

IMPLEMENTACE OBJEKTIVNÍHO MODELU HODNOCENÍ KVALITY ZVUKU PEMO-Q V PROSTŘEDÍ MATLAB SE ZAHRNUTÝM MODELEM SLUCHOVÉ CESTY A MODELEM CASP

M. Zalabák

Katedra radioelektroniky, ČVUT FEL v Praze

Abstrakt

Cílem objektivních modelů hodnocení kvality je být alternativou k subjektivnímu testování kvality zvuku. Tématem tohoto článku je popis implementace modelu hodnocení PEMO-Q využívajícího model sluchové cesty v prostředí MATLAB. Dále byl implementován model sluchové cesty CASP sloužící jako možná substituce zahrnutého modelu sluchové cesty použitého v PEMO-Q.

Kromě aspektů spojených s implementací modelu PEMO-Q jsou také probrány možnosti optimalizace vyhodnocení modelu v prostředí MATLAB, jako například možnosti paralelizace či delegace výpočetně kritických částí na dílčí funkce využívající kompilovaný kód.

1 Úvod

Objektivní modely hodnocení kvality slouží jako alternativa k subjektivním poslechovým testům. Pro takové testy je vhodné jejich provedení s velkým počtem posluchačů pro dostatek dat ke zhodnocení vzhledem k velké variabilitě vnímání napříč posluchači. S tím jsou spojené velké materiální a časové požadavky pro realizaci. To je motivací k vytvoření alternativy založené na signálové analýze.

Jednou z takových metod je implementovaný model PEMO-Q, definovaný v [5]. Tento model je založený na modelu sluchové cesty lidského ucha. Výstupní hodnoty popisující kvalitu zvuku jsou získány vzájemnou korelací hodnot sluchové cesty pro referenční, „dokonalý“ zvuk s hodnotami sluchové cesty testovaného zvuku.

V kapitole 2 je tento model popsán. Dále je též popsán model sluchové cesty CASP popsáný v [6], který je možná alternativa pro model sluchové cesty zahrnutý v PEMO-Q. V kapitole 3 je popsána samotná implementace v prostředí MATLAB a aspekty s ní spojené. V závěru jsou probrány potenciální možnosti optimalizace modelu. Shrnutí a komentáře o současném stavu v této problematice lze nalézt v [3].

2 Teoretická část

2.1 Model sluchové cesty z PEMO-Q

Model sluchové cesty zahrnutý v modelu PEMO-Q popsáný v [5] a [3] je složen z několika bloků modelujících konkrétní části sluchové cesty. První blok je banka 35 gammatónových filtrů čtvrtého řádu o středních kmitočtech od 235 Hz do 14.5 kHz napodobujících bazilární membránu. Odstupy jednotlivých středních hodnot a šířek pásem filtrů odpovídají hodnotě 1 ERB (přepočty mezi ERB a Hz lze nalézt v [2]). Jednotlivá výstupní pásma jsou dále zpracována nezávisle.

Po následné transformaci z mechanických kmitů na nervové impulzy půlvlnným usměrněním a filtrací dolní propustí o kmitočtu 1 kHz je signál limitován dolní mezí závislou na maximální úrovni vstupního signálu (pro amplitudu rovnou 1 hodnotou 10^{-5} dle [3]) a přiveden na blok adaptivní filtrace. Ten se skládá z pěti zpětnovazebných smyček zapojených do kaskády s dolnoproputným RC článkem ve zpětné vazbě (časové koeficienty o hodnotách od 5 do 500 ms). Výstupem zpětné vazby je signál vydělen. Stacionární signály jsou přibližně sníženy na úroveň 32. odmocniny a velké změny mají téměř lineární charakter.

