

VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra kybernetiky a biomedicínského inženýrství

Využití Soft Computingových metod pro zpracování řečového signálu
v reálném prostředí Smart Home
Speech Signal Processing in Application of Soft Computing Methods

Zadání diplomové práce

Student: **Bc. Tomáš Weiper**

Studijní program: N2649 Elektrotechnika

Studijní obor: 3901T009 Biomedicínské inženýrství

Téma: **Využití Soft Computingových Metod pro zpracování řečového signálu v reálném prostředí Smart Home**
Speech Signal Processing in Application of Soft Computing Methods

Jazyk vypracování: čeština

Zásady pro vypracování:

Pro komfortní ovládání automatizace inteligentního domu určeného pro podporu nezávislého života seniorů v domácím prostředí je seniory akceptována možnost hlasového ovládání provozně technických funkcí. Pro vlastní zpracování řečových signálů v reálném prostředí Smart Home lze využít vhodných Soft Computingových metod. Proveďte implementaci Umělých Neuronových sítí pro zpracování řečových signálů s aditivním šumem s využitím standardu ITU-T P863 pro klasifikaci kvality zpracování řečového signálu.

Cíl:

Porovnání simulace a praktické realizace zpracování řečového signálu pomocí Umělých Neuronových sítí s využitím standardu ITU-T P863.

Body zadání:

1. Rešerše současného stavu využití Soft Computingových metod pro zpracování řečového signálu.
2. Simulace Neuronových sítí pro zpracování řečového signálu s aditivním šumem v Matlabu (klasifikace kvality zpracování podle ITU-T P863).
3. Praktická implementace Neuronových sítí v reálném prostředí (klasifikace kvality zpracování podle ITU-T P863).
4. Porovnání výsledků simulace a praktické realizace - diskuze.
5. Celkové zhodnocení závěrečné práce.

Seznam doporučené odborné literatury:

- [1] TUČKOVÁ, Jana. *Vybrané aplikace umělých neuronových sítí při zpracování signálů*. Praha: ČVUT v Praze, 2009. ISBN 976-80-01-04229-8.
- [2] Jiří, Jan. *Číslicová filtrace, analýza a restaurace signálů*. Brno: VUTIUM, 2002. ISBN 80-214-1558-4
- [3] VASEGHI, Saeed V. *Advanced digital signal processing and noise reduction*. 3rd ed. Chichester: Wiley, c2006. ISBN 0-470-09494-X.
- [4] OGUNFUNMI, Tokunbo, Roberto TOGNERI a Madihally NARASIMHA. (eds.) *Speech and audio processing for coding, enhancement and recognition*. New York: Springer, 2015. ISBN 978-1-4939-1455-5.
- [5] MERZ, Hermann, Hansemann THOMAS a Christof HÜBNER. *Automatizované systémy budov: sdělovací systémy KNX/EIB, LON a BACnet*. 1. vyd. Přeložil Václav BARTOŠ. Praha: Grada Publishing, 2009. ISBN 978-80-247-2367-9.
- [6] VAŇUŠ, Jan. *Řízení provozu budov, učební text*. Ostrava: VŠB-TU Ostrava, 2013.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Jan Vaňuš, Ph.D.**

Datum zadání: 01.09.2016

Datum odevzdání: 28.04.2017



doc. Ing. Jiří Koziorek, Ph.D.
vedoucí katedry




prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Prohlášení

„Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.“

Bc. Tomáš Weiper

V Ostravě, dne 28. dubna 2017

Podpis: 

Poděkování

Rád bych tímto poděkoval panu Ing. Janu Vaňušovi Ph.D., za profesionální přístup, ochotu, vstřícnost při konzultacích a za pomoc v laboratoři při vypracování této diplomové práce.

Dále bych chtěl poděkovat paní Ing. Janě Belešové, za poskytnuté informace k její aplikaci, pomoc uvést ji do provozu a za možnost navázat na její práci.

Abstrakt

Cílem této diplomové práce je prozkoumat odvětví Soft Computingových metod v oblasti zpracování řečového signálu, nalézt vhodnou metodu pro potlačení šumu a tu následně použít v simulaci a praktické realizaci. Vybraná metoda bude popsána teoreticky i matematicky a její implementace bude použita pro potlačení okolního šumu z hlasového signálu pro ovládání chytré domácnosti Smart Home založené na sběrnicovém systému KNX. Hlasové řízení provozně technických funkcí tohoto komunikačního sběrnicového systému je předpokladem pro jednodušší správu domácnosti, vylučující jinak nutnou manuální manipulaci s ovládacím zařízením a to zejména pro seniory nebo postižené osoby. V domácnostech se ovšem nachází řada rušivých elementů, mezi které můžeme zařadit například šum domácích spotřebičů, nebo vlivy počasí, které mohou zapříčinit nefunkčnost tohoto ovládání a je tedy potřeba je odstranit. V dnešní době je k dispozici již množství filtrů schopných uspokojivě odstranit předem specifikovaný šum, avšak použití nelineárních adaptivních metod by mohlo tyto výsledky posunout na zcela jinou úroveň. Po úspěšné implementaci je nutné provést zhodnocení signálu, zda byla použitá metoda v potlačení šumu úspěšná.

Klíčová slova

Soft Computingové metody, zpracování řečového signálu, neuronové sítě, potlačení šumu, Smart Home, KNX, klasifikace řeči, hlasové ovládání

Abstract

The aim of this diploma thesis is investigate the field of Soft Computing methods in Speech Signal Processing and find an appropriate method for noise suppression and use it in simulation and practical implementation. The chosen method will be theoretically and mathematically described and used in implementation to suppress the ambient noise from speech signal for controlling a Smart Home based on the KNX bus system. Voice control of the operational technical features of this communication bus system is a prerequisite for a simpler household management, eliminating otherwise necessary manual handling of the control device, especially for seniors or disabled persons. However, in households there are a number of disturbing elements, including for example the noise of household appliances or weather conditions that may cause this control to malfunction and noise need to be removed. Nowadays, a number of filters are already available to satisfactorily remove pre-specified noise, but the use of non-linear adaptive methods could shift this results to a completely different level. After successful implementation, it is necessary to evaluate the signal if the used noise suppression method has been successful.

Key words

Soft Computing methods, speech signal processing, neural networks, noise suppression, Smart Home, KNX, classification of speech, voice control

Obsah

Seznam použitých zkratk	8
Seznam použitých symbolů	10
Seznam ilustrací	12
Seznam tabulek	13
Úvod	14
1 Současný stav využití Soft Computingových metod pro zpracování řečového signálu	15
2 Soft Computing	22
2.1 Soft Computingové metody	22
2.1.1 Fuzzy počítání	22
2.1.2 Genetické algoritmy (evoluční počítání)	23
2.1.3 Bayesovské sítě	24
2.1.4 Data mining	24
2.1.5 Neuronové sítě	25
3 Typy neuronových sítí	26
3.1 ADALINE	26
3.2 Neuronová síť s opakujícími se vrstvami	26
3.3 Vícevrstevný perceptron	26
3.4 ANFIS	26
4 Adaptivní neuro-fuzzy inferenční systém (ANFIS)	27
4.1 Použití Neuro-Adaptivní učení	27
4.2 FIS struktura a nastavení parametrů	27
4.3 Vstupní data	28
4.4 Architektura a základní pravidlo učení	28
5 Řeč za přítomnosti šumu	31
5.1 Typy šumů	31
5.1.1 Aditivní šumy	31
5.1.2 Konvoluční šumy	32
5.1.3 Měření úrovně aditivního šumu v řeči	32
5.1.4 Metody odhadu SNR řeči	34
6 Norma ITU-T P.863	36
6.1 Algoritmus ITU-T P.863	39
6.2 Algoritmus PESQ	42

7	Střední kvadratická chyba	43
8	Smart Home	44
8.1	Bytová výstavba	44
8.2	Účelové stavby	44
8.3	KNX	45
9	Software Hestia pro ovládání standardu KNX	46
9.1	Testování aplikace v reálném prostředí	47
9.1.1	Simulační sestava KNX	47
9.1.2	Výzkumné a inovační centrum	47
9.2	Experiment rozpoznání příkazů	48
10	Navržení softwarové realizace adaptivního programu	54
10.1	Návrh algoritmu	54
10.2	Simulace	56
10.3	Reálná implementace	62
11	Diskuze	64
	Závěr	66
	Literatura	I
	Seznam příloh	IV
	Příloha A	V
	Příloha B	IX
	Příloha C	XIX
	Příloha D	XXVI

Seznam použitých zkratk

AANCC	Adaptivní aktivní potlačení šumu
ABR	Rychlost odpovědi mozkového kmene na lidskou řeč
ADALINE	Adaptivní lineární neuron
AGC	Automatické zesílení ovládání
ANC	Aktivní regulace šumu
ANFIS	Adaptivní neuro-fuzzy inferenční systém
ANN	Umělé neuronové sítě
ASNR	Aritmetický segmentální poměr signál-šum
ASR	Automatické rozpoznání mluvčího
ATF	Adaptivní časově-frekvenční parameter
AVSR	Audiovizuální rozpoznání řeči
DEKF	Rozšířený duální Kalmanův filtr
DNN	Hluboké neuronové sítě
DPF	Diskriminativní post filtry
DTX	Přerušované vysílání
FAMNN	Neuronová síť Fuzzy Artmap
FF-ANN	Vícevrstevná umělá neuronová síť dopředného typu
FFNN	Zesílená dopředná neuronová síť
FIS	Fuzzy inferenční systém
GA	Genetický algoritmus
GSNR	Globální poměr signál-šum
GUI	Grafické uživatelské rozhraní
HMM	Skrytý Markovův model

LMS	Metoda nejmenších čtverců
LRN	Neuronová síť s opakujícími se vrstvami
MBSS	Adaptivně kompenzovaná vícepásmová spektrální subtrikce
MLP	Vícevrstevný perceptron
MOS	Průměrný posudek výsledku
NMF	Nezáporná faktorizace matic
NR	Redukce šumu
PESQ	Percepční posuzování kvality řečového signálu
RMSE	Střední kvadratická chyba
RQNN	Opakující se kvantová neuronová síť
RTF	Vylepšený časově-frekvenční parameter
SC	Soft computing
SDR	Poměr signál-zkreslení
SNR	Poměr signál-šum
SSNR	Segmentální poměr signál-šum
VAC	Detekce hlasové aktivity
VED	Zařízení pro zesílení hlasu
VŠB	Vysoká škola báňská
WHO	Světová zdravotnická organizace

Seznam použitých symbolů

$\hat{\sigma}_n^2$	Odhad výkonu šumu
E_p	Míra chyby cílového výstupního vektoru
O_i^k	Klasifikátor funkce a výstupu uzlu
σ_n^2	Výkon aditivního šumu
σ_s^2	Partikulární výkon řeči
Σ	Kovarianční matice rozkladu
A	Zobrazení definující aktivační funkci
E	Celková míra chyby
ex	Externí vstupní funkce
$H[k]$	Přenosová funkce signálu
l	Délka signálu (s)
N	Velikost populace
net	je zobrazení, definující vstupní funkci
O	Zobrazení definující výstupní funkci
$Rx[k]$	Autokorelace signálu
S	Soubor uzlů, jejichž výstupy jsou závislé na α
$s[n]$	Aditivní složka řeči
$s_i[n]$	Segmenty signálu
$Sx[k]$	Spektrální výkonová hustota signálu
U	Konečná množina výpočetních jednotek neuronu
$u[n]$	Aditivní složka šumu
$vad[n]$	Informace o řečové aktivitě pro daný vzorek
W	Struktura sítě neuronu

$X[k]$	Komplexní spektrum signálu
$x[n]$	Snímaný signál
α	Parametr dané adaptivní sítě
η	Parametr rychlosti učení
ϵ	Množina chybových vzorků
$\delta E_p / \delta \mathbf{O}$	Míra chyby pro p-tá tréninková data

Seznam ilustrací

Obr. 1: Příklad Bayesovské sítě [35].....	24
Obr. 2: Schéma biologického neuron [34]	25
Obr. 3: FIS - Fuzzy inferenční systém [38].....	28
Obr. 4: Adaptivní síť [38]	29
Obr. 5: Výskyt šumu při přenosu řeči [37].....	31
Obr. 6: Přehled základní filozofie použité v ITU-T P.863 [40]	40
Obr. 7: Všeobecný přehled algoritmu ITU-T P.863 [40]	41
Obr. 8: Informatické zasíťování zařízení systémové techniky budov sběrnici KNX/EIB [39].....	45
Obr. 9: Schéma zapojení softwaru Hestia	46
Obr. 10: Simulační sestava KNX systému	47
Obr. 11: Výzkumné a inovační centrum	48
Obr. 12: Zapojená sestava aplikace Hestia.....	48
Obr. 13: Blokové schéma navrženého systému.....	54
Obr. 15: Ukázka grafického uživatelského rozhraní programu.....	55
Obr. 15: Blokové schéma zapojení celé funkční struktury	56
Obr. 16: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem RMSE pro příkaz "light on"	57
Obr. 17: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem SSNR pro příkaz "light on"	58
Obr. 18: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem PESQ pro příkaz "light on"	59
Obr. 19: Časové (a), frekvenční (b) a časově frekvenční (c) spektrum příkazu "light on" se šumem (mf = 2, epoch = 2, type = "trimf")	60
Obr. 20: Časové (a), frekvenční (b) a časově frekvenční (c) spektrum referenčního příkazu "light on" (mf = 2, epoch = 2, type = "trimf")	60
Obr. 21: Časové (a), frekvenční (b) a časově frekvenční (c) spektrum příkazu "light on" s potlačeným šumem (mf = 2, epoch = 2, type = "trimf").....	61
Obr. 22: Časové (a), frekvenční (b) a časově frekvenční (c) spektrum šumu z příkazu "light on" (mf = 2, epoch = 2,	61

Seznam tabulek

Tab. 1: Příklad 1. Funkce [34].....	22
Tab. 2: Příklad 2. Funkce [34].....	23
Tab. 3: Faktory a aplikace zahrnující požadované specifikace použity ve výběrové fázi v algoritmu ITU-T P.863 [40]	36
Tab. 4: Algoritmus ITU-T P.863 není určen pro použití s následujícími proměnnými [40].....	38
Tab. 5: Testovací proměnné pro které je nutné další vyšetření [40]	38
Tab. 6: Faktory, technologie a aplikace, pro které ITU-T P.863 algoritmus nebyl v současné době potvrzen [40]	38
Tab. 7: Hodnoty MOS v algoritmu PESQ.....	42
Tab. 8: RMSE výsledky z náhodně generovaných nepravých chyb s nulovou střední hodnotou a jednotkovou odchylkou Gaussova rozdělení. [41].....	43
Tab. 9: Příkazy hlasového ovládání aplikace Hestia.....	46
Tab. 10: Úspěšnost rozpoznání pro 1. řečníka	49
Tab. 11: Úspěšnost rozpoznání pro 2. řečníka	50
Tab. 12: Úspěšnost rozpoznání pro 3. řečníka	51
Tab. 13: Úspěšnost rozpoznání pro 4. řečníka	52
Tab. 14: Průměrná úspěšnost rozpoznání pro všechny řečníky	53
Tab. 15: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem RMSE pro příkaz “light on”	57
Tab. 16: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem SSNR “light on”	58
Tab. 17: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem PESQ “light on”	59
Tab. 18: Srovnání referenčních, zašuměných a filtrovaných příkazů a jejich úspěšnost rozpoznání. ..	62
Tab. 19: Výsledky procesu v reálném prostředí.....	63

Úvod

V současné době jsme svědky velkého rozvoje v oblasti inteligentních budov a chytrých domácností. Důvody aplikace těchto standardizovaných systémů na vlastní bydlení jsou různé, ale těmi nejdůležitějšími jsou určitě pohodlí, bezpečnost a úspora energie. Ovládání těchto systémů je zcela na jiné úrovni, než známe u klasické stavby. Ovládací prvky lze v mnoha ohledech přizpůsobit požadavkům uživatele, aby pro něj byly co nejméně problematické. Jedním z možných usnadnění v této oblasti je hlasové ovládání. Ovládání budovy touto cestou, představuje velkou výhodu pro určitou skupinu lidí, do které patří zejména senioři a tělesně postižení. Pomocí hlasových příkazů lze například jednoduše rozsvítit a zhasnout světlo, ovládat okenní rolety, nebo zapnout topení či klimatizaci. Hlasové příkazy jdou nastavit prakticky na vše, čeho je konkrétní inteligentní systém schopen.

Hlasové ovládání však nemusí být vždy ideální. Každá místnost má svou akustiku, která zásadním způsobem může zkreslit vyřčený příkaz. Dalším a také závažnějším problémem je to, že v některých místnostech se lze setkat s různými typy šumu, které konkrétní hlasový příkaz znehodnotí a ten je prakticky nepoužitelný. Pro uživatele, kteří mají hlasové ovládání jen jako doplněk ovládání klasického, to nebude představovat tak velký problém jako například pro ty, kteří nemusí být schopni ze zdravotních důvodů manuální manipulace s ovladačem. V tomto případě je nutné doplnit do tohoto systému zařízení, které bude schopno zaznamenat hlasový příkaz i s konkrétním šumem a ten poté na základě předem dané struktury potlačit, aby byl systém pro hlasové ovládání schopen tento signál rozpoznat.

Cílem této diplomové práce je tedy vyvinout systém, který by byl schopen potlačovat šum zaznamenaný zároveň s řečí a zajistit tak větší procentuální úspěšnost rozpoznání příkazu, který je určen pro hlasové ovládání inteligentních budov.

Na základě této problematiky je nutné zpracovat rešeršní výzkum v oblasti zesílení hlasu a potlačení šumu pomocí adaptivních Soft Computingových metod. Tomuto tématu se bude věnovat celá první kapitola. V případě nalezení některých funkčních a v praxi úspěšných adaptivních metod, tuto metodu blíže prostudovat, zhodnotit a využít v realizaci.

Ve druhé kapitole budou přiblíženy adaptivní metody a pojem „Soft Computing“. Za tímto pojmem se skrývá velké množství konkrétních metod. Pro tuto práci byly vybrány umělé neuronové sítě, které se dále dělí do několika typů, popsaných v následující kapitole. Tyto typy je třeba zhodnotit, porozumět jejich strukturám a vybrat takový, který bude ideální pro potlačení šumu v řečovém signálu. Vybraný typ neuronové sítě je ve čtvrté kapitole popsán jak teoreticky tak i matematicky a implementace adaptivního algoritmu je provedena v interaktivním programovém prostředí MATLAB.

V následující kapitole jsou také popsány šумы, jejich typy a hlavně to jak ovlivňují řeč. Důležitým prvkem v této problematice je segmentální odstup signálu od šumu SSNR s exponenciálním průměrováním výkonu šumu v řečových pauzách, podle kterého lze řečový signál hodnotit. Dále je potřeba seznámit se s normou ITU-T P.863, která je pro hodnocení kvality řeči v této práci velmi důležitá a které se věnuje celá šestá kapitola. V následující kapitole je popsána poslední použitá hodnotící metoda a tou je střední kvadratická chyba.

V osmé kapitole jsou popsány inteligentní budovy a komunikační systém KNX. V praktické realizaci, která začíná devátou kapitolou, bude přiblížen software pro hlasové ovládání KNX systému, na který tato práce navazuje a pomocí kterého bylo možno zhodnotit dosažené výsledky. V desáté kapitole bude popsán navržený systém a také jeho implementace v reálném prostředí. Funkčnost programu bude odzkoušena simulací a také experimentálně s reálnými daty. Výsledky budou zhodnoceny pomocí klasifikačních kritérií, porovnány mezi sebou a podrobněji popsány v poslední kapitole.

1 Současný stav využití Soft Computingových metod pro zpracování řečového signálu

S. Agarwalla se ve své práci zaměřuje na techniky Automatického Rozpoznání Mluvčího (ASR) a Lidský vliv na počítače (HCI). Řeší se zde učení strojů používané pro extrakci relevantních vzorků z velkého datového objemu a jejich aplikaci pro ASR. Systém je testován pro několik vzorků šumu a rušení s ohledem na míru rozpoznání. [1]

V práci autora Asano, F., je navržen způsob detekce řeči z několikanásobného zvukového zdroje pomocí zvukové a vizuální informace. Detekce řeči je důležitá pro automatické rozpoznání řeči v reálném prostředí. Výkon separace zvuku pomocí adaptivního tvarování je výrazně silnější, pokud je známo umístění zdroje. Za účelem lokalizace řeči se používají mikrofony v kombinaci s Bayesovskou sítí. [2]

V práci autorů Czyżewski, A. a R. Królikowski je řešen problémem zpracování digitálního audio signálu pomocí hrubé neuro hybridizace. Kromě tohoto je zde popsána aplikace Soft Computingových metod na snížení nestacionárního šumu. Tento algoritmus na redukci šumu je založen na novém přístupu využívající některých vjemových vlastností lidského sluchového systému. Cílem je odhalit aktuální statistiky ohrožení sluchu na základě analýzy signálů přijatých z telekomunikačních kanálů. [3]

Autor Du, K. L. se ve své práci zabývá metodami založenými na konvenčním zpracování dat, které jsou výpočetně náročné a vyžadují znalosti odborníků pro modelování systému. Neuronové sítě představují důležité odvětví v Soft Computingových metodách. Tato práce poskytuje výkonné a univerzální paradigma pro zpracování informací. V každé kapitole jsou popsány výsledky výzkumu příslušných metod neuronových sítí. Je zde představeno několik relevantních témat jako například fuzzy logika, nebo evoluční algoritmy. Tohle jsou účinné nástroje pro učení neuronových sítí. Tento článek je cennou referencí pro všechny, kteří pracují na rozpoznávání, zpracování signálu nebo řeči, zpracování obrazu, analýze dat a umělé inteligenci. [4]

V práci autora Genaro, N. je uvedeno, že Světová zdravotnická organizace (WHO) klasifikuje šum jako nečistou část signálu a poté byly přijaty zákony a předpisy, které regulují hladinu hluku. Autor dále popisuje použití umělých neuronových sítí (ANN) k modelování městského hluku. Provedl několik aplikací na akusticky odlišných místech ve Španělsku a výsledky porovnal s matematickými modely. Bylo zjištěno, že systém ANN byl schopen předpovědět výskyt hluku s velkou přesností, a proto došlo u těchto modelů ke zlepšení. Dále tato práce obsahuje další výzkumné studie, které využívají Soft Computingové metody pro modelování městského hluku. [5]

Autor Gil-Pita, R. pojednává o využití Soft Computingových metod pro vytvoření energeticky účinných algoritmů pro binaurální sluchadla schopna rozlišení a oddělení řeči od ostatních slyšitelných zvuků, které jsou nežádoucí. K tomu je potřeba bezdrátového přenosu mezi levým i pravým sluchadlem pro předávání informací a parametrů, které jsou důležité pro algoritmus separace řeči. Hlavním problémem je zde snížit počet bitů použitých k reprezentaci informace, z důvodu velkého datového přenosu a tudíž snižování životnosti baterie. Autorův přístup je založen na kvantování parametrů, které mají být přeneseny a to tak, že výpočet založený na kombinaci neuronových sítí a genetických algoritmů provede ve snaze o nalezení rovnováhy mezi nízkou přenosovou rychlostí (a tedy i nízkou spotřebou energie) a dobrou separací řeči sumarizací kvantizačních bitů. [6]

Kasabov, N. K. se ve své práci zabývá fuzzy neuronovými sítěmi. Tyto sítě mají řadu vlastností díky nimž jsou vhodné pro širokou škálu technických aplikací. Mezi tyto výjimečné vlastnosti patří například rychlé a přesné učení, dobré generalizační a vysvětlovací schopnosti, vše v podobě sémanticky smysluplných fuzzy pravidel schopných pojmout jak data, tak i stávající odborné znalosti o konkrétním problému. Mezi dva alternativní způsoby optimalizace struktury patří genetický algoritmus a metoda učení se zapomináním. Taková struktura může být účinně použita pro rozpoznání řeči na bázi fonému. [7]

Autoři Malcangi, M. a P. Grew se ve své práci zabývají problémem zlepšení automatizovaných systémů pro automatické rozpoznání řeči. Toto téma je momentálně tím hlavním v oblasti zpracování řečového signálu, zvláště pokud jsou tyto systémy v provozu v hlučném prostředí. Tato práce navrhuje multimodální evoluční neuro-fuzzy přístup k rozvoji automatických systému pro rozpoznávání řeči. Tato technika je využita pro zpracování audiovizuální informace v převodu řeči na text. Používá se zde rozhodující síť EFuNN. Tato Soft Computingová metoda se ukázala jako solidní v přísných podmínkách a zároveň méně složitá oproti Hard Computingovým metodám. [8]

Malcangi, M. řeší téma zlepšení robustnosti rozpoznání řeči jako jedno z nejdůležitějších témat ve zpracování řečového signálu a to zejména při aplikaci v hlučném prostředí. Většina prací je soustředěna na kombinaci zvukových a obrazových dat v implementaci systému pro audiovizuální rozpoznání řeči (AVSR). Bimodální přístup prokázal zisk vynikajícího výkonu oproti rozděleným složkám. Tato práce navrhuje způsob, ve kterém se pomocí fuzzy logiky využívající schopnost rozpoznání dvou nezávislých pracovních systémů v kombinaci se systémem automatického rozpoznání řeči, slučují data. [9]

Článek S. Bhardwaj, navrhuje přístup založený na tvarové podobnosti vzorků pro odhad časové řady. Používají zde Skrytý Markovův model (HMM) pro shlukování a kombinaci se Soft Computingovými metodami (fuzzy inferenční systém, umělé neuronové sítě) pro odhad časové řady. Namísto použití funkce vzdálenosti jako indexu podobnosti, se zde tvar/vzorek sekvence používá jako index podobnosti pro shlukování, který překonává některé z nedostatků, které souvisí s přístupem založeným na vzdálenosti. Podkladové skryté vlastnosti časových řad jsou zachyceny pomocí HMM. Predikce je zde založena na identifikační zdatnosti vzorku HMM pro výběr shluku, zobecnění a nelineárních modelovacích schopností Soft Computingových výpočetních metod schopných předpovědět výstup systému. [10]

Malcangi, M. se ve své práci zabývá tématem dolování audia, které je v současné době předmětem usilovného výzkumu vzhledem k jeho potenciálu urychlit vyhledávání mluveného slova z audio nahrávek. Tato studie zkoumá tento způsob blíže z druhého konce. Navrhuje rámec založený na čipové logice, zejména fuzzy logice a na audio modelu aplikovatelném na jakýkoli druh zvukového záznamu, včetně hudby. [11]

Práce autora Teodorescu, H. N. se zabývá kritikou a recenzemi několika aplikací fuzzy logiky a fuzzy systémů v technice řeči. Mezi tyto techniky lze zařadit syntézu řeči, rozpoznání řeči a analýzu řeči. Krátké zachycení použití smíšených technik kombinujících fuzzy logiku, fuzzy klasifikaci a nelineární dynamiku. [12]

Autor S. Anissa, zkoumá zesílení řečového signálu použitím optimálního adaptivního filtru typu dolní propust, řízeného neuronovými sítěmi. Přítomnost aditivního šumu v řečovém signálu způsobuje degradaci jeho kvality a snižuje srozumitelnost. Filtrace tohoto zkráceného signálu v jeho prostorové oblasti je těžký úkol. Ještě těžší se tento úkol stává v případě, že jsou deformace způsobeny barevným šumem. Použití statického filtru není efektivní vzhledem k variabilitě řečového signálu. V jedné větě se může foném změnit jak tvarem, tak amplitudou. Díky těmto omezením je vhodnější použít filtr typu dolní propust s Gaussovým jádrem, řízený neuronovými sítěmi. Síla filtrace se neustále mění s variací fonémů pro generování variabilních filtrů, které jsou schopny změnit celou větu. [13]

Článek autora L. Behera, se zabývá inteligentním stochastickým filtrováním pomocí opakujícího se kvantového modelu neuronové sítě. Tento přístup nedává žádný předpoklad o povaze a tvaru signálů a šumu. Opakující se kvantová neuronová síť (RQNN) je navržena tak, aby modelovala jednotnou odezvu neuronové mřížky, zatímco ignoruje jednotlivé neuronové odpovědi. Průměrná odpověď neuronové mřížky je popsána pomocí Schrödingerovy vlnové rovnice. Bylo zjištěno, že uzavřené smyčky RQNN vykazují solitonové vlastnosti. Tento vlastnosti se využívá při stochastické filtraci v dnešní komunikaci. Výkonná účinnost RQNN je potvrzena při sledování amplitudové modulace sinusoida a čtvercových signálů. Je zde také prezentováno sledování řečového signálu vnořeného v Gaussovském šumu s nestacionárním rozptylem. Zesílení řečového signálu pomocí RQNN je také testováno v reálném čase. [14]

E. Cherifi řeší zlepšení kvality řečového signálu odstraněním hluku z potřebného, užitečného signálu. Toto zlepšení má několik aplikačních polí: Zlepšení výkonu hlasových komunikačních zařízení v hlučném prostředí, jako například letecká navigace, videokonference, nebo oblast mobilních telefonů. V této práci diskutují o využití vícevrstevné umělé neuronové sítě dopředného typu (FF-ANN). Síť se podrobí učicímu algoritmu využívajícího rozšířený duální Kalmanův filtr (DEKF) pro odstranění nestacionárního a barevného šumu z řečového signálu. Je zde také předvedena účinnost metody. [15]

Autor K. Daqrouq, se zabývá zvýrazněním řeči a zpracováním zašuměného řečového signálu s cílem zlepšit kvalitu nebo srozumitelnost signálu. Cílem je zesílit řečový signál, nebo ze signálu poškozeného hlukem získat čistý signál s vyšší kvalitou. Nicméně, přítomnost hluku v řečovém signálu přispěje ve vysoké míře k nepřesnostem v systému pro zpracování řeči. Tato myšlenka pro potlačení šumu v řečovém signálu je realizována pomocí neuronových sítí. Byly testovány tři metody: 1. Adaptivního lineárního neuronu (ADALINE), 2. Metoda zesílení dopředné neuronové sítě (FFNN), 3. Transformace vlny a vylepšená ADALINE metoda. Získané výsledky byly zhodnoceny jako velmi kvalitní díky rychlému zpracování a vysokému poměru signál-šum. Testované signály byly navýšeny o 10 dB u ADALINE, o 3 dB u FFNN a o 8 dB u Transformace vlny a vylepšené ADALINE metody. [16]

Práce autora L. Dehyadegary, se zabývá tvorbou průběžné dynamiky atraktoru v nelineárních opakujících se neuronových sítích. Tato nelineární metoda se používá k odstranění šumu z řečového signálu, za účelem rozpoznání fonémů a získání informací. Pro formaci dynamiky atraktoru v rekurentní neuronové síti se nejprve provádí výcvik čisté řeči jakožto kontinuálního atraktoru. Poté je neuronová síť schopna rozpoznat hlučný projev se stacionárním i nestacionárním šumem. V této práci je účinnost nelineární dopředné neuronové sítě srovnávána s podobnou, avšak navíc obsahující opakující se spojení ve skryté vrstvě. Struktura a učení tohoto opakujícího se připojení je navržena takovým způsobem, že je síť schopna odrušovat signál krok za krokem, s využitím vytvořených atraktorů. Pomocí těchto

technik se přesnost rozpoznání zlepšil o 21% u stacionárního signálu a o 14% u signálu nestacionárního s 0 dB SNR, ve vztahu k referenčnímu modelu, kterým je dopředná neuronová síť. [17]

Autor H. Erdogan, v této práci řeší náročný problém separace řeči na nestacionární rušení. Nedávno tento problém zaznamenal dramatické zlepšení díky metodě hlubokých neuronových sítí. Původní práce ukázaly, že maskovací funkce může být zlepšena pomocí aproximačního signálu založeném na objektivní funkci. Vytváření dynamiky skrze hluboké rekurentní sítě bylo prokázáno jako výrazně lepší z hlediska výkonu. Dále sledují oba tyto směry. Vyvíjí cílové funkce fázově citlivé na základě SNR z rekonstruovaného signálu. V experimentech je dokázáno, že mnohem lepších výsledků je dosaženo u signálu s poměrem signál-zkreslení (SDR). [18]

V této práci se autor L. B. Fah zabývá použitím neuronových sítí pro zlepšení řečového signálu a to tím způsobem, že z tohoto signálu odstraní šum. Čisté řečové signály zaznamenali ve formě číslovek v Malajštině a to v prostředí téměř bez šumu. Zašuměné signály získali tak, že na ty čisté přidali náhodný typ šumu. Potlačení šumu pak aplikují na zašuměné signály pomocí metody ADALINE. Hodnocení výkonosti metody ADALINE poté vychází ze dvou samostatných metod: SNR a audiovizuální kontrolou. Efektivita metody k potlačení šumu ADALINE je také srovnávána s metodou Vícevrstevného perceptronu (MLP) pro zvýraznění řeči. [19]

