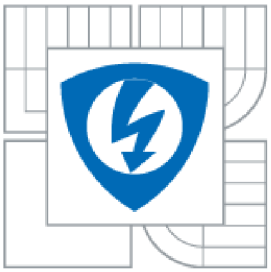
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

VYUŽITÍ PSYCHOAKUSTICKÉHO MODELU A TRANSFORMACE TYPU WAVELET PACKET PRO VODOZNAČENÍ AUDIO SIGNÁLŮ

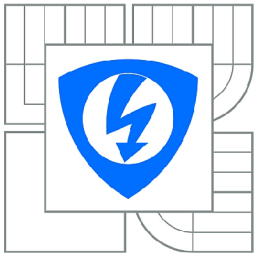
UTILIZING PSYCHOACOUSTICS MODEL AND WAVELET PACKET TRANSFORM FOR PURPOSES OF AUDIO SIGNAL WATERMARKING

DIPLOMOVÁ PRÁCE
MASTER´S THESIS

AUTOR PRÁCE
AUTHOR
Bc. TOMÁŠ HEITEL

VEDOUCÍ PRÁCE
SUPERVISOR
Mgr. PAVEL RAJMIC, Ph.D.

BRNO 2010



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky a komunikačních technologií

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student:	Bc. Tomáš Heitel	***ID:***	83764
Ročník:	2	***Akademický rok:***	2009/2010

NÁZEV TÉMATU:

Využití psychoakustického modelu a tranformace typu wavelet packet pro vodoznačení audio signálů

POKYNY PRO VYPRACOVÁNÍ:

Nastudujte a popište transformaci známou jako Discrete Wavelet Packet Transform (DWPT). Uveďte také způsob jejího využití pro psychoakustický model lidského ucha a následně tento psychoakustický model implementujte v prostředí Matlab.

Zvolte vhodnou metodu vodoznačení využívající vypočtený maskovací práh. Otestujte robustnost vytvořeného algoritmu, a srovnejte úroveň transparentnosti vodoznaku s obvyklými metodami.

DOPORUČENÁ LITERATURA:

[1] Carnero, B.; Drygajlo, A.: Perceptual Speech Coding Using Time and Frequency Masking Constraints. In Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)-Volume 2, 1997. ISBN 0-8186-7919-0
[2] Cvejic,N.; Seppanen, T.: Digital Audio Watermarking Techniques and Technologies (Applications and benchmarks). IGI Global, 2007. ISBN 978-159904513-9.
[3] Xing, H.; Watermarking in Audio: Key Techniques and Technologies. Cambria Press, 2008. ISBN 1604975016.

Termín zadání:	29.1.2010	***Termín odevzdání:***	26.5.2010

Vedoucí práce: Mgr. Pavel Rajmic, Ph.D.

prof. Ing. Kamil Vrba, CSc.
Předseda oborové rady

Abstrakt

Tato diplomová práce pojednává o metodě prosazující dodržování vlastnických práv a ochranu multimediálních dat proti nelegální manipulaci s jeho obsahem – Digitálním vodoznačením audio signálů. Hlavním cílem této práce je implementovat algoritmus pro digitální vodoznačení audio dat. V teoretické části jsou popsány základní pojmy, metody a postupy, které se vztahují k této oblasti digitálního zpracování dat.

V praktické části je realizován samotný proces vkládání tajné informace do originálního audia a následná možnost jejího zpětného vyjmutí. Algoritmus vodoznačení využívá metodu rozprostřeného spektra a psychoakustický model. Implementovaný psychoakustický model zahrnuje nedokonalosti lidského ucha, konkrétně jde o frekvenční maskování a dělení frekvenčního intervalu na kritická pásma. Tento model je založený na transformaci DWPT. Pomocí něho je vodoznak vkládán ke koeficientům vlnkové transformace ve vlnkové oblasti. Algoritmus vkládání a extrakce vodoznaku je implementován v programovém prostředí MATLAB. Část práce se zabývá testem robustnosti vloženého vodoznaku. Jsou použity běžné metody zpracování audio signálů: oříznutí audia, změna vzorkovacího kmitočtu, ztrátová komprese, filtrace, ekvalizace, vložení hudebního efektu a bílého šumu. V závěru diplomové práce jsou použity objektivní a subjektivní metody stanovení úrovně transparentnosti vloženého vodoznaku.

Klíčová slova: Digitální vodoznačení audia, DWPT, Psychoakustický model, metoda rozprostřeného spektra, MATLAB

Abstract

This Thesis deals with a method to enforce the intellectual property rights and protect digital media from tampering – Digital Audio Watermarking. The main aim of this work is implement an audio watermarking algorithm. The theoretical part defined basic terms, methods and processes, which are used in this area.

The practical part shows a process of embedding the digital signature into a host signal and her backward extraction. The embedding rule used spread spectrum technique and a psychoacoustic model. The implemented psychoacoustic model involves two properties of the human auditory system which are frequency masking and representation the frequency scale on limited bands called critical bands. The model is relatively new and based on the DWPT. In terms of above model is then the digital watermark embedded in the wavelet domain. This algorithm is implemented in technical software MATLAB. One part of this work focuses on robustness tests of the algorithm. Common signal processing modifications are applied to the watermarked audio as follows: Cutting of the audio, re-sampling, lossy compression, filtering, equalization, modulation effects, noise addition. The last part of the thesis presents subjective and objective methods usable in order to judge the influence of watermarking embedding on the quality of audio tracks called transparency.

Keywords: Digital audio watermarking, DWPT, Psychoacoustics model, Spread spectrum, MATLAB

HEITEL, T. *Využití psychoakustického modelu a tranformace typu wavelet packet pro vodoznačení audio signálů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2010. 85 s, 3 přílohy. Vedoucí diplomové práce Mgr. Pavel Rajmic, Ph.D.

Prohlášení:

Prohlašuji, že svou diplomovou práci na téma „Využití psychoakustického modelu a transformace typu wavelet packet pro vodoznačení audio signálů“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením tohoto projektu jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne …… ……

(podpis autora)

Poděkování:

Děkuji vedoucímu diplomové práce Mgr. Pavlu Rajmicovi, Ph.D. a bývalému odbornému asistentovi Ústavu telekomunikací Ing. Radku Zezulovi, Ph.D. za obětavý přístup a cenné rady při zpracování diplomové práce. Taktéž bych chtěl poděkovat za poskytnutí technické literatury, která mi pomohla při tvorbě této práce.

Dále bych chtěl poděkovat odborným asistentům a interním doktorandům Ústavu telekomunikací, kteří se podíleli na poslechových testech v rámci této diplomové práce. Jmenovitě: Ing. Jiří Schimmel, Ph.D., Ing. Petr Sysel, Ph.D., Ing. Ivan Míča, Mgr. Pavel Rajmic, Ph.D., Ing. Zdeněk Průša, Ing. Radek Beneš, Ing. Pavel Šilhavý, Ph.D., Ing. Radim Číž, Ph.D., Ing. Jan Karásek, Ing. Jan Šporik.

V Brně dne …… ……

(podpis autora)

Obsah

Seznam obrázků

Seznam tabulek

Úvod

Všestranný, jednoduše použitelný software a snižování cen digitálních zařízení (digitální kamery, fotoaparáty, CD a mp3 přehrávače, notebooky, PDA, …) umožnily uživatelům si libovolně vytvářet, upravovat a kopírovat multimediální data. Velkou výhodou je, že vícenásobným kopírováním neztrácejí digitální soubory na kvalitě, na rozdíl od analogové kazety, nebo VHS pásky. Vysokorychlostní internet a téměř bezchybný přenos dat umožňuje lidem přenášet si přes tuto síť nelegální kopie multimediálních souborů o velké velikosti. Tímto vznikají majitelům autorských práv nemalé finanční ztráty.

Tradiční metody ochrany autorských práv multimediálních dat začínají být nevyhovující. Jednoduché zabezpečovací mechanizmy jako například vložení informace do hlavičky před digitální data jsou již nepoužitelné. Hlavičku můžeme jednoduše odstranit změnou formátu dat, který ale nemá vliv na samotná data. Další metoda ochrany, např. šifrování multimediálních dat zabraňuje přístupu k obsahu jednotlivcům bez správného dešifrovacího klíče. Nicméně, každý uživatel, který zaplatí poskytovateli licenční poplatek, aby získal dešifrovací klíč, může tyto data používat. Jakmile byla ale data dešifrována, můžeme je dále nelegálně kopírovat. Tyto nedostatky v ochraně autorských práv multimediálních dat v digitální podobě (obrázky, video a audio) byly důvodem pro vytvoření nových metod, jako například digitální vodoznačení.

S digitálním vodoznakem se můžeme setkat denně v běžném životě. Například ochranný vodoznak na papírových bankovkách, nebo viditelná loga firem vložená do obrázku. Digitální vodoznačení je velmi podobné steganografii, kde se snažíme o to, aby nikdo nevěděl, že tajná informace existuje. V případě, že zjistíme přítomnost tajné informace, je velmi jednoduché ji dekódovat. Pokud vezmeme šifrování, jako další možnost ochrany dat, tak na rozdíl od ní vodoznak chrání obsah dat i po jeho dekódování. Původně byly algoritmy vodoznačení použity pro vodoznačení obrázků a videa. Až následně se využívalo vodoznačení i pro audio data. V této diplomové práci se budu zabývat pouze **digitálním vodoznačení audio dat**.

Cílem této práce je popsat transformaci známou jako Discrete Wavelet Packet Transform (DWPT) a uvést způsob jejího využití pro psychoakustický model lidského ucha. Implementovat algoritmus pro vodoznačení audio dat, který využívá tento psychoakustický model. Vložený vodoznak otestovat na robustnost a stanovit úroveň transparentnosti vloženého vodoznaku.

V první kapitole se snažím stručně a srozumitelně popsat základní vlastnosti sluchového ústrojí člověka. Existuje mnoho monografií a odborných článků, zabývajících se podrobně sluchovému ústrojí jako celek. V této práci popisuji pouze tzv. psychoakustické jevy. Širší náhled na tuto problematiku najdeme např. v monografii [1].

Druhá kapitola je zaměřena na používané metody vodoznačení. Dále je zde uvedena souvislost s kryptografickým systémem. Neméně důležitou částí jsou požadavky, které jsou kladeny na algoritmy vodoznačení. V závěru této kapitoly jsou uvedeny příklady použití digitálního vodoznačení audio dat v praxi.

Třetí kapitola tvoří úvod do vlnkové transformace, která se těší značné pozornosti. Základní přehled o vlnkové transformaci můžeme získat z literatury [2][13]. V této práci se zaměřuji hlavně na popis transformace DTWT a DWPT, kterou využívám v psychoakustickém modelu.

Úkolem čtvrté kapitoly je ukázat ucelený pohled na proces vodoznačení a operace se signály s tím spojené ve formě blokového schématu kodéru a dekodéru vodoznaku.

Pátá kapitola se věnuje problematice návrhu algoritmu pro vodoznačení audia, tedy praktickou částí diplomové práce. Je vybrána určitá metoda vodoznačení a ta je implementována v prostředí Matlab. Součástí je i extrakce vodoznaku.

Šestá kapitola se zabývá testem robustnosti vloženého vodoznaku, tedy skupinou algoritmů, které slouží k simulaci možných útoků vedených k poškození, nebo odstranění vodoznaku z vodoznačných audio dat. Pro test robustnosti vloženého vodoznaku byly vybrány tyto úpravy se signály: oříznutí, změna vzorkovacího kmitočtu, ztrátová komprese, filtrace, ekvalizace, hudební efekty a AWGN kanál.

Sedmá kapitola se zabývá stanovením úrovně transparentnosti vloženého vodoznaku pomocí objektivního a subjektivního testu. Dále je zde vybrán i volně stažitelný software pro srovnání úrovně transparentnosti s implementovaným algoritmem.

V osmé kapitole je popsán návod k ovládání programu, jedná se tedy o tzv. uživatelskou příručku. Kapitola je rozdělena na část vložení vodoznaku a extrakci vodoznaku.

V závěru jsou shrnuty dosažené výsledky vzhledem k zadání diplomové práce.

1 Psychoakustika a psychoakustické jevy

Psychoakustika je součást psychofyziky a zabývá se souvislostí mezi fyzikálními veličinami popisující akustický signál a následným vnímáním zvuku člověkem. Nedokonalosti ve vnímání člověka pomocí sluchového ústrojí popisují psychoakustické jevy. Psychoakustické jevy byly mnohdy zjištěny experimentálně a patří mezi ně například maskovaní, nebo rozdělení frekvenčního rozsahu na kritická pásma, které jsou popsány níže a jsou důležité pro vodoznačení audio signálů.

1.1 Oblast slyšení lidského ucha

Zvuk, který vnímáme sluchovým ústrojím, je pohyb hmotného prostředí, nejčastěji vzduchu, který je lidský sluch schopen v určitých mezích vnímat. Fyzikální veličiny charakterizující zvukové vlnění vyvolávají ve sluchovém orgánu určité fyziologické pochody, které se v lidském vědomí projevují jako zvukový vjem s jeho charakteristickými veličinami (hlasitost, výška, …) [12]. Oblast slyšení je plocha ohraničená co do dynamického rozsahu, tak frekvenčního rozsahu a vidíme ji na obr. 1.1.

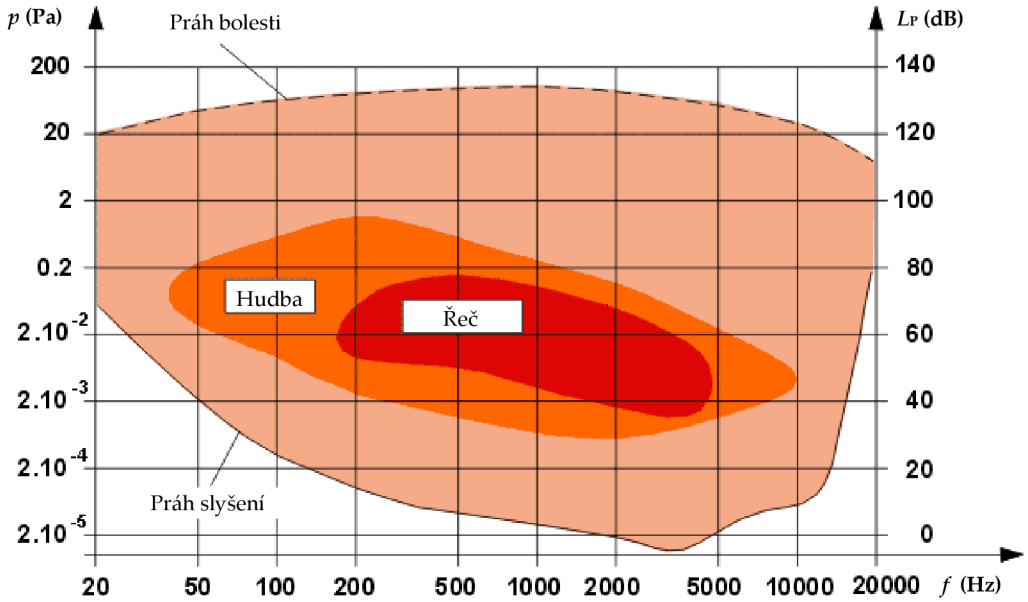
Dynamický rozsah slyšení je obrovský a sahá od hladiny akustického tlaku 0 dB až po 140 dB. Je to relativní hodnota, která je vztažena k referenční hodnotě p_0 a je dána vztahem [1]:

$$L_P = 20 \cdot \log \frac{p}{p_0} \ , \tag{2.1}$$

kde: $p_0 = 5.10^{-5}$ Pa; je to nejmenší možná změna akustického tlaku, kterou je lidské ucho schopno zaznamenat.

Míra subjektivního vjemu související s hladinou akustického tlaku je hlasitost. Lidský sluch má různou citlivost pro zvuky různých frekvencí, proto byla stanovena jednotka pro hladinu hlasitosti **Phon** (Ph). Při frekvenci 1000 Hz je hodnota ve Phonech rovna právě hladině akustického tlaku v dB. Pro ostatní frekvence je hladina hlasitosti definována subjektivním porovnáním s hladinou hlasitostí referenčního tónu. Takto vnikly křivky stejné hlasitosti (isofony). Hladina hlasitosti ve Phonech však nevyjadřuje správně subjektivně vnímané změny hlasitosti. Proto byla definována subjektivní míra zvukového vjemu a byla zavedena jednotka **son**. 1 son je hlasitost, kterou vnímá posluchač, naslouchá-li referenčnímu tónu o frekvenci 1000 Hz při hladině akustického tlaku 40 dB. Zvětšením této hladiny na 2 sony se jeví zvukový vjem dvakrát silnější. Subjektivní hlasitost je veličinou aditivní a lze tak sčítat současně působící zvuky, nevzniká-li jejich vzájemné maskování. [12]

Křivka nulové hlasitosti zvuku, tedy 0 Ph, udává sluchový práh, tzv. **práh slyšení**. Cokoliv pod touto hranicí je pro člověka neslyšitelné. V okolí kmitočtů 3 kHz až 4 kHz má sluch citlivost největší. Velká hladina akustického tlaku podráždí nejen smyslové buňky vnitřního ucha, ale i hmatová tělíska ve sluchovém orgánu [12]. **Hmatový práh** lze stanovit i u osob úplně hluchých. Tento práh leží při hladině hlasitosti 120 Ph. Zvyšujeme-li dále hladinu hlasitosti, vznikne při hladinách hlasitostí 130 až 140 Ph ve sluchovém orgánu pocit bolesti. Dosahuje se **prahu bolesti**.



Obr. 1.1: Oblast slyšení lidského ucha [12]

Lidské ucho je schopno vnímat zvuk v kmitočtovém rozsahu od 20 Hz do 20 kHz. Tomu odpovídají vlnové délky od 17,15 m do 17,15 mm. Tento rozsah je pouze teoretický a je u každého člověka individuální. Mění se například i s věkem. Kmitočty pod 20 Hz nazýváme infrazvukovými, nad 20 kHz ultrazvukovými. Míra subjektivního vjemu související s kmitočtem je výška. Můžeme určit absolutní, relativní, nebo subjektivní výšku. Subjektivní výška byla stanovena experimentálně, tím že ke každému tónu byl stanoven tón, který se zdá dvojnásobný, poloviční, atd. Tak vznikla lineární stupnice s jednotkou **mel**. Stupnice je od 0 do 2400 melů. Tato stupnice odráží fakt, že nižší kmitočty vnímáme subjektivně jako vyšší a naopak.

Kmitočtovou osu můžeme z praktického hlediska dělit lineárně, logaritmicky, na kritická pásma, atd. Rozlišovací schopnost ucha na kmitočtové ose je přibližně logaritmická. Tuto vlastnost právě odráží kritická pásma.

1.2 Maskování akustických signálů

Maskování je jev, během kterého dochází k tomu, že když souzní několik signálů a všechny jsou nad prahem slyšení, tak není slyšet některý z nich. Maskování zvuků je dáno způsobem činnosti vnitřního ucha a je závislé na rozdílu kmitočtů přijímaných zvuků a jejich intenzitě [12]. Důsledkem toho je zvednutí prahu slyšení.

Mezi základní pojmy maskování patří:

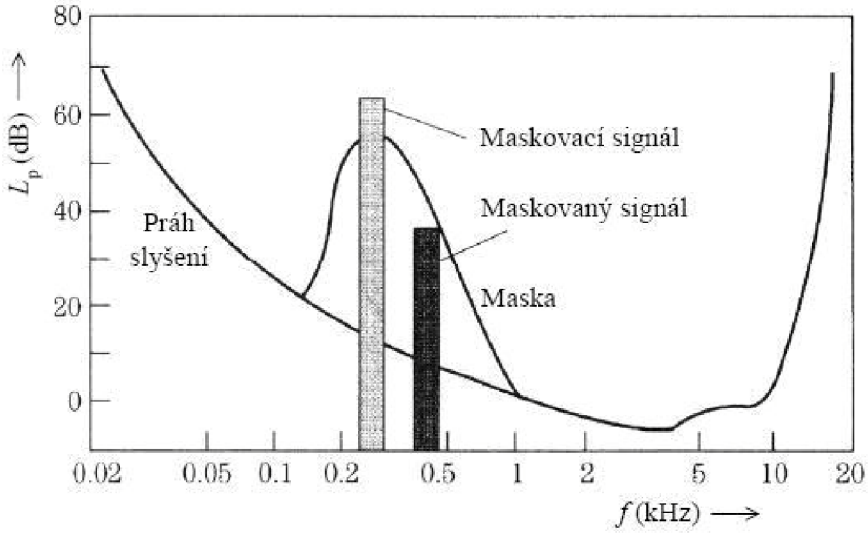
- **Maskovací signál**: signál, který maskuje ostatní signály,
- **Maskovaný signál**: signál, který je zamaskován a není slyšet,
- **Maska**: obálka maskovacího signálu, pod kterou jsou všechny signály zamaskovány.

Maskovaný signál má také svoji masku, která může zasáhnout do procesu maskování. Nedochází tedy ke sčítání masek. Masky se neustále mění v průběhu reprodukce zvuku. Maskovací signál může být čistý tón, bílí šum, komplexní tón. V závislosti na tom se mění tvar masky.

1.2.1 Maskování ve frekvenční oblasti

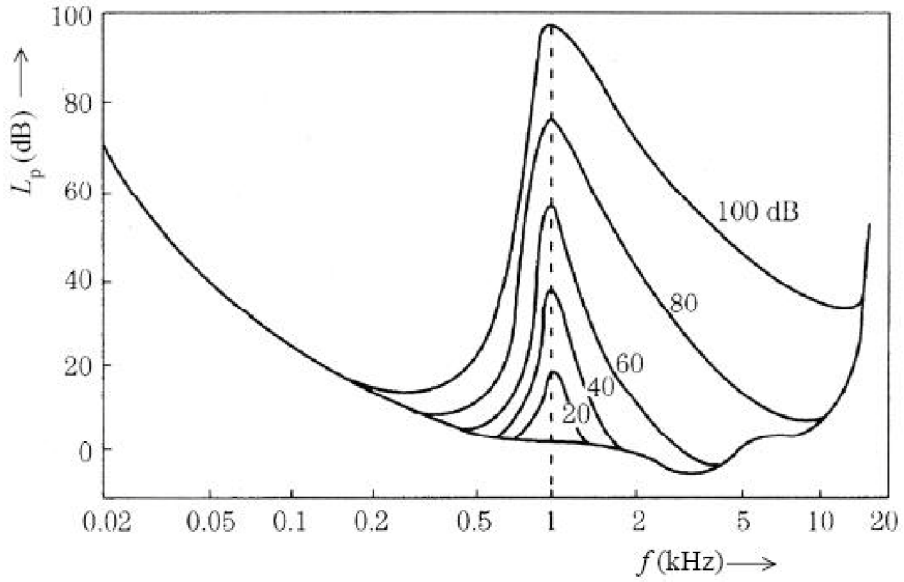
Při reprodukci dvou kmitočtově blízkých signálů, nebudeme slyšet signál s nižším akustickým tlakem. Maskovací účinek čistého tónu a pásmem šumu při téže hlasitosti se poněkud liší. Čisté tóny při téměř shodném kmitočtu tónu maskujícího a maskovaného tvoří zázněje. Bílý šum zasahuje širokou část sluchového analyzátoru, a proto maskuje v celém rozsahu slyšitelných kmitočtů, nejvíce však v oblasti 2000 až 3000 Hz. Maskuje-li bílý šum čistý tón, ovlivňuje maskování pouze určité pásmo šumu v okolí kmitočtu maskovaného tónu [1].

Na obrázku 1.2 vidíme maskovací práh, pokud byl čistý tón maskován úzkopásmovým šumem na kmitočtu 300 Hz.



Obr. 1.2: Princip maskování úzkopásmovým šumem ve frekvenční oblasti [11]

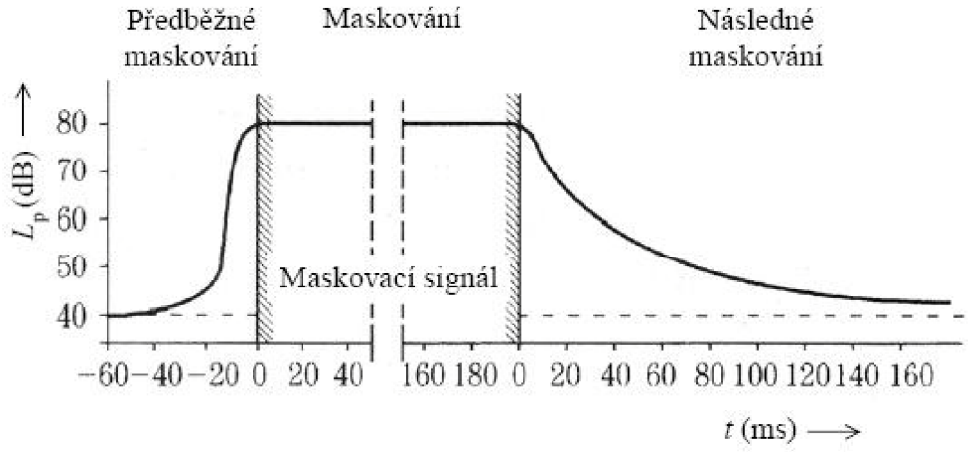
Maskovací účinek bílého šumu je největší okolo jeho středního kmitočtu. Při slabých tónech (20 až 40 dB nad prahem) je maskovací účinek rozložen souměrně na obě strany. Při zvedání hladiny jsou tóny vyšší než maskující maskovány více. Masky již nejsou souměrné vůči kmitočtu maskovaného tónu, viz obr. 1.3.



Obr. 1.3: Závislost tvaru masky na hladině maskovacího signálu [11]

1.2.2 Maskování v časové oblasti

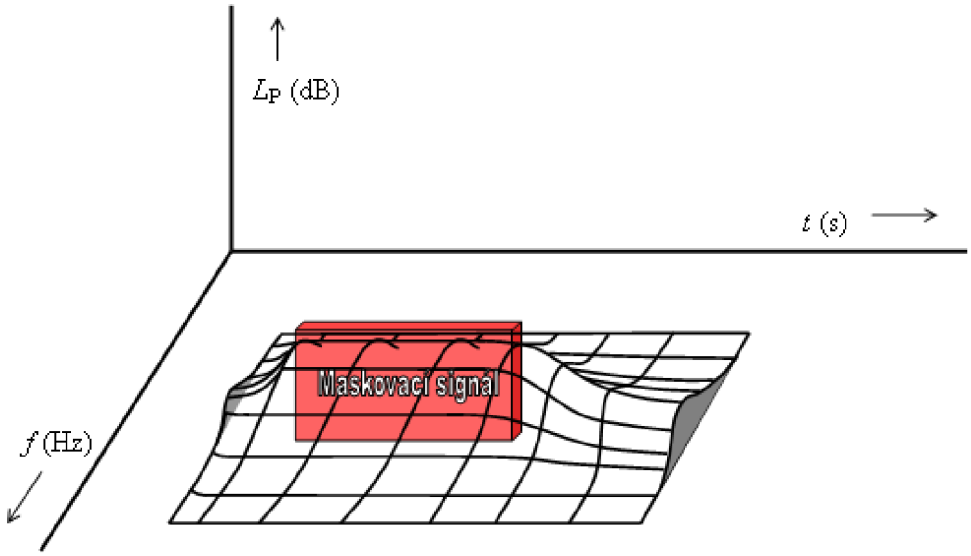
Maskovací jev však nastává i v případě, kdy maskovaný krátkodobý signál určité hladiny přichází až po ukončení maskujícího signálu vyšší hladiny, v době do 30 ms. Při delším intervalu než 30 ms maskování slábne a při intervalu 150 ms již zcela zaniká. Této oblasti se říká **následné maskování**. Maskován může být rovněž krátký zvukový impuls, následuje-li po něm nejdéle do 10 ms maskující signál. Této oblasti se říká **předběžné maskování**.



Obr. 1.4: Princip maskování v časové oblasti [11]

Následné maskování závisí hodně na délce maskovacího signálu. Mějme testovací tón délky 200 ms a 5 ms. U maskovacího signálu délky 5 ms je mnohem strmější pokles účinku následného maskovaní, než u tónu délky 200 ms. U předběžného maskování nelze jednoznačně rozhodnout, zda závisí na délce maskovacího signálu. Předběžné maskování trvá do 20 ms v každé situaci, proto nemusíme tento efekt uvažovat v praxi [1].

V praxi samozřejmě maskování v časové a frekvenční oblasti působí zároveň. Jak může vypadat maska, pokud maskovacím signálem bude úzkopásmový šum, můžeme vidět na obrázku 1.5. Všechny signály, které se nachází pod maskou, neslyšíme.



Obr. 1.5: Příspěvek masek v časové a frekvenční oblasti – 2D maska

1.3 Kritická pásma

Lidské ucho funguje jako krátkodobý frekvenční analyzátor mapující jednotlivé frekvence do tzv. kritických pásem, které korespondují určitému fyzickému umístění podél bazilární membrány. Bazilární membrána je součástí vnitřního ucha. Pokud tedy začne kmitat, vlivem dopadu zvuku, jsou indikovány největší změny výchylky, které potom stimulují příslušné vlasové buňky. Tyto vedou tento signál dále do lidského mozku [12]. Podle knihy [1] můžeme rozdělit kmitočtový rozsah na 25 frekvenčních intervalů, tedy kritických pásem. Tyto pásma byla stanovena experimentálně pomocí maskovacího efektu. Pokud je čistý tón maskován bílým šumem příslušné frekvence a tento tón je zamaskován, neslyšíme ho, potom se tato šířka pásma bílého šumu nazývá kritické pásmo [1]. Do Hodnoty 500 Hz je šířka pásma přibližně konstantní a pohybuje se kolem 100 Hz, nad 500 Hz je kolem 20 % aktuální frekvence.

Na základě konceptu kritických pásem byla vytvořena Barkova stupnice. Hodnota jednoho Barku představuje šířku právě jednoho kritického pásma nad celým

frekvenčním pásmem. Barkova stupnice odráží fakt, že nízké frekvence lidské sluchové ústrojí dokáže lépe rozlišovat než vysoké frekvence.

Tab. 1.1: Rozdělení kmitočtového rozsahu na kritická pásma [1]

Kritické pásmo	Dolní frekvence (Hz)	Šířka pásma (Hz)	Šířka pásma (%)	Bark
0.	0	100	200	0
1.	100	100	67	1
2.	200	100	40	2
3.	300	100	29	3
4.	400	110	24	4
5.	510	120	21	5
6.	630	140	20	6
7.	770	150	18	7
8.	920	160	16	8
9.	1080	190	16	9
10.	1270	210	15	10
11.	1480	240	15	11
12.	1720	280	15	12
13.	2000	320	15	13
14.	2320	380	15	14
15.	2700	450	16	15
16.	3150	550	16	16
17.	3700	700	18	17
18.	4400	900	19	18
19.	5300	1100	19	19
20.	6400	1300	19	20
21.	7700	1800	21	21
22.	9500	2500	24	22
23.	12000	3500	26	23
24.	15500			24

Pro převod mezi frekvencí v Hz a Barkovou stupnici můžeme použít následující vzorec [1]:

$$Bark = 13 \cdot \arctan(0{,}00076 \cdot f) + 3{,}5 \cdot \arctan\left(\frac{f}{7500}\right)^2, \tag{1.1}$$

kde: f je frekvence.

