



TECHNICKÁ UNIVERZITA V LIBERCI  
Fakulta mechatroniky, informatiky  
a mezioborových studií ■

# REDUKCE HLUKU PRO MOBILNÍ TELEFON SE DVĚMA MIKROFONY

## Diplomová práce

*Studijní program:* N2612 – Elektrotechnika a informatika

*Studijní obor:* 1802T007 – Informační technologie

*Autor práce:* **Bc. David Botka**

*Vedoucí práce:* doc. Ing. Zbyněk Koldovský, Ph.D.



## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: Bc. David Botka  
Osobní číslo: M11000229  
Studijní program: N2612 Elektrotechnika a informatika  
Studijní obor: Informační technologie  
Název tématu: Redukce hluku pro mobilní telefon se dvěma mikrofony  
Zadávací katedra: Ústav informačních technologií a elektroniky


### Z á s a d y p r o v y p r a c o v á n í :

1. Nastudujte způsoby návrhu filtru se dvěma vstupy pro potlačení cílového zdroje (target cancellation filter - CF).
2. Nastudujte vlastnosti CF a možnosti jejich přepínání pro různé polohy zdroje.
3. Vytvořte banku CF pro soustavu umělé hlavy a umělé ruky držící telefon a pro skutečnou osobu (sebe), která pokrývá nejpravděpodobnější pozice.
4. Pomocí naměřené banky CF realizujte algoritmus pro odstraňování okolního hluku z hlasu telefonující osoby.

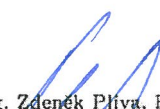
Rozsah grafických prací: Dle potřeby dokumentace  
Rozsah pracovní zprávy: cca 40 - 50 stran  
Forma zpracování diplomové práce: tištěná/elektronická  
Seznam odborné literatury:

- [1] B. Porat, "A Course in Digital Signal Processing", John Wiley & Sons, 1997.
- [2] Jeub, M.; Herglotz, C.; Nelke, C.; Beaugeant, C.; Vary, P.; , "Noise reduction for dual-microphone mobile phones exploiting power level differences," Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on , pp.1693-1696, 25-30 March 2012.

Vedoucí diplomové práce: doc. Ing. Zbyněk Koldovský, Ph.D.  
Ústav informačních technologií a elektroniky  
Konzultant diplomové práce: Ing. Jiří Málek, Ph.D.  
Ústav informačních technologií a elektroniky  
Datum zadání diplomové práce: 12. září 2013  
Termín odevzdání diplomové práce: 16. května 2014

  
prof. Ing. Václav Kopecký, CSc.  
děkan

L.S.

  
prof. Ing. Zdeněk Pliva, Ph.D.  
vedoucí ústavu

V Liberci dne 12. září 2013

## Prohlášení

Byl jsem seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé diplomové práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum: 15. 5. 2014

Podpis:

## **Poděkování**

Chtěl bych poděkovat všem, kteří mi pomohli s vypracováním diplomové práce. Děkuji především vedoucímu mé práce doc. Ing. Zbyňku Koldovskému, Ph.D. za cenné rady, konzultace a pomoc při tvorbě této práce.

## Abstrakt

Práce se zabývá metodou redukce hluku se zaměřením na mobilní telefony. Navržený systém používá k odhadu okolního hluku banku předem připravených target cancellation filtrů (CF), které při dodržení určitých podmínek potlačí cílový signál (hlas telefonující osoby). Vznikne tak signál obsahující především okolní hluk, který následně potlačujeme v původní nahrávce. Pro samotné potlačení hluku uvádíme některé již existující filtry. Systém může zpracovávat více vstupních kanálů, s ohledem na současné mobilní telefony klademe důraz na dvoukanálové zpracování. Popisujeme návrh CF se dvěma vstupními kanály a navrhujeme postupy pro vytváření bank CF. V experimentech uvažujeme dvě různá rozmístění mikrofonů. Zpracováváme nahrávky pořízené na umělé hlavě i nahrávky skutečných osob. Provedené experimenty ukázaly, že navržený systém má nízkou výpočetní náročnost a poskytuje možnosti dosažení kvalitních výsledků i v náročných situacích obsahujících hluk velmi výrazně. K navrženému systému dále popisujeme demonstrační aplikaci, kterou jsme vyvinuli v jazyce Java.

**Klíčová slova:** potlačení hluku, zlepšování řečového záznamu, mobilní telefon, target cancellation filtr, dvoukanálové zpracování

## Abstract

This thesis discusses a method of noise cancellation focused on mobile phones. For noise estimation the proposed system uses a bank of pre-measured target cancellation filters (CF), which under specific conditions cancels target signal (speaker's voice). Primarily, obtained signal contains background noise, which is then cancelled in original recording. To perform of cancellation we describe some existing filters. The system is able to process more input channels; with respect to current mobile phones we place emphasis on dual-channel processing. We describe proposal of CF with two input channels and we propose methods for CF banks creating. In experiments we consider two different microphone arrangements. We process recordings acquired on artificial head and also recordings of real persons. Performed experiments have shown, that system has low computational complexity and provides possibilities to reach high quality results also in difficult scenarios with very loud noise. To proposed system we also describe a demo application, which we have developed in the Java programming language.

**Key words:** noise cancellation, speech enhancement, mobile phone, target cancellation filter, dual-channel processing

# Obsah

Prohlášení . . . . .	3
Poděkování . . . . .	4
Abstrakt . . . . .	5
Abstract . . . . .	6
Obsah . . . . .	7
Seznam obrázků . . . . .	9
Seznam tabulek . . . . .	10
Seznam zkratek . . . . .	11
<b>1 Úvod</b>	<b>12</b>
<b>2 Redukce hluku</b>	<b>15</b>
2.1 Potlačení hluku v časové oblasti . . . . .	16
2.1.1 Least Mean Square (LMS) filtr . . . . .	17
2.1.2 Wienerův filtr . . . . .	18
2.1.3 Adaptivní LMS filtr . . . . .	19
2.2 Potlačení hluku ve frekvenční oblasti . . . . .	21
2.2.1 Wienerův filtr ve frekvenční oblasti . . . . .	22
2.2.2 Adaptivní LMS ve frekvenční oblasti . . . . .	23
2.2.3 Short-term MMSE . . . . .	24
2.2.4 Short-term Log-MMSE . . . . .	25
<b>3 Odhad hluku</b>	<b>26</b>
3.1 Target cancellation filtr (CF) . . . . .	26
3.1.1 CF se dvěma vstupy . . . . .	27
3.1.2 Návrh CF se dvěma vstupy . . . . .	29



<b>4</b>	<b>Systém redukce hluku s bankou CF</b>	<b>32</b>
4.1	Schéma systému . . . . .	32
4.2	Vytvoření banky CF . . . . .	34
4.2.1	Banka CF jako síť definovaných pozic . . . . .	35
4.2.2	Algoritmus učení banky CF . . . . .	36
<b>5</b>	<b>Experimenty</b>	<b>40</b>
5.1	Audio zařízení . . . . .	41
5.2	Kritéria vyhodnocení . . . . .	43
5.3	Nahrávky s umělou hlavou . . . . .	44
5.3.1	Potlačení cílového signálu . . . . .	45
5.3.2	Potlačení hluku . . . . .	56
5.4	Nahrávky skutečných osob . . . . .	60
5.4.1	Potlačení cílového signálu . . . . .	60
5.4.2	Potlačení hluku . . . . .	62
<b>6</b>	<b>Demonstrační aplikace</b>	<b>64</b>
6.1	Struktura aplikace . . . . .	65
6.2	Formát souboru s bankou CF . . . . .	67
6.3	Grafické uživatelské rozhraní . . . . .	68
6.3.1	Nastavení . . . . .	68
6.3.2	Režim online . . . . .	70
6.3.3	Režim offline . . . . .	73
6.3.4	Režim nahrávání . . . . .	74
6.3.5	Učení banky CF . . . . .	75
<b>7</b>	<b>Závěr</b>	<b>77</b>
<b>A</b>	<b>Obsah příloženého CD</b>	<b>80</b>
	<b>Literatura</b>	<b>81</b>

## Seznam obrázků

4.1	Blokové schéma navrženého systému redukce hluku . . . . .	35
4.2	Oblast nejpravděpodobnějších pozic mobilního telefonu vůči mluvčímu	37
5.1	Model mobilního telefonu . . . . .	44
5.2	Zvuková karta (vlevo) a předzesilovač (vpravo) . . . . .	45
5.3	Experimentální soustava . . . . .	47
5.4	Schéma experimentu potlačení cílového signálu . . . . .	48
5.5	CF pro pozici $K$ . . . . .	49
5.6	Graf potlačení cílového signálu ve fixní pozici bankou s jedním CF . .	50
5.7	Průběh potlačení cílového signálu bankami $B_1, \dots, B_5$ . . . . .	52
5.8	Pohyb modelu telefonu pro vytvoření bank CF algoritmem učení . . .	54
5.9	Oblasti vzniku CF během učení bank $B_6$ . . . . .	55
5.10	Průběh potlačení cílového signálu bankami $B_5, B_6, B_8, B_{10}$ a $B_{12}$ . . .	56
5.11	Schéma experimentu potlačení hluku . . . . .	59
5.12	Vyhodnocení experimentu potlačení hluku na experimentální soustavě	61
5.13	Vyhodnocení experimentu potlačení hluku se skutečnými osobami . .	66
6.1	Schéma struktury aplikace . . . . .	68
6.2	Balíčky a třídy v aplikaci . . . . .	69
6.3	Struktura hlavní nabídky aplikace . . . . .	71
6.4	Screenshot nastavení aplikace . . . . .	71
6.5	Screenshot aplikace v režimu online . . . . .	73
6.6	Výpis informací o bance CF . . . . .	74
6.7	Screenshot okna snadného ladění parametrů . . . . .	75
6.8	Screenshot aplikace v režimu offline . . . . .	76
6.9	Screenshot aplikace v režimu nahrávání . . . . .	77
6.10	Screenshot průvodce učním novou bankou CF . . . . .	78
6.11	Screenshot okna pro výpis informací o postupu učení banky CF . . . .	79

## Seznam tabulek

2.1	Výpočet aproximace short-term MMSE . . . . .	26
5.1	Banky CF (síť pozic) pro experiment potlačení cílového signálu . . . .	49
5.2	Celková potlačení cílového signálu bankami $B_1, \dots, B_5$ . . . . .	51
5.3	Banky CF (algoritmus učení) pro experiment potlačení cílového signálu	52
5.4	Celková potlačení cílového signálu bankami $B_5, \dots, B_{13}$ . . . . .	56
5.5	Banky CF v experimentu potlačení hluku na experimentální soustavě	58
5.6	Banky CF skutečných osob . . . . .	61
5.7	Celková potlačení bankami CF pro skutečné osoby a umělou hlavu . .	62
6.1	Formát binárního souboru s bankou CF . . . . .	68

## Seznam zkratek

<b>ATS</b>	Attenuation of Target Signal, potlačení cílového signálu
<b>CF</b>	target Cancellation Filtr, digitální filtr potlačující cílový signál
<b>FFT</b>	Fast Fourier Transform, rychlá Fourierova transformace
<b>FIR</b>	Finite Impulse Response, konečná impulsní odezva
<b>GUI</b>	Graphical User Interface, grafické uživatelské rozhraní
<b>ISTFT</b>	Inverse Short-Time Fourier Transform, inverzní krátkodobá Fourierova transformace
<b>LMS</b>	Least Mean Square, digitální filtr pro potlačení hluku
<b>OPS</b>	Overall Perceptual Score, percepční kritérium
<b>PP</b>	Přední Přední, rozmístění mikrofónů
<b>PZ</b>	Přední Zadní, rozmístění mikrofónů
<b>pSNR</b>	počáteční Signal to Noise Ratio, počáteční poměr signálu a šumu
<b>SDR</b>	Signal to Distortion Ratio, poměr signálu a zkreslení
<b>SNR</b>	Signal to Noise Ratio, poměr signálu a šumu
<b>STFT</b>	Short-Time Fourier Transform, krátkodobá Fourierova transformace
<b>VAD</b>	Voice Activity Detector, detektor řečové aktivity

# 1 Úvod

Mobilní telefony jsou již řadu let naprosto běžnou součástí života lidí po celém světě. S obrovským rozvojem mobilních telefonů se rozvíjejí i metody pro zvýšení kvality hovoru. Kromě vylepšování hardwarových komponent pro zpracování zvuku, jako jsou například mikrofony a součásti pro digitalizaci zvuku, se rozvíjejí i softwarové metody, jejichž cílem je co nejvíce potlačit okolní hluk a zároveň co nejméně zdeformovat hlas telefonující osoby. Na metody pro potlačování okolního hluku se zaměřením na mobilní telefony jsou ale kladeny specifické požadavky, především nízká výpočetní náročnost.

Donedávna byly mobilní telefony vybaveny zpravidla pouze jedním mikrofonem, tudíž pro potlačování okolního hluku bylo možné použít pouze jednokanálové metody zpracování. V současné době se ale na trhu běžně objevují telefony osazené více mikrofony. Vícekanálové zpracování signálů otevírá v určitých oblastech zcela nové možnosti, jak dosáhnout lepších výsledků než při jednokanálovém zpracování. Jednou z těchto oblastí je i potlačování okolního hluku v zarušených signálech. Již dva vstupní kanály poskytují široké možnosti dosažení kvalitních výsledků. Z toho důvodu je mnoho dnešních mobilních telefonů vybaveno právě dvěma mikrofony.

V případě telefonních hovorů představují okolní hluk všechny zaznamenané složky kromě hlasu telefonující osoby. Postup redukce hluku obecně sestává ze dvou kroků. Nejprve je třeba v pořízené nahrávce získat informace o obsaženém hluku a na jejich základě pak hluk potlačit. Oba tyto kroky ale obvykle vyžadují mnoho složitých operací.

Pro potlačování hluku v řečových nahrávkách, někdy označováno jako zlepšování řečového záznamu (*Speech enhancement*), bylo již navrženo mnoho různých postupů. Problematika řeči a hluku je ale natolik komplikovaná, že pro účinné potlačování

hluku není možné navrhnout zcela univerzální postup, který by bezchybně fungoval pro všechny možné situace. Obvykle je třeba do řešení zahrnout určité skutečnosti vyplývající z charakteristiky daného problému.

Některé metody jsou navrženy obecně pro více vstupních kanálů. S vyšším počtem vstupních kanálů lze sice někdy dosáhnout lepších výsledků, ovšem zvyšuje se tím i výpočetní náročnost zpracování. V případě mobilních telefonů by kromě výpočetní náročnosti mohl být navíc i problém s umístěním vyššího počtu mikrofónů po povrchu telefonu. Potíže by mohly například nastat, pokud by daná metoda vyžadovala určitou minimální vzdálenost sousedních mikrofónů. Také by se zvyšovalo riziko, že uživatel při uchopení telefonu některé mikrofony zakryje. Z různých důvodů se proto u mobilních telefonů klade důraz na dvoukanálové metody zpracování. V případě mobilních telefonů se potlačování hluku může od jiných aplikací značně lišit. Mobilní telefony se dvěma mikrofony mají obvykle primární mikrofón umístěný dole (pro záznam především řeči) a sekundární mikrofón umístěný nahoře někdy i na zadní straně (pro záznam především hluku). Některé systémy pro potlačování hluku tyto rozdíly využívají. Z existujících systémů zde uvádíme jen vybrané, především zaměřené na telefonní hovory.

Na rozdílech v signálech z primárního a sekundárního mikrofónu je založena metoda PLD [1], kterou autoři navrhli na základě měření. Při záznamu hluku naměřili hlasitost zaznamenaných signálů z obou mikrofónů přibližně stejnou, ovšem při záznamu řeči byly rozdíly v hlasitostech alespoň 10 dB. Podle těchto rozdílů pak detekují bloky signálu obsahující především hluk, na jehož základě navrhnou filtr pro potlačení hluku založený na Wienerovu filtru.

Na podobných předpokladech je i založena metoda uvedená v článku [2]. Oproti PLD ale zavádí k odhadu přítomnosti řeči PSNRD (*Posteriori Signal to Noise Ratio Difference*), které autoři mimo jiné popisují v porovnání s PLD jako robustnější a nezávislé na citlivosti mikrofónů. Pro redukci hluku pak používají dvoukanálový MVDR (*Minimum Variance Distortionless response*) filtr [3].

V článku [4] byla představena metoda potlačení hluku pro situaci *handsfree*, kde uživatel nedrží svůj telefon v ruce. Zde tedy nelze předpokládat výrazné rozdíly v signálech z primárního a sekundárního mikrofónu. Systém kombinuje jedno-kanálovou metodu zpracování pro odhad přítomnosti řeči a dvoukanálovou metodu

pro odhad hluku využívající souvislostí mezi řečí a hlukem. Ze získaných informací poté navrhnou výsledný filtr.

Redukci hluku pro telefonní hovory lze provádět i pomocí metod BSS (*Blind Source Separation* – slepá separace zdrojů) založených na ICA (*Independent Component Analysis* – analýza nezávislých komponent). Systém používající takový postup byl navržen v článku [5]. Navržený systém lze používat i s více vstupními kanály. Autoři také prezentovali zlepšení úspěšnosti rozpoznávání v experimentech se dvěma vstupními kanály.

Tato práce pojednává o metodě, která k odhadu okolního hluku používá banku předem připravených tzv. *target cancellation filtrů* (CF). Použije-li se vhodný CF na nahrávku telefonního hovoru, dojde k potlačení hlasu telefonující osoby, přičemž okolní hluk zůstane zachován (některé metody používají podobný přístup, ale část systému potlačující řeč nazývají *blocking matrix*). Pro kvalitní odhad hluku je ale obvykle nezbytné, aby banka obsahovala těchto CF více pro různé pozice telefonu vzhledem k mluvčímu. Během zpracování je pak nutné vybrat z banky takový CF, který poskytne nejvěrnější podobu okolního hluku. Pro samotné potlačení hluku na základě získaného odhadu pak existuje mnoho různých filtrů.

V kapitole 2 popisujeme filtry pro redukci hluku. Návrh CF se dvěma vstupními kanály uvádíme v kapitole 3, které používáme v systému redukce hluku navrženém v kapitole 4, kde dále navrhujeme i způsoby vytvoření bank CF. Provedené experimenty prezentujeme v kapitole 5. Demonstrační aplikaci navrženého systému jsme popsali v kapitole 6 a kapitola 7 obsahuje závěr a souhrn práce.

## 2 Redukce hluku

Reálná akustická nahrávka cílového zdroje (například lidské řeči) většinou obsahuje navíc i okolní hluk [6]. Proces nahrávání zvuku je lineární, tedy označíme-li v časové oblasti cílový signál  $x[n]$  a okolní hluk  $d[n]$ , zaznamenaný signál  $y[n] = x[n] + d[n]$ . Cílem procesu redukce hluku je z pozorovaného signálu  $y[n]$  co nejpřesněji odhadnout signál s cílovým zdrojem  $x[n]$ . Odhad cílového signálu označujeme  $\hat{x}[n]$ .

Pro potlačení hluku se většinou používají filtry, které se navrhují podle jejich výstupu pro konkrétní vstupní signál. Návrh takových filtrů lze provádět například na principu minimalizace kvadratických kritérií. Filtry minimalizující kvadratická kritéria popisujeme v následujících kapitolách.

Mnoho algoritmů pro zpracování audio signálů pracuje se signálem po blocích obvykle stejné délky, které se mohou překrývat. Patří mezi ně i metody pro potlačení hluku. Tyto metody většinou zpracovávají vstupní signál časově-frekvenční nebo časové oblasti. Při blokovém zpracování bývá obvykle filtr potlačující hluk časově proměnný a přepočítává se na každém bloku dat. Některé metody aplikují na aktuálních datech filtr zcela nový, jiné na základě aktuálních dat upravují filtr z předchozích bloků.

Existují i filtry, které zohledňují různá psychoakustická kritéria. Výsledný signál sice pak může být z hlediska určitých kritérií horší, nicméně poslechově se může zdát výsledek kvalitnější. Jsou-li výsledné signály určené především pro poslech, což je i případ telefonních hovorů, je vhodné takovéto filtry používat.

V literatuře se rozlišují pojmy *Noise Suppression* a *Noise Cancellation*. Algoritmy pro *Noise Suppression* používají k odhadu cílového signálu  $\hat{x}[n]$  pouze pozorovaný signál  $y[n]$ , zatímco algoritmy *Noise Cancellation* předpokládají i znalost okolního hluku, tedy signálu  $d[n]$  [6]. Přesná podoba signálu  $d[n]$  je ale známá pouze v experimentálních úlohách, proto je v praxi nutné nahradit  $d[n]$  jeho odhadem.



## 2.1 Potlačení hluku v časové oblasti

Redukci hluku provedeme zpracováním vstupního signálu filtrem  $h$  typu FIR a délky  $M$ , jehož výstup označený  $\hat{x}[n]$  (odhad cílového signálu) je [7]

$$\hat{x}[n] = \sum_{i=0}^{N-1} h[i] \cdot y[n-i], \quad (2.1)$$

vektorovým zápisem

$$\hat{x}[n] = \mathbf{h}^T \mathbf{y}_n, \quad (2.2)$$

kde

$$\mathbf{y}_n = \begin{bmatrix} y[n] \\ y[n-1] \\ \vdots \\ y[n-M] \end{bmatrix} \quad \mathbf{h} = \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[M] \end{bmatrix}. \quad (2.3)$$

Ve výstupu vzniká chyba odhadu cílového signálu, které odpovídá chybový signál

$$e[n] = \hat{x}[n] - x[n]. \quad (2.4)$$

Zavedeme kritérium, které bude funkcí hledaného filtru  $\mathbf{h}$  a rovno kvadrátu chyby v čase  $n$

$$J_n(\mathbf{h}) = e[n]^2. \quad (2.5)$$

Za účelem snížení chyby a tím i co nejpřesnějšího odhadu signálu  $x[n]$  budeme zavedené kritérium minimalizovat. K tomu použijeme gradient  $J_n(\mathbf{h})$ , který je vektorem parciálních derivací podle jednotlivých složek  $\mathbf{h}$ . Vektorově zapsáno

$$\Delta J_n(\mathbf{h}) = -2\mathbf{y}_n x[n] + 2\mathbf{y}_n \mathbf{y}_n^T \mathbf{h}. \quad (2.6)$$

Dále zavedeme značení

$$\begin{aligned}\mathbf{R}_n &= \mathbf{y}_n \mathbf{y}_n^T \\ \mathbf{p}_n &= \mathbf{y}_n x[n]\end{aligned}\tag{2.7}$$

pak gradient

$$\Delta J_n(\mathbf{h}) = -2\mathbf{p}_n + 2\mathbf{R}_n \mathbf{h}.\tag{2.8}$$

### 2.1.1 Least Mean Square (LMS) filtr

Filtrem LMS se minimalizuje průměrný kvadrát chyby v daném časovém úseku  $n = n_1, \dots, n_2$ , tedy průměrná hodnota kritéria  $J_n(\mathbf{h})$  na daném intervalu [7], zapsáno jako

$$J_{n_1, n_2}(\mathbf{h}) = \frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} J_n(\mathbf{h}) = \frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} e[n]^2.\tag{2.9}$$

Je-li uvažován celý úsek dat ( $n_1 = 1$  a  $n_2 = N$ ), pak

$$J_n(\mathbf{h}) = \frac{1}{N} \sum_{n=1}^N e[n]^2.\tag{2.10}$$

Jelikož je derivace lineární operace, gradient  $J(\mathbf{h})$  je roven průměru gradientů a platí

$$\Delta J(\mathbf{h}) = -2\mathbf{p} + 2\mathbf{R}\mathbf{h},\tag{2.11}$$

kde

$$\begin{aligned}\mathbf{R} &= \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^T \\ \mathbf{p} &= \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n d[n].\end{aligned}\tag{2.12}$$

Jak již bylo uvedeno, gradient  $J(\mathbf{h})$  je vektor parciálních derivací. Položíme-li gradient roven nule  $\Delta J(\mathbf{h}) = 0$ , dostaneme minimum kritéria (minimalizace chyby) vzhledem k  $\mathbf{h}$ . Potom je hledaný filtr

$$\mathbf{h} = \mathbf{R}^{-1} \mathbf{p}. \quad (2.13)$$

Matice  $\mathbf{R}$  obsahuje prvky autokovariance  $y$  a prvky  $\mathbf{p}$  představují vzájemnou kovarianci  $y$  a  $x$ . Je zřejmé, že pro výpočet rovnice 2.13 musí existovat inverzní matice  $\mathbf{R}$ . Bude-li například interval  $n = n_1, \dots, n_2$  (interval, na kterém se filtr vypočítává) příliš krátký, nemusí inverzní matice  $\mathbf{R}$  existovat.

