



# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

## FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

## ÚSTAV RADIOELEKTRONIKY

DEPARTMENT OF RADIOENGINEERING

## URČOVÁNÍ A ANALÝZA PERIODICITY V ŘEČOVÝCH SIGNÁLECH

ESTIMATION AND ANALYSIS OF SPEECH SIGNAL PERIODICITY

### BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

### AUTOR PRÁCE

AUTHOR

Jan Malucha

### VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. Milan Sigmund, CSc.

BRNO 2020



# Bakalářská práce

bakalářský studijní program **Elektronika a komunikační technologie**

Ústav radioelektroniky

**Student:** Jan Malucha

**ID:** 203286

**Ročník:** 3

**Akademický rok:** 2019/20

**NÁZEV TÉMATU:**

## Určování a analýza periodicity v řečových signálech

### POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s charakterem řečového signálu a metodami na určování krátkodobé periody a periodicity proměnných signálů. Z dostupných publikací vytvořte přehledovou studii o nejčastěji používaných metodách v této oblasti pro řečové signály. Naprogramujte vybrané algoritmy na určování periodicity řeči rozšířené o možnosti pro realizaci lineárního a nelineárního přizpůsobení časové osy. Algoritmy ověřte na reálném řečovém signálu.

Vytvořte a ověřte autonomní program na určování periodicity řečových signálů v časové oblasti alespoň třemi různými metodami. Do programu přidejte modul na shlukování řečových segmentů podle kritéria periodicity a následné ukládání na disk. Programem zpracujte řečové záznamy z určené databáze a výsledky porovnejte s podobnými výsledky získanými pomocí volně dostupných softwarových nástrojů jako Praat apod.

### DOPORUČENÁ LITERATURA:

[1] PSUTKA, J., MÜLLER, Z., MATOUŠEK, J., RADOVÁ, V. Mluvíme s počítačem česky. Praha: Academia, 2006.

[2] SIGMUND, M. Rozpoznávání řečových signálů. Skriptum FEKT VUT v Brně. Brno: MJ servis, 2007.

**Termín zadání:** 3.2.2020

**Termín odevzdání:** 4.6.2020

**Vedoucí práce:** prof. Ing. Milan Sigmund, CSc.

**prof. Ing. Tomáš Kratochvíl, Ph.D.**  
předseda rady studijního programu

### UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské 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**

V teoretické části bakalářské práce je provedeno seznámení s podstatou řečového signálu, metodou krátkodobé analýzy a je objasněn pojem periodičita. Následuje rozbor analyzovaných vlastností signálu čítajících znělost/neznělost, intonaci a podobnost sousedních period, dále je uvedena přehledová studie metod použitelných k jejich analýze. Tyto metody zahrnují ZCR, STE, HNR, ACF, NCC, AMDF a DTW. V rámci praktické části byly naprogramovány algoritmy využívající tyto metody v prostředí MATLAB, společně s obslužnými funkcemi pro segmentaci a předzpracování signálu. Všechny algoritmy byly otestovány na reálných řečových signálech a dosažené výsledky jsou rozebrány na konci práce společně s příklady využitelnosti v praxi.

## **Klíčová slova**

Znělost, intonace, podobnost, periodičita, krátkodobá analýza, autokorelace, AMDF

## **Abstract**

Theoretical part starts with a short introduction to speech signals and short time analysis method. Concept of periodicity is clarified. This is followed by analysis of the speech signal parameters - voiced/unvoiced classification, intonation and short time period similarity. Next the overview of methods used to analyze the mentioned parameters together with concepts of their algorithms is provided. These methods include ZCR, STE, HNR, ACF, NCC, AMDF and DTW. Algorithms using the methods with supporting preprocessing algorithms were programmed in MATLAB for practical part. All were tested on real speech signals and the results are discussed at the end of the thesis.

## **Keywords**

Voicing, intonation, similarity, periodicity, short time analysis, autocorrelation, AMDF

MALUCHA, Jan. *Určování a analýza periodicity v řečových signálech* [online]. Brno, 2020 [cit. 2020-05-27]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/126097>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky. Vedoucí práce Milan Sigmund.

## **Prohlášení**

„Prohlašuji, že svou bakalářskou práci na téma Určování a analýza periodicity v řečových signálech jsem vypracoval 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 uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si 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.

V Brně dne: 4. června 2020

.....  
podpis autora

# Obsah

1.	Úvod.....	12
2.	Charakter řečového signálu.....	13
2.1	Řeč jako diskrétní signál .....	13
2.2	Krátkodobá analýza.....	13
2.2.1	Mikrosegment .....	13
2.2.1.1	Fonetická klasifikace .....	13
2.2.1.2	Akustická klasifikace.....	14
2.3	Periodicita .....	14
2.4	Analyzované veličiny .....	15
2.4.1	Znělost a neznělost.....	15
2.4.2	Intonace.....	17
2.4.3	Podobnost.....	17
2.5	Existující studie.....	18
3.	Výpočetní metody .....	20
3.1	Zero Crossing Rate.....	20
3.2	Short Time Energy .....	21
3.3	Harmonics-to-Noise Ratio .....	21
3.4	Autokorelační funkce .....	22
3.5	Average Magnitude Difference Function.....	23
3.6	Normalized cross-correlation .....	24
3.7	Dynamic Time Warping.....	24
3.8	Přizpůsobování časové osy .....	26
3.8.1	Lineární přizpůsobení časové osy .....	26
3.8.2	Nelineární přizpůsobení časové osy .....	27
4.	Praktická část .....	28
4.1	Popis programu .....	28
4.1.1	Segmentace .....	29
4.1.2	Určení znělosti/neznělosti.....	29
4.1.2.1	ZCR .....	29
4.1.2.2	STE .....	29

4.1.2.3	HNR.....	30
4.1.3	Určení intonace.....	30
4.1.3.1	ACF .....	30
4.1.3.2	AMDF.....	31
4.1.3.3	NCC .....	31
4.1.4	Určení podobnosti.....	31
4.1.4.1	AMDF.....	32
4.1.4.2	NCC .....	33
4.1.4.3	DTW .....	33
4.1.5	GUI .....	33
4.2	Testování .....	35
4.2.1	Znělost/Neznělost .....	35
4.2.2	Intonace.....	39
4.2.3	Podobnost.....	41
4.3	Srovnání s programem Praat .....	45
4.4	Praktické využití.....	47
5.	Závěr .....	50

# Seznam symbolů a zkratek

## Zkratky:

ACF	...	Autocorrelation Function
AMDF	...	Average Magnitude Difference Function
DTW	...	Dynamic Time Warping
EKG	...	Elektrokardiogram
GUI	...	Graphic User Interface
HNR	...	Harmonic-to-Noise Ratio
NCC	...	Normalized Cross Correlation
STE	...	Short Time Energy
ZCR	...	Zero Crossing Rate

## Symboly:

$t$	...	čas	[s]
$n$	...	vzorek	[-]
$\tau(n)$	...	posun	[-]
$f$	...	frekvence	[Hz]
$F_s$	...	vzorkovací frekvence	[Hz]
$T_0$	...	základní perioda	[s]
$F_0$	...	základní (fundamentální) frekvence	[Hz]
$f(t), x(t)$	...	funkce	[-]
sgn	...	funkce signum	{-1,0,1}
$R(\tau)$	...	autokorelační funkce	[-]
$N$	...	počet vzorků okna	[-]
$A, B$	...	označení sousedních period signálu; B = delší z period	
$d$	...	Euklidovská vzdálenost	[-]
$p, q$	...	soubory dat	
$\log_a$	...	logaritmus o zákl. a	
$r(\tau_{peak})$	...	norm. hodnota 2. extrému ACF	
$Y_{xy}$	...	matice y krátkodobých segmentů o délce x vzorků	



## Seznam obrázků

Obr. 2-1 - Časová reprezentace hlásky /a/ .....	16
Obr. 2-2 - Časová reprezentace hlásky /s/ .....	16
Obr. 3-1 - Distribuce průchodů nulou pro znělou a neznělou řeč [7] .....	20
Obr. 3-2 - Jednostranná autokorelace dokonale periodického signálu .....	22
Obr. 3-3 - AMDF dokonale periodického signálu .....	23
Obr. 3-4 - Jednostranná normalizovaná cross-korelace dokonale periodického signálu .....	24
Obr. 3-5 - Znázornění roztažení funkce pomocí DTW, převzato z nápovědy MATLAB .....	25
Obr. 3-6 - Lineární přizpůsobení časové osy .....	26
Obr. 4-1 - Blokové schéma programu .....	28
Obr. 4-2 - Porovnání algoritmů pro lin. a nelin. časové přizpůsobení .....	32
Obr. 4-3 - Grafické uživatelské rozhraní (analýza podobnosti /a/, alg. NCC(lin), okno o 40 ms, práh 90% podobnost) .....	34
Obr. 4-4 - Grafické znázornění analýzy znělosti/neznělosti algoritmy ZCR, STE a HNR .....	35
Obr. 4-5 - Histogramy vyhodnocených segmentů .....	36
Obr. 4-6 - Vstupní signál a jeho verze po zpracování algoritmy ZCR, STE a HNR	37
Obr. 4-7 - Vstupní signál a jeho verze po zpracování algoritmy ZCR, STE a HNR + manuální rozčlenění věty .....	38
Obr. 4-8 - Histogramy základní frekvence segmentů .....	39
Obr. 4-9 - Grafické zobrazení průběhů F0, resp. intonace .....	40
Obr. 4-10 - Podobnost sousedních period dokonale periodického signálu pro algoritmy AMDF, NCC a DTW .....	41
Obr. 4-11 - Předzpracovaný signál hlásky /a/ .....	42
Obr. 4-12 - Podobnost sousedních period hlásky /a/ pro algoritmy AMDF, NCC a DTW .....	43
Obr. 4-13 - Analýza programem Praat .....	46
Obr. 4-14 - Histogramy vyhodnocených segmentů v klidu a ve stresu .....	47
Obr. 4-15 - Histogram F0 pro řeč v klidu a ve stresu .....	48

Obr. 4-16 - Průběh podobnosti sousedních period hlásky /a/ pro řeč v klidu a ve  
stresu ..... 49

## Seznam tabulek

Tab. 1 - Znělé a neznělé souhlásky, převzato z [3].....	15
Tab. 2 – Porovnání podobnosti sousedních period pro algoritmus AMDF .....	44

# 1. ÚVOD

Jazyk je multidimenzionální fenomén, v průběhu historie lidstva sloužící jako prostředek mezilidské komunikace. Existuje více definic pojmu „jazyk“; podle P. Mareše jej lze vymezit jako „soubor jednotek znakové povahy, které tvoří uspořádaný celek (systém) a soubor pravidel, jež určují, jak s jednotkami zacházet.“ [1] Jazyk obecně existuje ve dvou formách, a to v mluvené a psané. Jeho starší a pro člověka přirozenější forma mluvená (řeč) se s rozvojem informačních a výpočetních technologií stala předmětem širokého výzkumu, mimo jiné za účelem jejího využití v komunikaci se stroji. Ačkoli výsledky těchto snah jsou již poměrně široce rozšířené v moderních elektronických zařízeních, od možnosti hlasového vyhledávání v počítačích a mobilních telefonech až po zabezpečovací systémy, lékařské diagnostické systémy či kompenzační pomůcky pro sluchově postižené, stále neexistuje dokonalý strojový systém, jenž by byl schopen vést s člověkem plnohodnotný dialog. Do budoucna tak můžeme očekávat řadu nových metod, které společně s rostoucími možnostmi výpočetních technologií budou mluvenou řeč stále více přibližovat plně využitelnému prostředku ke komunikaci se strojem.

