ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická

Katedra radioelektroniky

Identifikace obsahu archivních zvukových záznamů

Archive Audio Record Content Identification

Bakalářská práce

Studijní program: Komunikace, Multimédia a Elektronika

Studijní obor: Multimediální technika

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

Ekaterina Koshkina

Praha 2015

České vysoké učení technické v Praze
Fakulta elektrotechnická

katedra radioelektroniky

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student: **Ekaterina Koshkina**

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

Název tématu: **Identifikace obsahu archivních zvukových záznamů**

Pokyny pro vypracování:

Seznamte se s problematikou identifikace obsahu audio stopy a se specifiky archivních analogových zvukových záznamů. Ověřte použitelnost vybrané metody pro identifikaci obsahu (řeč, hudba, šum, ...) archivních nahrávek (např. optická zvuková stopa u archivních filmů). Vzorky archivních nahrávek dodá vedoucí práce.

Seznam odborné literatury:

[1] J. Foote, Content-based retrieval of music and audio, In Multimedia Storage and Archiving Systems II, Proc. of SPIE, 1997, vol 3229, pp. 138-147
[2] Godsill, S. J., Rayner, P. J. W. Digital Audio restoration. 1st. ed., London: Springer 1998. ISBN 3-5407622-1

Vedoucí: Ing. František Rund, Ph.D.

Platnost zadání: do konce letního semestru 2015/2016

L.S.

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

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 10. 2. 2015

Prohlášení

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

Dne 21. května 2015 v Praze

...……...

Poděkování

Děkuji svému vedoucímu bakalářské práce Ing. Františku Rundovi, Ph.D. za věcné připomínky, dobré rady, vstřícnost při konzultacích a pomoc při vypracovávání bakalářské práce.

Abstrakt

Tato práce se zabývá problematikou identifikace (rozpoznávaní) obsahu archivních zvukových záznamů. Cílem je seznámení s danou problematikou a realizace vybraného algoritmu identifikace v programovém prostředí MATLAB. Hlavními kroky v realizovaném algoritmu jsou segmentace, parametrizace a klasifikace. Implementovaný algoritmus využívá klasifikátoru *k-Nearest Neighbors* (*KNN*). Výstupem této práce je systém, který je schopný identifikovat jednotlivé typy zvukových záznamů (řeč, hudba, hluk, ticho) ve vstupní archivní nahrávce. Systém odhaduje pravděpodobnost výskytu jednotlivých typů zvuku v závislosti na čase. Funkce algoritmu je ověřena na ukázkách nahrávek z archivních filmů.

Klíčová slova: identifikace, audio záznam, archivní záznam, segmentace, parametrizace, klasifikace, KNN, MATLAB

Abstract

This bachelor thesis introduces the problem of content-based identification (recognition) of archive audio records. The goal of this thesis is to familiarize with the problem and to implement the selected method of the identification in MATLAB environment. The main steps of the algorithm include segmentation, feature extraction and classification. The algorithm uses *k-Nearest Neighbors* (*KNN*) classifier. The outcome of this work is represented by a system which is capable of identifying different types of audio records (speech, music, noise, silence) in the input archive record. The system estimates the time relevant probability of occurrence of mentioned types of audio records. Functionality of the algorithm is verified with records from archive movies.

Key words: identification, audio record, archive record, segmentation, feature extraction, classification, KNN, MATLAB

Obsah

Seznam obrázků a tabulek

Seznam obrázků

Seznam tabulek

Seznam použitých zkratek

KNN k-Nearest Neighbors

NN Nearest Neighbor

ANN Artificial Neural Network

GMM The Gaussian Mixture Model

HMM The Hidden Markov Model

RMS Root-Mean-Square

ZCR Zero-Crossing Rate

SF Spectrum Flux

HCDF Harmonic Change Detection Function

EM Expectation Maximization

Úvod

Cílem této bakalářské práce je seznámení s problematikou identifikace obsahu audio stopy, se zaměřením se na archivní analogové zvukové záznamy a implementace vybrané metody identifikace.

Identifikací (rozpoznáváním) obsahu audio stopy se myslí schopnost rozlišování mezi různými typy zvuků za pomocí různých algoritmů z oblasti zpracování signálů. Identifikace je většinou založená na extrakci určitých parametrů signálu a dále třídění jeho částí do předem definovaných tříd (řeč, hudba, šum, apod.).

V současné době je rozpoznávání zvukových záznamů aktuálním tématem a může být velmi užitečnou pomůckou například při hromadné analýze velkých zvukových archivů, se kterými je možné se setkat například při zpracovávání starších analogových záznamů.

Specifikem analogových záznamů bývá, že se jejich kvalita používáním snižuje. Nejen, že analogové záznamy jsou náchylné na mechanické poškození, ale degradace může nastávat i samovolně časem. Pro delší zachování kvality zvukové informace se proto analogové záznamy digitalizují. Získané digitální záznamy se dají dále počítačově analyzovat a zpracovávat. Počítačové zpracování umožňuje zároveň restaurovat nahrávky, které byly v době digitalizace poškozené. Problematikou restaurace jsem se zabývala v individuálním projektu I. Při restauraci nahrávek může někdy být prospěšné zpracovávat pouze požadované části záznamu (řeč, hudba, šum, apod.), bez předběžného poslechu celých nahrávek. Proces rozpoznávání záznamu je jednou z možných etap analýzy digitalizovaného signálu, který mnohdy urychluje a zjednodušuje práci s nahrávkami, například při přípravě subjektivních testů.

Tato práce se zabývá především analyzováním zvukového signálu a jeho časovým rozdělením podle charakteru obsahu. V dnešní době existuje velké množství algoritmů na rozpoznávání zvukových záznamů založených na různých principech. Algoritmus, implementovaný v rámci této práce, těchto principu využívá. Výstupem implementovaného algoritmu je potom pravděpodobnostní rozložení typů signálu, které může být použito například pro efektivní vyhledávání v dlouhých zvukových záznamech.

Zbytek práce je členěn následovně. **Kapitola 1** se zabývá základním popisem charakteristik archivních zvukových záznamů a jejich odlišností od záznamů nearchivních.

Ve **2. kapitole** je probrána teorie rozpoznávání obsahu zvukových záznamů. Zprvu jsou zde zmíněny různé parametry charakterizující signál v časové a spektrální doméně a dále potom metody klasifikace, běžně užívané při rozpoznávání libovolných signálů. **Kapitola 3** popisuje praktickou část této práce, tedy zejména realizaci algoritmu v programovém prostředí MATLAB.

1 Archivní audio záznam

Archivním audio záznamem se myslí především starší analogová zvuková nahrávka (např.: optická analogová zvuková stopa na filmovém pásu, analogová magnetofonová nahrávka apod.). Velké množství těchto archivních záznamů je v dnešní době digitalizováno a rekonstruováno.

Archivní záznam má často své specifické nedokonalosti. Na rozdíl od záznamů nearchivních může archivní nahrávka obsahovat hodně nežádoucích zvukových artefaktů, vznikajících různým způsobem. Tyto artefakty mohou komplikovat identifikaci obsahu audio stopy.

Zvukové nedokonalosti mohou být zapříčiněny způsobem získávání záznamu v minulosti. Dříve byly používány jiné typy medií a řetězce zpracování, neumožňující uchování záznamu bez nedokonalostí. Dalším důvodem vzniku nedokonalostí může být stárnutí záznamu. Poruchy mohou vznikat i při digitalizaci nahrávky. Problematikou této kapitoly se zabývá např. literatura [10].

Mezi hlavní zvukové artefakty archivních zvukových záznamů patří zejména různé druhy šumů, zkreslení, lupanců a praskotu.

Šum

Šum je ve zvukových záznamech většinou nežádoucím širokopásmovým signálem. Jde o signál stochastický, tedy náhodný signál, u kterého nelze předem determinovat přesný tvar jeho vlny, a proto není jednoduché ho potlačit nebo zcela odstranit.

Lupance a praskot

Jde o rušivé artefakty impulsního charakteru. Lupance jsou nízkofrekvenční impulzy a praskotem se myslí impulzy o malé amplitudě a vysoké frekvenci.

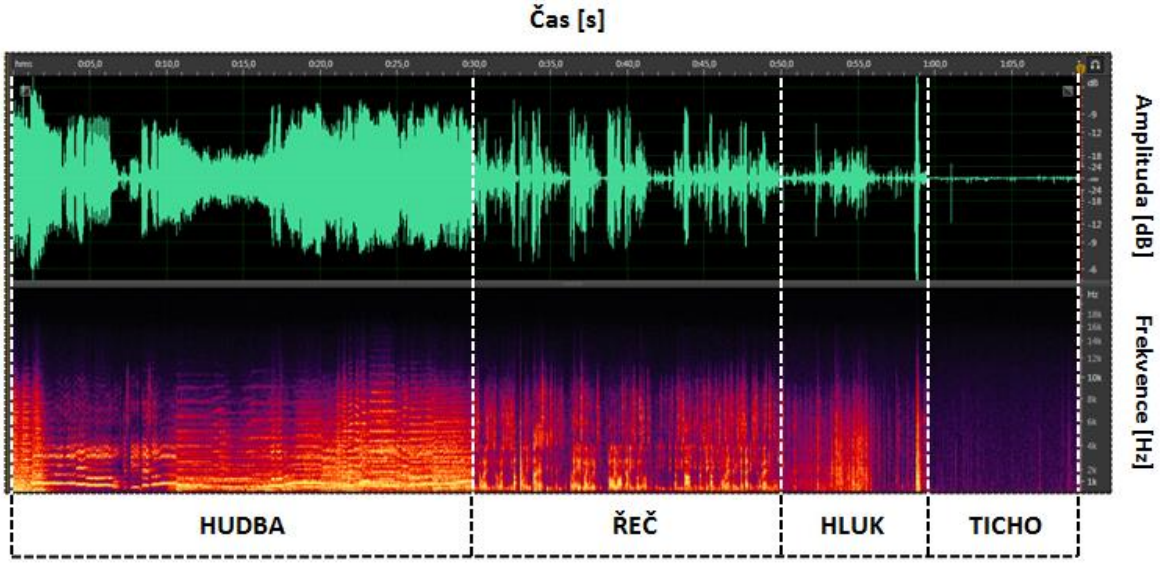
Zkreslení

Zkreslení lze rozdělit na lineární a nelineární. Lineární zkreslení je jevem spojeným s omezením amplitudy užitečného signálu na analogové nebo digitální úrovni. Nelineární zkreslení jsou charakterizované výskytem nových frekvenčních složek ve spektru signálu a je možné se s nimi často setkávat u analogových záznamů.

Omezení šířky pásma

Analogové záznamy mohou mít omezenou šířku pásma, neboli rozdíl mezi maximální a minimální frekvencí přenášeného signálu.

Na *obr. 1.1* je znázorněn starší archivní zvukový záznam extrahovaný z filmu „Božská Ema“ z roku 1979. V oblasti ticha je možné pozorovat některé zmíněné artefakty. V této části záznamu se nenachází žádný užitečný signál a průběh, který lze pozorovat v časovém vývoji obálky je signálem rušivým. Šum je dobře viditelný ve spektrální oblasti jako širokopásmový signál na malých úrovních. Lupance jsou patrné v časovém vývoji spektra jako tenké barevné špičky nebo jako „ostré štíty“ v průběhu zvukové obálky (waveform). Omezení šířky pásma je možné pozorovat ve spektru, kde se přibližně nad frekvencí 10 kHz vyskytuje už pouze nepatrný zlomek celkové energie signálu.



Obr. 1.1 Archivní záznam (horní část: průběh obálky (waveform), dolní část: spektrogram)

Pro obnovení některých chybějících parametrů signálu nebo naopak odstranění nežádoucích příměsí (vizte *obr. 1.1*) se provádí rekonstrukce a restaurace zvukového záznamu. Rekonstrukci digitálního záznamu lze realizovat pomocí číslicového zpracování signálu.

2 Identifikace obsahu zvukových záznamů

Identifikace (rozpoznáváni) obsahu audio stopy je, jak už bylo poznamenáno v úvodu, třídění části signálů do určitých tříd. Proces rozpoznávání zvuku se dá rozdělit do několika částí.

Za jakýsi přípravný krok se dá považovat předzpracování záznamu. Dle charakteru vstupního signálu a způsobu dalšího zpracování se mohou úpravy velmi lišit. Nejčastěji se ale těmito úpravami myslí změny ve frekvenční oblasti nebo dynamické úpravy, případně nějaké způsoby rekonstrukce poškozených nahrávek. Úkolem předzpracování je normalizovat signál pro další kroky systému.

Po předzpracování následuje **segmentace** a **parametrizace** (vizte kap. 2.1) audio signálu. Segmetací se myslí rozdělování signálu na kratší časové úseky, což je vhodné pro parametrizaci, tedy proces získávání určitých vhodných parametrů ze signálu. Důvodem segmentace je proměnnost struktury signálu v čase. Obecně však délka úseku záleží na účelu a použitém algoritmu identifikace. Typické délky jsou 1 s a 2,4 s [1, 6]. Například autor článku [1] uvádí, že používáním délky segmentů kratší 2,4 s nelze s jednoduchými parametry docílit dobrých výsledků. Získané segmenty jsou dále rozdělovány na další, ještě kratší časové úseky, aby bylo možně co nejpřesněji odhadnout parametry. Typické délky pod-segmentů se pohybují mezi 10 ms a 40 ms [25]. Pod-segmenty se obvykle váhují speciální okny potlačujícími okrajové hodnoty a volí se mezi nimi určitá míra překryvu (typicky polovina nebo čtvrtina délky pod-segmentu). V této kapitole jsou krátce popsány parametry, které dále budou využívány v praktické části (vizte kap. 3).

Posledním krokem rozpoznávání je proces **klasifikace**, tj. třídění částí signálu do předem definovaných tříd (řeč, hudba, šum, apod..). Klasifikace se týká oblasti strojového učení. Mezi způsoby strojového učení patří učení s učitelem a učení bez učitele. Toto rozdělení je dále popsáno dle [35].

Metoda **učení s učitelem** je založená na učení funkce na základě tzv. **trénovací množiny dat**. Tedy existuje trénovací množina vzorů, která je složená z rysů vstupních objektů (vektorů příznaků) a požadovaných výstupů (označení tříd vstupních objektů). Cílem je najít neznámou závislost mezi vstupy a výstupy.

Učení bez učitele znamená případ, kdy není algoritmu poskytována označená trénovací množina. Při takovém učení se atributy vstupních objektů rozdělují na neprotínající se podmnožiny, tzv. klastry. Každý klastr se skládá ze shodných atributů objektů a rysy objektů různých klastrů se značně liší. Tedy není předem určené, jestli je vstupní objekt známý a patří do nějakého předem známého klastru, tj. třída vstupního objektu není předem známá.

Přesnost klasifikačního modelu lze například určit pomocí tzv. metody ***Cross-Validation (křížové validace)***. Přesnost klasifikace vyjadřuje míru spolehlivosti modelu správně klasifikovat data, na která model nebyl trénován, tedy data jemu neznámá. Metoda *Cross-Validation* je založená na hodnocení přesnosti pomocí dat na základě tzv. **testovací množiny**. Testovací množinou se myslí množina vzorů, pomocí které se hodnotí použitelnost modelu klasifikace [36]. Přesnost klasifikace testovací množiny se porovnává s přesnosti klasifikace trénovací množiny. To znamená, že se zprvu testuje testovací množina za pomocí trénovací. Potom dojde prohození trénovacích dat za testovací a testuje se trénovací množina.

