VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

ODSTRANĚNÍ ZNÁMÉHO SIGNÁLU Z NAHRÁVKY

REMOVAL OF A KNOWN SIGNAL FROM A RECORDING

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE **PAVEL HOŠEK**
AUTHOR

VEDOUCÍ PRÁCE **Doc. Dr. Ing. JAN ČERNOCKÝ**
SUPERVISOR

BRNO 2019

Vysoké učení technické v Brně
Fakulta informačních technologií

Ústav počítačové grafiky a multimédií (UPGM) Akademický rok 2018/2019

Zadání bakalářské práce

Student: **Hošek Pavel**
Program: Informační technologie
Název: **Odstranění známého signálu z nahrávky**
Removal of a Known Signal from a Recording
Kategorie: Zpracování signálů

Zadání:

1. Seznamte se s metodami analýzy signálů a řeči.
2. Vytvořte simulovaná data, kdy budete k užitečnému řečovému signálu přimíchávat známé rušení prošlé simulací místnosti a bude uměle přidán šum. Rušivý signál pokládejte za známý.
3. Navrhněte a implementujte techniku pro synchronizaci nahraného a známého zvuku a pro odstranění známého zvuku z nahrávky.
4. Vyhodnoťte, např. pomocí poměru signálu k šumu (SNR).
5. Získejte malou datovou sadu reálných nahrávek se známými signály (dostupnými např. z archivu rozhlasu nebo televize) a otestujte Vaši techniku.
6. Výsledky vyhodnoťte, případně navrhněte zlepšení (např. využití neuronových sítí, známých technik dereverberace a odšumování, atd).
7. Vytvořte krátké video dokumentující Vaši práci.

Literatura:

- dle doporučení vedoucího.

Podrobné závazné pokyny pro vypracování práce viz http://www.fit.vutbr.cz/info/szz/

Vedoucí práce: **Černocký Jan, doc. Dr. Ing.**
Vedoucí ústavu: Černocký Jan, doc. Dr. Ing.
Datum zadání: 1. listopadu 2018
Datum odevzdání: 15. května 2019
Datum schválení: 2. listopadu 2018

Abstrakt

Cílem této bakalářské práce je návrh a implementace metody, která z audio nahrávky odstraní známý signál. V úvodní části práce je shrnuta teorie o vlastnostech zvuku, základech zpracování signálu a o identifikaci systému. Následně jsou představena data a je uvedeno jakým způsobem bude probíhat testování. Dále je popsán vývoj metody pro odstranění známého signálu z nahrávky. Na závěr jsou pak provedeny experimenty , které zahrnují i porovnání vybraných metod. Nejlepší metody dosahovaly poměrně kvalitních výsledků a známý signál většinou dokázaly v dostatečné míře odstranit.

Abstract

The goal of this bachelor thesis is to design and implement method for removing known signal from recorded sound. The introductory part of the thesis summarizes the theory of sound properties, basics of signal processing, and system identification. Subsequently, the data are presented and the way of testing is stated. Afterwards, the development of a method for removing a known signal from a recording is described. Finally, experiments are performed, which include a comparison of selected methods. The best methods achieved relatively good results and they were mostly able to remove the known signal sufficiently.

Klíčová slova

odstranění rušivého signálu, zpracování signálu, identifikace systému, impulzní odezva, časově-frekvenční analýza

Keywords

removal of interfering signal, signal processing, system identification, impulse response, time–frequency analysis

Citace

HOŠEK, Pavel. *Odstranění známého signálu z nahrávky.* Brno, 2019. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Doc. Dr. Ing. Jan Černocký

Odstranění známého signálu z nahrávky

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Jana Černockého. Další informace mi poskytli Kateřina Žmolíková, Jan Vlk a Jiří Jan. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Pavel Hošek
14. května 2019

Poděkování

Chtěl bych poděkovat zejména vedoucímu práce Janu Černockému za cenné rady, věcné připomínky a vstřícnost při konzultacích a vypracování bakalářské práce. Poděkování patří také Kateřině Žmolíkové, Jiřímu Janovi a Janu Vlkovi za cenné rady.

Obsah

Kapitola 1

Úvod

V dnešní době máme technologie, které nám umožňují nahrávání zvuku a potom jeho přehrání. Před 200 lety bylo jistě těžko představitelné, že bude možné poslouchat skladbu oblíbeného umělce z nějakého přístroje. Nahrát si svou dceru, která recituje báseň a poté si ji poslechnout třeba za několik let. V dnešní době se audio nahrávky využívají v mnoha oblastech. Hudební nahrávky nám zpříjemňují práci i náš volný čas. Nahrávky z černé skřínky umístěné v letadle pomáhají při objasnění příčiny havárie letadla. Nicméně, většina nahrávek obsahuje určité rušení, které znehodnocuje užitečný zvukový obsah. Příkladem může být rozhovor dvou teroristů, kteří se baví v místnosti a v pozadí mají zapnutou hlasitě televizi, či záměrně zapnutou rušičku. V nahrávce tedy není pouze jejich rozhovor, ale je tam obsažen i rušivý signál. Zjistit informace z takových dat je pak velmi obtížné jak pro lidského posluchače, tak pro případné automatické zpracování. Dalším reálným problémem je hlasové ovládání zařízení, které přijímá a zároveň vydává zvuky. Jako příklad lze uvést televizor, který by měl být ovládán pomocí hlasu. Televizor přijímá hlasové pokyny uživatele a zároveň produkuje zvuk probíhajícího televizního pořadu. Je tak velmi těžké analyzovat hlasové pokyny uživatele, které jsou zarušeny zvukem televize. Řešením těchto reálných problémů je rušivý signál z nahrávky odstranit. A právě odstranění rušivého signálu je cíl této práce. Rušivý signál může být náhodný signál, či signál který lze vyrobit nebo signál jehož obsah je známý (např. televize, rádio). Tato práce se bude zabývat odstraněním známého signálu z nahrávky. I když je rušivý signál předem známý, není možné ho jednoduše odečíst. Důvodem je to, že rušivý signál v nahrávce většinou vypadá jinak. Při průchodu signálu prostředím se signál zeslabuje. Může dojít také k odrazu signálu od pevné překážky, odražené vlny se smíchají a mění tak původní signál. Samotné přehrávání či nahrávaní zvuku může také pozměnit signál. Pokud je rušivý signál v místnosti přehráván reproduktorem a nahráván pomocí mikrofonu, každé toto zařízení signál jistým způsobem pozmění.

V kapitole 2 je podán úvod do problematiky. Dále jsou v kapitole obsaženy teoretické znalosti, které jsou potřeba pro řešení problému. V kapitole 3 jsou představena data, která byla vytvořena pro testování a vyhodnocení navržených metod. V kapitole 4 je popsán vývoj metody pro odstranění známého signálu ze zarušené nahrávky. V kapitole 5 jsou obsaženy experimenty s vybranými metodami. V kapitole 6 je představena demonstrační aplikace. V poslední kapitole je shrnuta provedená práce a je navrženo, jakým směrem by se mohlo ubírat pokračovaní této práce.

Kapitola 2

Odstranění známého signálu z nahrávky

Reálná nahrávka může vypadat následovně: V jedné místnosti u stolu spolu hovoří dva lidé. Mikrofon, který zaznamenává zvukový signál je umístěn v přední části místnosti. V pozadí je zapnutá televize, kde běží nějaký pořad. Zvuk, který je mikrofonem zaznamenán, obsahuje jak rozhovor dvou lidí, tak televizi. Jak již bylo zmíněno v úvodu, signál který projde prostředím se určitým způsobem změní v závislosti na prostředí. V nahrávce je tedy signál rozhovoru pozměněn místností. Signál televize je také pozměněn místností, ale změna signálu nemusí být stejná, jelikož televize je v jiné vzdálenosti od mikrofonu než hovořící lidé. Výsledný signál, který je zaznamenáván mikrofonem je tedy dán jako

$$c[n] = u[n] * h[n] + z[n] * h'[n], \tag{2.1}$$

kde c je signál zarušené nahrávky, u je užitečný signál rozhovoru dvou lidí, h a h' jsou impulzní odezvy prostředí a z je známý rušivý signál televize. Ilustrace viz obrázek 2.1. Užitečný signál po průchodu místností lze potom vyjádřit jako

$$(u[n] * h[n]) = c[n] - (z[n] * h'[n]), \tag{2.2}$$

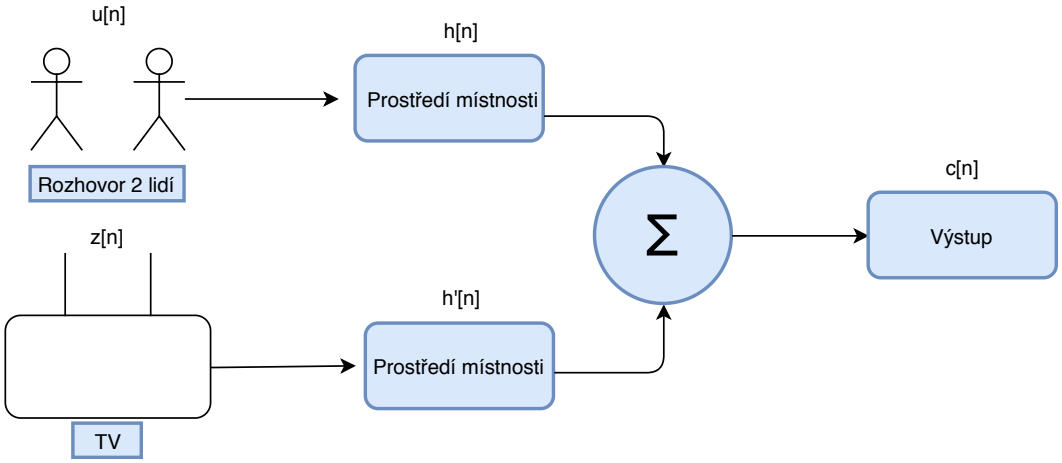
Jestliže by dané prostředí nijak neměnilo šířící se signál, lze potom získat užitečný signál jako

$$u[n] = c[n] - z[n]. \tag{2.3}$$

V reálných případech ale tento postup nelze aplikovat. Signál je zpravidla určitým způsobem modifikován a to při průchodu prostředím nebo při samotném nahrávání. Aby bylo možné rušivý signál odečíst, je potřeba odhadnout impulzní odezvu pro rušivý signál. Tato impulzní odezva bude poté aplikována na známý rušivý signál z jako

$$z'[n] = z[n] * h'[n], \tag{2.4}$$

kde z' značí rušivý signál, který je pozměněn průchodem místností. Jakmile je získán signál $z'[n]$, je možné ho odečíst od zarušené nahrávky, viz rovnice 2.2, a tím získat užitečný signál $u[n]$, který prošel prostředím.



Obrázek 2.1: Průchod signálu místností.

2.1 Zvuk a jeho šíření v prostoru

Zvuk je mechanické vlnění, které se šíří ze zdroje pouze pružným látkovým prostředím libovolného skupenství. Nejčastěji se jedná o vzduch, kde se zvuk šíří jako podélné postupné vlnění. Dochází k periodickému stlačování a rozpínání vzduchu, což se projevuje periodickými změnami tlaku vzduchu. Ve všech prostředích se zvuk šíří jako postupné podélné vlnění. Rychlost zvuku ve vzduchu je cca 343 metrů za sekundu v místnosti o pokojové teplotě.

Odraz zvuku

Pokud dopadne zvuková vlna na překážku může dojít k jejímu odrazu. Směr odražené vlny je opačný vůči původní vlně. Amplituda odražené vlny je menší a její fáze je posunuta. Odraz zvuku se řídí zákonem odrazu. Zákon odrazu říká, že úhel odrazu je stejný jako úhel dopadu a dopadající i odražený paprsek leží v rovině dopadu [8]. Ilustrace odrazu zvuku viz obrázek 2.2.

Útlum zvuku

Při průchodu zvuku prostředím dochází ke snižování intenzity zvuku. Intenzita zvuku se zmenšuje se vzdáleností od zdroje zvuku, vlivem rozdělování zvukové energie na vzrůstající plochu [9]. K útlumu zvuku dochází také, pokud se šíří nehomogenním prostředím. Pokud zvuk prochází uzavřenou místností, tak určitý vliv na celkový útlum zvuku má dopad zvukových vln na překážky. Při dopadu zvukové vlny na překážku se část její energie odrazí, část energie je pohlcena překážkou a část projde skrz. Pohlcování zvukové energie překážkou je závislé na vlnové délce zvukových vln a na vlastnostech překážky. Vlastnosti překážky mohou být ovlivněny např. materiálem ze kterého je vyrobena. Obecně platí, že pokud je vlnová délka vlny menší než tloušťka překážky, vlna se odrazí. Pokud je vlnová délka větší, dochází k průchodu vlny překážkou. Útlum je přímo závislý na frekvenci vlnění a je tak různý pro jiné frekvence.

Difuzní odraz

Difuzní odraz (rozptyl) odražené vlny je způsoben drsností povrchu a odporem povrchu. Odražený zvukový signál se skládá z přímo odražené vlny a difuzních odrazů. Difuzní odraz je určitým způsobem zdeformována odražená zvuková vlna, která má jistou část intenzity původní vlny. Ilustrace difuzního odrazu a dalších jevů viz obrázek 2.3.

Dozvuk

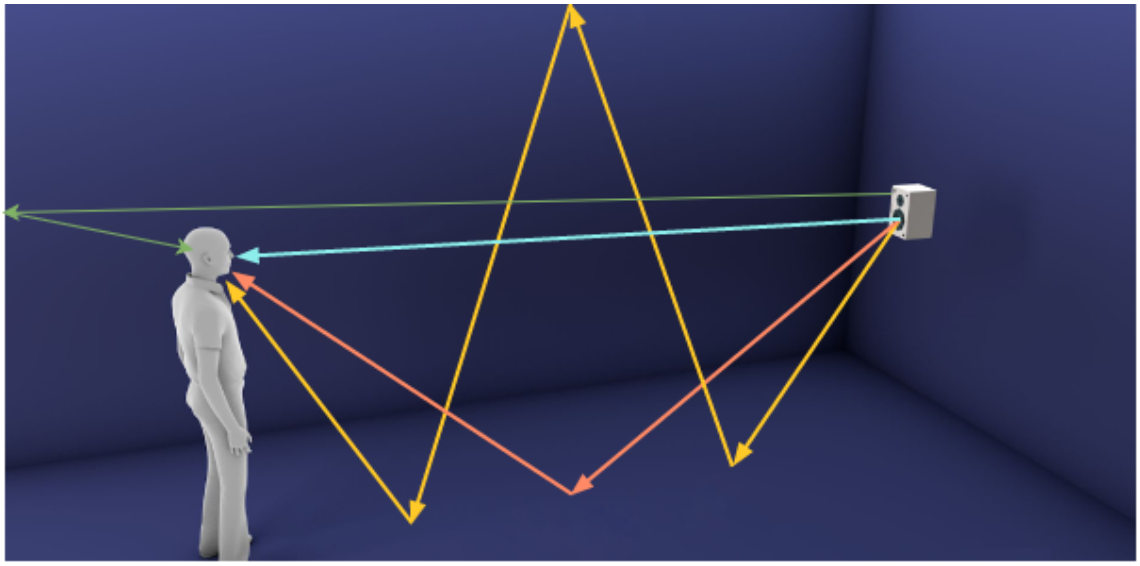
Pokud se zvuk šiří v uzavřeném prostoru vzniká tzv. dozvuk. Zvuková vlna je po dopadu na překážku částečně pohlcena a částečně se odrazí. Dochází k interferenci vlnění odraženého od stěny a vlnění postupujícího od zdroje zvuku a zvyšuje se tak celková hladina intenzity zvuku v prostoru [8]. Charakter dozvuku je určen typem prostředí, např. jeho velikostí či materiálovou nebo tvarovou skladbou prostředí.

Difrakce

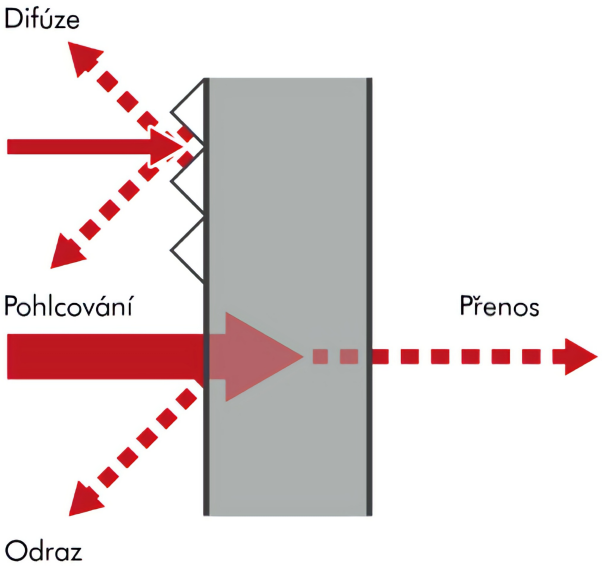
Difrakce (ohyb) je jev, kdy se vlnění za překážkou „ohýbá“ od svého původního směru vlnění a dostává se do oblasti geometrického stínu. Difrakce závisí na vlnové délce zvuku a na velikosti překážky. Pokud je překážka, na kterou narazí zvuková vlna podstatně menší než vlnová délka zvuku, zvuková vlna projde kolem překážky bez větších problémů. Pokud je překážka naopak podstatně větší než vlnová délka, zvuková vlna neprojde kolem překážky. Více informací viz [12].

