

ROZPOZNÁNÍ TITULU GRAMOFONOVÉ DESKY PODLE KRÁTKÉ UKÁZKY

V. Moldan, F. Rund

Katedra radioelektroniky, fakulta elektrotechnická
České vysoké učení technické v Praze, Česká republika

Abstrakt

Tento článek popisuje návrh a implementaci vlastního řešení pro úlohu identifikace nahrávky, se zaměřením na identifikaci nahrávek z gramofonové desky. Algoritmus je implementován v programovém prostředí MATLAB. Pro klasifikaci nahrávky a vyhledávání v databázi nahrávek je zvolena metoda strojového učení K-nearest neighbors (k-nejbližších sousedů, KNN). Pro parametrizaci nahrávky je použito již připravených parametrů z Music Information Retrieval (MIR) Toolboxu, nahrávka je klasifikována pomocí parametrů tempo a mode. Pro testování algoritmu byla jako vedlejší funkcionální program vytvořena možnost přidat typické zvukové atributy gramofonových desek do existujících nahrávek. Po natrénování na množině všech 269 dostupných masterů nahrávek byl program schopen rozpoznat 92,8 % nahrávek uměle znehodnocených atributy gramofonové desky (posunutí začátku, snížení amplitudy, přidání šumu a ztráta vysokých frekvencí) a zároveň všech 5 dostupných nahrávek pořízených z gramofonové desky. Pro ovládání programu bylo také vytvořené jednoduché grafické rozhraní.

1 Úvod

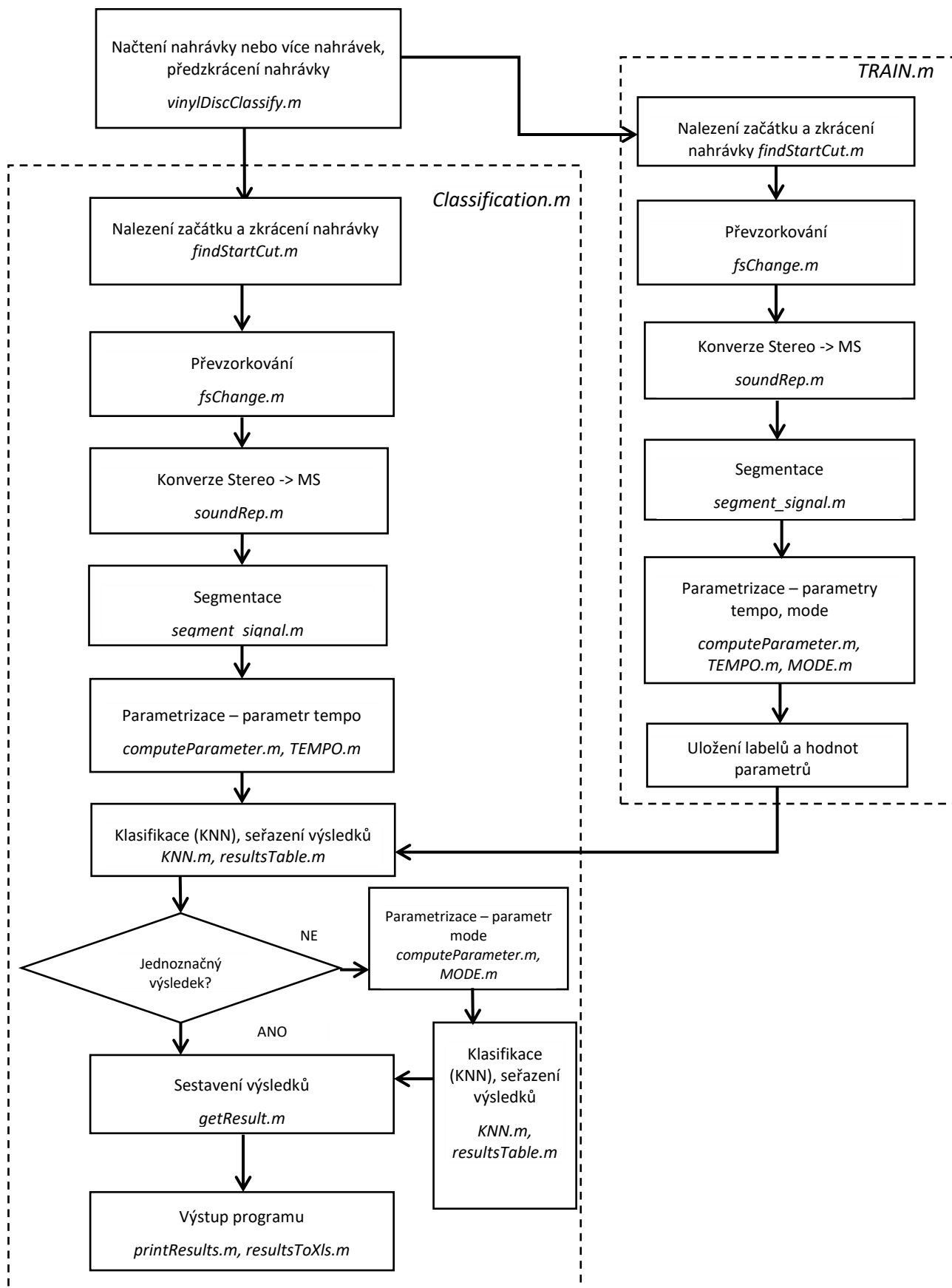
Téma rozpoznávání hudebních děl nachází v dnešní době stále větší popularitu. Stále dostupnějšími se stávají například mobilní aplikace se schopností rozpoznat nahrávku podle krátké zvukové ukázky. Možné využití těchto algoritmů je ale i v hromadné výrobě gramofonových desek, pro kontrolu správné identifikace jednotlivých nahrávek.

