



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická

Katedra radioelektroniky

**System pro detekci impulzních vad mechanického analogového
záznamu využívající model slyšení**

**System for Impulse Noise Detection in Gramophone Records using
Human Hearing Model**

Diplomová práce

Studijní program: Elektronika a komunikace

Studijní obor: Audiovizuální technika a zpracování signálů

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

Bc. Marek Semanský

Praha 2019

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Semanský** Jméno: **Marek** Osobní číslo: **434696**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávací katedra/ústav: **Katedra radioelektroniky**
Studijní program: **Elektronika a komunikace**
Studijní obor: **Audiovizuální technika a zpracování signálů**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Systém pro detekci impulzních vad mechanického analogového záznamu využívající model slyšení

Název diplomové práce anglicky:

System for Impulse Noise Detection in Gramophone Records using Human Hearing Model

Pokyny pro vypracování:

Seznamte se s metodami detekce impulzních vad ve zvukovém signálu. Zaměřte se na algoritmy, které berou do úvahy způsob zpracování zvuku v lidském uchu. Vybrané metody implementujte a porovnejte jejich výsledky - z hlediska úspěšnosti detekce i výpočetní náročnosti. Metodu s nejvhodnějšími vlastnostmi důkladně ověřte a optimalizujte pro detekci poruch v analogovém záznamu.

Seznam doporučené literatury:

- [1] Godsill, S, Rayner, P. Digital Audio Restoration. Springer-Verlag, London, 1998.
- [2] Rund, F., Vencovský, V., Bouše, J.: Detection of Clicks in Analog Recordings Using Peripheral-Ear Model. In: 19th International Conference on Digital Audio Effects (DAFx16). Brno. 2016, pp. 195-199.

Jméno a pracoviště vedoucí(ho) diplomové práce:

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

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

Datum zadání diplomové práce: **12.02.2019**

Termín odevzdání diplomové práce: **24.05.2019**

Platnost zadání diplomové práce: **20.09.2020**

Ing. František Rund, Ph.D.
podpis vedoucí(ho) práce

prof. Mgr. Petr Páta, Ph.D.
podpis vedoucí(ho) ústavu/katedry

prof. Ing. Pavel Ripka, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

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

Datum převzetí zadání

Podpis studenta

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

V Praze dne

.....

Poděkování

Děkuji svému vedoucímu práce, panu Ing. Františkovi Rundovi, Ph.D., za cenné podněty, rady a odborné připomínky při tvorbě této práce. Děkuji také společnosti GZ Media, a.s. za poskytnutí zvukových vzorků potřebných pro zpracování testů.

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

Abstrakt

Tato práce se zabývá detekcí impulzních degradací v analogovém záznamu zvuku. Jsou zde popsány typické modely pro detekci těchto vad. Dále jsou představeny modely slyšení, které jsou implementovány do pěti detekčních algoritmů v prostředí Matlab. Tyto algoritmy jsou testovány na signálech extrahovaných z degradovaných gramofonových desek. Je vyhodnocena jejich úspěšnost detekce a výpočetní náročnost. Dva algoritmy vykazují slibné výsledky s nízkou výpočetní náročností. Algoritmy dosahují až 94% úspěšných detekcí a pouze 5% falešných detekcí. V porovnání se současně preferovanými systémy tyto detekční metody vykazují srovnatelné či lepší výsledky. Díky těmto výsledkům se nabízí i možnost využití v reálném detekčním systému.

Klíčová slova

Analogový záznam zvuku, impulzní šum, detekce vad, model slyšení, ERBletová transformace, Matlab

Abstract

This thesis is devoted to detection of impulse noise in analog records. There are several methods for detecting this sort of degradation described in this paper. Furthermore models simulating signal processing in the human auditory system are presented. Based on these models, five detection algorithms are suggested and implemented in Matlab. These algorithms are tested on sound samples extracted from degraded gramophone records. Evaluation of performance and computational demands is presented. Two of these algorithms provide promising results with low computational requirements. Up to 94% of degradations are detected correctly and only about 5% of false detection occurred in the tests. Compared to currently preferred methods, these provide equal or better results. Due to these statistics, there is a potential for use in a real detection system.

Key words

Analog records, impulse noise, defect detection, human auditory system, ERBlet transform, Matlab

Obsah

Seznam obrázků	9
Seznam tabulek.....	10
1. Úvod.....	11
2. Záznam zvuku.....	12
2.1. Záznam zvuku na gramofonové desky	12
2.2. Proces výroby gramofonových desek	13
2.3. Vady gramofonových desek.....	13
3. Metody detekce impulzních vad.....	16
3.1. Detekce vad s využitím AR modelu	16
3.2. Rozšíření AR modelu o přizpůsobený filtr.....	17
3.3. Detekční metoda využívající vlnkové transformace	17
3.4. Detekce vad pomocí hydrodynamického modelu vnitřního ucha	18
3.5. Další metody	18
4. Modely slyšení	19
4.1. Dual Resonance Non-Linear	19
4.2. ERBletová transformace	21
4.3. Model slyšení dle S. Seneff	23
4.4. Kochleární model R. F. Lyona	25
5. Návrh a implementace detekčních algoritmů.....	27
5.1. Detekční algoritmy využívající model DRNL.....	27
5.2. Metoda detekce využívající ERBletové transformace.....	29
5.3. Detekční algoritmus implementující model R. F. Lyona	30
5.4. Detekce vad pomocí modelu S. Seneff	31
6. Testování vzorků	32
6.1. Testovací množiny vzorků	32
6.2. Popis testů.....	33
6.3. Test vlivu parametrů jednotlivých algoritmů.....	35
6.3.1. DRNL.....	35
6.3.2. Lopez-Poveda	37
6.3.3. ERBlet	39
6.3.4. Lyon	41
6.3.5. Seneff	43
6.3.6. Výsledky pro druhou rozhodovací úroveň	45
6.4. Porovnání modelů s komerčním softwarem.....	45
6.5. Test vlivu dynamiky signálu	48
6.6. Test citlivosti detekce a schopnosti lokalizace vad	50
6.7. Porovnání s dalšími detekčními metodami.....	52
7. Vyhodnocení výsledků	53
8. Závěr	55
Reference	56
Příloha	58

Seznam obrázků

Obrázek 1: Stereofonní záznam zvuku [1]	12
Obrázek 2: Poškrábaná deska [5].....	14
Obrázek 3: Nedolisovaná deska [5].....	14
Obrázek 4: Blokové schéma DRNL filtru [14]	19
Obrázek 5: Data z experimentů s činčilou (A, C) a vstupně výstupní funkce modelu (B, D) [15]	20
Obrázek 6: Spektrogram STFT (A) a ERBletové transformace signálu s impulzní vadou (B).....	22
Obrázek 7: Přiblížený spektrogram STFT (A) a ERBletové transformace signálu s impulzní vadou (B) 22	
Obrázek 8: Blokové schéma modelu S. Seneff [20]	23
Obrázek 9: Popis filtrů v komplexní rovině a jejich frekvenční charakteristiky [20].....	25
Obrázek 10: Kaskádní zapojení filtrů [21]	25
Obrázek 11: Blokové schéma demodulace a AGC [21]	26
Obrázek 12: Přenosové charakteristiky DRNL filtrů (kombinace pásmové a dolní propusti).....	28
Obrázek 13: Korelace signálu s impulzní odezvou (kanál 20) pro segment bez clicku a s clickem.....	28
Obrázek 14: Schéma automatického ovládní zisku [19]	30
Obrázek 15: Porovnání detekčního signálu s prahem	31
Obrázek 16: Výsledky subjektivních testů pro dvě rozhodovací úrovně	32
Obrázek 17: Grafické rozhraní pro detekci a korekci vad	33
Obrázek 18: DRNL - vliv délky segmentu impulzní odezvy	35
Obrázek 19: DRNL - vliv hodnoty konstanty pro násobení detekčního prahu.....	36
Obrázek 20: DRNL - vliv volby kanálů.....	36
Obrázek 21: Lopez-Poveda - vliv délky segmentu impulzní odezvy.....	37
Obrázek 22: Lopez-Poveda - vliv hodnoty konstanty pro násobení detekčního prahu	38
Obrázek 23: Lopez-Poveda - vliv volby kanálů	38
Obrázek 24: ERBlet - vliv délky okna mediánového filtru	39
Obrázek 25: ERBlet - vliv hodnoty konstanty pro násobení prahu	40
Obrázek 26: ERBlet - vliv volby kanálů	40
Obrázek 27: Lyon - vliv délky segmentu impulzní odezvy.....	41
Obrázek 28: Lyon - vliv konstanty pro násobení prahu	42
Obrázek 29: Lyon - vliv volby kanálů.....	42
Obrázek 30: Seneff - vliv délky segmentu impulzní odezvy.....	43
Obrázek 31: Seneff - vliv použitých kanálů a hodnoty pro násobení obálky signálu	44
Obrázek 32: Seneff - vliv délky okna mediánového filtru	44
Obrázek 33: Nedetekovaný click v prostředí Wavelab Pro.....	46
Obrázek 34: Falešné detekce bicích nástrojů detekované programem Wavelab Pro	46
Obrázek 35: Výstupy detekčních algoritmů pro signál s bicími nástroji	46
Obrázek 36: Falešná detekce v prostředí Wavelab Pro	47
Obrázek 37: Výstupy detekčních algoritmů.....	47
Obrázek 38: Kompresní funkce	48
Obrázek 39: Změna signálu vlivem kompresní funkce.....	48
Obrázek 40: Vliv komprese na detekci vad	49
Obrázek 41: Znázornění hudebního a šumového signálu	50
Obrázek 42: Detekce vad u korelačních algoritmů	51
Obrázek 43: Detekce vad u ERBletové transformace	51

Seznam tabulek

Tabulka 1: Typické vady gramofonových desek a jejich vlastnosti [2]	15
Tabulka 2: Vhodná volba parametrů detekčních algoritmů	45
Tabulka 3: Výsledky detekčních algoritmů pro druhou rozhodovací úroveň	45
Tabulka 4: Konfigurace použité pro detekci anomálií v softwaru Wavelab Pro	45
Tabulka 5: Statistické výsledky detekce v programu Wavelab Pro.....	47
Tabulka 6: Výsledky detekce po kompresi signálů (práh 0,3; poměr 1/3).....	49
Tabulka 7: Počet detekovaných vad	50
Tabulka 8: Výpočetní náročnost algoritmů.....	51
Tabulka 9: Souhrn výsledků pro detekční algoritmy testované v [8].....	52
Tabulka 10: Souhrn výsledků pro detekční algoritmus využívající hydrodynamický model slyšení [9].....	52
Tabulka 11: Účinnost detekce a výpočetní náročnost navržených algoritmů	53

1. Úvod

S rostoucí popularitou analogového záznamu zvuku stoupají i nároky výrobců na kvalitu jejich produktů. Společnosti, jako například GZ Media, a.s., se při výrobě mechanického záznamu potýkají s mnoha komplexními kroky, které, při nedokonalém postupu, mohou vyústit v degradovaný produkt. Při výrobě gramofonových desek takto mohou vzniknout například vady impulzního charakteru (clicky). Právě impulzní šum je typickou vadou mechanického záznamu zvuku, kterou je potřeba včas detekovat a eliminovat pro dosažení větší efektivity výroby a kvality produktů. Impulzní šum není problémem jen při výrobě gramodesek, ale např. i při restaurování archivních záznamů.

Tato práce je zaměřena na detekci takovýchto impulzních degradací. Jejím cílem je seznámit čtenáře s modely pro detekci impulzního šumu ve zvukových signálech se zaměřením na algoritmy, které pracují s modelem lidského slyšení. Dále také implementovat a otestovat tyto techniky z hlediska úspěšnosti detekce a výpočetní náročnosti.

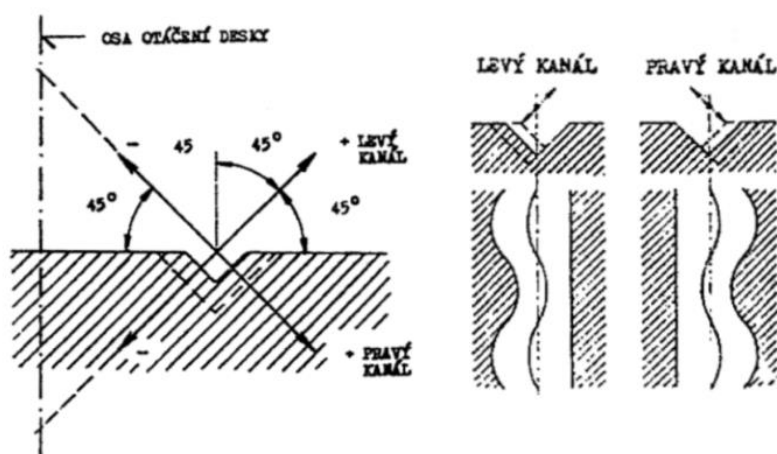
První část práce se zabývá teoretickým pozadím týkajícím se mechanického záznamu zvuku, postupem výroby gramofonových desek a popisem častých zástupců impulzních degradací. Ve 3. kapitole jsou představeny typické metody detekce clicků. Na metodu využívající model slyšení je dále v práci navázáno. V kapitole 4 jsou představeny jiné modely slyšení, které lze použít pro detekci vad a potenciálně vylepšit efektivitu detekce. Tyto modely popisují zpracování zvuku ve vnitřním uchu a vyhodnocují například rychlost výchylky bazilární membrány či pravděpodobnost vzruchů nervových zakončení. Je zde také představena možnost časově frekvenční reprezentace signálu založena na frekvenční analýze signálu uvnitř kochley. Na základě těchto modelů jsou dále implementovány (kapitola 5) detekční algoritmy, které vyhodnocují, zda vstupní signál obsahuje šum impulzního charakteru. Algoritmy jsou optimalizovány pro detekci slyšitelných vad v analogovém záznamu. V 6. kapitole jsou provedeny čtyři testy zkoumající úspěšnost detekce, výpočetní náročnost či stabilitu algoritmů při změnách vstupního signálu. Algoritmy jsou porovnány s komerčním softwarem, je testována jejich citlivost detekce a také je vyhodnocena jejich schopnost lokalizace degradací.

2. Záznam zvuku

Pro správnou orientaci v problematice detekce vad u mechanického záznamu zvuku je nutné znát podrobně veškeré aspekty související se záznamem zvukového signálu, produkcí kopií záznamových médií a reprodukce těchto médií. Každý z těchto kroků reprezentuje komplexní úlohu, která, v případě nepřesného postupu, může vyústit ve vadný finální produkt. Výrobní procesy nejsou ovšem nikdy dokonalé, a proto je nutné provádět důkladnou analýzu jednotlivých kroků ke zjištění příčiny vad ve finálním produktu. Tyto vady mohou být způsobeny mechanickými a fyzikálními vlastnostmi materiálů, subjektivní chybou obsluhy či nepřesností výrobních strojů. Analýza těchto aspektů nám může pomoci k pochopení vzniku chyb, jejich vlastností a dokonce i k optimalizaci detekčních algoritmů.

2.1. Záznam zvuku na gramofonové desky

Princip záznamu zvuku na gramodesky spočívá ve vychylování spirálovité drážky vůči její klidové poloze [1]. V minulosti byl záznam hloubkový. Drážka byla vychylována do hloubky materiálu. Později se přešlo na záznam stranový, kdy je drážka vychylována do strany s konstantní hloubkou. Tato výchylka je prováděna nožem, který ryje do materiálu a je úměrná okamžité hodnotě akustického tlaku zaznamenávané akustické vlny. Dnes je standardem stereofonní záznam, kdy se používá jeden nůž ovládaný dvěma akčními členy. Tyto členy svírají s deskou úhel 45 stupňů a oba zvukové kanály zapisují do jiné strany drážky (obr. 1). Pravý kanál je zaznamenán do vnější strany drážky. Při stejných amplitudách, frekvencích a fázi signálů v obou kanálech vzniká pouze stranový záznam [1]. V případě např. různé amplitudy může vzniknout i hloubková složka záznamu. Na obrázku 1 jsou drážky stereofonního stranového záznamu. Kompatibilita pro monofonní reprodukci je vyřešena použitím systému MS, kde M reprezentuje monaurální signál a složka S reprezentuje informaci o směrovém rozdělení do dvou kanálů. Součtem a rozdílem M a S složek lze získat dva binaurální signály.



Obrázek 1: Stereofonní záznam zvuku [1]

Pro snímání zvuku z gramofonových desek se využívají mechanicko-elektrické měniče založené na primárně dvou principech. Prvním z nich je piezoelektrický, který generuje výstupní napětí na okamžité výchylce přenosky a druhým je princip magnetodynamický. Tento princip je rychlostní a vyžaduje korekční předzesilovač. I přesto se v praxi využívá častěji než měnič piezoelektrický. Snímání zvuku je velmi precizní záležitost. Drážky jsou v průměru 60 μm široké a při záznamu referenčního sinusového signálu 0 dB do masteru desky vzniká výchylka v drážce 22,5 μm . Již z těchto čísel je patrné, že sebemenší nepřesnost v záznamu může způsobit akustické rušení při čtení záznamu. Tyto rozměry mají za následek, zejména u starších gramodesek, také omezené kmitočtové pásmo a relativně malý odstup signálu od šumu.

2.2. Proces výroby gramofonových desek

Prvním krokem při výrobě gramofonových desek je úprava samotných hudebních souborů dodaných zákazníkem [2]. Jedná se o mastering, ve kterém jsou nahrávky poskládány do správného pořadí, jsou vloženy mezery mezi skladby či např. prodlouženy nebo zkráceny skladby. Následuje mastering, který je velice důležitý zejména z hlediska potlačení zkreslení a zabránění poškození mechaniky výrobních strojů. Mistr zvuku má za úkol provést takové korekce, aby bylo možné skladbu vyřezat do desky a zároveň nebyl změněn charakter skladby. Součástí masteringu je také nastavení šířky drážky, zrcátka (rozteče mezi drážkami) a celkové hlasitosti nahrávky. Tato fáze výroby zahrnuje první kontrolu chyb.

Výstup masteringu přebírá technik zodpovědný za řezání desky. Dle zvolené techniky je řezání provedeno do měděného plechu či do vrstvy laku. Metoda řezání do plechu se nazývá direct metal mastering [3]. Tato metoda spočívá v přenosu dvou zvukových kanálů do dvou navzájem kolmých cívek, které svým pohybem ovládají řezací stylus, na jehož konci je broušený diamantový nůž tvaru písmene v. Nůž řeže do nemagnetického hladkého povrchu pokrytého tenkou vrstvou vysoce kvalitní mědi. Předlohy pro lisování desek jsou z těchto masterů vytvořeny elektrolytickým procesem. Výhodou této metody je vyšší přesnost záznamu vysokofrekvenčních složek, nižší úroveň šumu a omezení přeslechů mezi drážkami. Druhá metoda stojí na principu řezání do vrstvy laku. U této metody se používá nůž z broušeného safíru či rubínu, který je nahřátý pro snazší řezání. Pro vytvoření matrice je nutný trojnásobný elektrolytický proces. Mezi výhody této metody patří širší a hlubší drážky, které umožňují lepší reprodukci nízkofrekvenčních složek a jsou zároveň odolnější vůči jejich mechanickému poškození. Prvním krokem při řezání je provedení testovacího řezu, kde je zkontrolována šířka drážky, řez a šumové vlastnosti. Z měření šumu lze rozpoznat špatný plech či zhoršující se vlastnosti řezacích nožů. U druhé popsané metody se v tento moment definuje nahřívání nožů, které se musí během řezání měnit. V průběhu řezání je v reálném čase poslouchán záznam za účelem detekce vad v masteru či chyby při řezání. Poté probíhá jednoduchý či trojitý elektrolytický proces pro vytvoření matrice pro lisování desek.

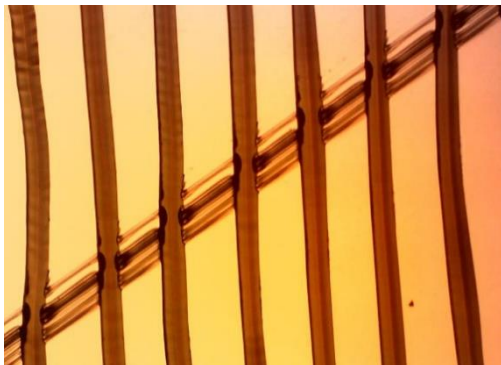
Po vylisování desek je provedena hlavní kontrola kvality produktu. Tato kontrola má tři fáze - informační, optickou a akustickou. Informační kontrola spočívá v kontrole průvodního listu kontroly. Tedy například matričního čísla, správnosti vytištění etikety apod. Optická kontrola se snaží odhalit vady mechanického charakteru. Kontrola je prováděna manuálně obsluhou tak, že se kontrolují odlesky světla od povrchu desky a hledají se nepřesnosti. K optické kontrole je využíván i mikroskop. Poslední kontrolou je ta akustická, která je námětem pro tuto práci. Deska je opět poslouchána školenou obsluhou. Jedná se tedy o subjektivní testy, které mohou odhalit pouze slyšitelné vady. Tyto vady jsou následkem např. nečistot v drážce, škrábance či špatného vylisování desky. Nejvíce rušivé jsou impulzní vady označované jako lupance či clicky, které se často vyskytují pravidelně v každé otáčce. Dle současných praktik ve výrobě firmy GZ Media a.s. jsou některé slyšitelné vady ignorovány a ponechány na finálním produktu, který obdrží zákazník. Tyto praktiky jsou brány v potaz při návrhu a realizaci algoritmů pro automatizovanou detekci vad na vinylových deskách. Kontrola je prováděna i porovnáním výsledného produktu s masterem, který vady neobsahuje. Tato metoda je velice efektivní a přináší dobré výsledky. Ovšem ne vždy je možné toto porovnání provést, např. z důvodu špatné synchronizace signálů či v případě nedostupného masteru a je nutné aplikovat algoritmy, které detekují vady pouze při znalosti jednoho signálu.

2.3. Vady gramofonových desek

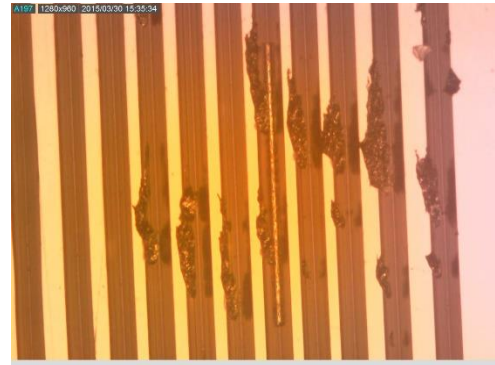
Dle předchozích kapitol je zřejmé, že se chyba může objevit v jakékoliv fázi výroby a může mít téměř jakýkoliv projev. Obecně se dají vady gramofonových desek rozdělit do čtyř kategorií. Jedná se o informační, mechanické, optické a akustické vady. Informační vady jsou vždy způsobeny lidským

faktorem. Například vložení špatné etikety, vynechání skladby apod. Mechanické vady vznikají v procesu lisování. Jde o křivost, zvlnění nebo špatnou hmotnost desky. Optické vady jako skvrny a nečistoty mohou vznikat téměř ve kterékoliv fázi výroby a mají pouze optický projev. Pro posluchače nejpodstatnější jsou vady akustické. Tyto vady mohou vznikat již u zákazníka před procesem výroby gramofonové desky, v jakékoliv fázi výroby či později při manipulaci s hotovou deskou. Právě akustické vady jsou zajímavé pro tuto práci. Akustické vady lze dle [4] rozdělit do kategorií globálních a lokálních. Globální vady, jako např. širokopásmový šum, chvění či zkreslení zvuku, se projevují v celém signálu. Lokální vady jsou nespojitosti v signálu s projevem pouze v některých vzorcích (např. lupance). Předmětem této práce jsou pouze lokální vady. Rozlišujeme dva typy lokálních degradací – clicky a nízkofrekvenční šumové přechody. Clicky jsou velmi krátké skoky signálu náhodné v čase a amplitudě, které se projevují jako lupance nebo prasknutí. Typická délka clicku je 1 až 5 milisekund, ovšem dle [5] může následovat doznívání dlouhé až 50 ms. Nízkofrekvenční šumové přechody jsou dle [4] generovány většími škrábanci a deformacemi na čteném médiu a mají tedy delší projev. Tyto deformace excitují nízkofrekvenční rezonance akusticky se projevující jako dupnutí či úder.

Subjektivní kontrola akustických vad je prováděna v několika krocích výroby a je snaha tuto část automatizovat. S využitím automatizace by mohlo dojít k zefektivnění detekce vad a zrychlení výrobního procesu. Automatizace má velký potenciál i v možnostech klasifikace vad a také identifikace špatně provedeného kroku výroby. Tabulka 1, upravená z [5], popisuje základní impulzní degradace signálů získaných z poškozených gramofonových desek. Obrázky 2 a 3 znázorňují mechanické poškození gramofonové desky pod mikroskopem. První z obrázků ukazuje typický škrábanec na desce, který degraduje oba zvukové kanály. Většina impulzních degradací má akustický projev v obou kanálech. Chyby způsobené například nedolisováním desky, viz obrázek 3, mají většinou akustický projev pouze v levém z kanálů.