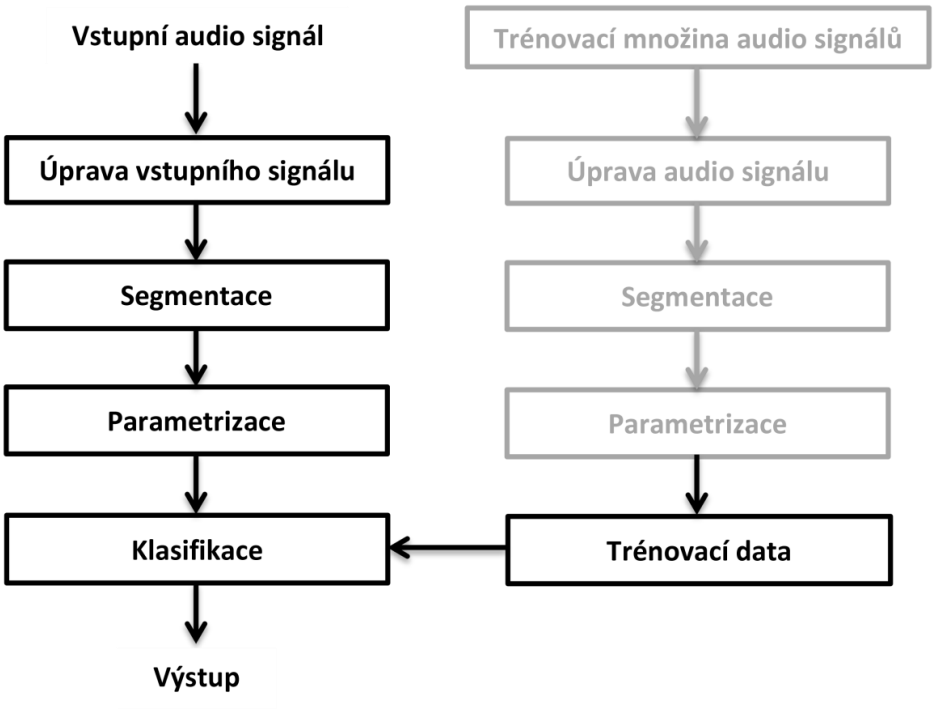
Rozdělení na trénovací a testovací sekvenci se provádí tak, že se vzorky dat rozdělí v určitém poměru. Trénovací množině se například přidělí dvě třetiny všech dat a testovací zbylá jedna třetina. [29, 36]

Dle literatury [1, 2, 25, 37, 41, 42] se pro klasifikaci zvukových signálů obvykle používají metody:

- *k-Nearest Neighbors (KNN)* [1, 25, 42],
- *Artificial Neural Network (ANN)* [41],
- *The Gaussian Mixture Model (GMM)* [2, 25, 37],
- *The Hidden Markov Model (HMM)* [2, 41].

Tyto metody jsou popsané v kapitole 2.2.

Základní princip klasifikace spočívá v porovnání vstupního signálu s trénovací množinou vzorů. Níže na *obr. 2.1* je zobrazeno ilustrační blokové schéma základního principu klasifikace zvukových záznamů.



Obr. 2.1 Základní princip identifikace audio záznamů

2.1 Parametrizace

Parametrizace je zásadním krokem v systémech rozpoznávání obsahu zvukových záznamů. V této kapitole jsou probrány jak jednoduché, běžně používané parametry pro rozpoznávání zvuku, tak parametry komplexnější, obvykle sloužící k charakterizaci hudebních signálů [4, 34].

Mezi jednodušší parametry se řadí například *Root-Mean Square Energy* (*RMS Energy*), *Zero-Crossing Rate* (*ZCR*) nebo *Delta Spectrum Magnitude* (*Spectrum Flux – SF*) [1, 21].

Popis všech parametrů probíraných v této kapitole vychází především z manuálu MIRtoolboxu 1.6.1 [4] a proto zde na něj nebude dále odkazováno.

2.1.1 Zero-Crossing Rate (ZCR)

Zero-Crossing Rate je užitečný parametr zejména pro klasifikaci řečových a hudebních signálů [1].

Jedná se o jednoduchou charakteristiku, udávající počet výskytů průchodů signálu nulou za jednotku času (tj. počet přechodů signálu od kladné hodnoty amplitudy k záporné a naopak [3]). *ZCR* zároveň částečně poskytuje informaci o frekvenčním obsahu signálu.

Zero-Crossing Rate je definován následujícím vzorcem:

$$ZCR = \frac{1}{N-1} \cdot \sum_{n=1}^{N-1} f\{s_n \cdot s_{n-1} < 0\}, \qquad (2.1)$$

kde *f* je charakteristická funkce, *s* je signál a *N* je jeho délka (ve vzorcích) [3].

Cílem algoritmu je najít v krátkém časovém úseku signálu všechny průchody nulou a uvést jejich počet. Průchody nulou lze nalézt vynásobením dvou po sobě jdoucích vzorků signálu. Záporný výsledek tohoto součinu značí průchod nulou v místě mezi danými vzorky.

U řečových signálů vychází *ZCR* obvykle větší než u signálů hudebních. Různé typy hudby mají *ZCR* na rozdíl od řeči podobné. Pro znělé a neznělé zvuky řeči se hodnota parametru liší. Hluk je oproti hudbě a řeči pomocí *Zero-Crossing Rate* mnohem těžší rozlišit. Existují různé druhy hluků a jejich charakteristiky mohou být dost rozdílné, tudíž mohou mít odlišné i *ZCR* [1, 2]. Například bílý šum je podle autorů článku [1] charakterizován menším počtem průchodů signálu nulou, než zvuk bubnu.

2.1.2 Root-Mean-Square Energy (RMS energy)

Jde o parametr vyjadřující energii signálu. Tato energie se dá jednoduše vypočítat pomocí kvadratického průměru. Jde o druhou odmocninu aritmetického průměru druhých mocnin daných hodnot signálu *s* v určitém časovém úseku. Parametr *RMS* pro signál délky *N* (ve vzorcích) je dán vztahem [5]:

$$RMS = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^{N} s(n)^2} \qquad (2.2)$$

Časový rozvoj energie signálu se dá získat pomocí nalezení *RMS* v krátkých časových úsecích (vizte kap. 2). Získaný průběh energie se dá označit jako obálka signálu.

2.1.3 High-Frequency Energy (Brightness)

Autor článku [8] definuje *Brightness* (*jasnost*) jako jednou z důležitých charakteristik barvy zvuku a považuje jí za silný vjemový rozdíl mezi zvuky. Zvuky s vysokou energií na vysokých frekvencích mají větší hodnotu *Brightness*, než zvuky s malou energií na vysokých frekvencích [8]. Metoda jednoduše spočívá v měření množství energie nad určitou frekvencí.

2.1.4 High-Frequency Energy (Spectral Roll-off)

Spectral Roll-off je charakteristikou spektra signálu principiálně podobnou parametru *Brightness*. Parametr je definován jako frekvence *X(R)*, pod kterou se nachází určité procento spektrálního výkonu *p* [17]. V literatuře [21] je poměr stanoven na hodnotu 85 % a v literatuře [17] na 95 %. Parametr *Spectral Roll-off* lze vypočítat pomocí vztahu:

$$\sum_{n=1}^{R} X(n) = p \cdot \sum_{n=1}^{N} X(n), \tag{2.3}$$

kde $X(n)$ je hodnota amplitudy vzorku spektra, R odpovídá hledané frekvenci *Spectral Roll-off* a N je počet vzorků spektra [17].

2.1.5 Delta Spectrum Magnitude (Spectrum Flux, SF)

Delta Spectrum Magnitude je charakteristikou signálu, která popisuje rychlost změn spektrální obálky. Parametr lze nalézt ve spektrální oblasti, porovnáním spektra jednoho časového rámce se spektrem předchozího [6]. Parametr *SF* se dá vypočítat jako součet kvadrátů rozdílů dvou časově sousedních spektrálních amplitud.

$$SF = \sum_{n=1}^{N/2} (X_t(n) - X_{t-1}(n))^2, \tag{2.4}$$

kde N je počet vzorků spektra, $X_t(n)$ je hodnota amplitudy vzorků spektra v době rámce t, a $X_{t-1}(n)$ je hodnota amplitudy v předchozím časovém rámci t-1 [21].

Dle literatury [6] má hudba vyšší rychlost změn, než řeč, a tudíž i hodnota *Spectrum Flux* je u hudby větší.

2.1.6 Sensory Dissonance (Roughness)

Odhad parametru *Roughness* (*drsnost*) je spojený se záznějovým jevem. Záznějový jev je interference dvou tonů, jejichž frekvence jsou podobné, a projevuje se v periodickém snižování a zvyšování amplitudy složeného signálu (periodické změny hlasitosti). Frekvence změny amplitudy složeného signálu se rovná rozdílu frekvencí původních signálů [20].

Hodnota *Roughness* se dá získat výpočtem všech vrcholů spektra a následným zprůměrováním disonancí mezi všemi možnými páry vrcholů [4].

2.1.7 Spectral Peaks Variability (Irregularity)

Parametrem *Irregularity* se dle literatury [4] myslí nerovnoměrnost tvaru spektra a značí míru změny po sobě jdoucích vrcholů. Signál s naprosto rovným spektrem bude mít tento parametr nulový, což vyplývá i z níže uvedeného vztahu [24]:

$$IRR = \sum_{k=2}^{N-1} \left| a_k - \frac{a_{k-1} + a_k + a_{k+1}}{3} \right|, \tag{2.5}$$

kde a_k je hodnota amplitudy vzorku spektra a *N* je počet vzorků. Další možnost výpočtu je definována ve [23] jako součet kvadrátů rozdílu amplitud sousedních vzorků, podělený sumou kvadrátů všech vzorků vztah (2.6).

$$IRR = \left(\sum_{k=1}^{N} (a(k) - a(k+1))^2\right) / \sum_{k=1}^{N} a(k)^2 \tag{2.6}$$

2.1.8 Mode

Mode (*modus*) je parametr popisující vztah sladěnosti mezi jednotlivými zvuky (tóny). Tónové vzdálenosti ve stupnicích závisejí na použitém ladění, které taky určuje frekvence jednotlivých tónů a poměry mezi nimi. Existují různá uspořádání tónů (tonalita) ve stupnici: v evropské hudbě jsou uspořádání založená na durových (major) a mollových (minor) stupnicích. [19]

2.1.9 Novelty Curve

Jedná se o konkrétní statistický popis, vztahující se k časové posloupnosti jednotlivých momentů charakterizovaných určitými hudebními vlastnostmi [13]. Parametr *Novelty Curve* reprezentuje časovou pravděpodobnost výskytu přechodů mezi jednotlivými po sobě

následujícími stavy (např. změna tempa), označených vrcholy, a taky jejich relativní význam, označený výškou vrcholů. Hodnoty *Novelty* odpovídají kombinaci dvou faktorů: časový rozsah předchozího konce segmentu a množství kontrastní změny před a po ukončení segmentu [12].

2.1.10 Pulse Clarity

Pulse Clarity je charakteristika, která udává, jak snadno mohou posluchači v dané hudební skladbě nebo v určitém časovém okamžiku této skladby, vnímat základní rytmickou nebo metrickou pulzaci [9].

2.1.11 Harmonic Change Detection Function (HCDF)

Parametr *HCDF* je definován jako celková rychlost změny vyhlazeného tónového těžiště signálu [7]. Je definován vztahem:

$$\xi_n = \sqrt{\sum_{d=0}^{5} (\varsigma_{n+1}(d) - \varsigma_{n-1}(d))^2}, \qquad (2.7)$$

kde ξ_n je Euklidova vzdálenost mezi vektory tónového těžiště ς_{n+1} a ς_{n-1}, a d je dimenze vektoru tónového těžiště.

Vrcholy v signálu ukazují přechody mezi tónovými oblastmi, které jsou harmonicky stabilní.

2.1.12 Attack Time

V časové oblasti mohou být elementy signálu (především hudebního) popsány tzv. ADSR obálkou (*Attack, Decay, Sustain, Release*), vyjadřující časový rozvoj amplitudy. *Attack* (*náběh*) je částí obálky, při které se s časem prudce zvyšuje amplituda signálu [16].

Parametr *Attack Time* je jedním ze způsobu popisu *Attack* (*náběhu*), jde o zjišťování časového trvání oblasti, ve které signál strmě narůstá a dostává se na své lokální maximum.

2.1.13 Attack Leap

Parametr *Attack Leap* je dalším ze způsobů popisu *Attack* (*náběhu*) a spočívá v odhadování velikosti jeho amplitudového skoku.

2.2 Metody klasifikace

V této kapitole jsou popsány metody klasifikace (klasifikátory), které se dle literatury [1, 2, 25, 37, 41, 42] obvykle používají při rozpoznávaní zvukových signálů. Princip všech vybraných metod klasifikace je proto v této kapitole popisován s ohledem na klasifikaci zvukových signálů.

2.2.1 *k*-Nearest Neighbors (KNN)

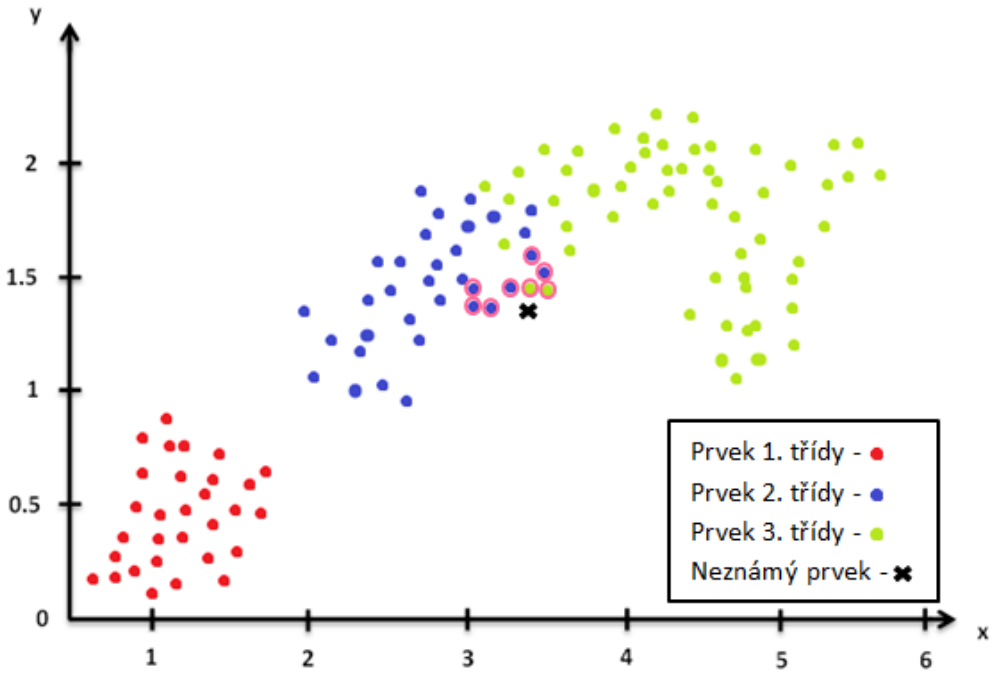
Metoda *k-Nearest Neighbors* (*k-nejbližších sousedů*) je statistickou metodou klasifikace objektů. Jedná se o metodu pro učení s učitelem, tj. metoda strojového učení pro učení funkce z trénovacích dat (vizte kap. 2). Učení (trénování) je důležitým přípravným procesem pro klasifikaci. Trénovací sekvence obsahuje určité množství dat, patřících do konkrétních tříd, do kterých jsou potom testované části audio signálu ve fázi klasifikace zařazovány. Vytvořená trénovací množina je umístěná do některého místa N-rozměrného prostoru. Dále následuje proces klasifikace, kdy se umísťuje testovaný prvek do téhož prostoru a měří se *k*-nejbližších vzdáleností od testovaného prvku k trénovacím prvkům, tj. nalezení *k*-nejbližších sousedů. Objekt je pak klasifikován do té třídy, kam patří většina z těchto nejbližších sousedů. [26, 27, 29]

Pro hledání nejbližší vzdálenosti ve množině lze použít různé metriky. Dle literatury [25] je nejobvyklejší *Euklidovská metrika*, která je daná vztahem:

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}, \tag{2.8}$$

kde $x = x_1, x_2, \ldots, x_m$ a $y = y_1, y_2, \ldots, y_m$ jsou hodnoty atributů.

