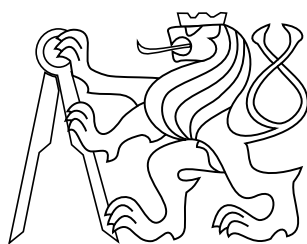


Diplomová práce

Separace akustických záznamů

Josef Vondrlík



Leden 2017

vedoucí práce: Ing. Radoslav Bortel, Ph.D.

České vysoké učení technické v Praze
Fakulta elektrotechnická, Katedra teorie obvodů

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Josef V o n d r l í k

Studijní program: Biomedicínské inženýrství a informatika

Obor: Biomedicínské inženýrství

Název tématu: Separace akustických záznamů

Pokyny pro vypracování:

1. Posuďte možnosti separace akustických signálů ze zdrojů s předem známou pozicí.
2. Navrhněte algoritmus pro separaci akustických signálů s využitím kalibračních dat.
3. Implementujte navržený algoritmus v prostředí MATLAB.
4. Navržený algoritmus otestujte na reálných akustických signálech.

Rozsah původní zprávy: cca 40 stran.

Seznam odborné literatury:

- [1] Van Trees, H. L.: Optimum Array Processing. Wiley, 2002.
- [2] Benesty, J., Chen, J., Huang Y.: Microphone Array Signal Processing. Springer, 2008.
- [3] Naidu, P. S.: Sensor Array Signal Processing, CRC Press, 2001.

Vedoucí diplomové práce: Ing. Radoslav Bortel, Ph.D.

Platnost zadání: do konce letního semestru 2016/2017

L.S.

prof. Ing. Pavel Sovka, CSc.
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 18. 12. 2015

Poděkování

Rád bych poděkoval hlavně Radkovi, mému vedoucímu, za neskutečnou trpělivost, kterou se mnou měl a za jeho lidský přístup. Dále bych chtěl poděkovat rodině a všem ostatním, kteří mě během studia podporovali.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně, a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze 9.1.2017

Abstrakt

Cílem této práce bylo navrhnout algoritmus pro separaci akustických záznamů pořízených mikrofonním polem, zejména záznamů řečových. Algoritmus využívá k separaci sadu filtrů s konečnou impulzní odezvou, jeden filtr pro každý kanál. K nastavení koeficientů těchto filtrů je využito kalibračních dat. Jako kalibrační data jsou v této práci použity signály ze zdroje s neměnnou pozicí. Tyto záznamy jsou následně segmentovány a s využitím signulárního rozkladu jsou z nich určeny optimální hodnoty koeficientů filtru. Filtrované záznamy z jednotlivých kanálů jsou následně sečteny. K testování algoritmu byla použita simulovaná data a data získaná z mikrofonního pole. Kvalita rekonstrukce signálu byla hodnocena na základě střední kvadratické chyby ve spektru a cepstru signálu. Kvalita separace byla posuzována na základě odstupu signál-šum. Algoritmus je schopný výrazně potlačit korelovaný šum při zachování srozumitelnosti řeči přicházející z požadovaného směru.

Klíčová slova

senzorová pole, tvarování přijímací charakteristiky, mikrofony, zpracování řeči

Abstrakt

The purpose of this work is to create an algorithm that separates acoustic signals, especially speech, acquired by a microphone array based on their incoming direction. The algorithm separates the signals using a set of filters with a finite impulse response – a filter per each channel. The filter coefficient setup is done via calibration data obtained from a signal source with fixed position. The optimal coefficients of the filters are acquired by separating the records using singular value decomposition, and the filtered records from each channel are summed up. The algorithm testing was done with simulated data and data obtained from the microphone array. The reconstruction quality was assessed by the mean square error in the spectra and cepstra of the signal. The separation quality was assessed by signal-to-noise-ratio. The algorithm is able to significantly attenuate the correlated noise while retaining the intelligibility of the speech coming from the desired direction.

Keywords

sensor array, beamforming, microphones, speech processing

Obsah

1	Úvod	1
2	Současný stav	3
2.1	Optimální filtrace	3
2.1.1	Wienerův filtr	3
2.1.2	Kálmánův filtr	3
2.2	Senzorová pole	4
2.2.1	Výhody, oblasti použití, tvary	4
2.2.2	Lineární pole	4
2.3	Zpracování signálů ze sensorových polí	5
2.3.1	Odhad směru příchodu	5
2.4	Tvarování přijímací charakteristiky	5
2.4.1	Delay-and-sum	5
2.4.2	Filter-and-sum	7
2.4.3	Optimální beamforming	7
	Minimální odchylka od požadovaného signálu	8
	Maximální odstup signál-šum	9
	Minimální výkon výstupního signálu	10
2.5	Využití v akustické oblasti	11
3	Navrhovaný algoritmus	12
3.1	Model přenosu signálu	12
3.1.1	Kalibrace	14
3.1.2	Odhad koeficientů separačního filtru	15
3.2	Testování algoritmu	16
3.2.1	Model mikrofonního pole	16
3.2.2	Testování rekonstrukce signálu	17
3.2.3	Odstranění šumu	18
3.3	Použitá data	18
4	Výsledky	20
4.1	Rekonstrukce signálu	20
4.2	Potlačení šumu na simulovaných datech	21
4.3	Potlačení šumu na reálných datech	23

4.4	Subjektivní hodnocení potlačení šumu	24
4.5	Zhodnocení výsledků	27
5	Závěr	29
	Literatura	30

1 Úvod

S vícekanálovými záznamy získanými pomocí různých typů mikrofonních polí se dnes setkáváme zcela běžně. V oblasti hudby můžeme například využít pole tvořeného dvěma mikrofony k pořízení stereo nahrávky. Pole mikrofonů různých směrových charakteristik se využívá ke snímání prostorového zvuku, který je nejčastěji používán ve filmové produkci. Tyto přístupy využití mikrofonního pole mají za cíl využít prostorové informace obsažené ve vícekanálovém záznamu ke zprostředkování pocitu z prostoru i pro posluchače.

Prostorovou informaci obsaženou ve vícekanálovém záznamu z mikrofonního pole je ale možné využít i k zesílení signálů přicházejících z konkrétního směru díky různému zpoždění signálů na jednotlivých mikrofonech. Toho může být využito například u naslouchadel k odstranění šumu prostředí a zaměření přijímací charakteristiky naslouchadla podle požadavků uživatele v danou chvíli. Stejně tak může prostorový filtr neúmyslně vytvořit nezkušený zvukař, který na řečnický pult umístí dva mikrofony, jejichž výstupy následně smíchá. Pokud řečník nestojí ve stejné vzdálenosti od obou mikrofonů, jsou tímto polem některé frekvence potlačeny, čímž je snížena zvuková kvalita.

Tato práce si klade za cíl otestovat možnosti tvarování přijímací charakteristiky mikrofonního pole a separaci signálů z jednotlivých zdrojů s předem danou pozicí. Díky rozdílné pozici zdrojů by mělo být možné signály odseparovat nebo alespoň nežádoucí signál potlačit.

K určení pozice jednotlivých zdrojů slouží kalibrační data, která jsou nahrávkami zvukového projevu těchto zdrojů. Činnost algoritmu tak nevyžaduje znalost geometrie mikrofonního pole. Jako typ akustického signálu zpracovávaného v této práci byly zvoleny nahrávky řečového projevu.

Tato práce má následující cíle

- navržení algoritmu pro separaci akustických signálů s využitím kalibračních dat
- implementace navrženého algoritmu v prostředí MATLAB

1 Úvod

- otestování implementovaného algoritmu na reálných a simulovaných datech

2 Současný stav

Tato kapitola shrnuje některé z dosavadních poznatků a postupů v oblasti zpracování signálu. V jednotlivých částech jsou popsány možnosti zlepšení kvality signálu použitelné i pro signály z jednotlivých senzorů. Dále jsou popsány základní principy sensorových polí, jejich výhody a využití. Popsány jsou také základní metody zpracování signálu z těchto polí.

2.1 Optimální filtrace

Metody popsané v této části umožňují filtraci šumu v samostatných signálech bez nutnosti využití sensorového pole. Tyto metody našly využití i oblastech zpracování obrazu či v navigaci.

2.1.1 Wienerův filtr

Wienerův filtr je FIR filtr využíván k odstranění aditivního šumu pomocí odhadu charakteristik signálu. Jako optimalizační kritérium využívá odhad minimální kvadratické odchylky. Má dobré výsledky v případě snížení množství šumu u méně zašuměných stacionárních signálů. Z toho důvodu není příliš vhodný pro zpracování řeči, která je signálem nestacionárním. Při nižším odstupu signál-šum navíc způsobuje zkreslení signálu. [1] [2]

2.1.2 Kálmánův filtr

Variantou Wienerova filtru upravenou pro nestacionární prostředí je Kálmánův filtr. Tento filtr provádí odhad signálu na základě statistického modelu a předchozích měření. Je vhodný pro zpracování řečových signálů, kde je schopný zvýšit srozumitelnost řečového projevu [3].

2.2 Senzorová pole

Senzorová pole našla využití v různých oblastech. Využívají je radary a sonary, našla využití v seismologii, radioastronomii či v komunikacích, ve formě anténních polí. Senzorové pole využívá i tomografii, princip zpracování získaných signálů se ale v tomto případě liší od metod, na které se zaměřuje tato práce. [4]

Využit lze i pole mikrofonů pro zpracování řeči či jiných akustických signálů. Tato problematika je více popsána v sekci 2.5.

