

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVA  
Fakulta elektroniky a informatiky  
Ústav elektroniky a fotoniky

## Rozvoj metód separácie rečových signálov

Development of methods for speech signals separation

Autoreferát dizertačnej práce  
na získanie vedecko-akademickej hodnosti Philosophiae Doctor (PhD.)

v odbore doktorandského štúdia: 5.2.13. Elektronika  
v študijnom programe: Mikroelektronika

Evidenčné číslo: FEI-104404-22382

autor: Ing. Vladimír Sedlák  
školiťka: prof. Ing. Viera Stopjaková, PhD.

Bratislava, júl 2018

Dizertačná práca bola vypracovaná na Ústave elektroniky a fotoniky Fakulty elektrotechniky a informatiky Slovenskej technickej univerzity v Bratislave.

doktorand: **Ing. Vladimír Sedlák**  
Ústav elektroniky a fotoniky  
Fakulta elektrotechniky a informatiky  
Slovenská technická univerzita v Bratislave  
Ilkovičova 3, 812 19 Bratislava

školiťka: **prof. Ing. Viera Stopjaková, PhD.**  
Ústav elektroniky a fotoniky  
Fakulta elektrotechniky a informatiky  
Slovenská technická univerzita v Bratislave  
Ilkovičova 3, 812 19 Bratislava

oponent: **prof. RNDr. Ľubica Beňušková, PhD.**  
Katedra aplikovanej informatiky  
Fakulta matematiky, fyziky a informatiky  
Univerzita Komenského  
Mlynská dolina, 842 48 Bratislava

**prof. Ing. Vladislav Musil, CSc.**  
Ústav mikroelektroniky  
Fakulta elektrotechniky a komunikačných technológií  
Vysoké učení technické v Brně  
Technická 10, Královo Pole  
61600 Brno, Česká republika

Autoreferát bol odovzdaný dňa: .....

Obhajoba dizertačnej práce sa koná dňa 23.8.2018 o 11 hod. na Fakulte elektrotechniky a informatiky Slovenskej technickej univerzity v Bratislave, Ilkovičova 3, 812 19 Bratislava.

prof. Dr. Ing. Miloš Oravec  
dekan FEI STU

# Obsah

|                                                                    |    |
|--------------------------------------------------------------------|----|
| Úvod.....                                                          | 5  |
| 1 Súčasný stav problematiky.....                                   | 6  |
| 2 Ciele dizertačnej práce .....                                    | 9  |
| 3 Kvalita ľudskej reči a jej hodnotenie .....                      | 10 |
| 3.1 Motivácia .....                                                | 10 |
| 3.2 Analýza .....                                                  | 10 |
| 3.3 Zhrnutie .....                                                 | 13 |
| 4 Návrh inovatívnej metódy pre separovanie rečových signálov ..... | 14 |
| 4.1 Generatívna kontradiktórna sieť (GAN).....                     | 14 |
| 4.2 Návrh neurónovej siete .....                                   | 14 |
| 4.3 Zhrnutie .....                                                 | 16 |
| 5 Vyhodnotenie metód .....                                         | 17 |
| 5.1 Analýza nezávislých podpriestorov (ISA) .....                  | 17 |
| 5.2 Nezáporná maticová faktorizácia (NMF) .....                    | 19 |
| 5.3 Vektorová kvantizácia (VQ) .....                               | 20 |
| 5.4 Separačná metóda na báze GAN .....                             | 21 |
| 5.5 Porovnanie separačných metód .....                             | 23 |
| 6 Zhrnutie dosiahnutých výsledkov a prínosov .....                 | 24 |
| 7 Záver.....                                                       | 25 |
| 8 Summary .....                                                    | 26 |
| Zoznam publikácií autora .....                                     | 27 |
| Zoznam literatúry použitej v autoreferáte.....                     | 30 |

# Úvod

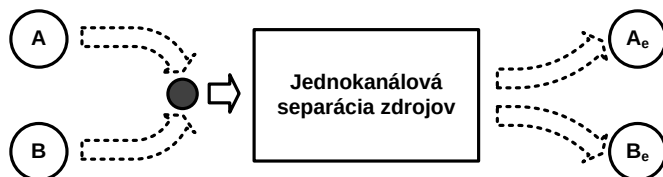
Ľudská reč ako dôležitý nástroj komunikácie čoraz viac nachádza uplatnenie vo viacerých vedeckých a priemyselných oblastiach, ktoré sa zameriavajú na interakciu človeka z elektronickými zariadeniami. Môžeme povedať, že tento trend je zreteľnejší práve v súčasnej dobe, keď sa ako užívatelia obklopujeme elektronikou, na ktorej ovládanie nám postačuje náš hlas. Dnes je úplne bežné ak svojmu mobilnému telefónu zadávame hlasové príkazy namiesto zadávania pokynov pomocou kláves, ovládame osvetlenie miestností pomocou hlasových povelov, vyhľadávame na internete alebo nastavujeme voľbu želaného kanálu na televíznom prijímači len vyslovením nášho priania. Veľký prielom a úspech v tejto oblasti zaznamenal osobný asistent od spoločnosti Apple nazvaný Siri. Postupom času sa začali objavovať nové a nové systémy a aplikácie, ktoré umožňujú podobný druh komunikácie.

Na to aby bolo ovládanie hlasom zaujímavé pre užívateľov, musí byť vykonané v reálnom čase a najmä dostatočne spoľahlivo. Oba tieto aspekty sú v moderných systémoch zaručené najmä pomocou metód strojového alebo hlbokého učenia. Ich úlohou je správne rozpoznať hlasový povel a následne odoslať príkaz na vykonanie daného procesu. Otázkou ale zostáva, ako zaručiť dostatočnú spoľahlivosť? V moderných systémoch sa na dosiahnutie najväčšej spoľahlivosti využívajú systémy na báze neurónových sietí, ktoré v ideálnych podmienkach dokážu veľmi spoľahlivo rozpoznať hlasový povel. Avšak ako je dobre známe, reálne podmienky sú častokrát značne odlišné od ideálnych. Vedci a vývojári sa musia zaoberať otázkou, či bude mať okolité prostredie vplyv na spoľahlivosť a ak hej, akým spôsobom je možné tento vplyv vyhodnotiť a eliminovať? A presne túto otázku sme si pred písaním tejto práce položili a rozhodli sme sa váš výskum zamerať práve na túto oblasť. Otázka, ako môžeme vyhodnotiť vplyv prostredia je opísaná v tretej kapitole, ktorá sa z prevažnej väčšiny skladá z experimentov zameraných na vyhodnotenie kvality reči pomocou rôznych metrík a ich schopnosti odzrkadľovať reálne podmienky. Otázka ako môžeme eliminovať vplyv okolitého prostredia je rozobratá v nasledujúcej kapitole. Sú tu predstavené a experimentálne overené rôzne konvenčné metódy, ktoré nahrádzajú štandardný prístup založený na frekvenčnom filtrovaní, ako aj inovatívne a moderné prístupy na báze neurónových sietí a kontradiktívne trénovania.

# 1 Súčasný stav problematiky

V súčasnej dobe sa proces spracovania reči dostáva čoraz viac do popredia. Tento trend je významne podporovaný vládnyimi a vojenskými inštitúciami, nadnárodnými korporáciami (napr. Apple, Microsoft, Heaton Research, atď.), ale taktiež malými špecializovanými spoločnosťami, ktoré sa zaoberajú výskumom a vývojom v oblasti digitálneho spracovania signálov (DSP). O atraktivnosti danej oblasti sa môže presvedčiť v článku [1], kde autori analyzujú príspevky prezentované na medzinárodnej konferencii *IEEE International Conference on Acoustics, Speech and Signal Processing*. Z ich práce môžeme vidieť, že viac ako štvrtina všetkých publikácií je venovaná spracovaniu reči.

Do oblasti spracovania rečových signálov patrí aj problematika ich separovania. Pod pojmom separovania rečových signálov si môžeme predstaviť či už extrakciu reči dominantného rečníka zo vstupného signálu, alebo rozdelenie vstupného signálu podľa príspevkov jednotlivých rečníkov. Základné rozdelenie metód a teda aj prístupov určených na separovanie signálov je definované na základe počtu senzorov a zdrojov signálu, ktoré sú zahrnuté do procesu. V prípade rečových signálov je senzor reprezentovaný mikrofónom a zdroj signálu rečou rečníka. Ak je výsledný (zmiešaný) signál zaznamenaný iba pomocou jedného mikrofónu, hovoríme o jednomikrofónovom alebo jednokanálovom prístupe (SCSS). V prípade viacerých mikrofónov je táto technika označovaná ako viacmikrofónová alebo viackanálová (MCSS). Výhodu MCSS prístupu je to, že môžeme pracovať aj s priestorovou informáciou, čo v prípade SCSS nie je možné. Z toho dôvodu tento prístup zvyčajne vykazuje vyššiu výkonnosť ako SCSS, avšak za cenu väčšej výpočtovej náročnosti (je potrebné spracovať viacej vzoriek) a väčšieho počtu senzorov. My sme sa v našom výskume zamerali na jednokanálové metódy. Proces separácie je graficky znázornení na obrázku 1.1.



Obr. 1.1: Konfigurácia systému pre separovanie signálov.

