

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky  
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2017

Bc. Lukáš Pospíšil



# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

## FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

## ÚSTAV RADIOELEKTRONIKY

DEPARTMENT OF RADIO ELECTRONICS

# ANALÝZA ROC KŘIVEK ZVUKOVÝCH SIGNÁLŮ A JEJICH SROVNÁNÍ

ANALYSIS AND COMPARISON OF ROC CURVES OF AUDIO SIGNALS

## DIPLOMOVÁ PRÁCE

MASTER'S THESIS

## AUTOR PRÁCE

AUTHOR

**Bc. Lukáš Pospíšil**

## VEDOUCÍ PRÁCE

SUPERVISOR

**doc. RNDr. Jitka Poměnková, Ph.D.**

**BRNO 2017**

# Diplomová práce

magisterský navazující studijní obor **Elektronika a sdělovací technika**

Ústav radioelektroniky

**Student:** Bc. Lukáš Pospíšil

**ID:** 110990

**Ročník:** 2

**Akademický rok:** 2016/17

**NÁZEV TÉMATU:**

## Analýza ROC křivek zvukových signálů a jejich srovnání

### POKYNY PRO VYPRACOVÁNÍ:

Cílem projektu je porovnat hodnocení účinnosti klasifikace prostřednictvím ROC křivek s přihlédnutím k rozsahu hodnoceného souboru. Pomocí tohoto posouzení rozhodnout o vhodnosti metody pro identifikaci a zvolených parametrech metody. V průběhu práce student vytvoří výpočtovou proceduru pro hodnocení a pro vzájemné srovnání ROC křivek.

Po prostudování problematiky klasifikace ROC křivek a možností jejich statistického porovnání zpracuje student zjištěné poznatky do metodické části práce. Následně navrhne metodický koncept pro hodnocení účinnosti klasifikace prostřednictvím ROC křivek s ohledem na rozsah hodnoceného souboru. Pro aplikaci student využije datové soubory z oblasti řečové aktivity. Student provede podrobný popis datových souborů, jejich zdrojů a provedených způsobů transformace dat. Na základě navržené metodiky a vytvořené výpočtové procedury provede analýzu dat. V závěru student vyhodnotí zjištěné poznatky a formuluje metodická doporučení.

### DOPORUČENÁ LITERATURA:

[1] ZOU, G. Y., YUE, L. Using confidence intervals to compare several correlated areas under the receiver operating characteristic curves, *Statistics in Medicine*, 2013, Vol. 32, pp. 5077-5090.

[2] WESTIN, L. K. Receiver operating characteristic (ROC) analysis. Evaluating discriminance effects among decision support systems. 28 pages. [online] Cited 2014-02-11. Available at: <http://nutkin.cs.umu.se/research/reports/2001/018/part1.pdf>.

**Termín zadání:** 6. 2. 2017

**Termín odevzdání:** 16.5.2017

**Vedoucí práce:** doc. RNDr. Jitka Poměnková, Ph.D.

**Konzultant:**



prof. Ing. Tomáš Kratochvíl, Ph.D.  
předseda oborové rady

### UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

## ABSTRAKT

Tato práce se zabývá možností využití ROC křivek pro ohodnocení metod, které pracují se zvukovými signály. Konkrétně se zaměřuje na možnosti detekce stresu v řečových signálech. Samotná detekce je prováděna ve frekvenční oblasti signálů. Z tohoto důvodu se práce snaží nalézt nejlepší metodu pro odhad spektra a řeší se také problematika velikosti spektra vstupních signálů. Je také navržen klasifikátor využívající ROC křivky, který rozhoduje, zda je vstupní signál stresový, nebo nestresový. Výstupem této práce jsou poznatky získané provedením analýz, z kterých jsou následně vyvozena doporučení.

## KLÍČOVÁ SLOVA

ROC, AUC, lichoběžníková metoda, odhad spektra, neparametrické metody odhadu spektra, parametrické metody odhadu spektra, periodogram, Bartlettova metoda, Welchův periodogram, AR model, Yule-Walkerova metoda, Burgova metoda, detekce stresu, zvukové signály

## ABSTRACT

This thesis deals with opportunity of ROC curve usage in the description of methods that work with sound signals. Specifically, it focuses on ways of detecting of stress in speech signals. The detection itselfs is done in a range of frequencies of the sound signal. There is also a classifier designed using ROC curves that decides whether the input signal is stressed or not. The output of this thesis are findings gathered from analyses and also some recommendation based on those analyses.

## KEYWORDS

ROC, receiver operating characteristic, AUC, trapezoidal rule, spectrum estimation, non-parametric methods for spectrum estimation, parametric methods for spectrum estimation, periodogram, Bartlett method, Welch method, AR model, Yule-Walker method, Burg method, detection of stress, audio signal

POSPÍŠIL, Lukáš. *Analýza ROC křivek zvukových signálů a jejich srovnání*. Brno, 2017, 52 s. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky. Vedoucí práce: doc. RNDr. Jitka Poměnková, Ph.D.



## PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Analýza ROC křivek zvukových signálů a jejich srovnání“ jsem vypracoval(a) samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor(ka) uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil(a) autorská práva třetích osob, zejména jsem nezasáhl(a) nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom(a) následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno .....

.....

podpis autora(-ky)

## PODĚKOVÁNÍ

Rád bych poděkoval vedoucí diplomové práce paní doc. RNDr. Jitkce Poměnkové, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno .....

.....

podpis autora(-ky)



Faculty of Electrical Engineering  
and Communication  
Brno University of Technology  
Purkynova 118, CZ-61200 Brno  
Czech Republic  
<http://www.six.feec.vutbr.cz>

## PODĚKOVÁNÍ

Výzkum popsany v této bakalářské práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

Brno .....

.....  
podpis autora(-ky)



EVROPSKÁ UNIE  
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ  
INVESTICE DO VAŠÍ BUDOUCNOSTI



# OBSAH

<b>1</b>	<b>Úvod</b>	<b>10</b>
<b>2</b>	<b>Teorie</b>	<b>11</b>
2.1	Počítačové zpracování řeči . . . . .	11
2.2	Odhad spektra . . . . .	12
2.2.1	Neparametrické metody odhadu spektra . . . . .	12
2.2.2	Parametrické metody odhadu spektra . . . . .	14
2.3	Teorie ROC křivek . . . . .	17
2.3.1	Základní pojmy . . . . .	17
2.3.2	ROC křivky . . . . .	19
2.3.3	AUC . . . . .	21
<b>3</b>	<b>Hodnocení účinnosti klasifikátoru prostřednictvím ROC křivek</b>	<b>23</b>
3.1	Zdrojové data a předzpracování signálů . . . . .	23
3.2	Analýza metod odhadu spektra . . . . .	24
3.3	Analýza vlivu frekvenčního rozsahu spektra vstupních signálů . . . . .	26
3.4	Klasifikace pomocí ROC křivek . . . . .	31
3.5	Vliv zašumění signálů na detekci . . . . .	36
<b>4</b>	<b>Struktura programu</b>	<b>39</b>
4.1	Předpřipravené analýzy . . . . .	40
4.2	Program pro vlastní analýzy . . . . .	41
<b>5</b>	<b>Závěr</b>	<b>45</b>
	<b>Literatura</b>	<b>48</b>
	<b>Seznam symbolů, veličin a zkratk</b>	<b>49</b>
	<b>Seznam příloh</b>	<b>50</b>
<b>A</b>	<b>Ukázka odhadu spekter</b>	<b>51</b>
<b>B</b>	<b>Obsah přiloženého CD</b>	<b>52</b>

# SEZNAM OBRÁZKŮ

2.1	Princip digitalizace (vzorkování a kvantizace). [7]	11
2.2	Vztahy mezi pravděpodobnostmi pozitivní diagnostiky a testu. [4]	18
2.3	Tři dvojice normálního rozdělení pravděpodobností pro velký, střední a malý překryv. [4]	20
2.4	Odpovídající ROC křivky pro tři dvojice normálního rozdělení pravděpodobností pro velký, střední a malý překryv. [4]	20
2.5	Ukázka jaký vliv má počet bodů ROC křivky na hodnotu AUC. [4]	21
3.1	Odhad spektra signálu <i>signStres7</i> pomocí periodogramu.	25
3.2	Tři ROC křivky pro dvojici signálů <i>signNestres2</i> a <i>signStres7</i> získaných různými metodami odhadu spektra.	25
3.3	Grafické znázornění kombinací vstupních signálů pro ROC křivky, na kterých byla analýza pro odhad metody spektra provedena.	26
3.4	Grafické znázornění kombinací vstupních signálů pro ROC křivky, na kterých byla provedena analýza pro vliv frekvenčního rozsahu spektra vstupních signálů. Jako vstupní data je vždy použit jeden stresový a jeden nestresový signál.	28
3.5	ROC křivky pro dvojici signálů <i>signStres6</i> a <i>signNestres1</i> s různou velikostí vstupního spektra.	29
3.6	Chybný průběh ROC křivky (modrá čára). Tato křivka by se neměla nacházet na žádném úseku pod červenou čarou.	30
3.7	Grafické znázornění, jak bylo získáno referenční spektrum pro ROC křivky <i>reference</i> .	31
3.8	Spektra signálů <i>maxNestres</i> , <i>minStres</i> a <i>reference</i> .	32
3.9	Vývojový diagram klasifikátoru.	33
3.10	Grafické znázornění dat, na kterých byl klasifikátor otestován.	34
3.11	ROC křivky testovaných dat. Červené ROC křivky jsou nestresových signálů a modré ROC křivky jsou stresových signálů.	34
3.12	Spektrum zašuměného a nezašuměného nestresového signálu <i>signNestres1</i> a stresového signálu <i>signStres8</i> . Plnou čarou je zobrazeno spektrum signálu <i>signNestres1</i> a čárkovaně spektrum signálu <i>signStres8</i> . Modrou čarou jsou spektra nezašuměných signálů a červenou čarou jsou spektra signálů s přidaným bílým gaussovským šumem $SNR = 3dB$	36
3.13	Vliv zašumění signálu na spektrum. Spektrum nezašuměného signálu <i>signNestres7.wav</i> ze zobrazeno zeleně, spektrum <i>reference</i> modře a zašuměný signál <i>signNestres7.wav</i> ( $SNR = 0.1 dB$ ) červeně.	37

3.14	Modrá ROC křivka je vytvořena z signálů <i>signNestres7</i> a <i>reference</i> . Červená ROC křivka je vytvořena ze zašuměného signálu <i>signNestres7</i> ( $SNR = 0, 1dB$ ) a <i>reference</i> . . . . .	38
4.1	Zjednodušený vývojový diagram programu s předpřipravenými analýzy.	41
4.2	GUI aplikace pro testování. . . . .	42
4.3	Zobrazení spektra audio signálu. . . . .	43
4.4	Přímé načtení spektra ze souboru. . . . .	43
4.5	Zobrazení ROC křivky a její AUC hodnoty. . . . .	44
4.6	Ukázky warningů, při chybně zadaných parametrech do aplikace. . . . .	44

# SEZNAM TABULEK

2.1	Hodnoty pásem prvních tří formantů pro české samohlásky. [1] . . . .	12
3.1	Vstupní data pro testování. . . . .	24
3.2	Hodnoty <i>AUC</i> pro jednotlivé kombinace <i>signNestres2</i> × <i>signStres1-11</i> . . . . .	27
3.3	Výpočetní náročnost metod odhadu spektra. . . . .	27
3.4	Závislost <i>AUC</i> na velikosti spektra vstupních signálů. . . . .	29
3.5	Hodnoty <i>AUC</i> jednotlivých ROC křivek stresových signálů. . . . .	35
3.6	Hodnoty <i>AUC</i> jednotlivých ROC křivek nestresových signálů. . . . .	35

# 1 ÚVOD

Počítačové zpracování řeči je v dnešní době velmi důležitou součástí řady systémů. Například se jedná o systémy pro rozpoznání mluvené řeči, počítačovou syntézu řeči, identifikaci a validaci člověka na základě jeho hlasu, přenos řečového signálu, atd. Na základě konkrétního použití se liší požadavky na zpracování signálů. U přenosu řečového signálu nás nejvíce zajímá přenosová rychlost. Při malé přenosové rychlosti může být zakódovaná informace přenesena kanálem s menší kapacitou a efektivně zpracována. Avšak při malé přenosové rychlosti může dojít k redukci informačního obsahu signálu a následné ztrátě kvality při rekonstrukci řeči. U problematiky rozpoznávání řeči se snažíme snížit informační redundanci zvukových signálů a to bez ztráty informačního obsahu signálu. Díky stále výkonnějším počítačům se používají robustnější a výpočetně náročnější algoritmy. Při syntéze řeči je kladen důraz na dokonalou srozumitelnost (vyhlazování spekter, atd.). Identifikaci řečníka je založena na zvýrazňování odlišností hlasů, z které opět plyne zcela odlišný způsob zpracování signálů.[1]

Možnost uplatnění zpracování zvukových signálů je tedy více než velká. Tato práce se věnuje problematice detekce hlasového záznamu ve stresovém emočním stavu (dále označováno jako detekce stresu). Na podobných principech pracuje i detekce jiných emocí, nebo atypických přízvuků ve zvukových signálech. Práce navazuje na diplomovou práci [6]. Cílem práce je zaměřit se na porovnání metod, které je možné při detekci využít a pomocí ROC křivek srovnat jejich výkonnost. Stanovit doporučení pro volbu konkrétních metod a částí klasifikátoru. Využití ROC křivek pro tuto problematiku je velmi vhodné, protože máme pouze omezenou (relativně malou) skupinu testovacích dat. A právě srovnání různých metod na stejné množině testovacích dat je jedna z mnoha možných využití ROC křivek.

Práce je rozdělena tří hlavních kapitol. V kapitole 2 získá čtenář znalosti potřebné k porozumění této problematice. V následující kapitole 3 jsou popsány jednotlivé analýzy, které byly provedeny. Kapitola 4 je věnována konkrétním řešením podpurných programů. V poslední kapitole 5 jsou shrnuté poznatky a doporučení.

Podpurné programy byly napsány v programovém prostředí MATLAB R2014b a práce byla vysázena systémem  $\text{\LaTeX}$ .

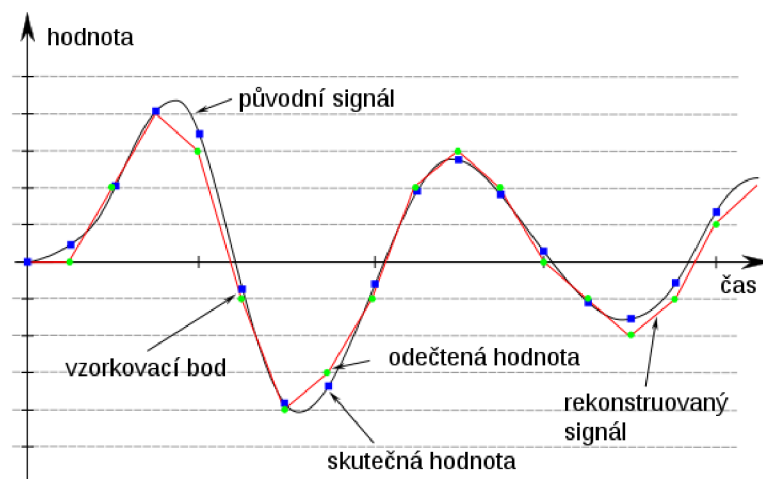


## 2 TEORIE

Po přečtení této kapitoly získá čtenář základní znalosti, které jsou potřeba pro pochopení problematiky této práce.

### 2.1 Počítačové zpracování řeči

Zvukové signály jsou spojité analogové signály, které byly získány nahráním pomocí mikrofónu. Abychom mohli provést jejich zpracování (analýzu) v počítači, musíme provést tzv. digitalizaci. Ta spočívá v převedení spojitého analogového signálu na digitální. To je provedeno procesem vzorkování a následným kvantováním (obr. 2.1). Procesem vzorkování získáme z původního spojitého signálu  $x(t)$  diskrétní signál, pro který platí  $x(n) = x(n \cdot 1/F_s)$ , kde  $F_s$  je vzorkovací frekvence. Kvantizace spočívá v zaokrouhlení hodnoty spojitého signálu v čase vzorkování k nejbližší kvantizační hladině.