Ilustrační příklad klasifikace *KNN* je znázorněn na *obr. 2.2.* Počet hledaných sousedů *k* lze určit pomocí statistické metody *Cross-Validation* (vizte kap. 2). Z principu metody *Cross-Validation* vyplývá, že pro určení vhodné hodnoty *k*, je potřeba provést tuto metodu s různými *k* a potom vybrat hodnotu optimální [26, 29]. Hodnota *k* je celé číslo a může se pohybovat od 1 do počtu vzorků dat trénovací sekvence. Pokud $k = 1$, jde o speciální zjednodušený případ, tzv. metodu *Nearest Neighbor* (*NN*) [27]. Na *obr. 2.3* je zobrazeno ilustrační blokové schéma algoritmu metody *KNN*.



Obr. 2.2 Ilustrace klasifikace k-Nearest Neighbors (Euklidovská metrika, k = 8)

Obr. 2.3 Algoritmus metody KNN

Metoda klasifikace *KNN* má své výhody a nevýhody [29, 38, 39].

Výhody:

- Velmi jednoduchý algoritmus
- Málo vstupních parametrů klasifikátoru (počet *k* a metrika)
- Učení je založeno na zapamatování trénovací množiny
- Efektivní a přesný

Nevýhody:

- Může být pomalým algoritmem (algoritmus musí vypočítat a roztřídit vzdálenosti mezi všemi prvky trénovací množiny, čím větší počet trénovacích dat, tím menší je rychlost klasifikace)
- Vyžaduje hodně paměti
- Výsledek zaleží na výběru metriky a počtu *k*

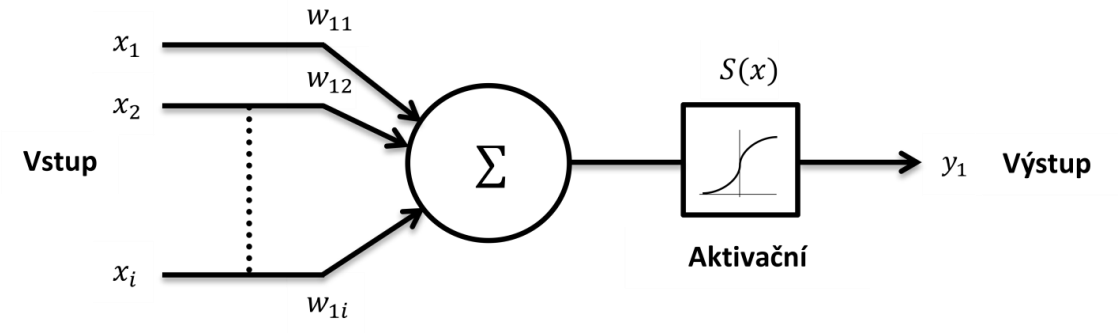
2.2.2 Artificial Neural Network (ANN)

Artificial Neural Network (*umělá Neuronová síť*) je matematický model, jehož algoritmus napodobuje ve svém stylu chování biologické struktury (sítě nervových buněk živého organismu) [33].

Neuronová síť znamená systém spojených a vzájemně působících jednoduchých procesů (umělých neuronů). Hlavní funkcí umělého neuronu je generování výstupního signálu v závislosti na signálech přivedených na jeho vstupy. Neuron má libovolný počet vstupů, ale pouze jeden výstup. Vstupní data prochazejí jednosměrnými vazbami tzv. synapsemi, kde každá synapse je chrakterizovaná vahovým koeficentem w_{ji}. Každá vstupní proměná x_i *j*-tého neuronu je násobena tímto váhovým koeficientem w_{ji}. Dále se tyto násobky sčítají a výsledný součet je v neuronu transferován určitou nelineární přenosovou funkcí $S(x)$, tzv. aktivační funkcí, pomocí které se počítá výstupní signál umělého neuronu. Výstup základního modelu neuronu y_j je dán vztahem (2.9) a grafické znázornění vztahu lze pozorovat na *obr. 2.4.*

$$y_j = S\left(\sum_{i=1}^{n} x_i \cdot w_{ji}\right), \tag{2.9}$$

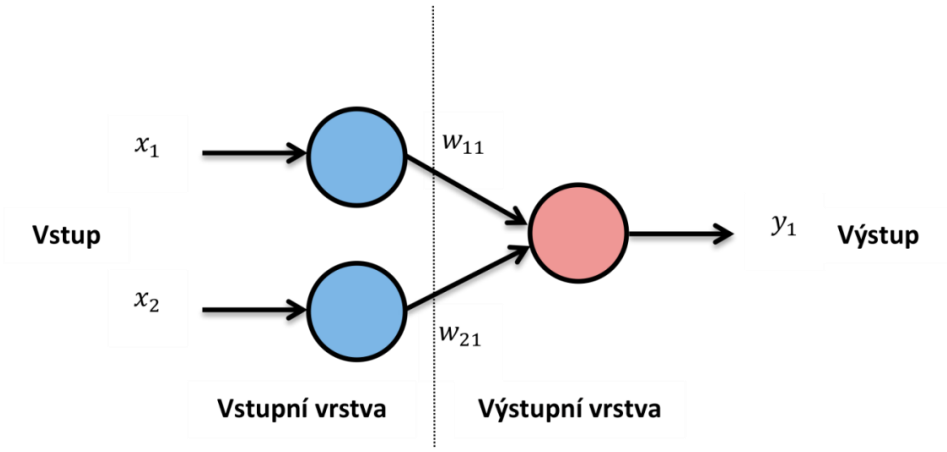
kde *n* je počet vstupních proměných x_i. [30, 31].



Obr. 2.4 Základní model umělého neuronu

Neuronová síť může být *jednovrstvá* a *vícevrstvá*.

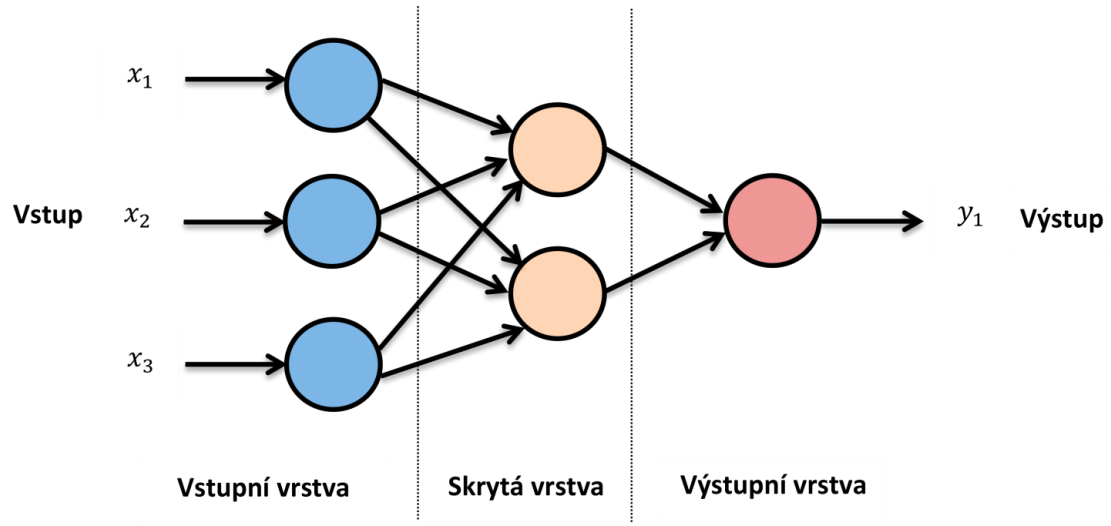
Jednovrsvtá Neuronová síť (*obr. 2.5*) má jenom jednu vrstvu, jedná se tedy o síť, ve které jsou všechny vstupní proměné přímo spojené s výstupy [29].



Obr. 2.5 Jednovrstvá Neuronová síť

Vícevrstvá síť (*obr. 2.6*) obsahuje několik vrstev (vstupní vrstva, několik skrytých vrstev a výstupní vrstva). Neurony se v jednotlivých vrstvách mezi sebou vzájemně neovlivňují, ale jsou závislé na neuronech z předchozí a následující vrstvy. Zároveň počet neuronů v jednotlivých vrstvách nezávisí na počtu neuronů ve vrstvách jiných. Data se v každé vrstvě zpracovávají paralelně.

Vstupní vrstva obsahuje vstupní neurony, které přijímají signál a posílají ho na vstupy neuronů následující skryté vrstvy. Transferovaný signál postupně přechází do dalších skrytých vrstev, dokud nedojde do poslední výstupní vrstvy, jejíž výstupy lze považovat za konečné výstupy celého modelu *Neuronové sítě*. [29]



Obr. 2.6 Vícevrstvá Neuronová síť (3 vstupní neurony, 1 výstupní neuron)

Důležitou fází principu funkce *Neuronových sítí* je učení. Pro účely klasifikace pomocí neuronových sítí se používá učení s učitelem (vizte kap. 2). Při učení s učitelem, je třída dat trénovací množiny předem známá. Tato trénovací množina je poslána do *Neuronové sítě*, která poskytne určitý výsledek. Tento výsledek se porovná s výsledkem požadovaným (třídy dat testovací množiny) a určí se chyba. Poté se upravují hodnoty váhových koeficientů tak, aby chyba byla minimální. [32, 29]

Přesnost klasifikace může být vyhodnocená pomocí metody *Cross-Validation* (vizte kap. 2) [29].

V literatuře [29] a [40] je uvedený popis kladů a záporů *ANN*.

Výhody:

- Algoritmus napodobuje funkci nervové soustavy živého organismu
- Umožňuje efektivně tvořit nelineární závislosti, které přesně popisují množinu dat
- Rychlý algoritmus (umožňuje paralelní zpracování informace)
- Odolný proti šumům vstupních signálů (model určí jejích škodlivost pro řešení klasifikace a automaticky zahodí tyto nepřínosné signály)

Nevýhody:

- Nelze zajistit opakovatelnost a jednoznačnost výsledků
- Není jasné, co se děje v rámci sítě (nelze prozkoumat kroky, jak byly vypočteny výstupní hodnoty, což komplikuje proces interpretace výsledků a modifikace sítě pro zlepšení přesnosti klasifikace)
- Mnoha krokové nastavení vnitřních prvků a vazeb mezi nimi
- Náročnost při vytvoření architektury *ANN* (pro různé příklady jsou různé architektury)

2.2.3 The Gaussian Mixture Model (GMM)

The Gaussian Mixture Model (*model směsí Gaussových křivek*) je obvykle stochastickým modelem, který je užitečný při vytváření systémů rozpoznávání a obvykle je založený na učení bez učitele [43]. V oblasti zpracování zvukových signálů je tento klasifikační model dle literatury [45] velmi efektivní pro rozpoznávání řeči.

Předpokládá se, že pro každou třídu existuje funkce hustoty pravděpodobnosti, která je vyjádřená ve tvaru tzv. směsí komponentů. Směsí komponentů se myslí skupina jednotlivých křivek hustoty normálního rozdělení pravděpodobnosti (Gaussových křivek). [44]

Směs Gaussových křivek je váženou sumou *K* komponentů a může být vyjádřená vztahem [43]:

$$p(\overrightarrow{x}|\lambda) = \sum_{i=1}^{K} w_i \cdot b_i(\vec{x}), \tag{2.10}$$

kde $\vec{x}$ je D – rozměrný vektor náhodných hodnot, $b_i(\vec{x})$ je funkce hustoty rozdělení pravděpodobnosti a w_i je váha komponentu.

Pro váhy komponentů musí být dodržená níže uvedená podmínka:

$$\sum_{i=1}^{K} w_i = 1 \tag{2.11}$$

GMM je určený tzv. kovarianční maticí (Σ_k), vektorem středních hodnot ($\overrightarrow{\mu_k}$) a váhou komponentů (w_i). Tyto parametry se dohromady značí [2]:

$$\lambda = \{w_i, \overrightarrow{\mu_k}, \Sigma_k\} \tag{2.12}$$

Cílem algoritmu je hledání optimálních parametrů modelu λ, aby pravděpodobnost $p(\overrightarrow{x}|\lambda)$ monotónně stoupala, dokud nedosáhne určité prahové hodnoty. Tento proces odhadu parametrů se nazývá učení. Pro postupné hledání optimálních parametrů se používá iterativní algoritmus *Expectation Maximization* (*EM*). Algoritmus spočívá v postupném hodnocení modelu (tj. výpočet parametrů modelu) tak, aby byla splněná nerovnost $p(\overrightarrow{x}|\lambda_i) \geq p(\overrightarrow{x}|\lambda_i)$, jinými slovy se děje proces přehodnocení parametrů. Proces přehodnocení probíhá, dokud hodnota $p(\overrightarrow{x}|\lambda)$ nedosáhne nějaké prahové hodnoty. [25]

Výhody:

- Získává se hustota pravděpodobnosti pro každý klastr (vizte kap. 2)
- Rychlý algoritmus učení modelu
- Při klasifikaci není nutné brát v potaz neznámé třídy sousedních bloků, což urychluje proces výpočtu [46]
- Model má jenom 3 parametry

Nevýhody:

- Může nastat problém spojený s parametry modelu kovarianční matice (Σ_k), kdy se bude hledat řešení s nekonečnou hodnotou pravděpodobností $p(\overrightarrow{x}|\lambda)$ [45]
- Algoritmus *GMM* může přestat fungovat při velkých dimenzích signálů (např. větší, než 6) [47]

2.2.4 The Hidden Markov Model (HMM)

The Hidden Markov Model (*Skrytý Markovův Model*) je statistickou metodou, založenou na matematických modelech simulujících proces přechodu mezi jednotlivými stavy. Tento model je stejně jako *GMM* (vizte kap. 2.2.3) široce používaný v oblasti analýzy procesů měnících se v čase (např. rozpoznávaní řeči) [41].