Pro problematiku rozpoznávání zvukových nahrávek existuje řada přístupů. Vyskytují se i různé variace této úlohy, kromě identifikace konkrétní nahrávky lze například zařazovat nahrávky do nějaké kategorie, například dle hudebního žánru, nebo rozpoznat, zda se jedná o hudební nahrávku, mluvené slovo nebo jiný typ nahrávky.

Zásadním problémem pro využití systému pro identifikaci nahrávky pro rozpoznání titulu gramofonové desky je změna vlastností nahrávky vlastním záznamem na gramofonovou desku. Je proto nutné vybrat takové parametry pro popis nahrávky, které nebudou ovlivněny záznamem na desku. Tato problematika byla rozpracována již v bakalářské práci [1], kde lze nalézt podrobnější teoretický popis zvolené klasifikační metody, jiných existujících řešení a popis vlastností nahrávek na gramofonové desce. Tento příspěvek se zaměřuje na popis implementace algoritmu v prostředí Matlab, včetně grafického uživatelského rozhraní.

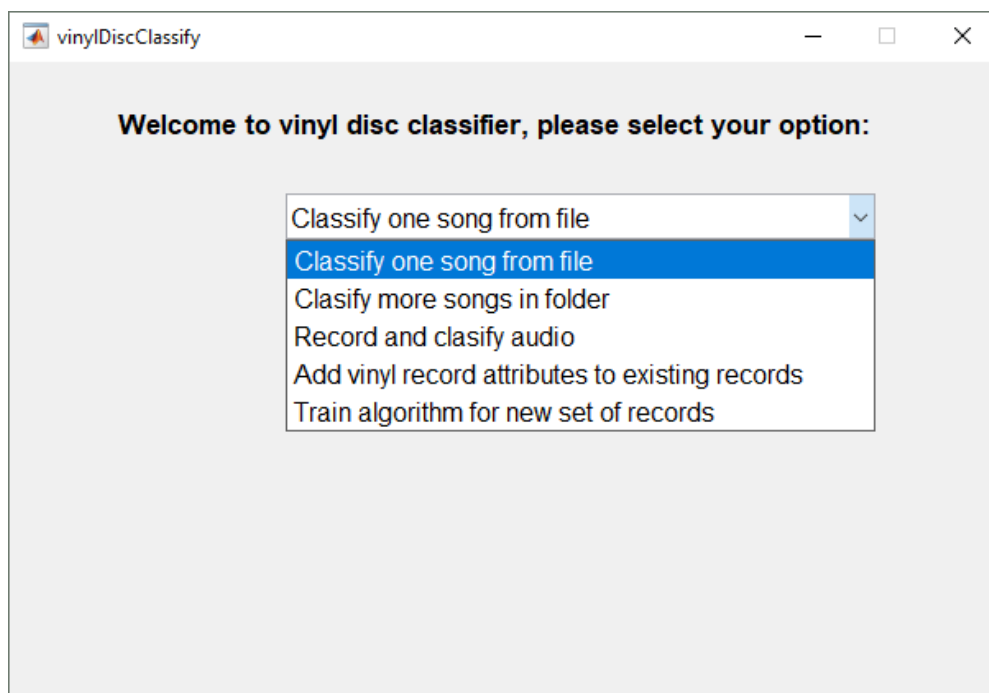
2 Popis systému

Cílem navrhovaného programového systému je identifikace předložené audio nahrávky, ke které je přiřazena nejpodobnější nahrávka z dostupné databáze. Jelikož bylo pro účely práce nutné algoritmus testovat i na nahrávkách s charakteristickými rysy gramofonových nahrávek, nabízí program jako vedlejší funkcionální také možnost přidat do existujících nahrávek simulované typické atributy gramofonových desek. Schéma systému je na obr. 1.



Obr. 1: Schéma navrhovaného systému

Program je řízen pomocí jednoduchého grafického rozhraní. Po spuštění si uživatel může vybrat, k jaké funkci program chce využít (obr. 2). Možnosti jsou identifikace nahrávky uložené v počítači, identifikace většího počtu nahrávek a uložení výstupních dat do souboru pro tabulkový procesor (MS