Impulzní odezva místnosti

Přenos zvuku mezi dvěma fyzickými body v místnosti lze popsat impulzní odezvou. Musí ale platit, že místnost je stabilní lineární časově invariantní systém. Délka impulzní odezvy pro určitou místnost je závislá na standardní době dozvuku. Standardní doba dozvuku neboli T_{60} je taková doba, za kterou klesne hladina intenzity zvuku v dané místnosti o 60 dB. Doba dozvuku v místnostech jako jsou kanceláře, učebny, či restaurace se pohybuje mezi 0.4 až 1 sekundou. Doba dozvuku velkých koncertních sálů může přesáhnout i 2 sekundy.



Obrázek 2.2: Ilustrace odrazu zvuku v místnosti, modrá značí přímý zvuk, červená, zelená a žlutá ilustrují vícenásobné odrazy.
Obrázek převzat s úpravami z https://accusonus.com.



Obrázek 2.3: Ilustrace jevů, které mohou nastat, pokud zvuk narazí na pevnou překážku.
Obrázek převzat z https://www.paroc.cz.

2.2 Základy zpracování signálu

Audio signály

Slyšitelný zvuk vzniká z tlakových změn vzduchu dopadajícího na ušní bubínek. Zvuk zachycený mikrofonem je časový průběh změn tlaku vzduchu v místě, kde je umístěn mikrofon. Digitální audio signál je získán vhodným vzorkováním a kvantizací elektrického výstupu mikrofonu. Komponenta, která zajišťuje převod do digitálního audio signálu se nazývá analogově digitální převodník.

Analýza audio signálů

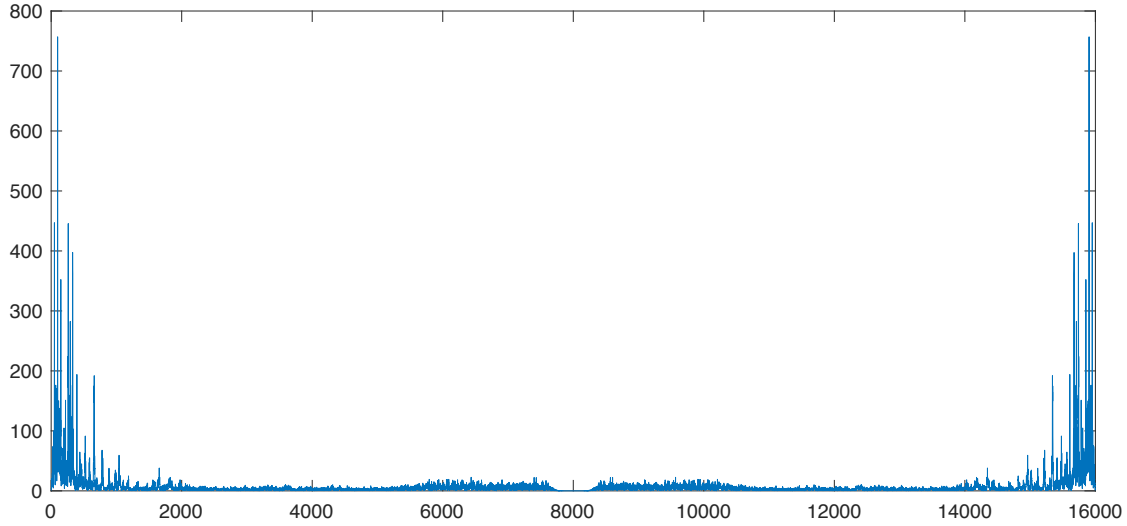
Audio signály jsou obecně nestacionární. Při analýze audio signálů se ovšem předpokládá, že audio signály se mění poměrně pomalu v čase. Proto je signál typicky rozdělen na úseky (rámce). Délka rámců se zvolí tak, že budou dostatečné malé aby byl signál v daném úseku stacionární, ale přitom dostatečně velké aby bylo možné odhadnout požadované parametry. Akustické vlastnosti zvukových událostí lze vizualizovat v časově frekvenčním „obrazu“. Lidské sluchové vnímání začíná frekvenční analýzou zvuku ve vnitřním uchu. Proto je časově frekvenční reprezentace zvuku přirozeným výchozím bodem pro strojovou klasifikaci [7]. V této práci bude použita frekvenční reprezentace signálu pro odhad impulzní odezvy místnosti.

Frekvenční spektrum signálu

Vztah mezi časovou a frekvenční oblastí signálu lze popsat pomocí Fourierovy transformace. Fourierovou transformací se dá vyjádřit jakýkoliv signál jako vážený součet harmonických průběhů (funkce sinus a cosinus). Jelikož audio signály jsou vzorkované signály, bude zde představena diskrétní Fourierova transformace (DFT).

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}, \tag{2.5}$$

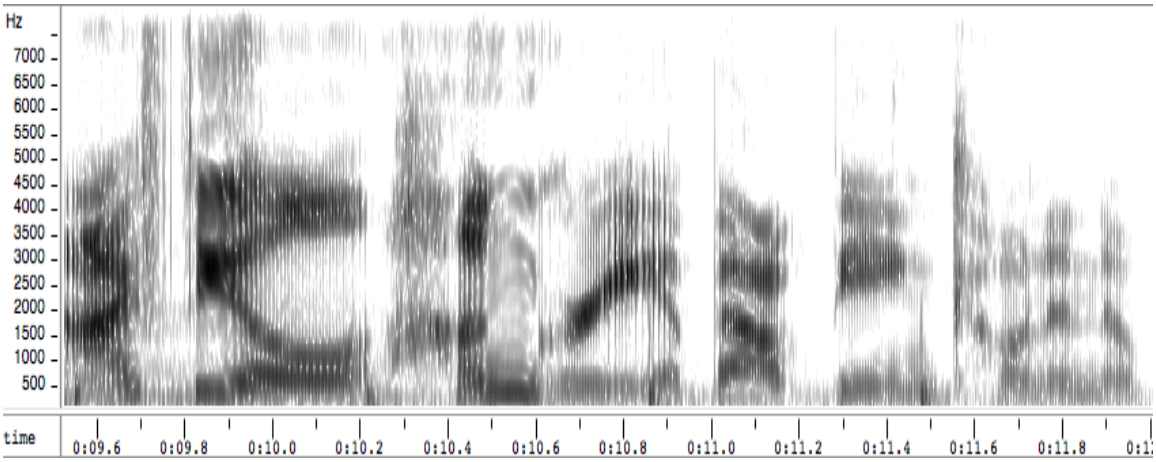
kde k nabývá hodnot 0 až $N-1$. Výsledek $X[k]$ je komplexní číslo, které udává zastoupení frekvence pro index k v analyzovaném signálu. Absolutní hodnota výsledku $|X[k]|$ udává amplitudu a úhel $arg\ X[k]$ udává časový posun. Z výsledků pro všechna k se bere pouze první polovina, jelikož druhá polovina je symetrická. Platí Shannonův teorém. Shannonův teorém říká, že je potřeba mít vzorkovací frekvenci vyšší než dvojnásobek nejvyšší harmonické složky vzorkovaného signálu. Fourierova transformace je poměrně výpočetně náročná operace. Zrychlení výpočtu lze realizovat rychlou Fourierovou transformací (FFT). FFT pracuje nejrychleji, pokud je počet vzorků zpracovávaného signálu mocninou dvou. Existuje i inverzní Fourierova transformace, která převede signál z frekvenční domény zpět do časové. Z koeficientů Fourierovy transformace dopočítá hodnoty signálu v čase. Na obrázku 2.4 je ukázka diskrétní Fourierovy transformace na jedné sekundě audio signálu.



Obrázek 2.4: Ukázka diskrétní Fourierovy transformace na jedné sekundě audio signálu. Na ose x je frekvence v Hz a na ose y je amplituda pro jednotlivé frekvence. Na obrázku je vidět že druhá polovina DFT je symetrická.

Spektogram

Spektrogram je vizuální reprezentace frekvencí obsažených v signálu, jak se mění s časem. Spektrální analýza signálu je získána jeho Fourierovou transformací. Podstatou spektrální analýzy je zjistit, nakolik jsou určité frekvence zastoupeny v analyzovaném signálu. Fourierova transformace produkuje funkci komplexních hodnot, ze kterých lze získat amplitudové spektrum a fázové spektrum. Pro sledování časově proměnných charakteristik signálu je aplikována Fourierova transformace na analyzovaný signál, postupně po krátkých úsecích. Tyto úseky se vybírají pomocí reálného symetrického okna (např. Hannovo okno). Krátkodobé fázové spektrum není považováno za významné a proto se vynechává při reprezentaci signálu [7].



Obrázek 2.5: Ukázka spektrogramu, cca. 3 sekundy lidské řeči. Mluvila žena a říkala „I feel hungry about being Conan“.

.

Konvoluce

Konvoluce je matematická operace vyjadřující vztah mezi vstupem a výstupem lineárně časově invariantního (LTI) systému. Může být spojitá či diskrétní. LTI systémy lze kompletně charakterizovat pomocí impulzní odezvy. Výstup LTI systému lze tedy určit ze vstupu a impulzní odezvy:

$$y[n] = x[n] * h[n], \tag{2.6}$$

kde y je výstupní signál, x je vstupní signál, $*$ značí konvoluci a h je impulzní odezva. Při diskrétní konvoluci jsou jednotlivé výstupy dané jako součet předchozích váhovaných vstupů. Výstupní signál lze vypočítat jako

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k]. \tag{2.7}$$

Korelace

Korelace, resp. vzájemná korelace (cross-correlation), je operátorem v oblasti zpracování signálu. Korelace určuje vzájemnou podobnost dvou signálů. Výstupem korelace dvou signálů jsou korelační koeficienty, které udávají podobnost signálů pro jednotlivé posuny. Autokorelace je operace, kdy je korelován signál sám se sebou. Největší hodnota autokorelace bude vždy pro posun roven 0. Pro diskrétní signály je korelace definována jako

$$R[k] = \sum_{n=-\infty}^{\infty} x[n]m[n-k], \tag{2.8}$$

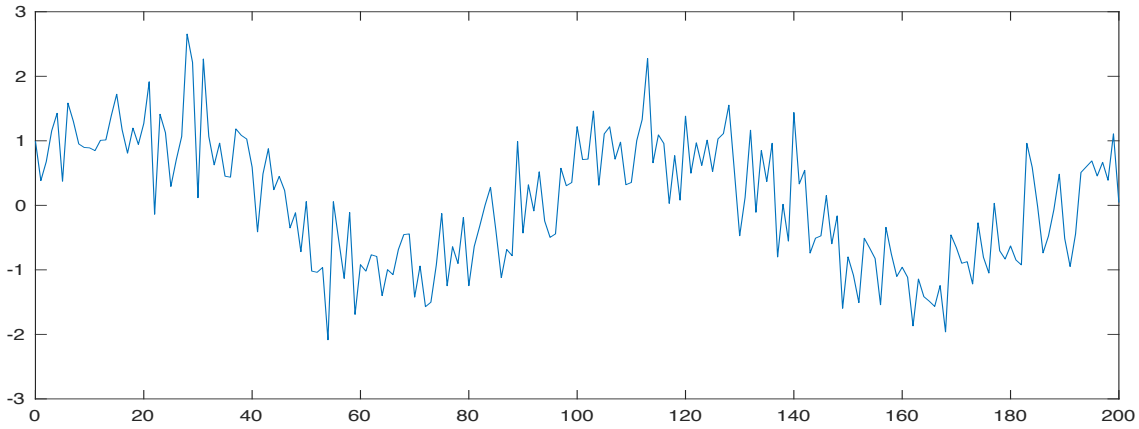
kde x a m jsou signály které se korelují. Normalizovaná korelace signálu x a m se vypočítá jako

$$R_{norm}[k] = \frac{R_{xm}}{\sqrt{E_x E_m}}, \tag{2.9}$$

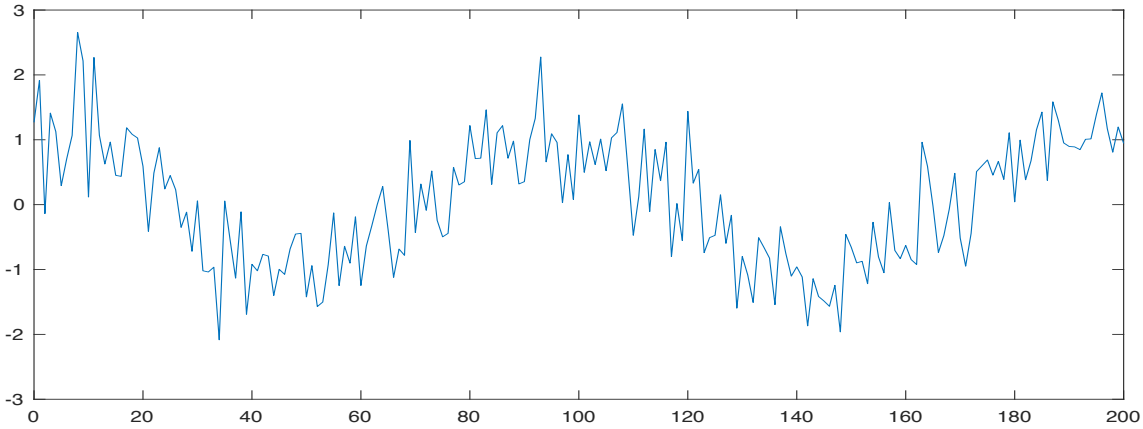
kde E_x a E_m značí energie signálu x a m, a R_{xm} značí korelaci signálu x a m. Korelace bude normalizována tak, že autokorelace s posunem 0 bude rovna 1. Energie signálu m se vypočítá jako

$$E = \sum_{n=-\infty}^{\infty} m^2[n]. \tag{2.10}$$

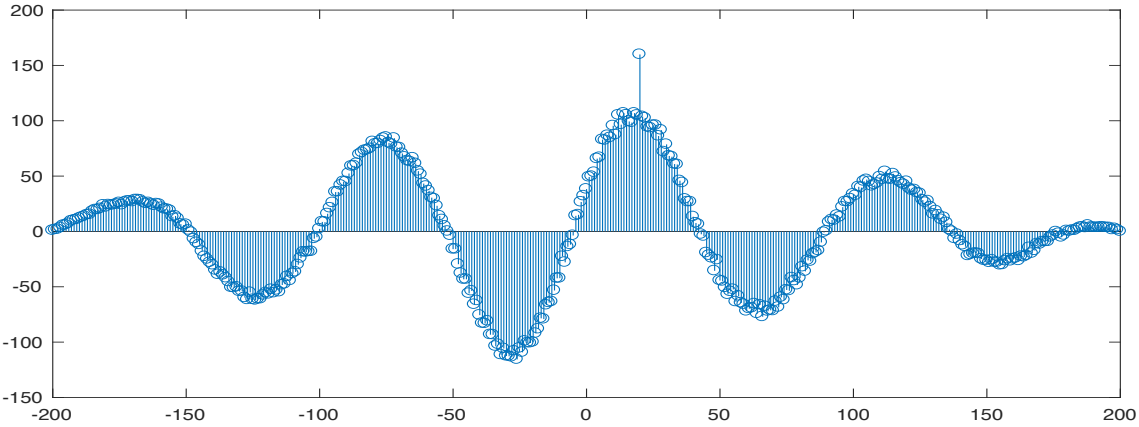
Na obrázcích 2.9 je ilustrován případ možného praktického využití korelace. Na prvním obrázku je původní signál. Na druhém obrázku je původní signál, který byl posunut o 20 vzorků. Cílem bylo najít vzájemné posunutí. To bylo nalezeno pomocí korelace. V korelačních koeficientech lze vidět jedno maximum. Toto maximum odpovídá koeficientu korelace pro posun 20. Což odpovídá vzájemnému posunutí obou signálů.



Obrázek 2.6: Původní signál.



Obrázek 2.7: Původní signál, který byl posunut o 20 vzorků.



Obrázek 2.8: Korelační koeficienty.

Obrázek 2.9: Nalezení vzájemného posunu dvou signálů pomocí korelace. Korelační koeficient je maximální pro hodnotu 20, což odpovídá vzájemnému posunutí obou signálů.

2.3 Identifikace systému

Pod pojmem identifikace systému se rozumí vytváření matematických modelů systému z naměřených dat. Existují tři přístupy pro identifikaci systému.

- white box model (tzv. bílá skříňka)
- grey box model (tzv. šedá skříňka)
- black box model (tzv. černá skříňka)

Tyto přístupy se liší na základě toho, jak moc vycházejí z matematicko-fyzikálních údajů identifikovaného systému.

White box model

Matematický model se sestavuje na základě matematicko-fyzikální analýzy objektu. Technologické, konstrukční a provozní údaje zkoumaného systému jsou výchozím bodem pro identifikaci. Vlastnosti systému se matematicky popisují dle fyzikálních, chemických a dalších zákonů. Matematickým popisem se získávají vztahy mezi sledovanými veličinami systému. Pomocí matematických rovnic (např. rovnice kontinuity) lze stanovit vztahy mezi vstupními a výstupními veličinami systému. Potom lze těmito vztahy vyjádřit vnitřní popis systému [11].