Výhodou sensorových polí je, že v naměřeném signálu získáváme prostorovou informaci. Díky té je možné zaměřit sensorové pole jedním směrem bez nutnosti fyzického zásahu do něj pouhým zpracováním naměřeného signálu. Stejně tak je možné určit směr příchodu měřeného signálu.

2.2.1 Výhody, oblasti použití, tvary

Další výhodou sensorového pole je redundance získané informace. Díky tomu je možné omezit v signálu šum vzniklý náhodnými procesy na senzorech, či jiný šum, který je na jednotlivých kanálech nekorelovaný. Za splnění určitých podmínek je pak možné omezit i šum, který je mezi kanály korelovaný. [5] [6] [4]

Senzorová pole je možné dělit podle způsobu rozložení senzorů na lineární, planární a volumetrické. Podle vzdáleností mezi jednotlivými senzory je možné pole dále rozdělit na uniformní, neuniformní a náhodné. [5] [6]

Fyzická struktura pole může hrát roli ve výběru použitého algoritmu pro zpracování signálu. V případě odhadu směru příchodu bez předchozí kalibrace je dokonce předchozí znalost fyzické struktury pole nutností. Z důvodu zjednodušení implementace těchto algoritmů jsou tak nejčastěji využívána pole uniformní.

Existují ale i aplikace, kde hraje struktura pole (v závislosti na zvoleném algoritmu) jen malou roli.

2.2.2 Lineární pole

Nejjednodušší strukturou, kterou může sensorové pole tvořit, je lineární pole. V tomto poli leží všechny senzory v jedné ose. Pokud je vzdálenost mezi jednotlivými senzory u takového pole konstantní, mluvíme o lineárním uniformním poli. [6] [5] [4]

Počet senzorů a jejich vzdálenost od sebe určují vlastnosti pole. Pokud jako způsob zpracování signálů zvolíme jejich zprůměrování, bude se směrová charak-

teristika soustavy chovat následujícím způsobem. S vyšším počtem sensorů roste směrovost pole a obráceně.

Další možností, jak ovlivnit charakteristiku pole je různý rozestup sensorů. V případě prostého zprůměrování signálů rostoucí vzdálenost mezi senzory způsobuje zúžení hlavního laloku. Pokud je ale vzdálenost mezi senzory větší než polovina snímané vlnové délky, objevují se ve směrové charakteristice další hlavní laloky. [6] [4]

2.3 Zpracování signálů ze sensorových polí

2.3.1 Odhad směru příchodu

Senzorové pole umožňuje odhad směru příchodu signálu (*direction-of-arrival* – *DOA*). To je možné díky zpoždění příchodu signálu mezi jednotlivými senzory (*time-difference-of-arrival* – *TDOA*). [7] [5]

V případě, že je zdroj signálu dostatečně blízko, aby se dala vlna, po které se šíří, považovat za radiální, je možné odhadnout i vzdálenost zdroje. Většina metod k tomu využívá triangulaci.

2.4 Tvarování přijímací charakteristiky

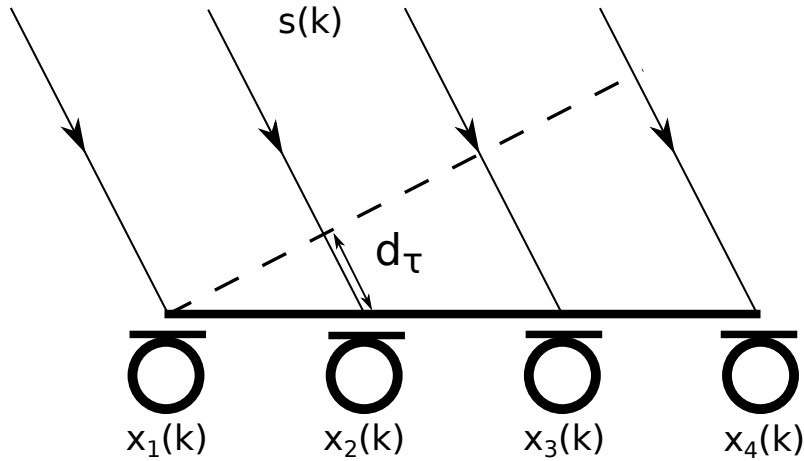
Metody zpracování signálu využívající prostorovou informaci obsaženou v signálech naměřených sensorovým polem jsou nazývány beamforming. Zde uvádím několik základních přístupů.

2.4.1 Delay-and-sum

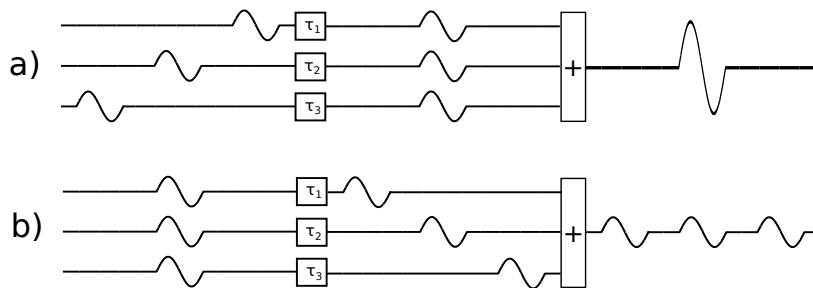
Tato metoda využívá rozdílného času příchodu signálu na jednotlivé senzory. Rozdílný čas příchodu signálu je kompenzován časovým zpožděním. Sesynchronizované signály jsou následně sečteny. Díky tomu je amplituda signálu přicházejícího z požadovaného směru zesílena. [5] [4] Tento princip je vidět na obrázcích 1 a 2.

Uvažujeme-li signál šířící se rovinnou vlnou, je možné zapsat podobu signálů na jednotlivých senzorech v čase k následovně:

$$x_i(k) = s(k - t - \tau_i) + n(k), \quad (1)$$



Obrázek 1 Znáznornění rozdílného času příchodů signálů. Vzdálenost, kterou signál překoná za čas τ je označena d_τ



Obrázek 2 Synchronizace a součet signálů delay-and-sum beamformerem. a) ukazuje signály přicházející ze směru, na který je beamformer nastaven, b) ukazuje signály přicházející z jiného směru.

kde $x_i(k)$ je signál přijatý i -tým senzorem v čase k , $s(k)$ je neznámý zdrojový signál, t je čas nutný pro signál k dosažení nejbližšího senzoru, τ_i je rozdíl v čase příchodu signálu mezi prvním a i -tým senzorem a $n(k)$ je náhodný signál na senzoru v čase k . Je-li hodnota časového zpoždění mezi prvním a druhým mikrofonom rovna τ , platí pro koeficienty τ_i následující: $\tau_1 = 0$, $\tau_2 = \tau$, $\tau_3 = 2 \cdot \tau$ až $\tau_N = (N - 1) \cdot \tau$.

Jednotlivé signály jsou upraveny tímto způsobem:

$$x_{i\text{-DEL}}(k) = x_i(k + \tau_i), \quad (2)$$

$$x_{i\text{-DEL}}(k) = s(k - t) + n(k), \quad (3)$$

kde $x_{i\text{-DEL}}(k)$ je signál x_i zpožděný o τ_i v čas k .

Zprůměrováním signálů tak získáváme výsledný signál y

$$y(k) = \frac{1}{N} \sum_{i=1}^N x_{i-\text{DEL}}(k), \quad (4)$$

$$y(k) = s(k-t) + \frac{1}{N} \cdot n(k), \quad (5)$$

kde N je počet senzorů.

Tento beamformer je nicméně velmi citlivý na korelovaný šum. V případě nekorelovaného šumu, stejného výkonu šumu na všech kanálech a správného nastavení snímaného směru zvyšuje poměr signál-šum úměrně počtu senzorů v poli, tedy

$$\text{SNR}_{\text{výstupní}} = N \cdot \text{SNR}_{\text{vstupní}}. \quad (6)$$

Problémem pro zpracování řečových signálů je u tohoto beamformery jeho úzký frekvenční rozsah. [4]

2.4.2 Filter-and-sum

Na podobném principu jako delay-and-sum beamformer fungují i filter-and-sum beamformery. Tyto beamformery signál zpožďují posunem fází jednotlivých frekvenčních složek pomocí FIR filtru. [6] [5] [4]

Princip těchto filtrů je tak možné popsat rovnicí

$$\mathbf{y} = \sum_{i=1}^N \mathbf{b}_i * \mathbf{x}_i, \quad (7)$$

[8] [9] kde y je výstupní signál beamformery, \mathbf{b}_i je vektor koeficientů FIR filtru pro i -tý kanál, \mathbf{x}_i je signál získaný z i -tého senzoru, N je počet senzorů (kanálů) a operátor $*$ značí konvoluci.

2.4.3 Optimální beamforming

Rovnici pro beamforming je možné zapsat ve tvaru

$$y = \mathbf{w}^H \mathbf{x} \quad (8)$$

kde y je výsledný signál po zpracování, \mathbf{x} je přijatý signál a \mathbf{w} jsou takzvané váhy beamformery. Operátor H značí hermitovskou transpozici, tedy transpozici matice, jejíž všechny prvky jsou komplexně sdružené.

Jednotlivé signály získané ze senzorů (mikrofonů) jsou tak každý přenásoben danými koeficienty a výsledné signály jsou následně sečteny.