Šířku pásma, která přísluší jednotlivým kritickým pásmům, můžeme vypočítat pomocí vzorce [1]:

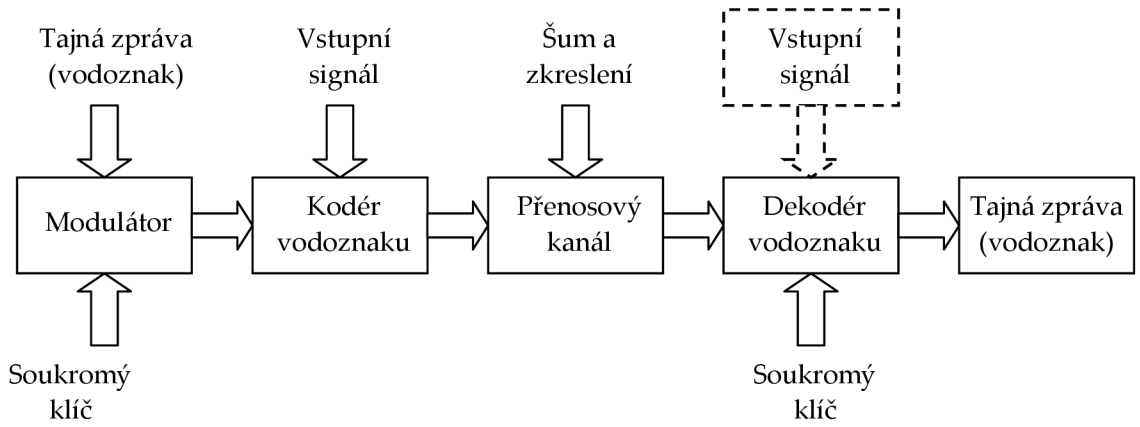
$$SP = 25 + 75 \cdot \left[1 + 1{,}4 \cdot \left(\frac{f}{1000}\right)^2\right]^{0{,}69}. \tag{1.2}$$

2 Digitální vodoznačení audio signálů

Digitální vodoznačení audio signálů je poměrně nová metoda prosazující dodržování vlastnickým práv a ochranu digitálních dat proti nelegální manipulaci s jeho obsahem. Cílem vodoznačení je vložit tajnou zprávu (vodoznak) do vstupního audio signálu s určitými požadavky (neslyšitelnost, robustnost, …). Vodoznak nesmíme tedy vložit jen např. do hlavičky datového toku, do samostatného souboru, nebo poslat samostatným bitovým tokem. Další podmínka je, abychom byli schopni na přijímací straně jednoznačně extrahovat vodoznak z vodoznačených dat. Některé metody vodoznačení využívají stejných jevů, jakých se využívá např. při ztrátové kompresi zvukových signálů. Jedná se právě o psychoakustické jevy. Snažíme se přiblížit co možná nejpřesněji sluchovým vlastnostem lidského ucha. Jeden ze standardů je právě MPEG-1 Audio Layer 3, který je více známý jako mp3 formát. Pro analýzu těchto signálů (diskrétních, spojitých) používáme různé transformace, podle vlastností multimediálních dat (stacionární, nestacionární, …). Může to být například Fourierova transformace (FT), krátkodobá Fourierova transformace (STFT), diskrétní kosinová transformace (DCT), vlnková transformace (WT), atd. Jaký bude vodoznak, jestli robustní, nevnímatelný, nebo bezpečný záleží na konkrétním použití.

2.1 Princip vodoznačení

Na vodoznačení se můžeme dívat, jako na sdělovací soustavu podobnou kryptografickému systému sestávají z tří hlavních bloků: vysílač tvořený kodérem vodoznaku, přenosovým kanálem a přijímačem tvořeným dekodérem vodoznaku. Jejich blokové schéma můžeme vidět na obrázku 2.1.



Obr. 2.1: Princip vodoznačení [3]

Na vstupu kodéru vodoznaku máme tajnou zprávu z rozsahu {0,1} a vstupní signál, na který se můžeme dívat jako na nosný signál, a který použijeme k přenosu vodoznaku. Jako další vstup je použit soukromý klíč, který se využívá na straně vysílače i přijímače. Jeho hlavním úkolem je neumožnit na straně přijímače neoprávněné osobě bez tajného klíče extrahovat vodoznak. Všechny prakticky použitelné systémy by měli používat alespoň jeden klíč, přičemž není vyloučeno ani použití kombinace několika klíčů. Výstupem z kodéru jsou chráněná data obsahující vodoznak, která prochází přenosovým kanálem k přijímači. Vodoznačený signál podléhá běžným metodám zpracování audio signálů (komprese, změna vzorkovacího kmitočtu, atd.). Mimo to dochází na přenosovém kanálu k rušení šumem, zkreslení, atd. Na přijímací straně pomocí soukromého klíče a vstupního signálu, pokud se jedná o privátní systém, nebo bez vstupního signálu, pokud se jedná o veřejný systém, získáme vodoznak.

Mějme m jako tajnou zprávu, k je soukromý klíč, x je vstupní signál, f je modulační funkce a w je vodoznak, tak potom můžeme psát:

$$w = f(m,k), \tag{3.1}$$

pokud je vodoznak nezávislý na vstupním signále, nebo

$$w = f(m,k,x), \tag{3.2}$$

kde jsou data ze vstupního signálu x připojená k vodoznaku.

Vodoznačený signál je popsán jako:

$$y = x + \alpha \cdot w, \tag{3.3}$$

kde α je činitel zaručující nevnímatelné vložení vodoznaku.

2.2 Metoda rozprostřeného spektra

Metoda rozprostřeného spektra (Spread Spectrum) se původně používala v rádiovém spojení k přenosu více nezávislých informačních signálů jedním kanálem. V dnešní době se metoda rozprostřeného spektra využívá v řadě aplikací: bezdrátové sítě dle standardu 802.11, radiokomunikační systém bluetooth, navigační systém GPS.

Podle způsobu rozšíření spektra rozlišujeme systémy:

- S přímým rozprostřením spektra (DSSS)
- S frekvenčním skákáním (FHSS)

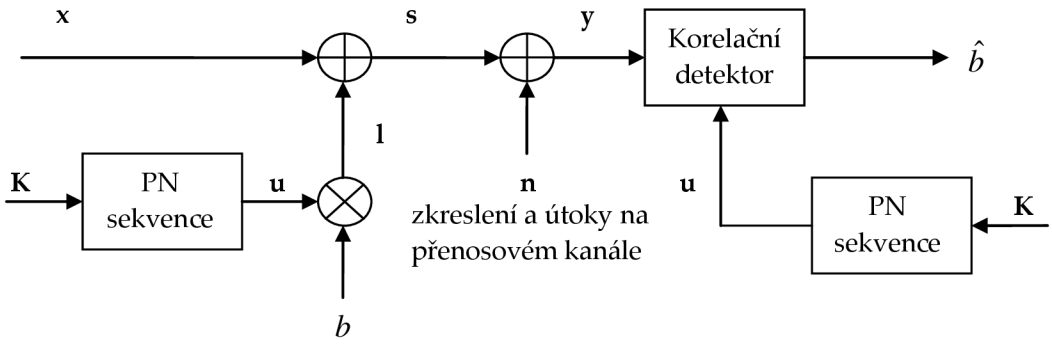
Největší výhodou metody rozprostřeného spektra je jejich odolnost vůči širokopásmovému a úzkopásmovému rušení. Pro tyto výhody se začala metoda

rozprostřeného spektra využívat i v digitálním vodoznačení, kde je vodoznak rozprostřen po celém spektru signálu, tímto také ztížíme útočníkům lokalizovat vodoznak.

V systému s využitím techniky DSSS je relativně úzkopásmový datový signál násoben pseudonáhodnou posloupností (dále jen PN posloupnost). Bitová rychlost pseudonáhodné posloupnosti se označuje jako *čipová rychlost* f_{ch}, která má o několik řádu vyšší bitovou rychlost, než původní signál s bitovou rychlostí f_b. Z tohoto důvodu je šířka výsledného signálu o mnohem větší než šířka pásma původního signálu. Můžeme stanovit činitel rozprostření definován jako [15]:

$$SF = \frac{f_{ch}}{f_b}. \tag{3.4}$$

Tento činitel udává přibližně poměr šířky pásma původního signálu před a po jeho vynásobením PN posloupností. PN posloupnost musí splňovat určité požadavky. Jejich autokorelační funkce musí mít výrazné maximum, umožňující rychlou synchronizace na straně přijímače, dále musí vykazovat minimální vzájemnou korelaci. Jako generátor pseudonáhodné posloupnosti lze použít posuvné registry se zpětnou vazbou a pamětí, nebo periodické signály se statistickými vlastnostmi, blízké bílému šumu [15]. Blokové schéma obecného vodoznačícího systému založeného na metodě přímého rozprostření spektra je zobrazeno na obr. 5.6.



Obr. 2.2: obecné schéma metody vodoznačení založené na rozprostřeném spektru [8]

Vektor **x** označuje originální vstupní signál v příslušné oblasti dle použité transformace. Může se tedy jednat o vzorky originálního audio signálu v časové oblasti, koeficienty získané pomocí FFT, kepstrální koeficienty, koeficienty vlnkové transformace, atd. Na přijímací straně získáme vektor **y**, který byl zkreslen na přenosovém kanálu. Ke generování pseudonáhodné posloupnosti je použit tajný klíč **K**. Na vysílací i přijímací straně musí být použit stejný klíč K. Takto vznikne sekvence **u** se střední hodnotou 0 a prvky $+\sigma_u$ nebo $-\sigma_u$. Bity vektoru b (vodoznak), nabývající hodnot +1 nebo –1, jsou rozprostřeny pomocí sekvence u. Rozprostřená sekvence l je potom přičtena, nebo odečtena od vektoru x. Signál **s** je vodoznačený audio signál.

K odhadu vloženého vektoru b na výstupu musíme definovat skalární součin a absolutní hodnotu [8]:

$$(x,u)=\sum_{i=0}^{N-1} x_i u_i \tag{3.5}$$

a

$$|x|=\sqrt{x\cdot x}\,, \tag{3.6}$$

kde N je délka vektorů **x**, **s**, **u**, **n** a **y**.

Pro jednoduchost předpokládejme, že vkládáme tajnou zprávu b, nabývající hodnot +1 nebo –1 a velikosti 1 bit.

$$s=x+b\cdot u\,. \tag{3.7}$$

Zkreslení D je na přijímací straně definováno jako $|s-x|$. Porovnáním s rovnicí 3.7 můžeme odvodit:

$$D=|b\cdot u|=|u|=\sigma_u\,. \tag{3.8}$$

Přenosový kanál můžeme definovat jako součet užitečného signálu a aditivního šumu n takto: $y=s+n$. Potom je extrakce vodoznaku dána výpočtem normalizované statistické hodnoty r:

$$r=\frac{(y,u)}{(u,u)}=\frac{(b\cdot u+x+n,u)}{\sigma_u^2}=b+x+n\,, \tag{3.9}$$

kde

$$x=\frac{(x,u)}{|u|}\ \text{a}\ n=\frac{(n,u)}{|u|}. \tag{3.10}$$

Odhad vloženého bitu b je definován jako:

$$\hat{b}=\mathrm{sgn}(r)\,, \tag{3.11}$$

kde sgn je funkce signum definována následujícím způsobem:

$$\mathrm{sgn}(r)=\begin{cases}+1 & pro\ r\geq 0\\ -1 & pro\ r<0\end{cases}. \tag{3.12}$$

Existují další metody, které se používají pro vodoznačení audio signálů a které jsou ve stručnosti popsány zde [9]:

- **Kódování LSB bitu**

Je to jedna z prvních metod využívaná v digitálním vodoznačení. Je založena na záměně LSB bitu vstupního signálu za bit vodoznaku. Nevyužívá žádný psychoakustický model za účelem nevnímatelného vložení vodoznaku. Tato metoda vyžaduje přesnou synchronizaci vodoznačného signálu na straně přijímače.

- **Kódování fáze**

Metoda založená na vkládání vodoznaku do fáze vstupního signálu. Nevyužívá efektů časového a frekvenčního maskování, ale využívá toho, že sluchové ústrojí člověka má malou citlivost na relativní změnu fáze.

- **Kódování echa**

U této metody se vložením vodoznaku do vstupního signálu přidá echo, a tak vznikne vodoznačený signál. Můžeme měnit dva parametry, abychom zaručili nevnímatelnost vodoznaku. Jedná se o změnu *doby zpoždění* a *činitel útlumu*, kterým ovlivňujeme význam echa.

2.3 Požadavky na vodoznak

Algoritmy pro vodoznačení můžou být popsány řadou vlastností. Podíl jednotlivých vlastností závisí na konkrétním použití. Nyní popíši několik požadavků, které mohou být kladeny na algoritmus vodoznačení [5].

- **Nevnímatelnost**

V mnoha aplikacích je požadováno, aby algoritmus, který vloží dodatečná data do vstupního audio signálu, neměl žádný vliv na jeho kvalitu. Ta je definována jako vjemový rozdíl originálního a vodoznačného audio signálu. Samotným přenosem audia po přenosové cestě k uživateli dojde ke zhoršení jeho kvality. Nevnímatelnost je lepší tedy definovat a vyhodnotit na straně vysílače.

- **Přenosová rychlost vodoznaku**

Přenosová rychlost vloženého vodoznaku je počet vložených bitů za jednotku času udávaných v b/s. Některé aplikace pro vodoznačení, jako například ochrana proti kopírování, které vkládají identifikační číslo autora, nebo sériové číslo, mají průměrnou přenosovou rychlostí 0,5 b/s. V rozhlasovém vysílání přenášíme identifikační údaje o zvukových nahrávkách během každých prvních sekund nahrávky a tak se dostáváme na přenosovou rychlost vodoznaku do 15 b/s. V některých aplikacích vkládáme vodoznak přenosovou rychlostí až 150 kb/s.

- **Robustnost**

Robustnost algoritmu je definována jako schopnost detekovat vložený vodoznak na přijímací straně. Aplikace většinou vyžadují vysokou robustnost, abychom mohli na přijímací straně detekovat jednoznačně vodoznak podléhající běžným

metodám zpracování signálů. Naopak v některých aplikacích je robustnost nežádoucí a takovéto algoritmy označujeme jako křehké.

- **Detekce vodoznaku**

V mnoho aplikacích využívá algoritmus pro detekci vodoznaku původní audio data (tzv. privátní systémy). Tímto zlepšíme přesnost demodulátoru v tom, že původní audio data můžeme odečíst od vodoznačného audia. V opačném případě demodulátor nepoužívá původní audio data k extrakci vodoznaku, čímž podstatně snížíme velikost dat, které můžeme utajit (tzv. veřejné systémy).

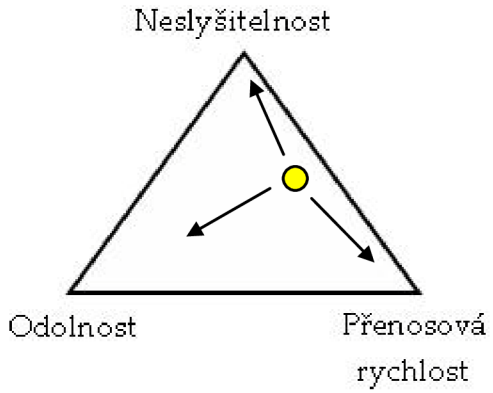
- **Bezpečnost**

Algoritmus pro vodoznačení musí být bezpečný v tom smyslu, že útočník nesmí vědět o vloženém vodoznaku, natož aby byl schopný ho odstranit. Bezpečnost vodoznačení je založena na stejném principu jako šifrování. Vodoznak nemůže nikdo vyjmout z audio dat bez soukromého klíče, který používáme i při vkládání. Neoprávněná osoba by neměla být schopna extrahovat vodoznak za určitý čas, i když by věděla, že audio data obsahují vodoznak a byla obeznámena s algoritmem pro jeho vkládání.

- **Výpočetní náročnost**

Základní problém z technického hlediska je výpočetní náročnost algoritmu pro vkládání a detekci vodoznaku a počet modulátorů a demodulátorů použitých v přenosové soustavě. U některých aplikací musí probíhat vše v reálném čase, jiné aplikace toto nevyžadují. Rozdíl ve výpočetní náročnosti bude, zda je vše implementováno hardwarově, nebo softwarově ve formě plug-in modulů.

Nejjednodušší zobrazení požadavků na vodoznačení se nazývá tzv. **magický trojúhelník**. Tento model zobrazuje závislosti mezi kapacitou, odolnosti proti útokům a nevnímatelnosti. Z tohoto trojúhelníku je patrné, že pokud jeden požadavek převládá, ostatní jsou na nižší úrovni. Pokud tedy požadujeme vysokou odolnost, přenosová rychlost vloženého vodoznaku je nízká atd. Například požadavek na vysokou odolnost může způsobit viditelné změny ve výsledném signále.



Obr. 2.3: Trojúhelník požadavků na digitální vodoznak

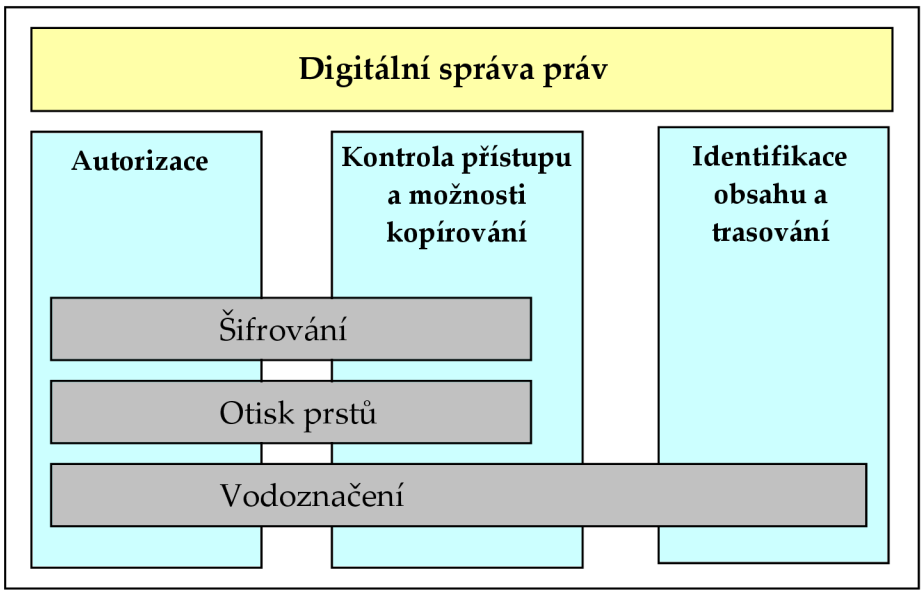
2.4 Oblast použití

Digitální vodoznačení audio signálů má v dnešní době poměrně mnoho oblasti užití. Je to především ochrana a kontrola dodržování autorských práv. Najdou se ale i jiné oblasti použití, které využívají toho, že vložená informace (vodoznak) je nevnímatelný.

2.4.1 DRM

Jedno z možných využití vodoznačení je v DRM. Tento akronym, vzniklý z anglického Digital Rights Management označuje souhrnně technická řešení umožňující kontrolu užití a rozmnožování autorskoprávně chráněného díla [3]. Český ekvivalent pro DRM je správa digitálních práv. DRM byla nutná reakce vydavatelů a majitelů autorských práv na nelegální šíření multimediálních dat jak přes internet, tak na nosičích CD, DVD. DRM jednak konkrétně stanoví, co má být s autorskoprávně chráněným dílem a jednak technicky vynucují, že tomu tak opravdu bude.

Technickými prostředky jsou vodoznačení, šifrování a otisk prstů, umožňující dílo identifikovat, kontrolovat přístup k němu, omezit jeho kopírování, nebo kontrolovat jeho užití. Jak je patrné z obrázku 2.4 pomocí metody vodoznačení jsme navíc schopni vypátrat původce pirátství. Uživatel si zakoupí audio data obsahující jednoznačné číslo identifikující právě tohoto uživatele. Pokud bude šířit obsah dat nelegálně dál a jedna z nelegálních kopií je naleznuta, tak pomocí identifikačního čísla originální jedinečné kopie jsme schopni zjistit vlastníka. Jedině tento uživatel mohl šířit nelegální data a je označen jako pirát. Tomuto se říká tzv. trasování.



Obr. 2.4: DRM technologie [3]

2.4.2 Jiné oblasti použití

S výše uvedeným příkladem se může setkat každý z nás, pokud si bude chtít např. doma přehrát zvukový soubor na příslušném přehrávači, který umí detekovat vodoznak. Pokud je vše v pořádku, můžeme si dopřát příjemný poslech hudby. Pokud jsou porušena pravidla, která stanovila DRM, nepovolí nám přehrávač zvukový soubor přehrát. V DRM můžeme vodoznačení využít více způsoby. Níže jsou obecněji uvedeny oblasti použití v DRM, ale i oblasti které nesouvisejí s DRM.

- **Ochrana autorských práv**

V ochraně autorských práv obsahuje vodoznak údaje o autorských právech, které jsou vloženy do vstupního signálu. Vodoznak, který zná pouze majitel autorského práva, musí být robustní a bezpečný proti útokům. Vlastníkovi umožní dokázat přítomnost vodoznaku v případě sporu o vlastnická práva. Při detekci vodoznaku musíme zaručit velmi nízkou pravděpodobnost chybného příjmu.

- **Prokázání vlastnictví**

Vodoznak se více používá jako prokázání skutečného vlastnictví. Tento problém nastane, pokud druhý uživatel použije např. software pro nahrazení originálního autorského práva a potom je prokazuje za svoje autorské práva. Každý kdo odhalí vodoznak, ho může nepochybně odstranit, nebo pozměnit. Pokud chceme tomuto zabránit, je nezbytné omezit přístup k demodulátoru vodoznaku. Namísto přímého prokázání vlastnictví vložením podpisu formou vodoznaku do vstupního signálu se používají algoritmy, které prokáží, že byla zvuková data odvozená z originálního zvukového souboru.

- **Zjištění a ověření neoprávněné manipulace s digitálním obsahem**

Vložením přídavných dat můžeme později stanovit, zda bylo s originálními daty manipulováno. Musíme zabránit falšování platného vodoznaku v neověřeném nebo zmanipulovaném vstupním signále. Detekci provádíme bez originálního vstupního signálu, protože audio signál není k dispozici. Tomuto druhu detekce vodoznaku se říká slepá detekce.

- **Identifikace**

Přídavná data vložená vodoznačením slouží k vyhledání původce nebo příjemce konkrétní kopie multimediálního souboru. Například, vodoznaky nesoucí rozdílné identifikační čísla jsou vloženy do různých kopií zvukových souborů na CD před tím, než jsou poskytnuty velkému množství zákazníků. Implementovaný algoritmus musí prokázat velkou odolnost proti úmyslnému útoku. Musíme tedy detekovat pouze jedno identifikační číslo, jinak není demodulátor schopný rozlišit, o kterou kopii se jedná.

- **Ochrana proti kopírování a kontrola přístupu**

Demodulátor vodoznaku je součástí software pro přehrávání, nebo nahrávání zvukových dat. Jakmile byl obsah vodoznaku dekódován, spustí se politika ochrany proti kopírování a kontroly přístupu. Děje se to pomocí hardware nebo software, který zakáže nebo povolí přístup ke zvukové nahrávce. Požadavek na algoritmus je, aby byl schopný tzv. slepé detekce vodoznaku.

- **Přenos doplňkových informací**

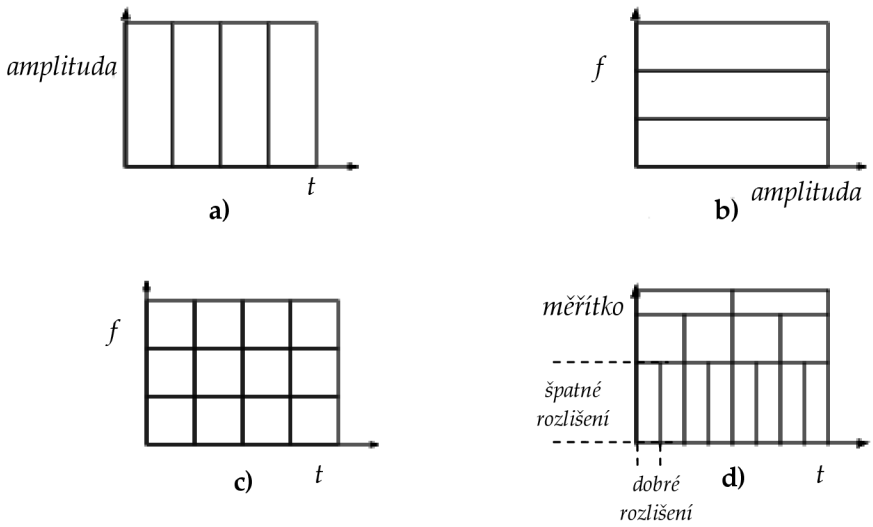
Vodoznak vložený do multimediálních dat se používá jako externí odkaz k databázi, kde jsou uloženy informace o samotném multimediálním souboru. Například to můžou být licenční podmínky nebo informace o autorských právech. Jiné využití může být například v přenosu metadat (informace o skladateli, žánru zvukové nahrávky, …) společně s multimediálními daty [3].

3 Analýza signálů

Analýza signálů slouží jako příprava k hodnocení signálu. Vstupem je analyzovaný signál (obraz, zvuk) a výstupem je co možná nejpřesnější popis signálu množinou vhodných parametrů. Typické jsou např. frekvenční, časově-frekvenční, spektrální a korelační analýza.

3.1 Úvod do transformací

Základními prostředky pro *frekvenční analýzu* spojitých signálu jsou **Fourierova řada a Fourierova transformace**. Je známo, že časově spojité periodické signály lze rozložit pomocí Fourierovy řady teoreticky do nekonečného počtu harmonických složek (sinusovek a kosinusovek). Periodické diskrétní signály (posloupnosti) je možné rozložit pouze do konečného počtu harmonických složek pomocí časově diskrétní Fourierovy řady. Základní nevýhodou Fourierovy transformace je, že nedokáže poskytnout informaci časové lokalizaci jednotlivých kmitočtů v analyzovaném signálu. Možným řešením je použití „okna", které v čase ohraničí krátký úsek signálu a umožní z něj určit spektrum v daném časovém intervalu. Jedná se o **krátkodobou Fourierovou transformací**. Je-li jako časové okno použít Gaussův impulz, pak se krátkodobá Fourierova transformace označuje jako **Gaborova transformace.** Z principu neurčitosti podle Heisenberga vyplývá, že nelze současně určit přesně kmitočtové rozlišení Δf a časové rozlišení Δt. Existuje mez dosažitelného rozlišení: $\Delta f \cdot \Delta t \geq konstanta$ pro zvolenou délku okna [13]. Prodloužení okénka má za následek zhoršení časového rozlišení, ale zlepšení kmitočtového rozlišení a naopak. Tento problém řeší **vlnková transformace**, kterou používáme především podobně jako krátkodobou Fourierovu transformaci pro analýzu nestacionárních signálů. Jedná se tedy o tzv. *časově frekvenční analýzu*.



Obr. 3.1: Princip neurčitosti: a) časová oblast (Schannon), b) frekvenční oblast (Fourier), c) STFT (Gabor), d) vlnková analýza

3.2 Vlnková transformace

Vlnková transformace (z anglického wavelet transform) je moderní transformace používaná pro analýzu především nestacionárních signálů, kompresi obrazu, detekci nespojitostí, odstranění šumu atd. Jde o jistý typ transformace, která se liší jen podle tvaru zvolené **bázové funkce**. Bázová funkce FT je definována vztahem

$$\psi(t) = e^{j\omega t}, \qquad \text{pro } \omega \in \mathbb{R}. \tag{3.1}$$

Tyto bázové harmonické funkce jsou nenulové na celé časové ose, takže každá spektrální hodnota je ovlivněna úplným průběhem signálu a nejsme schopni časově lokalizovat události ve spektru. U STFT je bázová funkce definována

$$\psi(t) = w(t-\tau) \cdot e^{j\omega t}, \tag{3.2}$$

kde: τ je časový parametr (poloha váhové funkce w).

Tvarem váhové funkce (okna) $w(t)$ lze ovlivnit rozlišovací schopnost v čase nebo kmitočtu. Rozkladem signálu dostaneme spektrogram. Tedy dvojrozměrnou funkci času a frekvence.

Vlnková transformace má bázové funkce (**vlnky**) definovány [13]:

$$\psi_{p,q}(t) = \frac{1}{\sqrt{|p|}} \psi \frac{(t-p)}{q}, \qquad \text{pro } p, q \in \mathbb{R} \tag{3.3}$$

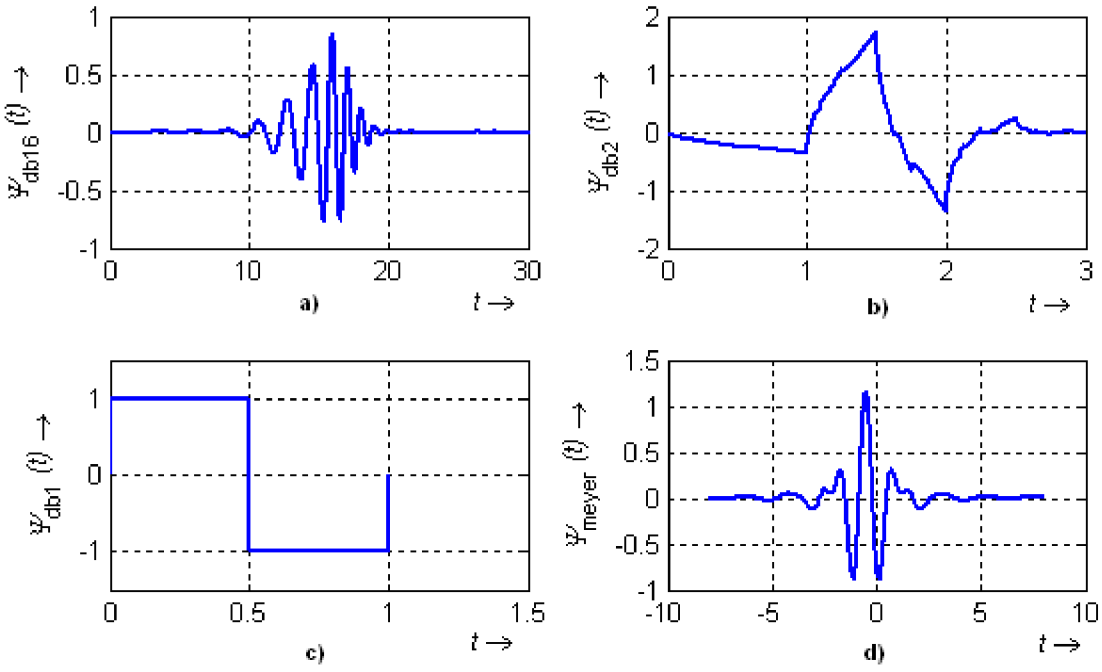
kde: p je dilatační parametr ($p < 1$ nastává komprese; $p > 1$ nastává expanze)
q je translační parametr
$\frac{1}{\sqrt{|p|}}$ je činitel, který zaručí, že energie rozšířené nebo zkrácené vlnky bude stále stejná.

Základní vlnka $\psi(t)$ se označuje jako **mateřská vlnka**, protože jsou z ní změnou měřítka (dilatace) a posunem podle časové osy (translace) odvozeny další vlnky. Na obrázku 3.2 jsou zobrazeny vybrané mateřské vlnky, které byly získány pomocí programu MATLAB.

Mateřská vlnka musí splňovat některé vlastnosti [18]:

- Musí mít nulovou střední hodnotu, z čehož plyne oscilatorický charakter.
- Může mít nenulovou hodnotu jen na konečném časovém intervalu (nebo nenulové hodnoty musí být zanedbatelné mimo tento konečný interval). Jedná se o vlnku s kompaktním nosičem.

