

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta
DIPLOMOVÁ PRÁCE



Viliam Sabol

Model kódování ve sluchové dráze

Katedra softwarového inženýrství

Vedoucí diplomové práce: RNDr. David Obdržálek

Studijní program: Softwarové systémy

2010

Na tomto mieste by som chcel poďakovať pánovi docentovi MUDr. RNDr. Petrovi Maršálkovi, Ph.D., ktorý mi ako konzultant počas písania tejto diplomovej práce venoval ochotu a čas, a poskytol mnoho užitočných rád a pripomienok. Ďalej chcem poďakovať pánovi RNDr. Davidovi Obdržádkovi za vedenie diplomovej práce a cenné rady pri implementácii jej softwarovej časti. V neposlednom rade patrí poďakovanie Janu Kohoutovi za jeho spoluprácu pri testovaní softwarovej časti práce.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 17.6.2010

Viliam Sabol

Názov práce: Model kódování ve sluchové dráze
Autor: Viliam Sabol
Katedra: Katedra softwarového inženýrství
Vedúci diplomovej práce: RNDr. David Obdržálek
E-mail vedúceho: david.obdrzalek@mff.cuni.cz

Abstrakt:

Cieľom tejto práce bolo preskúmanie funkcie častí sluchovej dráhy, ktoré sú zodpovedné za kódovanie akustických podnetov do neurálnej aktivity v sluchovom nerve. Tieto poznatky sú využívané pri konštrukcii kochleárnych implantátov. Kochleárny implantát čiastočne nahrádza funkciu stredného a vnútorného ucha, ktoré prevádza mechanickú energiu (zvuk) na postupnosť impulzov (akčných potenciálov) v sluchovom nerve elektrickou stimuláciou a umožňuje obnoviť sluch pri hluchote spôsobenej poruchou vnútorného ucha. Ako je zvuk zakódovaný v sluchovom nerve nie je presne popísané, ale teoretické poznatky z tejto oblasti umožňujú konštrukciu kochleárnych implantátov. Ďalšou úlohou tejto práce je oboznámenie sa s niektorými základnými metódami kódovania používaného v súčasných kochleárnych implantátoch a vytvorenie počítačovej simulácie implantátu na základe niektorej z týchto metód. Na implementáciu softwarovej časti práce bolo použité programové prostredie Matlab s podporou pre digitálne spracovanie signálu (DSP). DSP tvorí podstatnú časť kochleárneho implantátu. Výsledkom je aplikácia umožňujúca kódovanie a dekódovanie zvuku, s možnosťou nastavenia parametrov kódovania a vizualizácie základných akustických parametrov vstupu a výstupu. Užitočnou funkcionalitou aplikácie je možnosť porovnania výsledkov simulácie s rôznymi parametrami.

Kľúčové slová: kochleárny implantát, simulácia, kódovanie zvuku, sluchový nerv

Title: Model of coding in the auditory pathway
Author: Viliam Sabol
Department: Department of Software Engineering
Supervisor: RNDr. David Obdržálek
Supervisor's e-mail address: david.obdrzalek@mff.cuni.cz

Abstract:

The aim of this work is to explore parts of the human auditory pathway that are involved in encoding of acoustic stimuli to the neural activity in cochlear nerve. This knowledge is exploited in cochlear implant design. Cochlear implant bypasses the middle ear, and substitutes the inner ear by converting the mechanical energy (sound) to electric stimuli that evoke the sequence of impulses (action potentials) in cochlear nerve. Cochlear implants are used to partially restore hearing for patients with hearing loss caused by the damage of the inner ear. Today, we do not know exactly how the sound is encoded in cochlear nerve neural activity. However, there are methods used to stimulate the cochlear nerve, which are used in cochlear implants. Another aim of this work is to explore sound encoding methods used in present cochlear implants, and implementation of cochlear implant software simulation that uses one of these coding strategies. Simulation is implemented in Matlab environment that supports digital signal processing, which is the fundamental part of the cochlear implant. The result of this work is application that can encode and decode sound, configure and adjust cochlear implant parameters, and visualize basic acoustic parameters of the input and output sound. Results of the simulation for different parameter settings can be easily compared, which is very helpful feature of the application.

Keywords: cochlear implant, simulation, sound encoding, auditory nerve

Obsah

Zoznam obrázkov	6
Zoznam použitých skratiek.....	7
1 Úvod.....	8

Teoretická časť

2 Ucho a kódovanie zvuku v sluchovom nerve	9
2.1 Stavba sluchového orgánu a spôsob spracovávania zvukového podnetu	9
2.2 Prirodzené kódovanie vlastností akustického signálu	12
2.3 Vnímanie vlastností akustického signálu.....	13
3 Spracovanie akustického signálu v kochleárných implantátoch.....	19
3.1 Digitálne spracovanie signálu	19
3.2 Elektrostimulácia a jej vlastnosti	23
3.2.1 Vlastnosti impulznej stimulácie	23
3.2.2 Tonotopické usporiadanie a počet elektród.....	24
4 Stratégie kódovania v kochleárných implantátoch	26
4.1 Continuous interleaved sampling (CIS).....	27
4.1.1 Predspracovanie signálu	28
4.1.2 Frekvenčná analýza signálu.....	29
4.1.3 Detekcia obálky	30
4.1.4 Kompresia amplitúdy obálky.....	30
4.1.5 Generátor elektrických impulzov	31
4.2 Spectral Peak (SPEAK)	32
4.3 Ďalšie používané stratégie kódovania.....	33

Implementačná časť

5 Simulácia kochleárneho implantátu	35
5.1 Vývojové prostredie.....	35
5.2 Návrh aplikácie	36
5.3 Implementácia blokov metód kódovania	36
5.3.1 Načítavanie a predspracovanie zvuku	37
5.3.2 Návrh banky filtrov	39
5.3.3 Obálka a jej kompresia	43

5.3.4	Výber spektrálneho maxima.....	46
5.3.5	Impulzný výstup simulácie.....	46
5.4	Syntéza zvukového výstupu.....	47
5.5	Prezentácia dát	49
5.6	Implementácia GUI a ďalšie funkcie.....	50
Analýza vlastných výsledkov		
6	Výsledky a diskusia	51
6.1	Metódy porovnávania výsledkov	51
6.2	Konfigurácia testu.....	52
6.2.1	Testovacie vstupy	52
6.2.2	Testované parametre.....	53
6.3	Výsledky testovania vplyvu parametrov simulácie	54
6.4	Súhrn a porovnanie výsledkov z ďalších dostupných zdrojov	60
6.5	Možné vylepšenia simulácie	60
7	Záver	61
	Literatúra	62
	Prílohy	64
A.	Obsah priloženého CD.....	64

Zoznam obrázkov

2-1. Sluchový orgán človeka a jeho časti	9
2-2. Cortiho orgán a vlásokové bunky.....	10
2-3. Vzťah vnímanej frekvencie a pozície v rámci bazilárnej membrány.....	11
2-4. Synchronizácia v sluchovom nerve.....	13
2-5. Závislosť vnímanej výšky tónu na frekvencii	14
2-6. Závislosť vnímanej hlasitosti (prahu počuteľnosti) na frekvencii	15
2-7. Závislosť kritickej šírky pásma na frekvencii	16
2-8. Barkova stupnica	17
2-9. Spektrogram rečového signálu anglického slova ‘seven’	18
3-1. Hammingovo okno.....	23
4-1. Schéma vokodéra (hlasového kodéra/ dekodéra).....	26
4-2. Schéma stratégie kódovania CIS.....	27
4-3. Výstup simulácie stratégie CIS (rekonštrukcia a el. impulzy).....	28
4-4. Schéma stratégie kódovania SPEAK	32
4-5. Výstup simulácie stratégie SPEAK (rekonštrukcia a el. impulzy).....	33
5-1. Použitie DC filtra	37
5-2. Frekvenčná odpoveď preemfázového filtra	38
5-3. Frekvenčná odpoveď logaritmickkej banky filtrov.....	40
5-4. Frekvenčná odpoveď banky filtrov s Greenwoodovym rozložením	41
5-5. Frekvenčná odpoveď banky zložená z osemnástich kritických pásiem.....	41
5-6. Frekvenčná odpoveď FIR a IIR banky filtrov.....	43
5-7. Vzťah hranice integračného filtra a rýchlosti stimulácie	44
5-8. Extrakcia obálky.....	45
5-9. Kompresné krivky	46
5-10. Spektrogramy rekonštruovaných signálov reči po CIS simulácii	49
6-1. Vplyv rýchlosti stimulácie na výsledok simulácie.....	54
6-2. Vplyv šírky spracovávaného pásma na výstup simulácie	55
6-3. Vplyv počtu frekvenčných pásiem na výsledok simulácie	56
6-4. Vplyv metódy rozloženia hraničných frekvencií na výsledky simulácie.....	57
6-5. Vplyv zvoleného typu banky filtrov na výsledok simulácie.....	57
6-6. Vplyv typu kompresnej funkcie na výsledok simulácie	58
6-7. Vplyv typu extrakcie obálky na výsledok simulácie.....	59
6-8. Vplyv počtu spektrálnych maxím na výsledok simulácie.....	59

Zoznam použitých skratiek

ACE	advanced combination encoder	zdokonalený kombinačný kodér
AGC	automatic gain control	automatické riadenie zosilnenia
CA	compressed analog	komprimovaný analógový signál
CIS	continuous interleaved sampling	spojité prekladané vzorkovanie
DC	direct current	jednosmerná zložka signálu
DSP	digital signal processing/processor	digitálne spracovanie signálu/signálový procesor
DTFT	discrete-time Fourier transform	Fourierova transformácia diskrétného signálu
FBS	filter bank synthesis	syntéza pomocou banky filtrov
FFT	fast Fourier transform	rýchla Fourierova transformácia
FIR	finite impulse response	konečná impulzná odpoveď
GUI	graphical user interface	grafické užívateľské rozhranie
IIR	infinite impulse response	nekonečná impulzná odpoveď
LPF	low-pass filter	dolná priepusť
LTI	linear time-invariant	lineárny časovo-invariantný
MCL	most comfortable level	najpríjemnejšia úroveň
MCR	Matlab compiler runtime	knižnica pre samostatné Matlab aplikácie
MPEAK	multi-peak	viacnásobný spektrálny vrchol
PCM	pulse code modulation	pulzná kódová modulácia
PDA	personal digital assistant	vreckový počítač
PPS	pulses per second	pulzy za sekundu
SIMD	single instruction, multiple data	jedna inštrukcia, viac dát
SMSP	spectral maxima sound processor	zvukový procesor spektrálnych maxím
SPEAK	spectral peak	spektrálny vrchol
SPL	sound pressure level	hladina akustického tlaku
SPT	signal processing toolbox	knižnica pre spracovanie signálu
SSS	superimposed sinusoidal synthesis	syntéza pomocou zložených sinusoid
WAVE	waveform audio format	formát zvukových súborov

1 Úvod

Strata sluchu môže byť spôsobená chorobami ako je napr. meningitída, vrodenými vadami, alebo inými príčinami (nadmerné užitie antibiotík). Vo väčšine prípadov však dochádza k čiastočnému poškodeniu, alebo úplnému zničeniu sluchových receptorov, tzv. vlásokových buniek, zatiaľ čo sluchový nerv zostáva nepoškodený a je schopný aj naďalej spracovávať výstup z týchto receptorov. V takýchto prípadoch je možné stimulovať sluchový nerv elektrickými impulzmi pomocou elektród umiestnených v kochlee. Prístroj ktorý prevádza akustický signál až na jednotlivé impulzy ktorými sa stimuluje sluchový nerv sa volá kochleárny implantát. O stimuláciu sluchového nervu sa pokúsil už v osemnástom storočí taliansky fyzik Alessandro Volta [1], avšak prvé podrobnejšie popísané experimenty so stimuláciou sluchového nervu u pacienta postihnutého hluchotou sa objavili až v roku 1950, keď sa chirurgom (A. Djourno a C. Eyriès) podarilo počas operácie elektricky stimulovať sluchový nerv pacienta [1]. Problematika kochleárných implantátov teda existuje v audiológii už niekoľko desaťročí a za túto dobu bolo vyvinutých množstvo matematických modelov rôznych častí sluchovej dráhy, na ktorých základe boli implementované algoritmy kódujúce zvuk do postupnosti impulzov generovaných elektródami implantátu. O žiadnom z týchto prístupov sa ale nedá hovoriť ako o najlepšom, pretože každý sa špecializuje na určité vlastnosti vstupných akustických podnetov, ktoré sa snaží extrahovať a následne zakódovať do výstupu v podobe elektrickej stimulácie. Vo väčšine prípadov sú to algoritmy, ktoré umožňujú lepšie porozumenie hovorenej reči, a to aj v hlučnom prostredí. V súčasnosti sa pracuje aj na vylepšení kódovania hudby, alebo priestorovom počutí pomocou páru kochleárných implantátov. V prvej časti tejto práce sú načrtnuté teoretické základy vnímania zvuku, a tiež niekoľko základných metód kódovania zvuku, ktoré sa používajú v kochleárných implantátoch, a ktoré vznikli na základe poznatkov z audiológie [2], psychoakustiky, ale aj z oblasti digitálneho spracovania signálu. Ďalšou časťou práce je popis implementácie simulácie kochleárneho implantátu, ktorá umožňuje simulovať dve najčastejšie používané metódy kódovania CIS a SPEAK. V každej časti simulácie bolo navrhnutých niekoľko prístupov k spracovaniu vstupného signálu a do implementácie tak bolo zahrnutých viacero parametrov, ktorými možno výsledok simulácie ovplyvniť. Simulátor rečového procesora umožňuje prácu s viacerými vstupnými signálmi súčasne a tiež porovnanie výsledkov dvoch simulácií, pri ktorých je možné nastaviť odlišné typy stratégie kódovania a parametre simulácie. Vstupom a výstupom simulačnej aplikácie sú zvukové nahrávky, pomocou ktorých je v poslednej časti práce analyzovaný vplyv jednotlivých parametrov na výsledok simulácie.

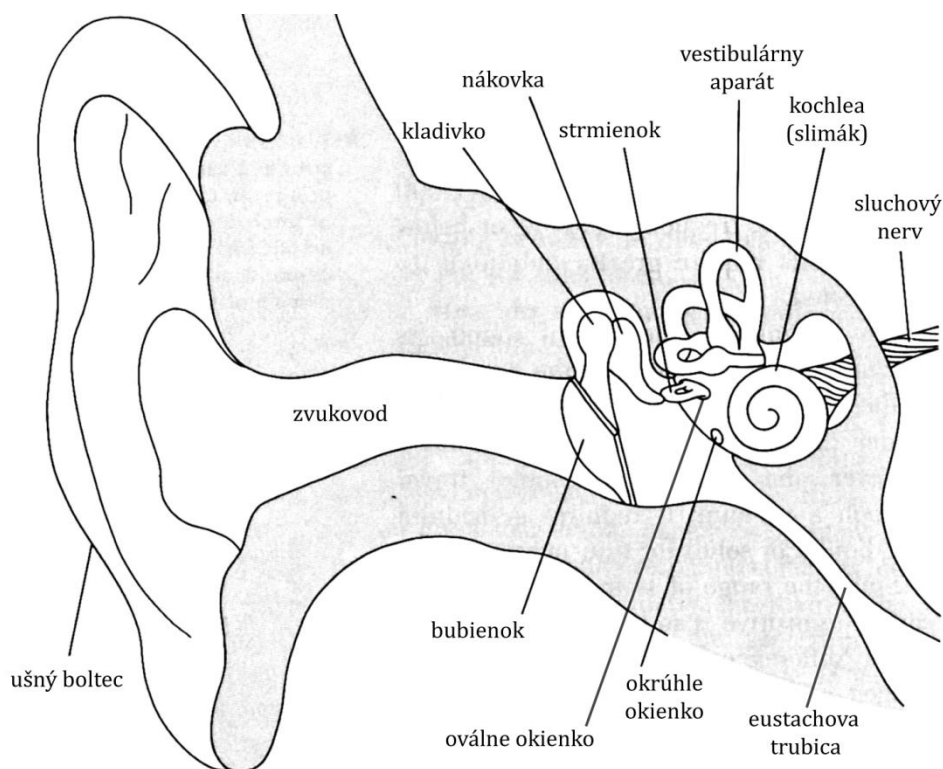
2 Ucho a kódovanie zvuku v sluchovom nerve

V nasledujúcej kapitole je vysvetlená problematika recepcie a percepcie akustických podnetov, teda príjem a vnímanie zvuku u človeka. Práve porozumenie týmto javom položilo základy myšlienky elektrostimulácie sluchového nervu, ktorá je využitá v kochleárných implantátoch.

2.1 Stavba sluchového orgánu a spôsob spracovávania zvukového podnetu

Vonkajšie ucho

Sluchový orgán – ucho, zobrazený na obrázku 2-1, sa skladá z vonkajšieho, stredného a vnútorného ucha. Akustický signál, ktorý sa šíri priestorom je zachytený a usmernený ušnicou a ďalej sa šíri zvukovodom až k membráne bubienka. Vonkajšie ucho hrá dôležitú úlohu tiež pri lokalizácii zvuku.



Obrázok 2-1 Sluchový orgán človeka a jeho časti. Na obrázku je znázornený prierez ľudským ušom, s vyznačením niektorých častí vonkajšieho, stredného a vnútorného ucha. (prevzaté z <http://www.psywww.com/intropsych>)

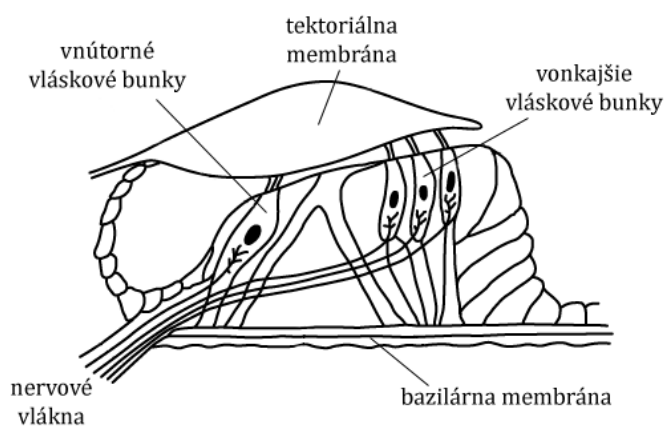
Stredné ucho

Bubienok, ktorý tvorí hranicu medzi vonkajším a vnútorným ušom, sa pri náraze zvukových vln rozkmitá a tieto vibrácie sú ďalej prenášané sluchovými kostičkami (kladivko, nákovka a strmienok) až k oválnemu okienku a ďalej do vnútra slímáka (kochlea). Pretože kochlea je naplnená tekutinou, je k prenosu vibrácií potrebné vyvinúť oveľa väčší tlak ako je hodnota akustického tlaku zvukového signálu. To je úloha sluchových kostičiek, ktoré zosilňujú

vibrácie bubienku, ale zároveň znižujú amplitúdu zákmitov. Sila zákmitov na oválnom okienku je až 20-krát vyššia ako zákmity bubienka vyvolané zmenami tlaku vzduchu. K strednému uchu patrí ešte Eustachova trubica spájajúca stredné ucho s hltanom, pomocou ktorej sa vyrovnáva tlak medzi stredným uchom a vonkajším prostredím. Pri veľmi silných vibráciách môže dôjsť k poškodeniu vnútorného ucha. V takejto situácii sa prejavuje tzv. stapediálny reflex, pri ktorom dochádza k sťahu špeciálnych svalov napojených na sluchové kostičky, čím sa celá sústava znehybní a prenos vibrácií na oválne okienko slimáka sa značne zníži, čo zabráni možnému poškodeniu častí vnútorného ucha.

Vnútorné ucho

Vnútorné ucho je zložené z kochley, ktorá obsahuje sluchové receptory, a labyrintu vestibulárneho aparátu, ktorého polkruhové kanáliky slúžia rovnovážnemu zmyslu. Ak by sme stočený tvar kochley narovnali, na priereze je možné vidieť, že kochlea obsahuje tri dutiny, scala vestibuli, scala media a scala tympani. Dutinu scala vestibuli a scala media oddeľuje veľmi pružná Reissnerova membrána. Dutiny scala media a scala tympani sú oddelené bazilárnou membránou, na ktorej je umiestnený tzv. Cortiho orgán obsahujúci sluchové receptory, ktoré sú zvrchnej strany chránené ďalšou (tektoriálnou) membránou. Dutiny scala vestibuli a scala tympani sú naplnené tekutinou tzv. perilymfou, zatiaľ čo scala media je vyplnená endolymfou. Obidve spomínané dutiny sú pri vrchole kochley prepojené otvorom ktorý sa nazýva helicotrema.



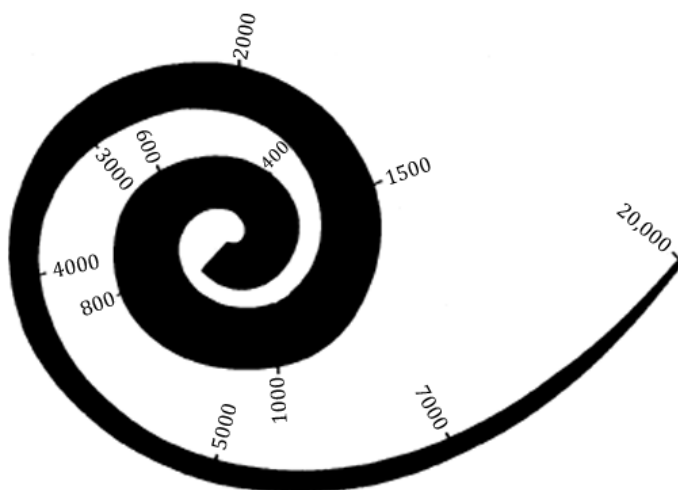
Obrázok 2-2 Cortiho orgán a vláskové bunky. Obrázok ukazuje rozmiestnenie vláskových buniek v Cortiho orgáne pozdĺž bazilárnej membrány v kochlee. (prevzaté z [3])

Vláskové bunky

Sluchové receptory - tzv. vláskové bunky, sú umiestnené rovnobežne v radoch, pozdĺž celej stredovej dutiny v kochlee. Cortiho orgán obsahuje dva druhy vláskových buniek, vonkajšie - majú vyššie zastúpenie a sú umiestnené v troch radoch, a vnútorné vláskové bunky v jednom rade. Prierez Cortiho orgánom je znázornený na obrázku 2-2. Názov vláskových buniek má pôvod v tvare jemných výbežkov (stereocilia) na vrchu každej z týchto buniek. Každá bunka obsahuje až 100 takýchto výbežkov. Práve tieto výbežky umožňujú generovanie akčného potenciálu tak, že pri ohybe, hoci aj veľmi miernom, sa otvoria pohybom ovládané draselné

kanály, ktoré bunku depolarizujú. Následná zmena potenciálu otvorí napätovo ovládané kalciové kanály, ktoré umožnia vypustenie neurotransmitteru na synaptické spoje gangliových buniek sluchového nervu.

Na zvuk reagujú časti vnútorného ucha nasledujúcim spôsobom. Strmienok cez oválne okienko rozkmitá tekutinu v kochlee, čo spôsobí vytvorenie vlny, ktorá sa šíri kochleou a vyvoláva kmitanie bazilárnej membrány. Bazilárna membrána sa smerom k vrcholu kochley (klenbe) rozširuje a zároveň sa znižuje jej tuhosť. To znamená, že vlny s vyššou frekvenciou rozkmitajú tuhšiu prednú časť membrány, čím sa rozptýli väčšia časť energie a vlnenie sa ďalej nedostáva. Naopak vlny s nižšou frekvenciou rozkmitajú bazilárnu membránu až pri klenbe kochley. Frekvenčný rozsah počuteľného spektra je tak daný vlastnosťami sluchového ústrojenstva. Zvuky s frekvenciou nižšou ako 20 Hz sú príliš slabé aby rozkmitali bazilárnu membránu, zatiaľ čo zvuky s frekvenciou vyššou ako 20 kHz sú utlmené už v strednom uchu. Rozloženie vnímania frekvencií na bazilárnej membráne je znázornené na obrázku 2-3.



Obrázok 2-3 Vzťah vnímanej frekvencie a pozície v rámci bazilárnej membrány. Zvuky s vysokou frekvenciou sú zaznamenávané receptormi v prednej časti membrány, zvuky s nižšou frekvenciou v oblasti klenby kochley. Znázornené rozloženie frekvencií (v Hz) pozdĺž membrány je nelineárne. (prevzaté z <http://www.medicine.mcgill.ca/physio/cullenlab>)

Pohyby bazilárnej membrány na určitej pozícii pozdĺž kochley vyvolávajú ohyby výbežkov vláskových buniek uchytených v pevnej tektoriálnej membráne a preto vláskové bunky v určitom mieste začnú generovať akčné potenciály. Pozícia, na ktorej dôjde k najvyššiemu vybudeniu vláskových buniek závisí priamo na frekvencii tónu, a preto hovoríme o tzv. tonotopickom usporiadaní buniek sluchových receptorov. Pre každú vláskovú bunku a jej odpovedajúce okolie bazilárnej membrány tak hovoríme o charakteristickej frekvencii. Tonotopické rozmiestnenie buniek môžeme nájsť okrem kochley aj v sluchovom nerve a vyšších sluchových vrstvách.