Black box model

Model vyjadřuje systém pomocí jeho vstupně-výstupního chování. Sbírají se informace vyšetřovaného systému jeho pozorováním v normálním provozu nebo při vhodně zvoleném experimentu. Matematický model systému je sestaven rozborem průběhů vstupních a výstupních veličin. Model vyjadřuje systém pomocí jeho vstupně-výstupního chování. Neumožňuje avšak pohled do jeho vnitřní struktury [11]. Více informací o tomto přístupu lze nalézt v [6].

Grey box model

Model kombinuje přístup black box modelu a white box modelu. Kombinuje tedy částečnou znalost vnitřního popisu systému a jeho vstupně-výstupního chování pro identifikaci.

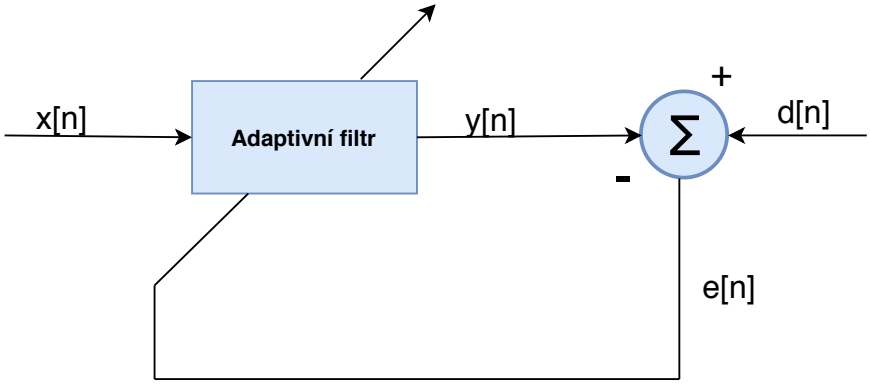
Odhad impulzní odezvy prostředí

Prvotní předpoklad je, že prostředí ve kterém je nahrávka pořízena, je lineárně časově invariantní systém. Tento systém lze potom kompletně charakterizovat pomocí impulzní odezvy. V této práci bude potřeba odhadnout impulzní odezvu prostředí. Známý je pouze rušivý signál. Nejsou k dispozici údaje o prostředí, ve které byla nahrávka vytvořena. Bude tedy použit přístup černé skříňky.

2.4 Adaptivní filtrování

Adaptivní filtr je digitální filtr, který je schopen se sám nastavovat. Automaticky upravuje své koeficienty filtru v závislosti na měnících se charakteristikách signálu. Při úpravě svých koeficientů se snaží minimalizovat tzv. chybovou funkci. Tato chybová funkce je vzdálenost mezi referenčním signálem a výstupem adaptivního filtru. Na obrázku 2.10 je blokové schéma základního adaptivního systému. Adaptivní filtrování se používá v aplikacích, které řeší tyto tři obecné problémy [10]:

- Odstranění šumu
- Odhad průběhu signálu
- Identifikace neznámého systému



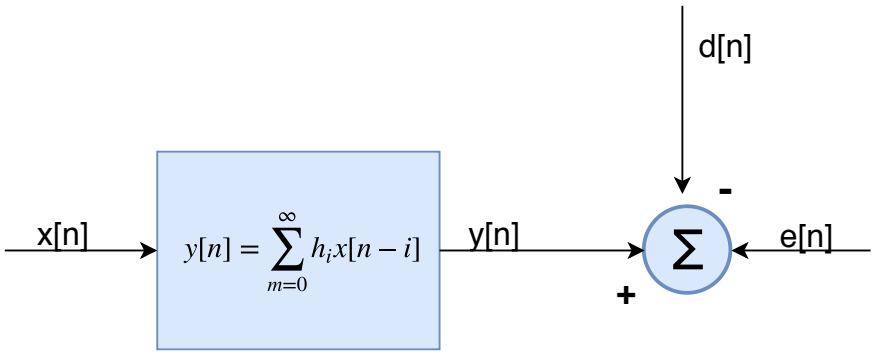
Obrázek 2.10: Blokové schéma adaptivního systému. Signál$x[n]$ značí vstupní signál, $y[n]$ je výstupní signál, $d[n]$ je referenční signál a $e[n]$ je chybový signál.

Mezi nejznámější adaptivní filtry patří LMS filter či RLS filtr. LMS (Least mean squares) filtr hledá požadované koeficienty filtru pomocí minimalizace střední kvadratické chyby. Více informací lze nalézt v [3]. RLS (Recursive least squares) filtr se oproti LMS snaží minimalizovat celou kvadratickou chybu. Více informací lze nalézt v [3] či [10]. Řešení problému pomocí LMS filtru s konečnou impulzní odezvou konverguje k optimálnímu Wienerovu filtru.

LMS filtr byl testován v této práci na odstranění šumu a pro odhad impulzní odezvy. Vstupním signálem $x[n]$ je známý rušivý signál. Při odhadu impulzní odezvy je referenčním signálem $d[n]$, známý rušivý signál ze zarušené nahrávky a výsledkem jsou koeficienty filtru. Při odstranění šumu je referenčním signálem zarušená nahrávka a výsledkem je chybový signál $e[n]$, který by měl obsahovat pouze užitečný signál.

2.5 Diskrétní Wienerův filtr

Diskrétní Wienerův filtr je založen na statistickém přístupu a stejně jako u LMS adaptivního filtru se snaží minimalizovat střední kvadratickou chybu. Pro nalezení optimálních koeficientů používá autokorelaci vstupního signálu a vzájemnou korelaci vstupního a referenčního signálu. Více informací ohledně Wienerova filtru lze nalézt v [3] či [4].



Obrázek 2.11: Blokový diagram diskrétního Wienerova filtru. Vstupní signál $x[n]$ je konvoluován s Wienerovým filtrem. Výstup $y[n]$ je pak porovnán vůči referenčnímu signálu $d[n]$, čímž vznikne chybový signál $e[n]$.

Wiener-Hopf rovnice

Tyto rovnice vycházejí z korelace vstupního signálu a vzájemné korelace vstupního a referenčního signálu. Za pomocí těchto rovnic lze odhadnout impulzní odezvu systému ze znalosti vstupu a výstupu systému. Výstupní signál systému je považován za referenční. Vstupní matice $\mathbf{X}$ je naplněna autokorelacemi vstupního signálu. Výstupní vektor $\mathbf{y}$ je naplněn vzájemnou korelací vstupního a výstupního signálu:

$$\underbrace{\begin{pmatrix} R_{xy}[0] \\ R_{xy}[1] \\ \vdots \\ R_{xy}[l] \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} R_{xx}[0] & R_{xx}[-1] & \dots & R_{xx}[-l] \\ R_{xx}[1] & R_{xx}[0] & \dots & R_{xx}[1-l] \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}[l] & R_{xx}[l-1] & \dots & R_{xx}[0] \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} h[0] \\ h[1] \\ \vdots \\ h[l] \end{pmatrix}}_{\mathbf{h}}, \tag{2.11}$$

kde R_{xx} značí autokorelaci vstupního signálu, R_{xy} značí vzájemnou korelaci vstupního a výstupního signálu, l značí délku impulzní odezvy a h je impulzní odezva systému. Impulzní odezva se pak vypočítá jako

$$\mathbf{h} = \mathbf{X}^{-1}\mathbf{y} \tag{2.12}$$

Wiener-Hopf rovnice se používají v této práci pro odhadování impulzní odezvy. Vynikají svojí výpočetní rychlostí.

Kapitola 3

Data

Data, se kterými se pracuje v této práci, jsou různé audio nahrávky. Tato data se využívají pro testování výsledného programu či jeho jednotlivých součástí. Jednotlivé nahrávky se liší v závislosti na jejich obsahu.

- Jaký rušivý signál obsahují?
- Jaký užitečný signál obsahují?
- V jakém prostředí jsou nahrávané?

Celkově se nahrálo přes 2 hodiny nahrávek. Užitečným signálem je lidská řeč. Jako nahrávací zařízení byl použit přenosný rekordér značky Zoom. Všechny nahrávky mají vzorkovací frekvenci 16000 Hz. Tato vzorkovací frekvence byla zvolena jako kompromis mezi kvalitou nahrávky a výpočetními požadavky pro její zpracování.

3.1 Rušivý signál v nahrávkách

Byly vybrány tři typy rušivého signálu.

- klasická hudba - neobsahuje hlas, pouze melodii
- pop, rock, rap - obsahuje melodii i hlas
- talk show - obsahuje hlavně mluvené slovo

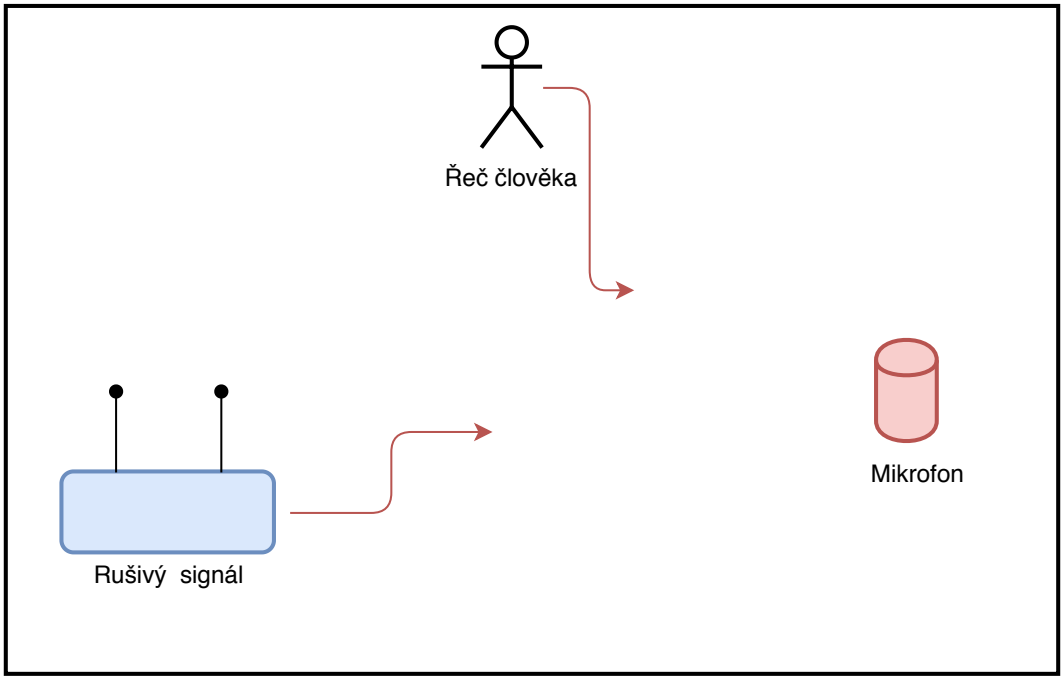
Tyto typy byly zvoleny na základě toho, že užitečným signálem je lidská řeč. Bude tak možnost otestovat odečítaní rušivého signálu pro tři různé složitosti úloh. Jednoduché úlohy jsou nahrávky, které obsahují jako rušivý signál klasickou hudbu. Klasická hudba má jiný charakter než užitečný signál. Středně složité úlohy jsou nahrávky, které obsahují jako rušivý signál pop rock či rap. Hudba, kde se zpívá, má z části podobný charakter jako užitečný signál. Složité úlohy jsou nahrávky, které obsahují jako rušivý signál talk show. Talk show má stejný charakter jako užitečný signál.

Ze žánru klasické hudby byly jako rušivé signály vybrány skladby převážně od Bedřicha Smetany (Šárka, Tábor, Vltava). Z popu a rocku byly vybrány známé skladby, jako je např. skladba We're Not Gonna Take It od Twisted Sister. Z rapu byla vybrána skladba Rap God od Eminema. Jako talk show byla vybrána část audio nahrávky z pořadu Conan O'Brien Needs a Friend.

3.2 Nahrávky s reálným mluvčím

V těchto nahrávkách je užitečným signálem řeč reálných osob, které jsou přítomny v místnosti. Tyto nahrávky slouží k testování jednotlivých metod při vývoji programu. Vyhodnocení probíhá sluchem, jelikož není k dispozici nezarušený užitečný signál pro objektivní vyhodnocení. Bylo nahráno celkem 10 nahrávek o průměrné délce cca. 3 minuty. Nahrávalo se ve dvou místnostech (středně velká kancelář, velká místnost). Kancelář měla velikost cca 40 m^3 a byla vybavena běžným kancelářským nábytkem. Velká místnost měla cca 70 m^3 a byla vybavena běžným nábytkem. Jako rušivé signály sloužily všechny typy popsané výše.

Zdrojem užitečného signálu je člověk, který předčítá knihu, či rozhovor dvou lidí. Zdrojem rušení je reproduktor, který přehrává rušivý signál. Člověk je blíže mikrofonu než reproduktor. Ilustrace viz obrázek 3.1



Obrázek 3.1: Schéma místnosti při nahrávaní dat s reálným mluvčím.

3.3 Simulované nahrávky

Jako užitečný signál v těchto nahrávkách je lidská řeč, která je přehrávána pomocí reproduktoru. Audio nahrávky lidské řeči byly získány z databáze SpeechDat-E[1]. Vyhodnocení testování programu pomocí těchto nahrávek lze ověřit sluchem či objektivně. Pro objektivní vyhodnocení musí být mimo jiné nahrán pouze užitečný signál. Byl také nahrán pouze známý rušivý signál. Nahrávky, které obsahují pouze rušivý signál, se využijí pro kvalitní odhad impulzní odezvy. Průměrná délka nahrávek obsahující pouze rušivý signál je cca 8 minut. Průměrná délka zarušených nahrávek a nahrávek obsahující pouze užitečný signál je cca 10 minut.

[1] http://www.fee.vutbr.cz/SPEECHDAT-E/

Nahrávalo se ve středně velké kanceláři viz. obrázek 3.2. Kancelář měla velikost cca 40 m^3 a byla vybavena běžným kancelářským nábytkem. Jako rušivé signály sloužily všechny typy popsané výše. Průměrné SNR těchto nahrávek bylo cca -2.2 dB. Což znamená, že energie rušivého signálu byla o něco málo větší, než energie užitečného signálu.

Vytvoření nahrávky probíhalo podobně jako v sekci 3.2. Avšak místo reálné osoby byl v místnosti další reproduktor, který přehrával lidskou řeč. V tabulce 3.1 jsou počty všech vytvořených nahrávek.

Typ nahrávky	**Klasická hudba**	**Pop,rock,rap**	**Talk Show**
S reálným mluvčím	4	3	3
Simulované	1	2	1

Tabulka 3.1: Přehled všech vytvořených nahrávek v závislosti na typu nahrávky a žánru rušivého signálu.

Obrázek 3.2: Fotografie kanceláře ve které se nahrávalo.

3.4 Způsoby vyhodnocení kvality zvukových signálů

Při vývoji jednotlivých metod pro odstranění známého signálu z nahrávky je potřeba vyvíjené metody určitým způsobem otestovat. Testování metod probíhá většinou porovnáním kvality odečtených nahrávek.

Vyhodnocení kvality zvukových dat je poměrně problematická záležitost a není definována jedna technika, která by zaručovala úspěšné vyhodnocení. Metody se dělí na objektivní,

kde se kvalita vyhodnocuje algoritmem, a subjektivní, kde je kvalita hodnocena lidskými posluchači. Více informaci viz [2].

Jako objektivní metodu pro vyhodnocení kvality jsem zvolil logaritmickou spektrální vzdálenost. Měří spektrální vzdálenost (v decibelech) mezi dvěma signály. Je definována jako

$$D_{LS} = \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[10\log_{10}\frac{P(\omega)}{P'(\omega)}\right]d\omega}, \tag{3.1}$$

kde $P(\omega)$ a $P'(\omega)$ jsou výkonová spektra porovnávaných signálů. Logaritmická spektrální vzdálenost se počítá mezi jednotlivými rámci porovnávaných signálů. Vzhledem ke vzorkovací frekvenci 16000 Hz se zvolila velikost rámců na 320 vzorků. Výsledná spektrální logaritmická vzdálenost je pak dána jako průměr všech vzdáleností. Logaritmickou spektrální vzdálenost lze použít při testování metod na simulovaných nahrávkách. K těmto nahrávkám je k dispozici užitečný signál, který byl nahrán ve stejné místnosti jako zarušená nahrávka. Vypočte se logaritmická spektrální vzdálenost mezi užitečným signálem a odečtenou nahrávkou.

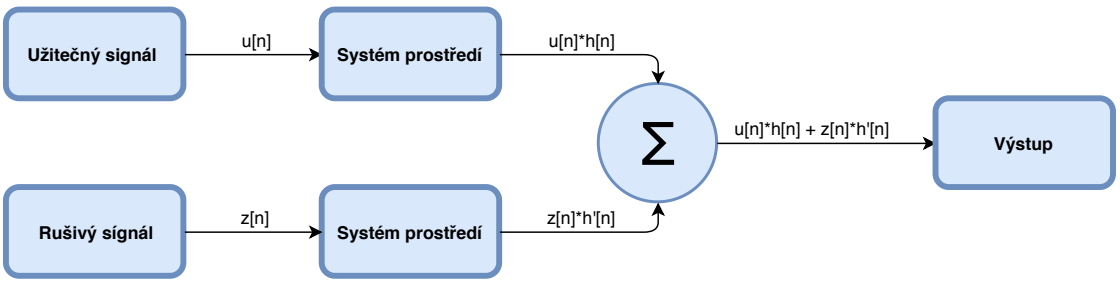
Pro určení kvality odečtené nahrávky se používá také subjektivní ohodnocení. Kvalita odečtené nahrávky se určuje poslechem. Při poslechu se zaměřuji na 3 hlavní rysy nahrávky.