Z hľadiska prístupu môže byť separácia rozdelená na tri základné skupiny:

- **Filtrácia** – využíva sa skupina špeciálne navrhnutých filtrov s cieľom extrahovať požadovaný signál z predkladaného zmiešaného signálu. Optimálne nastavenie týchto filtrov prebieha najčastejšie počas fázy tréovania.
- **Dekompozícia a zoskupenie** – vstupný signál je počas dekompozície (môže byť vykonaná napríklad pomocou STFT rozdelený na jednotlivé komponenty. Tie sú v nasledujúcom kroku zoskupené podľa pravidiel špecifických pre danú metódu s cieľom čo najväčšej zhody medzi odhadnutým a pôvodným signálom.
- **Modelovanie zdrojov** – pomocou štatistických metód sú počas tréovania vytvorené modely každého zdroja (rečníka) produkujúceho vstupný signál a taktiež aj model samotného procesu zmiešania signálov.

Okrem vyššie uvedeného rozdelenia je možné metódy pre separáciu signálov rozdeliť do skupín, čiastočne odvodených od používaných prístupov. Je potrebné ale poznamenať, že existujú aj metódy využívajúce kombináciu týchto prístupov:

- **Nedourčené „slepé“ metódy** – využívajú adaptačné techniky založené na informačno-teoretických princípoch s cieľom separovať odhady pôvodných signálov priamo zo vstupného (zmiešaného) signálu a zvyčajne bez akejkoľvek dodatočnej informácie o vlastnostiach pôvodných zdrojov alebo procesu zmiešania. Najčastejšie využívajú predpoklad je, že jednotlivé zdroje sú medzi sebou nezávislé, na čom sú založené algoritmy ako analýza nezávislých komponentov (ICA) [VS16], nezáporná maticová faktorizácia (NMF) [2], analýza nezávislých podpriestorov (ISA) [3] alebo riedke kódovanie (SP) [4].
- **Metódy založené na modeloch** – využívajú vopred známe informácie o charaktere a vlastnostiach pôvodných zdrojov. Separácia sa vykonáva vo viacerých krokoch, pričom prvým je vytvorenie, respektíve natréovanie modelov špecifických pre každý jeden zdroj signálu. Vo väčšine prípadov tieto modely pozostávajú z parametrov reči získaných z časovo-frekvenčnej (Č-F) reprezentácie. Na natréovanie modelov sa používajú známe metódy strojového učenia, ako napríklad: vektorová kvantizácia (VQ) [VS19], NMF

(využíva aj pri nedourčených metódach), skryté Markove modely (HMM) [5], Gausovské modely (GMM) [6].

- **Metódy založené na analýze sluchovej scény (CASA)** – sú inšpirované ľudským sluchovým ústrojenstvom a jeho schopnosťou vnímať signál, ktorý môže byť aj silne poškodený alebo maskovaný okolitým prostredím. Vo väčšine prípadov sú založené na Č-F reprezentácii vykonanej pomocou STFT alebo na základe banky filtrov. Tento proces sa v prípade CASA metód nazýva segmentácia. Za ňou nasleduje ďalšia veľmi náročná fáza, ktorú nazývame zoskupovanie. Úlohou zoskupovania je na základe zvoleného parametra (základná frekvencia, amplitúdová modulácia, ofset, onset) zlúčiť tie elementy Č-F reprezentácie, ktoré pravdepodobne pochádzajú od jedného zdroja [7, 8].
- **Metódy založené na hlbokom učení** – sú založené buď na priamom generovaní binárnej masky alebo priamej transformácii vstupného signálu. Za týmto účelom boli viac krát využité dopredné neurónové siete (FNN) [9], auto-enkóder (AE) [10], prípadne rekurentné neurónové siete (RNN) [11] a konvolučné neurónové siete (CNN) [12].

## 2 Ciele dizertačnej práce

Na základe dôkladnej analýzy aktuálneho stavu problematiky, ako aj z nej vyplývajúcich požiadaviek a potrieb v oblasti separovania akustických (rečových) signálov, boli ciele nášho výskumu stanovené nasledovne:

- Špecifikácia najnovších trendov a potrieb v oblasti separovania rečových signálov s cieľom zvýšenia kvality snímaného vstupného signálu.
- Stanovenie a porovnanie parametrov vhodných na objektívne a subjektívne vyhodnotenie kvality reči z dôrazom na odzrkadlenie vplyvu prostredia.
- Návrh metódy vhodnej na separáciu zdrojov akustických signálov s využitím adaptívnych výpočtových systémov a techník hlbokého učenia.
- Vyhodnotenie a porovnanie účinnosti navrhnutej metódy pre rôzne druhy vstupného signálu a parametre prostredia.
- Implementácia navrhnutej metódy separácie vo forme softvérovej knižnice pre možnosti jej ďalšieho využitia.



## 3 Kvalita ľudskej reči a jej hodnotenie

### 3.1 Motivácia

Na základe predchádzajúcej analýzy sme zaznamenali viacero publikovaných prác, ktoré definovali rôzne metriky na určenie kvality reči. Avšak väčšina z nich sa buď zameriavala sa hodnotenie kvality signálu ako takého (a nie signálu produkovaného separačným algoritmom) alebo nezohľadňovala vplyv akustického prostredia. Táto skutočnosť nás motivovala k vykonaniu komplexnej analýzy, ktorá by tento chýbajúci priestor vyplnila a poskytla nám dôležité informácie. Primárnym cieľom v prvej fáze výskumu je definovať metriky, ktoré dokážu odzrkadliť vplyv prostredia a sú aplikované na separovaný signál. Za týmto účelom sme využili ideálnu binárnu masku, ktorú sme aplikovali na vstupný signál a analyzovali zmenu hodnoty jednotlivých parametrov.

### 3.2 Analýza

Za účelom identifikovania vhodných metrík sme navrhli experiment využívajúci ideálnu binárnu masku (IBM), ktorú sme následne aplikovali na vstupný signál maskovaný rečou iného rečníka, hlukom, prípadne odrazmi signálu od stien miestnosti. Technika IBM je matematicky opísaná rovnicou (3.1), kde symbol  $j$  reprezentuje frekvenčný index,  $m$  časový index,  $LC$  je hodnota lokálnej prahovej úrovne SNR,  $X(j, m)$  je hodnota  $j$ -tého frekvenčného pásma  $m$ -tého segmentu užitočného signálu a  $N(j, m)$  zase maskovacieho signálu. V praxi je najčastejšie prahová úroveň nastavená na hodnotu 0 dB.

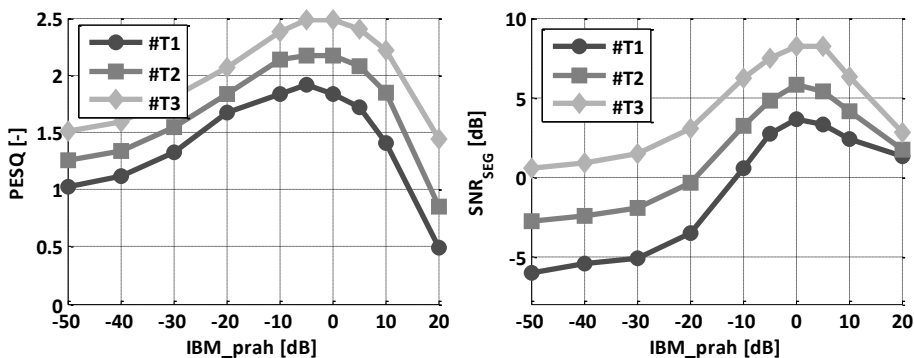
$$IBM = \begin{cases} 1, & \text{ak } \frac{X(j, m)}{N(j, m)} > LC \\ 0, & \text{inak} \end{cases} \quad (3.1)$$

Metriky, ktoré sme podrobili analýze sú zosumarizované v tabuľke 3.1, a pokrývajú širokú škálu prístupov. Naším cieľom bolo sledovať vplyv prostredia na relatívnu zmenu danej metriky, pričom vyhodnocovaný signál bol spracovaný pomocou IBM. Sledovali sme vplyv maskovacieho signálu (bľabot, reč jedného či dvoch rečníkov, hluk na letisku, na vlakovej stanici a v automobile), jeho hodnoty (-5 dB, 0 dB, 5 dB), parametrov ideálnej binárnej masky (chyba a prahová úroveň) a parametrov akustického prostredia (čas dozvuku, vzdialenosť medzi mikrofónom a rečníkom).

