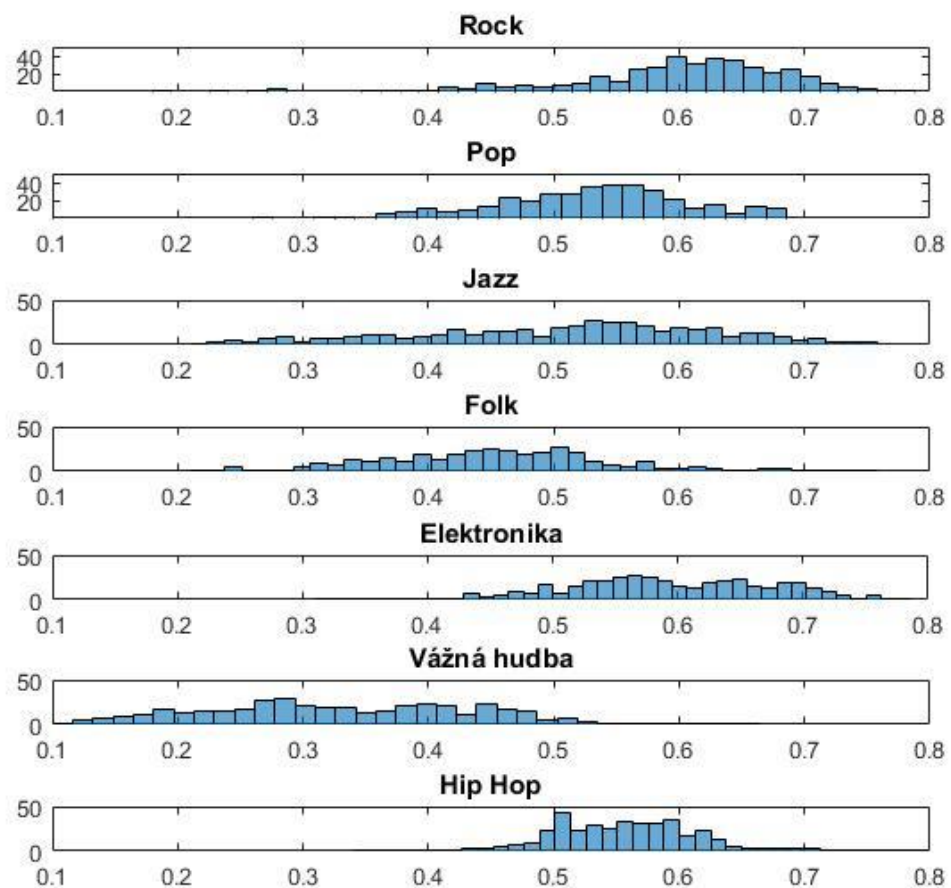
**Realizace:** V procesu trénování se určí klíčový parametr. Je třeba volit takový parametr, jehož výpočetní náročnost není moc vysoká, aby se výrazně nezpomalovala rychlost detekce. Vhodné mohou být například parametry ze skupiny A. Určí se ojedinělý žánr, kde se parametr nejvíce liší od ostatních. Poté se na základě výsledků vytvoří váhovací vektor, pomocí kterého se bude přepočítávat výsledná pravděpodobnost výskytu žánrů. Při detekci se poté souběžně s MFCC koeficienty vypočítá i klíčový parametr neznámé nahrávky. Pokud by jeho hodnota odpovídala právě ojedinělému žánru, zvýšila by se pravděpodobnost klasifikace dané nahrávky do ojedinělého žánru na základě váhovacího vektoru. Znázornění za použití upraveného základního schématu detekce je na obr. 17.

**Postup hledání parametru:** Pro analýzu byla zvolena trénovací množina TR2. Z každé nahrávky byly extrahovány *úseky*. Nejprve byla nastavena délka *úseku* 60s. Při této hodnotě se nejvíce lišil pro všechny parametry žánr Vážná hudba. Tento žánr ale už úspěšně systém detekuje, proto pro 60 s nebyla data shledána jako užitečná. Následně byla délka *úseku* změněna na 30 s, 15s, 1s a nakonec 60 ms. Ani při jednom z nastavení nebyl žádný parametr pro určitý žánr vyjma Vážné hudby dominantní. Bylo potřeba vyhledat parametr, který by odlišoval od ostatních např. Pop či Elektroniku. Toho nebylo dosaženo.

Možným důvodem neúspěšnosti je málo objemná trénovací množina či špatný výběr parametrů. Příklad výsledků pro parametr Brightness s délkou úseku 15 je na obr. 18.



Obrázek 17 Schéma detekce s klíčovým parametrem



Obrázek 18 Hodnoty parametru Brightness pro jednotlivé žánry

### 3.3.2 Koeficient změny

Díky Loudness war se předpokládá, že hodnota RMS bude vyšší u novějších nahrávek. Dynamická komprese také snižuje rozdíly mezi tichými a hlasitými pasážemi nahrávky. Pokud by se zjistilo, kolik alb reprezentujících jednotlivé žánry se vydalo v posledních letech, lze na základě těchto informací upřesnit detekci.

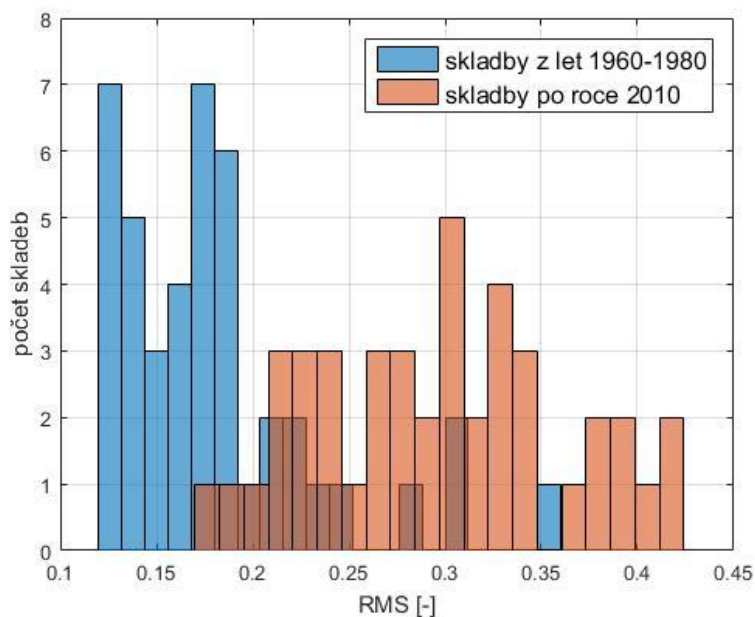
Internetový portál Wikipedia obsahuje záznamy z let 2010 – 2015 (kromě roku 2011) o vydaných albech v jednotlivých letech. U každého alba je též informace o žánru. Nejsou zde samozřejmě všechna vydaná alba daného roku. Bude ale předpokládáno, že na základě těchto dat lze charakterizovat celkový roční počet vydaných alb pro jednotlivé žánry. Data byla zpracována. Přehled je uveden v tabulce 7.