- čistota užitečného signálu
- míra potlačení rušivého signálu v nahrávce
- přítomnost „artefaktů“ v nahrávce

Kapitola 4

Vývoj metody

Cílem této práce je odstranění známého signálu z nahrávky. Problém je podrobně rozebrán v kapitole 2. Ilustrace problému viz obrázek 4.1. Implementace metod rozebraných v této kapitole je realizována v programovacím jazyce Matlab.



Obrázek 4.1: Průchod signálu místností.

4.1 Odhad impulzní odezvy

Výstupní signál lze vypočítat jako

$$z'[n] = z[n] * h'[n]. \tag{4.1}$$

kde z' je výstupní signál, z je vstupní signál a h' je impulzní odezva prostředí. Při odhadovaní impulzní odezvy lze tedy vycházet z rovnice 4.1 a rovnice konvoluce.

$$z'[n] = \sum_{k=-\infty}^{\infty} z[k]h'[n-k]. \tag{4.2}$$

Aby bylo možné impulzní odezvu odhadnout, je potřeba znát vstup a výstup. Nechť je vstupním signálem z známý rušivý signál. Impulzní odezva h' je odhadovaná impulzní odezva prostředí. Potom výstup z' bude odpovídat známému rušivému signálu z, který je pozměněn impulzní odezvou prostředí. Tento výstupní signál z' odpovídá úsekům v zarušené nahrávce, kde užitečný signál (např. mluva) není obsažen. Pokud se impulzní odezva nebude v nahrávce s časem měnit, lze ji odhadnout na úsecích nahrávky, kde užitečný signál není obsažen. Tyto úseky je třeba v nahrávce najít, postup hledání viz sekce 4.7.

4.2 Odhadnutí impulzní odezvy v časové oblasti

Jako výstupní signál budou označeny úseky v nahrávce, které obsahují pouze známý rušivý signál. Vstupní signál bude poté odpovídat stejným úsekům v známém rušivém signálu. Ze vstupních vzorků se vytvoří tzn. levá strana pro odhad. Z výstupních vzorků se vytvoří tzn. pravá strana pro odhad.

$$\mathbf{L} = \begin{pmatrix} x[l] & x[l-1] & \dots & x[1] \\ x[l+1] & x[l] & \dots & x[2] \\ \vdots & \vdots & \ddots & \vdots \\ x[p] & x[p-1] & \dots & x[p-l+1] \end{pmatrix} \mathbf{p} = \begin{pmatrix} y[l] \\ y[l+1] \\ \vdots \\ y[p] \end{pmatrix}, \tag{4.3}$$

kde $\mathbf{L}$ je levá strana pro odhad, $\mathbf{p}$ je pravá strana pro odhad, l je maximální délka odhadované impulzní odezvy, x značí vstupní signál, y značí výstupní signál a p je počet všech vzorků. Výstupní signál lze maticově vypočítat jako

$$\mathbf{p} = \mathbf{L}\ \mathbf{h}', \tag{4.4}$$

kde $\mathbf{h}'$ je impulzní odezva o délce l. Odhadovaná impulzní odezva se potom vypočte jako

$$\mathbf{h}' = \mathbf{L}^{+}\mathbf{p}, \tag{4.5}$$

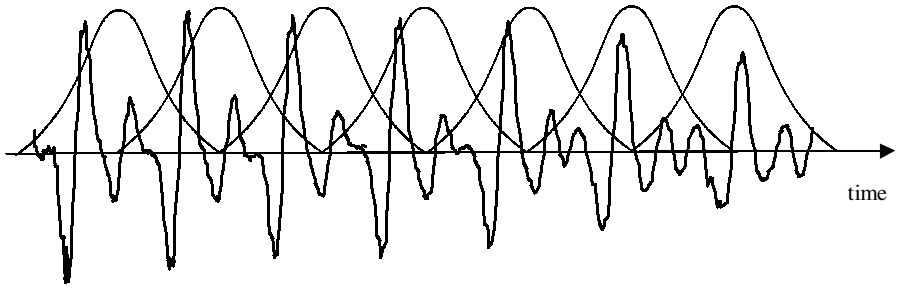
kde $\mathbf{L}^{+}$ značí pseudoinverzi matice $\mathbf{L}$. Odhadování impulzních odezev v časové oblasti není příliš prakticky využitelné. Doba výpočtu pro cca 3 minutovou nahrávku je extrémně dlouhá a přesnost odhadu impulzní odezvy je také velmi špatná. Přesnější a prakticky využitelné bude odhadování impulzní odezvy ve frekvenční oblasti.

4.3 Odhadnutí impulzní odezvy ve frekvenční oblasti

Vstupní signál (známý rušivý signál) i výstupní signál (vstupní signál pozměněn impulzní odezvou prostředí) se rozdělí na rámce, které se budou překrývat. Velikost jednotlivých rámců se zvolí na cca 30 ms. Na každý rámec je aplikováno Hannovo okno definováno jako

$$w[n] = 0.5\Big(1 - \cos\Big(2\pi\frac{n}{N}\Big)\Big), \tag{4.6}$$

kde N je počet vzorků, n nabývá hodnot 0 až N a délka okna L je $N + 1$. Ilustrace viz obrázek 4.2.



Obrázek 4.2: Ilustrace procesu rámcování, překrývání a váhovaní signálu
Obrázek převzat z [10].

Každý rámec se převede z časové oblasti do frekvenční pomocí diskrétní Fourierovy transformace. Jednotlivé rámce vstupního a výstupního signálu, které jsou již ve frekvenční oblasti, se naskládají za sebe do matic. Vzniknou dvě časově-frekvenční matice, jedna obsahuje rámce vstupu a druhá obsahuje rámce výstupu. Matice pro vstup a výstup vypadají následovně.

$$\mathbf{X} = \begin{pmatrix} x_0(1) & x_0(t) & \dots & x_0(d) \\ x_1(1) & x_1(t) & \dots & x_1(d) \\ x_k(1) & x_k(t) & \dots & x_k(d) \\ \vdots & \vdots & \ddots & \vdots \\ x_{\frac{P}{2}}(1) & x_{\frac{P}{2}}(t) & \dots & x_{\frac{P}{2}}(d) \end{pmatrix} \mathbf{Y} = \begin{pmatrix} y_0(1) & y_0(t) & \dots & y_0(d) \\ y_1(1) & y_1(t) & \dots & y_1(d) \\ y_k(1) & y_k(t) & \dots & y_k(d) \\ \vdots & \vdots & \ddots & \vdots \\ y_{\frac{P}{2}}(1) & y_{\frac{P}{2}}(t) & \dots & y_{\frac{P}{2}}(d) \end{pmatrix} \tag{4.7}$$

$\mathbf{X}$ značí vstupní matici, $\mathbf{Y}$ značí výstupní matici, t je čas v rámcích, $\frac{P}{2}$ je polovina počtu vzorků DFT a d značí počet všech rámců v signálu. Jednotlivé sloupce lze značit vektorově jako $\mathbf{x}(t)$, nebo $\mathbf{y}(t)$. Předpokládá se, že každý řádek $\mathbf{y}_k$ matice $\mathbf{Y}$ je dán konvolucí jako

$$\mathbf{y}_k = \mathbf{x}_k * \mathbf{h}'_k, \tag{4.8}$$

kde $\mathbf{h}'_k$ je impulzní odezva pro spektrální koeficient k. Pro každý koeficient k ze vstupní matice $\mathbf{X}$ se vytvoří tzn. levá strana pro odhad. Pro každý koeficient k z výstupní matice $\mathbf{Y}$ se vytvoří tzn. pravá strana pro odhad. Obě operace jsou podobné jako při počítaní v časové oblasti viz rovnice 4.3.

$$\mathbf{L}_k = \begin{pmatrix} x_k(l) & x_k(l-1) & \dots & x_k(1) \\ x_k(l+1) & x_k(l) & \dots & x_k(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_k(d) & x_k(d-1) & \dots & x_k(d-l+1) \end{pmatrix} \mathbf{p}_k = \begin{pmatrix} y_k(l) \\ y_k(l+1) \\ \vdots \\ y_k(d) \end{pmatrix}, \tag{4.9}$$

kde $\mathbf{L}_k$ je levá strana pro odhad, $\mathbf{p}_k$ je pravá strana pro odhad, l je maximální délka odhadované impulzní odezvy. Pro každý koeficient k se tedy vytvoří $\mathbf{L}_k$ a $\mathbf{p}_k$ a odhadovaná impulzní odezva se vypočte jako

$$\mathbf{h}'_k = \mathbf{L}_k^{+}\mathbf{p}_k, \tag{4.10}$$

kde $\mathbf{L}^{+}$ značí pseudoinverzi matice $\mathbf{L}$. Pro každý koeficient se odhadne impulzní odezva $\mathbf{h}'_k$. Tyto impulzní odezvy se naskládají do matice jako

$$\mathbf{H}' = \begin{pmatrix} \mathbf{h}'_0 \\ \mathbf{h}'_1 \\ \vdots \\ \mathbf{h}'_{\frac{P}{2}} \end{pmatrix} \tag{4.11}$$

kde **H**' značí celkovou impulzní odezvu ve frekvenční oblasti.

4.4 Aplikace impulzní odezvy na rušivý signál

Jakmile je k dispozici odhadnutá impulzní odezva $\mathbf{H}'$, aplikuje se na známý rušivý signál. Vytvoří se matice $\mathbf{Z}$, která bude obsahovat známý rušivý signál. Tato matice bude vytvořena stejným způsobem jako matice $\mathbf{X}$ viz rovnice 4.7. Na každý řádek matice $\mathbf{Z}$ bude aplikována impulzní odezva jako

$$\mathbf{z}'_k = \mathbf{z}_k * \mathbf{h}'_k, \tag{4.12}$$

kde $\mathbf{z}_k$ značí řádek matice $\mathbf{Z}$, $\mathbf{h}'_k$ značí impulzní odezvu pro spektrální koeficient k. Výsledkem aplikace impulzní odezvy $\mathbf{H}'$ na matici $\mathbf{Z}$ je matice $\mathbf{Z}'$, která obsahuje rušivý signál pozměněný odhadnutou impulzní odezvou ve frekvenční oblasti.

4.5 Spektrální odečítání

Spektrální odečítaní je metoda pro odečítaní signálů v jejich frekvenční oblasti. Signály jsou rozděleny do překrývajících se rámců o N vzorcích. Na každý rámec je aplikováno reálné symetrického okno (např. Hannovo okno), ilustrace viz obrázek 4.2. Poté jsou rámce transformovány pomocí Diskrétní Fourierovy transformace do frekvenční oblasti. Ve frekvenční oblasti jsou tyto rámce od sebe odečteny. Převod odečtených rámců, zpět do časové oblasti, je realizován pomocí inverzní diskrétní Fourierovy transformace. Výběr velikosti rámců je kompromis mezi požadavkem na časové a frekvenční rozlišení. Typicky je použit rámec o velikosti 5-50 ms. Více informací lze nalézt v [10]. Schéma spektrálního odečítání viz obrázek 4.3.

Cílem této práce je odečíst rušivý signál od celkové nahrávky a tím získat užitečný signál. Rušivý signál, na který již byla aplikována odhadnutá impulzní odezva, je obsažen v matici $\mathbf{Z}'$. Ze signálu zarušené nahrávky je potřeba vytvořit matici $\mathbf{C}$, která bude vytvořena stejným způsobem jako matice $\mathbf{X}$ v sekci 4.3. Tedy rozdělení signálu na rámce (30ms), překrytí, váhování, převedení do frekvenční oblasti. Spektrální odečítání může být implementováno jako odečítaní energií či magnitud ve spektrálních oblastech viz [10]. Ve výsledném řešení budou implementovány obě metody. Bude si tedy možné vybrat ten přístup, který bude u konkrétní nahrávky podávat lepší výsledky.

Spektrální odečítaní energií

Užitečný signál $\mathbf{U}$(f) je získán jako:

$$|\mathbf{U}(f)|^2 = \begin{cases} |\mathbf{C}(f)|^2 - \alpha|\mathbf{Z}'(f)|^2 & \text{jestliže } |\mathbf{C}(f)|^2 - \alpha|\mathbf{Z}'(f)|^2 > 0 \\ 0 & \text{jinak} \end{cases}, \tag{4.13}$$

kde $\mathbf{C}(f)$ je zarušený signál ve frekvenční oblasti, $\mathbf{Z}'(f)$ je známý rušivý signál po aplikaci odhadnuté impulzní odezvy prostředí, α je koeficient, který určuje množství odečítaného signálu, pro úplné odečtení $\alpha = 1$ a pro nadměrné odečtení $\alpha > 1$.

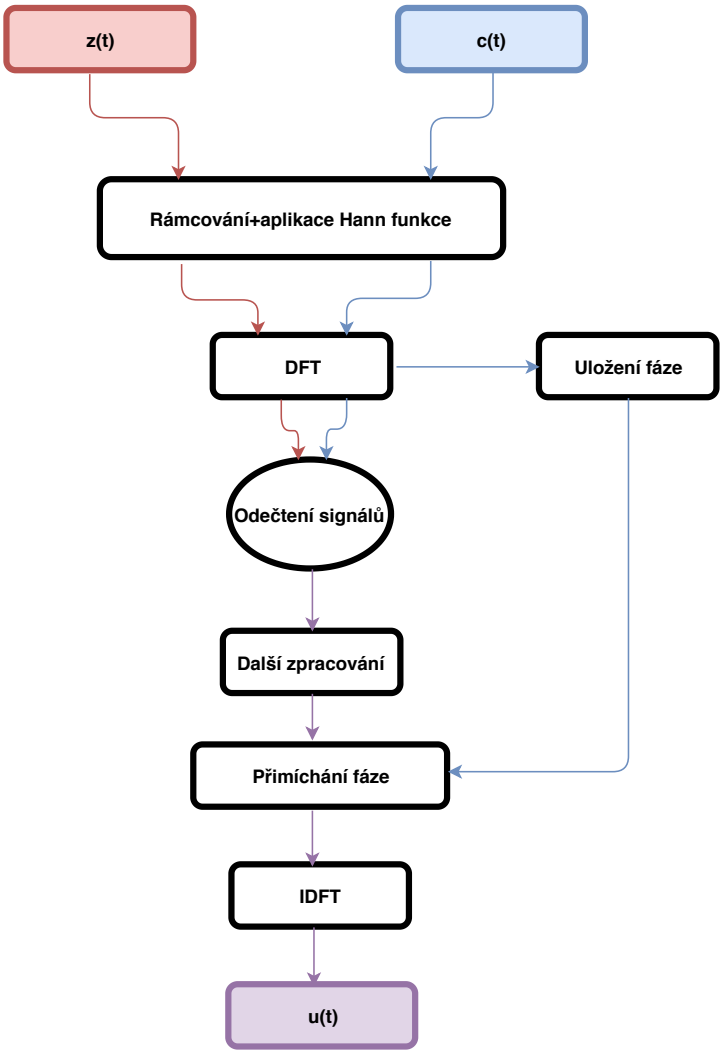
U tohoto přístupu se odečítají energie obou signálů ve frekvenční doméně. Před převedením do časové domény je nutné výsledek odmocnit. Tedy

$$|\mathbf{U}(f)| = \sqrt{|\mathbf{U}(f)|^2}. \tag{4.14}$$

Spektrální odečítaní magnitud

Užitečný signál $\mathbf{U}$(f) je získán jako:

$$|\mathbf{U}(f)| = \begin{cases} |\mathbf{C}(f)| - \alpha|\mathbf{Z}'(f)| & \text{jestliže } |\mathbf{C}(f)| - \alpha|\mathbf{Z}'(f)| > 0 \\ 0 & \text{jinak} \end{cases}, \tag{4.15}$$

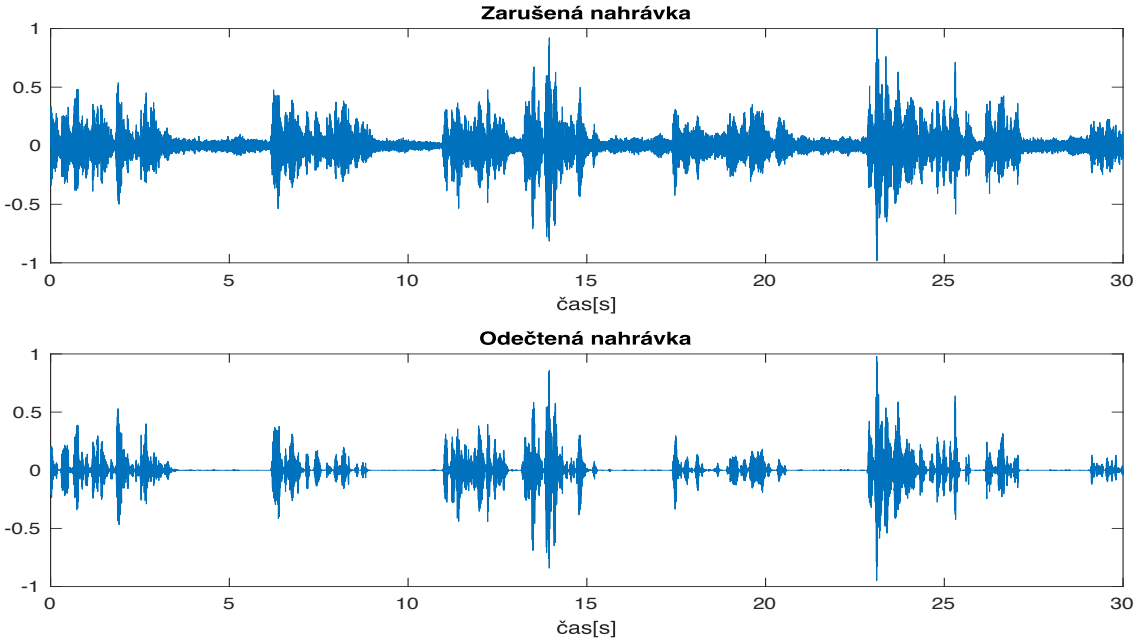


Obrázek 4.3: Schéma spektrálního odečítaní

4.6 Odhadnutí impulzní odezvy z celých signálů

Na začátku kapitoly 4 bylo zmíněno, že pro odhad impulzní odezvy prostředí je potřeba najít úseky v zarušené nahrávce, kde je obsažen pouze rušivý signál. Tyto úseky slouží jako výstupní signál při odhadu impulzní odezvy viz sekce 4.3. Při odhadu impulzní odezvy lze ale také použít jako výstup celou zarušenou nahrávku. Užitečný signál je při odhadu impulzní odezvy částečně vyrušen a odhadnutá impulzní odezva odpovídá s poměrně dobrou přesností reálné impulzní odezvě místnosti. Nicméně tato impulzní odezva je částečně zkreslená užitečným signálem, který byl obsažen v zarušené nahrávce. Na obrázku 4.4 je zobrazena zarušená nahrávka (rušivý signál klasická hudba) a odečtená nahrávka. Impulzní odezva byla odhadnuta z celých signálů. Bylo aplikováno spektrální odečítání energií a odečítací koeficient byl zvolen $\alpha = 3$. Z obrázku odečtené nahrávky lze určit úseky, kde se vyskytoval pouze rušivý signál. Tyto úseky mají téměř nulovou energii. Podařilo se tedy odečíst rušivý signál, nicméně užitečný signál byl z hlediska sluchové kvality značně deformován.