Následkem druhé vlastnosti, kterou musí splňovat mateřská vlnka je to, že kterákoli hodnota spektra je ovlivněna pouze odpovídajícím úsekem analyzovaného signálu. Vlnkové bázové funkce ovšem pokrývají po částech celý časový rozsah analyzovaného signálu, takže je zachována úplná informace.



Obr. 3.2: Příklad mateřských vlnek: a) db16, b) db2, c) Haar (db1), d) Meyer

Další zajímavou vlastností vlnkové transformace je, že umožňuje analýzu signálu s vícenásobným rozlišením. Znamená to, že pro malé hodnoty měřítka poskytuje velké rozlišení v čase, ale malé rozlišení v kmitočtu. A naopak, pro velké hodnoty měřítka poskytuje malé rozlišení v čase, ale velké rozlišení v kmitočtu. Jde o princip neurčitosti.

Výběr tvaru vlnky pro analýzu signálu skýtá velký stupeň volnosti. V mnoha aplikacích je proveden výběr náhodně. Přesto však vlnky musí splňovat určité požadavky [18]:

- Pokud má být transformace inverzibilní, bázové funkce musí být vzájemně ortogonální a musí mít nulovou střední hodnotu.
- Pro časově frekvenční analýzu musí být vlnky kompaktní jak v časové, tak i ve frekvenční reprezentaci.

V tabulce 3.1 jsou vypsány vlastnosti vybraných mateřských vlnek. Tyto vlastnosti byly zjištěny v programu MATLAB.

Tab. 3.1: Vlastnosti vybraných mateřských vlnek

Název vlnky	Vlastnosti vlnky			
	Ortogonální	Symetrická	Komplexní	Kompaktní Nosič
Daubechies	ANO	NE	NE	ANO
Haar	ANO	ANO	NE	ANO
Morlet	NE	ANO	ANO	NE
Meyer	ANO	ANO	NE	NE
Mexican hat	NE	ANO	NE	NE

Obecně je možné funkci $s(t)$ vytvořit jako lineární kombinací bázových vlnek [13]:

$$s(t) = \sum_{p,q} b_{p,q} \cdot \psi_{p,q}(t), \tag{3.4}$$

kde: $b_{p,q}$ udávají velikost, s jakou se jednotlivé vlnky $\psi_{p,q}(t)$ podílí na vytváření signálu.

Rozkladem signálu dostaneme scalegram. Tedy dvojrozměrná funkce času a měřítka. V případě vlnkové transformace nemůžeme hovořit o kmitočtu, protože základní bázové funkce již nejsou harmonické signály násobené časovým oknem, ale jsou to krátké vlnky, u nichž se mění časové měřítko, a které jsou ještě v čase posouvány. Měřítko je nepřímo úměrné frekvenci, přičemž nízké frekvence odpovídají velkým měřítkům a naopak.

Rozlišujeme tři druhy vlnkových transformací [13]:

- **Spojitá vlnková transformace (WT)**, u níž je spojitý vstupní signál, vlnková funkce, změna měřítka a posunutí
- **Diskrétní vlnková transformace (DWT)**. Zde jsou spojité vlnkové funkce. Ale vstupní signál, změna měřítka a posunutí se mění diskrétně.
- **Diskrétní vlnková transformace s diskrétním časem (DTWT)**. Vlnkové funkce, vstupní signál, změna měřítka a posunutí se mění diskrétně.

3.3 DTWT

Abychom získali číslicově vyjádřitelnou diskrétní spektrální reprezentaci, je nutno spojitou transformaci vzorkovat. Nejčastěji se volí parametry p a q, jako násobky 2 takto:

$$p = 2^l ,$$

$$q = i \cdot 2^l = i \cdot p \qquad l, i \in \mathbb{Z}.$$

Tato volba znamená, že měřítko p je vzorkováno v dyadické posloupnosti, zatímco časové posunutí q je děleno rovnoměrně. V tomto případě se l označuje jako úroveň rozkladu. Potom se jedná o tzv. **diskrétní dyadickou vlnkovou transformaci**. Jestliže dosadíme tyto parametry do bázové funkce (3.3), tak dostaneme [13]:

$$\psi_{p,q}(n) = \frac{1}{\sqrt{p}} \cdot \psi\left(\frac{n-q}{p}\right) = \frac{1}{\sqrt{2^l}} \cdot \psi\left(\frac{n - i \cdot 2^l}{2^l}\right) = \frac{1}{\sqrt{2^l}} \cdot \psi\left(2^{-l} \cdot n - i\right). \tag{3.5}$$

Dyadická diskrétní vlnková transformace s diskrétním časem je definována [13]:

$$S_{DTWT}(l,i) = \frac{1}{\sqrt{2^l}} \sum_{n=-\infty}^{\infty} x(n) \cdot \psi^*(2^{-l} n - i). \tag{3.6}$$

Jedná se tedy o korelaci signálu $x(n)$ s jednotlivými vlnkami ψ. Lze dokázat, že při časové expanzi mateřské vlnky na 2^lnásobnou délku se odpovídající spektrum vlnky změní na $1/2^l$násobek výchozí šířky s posunem k nižším frekvencím se střední kmitočtem na $1/2^l$násobek výchozího. Dyadická DTWT je tedy charakterizována oktávovou podobou spekter soustavy vlnek [2].

DTWT lze realizovat diskrétní konvolucí diskrétního signálu x(n) s časově reverzními vlnkovými funkcemi [13]:

$$S_{DTWT}(l,i) = x(n) * \psi(2^l i - n) = x(n) * h_l(2^l i - n). \tag{3.7}$$

Dyadická DTWT může být potom vyjádřena jako:

$$S_{DTWT}(l,i) = \sum_{n=-\infty}^{\infty} x(n) \cdot h_l(2^l i - n), \tag{3.8}$$

a lze ji realizovat rozkladem signálu bankou zrcadlových filtrů s impulsními charakteristikami $h_l(n)$.

3.4 Realizace vlnkové transformace bankou zrcadlových filtrů

DTWT a IDTWT je často formulována rychlými algoritmy, realizovanými jistým typem bank filtrů. Banku číslicových filtrů můžeme rozdělit na dvě skupiny:

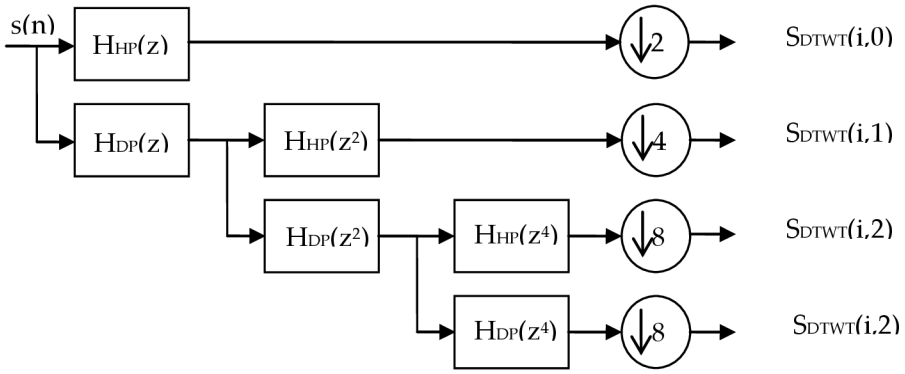
- Rozkladová banka číslicových filtrů
- Rekonstrukční banka číslicových filtrů

Základem pro rozkladovou banku číslicových filtrů je dvoukanálová banka číslicových filtrů tvořena zrcadlovými filtry typu dolní propust (dále jen DP) s přenosovou funkcí H_{HP} a horní propust (dále jen HP) s přenosovou funkcí H_{DP}. Rozdělíme tak kmitočtové pásmo na dvě stejně široká pásma.

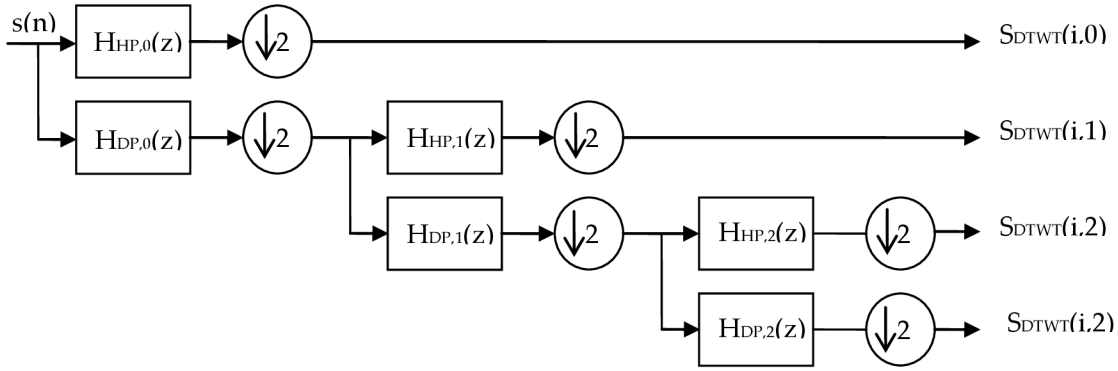
Předpokládejme dvojici zrcadlových filtrů [18].

$$\left|H_{DP}(e^{j\omega})\right| = \begin{cases} 1 & pro \quad \omega \in \langle 0, \frac{\pi}{2}) \\ 0.5 & pro \quad \omega = \frac{\pi}{2} \\ 0 & pro \quad \omega = (\frac{\pi}{2}\rangle \end{cases}, \quad \left|H_{HP}(e^{j\omega})\right| = \begin{cases} 0 & pro \quad \omega \in \langle 0, \frac{\pi}{2}) \\ 0.5 & pro \quad \omega = \frac{\pi}{2} \\ 1 & pro \quad \omega = (\frac{\pi}{2}\rangle \end{cases}. \tag{3.9}$$

Dosadíme-li $e^{j\omega} = z$ do rovnice (3.9) a provedeme-li dále substituci $z \to z^k$, obdržíme systém $H(z^k)$ s k-krát stlačenou frekvenční charakteristikou. Tato substituce se projeví zředěním impulsní charakteristiky filtru. Tedy vložením $k-1$ nulových vzorků mezi jednotlivé impulsní charakteristiky výchozího filtru. Na obr. 3.3 je realizace třístupňové dyadické DTWT s filtry odvozenými od zrcadlových filtrů. Příslušná modulová kmitočtová charakteristika je na obr. 3.5. Z hlediska implementace je velmi vhodné realizovat podvzorkování postupně tak, že výstup každého filtru v sérii podvzorkujeme s faktorem 2. Dosáhneme toho, že všechny použité filtry budou z dvojice zrcadlových filtrů H_{DP} a H_{HP}. Schéma tedy můžeme překreslit do podoby na obr. 3.4 [18].



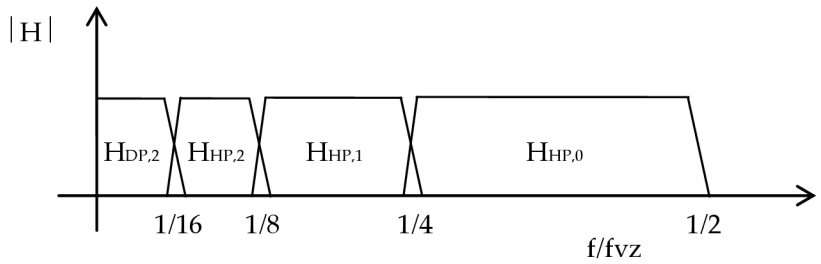
Obr. 3.3: Realizace DTWT s filtry odvozenými od zrcadlových filtrů typu DP a HP [18]



Obr. 3.4: Třístupňová dyadická DTWT realizována zrcadlovými filtry typu DP a HP [18]

Koeficienty DTWT jsou tvořeny výstupními vzorky banky filtrů. Jelikož jsou výstupní filtry podvzorkovány, tak je počet koeficientů na výstupu shodný s počtem vzorků vstupního signálu. Na počáteční úrovni, tedy pro $l = 0$, můžeme počítat spektrální koeficienty $S_{DTWT}(i,0)$ jednoduše filtrací signálu, po níž následuje podvzorkovaní s činitelem dva. Podobně obdržíme koeficienty na dalších úrovních. Pomocí HP přitom získáme tzv. **detailní** koeficienty a pomocí DP získáme tzv. **aproximační** koeficienty.

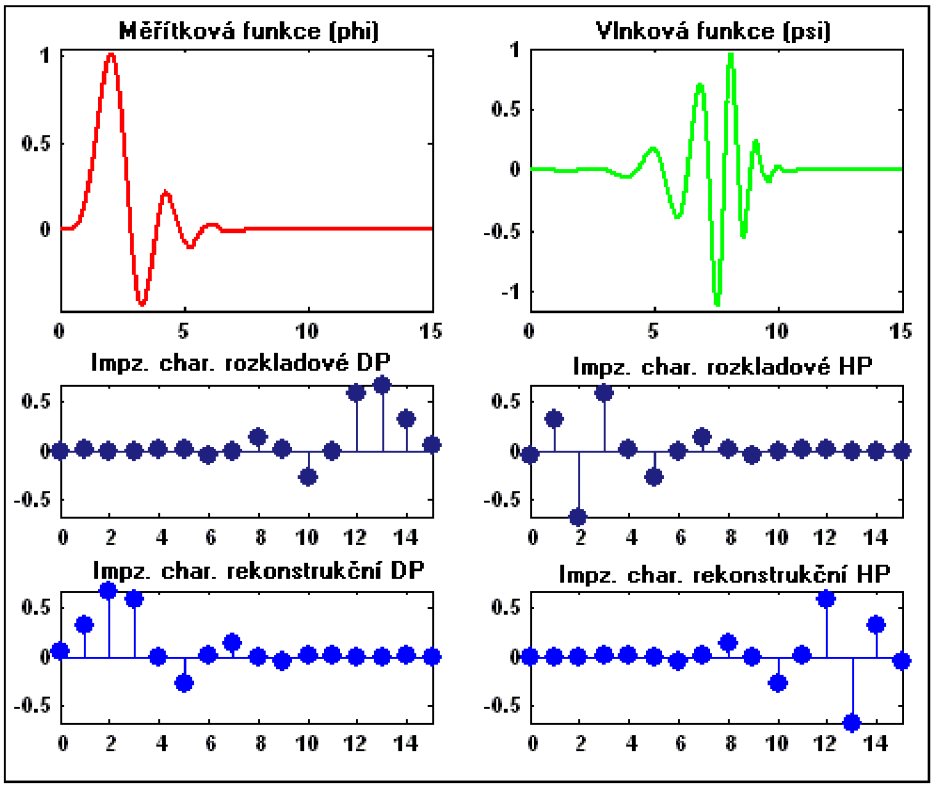
Na obr. 3.5 vidíme idealizovanou modulovou kmitočtovou charakteristiku třístupňové DTWT vycházející ze zrcadlových filtrů.



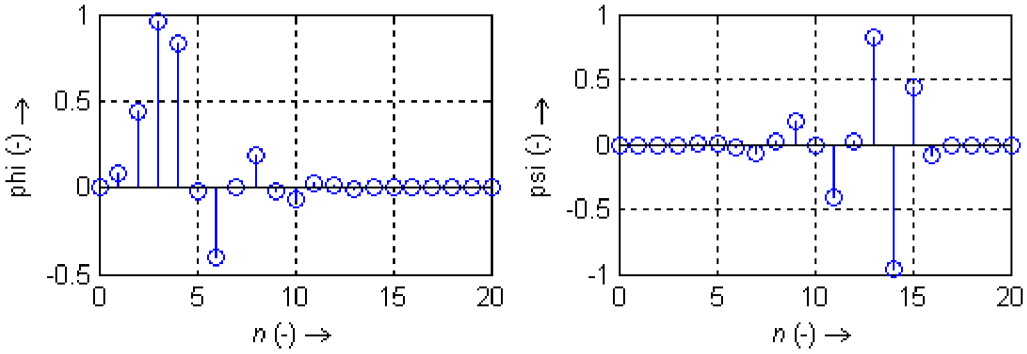
Obr. 3.5: Idealizované modulové kmitočtové charakteristiky odpovídající třístupňovému rozkladu odvozené od zrcadlových filtrů [18]

Postup rekonstrukce je opačný oproti rozkladové bance číslicových filtrů. Jsou zde bloky pro nadvzorkování s činitelem dva a zrcadlové filtry H_{DP} a H_{HP}.

Souvislost mezi časovým průběhem vlnky a impulsní charakteristikou zrcadlové banky číslicových filtrů s perfektní rekonstrukcí lze ukázat například na Daubechies vlnky 8. řádu (db8). Vlnková funkce (psi) a měřítková funkce (phi) má časový průběh zobrazený na obr. 3.6. Její diskrétní hodnoty jsou zobrazeny na obr. 3.7. Právě tyto hodnoty odpovídají impulsní charakteristice rekonstrukční DP a rekonstrukční HP. Dále si můžeme povšimnout na obr. 3.6, že počet nenulových hodnot impulsní charakteristiky je roven dvojnásobku řádu vlnkové funkce.



Obr. 3.6: Časový průběh Daubechies vlnky 8. řádu (db8) a odpovídající impulsní charakteristiky rozkladové a rekonstrukční DP a HP



Obr. 3.7: Diskrétní hodnoty měřítkové (phi)a vlnkové funkce (psi)

Mezi impulsní charakteristikou rozkladové (rekonstrukční) DP a HP platí vztah [13]:

$$h_{HP}(n) = (-1)^{N-1-n} \cdot h_{DP}(N-1-n), \tag{3.10}$$

Obě impulsní charakteristiky jsou souměrné kolem počátku a otočením lichých členů impulsní charakteristiky DP získáme HP. Srovnáním numerických hodnot, které si můžeme nechat vypsat programem MATLAB, impulsních charakteristik banky číslicových zrcadlových filtrů s vlnkovými a měřítkovými funkcemi DTWT zjistíme, že platí [13]:

$$h_{HP,R}(n) = \frac{1}{\sqrt{2}} \cdot \psi(n)\ , \tag{3.11}$$

$$h_{DP,R}(n) = \frac{1}{\sqrt{2}} \cdot \phi(n), \tag{3.12}$$

kde: $h_{HP,R}$ je impulsní charakteristika rekonstrukční horní propusti,

$h_{DP,R}$ je impulsní charakteristika rekonstrukční DP,

$\psi(n)$ je vlnková funkce,

$\phi(n)$ je měřítková funkce.

Souvislost lze najít i mezi impulsními charakteristikami rozkladové, rekonstrukční DP a HP a lze tedy psát [13]:

$$h_{DP,R}(n) = h_{DP,r}(N-1-n), \tag{3.13}$$

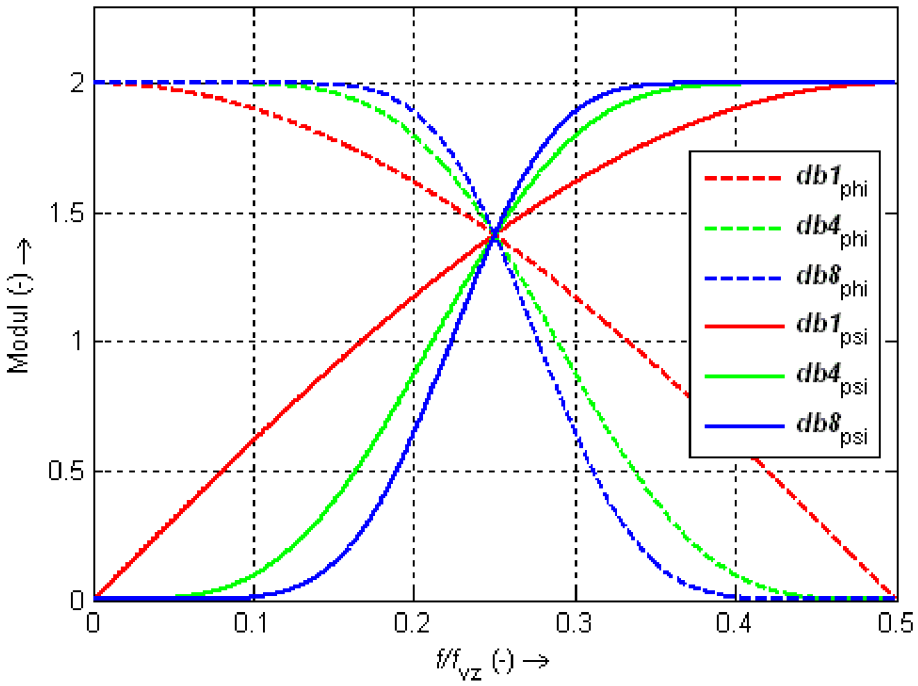
$$h_{HP,R}(n) = (-1)^n \cdot h_{DP,r}(n), \tag{3.14}$$

kde: $h_{DP,r}$ je impulsní charakteristika rozkladové DP,

N je délka impulsní charakteristiky.

Na obrázku 3.8 vidíme modulové kmitočtové charakteristiky vybraných mateřských vlnek (db1, db4, db8). S rostoucím řádem Daubechies vlnky roste strmost přechodového pásma HP a DP. Dále si můžeme povšimnout,

že modulové kmitočtové charakteristiky jsou symetrické podle hodnoty f/f_{VZ} = 0,25. Proto se těmto filtrům říká *zrcadlové filtry*.



Obr. 3.8: Modulové kmitočtové charakteristiky DP a HP odpovídající měřítkové a vlnkové funkci Daubechies vlnky

Všechny hodnoty a průběhy v obr. 3.6, 3.7, 3.8 byly získány pomocí programu MATLAB.

3.5 DWPT

DWPT (z anglického Discrete Wavelet Packet Transform) je transformace, kterou lze chápat jako zobecnění klasického pojetí vlnkové transformace. Vlastnosti, které sdílí s vlnkovou transformací jsou tak popsány v kap. 3.2–3.4 Bázové funkce $\psi_{j,b,k}$ jsou odvozeny od základní vlnky ψ_b, takto [18]:

$$\psi_{j,b,k}(n) = 2^{j/2} \cdot \psi_b \cdot (2^{-j}(n-k)), \tag{3.15}$$

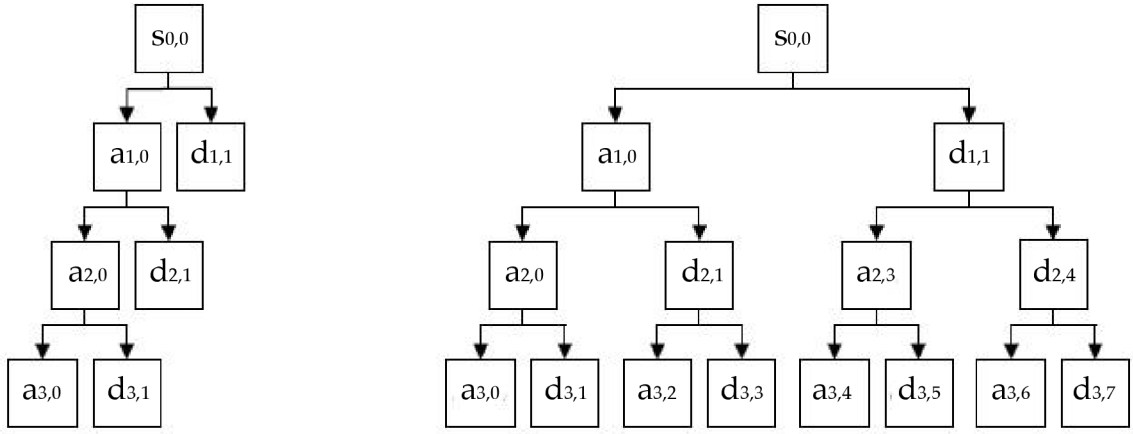
kde: j je aktuální úroveň rozkladu,
b je počet kmitů,
k je posun v čase.

Při paketové vlnkové analýze je signál $s(n)$ popsán lineární kombinací bázových funkcí $\psi_{j,b,k}(n)$ s rozdílným měřítkem, počtem kmitů a polohou [18]:

$$s(n) \approx \sum_j \sum_b \sum_k b_{j,b,k} \cdot \psi_{j,b,k}(n), \tag{3.16}$$

kde: $b_{j,b,k}$ je míra s jakou se jednotlivé bázové funkce podílejí na výsledném signále.

U DWPT tak postupně rozkládáme aproximační a detailní koeficienty úrovně j a získáme tak aproximační a detailní koeficienty úrovně $j + 1$. Vzniká tak úplná stromovitá struktura, jak je patrné z třístupňového rozkladu na obr. 3.9. Koeficient $\mathbf{a}_{1,0}$ značí aproximační koeficient v první úrovni. Koeficient $\mathbf{d}_{1,1}$ značí detailní koeficient v první úrovni. Všechna kmitočtová pásma, na která se signál rozkládá, jsou stejně široká. Ve svém důsledku je ale možné rozkládat libovolnou složku na libovolné hladině do tzv. neúplného stromu.

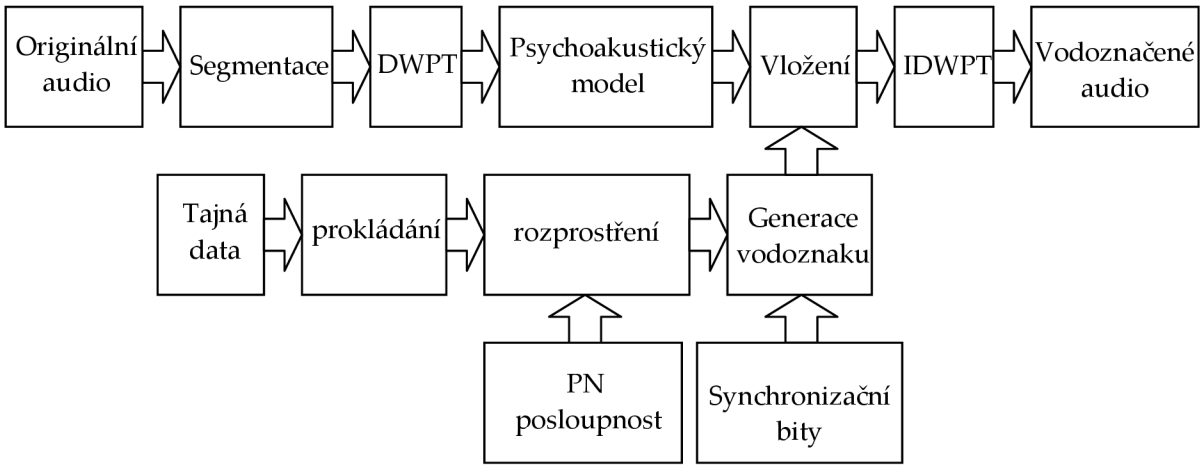


Obr. 3.9: Rozkladové stromy: a) WT, b) WPT (úplná)

Rozkladem pomocí WPT dostáváme tedy větší počet koeficientů reprezentující původní signál. Z této množiny jsme schopni vybrat pomocí vhodné výběrové funkce nejlepší reprezentaci signálu, která se označuje termínem nejlepší báze (best basis). Jednoduchým příkladem výběrové funkce je prahovací funkce. Tato funkce vrátí určitý počet koeficientů ležících nad stanoveným prahem, který nejlépe vystihuje chování signálu [2].

4 Blokové schéma kodéru a dekodéru procesu vodoznačení

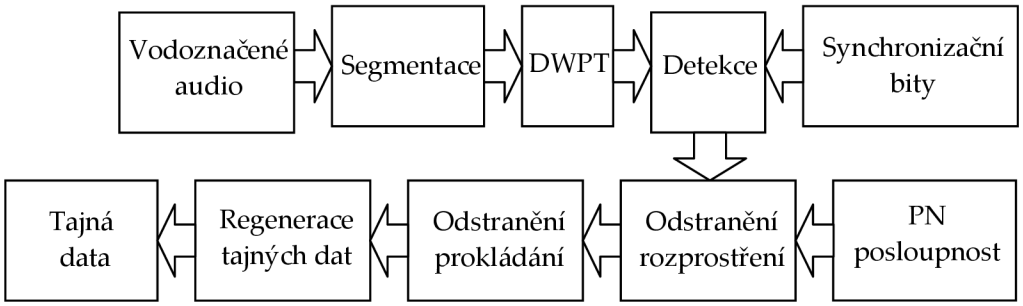
Celkový pohled na proces vodoznačení využívající metodu rozprostřeného spektra, paketovou vlnkovou transformaci a psychoakustický model je zobrazen na obr. 4.1 a 4.2.



Obr. 4.1: Kodér procesu vodoznačení využívající metodu rozprostřeného spektra [4]

Funkci kodéru lze shrnout do následujících kroků:

1. Vstupní audio v CD kvalitě je rozdělen na překrývající se segmenty.
2. Každý segment je pomocí DWPT rozložen na 25 subpásem.
3. V každém subpásmu se pomocí psychoakustického modelu stanovení maskovacího prahu.
4. Data, která mají být vkládána (tajná data) jsou podrobeny tzv. prokládání.
5. Tyto data jsou rozprostřena pomocí PN posloupnosti.
6. Synchronizační bity jsou vloženy před rozprostřená data, tímto vznikne výsledný vodoznak.
7. Způsob vkládání vodoznaku do originálního audia je závislý na maskovacím prahu.
8. Pro převod vodoznačeného audia zpět do časové oblasti je použita inverzní diskrétní paketová vlnková transformace.



Obr. 4.2: Dekodér procesu vodoznačení využívající metodu rozprostřeného spektra [4]

Dekodér pracuje následujícím způsobem:

1. Vodoznačené audio je nejprve rozděleno na překrývající se segmenty.
2. Každý segment je pomocí DWPT rozložen na 25 subpásem.
3. Proces synchronizace se děje na základě hledání synchronizační sekvence.
4. Nalezená data jsou následně podrobena inverzní operaci k operaci rozprostření a prokládání za účelem obnovení tajných dat.

5 Implementace algoritmu pro vodoznačení audia

5.1 Psychoakustický model

Pomocí tzv. psychoakustického modelu stanovíme celkový maskovací práh, pod který můžeme vkládat vodoznak. Je to tedy matematický model popisující chování lidského ucha. Tento model je velmi důležitou součástí procesu vodoznačení. V této práci využívám psychoakustický model založený na DWPT.

Postup při výpočtu celkového maskovacího prahu můžeme shrnout do následujících kroků [4]:

1. Vstupní audio signál v CD kvalitě je rozdělen na překrývající se segmenty.
2. Každý segment je pomocí DWPT rozložen na 25 subpásem. (Podrobněji viz kapitola 5.1.1).
3. Energie signálu v každém subpásmu se vypočítá ve vlnkové oblasti takto:

$$E(z) = \sum_i x_i^2 \,, \tag{5.1}$$

   kde: z je index subpásma ($1 \le z \le 25$),
   x_i je i-tý koeficient získaný pomocí DWPT v subpásmu z.

4. Identifikace tonálních a netonálních komponentů v signálu. (Podrobněji viz kapitola 5.1.2). Výsledkem je tonální faktor $a(z)$.
5. Přepočet energie podle tonálního faktoru dle vzorce:

$$SE(z) = E(z) \cdot 10^{\frac{a(z)}{10}} \,, \tag{5.2}$$

6. Rozprostření energie do vedlejších kritických pásem lze spočítat jako konvoluci energie $SE(z)$ s funkcí rozprostření $SP(z)$. (Podrobněji viz kapitola 5.1.3). Ke stanovení skutečného maskovacího prahu subpásma z použijeme vzorec:

$$C(z) = SE(z) * SP(z) \,. \tag{5.3}$$

7. Maskovací práh v každém subpásmu je normalizován jeho šířkou a následně porovnán s prahem slyšení (Podrobněji viz kapitola 5.1.4). Výběrem maximální hodnoty se stanoví celkový maskovací práh. (Podrobněji viz kapitola 5.1.5).

$$T(z) = \max\left(\frac{C(z)}{L(z)}, T_{abs}(z)\right), \tag{5.4}$$

   kde: T_{abs} (z) je hodnota prahu slyšení pro subpásmo z,
   L (z) je šířka subpásma z.