### 2.1.2 Wienerův filtr

Definice Wienerova filtru předpokládá, že pozorovaný signál  $y[n]$  a cílový signál  $x[n]$  jsou slabě stacionární [7]. Z toho vyplývá, že slabě stacionární je výstupní signál  $\hat{x}[n]$  i chybový signál  $e[n]$ . Minimalizované kritérium je

$$J(\mathbf{h}) = E\{e[n]^2\}. \quad (2.14)$$

Odvození Wienerova filtru je podobné jako odvození LMS. Výpočty průměrných hodnot se nahradí operátorem střední hodnoty  $E\{\cdot\}$  a hledaný filtr je

$$\mathbf{h} = \mathbf{R}^{-1} \mathbf{p}, \quad (2.15)$$

kde

$$\begin{aligned} \mathbf{R} &= E\{\mathbf{y}_n \mathbf{y}_n^T\} \\ \mathbf{p} &= E\{\mathbf{y}_n d[n]\}. \end{aligned} \quad (2.16)$$

Pokud by se odvozoval Wienerův filtr z konkrétního naměřeného signálu, nahradily by se operátory středních hodnot aritmetickými průměry. Tím se dostane filtr LMS. Filtr LMS je tedy odhadem Wienerova filtru, jsou-li signály  $y$  a  $x$  slabě stacionární. Se zvětšováním délky dostupných dat filtr LMS konverguje k Wienerovu

filtru. Pokud jsou ale signály  $y$  a  $x$  na uvažovaném intervalu nestacionární, Wienerův filtr na tomto intervalu neexistuje. Mnohdy jsou ale signály stacionární alespoň na krátkých intervalech, například řeč je stacionární na intervalech 20–40 ms. Tam tedy Wienerův filtr teoreticky existuje a LMS je jeho odhadem [7].

### 2.1.3 Adaptivní LMS filtr

Signál lze zpracovávat blokově a filtr LMS vždy přepočítávat na aktuálním bloku dat [7]. Nastává však problém prudkých změn filtru, které jsou způsobené malou délkou uvažovaného intervalu  $n = n_1, \dots, n_2$ . Čím je interval kratší, tím jsou navíc změny výraznější. Pokud bude interval příliš krátký, nemusí navíc existovat inverzní matice  $\mathbf{R}$  (vztah 2.13).

Adaptivní LMS je proto odvozen jiným způsobem. Upravuje  $\mathbf{h}$  tak, aby byla minimalizována chyba v aktuálním čase  $n$  daná kritériem  $J_n(\mathbf{h})$ . Zde je  $\mathbf{h}$  časově proměnné, proto zavádíme dolní index pro daný čas  $\mathbf{h}_n$ . Jak již bylo popsáno, LMS provádí minimalizaci položením gradientu rovno nule. V tomto případě by ale neexistovala inverze k matici  $\mathbf{R}$ , proto se zde používá jiný postup – metoda největšího spádu

$$\mathbf{h}_{n+1} \leftarrow \mathbf{h}_n - \mu \Delta J_n(\mathbf{h}), \quad (2.17)$$

kde  $\mu$  je délka kroku. Po dosazení je krok pro adaptaci

$$\mathbf{h}_{n+1} \leftarrow \mathbf{h}_n - \mu \mathbf{y}_n e[n]. \quad (2.18)$$

Provedením každé iterace se teoreticky zmenšuje chyba  $J_n(\mathbf{h})$ . Filtr se tak neustále mění podle aktuální chyby. Opět zde platí předpoklad, že signály  $y$  a  $x$  jsou stacionární. Pak na uvažovaném intervalu existuje Wienerův filtr a adaptivní LMS k němu za určitých podmínek konverguje [7].

Velmi důležitá je volba kroku  $\mu$ . Bude-li  $\mu$  příliš malé, filtr se nezadaptuje dostatečně rychle z důvodu příliš pomalé konvergence. Příliš velké  $\mu$  může zase vést až k divergenci, což způsobí špatné fungování filtru. Správná volba  $\mu$  je složitá a závisí na mnoha faktorech, proto se často určuje experimentálně.

Podmínka konvergence je obecně u metody největšího spádu dána jako

$$0 < \mu < \frac{2}{\lambda_{max}}, \quad (2.19)$$

kde  $\lambda_{max}$  je největší vlastní číslo matice  $\mathbf{R}_n$  [7]. V případě adaptivního LMS se ovšem v každém kroku matice  $\mathbf{R}_n$  mění.

Pro filtry, které v aktuálním kroku upravují filtr z předchozí kroku, což je i případ adaptivního LMS, je dále nutné správně rozhodnout, zda na aktuálním bloku dat filtr adaptovat nebo raději nijak neměnit. Adaptovat filtr má význam v situacích, je-li v aktuálním bloku dat obsažen pouze okolní hluk, který je potlačován. Pokud by se prováděla adaptace i když je aktivní cílový zdroj (například řeč), může dojít k nežádoucím úpravám filtru a tím i ke zhoršení výsledků. Zde je třeba, pro rozhodnutí zda filtr neadaptovat z důvodu přítomnosti cílového zdroje, použít systémy pro jejich detekci, v případě řeči se jedná o detektor řečové aktivity (VAD – Voice Activity Detector), které o přítomnosti řeči rozhodnou.

### Normalizovaný adaptivní LMS filtr

Chybový signál, který odpovídá chybě v čase  $n$ , lze zapsat jako [7]

$$e[n] = x[n] - \mathbf{h}_n^T y[n]. \quad (2.20)$$

Jestliže se použije ve stejném čase  $n$  již adaptovaný filtr  $\mathbf{h}_{n+1}$  (filtr, který se aplikuje v následujícím okamžiku  $n + 1$ ), vznikne chybový signál

$$\varepsilon[n] = x[n] - \mathbf{h}_{n+1}^T y[n]. \quad (2.21)$$

Normalizovaný adaptivní LMS provádí minimální změnu definovanou jako

$$\delta \mathbf{h}_{n+1} = \mathbf{h}_{n+1} - \mathbf{h}_n, \quad (2.22)$$

tedy minimalizuje  $\|\delta \mathbf{h}_{n+1}\|_2^2$  s podmínkou, že

$$\varepsilon[n] = 0. \quad (2.23)$$

Po vyřešení dostaneme vztah pro adaptaci

$$\mathbf{h}_{n+1} \leftarrow \mathbf{h}_n - \mu \frac{\mathbf{y}_n}{\|\mathbf{y}_n\|_2^2} e[n]. \quad (2.24)$$

## 2.2 Potlačení hluku ve frekvenční oblasti

Jak již bylo uvedeno, některé metody zpracovávají vstupní signál v časově-frekvenční oblasti. Pro transformaci signálu do časově-frekvenční oblasti se používá krátkodobá Fourierova transformace (STFT – *Short-time Fourier Transform*). Operace STFT je lineární, takže signál  $y[n] = x[n] + d[n]$  transformovaný do časově-frekvenční oblasti je  $Y_k^{(n)} = X_k^{(n)} + D_k^{(n)}$ , kde  $k$  je index frekvence a  $n$  index bloku.

Lidský sluch není schopný rozlišovat fázi signálu [6]. Proto při potlačování hluku nemá význam jakkoliv fázi upravovat. Potlačování hluku ve frekvenční oblasti se tedy provádí reálným filtrem  $H_k^{(n)}$ . Aplikace filtru  $H_k^{(n)}$  na původní signál  $Y_k^{(n)}$  lze zapsat jako

$$\widehat{X}_k^{(n)} = H_k^{(n)} \cdot Y_k^{(n)} = H_k^{(n)} \cdot (X_k^{(n)} + D_k^{(n)}). \quad (2.25)$$

Jelikož filtr  $H_k^{(n)}$  je reálný, dochází v aktuálním bloku pouze k zesílení nebo zeslabení amplitud jednotlivých frekvencí. Fáze výstupního signálu  $\widehat{X}_k^{(n)}$  zůstává stejná jako u původního signálu  $Y_k^{(n)}$ .

Mnoho algoritmů redukce hluku v časově-frekvenční oblasti zavádí pro výpočet filtru  $H_k$  hodnoty a-priorní SNR (*Signal to Noise Ratio* – poměr signálu a hluku) a a-posteriorní SNR, které byly podle [6] poprvé definovány v článku [8].

Autoři zavedli  $\lambda_d(k) = E\{|D_k|^2\}$  a  $\lambda_x(k) = E\{|X_k|^2\}$ . Pak definovali a-priorní SNR  $\xi_k$  a a-posteriorní SNR  $\gamma_k$  jako

$$\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)} \quad (2.26)$$

$$\gamma_k = \frac{|Y_k|^2}{\lambda_d(k)}. \quad (2.27)$$

Hodnota  $\xi_k$  představuje průměrné SNR a  $\gamma_k$  lze chápat jako okamžité SNR.

### 2.2.1 Wienerův filtr ve frekvenční oblasti

Chyba výstupu je ve frekvenční oblasti [9]

$$C_k = X_k - H_k Y_k. \quad (2.28)$$

Wienerův filtr minimalizuje kritérium dané kvadrátem chybového signálu, tedy ve frekvenční oblasti

$$J(H_k) = E\{|C_k|^2\}. \quad (2.29)$$

Wienerův filtr je dán vztahem

$$H_k = \frac{E\{|X_k|^2\}}{E\{|X_k|^2\} + E\{|D_k|^2\}}. \quad (2.30)$$

S využitím zavedených pojmů a-priorní SNR a a-posteriorní SNR je Wienerův filtr

$$H_k = \frac{\xi_k}{1 + \xi_k}. \quad (2.31)$$

#### Aplikace Wienerova filtru v praxi

Wienerův filtr, jak byl výše popsán, naráží při nasazení ve skutečných systémech pro potlačování hluku na určité potíže [6]. Jedná se především o odhad a-priorního SNR  $\xi_k$ . V některých případech je ale možné provést odhad hluku a tím i vypočítat odhad a-posteriorního SNR. Se znalostí a-posteriorního SNR pak lze odhadnout a-priorní SNR, což vede k přibližnému řešení

$$H_k = \frac{\xi_k}{1 + \xi_k} \approx \frac{\gamma_k - 1}{\gamma_k} = \frac{|Y_k|^2 - \lambda_d(k)}{|Y_k|^2}. \quad (2.32)$$

Filtr  $H_k$  je definován jako nezáporný a reálný, to ovšem platí pouze pokud  $\lambda_d(k) \leq |Y_k|^2$ . Mohla by tedy nastat situace, kdy by filtr  $H_k$  obsahoval záporné hodnoty. Z toho důvodu se vztah (2.32) doplňuje o pravidlo limitující hodnoty na

nezáporné. Dále existuje potenciální riziko dělení nulou, pokud by byl signál na dané frekvenci nulový. Proto se do jmenovatele navíc přidává malá nenulová konstanta  $\epsilon$  zabráňující dělení nulou. Vztah pro Wienerův filtr široce využívaný v praktických aplikacích je

$$H_k = \frac{\max[0, |Y_k|^2 - \lambda_d(k)]}{|Y_k|^2 + \epsilon}. \quad (2.33)$$

Ve výsledném signálu sice dochází k potlačení hluku, nicméně se v něm mohou objevovat i různé artefakty a zkreslení. Některé artefakty vznikají hlavně během tichých segmentů, kdy je ve výsledném signálu slyšitelný specifický druh šumu nazývaný hudební šum (*musical noise*). Pozorování ukázala, že nejvíce hudebního šumu vzniká po vynulování frekvencí v daném bloku dat [6]. Z rovnice (2.33) je zřejmé, že vynulování frekvence nastane Wienerovým filtrem pokud  $\lambda_d(k) \geq |Y_k|^2$ . Hudební šum ale nevzniká pouze použitím Wienerova filtru, jedná se obecně o problém filtrů pro potlačení hluku.

## 2.2.2 Adaptivní LMS ve frekvenční oblasti

Jak již bylo uvedeno, chybový signál je ve frekvenční oblasti [10]

$$C_k = X_k - H_k Y_k. \quad (2.34)$$

Vztah pro adaptivní LMS je

$$H_k \leftarrow H_k + \mu Y_k C_k \quad (2.35)$$

a normalizovaný adaptivní LMS

$$H_k \leftarrow H_k + \mu \frac{Y_k}{|Y_k|^2 + \epsilon} C_k, \quad (2.36)$$

kde  $\epsilon$  je opět malá nenulová konstanta zabráňující dělení nulou.



### 2.2.3 Short-term MMSE

Velmi kvalitní, ovšem relativně výpočetně náročný filtr byl odvozen v článku [11]. Je dán vztahem

$$H_k = \frac{\sqrt{\pi v_k}}{2\gamma_k} \left[ (1 + v_k) I_0 \left( \frac{v_k}{2} \right) + v_k I_1 \left( \frac{v_k}{2} \right) \right] \exp \left( \frac{v_k}{2} \right), \quad (2.37)$$

kde  $I_0$  a  $I_1$  značí modifikované Besselovy funkce ( $I_0$  nultého řádu a  $I_1$  prvního řádu) a

$$v_k = \frac{\xi_k}{1 + \xi_k} \gamma_k. \quad (2.38)$$

V porovnání s Wienerovým filtrem lze pomocí filtru Short-term MMSE dosáhnout v potlačení hluku srovnatelných výsledků a zároveň menšího zkreslení signálu [6]. Hlavní nevýhoda ale spočívá v relativně vysoké výpočetní náročnosti, což pro systémy vyžadující výpočty v reálném čase může představovat značné problémy.

#### Efektivnější řešení

Výpočet filtru Short-term MMSE je i přes různé optimalizace stále relativně náročný. V článku [12] byly navrženy výpočetně efektivnější alternativy. Autoři odvodili tři způsoby, které optimalizují různá kritéria:

- Joint maximum a-posteriori spectral and phase (JMAP SAE) estimator
- Maximum a-posteriori spectral amplitude (MAP SA) estimator
- Minimum mean-square-error spectral power (MMSE SP) estimator

Všechny tyto způsoby jsou výrazně výpočetně úspornější než výpočet Short-term MMSE (2.37). Ve stejném článku dále autoři provedli srovnání výsledků dosažených těmito třemi způsoby s Short-term MMSE. Tabulka 2.1 [6] zobrazuje vztahy pro výpočet filtru  $H_k$ , průměrnou a maximální odchylku od výsledku Short-term MMSE pro  $(\gamma_k - 1, \xi_k) \in [-30, 30]$  dB.

Tabulka 2.1: Výpočet aproximace short-term MMSE

Optimalita	Vztah výpočtu	Prům. odch. [dB]	Max. odch. [dB]
JMAP SAE	$H_k = \frac{\xi_k + \sqrt{\xi_k^2 + 2(1+\xi_k)\frac{\xi_k}{\gamma_k}}}{2(1+\xi_k)}$	+0,52	+1,77
MAP SA	$H_k = \frac{\xi_k + \sqrt{\xi_k^2 + (1+\xi_k)\frac{\xi_k}{\gamma_k}}}{2(1+\xi_k)}$	+1,26	+4,70
MMSE SP	$H_k = \sqrt{\frac{\xi_k}{1+\xi_k} \left( \frac{1+v_k}{\gamma_k} \right)}$	+0,69	-1,05

## 2.2.4 Short-term Log-MMSE

Lidský sluch vnímá frekvence a hlasitosti v logaritmickém měřítku. V článku [13] byl navržen filtr pro potlačení hluku zohledňující tuto skutečnost. Tento filtr je dán vztahem

$$H_k = \frac{\xi_k}{1 + \xi_k} \exp \left\{ \frac{1}{2} \int_{v_k}^{\infty} \frac{\exp(-t)}{t} dt \right\}, \quad (2.39)$$

kde  $v_k$  je definováno stejně jako ve vztahu (2.38). Podobně jako u Short-term MMSE je zde pro systémy vyžadující zpracování v reálném čase problém relativně vysoké výpočetní náročnosti. Hlavní problém z hlediska výpočetní náročnosti představuje ve vztahu (2.39) integrál. Jelikož Short-term Log-MMSE patří k nejlepším filtrům pro potlačování hluku [6], pro výpočet integrálu v reálném čase bylo navrženo mnoho interpolací a aproximací, které umožňují rychlý výpočet.

## 3 Odhad hluku

V systémech pro redukci hluku, které podle zavedené terminologie patří do skupiny *Noise Cancellation*, je důležitým krokem během zpracování získání referenčního signálu, který koresponduje s okolním hlukem obsaženým ve zpracovávané nahrávce. Získaný referenční signál lze pak použít pro redukci hluku. Jak již bylo zmíněno, přesná podoba hluku bývá ale známá pouze v experimentálních úlohách, proto je třeba při zpracování nahradit hluk jeho odhadem. Obecně ani není nezbytně nutné odhadnout okolní hluk jako signál, k redukci může stačit znát jen některé jeho vlastnosti. Jedná-li se například o jednoduchou situaci, ve které se z hlediska obsažených frekvencí okolní hluk s cílovým signálem nijak výrazně nepřekrývá, stačí pro redukci hluku znát pouze frekvenční pásmo, kde se hluk vyskytuje.

Ve složitějších případech je ale nutné ke kvalitnímu odhadu hluku a tím účinné redukci hluku použít jiné postupy. Jak již bylo zmíněno, tato problematika je velmi komplikovaná a obvykle je třeba do systémů redukce hluku zanést skutečnosti zohledňující charakteristiku daného problému.

Je-li například možné zohlednit pozici cílového zdroje vůči nahrávacímu zařízení, což je i případ telefonních hovorů, je jednou z možností pro získávání referenčního signálu použití banky předem připravených *target cancellation filtrů* (CF), které za určitých podmínek cílový signál potlačí, ale okolní hluk propustí.

### 3.1 Target cancellation filtr (CF)

Target cancellation filtr (CF) je digitální filtr typu FIR s několika vstupy, který v nahrávce potlačuje akustické zdroje především v závislosti na jejich pozici vůči mikrofonom nahrávacího zařízení. Je-li použit CF pro správnou pozici, lze v nahrávce docílit potlačení cílového zdroje. Výpočet impulsní odezvy CF se provádí z vícekanálové akustické nahrávky. Nahrávka vhodná pro výpočet CF obsahuje

pouze jeden akustický (cílový) zdroj bez okolního hluku. CF vypočítaný z takové nahrávky pak výrazně potlačuje cílový zdroj pro pozici, ve které se zdroj nacházel v nahrávce použité pro výpočet CF.

### 3.1.1 CF se dvěma vstupy

Pro CF je třeba uvažovat více vstupních kanálů, minimálně dva. Jak již bylo zmíněno, u metod pro mobilní telefony je kladen důraz na dvoukanálové zpracování. Odvození výpočtu pro CF se dvěma vstupními kanály jsme popsali v článku [14]. Dvoukanálové nahrávání s fixní pozicí cílového zdroje a obecným okolním hlukem popisujeme jako

$$\begin{aligned} y_L[n] &= \{h_L * x\}[n] + d_L[n], \\ y_P[n] &= \{h_P * x\}[n] + d_P[n]. \end{aligned} \tag{3.1}$$

Značení:

$y_L[n], y_P[n]$	...	signály zaznamenané levým a pravým mikrofonem
$h_L[n], h_P[n]$	...	impulsní odezvy ovlivňující záznam cílového signálu na levém a pravém mikrofonu
$d_L[n], d_P[n]$	...	signály s hlukem na levém a pravém mikrofonu
$x[n]$	...	cílový signál
$n$	...	časový index, $n = 1 \dots N$
$*$	...	konvoluce

Jelikož předpokládáme, že se cílový zdroj (například mluvčí) vůči mikrofonům nahrávacího zařízení nepohybuje, popisujeme záznam cílového signálu jako  $\{h_L * x\}[n]$  a  $\{h_P * x\}[n]$ , tedy impulsní odezvy  $h_L$  a  $h_P$  ovlivňující záznam cílového signálu se během nahrávání nemění.

Pokud bychom uvažovali pro okolní hluk pouze jeden bodový zdroj, bylo by možné popsat záznam okolního hluku například jako  $\{k_L * d\}[n]$  a  $\{k_P * d\}[n]$ , kde  $k_L$  a  $k_P$  by byly impulsní odezvy vyjadřující vliv na záznam okolního hluku. Zdrojů hluku ale bývá v reálných situacích více a navíc se nemusí jednat pouze o bodové zdroje. Proto v (3.1) popisujeme okolní hluk jako signály  $d_L[n]$  a  $d_P[n]$ , které obecně mohou obsahovat směs jakýchkoliv signálů.

Hlavním požadavkem na CF je potlačení cílového signálu  $x[n]$ . Jak bylo popsáno v (3.1), cílový signál zaznamenaný levým mikrofonom není totožný se signálem zaznamenaným na pravém mikrofону. Ideální CF zcela potlačující cílový signál  $x[n]$  proto obsahuje dva nenulové filtry  $g_L$  a  $g_P$  (filtry s jedním vstupem a jedním výstupem), pro které platí

$$\{g_L * h_L * x\}[n] = \{g_P * h_P * x\}[n]. \quad (3.2)$$

Jak uvádí vztah (3.2), filtry  $g_L$  a  $g_P$  provedou se záznamem takové úpravy, že cílový signál zaznamenaný levým mikrofonom po zfiltrování filtrem  $g_L$  je stejný jako cílový signál zaznamenaný pravým mikrofonom po zfiltrování filtrem  $g_P$ . Pokud odečteme tyto signály od sebe, dojde k úplnému potlačení cílového signálu a tedy výstup ideálního CF je

$$\begin{aligned} z[n] &= \{g_L * y_L\}[n] - \{g_P * y_P\}[n] \\ &= \{g_L * h_L * x\}[n] + \{g_L * d_L\}[n] - \{g_P * h_P * x\}[n] + \{g_P * d_P\}[n] \quad (3.3) \\ &= \{g_L * d_L\}[n] - \{g_P * d_P\}[n]. \end{aligned}$$

Jelikož v signálu  $z[n]$  (výstup ideálního CF) došlo k úplnému potlačení cílového signálu  $x[n]$ , signál  $z[n]$  obsahuje pouze okolní hluk. Ze vztahu (3.3) je ale vidět, že s potlačením cílového signálu  $x[n]$  dochází i ke změnám na signálech s hlukem. Signál s hlukem z levého mikrofону  $d_L[n]$  je zpracován filtrem  $g_L$  a podobně je i signál  $d_P[n]$  zpracován filtrem  $g_P$ .

Hlavní požadavek na CF (potlačení cílového signálu) je sice splněn, ovšem filtry  $g_L$  a  $g_P$  mohou signály s hlukem velmi výrazně pozměnit, což může zkomplikovat další zpracování (například potlačení hluku s použitím výstupu CF jako referenčního signálu).