Obrázek 2: Poškrábaná deska [5]



Obrázek 3: Nedolisovaná deska [5]

Tabulka 1: Typické vady gramofonových desek a jejich vlastnosti [2]

Název vady	Popis příčiny vzniku a projevu vady
Narážka	Obecná porucha, která se projevuje lupancem. Jeho intenzita je různá. Důvodem této poruchy může být: elektrostatický lupanec, nečistota v drážce, prořez, přeskok, škrábanec a dokonce i chyba v podkladu od zákazníka.
Elektrostatický lupanec	Tato vada se projevuje silným lupancem. Když však dané místo přehrajeme, znovu na daném místě ho nepozorujeme vůbec nebo se přemístí na jiné místo (přesun náboje). Deska v takovém případě obsahuje silný statický náboj, po přehrání se náboj vybijí a vznikají tak lupance. Tyto lupance jsou typické svým akustickým projevem, můžeme slyšet velmi ostrý zvuk.
Nečistota v drážce	Cizí částice, která je umístěna v drážce. Přenoska do ní narazí a dojde k akustickému projevu. Rozlišujeme dva typy nečistot. První nečistota je taková, která po přehrání přenoskou zmizí. Přenoska jí vytlačí z drážky pryč. Druhý typ nečistoty je takový, že nečistota zůstane v drážce i po přejetí daného místa. Zde se jedná o nečistotu, která je zalisována v drážce.
Poškrábaná matrice	Při škrábnutí desky můžeme mít štěstí, že se poškodí pouze zrcátko a deska se pak nebude projevovat lupancem na otáčku. U škrábnuté matrice takové štěstí již nemáme. Matrice je totiž negativ a jakýkoliv škrábanec poškodí výstupky, z kterých jsou po vylisování dna drážek. Důsledkem toho pak vznikají různé vady, jako jsou například lupance na otáčku, samostatný lupanec a praskot. Tato vada je způsobena především špatnou manipulací s matricí.
Nedolisovaná deska	Tato vada se především projevuje rušivým praskáním v jednom kanálu. Nedolisy jsou fyzicky umístěny v závětrí hran matrice. Ve většině případů je vada umístěna přes několik drážek a praskání tak můžeme slyšet na otáčku. Příčinou je lisování, kde hmota při lisování nedostatečně obteče matrici. Tento problém se obvykle objevuje ke konci stran. Opticky můžeme nedolisy pozorovat jako shluky bílých teček.
Prořez	Špatně nastavené parametry od operátora nebo porucha řezací aparatury způsobí, že se drážky proříznou do sebe. Důsledkem toho může dojít k přeskoku, praskotu, lupanci nebo k zacyklení celé desky. Vznikne tak nekonečná smyčka.
Smeť	Smeť je vada, která ve svém důsledku vzniká již při přípravě na lisování. Příčinou je nečistota usazená na matrici, která se obtiskne na výsledné desky. Opticky se smetě projevuje jako malý důlek. Akusticky se smetě, projevuje zvukem obsahujícím nízké frekvence (bouchání).
Praskot	Posloupnost lupanců s různou intenzitou. Praskot může být způsoben jak poruchami z galvanoplastiky, tak i kvůli poruchám z lisování.

3. Metody detekce impulzních vad

Některé detekční algoritmy jsou již dobře známé a důkladně otestované. Jsou to například metody využívající AR modelu, jeho rozšíření přizpůsobeným filtrem, vlnkové transformace a hydrodynamického modelu ucha, které byly publikovány v [6] [7] [8] [9]. Tyto detekční modely jsou zde v krátkosti popsány kvůli porovnání výsledků a výhod či nevýhod jednotlivých metod. Prostor pro jejich vylepšení se nabízí v oblasti úspěšnosti detekce i výpočetní náročnosti. Některé z těchto algoritmů detekují všechny impulzní vady bez ohledu na jejich slyšitelnost. V různých aplikacích je ovšem žádané detekovat pouze slyšitelné degradace.

3.1. Detekce vad s využitím AR modelu

Metoda autoregresního modelu je založena na odhadu amplitudy vzorku signálu váhovým součtem p předchozích vzorků. Tento model byl navržen a popsán v [10]. Princip spočívá v neschopnosti predikce impulzních zvuků a následném určení rozdílů mezi původním predikovaným signálem a jeho predikcí. Metoda předpokládá, že je vstupní signál s_t porušen aditivním šumem n_t

$$x_t = s_t + n_t. \quad (1)$$

Odhad vstupního signálu s_t lze zapsat rovnicí

$$s_t = - \sum_{k=1}^p a_k s_{t-k} + e_t, \quad (2)$$

kde a_k jsou koeficienty AR modelu získané řešením Yule-Walkerových rovnic, e_t je bílý šum s nulovou střední hodnotou a p je řád predikce. Ze dvou předchozích rovnic můžeme vytvořit rovnice

$$x_t = - \sum_{k=1}^p a_k s_{t-k} + e_t + n_t, \quad (3)$$

$$x_t = - \sum_{k=1}^p a_k (x_{t-k} - n_{t-k}) + e_t + n_t. \quad (4)$$

Z rovnice 4 lze vyjádřit detekční signál d_t jako součet známého (porušeného) signálu a jeho predikce:

$$d_t = x_t + \sum_{k=1}^p a_k x_{t-k} = e_t + n_t + \sum_{k=1}^p a_k n_{t-k} + e_t + n_t. \quad (5)$$

Detekční signál se tedy skládá z bílého šumu, impulzního šumu a efektu předchozích p vzorků. Levá strana rovnice 5 odpovídá detekčnímu signálu získanému filtrací porušených dat filtrem

$$H(z) = 1 + \sum_{k=1}^p a_k z^{-k}. \quad (6)$$

Předpokladem pro detekci je konstantně nízká amplituda bílého šumu v celém signálu a vysoká amplituda impulzního šumu.

3.2. Rozšíření AR modelu o přizpůsobený filtr

Rozšíření autoregresivní metody navržené v [11] umožňuje zvýraznění impulzního šumu, nicméně zhoršuje časovou lokalizaci. Toto rozšíření spočívá v implementaci přizpůsobeného filtru, který má opačnou impulzní odezvu oproti filtru popsanému v předchozí kapitole. Celkový filtr lze tedy zapsat jako $H(Z)H(Z^{-1})$. Jedná se o nekauzální filtr s $2p+1$ koeficienty. Výhody tohoto filtru jsou znatelné pouze pro určitý typ impulzní poruchy. Silně závisí na délce intervalů mezi clicky a na délce trvání jednotlivých degradací.

Důležitými parametry pro metody založené na autoregresivní predikci signálu jsou:

- Délka segmentu zvukového signálu
- Řád predikce
- Detekční práh (konstanta k)

Testování vhodné volby těchto parametrů bylo provedeno v [8].

3.3. Detekční metoda využívající vlnkové transformace

Vlnková transformace (Wavelet Transform [12]) v diskrétním čase dělí signál na aproximační a detailní část za pomoci kvadratických filtrů. Signál je filtrován hornopropustným a dolnopropustným filtrem a následně decimován. Výstup dolní propusti je dále filtrován dvěma filtry. Toto lze iterativně opakovat do více úrovní. Z těchto větví můžeme získat informace o frekvenčním složení signálu a zároveň o časové lokalizaci jednotlivých složek. S využitím znalostí o časovém a frekvenčním charakteru impulzních vad byl v [13] navržen detekční algoritmus využívající vlnkovou transformaci. Princip je založen na porovnání koeficientů transformace s detekčním prahem. Při detekci se předpokládá, že koeficienty transformace budou podstatně vyšší v místech s výskytem impulzního šumu než v místech širokopásmového šumu.

Detekční práh je v [13] definován jako

$$\Gamma(n, s) = k \text{Env}[|Wf(n, s)|], \quad (7)$$

kde $Wf(n, s)$ jsou koeficienty transformace signálu $f(n)$ v úrovni s , k je empiricky určená konstanta a Env je operátor pro určení obálky tak, aby nebyla narušena impulsy. Navrhovanou metodou pro určení obálky je použití mediánového filtru vhodné délky okna na koeficienty transformace. Obdobně jako u předchozích metod je chyba detekována tam, kde platí

$$|Wf(n, s)| > \Gamma(n, s). \quad (8)$$

Důležité parametry pro tuto metodu a jejich vliv na detekci byl testován v [8]. Jedná se o:

- Délka okna mediánového filtru
- Hodnota konstanty k
- Typ vlnky
- Řád vlnky

3.4. Detekce vad pomocí hydrodynamického modelu vnitřního ucha

Předešlé detekční metody se vztahují na všechny impulzní degradace bez ohledu na jejich akustický projev. V některých případech zpracování zvuku jsou zajímavé pouze slyšitelné vady. Z toho to důvodu byl v [9] navržen detekční algoritmus využívající modelu slyšení. Autoři zde využívají komplexního hydrodynamického modelu, který byl v [14] navržen pro studie otoakustických emisí ve vnitřním uchu. V tomto detekčním modelu je využito silné odezvy bazilární membrány na impulzní zvuky, zejména ve vyšších frekvenčních pásmech. Teoretickým předpokladem pro detekci je podoba impulzní odezvy bazilární membrány a její odezva na impulzní šum [9]. Porovnáním korelací těchto odezev s detekčním prahem je možno vyhodnocovat přítomnost impulzních degradací.

Obdobně jako ostatní modely i tento převádí vibrace třmínku na vibrace segmentů bazilární membrány uvnitř kochley. Tento proces je aproximován oscilátory vzájemně vázanými okolní kapalinou. Výstupem modelu je výchylka jednotlivých oscilátorů v čase. Celý výpočet lze popsat pohybovou rovnicí bazilární membrány. Popis metody je převzat z [9]. Výchylka ξ_i i -tého oscilátoru je dána vztahem

$$m_i \ddot{\xi}_i(t) + h_i \dot{\xi}_i(t) + s_i [2\dot{\xi}_i(t) - \dot{\xi}_{i-1}(t) - \dot{\xi}_{i+1}(t)] + k_i \xi_i(t) = f_{H_i}(t) + f_{OHC_i}[\eta_i(t)], \quad (9)$$

kde m_i , h_i , s_i a k_i jsou hmotnost, lokální viskozita, sdílená viskozita a tuhost bazilární membrány. Hydrodynamická síla f_{H_i} pohybující oscilátory je definována rovnicí

$$f_{H_i}(t) = -G_{S_i} a_{S_i}(t) - \sum_{j=1}^N G_i^j \ddot{\xi}_i(t), \quad (10)$$

kde $a_{S_i}(t)$ je zrychlení třmínku, $\ddot{\xi}_i(t)$ je zrychlení oscilátoru, G_{S_i} a G_i^j jsou přenosové funkce popisující viskózní vazbu mezi třmínkem a bazilární membránou. Člen f_{OHC_i} popisuje sílu působící na zevní vláskové buňky a je funkcí jejich odrazivosti η_i definované takto:

$$\bar{m}_i \ddot{\eta}_i(t) + \bar{h}_i \dot{\eta}_i(t) + \bar{k}_i \eta_i(t) = -D_i \ddot{\xi}_i(t), \quad (11)$$

kde D_i je konstanta. Z odrazivosti vláskových buněk lze vypočítat f_{OHC_i} pomocí sigmoidní funkce S :

$$f_{OHC_i}[\eta_i(t)] = -aS[b\eta_i(t)], \quad (12)$$

kde a a b jsou vypočteny z vybraných vlastností Cortiho orgánu [14].

3.5. Další metody

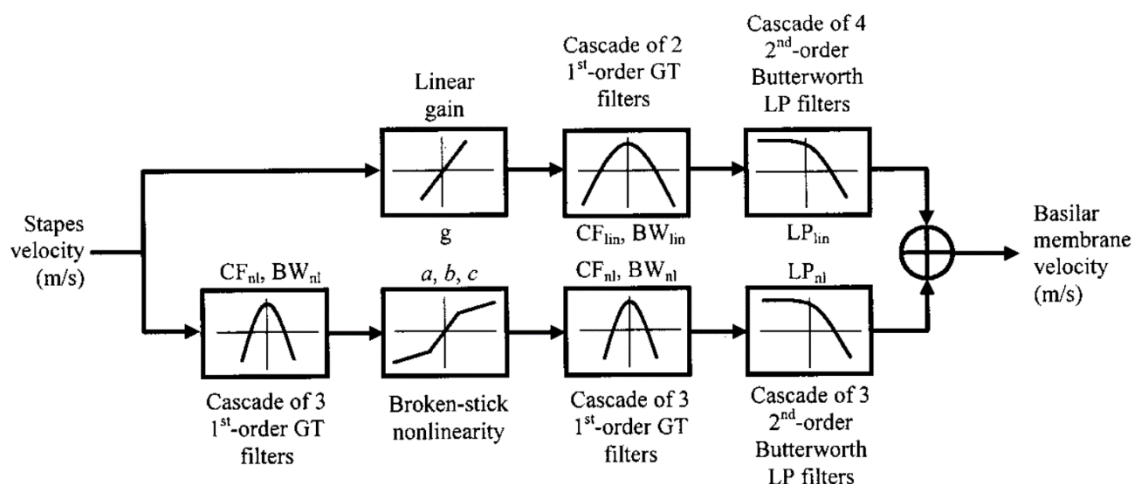
Mezi další metody patří například použití jednoduchého filtru typu horní propust, který ponechá pouze vysokofrekvenční složky signálu, jejichž častým zástupcem jsou impulzní šумы. Možnost detekce těchto degradací nabízí i filtrace signálu průměrovacím či mediánovým filtrem. Tyto filtry jsou schopny odstranit rapidní změny signálu a ponechat pouze hladkou obálku. Porovnáním této obálky s původním signálem lze nalézt místa s impulzním charakterem. Další možností je využití derivace signálu, která naopak zvýrazní náhlé změny [7]. Opakovanou derivací signálu lze tento efekt znásobit a získat hladký signál s pulzním charakterem pouze v místě clicků. Takovýto signál lze porovnat s vyhlazenou obálkou původního signálu a získat informaci o poloze šumu.

4. Modely slyšení

Metody pro detekci impulzních vad nově popsané a implementované v této práci se snaží modelovat vlastnosti lidského sluchu a zpracování zvukového signálu lidským sluchovým ústrojím. Literatura [9] navrhuje detekci slyšitelných vad korelací impulzní odezvy modelu kochley s odezvou na degradovaný signál. Tento model kochley je ovšem výpočetně náročný a detekci nelze použít v reálném čase. V následujících kapitolách jsou představeny další modely zpracování zvuku v uchu. Tyto modely nabízí simulace sluchového ústrojí obdobně jako hydrodynamický model ovšem s nižší výpočetní náročností, a proto je jím v této práci věnována pozornost. Je zde zaměřeno na modely, k nimž byla dostupná implementace v prostředí Matlab. Detekční signál je získáván z rychlosti vibrací bazilární membrány, synapsí sluchových nervů či z frekvenční analýzy signálů dle charakteristiky kochley.

4.1. Dual Resonance Non-Linear

U modelu pod zkratkou DRNL se jedná o banku filtrů, kterou autoři modelovali tak, aby simulovala excitaci bazilární membrány na základě výchylky třmínku vůči jeho klidové poloze. Princip funkčnosti a matematický popis modelu jsou převzaty z [15] a [16]. Snahou při modelování DRNL bylo vytvořit výpočetní model frekvenční selektivity lidského sluchového ústrojí využívající v té době preferované lineární gammatónové banky filtrů a zároveň znalosti o nelineárních aspektech lidského sluchu. Objektivní metrikou pro ověření přenosové charakteristiky systému byla data z experimentů prováděných na zvířecím sluchovém ústrojí (kochlea činčily). Vstupem tohoto modelu je rychlost pohybu třmínku $x(t)$ a výstupem je rychlost bazilární membrány $y(t)$ v definovaných frekvenčních pásmech podél kochley.



Obrázek 4: Blokové schéma DRNL filtru [14]

Každé pásmo systému je reprezentováno dvěma paralelními větvemi, jednou lineární a druhou nelineární. Blokové schéma obou větví je znázorněno na obrázku 4. Lineární větev se skládá z bloku lineárního zesílení, pásmové propusti a dolní propusti v sérii. Nelineární větev obsahuje bloky pásmové propusti, kompresní funkce, druhé pásmové propusti a dolní propusti. Výstupem DRNL jednotek jsou sumy jejich dvou větví v určitém frekvenčním pásmu. Parametry pásmových propustí CF (centrální frekvence) a BW (šířka pásma) jsou různé pro obě větve. Parametr LP (mezní frekvence) dolní propusti je nastaven na hodnotu centrální frekvence předcházející pásmové propusti. Obě pásmové propusti v nelineární větvi mají stejné parametry. Parametry filtrů a kompresních funkcí byly

odvozeny z testů na sluchovém ústrojí zvířat (čičila, morče). Tvar křivky kompresní funkce má dle [16] lineární průběh pro signály s nižší úrovní,

$$y_a = a \cdot x(t), \quad (13)$$

kde a určuje strmost křivky a $x(t)$ je výstup první pásmové propusti. Pro vyšší úrovně stimulu má kompresní funkce odezvu nelineární,

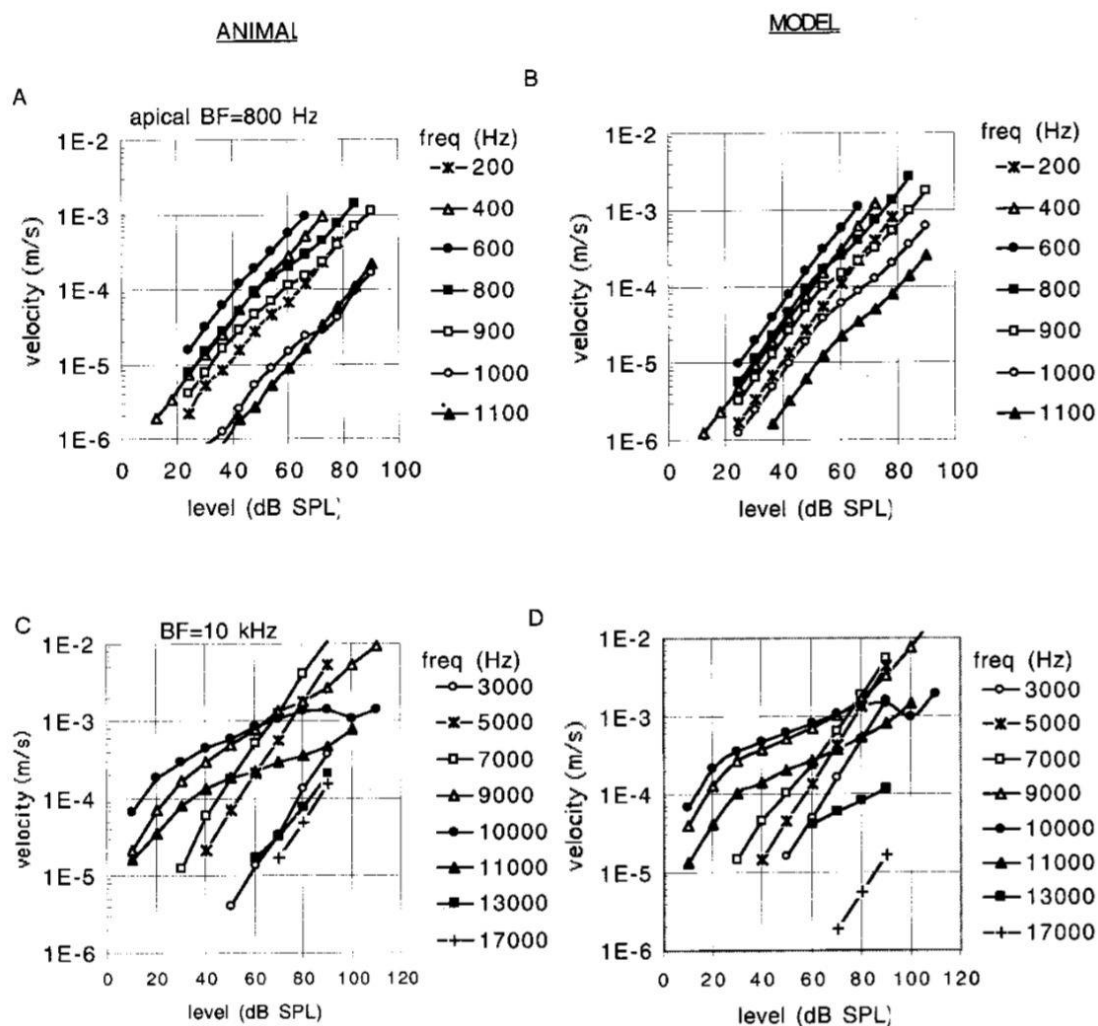
$$y_b = b|x(t)|^c \cdot \text{sign}(x(t)), \quad (14)$$

Kde konstanty b a c definují tvar křivky. Pro všechny stimuly je tvar výsledné křivky dán vztahem:

$$y = \text{sign}(x(t)) \cdot \min(a|x(t)|, b|x(t)|^c), \quad (15)$$

kde parametry a , b , c určují celkový zisk nelineární větve.

Celková charakteristika DRNL je po součtu obou paralelních větví přibližně lineární v nízkých úrovních. Při stimulu o frekvenci 800 Hz je to v úrovních zhruba do 60dB a opět přibližně od 75 dB výš. Pro stimuly o vyšších frekvencích je nelineární pásmo podstatně širší. Mezi těmito hladinami převládá nelineární větev, a proto je i celkový charakter filtru nelineární. Výsledky měření pro experimenty s čičilou a výstupy modelu jsou graficky znázorněny na obrázku 5. Vstupní hladina akustického tlaku je převoditelná na rychlost kmitání třmínku.



Obrázek 5: Data z experimentů s čičilou (A, C) a vstupně výstupní funkce modelu (B, D) [15]

4.2. ERBletová transformace

Zvukové signály jako řeč či hudbu je často výhodné zpracovávat ve frekvenčním spektru, tedy provádět jejich dekompozici na sady elementárních funkcí a pracovat s nimi. Většina aplikací pracuje s teorií založenou na Fourierově transformaci. Problém této transformace je konstantní časově frekvenční rozlišení. Pokud tedy použijeme transformaci odvozenou z Fourierovy, zjistíme přesnou polohu nebo přesné frekvenční rozlišení, ale nikdy nezískáme obojí najednou. Algoritmy jako je Fourierova či kosinová transformace jsou velice užitečné nástroje, nicméně jsou postaveny na matematickém základu bez bližšího ohledu na fyziologický princip vnímání zvuku. Pro analýzu hudby je často vhodné hledat jiné řešení, které více odpovídá parametrům slyšení a vnímání zvuku lidským sluchovým ústrojím. Časově frekvenční analýzu s ohledem na vjem zvuku nabízí ERBletová transformace.

Sluchové ústrojí lze modelovat pomocí banky pásmových propustí, jejichž šířky pásma odpovídají spektrálnímu rozlišení kochley. V případě ERBletové transformace se šířky pásma určují pomocí jednotky označované jako ERB, neboli equivalent rectangular bandwidth. ERB definuje šířku kritického pásma [17] v Hz v závislosti na frekvenci dle vztahu z [18]

$$ERB(f) = 24,7 + \frac{f}{9,265}, \quad (16)$$

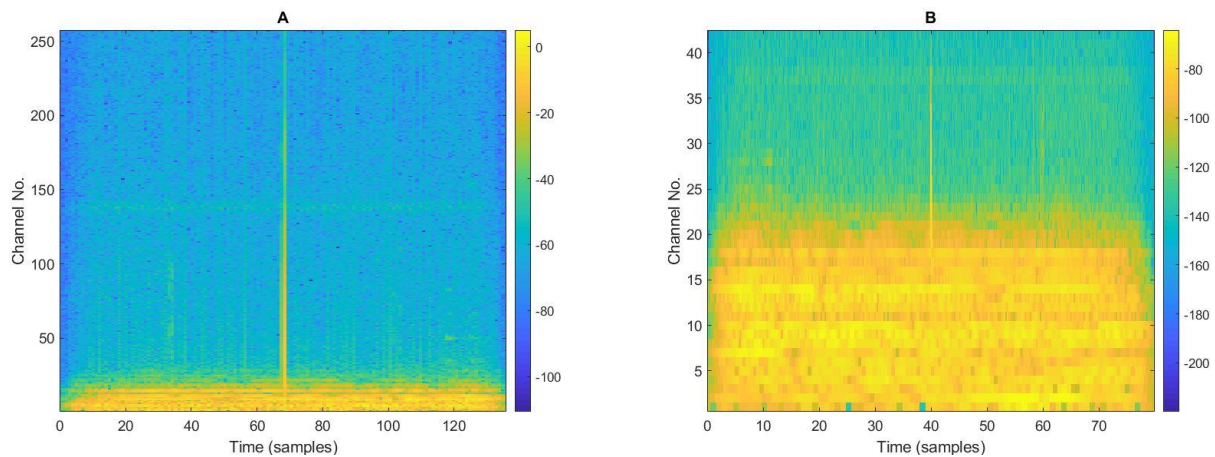
kde f je frekvence. Šířku pásma lze také určit přímo v počtu ERBů dle [18]

$$ERB_{num}(f) = 9,265 \ln \left(1 + \frac{f}{228,8455} \right). \quad (17)$$

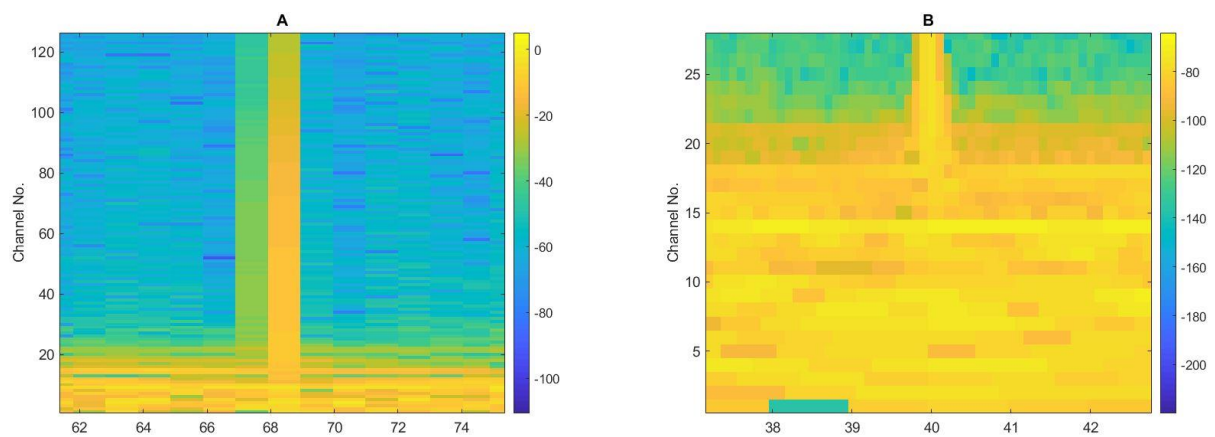
Rozdělení frekvenční osy do pásma dle banky filtrů je nepřímo úměrné rozdělení časové osy do úseků odpovídajícím centrální frekvenci těchto filtrů. Ze spektrogramů získaných pomocí krátkodobé Fourierovy transformace a ERBletové transformace na obrázku 6 je toto rozdělení evidentní. V případě použití krátkodobé Fourierovy transformace je signál rozdělen 256 frekvenčních pásma, z nichž všechny mají stejné časové rozlišení (obr. 6A). Filtry ERBletové transformace mají dle rovnice 16 užší šířku pásma v nižších frekvencích a tím pádem mají lepší frekvenční rozlišení v těchto pásmech (obr. 6B). Naopak ve vyšších pásmech mají velkou šířku pásma a z toho důvodu velmi dobré časové rozlišení. Tato vlastnost aproximuje vlastnosti kochley. Na obrázku 7 jsou zobrazeny detaily spektrogramů a lze na nich pozorovat měnící se časově frekvenční rozlišení ERBletové transformace (obr. 7B). Okna jednotlivých filtrů jsou definována rovnicí [18]

$$g_k[m] = \Gamma_k^{-\frac{1}{2}} e^{-\pi \left[\frac{m-v_k}{\Gamma_k} \right]^2}. \quad (18)$$

kde m je index frekvence, v je centrální frekvence a Γ je tvarovací faktor, který určuje efektivní šířku pásma okna g . Člen $\Gamma_k^{-1/2}$ zajišťuje, že všechna okna mají stejnou energii. Samotná transformace s těmito okny je realizována nestacionární diskretní Gaborovou transformací [19]. Výhodou použití těchto vlastností je dokonalá rekonstrukce s minimální redundancí a efektivním numerickým výpočtem pro její realizaci.



