

MASKOVÁNÍ AKUSTICKÝCH SIGNÁLŮ

Lukáš Pelant

oddělení multimediální techniky, katedra radioelektroniky FEL ČVUT

Abstrakt

Pro potřeby maskování zvuku v oddělení multimediální techniky na katedře radioelektroniky FEL ČVUT byl vyvinut měřicí program v prostředí Matlab. Program umožňuje měřit maskování s harmonickými signály a úzkopásmovými šумы, dále měřit práh slyšení a ověřovat funkčnost současně dostupných maskovacích filtrů. Novým přínosem programu je možnost měření s prostorovými signály. Následující příspěvek se zabývá postupem implementace těchto funkcí.

1 Úvod

Maskování zvuku nabývá v posledních letech graduálně na významu. Právě na něm jsou založeny všechny metody ztrátové komprese zvuku, jako například MP3 a ATRAC. Oblast spotřební elektroniky je dnes na výsledcích výzkumu komprese zvuku, potažmo maskování, závislá, a naopak, můžeme říci, že k rozvoji zvukových kompresí přispívá největším dílem. Dále se maskování využívá v telekomunikacích a přináší nové možnosti archivace.

Maskovací děje patří mezi psychoakustické jevy ve sluchovém aparátu, které způsobují, že za přítomnosti společného výskytu určitého zvuku nebo jeho složky posluchač nedokáže vnímat jiný zvuk nebo jeho složku. z maskování zvuku vycházejí všechny algoritmy ztrátové komprese. Druh i účinek maskovacího děje je závislý na mnoha vlastnostech podnětu nebo vlastnostech, které se vyskytují mezi podněty navzájem. Mezi tyto vlastnosti patří například úroveň maskovacího zvuku, jeho frekvence a charakter (spektrální složení).

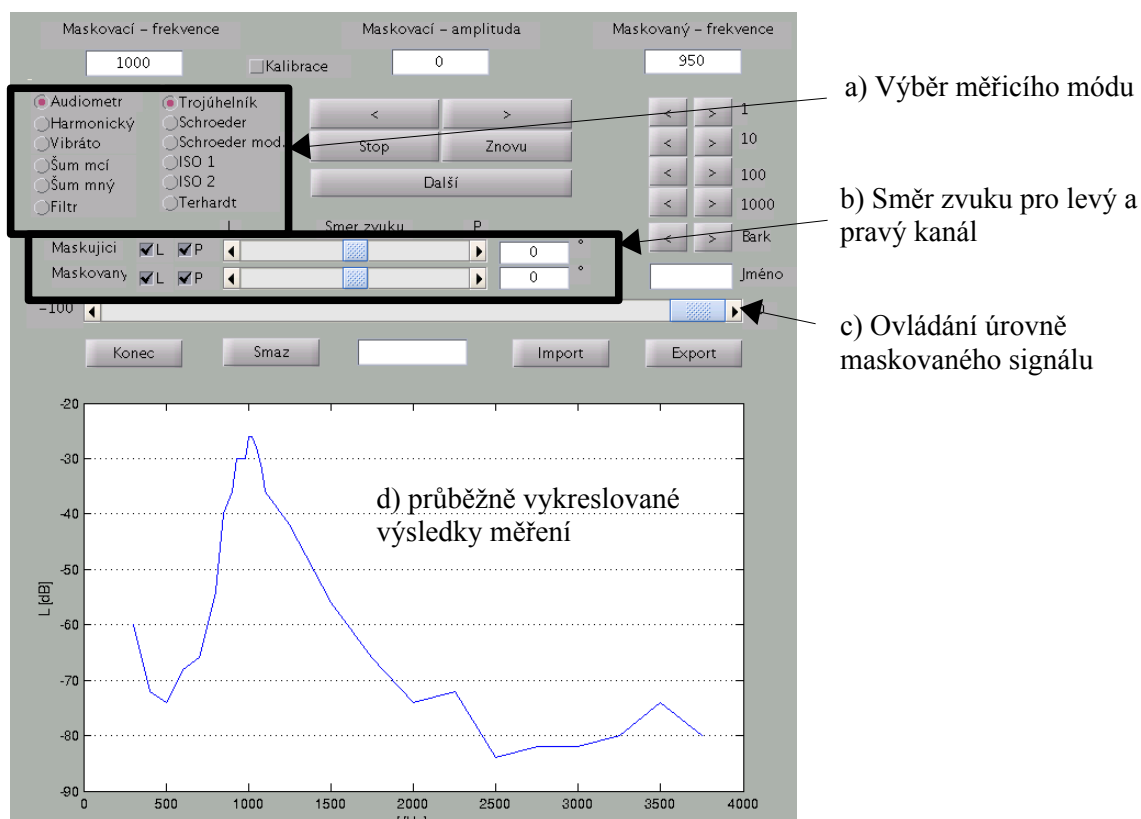
Ačkoliv je jev maskování zdám již poměrně dlouhou dobu, výzkum prostorových aspektů maskování začal teprve nedávno. Pokud maskovací signál přichází z jiného směru než maskovaný, rozdíl v úrovni maskovacích křivek může dosahovat až 20dB. Tento jev je v anglické literatuře označován jako Binaural Masking Level Difference (BMLD), tedy binaurální maskování, popřípadě prostorové maskování. Rozdíl spočívá v tom, že zatímco binaurální maskování bere v úvahu pouze zvuky ležící v rovině hlavy, prostorové maskování zohledňuje všechny tři dimenze. Většina dosud provedených experimentů se tímto binaurálním ani prostorovým maskováním nezabývala, popřípadě se zaměřila pouze na maskování šumem, zpravidla širokopásmovým.