Tabulka 7 Orientační přehled vydaných alb dle žánrů v období 2010 - 2015

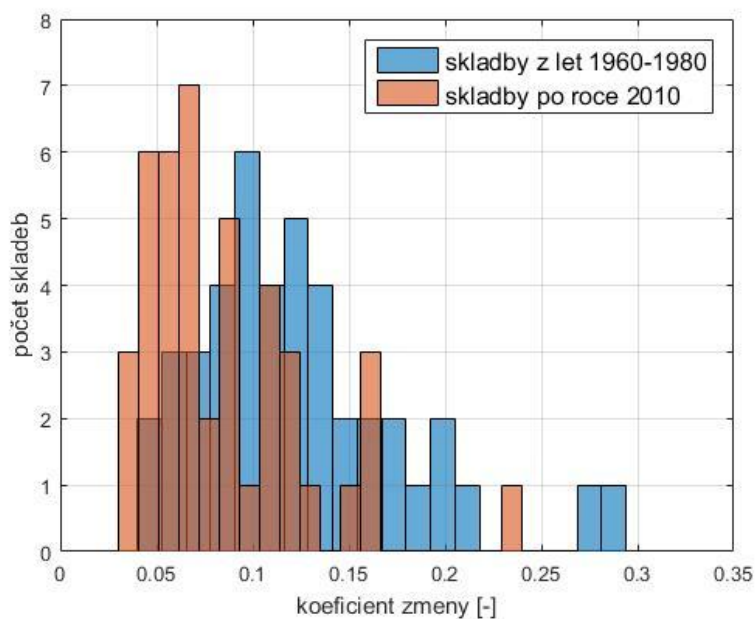
skladeb v žánru za rok	2015	2014	2013	2012	2010	celkem	podíl [%]
<b>ROCK</b>	332	333	337	247	208	<b>1457</b>	51,2
<b>POP</b>	174	194	151	90	94	<b>703</b>	24,7
<b>JAZZ</b>	15	13	7	6	1	<b>42</b>	1,5
<b>FOLK</b>	32	34	30	21	9	<b>126</b>	4,4
<b>ELEKTRO</b>	45	40	48	31	19	<b>183</b>	6,4
<b>KLASIKA</b>	4	5	6	2	4	<b>21</b>	0,7
<b>HIPHOP</b>	73	76	74	51	42	<b>316</b>	11,1
<b>skladeb za rok</b>	<b>675</b>	<b>695</b>	<b>653</b>	<b>448</b>	<b>377</b>	<b>2848</b>	

Následně byla vypočítána hodnota RMS pro 43 nahrávek pořízených po roce 2010 a 43 nahrávek pořízených v letech 1960-1980. Graf na obr. 19 potvrzuje domněnku, že RMS bude u novějších skladeb vyšší.

Předpokládá se, že RMS se v průběhu nahrávky bude více měnit u nahrávek vydaných v letech 1960 – 1980. Nahrávky byly rozděleny na 5 vteřinové úseky (s délkou úseku bylo experimentováno a tato přinesla nejlepší výsledky) a z každého úseku bylo vypočítáno RMS. Poté byla pro každou skladbu vypočítána procentuální změna RMS mezi sousedními úseky a tento údaj byl pro každou skladbu sečten a vydělen počtem úseků. Výsledkem tak je 43 koeficientů pro obě skupiny, které popisují, jak se v průběhu času mění RMS nahrávky. Jinými slovy bylo vypočítáno, jak intenzivně se mění obálka signálu v průběhu skladby. Tyto koeficienty jsou nazvány *koeficienty změny*. Koeficienty pro jednotlivé nahrávky jsou znázorněny na obr. 20. Koeficient změny popisuje, jak se dynamicky skladba mění ve větších časových úsecích. U dynamicky kompresovaných nahrávek by tak tento koeficient měl být menší, což bylo potvrzeno.



Obrázek 19 RMS skladeb v různém období



Obrázek 20 Koeficienty změny pro vybrané nahrávky

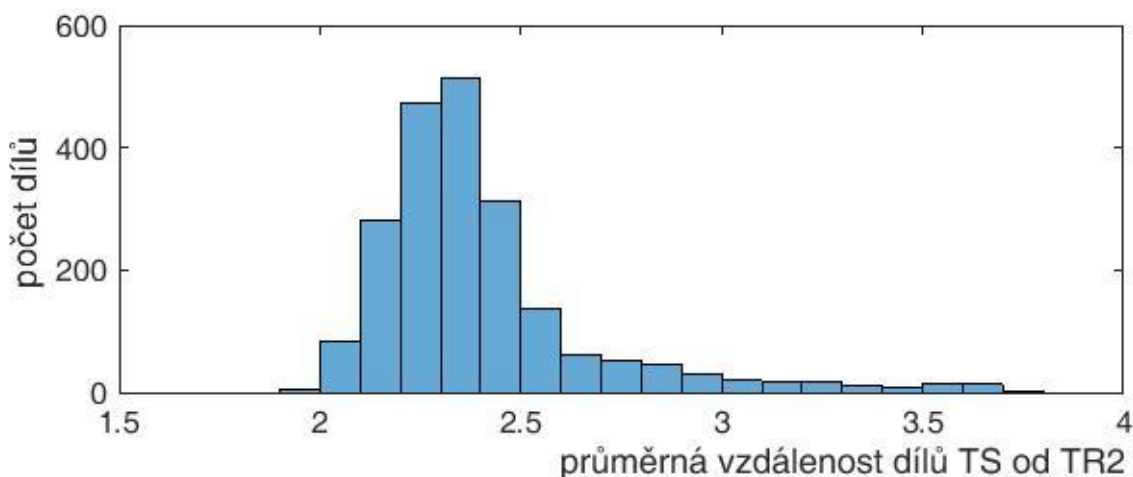
Z podílů vydání alb v jednotlivých letech v tabulce 7. byl vytvořen váhovací vektor. Z výsledků bylo stanoveno, že pokud RMS neznámé skladby přesáhne hodnotu 0,25 (hodnota určena odhadem z grafu na obr. 19) a koeficient změny bude menší než 0,1 (hodnota určena odhadem z grafu na obr. 20), poté se k výsledné pravděpodobnosti přičtou hodnoty vektoru. Tím se změní výsledek klasifikace. Výsledek je poté upraven, aby součet jednotlivých pravděpodobností nebyl větší než 1. Jelikož je tato metoda založená na odhadu a malému vzorku dat, byla vzata pouze poloviční hodnota vektoru.