Excel), nahrání zvuku z externího zdroje (pomocí například mikrofону nebo gramofonu připojeného k PC) a jeho identifikace, vytvoření nahrávek se simulovanými atributy gramofonových desek nebo natrénování programu na novou množinu nahrávek (vytvoření nové databáze nahrávek).



Obr. 2. Výběr požadované funkcionality programu v implementovaném grafickém rozhraní

Program je spuštěn pomocí funkce *vinylDiscClassify.m*, která spustí grafické rozhraní. Po výběru funkcionality identifikaci nahrávky a po načtení požadované nahrávky dojde okamžitě ke zkrácení nahrávky na 60 s, čímž zabráníme přetížení programu, které by mohlo nastat při práci s příliš dlouhou nahrávkou.

Po ukončení identifikace je v GUI vypsán název nahrávky, o kterou se pravděpodobně jedná. V případě, že program identifikuje stejné pravděpodobnosti u více nahrávek, jsou v GUI vypsány všechny tyto nahrávky.

3 Implementace systému

3.1 Předzpracování nahrávky

3.1.1 Převzorkování

Pro další výpočty je výhodné zkoumanou nahrávku vždy převést na stejný referenční kmitočet. V programu je proto implementována jednoduchá funkce *fsChange.m*, která je založena na návrhu z bakalářské práce [2]. Pro samotné převzorkování je použita funkce MATLABu *resample*, ve funkci *fsChange.m* dojde k přepočtení jejích vstupních parametrů (původní vzorkovací frekvence a nová vzorkovací frekvence) na vstupní parametry požadované funkcí *resample*.

Jako referenční kmitočet je zvolen vzorkovací kmitočet 48 kHz.

3.1.2 Segmentace

Dalším krokem je segmentace nahrávky na kratší časové úseky, pro které bude prováděn výpočet parametrů. Segmentace je v programu implementována ve funkci *segment_signal.m*, která je převzata z bakalářské práce [2]. Je zvolena délka použitých segmentů 4,5 s a délka překryvu 1,5 s.