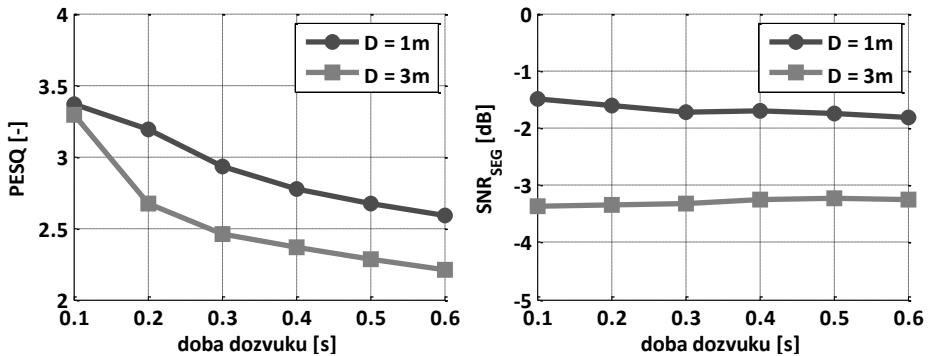
Tab. 3.1: Metriky hodnotenia kvality reči podrobené analýze

| Skratka            | Názov                                              |
|--------------------|----------------------------------------------------|
| SNR <sub>SEG</sub> | Segmentálny odstup užitočného signálu od šumu      |
| STOI               | Krátkočasové objektívne hodnotenie zrozumiteľnosti |
| PESQ               | Percepčné hodnotenie kvality reči                  |
| SDR                | Odstup užitočného signálu a celkového skreslenia   |
| SIR                | Odstup užitočného signálu od interferencií         |
| SAR                | Odstup užitočného signálu od artefaktov            |
| OPS                | Celkové percepčné skóre                            |
| TPS                | Percepčné skóre užitočného signálu                 |
| IPS                | Percepčné skóre interferencií                      |
| APS                | Percepčné skóre artefaktov                         |

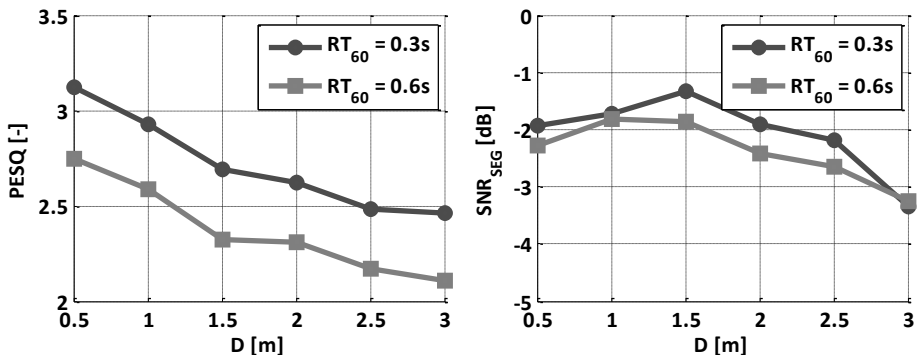
Za účelom overenie IBM sme na úvod vykonali jednoduchú analýzu s cieľom identifikovať vplyv správnej voľby hodnoty prahu. Výsledky sú graficky znázornené na obrázku 3.1, pričom v prvom prípade bola kvalita výstupného signálu určená pomocou PESQ a v druhom pomocou SNR<sub>SEG</sub>. Pre tieto parametre sme sa rozhodli z dôvodu, že pokrývajú pomerne širokú škálu prístupov hodnotenia kvality a metodika ich výpočtu je značne rozdielna. V nasledujúcej analýze sme zohľadňovali vplyv maskovacieho signálu na relatívnu zmenu analyzovaných metrík. Z výsledkov, ktoré sme z dôvodu eliminovania počtu strán ukázali iba v dizertačnej práci, vidieť že všetky metriky dokázali úspešne odzrkadliť pôvod maskovacieho signálu.



Obr. 3.1: Výkonnosť IBM v závislosti od hodnoty prahu pre maskovací signál #T1 (reč iného rečníka), #T2 (hluk), #T3 (reč + hluk).



Obr. 3.2: Výkonnosť IBM v závislosti od čas dozvu pre vzdialenosť medzi mikrofónom a rečníkom 1 m a 3 m.



Obr. 3.3: Výkonnosť IBM v závislosti od vzdialenosti medzi mikrofónom a rečníkom pre čas dozvu 300 ms a 600 ms.

Nasledovala analýza, ktorej cieľom bolo zohľadniť vplyv chyby IBM na schopnosť analyzovaných metrík vyhodnotiť kvalitu reči. Rovnako ako v predchádzajúcom prípade, sú dosiahnuté výsledky prezentované iba v dizertačnej práci, nakoľko všetky metriky boli schopné tento typ poruchy odhaliť. Poslednou analýzou bolo sledovanie parametrov prostredia, v ktorom sa reč šíri, na relatívnu zmenu metrík. Pomocou RIR generátora [13] sme vytvorili model fiktívnej miestnosti o rozmeroch  $6 \times 3 \times 4$  metrov (dĺžka  $\times$  výška  $\times$  šírka), pomocou ktorého sme získali jej impulznú odozvu. Počet vzoriek tejto odozvy sme zvolili na 1024, filtrácia vysokých frekvencií nebola použitá a taktiež nebolo nastavené obmedzenie počtu odrazov. Výsledný vstupný signál bol potom vytvorený konvolúciou odozvy a pôvodného bezodrazového signálu. Z dosiahnutých výsledkov vyplýva, že v tomto prípade už nie všetky metriky dokážu správne identifikovať kvalitu reči. Ako príklad sme uviedli porovnanie medzi PESQ a SNR<sub>SEG</sub> pre

rôzne hodnoty času dozvuku a vzdialenosti medzi mikrofónom. V prípade ak sa zameriame na metriku PESQ vidíme klesajúci trend, čo však spíňa predpoklady, nakoľko narastajúca doba dozvuku prípadne vzdialenosť zvyšuje množstvo interferencií a odrazov, ktoré negatívne vplyvajú na kvalitu reči. Iné správanie môžeme pozorovať v prípade  $SNR_{SEG}$  nakoľko zmena času dozvuku nemá výrazný vplyv na hodnotu tohto parametra. S toho dôvodu prichádzame k záveru, že  $SNR_{SEG}$  patrí medzi metriky, ktoré nebudeme v našej záverečnej analýze využívať. Zo všetky analyzovaných metrík z tabuľky 3.1 sa ako najlepšie prejavili PESQ a STOI, ktoré budeme používať pre porovnanie výkonnosti jednotlivých separačných metód.

### 3.3 Zhrnutie

Cieľom tejto kapitoly bolo stanovanie a analýza metrík vhodných na vyhodnotenie kvality rečových signálov, čo je jedným z cieľov tejto dizertačnej práce. Za týmto účelom sme navrhli a vykonali experiment, ktorého cieľom bolo preveriť schopnosti zvolených metrík odzrkadľovať vplyv akustického prostredia a typu maskovacieho signálu na kvalitu reči. Samotná analýza prestavovala v prvom kroku výpočet ideálnej binárnej masky, ktorá bola použitá na separovanie testovacieho signálu. Výsledný signál bol následne vyhodnotený pomocou zvolených meraní s cieľom zistiť či dokázali odhaliť zmenu vstupných parametrov pri zmiešavaní signálov ako napríklad: hodnota maskovacieho signálu, typ maskovacieho signálu, presnosť odhadu (výpočtu) binárnej masky a taktiež typ akustického prostredia v ktorom sa zvuk šíri. Výsledky dosiahnuté v rámci vykonaných experimentov potvrdzujú, že metriky ako PESQ, STOI alebo  $BSS_{EVAL}$  sú schopné vyhodnocovať aj rôzne vplyvy na proces separácie rečových signálov. Získané poznatky a výsledky tejto analýzy boli publikované v našich prácach [VS01] a [VS18].

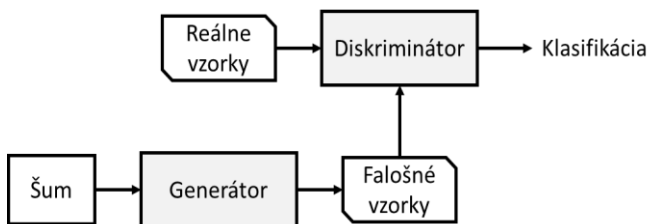
PESQ nebolo pôvodne navrhnuté pre určovanie kvality reči obsahujúcej odrazy, avšak vo vykonaných experimentoch bolo preukázané, že dokáže čiastočne odhaliť aj tieto vlastnosti reči. So vzrastajúcim časom dozvuku, respektíve vzdialenosťou medzi zdrojom signálu a PESQ skóre mierne klesá. Avšak toto meranie je veľmi citlivé na ostatné druhy skreslenia, ktoré majú vplyv na zrozumiteľnosť reči. Rovnako ako PESQ, aj STOI meranie patrí do kategórie percepčných metód, a taktiež vykazuje obdobné vlastnosti pri hodnotení kvality reči, či už sa jedná o hodnotenie bez odrazového signálu alebo s odrazmi.

## 4 Návrh inovatívnej metódy pre separovanie rečových signálov

### 4.1 Generatívna kontradiktórna sieť (GAN)

GAN sa skladá z dvoch hlbokých neurónových sietí, ktoré sa medzi sebou snažia súťažiť. Boli vynájdené v roku 2014 na univerzite v Montreale a vo vedeckej obci sa pokladajú za „najzaujímavejší nápad“ v oblasti strojového učenia za posledných 10 rokov. Majú obrovský potenciál, pretože sú teoreticky schopné naučiť sa generovať akúkoľvek distribúciu dát (obraz, zvuk, reč a pod.). Častokrát sú označované ako roboty s umeleckým cítením.

Blokový diagram základného princípu GAN siete je zobrazený na obrázku 4.1. Úlohou generátora je generovať vzorky zatiaľ čo úlohou diskriminátora je vyhodnotiť ich pôvod. Vygenerovaná vzorka sa spoločne so vzorkou zo vstupného dátového setu privedie na vstupy diskriminátora. Jeho úlohou je vyhodnotiť či vzorka privedená na jeho vstupy je pravá alebo falošná. Cieľom diskriminátora je čo najlepšie sa natrénovať na odhalenie falošných vzoriek a naopak cieľom generátora je produkovať vzorky, ktoré budú na nerozoznanie od pôvodných vzoriek. Diskriminátor je väčšinou konvolučná neurónová sieť, ktorá klasifikuje vstupné vzorky. Generátor je štandardne vytvorený pomocou inverznej konvolučnej siete. Obe siete sa navzájom snažia optimalizovať protikladnú chybovú funkciu a vzájomne sa ovplyvňujú.