Základem pro takovou komunikaci je záznam akustické reprezentace řeči – řečový signál. Jeho analýzou může stroj určit nejen obsah řeči z hlediska sémantického, ale také doplňující informace, např. o emocionálním stavu řečníka či dokonce řečnickovu identifikaci. Všechny tyto informace lze získat na základě využití matematických metod aplikovaných na předzpracovaný signál.

Předmětem této práce jsou metody určující základní vlastnosti řečového signálu v jeho časové reprezentaci. Těmito vlastnostmi jsou znělost/neznělost, intonace a periodičita v čase. Všechny tyto vlastnosti jsou spojeny s analýzou krátkodobých period řečového signálu a jejich znalost je základním předpokladem pro získání dalších informací o řeči či řečnickovi. V teoretické části budou uvedeny a podrobněji rozebrány používané metody či jejich modifikace, cílem praktické práce je jejich realizace v prostředí MATLAB a následné ověření jejich vlastností a funkcionality.

## 2. CHARAKTER ŘEČOVÉHO SIGNÁLU

### 2.1 Řeč jako diskrétní signál

Z technického hlediska můžeme pojem signál chápat jako fyzikální veličinu závislou na čase. Řečový signál je potom mechanické vlnění elastického prostředí, běžně vzduchu, přičemž toto vlnění je vybuzeo mluvním ústrojím. Na výsledné podobě produkované řeči, potažmo řečového signálu, se podílí jednotlivé části tohoto ústrojí. Obecně jej členíme na část respirační, fonační a artikulační [2]. Základem pro tvorbu řeči je výdechový proud vzduchu z plic (respirační část), který je nadále modifikován hlasivkami (fonační část) a artikulačním ústrojím nad hlasivkami, tzn. nadhrtanovými dutinami a artikulačními orgány.

Vyprodukovaný řečový signál ve formě mechanického vlnění zachycujeme mikrofonem, čímž jej převádíme na spojitý elektrický signál a ten dále zesilujeme a vzorkujeme. Výslednou posloupnost hodnot již můžeme počítačově analyzovat.

### 2.2 Krátkodobá analýza

Řečový signál je z celkového pohledu signál nestacionární. V čase se však mění pomalu, což nám dává možnost vycházet z metod krátkodobé analýzy spočívající v analýze krátkodobého úseku řečového signálu o délce maximálně několika desítek ms, kdy signál považujeme za stacionární a můžeme přibližně určovat jeho vlastnosti. Jak bylo zjištěno, délka úseků má vliv na výsledky zpracování daného signálu. Po segmentaci je zde možnost aplikovat váhovací okna (při zpracování řeči často používáno pravoúhlé a Hammingovo), která mají za cíl omezit problémy při spektrálním prosakování energie.

#### 2.2.1 Mikrosegment

Stacionarita je základním předpokladem pro relevanci matematické analýzy segmentu. U některých metod analýzy periodicity řečového signálu je tedy pro získání určitých informací nutné signál rozčlenit na specifické mikrosegmenty s předem danými vlastnostmi. Zkoumáme-li řeč z fonetického či akustického hlediska, můžeme jej členit na základní řečové jednotky.

##### 2.2.1.1 Fonetická klasifikace

Zkoumáme-li řeč z fonetického hlediska, můžeme řečový signál rozdělit na základní řečové jednotky zvané hlásky, jež jsou definovány jako soubor foneticky podobných zvuků vyslovených člověkem [3]. Každá z těchto hlásek je vytvářena specifickým stavem či procesem řečového ústrojí a z hlediska fonologického se dále dá členit na nejmenší lingvistické jednotky – fonémy, které postihují jemné rozdíly v rámci použití jedné hlásky, např. [n] a [ŋ]. Řečový signál jako celek tedy můžeme v časové oblasti

rozdělit na celky odpovídající jednotlivým hláskám, místy oddělenými úseky ticha. Hlásky podle jejich akustického charakteru dělíme na souhlásky a samohlásky, viz kap. 2.4.1.

### 2.2.1.2 Akustická klasifikace

Pohlížíme-li na signál z akustického hlediska, zajímají nás jeho vlastnosti v časové oblasti. Pro určení podoby mikrosegmentu zde můžeme vycházet ze znalosti fonetické klasifikace - určité hlásky vykazují kvaziperiodický průběh a skládají se z krátkodobých period, jež se v čase poněkud liší tvarem, a mírně také délkou. Tato periodicitu vzniká ve fonačním ústrojí kmitáním hlasivek. Následná artikulace dané hlásky dodá vyšší harmonické složky a diferencuje ji od jiných hlásek. Můžeme tedy určité hlásky rozdělit na jejich základní periody.

## 2.3 Periodicita

Za ideálních podmínek má pojem periodicita binární charakter, tzn. signál buďto periodický je, nebo není. V reálném světě je však nutné vzít v potaz i veškeré vlivy při buzení či přenosu signálu. Jejich působením vzniknou odchylky v periodách signálu, jmenovitě odchylky v trvání jednotlivých period a v jejich okamžitých hodnotách. Perturbace trvání a amplitudy periody jsou způsobeny charakterem reálného lidského mluvního ústrojí, zejména přechodovými jevy při tvoření jednotlivých hlásek či přechody mezi nimi (tzv. koartikulace). Také se zde projevuje kolísání dané vedlejšími vlivy, např. emocionálním stavem řečníka, jehož projevem může být „roztřesený hlas“, nebo jeho zdravotním stavem či věkem. V zahraniční literatuře se tyto odchylky označují jako jitter a schimmer [3]. Reálné signály tudíž vlivem těchto odchylek místo ideální periodicity vykazují kvaziperiodický charakter.

Kvaziperiodicitou rozumíme téměř periodický proces, jenž se projevuje periodami vykazujícími jisté odchylky ve svých průbězích, a je zcela typická pro neideální prostředí reálného světa; řečový signál znělých (kvaziperiodických) hlásek můžeme na nejnižší úrovni rozdělit na mikrosegmenty o jednotlivých krátkodobých periodách. Tyto odchylky se dají kvantifikovat a změřit a tímto pojem periodicita získává nový význam – spíše než jako binární vlastnost ji lze vnímat jako míru přiblížení se ideální periodicitě.

Pro bližší zkoumání kvaziperiodických řečových signálů a jejich periodicity je třeba stanovit veličiny, jež mohou popsat jitter a schimmer. Jitter, neboli odchylky v trvání jednotlivých period, postihuje parametr intonace, v zahraniční literatuře označované jako pitch. V mluvené řeči je vnímána jako melodie a pohybuje se v rozmezí 50 – 600 Hz. Schimmer, neboli odchylky v okamžitých hodnotách jednotlivých period, lze postihnout parametrem vzájemné podobnosti sousedních period. V mluvené řeči lze schimmer často zaznamenat jako tzv. třesoucí se hlas.

## 2.4 Analyzované veličiny

### 2.4.1 Znělost a neznělost

České hlásky můžeme dělit na znělé a neznělé podle míry jejich periodicity, resp. podle míry přítomnosti periodického průběhu dané hlásky. Znělé hlásky jsou charakteristické kvaziperiodickým průběhem, tzn. přítomností krátkodobé periody vznikající přibližně pravidelným kmitáním hlasivek při jejich vytváření a volným průchodem vzduchového proudu rezonančními dutinami řečového ústrojí. Obsahují tedy základní frekvenci  $F_0$ , tedy tón. Pro neznělé hlásky je typická neúčast hlasivek na jejich tvorbě. Z akustického hlediska má jejich průběh charakter šumu, tzn. nejsou složeny z krátkodobých period. K tomu dochází vlivem znesnadňování průchodu vzduchového proudu řečovým ústrojím, zpravidla buďto stavěním překážky (např. jazyku) nebo zúžením hlasového traktu [3].