Pro měření maskování byl vyvinut měřicí program, který generuje měřené signály, poskytuje grafické rozhraní pro ovládání měření a zobrazuje naměřené výsledky. Program umožňuje jak základní měření maskování, tak měření prostorového maskování. Pro účely prostorového maskování byla využita metoda založená na HRTF, díky čemuž je možné provádět měření se sluchátky. Pro implementaci programu bylo rozhodnuto využít Matlab pro jeho nezávislost na platformě, relativně jednoduchý jazyk a především šířku jeho nasazení na FEL, kde je v podstatě standardním programovacím jazykem, což umožní případnou spolupráci s kolegy či jejich navázání na výsledky této práce.

2 Průběh měření

Program je určen k měření maskování zvukových signálů. Před vlastním měření měřicí technik nejprve vybere měřicí mód, tedy charakter maskovaného a maskujícího signálu. Poté provede kalibraci měřicího řetězce, tedy nastavení zesilovače tak, aby při maximálním signálu dosahoval tlak pod mušlemi sluchátek požadované referenční úrovně, zpravidla 80 dB. Následně zvolí úroveň maskovacího signálu relativně k referenční úrovni. Posledním krokem je nastavení směrů, ze kterých bude maskovací a maskovaný signál přehráván.

Měření samotné spočívá ve zvolení aktuálně měřené frekvence maskovaného signálu a opakovaném přehrávání obou zvuků, tedy maskovaného i maskujícího, s měnící se úrovní



Obr.1: Grafické uživatelské rozhraní vytvořeného měřicího programu

maskovaného. V okamžiku, kdy právě přestane (popřípadě začne, podle zvolené metodiky měření) být maskovaný signál slyšet, je daná frekvence považována za proměřenou, výsledná hodnota se uloží do paměti a vykreslí do grafu a postup se opakuje pro další frekvence, dokud není proměřeno celé požadované frekvenční pásmo. Po změření celé charakteristiky se naměřená data exportují tlačítkem *Export* do textového formátu, který se dále zpracovává v tabulkovém kalkulátoru nebo specializovaném nástroji pro vykreslování grafů. Pro rychlé prohlédnutí již dříve naměřených dat nebo pro pokračování v nedokončeném měření po krátkém odpočinku měřeného subjektu (psychoakustická měření by neměla trvat déle než 20 minut) je možné data zpětně importovat z uloženého souboru.