Obr. 4.1: Blokový diagram GAN siete.

### 4.2 Návrh neurónovej siete

Návrh siete pre separáciu reči prezentovaný v tejto kapitole bol inšpirovaný sieťou prezentovanou v práci [14], ktorá patrí medzi priekopníkov v danej oblasti. Je založená

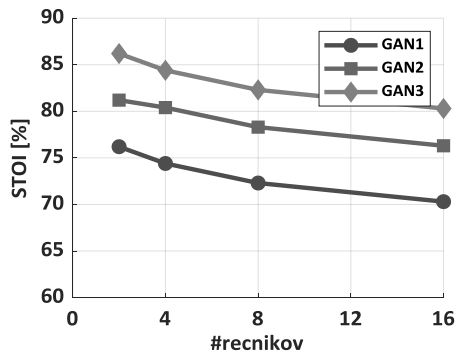
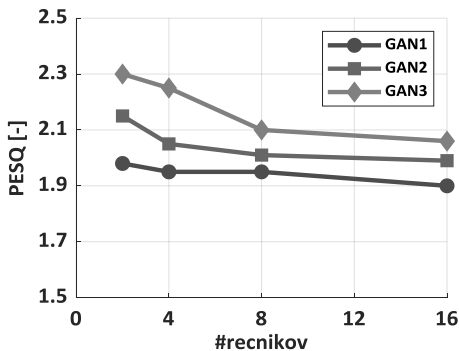
na priamej transformácii, pri ktorej je úlohou generátora na základe vstupného vektora vygenerovať signál bez prídavného maskovacieho signálu. Pre overenie efektívnosti GAN ako separátora rečových signálov sme navrhli rôzne architektúry siete, ktoré sú zhrnuté v tabuľke 4.1. Generátor má štruktúru auto-enkódera, pričom kóder a dekóder majú filtre s rovnakou šírkou, takže tento auto-ekodér má symetrickú štruktúru. Jedná sa o podobnú štruktúru akú využíva auto-ekodér navrhnutý pre eliminovanie šumu avšak s tým rozdielom, že v tomto prípade využívame kontradiktórne tréningovanie.

Tab. 4.1: Konfigurácia auto-enkódera pre GAN.

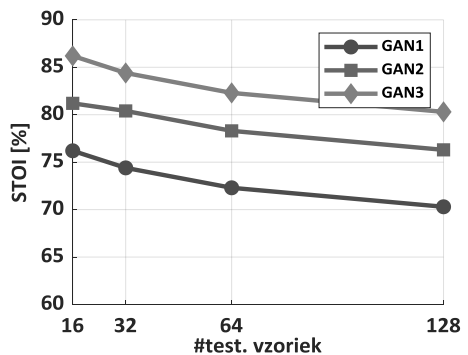
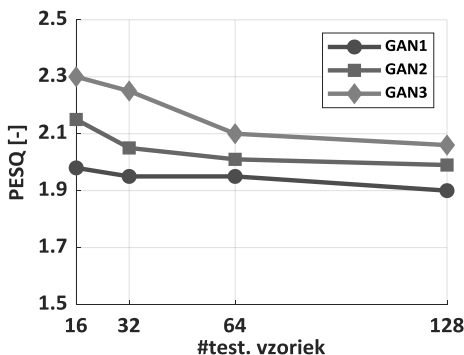
| Názov       | Štruktúra generátora                                                                                             |
|-------------|------------------------------------------------------------------------------------------------------------------|
| <b>GAN1</b> | 4096x1 – 2048x16 – 1024x32 – 512x64 – 256x128 – 128x128 – 64x256 – 32x256 – 16x512 – 8x512                       |
| <b>GAN2</b> | 8192x1 – 4096x16 – 2048x32 – 1024x64 – 512x64 – 256x128 – 128x128 – 64x256 – 32x256 – 16x512 – 8x1024            |
| <b>GAN3</b> | 16384x1 – 8192x16 – 4096x32 – 2048x32 – 1024x64 – 512x64 – 256x128 – 128x128 – 64x256 – 32x256 – 16x512 – 8x1024 |

Za účelom separovania rečových signálov sa využíva segmentovanie vstupného signálu pomocou oknovej funkcie, ktorej dĺžka je totožná s dĺžkou vstupného vektora neurónovej siete. V prípade siete GAN1 a 16 kHz frekvencie, je veľkosť okna stanovená na 256 ms, pre GAN2 je to 512 ms a pre GAN3 bolo okno 1.024 ms. Pri posune oknovej funkcie je využité 50% prekrytie medzi segmentami. Pred procesom tréningovania je potrebné vytvoriť dve sady tréningových vzoriek, pričom prvá skupina bude obsahovať iba reč, ktorá nie je poškodená hlukom prípadne rečou iného rečníka. Druhá skupina naopak obsahuje reč maskovanú rečou iného rečníka. Je nutné zaručiť, aby v tom istom okamihu sieť spracovávala vzorky, ktoré obsahujú reč toho istého rečníka, pretože inak by sa nebola schopná správne natréningovať.

Na obrázkoch 4.2 a 4.3 je zobrazený vplyv testovacej sady na výkonnosť separačného procesu. Ako je možné vidieť najlepšie výsledky dosiahneme ak do procesu zahrnieme malý počet rečníkov (čím však výrazne obmedzíme flexibilitu riešenia) a veľký počet vzoriek, čo však výrazne predlžuje čas tréningovania. Ak analyzujeme dosiahnuté výsledky na základe štruktúry siete vidíme, že sieť s najväčším počtom vrstiev – GAN3 dosahuje najlepšie výsledky avšak rozdiel medzi ňou a sieťou GAN2 nie je markantný. Podrobnejšia analýza výkonnosti GAN bude prezentovaná v nasledujúcej kapitole.



Obr. 4.2: Vplyv počtu rečníkov v trénovacej sade na výkonnosť navrhnuťej GAN siete.



Obr. 4.3: Vplyv počtu vzoriek v trénovacej sade na výkonnosť navrhnuťej GAN siete.

### 4.3 Zhrnutie

Primárne sú GAN siete využívané na spracovanie obrazu, je ich ale možné výhodne uplatniť aj v prípade rečových signálov, ktoré je možné transformovať do grafickej podoby pomocou rôznych transformácií, čo bolo predmetom našich prác [VS13] a [VS14]. Našou prvou úlohou bolo overenie vplyvu architektúry na parametre separovaného signálu. Na počiatku sme zvolili sieť s menším počtom vrstiev, ktoré sme postupne zvyšovali. Okrem kvality výstupu sme monitorovali aj čas potrebný na natréovanie siete a chybovú funkciu. Zistili sme, že veľkosť siete má podľa očakávania pozitívny vplyv na kvalitu separácie, avšak od určitej veľkosti siete sme nezaznamenali výraznejšie zlepšenie. V ďalšom kroku sme experimentovali s parametrami vstupných signálov. Potvrdilo sa, že dĺžka vstupnej sekvencie ovplyvňovala najmä čas potrebný na natréovanie siete.

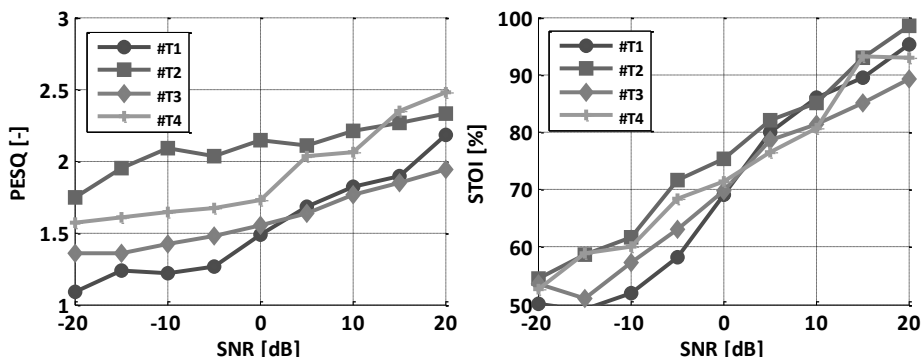
## 5 Vyhodnotenie metód

Nasledujúca analýza ma za cieľ overiť a porovnať schopnosť separačných metód eliminovať vplyv okolitého prostredia na kvalitu separovaného signálu. Prvý test je zameraný na overenie schopnosti separovať signály zmiešané na rôznych úrovniach SNR maskované signálmi uvedenými v tabuľke 5.1. V tomto prípade sa SNR menilo od -20 dB po 20 dB s krokom 5 dB a rovnako ako v predchádzajúcich experimentoch mali testovacie signály dĺžku 2 sekundy. V ďalšom kroku sme sa zamerali na analýzu vplyvu odrazov a interferencií vo zmiešanom signále. Za týmto účelom sme vytvorili impulzovú odozvu fiktívnej testovacej miestnosti o rozmeroch  $6 \times 3 \times 4$  metrov (dĺžka  $\times$  výška  $\times$  šírka), počet vzoriek sme zvolili 1024. V experimentoch sme najskôr menili čas odozvy od 0,1 s do 0,6 s. Následne sme sledovali vplyv vzdialenosti medzi mikrofónom a rečníkom, ktorá sme menili od jedného metra do 3 metrov s krokom pol metra.