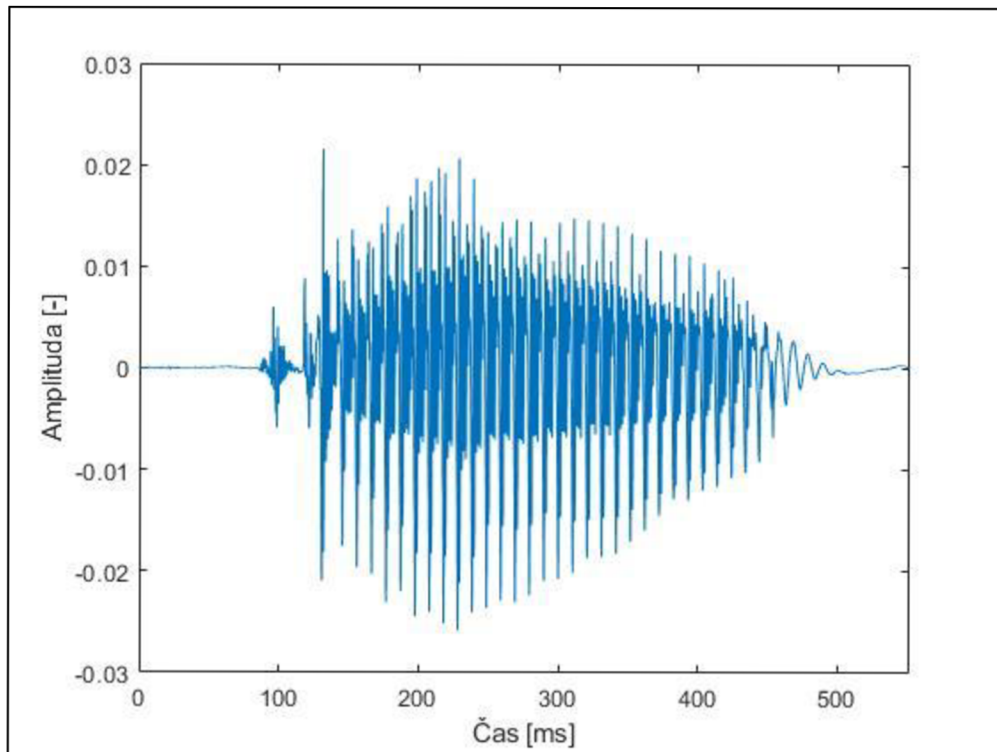
Český jazyk dělí hlásky na samohlásky a souhlásky. Všechny samohlásky jsou znělé. Oproti tomu souhlásky, ačkoli je pro ně typický šumový charakter a bývají neznělé, mohou obsahovat také periodickou složku a být klasifikovány jako znělé [3].

Párové	neznělé	/p/	t/	/t'/	/k/	/f/	/s/	/š/	/ch/	/c/	/č/	/ř/
	znělé	/b/	/d/	/d'/	/g/	/v/	/z/	/ž/	/h/	/dz/	/dž/	/ř/
Nepárové	znělé	/m/, /n/, /ň/, /l/, /j/, /r/										

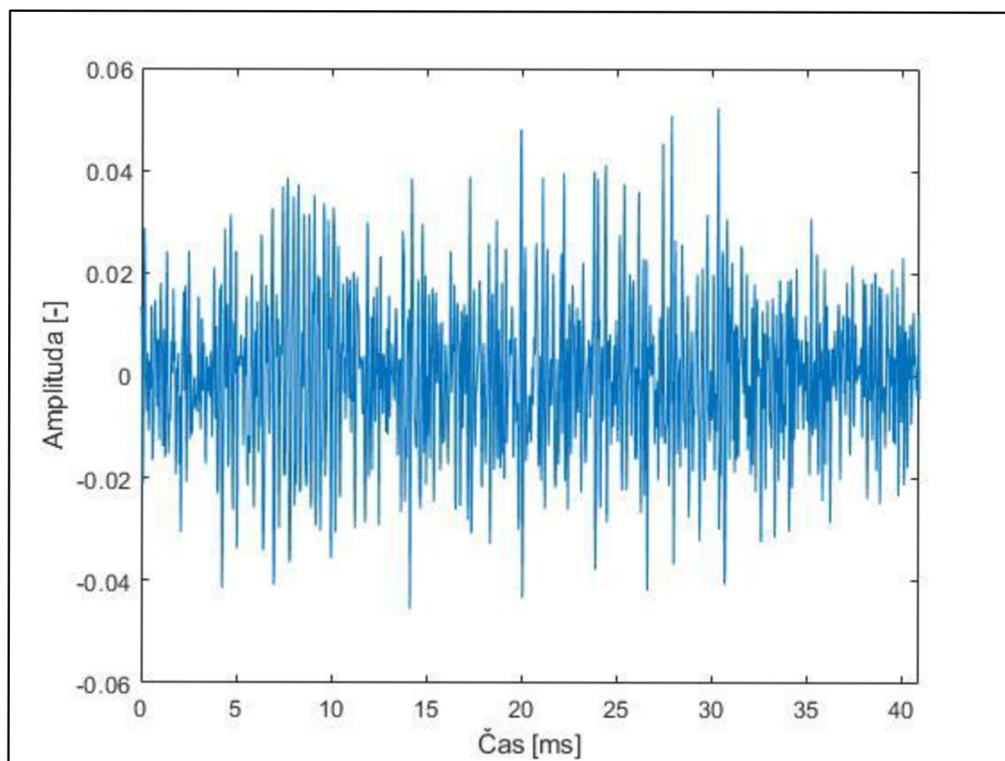
Tab. 1 - Znělé a neznělé souhlásky, převzato z [3]

Na obr. 2-1 můžeme vidět signál hlásky /a/ trvající přibližně 400 ms. Hlávka /a/ je znělá – v čase vykazuje kvaziperiodický průběh s počátečním náběhem do cca 230 ms a následným poklesem a odezníváním. Zároveň můžeme pozorovat výrazně vysokou hodnotu amplitudy periody na cca 130. milisekundě. Tato výrazná hodnota je způsobena uvolněním okluze v hlasívkách před jejich vybuzením (tzv. exploze) [3].

Obr. 2-2 je časovou reprezentací neznělé souhlásky /s/. Odlišnost od znělé souhlásky /a/ je zřejmá – signál je šumového charakteru, jeho hodnoty v čase jsou náhodné. Jde tedy o hlávku neznělou. Artikulací se /s/ řadí mezi frikativa neboli třené souhlásky. Vzniká vytvořením neúplné překážky přiblížením dvou artikulačních orgánů, zde jazyku a horního patra. Takto vznikne úžina a výdechový proud nabude charakteru náhodného šumu vlivem turbulentního proudění [4].



**Obr. 2-1 - Časová reprezentace hlásky /a/**



**Obr. 2-2 - Časová reprezentace hlásky /s/**



## 2.4.2 Intonace

Jitter je měřítkem odlišnosti časového trvání mezi jednotlivými krátkodobými periodami. Projevuje se intonací, již rozumíme melodii řeči. Melodie, resp. tón úseku řečového signálu je dán základní frekvencí  $F_0$ . Z toho tedy lze vyvodit, že určování intonace řeči má smysl pouze pro úseky znělé, které mají periodický charakter. Tónový rozsah řeči je udáván v rozmezí 60–400 Hz, v krajních případech může frekvence klesnout k 50 Hz u hlubokých hlasů a u dětského hlasu stoupnout až k 600 Hz [3]. Intonace je během řeči modifikována jak vědomě, např. formou přízvuku, tak nevědomě, např. při zvýšení hlasu vlivem stresu. V těchto případech bývá způsoben nedostatečnou kontrolou kmitání hlasivek [5]. Intonace řeči je tedy parametr, z nějž lze vyvozovat další informace o mluvčí a promluvě, např. emocionální stav řečníka.

## 2.4.3 Podobnost

Měření podobnosti dvou souborů dat je obecně věcí měřítka podobnosti, které může mít různé definice, a jednotlivá měřítka, resp. metody měření podobnosti, mohou udávat různé výsledky. Výsledkem by však vždy mělo být číslo pro daná data vyjádřitelné v procentech, které nám o vzájemné podobnosti dá relevantní informaci.

Hledání odlišností krátkodobých period v průběhu signálu má velké praktické využití zejména ve zdravotnictví při analýze kvality signálu EKG, kdy je takřka nemožná manuální kontrola každé krátkodobé periody signálu z důvodu jejich velkého počtu. Jak uvádí [6], podobnost dvou objektů, v tomto případě krátkodobých period signálu, lze určit pomocí koeficientů korelace.

Schimmer je měřítkem odlišnosti hodnot amplitudy mezi krátkodobými periodami. Z toho vyplývá, že v rámci řečového signálu má tedy smysl se zabývat znělými úseky signálu, které mají kvaziperiodický charakter a obsahují měřitelné krátkodobé periody. Budeme-li předpokládat, že analyzovaná data, tedy jednotlivé periody signálu, nejsou ovlivněna jitterem, můžeme analyzovat jejich vzájemnou podobnost.

Pro praktické využití (zejména ve vztahu k periodicitě řeči) má velký význam analýza podobnosti sousedních period signálu. Můžeme tímto získat informaci o závislosti perturbací amplitud na čase. Odtud lze vyvozovat další informace o řečníkovi a jeho emocionálním či zdravotním stavu, případně o jeho věku; schimmer je spojen se zvýšeným odporem hlasivek vzduchovému proudu a jejich onemocněním [5]. Zároveň popisuje zvyšování intenzity hlasu při promluvě, ať už vědomě kvůli např. přízvuku, případně nevědomě při roztřeseném hlasu způsobeného stresem. Obecně - čím je vyšší podobnost sousedních period v průběhu celého signálu, tím blíže signál je ke stavu ideální periodicity, tedy jeho periodičita jako míra přiblížení se ideálnímu stavu se zvyšuje.

## 2.5 Existující studie

Vzhledem k praktickému významu zmíněných parametrů při analýze řečových signálů je logické, že jim byl věnován poměrně rozsáhlý výzkum, jehož výsledkem je množství algoritmů sloužících k jejich určování. Tyto algoritmy jsou založeny na vícero různých matematických metodách a jejich výsledky se mohou lišit.

Studie zabývající se parametrem znělosti/neznělosti jsou poměrně četné, např. [8], kde jsou mimo jiné prezentovány metody STE a ZCR, porovnávající neznělé segmenty s šumem. Tyto metody jsou zmíněny také v české publikaci Mluvíme s počítačem česky [3] a v práci [7] autorů Bachu a kol., která tyto dvě metody kombinuje v jednu. Metoda STE je prezentována jako citlivá na dynamiku signálu a při jeho velkých změnách i nepřesná, ZCR ve svých průměrných hodnotách nemusí být precizní z důvodu širokopásmovosti řečového signálu. Kombinace těchto metod v [7] však vzhledem k jejich jednoduchosti vrací dobré výsledky. V práci J. Heranové [9] je k určování znělosti/neznělosti užitá metoda HNR, jenž, podobně jako předchozí dvě, pracuje s analogií šumu a neznělosti, přičemž vrací přesné výsledky. Zároveň ji využívá softwarový nástroj Praat pro určení míry aditivního šumu. Všechny zmíněné metody jsou charakteristické svou nízkou výpočetní náročností, nicméně vzhledem k jejich podstatě, podrobněji popsané v kap. 3, může být správnost výsledků ovlivněna i výrazným aditivním šumem.