T. Gao ve své práci řeší důležité nastavení ve zpracování řeči a to konkrétně zvýraznění řeči a její separaci pomocí hlubokých neuronových sítí (DNN). V reálných situacích je lidský hlas ve směsi s ostatními rušivými zvuky v pozadí. DNN jsou brány jako nelineární mapovací funkce mezi čistými a zašuměnými řečovými signály. Zvuky pocházející z reproduktorů jsou brány jako jeden z univerzálních druhů rušivých signálů. Aby bylo možné zvládnout filtrovat velikou škálu aditivních šumů, vyskytujících se v reálném prostředí, je zapotřebí velká tréninková souprava, která zahrnuje mnoho možných kombinací řeči a zvuků. Experimentální výsledky ukazují, že pomocí této metody lze ze zašuměných signálů získat zpět téměř čisté. Kromě toho, že výsledný DNN model pracoval pouze s uměle syntetizovanými daty, je účinný i při řešení hlučných situací zaznamenaných v reálných podmínkách. [20]

Práce autora S. Gholami-Boroujeny pojednává o měření rychlosti odpovědi mozku na lidskou řeč (ABR). Tato metoda je slibná pro vyhodnocování sluchových funkcí. Nicméně, tyto vjemy jsou silně znečištěny na pozadí šumem. Nejčastěji používanou metodou pro zvýšení SNR je koherentní zpřůměrování, zatímco v poslední době začíná být také využívána adaptivní filtrace. Všechny použité metody jsou založeny na lineárních operacích, ale protože předpoklad linearity nemusí platit pro nervovou aktivitu, lineární metody nemusí být vždy dostatečné. Autoři v článku prezentují nový nelineární způsob odstranění adaptivního šumu, založený na perceptronové neuronové síti vedoucí ke zlepšení ABR a k porovnání výkonu s lineárním algoritmem. [21]

L. Girin, se v práci zabývá klíčovým problémem pro zlepšení telekomunikačních přístrojů, založených na hlasovém ovládní. Tento problém je jednoduchý – šum. V této oblasti již existuje několik technik na jeho odstranění, všechny jsou však založené na akustickém přístupu, to znamená, že ke zpracování zvukového signálu dochází pouze skrze poškozený signál, nebo přes aditivní zvukový signál. V této práci je upřednostňován audiovizuální přístup, protože je prokázáno, že sledování obličejové mluvího při mluvení, zlepšuje srozumitelnost zprávy a to zejména v hlučném prostředí. Je vyvíjen prototyp zařízení

pro zvýraznění řečového signálu s vizuálním vstupem. Klíčovým aspektem v tomto rozpoznání je pohyb lidských rtů. Proces rozpoznání je založen na jednoduché lineární regresi, nebo na neuronových sítích využívajících tréninkových možností. [22]

Autor T. G. Kang řeší jednu z nejnámějších technik pro oddělení požadovaného zdroje z většího množství dat. Tato technika se nazývá Nezáporná faktorizace matic (NMF). V rámci NMF je soubor dat faktorizován do základní a kódovací matice. Základní matice pro směsici dat je obvykle konstruována rozšířením základních matic pro nezávislé zdroje. Nicméně oddělení cílového zdroje s rozšířenou základní maticí se ukázalo jako problematické, pokud existuje nějaké překrytí mezi podprostory, tak je to báze pro individuální rozsah zdrojů. V této práci autoři navrhují nový přístup ke zlepšení kódování vektorového odhadu pro cílovou extrakci signálu. Odhad kódování vektorů ze směsice dat je považován za regresní problém. Hluboké neuronové sítě DNN se používají k učení mapování mezi směsici dat a odpovídajícími kódovacími vektory. K demonstraci funkce navrženého algoritmu, byly prováděny experimenty za účelem zvýraznění řečového signálu. Experimenty ukázaly, že navržený algoritmus překonává konvenční kódování vektorů. [23]

W. G. Knecht, se zabývá potlačením šumu pomocí neuronových sítí. V adaptivním potlačení šumu, byly použity lineární digitální filtry pro minimalizaci středního čtvercového rozdílu mezi filtračními výstupy a požadovaným signálem. Nicméně, pro ne-Gaussovské pravděpodobnostní funkce hustoty použitých signálů, nelineární filtry mohou dále snížit množství šumu, čímž se zlepšuje SNR na výstupu systému. K aproximaci optimálního výkonu, je potřeba realizovat dva nelineární filtry, Volterrův filtr a vícevrstevný perceptron. Volterrův filtr také zaznamenává rušení v řeči. Kritéria požadují, aby na výstupu byl poměr signál-šum. Pro nelineární filtr je to splněno pouze v případě, když neexistuje žádná cílová složka v referenčním kanálu filtru rušení, takže požadovaný signál lze přenášet bez jakéhokoli zkreslení nebo rušení. Za těchto ideálních podmínek a ve stejných délkách filtru, kvadratický Volterrův filtr zlepšuje vážený zisk o maximálně 2 dB oproti lineárnímu filtru. [24]

Y. H. Lai, se v práci věnuje redukci šumu (NR). Redukce šumu je klíčová procedura ke zlepšení řečového signálu ve hlučném prostředí. Přesněji řečeno, NR může zlepšit úroveň poměru signál-šum u výstupního signálu. Tímto se také zvýší srozumitelnost řeči a kvalita zvuku. V poslední době byly vyvinuty diskriminativní post-filtry (DPF) a předběžné výsledky ukazují, že DPF mohou dále zlepšit výkon výstupu SNR. Je důležité zadat vhodné parametry pro DPF, zůstává však otázkou, jak tyto parametry nastavit. Proto se tato studie zaměřuje především na genetický algoritmus (GA), aby prošetřila tyto parametry pro DPF, vedoucí k dalšímu navýšení výkonosti SNR na výstupu. Výsledky objektivního hodnocení ukazují, že v hlučném prostředí, zkoumané parametry DPF pomocí algoritmu GA jsou udrženy a to vede k vyšším výkonům SNR a to ve srovnání s klasickými NR metodami. Co se týče výkonu, navržená metoda může zvýšit výstupní SNR oproti klasickým metodám NR až o 5 dB. Z toho vyplývá, že tato metoda může pomoci s rozpoznáním řečového signálu v hlučném prostředí. [25]

Práce autora C. T. Lin, řeší problém segmentace řeči a její zesílení v přítomnosti šumu. Nejprve je navržen nový algoritmus detekce slova pomocí Fuzzy neuronové sítě (tzv. ATF SONFIN algoritmus). Tento algoritmus je schopný v signálu identifikovat centra slovních signálů v prostředí, které podléhá také šumu. Dále zde navrhují nový algoritmus založený na RSONFIN algoritmu, kde se mění hladina hluku v pozadí během nahrávání. Adaptivní časově-frekvenční parametr (ATF) a Vylepšený časově-frekvenční parametr (RTF) rozšiřují časově-frekvenční analýzu z jednopásmové na

vícepásmovou. Dále jsou schopny rozpoznat řeč a rušivé signály. Algoritmus RSONFIN RTF je dále schopen najít na pozadí hladinu hluku a detekovat správnou hranici slova. Experimentální výsledky ukazují, že navržené použité algoritmy dosáhly lepší úspěšnosti v rozpoznání slov než běžně používané detekční algoritmy a snižují tedy chybovost rozpoznání konkrétních příkazů. [26]

B. Marungsri, ve svém článku prezentuje metodu ke snížení šumu v řečovém signálu, který je určen pro ovládání automatického kolečkového křesla. Tato metoda je založena na použití techniky neuronové sítě Fuzzy Artmap (FAMNN). Měřitelné výstupy zašuměného signálu s 5dB, 1dB, -5dB a -10dB. Úroveň SNR se získá ve formě kontaminovaného signálu, ve srovnání s výstupními, filtrovanými daty. Jako vstup je použit zdroj bílého šumu. Nakonec po tréninku byl prokázán Fuzzy ARTMAP výstup. Poté je informační signál vypočten jako rozdíl mezi měřeným signálem a předpokládaným rušením. Fuzzy ARTMAP může tedy prakticky řešit situace adaptivního potlačování rušivých složek z hlasových příkazů pro automatický systém invalidních vozíků s nelineární charakteristikou. [27]

Autor S. Pal, se v této práci věnuje hlasovým hands-free terminálům, které musí být funkční v různých situacích, ve kterých adaptivní filtr (použit pro redukci ozvěny) není dostatečný. K takovým situacím patří například změna cesty ozvěny, příliš hlasitý vzdálený konec, nelinearita hlasu mluvčího, změna hlasitosti a další. V těchto situacích by měla být zbytková ozvěna potlačena přidáním řídicím modulem zesílení. V této práci je navrženo řešení pro zesílení pomocí neuronové sítě se dvěma cíli: a) Potlačení zbytkové ozvěny vzniklé z různých důvodů, b) Výpočet optimálního zesílení pro odeslání signálu. Neuronová síť je schopna rozpoznat situace, ve kterých může adaptivní filtr účinně potlačit akustickou ozvěnu hlasu. Navržený algoritmus je vhodný pro implementaci v reálném čase a byl testován v malé kancelářské místnosti. Experiment ukázal, že tato metoda měla lepší výsledky, než podobné Voice-over-IP řešení. [28]

B. P. A. Rohman, se zabývá adaptivně kompenzovanou vícepásmovou spektrální subtrikcí (MBSS). V tomto výzkumu adaptivní kompenzace jsou využity umělé neuronové sítě. Účelem je zlepšit kvalitu řečového signálu po odstranění šumu. Byla použita neuronová síť složená z vícevrstevného perceptronu a to konkrétně ze tří skrytých vrstev. Neuronová síť byla testována na třech řečových signálech znečištěných pomocí bílého Gaussovského šumu s poměrem signál-šum 0 dB a 30 dB. Pro zjištění výkonu byla tato navržená metoda testována pěti zašuměnými řečovými signály s poměrem signál-šum 0 dB až 10 dB. Na základě tohoto experimentu lze říci, že navržená metoda pro zvýraznění řeči prokázala vyšší výkon řečového signálu oproti původnímu algoritmu. [29]

Autor O. Sharifi-Tehrani se v tomto článku věnuje aktivní regulaci hluku (ANC), která se stala v posledních desetiletích důležitou oblastí výzkumu. Řízení hluku v průmyslových prostředích, v konferenčních sálech a v komunikačních systémech bylo zkoumáno jako Adaptivní aktivní potlačení šumu (AANCC) a výsledky těchto studií byly využity v praktických aplikacích. Snížení hluku v závislosti na čase je jedním ze způsobů, vedoucích ke zvýraznění řeči. Byly zde zavedeny umělé neuronové sítě ADALINE jako inteligentní duální mikrofon aktivního potlačování šumu v systému. Tato umělá neuronová síť identifikuje zdroje šumu a rušení v průběhu tréninkové fáze a dále se upravuje. Tímto způsobem systém snižuje šum vstupního signálu. Testy a implementace zde uvedené jsou v perském jazyce za použití bílého Gaussovského šumu. [30]

E. A. Wan se v této práci věnuje redukci šumu z řečového signálu. Tento proces zesílení hlasu má využití v komunikacích, konkrétně v rozpoznání řeči. Je zde představena neuronová síť založená na metodě časové domény DEKF pro odstranění nestálého a barevného šumu z řečového signálu. Tento článek popisuje algoritmus a poskytuje sadu experimentálních výsledků. [31]

V práci autora Y. S. Xia je navržena nová metoda opakujících se neuronových sítí založena na Kalmanově filtru pro zvýraznění řeči, odhadující a omezující šum. Parametry řečového signálu modelovány jako autoregresní proces, se nejprve odhadnou pomocí navrhované rekurentní neuronové sítě a signál řeči se poté získá pomocí Kalmanovy filtrace. Navržená rekurentní neuronová síť je globálně asymptoticky stabilní vůči šumu, který je schopna omezit. Navržený algoritmus zvýraznění řeči pomocí této neuronové sítě, je mnohem rychlejší než aktuální algoritmy. Výsledky simulace ukazují, že navržený algoritmus opakujících se neuronových sítí, je schopen výkonně a rychle redukovat šum. [32]

Y. Xu ve své práci představuje regresně založenou metodu zvýraznění řeči s využitím hlubokých neuronových sítí (DNN) s vícevrstevnou hlubokou architekturou. V procesu učení DNN se nachází výkonná modelovací schopnost odhadnout složitá nelineární mapování a ze zašuměného signálu získat signál čistý. Byl zde nalezen akustický kontext vedoucí ke zlepšení plynulosti řeči a k úspěšnému oddělení šumu z pozadí bez nepříjemného zvukového artefaktu, který je pozorován u běžných algoritmů. Řada experimentů byla provedena na více než 100 hodin dlouhých řečových záznamech, což má za následek dobrou generalizační schopnost i ve špatných testovacích podmínkách. Experimenty prokázaly lepší výsledky oproti konvenčním algoritmům. [33]

V této rešerši se vyskytuje množství článků, které řeší problém rozpoznání řeči v zašuměném prostředí a následnou proceduru pro odstranění šumu z podstatného řečového signálu. Metody pro odstranění šumu využívají Umělých neuronových sítí popřípadě jiných, podobných Soft Computingových metod. Ve všech experimentech se setkáváme s dobrými a mnohdy lepšími výsledky než u konvenčních filtračních metod pro odstranění šumu. Na základě několika článků jsem se rozhodl použít pro potlačení šumu z hlasových příkazů umělé neuronové sítě.

2 Soft Computing

Soft computing [43] je termín z angličtiny, který nemá přesný a výstižný český překlad. Slovo computing můžeme chápat jako výpočetní metody, algoritmy a teorie. Slovo soft v tomto smyslu může znamenat jemný, nebo mírný.

Soft computing (SC) je spojení výpočetních metodologií, jako jsou zejména fuzzy logika, neuropočítání, genetické počítání a pravděpodobnostní počítání. Metodologie zahrnující Soft Computing jsou ponejvíce komplementární a synergistické, spíše než konkurenční. Vedoucím principem SC je využití tolerance vůči nepřesnosti, nejistotě, částečné pravdě a aproximaci k dosažení zvládnutelnosti, robustnosti, nižší ceny řešení a lepšího souhlasu s realitou. Jedním z hlavních cílů SC je vytvořit základy pro navrhování, vytváření a aplikování inteligentních systémů uplatňujících své složky symbioticky a ne izolovaně.

Francouzští odborníci na fuzzy logiku Didier Dubois a Henri Prade upozorňují, že se jedná o obecně tvůrčí propojení metod, které využívají numerická data, dále metod pracujících se znalostmi a metod symbolického usuzování a matematického modelování. Nejde tedy o relativizování významu metod klasické matematiky, ale o to, aby byly nalézány účinné výpočtové metody pro situace, kde jsou klasické výpočetní metody nepoužitelné. [34]

Autor termínu Soft Computing a teorie fuzzy množin, profesor L. Zadeh zdůrazňuje, že neurčité, či příliš složité problémy vyhodnotí jako absurdní, nebo exaktní řešení vylučují. Sledování reálného světa z lidského pohledu je nepřesné či neúplně definované. Autor to popisuje jako „princip neslučitelnosti“ a vysvětluje ho tak, že problémy nastíněného typu se neslučují s ideálem exaktního řešení, které je optimální. Tato optimální řešení nemusí existovat, jsou nákladná, nebo nezajímavá. Je třeba proto užívat toleranci a snažit se najít řešení, která jsou dostatečná. [34]

2.1 Soft Computingové metody

2.1.1 Fuzzy počítání

Tento výraz je příkladem dalšího anglického termínu, pro který zatím bohužel nebyl nalezen vhodný ekvivalent v českém jazyce. Fuzzy podmnožina nějakého oboru je určena tím, že každý prvek toho oboru patří do fuzzy podmnožiny v jistém určitém stupni (obvykle se jedná o reálné číslo mezi nulou a jedničkou). Tyto fuzzy množiny modelují velmi dobře nejasné pojmy, jako je např. „vysoká teplota“. Lze si představit, že každé konkrétní hodnotě teploty odpovídá nějaké číslo, které říká, do jaké míry je tato teplota vysoká. Na následujících příkladech bude vysvětleno, jak lze s fuzzy množinami počítat.

Tab. 1: Příklad 1. Funkce [34]

x	2	3	4	5
y	4	6	8	10

(První funkce přiřazuje každému přirozenému číslu od 2 do 5 jeho dvojnásobek.) Ve druhé tabulce se místo čísel vyskytují fuzzy pojmy velmi malý, malý, střední, velký, velmi velký.

Tab. 2: Příklad 2. Funkce [34]

x	Velmi malý	Malý	Střední	Velký	Velmi velký
y	Malý	Střední	Velký	Střední	Malý

Tato druhá tabulka, popisuje nějakou funkci pouze velmi neurčitě, ale může být také přirozeným popisem, který v dané situaci může stačit, a dokonce – známe-li významy uvedených pojmů, tj. jsou-li dány odpovídající fuzzy množiny – lze pomocí takové tabulky (obvykle vyjadřované ve formě pravidel tvaru „když x je velmi malé, pak y je malé“) spočítat přesné hodnoty funkce, která naši informaci vystihne. Toho se podstatně využívá v metodách fuzzy řízení (fuzzy regulace) při formalizaci empirických znalostí o chování řízeného systému. [34]

2.1.2 Genetické algoritmy (evoluční počítání)

Jde o řetězec znaků (gen) popisující řadu různých řešení určitého problému, která jsou k dispozici. Lze také nalézt nová řešení a to obměnou, či křížením již známých řešení, podobně jako selekce v genetice, čímž vzniká nová populace různých řešení. [34]

Myšlenka skryta za vytvořením genetických algoritmů, byla použít právě evolučních principů, založených na metodách optimalizace funkcí a umělé inteligence, pro hledání řešení nějaké úlohy.

Níže je popsána základní struktura genetického algoritmu. [35]

Genetický algoritmus (fit, N, K, M)

Inicializace

1. přiřad' $t := 0$ (počítadlo generací)
2. náhodně vytvoř populace $P(t)$ velikosti N
3. urči hodnoty funkce fit pro každého jedince b v $P(t)$

Hlavní cyklus

1. dokud není splněna podmínka pro zastavení
 - 1.1. proved' selekci:
 - 1.1.1. vyber z $P(t)$ jedince kteří se přímo přenesou do $P(t+1)$
 - 1.2. Proved' křížení:
 - 1.2.1. vyber z $P(t)$ dvojice jedinců určených k reprodukci
 - 1.2.2. aplikuj na každou dvojici operaci křížení
 - 1.2.3. zařad' potomky do $P(t+1)$

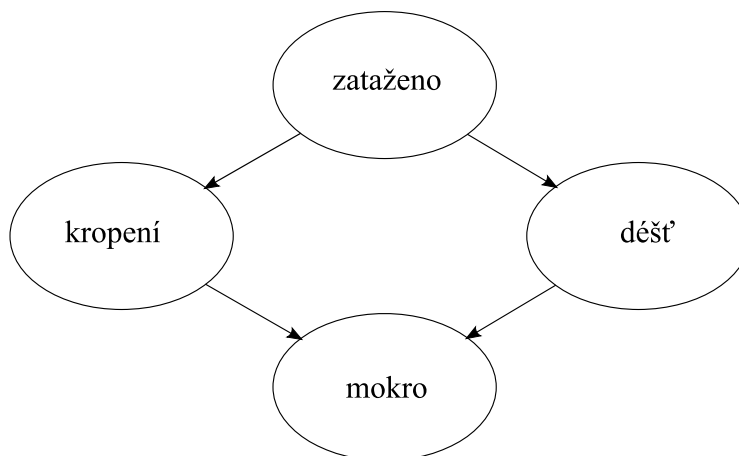
- 1.3. proved' mutaci
 - 1.3.1. vyber z $P(t+1)$ jedince určené k mutaci
 - 1.3.2. aplikuj na každého jedince operaci mutace
- 1.4. přiřaď $t := t+1$ (nová populace má opět velikost N)
- 1.5. spočítej pro každé $b \in P(t)$ hodnotu $fit(b)$
2. vrať jedince b s nejvyšší hodnotou $fit(b)$ [35]

Tento algoritmus, začne náhodně pracovat s nějakou zvolenou populací, kterou se snaží postupně zlepšovat, modifikovat. Genetická výbava jedinců přechází do následující populace přímo, nebo pomocí potomků. Snaha tohoto algoritmu je přenést ze stávající do následující populace jen to nejlepší. Činnost algoritmu končí po splnění určité podmínky. Výběr, mutace a křížení jsou operace, díky kterým je možné vytvořit novou populaci. [35]

2.1.3 Bayesovské sítě

Bayesovská síť je acyklický orientovaný graf zachycující pomocí hran pravděpodobnostní závislost mezi náhodnými veličinami. Vyskytují se zde uzly, představující náhodné veličiny a ke každému z nich je přiřazena pravděpodobnostní distribuce. K výpočtům této distribuce je potřeba znát čtyři dílčí podmíněné pravděpodobnostní distribuce. V případě diskretních veličin jsou to tabulky podmíněných pravděpodobností. Příklad Bayesovské sítě je uveden na obrázku. [35]

$$P(Z,K,D,M) = P(Z) P(K|Z) (D|Z) P(M|K,D)$$



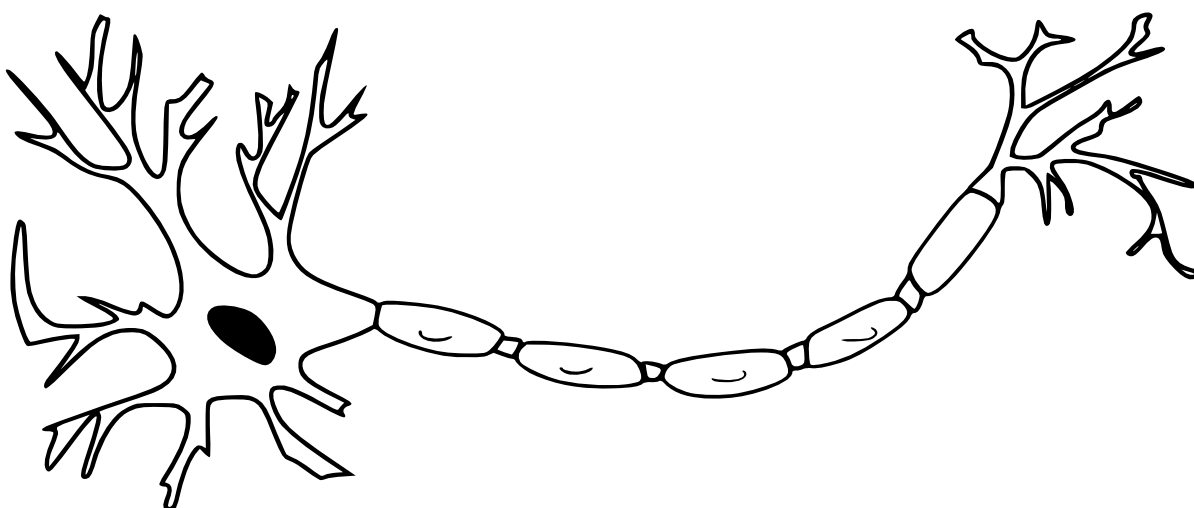
Obr. 1: Příklad Bayesovské sítě [35]

2.1.4 Data mining

Tento termín lze volně přeložit jako „těžení z dat“. Vznikl na základě toho, že se objevují stále větší soubory dat, v nichž se potenciálně nachází cenná informace. Cílem této metody je tuto informaci z velkého datového obsahu získat. Tuto metodu lze zadávat a realizovat kombinací různých podobných metod a přístupů jako např. pomocí fuzzy logiky či genetického algoritmu. [34]

2.1.5 Neuronové sítě

Neurony jsou nervové buňky tvořící lidský mozek, kde se jich nachází přibližně 10^{10} . Tyto buňky jsou vzájemně propojeny řádově ještě vyšším počtem vazeb, takzvanými synapsemi. Na obrázku je popsán typický neuron tvořený tělem (soma), ze kterého vybíhají desítky až stovky kratších výběžků (dendrity) a jeden dlouhý výběžek (axon). Tělo neuronu je veliké řádově několik mikrometrů, dendrity několik milimetrů, přičemž axon dosahuje délek od několika desítek centimetrů až do jednoho metru. Přes synapse jednotlivých neuronů dochází k přenosu vzruchů. K přenosu dochází díky chemickým reakcím a závisí na změně propustnosti buněčné membrány neuronu. Změnou propustnosti se lokálně mění koncentrace kladných a záporných iontů vně a uvnitř buňky a tedy dochází mimo jiné také ke změně membránového potenciálu. U vazeb se setkáváme s excitačním a inhibičním charakterem, v závislosti na snížení či zvýšení membránového potenciálu. Ve chvíli, kdy celkový membránový potenciál přesáhne určitý práh, dojde k aktivaci neuronu a ten přes svou synapsi začíná působit na další neurony, se kterými je spojen. [35]



Obr. 2: Schéma biologického neuron [34]

Lze říci, že neuron přijímají od jiných neuronů kladné, nebo záporné podněty. Pokud tyto podněty překročí určitý daný práh, dojde k samovolné aktivaci neuronu. Výstupní hodnotou neuronu bývá obvykle nějaká nelineární transformace souhrnu podnětů.

Obecná definice neuronové sítě:

Neuronová síť je n -tice (U, W, A, O, net, ex) , kde

1. U je konečná množina výpočetních jednotek (neuronu),
2. W je struktura sítě, tedy zobrazení z $U \cdot U$ do R (vazby mezi neurony),
3. A je zobrazení definující aktivační funkci $A_u : R \rightarrow R$ pro každou jednotku u ,
4. O je zobrazení definující výstupní funkci $O_u : R \rightarrow R$ pro každou jednotku u ,
5. net je zobrazení, definující vstupní funkci (přenosovou funkci) $Iu : (R \cdot R)^U \rightarrow R$ pro každou jednotku u ,
6. ex je externí vstupní funkce $ex : U \rightarrow R$, která každé jednotce u přiřadí externí vstup jakožto reálnou hodnotu $ex_u = ex(u) \in R$.

Neuronová síť se tedy skládá z uzlů (neuronů) a hran (synapsí). Každá hrana vede od nějakého neuronu k jinému. Některé neurony jsou vstupní, nevede od nich žádná hrana, ostatní jsou vnitřní. Neuronům a hranám přísluší různé přirozeně popsané charakteristiky. Tyto charakteristiky zapříčiňují volbu neuronu pro hodnotu, kterou pak dále posílá po hranách, jež z něj vycházejí. Fungování sítě si lze představit tak, že konkrétní hodnoty nezávisle proměnných vstoupí z okolního prostředí do vstupního neuronu, síť je zpracovává a na výstupních neuronech se objeví hodnoty závisle proměnných. Přáním uživatele je, aby tyto výstupní hodnoty byly v dobré shodě se vstupními daty. Neuronová síť je schopna tyto charakteristiky měnit formou učení, aby se shoda s daty zlepšovala. [34]

3 Typy neuronových sítí

V následujícím textu bude představeno několik typů neuronových sítí, které se v praxi používají zejména v softwaru MATLAB, a které by mohly pomoci při realizaci celé struktury schopné odstranit šum z řečového signálu.

3.1 ADALINE

ADALINE, neboli adaptivní lineární neuron, je neuronová síť, podobná perceptronu s lineární funkcí převodu. Linearita v tomto případě umožňuje, že vstup může nabývat jakékoli hodnoty, zatímco výstup perceptronu pouze 0 nebo 1. ADALINE i perceptron mohou vyřešit pouze lineárně oddělitelné problémy. Je zde použita metoda učení nejmenších čtverců (LMS), která je účinnější, než pravidlo učení perceptronu. Metoda LMS nebo metoda „Windrow-Hoff“ minimalizují střední kvadratické chyby a posouvá tak rozhodovací hranice tak daleko, jak je to možné z tréninkových dat. [38]

3.2 Neuronová síť s opakujícími se vrstvami

V neuronové síti s opakujícími se vrstvami (LRN) se vyskytuje zpětná vazba s jednoduchým zpožděním kolem každé vrstvy sítě, s výjimkou té poslední. Původní síť měla pouze dvě vrstvy a pro skrytou vrstvu používala přenosovou funkci „tansing“ a pro výstupní vrstvu přenosovou funkci „purelin“. Původní síť byla také trénována pomocí aproximace zpětného šíření „backpropagation“ algoritmu. [38]

3.3 Vícevrstevný perceptron

Vícevrstevný perceptron (MLP) je model dopředné neuronové sítě, která mapuje sadu vstupních dat na sadu příslušných výstupů. Procedura MLP se skládá z několika vrstev uzlů v orientovaném grafu, přičemž každá vrstva je připojena k další. S výjimkou vstupních uzlů, je každý ostatní uzel samostatným neuronem s nelineární aktivační funkcí. Neuronová síť MLP využívá k učení techniku zvanou „backpropagation“. Tato síť je modifikací standardního lineárního perceptronu a dokáže zpracovat údaje, u kterých to není za podmínek linearit možné. [38]

3.4 ANFIS

Jde o druh neuronové sítě, která je založena na „Takagi-Sugeno fuzzy“ inferenčním systému. Tato technika byla vyvinuta na počátku roku 1990. Tento model pracuje na principu jak neuronových sítí, tak na zásadách fuzzy logiky a má tedy potenciál pro řešení problémů díky výhodám obou těchto rámců. Výsledek systému odpovídá souboru fuzzy pravidel „IF/THEN“, díky kterým je systém schopen učení a chová se jako nelineární funkce. [38]

4 Adaptivní neuro-fuzzy inferenční systém (ANFIS)

ANFIS, neboli adaptivní neuro-fuzzy inferenční systém, používá hybridní algoritmus učení k identifikaci funkčních parametrů jednoho výstupu. Dále dojde k vytvoření fuzzy inferenčního systému (FIS), jehož funkční parametry jsou laděny algoritmem zpětného šíření „backpropagation“ nebo také v kombinaci s metodou nejmenších čtverců. Tento model umožňuje to, že systém je schopen učení z dat. Funkci ANFIS lze vyvolat příkazem v příkazovém řádku, nebo pomocí Neuro-Fuzzy designeru. Nejdříve je nutné navrhnout hypotézu parametrizované modelové struktury. Dále je nutný výběr vstupně výstupních dat, která budou vhodná pro trénink tohoto modelu. Po těchto krocích lze začít s učením sítě podle předložených dat s tím, že změna parametrů je možná, pomocí předem zvoleného chybového kritéria. [38]

4.1 Použití Neuro-Adaptivní učení

Základní struktura „Mamdani-Fuzzy“ inferenčního systému je model, který mapuje vstupní charakteristiky pro vstupní funkce příslušnosti, vstupní funkce příslušnosti pro pravidla, pravidla pro sadu výstupních charakteristik, výstupní charakteristiky pro výstupní funkce příslušnosti a výstupní funkce příslušnosti pro výstup o jedné hodnotě, nebo rozhodnutí, které je spojeno s výstupem. Takový systém používá pevné členské funkce, které jsou vybírány náhodně a strukturované pravidlo, které je v podstatě v tomto modelu uživatelskou interpretací charakteristik proměnných.

ANFIS a „Neuro-Fuzzy“ konstruktér používají fuzzy inferenční techniky k modelování dat. Tvar členských funkcí se mění v závislosti na vstupních parametrech. Namísto pouhého pohledu na data pro výběr parametrů členské funkce, lze tyto parametry vybrat automaticky použitím aplikace „Fuzzy Logic Toolbox“. V některých modelových situacích, totiž není možné rozeznat, jak by měla členské funkce vypadat pouze z jednoduchého pohledu na data.

Metoda učení tohoto systému funguje podobně jako u neuronové sítě. Technika neuro-adaptivního učení poskytuje způsob zisku informací o sadě dat. „Fuzzy Logic Toolbox“ počítá parametry členské funkce, které nejlépe umožňují fuzzy inferenčnímu systému sledovat daná vstupní/výstupní data. Funkce „Fuzzy Logic Toolbox“, která dosáhne parametrů této členské funkce, se nazývá ANFIS. K funkci ANFIS lze přistupovat v programu MATLAB buď z příkazového řádku, nebo pomocí „Neuro-Fuzzy“ konstruktéru.