Tab. 5.1 Tabuľka maskovacích signálov.

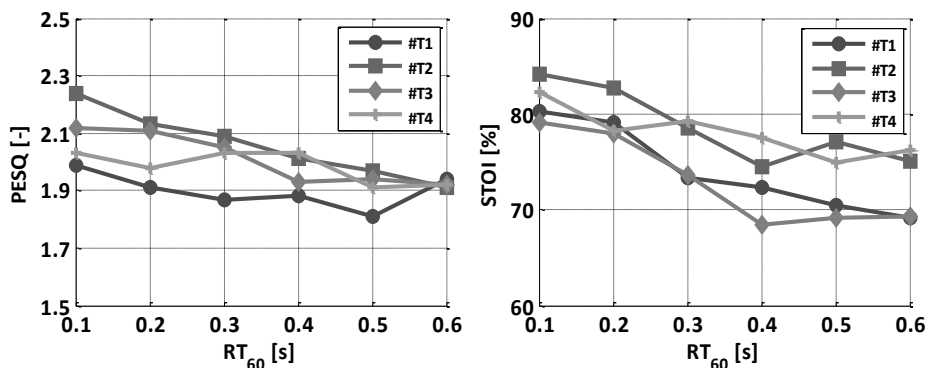
| ID  | Typ maskovacieho signálu |
|-----|--------------------------|
| #T1 | blábot                   |
| #T2 | reč jedného rečníka      |
| #T3 | hluk na stanici          |
| #T4 | hluk v aute              |

### 5.1 Analýza nezávislých podpriestorov (ISA)

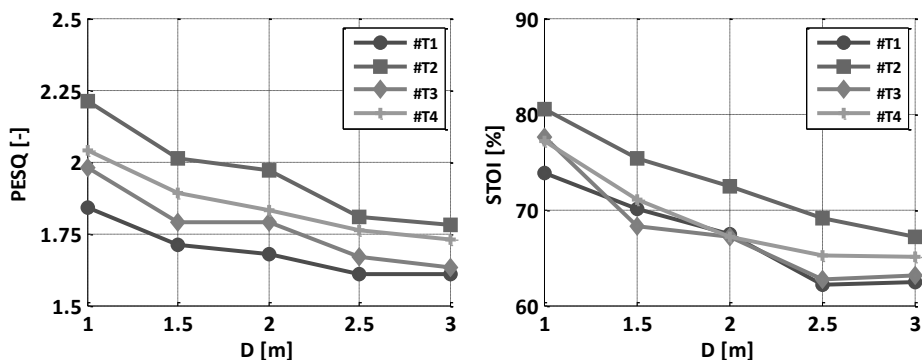


Obr. 5.1: Výkonnosť ISA v závislosti od typu maskovacieho signálu a jeho hodnoty.





Obr. 5.2: Výkonnosť ISA v závislosti od času dozvuku pre rôzne typy maskovacieho signálu.

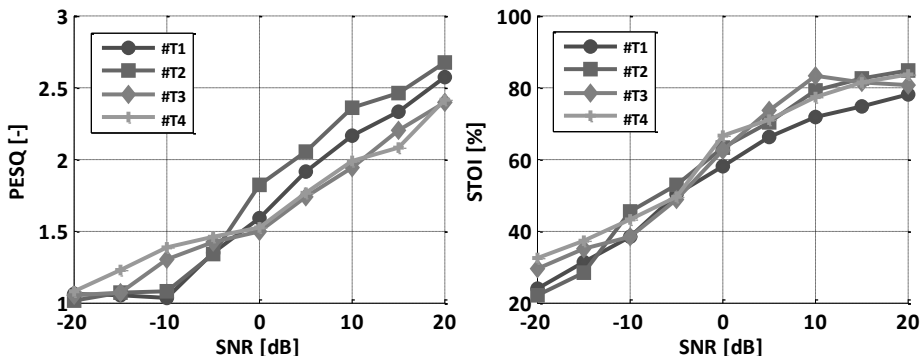


Obr. 5.3: Výkonnosť ISA v závislosti od vzdialenosti pre rôzne typy maskovacieho signálu.

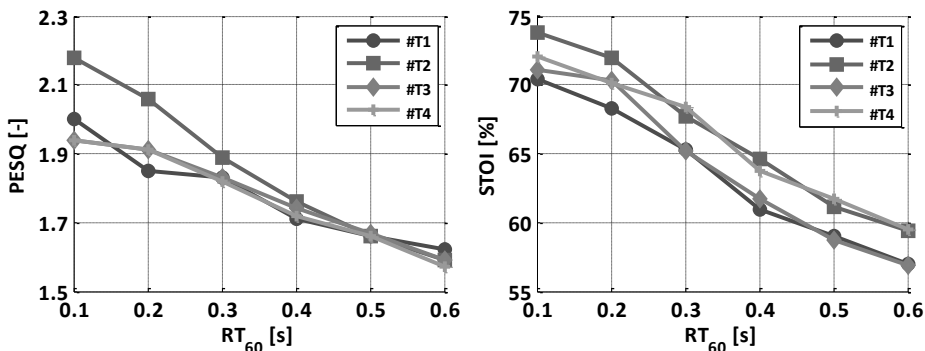
Metóda ISA využíva štatistickú nezávislosť zmiešaných signálov, čo je aj dôvod prečo pri všetkých experimentoch bola najnižšia kvalita zistená v prípade maskovania bľabotom. Naopak väčšinou najlepšie výsledky vykazovala reč maskovaná rečou iného rečníka, pričom maskovaný signál bol zámerne zvolený tak, aby bol čo najmenej zhodný s pôvodným. Pri analýzach zameraných na akustiku miestnosti sa prejavil teoretický limit tejto metódy. Nakoľko aplikovaním impulznej odozvy miestnosti boli do analyzovaného signálu zavedené odrazy, ktoré sú štatisticky zhodné s pôvodným signálom. Separačná metóda teda predpokladá, že tieto odrazy sú súčasťou pôvodného signálu, a z toho dôvodu klesá úspešnosť separácie priamoúmerne s nárastom odrazov. Celkovo však dosiahnuté výsledky nevykazovali výraznú úspešnosť (efektívnosť) separácie signálov. Výhodou tejto metódy je fakt, že dokáže pracovať len s jedným kanálom (nie je potreba využívať mikrofónové polia) a nevyžaduje žiadne informácie o pôvode signálov a procese zmiešania

## 5.2 Nezáporná maticová faktorizácia (NMF)

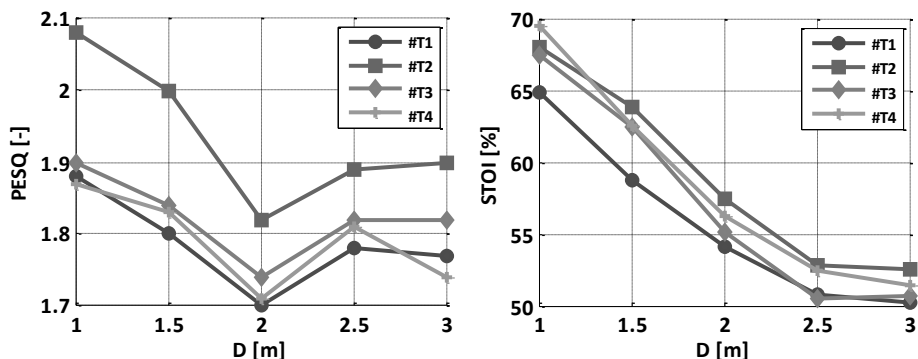
Pri prvom pohľade na dosiahnuté výsledky môžeme poznamenať, že metóda vykazuje podobné vlastnosti ako ISA. Výkonnosť metódy klesá takmer priamo úmerne so znižujúcou sa kvalitou vstupného signálu. Ak sa zameriame na typ maskovacieho signálu, môžeme pozorovať, že opäť najlepšie výsledky podľa očakávania sú dosiahnuté pre reč rečníka, keďže tento maskovací signál bol zámerne zvolený tak, aby bol čo najviac odlišný od vstupného signálu. Celkovo je však výkonnosť tejto metódy nižšia ako pri aplikovaní ISA a od úrovne SNR = -10 dB a nižšie je separovaný signál pomerne degradovaný. Podobný trend vykazuje aj test pri aplikovaní impulzovej odozvy miestnosti. Domnievame sa, že pokles kvality separovaného signálu pre 2 metrovú vzdialenosť medzi rečníkom a mikrofónom je spôsobený zložitou voľbou báзовých vektorov, avšak iba v prípade PESQ metriky.



Obr. 5.4: Výkonnosť NMF v závislosti od typu maskovacieho signálu a jeho hodnoty.



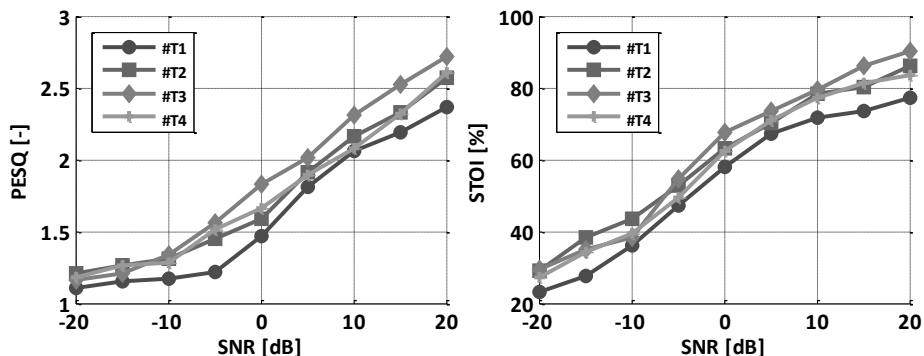
Obr. 5.5: Výkonnosť NMF v závislosti od čas dozvuku pre rôzne typy maskovacieho signálu.