Výstupní signál beamformeru je možné rozepsat do rovnice

$$y = \mathbf{w}^H \mathbf{x} = \mathbf{w}^H \mathbf{a} s \quad (9)$$

kde

$$\mathbf{a} = \mathbf{a}(\mathbf{r}) \quad (10)$$

\mathbf{a} je tak funkcí polohového vektoru \mathbf{r} a tvoří prostorové koeficienty.

Rovnice (9) ale platí pouze v ideálním, kdy není ve vstupním signálu přítomen šum. V případě signálu se šumem pak platí pro vstupní signál

$$\mathbf{x} = \mathbf{a} s + \mathbf{n}, \quad (11)$$

kde s značí zdrojový signál a \mathbf{n} značí šum. Pro výstup beamformeru pak můžeme psát

$$y = \mathbf{w}^H \mathbf{x} = \mathbf{w}^H \mathbf{a} s + \mathbf{w}^H \mathbf{n} \quad (12)$$

K nastavení vah beamformeru existuje několik praktikovaných přístupů optimálního nastavení vah. Jsou jimi kritérium minimální odchylky od požadovaného signálu (*minimal mean square error – MMSE*), maximální odstup signál šum a minimální výkon výstupního signálu.

Minimální odchylka od požadovaného signálu

Pro kritérium minimální odchylky od požadovaného signálu je nutné nejprve zavést referenční, tedy požadovaný signál. Ten označme jako d . Pro střední kvadratickou odchylku výstupu beamformeru od tohoto signálu platí

$$\text{MSE} = \text{E} \left[\left| \mathbf{w}^H \mathbf{x} - d \right|^2 \right]. \quad (13)$$

Cílem je tedy nalézt takové hodnoty \mathbf{w}^H , pro které je tato hodnota minimální. [9]

Rovnici je dále možné rozepsat tímto způsobem

$$\begin{aligned} \text{MSE} &= \text{E} \left[(\mathbf{w}^H \mathbf{x} - d)(\mathbf{x}^H \mathbf{w} - d^*) \right] = \\ &= \text{E} \left[\mathbf{w}^H \mathbf{x} \mathbf{x}^H \mathbf{w} - \mathbf{x}^H \mathbf{w} d - \mathbf{w}^H \mathbf{x} d^* + d d^* \right] = \\ &= \mathbf{w}^H \mathbf{R}_{xx} \mathbf{w} - \mathbf{r}_{dx}^H \mathbf{w} - \mathbf{w}^H \mathbf{r}_{dx} + d d^*, \end{aligned} \quad (14)$$

kde \mathbf{R}_x je autokorelační matice signálu a $\mathbf{r}_{dx} = E[\mathbf{x}d]$ je korelací mezi přijatým signálem a požadovaným signálem.

Po zderivování této funkce podle \mathbf{w}^H získáváme

$$\frac{\partial \text{MSE}}{\partial \mathbf{w}^H} = \mathbf{R}_x \mathbf{x} - \mathbf{r}_{dx}. \quad (15)$$

Tuto derivaci položíme rovnou nule a získáme tak

$$\mathbf{w} = \mathbf{R}_x^{-1} \mathbf{r}_{dx} \quad (16)$$

[8] [9]

Tento typ beamformeru tedy nevyžaduje znalost prostorových koeficientů, ale pouze vstupního signálu.

Maximální odstup signál-šum

Pro tento případ využijeme rovnici pro tvar výstupního signálu (12).

Odstup signál-šum pak můžeme spočítat jako

$$\begin{aligned} \text{SNR} &= \frac{E \left[|\mathbf{w}^H \mathbf{a} s|^2 \right]}{E \left[|\mathbf{w}^H \mathbf{n}|^2 \right]} \\ &= \frac{E \left[|\mathbf{w}^H \mathbf{a}|^2 \right]}{E \left[|\mathbf{w}^H \mathbf{n}|^2 \right]} E \left[|s|^2 \right] \end{aligned} \quad (17)$$

S hodnotou $\mathbf{w}^H \mathbf{a}$ můžeme dále pokračovat jako s konstantou. Vektor \mathbf{a} je totiž pevně daný a přenásobením vektoru \mathbf{w}^H konstantou se hodnota SNR nezmění. Bez újmy na obecnosti budeme pro zjednodušení dále počítat s hodnotou 1. Výsledný vektor \mathbf{w} je ale možné přenásobit libovolnou konstantou pro stejný výsledek.

Dále je tedy nutné jen minimalizovat hodnotu ve jmenovateli. Tu můžeme rozepsat jako

$$\begin{aligned} E \left[|\mathbf{w}^H \mathbf{n}|^2 \right] &= E \left[\mathbf{w}^H \mathbf{n} \mathbf{n}^H \mathbf{w} \right] \\ &= \mathbf{w}^H \mathbf{R}_n \mathbf{w}, \end{aligned} \quad (18)$$

kde \mathbf{R}_n je autokorelační matice šumu.

Jelikož máme podmínku $\mathbf{w}^H \mathbf{a} = 1$, provedeme minimalizaci pomocí Lagrangeových multiplikátorů.

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}^H} (\mathbf{w}^H \mathbf{R}_n \mathbf{w} + \lambda \mathbf{w}^H \mathbf{a}) &= 0 \\ \mathbf{R}_n \mathbf{w} + \lambda \mathbf{a} &= 0 \end{aligned} \quad (19)$$

Z toho už je možné vyjádřit \mathbf{w} jako

$$\mathbf{w} = -\lambda \mathbf{R}_n^{-1} \mathbf{a} \quad (20)$$

a s využitím dané podmínky $\mathbf{w}^H \mathbf{a} = 1$ je možné vyjádřit rovnici jako

$$\mathbf{w}^H \mathbf{a} = -\lambda \mathbf{a}^H \mathbf{R}_n^{-1} \mathbf{a} = 1 \quad (21)$$

z čehož je možné vyjádřit $-\lambda$ jako

$$-\lambda = \frac{1}{\mathbf{a}^H \mathbf{R}_n^{-1} \mathbf{a}} \quad (22)$$

Výsledný vztah pro váhy beamformeru je tak

$$\mathbf{w} = \frac{\mathbf{R}_n^{-1} \mathbf{a}}{\mathbf{a}^H \mathbf{R}_n^{-1} \mathbf{a}} \quad (23)$$

Minimální výkon výstupního signálu

Tento přístup nastavení vah beamformeru pracuje na základě minimalizace celkového výkonu výstupního signálu při udržení konstantního zesílení požadovaného signálu. Minimalizace tak omezuje pouze šum ve výstupním signálu. [8]

Hledáme tedy váhy \mathbf{w} tak, aby platilo

$$\mathbf{w} = \arg \min_{\mathbf{w}} E[|y|^2], \quad \mathbf{w}^H \mathbf{a} = 1 \quad (24)$$

Konstanta 1 je v tomto případě zvolena pro zjednodušení celého vztahu. Její místo může zaujmout jakákoliv jiná konstanta.

$$\begin{aligned} E[|y|^2] &= E[\mathbf{w}^H \mathbf{x} \mathbf{x}^H \mathbf{w}] \\ &= \mathbf{w}^H \mathbf{R}_x \mathbf{w} \end{aligned} \quad (25)$$

dále postupujeme obdobným způsobem jako v minulém případě s využitím Lagrangeových multiplikátorů.

$$\frac{\partial}{\partial \mathbf{w}^H} (\mathbf{w}^H \mathbf{R}_x \mathbf{w} + \lambda \mathbf{w}^H \mathbf{a}) = 0 \quad (26)$$

z toho plyne, že

$$\mathbf{w} = -\lambda \mathbf{R}_x^{-1} \mathbf{a} \quad (27)$$

a tedy

$$\mathbf{w} = \frac{\mathbf{R}_x^{-1} \mathbf{a}}{\mathbf{a}^H \mathbf{R}_x^{-1} \mathbf{a}} \quad (28)$$

Tento beamformer se také nazývá Caponův beamformer nebo *minimum variance distortionless response beamformer (MVDR)*. [8] [4]

2.5 Využití v akustické oblasti

Mikrofonní pole přinášejí následující možnosti použití: snížení šumu, odstranění echa a odrazů, lokalizace zdroje zvuku, odhad počtu zdrojů zvuku či separace záznamů podle jednotlivých zdrojů. V této práci se budu zabývat zejména zpracováním řeči. Jednotlivé zdroje zvuku tak budu označovat, jako řečníky.

Omezením problému na zpracování řeči se také sníží zpracováváný frekvenční rozsah do 4000Hz. Tohoto omezení využívá například i veřejná telefonní síť (*public switched telephone network – PSTN*) s vzorkovací frekvencí 8kHz.

Při využití mikrofonních polí pro zpracování řeči se musíme potýkat s několika problémy [4]:

- řeč má široké frekvenční pásmo
- výskyt odrazů v místnostech
- požadovaný signál může mít stejné spektrum jako šum (např. jiný řečník)
- vysoké požadavky na kvalitu rekonstrukce signálu
- řeč je nestacionární signál

3 Navrhovaný algoritmus

V této sekci popíši algoritmus použitý a testovaný v této práci včetně jeho základního matematického odvození. Výsledkem algoritmu je sada FIR filtrů použitých ke zpracování vícekanálového signálu. To je typický přístup k tomuto problému, samotný algoritmus je tak specifický ve způsobu nalezení koeficientů těchto filtrů.