3.1.3 Zkrácení nahrávky

Pro zajištění shody úseku nahrávky uloženého v databázi a zkoumaného úseku nahrávky byla do programu zařazena funkce pro nalezení začátku a zkrácení nejen testovaných nahrávek, ale i nahrávek z trénovací množiny. Implementovaná funkce *findStartCut.m* využívá návrhu, který je popsán v bakalářské práci [2].

Začátek skladby je detekován tam, kde hodnota nahrávky přesáhne daný práh. V bakalářské práci [2] je použit práh 0,25 hodnoty maximální amplitudy přítomné na předzkrácené 40 s nahrávce. Tato hodnota prahu snížena na 0,15 maximální nalezené amplitudy, pro lepší zajištění detekce začátku i v případě většího rozdílu mezi maximální amplitudou a začátkem hudby. Zároveň je to hodnota, která by měla být ještě stále výrazně větší než hodnota šumu na nahrávce.

Nicméně ani po využití této funkcionality nemůžeme očekávat naprosto přesnou shodu úseků, ale pro identifikace je důležité, aby posunutí bylo výrazně menší, než je délka segmentů, čehož se touto funkcí téměř vždy podařilo dosáhnout.

3.1.4 Převod nahrávky do libovolné reprezentace signálu

Pro účely rozpoznání desky se nabízí možnost využít i jiných možných reprezentací signálu, než je tradiční stereofonní či monofonní reprezentace. Možností je využít méně typický tvar mid-side (MS), kdy se v kanálu mid nachází součet L+R a v kanálu side rozdíl L-R. Pro tento účel je v programu napsaná funkce *soundRep.m*. Výstupem funkce je zvuk v libovolném tvaru, buď stereo, mono nebo ve tvaru MS.

Na základě testování všech zmíněných možných reprezentací signálů byla vybrána právě reprezentace mid side. Výsledky testování jsou uvedené v tabulce 1.

Tabulka 1: Vliv zvolené reprezentace signálu na úspěšnost identifikace (pozn.: číslo v závorce udává počet správně rozpoznávaných nahrávek z 264 testovaných)

Reprezentace signálu	Úspěšnost identifikace
Mono	88,6 % (234 nahrávek)
Stereo	88,6 % (234 nahrávek)
MS	92,8 % (245 nahrávek)

3.2 Parametrizace

Pro identifikaci nahrávky je možno použít celou řadu parametrů. Jednou často používanou skupinou jsou signálové parametry, jako např. parametry odvozené ze spektra nahrávky, energie a výkon signálu apod. Další skupinou jsou hudební parametry, jako například tempo skladby na nahrávce nebo její tónina.

Při implementaci tohoto algoritmu byly využity již vytvořené parametry z Music Information Retrieval Toolboxu [4] (dále MIR Toolbox). Tento volně dostupný toolbox umožňuje snadno spočítat mnoho parametrů audio nahrávky.

Pro výpočet konkrétního parametru pro všechny segmenty nahrávky je v programu implementována funkce *computeParameter.m*. Prvním vstupním argumentem této funkce je matice jednotlivých segmentů nahrávky a druhým je řetězec udávající požadovaný parametr.

3.2.1 Výběr parametrů

Z velkého množství parametrů, které MIR Toolbox nabízí, bylo nutné najít vhodnou kombinaci parametrů, která se hodí pro identifikaci konkrétní nahrávky a zejména nahrávky na gramofonové desce.

Na základě testování byly vybrány parametry z hudební oblasti tempo a mode. Tyto parametry jsou implementovány v hierarchii, ve funkci *getResults.m* dojde k porovnání výsledků metody získaných metodou KNN pro každý parametr zvlášť a získání konečného výsledků tak, že z matice

výsledků získaných podle druhého parametru (mode) jsou vybrány pouze výsledky, které jsou první v pořadí podle hlavního parametru (tempo). Z takto získané matice je vybrána nahrávka na prvním místě, případně více nahrávek se stejnou pravděpodobností.