Toto je pouze experimentální obohacení detekce, které bere v úvahu fenomén Loudness war. Jedná se o návrh. Pro přesnější využití této problematiky by bylo zapotřebí přesnějších výpočtů a více dat.

### 3.3.3 Klasifikace nehuděbních nahrávek

Komplexnost detekce vyžaduje rozlišení hudební produkce od ostatních druhů signálů, proto byl navržen experimentální systém, který na základě vzdáleností TR2 a TS rozpozná, zda signál na vstupu je hudebního charakteru.

Pro výpočet byl použit systém vzniklý v této práci. Pouze počet extrahovaných *dílů* z každé z nahrávek TS byl kvůli vysoké výpočetní náročnosti snížen na 30. Vypočítal se průměr vzdálenosti každého *dílu* skladeb z TS od TR2. Celkem vzniklo 2100 průměrů vzdáleností. Výsledek je na obr. 21



Obrázek 21 Průměrná vzdálenost dílů TS od TR2

Průměrná vzdálenost 2,502 bylo zvoleno jako referenční hodnota. 80 % *dílů* má menší průměrnou vzdálenost od TR2. Pokud u poloviny a více *dílů* detekovaného signálu bude průměrná vzdálenost větší než 2,502, systém upozorní, že se pravděpodobně nejedná o klasifikovatelnou nahrávku. Klasifikace se přesto provede. S těmito hodnotami bylo experimentováno na 4 nahrávkách obsahujících mluvené slovo. U všech bylo určeno, že se nejedná o hudbu.

Toto je pouze experimentální metoda. Pro její upřesnění je třeba podrobit jí důkladnější analýze. Není součástí zadání BP. Jedná se pouze o návrh řešení. Cílem této práce není rozlišit hudební nahrávky od ostatních signálů. Při užití systému detekce se nepředpokládá, že budou detekovány nehuděbní signály.

## 4 Závěr

Tématem této práce je vytvoření systému pro detekci hudebních žánrů pro účely masteringu gramofonových desek. Tento úkol byl splněn. Systém lze navíc použít i v dalších oblastech.

Nejprve byl definován hudební žánr, poté popsána aktuální situace automatické detekce hudebních žánrů, a MIR. Jsou zmíněny problémy při detekci hudby, je rozebrán též fenomén Loudness war. Bylo popsáno obecné vytvoření systému automatické detekce hudebních žánrů a byly popsány používané metody, modely a parametry. Následně byl dle kapitoly 2 kompletně navržen a implementován systém pro automatickou detekci žánrů, za použití modelu *k-nearest neighbours*. Nejprve bylo porovnáváno, která ze dvou skupin parametrů poskytne lepší výsledky detekce a zda záleží na volbě trénovací množiny. Lepší výsledky poskytovaly MFCC. U trénovacích množin nebyl na výsledek větší vliv. Dále bylo detailněji prozkoumáno použití koeficientů MFCC a také objem dat, která budou extrahována z neznámé nahrávky. Byl hledán klíčový parametr ze skupiny parametrů A, který by zlepšil výslednou detekci. Tento parametr nebyl nalezen. Pravděpodobně z důvodu chybně volených parametrů či nedostatečně obsáhlé trénovací množiny. Dále bylo zkoumáno, jak by fenomén Loudness war mohl ovlivnit výslednou detekci. Výsledkem je nalezení experimentálního koeficientu, který kombinuje parametr RMS a množství vydaných alb v letech 2010 – 2015. Následně bylo experimentálně zajištěno, aby systém dokázal rozlišit hudební nahrávku od nehudební. Vytvořený systém byl ověřen pomocí 70 nahrávek. Systém dosáhl úspěšnosti 64,3 %. To je možné považovat za úspěšný výsledek vzhledem k tomu, že ani člověk někdy nedokáže jednoznačně hudební žánr identifikovat.

Systém od sebe dokázal bezpečně rozlišit nejrozdílnější žánry, což je velice důležité. Výhoda systému spočívá v tom, že neinformuje pouze o zařazení do jednoho žánru, ale udává pravděpodobnost, s jakou nahrávka do jednotlivých žánrů patří. Vyvinutý systém byl nazván GENDET a bylo k němu vytvořeno jednoduché uživatelské rozhraní. Systém je popsán v příloze D.

Na tuto práci bude navázáno nejprve tím, že bude vyzkoušeno, jaký vliv má jiný počet MFCC na detekci. Dále bude postupováno zaměřením opět na množinu parametrů A, bude hledáno lepší řešení a změny parametrů. Zajímavé výsledky by mohlo přinést studium harmonických parametrů. Bude kombinována skupina parametrů A a B. Bude zvětšen objem trénovací množiny a lépe definován koeficient změny. Cílem bude také větší přesnost rozlišení hudební nahrávky od nehudební.

V oboru zpracování signálů je zpracování hudby stále aktuálnější, byl nalezen velký počet článků, referujících o automatické detekci. Tato práce se volně inspiruje postupy provedenými v některých z nich a snaží se hledat nové postupy. Práce se od ostatních liší výstupem. Kromě zařazení nabídne i pravděpodobnost, s jakou skladba patří do ostatních žánrů.

# Citovaná literatura

- [1] GJERDINGEN, Robert O.; PERROTT, David. Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, 2008, 37.2: 93-100.
- [2] Music Genre List. *Different types of music genres*.
- [3] Genres. *All music*. [Online] [Citace: 18. 4 2016.] <http://www.allmusic.com/genres>.
- [4] TZANETAKIS, George, and Perry COOK. "Musical genre classification of audio signals." *Speech and Audio Processing, IEEE transactions on* 10.5 (2002): 293-302..
- [5] NAM, Unjung; SMITH III, Julius O.; BERGER, Jonathan. Automatic Music Style Classification: towards the detection of perceptually similar music. *Proceedings of Japanese Society of Music Cognition and Perception, Fukuoka, Japan, 2001..*
- [6] ROCK MUSIC IS TWICE AS POPULAR AS POP IN AMERICA – BUT R&B RULES STREAMING. *Musicbusinessworldwide*. [Online] [Citace: 15. 5 2016.] <http://www.musicbusinessworldwide.com/rock-music-twice-popular-pop-america-rb-rules-streaming/>.
- [7] Wikipedia. List of music styles. *Wikipedia*. [Online] [Citace: 18. 4 2016.] [https://en.wikipedia.org/wiki/List\\_of\\_music\\_styles](https://en.wikipedia.org/wiki/List_of_music_styles).
- [8] DOWNIE, J. Stephen. 2003. Music information retrieval. *Annual Review of Information Science and*.
- [9] FOOTE, Jonathan T. Content-based retrieval of music and audio. In: *Voice, Video, and Data Communications. International Society for Optics and Photonics, 1997. p. 138-147..*
- [10] ZHANG, Tong, and CC Jay KUO. "Hierarchical classification of audio data for archiving and retrieving." *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on. Vol. 6. IEEE, 1999..*
- [11] FUJINAGA, Ichiro, and Karl MACMILLAN. "Realtime recognition of orchestral instruments." *Proceedings of the international computer music conference. Vol. 141. 2000..*
- [12] LI, Tao; OGIHARA, Mitsunori; LI, Qi. A comparative study on content-based music genre classification. In: *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, 2003. p. 282-289..*
- [13] NILSSON, Martin a SUNDSTROM, Johan. *ID3*. [Online] [Citace: 10. 5 2016.] <http://id3.org/>.
- [14] HASTIE, Trevor, et al. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 2005, 27.2: 83-85.
- [15] KANTOR, Jan. Učení bez učitele. *Diplomová práce*. [Online]. Brno : VUT, 2008. Dostupné online: <https://dspace.vutbr.cz/bitstream/handle/11012/8600/U%C4%8Den%C3%AD%20bez%20u%C4%8Ditele.pdf?sequence=1>.