Po bloku adaptivní filtrace zbývá modulační banka filtrů modelující schopnost rozpoznání amplitudově modulovaného zvuku složené z osmi filtrů. Prvním je dolní propust druhého řádu nastavené na kmitočet 2.5 Hz. Další dva jsou pásmové filtry s konstantní šířkou pásma 5 Hz a středními kmitočty 5 a 10 Hz. Zbylé filtry jsou rovněž pásmové propusti, konkrétně filtry s konstantním koeficientem Q rovným 2 a překryvy na -3 dB (nejnižší filtr se na tomto poklesu překrývá s filtrem na 10 Hz). Nejvyšší střední kmitočet odpovídá hodnotě 129 Hz. Pro redukci redundantních dat je pro další zpracování spočtena Hilbertovská obálka signálů a jejich podvzorkování nejméně na šestinásobek středního resp. mezního kmitočtu filtrů. Další detaily o tomto modelu sluchové cesty lze nalézt v [5] a [3] a blokové schéma je zobrazeno na obrázku 1.



Obrázek 1: Blokové schéma modelu sluchové cesty z PEMO-Q

2.2 Model sluchové cesty CASP

Alternativou pro výše popsaný model je model sluchové cesty CASP popsaný v [6], který obsahuje identické bloky vlasových buněk (usměrnění a dolní propust), adaptivní kaskádu a modulační banku filtrů.

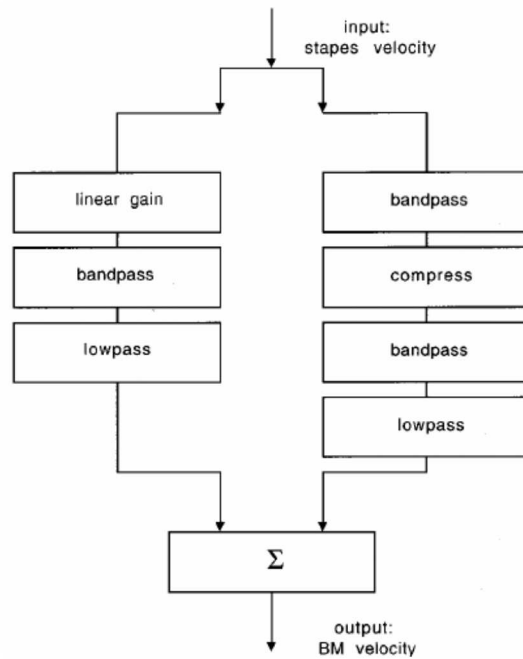
Na vstup modelu je ovšem přidán blok modelující vnější a střední ucho transformující vstupní signál v podobě výchylky tlaku na hodnotu odpovídající rychlosti pohybu třmínku skrze dva 512-ti bodové FIR filtry vycházejících z publikovaných měření (detaily viz [6]).

Alternativně lze použít hodnotu výchylky třmínku skrze kaskádu dvou rezonančních filtrů se středními hodnotami $f_c = \{2.5, 4.75\}$ kHz, šířkami pásma $BW = \{1.5, 4.5\}$ kHz a zisky 10 a 25 dB. Jejich výstup je sečten s originálním signálem, filtrován Butterworthovou dolní propustí prvního řádu o kmitočtu 50 Hz, vynásoben skalárem $45 \cdot 10^{-9}$ a filtrován Butterworthovou horní propustí prvního řádu o mezním kmitočtu 1 kHz. Tato metoda je popsána v [9].

V případě modelu bazilární membrány byla místo banky gammatónových filtrů použita DRNL banka složená z lineární a nelineární větve. Lineární část obsahuje lineární zisk, kaskádu identických gammatónových filtrů prvního řádu se středními kmitočty a šířkami pásma rovnými 1 ERB (jako v případě filtrů v PEMO-Q) a kaskádu identických dolních propustí s mezním kmitočtem rovným střednímu kmitočtu gammatónových filtrů. Pokud byla v bloku vnějšího a středního ucha využita rychlost pohybu třmínku, je lineární zisk na vstupu frekvenčně závislý.