### 3.1.2 Návrh CF se dvěma vstupy

Pro návrh CF zavádíme značení, kde symboly  $\mathbf{Y}_i$ ,  $i \in \{L, P\}$  označujeme signály  $y_i[n]$  v Toeplitzovských maticích, které jsou ve tvaru

$$\mathbf{Y}_i = \begin{bmatrix} y_i[0] & y_i[1] & y_i[2] & \cdots & \cdots & y_i[N] \\ 0 & y_i[0] & y_i[1] & \cdots & \cdots & y_i[N-1] \\ 0 & 0 & y_i[0] & \cdots & \cdots & y_i[N-2] \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & y_i[0] & \cdots & y_i[N-M] \end{bmatrix}. \quad (3.4)$$

Podobně zavádíme Toeplitzovské matice  $\mathbf{D}_i$  pro signály  $d_i[n]$ . Vektory  $\mathbf{g}_L$  a  $\mathbf{g}_P$  obsahují koeficienty filtrů  $g_L$  a  $g_P$  délky  $M$ .

Jak již bylo uvedeno, CF vypočítáváme z nahrávky cílového zdroje bez okolního hluku. Podle zavedeného značení tedy nyní předpokládáme, že signály s hlukem  $d_i[n]$  jsou nulové a zaznamenané signály  $y_i[n]$  tak obsahují pouze cílový signál.

#### Návrh $\mathbf{CF}_1$

Jeden ze způsobů návrhu CF, který dále označujeme jako  $\mathbf{CF}_1$ , spočívá v definici vektoru  $\hat{\mathbf{g}}_P = \mathbf{e}_I$  [15], kde  $\mathbf{e}_I$  je jednotkový vektor délky  $M$  obsahující jedna na  $I$  té pozici a  $I$  je celé číslo, které udává celkové zpoždění výstupu CF.  $\mathbf{CF}_1$  je vypočítáván jako

$$\mathbf{CF}_1 : \quad \hat{\mathbf{g}}_L = \arg \min_{\mathbf{g}_L} \|\mathbf{g}_L^T \mathbf{Y}_L - \hat{\mathbf{g}}_P^T \mathbf{Y}_P\|_2^2. \quad (3.5)$$

Řešením je

$$\hat{\mathbf{g}}_L = \mathbf{W}^{-1} \mathbf{h}, \quad (3.6)$$

kde

$$\begin{aligned} \mathbf{W} &= \mathbf{Y}_L \mathbf{Y}_L^T \\ \mathbf{h} &= \mathbf{Y}_L \mathbf{y} \end{aligned} \quad (3.7)$$

a vektor  $\mathbf{y}$  obsahuje signál  $y_P[n - I]$ . Jak je vidět ze vztahu (3.6), návrh  $CF_1$  obsahuje výpočet inverzní matice  $\mathbf{W}$ , což je obecně velmi náročná operace. Jelikož je ale  $\mathbf{W}$  symetrická Toeplitzovská matice, lze pro vyřešení vztahu (3.6) použít algoritmus Levinson-Durbin [16] [17], jehož složitost je  $\mathcal{O}(M^2)$ .

Návrh  $CF_1$  má nevýhodu, že nijak nezohledňuje účinek CF na výstupní signál. Výstupní signál CF by měl obsahovat především okolní hluk, který ovšem po aplikaci  $CF_1$  může být různě zabarven. Vznik zabarvení je zřejmý ze vztahu (3.3).

## Návrh $CF_2$

Jiný způsob návrhu umožňuje přizpůsobit CF na určitý druh hluku, kterým tak lze ovlivnit zabarvení výstupu CF. Tento způsob jsme navrhli v článku [14] a dále ho označujeme jako  $CF_2$ .

Signály  $y_i[n]$  opět obsahují pouze záznam cílového signálu, ale signály  $d_i[n]$  nyní představují referenční hluk, na který bude filtr přizpůsoben. Referenční hluk může být libovolná nahrávka neobsahující cílový signál. Lze ji pořídít buď zvlášť nebo i z části nahrávky, při které není cílový zdroj aktivní. Výpočet  $CF_2$  představuje nalezení řešení

$$CF_2: \quad \hat{\mathbf{g}}_L, \hat{\mathbf{g}}_P = \arg \min \|\mathbf{g}_L^T \mathbf{Y}_L - \mathbf{g}_P^T \mathbf{Y}_P\|_2^2 + \epsilon \|\mathbf{g}_L^T \mathbf{D}_L - \mathbf{g}_P^T \mathbf{D}_P - \mathbf{d}\|_2^2, \quad (3.8)$$

kde  $\epsilon$  je kladný regularizační parametr a  $\mathbf{d}$  představuje signál na výstupu CF (signál  $z[n]$  v rovnici (3.3)). Vektor  $\mathbf{d}$  může obsahovat různé signály, vhodná volba je například signál  $d_L[n - I]$ , kde  $I$  udává zpoždění jako u (3.5). Zde je navíc minimalizován rozdíl mezi výstupem CF ( $\mathbf{g}_L^T \mathbf{D}_L - \mathbf{g}_P^T \mathbf{D}_P$ ) a signálem s požadovaným výstupem CF (vektor  $\mathbf{d}$ ). Zmíněná možnost ovlivnění zabarvení výstupu  $CF_2$  se tedy provádí obsahem vektoru  $\mathbf{d}$ . Řešením vztahu (3.8) je

$$\begin{bmatrix} \hat{\mathbf{g}}_L \\ \hat{\mathbf{g}}_P \end{bmatrix} = \mathbf{W}^{-1} \mathbf{h}, \quad (3.9)$$

kde

$$\begin{aligned}\mathbf{W} &= \begin{bmatrix} \mathbf{Y}_L \\ -\mathbf{Y}_P \end{bmatrix} \begin{bmatrix} \mathbf{Y}_L^T, -\mathbf{Y}_P^T \end{bmatrix} + \epsilon \begin{bmatrix} \mathbf{D}_L \\ -\mathbf{D}_P \end{bmatrix} \begin{bmatrix} \mathbf{D}_L^T, -\mathbf{D}_P^T \end{bmatrix} \\ \mathbf{h} &= \epsilon \begin{bmatrix} \mathbf{D}_L \\ -\mathbf{D}_P \end{bmatrix} \mathbf{d}.\end{aligned}\tag{3.10}$$

Podobně jako u  $\text{CF}_1$  (3.6) obsahuje vztah (3.9) pro návrh  $\text{CF}_2$  výpočet inverzní matice  $\mathbf{W}$ . V tomto případě se ale jedná o symetrickou blokovou Toeplitzovskou matici, tudíž pro vyřešení vztahu (3.9) lze použít zobecněný algoritmus Levinson-Durbin [18] pro symetrické blokové Toeplitzovské matice se složitostí  $\mathcal{O}(aM^2)$ , kde  $a$  je počet bloků (pro  $\text{CF}_2$   $a = 2$ ).

Oproti  $\text{CF}_1$  má  $\text{CF}_2$  výhodu, že umožňuje zmírnit nežádoucí účinky CF na okolní hluk.  $\text{CF}_1$  má ale zase nižší výpočetní nároky (pro výpočet i aplikaci). Pro aplikaci  $\text{CF}_1$  i  $\text{CF}_2$  je sice nutné dvakrát provést konvoluci, ovšem u  $\text{CF}_1$  provádí  $\hat{\mathbf{g}}_P$  pouze zpoždění, které lze realizovat efektivněji než pomocí konvoluce.



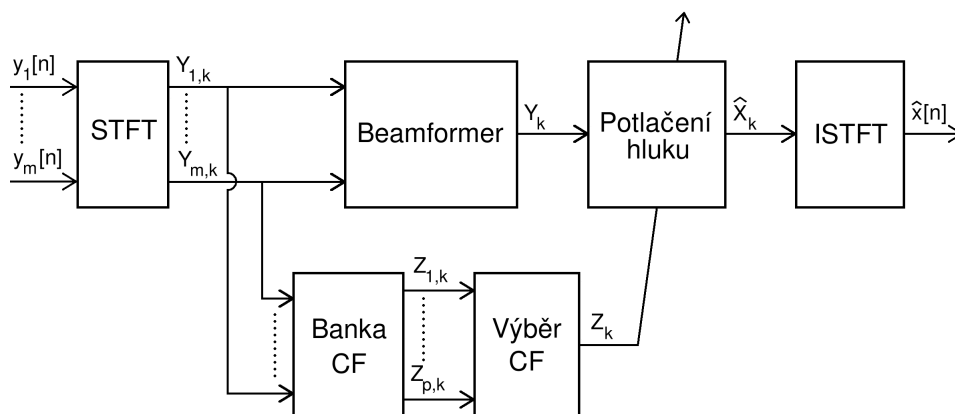
## 4 Systém redukce hluku s bankou CF

Navrhujeme systém redukce hluku, který lze kromě jiných využití implementovat i do mobilních telefonů osazených alespoň dvěma mikrofony. Tento systém jsme prezentovali v článku [14]. Originalita systému spočívá ve způsobu získávání referenčního signálu, který následně použijeme pro redukci okolního hluku. Referenční signál získáváme pomocí banky předem připravených CF.

Aplikací kvalitní banky CF na zaznamenaný signál dojde k velmi výraznému potlačení cílového signálu (hlasu telefonující osoby) a vznikne tak signál obsahující především okolní hluk. V praxi zpravidla nedochází k úplnému potlačení cílového signálu a jak bylo uvedeno v kapitole 3.1, banka CF navíc částečně ovlivní ve zpracovávaném signálu i složku s hlukem. Dojde-li ale bankou CF k dostatečnému potlačení cílového signálu, lze takto získaný referenční signál použít k účinné redukci okolního hluku.

### 4.1 Schéma systému

Postup zpracování navrženého systému je zobrazen schématem na obrázku 4.1.



Obrázek 4.1: Blokové schéma navrženého systému redukce hluku

## Význam bloků

Signály zaznamenané mikrofony  $y_1[n], \dots, y_m[n]$  jsou transformovány operací STFT do časově-frekvenční oblasti na signály  $Y_{1,k}, \dots, Y_{m,k}$  (uvažujeme obecně více mikrofónů, minimálně dva).

Vstupní signál je zpracován všemi CF v bance (banka obsahuje FFT obrazy všech CF, protože zpracování provádíme v časově-frekvenční oblasti). Jelikož zpracovávání vstupního signálu jednotlivými CF není na sobě nijak závislé, lze tuto operaci provádět paralelně. Z výstupů všech CF (signály  $Z_{1,k}, \dots, Z_{p,k}$ ) se vybere referenční signál  $Z_k$  (nejvěrnější podoba hluku). Za předpokladu, že CF potlačí především cílový signál a okolní hluk zůstane bez výrazných změn, může být jako referenční signál  $Z_k$  vybrán signál s nejnižší energií. Nicméně referenční signál vybraný tímto způsobem nemusí obecně představovat signál s nejvyšším potlačením cílového signálu. Pro spolehlivější výběr správného výstupu CF lze například použít složitější metody z článku [19].

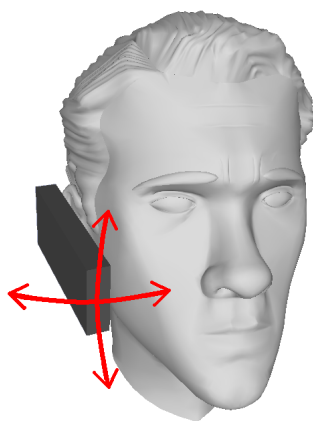
Z bloku *Beamformer* vychází počáteční odhad cílového signálu. Je možné použít signál přímo z vybraného mikrofónu. Výběr lze provést například na základě umístění mikrofónů nebo i podle energie zaznamenaných signálů. Výstupem zde může být signál s nejvyšší energií, protože ten bude pravděpodobně obsahovat cílový signál výrazněji (za předpokladu, že okolní hluk má na všech mikrofónech stejnou energii).

V počátečním odhadu cílového signálu  $Y_k$  se následně provede potlačení referenčního signálu  $Z_k$ . Metod pro tuto operaci existuje mnoho, může se jednat například o filtry popsané v kapitole 2.2. Pro zvýšení poslechové kvality je ještě možné od určité frekvence (například 2 až 3 kHz) omezovat potlačování hluku. K výpočtu filtru  $H_k$  se přidá od dané frekvence podmínka, která limituje jeho minimální hodnoty na zadanou mez.

Poslední krok představuje operaci ISTFT, kterou se převede signál s potlačeným hlukem  $\widehat{X}_k$  z časově-frekvenční oblasti zpět do časové oblasti. Signál  $\hat{x}[k]$  je výstupem systému redukce hluku.

## 4.2 Vytvoření banky CF

Pro navržený systém redukce hluku je nutné před použitím nejprve vytvořit banku CF. Během běžného telefonního hovoru se telefon pohybuje vůči mluvčímu pouze v určité oblasti, kterou znázorňuje obrázek 4.2 a právě tuto oblast je třeba pokrýt bankou CF. Obrázek 4.2 jsme vytvořili v aplikaci *Google SketchUp* verze 7.0.8657. Model hlavy pochází z databáze modelů *3D Warehouse* [20].



Obrázek 4.2: Oblast nejpravděpodobnějších pozic mobilního telefonu vůči mluvčímu

Vytvoření banky CF lze provést různými způsoby, například definováním sítě pozic, pro které se CF vypočítají (kapitola 4.2.1) nebo i algoritmem učení (kapitola 4.2.2). Vždy je ale nutné zohlednit alespoň ty nejdůležitější skutečnosti, na kterých nejvíce závisí potlačení cílového signálu pomocí CF. Jedná se především o závislost na pozici a závislost na mluvčím.

Jak bude vidět z experimentu v kapitole 5.3.1, potlačení CF je na pozici velmi závislé. Banka CF tedy musí být dostatečně obsáhlá a co nejlépe pokrývat uvažovanou oblast. Dále jsou CF závislé na akustickém zdroji, z jehož nahrávky byly vypočítány. Jedná-li se o záznamy skutečných osob, potlačení CF závisí i na konkrétních mluvčích. Z experimentu v kapitole 5.4.1 je zřejmé, že získání nejlepší podoby referenčního signálu je možné pouze s bankou CF přizpůsobenou konkrétnímu uživateli. I když je banka CF vytvořená z nahrávky přímo od skutečné, ovšem jiné osoby, může dojít k výraznému poklesu potlačení cílového signálu a tím i ke znatelnému snížení kvality výstupního signálu. Pravděpodobně tedy neexistuje

jakási „univerzální“ banka CF vhodná pro použití ve všech mobilních telefonech. Do telefonů je nutné implementovat způsoby, kterými si uživatel bude moci snadno přizpůsobit banku CF na sebe. Právě k tomu jsme navrhli algoritmus učení banky CF popsany v kapitole 4.2.2.

Banka CF sice musí dostatečně pokrývat uvažovanou oblast, nicméně zvyšování počtu CF v bance může způsobit i zhoršení výsledků. Uvažujeme uvedený způsob pro výběr referenčního signálu (výstup CF s minimální energií). Jedná-li se pouze o potlačování cílového signálu v nahrávce bez hluku, přidávání dalších CF může výsledky pouze zlepšit, protože každý nový CF může potlačit cílový signál buď více, a pak bude jeho výstup vybrán jako referenční signál, nebo méně a tím výběr referenčního signálu nijak neovlivní. Při zpracování nahrávek obsahujících i hluk je ovšem situace jiná. Jak již bylo zmíněno, CF neovlivňuje pouze cílový signál. Může tedy nastat situace, kde nový CF přidaný do banky bude mít výstup sice s nejnižší energií, ovšem v bance bude jiný CF, v jehož výstupu bude cílový signál potlačen výrazněji. Výstup nově přidaného CF tedy bude chybně vybrán jako referenční signál, což následně způsobí výraznější deformace ve výstupním signálu (v původním signálu se potlačí výraznější část cílového signálu). S obsáhlejší bankou CF se toto riziko zvyšuje.

### 4.2.1 Banka CF jako síť definovaných pozic

Základní způsob pro vytvoření banky CF spočívá v definování pozic, ve kterých se pořídí nahrávky pro výpočty CF. Tento postup je vhodný pro určitý druh experimentů a studování vlastností CF. Banky CF jsme vytvářeli tímto způsobem v rámci některých experimentů v kapitole 5.3.1, kde mimo jiné zkoumáme vliv potlačení na pozici.

Pro použití v mobilních telefonech by taková síť CF měla různou hustotu. Nejvíce CF by se vyskytovalo v nejpravděpodobnějších pozicích, zatímco na okraji uvažované oblasti by byla vzdálenost bezprostředně sousedních pozic větší. Snížil by se tak počet CF v bance a s tím i výpočetní náročnost, protože jak bylo uvedeno v popisu systému, každý blok vstupního signálu je nutné zpracovat všemi CF v bance.

Vytváření banky CF na základě definovaných pozic představuje pro použití v praxi určité potíže. Jelikož je třeba umožnit přizpůsobení banky CF běžným

uživatelům (CF jsou příliš závislé na konkrétním mluvčím), postup přizpůsobení banky musí být z hlediska uživatele co nejjednodušší. Běžným uživatelům by tento způsob vytváření banky CF jistě působil příliš mnoho problémů.

#### 4.2.2 Algoritmus učení banky CF

Navrhujeme metodu pro vytváření bank CF, která je pro běžné uživatele dostatečně snadná a nevyžaduje ani jakékoliv odborné znalosti. Tuto metodu jsme prezentovali v článku [21] a dále ji označujeme jako algoritmus učení banky CF.

Jediné co zde bude muset uživatel pro vytvoření banky CF udělat sám, je spustit nahrávání, držet svůj telefon jako při hovoru a mluvit (v tichém prostředí). Během nahrávání je třeba telefonem pomalu pohybovat, aby pořizovaná nahrávka pokrývala celou oblast možných pozic telefonu vůči mluvčímu. Po ukončení nahrávání se tento záznam zpracuje, čímž vznikne nová banka CF přizpůsobená danému uživateli.

Nahráný signál je rozdělen do bloků, které se mohou překrývat. Na každém bloku nebo jeho části je cílový signál (hlas mluvčího) potlačen všemi CF obsaženými v bance, z nichž se vybere CF s nejvyšším potlačením cílového signálu. Podobně jak bylo uvedeno v popisu navrženého systému, signál s nejvyšším potlačením cílového signálu lze vybrat na základě nejnižší energie výstupů CF nebo i pomocí jiných složitějších metod [19]. Není-li potlačení na aktuálním bloku již vytvořenou bankou CF dostatečné, vypočítá se na tomto bloku nový CF podle (3.5) nebo (3.8). Tento nový CF je následně aplikován na aktuální blok a je přidán do banky, pokud jeho potlačení je vyšší než u ostatních CF. Potom se pokračuje na další blok a celý proces je opakován až do dosažení konce trénovací nahrávky.

Tento základní postup může být dále rozšiřován. Jak bylo popsáno, nový CF bude přidán vždy, pokud v jeho výstupu nastalo vyšší potlačení cílového signálu než u ostatních CF. Banka tudíž pak může obsahovat některé CF velmi podobné. Jednou z možností jak tomu předejít, je nepřidávat do banky nový CF vždy, když dosáhne vyššího potlačení, ale jen pokud bude jeho potlačení vyšší o stanovenou mez (například 1 dB). Pokud nový CF má sice vyšší potlačení než dosavadní CF s nejvyšším potlačením, ovšem méně než o stanovenou mez, je také možné dosavadní CF s nejvyšším potlačením tímto novým nahradit.

Dále lze z důvodu výpočetní náročnosti limitovat počet CF v nové bance. Pokud by se ale přerušilo zpracování vstupní nahrávky, jakmile byl limit dosažen, může se stát, že nedojde ke zpracování celé nahrávky a tím i k nepokrytí celé uvažované oblasti bankou CF. Pro tuto situaci navrhuje postup zpracování ve dvou krocích, kde v prvním kroku přidáváme nový CF vždy, když byly splněny všechny podmínky pro přidání. V druhém kroku (po zpracování celé trénovací nahrávky) vyhodnotíme pro všechny CF skóre, na základě kterého jeden odstraníme. Skóre pro CF lze vypočítávat například jako pokles potlačení v trénovací nahrávce bankou CF bez účasti daného CF. Po odstranění vybraného CF je třeba pro zbylé CF opět vyhodnotit skóre, protože po odstranění jediného CF se může takto vypočítávané skóre pro ostatní CF změnit. Postup se pak opakuje až do dosažení požadovaného počtu.

Celá navržená metoda je popsána pseudokódem v algoritmu 1, jehož vstupní parametry jsme označili:

*trainData* ... záznam hlasu mluvčího  
*attReq* ... požadované potlačení na aktuálním bloku  
*NLearn* ... délka bloku pro výpočet CF  
*WLen* ... délka bloku pro test účinnost CF ( $WLen \leq NLearn$ )  
*shift* ... velikost posunu na další blok ( $shift \leq NLearn$ )  
*btr* ... hodnota vyjadřující o kolik musí mít CF vyšší potlačení  
oproti dosavadnímu nejvyššímu, aby byl přidán do banky  
*C* ... limit na počet CF v bance (nepovinný parametr)

V popisu algoritmu jsme dále zavedli značení:

$len(trainData)$  ... délka trénovacích dat  
 $trainData(i_1 :: i_2)$  ... blok trénovacích dat od indexu  $i_1$  do  $i_2$   
 $attTarg(banka, blok, WLen)$  ... nejvyšší dosažitelné potlačení bankou *banka*  
na části dlouhé *WLen* z bloku dat *blok*  
 $cmpCF(blok)$  ... výpočet CF z bloku dat *blok*

---

**Algoritmus 1** Učení banky CF

---

**Vstup:** trainData, attReq, NLearn, WLen, shift, btr, (C)

```
bank = ∅; istart = 0;
while (istart + NLearn - 1) < len(trainData) do
    block = trainData(istart :: [istart + NLearn - 1]);
    attHi = attTarg(bank, block, WLen);
    if attReq > attHi then
        newCF = cmpCF(block);
        attNew = attTarg(newCF, block, WLen);
        if (attNew + btr) > attHi then
            bank = bank ∪ newCF;    //přidání newCF do banky
            if (attNew - attHi) ≤ btr then
                volitelné: odebrat předchozí CF s nejvyšším potlačením;
            end if
        end if
    end if
    istart = istart + shift;
end while
volitelné: vybrat C nejdůležitějších CF;
return bank
```

---

### Vliv vstupních parametrů

Všechny parametry mohou výrazně ovlivnit průběh učení banky a tím i její kvalitu. Nižší hodnota parametru požadované potlačení (*attReq*) způsobí častější posun na další blok bez výpočtu nového CF. Nastaví-li se tento parametr na vyšší hodnotu, bude docházet k častějším výpočtům CF z aktuálních bloků. Pravděpodobně se tak budou objevovat CF s vyšším potlačením a tím se i budou přidávat do banky. Účinnost banky (potlačení cílového signálu) se sice může zvýšit, ovšem za cenu vyššího počtu CF. Může ale také nastat situace, kde tyto nové CF nebudou dostatečně účinné pro přidání a vyšší hodnota parametru *attReq* tak nebude mít žádné důsledky.

Znatelný vliv má nastavení délky bloků pro výpočet CF (parametr *NLearn*). V experimentech v kapitole 5.3.1 vznikaly banky CF obsáhlejší, pokud jsme nastavili pro výpočty CF kratší bloky než v případech s delšími bloky. Rozdíly v účinnostech

bank ale nebyly nijak zásadní. Jako velmi důležité se ukázalo i nastavení délky bloků pro test účinnosti CF (parametr  $WLen$ ). Banky v experimentech vznikaly kvalitnější, pokud jsme testovali účinnost CF jen na části aktuálního bloku. Tento parametr výrazně ovlivňoval průběhy učení bank a tím pak i jejich účinnost.