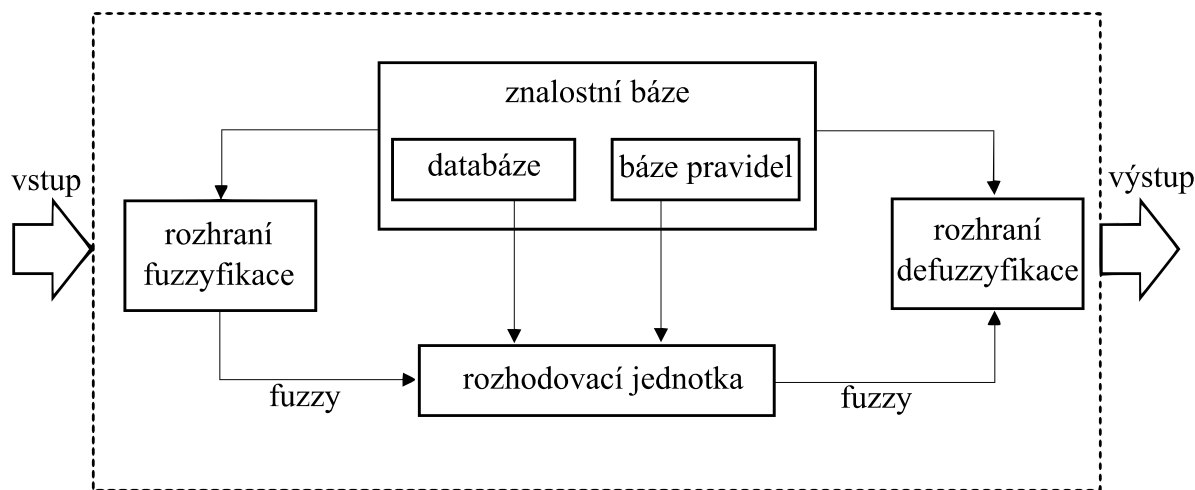
4.2 FIS struktura a nastavení parametrů

Konstrukce a typ sítě je velmi podobný jako u sítě neuronové, které mapuje vstupy pomocí členských vstupních funkcí a příslušných parametrů a poté přes výstupní funkce příslušnosti pomocí přidružených parametrů na výstupu. Parametry související s členskou funkcí se mění v průběhu procesu učení. Výpočet těchto parametrů (nebo jejich seřízení) je umožněno pomocí spádového vektoru. Spádový vektor poskytuje měřítko toho, jak dobře je fuzzy inferenční systém schopen modelovat vstupní/výstupní data pro danou sadu parametrů. Po získání spádového vektoru, může být použita některé z optimalizačních rutin pro nastavení parametrů za účelem snížení chyby procesu. Chyba procesu je obvykle definována součtem rozdílů čtverců, mezi skutečnými a požadovanými výstupy. ANFIS používá metodu zpětného šíření „backpropagation“ samotnou, nebo v kombinaci s metodou nejmenších čtverců pro odhad parametrů členské funkce.

4.3 Vstupní data

Přístup modelování u systému ANFIS je podobný, jako u mnoha dalších systémových identifikačních technik. Za prvé lze předpokládat parametrizovanou modelovou strukturu (vstupy se vztahují k členským funkcím atd.) Dále je potřeba sbírat vstupní/výstupní data ve formě, která bude použitelná pro výcvik systému ANFIS. Po splnění těchto kritérií lze použít ANFIS k tréninku FIS modelu k emulaci tréninkových dat předložených ke změně parametrů členské funkce podle předem zvoleného kritéria chyb.

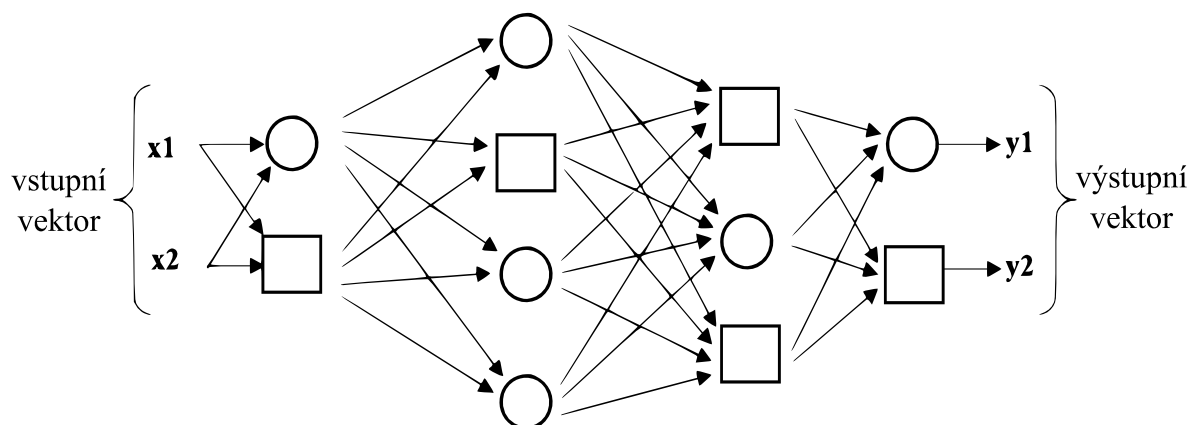
Obecně platí, že tento typ modelování funguje dobře, pokud data určená pro trénování parametrů členské funkce systému ANFIS jsou vzorově přesná k modelování FIS struktury. Avšak v některých případech jsou data shromažďována zašuměným měřením a tréninková data tudíž nemohou být použita. V těchto případech je vhodné data přezkoumat a zhodnotit.



Obr. 3: FIS - Fuzzy inferenční systém [38]

4.4 Architektura a základní pravidlo učení

Adaptivní síť (Obr. 4) je vícevrstevná dopředná síť, ve které každý uzel plní určitou funkci s ohledem na příchozí signály, stejně jako na sadu parametrů týkajících se tohoto uzlu. Vzorce pro funkce uzlu se mohou lišit od uzlu k uzlu a výběr jednotlivých funkcí uzlu závisí na celkové vstupně-výstupní funkci, které adaptivní síť pouze indikuje směr toku signálů mezi uzly; žádné váhy nejsou ve spojení. K odražení různých adaptivních schopností, jsou v adaptivní síti použity uzly ve tvaru kruhů a čtverců. Čtvercový uzel (adaptivní) má určité parametry, zatímco kruhový uzel (pevný) nemá žádné. Sada parametrů adaptivní sítě je spojení sady parametrů každého adaptivního uzlu. Aby bylo možné dosáhnout výsledků podle daných tréninkových dat a učící procedury na principu gradientního spádu, je třeba dodržet pravidla „IF/THEN“ a uvažovací fuzzy mechanismy.



Obr. 4: Adaptivní síť [38]

Lze předpokládat, že adaptivní síť má L vrstev a k -tá vrstva má počet $\#(k)$ uzlů. Je možno označit uzel v i -té části k -té vrstvy klasifikátory (k, i) a funkci uzlu (nebo výstup uzlu) klasifikátorem O_i^k . Zatímco výstup uzlu závisí na příchozích signálech a nastavených parametrech, je určeno

$$O_i^k = O_i^k(O_i^{k-1}, \dots, O_{\#(k-1)}^{k-1}, a, b, c, \dots) \quad (4.1)$$

kde a, b, c , atd., jsou parametry týkající se tohoto uzlu. (O_i^k je použito jako výstup uzlu a také jako funkce uzlu.) Za předpokladu, že daná tréninková data mají P vstupů, lze definovat míru chyb (nebo energii funkce) pro p -tý ($1 \leq p \leq P$) vstup vstupu tréninkových dat jako součet kvadratických chyb:

$$E_p = \sum_{m=1}^{\#(L)} (T_{m,p} - O_{m,p}^L)^2 \quad (4.2)$$

kde $T_{m,p}$ je m -tá součást p -tého cílového výstupního vektoru a $O_{m,p}^L$ je m -tá součást skutečného výstupního vektoru vytvořeného představením p -tého vstupního vektoru. Proto je celková míra chyb $E = \sum_{p=1}^P E_p$.

S cílem rozvíjet proceduru učení, která implementuje sestupný gradient E přes prostor parametru, je nejdříve nutné vypočítat míru chyb $\delta E_p / \delta O$ pro p -tá tréninková data a pro každý výstup uzlu O . Míra chyb pro výstup uzlu (L, i) může být vypočítána ze vztahu:

$$\frac{\delta E_p}{\delta O_{i,p}^k} = \sum_{m=1}^{\#(k+1)} \frac{\delta E_p}{\delta O_{m,p}^{k+1}} \frac{\delta O_{m,p}^{k+1}}{\delta O_{i,p}^k} \quad (4.3)$$

kde $1 \leq k \leq L - 1$. To znamená, že míra chyb vnitřního uzlu může být vyjádřena jako lineární kombinace četnosti chyb uzlů v další vrstvě. Proto pro všechny $1 \leq k \leq L$ a $1 \leq i \leq \#(k)$, kde lze najít $\delta E_p / \delta O_{i,p}^k$ jako v předchozích dvou rovnicích.

Pokud je α parametr dané adaptivní sítě, potom

$$\frac{\delta E_p}{\delta \alpha} = \sum_{O^* \in S} \frac{\delta E_p}{\delta O^*} \frac{\delta O^*}{\delta \alpha}, \quad (4.4)$$

kde S je soubor uzlů, jejichž výstupy jsou závislé na α . Potom se derivace celkové míry chyby E s váhou na α rovná

$$\frac{\delta E_p}{\delta \alpha} = \sum_{p=1}^P \frac{\delta E_p}{\delta \alpha}. \quad (4.5)$$

V důsledku toho, dojde k aktualizaci vzorce pro obecný parametr

$$\Delta \alpha = -\eta \frac{\delta E}{\delta \alpha} \quad (4.6)$$

kde η je parametr rychlosti učení, který může být dále vyjádřen jako

$$\eta = \frac{k}{\sqrt{\sum_{\alpha} \left(\frac{\delta E}{\delta \alpha}\right)^2}} \quad (4.7)$$

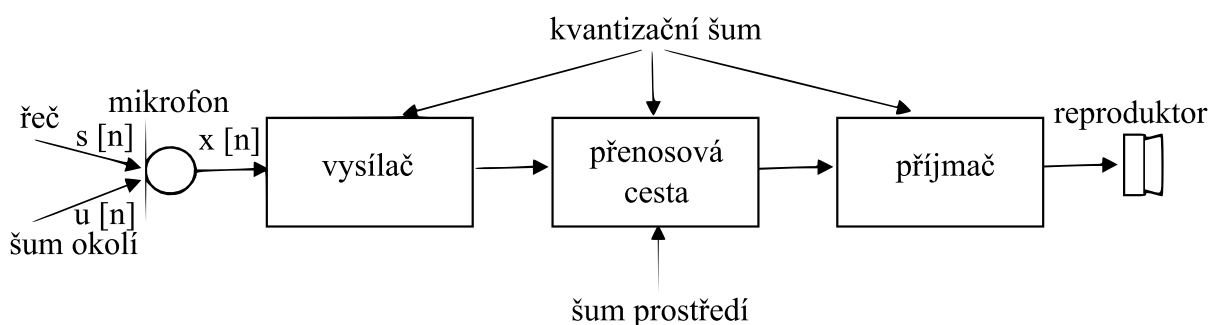
kde k je velikost kroku, délka každého přechodového gradientu v parametrickém prostoru. Obvykle můžeme změnit hodnotu k jako rychlost konvergence. Pro adaptivní sítě existují dva učící vzory a to pravidlo hybridní „Hybrid“ a zpětnovazební „Backpropagation“.

5 Řeč za přítomnosti šumu

Poslední dobou nastává rozmach aplikací pro zpracování řečového signálu v reálném prostředí. S příchodem těchto technologií nastávají situace, kdy analyzovaný signál obsahuje rušivé pozadí. V následujícím textu je popsána kategorizace možného rušení, měření úrovně aditivního šumu v řečovém signálu a detekce řečové aktivity v rušivém signálu. Tento text pomůže pochopit problematiku rušení řečových příkazů určených k ovládání inteligentní budovy.

5.1 Typy šumů

Při analýze řečového signálu lze pozorovat dvě základní možná rušení: aditivní šum a konvoluční zkreslení. Aditivní šum může být výsledkem operací v pevné řádové čarce při kódování a zpracování přenášeného signálu, tzv. kvantizační šum, nebo může být k řeči přičten jako šumové pozadí, ve kterém je řečový signál zaznamenáván, popřípadě může představovat šum přenosové cesty. Konvoluční šum bývá způsoben změnami parametrů přenosového prostředí, nebo přeslechy. [37]



Obr. 5: Výskyt šumu při přenosu řeči [37]

5.1.1 Aditivní šumy

Pro mikrofonom snímaný signál $x[n]$, který tvoří aditivní složky šumu $u[n]$ a řeči $s[n]$ platí následující vztah

$$x[n] = s[n] + u[n]. \quad (5.1)$$

Toto platí také pro komplexní spektrum, neboť DFT je lineární transformace

$$X[k] = S[k] + U[k]. \quad (5.2)$$

Pokud je řeč se šumem nekorelovaná, je rovněž autokorelace signálu $R_x[k]$ tvořena součtem korelací řeči $R_s[k]$ a šumu $R_u[k]$

$$R_x[k] = R_s[k] + R_u[k] \quad (5.3)$$

a také pro spektrální výkonové hustoty, pro které platí

$$S_x[k] = S_s[k] + S_u[k]. \quad (5.4)$$

V systémech, ve kterých se nachází více mikrofonů, lze dále rozeznat ještě tři základní typy aditivních šumů. Jejich konkrétní vlastnosti jsou určeny typem prostředí a způsobem generování. Patří sem koherentní šum, nekorelovaný šum a šum difúzní. Tyto šумы se liší hodnotou MSC a závislostí na frekvenci. Difúzní šum je generován odrazy v místnostech, nebo větším počtem nezávislých zdrojů, které jsou rozloženy v prostoru. [37]

5.1.2 Konvoluční šумы

Konvoluční šумы vznikají modifikací signálu $s[n]$ při jeho přenosu akustickým prostředím s impulsovou odezvou $h[n]$, nebo akustickým prostředím. Pro vznik konvolučních šumů platí vztah

$$x[n] = s[n] * h[n]. \quad (5.5)$$

Pro komplexní spektrum plyne z vlastností DFT vztah mezi spektrem řeči $S[k]$ a přenosovou funkcí cesty $H[k]$

$$X[k] = S[k]H[k]. \quad (5.6)$$

Pro spektrální hustotu platí

$$S_x[k] = S_x[k] = E[X^*[k]H^*[k]X[k]H[k]] = S_s[k]|H[k]|^2, \quad (5.7)$$

Protože přenosová funkce je nenáhodná veličina. Zpětnou transformací předchozího vztahu získáme pro autokorelaci signálu $R_x[k]$

$$R_x[k] = R_s[k] * R_u[k]. \quad (5.8)$$

Důležitou vlastností konvolučního šumu je jeho aditivní charakter v spektrální oblasti v důsledku aplikace logaritmu na amplitudové spektrum. Platí

$$C_x[n] = C_s[n] + C_h[n]. \quad [37] \quad (5.9)$$

5.1.3 Měření úrovně aditivního šumu v řeči

Standardním kritériem pro měření úrovně šumu v signálu je odstup signálu od šumu – SNR. Řečový signál je však silně nestacionární a v tomto důsledku není prostá aplikace tohoto kritéria na řeč optimální, proto se lze setkat s různými definicemi SNR pro účely měření šumu v řeči. [37]

Je-li $s[n]$ řečový signál s výkonem σ_s^2 a $n[n]$ aditivní šum s výkonem σ_n^2 představující rušivé pozadí, lze pro jejich směs $x[n] = s[n] + n[n]$ mající výkon σ_x^2 definovat následující kritéria:

- GSNR – globální SNR počítané přes celý signál,

$$\text{GSNR} = 10 \log \frac{\sigma_s^2}{\sigma_n^2} = 10 \log \frac{\sum_{n=0}^{l-1} s^2[n]}{\sum_{n=0}^{l-1} n^2[n]} \quad (5.10)$$

kde l je délka signálu,

- SNR – globální SNR počítané z úseků s řečovou aktivitou,

$$\text{SNR} = 10 \log \frac{\sum_{n=0}^{l-1} s^2[n] \text{vad}[n]}{\sum_{n=0}^{l-1} n^2 \text{vad}[n]} \quad (5.11)$$

kde $\text{vad}[n]$ nese informaci o řečové aktivitě pro daný vzorek signálu (0 – pauza, 1 – řeč).

- SNR_i – lokální (krátkodobé) SNR počítané v i -tém segmentu konečné délky při segmentaci řeči na kvazistacionární úseky (typicky 10-30 ms),

$$\text{SNR}_i = 10 \log \frac{\sigma_{s,i}^2}{\sigma_{n,i}^2} = 10 \log \frac{\sum_{n=0}^{M-1} s_i^2[n]}{\sum_{n=0}^{M-1} n_i^2[n]} \quad (5.12)$$

kde $s_i[n] = s[mi + n]$, $n_i[n] = n[mi + n]$, jsou segmenty signálu délky M vybírané s krokem m

- SSNR – segmentální SNR, dané průměrem lokálních SNR v segmentech s řečovou aktivitou,

$$\text{SSNR} = \frac{1}{K} \sum_{i=0}^{L-1} \frac{\sigma_{s,i}^2}{\sigma_{n,i}^2} * \text{VAD}_i, \quad (5.13)$$

Kde L je celkový počet segmentů signálu, K počet segmentů s řečovou aktivitou a hodnota VAD_i , nabývá opět hodnot 0 a 1, nese informaci o řečové aktivitě v i -tém segmentu

- ASNR – aritmetické segmentální SNR, dané průměrem lineárních lokálních SNR v segmentech s řečovou aktivitou a následným přepočtem na dB,

$$\text{ASNR} = 10 \log \frac{1}{K} \sum_{i=0}^{L-1} \frac{\sigma_{s,i}^2}{\sigma_{n,i}^2} * \text{VAD}_i. \quad (5.14)$$

Název aritmetické segmentální SNR je odvozen od způsobu průměrování lineárních lokálních SNR. Pro standardní segmentální SNR totiž na základě vlastností logaritmu realizujeme vlastně geometrický průměr lineárních lokálních SNR.

Opět je třeba zdůraznit, že se segmentují pouze segmenty s řečovou aktivitou. Ve výše uvedeném vztahu je tedy v mezích sumace celkový počet segmentů signálu L , do průměru se však započítává pouze K segmentů, pro které platí $\text{VAD}_i = 1$. [37]

Pro takto definovaná kritéria lze na základě teoretických rozborů i provedených experimentů, udělat následující nejdůležitější závěry:

GSNR není vhodným kritériem, neboť je výrazně ovlivněno různě trvajících pauzami v promluvách. Z tohoto důvodu je pro přesnější kvantifikaci úrovně šumu v řeči vhodné použít detektor řeči a vyčíslit výkony globálně, avšak pouze v úsecích s řečovou aktivitou, stejně jako v kritériu SNR.

Vhodnějším kritériem je SSNR, které je schopno nejlépe postihnout měnící se dynamiku řeči či měnící se intenzitu rušivého pozadí. SSNR však poskytuje mírně vychýlené hodnoty. Pro signál se SNR okolo 0dB vychází SSNR přibližně o 5dB menší.

Algoritmus funkce SSNR porovnává zašuměný signál S se signálem referenčním R a vypočte segmentální odstup signálu k šumu (SNR) v decibelech. Tyto signály, které jsou porovnávány, musejí mít stejnou délku a musejí být rozděleny do nepřekrývajících se rámců, ze kterých se počítají jednotlivé hodnoty SNR. [37, 38]

ASNR je kritérium definované rovněž na segmentální bázi jako SSNR avšak poskytuje podobné hodnoty jako globálně počítané SNR. [37]

5.1.4 Metody odhadu SNR řeči

Pro účely měření SNR v reálných signálech však nejsou výše uvedená kritéria přímo použitelná. Problémem je dostupnost pouze směsi řeči a rušivého pozadí $x[n]$. Partikulární výkony řeči, σ_s^2 , a šumu, σ_n^2 , je nutné odhadnout z jednoho signálu. To ale není vždy jednoduchá úloha, zejména pak v případech, kdy je úroveň rušivého pozadí velmi vysoká, např. v signálech snímaných v jedoucím automobilu. [37]

Standardním postupem je v této situaci obvykle nalezení odhadu výkonu šumu $\hat{\sigma}_n^2$ a získání výkonu řečového signálu odečtením této hodnoty od výkonu směsi, tj. $\hat{\sigma}_s^2 = \sigma_x^2 - \hat{\sigma}_n^2$. Tento postup je možný za předpokladu, že řečový signál a šum jsou nekorelované. V tomto případě pak platí

$$\sigma_x^2 = \sigma_s^2 + \sigma_n^2. \quad (5.15)$$

Na základě výsledků prezentovaných dříve jsou prakticky použitelné odhady výše definovaných kritérií následující.

I. Odhad SNR s detektorem řečové aktivity

$$\widehat{SNR} = 10 \log \frac{\hat{\sigma}_s^2}{\hat{\sigma}_n^2}, \quad (5.16)$$

$$\hat{\sigma}_n^2 = \frac{1}{l_n} \sum_{n=0}^{l-1} x^2[n] * |1 - vad[n]|, \quad (5.17)$$

$$\hat{\sigma}_s^2 = \frac{1}{l_s} \sum_{n=0}^{l-1} x^2[n] * vad[n] - \hat{\sigma}_n^2. \quad (5.18)$$

Tato metoda poskytuje relativně dobrý odhad SNR v případě, že detekce řeči je spolehlivá. V odhadu však bude znatelná určitá chyba. Chyba se projevuje především díky konečné délce analyzovaných signálů, kdy je podmínka splněna pouze přibližně. [37]

- II. Odhad SSNR resp. ASNR s detektorem řečové aktivity a exponenciálním průměrováním výkonu šumu v řečových pauzách

$$\widehat{SSNR} = \frac{1}{K} \sum_{i=0}^{L-1} 10 \log \left(\frac{\sigma_{x,i}^2 - \hat{\sigma}_{n,i}^2}{\hat{\sigma}_{n,i}^2} * VAD_i \right), \quad (5.19)$$

$$\widehat{ASNR} = 10 \log \left(\frac{1}{K} \sum_{i=0}^{L-1} \frac{\sigma_{x,i}^2 - \hat{\sigma}_{n,i}^2}{\hat{\sigma}_{n,i}^2} * VAD_i \right) \quad (5.20)$$

$$\hat{\sigma}_{n,i}^2 \begin{cases} p * \hat{\sigma}_{n,i-1}^2 + (1-p) * \sigma_{x,i}^2, & \text{jestliže } VAD_i = 0 \\ \sigma_{n,i-1}^2, & \text{jestliže } VAD_i = 1. \end{cases} \quad (5.21)$$

$$\sigma_{x,i}^2 = \frac{1}{M} \sum_{i=0}^{M-1} x_i^2[n]. \quad [37] \quad (5.22)$$

6 Norma ITU-T P.863

Norma ITU-T P.863 slouží k posuzování kvality řečového signálu, poskytuje objektivní metodu měření kvality řečového signálu a to hlavně v oblasti telefonních hovorů. Lze ji ovšem použít i jako hodnotící parametr pro posouzení kvality řečového signálu, jenž je určen k ovládní prvků v chytré domácnosti založené na standardu KNX.

Tato norma, popisuje objektivní metodu předpovídání celkového poslechu kvality hovoru z úzkého pásma (300 – 3400 Hz) do super-širokého pásma (50 – 14000 Hz). Doporučení ITU-T P.863 podporuje dva provozní režimy pro úzké a super-široké pásmo. Toto doporučení představuje vysokou úroveň popisu metody a rady, jak ji použít. ITU-T P.863 definuje algoritmus pro hodnocení kvality řeči současných a budoucích telefonních systémů, které využívají širokou škálu kódování, transportu a zesilujících technologií. Algoritmus měření je plný referenční model, který je založen na srovnání mezi známým referenčním signálem a zachyceným zhoršeným signálem. Vše je v souladu s algoritmy popsány v doporučeních ITU-T P.861 a ITU-T P.862. Doporučení ITU-T P.863 obsahuje aktuální požadavky průmyslu a zejména umožňuje zhodnotit super-širokopásmový řečový signál, stejně jako sítě a kodeky, které zavádějí deformaci času.

Na základě referenčních výsledků prezentovaných v rámci studií ITU-T, přehledu testovaných faktorů, kódovacích technologií a aplikací, ke kterým se toto doporučení vztahuje, jsou uvedeny v následujících tabulkách. Tabulka 1 (Tab. 1) představuje faktory a aplikace zahrnuté do specifikace požadavků, které byly použity při fázovém výběru algoritmu ITU-T P.863. Je třeba poznamenat, že výkon algoritmu ITU-T P.863 za každého jednotlivého stavu v tabulce 1 se v této tabulce neodráží. Tabulka 2 (Tab. 2) uvádí seznam podmínek, pro které toto doporučení není vhodné použít. V tabulce 3 (Tab. 3) jsou testovací proměnné, kterým musí být věnován další výzkum, nebo při kterém se ITU-T P.863 vztahuje na tvrzení ve spojení s předpověďmi. Tabulka 4 (Tab. 4) uvádí faktory, technologie a aplikace pro které algoritmus ITU-T P.863 není v současné době ověřen. Je třeba také poznamenat, že algoritmus ITU-T P.863 neposkytuje komplexní hodnocení kvality přenosu. Doporučení měří pouze dopady jednosměrného zkreslení řeči a zašumění řečové kvality.

Tab. 3: Faktory a aplikace zahrnující požadované specifikace použité ve výběrové fázi v algoritmu ITU-T P.863 [40]

Testovací faktory
Hladiny hlasových vstupů ke kodeku
Chyby přenosového kanálu
Ztráty paketu a ukrytí ztráty paketu
Přenosová rychlost v případě, že kodek má více, než 1 mód přenosové rychlosti
Překódování
Akustický šum ve vysílacím prostředí
Vliv kolísajícího zpoždění pouze v poslechových testech

Krátkodobá deformace zvukového signálu
Dlouhodobá deformace zvukového signálu
Poslechové úrovně mezi 53 a 78 dB SPL v super-širokopásmovém módu
Ztráty paketu a ukrytí ztráty paketu s typovými kodeky PCM
Časové a amplitudové výstřihy řeči
Testovací faktory
Lineární zkreslení včetně omezení šířky pásma a spektrálního tvarování
Frekvenční odezva
Kódovací technologie
TU-T G.711, ITU-T G.711 PLC, ITU-T G.711.1
ITU-T G.718, ITU-T G.719, ITU-T G.722, ITU-T G.722.1, ITU-T G.723.1, ITU-T G.726, ITU-T G.728, ITU-T G.729
GSM-FR, GSM-HR, GSM EFR
AMR-NB, AMR-WB (ITU-T G.722.2), AMR-WB+
PDC-FR, PDC-HR
EVRC (ANSI/TIA-127-A), EVRC-B (TIA-718-B)
Skype (SILK V3, iLBC, iSAC and ITU-T G.729)
Speex, QCELP (TIA-EIA-IS-733), iLBC, CVSD (64 kbit/s, "Bluetooth")
MP3, AAC, AAC-LD
Aplikace
Vyhodnocovací kodek
Testovací terminál, vliv akustické cesty v odesílaném a přijímaném směru
Rozšíření pásma
Testování živé sítě pomocí digitálního nebo analogového připojení k síti
Testování emulovaných a prototypových sítí
UMTS, CDMA, GSM, TETRA, WB-DECT, VoIP, POTS, PSTN, Video Telephony, Bluetooth
Detekce hlasové aktivity (VAC), automatické zesílení ovládání (AGC)
Zařízení pro zesílení hlasu (VED), redukce šumu
Přerušované vysílání (DTX)

Tab. 4: Algoritmus ITU-T P.863 není určen pro použití s následujícími proměnnými [40]

Testovací faktory
Účinek zpoždění v testech konverzace
Odezva mluvčího
Boční tón
Akustický hluk v přijímacím prostředí
Aplikace
Nerušivá měření
Obousměrný výkon komunikace

Tab. 5: Testovací proměnné pro které je nutné další vyšetření [40]

Testovací faktory
Akustické nahrávky využívající volná pole mikrofону bez simulátorů ušního kanálu

Tab. 6: Faktory, technologie a aplikace, pro které ITU-T P.863 algoritmus nebyl v současné době potvrzen [40]

Testovací faktory
Závislost na mluvčím
Více mluvčích najednou
Bitová nesouhlasná rychlost mezi kodérem a dekodérem, pokud má kodek více, než 1 mód přenosové rychlosti
Informační síť signálu jako vstup ke kodeku
Umělé řečové signály jako vstup ke kodeku
Hudba jako vstup ke kodeku
Odezva posluchače
Kódovací technologie
Kódovací technologie pracují pod 4 kbit/s
Aplikace
Živé VoLTE sítě

Toto doporučení, je více citlivé na velmi nízkou úroveň šumu v super-širokém pásmu, než v pásmu úzkém. Testovací sada ITU-T P.863 je schopna pracovat s těmito jazyky: Americká angličtina, britské angličtina, čínština (mandarínština), čeština, holandština, francouzština, němčina, italština, japonština švédština a švýcarská němčina. Subjektivní experimenty byly prováděny ve zkušebních laboratořích v těchto jednotlivých zemích. ITU-T P.863 je budoucnost technologií pro testování kvality hlasu. Tato norma byla vybrána jako nový ITU-T standard pro testování hlasové kvality. Toto doporučení bylo vyvinuto v letech 2006 – 2010. Účelem objektivního ITU-T P.863 modelu je predikovat z celkového poslechu kvalitu hlasu z úzkého pásma (300 až 3400 Hz) do super-širokého pásma (50 až 14000 Hz). Jsou zde zahrnuty všechny řečové složky, které jsou obvykle součástí v zašuměných podmínkách a v podmínkách bez šumu. Termín „poslouchání řečové kvality“ znamená hodnocení celkové kvality řeči a také to jak je vnímána. [40]

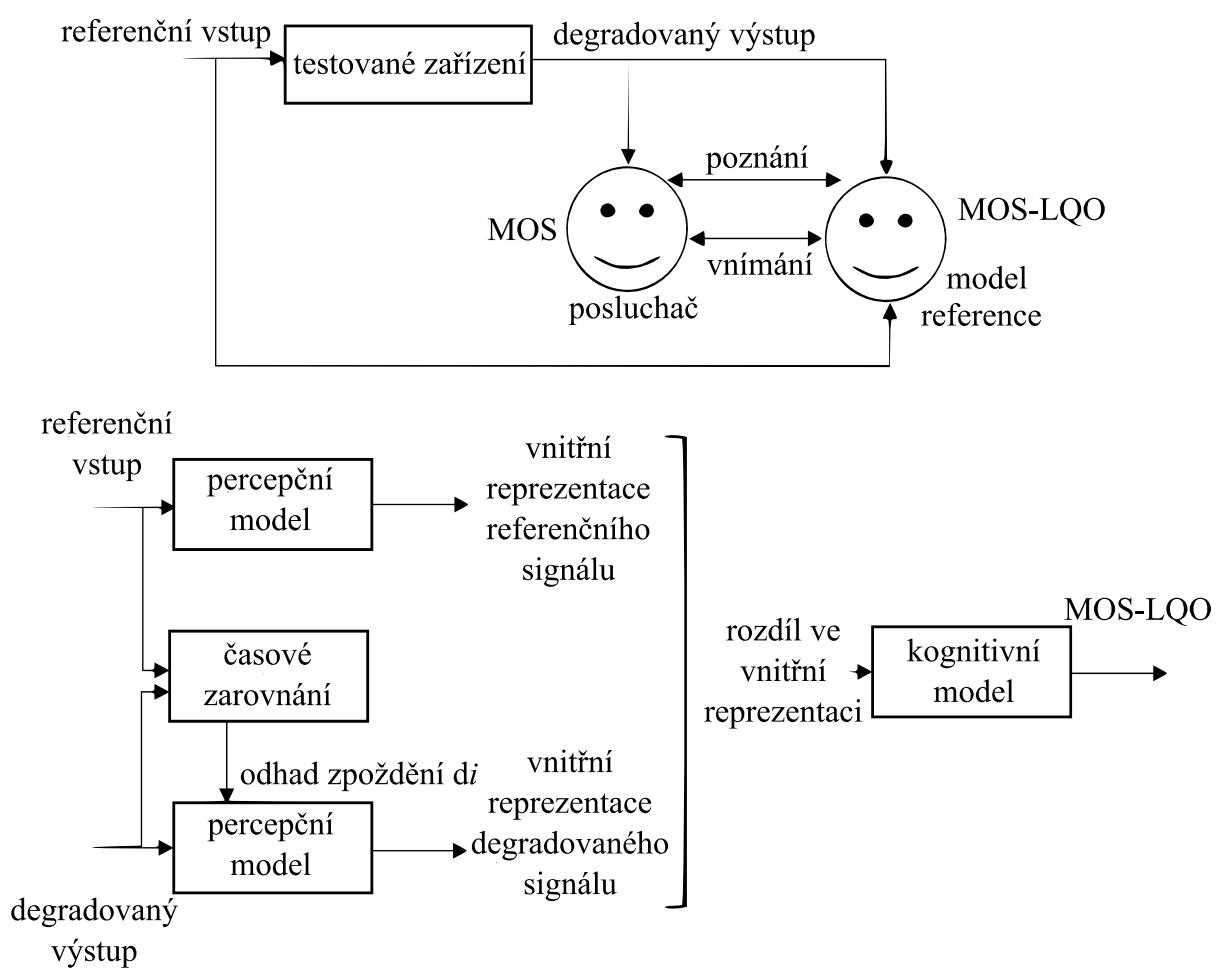
6.1 Algoritmus ITU-T P.863

Algoritmus ITU-T P.863 porovnává referenční signál $X(t)$ s degradovaným signálem $Y(t)$, kde $Y(t)$ je výsledným signálem prošlým reálným prostředím, nebo komunikačním systémem. Výstupem tohoto algoritmu je předpověď vnímané kvality, kterou by měl mít signál $Y(t)$ v subjektivním poslechovém testu. V prvním kroku, jsou oba signály rozděleny do velmi malých časových úseků – dále popisovaných jako rámce. Poté se vypočte zpoždění referenčního signálu každého rámce vůči přidruženému rámci signálu degradovaného. Na základě tohoto zpoždění je odhadnuta vzorkovací frekvence degradovaného signálu. Pokud se tato vzorkovací frekvence významně liší od vzorkovací frekvence referenčního signálu, nebo pokud má jakýkoli signál vyšší vzorkovací frekvenci, je tato frekvence převzorkována a zpoždění je přehodnoceno. [40]