Nelineární část obsahuje rovněž kaskády gammatónových filtrů prvního řádu a dolní propusti na stejném kmitočtu s gammatónovými filtry, avšak jsou využity rozdílné kmitočty v porovnání s lineární částí. Dále se v nelineární části nachází kompresní prvek využívající nelinearitu „zlomené tyče“: hodnoty pod určitou mezní úrovní mají lineární přenos, avšak nad touto úrovní je signál komprimován konstantním koeficientem. Blokové schéma DRNL banky s konkrétním zapojením lze nalézt v obrázku 2 a konkrétní hodnoty, popř. metody jejich výpočtu lze nalézt v [8], [6] a [9].

Zbylé odlišnosti modelu CASP od modelu zahrnutého v PEMO-Q jsou zisk (50 dB z [6], popř. 38 dB z [3]) a následná druhá mocnina mezi dolní propustí modelu vlasových buněk a adaptivní kaskádou. V některých zdrojích (např. v [3] je uvedena též dolní propust prvního řádu o mezním kmitočtu 150 Hz na vstupu modulační banky. Blokové schéma lze též vidět na obrázku 3.



Obrázek 2: Blokové schéma DRNL bloku

2.3 Vyhodnocující část

Třídídimenzionální výstupy modelu sluchové cesty pro referenci a testovaný zvuk představuje vnitřní reprezentaci vstupních signálů. Před samotnou korelací je v modelu PEMO-Q z definice v [5] zahrnuta asimilace výstupu zkresleného signálu,

$$\hat{y}_{tfm} = \begin{cases} \frac{y_{tfm} + x_{tfm}}{2}, & |y_{tfm}| < |x_{tfm}| \\ y_{tfm}, & |y_{tfm}| \geq |x_{tfm}| \end{cases} \quad (1)$$

kde x představuje hodnotu interní reprezentace referenčního signálu a y hodnotu zkresleného. Asimilace je provedena ve všech třech dimenzích a vychází z předpokladu, že „chybějící“ prvky ve zvuku jsou méně subjektivně rušivé než „přidané“.

Korelace je dle [5] počítaná pro každý modulační kanál zvlášť přes všechny hodnoty v čase a frekvenčním kanálu tímto vztahem:

$$r = \frac{\sum_{t,f} (x_{tf} - \bar{x})(y_{tf} - \bar{y})}{\sqrt{\sum_{t,f} (x_{tf} - \bar{x})^2 \sum_{t,f} (y_{tf} - \bar{y})^2}} \quad (2)$$

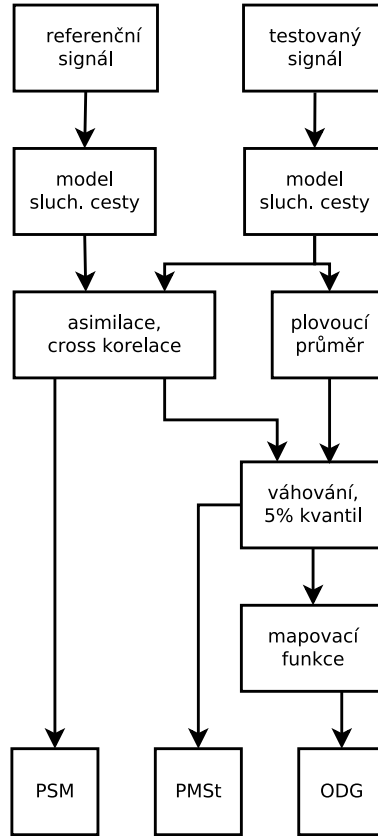
Hodnoty \bar{x} , \bar{y} představují průměrné hodnoty reprezentací přes čas a frekvenční pásma.