Priemerný počet vnútorných vláskových buniek je u človeka 3500, zatiaľ čo vonkajších buniek je 12000 až 20000. Napriek tejto disproporcii je až 95% neurónov sluchového nervu synapticky prepojených s vnútornými vláskovými bunkami. Zvyšných 5% inervuje vonkajšie vláskové bunky. Každá vnútorná vlásková bunka je spojená až s desiatimi neurónmi Cortiho gangliónu (spiral ganglion) a každý neurón, ktorý je prepojený s vnútornou vláskovou bunkou, je spojený len s touto jedinou vláskovou bunkou. Kvôli početnej prevahe je situácia

opačná pre vonkajšie vláskové bunky, kde jeden neurón Cortiho gangliónu inervuje niekoľko vláskových buniek. To znamená, že väčšina informácií prichádza do sluchového nervu len z vnútorných vláskových buniek a preto vonkajšie bunky musia mať ešte inú úlohu. Touto úlohou je zosilnenie vlnenia bazilárnej membrány pri stimulácii s nízkou intenzitou. Vonkajšie vláskové bunky sú schopné vďaka špeciálnemu proteínu (prestin) veľmi rýchlo reagovať na malé vychýlenie bazilárnej membrány stiahnutím alebo rozťahnutím bunkového tela a tým sa dosiahne až 100-násobne vyššie vychýlenie bazilárnej membrány. Tento aparát je regulovaný výstupnými nervovými vláknami z vyšších vrstiev sluchovej kôry.

Ďalšie vyššie časti sluchovej dráhy ako sú kochleárne a olivárne jadrá v mozgovom kmeni, inferior colliculus a sluchová kôra, popisovať nebudeme, pretože v nich dochádza k vyššiemu spracovaniu impulzov zo sluchového nervu a nie sú oblasťou pôsobenia kochleárneho implantátu.

2.2 Prirodzené kódovanie vlastností akustického signálu

Pre kódovanie zvuku v kochleárnych implantátoch je dôležité vedieť ako sú reprezentované základné veličiny, ktorými možno popísať akustický signál. Intenzita a frekvencia sú dve hlavné vlastnosti signálu, ktoré sú zakódované do postupností akčných potenciálov v sluchovom nerve. Tie sa kochleárny implantát snaží vyvolať elektrostimuláciou. Pretože pri šírení zvuku prostredím dochádza k odrazom a interferenciám vlnenia akustického signálu, ktoré spôsobuje fázový posun, nie je pri rozpoznávaní zvuku fáza až taká dôležitá a preto ani nie je táto vlastnosť zvukového stimulu nijak špeciálne v kochleárnych implantátoch kódovaná, pozornosť sa sústreďuje hlavne na intenzitu a frekvenciu signálu.

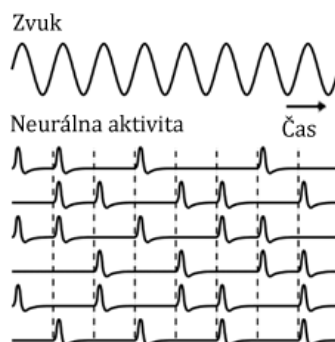
Frekvencia

Ľudské ucho je schopné vnímať zvuky o frekvencii približne od 20 Hz do 20 kHz. Frekvencia signálu sa prevedie na vlnenie bazilárnej membrány, vďaka ktorej vlastnostiam dochádza k jej vychýleniu len na určitom mieste. Môžeme teda povedať, že frekvencia je kódovaná do pozície vláskových buniek na bazilárnej membráne, ktoré neurálnu aktivitu vyvolávajú. Bližšie pri vstupe vlnenia, teda pri oválnom okienku, je bazilárna membrána tuhšia a vláskové bunky tu odpovedajú vyšším frekvenciám. Ďalej od tohto miesta smerom ku klenbe slimáka je bazilárna membrána pružnejšia a vláskové bunky odpovedajú nižším frekvenciám.

Pri spozorovaní neprítomnosti buniek odpovedajúcich veľmi nízkym frekvenciám pod 200 Hz v tonotopických mapách sluchového nervu sa zistilo, že frekvencia nemôže byť kódovaná len pozíciou vláskovej bunky alebo jej prislúchajúcemu neurónu v sluchovom nerve. Pre nízkofrekvenčné signály je informácia o frekvencii kódovaná v časovaní akčných potenciálov, ktorých frekvencia odpovedá frekvencii vlnenia. Ide o tzv. temporálne kódovanie frekvencie. Akustický signál a signál tvorený akčnými potenciálmi bunky je fázovo synchronizovaný. Znamená to, že vrcholy periodického signálu a akčné potenciály sú vždy v rovnakom fázovom posuve. Série pokusov [4] ukázali, že pri postupnom zvyšovaní vstupnej frekvencie

sa fázová synchronizácia zachováva, ale kvôli refrakčnej perióde frekvencia akčných potenciálov jednotlivých vlásokových buniek už nezodpovedá vstupnej frekvencii.

Zložený akčný potenciál sluchového nervu (na obrázku 2-4) je ale stále fázovo synchronizovaný, čo platí zhruba po hranicu 500 Hz. Pri ďalšom zvyšovaní frekvencie sa stráca aj synchronizácia zloženého akčného potenciálu. Tento druh kódovania frekvencie sa snažili využiť prvé jednokanálové kochleárne implantáty, a pretože na základné porozumenie reči je potrebné vnímať rozsah frekvencií aspoň do 4 kHz [4], nemohli byť takéto implantáty pri kódovaní reči úspešné.



Obrázok 2-4 Synchronizácia v sluchovom nerve. Zložený akčný potenciál (dole) je fázovo synchronizovaný so vstupným akustickým signálom (hore). (prevzaté z [4])

Intenzita

Akustický signál s vyššou intenzitou spôsobí jednak vyššie vychýlenie bazilárnej membrány a zároveň vychýlenie pokryje (aj vďaka zosíleniu pomocou vonkajších vlásokových buniek) väčšiu časť bazilárnej membrány a tým aj väčší rozsah charakteristických frekvencií. Väčšie vychýlenie membrány má za následok silnejší ohyb vlásokov na sluchových receptoroch, čo spôsobí väčšiu depolarizáciu a vyššiu frekvenciu generovaných akčných potenciálov. Intenzita je teda kódovaná jednak počtom aktívnych buniek, ale aj priemernou frekvenciou generovaných akčných potenciálov.

2.3 Vnímanie vlastností akustického signálu

Spolu so spomínanými parametrami zvuku úzko súvisí aj ich vnímanie, ktoré skúma oblasť psychoakustiky. Vzťah medzi týmito vnímanými hodnotami a im odpovedajúcimi veličinami nemusí byť lineárny, čoho príkladom sú aj spomínaná intenzita a frekvencia.

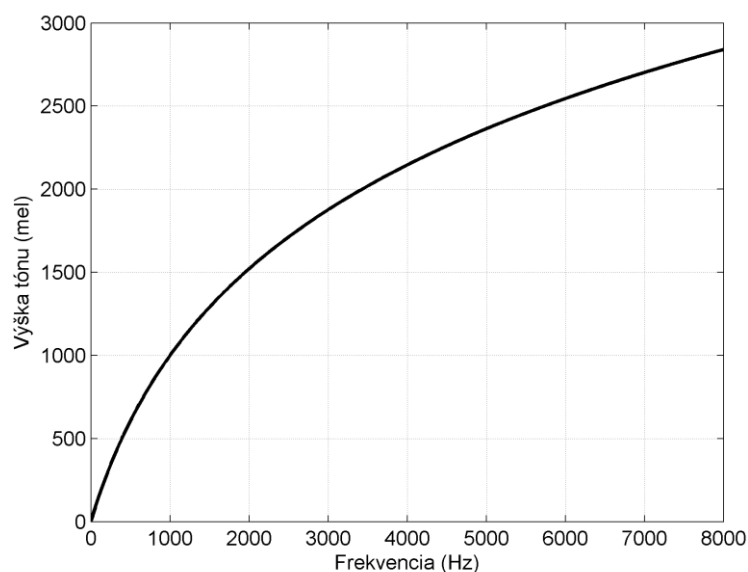
Výška tónu a frekvencia

Rozmiestnenie charakteristických frekvencií na bazilárnej membráne znázornené na obrázku 2-3 je nelineárne, s vyšším rozlíšením pre nižšie frekvencie. Človek vníma frekvenciu ako výšku tónu (pitch). Kvôli nelineárnej závislosti na frekvencii, a tiež aj na intenzite bola

zavedená tzv. melova stupnica (mel scale) s jednotkou mel [5]. Nárast o určitú hodnotu na tejto stupnici je priamo úmerný nárastu výšky tónu. Ako referenčná hodnota bola zvolená hodnota 1000 melov, ktorá odpovedá frekvencii 1000 Hz pri hodnote akustickej intenzity 40 dB. Závislosť medzi výškou tónu (f_{mel}) a frekvenciou (f) je daná nasledujúcim logaritmickým vzťahom [6]:

$$f_{mel} = 2595 \log_{10} \left(\frac{f}{700} + 1 \right) \quad (2-1)$$

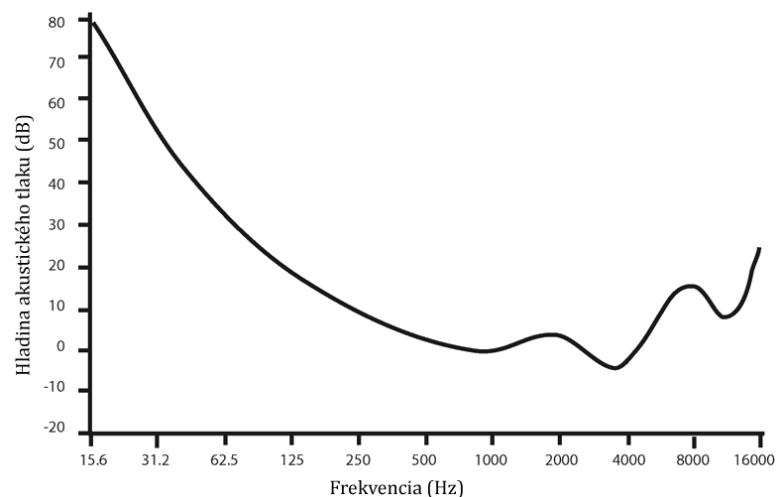
Na obrázku 2-5 je znázornená závislosť výšky tónu a hlasitosti, ktorá sa v praxi často aproximuje lineárnym vzťahom pre frekvencie do 1 kHz, a logaritmickým pre vyššie frekvencie.



Obrázok 2-5 Závislosť vnímanej výšky tónu na frekvencii. Na osi x je zobrazená frekvencia akustického signálu a na osi y zodpovedajúca vnímaná výška tónu v melovej stupnici.

Hlasitosť a intenzita

Intenzita akustického signálu je vnímaná ako hlasitosť, ktorá je podľa Weber-Fechnerovho zákona priamo úmerná logaritmu jeho intenzity. Hlasitosť môžeme vyjadriť pomocou hladiny akustického tlaku (sound pressure level – SPL). Vnímanie intenzity závisí tiež na frekvencii signálu a táto závislosť je znázornená na obrázku 2-6 pre hodnotu prahu počutia.



Obrázok 2-6 Závislosť vnímanej hlasitosti (prahu počuteľnosti) na frekvencii. Na osi x je zobrazená frekvencia akustického signálu a na osi y hladina akustického tlaku odpovedajúca ľudskému prahu počutia pre konkrétnu frekvenciu. (prevzaté z <http://en.wikipedia.org/wiki/Psychoacoustics>)

Na vyjadrenie intenzity akustického signálu sa často používa jednotka Bel. Je to bezrozmerná logaritmická jednotka, ktorou sa vyjadruje intenzita veličiny vzhľadom k nejakej referenčnej hodnote. V praxi sa ale používa čiastková jednotka decibel (dB). Použitie decibelovej stupnice pri meraní úrovni akustického signálu je vhodné hlavne preto, že rozsah hodnôt akustického tlaku vnímaného človekom je veľmi rozsiahly, a pomer medzi hodnotou prahu počuteľnosti (20 μ Pa rms) a prahom bolesti je preto veľmi veľký ($> 10^{12}$). Hodnota sa v decibeloch vyjadruje nasledujúcim vzťahom.

$$L_{dB} = 10 \log_{10} \left(\frac{P_1}{P_0} \right) \quad (2-2)$$

kde P_1 a P_0 sú intenzity meranej veličiny. Vo väčšine prípadov je intenzita veličiny rovná štvorcu amplitúdy, a preto ak sa v decibelovej stupnici merajú úrovne signálu, používa sa vzťah 2-3.

$$L_{dB} = 20 \log_{10} \left(\frac{A_1}{A_0} \right) \quad (2-3)$$

Maskovanie a kritické pásmo

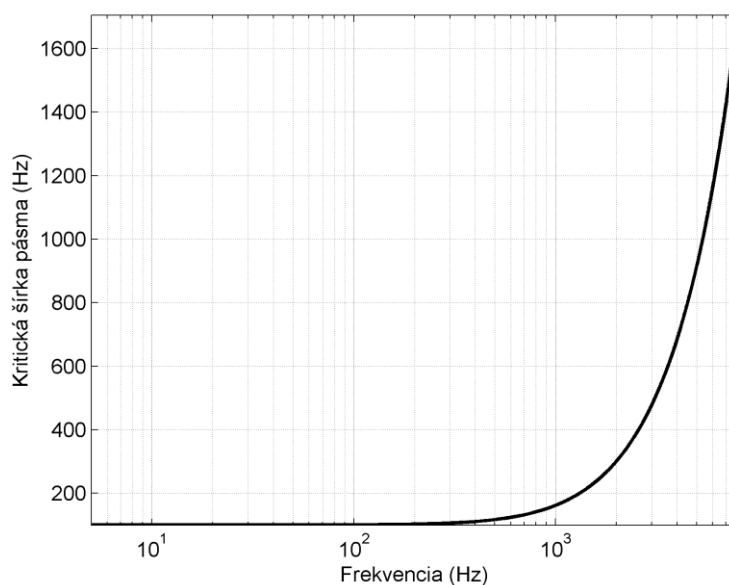
Pri akustickej stimulácii sluchového ústrojenstva súčasne dvoma zvukmi s podobnou frekvenciou dochádza k potlačeniu vnemu jedného zo zvukov. Tento jav sa nazýva maskovanie a dôvodom jeho vzniku je také vybudenie nervových buniek v okolí miesta (v rámci tonotopického usporiadania) prislúchajúceho frekvencii zvuku, že tieto bunky kvôli refrakčnej perióde nie sú schopné po určitú dobu reagovať na druhý zvuk. Ak máme daný čistý tón s určitou frekvenciou a signál bieleho šumu ktorý maskuje tento tón, potom kritickým pásmom je minimálna oblasť frekvenčného pásma šumu, ktorá sa podieľa na maskovaní. Šírku tohto pásma nazývame kritickou šírkou pásma, ktorá nám vyjadruje rozlišovaciu schopnosť sluchu pre jednotlivé tóny. Závislosť kritickej šírky pásma na

frekvencii (obrázok 2-7) vyjadruje vzťah z [6]:

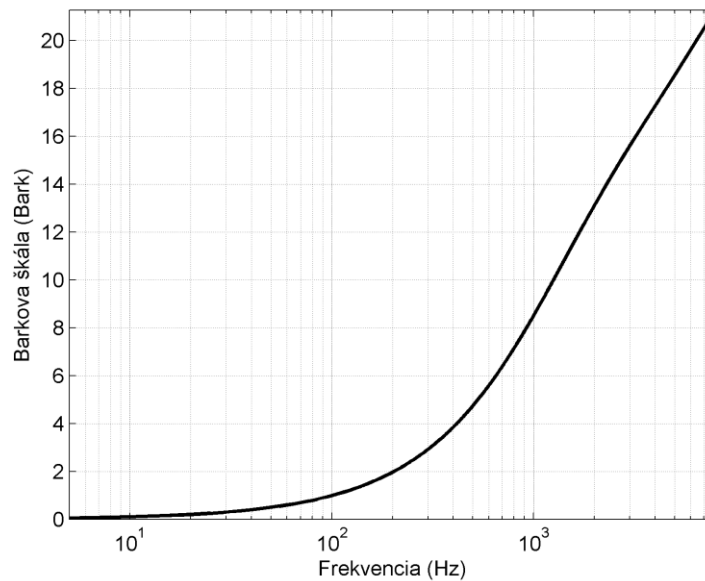
$$cbw = 25 + 75 \left(1 + 1.4 \left(\frac{f}{1000} \right)^2 \right)^{0.69} \quad (2-4)$$

Ak zoradíme jednotlivé kritické pásma postupne za seba, získame tak Barkovu [7] stupnicu s jednotkou Bark, ktorá zodpovedá vzdialenosti centrálnych frekvencií jednotlivých kritických pásiem. Barkova škála obsahuje prvých 24 kritických pásiem a prevod z nej do lineárnej frekvenčnej stupnice vyjadruje nasledujúci vzťah [6], ktorý je znázornený na obrázku 2-8.

$$f_{bark} = 13 \tan^{-1} \left(\frac{0,76f}{1000} \right) + 3,5 \tan^{-1} \left(\frac{f}{7500} \right)^2 \quad (2-5)$$



Obrázok 2-7 Závislosť kritickej šírky pásma na frekvencii. Na osi x je v logaritmickej škále znázornená frekvencia zodpovedajúca umiestneniu pozdĺž kochley (v rámci tonotopického usporiadania), a na osi y frekvenčný rozsah kritického pásma pre dané umiestnenie.

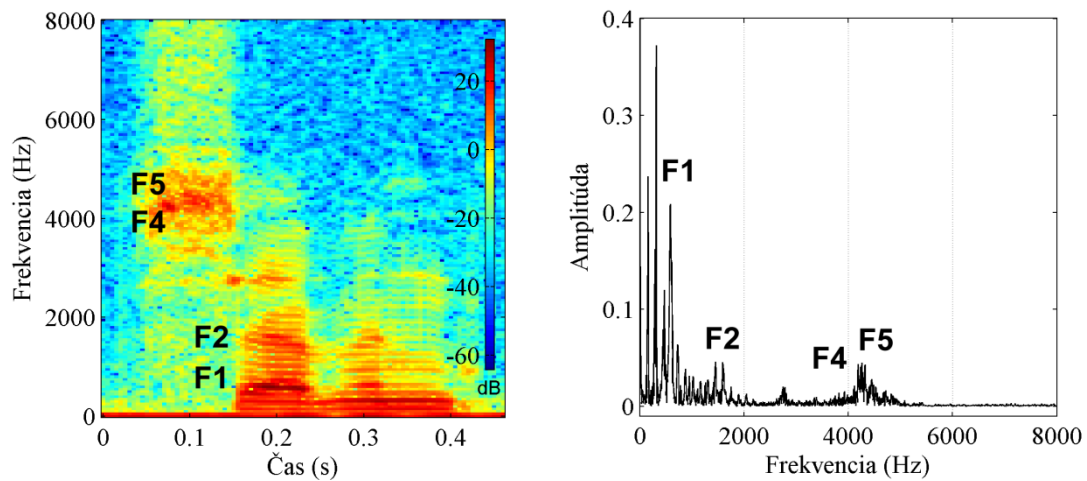


Obrázok 2-8 Barkova stupnica. Na osi y sú zobrazené centrálnne frekvencie pre prvých 24 kritických pásiem a na osi x im zodpovedajúce frekvencie v logaritmickú škále.

Formanty v signáli reči

Čím viac informácií zo vstupného signálu sa nám podarí v kochleárnom implantáte extrahovať a zakódovať, tým bude výsledok stimulácie pre pacientov zrozumiteľnejší a vernejší. Signály reči, na ktorých kódovanie sa implantáty sústreďujú najviac, môžeme charakterizovať tzv. formantmi. Formant definujeme ako koncentráciu energie akustického signálu v oblasti určitých frekvencií. Tieto koncentrácie sú spôsobené rezonanciami v hlasovom ústrojenstve človeka. Prítomnosť formantov môžeme sledovať v širokopásmovom spektrograme na obrázku 2-9 (vľavo), kde sú zobrazené ako miesta so sýtejšou farbou. Formanty označujeme číslom postupne od najnižších frekvencií. Špeciálnym prípadom je základný kmitočet F_0 , ktorý udáva základnú výšku tónu, ďalej je to prvý formant F_1 , ktorý sa objavuje v pásme približne 0 až 1 kHz, druhý formant F_2 má rozsah 1 až 3 kHz, atď.

Existujú dva rôzne varianty vzniku rečového signálu. Prvým zdrojom je periodické kmitanie hlasiviek modulované hlasovým traktom a ide o tzv. znelé časti reči, druhým je vysokofrekvenčný šum tvorený prúdom vzduchu vytláčaného z pľúc. Ten je taktiež modulovaný hlasovým ústrojenstvom a predstavuje neznelú časť reči. Z toho je zrejmé, že frekvenčné spektrá pre tieto dva druhy signálu budú odlišné. Zatiaľ čo znelé časti (samohlásky) obsahujú hlavne formanty nižšie (s nízkou frekvenciou), neznelé časti (spoluhlásky) obsahujú formanty vyššie, a niektoré nízke formanty ako F_0 a F_1 úplne chýbajú. Rozmiestnenie základných formantov sa tak stalo jedným z kritérií pri frekvenčnej analýze signálu v implantátoch.



Obrázok 2-9 Spektrogram rečového signálu anglického slova ‘seven’. Na obrázku vľavo je spektrogram vyznačujúci miesta signálu s najvyššou amplitúdou červenou farbou. Vyššie formanty F4 a F5 sú výrazné v začiatkovej, neznelnej časti slova. Nižšie formanty F1 a F2 sú charakteristické pre zvyšnú, znelú časť slova. Vpravo je zobrazené frekvenčné spektrum (s relatívnou amplitúdou) rovnakého signálu a jeho význačných formantov.

3 Spracovanie akustického signálu v kochleárných implantátoch

Pri niektorých senzorineurálnych poruchách sluchu sú poškodené len sluchové receptory – vlásokové bunky. Ak sú ale nervové bunky sluchového nervu a ich synaptické spojenia k vlásokovým bunkám v poriadku, je možné čiastočne nahradiť funkciu sluchových receptorov stimuláciou elektrickým prúdom. Zariadenie, ktoré je schopné takto stimulovať bunky sluchového nervu, sa nazýva kochleárny implantát. Implantát vlastne nahrádza funkciu celého vonkajšieho a stredného ucha až po vlásokové bunky v kochlee - všetky tieto orgány spoločne prevádzajú mechanickú energiu, ktorá vzniká vplyvom akustického tlaku na bubienok, na akčné potenciály. Úlohou implantátu je teda mikrofónom nasnímať zvuk z okolia a v reálnom čase ho transformovať na sériu elektrických impulzov, ktoré sú privedené k elektróde a stimulujú tkanivo v okolí nervových zakončení zo sluchového nervu. Stimulácia tkaniva, ktoré obsahuje aj nervové zakončenia sluchového nervu, spôsobí depolarizáciu v nervových bunkách a ak zmena potenciálu prekročí potrebný prah, otvoria sa napäťovo regulované sodíkové kanály a vygeneruje sa akčný potenciál.

Kochleárny implantát pozostáva z rečového (signálového) procesora a poľa elektród, ktoré sa vkladá do dutiny scala tympani v kochlee. Procesor sníma zvuk z okolitého prostredia a stará sa o kódovanie a prenos impulzov až k poľu elektród. Prenos medzi oboma časťami implantátu je bezdrôtový a zabezpečuje ho vysielacia cievka umiestnená nad prijímacou časťou implantátu pod pokožkou.

3.1 Digitálne spracovanie signálu

Jedným z najdôležitejších oborov využívaných pri návrhu kochleárných implantátov a ďalších zariadení, ktoré spracovávajú akustický signál je digitálne spracovanie signálu (digital signal processing – DSP). Táto oblasť poskytuje nielen matematické základy, nástroje a metódy, ale tiež algoritmy schopné riešiť úlohy spracovania signálu strojovo v diskretnom priestore hodnôt. Samotná oblasť DSP je veľmi obširná, a preto sú v nasledujúcej sekcii spomenuté len niektoré základy, potrebné pre pochopenie činnosti signálového procesora implantátu.

Digitálny signál a vzorkovanie

Signál definujeme ako hodnotu závislú na parametri, ktorým môže byť ľubovoľná veličina, vo väčšine prípadov je to ale čas. Analógový signál značíme $x_a(t)$, kde parameter t je reálne číslo. Vzorkovaním analógového signálu v pravidelných intervaloch dĺžky T_s a následnou normalizáciou časovej osi a kvantovaním, získame digitálny diskretný signál $x[n]$ definovaný v bodoch $n = 0, 1, 2, \dots$. Hodnotu $F_s = 1/T_s$ nazývame vzorkovacia frekvencia.

Pri digitalizácii signálu s použitím nízkej vzorkovacej frekvencie môže dôjsť k tzv. aliasingu, keď sa do frekvenčného spektra signálu dostávajú frekvencie, ktoré v pôvodnom signáli neboli a z takto navzorkovaného signálu nie je možné pôvodný signál rekonštruovať. Preto je

pri vzorkovaní nutné dodržať Nyquist-Shannonovo pravidlo [8], ktoré hovorí, že každý analógový signál $x_a(t)$ s obmedzeným frekvenčným pásmom a šírkou pásma F_0 je možné spätne rekonštruovať z jeho vzorkovania $x[n] = x_a(nT_s)$ práve vtedy, ak je vzorkovacia frekvencia aspoň dvakrát vyššia ako šírka pásma signálu, teda ak platí $F_s > 2F_0$. Hodnotu $2F_0$ nazývame tiež Nyquistova frekvencia. Ak nie je možné zvoliť takú vysokú vzorkovaciu frekvenciu (napr. kôli nedostatočnému výkonu zariadenia), aby sa aliasingu predišlo, používa sa tzv. anti-aliasing filter. Tento filter typu dolnej priepuste odstraňuje zo signálu frekvencie, ktoré by aliasing spôsobili.