Obr. 2.1: Princip digitalizace (vzorkování a kvantizace). [7]

Zvukový signál může být zobrazen v časové, nebo frekvenční oblasti. Časové průběhy zobrazují vývoj amplitudy v čase a jsou kvaziperiodické. Ve frekvenční oblasti je signál zobrazen jako závislost amplitudy signálu ve frekvenci (amplitudové spektrum). Tvar spektra zobrazuje široké vrcholky odpovídající jednotlivým formantům hlasového traktu.

Nejjednodušší metoda výpočtu spektra je pomocí krátkodobé Fourierovi transformace  $S(\omega, n)$ , která je pro diskrétní signál  $s(k)$  definována vztahem [1]

$$S(\omega, n) = \sum_{k=-\infty}^{\infty} s(k)h(n-k)e^{-j\omega k} \quad (2.1)$$

kde  $\omega$  je kruhová frekvence,  $n$  je diskétní proměnná času a  $h(n)$  je okénková funkce, kterou vybereme úsek signálu.

Z hlediska detekce stresu v hlasu můžeme s výhodou využít detekce samohlásek, které se vyznačují vysokou amplitudou a delším trváním. Pro české samohlásky jsou podstatné první tři formanty (oblasti koncentrace akustické energie). V tabulce 2.1 jsou uvedeny frekvence prvních tří formantů [1].

Tab. 2.1: Hodnoty pásem prvních tří formantů pro české samohlásky. [1]

	$F_1[Hz]$	$F_2[Hz]$	$F_3[Hz]$
/i/, /í/	300 - 500	2000 - 2800	2600 - 3500
/e/, /é/	480 - 700	1560 - 2100	2500 - 3500
/a/, /á/	700 - 1100	1100 - 1500	2500 - 3000
/o/, /ó/	500 - 700	850 - 1200	2500 - 3000
/u/, /ú/	300 - 500	600 - 1000	2400 - 2900

## 2.2 Odhad spektra

Z literatury [1] je patrné, že pro detekci stresu v hlasových signálech je výhodnější detekce v frekvenční oblasti. Budeme potřebovat získat spektrum zvukových signálů. Této problematice budou věnovány následující kapitoly. Metody odhadu spektra lze rozdělit na dva základní přístupy: neparametrické a parametrické metody. Nejprve bude představeno několik metod využívajících neparametrický odhad spektra. V další kapitole jsou parametrické metody založené na odhadu autoregresního modelu. Následující kapitoly byly vytvořeny volným překladem z literatury [3] a [2].

### 2.2.1 Neparametrické metody odhadu spektra

U těchto metod je odhad spektra založen pouze na konečném počtu vzorků signálu. Nepotřebujeme žádné znalosti o tom, jak byla data vytvořena. Proto jsou tyto metody nazývány neparametrické. Jejich nevýhodou je, že snižují frekvenční rozlišení. Základní metodou je periodogram, z kterého jsou následně odvozeny další metody jako např. Welch, Bartlett, Blackman a Tukey.

## Metoda periodogramu

Předpokládáme odhad spektra hustoty výkonu (dále označováno jako spektrum) z jednoho vzorku náhodného procesu. V praxi je spojité signál  $x_a(t)$  navzorkován frekvencí splňující vzorkovací teorém  $F_s > 2B$ , kde  $B$  je nejvyšší frekvence obsažená ve spektru náhodného procesu. Také předpokládáme, že tento navzorkovaný signál je konečný,  $x(n), 0 \leq n \leq N-1$ . Z takového signálu můžeme vypočítat jeho časově průměrovanou autokorelaci [3]

$$r_{xx}(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x^*(n)x(n+m), \quad m \geq 0 \quad (2.2)$$

Tuto autokorelaci můžeme využít pro odhad spektra

$$P_{xx}(f) = \sum_{m=-(N-1)}^{N-1} r_{xx}(m)e^{-j2\pi fm} \quad (2.3)$$

Dosažením vztahu 2.2 za  $r_{xx}(m)$  dostaneme výsledný vztah pro odhad spektra pomocí periodogramu.

$$P_{xx}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n)e^{-j2\pi fn} \right|^2 = \frac{1}{N} |X(f)|^2 \quad (2.4)$$

kde  $X(f)$  je Fourierova transformace vzorku signálu  $x(n)$ .

## Bartlettova metoda

Tato metoda pro redukci odchylek v periodogramu využívá tři kroky [3]. Nejdříve signál o  $N$  vzorcích rozdělí na  $K$  nepřekrývajících se segmentů, které mají všechny stejnou délku  $M$ .

$$x_i(n) = x(n+iM), \quad \begin{array}{l} i = 0, 1, \dots, K-1 \\ n = 0, 1, \dots, M-1 \end{array} \quad (2.5)$$

Pro každý segment je vypočten periodogram

$$P_{xx}^{(i)}(f) = \frac{1}{M} \left| \sum_{n=0}^{M-1} x_i(n)e^{-j2\pi fn} \right|^2, \quad i = 0, 1, \dots, K-1 \quad (2.6)$$

Nakonec zprůměrujeme periodogramy všech  $K$  segmentů a získáme výsledné spektrum hustoty výkonu

$$P_{xx}^B(f) = \frac{1}{K} \sum_{i=0}^{K-1} P_{xx}^{(i)}(f) \quad (2.7)$$

## Welchovův periodogram

Tato metoda rozšiřuje Bartlettovu metodu o dvě modifikace [3]. Původní signál je opět rozdělen na segmenty, ale ty se nyní mohou překrývat. Data segmentu mohou být reprezentována následovně

$$x_i(n) = x(n + iD), \quad \begin{array}{l} n = 0, 1, \dots, M - 1 \\ i = 0, 1, \dots, L - 1 \end{array} \quad (2.8)$$

kde  $iD$  je počátek  $i$ -té sekvence. Všimněte si, že pokud je  $D = M$ , tak se segmenty nepřekrývají a  $L$  odpovídá  $M$  v Bartlettově metodě. Pokud  $D = L/2$ , je překryv 50% mezi následujícími segmenty.

Druhou modifikací je váhování vzorků v každém segmentu oknem ještě před výpočtem periodogramu,

$$\tilde{P}_{xx}^{(i)}(f) = \frac{1}{MU} \left| \sum_{n=0}^{M-1} x_i(n)w(n)e^{-j2\pi fn} \right|^2, \quad i = 0, 1, \dots, L - 1 \quad (2.9)$$

kde  $U$  je normalizační faktor okna

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n) \quad (2.10)$$

Výsledné spektrum je dáno průměrem těchto modifikovaných periodogramů

$$P_{xx}^W(f) = \frac{1}{L} \sum_{n=0}^{L-1} \tilde{P}_{xx}^{(i)}(f) \quad (2.11)$$

## 2.2.2 Parametrické metody odhadu spektra

Hlavním limitem neparametrických metod je předpoklad, že autokorelace  $r_{xx}(m)$  je nula pro  $|m| \geq N$ . Tento předpoklad silně omezuje frekvenční rozlišení a kvalitu spektra. Dalším předpokladem je, že se data opakují s periodou  $N$ . Ani jeden z těchto předpokladů není reálný [3] a právě parametrické metody těchto nereálných předpokladů nevyužívají. Pro využití parametrických metod je potřeba znát informace, jak byla data vygenerována. Pak můžeme vytvořit autoregresní (dále pouze AR) model s několika parametry, které jsou odhadnuty ze zkoumaných dat. Z AR modelu a odhadnutých parametrů je vypočtena spektrální výkonová hustota. [3]

### AR model

Pomocí parametrů AR modelu lze provést metody odhad spektra signálu. AR model nám udává závislost aktuálního vzorku na hodnotách předchozích vzorků. Pokud

zpoždění o jeden vzorek budeme značit  $q^{-1}$  a obdobně zpoždění o  $k$  vzorků  $q^{-k}$ , pak závislost aktuálního vzorku na předchozích můžeme popsat vztahem [8]

$$(1 - a_1q^{-1} - a_2q^{-2} - \dots - a_Mq^{-M})y_t = e_t \quad (2.12)$$

kde  $M$  je řád modelu,  $a_1, a_2, \dots, a_M$  jsou parametry modelu,  $y_t$  je aktuální vzorek a  $e_t$  je náhodná chyba typu nekorelovaného šumu. Právě tato rovnice nám popisuje AR model.

Pokud známe parametry AR modelu, můžeme z nich vypočítat spektrum následovně

$$P(f) = \frac{2\sigma^2T}{\left|1 - \sum_{m=1}^M a_m e^{-j2\pi fm}\right|^2} \quad (2.13)$$

kde  $2\sigma^2T$  odpovídá spektrální výkonové hustotě bílého šumu.

### Yule-Walkerova metoda

Yule-Walkerova metoda využívá pro odhad parametrů AR modelu autokorelace. Autokorelace je odhadnuta z dat a vyřešením následujících rovnic v maticovém zápisu pomocí Levinson-Durbinova algoritmu získáme parametry AR modelu.[3]

$$\begin{bmatrix} \gamma_{xx}(0) & \gamma_{xx}(-1) & \cdots & \gamma_{xx}(-p+1) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \cdots & \gamma_{xx}(-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{xx}(p-1) & \gamma_{xx}(p-2) & \cdots & \gamma_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} \gamma_{xx}(1) \\ \gamma_{xx}(2) \\ \vdots \\ \gamma_{xx}(p) \end{bmatrix} \quad (2.14)$$

kde  $\gamma_{xx}(m)$  můžeme nahradit vyváženým autokorelačním odhadem

$$r_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x^*(n)x(n+m), \quad m \geq 0 \quad (2.15)$$

Tím zajistíme, že autokorelační matice bude pozitivně semidefinitní a výsledný AR model bude stabilní.

Odpovídající spektrum získáme následovně

$$P_{xx}^{YW}(f) = \frac{\hat{\sigma}_{wp}^2}{\left|1 + \sum_{k=1}^p \hat{a}_p(k)e^{-j2\pi fk}\right|^2} \quad (2.16)$$

kde  $\hat{a}_p(k)$  jsou odhadnuté parametry AR modelu a

$$\hat{\sigma}_{wp}^2 = \hat{E}_p^f = r_{xx}(0) \prod_{k=1}^p [1 - |\hat{a}_p(k)|^2] \quad (2.17)$$

je odhadnutá minimální střední kvadratická hodnota pro  $p$ -tou predikci.

## Burgova metoda

Burgova metoda odhadu AR parametrů je založená na minimalizaci dopředné a zpětné chyby lineární predikce s tím, že AR parametry musí splňovat Levinson-Durbinovu rekurzi. Předpokládejme data  $x(n), n = 0, 1, \dots, N-1$  a berme v potaz odhad dopředné a zpětné lineární predikce řádu  $m$ , který je definován

$$\begin{aligned}\hat{x}(n) &= -\sum_{k=1}^m a_m(k)x(n-k) \\ \hat{x}(n-m) &= -\sum_{k=1}^m a_m^*(k)x(n+k-m)\end{aligned}\quad (2.18)$$

a odpovídající dopřednou a zpětnou chybu  $f_m(n)$  a  $g_m(n)$  definovanou jako  $f_m(n) = x(n) - \hat{x}(n)$  a  $g_m(n) = x(n-m) - \hat{x}(n-m)$ , kde  $a_m^*(k), 0 \leq k \leq m-1, m-1, 2, \dots, p$  jsou koeficienty predikce. Chyba nejmenších čtverců je

$$\mathcal{E}_m = \sum_{n=m}^{N-1} [|f_m(n)|^2 + |g_m(n)|^2] \quad (2.19)$$

Tato chyba je minimalizována výběrem predikčních koeficientů, které splňují Levinson-Durbinovu rekurzi danou vztahem

$$a_m(k) = a_{m-1}(k) + K_m a_{m-1}^*(m-k), \quad \begin{array}{l} 1 \leq k \leq m-1 \\ 1 \leq m \leq p \end{array} \quad (2.20)$$

kde  $K_m = a_m(m)$  je  $m$ -tý koeficient odrazu filtru predikce. Minimalizací  $\mathcal{E}_m$  s ohledem na komplexní hodnotu koeficientu odrazu  $K_m$  získáme vztah

$$\hat{K}_m = \frac{-\sum_{n=m}^{N-1} f_{m-1}(n)g_{m-1}^*(n-1)}{\frac{1}{2}[\hat{E}f_{m-1} + \hat{E}b_{m-1}]}, \quad m = 1, \dots, p \quad (2.21)$$

kde  $\hat{E}f_{m-1} + \hat{E}b_{m-1}$  je odhad celkové chyby nejmenších čtverců  $\mathcal{E}_m$ .

Burgův algoritmus tedy nejdříve vypočítá koeficienty odrazu a Levinson-Durbinův algoritmus je použit pro získání parametrů AR modelu. Z odhadu AR parametrů pak získáváme vztah pro odhad spektra

$$P_{xx}^{BU}(f) = \frac{\hat{E}_p}{|1 + \sum_{k=1}^p \hat{a}_p(k)e^{-j2\pi fk}|^2} \quad (2.22)$$

Hlavní výhodou této metody je vysoké frekvenční rozlišení, stabilní AR model a metoda je výpočetně efektivní. Tato kapitola byla vytvořena volným překladem z literatury [3].

## 2.3 Teorie ROC křivek

ROC je zkratka z anglického Receiver Operating Characteristic. Tento typ analýzy vychází ze statistické teorie rozhodování a poprvé byla použita během druhé světové války pro analýzu radarových dat [4]. ROC analýza je často využívána v medicíně. Kromě zdravotnictví lze tuto analýzu použít i v mnoha dalších oborech a stále častěji se používá pro ohodnocení metod, nebo jejich porovnání. Následující podkapitoly vznikly volným překladem článku [4].

### 2.3.1 Základní pojmy

Zavést základní pojmy není úplně snadné, protože téměř každý obor má své zavedené názvosloví. Pro naše potřeby zavedeme následující názvosloví. Nejdříve je potřeba určit, jaké objekty (ať už osoby či neživé předměty) budeme testovat (pozorovat). Soubor objektů nám tvoří vzorek. Dalším ze základních pojmů je tzv. selhání, tj. stav nebo vlastnost, kterou u objektů pozorujeme, resp. testujeme (např. mechanická závada palivového článku, účast ve volbách, onemocnění). Dalším důležitým pojmem je tzv. diagnóza (predikce), tj. námi odhadnutý stav či vlastnost objektu (predikce může být založena např. na matematických modelech či na předchozím pozorování). Většina testů pouze porovnává, jak moc výsledek testu odpovídá dané diagnóze, která byla predikována. Diagnóza je provedená jedním nebo více testy. Pokud je použit pouze jeden test, je prohlášen za tzv. zlatý standard. Pokud je nalezen nový test, který funguje lépe, stává se novým zlatým standardem.

V článku [4] je uvedeno, že pro ohodnocení testu musí být splněny následující podmínky:

1. Doba testování musí být tak dlouhá, aby po celou dobu měl testovaný objekt pozorovanou vlastnost, nebo naopak, aby po celou dobu vlastnost neměl.
2. Účelem testu je pomoci rozhodnout, zda objekt má, nebo nemá specifickou vlastnost.
3. Objekty, které mají větší pravděpodobnost pozitivního testu mají s větší pravděpodobností pozorovanou vlastnost.

Velmi důležitý je i výběr vzorku. Existují tři základní způsoby, jak se k této problematice postavit [4]. Máme výběr tzv.:

- *Přirozený*

Výběr je proveden z  $N_0$  objektů, které jsou reprezentativní skupinou. Každý je diagnostikován a je na něm proveden test. Tento typ výběru je přirozený, ale v praxi někdy těžko realizovatelný.

- *Retrospektivní*

Pokud je proces testování drahý, nebo riskantní, je vybráno  $N_0$  objektů a

u všech je provedena diagnostika. Z nich je náhodně vybráno  $N_1$  objektů s pozitivní diagnózou a  $N_2$  objektů s negativní diagnózou. Pouze na těchto  $N_1 + N_2$  objektech je proveden test.