Obrázek 6: Spektrogram STFT (A) a ERBletové transformace signálu s impulzní vadou (B)

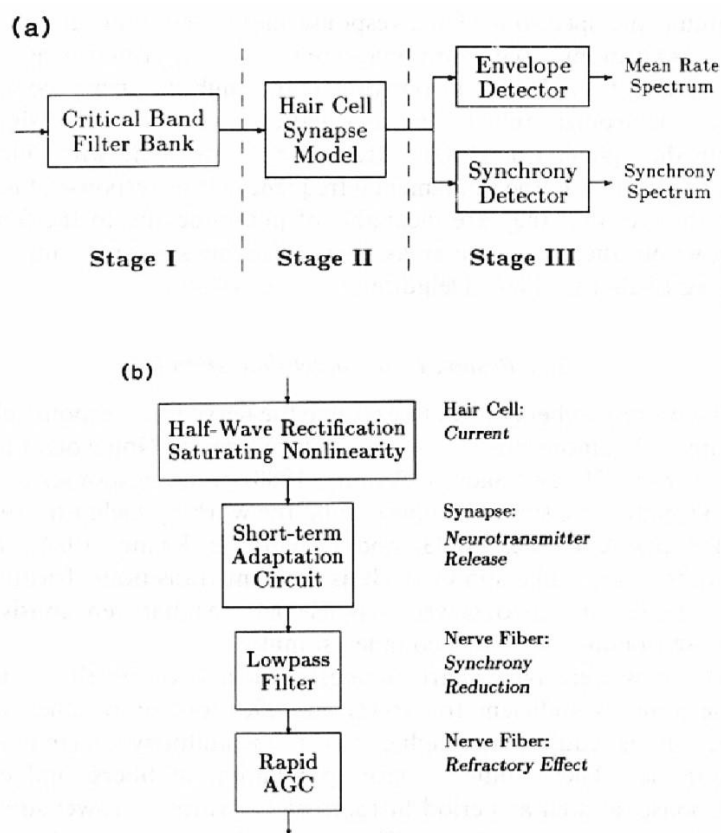


Obrázek 7: Přibližný spektrogram STFT (A) a ERBletové transformace signálu s impulzní vadou (B)

4.3. Model slyšení dle S. Seneff

Pro účely zpracování lidské řeči a rozpoznávání jejich elementárních prvků byl v [20] navržen model lidského sluchového ústrojí. Snahou bylo vylepšit systémy pro zpracování lidské řeči, které jsou založeny na nebiologických kritériích. Ze studií zabývajících se fyziologickými procesy v uchu je zřejmé, že při probíhající frekvenční analýze signálu je nutno brát v potaz vliv nelinearity, saturací a adaptačních mechanismů. Adaptační mechanismy mohou být různě rychlé a intenzivní. Způsobují utlumení synapsí nervových zakončení po příchodu nestacionárního signálu. Znatelně nejvíc synapsí probíhá přibližně prvních 15 milisekund po příchodu stimulu. Poté následuje rapidní krátkodobý pokles (adaptace) a dále déletrvající a méně intenzivní pokles až na stacionární úroveň zhruba 50 milisekund po prvotní stimulaci. Dalším důležitým aspektem je časové maskování.

Blokové schéma tohoto modelu je znázorněno na obrázku 8. Systém byl navržen jako soustava 40 nezávislých kanálů pokrývajících frekvence od 130 do 6400 Hz. Šířka pásma jednotlivých filtrů je zhruba 0,5 barku. Model využívá preemfáze k simulaci rezonancí ve vnějším zvukovodu. Počet kanálů byl omezen z důvodu výpočetní náročnosti na úkor přesnosti simulace. Každý kanál se skládá z pásmové propusti a části reprezentující přenos vibrací bazilární membrány na nervová vlákna. Druhá jmenovaná část zahrnuje kompresi dynamického rozsahu, usměrnění signálu, adaptaci na vstupní signál či dopředné maskování. Výstupem této části je pravděpodobnost vzniku vzruchu na nervových zakončeních jako funkce času. Celkový výstup modelu je rozdělen na dvě části, z nichž první je obálka signálu, korespondující s průměrnou četností vzruchů nervových zakončení. Druhá část určuje míru vlivu signálu okolo jednotlivých centrálních frekvencí filtrů na výstup modelu.



Obrázek 8: Blokové schéma modelu S. Seneff [20]

Nelineární část modelu se skládá ze čtyř komponent znázorněných na obr. 8b. První v pořadí je usměrňovač definován jako [20]:

$$y = 1 + A \tan^{-1}(Bx) \text{ pro } x > 0 \quad (19)$$

$$y = e^{ABx} \text{ pro } x \leq 0, \quad (20)$$

kde A a B jsou konstanty upravující zisk. Krátkodobá adaptace je založena na modelu toku látky ze zdroje skrze membránu. Tato membrána umožňuje přesun látky (neurotransmiteru) v závislosti na koncentraci dané látky před ní a za ní. Adaptace také zahrnuje simulaci samovolné ztráty látky. Model lze popsat rovnicemi [20]

$$\frac{dC(t)}{dt} = \mu_a[S(t) - C(t)] - \mu_b C(t) \text{ pro } C(t) < S(t) \quad (21)$$

$$\frac{dC(t)}{dt} = -\mu_b C(t) \text{ pro } C(t) \geq S(t), \quad (22)$$

kde C(t) je koncentrace látky za membránou a S(t) je koncentrace před membránou. Členy μ_a a μ_b jsou konstanty. Následuje n_{lp} dolnoproputných filtrů, které redukuje odezvu na vysokofrekvenční signály a vyhlazují signál z usměrňovače v případě saturace. Celková přenosová funkce je definována rovnicí [20]

$$H(z) = \left(\frac{1 - \alpha}{1 - \alpha z^{-1}} \right)^{n_{lp}}, \quad (23)$$

kde α je pozice pólu na reálné ose z-roviny. Posledním blokem je AGC, neboli automatická změna zisku, který simuluje časové maskování. Tento blok je v modelu popsán rovnicí [20]

$$y(t) = \frac{x(t)}{1 + \int_{t-\Delta}^t x(\alpha) d\alpha}, \quad (24)$$

kde Δ je trvání maskovacího efektu.

4.4. Kochleární model R. F. Lyona

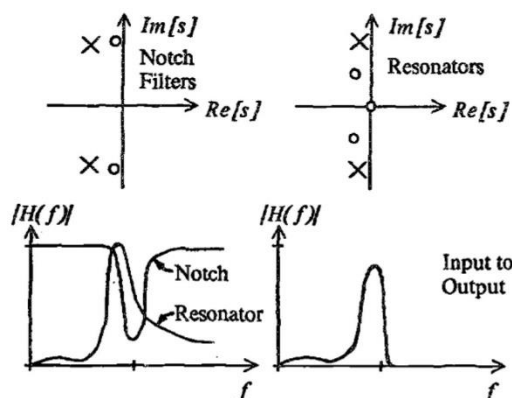
Obdobně jako u předchozího modelu i v [21] se autor snažil modelovat přenos zvukového signálu lidským sluchovým ústrojím za účelem zpracování řeči. Celý model je rozdělen do tří částí, které provádí postupné zpracování signálu a jeho převod do oblasti nervových buněk. Autor popisuje filtraci, detekci a kompresi (AGC) signálů v kochleě a navrhuje diskrétní výpočetní model pro aproximaci těchto aspektů. Prvotní filtrace signálu je navržena jako lineární a časově invariantní. Je zde využito kaskády filtrů, které dělí signál do frekvenčních pásem dle fyziologické funkce kochleý. Každý z těchto filtrů je definován komplexní přenosovou funkcí tlakové vlny [21]

$$\frac{P}{P_0} = e^{-ikdx}, \quad (25)$$

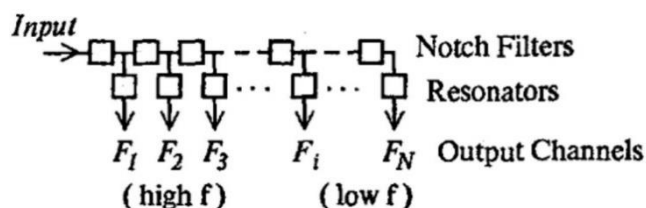
kde dx je délka segmentu bazilární membrány přiřazená danému filtru a k je vlnové číslo definováno rovnicí [21]

$$k = \frac{c}{\sqrt{\omega_R^2 + \frac{i\omega\omega_R}{Q} - \omega^2}}, \quad (26)$$

kde c je konstanta, ω_r je rezonanční frekvence, Q činitel jakosti a ω je frekvence tlakové vlny. Pro delší segment je k rovno integraci podél dx . Výsledkem je pásmová zádrž, kterou lze pro dostatečně malé segmenty aproximovat bikvadrálním filtrem, tedy filtrem s dvěma póly a dvěma nulami na blízkém kmitočtu, kde nuly jsou blíže jednotkové kružnici. Pro převod akustického tlaku na pohyb bazilární membrány je navíc vložen rezonátor, modelovaný jako filtr s nulami ve stejnosměrné složce a párem pólů mezi nulami předchozí a následující pásmové zádrže. Pro zpřesnění výpočtu je přidána rezonance Cortiho orgánu, která je modelována přidáním páru nul do rezonátoru, zhruba oktávu pod rezonanční frekvenci bazilární membrány pro daný segment. Popis filtrů v komplexní rovině a jejich frekvenční charakteristiky jsou zobrazeny na obrázku 9. Schéma kaskády filtrů je znázorněno na obrázku 10.



Obrázek 9: Popis filtrů v komplexní rovině a jejich frekvenční charakteristiky [20]



Obrázek 10: Kaskádní zapojení filtrů [21]

Autor přistupuje k výstupu z části filtrace jako k nosným vlnám, na kterých je modulován signál. Pro lepší práci se signálem byla navržena amplitudová demodulace pomocí jednocestného usměrňovače. Nevýhodou tohoto řešení je generování vyšších harmonických. Proto je signál před zpracováním interpolován na minimálně dvojnásobnou vzorkovací frekvenci. Po zpracování je možno signál filtrovat dolnopropustným filtrem a decimovat zpět na požadovanou frekvenci.

Aby bylo možné přesně znázornit a vyhodnotit výstup modelu, je nutné implementovat kompresní funkci. Hodnoty akustického tlaku vstupujícího do lidského ucha jsou od prahu slyšení do prahu bolesti v rozmezí 12 řádů. Četnost vzruchů nervových zakončení na základě vstupního akustického tlaku je v rozmezí pouze dvou řádů. Koncept automatické změny zisku systému je implementován z důvodu snahy udržet výstupní úroveň v minimálním rozsahu. Problém kontroly zisku pomocí nelineárních funkcí je v tom, že generují příliš mnoho šumu při nízkých úrovních vstupního signálu a naopak zarovnávají špičkové hodnoty. Pro eliminaci těchto efektů byl navržen model adaptivního chování vláskových buněk, který sdílí konstanty pro změnu zisku s ostatními kanály. Zisk je kontrolován třemi parametry, z nichž jeden je globální pro celý signál a dva jsou definovány pro každý kanál zvlášť. Změna zisku je definována rovnicemi [21]

$$O_i = L(D_i G_A G_{B,i} G_{C,i}) \quad (27)$$

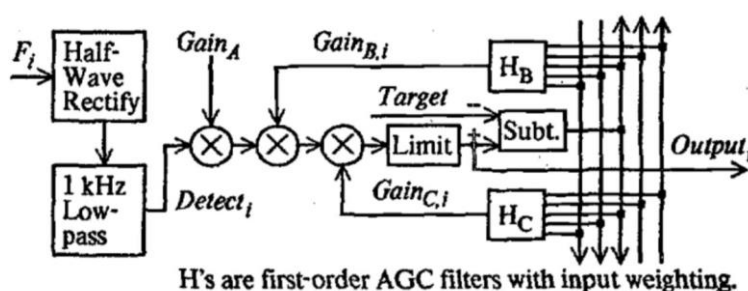
$$E = O_i - T \quad (28)$$

$$G_A = Z^{-1}[(1 - \varepsilon_A)G_A - \varepsilon_A(W_{T_A}E)] \quad (29)$$

$$G_{B,i} = Z^{-1}[(1 - \varepsilon_B)G_{B,i} - \varepsilon_B(W_{T_{B,i}}E)] \quad (30)$$

$$G_{C,i} = Z^{-1}[(1 - \varepsilon_C)G_{C,i} - \varepsilon_C(W_{T_{C,i}}E)], \quad (31)$$

kde O_i je výstup i -tého kanálu, L je kompresní nelineární funkce, D_i je decimovaný výstup z jednocestného usměrňovače, G_a je celkový zisk, $G_{b,i}$ a $G_{c,i}$ jsou vektory zisků pro jednotlivé kanály, E je zpětnovazební vektor pro kontrolu zisku, T je požadovaná úroveň výstupu, W_{T_A} je vektor vah jednotlivých kanálů pro zisk G_a . $W_{T_{B,i}}$ a $W_{T_{C,i}}$ jsou vektory vah jednotlivých kanálů pro zisky $G_{A,i}$ a $G_{B,i}$. Časové konstanty filtrů AGC jsou definovány jako T/ε_A , T/ε_B a T/ε_C , kde T je vzorkovací perioda. Obrázek 11 znázorňuje celkové blokové schéma demodulace a AGC.



Obrázek 11: Blokové schéma demodulace a AGC [21]

5. Návrh a implementace detekčních algoritmů

Detekční algoritmus stručně popsáný v kapitole 3.4 [9] nabízí velmi slibné výsledky pro detekci slyšitelných vad. Oproti ostatním popsaným metodám vykazuje minimum falešných detekcí, ovšem nevýhodou tohoto modelu je jeho výpočetní náročnost. Z těchto důvodů jsou implementace čtyř detekčních algoritmů využívající modelů ze třetí kapitoly založeny na stejné myšlence jako metoda využívající hydrodynamický model [9]. Zásadní kroky všech algoritmů jsou vždy obdobné. Jedná se o načtení signálů a jejich předzpracování, filtrace určitým modelem, další zpracování (korelace, filtrace) a vyhodnocení detekce. Klíčové parametry těchto úprav jsou testovány v prvním testu (kapitola 6.3). Funkce realizující jednotlivé modely byly převzaty a nebyly vytvořeny v rámci této práce. Detekční algoritmy využívající těchto modelů byly programovány v prostředí Matlab R2018a. Celkem bylo vytvořeno pět funkcí, které provádějí detekci vad s využitím čtyř modelů, jedna funkce pro spuštění detekce a jedna funkce pro vyhodnocení detekce chyb v porovnání s výsledky subjektivních testů. Jedná se o funkce:

- *detekce*
- *semansky_DRNL*
- *semansky_lopezpoveda*
- *semansky_erblet*
- *semansky_lyon*
- *semansky_seneff*
- *vyhodnoceni*

Detekční algoritmy mají vždy stejné vstupní a výstupní proměnné. Vstupní proměnné jsou:

- *sada* – adresa složky se zvukovými soubory určenými k analýze
- *start_sig* – index prvního souboru
- *num_sig* – počet testovaných souborů

Výstupní proměnné jsou:

- *det* – vektor s výsledky algoritmu
- *elapsed_time* – doba běhu algoritmu

Výstupní parametr *det* obsahuje výsledky algoritmů. Jeho délka je shodná s počtem testovaných vzorků. V případě, že algoritmus detekuje vadu v signálu, proměnná *det* bude mít pro daný vzorek hodnotu 1. Doba běhu algoritmu je počítána pomocí instrukcí *tic* a *toc*. Instrukce *tic* je volána jako první v detekčním algoritmu a spouští měření času. Na konci vyhodnocení každého ze vzorků je instrukce *toc*, která měří dobu od spuštění měření. Po posledním vzorku, kdy je instrukce *toc* volána naposledy, je zobrazen celkový čas běhu algoritmu a je zapsán do proměnné *elapsed_time*.

5.1. Detekční algoritmy využívající model DRNL

Detekční algoritmus využívající model DRNL (kapitola 4.1) byl implementován ve dvou různých verzích. Rozdíl mezi implementacemi je ve funkci, která realizuje zpracování signálu. První z těchto algoritmů bude dále v textu označován jako DRNL a má tvar:

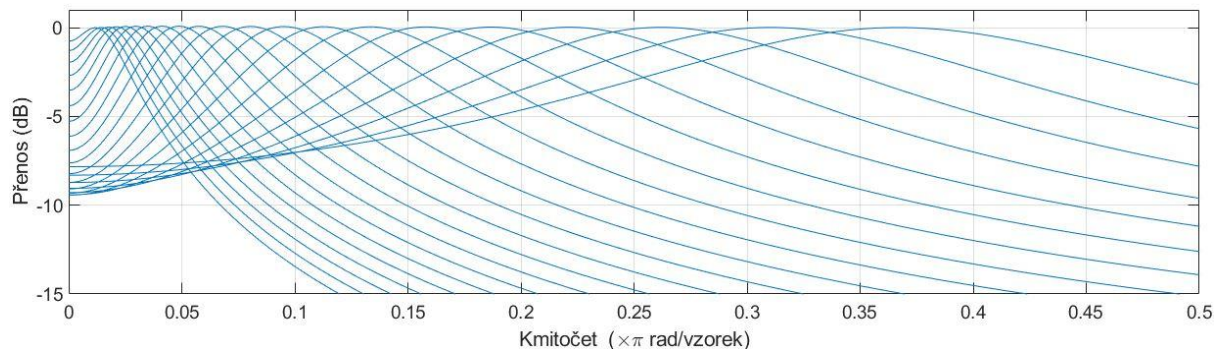
$$[det, elapsed_time] = semansky_DRNL(sada, start_sig, num_sig)$$

Funkce nejprve načte impulzní odezvu modelu (výstup funkce *DRNL_MAP1_14* [22]) z uložené proměnné *semansky_MAP_DRNL_impBM.mat*. Z impulzní odezvy je vyříznuta jen část odpovídající délce trvání průměrného clicku. Následně je načten seznam souborů ze složky *sada* a poté se ve *for* cyklu jednotlivé vzorky postupně vyhodnocují. Po načtení vzorku je odstraněn pravý kanál. Toto

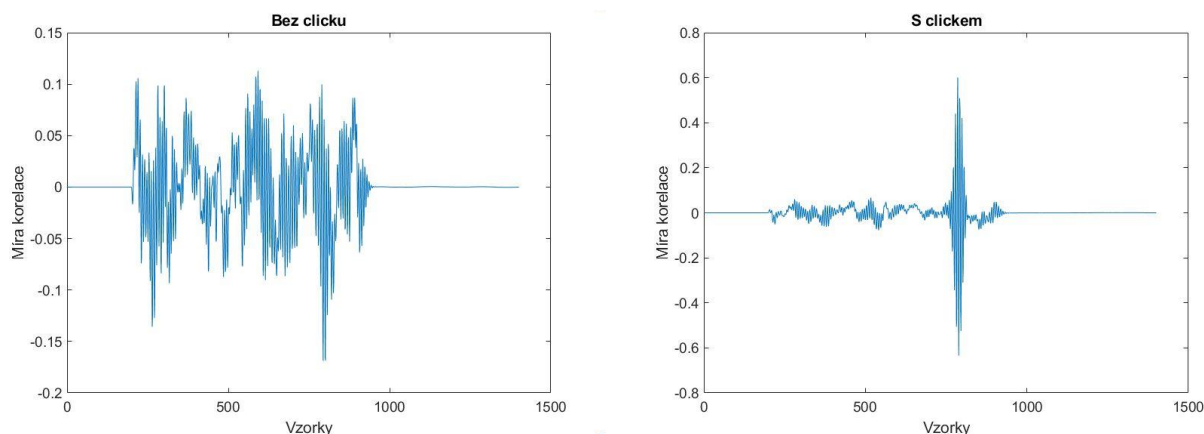
opatření je nutné z důvodu provedení subjektivních testů pouze na levém kanálu. Pro případ testování signálů s různými vzorkovacími frekvencemi a amplitudami je implementováno jejich automatické převzorkování na 44,1 kHz a normalizace.

Takto upravený signál a jeho vzorkovací frekvence jsou vstupními parametry funkce *DRNL_MAP1_14* [22], která simuluje průchod zvuku sluchovým ústrojím. Vstupní signál je segmentován a v několika krocích filtrován. První filtr simuluje rezonance vnějšího ucha, které jsou následně přičteny k původnímu signálu. Druhá filtrace převádí akustický tlak na výchylku bubínku. Třetí filtr typu horní propust simuluje setrvačnost třmínku. Posledním krokem ve středním uchu je simulace akustického reflexu, který potlačuje pohyb třmínku. Následný *for* cyklus dělí výpočet na 21 frekvenčních pásem kochley, jejichž centrální frekvence jsou generovány funkcí *logspace*, tedy v logaritmickém rozložení. V každém cyklu se mění parametry použitých filtrů. Vychýlení třmínku je poté zpracovááno ve dvou paralelních větvích. První z nich aplikuje lineární zisk a následně filtraci gammatónovým filtrem. Dva filtry znázorněné na obrázku 4 jsou zde implementovány pomocí jediné funkce, jejíž parametry kombinují vlastnosti obou filtrů. Přenosové charakteristiky těchto filtrů jsou znázorněny na obrázku 12. Druhá, nelineární větev aplikuje postupně gammatónový filtr, nelineární funkci a další filtr, který opět kombinuje parametry dvou filtrů obdobně jako v lineární větvi. Následným sečtením obou větví je získána rychlost bazilární membrány pro jedno frekvenční pásmo.

Výstupem funkce *DRNL_MAP1_14* je matice pásmových signálů, která je opět zpracovávána ve *for* cyklu. V tomto cyklu je každý signál segmentován a korelován s impulzní odezvou v odpovídajícím frekvenčním pásmu (obrázek 13). Délka jednoho segmentu je stejná jako délka impulzní odezvy. Všechny segmenty korelací ve všech pásmech jsou poté zapsány do matice *corr_out*. Pro vyhodnocení chybovosti signálu jsou sečteny korelace pro několik empiricky zvolených pásem signálu. Tento signál je porovnávána s detekčním prahem, odvozeným ze střední hodnoty odpovídajících pásem. Práh je násoben konstantou, která určuje citlivost systému.