Skrytý Markovův Model je určený následujícími parametry [2]:

- $S = \{s_1, s_2, \dots, s_N\}$ – stavový prostor,
- $V = \{v_1, v_2, \dots, v_M\}$ – abeceda posloupnosti možných pozorování modelu,

- $\pi = \{\pi_i\}$ – rozdělení pravděpodobnosti počátečního stavu,
- $A = \{a_{ij}\}$ – matice pravděpodobnosti přechodů,
- $B = \{b_j(k)\}$ – rozdělení pravděpodobnosti výstupů.

Množina parametrů $\lambda = \{\pi, A, B\}$ generuje posloupnost pozorování událostí $O = O_1 O_2 \dots O_T$, kde O_t je jedním ze symbolu abecedy posloupnosti možných pozorování *V*. Termín pozorování je statistickým termínem, kterým se označuje pozorovaný výsledek při známých nebo neznámých vstupních datech. Model představuje tzv. *Markovův řetězec*, pro který nejsou známé parametry, a jeho úkolem je hledání těchto neznámých parametrů na základě posloupnosti pozorování modelu, které se pak používají při další analýze (např. při rozpoznávání zvukových signálů). [2]

Markovův řetězec je určený posloupností náhodných procesů, kde matice pravděpodobnosti přechodů budoucích stavů *S* závisí na současných stavech, bez ohledu na stavy předchozí (minulé). Matice přechodů A charakterizuje přechody ze stavu i do stavu j s pravděpodobnosti a_{ij}. [48]

Existují tři algoritmy, které využívají *HMM* ve svůj prospěch v různých aplikacích [49]. První algoritmus spočívá v nalezení pravděpodobnosti $p(O|\lambda)$ příslušící modelu λ ke známé posloupnosti pozorování *O*. Další algoritmus je určený pro zjištění procesu, probíhajícího ve skryté části modelu. Je založený na hledání posloupnosti, která je nejvíce shodná s posloupností pozorování *O*. Principem posledního algoritmu je optimalizace modelu (optimalizace parametrů) tak, aby pravděpodobnost $p(O|\lambda)$ byla maximální. [48]

Výhody:

- Není nutné pochopení vnitřního procesu dynamických změn v systému (existuje konkrétní výstup pro další využití) [18]
- Výsledky analýzy *HMM* je jednoduché znázornit graficky [18]
- Umožňuje simulace dlouhých časových závislostí [18]
- Model dokáže pracovat s algoritmy s polynomiální složitostí [49]

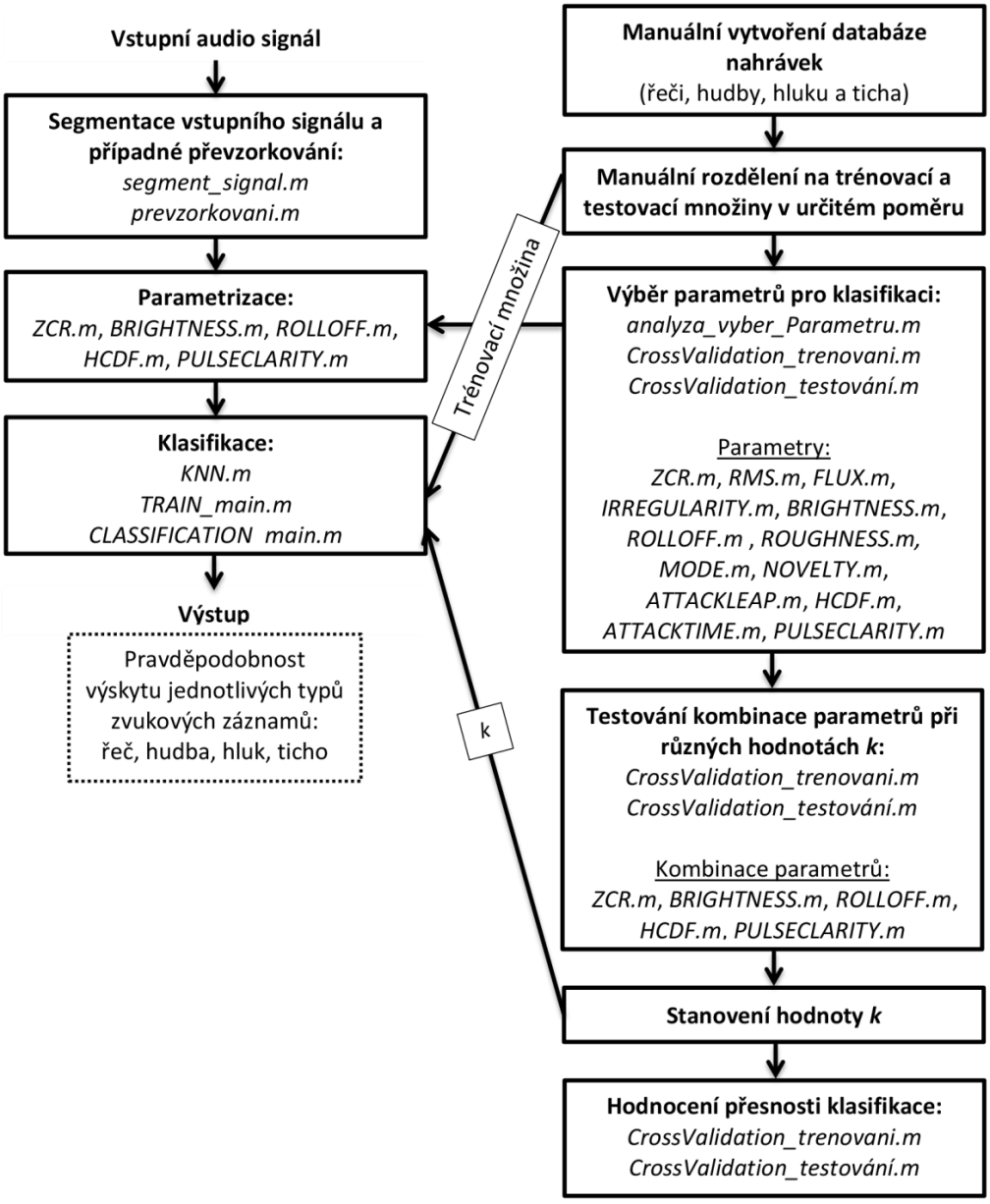
Nevýhody:

- Je nutné vytvořit komplexní *HMM* při práci s akustickou informací (akustická informace vyžaduje velký objem paměti pro uchování parametrů modelu a velký počet trénovacích dat) [49]

3 Realizace identifikace zvukových záznamů

Obsahem této kapitoly je návrh a popis experimentální implementace algoritmů jednoduchých metod identifikace (rozpoznávání) obsahu zvukových záznamů v programovém prostředí MATLAB. Realizace algoritmů využívá kombinace charakteristik zvukového signálu a klasifikační metodu *KNN*.

V úvodu bylo poznamenáno, že existuje velké množství algoritmů sloužících k rozpoznávání zvukových záznamů. Na jejich základě jsem implementovala algoritmus vlastní, jehož základní blokové schéma je znázorněno na *obr. 3.1*.



Obr. 3.1 Blokové schéma algoritmu identifikace obsahu audio záznamu

3.1 Implementace metody identifikace v MATLAB

Před implementací metody identifikace jsem připravila databázi zvukových signálů o celkové délce přibližně 20 min, která obsahuje řeč, hudbu, hluk a ticho. Databázi jsem vytvořila ze vzorků zvukové stopy archivních filmů dodaných vedoucím práce.

Níže je uvedená *tab. 3.1* s názvy použitých filmů a jejich charakteristikami.

Název filmu	**Rok výroby**	**Typ primárního / distrib. zvukového záznamu**
Starci na chmelu	1964	optický plochový
Božská Ema	1979	optický plochový
Adéla ještě nevečeřela	1977	optický plochový

Tab. 3.1 Původ využívaných archivních nahrávek

Vytvořenou databázi jsem rozdělila na trénovací a testovací množiny, kde trénovací množina obsahuje dvě třetiny a testovací jednu třetinu všech dat. Tento poměr jsem určila na základě literatury [29] (vizte kap. 2). Testovací množina je nutná pro metodu *Cross-Validation.*

Pro vytvoření databáze jsem použila program Adobe Audition CS6, ve kterém jsem od sebe oddělovala jednotlivé typy zvuků (řeč, hudba, hluk, ticho). Za hluk jsem považovala zvukový signál, který neobsahoval řeč, hudbu ani ticho. Hudba může být čistě instrumentální, nebo i s vokálním partem. Nahrávky jsem pojmenovala podle obsahu. První písmeno názvu souboru v databázi definuje typ zvukového signálu (h-hudba, r-řeč, s-hluk, t-ticho).

Podrobný popis navrhnutých skriptů či funkcí není obsahem následujících kapitol, ten je uveden v nápovědách jednotlivých funkcí (vizte přílohu A).

3.1.1 Segmentace

Navrhla jsem funkci ***segment_signal.m***, která umožňuje automatickou segmentaci signálu na segmenty požadované délky. Referenční délku úseku jsem zvolila 2,4 s, s touto volbou jsem se opírala o poznatky z literatury [6] (vizte kap. 2). Délka segmentu by měla být taková, aby se i poslechem dalo rozlišit, o jaký typ zvukového záznamu se jedná.

Funkce zároveň dovoluje nastavovat překryv těchto segmentů. Překryv je zaveden pro zvýšení přesnosti klasifikace. Používala jsem délku překryvu 1,5 s.

Ze segmentů konstantní délky, na které je signál rozdělen, lze dále určovat parametry. MATLAB funkce těchto parametrů jsou popsané v následující kapitole.

3.1.2 Parametrizace

Zprvu bylo nutné vybrat parametry, které budou využity pro klasifikaci audio záznamu. Vybírala jsem z velkého množství parametrů, které nabízí MIRtoolbox 1.6.1 (vizte kap. 2.1) [4].

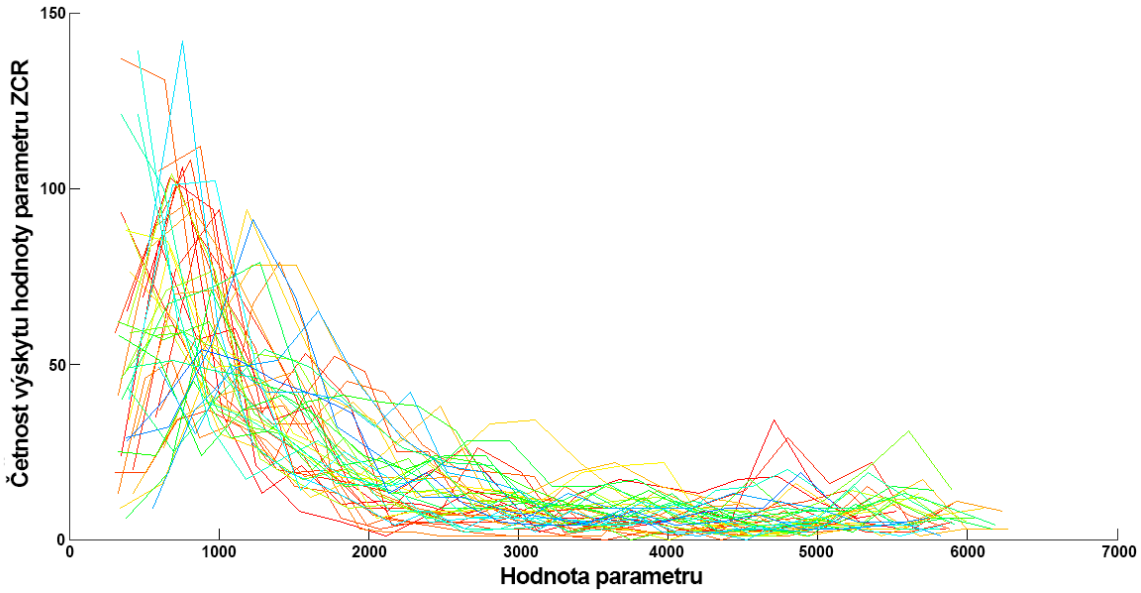
Výběr jednotlivých parametrů jsem uskutečnila na základě jejich postupné analýzy pomocí skriptu ***analyza_vyber_Parametru.m***.

Vstupem pro skript ***analyza_vyber_Parametru.m*** je zvukový signál libovolné délky, který je automaticky rozdělen na segmenty požadované délky, pomocí funkce ***segment_signal.m*** (vizte kap. 3.1.1). Pokud mají nějaké signály různé vzorkovací kmitočty, dojde navíc k jejich převzorkování. Převzorkováním se myslí změna hodnoty vzorkovacího kmitočtu signálu na stejnou hodnotu pro všechny signály a provádí se pomocí funkce ***prevzorkovani.m***.

Dále se v každém obdrženém úseku vypočítají hodnoty parametru a na výstupu vznikne histogram pro každý segment. Histogram ukazuje četnost výskytu nějaké hodnoty příslušného parametru v jednotlivých segmentech signálu.

Analýza každého parametru byla provedena pro všechny jednotlivé typy zvuků (řeč, hudba, hluk, ticho). Níže na *obr. 3.2* je uveden příklad několika histogramů vzešlých z analýzy parametru *ZCR*, kde vstupem byl řečový signál.

Parametry jsem vybírala na základě podobnosti histogramů pro různé vzorky stejné třídy. Pro realizaci vybraných parametrů (vizte *tab. 3.2*) v MATLAB jsem využila implementovaných funkcí převzatých z MIRtoolboxu [4].



Obr. 3.2 Výsledný histogram (analýza parametru ZCR, vstupem je řečový signál)

Zprvu se vstupní signál (vstupní segment signálu) rozdělí na další, ještě kratší pod-segmenty a váhuje se Hammingovým oknem (vizte kap. 2). Zvolila jsem délku pod-segmentu 25 ms a délku okna 1/4 délky pod-segmentu (vizte kap. 2). Nakonec se počítá příslušný parametr.

Název parametru (vizte kap. 2.1)	**Název funkce v MATLAB**
Zero-Crossing Rate (ZCR)	*ZCR.m*
Root-Mean-Square Energy (RMS energy)	*RMS.m*
Brightness	*BRIGHTNESS.m*
Spectral Roll-of	*ROLLOFF.m*
Spectrum Flux (SF)	*FLUX.m*
Sensory Dissonance (Roughness)	*ROUGHNESS.m*
Spectral Peaks Variability (Irregularity)	*IRREGULARITY.m*
Mode	*MODE.m*
Novelty Curve	*NOVELTY.m*
Pulse Clarity	*PULSECLARITY.m*
Harmonic Change Detection Function (HCDF)	*HCDF.m*
Attack Time	*ATTACKTIME.m*
Attack Leap	*ATTACKLEAP.m*

Tab. 3.2 Seznam vybraných parametrů a implementovaných funkcí

3.1.3 Metoda Cross-Validation

Metodu *Cross-Validation* (vizte kap. 2) využívám v případě implementovaného algoritmu pro hodnocení přesnosti klasifikace a ke stanovení optimální hodnoty *k* klasifikátoru *KNN* (počet nejbližších sousedů). Tuto metodu jsem rovněž používala k dalšímu výběru parametrů audio signálu pro klasifikaci.