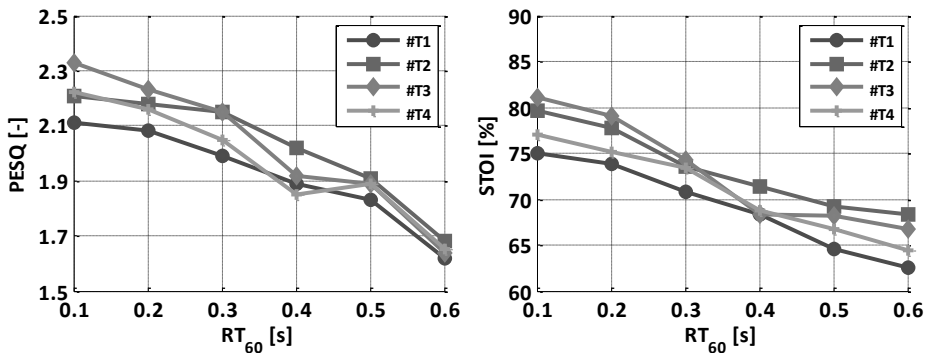
Obr. 5.6: Výkonnosť NMF v závislosti od vzdialenosti pre rôzne typy maskovacieho signálu.

### 5.3 Vektorová kvantizácia (VQ)

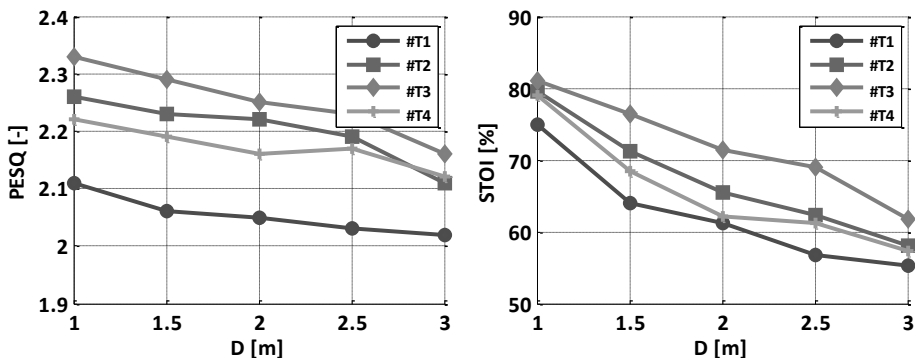
Táto separačná metóda je založená na generovaní modelov jednotlivých rečníkov. Ani pri tejto metóde neexistuje jednoznačné pravidlo ako dosiahnuť najvyššiu kvalitu výstupného signálu, keďže v celkom procese je zahnutých niekoľko činiteľov, ktoré sa môžu prejaviť a je potrebné s nimi počítať. K dosiahnutým výsledkom môžeme povedať, že celková kvalita výstupného signálu je porovnateľná s predchádzajúcim metódami, hoci patrili do inej kategórie. Najvýznamnejší rozdiel oproti predchádzajúcej metóde bol v tom, že ak bola vstupná reč maskovaná hlukom, kvalita výstupného signálu bola vyššia ako pri maskovaní ľudskou rečou. Tento fakt bol pravdepodobne spôsobený tým, že hoci sa jednalo o reč iného rečníka, jej model bol oveľa zhodnejší s modelom pôvodnej reči ako modely hluku. Najmenej významný vplyv na kvalitu výstupného signálu má pri tejto metóde vzdialenosť rečníka a poslucháča.



Obr. 5.7: Výkonnosť VQ v závislosti od typu maskovacieho signálu a jeho hodnoty.



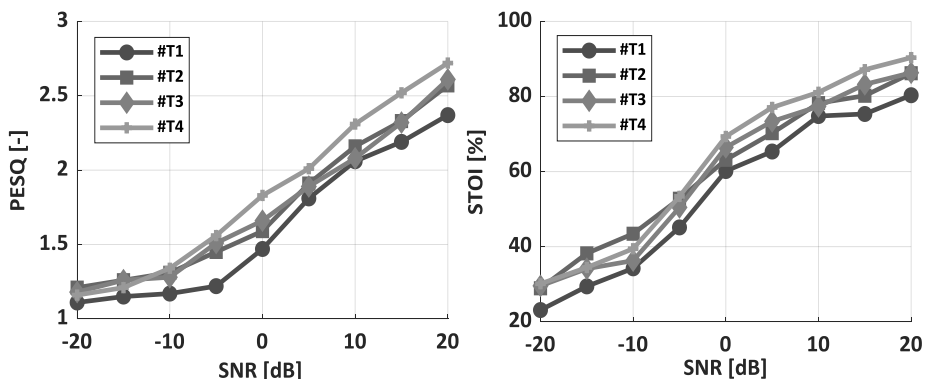
Obr. 5.8: Výkonnosť VQ v závislosti od času dozvuku pre rôzne typy maskovacieho signálu.



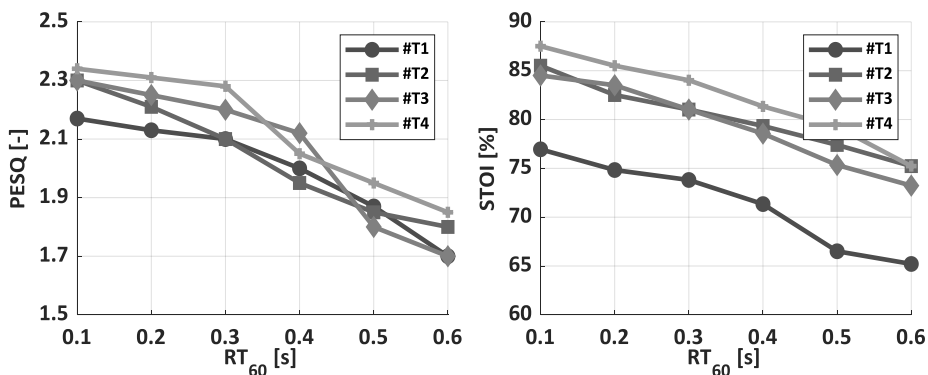
Obr. 5.9: Výkonnosť VQ v závislosti od vzdialenosti pre rôzne typy maskovacieho signálu.

## 5.4 Separačná metóda na báze GAN

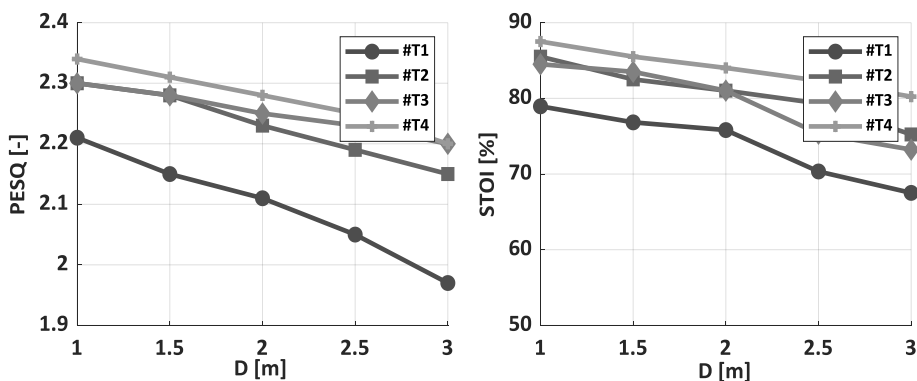
Hlavným problémom systémov na báze hlbokého alebo strojového učenia je skutočnosť, že neexistuje reálne pravidlo či postup ako dosiahnuť minimálnu chybu výstupu, respektíve maximálnu “účinnosť”. Ak porovnáme dosiahnuté výsledky z prvého testu môžeme vidieť, že separačná metóda na báze GAN mierne prekonáva ostatné prezentované metódy. Na rozdiel od ostatných metód však ponúka lepšie výsledky pre vstupnú reč, ktorá je maskovaná ľudskou rečou. Tento fakt pripisujeme spôsobu akým bola neurónová sieť trénovaná, pretože jej úlohou počas tohto procesu bolo správne separovanie reči poškodenej práve týmto spôsobom. Pri analýze času dozvuku miestnosti je možné pozorovať podobný trend ako v predchádzajúcich experimentoch, čo nám napovedá, že GAN sieť nie je imúnna voči tomuto typu poruchy. To isté platí aj pri analýze vplyvu vzdialenosti medzi mikrofónom a rečníkom.



Obr. 5.10: Výkonnosť GAN v závislosti od typu maskovacieho signálu a jeho hodnoty.



Obr. 5.11: Výkonnosť GAN v závislosti od doby dozvuku pre rôzne typy maskovacieho signálu.



Obr. 5.12: Výkonnosť GAN v závislosti od vzdialenosti pre rôzne typy maskovacieho signálu.