5.1.1 Rozložení signálu pomocí DWPT

V psychoakustickém modelu doposud převládaly metody založená na Fourierově transformaci. Použití vlnkové transformace je poměrně nové. Pomocí DWPT rozložíme každý segment na 25 subpásem, které aproximují kritická pásma, a tímto pokryjeme celou slyšitelnou kmitočtovou oblast. Použitý rozkladový strom vidíme na obr. 5.1. Index kritických pásem je číslován od 1 do 25. Rozdělení do klasických kritických pásem je uvedeno v tab. 1.1. V tomto algoritmu používám dělení uvedené v tab. 5.1.

Tab. 5.1: Kritická pásma navrhovaného modelu [4]

Kritická Pásma	Hranice pásma (Hz)	Střední frekvence (Hz)	Hranice pásma (Hz)	Šířka pásma (Hz)
1.	0	43	86	86
2.	86	129	172	86
3.	172	258	344	172
4.	344	387	430	86
5.	430	473	516	86
6.	516	602	687	171
7.	687	730	773	86
8.	773	816	859	86
9.	859	945	1031	172
10.	1031	1203	1375	344
11.	1375	1462	1549	174
12.	1549	1634	1719	170
13.	1719	1891	2062	343
14.	2062	2234	2406	344
15.	2406	2578	2750	344
16.	2750	2922	3093	343
17.	3093	3265	3437	344
18.	3437	3781	4125	688
19.	4125	4813	5512	1387
20.	5512	5968	6437	925
21.	6437	6906	7375	938
22.	7375	8313	9250	1875
23.	9250	10137	11024	1774
24.	11024	13762	16538	5514
25.	16538	19294	22050	5512

Výběr mateřské vlnky je důležitý pro splnění požadovaného časového rozlišení, tedy méně než 10 ms u vysokých kmitočtů a do 100 ms u nízkých kmitočtů. Z těchto důvodů je vybrána, pro audio v CD kvalitě, mateřská vlnka **Daubechies 8. řádu** (Počet nenulových hodnot L = 16 vzorků).

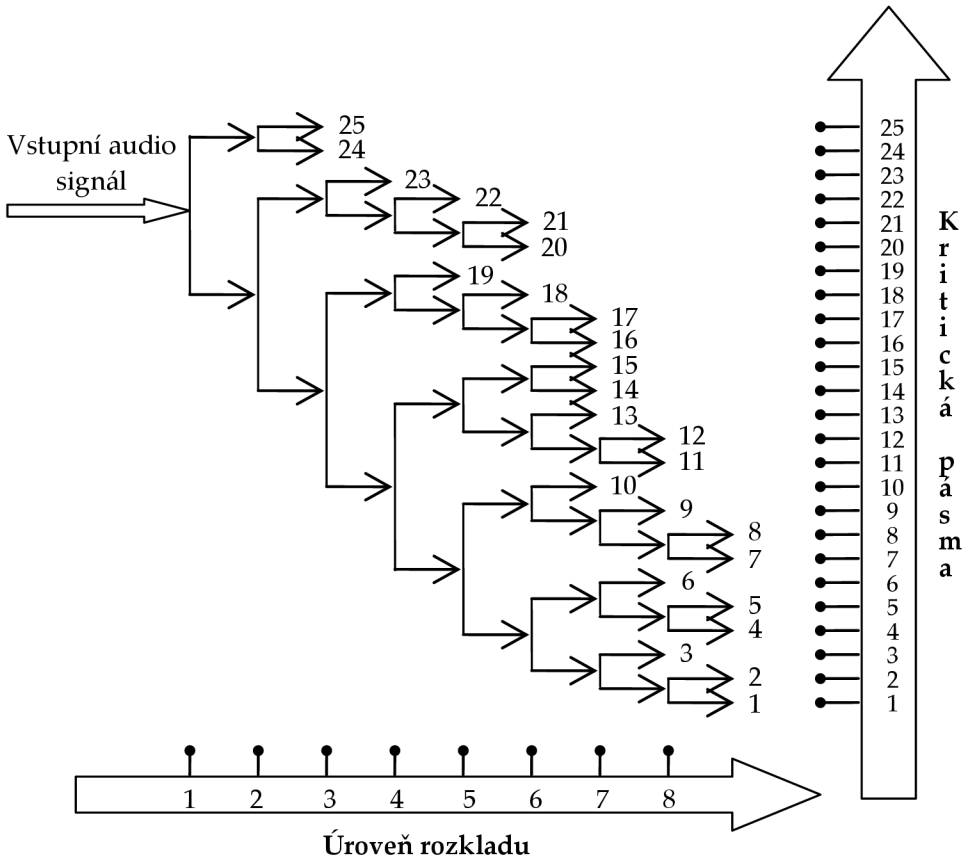
Šířka subpásma (ve vzorcích) na úrovni j ($2 \le j \le 8$) je dána:

$$F_j = 2^j . \tag{5.5}$$

Doba trvání analyzujícího okna (ve vzorcích) na úrovni *j* je dána:

$$W_j = (L-1)\cdot(F_j - 1) + 1. \tag{5.6}$$

Pro audio signál s šířkou pásma 22 kHz je maximální délka subpásma $F_{\max} = 2^8 = 256$ vzorků, které odpovídá frekvenčnímu rozlišení 22000/256 = 86 Hz. Minimální délka subpásma je $F_{\min} = 2^2 = 4$ vzorky, které odpovídají frekvenčnímu rozlišení 22000/4 = 5,5 kHz. Maximální doba trvání analyzujícího okna je $W_{\max} = 15\cdot(256-1)+1 = 3826$ vzorků. Při vzorkovacím kmitočtu 44,1 kHz to odpovídá 87 ms. Minimální doba trvání analyzujícího okna je $W_{\min} = 15\cdot(4-1)+1 = 46$ vzorků, neboli 1 ms. [4]



Obr. 5.1: Rozložení vstupního signálu pomocí DWPT [4]

5.1.2 Identifikace tonálních a netonálních složek

Identifikace tonálních (periodických) a netonálních (šumových) složek v signálu je velmi důležitá. Každý z těchto typů složek audia vyžaduje rozdílné maskovací úrovně, a proto jsou zpracovány odděleně.

Parametrem $a(z)$ určujeme, zda se jedná o složku tonální, nebo netonální, a to podle vzorce:

$$a(z) = \lambda \cdot a_{tmn}(z) + (1-\lambda) \cdot a_{nmn}(z), \tag{5.7}$$

kde: $a_{tmn}(z)$ je ukazatel tonální složky,
$a_{nmn}(z)$ je ukazatel netonální složky.

Pro všechny subpásma z je ukazatel netonální složky $a_{nmn}(z) = -9$. Ukazatel tonální složky stanovíme jako:

$$a_{tmn}(z) = -0{,}275 \cdot z - 15{,}025. \tag{5.8}$$

V rovnici 4.7 se vyskytuje parametr λ. Jedná se o tonální koeficient definovaný:

$$\lambda = \min\left(\frac{sfm}{sfm_{\min}},\ 1\right). \tag{5.9}$$

Koeficient λ je menší hodnota ze dvou parametrů uvedených v závorce rovnice 5.9. Hodnota $sfm_{\min}$ je -25 dB. Hodnota sfm je definována jako poměr geometrického průměru a aritmetického průměru rozložení energie signálu ve spektru. Lze ji vypočítat pomocí vztahu:

$$sfm = \frac{\left(\prod_{i=1}^{N} x_i^2\right)^{1/N}}{\frac{1}{N}\sum_{i=1}^{N} x_i^2}, \tag{5.10}$$

Proměnná N představuje počet koeficientů v subpásma z. sfm může nabývat hodnot z intervalu $\langle 0,1\rangle$. Hodnoty sfm blíže k jedné značí, že se jedná o netonální komponentu. Naopak hodnoty blíže k nule, značí tonální komponentu. [4]

5.1.3 Funkce rozprostření

Maskování, vycházející z reálného audio signálu, který obsahu složky tonální a netonální na různých kmitočtech a intenzitách, není jednoduše rovno součtu jednotlivých masek signálu. Maskovací efekt tedy není omezen šířkou kritického pásma, rozprostírá se i do sousedních pásem. Pomocí tzv. **funkce rozprostření** zohledníme vliv maskování na sousední kritická pásma, a tím jsme schopni stanovit skutečný maskovací práh v každém z těchto pásem, označovaných v rovnici 5.11 proměnnou z. U nízkých kmitočtů klesá okolo 25 dB/Bark. Její definici najdeme v normě [6]:

$$SP_{dB}(z) = 15{,}81 + 7{,}5 \cdot (z + 0{,}474) - 17{,}5 \cdot \sqrt{1 + (z + 0{,}474)^2}. \tag{5.11}$$

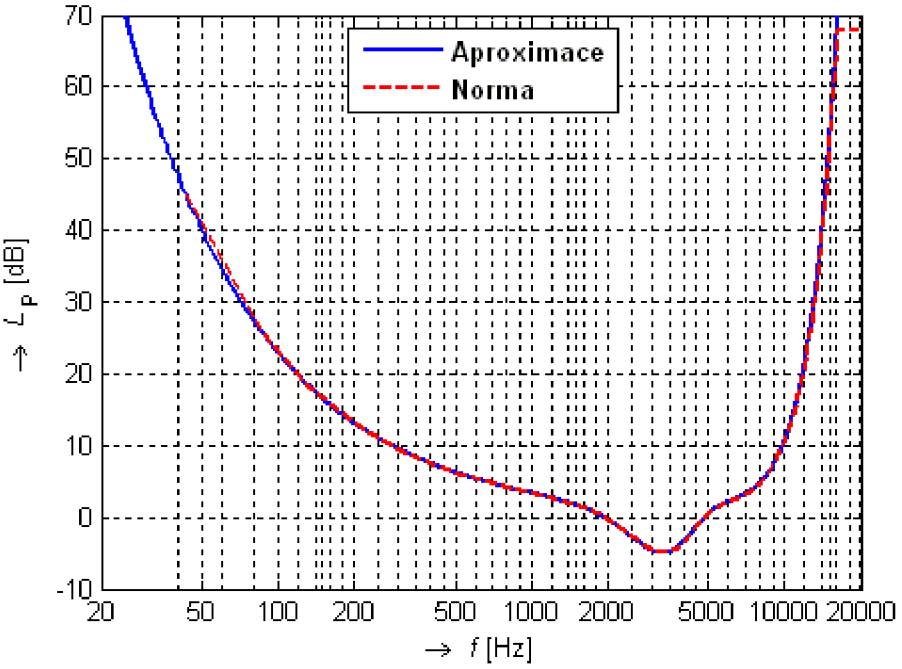
5.1.4 Práh slyšitelnosti

Práh slyšitelnosti představuje minimální hodnoty intenzity zvuku, při které začínáme vnímat čistý tón. Tuto prahovou křivku můžeme aproximovat v kmitočtové oblasti podle vzorce [5]:

$$ATH(f) = 3{,}64 \cdot \left(\frac{f}{1000}\right)^{-0{,}8} - 6{,}5 \cdot e^{-0{,}6\cdot\left(\frac{f}{1000}-3{,}3\right)^2} + 10^3 \cdot \left(\frac{f}{1000}\right)^4. \qquad (5.12)$$

Na obr. 5.2 vidíme závislosti hladiny akustického tlaku na kmitočtu. O aproximaci dle vzorce 5.12 se jedná v případě modré křivky. Druhá křivka představuje přímo hodnoty z normy [6] pro Layer-II a Fvz = 44100 Hz. Průběh byl vytvořen v programu MATLAB.

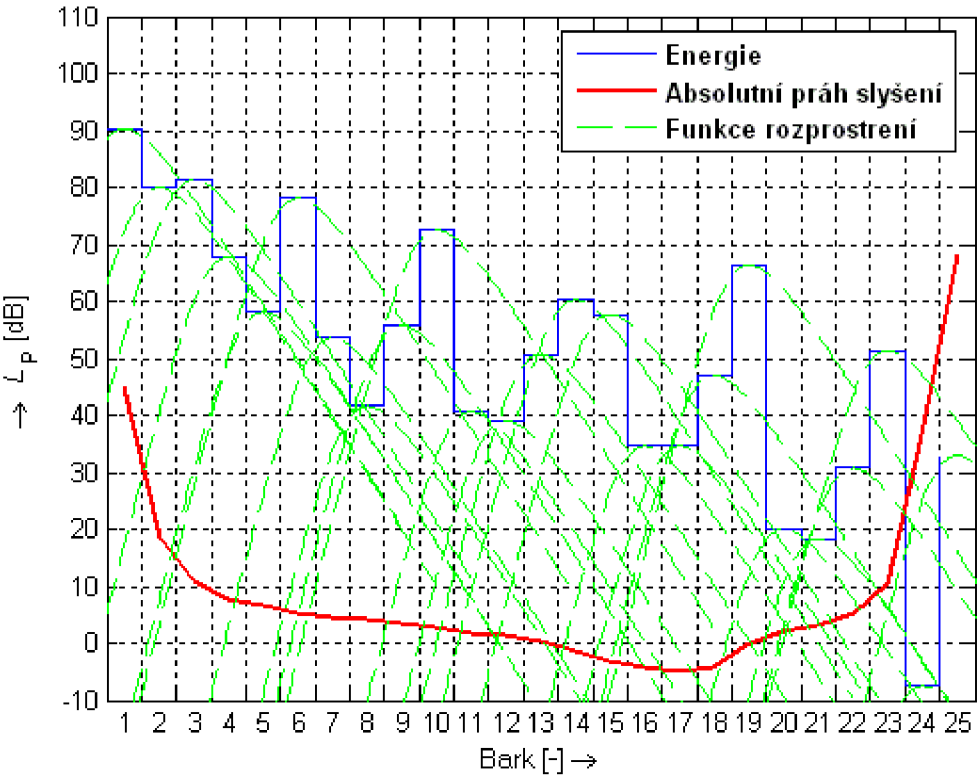
Hodnoty prahu slyšení v dB pro dané kmitočty je možné získat z normy [6] pro různé vzorkovací kmitočty a úrovně I, II.



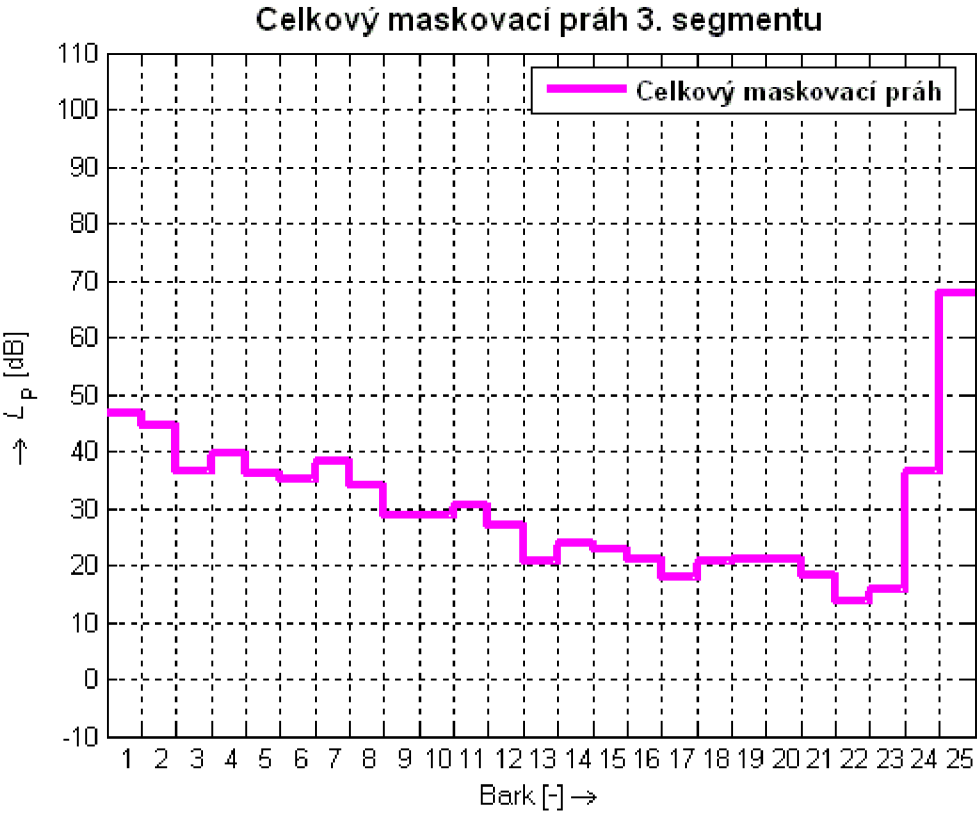
Obr. 5.2: Křivky absolutního prahu slyšitelnosti lidského sluchu

5.1.5 Celkový maskovací práh

Celkový maskovací práh se stanovuje pro každý segment zvlášť. Na obr. 5.3 vidíme rozložení energie v příslušném segmentu jako funkcí kritických pásem. Jde vidět, že největší energie u audia je obsažena v první polovině frekvenčního pásma. Dále je zde zobrazen absolutní práh slyšení a funkce rozprostření pro každé kritické pásmo. Na obr. 5.4 vidíme celkový maskovací práh pro určitý segment audia.



Obr. 5.3: Funkce rozprostření



Obr. 5.4: Celkový maskovací práh pro 3. segment

5.2 Generace vodoznaku

K vytvoření vodoznaku {w} je použita metoda přímého rozprostření spektra. Tato bitová posloupnost tvořící vodoznak je následně vkládána podle psychoakustického modelu ke koeficientům vlnkové transformace.

5.2.1 Prokládání

Vzniku chyb nemůžeme nikdy úplně zabránit. Jejich uplatnění ve výsledné přijaté zprávě však můžeme částečně předejít kódovým zabezpečením. Jedno z možností je kanálové kódování, které zvětší počet bitů ve výsledné zprávě. Prokládání se používá jako doplněk kanálového kódování, kvůli ochraně proti **shlukům chyb**. Prokládání je postup, při kterém se původní bity přeskládají do jiného pořadí.

Předpokládejme, že zpráva $\{z_n\}$ ($1 \le n \le M$), která se má vkládat, obsahuje M bitů sestávají z „1“ a „0“ („0“ se v tajné zprávě převádí na „–1“). Převod z binární na bipolární posloupnost se provádí podle vztahu:

$$m_n = 2.z_n - 1. \tag{5.13}$$

Nejprve použijeme opakovacího kódování. Každý bit zprávy se opakuje *N*-krát. Dostaneme tedy sekvenci bitů: $\{m_1 m_1 m_1 ... m_2 m_2 m_2 ... m_M m_M ... m_M\}$. Tento nový bitový tok je následně v prokladači zpožděn o určitý čas, který souvisí s hloubkou prokládání M. Je použito **blokové prokládání** s délkou kódového slova N a hloubkou prokládání M. Čím větší je hloubka prokládání, tím větší může být skupinová chyba, kterou je prokladač schopen rozprostřít. Délka kódového slova udává počet bitů, po kterých se budou opakovat vzniklé ojedinělé chyby. V kodéru (přímé prokládání) se provádí zápis do matice po řádcích a čtení z matice po sloupcích. V dekodéru (inverzní prokládání) se provádí zápis do matice po sloupcích a čtení po řádcích. Proto se při přenosu nevyskytují jednotlivé symboly jednoho kódového slova těsně za sebou a případný shluk chyb je tudíž rozprostřen mezi více kódových slov. Tyto chyby lze potom opravit. Bitový tok za prokládací maticí u kodéru má tvar: $\{m_1 m_2 ... m_M m_1 m_2 ... m_M ... m_1 m_2 ... m_M\}$.

Pro názornou ukázku si uveďme tento příklad. Předpokládejme zprávu, která obsahuje osm bitů:

m = { 1 –1 –1 1 1 –1 –1 1 }.

Následně je každý bit vstupní posloupnosti opakován *N*-krát (v tomto případě N = 5). Dostaneme bitový tok:

m = { 1 1 1 1 1 –1 –1 –1 –1 –1 –1 –1 –1 –1 –1 1 1 1 1 1
1 1 1 1 1 –1 –1 –1 –1 –1 –1 –1 –1 –1 –1 1 1 1 1 1}.

Tento bitový tok je načítán, způsobem uvedeným výše, do prokládací matice M*N, viz obr. 5.2.

Tab. 5.2: Příklad přímé prokládací matice o rozměrech M = 8, N = 5.

1	1	1	1	1
–1	–1	–1	–1	–1
–1	–1	–1	–1	–1
1	1	1	1	1
1	1	1	1	1
–1	–1	–1	–1	–1
–1	–1	–1	–1	–1
1	1	1	1	1

Bity se nyní vyčítají ze sloupců matice do nové sekvence:

$$m = \{\, 1\ {-1}\ {-1}\ \ 1\ \ 1\ {-1}\ {-1}\ \ 1\ \ 1\ {-1}\ {-1}\ \ 1\ \ 1\ {-1}\ {-1}\ \ 1\ \ 1\ {-1}\ {-1}\ \ 1$$
$$1\ {-1}\ {-1}\ \ 1\ \ 1\ {-1}\ {-1}\ \ 1\ \ 1\ {-1}\ {-1}\ \ 1\ \ 1\ {-1}\ {-1}\ \ 1\ \ 1\ {-1}\ {-1}\ \ 1\}.$$

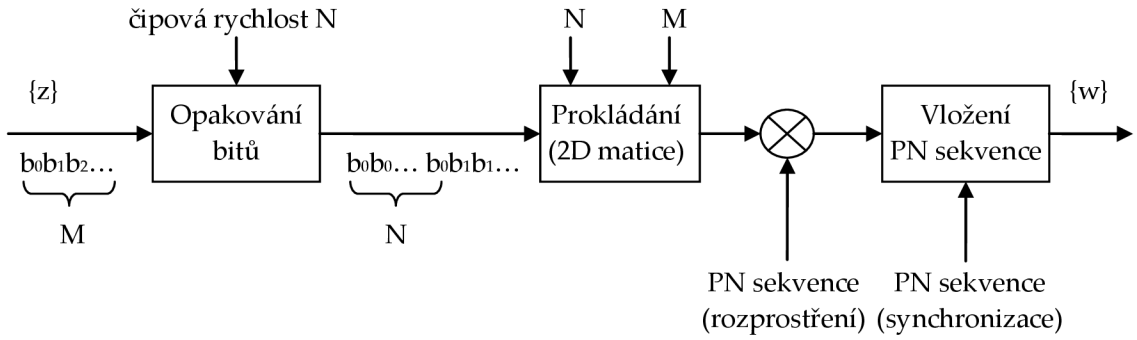
5.2.2 Vkládání vodoznaku

Za prokladačem se tajná zpráva {m} rozprostře pomocí PN posloupnosti. Tato posloupnost musí splňovat podmínky uvedené v kap. 2.2 a dále musí být zaručeno, že délka PN posloupnosti je delší nebo rovna délce zprávy {m}. V praxi je výhodné, aby tyto signály byly bipolární, a to typu NRZ (Not Return to Zero). Operaci sčítání modulo 2 je možno nahradit prostým násobením. Jedná se tedy o realizaci systému s přímým rozprostřením spektra (DSSS). Před tuto rozprostřenou sekvenci se vkládají synchronizační bity. Tvoří ji typicky PN posloupnost, která musí splňovat následující podmínky [4]:

- Délka PN posloupnosti by měla být co možná nejkratší, abychom nevyprodukovali slyšitelný šum. Na druhou stranu musí být dostatečně dlouhá, aby korelační funkce měla na straně přijímače jednoznačnou špičku.
- PN posloupnost musí být vložena tak, aby byla odolná vůči desynchronizačním útokům.

Z těchto důvodů má PN posloupnost použitá v tomto systému pro synchronizaci délku 2560 vzorků, nebo asi 58 ms při vzorkovacím kmitočtu 44100 Hz [4]. Důležitým faktorem je to, aby měl kodér i dekodér **totožné kopie obou PN posloupností**.

Proces tvorby vodoznaku za účelem zvětšení jeho robustnosti vidíme na obr. 5.5. Bitová posloupnost tvořící vodoznak má následující tvar: $w = \{pns_1 pns_2 ... pns_{2560} s_1 s_2 ... s_M s_1 s_2 ... s_M ... s_1 s_2 ... s_M\}$. pns značí PN synchronizační bity, s značí rozprostřenou sekvenci bitů.



Obr. 5.5: Obecné blokové schéma generace vodoznaku [21]

Vodoznak {w} je vícekrát vkládán ke koeficientům vlnkové transformace následujícím způsobem [4]:

$$c_k = \begin{cases} c_k + \alpha \cdot w_i & , je\text{-}li\ c_k^2 > T \\ \sqrt{T} \cdot w_i & , je\text{-}li\ c_k^2 \le T \end{cases}, \tag{5.14}$$

kde c_k je hodnota k-tého koeficientu vlnkové transformace,

α je parametr určující intenzitu vodoznaku z rozsahu ($0 \le \alpha \le 1$),

w_i je i-tý bit vodoznaku, který se má vkládat,

T je hladina maskovacího prahu pro dané subpásmo.

Zvětšováním hodnoty α zvyšujeme robustnost vloženého vodoznaku. Nicméně hrozí nebezpečí narušení kvality audio signálu. Zvolením $\alpha = 0$ zajistíme to, že vodoznak bude nevnímatelný.

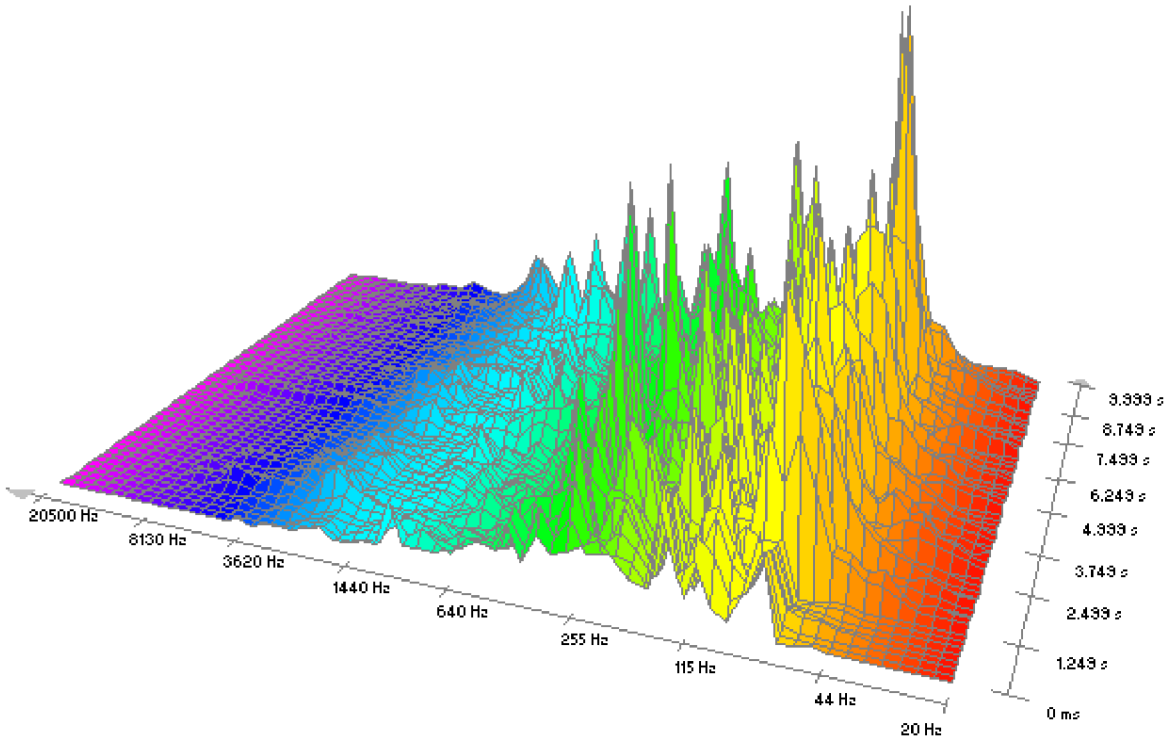
Pro zajímavost je v tab. 5.3 uveden příklad udávající počet vložených vodoznaků w = „HEITEL“ do audio souborů různých žánrů délky 10 s, koeficientem intenzity α = 0,0.

Tab. 5.3: Počet vložených vodoznaků do audio souborů vybraných žánrů

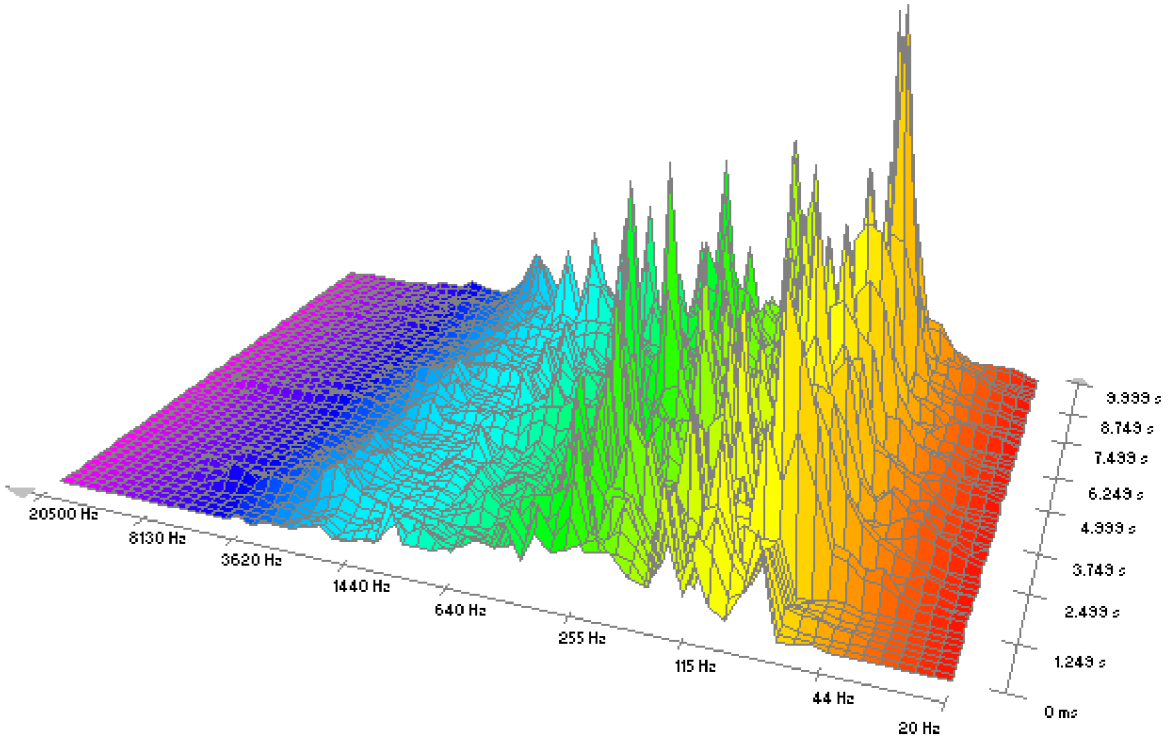
Žánr	Pop	Rock	Country	Klasika	Disco	Swing
Počet w	103	50	92	142	73	98

5.3 Výsledné vodoznačené audio

Transformací z vlnkové oblasti do časové oblasti užitím inverzní paketové vlnkové transformace získáme segment vodoznačeného audio signálu. Složením těchto segmentů vytvoříme výsledné vodoznačené audio. Na obrázku 5.6 a 5.7 jsou k porovnání frekvenční analýzy originální a vodoznačené nahrávky.