LTI Systém

Ľubovoľnú transformáciu, ktorá modifikuje vstupný signál na výstupný signál, nazývame systém. Ak vstupom aj výstupom systému je digitálny signál, hovoríme o systémoch diskrétného času alebo diskrétnych systémoch, ktoré značíme operátorom zložených zátvoriek $T\{\cdot\}$, teda pre vstupný signál $x[n]$ (excitácia) a výstupný signál $y[n]$ (odpoveď systému) platí $y[n] = T\{x[n]\}$. Špeciálnym vstupným signálom je impulz:

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \quad (3-1)$$

Výstupný signál, ktorý zodpovedá tomuto impulznému vstupu voláme impulzná odpoveď $h[n]$.

Na spracovanie signálu sa v oblasti DSP používajú prevažne lineárne časovo-invariantné (linear time-invariant – LTI) systémy a to hlavne kvôli svojim vhodne zvoleným vlastnostiam. Ako už z názvu vyplýva, ide o systém invariantný voči násobeniu vstupného signálu konštantou a sčítaniu vstupných signálov (linearita) a navyše je invariantný aj voči posunutiu vstupnej sekvencie v čase. Vďaka týmto vlastnostiam vieme pre každý LTI systém so známou impulznou odpoveďou určiť odpoveď systému na ľubovoľnú excitáciu. Ľubovoľnú vstupnú sekvenciu môžeme rozložiť na jednotlivé prvky, ktoré reprezentujeme ako naškálované a v diskrétnom čase posunuté impulzy. Ak je teda vstupný signál rovný súčtu takto modifikovaných impulzov, bude výstup podľa vlastností LTI systému rovný súčtu rovnakým spôsobom modifikovaných impulzných odpovedí. Tento postup je vyjadrený nasledujúcim vzťahom, ktorý v DSP nazývame konvolúcia a označujeme symbolom $*$.

$$y[n] = T\{x[n]\} = h[n] * x[n] = \sum_{i=-\infty}^{\infty} h[n-i]x[i] \quad (3-2)$$

LTI systém môžeme vyjadriť tiež pomocou tzv. lineárnej diferenčnej rovnice s konštantnými koeficientami [6]. Pomocou tejto rovnice dokážeme popísať výstup systému v určitom čase ako hodnotu závislú nielen na aktuálnej hodnote vstupu, ale aj na predchádzajúcich hodnotách vstupu aj výstupu systému.

$$y[n] = \sum_{l=0}^L b_l x[n-l] - \sum_{k=1}^K a_k y[n-k], a_0 = 1 \quad (3-3)$$

Hovoríme o diferenčnej rovnici rádu K , ak $a_K \neq 0$. Podmnožinu LTI systémov, ktoré môžeme vyjadriť pomocou diferenčnej rovnice nazývame množina racionálnych systémov.

Filter

V oblasti digitálneho spracovania signálu nazývame LTI systém s určitými konkrétnymi vlastnosťami filter. Najčastejšie používané typy filtrov sú dolná, horná a pásmová priepusť alebo zádrž (high-pass, low-pass a band-pass, resp. band-stop filter). LTI systém s diskretným vstupným signálom (discrete-time LTI system) nazývame digitálny filter. Rozlišujeme dve základné skupiny digitálnych filtrov:

- *Filter s konečnou impulznou odpoveďou* označujeme tiež ako FIR (finite impulse response) filter. Ako už z názvu vyplýva, ide o filter ktorého impulzná odpoveď je konečná postupnosť a výstup filtra závisí len na vstupe. Ak použijeme na popis FIR filtra diferenčnú rovnicu 3-3, potom je výstup definovaný ako $y[n] = \sum_{l=0}^L b_l x[n-l]$, kde $b[l]$ je konečná postupnosť, $a_0 = 1$ a všetky ostatné koeficienty $a[k]$ sú rovné nule.
- *Filter s nekonečnou impulznou odpoveďou* – IIR (infinite impulse response) filter. Výstup systému je rekurzívne závislý na predchádzajúcich výstupoch a filter je určený diferenčnou rovnicou 3-3 s nenulovými prvkami postupnosti $a[k]$. Tento typ filtra sa označuje tiež ako autoregresívny filter. Pretože nekonečnú impulznú odpoveď nie je možné reprezentovať ako nekonečný zoznam hodnôt, v praxi sa používajú IIR filtre, ktoré zároveň patria do množiny racionálnych systémov.

Frekvenčná analýza systému

Pri analýze LTI systému nás často zaujímajú zmeny vstupného signálu vo frekvenčnej doméne. Tieto vlastnosti systému sú často najdôležitejšie pri návrhu zariadení pre filtrovanie signálu, kde je potrebné určité časti frekvenčného spektra zo signálu odobrať alebo modifikovať. Na analýzu vo frekvenčnej oblasti sa používa Fourierova transformácia diskretného signálu (discrete-time Fourier transform – DTFT) [8], ktorú pre vstupný signál $x[n]$, $\sum_{n=-\infty}^{\infty} |x[n]| < \infty$, definujeme nasledujúcim vzťahom, kde $\omega = 2\pi f$ je uhlová rýchlosť.

$$X(e^{i\omega}) = \mathcal{F}[x[n]] = \sum_{n=-\infty}^{\infty} x[n]e^{-i\omega n} \quad (3-4)$$

Zo symetrie a periodicity DTFT vyplýva, že pre účely frekvenčnej analýzy stačí vyčíslit hodnotu $X(e^{i\omega})$ len pre polovicu jej periódy. Pre konečnú reálnu postupnosť tak môžeme vyčíslit hodnotu DTFT priamo z definície, v bodoch z intervalu $[0, \pi]$. V praxi sa však často používa omnoho efektívnejšia metóda výpočtu pomocou algoritmu rýchlej Fourierovej transformácie (fast Fourier transform – FFT) [8].

Charakteristiku LTI systému vo frekvenčnej oblasti určuje tzv. frekvenčná odpoveď systému, čo je vlastne Fourierova transformácia impulznej odpovede systému, teda lineárna kombinácia bázy komplexných exponenciál. Ide o komplexnú funkciu, ktorú označujeme ako $H(e^{i\omega})$ a pre väčšiu názornosť sa v praxi zobrazuje ako amplitúda komplexných čísel v goniometrickom tvare. Takéto znázornenie je zobrazené napr. na obrázku 2-9 (vpravo).

Frekvenčnú odpoveď systému je možné odvodiť tiež z diferenciálnej rovnice systému tak, že za vstupný signál dosadíme postupnosť komplexných exponenciál a ako výstup použijeme súčin vstupu a frekvenčnej odpovede.

Dôležitým pozorovaním je to, že operácii konvolúcie signálov v časovej doméne odpovedá násobenie ich Fourierových transformácií vo frekvenčnej doméne. Vďaka dualite časovej a frekvenčnej domény vzhľadom k operácii Fourierovej transformácie zároveň platí aj opačná implikácia [8], a teda:

$$y[n] = a[n]b[n] \stackrel{\mathcal{F}}{\leftrightarrow} Y(e^{i\omega}) = A(e^{i\omega}) * B(e^{i\omega}) \quad (3-5)$$

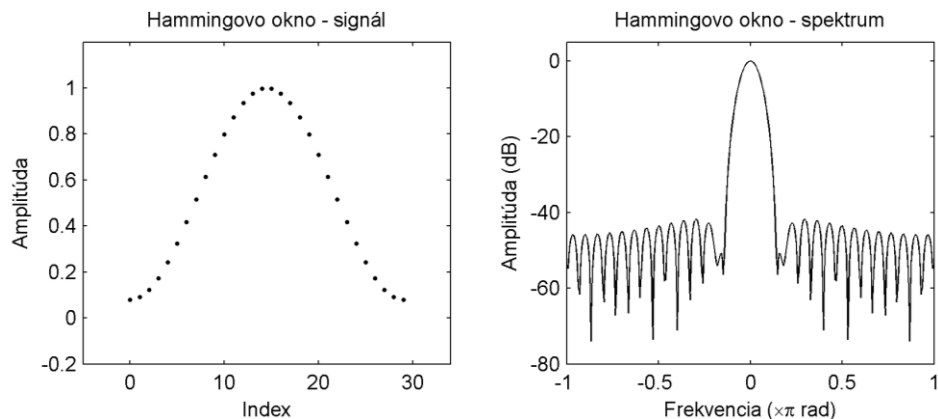
Krátkodobá frekvenčná analýza a spektrogram

Fourierova transformácia nám umožňuje zistiť spektrálnu charakteristiku signálu, teda zastúpenie rôznych frekvencií vo frekvenčnom spektre, ale túto informáciu nemôžeme upresniť v čase, pretože výstup frekvenčnej analýzy odpovedá časovému úseku medzi prvou a poslednou vzorkou vstupného diskrétného signálu. Ak potrebujeme zvýšiť rozlíšenie frekvenčnej analýzy v čase, použijeme metódu tzv. krátkodobej frekvenčnej analýzy, ktorá spočíva v tom, že výsledky frekvenčnej analýzy vyhodnocujeme osobitne pre jednotlivé časti vstupného signálu. Zobrazením výsledku krátkodobej frekvenčnej analýzy vzniká spektrogram, jeden zo základných spôsobov zobrazenia frekvenčného spektra signálu, ktorý obsahuje vedľa seba navrstvené frekvenčné spektrá odpovedajúce rovnomerne rozdeleným časovým intervalom signálu. Spektrogram signálu je znázornený napr. na obrázku 2-9. Na zjemnenie zobrazenia sa používa technika prekrývania, keď sa susedné časové intervaly prekrývajú a jeden interval tak prispieva do viacerých susedných frekvenčných spektier podľa miery prekrytia. V praxi sa používa často prekrytie 50% až 75% z dĺžky intervalu.

Jednotlivé časti signálu vyberáme postupne do fáz krátkodobej analýzy pomocou metódy okna. Okno definujeme ako diskretný signál s dĺžkou (počtom nenulových prvkov) menšou ako je dĺžka vstupného signálu. Vynásobením postupnosti signálu postupnosťou okna tak získame časť signálu na ktorej prevádzame frekvenčnú analýzu. Keďže násobenie v časovej doméne odpovedá konvolúcii spektier signálov v doméne frekvenčnej, prejaví sa aplikácia okna ako rozmazanie spektra pôvodného signálu frekvenčným spektrom okna. Výsledok krátkodobej frekvenčnej analýzy tak závisí na tvare frekvenčnej odpovede okna a jeho dĺžke.

- Pri výbere okna sledujeme šírku tzv. hlavného laloku a relatívny odstup a tvar bočných lalokov vo frekvenčnom spektre okna. Dôvodom vzniku týchto lalokov je tzv. leakage efekt [8], ktorý vzniká pri vzorkovaní DTFT pomocou diskretnéj Fourierovej transformácie. Čím užší je hlavný lalok, tým lepšie je okno schopné rozlíšiť jednotlivé frekvencie. Medzi najčastejšie používané typy okien patrí Hammingovo, Kasierovo, prípadne Hannovo okno. Na obrázku 3-1 je znázornený tvar Hammingovho okna a jeho frekvenčná charakteristika.
- Šírkou okna volíme kompromis medzi rozlíšením v čase a vo frekvencii. Ak použijeme dlhšie okno, frekvenčné rozlíšenie stúpa, ale rozlíšenie v čase sa znižuje, pretože nedokážeme presne určiť ktorej časti časového intervalu okna odpovedá výsledok

frekvenčnej analýzy. Čím je okno užšie, tým sa znižuje frekvenčné rozlíšenie a nemôžeme spracovávať nižšie frekvencie spektra, na druhej strane máme ale menší časový interval, ku ktorému sa krátkodobá frekvenčná analýza vzťahuje.



Obrázok 3-1 Hammingovo okno. Na obrázku vľavo je zobrazený tvar signálu Hammingovho okna. Vpravo je znázornené frekvenčné spektrum okna s relatívnou frekvenciou na osi x a amplitúdou v decibelovej škále na osi y.

3.2 Elektrostimulácia a jej vlastnosti

Väčšina kochleárných implantátov v súčasnosti používa impulznú elektrickú stimuláciu. To znamená, že vstupný signál je prevedený na postupnosť diskretných impulzov. Od používania impulzov spojitého charakteru sa upustilo práve kôli tomu, že nepretržitý tok elektrického prúdu pri stimulácii viacerých elektród naraz spôsoboval interferencie medzi týmito signálmi, ktoré by za normálnych okolností putovali oddelene v rôznych nervových vláknach. Impulzná stimulácia umožňuje veľmi rýchle striedanie impulzov tak, že vždy je aktívna len jedna elektróda a v jej okolí už nepôsobia žiadne iné elektródy ktoré by spôsobili interferenciu. Táto metóda stimulácie s časovo prekladanými stimulmi sa používa napríklad v stratégii CIS.

3.2.1 Vlastnosti impulznej stimulácie

Dôležitý je tvar, amplitúda, trvanie a frekvencia týchto impulzov. Počas obdobia od prvých implantátov až po súčasnosť sa vystriedalo niekoľko metód stimulácie, ktoré používali impulzy s rôznymi vlastnosťami. Medzi dôležité hodnoty, ktoré sa pri elektrostimulácii sluchového nervu sledujú, sú prahová hodnota stimulu (threshold level), ktorá udáva minimálnu hodnotu prúdu, ktorý je potrebný nato, aby subjekt zaznamenal tento impulz. Ďalšou meranou hodnotou je MCL (most comfortable level) úroveň, ktorá odpovedá sile impulzu s najvyššou hodnotou prúdu, pri ktorej subjekt ešte nepocítiťuje nepohodlie.

Tvar impulzov

Ako prvé boli používané jednofázové prúdové impulzy, ktoré sa ukázali byť nevyhovujúce. Napriek dobrým vlastnostiam týkajúcich sa generovaných impulzov nemohli byť používané na dlhodobú stimuláciu, pretože spôsobovali nežiaduce javy ako rozrušenie materiálu elektród, hromadenie náboja v okolí elektródy, vychýlenie hodnoty pH a pri vyšších hodnotách prúdu mohlo dôjsť k poškodeniu tkaniva a nervových buniek. V tejto oblasti sa chovajú omnoho lepšie tzv. nábojovo vyvážené impulzy dvojfázové impulzy. Tie začínajú katódovou fázou, ktorá spôsobí depolarizáciu a vyvolá akčný potenciál. Nasleduje anódová fáza, ktorá obnoví rovnováhu náboja narušenú prvou fázou. Dvojfázové impulzy nemusia byť nutne symetrické. Jediný predpoklad je, aby súčin amplitúdy a dĺžky trvania fázy bol rovnaký pre obidve fázy. V novších modeloch implantátov sa používajú tiež asymetrické impulzy, s krátkou katódovou fázou, s vysokou amplitúdou, za ktorou nasleduje dlhšia anódová fáza s menšou amplitúdou. Takéto impulzy nazývame pseudomonofázové a vykazujú nižšie prahové a MCL hodnoty. Tým sa dosiahne menšia spotreba zariadenia a teoreticky (kvôli nízkym hodnotám prúdu) aj menšia interferencia pri stimulácii viacerých elektród. Ďalšou dôležitou vlastnosťou dvojfázových impulzov je tzv. medzifázová pauza. Ak by anódová fáza nastúpila hneď za katódovým impulzom, mohlo by dôjsť k repolarizácii membrány nervovej bunky ešte pred samotným dosiahnutím prahovej hodnoty a akčný potenciál by sa nevygeneroval.

Magnitúda stimulácie

Veľkosť prúdu ktorým je tkanivo stimulované je závislé na tvare jednotlivých impulzov. Tie však bývajú často pevne stanovené pre konkrétnu stratégiu kódovania a počas stimulácie sa mení amplitúda prúdu, ktorá priamo reprezentuje hlasitosť rozpoznávaného zvuku. Čím je intenzita prúdu vyššia, tým je viac stimulovaných nervových buniek a je generovaných viac akčných potenciálov. Bežné hodnoty prúdu pre jednotlivé impulzy sú od 10 μA do 1700 μA . Aby bolo možné vnímať jednotlivé kroky hlasitosti rovnomerne, je tento rozsah rozdelený rovnomerne na logaritmickú stupnicu.

3.2.2 Tonotopické usporiadanie a počet elektród

Pri použití stimulácie za pomoci diskretných impulzov udáva frekvencia impulzov frekvenciu generovaných akčných potenciálov. To platí len zhruba do frekvencie 1 kHz a preto sa na kódovanie frekvencie používa tonotopické usporiadanie vlákien sluchového nervu. To viedlo k vývoju kochleárnych implantátov s poľom elektród ktoré sú pravidelne rozmiestnené pozdĺž celej bazilárnej membrány a môžu stimulovať každú jej časť. Prvé jednoelektródové (jednokanálové) implantáty využívali fakt, že pre veľmi nízke frekvencie (pod 200 Hz) nezáleží na oblasti v ktorej sluchový nerv stimulujeme a subjekt rozpoznáva impulz vždy ako tón s frekvenciou zodpovedajúcou frekvencii impulzov. Pre kochleárne implantáty s viacerými elektródami (multikanálové) je frekvencia zvuku kódovaná do pozície aktívnej elektródy. Frekvencia stimulácie daného úseku je často pre konkrétnu metódu kódovania nastavená pevne a je obmedzená už len dobou (absolute refractory period), počas ktorej nervová bunka nemôže vygenerovať nový akčný potenciál. Po tejto fáze je bunka ešte určitý

čas (relative refractory period) v stave, keď je potrebné ju stimulovať impulzom s vyššou amplitúdou, aby sa vytvoril akčný potenciál.

Jednokanálová stimulácia

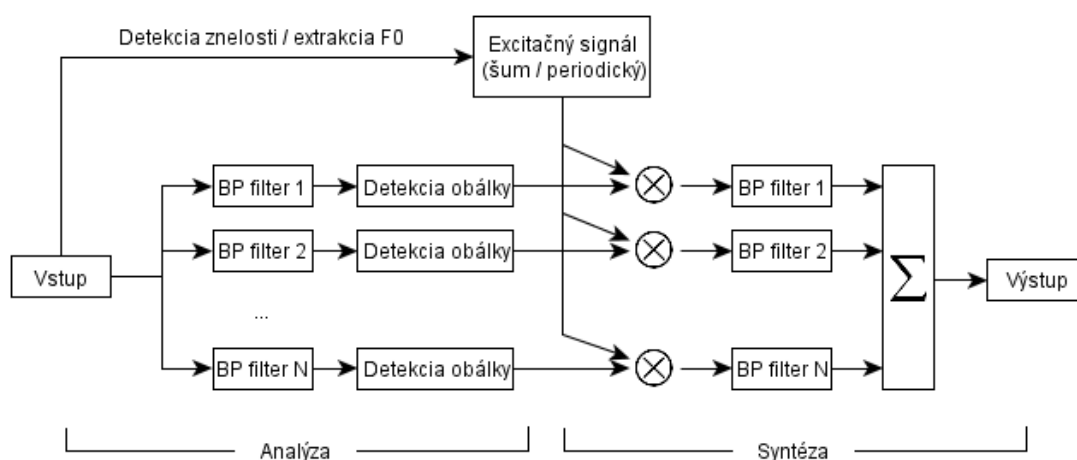
Prvé verzie kochleárných implantátov používali iba jednu stimulačnú elektródu, pretože sa predpokladalo, že všetky dôležité informácie dôležité pre rozpoznanie reči sú zakódované do časového priebehu stimulov. Ďalej sa vychádzalo z poznatku, že pre rozpoznanie reči sú najdôležitejšie frekvencie v pásme od 100 Hz do 4 kHz. Vstupný signál z mikrofónu prechádzal kompresným filtrom, ktorý umožňuje upraviť signál podľa dynamického rozsahu pre konkrétneho pacienta. V nasledujúcej fáze sa na signál aplikuje ekvalizačný filter, ktorého úlohou je tlmenie nízkych a posilnenie vyšších frekvencií, pretože prahové hodnoty prúdu sú nižšie pre nízke a vyššie pre vysoké frekvencie. V ďalšej fáze signál prechádza modulačným filtrom a je bezdrôtovo prenášaný z cievky rečového procesora až k implantátu, kde je opäť demodulovaný a použitý priamo na stimuláciu elektródy. Tento typ implantátu sa v praxi veľmi neosvedčil hlavne kvôli tomu, že na kódovanie výšky tónu používa iba frekvenciu stimulov a nevyužíva tonotopické usporiadanie sluchového nervu. Ďalej nezohľadňuje ani maximálnu možnú frekvenciu stimulácie danú refrakčnou periódou nervových buniek.

Multikanálová stimulácia

Vzhľadom na problémy jednokanálovej stimulácie používajú všetky implantáty v súčasnosti multikanálovú stimuláciu pomocou poľa elektród. Metódy využívajúce viacero elektród využívajú filtrovanie frekvenčného spektra vstupného signálu do jednotlivých úsekov, ktoré zodpovedajú frekvenčnému pásmu stimulovanému konkrétnou elektródou. Parametre filtrov a ich rozmiestnenie v rámci sluchom vnímaného frekvenčného spektra môže byť pevné alebo sa filtre môžu adaptovať podľa potreby. Medzi stratégie kódovania, ktoré využívajú druhý zo spomínaných prístupov patria metódy založené na modeloch sluchového nervu a vychádzajú z teórie spracovania reči a extrahujú časti spektra, ktoré sú význačné pre rozpoznanie reči.

4 Stratégie kódovania v kochleárných implantátoch

Takmer všetky signálové procesory kochleárných implantátov súčasnosti sú založené na technológii kódovania hlasu, ktorá vznikla už v tridsiatych rokoch minulého storočia a jej autorom je Homer Dudley, inžinier laboratórií Bell. Pôvodným cieľom bolo vytvoriť technológiu, ktorou by bolo možné zakódovať reč pred jej prenosom cez telefónnu linku. Výsledkom bol tzv. vokodér (z angl. voice-encoder), hlasový kodér a dekodér, ktorý umožnil zakódovanie analógového zvukového signálu, ktorý v tomto stave spotreboval len časť šírky pásma komunikačnej linky. Tak bolo možné zvýšiť kapacitu linky, na ktorej mohlo prebiehať viacero hovorov zároveň. Na obrázku 4-1 je znázornená schéma vokodéra a princíp jeho funkcie.



Obrázok 4-1 Schéma vokodéra (hlasového kodéra/ dekodéra). Schéma zobrazuje jednotlivé moduly spracovania signálu v časti analýzy aj spätnej syntézy zvuku. Vstupom aj výstupom vokodéra je zdigitalizovaný signál.

Zakódovanie (fáza analýzy) začína filtrovaním vstupného signálu pomocou sústavy analógových band-pass filtrov, ktorých prepúšťané pásma na seba plynule nadväzujú tak, že je pokrytá potrebná časť spektra signálu. Filtráciou dostávame n signálov (n je počet filtrov sústavy) s obmedzenou šírkou pásma, z ktorých sa extrahuje obálka signálu, ktorá je po decimácii a kvantizácii prenášaná linkou. Na extrakciu obálky signálu je použité usmernenie a filtrovanie signálu dolnou priepusťou (low-pass filtrom). Spolu s obálkami signálov sa prenáša aj informácia o tom, či ide o znelú alebo neznelú časť reči a pre znelé časti aj výška tónu, ktorú získame extrakciou formantu F_0 . Tieto informácie sa použijú pri dekódovaní signálu.

Vo fáze dekódovania (syntézy) sú obálky použité na amplitúdovú moduláciu excitačných signálov. Pre znelé časti reči je excitačným signálom postupnosť impulzov s periódou rovnou základnej perióde reči, ktorá bola extrahovaná pri kódovaní reči. Pre neznelé časti reči je excitačným signálom biely šum.

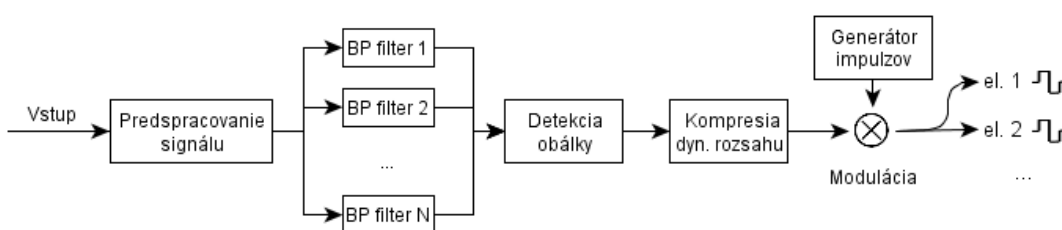
Kochleárne implantáty, na rozdiel od vokodéra, nemajú dekódovaciu časť. Ich výstupom sú postupnosti elektrických impulzov s amplitúdou modulovanou obálkami signálu z fázy kódovania. Tieto impulzy sa následne vysielajú do poľa elektród implantátu, ktoré potom

elektricky stimulujú tkanivo v okolí vlásokových buniek, a tým aj sluchový nerv. Každá elektróda je zásobovaná signálom odpovedajúcim konkrétnemu filteru vo fáze analýzy a umiestnenie elektródy v kochlee približne odpovedá stredovej frekvencii filterom prepúšťaného pásma. Vďaka tonotopickému usporiadaniu vlásokových buniek v kochlee to znamená, že stimulácia každej z elektród poskytuje pacientovi s kochleárnym implantátom informáciu z konkrétneho frekvenčného pásma. Signálové procesory v kochleárných implantátoch pracujú na rozdiel od pôvodného vokodéra s digitálnym signálom, čo umožnilo zjednodušiť proces výroby týchto zariadení a zmenšiť ich rozmery.