První výstupní hodnota celého modelu, PSM , je spočtena pomocí pomocného koeficientu w_m spočteného jako poměr sumy kvadrátů všech vzorků pro jednotlivé modulační kanály ku jejich celkovému součtu,

$$w_m = \frac{\sum_{t,f} y_{tfm}^2}{\sum_{t,f,m} y_{tfm}^2} \quad (3)$$



Obrázek 3: Blokové schéma modelu sluchové cesty CASP



Obrázek 4: Blokové schéma implementovaného modelu PEMO-Q

a sumy jejich násobků s výsledky vzájemných korelací:

$$PSM = \sum_m w_m r_m \quad (4)$$

Další výstupní hodnotou je časově závislá hodnota PSM_t dle [5] a [3] spočtená identickou vzájemnou korelací, avšak pro 10 ms rámce interních reprezentací. Tyto dílčí korelace jsou váhovány pohyblivým průměrem časového průběhu interní reprezentace, neboli „okamžitou aktivitou“. Výsledná hodnota je 5%-ní kvantil váhovaných krátkodobých korelací.

Poslední krok je přepočtení hodnoty PSM_t na hodnotu ODG , neboli „Objective Difference Grade“. Jedná se o hodnotu srovnatelnou se subjektivními testy, tedy v mezích od 0 (identita) do -4 (velmi rušivé zkresení) oproti hodnotám od 1 do 0 v případech PSM a PSM_t , kde hodnota 1 vyjadřuje identitu:

$$ODG(x) = \begin{cases} \max\{-4, \frac{a}{x-b} + c\}, & x < x_0 \\ d \cdot x - d, & x \geq x_0 \end{cases} \quad (5)$$

Konstanty použité při přepočtu jsou $a = -0.22$, $b = 0.98$, $c = -4.13$, $d = 16.4$ a $x_0 = 0.864$. Blokové schéma celého modelu je znázorněno na obrázku 4.

3 Implementace

Model vnějšího a středního ucha pro model CASP byl realizován pomocí metody výchylky třmínku. Koeficienty rezonančních filtrů byly spočteny těmito vztahy z [9]:

$$b_0 = -b_2 = \frac{1}{1 + \cot q} \quad q = \pi \frac{1}{f_s} BW \quad (6)$$

$$a_1 = -\frac{2 \cdot \cos(2\pi f_c / f_s)}{(1 + \tan q) \cos q} \quad (7)$$

$$a_2 = -\frac{\tan q - 1}{\tan q + 1} \quad (8)$$

kde koeficient b_1 je nulový. Následuje Butterworthova dolní propust prvního řádu na 50 Hz, vynásobení hodnotou $45 \cdot 10^{-9}$ a dalším butterworthovým filtrem prvního řádu, horní propustí na 1 kHz. Výpočty koeficientů Butterworthových filtrů byly provedeny funkcí `butter`.

K výpočtu hodnot gammatónových filtrů pro oba modely sluchové cesty byl použit postup definovaný v [10] včetně přiloženého zdrojového kódu včetně výpočtů kmitočtů odpovídajících 1 ERB pro daná pásma. Výsledné IIR filtry ovšem vykazovaly nestabilitu pro malé poměry středního kmitočtu vůči vzorkovacímu. Tento problém byl zmírněn skrze decimaci na poloviční kmitočet před filtrací a zpětné interpolaci za pomoci funkcí `decimate` a `interp` s využitím implicitních filtrů.

V případě DRNL banky byly dolní propusti realizovány rovněž pomocí Butterworthových filtrů získaných funkcí `butter`, lineární zisk byl proveden prostým násobením a komprese pomocí podmíněného indexování dle tohoto vzoru v kódu MATLAB (s mezí v proměnné `threshold` a poměrem `alpha`):

```
x(x>threshold)=threshold+alpha.*(x(x>threshold)-threshold);
x(x<-threshold)=-threshold+alpha.*(x(x<-threshold)+threshold);
```

Pro plynulé usměrnění byla využita vlastnost interpretu MATLAB, že výstupem podmíněných výrazů je 1 pro pravdu a 0 pro nepravdu, a výstup podmínky byl využit jako násobící koeficient:

```
y=x.*(x>0);
```

Koeficienty pro 1 kHz dolní propust byly spočteny následujícími vztahy dle [11]:

$$K = \tan \frac{\pi f_c}{f_s} \quad (9)$$

$$b_0 = b_1 = \frac{K}{K + 1} \quad (10)$$

$$a_1 = \frac{K - 1}{K + 1} \quad (11)$$

Zisk a umocnění pro CASP byly provedeny základními operátory. Aplikace minimální hodnoty byla realizována funkcí `max`.