## 5.5 Porovnanie separačných metód

Kvalita, respektíve výkonnosť separačnej metódy je vo všeobecnosti nepriamo úmerná všestrannosti aplikovanej metódy. Z toho dôvodu, pri voľbe správnej separačnej metódy musíme zohľadniť preferencie ako aj konkrétne požiadavky navrhovaného systému. Prehľadné porovnanie separačných metód na základe dosiahnutých výsledkov je znázornené v tabuľke 5.2, v ktorej je pre úplnosť zaradená aj metóda ICA hoci sa jedná o viackanálovú metódu. Vzhľadom na dosiahnuté výsledky a publikované závery, za najperspektívnejšiu metódu separácie považujeme v práci navrhnutú metódu na báze GAN sietí, hoci je potrebné počítať s nutnou podmienkou dostatočného množstva tréningových dát.

Tab. 5.2: Zhrnutie výhod a nevýhod metód separácie rečových signálov.

| Metóda | Typ                                                               | Výhody                                                                                                          | Nevýhody                                                                                                                                       |
|--------|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| ICA    | <ul style="list-style-type: none"> <li>• Viackanálová</li> </ul>  | <ul style="list-style-type: none"> <li>• Efektívnosť</li> </ul>                                                 | <ul style="list-style-type: none"> <li>• Potreba viacerých kanálov</li> <li>• Vstupné dáta musia byť plne definované</li> </ul>                |
| ISA    | <ul style="list-style-type: none"> <li>• Jednokanálová</li> </ul> | <ul style="list-style-type: none"> <li>• Mierne vyššia efektívnosť ako NMF</li> </ul>                           | <ul style="list-style-type: none"> <li>• Výpočtová náročnosť</li> <li>• Nejednoznačnosť optimálneho nastavenia</li> </ul>                      |
| MNF    | <ul style="list-style-type: none"> <li>• Jednokanálová</li> </ul> | <ul style="list-style-type: none"> <li>• Nižšia výpočtová náročnosť ako ISA</li> <li>• Univerzálnosť</li> </ul> | <ul style="list-style-type: none"> <li>• Náročnosť zoskupovania bazových vektorov</li> <li>• Nejednoznačnosť optimálneho nastavenia</li> </ul> |
| VQ     | <ul style="list-style-type: none"> <li>• Jednokanálová</li> </ul> | <ul style="list-style-type: none"> <li>• Jednoduchá implementácia</li> <li>• Rýchlosť výpočtu</li> </ul>        | <ul style="list-style-type: none"> <li>• Závislosť na kvalite vzoriek</li> <li>• Nejednoznačnosť optimálneho nastavenia</li> </ul>             |
| GAN    | <ul style="list-style-type: none"> <li>• Jednokanálová</li> </ul> | <ul style="list-style-type: none"> <li>• Kvalita výstupných vzoriek</li> </ul>                                  | <ul style="list-style-type: none"> <li>• Závislosť na kvalite vzoriek</li> <li>• Náročný proces tréningovania</li> </ul>                       |

## 6 Zhrnutie dosiahnutých výsledkov a prínosov

Hlavné prínosy dizertačnej práce z hľadiska rozvoja a napredovania príslušného vedného odboru, ako aj praktického využitia metód pre separáciu rečových signálov ako aj ďalších dosiahnutých výsledkov nášho výskumu sú nasledovné:

- Prínosom tejto práce v oblasti vyhodnocovania kvality separovaných rečových signálov je vyšetrenie a porovnanie vlastností a výkonnosti existujúcich metrík z hľadiska schopnosti odzrkadľovať vplyv okolitého prostredia na signál spracovaný separačným algoritmom.
- Jedným z najdôležitejších prínosov je využitie GAN siete v inovatívnej metóde separovania rečových signálov, nakoľko doteraz bol známy a publikovaný iba základný koncept tohto prístupu. Dosiahnuté výsledky dávajú veľký príslub, že práve využitie metód hlbokého učenia v procese separácie reči pomôže významne zlepšiť kvalitu spracovanej reči.
- Nezanedbateľným prínosom je taktiež vykonaná štúdia zameraná na monitorovanie procesu tréningu a voľby tréningovej sady za cieľom dosiahnutia optimálneho výkonu GAN separátora s ohľadom na čas potrebný pre natréningovanie siete. Dosiahnuté poznatky a výsledky môžu byť následne využité pri voľbe optimálneho nastavenia a architektúry siete.
- Dôležitým aspektom práce je vyšetrenie zohľadnenia vplyvu okolitého prostredia, nakoľko vo väčšine existujúcich prác sa hodnotenie výkonnosti separačných algoritmov obmedzuje iba na analýzu vplyvu typu a hodnoty maskovacieho signálu. V tejto práci boli prezentované aj výsledky zohľadňujúce čas dozvuku, prípadne vzdialenosť medzi mikrofónom a rečníkom.
- Bola vykonaná analýza a porovnanie separačnej metódy na báze GAN s konvenčnými metódami, ktoré boli zvolené tak, aby pokrývali široké spektrum prístupov využívaných pre separáciu rečových signálov.
- Pre potreby analýzy jednotlivých separačných prístupov bola vykonaná ich implementácia a následne integrovanie do knižnice, čo umožňuje ich využitie pre potreby ďalšieho výskumu.

## 7 Záver

Predložená dizertačná práca sa zaoberá metódami separovania rečových signálov. Ako bolo v práci preukázané, tento prístup umožňuje či už extrakciu reči dominantného rečníka alebo rozdelenie vstupného signálu podľa príspevkov jednotlivých rečníkov. Za týmto účelom bolo vyvinutých a publikovaných viacero prístupov. Stretávame sa s metódami založenými na štatistickej nezávislosti jednotlivých zdrojov, tvorbe modelov rečníkov, prípadne metódami inšpirovanými ľudským sluchovým ústrojenstvom. Každá z nich má určité špecifiká ako aj výhody či nevýhody, ktoré sa následne prejavujú aj na kvalite spracovaného signálu. Cieľom tejto práce bolo rozšíriť poznanie o týchto metódach a poukázať na ich silné a slabé stránky s ohľadom na vplyv prostredia, v ktorom sa reč šíri.

Jadrom práce je implementácia a porovnanie výkonnosti konvenčných metód a novej navrhutej metódy na báze hlbokého učenia. Pred samotným porovnaním však skúmame možnosti vyhodnocovania kvality reči s cieľom čo najlepšie odzrkadliť reálne podmienky. Za týmto účelom sme vykonali analýzu rôznych metrík, ktoré sme aplikovali na reč spracovanú ideálnou binárnou maskou a zámerne vystavili rôznym druhom hluku a interferencií. Na jej základe sme zvolili dva príznaky (PESQ, STOI), ktoré sme sa rozhodli využívať na porovnanie jednotlivých separačných metód, nakoľko najlepšie odzrkadľovali reálne podmienky. V nasledujúcej časti sme implementovali a porovnali niekoľko konvenčných metód so separačnou metódou na báze GAN, ktorú sme navrhli v rámci práce. Tento inovatívny prístup v niektorých prípadoch ako je typ maskovacieho signálu alebo množstvo odrazov signálu, dokázal prekonať konvenčné metódy, avšak za cenu času potrebného na natrénovanie siete. Dosiahnuté výsledky v rámci vykonaného výskumu poukázali aj na fakt, ako môžu parametre okolitého prostredia ovplyvniť kvalitu separovaného signálu.

Nakoľko sa oblasť hlbokého učenia neustále rozvíja, pričom vzniká veľké množstvo nových alebo vylepšených metód a prístupov, radi by sme náš ďalší výskum zamerali práve týmto smerom. Rovnako ako v tejto práci má ich aplikovanie smerovať k zlepšeniu výkonnosti separačných metód s ohľadom na analýzu vplyvu okolitého prostredia. Príkladom môže byť napríklad Wasserstein GAN, ktorá v porovnaní so štandardnou GAN vykazuje lepšiu konvergenciu tréningu a redukciu problému s miznúcim gradientom.



## 8 Summary

The presented dissertation deals with methods of speech signal separation. This approach allows either the extraction of the dominant speaker speech or the separation of the input signal according to the contributions of the individual speakers. There are methods based on the statistical independence of individual sources, the creation of speaker models, or methods inspired by human hearing aids. Each of them has its own specificities as well as advantages or disadvantages that affect the quality of the processed signal. The aim of this work was to extend the knowledge about these methods and to point out their strengths and weaknesses with regard to the influence of the environment in which speech is spread.

The core of the work is the implementation and comparison of the performance of conventional methods and the proposed method based on deep learning. Methods of speech quality evaluation in order to best reflect real conditions were investigated as the first part of our work. We analyzed the various metrics that we applied to speech processed by an ideal binary mask and deliberately exposed to various types of noise and interference for this purpose. Based on this, we chose two metrics (PESQ, STOI) that we decided to use in our next research. In the following section, we implemented and compared several conventional GAN-based separation methods that we proposed in the work. This innovative approach was able to overcome conventional methods in some cases, such as the type of masking signal or the amount of signal reflections. The drawback is computational time needed to train the network. The results obtained in our research have also highlighted how the parameters of the surrounding environment can affect the quality of the separated signal.

As the field of deep learning is constantly evolving, we would like to focus our further research in this direction. As in this work, their application is intended to improve the performance of separation methods with respect to the environmental impact analysis. An example may be, for example, Wasserstein GAN, which shows better co-convergence compared to the standard GAN, and reducing the problem with the disappearing gradient.