Což je způsobeno poměrně vysokou hodnotou odečítacího koeficientu a také kvůli odhadu impulzní odezvy z celých signálů.



Obrázek 4.4: Porovnání zarušené a odečtené nahrávky.

4.7 Hledání úseků obsahujících pouze rušivý signál

Pro řešení tohoto problému jsem navrhl dva různé přístupy. První přístup je založen na myšlence, že úseky v zarušené nahrávce, které obsahují pouze rušivý signál, jsou podobné odpovídajícím úsekům známého rušivého signálu. Druhý přístup je založen na myšlence, že pokud bude alespoň částečně odstraněn rušivý signál z nahrávky, tak v místech, kde byl původně rušivý signál, bude energie signálu podstatně menší. Oba přístupy byly implementovány, ale druhý vykazuje lepších výsledků, takže bude použit ve finálním řešení.

Přístup založený na korelaci signálů

Zarušená nahrávka a známý rušivý signál jsou zarovnány a rozděleny na rámce o délce trvání 1 sekundy. Tyto rámce mají překryv 0.75 sekundy. Pro každou dvojici odpovídajících rámců je vypočtena křížová korelace.

$$\mathbf{R} = \begin{pmatrix} \mathbf{r}_{cz1} \\ \mathbf{r}_{cz2} \\ \vdots \\ \mathbf{r}_{czD} \end{pmatrix}, \tag{4.16}$$

kde $\mathbf{r}_{cz1}$ značí křížovou korelaci známého rušivého signálu a zarušené nahrávky pro první rámec, D je počet všech rámců.

Z jednotlivých křížových korelací ($\mathbf{r}_{cz}$) se vybere určitý počet (např. 10) největších koeficientů v absolutní hodnotě.

$$\mathbf{R}max10 = \begin{pmatrix} \mathbf{r}max10_{cz1} \\ \mathbf{r}max10_{cz2} \\ \vdots \\ \mathbf{r}max10_{czD} \end{pmatrix}, \tag{4.17}$$

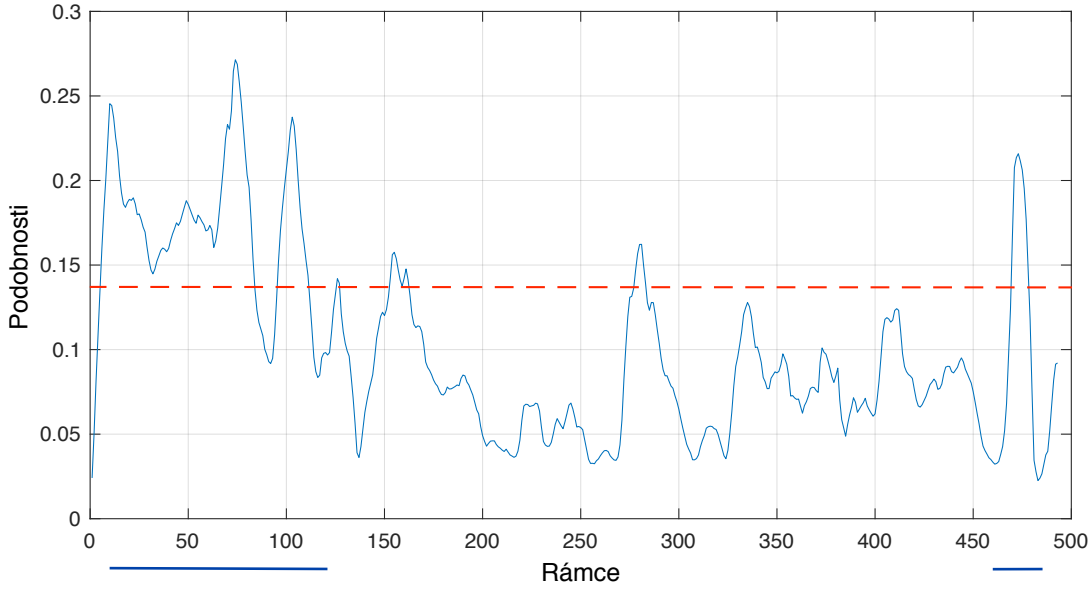
kde $\mathbf{r}max10_{cz1}$ značí 10 největších koeficientů křížové korelace známého rušivého signálu a zarušené nahrávky pro první rámec. Z těchto největších koeficientů pro jednotlivé rámce se vytvoří průměr a tyto průměry udávají podobnost rámců zarušené nahrávky a rámců rušivého signálu.

$$\mathbf{p} = \begin{pmatrix} prmax10_{cz1} \\ prmax10_{cz2} \\ \vdots \\ prmax10_{czD} \end{pmatrix}, \tag{4.18}$$

kde $prmax10_{cz1}$ značí průměr z $\mathbf{r}max10_{cz1}$ a v $\mathbf{p}$ jsou obsaženy podobnosti jednotlivých rámců zarušené nahrávky a rušivého signálu.

Hlavní myšlenkou tohoto přístupu je předpoklad, že rámce v zarušené nahrávce, kde je obsažen pouze rušivý signál a odpovídající rámce rušivého signálu, budou mít větší podobnost. Tato metoda byla testována na testovacích nahrávkách a dokázala určit cca 80 procent úseků ticha. Bohužel ale určila i některé úseky, kde byl obsažen užitečný signál. To bylo pravděpodobně způsobeno podobností užitečného signálu s rušivým. Proto tato metoda není úplně ideální. Ukázka jednotlivých podobností rámců reálné zarušené nahrávky a rušivého signálu viz obrázek 4.5.

Hodnoty na ose y značí podobnosti rámců, hodnoty na ose x jsou jednotlivé rámce. Úseky v nahrávce, kde je obsažen pouze rušivý signál, jsou cca od 15. rámce do 125. rámce a poté mezi 460. až 480. rámcem. Tyto rámce jsou v obrázku označené modrou čarou. Z obrázku lze vypozorovat, že na hledaných úsecích je podobnost větší a při volbě vhodné hodnoty podobnosti (např. 0.14), která by určovala spodní hranici hledaných úseků, by tento postup nalezl většinu úseků. Na obrázku je spodní hranice označena červenou přerušovanou čarou. Rámce které mají podobnost větší jak je spodní hranice (červená přerušovaná čára) by se považovaly za úseky kde je obsažen pouze rušivý signál. Tento postup ale nenalezl všechny hledané úseky (např. mezi cca 80. až 100. rámcem) a hlavně nalezl i úseky, kde je obsažen užitečný signál.



Obrázek 4.5: Ukázka podobnosti jednotlivých rámců. Modrá čára značí rámce, kde je obsažen pouze rušivý signál. Červená přerušovaná čára značí spodní hranici hledaných úseků. Rámce s větší hodnotou podobnosti než je spodní hranice, jsou hledané rámce.

Přístup založený na zbytkové energii po částečném odečtení

Nejprve bylo implementováno částečné odečítání rušivého signálu pomocí LMS adaptivního filtru. Toto odečítání probíhalo v časové doméně. Nicméně výsledky nebyly ideální. Rušivý signál nebyl odečten dostatečně. Lepší výsledky měl postup, kdy pro odhad impulzní odezvy byl použit celý zarušený signál viz sekce 4.6. Odhadnutá impulzní odezva byla aplikována na známý rušivý signál a tento rušivý signál byl odečten metodou spektrálního odečítaní viz schéma 4.3.

Výsledný odečtený signál se potom rozdělí na úseky o délce trvání cca 0.25 sekundy. Z těchto úseků se vypočítá energie a podle hodnoty této energie se určí, zda tento úsek obsahuje pouze rušivý signál. Energie úseků, kde byl obsažen pouze rušivý signál, by měla být podstatně menší než energie úseků, kde je obsažen i užitečný signál. Je určena horní hranice (hodnota energie). Pokud je energie rámce menší než je horní hranice, tak je brán jako úsek, kde je obsažen pouze rušivý signál.

$$\begin{cases} \text{úsek obsahuje pouze rušivý signál} & \text{jestliže } E_r < hh \\ \text{úsek obsahuje i užitečný signál} & \text{jinak} \end{cases}, \tag{4.19}$$

kde E_r je energie rámce a hh je horní hranice. Tento způsob funguje poměrně spolehlivě. Pokud je správně zvolena horní hranice, tak tato metoda nalezne téměř všechny hledané úseky a nenalezne úsek, kde je obsažen užitečný signál. U každé reálné nahrávky může být tato horní hranice trochu jiná. Záleží na mnoha faktorech jako je přesnost odhadu impulzní odezvy či na volbě odečítacího koeficientu α v rovnici 4.13. Proto byl ještě implementován algoritmus, který nalezené úseky uloží za sebou do jednoho vektoru. Tento vektor je možné poslechnout a pokud by bylo slyšet, že je ve vektoru obsažen i užitečný signál, musí se hodnota horní hranice zmenšit.

Při testování této metody dobře fungoval přístup, kdy se jako horní hranice nastavila nízká hodnota a odečítací koeficient byl zvolen poměrně vysoký (α=10). Testovalo se na 8 nahrávkách a algoritmus vždy našel většinu hledaných úseků, aniž by se musela měnit horní hranice.

4.8 Odhadnutí impulzní odezvy z nespojitých úseků signálu

Jakmile jsou k dispozici úseky ze zarušené nahrávky, které obsahují pouze známý rušivý signál, lze z nich odhadnout impulzní odezvu. Pokud bude nalezený úsek spojitý (např. od 2. sekundy do 40. sekundy v nahrávce), tak lze odhadnout impulzní odezvu jak je popsané v sekci 4.3.

Nechť zarušená nahrávka obsahuje jako užitečný signál rozhovor více lidí. Hledané úseky, kde není obsažen užitečný signál (rozhovor), mohou být na začátku nahrávky, v částech nahrávky mezi řečí nebo případně na konci nahrávky. Na začátku či na konci nahrávky může být hledaný úsek v případě, že se nahrávalo ještě před tím než rozhovor začal nebo poté co skončil. Úseky mezi řečí budou v nahrávce téměř vždy. Tyto úseky jsou tedy v různých částech nahrávky a nenavazují na sebe.

Je tedy potřeba pro každý nalezený úsek nalézt odpovídající úsek rušivého signálu. Odpovídající úsek rušivého signálu musí ještě navíc obsahovat určitý počet zpětných vzorků, který je určen délkou odhadované impulzní odezvy. Z nalezeného úseku se pak sestaví levá a pravá strana pro odhad stejně jako v sekci 4.3. Ze vzniklé levé a pravé strany se ale neodhaduje impulzní odezva ihned. Z dalších nalezených úseků, se vytvoří znovu levá a pravá strana a připojí se k již existující levé a pravé straně.

$$\mathbf{L}_k = \begin{pmatrix} x1_k(l) & x1_k(l-1) & \dots & x1_k(1) \\ x1_k(d) & x1_k(d-1) & \dots & x1_k(d-l) \\ x2_k(l) & x2_k(l-1) & \dots & x2_k(1) \\ x2_k(d) & x2_k(d-1) & \dots & x2_k(d-l) \\ \vdots & \vdots & \ddots & \vdots \\ xp_k(l) & xp_k(l-1) & \dots & xp_k(1) \\ xp_k(d) & xp_k(d-1) & \dots & xp_k(d-l) \end{pmatrix} \mathbf{p}_k = \begin{pmatrix} y1_k(l) \\ y1_k(d) \\ y2_k(l) \\ y2_k(d) \\ \vdots \\ yp_k(l) \\ yp_k(d) \end{pmatrix} \tag{4.20}$$

kde $\mathbf{L}$ je levá strana pro odhad, $\mathbf{p}$ je pravá strana pro odhad, l je maximální délka odhadované impulzní odezvy, $x1$ značí vstupní rámce prvního úseku, $x2$ značí vstupní rámce druhého úseku, xp značí vstupní rámce posledního úseku , $y1$ značí výstupní rámce prvního úseku, $y2$ značí výstupní rámce druhého úseku, yp značí výstupní rámce posledního úseku a d je počet všech rámců v daném úseku.

Z takto sestavených levých a pravých stran se pak pro každý spektrální koeficient k odhadne impulzní odezva stejně jako v sekci 4.3.

4.9 Další metody pro odhad impulzní odezvy

Zatím pro odhad impulzní odezvy byl použit přístup, kdy matice $\mathbf{L}$ je vytvořena ze vstupního signálu. Vektor $\mathbf{p}$ obsahuje výstupní signál. Impulzní odezva se pak odhadne na základě matice $\mathbf{L}$ a vektoru $\mathbf{p}$.

Jiný přístup pro odhad impulzní odezvy je založen na Wienerově filtru. Tento přístup pro odhad impulzní odezvy využívá statistické vlastnosti vstupního a výstupního signálu. Levá strana $\mathbf{L}$ je vytvořena z auto korelace vstupního signálu. V pravé straně $\mathbf{p}$ je obsažena

křížová korelace vstupního a výstupního signálu. Rovnice, které vzniknou se označují jako Wiener-Hopf rovnice. Více informací ohledně tohoto přístupu viz [3] a sekce 2.5.

Odhad probíhá ve frekvenční doméně. Ze vstupního signálu je vytvořena matice **X**. Z výstupního signálu je vytvořena matice **Y**. Tyto matice jsou vytvořeny stejným způsobem jako v sekci 4.3. Každý řádek matice **X** slouží jako vstupní signál a každý řádek matice **Y** slouží jako výstupní signál. Ze vstupního a z výstupního signálu se sestaví Wiener-Hopf rovnice. Každý řádek odpovídá jistému koeficientu Fourierovy transformace viz rovnice 4.7. Pro každý koeficient je odhadnuta impulzní odezva a je vytvořena matice **H**' stejně jako v rovnici 4.11. Tento přístup se jevil jako velmi rychlý a kvalitní, viz sekce 5.1.

Další přístup, který jsem aplikoval pro odhad impulzní odezvy, byl normalizovaný LMS adaptivní filtr[1]. Tento filtr implementovaný v rámci knihoven programu Matlab ze vstupu a výstupu odhadne impulzní odezvu systému. Vstup i výstup byl zvolen stejný jako u výše popsaného přístupu. Řád filtru se zvolil v závislosti na délce odhadované impulzní odezvy.

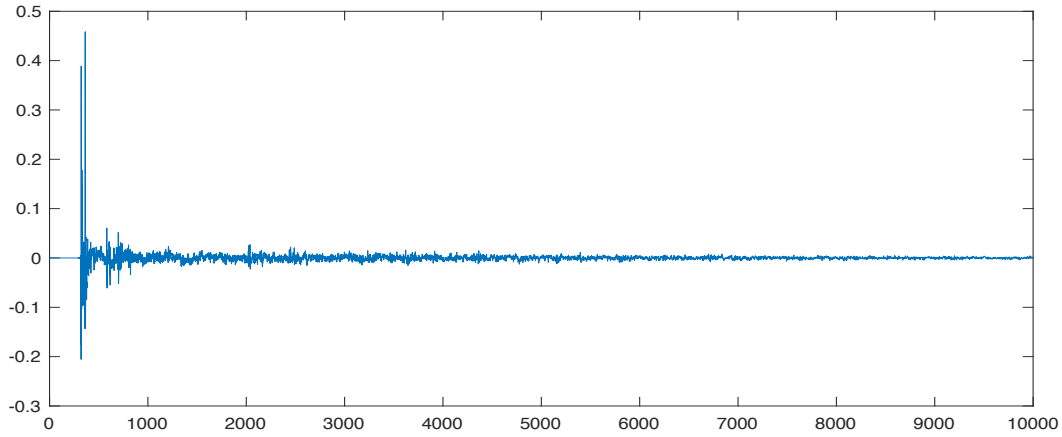
Tento přístup byl otestován na testovacích nahrávkách. Z hlediska rychlosti nepředstavoval zlepšení, spíše naopak. Z hlediska kvality odečtených nahrávek, při odhadu impulzní odezvy z celých signálů, byla kvalita srovnatelná s přístupem popsaným v sekci 4.6. Nepředstavoval tedy žádné zlepšení a proto jsem s ním dále nepracoval a neexperimentoval.