Pro metodu *Cross-Validation* jsem vytvořila skripty ***CrossValidation_trenovani.m*** a ***CrossValidation_testovani.m***. Blokové schéma implementovaných algoritmů založených na metodě *Cross-Validation* ilustruje *obr. 3.3*.

Při výběru parametru vhodného pro účely klasifikace se implementovaný algoritmus odehrává ve dvou krocích. Cílem je najít takové parametry, s pomocí kterých bude možné dostatečně úspěšně klasifikovat různé typy zvukových signálů.

První krok (na obrázku *obr. 3.3* označený fialovou barvou):

- Ve skriptu ***CrossValidation_trenovani.m*** dochází k úpravě vstupní trénovací množiny. Signály se segmentují (vizte 3.1.1) a případně se převzorkují (vizte 3.1.2). Zároveň se z každého segmentu získávají hodnoty testovaného parametru.
- Získané hodnoty testovaného parametru se ukládají do souboru *train_CV.mat.*
- Uložená data v souboru *train_CV.mat* se předávají na vstup skriptu ***CrossValidation_testovani.m.***
- Druhým vstupem pro tento skript jsou data z testovací množiny, které jsou předem rozdělené do tříd (hudba, řeč, hluk, ticho)
- Tyto data se ve skriptu ***CrossValidation_testovani.m*** zpracovávají stejným způsobem jako ve skriptu pro trénování (segmentace, převzorkování, získávání hodnot testovaného parametru)
- Získané hodnoty se porovnávají s hodnotami, ze souboru *train_CV.mat* pomocí klasifikátoru *KNN* s hodnotami $k = 9; 7; 5$ (implementace klasifikátoru je popsaná v kapitole 3.1.5).
- Výsledkem klasifikace jsou hodnoty pravděpodobnosti výskytu jednotlivých typů zvuku, které se dále ukládají do Excel souboru.
- Tento celý proces se opakuje pro všechny parametry z *tab. 3.2.*

Druhý krok (na *obr. 3.3* označený zelenou barvou):

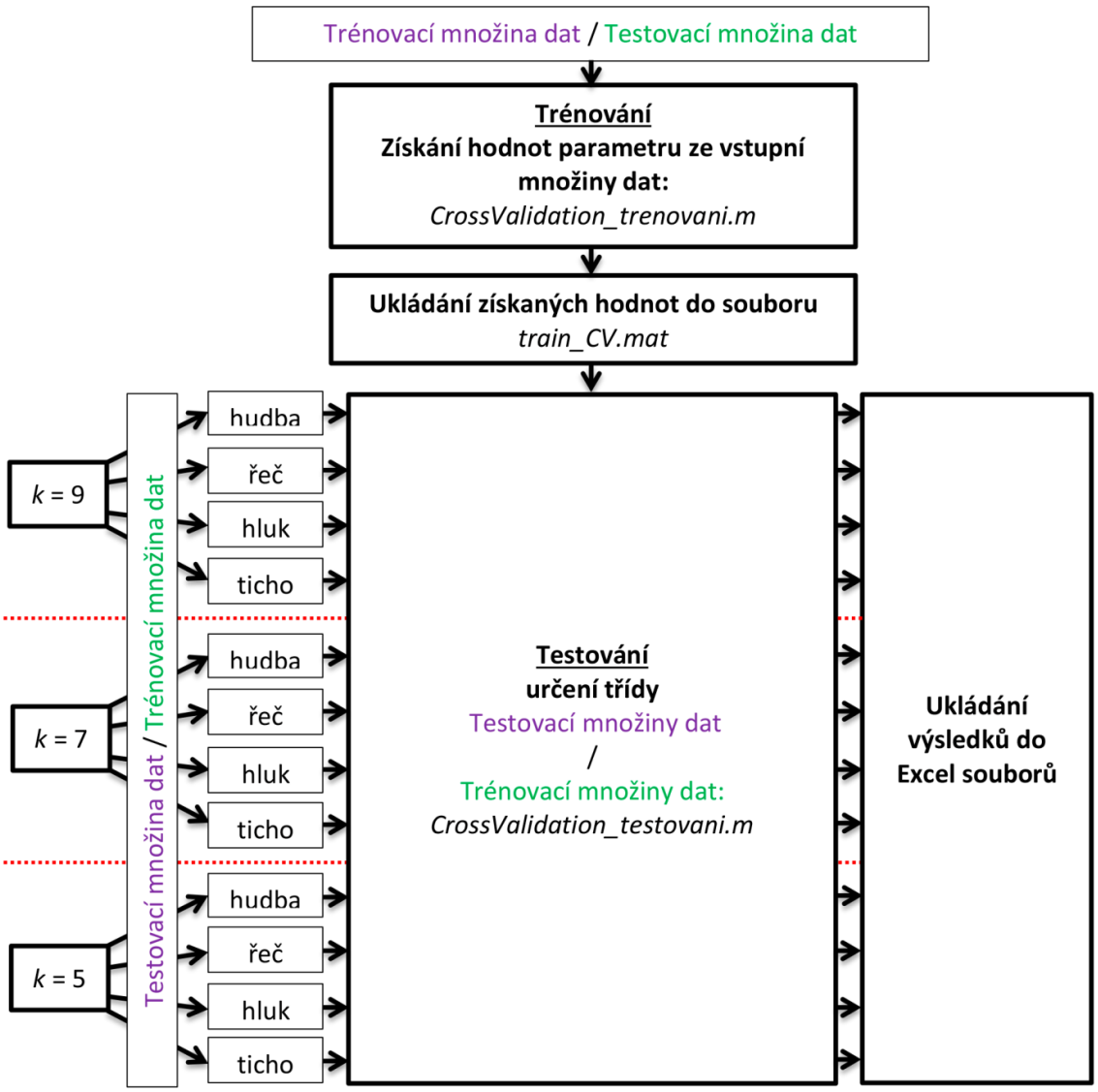
- Trénovací množina se prohodí s testovací a celý proces popsaný v prvním kroku se opakuje.

Výsledky testování parametrů je možné najít v Excel souboru *Testování parametrů.xls* v příloze A. Ze získaných hodnot jsem vybrala ty parametry, pro které platila úspěšnost identifikace hudby i řeči vyšší než 58 %.

Při testování nahrávek hluku a ticha jsem nedokázala s používanými parametry docílit použitelných výsledků. Pro většinu parametrů se úspešnost pohybovala pod 50 %. To může být způsobeno nevhodným obsahem trénovací a testovací množiny. Možným důvodem nespolehlivosti je malý počet nahrávek těchto typů zvuků v množinách dat. Pro záznamy hluku je důležitý velký počet nahrávek v trénovací množině, protože existují různé druhy hluků, a jejich charakteristiky mohou být dost rozdílné. Množiny obsahující ticho se mohou na první pohled zdát jako nahrávky bez zvuku, ale často tomu tak v případě archivních záznamů obsahujících artefakty není (vizte kap. 1).

Z výše uvedeného experimentování jsem se rozhodla pro vyžití následujících parametrů: *Zero-Crossing Rate (ZCR)*, *Harmonic Change Detection Function (HCDF)*, *Pulse Clarity*, *Spectral Roll-of* a *Brightness.* Z vybraných těchto parametrů jsem vytvořila jejich kombinaci, pomocí které probíhala klasifikace obsahu audio záznamů.

Výše popsaný princip implementovaného algoritmu lze analogicky použít při zjišťování optimální hodnoty *k* a při hodnocení přesnosti klasifikátoru, ale místo jednotlivých parametrů se testuje vytvořená kombinace parametrů.



Obr. 3.3 Blokové schéma implementovaného algoritmu metody Cross-Validation

3.1.4 Klasifikace

V kapitole 2.2 byla zmíněna existence různých metod pro rozpoznávání audio záznamu. Pro klasifikaci jsem vybrala metodu *k-Nearest Neighbors* (*KNN*).

Metody *ANN*, *HMM* a *GMM*, které jsou uvedené v kapitole 2.2, mohou být uplatněné pro řešení složitějších příkladů, kde se využívá hodně parametrů a klasifikačních tříd (např. rozpoznávání řeči, klasifikace hudby dle žánrů). Použití těchto metod pro řešení úkolu v této práci zřejmě není kvůli jejich komplexnosti optimální. Proto jsem využila klasifikátor *KNN*, který je jednoduchý na implementaci a v daném případě může být dostatečně přesný (vizte kap. 2.2.1).

Pro vybranou metodu klasifikace jsem implementovala funkci ***KNN.m***. Princip funkce spočívá v porovnání trénovací a testovací množiny. Počítá se Euklidova vzdálenost

(vizte kap. 2.2.1) mezi prvky z těchto množin a hledá se *k* minimálních vzdáleností a jím odpovídající názvy tříd (labels). Základní princip klasifikátoru *KNN* je znázorněný na *obr. 2.3.*

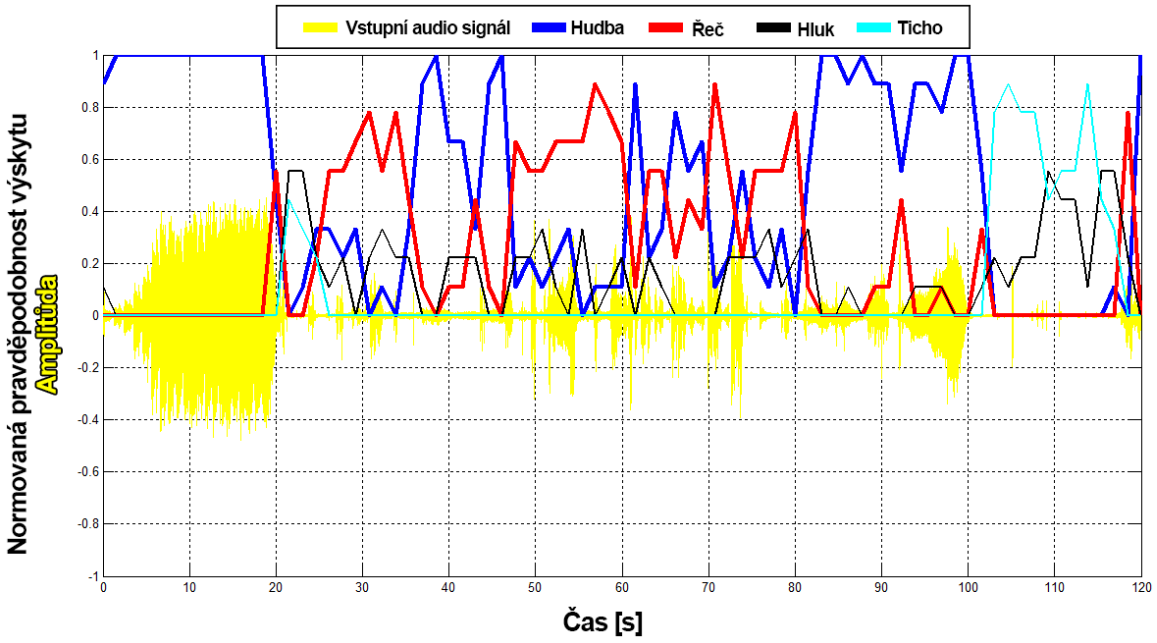
Zvolila jsem *k* = 9 na základě metody *Cross-Validation* (vizte *tab. 3.3*). Hlavním kritériem byla přesnost testování. V *tab. 3.3* jsou uvedeny střední hodnoty přesnosti testování všech zkoumaných typů zvuku a odpovídající střední kvadratická odchylka *σ*. Jako pomocné kritérium jsem použila chybu metody. Chybou se myslí průměrný rozdíl hodnoty při testování jednotlivých typů zvuku testovací množiny a hodnoty při testování jednotlivých typů zvuku trénovací množiny vydělený dvěma. Tyto hodnoty jsou uložené v Excel souboru *Hodnocení spolehlivosti klasifikátoru.xls* v příloze A.

Počet nejbližších sousedů	**Přesnost testování testovací množiny [%]**	***σ* [%]**	**Přesnost testování trénovací množiny [%]**	***σ* [%]**	**Chyba metody *Cross-Validation* [%]**	***σ* [%]**
k = 9	71	30	63	25	6	4
k = 7	70	34	61	27	8	6
k = 5	65	33	59	26	5	3

Tab. 3.3 Hodnocení přesnosti klasifikace při různých hodnotách k

Pro rozpoznání obsahu zvukového záznamu jsem vytvořila další skripty: ***TRAIN_main.m*** a ***CLASSIFICATION_main.m***. Algoritmy těchto skriptů jsou podobné již popsaným skriptům ***CrossValidation_trenovani.m*** a ***CrossValidation_testovani.m***.

Skript ***TRAIN_main.m*** je určený pro natrénování klasifikátoru daty, která jsou určená kombinací vybraných parametrů zvukových signálů (vizte 3.1.3). Po natrénování může už následovat klasifikace vstupní audio nahrávky, která probíhá ve skriptu ***CLASSIFICATION_main.m***. Výstupem tohoto skriptu je pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase. Příklad výstupu pro dvouminutovou archivní nahrávku z filmu „Lev s bílou hřívou“ z roku 1986 je znázorněn na *obr. 3.4.*



Obr. 3.4 Příklad pravděpodobnosti výskytu jednotlivých typů zvukových záznamů v závislosti na čase

Pro hodnocení přesnosti rozpoznání obsahu zvukového záznamu jsem využila metodu *Cross-Validation* (vizte kap. 2), pomocí funkcí ***CrossValidation_trenovani.m*** a ***CrossValidation_testovani.m***, popsaných v kapitole 3.1.3. Výsledky hodnocení jsou k nalezení v Excel souboru *Hodnocení spolehlivosti klasifikátoru.xls* v příloze A. V *tab. 3.4* jsou uvedeny hodnoty, charakterizující přesnost klasifikace jednotlivých typů zvuku, při využití hodnoty $k = 9$.

	Hudba	**Řeč**	**Hluk**	**Ticho**
k = 9	84 %	71 %	22 %	91 %

Tab. 3.4 Hodnocení přesnosti implementovaného klasifikátoru KNN, kde k = 9

V *tab. 3.4* je možné pozorovat neschopnost implementovaného klasifikátoru úspěšně rozpoznávat zvukové záznamy, které obsahují hluk. Podobný trend jsem už zaznamenala při testování jednotlivých parametrů (vizte kap. 3.1.3). Jak již bylo uvedeno, příčinami mohou být velké rozdíly mezi charakteristikami nahrávek hluků a s tím spojená nedostatečná velikost trénovací množiny.