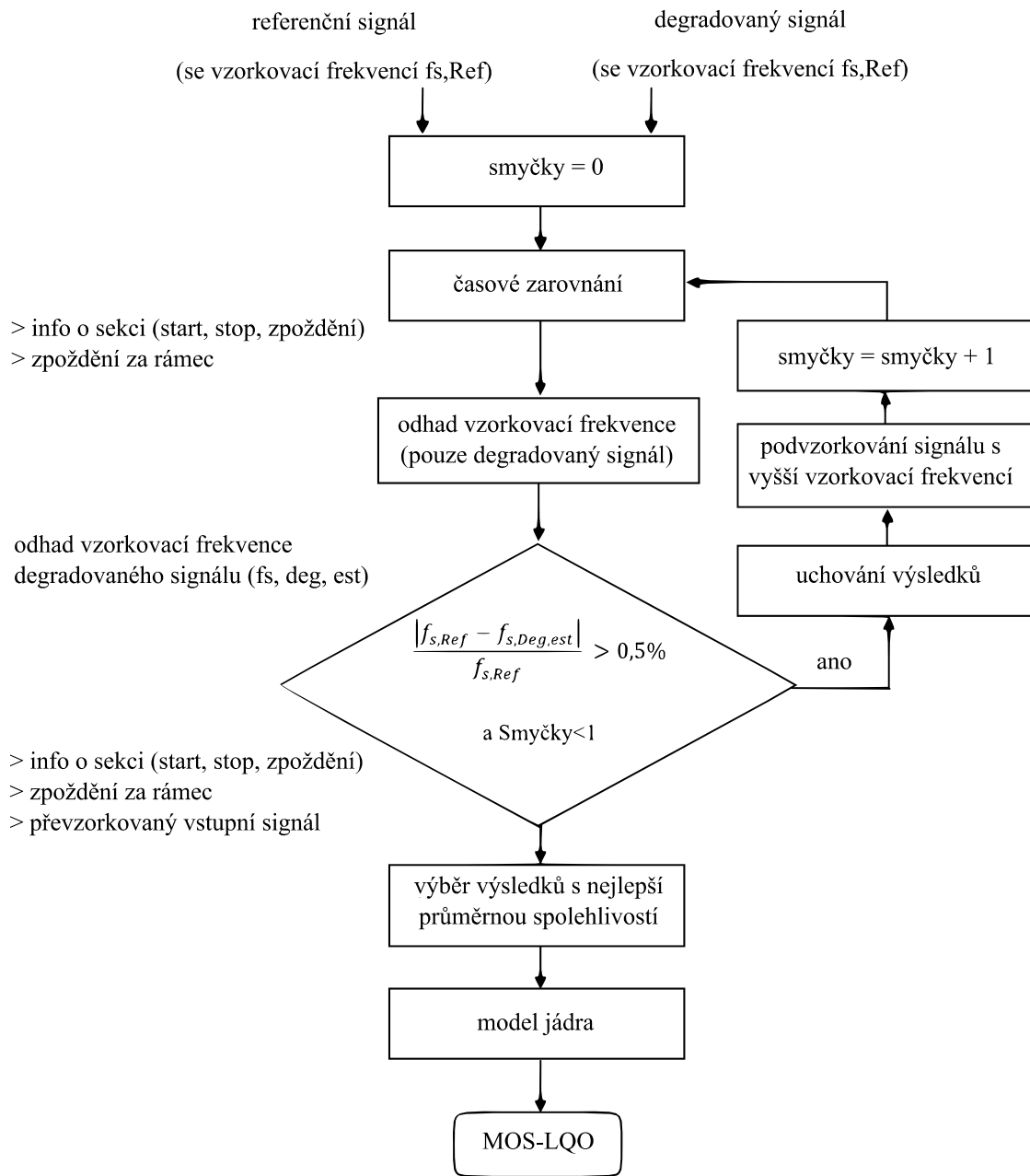
Na základě množiny zpoždění, které jsou nalezeny, algoritmus ITU-T P.863 porovnává referenční (vstupní) signál s degradovaným (výstupním) signálem testované soustavy jak je uvedeno na obrázku 7. Klíčem k celému tomuto procesu je transformace těchto dvou signálů na vnitřní znázornění, které je analogické psychofyzickému znázornění zvukových signálů v lidském sluchovém systému, kde je přihlédnuto k vnímání frekvence a hlasitosti. Toho je dosaženo v několika fázích:

- Časové zarovnání.
- Úroveň zarovnání ke kalibrované úrovni poslechu.
- Časově-frekvenční mapování
- Frekvenční deformace a kompresivní změna velikosti hluku

Algoritmus ITU-T P.863 je navržen tak, aby bral v úvahu dopad úrovně přehrávání pro vnímanou kvalitu předpovědi v super-širokopásmovém módu. Úroveň přehrávání je počítána vzhledem k nominální úrovni -26 dBov, které představují 73 dB(A) v diotické úrovni. [40]



Obr. 6: Přehled základní filozofie použité v ITU-T P.863 [40]



Obr. 7: Všeobecný přehled algoritmu ITU-T P.863 [40]

6.2 Algoritmus PESQ

Algoritmus PESQ, neboli percepční posuzování kvality řečového signálu, je objektivní metoda měření kvality řečového signálu stanovená normou ITU-T P.863. PESQ v podstatě počítá subjektivní MOS hodnoty, srovnáním hlasových nahrávek, které byly zaznamenány. Jedná se vždy o srovnání čisté, kvalitní referenční nahrávky a degradované, obsahující šum. Je nutno poznamenat, že PESQ měří pouze zkreslení a šum, jenž ovlivňují kvalitu signálu a nestará se o ostatní aspekty, jako jsou ztráta hlasitosti, zpoždění, boční tón a ostatní poruchy ve zvukovém signálu.

PESQ bylo vytvořeno pro výpočet hodnot MOS k posouzení kvality přeneseného či zaznamenaného zvuku. Tuto kvalitu lze posoudit v určité škále výběrem z pěti možností: „špatný“, „slabý“, „ucházející“, „dobrý“ a „vynikající“. Tyto kritéria jsou pak přiřazena k číselným hodnotám (1, 2, 3, 4, 5) a to ve stejném pořadí. Výsledné číslo je bráno právě jako „Průměrný posudek výsledku“ (MOS). Například výsledné číslo 3,2 bude tímto algoritmem posouzeno jako „ucházející“.

Přesnost algoritmu PESQ měřící MOS, musí být stanovena porovnáním výsledků referenčního a degradovaného signálu. Například předpokládejme, že MOS pro konkrétní referenční signál je "3.2". Je-li výsledný degradovaný signál téměř shodný s referenčním, pak PESQ hodnota bude velmi blízká číslu "3.2".

Tab. 7: Hodnoty MOS v algoritmu PESQ

Hodnota MOS	Hodnocení signálu
1	špatný
2	slabý
3	ucházející
4	dobrý
5	vynikající

7 Střední kvadratická chyba

„Root Mean Square Error” neboli střední kvadratická chyba je často používaným měřícím parametrem v hodnotících modelových studiích. V tomto případě bude tento posuzovací parametr využit k posouzení rozdílu mezi referenčním a zašuměným příkazem určených k hlasovému ovládní KNX systému v inteligentní budově.

Střední kvadratická chyba (RMSE) byla použita jako standard statistické metriky. Pro pochopení tohoto algoritmu předpokládejme, že máme k dispozici n -vzorků modelové chyby ϵ vypočtené jako $(e_i, i = 1, 2, \dots, n)$. Nejistoty zjištěny pozorovanými chybami, nebo použitými srovnávacími metodami k pozorování, zde nejsou zahrnuty. Také lze předpokládat, že množina chybových vzorků ϵ je objektivní. RMSE je vypočteno na základě následujícího vzorce:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (7.1)$$

Základní předpoklad při výpočtu RMSE je, že chyby jsou objektivní a je dodrženo normální rozdělení.

Tab. 8: RMSE výsledky z náhodně generovaných nepravých chyb s nulovou střední hodnotou a jednotkovou odchylkou Gaussova rozdělení. [41]

n	RMSE				
4	0,92	0,65	1,48	1,02	0,79
10	0,81	1,10	0,83	0,95	1,01
100	1,05	1,03	1,03	1,00	1,04
1000	1,04	0,98	1,01	1,00	1,00
10 000	1,00	0,98	1,01	1,00	1,00
100 000	1,00	1,00	1,00	1,00	1,00
1 000 000	1,00	1,00	1,00	1,00	1,00

Použití RMSE v tomto případě pomáhá zajistit ucelený obraz o rozložení chyb. Tabulka zobrazuje hodnoty RMSE pro náhodně generované nepravé chyby s nulovou střední hodnotou a jednotkovou odchylkou Gaussova rozdělení. Pokud velikost vzorku dosáhne 100 nebo více, lze pomocí vypočteného RMSE rekonstruovat chybové rozdělení blízké přesnému měření se standardní odchylkou 5%. Čím více vzorků je k dispozici, tím je rekonstrukce rozložení chyb pomocí RMSE spolehlivější. [41]

8 Smart Home

V současné době dochází k velmi rychlému rozvoji chytré domácnosti, neboli „Smart Home“. S rozmachem technologií rostou také nároky uživatelů na pohodlí, bezpečnost a úsporu energie. Proto také narůstá podíl automatizace jak v bytové výstavbě, tak i v účelových stavbách. V neposlední řadě nelze opomenout také aspekt zajištění bezpečnosti. Došlo také k vývinu samostatného a významného aplikačního oboru automatizační techniky, který poskytuje zákaznický orientovaná řešení jak provozovatelům, tak uživatelům všech druhů budov. Mezi automatizační prvky lze zařadit snímače, akční členy, vizualizace procesů a regulátory. Tyto prvky zpravidla fungují decentralizovaně. Aby tato zařízení mohla provádět složité operace, musejí být vybaveny komunikačním systémem pro vzájemnou a oboustrannou výměnu dat. Pro tento účel jsou použity sběrnice a sítě. Jako jedním z těchto sběrnicových systémů je Evropská instalační sběrnice KNX. [39] Ovládání tohoto systému je velmi variabilní a každý uživatel si ho může zařídit podle svých představ. Jednou z novinek je právě i hlasové ovládání, které hranici komfortu posunuje ještě dále.

8.1 Bytová výstavba

V soukromé bytové výstavbě se automatizované funkce se postupně stává určitým standardem. Mezi základní prvky lze jednoznačně zařadit regulaci spotřeby energie řízeným vytápěním a regulací teploty v jednotlivých místnostech. Dalším příkladem automatizace funkcí lze uvést řízení osvětlení. Hlavní roli zde má samočinný tepelný spínač, který detekuje vstup osoby a snímač intenzity osvětlení, který sepne v případě, že je již dostatečně šero. Dále nelze opominout bezpečnostní prvky a těmi jsou alarmy, založené na souboru různých čidel.

Důvody k automatizaci domácnosti lze shrnout ve třech základních oblastech:

- komfort
- bezpečnost
- hospodárnost a úspora energie. [39]

8.2 Účelové stavby

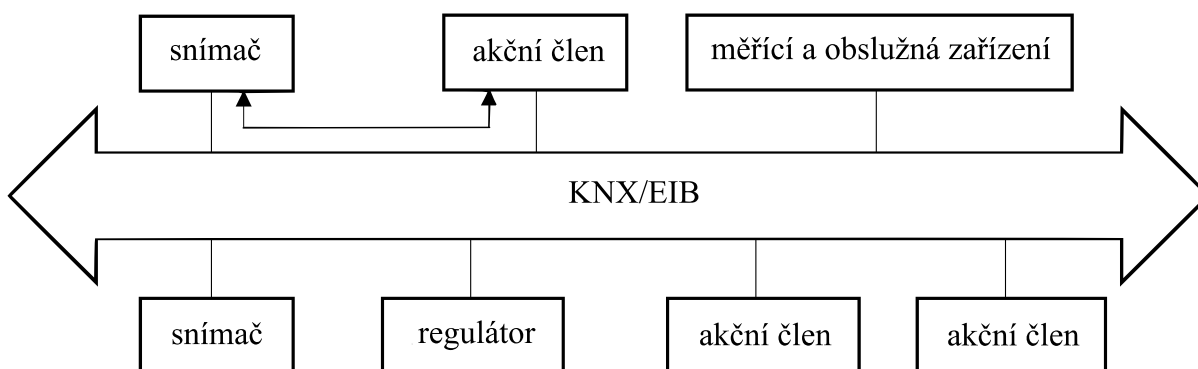
Mezi tyto budovy lze zařadit kancelářské budovy, nemocnice, nákupní centra, školy a podobně. Tento druh budov se od bytových výstaveb odlišuje přesně vymezeným zaměřením. V těchto budovách se nachází mnohem více automatizačních systémů jako například chladicí přístroje a vzduchotechnika. Tyto stavby jsou zpravidla vybaveny zpravidla náročnými řídicími a regulačními systémy. Vedle optimalizace spotřeby energie se projevuje úspora financí i ve snížení počtu pracovníků obsluhy. Všechny prvky jsou ovládány z přístupných míst jako například žaluzie, nebo osvětlovací tělesa a to přispívá ke komfortu a vyšší výkonnosti pracovníků

Automatizace v účelové výstavbě nabývá na významu zejména v těchto oblastech:

- komfort
- flexibilita
- komunikace prostřednictvím sběrnicových systémů a sítí
- hospodárnost a úspora energie. [39]

8.3 KNX

Evropská instalační sběrnice KNX/EIB je komunikační průmyslový systém, který je používán v systémové technice budov pro síťové inforatické spojení různých zařízení (snímače, regulátory, akční členy, řídicí a regulační zařízení). Jedná se o standard rozšířený na celém světě, který je určený pro řízení provozně technických funkcí v budovách. Vlastníkem a zakladatelem této technologie je Asociace KNX. Na základě certifikace produktů je zaručena vzájemná kompatibilita výrobků dodávaných od odlišných firem. Tato kompatibilita poskytuje flexibilitu na vysoké úrovni. Tento inovativní systém poskytuje uživatelům různé aplikační možnosti sjednocení různých technologií. [39]

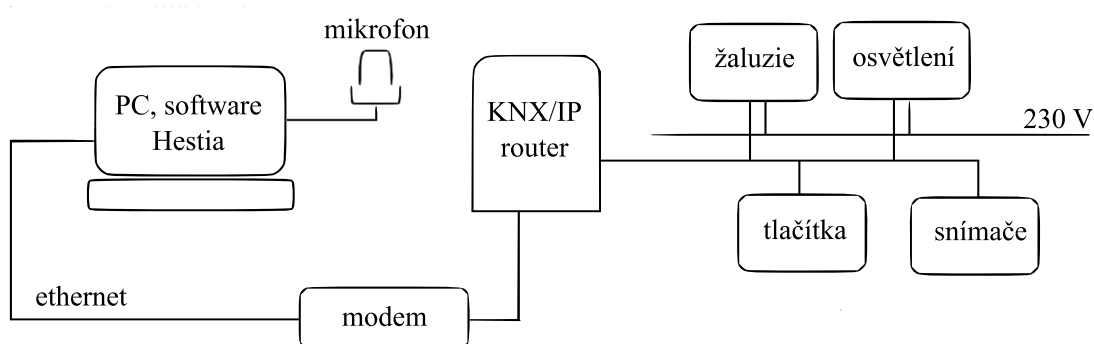


Obr. 8: Informatické zasíťování zařízení systémové techniky budov sběrnici KNX/EIB [39]

Komunikační data jsou vložena do datového telegramu a prostřednictvím instalační sběrnice digitálně přenesena. Sběrnici je možno realizovat různými způsoby, například u KNX/TP je to „Twisted Pair“ kabel, neboli kroucený pár metalických vodičů, u KNX.PL jde o silový kabel a u KNX.RF se používá radiové spojení. Pro přenos dat lze také využít optický kabel. [39]

9 Software Hestia pro ovládání standardu KNX

Tato aplikace byla navržena a naprogramována paní Ing. Janou Bešešovou a byla realizována v rámci diplomové práce v roce 2016 na VŠB – TUO. Software se nazývá Hestia 11 [42] a umožňuje ovládání inteligentních budov, které jsou vybaveny sběrníkovým standardem KNX. Aplikace se skládá z několika obrazovek. Pro uživatele a pro testování funkčnosti je klíčová obrazovka s ovládacími prvky. Na této obrazovce je zobrazen seznam místností a v jednotlivých místnostech nakonfigurovaná zařízení. Pomocí hlasových povelů lze ovládat osvětlení a žaluzie. V případě osvětlení je možno ovládat rozsvěcování a zhasínání světel a lze také měnit jas světel pomocí tlačítek pro stmívání. Žaluzie lze stahovat nahoru a dolů. Dále je zde možnost žaluzie umístit do určité výšky, a také poslední naklopit lamely žaluzií do požadovaného úhlu. Tyto jednotlivé úkony lze provádět manuálně přímo v aplikaci, ale také pomocí hlasových příkazů v angličtině.



Obr. 9: Schéma zapojení softwaru Hestia

Tab. 9: Příkazy hlasového ovládání aplikace Hestia

Příkaz	Provedení
"Go to control page"	návrat do hlavního menu
"Select room 0-9"	výběr místnosti 0-9
"Select device 0-9"	výběr zařízení 0-9
"Light on"	rozsvítit světlo
"Light off"	zhasnout světlo
"Blind up"	vytažení rolet
"Blind down"	zatažení rolet
"Change blind height"	změna výšky rolet
"Write height value 0-100"	výška rolet 0-100

9.1 Testování aplikace v reálném prostředí

9.1.1 Simulační sestava KNX

Funkčnost aplikace Hestia byla testována na výukových modulech vytvořených na Vysoké škole báňské a určených pro výuku studentů o technologiích o sběrníkovém systému KNX. Tento výukový modul sestává ze čtyřech samostatných panelů (Obr. 10), kde každý z nich zastává svou funkci. Hlavní panel zobrazuje půdorys bytu, který simuluje reálnou stavbu, ve které jsou rozmístěny aktory. Další panel sestává z tlačítek, pomocí kterých lze manuálně ovládat jednotlivé aktory. Aby byl systém funkční, je třeba komunikaci naprogramovat v KNX softwaru ETS. Na dalším panelu je umístěna žaluzie, snímač jasu a teploty a zásuvka. Poslední panel nese dotykovou obrazovku, přes kterou můžeme komunikovat se systémem KNX. Aplikace Hestia byla na tomto simulačním modulu úspěšně otestována.



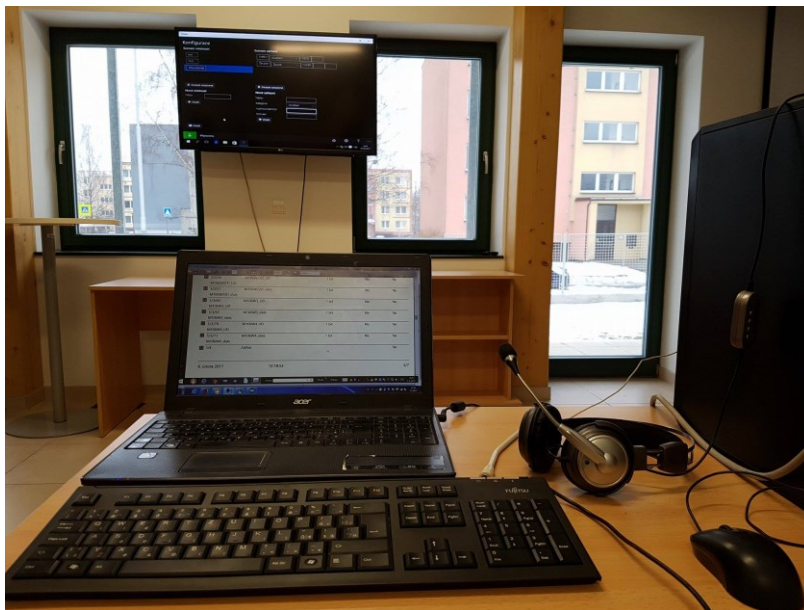
Obr. 10: Simulační sestava KNX systému

9.1.2 Výzkumné a inovační centrum

Po otestování aplikace na simulačních panelech, bylo nutné ji také vyzkoušet v reálném prostředí budovy. Pan Ing. Jan Vaňuš Ph.D. mi umožnil přístup do výzkumného a inovačního centra, které je umístěno v areálu stavební fakulty VŠB. V této budově, je nainstalována spousta inovativních technologií, včetně sběrníkového systému KNX. Nejprve bylo nutné napojit počítač s aplikací Hestia na KNX-IP router, poté zjistit skupinové adresy jednotlivých aktorů. Ty nastavit v aplikaci Hestia a poté bylo možno přejít k testování. Testovali jsme ovládání světel a žaluzií, video je dostupné v příloze IV na CD. Samotné testování bylo úspěšné, aplikace byla schopna rozpoznat příkazy a ovládat jednotlivé aktory. Netestovali jsme zde ovšem funkčnost za přítomnosti šumu.



Obr. 11: Výzkumné a inovační centrum



Obr. 12: Zapojená sestava aplikace Hestia

9.2 Experiment rozpoznání příkazů

Následuje experiment s úspěšností rozpoznání konkrétních příkazů za určitých podmínek. Pro tento experiment byly vybrány čtyři základní příkazy pro ovládání světel a žaluzií: „*Light on*“, „*light off*“, „*blind up*“ a „*blind down*“. Pro záznam zvuku byl vybrán a použit mikrofon „Trust All-round USB Microphone“ s vhodnými parametry pro záznam zvuku v prostoru. Vzdálenost řečníků od mikrofonu byla stanovena na 0,5 m. Každý konkrétní příkaz byl vyřčen stokrát, ve třech případech - bez šumu, který je brán jako referenční a se dvěma různými reálnými šumy. Experimentu se zúčastnili 4 řečníci, dva muži a dvě ženy.

Tab. 10: Úspěšnost rozpoznání pro 1. řečníka

Mikrofon:	Trust All-round USB Microphone	
Řečník:	Řečník 1 - muž	
Příkaz:	"Light on"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	33 %	17 %
SSNR (dB)	7,36	10,57
Příkaz:	"Light off"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	37 %	20 %
SSNR (dB)	8,12	11,73
Příkaz:	"Blind up"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	35 %	39 %
SSNR (dB)	13,62	12,83
Příkaz:	"Blind down"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	28 %	25 %
SSNR (dB)	10,54	11,86

Tab. 11: Úspěšnost rozpoznání pro 2. řečníka

Mikrofon:	Trust All-round USB Microphone	
Řečník:	Řečník 2 - muž	
Příkaz:	"Light on"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	29 %	21 %
SSNR (dB)	6,28	8,36
Příkaz:	"Light off"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	34 %	23 %
SSNR (dB)	7,19	6,39
Příkaz:	"Blind up"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	37 %	35 %
SSNR (dB)	8,42	8,17
Příkaz:	"Blind down"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	31 %	28 %
SSNR (dB)	7,49	6,83

Tab. 12: Úspěšnost rozpoznání pro 3. řečníka

Mikrofon:	Trust All-round USB Microphone	
Řečník:	Řečník 3 - žena	
Příkaz:	"Light on"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	25 %	17 %
SSNR (dB)	6,93	6,13
Příkaz:	"Light off"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	30 %	19 %
SSNR (dB)	7,89	6,24
Příkaz:	"Blind up"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	31 %	34 %
SSNR (dB)	8,34	9,15
Příkaz:	"Blind down"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	32 %	24 %
SSNR (dB)	9,81	7,56

Tab. 13: Úspěšnost rozpoznání pro 4. řečníka

Mikrofon:	Trust All-round USB Microphone	
Řečník:	Řečník 4 - žena	
Příkaz:	"Light on"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	31 %	19 %
SSNR (dB)	8,29	6,93
Příkaz:	"Light off"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	32 %	21 %
SSNR (dB)	8,94	7,25
Příkaz:	"Blind up"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	32 %	31 %
SSNR (dB)	9,86	8,49
Příkaz:	"Blind down"	
Typ signálu:	zašuměný	zašuměný
Typ šumu:	mikrovlnná trouba	rychlovarná konvice
Úspěšnost rozpoznání:	34 %	32 %
SSNR (dB)	10,64	9,84

Vzhledem k výsledkům tohoto experimentu lze říci, že nezáleží na pohlaví řečníka. Dále je patrné, že foneticky výraznější příkazy obsahující na začátku párové znělé souhlásky jako „*blind up*“ a „*blind down*“ mají lepší statistiku úspěšného rozpoznání oproti příkazům „*light on*“ a „*light off*“ (viz Tabulka 13). Nejvíce záleží na správném a co nejlépe srozumitelném vyřčení konkrétního příkazu. Hlasitost povelu je také velmi důležitá, nesmí být příliš nízká ani vysoká, poté nedojde k rozpoznání, i když není přítomen jakýkoli šum. Z dosažených výsledků je však jednoznačné, že za přítomnosti okolního šumu klesá úspěšnost rozpoznání příkazu až o 71%. Závěrem tohoto experimentu lze říci, že aplikace pro hlasové ovládání chytré domácnosti je za některých podmínek velmi neefektivní a je jí nutno vyřešit vhodnou strukturou pro odstranění šumu.

Tab. 14: Průměrná úspěšnost rozpoznání pro všechny řečníky

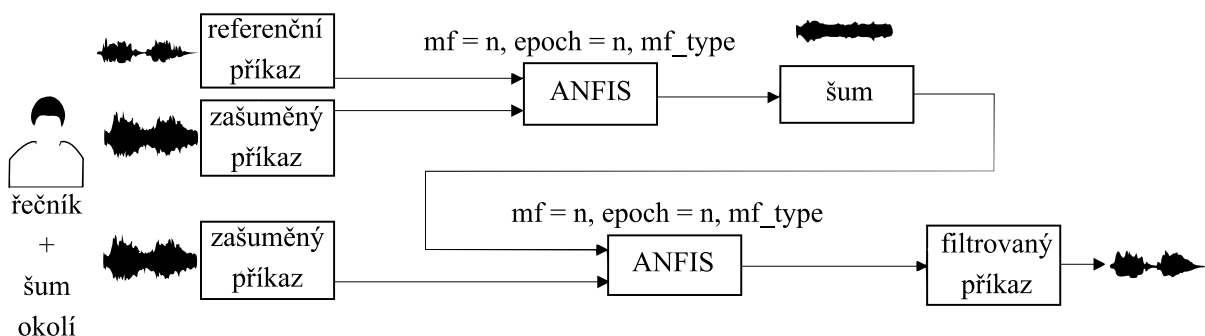
Mikrofon:	Trust All-round USB Microphone	
Řečník:	Všichni řečníci	
Příkaz:	"Light on"	
Typ signálu:	<i>zašuměný</i>	<i>zašuměný</i>
Typ šumu:	<i>mikrovlnná trouba</i>	<i>rychlovarná konvice</i>
Úspěšnost rozpoznání:	29,50 %	18,50 %
SSNR (dB)	7,215	8,2475
Příkaz:	"Light off"	
Typ signálu:	<i>zašuměný</i>	<i>zašuměný</i>
Typ šumu:	<i>mikrovlnná trouba</i>	<i>rychlovarná konvice</i>
Úspěšnost rozpoznání:	33,25 %	20,75 %
SSNR (dB)	8,035	7,9025
Příkaz:	"Blind up"	
Typ signálu:	<i>zašuměný</i>	<i>zašuměný</i>
Typ šumu:	<i>mikrovlnná trouba</i>	<i>rychlovarná konvice</i>
Úspěšnost rozpoznání:	33,75 %	34,75 %
SSNR (dB)	10,06	9,675
Příkaz:	"Blind down"	
Typ signálu:	<i>zašuměný</i>	<i>zašuměný</i>
Typ šumu:	<i>mikrovlnná trouba</i>	<i>rychlovarná konvice</i>
Úspěšnost rozpoznání:	31,25 %	27,25 %
SSNR (dB)	9,62	8,91

10 Navržení softwarové realizace adaptivního programu

Na základě rešerše byly vybrány pro potlačení šumu z konkrétního užitečného signálu neuronové sítě a konkrétně struktura ANFIS, která je pro tuto vhodná a je realizovatelná v prostředí softwaru MATLAB. Jako vstupy do této sítě za účelem úspěšného rozdělení signálu od šumu je nutné přivést dva signály o stejné délce vzorků. Těmito signály je referenční signál a signál s jakýmkoli aditivním šumem. Jako aditivní šumy byly v simulaci i v reálném testování použity jak uměle generované šumy (Gaussův šum, růžový a bílý šum), tak šumy zaznamenané z reálného prostředí (mikrofonná trouba, varná konvice, tekoucí voda, atd.) Hlasové příkazy, se kterými se v realizaci této diplomové práce pracuje, jsou vybrané na základě softwarové aplikace Hestia pro Windows 10, která je schopná ovládat strukturu KNX.

10.1 Návrh algoritmu

Na obrázku 10, lze vidět navrženou strukturu adaptivního systému pro potlačení šumu. Struktura se skládá ze dvou ANFIS vrstev, které nezávisle na sobě zpracovávají přijaté signály. Do každé ANFIS struktury jsou přivedeny dva signály. Jako první vstup pro první ANFIS strukturu je referenční příkaz bez šumu a jako druhý je zašuměný příkaz obsahující aditivní přičtený šum. Na základě přivedených signálů, je proveden algoritmus a je získán signál, obsahující konkrétní šum extrahovaný ze zašuměného příkazu. Tento signál je nezbytný pro další výpočet. Do druhé ANFIS struktury je opět přiveden stejný signál jako v předchozím případě a to ten obsahující šum. Dalším vstupem v této struktuře je extrahovaný šum. Referenční signál zde hraje důležitou roli, protože na jeho základě je možno potlačit jakýkoli okolní šum, struktura je tedy vysoce variabilní. Na konci celého výpočetního procesu, získáváme příkaz s potlačeným šumem, který by měl být dostatečně čistý na to, aby byl přijat aplikací Hestia pro hlasové ovládání KNX systému a provedl požadovaný proces. Tento signál by se měl co nejméně lišit od signálu referenčního. Aby výsledky procesu byly co nejlepší, je možné nastavovat parametry ANFIS struktury. Mezi tyto parametry patří počet tréninkových epoch (*epoch*), počet členských funkcí (*mf*) a typ celé struktury (*mf_type*). Epochy a členské funkce lze nastavovat až do několika desítek, již ale po přesáhnutí počtu 10, výpočet trvá přibližně 30 sekund, což je pro konkrétní aplikaci nevhodné, navíc jsou výsledky při tak vysokých hodnotách nastavení zkreslené (Tab. 15 - Tab. 17), proto jsou tyto parametry nastaveny na nejnížší možnou úroveň. Pro porovnání výsledků jsou použita různá hodnocení a to konkrétně RMSE, SSNR a PESQ.



Obr. 13: Blokové schéma navrženého systému

Následuje část kódu z MATLABu, ve kterém dochází k nastavení ANFIS struktury. Typ, počet epoch a počet členských funkcí jsou proměnlivé a volí se na základě vybraných hodnot v nastavení programu (Obr. 15 – ANFIS settings). Pracuje se se signály, které jsou vybrány na začátku, a které jsou následně také upraveny na stejnou délku. Poté dojde k procesu odlišení řečového signálu a šumu a následně k potlačení šumu v degradovaném signálu.