Dve v súčasnosti najpoužívanejšie stratégie kódovania v kochleárných implantátoch sú *Continuous Interleaved Sampling (CIS)*, a *Spectral Peak (SPEAK)*. Obidve metódy vychádzajú z myšlienky rozkladu vstupného signálu použitej v Dudleyho vokodéri, a pretože ide o stratégie rozšírené aj v komerčných implantátoch, stali sa základnými stratégiami v aplikácii simulácie kochleárneho implantátu, ktorej implementácia je súčasťou tejto práce. Aj keď sa v praxi stretávame s množstvom rôznych konfigurácií parametrov pre tieto konkrétne metódy, princíp ich funkcie sa pre jednotlivé varianty týchto metód väčšinou nemení a podrobnejšie ho popíšem v nasledujúcich odsekoch.

4.1 Continuous interleaved sampling (CIS)

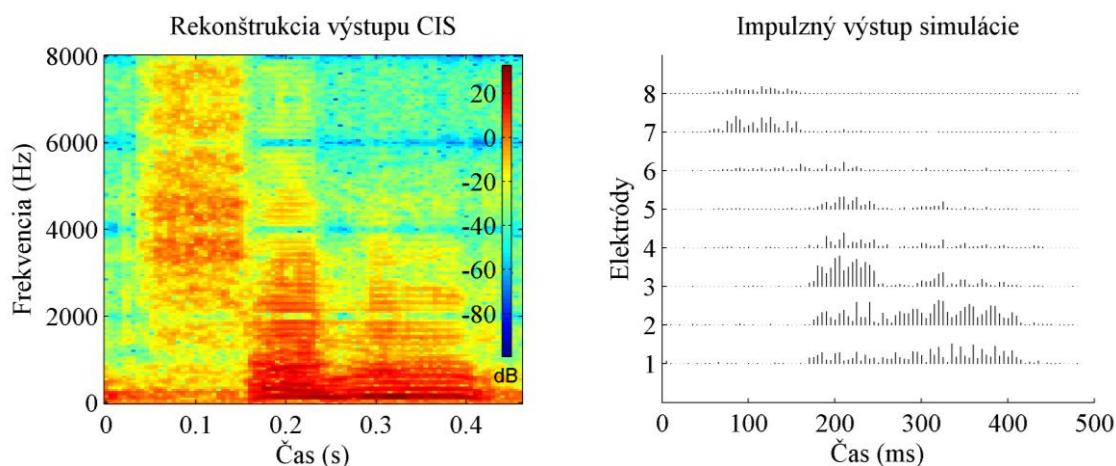
Jedna z prvých úspešných stratégií používaných v kochleárných implantátoch s multikanálovou stimuláciou je stratégia CIS, ktorej schéma je znázornená na obrázku 4-2. Vstupný signál z mikrofónu prechádza najprv modulom predspracovania, ktorý sa stará o odstránenie jednosmernej zložky zo vstupného signálu, zvýraznenie hornej časti frekvenčného spektra, prípadne o automatické zosilnenie vstupného signálu. Ďalšou a kľúčovou časťou je banka filterov, ktorou sa frekvenčné spektrum signálu rozdelí do pásiem odpovedajúcim jednotlivým elektródam implantátu a ďalšie fázy spracovania prebiehajú osobitne v každom z týchto pásiem. Nasleduje extrakcia obálky signálu a aplikácia kompresnej funkcie, ktorá má za úlohu na seba naškálovať dynamické rozsahy akustickej a elektrickej stimulácie. Takto získanou obálkou signálu je následne modulovaný výstup z generátora elektrických impulzov.



Obrázok 4-2 Schéma stratégie kódovania CIS. Schéma zobrazuje základné bloky kódovania zvuku v implantáte. Vstupom je digitalizovaný signál zvuku, výstupom sú elektrické impulzy, ktoré stimulujú sluchový nerv.

Pretože kochleárny implantát spracováva vstupný signál v reálnom čase a musí spĺňať požiadavky na minimálnu latenciu výstupného signálu, prebieha spracovanie signálu po

krátkych blokoch, ktoré sú postupne načítavané zo vstupného zariadenia. Celá schéma kódovania sa tak opakovane aplikuje na každý zo vstupných blokov signálu. Výstup CIS stratégie v podobe rekonštruovaného rečového signálu, ale aj ako impulzný výstup pre elektródy implantátu je zobrazený na nasledujúcom obrázku.



Obrázok 4-3 Výstup simulácie stratégie CIS (rekonštrukcia a elektrické impulzy). Na obrázku vľavo je zobrazený spektrogram rekonštruovaného zvukového signálu, ktorý je výstupom simulácie. Vstupný zvukový signál slova 'seven' je zobrazený na obr. 2-9. Vpravo je zobrazený impulzný výstup (prúdové amplitúdy) simulácie v aplikácii „CIsim“ pre 8 elektród.

4.1.1 Predspracovanie signálu

Do oblasti predspracovania signálu patrí odstránenie nenulovej jednosmernej zložky signálu, ktorá sa tiež označuje ako zložka s nulovou frekvenciou (direct current offset). V podstate ide o nenulovú strednú hodnotu amplitúdy signálu, ktorá je pri spracovaní audio signálu nežiadúca. Tento jav vzniká už pri nahrávaní signálu v hardware zariadení a spôsobuje problémy napríklad pri normalizácii amplitúdy signálu, keď sa kvôli tomuto posunu signál nesprávne naškáluje. V prípade kochleárných implantátov zase musí platiť, že ak nie je zachytený žiadny akustický signál, nesmie byť do elektród vysielaný žiadny stimulačný signál. Na odstránenie jednosmernej zložky sa používa najčastejšie high-pass filter, ktorý zo spektra signálu vynechá okrem jednosmernej zložky aj časť s veľmi nízkymi frekvenciami (tzv. slowly changing DC).

Ďalšou časťou predspracovania v kochleárných implantátoch je zvýraznenie vyšších frekvencií spektra rečového signálu tak, aby bolo rozdelenie energie v rámci celého rečového spektra čo najviac rovnomerné. Frekvenčné spektrum rečového signálu je špecifické útlmom 6 dB na oktávu [9], čo znamená, že amplitúda signálu klesá o polovicu pri každom ďalšom zdvojnásobení frekvencie. Tento útlm sa rieši v praxi filtrom predbežnej korekcie (pre-emphasis filter), ktorý má krivku frekvenčnej odpovede podobnú rastúcej logaritmickému funkcii, takže amplitúdy nízkych frekvencií sú zmenšené a naopak.

Aby bolo možné používať implantát v bežnom prostredí kde sa neustále mení úroveň hlasitosti akustického signálu, ktorý je navyše zaťažený šumom z prostredia, obsahujú procesory implantátov tiež moduly na automatickú reguláciu intenzity (Automatic Gain

Control) vstupného signálu a na redukciu šumu (Noise reduction). AGC moduly využívajú reguláciu na princípe spätnej väzby a regulujú zosilnenie vstupu na základe porovnania aktuálnej amplitúdy signálu s priemernou hodnotou.

4.1.2 Frekvenčná analýza signálu

Každá elektróda kochleárneho implantátu pôsobí na oblasť bazilárnej membrány, ktorá odpovedá určitému frekvenčnému pásmu. Najdôležitejšou informáciou, ktorú je treba zo vstupného signálu získať, je teda rozloženie energie signálu v rámci frekvenčného spektra. Množstvo energie obsiahnuté v pásmach spektra, ktoré odpovedajú elektródam sa potom využije na moduláciu impulzov ktoré sú posielané do elektródového poľa. Jedným zo spôsobov ako analyzovať distribúciu energie je použitie tzv. banky filtrov, teda sústavy band-pass filtrov, ktoré pokrývajú časť frekvenčného spektra obsahujúceho významné vlastnosti rečového signálu. Práve pásmo od 100 Hz do 8000 Hz obsahuje podstatnú časť informácie rečového signálu [10], preto aj väčšina implantátov dokáže pracovať so signálom s vzorkovacou frekvenciou 16 kHz, takže môžu analyzovať spomínané pásmo po hranicu 8 kHz, čo je Nyquistova hodnota pre danú vzorkovaciu frekvenciu.

Rôzne typy implantátov používajú rôzne varianty banky filtrov. Počet filtrov v banke určuje rozlíšenie stimulácie a tým aj ovplyvňuje kvalitu elektricky evokovaného počutia. Čím viac filtrov stimuluje im prislúchajúce elektródy, tým viac informácie zo signálu dodávame sluchovému nervu. Jeden z prvých implantátov so stratégiou CIS používal rozklad len do štyroch pásiem, súčasné implantáty používajú rozklad do 8, 12, až 16 pásiem [11]. Čo sa týka šírky spracovávaného pásma a šírky pásiem jednotlivých filtrov, používa sa nelineárne rozmiestnenie vychádzajúce z percepcie frekvencie. Najčastejšie používanými typmi rozmiestnenia centrálnych frekvencií filtrov je logaritmické, ďalej je to rozmiestnenie vychádzajúce s melovej a Barkovej škály. Rozmiestnenie podľa Barkovej škály, často označované ako banka filtrov kritického pásma (critical-band filter bank), je v niektorých prípadoch aproximované lineárnym rozmiestnením po hranicu 1300 Hz a nad touto hranicou je použité logaritmické rozdelenie.

Experimentálne sa preukázalo [11], že jedným z faktorov pre porozumenie reči je pokrytie frekvenčného rozsahu formantov F1 (0 – 1 kHz) a F2 (1 – 3 kHz). Najlepšie pokrytie týchto oblastí má banka filtrov kritických pásiem, ktorá oblasť formantu F1 pokrýva až šiestimi filtrami, zatiaľ čo rozmiestnenie podľa melovej stupnice obsahuje v oblasti pre F1 len 4 pásma. Pre oblasť formantu F2 je banka filtrov kritických pásiem tiež lepšia v porovnaní s ostatnými typmi rozmiestnenia. Viaceré súčasné komerčné implantáty používajú rozdelenie frekvenčného pásma s ohľadom na dobré pokrytie pásma nižších formantov, čo prinieslo výrazné zlepšenie porozumenia samohlások, ktoré sú charakteristické prítomnosťou nižších formantov.

4.1.3 Detekcia obálky

Po vytvorení série pásmovo obmedzených signálov, potrebujeme získať informáciu o sile signálu v čase, ktorú použijeme na moduláciu elektrických stimulov. Každý z týchto signálov môžeme vyjadriť nasledujúcim súčinom:

$$x_i(t) = a_i(t) \cos(\omega t + \phi_i(t)) \quad (4-1)$$

kde výraz $a_i(t)$ nazývame obálkou signálu, ktorá nám poskytuje hľadanú informáciu a druhý činiteľ tzv. jemnú štruktúru signálu. V kochleárných implantátoch sa používajú dve metódy na extrakciu obálky. Prvý algoritmus zahŕňa usmernenie signálu a filtrovanie low-pass filtrom. Používa sa buď dvojcestné (full-wave) alebo jednocestné (half-wave) usmernenie. Dvojcestné usmernenie je vlastne absolútna hodnota signálu a pri jednocestnom usmernení vynulujeme všetky vzorky signálu so zápornou hodnotou. Následné filtrovanie dolnou priepusťou nám vyhladí signál a vytvorí odhad obálky. Low-pass filter, ktorý slúži zároveň aj ako tzv. antialias filter, musí mať ale hraničnú frekvenciu rovnú najviac polovici hodnoty (Nyquistova hodnota) frekvencie stimulácie. V praxi sa používajú na extrakciu obálky filtre s hraničnou frekvenciou 200 až 400 Hz. Niektoré výskumy [11] v oblasti psychofyziky dospeli k záveru, že frekvencia stimulácie by mala byť až 4-krát vyššia ako hraničná frekvencia dolnej priepuste pre detekciu obálky. Čím je frekvencia vyhladzovacieho filtra vyššia, tým je odhad obálky presnejší.

Druhou metódou detekcie je použitie tzv. Hilbertovej transformácie, ktorou je možné vyjadriť signál ako súčin daný vzťahom 4-1. Obálka získaná Hilbertovou transformáciou je presnejšia ako odhad usmernením a vyhladením, ale pre vyššie hraničné frekvencie vyhladzovacieho filtra sa rozdiel stráca.

4.1.4 Kompresia amplitúdy obálky

Neoddeliteľnou súčasťou modulu pre detekciu obálky je kompresia amplitúdy obálky. Dôvodom je veľký nepomer medzi dynamickým rozsahom hodnôt obálky akustického signálu a amplitúdy elektrických impulzov. Dynamický rozsah je definovaný ako rozsah od minimálnej rozoznávanej hodnoty, tiež prahovej hodnoty (threshold level), po maximálnu znesiteľnú hodnotu (maximum comfortable level). Hodnoty dynamického rozsahu sú obyčajne stanovené experimentálne a preto sa často v rôznych zdrojoch líšia. Dynamický rozsah akustického signálu je 30 až 50 dB SPL [11] (40 až 70 dB SPL podľa [12]). Pre elektrickú stimuláciu je dynamický rozsah len 5dB [11] (8dB podľa [12]). Tento nepomer sa rieši použitím kompresnej funkcie, ktorá sa aplikuje na obálku každého z analyzovaných frekvenčných pásiem. Jednotlivé kanály implantátu majú vlastný kompresný modul, takže je možné stanoviť mieru kompresie osobitne pre každú elektródu.

Najčastejšie používanou funkciou je logaritmická, pretože vystihuje logaritmický vzťah medzi hlasitosťou akustických stimulov v decibeloch a elektrických impulzov v mikroampéroch. Logaritmická kompresná funkcia je daná nasledujúcou rovnicou:

$$y = a \log_{10}(1 + kx) + b, \quad k > 0 \quad (4-2)$$

kde y je amplitúda elektrického signálu, x je amplitúda obálky akustického signálu, k je tzv. kompresný koeficient a parametre a a b sú volené tak, aby výsledné hodnoty spadali do cieľového dynamického rozsahu.

Druhou používanou kompresnou funkciou je univerzálnejšia mocninná funkcia daná rovnicou 4-3, ktorá je používaná aj v iných oblastiach digitálneho spracovania signálu. Vlastnosti kompresie môžeme modifikovať parametrom k .

$$y = ax^k + b, \quad k \in (0, 1) \quad (4-3)$$

Pri pokusoch v štúdií [11], s použitím mocninnej kompresie, sa najviac osvedčili funkcie s vyššími kompresnými vlastnosťami a naopak takmer lineárne kompresné funkcie poskytovali najhoršie rozpoznanie vstupného zvukového podnetu.

4.1.5 Generátor elektrických impulzov

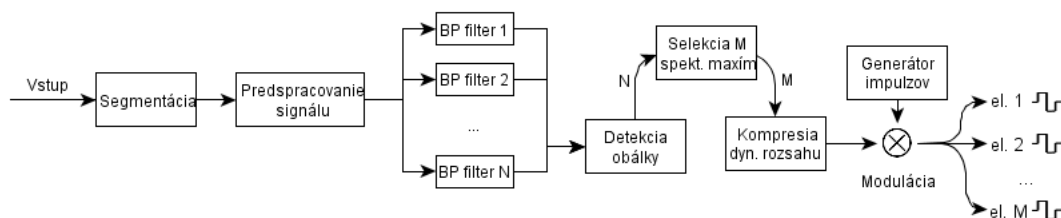
Modul generátoru elektrických impulzov implantátu sa stará o kontinuálne generovanie elektrických impulzov modulovaných obámkami signálov spracovávaných pásiem a dodáva ich priamo do elektródového poľa implantátu. Hlavným parametrom, ktorý sledujeme, je frekvencia impulzov (stimulation rate), od ktorej sa odvádza aj maximálna možná šírka fázy impulzov. CIS stratégia používa stimuláciu dvojfázovými impulzmi a frekvenciou stimulácie od 250 až po 5000 impulzov za sekundu (označujeme pps, z angl. pulses per second). Čím je počet impulzov za sekundu vyšší, tým väčšie rozlíšenie stimulácia v časovej oblasti poskytuje. To ale nemusí platiť vždy, pretože napríklad pri veľmi vysokých hodnotách pps a nízkej vzorkovacej frekvencii vstupného akustického signálu, či nízkom rozlíšení frekvenčnej analýzy sa duplikujú hodnoty amplitúd impulzov a do stimulácie sa zvýšením pps neprináša nová informácia o signáli. Aj keď hodnota pps bola predmetom výskumu viacerých štúdií (napr.: v [11] sa preukázalo že stimulácia s hodnotou 2100 pps umožňuje lepšie rozpoznanie slov ako stimulácia s 800 pps), nepodarilo sa nájsť univerzálnu hodnotu, ktorá by uspokojovala všetky kritériá. Voľba frekvencie stimulácie závisí na použitom prístroji, pacientovi, ale aj na prostredí, v ktorom sa implantát používa. Dnešné implantáty dokážu pracovať s frekvenciami stimulácie od 800 do 2500 pps a šírkou impulzu 33 μ s na fázu.

Pri stimulácii tkaniva v okolí zakončení sluchového nervu dochádza k tomu, že stimuly jednotlivých elektród zasahujú okrem svojho blízkeho okolia aj pásma susedných elektród a tým dochádza k porušeniu informácie, ktorú sa snažíme zakódovať do postupností impulzov. Miera interakcie závisí priamo na vzájomnej vzdialenosti aktívnych elektród. Tu sa dostávame k pôvodu vzniku názvu pre metódu CIS (Continuous Interleaved Sampling). Aby sa zabránilo interakcii medzi elektródami, je aktívna v jednom čase vždy len jedna z elektród a impulzy sa v čase postupne prekladajú. Rôzne druhy implantátov používajú tiež rôzne stratégie prekladania. Ide napríklad o postupné striedanie všetkých elektród alebo prekladanie vždy s dvojicou (n -ticou) aktívnych elektród, ktorých vzájomná vzdialenosť je dostatočne veľká nato, aby nedošlo k interferencii medzi nimi.

V poslednej dobe CIS stratégie používajú tiež veľmi vysokú frekvenciu stimulácie až 5000 pps. Pri týchto rýchlostiach stimulácie dochádza k javu obnovenia stochastickej aktivity v sluchovom nerve [13], ktorú pozorujeme v sluchovom nerve aj u zdravých jedincov bez implantátu. Práve tomuto náhodnému druhu aktivity sa pripisuje významná úloha pri kódovaní v sluchovom nerve, ktorú sa však zatiaľ nepodarilo popísať. Pri elektrickej stimulácii nervu, a to hlavne nízkymi frekvenciami stimulácie, dochádza k synchronizácii neurálnej aktivity a tým aj k odstráneniu spomínanej informácie. Okrem metódy používajúcej vysoké hodnoty pps na dosiahnutie stochastickej aktivity sa používa experimentálne tiež metóda vkladania náhodných, obálkou nemodulovaných impulzov [11].

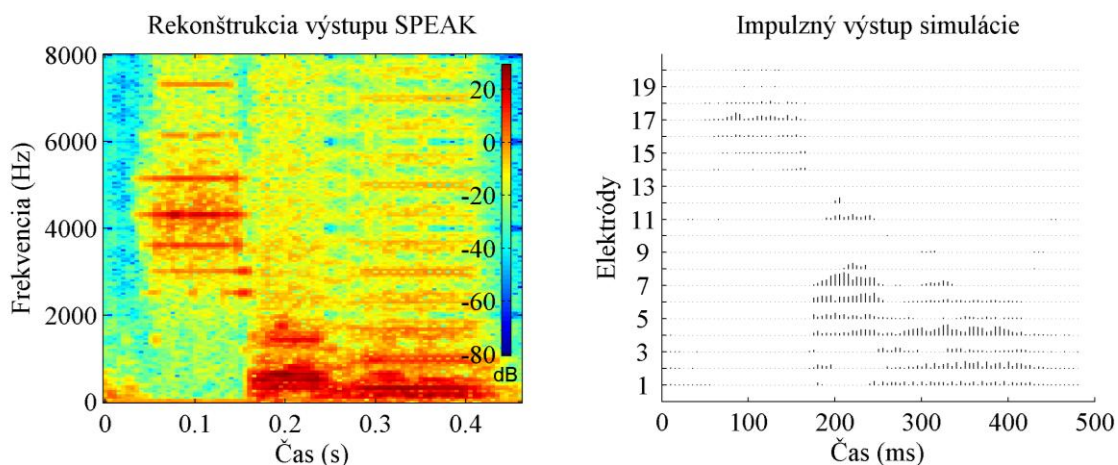
4.2 Spectral Peak (SPEAK)

Druhou stratégiou založenou na princípe vokodéra, ktorá sa v implantátoch používa, je SPEAK [14] a vychádza z pozorovania, že na postačujúce zakódovanie zvuku do impulzov nemusí výstupný signál obsahovať informácie z úplne celého frekvenčného spektra vstupného signálu. Pre zrozumiteľné zakódovanie rečového signálu stačí vybrať len oblasti frekvenčného spektra s maximálnou energiou, od čoho je odvodený aj súhrnný názov pre stratégie podobného typu, tzv. výber spektrálneho maxima (Spectral Maxima Sound Processor, SMSP). Zo schémy kódovania stratégie SPEAK na obrázku 4-4 vidíme, že na rozdiel od CIS stratégie, SPEAK obsahuje modul na výber n z m kanálov s maximálnou energiou na stimuláciu, kde m je počet kanálov použitej banky filtrov a n je maximálny počet vybraných kanálov.



Obrázok 4-4 Schéma stratégie kódovania SPEAK. Schéma zobrazuje základné bloky kódovania zvuku v implantáte. Vstupom je digitalizovaný signál zvuku, výstupom sú elektrické impulzy, ktoré stimulujú sluchový nerv.

Banka filtrov je oproti stratégii CIS zložená z väčšieho počtu filtrov, z ktorých je možno vybrať maximum, a tento počet sa v súčasných implantátoch pohybuje od 16 do 22 kanálov. Maximálny počet stimulovaných elektród sa pohybuje v rozmedzí od 6 do 12, podľa typu implantátu a celkového počtu kanálov. Typy rozloženia stredových frekvencií sú rovnaké ako pri CIS stratégii. Z pohľadu na spektrogram rečového signálu na obrázku 2.9 je zrejmé, že rozloženie zhukov v okolí spektrálnych maxim sa v čase mení, a preto je nutné frekvenčnú analýzu a následnú selekciu stimulačných kanálov prevádzať na menších časových úsekoch signálu, keď sa charakter signálu nemení. Postačujúce rozlíšenie tak predstavuje rozdelenie na znelé a neznelé časti reči, kde sa rozloženie formantov nemení, čomu zodpovedá segmentácia na bloky s dĺžkou 4 až 20 ms. Výstup stratégie SPEAK s výberom maxim v pomere 6/20 je znázornený na nasledujúcom obrázku.



Obrázok 4-5 Výstup simulácie stratégie SPEAK (rekonštrukcia a elektrické impulzy). Na obrázku vľavo je zobrazený spektrogram rekonštruovaného zvukového signálu, ktorý je výstupom simulácie. Vstupný zvukový signál slova 'seven' je zobrazený na obr. 2-9. Vpravo je zobrazený impulzný výstup (prúdové amplitúdy) simulácie v aplikácii „CIsim“ pre 20 elektród.

Hlavnou výhodou SPEAK stratégie je relatívne malý počet kanálov, ktoré stimulujeme, čo potvrdzuje súhrn viacerých štúdií v [11]. Napríklad konfigurácia s maximálne tromi kanálmi z dvadsiatich poskytuje postačujúce porozumenie pri rozoznávaní hlások, ale aj slov a viet.

Nevýhodou metód s premenlivým počtom stimulovaných kanálov je zhoršená kvalita kódovania šumom zaťaženého signálu. Podľa druhu a frekvenčného spektra šumu dochádza k vybudeniu signálu v rôznych oblastiach spektra a následnej nesprávnej selekcii kanálov pre stimuláciu obsahujúcich tieto vybudené časti spektra. Riešením takéhoto problému môže byť jednak zvýšenie počtu stimulovaných kanálov, čím sa do výberu dostanú ďalšie dôležité časti spektra pôvodného signálu alebo použitie jednej z množstva používaných metód na odstránenie šumu zo signálu [15].

Rýchlosť stimulácie sa pre SPEAK stratégiu vďaka lepšiemu výberu informácie zo signálu pohybuje na pomerne nízkej úrovni od 250 pps. Nástupca tejto stratégie, tzv. ACE metóda (Advanced Combination Encoder), používa vyššiu rýchlosť stimulácie, 900 pps až 1800 pps, čím sa dosiahne vyššie temporálne rozlíšenie stimulačného signálu a vyššie hodnotenie u pacientov s touto stratégiou kódovania.

4.3 Ďalšie používané stratégie kódovania

Pre doplnenie zoznamu a ilustráciu odlišných typov stratégií, môžeme spomenúť napríklad stratégie s analógovým typom elektrickej stimulácie, medzi ktoré patrí metóda Compressed Analog (CA). CA využíva súbor štyroch pevných pásmových priepustí v rozmedzí od 500 do 3400Hz. Vstupný signál najprv prechádza fázou predspracovania, potom nasledujú spomínané pásmové priepuste s regulovateľným zosilnením. Výstup z týchto filtrov sa používa priamo na spojitú analógovú stimuláciu jednotlivých elektród. Metóda CA bola používaná hlavne v minulosti, a kvôli spojitosti elektrostimulácii, ktorá vykazovala interferenciu stimulácie medzi jednotlivými elektródami, ju nahradili súčasné metódy tzv. diskretnej elektrostimulácie, ktorá umožňuje dokonca prekladanie jednotlivých impulzov.