Určování intonace je rovněž často studováno, existuje množství prací zabývajících se metodami k jejímu určování, a to nejen v rámci řečových signálů, ale i např. v rámci signálů hudebních, kde je typickým praktickým příkladem běžná ladička hudebních nástrojů. Některé z metod uvádí článek [14] a testuje P. Motlíček v [15]; základní metodou v této oblasti je ACF a její modifikace, schopné určit dobu krátkodobé periody a tím i intonaci. Jejich nepřesnosti mohou záviset na částech signálu, u nichž byla předtím nepřesně určena znělost či neznělost. Ačkoli je metoda jednoduchá, pro aplikace v reálném čase je výpočetně náročná. Tento problém eliminuje algoritmus na bázi metody AMDF. Tou se v praktickém použití zabývá také [13] a popisuje ji jako nenáročnou a snadno implementovatelnou, ovšem podle [14] je náchylná na podobné nepřesnosti jako ACF a vykazuje i méně přesné výsledky. Za velmi přesnou metodu je v [15] označena metoda NCC, která vrací lepší výsledky díky schopnosti lépe sledovat strmé změny v signálu [14].

Výběr metod ZCR, STE a HNR byl dán zejména jejich celkovou jednoduchostí, zároveň pro ně existují příslušné publikace v českém jazyce. Metody ACF a AMDF byly vybrány na základě jejich rozšířenosti v aplikacích k určování základní frekvence, NCC díky předpokládanému rapidnímu zlepšení performance algoritmu za cenu jednoduchých úprav.

Publikace ohledně určování podobnosti sousedních period se takřka nevyskytují, metody pro analýzu jitteru touto cestou v časové oblasti tedy zřejmě nejsou příliš rozšířeny. Při návrhu metod pro tuto práci byla výchozím bodem práce S. Pospíšila [6],

která rozvádí možnost určení podobnosti mezi dvěma objekty pomocí měřítek podobnosti. Ohledně těch je již více dostupných publikací; v [10] a [11] je jako jedno z měřítek zmíněn korelační koeficient, který je v oblasti zpracování signálů v souvislosti s podobností velmi rozšířen. Podobně lze využít také absolutní diferenci signálů, získanou metodou AMDF [12]. Práce [16] jako další metodu měření podobnosti zmiňuje DTW, jenž byla navržena speciálně pro řečové signály.

Pro správné fungování naprogramovaných algoritmů (popsaných v kap. 4.1) byly zároveň užity také výše zmíněné metody určené pro analýzu základní frekvence.

## 3. VÝPOČETNÍ METODY

V následujících podkapitolách budou představeny metody užívané v naprogramovaných algoritmech pro analýzu periodicity řeči.

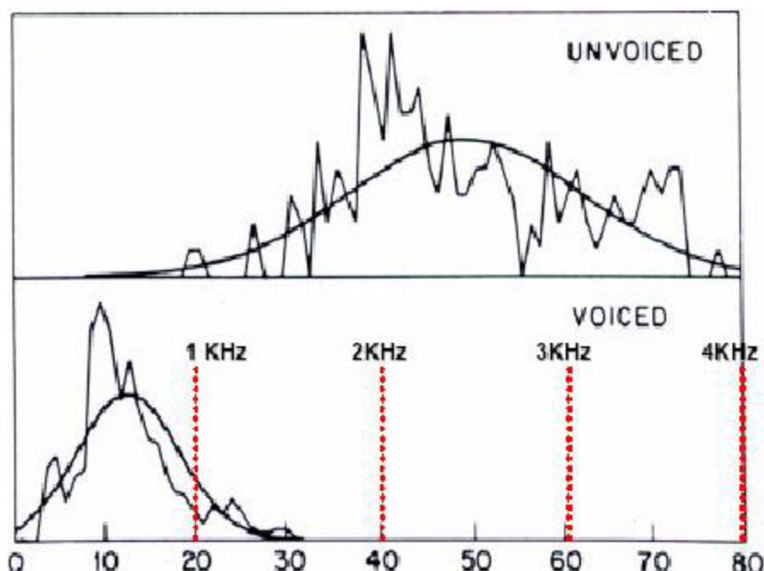
### 3.1 Zero Crossing Rate

Zero Crossing Rate, obvykle značeno ZCR a česky označováno jako *funkce středního počtu průchodů signálu nulou*, je často používaným nástrojem v oblasti zpracování řečových signálů. Jedná se o měřítko počtu průchodů nulou, resp. vyhodnocení, jak často v daném úseku řečového signálu dochází ke změně znaménka okamžité hodnoty tohoto signálu.

$$ZCR = \sum_{n=0}^{N-1} |\text{sgn}[x(n)] - \text{sgn}[x(n-1)]| \quad (1)$$

kde  $N$  je počet vzorků okna,  $\{x(n)\}$  je  $n$ -tý vzorek okna a  $\text{sgn}$  je funkce signum.

Neznělé hlásky mají šumový charakter, přičemž šum se typicky nachází na vyšších frekvencích. Jak uvádí Bachu a kol. v [7], energie znělé řeči je koncentrována pod 3 kHz. Můžeme tedy předpokládat, že hlásky šumového charakteru budou vykazovat vyšší míru průchodu nulou než na frekvenční ose níže položené hlásky znělé, a lze je tímto rozlišit. Obr. 3-1 zobrazuje histogram průměrného počtu průchodů nulou pro znělou a neznělou řeč nad intervaly o 10 ms.



Obr. 3-1 - Distribuce průchodů nulou pro znělou a neznělou řeč [7]

### 3.2 Short Time Energy

Short Time Energy, obvykle značeno STE a česky označováno jako funkce krátkodobé energie, udává hodnotu energie krátkodobého segmentu řečového signálu. Jedná se o další z užívaných metod pro určování znělosti/neznělosti v řečových signálech [3] [8].

$$STE = \sum_{n=0}^{N-1} [x(n)]^2 \quad (2)$$

kde  $N$  je počet vzorků okna a  $\{x(n)\}$  je  $n$ -tý vzorek okna.

Vzhledem k tomu, že amplituda znělé části signálu je zpravidla vyšší než amplituda části neznělé, lze tuto metodu využít k jejich odlišení, přičemž předpokládáme, že pro znělou část získáme výrazně vyšší hodnotu STE. Metoda je však citlivá na dynamiku signálu a při velkých změnách může být nepřesná [3].

Podobně jako u funkce ZCR bude vstupem algoritmu vstupní signál, vzorkovací frekvence a délka analyzovaného okna. Signál bude segmentován a pro každý segment bude vypočteno STE podle rovnice xy. Pro prahování lze opět využít bílý šum - proběhne výpočet hodnoty jeho STE, která budiž považována za práh. Výstupem bude binární funkce znělosti/neznělosti a vstupní signál s vyfiltrovanými neznělými segmenty.

### 3.3 Harmonics-to-Noise Ratio

Harmonics-to-Noise Ratio (HNR) je parametrem popisujícím harmonicitu signálu, tedy poměr harmonických a šumových složek v jeho spektru. Podle [9] lze tento poměr vyjádřit pomocí autokorelační funkce daného signálu. Existuje-li výrazné lokální maximum autokorelační funkce daného signálu mimo maximum pro nulový posun, předpokládáme u signálu periodický charakter, přičemž normalizovaná hodnota tohoto lokálního maxima udává relativní výkon harmonické složky a její komplementární hodnota relativní výkon složky šumové. Logaritmus poměru těchto dvou hodnot získáme výsledné HNR. Matematicky jej lze definovat jako:

$$HNR = 10 \log_{10} \frac{r(\tau_{peak})}{1 - r(\tau_{peak})} \quad (3)$$

kde  $r(\tau_{peak})$  je normalizovaná hodnota druhého extrému autokorelační funkce,  $\tau$  je posun ve vzorcích.

Metoda HNR je velmi rozšířená v oblasti zkoumání parametrů hlasu za účelem lékařské diagnostiky, např. chrapotu a indexu stárnutí hlasu [5]. V [9] je metoda úspěšně použita pro určení hranic mezi znělými a neznělými částmi signálu.

### 3.4 Autokorelační funkce

Autokorelační funkce (ACF) je funkcí v oblasti zpracování řeči typicky užívanou k detekci periodických struktur ve vstupním signálu. Je definována jako:

$$R(\tau) = \sum_{n=0}^{N-1-\tau} x(n)x(n + \tau) \quad (4)$$

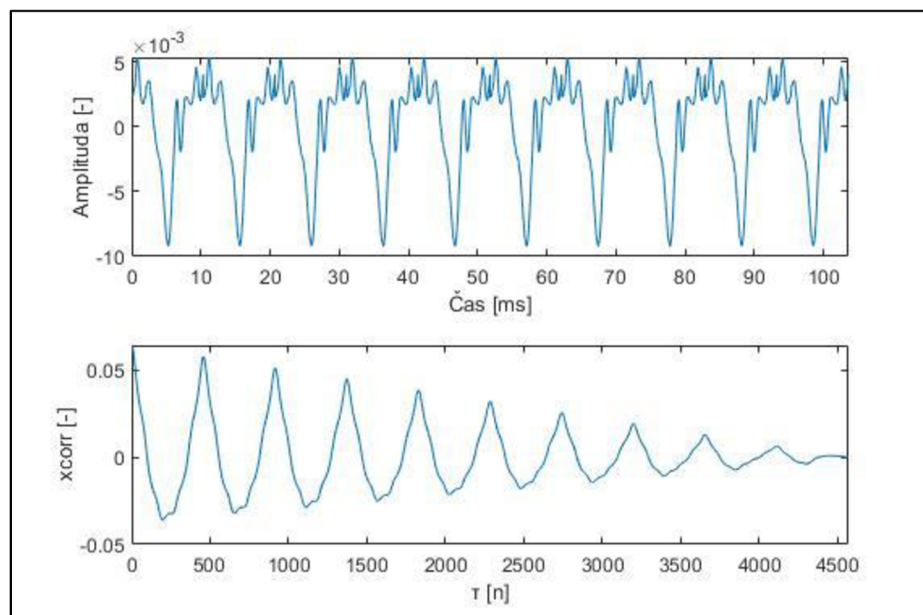
kde  $N$  je počet vzorků okna,  $\{x(n)\}$  je  $n$ -tý vzorek okna a  $\tau$  je posun ve vzorcích.