3 Měřicí módy

Míra maskování je silně závislá na mnoha proměnných, jako je úroveň maskovacího zvuku, jeho frekvence a v neposlední řadě i jeho charakter, tedy spektrální složení. Maskovací křivky, tedy průběhy hranice slyšitelnosti maskovaného signálu při měnící se frekvenci maskovaného signálu, zatímco frekvence i amplituda maskujícího signálu zůstávají stále, budou mít nejen odlišně vysoké maximum v závislosti na charakteru obou zvuků, ale i lišit se bude i tvar celého průběhu maskovací křivky.

Program umožňuje měřit v několika módech. Jednotlivé módy se od sebe liší charakterem maskovacího a maskovaného zvuku. Jedná se o kombinaci harmonických signálů a různých druhů šumů.

3.1 Měření prahu slyšení

Při tomto druhu měření není přehráván maskovací zvuk. Postup měření je shodný s ostatními módy. Mód slouží k úvodnímu proměření testovaného subjektu. Bylo by totiž nevhodné zahrnout do souboru naměřených hodnot křivky ovlivněné sluchovou vadou.

3.2 Měření maskování harmonického signálu harmonickým signálem

Při tomto druhu měření jsou oba signály, tedy maskovací $m(t)$ a maskovaný $n(t)$, harmonické. Pro harmonický signál $x(t)$ obecně platí

$$x(t) = A \cdot \sin(2\pi f \cdot t)$$

kde A je amplituda signálu, f je jeho frekvence a t je čas.

Pro maskovací signál $m(t)$ s úrovní L_M a frekvencí f_M a maskovací signál $n(t)$ s úrovní L_{mny} a frekvencí f_{mny} můžeme tedy psát

$$m(t) = 10^{\left(\frac{L_M}{20}\right)} \cdot \sin(2 f_M t)$$

$$n(t) = 10^{\left(\frac{L_{mny}}{20}\right)} \cdot \sin(2 f_{mny} t)$$

$$s(t) = m(t) + n(t)$$

kde $s(t)$ je výsledný signál vzniklý součtem obou dílčích signálů.

Tato varianta nepředstavuje při implementaci větší potíže. Jediné, nač je třeba pomyslet, aby nedošlo k přebuzení převodníku, je fakt, že maximální amplituda výsledného signálu může být součet maximální amplitudy každého signálu, tedy za předpokladu rovnosti obou signálů (nejnepříznivější případ) je to dvojnásobek amplitudy maskujícího signálu. K tomu se však ještě dále přičítá další případné zesílení vlivem simulace prostorovosti, viz dále.

3.3 Kombinace harmonického signálu a úzkopásmového šumu

Maskování harmonického signálu harmonickým signálem je velmi specifický případ, který v reálném prostředí nastává velice zřídka. Pokud mají být výsledky použity v praxi, je třeba použít k měření signály přirozenější. Proto byla implementována funkce měření úzkopásmovým šumem.

Úzkopásmový šum je definován jako šum s šířkou pásma maximálně 1 bark. Na tvar maskovací křivky u takového signálu by šířka pásma šumu neměla mít vliv.

Při tvorbě úzkopásmového šumu je nutné nejprve vypočítat požadovanou šířku pásma, tedy přepočítat danou frekvenci na barky.

Barková stupnice je dělení frekvence podle vzdáleností na basilární membráně. Každý bark odpovídá stejně dlouhému úseku na této membráně, zároveň má každý bark právě kritickou šířku pásma. Proto jsou vlastnosti sluchu spojené s maskováním při použití barkové stupnice prakticky nezávislé na frekvenci.

Hranice a střední frekvence barků jsou standardizované, tabulované. Je třeba si však uvědomit, že tyto frekvence jsou pouze definované, dané konvencí. Skutečné jevy vázané na barková pásma vznikají plynule kolem budícího signálu, bez vazby na toto konkrétní diskrétní rozdělení pásem. Proto je v programu použit pro přepočet frekvence f na šířku barku df vzorec podle Zwickera, který se podle místa původu označuje také jako Mnichovský vzorec [1]:

$$dz = 25 + 75 \cdot \left(1 + 1,4 \cdot \left(\frac{f}{1000}\right)^{2 \cdot 0,69}\right)$$