3.2.2 Parametr tempo

Tempo popisuje nahrávku z hlediska rytmu a rychlosti, udává se v jednotkách bpm (beats per minute), tj. počet čtvrtových not za minutu. Výpočet v MIR Toolboxu [4] je založen na autokorelační funkci, kdy jsou nalezeny její vrcholy a z nich je poté odhadnuta hodnota tempa v bpm.

3.2.3 Parametr mode

Mode (česky modus, mód) popisuje nahrávku z hlediska tónového uspořádání. Výpočet tohoto parametru v MIR Toolboxu [4] je omezen na odlišení molové a durové tóniny. Čím je hodnota parametru větší než 0, tím více durová je očekávaná tónina skladby, čím je menší než 0, tím více molová je očekávaná tónina skladby.

3.3 Klasifikační algoritmus KNN

KNN je algoritmus strojového učení, učení probíhá z trénovacích dat. Trénovací data jsou rozdělena do konkrétních tříd, do těch jsou poté při klasifikaci zařazována i testovaná data. Každý prvek trénovací množiny je poté umístěn do některého místa v N-rozměrném prostoru. Následně přichází na řadu samotná klasifikace, kdy je testovaný objekt zařazen do té třídy, do které patří většina jeho nejbližších sousedů. Pomocí parametru k můžeme určit, kolik nejbližších sousedů se bude při vyhodnocování uvažovat.

Pro implementaci metody strojového učení KNN je implementována funkce $KNN.m$. Základ této funkce je převzat z bakalářské práce [2] a je upraven pro použití pro úlohu identifikace konkrétní nahrávky.

Použití metody strojového učení KNN pro úlohy identifikace konkrétní nahrávky je specifické tím, že každá nahrávka je sama o sobě třídou, máme tedy v našem případě 269 tříd, a pro natrénování klasifikátoru máme od každé třídy k dispozici pouze jednu nahrávku.

Na základě výsledků testování byla zvolena hodnota parametru $k=4$. Výsledky testování jsou uvedeny v tabulce 2.

Tabulka 2: Vliv hodnoty k na úspěšnost identifikace (pozn.: číslo v závorce udává počet správně rozpoznaných nahrávek z 264 testovaných, testování bylo prováděno při mono reprezentaci signálu)

k	Úspěšnost identifikace
3	86,7 % (229 nahrávek)
4	88,6 % (234 nahrávek)
5	87,8 % (232 nahrávek)
6	88,2 % (233 nahrávek)
7	88,2 % (233 nahrávek)

3.4 Simulace vlastností vinylových desek

Pro ověření funkčnosti výsledného programu pro použití na nahrávkách z vinylových desek, které mají charakteristické zvukové vlastnosti, bylo vhodné mít k dispozici dostatečné množství nahrávek s těmito vlastnostmi. Protože se pro testování podařilo získat pouze 5 reálných nahrávek

z gramofonových desek, bylo nutné simulovat vlastnosti nahrávek z gramofonových desek a přidat je do masterů nahrávek, pro zvětšení testovací skupiny.

Program tedy nabízí funkcionalitu vytvoření nahrávek s přidávanými atributy gramofonových desek. Po vybrání této funkcionality a spuštění programu dojde v aktuální složce MATLABu k vytvoření nových nahrávek, do kterých jsou přidány atributy gramofonových desek.

Přidání těchto atributů je řízeno ve funkci *vinylSim.m*, ve které jsou dále volány funkce, které řeší konkrétní atributy. Ty jsou dále popsány v následujících odstavcích.

3.4.1 Přidání šumu, snížení amplitudy

Pro přidání šumu do nahrávek je vytvořena funkce *addNoise.m*. Vzhledem k charakteristikám šumu gramofonového záznamu nebyl tento šum generován v MATLABu ale byl použit šum extrahovaný z nahrávek z reálných gramofonových desek.