Obrázek 12: Přenosové charakteristiky DRNL filtrů (kombinace pásmové a dolní propusti)



Obrázek 13: Korelace signálu s impulzní odezvou (kanál 20) pro segment bez clicku a s clickem

Druhý z algoritmů implementujících DRNL model má tento tvar:

$$[det, elapsed_time] = semansky_lopezpoveda(sada, start_sig, num_sig)$$

Tento detekční algoritmus, dále jen „Lopez-Poveda“, využívá jiné funkce k simulaci DRNL modelu. Jedná se o funkci *lopezpoveda2001* [23], která oproti funkci *DRNL_MAP1_14* zanedbává rezonance vnějšího ucha a střední ucho je modelováno jediným FIR filtrem, jehož parametry jsou generovány funkcí *middleearfilter* [24]. Další změnou je použití ERBové stupnice pro generování centrálních frekvencí filtrů vnitřního ucha. Namísto 21 filtrů s logaritmičtými rozmístěnými centrálními frekvencemi je použito 31 filtrů s mírnějším vzestupem centrálních frekvencí. Dva filtry, které v obou větvích DRNL modelu byly u funkce *DRNL_MAP1_14* modelovány jedním filtrem jsou zde realizovány dvěma separátními filtry. Většina parametrů se u těchto dvou funkcí liší, ovšem celkové schéma modelu, včetně vyhodnocení pomocí korelace s impulzní odezvou a porovnání s detekčním prahem, je stejná a výstup modelů je podobný.

5.2. Metoda detekce využívající ERBletové transformace

Metoda využívající spektrální analýzu založenou na ERBletové transformaci (kapitola 4.2) byla implementována ve funkci, dále značené jako ERBlet:

$$[det, elapsed_time] = semansky_erblet(sada, start_sig, num_sig).$$

Funkce nejprve načte seznam souborů ze složky *sada* a ve *for* cyklu čte postupně tyto soubory a zpracovává je. Prvními kroky jsou odstranění pravého kanálu, převzorkování signálu na 44.1 kHz a jeho normalizace.

Následně je provedena samotná ERBletová transformace pomocí funkce *erblet* [25]. Tato funkce provádí pouze dva kroky. Prvním z nich je generování oken pro analýzu signálu funkcí *erbletwin* [25]. Každé okno je definováno parametrem centrální frekvence a šířky okna. Celkem je vygenerováno 84 oken, z nichž první má centrální frekvenci 26 Hz a prostřední okno má centrální frekvenci 22050 Hz. Okna jsou symetrická a pro analýzu signálu je tedy postačujících prvních 42 oken. Druhým krokem je výpočet nestacionární diskrétní Gaborovy transformace pomocí funkce *nsdgt* [26], jejíž parametrem jsou předem vygenerovaná okna. Výstupem této funkce je matice *coef* s komplexními koeficienty transformace. Pro první kanál je vygenerováno 80 koeficientů a pro kanál s nejvyššími frekvencemi 7696 koeficientů.

Výstupní koeficienty jsou uspořádány v opačném pořadí, než je vstupní signál, což je řešeno otočením finálního detekčního signálu. Toto otočení je pro detekci výhodné, jelikož například perkusní nástroje mají velmi rychlý nástup a pomaleji odeznívají. Dále použitý mediánový filtr je tedy schopen včas reagovat na tyto zvuky a ve finále nejsou vyhodnoceny jako falešné detekce. Aby bylo možné koeficienty transformace snadno sčítat v co nejkratších úsecích, byla provedena lineární interpolace. Po této interpolaci bylo dosaženo matice s 7600 koeficienty ve všech frekvenčních pásmech. Reálná část těchto komplexních koeficientů byla sečtena napříč vybranými frekvenčními pásmy a uložena do proměnné *coef_interp_real_abs_sum*. Z této proměnné byl následně určen detekční práh. K tomu bylo využito mediánové filtrace (*medfilt1*) a empiricky určeného koeficientu. Detekční práh je posunut o dvacet vzorků směrem doleva z důvodu zpoždění oproti detekčnímu signálu, které je generováno mediánovou filtrací. Prvních šedesát vzorků detekčního signálu je vynulováno kvůli příliš nízké amplitudě obálky. Tento úsek signálu je tedy v detekci vynechán. Detekce impulzních vad byla prováděna porovnáním proměnné *coef_interp_real_abs_sum* a detekčního prahu násobeného konstantou určující citlivost systému. Jedná se tedy o jedinou detekční metodu, ve které není použita korelace s impulzní odezvou, čímž tím se liší od metody popsané v kapitole 3.4 [9].

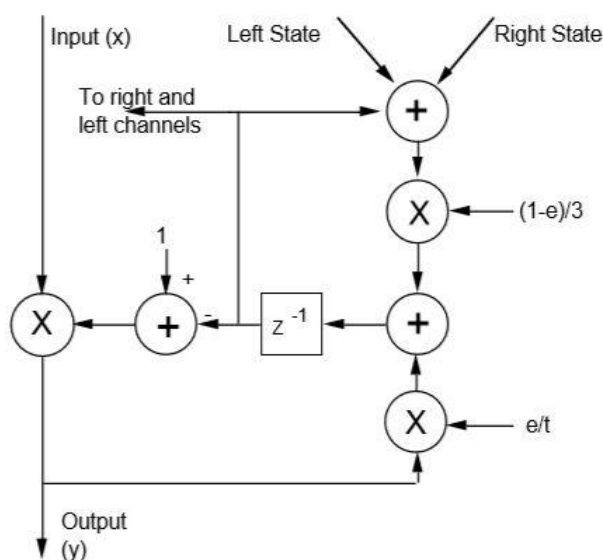
5.3. Detekční algoritmus implementující model R. F. Lyona

Model popsaný v kapitole 4.4 byl implementován ve funkci

$$[det, elapsed_time] = semansky_lyon(sada, start_sig, num_sig),$$

kteřá má identickou strukturu jako funkce *semansky_DRNL*. V dalším textu je tento model označován jako Lyon. V tomto případě modeluje sluchové ústrojí funkce *LyonPassiveEar* [27]. Nejprve jsou funkcí *DesignLyonFilters* [27] vygenerovány parametry centrální frekvence a šířky pásma kaskádních filtrů. Celkem je vygenerováno 120 filtrů (frekvenčních pásem). Centrální frekvence se pohybují od 88Hz do 21037Hz a šířka pásma filtrů je v rozsahu od 125 Hz do 2632 Hz. Z těchto parametrů jsou vypočteny nuly a póly filtrů, jejich činitelé jakosti, zisk každého kanálu a následně jsou navrženy filtry druhého řádu. Samotná filtrace není implementována v Matlabu, ale v externích souborech v jazyce C kompilovaných do MEX souboru, aby je bylo možné použít v Matlabu. První filtraci provádí funkce *soscascade* [24], která dělí signál do 120 pásem pomocí navržených pásmových propustí. Následuje usměrnění signálu a změna zisku pomocí funkce *agc* [24]. Tato funkce upravuje amplitudu výsledného signálu dle současné amplitudy daného kanálu a okolních dvou kanálů. Je zde implementován zpětnovazební obvod (obrázek 14) zahrnující filtraci dvou nejbližších kanálů a kompresi výstupu. Tyto tři signály slouží pro určení potlačení následujícího výstupu. Funkce *agc* vrací komprimovaný signál z jednotlivých segmentů kochley. Tento signál je převeden pomocí paralelních rezonátorů implementovaných ve funkci *sosfilters* [24] na rychlost kmitání bazilární membrány. Autoři algoritmu z neznámého důvodu odstranili první dva kanály. Při detekci clicků se však používají jen vyšší frekvenční pásma a z toho důvodu toto nepřináší žádná omezení. Výstupem funkce *LyonPassiveEar* je matice s informací o kmitání bazilární membrány ve 118 segmentech.

Řádky této matice jsou následně korelovány s impulzní odezvou v odpovídajícím pásmu. Korelace je následně sečtena přes vybraná frekvenční pásma a porovnána s detekčním prahem. Detekční práh je určen jako střední hodnota uvedeného součtu korelací a pro nastavení citlivosti algoritmu je násoben zvolenou konstantou.



Obrázek 14: Schéma automatického ovládání zisku [19]

5.4. Detekce vad pomocí modelu S. Seneff

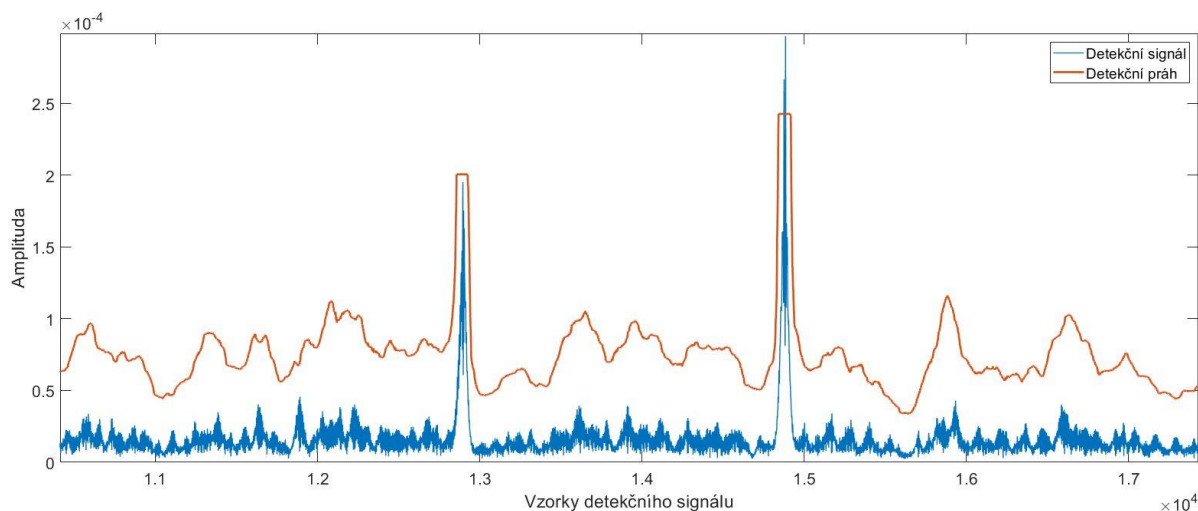
Detekční algoritmus využívající modelu Seneff (kapitola 4.3) byl implementován ve funkci

$$[det, elapsed_time] = semansky_seneff(sada, start_sig, num_sig).$$

Detekční algoritmus je pro lepší orientaci označován jako Seneff. Struktura algoritmu je kombinací předchozích. Je zde využito korelace a také porovnání s detekčním prahem získaným mediánovým filtrem, obdobně jako u algoritmu využívajícího ERBletovou transformaci. Nejprve je načtena impulzní odezva modelu, která je omezena na určitou délku. Následně jsou ve *for* cyklu čteny a zpracovávány jednotlivé zvukové vzorky. Před filtrací modelem Seneff je zvukový vzorek převzorkován a normalizován.

Funkce simulující sluchové ústrojí se nazývá *SeneffEar* [27] a modeluje všechny bloky znázorněné na obrázku 8. Prvním krokem této funkce je načtení všech parametrů z globálních proměnných popřípadě jejich vygenerování funkcí *SeneffEarSetup* [27]. Parametry jsou převzaty z [28]. Vstupní signál je nejprve upraven preemfázovým filtrem a poté, stejně jako u funkce *semansky_lyon*, dělen do frekvenčních pásem pomocí funkce *soscascade*. Každé pásmo je následně filtrováno rezonančním filtrem se dvěma páry nula-pól. Signál je dále usměrněn podle rovnic 19 a 20. Krátkodobá adaptace je implementována v následujícím *for* cyklu. Předposledním krokem je simulace ztráty synchronní odezvy pro vyšší frekvence dosažená dolnopropustným filtrem. Posledním blokem je automatické řízení zisku (AGC), které je definováno rovnicí $y = y / (1 + kagc * averageY)$, kde *kagc* je konstanta a *averageY* je výstup průměrovacího filtru. Výstupem algoritmu je pravděpodobnost synapsí sluchových nervů.

Jednotlivé kanály jsou poté korelovány s částí impulzní odezvy. Na rozdíl od přechozích metod není korelace prováděna po segmentech, ale je korelován celý signál s impulzní odezvou. Toto opatření je z důvodu implementovaného řízení zisku, které znemožní detekci chyby pomocí jednoduchého prahu. Výstup korelace je omezen pouze na část s konstantním ziskem, aby bylo možné snadno detekovat chyby. Tímto je sice odebrána krátká část signálu, nicméně toto se dá snadno řešit překrýváním zvukových vzorků. Blok usměrňovače způsobil posun stejnosměrné složky do kladných hodnot, což je řešeno odečtením střední hodnoty signálu. Následně je provedena sumace korelací ve vybraných kanálech a vynulování prvních dvaceti vzorků. Toto opatření je nutné kvůli použití mediánového filtru, který má na začátku signálu nižší hodnotu a tím způsobuje výskyt falešných detekcí. Krok detekce spočívá v porovnání sumy korelací přes vybrané kanály s obálkou signálu získanou mediánovým filtrem a vynásobenou empiricky určenou konstantou.



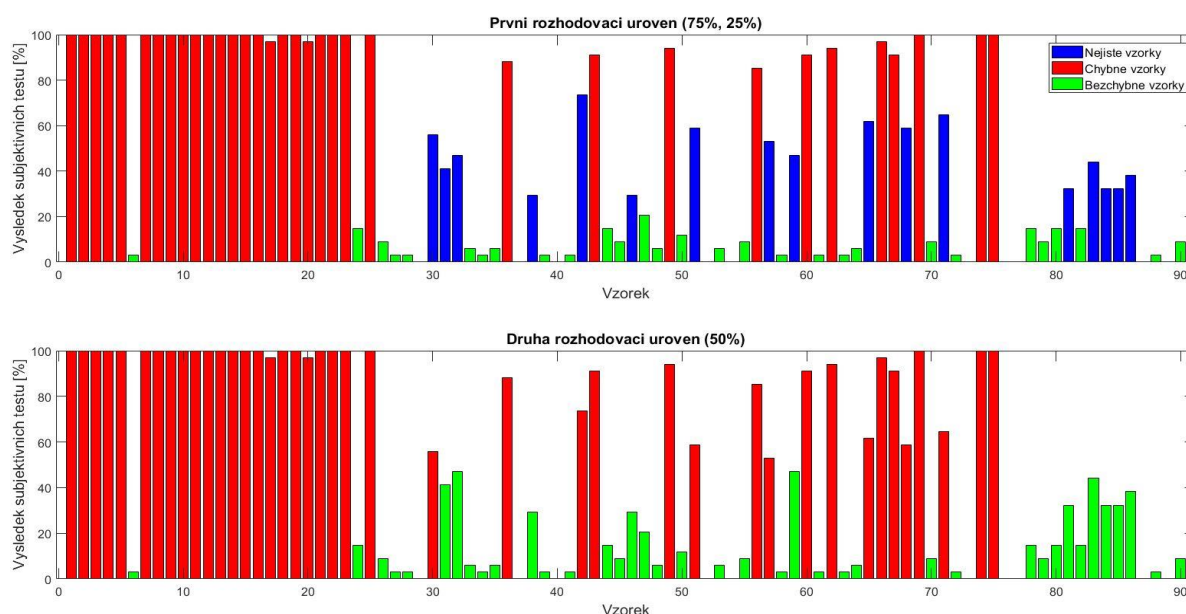
Obrázek 15: Porovnání detekčního signálu s prahem

6. Testování vzorků

6.1. Testovací množiny vzorků

Jak již bylo zmíněno v první kapitole, v současné době probíhá zejména subjektivní kontrola audiozáznamu na konci výrobního procesu gramofonových desek. Z tohoto důvodu je v této práci cíleno na detekci slyšitelných defektů. Pro testování detekčních algoritmů bylo využito sady vzorků dodaných firmou GZ Media, a.s. Tato sada obsahuje devadesát úryvků extrahovaných z vinylových desek a digitalizovaných do formátu WAV s vzorkovací frekvencí 44.1 kHz a 16 bitovou kvantizací. Každý ze vzorků je 800 milisekund dlouhý. Obsahem nahrávek je například vážná hudba, bubnování, šum apod. O nahrávkách nejsou dostupné žádné informace týkající se polohy a vlastnosti degradací. Kvůli možnosti provádět testy detekčních algoritmů na těchto signálech byla v [8] provedena jejich analýza s využitím subjektivních poslechových testů. Celkem sedmnáct subjektů ve věku od 21 do 47 let (s mediánem 22 let) se zúčastnilo těchto testů, ve kterých hodnotili přítomnost akustických degradací. Každý posluchač prováděl test dvakrát. Celkem bylo tedy vyhodnoceno 34 testů.

Během testů každý subjekt poslouchal postupně jednotlivé vzorky s možností jejich zopakování. U každého vzorku bylo vyhodnoceno, zda je v něm slyšitelná vada. Výstupem těchto testů bylo procentuální vyjádření chybovosti vzorků, tedy kolik procent posluchačů ve vzorku vnímalo impulzní akustické rušení. Pro účely automatizovaného testování byly zvoleny dvě rozhodovací úrovně. První rozhodovací úroveň nebrala v potaz vzorky, u kterých subjektivní testy vyšly mezi 25% až 75%. Tyto vzorky jsou brány jako nejisté. Naopak vzorky nad 75% jsou brány jako vadné a pod 25% jsou považovány za bezchybné. Druhá úroveň pracuje se všemi vzorky a rozhodovací práh mezi chybovostí a bezvadností je 50%. Obrázek 16 graficky znázorňuje výsledky subjektivních testů pro obě rozhodovací úrovně chybovosti.



Obrázek 16: Výsledky subjektivních testů pro dvě rozhodovací úrovně

Ve čtvrtém testu jsou využity dva zvukové vzorky převzaté z [29]. Tyto vzorky formátu WAV jsou 22 sekund dlouhé, vzorkované frekvencí 48 kHz a kvantované 16 bity. Jedná se o hudební vzorky, do kterých byl aditivně přidán impulzní šum extrahovaný z tichých částí gramofonových desek. Celkem je v signálu 2175 pulzů šumu se šířkou mezi 4 až 31 vzorky. Celkový počet poškozených vzorků je 23898, tedy 2,49% všech vzorků v testovaných hudebních úryvcích. K dispozici je hudební signál bez vad, samotný šumový signál i hudební signál kontaminovaný tímto šumem. Není k dispozici informace o akustickém projevu jednotlivých vad.

6.2. Popis testů

Prvním z testů je vyhodnocení statistické úspěšnosti detekčních algoritmů. Detekční algoritmy jsou aplikovány na vzorky dodané firmou GZ Media, a.s., na kterých byly v [8] provedeny subjektivní poslechové testy. Výsledky detekčních algoritmů jsou porovnány s výsledky subjektivních testů popsanych v předchozí kapitole. Testy se vyhodnocují ve dvou rozhodovacích úrovních. Kvalita detekčních algoritmů je určena dle tří kritérií. Tato kritéria jsou procento detekovaných vad, procento falešných detekcí a doba běhu algoritmu. První jmenované kritérium je vypočteno dle rovnice

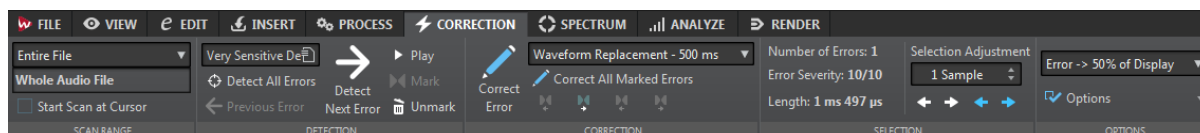
$$\% \text{ detekovaných vad} = \frac{\text{spravne_detekce}}{\text{subjektivne_chybne}} * 100, \quad (32)$$

kde *spravne_detekce* je počet chybných vzorků dle detekčního algoritmu z množiny vzorků subjektivními testy označených za chybné a *subjektivne_chybne* je celkový počet chybných vzorků dle subjektivních testů. Toto kritérium určuje procento úspěšnosti detekce u vzorků, ve kterých je slyšitelné akustické rušení. Druhým kritériem je procento falešných detekcí definované rovnicí

$$\% \text{ falešných detekcí} = \frac{\text{falesne_detekce}}{\text{subjektivne_bezchybne}} * 100, \quad (33)$$

kde *falesne_detekce* je počet chybných vzorků dle detekčního algoritmu z množiny vzorků subjektivními testy označených za bezchybné a *subjektivne_bezchybne* je celkový počet bezchybných vzorků dle subjektivních testů. Toto kritérium určuje chybovost systému jako procento vzorků bez slyšitelného impulzního rušení, které byly algoritmem označeny za vadné. Třetím kritériem je doba běhu algoritmu, která je vyhodnocena jako doba od zavolání funkce detekčního algoritmu v prostředí Matlab až do poslední vykonané instrukce v této funkci. Použité instrukce pro měření času dokáží určit čas s přesností na více desetinných míst, nicméně čas je ovlivněn procesy na pozadí PC, a proto jsou směrodatným údajem pouze celé sekundy. V následujících kapitolách jsou popsány parametry, které mají vliv na výsledky detekčních algoritmů. Parametry modelů využitých v detekčních algoritmech jsou ponechány dle návrhu autorů a jejich vliv na výsledky není v této práci testován. Testovanými parametry jsou například detekční práh, počet kanálů pro vyhodnocení či délka okna mediánového filtru.

Druhý test spočívá v porovnání detekčních algoritmů navržených a implementovaných v této práci s komerčním softwarem, který obsahuje nástroj pro detekci impulzních vad. Pro tyto testy byl použit software Wavelab Pro 9.5.15 (build 45). Tato verze je oproti předchozím rozšířena o záložku, ve které jsou nástroje pro detekci a korekci vad (obr.17). Software nenabízí výběr více detekčních metod, ovšem nabízí tři parametry, které lze uživatelsky měnit. Těmito parametry jsou „Threshold“, „Detection“ a „Detection Width“. Bohužel není dodán bližší popis těchto parametrů ani jejich vliv na detekci. Jsou dostupné tři přednastavené konfigurace těchto parametrů s názvy „Click detection“, „Sensitive Detection“ a „Very Sensitive Detection“.



Obrázek 17: Grafické rozhraní pro detekci a korekci vad

Detekci vad v programu Wavelab Pro není možné automatizovat. Z toho důvodu byly testovány pouze tři přednastavené konfigurace pro detekci. K testování byly použity stejné hudební vzorky jako u testu prvního a výsledky detekce byly porovnány se subjektivními testy. Nelze také

vyhodnocovat výpočetní náročnost, jelikož veškeré testování bylo prováděno manuálně pro každý hudební vzorek. Vyhodnoceny jsou tedy pouze kritéria procentuální efektivity detekce.

Účelem třetího testu je ověřit robustnost jednotlivých algoritmů. Zvukové vzorky dodané firmou GZ Media, a.s. obsahují signály s různým dynamickým rozsahem a různou intenzitou vad. Algoritmy byly navrženy tak, aby se jejich parametry měnily v závislosti na vstupním signálu. Takovým parametrem je například detekční práh, který silně závisí na charakteru vstupního signálu. Pro tento test bylo navrženo šest kompresních funkcí, které mění dynamiku signálu i vad v něm přítomných. Výstupem testu je informace o vlivu změny dynamiky na výsledky detekce. Pro všech šest funkcí byly otestovány detekční algoritmy a zmapovány tři výstupní kritéria.

Ve čtvrtém testu je studována odezva detekčních algoritmů na dva hudební vzorky převzaté z [29]. Tyto vzorky lze analyzovat objektivně bez použití subjektivních poslechových testů. Jelikož známe přesnou pozici, délku i počet všech impulzních degradací v těchto vzorcích, může se detailně zkoumat výstup detekčních algoritmů a porovnat jej se samotným šumovým signálem. V tomto testu již nelze použít tři předchozí kritéria. Vzhledem k velkému počtu a charakteru vad v signálu lze očekávat velký počet nedetekovaných vad. Mnoho vad v signálu je ovšem bez postřehnutelného akustického projevu. Důležitým výstupem tohoto testu bude množství detekovaných vad jednotlivými algoritmy, doba běhu algoritmu a také schopnost časové lokalizace vad. Tedy synchronizace detekčního signálu s chybovým signálem.

V poslední části této kapitoly jsou výsledky detekčních algoritmů srovnány s metodami popsány v kapitole 3. Tyto metody byly v [8] podrobeny testu vlivu jednotlivých parametrů. V tomto testu byly vyhodnoceny stejné procentuální kritéria jako v této práci a byly použity i stejné hudební vzorky pro testování. Nebyla ovšem vyhodnocena výpočetní náročnost těchto algoritmů.

6.3. Test vlivu parametrů jednotlivých algoritmů

První test spočívá ve vyhodnocení kvality detekčního algoritmu. Aby bylo možné provést objektivní hodnocení, je potřeba vyhledat ideální nastavení parametrů a jejich vliv na výsledky detekce. Tyto parametry jsou v následujících kapitolách testovány. Vždy pouze jeden parametr je testován v určitém rozsahu hodnot. Hodnoty ostatních parametrů byly určeny empiricky a jsou statické. V následujících kapitolách je popsán vliv volby parametrů na výsledky detekce v porovnání se subjektivními testy při použití první rozhodovací úrovně. Výsledky pro druhou rozhodovací úroveň jsou shrnuty v závěru bez detailní analýzy.