- *Prospektivní*

Používá se, pokud je  $N_0$  objektů vybráno a otestováno. Z nich je náhodně vybráno  $N_1$  objektů s pozitivním testem a  $N_2$  objektů s negativním testem. Pouze těchto  $N_1 + N_2$  objektů je diagnostikováno.

Pokud máme vybranou skupinu objektů a provedeme na nich diagnózu a testy, můžeme na nich provést základní statistické zpracování. Nechť  $p_i$  je pravděpodobnost, že objekt  $i$  bude mít pozitivní diagnózu a  $q_i$  je pravděpodobnost, že objekt  $i$  bude mít pozitivní test. Potom  $P = \text{mean}(p_i)$  je pravděpodobnost pozitivní diagnostiky reprezentativní skupiny. Obdobně  $Q = \text{mean}(q_i)$  je pravděpodobnost pozitivního testu reprezentativní skupiny. Také definujeme  $P' = 1 - P$  a  $Q' = 1 - Q$ . Vztahy těchto pravděpodobností jsou zobrazeny na obr. 2.2.

	Test result		
	Positive	Negative	
Diagnosis Positive	TP	FN	$P$
Negative	FP	TN	$P'$
	$Q$	$Q'$	1

Obr. 2.2: Vztahy mezi pravděpodobnostmi pozitivní diagnostiky a testu. [4]

Z obr. 2.2 je patrné, že mohou nastat pouze čtyři možnosti výsledku. Z nich dvě rozhodnutí jsou chybná. Pokud je diagnostika i test pozitivní, jedná se o tzv. *true positive* ( $TP$ ). Pravděpodobnost  $TP$  se spočítá jako počet  $TP$  vzorků podělených celkovým počtem vzorků. Obdobně je definován tzv. *true negative* ( $TN$ ), kde diagnostika i test jsou negativní. Chybná rozhodnutí jsou tzv. *false negative* ( $FN$ ) a tzv. *false positive* ( $FP$ ). Tyto pojmy se používají při hodnocení kvality testů. Dalším důležitým pojmem je *senzitivita*  $SE$  testu, která nám udává pravděpodobnost pozitivních testů mezi pozitivními diagnostikami.

$$SE = TP / (TP + FN) = TP / P \quad (2.23)$$

*Specificita*  $SP$  testu je pravděpodobnost negativních testů mezi negativními diagnostikami.



$$SP = TN/(FP + TN) = TN/P' \quad (2.24)$$

*Efektivita* EFF je dána vztahem  $EFF = TP + TN$ . Kromě těchto třech základních vztahů se můžete setkat s pojmy předpoklad pozitivního testu (predictive value of a positive test)  $PVP = TP/(TP + FP) = TP/Q$  a předpoklad negativního testu (predictive value of a negative test)  $PVN = TN/(TN + FN) = TN/Q'$ .

Hodnoty směrodatné odchylky pro  $P$ ,  $Q$ ,  $TP$ ,  $FN$ ,  $FP$ ,  $TN$  a  $EFF$  mohou být vypočteny pomocí následujícího vztahu

$$SERR(\widehat{X}) = \sqrt{(\widehat{X}(1 - \widehat{X}))/N_0} \quad (2.25)$$

kde  $X$  je jedna ze zmiňovaných hodnot. Pro odhad směrodatné odchylky můžeme zaměnit  $X$  za  $\widehat{X}$ . U senzitivity a specificity je to obtížnější. Pro jejich odhady platí

$$\widehat{SE} = \widehat{TP}/\widehat{P}_a \widehat{SP} = \widehat{TN}/\widehat{P}'$$

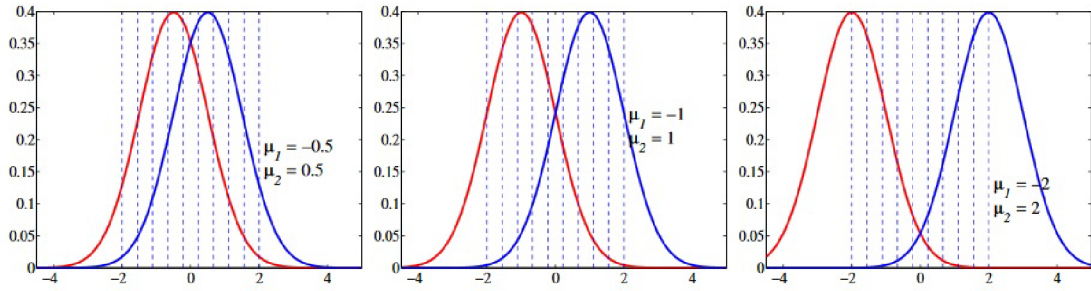
Tyto odhady ale nejsou objektivní. Objektivita záleží jak na počtu vzorků  $N_0$ , tak na  $P$ . S rostoucím počtem vzorků tedy roste objektivita. Pokud je počet vzorků  $N_0$  dostatečně velký, lze pro odhad směrodatné odchylky použít následující vztahy

$$\begin{aligned} SERR(\widehat{SE}) &= \sqrt{(SE(1 - SE))/(N_0 \cdot P)} \\ SERR(\widehat{SP}) &= \sqrt{(SP(1 - SP))/(N_0 \cdot (1 - P))} \end{aligned} \quad (2.26)$$

### 2.3.2 ROC křivky

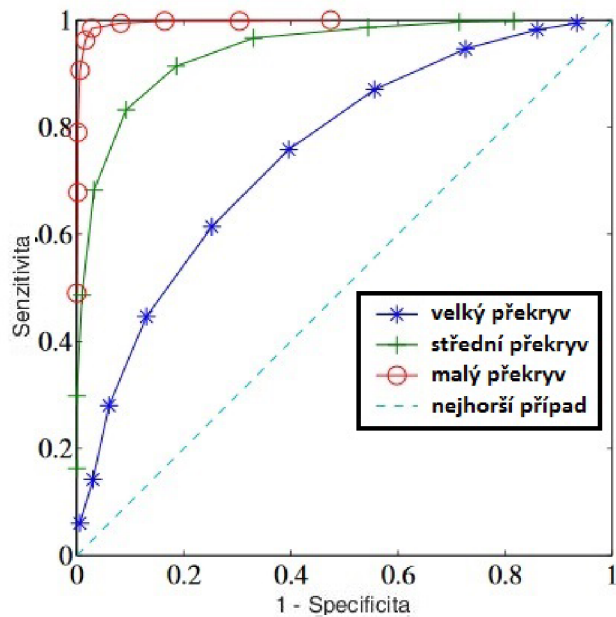
Výstup ze systému pro podporu rozhodování je v intervalu  $[0, 1]$ , kde 0 označuje negativní a 1 pozitivní diagnostiku. Zavedením rozhodovací úrovně (práhu) v intervalu se převede výstup na binární a je porovnán s pravou diagnostikou. Každá rozhodovací úroveň odpovídá bodu na ROC křivce. Na ose  $y$  ROC křivky je vynesena senzitivita a na ose  $x$  je vynesena doplněk specificity ( $1 - SP = 1 - \frac{TN}{P'} = \frac{FP+TN-TN}{FP+TN} = \frac{FP}{FP+TN}$ ). Proto je někdy osa  $x$  nazývaná jako *false positive rate*. Senzitivita a specificita, a proto také výkon systému, se liší použitou rozhodovací úrovní (práhem). Pokud je několik testů provedeno na stejných datech, může být ROC křivka využita pro jejich porovnání. Dalším způsobem, jak lze využít ROC křivky, je zobrazení výkonu podpůrného rozhodovacího systému. Korespondence mezi rozhodovací úrovní a výkonem je ukázána na příkladu.

Předpokládejme výsledek z podpůrného rozhodovacího systému jako dvě normální rozdělení (např. jedna pro zdravého a druhá pro nemocného pacienta). Rozhodovací úroveň je postupně posouvána. Pro každou polohu práhu je vypočtena senzitivita a specificita. Z těchto hodnot je získán bod na ROC křivce (z každé polohy



Obr. 2.3: Tři dvojice normálního rozdělení pravděpodobností pro velký, střední a malý překryv. [4]

rozhodovací úrovně). Křivka nám dává přehled o výkonu systému. Na obrázku 2.3 je ukázka tří případů, kde na každém z nich je deset různých rozhodovacích úrovní. Výsledné ROC křivky pro tři dvojice normálního rozdělení s různým překryvem z obr. 2.3 jsou zobrazeny na obr. 2.4.

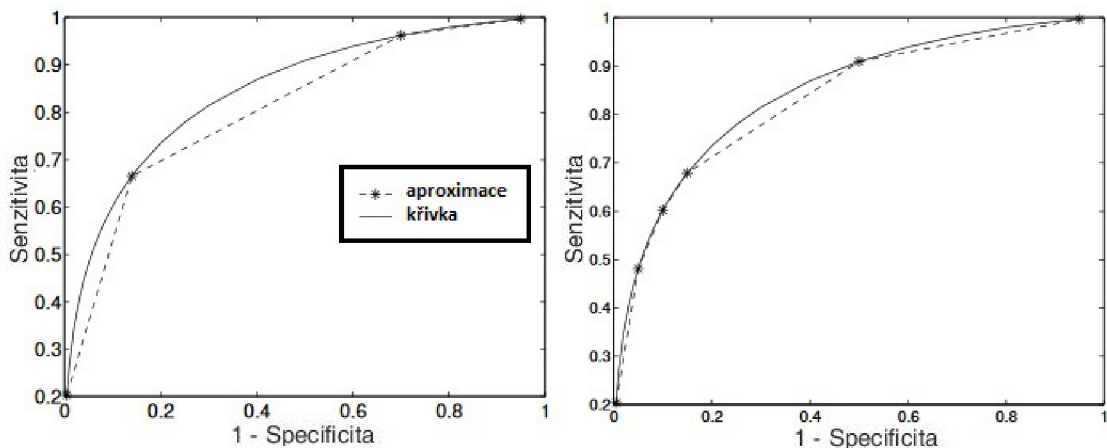


Obr. 2.4: Odpovídající ROC křivky pro tři dvojice normálního rozdělení pravděpodobností pro velký, střední a malý překryv. [4]

### 2.3.3 AUC

AUC je zkratka anglického "area under the curve" a značí plochu pod ROC křivkou. Celková plocha pod ROC křivkou nám udává výkon diagnostiky při všech možných prahových hodnotách. Plocha je v intervalu  $[0.5, 1]$  a čím větší plocha je, tím je lepší výkonnost. Předpokládejme, že vysoká hodnota AUC udává, že diagnostika metody je *positive* (kladná) a nízká hodnota naopak *negative* (negativní). Plocha odpovídá míře pravděpodobnosti, že rozdělení pozitivní diagnostiky je statisticky větší než rozdělení negativní diagnostiky.

Při experimentech máme zpravidla pouze konečnou množinu bodů na ROC křivce. Proto je možné najít pouze aproximaci plochy pod křivkou (příklad na obr. 2.5). Je zřejmé, že čím více bodů na křivce známe, tím přesnější bude aproximace. Existuje několik možností výpočtu AUC. Jedním ze základních je výpočet pomocí lichoběžníkového pravidla. Další možností je aproximace pomocí binormálního modelu s odhadem pomocí maximum-likelihood a z tohoto modelu následně přesněji určit hodnotu AUC. Jiný pohled na výpočet AUC je pomocí Mann-Whitney U test (také známého jako neparametrický Wilcoxonův test).



Obr. 2.5: Ukázka jaký vliv má počet bodů ROC křivky na hodnotu AUC. [4]

#### Lichoběžníkové pravidlo

Lichoběžníkové pravidlo je metoda numerické integrace používaná pro aproximaci integrálu nebo plochy pod křivkou.[5] Integrál na intervalu  $[a, b]$  funkce  $f(x)$  je rovnoměrně rozdělen na  $n$  stejných intervalů. Plochy těchto intervalů jsou nahrazeny lichoběžníky (horní křivka na intervalu je nahrazena tětivou). Součtem dílčích aproximovaných částí získáme výslednou hodnotu AUC.

$$\int_a^b f(x)dx$$

je určitý integrál funkce  $f(x)$ , kterou rozdělíme na rovnoměrné intervaly a body podintervalu označíme  $x_0, x_1, \dots, x_n$ , kde  $x_0 = a, x_n = b, x_r = x_0 + r(b - a)/n$ .

Při aplikaci lichoběžníkového pravidla na ROC křivky je výhodné využít přímo známých bodů ROC křivky, takže jednotlivé podintervaly nebudou stejně široké. Nemá smysl dělit ROC křivku na stejně velké podintervaly a následně určovat hodnotu funkcí z aproximace.

## 3 HODNOCENÍ ÚČINNOSTI KLASIFIKÁTORU PROSTŘEDNICTVÍM ROC KŘIVEK

Jak již bylo uvedeno v teoretické části, ROC křivky jsou ideálním prostředkem pro porovnání účinnosti různých metod aplikovaných na stejných testovacích datech. Tato práce je konkrétně věnována detekci stresu v hlasových signálech. Práce volně navazuje na diplomovou práci [6], která se věnovala odhadu identifikace významných spektrálních složek ve stresovém řečovém signálu. Výstupem této práce byly mimo jiné získány intervaly, ve kterých by se měly pohybovat výkonové spektra stresových a nestresových signálů v oblasti třetího formantu ( $2500 - 3000Hz$ ). Z diplomové práce [6] jsou převzaty nahrané stresové a nestresové audio soubory a předzpracování vstupních signálů.

V této kapitole (včetně podkapitol) jsou uvedeny veškeré problematiky, kterými se práce zabývala. Řeší se zde problematika volby odhadu spektra audio souboru, vstupní rozsah spekter signálů pro ROC křivky (včetně předzpracování). Je také navrhnut klasifikátor na základě ROC křivek. V kapitole 3.5 je částečně nastíněna i problematika zašumění signálů, ale toto téma by rozsahem pokrylo další samostatnou práci. Proto tato problematika zde není podrobně rozebrána, ale jsou zde pouze nastíněny základní problémy z této oblasti. Kapitola 4 je věnována konkrétním řešením programů, které jsou součástí této práce.

### 3.1 Zdrojové data a předzpracování signálů

Abychom mohli provádět analýzy na datech, potřebujeme o nich vědět základní informace. Na základě těchto informací můžeme začít vymýšlet postupy práce s těmito daty. Pro účely této práce bylo převzato 23 audio souborů ve formátu *wav* (stejně jako v práci [6]). Z nich je 12 souborů nestresových a 11 stresových souborů. Pojmenování signálů nalezneme v tabulce 3.1. Všechny audio soubory byly nahrány pomocí stejného zařízení s nastavenou stejnou intenzitou na FEKT VUTBR. Záznamy jsou namluveny pouze muži (studenty) a jednotlivé nahrávky nemají stejnou úroveň hlasitosti. Každý záznam je namluvený jinou osobou. Neplatí tedy např. že nahrávku *signStres1.wav* a *signNestres1.wav* namluvila stejná osoba. Audio nahrávky obsahují různý text a jsou různě dlouhé.

Audio soubory jsou načteny pomocí funkce `wavread` s vzorkovacím kmitočtem  $F_s = 8000Hz$ . Následně je aplikováno předzpracování, při kterém je vstupní signál rozdělen na několik segmentů o délce 160 vzorků. Z těchto segmentů jsou pro další zpracování vybrány pouze segmenty, které jsou pro nás zajímavé (obsahují formanty samohlásek, specifikováno rozsahem frekvencí). Z původního signálu jsou odstraněny

Tab. 3.1: Vstupní data pro testování.