Zaujímavý je typ stratégií, ktoré sa snažia o lepšiu interpretáciu základného kmitočtu reči daného formantom F0. Predchádzajúce stratégie sa sústredili na zrozumiteľnosť reči a základná výška tónu nebola zachovaná a závisela od parametrov použitej stratégie. To sa ukázalo ako problém pre používateľov implantátov hovoriacich niektorým z tonálnych jazykov, kde výška tónu reči vplýva na lexikálny význam slova. Na týchto poznatkoch je založená metóda extrakcie formantov F0 a F2. Formant F0 je extrahovaný dolnou priepusťou s hranicou 270Hz. Formant F2 pásmovou priepusťou 1 až 4kHz. Frekvenciu F0, v ktorej je kódovaná výška tónu, udáva priamo rýchlosť elektrickej stimulácie a frekvencia F2 nám určuje poradie elektródy z elektródového poľa, rozmiestneného pozdĺž celej bazilárnej membrány, ktorá sa na stimuláciu použije. Amplitúdu impulzov určíme z obálky výstupného signálu filtra ktorý extrahuje formant F2. Vylepšením stratégie je ešte extrakcia formantu F1. Metóda s použitím F0, F1 a F2 stimuluje naraz dve elektródy odpovedajúce frekvenciám F1 a F2. Obidve metódy sa u pacientov prejavili lepším rozoznávaním samohlások. Príkladom tohto typu stratégie je tzv. MultiPeak (MPEAK) stratégia.

5 Simulácia kochleárneho implantátu

Metódy kódovania zvuku v kochleárných implantátoch sú v súčasnosti predmetom záujmu viacerých výskumných skupín, či už v oblasti akademickej alebo komerčnej sféry. Väčšina z týchto výskumov sa sústreďuje len na jednu metódu a sleduje sa vplyv úzkeho počtu parametrov na úspešnosť metódy. Navyiac len k veľmi malému počtu z týchto výskumov je zverejnený úplne podrobný popis metódy a použitých konfigurácií, prípadne zdrojový kód, čo by mohlo odhaliť dôvody vzniku odlišných alebo dokonca protichodných výsledkov, ktoré sa občas objavujú medzi jednotlivými štúdiami. Simulátor funkcie kochleárneho implantátu nazvaný „CIsim“, ktorý vznikol ako softwarová časť tejto práce, podporuje simuláciu viacerých kódovacích stratégií, konfiguráciu a možnosť porovnania vplyvu jednotlivých parametrov na výsledok simulácie, ktorým je signál impulzov, ktorými sa stimuluje elektródové pole. Tento signál je možné modulom syntézy previesť na zvukový výstup, ktorý môže byť použitý na subjektívne či objektívne porovnanie výsledkov.

5.1 Vývojové prostredie

Pri implementácii simulácie bolo použité programové prostredie Matlab. Hlavným dôvodom pre výber tohto prostredia je podpora pre digitálne spracovanie signálu, ktoré je základom signálových procesorov kochleárných implantátov. Výhodou tohto prostredia je tiež prepracovaná manipulácia s maticami, a nakoniec práve Matlab je veľmi rozšíreným a obľúbeným nástrojom pre podobné účely, takže prípadné rozšírenie, či úprava kódu pre skúsených používateľov nie je problém.

Aplikácia je určená pre verziu *Matlab 7.0.1 (R14SP1)* a používa tiež balík funkcií *Signal Processing Toolbox (SPT)*, ktorý je ale súčasťou inštalácie. Pri návrhu aplikácie sa bral ohľad aj na používateľov iných verzií Matlabu a nie sú v nej použité žiadne funkcie špecifické pre vývojovú verziu, prípadne iné systémovo závislé volania, takže je aplikácia použiteľná aj v novších verziách Matlabu (*R2010a*).

Prostredie Matlab umožňuje okrem distribúcie priamo zdrojových kódov aplikácie aj prevod zdrojových súborov „*.m“ do natívneho C, prípadne C++ kódu a kompiláciu, čím je možné vytvoriť distribúciu pre konkrétny operačný systém, bez nutnosti inštalácie celého prostredia Matlab. To je využité aj v tomto projekte, ktorého súčasťou je samostatne spustiteľná aplikácia určená pre platformu Windows. Aby bolo možné spustiť takto vytvorenú aplikáciu, musí byť pred jej spustením nainštalovaný *Matlab Compiler Runtime (MCR)*, čo je vlastne samostatný balík zdieľaných knižníc, ktoré obsahujú základné funkcie Matlabu. Výhodou takejto distribúcie je okrem dobrej portability aj jej zrýchlenie, pretože samotný Matlab je interpretovaný jazyk. Zdrojové kódy ako aj samostatná distribúcia aplikácie a *MCR* určené pre platformu Windows sú súčasťou príloh tejto práce.

V kochleárných implantátoch, či už komerčných alebo z oblastí akademického výskumu, je predpokladom pre ich fungovanie spracovanie signálu v reálnom čase, preto sa často využívajú špeciálne signálové procesory umožňujúce rýchle spracovanie signálu, čo však

často znamená, že priestor konfigurácií parametrov je značne obmedzený. Hlavný dôraz pri vývoji aplikácie „CIsim“ bol kladený na demonštračné účely kódovania a dekódovania zvuku v kochleárnom implantáte a vplyv jednotlivých parametrov na výsledky. Samotné prostredie Matlab nemá podporu pre real-time spracovanie a prehrávanie signálu. To umožňujú až niektoré rozšírenia ako napr. Simulink. Preto sú v aplikácii „CIsim“ použité tzv. offline simulácie.

5.2 Návrh aplikácie

Matlab je interpretovaný jazyk určený prevažne na vedecké výpočty a vizualizáciu dát a tomu je prispôsobený aj vývoj software. Základným stavebným prvkom akéhokoľvek programu v Matlabe je funkcia alebo skript, čo je vlastne sekvencia volaní jednotlivých funkcií. Podpora objektového návrhu sa objavuje až v novších verziách Matlabu, a preto bol pri implementácii, s ohľadom na spätnú kompatibilitu, zvolený radšej klasický prístup. Aplikácia „CIsim“ nie je len samotná simulácia, ale ide o súbor základných stavebných modulov v podobe funkcií a skriptov, ktoré je možné samostatne použiť aj v iných konfiguráciách podobne ako to je u toolboxov Matlabu. Jednotlivé moduly, ktoré odpovedajú prevažne fázam stratégií kódovania vymenovaným v predchádzajúcej kapitole, sú spolu s ostatnými pomocnými funkciami rozdelené do samostatných zdrojových súborov Matlabu („*.m“). Medzi pomocné funkcie patria napríklad moduly na načítavanie vstupných zvukových súborov alebo na vizualizáciu nastavení a výstupov simulácie. Aplikácia je navrhnutá tak, aby do nej bolo možné pridávať aj ďalšie moduly spracovania signálu alebo dokonca nové stratégie kódovania, ktoré je potom možné začleniť aj do užívateľského rozhrania aplikácie. Väčšina stratégií kódovania používa rovnaký spôsob frekvenčnej analýzy a generovania elektrických impulzov, takže tieto už implementované moduly je možné znovu využiť. Dôležitou časťou implementácie bolo vytvorenie grafického užívateľského rozhrania, pomocou ktorého je možné spúšťanie simulácie a jednoduché nastavovanie parametrov s možnosťou výberu bez toho, aby používateľ musel poznať implementačné detaily. Aplikácia podporuje tiež načítavanie a ukladanie nastavení simulácie a výstupov simulácie do súborov vlastného formátu. Konzolová verzia simulácie, ako aj verzia s grafickým rozhraním, užívateľská dokumentácia k aplikácii a vzorové konfigurácie a zvukové vstupy sú súčasťou prílohy tejto práce.

5.3 Implementácia blokov metód kódovania

Ako už bolo spomenuté, aplikácia dokáže simulovať metódy CIS a SPEAK popísané v predchádzajúcej kapitole. Simulácia stratégie ACE spadá pod typ SPEAK a líši sa len rozložením banky filtrov a rýchlosťou stimulácie. Zásadným rozdielom v spracovaní signálu týchto dvoch simulácií je to, že zatiaľ čo CIS simulácia spracováva naraz celý vstupný signál, SPEAK používa rozdelenie na bloky s nastaviteľnou veľkosťou. Minimálna hodnota dĺžky bloku je 4 ms. Maximálna dĺžka bloku je zase daná veľkosťou vstupného signálu. Prednastavená hodnota tohto parametra pre SPEAK simuláciu je 10 ms, čo je kompromisom

medzi časovým rozlíšením a celkovým počtom blokov, ktoré je treba spracovať. V nasledujúcich sekciách popíšeme presnú funkciu a rozsah parametrov všetkých dôležitých častí simulácie.

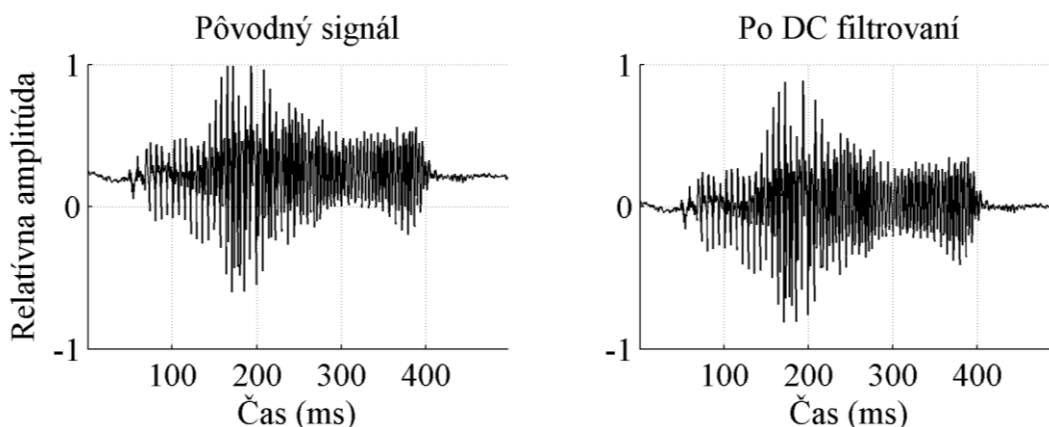
5.3.1 Načítavanie a predspracovanie zvuku

Formát vstupu

Vstupom pre simuláciu je súbor zdigitalizovaného zvuku vo formáte WAVE PCM [16], so vzorkovacou frekvenciou od 8000 Hz do 22050 Hz. Tento frekvenčný rozsah bol zvolený s ohľadom na šírku spektra, ktoré je možné spracovávať a tá sa teda pohybuje od 4000 do 11025 Hz (Nyquistova frekvencia). Pretože sa jedná o simuláciu tzv. monoaurálneho kochleárneho implantátu, je pri načítavaní stereofónneho zvukového signálu vybratý len jeden z kanálov (výber kanálu prípadne zlúčenie stereo signálu do jedného kanálu je možné nastaviť). Po načítaní je zvuk reprezentovaný ako pole amplitúd jednotlivých vzoriek, ktorých hodnoty v Matlabe musia byť navyše normalizované do intervalu $(-1; 1)$, pretože v opačnom prípade dochádza k orezávaniu zvuku a tým strate informácie. Simulácia neobsahuje žiaden AGC modul, ktorý by automaticky upravoval hlasitosť a preto je pri načítaní zvuku signál naškálovaný do maximálneho pracovného intervalu amplitúd $(-1; 1)$.

DC filter

Menej kvalitné vstupné nahrávky zvuku sa vyznačujú nenulovou jednosmernou zložkou signálu (DC), ktorej odstránenie je možné v aplikácii „CIsim“ previesť dvoma spôsobmi. Prvým je aplikovanie hornej priepuste s nastaviteľnou hraničnou frekvenciou, ktorej prednastavená hodnota 30 Hz je pre tento účel postačujúca. Vyššie hodnoty hraničnej frekvencie filtra môžu dobre poslúžiť pri redukcii nízkofrekvenčného šumu v nahrávkach. Druhou možnosťou je jednoduché odčítanie strednej hodnoty amplitúdy od signálu, čím sa dosiahne rovnaký účinok. Tento druh DC filtrovania je povolený len v CIS simulácii, ktorá spracováva vstupný signál ako celok, pretože stredná hodnota amplitúd signálu by sa pre jednotlivé bloky protokolu SPEAK líšila.



Obrázok 5-1 Použitie DC filtra. Na obrázku vľavo je znázornený priebeh signálu s nenulovou jednosmernou zložkou. Vpravo je zobrazený rovnaký signál po aplikácii DC filtra.

Predbežná korekcia

Pri simulácii je možný výber z dvoch filtrov predbežnej korekcie. Prvým z nich je často používaný filter s konečnou impulznou odpoveďou. Frekvenčná odpoveď filtra pre rôzne koeficienty zvýraznenia je znázornená na obrázku 5-2. Koeficienty jeho prenosovej funkcie sú definované takto:

$$b = [1, p], \quad p \in (-0.99; -0.7) \quad (5-1)$$

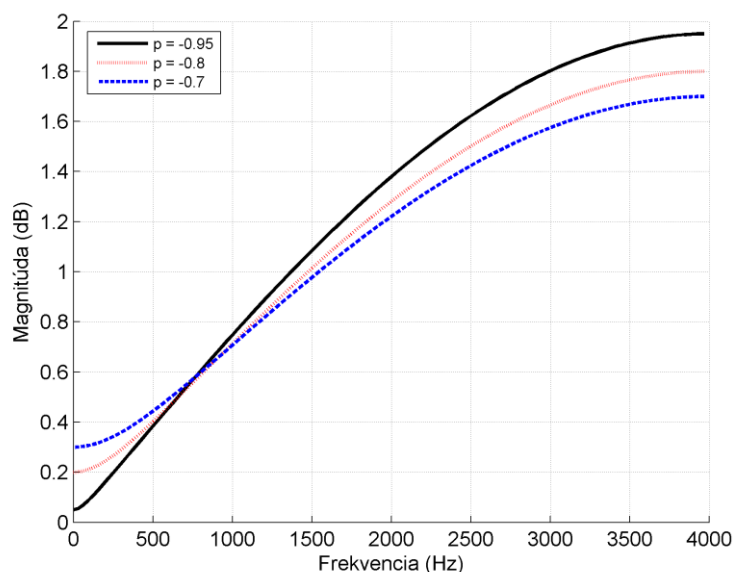
$$a = [1]$$

Prednastavená hodnota v simulácii, $p = -0,95$, je nastavená tak, aby útlm signálu nastával až pre hodnoty pod jednu tretinu vzorkovacej frekvencie. Druhý z filtrov, ktorý sa v simulácii používa je filter s nekonečnou impulznou odpoveďou, prevzatý z [17], ktorého tvar sa odvíja od vzorkovacej frekvencie signálu. Jeho koeficienty sú definované nasledujúcim vzťahom:

$$b = \left[1 - e^{\left(-1200 \cdot \frac{2\pi}{f_s}\right)} \right] \quad (5-2)$$

$$a = \left[1 - e^{\left(-3000 \cdot \frac{2\pi}{f_s}\right)} \right]$$

Útlm tohto typu filtra je rovnomernejší a preto je prednastavený v oboch typoch simulácie.



Obrázok 5-2 Frekvenčná odpoveď preemfázového filtra. Na obrázku je znázornená frekvenčná charakteristika preemfázového filtra daného vzťahom 5-1, pre rôzne hodnoty koeficientu p . Z obrázku je zrejmé, že so stúpajúcou frekvenciou rastie zosilnenie signálu.

Aby bolo možné v simulácii sledovať efekt jednotlivých stupňov predspracovania, je možné v nastavení zvoliť vynechanie každej z fáz. V štandardnom nastavení simulácie sú ale DC filtrovanie a preemfázový filter zapnuté.

5.3.2 Návrh banky filtrov

Počet kanálov

Banka filtrov určená na frekvenčnú analýzu vstupného zvuku je najdôležitejšou časťou simulácie a má kritický vplyv na kvalitu výsledného rekonštruovaného zvuku. Je zložená zo sústavy na seba nadväzujúcich digitálnych filtrov typu pásmovej priepuste a počet týchto filtrov sa v simulácii dá nastaviť v rozsahu od 2 do 22, pričom výstup každého filtra sa použije na stimuláciu jednej elektródy, takže tento parameter určuje aj počet stimulačných elektród. Rozsah počtu pásiem bol zvolený tak, aby vyhovoval obidvom podporovaným typom simulácií. Pre CIS simuláciu je počet elektród prednastavený na 8, pre SPEAK simuláciu je prednastavená hodnota 20.

Rozsah analyzovaného pásma

Maximálny možný rozsah analyzovaného frekvenčného spektra je v aplikácii „CIsim“ 100 až 11025 Hz a vychádza z hornej hranice pre vzorkovaciu frekvenciu (22 kHz). Rozsah bol zvolený tak, aby pokrýval pásmo rečového signálu (0,1 až 8 kHz). Rozsah analyzovaného pásma je automaticky nastavený na maximálny možný rozsah podľa vzorkovacej frekvencie vstupného signálu, ale je možné tento rozsah nastaviť aj ručne. Jediným obmedzením je minimálna šírka spracovávaného pásma stanovená na 1 kHz, čo ale nijak vážne neovplyvňuje návrh banky filtrov, pretože väčšina implantátov používa šírku analyzovaného pásma aspoň 3 kHz. Toto obmedzenie je nastavené kvôli samotnému návrhu digitálnych filtrov. Pri filtroch s konečnou impulznou odpoveďou sa pri znižovaní šírky prepúšťaného pásma zvyšuje počet koeficientov prenosovej funkcie, čo jednak zvyšuje výpočtovú náročnosť, a tiež vnáša do výstupného signálu oneskorenie, čo je jedna z vlastností FIR filtrov, ktorá by mohla mať na výsledky simulácie nežiadúci vplyv. Pre filtre s nekonečnou impulznou odpoveďou sa spolu so zúžením šírky prepúšťaného pásma zvyšuje ich rád, čoho následkom je ich nestabilita, ktorá sa prejaví nežiadúcim poškodením ich frekvenčnej odpovede. Povolenský rozsah spracovávaného pásma je preto nastavený tak, aby k spomínaným javom nedochádzalo pri použití žiadneho z typov rozloženia hraničných frekvencií.

Typy rozloženia hraničných frekvencií

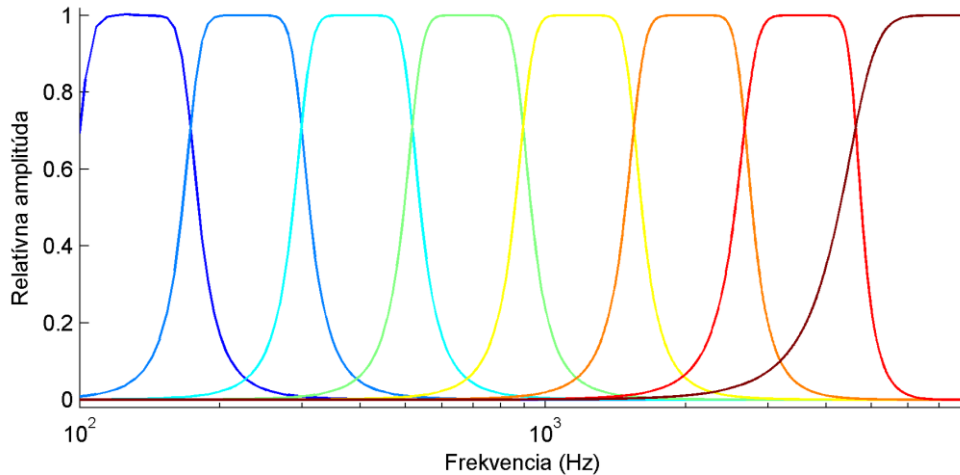
Ako už bolo spomenuté v odstavci 2.3, závislosť výšky tónu a frekvencie je nelineárna a existuje hneď niekoľko modelov závislosti, ktoré z percepcie frekvencie vychádzajú. V simulácii je možné zvoliť jednu z troch metód pre rozloženie hraničných frekvencií banky filtrov a všetky tri sú odvodené z poznatkov v odstavci 2.3.

Prvá z použitých metód navrhuje logaritmické rozloženie hraníc, ktoré závisí na použitej vzorkovacej frekvencii a tiež na šírke a pozícii pásma v rámci maximálneho rozsahu frekvenčnej analýzy. Táto metóda návrhu je prevzatá a používaná tiež v simulácii [17]. Pre odvodenie hraničných frekvencií platí vzťah:

$$f_i = f_{start} \cdot 10^{w \cdot i}, i = 0..n \quad (5-3)$$

$$w = \frac{\log_{10}(f_{stop}/f_{start})}{n}$$

kde n je počet pásiem, f_{start} a f_{stop} sú hranice spracovávaného pásma a w je šírka pásma (bandwidth) pre jeden kanál. Frekvenčná odpoveď banky filtrov zo simulácie, ktorá používa tento typ rozdelenia, je znázornená na obrázku 5-3.

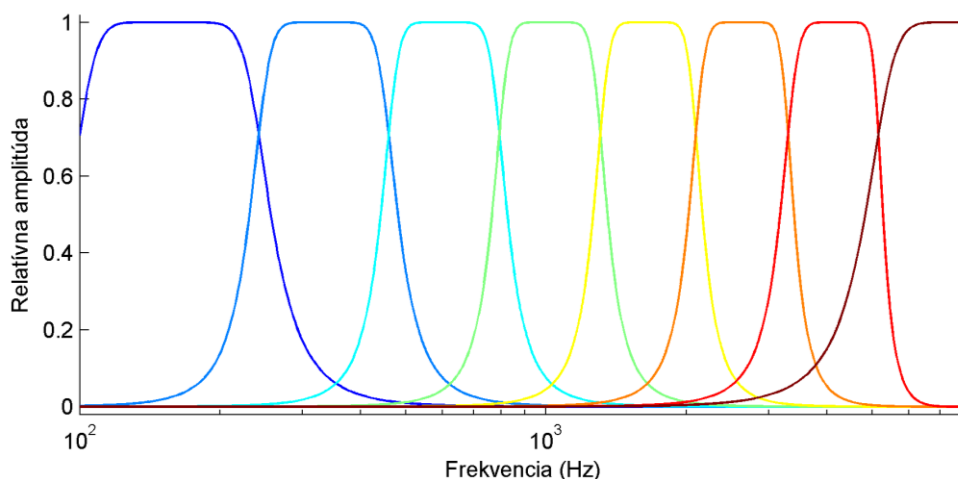


Obrázok 5-3 Frekvenčná odpoveď logaritmickej banky filtrov. Obrázok znázorňuje frekvenčné odozvy banky filtrov s logartmickým rozložením frekvencií. Na ose x je v logaritmickej škále zobrazené frekvenčné pásmo 0,1 až 8 kHz, na ose y relatívna amplitúda. V logaritmickej škále je rozloženie filtrov rovnomerné.

Druhý, pre simuláciu navrhnutý typ rozloženia frekvencií vychádza z výsledkov experimentu D. Greenwooda [18], ktorému sa podarilo na základe psychoakustických meraní určiť súvislosť medzi pozíciou na bazilárnej membráne a frekvenciou, pri ktorej sú vlásokové bunky na tejto pozícii aktívne. Bol vytvorený všeobecný model tejto závislosti a za predpokladu, že priemerná dĺžka rozvinutej bazilárnej membrány je u človeka 35 mm, platí nasledujúci vzťah:

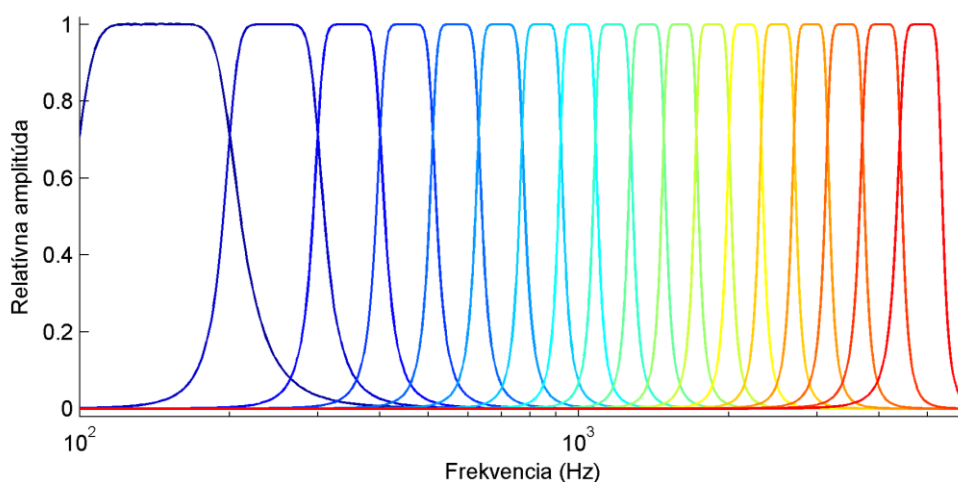
$$f = 165,4(10^{0,06 \cdot (35-d)} - 1) \quad (5-4)$$

kde f je frekvencia v Hz, a d je vzdialenosť skúmanej pozície od prednej časti bazilárnej membrány pri vstupe do kochley. Pre určenie hraničných frekvencií na základe tohto vzťahu predpokladáme, že v nastavení špecifikovaný počet elektród je rozmiestnený rovnomerne pozdĺž bazilárnej membrány, v pásme definovanom krajnými frekvenciami spracovávaného pásma. Z pozícií jednotlivých elektród potom pomocou vzťahu 5-4 dopočítame hraničné frekvencie pre jednotlivé pásma. Tento typ rozdelenia je prednastavený pri oboch typoch simulácie. Obrázok 5-4 znázorňuje banku filtrov využívajúcu toto rozdelenie.



Obrázok 5-4 Frekvenčná odpoveď banky filtrov s Greenwoodovým rozložením. Obrázok znázorňuje frekvenčné odozvy banky filtrov s rozložením frekvencií podľa vzťahu 5-4. Na ose x je v logaritmickej škále zobrazené frekvenčné pásmo 0,1 až 8 kHz, na ose y relatívna amplitúda.