## Zoznam publikácií autora

- [VS01] Vladimír Sedlák, Daniela Ďuračková, Roman Zálusky, Tomáš Kováčik. Intelligibility assessment of ideal binary-masked noisy speech with acceptance of room acoustic. *Journal of Electrical Engineering*. Vol. 65, No. 6, s. 325-332, 2014.
- [VS02] Juraj Brenkuš, Viera Stopjaková, Viera Čerňanová, Daniel Arbet, Lukáš Nagy, Vladimír Sedlák. A novel method towards time-efficient fault analysis of analog and mixed-signal circuits. *Journal of Circuits Systems and Computers*. Vol. 26, No. 8, s.120-131, 2017.
- [VS03] Vladimír Sedlák, Viera Stopjaková, Juraj Brenkuš. A real-time method for smoke detection in monitored forest areas. *Applied electronics 2017*. s. 162-165. 2017.
- [VS04] Roman Zálusky, Daniela Ďuračková, Vladimír Sedlák. Reduction of Chip Area for Feed-Forward Neural Networks with Use the Special Multiplication by and Gate. *IN-TECH 2012 : Proceedings of International Conference on Innovative Technologies*. s.83-86, 2012.
- [VS05] Roman Zálusky, Daniela Ďuračková, Viera Stopjaková, Lukáš Nagy, Vladimír Sedlák. Novel architecture of a digital neuron for FFNN employing special multiplication. *ECAI 2014 : 21st European Conference on Artificial Intelligence*. s. 933-938, 2014.
- [VS06] Marcel Černák, Daniela Ďuračková, Tomáš Kováčik, Vladimír Sedlák, Roman Zálusky. Image segmentation methods. *APCOM 2014. Applied Physics of Condensed Matter*. s. 328-331, 2014.
- [VS07] Marcel Černák, Daniela Ďuračková, Tomáš Kováčik, Vladimír Sedlák, Roman Zálusky. Image segmentation methods survey. *ELITECH'14 [elektronický zdroj] : 16th Conference of Doctoral Students*. str [6] s. 2014.
- [VS08] Tomáš Kováčik, Daniela Ďuračková, Vladimír Sedlák, Roman Zálusky. Accurate Time Measure on Spartan 3E FPGA. *APCOM 2013. Applied Physics of Condensed Matter*. s.262-265, 2013.
- [VS09] Tomáš Kováčik, Daniela Ďuračková, Vladimír Sedlák, Roman Zálusky, Marcel Černák. Implementation of sobel edge detector in to FPGA. *APCOM 2014. Applied Physics of Condensed Matter*. s. 332-335, 2014.

- [VS10] Tomáš Kováčik, Daniela Ďuračková, Marcel Černák, Vladimír Sedlák. Skin color detection in YCbCr and HSV color space. *ELITECH'14*, str [6] s. 2014.
- [VS11] Tomáš Kováčik, Daniela Ďuračková, Vladimír Sedlák, Roman Záluský. Handwriting recognition on image. *APCOM 2015*. str. 360-364, 2015.
- [VS12] Vladimír Sedlák, Daniela Ďuračková, Roman Záluský. Detection of Signals in Noisy Environment. *APCOM 2012. Applied Physics of Condensed Matter*. str. 333-336. 2012.
- [VS13] Vladimír Sedlák, Daniela Ďuračková. Methods for Speech Visualization. *ELITECH'11 : 13th Conference of Doctoral Students*, str.1-4, 2011.
- [VS14] Vladimír Sedlák, Daniela Ďuračková. Speech Signals Visualization. *APCOM 2011*. str.316-319, 2011.
- [VS15] Vladimír Sedlák. Adaptive Filters for Noise Canceling. *Počítačové architektúry a diagnostika PAD 2011*, str.50-54, 2011.
- [VS16] Vladimír Sedlák, Daniela Ďuračková, Roman Záluský. Investigation Impact of Environment for *Performance* of ICA for Speech Separation. *Elektro 2012 : 9th International Conference*, str.89-93, 2012.
- [VS17] Vladimír Sedlák, Daniela Ďuračková, Roman Záluský. Performance Comparison of Adaptive Filters and Neural Networks for Noise Cancelling. *ELITECH'12 [elektronický zdroj] : 14th Conference of Doctoral Students*, str. [3] s. 2012.
- [VS18] Vladimír Sedlák, Daniela Ďuračková, Tomáš Kováčik, Roman Záluský. Quality Assessment for Single Channel Source Separation. *APCOM 2013*, str. 270-273, 2013.
- [VS19] Vladimír Sedlák, Daniela Ďuračková, Tomáš Kováčik, Roman Záluský. Simulation and Performace Analysis of Model-Based Single Channel Speech Separation. *ADEPT 2013 : 1st International Conference on Advances in Electronic and Photonic Technologies*, str. 246-249, 2013.
- [VS20] Vladimír Sedlák, Daniela Ďuračková, Tomáš Kováčik. Investigation of an impact of room acoustics on performance of ideal binary mask. *ELEKTRO 2014 : 10th International Conference*, str. 90-93, 2014.
- [VS21] Vladimír Sedlák, Daniela Ďuračková, Roman Záluský, Tomáš Kováčik, Marcel Černák, Frank Schwierz. A survey on separation methods for quality enhancement of affected signals. *APCOM 2014. Applied Physics of Condensed Matter*, str. 194-197, 2014.

- [VS22] Vladimír Sedlák, Daniela Ďuračková, Roman Záluský, Tomáš Kováčik. An experimental survey on non-negative matrix factorization for separation of signals. *APCOM 2015*, str. 314-317, 2015.
- [VS23] Roman Záluský, Daniela Ďuračková, Vladimír Sedlák. The Effectiveness of Dactyl Alphabet Recognition of Neural Network with New Architecture. *APCOM 2012*, str. 337-340, 2012.
- [VS24] Roman Záluský, Mario Krajmer, Vladimír Sedlák, Daniela Ďuračková. The Recognition Characters for Dactyl Alphabet with use the Feed Forward Neuralnetwork. *APCOM 2011*, str. 320-323, 2011.
- [VS25] Roman Záluský, Daniela Ďuračková, Vladimír Sedlák, Tomáš Kováčik. The Use of Neural Network for Data Encryption Standard (DES). *APCOM 2013*, str. 266-269, 2013.
- [VS26] Roman Záluský, Daniela Ďuračková, Viera Stopjaková, Juraj Brenkuš, Jozef Mihálov, Libor Majer, Vladimír Sedlák, Tomáš Kováčik, Marcel Černák Parametric test of antennas using the neural networks. *APCOM 2014*, str. 190-193, 2014.
- [VS27] Roman Záluský, Daniela Ďuračková, Viera Stopjaková, Daniel Arbet, Lukáš Nagy, Vladimír Sedlák, Tomáš Kováčik. Automated evaluation and test of readout interface for MEMS microphone prototype chips. *APCOM 2015*, str. 212-215, 2015.

## Zoznam literatúry použitej v autoreferáte

- [1] J. Feng, B. Ramabhadran, J. Hansen a J. D. Williams, Trends in Speech and Language Processing, *IEEE Signal Processing Magazine*, zv. 29, 1. vyd.1, pp. 177-179, 2012.
- [2] E. Grais a H. Ergodan, Single channel speech separation using nonnegative matrix factorization and spectral masks, *Proceeding of International Conference on DSP*, Corfu, 2011.
- [3] N. Hamdoui a A. Abid, Single mixture audio source separation using ISA technique in EMD domain, *Proceedings of International Symposium on I/V Communications and Mobile Network*, Rabat, 2010.
- [4] M. Shashanka a B. Smaragdis, Sparse overcomplete decomposition for single channel speaker separation, *Proceedings of ICASSP 2007*, Honolulu, 2007.
- [5] R. Weiss a D. Ellis, Monaural speech separation using source-adapted models, *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, 2007.
- [6] K. Wilson, Speech source separation by combining localization cues with mixture models of speech spectra, *Proceedings of ICASSP 2007*, Honolulu, 2007.
- [7] T. Beierholm, B. Pedersen a O. Winther, Low complexity bayesian single channel source separation, *Proceedings of ICASSP 2004*, Montreal, 2004.
- [8] L. Gu a R. Stern, Single-channel speech separation based on modulation frequency, *Proceedings of ICASSP 2008*, Las Vegas, 2008.
- [9] A. N. a. D. W. Y. Wang, On training targets for supervised speech separation, *IEEE/ACM Trans. Audio Speech Lang. Proc.*, zv. 22, 1. vyd.1, pp. 1849-1858, 2014.
- [10] Y. T. S. M. a. C. H. X. Lu, Speech enhancement based on deep denoising autoencoder, *Proceedings of Interspeech*, Paris, 2013.
- [11] J. H. S. W. a. J. L. R. H. Erdogan, Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks, *Proceedings of ICASSP*, London, 2015.
- [12] L. Hui, Convolutional maxout neural networks for speech separation, *Proceedings of ISSPIT*, Boston, 2015.
- [13] E. Habets, Room Impulse Response Generator for MATLAB, 2010. [Online]. Available: [http://home.tiscali.nl/ehabets/rir\\_generator.html](http://home.tiscali.nl/ehabets/rir_generator.html). [Cit. 1 9 2016].
- [14] S. Pascual, A. Bonafonte a J. Serra, SEGAN: Speech enhancement generative adversarial network, arXiv:1703.09452v3, 2017.