Původní podoba implementace zahrnovala v kódu MATLAB přímo implementovanou adaptivní kaskádu, pro které byly koeficienty filtrů z časových konstant získány těmito vztahy z [1]:

$$b_0 = \frac{dt}{\tau + dt}; \quad dt = \frac{1}{f_s} \quad (12)$$

$$a_1 = 1 - b_0 \quad (13)$$

kde f_s je vzorkovací kmitočet a τ časová konstanta. Přímá realizace ovšem představovala z celého modelu největší výpočetní zátěž (viz tabulka 1), tudíž byla z [7] převzata implementace využívající kompilované mex funkce.

První filtr modulační banky byl sestaven jako Butterworthův filtr druhého řádu s mezním kmitočtem 2.5 Hz stejným způsobem, jako u předcházejících filtrů. Koeficienty dalších filtrů s definovanými středními kmitočty f_c a vzorkovacím kmitočtem f_s byly spočteny dle [4] těmito vztahy:

$$w_0 = \frac{2\pi f_c}{f_s} \quad (14)$$

$$e_0 = \exp\left(-\frac{w_0}{Q}/2\right) \quad (15)$$

$$b_0 = 1 - e_0 \quad (16)$$

$$a_1 = -e_0 \cdot e^{jw_0} \quad (17)$$

Po filtraci jsou signály decimovány funkcí `decimate` na kmitočty $f'_s = \frac{f_s}{n}$, kde n je dolů zao-krouhlený podíl originálního vzorkovacího kmitočtu a desetinásobku středního, v případě dolní propusti mezního kmitočtu filtru modulační banky (s pro každé pásmo banky jinou hodnotou n). Výsledný filtrovaný signál při implicitní filtraci Čebyševovým filtrem roste nad všechny meze, tudíž bylo pomocí přepínače `'fir'` použit před decimací FIR filtr.

V tabulce 1 jsou vypsané typické doby výpočtu jednotlivých částí na testovaném počítači (MATLAB r2012a, Intel 3210m, Linux 4.2.2 x86_64), obdržené pomocí funkcí `tic` a `toc`.

Asimilace ze vztahu (1) byla realizována pomocí následujícího jednořádkového vztahu:

```
ya=y+(+(y<x).*((x-y)./2));
```

Součtový operátor před závorkou s logickým výrazem zajišťuje konverzi na datový typ reálné číslo 1 či 0, což umožňuje další aritmetické operace. Vzájemná korelace pro jednotlivá modulační pásma vyjádřená v (2) implementována tímto algoritmem pro celé matice `ytf` a `xtf`:

```
buf1=xtf-mean(xtf(:));
buf2=ytf-mean(ytf(:));
buf3=buf1.*buf2;
buf12=buf1.^2;
buf22=buf2.^2;
rm=sum(buf12(:))/sqrt(sum(buf22(:))*sum(buf3(:)));
```

Realizace hodnoty PSM_t z krátkodobých vzájemných korelací a pohyblivých průměrech (získaných dělením výstupů funkcí `sum` a `length`) byla provedena seřazením pohyblivých průměrů podle hodnot korelací pro dané rámce. Tyto průměry byly dále postupně integrovány, dokud celková hodnota nepřekročila 5% celkové sumy. Index, při kterém byla hodnota překročena, byl použit k identifikaci rámce s danou hodnotou korelace, tedy výsledné hodnoty. Poslední úryvek kódu popisuje přepočítání na hodnotu ODG .

```
ODGa=-0.22;ODGb=0.98;ODGc=-4.13;ODGd=16.4;ODGx0=0.864;
if (psmt<ODGx0)
    odg=max(-4,(ODGa/(psmt-ODGb))+ODGc);
else
    odg=ODGd*psmt - ODGd;
end
```