Ve funkci je náhodně vybrán jeden ze tří krátkých úseků šumu extrahovaných z nahrávky z gramofonové desky, ten je načten ze souboru a přidán do nahrávky. Zároveň je pro každý z těchto šumů ve funkci zapsána hodnota amplitudy, která se vyskytovala na nahrávce, ze které byl šum pořízen, a na tuto hodnotu je amplituda výsledné nahrávky snížena. Tím docílíme přibližného zachování poměru signálu a šumu SNR.

3.4.2 Úprava frekvenčního rozsahu

Pro gramofonovou desku je typické, že s rostoucím počtem přehrání dochází k jejímu opotřebení a postupnému snižování frekvenčního rozsahu gramofonové desky. Tento úbytek je znát především na vysokých frekvencích.

Za tímto účelem byla v práci implementována dolní propust pomocí funkce *lowPass.m*. Funkce využívá zabudované funkce MATLABu pro Butterworthův filtr *butter*. Funkce *lowpass.m* umožňuje vybrat libovolnou zlomovou frekvenci pomocí vstupního parametru *cutOff*, a tu poté přepočte na vstupní argumenty zabudované funkce MATLABu *butter*. Zlomová frekvence byla nastavena jako 15 kHz.

3.4.3 Posunutí začátku nahrávky

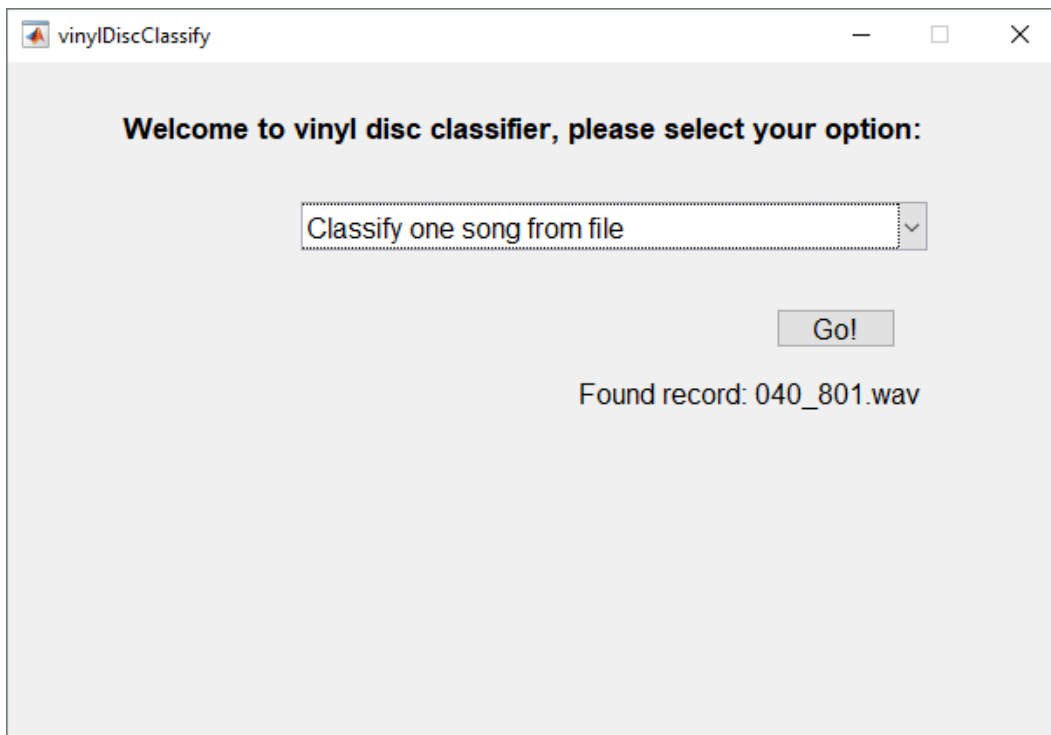
Ačkoliv se posunutí začátku nahrávky dá těžko řadit mezi typické atributy gramofonové desky, je pro ověření správné funkcionality výsledného programu nutné zařadit do simulace nahrávky z gramofonové desky i tento prvek.