Po výpočtu šířky pásma můžeme vytvořit úzkopásmový šum filtrací šumu širokopásmového, například funkcí fir2. Přechodové pásmo přitom volíme tak, aby strmost filtru byla větší než 27 dB na bark, což je teoretická maximální strmost maskovacích křivek. Odstup signálu mezi propustným a nepropustným pásmem dále volíme tak, aby signál v nepropustném pásmu nemohl být slyšet ani mimo oblast maskování. Práh slyšení závisí silně na frekvenci, přičemž maximální citlivost je kolem 1 - 4 kHz, a to několik decibelů pod úrovní 0 dB (přesná hodnota je individuální). Při měření se signály o úrovní maximálně 80 dB tedy stačí odstup kolem 90 dB. Při splnění těchto podmínek

máme jistotu, že vlastní maskování bude ovlivňovat pouze signál nacházející se uvnitř propustného pásu filtru.

Dále je také nutné zohlednit fakt, že energie šumu je úměrná čtverci šířky jeho pásma, a příslušně korigovat amplitudu signálu.

4 Maskovací filtry

Jedním z požadavků na funkcionalitu programu byla možnost ověření funkčnosti současně dostupných maskovacích filtrů. Toho je možné dosáhnout současným přehráváním šumu tvarovaného podle měřeného filtru a harmonického signálu. Přitom se sleduje, kdy zanikne harmonický signál nebo kdy začne být slyšet šum, v závislosti na tom, zda se harmonický signál použije jako maskovaný nebo maskovací zvuk.

Pro toto měření, vzhledem ke komplexnímu tvaru filtrů a nutnosti najít univerzální řešení, byl zvolen následující postup výroby tvarovaného šumu: Nejprve je širokopásmový šum převeden na své spektrum. Poté je jeho amplituda upravena příslušnou tvarovací funkcí (maskovacím filtrem) a signál je následně převeden zpět do časové oblasti. Přitom je nezbytné vzít v úvahu zrcadlení spektra – amplitudové spektrum reálného signálu je sudé, proto je nutné i tvarovací funkci zrcadlit podle nulového kmitočtu (respektive, v případě matlabu, podle vzorku odpovídajícího $fs/2$). Fázové spektrum zůstává nezměněno.

V tomto případě byly pro tvorbu maskovaného signálu použity následující vzorce:

$$w(t) = \begin{cases} f_{okno}(t) & \text{pro } 0 \leq t < T \\ f_{okno}(2T - t) & \text{pro } T < t < 2T \end{cases}$$

$$N_{filt}(t) = \text{ifft} \{ \text{fft} [N(t)] \cdot w(t) \}$$

$$n(t) = 10^{\frac{(L_M + konst)}{20}} \cdot N_{filt}(t)$$

kde $w(t)$ je příslušná tvarovací funkce vyjádřená ve spektru. Funkce $f_{okno}(t)$ jsou okénkovací funkce mající tvar maskovacích filtrů, fft je Fourierova transformace a ifft pak zpětná Fourierova transformace a $N(t)$ je širokopásmový šum vytvořený funkcí `randn(délka, 1)`.

5 Simulace prostorovosti sluchátky

Lidské slyšení používá pro rozlišení směru zvuku dvě metody, které působí současně. První je založena na zpracování časového posunu mezi levým a pravým uchem. Má se zato, že časový rozdíl mezi uchy se v mozku spočítá na základě korelace obou signálů nebo obdobného principu. Druhá metoda lateralizace je založena na rozdílu úrovně. Úroveň akustického tlaku zvuku je dána jeho vzdáleností od zdroje. Je tedy zjevné, že ucho umístěné dále od zdroje bude snímat zvuk o menší úrovni než ucho přivrácené. Dále do hry vstupuje hlava, která působí jako překážka v cestě zvuku, odvrácené ucho se nachází v jejím akustickém stínu

Ani jedna z popsaných metod neposkytuje úplné řešení problému modelování prostorovosti. Obě metody jsou frekvenčně omezené. Jelikož při studiu prostorového maskování je třeba operovat s širokou škálou frekvencí, ani jedna z nich neposkytuje úplné a dobře použitelné řešení. Je nezbytně nutné obě metody použít současně.