Rozdíly v bankách CF mohou nastat i s pravidly pro přidání nového CF (parametr  $btr$ ). V experimentech v kapitole 5.3.1 jsme také testovali dvě nastavení, kde v prvním jsme přidávali nový CF vždy, když byl účinnější než dosavadní nejúčinnější, a v druhém, pokud byl nový CF účinnější alespoň o 1 dB. Zde se ukázalo, že zavedení tohoto pravidla může účinnost banky zvýšit i snížit.

Podle experimentů provedených v kapitolách 5.3 a 5.4 jsou ale banky vytvořené algoritmem učení v potlačení cílového signálu dostatečně účinné a lze je tak použít pro získávání referenčního signálu. V porovnání s obsáhlejšími bankami vytvořenými jako síť pozic byly banky vytvořené algoritmem učení v některých případech i účinnější. Velmi ale záleží na nastavení všech parametrů a v neposlední řadě i na samotné trénovací nahrávce. Hlavní výhodou tvorby bank CF algoritmem učení je především nízká obtížnost z hlediska uživatele, což je velmi důležité pro nasazení systému do skutečných mobilních telefonů.



## 5 Experimenty

Provedli jsme sérii experimentů, ve kterých jsme otestovali systém redukce hluku navržený v kapitole 4. Nejprve testujeme potlačení cílového signálu pomocí bank CF vytvořených různými způsoby, v dalších experimentech pak potlačujeme i hluk.

V kapitole 5.3 používáme experimentální soustavu obsahující umělou hlavu pro simulaci mluvčích. Nejprve provádíme potlačování cílového signálu (kapitola 5.3.1), kde vytváříme banky CF jako sítě definovaných pozic a zpracováváme nahrávky ve fixní i proměnlivé pozici. Dále zde vytváříme banky CF algoritmem učení a srovnáváme výsledky dosažené s bankami vytvořenými jako síť pozic. Následně provádíme i potlačování hluku (kapitola 5.3.2). Schéma experimentů potlačení cílového signálu z kapitoly 5.3.1 jsme také použili pro část experimentů provedených v rámci článku [21].

Experimenty s nahrávkami skutečných osob jsme popsali v kapitole 5.4. Opět nejprve provádíme potlačení cílového signálu (kapitola 5.4.1) a pak i potlačování hluku (kapitola 5.4.2).

Zavádíme označení, kde sadu nahrávek, která slouží pro vytvoření banky CF, označujeme jako sadu *trénovací*. Sadu nahrávek, kterou poté zpracováváme (potlačení cílového signálu a potlačení hluku) označujeme jako sadu *testovací*.

V kapitole 5.1 jsou popsána používaná zařízení pro záznam zvuku. Mezi nimi je i model telefonu, kterým simulujeme mobilní telefon se dvěma předními mikrofony, nebo jedním předním a druhým zadním mikrofonom. Jeden mikrofón je společný pro obě rozmístění mikrofónů. Experimenty provádíme vždy pro obě rozmístění mikrofónů a abychom mohli tyto výsledky porovnávat, cílový signál potlačujeme v signálu zaznamenaném právě tímto společným mikrofonom.

Ve všech experimentech vypočítáváme CF vždy délky 1000 a parametr zpoždění ze vztahů (3.5) a (3.8) nastavujeme na  $I = 20$ . V experimentech potlačení cílového signálu používáme jen  $CF_1$ , při potlačování hluku i  $CF_2$ . Jak bylo uvedeno v ka-

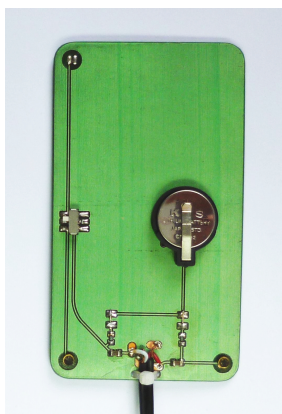
pitole 3.1,  $CF_2$  lze přizpůsobit na určitý signál s hlukem. Je ale rozdíl, jestli se pro přizpůsobení použije přímo signál pocházející z externí databáze nahrávek, nebo jestli se tentýž signál přehraje reproduktorem a pro přizpůsobení použije jeho záznam. V našich experimentech používáme druhou variantu, kde všechny  $CF_2$  přizpůsobujeme na signály zaznamenané modelem telefonu.

Signály vždy pořizujeme a zpracováváme se vzorkovací frekvencí 16 kHz a 16 bity na vzorek. Zpracovávání provádíme po blocích délky 3000 vzorků s překryvem 2500 vzorků. Jako referenční signál používáme výstup CF s nejnižší energií.

## 5.1 Audio zařízení

### Model mobilního telefonu

Mobilní telefon simulujeme modelem zobrazeném na obrázku 5.1.



Obrázek 5.1: Model mobilního telefonu

Model telefonu obsahuje zapojení tří elektretových mikrofونů, přičemž dva jsou umístěny vpředu vedle sebe a třetí vzadu vlevo nahoře. Přepínačem lze vybrat, zda budou zapojeny oba přední mikrofony (toto rozmístění dále označujeme jako PP), nebo přední a zadní mikrofون (označujeme PZ). U rozmístění PZ je zapojen pravý přední mikrofон, aby vzdálenost mezi předním a zadním mikrofonom byla co největší. Pravý přední mikrofон je tedy společný pro obě rozmístění. Tento typ mikrofонů byl použit v mobilním telefonu Sony Ericsson K850.

Obě rozmístění mikrofونů mají své výhody i nevýhody. Jedná-li se o běžný telefonní hovor, v případě rozmístění mikrofонů PZ bude signál zaznamenaný předním mikrofonom obsahovat především cílový signál (hlas telefonující osoby) a signál zaznamenaný zadním především okolní hluk, zatímco při rozmístění mikrofонů PP si budou zaznamenané signály výrazně podobnější. Rozmístění mikrofонů PZ poskytuje výhodnější výchozí podmínky pro potlačování hluku právě kvůli přesnější počáteční podobě cílového signálu a okolního hluku. Pro řadu jiných aplikací je ale výhodnější rozmístění PP (například lokalizace zdroje na základě zpoždění signálu mezi mikrofony). Jak jsme uvedli v kapitole 1, některé metody vyžadují rozmístění PZ a jsou založené právě na předpokladech rozdílů v signálech.

Jelikož elektretové mikrofony potřebují pro svou činnost elektrické napájení, deska modelu telefonu je opatřena patičí pro baterii. K napájení používáme univerzální mincovou baterii typu CR2032 s nominálním napětím 3 volty. Zapojení mikrofонů dále ještě obsahuje čtyři rezistory.

Z desky modelu telefonu je vyveden kabel, který je zakončený dvěma konektory Jack 6,3 mm (oba kanály mají vlastní konektor).

### **Předzesilovač a zvuková karta**

Signály z modelu telefonu zesilujeme předzesilovačem M-Audio AudioBuddy a zaznamenáváme externí zvukovou kartou Asus XONAR U3. Obě uvedená zařízení jsou zobrazena na obrázku 5.2.



Obrázek 5.2: Zvuková karta (vlevo) a předzesilovač (vpravo)

## 5.2 Kritéria vyhodnocení

Podle značení signálů zavedeného v kapitole 2.2 je zaznamenaný signál  $Y_k^{(n)}$  složený z cílového signálu  $X_k^{(n)}$  a okolního hluku  $D_k^{(n)}$ , tedy  $Y_k^{(n)} = X_k^{(n)} + D_k^{(n)}$ . Jak již bylo uvedeno, ideální CF zcela potlačuje cílový signál  $X_k^{(n)}$ , ovšem v praxi zpravidla k úplnému potlačení cílového signálu nedojde a část z něj v nahrávce zůstane. Tuto zbylou část signálu označujeme  $\overline{X_k^{(n)}}$ .

V experimentech, ve kterých testujeme potlačení cílového signálu, vyhodnocujeme potlačení jako kritérium ATS (*Attenuation of Target Signal*) podle vztahu 5.1.

$$ATS_k^{(n)} = 10 \cdot \log_{10} \left( \frac{E\{|X_k^{(n)}|^2\}}{E\{|\overline{X_k^{(n)}}|^2\}} \right) \quad (5.1)$$

V dalších experimentech, kde provádíme i potlačování hluku, vyhodnocujeme kritéria, pomocí kterých lze porovnávat kvalitu dosažených výsledků. Z objektivních kritérií se jedná o SNR (*Signal to Noise Ratio*) definované vztahem (5.2), které vyjadřuje poměr, jak je v dané nahrávce výrazný cílový signál vůči okolnímu hluku.

$$SNR_k^{(n)} = 10 \cdot \log_{10} \left( \frac{E\{|X_k^{(n)}|^2\}}{E\{|D_k^{(n)}|^2\}} \right) \quad (5.2)$$

Proces potlačování hluku ovlivňuje i cílový signál. Systém pro potlačování hluku je třeba nastavit tak, aby deformoval cílový signál co nejméně. V krajním případě může dojít k tak výrazným deformacím cílového signálu, že výsledný signál bude prakticky nepoužitelný. Pro měření deformace cílového signálu zavádíme druhé objektivní kritérium SDR (*Signal to Distortion Ratio*) zapsané ve vztahu (5.3). Definujeme ho jako poměr výsledného signálu vůči zkreslení, kde zkreslení udáváme jako vzdálenost cílového signálu od výsledného.

$$SDR_k^{(n)} = 10 \cdot \log_{10} \left( \frac{E\{|\widehat{X}_k^{(n)}|^2\}}{E\{|\widehat{X}_k^{(n)} - X_k^{(n)}|^2\}} \right) \quad (5.3)$$

Nejlepší výsledky vybrané pouze na základě objektivních kritérií SNR a SDR nemusí vždy poslechově působit jako ty nejlepší. Některé filtry, jako je například Short-term Log-MMSE popsany v kapitole 2.2.4, potlačují okolní hluk s ohledem

právě na poslechový dojem. Výsledný signál sice pak může mít horší SNR a SDR, ovšem u posluchačů může vyvolávat dojem vyšší kvality. Tento fakt je vhodné zahrnout především do systémů, které zpracovávají signály určené výhradně pro poslech (například telefonní hovory). Proto v experimentech vyhodnocujeme také kritérium OPS (*Overall Perceptual Score*) balíku PEASS verze 2.0 [22], které toto subjektivní vnímání napodobuje.

### 5.3 Nahrávky s umělou hlavou

V první sérii experimentů používáme experimentální soustavu zobrazenou na obrázku 5.3.



Obrázek 5.3: Experimentální soustava

Hlavním účelem experimentální soustavy je simulace mluvčího. K tomu slouží umělá hlava, která je vyrobena ze sádky a uvnitř osazena reproduktorem. Kabel z tohoto reproduktoru je připojen přes konektor Cinch do ovládacího zařízení připevněného v pravé části soustavy, kterým lze regulovat hlasitost signálu přehrávaného z umělé hlavy. Z ovládacího zařízení je dále vyveden kabel zakončený konektorem Jack 3,5, přes který se připojuje přehrávací zařízení.

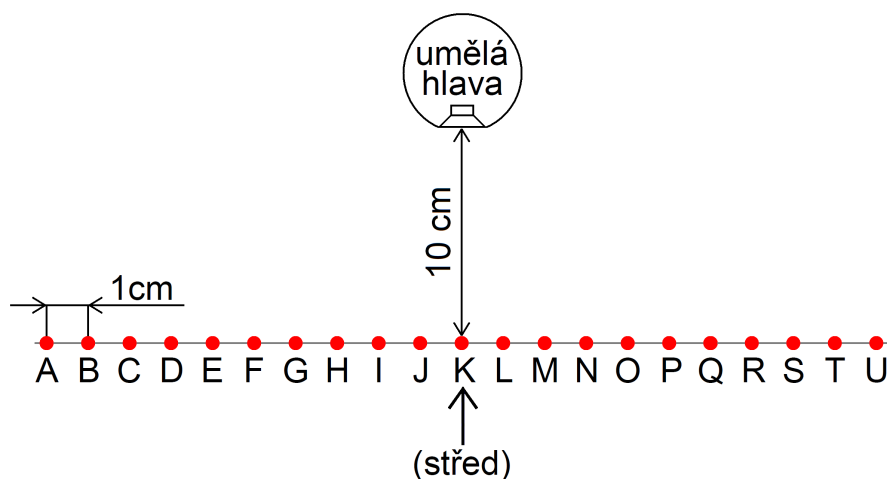
Okolo umělé hlavy je umístěna průhledná podložka, pod kterou jsme zakreslovali definované pozice sítě CF. V některých experimentech bylo nutné zopakovat

co nejpřesnější umístění modelu telefonu v dané pozici, proto jsme pro tyto experimenty uchytili model telefonu do držáku PDA, díky čemuž jsme mohli zopakovat umístění modelu telefonu v dané pozici s dostatečnou přesností.

### 5.3.1 Potlačení cílového signálu

Jak jsme již uvedli v úvodu kapitoly, experimenty potlačení cílového signálu jsme provedli pomocí bank obsahující  $CF_1$  (vztah 3.5). Definovali jsme síť pozic, podle které jsme vytvořili banky CF s různou hustotou. Nejprve zpracováváme nahrávky pořízené ve fixní pozici modelu telefonu vůči umělé hlavě. Poté testujeme potlačení v nahrávce s proměnlivou pozicí. Banky CF dále vytváříme i algoritmem učení a porovnáváme výsledky dosažené pomocí bank CF vytvořených oběma způsoby. Zaznamenávané signály obsahují řečové nahrávky z databáze TIMIT.

Schéma experimentů potlačení cílového signálu znázorňuje obrázek 5.4. Definovali jsme 21 pozic, které jsou na obrázku znázorněné červeně a označené písmeny A – U. Vzdálenost bezprostředně sousedních pozic je 1 cm. Umělá hlava je vodorovně umístěna uprostřed a svisle od pozice K vzdálena 10 cm.

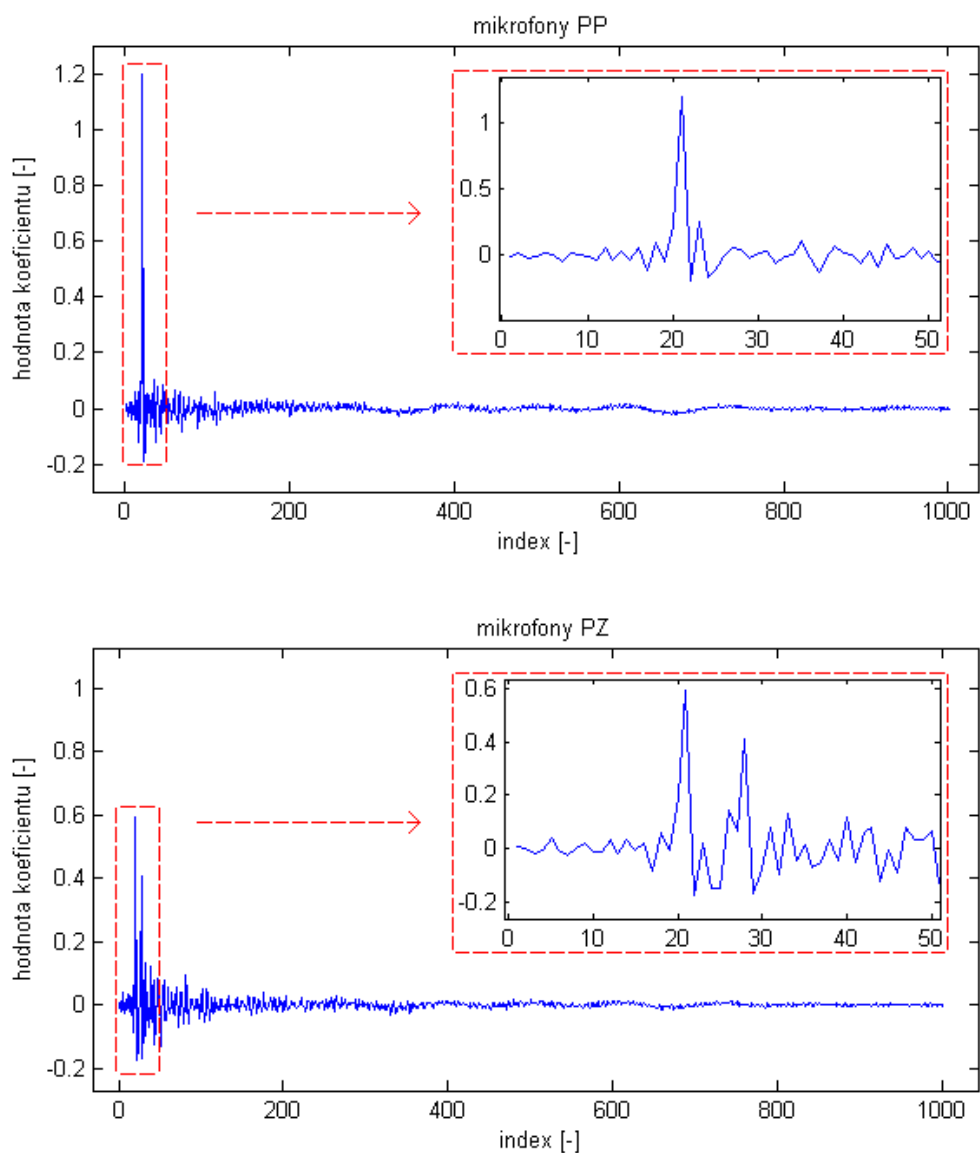


Obrázek 5.4: Schéma experimentu potlačení cílového signálu

Ve všech definovaných pozicích jsme pořídili v tichém prostředí trénovací a testovací nahrávky pro obě rozmístění mikrofونů. Model telefonu byl uchycen v držáku a během nahrávání se nijak nepohyboval. Z umělé hlavy jsme vždy přehrávali záznam mužského hlasu délky 10 sekund.

## Potlačení ve fixní pozici

Ze sady trénovacích nahrávek jsme vytvořili pro obě rozmístění mikrofonů banky CF, které nejprve obsahovaly CF pouze pro pozici  $K$  (pozice uprostřed). Koeficienty obou těchto CF jsou zobrazené na obrázku 5.5, na kterém je i detail na prvních 50 koeficientů.

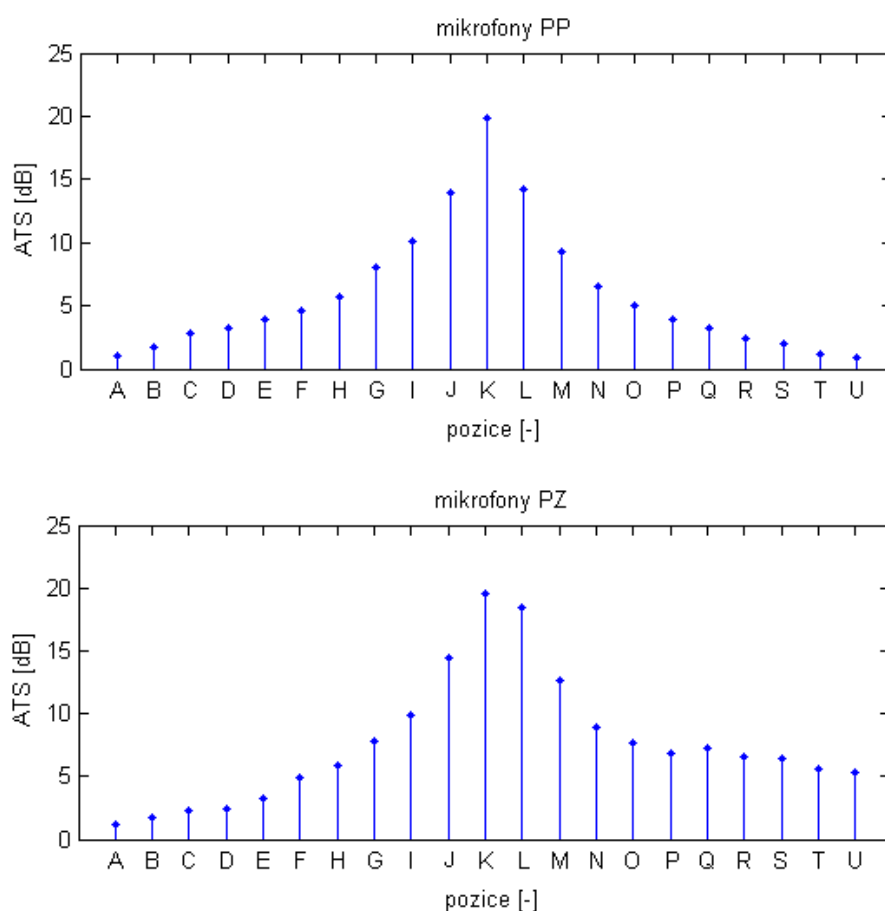


Obrázek 5.5: CF pro pozici  $K$

Jak bylo zmíněno v úvodu kapitoly, všechny CF vypočítáváme s parametrem zpoždění  $I = 20$ , proto jsou hlavní špičky obou CF na indexu 21. V místnosti, ve které byly všechny nahrávky pořizovány, nastal během přehrávání silný odraz zvuku,

což se hlavně projevilo na CF pro PZ výraznou špičkou na indexu 28. Efekt dozvuku dále způsobil, že další koeficienty CF následující po hlavních špičkách nejsou blíže k nule. Od dalších indexů je vidět, že zobrazení CF jsou zvlněná. Tato zvlnění způsobil nízkofrekvenční šum obsažený v nahrávkách použitých pro výpočet CF.

Bankami obsahující CF pro pozici  $K$  jsme zpracovali všechny testovací nahrávky a následně vyhodnotili ATS (potlačení cílového signálu) podle vztahu (5.1). Výsledky experimentu jsou zobrazeny v grafech na obrázku 5.6.



Obrázek 5.6: Graf potlačení cílového signálu ve fixní pozici bankou s jedním CF

## Zhodnocení

Z grafů je patrné, že největší potlačení cílového signálu bylo dosaženo v pozici  $K$ , tedy v pozici, pro kterou banky obsahovaly CF. V ostatních pozicích potlačení klesá, u rozmístění mikrofónů PP je pokles potlačení znatelnější než u PZ, z čehož plyne, že rozmístění mikrofónů PP je na pozici citlivější. Nejvyšší potlačení (v pozici  $K$ ) bylo u rozmístění mikrofónů PP 19,82 dB, u PZ: 19,58 dB.



Grafy by byly v ideálním případě symetrické. Z důvodu způsobu šíření zvuku po místnosti a jeho odrazů tomu tak ale není. U rozmístění mikrofónů PP nejsou rozdíly v symetrii příliš výrazné, ale u rozmístění PZ je tento jev zřetelnější právě kvůli umístění druhého mikrofónu vzadu. Zvuková vlna nedorazí k zadnímu mikrofónu přímo jako k přednímu mikrofónu. Zde se tato skutečnost navíc projevila vyšším potlačením v pravé části grafu (pozice  $L - U$ ) než v levé části (pozice  $A - J$ ).

Z tohoto experimentu je tedy zřejmé, že potlačení cílového signálu pomocí CF je vysoce citlivé na pozici. V pozici  $I$  (od pozice  $K$  vzdálena jen 2 cm) pokleslo potlačení pro rozmístění mikrofónů PP z 19,82 dB na 10,04 dB (pokles o 9,78 dB), pro PZ nastal pokles z 19,58 dB na 9,90 dB (pokles o 9,68 dB). Potlačení cílového signálu okolo 10 dB sice může být pro určité aplikace dostačující, ovšem v systému potlačování hluku je takto velký pokles nepřijatelný.