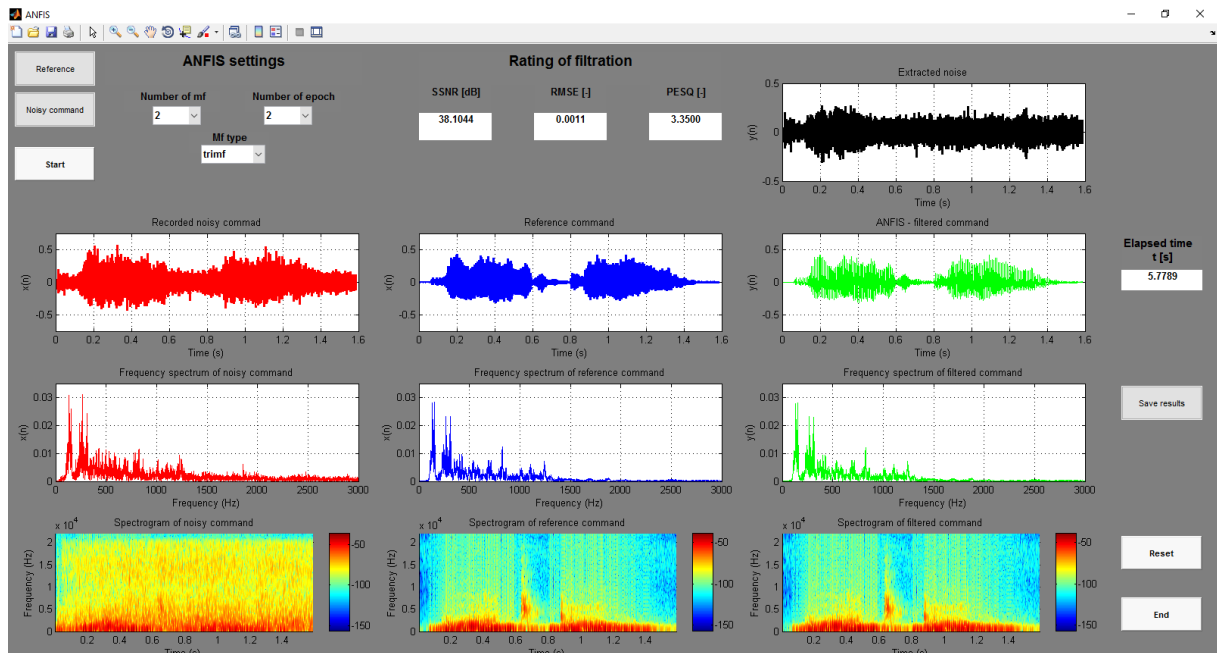
```
%% FIS struktura 1
mf_n = selectedIndex;
ss = 0.2;
num_epoch = selectedIndex1;
inmftype = selectedIndex2;
optMethod = 1

in_fismat = genfis1(trn_data,mf_n,inmftype);
out_fismat = anfis(trn_data, in_fismat,num_epoch);
estimated_n = evalfis(trn_data(:, 1:1), out_fismat);
e = noisyscomm - estimated_n;

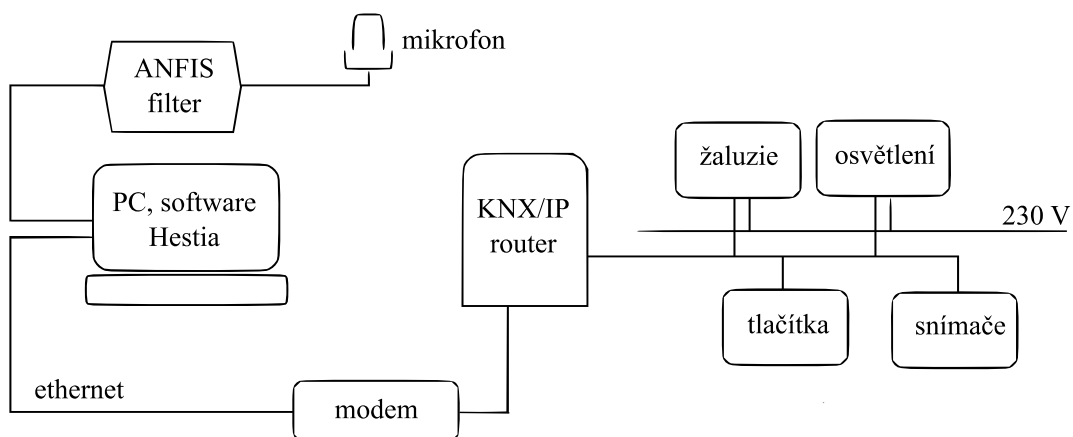
bb = e;
bbl = length(e);
cc = noisyscomm;
ccl = length(cc);
cas = (1:length(e))/Fs;

maxvychylka_e = max(e)
maxvychylka_noisyscomm = max(noisyscomm)
minvychylka_e = min(e)
minvychylka_noisyscomm = min(noisyscomm)

trn_data1 = [e noisyscomm];
```



Obr. 14: Ukázka grafického uživatelského rozhraní programu



Obr. 15: Blokové schéma zapojení celé funkční struktury

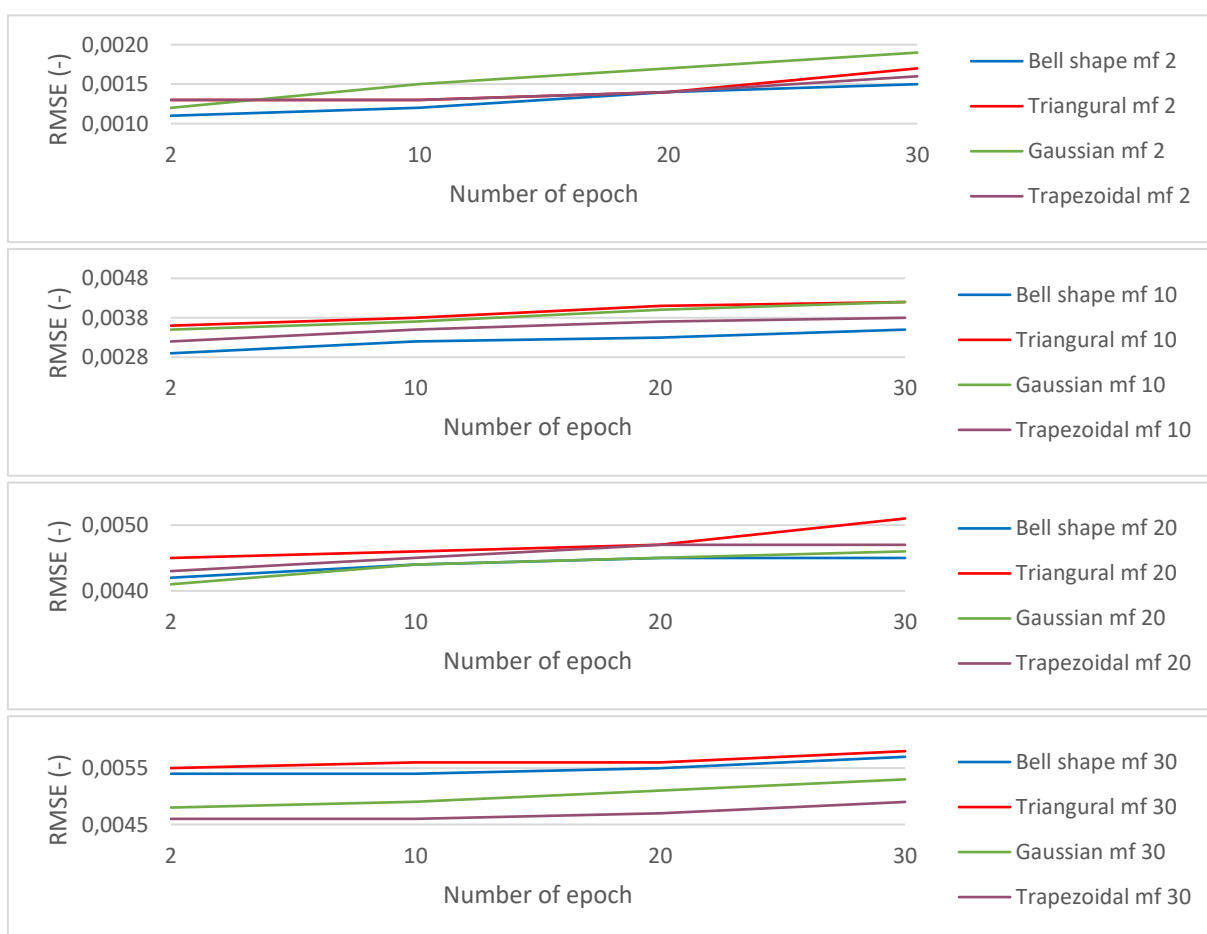
10.2 Simulace

Software pro potlačení šumu byl realizován v prostředí MATLAB na základě vestavěné funkce ANFIS a jak již bylo dříve zmíněno, skládá se ze dvou těchto struktur. Do celkové struktury je tedy zapotřebí přivést dva signály. Pro simulaci byly vybrány již testované čtyři signály pro ovládání světel a žaluzií „light on“, „light off“, „blind up“ a „blind down“. Tyto signály byly zaznamenány bez jakéhokoli šumu a byly tedy považovány za referenční. Jako degradované signály byly použity referenční s aditivními šumy jako uměle vygenerovaný Gaussův šum a zaznamenané šumy okolí sečteny do jednoho. Výsledky simulace pro příkaz „light on“ jsou uvedeny v následujících tabulkách (Tab. 15 - Tab. 17). Byly zde použity různé parametry nastavení a to počet epoch, počet členských funkcí a typy struktur. Výpočetní čas se významně lišil na základě nastavené struktury a to zejména podle počtu členských funkcí a epoch. Typ vybrané struktury výpočetní čas neovlivňoval, dále závisel spíše na výpočetním výkonu zařízení a na základě tohoto faktoru byl výpočetní čas při stejně nastavených parametrech pokaždé odlišný. Nejvíce potom výpočetní čas ovlivňoval nastavovaný parametr počtu členských funkcí (*Number of mf*).

Z výsledků simulace, lze zhodnotit nastavení celé sítě a určit, které parametry jsou pro tuto strukturu nejlepší. Pro nejlepší výsledky, je vhodné nastavit počet členských funkcí a počet epoch na co nejmenší hodnotu, tedy na dvě. U typu struktury se podle hodnotících parametrů jako nejlepší jeví typ „Triangular“ (*trimf*). Toto nejjednodušší nastavení má nejlepší výsledky v rámci hodnocení metodami RMSE, SSNR i PESQ. V poslední řadě je za nejjednodušších podmínek také nejkratší čas výpočtu, což je pro tento typ programu klíčová hodnota.

Tab. 15: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem RMSE pro příkaz "light on"

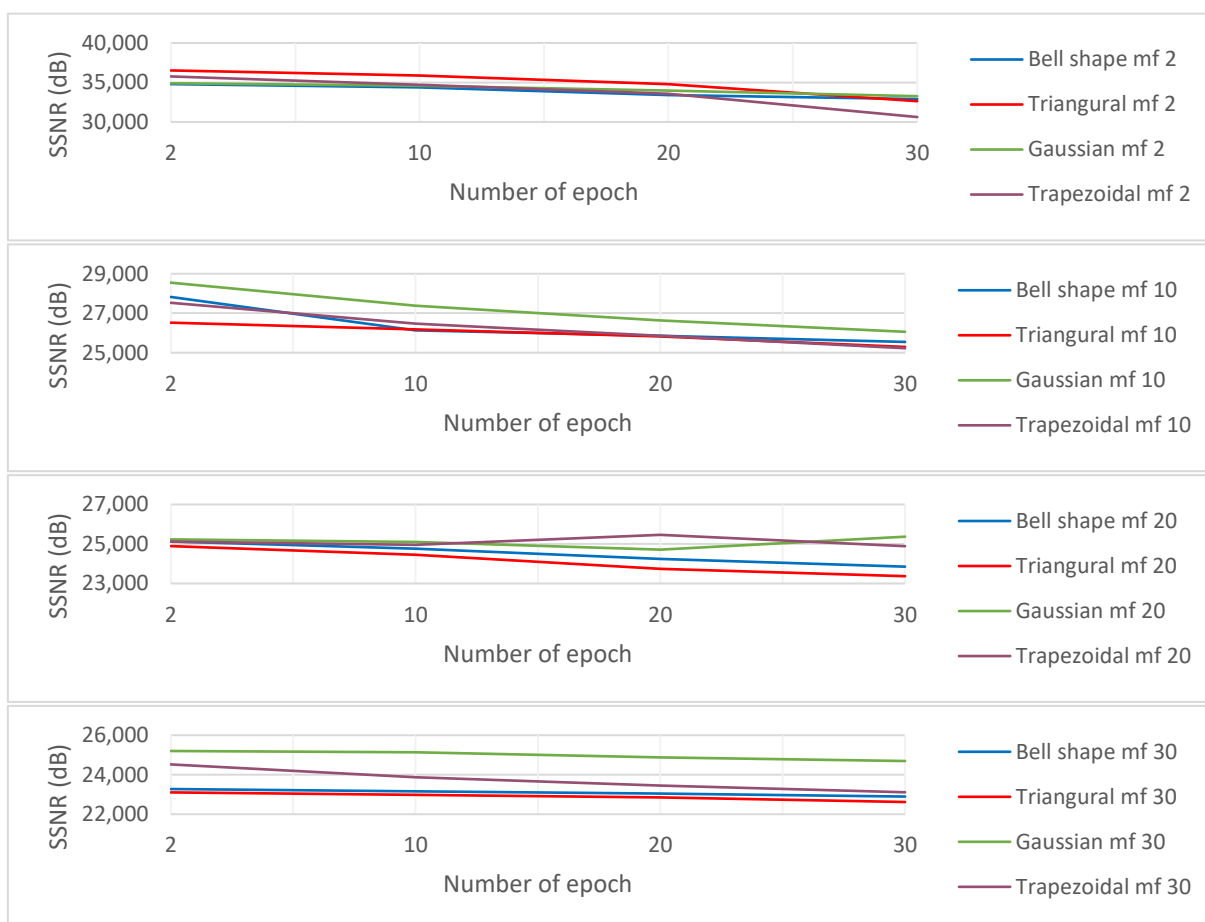
number of epoch	number of mf	elapsed time	RMSE (-)			
			Bell shape	Triangular	Gaussian	Trapezoidal
2	2	4,7	0,0011	0,0013	0,0012	0,0013
10	2	6,5	0,0012	0,0013	0,0015	0,0013
20	2	8,6	0,0014	0,0014	0,0017	0,0014
30	2	12,4	0,0015	0,0017	0,0019	0,0016
2	10	49,8	0,0029	0,0036	0,0035	0,0032
10	10	51,5	0,0032	0,0038	0,0037	0,0035
20	10	54,9	0,0033	0,0041	0,0040	0,0037
30	10	56,8	0,0035	0,0042	0,0042	0,0038
2	20	58,7	0,0042	0,0045	0,0041	0,0043
10	20	59,8	0,0044	0,0046	0,0044	0,0045
20	20	61,5	0,0045	0,0047	0,0045	0,0047
30	20	118,9	0,0045	0,0051	0,0046	0,0047
2	30	279,4	0,0054	0,0055	0,0048	0,0046
10	30	286,3	0,0054	0,0056	0,0049	0,0046
20	30	304,1	0,0055	0,0056	0,0051	0,0047
30	30	316,5	0,0057	0,0058	0,0053	0,0049



Obr. 16: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem RMSE pro příkaz "light on"

Tab. 16: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem SSNR "light on"

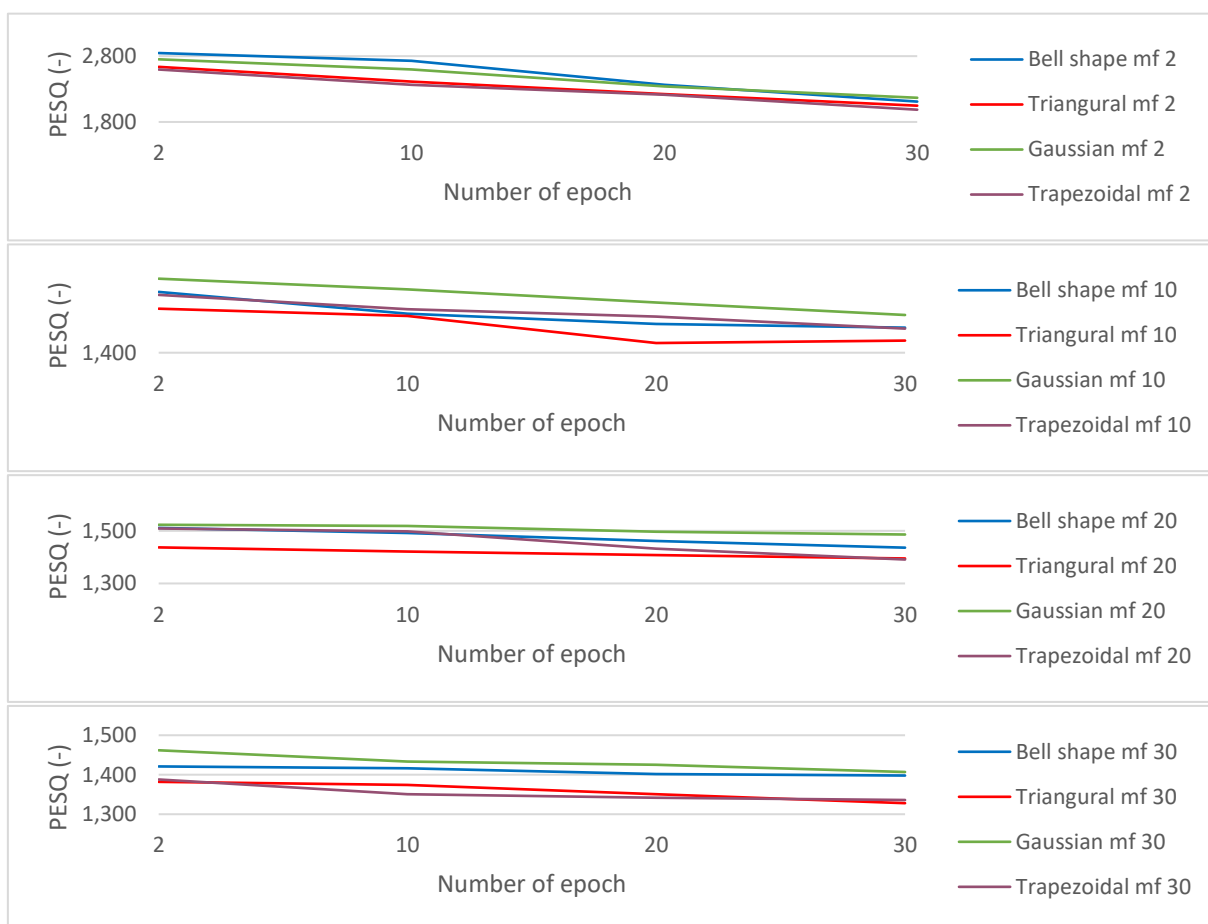
number of epoch	number of mf	elapsed time	SSNR (dB)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	4,7	34,792	36,523	34,896	35,764
10	2	6,5	34,357	35,873	34,567	34,693
20	2	8,6	33,419	34,784	33,966	33,547
30	2	12,4	32,897	32,631	33,258	30,619
2	10	49,8	27,824	26,522	28,547	27,528
10	10	51,5	26,134	26,186	27,375	26,473
20	10	54,9	25,863	25,830	26,641	25,861
30	10	56,8	25,546	25,291	26,059	25,227
2	20	58,7	25,129	24,892	25,238	25,118
10	20	59,8	24,756	24,447	25,095	24,955
20	20	61,5	24,238	23,745	24,714	25,463
30	20	118,9	23,851	23,369	25,369	24,889
2	30	279,4	23,276	23,105	25,203	24,523
10	30	286,3	23,161	22,987	25,142	23,869
20	30	304,1	23,052	22,845	24,879	23,452
30	30	316,5	22,894	22,619	24,695	23,112



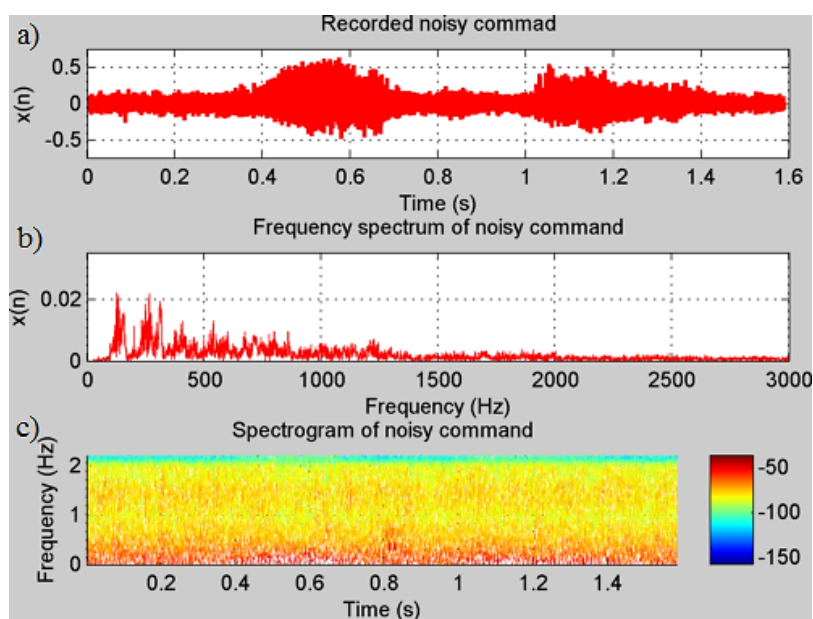
Obr. 17: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem SSNR pro příkaz "light on"

Tab. 17: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem PESQ "light on"

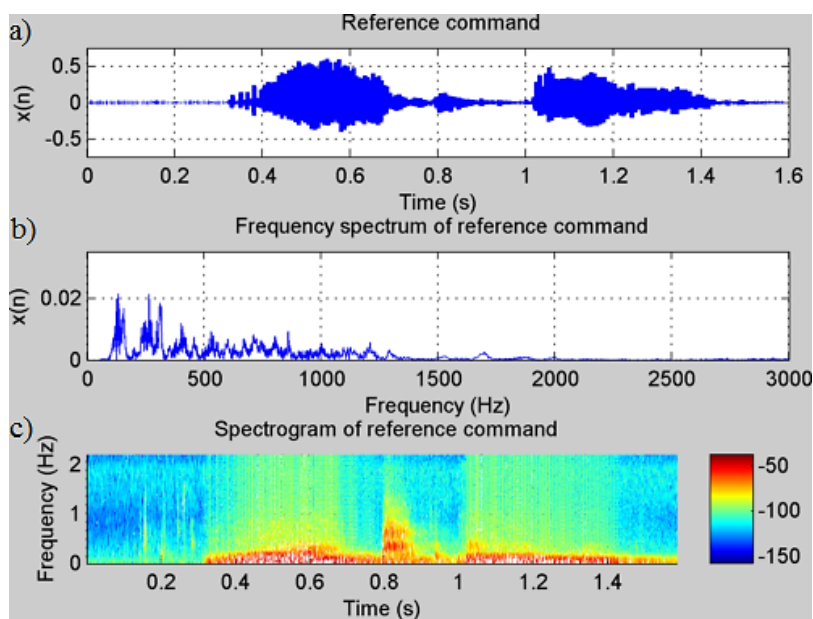
number of epoch	number of mf	elapsed time	PESQ (-)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	4,7	2,847	2,637	2,752	2,597
10	2	6,5	2,729	2,415	2,599	2,362
20	2	8,6	2,365	2,224	2,341	2,213
30	2	12,4	2,109	2,047	2,167	1,986
2	10	49,8	1,708	1,623	1,775	1,693
10	10	51,5	1,597	1,586	1,721	1,620
20	10	54,9	1,546	1,449	1,654	1,583
30	10	56,8	1,527	1,462	1,591	1,522
2	20	58,7	1,511	1,437	1,523	1,509
10	20	59,8	1,492	1,421	1,518	1,498
20	20	61,5	1,461	1,408	1,497	1,432
30	20	118,9	1,436	1,395	1,486	1,391
2	30	279,4	1,421	1,382	1,462	1,388
10	30	286,3	1,416	1,374	1,433	1,351
20	30	304,1	1,402	1,351	1,425	1,342
30	30	316,5	1,398	1,328	1,407	1,336



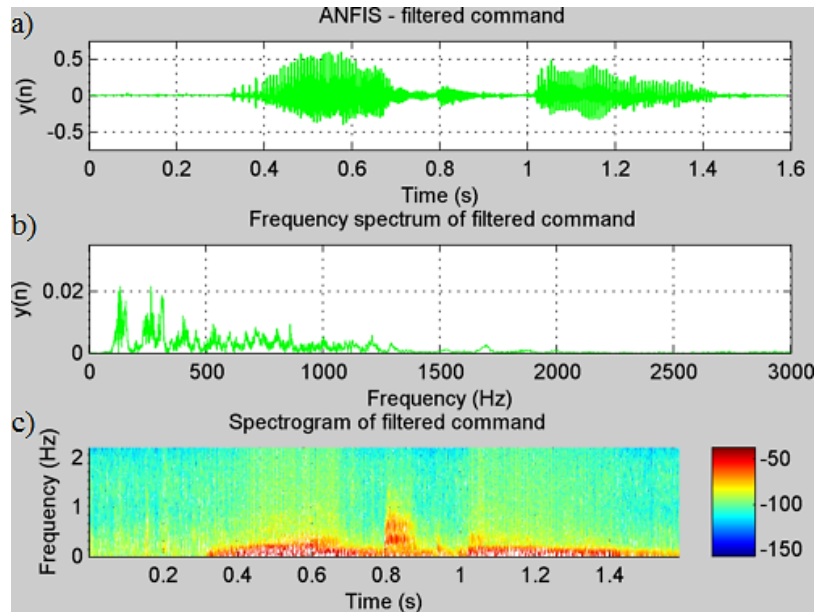
Obr. 18: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem PESQ pro příkaz "light on"



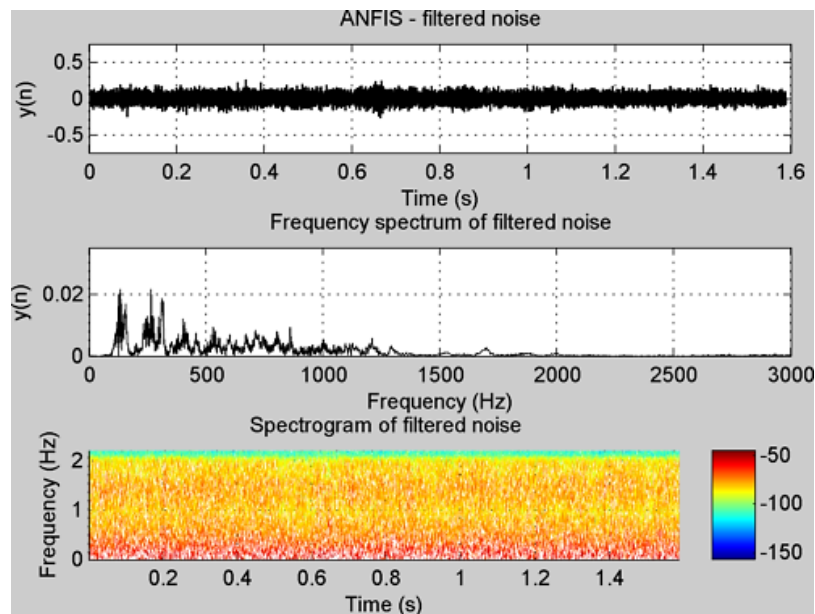
Obr. 19: Časové (a), frekvenční (b) a časově frekvenční (c) spektrum příkazu "light on" se šumem ($mf = 2$, $epoch = 2$, $type = "trimf"$)



Obr. 20: Časové (a), frekvenční (b) a časově frekvenční (c) spektrum referenčního příkazu "light on" ($mf = 2$, $epoch = 2$, $type = "trimf"$)



Obr. 21: Časové (a), frekvenční (b) a časově frekvenční (c) spektrum příkazu "light on" s potlačeným šumem ($mf = 2$, $epoch = 2$, $type = "trimf"$)



Obr. 22: Časové (a), frekvenční (b) a časově frekvenční (c) spektrum šumu z příkazu "light on" ($mf = 2$, $epoch = 2$,

Na předchozích obrázcích (Obr. 19 - Obr. 22) lze vidět srovnání všech signálů, se kterými zvolená struktura pracuje. Tyto signály jsou zde zobrazeny v časovém, frekvenčním a časově frekvenčním spektru. Jedná se o referenční signál a zašuměný signál, na základě kterých je proveden první výpočet, jehož výstupem je šum okolí, který je obsažen v zašuměném signálu. Když je k dispozici tento signál se šumem, je možné jej odstranit ze zašuměného signálu a získat příkaz, ideálně shodný se signálem referenčním, ale tak dobrých výsledků nelze nikdy dosáhnout.

10.3 Reálná implementace

Proběhlé simulace měly velmi dobré výsledky, proto je potřeba otestovat strukturu v praxi. ANFIS struktura pro potlačení šumu byla spuštěna na jiném počítači než aplikace Hestia a počítače byly propojené pomocí 3,5 mm jack konektoru, aby příkaz s potlačeným šumem mohl bez dalšího zkreslení okolím dorazit k rozpoznávací hlasu softwaru Hestia. Jako data k tomuto experimentu byly použity záznamy pro příkazy pro 1. řečníka z experimentu, který je popsán v kapitole 10. Referenční signály zůstaly stejné a jako aditivní šum byl použit součet několika reálných zaznamenaných šumů v jednom.

Následující tabulka (Tab. 18) zobrazuje statistické výsledky a porovnání referenčních, zašuměných a filtrovaných příkazů pro 1. řečníka. Podstatnými výsledky experimentu je procentuální rozpoznání příkazů před a po potlačení šumu.

Tab. 18: Srovnání referenčních, zašuměných a filtrovaných příkazů a jejich úspěšnost rozpoznání.

Mikrofon:	Trust All-round USB Microphone	
Řečník:	Řečník 1 - muž	
Příkaz:	"Light on"	
Typ signálu:	<i>zašuměný</i>	<i>filtrovaný</i>
Typ šumu:	<i>směs šumů</i>	X
Úspěšnost rozpoznání:	11 %	52 %
SSNR (dB)	2,15	27,90
Příkaz:	"Light off"	
Typ signálu:	<i>zašuměný</i>	<i>filtrovaný</i>
Typ šumu:	<i>směs šumů</i>	X
Úspěšnost rozpoznání:	9 %	49 %
SSNR (dB)	1,84	24,36
Příkaz:	"Blind up"	
Typ signálu:	<i>zašuměný</i>	<i>filtrovaný</i>
Typ šumu:	<i>směs šumů</i>	X
Úspěšnost rozpoznání:	14 %	69%
SSNR (dB)	2,95	29,81
Příkaz:	"Blind down"	
Typ signálu:	<i>zašuměný</i>	<i>filtrovaný</i>
Typ šumu:	<i>směs šumů</i>	X
Úspěšnost rozpoznání:	17 %	73%
SSNR (dB)	3,76	28,29

Všechny výpočty a experimenty byly doposud prováděny se záznamy, ke kterým byl přičten aditivní šum. Bylo nutné tedy provést další experiment, ve kterém bude šum zaznamenán současně s vyřčeným příkazem. V následujících tabulkách a grafech jsou uvedena všechna zjištěná data pro 1. řečníka. Jako zdroj rušení byla použita rychlovarná konvice vzdálená od mikrofonu stejně jako řečník a to 1m.

Na základě výsledků ze simulace zůstalo nastavení následující:

- *Number of mf* = „2“
- *Number of epoch* = „2“
- *Mf type* = „trimf“

Tab. 19: Výsledky procesu v reálném prostředí

Mikrofon:	Trust All-round USB Microphone	
Řečník:	Řečník 1 - muž	
Příkaz:	"Light on"	
Typ signálu:	zašuměný	filtrovaný
Typ šumu:	reálný	X
SSNR (dB)	-3,1428	-0,0329
RMSE (-)	0,1277	0,0893
PESQ (-)	1,1590	1,2610
Příkaz:	"Light off"	
Typ signálu:	zašuměný	filtrovaný
Typ šumu:	reálný	X
SSNR (dB)	-4,4603	-0,1183
RMSE (-)	0,0956	0,0580
PESQ (-)	1,0820	1,2050
Příkaz:	"Blind up"	
Typ signálu:	zašuměný	filtrovaný
Typ šumu:	reálný	X
SSNR (dB)	-3,9977	0,0361
RMSE (-)	0,0887	0,0558
PESQ (-)	1,1450	1,3602
Příkaz:	"Blind down"	
Typ signálu:	zašuměný	filtrovaný
Typ šumu:	reálný	X
SSNR (dB)	-2,9317	-0,0319
RMSE (-)	0,1284	0,0919
PESQ (-)	1,0780	1,2490

11 Diskuze

Hlasové ovládání systému KNX se po praktickém testování v laboratoři a ve výzkumném a inovačním centru osvědčilo jako velmi efektivní a komfortní. Za přítomnosti šumu byl však tento systém velmi neefektivní, proto bylo rozhodnuto o realizaci adaptivního softwaru, který by byl schopen tento šum z řečového signálu potlačit.