Tretím rozdelením hraničných frekvencií je rozdelenie podľa kritických pásiem (dané Barkovou škálou), ktoré definuje hraničné frekvencie pre prvých 22 kritických pásiem. Rozdielom oproti predchádzajúcim typom rozmiestnenia hraníc je to, že v tomto prípade sú hodnoty frekvencií pevne definované a začiatková frekvencia spracovávaného pásma je pevne stanovená na hodnotu 100 Hz (dolná hranica frekvenčného rozsahu simulácie) a určuje sa len koncová frekvencia spracovávaného pásma. Kvôli týmto pevným hraniciam je odvodený aj počet kanálov, ktoré pásmo od 100 Hz po určenú koncovú frekvenciu zaberá. Frekvenčné odpovede banky filtrov pre toto rozdelenie sú zobrazené na obrázku 5-5. Práve toto rozdelenie sa používa vo väčšine implantátov, jednak preto, že berie do úvahy tonotopické usporiadanie receptorov a jav maskovania, ale aj preto, že takmer lineárne rozloženie pásiem v oblasti nízkych frekvencií dobre pokrýva oblasť prvého formantu a umožňuje lepšie rozoznanie znejšej časti rečového signálu.



Obrázok 5-5 Frekvenčná odpoveď banky zložená z osemnástich kritických pásiem. Obrázok znázorňuje frekvenčné odozvy banky filtrov s rozložením frekvencií podľa vzťahu 2-5. Na ose x je v logaritmickej škále zobrazené frekvenčné pásmo 0,1 až 6 kHz, na ose y relatívna amplitúda.

Typy použitých filtrov

Nastavenia simulácie umožňujú voľbu medzi bankou filtrov s konečnou (FIR) alebo nekonečnou (IIR) impulznou odpoveďou, aby bolo možné overiť prínosy, prípadne nežiadúce javy pre obidva druhy filtrov.

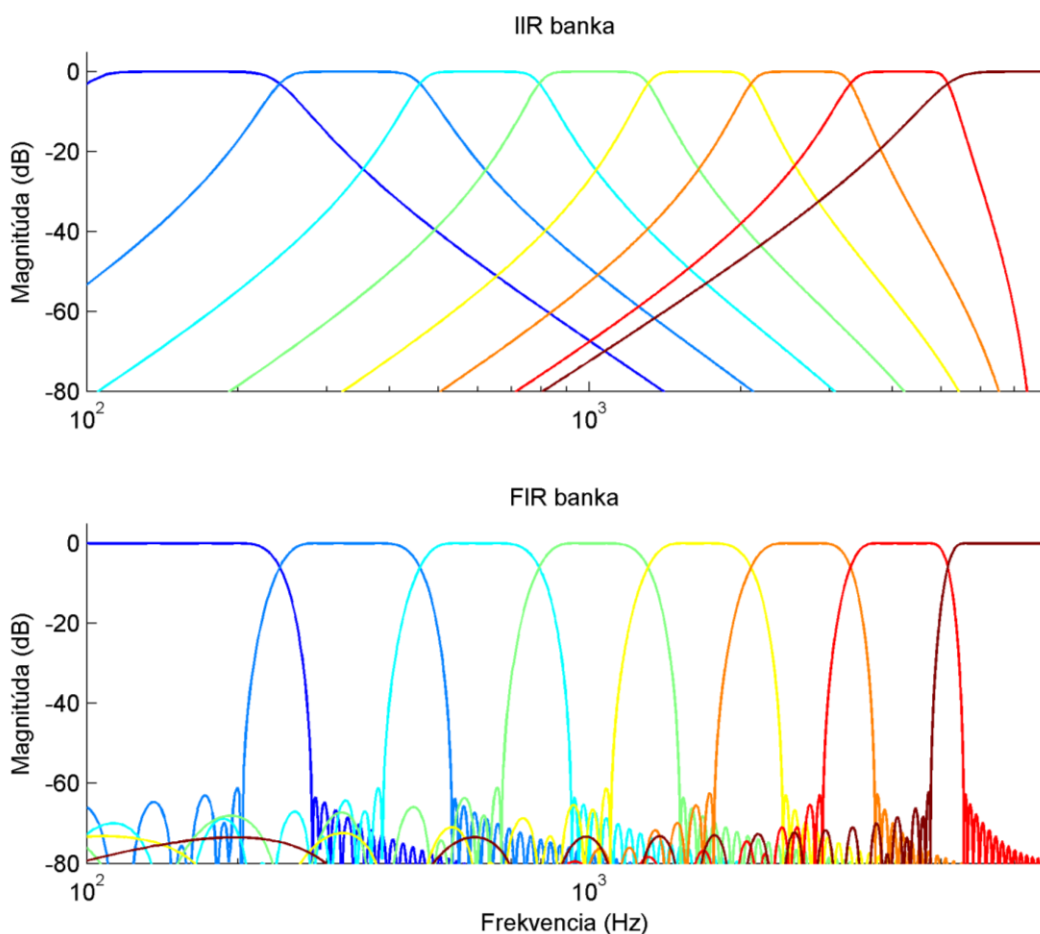
Banka FIR filtrov je navrhnutá metódou okna (FIR windowing design) s pomocou funkcie *fir1*, ktorá je súčasťou SPT. Pri návrhu bolo použité Kaiserovo okno a filtre sú navrhnuté tak, že útlm v oblasti centrálnych frekvencií filtrov má hodnotu 0 dB a v oblasti hraničných frekvencií asi 6 dB. Útlm zadržovaného pásma je nastavený parametrom *beta* vstavanej funkcie *kaiser* na hodnotu 6, ktorá pre návrh pomocou Kaiserovho okna znamená útlm asi 43 dB, čo je pre spracovanie audio signálu dostačujúce. Samotný návrh banky prebieha v cykle. V každom kroku sa najprv vytvorí horná priepusť spektrálnou inverziou (vertikálne otočenie frekvenčného spektra) dolnej priepuste z predchádzajúceho kroku, a následne je navrhnutá nová dolná priepusť pre ďalšiu z hraničných frekvencií. Konvolúciou impulzných odpovedí obidvoch filtrov (označovanou tiež ako single stage band-pass filter design [8]) dostávame pásmovú priepusť pre jeden kanál a pokračujeme ďalším cyklom. Výhoda použitia tejto FIR banky tkvie v tom, že FIR filtre sú stabilné. To znamená, že ich výstup v čase konverguje k jednej hodnote, prípadne nevykonverguje z určitého obmedzeného intervalu, zatiaľ čo výstup nestabilných filtrov môže rásť nad všetky medze. Ďalšou výhodou je lineárna fázová odpoveď, ktorá zaručuje konštantné oneskorenie vo výstupnom signáli. Nevýhodou FIR filtrov je dĺžka ich impulznej odpovede, ktorá je označovaná tiež ako rád filtra. Rád filtra rastie spolu s požiadavkami na vlastnosti filtra ako sú šírka prechodového pásma a útlm v zadržovanom/prepúšťanom pásme. V simulácii s nastavením 8 pásiem pri maximálnom možnom rozsahu spracovávaného pásma dosahuje dĺžka impulznej odpovede filtra hodnoty až 200 vzoriek. Z oblasti digitálneho spracovania signálu vieme [8], že oneskorenie filtra vo výstupnom signále je konštantné, a rovné

$$d = (ord - 1)/2 \quad (5-5)$$

kde *ord* je rád filtra. To znamená, že pri spracovaní signálu s vzorkovacou frekvenciou 8 kHz bankou filtrov dôjde k oneskoreniu asi 25 ms. Keďže sa jedná o offline simuláciu, môžeme si dovoliť oneskorenie z výstupného signálu po filtrovaní odstrániť, čo je prednastavená možnosť pri použití FIR banky, ale len v prípade CIS simulácie, keď spracovávame celý vstupný signál v jednom priechode. Simulácia SPEAK používa rozčlenenie signálu na bloky, ktorých trvanie môže byť menšie ako dĺžka samotného oneskorenia, a preto filtrovanie pomocou FIR banky prebieha bez odstránenia oneskorenia.

Druhým typom banky filtrov ktorú je možné použiť je banka Butterworthových IIR filtrov navrhnutá pomocou funkcie *butter* z balíku SPT. Rád filtrov je pevne stanovený na hodnotu 3, ktorá sa ukázala ako postačujúca. Vďaka použitému druhu filtrov je útlm v oblasti centrálnych frekvencií prepúšťaných pásiem 0 dB a v oblasti hraničných pásiem 3 dB. Na rozdiel od FIR banky majú tieto filtre pomalšie rastúci útlm (slow spectral roll-off), čo môže mať negatívny vplyv na úspešnosť kódovania v simulácii pri ich použití. Výhodou je oproti FIR filtrom veľmi malý rád a takmer zanedbateľné oneskorenie výstupného signálu. IIR filtre sa vyznačujú nelineárnou fázovou odpoveďou, čo ale našťastie pri spracovaní audio signálu nie je kritické a stratégie kódovania kochleárných implantátov sa na rekonštrukciu pôvodnej

fázy signálu nesústredujú. Na nasledujúcom obrázku je pre porovnanie zobrazená frekvenčná odpoveď pre FIR aj IIR banku filtrov, kde je môžeme pozorovať výrazný rozdiel v útlme zdržiavaného pásma filtrov na decibelovej stupnici a tiež aj tzv. spectral leakage efekt v podobe bočných lalokov FIR filtrov. Rovnaké banky sa používajú aj v simulácii.



Obrázok 5-6 Frekvenčná odpoveď FIR a IIR banky filtrov. Na ose x je zobrazené v logaritmickej škále frekvenčné pásmo 0,1 až 8 kHz, na ose y magnitúda filtrov v decibelovej škále. Na obrázku je výrazný rozdiel v strmosti frekvenčnej odpovede FIR a IIR filtrov mimo prepúšťané pásmo. Pre FIR banku je viditeľný výskyt bočných lalokov filtrov (leakage efekt).

5.3.3 Obálka a jej kompresia

Simulácia „CIsim“ podporuje všetky bežne používané algoritmy na detekciu obálky popísané v sekcii 4.1.3. V oboch typoch simulácií je možné zvoliť buď extrakciu usmernením a vyhladením tzv. integračným filtrom, alebo použitie Hilbertovej transformácie.

Usmernenie a filtrácia

Pri zvolenom type extrakcie usmernením sa dá použiť usmernenie jednocestné alebo dvojcestné a pre integračný (vyhladzovací dolno-priepustný) filter môžeme nastaviť hraničnú frekvenciu ručne alebo automaticky podľa rýchlosti použitej elektrickej stimulácie. Pri

manuálnej voľbe musíme dbať nato, aby nedochádzalo k aliasingu a hodnota hraničnej frekvencie by nemala presahovať polovicu stimulačnej frekvencie. Pri automatickom nastavení je hranica pre vyhladzovací filter nastavená podľa nasledujúcej tabuľky, ktorej hodnoty sú zvolené tak, aby nedošlo k aliasingu a zároveň sa berie ohľad na odporúčania z oblasti psychoakustiky uvedené v sekcii 4.1.3 a v [11].

<i>Rýchlosť stimulácie (pps)</i>	100 – 400	400 – 800	800 – 2000	2000 – 5000
<i>Hranica integračného filtra (Hz)</i>	100	200	400	800

Obrázok 5-7 Vzťah hranice integračného filtra a rýchlosti stimulácie.

Hodnota hraničnej frekvencie vyhladzovania priamo ovplyvňuje temporálne rozlíšenie informácie, ktorá sa na stimuláciu použije. Preto je samozrejmé, a to sa aj pri porovnaní výsledkov simulácie prejavilo, že so stúpajúcou frekvenciou vyhladenia došlo k zlepšeniu výsledkov kódovania, pretože bola podrobnejšie zachytená jemná štruktúra pôvodného signálu. Pre porovnanie sú extrahované obálky s rôznou frekvenciou vyhladenia a originálny signál znázornené na obrázku 5-8. V simulácii je prednastavená detekcia s použitím dvojcestného usmernenia a hodnota vyhladzovacej frekvencie sa volí automaticky, pretože práve toto nastavenie sa pri experimentoch najviac osvedčilo.

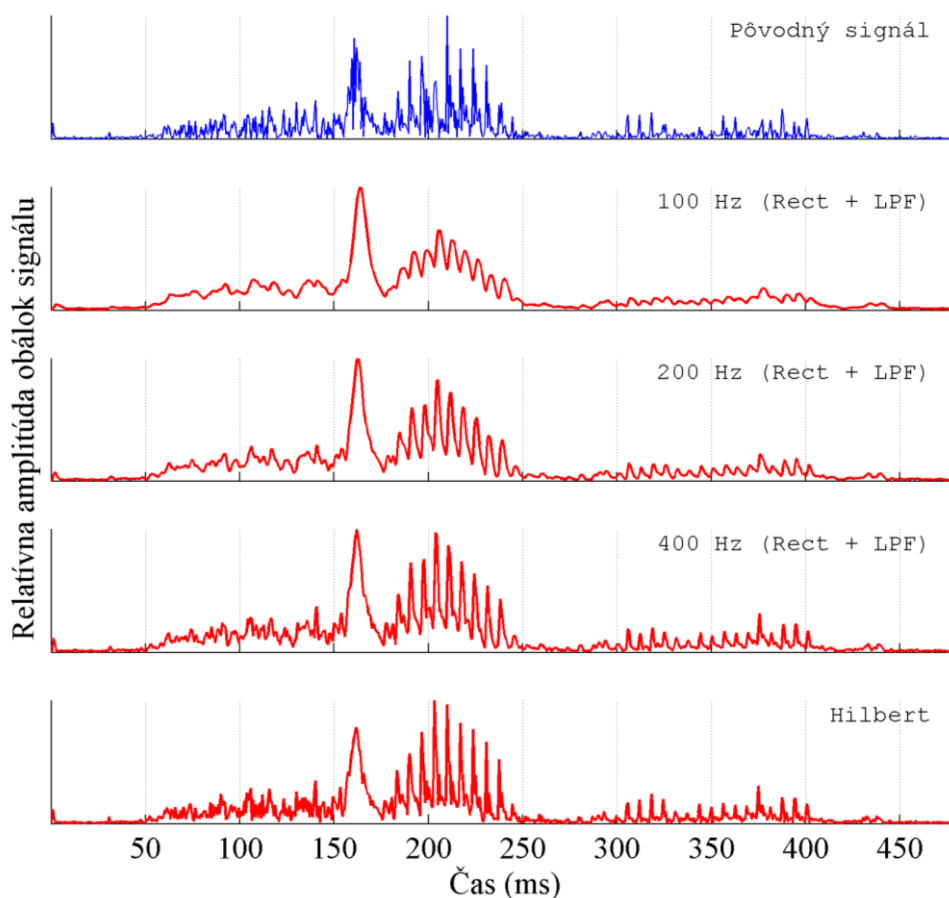
Forward-backward filtrovanie

Simulácia umožňuje nastaviť aj typ integračného filtra s voľbou medzi Butterworthovým IIR filtrom rádu 2 alebo FIR filtrom automaticky navrhnutým metódou okna. Kvôli rýchlejšiemu spracovaniu je v simulácii predvolené vyhladzovanie IIR filtrom. Ďalej je v simulácii naprogramovaná voľba možnosti filtrácie. Offline simulácia opäť umožňuje oproti normálnemu spôsobu filtrovania použiť aj tzv. forward-backward filtering, keď sa filtruje sekvencia signálu 2-krát, najprv spredu a potom zozadu. Takéto filtrovanie môžeme použiť len pri CIS simulácii s jednopriechodovým spracovaním signálu a metóda má výhodu v tom, že spätná filtrácia spôsobí obnovenie pôvodnej fázy signálu a nedochádza k poškodeniu informácie z fázy, čo má zmysel hlavne pri použití IIR filtrov, ktoré sa skreslením fázy vyznačujú.

Detekcia s použitím Hilbertovej transformácie

Na výpočet obálky touto metódou (načrtnutou v sekcii 4.1.3) bola v simulácii použitá funkcia balíku SPT – *hilbert*, ktorá počíta Hilbertovu transformáciu s pomocou FFT a inverznej FFT. Výsledkom je tzv. analytický signál, ktorý obsahuje originálny signál v reálnej zložke a Hilbertovu transformáciu v komplexnej zložke. Obálku signálu získavame zo vzťahu $env = abs(hilbert(y))$, kde y je vstupný signál ako parameter spomínanej funkcie matlabu a abs je funkcia absolútnej hodnoty (v tomto prípade ide o veľkosť komplexného čísla). Obálka signálu získaná touto metódou je zobrazená na nasledujúcom

obrázku, z ktorého je možné vidieť, že dosahuje najvyššieho rozlíšenia pri porovnaní s ostatnými typmi extrakcie.



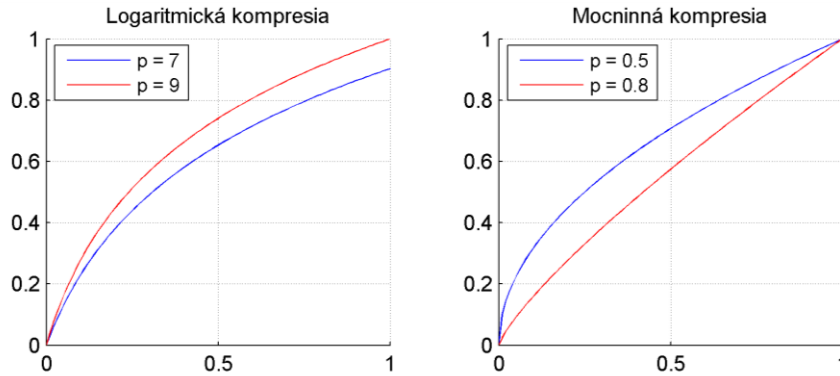
Obrázok 5-8 Extrakcia obálky. Obrázok obsahuje časový priebeh pôvodného signálu (hore), a jeho obálky extrahované usmernením (Rect) a filtrovaním LPF (low-pass) filtrom s rôznou frekvenciou (v strede) a Hilbertovou transformáciou (dole).

Kompresia amplitúdy

Fáza kompresie amplitúdy, tak ako bola popísaná v sekcii 4.1.4, nemá v prípade simulácie kochleárneho implantátu úplný význam, hlavne čo sa týka škálovania dynamických rozsahov akustického a elektrického počutia. Na rozdiel od skutočného implantátu, keď impulzy stimulujú sluchový nerv, je v simulácii použitá syntéza zvuku priamo z amplitúd elektrických impulzov, a preto má v simulácii zmysel hlavne tvar použitej krivky. Zatiaľ čo niektoré simulácie implantátov dokonca ani neumožňujú kompresiu amplitúd, „CIsim“ implementuje logaritmickú aj mocninnú kompresiu danú vzt'ahmi 4-2 a 4-3 v odstavci 4.1.4. Koeficienty kompresnej krivky je možné nastaviť osobitne pre obálku každého kanálu, alebo nastaviť použitie rovnakej krivky pre všetky kanály. Kompresiu amplitúdy je tiež možné zo simulácie úplne vyradiť. Porovnaním výsledkov viacerých simulácií sa ukázala logaritmická kompresia ako výhodnejšia a preto je prednastaveným typom pri oboch simuláciách. Hodnoty konštánt a , a b , sú pre obidva typy kriviek zvolené tak, aby kompresná krivka pokrývala celý interval $(-1; 1)$ pre hodnotu amplitúdy. Prednastavené hodnoty koeficientov logaritmickej kompresnej krivky sú zvolené nasledovne:

$$a = 1, b = 0, k = 9$$

Zmenou koeficientov a a b je možné sledovať skreslenie signálu spôsobené zmenou dynamického rozsahu amplitúd impulzov. Pre mocninnú kompresnú krivku je prednastavená hodnota kompresného koeficientu $k = 0,6$. Tvary kompresných kriviek s hodnotami kompresných koeficientov testované v simulácii sú zobrazené na nasledujúcom obrázku.



Obrázok 5-9 Kompresné krivky. Na obrázku vľavo je zobrazený priebeh logaritmickej kompresnej krivky pre kompresné koeficienty testované v simulácii. Vpravo je zobrazený priebeh mocninatej kompresnej krivky.

5.3.4 Výber spektrálneho maxima

Pri použití stratégie SPEAK, predchádza kompresii amplitúdy ešte modul výberu spektrálneho maxima. Jedná sa o veľmi jednoduchú operáciu, keď sú kanály v konkrétnom bloku zoradené zostupne podľa energie signálu danej súčtom amplitúd v rámci bloku. Prvých m kanálov z tejto množiny vyberieme na stimuláciu a výstupy ostatných kanálov sa nepoužijú. Minimálna hodnota tohto parametru je 1, maximálna hodnota určujúca počet vybraných kanálov je rovná celkovému počtu kanálov simulácie. Prednastavená hodnota pre SPEAK stratégiu používajúcu 20 kanálov je $m = 6$. Obmedzenie počtu kanálov tak umožní zredukovať nepodstatnú časť spektra.

5.3.5 Impulzný výstup simulácie

Posledným stupňom spracovania signálu v implantáte je generovanie impulzov. Podobne ako v implantáte aj v tejto simulácii je vygenerovaná jednotková sekvencia impulzov s nastavenou frekvenciou, ktorej amplitúda je modulovaná obáčkami. Táto operácia je v simulácii pre jednoduchosť a efektívnosť naprogramovaná ako prevzorkovanie signálu s použitím tvarovača nultého rádu (angl. zero-order hold resampling), podobné metódy sa používajú aj v kochleárných implantátoch. Rýchlosť stimulácie je možné nastaviť od 100 do 5000 impulzov za sekundu, čo je rozsah používaný v súčasných kochleárných implantátoch. Prednastavená hodnota v oboch simuláciách je 2000 pps. Hlavný rozdiel simulácie od reálnych implantátov je v tom, že impulzy sú skutočne iba amplitúdy v diskretnom čase

vyjadrujúce relatívnu hodnotu amplitúdy elektrického prúdu ktorý by sa na stimuláciu mal použiť, zatiaľ čo v reálnych implantátoch sa generujú dvojfázové pulzy s pevne stanovenou dĺžkou trvania v čase. Namodelovanie presného šírenia elektrického prúdu v závislosti na dĺžke impulzu by vyžadovalo pokročilé znalosti z oblasti fyziky a preto sa v práci touto problematikou nezaobrám. V priebehu implementácie bola vyskúšaná aj metóda prekladania jednotlivých impulzov v čase tak, ako sa to robí v CIS stratégii, aby sa predišlo interferenciám signálu z jednotlivých elektród. Vplyv prekladania na výsledok simulácie bol však zanedbateľný a preto bol tento prístup zo simulácie vyradený. Prekladanie impulzov by malo význam v prípade, že by simulácia obsahovala modul, ktorý by na základe vlastností šírenia prúdu v tkanive ľudského tela počítal vzájomný vplyv elektród, čo by sa prejavilo skreslením vo výstupnom impulznom signáli. Pre názornosť aplikácia „CIsim“ dokáže zobrazit' svoj impulzný výstup v podobe jednofázových diskretných impulzov. Impulzný výstup simulácie je možné tiež uložit' buď do textového („*.txt“), alebo binárneho („*.mat“) súboru, a následne použit' pri ďalšej analýze.

5.4 Syntéza zvukového výstupu

Okrem výstupu simulácie v podobe impulzov pre stimulačné elektródy, bolo nutné do aplikácie „CIsim“ doplnit' modul na spätnú syntézu reči z vygenerovaných impulzov. Pomocou neho môže užívateľ aplikácie získať predstavu o tom, aký vnem vyvoláva elektrická stimulácia u pacientov s implantátom. Výsledok simulácie v podobe rekonštruovaného zvukového signálu je možné v aplikácii prehrať, alebo uložit' do súboru WAVE formátu. Modul syntézy má aj druhý dôvod vzniku a tým je zvolená metóda na objektívne porovnanie vplyvu parametrov na výsledok simulácie. Jedná sa o metódu tzv. keprálnej analýzy, ktorá bude podrobnejšie popísaná v nasledujúcej kapitole. V literatúre zaoberajúcej sa spracovaním rečového signálu je popísaných viacero rekonštrukčných modelov. V tomto projekte je implementovaná metóda syntézy pomocou banky filtrov (Filter Bank Synthesis - FBS) a syntéza pomocou zložených sinusoid (Superimposed Sinusoidal Synthesis - SSS). Vstupným signálom pre obidva rekonštrukčné moduly je impulzný signál z generátora impulzov, čím v rekonštruovanom signáli simulujeme pokles časového rozlíšenia, ktoré vzniká prevzorkovaním na frekvenciu impulznej stimulácie. Väčšina simulácií kochleárných implantátov túto možnosť ale nemá a ako vstup pre rekonštrukčný modul sa používa priamo obálka extrahovaná z jednotlivých kanálov. Rekonštrukciu signálu za pomoci obálky umožňuje aj aplikácia „CIsim“, ale pre dodržanie presnosti simulácie je prednastaveným vstupom rekonštrukčných modulov impulzný signál.