3.1 Model přenosu signálu

Přenos signálu od zdroje na jednotlivé mikrofony je možné popsat jako průchod FIR filtrem. Pokud uvažujeme více vstupních signálů přicházejících z různých směrů, získáváme systém s více vstupy (počet zdrojů signálu, zvuku) a více výstupy (počet mikrofonů, kanálů). Takový systém je možné popsat rovnicí

$$\mathbf{x}_i = \sum_{j=1}^n \mathbf{a}_{i,j} * \mathbf{s}_j + \mathbf{n}, \quad (29)$$

kde \mathbf{x}_i je výstupní signál z i -tého mikrofonu, n je počet vstupních signálů, \mathbf{s}_j je j -tý vstupní signál, $\mathbf{a}_{i,j}$ jsou koeficienty FIR filtru modelujícího přenos j -tého vstupního signálu na i -tý mikrofón a operátor $*$ značí konvoluci.

\mathbf{n} pak značí náhodný šum, například termální šum vzniklý v elektronice snímacího zařízení. Pokud budeme jako šum uvažovat například promluvu jiného řečníka, než jehož snímání je požadováno, lze tento šum obsáhnout v modelu jako jeden ze zdrojů signálu.

Celý model je následně možné popsat ve spektrální oblasti pomocí věty o konvoluci [10]

$$\mathcal{F}\{\mathbf{f} * \mathbf{g}\} = \mathcal{F}\{\mathbf{f}\} \cdot \mathcal{F}\{\mathbf{g}\}, \quad (30)$$

kde \mathcal{F} značí diskrétní Fourierovu transformaci.

Vztah (29) je tak možné zapsat jako

$$\mathcal{F}\mathbf{X} = \sum_{j=1}^n \mathcal{F}\mathbf{A}_j \circ \mathcal{F}\mathbf{S}_j + \mathcal{F}\mathbf{N}, \quad (31)$$

kde $\mathcal{F}\mathbf{X}$ je matice obsahující v řádcích spektra všech výstupních signálů, $\mathcal{F}\mathbf{A}_j$ je matice obsahující v řádcích frekvenční charakteristiky FIR filtrů pro přenos j -tého signálu na jednotlivé mikrofony, $\mathcal{F}\mathbf{S}_j$ je matice obsahující v řádcích spektrum j -tého signálu, $\mathcal{F}\mathbf{N}$ je matice náhodného šumu a operátor \circ je Hadamardův součin. Tento operátor vzájemně vynásobí odpovídající prvky dvou matic stejného rozměru. Index před označením matice \mathcal{F} značí, že matice obsahuje prvky spektra získané odpovídající diskrétní Fourierově transformací signálu.

Rovnice má tedy následující strukturu

$$\begin{bmatrix} \mathcal{F}\{\mathbf{x}_1\} \\ \vdots \\ \mathcal{F}\{\mathbf{x}_k\} \end{bmatrix} = \sum_{j=1}^n \begin{bmatrix} \mathcal{F}\{\mathbf{a}_{1,j}\} \\ \vdots \\ \mathcal{F}\{\mathbf{a}_{k,1}\} \end{bmatrix} \circ \begin{bmatrix} \mathcal{F}\{\mathbf{s}_j\} \\ \vdots \\ \mathcal{F}\{\mathbf{s}_j\} \end{bmatrix}, \quad (32)$$

kde k značí počet kanálů (mikrofonů).

Pro další odvozování si převedeme tuto rovnici tak, aby vyjadřovala výsledek pro všechny vstupní i výstupní signály najednou, ale zvlášť pro každou frekvenci

$$\mathcal{F}\mathbf{X}[m] = \mathcal{F}\mathbf{A}[m] \mathcal{F}\mathbf{S}[m]. \quad (33)$$

Vektor $\mathcal{F}\mathbf{X}[m]$ tak v této rovnici obsahuje složku pro frekvenci m pro všechny kanály, tedy m -tý sloupec matice $\mathcal{F}\mathbf{X}$ a vektor $\mathcal{F}\mathbf{S}[m]$ obsahuje složku pro frekvenci m pro všechny vstupní signály. Matice $\mathcal{F}\mathbf{A}[m]$ pak obsahuje jednotlivé sloupce matic $\mathcal{F}\mathbf{A}_j$ příslušející frekvenci m .

Teoreticky je tak možné pomocí jednoduché inverze matice $\mathcal{F}\mathbf{A}[m]$ získat z vícekanalového záznamu vektor $\mathcal{F}\mathbf{S}[m]$.

$$\mathcal{F}\mathbf{A}[m]^{-1} \mathcal{F}\mathbf{X}[m] = \mathcal{F}\mathbf{S}[m] \quad (34)$$

K tomuto postupu je ale nutné získat samotnou matici $\mathcal{F}\mathbf{A}[m]^{-1}$ a také je nutná její invertovatelnost.

Omezme se pro tuto chvíli na situaci, kdy uvažujeme pouze jeden zdroj signálu. Matice $\mathcal{F}\mathbf{A}^{-1}[m]$ je tak vektorem (invertovatelnost takovéto matice je řešena Moore-Penroseovou inverzí [11]). Z těchto vektorů pro všechny m sestavíme matici

$$\mathcal{F}\mathbf{B}^T = \begin{bmatrix} \mathcal{F}\mathbf{A}[1]^{-1} \\ \mathcal{F}\mathbf{A}[2]^{-1} \\ \vdots \\ \mathcal{F}\mathbf{A}[m_{max}]^{-1} \end{bmatrix}, \quad (35)$$

Pokud bychom uvažovali více zdrojů signálu, bude každý řádek matice $\mathcal{F}\mathbf{A}[m]^{-1}$ odpovídat jednomu zdroji. Výsledkem postupu daného rovnicí (35) by tak byla množina matic $\mathcal{F}\mathbf{B}_j^T$, jedna pro každý signál, kde j odpovídá indexu daného signálu.

Úpravou vztahu (31) je tak možné získat matici $\mathcal{F}\mathbf{S}'_j$ definovanou vztahem

$$\mathcal{F}\mathbf{S}'_j = \mathcal{F}\mathbf{B} \circ \mathcal{F}\mathbf{X}. \quad (36)$$

Po sečtení jednotlivých řádků matice $\mathcal{F}\mathbf{S}'_j$ získáváme vektor $\mathcal{F}\hat{\mathbf{S}}$, který je odhadem spektra požadovaného signálu.

Tento proces je ale praktičtější provést v časové oblasti. Vztahem

$$\mathbf{b}_{i,j} = \mathcal{F}^{-1}\{\mathcal{F}\mathbf{B}_{i,j}\}, \quad (37)$$

kde \mathcal{F}^{-1} značí inverzní diskretní Fourierovu transformaci a $\mathcal{F}\mathbf{B}_{i,j}$ značí i -tý řádek matice $\mathcal{F}\mathbf{B}_j$, získáváme koeficienty $\mathbf{b}_{i,j}$, které jsou koeficienty separačního FIR filtru.

Následně pomocí vztahu

$$\hat{\mathbf{s}}_j = \sum_{i=1}^k \mathbf{b}_{i,j} * \mathbf{x}_i \quad (38)$$

získáme odhad požadovaného signálu $\hat{\mathbf{s}}_j$.

3.1.1 Kalibrace

K získání matice $\mathcal{F}\mathbf{A}[m]$ je možné využít kalibraci. K té je, pro správné fungování beamformeru, nutné využít signál se stejným nebo širším frekvenčním rozsahem než signál, který bude ve výsledku snímán. K tomu je možné využít bílý šum, případně promluvu řečníka, pro kterého se systém kalibruje.

Jednou z možností je použít jako odhad matice $\mathcal{F}\mathbf{A}$ přímo spektrum kalibračního signálu, tedy matici $\mathcal{F}\mathbf{X}$ získanou z kalibračního signálu.

Problémem může být nestacionarita signálu. Současně by bylo nutné zvolit řád separačního filtru odpovídající délce tohoto signálu. Dále je potřeba uvážit možnou přítomnost šumu během kalibračního procesu. Pokud je po celou dobu kalibrace přítomný korelovaný šum, pak je tento postup obecně nevhodný. Přítomnost nekorelovaného šumu je možné vyřešit postupem popsáním dále.

3.1.2 Odhad koeficientů separačního filtru

Pro odhad koeficientů je nutné nejprve získat kalibrační signál. Kalibračním signálem se rozumí signál ze zdroje s pevně danou pozicí se stejným frekvenčním rozsahem jako je signál, který bude následně zpracováván. Kalibrační signál by neměl obsahovat šum. V praxi může jít o promluvu jednoho řečníka.

Pro odhad matice $\mathcal{F}\mathbf{A}[m]$ je v rámci algoritmu použitého v této práci použit následující postup. Kalibrační signál je nejprve rozsegmentován tak, aby délka jednotlivých segmentů odpovídala řádu požadovaných separačních filtrů. Každý ze segmentů tak vypadá následovně

$$\text{Segment}_{i,j} = [\mathbf{x}_i[j \cdot k + 1], \mathbf{x}_i[j \cdot k + 2], \dots, \mathbf{x}_i[j \cdot k + k]], \quad (39)$$

kde $\text{Segment}_{i,j}$ je j -tý segment signálu \mathbf{x}_i , k je délka požadovaného filtru a $\mathbf{x}_i[n]$ je n -tý vzorek signálu \mathbf{x}_i .

Každý ze segmentů signálu je následně převeden na spektrum pomocí diskrétní Fourierovy transformace.