## Potlačení v proměnlivé pozici bankou definovaných pozic

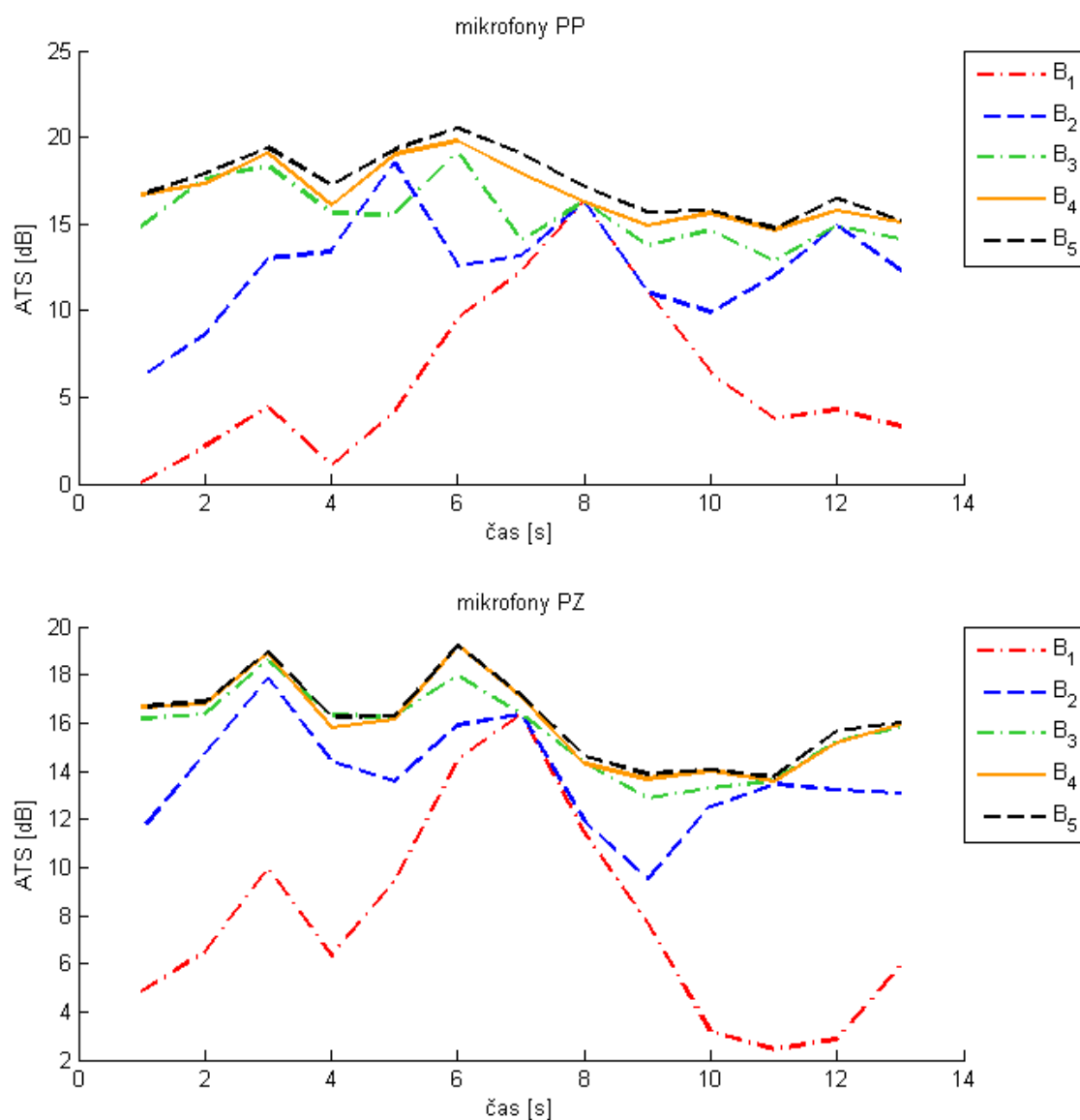
Pořídili jsme novou testovací nahrávku, ve které jsme modelem telefonu uchyceném v držáku pomalu pohybovali během nahrávání z pozice  $U$  do pozice  $A$  (rovný pohyb po úsečce). Z umělé hlavy jsme přehrávali stejný záznam jako v předchozím experimentu. Tato nahrávka délky 13 sekund tedy obsahuje záznam ze všech pozic.

Pro obě rozmístění mikrofónů jsme vytvořili několik různých bank CF, kterými jsme tuto testovací nahrávku zpracovali. Nejprve jsme použili banky popsané v tabulce 5.1 označené jako  $B_1, \dots, B_5$ , které vznikly jako různě hustá síť CF v definovaných pozicích (z trénovacích nahrávek ve fixních pozicích).

Tabulka 5.1: Banky CF (síť pozic) pro experiment potlačení cílového signálu

Banka	Obsažené pozice
$B_1$	$K$
$B_2$	$E, K, P$
$B_3$	$B, E, G, K, N, Q, T$
$B_4$	$A, C, E, H, I, K, M, O, Q, S, U$
$B_5$	všechny definované pozice ( $A - U$ )

Obrázek 5.7 zobrazuje průběh potlačení cílového signálu bankami  $B_1, \dots, B_5$ . Potlačení je vyhodnoceno na úsecích délky jedné sekundy (16000 vzorků).



Obrázek 5.7: Průběh potlačení cílového signálu bankami  $B_1, \dots, B_5$

## Zhodnocení

U zpracování bankami  $B_1$  je v obou případech vysoké potlačení pouze v krátkém okamžiku. Jak bylo popsáno v tabulce 5.1, banky  $B_1$  obsahovaly CF pouze pro pozici  $K$ . Vysokého potlačení bylo proto dosaženo právě v okamžiku, kdy se model telefonu pohyboval velmi blízko této pozice. Při rozmístění mikrofونů PZ je v první části (0 – 6 sekund) vyšší potlačení než v druhé (8 – 13 sekund). Jak již bylo

zmíněno, v testovací nahrávce jsme modelem telefonu pohybovali z pozice  $U$  do  $A$ , tedy v čase 0 – 6 sekund se model telefonu pohyboval mezi pozicemi  $U - L$ . Skutečnost, že v této fázi je vyšší potlačení než mezi pozicemi  $J - A$  koresponduje s výsledky předchozího experimentu vyhodnoceném na obrázku 5.6, kde je také vidět vyšší potlačení v oblasti  $U - L$  než v  $J - A$ .

Podobné jevy nastaly i při zpracování bankami  $B_2$ , které obsahovaly CF pro pozice  $E$ ,  $K$  a  $P$ . Zde jsou také patrné okamžiky vysokého potlačení, kdy se model telefonu pohyboval v pozicích, pro které banka obsahovala CF. U rozmístění mikrofonů PZ opět nastalo vyšší potlačení v oblasti  $U - L$  než v  $J - A$ .

Výsledky získané bankami  $B_3$  se již v mnoha okamžicích přibližují nejlepším dosaženým výsledkům. Z grafů je ale vidět, že potlačení je v některých segmentech stále nižší.

Potlačení bankami  $B_4$  a  $B_5$  se již nijak výrazně neliší. Potlačení bankami  $B_5$  je sice vyšší než bankami  $B_4$ , ale rozdíl není nijak výrazný, přestože banky  $B_5$  obsahovaly 21 CF, zatímco banky  $B_4$  jen 12 CF.

V tabulce 5.2 jsou zapsána potlačení cílového signálu vyhodnocená přes celý signál pomocí bank  $B_1, \dots, B_5$ .

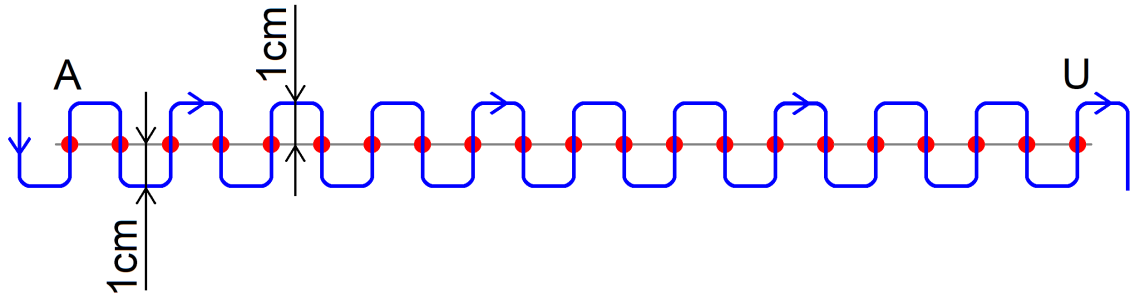
Tabulka 5.2: Celková potlačení cílového signálu bankami  $B_1, \dots, B_5$

Banka	Potlačení – mikr. PP [dB]	Potlačení – mikr. PZ [dB]
$B_1$	3,66	6,44
$B_2$	11,41	13,55
$B_3$	15,49	15,72
$B_4$	16,74	15,94
$B_5$	17,29	16,14

Opět se zde projevila vysoká závislost potlačení na pozici a CF při rozmístění mikrofonů PP byly na pozici znovu citlivější než při PZ. Je to patrné z průběhů potlačení bankami  $B_1$  a  $B_2$  na obrázku 5.7, kde u PP je průběh potlačení strmější oproti PZ. Také podle tabulky 5.2 je celkové potlačení bankami  $B_1$  a  $B_2$  vyšší u PZ než u PP. Byla-li ale banka dostatečně obsáhlá, vyššího potlačení jsme dosáhli s PP.

## Potlačení v proměnlivé pozici bankou z algoritmu učení

Dále jsme vytvářeli banky CF algoritmem učení. Tyto banky vznikly z nahrávek délky 60 sekund pořízených v tichém prostředí, ve kterých jsme modelem telefonu uchyceném v držáku pomalu pohybovali v uvažované oblasti. Vykonávaný pohyb znázorňuje obrázek 5.8. Z umělé hlavy jsme přehrávali opět stejný záznam jako v předchozích případech.



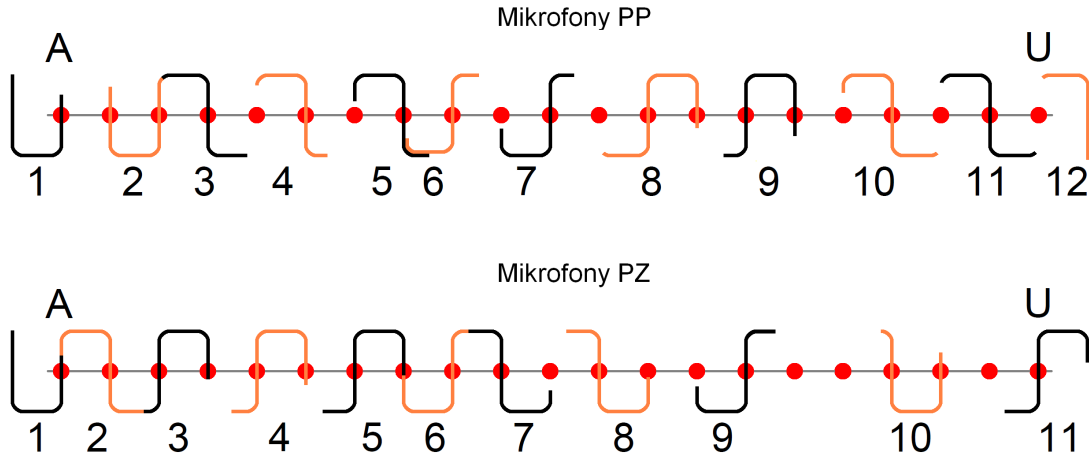
Obrázek 5.8: Pohyb modelu telefonu pro vytvoření bank CF algoritmem učení

Z těchto trénovacích nahrávek jsme vytvořili banky CF označené  $B_6, \dots, B_{13}$ , které se liší nastavením některých parametrů učení. Pro všechny banky jsme nastavili požadované potlačení  $AttReq = 15$  dB a posun na čtvrtinu délky bloku pro výpočet nového CF ( $shift = \frac{NLearn}{4}$ ). Do bank označených  $B_6, \dots, B_9$  jsme přidávali nový CF vždy, když byl účinnější než předchozí (parametr  $btr = 0$ ) a do  $B_{10}, \dots, B_{13}$  jen pokud byl nový CF účinnější alespoň o 1 dB (parametr  $btr = 1$ ). Banky  $B_6$  a  $B_{10}$  (podobně i další dvojice) se tedy liší pouze v nastavení parametru  $btr$ . Vytvořené banky a jejich parametry jsou popsány v tabulce 5.3.

Tabulka 5.3: Banky CF (algoritmus učení) pro experiment potlačení cílového signálu

Banka	$NLearn$ [s]	$WLen$ [s]	$btr$ [dB]	počet CF – PP	počet CF – PZ
$B_6; B_{10}$	4	1	1; 0	12; 11	11; 10
$B_7; B_{11}$	4	4	1; 0	17; 17	17; 14
$B_8; B_{12}$	1	$\frac{1}{4}$	1; 0	35; 28	40; 36
$B_9; B_{13}$	1	1	1; 0	43; 37	48; 40

Na obrázku 5.9 jsou znázorněné a očíslované oblasti, ve kterých vznikly CF v bankách  $B_6$ .



Obrázek 5.9: Oblasti vzniku CF během učení bank  $B_6$

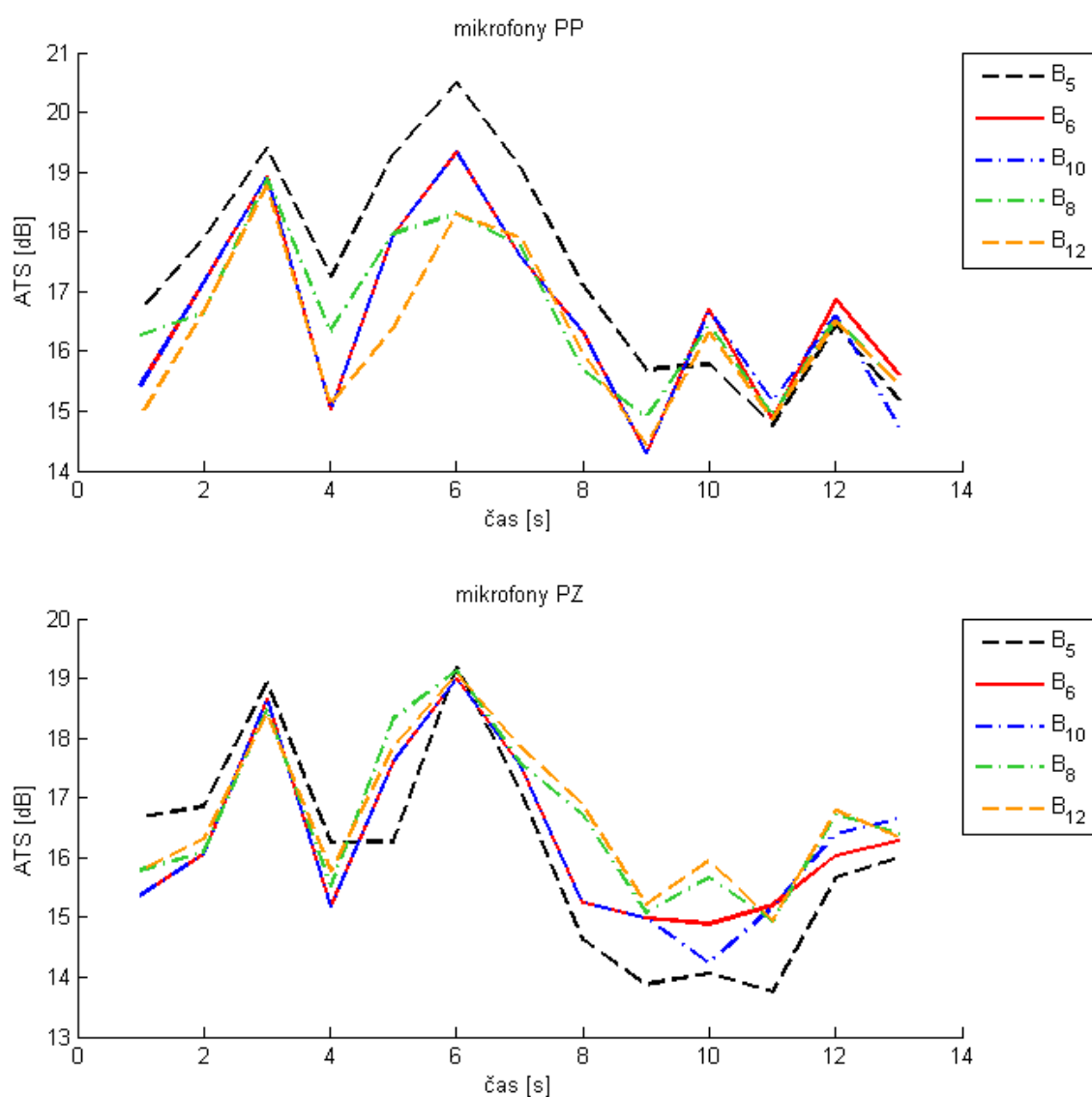
Z obrázku 5.9 je vidět, že při rozmístění mikrofónů PP jsou v bance  $B_6$  vzniklé CF relativně rovnoměrně rozmístěné po celé oblasti. Ve většině případů je mezi bezprostředně sousedními CF mezera délky 1 – 2 sekundy (3 sekundy mezi CF 9 a 10), pouze CF označené 2 a 3 vznikly hned po sobě a jediné CF 5 a 6 se v jedné sekundě překrývají. CF vypočítán z pohyblivé nahrávky tedy účinkuje i v blízkosti své oblasti. Z toho důvodu jsou mezi většinou CF zmíněné mezery, které vznikly právě v důsledku dostatečné účinnosti předchozího CF.

Banka  $B_6$  pro rozmístění mikrofónů PZ je v levé polovině (pozice  $A - J$ ) podobná bance pro PP. Častěji ale nastala situace, kdy CF vznikly za sebou bez mezery (CF 1, 2, 3 a 5, 6, 7). Pravá část (pozice  $L - U$ ) se již ale oproti PP značně liší. CF zde vznikly s mnohem většími mezerami (5 sekund mezi CF 9, 10 a 4 sekundy mezi CF 10, 11), z čehož plyne, že byly dostatečně účinné i ve vzdálenější oblasti, než ve které byly vypočítány. Tento jev opět souvisí s předchozím experimentem vyhodnoceném na obrázku 5.6, kde je vidět v případě rozmístění mikrofónů PZ vyšší potlačení v pozicích  $L - U$  než v pozicích  $A - J$ .

Podobné rozmístění CF jako u banky  $B_6$ , kde se bezprostředně sousední CF příliš často nepřekrývají, nastalo i pro banky  $B_8$ ,  $B_{10}$  a  $B_{12}$ , tedy banky, kterým jsme pro vytvoření také nastavili parametr  $WLen$  (délka bloku pro test účinnosti CF) na čtvrtinu  $NLearn$  (délka bloku pro výpočet CF). Naopak u bank  $B_7$ ,  $B_9$ ,  $B_{11}$

a  $B_{13}$  nastával překryv sousedních CF mnohem častěji, v některých případech se překrývala i více než polovina obsažených CF. Pro rozmístění mikrofonů PZ opět vznikaly mezi CF větší mezery v oblasti  $L - U$  než v  $A - J$ .

Testovací nahrávku jsme zpracovali bankami  $B_6, \dots, B_{13}$  a opět vyhodnotili ATS na úsecích délky jedné sekundy. Pro zobrazení průběhů potlačení na obrázku 5.10 jsme vybrali výsledky získané pomocí bank  $B_6, B_8, B_{10}$  a  $B_{12}$ . Výsledky potlačení také porovnáváme s bankou  $B_5$  (banka obsahující CF vypočítané z nahrávek ve fixní pozici pro všechny definované pozice  $A - U$ ), pomocí které jsme v předchozím experimentu dosáhli nejlepších výsledků.



Obrázek 5.10: Průběh potlačení cílového signálu bankami  $B_5, B_6, B_8, B_{10}$  a  $B_{12}$

## Zhodnocení

Pro rozmístění mikrofونů PP byla banka  $B_5$  až do deváté sekundy účinnější než všechny ostatní banky vytvořené algoritmem učení. Největší rozdíly ale byly okolo 2 dB. U rozmístění mikrofонů PZ jsou naopak většinou účinnější banky vytvořené algoritmem učení.

Jak bylo zapsáno v tabulce 5.3, banka  $B_{10}$  se liší oproti  $B_6$  v pravidle přidání nového CF (do  $B_6$  přidat nový CF vždy, když je účinnější než dosavadní nejúčinnější, do  $B_{10}$  přidat, pokud je účinnější alespoň o 1 dB). Toto pravidlo se jednak projevilo v počtu CF ( $B_6$  obsahovala pro PP 12 CF, pro PZ 11 CF a  $B_{10}$  obsahovala pro PP 11 CF, pro PZ 10 CF) a i v průběhu potlačení. Rozdíly nastaly až po deváté sekundě, přičemž do této chvíle byly průběhy totožné. Zřetelnější rozdíl nastal při rozmístění mikrofонů PZ, kde mezi devátou až jedenáctou sekundou je účinnější  $B_6$ , ale od jedenácté sekundy má vyšší potlačení  $B_{10}$ . Procesy vytváření bank  $B_6$  a  $B_{10}$  byly do určité chvíle zcela stejné, proto jsou i průběhy potlačení zčásti totožné. Rozdíl nastal ve chvíli, kdy do banky  $B_6$  byl přidán nový CF, který byl účinnější pouze o 0,1 dB než předchozí nejúčinnější. Tento nepatrný rozdíl pak změnil další průběh učení bank a při zpracování testovací nahrávky způsobil v desáté sekundě u banky  $B_{10}$  pokles potlačení oproti  $B_6$  okolo 1 dB. Od jedenácté sekundy byla ale o cca 0,5 dB účinnější banka  $B_{10}$ .

Rozdíly jsou také zřetelné ve zpracování bankami  $B_8$  a  $B_{12}$ . U rozmístění mikrofонů PP je  $B_8$  účinnější téměř vždy.  $B_8$  zde ale obsahovala 35 CF a  $B_{12}$  28 CF. Jiná situace nastala u rozmístění mikrofонů PZ. Potlačení bankou  $B_{12}$  je většinou vyšší než bankou  $B_8$  a segmentech, kde je  $B_8$  účinnější, jsou rozdíly v potlačení v řádu desetin dB, přestože  $B_8$  obsahovala pro PZ 36 CF a  $B_{12}$  40 CF. Pravidlo přidání nového CF, pokud je účinnější alespoň o 1 dB, tedy způsobilo, že výsledná banka obsahuje nejen méně CF, ale i celkově poskytuje vyšší potlačení.

Pro přehlednost grafů na obrázku 5.10 jsme nezobrazili průběhy potlačení bankami  $B_7$ ,  $B_{11}$ ,  $B_8$  a  $B_{12}$ . Výsledky potlačení pomocí bank  $B_7$  a  $B_{11}$  byly v porovnání s ostatními výsledky průměrné, ale výsledky získané pomocí bank  $B_9$  a  $B_{13}$  byly v porovnání s ostatními nejhorší, přestože tyto banky byly ohledně počtu CF nejobsáhlejší. Maximální rozdíly v průbězích potlačení ale oproti průměrným nepřesahovaly 1 dB.

Potlačení cílového signálu vyhodnocená přes celý signál pomocí bank  $B_5, \dots, B_{13}$  jsou zapsána v tabulce 5.4.

Tabulka 5.4: Celková potlačení cílového signálu bankami  $B_5, \dots, B_{13}$

Banka	Potlačení – PP [dB]	Potlačení – PZ [dB]
$B_5$	17,29	16,14
$B_6; B_{10}$	16,44; 16,39	16,19; 16,18
$B_7; B_{11}$	16,39; 16,39	16,36; 16,36
$B_8; B_{12}$	16,65; 16,10	16,57; 16,64
$B_9; B_{13}$	16,23; 16,22	16,12; 16,12

Rozdíly v potlačení cílového signálu v tabulce 5.4 nejsou nijak výrazné, přesto jsme ve většině případů dosáhli vyššího potlačení při rozmístění mikrofonů PP než u PZ. Je ale nutné podotknout, že zpracovávané nahrávky obsahovaly pouze cílový signál bez okolního hluku. Jak bude vidět v následujícím experimentu, pro potlačování hluku bylo výhodnější rozmístění mikrofonů PZ.