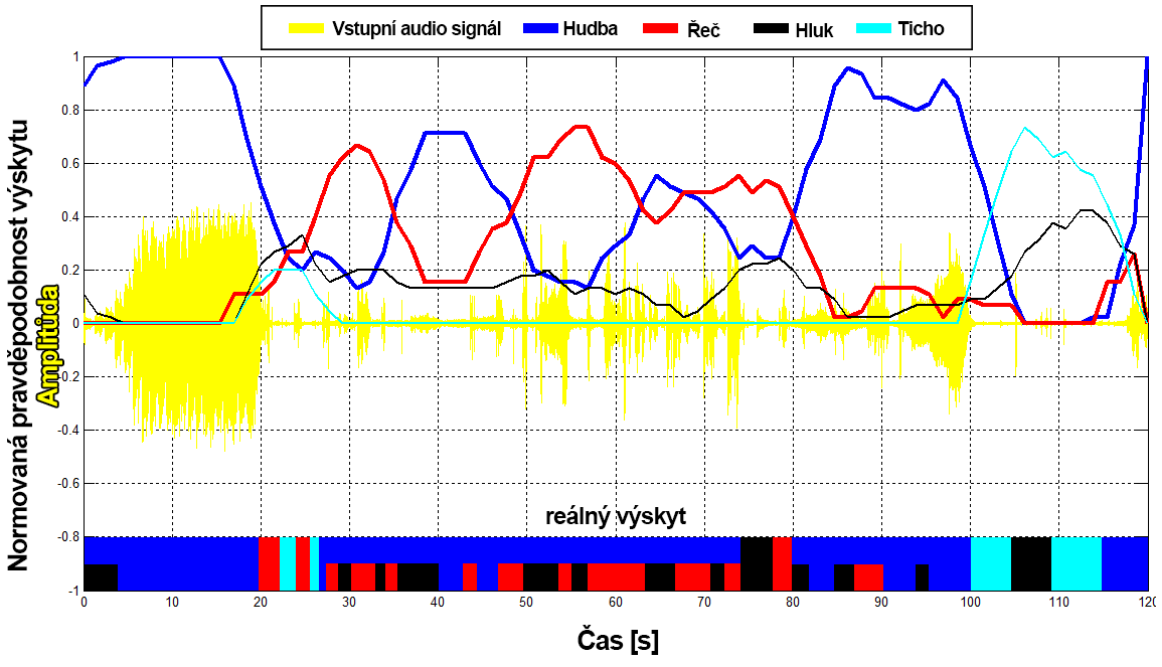
Úspěšnost identifikace hudby, řeči a ticha je vyšší 70 % a klasifikaci tak mohu posoudit jako relativně úspěšnou.

3.2 Výsledky identifikace

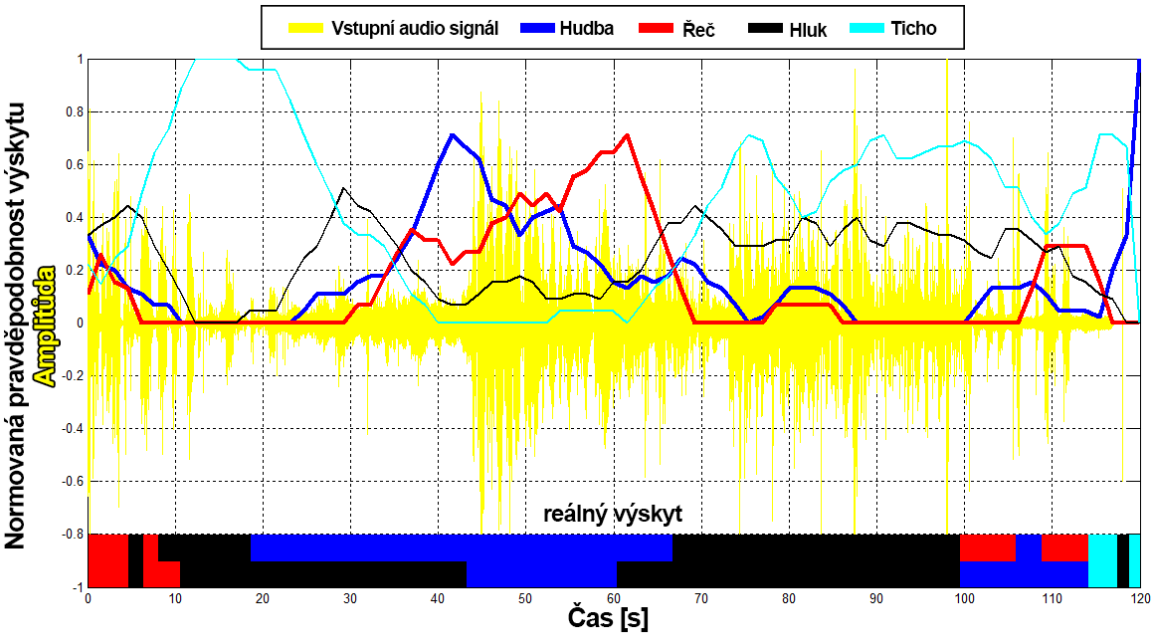
V této kapitole uvádím výsledky realizace identifikace zvukových záznamů, kde vstupem implementovaného algoritmu byly nahrávky v délce okolo dvou minut vyjmuté z archivních filmů. Testovanými nahrávkami byly výřezy z filmů: „Lev s bílou hřívou“ z roku 1986 (*obr. 3.5*), „Slavnosti sněženek“ z roku 1983 (*obr. 3.6*) a „Valčík pro milión“ z roku 1960 (*obr. 3.7*), „Jedenácté přikázání“ z roku 1935 (*obr. 3.8* a *obr. 3.9*). Na obrázcích pro příslušné filmy jsou znázorněny křivky udávající pravděpodobnost výskytu jednotlivých typů zvuku v závislosti na čase spolu s obálkou vstupního signálu a naznačením opravdového výskytu typů zvuku.

Pro lepší znázornění křivek bylo provedeno jejich vyhlazení, které potlačilo úzké špičky. Grafické výstupy bez vyhlazení je možné najít v příloze B.

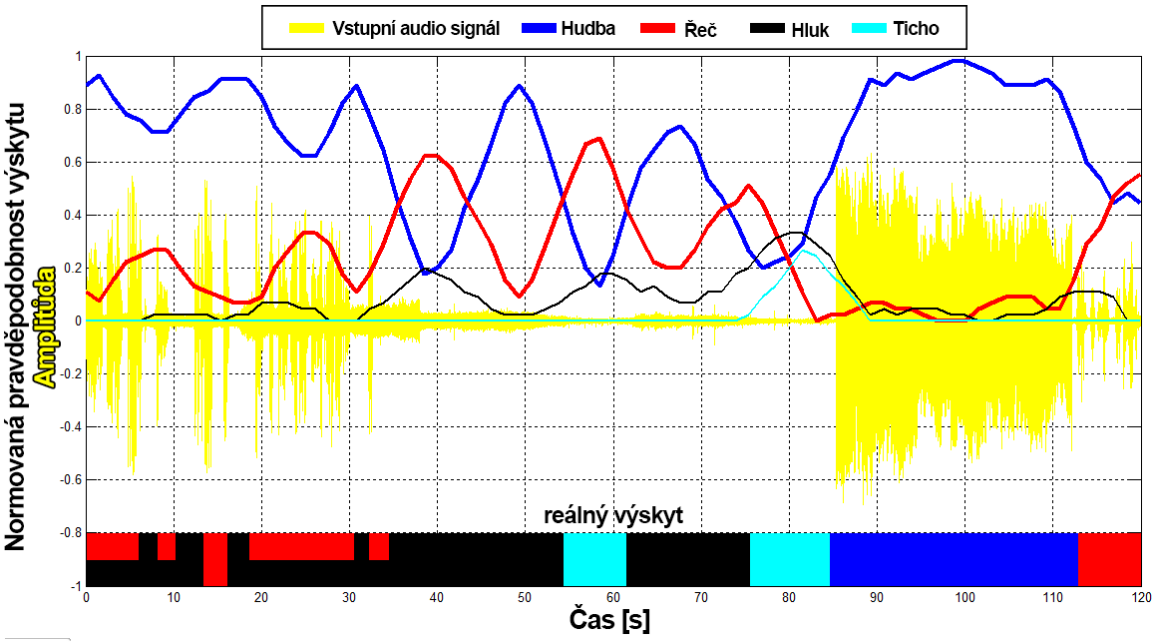
Specifikem nahrávek z filmů komplikujících identifikaci jsou zvuky na pozadí. Jak je v grafech uvedeno, téměř po celou dobu nahrávky se v jednotlivých časových úsecích vyskytuje několik typů zvuku a z výsledných křivek není jasné, zda jde o chybu algoritmu, nebo jde o přítomnost několika typů zvuku.



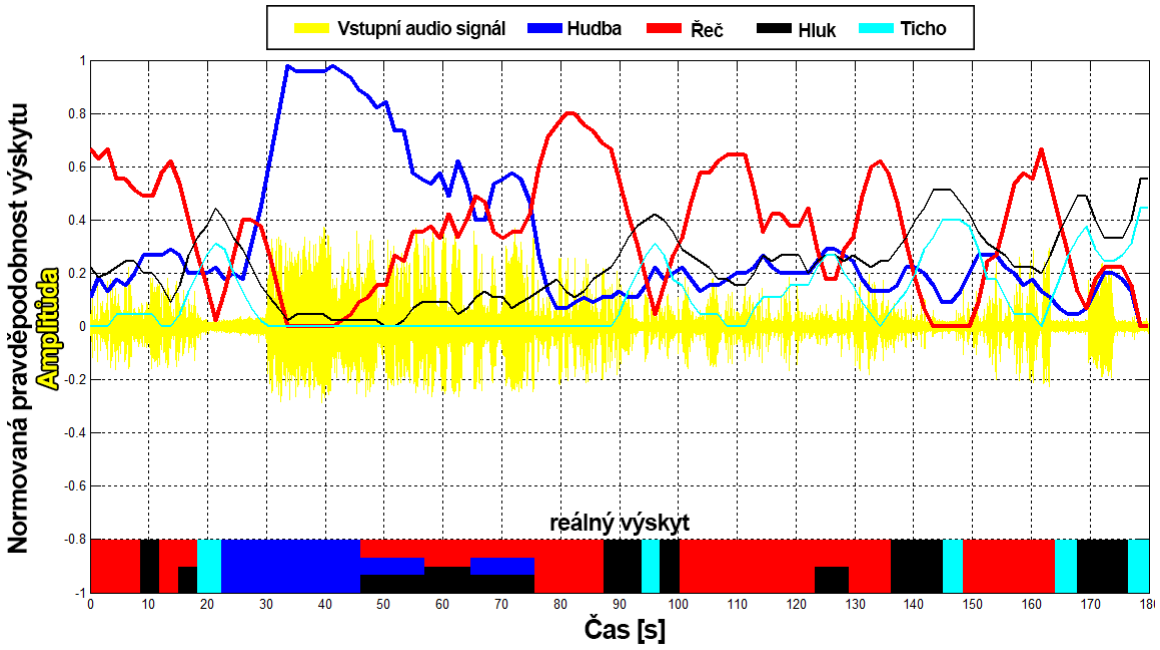
Obr. 3.5 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Lev s bílou hřívou“



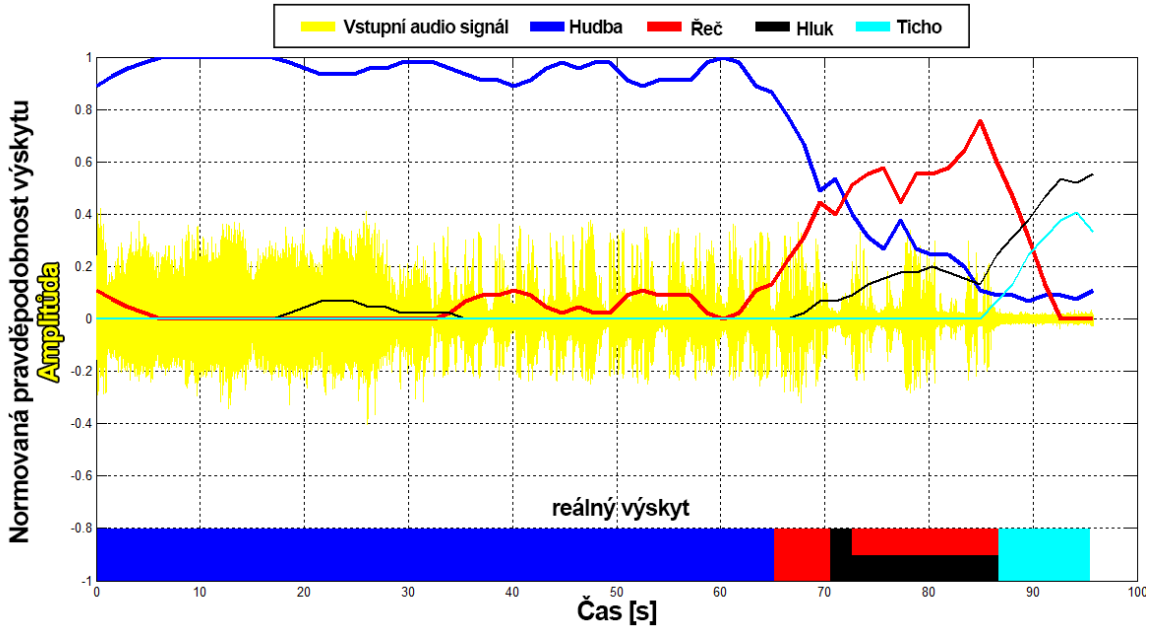
Obr. 3.6 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Slavnosti sněženek“



Obr. 3.7 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Valčík pro milión“



Obr. 3.8 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Jedenácté přikázání“ (výřez 1)

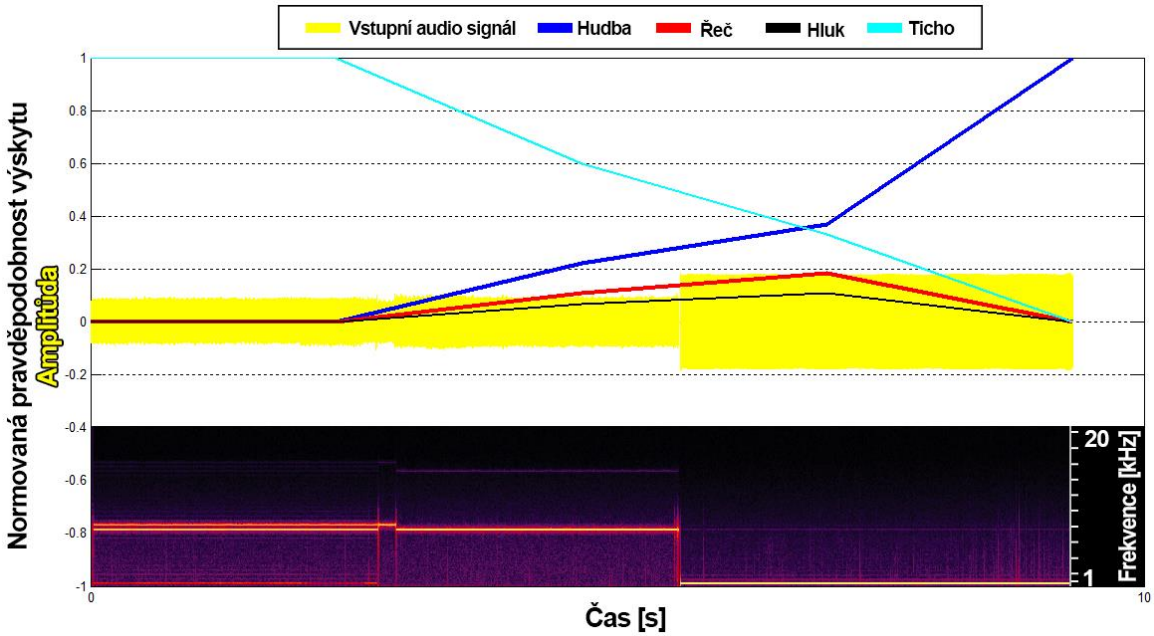


Obr. 3.9 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Jedenácté přikázání“ (výřez 2)

Z grafů je možné pozorovat, že se úspěšnost klasifikace pro jednotlivé nahrávky značně liší. Ve třech případech (*obr. 3.5*, *obr. 3.8* a *obr. 3.9*) proběhlo rozpoznání relativně přesně. V ostatních dvou případech výsledky byly horší. Například v druhém případě (*obr. 3.6)* nastal problém s klasifikací hluku, který byl většinu času považován za ticho. V případě

třetím (*obr. 3.7*) byla řeč s hlukem na začátku nahrávky klasifikována s velkou jistotou jako hudba.