- [16] PANIGRAHI Kiran Kumar. Artificial Intelligence - Neural Networks. *www.tutorialspoint.com*. [Online] 2016. [Citace: 20. 4 2016.] [http://www.tutorialspoint.com/artificial\\_intelligence/artificial\\_intelligence\\_neural\\_networks.htm](http://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm).
- [17] KOSHKINA, Ekaterina. "Identifikace obsahu archivních zvukových záznamů". *Bakalářská práce*. Praha : ČVUT v Praze, 2015. Dostupné online: [https://dspace.cvut.cz/bitstream/handle/10467/61652/F3-BP-2015-Koshkina-Ekaterina-identifikace\\_obsahu\\_archivnich\\_zvukovych\\_zaznamu.pdf?sequence=2&isAllowed=y](https://dspace.cvut.cz/bitstream/handle/10467/61652/F3-BP-2015-Koshkina-Ekaterina-identifikace_obsahu_archivnich_zvukovych_zaznamu.pdf?sequence=2&isAllowed=y).
- [18] Wikipedia. Expectation–maximization algorithm. *en.wikipedia.org*. [Online] 8. 4 2016. [Citace: 20. 4 2016.] [https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization\\_algorithm](https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm).
- [19] NickGillianWiki. GMM Classifier. *NickGillianWiki*. [Online] [Citace: 30. 4 2016.] <http://www.nickgillian.com/wiki/pmwiki.php/GRT/GMMClassifier>.
- [20] STERGIOPOULOS, Panagiotis S.; EFREMIDES, Odysseas B. Optimizing ANN's Architecture for Audio Music Genre Classification. In: *Proceedings of International Conference on Machine Learning and Computing (ICMLC 2009)*. 2009.
- [21] HLAVÁČ, Václav. Rozpoznávání s markovskými modely. <http://cmp.felk.cvut.cz/>. [Online] [Citace: 22. 4 2016.] <http://cmp.felk.cvut.cz/~hlavac/TeachPresCz/31Rozp/61MarkovianPR.pdf>.
- [22] RABINER, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE 77.2* (1989): 257-286.
- [23] HOUDEK, Michal, SVOBODA, Tomáš a PROCHÁZKA, Tomáš. Klasifikace podle nejbližších sousedů. <http://cmp.felk.cvut.cz/>. [Online] [Citace: 11. 4 2016.] [http://cmp.felk.cvut.cz/cmp/courses/recognition/zapis\\_prednasky/zapis\\_01/4/rpz4.pdf](http://cmp.felk.cvut.cz/cmp/courses/recognition/zapis_prednasky/zapis_01/4/rpz4.pdf).
- [24] BURRED, Juan José. An Objective Approach to Content-Based Audio Signal Classification. *Masterthesis. Technische Universitat Berlin, Berlin*, 2003.
- [25] Mindtools. Decision Trees. *mindtools.com*. [Online] [Citace: 24. 4 2016.] <https://www.mindtools.com/dectree.html>.
- [26] LAVNER, Yizhar; RUINSKIY, Dima. A decision-tree-based algorithm for speech/music classification and segmentation. *EURASIP Journal on Audio, Speech, and Music Processing*, 2009, 2009: 2.
- [27] LARTILLOT, Olivier. MIRtoolbox 1.6. *jyu.fi*. [Online] 30. 11 2014. [Citace: 25. 2 2016.] <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/MIRtoolbox1.6guide>.
- [28] DEKKER, Marcel a CHEN, C. H. *Signal Processing Handbook*. New York : 270 Madison Avenue, 1988. ISBN 0-8247-7956-8.

- [29] Wikipedia. Root mean square. *Wikipedia*. [Online] [Citace: 3. 4 2016.] [http://en.wikipedia.org/wiki/Root\\_mean\\_square](http://en.wikipedia.org/wiki/Root_mean_square).
- [30] MUDA, Lindasalwa; BEGAM, Mumtaj; ELAMVAZUTHI, I. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *arXiv preprint arXiv:1003.4083*, 2010.
- [31] AHREND, Peter. *Music genre classification systems*. Diss. Ph. D. dissertation, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2006.
- [32] HAGGBLADE, Michael; HONG, Yang; KAO, Kenny. Music genre classification. *Department of Computer Science, Stanford University*, 2011.
- [33] LYONS, James. Mel Frequency Cepstral Coefficient (MFCC) tutorial. <http://practicalcryptography.com/>. [Online] [Citace: 16. 4 2016.] <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>.
- [34] Dynamic Compression. *Bit Perfect Sound*. [Online] 2015. [Citace: 9. 5 2016.] <http://bitperfectsound.blogspot.cz/2015/06/dynamic-compression.html>.
- [35] DERUTY, Emmanuel. 'Dynamic Range' & The Loudness War. *Soundonsound*. [Online] 9 2011. [Citace: 11. 5 2016.] <https://www.soundonsound.com/sos/sep11/articles/loudness.htm>.
- [36] LARTILLOT, Olivier; TOIVIAINEN, Petri; EEROLA, Tuomas. A matlab toolbox for music information retrieval. In: *Data analysis, machine learning and applications*. Springer Berlin Heidelberg, 2008. p. 261-268.
- [37] Preallocation. *MathWorks*. [Online] [Citace: 12. 4 2016.] [http://www.mathworks.com/help/matlab/matlab\\_prog/preallocating-arrays.html](http://www.mathworks.com/help/matlab/matlab_prog/preallocating-arrays.html).
- [38] Wikipedia. Gaussova funkce. *Wikipedia*. [Online] [Citace: 19. 4 2016.] [https://cs.wikipedia.org/wiki/Gaussova\\_funkce](https://cs.wikipedia.org/wiki/Gaussova_funkce).
- [39] —. HiddenMarkovModel. *www.en.wikipedia.org*. [Online] [Citace: 24. 4 2016.] <https://commons.wikimedia.org/wiki/File:HiddenMarkovModel.svg>.
- [40] Bitperfectsound. [Online] [Citace: 9. 5 2016.] [http://bitperfectsound.blogspot.cz/2015\\_06\\_01\\_archive.html](http://bitperfectsound.blogspot.cz/2015_06_01_archive.html).