	PEMO-Q [s]	CASP [s]
Vnější a střední ucho	X	0.03
Bazilární membrána	1.13	2.71
Vlásečnice	0.47	0.47
Adaptivní kaskáda (přímá)	1.93	1.93
Adaptivní kaskáda (kompil.)	0.33	0.29
Modulační banka	6.31	5.8

Tabulka 1: Typické hodnoty jednotlivých částí bloků sluchové cesty (5 s signál)

V rámci implementace nebyla realizována žádná metoda paralelizace. Jediná možnost paralelizace jde uskutečnit u úloh s větším počtem realizací modelů za pomoci rozdělení úlohy do několika skriptů, přičemž každý bude spuštěn zvlášť ve vlastní instanci interpretu MATLAB v jeden okamžik.

4 Závěr

V modelech sluchové cesty je počínaje blokem bazilární membrány výpočet rozdělen na 35 nezávislých pásem (když zanedbáme počty možné dělat maticově přes všechny pásma najednou, např. usměrnění) a v případě modulační banky i osmkrát více. Model má tedy poměrně rozsáhlý potenciál pro paralelizaci výpočtu. Podobně široké možnosti se vyskytují i v případě vyhodnocování vnitřních reprezentací, tam je však v porovnání s modelem sluchové cesty výpočet poměrně rychlý skrze výrazně menší objem dat získaný decimací a výpočet všech náročných částí přes maticové operace.

Co se objemu dat týče nejnáročnější částí je filtrace a následná decimace v části modulační banky. Je možné, že by tato část, podobně jako realizovaná adaptivní kaskáda, mohla též být vhodná k realizaci mimo MATLAB, např. pomocí kompilované mex funkce napsané v jazyce C. V případě nevyužití či nepřítomnosti Parallel Computing Toolbox by to též mohlo umožnit paralelizaci na úrovni C kódu či kompilátoru, např. pomocí rozhraní OpenMP.

Reference

- [1] Ken Chapman. Digitally removing a DC offset: DSP without mathematics. *Xilinx white paper*, 279:134, 2008.
- [2] Brian R Glasberg and Brian CJ Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1):103–138, 1990.
- [3] Niklas Harlander, Rainer Huber, and Stephan D. Ewert. Sound quality assessment using auditory models. *J. Audio Eng. Soc.*, 62(5):324–336, 2014.
- [4] E. Hogenauer. An economical class of digital filters for decimation and interpolation. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(2):155–162, Apr 1981.
- [5] Rainer Huber and Birger Kollmeier. PEMO-Q - a new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio, Speech & Language Processing*, 14(6):1902–1911, 2006.
- [6] ML Jepsen, Stephan D. Ewert, and Torsten Dau. A computational model of human auditory signal processing and perception. *Journal of the Acoustical Society of America*, 124:422–438, 2008.
- [7] Soendergaard. Peter L. Computational auditory signal processing, 2013.
- [8] Enrique A Lopez-Poveda and Ray Meddis. A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*, 110(6):3107–3118, 2001.
- [9] Ray Meddis. *Matlab Auditory Periphery (MAP), Model technical description*. Essex, 2011.
- [10] Malcolm Slaney et al. An efficient implementation of the patterson-holdsworth auditory filter bank. *Apple Computer, Perception Group, Tech. Rep.*, 35:8, 1993.
- [11] Udo Zoelzer, editor. *DAFX: Digital Audio Effects*. John Wiley & Sons, Inc., New York, NY, USA, 2002.

Martin Zalabák

Katedra radioelektroniky, ČVUT FEL v Praze, Technická 2, 166 27 Česká Republika

Podpořeno grantem Studentské grantové soutěže ČVUT č. SGS14/204/OHK3/3T/13.