Nelze zcela jednoznačně určit nejlepší nastavení pro vytváření bank CF algoritmem učení. Banky  $B_9$  a  $B_{13}$  ukázaly, že vyšší počet CF nemusí vždy vést k lepším výsledkům. Podle bank  $B_6$  s  $B_{10}$  a  $B_8$  s  $B_{12}$  pro PZ není ani jednoznačné, zda je výhodnější přidat nový CF vždy, když poskytne jen o trochu vyšší potlačení než předchozí nejúčinnější CF, nebo jestli přidávat nový CF jedině pokud bude mít vyšší potlačení alespoň o definovanou konstantu například 1 dB.

Učení bank CF také výrazně ovlivnil parametr  $NLearn$  (délka bloku pro výpočet CF). Při nastavení  $NLearn = 1$  s vznikaly banky značně obsáhlejší než při  $NLearn = 4$  s, přičemž rozdíly v účinnostech nebyly nijak zásadní. Zajímavé rozdíly nastávaly také s parametrem  $WLen$  (délka bloku pro test účinnosti CF). Ukázalo se, že pokud jsme nastavili  $WLen = \frac{NLearn}{4}$ , vznikaly banky kvalitnější než při  $WLen = NLearn$ . Tento parametr ztelně ovlivňoval průběhy učení bank, což následně vedlo i ke změnám účinnosti bank.

Všechny parametry učení ovlivňují kvalitu i počet CF ve výsledné bance. Pro použití banky CF v reálném čase je třeba, aby banka nebyla příliš obsáhlá. Pokud vznikne banka s příliš mnoha CF, lze některé sice odstranit například druhým



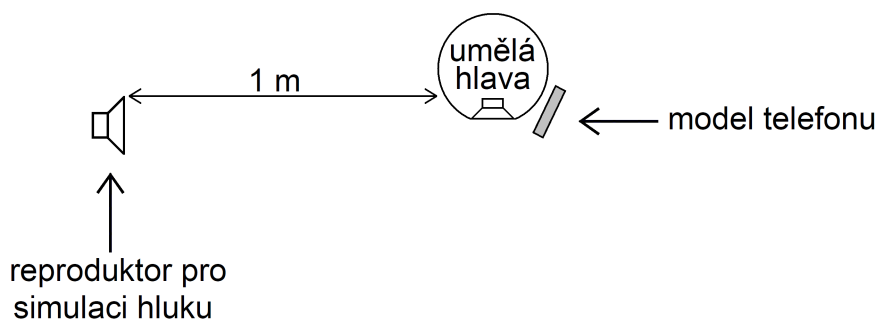
průchodem banky zmíněném v kapitole 4.2.2, ovšem výhodnější by bylo takové nastavení, které produkuje samo o sobě banku CF dostatečně účinnou a ne příliš obsáhlou. Z testovaných bank má k této vlastnosti nejblíže nastavení pro banku  $B_6$ , tedy  $AttReq = 15$  dB,  $NLearn = 4$  s,  $WLen = 1$  s,  $shift = 1$  s,  $btr = 0$ . V tomto experimentu se zdá toto nastavení jako nejvýhodnější, ovšem v jiných situacích a s použitím jiných zařízení tomu obecně tak vždy být nemusí.

Banky CF vytvořené algoritmem učení jsou ale dostatečně účinné a v případě rozmístění mikrofonů PZ dokonce i účinnější než obsáhlejší banky vytvořené jako síť pozic. Nejúčinnější banky  $B_5$  vytvořené jako síť obsahovaly 21 CF, zatímco banka  $B_6$  pro PP 12 CF (o 9 CF méně) a pro PZ 11 CF (o 10 CF méně).

### 5.3.2 Potlačení hluku

Provedli jsme další experimenty, které se více přiblížily reálným situacím. Modelem telefonu jsme nyní pohybovali okolo umělé hlavy v oblasti jako při telefonním hovoru. Zpracovávané signály navíc obsahovaly okolní hluk. Experimenty provádíme opět pro obě rozmístění mikrofonů a navíc srovnáváme výsledky dosažené pomocí bank obsahující  $CF_1$  a  $CF_2$ . K porovnání výsledků vyhodnocujeme kritéria SNR, SDR a OPS uvedené v kapitole 5.2.

Pro simulaci okolního hluku používáme reproduktor, který jsme umístili 1 metr od experimentální soustavy, jak znázorňuje schéma experimentu na obrázku 5.11.



Obrázek 5.11: Schéma experimentu potlačení hluku

Cílové signály obsahovaly opět řečové nahrávky z databáze TIMIT, jako okolní hluk jsme používali hluk restaurace, hluk provozu na ulici a také řečové nahrávky z databáze TIMIT jiných mluvčích, než kteří představovali mluvčí v trénovacích

a testovacích nahrávkách. Hluk restaurace a hluk provozu na ulici pochází z [23]. Testovací nahrávky cílových signálů a signálů s hlukem jsme pořídili zvlášť a před zpracováním je míchali v poměru  $-15$  dB až  $+15$  dB s krokem  $2,5$  dB. Tento poměr vyjadřuje parametr počáteční SNR (pSNR). Všechny testovací nahrávky byly délky 10 sekund.

Jelikož předchozí experiment ukázal, že banka CF vytvořená algoritmem učení je dostatečně účinná, vytvářeli jsme zde banky CF již jen tímto způsobem. Simulujeme tím i situaci, kdy si uživatel přizpůsobuje na sebe banku CF zcela sám. Pro vytvoření bank CF jsme v tichém prostředí pořídili trénovací nahrávky délky 60 sekund pro mužský a ženský hlas. Během nahrávání jsme model telefonu drželi v ruce a nahodile jím pohybovali v oblasti pravděpodobných pozic vůči umělé hlavě. Nejednalo se tedy o pohyb po předem stanovené dráze jako v předchozím experimentu.

Parametry učení jsme nastavili stejně jako pro banku  $B_6$  z předchozího experimentu ( $AttReq = 15$  dB,  $NLearn = 4$  s,  $WLen = 1$  s,  $shift = 1$  s,  $btr = 0$ ). Banky obsahující  $CF_2$  jsme přizpůsobili na hluk restaurace, kde  $\epsilon = 0,5$  (vztah 3.8).

Potlačení hluku jsme provedli pomocí Wienerova filtru (vztah 2.33). Filtry pro potlačení hluku lze doplnit o parametr  $\tau$ , který vyjadřuje, jak intenzivně se potlačení hluku provede. Obvykle se jedná o kladné reálné číslo, kterým se odhadnutý hluk zesílí. Vztah pro Wienerův filtr rozšířený o parametr  $\tau$  je zapsán ve vztahu (5.4). Pro tyto experimenty jsme při použití  $CF_1$  nastavili  $\tau = 30$ , pro  $CF_2$   $\tau = 20$ .

$$H_k = \frac{\max[0, |Y_k|^2 - \tau \lambda_d(k)]}{|Y_k|^2 + \epsilon}. \quad (5.4)$$

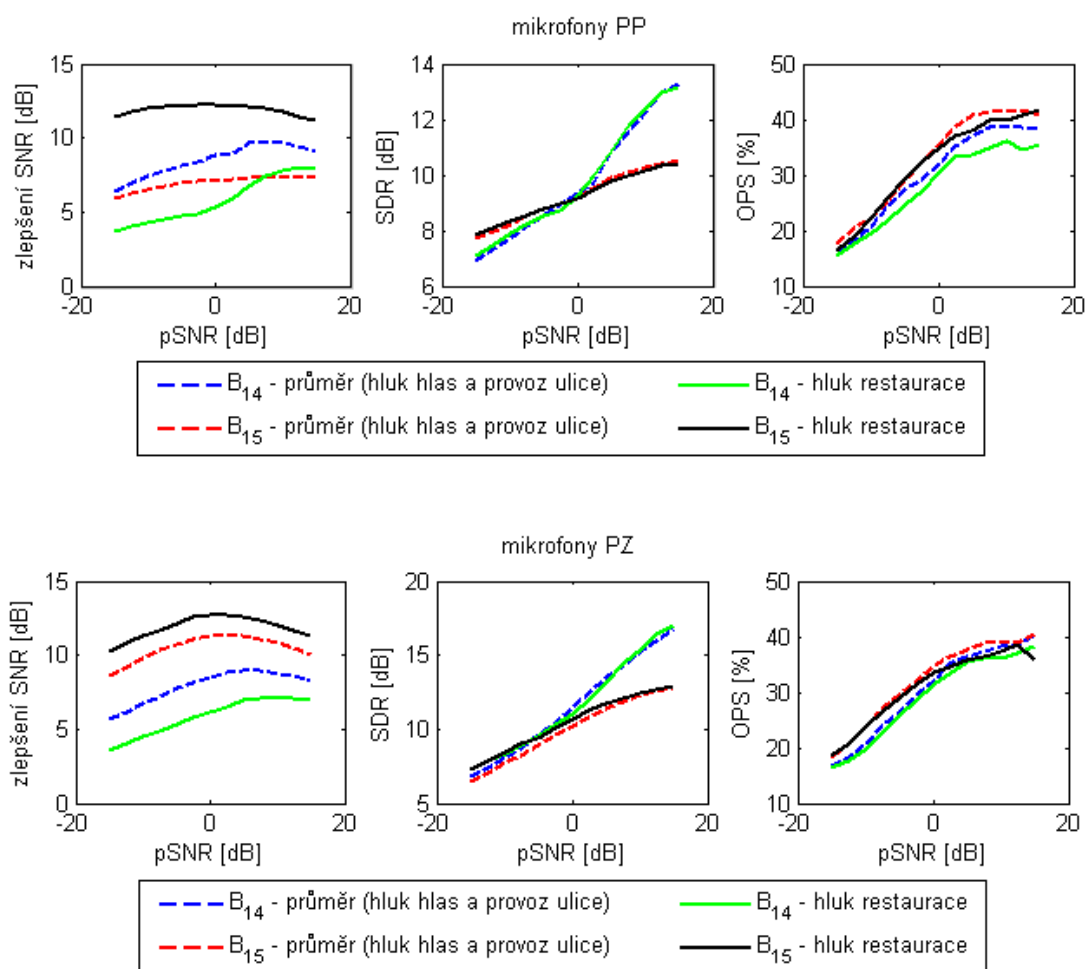
Z trénovacích nahrávek jsme vytvořili dohromady 8 bank CF (mužský a ženský hlas, rozmístění mikrofónů PP a PZ,  $CF_1$  a  $CF_2$ ). V tabulce 5.5 jsou zapsány počty CF pro banky vytvořené z trénovacích nahrávek mužského hlasu.

Tabulka 5.5: Banky CF v experimentu potlačení hluku na experimentální soustavě

Banka	Typ CF	Počet CF – PP	Počet CF – PZ
$B_{14}$	$CF_1$	16	9
$B_{15}$	$CF_2$	9	7

Ke zpracování testovacích nahrávek jsme použili banky CF odvozené z trénovacích nahrávek obou mluvčích. Testujeme i situace, kdy pro potlačení cílového signálu používáme „nesprávnou“ banku CF (banku odvozenou z trénovací nahrávky druhého mluvčího).

Výsledky zde uvedené jsou pro případ, kdy jsme použili banky popsané v tabulce 5.5 (banky pro mužský hlas). Porovnali jsme výsledky, ve kterých jsme těmito bankami zpracovali nahrávky obsahující jako cílový signál mužský hlas (mluvčí, pro kterého jsou banky přizpůsobené) s výsledky, kde cílový signál představoval ženský hlas. Jelikož rozdíly nebyly nijak výrazné, prezentujeme zde tu náročnější situaci, ve které bankami CF odvozenými pro mužský hlas zpracováváme nahrávky obsahující jako cílový signál ženský hlas. Vyhodnocení experimentu je zobrazeno v grafech na obrázku 5.12. Ukázky k tomuto experimentu jsou k dispozici na přiloženém CD.



Obrázek 5.12: Vyhodnocení experimentu potlačení hluku na experimentální soustavě

## Zhodnocení

Průběhy znázorněné červeně a modře představují průměrné hodnoty pro dvě situace, kde signál s hlukem obsahoval řečovou nahrávku a poté hluk provozu na ulici. Výsledky pro tato dvě nastavení byly podobné, proto pro přehlednost grafů je nezobrazujeme zvlášť, ale uvádíme jejich průměr. Zeleně a černě jsou zobrazeny průběhy pro jedno konkrétní nastavení, kde okolní hluk představoval hluk restaurace.

Ve všech případech jsme dosáhli zlepšení SNR, přičemž výraznější deformace cílového signálu vyjádřené kritériem SDR nastaly jen v těch nejnáročnějších situacích při velmi nízkém pSNR. Z grafů je ale vidět, že systém poskytuje určité zlepšení pro všechny uvažované situace, tedy i pro tu pro nejnáročnější, kdy pSNR bylo  $-15$  dB (hluk mnohem výraznější než cílový signál).

Nejkvalitnějších výsledků jsme dosáhli s bankami obsahující  $CF_2$  (banky  $B_{15}$ ), pokud byl ve zpracovávané nahrávce hluk restaurace. V ostatních situacích, obsahujících jako okolní hluk jiný signál, jsou výsledky horší. Zhoršení je zřetelnější u rozmístění mikrofónů PP. Jak jsme uvedli v popisu experimentu,  $CF_2$  jsme přizpůsobovali na hluk restaurace a z toho důvodu jsme zde dosáhli lepších výsledků právě v případech, kde okolní hluk představoval hluk restaurace.

Zlepšení SNR za použití bank obsahující  $CF_1$  (banky  $B_{14}$ ) je sice nižší než při použití bank s  $CF_2$ , nicméně lze tyto výsledky stále považovat za úspěšné. V situacích, kde bylo pSNR  $> 0$ , poskytly navíc banky obsahující  $CF_1$  lepší SDR (menší deformace cílového signálu) než banky obsahující  $CF_2$ . Pomocí bank s  $CF_2$  jsme zase ale dosáhli ve většině případů lepších hodnot perceptuálního kritéria OPS.

Hodnoty perceptuálního kritéria OPS jsou pro náročné situace pSNR  $< 0$  relativně nízké. Znamená to tedy, že výsledné signály nemusí poslechové působit příliš příjemně, přestože jsme v některých situacích dosáhli velmi výrazného zlepšení SNR.

Jelikož je většina průběhů relativně strmá, neuvádíme zde průměrné hodnoty kritérií vyhodnocené přes všechny pSNR.

Jak již bylo zmíněno, uvedené výsledky jsou pro nastavení, ve kterém jsme použili banku CF odvozenou pro mužský hlas, ale cílové signály ve zpracovávaných nahrávkách obsahovaly ženský hlas. Z tohoto experimentu je tedy zřejmé, že při simulaci mluvčího umělou hlavou (reproduktorem) není nezbytně nutné pro dosažení kvalitních výsledků použít banku CF odvozenou ze záznamu hlasu daného mluvčího.

## 5.4 Nahrávky skutečných osob

V následujících experimentech simulujeme již reálnou situaci telefonního hovoru. Pro záznam zvuku používáme stejná zařízení jako v předchozích experimentech, ale cílové zdroje nyní představují skutečné osoby (uvažujeme zde dvě testovací osoby – muže a ženu). V tichém prostředí jsme pořídili minutové trénovací nahrávky od obou osob pro vytvoření bank CF algoritmem učení a testovací nahrávky délky 10 sekund. Ve všech nahrávkách držely osoby model telefonu ve své ruce a nahodile jím pohybovaly v oblasti, ve které se mobilní telefon při běžném hovoru vyskytuje.

Použili jsme stejné nastavení parametrů učení bank CF jako v předchozím experimentu. Banky obsahující  $CF_2$  jsme přizpůsobili na řečové nahrávky, kde parametr  $\epsilon = 0,3$  (rovnice 3.8). Vytvořené banky CF jsou popsány v tabulce 5.6.

Tabulka 5.6: Banky CF skutečných osob

Banka	Osoba	Typ CF	Počet CF – PP	Počet CF – PZ
$B_{16}$	Muž	$CF_1$	14	7
$B_{17}$	Muž	$CF_2$	10	7
$B_{18}$	Žena	$CF_1$	22	5
$B_{19}$	Žena	$CF_2$	17	10

### 5.4.1 Potlačení cílového signálu

Jako v kapitole 5.3.1 jsme nejprve provedli experiment potlačení cílového signálu bankami obsahující  $CF_1$ . Testovací nahrávky obou osob jsme zpracovali bankami popsanými v tabulce 5.6 a i bankami  $B_{14}$  z předchozího experimentu, které jsme vytvořili z nahrávek umělé hlavy obsahující mužský hlas.

V tabulce 5.7 jsou zapsána potlačení cílových signálů vyhodnocená přes celý signál pomocí zmíněných bank.

Tabulka 5.7: Celková potlačení bankami CF pro skutečné osoby a umělou hlavu

Scénář	Banka	Test. nahr.	Potl. – PP [dB]	Potl. – PZ [dB]
$S_1$	$B_{16}$ (muž)	muž	15,20	17,81
$S_2$	$B_{16}$ (muž)	žena	13,28	15,51
$S_3$	$B_{18}$ (žena)	žena	14,40	19,92
$S_4$	$B_{18}$ (žena)	muž	7,12	13,00
$S_5$	$B_{14}$ (um. hlava)	muž	2,61	6,24
$S_6$	$B_{14}$ (um. hlava)	žena	3,48	11,75

## Zhodnocení

Nejlepší výsledky nastaly podle předpokladů pro scénáře  $S_1$  a  $S_3$ , ve kterých jsme pro potlačení cílového signálu použili banky CF odvozené pro osobu v testovací nahrávce. V ostatních případech nastal v potlačení znatelný pokles. S použitím nesprávné banky nejlépe fungoval scénář  $S_2$  (banka – muž, testovací nahrávka – žena), kde pro rozmístění mikrofonů PP pokleslo potlačení o 1,92 dB, pro PZ o 2,3 dB. Scénář  $S_2$  by naznačoval, že banka odvozená pro jiného mluvčího nezpůsobí nijak výrazný pokles potlačení, ovšem opačném případě provedeném v rámci scénáře  $S_4$  (banka – žena, testovací nahrávka – muž) pokleslo potlačení u PP o 7,28 dB, u PZ o 6,92.

Největší pokles potlačení ale nastal při scénářích  $S_5$  a  $S_6$ , kde byla použita banka CF pro umělou hlavu. Ve scénáři  $S_5$  (testovací nahrávka – muž) nastal pokles potlačení pro PP 12,59 dB, pro PZ 11,57 dB, ve scénáři  $S_6$  (testovací nahrávka – žena) pokleslo potlačení u PP o 9,28 dB, u PZ 8,17 dB. Scénář  $S_6$  navíc dopadl hůře než  $S_5$ , přestože použité banky  $B_{14}$  odvozené z nahrávek umělé hlavy obsahovaly mužský hlas.

Negativní vliv na potlačení cílového signálu při použití nesprávné banky CF je z tohoto experimentu zřejmý. Pro dosažení nejlepších výsledků je nezbytně nutné použít banku CF odvozenou pro konkrétní osobu. Jak bylo vidět, použití nevhodné banky CF sice odvozené pro skutečnou osobu, ovšem jinou, může způsobit velmi výrazné zhoršení kvality referenčního signálu.

Ve výsledcích oproti předchozím experimentům potlačení cílového signálu je zde ještě navíc rozdíl, že vyššího potlačení jsme nyní dosáhli pro rozmístění mikrofonů PZ než pro PP.

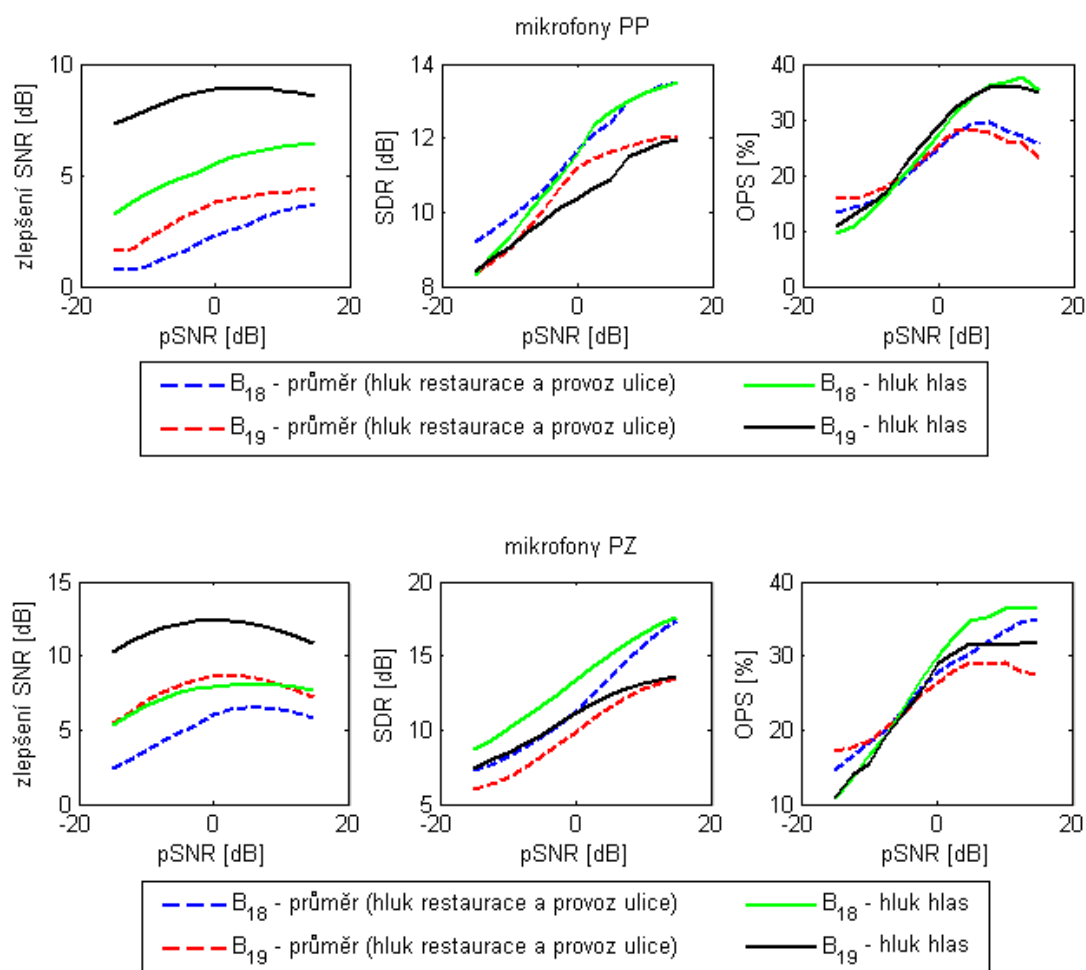
### 5.4.2 Potlačení hluku

Podobně jako v kapitole 5.3.2 jsme provedli experimenty potlačení hluku. Tyto experimenty měly podobné schéma (banky obsahující  $CF_1$  a  $CF_2$ , tři různé signály s hlukem, reproduktor pro simulaci hluku vzdálený od osob jeden metr, míchání cílových signálů se signály s hlukem v poměru  $-15$  dB až  $15$  dB s krokem  $2,5$  dB atd). Pořídili jsme nové nahrávky hluku, které obsahovaly stejné záznamy jako v experimentu v kapitole 5.3.2, ale nyní pro simulaci reálnějších podmínek držela osoba model telefonu během nahrávání jako při hovoru.

K získávání referenčního signálu jsme použili banky již uvedené v tabulce 5.6. Jak bylo vidět z předchozího experimentu, použitím nesprávné banky CF se při zpracovávání nahrávek skutečných osob může kvalita referenčního signálu velmi výrazně zhoršit. Prezентujeme zde tedy výsledky pro situaci, ve které jsme použili banku CF odvozenou pro stejnou osobu jako v testovacích nahrávkách.