# Přílohy

Součástí práce jsou 4 přílohy. Příloha A (CD), Příloha B (dodatečné tabulky), Příloha C (přehled funkcí a skriptů vytvořených v Matlabu), Příloha D (popis GENDET)

## Příloha A

CD obsahuje systém GENDET, jeho popis a návod. Dále jsou zde veškeré vytvořené kódy a důležitá data z výpočtů ve složce MATLAB. Důkladnější popis je umístěn na CD v textovém souboru INFO.txt.

## Příloha B

Příloha B obsahuje některé tabulky vybraných výsledků z kapitoly 3. Dále je zde uveden seznam všech vytvořených kódů v Matlabu a popis nejdůležitějších z nich. Také je zde zobrazeno uživatelské rozhraní systému GENDET.

Tabulka 8 Úspěšnosti jednotlivých variant v 1. části

		k																				aritm.pr			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	40	50
úspěšnost [%]	TR1, A, 1s	17,3	19,4	14,1	20,1	12,3	21,6	19,3	16,3	17,3	19,7	17,3	19,3	19,7	19,3	19,7	18,3	14,3	16,3	17,3	14,3	15,3	15,3	13,3	17,3
	TR1, A, 3s	15,3	14,3	12,3	11,3	19,3	15,3	13,3	14,3	14,3	15,3	14,3	15,3	15,3	17,3	14,3	13,3	12,3	17,3	15,3	14,3	15,3	12,3	12,3	18,3
	TR1, A, 5s	13,3	11,3	12,3	13,3	14,3	16,3	11,3	11,3	16,3	15,3	18,3	15,3	17,3	18,3	14,3	18,3	17,3	12,3	18,3	17,3	12,3	17,3	14,3	14,3
	TR1, B, 10 ms	47,1	44,3	47,1	45,7	48,6	52,9	52,9	47,1	44,3	42,9	44,3	42,9	44,3	45,7	45,7	42,9	41,4	41,4	42,9	41,4	41,4	41,4	40,0	40,0
	TR1, B, 30 ms	44,3	54,3	48,6	54,3	55,7	51,4	48,6	47,1	48,6	51,4	50,0	47,1	45,7	44,3	45,7	47,1	45,7	44,3	45,7	42,9	44,3	40,0	41,4	42,9
	TR1, B, 50 ms	41,4	47,1	47,1	45,7	47,1	51,4	51,4	44,3	44,3	42,9	44,3	44,3	45,7	47,1	44,3	44,3	44,3	45,7	45,7	44,3	44,3	47,1	47,1	47,1
	TR2, A, 1s	18,3	15,3	11,3	18,3	12,3	11,3	11,3	15,3	19,3	15,3	19,3	18,3	12,3	15,3	16,3	11,3	17,3	19,3	18,3	13,3	14,3	11,3	14,3	12,3
	TR2, A, 3s	11,3	17,3	18,3	18,3	16,3	12,3	19,3	16,3	15,3	15,3	13,3	19,3	11,3	17,3	17,3	13,3	12,3	12,3	11,3	18,3	12,3	15,3	15,3	13,3
	TR2, A, 5s	11,3	19,3	13,3	11,3	11,3	14,3	19,3	11,3	18,3	15,3	16,3	12,3	15,3	16,3	17,3	17,3	13,3	17,3	16,3	17,3	14,3	13,3	16,3	12,3
	TR2, B, 10 ms	44,3	47,1	50,0	50,0	58,6	52,9	52,9	54,3	51,4	51,4	50,0	50,0	52,9	51,4	55,7	55,7	54,3	51,4	52,9	52,9	50,0	51,4	51,4	51,4
	TR2, B, 30 ms	47,1	50,0	52,9	50,0	57,1	60,0	54,3	55,7	55,7	58,6	58,6	58,6	58,6	55,7	57,1	55,7	57,1	57,1	57,1	57,1	57,1	55,7	54,3	52,9
	TR2, B, 50 ms	50,0	52,9	50,0	55,7	52,9	55,7	58,6	60,0	61,4	60,0	55,7	57,1	55,7	52,9	55,7	55,7	52,9	50,0	52,9	51,4	52,9	54,3	52,9	52,9

Tabulka 9 Výsledky úspěšnosti při hledání nejlepšího nastavení systé-

mu	počet dílů	10	30	50	75	100	200	průměrná úspěšnost [%]		počet dílů	10	30	50	75	100	200	průměrná úspěšnost [%]	
		úspěšnost [%]									úspěšnost [%]							
délka úseku [ms]	k = 1							47,4		k = 5							51,9	
	20	32,9	42,9	45,7	54,3	55,7	52,9				20	50,0	47,1	48,6	55,7	54,3		55,7
	30	34,3	41,4	48,6	58,6	52,9	57,1				30	45,7	60,0	55,7	61,4	57,1		57,1
	40	42,9	54,3	51,4	58,6	52,9	58,6				40	45,7	50,0	57,1	55,7	55,7		55,7
	50	37,1	45,7	52,9	51,4	58,6	64,3				50	45,7	52,9	52,9	60,0	61,4		64,3
	60	38,6	50,0	54,3	57,1	60,0	61,4				60	48,6	55,7	57,1	57,1	64,3		64,3
100	55,7	55,7	55,7	55,7	57,1	62,9	100	47,1	54,3	55,7	60,0	57,1	64,3					
	počet dílů	10	30	50	75	100	200	průměrná úspěšnost [%]		počet dílů	10	30	50	75	100	200	průměrná úspěšnost [%]	
délka úseku [ms]	k = 7							51,0		k = 9							51,7	
	20	48,6	44,3	51,4	51,4	55,7	54,3				20	50,0	47,1	54,3	52,9	51,4		54,3
	30	47,1	62,9	55,7	57,1	55,7	58,6				30	51,4	55,7	58,6	55,7	55,7		58,6
	40	44,3	51,4	57,1	58,6	57,1	54,3				40	47,1	54,3	55,7	57,1	57,1		58,6
	50	48,6	50,0	58,6	58,6	61,4	58,6				50	48,6	57,1	61,4	61,4	61,4		58,6
	60	45,7	60,0	62,9	58,6	64,3	62,9				60	47,1	60,0	58,6	60,0	61,4		60,0
100	50,0	58,6	54,3	61,4	58,6	60,0	100	50,0	61,4	54,3	61,4	60,0	61,4					
	počet dílů	10	30	50	75	100	200	průměrná úspěšnost [%]		počet dílů	10	30	50	75	100	200	průměrná úspěšnost [%]	
délka úseku [ms]	k = 11							52,4		k = 13							51,7	
	20	50,0	50,0	54,3	50,0	54,3	55,7				20	50,0	48,6	54,3	50,0	52,9		54,3
	30	48,6	54,3	60,0	55,7	55,7	58,6				30	51,4	55,7	60,0	55,7	55,7		57,1
	40	47,1	54,3	54,3	52,9	57,1	55,7				40	51,4	55,7	54,3	52,9	58,6		54,3
	50	50,0	55,7	61,4	60,0	58,6	57,1				50	50,0	55,7	58,6	60,0	57,1		54,3
	60	48,6	60,0	58,6	60,0	60,0	58,6				60	45,7	60,0	58,6	58,6	61,4		58,6
100	48,6	60,0	55,7	61,4	57,1	61,4	100	47,1	58,6	57,1	60,0	58,6	61,4					
	počet dílů	10	30	50	75	100	200	průměrná úspěšnost [%]		počet dílů	10	30	50	75	100	200	průměrná úspěšnost [%]	
délka úseku [ms]	k = 15							51,2		k = 50							51,7	
	20	51,4	45,7	54,3	50,0	52,9	52,9				20	48,6	45,7	57,1	52,9	54,3		51,4
	30	50,0	54,3	57,1	54,3	57,1	55,7				30	51,4	47,1	54,3	54,3	57,1		54,3
	40	50,0	52,9	54,3	52,9	55,7	54,3				40	51,4	48,6	55,7	52,9	54,3		54,3
	50	51,4	58,6	54,3	60,0	57,1	54,3				50	48,6	51,4	52,9	54,3	52,9		54,3
	60	48,6	57,1	60,0	57,1	58,6	57,1				60	50,0	54,3	55,7	55,7	52,9		54,3
100	48,6	57,1	55,7	60,0	58,6	60,0	100	45,7	54,3	54,3	57,1	54,3	55,7					