V poslední době se jako vhodné řešení prosazuje metoda založená na impulzních odezvách lidské hlavy (HRIR nebo HRTF). Vychází z úvahy, že pokud předem změříme impulzní odezvu lidské hlavy pro zvuky dopadající z různých směrů, můžeme pak při modelování prostorového zvuku použít příslušnou impulzní odezvu pro daný směr a pomocí konvoluce s původním zvukem dostat požadovaný rozdíl úrovně i časový posun. Pro signály procházející soustavou obecně platí, že

$$y(t) = h(t) * x(t)$$

kde $h(t)$ je impulzová odezva, $x(t)$ vstupní signál a $y(t)$ výstupní signál. Tohoto jednoduchého vztahu bylo přímo využito pro simulaci prostorovosti sluchátky.

Zadání problému bylo vytvořit takový signál, který po přehrání na sluchátkách bude znít, jako by vycházel ze zadaného směru. Pro výpočet byla využity data volně dostupná na [2]. Jedná se o tabulované impulzní odezvy (HRIR) změřené v horizontální rovině s odstupem 5° . Při použití v programu je nejprve nutné data načíst

```
load large_pinna_final.mat
```

Poté se vstupní směr zvuku `uhel` vybraný pomocí ovládacích prvků (viz obr. 1, bod b) zaokrouhlí na intervaly po 5° .

```
index=round(mod(uhel+360,360)/5)+1;
```

Výsledná hodnota `index` se použije jako index pro vyčítání příslušné impulzové odezvy z tabulovaných hodnot HRIR pro levé (pole `left`) a pravé (pole `right`) ucho.

Nakonec se vypočítá konvoluce impulzové odezvy a zdrojového zvuku $s(t)$ v proměnné `s`, který se snažíme umístit do prostoru, a to jak pro levé, tak pro pravé ucho. Výsledkem je dvourozměrné pole `vystup` s dvoukanalovým výsledným prostorovým zvukem

```
vystup(:,:)= [conv(s,left(:,index));  
              conv(s,right(:,index))]';
```

6 Přehrávání zvuku

Původně byl program koncipován jako multiplatformní. Předpokládalo se, že Matlab, jakožto skriptovací jazyk, bude na operačním systému zcela nezávislý. Jak se však ukázalo, implementace výstupu zvuku v jádru Matlabu se mezi operačními systémy liší. Na MS Windows byl použit vzorkovací kmitočet $fs=44100$ Hz a k přehrání zvuku `vystup` se užívá funkce

```
wavplay (vystup.*B,fs,'sync');
```

Na Linuxu se bylo omezit na vzorkovací kmitočet $fs=8192$, a jelikož funkce `wavplay` není dostupná, k přehrávání se užívá funkce

```
sound (vystup,fs);
```

Tato funkce však neumožňuje stereo výstup, tedy není možné pracovat s prostorovostí. K rozlišení mezi systémy je možné užít funkce `isunix`, díky čemuž je možné na Linuxu provádět alespoň vývoj a základní testování. Měření samotné se pak musí provést na počítači s OS MS Windows.

7 Závěr

V prostředí MATLAB byl vytvořen program pro psychoakustická měření maskování zvuku s přihlédnutím k možnosti měřit prostorovou závislost maskování. Program umožňuje měření za použití harmonických signálů i kombinace harmonického signálu a úzkopásmového šumu. Dále je možné ověřovat funkčnost současně dostupných maskovacích filtrů a změřit práh slyšení pokusného subjektu. Ačkoliv implementace přinesla některé problémy, ukázalo se použití MATLABu jako výhodné.

8 Poděkování

Práce byla podpořena grantem GAČR č. 102/05/2054 *Kvalitativní aspekty zpracování audiovizuální informace v multimediálních systémech*

Reference

- [1] Smith J. O. III, Abel J. S., *The Bark and ERB Bilinear Transforms*. IEEE Transactions on Speech and Audio Processing, prosinec 1999.
- [2] *The CIPIC HRTF Database* [online]. Center for Image Processing and Integrated Computing University of California, 2001, URL: <http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm>

Lukáš Pelant
pelanl1@fel.cvut.cz