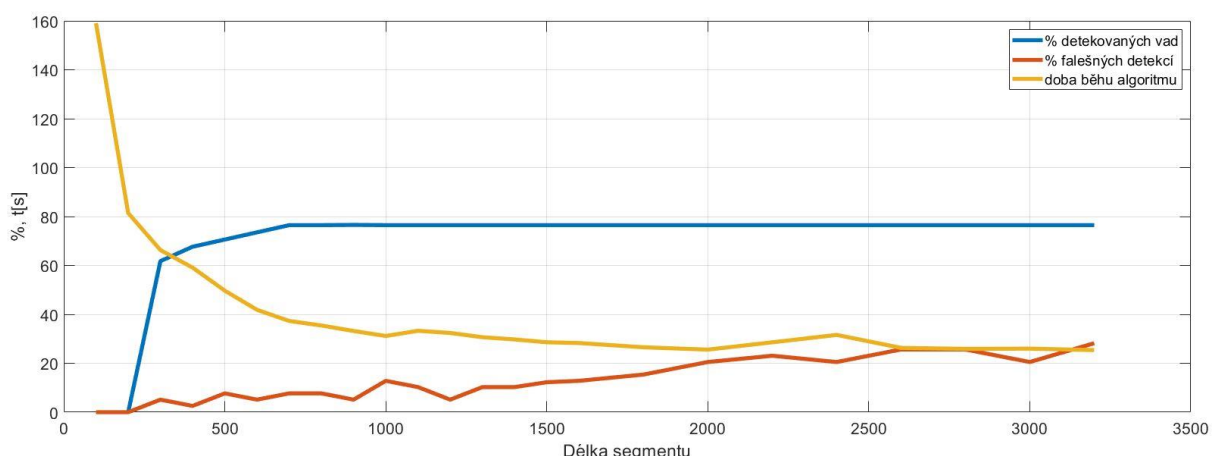
6.3.1. DRNL

V případě detekčního algoritmu využívajícího DRNL model jsou pro detekci důležité tyto tři parametry:

- Délka použitého segmentu impulzní odezvy: 900
- Použité kanály: 15 až 21
- Hodnota detekčního prahu – konstanta pro násobení: 22

Uvedená čísla jsou statické hodnoty jednotlivých parametrů pro testování. Tyto parametry jsou stejné pro všechny detekční algoritmy využívající korelaci s impulzní odezvou.

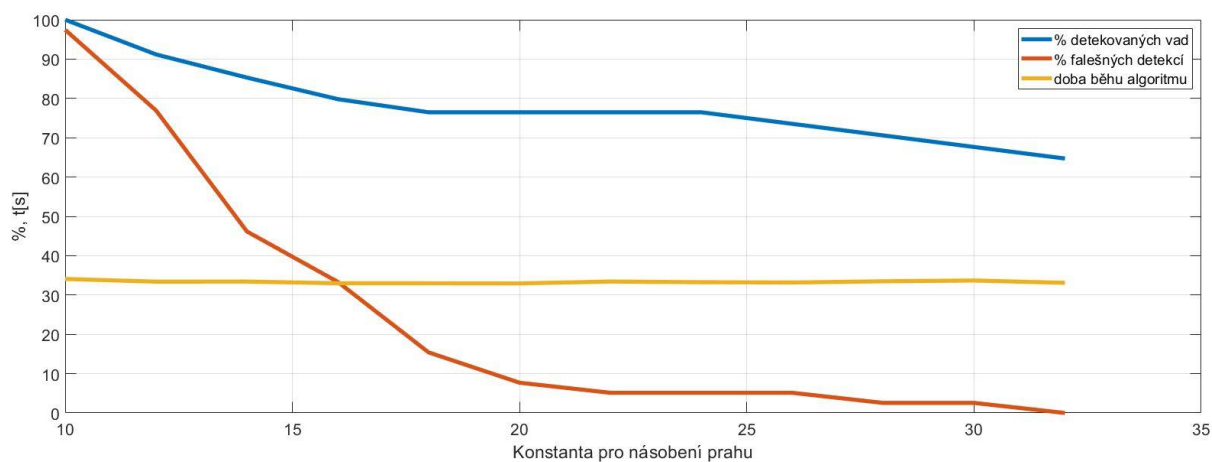
Prvním zkoumaným parametrem je délka segmentu impulzní odezvy použitého pro korelaci. Impulzní odezva modelu je dle předpokladů dlouhá okolo 17 milisekund, což zhruba odpovídá teoretickému předpokladu o impulzních vadách na gramofonových deskách. Po této době klesne její amplituda na 1% maximální hodnoty. Z obrázku 18 je vidět, že s rostoucí délkou použitého segmentu impulzní odezvy klesá čas potřebný pro vykonání algoritmu. To je způsobeno snížením počtu iterací ve *for* cyklu, který segmentuje a koreluje signál s impulzní odezvou. S rostoucí délkou segmentu také roste procento detekovaných chyb. Při délce 700 vzorků se tato hodnota ustálí na 76,47%. Při dalším zvyšování narůstá počet falešných detekcí. To je pravděpodobně způsobeno narůstajícím vlivem nízkofrekvenčních složek hudby. V nižších frekvenčních pásmech je delší impulzní odezva a také se zde vyskytují zvuky perkusních nástrojů. V těchto případech je tedy vzájemná korelace vyšší než práh a vyskytne se falešná detekce. Ideální volbou parametru délky segmentu impulzní odezvy je 900 vzorků.



Obrázek 18: DRNL - vliv délky segmentu impulzní odezvy

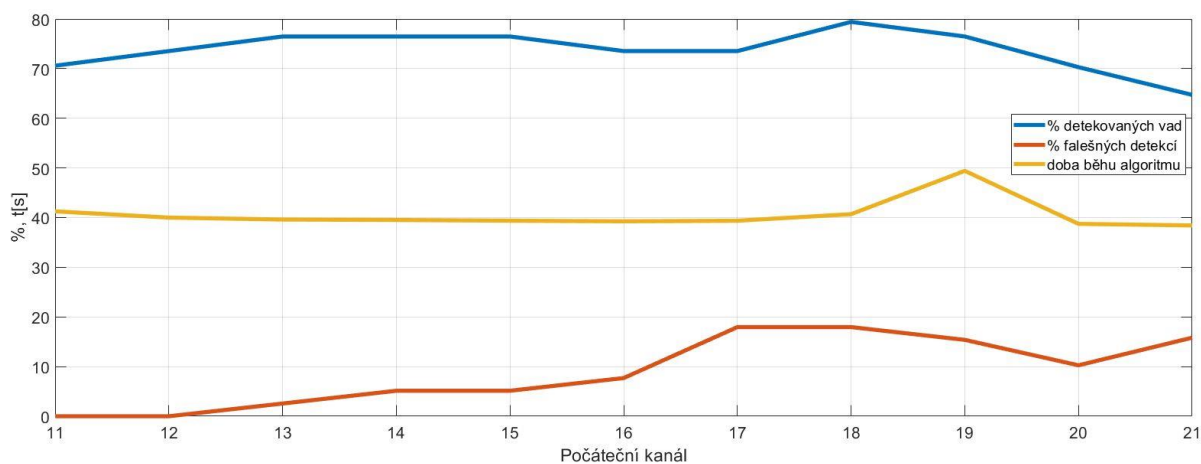
Druhým parametrem je konstanta pro násobení detekčního prahu získaného ze střední hodnoty korelace v použitých kanálech. Tento parametr přímo ovlivňuje citlivost detekčního algoritmu. Volba tohoto parametru musí být kompromisem mezi vysokým procentem správně detekovaných vad a falešně detekovaných vad. Z obrázku 19 je patrné, že při vysoké hodnotě

detekčního prahu je nalezeno příliš málo chyb a při nízkém prahu je hlášeno příliš mnoho falešných detekcí. Hodnota konstanty pro násobení prahu, nejvíce se blíží výsledkům subjektivních testů, je mezi 22 a 24.



Obrázek 19: DRNL - vliv hodnoty konstanty pro násobení detekčního prahu

Volba kanálů pro detekci je třetím zkoumaným parametrem. Jelikož mají impulzní vady ploché spektrum, relativně snadno se dají detekovat ve vyšších frekvenčních pásmech, kde je hudební složka signálu minimální. Z tohoto důvodu je zkoumána pouze spodní hranice použitých kanálů a horní zůstává na 21. kanálu obsahujícím nejvyšší frekvence. Detekční práh je závislý na počtu použitých kanálů. Pokud je použito méně kanálů, je nižší i práh (při zachování konstanty, která ho násobí) a tím pádem se můžou vyskytovat falešné detekce. Tento efekt se projevuje při použití čtyř a méně kanálů. Nejlepších výsledků bylo dosaženo při použití kanálů 13 až 21. Při použití kanálů 19 až 21 trvalo vykonání algoritmu neobvykle dlouho. Tento jev je připisován nestabilní výkonnosti použitého hardwaru. Při použití pouze dvou nebo jednoho pásma je z obrázku 20 evidentní pokles detekovaných chyb způsobený minimální korelací impulzní odezvy a reálného clicku.



Obrázek 20: DRNL - vliv volby kanálů

Z výše popsaných vlastností jednotlivých parametrů lze zvolit jejich ideální hodnoty. Nejlepších výsledků bylo pomocí tohoto algoritmu dosaženo při následujících hodnotách parametrů:

- Délka použitého segmentu impulzní odezvy: 900
- Použité kanály: 13 až 21
- Hodnota detekčního prahu – konstanta pro násobení: 22-24

V tomto případě byla kritéria vyhodnocující efektivitu algoritmu následující:

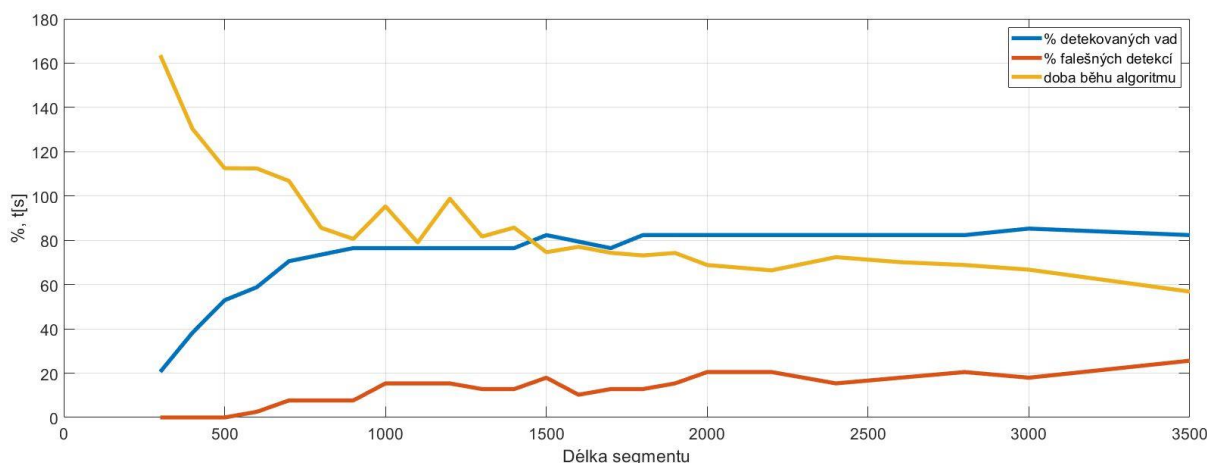
- Detekovaných chyb: 76,47 %
- Falešných detekcí: 2,56 %
- Doba běhu algoritmu: 43 sekund

6.3.2. Lopez-Poveda

Druhý detekční algoritmus založený na modelu DRNL, využívající jiné funkce oproti předchozímu testovanému algoritmu, je již z teoretického pohledu výpočetně náročnější. Jelikož se jedná o stejný model, je zde také využita korelace pro detekci clicků. Oproti předchozímu modelu je zde použito 31 kanálů, což znamená, že výpočetně nejnáročnější operace, korelace, musí proběhnout o deset cyklů vícekrát. Samotná funkce *lopezpoveda2001* simulující model slyšení využívá velké množství parametrů a funkcí, které způsobují, že oproti funkci *DRNL_MAP1_14* je více než 2x pomalejší. Celková charakteristika obou modelů by měla být stejná. Jelikož se v tomto případě jedná o model komplexnější, lze očekávat méně artefaktů a neočekávaných detekcí při minimální změně hodnoty některého z parametrů. Testované parametry a volba jejich statických hodnot je následující:

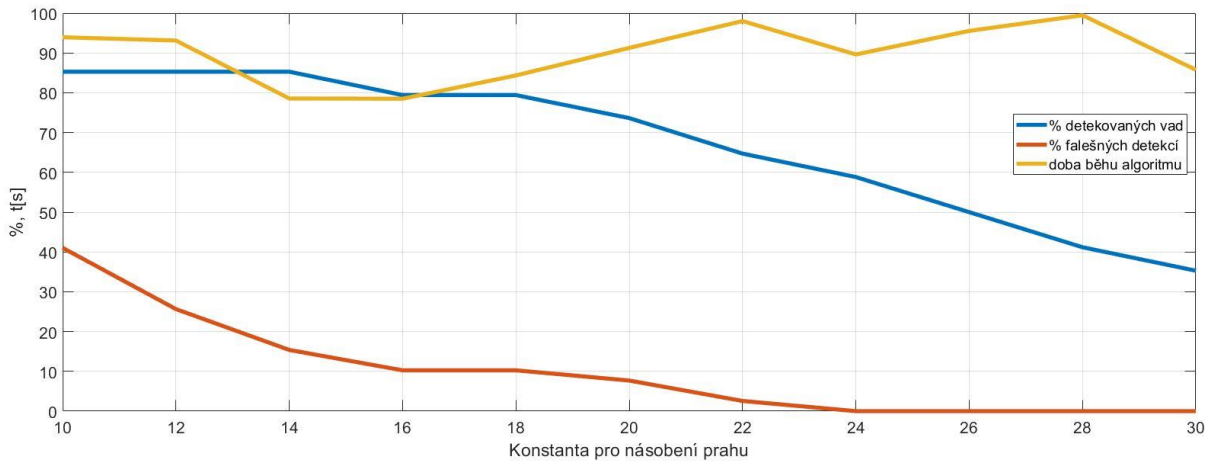
- Délka použitého segmentu impulzní odezvy: 900
- Použité kanály: 25 až 31
- Hodnota detekčního prahu – konstanta pro násobení: 19

Z obrázku 21, který znázorňuje vliv změny délky segmentu impulzní odezvy, lze odečíst velice podobný průběh hodnot jako u předchozího algoritmu. Tímto se opět potvrzuje, že impulzní odezva nesmí být omezena příliš, aby nedošlo k tomu, že bude kratší než je samotný click. I v tomto případě začne procento detekovaných vad narůstat až zhruba při použití 700 vzorků, tedy zhruba 16 milisekund signálu. Ideální hodnota je identická s předchozím modelem, tedy 900 vzorků (20 milisekund). Z trendu popisujícího dobu běhu algoritmu je zřejmý pokles při snížení délky segmentu. Tento efekt je způsoben poklesem počtu iterací *for* cyklu, ve kterém je počítána korelace signálů.



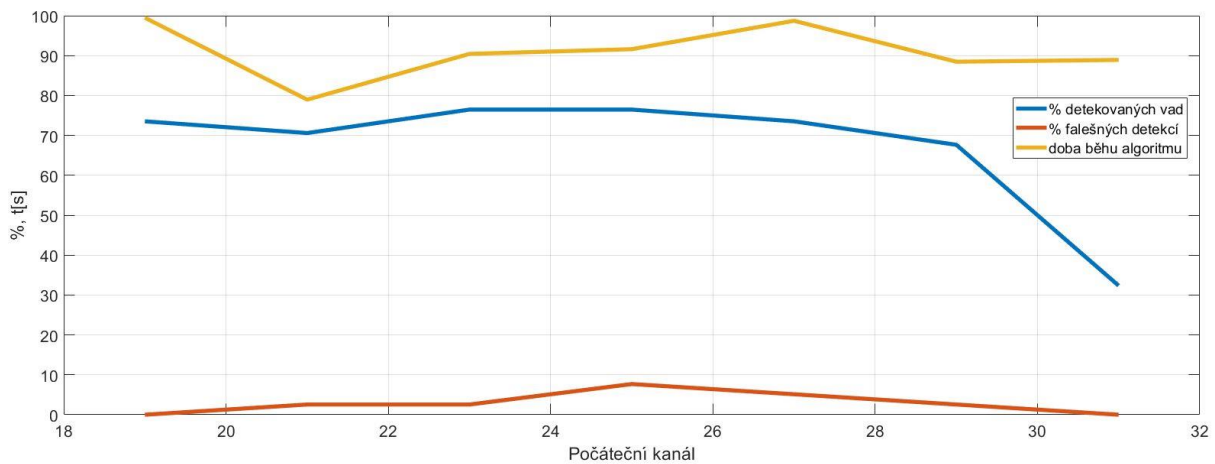
Obrázek 21: Lopez-Poveda - vliv délky segmentu impulzní odezvy

Druhým testovaným parametrem je hodnota konstanty pro násobení prahu, získaného ze střední hodnoty korelací ve vybraných kanálech. Výsledný trend značí vliv parametru na citlivost systému a je stejný jako u předchozího modelu. Lze předpokládat, že tento trend bude stejný u všech detekčních algoritmů testovaných v této práci. Volba tohoto parametru je kompromisem a v tomto případě se jako ideální jeví hodnoty 16 až 18.



Obrázek 22: Lopez-Poveda - vliv hodnoty konstanty pro násobení detekčního prahu

Kanály, ve kterých je prováděna detekce clicků, je potřeba zvolit s ohledem na hudební složku. Ideální volbou jsou jen kanály, kde je minimální obsah hudební složky, jelikož je impulzní rušení výraznější. V případě tohoto algoritmu nemá volba výrazný vliv na výsledky detekce. V obou extrémech je znatelné minimum detekovaných chyb i falešných detekcí. Při použití kanálů od 24. až od 27. do 31 je mírný nárůst falešných detekcí, nicméně stále se jedná pouze o jeden až dva špatně detekované vzorky. Nejlepších výsledků bylo dosaženo při použití kanálů 23 až 31. Ve všech případech velmi kolísá doba běhu algoritmu. V případě grafu znázorňujícího vliv hodnoty konstanty pro násobení prahu detekce je znatelné, že tato doba je náhodná a nemá spojitost se změnou tohoto parametru. V detekčním algoritmu je tento parametr použit pouze v jedné instrukci, která porovnává vektor čísel (sumu korelací přes použité kanály) s konstantou. Je tedy nepravděpodobné, že by ovlivňoval výpočetní náročnost algoritmu a tento efekt tedy můžeme přisoudit výkonnosti nestabilitě PC.



Obrázek 23: Lopez-Poveda - vliv volby kanálů

Z předchozích grafů lze odvodit, že ideální volbou parametrů jsou tyto hodnoty:

- Délka použitého segmentu impulzní odezvy: 900
- Použité kanály: 23 až 31
- Hodnota detekčního prahu – konstanta pro násobení: 19

V tomto případě byla kritéria vyhodnocující efektivitu algoritmu následující:

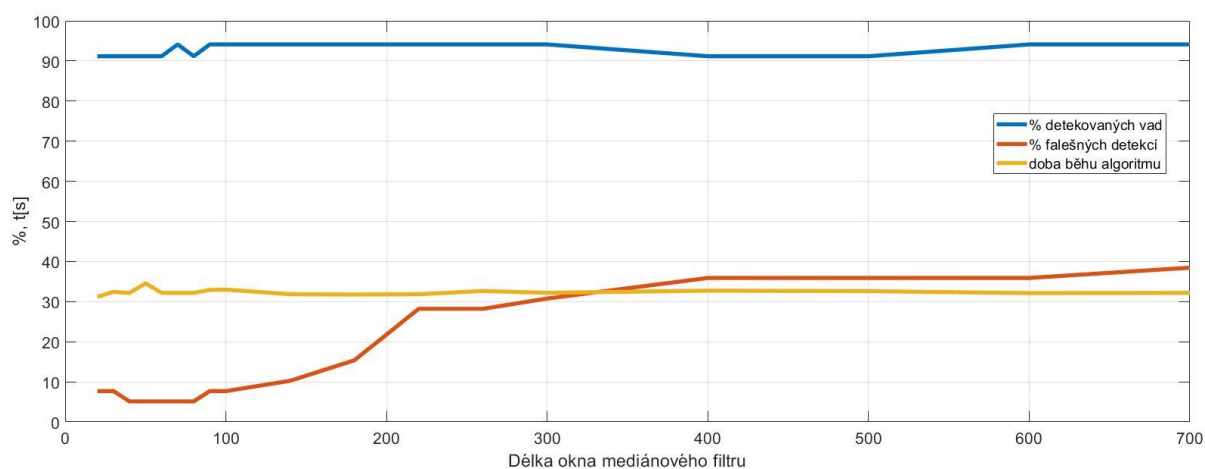
- Detekovaných chyb: 76,47 %
- Falešných detekcí: 2,56 %
- Doba běhu algoritmu: 94,99 sekund

6.3.3. ERBlet

Detekční algoritmus využívající ERBletové transformace již nepoužívá korelaci jako metody předchozí. Z tohoto důvodu by měl být algoritmus relativně rychlý, jelikož instrukce vykonávající ERBletovou transformaci není výpočetně příliš náročná. Tento algoritmus jednoduše sečte koeficienty transformace v určitých pásmech a porovná je s detekčním prahem. Z tohoto důvodu jsou i parametry algoritmu jiné. Výsledky detekce mohou ovlivnit tři základní parametry, kterými jsou délka okna mediánového filtru, konstanta pro násobení prahu a pásma použitá k detekci. Jejich statické hodnoty jsou tyto:

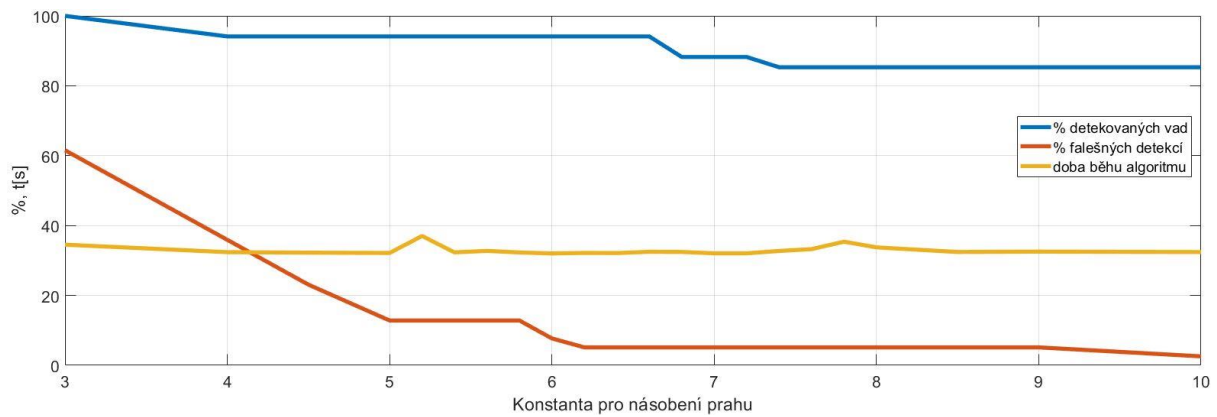
- Délka okna mediánového filtru: 70
- Detekční práh – konstanta pro násobení prahu: 6,4
- Pásma použitá k detekci: 30 až 42

Mediánový filtr je schopný vyhladit signál a získat jeho obálku, která nebude zahrnovat známky impulzního rušení. Toho je dosaženo, pokud je délka okna filtru větší, než je délka samotného rušení. V případě, že je okno výrazně delší, dokáže zamaskovat i například perkusní nástroje a jiné zdroje impulzních zvuků, které nepovažujeme za rušivé. Pokud by bylo dosaženo takto vyhlazeného detekčního prahu a následně by s ním byly porovnány koeficienty transformace, docházelo by k falešným detekcím i např. zvuků perkusních nástrojů. Je tedy potřeba volit okno tak dlouhé, aby byly potlačeny jen ty nejkratší impulzy v signálu. Z obrázku 24 je patrné, že hranice, při které se tento efekt začíná projevovat, je v okolí 150 vzorků. Procento korektně detekovaných vzorků kolísá jen nepatrně a není závislé na volbě tohoto parametru. Z trendu popisujícího dobu běhu algoritmu je patrné, že tento parametr nemá vliv ani na výpočetní náročnost.



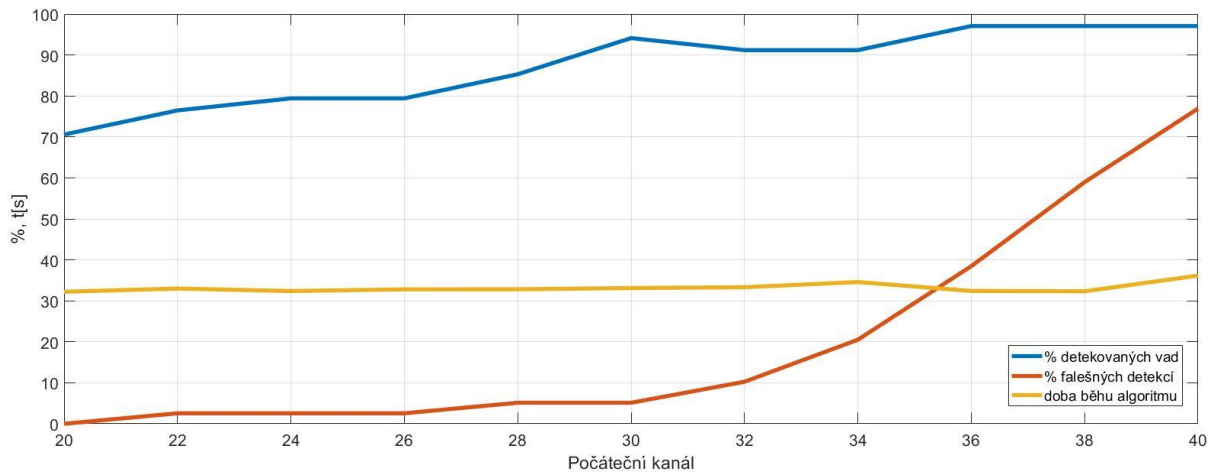
Obrázek 24: ERBlet - vliv délky okna mediánového filtru

Z výsledků předchozích detekčních algoritmů je možno vidět, že hodnota detekčního prahu ovlivňuje výsledky velmi výrazně. Detekční práh je parametr, který mimo ovlivnění výsledků detekce také dokáže popsat robustnost detekce. Možnost změny hodnoty parametru ve větším rozsahu bez vlivu na výsledky detekce značí stabilitu systému. Oproti výše popsaným metodám detekce se v případě ERBletové transformace procento detekovaných vad i procento falešných detekcí mění od určité úrovně jen minimálně. Tento fakt naznačuje, že ERBletová transformace, přesněji součet jejich koeficientů ve vyšších pásmech zvýrazňuje impulzní rušení oproti zbytku signálu a také, že mediánová filtrace se správně zvolenou délkou okna dokáže potlačit pouze cliky a ponechat impulzní zvuky přítomné v signálu.