Tabulka 10 Nahrávky použité v TR2

Žánr	ROCK	POP	JAZZ	FOLK
Interpret - skladba	Deep Purple - Smoke on the water	Meghan Trainor - All about that bass	Ray Charles - Hit the Road Jack	Bob Dylan - The times They Are A changin
	Metallica - Master of puppets	Adelle - Hello	Frank Sinatra - I've Got You Under My Skin	Simon and Garfunkel - The sound of silence
	Iron maiden - Fear of the dark	Taylor Swift - Shake It Off	Nat King Cole - Unforgettable	Simon and Garfunkel - El Condor Pasa
	Iron maiden - Number of the beast	Madonna - Hung Up	Louis Prima - Just a Gigolo	Jaromir Nohavica - Tri cunici
	Metallica - St Anger	Britney Spears - Baby One more time	Nina Simone - My Baby Just Cares for Me	Johnny Cash - One
	Halloween - I want out	Ne-Yo - One in a million	Louis Armstrong, Ella Fitzgerald - Dream a Little Dream of Me	Neil Young - Heart Of Gold
	Slipknot - Killpop	Michael Jackson - The Don't care about as	Diana Krall - Cry Me A River	Little Big Town - Girl Crush
	Black Sabbath - Black Sabbath	Michael Jackson - Billie Jean	Marilyn Monroe - I Wanna Be Loved By You	Doing My Time - Flatt & Scruggs
	AC DC - Highway to hell	Michael Jackson - You are not alone	Dave Brubek - Take Five	Walking in My Sleep - Bill Clifton
	Bon Jovi - It's my life	Michael Jackson - Earth song	Dean Martin - Sway	Great Balls of Fire - New Grass Revival
	U2 - Beautiful day	Bruno Mars - Just the way you are	Louis Armstrong, Ella Fitzgerald - Cheek to Cheek	Truck Driver's Queen - Moore/Bill Napier
	Blink 182 - All the small things	Shakira - Waka Waka	Perry Como - Papa Loves Mambo	Hard Hearted - Jim & Jesse
	Green day - american idiot	Katy Perry - Roar	Peggy Lee - Fever	Gary Allan - Best I ever Had
	Bloodspot - Volcanos	Miley Cyrus - Wrecking ball	Judy Garland - Over the Rainbow	Jim Croce - I'll Have to say I love you in a song
	Eluveitie - the call of the mountaion	Passenger - Let her go	Louis Prima - The Whole World Smiles With You	Cechomor - Promeny
	Cannibal Corpse - hammer smashed face	Robbie Williams - Candy	Dinah Washington - Mad About the Boy	Lidová - Cechy krasne Cechy me
	Rammstein - Ich will	Pharrell Williams - Happy	Frank Sinatra - Let It Snow! Let It Snow! Let It Snow!	Budvarka - Pivovarsti kone
	Rolling stones - Satisfaction	John Newman - Love me again	John Coltrane - My Favorite Things	Buty - Nad stadem koni
Pink floyd - Another brick in the wall	MAGIC! - Rude	Charles Mingus - Peggy's Blue Skylight	First Aid Kit - The Lion's Roar	
	Lunetic - Mama	Jazz mix	Jožka Černý - 8 lidových písní	
	No name - Život			
	BLACK MILK-Nechci tě trápit			
Žánr	ELEKTRO	KLASIKA	HIPHOP	
Interpret - skladba	Skrillex - Bangarang	Madagascar (2005, composed by Hans Zimmer)	Dr. Dre - Still dree	
	Britney spears - Work bitch	Vivaldi	eminem - lose yourself	
	Iggy Azalea - Team	Once Upon a Time in America (1984, composed by Ennio Morricone)	Coolio - Gangsters paradise	
	Will.i.am - Scream	Bach	Cipress hills - Hits from the bong	
	The ecstasy of sain theresa - Dumb it	Harry Potter and the Philosopher's Stone (2001, composed by John Williams)	Put it on - Big L	
	Sigma - Nobody to love	Inception (2010, composed by Hans Zimmer)	2Pac - Ambitionz az a ridah	
	LMFAO - Party rock anthem	Gladiator (2000, composed by Hans Zimmer)	50 Cent - in da club	
	Psy - Gangnam style	Road to Perdition (2002, composed by Thomas Newman)	eminem - rap god	
	Rihanna - Bitch better have my money	Bethoven	eminem - the real slim shady	
	DJ ross Aoky - Lost in disco	Mozart	eazy e - how we do	
	Avicii - Waiting for love	Smetana	snoop dogg - smoke weed every day	
	DIE ANTWOORD - Baby's on fire		wizz khalifa - black and yellow	
	Macklemore - and we danced		ghostface hillah - 2getha baby	
	Bloodhoung gang - The bad touch		the notorious B.I.G. - Juicy	
	Gorillaz - Dare		Jay Z - Anything	
	Krewella - Human		Macklemore - Trift shop	
	Tiesto - Chemicals Twisterz		N.W.A.-Straight Outta Comton	
	Electro & House 2012 Summer Club Mix		Ice Cube - Gangsta Rap Made Me Do It	
Razihel - Renzokuken		Chaozz - Televize		
SPACE-LACES-Say it Aint So		Cro - Bad-Chick		
New-Cool-Collective - Staying-alive		The-Notorious-B.I.G. - Big-Poppa		
Alan Walker - Faded				