Jednou z možností dalšího zpracování je zprůměrování hodnot jednotlivých segmentů. V této práci je ale k odstranění šumu v tomto kalibračním signálu využito singulárního rozkladu matice. Ten se k odstranění šumu osvědčil v mnoha oblastech. [12]

Pro další zpracování zavedeme množinu pomocných matic $\mathcal{F}\mathbf{C}[m]$ s následující strukturou

$$\mathcal{F}\mathbf{C}[m] = \begin{bmatrix} \mathcal{F}\mathbf{C}_{1,1}[m] & \cdots & \mathcal{F}\mathbf{C}_{1,n}[m] \\ \vdots & \ddots & \vdots \\ \mathcal{F}\mathbf{C}_{k,1}[m] & \cdots & \mathcal{F}\mathbf{C}_{k,n}[m] \end{bmatrix}, \quad (40)$$

kde $\mathcal{F}\mathbf{C}_{i,j}[m]$ odpovídá hodnotě spektra pro frekvenci m j -tého segmentu i -tého kanálu. Počet řádků matice tak odpovídá počtu jednotlivých mikrofonů a počet sloupců počtu segmentů. Matic je stejné množství, jako je řád separačního filtru.

Každou z těchto matic podrobíme singulárnímu rozkladu.

Pro singulární rozklad matice platí následující vztah

$$\mathcal{F}\mathbf{C}[m] = U\Sigma V^H, \quad (41)$$

kde matice U a V jsou unitární matice, tedy matice, pro které platí $U^{-1} = U^H$. Matice Σ je diagonální matice, která na diagonále obsahuje singulární hodnoty matice \mathbf{C} (pro tento odstavec nebudou pro přehlednost zápisu použity indexy).

Singulární hodnoty matice \mathbf{C} jsou druhé odmocniny vlastních čísel matice $\mathbf{C}\mathbf{C}^H$. Operátor H značí hermitovské sdružení.

Pro omezení šumu nahradíme matici Σ maticí σ [12], pro jejíž prvky platí

$$\begin{aligned}\sigma_{1,1} &= 1 \\ \sigma_{i,j} &= 0, \quad \text{kde } i \neq 1 \wedge j \neq 1.\end{aligned}\tag{42}$$

Rovnici (41) tak upravíme pro zpětnou rekonstrukci na

$$\mathcal{F}\mathbf{C}'[m] = U\sigma V^H.\tag{43}$$

Výsledkem takovéto rekonstrukce je singulární matice $\mathcal{F}\mathbf{C}'[m]$. Všechny sloupce této matice jsou vůči sobě lineárně závislé. Obsahují odhad těchto koeficientů kalibračního signálu nezatíženého šumem. Každý z těchto sloupců je možné použít jako sloupec matice $\mathcal{F}\hat{\mathbf{A}}$, která je odhadem matice $\mathcal{F}\mathbf{A}$. Pokud tento postup zopakujeme pro celou množinu matic $\mathcal{F}\mathbf{C}'[m]$, získáme tak celou matici $\mathcal{F}\hat{\mathbf{A}}$.

Problémem tohoto postupu je, že separační matice $\mathcal{F}\mathbf{B}$ získaná z této matice $\mathcal{F}\hat{\mathbf{A}}$ by u rekonstruovaného signálu způsobovala ekvalizaci spektra. To je dáno různou amplitudou jednotlivých frekvencí v kalibračním signálu, které jsou algoritmem interpretovány jako vliv prostředí, který je nutné vyrovnat. Proto je před převodem na matici $\mathcal{F}\mathbf{B}$ nutné nanormovat jednotlivé sloupce matice $\mathcal{F}\hat{\mathbf{A}}$ tak, aby jimi daný vektor měl jednotkovou velikost.

Z toho vyplývá, že krok daný rovnicí (43) je možné upravit na vyjmutí prvního sloupce matice U , který je následně využit pro tvorbu matice $\mathcal{F}\hat{\mathbf{A}}$ stejným způsobem, jako sloupce matice $\mathcal{F}\mathbf{C}'[m]$. Matice V ani škálovací matice Σ se v tomto případě neuplatní.

3.2 Testování algoritmu

Testování navrženého algoritmu bylo provedeno jednak na datech naměřených na reálném mikrofonním poli a jednak na datech vygenerovaných pomocí modelu mikrofonního pole.

3.2.1 Model mikrofonního pole

V rámci této práce byl v prostředí MATLAB naimplementován model mikrofonního pole. Tento model pracuje na základě popisu přenosu signálu od zdroje na jednotlivé mikrofony popsáno v sekci 3.1.

Model nejprve náhodně vytvoří koeficienty FIR filtrů \mathbf{a} . Tyto koeficienty jsou pro pozdější použití Fourierovou transformací převedeny na matici A . Vstupem pro tvorbu modelu je tak jen počet senzorů a řád FIR filtrů modelu.

Vzhledem k tomu, že je inicializace koeficientů náhodná, neodpovídá model uniformnímu lineárnímu sensorovému poli, ale poli obecnému. To není problém vzhledem k tomu, že algoritmus navržený v této práci informaci o geometrické struktuře pole neuplatňuje.

Těmito vygenerovanými koeficienty je následně přefiltrován požadovaný signál. Výstupem jsou signály na jednotlivých kanálech pole.

3.2.2 Testování rekonstrukce signálu

Samotnou kvalitu rekonstrukce signálu není možné objektivně posoudit bez znalosti zdrojového signálu. Reálná data naměřená jako promluva řečníka tak nebudou tímto způsobem testována.

Výstupní data testovací modelu je tímto způsobem ale možné otestovat, jelikož je původní signál známý. Jako metrika je použita střední kvadratická odchylka (*MSE – mean squared error*)

$$e = \frac{1}{N} \sum_{i=1}^N (x_{1i} - c \cdot x_{2i})^2, \quad (44)$$

kde x_{1i} je i -tý vzorek prvního (vstupního) signálu, x_{2i} je i -tý vzorek druhého (výstupního) signálu, c je škálovací koeficient a N je počet vzorků signálu.

Škálovací koeficient je nutný vzhledem k rozdílné amplitudě vstupního a výstupního signálu. Hodnota škálovacího koeficientu je určena tak, aby byla hodnota střední kvadratické odchylky minimalizována. Rovnici (44) zderivujeme podle c

$$\frac{\partial}{\partial c} e = \frac{2}{N} \sum_{i=1}^N (c \cdot x_{2i} - x_{1i} \cdot x_{2i}). \quad (45)$$

Pokud položíme tuto derivaci rovnou nule, dostáváme se po drobných úpravách k hodnotě koeficientu

$$c = \frac{\sum_{i=1}^N x_{1i} \cdot x_{2i}}{\sum_{i=1}^N x_{2i}^2}. \quad (46)$$

Vzhledem k tomu, že hodnota střední kvadratické chyby je úměrná amplitudě signálů, bude v této práci uváděna i hodnota normovaná průměrným výkonem signálu

$$e_{\text{norm}} = \frac{N \cdot e}{\sum_{i=1}^N x_{1i}^2}. \quad (47)$$

Střední kvadratická odchylka je v rámci testování rekonstrukce vypočítána jednak pro spektrum signálu získané diskretní Fourierovou transformací a jednak pro reálné cepstrum signálu.

Reálné cepstrum signálu \mathbf{x} [13] je definováno jako

$$\text{Cepstrum}(\mathbf{x}) = \mathcal{F}^{-1} \{ \log_{10} (|\mathcal{F} \{ \mathbf{x} \}|) \}, \quad (48)$$

kde \mathcal{F} je diskretní Fourierova transformace a \mathcal{F}_{-1} inverzní diskretní Fourierova transformace.

Střední kvadratická odchylka mezi cepstry dvou signálů je tak vlastně průměrem kvadrátu Euklidovské cepstrální vzdálenosti. Tu je možné využít jako měřítko zkreslení řečového signálu.

3.2.3 Odstranění šumu

Testování potlačení korelovaného šumu je možné provést na reálných i simulovaných datech. Jako metrika pro vyhodnocení úspěšnosti je použit odstup signál-šum (*SNR – signal-to-noise ratio*) [10]

$$\text{SNR}_{\text{dB}} = 10 \cdot \log_{10} \frac{P_{\text{signál}}}{P_{\text{šum}}}, \quad (49)$$

kde $P_{\text{signál}}$ je výkon požadovaného signálu a $P_{\text{šum}}$ je výkon šumu.

Jelikož je proces beamformingu v rámci algoritmu navrženého v této práci lineární, je možné pro testování použít dva samostatné signály, u kterých je SNR porovnáno před průchodem filtry beamformeru a po něm.

Co se nekorelovaného šumu týče, je možné očekávat výsledek stejný nebo lepší než u šumu korelovaného.

3.3 Použitá data

Pro testování byla použita simulovaná data vytvořená modelem popsáním v sekci 3.2.1. Řád náhodně generovaných modelovacích FIR filtrů byl v tomto případě 500, počet kanálů modelu pak 8.

Jako zdrojové signály pro modelovaná data byly použity 3 nahrávky řeči o trvání 30 sekund nahrané s vzorkovací frekvencí 8000 Hz. Všechny tři nahrávky obsahují promluvu autora této práce, obsahem promluvy je předčítaná poezie.