Obrázek 25: ERBlet - vliv hodnoty konstanty pro násobení prahu

U ERBletové transformace, stejně jako u předchozích modelů, bylo použito pouze několik frekvenčních pásem pro detekci. Z obrázku 26 je patrné, že ideální volbou je použití pásem 30 až 42. V tomto případě nejsou detekovány pouze dva vzorky a pouze dva vzorky jsou detekovány falešně.



Obrázek 26: ERBlet - vliv volby kanálů

Ideální volbou parametrů v případě ERBletové transformace jsou tyto:

- Délka okna mediánového filtru: 70
- Detekční práh – konstanta pro násobení prahu: 6,5
- Pásmo použitá k detekci: 30 až 42

Při takovéto konfiguraci je tento detekční algoritmus z doposud popsaných metod nejefektivnější, nejrychlejší a nejrobustnější. Tři posuzovaná kritéria dosahují těchto hodnot:

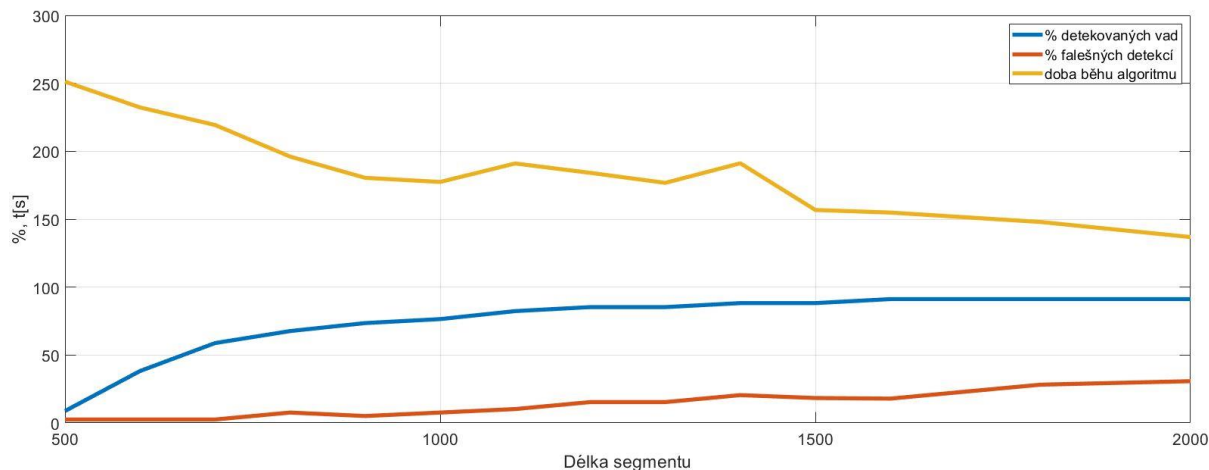
- Detekovaných chyb: 94,12 %
- Falešných detekcí: 5,13 %
- Doba běhu algoritmu: 34,68 sekund

6.3.4. Lyon

Čtvrtý detekční algoritmus opět využívá korelace s impulzní odezvou. Testovanými parametry jsou tedy délka impulzní odezvy, konstanta pro násobení prahu a použité kanály. Z dříve popsaných algoritmů je tento výpočetně nejnáročnější. Tento algoritmus nabízí potenciál pro přesnější detekci díky rozkladu signálu do 120 pásem v celém slyšitelném rozsahu. Vysoký počet pásem je dán také tím, že se navzájem překrývají. Dle původního návrhu autora je překryv nastaven na 25% šířky pásma. V případě snížení počtu pásem či odstranění jejich překryvu by bylo možno algoritmus zrychlit. Statické hodnoty testovaných parametrů jsou tyto:

- Délka použitého segmentu impulzní odezvy: 900
- Použité kanály: 2 až 40
- Hodnota detekčního prahu – konstanta pro násobení: 16

Změna délky impulzní odezvy má dle očekávání podobný průběh jako u předchozích algoritmů. Celková délka impulzní odezvy je přibližně 1000 vzorků, tedy necelých 23 milisekund. Důležité je uvědomit si, že detekční algoritmus provádí segmentaci signálů dle délky impulzní odezvy. Tedy čím delší je impulzní odezva, tím méně segmentů bude korelováno. Z toho důvodu je nutné testovat detekční algoritmus i pro případy, že je použit segment impulzní odezvy delší, než je samotná odezva. Z obrázku 27 je patrné, že od délky 900 vzorků začíná narůstat procento falešných detekcí. Nicméně až do délky 1600 vzorků narůstá i procento správně detekovaných chyb. Volba použité délky segmentu musí být kompromisem mezi těmito dvěma kritérii. Z hlediska výpočetní náročnosti se jeví lepší volba delšího segmentu, kvůli menšímu počtu korelací. Z praktického hlediska je lepší detekovat více chyb a falešné detekce ignorovat než nedetekovat chybné signály. Dle těchto kritérií bylo nejlepších výsledků dosaženo při použití impulzní odezvy o délce 1100 vzorků.

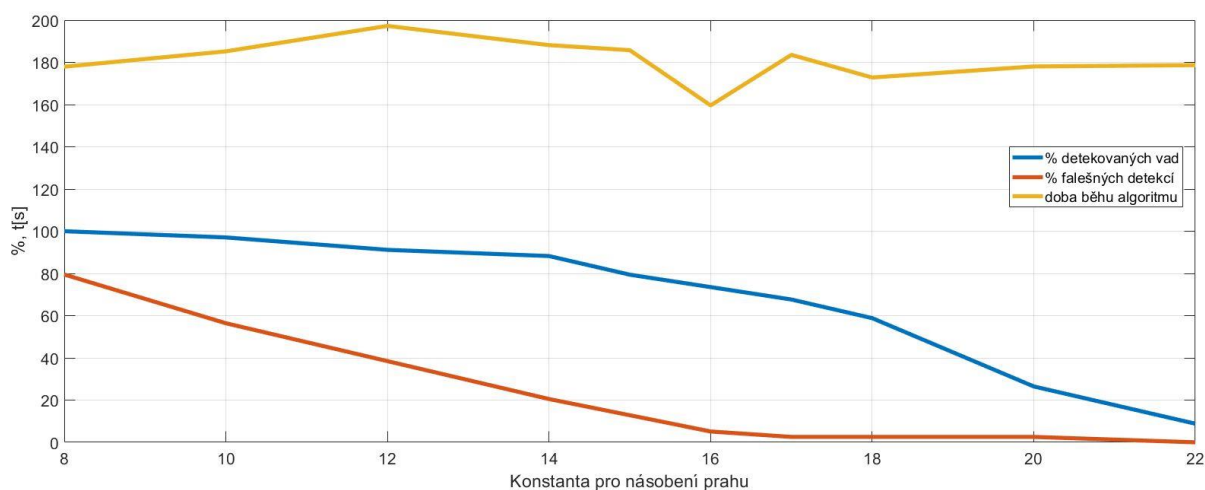


Obrázek 27: Lyon - vliv délky segmentu impulzní odezvy

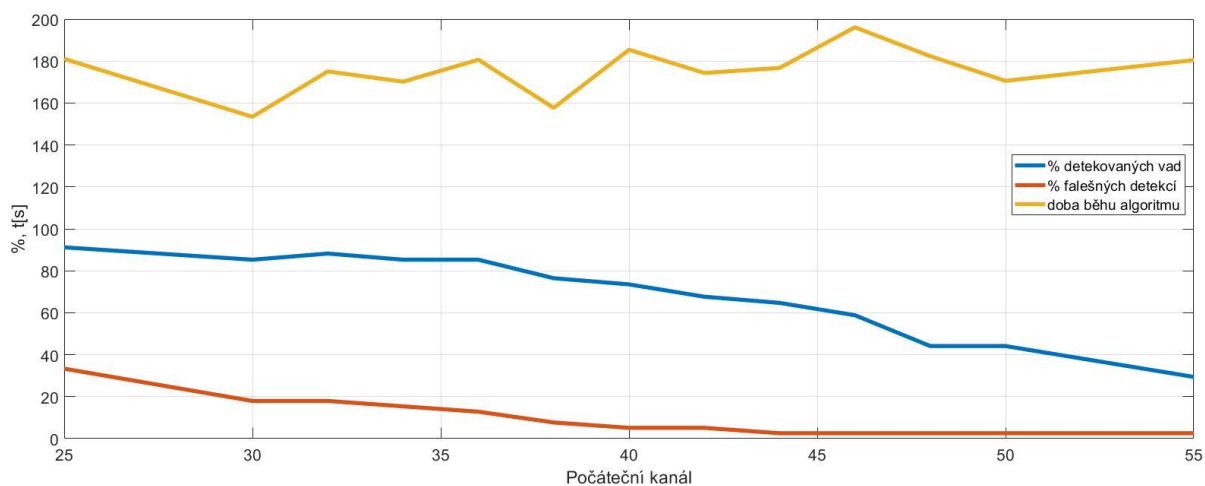
Vliv konstanty pro násobení prahu teoreticky nemá žádný vliv na výpočetní náročnost. Křivka popisující toto kritérium kolísá z neznámého důvodu. Pravděpodobně se jedná o výpočetní nestabilitu PC. Klesající trend detekovaných chyb a falešných detekcí značí, oproti algoritmu využívající ERBletovou transformaci, mírnou nestabilitu systému. I drobná změna této konstanty může způsobit výraznou změnu hodnocených kritérií. Kompromisem těchto kritérií je hodnota konstanty 16, kdy byly detekovány tři čtvrtiny chybných signálů a přibližně jedna osmina byla detekována nesprávně.

V případě použití funkce *LyonPassiveEar* k modelování sluchového ústrojí jsou jednotlivé kanály řazeny od nejvyššího frekvenčního pásma po nejnižší. Tato funkce z neznámého důvodu vrací v prvním kanále nečíselné hodnoty. Nejvyšším použitelným pásmem pro detekci je tedy až to druhé.

Při použití pouze nejvyšších pásem jsou detekovány téměř všechny chybné signály, nicméně se vyskytuje velké množství falešných detekcí. Algoritmus detekuje i nepatrné a akusticky nevýrazné zvuky impulzního charakteru. Při snaze omezit počet falešných detekcí a zachovat relativně vysoké procento správných detekcí je vhodné volit kanály 2 až 36.



Obrázek 28: Lyon - vliv konstanty pro násobení prahu



Obrázek 29: Lyon - vliv volby kanálů

Ideální volba parametrů v porovnání s výsledky subjektivních testů je tato:

- Délka použitého segmentu impulzní odezvy: 1100
- Použité kanály: 2 až 36
- Hodnota detekčního prahu – konstanta pro násobení: 16

Při této volbě parametrů bylo dosaženo těchto výsledků:

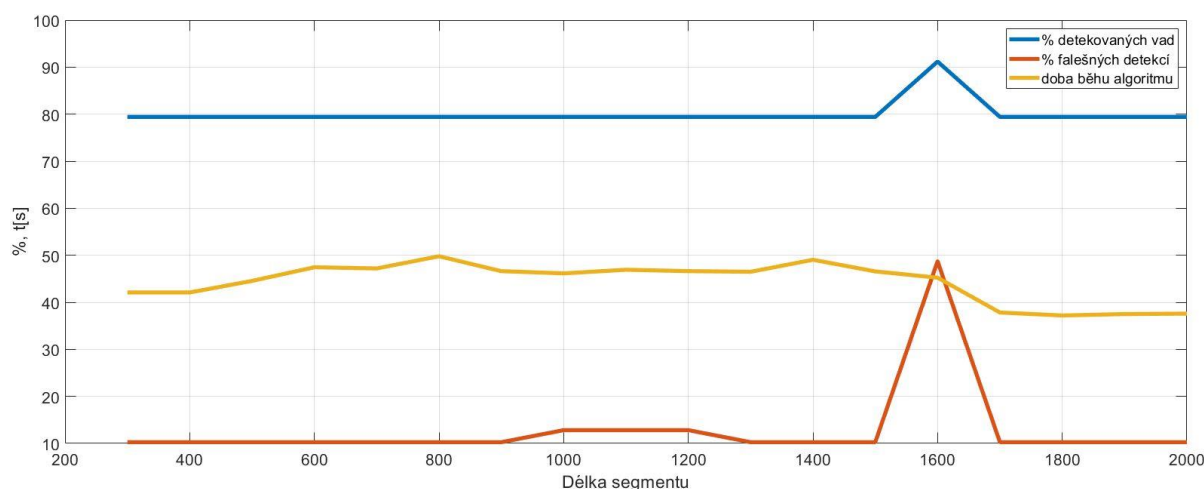
- Detekovaných chyb: 88,24 %
- Falešných detekcí: 15,38 %
- Doba běhu algoritmu: 187,79 sekund

6.3.5. Seneff

Tento detekční algoritmus využívá korelace k určení detekčního signálu a mediánové filtrace k určení detekčního prahu. Na rozdíl od předchozích algoritmů neprovádí korelaci po segmentech signálu stejně dlouhých jako je impulzní odezva, ale je korelován celý signál s impulzní odezvou. Nelze tedy předpokládat změnu výpočetní náročnosti se změnou délky impulzní odezvy. Lze očekávat, že výpočet proběhne rychleji než u ostatním modelů využívajících korelaci. Signál je zde rozdělen do 40 kanálů, z nichž první obsahuje nejvyšší frekvence. S použitím mediánové filtrace je přidán parametr délky okna mediánového filtru. Statické hodnoty testovaných parametrů jsou tyto:

- Délka použitého segmentu impulzní odezvy: 900
- Použité kanály: 1 až 20
- Hodnota detekčního prahu – konstanta pro násobení: 5
- Délka okna mediánového filtru: 150

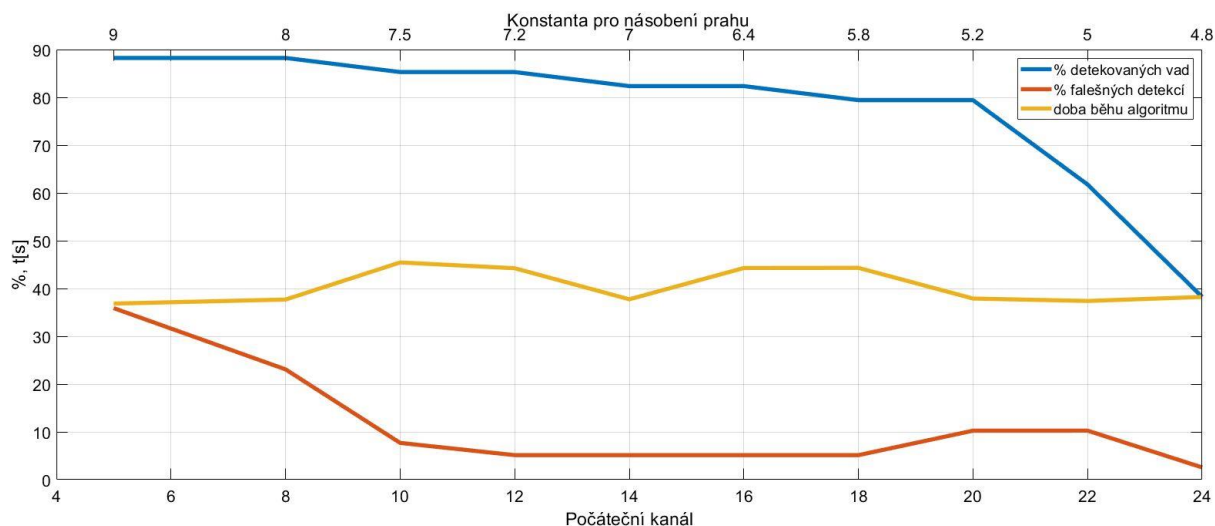
Z grafu 30 je patrné, že volba délky segmentu impulzní odezvy nemá v případě tohoto modelu téměř žádný vliv. V případě délky 1600 vzorků se vyskytla neočekávaná anomálie, jejíž příčina je neznámá. Výpočetní náročnost není ovlivněna volbou tohoto parametru.



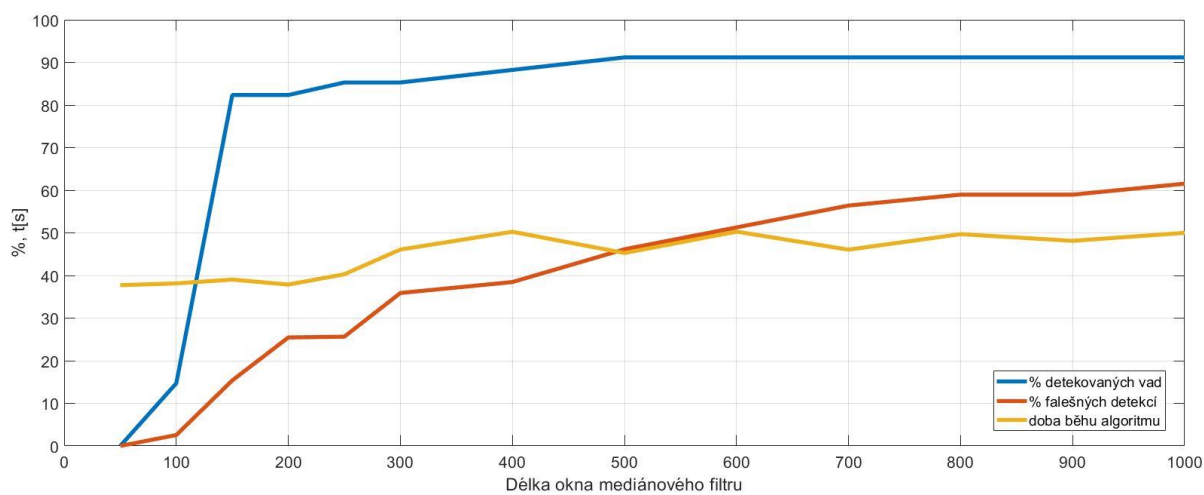
Obrázek 30: Seneff - vliv délky segmentu impulzní odezvy

V případě předchozích modelů byl detekční práh určen pomocí střední hodnoty detekčního signálu. U tohoto algoritmu je detekční práh určen obálkou signálu násobenou konstantou. Problém nastává při změně počtu kanálů, ze kterých se počítá detekční práh. Obálka signálu změní svou amplitudu, nicméně musí se dynamicky měnit i konstanta pro násobení obálky. V případě použití většího počtu kanálů by docházelo k velkému počtu falešných detekcí a naopak při použití malého počtu kanálů by algoritmus detekoval jen málo chyb. V grafu 31 je zobrazen průběh tří vyhodnocovaných kritérií v závislosti na volbě použitých kanálů a hodnoty, která násobí obálku signálu. Na osách x je zobrazen rozsah použitých kanálů (spodní osa), začínající vždy prvním kanálem a odpovídající hodnoty, kterou je násobena obálka (horní osa). Z grafu vyplývá, že ideální volbou je použít kanály 1 až 12 a obálku signálu násobit hodnotou 7,2.

Posledním parametrem je délka okna mediánového filtru, který určuje obálku signálu. V případě použití ERBletové transformace bylo ideální okno dlouhé 70 vzorků, nicméně celý detekční signál měl necelých 8000 koeficientů v každém kanále. V případě tohoto modelu má detekční signál 29000 vzorků a okno mediánového filtru musí být delší. S rostoucí délkou ovšem narůstá počet falešných detekcí. Ideální volbou je délka okna o 150 vzorcích.



Obrázek 31: Seneff - vliv použitých kanálů a hodnoty pro násobení obálky signálu



Obrázek 32: Seneff - vliv délky okna mediánového filtru

V porovnání výsledků detekčního algoritmusu se subjektivními testy byly zvoleny tyto hodnoty parametrů:

- Délka použitého segmentu impulzní odezvy: 900
- Použité kanály: 1 až 12
- Hodnota detekčního prahu – konstanta pro násobení: 7,2
- Délka okna mediánového filtru: 150

Při této volbě parametrů měla výsledná kritéria tyto hodnoty:

- Detekovaných chyb: 85,29 %
- Falešných detekcí: 5,13 %
- Doba běhu algoritmu: 40,5 sekund

Tabulka 2: Vhodná volba parametrů detekčních algoritmů

	Délka imp. odezvy	Počáteční kanál	Konstanta pro násobení prahu	Délka mediánového filtru
DRNL	900	13	22-24	-
Lopez-Poveda	900	23	19	-
ERBlet	-	30	6,5	70
Lyon	1100	36	16	-
Seneff	900	12	7,2	150

6.3.6. Výsledky pro druhou rozhodovací úroveň

Druhá rozhodovací úroveň říká, že chybné signály jsou ty, které více jak 50% posluchačů označilo za vadné a bezchybné jsou ty, které méně jak polovina posluchačů označila za chybné. V případě ponechání hodnot parametrů odvozených z testů popsaných v předchozích kapitolách jsou výsledky znatelně horší. Tabulka č. 2 zobrazuje přehled hodnocených kritérií. Z výsledků vychází najevo, že nejlepším modelem je, stejně jako u první rozhodovací úrovně, algoritmus využívající ERBletové transformace. U všech algoritmů došlo k výraznému zhoršení obou procentuálních kritérií. Detekce pomocí ERBletové transformace se opět ukázala jako nejrobustnější, jelikož i při této rozhodovací úrovni má lepší výsledky než ostatní algoritmy při první rozhodovací úrovni. Obě hodnocená kritéria se zhoršila, nicméně zhoršení není tak výrazné jako u ostatních modelů. Jelikož se jedná o testy založené na subjektivním hodnocení, je nutné brát v úvahu, že tato rozhodovací úroveň nemusí být směrodatná. Velký počet testovaných vzorků obsahuje perkusní nástroje a je pravděpodobné, že je někteří z posluchačů považovali za chybu. Zvukové úryvky jsou také relativně krátké a posluchač z nich nemusí mít dostatek informace o celkovém charakteru nahrávky. V dalších testech jsou z tohoto důvodu výsledky vyhodnocovány pouze pro první rozhodovací úroveň.

Tabulka 3: Výsledky detekčních algoritmů pro druhou rozhodovací úroveň

	Detekovaných chyb	Falešných detekcí	Doba běhu algoritmu
DRNL	65,8 %	4,08 %	35,39 s
Lopez-Poveda	65,8 %	4,08 %	102,68 s
ERBlet	90,24 %	12,24 %	35,33 s
Lyon	75,6 %	18,37 %	192,21 s
Seneff	80,49 %	16,33 %	38,03 s

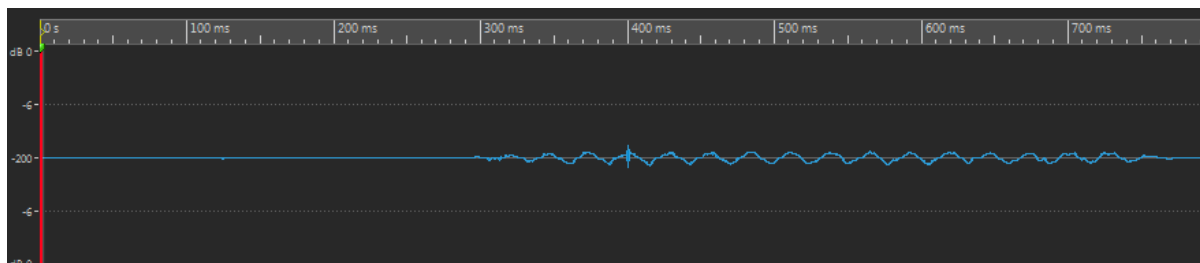
6.4. Porovnání modelů s komerčním softwarem

Účelem tohoto testu je porovnat nově navržené a implementované detekční algoritmy s placeným softwarem, který má stejnou či podobnou funkcionalitu. V našem případě se jedná o program Wavelab Pro, který má implementované funkce pro detekci a korekci anomálií v signálu. Jelikož je hlavním cílem této funkcionality pomoci uživateli v části zpracování signálu, kterou by jinak musel provádět poslechem nahrávky, je smysluplné porovnat výsledky detekce se subjektivními testy. Tento test je proveden ve třech konfiguracích pro všech devadesát vzorků. Tyto konfigurace jsou přednastaveny v softwaru a jejich parametry jsou popsány v tabulce 3. Některé z konfigurací jsou definovány pouze jedním či dvěma parametry. Je zde použita první rozhodovací úroveň, která nepočítá se vzorky, které mají nejistý výsledek (25% až 75%) z poslechových testů.