Pro posunutí začátku je implementována funkce *shiftStart.m*, která umožní posunutí začátku (tedy přidání ticha před začátek samotné nahrávky) o zvolený čas a zároveň zkrácení výsledné nahrávky na požadovanou délku. Protože i na nahrávkách pořízených z vinylů byla doba posunutí až 10 s, rozhodl jsem se začátek všech nahrávek také posunout o 10 s.

3.5 Grafické rozhraní, řízení programu

Jednoduché grafické rozhraní je vytvořené pomocí zabudovaného průvodce grafickým rozhraní MATLABu *guide*.

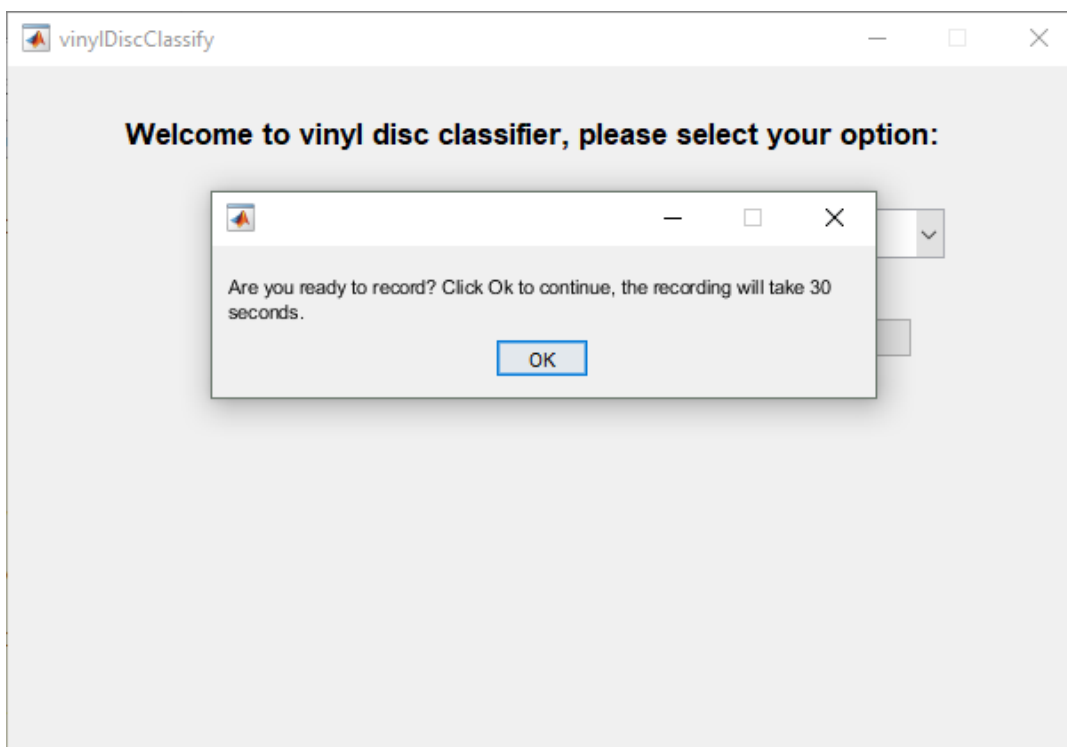
Pomocí pop-up menu je možná vybrat požadovanou funkcionalitu a volbu potvrdit stisknutím tlačítka "GO!". O průběhu celého procesu informuje textové pole (na obrázku níže například informuje o tom, že zkoumaná nahrávka byla identifikována jako nahrávka *040_801.wav*).



Obr. 4: Výpis nalezené nahrávky v grafickém rozhraní.