Pro testování reálných dat byly využity 4 nahrávky různé délky. Tyto nahrávky byly získány za pomoci lineárního mikrofonního pole s osmi všesměrovými mikrofony. Vzorkovací frekvence nahrávek je 8000 Hz. Každá z nahrávek obsahuje řečový projev jednoho, vždy toho samého řečníka. Pozice řečníka vůči mikrofonnímu poli je pro každou nahrávku rozdílná.

4 Výsledky

Tato kapitola shrnuje výsledky testování algoritmu.

4.1 Rekonstrukce signálu

Kvalita rekonstrukce signálu byla testována pomocí výpočtu normované minimální střední kvadratické odchylky mezi amplitudovým spektrem zdrojového signálu a amplitudovým spektrem signálu získaného rekonstrukcí signálu získaného pomocí modelu mikrofonního pole. Mezi stejnými signály byla počítána i normovaná minimální střední kvadratická odchylka v cepstru.

Bylo použito celkem 5 různých, náhodně inicializovaných modelů řádu 500.

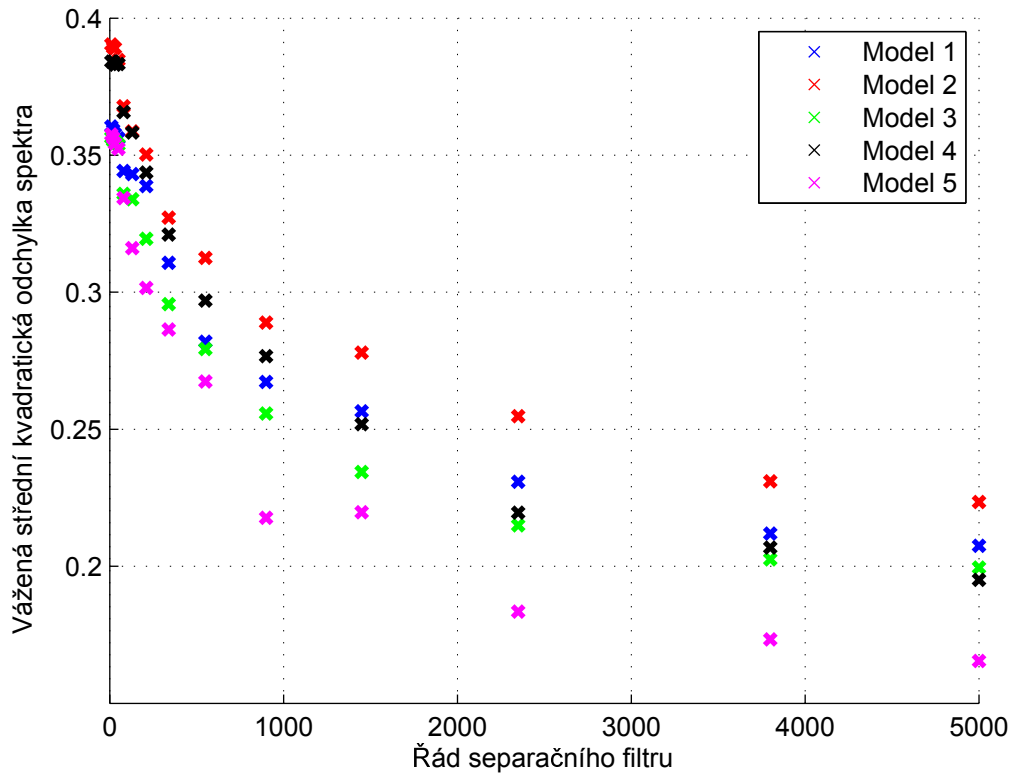
Pro každý z těchto modelů byla následně generována simulovaná data pro každou z použitých nahrávek (viz sekce 3.3). Jednotlivé nahrávky jsou označeny jako *Signál 1*, *Signál 2* a *Signál 3*.

Pro každý model byl testovací scénář následující:

- *Signál 1* je použitý jako kalibrační signál, získanými koeficienty jsou rekonstruovány signály 2 a 3
- *Signál 2* je použitý jako kalibrační signál, získanými koeficienty jsou rekonstruovány signály 1 a 3
- *Signál 3* je použitý jako kalibrační signál, získanými koeficienty jsou rekonstruovány signály 1 a 2
- jsou vypočítány normované minimální střední kvadratické odchylky v cepstru a amplitudovém spektru

Tento postup je proveden pro různý řád separačních filtrů. Výsledky jsou následně pro každý model a řád zprůměrovány.

Hodnoty minimální střední kvadratické odchylky amplitudových spekter jsou pro každý model a řád separačního filtru zobrazeny v grafu na obrázku 3 a také v tabulce 1. Hodnoty minimální střední kvadratické odchylky cepster jsou pro



Obrázek 3 Graf zobrazuje závislost hodnoty normované střední kvadratické odchylky spektra původního a rekonstruovaného signálu na řádu separačního filtru.

každý model a řád separačního filtru zobrazeny v grafu na obrázku 4 a také v tabulce 2.

Jak je z grafů patrné, zatímco normovaná střední kvadratická odchylka ve spektru s rostoucím řádem filtru klesá, normovaná střední kvadratická odchylka v cepstru naopak stoupá. To je možné interpretovat tak, že filtry s rostoucím řádem zvýrazňuje frekvence s vyšší amplitudou, zatímco ostatní potlačuje.

4.2 Potlačení šumu na simulovaných datech

Pro testování zvýšení odstupů signál-šum na simulovaných datech byly použity stejné modely jako pro testování rekonstrukce signálu. *Model 1* byl použit pro tvorbu signálů přicházejících z požadovaného směru. Zbylé 4 modely pak byly použity pro tvorbu korelovaného šumu.

Testovací scénář:

- ze *signálu 1* jsou vygenerována pomocí *modelu 1* kalibrační data, pomocí

4 Výsledky

Tabulka 1 Tato tabulka uvádí vypočtené hodnoty normované střední kvadratické odchylky mezi spektrem původního signálu a signálu rekonstruovaného separačním filtrem pro jednotlivé modely. Jednotlivé sloupce odpovídají danému řádu filtru. Tyto hodnoty jsou také zobrazeny v obrázku 3

Model \ Řád filtru	10	20	30	50	80	130	210
Model 1	0.3605	0.3592	0.3578	0.3561	0.3442	0.3431	0.3387
Model 2	0.3903	0.3893	0.3891	0.3847	0.3678	0.3587	0.3502
Model 3	0.3564	0.3552	0.3552	0.3538	0.3358	0.3339	0.3195
Model 4	0.3841	0.3841	0.3830	0.3832	0.3657	0.3583	0.3437
Model 5	0.3576	0.3564	0.3543	0.3523	0.3344	0.3161	0.3016

Model \ Řád filtru	340	550	900	1450	2350	3800	5000
Model 1	0.3107	0.2819	0.2672	0.2566	0.2308	0.2119	0.2075
Model 2	0.3272	0.3125	0.2889	0.2780	0.2547	0.2310	0.2235
Model 3	0.2956	0.2793	0.2557	0.2344	0.2149	0.2025	0.1995
Model 4	0.3210	0.2970	0.2767	0.2518	0.2195	0.2068	0.1950
Model 5	0.2864	0.2674	0.2177	0.2197	0.1834	0.1734	0.1655

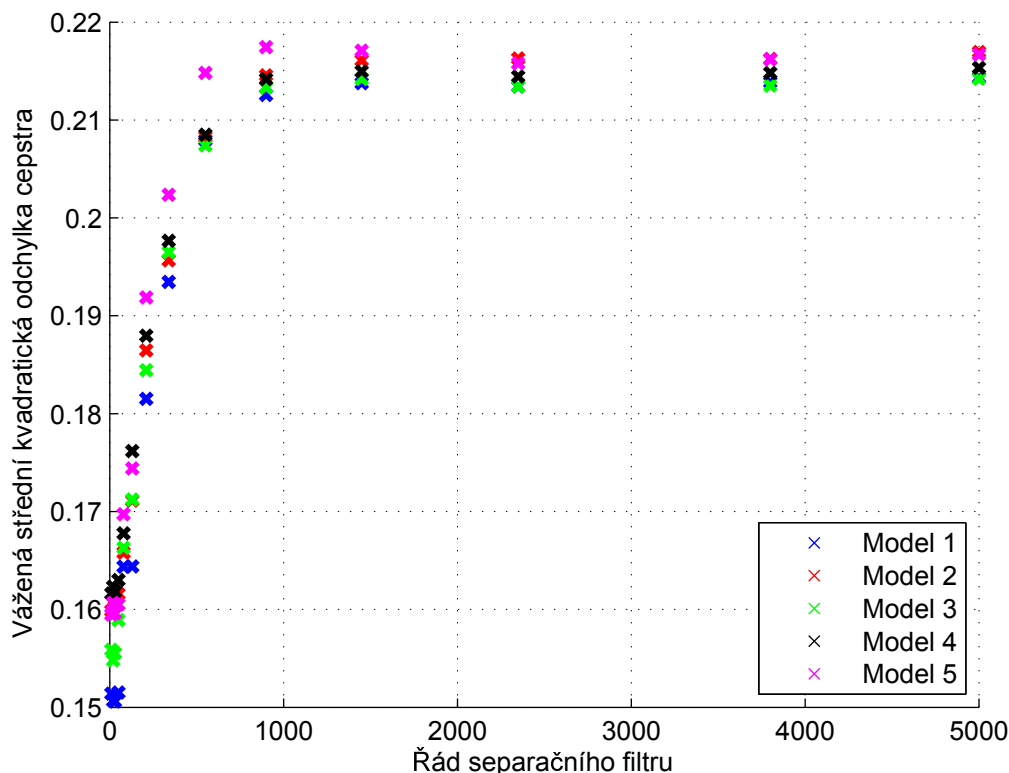
nichž jsou vygenerovány separační filtry

- ze *signálu 2* jsou pomocí *modelu 1* vygenerována testovací data
- ze *signálu 3* jsou pomocí *modelů 2, 3, 4 a 5* vygenerována testovací data
- testovací data jsou filtrována separačními filtry, kanály jednotlivých testovacích dat jsou následně sečteny
- je vypočítán průměrný výkon signálů na vstupu P_{i-IN} a na výstupu P_{i-OUT} , kde i odpovídá číslu použitého modelu (tím je podle předchozích kroků určen i použitý signál)
- změna poměru signál-šum je vypočítána následujícím vzorcem

$$\text{SNR}_{i-\text{dB}} = 10 \cdot \log_{10} \frac{P_{1-OUT} \cdot P_{i-IN}}{P_{i-OUT} \cdot P_{1-IN}},$$

kde $i = 2, 3, 4, 5$

Takto získané hodnoty jsou vyneseny v grafu na obrázku 5 a v tabulce 3.