4.10 Analýza odhadnutých impulzních odezev

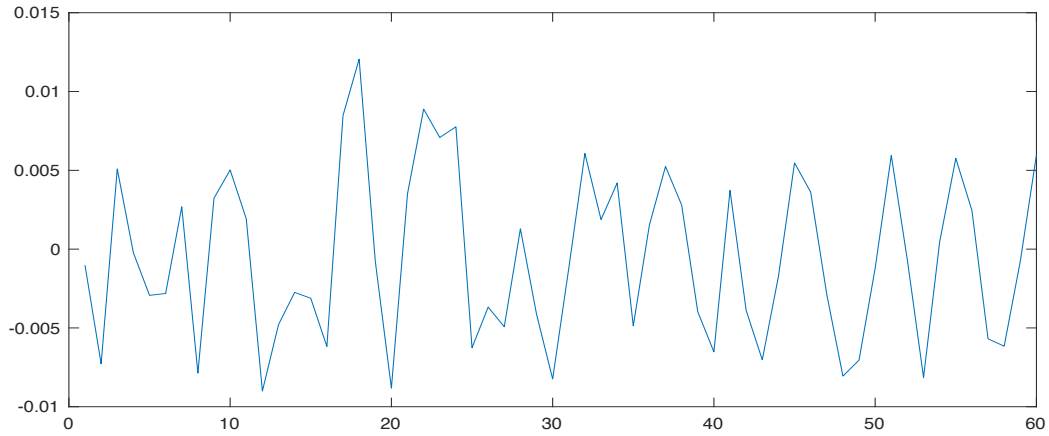
Impulzní odezva reálného prostředí (např. kanceláře) by měla mít největší hodnotu pro prvních pár vzorků (první vzorky odpovídají přímému vstupu). Absolutní hodnoty dalších vzorků by se měly postupně zmenšovat až k hodnotě 0. Více informací viz [1]. Ilustrace reálné impulzní odezvy kanceláře viz obrázek 4.6.

Odhadnuté impulzní odezvy ve frekvenční oblasti by měly mít podobný průběh. Byla tedy provedena analýza odhadnutých impulzních odezev. Odhadování impulzních odezev proběhlo na testovacích nahrávkách, které obsahovaly pouze nahraný rušivý signál. Rušivý signál byl nahrán v kanceláři. Většina impulzních odezev vypadala v pořádku viz obrázek 4.8, ale impulzní odezvy pro nejnižší a nejvyšší frekvence nebyly odhadnuty správně. Jejich průběh se zdaleka nepodobal očekávané impulzní odezvě místnosti viz obrázek 4.7. Aplikace takto chybně odhadnutých impulzních odezev zhoršuje kvalitu výsledného odečteného signálu. Ve výsledném programu se bude pracovat s nahrávkami, které budou mít vzorkovací frekvenci 16000 Hz. Ve frekvenční oblasti je k dispozici rozlišení od 0 do 8000 Hz. Navrhl jsem tedy, že se frekvence od 0 do cca 200 Hz a frekvence od cca 6500 do 8000 Hz se ve výsledné nahrávce podstatně ztlumí. Efekt většiny spatně odhadnutých impulzních odezev se tímto krokem eliminuje.

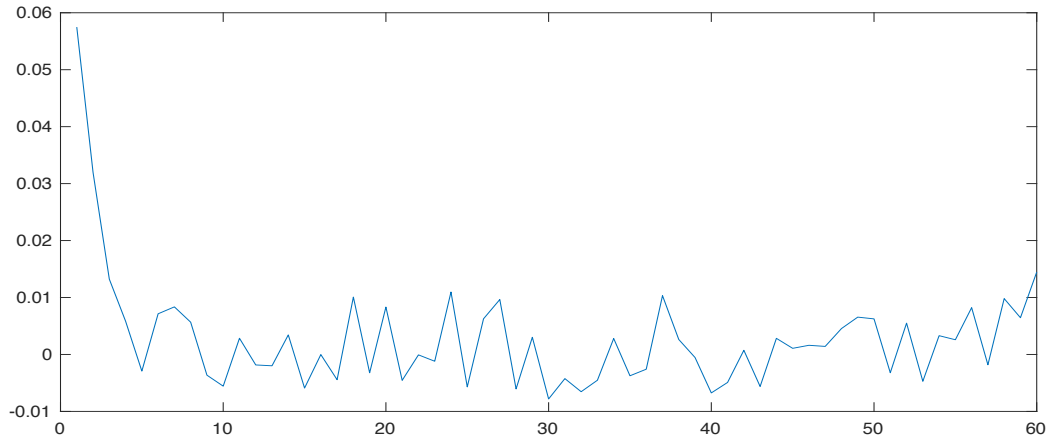
[1] https://www.mathworks.com/help/dsp/ug/lms-adaptive-filters.html#f1-5839



Obrázek 4.6: Ilustrace reálné impulzní odezvy kanceláře.
Data získána z databáze univerzity Aachen. https://www.iks.rwth-aachen.de



Obrázek 4.7: Chybný průběh odhadnuté impulzní odezvy pro nejvyšší frekvenci (8000 Hz).



Obrázek 4.8: Očekávaný průběh odhadnuté impulzní odezvy pro frekvenci 2100 Hz.

4.11 Analýza odečtených nahrávek

Jakmile je k dispozici impulzní odezva, odhadnutá některou z metod popsaných v této kapitole, je aplikována na známý rušivý signál. Tento rušivý signál se pak odečte od zarušené nahrávky pomocí spektrálního odečítání. Výsledkem je odečtená nahrávka, která by již neměla obsahovat rušivý signál.

Ve spektrogramu odečtené nahrávky se občas vyskytovaly krátké úseky, které obsahovaly všechny frekvence. Tyto úseky jsou ve spektrogramu reprezentovány jako dlouhé úzké sloupce. Poslechově se tyto úseky jeví jako „praskání“. Tento jev je způsoben poměrně velkou změnou hodnoty mezi jednotlivými vzorky odečtené nahrávky v časové oblasti. Při spektrálním odečítaní, pokud vyjde výsledný koeficient záporný, nastaví se na 0 viz rovnice 4.13. Je pravděpodobné, že tato skutečnost způsobuje jisté „nespojitosti“ v časové doméně. Na základě rady vedoucího práce, pana Černockého, byl vyzkoušen nový přístup spektrálního odečítání. Ilustrace původního spektrálního odečítaní viz schéma 4.3.

V novém přístupu se aplikuje Hann funkce až na konci procesu. Tedy až po zpětné Fourierově transformaci. Porovnání spektrogramů odečtených nahrávek, získané původním spektrálním odečítáním a novým přístupem viz obrázky 4.9, 4.10. Z obrázků lze vidět, že se ve spektru již neobjevují dlouhé sloupce značící „nespojitosti“. Z poslechu obou nahrávek šlo poznat, že v nahrávce odečtené novým přístupem „praskání“ již není. Nicméně v této nahrávce bylo slyšet nepatrně více rušivého signálu než v nahrávce odečtené původním přístupem. Proto byl ještě vyzkoušen další přístup.

Hann funkce na začátku i na konci

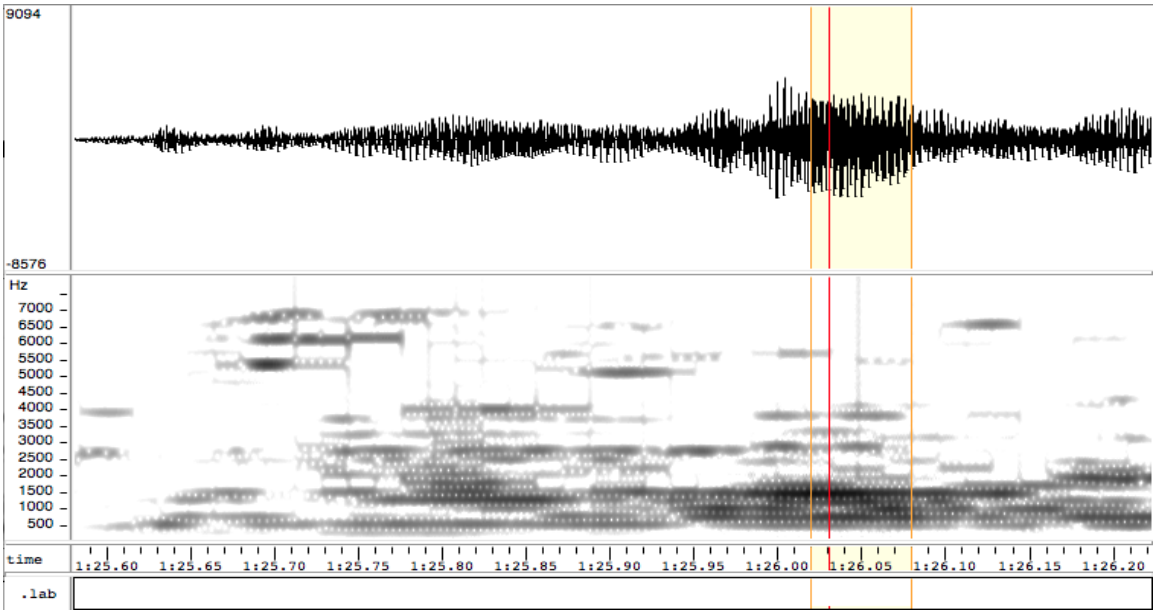
Tento přístup je založen na myšlence zkombinovat původní a nový přístup. Hann funkce se tedy bude aplikovat na začátku jako v původním přístupu, ale také na konci jako v novém přístupu. V rámci zachování energie signálu se ale nebude aplikovat celá. Bude se aplikovat její odmocnina. Tento přístup z hlediska poslechu vykazoval nejlepší výsledky. Praskání vymizelo a míra rušivého signálu byla podobná jako v původním přístupu. Z hlediska kvality byl lehce pozměněn i užitečný signál. Užitečný signál je více „vyhlazen“ a je tedy příjemnější na poslech.

4.12 Použití dereverbace pro lepší odhad impulzní odezvy

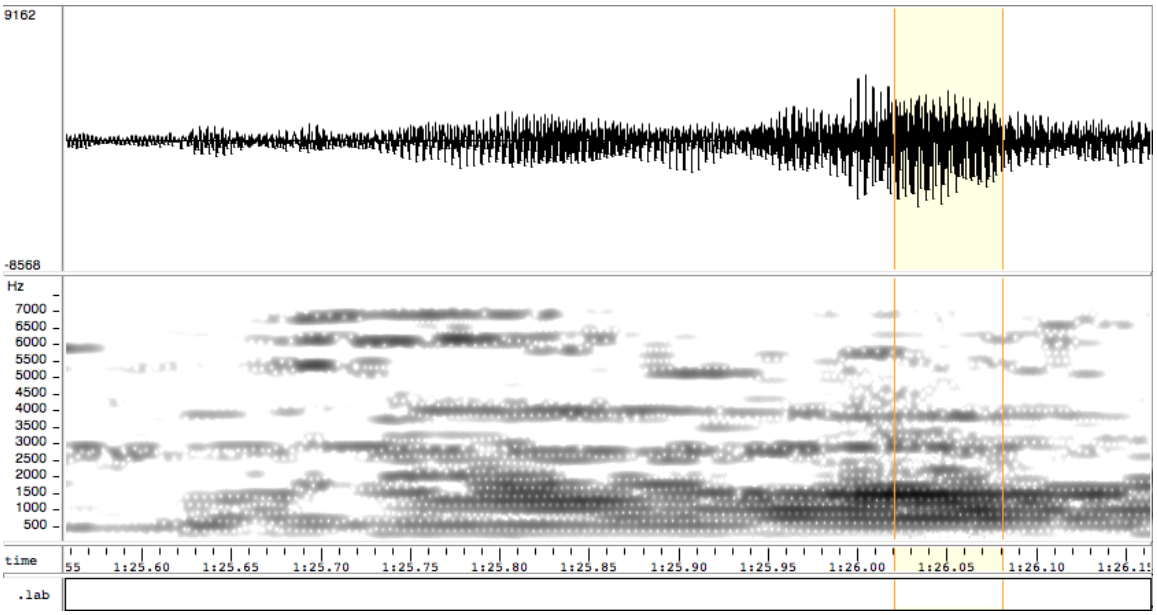
Dereverbace je proces, který se snaží odstranit reverbaci, která je obsažena v signálu. Myšlenka použití dereverbace pro lepší odhad impulzní odezvy je následující. Dereverbace se použije na zarušenou nahrávku. Tím se zvýší kvalita zarušené nahrávky a rušivý signál, který je v ní obsažen, bude podstatně méně změněn prostředím. Tato skutečnost by mohla mít vliv na odhad impulzní odezvy a na případné zlepšení kvality odečteného signálu. Pro dereverbaci byl použit program[2], který používá metodu známou jako weighted prediction error (WPE). Tato metoda zpracovává signál v časově frekvenční doméně.

Testování proběhlo na 5 testovacích nahrávkách s reálným mluvčím. Dereverbace se aplikovala na zarušenou nahrávku. Kvalita výsledných odečtených nahrávek byla podobná jako v případě, kdy dereverbace nebyla použita. Aplikace dereverbace na zarušenou nahrávku tedy nezlepšila kvalitu výsledných odečtených nahrávek. Pokud se ale aplikuje dereverbace na již odečtenou výslednou nahrávku, tak kvalita užitečného signálu se podstatně zlepší. Zlepšení kvality závisí na množství reverbace obsažené v zarušené nahrávce.

[2] https://github.com/helianvine/fdndlp



Obrázek 4.9: Spektogram odečtené nahrávky získané pomocí spektrálního odečítaní viz schéma 4.3.



Obrázek 4.10: Spektogram odečtené nahrávky získané pomocí nového spektrálního odečítaní viz sekce 4.11.

Kapitola 5

Experimenty s vybranými metodami

5.1 Porovnání metod pro odhad impulzní odezvy

Cílem tohoto experimentu je porovnat vybrané metody pro odhad impulzní odezvy. Odhad impulzní odezvy má zásadní vliv na výslednou odečtenou nahrávku. Proto je třeba metody důkladně otestovat, zjistit jak se chovají pro různé typy nahrávek a doporučit ty metody, které budou mít nejlepší výsledky. Jednotlivé metody jsou porovnávány podle výsledných (odečtených) nahrávek. Kvalita výsledných nahrávek se vyhodnocuje objektivně (spektrální logaritmická vzdálenost) a subjektivně poslechem. Jako testovací nahrávky se zvolí simulované nahrávky viz sekce 3.3. V těchto nahrávkách je lidská řeč (užitečný signál) přehráván pomocí reproduktoru. Rušivými signály v nahrávkách jsou klasická hudba, pop hudba a talk show. Důvod pro výběr těchto rušivých signálů je popsán v sekci 3.1. Pro jednotlivé nahrávky obsahující určitý typ rušivého signálu jsou k dispozici další dvě související nahrávky. První obsahuje cca 8 minut pouze rušivého signálu. Druhá obsahuje pouze užitečný signál viz sekce 3.3.

Chování jednotlivých metod pro různé typy testovacích nahrávek bude porovnáváno v závislosti na kvalitě odečtených nahrávek. Bude dobré pokud tyto nahrávky budou odečteny s optimálním odečítacím koeficientem. Tedy koeficientem, pro který bude mít odečtená nahrávka nejlepší kvalitu. Nahrávka s nejlepší kvalitou by měla obsahovat co nejméně deformovaný užitečný signál a zároveň obsahovat co nejmenší množství rušivého signálu. Odečítací koeficienty se používají při spektrálním odečítaní a určují množství odečítaného signálů viz sekce 4.5. Metody aplikované na jednotlivé testovací nahrávky se tedy budou porovnávat podle kvality odečtené nahrávky, odečtené pomocí optimálního odečítacího koeficientu. Chování metod bude také porovnáváno podle toho, jak se mění kvalita odečtené nahrávky při změně odečítacího koeficientu.

Pro porovnání byly vybrány tři metody pro odhad impulzní odezvy:

- První metoda odhaduje impulzní odezvu z celých signálů viz sekce 4.6, způsobem popsaným v sekci 4.3 (odhad ve frekvenční oblasti), označí se jako metoda **A**.
- Druhá metoda odhaduje impulzní odezvu z celých signálů, pomocí Wienerova filtru viz sekce 4.9, označí se jako metoda **B**.
- Třetí metoda odhaduje impulzní odezvu z úseků obsahující pouze rušivý signál, které byly nalezené přístupem popsaným viz sekce 4.7. Samotný odhad impulzní odezvy je

popsán viz sekce 4.8 (odhadování impulzní odezvy z nespojitých úseků), metoda bude označena jako metoda **C**.