Zkusila jsem rovněž otestovat kalibrační úsek s obsahem jednoduchých tonů (*obr. 3.10*), který byl klasifikován nejvíce jako ticho nebo hudba.



Obr. 3.10 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro kalibrační úsek

Důvodů výsledných nepřesností klasifikace záznamů může být mnoho. Prvním z nich je například fakt, že trénovací množina dat obsahuje relativně malé množství nahrávek. Archivní záznam může mít velký počet různých artefaktů, měnících parametry signálu. Změněné parametry signálu pak mohou představovat pro klasifikátor data, na které nebyl naučen.

Dalším možným nedostatkem je nevhodně zvolená délka segmentů signálu před klasifikací. Při výběru délky segmentů jsem zkoušela různé hodnoty. Ve finálním algoritmu jsem zvolila délku segmentu 2,4 s, protože jsem pro ni získávala nejlepší výsledky.

Pro klasifikaci je také nesmírně důležitý výběr parametrů, protože jsou to data, na základě kterých se signály klasifikují. Použití kombinace vybraných parametrů může rovněž ovlivnit výsledek. Nakonec neopomenutelným aspektem je volba klasifikátoru a jeho konfigurace.

Závěr

Cílem této bakalářské práce bylo seznámení s problematikou identifikace archivních zvukových záznamů a implementace vybrané metody v programovém prostředí MATLAB. Implementovala jsem algoritmus, jehož princip spočívá v extrakci vybraných parametrů signálu a v jejich následném srovnávání s daty, na které byl systém natrénován. Jednotlivé parametry jsem vybírala na základě experimentálního pozorování vlastností zkoumaných tříd zvuku (řeč, hudba, hluk, ticho). Samotné srovnávání jsem realizovala klasifikátorem *k-Nearest Neighbors* (*KNN*) s hodnotou $k = 9$.

Na tento klasifikátor jsem se zaměřila, protože je jednoduchý na implementaci a efektivní v případě klasifikace do malého počtu tříd. Původně jsem zvažovala použití metody *Artificial Neural Network* (*ANN*), ale ta je komplexnější a uplatnění nachází spíše při klasifikaci signálů do velkého počtu tříd (např.: žánry hudby, typ hlasu řečníka), což není případ této práce.

Výslednou spolehlivost implementovaného algoritmu identifikace zvukových záznamů jsem otestovala pomocí metody *Cross-Validation.* Pro vytvoření dat trénovací a testovací množiny jsem využívala pouze archivních nahrávek, které mi dodal vedoucí práce.

Výsledky testování zmíněnou metodou, s přibližně 20 min archivních nahrávek, dopadly relativně úspěšně pro klasifikování řeči, hudby a ticha. Úspešnost algoritmu se pro tyto tři typy zvuků pohybovala okolo 70-90 %. Rozpoznání hluku nebylo s použitou metodou klasifikace úspěšné. To lze vysvětlit tím, že užívané nahrávky hluků mohou mít velmi odlišné charakteristiky. Pro řešení tohoto problému se naskytuje několik možností. Jednou z nich je například vytvoření větší trénovací množiny hluků. Dále je také možné nevytvářet trénovací množinu hluků vůbec a považovat za hluk všechny části signálu, u kterých je například vzdálenost nejbližších sousedů větší než určitá mez.

Výstupem této práce je algoritmus určující časové rozložení pravděpodobnosti jednotlivých typů zvukového záznamu. Realizovaný algoritmus jsem vyzkoušela na různých archivních nahrávkách. Úspěšnost klasifikace se pro různé nahrávky bohužel někdy i výrazně liší. V některých případech určí algoritmus časové rozložení typů záznamu úspěšně, jindy podá výsledky zavádějící.

Úspešnost klasifikace je spojená s nedostatky algoritmu, které jsou zmíněné v kapitole 3.2 a lze je zohlednit do budoucnosti, při dalším řešení této problematiky. V prvním kroku

by se bylo zřejmě nejlepší zaměřit na výběr kombinace parametrů, které jednoznačněji specifikují jednotlivé typy zvuků. Dalším krokem je vylepšení trénovací množiny, tedy zaměření se více na její obsah i objem. Konečně pokročilejším klasifikátorem by šlo dosáhnout lepších časů zpracování a výsledků klasifikace.

Vytvořený algoritmus v MATLAB je svou strukturou připravený na implementaci většiny zmíněných vylepšení. Do budoucna by se tato práce dala dále rozšířit například o přidání dalších tříd, do kterých lze klasifikovat.

Stanoveného cíle, tedy implementace klasifikátoru pro rozpoznávání jednotlivých typů zvuku v archivních nahrávkách bylo dosaženo a je možné jej využít pro účely uvedené v úvodu této práce.

Seznam použité literatury a zdrojů

[1] Hao J., Lie L., Zhang H., „A Robust Audio Classification and Segmentation Method“, *Microsoft Research*, 2001

[2] Tong Z., C.-C. Jay K., „Hierarchical System for Content-based Audio Classification and Retrieval”, *Proceedings, 1999 IEEE International Conference on*, vol. 6, pp. 3001-3004, 1999

[3] Wikipedia, „Zero-Crossing rate“, [cit. 17.04.2015], [online], dostupné z: http://en.wikipedia.org/wiki/Zero-crossing_rate

[4] Lartillot O., Toiviainen P., Eerola T., A Matlab Toolbox for Music Information Retrieval”, *Data Analysis, Machine Learning and Applications*, pp. 261-268, 2008

[5] Wikipedia, „Root mean square“, [cit. 17.04.2015], [online], dostupné z: http://en.wikipedia.org/wiki/Root_mean_square

[6] Scheirer E., Slaney M., „Construction and evaluation of a robust multifeature speech/music discriminator”, *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 2, pp. 1331-1334, 1997

[7] Harte C., Mark S, Gasser M., „Detecting harmonic change in musical audio”, *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, pp. 21-26, 2006

[8] McDermott JH, Lehr AJ, Oxenham AJ, „Is relative pitch specific to pitch?”, *Psychological Science*, pp. 1263-1271, 2008

[9] Lartillot O., Eerola T., Toiviainen T., Fornari J., „Multi-feature modeling of pulse clarity: Design, validation, and optimization”, *Proceedings of the 9th International Conference on Music Information Retrieval*, pp. 14-18, 2008

[10] Godsill, S. J., Rayner, P. J. W., „Digital Audio Restoration”, 1998

[11] Müller M., Prätzlich T., Driedger J., „A cross-version approach for stabilizing tempo-based novelty detection“, *Proceedings of the 13th International Society for Music Information Retrieval Conference*, 2012

[12] Olivier L., Donato C., Kim E., Didier G., „A simple, high-yield method for assessing structural novelty“, *Proceedings of the 3rd International Conference on Music & Emotion (ICME3)*, 2013

[13] Grosche, P., Muller, M., Kurth, F., „Cyclic tempogram-A mid-level tempo representation for musicsignals”, *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 5522-5525, 2010

[14] Foote, J., „Automatic audio segmentation using a measure of audio novelty”, *Multimedia and Expo, 2000 IEEE International Conference on*, vol. 1, pp. 452-455, 2000

[15] Rodet X., Jaillet F, „Detection and modeling of fast attack transients”, *Proceeding of the International Computer Music Conference*, 2001

[16] Bello, J.P., Daudet, L., Abdallah, S., Duxbury, C., „A Tutorial on Onset Detection in Music Signals“, *Speech and Audio Processing, IEEE Transactions on*, vol. 13, pp. 1035-1047

[17] Kos M., Kačič Z., Vlaj D., „Speech bandwidth classification using general acoustic features, modified spectral roll-off and artificial neural network“, *Mathematical models and methods in modern science Conference; 14th, Mathematical models and methods in modern science*, pp. 212-217, 2012

[18] Маковкин К.А., „Гибридные модели: скрытые марковскиемодели и нейронные сети, их применение всистемах распознавания речи“, *Сборник трудов, Вычислительный центр им. А.А.Дородницына, Российской академии наук,* с.40-54, 2006

[19] Wikipedia, „Mode (music)“, [cit. 19.04.2015], [online], dostupné z: http://en.wikipedia.org/wiki/Mode_(music)#Major_modes

[20] Wikipedia, „Beat (acoustic)“, [cit. 19.04.2015], [online], dostupné z: http://en.wikipedia.org/wiki/Beat_(acoustics)

[21] Ericsson L., „Automatic speech/music discrimination in audio files“, *Master's thesis, School of Media Technology Royal Institute of Technology, Sweden*, 2009

[22] Tzanetakis, G., Cook P., „Musical genre classification of audio signals“, *IEEE Transactions on Speech and Audio Processing*, pp. 293-302, 2002.

[23] Jensen K., „Timbre Models of Musical Sounds Kristoffer Jensen“, *DIKU, University of Copenhagen*, 1999

[24] Krimphoff J., Mcadams S., Winsberg S., „Caractérisation du timbre des sons complexes.II. Analyses acoustiques et quantification psychophysique*“, Journal de Physique IV, 04 (C5),* pp. 625-628, 1994.

[25] Juan J., Burred J.J., „An Objective Approach to Content-Based Audio Signal Classification“, *Diplomarbeit eingereicht*, 2003

[26] Larose D. T., „Discovering Knowledge in Data: An Introduction to Data Mining“, *John Wiley & Sons, Inc., ISBN 9780471666578*, 2005

[27] Devroye L., Györfi L., Lugosi G. „A probabilistic theory of pattern recognition“, *Springer-Verlag New York Berlin Heidelberg, ISBN 0-387-94618-7*, 1996

[28] Duda R. O, Hart P.E., Stork D. G., „Pattern Classification and Scene Analysis: Part I Pattern Classification”, *to be published in 1998 by John Wiley & Sons, Inc., ISBN 0-471-05669-3*, 1998

[29] Чубукова И.А., „Data Mining: учеб. Пособие Основы информационных технологий“, *Интернет-ун-т информ. технологий*, 2008

[30] Wikipedia, „Biological neuron model“, [cit. 29.04.2015], [online], dostupné z: http://en.wikipedia.org/wiki/Biological_neuron_model

[31] Rojas R., „Neural Networks - A Systematic Introduction“, *Springer-Verlag, Berlin, New-York, ISBN-10: 3540605053*, 1996

[32] Gershenson C., „Artificial Neural Networks for Beginners“,*arXiv preprint cs/0308031*, 2003

[33] Wikipedia, „Artificial neural network“, [cit. 29.04.2015], [online], dostupné z: http://en.wikipedia.org/wiki/Artificial_neural_network

[34] Syrový V., „Hudební akustika“, *ISBN 80-7331-901-2, AMU, Praha* 2003

[35] Hastie T., Tibshirani R., Friedman J., „The Elements of Statistical Learning“, *Springer-Verlag New York, Second Edition, eBook ISBN 978-0-387-84858-7*, 2009

[36] Alpaydın E., „Introduction to Machine Learning“, *The MIT Press Cambridge, Massachusetts London, England, Second Edition, ISBN 978-0-262-01243-0,* 2010

[37] Balabko P., „Speech and music discrimination based on signal modulation spectrum”, *Tech. Rep.*, 1999

[38] Матяско, А. А., Хаустов В.А., „Классификация документов в векторном пространстве. Сравнение методов Роккио и метода k-ближайших соседей”*, БГУИР, ISBN 978-985-488-926-9,* 2012, [cit. 05.04.2015], [online], dostupné z: http://libeldoc.bsuir.by/handle/123456789/2979

[39] Wikibooks, „Data Mining Algorithms In R/Classification/kNN“, [cit. 08.05.2015], [online], dostupné z: http://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Classification/kNN

[40] Корнеев Д.С., „Использование аппарата нейронных сетей для создания модели оценки и управления рисками предприятия“, *Управление большими системами: сборник трудов, №17,* с. 81-102, 2007

[41] Namrata Dave, „Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition”, *International Journal for Advance Rresearch in Engineering and Technology,* vol. 1, 2013

[42] El-Maleh K., Klein M., Petrucci G., Kabal P., „Speed/music discrimination for multimedia applications“, *Dept. Electrical & Computer Engineering McGill University, Montreal,* 2000

[43] Садыхов Р.Х., Ракуш В.В., „Модели гауссовых смесей для верификации диктора по произвольной речи“, *Доклады БГУИР № 4,* с. 95-103, 2003

[44] Ромацкий Д.Б, *„*Автоматическая классификация музыкальных произведений по жанрам*“, Томск. гос. ун-т. Факультет информатики,* 2014

[45] Yu D., Deng L., „Automatic Speech Recognition: A Deep Learning Approach“, *Springer-Verlag London, ISBN 1447157796,* pp. 13-20, 2014

[46] Permuter H. „A study of Gaussian mixture models of color and texture features for image classification and segmentation“, *Pattern Recognition,* vol. 39, pp. 695-706, 2006

[47] NickGillianWiki, „GMM Classifier“, [cit. 10.05.2015], [online], dostupné z: http://www.nickgillian.com/wiki/pmwiki.php/GRT/GMMClassifier

[48] Гультяева Т.Ф., Попов А.А., „Классификация последовательностей с использованием скрытых марковских моделей в условиях неточного задания их структуры“, *Вестник Томск. гос. ун-та, Управление, вычислительная техника и информатика, № 3(24),* с. 57-63, 2013

[49] Rabiner L., „A tutorial on hidden Markov models and selected applications in speech recognition“, *Proceedings of the IEEE,* vol. 77, pp. 257-286, 1989

Přílohy

Příloha A (CD)

Příloha B (dodatečné grafy, ukázka tabulky)

Příloha A

Přiložené CD obsahuje 3 adresáře:

- ***MATLAB***
- ***Výsledky testování***
- ***Výsledky identifikace***

Ve složce ***MATLAB*** jsou implementované skripty, funkce a uložená trénovací data.

Složka ***Výsledky testování*** obsahuje dva Excel soubory (*Testování parametrů.xls* a *Hodnocení spolehlivosti klasifikátoru.xls*). V souboru *Testování parametrů.xls* jsou uvedené tabulky výsledků metody *Cross-Validation* při testování parametrů (hodnoty pravděpodobnosti výskytu řeči, hudby, hluku a ticha při testování jednotlivých parametrů). V souboru *Hodnocení spolehlivosti klasifikátoru.xls* se nachází výsledky metody *Cross-Validation* při hodnocení přesnosti klasifikace.

Ve složce ***Výsledky identifikace*** jsou grafické výstupy implementovaného algoritmů.

Příloha B

V příloze B jsou grafické výstupy implementovaného algoritmu bez vyhlazení a ukázka tabulky ze souboru *Testování parametrů.xls* (vizte přílohu A).