Jedná se tedy o speciální případ korelační funkce, kdy vstupní signály jsou totožné. Korelační funkce je jedno z nejběžnějších měřítek podobnosti [10] [11] a při daném posuvu nám udává číslo reprezentující hodnotu podobnosti mezi dvěma signály. Lze jí tedy využít při určování perturbací v amplitudách sousedních period.

Je-li tedy periodická struktura ve zkoumaném signálu skutečně přítomna, podle [14] se v její výsledné autokorelační funkci v polohách odpovídajícím násobkům hodnoty základní periody  $T_0$  nacházejí lokální maxima, protože na těchto pozicích jsou vynásobeny hodnoty dvou zcela překrývajících se totožných krátkodobých period signálu.

Výsledná autokorelační funkce nám tedy může dát velmi užitečnou informaci o přítomnosti periodické struktury a zejména o hodnotě základní periody  $T_0$  vstupního signálu. Jsou-li správně určeny vzdálenosti periodických maxim autokorelační funkce v počtu vzorků posunu  $\tau$ , odpovídají svou číselnou hodnotou velikosti základní periody daného vstupního signálu.

V prostředí MATLAB ver. 2019 je tato funkce zahrnuta, viz *autocorr*.



Obr. 3-2 - Jednostranná autokorelace dokonale periodického signálu

### 3.5 Average Magnitude Difference Function

Average Magnitude Difference Function (AMDF) je variantou autokorelační funkce. Byla navržena jako alternativa ke korelační funkci, vhodná pro strojové výpočty, protože operaci násobení nahrazuje operací rozdílu, což ji činí jednodušší. Je definována jako:

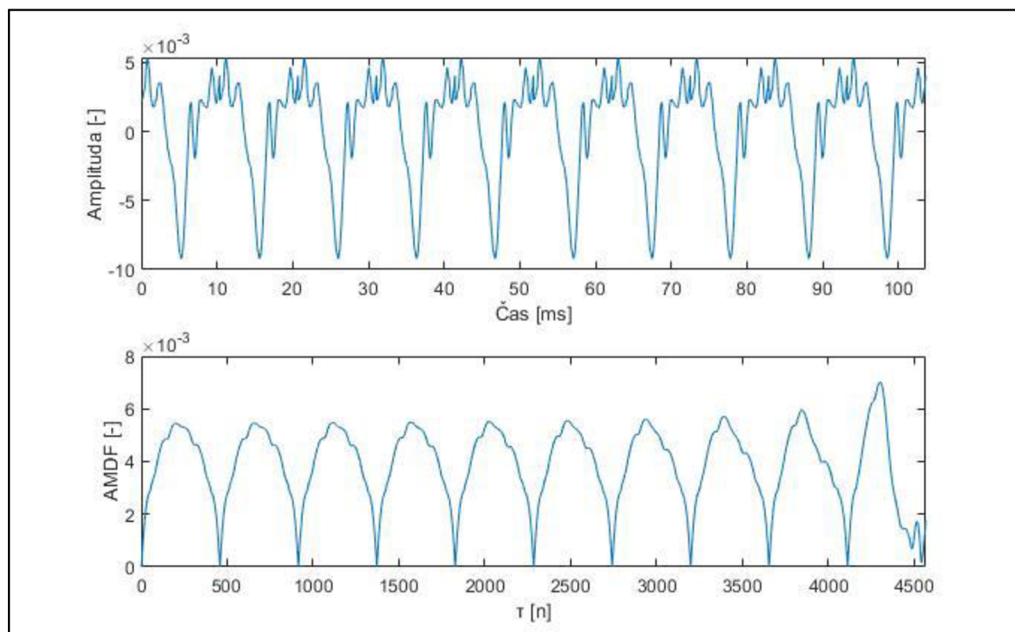
$$\text{AMDF}(\tau) = \sum_{n=0}^{N-1-\tau} |x(n) - x(n + \tau)| \quad (5)$$

kde  $N$  je počet vzorků okna,  $\{x(n)\}$  je  $n$ -tý vzorek okna a  $\tau$  je posun ve vzorcích.

Rozdíl mezi AMDF a ACF spočívá v tom, že AMDF udává vzdálenost mezi dvěma objekty [12]. Je jí možné použít pro určení absolutní difference dvou signálů, jejíž hodnotu získáme pro nulový vzájemný posun vstupních signálů. Pokud jsou oba signály totožné, bude výsledná rozdílová funkce nulová. Při odchylkách ve funkčních hodnotách mezi oběma signály, tzn. za přítomnosti jitteru, získáme hodnotu absolutní difference, již můžeme použít pro kvantifikaci jejich podobnosti.

Narozdíl od ACF se ve výsledné rozdílové funkci na násobcích hodnoty základní periody  $T_0$  nenacházejí lokální maxima, ale minima. Tato funkce nám tedy opět může dát informaci o výskytu periodicity ve vstupním signálu [13]. Pokud je průběh vstupní funkce dokonale periodický (se shodnými krátkodobými periodami), hodnota tohoto minima, resp. absolutní difference, je nulová.

Tato funkce není součástí MATLABu, byly naprogramovány normalizované verze *AMDF.m* a *twosignalAMDF.m*.



Obr. 3-3 - AMDF dokonale periodického signálu

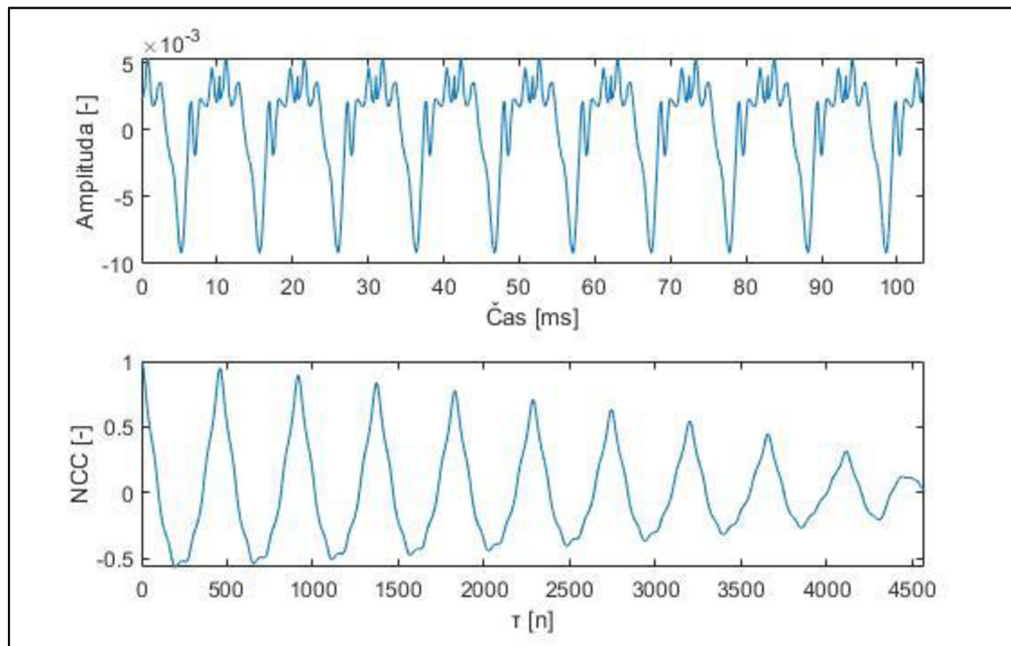
### 3.6 Normalized cross-correlation

Normalized cross-correlation (NCC) je další variantou ACF. Je jí velmi podobná, podle [14] však nabízí přesnější výsledky než ACF díky normalizaci, která snižuje vliv poklesu autokorelační funkce a je tak možný přesnější odhad poloh lokálních maxim a tedy i základní frekvence signálu [15]. Je definována jako:

$$NCC(\tau) = \frac{\sum_{n=0}^{N-\tau-1} x(n)x(n+\tau)}{\sqrt{\sum_{n=0}^{N-\tau-1} x^2(n) \times \sum_{n=0}^{N-\tau-1} x^2(n+\tau)}} \quad (6)$$

kde  $N$  je počet vzorků okna,  $x(n)$  je  $n$ -tý vzorek okna a  $\tau$  je posun ve vzorcích.

V prostředí MATLAB ver. 2019 je tato funkce integrována (viz *normxcorr2*), byla však naprogramována podobná funkce *NCC.m*.