Realizace porovnání metod

Nejprve se vybere simulovaná nahrávka. Poté se některou metodou odhadne impulzní odezva. Tato impulzní odezva se aplikuje na známý rušivý signál. Následuje spektrální odečítání rušivého signálu od zarušené nahrávky. Spektrální odečítaní se provede s koeficienty od α=0.1 až po α=8. Koeficienty se zvětšují o 0.1. Bude tedy vytvořeno 80 odečtených nahrávek. Pro každou nahrávku se zjistí její spektrální logaritmická vzdálenost vůči nahrávce, která obsahuje pouze užitečný signál. Nahrávka pro kterou vyjde nejmenší logaritmická vzdálenost bude považována za optimální. Koeficient, pomocí kterého se odečítala optimální nahrávka, bude považován za optimální koeficient. Celý tento postup se provede pro všechny metody (výše zmíněné tři metody pro odhad impulzní odezvy).

Porovnání metod na nahrávce obsahující jako rušivý signál klasickou hudbu

Na obrázku 5.1 lze vidět průběhy jednotlivých logaritmických vzdáleností pro vybrané metody, kde jako testovací nahrávka byla použita nahrávka, která obsahuje jako rušivý signál **klasickou hudbu** . Z obrázků lze vypozorovat chování jednotlivých metod, které odhadovaly impulzní odezvu z dané nahrávky. Na obrázcích jsou na ose x jednotlivé odečítací koeficienty a na ose y jsou jednotlivé spektrální logaritmické vzdálenosti v decibelech. Optimálním koeficientem pro metodu **A** je koeficient α=0.6 se vzdáleností 7.4, pro metodu **B** je to koeficient α=2.1 se vzdáleností 6.65 a pro metodu **C** je to α=1.6 se vzdáleností 6.60.

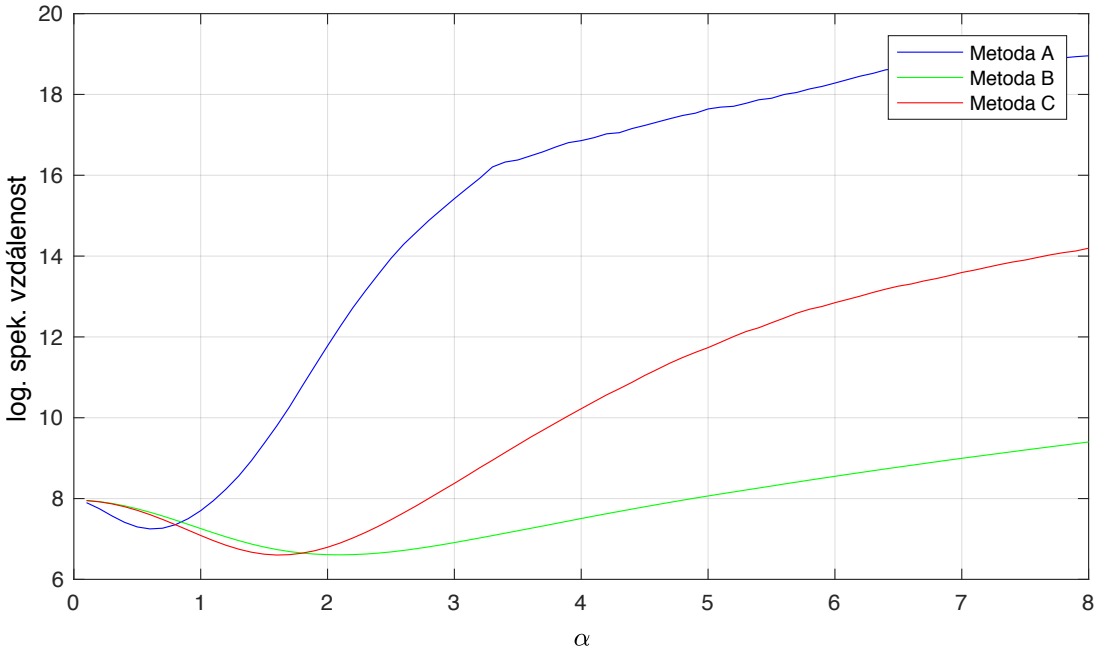
Metoda **A** má viditelně rozdílný průběh oproti ostatním metodám a optimální koeficient je roven 0.6. Při poslechu optimální nahrávky (nahrávka odečtená optimálním koeficientem) je užitečný signál v téměř v originálním stavu. Rušivý signál nicméně není dostatečně odečten. V nahrávce odečtené s koeficientem 1.5 je rušivý signál slyšet podstatně méně, užitečný signál je ovšem slyšitelně deformovaný. V nahrávce odečtené s koeficientem 3.0 rušivý signál téměř není, nicméně užitečný signál je deformovaný natolik, že je obtížné mu rozumět. Hodnocení kvality odečtené nahrávky sluchem koresponduje s průběhem jeho spektrálních logaritmických vzdáleností. Při zvyšování koeficientu rapidně klesá kvalita užitečného signálu. Toto je pravděpodobně způsobeno tím, že tato metoda odhaduje impulzní odezvu z celého signálu. V odhadnuté impulzní odezvě je částečně „obsažen“ užitečný signál.

Metoda **C** nemá tak strmý průběh spektrálních logaritmických vzdáleností jako metoda **A**. Optimální koeficient je roven 1.6. Při poslechu optimální nahrávky je stejně jako u metody **A** užitečný signál v téměř originálním stavu. Rušivý signál také není úplně odečten. Rozdíl metod je ovšem patrný při zvyšování koeficientu. V nahrávce odečtené s koeficientem 2.2 je rušivý signál slyšet podstatně méně, užitečný signál má stále dobrou kvalitu. V nahrávce odečtené s koeficientem 3.3 rušivý signál v nahrávce téměř neexistuje. Užitečný signál je částečně deformovaný, nicméně lze mu bez problému rozumět. Hodnocení kvality odečtené nahrávky sluchem opět koresponduje s průběhem jeho spektrálních logaritmických vzdáleností. Při zvyšování koeficientu klesá množství rušivého signálu v nahrávce a zároveň neklesá tak rapidně kvalita užitečného signálu jako u metody **A**.

Metoda **B** má nejplynulejší průběh ze všech metod. Optimální koeficient je roven 2.1. Kvalita odečtených nahrávek je velmi podobná jako u metody **C**. Metoda **C** má ovšem z hlediska poslechu o něco lepší kvalitu. Metoda **B** má také vlastnost, že při zvyšování koeficientu klesá množství rušivého signálu v nahrávce a zároveň neklesá tak rapidně kvalita užitečného signálu. Pokud se vezme v potaz, že metoda **B** odhaduje impulzní odezvu z celého

signálu, je její chování pro testovací nahrávku velmi dobré. Navíc z hlediska výpočetní rychlosti je na tom velmi dobře viz tabulka 5.2.

Vytvořil jsem webovou stránku, kde je možné si poslechnout vybrané odečtené nahrávky. Odkaz je přístupný na adrese http://www.stud.fit.vutbr.cz/~xhosek11/index.html .



Obrázek 5.1: Průběhy jednotlivých logaritmických vzdáleností pro vybrané metody, kde jako testovací nahrávka byla použita nahrávka, která obsahuje jako rušivý signál klasickou hudbu.

Porovnání metod na nahrávce obsahující jako rušivý signál pop hudbu

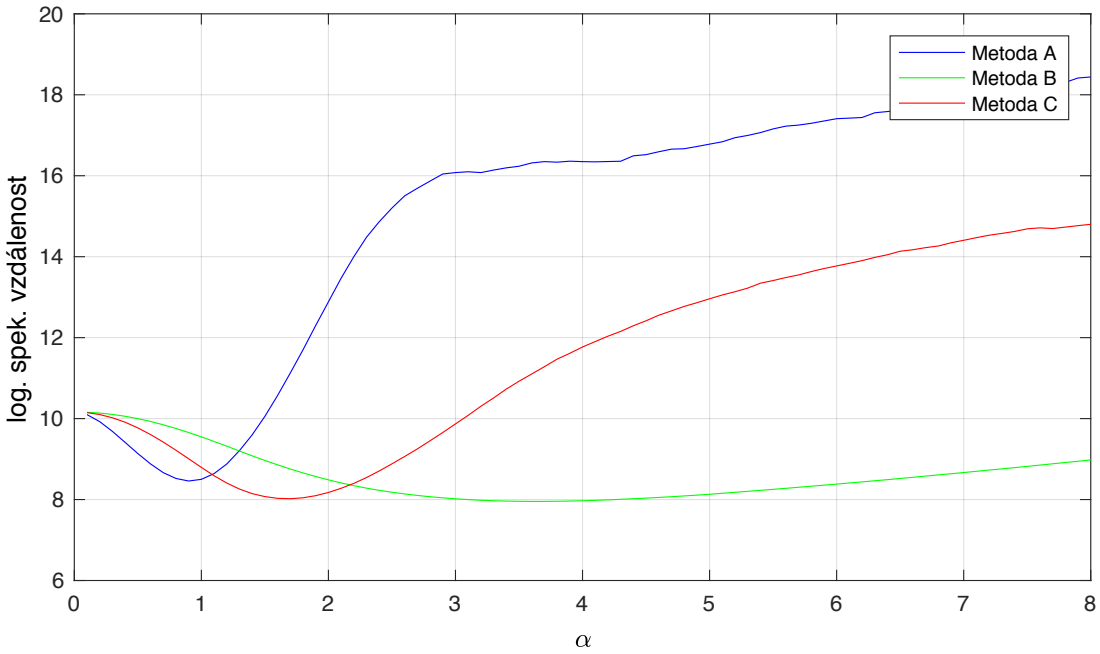
Na obrázku 5.2 lze vidět průběhy jednotlivých logaritmických spektrálních vzdáleností pro vybrané metody. Jako testovací nahrávka byla použita nahrávka, která obsahuje jako rušivý signál **pop hudbu**. Na obrázku jsou na ose x jednotlivé odečítací koeficienty a na ose y jsou jednotlivé logaritmické vzdálenosti v decibelech. Optimálním koeficientem pro metodu **A** je koeficient α=0.9 se vzdáleností 8.46, pro metodu **B** je to koeficient α=4.1 se vzdáleností 8.1 a pro metodu **C** je to α=1.7 se vzdáleností 8.01.

Metoda **A** má stejně jako u nahrávky s klasickou hudbou nejstrmější průběh oproti ostatním metodám. Optimální koeficient je roven 0.9. Při poslechu optimální nahrávky je užitečný signál v kvalitním stavu, těžko rozlišitelný od originálu. Rušivý signál není ovšem dostatečně odečten. V nahrávce odečtené s koeficientem 1.5 je rušivý signál slyšet méně, nicméně užitečný signál se začíná deformovat. V nahrávce odečtené s koeficientem 3.0 se rušivý signál objevuje v malé míře. Užitečný signál je ovšem deformovaný natolik, že je obtížné mu rozumět. Kvalita nahrávek s klasickou hudbou je lepší než kvalita těchto nahrávek. Průběh vzdáleností je podobný jako u nahrávek s klasickou hudbou. Platí stejné pravidlo, že při zvyšování koeficientu rapidně klesá kvalita užitečného signálu.

Metoda **C** nemá opět tak strmý průběh spektrálních logaritmických vzdáleností jako metoda **A**. Optimální koeficient je roven 1.7. Při poslechu optimální nahrávky je užitečný

signál v téměř v originálním stavu. Rušivý signál je ve slyšitelně menší míře než u metody **A**. Nicméně stále není dostatečně odečten. V nahrávce odečtené s koeficientem 2.5 je rušivý signál slyšet méně, užitečný signál je pořád v dobrém stavu. V nahrávce odečtené s koeficientem 3.5 je rušivý signál obsažen v malé míře, ale užitečný signál se již začíná deformovat. Kvalita těchto nahrávek je také horší než kvalita nahrávek s klasickou hudbou. Metoda **C** se ovšem chová podobně jako u nahrávek s klasickou hudbou. Platí pravidlo, že při zvyšování koeficientu klesá množství rušivého signálu v nahrávce a zároveň kvalita užitečného signálu neklesá tak rapidně.

Metoda **B** má opět nejplynulejší průběh ze všech metod. Optimální koeficient je roven 3.6. Kvalita odečtených nahrávek je velmi podobná jako u metody **C**. Sluchem nebylo možné rozeznat rozdíl kvalit. U metody **B** také platí pravidlo, že při zvyšování koeficientu klesá množství rušivého signálu v nahrávce a zároveň kvalita užitečného signálu neklesá tak rapidně.



Obrázek 5.2: Průběhy jednotlivých logaritmických vzdáleností pro vybrané metody, kde jako testovací nahrávka byla použita nahrávka, která obsahuje jako rušivý signál pop hudbu.

Porovnání metod na nahrávce obsahující jako rušivý signál talk show

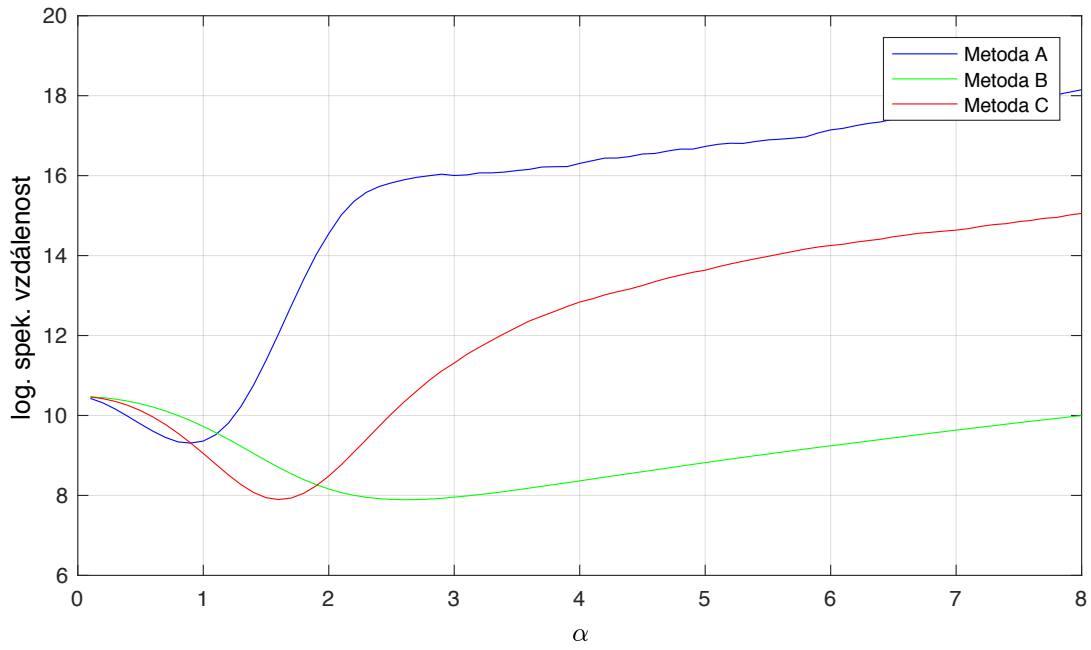
Na obrázku 5.3 lze vidět průběhy jednotlivých logaritmických spektrálních vzdáleností pro vybrané metody. Jako testovací nahrávka byla použita nahrávka, která obsahuje jako rušivý signál **talk show**. Tento rušivý signál obsahuje z velké části pouze lidskou řeč. Na obrázku jsou na ose x jednotlivé odečítací koeficienty a na ose y jsou jednotlivé spektrální logaritmické vzdálenosti v decibelech. Optimálním koeficientem pro metodu **A** je koeficient α=0.9 se vzdáleností 8.46, pro metodu **B** je to koeficient α=4.1 se vzdáleností 7.98 a pro metodu **C** je to α=1.6 se vzdáleností 7.95.

Metoda **A** má podobný průběh jako u nahrávek s klasickou hudbou a pop hudbou. Optimální koeficient je roven 0.9. Při poslechu optimální nahrávky je užitečný signál v

kvalitním stavu, lze mu perfektně rozumět. Rušivý signál je obsažen v malé míře. To je první případ, kdy pro optimální koeficient je rušivý signál přítomen v malé míře. V nahrávce odečtené s koeficientem 1.5 je rušivý signál obsažen ve velmi malé míře, užitečný signál je jen lehce deformován. V nahrávce odečtené s koeficientem 3.0 rušivý signál téměř neexistuje. Užitečný signál je deformován, nicméně deformace není nijak závratná. Užitečnému signálu jde bez problému rozumět. Metoda **A** má jednoznačné nejlepší výsledky pro tento typ nahrávek (rušivý signál je talk show). Průběh spektrálních logaritmických vzdáleností je sice podobný jako u předchozích dvou typů nahrávek, nicméně kvalita z hlediska poslechu je podstatně lepší. Může to být způsobeno tím, že rušivý signál obsahuje podobné frekvence jako užitečný signál.

Metoda **C** nemá opět tak strmý průběh jako metoda **A**. Optimální koeficient je roven 1.6. Poslechová kvalita je ještě lepší než u metody **A**. Při poslechu optimální nahrávky, je užitečný signál v téměř originálním stavu. Rušivý signál je obsažen v malé míře. V nahrávce odečtené s koeficientem 2.5 je rušivý signál slyšet méně, užitečný signál je stále v téměř originálním stavu. V nahrávce odečtené s koeficientem 3.8 rušivý signál téměř neexistuje. Užitečný signál je ale stále ve velmi dobré kvalitě.