Program byl realizován v programovém prostředí MATLAB. Jeho struktura je založena na adaptivním neuro-fuzzy inferenčním systému ANFIS. Byl navržen tak, aby bylo možné celou jeho podstatnou strukturu před každým procesem potlačení šumu nastavit. Na tomto nastavení však následně závisí celý výpočetní proces. Vzhledem k využití tohoto programu je vhodné, aby byl proces výpočtu co nejkratší a zároveň bylo dosaženo dobrých výsledků. Jak již bylo v této práci zmíněno, nastavit bylo možné počet členských funkcí (*mf*), počet tréninkových epoch (epoch) a typ struktury (*mf_type*), kde je možné výběr z následujících možností: Zvonovitá „*gbellmf*“, Trojúhelníková „*trimf*“, Gaussova „*gaussmf*“, nebo trapézová „*trapmf*“. Na výběru těchto parametrů velmi záleží z hlediska výpočetního času, který se navyšuje podle složitosti zvolené struktury.

Na základě simulačních experimentů, které byly provedeny pro jednotlivé hlasové příkazy pro ovládání světel („*light on*“, „*light off*“) a žaluzií („*blind up*“, „*blind down*“) lze vyvodit jistý závěr o tom, které nastavení je pro tuto adaptivní strukturu vhodné.

Výsledky ze simulace v tabulce s hodnotícím parametrem RMSE pro příkaz „*light on*“ (Tab. 15) jsou následující: Nejlepší hodnota procesu potlačení šumu je pro nastavení „*mf* = 2“, „epoch = 2“ a typ struktury „*gbellmf*“ a to konkrétně 0,0011. Pro další parametr SSNR (Tab. 16) je nejlepší hodnota 36,523 dB a té bylo dosaženo se stejným nastavením členských funkcí a epoch ale s typem struktury „*trimf*“. Pro poslední hodnotící parametr tohoto příkazu PESQ (Tab. 17), jde o hodnotu 2,847, kde nastavení členských funkcí a epoch bylo opět stejné („*mf* = 2“, „epoch = 2“), přičemž typ struktury byl „*gbellmf*“.

Pro příkaz „*light off*“ byly nejlepší výsledky u všech hodnotících parametrů zaznamenány opět v nastavení „*mf* = 2“, „epoch = 2“ a ke změnám docházelo pouze při změně struktury. Pro hodnocení RMSE (Tab. II. 2) to byla struktura „*gbellmf*“ a „*gaussmf*“, kde byla hodnota výpočtu stejná a to 0,0014. Pro parametr SSNR (Tab. II. 3) byla nejlepší vypočtená hodnota 33,214 dB, kde byla nastavena struktura „*trimf*“. Stejná struktura („*trimf*“) byla také u nejlepšího výsledku pro parametr PESQ (Tab. II. 4), kde šlo o hodnotu 3,652.

Nejlepší hodnoty u hlasového příkazu „*blind up*“ byly dosažena opět při shodném nastavení členských funkcí a epoch („*mf* = 2“, „epoch = 2“). Výsledek pro RMSE (Tab. II. 5) byl 0,0008 a to při struktuře „*trimf*“. Nejlepší výsledek u segmentálního odstupu signálu od šumu (Tab. II. 6) byl 36,700 dB a to opět při struktuře „*trimf*“. Pro poslední hodnotící parametr PESQ (Tab. II. 7) vyšla nejlépe hodnota 3,485, které bylo dosaženo na základě struktury „*trapmf*“.

Při testování posledního příkazu „*blind down*“ bylo za nejlepšími výsledky stejné nastavení členských funkcí a epoch („*mf* = 2“, „*epoch* = 2“). Struktury s nejlepšími výsledky u hodnocení RMSE (Tab. II. 8) byly dvě a to „*trimf*“ a „*gaussmf*“ s výsledky 0,0012. Pro SSNR (Tab. II. 9) byl nejlepší výsledek pozorován u struktury „*trimf*“ a to 38,106 dB. Nejlepší hodnota hodnocení PESQ (Tab. II. 10) byla 3,503 a zvolená struktura pro tuto hodnotu byla „*trimf*“.

Výsledky těchto experimentů jsou také vykresleny graficky pro signál „*light on*“ (Obr. 16 - Obr. 18), „*light off*“ (Obr. II. 1 - Obr. II. 3), „*blind up*“ (Obr. II. 4 - Obr. II. 6) a „*blind down*“ (Obr. II. 7 - Obr. II. 9). Při této simulaci, při které byly použity reálné záznamy příkazů s přičtenými různými typy šumů, je jednoznačně nejlepší pro budoucí procesy zvolit nejjednodušší strukturu „*mf* = 2“, „*epoch* = 2“. Jako nejvhodnější typ struktury se podle výsledků z tohoto experimentu jeví trojúhelníková „*trimf*“, která byla nejčtenější strukturou s nejlepšími výsledky („*trimf*“ – 8x, „*gbellmf*“ – 3x, „*gaussmf*“ – 2x, „*trapmf*“ – 1x). Při tomto nastavení bylo dosaženo statisticky nejlepších výsledků. Jednotlivé signály, které byly zpracované při tomto nastavení pro všechny čtyři příkazy lze pozorovat na vykreslených grafech. Pro příkaz „*light on*“ se jedná o zobrazení zašuměného signálu v časovém, frekvenčním a časově frekvenčním spektru (Obr. 19). Pro stejný příkaz je zde zobrazen také referenční signál (Obr. 20), signál s potlačeným šumem (Obr. 21) a také konkrétní šum ze signálu (Obr. 22). Pro ostatní příkazy („*light off*“, „*blind up*“ a „*blind down*“) jsou všechny tato spektra vykreslena také (Příloha).

Toto nastavení a signály byly použity v experimentu, kdy byla v provozu celá funkční struktura a výsledky rozpoznání byly oproti zašuměným signálům velmi dobré (Tab. 18). Problém však nastal při testování v reálných podmínkách, kdy byl v místnosti se záznamovým zařízením generován reálný šum, stejně vzdálený od mikrofonu jako řečník. Za těchto podmínek byl zaznamenaný signál mnohem více znehodnocen, a i když došlo po potlačení šumu podle hodnotících parametrů ke zlepšení, nebylo to dostačující k tomu, aby byl řečový signál softwarem Hestia úspěšně rozpoznán. Na základě těchto výsledků byl proveden experiment s náhodně upravenou ANFIS strukturou, kdy byly parametry nastavení zvoleny vyšší než 2, výsledky však byly podle očekávání a zkušeností ze simulace vždy ještě horší.

Závěr

V této práci bylo nutné provést rešerši na téma využití Soft Computingových metod pro zpracování řečového signálu. Na základě článků zabývajících se touto problematikou byly vybrány neuronové sítě, aby mohly být použity jako adaptivní metoda pro potlačení šumu v řečovém signálu. Po nastudování obecné problematiky neuronových sítí, byla zvolena vhodná struktura, založená jak na neuronových sítích tak také na fuzzy logice. Vybraná adaptivní struktura se nazývá adaptivní neuro-fuzzy inferenční systém, neboli ANFIS.

V teoretické části byly popsány všechny náležitosti, které se týkají zadané problematiky. Byl zde přiblížen pojem „Soft Computing“ a vše co je s tímto pojmem spojeno. Dále byly v práci popsány neuronové sítě, které byly zvoleny jako základ pro vytvoření struktury adaptivního softwaru určeného pro potlačení šumu.

Další samostatná kapitola se zabývala řečí za přítomnosti šumu. V záznamu obsahující řečový signál lze pozorovat různé typy šumů. Tyto šумы byly matematicky popsány a u jednotlivých typů byly uvedeny příklady. Dalším tématem v této kapitole bylo měření odstupů řečového signálu od šumu, které se liší od klasického SNR.

Dále bylo nutné zhodnotit výsledky proběhlého procesu potlačení u jednotlivých řečových signálů. Byla zde proto vysvětlena a matematicky popsána norma ITU-T P.863, která je standardem poskytujícím objektivní metodu měření kvality řečového signálu. Podle této normy byl vytvořen algoritmus PESQ, pomocí kterého lze na základě referenčního a zašuměného signálu určit hodnotu MOS, podle které je možno určit kvalitu řečového signálu, u kterého došlo k potlačení šumu a porovnat jej s původní hodnotou. Další metodou byl segmentální odstup signálu od šumu neboli SSNR, které je dané průměrem lokálních SNR v segmentech s řečovou aktivitou. Jako poslední hodnotící parametr byla vybrána metoda RMSE, neboli střední kvadratická chyba, která slouží k posouzení rozdílu mezi referenčním a degradovaným signálem.

Praktická část této práce se skládala z několika částí. Nejdříve bylo nutné zprovoznit aplikaci Hestia pro hlasové ovládání KNX systému. Důležité bylo správné nastavení skupinových adres, aby byla zřízena komunikace mezi softwarem a sběrnici. Funkčnost aplikace byla demonstrována na simulační sestavě KNX v laboratoři EB312 na fakultě elektrotechniky a informatiky a také v budově výzkumného a inovačního centra, v areálu stavební fakulty Vysoké školy báňské vybavené systémem KNX. Bylo potřeba také zjistit statistickou úspěšnost rozpoznání jednotlivých příkazů určených k ovládání světel a žaluzií. Byl tedy realizován experiment, kterého se zúčastnili čtyři lidé, z toho dva muži a dvě ženy. Výsledek experimentu jasně stanovil, že za přítomnosti okolního šumu je aplikace prakticky nepoužitelná a je tedy nutné tento problém řešit zařízením na jeho potlačení.

V následující části byl realizován návrh adaptivní struktury ANFIS, kde byla popsána její funkčnost a zapojení do celého systému hlasového ovládání KNX. Program byl vytvořen v softwarovém prostředí MATLAB, kde bylo vytvořeno také grafické uživatelské rozhraní. Následně bylo nutné provést testování a to jak simulaci, tak implementaci v reálném prostředí. Díky rozsáhlé simulaci, při které byla struktura testována s různým nastavením, byly získány výsledky, podle kterých bylo možné určit optimální parametry této struktury. Simulace proběhla pro jednotlivé příkazy „light on“, „light off“,

„blind up“ a „blind down“ a s různými typy šumů, které byly přiřčeny k jednotlivým příkazům pomocí softwaru Audacity. Nastavení struktury bylo pro všechny tyto příkazy shodné a výstupy byly porovnávány podle zmíněných hodnotících parametrů. Nejjednodušší nastavení ANFIS struktury přineslo paradoxně ty nejlepší výsledky potlačení šumu. Toto nastavení bylo použito také při testování v reálném prostředí. Pro ovládání prvků KNX byly použity záznamy ze simulace. Signály znehodnoceny šumem byly přijaty hlasovým rozpoznávačem jen velmi zřídka. ANFIS struktura však byla schopna na základě referenčního signálu rozpoznat šum, který poté dokázala potlačit v zašuměném příkazu. Příkaz s potlačeným šumem byl mnohem zřetelnější a u aplikace Hestia se zvýšila úspěšnost rozpoznání. Následovalo testování s reálným šumem, kde už byly výsledky horší. Hlasový příkaz byl zaznamenán zároveň s reálným šumem okolí, konkrétně byl použit zvuk mikrovlnné trouby a rychlovarné konvice jako v simulaci. Takto zaznamenaný signál představoval pro vytvořenou strukturu větší problém. U signálu, zaznamenaného v reálných podmínkách byly znatelně horší parametry hodnocení než při simulaci. Poté, co proběhl proces potlačení šumu, se výsledky hodnotících parametrů prokazatelně zlepšily, nebyly však tak dobré, aby byla aplikace pro hlasové ovládání schopna příkaz úspěšně rozpoznat. Lze tedy říci, že použití adaptivní Soft Computingové metody ANFIS při potlačení šumu z řečového signálu bylo prokazatelně úspěšné, ne však natolik aby byl tento software použitelný v zašuměném prostředí, kde je použito k ovládání inteligentní budovy softwaru Hestia. Přínosem této práce je tedy zjištění, že zde je možnost použití Soft Computingových metod, konkrétně adaptivního neuro-fuzzy inferenčního systému v praxi, jako médium pro potlačení šumu z řečového signálu. K tomu účelu lze využít software vytvořený v programovém prostředí MATLAB, který byl vytvořen na základě zadání této diplomové práce.

Literatura

- [1] AGARWALLA, S. AND K. K. SARMA Machine learning based sample extraction for automatic speech recognition using dialectal Assamese speech. *Neural Networks*, 6// 2016, 78, 97-111.
- [2] ASANO, F., Y. MOTOMURA, S. NAKAMURA, H. ASOH, et al. Fusion of audio and video information for detecting speech events. In *Proceedings of the 6th International Conference on Information Fusion, FUSION 2003*. 2003, vol. 1, p. 386-393.
- [3] CZYŻEWSKI, A. AND R. KRÓLIKOWSKI Neuro-rough control of masking thresholds for audio signal enhancement. *Neurocomputing*, 2// 2001, 36(1-4), 5-27.
- [4] DU, K. L. AND M. N. S. SWAMY *Neural networks in a softcomputing framework*. Edition ed., 2006. 1-566 p.
- [5] GENARO, N., A. TORIJA, A. RAMOS, I. REQUENA, et al. Modeling environmental noise using artificial neural networks. In *ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications*. 2009a, p. 215-219.
- [6] GIL-PITA, R., L. CUADRA, E. ALEXANDRE, D. AYLLÓN, et al. Enhancing the energy efficiency of wireless-communicated binaural hearing aids for speech separation driven by soft-computing algorithms. *Applied Soft Computing*, 7// 2012, 12(7), 1939-1949.
- [7] KASABOV, N. K., R. KOZMA AND M. J. WATTS Phoneme-based speech recognition via fuzzy neural networks modeling and learning. *Information Sciences*, 9// 1998, 110(1-2), 61-79.
- [8] MALCANGI, M. AND P. GREW. Evolving fuzzy-neural method for multimodal speech recognition. In *Communications in Computer and Information Science*. 2015b, vol. 517, p. 216-227.
- [9] MALCANGI, M., K. OUZZANE AND P. PATEL. Audio-visual fuzzy fusion for robust speech recognition. In *Proceedings of the International Joint Conference on Neural Networks*. 2013.
- [10] BHARDWAJ, S., S. SRIVASTAVA AND J. R. P. GUPTA Pattern-similarity-based model for time series prediction. *Computational Intelligence*, 2015, 31(1), 106-131.
- [11] MALCANGI, M. Employing smart logic to spot audio in real time on deeply embedded systems. 2011. IFIP Advances in Information and Communication Technology.
- [12] TEODORESCU, H. N. Fuzzy logic in speech technology – introductory and overviewing glimpses. 2015. *Studies in Fuzziness and Soft Computing*.
- [13] ANISSA, S., S. HASSENE, M. ZOUHAIR AND IEEE Efficient Speech De-noising Applied to Colored Noise Based Dynamic Low-pass Filter Supervised by Cascade Neural Networks. 2013 International Conference on Electrical Engineering and Software Applications (Iceesa), 2013, 659-663.
- [14] BEHERA, L., B. SUNDARAM AND IEEE Stochastic filtering and speech enhancement using a recurrent quantum neural network. *Proceedings of International Conference on Intelligent Sensing and Information Processing*, 2004, 165-170.

- [15] CHERIFI, E., M. GUERTI AND IEEE Speech Signal Filtering by Artificial Neural Networks. 2015 4th International Conference on Electrical Engineering (Icee), 2015, 306-+.
- [16] DAQROUQ, K., I. N. ABU-ISBEIH, M. ALFAUORI AND IEEE Speech Signal Enhancement Using Neural Network and Wavelet Transform. 2009 6th International Multi-Conference on Systems, Signals and Devices, Vols 1 and 2, 2009, 826-831.
- [17] DEHYADEGARY, L., S. A. SEYYEDSALEHI AND I. NEJADGHOLI Nonlinear enhancement of noisy speech, using continuous attractor dynamics formed in recurrent neural networks. *Neurocomputing*, Oct 2011, 74(17), 2716-2724.
- [18] ERDOGAN, H., J. R. HERSHEY, S. WATANABE, J. LE ROUX, et al. PHASE-SENSITIVE AND RECOGNITION-BOOSTED SPEECH SEPARATION USING DEEP RECURRENT NEURAL NETWORKS. In *40th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Brisbane, AUSTRALIA: Ieee, 2014, p. 708-712.
- [19] FAH, L. B., A. HUSSAIN, S. A. SAMAD AND IEEE. Speech enhancement by noise cancellation using neural network. In *10th International IEEE Tencon Conference*. Kuala Lumpur, Malaysia: Ieee, 2000, p. 39-42.
- [20] GAO, T., J. DU, L. XU, C. LIU, et al. A UNIFIED SPEAKER-DEPENDENT SPEECH SEPARATION AND ENHANCEMENT SYSTEM BASED ON DEEP NEURAL NETWORKS. 2015 Ieee China Summit & International Conference on Signal and Information Processing, 2015a, 687-691.
- [21] GHOLAMI-BOROUJENY, S., A. FALLATAH, B. P. HEFFERNAN AND H. R. DAJANI Neural network-based adaptive noise cancellation for enhancement of speech auditory brainstem responses. *Signal Image and Video Processing*, Feb 2016, 10(2), 389-395.
- [22] GIRIN, L., J. L. SCHWARTZ AND G. FENG Audio-visual enhancement of speech in noise. *Journal of the Acoustical Society of America*, Jun 2001, 109(6), 3007-3020.
- [23] KANG, T. G., K. KWON, J. W. SHIN AND N. S. KIM NMF-based Target Source Separation Using Deep Neural Network. *Ieee Signal Processing Letters*, Feb 2015, 22(2), 229-233.
- [24] KNECHT, W. G., M. E. SCHENKEL AND G. S. MOSCHYTZ NEURAL-NETWORK FILTERS FOR SPEECH ENHANCEMENT. *Ieee Transactions on Speech and Audio Processing*, Nov 1995, 3(6), 433-438.
- [25] LAI, Y. H., C. H. CHEN, S. T. TANG, Z. M. YEH, et al. Improving the Performance of Noise Reduction in Hearing Aids Based on the Genetic Algorithm. In *14th Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON)*. Paphos, CYPRUS: Springer, 2016, vol. 57, p. 209-214.
- [26] LIN, C. T., R. C. WU AND G. D. WU Noisy speech segmentation/enhancement with multiband analysis and neural fuzzy networks. *International Journal of Pattern Recognition and Artificial Intelligence*, Nov 2002, 16(7), 927-955.
- [27] MARUNGSRI, B. AND A. OONSIVILAI Fuzzy ARTMAP technique for speech noise reduction. *Lecture Notes in Signal Science, Internet and Education (Ssip'07/Miv'07/Diweb'07)*, 2007, 20-25.

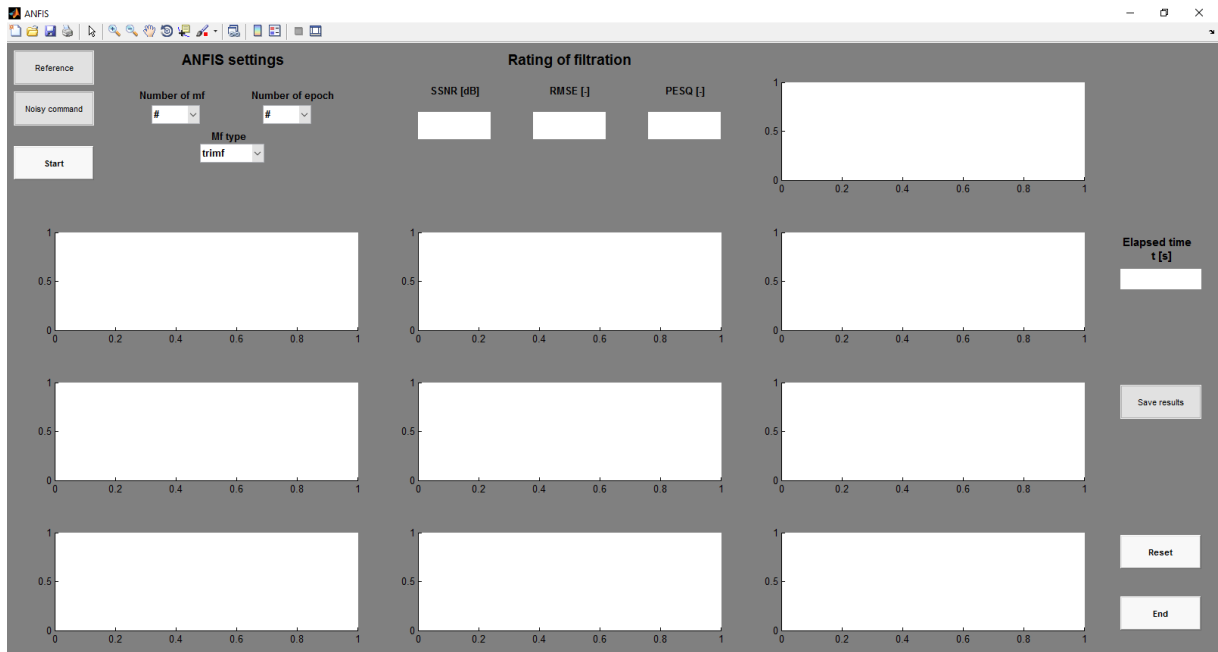
- [28] PAL, S., Z. SARIC, S. OCOVAJ, I. PAPP, et al. Hands free VoIP terminal with gain control based on neural network. 2012 Ieee 19th International Conference and Workshops on Engineering of Computer Based Systems (Ecbs), 2012, 173-181.
- [29] ROHMAN, B. P. A., C. B. WAEL, K. PARAMAYUDHA AND IEEE Adaptively Compensated Multiband Spectral Subtraction for Robust Noise Reduction. 2015 International Symposium on Intelligent Signal Processing and Communication Systems (Ispacs), 2015a, 139-142.
- [30] SHARIFI-TEHRANI, O. AND E. GHAFARIOUN Design and Implementation of Dual-Microphone Adaptive-Active Noise Cancellation System for De-Noising Speech Signal. *Przeglad Elektrotechniczny*, 2012, 88(2), 258-261.
- [31] WAN, E. A., A. T. NELSON AND IEEE. Removal of noise from speech using the Dual EKF algorithm. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 98)*. Seattle, Wa: Ieee, 1998, p. 381-384.
- [32] XIA, Y. S. AND J. WANG Low-dimensional recurrent neural network-based Kalman filter for speech enhancement. *Neural Networks*, Jul 2015, 67, 131-139.
- [33] XU, Y., J. DU, L. R. DAI AND C. H. LEE An Experimental Study on Speech Enhancement Based on Deep Neural Networks. *Ieee Signal Processing Letters*, Jan 2014, 21(1), 65-68.
- [34] HÁJEK, P. *Soft Computing*, 2000. Available from Internet:< <http://casopis.vesmir.cz/clanek/soft-computing> >.
- [35] BERKA, P. 2016. Available from Internet:<<http://sorry.vse.cz/~berka/docs/4iz430/P07-Soft%20Computing.pdf>>.
- [37] UHLÍŘ, J. *Technologie hlasových komunikací: revue littéraire mensuelle*. Praha: Nakladatelství ČVUT, 2007. ISBN 978-80-01-03888-8.
- [38] MathWorks. In. United States, 1994-2016, p. Explore products for MATLAB , the language of technical computing, and Simulink , for simulation and Model-Based Design.
- [39] MERTZ, H., T. H. HANSEMANN AND C. HÜBNER *Automatizované systémy budov*. Translated by I.V. BARTOŠ. Edtion ed. Munich: Carl Hanser Verlag, 2007. 264 p. Translation of: *Automatizované systémy budov*. ISBN 978-80-247-2367-9.
- [40] ITU-T P.863, Perceptual objective listening quality assessment. In.: International Telecommunication Union, 2014, p. 82.
- [41] CHAI, T. AND R. R. DREXLER. Root mean square error (RMSE) 2014.
- [42] BELEŠOVÁ, J. *Tvorba prostředí pro ovládání provozně technických funkcí v Smart Home Care*. Vysoká škola báňská - Technická univerzita Ostrava., 2016.
- [43] ZADEH, L., B. BOUCHON-MEUNIER AND R. YAGER *Fuzzy Logic and Soft Computing*. Edtion ed.: World Scientific Publishing Co. Pte. Ltd., 1995. ISBN 981-02-2345-5.

Seznam příloh

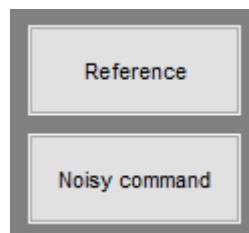
<u>Příloha A</u>	Příručka pro použití vytvořeného systému	Chyba! Záložka není definována.
<u>Příloha B</u>	Parametry mikrofonu, výsledky filtrace.....	Chyba! Záložka není definována.
<u>Příloha C</u>	Grafy z procesů potlačení šumu pro zbylé příkazy	Chyba! Záložka není definována.
<u>Příloha D</u>	Obsah CD.....	XXVI

Příloha A

Grafické uživatelské rozhraní aplikace je navrženo tak, aby bylo intuitivní a přehledné. Po spuštění programu, se otevře okno, které obsahuje šest tlačítek, deset grafických oken pro vykreslení konkrétních příkazů, nastavení struktury a hodnocení proběhlého procesu potlačení šumu.

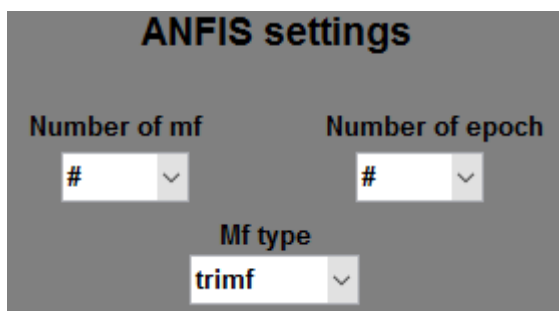


Obr. I. 1: Grafické uživatelské rozhraní programu po spuštění



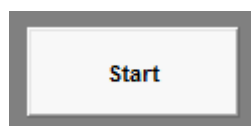
Obr. I. 2: Tlačítka pro načtení příkazů

Na předchozím obrázku (Obr. I. 2) je vidět tlačítka sloužící k načtení referenčního signálu „Reference“ a zašuměného signálu „Noisy command“. Po vybrání potřebných signálů je nutné nastavit parametry ANFIS struktury.



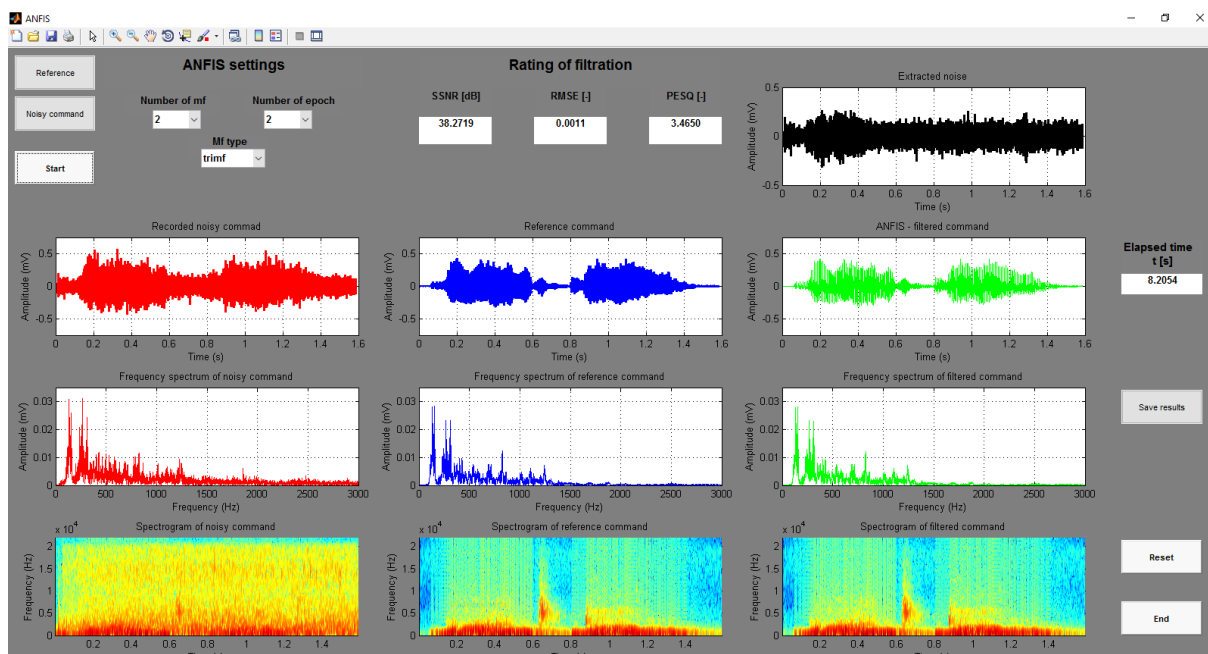
Obr. 1. 3: Nastavení parametrů struktury

Nastavení aplikace se skládá ze třech atributů a to z počtu členských funkcí „*Number of mf*“, z počtu epoch „*Number of epoch*“ a z typu struktury. Struktura systému ANFIS může být buď Zvonovitá „*gbellmf*“, Trojúhelníková „*trimf*“, Gaussova „*gaussmf*“, nebo trapézová „*trapmf*“. Počty členských funkcí a epoch lze nastavit na tyto hodnoty: 2, 3, 4, 5, 6, 8, 10 a 15. Po nastavení všech těchto parametrů je struktura připravena k přijetí signálů a k potlačení šumu.



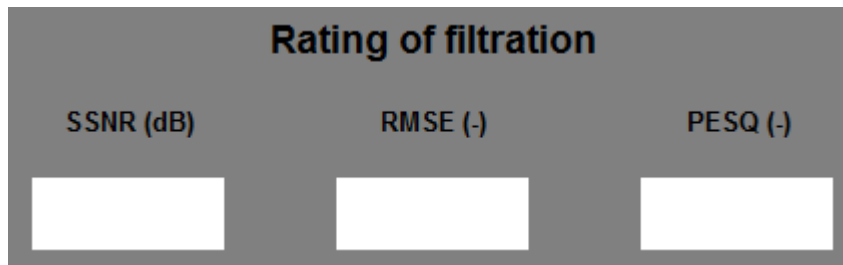
Obr. 1. 4: Spouštěcí tlačítko procesu

Proces potlačení následně spustíme pomocí tlačítka „*Start*“, které se nachází hned pod tlačítky sloužící pro načtení signálů.



Obr. 1. 5: GUI po procesu potlačení šumu

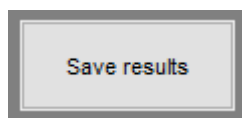
Předchozí obrázek (Obr. I. 5) zobrazuje proces potlačení, který proběhl na základě vybraných signálů (referenčního a zašuměného) a nastavené ANFIS struktury. Jednotlivé grafy je možno přímo v oknech libovolně přibližovat, či oddalovat a jsme také schopni všechny signály porovnat mezi sebou z hlediska časového, frekvenčního a časově frekvenčního.



The image shows a dark gray rectangular panel titled "Rating of filtration". Below the title, there are three white rectangular input fields arranged horizontally. Above each field is a label: "SSNR (dB)" for the first, "RMSE (-)" for the second, and "PESQ (-)" for the third.

Obr. I. 6: Hodnocení výsledného signálu

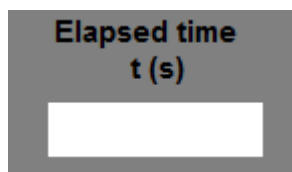
Poté, co dojde k procesu potlačení šumu, je potřeba zhodnotit výsledný signál. K tomu dochází v části okna, kde jsou k dispozici 3 hodnotící parametry a to SSNR, RMSE a PESQ.



A rectangular button with a gray border and a lighter gray background. The text "Save results" is centered on the button.

Obr. I. 7: Tlačítko pro uložení výsledků

Pokud chceme přejít k dalšímu procesu, lze využít tlačítka „Save results“, pomocí kterého uložíme výsledky posledního zpracovávaného signálu.



A dark gray rectangular panel. At the top, it says "Elapsed time" in bold, followed by "t (s)" in a smaller font. Below this text is a white rectangular input field.

Obr. I. 8: Zobrazení proběhlého času

Posledním parametrem, které se týká procesu potlačení šumu mimo hodnotící metody, je čas potřebný na konkrétní výpočty. Délka procesu závisí na složitosti zvolené ANFIS struktury.



A vertical stack of two rectangular buttons with gray borders and light gray backgrounds. The top button is labeled "Reset" and the bottom button is labeled "End".

Obr. I. 9: Tlačítka pro vymazání současných výsledků a ukončení program

Jako posledními prvky programu jsou dvě tlačítka. Tlačítko „*Reset*“ slouží k vymazání signálu posledního proběhlého procesu ze všech grafických oken a z paměti načtených signálů. Aplikaci lze ukončit klasicky v pravém horním rohu ale také pomocí tlačítka „*End*“.

Příloha B

Tato příloha se skládá z informací o technických parametrech mikrofonu **Trust All-round USB Microphone**, který byl použit na experimenty.