Obr. 3-4 - Jednostranná normalizovaná cross-korelace dokonale periodického signálu

### 3.7 Dynamic Time Warping

Dynamic Time Warping (DTW), v překladu metoda dynamického borcení času, je starší, ale stále používaná metoda, původně navržená pro porovnávání dvou souborů časových dat reprezentujících právě řečové signály [16]. Její typickou vlastností je nelineární roztažení jednoho ze souborů dat tak, aby vzdálenosti mezi odpovídajícími

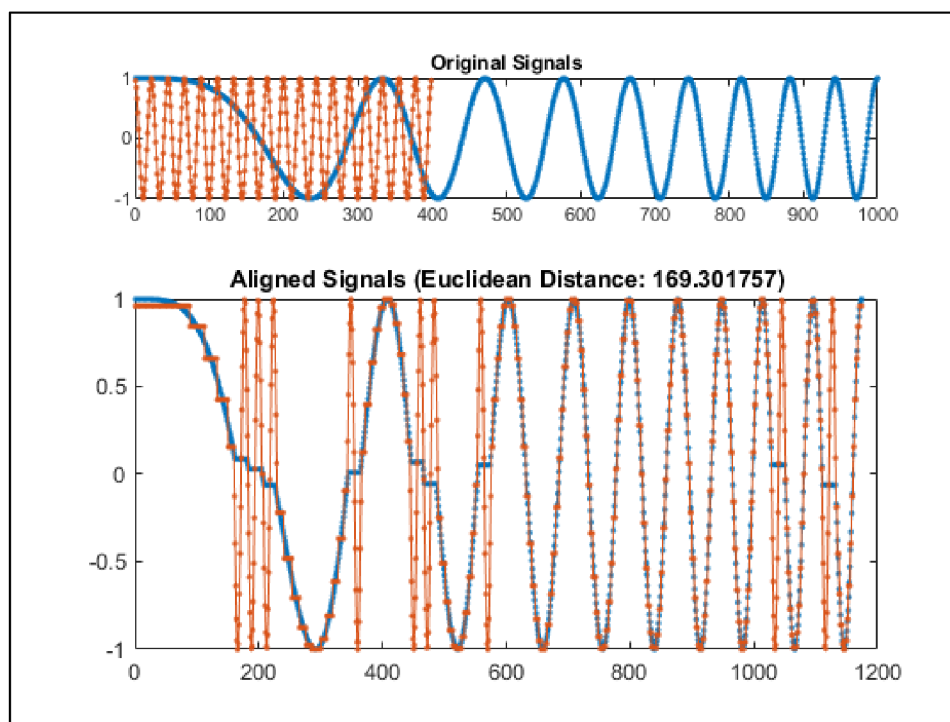


dvojicemi bodů v obou souborech vykazovaly nejkratší vzdálenost, a to zpravidla Euklidovskou, ačkoli existují další varianty této metody používající jiné metriky; tímto můžeme získat číselný výsledek reprezentující podobnost mezi porovnávanými daty [10]. Roztažení probíhá na základě jistých pravidel, jeho podrobný rozbor je však svým rozsahem nad rámec této práce. Obr. 3-5 ilustruje roztažení souboru dat na délku druhého při zajištění nejkratší Euklidovské vzdálenosti.

$$d(p, q) = \sqrt{\sum_{n=0}^N (p(n) - q(n))^2} \quad (7)$$

kde  $d$  je Euklidovská vzdálenost,  $p$ ,  $q$  jsou soubory dat,  $N$  je počet vzorků souboru dat a  $p(n)$ ,  $q(n)$  jsou odpovídající dvojice bodů vyhodnocené algoritmem nelineárního roztažení.

Funkce je v prostředí MATLAB zahrnuta (*viz dtw*), vstupem jsou dva porovnávané soubory dat a výstupem je hodnota nejkratší Euklidovy vzdálenosti, tzn. skalární hodnota podobnosti mezi nimi, již je dále možné normalizovat. Jsou-li porovnávané signály totožné, je vzdálenost nulová a podobnost maximální.



**Obr. 3-5 - Znárodnění roztažení funkce pomocí DTW, převzato z nápovědy MATLAB**

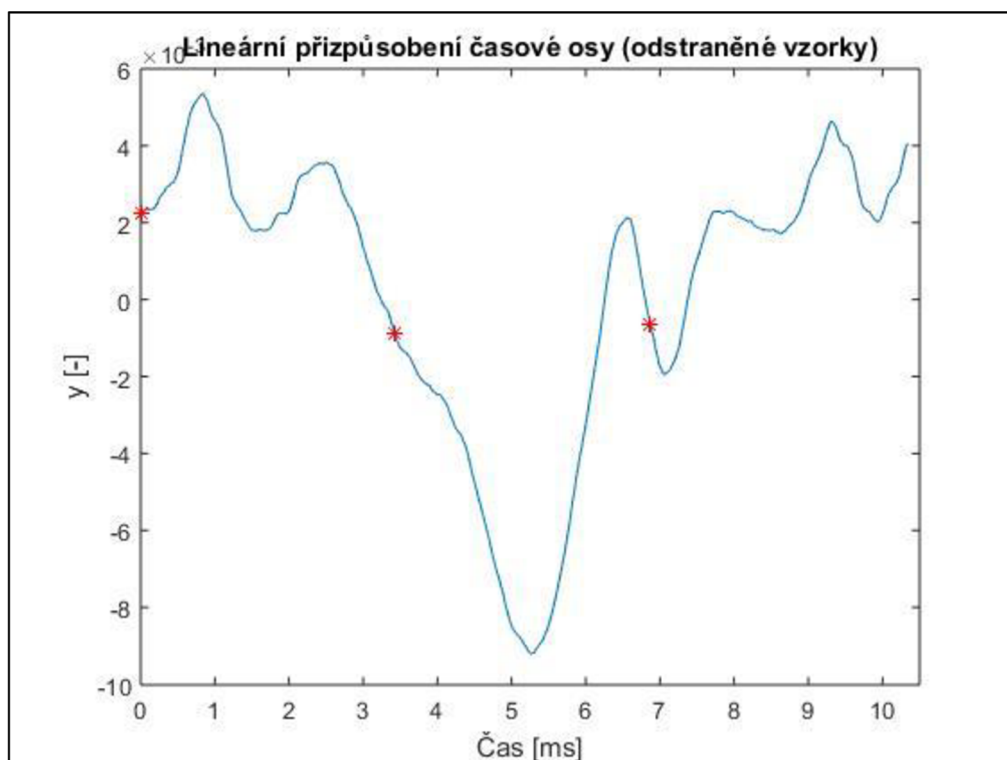
### 3.8 Přizpůsobování časové osy

Bude-li třeba porovnávat dva soubory dat v časové doméně pomocí funkcí pro určení vzájemné podobnosti za účelem určení odchylek v schimmeru, bude v řadě případů nutné tato data zpracovávat v úplném překryvu, tzn. pro nulovou hodnotu vzájemného posunu. Zároveň bude nutné zajistit, aby délka těchto dvou souborů dat byla zcela totožná. Vstupní soubory dat však takové zpravidla nejsou.

V této práci budou vstupními soubory dat pro analýzu schimmeru zpravidla dvě sousední periody kvaziperiodického úseku signálu; označme je jako A a B. Je-li jejich délka odlišná, nelze získat kýženou diferenci, protože se tyto periody stoprocentně nepřekrývají. Abychom problém možného jitteru vyřešili, je třeba delší z period B zkrátit na délku kratší periody A. Toho můžeme docílit více způsoby.

#### 3.8.1 Lineární přizpůsobení časové osy

Lineární přizpůsobení spočívá v jednoduchém „vypuštění“ určitého počtu vzorků dané periody B. Nadbývá-li např. o  $n$  vzorků, vypustíme  $n$  jejich vzorků lineárně rozložených v jejím průběhu. V naší metodě to provedeme rovnoměrným rozdělením periody na  $n$ -tiny a odstraníme vždy první vzorek tohoto segmentu, viz obr. 3-6.



Obr. 3-6 - Lineární přizpůsobení časové osy

### 3.8.2 Nelineární přizpůsobení časové osy

Nelineární přizpůsobení je sofistikovanější způsob zkrácení dané periody o  $n$  bodů. Perioda je opět rovnoměrně rozdělena na  $n$ -tiny, následně jsou postupně vypouštěny body v této  $n$ -tině a při každém vypuštění je testována podobnost s periodou A. Pro periodu B s nadbývajícím např. třemi vzorky je postup je následující:

- 1) Perioda je rovnoměrně rozdělena na třetiny.
- 2) Jsou odstraněny první vzorky druhého a třetího segmentu, podobně jako při lineárním přizpůsobení.
- 3) Postupně jsou odstraňovány vzorky v prvním segmentu, pro každé odstranění je pomocí AMDF testována podobnost s periodou A, výsledné hodnoty jsou ukládány. Ze souboru hodnot podobností je nakonec vybrána nejnižší hodnota a je odstraněn vzorek s jejím indexem.
- 4) Totéž se provádí ve druhém a třetím segmentu, ovšem v již zpracovaných segmentech (zde v prvním) je vzorek již odstraněn, neprobíhá v nich tudíž odstranění prvního vzorku podle bodu č. 2.

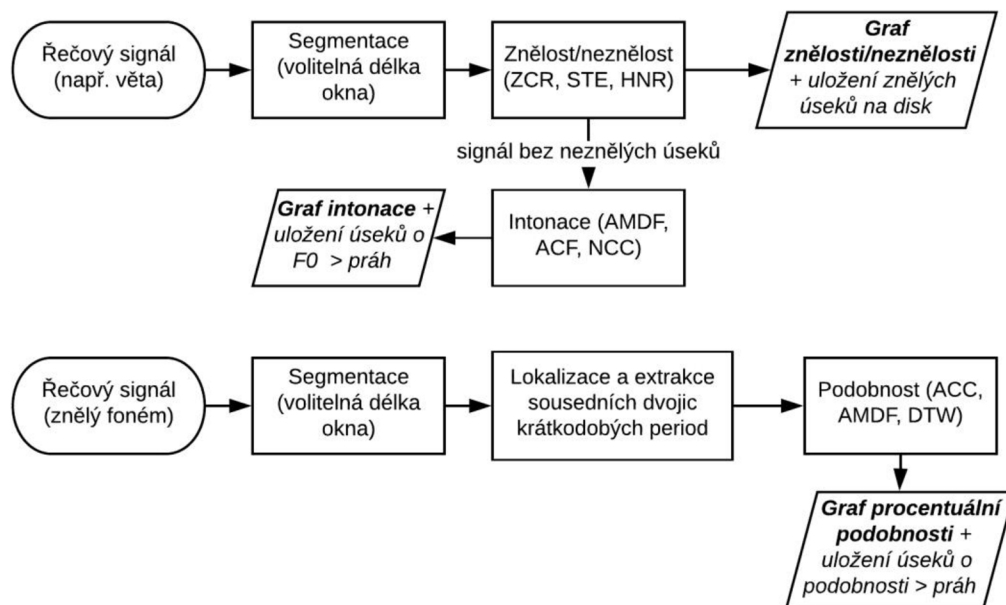
Výsledkem této procedury je přizpůsobení délky periody B na kýženou délku periody A odstraněním vzorku v každé  $n$ -tině této periody, získaném rovnoměrným rozdělením, přičemž jsou odstraněny takové vzorky, aby byla co nejvíce zvýšena podobnost period A a B.