FBS syntéza a použitie úzkopásmového šumu

Schéma syntézy FBS je totožná s fázou syntézy vokodéra na obrázku 4-1. Na rekonštrukciu zvukového signálu používa banku filtrov z fázy analýzy a to tak, že každý kanál prejde odpovedajúcim filtrom z banky. Tým sa odstráni časť frekvenčného spektra, ktoré do signálu jednotlivých kanálov vniesli moduly kompresie, generátor impulzov a generátor bieleho šumu, ktorý je pred filtráciou použitý ako tzv. excitačný signál, ktorý má

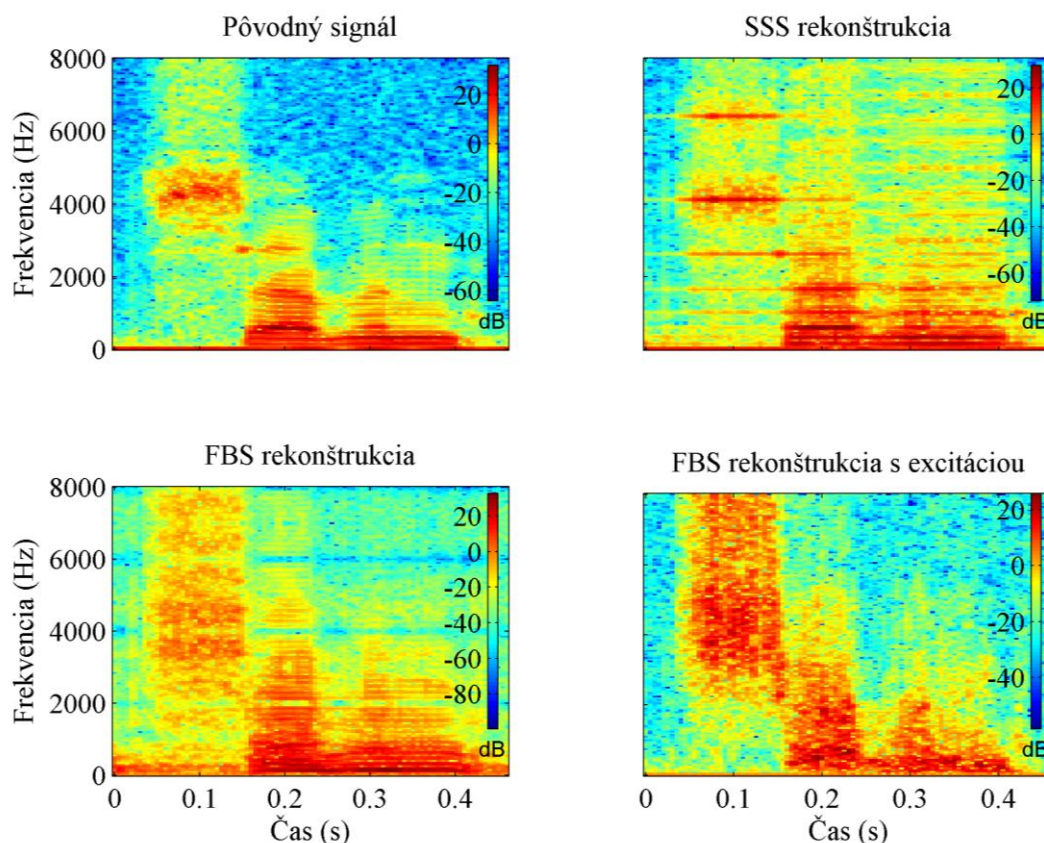
napodobňovať stochastickú nervovú aktivitu vyvolanú v sluchovom nerve pri stimulácii. Aj keď tento excitačný signál je možné v nastavení simulácie vypnúť, prednastavená hodnota povoľuje použitie excitácie pri rekonštrukcii. Rekonštrukcia pomocou banky filtrov za pomoci excitácie šumom určitej spektrálnej charakteristiky, sa v literatúre označuje tiež ako rekonštrukcia s pomocou úzkopásmového šumu (narrow-band noise synthesis).

SSS syntéza

Rekonštrukčný modul typu SSS generuje výstupný signál ako súčet sinusoíd, z ktorých každá má frekvenciu rovnú centrálnej frekvencii odpovedajúceho pásma z fázy analýzy. Amplitúda sinusoíd je modulovaná vstupným signálom. Na výpočet stredových frekvencií pásiem bol použitý geometrický priemer hraničných frekvencií filtrov. Výsledný signál syntézy je daný vzťahom 5-6, kde e_i je obálkou i -tého kanálu, a f_i je jeho stredová frekvencia.

$$s_{syn}(x) = \sum_{i=1}^N e_i \cdot \sin(2\pi f_i x) \quad (5-6)$$

Nasledujúci obrázok obsahuje spektrogram originálneho signálu a signálu vygenerovaného metódami SSS a FBS, s vypnutým a zapnutým excitačným signálom na porovnanie. Na vygenerovanie spektrogramov bola použitá simulácia CIS. Najmenej frekvencií z pôvodného spektra signálu obsahuje rekonštrukcia SSS, zatiaľ čo najviac z pôvodného spektra obsahuje FBS. FBS so zapnutým excitačným signálom zaplní šum celé spracovávané pásmo, a preto sa najviac podobá pôvodnému signálu.



Obrázok 5-10 Spektrogramy rekonštruovaných signálov reči po CIS simulácii. Obrázky zobrazujú frekvenčné spektrum pôvodného signálu anglického slova 'seven', a rôznych typov rekonštrukcií. Vo všetkých výstupoch pozorujeme podobnosť s pôvodným signálom, avšak najčistejšie spektrum má SSS rekonštrukcia, naopak, najviac vybudené spektrum má výstup rekonštrukcie s excitáciou.

5.5 Prezentácia dát

Výhodou použitia prostredia Matlab pri implementácii aplikácie je vstavaná podpora pre vizualizáciu dát. Napriek tomu sa objavili situácie, keď vlastnosti vstavaných funkcií úplne nevyhovovali a museli byť doplnené prípadne nahradené novým kódom. Tieto funkcie sú implementované s ohľadom na maximálnu rozšíriteľnosť a je možné ich použiť na prezentáciu dát aj v iných aplikáciách. Nasledujúci zoznam obsahuje niektoré zo spomínaných prezentačných funkcií:

- *spectrogram* – náhrada vstavenej funkcie s možnosťou zmeny farebnej schémy, typu spektrogramu (2D/3D) a hlavne s možnosťou nastavenia rovnakého rozsahu farieb pre spektrogramy, ktorých obsah treba porovnať.
- *electrodiagram* – vykreslenie impulzného signálu pre všetky kanály prehľadne do jedného grafu.
- *waveplot* – vykreslenie časového priebehu signálu s možnosťou škálovania
- *fspectrum* – vykreslenie frekvenčného spektra signálu do grafu. Na výpočet spektra je použitá vstavaná funkcia *fft* na rýchlu Fourierovu transformáciu. Je

možné zvolit' tiež zobrazenie v dB stupnici, prípadne normalizovať magnitúdu frekvenčného spektra.

- *fresponse*, *fbfresponse* – je obalením vstavanej funkcie *freqz*, ktorá počíta frekvenčnú odpoveď systému. Tieto funkcie umožňujú vykreslenie frekvenčnej odpovede pre samostatné filtre, ale aj pre celú banku filtrov naraz do prehľadného grafu.
- *sndplayer* – je obalením vstavanej funkcie *audioplayer* a predstavuje prehrávač zvukových signálov s možnosťou prehrávania v slučke a zmenou hlasitosti, čo sa použilo hlavne v užívateľskom rozhraní aplikácie „CIsim“.

5.6 Implementácia GUI a ďalšie funkcie

Grafické užívateľské rozhranie k aplikácii bolo jednou z požiadaviek v zadaní a umožňuje jednoduché subjektívne aj objektívne porovnávanie výsledkov simulácie a nastavovanie parametrov. Aj keď je možnosť vytvárania GUI v Matlabe (a hlavne jeho starších verziách), značne obmedzená, podarilo sa vytvoriť dostatočne pohodlné rozhranie na spomínané účely. Keďže stavebným prvkom v Matlabe je funkcia, v implementácii GUI je často použitý návrhový vzor, ktorý združuje logicky súvisiace metódy do jednej funkcie. Tieto vnorené funkcie je možné identifikovať ich názvom a volať ich z hlavnej funkcie. Príkladom komponentov rozhrania, ktoré bolo nutné implementovať je sústava záložiek a panelov, ktorá je bežná v iných vývojových prostrediach alebo tzv. zoom-box, použitý na prehľadný pohyb a škálovanie v rámci vizualizácie zvukového signálu.

Medzi ďalšie funkcie aplikácie patrí načítavanie a ukladanie nastavení simulácie do textového súboru jednoduchého formátu, takže je možné ľahko meniť konfiguráciu aj pri použití konzolovej verzie simulácie. Výstup simulácie v podobe impulzov je tiež možné uložiť do súboru a následne použiť na ďalšie spracovanie alebo vizualizáciu. Popis formátu týchto súborov je súčasťou užívateľskej dokumentácie k aplikácii v elektronickej podobe na priloženom médiu.

6 Výsledky a diskusia

Výsledkom tejto práce je implementácia modelu kódovania zvuku v nižších častiach sluchovej dráhy, na ktorom je založený princíp fungovania kochleárneho implantátu. Neoddeliteľnou súčasťou takejto práce je teda aj zhrnutie vplyvu parametrov modelu na jeho výstup a validácia modelu, ktorou skúmame, či výsledky implementácie modelu (počítačovej simulácie) zodpovedajú pozorovaniu reálneho sveta. Na tomto mieste obsahuje väčšina štúdií zaoberajúcich sa kódovaním v kochleárnych implantátoch porovnanie úspešnosti navrhovaného modelu. Táto úspešnosť sa meria ako skóre metódy alebo súboru hodnôt parametrov v testoch zrozumiteľnosti, keď sú priamo pacientom s kochleárnym implantátom predložené impulzné výstupy z danej metódy a ide teda o subjektívne porovnanie.

Ako už bolo spomenuté, výsledkom práce je simulácia, ktorá umožňuje spätnú syntézu vstupného zvuku, a preto bola na porovnávanie vplyvu parametrov modelu zvolená metóda porovnávania rekonštruovaných zvukových výstupov pre jednotlivé parametre s pôvodným vstupným signálom. Metódy porovnania sú podrobne popísané v nasledujúcom odseku.

6.1 Metódy porovnávania výsledkov

Objektívne porovnanie

Na porovnanie dvoch sekvencií zvukového signálu je možné použiť aj objektívne metódy. V oblasti spracovania a rozpoznávania rečového signálu sa často používa metóda objektívneho porovnávania pomocou tzv. kepstrálnych vzdialeností. Táto metóda je súčasťou kepstrálnej analýzy, teda analýzy kepstra, ktoré je definované ako inverzná Fourierova transformácia logaritmu modulu spektrálnej funkcie signálu. Na výpočet kepstra môžeme teda použiť diskretnú fourierovu transformáciu počítanú pomocou algoritmu FFT. Pre výpočet kepstrálnej vzdialenosti použijeme len reálnu časť koeficientov kepstra určenú nasledujúcim vzťahom [19], kde x je vstupný signál a c postupnosť reálnych koeficientov kepstra.

$$c[k] = \text{Re}\{\mathcal{F}^{-1}[\ln|\mathcal{F}[x[n]]|\}]\} \quad (6-1)$$

Kepstrálnu vzdialenosť (d) rádu p , vstupných signálov x_a a x_b definujeme podľa [20] nasledujúcim vzťahom, kde c_a a c_b sú reálne koeficienty kepstra signálov podľa vzťahu 6-1, a výsledná vzdialenosť je normalizovaná do decibelovej škály.

$$d_{(dB)} = 4.3429 \sqrt{\sum_{k=1}^p (c_a[k] - c_b[k])^2} \quad (6-2)$$

Koeficienty kepstra s nulovým indexom reprezentujú energiu signálu pre daný segment a preto nie sú zahrnuté do výpočtu vzdialenosti. Podobne, ako v práci [21], bola aj tu použitá kepstrálna vzdialenosť rádu 30, čo je vhodná voľba z hľadiska náročnosti aj dostatočnej presnosti výpočtu. Spektrálna charakteristika signálu sa v čase mení, a preto je potrebné vstupný signál rozčleniť na kratšie časové segmenty, v ktorých je spektrum signálu ustálené. Na druhej strane sa ale pri segmentácii na úseky s dĺžkou, ktorá neodpovedá perióde signálu

dopúšťame chyby, a preto je potrebné zvoliť vhodný kompromis. Aby sa dosiahlo jemnejšie rozlíšenie, môžu sa jednotlivé segmenty signálu prekrývať. Pre účely porovnania výstupov z jednotlivých simulácií bola použitá segmentácia s dĺžkou okna približne 6 milisekúnd (128 vzoriek pri vzorkovacej frekvencii 22 kHz), prekrývaním 50%, a ako váhová funkcia pre jednotlivé segmenty bolo použité Hammingovo okno. Kepstrálne vzdialenosti pre jednotlivé segmenty sú vypočítané ako priemer zodpovedajúcich segmentov v každom frekvenčnom pásme (pásma sú určené bankou filtrov použitých vo fáze analýzy) a výsledná vzdialenosť je určená ako priemer vzdialeností všetkých segmentov. Čím je kepstrálna vzdialenosť dvoch signálov nižšia, tým sú si viac podobné. Pre dva totožné signály je táto hodnota rovná nule.

Ak teda máme porovnať vplyv nejakého parametra simulácie, jednoducho vygenerujeme zvukový výstup simulácie pre viacero hodnôt testovaného parametra a vypočítame kepstrálnu vzdialenosť týchto výstupov od pôvodného signálu. Tieto vzdialenosti nám tak určia, ktorá z hodnôt parametra dáva výstup najviac podobný pôvodnému signálu, čím získame najvhodnejšie hodnoty parametrov simulácie.

Subjektívne porovnanie

Druhou možnosťou použitou pre porovnanie zvukového výstupu simulácie je subjektívne porovnanie zvukových signálov. Pre jednotlivé nastavenia daného parametra simulácie bola porovnávaná zrozumiteľnosť zvukového výstupu.

6.2 Konfigurácia testu

Pri testovaní vplyvu parametrov bola využitá modularita projektu „CIsim“ a testy boli naprogramované pomocou skriptov v Matlabe, ktoré volajú funkcie simulácie a načítavania vstupov z balíku funkcií „CIsim“. Skript testu automaticky vygeneroval testovacie konfigurácie simulácie a následne ich spúšťal na množine zvukových súborov tak, aby sa časovo náročné operácie ako napríklad návrh banky filtrov nevykonávali zbytočne viackrát. Výsledky v podobe kepstrálnych vzdialeností pre objektívne porovnanie, ale aj rekonštruovaných zvukov pre subjektívny posluš boli po ukončení testu uložené do súborov. Hodnoty výstupných kepstrálnych vzdialeností boli spracované ďalším analyzačným skriptom, ktorý zozbierané údaje prehľadne vykreslil. Všetky testovacie skripty sú súčasťou prílohy tejto práce, a tak je možné jednoducho upravovať alebo pridávať testy skúmajúce vplyvy ďalších parametrov.

6.2.1 Testovacie vstupy

Ako testovacie vstupy boli použité zvukové nahrávky formátu WAVE, z voľne dostupnej audiometrickej databázy z Carnegie Mellon University [22]. Z tejto databázy bolo na testovanie použitých 25 nahrávok od štyroch anglicky hovoriacich respondentov (dva muži,

dve ženy). Nahrávky boli vytvorené pomocou vstavaného mikrofónu PDA zariadenia, a obsahujú niekoľko krátkych viet. Aj keď nahrávky obsahovali nízkofrekvenčný šum, kvalita bola dostatočná a nahrávky nemuseli byť predspracované. Vzorkovacia frekvencia nahrávok je 16 kHz, a tak bolo možné pracovať s frekvenčným pásmom až do 8 kHz.

Pre rozdielny spektrálny charakter signálu znej a neznej časti reči boli do testov osobitne zahrnuté tiež nahrávky samohlások a spoluhlások. Takéto rozdelenie nám umožní sledovať vplyv parametrov výhradne na znelé a neznelé časti reči. Databáza [22] neobsahovala osobitné nahrávky samohlások a spoluhlások, preto boli vytvorené zvukové sekvencie špeciálne pre tento test (nahrávky samohlások a spoluhlások od dvoch respondentov). Nahrávanie prebehlo pomocou elektretového mikrofónu a bežnej zvukovej karty, pri vzorkovacej frekvencii 22,05 kHz. Na odstránenie šumu a jednosmernej zložky zo signálu musel byť použitý high-pass filter s hraničnou frekvenciou 250 Hz. Pri strihu nahrávok bol použitý software GoldWave [23].

6.2.2 Testované parametre

Aj keď simulácia „CIsim“ umožňuje nastaviť až 25 rôznych parametrov spracovania vstupného súboru, rozhodol som sa otestovať len najdôležitejšie parametre, ktoré majú kritický vplyv na výsledok. Zároveň bol kladený dôraz na možnosť porovnania výsledkov z podobnými prácami a preto boli testované len tieto parametre:

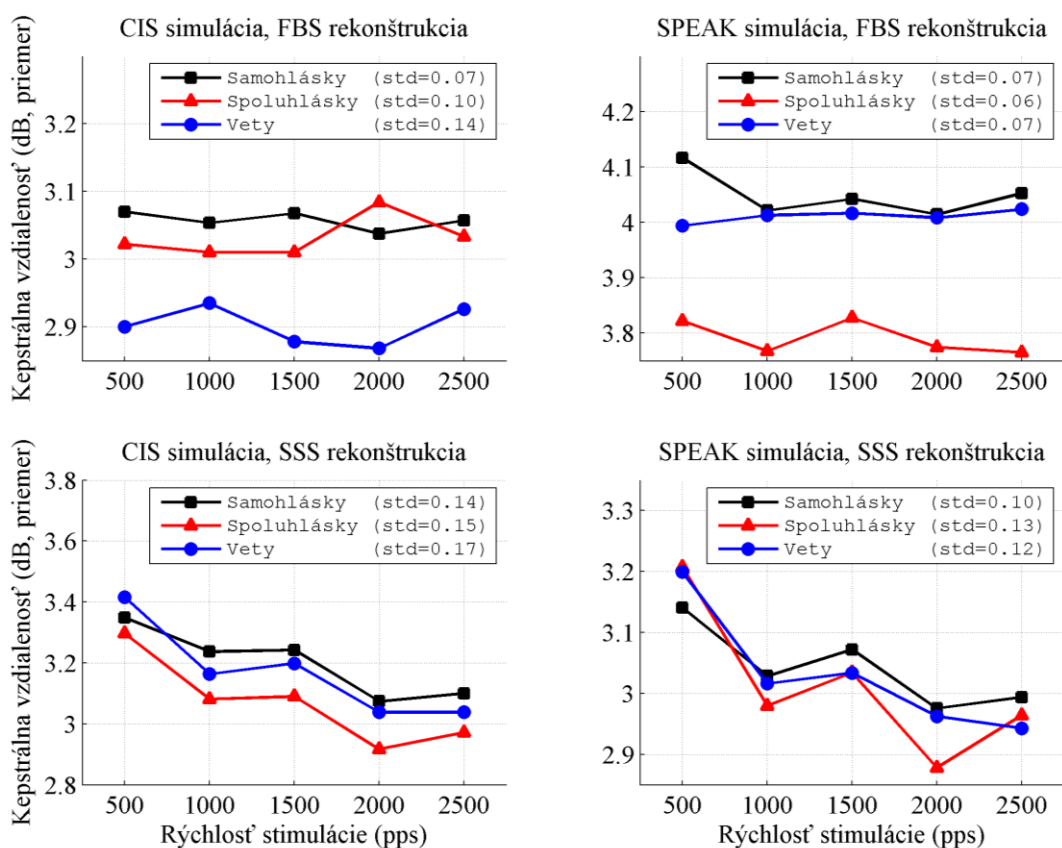
- rýchlosť stimulácie
- frekvenčný rozsah spracovávaného pásma
- počet frekvenčných pásiem
- metóda rozloženia hraničných frekvencií
- typ banky filtrov
- typ kompresnej funkcie a hodnota koeficientu kompresie
- typ extrakcie obálky signálu
- počet spektrálnych maxím stratégie SPEAK

Celkovo sa vykonalo 165 testov v ôsmich častiach, s výstupom 2090 zvukových nahrávok. Pre výstupy spektrálnych vzdialeností boli pre jednotlivé testy vypočítané priemerné hodnoty a rozptyl hodnôt v rámci testovanej množiny súborov.

6.3 Výsledky testovania vplyvu parametrov simulácie

Rýchlosť simulácie

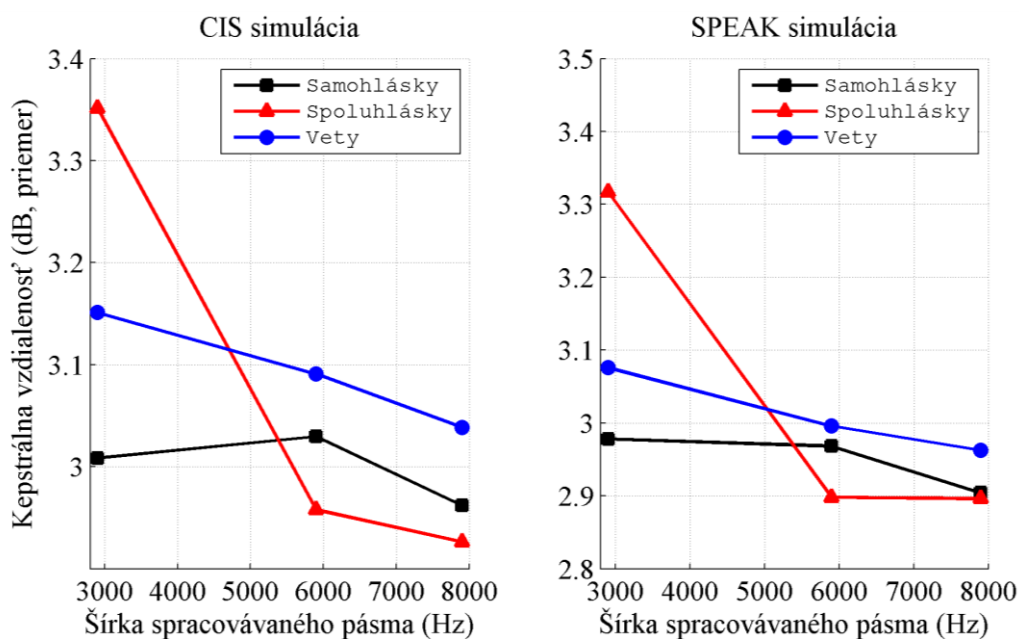
Test vplyvu rýchlosti simulácie prebiehal pre hodnoty 500 pps až 2500 pps s krokom 500 pps, osobitne pre CIS aj SPEAK simuláciu a syntézu typu FBS, aj SSS. Nastavenie simulácie bolo štandardné, čiže 8-pásmová banka typu IIR a extrakcia obálky pomocou usmernenia a filtrovania. Výsledok testu je zobrazený na obrázku 6-1. Predpokladané bolo znižovanie kepstrálnej vzdialenosti pri zvyšovaní rýchlosti simulácie, ktorá reprezentuje časové rozlíšenie informácie vysielanej do výstupu. Pre FBS rekonštrukciu je pokles kepstrálnej vzdialenosti výrazne menší, pre SSS rekonštrukciu je klesajúca tendencia viac viditeľná. Pri hranici 2500 pps môžeme pre niektoré výsledky pozorovať aj nárast. Pre výsledky, kde sú vstupom vety, je smerodajná odchýlka vyššia, čoho príčinou je vyššia variabilita vstupov (muži/ženy, znelá/neznelá zložka) ako pri samohláskach a spoluhláskach. Najmenšie hodnoty kepstrálnych vzdialeností sú dosiahnuté pomocou CIS simulácie s FBS rekonštrukciou, najvyššie zase pre SPEAK, čo môže byť dôsledkom práve výberu maxim do výstupného signálu, čím sa zvyšná časť frekvenčného spektra zanedbáva.



Obrázok 6-1 Vplyv rýchlosti simulácie na výsledok simulácie. Obrázok obsahuje výsledky pre kombinácie oboch stratégií CIS a SPEAK a rekonštrukčných metód FBS a SSS. Na osi x je znázornený skúmaný rozsah rýchlosti simulácie, na osi y priemerná hodnota normalizovanej kepstrálnej vzdialenosti.

Frekvenčný rozsah spracovávaného pásma

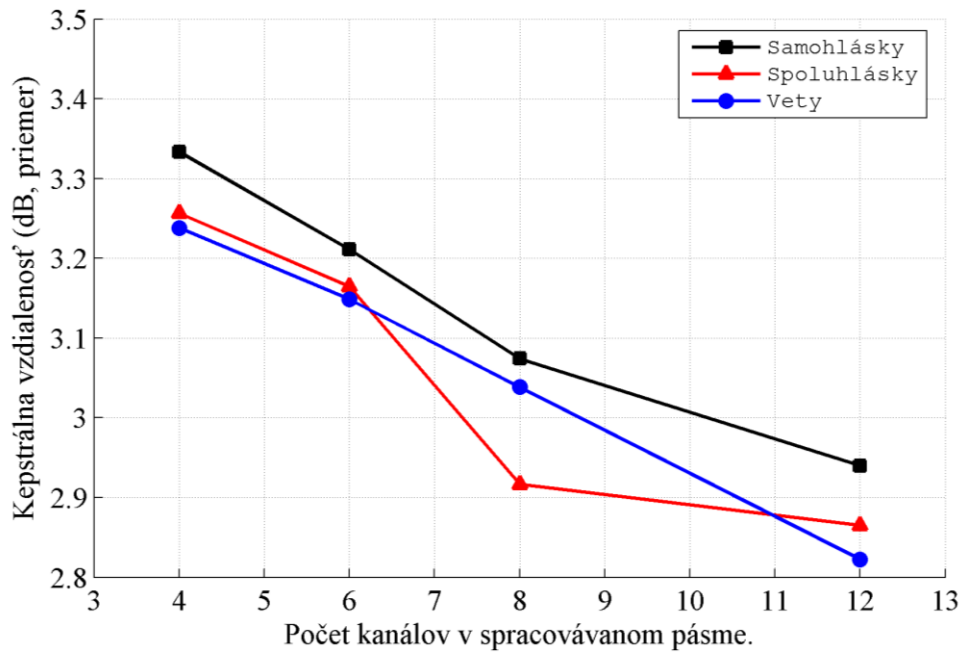
Ďalší z testov sa venuje vplyvu šírky pásma, z ktorého sa berie informácia o frekvencii v signáli na výsledok simulácie. Predpokladom je samozrejme znižovanie kepstrálnej vzdialenosti pri zvyšovaní šírky tohto pásma. Test prebiehal samostatne pre CIS a SPEAK simuláciu so štandardným nastavením. Šírka pásma sa pohybovala od 3 kHz, čo je podľa [24] najmenšia hodnota tohto parametru pre porozumenie rečového signálu, po 8 kHz, čo bola maximálna možná hodnota pre nahrávky danej kvality. Na obrázku 6-2, kde je znázornený výsledok testu, pozorujeme výrazné zlepšenie pre vstupy spoluhlások pri oboch simuláciách. To je spôsobené tým, že do spracovania bolo zahrnuté aj frekvenčné pásmo vyšších formantov, ktorými je charakteristická neznelá časť reči.