3.6 Nahrání nahrávky z externího zdroje

Nahrání nahrávky z externího zdroje je prováděno pomocí zabudované funkce MATLABu *audiorecorder*. Začátek nahrávání spustí uživatel stisknutím tlačítka “Ok” v dialogovém okně, které se zobrazí po vybrání této funkcionality. Po 30 sekundách je nahrávání ukončeno a s nahrávkou se dále pracuje stejným způsobem, jako při identifikaci nahrávky ze souboru.



Obr. 5: Dialogové okno, které uživateli umožní zahájit nahrávání nahrávky z externího zdroje.

3.7 Úspěšnost algoritmu

Při testování úspěšnosti programu se podařilo dosáhnout úspěšnosti 92,8 % (vizte Tabulka 1) při identifikaci masterů s přidávanými simulovanými vlastnostmi gramofonových nahrávek a zároveň se podařilo správně rozpoznat všech 5 reálných vinylových nahrávek, které byly pro testování k dispozici.

Vzhledem k tomu, že program v obou případech musel rozpoznávat nahrávky, které měly od původních značně odlišné vlastnosti, dá se výsledek testu označit za dobrý.

3.8 Výpočetní náročnost

Identifikace jedné nahrávky ze souboru zabere programu poměrně dlouhou dobu, přibližně 10 s. Při detailnějším zkoumání se ukázalo, že nejvíce času zabírá výpočet tempa v MIR Toolboxu, který se pohybuje pro 20 s dlouhou nahrávku mezi 5 a 8 s.

Tyto časy byly naměřeny na PC s dvoujádrovým procesorem a taktovací frekvencí 2,7 GHz.

Čas nutný k výpočtu tempa je dán složitým výpočtem nutným k získání tohoto parametru a jeho výpočtem pro všechny segmenty zkoumaného úseku nahrávky. Tuto dobu by bylo možné zkrátit, pokud by tempo nebylo stanovováno pro všechny segmenty nahrávky, ale pouze pro celý zkoumaný úsek. Tím by ale zanikla informace o vývoji tempa nahrávky v čase. Zároveň by bylo možné na základě již napsané funkce pro tempo v MIR Toolboxu implementovat vlastní podobnou funkci a pokusit se ji optimalizovat pro účely tohoto programu.

4 Závěr

V bakalářské práci [1] se úspěšně podařilo implementovat vlastní algoritmus se schopností rozpoznat nahrávku z databáze. Při návrhu algoritmu byly zohledněny a modelovány typické zvukové atributy gramofonových desek. Tento článek popisuje rozšíření tohoto algoritmu o jednoduché grafické rozhraní.

Výsledek 92,8 % správně rozpoznaných nahrávek s upravenými vlastnostmi se dá označit jako dobrý. Nevýhodou vytvořeného programu zůstává jeho velká výpočetní náročnost. V další práci by bylo vhodné zaměřit se zejména na snížení této náročnosti, buď nalezením vhodné doby segmentu pro výpočet parametru tempo, případně nahrazením funkce pro tempo z MIR Toolboxu vlastní funkcí.

Poděkování

Tato práce byla podpořena grantem Studentské grantové soutěže ČVUT č. SGS17/190/OHK3/3T/13.

Literatura

- [1] Moldan, V., Rozpoznání titulu gramofonové desky podle krátké ukázky, FEL ČVUT, 2017
- [2] Koshkina, E., Identifikace obsahu archivních zvukových záznamů, FEL ČVUT, 2015
- [3] Král, V., Identifikace gramofonové desky podle krátké zvukové ukázky, FEL ČVUT, 2016
- [4] Lartillot, O., MIRtoolbox 1.5 User's Manual, Finnish Centre of Excellence in Interdisciplinary Music Research, 2013

Bc. Václav Moldan

Katedra radioelektroniky, FEL ČVUT v Praze, Technická 2, 166 27 Praha 6, ČR
email: moldavac@fel.cvut.cz

Ing. František Rund, Ph.D.

Katedra radioelektroniky, FEL ČVUT v Praze, Technická 2, 166 27 Praha 6, ČR
email: xrund@fel.cvut.cz