<b>nestresové signály</b>	<b>stresové signály</b>
<i>signNestres1.wav</i>	<i>signStres1.wav</i>
<i>signNestres2.wav</i>	<i>signStres2.wav</i>
<i>signNestres3.wav</i>	<i>signStres3.wav</i>
<i>signNestres4.wav</i>	<i>signStres4.wav</i>
<i>signNestres5.wav</i>	<i>signStres5.wav</i>
<i>signNestres6.wav</i>	<i>signStres6.wav</i>
<i>signNestres7.wav</i>	<i>signStres7.wav</i>
<i>signNestres8.wav</i>	<i>signStres8.wav</i>
<i>signNestres9.wav</i>	<i>signStres9.wav</i>
<i>signNestres10.wav</i>	<i>signStres10.wav</i>
<i>signNestres11.wav</i>	<i>signStres11.wav</i>
<i>signNestres12.wav</i>	

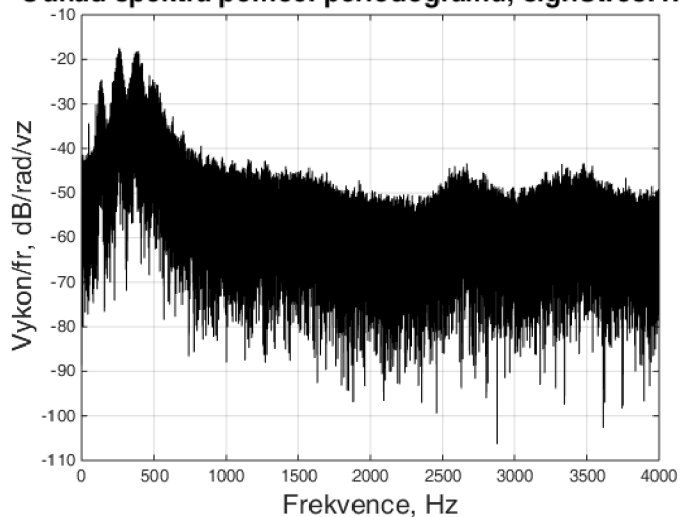
tzv. „prázdné oblasti“ (nezajímavé úseky z pohledu detekce stresu). Algoritmus pro hledání segmentů, které obsahují pro nás zajímavé data, je implementován a popsán v literatuře [9]. Algoritmus využívá LPC-analýzu (Linear Coded Predicting) a sestavení fázové charakteristiky závislé na kořenech získaných v této analýze. Po předzpracování získáme data obsahující pouze segmenty s definovanými rozsahy frekvencí. Signál který projde předzpracováním samozřejmě obsahuje i nějaké množství dat, která jsou pro nás nezajímavá. A to v případě, pokud jsou v rámci jednoho segmentu „zajímavé“ i „prázdné oblasti“. Ale např. úseky, kdy řečník nemluví a je nahráván pouze šum pozadí, budou odstraněny. Z tohoto signálu následně provádíme odhad spektra.

## 3.2 Analýza metod odhadu spektra

Velmi důležitou problematikou detektoru stresu v hlasovém záznamu je odhad spektra, protože detekce probíhá ve frekvenční oblasti. Odhad spektra je proveden jak pomocí neparametrických, tak parametrických metod. Z neparametrických metod je použit Welchův periodogram (kapitola 2.2.1).

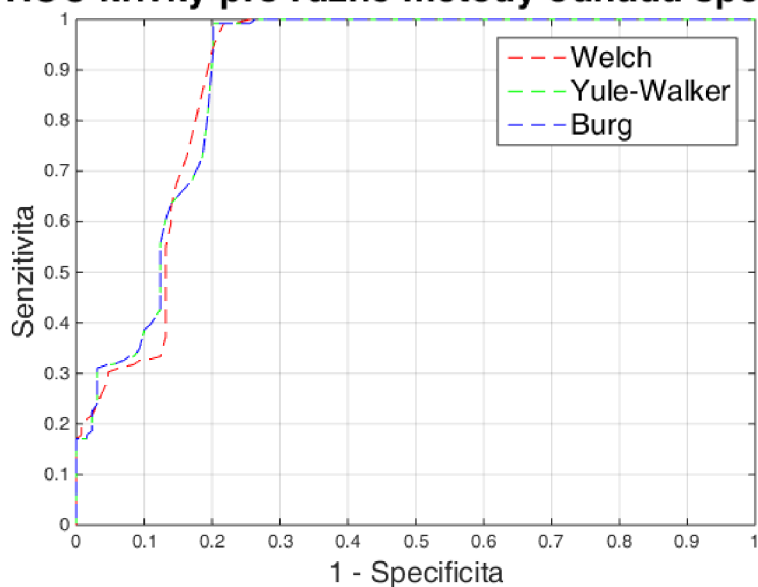
Nemodifikovaný periodogram nebyl použit, protože spektrum získané touto metodou je velmi husté a funkční hodnoty na jednotlivých frekvencích se těžko určují (viz. obrázek 3.1). Z parametrických metod odhadu spektra prostřednictvím odhadu parametrů AR procesu byly použity metody Yule-Walker (kapitola 2.2.2) a Burg (kapitola 2.2.2).

**Odhad spektra pomocí periodogramu, signStres7.wav**



Obr. 3.1: Odhad spektra signálu *signStres7* pomocí periodogramu.

**ROC křivky pro různé metody odhadu spektra**



Obr. 3.2: Tři ROC křivky pro dvojici signálů *signNestres2* a *signStres7* získaných různými metodami odhadu spektra.

Porovnání metod odhadu spektra je provedeno na kombinaci jednoho náhodně vybraného nestresového signálu (konkrétně *signNestres2.wav*) a postupně všech stresových signálů (viz. obrázek 3.3). Ukázky odhadnutých spekter signálů jednotlivými metodami nalezneme v příloze A. Pro všechny kombinace  $signNestres2 \times signStres1-11$  byly vypočteny ROC křivky a následně určeny odpovídající hodnoty AUC.

Výsledky analýzy nalezneme v tabulce 3.2. Z tabulky je na první pohled vidět, že



Obr. 3.3: Grafické znázornění kombinací vstupních signálů pro ROC křivky, na kterých byla analýza pro odhad metody spektra provedena.

AR Yule-Walker a AR Burg jsou stejně výkonné. Tento fakt potvrzují také příslušné ROC křivky (na obrázku 3.2 překrývající se ROC křivky pro AR Yule-Walker a AR Burg). Z tabulky 3.2 můžeme také konstatovat, že metody které odhadují spektrum na základě AR procesu jsou výkonnější. Abychom se mohli rozhodnout, která z metod bude dále využita, byla provedena doplňující analýza výpočetní náročnosti. Na stejné testovací množině bylo provedeno další ohodnocení (výsledky zobrazeny v tabulka 3.3). Na základě výpočetní náročnosti byla zvolena za nejlepší metodu metoda odhadu spektra pomocí AR procesu metodou Yule-Walker, která je ze všech testovaných metod jednoznačně nejrychlejší. Výpočetní náročnost byla měřena pomocí funkcí `tic` a `toc`. Během této analýzy byla měřena i výpočetní náročnost metod odhadu spektra pro signál *signNestres2*. Tato hodnota byla změřena  $10\times$  pro všechny metody a z nich byly vypočteny hodnoty aritmetického průměru,  $\bar{t}_{\text{Welch}} = 3,0153s$ ,  $\bar{t}_{\text{Yule-Walker}} = 0,0965s$  a  $\bar{t}_{\text{Burg}} = 0,6272s$ . Aritmetické průměry částečně eliminují případné náhodné výchylky. Při dalších analýzách bylo využito pouze odhadu spektra na základě odhadu parametrů AR procesu metodou Yule-Walker.

Tab. 3.2: Hodnoty  $AUC$  pro jednotlivé kombinace *signNestres2*  $\times$  *signStres1-11*.

signál	$AUC_{\text{Welch}} [-]$	$AUC_{\text{Yule-Walker}} [-]$	$AUC_{\text{Burg}} [-]$
signStres1	0,9376	0,9434	0,9434
signStres2	0,9106	0,9119	0,9119
signStres3	0,8885	0,8897	0,8897
signStres4	0,9133	0,9170	0,9170
signStres5	0,8734	0,8724	0,8724
signStres6	0,9101	0,9153	0,9153
signStres7	0,8885	0,8983	0,8983
signStres8	0,9578	0,9633	0,9633
signStres9	0,9085	0,9091	0,9091
signStres10	0,9122	0,9195	0,9195
signStres11	0,9106	0,9119	0,9119



Tab. 3.3: Výpočetní náročnost metod odhadu spektra.

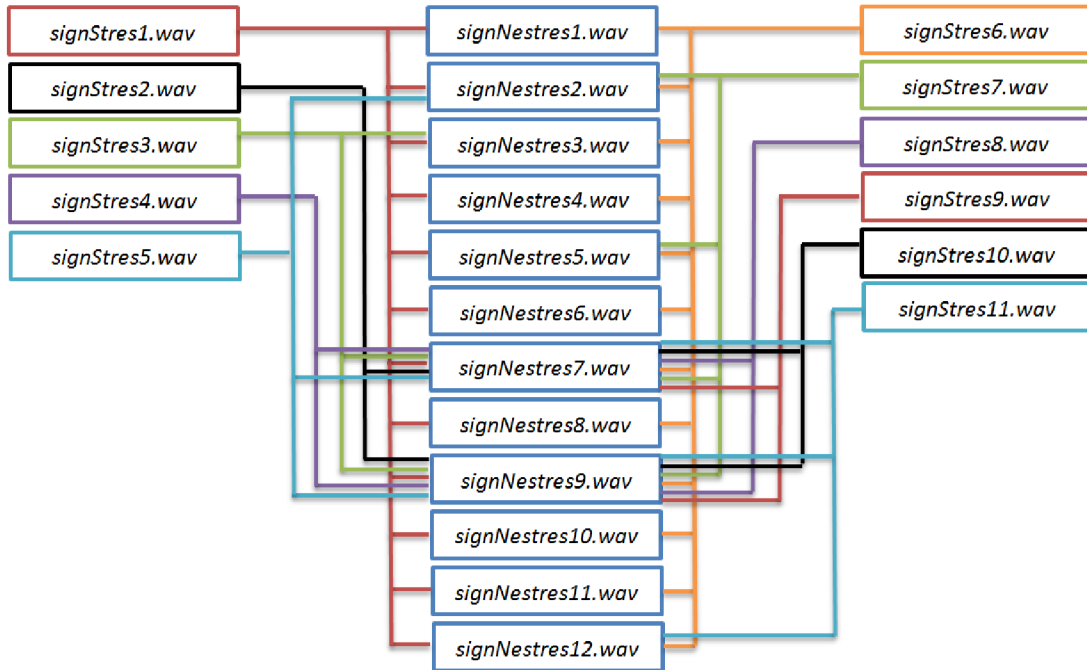
signál	$t_{\text{Welch}}$ [ s ]	$t_{\text{Yule-Walker}}$ [ s ]	$t_{\text{Burg}}$ [ s ]
signStres1	6,5552	0,2982	1,4883
signStres2	13,0374	0,6668	2,8240
signStres3	1,6843	0,0770	0,3244
signStres4	9,1379	0,2947	1,9125
signStres5	1,6625	0,0689	0,3238
signStres6	14,1621	0,6389	2,9979
signStres7	1,6698	0,0771	0,3265
signStres8	3,9231	0,1040	0,8304
signStres9	9,3326	0,2976	1,9803
signStres10	7,7342	0,2744	1,6858
signStres11	13,0311	0,6533	2,7507

### 3.3 Analýza vlivu frekvenčního rozsahu spektra vstupních signálů

Po zvolení metody odhadu spektra se můžeme zabývat problematikou vstupních dat pro ROC křivky. V teorii bylo napsáno, že detekci budeme provádět ve frekvenční oblasti. Je ale potřeba určit, jak velký rozsah frekvencí bude mít vstupní spektrum. Analýzu lze popsat v několika krocích:

1. Odhad spektra pomocí AR Yule-Walker (stresový a nestresový signál).
2. Výřez spektra podle rozsahů frekvencí v tabulce 3.4 (stresový a nestresový signál).
3. Výpočet ROC a AUC

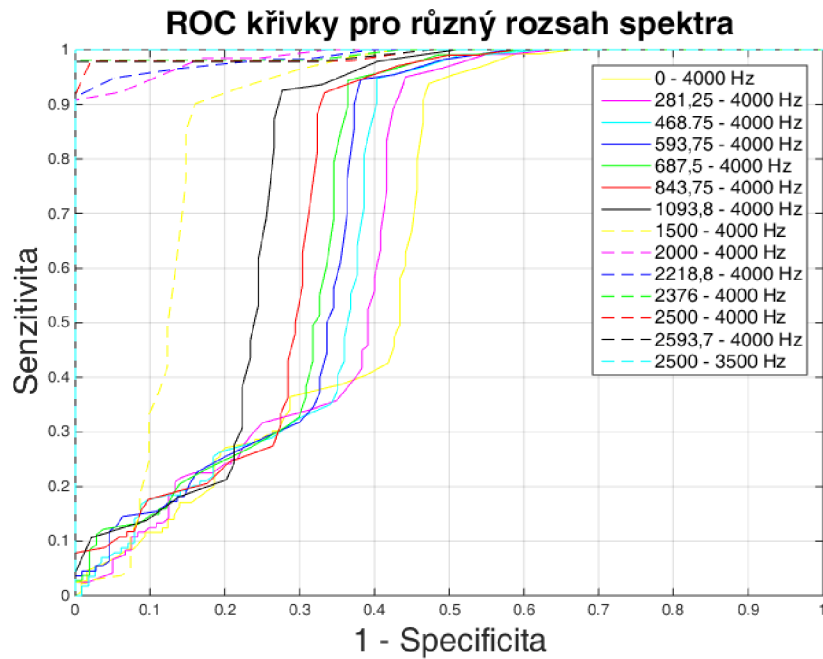
Bylo testováno 13 různých frekvenčních rozsahů spekter vstupních signálů. Jako vstupní množinu bylo náhodně vybráno 47 různých kombinací stresového a nestresového signálu (viz. obrázek 3.4). Pro všechny tyto dvojice byly vypočteny ROC křivky pro měnící se rozsah vstupního spektra signálů (celkem tedy 523 simulací). S ohledem na velikost vstupního spektra bylo provedeno i předzpracování signálů, které vyhledává v signálu úseky obsahující tyto frekvence (viz. kapitola 3.1 věnovaná předzpracování signálů). Ukázkou ROC křivek pro jednu kombinaci stresového a nestresového signálu (*signStres6*, *signNestres1*) můžeme vidět na obrázku 3.5. Hodnoty AUC pro tyto kombinace jsou v tabulce 3.4. Frekvenční rozsahy spekter vstupních dat byly voleny s ohledem na výskyt formantů. Z této tabulky 3.4 i ROC křivek (obrázek 3.5) je pěkně vidět, že čím užší spektrum vstupních signálů je, tím výkonnější je klasifikátor. Nejlepších výsledků bylo dosaženo u užších rozsahů spektra, protože



Obr. 3.4: Grafické znázornění kombinací vstupních signálů pro ROC křivky, na kterých byla provedena analýza pro vliv frekvenčního rozsahu spektra vstupních signálů. Jako vstupní data je vždy použit jeden stresový a jeden nestresový signál.

v těchto případech vstupní spektrum kopíruje teoretické rozložení třetích formantů samohlásek. Ve spektru je tedy nejméně složek, které pro nás nejsou významné. Během testování byla  $AUC$  hodnota několikrát rovna 1, což značí ideální klasifikátor. Rozhodovací úroveň (práh) rozděluje jednoznačně vstupní data (není žádný překryv dat z pohledu rozhodovací úrovně). Z testovaných rozsahů během provedených analýz dosahoval nejlepších výsledků rozsah  $2500 - 3500 Hz$ . Tento rozsah potvrzuje teoretické předpoklady o třetích formantech samohlásek (viz. kapitola 2.1). V dalších analýzách budeme vycházet z těchto poznatků a budeme dále pracovat se spektry signálů v tomto rozsahu.