Obr. 5.6: 3D frekvenční analýza originálního audio souboru



Obr. 5.7: 3D frekvenční analýza vodoznačeného audio souboru; w = „HEITEL“, $\alpha = 0{,}0$

5.4 Extrakce vodoznaku

Extrakce vodoznaku je proces detekování a následného dekódování vodoznaku z vodoznačených dat tak, aby ho bylo možné porovnat s vloženým vodoznakem. Shoda vodoznaků zajišťuje autenticitu dat. Extrakce zahrnuje operace jako segmentace, dále využívá transformaci DWPT a psychoakustický model. Je zde použit stejný model a stejný rozkladový strom, jako v případě vkládání vodoznaku (popsáno v kap. 5.1). Velmi důležitý je proces synchronizace a samotné dekódování vodoznaku, které jsou popsány níže. Celkový pohled na dekodér vodoznaku je zobrazen v kap. 4.

Pro extrakci vodoznaku **není** potřeba originální audio soubor, jedná se tedy o tzv. slepou detekci vodoznaku. V případě detekce vodoznaku můžou nastat následující možnosti:

- Algoritmus rozhodne, že audio soubor obsahuje vodoznak. Dekódovaný vodoznak je identický s vloženým vodoznakem.
- Algoritmus rozhodne, že audio soubor obsahuje vodoznak. Dekódovaný vodoznak ale není identický s vloženým vodoznakem.
- Algoritmus rozhodne, že audio soubor neobsahuje vodoznak. V případě, že vodoznak nebyl vložen, nebo nebyla nalezena synchronizační sekvence a tudíž nebyla možnost vodoznak dekódovat.

5.4.1 Synchronizace

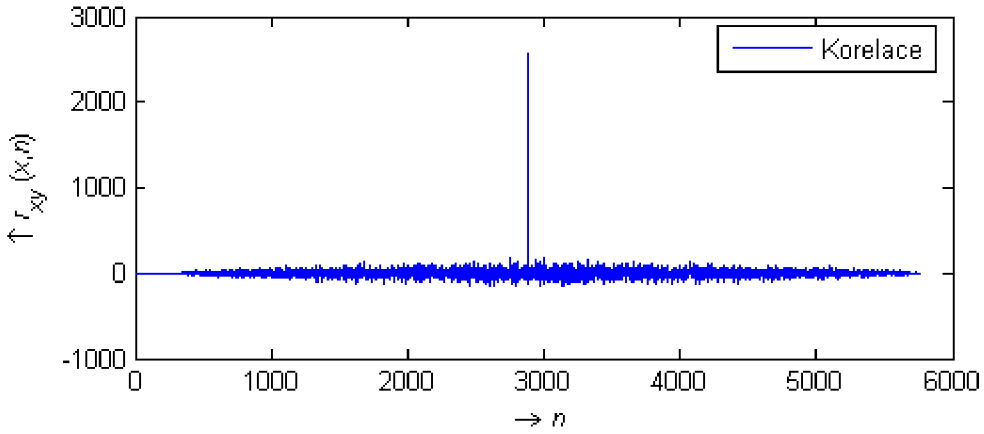
Mnoho systémů pro vodoznačení audia se potýká s problémem synchronizace. Aby mohl dekodér extrahovat vodoznak z vodoznačeného audia, musí nejprve znát umístění začátku vodoznaku.

Existuje mnoho řešení synchronizace. Ve stručnosti uveďme některé z nich:

- Použití synchronizačních značek
- Vytvoření vodoznaku, který umožňuje přímo synchronizaci
- Vyhledáním charakteristických míst v audiu (synchronizace na základě obsahu audia)

V této diplomové práci je použita synchronizace na základě **synchronizačních značek**. Synchronizační značky jsou kódy s určitou vlastností, které znají jak kodér, tak i dekodér vodoznaku. Tyto kódy nenesou žádnou informaci, jsou použity pouze pro potřeby synchronizace. Jako synchronizační kód je použita **PN posloupnost** popsána v kap. 5.2.2. Na obr. 5.8 vidíme křížovou korelační funkci PN posloupnosti délky 2560 vzorků a dekódovaných bitů (dekódován byl jeden vodoznak {w}; tajná zpráva má délku 8 Bytů, N = 5) délky 2880 vzorků. Hodnoty korelační funkce jsou závislé na podobnosti či nepodobnosti obou signálů při jejich vzájemném posunutí o τ. Jak je patrné z obr. 5.8, maximální

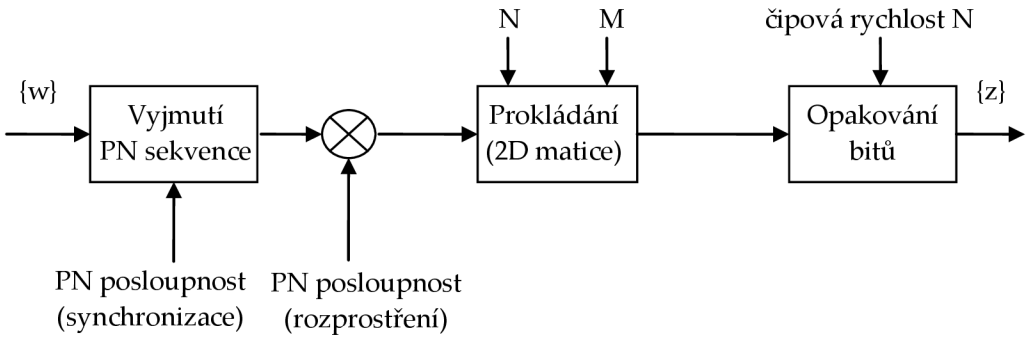
korelace nastává na indexu $n = 2880$. Lze tedy konstatovat, že na indexu $n + 2560$ (délka PN posloupnosti) začínají bity vodoznaku W. Je nutné přepočítat index n na index i dekódovaných bitů. Délka korelační funkce je $(L \cdot 2) - 1$, kde L je délka dekódovaných bitů.



Obr. 5.8: Odhad křížové korelační funkce PN posloupnosti a dekódovaných bitů

5.4.2 Dekódování vodoznaku

Proces dekódování obsahuje operace inverzní vůči vkládání vodoznaku. Postup je naznačen na obr. 5.9. Dekodér tedy musí znát od kodéru vodoznaku bity PN posloupnosti použité pro synchronizaci a rozprostření, hodnotu α, hodnoty maskovacího prahu T, vodoznak w (pro výpočet bitové chybovosti) a koeficienty vlnkové transformace originálního signálu c_{ke}.



Obr. 5.9: Blokové schéma dekódování vodoznaku

Po nalezení synchronizační PN posloupnosti se při dékódování vodoznaku musí rozlišit dva případy.

V prvním případě ($\alpha = 0$) je vodoznak vkládán ke koeficientům vlnkové transformace, pokud je splněna podmínka $c_k^2 \le T$. Nyní definujme $d = c_k$, kde c_k je k-tý koeficient vlnkové transformace. Potom je bitový tok dekódován podle následujícího pravidla:

$$w_i = \begin{cases} 1 & , je\text{-}li\ \ d > 0 \\ -1 & , je\text{-}li\ \ d \le 0 \end{cases}. \qquad (5.15)$$

V druhém případě byl vodoznak vkládán ke všem koeficientům vlnkové transformace ($\alpha \neq 0$). V případě podmínky $c_k^2 \leq T$ je bitový tok dekódován způsobem uvedeným výše, dle rovnice 5.15. V případě podmínky $c_k^2 > T$ je bitový tok dekódován podle následujícího pravidla:

$$w_i = \begin{cases} 1 & , je\text{-}li\ c_k - c_{ke} > 0 \\ -1 & , je\text{-}li\ c_k - c_{ke} \leq 0 \end{cases}, \tag{5.16}$$

kde: c_{ke} je *k*-tý koeficient vlnkové transformace originálního audio souboru,
c_k je *k*-tý koeficient vlnkové transformace vodoznačeného audio souboru.

Bitový tok $\{w\}$ je následně vynásoben s totožnou kopií PN posloupností jako u kodéru vodoznaku. Tento bitový tok je načítán do sloupců prokládací matice o rozměrech M*N. Čtení se provádí po řádcích (inverzní prokládání). Dostaneme nový bitový tok $\{w\}$. Podle následujícího pravidla získáme *n*-tý bit vodoznaku:

$$W_n = \begin{cases} 1 & , je\text{-}li\ r_n > 0 \\ -1 & , je\text{-}li\ r_n \leq 0 \end{cases}, \tag{5.17}$$

kde

$$r_n = \sum_{i=(n-1)\cdot N+1}^{n\cdot N} w_i \,. \tag{5.18}$$

N přestavuje hodnotu opakování použitého u kodéru vodoznaku, w_i je *i*-tý bit výše uvedené sekvence $\{w\}$.

Předpokládejme, že bylo dekódováno **H** vodoznaků a délka každého z nich je M. Potom je *n*-tý bit **výsledného vodoznaku** definován jako [4]:

$$W_n = \mathrm{sgn}\left(\sum_{i=1}^{H} w_{i,n} \right), \tag{5.19}$$

kde $w_{i,n}$ je *n*-tý bit *i*-tého dekódovaného vodoznaku a sgn je funkce signum definována:

$$\mathrm{sgn}(n) = \begin{cases} 1 & , je\text{-}li\ n > 0 \\ -1 & , je\text{-}li\ n \leq 0 \end{cases}. \tag{5.20}$$

Na základě rovnice 5.19 je dekodér schopen správně obnovit vodoznak i v případech, kdy některý z dekódovaných bitů vodoznaku byl přijat s chybou. Pro převod zpět z bipolární na binární posloupnost použijeme následující vzorec:

$$z_n' = \frac{(W_n + 1)}{2}. \tag{5.21}$$

Pokud byla zpráva z_n' dekódována správně, mělo by platit $\{z_n'\} = \{z_n\}$, $1 \leq n \leq M$.

Mějme příklad. Po nalezení synchronizační PN posloupnosti dostaneme, vlivem chyb na přenosovém kanále, následující bitový tok:

$$w = \{\, x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ x\ \ 1\ -1\ -1\ \ 1$$
$$1\ -1\ -1\ \ 1\ \ 1\ -1\ -1\ \ 1\ \ 1\ -1\ -1\ \ 1\ \ 1\ -1\ -1\ \ 1\ \ 1\ -1\ -1\ \ 1\,\}.$$

Symbol x značí neznámý bit z rozsahu „1“, nebo „–1“. Následně je použita inverzní prokládací matice se stejnou hloubkou prokládání M a délkou kódového slova N, jako v případě vkládání vodoznaku (v našem případě M = 8, N = 5). Bity jsou načítány do matice způsobem uvedeným v tab. 5.4. Funkce inverzního prokládání je popsána v kap. 5.2.1.

Tab. 5.4: Inverzní prokládací matice o rozměrech M = 8, N = 5.

x	x	1	1	1
x	x	–1	–1	–1
x	x	–1	–1	–1
x	x	1	1	1
x	x	1	1	1
x	x	–1	–1	–1
x	x	–1	–1	–1
x	x	1	1	1

Na výstupu prokládací matice dostaneme bitový tok:

$$w = \{\, x\ \ x\ \ 1\ \ 1\ \ 1\ \ x\ \ x\ -1\ -1\ -1\ \ x\ \ x\ -1\ -1\ -1\ \ x\ \ x\ \ 1\ \ 1\ \ 1$$
$$x\ \ x\ 1\ \ 1\ \ 1\ \ x\ \ x\ -1\ -1\ -1\ \ x\ \ x\ -1\ -1\ -1\ \ x\ \ x\ \ 1\ \ 1\ \ 1\,\}.$$

Vodoznak w získáme na základě rovnice (5.17) a (5.18) následujícím způsobem:

$$r_1 = r_4 = r_5 = r_8 = x + x + 1 + 1 + 1 = 3 + 2 \cdot x > 0$$
$$r_2 = r_3 = r_6 = r_7 = x + x - 1 - 1 - 1 = -3 + 2 \cdot x < 0.$$

V tomto případě byl dekódován pouze jeden vodoznak, tudíž vodoznak W tvoří přímo výsledný vodoznak:

$$W = \{\, 1\ -1\ -1\ \ 1\ \ 1\ -1\ -1\ \ 1\,\}.$$

Vodoznak W se shoduje s vloženým vodoznakem (viz kap. 5.2.1). Dekódování vodoznaku proběhlo tedy v pořádku.

6 Robustnost vodoznaku

Jednou z podmínek vodoznačení audia je, aby vložený vodoznak v příslušném audiu neměl vliv na jeho poslechovou kvalitu. Další velmi důležitou podmínkou je robustnost vodoznaku. Tato část diplomové práci popisuje skupinu algoritmů, které slouží k simulaci možných útoků vedených k poškození nebo odstranění vodoznaku z vodoznačných audio dat. Pro test robustnosti jsou vybrány tyto úpravy se signálem: oříznutí, změna vzorkovacího kmitočtu, ztrátová komprese, filtrace, ekvalizace, vložení hudebního efektu a AWGN kanál. Tyto algoritmy byly vytvořeny v programu MATLAB a jeho simulační nadstavbě Simulinku za pomocí skriptů a modelů. V případě ztrátové komprese byl použit volně šiřitelný mp3 enkodér LAME.

Pro test byly vybrány zvukové nahrávky formátu *.wav s vzorkovacím kmitočtem 44100 Hz a bitovou hloubkou 16 bit. Bylo vybráno šest různých žánrů zahrnující klasickou hudbu, pop, rock, disco, swing a country. Všechny zvukové nahrávky pro test mají délku 2 s.

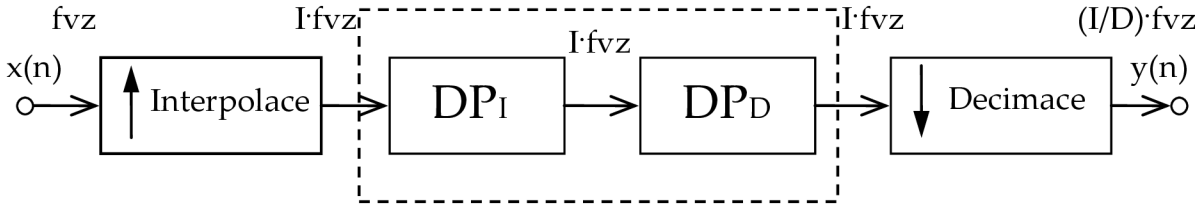
6.1 Oříznutí audio signálu

Jako nejjednodušší způsob útoku na vodoznačené audio se jeví z pohledu běžného uživatele jeho oříznutí. Volně stažitelný software pro práci s auditem umožňuje tyto jednoduché operace provést. Jelikož je vodoznak vkládán do segmentů délky 2048 vzorků, tedy každých 46 ms, tak by neměl mít tento způsob útoku žádný vliv na extrakci vodoznaku. Pro test robustnosti je vodoznačené audio oříznuto na délku 1 s.

6.2 Změna vzorkovacího kmitočtu

Různé systémy můžou používat různé vzorkovací kmitočty. Při záznamu na CD se používá vzorkovací kmitočet 44,1 kHz, v digitálním hudebním studiu se používá vzorkovací kmitočet 48 kHz. V této diplomové práci využiji změnu vzorkovacího kmitočtu pro test robustnosti.

Podvzorkování nebo nadvzorkování v poměru racionálního čísla obecně nelze realizovat prostým výběrem nebo vkládáním odpovídajícího počtu vzorků. Každé racionální číslo lze ale převést na podíl dvou celých čísel I/D. Signál nejprve nadvzorkujeme (interpolujeme) s činitelem I a poté podvzorkujeme (decimujeme) s činitelem D. Operace se signály jsou prováděny v tomto pořadí, protože obráceně bychom zbytečně při podvzorkování odfiltrovali složky, které se při nadvzorkování snažíme získat. Pokud lze nalézt největší společný dělitel čísla I a D, je vhodné jim oba činitele podělit, jinak rostou nároky na strmost přechodového pásma filtrů typu DP naznačených v obr. 6.1.



Obr. 6.1: Změna vzorkovacího kmitočtu v poměru racionálního čísla [14]

Při nadvzorkování chceme zvýšit vzorkovací kmitočet v poměru čísla I. Mezi dva vzorky vstupní posloupnosti je přidáváno $I-1$ nových vzorků. Hodnotu těchto vzorků, které mají být vloženy, neznáme a tudíž doplníme hodnotou nula, neprovádí se žádná interpolace. Jelikož kmitočtové spektrum nadvzorkovaného signálu obsahuje i zperiodizované složky původního signálu, musíme proto použít antialiasingový filtr. Konkrétně DP s mezním kmitočtem $f_m = \frac{f_{vz}}{2}$ [14].

Při podvzorkování jsou ze signálu rovnoměrně ponechány jen některé vzorky. Nový signál má tedy *D*-krát nižší počet vzorků než původní signál. Aby nedošlo k aliasingu, musí signál splňovat vzorkovací teorém i pro nový vzorkovací kmitočet ve tvaru:

$$f_{\max} \leq \frac{f_{vz(y)}}{2}, \tag{6.1}$$

kde $f_{\max}$ je největší kmitočet spektra původního signálu a $f_{vz(y)}$ je snížená hodnota vzorkovacího kmitočtu.

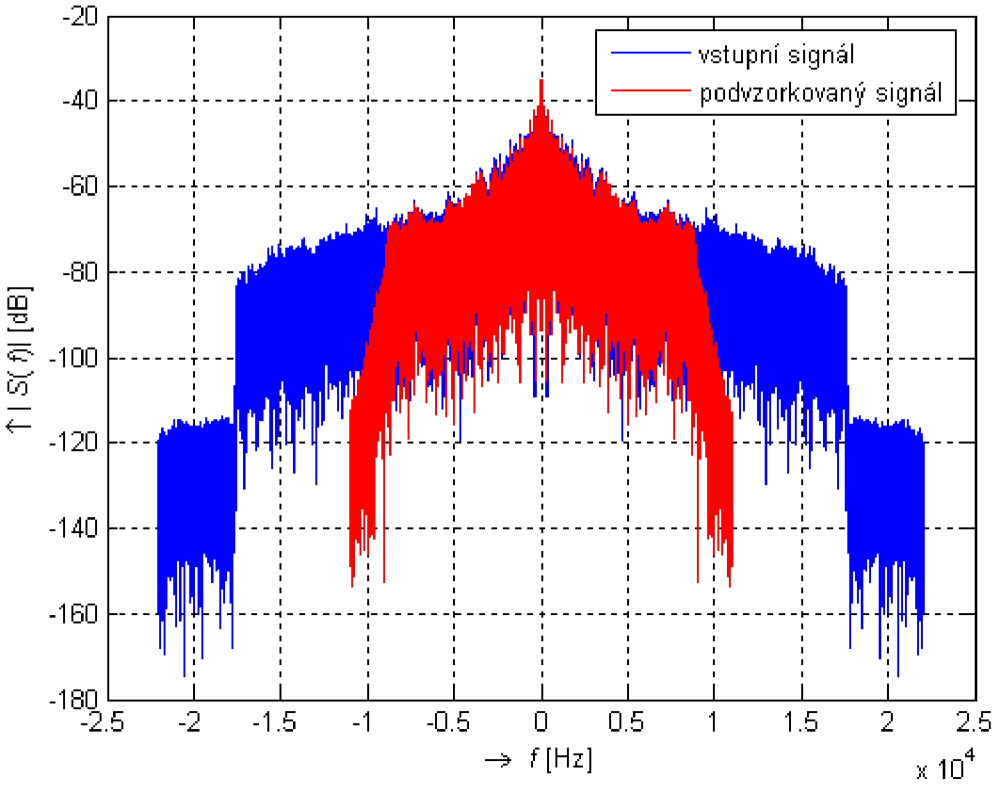
Jelikož nelze předpokládat splnění vzorkovacího teorému, je nutné před samotným podvzorkováním omezit kmitočtové spektrum filtrem typu DP s mezním kmitočtem[14]:

$$f_{cx} = \frac{f_{vz(y)}}{2} = \frac{f_{vz(x)}}{2 \cdot D} = \frac{1}{D}. \tag{6.2}$$

V tomto případě je zbytečné provádět nejprve filtraci interpolačním filtrem a následně antialiasingovým filtrem, ale je možné obě sloučit do jedné DP (naznačeno čárkovaně v obr. 6.1). Mezní kmitočet určíme z rovnice [14]:

$$f_c = \min\left(\frac{1}{I}, \frac{1}{D}\right). \tag{6.3}$$

Na obr. 6.2 vidíme modulové kmitočtové spektrum původního audio signálu v CD kvalitě a podvzorkovaného signálu na 22050 Hz.



Obr. 6.2: Modulové spektrum původního a podvzorkovaného signálu

V testu požiji tyto změny vzorkovacího kmitočtu:

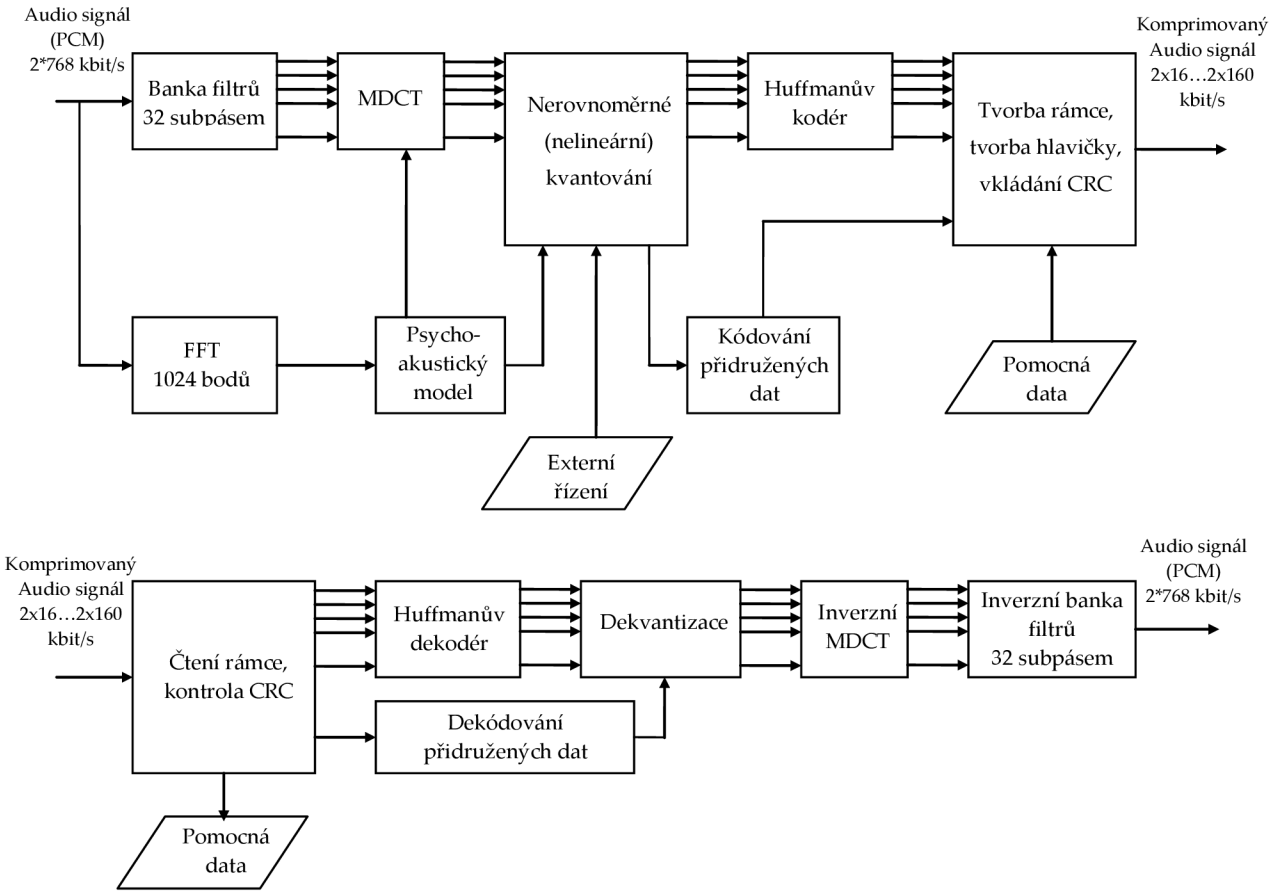
- 44100 Hz → 22050 → 44100 Hz, tedy v poměru ½ (nadvz. / podvz.)
- 44100 Hz → 17640 → 44100 Hz, tedy v poměru ⅖ (nadvz. / podvz.)

6.3 Ztrátová komprese

Jedním z hlavních požadavků spojených s multimediálními daty v minulých letech bylo snížení jejich objemu při archivaci z důvodu malých kapacit datových disků. Takto vznikaly kompresní algoritmy, které toto řešily. V dnešní době jsou použity v podstatě v každém hudebním souboru a filmu na internetu, v digitálním fotoaparátu a digitální televizi. Lze je rozdělit na dvě hlavní skupiny – ztrátové a bezeztrátové. Jedno z nejpopulárnějších a nejpoužívanějších ztrátových kompresí audio signálů je MPEG-1 Audio Layer 3, neboli mp3.

Vstupní PCM (Pulse Code Modulation) signál (každý vzorek signálu je popsán stejným počtem bitů se stejnou přesností) je rozložen bankou filtrů na 32 frekvenčních subpásem. Jelikož tato banka filtrů neposkytuje dostatečné frekvenční rozlišení, je na výstupu doplněna o MDCT (Modified Discrete Cosine Transform). Banka filtrů a MDCT tvoří tzv. hybridní banku filtrů. Pomocí psychoakustického modelu, který popisuje chování lidského ucha, se stanoví hladina maskovaného kvantizačního šumu a přidělí se počet bitů na kvantování každého pásma. Kvantovací obvod je nelineární a navíc je použito Huffmanovo kódování, které snižuje objem datového toku. Pomocí externího řízení volíme

bitovou rychlost. Na výstupu kodéru je komprimovaný zvukový signál s proměnným datovým tokem (VBR), nebo konstantním datovým tokem (CBR) v rozmezí 16–160 kbit/s. V dekodéru se provádí inverzní operace vůči kodéru. Na obr. 6.3 je blokové schéma kodeku pro standard MPEG-1 Audio Layer 3.



Obr. 6.3: Blokové schéma kodéru a dekodéru MPEG-1 Audio Layer 3 [11]

Pro test robustnosti byly provedeny tyto operace:

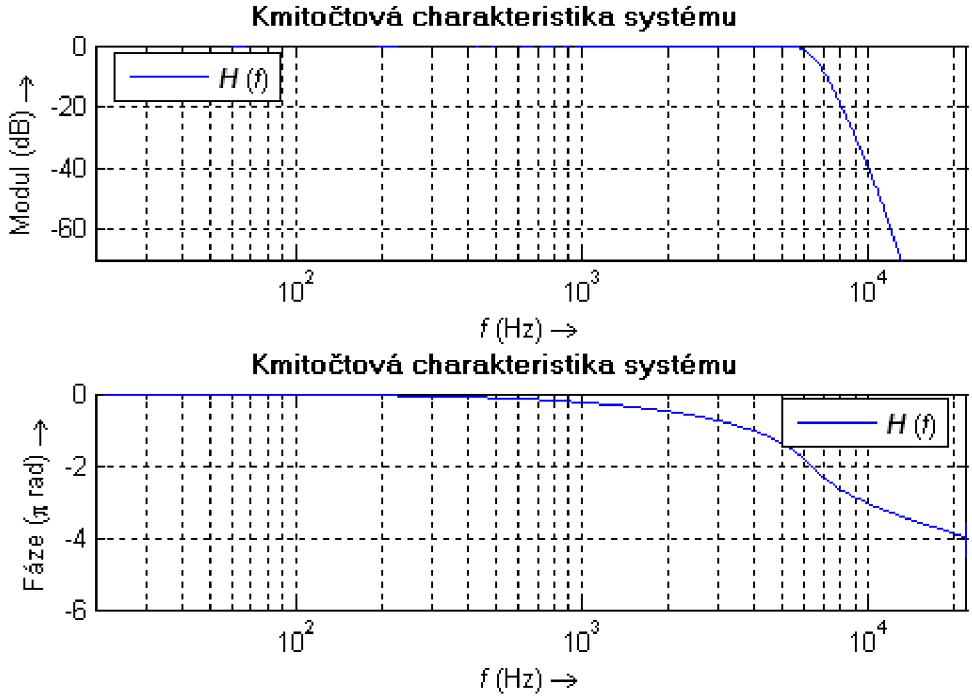
- wav → mp3 (CBR 96 kbit/s) → wav
- wav → mp3 (CBR 128 kbit/s) → wav

6.4 Filtrace

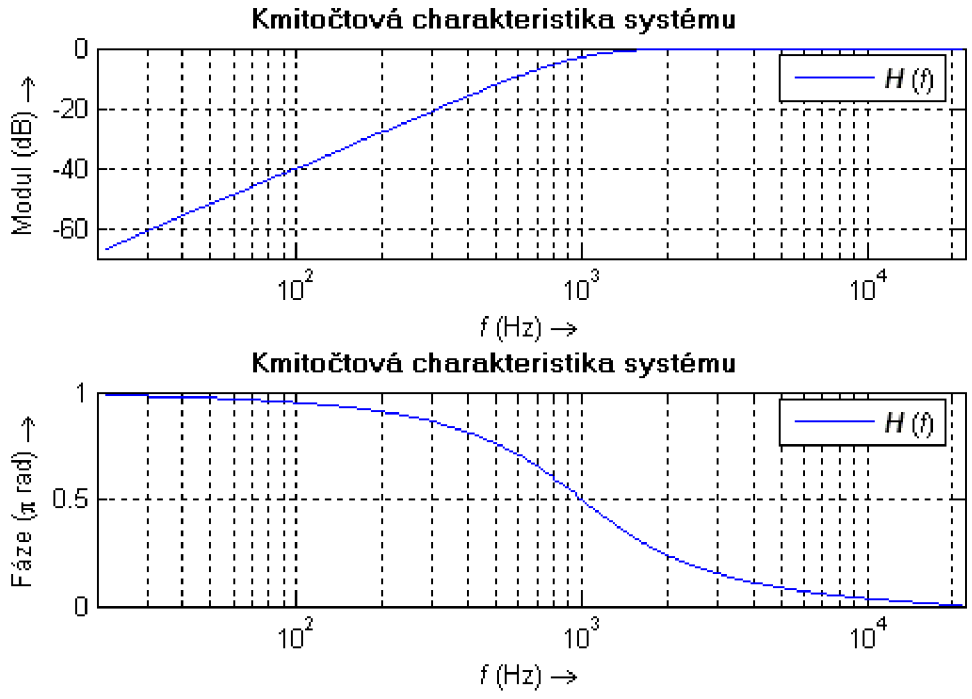
Úkolem číslicového filtru je požadovaným způsobem ovlivnit kmitočtové spektrum vstupního signálu. To znamená buď potlačit složky o určitých kmitočtech, zatímco složky s jinými kmitočty chceme ponechat, nebo tvarovat kmitočtovou charakteristiku, apod. Mezi základní typy filtrů patří: dolní propust (DP), horní propust (HP), pásmová zádrž (PZ) a pásmová propust (PP).