Metoda **B** má nejplynulejší průběh ze všech metod. Optimální koeficient je roven 2.6. Při poslechu optimální nahrávky je užitečný signál v téměř originálním stavu. Rušivý signál ale není tak dobře odečten jako u ostatních metod. Při zvyšování koeficientů se rušivý signál odečítá ve větším množství a užitečný signál si poměrně drží kvalitu. V nahrávce odečtené koeficientem 5.5 je užitečný signál v dobrém stavu a rušivý signál je obsažen v minimální míře. Nahrávky mají podobnou kvalitu jako u metody **A**, nicméně nedosahují kvalit metody **C**.



Obrázek 5.3: Průběhy jednotlivých logaritmických vzdáleností pro vybrané metody, kde jako testovací nahrávka byla použita nahrávka, která obsahuje jako rušivý signál talk show.

Celkové zhodnocení metod

Souhrn minimálních logaritmických spektrálních vzdáleností pro jednotlivé metody a typy nahrávek je v tabulce 5.1. V tabulce je také metoda, která odhaduje impulzní odezvu z cca 8 minut rušivého signálu. Tato metoda simuluje ideální případ, kdy je v zarušené nahrávce k dispozici dlouhý úsek, kde je obsažen pouze známý rušivý signál. Z tohoto úseku se odhadne impulzní odezva přístupem popsaným v sekci 4.3. V tabulce je tato metoda označena jako referenční.

Typ nahrávky	Referenční metoda (dB)	Metoda C (dB)	Metoda B (dB)	Metoda A (dB)
Klasická hudba	6.45	**6.60**	6.65	7.4
Pop	7.83	**8.01**	8.1	8.46
Talk show	7.84	**7.88**	7.89	9.31

Tabulka 5.1: Souhrn minimálních logaritmických spektrálních vzdáleností pro jednotlivé metody a typy nahrávek.

Tabulka 5.2 obsahuje pro každou porovnávanou metodu průměrný čas potřebný pro odhad impulzní odezvy. Průměrný čas byl vypočten pro simulované nahrávky. Délka těchto nahrávek je cca 10 minut. Metoda **C** použila pro „částečné“ odečtení metodu **B**. Výpočet probíhal na Macbooku Pro, model z poloviny roku 2012. Macbook má dvou-jádrový procesor Intel Core i5 s taktovací frekvencí 2.5 GHz. Velikost paměti je 4 GB.

Typ nahrávky	Průměrný čas výpočtu (sekundy)
Metoda A	37
Metoda B	**5.6**
Metoda C	106.2

Tabulka 5.2: Průměrná doba pro výpočet impulzní odezvy pro jednotlivé metody (délka nahrávek cca 10 minut).

Pro nahrávky, kde je rušivým signálem klasická hudba, dosahuje nejlepších výsledků metoda **C**. Následuje metoda **B**, která dosahuje lehce horších výsledků. Metoda **A** má nejhorší výsledky.

Pro nahrávky, kde je rušivým signálem hudba se zpěvem, dosahuje nejlepších výsledků metoda **C** a metoda **B**. Mezi těmito metodami nešlo rozlišit z hlediska poslechu a z hlediska minimálních spektrálních vzdáleností byl rozdíl minimální. Nejhorší výsledky má metoda **A**.

Pro nahrávky, kde je rušivým signálem lidská řeč, dosahuje nejlepších výsledků metoda **C**. Metoda **A** dosahuje také velmi kvalitních výsledků. Dokáže rušivý signál ve velké míře odstranit při zachování dobré kvality užitečného signálu. Metoda **B** dosahuje podobných kvalit jako metoda **A**.

Jelikož metoda **C** dosahovala nejlepších výsledků u všech typů nahrávek, lze ji považovat za nejlepší metodu. Metoda **B** dosahovala buď stejných nebo lehce horších výsledků než metoda **C**. Metoda **A** dosahovala srovnatelných výsledků jako metoda **B** pro nahrávku, kde je užitečný signál velmi podobný rušivému. U jiných typů nahrávek měla nejhorší výsledky. Z hlediska rychlosti výpočtu je jednoznačně nejlepší metoda **B**.

Výsledný program proto bude používat pro odhad impulzní odezvy metody **C** a **B**. Pro „částečné“ odečtení, které metoda **C** potřebuje, se použije metoda **B** vzhledem k její přesnosti a rychlosti. Pokud bude uživatel chtít, bude možné odhadnout impulzní odezvu také pomocí metody **A**.

5.2 Testování programu na nahrávkách s reálným mluvčím

Cílem tohoto experimentu je otestovat program na nahrávkách s reálným mluvčím viz sekce 3.2. K těmto nahrávkám je k dispozici pouze známý rušivý signál. Kvalitu odečtených nahrávek lze zhodnotit pouze sluchem. Zarušené nahrávky jsou dlouhé cca 3 minuty, rušivými signály jsou klasická hudba, rock hudba a talk show. Snažil jsem se nalézt odečítací koeficient tak, aby byl ve výsledné nahrávce obsažen co nejméně rušivý signál a zároveň užitečný signál byl co nejméně pozměněn.

Kvalita odečtených nahrávek je velmi podobná jako u simulovaných nahrávek. Nejlepší kvalitu má opět nahrávka s talk show a nahrávka s klasickou hudbou. O něco horší kvalitu má nahrávka s rock hudbou.

5.3 Hledání dostatečné délky úseků obsahujících pouze rušivý signál pro metodu C

Metoda **C** vykazovala nejlepších výsledků a je použita i ve výsledném programu. Tato metoda odhaduje impulzní odezvu z úseků, které obsahují pouze rušivý signál. Není zatím jasné jaká je dostačující délka těchto úseků obsažených v zarušené nahrávce, aby metoda **C** fungovala spolehlivě.

Cílem tohoto experimentu je zjistit dostačující délku úseků, ze kterých metoda **C** odhaduje impulzní odezvu. Testovat se bude na 3 zarušených simulovaných nahrávkách. Nahrávky budou obsahovat jako rušivý signál klasickou hudbu, pop hudbu a talk show.

Realizace experimentu

Vybere se simulovaná nahrávka. Pro vybranou nahrávku se nalezenou úseky, které obsahují pouze rušivý signál. Z nalezených úseků, metoda **C** odhaduje impulzní odezvu. Tyto úseky se budou postupně zkracovat. Budou tedy odhadovány impulzní odezvy pro jednotlivé zkrácené nalezené úseky. Chování metody se bude zjišťovat v závislosti na kvalitě výsledných odečtených nahrávek. Kvalita se určí spektrální logaritmickou vzdáleností a poslechem. Tento postup se zopakuje pro všechny testovací zarušené nahrávky.

Zhodnocení experimentu

V tabulce 5.3 jsou obsaženy spektrální logaritmické vzdálenosti jednotlivých výsledných nahrávek vůči nahranému užitečnému signálu. Grafová ilustrace dat z tabulky 5.3 je na obrázku 5.4. Odečítací koeficient pro všechny nahrávky byl zvolen jako α=3. Délka všech úseků obsahujících pouze rušivý signál se v testovacích nahrávkách pohybovala okolo dvou minut. V tabulce jsou uvedeny spektrální logaritmické vzdálenosti pro úseky zkrácené na 60, 30, 15, 7 a 1 sekundu. Maximální délka odhadované impulzní odezvy byla jedna sekunda. Tudíž úsek délky jedné sekundy je nejmenší možný úsek, ze které může metoda odhadnout impulzní odezvu. Z tabulky lze vypozorovat, že odhad impulzní odezvy z kratších úseků, zhoršuje kvalitu výsledné nahrávky. Největší zhoršení lze pozorovat mezi úseky celkové

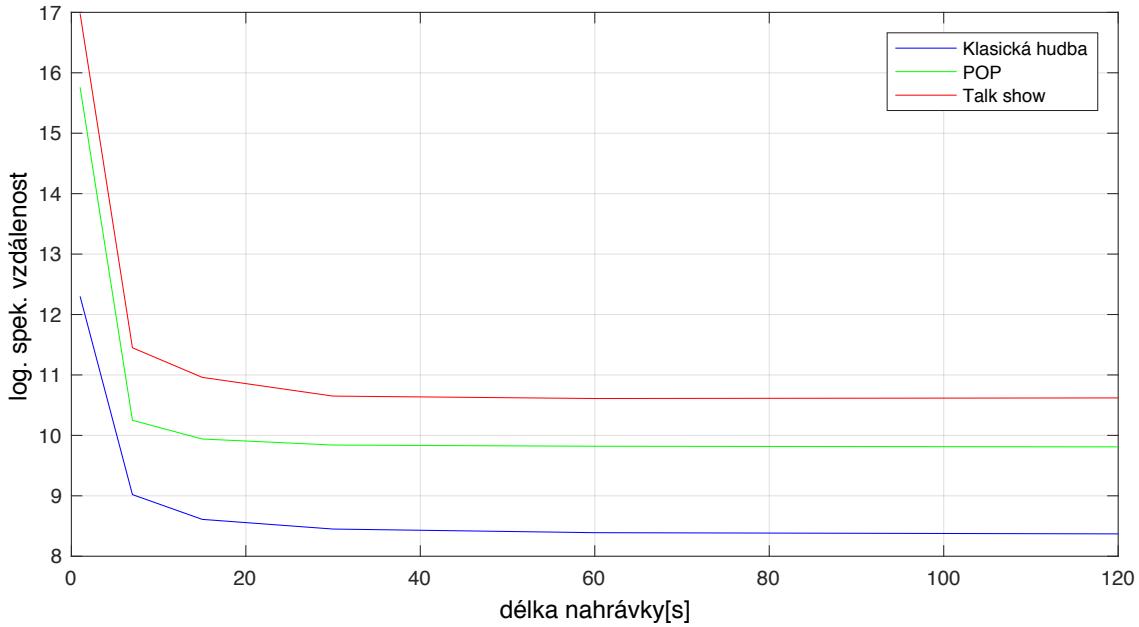
délky sedmi sekund a úseky celkové délky jedné sekundy. Rozdíl mezi všemi úseky (cca dvě minuty) a úseky celkové délky třiceti sekund je minimální.

Z hlediska poslechové kvality výsledné nahrávky nebylo možné rozeznat rozdíl až do úseku 15 sekund. Toto platilo pro všechny testovací nahrávky. Výsledná nahrávka pro úsek sedmi sekund již měla slyšitelně horší kvalitu, nicméně užitečnému signálu šlo stále rozumět. Nejhorší kvalitu měla nahrávka pro úsek jedné sekundy. Užitečný signál byl většinou nesrozumitelný. Tudíž kvalita byla velmi špatná.

Pokud je tedy v zarušené nahrávce nalezeno alespoň 30 sekund úseků, které obsahují pouze rušivý signál, metoda **C** bude pracovat spolehlivě. Pokud je délka nalezených úseků od 15 do 30 sekund, metoda bude pracovat stále velmi dobře. Pokud ovšem bude v zarušené nahrávce méně jak 10 sekund nalezených úseků, je vhodné zvolit jinou metodu, např. metodu **B**, která odhaduje impulzní odezvu z celého signálu.

Typ nahrávky	**Všechny úseky (dB)**	**Úsek 60 sec (dB)**	**Úsek 30 sec (dB)**	**Úsek 15 sec (dB)**	**Úsek 7 sec (dB)**	**Úsek 1 sec (dB)**
Klasická hudba	8.37	8.39	8.45	8.61	9.02	12.30
Pop	9.81	9.82	9.84	9.94	10.25	15.79
Talk show	10.62	10.61	10.65	10.96	11.45	16.98

Tabulka 5.3: Logaritmické spektrální vzdálenosti pro zkrácené úseky rušivého signálu.



Obrázek 5.4: Průběhy jednotlivých logaritmických vzdáleností pro zkrácené úseky rušivého signálu.

Kapitola 6

Demonstrační aplikace

Obrázek 6.1: Grafické uživatelské prostředí aplikace pro odstranění známého rušivého signálu z nahrávky.

6.1 Popis funkcionality

Na obrázku 6.1 je ukázka demonstrační aplikace. Tato aplikace byla implementována ve vývojovém prostředí App designer[1]. Práce s aplikací je následující: Nejprve se vybere zarušená nahrávka a rušivý signál ve formátu WAV. Vzorkovací frekvence nahrávek musí být 16000 Hz. Pokud nejsou nahrávky zarovnané, je potřeba je zarovnat (kliknutí na tlačítko

[1] https://www.mathworks.com/products/matlab/app-designer.html

„Zarovnat signály“). Stisknutím tlačítka „Rychlá metoda“ bude odstraněn známý rušivý signál metodou **B**. Je možné zvolit odečítací koeficient α. Odečtenou nahrávku lze poslechnout či ji případně uložit na disk počítače. Pro odstranění známého signálu z nahrávky metodou **C** se klikne na tlačítko „Klasická metoda“. Pro optimální výsledek je doporučené zkontrolovat dvě věci:

- Poslechnout si nalezené úseky, ze kterých metoda **C** odhaduje impulzní odezvu (stisknutí tlačítka „Poslech“). Nesmí být slyšet užitečný signál.
- Musí být nalezeno více jak 10 sekund rušivého signálu (příslušné zaškrtávací pole bude zaškrtnuté).

Pokud jsou tato kritéria splněna, měla by metoda pracovat optimálně. I u metody **C** může uživatel zadat odečítací koeficient α. Tlačítka „ZEK“ a „ZVEK“ pak slouží pro ladění hodnoty horní hranice energie viz sekce 4.7.

Kapitola 7

Závěr

7.1 Shrnutí provedené práce

Cílem této práce bylo odečíst známý rušivý signál z nahrávky. Pro odečtení rušivého signálu bylo třeba odhadnout impulzní odezvu místnosti. Byly představeny metody pro odhad impulzní odezvy místnosti. Kladl se důraz zejména na odhad impulzní odezvy v časově frekvenční oblasti. Jednotlivé metody byly vyhodnoceny na testovacích nahrávkách a byly doporučeny metody, které dosahovaly nejlepších výsledků. Na základě analýz odečtených nahrávek se modifikovaly metody, což pak vedlo k lepším výsledkům. Experimentovalo se také s metodou, která dosahovala nejlepších výsledků. Tyto experimenty vedly k nalezení podmínek, které musí být splněny, aby metoda pracovala optimálně. Testoval se také vliv dereverbace zarušených nahrávek na odhad impulzní odezvy. Byl vyvinut přístup pro nalezení úseků, které obsahují pouze rušivý signál. Vytvořilo se také základní grafické uživatelské rozhraní , které umožňuje uživateli měnit parametry metod, čímž je zajištěna větší univerzálnost.

7.2 Výhled do budoucna

V blízké budoucnosti bych navrhoval se zaměřit na odhad impulzní odezvy. Mohly by se otestovat neuronové sítě. Místo impulzní odezvy by se mohl odhadovat ARMAX či ARIMAX model [5]. Impulzní odezva by se také mohla odhadovat postupně. Tím by se zlepšila kvalita odečtených nahrávek, pokud by se charakteristika prostředí s časem měnila. Pro vylepšení již odhadnutého modelu, by se mohla použít např. metoda PEM (prediction error method) viz [5].

V delším časovém horizontu by se mohl vytvořit rekordér, který by uměl odečítat rušivý signál v reálném čase. Nahraný audio signál by již obsahoval pouze užitečný signál. Pro realizaci odčítání rušivého signálu v reálném čase by se mohly aplikovat adaptivní filtry, které by byly modifikovány pro tento problém.

Literatura

[1] Habets, E.: *Room Impulse Response Generator*. Internal Report, 2010.

[2] Černocký Jan: *Zpracování řečových signálů, studijní opora*. FIT VUT v Brně, 2006.

[3] Jiří, J.: *Číslicová filtrace, analýza a restaurace signálů*. Brno: Vysoké učení technické, 1997, ISBN 80-214-0816-2.

[4] Keesman, K.: *System identification: An introduction* . Springer Science Business Media, 01 2011, ISBN 978-0-85729-522-4.

[5] Ljung, L.: *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall PTR, 1999, ISBN 978-0-136-56695-3.

[6] Ljung, L.: *Black-box models from input-output measurements*. Instrumentation and Measurement Technology Conference, 06 2001, [Online; navštíveno 15.3.2019]. URL https://www.researchgate.net/publication/3899487_Black-box_models_from_input-output_measurements

[7] Prasad, B.; Prasanna, S.: *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*. Springer-Verlag Berlin Heidelberg, 2008, ISBN 978-3-540-75398-8.

[8] Reichl, J.; Všetička, M.: *Dozvuk, doba dozvuku*. 2006-2009, [Online; navštíveno 02.03.2019].
URL
http://fyzika.jreichl.com/main.article/view/1182-dozvuk-doba-dozvuku

[9] SCHIMMEL, J.: *Akustika uzavřených prostorů, Studijní text k předmětu Elektroakustika*. Brno, VUT v Brně, 2009.

[10] Vaseghi, S. V.: *Advanced Digital Signal Processing and Noise Reduction*. Department of Electronics Computer Engineering Brunel University, London, UK, 2008, ISBN 978-0-470-75406-1.

[11] Vrožina, M.; Jančíková, Z.; David, J.: *IDENTIFIKACE SYSTÉMŮ*. Vysoká škola báňská – Technická univerzita Ostrava, 2012.

[12] Černý, F.: *Simulace šíření zvukové vlny v uzavřeném prostoru: diplomová práce*. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií,Ústav telekomunikací, 2013.