U některých kombinací vstupních signálů dochází při velkém rozsahu spektra k chybným průběhům ROC křivek. Část ROC křivky se nacházela pod úsečkou definovanou body  $[0, 0]$ ,  $[1, 1]$ . Tento případ by neměl nastat. Ale jelikož v této části spektra nejsou významné složky samohlásek, na kterých je klasifikátor postaven, je zde spektrum naprosto náhodné z pohledu klasifikátoru. „Nezajímavá“ část spektra převažuje, takže může dojít k těmto případům ROC křivek (obrázek 3.6). Eliminací této části „nezajímavého“ spektra z pohledu klasifikátoru stresu získáme očekávaný průběh ROC křivky. Tedy i kvalitnější klasifikátor. Pokud by tento průběh měly všechny ROC křivky, stačilo by pouze otočit rozhodovací kritérium na této části

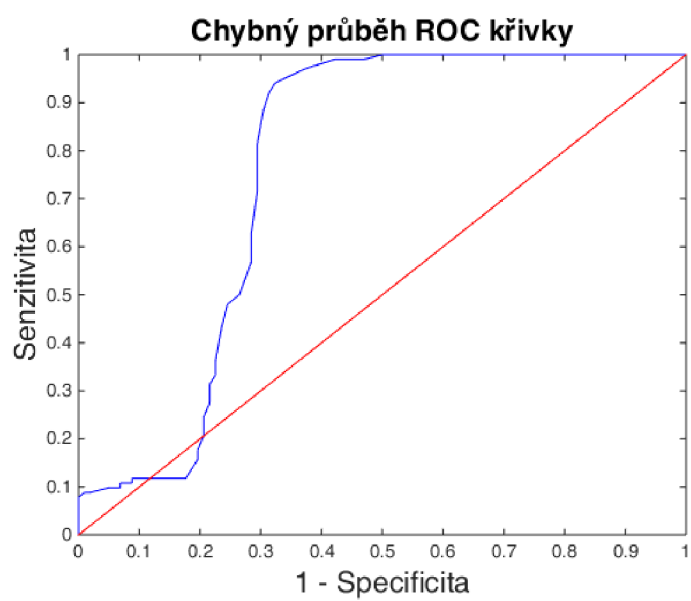


Obr. 3.5: ROC křivky pro dvojici signálů *signStres6* a *signNestres1* s různou velikostí vstupního spektra.

Tab. 3.4: Závislost *AUC* na velikosti spektra vstupních signálů.

velikost spektra [Hz]	<i>AUC</i> [ - ]
0 - 4000	0,6589
281,25 - 4000	0,6826
468,75 - 4000	0,6913
593,75 - 4000	0,7167
687,5 - 4000	0,7290
843,75 - 4000	0,7456
1093,8 - 4000	0,7853
1500 - 4000	0,8704
2000 - 4000	0,9879
2218,8 - 4000	0,9880
2376 - 4000	0,9929
2500 - 4000	0,9908
2593,7 - 4000	0,9913
2500 - 3500	1,0000

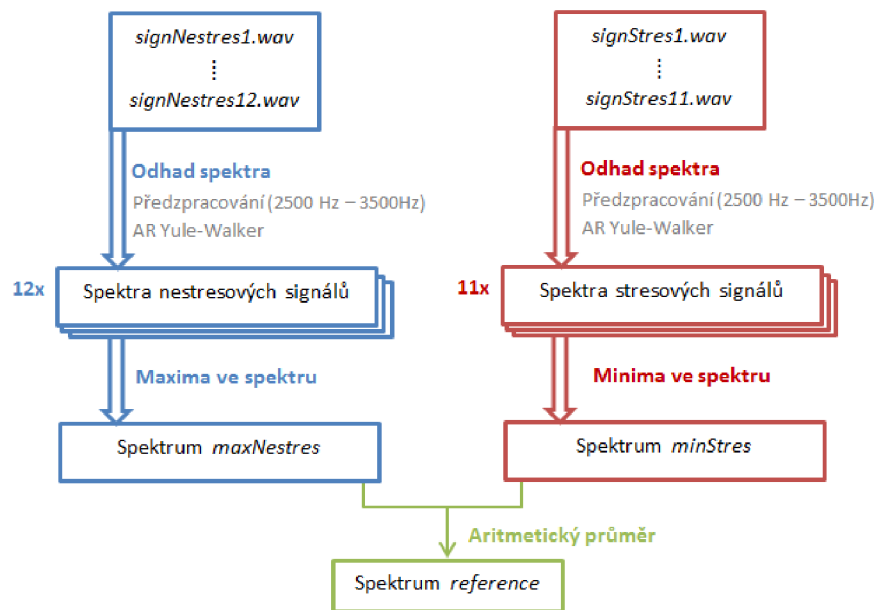
spektra. Ale jelikož je tento jev náhodný, tak je potřeba tuto část spektra odstranit ze vstupu.



Obr. 3.6: Chybný průběh ROC křivky (modrá čára). Tato křivka by se neměla nacházet na žádném úseku pod červenou čarou.

### 3.4 Klasifikace pomocí ROC křivek

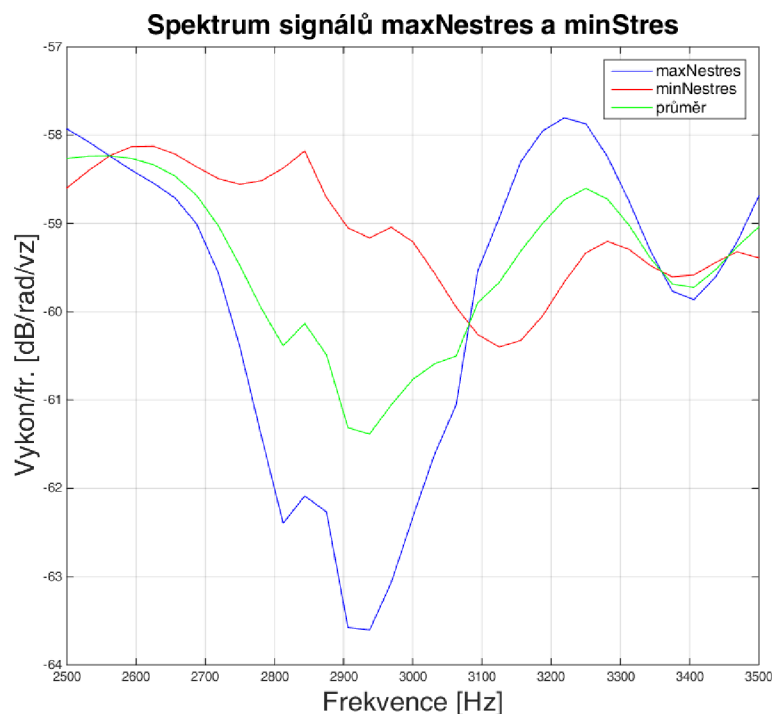
Jednou z možností využití ROC křivek je ohodnocování výkonu. Můžeme tedy vytvořit detektor stresu v hlasu, který bude rozhodovat na základě tvaru ROC křivky (AUC) zda je signál stresový. V diplomové práci [6] byl mimo jiné vytvořen min, max interval pro nestresové a stresové signály ve frekvenční oblasti. Na základě poznatků o těchto intervalech bychom měli být schopni detekovat stres v signálech na základě úrovně výkonu ve spektru. Stresové signály mají ve spektru většinou vyšší úroveň výkonu v naší oblasti zájmu (frekvence od  $2500\text{Hz}$  do  $3500\text{Hz}$ ).



Obr. 3.7: Grafické znázornění, jak bylo získáno referenční spektrum pro ROC křivky *reference*.

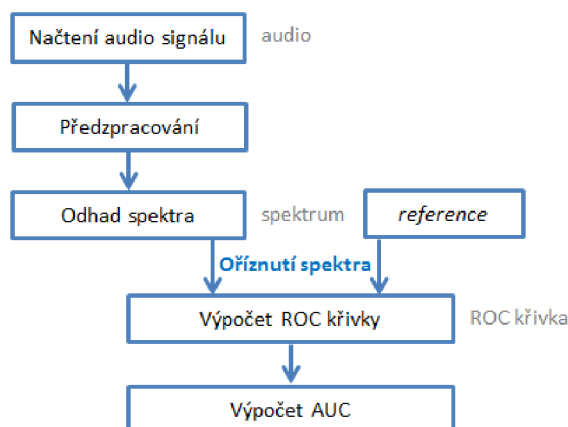
Abychom mohli provádět klasifikaci stresový/nestresový signál, potřebujeme nějaký referenční signál. Tento referenční signál by měl být tzv. prahem (rozhodovací úrovní) mezi stresovými a nestresovými signály. Jelikož provádíme detekci ve frekvenční oblasti, tak referenční signál bude spektrum. Toto spektrum získáme pomocí dvou pomocných spekter. Nejdříve ze všech nestresových signálů odhadneme jejich spektra (*signNestres1.wav* až *signNestres12.wav*). Při odhadu spektra využijeme poznatků z předchozích kapitol, takže využijeme předzpracování s vyhledáváním frekvencí  $2500\text{Hz} - 3500\text{Hz}$  a spektrum odhadneme pomocí AR Yule-Walker. První pomocné spektrum pak získáme jako maximum ze všech nestresových spekter na jednotlivých frekvencích (dále toto spektrum bude nazýváno pouze jako *maxNestres*). Druhé pomocné spektrum získáme obdobně ze stresových signálů (*signStres1.wav*

až *signStres11.wav*). Jednotlivá spektra stresových signálů získáme stejně jako u nestresových (předzpracování  $2500\text{Hz} - 3500\text{Hz}$ , AR Yule-Walker). Druhé pomocné spektrum získáme jako minimum ze všech stresových spekter na jednotlivých frekvencích (dále toto spektrum bude nazýváno pouze jako *minStres*). Referenční spektrum (dále toto spektrum bude nazýváno pouze jako *reference*) pro ROC křivky získáme jako průměr z těchto pomocných spekter (*maxNestres* a *minStres*). Postup získání *reference* je graficky znázorněno na obrázku 3.7. Na obrázku 3.8 je zobrazeno spektrum *maxNestres*, *minStres* a referenční spektrum pro ROC křivky *reference*. Z obrázku 3.8 je také vidět, že spektra stresových a nestresových signálů se prolínají, ale zde jsou zobrazeny nejhorší možné situace a při testování klasifikátor podával dobré výsledky.



Obr. 3.8: Spektra signálů *maxNestres*, *minStres* a *reference*.

Detekce stresu (klasifikace stresový/nestresový signál) je provedena pomocí ROC křivky, která má jako vstupní data spektrum analyzovaného signálu a spektrum *reference*. Z pohledu uživatele je tedy požadován pouze signál, který chceme klasifikovat (spektrum *reference* zůstává při všech analýzách). Tvar ROC křivek nám klasifikuje o jaký signál se jedná. Pokud se ROC křivka nachází nad diagonálou, signál je pravděpodobněji stresový. Naopak pokud se ROC křivka nachází pod diagonálou, jedná se pravděpodobně o nestresový signál. Hodnota AUC jednotlivých ROC křivek přímo odpovídá pravděpodobnosti, s jakou je signál stresový.

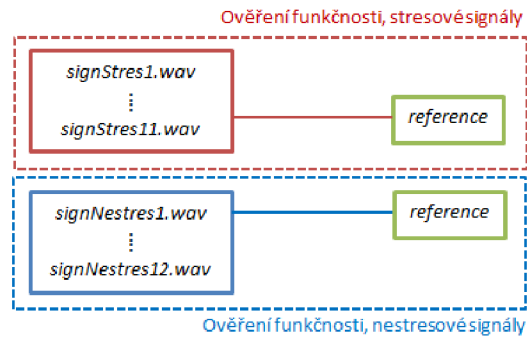


Obr. 3.9: Vývojový diagram klasifikátoru.

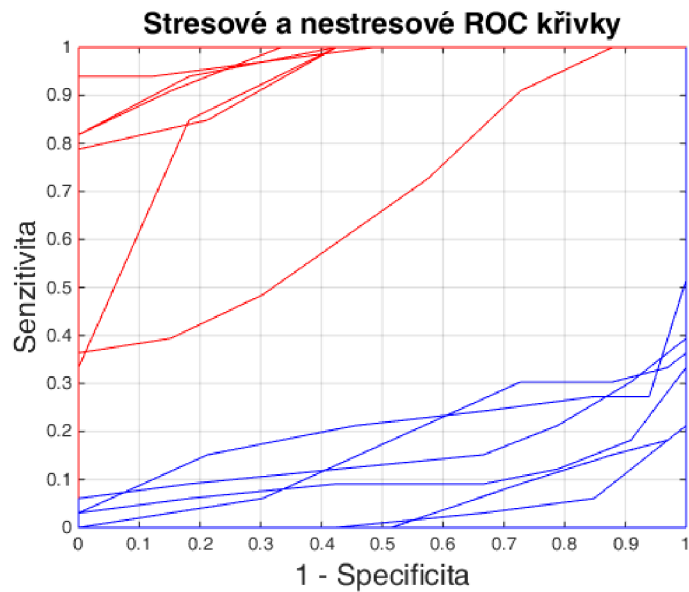
Hodnota AUC se nabízí jako vhodná veličina pro samotnou klasifikaci. Pro nastavení kritéria AUC je potřeba vzít v potaz následující fakta. V ideálním případě by celé spektrum stresového signálu na rozsahu frekvencí  $2500\text{Hz} - 3500\text{Hz}$  mělo větší hodnoty výkonu, takže by nám vyšla  $AUC = 1$ . V reálu tato podmínka ale splněna není, takže je potřeba určit hodnotu, kdy můžeme prohlásit signál za stresový, nebo nestresový. Pro stresové signály bylo určeno kritérium  $AUC > 0,8$ . Je potřeba brát v potaz, že u některých signálů může dojít k překryvu části spektra. Zároveň ale nesmí mít kritérium moc nízkou hodnotu, aby nebyly spatně klasifikovány nestresové signály. Pro nestresové signály bylo navrženo kritérium  $AUC < 0,5$ . Signály s  $0,5 < AUC < 0,8$  jsou potenciální stresové signály. Ale v této oblasti se mohou nacházet i nestresové signály, jelikož kritérium bylo vytvořeno a testováno na relativně malé množině stresových a nestresových signálů. Klasifikátor byl otestován všemi stresovými i nestresovými signály. Na vstupní data byla aplikována všechna doporučení z předchozích kapitol. ROC křivky pracují se spektry signálů (analyzované, referenční), takže vstupní audio signál je nejdříve předzpracován (vyhledávání třetích formantů) a následně je pomocí AR Yule-Walker odhadnuto jeho spektrum. Jako vstupní data pro ROC křivku slouží takto získaná spektra z audio nahrávek v rozsahu frekvencí  $2500\text{Hz} - 3500\text{Hz}$  spolu se spektrem *reference*.

Vývojový diagram klasifikátoru je zobrazen na obrázku 4.1. Klasifikátor byl otestován na všech stresových i nestresových signálech (viz. obrázek 3.10, jednotlivé ROC křivky jsou zobrazeny na obrázku 3.11), Výsledné hodnoty AUC jednotlivých signálů jsou zobrazeny v tabulce 3.5 a 3.6.

Na základě ROC křivek zobrazených na obrázku 3.11 a hodnotách AUC zobrazených v tabulce 3.5 a 3.6 vidíme, že všechny nestresové signály jsou správně klasifikovány ( $AUC < 0,5$ ). Ze stresových signálů je pouze signál *signStres9.wav*



Obr. 3.10: Grafické znázornění dat, na kterých byl klasifikátor otestován.



Obr. 3.11: ROC křivky testovaných dat. Červené ROC křivky jsou nestresových signálů a modré ROC křivky jsou stresových signálů.

klasifikován jako potenciálně stresový ( $0,5 < AUC < 0,8$ ). Zbýlých deset stresových signálů bylo správně klasifikováno jako stresové ( $AUC > 0,8$ ).



Tab. 3.5: Hodnoty  $AUC$  jednotlivých ROC křivek stresových signálů.