Tabulka 4: Konfigurace použité pro detekci anomálií v softwaru Wavelab Pro

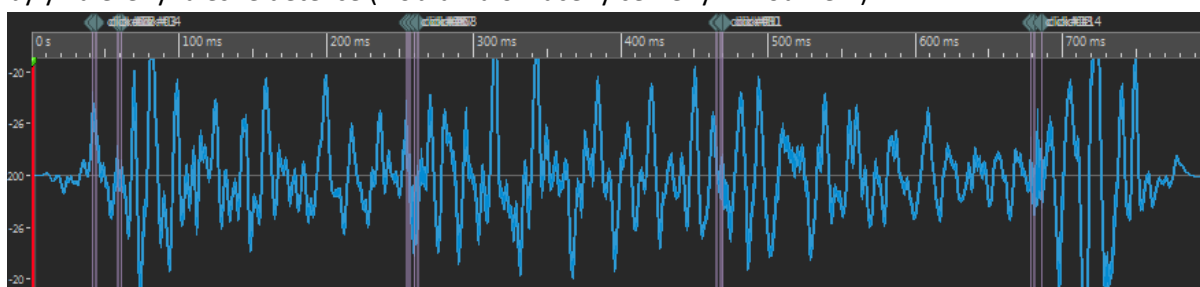
	Threshold	Detection	Detection Width
Click detection	20 %	-	-
Sensitive detection	20 %	50 %	-
Very sensitive detection	20 %	50 %	50 %

První z uvedených konfigurací se nazývá „Click detection“, ovšem po detekci s tímto nastavením je detekována pouze přibližně čtvrtina vadných vzorků. Do toho jsou započteny i jasně akusticky zřetelné lupance, které slyšeli všichni posluchači. Takový případ je znázorněn na obrázku 33. 100% posluchačů tuto vadu vnímalo a všechny algoritmy navržené v této práci ji detekovaly. Navzdory takto nízké úspěšnosti detekce byly vygenerovány i tři falešné detekce.

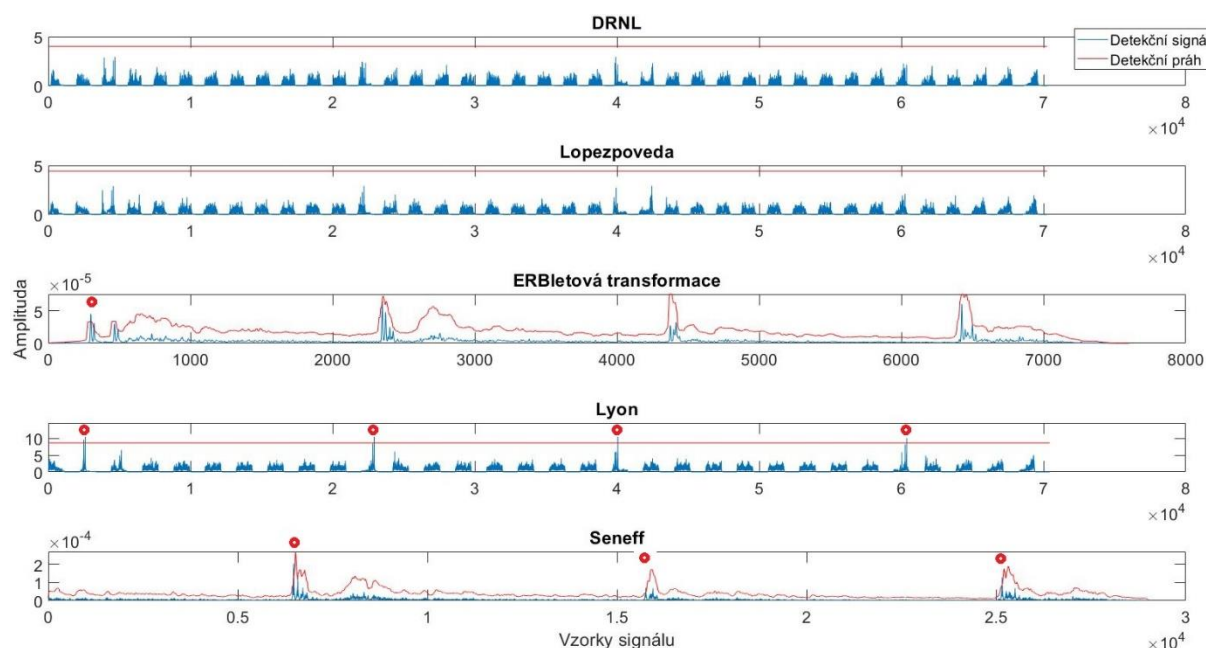


Obrázek 33: Nedetekovaný click v prostředí Wavelab Pro

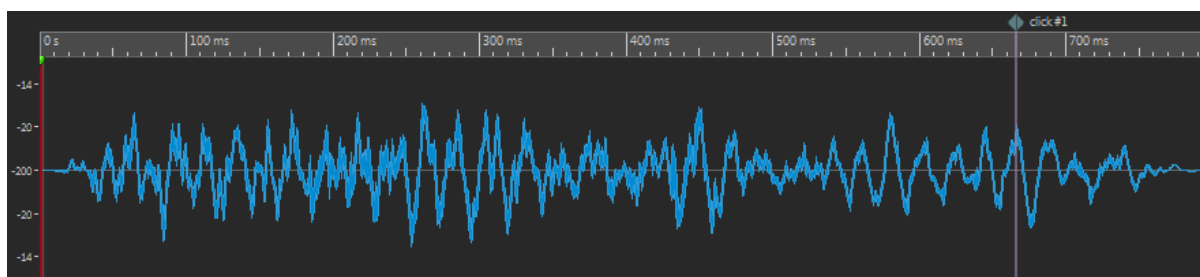
Při použití citlivější konfigurace „Sensitive detection“ je detekováno více než 80% chybných vzorků, ovšem procento falešných detekcí vzroste na necelých 40%. Toto nastavení je citlivé i na perkusní nástroje, což je zdrojem velkého množství falešných detekcí. Na obrázku 34 je zobrazen signál obsahující čtyři výrazné údery do bicích nástrojů, které jsou programem Wavelab Pro vyhodnoceny jako čtrnáct vad rozdělených do čtyř skupin. Tyto údery nemají akustický projev podobný clicku, připomíná spíše tympán. Tento signál byl označen za chybný pouze 15% respondentů. Na obrázku 35 jsou detekční signály algoritmů. Pouze DRNL a Lopez-Poveda nedetekovaly bicí. U ostatních algoritmů byly nalezeny falešné detekce (v obrázku označeny červeným kroužkem).



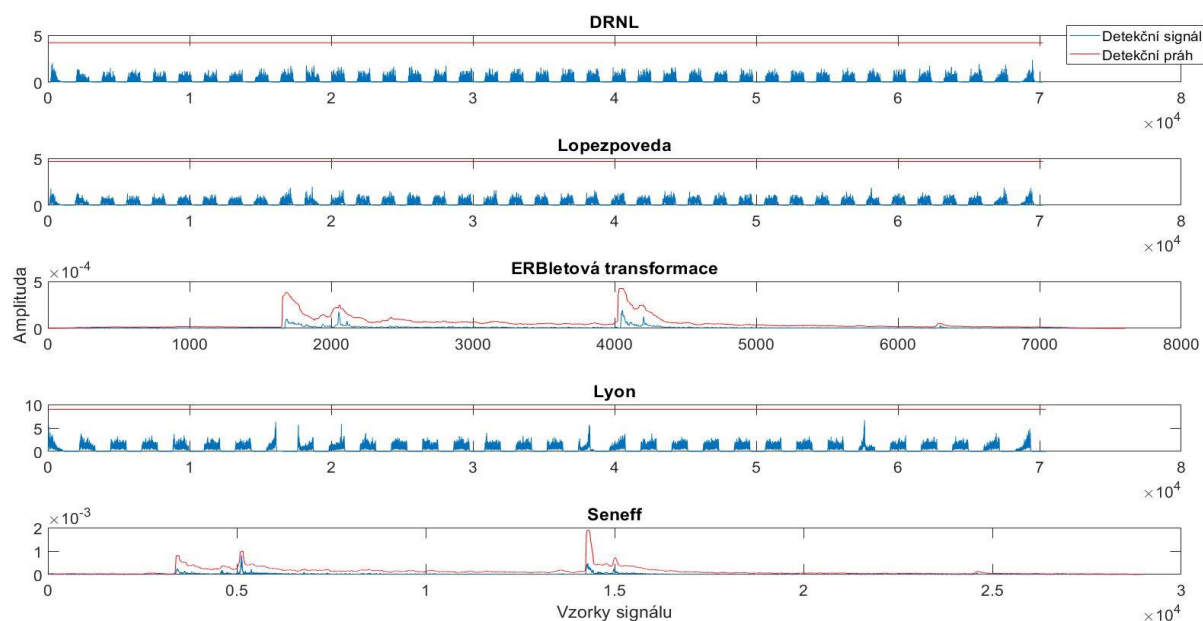
Obrázek 34: Falešné detekce bicích nástrojů detekované programem Wavelab Pro



Obrázek 35: Výstupy detekčních algoritmů pro signál s bicími nástroji



Obrázek 36: Falešná detekce v prostředí Wavelab Pro



Obrázek 37: Výstupy detekčních algoritmů

Podobný případ signálu je zobrazen na obrázku 36. Zvukový vzorek obsahuje dva údery bicích nástrojů a tři impulzy podobajících se zvuku kastanět. Dva z těchto pulzů jsou maskovány bubny a nejsou detekovány. Třetí z nich je ovšem programem Wavelab Pro falešně detekován jako chyba. Algoritmy implementované v Matlabu správně celý signál vyhodnocují jako bezvadný. Zvuk kastanět zvyšuje amplitudu detekčního signálu, např. u modelu Lyon okolo vzorku 57600, nicméně správným nastavením parametrů je detekční práh nastaven na ještě vyšší hodnotu.

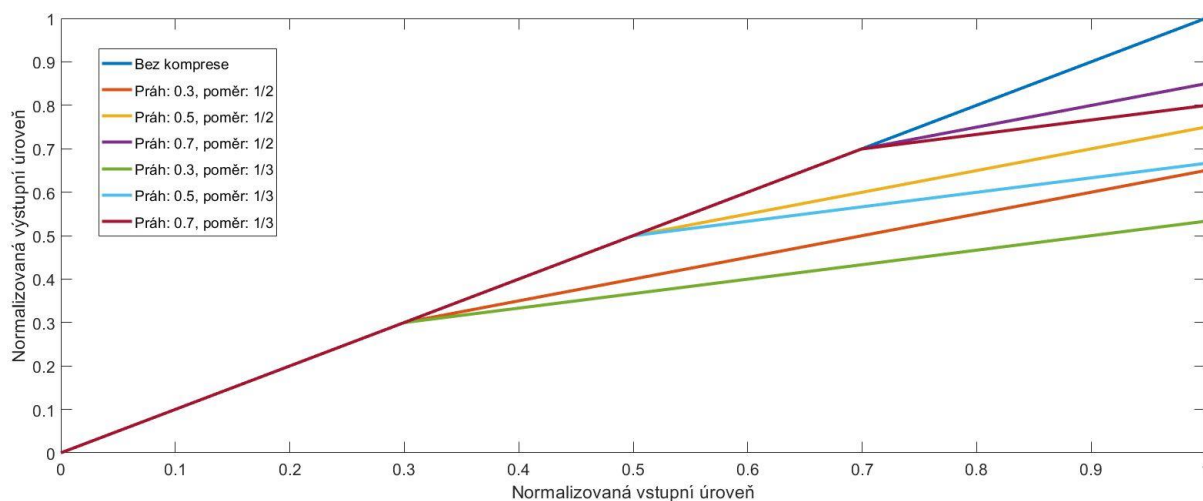
Při použití konfigurace „Very Sensitive Detection“ je výsledek procentuálně téměř stejný jako při konfiguraci „Sensitive Detection“. Nicméně nejsou vždy detekovány stejné chyby a generovány stejné falešné detekce. Z důvodu nedostatečné dokumentace není možné blíže vysvětlit tento jev ani popsat možnosti nastavení jednotlivých parametrů. Statistické vyhodnocení detekce v programu Wavelab Pro je popsáno v tabulce 4. V porovnání s algoritmy navrženými v této práci je program Wavelab Pro citlivější na anomálie, které nemají akustický projev. Současně ale dosahují horších výsledků detekce slyšitelných vad. Další nevýhodou tohoto řešení jsou omezené možnosti automatizace práce. Nelze například automaticky načíst signály, dynamicky měnit parametry detekce v závislosti na vstupním signálu či generovat statistické výsledky.

Tabulka 5: Statistické výsledky detekce v programu Wavelab Pro

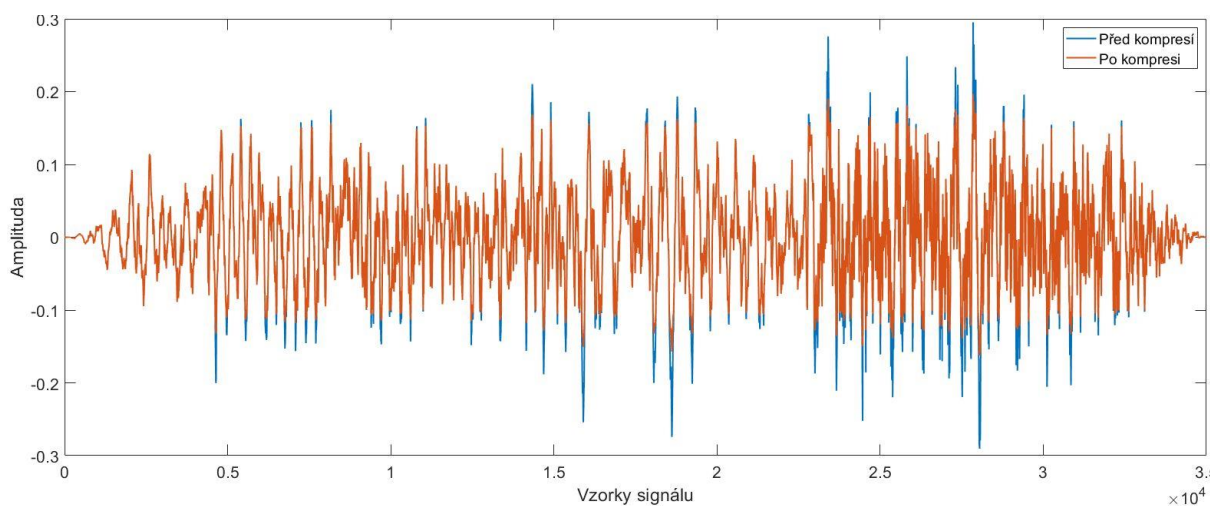
	Detekovaných chyb	Falešných detekcí
Click Detection	26,47 %	7,69 %
Sensitive Detection	85,29 %	38,46 %
Very Sensitive Detection	82,35 %	38,46 %

6.5. Test vlivu dynamiky signálu

V tomto testu byly detekční algoritmy rozšířeny o kompresní funkce, které mění dynamiku vstupního signálu. Kompresní funkce má dva základní parametry, které určují míru komprese. Prvním z nich je práh. Amplituda je zde normalizována k maximu signálu. Při prahu rovnému jedné polovině budou komprimovány všechny vzorky s vyšší absolutní hodnotou amplitudy, než je polovina maxima absolutní hodnoty signálu. Druhým parametrem je kompresní poměr, který určuje strmost kompresní funkce. Všech šest kompresních funkcí je znázorněných na obrázku 38. Kompresní poměry byly voleny jedna polovina a jedna třetina. Prahy komprese byly zvoleny v úrovních 0.3, 0.5 a 0.7.



Prvním ze dvou důležitých efektů komprese je snížení amplitudy výrazných clicků způsobených například škrábanci na desce. Tyto clicky mají často větší amplitudu než okolní signál a kompresí lze, v případě nedostatečně robustních algoritmů, zamezit jejich detekci. Druhým efektem je snížení celkového dynamického rozsahu a tím pádem i změna detekčního prahu. Efekt komprese je znázorněn na obrázku 39. Jelikož komprese ovlivňuje signál i rušení, jedná se o nejhorší možný případ. V případě reálné komprese by došlo k utlumení pouze signálu a poměr signálu k šumu by klesnul, což by teoreticky znamenalo lepší výsledky detekčních algoritmů.

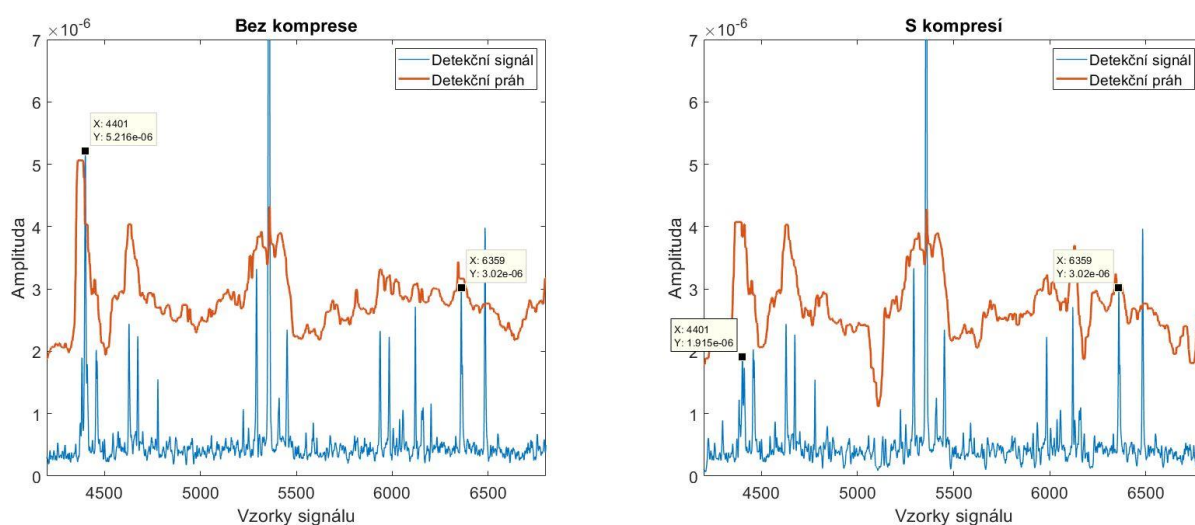


Tabulka 5 popisuje výsledky detekčních algoritmů při použití komprese s prahem 0,3 a kompresním poměrem 1/3. V tomto případě dochází k největší deformaci testovaných signálů a výsledky jsou tedy nejhorší. Detekční algoritmy DRNL a Lopez-Poveda mají zhruba o 12% (4 vzorky) horší výsledek v množství detekovaných vad a jeden vzorek navíc byl detekován falešně. Při použití méně výrazné komprese se tyto změny neprojevují vůbec nebo se jedná o maximálně dva nedetekované vzorky navíc a jednu falešnou detekci. V případě modelu Lyon má komprese pozitivní vliv na procento detekovaných vad. Tři ze šesti použitých kompresních funkcí zvýšily počet detekovaných vad o jeden a počet falešně detekovaných vzorků o jeden až dva. Všechny tři metody mají implementovaný stejný způsob detekce, pouze používají jiný model pro simulaci sluchového ústrojí. U prvních dvou výsledky naznačují, že je detekční práh příliš vysoko a kompresí lze výsledky negativně ovlivnit. Třetí model má práh nastaven níže, což znamená více falešných detekcí, avšak komprese má na výsledky detekce minimální vliv. Ostatní kompresní funkce nemění výsledky detekce tohoto modelu téměř vůbec.

Tabulka 6: Výsledky detekce po kompresi signálů (práh 0,3; poměr 1/3)

	Detekovaných vad	Falešných detekcí	Doba běhu algoritmu
DRNL	64,71 %	5,13 %	35,54 s
Lopez-Poveda	64,71 %	5,13 %	98,9 s
ERBlet	85,29 %	23,08 %	44,47 s
Lyon	91,18 %	20,51 %	207,64 s
Seneff	76,47 %	15,38 %	44,64 s

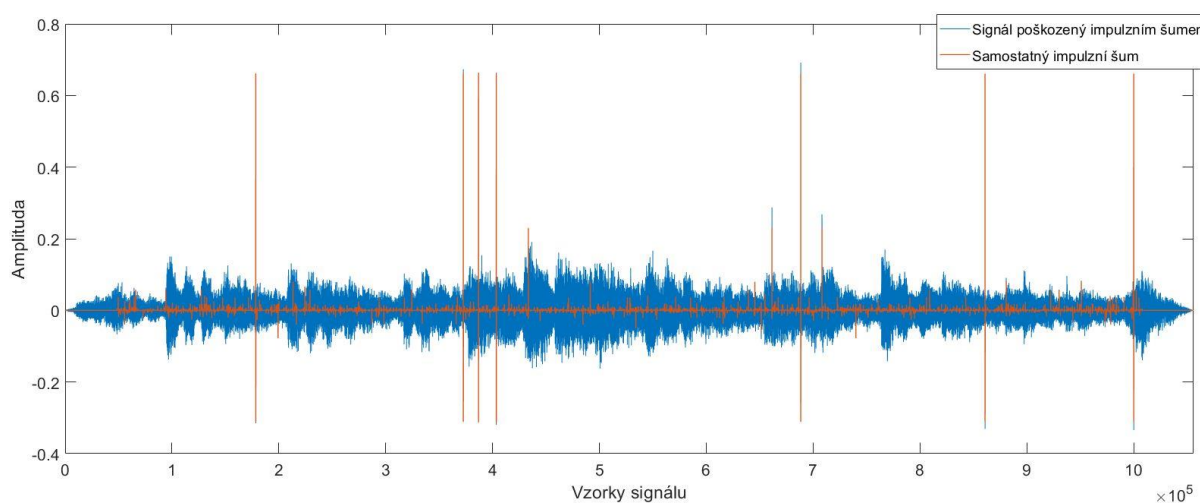
Detekční algoritmus využívající ERBletové transformace je kompresí ovlivněn nejvíce. Komprese má negativní vliv na procento detekovaných vad i na procento falešných detekcí. Z obrázku 40 je vidět, že v některých případech je impulzní šum, normálně detekovaný jako vada, výrazně potlačen (vzorek č. 4401) pod detekční práh. Toto je v důsledku mírné změny frekvenčního složení signálu po kompresi. ERBletová transformace je schopna tuto změnu zaznamenat a v některých případech se to významně projeví na detekci. V jiných případech (např. vzorek 6359) zůstává hodnota detekčního signálu stejná a pouze se sníží detekční práh. To je důsledkem snížení celkové dynamiky signálu. V případě modelu Seneff se vliv projevuje zhoršením obou procentuálních kritérií, ale pouze při použití této kompresní funkce. Ostatní funkce nemají na výsledky detekce téměř žádný vliv. Kompletní výsledky třetího testu jsou dostupné v příloze.



Obrázek 40: Vliv komprese na detekci vad

6.6. Test citlivosti detekce a schopnosti lokalizace vad

V předchozích kapitolách byly popsány statistické výsledky detekčních algoritmů, ovšem nebylo možné vyhodnotit přesný počet detekovaných vad či falešně detekovaných vad. V tomto testu jsou použity vzorky, kde máme přesné informace o rušení a je možné tyto informace vyhodnotit. Jsou zde použity dva signály, z nichž první je znázorněn na obrázku 41. Druhý signál obsahuje jinou hudební složku a stejnou šumovou složku. Jsou také dostupné oba signály bez přidaného šumu a samostatná šumová složka. Z grafického znázornění je patrné, že několik šumových pulzů je velice výrazných, ale drtivá většina z nich má oproti hudební složce signálu výrazně menší amplitudu. Všechny algoritmy jsou testovány s hodnotami parametrů odvozenými v prvním testu. Tím je možno zjistit nejen celkový počet chyb, ale i o které chyby se jednalo. Jelikož jsou algoritmy nastaveny na detekci pouze slyšitelných vad, lze očekávat velký počet nedetekovaných vad.



Obrázek 41: Znárodnění hudebního a šumového signálu

Důležitou vlastností detekčních algoritmů je jejich citlivost, kterou lze odvodit z počtu detekovaných vad. V tabulce 6 je přehled výsledků detekce všech algoritmů na dvou testovaných signálech. Celkem je v signálech 2175 impulzních defektů. První dva algoritmy detekují pouze minimum vad v řádu desítek. Ostatní algoritmy jsou znatelně citlivější. Celkový počet detekovaných vad je v řádu stovek. Nejvíce vad detekoval model Seneff na druhém z testovaných signálů. Celkem detekoval 657, tedy zhruba 30% všech vad, což odhadem odpovídá slyšitelné části všech chyb. Zároveň nebyla detekována jediná vada falešně. Obecně je u těchto dvou signálů naprosté minimum falešných detekcí. Na prvním signálu model Seneff detekoval 341 vad správně a na tom stejném signálu bez přidané šumové složky pouze jednu falešně. Tuto falešnou detekci vykazují i výsledky detekčních algoritmů ERBletové transformace a Lyon. Poslechem byla ověřena přítomnost této vady v údajně bezvadném signálu a nejedná se tedy o falešnou detekci.