Pro test robustnosti jsem si vybral dva číslicové filtry, a to DP a HP. Oba jsou typu IIR (Infinite Impulse Response). Možný přístup při návrhu IIR filtru je převzetí aproximací používaných při návrhu analogových filtrů, které aproximují ideální požadovanou kmitočtovou charakteristiku. Návrh filtru vychází z tolerančního schématu modulové kmitočtové charakteristiky pro normovanou DP (modul

kmitočtové charakteristiky v propustném pásmu $|H(\omega)|=1$ a modul kmitočtové charakteristiky v nepropustném pásmu $|H(\omega)|=0$). Kmitočtová charakteristika normované HP je zrcadlena vůči DP. Vybral jsem si Butterworthovu aproximaci, jehož modulová kmitočtová charakteristika je monotónně klesající funkce a fázová charakteristika se blíží lineárnímu průběhu. Na obr. 6.4 a 6.5 vidíme kmitočtové charakteristiky použité DP a HP. Podrobnější náhled na návrh číslicových filtrů najdete v literatuře [14].



Obr. 6.4: Kmitočtová a fázová charakteristika DP



Obr. 6.5: Kmitočtová a fázová charakteristika HP

Pro zjištění nejmenšího možného řádu filtru (*Nbutt*) a normovaného mezního kmitočtu při poklesu o 3 dB (*wnbutt*) použiji pro zvolenou aproximaci v prostředí Matlab funkci: $[Nbutt, wnbutt] = buttord(fp, fs, rp, rs)$. Vstupními parametry této funkce jsou hodnoty z tolerančního schématu:

- r_P – maximální zvlnění v propustném pásmu,
- r_S – minimální útlum v nepropustném pásmu,
- f_P – normovaný mezní kmitočet propustného pásma vhledem k $f_{VZ}/2$,
- f_S – normovaný mezní kmitočet nepropustného pásma vhledem k $f_{VZ}/2$.

Koeficienty čitatele a jmenovatele přenosové funkce filtru určíme podle funkce: $[a, b] = butter(Nbutt, wnbutt, typ)$. Pomocí této funkce jsme schopni navrhnout digitální filtr řádu *Nbutt* s Butterworthovu aproximací a normovaným mezním kmitočtem *wnbutt*. Funkce vrátí koeficienty filtru do řádkového vektoru *a* a *b* délky *Nbutt* + 1, pomocí kterých můžeme sestavit přenosovou funkci:

$$H(z) = \frac{B(z)}{A(z)} = \frac{b(1) + b(2).z^{-1} + ... + b(n+1).z^{-Nbutt}}{1 + a(2).z^{-1} + ... + a(n+1).z^{-Nbutt}}. \tag{6.4}$$

Cílem číslicového filtru použitého v této diplomové práci je potlačit složky o určitých kmitočtech. Parametry mezního kmitočtu v propustném pásmu f_{mpp} a mezního kmitočtu v nepropustném pásmu f_{mnp} DP a HP jsou:

- f_{mppDP} = 6000 Hz; f_{mnpDP} = 10000 Hz
- f_{mppHP} = 1000 Hz; f_{mnpHP} = 100 Hz

6.5 Ekvalizace

Z anglického equalization vznikl český název ekvalizace, nebo-li korekce. Ekvalizéry slouží k úpravě frekvenční charakteristiky audio signálu. Základní dělení ekvalizérů je na grafický a parametrický. Velkou výhodou parametrického ekvalizéru je možnost plynulé změny parametrů u jednotlivých filtrů: *fc* (střední kmitočet/mezní kmitočet), G (zesílení/potlačení) a Q (kvalita filtru). Parametrické ekvalizéry mohou být realizovány kaskádním zapojením filtrů typu low frequency shelving, několika filtrů typu peak a filtru typu high frequency shelving. Filtry typu shelving slouží pro váhování určité části kmitočtového spektra nad nebo pod mezním kmitočtem. Parametrický filtr typu peak slouží pro váhování (zesílení nebo potlačení) určité části kmitočtového spektra v okolí středního kmitočtu *fc*. Kmitočtovou charakteristiku peak filtru tak můžeme rozdělit na tři oblasti. Dvě oblasti, kde je přenos 0 dB a prostřední oblast, kde je přenos nastavitelný.

Pro test robustnosti je použit **10-ti pásmový ekvalizér** tvořený filtry typu **Peak filter** (dále jen PF), neboli **oktávový ekvalizér** s parametry uvedenými v tab. 6.1. Celková přenosová funkce parametrického ekvalizéru je dána [13]:

$$H(z) = \prod_{i=1}^{n} H_{PFi}(z), \tag{6.5}$$

kde $H_{PFi}(z)$ je přenosová funkce *i*-tého filtru typu PF.

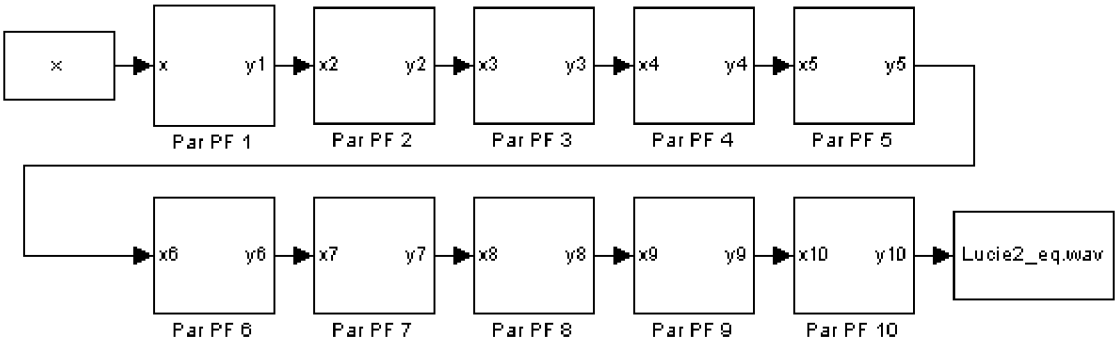
Tab. 6.1: Parametry oktávového ekvalizér [19]

f_C **[Hz]**	31	62	125	250	500	1k	2k	4k	8k	16k
***G* [dB]**	–6	6	–6	6	–6	6	–6	6	–6	6
***Q* [-]**	8	8	8	8	8	8	8	8	8	8

V prostředí Matlab je realizován tento algoritmus pomocí modelu a příslušného skriptu. Ve skriptu, běžícím na popředí v průběhu simulace, jsou definovány počáteční podmínky. Digitální filtr není definován přímo koeficienty filtru, ale parametry na určitých úrovních.

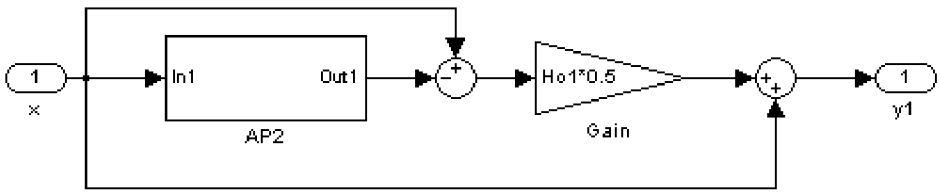
Vstupními parametry každého PF jsou: f_C, G, Q.

Při změně určitého parametru tak nemusíme přepočítat kompletní bilineární transformaci pro získání nových koeficientů filtru. Měníme-li parametr jedné substruktury tak, že je stabilní, bude stabilní i celá struktura. Na obr. 6.6 je zobrazeno kaskádní zapojení deseti parametrických filtrů typu PF tvořící oktávový ekvalizér.



Obr. 6.6: Blokové schéma 10-ti pásmového ekvalizér

Na obr. 6.7 je zobrazeno blokové schéma PF.



Obr. 6.7: Blokové schéma peak filtru [11]

Jedná se o paralelní spojení systému s přenosem 1 a fázovacího článku 2. řádu (AP2) paralelně spojeného se systémem s přenosem 1 násobeným koeficientem Ho1*0.5. U tohoto filtru souvisí parametr kvalita Q se strmostí přechodového pásma a šířkou propustného pásma f_B takto:

$$f_B = \frac{f_c}{Q}. \tag{6.6}$$

Přenosová funkce je dána rovnicí:

$$H(z) = 1 + \frac{H_0}{2} \cdot \left(1 - A_2(z)\right), \tag{6.7}$$

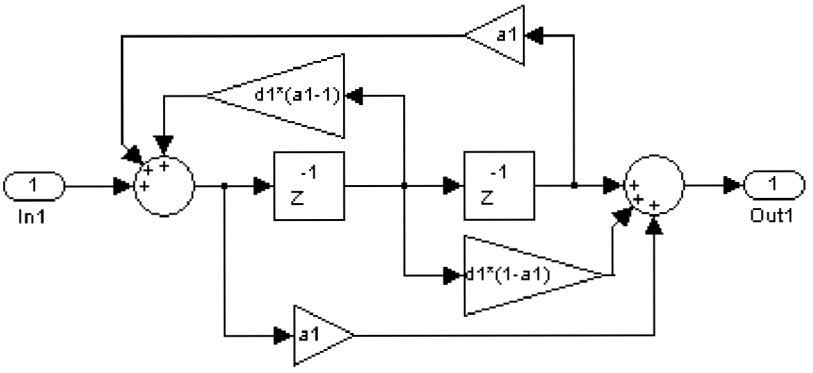
kde H_0 je váhovací koeficient v přenosové funkci vypočtený dle:

$$H_0 = V_0 - 1. \tag{6.8}$$

V_0 je lineární měřítko požadovaného zesílení G, které se zadává v jednotkách dB jako:

$$V_0 = 10^{\frac{G}{20}}. \tag{6.9}$$

Na nejnižší úrovni modelu parametrického filtru je fázovací článek 2. řádu, viz obr. 6.8.



Obr. 6.8: Blokové schéma fázovacího článku 2. řádu [11]

Hodnoty čitatele a jmenovatele přenosové funkce jsou zrcadleny, jak je patrné z rovnice 6.10.

$$A_2(z) = \frac{-a_1 + d1 \cdot \left(1 - a_1\right) \cdot z^{-1} + z^{-2}}{1 + d1 \cdot \left(1 - a_1\right) \cdot z^{-1} - a_1 \cdot z^{-2}} \tag{6.10}$$

Koeficient filtru $a1$ vypočítáme jako [11]:

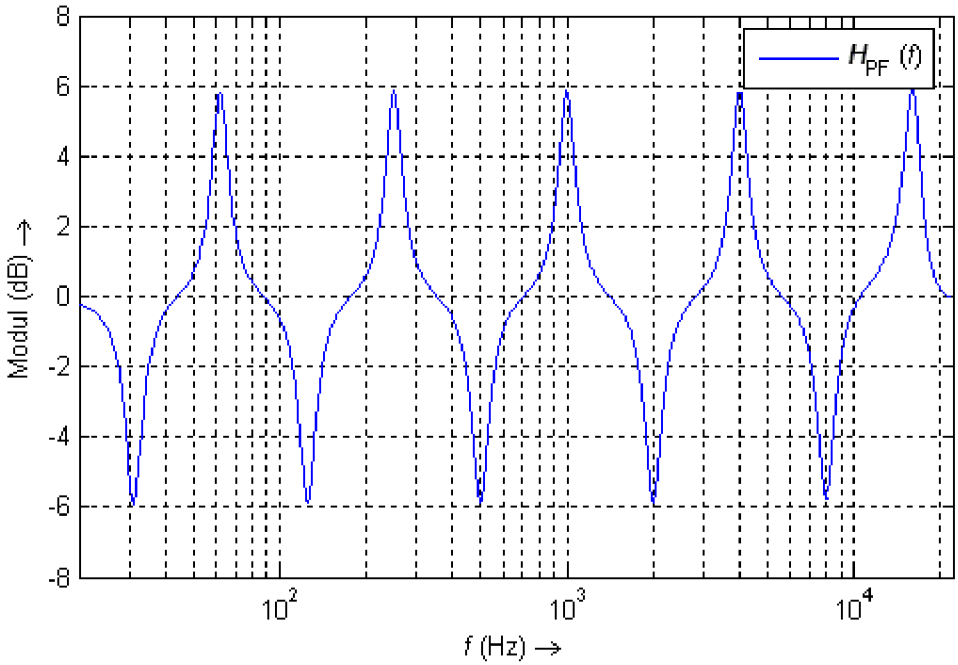
$$a1 = \frac{\tan\left(\pi \cdot f_B / f_{vz}\right) - 1}{\tan\left(\pi \cdot f_B / f_{vz}\right) + 1} \qquad , pro \quad G > 0 \tag{6.11}$$

$$a1 = \frac{\tan(\pi \cdot f_B / f_{vz}) - V_0}{\tan(\pi \cdot f_B / f_{vz}) + V_0} \qquad , pro \quad G < 0 \tag{6.12}$$

Je potřeba rozlišit pro jaké G se hodnoty počítají, jinak by se nám měnit skutečný střední kmitočet f_c. Koeficient filtru $d1$ vypočítáme jako [11]:

$$d1 = -\cos(2 \cdot \pi \cdot f_c / f_{vz}) \tag{6.13}$$

Modulové kmitočtové spektrum ekvalizéru tvořeného PF vidíme na obr. 6.9.



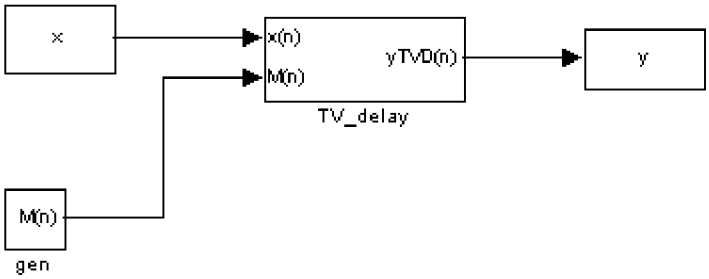
Obr. 6.9: Modulová kmitočtová charakteristika ekvalizéru

6.6 Hudební efekty

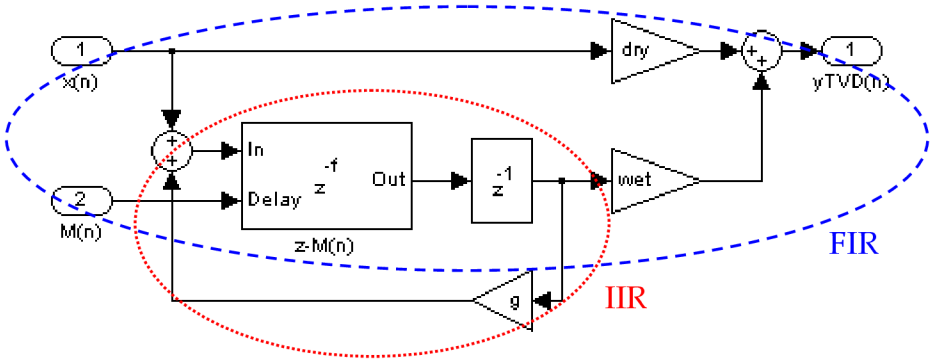
Jednou z dalších možností testování robustnosti vodoznaku je vložení hudebního efektu. Vybral jsem si hudební efekt se zpožďovací linkou, jehož doba zpoždění se periodicky mění. Změna délky zásobníku provádí frekvenční modulaci, proto se těmto efektům také někdy říká modulační efekty. Zpoždění zpožďovací linky je řízeno generátorem, většinou pomocí LFO (Low-Frequency Oscillator) s kmitočtem řádově od 0,1 do 10 Hz. Používají se harmonické, pilové i náhodné průběhy LFO. Nejznámější efekty jsou vibráto, chorus, phaser a flanger. Hudební efekt **flanger** využívám právě v mojí diplomové práci.

Tento algoritmus je vytvořený pomocí skriptu a příslušného modelu v simulinku, protože pouze skript nemá vlastnost na časově variantní systémy. Parametry simulace jsou definovány ve skriptu, který je v popředí a model je v pozadí simulace. Tento efekt není definován vnějším popisem, ale jeho strukturou. Model tak tvoří strukturu a podstrukturu.

V modelu, na obr. 6.10, najdeme blok *gen* s funkcí LFO. Používám sinusový průběh s parametry *m* a *fz*. Dále jsou zde bloky pro vstupní (x) a výstupní (y) vektor dat. Blok *TV_delay* obsahuje blokové schéma efektu flanger, viz obr. 6.11.



Obr. 6.10: Blokové schéma hlavní struktury efektu flanger [11]



Obr. 6.11: Blokové schéma efektu flanger [11]

Efekt flanger tvoří IIR hřebenový filtr vložený do FIR hřebenového filtru. Popis a návrh číslicových filtru typu IIR a FIR lze najít v literatuře [13]. Vstupní signál je v přímé větvi násoben konstantou *dry* a sečtený se signálem zpožděným o M vzorků násobeným konstantou *wet*. Z výstupu zpožďovací linky je na její vstup opět přiváděn zpětnovazební signál násobený konstantou *g*. Okamžitá hodnota výstupu $yTVD(n)$ je dána rovnicí:

$$yTVD(n) = dry \cdot x(n) + wet \cdot x(n - M(n)) + wet \cdot g \cdot x(n - M(n)). \qquad (6.14)$$

Blok z^{-1} ve schématu na obr. 6.11 nám zaručí to, že výstupní hodnoty budou v násobcích vzorkovacího kmitočtu. Zpoždění je závislé na proměnné M(n).

Parametry simulace digitálního hudebního efektu *flanger* jsou nastaveny takto:

- Ms = 200 vzorků (střední doba zpoždění)
- dry = 0,5 (přímá cesta)
- wet = 0,3 (efektová cesta)
- g = 0,5 (váhování zpětné vazby)
- m = 0,5 (hloubka modulace)
- fz = 0,7 (kmitočet změny)

6.7 AWGN kanál

Rušivé signály se objevují, nebo vznikají ve všech částech sdělovací soustavy. V praxi si nahrazujeme všechny zdroje rušení jediným zdrojem. Pokud má rušení vlastnosti blízké bílému šumu s normálním rozdělením, mluvíme potom o kanále AWGN (Additive white Gaussian noise). Signál $s(t)$ narušený aditivním bílým šumem je popsán rovnicí:

$$x(t) = s(t) + n(t), \tag{6.15}$$

kde $s(t)$ je užitečná složka signálu a $n(t)$ je rušivá složka signálu.

Rušivý signál tedy pro jednoduchost přičítáme přímo k užitečnému signálu. Hodnoty rušivého signálu se řídí normálním (Gaussovým) rozdělením s parametry μ a $\sigma^2 > 0$ [15]:

$$p(n) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \exp\left(- \frac{(x-\mu)^2}{2 \cdot \sigma^2} \right), \tag{6.16}$$

kde μ je střední hodnota a σ^2 je rozptyl.

Jedná se tedy o náhodný signál s rovnoměrnou výkonovou spektrální hustotou. V prostředí Matlab jsem použil pro simulaci AWGN kanálu funkci: $y = awgn(x, snr, 'measured')$, která přidá bílý šum ke vstupnímu vektoru x, číslo snr specifikuje výsledný odstup signál-šum v dB.

6.8 Shrnutí výsledků

Důležitým požadavkem kladeným na vodoznačící algoritmy je, aby měli vysokou odolnost proti poškození vloženého vodoznaku. Poškození může být jak úmyslné, tak k němu může dojít i při běžném zpracování audio signálů. Odolnost vodoznaku proti poškození lze stanovit objektivně při extrakci vodoznaku. Užívá se k tomu algoritmus BER (Bit Error Rate). BER je definován [19]:

$$BER = \frac{100}{N} \sum_{n=1}^{N} \begin{cases} 1, \; je\text{-}li \; \tilde{w}_n \neq w_n \\ 0, \; je\text{-}li \; \tilde{w}_n = w_n \end{cases}, \tag{6.17}$$

kde: N je délka vodoznaku,

$\tilde{w}_n$ je n-tý bit extrahovaného vodoznaku,

w_n je n-tý bit vkládaného vodoznaku.

Na vodoznačený audio soubor byly aplikovány algoritmy (popsané v kap. 6.1 až 6.7) simulující útok vedený k poškození vodoznaku. Tyto základní útoky a hodnoty BER pro vybrané žánry audia jsou uvedeny v tab. 6.2 a 6.3.

Tab. 6.2: Test robustnosti vloženého vodoznaku pro $\alpha = 0.0$

ÚTOKY NA AUDIO			Klasická hudba	Pop	Rock	Country	Disco	Swing	PRŮMĚR
			BER (%)						
Žádné		–	18,75	12,40	14,38	12,50	15,18	10,31	13,92
Oříznutí audia na délku		0:01 s	23,28	19,28	11,11	13,75	16,67	17,36	16,90
Změna vzorkovacího kmitočtu		22050 Hz	53,21	49,23	51,62	48,26	51,92	50,81	50,84
		17640 Hz	51,04	50,94	50,23	47,18	52,08	51,10	50,43
Ztrátová komprese – mp3		96 kbit/s	38,15	49,69	46,06	46,53	45,35	51,74	46,25
		128 kbit/s	32,86	42,32	46,99	41,55	45,03	44,52	42,21
Filtrace	DP	f_m = 6345 Hz	49,08	48,85	48,84	49,02	49,68	48,90	49,06
	HP	f_m = 990 Hz	20,06	19,48	26,62	23,15	30,21	16,55	22,67
Ekvalizace	PF	10-ti pásm.	21,37	32,71	41,44	40,28	40,92	34,65	35,23
Hudební efekty – flanger		m = 0,5 fz = 0,7 Hz Ms = 200	22,86	36,25	40,28	41,09	43,75	34,72	36,49
AWGN		SNR = 25 dB	49,63	49,78	48,84	53,06	48,08	51,32	50,12
		SNR = 35 dB	38,73	45,94	48,84	46,76	51,44	50,33	47,00
PRŮMĚR			34,92	38,07	39,60	38,59	40,86	38,52	

Tab. 6.3: Test robustnosti vloženého vodoznaku pro $\alpha = 0.9$

ÚTOKY NA AUDIO			Klasická hudba	Pop	Rock	Country	Disco	Swing	PRŮMĚR
			BER (%)						
Žádné		–	16,04	15,72	17,58	17,73	16,12	18,01	16,87
Oříznutí audia na délku		0:01 s	16,20	16,15	17,13	17,88	16,90	18,52	17,13
Změna vzorkovacího kmitočtu		22050 Hz	49,32	50,67	49,57	51,27	49,05	48,92	49,80
		17640 Hz	49,74	51,13	49,51	51,34	50,23	49,68	50,27
Ztrátová komprese – mp3		96 kbit/s	30,79	34,33	40,44	32,36	40,95	29,85	34,78
		128 kbit/s	28,28	30,14	36,25	28,74	34,39	26,94	30,79
Filtrace	DP	f_m = 6345 Hz	50,00	49,28	50,23	50,43	50,28	47,60	49,64
	HP	f_m = 990 Hz	24,37	32,82	34,22	30,96	32,73	28,91	30,67
Ekvalizace	PF	10-ti pásm.	29,97	34,50	39,81	31,42	40,18	31,22	34,52
Hudební efekty – flanger		m = 0,5 fz = 0,7 Hz Ms = 200	30,54	41,15	44,86	39,27	43,81	37,39	39,50
AWGN		SNR = 25 dB	37,10	30,57	39,95	34,73	38,76	34,05	35,86
		SNR = 35 dB	33,88	22,80	27,97	28,91	29,48	26,20	28,21
PRŮMĚR			33,02	34,10	37,29	34,58	36,91	33,11	

Hodnoty BER v tab. 6.2 a 6.3 označené zeleně značí, že se extrahovaný vodoznak shodoval s vloženým vodoznakem. V ostatních případech se vodoznak částečně, nebo zcela lišil od vloženého vodoznaku.

Výpočetní náročnost:

Výpočetní náročnost může tvořit jedno z dalších kritérií při pohledu na vodoznačící algoritmy. Tento parametr je velmi důležitý pro aplikace pracující v reálném čase. Implementovaný algoritmus slouží pro vložení vodoznaku z důvodu ochrany autorských práv, kde zpracování v reálném čase nehraje velkou roli. Výpočetní náročnost je tedy doba potřebná pro vložení $t_{vlož}$ a extrakci t_{extra} vodoznaku. Pro srovnání byly vybrány následující tři případy:

Audio soubor délky 1 s, vodoznak „HEITEL“ a α = 0,0:

- $t_{vlož}$ = 43,5 s
- t_{extra} = 152,9 s

Audio soubor délky 10 s, vodoznak „HEITEL“ a α = 0,0:

- $t_{vlož}$ = 2280,1 s
- t_{extra} = 3757,4 s

Audio soubor délky 10 s, vodoznak „HEITEL“ a α = 0,2:

- $t_{vlož}$ = 2360,5 s
- t_{extra} = 9530,7 s

7 Transparentnost vodoznaku

Jedním ze základních požadavků kladených na metody digitálního vodoznační audio signálů je vkládat vodoznak do audio dat takovým způsobem, aby byl dostatečně odolný proti poškození běžnými metodami zpracování audio signálu. Určuje se tedy tzv. robustnost vodoznaku. Přitom ale nesmí dojít k narušení původní kvality audio signálu. Stanovuje se tzv. **transparentnost vodoznaku**, tedy míra nevnímatelnosti vloženého vodoznaku. Existují objektivní a subjektivní metody vyhodnocení, které jsou podrobněji popsány v této kapitole.

7.1 Subjektivní metody

Mezi subjektivní metody patří tzv. poslechové testy. Pro ně musíme vybrat posluchače, stanovit jasné pokyny k provádění testů a ty následně zpracovat vhodnými statistickými metodami. Mezi ně patří například ABX test, trojúhelníkový test, duo-trio test, nebo metoda testování dle normy ITU-R BS.1116 a stupnice hodnocení dle ITU-R BS.562 [19].

Počet posluchačů, kteří se aktivně účastnili poslechových testů je **10**, počet opakování je **1**. K realizaci a vyhodnocení testu jsem použil software ABC/HR stažitelný z [22]. Program umožní realizovat **kombinovanou metodu** subjektivního psychoakustického měření. Jedná se tedy o **metodu podobnosti zvukových podnětů** a **metodu posuzování**. Konkrétně tento software umožňuje srovnání audio souborů **ABX dvojitým slepým testem** a stanovení **míry degradace kvality** u vodoznačeného audia dle ITU-R BS.562. Seznam referenčních originálních audio souborů je uveden v tab. 7.1. Do těchto audio souborů byl vložen vodoznak „**HEITEL**“ s koeficientem intenzity vodoznaku α **= 0,0** a α **= 0,2**.

Tab. 7.1: Seznam použitých audio souborů pro test transparentnosti

Číslo audia	Interpret	Žánr
1	Wolfgang Amadeus Mozart - Symphony no. 40	Klasika
2	Kabát - Dole v dole	Rock
3	Roxette - How Do You Do	Pop
4	Pavel Bobek - Dokud budu žít	Country
5	Global Deejays - What a Feeling (Flashdance)	Disco
6	Big Band play along - In The Mood	Swing

ABX dvojitý slepý test je realizován následovně: Posluchači mají přístup ke třem nahrávkám s označením A, B a X. Nahrávky s označením A a B jsou referenční. **A** je **originální** audio soubor, zatímco **B** je **vodoznačený** audio soubor. **X** představuje nahrávku **náhodně** vybranou z A a B. V průběhu testu se posluchač rozhoduje, zda nahrávka X odpovídá nahrávce A nebo B. Jelikož může odpověď od posluchače nabývat pouze dvou možností (X odpovídá A, X odpovídá B), říká se tomuto testu „Bernoulliho test“. U tohoto testu o n pokusech je důležité,

aby odpovědi od všech pokusů byly navzájem nezávislé. Potom je možno stanovit pravděpodobnost špatného rozhodnutí v poslechovém testu a eliminovat tak statistickou chybu testu [19].

Definujme statistickou hypotézu $\mathbf{H_0}$ a $\mathbf{H_1}$:

$\mathbf{H_0}$: Rozdíl mezi originálním a vodoznačeným audio souborem není slyšitelný.

$\mathbf{H_1}$: Rozdíl mezi originálním a vodoznačeným audio souborem je slyšitelný.

Nelze zaručit bezchybnost rozhodnutí hypotéz H_0 a H_1, proto je nutné definovat chyby:

Chyba 1. druhu (hladina významnosti α), když zamítneme platnou hypotézu H_0.

Chyba 2. druhu (chyba β), když nezamítneme neplatnou hypotéza H_0.

Čím větší je chyba 2. druhu (chyba 1. druhu se zmenšuje), tím menší je schopnost statistického testu odhalit přítomnost skutečně existujících rozdílů – tzv. síla testu [10]. Před testem je tak potřeba stanovit hladinu významnosti α, statistickou chybu β, celkový počet hodnocení n a práh T ($0 \leq T \leq n$). Podrobný popis, výpočet a odvození parametrů potřebných pro tento test najdeme v knize [3]. (Práh T je v programu [22] definován nepřímo pomocí parametru *theta*, jako poměr požadovaných správně identifikovaných pokusů (práh T) k celkovému počtu pokusů n.

Při vyhodnocení testu se porovnává hodnota správně identifikovaných nahrávek k s prahem T a výsledek lze následovně interpretovat:

$$\begin{aligned} &\text{je-li } k < T \text{ , potom je vodoznak neslyšitelný,} \\ &\text{je-li } k \geq T \text{ , potom je vodoznak slyšitelný.} \end{aligned} \tag{7.1}$$

Při přípravě, realizaci a vyhodnocení poslechových testů z oblasti experimentální psychoakustiky doporučuji postupovat dle knihy [10]. Metodika tohoto testu je uvedena v **příloze B**. Podle této přílohy postupovali posluchači při testech. Výsledky ABX testu pomocí programu ABC/HR pro zvoleného jednoho posluchače jsou uvedeny v tab. 7.2 a 7.3.

Tab. 7.2: Výsledky ABX testu pro $\alpha = 0{,}0$

Číslo audia	**Počet správných identifikací** k	**Práh** T	**Celkový počet hodnocení** n	**Pravděpodobnost nesprávného tvrzení: „A je rozdílné od B“**	**Je vložený vodoznak transparentní?**
1	5	7	8	0,363	ANO
2	5	7	8	0,363	ANO
3	8	7	8	0,004	NE
4	3	7	8	0,855	ANO
5	3	7	8	0,855	ANO
6	4	7	8	0,637	ANO

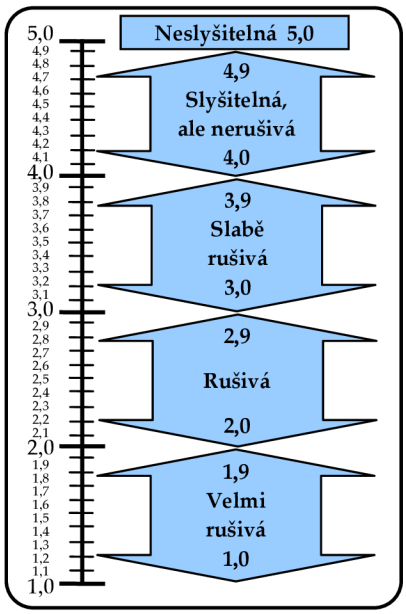
Tab. 7.3: Výsledky ABX testu pro $\alpha = 0{,}2$

Číslo audia	Počet správných identifikací k	Práh T	Celkový počet hodnocení n	Pravděpodobnost nesprávného tvrzení: „A je rozdílné od B“	Je vložený vodoznak transparentní?
1	8	7	8	0,004	NE
2	7	7	8	0,035	NE
3	8	7	8	0,004	NE
4	7	7	8	0,035	NE
5	3	7	8	0,855	ANO
6	6	7	8	0,145	ANO

Pro všechny zvolené audio soubory je uveden počet správně identifikovaných nahrávek X, práh T a celkový počet hodnocení n. Celkový počet hodnocení je dán přednastaveným režimem **Moderate difference** (mírný rozdíl). Vstupní parametry testu byly stanoveny následovně: *alpha* = 0,05, *beta* = 0,2 a *theta* = 0,90 (vyplívá z: $n = 8$ a $T = 7$). Pátý sloupec představuje statistické vyhodnocení nesprávného tvrzení hypotézy H_1. V posledním sloupci je vysloven verdikt, zda je transparentnost vloženého vodoznaku dostačující, nebo nedostačující (vodoznak je slyšitelný). Je rozhodnuto na základě **rovnice 7.1**.