Obrázek 4 Graf zobrazuje závislost hodnoty normované střední kvadratické odchylky cepstra původního a rekonstruovaného signálu na řádu separačního filtru.

4.3 Potlačení šumu na reálných datech

Testování na reálných nahrávkách pořizovaných za pomoci mikrofonního pole se ukázalo být mírně problematické vzhledem k tomu, že v každé z nahrávek se zdroj vyskytuje v jiné pozici vůči mikrofonnímu poli. Nakonec tak byl jako kalibrační signál zvolen stejný signál, na kterém je následně testováno potlačení šumu.

Vzhledem ke způsobu kalibrace, je tento postup možný bez přeučení pro signály s řádově větším počtem vzorků, než je řád požadovaných separačních filtrů. Signál použitý ke kalibraci v tomto případě obsahuje přibližně 55000 vzorků, což je více než desetkrát více než je největší testovaný řád filtru.

Reálné signály použité pro toto testování jsou označeny jako *Signál 1*, *Signál 2*, *Signál 3* a *Signál 4*.

Testovací scénář:

- jsou vygenerovány koeficienty filtrů, jako kalibrační data je použit *Signál 1*
- všechny signály jsou filtrovány touto sadou filtrů, kanály každého signálu

Tabulka 2 Tato tabulka uvádí vypočtené hodnoty normované střední kvadratické odchylky mezi cepstrem původního signálu a signálu rekonstruovaného separačním filtrem pro jednotlivé modely. Jednotlivé sloupce odpovídají danému řádu filtru. Tyto hodnoty jsou také zobrazeny v obrázku 4

Řád filtru \ Model	10	20	30	50	80	130	210
Model 1	0.1514	0.1507	0.1506	0.1515	0.1644	0.1644	0.1815
Model 2	0.1599	0.1596	0.1603	0.1616	0.1658	0.1711	0.1865
Model 3	0.1559	0.1548	0.1555	0.1589	0.1663	0.1712	0.1844
Model 4	0.1616	0.1623	0.1618	0.1630	0.1678	0.1762	0.1880
Model 5	0.1595	0.1607	0.1596	0.1605	0.1697	0.1744	0.1919

Řád filtru \ Model	340	550	900	1450	2350	3800	5000
Model 1	0.1935	0.2080	0.2126	0.2138	0.2134	0.2139	0.2143
Model 2	0.1957	0.2083	0.2146	0.2161	0.2163	0.2162	0.2169
Model 3	0.1965	0.2074	0.2133	0.2142	0.2134	0.2135	0.2142
Model 4	0.1977	0.2085	0.2142	0.2150	0.2144	0.2148	0.2153
Model 5	0.2024	0.2148	0.2174	0.2171	0.2158	0.2162	0.2167

jsou následně sečteny

- je vypočítán průměrný výkon signálů na vstupu P_{i-IN} a na výstupu P_{i-OUT} , kde i odpovídá číslu signálu
- změna poměru signál-šum je vypočítána následujícím vzorcem

$$\text{SNR}_{i-\text{dB}} = 10 \cdot \log_{10} \frac{P_{1-OUT} \cdot P_{i-IN}}{P_{i-OUT} \cdot P_{1-IN}},$$

kde $i = 2, 3, 4$

Takto získané hodnoty jsou vyneseny v grafu na obrázku 6 a v tabulce 4.

4.4 Subjektivní hodnocení potlačení šumu

Potlačení šumu bylo dobře slyšitelné jak pro reálná, tak pro simulovaná data.

Co se simulovaných dat týče, byl pro subjektivní hodnocení testován scénář, kdy je vstupem součet dvou různých promluv, každé upravené jedním z modelů. Signál připravený k separaci tak obsahuje dvě různé promluvy znějící zároveň.

Tabulka 3 Tato tabulka uvádí vypočtené hodnoty poměru signál-šum uvedené v decibelech, tedy hodnoty poměru požadovaného signálu ku signálu získanému modelem uvedeném v prvním sloupci. Jednotlivé sloupce odpovídají danému řádu filtru. Tyto hodnoty jsou také zobrazeny v obrázku 5

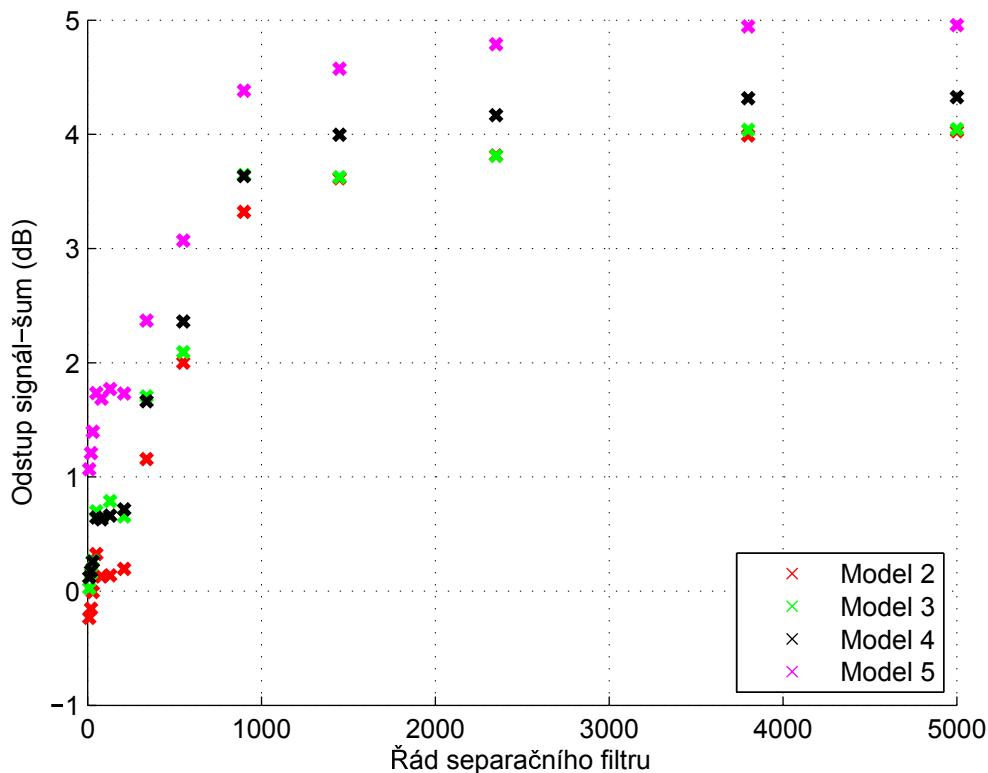
Model \ Řád filtru	Řád filtru						
	10	20	30	50	80	130	210
Model 2	-0,2348	-0,1543	-0,0084	0,3256	0,1272	0,1386	0,1937
Model 3	0,0246	0,1558	0,2661	0,6989	0,6417	0,7884	0,6528
Model 4	0,1196	0,1757	0,2585	0,6435	0,6281	0,6606	0,7167
Model 5	1,0667	1,2079	1,3953	1,7334	1,6863	1,7687	1,7298

Model \ Řád filtru	Řád filtru						
	340	550	900	1450	2350	3800	5000
Model 2	1,1554	2,0013	3,3207	3,6155	3,8149	3,9900	4,0277
Model 3	1,7042	2,0928	3,6443	3,6243	3,8108	4,0386	4,0458
Model 4	1,6630	2,3607	3,6334	3,9958	4,1678	4,3159	4,3256
Model 5	2,3698	3,0699	4,3821	4,5743	4,7884	4,9450	4,9568

Tabulka 4 Tato tabulka uvádí vypočtené hodnoty poměru signál-šum uvedené v decibelech, tedy hodnoty poměru požadovaného signálu ku signálu uvedenému v prvním sloupci. Jednotlivé sloupce odpovídají danému řádu filtru. Tyto hodnoty jsou také zobrazeny v obrázku 6

Šum \ Řád filtru	Řád filtru						
	10	20	30	50	80	130	210
Signál 2	8,3557	8,7141	8,6078	6,1724	7,1811	7,3741	8,2380
Signál 3	3,8487	4,8326	5,1761	4,6529	5,3243	5,1013	5,9702
Signál 4	2,6377	3,0664	3,6001	3,6546	3,8722	4,6266	6,1674