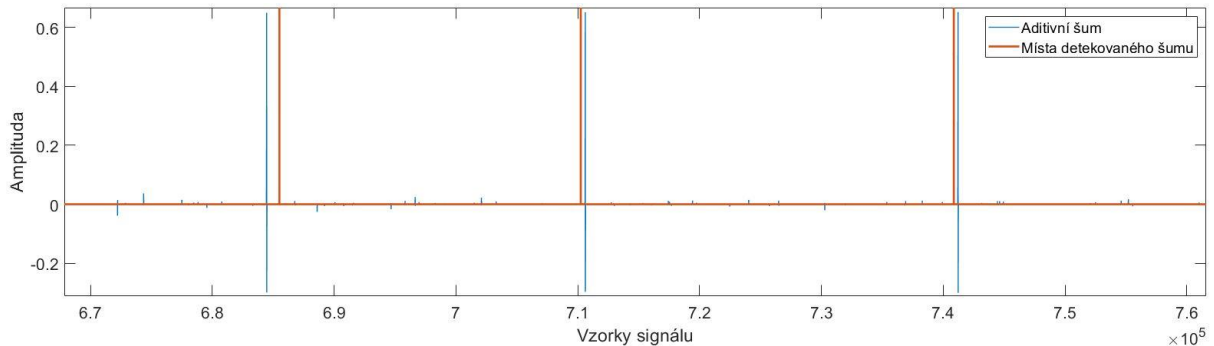
Tabulka 7: Počet detekovaných vad

	DRNL	Lopez-Poveda	ERBlet	Lyon	Seneff
1. signál	49	26	284	488	341
2. signál	58	53	355	476	657

V případě testování jednoho delšího vzorku můžeme, oproti testování devadesáti krátkých vzorků, pozorovat změnu ve výpočetní náročnosti. V případě metod využívajících korelaci po segmentech (DRNL, Lopez-Poveda, Lyon) je v případě jednoho signálu o délce 22 sekund znatelný nárůst výpočetní doby oproti zpracování devadesáti vzorků o délce 800 ms. V případě zbylých dvou algoritmů je výpočet detekce v jednom delším signálu naopak rychlejší.

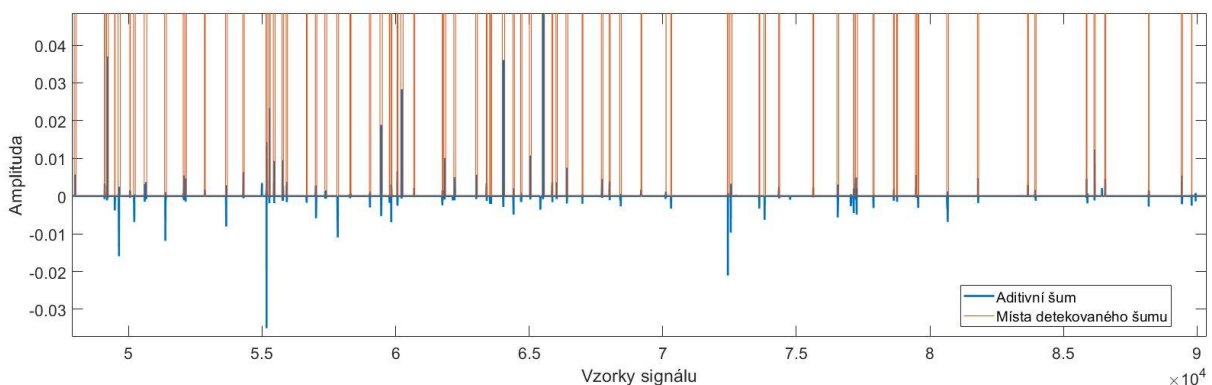
Tabulka 8: Výpočetní náročnost algoritmů

Délka signálů	DRNL	Lopez-Poveda	ERBlet	Lyon	Seneff
90x 800 ms	43 s	95 s	35 s	188 s	41 s
1x 22 s	87 s	276 s	11 s	467 s	16 s



Obrázek 42: Detekce vad u korelačních algoritmů

Důležitým aspektem detekčních algoritmů je jejich schopnost lokalizace vad. Detekční algoritmy generují výstupy porovnávané s detekčním prahem, které jsou různé délky oproti vstupnímu signálu. Z tohoto důvodu je nutno signály interpolovat tak, aby měly stejnou délku a bylo možné je vzájemně porovnat a lokalizovat detekovaný šum. U modelů využívajících korelaci po segmentech je toto prakticky nemožné. Přesnost lokalizace je tedy velmi omezena. Obrázek 42 zobrazuje část signálu, ve kterém jsou tři velmi výrazné clicks a označení míst, ve kterých detekční signál překročil detekční práh. Impulzní šum je lokalizován nepřesně a pokaždé v jiné vzdálenosti od správné pozice. Z tohoto důvodu nelze oba signály synchronizovat. Pozici vady lze tedy specifikovat pouze v určitém tolerančním pásmu, kde se ovšem může vyskytovat více dalších vad. V případě ERBletové transformace a modelu Seneff lze zpřesnit lokalizaci interpolací signálů. Výsledek lokalizace po interpolaci detekčního signálu vygenerovaného algoritmem s ERBletovou transformací je znázorněn na obrázku 43. Lokalizace je zde přesná na jednotky vzorků. Případnou nepřesnost způsobuje mediánový filtr.



Obrázek 43: Detekce vad u ERBletové transformace

6.7. Porovnání s dalšími detekčními metodami

Detekční algoritmy popsané v kapitolách 3.1, 3.2 a 3.3 byly v [8] testovány na stejné sadě vzorků jako algoritmy v této práci. Vyhodnoceny byly pouze procentuální kritéria, nikoliv výpočetní náročnosti. Výsledky těchto testů, včetně navrženého vylepšení (v závorce), jsou popsány v tabulce 9. Z této tabulky vyplývá, že dvě nejefektivnější detekční metody využívají ERBletovou a vlnkovou transformaci. Tyto dvě metody dosahují stejných výsledků na první sadě testovaných vzorků. V literatuře týkající se detekce impulzních degradací je nejčastěji skloňována metoda využívající autoregresní predikce. Tato metoda ovšem dosahuje horších výsledků, které jsou srovnatelné s algoritmy Lyon či Seneff.

Tabulka 9: Souhrn výsledků pro detekční algoritmy testované v [8]

	Detekovaných vad	Falešných detekcí
AR model	88,24 %	15,38 %
AR + přizpůsobený filtr	82,35 %	23,08 %
Vlnková transformace	94,12 %	10,26 % (5,13 %)

Metoda detekce popsaná v kapitole 3.4 [9] byla testována na vzorcích dodaných společností GZ Media, a.s. Jedná se o vzorky extrahované z gramofonových desek, podobného charakteru jako vzorky první sady použité v této práci. Autoři použili 30 vzorků pro natrénování parametrů a 59 vzorků pro testování algoritmu. Tento model dosáhl, při použití první rozhodovací úrovně, úspěšnosti detekce 87,5% a 3,7% chyb detekoval falešně. V případě druhé rozhodovací úrovně dle očekávání nastal pokles detekovaných vad. Překvapivě se nezvýšilo procento falešných detekcí. Největší nevýhodou použití hydrodynamického modelu je jeho výpočetní náročnosti. Při aplikování tohoto algoritmu na první sadu vzorků je doba výpočtu přibližně 2400 sekund. V této práci byly navrženy detekční algoritmy (DRNL, Lopez-Poveda, Lyon), které fungují na totožném principu, ovšem s využitím jiného modelu. Všechny tyto algoritmy jsou výpočetně méně náročné, ovšem dosahují znatelně horších výsledků, než je publikováno v [9]. Model Seneff se svými výsledky blíží této efektivitě a zároveň je více než padesátkrát rychlejší. ERBletová transformace dosahuje dokonce lepších výsledků s výrazně menšími výpočetními požadavky.

Tabulka 10: Souhrn výsledků pro detekční algoritmus využívající hydrodynamický model slyšení [9]

	Detekovaných vad	Falešných detekcí
1. rozhodovací úroveň	87,5 %	3,7 %
2. rozhodovací úroveň	78,1 %	3,8 %

7. Vyhodnocení výsledků

Pět algoritmů pro detekci impulzních vad, navržených a implementovaných v této práci, bylo testováno z hlediska úspěšnosti detekce i výpočetní náročnosti. První test umožnil analyzovat vliv jednotlivých parametrů detekčních algoritmů na jejich efektivitu. Nastavení parametrů bylo optimalizováno pro detekci pouze slyšitelných degradací. Porovnáním výstupů detekčních algoritmů se subjektivními testy byly zjištěny výsledky tří hodnocených kritérií (tabulka 8). Prvním testovaným kritériem je procento detekovaných vad. V tomto ohledu jsou nejhorší algoritmy využívající model popsany v kapitole 4.1, tedy DRNL a Lopez-Poveda. Tento model je navržen ve frekvenčním pásmu omezeném do 8 kHz. Z tohoto důvodu je, již dle teoretického předpokladu o spektrálních vlastnostech impulzních vad, tento model nejméně vhodný. Tyto algoritmy úspěšně detekovaly přibližně tři čtvrtiny chybných vzorků. Vliv použité banky filtrů či způsobu implementace filtrů modelu DRNL neměl na procentuální úspěšnost detekce žádný vliv a výsledky obou algoritmů jsou identické. Detekční model využívající ERBletové transformace má nejvyšší procento detekovaných vad. Dle prvního testu je oproti ostatním algoritmům odolný na změnu volby jednotlivých parametrů, tedy i na celkový charakter signálu. Detekční metody Lyon a Seneff mají úspěšnost velice podobnou, přibližně 85%, nicméně stále podstatně horší než při použití ERBletové transformace.

Procento falešných detekcí je u většiny detekčních metod minimální. Jedná se o jeden či maximálně dva falešně detekované hudební vzorky z 39 bezchybných vzorků. V případě modelu Lyon je celkem šest testovaných vzorků falešně označeno jako chybné. V některých aplikacích může být fi tento výsledek akceptovatelný s přihlédnutím k vysokému procentu správně detekovaných vad.

V prvním testu bylo analyzováno devadesát krátkých zvukových vzorků. Výpočetně nejnáročnějším je algoritmus Lyon. Analýza sady vzorků o celkovém trvání 72 sekund proběhla za více než tři minuty. Tento model je tedy nepraktický pro aplikace v reálném čase. Druhým výpočetně nejnáročnějším algoritmem je Lopez-Poveda, který detekci vyhodnotil v průměru za 95 sekund. Zbylé tři algoritmy jsou podstatně rychlejší. Průměrná délka zpracování se pohybuje v okolí 40 sekund. Při použití ERBletové transformace pro detekci byl výpočet nejrychlejší. To je především z toho důvodu, že tento algoritmus nevyužívá korelace, jejíž implementace v prostředí Matlab je výpočetně náročná.

Tabulka 11: Účinnost detekce a výpočetní náročnost navržených algoritmů

	Detekovaných vad	Falešných detekcí	Doba běhu algoritmu
DRNL	76,47 %	2,56 %	43 s
Lopez-Poveda	76,47 %	2,56 %	95 s
ERBlet	94,12 %	5,13 %	35 s
Lyon	88,24 %	15,38 %	188 s
Seneff	85,29 %	5,13 %	41 s

Druhý test poskytl informaci o tom, jak efektivní jsou jednotlivé algoritmy v porovnání s komerčním softwarem Wavelab Pro. Nástroj pro detekci impulzních vad je pomůckou pro uživatele, který by bez něj byl nucen hledat chyby poslechem zvukové stopy. Měl by tedy být schopen detekovat všechny slyšitelné degradace. Testovány byly tři přednastavené konfigurace detekčního nástroje. Výsledky ovšem ukazují, že ani při nejcitlivějším nastavení není software Wavelab Pro schopen detekovat více jak 85% slyšitelných vad. V tomto ohledu jsou výsledky srovnatelné s detekčními modely Lyon a Seneff. Citlivé nastavení tohoto nástroje je ovšem důvodem pro téměř 40% falešných detekcí, což je v porovnání se všemi algoritmy navrženými v této práci znatelně horší výsledek. Výpočetní náročnost nebylo možno testovat z důvodu nutnosti manuálního provedení detekce.

Odhadem z prováděných experimentů lze tvrdit, že samotná detekce v programu Wavelab Pro je velice rychlá a minimálně srovnatelná s algoritmem ERBlet. Ovšem není zde zahrnuto převzorkování signálu, jeho normalizace a vyhodnocení výsledků implementované v detekčních algoritmech v Matlabu.

Ve třetím testu byla sada devadesáti testovaných signálů podrobena kompresním funkcím, které upravily dynamiku signálu. Tímto byla rozšířena množina testovaných zvuků a otestována robustnost detekčních algoritmů. Testy prokázaly, že většina kompresních funkcí ovlivní výsledky pouze minimálně. Většinou se jedná pouze o zhoršení hodnocených kritérií v rozmezí pěti procent. Pouze kompresní funkce s prahem 0,3 a kompresním poměrem 1/3 ovlivnila všechny výsledky výrazněji. Při takové kompresi je signál velmi deformován, je výrazně změněn jeho akustický projev a v reálné aplikaci je podobný případ velice nepravděpodobný. Celkový vliv kompresních funkcí na výsledky detekce je překvapivě malý a algoritmy lze v tomto ohledu považovat za robustní.

Výstupem čtvrtého testu je informace o celkové citlivosti algoritmů a schopnosti lokalizace impulzních degradací. Test byl proveden na dvou hudebních signálech s reálnými vadami, jejichž počet přesahoval 2000. V Tabulce 6 je celkový počet vad nalezených jednotlivými algoritmy. Obdobně jako u prvního testu i zde je detekováno malé množství vad algoritmy založenými na modelu DRNL. Ostatní algoritmy detekují až 30% všech chyb. Toto procento lze považovat za vysoké s přihlédnutím k tomu, že jsou algoritmy nastaveny na detekci pouze slyšitelných vad. Z celkového počtu vad je akusticky výrazných pouze menšina. Zároveň byla algoritmy nalezena maximálně jedna falešná detekce v celém signálu. Při testu na údajně bezchybných hudebních vzorcích byly nalezeny impulzní vady, které se akusticky projevují. Tyto testy potvrzují robustnost a univerzálnost navržených detekčních algoritmů. Důležitým výstupem tohoto testu je informace o neschopnosti přesné lokalizace vad algoritmy, které využívají korelace segmentů signálu s impulzní odezvou použitého modelu. Z obrázku 42 je vidět posuv detekčních signálů a detekovaných chyb vůči testovanému signálu způsobený korelací po segmentech. Pouze algoritmy ERBlet a Seneff je možno synchronizovat s testovaným signálem a přesně určit polohu impulzních degradací. Tento test také odhalil výrazný nárůst výpočetní náročnosti algoritmů DRNL, Lopez-Poveda a Lyon při testování delších hudebních vzorků. Naopak u algoritmů ERBlet a Seneff je doba výpočtu relativně kratší než při testu na první sadě vzorků.

V porovnání s metodami popsány ve třetí kapitole dosahují některé algoritmy implementované v této práci stejných či lepších výsledků. Konkrétně se jedná o metody Seneff a ERBlet. V případě metody Seneff se jedná zejména o výrazně nižší výpočetní nároky. Tato metoda dosahuje lepších výsledků než autoregresní model, ovšem nepatrně horších než metoda využívající hydrodynamický model. Použití hydrodynamického modelu znamená více než padesátkrát pomalejší výpočet. Algoritmus založený na ERBletové transformaci dosahuje lepších výsledků a nižších výpočetních nároků. Tento algoritmus vykazuje identické výsledky jako detektor založený na vlnkové transformaci.

Při zohlednění všech tří hodnocených kritérií se jako nejefektivnější jeví detekční algoritmus využívající ERBletovou transformaci. Dosahuje největšího procenta detekovaných vad, malého procenta falešných detekcí a je ze všech testovaných metod detekce nejméně výpočetně náročný. Svou efektivitou se rovná detekčnímu algoritmu využívajícímu vlnkovou transformaci [8].

8. Závěr

Cílem této práce bylo seznámit se s metodami detekce impulzních degradací v analogovém záznamu zvuku, se zaměřením na algoritmy využívající model slyšení, vybrané metody implementovat, otestovat je z hlediska efektivity detekce a výpočetní náročnosti a výsledky porovnat. V první části práce je popsán analogový záznam zvuku, proces výroby gramofonových desek a vlastnosti impulzního šumu. V druhé části byly představeny dnes již dobře známé i netradiční metody detekce. Jedna z těchto metod využívá komplexního modelu slyšení k filtraci signálu a následné korelace s impulzní odezvou tohoto modelu k detekci ruchů impulzního charakteru [9]. Tato metoda poskytuje slibné výsledky detekce, ovšem velice výpočetně náročnou implementaci. Na tuto metodu je navázáno se snahou nalézt méně výpočetně náročný model, který by poskytl stejné či lepší výsledky detekce.

Ve čtvrté kapitole jsou představeny modely slyšení, které lze využít k filtraci signálu a následné detekci impulzních degradací obdobně jako v [9]. Jsou zde představeny tři modely simulující průchod zvuku skrze sluchové ústrojí až k bazilární membráně či na úroveň nervových zakončení. Je zde popsána i možnost časově frekvenční reprezentace signálu pomocí ERBletové transformace [18], která umožňuje analýzu signálu pomocí neuniformní banky filtrů založené na fyziologickém vnímání zvuku. Pro každý z těchto modelů je následně navržen a implementován detekční algoritmus v prostředí Matlab. Tyto algoritmy jsou optimalizované pro detekci slyšitelných vad v analogové záznamu zvuku.

V šesté kapitole jsou představeny zvukové vzorky použité pro ověření funkce uvedených algoritmů. Pro testování je použito devadesát vzorků, které byly extrahovány z poškozených gramofonových desek a dodány společností GZ Media, a.s. Jedná se o vzorky, na kterých byly v [8] provedeny subjektivní poslechové testy, jejichž výsledky jsou referencí pro objektivní vyhodnocení kvality detekčních algoritmů. V testech byla vyhodnocována tři kritéria – procento detekovaných chyb, procento falešných detekcí a doba běhu algoritmu.

Celkem byly provedeny čtyři testy. První z nich ukázal, že metoda detekce založená na ERBletové transformaci má nejvyšší úspěšnost detekce a je nejméně výpočetně náročná. Úspěšnost detekce dosáhla 94% a pouze okolo 5% vad bylo detekováno falešně. Všech devadesát hudebních vzorků bylo vyhodnoceno za přibližně 35 sekund. Druhý nejlepší algoritmus byl založen na modelu S.Seneff. Úspěšnost detekce 85%, pět procent falešných detekcí a běh algoritmu okolo 41 sekund jsou velmi dobrými výsledky. V porovnání s komerčním softwarem Wavelab Pro jsou tyto algoritmy citlivější na slyšitelné degradace a zároveň vykazují méně falešných detekcí. Změnou dynamiky signálu se výsledky detekce mění pouze minimálně a lze tedy algoritmy považovat za robustní. Pouze dva dříve uvedené algoritmy jsou schopné přesně lokalizovat impulzní šum. Tyto algoritmy prokázaly schopnost detekovat nepatrné degradace i na studiových, údajně bezvadných nahrávkách.

Úspěšnost detekce jednotlivých algoritmů byla porovnána i s běžně používanými modely (autoregresivní model, vlnková transformace) a bylo zjištěno, že poskytují stejnou či lepší účinnost. V porovnání s detekční metodou prezentovanou v [9] jsou výsledky algoritmu využívajícího ERBletovou transformaci lepší i s nižší výpočetní náročností. Metoda detekce impulzních degradací s využitím modelu S. Seneff výsledky zaostává nepatrně za modelem z [9], ovšem nabízí více jak 50krát nižší výpočetní nároky. Modely založené na simulaci sluchového ústrojí a použití korelace s jejich impulzní odezvou tedy neposkytují tak dobré výsledky jako model s ERBletovou či vlnkovou transformací. Zároveň jsou mnohdy výrazně komplikovanější a výpočetně náročnější. V dalších pracích bude tedy pozornost věnována modelům s časově frekvenční reprezentací signálu a to nejen k detekci impulzního šumu na gramofonových deskách, ale například k restaurování archivních záznamů, detekci kompresních artefaktů, kontrole studiových nahrávek či klasifikaci degradací v signálech.

Reference

- [1] I. Bašta, Zpracování a záznam signálu, Praha: ČVUT, 1994.
- [2] GZ Media, a.s., *Interní zdroje*.
- [3] GZ Media, a.s., „DMM technology,“ GZ Media, a.s., 2017. [Online]. Dostupné z: <http://www.gzvinyl.com/Manufacturing/Mastering/DMM-cutting.aspx>. [Přístup získán 10. 2. 2019].
- [4] S. J. Godsil a P. J. Rayner, Digital Audio Restoration - a statistical model based approach, Michigan: Springer, 1998.
- [5] S. V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, London: John Wiley & Sons, Ltd., 2008.
- [6] S. Godsil a P. Rayner, Digital Audio Restoration, London: Springer-Verlag, 1998.
- [7] M. Bolek, Detekce vad u mechanického záznamu zvuku, diplomová práce, Praha: ČVUT, 2016.
- [8] M. Semanský, Detekce impulzních vad v záznamu zvuku na gramofonových deskách, bakalářská práce, Praha: ČVUT, 2017.
- [9] F. Rund, V. Vencovský a J. Bouše, „Detection of Clicks in Analog Records Using Peripheral-Ear Model,“ v *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16)*, Brno, 2016.
- [10] L. Oudre, „Automatic Detection and Removal of Impulsive Noise in Audio Signals,“ *Image Processing On Line*, pp. 267-281, 21. listopad 2015.
- [11] S. Vaseghi a P. Rayner, „A New Application of Adaptive Filters for Restoration of Archived Gramophone Records,“ v *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, New York, 1988.
- [12] C.-L. Liu, „A Tutorial of the Wavelet Transform,“ 23. únor 2010. [Online]. Dostupné z: <http://disp.ee.ntu.edu.tw/tutorial/WaveletTutorial.pdf>. [Přístup získán 17. květen 2019].
- [13] R. Nongpiur a D. Shpak, „Impulse-noise suppression in speech using stationary wavelet transform,“ *Journal of the Acoustical Society of America*, pp. 866-879, únor 2013.
- [14] R. Nobili, A. Vetešnik, L. Turicchia a F. Mammano, „Otoacoustic emissions from residual oscillation of the cochlear basilar membrane in a human ear model,“ *Journal of the Association for Research in Otolaryngology*, pp. 478-494, 2003.
- [15] E. A. Lopez-Poveda a R. Meddis, „A human nonlinear cochlear filterbank,“ *The Journal of the Acoustical Society of America*, pp. 3107-3118, prosinec 2001.
- [16] R. Meddis, L. P. O'Mard a E. A. Lopez-Poveda, „A computational algorithm for computing nonlinear auditory,“ *Journal of the Acoustical Society of America*, pp. 2852-2861, červen 2001.

- [17] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Cambridge: Emerald Group Publishing Limited, 2012.
- [18] T. Necciari, P. Balasz, N. Holighaus a P. Sondergaard, „The Erblet Transform: An Auditory-Based Time-Frequency Representation with Perfect Reconstruction,” v *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, 2013.
- [19] P. Balasz, M. Dörfler, F. Jeillet, N. Holighaus a G. Velasco, „Theory, implementation and applications of nonstationary Gabor frames,” *Journal of Computational and Applied Mathematics*, pp. 1481-1496, 2011.
- [20] S. Seneff, „A joint synchrony/mean-rate model of auditory speech processing,” *Journal of Phonetics*, pp. 55-76, 1988.
- [21] R. F. Lyon, „A Computational Model of Filtering, Detection and Compression in the Cochlea,” v *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Paris, 1982.
- [22] R. Meddis, „Matlab auditory periphery (MAP),“ 2011. [Online]. Dostupné z: <https://github.com/rmeddis/MAP>. [Přístup získán 17 Květen 2019].
- [23] „lopezpoveda2001 - Dual Resonance Nonlinear Filterbank,” [Online]. Dostupné z: <http://amtoolbox.sourceforge.net/amt-0.9.9/doc/models/lopezpoveda2001.php>. [Přístup získán 17 Květen 2019].
- [24] „The Auditory Modelling Toolbox,” [Online]. Dostupné z: <http://amtoolbox.sourceforge.net/download.php>. [Přístup získán 17 Květen 2019].
- [25] „ICASSP 2013: The ERblet Transform,” Austrian Academy of Sciences, [Online]. Dostupné z: https://www.kfs.oeaw.ac.at/index.php?option=com_content&view=article&id=672:icassp2013-erblets&catid=198&lang=en&Itemid=794. [Přístup získán 17 Květen 2019].
- [26] „LTFAT Reference manual, strana 225,” [Online]. Dostupné z: <http://lftfat.github.io/doc/lftfat.pdf>. [Přístup získán 17 Květen 2019].
- [27] M. Slaney, „Auditory Toolbox,” Interval Research Corporation, [Online]. Dostupné z: <https://engineering.purdue.edu/~malcolm/interval/1998-010/AuditoryToolboxTechReport.pdf>. [Přístup získán 17 Květen 2019].
- [28] B. D. Bryant a J. D. Gowdy, „Simulation of Stages I and II of Seneff's Auditory Model (SAM) Using Matlab,” v *Proceedings of the 1993 Matlab User's Group Conference*, 1993.
- [29] M. Ciolek a M. Niedźwiecki, „Detection of Impulsive Disturbances in Archive Audio Signals,” v *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, 2017.

Příloha

Příložené DVD obsahuje .m soubory se všemi implementovanými algoritmy a funkcemi potřebnými pro jejich spuštění v prostředí Matlab. Obsahuje také soubor *Výsledky 3. testu.xlsx* popisující výsledky 3. testu. Z [8] jsou přiloženy výsledky subjektivních testů.

Skripty a funkce:

- detekce.m (skript pro výběr algoritmu a spuštění detekce, dokumentace potřebných funkcí)
- semansky_DRNL.m
- semansky_lopezpoveda.m
- semansky_erblet.m
- semansky_lyon.m
- semansky_seneff.m
- vyhodnoceni.m (funkce pro vyhodnoceni výsledků)

Impulzní odezvy:

- semansky_MAP_DRNL_impBM.mat
- semansky_lopezpoveda_impBM.mat
- semansky_lyon_impBM_dec_1.mat
- semansky_seneff_impBM.mat

Výsledky subjektivních testů

- subjective_results_1_sada.mat [8]

Výsledek testu vlivu změny dynamiky signálu

- Výsledky 3. testu (dynamika signalu).xlsx