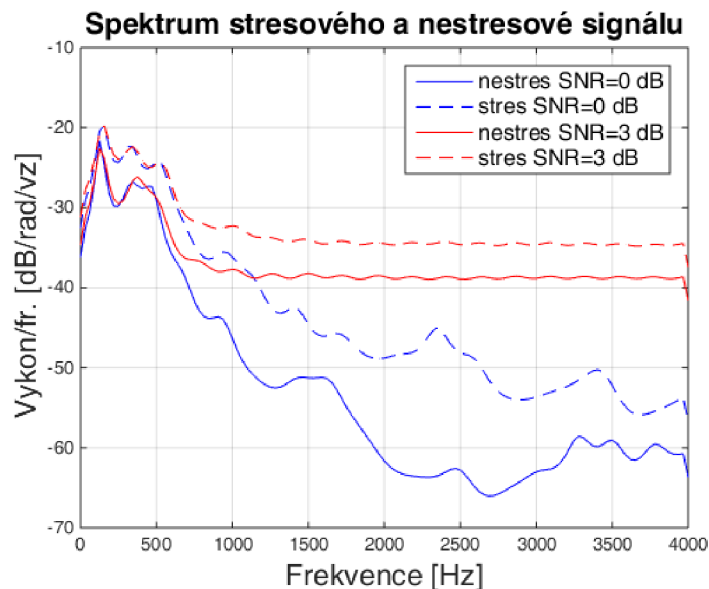
<b>analyzovaný signál</b>	$AUC$ [ - ]
<i>signStres1.wav</i>	1,0000
<i>signStres2.wav</i>	0,9711
<i>signStres3.wav</i>	0,9454
<i>signStres4.wav</i>	0,9706
<i>signStres5.wav</i>	0,9073
<i>signStres6.wav</i>	1,0000
<i>signStres7.wav</i>	0,9454
<i>signStres8.wav</i>	1,0000
<i>signStres9.wav</i>	0,6791
<i>signStres10.wav</i>	0,9816
<i>signStres11.wav</i>	0,9711

Tab. 3.6: Hodnoty  $AUC$  jednotlivých ROC křivek nestresových signálů.

<b>analyzovaný signál</b>	$AUC$ [ - ]
<i>signNestres1.wav</i>	0,1717
<i>signNestres2.wav</i>	0,0000
<i>signNestres3.wav</i>	0,0000
<i>signNestres4.wav</i>	0,0000
<i>signNestres5.wav</i>	0,0491
<i>signNestres6.wav</i>	0,0326
<i>signNestres7.wav</i>	0,2071
<i>signNestres8.wav</i>	0,0000
<i>signNestres9.wav</i>	0,0000
<i>signNestres10.wav</i>	0,1575
<i>signNestres11.wav</i>	0,0000
<i>signNestres12.wav</i>	0,0000

### 3.5 Vliv zašumění signálů na detekci

Zašumění vstupních signálů má na klasifikátor velký vliv. Klasifikátor rozhoduje zda je audio signál stresový na základě úrovně spektra. Nestresové signály mají v rozsahu frekvencí 2500 – 3500 Hz ve spektru nižší úroveň výkonu než signály stresové. Tento fakt nám umožňuje detekovat stres. Klasifikace je tedy možná pouze v případě, že si nejdříve vytvoříme nestresový referenční signál systému a za předpokladu, že charakter zašumění se výrazně nemění. Vliv zašumění byl v práci simulován přidáním bílého gaussovského šumu ke zdrojovým audio souborům pomocí funkce `awgn`. Pokud ke stresovému i nestresovému signálu přidáme „stejný“ bílý gaussovský šum (stejnou hodnotu SNR, poměr signál šum), tak jsme schopni nadále detekovat stres, protože spektrum stresového a nestresového signálu v naší oblasti zájmu bude mít opět rozdílné úrovně výkonu (viz. obrázek 3.12).

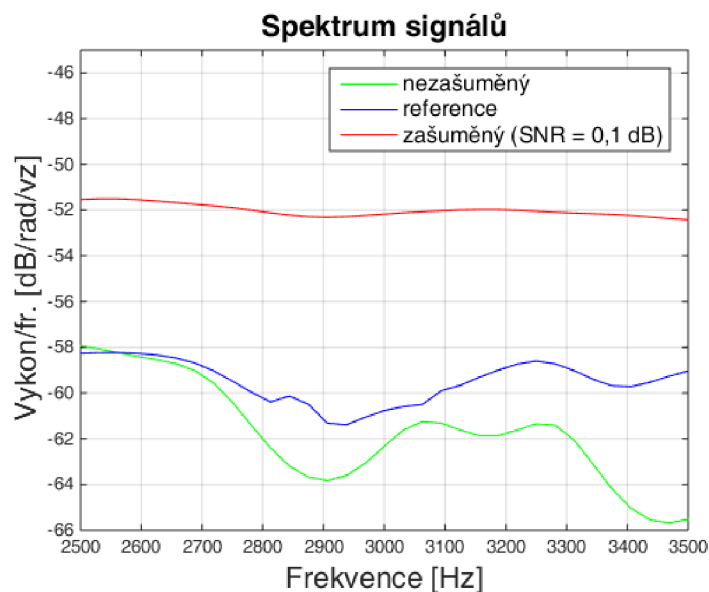


Obr. 3.12: Spektrum zašuměného a nezašuměného nestresového signálu *signNestres1* a stresového signálu *signStres8*. Plnou čárou je zobrazeno spektrum signálu *signNestres1* a čárkovaně spektrum signálu *signStres8*. Modrou čárou jsou spektra nezašuměných signálů a červenou čárou jsou spektra signálů s přidáním bílým gaussovským šumem  $SNR = 3dB$

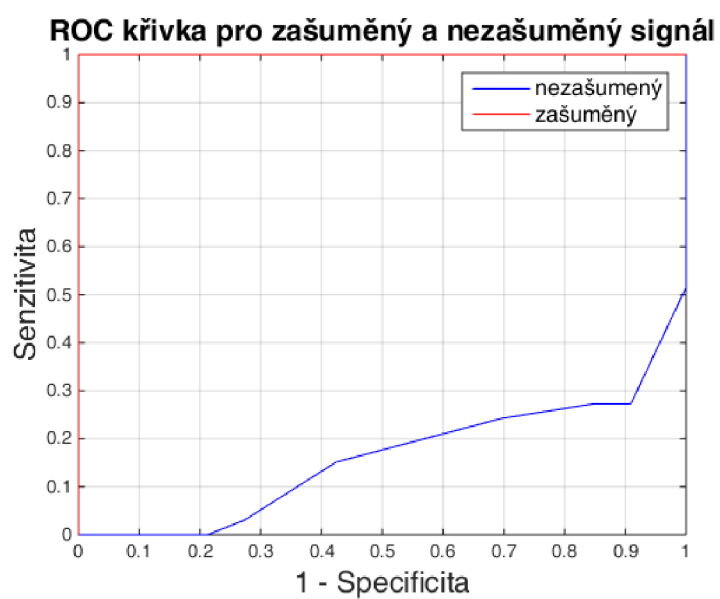
Problém nastává v případě, pokud systém mění charakter zašumění. Například pokud bychom vytvořili referenční nestresový signál a následně do klasifikátoru poslali nestresový signál, který by byl více zašuměný (větší SNR). Tento signál by byl klasifikován jako stresový. V reálné situaci je tedy ideální, pokud jsou všechny audio nahrávky pořízeny stejným zařízením a ve stejném prostředí. Zhotovit několik audio nahrávek a vytvořit si referenční spektrum, např. podobně jak bylo naznačeno

v kapitole 3.4. Výpočtem ROC křivky z audio nahrávky, u které chceme provést klasifikaci a zmíněné referenční nestresové nahrávky jsme schopni provést klasifikaci signálů.

Na obrázku 3.13 je pěkně vidět, jaký má vliv na výsledné spektrum i nepatrné přidání bílého gaussovského šumu ( $SNR = 0,1dB$ ). Zeleně je zobrazeno spektrum nezašuměného signálu. Modře je zobrazeno spektrum referenčního signálu (podrobně popsán v kapitole 3.4) a červeně je zobrazeno spektrum signálu s přidáním bílým gaussovským šumem. Z pohledu detekce stresu nás zajímá pouze spektrum od  $2500Hz$  do  $3500Hz$ . V tomto rozsahu je úroveň výkonu spektra mnohem vyšší než maximální nestresový signál. ROC křivka z těchto signálů tento zašuměný nestresový signál detekuje jako stresový,  $AUC = 1$  (viz. obrázek 3.14).



Obr. 3.13: Vliv zašumění signálu na spektrum. Spektrum nezašuměného signálu *signNestres7.wav* ze zobrazeno zeleně, spektrum *reference* modře a zašuměný signál *signNestres7.wav* ( $SNR = 0.1$  dB) červeně.



Obr. 3.14: Modrá ROC křivka je vytvořena z signálů *signNestres7* a *reference*. Červená ROC křivka je vytvořena ze zašuměného signálu *signNestres7* ( $SNR = 0, 1dB$ ) a *reference*

## 4 STRUKTURA PROGRAMU

Hlavním výsledkem této práce jsou doporučení získaná analýzou. Aby bylo možné data zanalyzovat, bylo potřeba vytvořit jednoduchý program, který je možné využít pro analýzu. V rámci práce byly vytvořeny dva programy. První obsahuje předpřipravené analýzy. Pokud by chtěl uživatel provést analýzy s jinými daty, musí provést úpravu přímo ve zdrojovém kódu. Druhý program byl vytvořen včetně grafického uživatelského rozhraní (GUI). U tohoto programu uživatel nemusí zasahovat do zdrojového kódu. Oba programy využívají stejné funkce. Hlavním souborem programu s předpřipravenými analýzami je `main.m`. Hlavním souborem programu s GUI je `gui.m`. V následujících odstavcích budou popsány jednotlivé funkce, které programy používají.

Funkce `spectSign` provádí převod signálu z časové oblasti do oblasti frekvenční (odhaduje spektrum signálu). Funkce má šest povinných vstupních parametrů, `spectSign(audio, metoda, fmin, fmax, cas, snr)`. Parametr *audio* udává cestu k audio souboru, kterou lze zadat pomocí celé cesty, nebo i relativní cestou. Samotné načtení souboru je provedeno pomocí funkce `wavread`, která podporuje pouze audio formát *wav*. Parametr *metoda* určuje metodu odhadu spektra. Parametr může nabývat pouze následujících textových řetězců: `welch/yulewalker/burg`. Parametry *fmin* a *fmax* slouží pro předzpracování signálů (viz. kapitola 3.1). Parametr *cas* nabývá hodnot 0, nebo 1. Pokud je hodnota parametru 1, je změřena doba odhadu spektra pomocí funkcí `tic` a `toc`. Poslední parametr *snr* slouží pro přidání bílého gaussovského šumu k audio signálu. Pokud má hodnotu 0, není přidán žádný šum. Pokud má hodnotu různou od 0, je šum přidán pomocí funkce `awgn`. Funkce `spectSign` vrací odhadnuté spektrum v matici  $129 \times 2$ , kde v prvním sloupci jsou hodnoty výkonu a v druhém sloupci jsou příslušné frekvence. Kromě spektra vrací funkce také dobu jak dlouho trvalo odhadnutí spektra. Pokud nebyl nastaven příslušný parametr pro měření doby, vrací se v této proměnné 0. Tato funkce byla převzata z diplomové práce [6] a byla upravena pro potřeby této práce. Z původní funkce bylo kompletně převzato předzpracování signálů, které není problematikou této práce.

Funkce `roc` počítá body ROC křivky. Funkce má dva povinné vstupní parametry, `roc(data1,data2)`. *Data1* a *data2* jsou funkční hodnoty spektra (vektor). Tyto vstupní data musí mít stejnou délku. Výstupem funkce jsou body ROC křivky ve dvou vektorech (souřadnice *y* a souřadnice *x* všech bodů ROC křivky). Tato funkce byla převzata ze studijních materiálů pro předmět MSTK vyučovaný na FEKT VUTBR a jejím autorem je doc. RNDr. Jitka Poměnková, Ph.D. a byla upravena pro potřeby této práce.

Funkce `auc` počítá plochu pod křivkou. Funkce má dva povinné vstupní parametry, `auc(Pfa,Pd)`. Parametry *Pfa* a *Pd* odpovídají souřadnicím *x* a *y* všech bodů

křivky jejíž plochu chceme vypočítat. V našem případě jsou tyto parametry přímo výstupy z funkce `roc`. Plocha pod křivkou je vypočtena pomocí lichoběžníkové metody.

Kromě zdrojových souborů k programům a audio souborům jsou součástí práce soubory `data.mat` a `prum.mat`. Soubor `data.mat` obsahuje pomocné proměnné *maxNestres* a *minStres*. V těchto proměnných je uloženo spektrum ve stejném formátu, jaké vrací i funkce `spectSign`. Jak byly tyto proměnné vytvořeny bylo zmíněno v kapitole 3.4. Z těchto proměnných je vypočteno referenční nestresové spektrum, *reference* (viz. kapitola 3.4). *Reference* je použita při ověřování klasifikace pomocí ROC křivek.

## 4.1 Předpřipravené analýzy

V rámci práce bylo provedeno několik různých typů analýz. Soubor `main.m` obsahuje program s ukázkovými analýzami. Program se spouští s jedním povinným parametrem, `main(mod)`. Parametr *mod* může nabývat hodnot 1, 2, nebo 3. Program využívá naprogramovaných funkcí, které byly popsány v kapitole 4.

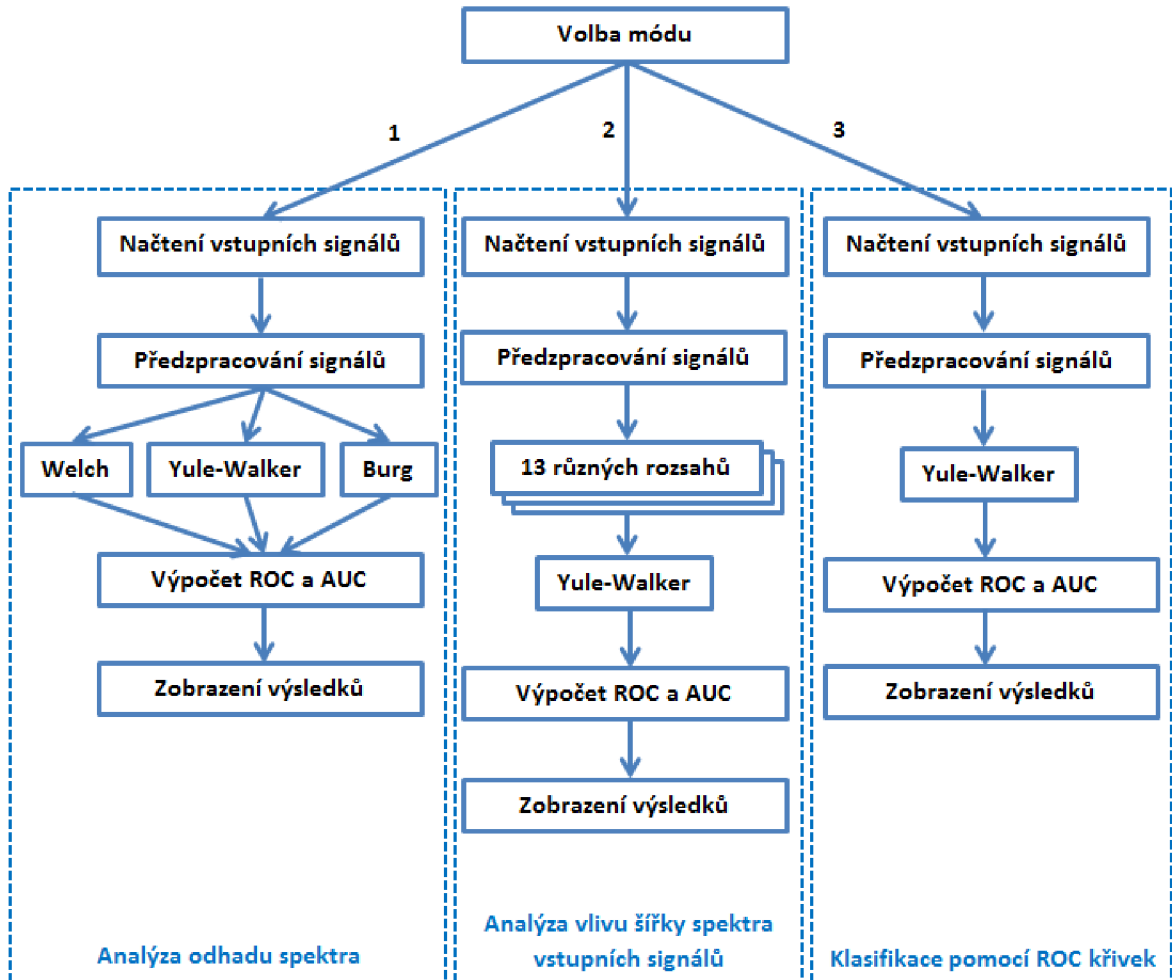
V případě spuštění programu s parametrem 1 (`main(1)`), bude provedena ukázková analýza metod odhadu spektra (viz. kapitola 3.2). Analýza je provedena na signálech *signNestre2.wav* × *signStre1-11.wav*. Pro případné provedení analýzy na jiných datech je potřeba upravit parametr funkcí `spectSign` (viz. kapitola 4). Po spuštění programu budou do *Command Window* postupně vypsány hodnoty *AUC* a doby trvání odhadu spektra *t* (pro všechny kombinace). Do jednoho grafu se vykreslí tři ROC křivky (odhad spektra pomocí Welchova periodogramu, AR Yule-Walker, AR Burg) pro jednu náhodně zvolenou kombinaci signálů (*signNestre2.wav* × *signStre7.wav*).