Tabulka 11 Nahrávky použité v TS

žánr	ROCK	POP	JAZZ	FOLK
Interpret - skladba	Black sabbath - Ioner	Carly-Rae-Jepsen---Call-Me-Maybe	Chet Baker - My Funny Valentine	Jozka cerny
	Green day - Boulevard of broken dreams	Robbie-Williams---Candy	Nat King Cole - Quizas Quizas Quizas	Jozka cerny
	Heartbreaker-Led Zeppelin	What-Makes-You-Beautiful - One direction	Miles-Davis,-Portia	Jack Johnson - do you remember
	Guns N' Roses - 01- Welcome To The Jungle	Justin-Bieber-Sorry	Dean Martin - Volare	Jack Johnson - Sitting, Waiting, Wishing
	House of Death - Manowar	Alan-Walker---Faded	Frank Sinatra - I Won't Dance	Nedvědi---Stánky
	Steppenwolf - Bom To Be Wild	Gotye - someone that I use to know	Kenny Dorham - Una mas	Doc Watson House Of The Rising Sun
	Megadeth - Holy Wars... The Punishment Due (Live 1992)	Ellie Goulding - Burn	Art Pepper - You'd Be so nice to come home	Stackolee Mississippi John Hurt
	Pantera - Walk	Pharrell Williams - Happy	Bud Powell - Collard Greens and Black Eyed Peas	Ruth Moody - Trouble and Woe
	Judas Priest - Painkiller	Miley Cyrus - Wrecking Ball	Duke Ellington - Take the a train	City of New Orleans -- Steve Goodman
	Red Hot Chili Peppers - Can't Stop	Coldplay - Paradise	John Coltrane - Giant Steps	Woody Guthrie-This Land Is Your Land
žánr	ELEKTRONIKA	VÁŽNÁ HUDBA	HIPHOP	
Interpret - skladba	Psy - Gentleman	W.-A.-Mozart---Lacrimosa	Eminem - brain damage	
	madagascar---i-like-to-move-it	Chopin---Spring-Waltz	2Pac---Ghetto-Gospel	
	Lost-Frequencies---Are-You-With-Me	Tchaikovsky---Waltz-of-The-Flowers	Eminem-feat.-Dr.-Dre---I-Need-A-Doctor	
	DJ-Isaac-&-Crystal-Lake---Stick-Em-(Hard-Mix)	Johann-Sebastian-Bach---Air	Lords-of-the-Underground---Here-Come-The-Lords	
	Noisecontrollers - Only You (Original Mix)	Schubert---Ave-Maria	Dr.-Dre---Keep-Their-Heads-Ringin	
	Milo Bice - I'm Not A Monkey (Casino Edit) [Monkey Jobs]	Wolfgang Amadeus Mozart - Piano Concerto No. 21 - Andante	KRS One - MC's Act Like They Don't Know	
	Milo Bice - Two million years and more	Beethoven -5.	Redman - Blow Your Mind -- - Bohemia After Dark	
	Major Lazer - Light It Up (feat. Nyla & Fuse ODG) [Remix]	Prokofiev - Dance of the Knights	Gangstarr - Full Clip - HD	
	Milo Bice - Kijucko	Dmitri Shostakovich - Waltz No. 2	2Pac feat. Eazy-E - Real Thugs	
	Imany – Don't be so shy (Filatov & Karas Remix)	Entry Of The Gladiators - Julius Fucik	Tupac ft. Notorious B.I.G, Big L - Deadly Combination	

## Příloha C

### Abecední seznam všech vytvořených skriptů a funkcí

Všechny kódy se nacházejí v příloze A v adresáři MATLAB

*analiza\_parametru.m* – použito pro hledání *klíčového parametru* (viz. kap. 3.3.1)  
*automatickaDetekceZanru.m* – hlavní skript pro rozhraní GENDET  
*Databaze1.m* – názvy nahrávek v TR1  
*Databaze2.m* – názvy nahrávek v TR2  
*delkadatabaze.m* – vypočítá celkový čas skladeb a další časy pro *Databaze1.m* a *Databaze2.m*  
*existence.m* – zjišťuje, zda skladba existuje  
*KNN.m* – volá *vzd\_tr\_nezn.m*, *knn.m* a *vysledek.m* pro všechny nahrávky v TS  
*knn.m* – počítá k-nearest neighbours  
*KNNone.m* – stejné jako *KNN.m*, počítá pouze pro jednu nahrávku v TS  
*koeficientZmeny.m* – výpočet *koeficientu změny* – viz. 3.3.2  
*maximum.m* – kontrola, pokud je nahrávka pro detekci příliš krátká  
*maxpocetdilu.m* – jaký je maximální počet dílů, které lze extrahovat z nahrávky  
*NacteniTR.m* – načte TR1 či TR2 – součást *TRA.m* a *TRB.m*  
*NacteniTSA.m* – výpočet parametrů A pro všechny členy TS  
*NacteniTSB.m* – výpočet parametrů B pro všechny členy TS  
*normalizace.m* – normalizace signálu  
*odstraneniNaN.m* – odstranění vadných vzorků  
*ostatni.m* – pro klasifikaci neznámých skladeb (viz. kap. 3.3.3)  
*parametry.m* – součást systému GENDET – počítá MFCC pro všechny detekované nahrávky  
*Spustit\_GENDET.m* – spustí systém GENDET  
*TRA.m* – výpočet parametrů A pro TR  
*TRaTS.m* – skript pro kap. 3.2 – slučuje jednotlivé TS a TR  
*TRB.m* – výpočet parametrů B pro TR  
*TSA.m* – výpočet parametrů A pro jednu nahrávku v TS  
*TSB.m* – výpočet parametrů B pro jednu nahrávku v TS  
*upravaNahravky.m* – základní úpravy nahrávky  
*vypocetParametru.m* – součást GENDET – počítá MFCC pro jednu detekovanou nahrávku  
*vysledek.m* – výsledné zařazení skladby do žánrů  
*vysledekDetekce.m* – součást GENDET – modifikace *vysledek.m*  
*vzd\_tr\_nezn.m* – počítá vzdálenost dílů neznámé nahrávky od TR  
*zanrNejPst.m* – součást GENDET – určí žánr s největší pravděpodobností

### Popis některých kódů vytvořených v Matlabu

#### Výpočet parametrů TR

Byly vytvořeny skripty *Databaze1.m* a *Databaze2.m*. V těchto skriptech jsou informace o nahrávkách pro TR1 a TR2. Jejich výstupem je cell *Data*, v které jsou uloženy názvy nahrávek.