Potlačení referenčního signálu provádíme opět pomocí Wienerova filtru (5.4), kde jsme nyní nastavili pro banky obsahující  $CF_1$  i  $CF_2$   $\tau = 20$ .

Vyhodnocení tohoto experimentu (testovací osoba žena) je zobrazeno v grafech na obrázku 5.13. Jelikož jsme nyní  $CF_2$  přizpůsobovali na řečovou nahrávku, uvádíme zvlášť průběhy výsledků pro nastavení obsahující jako hluk řeč a pro přehlednost grafů zobrazujeme průměry pro další dvě nastavení obsahující hluk restaurace a hluk provozu na ulici. Ukázky z tohoto experimentu jsou k dispozici na příloženém CD.



Obrázek 5.13: Vyhodnocení experimentu potlačení hluku se skutečnými osobami

## Zhodnocení

Výsledky jsou podobné jako u experimentu v kapitole 5.3.2, kde jsme pro simulaci mluvčího používali umělou hlavu. Určitého zlepšení jsme opět dosáhli ve všech uvažovaných situacích včetně těch nejnáročnějších s  $p\text{SNR} = -15$  dB. Nejvyšší zlepšení SNR poskytly banky obsahující  $\text{CF}_2$ , pokud hluk obsahoval řečovou nahrávku. Jako v experimentech v kapitole 5.3.2 jsme tedy i zde dosáhli vyššího zlepšení pro situaci obsahující hluk, na který jsme  $\text{CF}_2$  přizpůsobovali. Zde navíc většině situací poskytly banky obsahující  $\text{CF}_1$  lepší SDR než banky s  $\text{CF}_2$ .

Ukázali jsme tedy, že navržený systém redukce hluku poskytuje možnosti dosažení kvalitních výsledků i v situacích, kde mluvčí představují skutečné osoby. Tyto výsledky jsou velmi důležité pro úvahy o nasazování systému v praxi.



## 6 Demonstrační aplikace

Demonstrační aplikaci systému redukce hluku popsaného v kapitole 4 jsme nazvali *Noise Cancellor* a naimplementovali pro PC v jazyce Java SE (Standard Edition). Během vývoje jsme používali *Java Development Kit* (JDK) verze 7 od společnosti *Oracle* s vývojovým prostředím *Eclipse IDE* verze *Kepler*.

Vyvinutá aplikace umožňuje provádět redukci hluku v režimu online i offline. Při režimu online jsou zpracovávána data zaznamenaná ze zvukové karty, zatímco při režimu offline je vstupem již existující soubor typu *wav*. Aplikace dále obsahuje modul pro vytváření bank CF algoritmem učení popsaném v kapitole 4.2.2. Nejsou-li k dispozici potřebné nahrávky pro vytvoření banky CF, je možné přepnout aplikaci do dalšího režimu, který je určený pouze pro nahrávání zvuku a potřebné nahrávky zde pořídit.

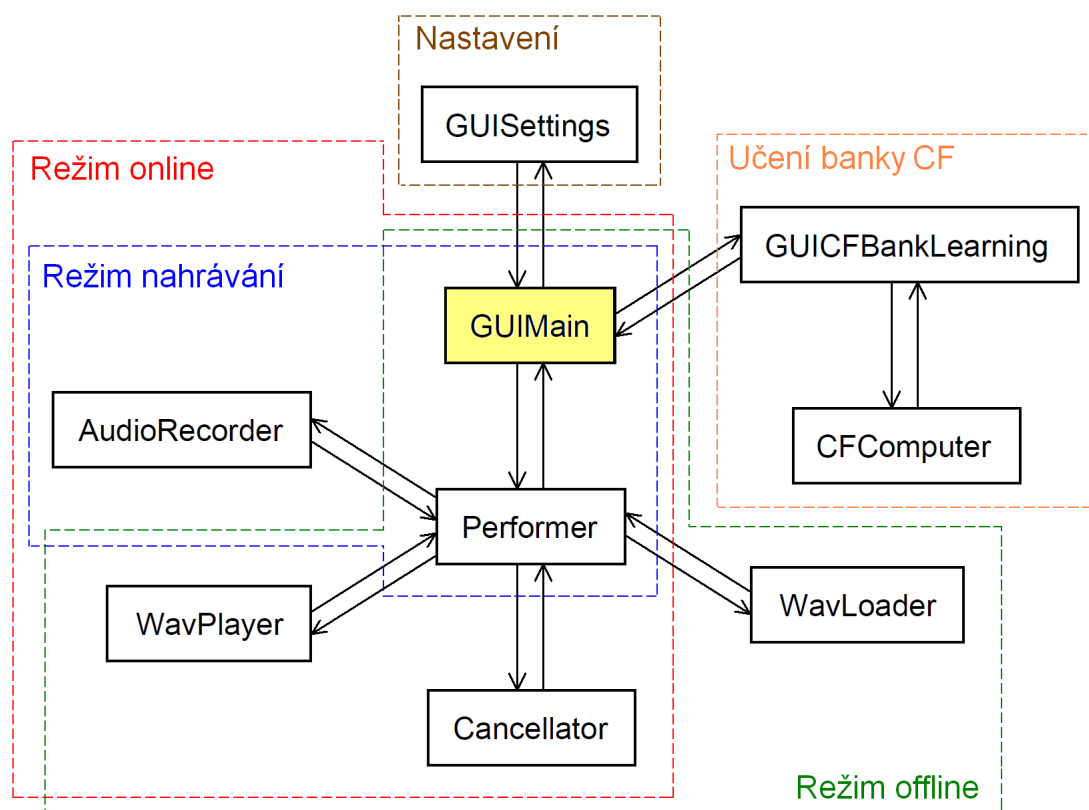
K dalším funkcím aplikace patří možnost přehrávání původního signálu, získaného referenčního signálu (odhad okolního hluku) i výsledného signálu s potlačeným hlukem. Přehrávání je možné v režimu online i offline.

Ovládání aplikace se provádí přes grafické uživatelské rozhraní (GUI), které jsme vytvořili pomocí knihovny *Swing* [24].

Podobně jako v experimentech pořizujeme a zpracováváme všechny signály vždy se vzorkovací frekvencí 16 kHz a 16 bity na vzorek. Do souborů typu *wav* ukládáme signály podle specifikace RIFF definované společnostmi IBM a Microsoft [25].

## 6.1 Struktura aplikace

Pro aplikaci *Noise Cancellator* jsme vytvořili sedmnáct tříd, které jsme rozdělili do dvou balíčků. Předávání dat mezi třídami jsme realizovali podle návrhového vzoru *Observer*, kde třídy generují definované události (např. data zpracována, nastala chyba atd.) a jiné třídy na ně reagují. Strukturu aplikace zobrazuje blokové schéma hlavních tříd na obrázku 6.1, kde je znázorněno i jejich použití.



Obrázek 6.1: Schéma struktury aplikace

Ovládání aplikace se provádí prostřednictvím třídy *GUIMain*, která po spuštění vytvoří a zobrazí hlavní okno GUI. Zde lze mimo jiné přepínat mezi pracovními režimy, otevřít okno s nastavením a i spustit průvodce učením nové banky CF.

Nastavení všech parametrů se provádí v samostatném okně, které vytváří a obsluhuje třída *GUISettings*. Zde lze zvolit metodu a parametry pro potlačování hluku, zadat pracovní složku a názvy výstupních souborů, vybrat audio zařízení pro nahrávání a přehrávání a i načíst do paměti banku CF. Po potvrzení změn se zadané hodnoty předají třídě *GUIMain*.

Jsou-li splněny všechny nutné podmínky (zmíněné v následujících kapitolách) pro běh programu, zadané nastavení se předá třídě *Performer*, která ovládá ostatní části aplikace a zajišťuje i ukládání signálů na místo zadané v nastavení. V režimu online a režimu nahrávání jsou vstupní data získávána ze zvukové karty, jejichž nahrávání provádí třída *AudioRecorder*. V případě režimu online jsou nahraná data následně předána třídě *Cancellator*, která podle zadaných parametrů provádí samostatné potlačování hluku. Původní signál, referenční signál a signál s potlačeným hlukem je možné v režimu online přehrávat pomocí třídy *WavPlayer*. Je-li zvolen režim offline, zadaný vstupní soubor typu *wav* je načítán prostřednictvím třídy *WavLoader*. Načtená data jsou poté předána třídě *Cancellator*. Signály lze v režimu offline také přehrávat, opět pomocí třídy *WavPlayer*.

Třída *GUICFBankLearning* vytváří nové okno průvodce učení banky CF, který v několika krocích provede uživatele celým procesem učení. Zadané parametry jsou pak předány třídě *CFComputer*, která učení banky CF provede. O průběhu je uživatel informován v samostatném okně.

Seznam všech tříd v aplikaci a jejich rozdělení do balíčků je znázorněno na obrázku 6.2. V balíčku *mainControl* jsou obsaženy třídy pro ovládání a v balíčku *spFunctions* ostatní a obecně použitelné třídy.

mainControl	spFunctions
GUICFbankLearning	AudioRecorder
GUIMain	Cancellator
GUIQuickTuning	CFBank
GUISettings	CFComputer
Main	FFT
Performer	GF
	MatrixOp
	ProgressDialog
	Settings
	WavLoader
	WavPlayer

Obrázek 6.2: Balíčky a třídy v aplikaci

Kromě hlavních tříd zobrazených na obrázku 6.1 aplikace dále obsahuje třídu *GUIQuickTunning* (okno pro snadné ladění parametru  $i$  za běhu), *MatrixOp* (implementace operací s maticemi používané během výpočtu koeficientů CF), *FFT* [26] (rychlá Fourierova transformace), *Filters* (reprezentace banky CF), *Settings* (třída obsahující parametry nastavení), *ProgressDialog* (výpis informací o postupu výpočtu v samostatném okně), *GF* (ostatní metody jako je například kontrola hlavičky souboru *wav*, metoda pro výběr souboru z disku, příprava grafických komponent nebo i definice konstant) a samozřejmě i třídu *Main* obsahující metodu *main*, která se volá při spuštění aplikace.

## 6.2 Formát souboru s bankou CF

Banky CF ukládáme do binárních souborů, pro které jsme definovali vlastní formát popsáný v tabulce 6.1. Ve sloupci velikost uvádíme i názvy datových typů používaných v jazyce Java.

Tabulka 6.1: Formát binárního souboru s bankou CF

Název	Velikost	Popis
Endian	1 B (byte)	znak „b“ pro Big Endian, „l“ pro Little Endian
Typ CF	1 B (byte)	„0“ pro $CF_1$ , „1“ pro $CF_2$
Délka CF	2 B (short)	délka CF v časové oblasti
Délka FFT	2 B (short)	délka FFT obrazu CF (ozn. $N$ )
Počet CF	2 B (short)	počet CF obsažených v bance (ozn. $C$ )
Zpoždění CF	1 B (byte)	parametr $I$ ze vztahů (3.5) nebo (3.8)
Koeficienty	$n^*$ (double)	koeficienty FFT obrazů všech CF

\* Koeficienty FFT obrazů CF jsou komplexní čísla. Každé komplexní číslo je uloženo v 16 bajtech (8 bajtů (double) pro reálnou část a 8 bajtů (double) pro imaginární část). Ukládáme ale pouze prvních  $\frac{N}{2} + 1$  koeficientů obrazů FFT, protože ostatní koeficienty jsou k nim komplexně sdružené. Koeficienty tedy budou v souboru zabírat  $2 \cdot 8 \cdot C \cdot \left(\frac{N}{2} + 1\right)$  bajtů pro  $CF_1$ , v případě  $CF_2$  dvakrát více, kde jsou vždy po sobě uloženy dvojice FFT obrazů vektorů  $\mathbf{g}_L$  a  $\mathbf{g}_P$  (vztah 3.8).

## 6.3 Grafické uživatelské rozhraní

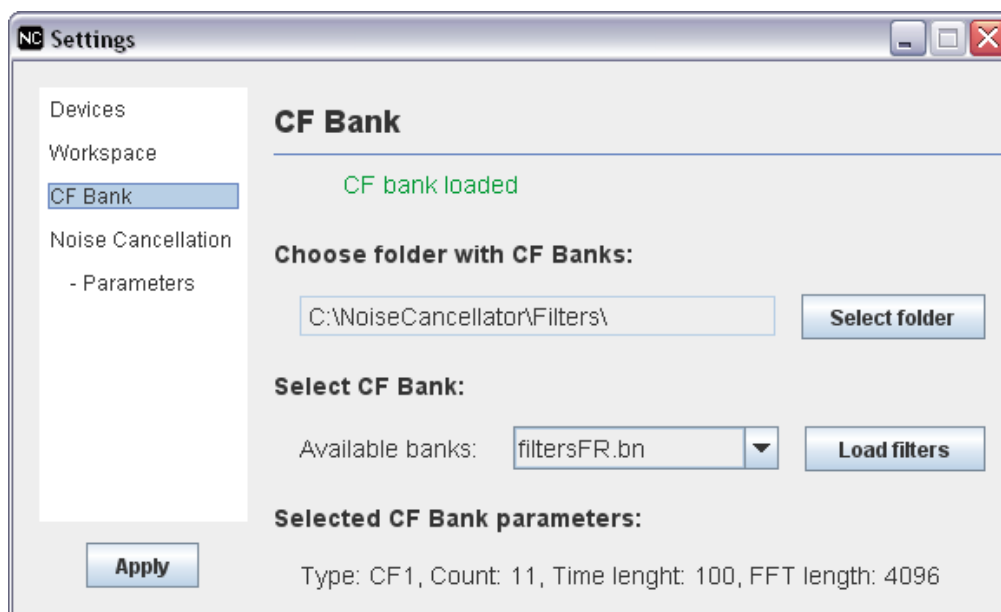
Jak již bylo zmíněno, ovládání aplikace se provádí přes grafické uživatelské rozhraní. V horní liště je umístěna hlavní nabídka, jejíž struktura s významem jednotlivých položek je znázorněna na obrázku 6.3.



Obrázek 6.3: Struktura hlavní nabídky aplikace

### 6.3.1 Nastavení

Screenshot okna nastavení je na obrázku 6.4, kde jsou zobrazeny volby banky CF.



Obrázek 6.4: Screenshot nastavení aplikace

Nastavení je rozděleno do skupin zobrazených v levé části okna a obsahuje položky

- Devices – výběr z dostupných audio zařízení pro nahrávání a přehrávání. V nabídce se zobrazí pouze ta zařízení, která podporují uvažované parametry (vzorkovací frekvence 16 kHz, 2 kanály, 16 bitů na vzorek).
- Workspace – nastavení pracovní složky a názvů souborů pro uložení původního signálu, referenčního signálu a signálu s potlačeným hlukem.
- CF Bank – výběr a načtení banky CF do paměti. Ve zvolené složce se vyberou soubory s příponou *.bn*, jejichž názvy se přidají do nabídky dostupných bank. Pro zvolenou banku se v dolní části vypíše informace z hlavičky souboru. Samotné načtení banky se provede stisknutím příslušného tlačítka. Načítání se provádí v samostatném vlákne a jeho průběh je graficky zobrazován pomocí prvku *JProgressBar* (grafický ukazatel průběhu zobrazující procentuální část již provedených operací) v horní části okna. Proběhne-li načtení banky CF úspěšně, vypíše se zeleně v horní části okna *CF bank loaded*, jinak je červeně vypísáno *no CF bank loaded*.
- Noise Cancellation – výběr metody potlačování hluku. V možnostech je Wienerův filtr (kapitola 2.2.1), adaptivní normalizovaný LMS filtr (kapitola 2.2.2) a JMAP SAE (kapitola 2.2.3).
- Parameters – parametry pro potlačování hluku
  - Time shift – posun v časové oblasti (parametr pro blokové zpracování).
  - Tau – parametr  $\tau$  udávající jakou intenzitou se provede potlačení referenčního signálu (použití  $\tau$  jsme uvedli v (5.4) pro Wienerův filtr).
  - Cut-off frequency, Value – frekvence, od které se omezuje potlačování hluku (jak jsme zmínili v kapitole 4.1, k výpočtu filtru, který se aplikuje na původní signál, přidáváme podmínku limitující minimální hodnotu na stanovenou mez určenou hodnotou Value).
  - Voice Activity Detector Limit, mi – mezní hodnota detektoru řečové aktivity (VAD) a parametr  $\mu$  ze vztahu (2.24). Volba těchto parametrů je možná jen v případě, je-li zvolena metoda potlačování hluku adaptivní normalizovaný LMS filtr. Pokud VAD rozhodne, že je hlas neaktivní, provádí se adaptace filtru. Tento VAD pracuje na jednoduchém principu,

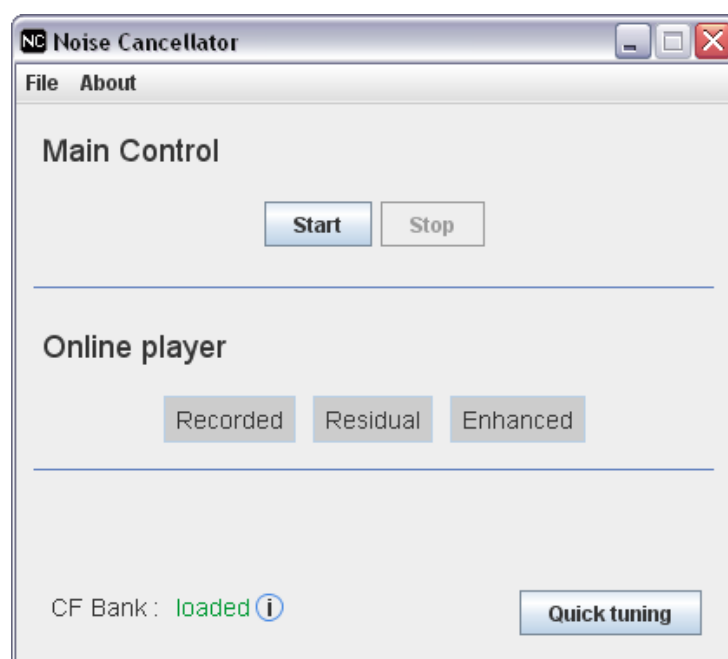
kde porovnává hlasitost vstupního signálu s nastavenou mezní hodnotou (předpokládáme, že je-li hlas aktivní, výrazně stoupne hlasitost, protože mluvčí je blízko mikrofonů).

Všechna textová pole mají nastavenou metodu pro kontrolu obsahu, která reaguje na změnu textu. Jsou-li zadány neplatné hodnoty (např. prázdný řetězec, písmena, kde jsou očekávána čísla nebo desetinné číslo v místě, kde má být číslo celé atd.), pozadí pole zčervená. V takové situaci nelze potvrdit změny ani zobrazit jinou skupinu nastavení. Pokud uživatel zvolí v nabídce jinou skupinu nastavení, zobrazí se chybové okno s výzvou k opravě hodnot ve zvýrazněných polích.

Potvrzení změn se provede tlačítkem *Apply*. Než dojde k zavření okna, zkontrolují se všechny zadané hodnoty. Nenastane-li nikde chyba, obsahy textových polí se převedou na příslušné proměnné. Dále se otestuje existence zadaných složek a pokud neexistují, aplikace se je pokusí vytvořit. V případě neúspěchu (například z důvodu nedostatečných práv k zápisu) je uživatel informován opět chybovým oknem.

### 6.3.2 Režim online

Screenshot aplikace v režimu online je na obrázku 6.5.

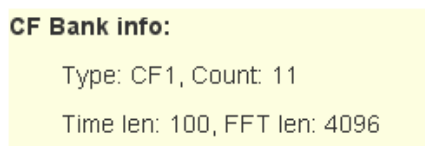


Obrázek 6.5: Screenshot aplikace v režimu online

V horní části okna se nacházejí hlavní ovládací prvky – tlačítka *Start* (spuštění zpracování) a *Stop* (ukončení zpracování). Tlačítko *Start* se zpřístupní tehdy, jsou-li splněny všechny nutné podmínky. Jedná se o načtení do paměti banky CF, výběr audio zařízení pro nahrávání a přehrávání a zadání všech parametrů popsanych v nastavení (kapitola 6.3.1).

Uprostřed je umístěný přehrávač pro režim online, který je zpřístupněný pouze za běhu (po stisknutí tlačítka *Start*), pokud se podařilo vytvořit instanci třídy *WavPlayer* se zadaným zařízením pro přehrávání. Je-li zvolené audio zařízení nedostupné nebo nastane jiná chyba, přehrávač zůstane nepřístupný a uživatel je o této chybě informován chybovým oknem. V opačném případě lze přehrávání spustit, aktivní přehrávání je znázorněné zelenou barvou pozadí tlačítka a neaktivní červenou barvou.

Ve spodní části okna jsou informace o bance CF. Není-li ještě načtena do paměti, je pouze červeně vypsáno *no CF Bank loaded*. Pokud v nastavení proběhne načtení banky CF úspěšně, nápis se změní na *CF Bank loaded* zelenou barvou a přibude navíc možnost vypsání informace o načtené bance CF, které se zobrazí kliknutím na ⓘ. Výpis informací o bance CF (typ CF, počet, délka v časové oblasti a délka FFT obrazu) zobrazený na obrázku 6.6 se pak objeví v hlavním okně vpravo dole.



**CF Bank info:**  
Type: CF1, Count: 11  
Time len: 100, FFT len: 4096

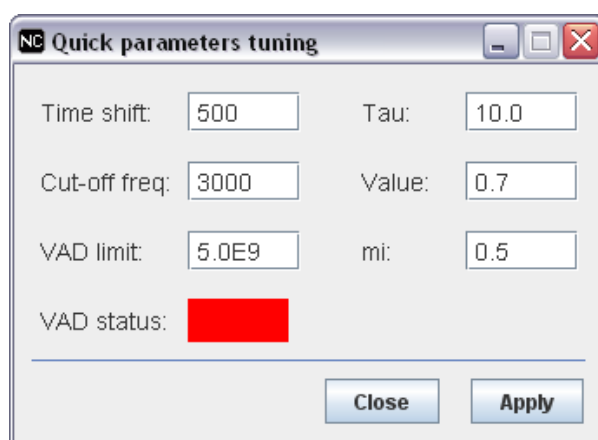
Obrázek 6.6: Výpis informací o bance CF

Vpravo dole je ještě umístěno tlačítko *Quick tuning*, které zobrazí okno snadného ladění parametrů pro potlačování hluku. Zde jsou zobrazeny jen ty parametry, které náleží zvolené metodě potlačování hluku. Kromě parametru časový posun (*Time shift*) lze ostatní parametry měnit i za běhu. Změny nastanou až po stisknutí tlačítka *Apply*. Podobně jako u okna s nastavením (kapitola 6.3.1) mají všechna textová pole kontrolu obsahu. Je-li v některém poli zadána neplatná hodnota, pozadí pole zčervená a pokud se stiskne tlačítko *Apply*, ke změnám parametrů nedojde. Také se opět zobrazí chybové informační okno.



Některé parametry lze tedy měnit z okna nastavení i z okna snadného ladění. Pokud jsou obě tato okna zobrazena a v jednom okně dojde ke změně parametrů, změny se projeví i v druhém okně.

Screenshot okna pro snadné ladění je na obrázku 6.7, kde jako metoda potlačování hluku je zvolen adaptivní LMS filtr. Pro adaptivní LMS filtr je v tomto okně navíc položka *VAD Status*, která signalizuje stav VAD (zelená – hlas aktivní, červená – hlas neaktivní). Pomocí této položky lze snadno experimentálně hledat vhodnou prahovou hodnotu pro VAD, na které velmi závisí správná funkce adaptivního LMS filtru.