## 4. PRAKTICKÁ ČÁST

Cílem této práce je bližší prozkoumání jevu kvaziperiodicity v řečovém signálu a v rámci praktické části vytvoření autonomního programu pro analýzu tohoto jevu v reálných nahrávkách řeči, jehož výstupem má být grafické znázornění analyzované veličiny popisující kvaziperiodicitu. Tento autonomní program má k analýze každého z uvedených parametrů periodicity nabízet možnost výběru tří metod, jež jsou široce používány v oblasti zpracování řečových signálů, a má být doplněn o modul ukládání segmentů řečového signálu na disk. Celkově se tedy jedná o 3 metody k určení znělosti/neznělosti + 3 metody pro analýzu jitteru a 3 pro analýzu schimmeru.

### 4.1 Popis programu

Program byl vytvořen v prostředí MATLAB ver. 2015 a 2019. Jeho strukturu popisuje blokové schéma, viz obr. 4-1. Vstupy programu jsou příslušné řečové signály; pro zjištění znělosti a intonace může být vstupem jakýkoliv řečový signál. Pro analýzu podobnosti je vstupem předzpracovaný signál bez neznělých částí, u něhož předpokládáme kvaziperiodický průběh, zpravidla tedy signál reprezentující znělou hlásku. Kýženým výstupem je vždy graf reprezentující analyzovanou vlastnost signálu.



Obr. 4-1 - Blokové schéma programu

Součástí výstupu je zároveň uložení určitých úseků zpracovaného signálu na disk. Tyto úseky jsou pro nebinární parametry, tzn. pro podobnost a intonaci, určeny volitelným prahem; uloženy jsou tedy pouze ty úseky, v nichž daný parametr spojitě přesahuje nastavený práh, např. úseky, kde fundamentální frekvence přesahuje 90 Hz nebo kde podobnost sousedních period přesahuje 90%. U analýzy znělosti/neznělosti, která má binární charakter, jsou uloženy pouze spojitě vykazující znělost.

## 4.1.1 Segmentace

Po načtení signálu a jeho vzorkovací frekvence do proměnných  $y$  a  $F_s$  pomocí integrované funkce *audioread* je nutné vstupní signál segmentovat, protože vycházíme z metody krátkodobé analýzy. Za tímto účelem byla naprogramována funkce *SEG\_A\_DOPLN.m*, jenž na základě zvolené délky segmentu v milisekundách vstupní signál nejprve doplní nulami a poté rozdělí na matici  $Y$ , do jejíž sloupců jsou uloženy části signálu o zvolené délce.

## 4.1.2 Určení znělosti/neznělosti

### 4.1.2.1 ZCR

Jak uvádí Bachu a kol. [7], neznělé hlásky jsou na frekvenční ose energeticky koncentrovány výše, protože mají charakter šumu a ten obecně bývá vysokofrekvenční; můžeme tedy podle něj hlásky znělé a neznělé rozlišit pomocí metody ZCR. Byla naprogramována funkce *ZCR\_final.m*, jejímž vstupem je daný řečový signál. Ten je pomocí funkce *SEG\_A\_DOPLN.m* segmentován. Následně je vypočítána hodnota ZCR pro každý segment a pro šumový signál o stejné délce. ZCR šumu je prahováno na experimentálně zništěnou optimální hodnotu rovnající se desetina původní hodnoty. Korespondující hodnoty pro řečový signál a šum jsou porovnávány, dosahuje-li hodnota ZCR segmentu řeči stejné nebo vyšší hodnoty než odpovídající ZCR šumu, je tento segment vyhodnocen jako neznělý a jeho sloupec v matici  $Y$  je vynulován.

Výstupem funkce je tedy matice  $Y$ , u níž jsou zachovány pouze sloupce zaznamenávající znělé segmenty signálu, zbytek je vynulován. Dalšími výstupy je integrovanou funkcí *reshape* složený časový průběh vstupního signálu, jehož neznělé úseky mají nulovou hodnotu, a vektor *znelest* reprezentující znělost či neznělost jednotlivých segmentů hodnotami 1 a 0.

### 4.1.2.2 STE

Podle [3] je pro určení znělosti/neznělosti další vhodnou metodou STE. Vychází ze stejné myšlenky jako ZCR, tzn. že neznělý signál má charakter šumu, který má zpravidla nižší energii než signál o periodickém charakteru. Byla naprogramována funkce *STE\_final.m*, která pracuje velmi podobně jako předchozí funkce *ZCR\_final.m*; vstupní signál je opět zpracován pomocí funkce *SEG\_A\_DOPLN.m*, pro dané segmenty

je vypočtena hodnota krátkodobé energie a ta je následně porovnávána s krátkodobou energií šumového signálu o délce segmentu a v řádovém rozpětí jeho funkčních hodnot. Je-li energie segmentu rovna či větší než energie šumu, je segment vyhodnocen jako znělý. Výstupem je opět matice  $Y$ , z ní složený řečový signál bez neznělých částí a vektor *znelost*.

### 4.1.2.3 HNR

Metoda HNR je v [9] úspěšně využita pro určení hranic mezi znělými a neznělými částmi řečového signálu, zároveň patří mezi metody užívané programem Praat k určení míry aditivního šumu. Byla naprogramována funkce *HNR\_final*. Vstupní signál je opět segmentován funkcí *SEG\_A\_DOPLN.m*. Následná výpočetní procedura zahrnuje použití funkce *xcorr* k získání autokorelační funkce každého segmentu. Z této funkce je extrahována hodnota druhého maxima a ta je následně použita k výpočtu HNR, jež nabývá hodnot z intervalu  $(0,1)$ . Dosahuje-li pro daný segment hodnoty větší než 0,5, je segment vyhodnocen jako znělý, v opačném případě jako neznělý. Výstupem je matice  $Y$ , z ní složený řečový signál bez neznělých částí a vektor *znelost*.

## 4.1.3 Určení intonace

### 4.1.3.1 ACF

Vlastnosti autokorelační funkce jsou pro odhalení periodicity a určení hodnoty  $F_0$  velmi často využívány, např. [11]. Délka periody v počtu vzorků je dána vzdáleností prvních dvou lokálních maxim této funkce. Byla naprogramována funkce *Pitch\_ACF\_final.m*, jejímž hlavním vstupem je řečový signál. Ten je nejprve filtrován dolní propustí typu Butterworth s experimentálně zjištěnými optimálními parametry. Protože určování intonace má smysl pouze pro znělé části signálu, je vstupní řečový signál nejprve předzpracován funkcí na určování znělosti/neznělosti. Za tuto funkci je defaultně zvolena metoda STE, v programu ji však lze jednoduše přepsat na jakoukoliv z již dostupných funkcí.

Dále je pomocí integrované autokorelační funkce *autocorr* analyzován řečový signál s vynulovanými neznělými částmi. Z autokorelační funkce jsou následně pomocí funkce *findpeaks* zjištěny lokace prvních dvou maxim a z jejich vzdálenosti je výpočtem určena hodnota  $F_0$  v Hz. Pokud v autokorelační funkci dvě maxima nejsou, je  $F_0$  vyhodnocena jako nulová. Výstupem funkce je vektor hodnot  $F_0$  pro každý segment vstupního signálu.

Z popisu funkce vyplývá, že intonace daného segmentu je určena pouze z velikosti první krátkodobé periody segmentu. Algoritmus tedy není dokonale přesný a údaj nabývá na přesnosti se snižující se délkou segmentu, resp. dobou jeho trvání.

#### 4.1.3.2 AMDF

AMDF je variantou ACF a poskytuje tudíž podobné informace o vlastnostech analyzovaného signálu. Byla naprogramována funkce *Pitch\_AMDF\_final.m*, jejíž vstupy, výstupy a výpočetní úkony jsou stejné jako u předchozí funkce *Pitch\_ACF\_final.m*, pouze je zde místo integrované funkce *autocorr* využita naprogramovaná funkce *AMDF.m*, jejíž výstupem je diferenční funkce vstupního signálu se svým duplikátem a z jeho převrácené hodnoty jsou extrahovány polohy kýžených maxim.

#### 4.1.3.3 NCC

NCC je modifikovaná verze autokorelační funkce, jenž podle [15] daleko lépe sleduje prudké změny vstupního signálu a dává lepší odhad periodické struktury vstupního signálu. Ačkoli tato funkce je v MATLABu ver. 2019 integrována jako funkce *normxcorr2*, byla naprogramována funkce *NCC.m* se stejnou úlohou. Samotná funkce *Pitch\_NCC\_final.m* opět pracuje na stejném principu, jako předchozí dvě funkce pro určování intonace, pouze je místo funkcí *autocorr* a *AMDF.m* použita právě funkce *NCC.m* a polohy maxim jsou extrahovány z výstupní normalizované autokorelační funkce.

### 4.1.4 Určení podobnosti

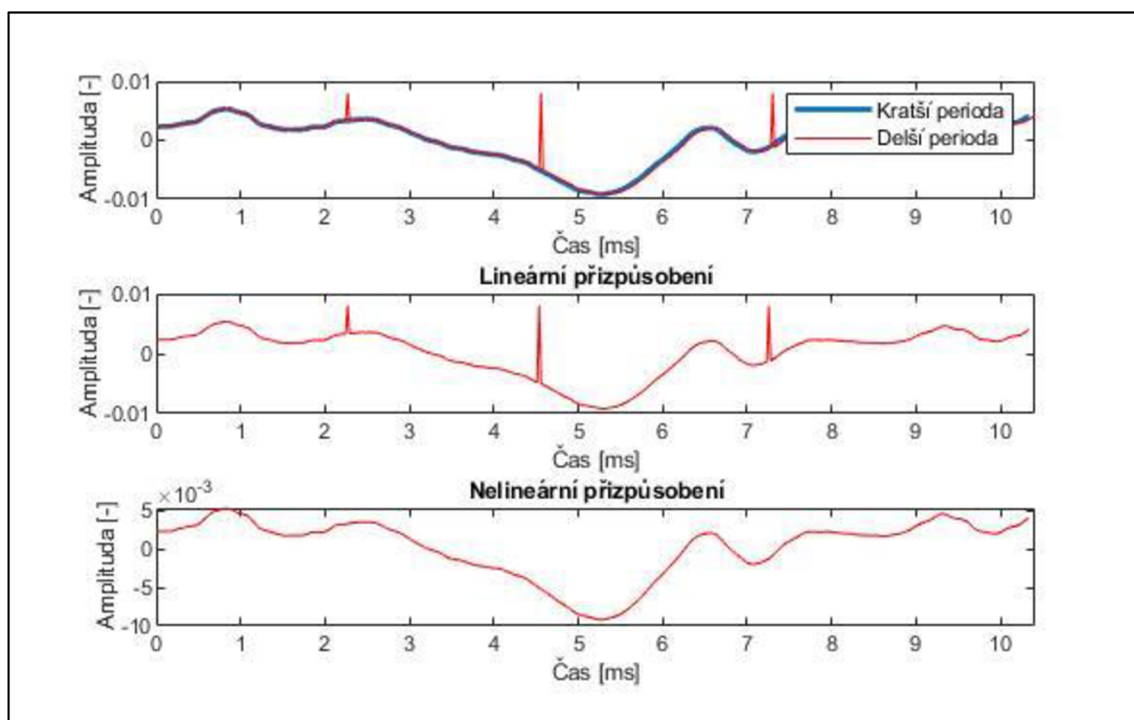
Naprogramované algoritmy určování vzájemné podobnosti sousedních period mají komplikovanější strukturu, než výše uvedené funkce pro určování znělosti/neznělosti a intonace. Pro lepší přehlednost byly naprogramovány ve formě stavového automatu s využitím cyklů switch-case. Protože zkoumáme odchylky v schimmeru, je součástí těchto algoritmů možnost lineárního a nelineárního přizpůsobení časové osy.