Výsledky od všech deseti posluchačů jsou zaneseny přehledně do tabulek a uvedeny v příloze C.

Součástí programu ABC/HR je i hodnocení míry degradace kvality vodoznačené nahrávky vůči originální nahrávce dle normy ITU-R. Metodika tohoto testu je uvedena v **příloze B**. Stupnice hodnocení dle této normy je na obr. 7.1. Jednotlivé hodnocení pro vybrané audio soubory vidíme v tab. 7.4 a 7.5, kde je i vyhodnocení všech výsledků (viz příloha C) první části testu z programu ABC/HR a je vysloven verdikt, zda je vodoznak neslyšitelný – transparentní (ANO), nebo slyšitelný (NE).



Obr. 7.1: Stupnice míry degradace kvality audio nahrávky dle normy ITU-R BS.562

Tab. 7.4: Vyhodnocení výsledků z programu ABC/HR pro $\alpha = 0{,}0$

Posluchač	Číslo audia					
	1	2	3	4	5	6
P. Rajmic	5,0/ANO	5,0/ANO	4,6/NE	5,0/ANO	5,0/ANO	5,0/ANO
P. Sysel	4,4/NE	5,0/ANO	5,0/ANO	5,0/ANO	4,0/NE	5,0/ANO
Z. Průša	2,9/NE	5,0/ANO	4,4/NE	4,7/NE	4,2/NE	5,0/ANO
J. Schimmel	1,4/NE	4,4/NE	3,7/NE	5,0/ANO	4,2/NE	5,0/ANO
I. Míča	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO
P. Šilhavý	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO
R. Číž	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO
R. Beneš	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO
J. Šporik	3,9/NE	5,0/ANO	3,8/NE	4,7/NE	5,0/ANO	5,0/ANO
J. Karásek	1,0/NE	2,1/NE	1,5/NE	3,9/NE	5,0/ANO	5,0/ANO
PRŮMĚR	3,9	4,7	4,3	4,8	4,7	5,0

Tab. 7.5: Vyhodnocení výsledků z programu ABC/HR pro $\alpha = 0{,}2$

Posluchač	Číslo audia					
	1	2	3	4	5	6
P. Rajmic	2,7/NE	4,7/NE	3,7/NE	4,8/NE	5,0/ANO	5,0/ANO
P. Sysel	2,5/NE	4,5/NE	4,5/NE	5,0/ANO	2,5/NE	5,0/ANO
Z. Průša	2,7/NE	5,0/ANO	4,2/NE	4,1/NE	3,8/NE	5,0/ANO
J. Schimmel	1,0/NE	4,2/NE	2,1/NE	5,0/ANO	2,9/NE	5,0/ANO
I. Míča	2,0/NE	5,0/ANO	3,0/NE	3,5/NE	2,5/NE	5,0/ANO
P. Šilhavý	2,1/NE	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO	5,0/ANO
R. Číž	3,0/NE	5,0/ANO	3,5/NE	3,9/NE	3,7/NE	5,0/ANO
R. Beneš	3,0/NE	5,0/ANO	3,8/NE	5,0/ANO	3,7/NE	5,0/ANO
J. Šporik	2,0/NE	5,0/ANO	3,8/NE	4,5/NE	3,9/NE	4,2/NE
J. Karásek	1,0/NE	2,1/NE	1,3/NE	1,8/NE	1,3/NE	3,8/NE
PRŮMĚR	2,2	4,5	3,5	4,3	3,4	4,8

U obou testů se jedná o tzv. **částečně řízený test**. Uživatel tak má možnost volit délku svého rozhodování, oblast přehrávání, nebo např. opětovné přehrání nahrávek. Výhodou tohoto testu je, že posluchač může pružně přizpůsobit experimentální proceduru jeho momentální situaci, metodika viz příloha B.

7.2 Objektivní metody

Cílem objektivních metod je nahrazení člověka autonomním systémem. Pro širokopásmové audio nahrávky existují např. metody NMR (Noise to Mask Ratio), nebo PEAQ (Perceptual Evaluation of Audio Quality). Úkolem této diplomové práce ale není implementovat tuto sofistikovanou metodu vyhodnocení kvality audio signálů. V tomto případě jsem použil výpočet SNR (Signal to Noise Ratio), který patří mezi diferenční metriky. SNR slouží k výpočtu diference mezi nezkresleným (originálním) audio signálem a zkresleným (vodoznačeným) audio signálem. V praxi se většinou udává v jednotkách decibel (dB) a definuje se jako [19]:

$$SNR[dB] = 10\log_{10}\left[\frac{\sum_n x_n^{\,2}}{\sum_n (\tilde{x}_n - x_n)^2}\right], \tag{7.2}$$

kde: x_n je n-tý vzorek originálního signálu,

$\tilde{x}_n$ je n-tý vzorek zašumělého (vodoznačeného) signálu.

Větší hodnota SNR nám prozradí to, že se vodoznačené audio blíží více originálnímu audiu. Výsledky měření pro různé hodnoty intenzity vodoznaku (α) jsou shrnuty v tab. 7.6. Jedná se podstatě o průměrnou hodnotu SNR přes celou délku audia. Výsledky tedy nejsou zcela vypovídající, protože energie audio signálů je velmi proměnlivá s časem. Pokud se hodnota SNR pohybuje přibližně nad 35 dB, lze považovat vodoznak za transparentní.

Tab. 7.6: Výsledky objektivního testu transparentnosti vodoznaku

Číslo audia		**1**	**2**	**3**	**4**	**5**	**6**	**Průměr**
α = 0,0	**SNR (dB)**	40,54	50,63	41,58	40,26	49,04	41,09	43,86
α = 0,2		33,40	44,91	37,49	37,01	43,86	37,84	39,08
α = 1,0		20,30	32,23	25,53	25,65	31,39	26,45	26,93

Byl použit počítač s konfigurací:

- CPU: Intel Pentium M; 1,5 GHz
- Operační paměť: 1 GB
- Operační systém: Microsoft Windows XP Home Edition
- Matlab: Version 7.0.1 (R14); SP 1
- Signal Processing Toolbox: Version 6.2.1

8 Uživatelské rozhraní v prostředí MATLAB

Pro zpracování a analýzu audio signálu využívám vývojové prostředí MATLAB (více se můžete dozvědět o zpracování signálu pomocí tohoto programu v literatuře [7]) a především knihovnu funkcí Signal Processing Toolbox. Je to knihovna, která významně rozšiřuje základní možnosti MATLABu ve vztahu k práci se signály. Níže je popsán návod k použití programu pro vložení a extrakci vodoznaku.

Pro správnou funkci algoritmu je zapotřebí mít v jedné složce uloženy všechny m-file, které využívá algoritmus pro vložení a extrakci vodoznaku. Hierarchie zdrojových souborů je uvedena v příloze A. Po přeložení (Debug) skriptu *spousteni.m* je uživateli do Command Window zobrazena možnost jak dále pokračovat. Úvodní nabídka je rozdělena do dvou částí, jak je patrné z obr. 8.1, a to pro vkládání a extrakci vodoznaku.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%  Pro spuštění programu pro vložení vodoznaku zadejte do Command Window:   %%
%%                              run wat_enc                                 %%
%%        pokud chcete zobrazit NÁPOVĚDU, zadejte do Command Window:        %%
%%                              help wat_enc                                %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%  Pro spuštění programu pro extrakci vodoznaku zadejte do Command Window:  %%
%%                              run wat_dec                                 %%
%%        pokud chcete zobrazit NÁPOVĚDU, zadejte do Command Window:        %%
%%                              help wat_dec                                %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

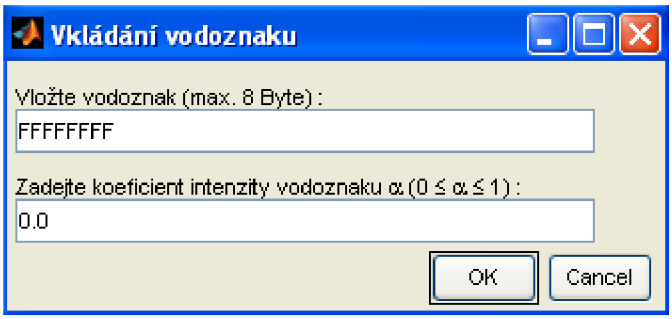
Obr. 8.1: Úvodní nabídka po přeložení skriptu spousteni.m

Pokud si uživatel není jistý, jak postupovat při vkládání vodoznaku, může si zobrazit nápovědu zadáním: **help wat_enc**. Zadáním **run wat_enc** se spustí samotný algoritmus pro vkládání vodoznaku, viz kapitola 8.1. Program Vás jednoduchými kroky vede k cíli. Obdobně funguje část pro extrakci vodoznaku, viz kapitola 8.2.

8.1 Vložení vodoznaku

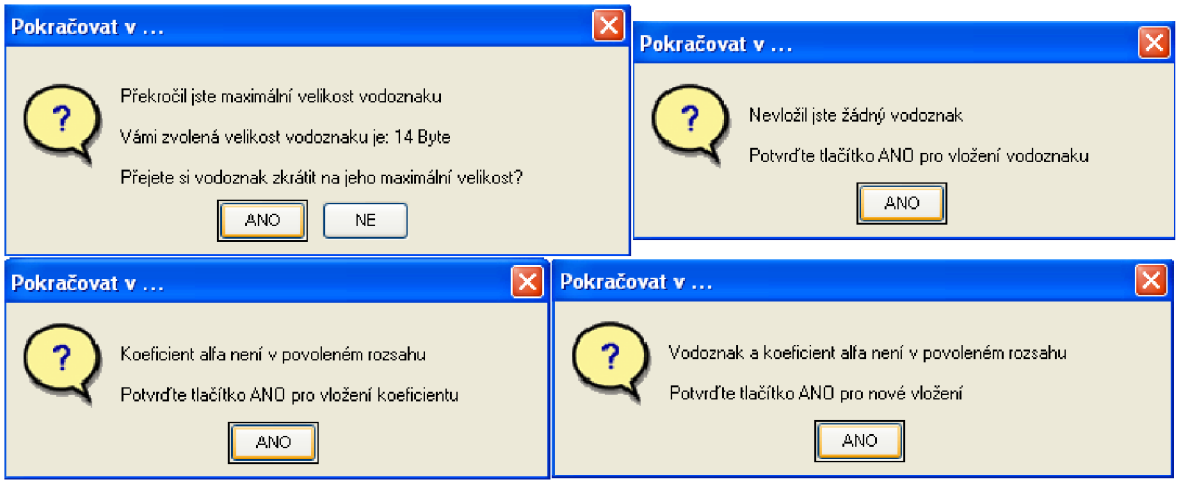
Uživatel je nejprve vyzván k načtení **originálního** audio souboru zmáčknutím libovolného tlačítka na klávesnici. Je podporován soubor typu ***.wav** v **CD** kvalitě s **jedním** (mono) i **dvěma** (stereo) kanály. Ze stereo nahrávky je vybrán pouze **jeden** kanál, do kterého se vkládá vodoznak. Při načítání audia je kontrolován vzorkovací kmitočet. Pokud je otevřen audio soubor v jiné než CD kvalitě, jste o tom informováni prostřednictvím dialogového okna. Potvrzením tlačítka **ANO** se Vám opět zobrazí okno pro načtení audia. Základní informace jako délka audio nahrávky, počet kanálů, vzorkovací kmitočet, bitová hloubka a vybraný audio

soubor se vypíší do Command Window prostředí MATLAB. Hned poté se zobrazí dialogové okno pro vložení vodoznaku, viz obr. 8.2.



Obr. 8.2: Dialogové okno pro vkládání vodoznaku

Do prvního řádku zadáváte vodoznak délky **1–8 Byte**. Vodoznak se zadává ve znacích. Vybírat můžete z kompletní sady **tisknutelných** znaků, které jsou uvedeny v ASCII tabulce (doporučeno: A–Z, a–z, 0–9). **Nelze** tedy zvolit znaky **s diakritikou**. Implicitně je zvoleno „FFFFFFFF“. Do druhého řádku zadáváte koeficient intenzity vodoznaku (α) v rozsahu **0–1**. Implicitně je nastavena hodnota 0,0. Pokud si nejste jisti, jak může parametr α ovlivnit vkládání vodoznaku, je nutné nahlédnout do kap. 5.3.2. Při vkládání vodoznaku a parametru alfa je kontrolován jejich povolený rozsah. Pokud se pohybujete mimo tento rozsah, jste o tom informováni pomocí dialogových oken, viz obr. 8.3.



Obr. 8.3: Ošetření velikosti vodoznaku a koeficientu α při jeho zadávání

V případech, kdy můžete odpovědět pouze **ANO**, se po potvrzení tohoto tlačítka zachová ten z Vámi zadaných parametrů, který byl v povoleném rozsahu. Ten z nepovoleného rozsahu je zpět nastaven na implicitní hodnotu. Pouze v případě překročení maximální velikosti vodoznaku máte na výběr, zda chcete zkrátit vodoznak na jeho **maximální** velikost (8 Byte) potvrzením tlačítkem ANO, nebo zobrazit dialogové okno pro nové vložení vodoznaku potvrzením tlačítka NE.

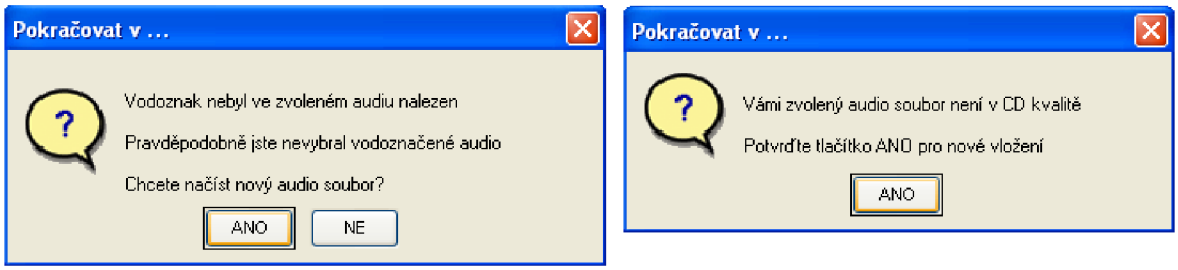
Potvrzením tlačítka **OK** se vodoznak vloží do originálního audio souboru. O průběhu vkládání vodoznaku jste informováni v Command Window. Po určité

časové prodlevě (doba potřebná pro vložení vodoznaku se zobrazí do Command Window) je zobrazeno dialogové okno pro uložení **vodoznačeného audia** opět typu *.wav s **jedním** kanálem (mono). Implicitní název audia tvoří název **orig. audia** doplněné o „**_waterm**“ jenž označuje vodoznačené audio. Název souboru můžete ale zvolit libovolný.

Po uložení **vodoznačeného audio souboru nesmíte** již tento audio soubor **přejmenovat**. Při extrakci vodoznaku z audia jsou zapotřebí určité pomocné soubory, jejichž název souvisí s názvem vodoznačeného audio souboru při jeho ukládání. Toto opatření zajišťuje uživateli při extrakci vodoznaku větší komfort, jinak by musel např. pokaždé hledat soubor uložený někde na disku, ve kterém by byly pomocné informace pro extrakci vodoznaku. Na druhou stranu je zapotřebí mít uložené pomocné soubory (*_“název vodoznačeného audio souboru“.mat) vždy ve stejné složce, jako je m-file pro extrakci vodoznaku. Název souboru by měl být **jedinečný** i vhledem ke všem ostatním vodoznačeným audio souborům uložených v adresářích od nejnižšího směrem ke kořenovému adresáři.

8.2 Extrakce vodoznaku

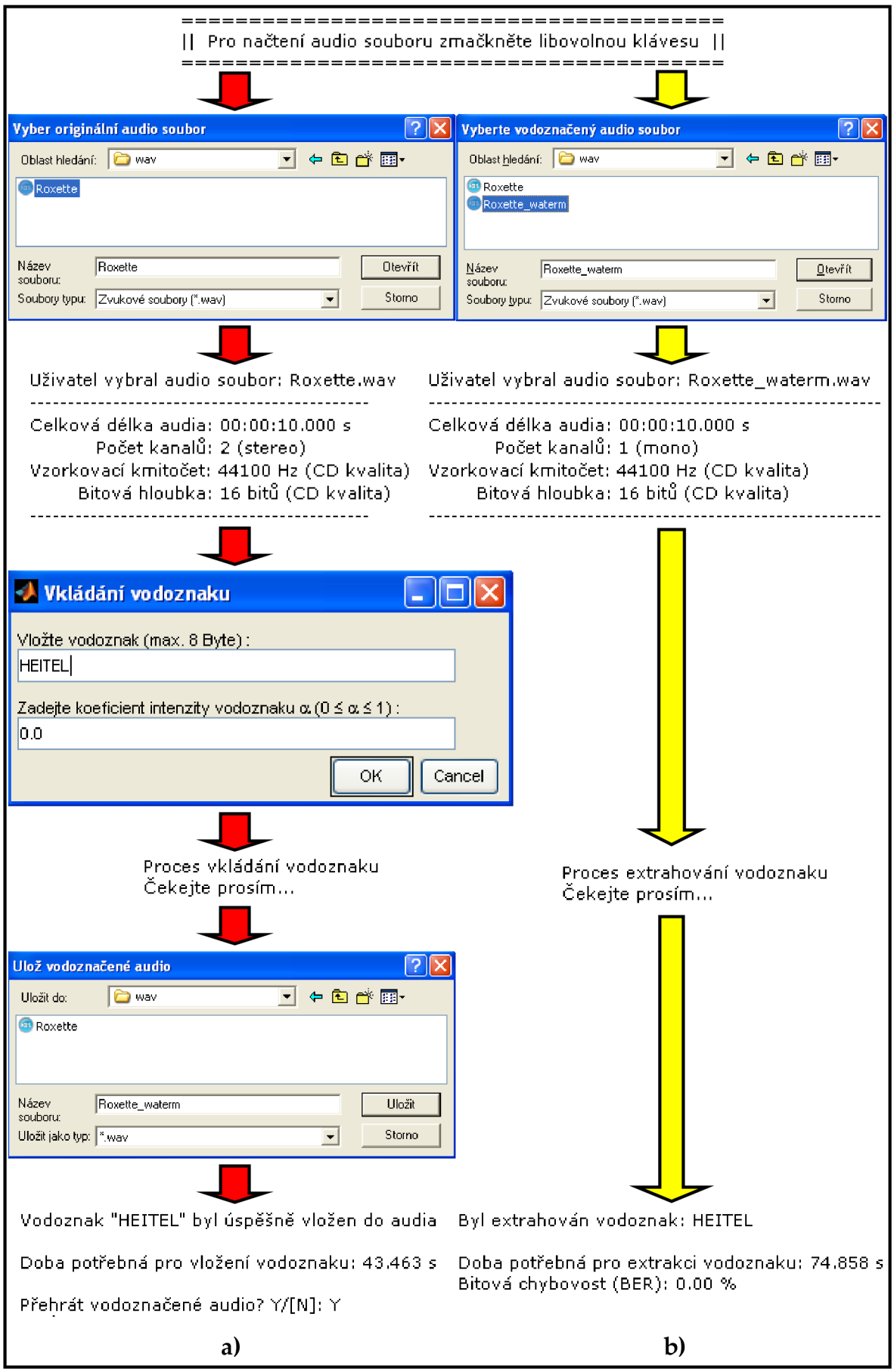
Uživatel je nejprve vyzván k načtení **(vodoznačeného)** audio souboru zmáčknutím libovolného tlačítka na klávesnici. Je podporován soubor typu ***.wav** v **CD** kvalitě a **jedním** kanálem (mono). Při načítání audia je kontrolován vzorkovací kmitočet. Pokud je otevřen audio soubor v jiné než CD kvalitě, jste o tom informováni prostřednictvím dialogového okna, viz obr. 8.4. Potvrzením tlačítka **ANO** se Vám zobrazí okno pro opětovné načtení audia. Základní informace jako vzorkovací kmitočet, bitová hloubka, počet kanálů, délka audio nahrávky a vybraný audio soubor se vypíší do Command Window prostředí MATLAB. V případě, kdy vodoznak není nalezen, se zobrazí dialogové okno, viz obr. 8.4. Máte možnost načíst nový audio souboru, nebo program ukončit.



Obr. 8.4: Zobrazení dialogových oken při extrakci vodoznaku

O průběhu extrakce vodoznaku jste informováni v Command Window. Po určité časové prodlevě (doba potřebná pro extrakci vodoznaku se zobrazí do Command Window) je zobrazen **extrahovaný vodoznak**. Je zobrazena i bitová chybovost (BER) udávaná v procentech.

Na obr. 8.5 je zobrazen postup při vkládání a extrakci vodoznaku.



Obr. 8.5: Hlavní kroky při a) vkládání vodoznaku b) extrakci vodoznaku

9 Závěr

Diplomová práce je zaměřena na digitální vodoznačení audio signálů. V teoretické části této práce jsou popsány psychoakustické jevy, které hrají významnou roli v procesu vodoznačení. Dále jsou zde popsány používané metody vodoznačení a především požadavky na ně kladené. Následně jsou uvedeny příklady využití metod digitálního vodoznačení audio signálů v praxi. V závěru teoretické části je podrobně popsána transformace DTWT, především DWPT, a vysvětlena jejich souvislost s bankou číslicových filtrů.

Samotný popis implementace algoritmu pro vkládání a vyjmutí vodoznaku je popsán v kapitole pět. Postupuji dle autora Xing He. V této knize se objevuje řada nesrovnalostí a některé procesní kroky mi byly zcela utajeny, proto jsem byl nucen doplnit tyto vynechané kroky o své vlastní výpočty. Pro digitální vodoznačení audio signálů jsem si vybral metodu přímého rozprostření spektra. Jedná se o metodu, která je robustní a má dobré vlastnosti při potlačení rušení. Vodoznak je rozprostřen pomocí PN sekvence a podle psychoakustického modelu je přidán ke koeficientům vlnkové transformace. Tento model je založený na transformaci DWPT. Tímto způsobem je získán vodoznačený audio signál. Na přijímací straně přenosového kanálu jsem pro detekci vodoznaku použil synchronizační posloupnost. Následně jsou dekódované bity převedeny zpět na vodoznak.

V kapitole šest testuji robustnost vloženého vodoznaku. Jedná se o skupinu algoritmů, které slouží k simulaci možných útoků vedených k poškození, nebo odstranění vodoznaku z vodoznačených audio dat. Pro test robustnosti jsou vybrány tyto úpravy se signálem: oříznutí audio signálu, změna vzorkovacího kmitočtu, ztrátová komprese, filtrace, ekvalizace, vložení hudebního efektu a bílého šumu s gaussovským rozdělením. Algoritmy jsou implementovány v prostředí Matlab a jeho simulační nadstavbě Simulinku. Pro převod z formátu *wav* do formátu *mp3* a zpět jsem použil volně stažitelný enkodér LAME. Útoky musely být samozřejmě navrženy tak, aby zcela neznehodnotily audio (jeho poslechovou kvalitu). Jelikož je použita metoda rozprostřeného spektra, tak musí být na straně přijímače zaručena přesná synchronizace a vhodná metoda dekódování vodoznaku. V opačném případě nemůže být zaručena robustnost vloženého vodoznaku. Z časových důvodů byly vybrány audio nahrávky délky 2 s. Vložený vodoznak pro $\alpha = 0{,}0$ nelze považovat za velmi robustní (odolal útokům: oříznutí audia, filtrace HP a v jednom případě vložení hudebního efektu). Lze ale říci, jak bylo naznačeno v teoretické části práce, že se pro $\alpha = 0{,}9$ značně zvýšila robustnost vodoznaku (odolal útokům: oříznutí audia, mp3 komprese, filtrace HP, vložení bílého šumu a v jednom případě vložení hudebního efektu a ekvalizace). Algoritmus je schopen dekódovat správně vodoznak do chybovosti cca. 35 %. Jako nejvíce robustní, pro obě hodnoty α, se jevil vodoznak vložený do klasické hudby. Dále lze říci, že byl vodoznak nejvíce

odolný proti oříznutí audia. Hodnot BER 0,0 % se mi při zajištění přesné synchronizace nepodařilo dosáhnout, protože nebylo možné na základě vypočtených koeficientů vlnkové transformace přesně dekódovat vodoznak (Pro realizaci DWPT používám Matlabovskou funkci *wpdec*). Při načtení upravených koeficientů vlnkové transformace na straně dekodéru bylo dosaženo při dekódování vodoznaku hodnot BER 0,0 %. Algoritmus tedy funguje správně. Synchronizace v časové oblasti se neprojevila jako velmi robustní vůči útokům.

Poslední část diplomové práce se zabývá stanovením úrovně transparentnosti vloženého vodoznaku. Vybral jsem si jednoho zástupce z objektivních metod a jednoho ze subjektivních metod. Z objektivních metod jsem si zvolil ABX dvojitý slepý poslechový test, jehož součástí bylo i stanovení míry degradace kvality u vodoznačeného audia dle ITU-R BS.562. Test byl realizován na počítačích se software *ABC/Hidden Reference Audio Comparison Tool* (ABC/HR) v prostorách UTKO (PA-327). Pro test byly k dispozici profesionální uzavřená sluchátka *Sennheiser HD 280 PRO* a polootevřená sluchátka *AKG K66*. Testu se účastnilo 10 posluchačů, kteří měli k dispozici šest audio nahrávek různých žánrů (klasická hudba, rock, pop, country, disko a swing), do kterých se vkládal postupně vodoznak s koeficientem intenzity $\alpha = 0{,}0$ a $\alpha = 0{,}2$. Test byl poměrně časově náročný (cca 40 min. pro jednoho posluchače). Pokud vezmeme výsledky od všech posluchačů, tak lze v průměru říci, že pro $\alpha = 0{,}0$ bylo 5 z 6 nahrávek transparentní. Zatímco pro $\alpha = 0{,}2$ lze považovat za transparentní již jen 2 z 6 nahrávek. Na výsledek poslechových testů má samozřejmě vliv také to, jaká byla zvolena část audia (10 s). Snažil jsem se pro objektivnost výsledků testů vybírat části energeticky slabší i silnější. V tišších pasážích mohl být vodoznak slyšet a projevit se jako šum. Tento problém by mohl být pravděpodobně odstraněn časovým maskováním. Hodnota intenzity vodoznaku $\alpha = 0{,}2$ byla zvolena jako kompromis mezi transparentností vodoznaku a kvalitou audio nahrávky. Vyšší hodnota zvyšovala degradaci kvality audia. Ze subjektivních metod jsem si vybral výpočet SNR. Výsledky korespondují s výsledky poslechových testů.

Jak je patrné z výsledků robustnosti a transparentnosti. Intenzita vodoznaku $\alpha = 0{,}0$ nám zaručí jeho větší transparentnost, ale menší robustnost. Naproti tomu $\alpha = 0{,}2$ (a vyšší) nám zaručí větší robustnost za cenu nižší transparentnosti vloženého vodoznaku.

Z výsledků výpočetní náročnosti je patrné, že zvolená metoda vkládání a vyjmutí vodoznaku je časově náročná. Matlab je optimalizován pro maticové operace, kdežto v algoritmu jsou použity často cykly *while* a *for*, které mohou dobu potřebnou pro výpočet značně prodloužit. Na optimalizaci algoritmu by bylo potřeba více času, avšak jsem byl omezen rozsahem a náročností této závěrečné práce.

Přímé srovnání úrovně transparentnosti vodoznaku s obvyklými metodami nelze provést, protože není k dispozici volně stažitelný algoritmus, nebo přímo program, pomocí něhož by bylo možné vkládat vodoznak do audio souboru.

Výsledky, kterých bylo dosaženo, lze shrnout do následujícího odstavce. V prostředí Matlab jsem úspěšně implementoval algoritmus pro vkládání a vyjmutí vodoznaku. Dále jsem tento vodoznak otestoval na robustnost pomocí algoritmů, které jsem implementoval opět v prostředí Matlab. V závěru jsem stanovil úroveň transparentnosti vloženého vodoznaku. Části stanovení úrovně transparentnosti, především poslechovým testům, byla věnována značná část práce.

Seznam použitých zdrojů

Monografie

[1] FASTL, H.; ZWICKER, E. *Psychoacoustics : Facts and Models*. 3rd edition. Berlin: Springer, 2006. 462 s. ISBN 978-3-540-23159-2.

[2] JAN, J. *Číslicová filtrace, analýza a restaurace signálů*. 2. vydání. Brno: VUTIUM, 2002. 427 s. ISBN 80-214-1558-4.

[3] CVEJIC, N.; SEPPÄNEN, T. *Digital Audio Watermarking Techniques and Technologies : Applications and Benchmarks*. 1st edition. New York: IGI Global, 2007. 328 s. ISBN 978-159904513-9.

[4] HE, X. *Watermarking in Audio: Key Techniques and Technologies*. 1st edition. New York: Kambria Press, 2008. 182 s. ISBN 978-1-60497-501-7.

[5] COX, I.; MILLER, M. L.; BLOOM, J. A. *Digital Watermarking and Steganography*. 2nd edition. Burlington: Elsevier, 2008. 593 s. ISBN 978-0-12-372585-1.

[6] ČSN EN ISO/IEC 11172-3 Informační technologie - Kodování pohyblivých obrazů včetně doprovodného zvuku pro číslicový záznam do rychlosti 1,5 Mbit/s - Část 3: Zvuk (audio), 1997.

[7] ZAPLATÍLEK, K.; DOŇAR, B. *Matlab: začínáme se signály*. 1. vydání. Praha: BEN, 2006. 271 s. ISBN 80-7300-200-0.

[8] SEITZ, J. *Digital Watermarking for Digital Media*. 1st edition. Hershey: Information Science Publishing, 2005. 262 s. ISBN 159140518-1.

[9] ARNOLD, M.; SCHMUCKER, M.; WOLTHUSEN, S. D. *Techniques and Applications of Digital Watermarking and Content Protection*. 1st edition. Norwood: Artech House, 2003. 296 s. ISBN 1-58053-111-3.

[10] MELKA, A. *Základy experimentální psychoakustiky*. 1. vydání. Praha: AMU, 2005. 327 s. ISBN 80-7331-043-0

Skripta

[11] BALÍK, M. *Číslicové zpracování akustických signálů*. Skripta VUT v Brně, 2008.

[12] KÁŇA, L. *Elektroakustika*. Skripta VUT v Brně, 2002

[13] SMÉKAL, Z. *Číslicové zpracování signálů*. Skripta VUT v Brně, 2008.

[14] SMÉKAL, Z.; SYSEL, P. *Číslicové filtry*. Skripta VUT v Brně, 2004.

[15] HANUS, S. *Rádiové a mobilní komunikace*. Skripta VUT v Brně, 2003.