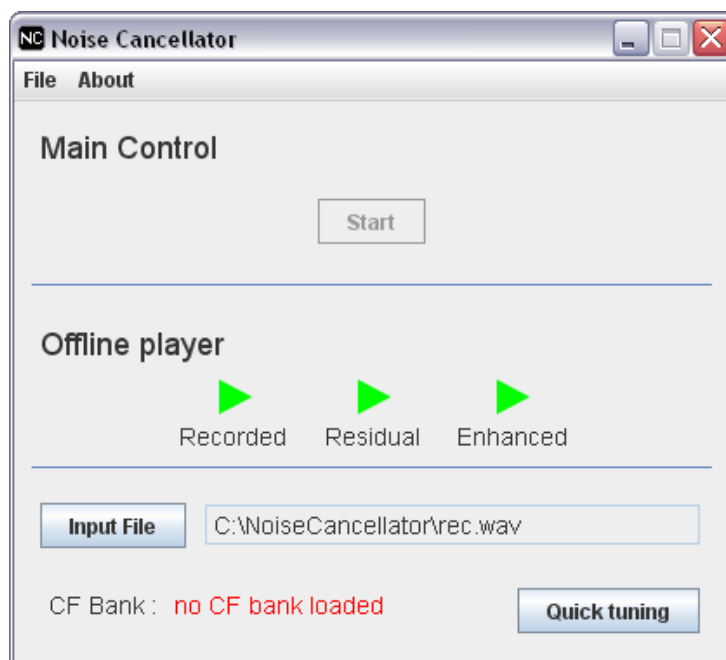
Obrázek 6.7: Screenshot okna snadného ladění parametrů

Stisknutím tlačítka *Start* v hlavním okně se předá nastavení instanci třídy *Performer* a poté se spustí zpracování v režimu online, jehož postup byl již popsán v kapitole 6.1.

Po stisknutí tlačítka *Stop* se ukončí činnost instance třídy *Performer*, která před ukončením zpracování zapíše hlavičky do výstupních souborů typu *wav* (hlavička obsahuje informaci o velikosti, která při režimu online není dopředu známá). Dále se zastaví přehrávání, je-li aktivní, a celý přehrávač režimu online se znepřístupní. Nakonec se povolí tlačítko *Start* pro opětovné spuštění.

### 6.3.3 Režim offline

Screenshot aplikace v režimu offline je na obrázku 6.8.



Obrázek 6.8: Screenshot aplikace v režimu offline

V tomto režimu jsou vstupní data předem známá. Proces potlačování hluku skončí po zpracování celé vstupní nahrávky, proto je v hlavním ovládání pouze tlačítko *Start* (spuštění zpracování), které je opět přístupné až po splnění všech nutných podmínek. Kromě podmínek jako při zpracování v režimu online zde musí být navíc zadán vstupní soubor.

Cesta ke vstupnímu souboru je zadávána ve spodní části okna, kde jsou také vypisovány informace o bance CF. Výpis informací o bance CF je stejný jako v režimu online. Je zde opět také tlačítko pro zobrazení okna snadného ladění parametrů.

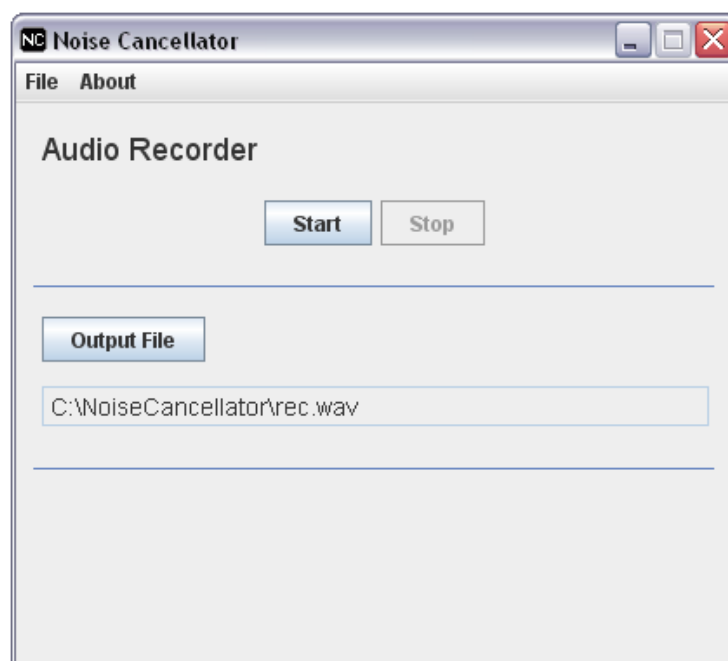
Uprostřed okna je umístěný přehrávač pro režim offline. Při přepnutí do režimu offline se provede kontrola existence zadaných souborů (původní signál, referenční signál a výstupní signál s potlačeným hlukem). Pokud soubory existují, jsou v přehrávači u jejich popisků zelené trojúhelníky, kterými se spouští přehrávání. V případě, že daný soubor neexistuje, je trojúhelník vykreslený šedě a na kliknutí nereaguje. Existuje-li daný soubor a uživatel spustí jeho přehrávání, před samotným přehráváním se znovu provede kontrola, zda soubor stále existuje, a pokud ano,

zkontroluje se jeho hlavička. Soubor musí mít v hlavičce všechny údaje správně, jedná se například o vzorkovací frekvenci (16 kHz), počet kanálů nebo i zda velikost souboru zapsaná v hlavičce koresponduje se skutečnou velikostí souboru. Během přehrávání se změní zelený trojúhelník na červený čtverec, kterým lze přehrávání zastavit.

Po stisknutí tlačítka *Start* se zkontroluje vstupní soubor stejně jako před přehráváním. Poté se předá nastavení instanci třídy *Performer*, která zadaný soubor zpracuje postupem popsaném v kapitole 6.1. O postupu zpracování je uživatel informován pomocí prvku *JProgressBar*, který se po dobu zpracování zobrazuje vpravo nahoře. Během zpracování se dále zneprístupní přehrávač režimu offline a po dokončení se opět povolí tlačítko *Start*.

### 6.3.4 Režim nahrávání

Screenshot aplikace v režimu nahrávání je na obrázku 6.9.

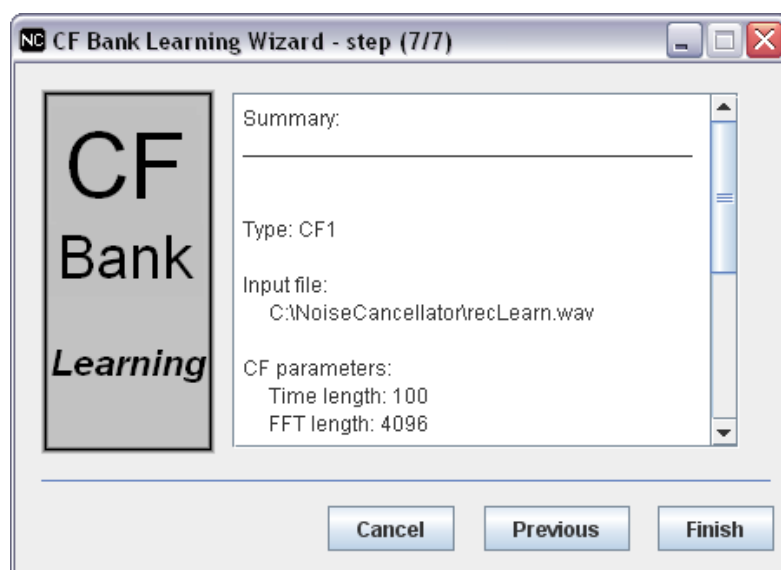


Obrázek 6.9: Screenshot aplikace v režimu nahrávání

Jak jsme již uvedli, režim nahrávání je v aplikaci obsažený především pro pořizování nahrávek určených k vytváření nových bank CF algoritmem učení. Ovládání zde obsahuje pouze tlačítka *Start* (spuštění nahrávání), *Stop* (ukončení nahrávání) a prvky pro zadání cesty k výstupnímu souboru.

### 6.3.5 Učení banky CF

Pro vytvoření nové banky CF algoritmem učení obsahuje aplikace průvodce, který uživateli celý tento proces usnadní. V několika krocích se zadají všechny potřebné parametry a po zpracování celé nahrávky se nová banka CF uloží na zadané místo. Vytvořenou banku je dále možné předat do paměti aplikace (novou banku lze tak použít bez nutnosti načítání přes okno nastavení). Screenshot průvodce učením nové banky CF je na obrázku 6.10, kde je zobrazen poslední krok – souhrn nastavení.



Obrázek 6.10: Screenshot průvodce učením nové banky CF

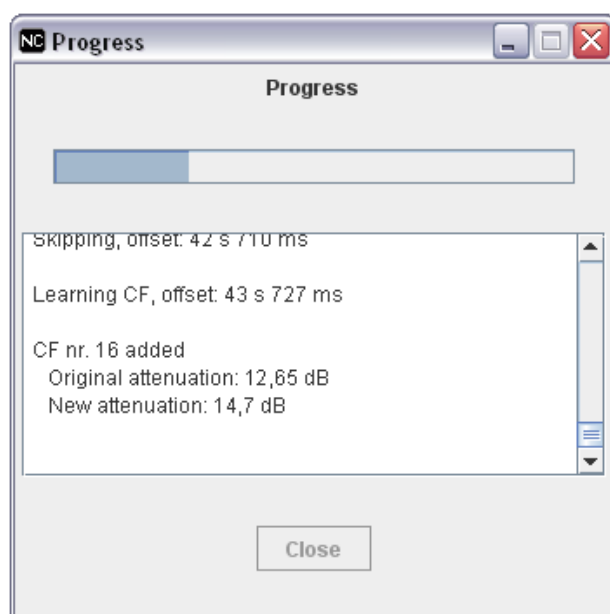
Kroky průvodce:

- Krok 1 – úvodní informace
- Krok 2 – volba typu CF ( $CF_1$  nebo  $CF_2$ )
- Krok 3 – zadání vstupní nahrávky, v případě  $CF_2$  i zadání nahrávky s hlukem
- Krok 4 – volba parametrů CF (délka v časové oblasti, délka FFT obrazů, u  $CF_2$  parametr  $\epsilon$  ze vztahu 3.8)

- Krok 5 – volba parametrů učení (*AttReq*, *NLearn*, *WLen*, *shift*, *btr*), které jsme popsali v kapitole 4.2.2
- Krok 6 – zadání výstupního souboru pro novou banku CF, možnost volby ukládání do souboru výstupů CF s nejvyšším potlačením a zadání cesty k tomuto souboru
- Krok 7 – souhrn nastavení

Během jednotlivých kroků se navíc provádí kontrola zadaných parametrů, například v kroku 3 – kontrola existence souborů a jejich hlaviček, v kroku 5 musí být zadáno *AttReq* a *btr* jako číslo atd. Pokud nastane chyba, průvodce neumožní pokračovat dál. U kroku 6 se ještě otestuje, zda zadaný výstupní soubor pro novou banku CF již existuje. Pokud ano, lze se vrátit a cestu změnit, nebo zvolit přepsání starého souboru.

Po spuštění učení banky CF se okno průvodce zavře a zobrazí se nové okno vytvořené pomocí třídy *ProgressDialog*, kam se vypisují informace o postupu. Screenshot tohoto okna je na obrázku 6.11.



Obrázek 6.11: Screenshot okna pro výpis informací o postupu učení banky CF

Postup výpočtů je graficky zobrazován pomocí prvku *JProgressBar* v horní části a pomocí textových zpráv uprostřed okna. Okno lze zavřít až po skončení všech operací.

## 7 Závěr

Navržený systém redukce hluku dosahuje při splnění určitých podmínek kvalitních výsledků. V provedených experimentech docházelo k určitému potlačení hluku při všech uvažovaných situacích, tedy i těch nejnáročnějších s počátečním SNR  $-15$  dB, přičemž deformace cílového signálu byly v přijatelných mezích. Jelikož je navržený systém určený pro mobilní telefony, příznivé výsledky experimentů uvažující jako cílové zdroje skutečné osoby jsou velmi důležité. Reálné situace se totiž mohou značně lišit od těch experimentálních, které používají reproduktory k simulacím mluvcích.

Systém poskytuje kvalitní výsledky pro obě uvažovaná rozmístění mikrofونů. O něco lepších výsledků potlačování hluku jsme však dosahovali s variantou rozmístění mikrofонů přední a zadní.

Hlavní slabina systému spočívá ve vysoké závislosti používané banky CF na konkrétním uživateli. Jak bylo vidět z experimentů, při použití nesprávné banky CF může dojít k výraznému zhoršení kvality referenčního signálu. Zmíněné zhoršení nastane v důsledku nedostatečného potlačení cílového signálu, jehož výrazná část se pak objeví v referenčním signálu. Ovšem i s použitím velmi kvalitní banky CF se malá část cílového signálu dostane do referenčního signálu a tento jev pak způsobuje při fázi samotného potlačení hluku nejvíce zkreslení a vznik různých artefaktů. Je-li ale banka CF dostatečně kvalitní, zkreslení ve výsledném signálu jsou zanedbatelná nebo alespoň přijatelná.

Banka CF by měla být přizpůsobena konkrétnímu uživateli, což pro použití systému v praxi nepředstavuje problém díky algoritmu učení navrženému v kapitole 4.2.2. Nelze ale po uživateli vyžadovat, aby si parametry učení banky CF nastavovali sami. Parametry učení by tedy bylo nutné přednastavit. Pravděpodobně ale nebude existovat takové nastavení, které by bylo možné označit ve všech ohledech jako to nejlepší. Určité nejednoznačnosti nejlepšího nastavení parametrů se projeví

i během našich experimentů. Před nasazením v praxi by bylo třeba provést s daným mobilním telefonem rozsáhlé série testů zahrnující různé scénáře a různé testovací osoby a na základě výsledků pak vybrat kompromis zohledňující všechny požadavky na banku CF.

Navržený systém má malé výpočetní nároky a pro současné mobilní telefony nepředstavuje ani z tohoto hlediska potíže. Vyvinutou demonstrační aplikaci (v jazyce Java) by bylo možné dále optimalizovat na výpočetní náročnost, například k proměnným přistupovat globálně, dynamicky alokovat paměť jen v těch nejnutnějších případech, produkovat co nejméně nepoužívaných objektů (aby příliš nezasahoval *garbage collector*) atd. Optimalizace tohoto typu jsme provedli jen v těch nejkritičtějších částech aplikace. Pokud by se ale jednalo o skutečné nasazení v mobilních telefonech, bylo by vhodné do telefonů implementovat specializovaný obvod vyhrazený jen pro činnost systému redukce hluku.

V rámci experimentů jsme prováděli pokusy i se dvěma skutečnými mobilními telefony. Jednalo se o Sony Ericsson Xperia Mini a Samsung Galaxy S3, oba s operačním systémem Android. Jelikož se aplikace pro Android programují v jazyce Java, použili jsme ve zpracování stejnou část kódu jako pro demonstrační aplikaci popsanou v kapitole 6. Zatímco v aplikaci pro PC potlačovala banka CF cílový signál spolehlivě, v případě telefonu Sony Ericsson Xperia Mini byly výsledky potlačení značně nahodilé. Obvykle se do referenčního signálu dostala podstatná část cílového signálu, přestože jsme testovali mnoho bank CF vytvořené různými způsoby, které ovšem v aplikaci pro PC vždy fungovaly správně. Ve skutečných telefonech mohou být implementované různé nelineární operace (například automatická regulace hlasitosti vstupního signálu), které mají ovšem na činnost CF velmi nepříznivé účinky. Špatná funkce banky CF naznačuje, že vstupní signál je nějakým způsobem nejprve předzpracován a není tak možné se z úrovně aplikace běžící pod operačním systémem dostat k signálům přímo z mikrofónů. V případě telefonu Samsung Galaxy S3 se nám ani nepodařilo pořídit dvoukanálovou nahrávku, přestože jsme použili stejný kód jako pro druhý telefon. Zde se tedy pravděpodobně jednalo o situaci, kdy telefon má sice dva mikrofony, ale programově dostupný je až výsledek zpracování. Z těchto důvodů jsme se experimenty se skutečnými telefony dále již nezabývali.

Systém je možné dále rozšiřovat. Jelikož jednu z důležitých částí (vytvoření banky CF) si provádí uživatel sám, je třeba zohlednit s tím spojená rizika. Každou vytvořenou banku CF by bylo třeba ohodnotit, zda je dostatečně kvalitní, což ovšem nemusí být hned po vytvoření banky proveditelné. Proto by bylo třeba do systému implementovat postupy, které by rozhodly, zda systém funguje správně a má tak smysl na výstup poslat výsledný signál s potlačeným hlukem, nebo jestli je výhodnější použít jako výstup původní signál z mikrofونů. Potom by také bylo možné adaptivně nastavovat parametr  $\tau$  vyjadřující intenzitu potlačení referenčního signálu v původní nahrávce.

Jednou z dalších možných vylepšení jsou adaptivní úpravy banky CF během zpracování. Zde by mohla být například banka vytvořená algoritmem učení neměnná a k ní by se adaptivně přidávaly nebo odebíraly CF vytvořené během hovoru. Bylo by ale nutné správně a spolehlivě rozhodovat, zda je aktuální část nahrávky vhodná pro výpočet CF a i vyřešit možné situace, kdy majitel telefonu půjčí svůj telefon někomu jinému.

Jak již bylo zmíněno, navržený systém umožňuje dosáhnout kvalitního potlačování hluku s nízkými výpočetními nároky. Další výhody systému spočívají v různých možnostech rozmístění mikrofونů a i ve schopnosti systému potlačit hluk v signálech obsahující okolní hluk značně výrazněji než cílový signál, aniž by došlo k závažným deformacím na cílovém signálu. Před implementací do skutečných telefonů by bylo ještě vhodné vyřešit uvedené možnosti vylepšení. Systém má jistě velký potenciál pro použití nejen v mobilních telefonech.



## A Obsah přiloženého CD

**/Demo Aplikace/** – Demonstrační aplikace popsaná v kapitole 6

**/NoiseCancellor/** – Zdrojové kódy (projekt z *Eclipse IDE* verze *Kepler*)

**noiseCancellor.jar** – spustitelný soubor

**/Experimenty/** – ukázky experimentů potlačení hluku

**/Skutecna osoba/** – ukázky experimentu z kapitoly 5.4.2 (skutečná osoba)

**/Umela hlava/** – ukázky experimentu z kapitoly 5.3.2 (umělá hlava)

(Všechny ukázky jsou rozdělené do složek podle počátečního SNR)

**diplomova\_prace.pdf** – práce v elektronické podobě

## Literatura

- [1] M. Jeub, C. Herglotz, C. M. Nelke, C. Beaugeant, P. Vary, *Noise Reduction for Dual-Microphone Mobile Phones Exploiting Power Level Differences*, Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), pp. 1693–1696, Mar. 2012.
- [2] Z. Fu, F. Fan, J. Huang, *Dual-microphone noise reduction for mobile phone application*, Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), pp. 7239–7243, May. 2013.
- [3] E. Habets, J. Benesty, and J. Chen, *Multi-microphone noise reduction using interchannel and interframe correlations*, Proc. of International Conference on Acoustic, Speech, and Signal Processing (ICASSP 2012), pp. 305–308, Mar. 2012.
- [4] C. Nelke, C. Beaugeant, P. Vary, *Dual microphone noise PSD estimation for mobile phones in hands-free position exploiting the coherence and speech presence probability*, Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), pp. 7279–7283, May 2013.
- [5] Z. Zhang , M. Etoh, *ICA-based Noise Reduction for Mobile Phone Speech Communication*, Proc. of 16th International Conference on Computer Communications and Networks (ICCCN 2007), pp. 470–473, Aug. 2007.
- [6] I. Tashev, *Sound Capture and Processing: Practical Approaches*, John Wiley & Sons Ltd., 2009.
- [7] Z. Koldovský, *Biologické a akustické signály (BSI), studijní materiály – Poznámky k předmětu* [online]. [vid. 6. 12. 2013].  
Dostupné z: <http://itakura.ite.tul.cz/zbynek/pubs/BSI.pdf>

- [8] R. McAulay, M. Malpass, *Speech enhancement using a soft-decision noise suppression filter*, IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 28, no. 2, pp. 137–145, Apr. 1980.
- [9] Z. Koldovský, *Biologické a akustické signály (BSI), studijní materiály – přednáška č. 4* [online]. [vid. 6. 12. 2013].  
Dostupné z: <http://itakura.ite.tul.cz/zbynek/dwnld/BSI/prednaska4.pdf>
- [10] Z. Koldovský, *Biologické a akustické signály (BSI), studijní materiály – přednáška č. 5* [online]. [vid. 6. 12. 2013].  
Dostupné z: <http://itakura.ite.tul.cz/zbynek/dwnld/BSI/prednaska5.pdf>
- [11] Y. Ephraim, D. Malah, *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*, IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
- [12] P. Wolfe, S. Godsill, *Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement*, Proc. of 11th Workshop on Statistical Signal Processing, pp. 496–699, Aug. 2001.
- [13] Y. Ephraim, D. Malah, *Speech enhancement using a minimum mean-square error log-spectral amplitude estimator*, IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 33, no. 2, pp. 443–445, Apr. 1985.
- [14] Z. Koldovský, P. Tichavský, D. Botka, *Noise Reduction in Dual-Microphone Mobile Phones Using A Bank of Pre-Measured Target-Cancellation Filters*, Proc. of International Conference on Acoustic, Speech, and Signal Processing (ICASSP 2013), pp. 679–683, May 2013.
- [15] J. Li, S. Sakamoto, S. Hongo, M. Akagi, Y. Suzuki, *Twostage binaural speech enhancement with Wiener filter based on equalization-cancellation model*, Proc. of Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009), pp. 133–136, Oct. 2009.
- [16] N. Levinson, *The Wiener RMS (root mean square) error criterion in filter design and prediction*, J. Math. Phys., Vol. 25, No. 4., pp. 261–278, 1947.

- [17] J. Durbin, *Efficient estimation of parameters in moving-average models*, Biometrika 46, pp. 306–316, 1959.
- [18] H. Akaike, *Block Toeplitz Matrix Inversion*, SIAM Journal on Applied Mathematics, vol. 24, no. 2, pp. 234–241, Mar. 1973.
- [19] J. Málek, Z. Koldovský, P. Tichavský, *Semi-Blind Source Separation Based on ICA and Overlapped Speech Detection*, Proc. of The 10th International Conference on Latent Variable Analysis and Source Separation (LVA/ICA 2012), pp. 462–469, Mar. 2012.
- [20] *SketchUp 3D Warehouse* [online]. [vid. 30. 4. 2014].  
Dostupné z: <https://3dwarehouse.sketchup.com/model.html?id=2dbfd136ab89ce5967fd64f19325>
- [21] J. Málek, D. Botka, Z. Koldovský, S. Gannot, *Methods to learn bank of filters steering nulls toward potential positions of a target source*, Proc. of the 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA 2014), May 2014.
- [22] V. Emiya, E. Vincent, N. Harlander, V. Hohmann, *Subjective and objective quality assessment of audio source separation*, IEEE Transactions on Audio, Speech and Language Processing, vol. 19, no. 7, pp. 2046–2057, Sep. 2011.
- [23] ETSI 202 396-1, *Speech and multimedia Transmission Quality (STQ); Part 1: Background noise simulation technique and background noise database*, 03 2009, V 1.2.3.
- [24] *Trail: Creating a GUI With JFC/Swing* [online]. [vid. 10. 3. 2014].  
Dostupné z: <http://docs.oracle.com/javase/tutorial/uiswing/>
- [25] *WAVE PCM soundfile format* [online]. [vid. 10. 3. 2013].  
Dostupné z: <https://ccrma.stanford.edu/courses/422/projects/WaveFormat/>
- [26] *JH Labs* [online]. [vid. 23. 4. 2013].  
Dostupné z: <http://www.jhlab.com/ip/FFT.java>