Následující tabulky (vizte *tab. 1)* jsou obsahem listu „BRIGHTNESS“ v souboru *Testování parametrů.xls*, které ukazují výsledky testování parametru *Brightness* (vizte kap. 2.1.3). Jednotlivé listy souboru se vztahují k testování jednotlivých parametrů.

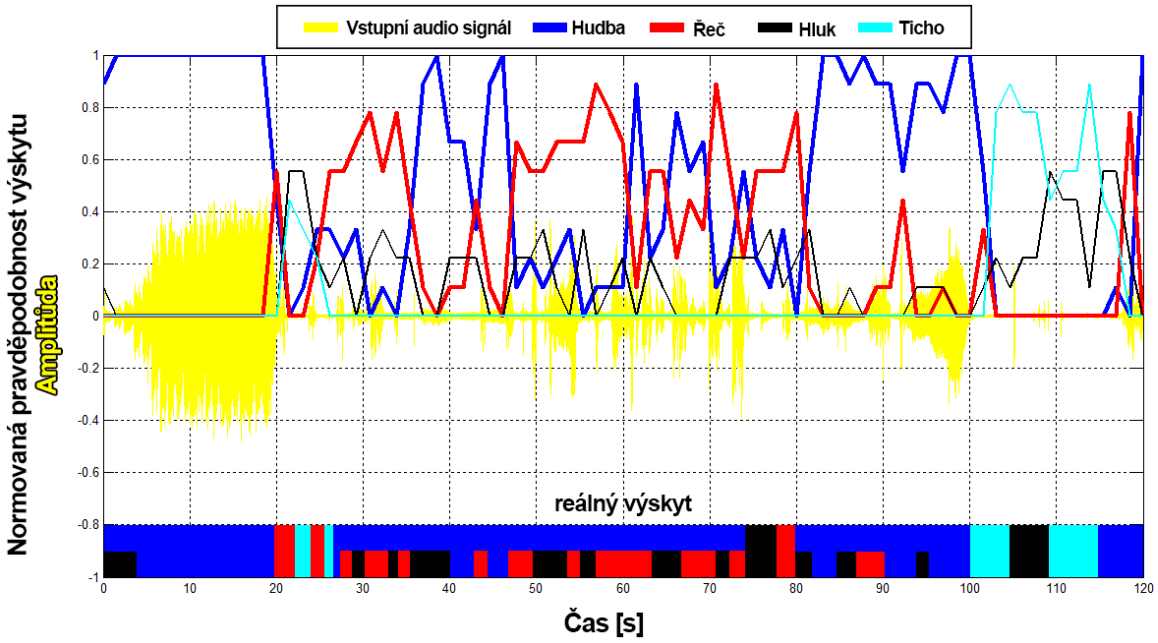
V tabulkách jsou uvedené hodnoty pravděpodobností výskytů jednotlivých typů zvuku, při testování s různými hodnotami parametru klasifikátoru *k*. První tři tabulky ukazují výsledky testování testovací množiny a další tři ilustrují výsledky testování trénovací množiny. Modrá barva označuje typ testovaného zvuku.

Testování testovací množiny dat								
hudba	**82,2**	hudba	16,3	hudba	40,0	hudba	0,0	**k = 9**
řeč	17,8	**řeč**	**81,4**	řeč	46,7	řeč	14,3	
hluk	0,0	hluk	0,0	**hluk**	**0,0**	hluk	0,0	
ticho	0,0	ticho	2,3	ticho	13,3	**ticho**	**85,7**	
hudba	**86,3**	hudba	27,9	hudba	40,0	hudba	0,0	**k = 7**
řeč	12,3	**řeč**	**60,5**	řeč	40,0	řeč	14,3	
hluk	0,0	hluk	2,3	**hluk**	**0,0**	hluk	0,0	
ticho	1,4	ticho	9,3	ticho	20,0	**ticho**	**85,7**	
hudba	**87,7**	hudba	27,9	hudba	40,0	hudba	0,0	**k = 5**
řeč	11,0	**řeč**	**60,5**	řeč	33,3	řeč	14,3	
hluk	0,0	hluk	2,3	**hluk**	**0,0**	hluk	0,0	
ticho	1,4	ticho	9,3	ticho	26,7	**ticho**	**85,7**	

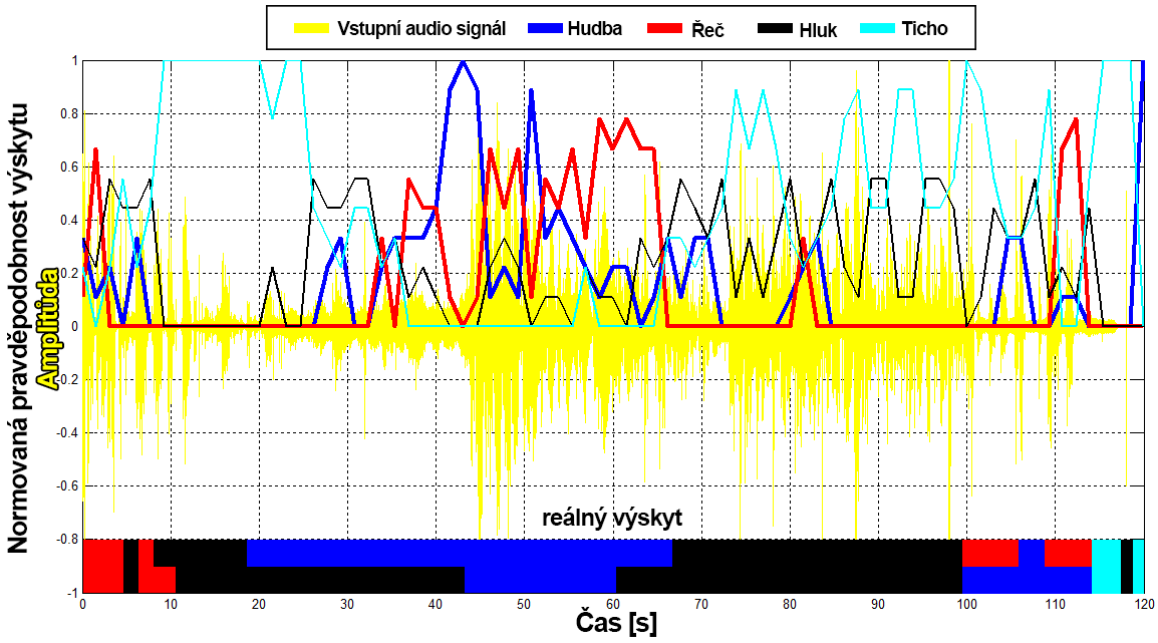
	Testování trénovací množiny dat							
k = 9	**hudba**	**90,8**	hudba	39,1	hudba	33,3	hudba	0,0
	řeč	9,2	**řeč**	**59,8**	řeč	40,0	řeč	20,7
	hluk	0,0	hluk	0,0	**hluk**	**0,0**	hluk	0,0
	ticho	0,0	ticho	1,1	ticho	26,7	**ticho**	**79,3**
k = 7	**hudba**	**90,2**	hudba	34,5	hudba	33,3	hudba	0,0
	řeč	9,8	**řeč**	**62,1**	řeč	43,3	řeč	24,1
	hluk	0,0	hluk	2,3	**hluk**	**0,0**	hluk	0,0
	ticho	0,0	ticho	1,1	ticho	23,3	**ticho**	**75,9**
k = 5	**hudba**	**88,4**	hudba	32,2	hudba	33,3	hudba	0,0
	řeč	11,6	**řeč**	**66,7**	řeč	43,3	řeč	24,1
	hluk	0,0	hluk	0,0	**hluk**	**0,0**	hluk	0,0
	ticho	0,0	ticho	1,1	ticho	23,3	**ticho**	**75,9**

Tab. 1Testování parametrů

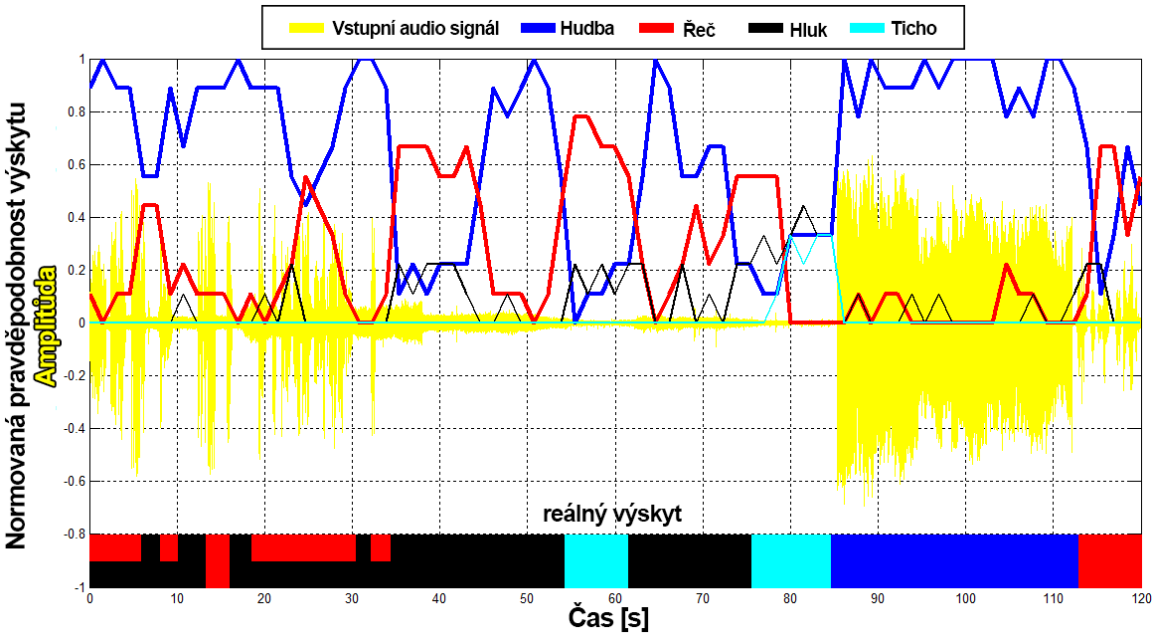
Následující grafy níže ilustrují pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase s označením reálného výskytu těchto typů.



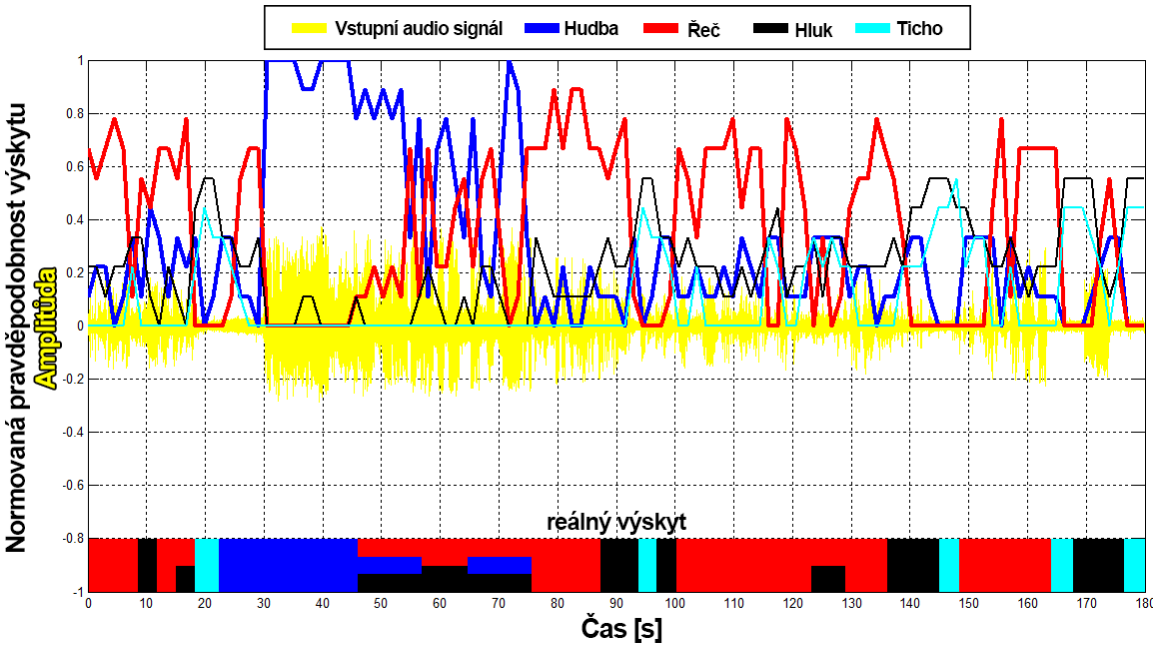
Obr. 1 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Lev s bílou hřívou“



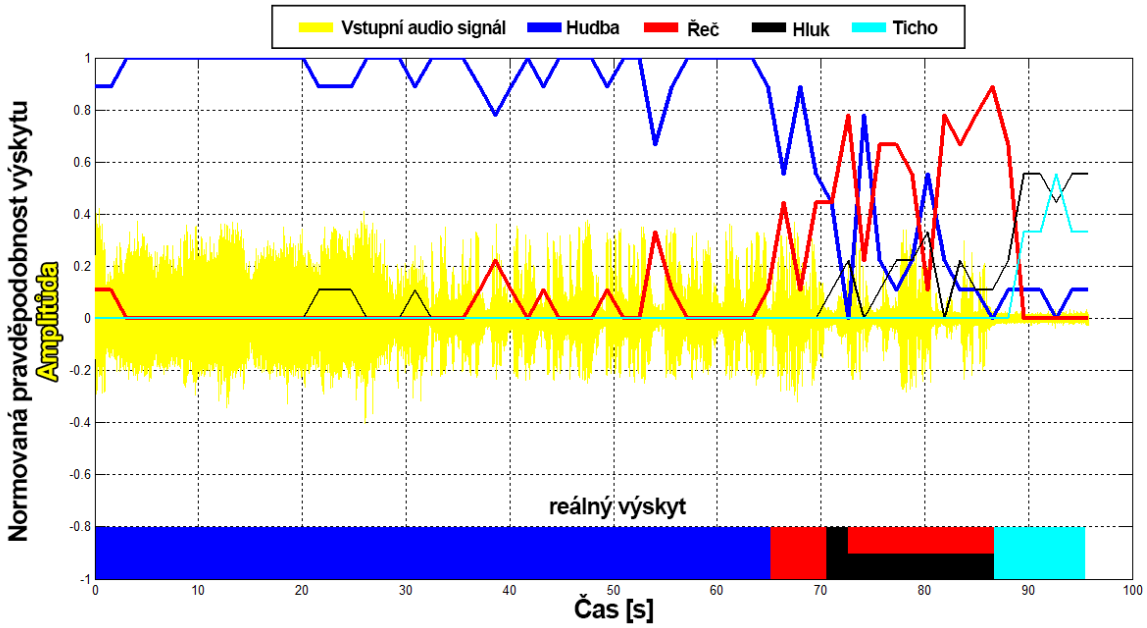
Obr. 2 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Slavnosti sněženek“



Obr. 3 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Valčík pro milión“



Obr. 4 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Jedenácté přikázání“ (výřez 1)



Obr. 5 Pravděpodobnost výskytu jednotlivých typů zvukových záznamů v závislosti na čase pro nahrávku z filmu „Jedenácté přikázání“ (výřez 2)