Elektronické dokumenty

[16] BOLDIŠ, Petr. *Bibliografické citace dokumentu podle ČSN ISO 690 a ČSN ISO 690-2: Část 2 – Modely a příklady citací u jednotlivých typů dokumentu*. Verze 3.0 (2004). © 1999–2004, poslední aktualizace 11. 11. 2004. Dostupné z: <http://www.boldis.cz/citace/citace2.pdf>.

[17] BOLDIŠ, Petr. *Bibliografické citace dokumentu podle ČSN ISO 690 a ČSN ISO 690-2: Část 1 – Citace: metodika a obecná pravidla*. Verze 3.3. © 1999–2004, poslední aktualizace 11.11. 2004. Dostupné z: <http://www.boldis.cz/citace/citace1.pdf>.

Vědecko-kvalifikační práce

[18] KOZUMPLÍK, J. *Vlnkové transformace a jejich využití pro filtraci signálu EKG*. [online] Brno, 2004. 81 s. Habilitační práce na Fakultě elektrotechniky a komunikačních technologií Vysokého učení technického [cit. 2009-11-03]. Dostupné z: <http://www.dbme.feec.vutbr.cz/~kozumplik/habilitace.pdf>.

Článek v elektronickém seriálu

[19] ZEZULA, R. Objektivní a subjektivní metody vyhodnocování kvality vodoznačných audio signálů. *Elektrorevue* [online]. 2008 [cit. 2010-03-20]. Dostupné z: <http://www.elektrorevue.cz/cz/clanky/zpracovani-signalu/15/objektivni-a-subjektivni-metody-vyhodnocovani-kvality-vodoznacnych-audio-signalu/>. ISSN 1213-1539.

[20] CARNERO, B.; DRYGAJLO, A. Perceptual speech coding using time and frequency masking constraints. *IEEE International Conference on Acoustics, Speech and Signal Processing* [online]. 1997, vol. 2 [cit. 2009-12-01], pp. 1363-1366. Dostupné z: <http://ieeexplore.ieee.org/iel3/4635/13030/00596200.pdf?arnumber=596200>. ISBN 0-8186-7919-0.

[21] SU, J. K.; HARTUNG, F. Digital watermarking of text, image, and video documents. *Science Direkt: Computers & Graphics* [online]. 1998, vol. 22, no. 6 [cit. 2010-03-02], pp. 687-695. Dostupné z: <linkinghub.elsevier.com/retrieve/pii/S0097849398000892>. ISSN 0097-8493.

Počítačový program

[22] MIYAGUCHI, Darryl. *ABC/Hidden Reference Audio Comparison Tool* [počítačový program]. Ver. 1.1 for Windows. 2004 [cit. 2010-03-29]. Dostupné z: <http://ff123.net/abchr/abchr.html>.

Seznam použitých zkratek, veličin a symbolů

$\mathbb{R}$	množina reálných čísel
$\mathbb{Z}$	množina celých čísel
ASCII	American Standard Code for Information Interchange
AWGN	kanál s rovnoměrně rozprostřeným bílým šumem (Additive white Gaussian noise)
BER	bitová chybovost (Bit Error Rate)
CBR	konstantní datový tok (Constant Bit Rate)
CRC	cyklická redundantní kontrola (Cyclic Redundancy Check)
DSSS	systém s přímým rozprostřením spektra (Direct Sequence Spread Spectrum)
D/A	digitálně analogový převodník
dB	decibel
DCT	Diskrétní kosinová transformace (Discrete Cosine Transform)
DRM	Správa digitálních práv (Digital Rights Management)
DTWT	Diskrétní vlnková transformace s diskrétním časem (Discrete Time Wavelet Transform)
DWPT	Diskrétní paketová vlnková transformace (Discrete Wavelet Packet Transform)
DWT	Diskrétní vlnková transformace (Discrete Wavelet Transform)
FHSS	systém s frekvenčním skákáním (Frequency Hopping Spread Spectrum)
FIR	filtr s konečnou délkou impulsové odezvy (Finite Impulse Response)
FT	Fourierova transformace (Fourier Transform)
GPS	družicový polohový systém (Global Positioning System)
Hz	Hertz
IDTWT	Inverzní diskrétní vlnková transformace s diskrétním časem (Inverse Discrete Time Wavelet Transform)
IIR	filtr s nekonečnou délkou impulsové odezvy (Infinite Impulse Response)
ITU-R	International Telecommunication Union Radiocommunication sector

LFO	sub-akustický oscilátor (Low-Frequency Oscillator)
LSB	nejméně významný bit (Least Significant Bit)
MPEG	Motion Picture Expert Group
NMR	odstup šumu od maskovacího signálu (Noise to Mask Ratio)
PCM	pulzní kódová modulace (Pulse Code Modulation)
PEAQ	Perceptual Evaluation of Audio Quality
PF	Peak Filter
PN	pseudonáhodná (pseudonoise) posloupnost
SNR	odstup signál-šum (Signal to Noise Ratio)
STFT	krátkodobá Fourierova transformace (Short-Time Fourier Transform)
VBR	proměnný datový tok (Variable Bit Rate)
WT	vlnková transformace (Wavelet Transform)
f	kmitočet (frekvence) (Hz)
f_{vz}	vzorkovací kmitočet (Hz)
$H(z), h(n)$	přenosová funkce a impulsní charakteristika diskrétního systému
L_p	hladina akustického tlaku (dB)
p_0	akustický tlak (Pa)
$r_{xy}(n)$	odhad křížové korelační posloupnosti
t	čas (s)
$s(t)$	spojitý signál
x(n)	diskrétní signál
$\psi(t)$	mateřská vlnka
$\psi_{p,q}(t)$	časově posunutá a dilatovaná vlnka
$\phi(n)$	měřítková funkce
ω	úhlový kmitočet (rad. s^{-1}), $\omega = 2.\pi.f$

Seznam příloh

Příloha A: Hierarchie zdrojových souborů.

Příloha B: Metodika poslechového testu v programu ABC/HR.

Příloha C: Výsledky ABX poslechového testu

C.1 pro $\alpha = 0{,}0$

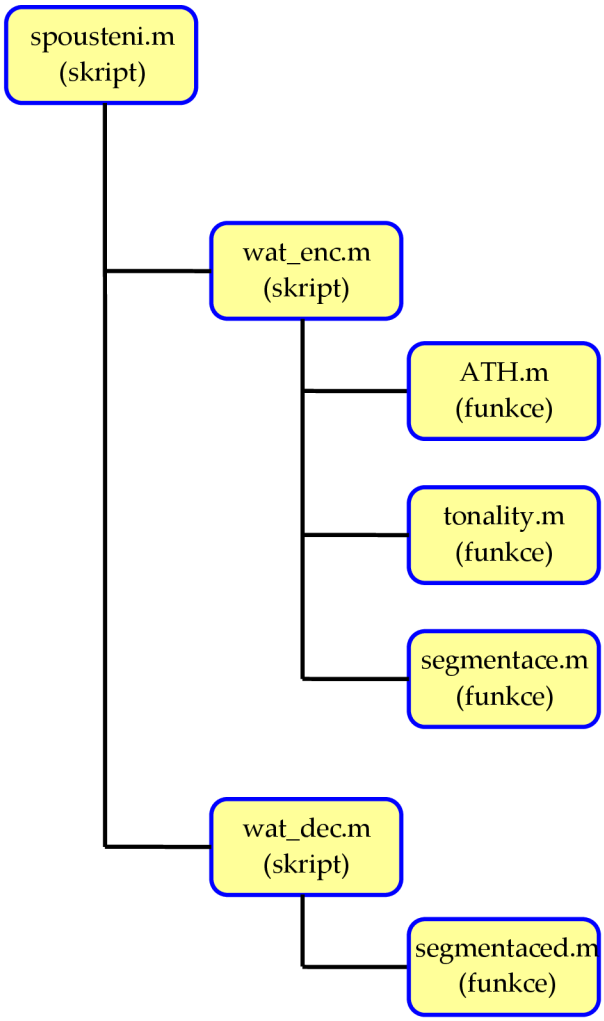
C.2 pro $\alpha = 0{,}2$

Příloha D: CD nosič obsahující pdf verzi diplomové práce a tři adresáře:

- adresář **m_file** – obsahuje soubory spustitelné v Matlabu (*.m), jedná se o skripty a funkce využívané v algoritmu pro vkládání a extrakci vodoznaku.
- adresář **audio** – obsahuje soubor ve formátu wave (*.wav), jedná se o originální a vodoznačené audio nahrávky používané v této DP.
- adresář **watermark_test** – obsahující soubory spustitelné v Matlabu (*.m, *.mdl), jedná se o skripty a modely používané v algoritmech simulující útok na vložený vodoznak – test robustnosti.

Přílohy

A Hierarchie zdrojových souborů

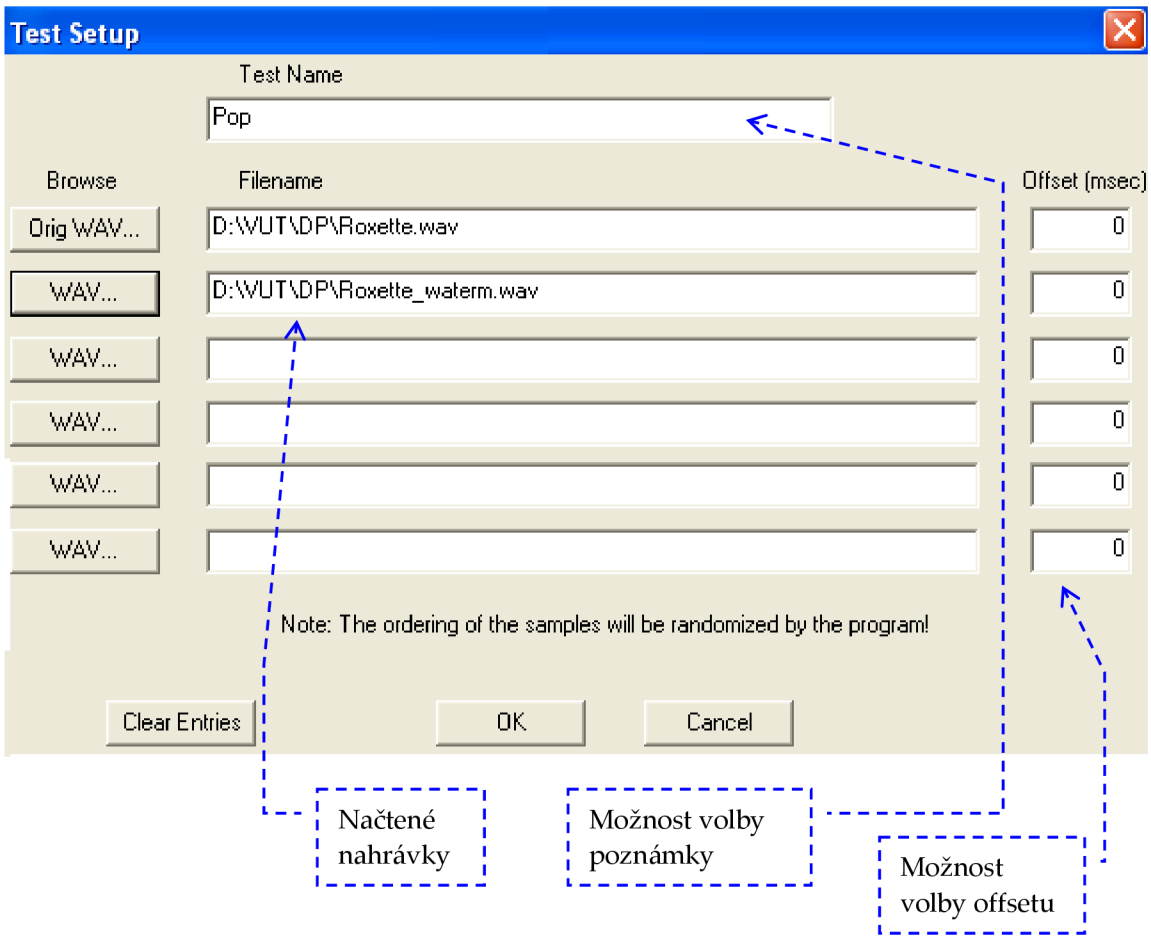


B Metodika poslechového testu v programu ABC/HR

1. Body 1–5 a 12–13, označené **červeně**, má na starost **administrátor testu**. Pokud chcete hodnotit audio nahrávky (funkce posluchače), přejděte k bodu 6.

2. Progam *ABC/Hidden Reference Audio Comparison Tool* spusťte pomocí **abchr.exe**

3. Nejprve je potřeba načíst originální a vodoznačené audio. V horní liště menu zvolte nabídku **File** → **Setup Test**. Dostanete se tak do nastavení testu. Stiskněte tlačítko **Orig WAV…** pro načtení originálního audia. Dále stiskněte tlačítko **WAV…** pro načtení vodoznačeného audia. Volitelně můžete vyplnit textové pole **Test Name** a **Offset**. V poli **Offset (msec)** máte možnost pro každé audio zvolit offset, udávaný v milisekundách. V poli **Test Name** si můžete pojmenovat název aktuálního testu (např. podle aktuálně testovaného žánru nahrávek). Potvrďte tlačítkem **OK**.



4. Zmáčkněte tlačítko **ABX…**. Otevře se Vám úvodní okno ABX testu. Abyste se dostali k parametrům testu, musíte zvolit v části **Select A** z rozbalovacího seznamu možnost **Original**, nebo **Sample 1**. V části **Select B** vyberte zbývající možnost. Nyní můžete přejít k části **Test Parameters**. Z rozbalovacího seznamu zvolte jednu z přednastavených možností testu: **Moderate difference** (mírný rozdíl), **Obvious diference** (zřetelný rozdíl), **Subtle diference** (jemný rozdíl). Jedná se o rozdíl kvality vodoznačeného audia oproti originálnímu audiu. Po zvolení určité možnosti se parametry **alpha**, **beta**, **theta** nastaví na defaultní hodnoty. Samozřejmě můžete poupravit hodnotu těchto parametrů. Pokud si nejste jisti, jak můžou tyto parametry ovlivnit test, nebo jaký mají význam, přečtěte si nejprve kapitolu 7.2.1. Na závěr zmačkněte tlačítko **Close**.

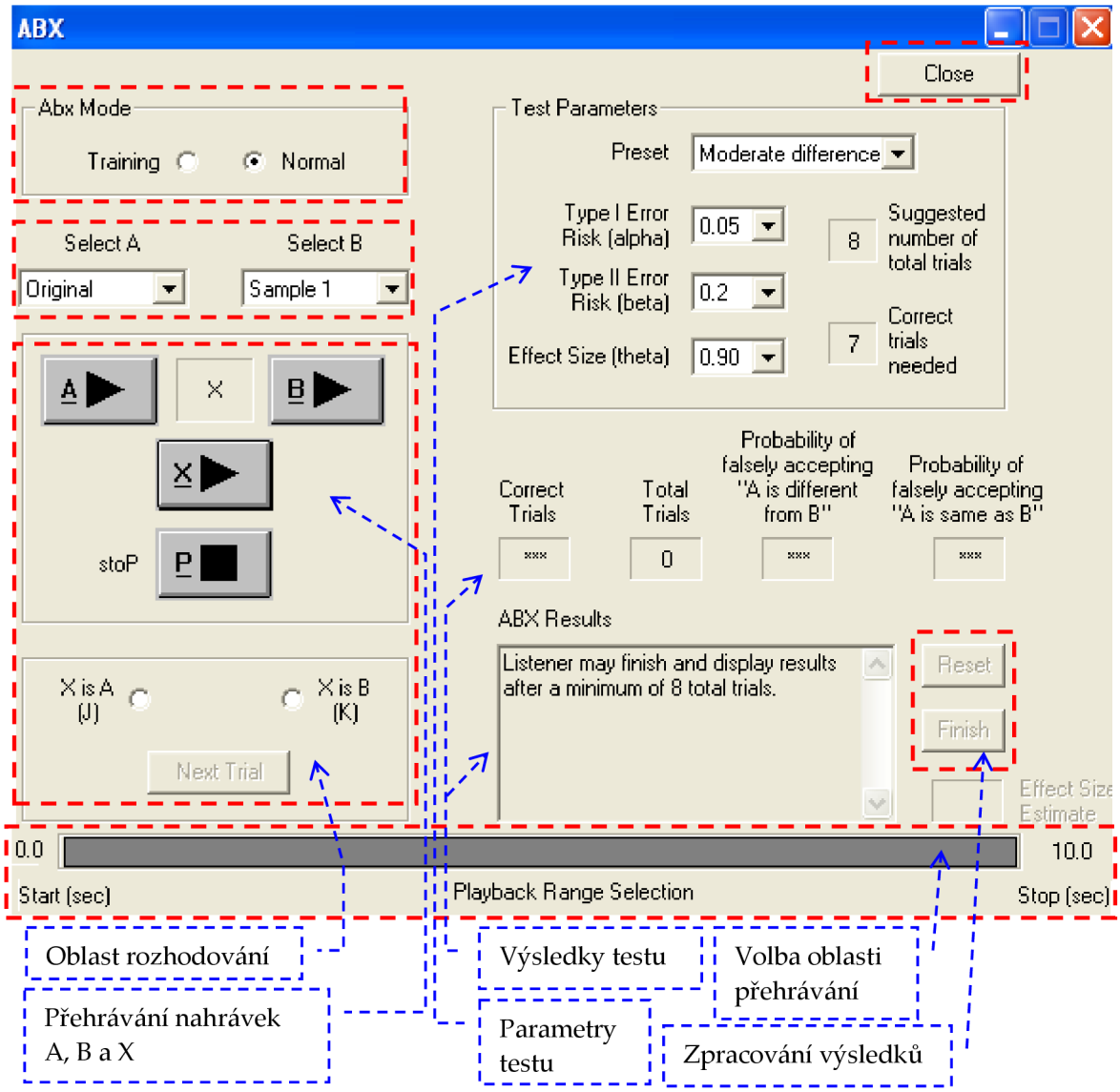
5. Program nechte otevřený. V tuto chvíli můžou posluchači začít hodnotit vybrané nahrávky.

6. V hlavním okně programu ABC/HR zmačkněte tlačítko **ABX…** (vpravo nahoře). Zobrazí se Vám úvodní okno ABX poslechového testu.

7. V části **Abx Mode** zvolte režim **Training** nebo **Normal**.

8. Režim **Training**. Jak již název režimu napovídá, jedná se pouze o poslechové cvičení. Nejprve zvolte v části **Select A** z rozbalovacího seznamu možnost **Original** (originální audio), nebo **Sample 1** (vodoznačené audio). V části **Select B** vyberte zbývající možnost. Následně můžete přejít k samotnému cvičení. Pro přehrání audia A/B zmačkněte tlačítko **Play** (*A* nebo *B*). Kde písmeno A a B značí originální, nebo vodoznačené audio dle výběru výše. Při zmáčknutí tlačítka **Play** *X* se **náhodně** přehraje buď nahrávka A nebo B. Při zmáčknutí jiného tlačítka **Play** v průběhu přehrávání nebude zvolená nahrávka přehrána od začátku. Pokud Vám tato funkce nevyhovuje a nechcete přehrát celou nahrávku, tak musíte vždy použít tlačítko **Stop** *P*. U všech třech nahrávek máte možnost volit oblast přehrávání pomocí výběrové lišty, umístěné v dolní části okna (podržením a tažením myší volíte Start a Stop oblast udávanou v sekundách). Vašim úkolem je zvolit, zda nahrávka **X** odpovídá nahrávce **A** nebo **B**, zaškrtnutím políčka **X is A**, nebo **X is B**. Svoji volbu musíte potvrdit tlačítkem **Next Trial**. Pokud tuto volbu nepotvrdíte, tak se pod písmenem X bude skrývat stále ta stejná náhodně vybraná nahrávka. Tímto jste provedli jeden ze svých pokusů. Tyto kroky můžete opakovat v libovolném počtu. V průběhu trénování vidíte výsledky Vašeho snažení ve spodní části okna v poli *Correct Trials* (počet správných identifikací) a *Total Trials* (celkový počet hodnocení). Dále je zde vypočtena pravděpodobnost *p* tvrzení A je stejné s B. Tlačítkem **Reset** vymažete výsledek Vašeho testu. Tento režim můžete přeskočit a přejít do režimu Normal.

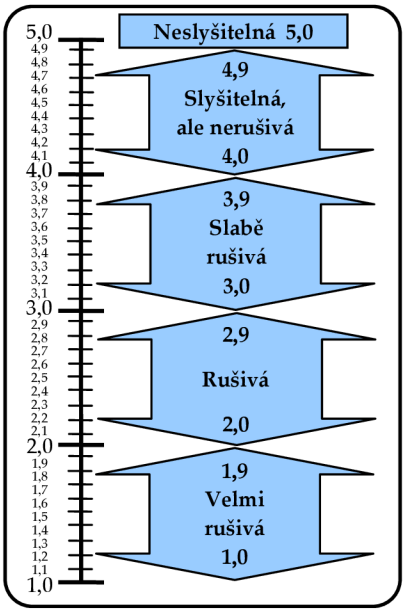
9. Režim **Normal**. V režimu Normal přistupujete k testu obdobně jako v režimu Training. Při vyhodnocení nahrávky **X** postupujete dle bodu 8. U tohoto režimu nepoužívejte tlačítko **Reset**. Po **osmém pokusu** (počet pokusů je zobrazen v textovém poli *Total Trials*) zmačkněte tlačítko **Finish**. Následně zmačkněte tlačítko **Close** (vpravo nahoře). Dostanete se zpět do úvodního okna programu ABC/HR, kde je druhá část poslechového testu.

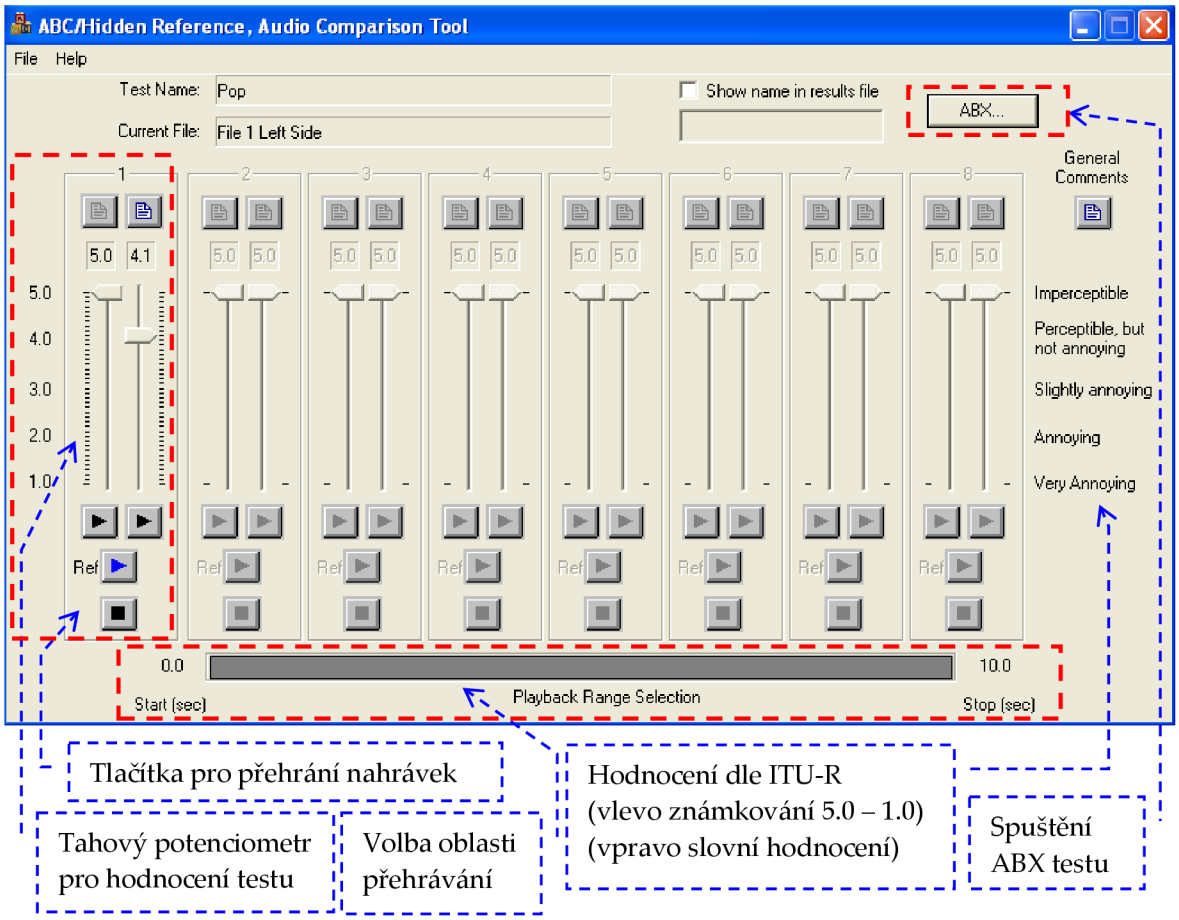


10. Cílem tohoto testu je ohodnotit dle normy ITU-R míru degradace kvality vodoznačené nahrávky vůči originální nahrávce.

11. V úvodním okně programu je aktivní skupina s označením **1**. Nacházejí se zde dva tahové potenciometry, tři tlačítka Play a jedno tlačítko Stop. Modré tlačítko Play s označením „Ref" umožňuje přehrát **originální audio**. Jedno ze zbylých tlačítek Play (pod tahovými potenciometry) přehraje **vodoznačené audio**, druhé **originální audio**. Program **náhodně** přiřadí nahrávky těmto tlačítkům. Reakce na opakované zmáčknutí tlačítek Play během přehrávání je popsána v bodě 8. U všech třech nahrávek máte možnost volit oblast přehrávání stejným způsobem jako v bodě 8. Vašim

úkolem je odhalit, u které z těchto nahrávek, ukrytými pod tlačítkem Play (**levé/pravé**), došlo k degradaci kvality a určit její míru pomocí **tahového potenciometru**. Pokud je jedno z tlačítek Play zašedlé, vodoznak byl vyhodnocen jako slyšitelný, a tudíž můžete přímo hodnotit míru degradace vodoznačeného audia. Stupnice se nachází vedle potenciometrů. Na závěr máte možnost vložit poznámku (umístěná nad tahovým potenciometrem) k nahrávce, u které si myslíte, že došlo k degradaci kvality (např. popsat typ degradace). Program nevypínejte a výsledky testu předejte **administrátorovi testu**.





12. Pokud jste vyplnili pole „Test Name“ a chcete, aby se tento název objevil ve výsledcích, tak zaškrtněte pole „Show name in results file“. Následně zvolte v horní liště menu **File** → **Save Tests Results**. Výsledky tohoto testu uložte do textového souboru *.txt.

13. Kroky 3–12 opakujte dle počtu dvojic nahrávek (vodoznačené, originál), které chcete podrobit poslechovým testům.

C Výsledky ABX poslechového testu

C.1 pro α = 0,0

Posluchač	Číslo audia	Počet správných identifikací k	Práh T	Celkový počet hodnocení n	Pravděpodobnost nesprávného tvrzení: „A je rozdílné od B“	Posluchač	Číslo audia	Počet správných ident. k	Práh T	Celkový počet hodnocení n	Pravděpodobnost nesprávného tvrzení: „A je rozdílné od B“
Ivan Míča	1	2	7	8	0,965	Pavel Rajmic	1	5	7	8	0,363
	2	5	7	8	0,363		2	5	7	8	0,363
	3	2	7	8	0,965		3	8	7	8	0,004
	4	5	7	8	0,363		4	3	7	8	0,855
	5	3	7	8	0,855		5	3	7	8	0,855
	6	5	7	8	0,363		6	4	7	8	0,637
Jan Šporik	1	8	7	8	0,004	Zdeněk Průša	1	8	7	8	0,004
	2	5	7	8	0,363		2	3	7	8	0,855
	3	7	7	8	0,035		3	8	7	8	0,004
	4	8	7	8	0,004		4	8	7	8	0,004
	5	6	7	8	0,145		5	8	7	8	0,004
	6	1	7	8	0,996		6	4	7	8	0,637
Radek Beneš	1	6	7	8	0,145	Pavel Šilhavý	1	5	7	8	0,363
	2	2	7	8	0,965		2	4	7	8	0,637
	3	5	7	8	0,363		3	4	7	8	0,637
	4	6	7	8	0,145		4	4	7	8	0,637
	5	5	7	8	0,363		5	2	7	8	0,965
	6	4	7	8	0,637		6	3	7	8	0,855
Radim Číž	1	2	7	8	0,965	Jiří Schimmel	1	8	7	8	0,004
	2	6	7	8	0,145		2	8	7	8	0,004
	3	3	7	8	0,855		3	8	7	8	0,004
	4	4	7	8	0,637		4	6	7	8	0,145
	5	3	7	8	0,855		5	8	7	8	0,004
	6	2	7	8	0,965		6	4	7	8	0,637
Petr Sysel	1	7	7	8	0,035	Jan Karásek	1	8	7	8	0,004
	2	4	7	8	0,637		2	8	7	8	0,004
	3	4	7	8	0,637		3	8	7	8	0,004
	4	3	7	8	0,855		4	8	7	8	0,004
	5	7	7	8	0,035		5	5	7	8	0,363
	6	5	7	8	0,363		6	6	7	8	0,145

C.2 pro α = 0,2

Poslu-chač	Číslo audia	Počet správných identifikací k	Práh T	Celkový počet hodnocení n	Pravděpodobnost nesprávného tvrzení: „A je rozdílné od B“	Poslu-chač	Číslo audia	Počet správných ident. k	Práh T	Celkový počet hodnocení n	Pravděpodobnost nesprávného tvrzení: „A je rozdílné od B“
Ivan Míča	1	8	7	8	0,004	Pavel Rajmic	1	8	7	8	0,004
	2	6	7	8	0,145		2	7	7	8	0,035
	3	8	7	8	0,004		3	8	7	8	0,004
	4	8	7	8	0,004		4	7	7	8	0,035
	5	8	7	8	0,004		5	3	7	8	0,855
	6	5	7	8	0,363		6	6	7	8	0,145
Jan Šporik	1	8	7	8	0,004	Zdeněk Průša	1	8	7	8	0,004
	2	2	7	8	0,965		2	5	7	8	0,363
	3	8	7	8	0,004		3	8	7	8	0,004
	4	8	7	8	0,004		4	8	7	8	0,004
	5	7	7	8	0,035		5	8	7	8	0,004
	6	7	7	8	0,035		6	2	7	8	0,965
Radek Beneš	1	7	7	8	0,035	Pavel Šilhavý	1	8	7	8	0,004
	2	5	7	8	0,363		2	1	7	8	0,996
	3	8	7	8	0,004		3	2	7	8	0,965
	4	5	7	8	0,363		4	5	7	8	0,363
	5	7	7	8	0,035		5	6	7	8	0,145
	6	4	7	8	0,637		6	4	7	8	0,637
Radim Číž	1	8	7	8	0,004	Jiří Schimmel	1	8	7	8	0,004
	2	3	7	8	0,855		2	8	7	8	0,004
	3	8	7	8	0,004		3	8	7	8	0,004
	4	8	7	8	0,004		4	5	7	8	0,363
	5	8	7	8	0,004		5	8	7	8	0,004
	6	4	7	8	0,637		6	6	7	8	0,145
Petr Sysel	1	8	7	8	0,004	Jan Karásek	1	8	7	8	0,004
	2	8	7	8	0,004		2	8	7	8	0,004
	3	8	7	8	0,004		3	8	7	8	0,004
	4	4	7	8	0,637		4	8	7	8	0,004
	5	8	7	8	0,004		5	8	7	8	0,004
	6	6	7	8	0,145		6	8	7	8	0,004