Obrázok 6-2 Vplyv šírky spracovávaného pásma na výstup simulácie. Na obrázku sú znázornené výsledky testov pre CIS a SPEAK simuláciu. Na osi x je šírka spracovávaného frekvenčného pásma, na osi y priemerná hodnota normalizovanej kepstrálnej vzdialenosti.

Počet frekvenčných pásiem

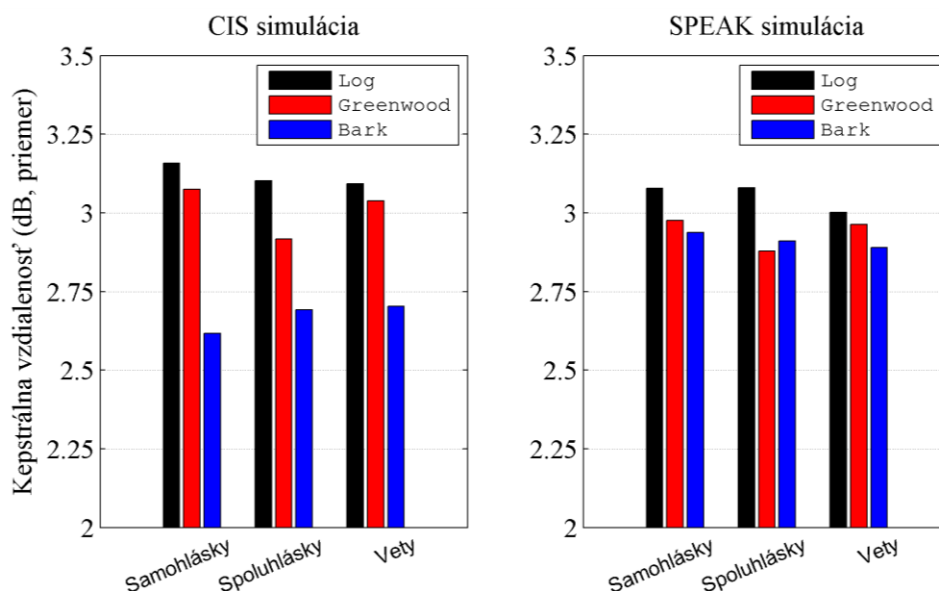
Očakávaný výsledok zlepšenia kódovania bol pozorovaný aj pri zvyšovaní frekvenčného rozlíšenia v podobe počtu kanálov v spracovávanom pásme. Test prebiehal pri štandardnom nastavení CIS simulácie a počet kanálov stúpал od štyroch po dvanásť. Zlepšenie výsledku (obrázok 6-3) pre vety a samohlásky je lineárne, pre spoluhlásky sa objavuje opäť výraznejšie zlepšenie pri simulácii s počtom kanálov 8.



Obrázok 6-3 Vplyv počtu frekvenčných pásiem na výsledok simulácie. Obrázok znázorňuje klesajúcu tendenciu priemernej kepstrálnej vzdialenosti (os y) výstupov CIS simulácie v závislosti na stúpajúcom počte kanálov (os x).

Metóda rozloženia hraničných frekvencií

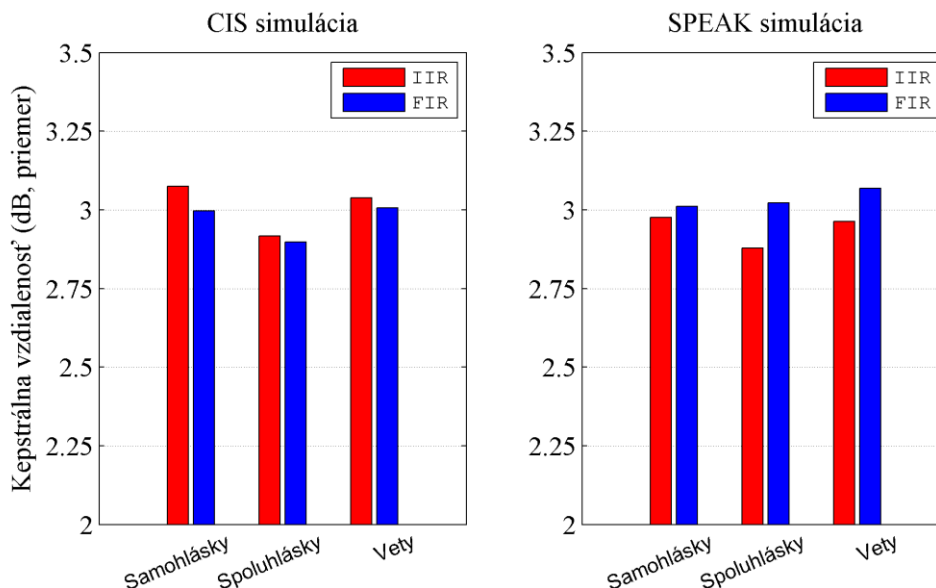
Pri testovaní sa ukázalo, že v porovnaní s ostatnými rozdeleniami spracovávaného pásma, je rozdelenie podľa kritických pásiem (na obrázku 6-4 označené ako Bark) najlepšie, druhým v poradí je rozdelenie vychádzajúce z Greenwoodovej vzdialenosti, a nakoniec je to logaritmické rozdelenie. Výsledky dopadli pre všetky testované množiny, až na jednu výnimku (spoluhlásky a SPEAK simulácia) rovnako. Pri simulácii boli opäť použité štandardné nastavenia pre CIS aj SPEAK.



Obrázok 6-4 Vplyv metódy rozloženia hraničných frekvencií na výsledky simulácie. Na obrázku je znázornená priemerná kepstrálna vzdialenosť (os y), pre obidva typy kódovania, ako aj pre jednotlivé typy rozloženia hraničných frekvencií použitých v simulácii.

Typ banky filtrov

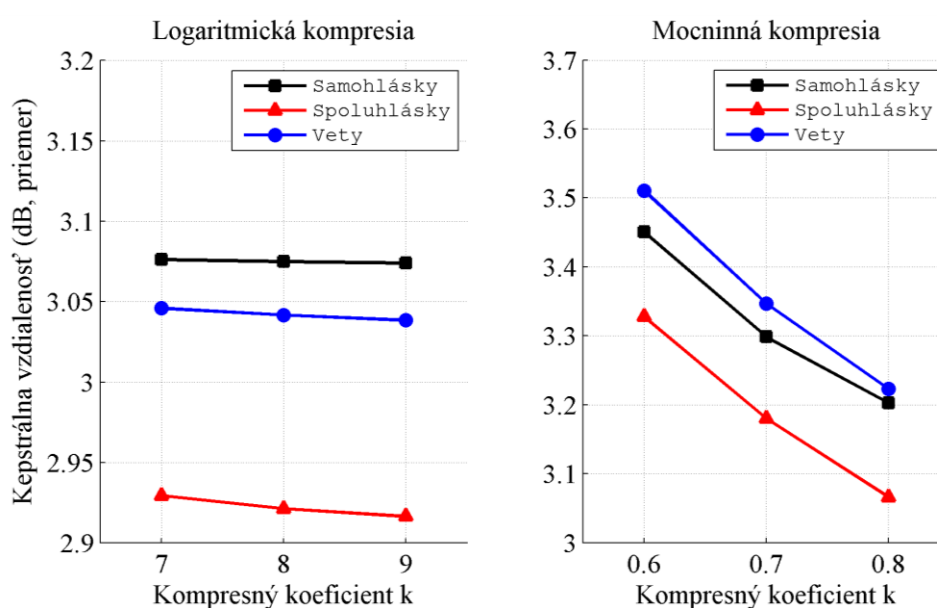
Výsledky vplyvu použitého typu banky filtrov na výsledok simulácie zobrazené na obrázku 6-5 ukazujú, že rozdiel medzi typom použitej banky je len veľmi malý. Pre CIS simuláciu bolo výhodnejšie použitie FIR banky, zatiaľ čo pri SPEAK simulácii je výsledok pre všetky tri testované skupiny opačný.



Obrázok 6-5 Vplyv zvoleného typu banky filtrov na výsledok simulácie. Na obrázku je znázornená priemerná kepstrálna vzdialenosť (os y), osobitne pre FIR a IIR typ banky filtrov, samostatne pre každý typ kódovania použitý v simulácii.

Typ kompresnej funkcie

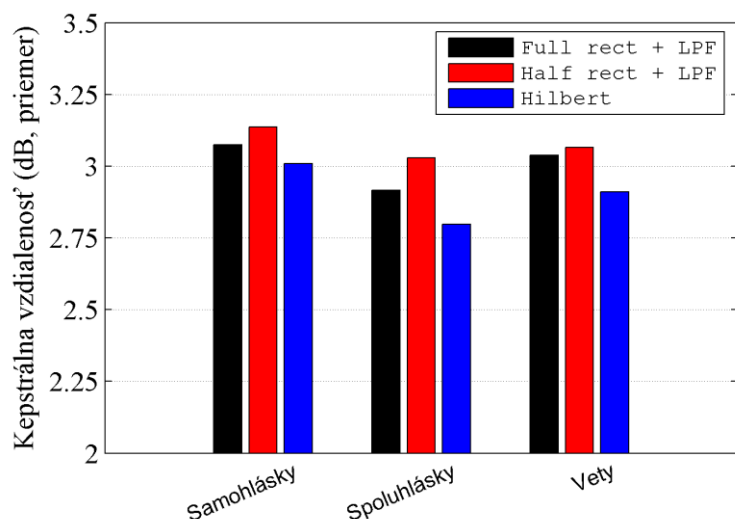
Výsledky testu kompresných funkcií pre CIS simuláciu na obrázku 6-6 ukazujú jednoznačne v prospech logaritmickej kompresnej funkcie. Hodnoty kepstrálnej vzdialenosti pre obidva typy použitej kompresie majú klesajúcu tendenciu so zvyšujúcou sa hodnotou kompresného koeficientu. Klesajúcu tendenciu je možné pozorovať pre všetky skupiny testovaných vstupov. Pre logaritmickú kompresnú funkciu sa najviac osvedčila hodnota kompresného koeficientu $k \cong 9$. Aj keď sa v niektorých prácach podobného zamerania používa mocninná kompresná funkcia, ako napr. v [12], alebo [25], v tejto práci sa jej použitie neosvedčilo a kepstrálna vzdialenosť sa spolu so znižovaním kompresie (hodnota k sa blíži 1) znižuje. Porovnanie zvukových výstupov hovorí tiež v prospech logaritmickej funkcie. Zvukové výstupy odpovedajúce mocninnej funkcii obsahujú zašumenie spôsobené pravdepodobne zosilnením nízkofrekvenčného šumu, ktorý sa v nahrávkach nachádzal.



Obrázok 6-6 Vplyv typu kompresnej funkcie na výsledok simulácie. Obrázok znázorňuje pokles priemernej kepstrálnej vzdialenosti, pre logaritmickú a mocninnú kompresnú funkciu so stúpajúcim kompresným koeficientom. Výsledky boli namerané pri CIS simulácii.

Typ extrakcie obálky signálu

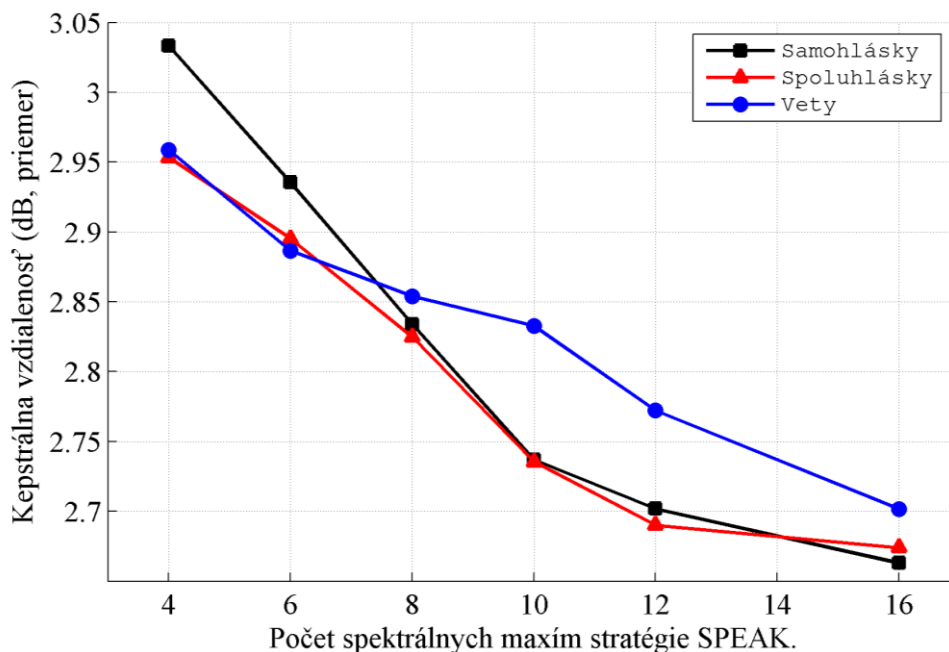
Test monitorujúci vplyv použitého algoritmu na extrakciu obálky (obrázok 6-7) dopadol podľa očakávania a najnižšie hodnoty kepstrálnej vzdialenosti boli dosiahnuté použitím extrakcie pomocou Hilbertovej transformácie. Druhým v poradí úspešnosti je jednocestné usmernenie s filtrovaním a ako posledné dvojcestné usmernenie. Testovanie prebiehalo s použitím štandardnej konfigurácie simulácie CIS. Napriek nameraným hodnotám sa pri subjektívnom porovnaní výstupných zvukov javili výsledky simulácie s dvojcestným usmernením zrozumiteľnejšie ako pri použití jednocestného usmernenia. Pre všetky testované skupiny je hodnota kepstrálnej vzdialenosti približne rovnaká.



Obrázok 6-7 Vplyv typu extrakcie obálky na výsledok simulácie. Na obrázku sú znázornené priemerné hodnoty kepstrálnej vzdialenosti (os y), pre tri použité druhy extrakcie obálky signálu. Hodnoty boli namerané pri CIS simulácii.

Poččet spektrálnych maxim stratégie SPEAK

Posledným z testovaných parametrov bol počet vybraných spektrálnych maxim SPEAK stratégie so štandardným nastavením výberu 4 až 16 maxim z dvadsiatich spracovávaných kanálov. Spolu so zvýšením tohto počtu sa do výstupného signálu dostáva väčšia časť informácie z frekvenčného spektra a to spôsobuje predpokladaný pokles kepstrálnej vzdialenosti, ktorý pozorujeme pre všetky testované skupiny vstupov na obrázku 6-8.



Obrázok 6-8 Vplyv počtu spektrálnych maxim na výsledok simulácie. Obrázok znázorňuje očakávaný pokles kepstrálnej vzdialenosti (os y), spolu s rastúcim počtom vybraných spektrálnych maxim SPEAK stratégie.

6.4 Súhrn a porovnanie výsledkov z ďalších dostupných zdrojov

Výsledky testovania dopadli prevažne podľa očakávaní, až na spomínanú výnimku pri testovaní kompresnej funkcie. Aj keď sa výsledky metódy SPEAK javili pri subjektívnom posluhu zrozumiteľnejšie, väčšinou získala lepšie objektívne hodnotenie simulácia metódy CIS a v subjektívnom hodnotení dopadla lepšie tiež metóda rekonštrukcie signálu typu SSS. Výsledky zo simulácie „CIsim“ môžeme porovnať napríklad z výsledkami v práci [21], kde sa používa rovnaký typ objektívneho hodnotenia pomocou kepstrálnej vzdialenosti. V prípade testov vplyvu rýchlosti stimulácie a aj počtu kanálov metód CIS a SPEAK boli namerané podobné hodnoty kepstrálnej vzdialenosti, a tiež bol pozorovaný rovnaký vplyv týchto parametrov na výsledok. Výsledky testov šírky pásma a rýchlosti stimulácie môžeme porovnať tiež s výsledkami štúdie [25], kde bola zistená rovnaká forma závislosti úspešnosti kódovania na týchto parametroch. Zistenia v práci [25] boli založené na subjektívnych testoch zrozumiteľnosti predložených pacientom používajúcim kochleárny implantát. V prípade práce [21], podobne ako v tejto práci, ide len o simuláciu založenú na modeli kódovania v sluchovej dráhe človeka.

6.5 Možné vylepšenia simulácie

Aj keď aplikácia „CIsim“ obsahuje oproti ostatným podobným implementáciám výber z väčšieho množstva parametrov, ktorými je možné ovplyvňovať výsledky simulácie, existuje mnoho faktorov, ktoré sa v simulácii nezohľadňujú. Jedným z týchto faktorov je interakcia medzi elektródami, ktorá by mohla byť v simulácii realizovaná ako modul zapojený medzi generátor elektrických impulzov a modul syntézy signálu. Na základe modelu šírenia elektrických impulzov v tkanive by sa tak vypočítala miera ovplyvnenia kanálov. Na toto vylepšenie by automaticky mohlo byť nasadené prekladanie impulzov v čase, ktoré by účinky interferencie medzi kanálmi malo potlačiť.

Z hľadiska vylepšenia implementácie by sa mohla uplatniť paralelizácia výpočtov, pretože vo väčšine aplikácií pre digitálne spracovanie signálu sa na jednotlivých blokoch dát vykonáva ten istý druh operácie a hovoríme o tzv. SIMD (Single Instruction Multiple Data) inštrukciách. Vylepšením, kde by paralelizácia výpočtu hrala dôležitú úlohu je simulácia kochleárneho implantátu v reálnom čase. Jedinou zmenou potrebnou na real-time spracovanie signálu simuláciou je rozdelenie vstupného signálu na bloky, ktorých spracovanie by trvalo kratší čas, ako je doba ich prehrávania. Základné funkcie Matlabu však neumožňujú prácu so zvukovým streamom, a preto by bolo nutné použiť niektoré z riešení pre sprístupnenie nižších vrstiev systému spracovania zvuku. Príkladom takého riešenia by mohla byť utilita PlayRec [26], ktorej funkcionálna je však úzko zviazaná s použitím konkrétnej platformy.

7 Záver

Cieľom tejto práce bolo implementovať model kódovania zvuku v sluchovej dráhe a vytvoriť simuláciu kochleárneho implantátu. Výsledkom je aplikácia „CIsim“, ktorá je porovnateľná s ostatnými existujúcimi štúdiami, čo sa týka veľkosti konfiguračného priestoru, ale aj z hľadiska miery úspešnosti kódovania, ktorá bola určená na základe objektívneho porovnania výstupov simulácie. Navyše riešenie v tejto práci je jedným z mála projektov, ktoré umožňujú nastavenie, ale aj vizualizáciu a súčasné porovnanie výsledkov simulácií v užívateľskom rozhraní. Tým je značne zjednodušená práca s aplikáciou simulácie a je možné ju využiť aj na demonštračné účely a vysvetlenie princípu spracovania zvuku v kochleárnych implantátoch. Veľké množstvo podobných prác v oblasti spracovania zvuku pre kochleárne implantáty pochádza z komerčnej sféry a zverejňuje len výsledky výskumu, zatiaľ čo implementačné detaily, ktoré by mohli často objasniť variabilitu vo výsledkoch výskumov, nie sú známe. Zdrojové súbory simulácie „CIsim“ sú voľne dostupné a aplikácia je navrhnutá tak, aby bolo možné do simulácie jednoducho pridávať nové moduly rozširujúce model kódovania. Výber vhodného a hlavne vo výskumnej sfére často používaného prostredia Matlabu len podporuje celkový princíp rozširiteľnosti.

Literatúra

- [1] Niparko J. K.: Cochlear Implants: Principles and Practices. Philadelphia: Lippincott Williams & Wilkins, 2000.
- [2] Maršálek P., Kvasnička V., Pospíchal J.: Simulace v neurovědách, příklad modelu prostorového slyšení. *Umelá inteligencia a kognitívna veda I*, STU, Bratislava, Slovakia, 2009, p.281-298.
- [3] Flanagan J. L.: Speech Analysis and Perception. Berlin: Springer-Verlag, 1965.
- [4] Clark G.M.: Personal reflections on the multichannel cochlear implant and a view of the future. *Journal of Rehabilitation Research and Development*, 2008, vol.45, p.651-693.
- [5] Stevens S., Volkman J., Newman E.: A scale for the measurement of the psychological magnitude of pitch. *Journal of the Acoustical Society of America*, 1937, vol.8, p.185–190.
- [6] O'Shaughnessy D.: Speech communication: human and machine. Addison-Wesley, 1987.
- [7] Zwicker E., Fastl H.: Psychoacoustics, Facts and Models. Berlin: Springer-Verlag, 1990.
- [8] Ingle V.K., Proakis J.G.: Digital Signal Processing Using Matlab. Thomson Learning, 1999.
- [9] Schimmel S.: SPHSC 503: Speech Signal Processing. University of Washington, 2006. (<http://isd1.ee.washington.edu/people/stevenschimmel/sphsc503>)
- [10] Clark G.M.: Cochlear implants: fundamentals and applications. Springer, 2003.
- [11] Loizou P.C.: Speech processing in vocoder-centric cochlear implants. *Advances in Oto-Rhino-Laryngology*, 2006, vol.64, p.109–143.
- [12] Sen D., Li W., Chung D., Lam, P.: Exploring the parameter space of Cochlear Implant Processors for consonant and vowel recognition rates using normal hearing listeners. Proc. Eleventh Australasian International Conference on Speech Science and Technology, 2006.
- [13] Rubinstein J., Wilson B., Finley C., Abbas P.: Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation. *Hearing Research*, 1999, vol.127, p.108-118.
- [14] McDermott H.J., McKay C.M., Vandali A.E.: A New Portable Sound Processor for the University of Melbourne/Nuclues Limited Multielectrode Cochlear Implant. *Journal of the Acoustical Society of America*, 1992, vol.91, p.3367-3371.
- [15] Loizou P.C., Lobo A., Hu Y.: Subspace algorithms for noise reduction in cochlear implants. *Journal of the Acoustical Society of America*, 2005, vol.118, p.2791-2793.

- [16] Kabal P.: Audio File Format Specifications - WAVE or RIFF WAVE sound file. McGill University, 2006.
(<http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE>)
- [17] Kehtarnavaz N.: Digital Signal Processing System Design: LabVIEW-Based Hybrid Programming. Elsevier/Academic Press, 2008.
- [18] Greenwood D.D.: A cochlear frequency-position function for several species–29 years later. *Journal of the Acoustical Society of America*, 1990, vol.87, p.2592-2605.
- [19] Rabiner L., Juang B.: Fundamentals of Speech Recognition. Prentice Hall Signal Processing Series. Prentice Hall, 1993.
- [20] Pollák P., Sovka P.: Cepstral speech/pause detectors. Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing, 1995, p.388-391.
- [21] Vondrášek, M.: Speech Preprocessing for Cochlear Implants. (Dizertačná práca) Praha, ČVUT, 2008.
- [22] Obuchi, Y.: CMU PDA Speech Database.
(<http://www.speech.cs.cmu.edu/databases/pda>)
- [23] GoldWave Digital Audio Editor. (<http://www.goldwave.com>)
- [24] Peterson G., Barney H.: Control methods used in a study of vowels. *Journal of Acoustical Society of America*, 1952, vol.24, p.175-184.
- [25] Loizou P., Poroy O., Dorman M.: The effect of parametric variations of cochlear implant processors on speech understanding. *Journal of Acoustical Society of America*, 2000, vol.108, p.790-802.
- [26] Playrec Multi-channel Matlab Audio. (<http://www.playrec.co.uk>)

Prílohy

A. Obsah priloženého CD

V adresárovej štruktúre na priloženom médiu (členenej podľa obsahu) sú uložené zdrojové súbory aplikácie, užívateľská dokumentácia, generovaná dokumentácia, vzorové konfiguračné súbory a zvukové vstupy, testovacie skripty a distribúcia kompilovanej aplikácie „CIsim“ pre systém Windows.

/	
readme.txt	- textová dokumentácia k obsahu CD
index.html	- hlavná stránka
dist/	- adresár obsahuje spustiteľné súbory aplikácie CIsim a inštalátor MCR
src/	- zdrojové súbory aplikácie CIsim
plots/	- skripty generujúce grafy do textu práce
res/	- ostatné súbory potrebné pre chod aplikácie
data/	- vzorové konfiguračné súbory a testovacie zvukové vstupy
cons/	- zvukové nahrávky spoluhlások
vovs/	- zvukové nahrávky samohlások
sents/	- zvukové nahrávky viet
test/	- adresár obsahuje testovacie skripty
doc/	- užívateľská dokumentácia, generovaná dokumentácia a text práce