Je třeba rozhodnout, zda se budou z TR počítat parametry A nebo B - pole toho je třeba spustit funkce *TRA.m* nebo *TRB.m*. Funkce vypočítají parametry pro celou TR.

VSTUP funkce

*DelkaUseku* – délka úseku (s)

*Hop* – hop faktor, který určuje, v kolika procentech předchozího fragmentu bude začínat nový fragment

*D* – délka jednoho fragmentu (s)

*CutOff* - definuje, kolik vteřin se vyjme na začátku a na konci nahrávky (s)

*databaze* – zda se použijí nahrávky z *Databaze1.m* nebo *Databaze2.m*.

VÝSTUP funkce

*trenovacimnozina* – matice velikosti *počet parametrů x počet úseků x počet žánrů*. Obsahuje vypočítaná data parametrů. Výstup je přejmenováván dle vzoru *TR\_N\_X\_D*, kde N je 1 nebo 2 pro různé trénovací množiny (1,2), X je počítaná skupina parametrů (A,B) a D je délka úseku (např. 1s)

Je zde obsažena funkce *nacteniTR.m*. Funkce určí, kolik *fragmentů* se vyjme z každé nahrávky tak, aby byl stejný počet pro všechny žánry.

Tím je dokončen proces učení.

### Výpočet parametrů TS

Funkce *nacteniTSA.m* a *nacteniTSB.m* za použití funkcí *TSA.m* resp. *TSB.m* vytvoří testovací množinu TS.

VSTUP funkce

*d* – počet extrahovaných dílů

*c* – CutOff (s)

*b* – délka dílu (s)

VÝSTUP funkce

*TS* – cell(1, počet nahrávek), v každé buňce je obsažena matice (*d*, počet parametrů). Jednotlivé výstupy jsou pojmenovávány obdobně jako v případě trénovací množiny.

Tím je dokončena tvorba TS.

### Vyhodnocení

Funkce *KNN.m* Vyhodnotí TS a TR. Je složena ze tří pomocných funkcí. U každé skladby vypočte vzdálenosti parametrů jednotlivých dílů od úseků pomocí funkce *vzd\_tr\_nezn.m*.

Poté funkce *knn.m* určí *k-nejbližších sousedů*. Funkce *vysledek.mat* následně vyhodnotí

*knn.m* a jejím výstupem je pravděpodobnost výskytu žánrů v nahrávce.

VSTUP funkce

*TR* – viz výše

*TS* – viz výše

*k* – počet k pro model *k – nearest neighbours*

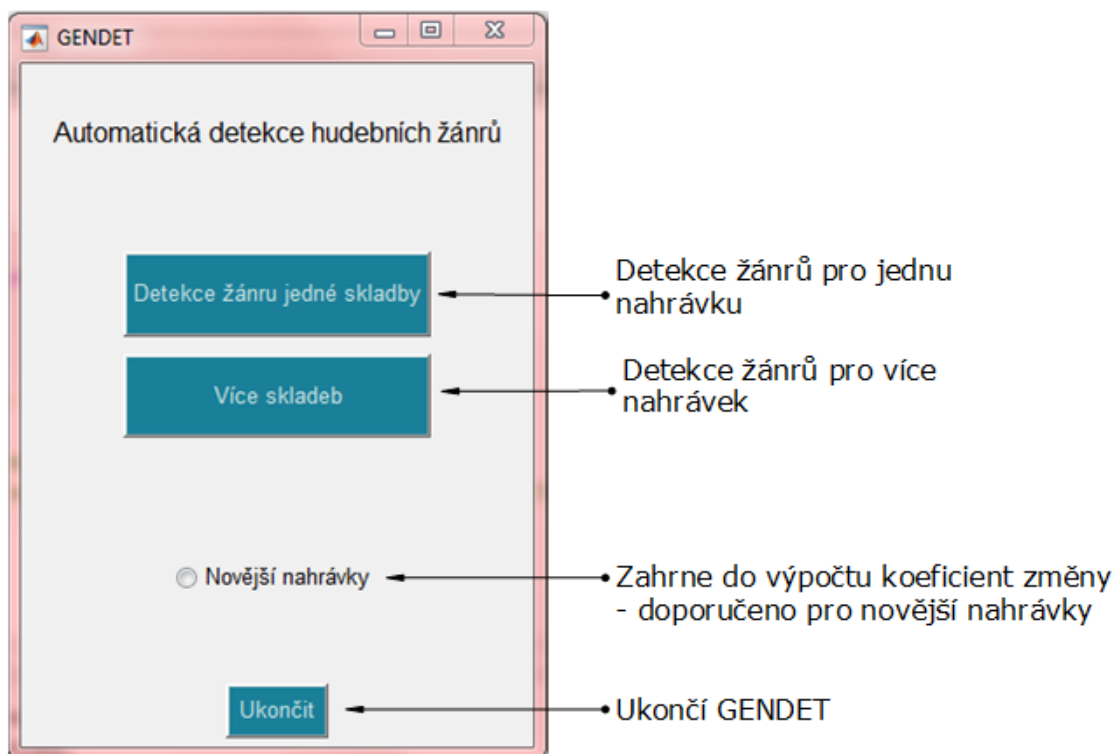
VÝSTUP funkce

*zarazeni* – cell(1, počet nahrávek v TS). V každé buňce je v procentech vyjádřeno pravděpodobnostní zastoupení jednotlivých žánrů nahrávky



## Příloha D

System GENEDET se nachází v příloze A v adresáři GENEDET, kde je též návod k použití. Spouští se pomocí *Spustit\_GENEDET.m*. Na obr. 22 je popis uživatelského prostředí.



Obrázek 22 Popis rozhraní GENEDET

	A	B	C	D	E	F	G	H	I	J
1	Nazev Skladby	Rock	Pop	Jazz	Folk	Elektronika	Vazna hudba	Hip Hop	Maximum	Poznamka
2	Iron Man.mp3	64,71	6,00	2,14	5,71	5,14	0,86	15,43	Rock	
3										

Obrázek 23 Výstup GENEDET v Microsoft Excel

Na obr. 23 jednotlivá čísla představují s jakou pravděpodobností (%) patří skladba do daného žánru. V sekci „Maximum“ je nejpravděpodobnější žánr. Pokud je v sekci „Poznamka“ obsažen „!“; nahrávka pravděpodobně nepatří ani do jednoho žánru (viz. kap. 3.3.3).