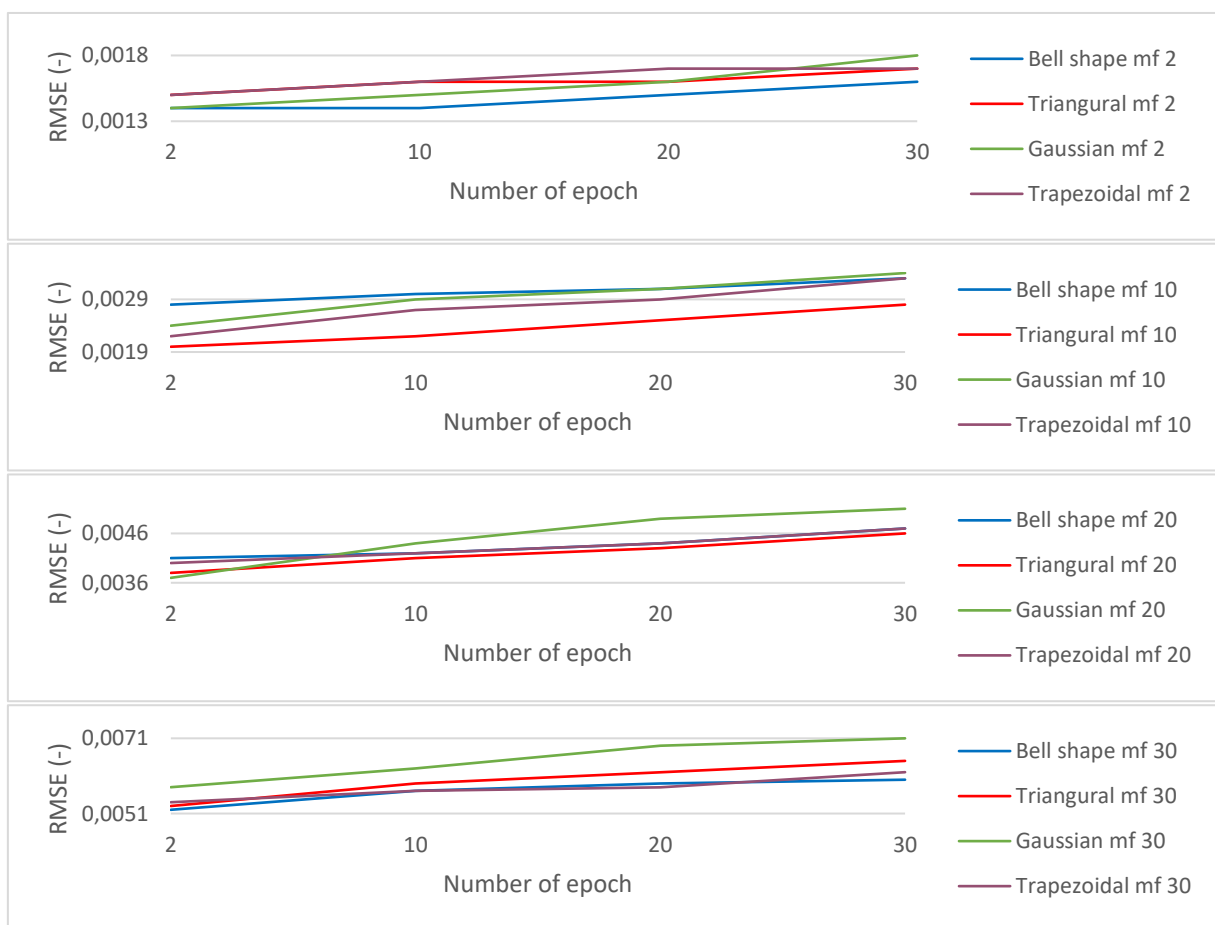
Tab. II. 1: Technické parametry použitého mikrofonu

Název	Trust All-round USB Microphone
Typ	kondenzátorový
Konektor	3.5mm + USB
Rozsah frekvence	50Hz-16kHz
Impedance	2.2kOhm
Směr	omni
Délka kabelu	1.8m
Stojan	tripod
Příslušenství	USB adapter
Velikost	130x43x35mm
Váha	150g

Následující tabulky obsahují výsledky hodnocení simulace pro příkazy s aditivními šumy a to konkrétně pro příkazy „*light off*“, „*blind up*“ a „*blind down*“.

Tab. II. 2: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem RMSE pro příkaz "light off"

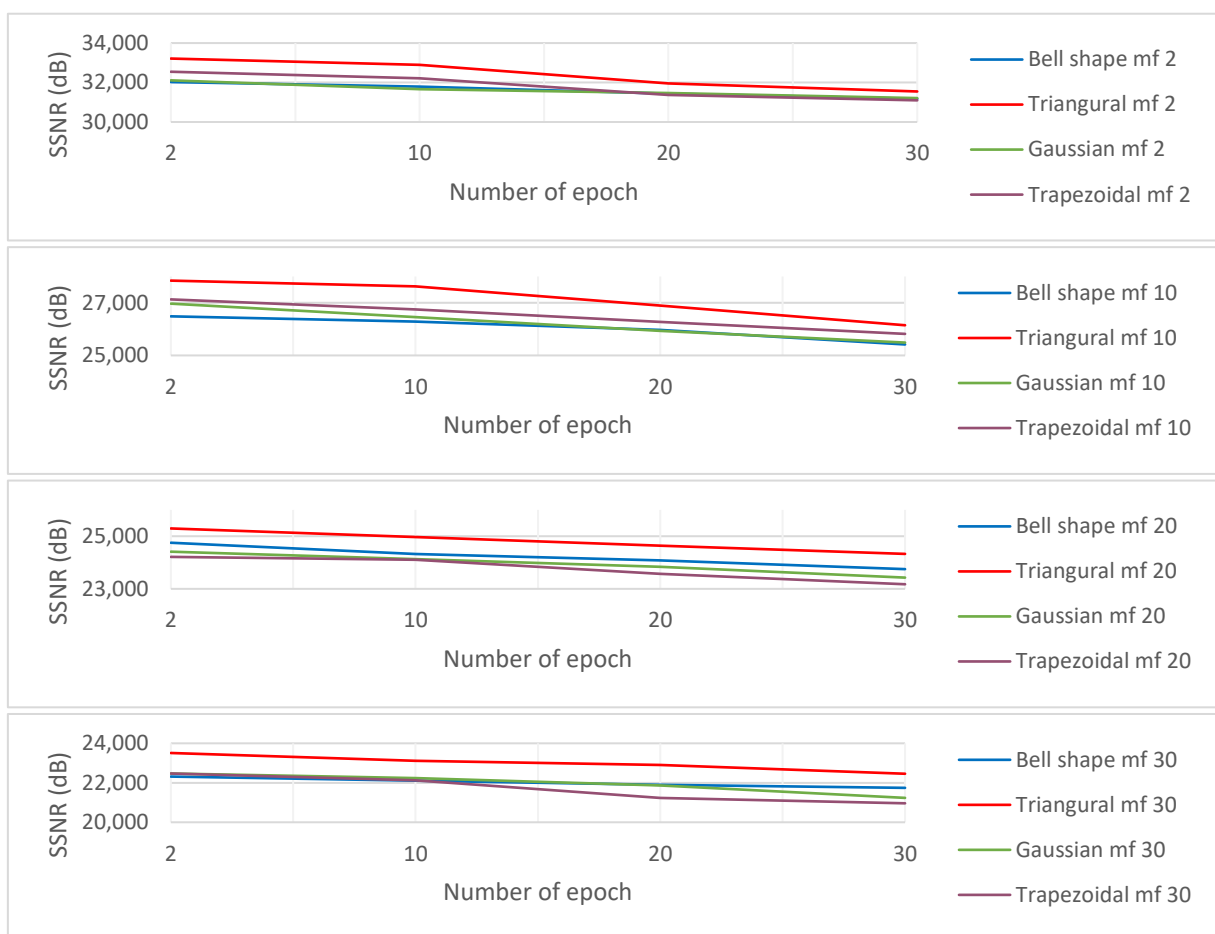
number of epoch	number of mf	elapsed time	RMSE (-)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	5,1	0,0014	0,0015	0,0014	0,0015
10	2	7,0	0,0015	0,0016	0,0015	0,0016
20	2	9,2	0,0015	0,0016	0,0016	0,0017
30	2	12,9	0,0016	0,0017	0,0018	0,0017
2	10	51,7	0,0028	0,0020	0,0024	0,0022
10	10	52,3	0,0030	0,0022	0,0029	0,0027
20	10	56,8	0,0031	0,0025	0,0031	0,0029
30	10	58,4	0,0033	0,0028	0,0034	0,0033
2	20	59,6	0,0041	0,0038	0,0037	0,0040
10	20	61,2	0,0042	0,0041	0,0044	0,0042
20	20	63,8	0,0044	0,0043	0,0049	0,0044
30	20	121,5	0,0047	0,0046	0,0051	0,0047
2	30	285,7	0,0052	0,0053	0,0058	0,0054
10	30	289,1	0,0057	0,0059	0,0063	0,0057
20	30	308,6	0,0059	0,0062	0,0069	0,0058
30	30	321,2	0,0060	0,0065	0,0071	0,0062



Obr. II. 1: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem RMSE pro příkaz "light off"

Tab. II. 3: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem SSNR pro příkaz "light off"

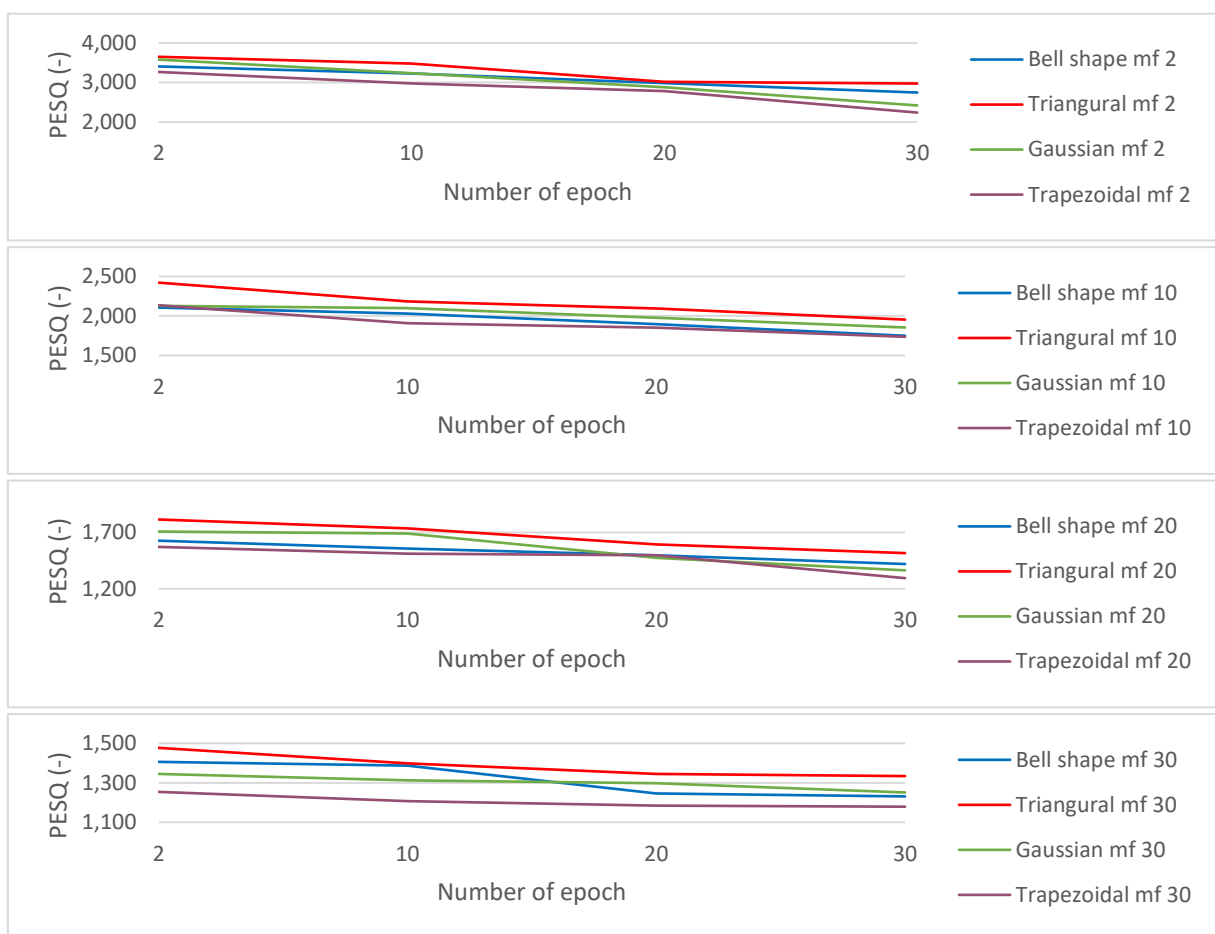
number of epoch	number of mf	elapsed time	SSNR (dB)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	5,1	32,025	33,214	32,110	32,547
10	2	7,0	31,787	32,896	31,655	32,213
20	2	9,2	31,439	31,947	31,472	31,376
30	2	12,9	31,174	31,546	31,212	31,098
2	10	51,7	26,485	27,843	26,968	27,129
10	10	52,3	26,286	27,624	26,452	26,743
20	10	56,8	25,973	26,892	25,931	26,268
30	10	58,4	25,417	26,147	25,488	25,818
2	20	59,6	24,752	25,295	24,412	24,215
10	20	61,2	24,319	24,971	24,133	24,102
20	20	63,8	24,083	24,639	23,843	23,564
30	20	121,5	23,749	24,328	23,427	23,175
2	30	285,7	22,316	23,510	22,471	22,471
10	30	289,1	22,109	23,107	22,239	22,120
20	30	308,6	21,893	22,896	21,858	21,237
30	30	321,2	21,744	22,459	21,236	20,962



Obr. II. 2: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem SSNR pro příkaz "light off"

Tab. II. 4: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem PESQ pro příkaz "light off"

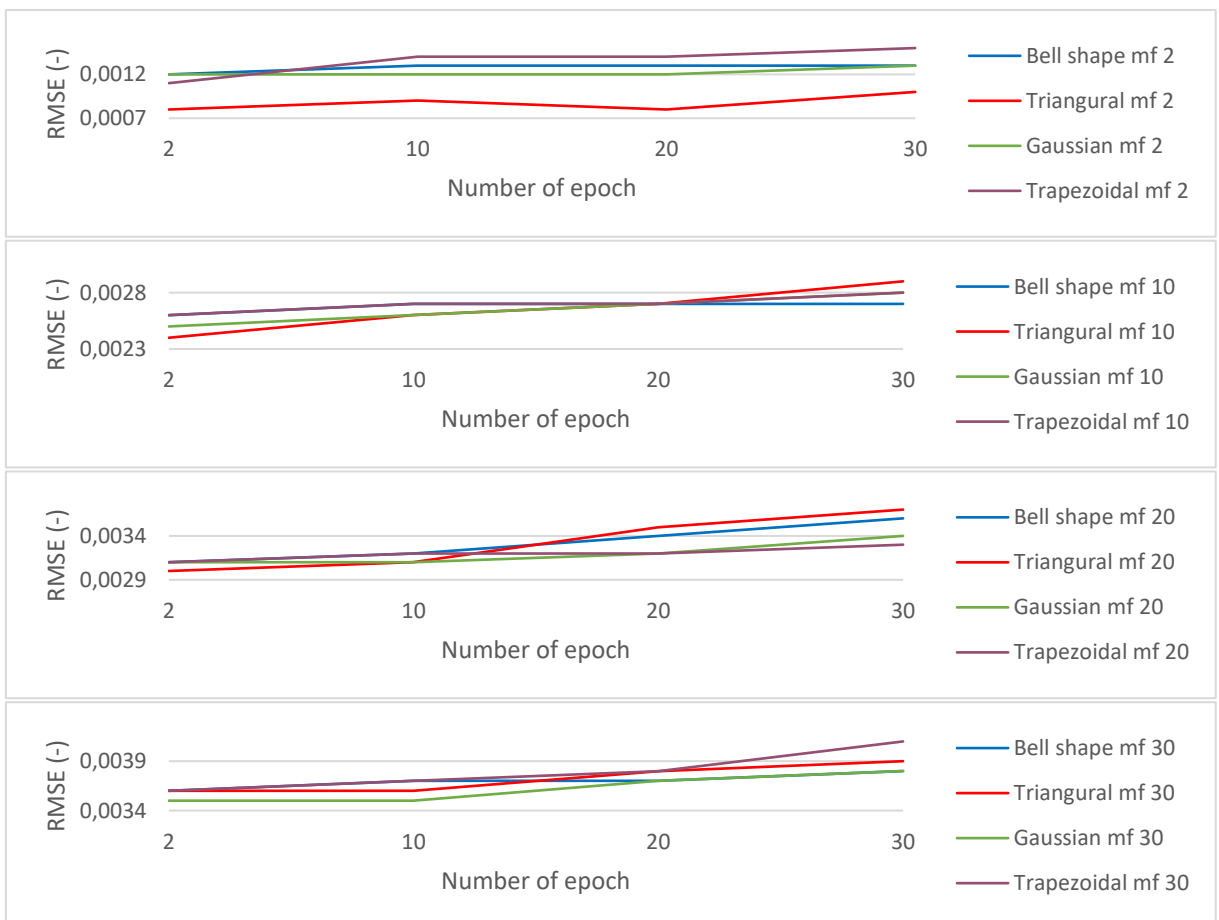
number of epoch	number of mf	elapsed time	PESQ (-)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	5,1	3,408	3,652	3,581	3,267
10	2	7,0	3,231	3,481	3,238	2,980
20	2	9,2	2,983	3,014	2,878	2,785
30	2	12,9	2,746	2,975	2,419	2,239
2	10	51,7	2,106	2,422	2,128	2,135
10	10	52,3	2,031	2,186	2,098	1,908
20	10	56,8	1,895	2,093	1,976	1,851
30	10	58,4	1,749	1,954	1,854	1,736
2	20	59,6	1,627	1,815	1,708	1,572
10	20	61,2	1,558	1,736	1,690	1,512
20	20	63,8	1,497	1,593	1,476	1,497
30	20	121,5	1,420	1,517	1,364	1,295
2	30	285,7	1,406	1,477	1,345	1,254
10	30	289,1	1,387	1,398	1,312	1,207
20	30	308,6	1,246	1,345	1,298	1,184
30	30	321,2	1,231	1,334	1,251	1,179



Obr. II. 3: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem PESQ pro příkaz "light off"

Tab. II. 5: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem RMSE pro příkaz "blind up"

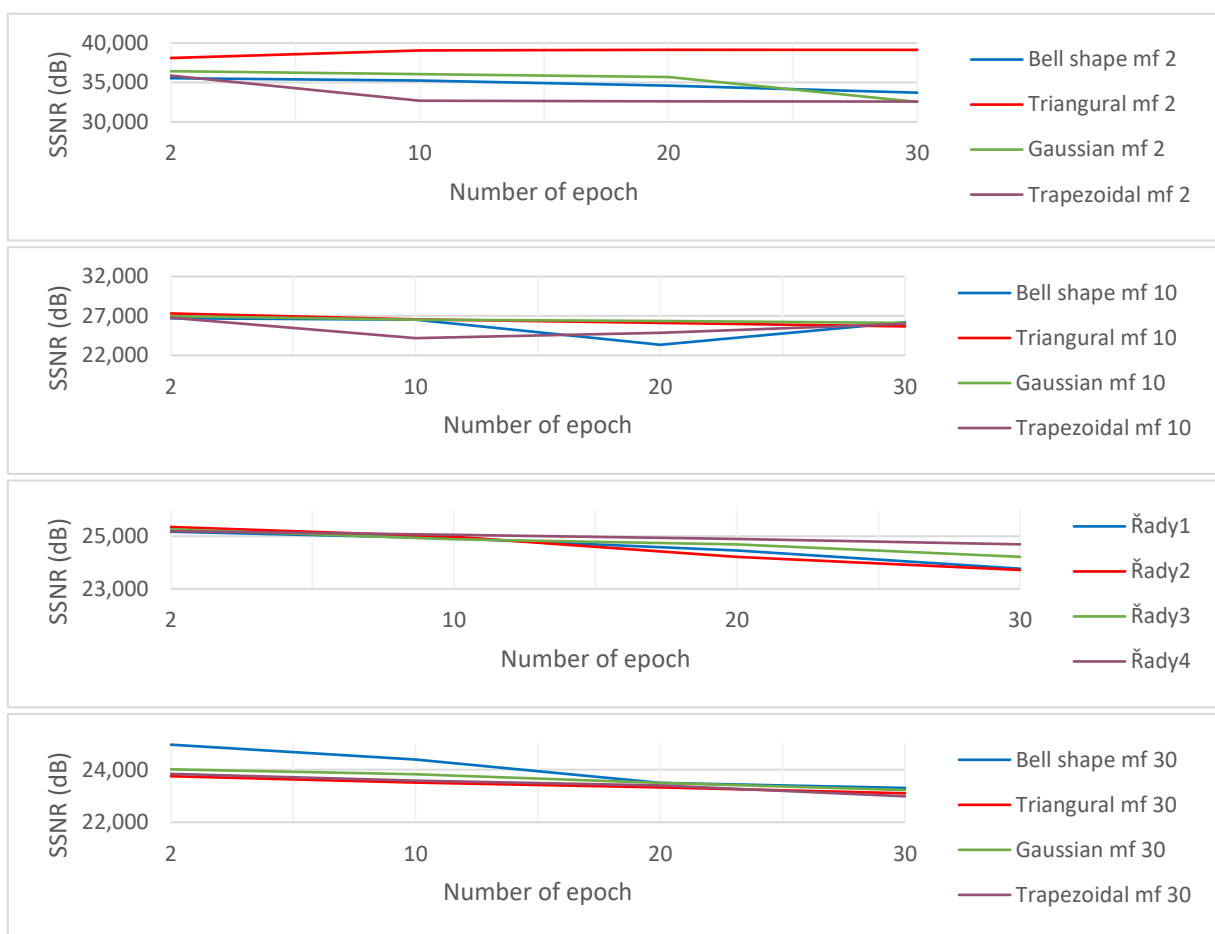
number of epoch	number of mf	elapsed time	RMSE (-)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	4,1	0,0012	0,0008	0,0012	0,0011
10	2	5,9	0,0013	0,0009	0,0012	0,0014
20	2	7,7	0,0013	0,0008	0,0012	0,0014
30	2	11,6	0,0013	0,0010	0,0013	0,0015
2	10	48,7	0,0026	0,0024	0,0025	0,0026
10	10	49,3	0,0027	0,0026	0,0026	0,0027
20	10	53,7	0,0027	0,0027	0,0027	0,0027
30	10	54,6	0,0027	0,0029	0,0028	0,0028
2	20	56,9	0,0031	0,0030	0,0031	0,0031
10	20	57,8	0,0032	0,0031	0,0031	0,0032
20	20	59,2	0,0034	0,0035	0,0032	0,0032
30	20	115,1	0,0036	0,0037	0,0034	0,0033
2	30	263,6	0,0036	0,0036	0,0035	0,0036
10	30	279,5	0,0037	0,0036	0,0035	0,0037
20	30	296,3	0,0037	0,0038	0,0037	0,0038
30	30	308,7	0,0038	0,0039	0,0038	0,0041



Obr. II. 4: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem RMSE pro příkaz "blind up"

Tab. II. 6: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem SSNR pro příkaz "blind up"

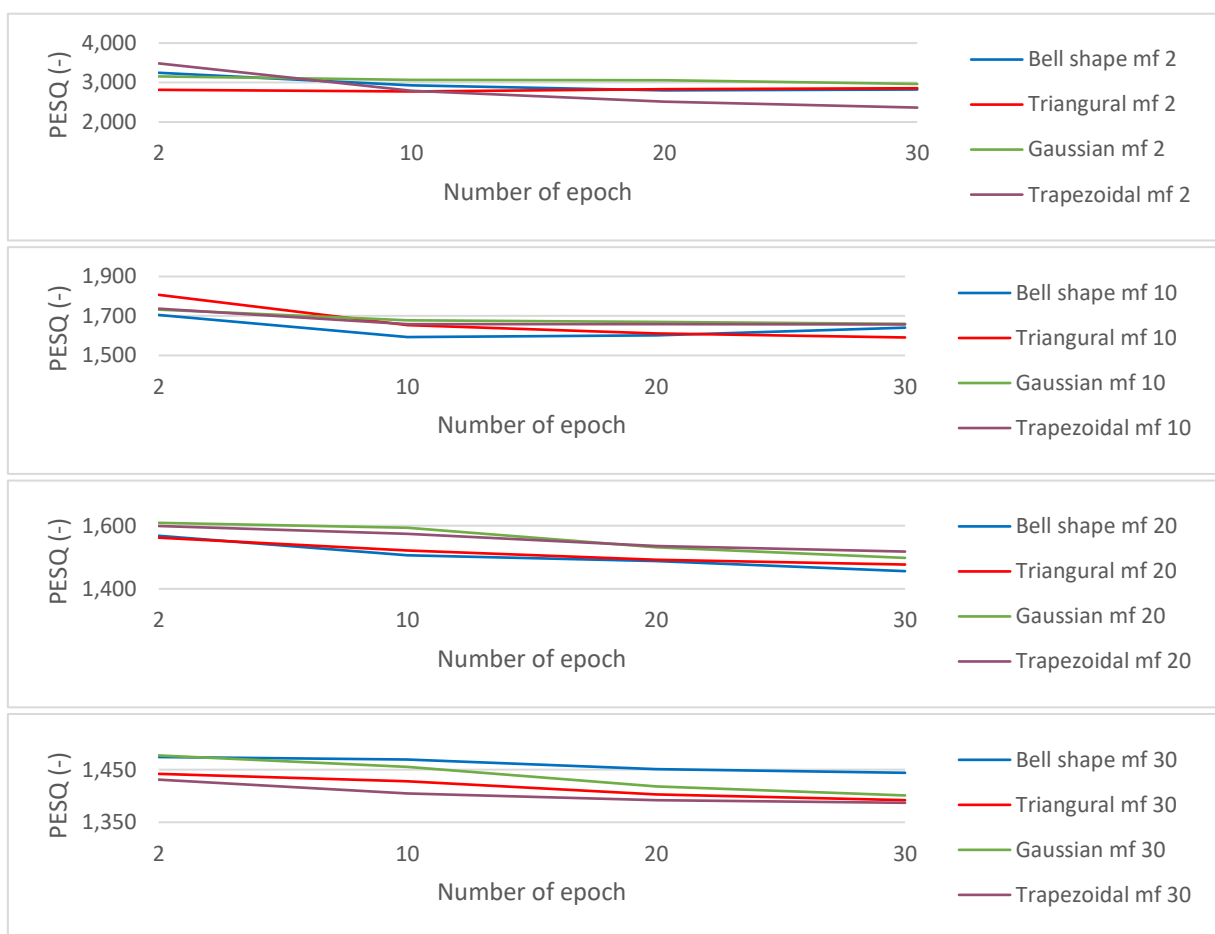
number of epoch	number of mf	elapsed time	SSNR (dB)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	4,1	33,251	36,700	33,607	33,921
10	2	5,9	32,993	35,697	33,177	32,290
20	2	7,7	32,907	36,121	33,037	31,965
30	2	11,6	32,892	34,950	32,994	31,720
2	10	48,7	26,706	27,315	26,974	26,774
10	10	49,3	26,528	26,577	26,513	24,192
20	10	53,7	23,349	26,128	26,385	24,863
30	10	54,6	26,194	25,679	26,121	26,012
2	20	56,9	25,171	25,351	25,265	25,192
10	20	57,8	25,918	24,986	24,878	25,056
20	20	59,2	24,457	24,213	24,695	24,893
30	20	115,1	23,767	23,717	24,217	24,691
2	30	263,6	24,949	23,751	24,011	23,839
10	30	279,5	24,384	23,509	23,825	23,577
20	30	296,3	23,502	23,328	23,495	23,396
30	30	308,7	23,304	23,101	23,225	22,987



Obr. II. 5: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem SSNR pro příkaz "blind up"

Tab. II. 7: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem PESQ pro příkaz "blind up"

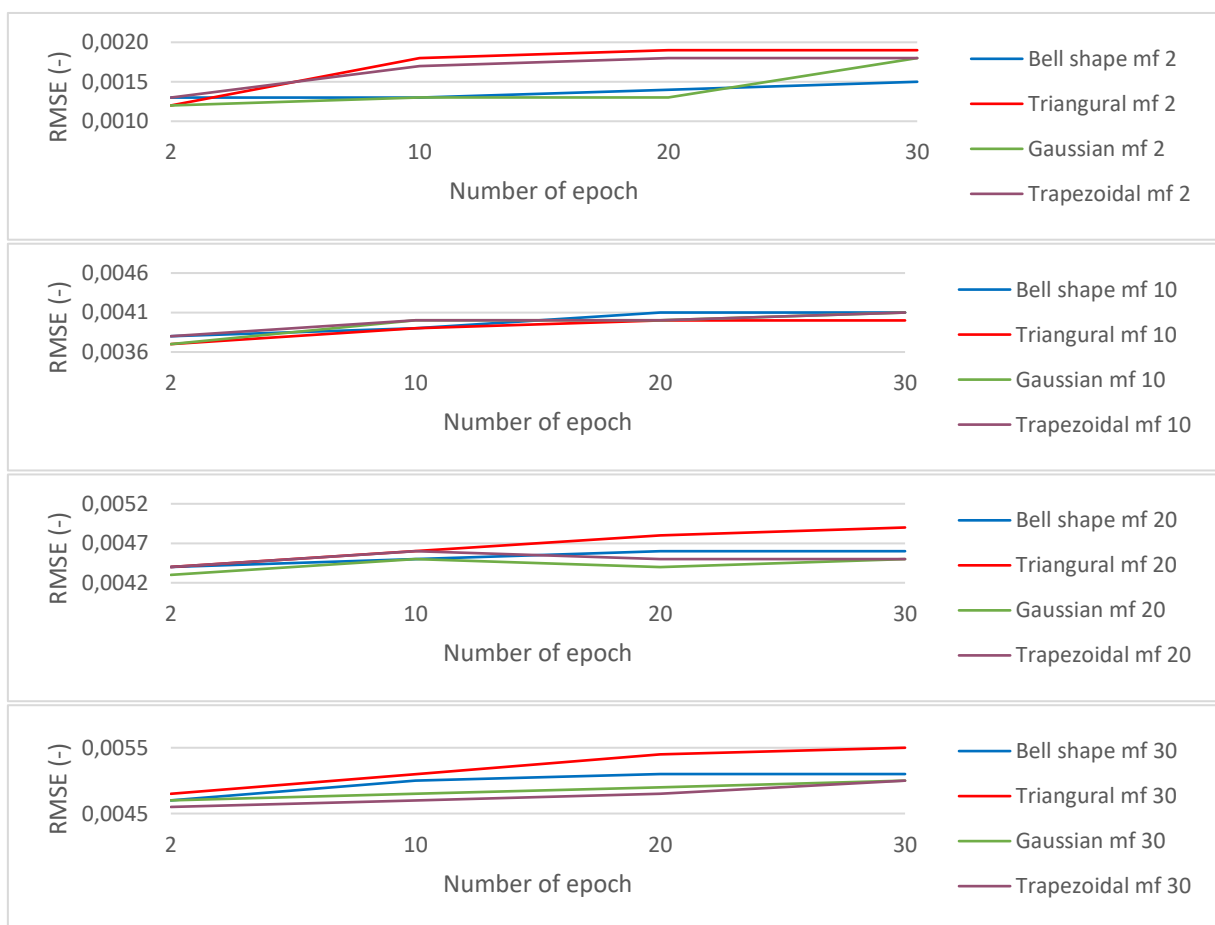
number of epoch	number of mf	elapsed time	PESQ (-)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	4,1	3,247	2,817	3,155	3,485
10	2	5,9	2,930	2,773	3,066	2,793
20	2	7,7	2,801	2,833	3,056	2,517
30	2	11,6	2,820	2,856	2,964	2,365
2	10	48,7	1,705	1,807	1,732	1,737
10	10	49,3	1,593	1,653	1,678	1,659
20	10	53,7	1,602	1,611	1,669	1,658
30	10	54,6	1,641	1,591	1,660	1,657
2	20	56,9	1,568	1,562	1,609	1,599
10	20	57,8	1,506	1,521	1,593	1,574
20	20	59,2	1,488	1,492	1,532	1,536
30	20	115,1	1,456	1,477	1,498	1,518
2	30	263,6	1,474	1,442	1,477	1,431
10	30	279,5	1,469	1,428	1,455	1,405
20	30	296,3	1,451	1,403	1,418	1,392
30	30	308,7	1,444	1,392	1,401	1,387



Obr. II. 6: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem PESQ pro příkaz "blind up"

Tab. II. 8: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem RMSE pro příkaz "blind down"