Pokud je program spuštěn s parametrem 2 (`main(2)`), bude provedena ukázková analýza pro vliv frekvenčního rozsahu spektra vstupních signálů (viz. kapitola 3.3). Analýza je provedena na signálech *signNestre1.wav* × *signStre6.wav*. S těmito soubory je vypočteno 13 ROC křivek pro různý rozsah spektra (s odpovídajícím předzpracováním). Po spuštění programu budou do *Command Window* postupně vypsány hodnoty *AUC* a všechny ROC křivky (pro jednotlivé rozsahy) budou vykresleny do společného grafu. Změna vstupních audio signálů se provádí stejně jako při předchozí analýze.

Poslední možnost spuštění programu je s parametrem 3 (`main(3)`). V tomto případě bude provedeno ověření klasifikátoru na všech stresových a nestresových audio signálech. Princip klasifikátoru je podrobně popsán v kapitole 3.4. Po spuštění programu budou do *Command Window* postupně vypsány hodnoty *AUC* na

základě kterých se provádí klasifikace a všechny ROC křivky (*analyzovaný\_sgnál* × *reference*) budou vykresleny do společného grafu.

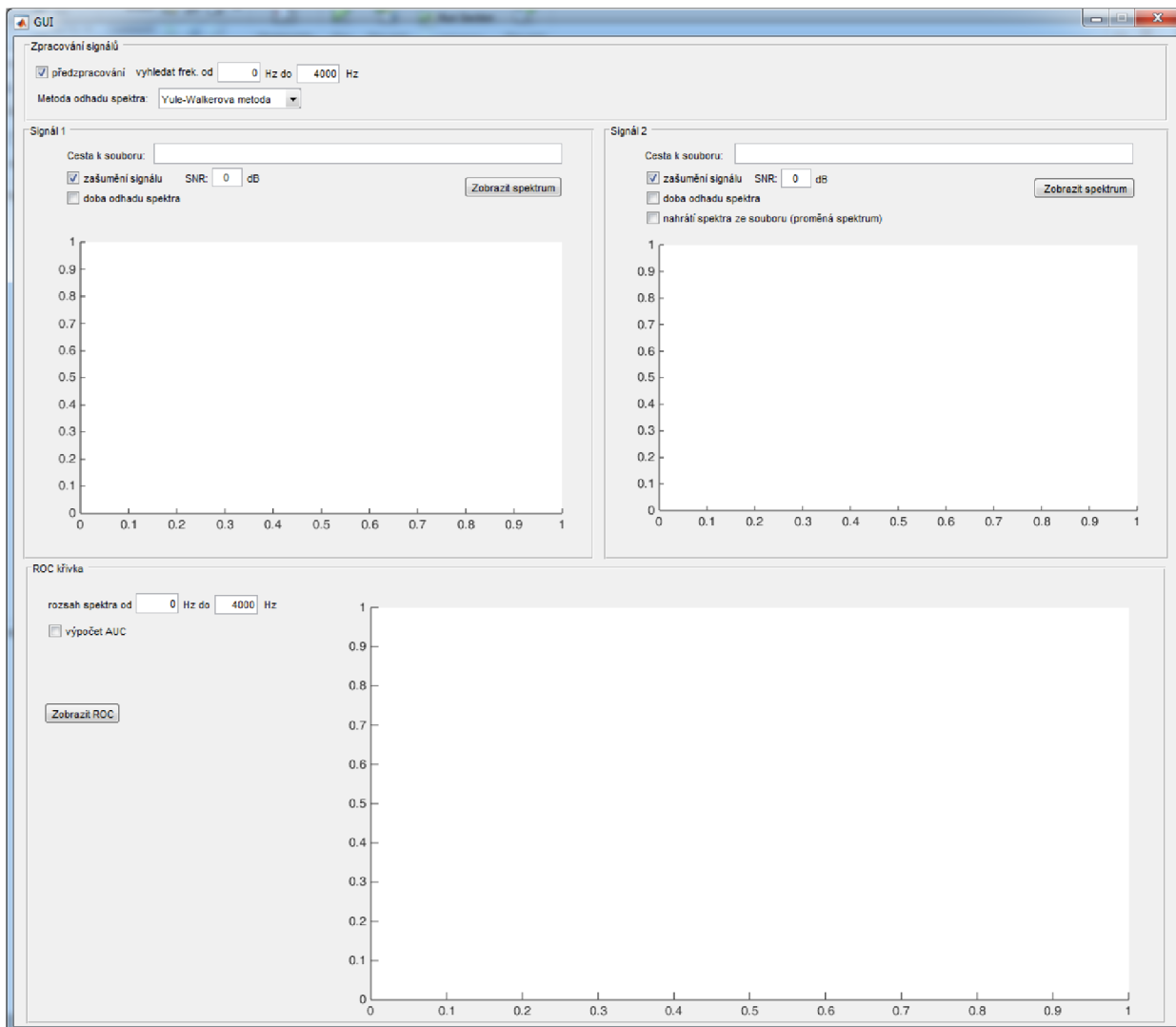
Zjednodušený vývojový diagram programu je zobrazen na obrázku 4.1.



Obr. 4.1: Zjednodušený vývojový diagram programu s předpřipravenými analýzy.

## 4.2 Program pro vlastní analýzy

V rámci práce byl vytvořen také program s GUI (grafické uživatelské rozhraní), který umožňuje provádět dílčí analýzy, aniž by byl potřebný zásah do zdrojového kódu. V kapitole 4.1 jsou popsány před připravené ukázkové analýzy. Pokud by chtěl uživatel provést změnu analýzy, musel by upravit zdrojový kód (např. jiné zdrojové soubory). Pro testování bez zásahu do zdrojového kódu byla vytvořena tato aplikace. Vzhled GUI je zobrazen na obrázku 4.2.

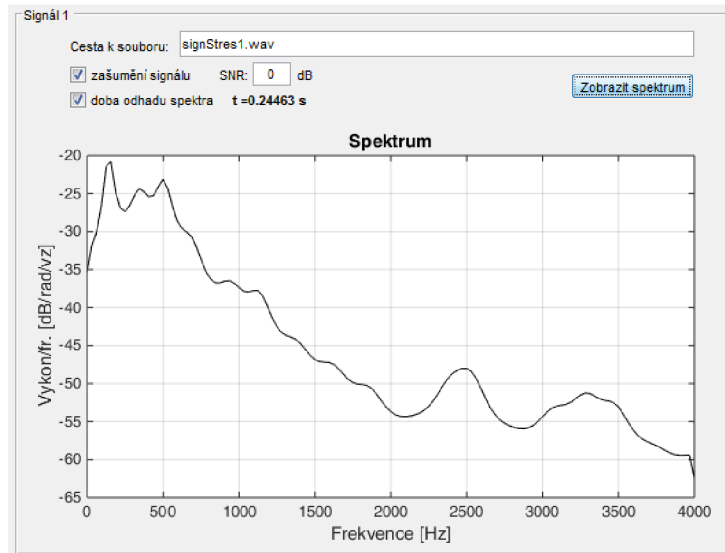


Obr. 4.2: GUI aplikace pro testování.

V sekci *Zpracování signálů* může uživatel zapnout předzpracování, které bude ve zdrojovém signálu vyhledávat oblasti zadaných frekvencí (předzpracování signálu se věnuje kapitola 3.1). Uživatel také může zvolit metodu pro odhad spektra (Welschův periodogram, Yule-Walkerova metoda, nebo Burgova metoda).

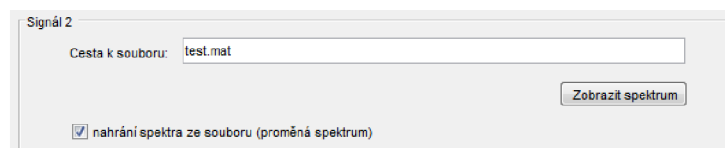
Sekce *Signál 1* a *Signál 2* slouží pro nahrávání analyzovaných souborů. Jediný rozdíl mezi sekcí *Signál 1* a *Signál 2* je přímá možnost nahrání spektra *Signál 2*. *Cesta k souboru* slouží ke specifikaci vstupních dat. Pokud není zaškrtnuto pole *nahrání spektra ...* je očekávána cesta ke zdrojovému audio souboru. Cesta může být zadána jak relativně (pokud je soubor ve stejné složce jako zdrojové soubory programu stačí zadat pouze název souboru včetně přípony), nebo plnou cestou (např. C:\audio\signal.wav). Pokud je pole *nahrání spektra ...* zaškrtnuto, zmizí možnosti pro zašumění signálu a měření doby odhadu spektra (viz. obrázek 4.4). V tomto





Obr. 4.3: Zobrazení spektra audio signálu.

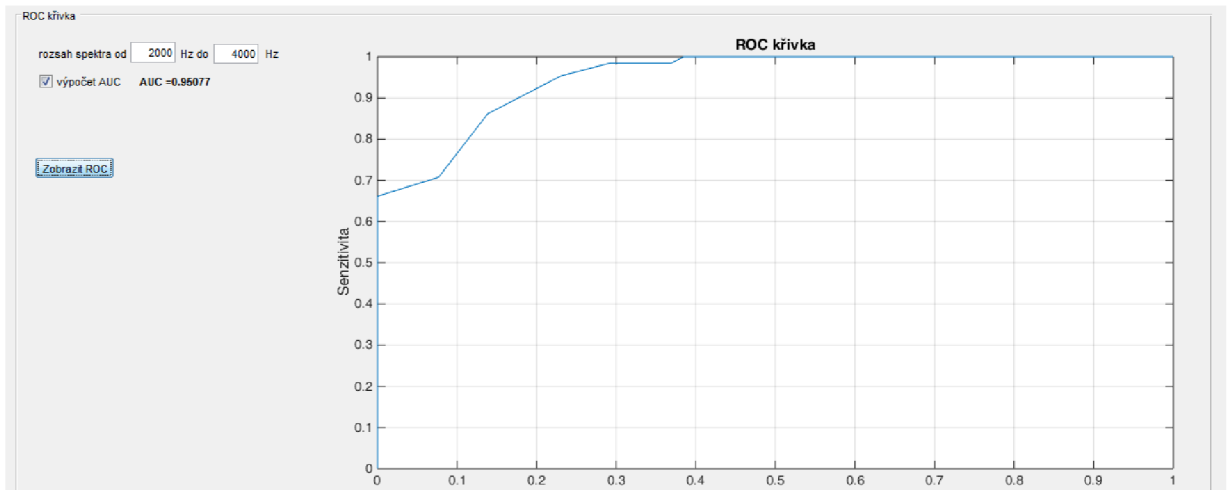
případě není očekáván audio soubor, ale soubor s příponou *.mat*, který obsahuje proměnnou *spektrum*. Proměnná *spektrum* je matice o rozměrech  $129 \times 2$ , kde první sloupec obsahuje hodnoty výkonu spektra a druhý sloupec obsahuje dané frekvence. Pokud není pole *nahrání spektra . . .* zaškrtnuto můžeme k audio signálu přidat bílý gaussovský šum (zaškrtnutím *zašumění signálu*) a nechat si zobrazit dobu trvání odhadu spektra (zaškrtnutím *doba odhadu spektra*). Po kliknutí na tlačítko *Zobrazit spektrum* se nejdříve provede načtení audio signálu, zašumění (pokud bylo zvoleno), předzpracování (pokud bylo zvoleno), provede se odhad spektra pomocí zvolené metody a zobrazí se spektrum signálu spolu s dobou odhadu spektra (pokud bylo zvoleno), viz. obrázek 4.3. Všechny tyto kroky provádí funkce `spectSign` s příslušnými parametry. Pokud byla zvolena možnost přímého nahrání spektra, nahraje se zadaný soubor a zobrazí se spektrum z proměnné *spektrum*.



Obr. 4.4: Přímé načtení spektra ze souboru.

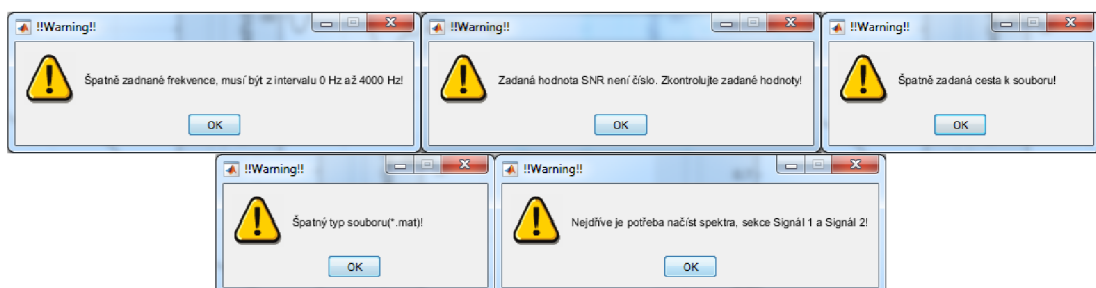
Sekce *ROC křivka* slouží pro výpočet bodů ROC křivky a její vykreslení. Výpočet ROC křivky provádí funkce `roc`. Aby bylo možné vykreslit ROC křivku, je potřeba nejdříve zobrazit spektra signálů (sekce *Signál 1* a *Signál 2*). Jejich hodnoty jsou uloženy do globálních proměnných *data1* a *data2*. Pro práci s globálními proměnnými jsou vytvořeny funkce `setGlobalData1`, `setGlobalData2` pro uložení

dat. Pro čtení dat z globálních proměnných jsou využity funkce `getGlobalData1`, `getGlobalData2`. Velikost vstupních spekter signálů je volena podle nastavení uživatele. Zaškrtnutím pole *výpočet AUC* bude vypočtena také hodnota AUC a zobrazena uživateli. Výpočet AUC provádí funkce `auc`. Po kliknutí na tlačítko *Zobrazit ROC* je proveden výpočet ROC křivky a její zobrazení. Pokud byla zaškrtnuta volba výpočtu AUC, je proveden výpočet a zobrazení AUC hodnoty ROC křivky (viz. obrázek 4.5).



Obr. 4.5: Zobrazení ROC křivky a její AUC hodnoty.

Všechny parametry zadávané uživatelem jsou kontrolovány a pokud v daných polích nejsou očekávané hodnoty, je na to uživatel upozorněn warningem s příslušnou hláškou (viz. obrázek 4.6). U předzpracování a rozsahu spektra je hlídáno, že uživatel zadal číslo z rozsahu 0 – 4000. U zašumění je kontrolováno zda bylo zadáno číslo. Při načítání zdrojových souborů je prováděna kontrola existence souborů. Při přímém načítání spektra je prováděna i kontrola správného typu souboru (\*.mat) a existence proměnné *spektrum*. Při kliknutí na tlačítko *Zobrazit ROC* je provedena kontrola, zda jsou data v globálních proměnných *data1* a *data2*



Obr. 4.6: Ukázky warningů, při chybně zadaných parametrech do aplikace.