Šum \ Řád filtru	Řád filtru						
	340	550	900	1450	2350	3800	5000
Signál 2	8,8183	9,427	9,0569	9,1980	9,4645	9,5796	9,8543
Signál 3	6,9282	7,2555	7,2069	7,5576	7,1232	7,2964	7,6472
Signál 4	6,4604	8,3299	8,0708	8,4581	7,9345	8,6067	9,2628



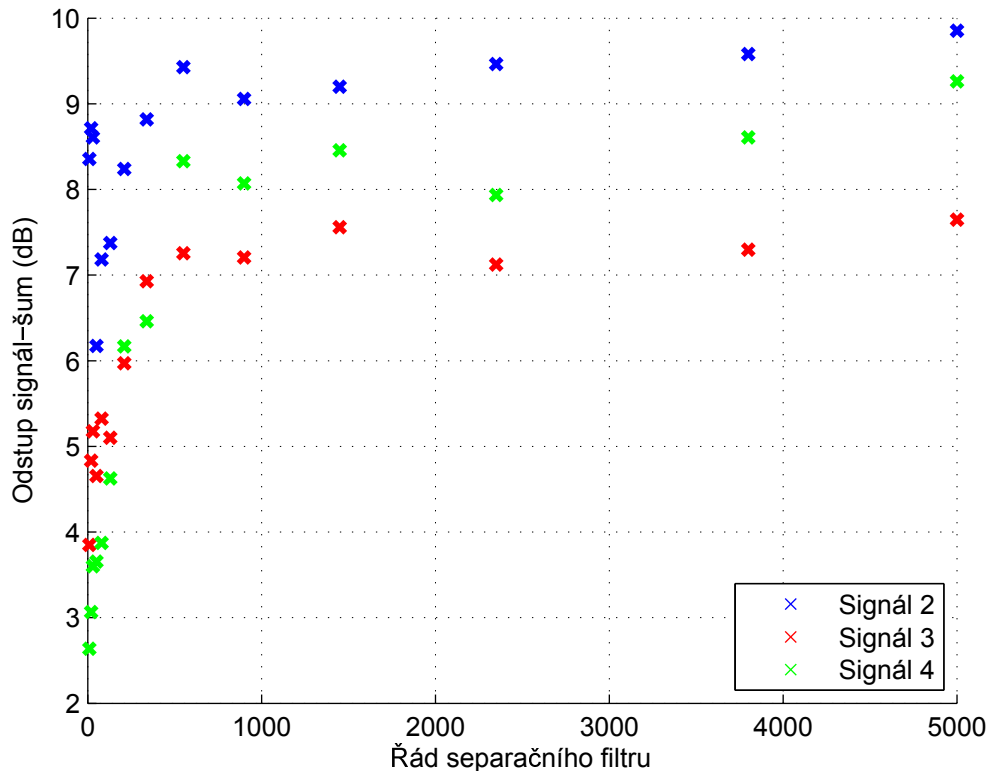
Obrázek 5 Graf zobrazuje závislost hodnoty odstup signal-šum (dB) na řádu separačního filtru získané postupem popsaným v sekci 4.2.

V takto získaném signálu je velmi obtížné rozlišit jednotlivá slova a porozumět obsahu promluvy.

Separací algoritmus je kalibrován jiným signálem získaným pomocí jednoho z těchto dvou modelů. Po separaci je promluva upravená druhým z těchto modelů v signálu potlačena dostatečně na to, aby byla požadovaná promluva srozumitelná a bylo možné rozlišit její obsah. Nežádoucí promluva je nicméně v signálu stále slyšet, její obsah je ale nesrozumitelný.

Je nutné podotknout, že v tomto konkrétním případě je srozumitelnost promluv snížena i tím, že jsou obě nahrávky promluvou stejné osoby (autora této práce). Pokud by byl hlasový rejstřík řečníka v každé z nahrávek jiný, byla by srozumitelnost lepší jak před separací, tak po ní.

Co se reálných dat týče, je výsledek obdobný, jako u dat simulovaných. Oproti simulovaným datům je ale možné navíc slyšet velké potlačení nekorelovaného šumu mikrofonního pole, čímž je promluva v těchto nahrávkách výrazně zvýrazněna.



Obrázek 6 Graf zobrazuje závislost hodnoty odstup signál-šum (dB) na řádu separačního filtru získané postupem popsáním v sekci 4.3.

4.5 Zhodnocení výsledků

Navržený algoritmus prokázal schopnost zvýšit odstup signál-šum pro korelovaný šum, současně je schopen potlačit i šum nekorelovaný a zachovat srozumitelnost řečového projevu.

Úroveň potlačení korelovaného šumu však nedosahuje takových hodnot, aby bylo možné při poslechu záznamu mluvit o jeho odstranění. Algoritmus tak může dobře sloužit ke zvýraznění signálu přicházejícího od konkrétního zdroje, ne k jeho naprosté separaci.

Schopnost algoritmu potlačit šum roste s řádem použitého filtru. Zvyšování řádu filtru ale také zvyšuje výpočetní náročnost jak samotné kalibrace, tak filtrace. Současně se s rostoucím řádem filtru zvyšuje zpoždění signálu, což může způsobit problémy při aplikaci v reálném čase.

Pro vyšší odstup signál-šum by bylo vhodnější využít algoritmu kalibrovaného šumem. Tato úloha ale vyžaduje jiné vstupní podmínky než tato práce, konkrétně

4 Výsledky

známou pozici zdroje šumu, nikoliv zdroje požadovaného signálu.

Pro lepší vyhodnocení účinnosti algoritmu navrženého v této práci by bylo možné využít více reálných dat. Další možností testování je také otestovat schopnost algoritmu reagovat na korelovaný i nekorelovaný šum přítomný v kalibračních datech.

5 Závěr

V rámci této diplomové práce byl navržen a otestován algoritmus pro separaci akustických záznamů pořízených mikrofonním polem. Tento algoritmus využívá tvarování přijímací charakteristiky pole za pomoci sady filtrů s konečnou impulzní odezvou. Tato technika je jinak známá také jako *filter-and-sum beamforming*.

Specifikou algoritmu navrženého v této práci je způsob odhadu koeficientů separačních filtrů, kdy je pro zpřesnění tohoto odhadu využito singulárního rozkladu matic. Vstupem pro tento algoritmus jsou kalibrační data. Kalibračními daty se v rámci této práce rozumí nahrávka akustického signálu, jehož zdroj zachovává vůči poli stálou pozici. Navržený algoritmus byl implementován v prostředí MATLAB.

Testování algoritmu bylo provedeno na simulovaných datech a na reálných datech naměřených na mikrofonním poli. Tato data obsahují řečové záznamy.

V rámci testování bylo prokázáno, že navržený algoritmus je pro korelovaný šum schopný zvýšit odstup signál-šum. Není však schopný tento šum zcela odstranit. Algoritmus je současně schopný zachovat srozumitelnost řečového záznamu.

Literatura

- [1] I. Cohen et al. *Noise Reduction in Speech Processing*. Springer Berlin Heidelberg, 2009. ISBN: 978-3-642-00295-3.
- [2] J. Benesty, J. Chen a Y. Huango. “On widely linear Wiener and tradeoff filters for noise reduction”. In: *Speech Communication* 52.5 (2010), s. 427–439. ISSN: 0167-6393.
- [3] K. Paliwal a A. Basu. “A speech enhancement method based on Kalman filtering”. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Institute of Electrical a Electronics Engineers (IEEE).
- [4] J. Benesty, J. Chen a Y. Huango. *Microphone Array Signal Processing*. Springer, 2008. ISBN: 978-3-540-78611-5.
- [5] H.L. Van Trees. *Optimum Array Processing*. Wiley, 2002. ISBN: 9780471093909.
- [6] P.S. Naidu. *Sensor Array Signal Processing*. CRC Press, 2001. ISBN: 0-8493-1195-0.
- [7] P. Stoica a A. Nehorai. “Performance study of conditional and unconditional direction-of-arrival estimation”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38.10 (1990), s. 1783–1795.
- [8] P. Chevalier a A. Blin. “Widely Linear MVDR Beamformers for the Reception of an Unknown Signal Corrupted by Noncircular Interferences”. In: *IEEE Transactions on Signal Processing* 55.11 (2007), s. 5323–5336.
- [9] P. Chevalier, J. Delmas a A. Oukaci. “Properties, performance and practical interest of the widely linear MMSE beamformer for nonrectilinear signals”. In: *Signal Processing* 97 (2014), s. 269–281.
- [10] Alan V. Oppenheim, Ronald W. Schafer a John R. Buck. *Discrete-time signal processing*. Upper Saddle River, N.J. Prentice Hall, 1999. ISBN: 0-13-754920-2.

- [11] A. Ben-Israel a T Greville. *Generalized Inverses: Theory and Applications (CMS Books in Mathematics)*. Springer, 2003. ISBN: 0387002936.
- [12] W. Chen et al. “SVD-Based Technique for Interference Cancellation and Noise Reduction in NMR Measurement of Time-Dependent Magnetic Fields”. In: *Sensors* 16.3 (2016), s. 323.
- [13] D. G. Childers, D. P. Skinner a R. C. Kemerait. “The cepstrum: A guide to processing”. In: *Proceedings of the IEEE* 65.10 (říj. 1977), s. 1428–1443. ISSN: 0018-9219.