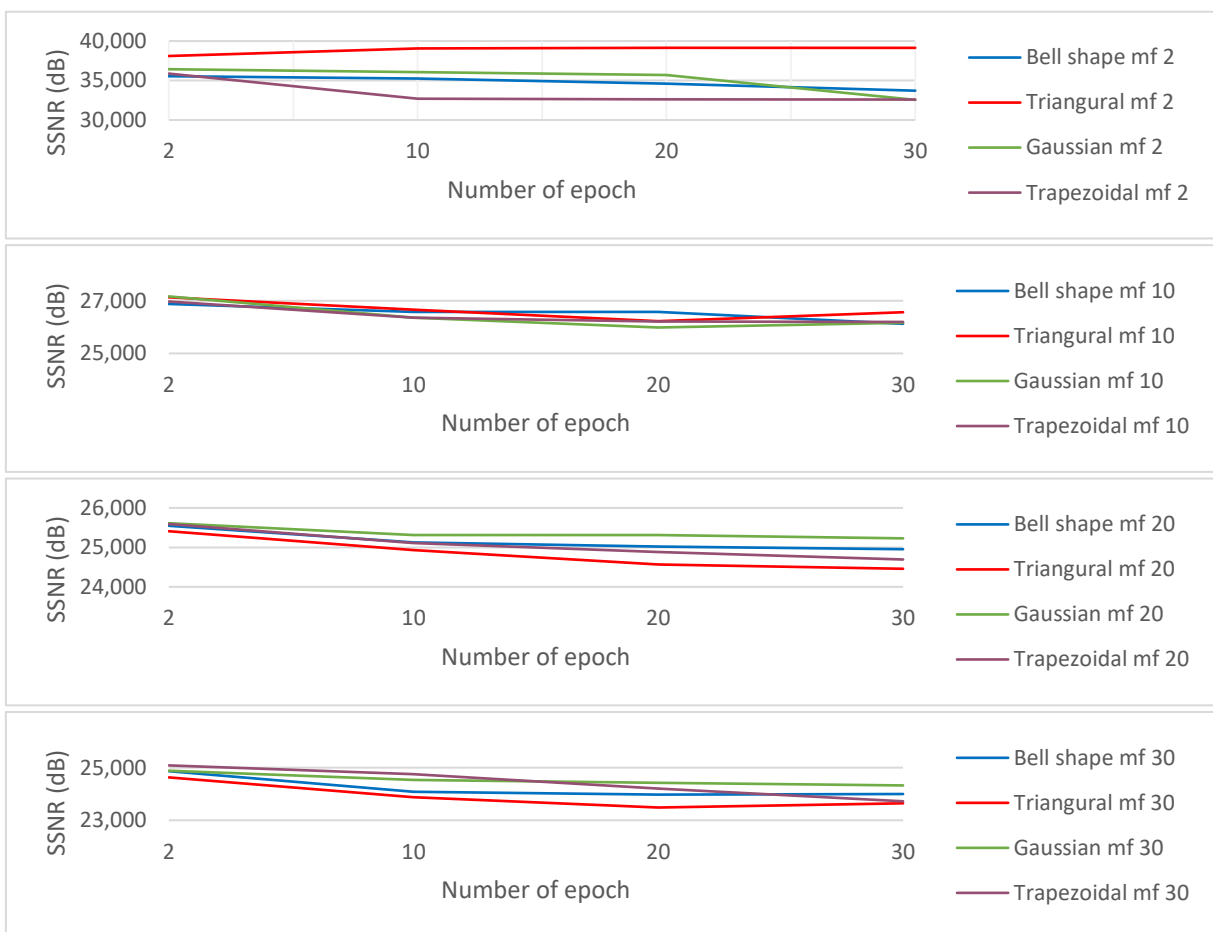
number of epoch	number of mf	elapsed time	RMSE (-)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	4,5	0,0013	0,0012	0,0012	0,0013
10	2	6,9	0,0013	0,0018	0,0013	0,0017
20	2	7,1	0,0014	0,0019	0,0013	0,0018
30	2	12,9	0,0015	0,0019	0,0018	0,0018
2	10	50,7	0,0038	0,0037	0,0037	0,0038
10	10	52,5	0,0039	0,0039	0,0040	0,0040
20	10	55,8	0,0041	0,0040	0,0040	0,0040
30	10	57,3	0,0041	0,0040	0,0041	0,0041
2	20	59,7	0,0044	0,0044	0,0043	0,0044
10	20	62,0	0,0045	0,0046	0,0045	0,0046
20	20	63,7	0,0046	0,0048	0,0044	0,0045
30	20	120,1	0,0046	0,0049	0,0045	0,0045
2	30	283,6	0,0047	0,0048	0,0047	0,0046
10	30	290,5	0,0050	0,0051	0,0048	0,0047
20	30	308,3	0,0051	0,0054	0,0049	0,0048
30	30	319,4	0,0051	0,0055	0,0050	0,0050



Obr. II. 7: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem RMSE pro příkaz "blind down"

Tab. II. 9: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem SSNR pro příkaz "blind down"

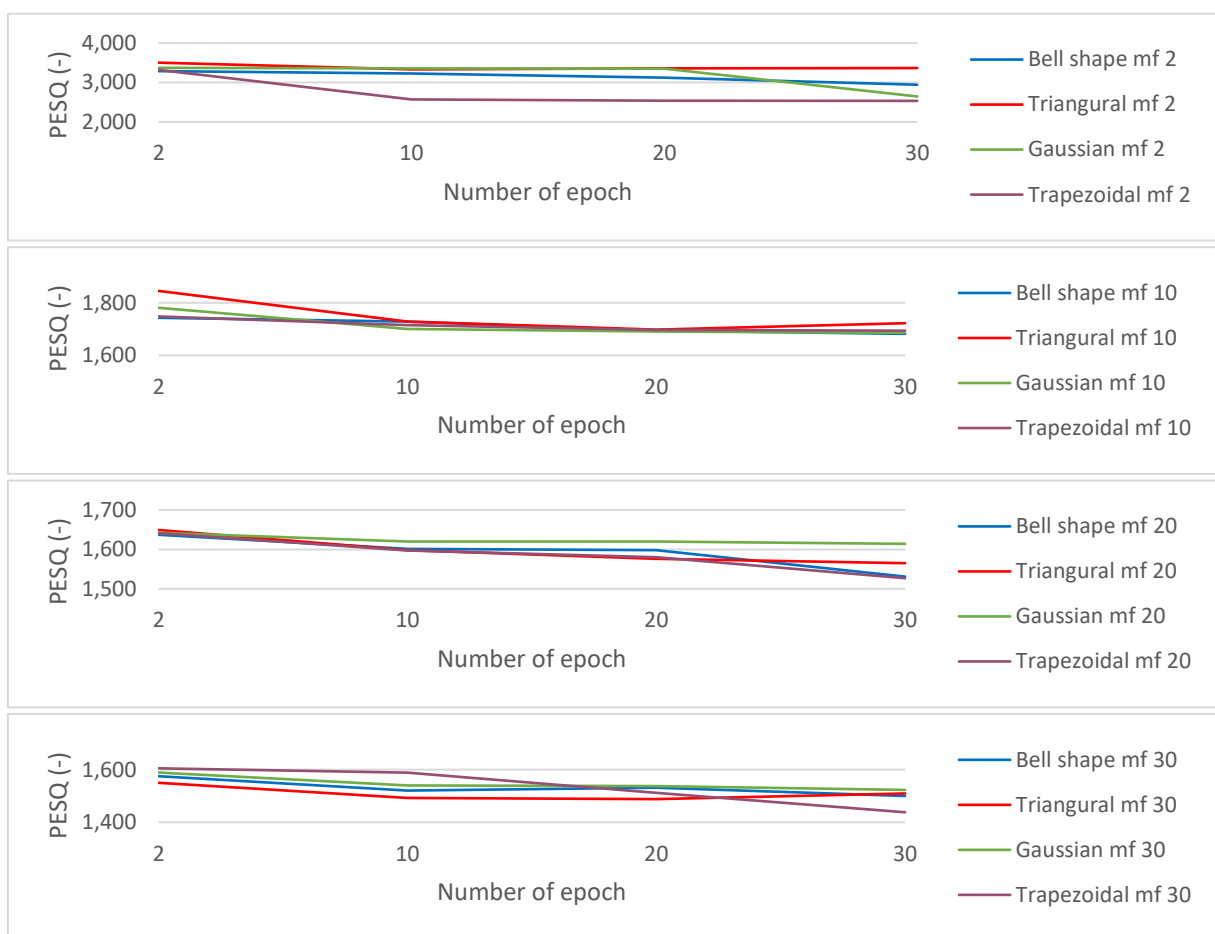
number of epoch	number of mf	elapsed time	SSNR (dB)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	4,5	35,547	38,106	36,436	35,876
10	2	6,9	35,267	39,065	36,046	32,710
20	2	7,1	34,583	39,124	35,691	32,609
30	2	12,9	33,716	39,139	32,548	32,588
2	10	50,7	26,880	27,132	27,166	26,972
10	10	52,5	26,576	26,657	26,361	26,364
20	10	55,8	26,579	26,226	25,986	26,214
30	10	57,3	26,124	26,571	26,165	26,196
2	20	59,7	25,547	25,409	25,612	25,591
10	20	62,0	25,125	24,929	25,309	25,109
20	20	63,7	25,018	24,569	25,311	24,886
30	20	120,1	24,954	24,458	25,229	24,693
2	30	283,6	24,865	24,629	24,883	25,085
10	30	290,5	24,083	23,871	24,527	24,747
20	30	308,3	23,975	23,485	24,424	24,203
30	30	319,4	23,998	23,651	24,324	23,721



Obr. II. 8: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem SSNR pro příkaz "blind down"

Tab. II. 10: Výsledky procesu pro různé parametry struktury hodnocené klasifikátorem PESQ pro příkaz "blind down"

number of epoch	number of mf	elapsed time	PESQ (-)			
			Bell shape	Triangural	Gaussian	Trapezoidal
2	2	4,5	3,289	3,503	3,372	3,319
10	2	6,9	3,232	3,333	3,360	2,572
20	2	7,1	3,125	3,359	3,354	2,543
30	2	12,9	2,944	3,365	2,647	2,535
2	10	50,7	1,743	1,845	1,781	1,748
10	10	52,5	1,728	1,728	1,700	1,715
20	10	55,8	1,695	1,698	1,691	1,697
30	10	57,3	1,682	1,722	1,686	1,694
2	20	59,7	1,637	1,649	1,642	1,641
10	20	62,0	1,601	1,598	1,620	1,597
20	20	63,7	1,598	1,576	1,620	1,580
30	20	120,1	1,531	1,565	1,614	1,527
2	30	283,6	1,575	1,550	1,589	1,605
10	30	290,5	1,520	1,493	1,540	1,589
20	30	308,3	1,532	1,488	1,538	1,512
30	30	319,4	1,500	1,510	1,523	1,438

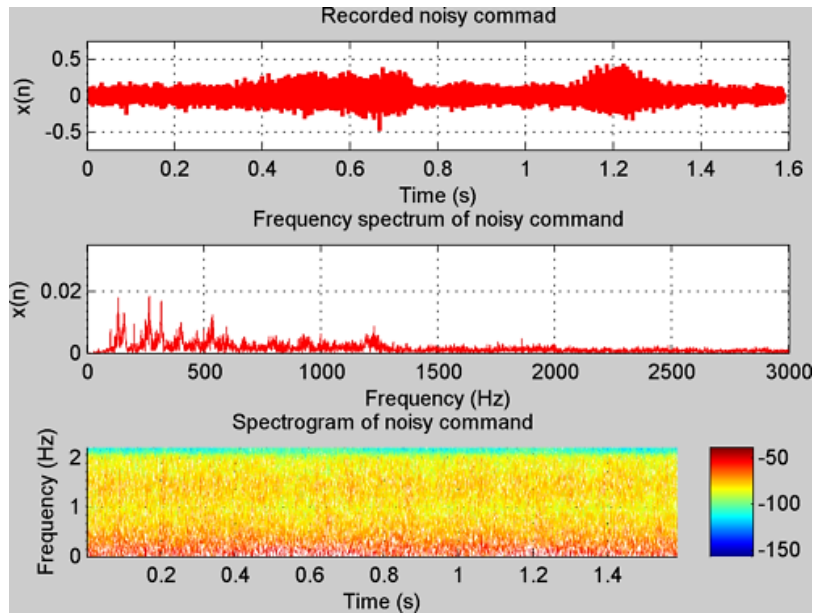


Obr. II. 9: Porovnání výsledků procesu pro různá nastavení struktury hodnocené klasifikátorem PESQ pro příkaz "blind down"

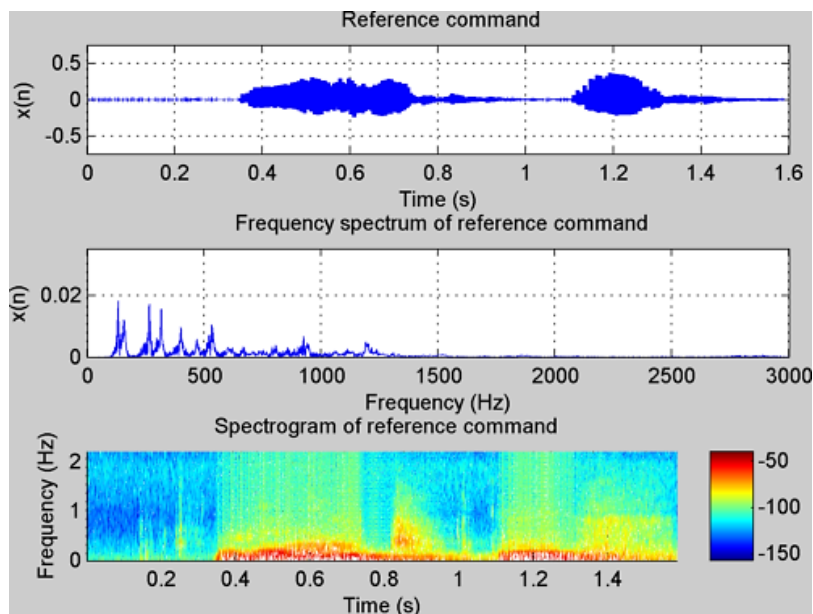
Příloha C

Následující obrázky představují vizuální výsledky procesu potlačení šumu příkazů „*light off*“, „*blind up*“ a „*blind down*“. Signály jsou vykresleny v časovém, frekvenčním a časově frekvenčním spektru. Nastavení ANFIS struktury je pro všechny příkazy jednotné:

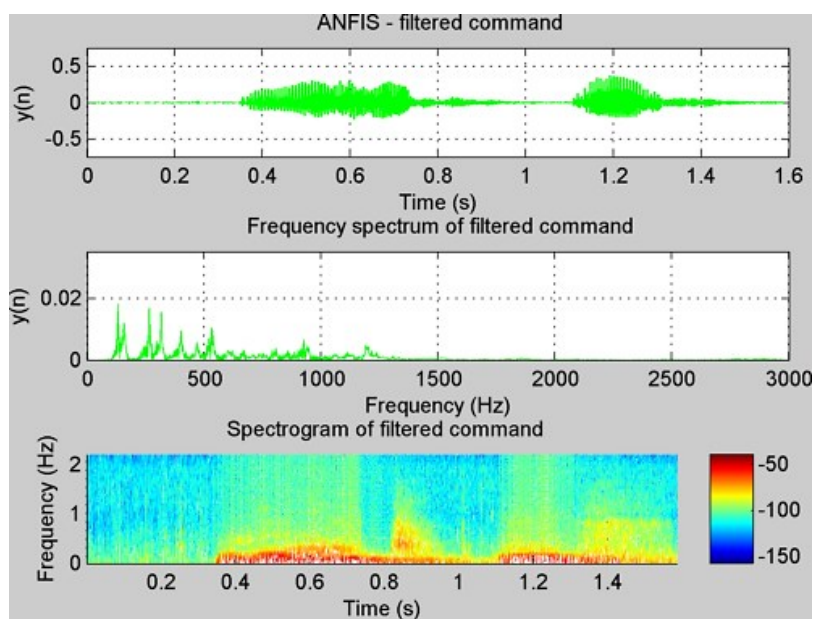
- Number of mf = „2“
- Number of epoch = „2“
- Mf type = „trimf“



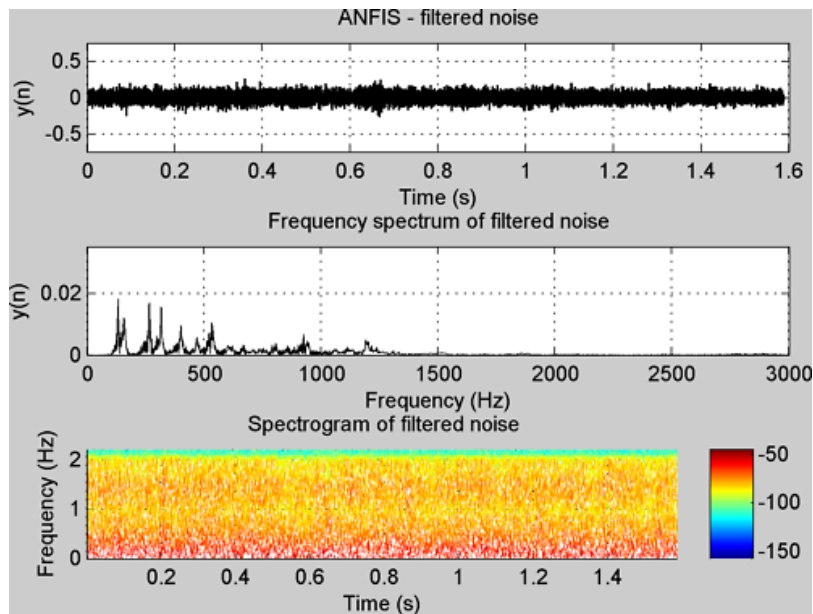
Obr. III. 1: Časové, frekvenční a časově frekvenční spektrum příkazu "light off" se šumem ($mf = 2$, $epoch = 2$, $type = "trimf"$)



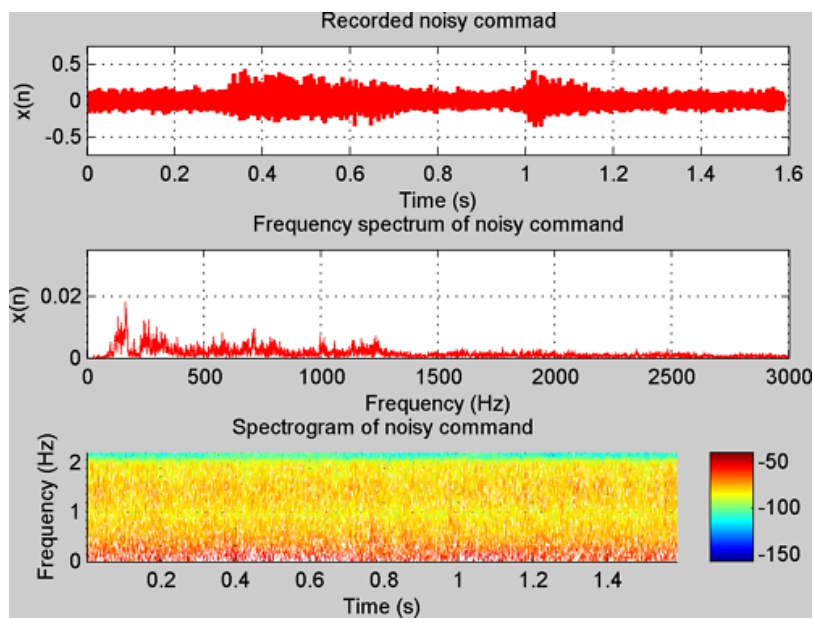
Obr. III. 2: Časové, frekvenční a časově frekvenční spektrum referenčního příkazu "light off" ($mf = 2$, $epoch = 2$, $type = "trimf"$)



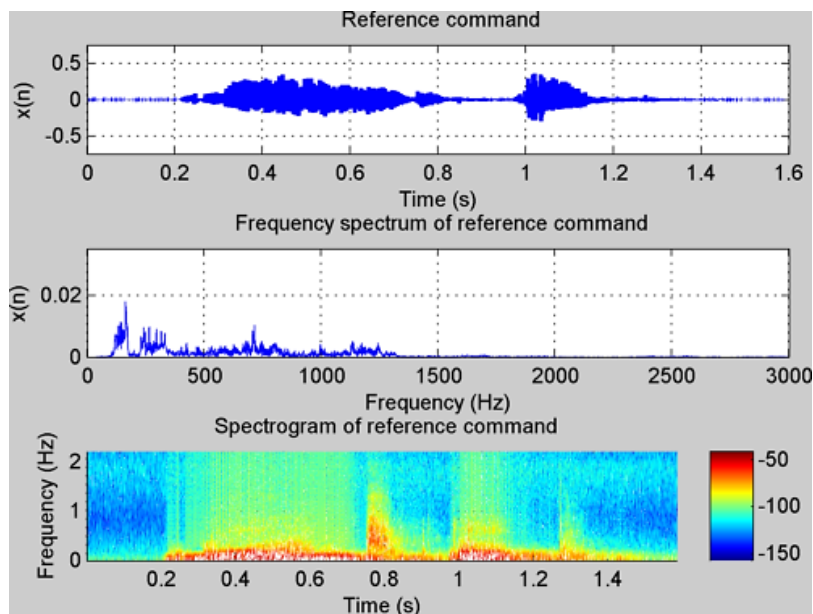
Obr. III. 3: Časové, frekvenční a časově frekvenční spektrum příkazu "light off" s potlačeným šumem ($mf = 2$, $epoch = 2$, $type = "trimf"$)



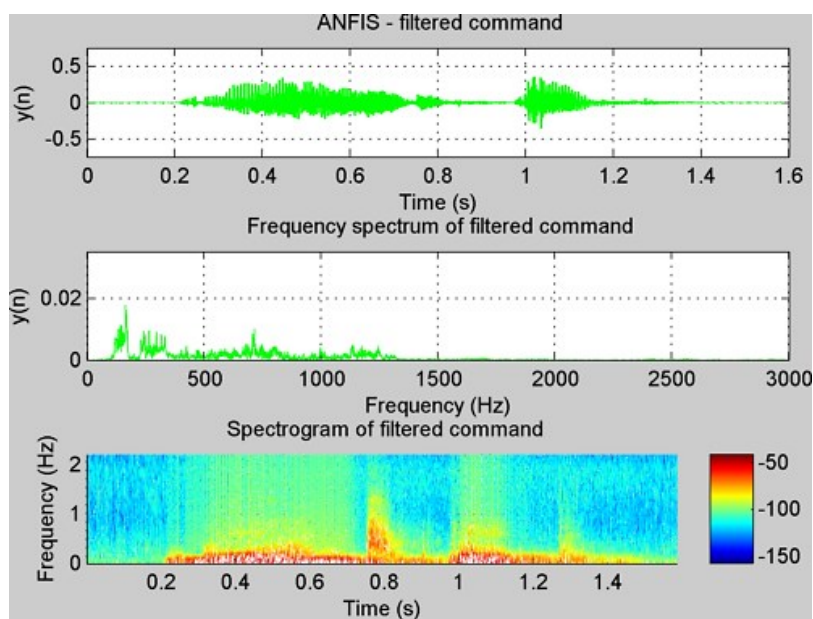
Obr. III. 4: Časové, frekvenční a časově frekvenční spektrum šumu z příkazu "light off" ($mf = 2$, $epoch = 2$, $type = "trimf"$)



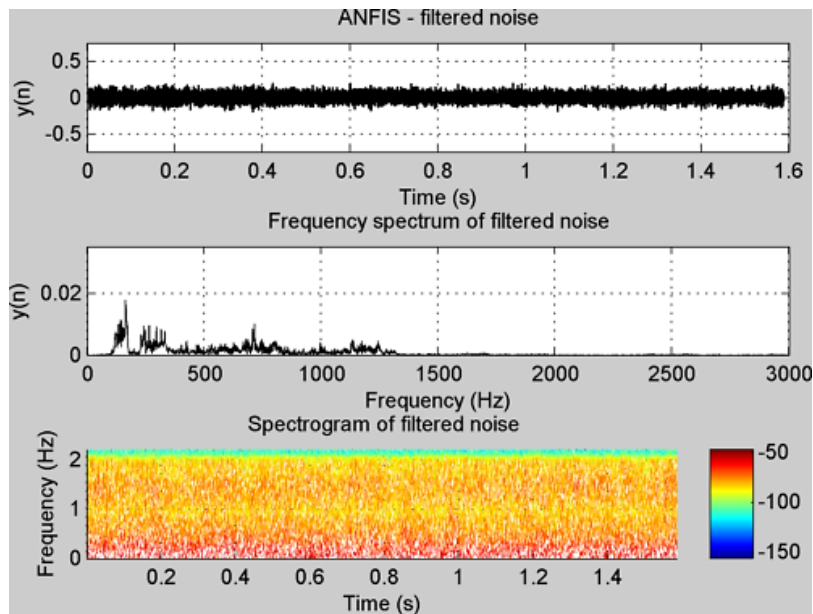
Obr. III. 5: Časové, frekvenční a časově frekvenční spektrum příkazu "blind up" se šumem ($mf = 2$, $epoch = 2$, $type = "trimf"$)



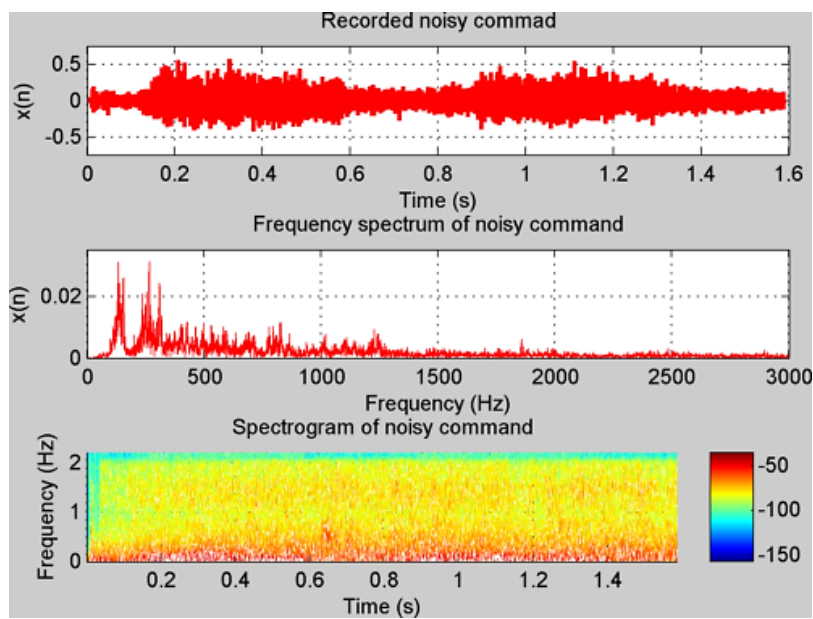
Obr. III. 6: Časové, frekvenční a časově frekvenční spektrum referenčního příkazu "blind up" ($mf = 2$, $epoch = 2$, $type = "trimf"$)



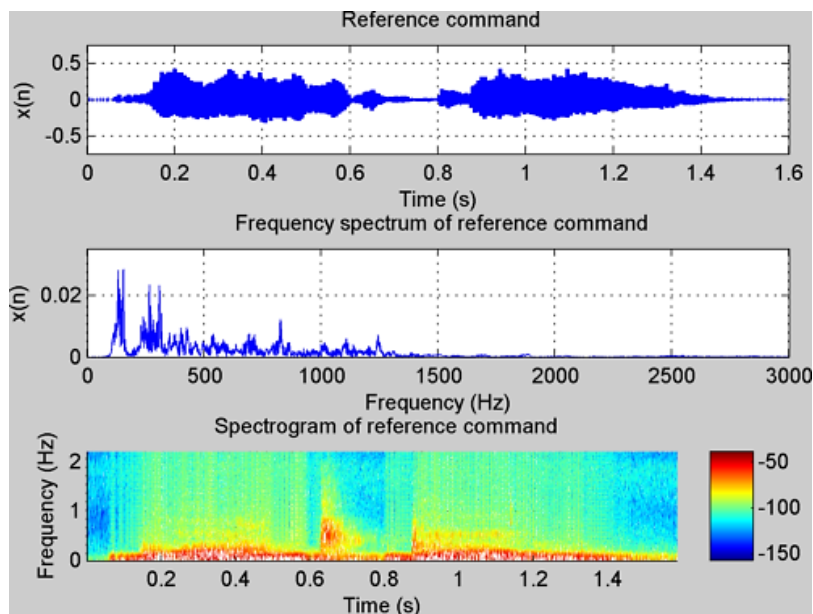
Obr. III. 7: Časové, frekvenční a časově frekvenční spektrum příkazu "blind up" s potlačeným šumem ($mf = 2$, $epoch = 2$, $type = "trimf"$)



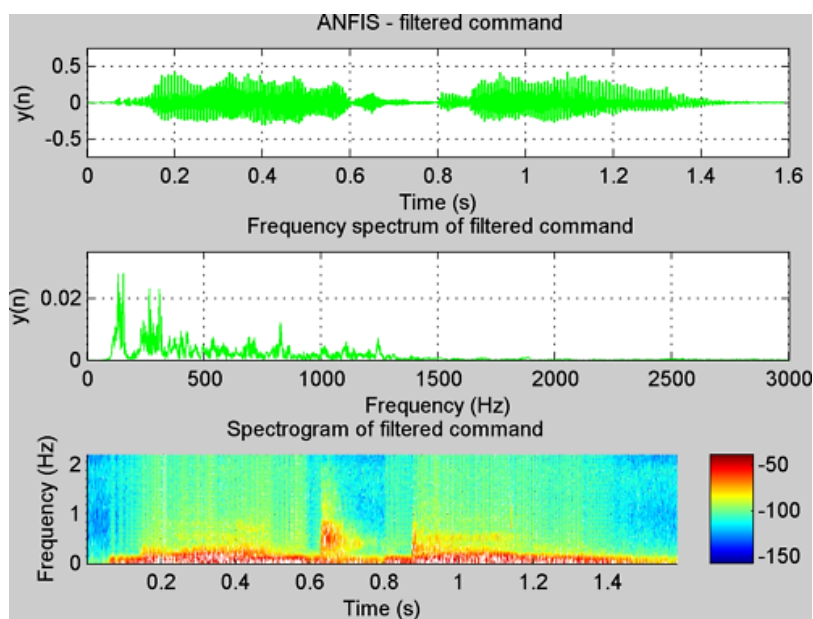
Obr. III. 8: Časové, frekvenční a časově frekvenční spektrum šumu z příkazu "blind up" ($mf = 2$, $epoch = 2$, $type = "trimf"$)



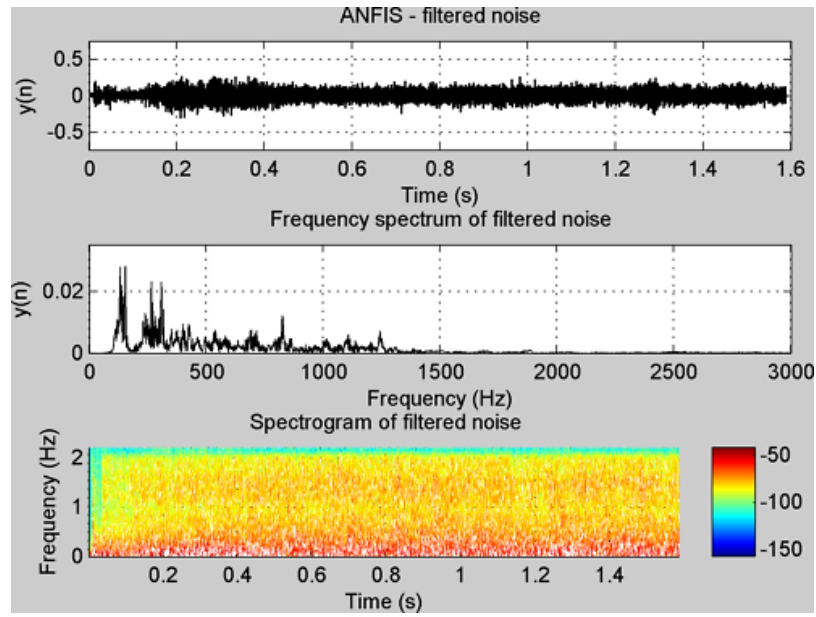
Obr. III. 9: Časové, frekvenční a časově frekvenční spektrum příkazu "blind down" se šumem ($mf = 2$, $epoch = 2$, $type = "trimf"$)



Obr. III. 10: Časové, frekvenční a časově frekvenční spektrum referenčního příkazu "blind down" ($mf = 2$, $epoch = 2$, $type = "trimf"$)



Obr. III. 11: Časové, frekvenční a časově frekvenční spektrum příkazu "blind down" s potlačeným šumem ($mf = 2$, $epoch = 2$, $type = "trimf"$)



Obr. III. 12: Časové, frekvenční a časově frekvenční spektrum šumu z příkazu "blind down" ($mf = 2$, $epoch = 2$, $type = "trimf"$)

Příloha D

Obsah CD:

- **Experimenty** – Tato složka obsahuje data a výsledky z provedených experimentů
- **Program** – V této složce se nachází všechny použité funkce a vytvořený software
- **Text** – Tato složka obsahuje text této práce v elektronické podobě