## 5 ZÁVĚR

V této práci bylo představeno základní využití ROC křivek pro hodnocení metod. Tato problematika je vztažena na detekci stresu ve zvukových signálech. Výsledkem jsou doporučení pro jednotlivé dílčí části klasifikátoru stresu v hlasu. Práce volně navazuje na diplomovou práci [6]. Z této práce bylo převzato předzpracování vstupních signálů. Také byly převzaty parametry pro odhad spektra pomocí Welchova periodogramu a AR modelu metody Yule-Walker a Burg. Touto problematikou se práce zabývala a bylo by tedy zbytečné tuto problematiku řešit znovu.

Tato práce se zaměřila na porovnání metod odhadu spektra, protože klasifikace je prováděna ve frekvenční oblasti a také rozsahu spektra vstupních dat. Na základě těchto poznatků byl navržen klasifikátor stresu v hlasu.

Pro odhad spektra byly porovnány následující metody: neparametrický Welchův periodogram a parametrické metody využívající AR model s odhadem koeficientů pomocí Yule-Walkerovy a Burgovy metody. Nejdříve byly porovnány tyto tři metody pomocí ROC křivek se vstupními spektry v rozsahu  $0 - 4000\text{ Hz}$ . Testování bylo provedeno na testovací množině o velikosti 11 vstupních dvojic signálů. Ohodnocení bylo provedeno výpočtem  $AUC$  pomocí lichoběžníkového pravidla. Metody využívající parametrické metody odhadu spektra vykazovaly lepší výsledky než Welchův periodogram. Jelikož Yule-Walkerova a Burgova metoda byly ohodnoceny naprosto stejně, bylo přidáno další kritérium na základě kterého bychom mohli metody porovnat. Za toto kritérium byla zvolena výpočetní náročnost metod. Pro úplnost byla měřena i výpočetní náročnost Welchova periodogramu. Pokud by byl výrazně rychlejší než metody využívající AR model pro odhad spektra, bylo by potřeba provést podrobnější vyhodnocení pomocí  $AUC$ . Jednoznačně nejrychlejší metoda byla Yule-Walkerova (o řád rychlejší než Burg). Welchův periodogram je výpočetně nejnáročnější metodou. Jako nejvýhodnější metoda odhadu spektra byla zvolena metoda odhadující spektrum získané z **AR modelu metodou Yule-Walker**.

Další důležitou problematikou, kterou se práce zabývá, byla velikost spektra vstupních signálů. V tomto bodě práce vychází z předchozího doporučení a pro odhad spektra, kdy byla zvolena metoda využívající odhad AR modelu metodou Yule-Walker. Na základě hodnot pásem formantů (tabulka 2.1) bylo zvoleno 13 rozsahů zohledňující různé kombinace prvních tří formantů samohlásek. Z této tabulky také předpokládáme, že nejlépe by měl klasifikátor pracovat pro třetí formanty, kde je frekvenční rozsah všech samohlásek velmi podobný. Testování proběhlo na 47 náhodných kombinacích stresových a nestresových signálů. Nejlepších výsledků bylo dosaženo s rozsahem **vstupního spektra 2500 - 3500 Hz**. Hodnocení pomocí  $AUC$  bylo nejvyšší a ve více než polovině případů se dokonce jednalo o ideální klasifikátor (ohodnocení  $AUC = 1$ ). Tento rozsah je tedy doporučen pro další vy-

užití. V tomto rozsahu se pohybují třetí formanty všech samohlásek. Výsledek této analýzy se tedy shoduje s teoretickými očekávaními.

V poslední části byl navržen detektor stresu využívající ROC křivky pro klasifikaci. Základním faktem, který umožňuje klasifikaci jsou rozdílné úrovně výkonu spekter nestresových a stresových signálů. Aby klasifikátor správně pracoval je nejdříve potřeba vytvořit referenční nestresový signál, jehož spektrum spolu se spektrem analyzovaného signálu jsou vstupy pro ROC křivky. Referenční nestresový signál byl získán jako aritmetický průměr ze spektra maximálního nestresového signálu (získán ze všech nestresových signálů) a spektra minimálního stresového signálu (získán ze všech stresových signálů). Pro klasifikátor byly určeny kritéria pro stresový signál a nestresový signál na základě plochy pod příslušnou ROC křivkou. Pro **stresové signály** bylo navrženo kritérium  $AUC > 0,8$  a pro **nestresové signály** kritérium  $AUC < 0,5$ . Signály s  $0,5 < AUC < 0,8$  nelze s jistotou klasifikovat. Jsou tedy klasifikovány jako potenciální stresové signály, ale v této oblasti se mohou pohybovat také zašuměné nestresové signály. Aby klasifikátor pracoval správně je potřeba se snažit získat analyzované signály pomocí stejného nahrávacího zařízení a nahrávky získat ve stejném prostředí. Čím méně se budou lišit podmínky pořízení analyzovaných nahrávek, tím lépe bude klasifikátor pracovat. V ideálním případě jsou v daném prostředí s daným nahrávacím zařízením pořízeny testovací stresové a nestresové signály a z nich je vytvořen referenční signál. Analyzované signály jsou potom pořizovány na stejném místě stejným zařízením. Tím budou zajištěny ideální podmínky pro správnou klasifikaci. Navržený klasifikátor byl otestován na všech stresových i nestresových signálech, které byly pro účely této práce získány. Pouze jeden stresový signál byl chybně klasifikován. Tento stresový signál byl klasifikován jako potenciální stresový signál. Po otestování správné funkčnosti klasifikátoru s těmito parametry, byl vyzkoušen ještě klasifikátor, který by pracoval s rozsahem spekter  $2500 - 3000 Hz$ . Při použití tohoto rozsahu sice přijdeme o část rozsahu třetího formantu samohlásek  $i$  a  $e$ , ale analyzovaný signál nebude ovlivněn rušením v tomto rozsahu, kde ostatní samohlásky nejsou přítomny. Touto úpravou ovšem nedošlo ke zlepšení klasifikátoru. Klasifikátor tedy využívá pro předzpracování a rozsah vstupních spekter doporučení  $2500 - 3500 Hz$ .

Práci je možné rozšířit v několika oblastech. Pro vyhodnocení ROC křivek byla využita lichoběžníková metoda výpočtu  $AUC$ . Její výhodou je nenáročný rychlý algoritmus, ale tato metoda má tendence k podhodnocování  $AUC$ . Jako velmi zajímavé se jeví porovnání s jinými metodami výpočtu  $AUC$ . Například odhad  $AUC$  pomocí neparametrické metody Mann-Whitney U test, která využívá zcela jiný přístup pro výpočet. Případně výpočet  $AUC$  pomocí aproximace binormalním modelem s odhadem pomocí maximum-likelihood. Dalším možným zaměřením je vliv zašumění, který už byl zmíněn. Tato oblast je ovšem velmi problematická a je možné, že při

různém zašumění signálů by bylo potřeba použít zcela jiný přístup pro klasifikaci. Tato problematika se nabízí jako možné další téma pro nezávislé zpracování.

# LITERATURA

- [1] PSUTKA, Josef *Mluvíme s počítačem česky*. Praha: Academia, 2006. Česká matice technická (Academia). ISBN 80-200-1309-1.
- [2] STOICA, Petre. a Randolph L. MOSES. *Spectral analysis of signals*. Upper Saddle River, N.J.: Pearson/Prentice Hall, c2005. ISBN 0131139568. Dostupné z URL: <<http://user.it.uu.se/~ps/SAS-new.pdf>>.
- [3] PROAKIS, John G. *Algorithms for statistical signal processing*. Upper Saddle River, N.J.: Prentice Hall, c2002. ISBN 0130622192.
- [4] WESTIN, L. K. *Receiver operating characteristic (ROC) analysis: Evaluating discriminance effects among decision support systems* [online]., 28 [cit. 2016-11-04] Dostupné z URL: <<https://www8.cs.umu.se/research/reports/2001/018/part1.pdf>>.
- [5] YEH, Shi-Tao. *Using Trapezoidal Rule for the Area Under a Curve Calculation* [online]. , 4 [cit. 2016-11-14]. Dostupné z URL: <<http://www2.sas.com/proceedings/sugi27/p229-27.pdf>>.
- [6] DULESOV, E. *Identifikace významných spektrálních složek ve stresovém řečovém signálu*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky, 2016. 58 s., 20 s. příloh. Diplomová práce. Vedoucí práce: doc. RNDr. Jitka Poměnková, Ph.D.
- [7] Wikipedia. *Pulzně kódová modulace* [online]. 2015 [cit. 2016-10-06]. Dostupné z URL: <[https://cs.wikipedia.org/wiki/Pulzn%C4%9B\\_k%C3%B3dov%C3%A1\\_modulace](https://cs.wikipedia.org/wiki/Pulzn%C4%9B_k%C3%B3dov%C3%A1_modulace)>.
- [8] TŮMA, Jiří. *Parametrická metoda výpočtu frekvenčních spekter signálů*. 7 th International Scientific - Technica l Conference - PROCESS CONTROL 2006 [online]. 2016, , 9 [cit. 2016-11-07]. Dostupné z URL: <[http://home1.vsb.cz/~tum52/publications/TumaFt\\_CZ.pdf](http://home1.vsb.cz/~tum52/publications/TumaFt_CZ.pdf)>.
- [9] MathWorks. *Formant Estimation with LPC Coefficients*. *MathWorks* [online]. 2015 [cit. 2017-4-06]. Dostupné z URL: <<http://www.mathworks.com/examples/signal/mw/signal-ex82230229-formant-estimation-with-lpc-coefficients>>.

## SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

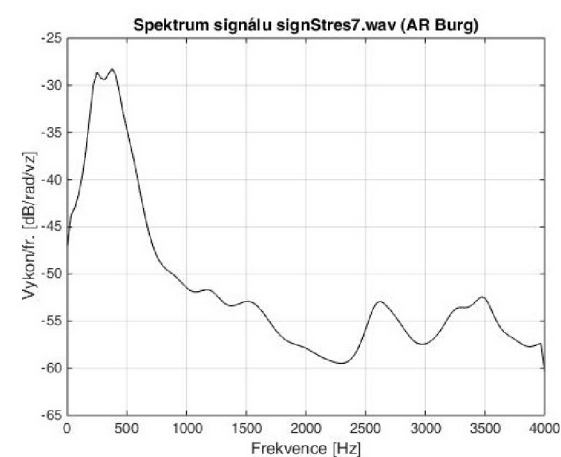
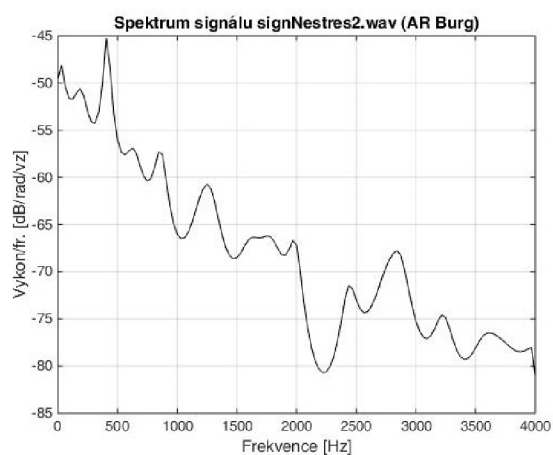
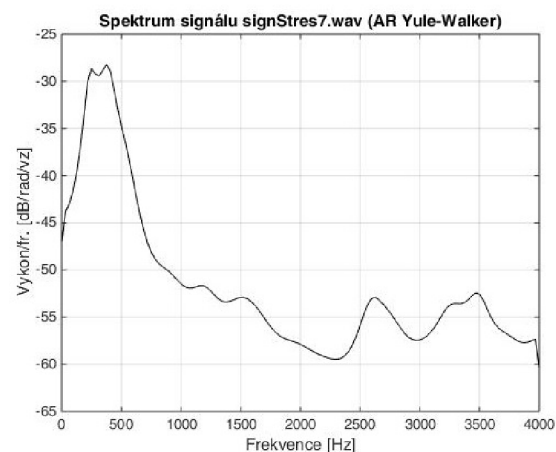
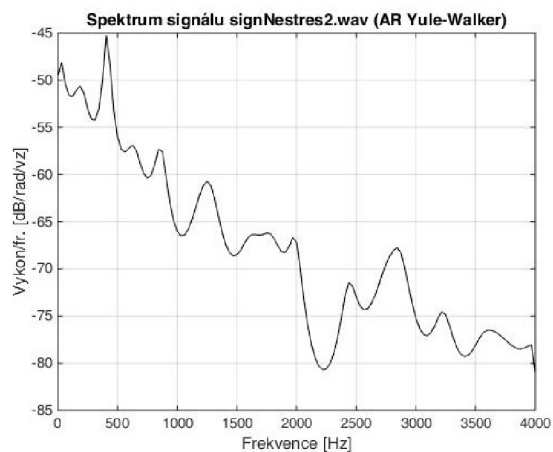
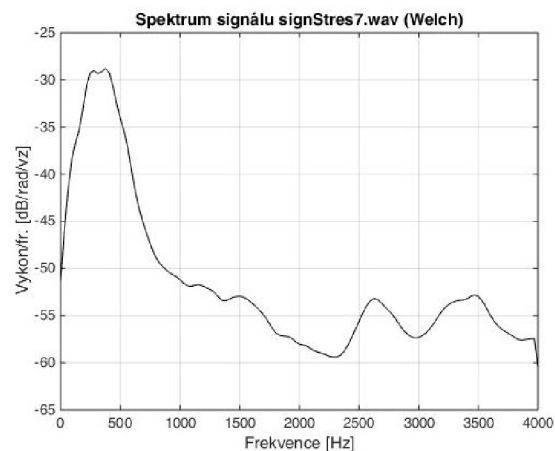
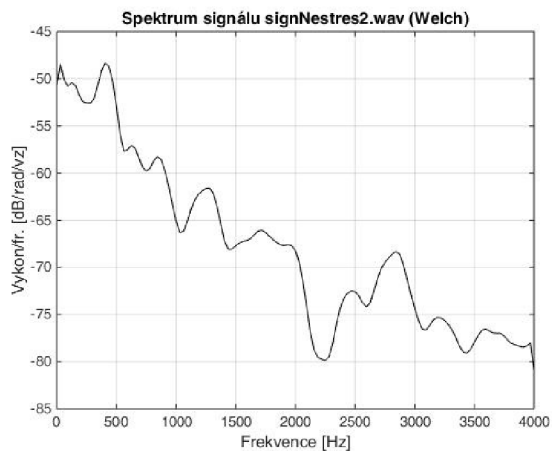
AR	autoregresní proces
AUC	plocha pod křivkou
EFF	efektivita
FN	false negative
FP	false positive
$F_s$	vzorkovací frekvence
GUI	grafické uživatelské rozhraní
PVP	předpoklad pozitivního testu
ROC	receiver operating characteristic, nástroj pro hodnocení a optimalizaci binárního klasifikačního systému
SE	senzitivita
SERR	směrodatná odchylka
SNR	poměr signál šum
SP	specificita
TN	true negative
TP	true positive

# SEZNAM PŘÍLOH

A Ukázka odhadu spekter	51
B Obsah přiloženého CD	52



# A UKÁZKA ODHADU SPEKTER



## B OBSAH PŘILOŽENÉHO CD

K práci je přiloženo CD, na kterém jsou všechny programy, audio soubory a práce v elektronické podobě včetně zdrojových kódů pro  $\text{\LaTeX}$ . V kořenovém adresáři je uložena práce ve formátu *pdf*, složky *latex* a *matlab* a textový dokument `readme.txt`, ve kterém je struktura CD.

Složka *latex* obsahuje všechny zdrojové kódy pro vygenerování pdf. Hlavní soubor je *sablona-prace.tex*, který obsahuje odkazy na jednotlivé kapitoly (podsložka *text*). Všechny obrázky práce jsou uloženy v podsložce *obrazky*.

Složka *matlab* obsahuje všechny zdrojové kódy k programům (`main.m`, `GUI.m`, `GUI.fig`, `spectSign.m`, `roc.m`, `auc.m`, `setGlobalData1.m`, `setGlobalData2.m`, `getGlobalData1.m` a `getGlobalData2.m`). Zdrojové kódy byly vytvořeny a testovány v prostředí MATLAB R2014b. Kromě kódů programů jsou zde audio soubory (`signNestres1-12.wav`, `signStres1-11.wav`) a datové soubory `prum.mat`, `data.mat`.