Obecná funkce, resp. kostra algoritmů určování podobnosti, je převážně stejná. Vstupem je úsek znělého signálu vykazujícího kvaziperiodicitu, ideálně např. signál reprezentující jednu znělou hlásku, a cílem je postupné porovnávání sousedních krátkodobých period tohoto signálu, tzn. dvojice první-druhá, druhá-třetí atd. Stěžejním úkolem je tedy nalezení začátku a průběhu jedné krátkodobé periody a toho samého pro jí odpovídající následující periodu. Máme-li tyto periody vymezeny, musí být uloženy do vektorů, přizpůsobeny na délku kratší z period a nakonec porovnány jednou z níže uvedených metod. Výsledkem má být vektor procentuálních hodnot jejich podobnosti.

V prvním kroku je ze vstupního signálu extrahováno okno o volitelné délce, které však musí obsahovat alespoň dvě krátkodobé periody. Následně je tento segment zpracován funkcí *AMDF.m* a je získána jeho diferenční funkce, z níž jsou pomocí funkce *findpeaks* zjištěny polohy lokálních maxim. Z maxim lze podobně jako při určování intonace

zjistit počátky a délky sousedních period a na základě těchto informací lze tyto periody přímo extrahovat ze vstupního signálu. Pokud je vzdálenost příslušných maxim příliš velká nebo naopak příliš malá, je úsek vyhodnocen jako aperiodický a algoritmus je ukončen. Proto je nezbytně nutné, aby byl vstupní signál vhodně předzpracován a neobsahoval např. hlasivkové náběhy.

Před porovnávacími algoritmy pro určení podobnosti je nutné přizpůsobit délku. Při nelineárním přizpůsobení je z delší periody vypuštěno rovnoměrně rozložené množství bodů odpovídající přebytku oproti kratší z period. Pro nelineární přizpůsobení byla naprogramována funkce *Nelin\_prizp.m*, která délku přizpůsobuje způsobem popsáným v kapitole 3.8.2, přičemž vzájemná podobnost při vypouštění vzorků je testována metodou AMDF. Funkci těchto procedur ilustruje obr. 4-2. Bylo testováno přizpůsobení délky signálu krátkodobé periody hlásky /a/ rozšířené o tři vzorky navíc (v každé třetině tohoto signálu) na délku původní periody o 457 vzorcích. Při lineárním přizpůsobení byly odebrány 3 rovnoměrně rozložené vzorky. Algoritmus nelineárního přizpůsobení detekoval a odstranil právě tři vložené vzorky, které svou velikostí výrazně narušovaly vzájemnou podobnost period, čímž bylo dosaženo větší (v tomto případě maximální) podobnosti.



**Obr. 4-2 - Porovnání algoritmů pro lin. a nelin. časové přizpůsobení**

Po získání dvou signálů sousedních period je provedena analýza jejich podobnosti.

#### 4.1.4.1 AMDF

První vzorek diferenční funkce dvou signálů, odpovídající nulovému vzájemnému posunu, nám dává informaci o jejich vzájemné podobnosti. Jedná se o jejich absolutní



diferenci, tedy průměrnou odchylku v amplitudách. Za tímto účelem byla naprogramována funkce *twosignalAMDF.m*, jejímiž vstupy jsou přizpůsobené signály sousedních krátkodobých period a výstupem je hodnota absolutní difference. Ta musí být následně normalizována a převedena na procenta.

Finálními algoritmy jsou *AMDF\_periodicita\_lin3* a *AMDF\_periodicita\_nelin*.

#### 4.1.4.2 NCC

Podobným způsobem lze využít také metody NCC, která při nulovém posunu analyzovaných funkcí v prvním vzorku vrátí hodnotu udávající vzájemnou podobnost v intervalu  $(0,1)$ . Při testování se ukázalo, že při určování procentuální podobnosti naprogramovaná funkce *NCC.m* vykazuje chybu o několika procentech. Byla proto využita integrovaná funkce *normxcorr*, která je zcela přesná. Výsledná hodnota je následně převedena na procenta.

Finálními algoritmy jsou *NCC\_periodicita\_lin* a *NCC\_periodicita\_nelin2*.

#### 4.1.4.3 DTW

Podle [16] je DTW jedna ze starších metod přímo určených k určování podobnosti mezi soubory dat. Byla využita integrovaná funkce *dtw*, jejímiž vstupy jsou porovnávané signály a výstupem je hodnota nejkratší Euklidovy vzdálenosti, která je dále normalizována a převedena na procenta.

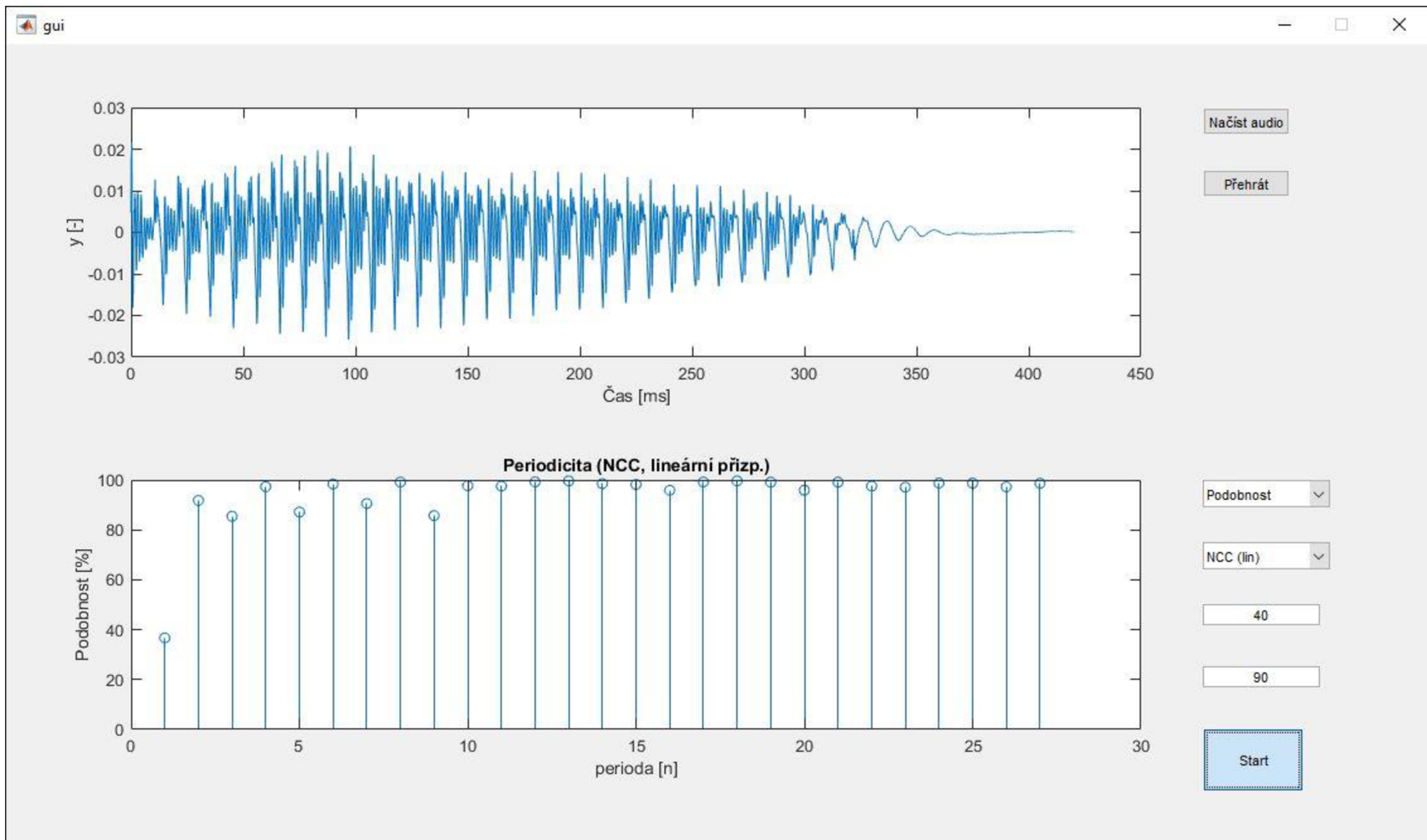
Po výpočtu podobnosti sousedních period se cyklus opakuje, tentokrát se okno posune na začátek druhé z krátkodobých period porovnávaných v uplynulém cyklu.

Finálními algoritmy jsou *DTW\_periodicita\_lin* a *DTW\_periodicita\_nelin*.

### 4.1.5 GUI

Součástí programu je jednoduché grafické uživatelské rozhraní vytvořené v modulu *Guide* a soubor *gui.m* sdružující vytvořené výpočetní a obslužné funkce. Hlavním prvkem GUI jsou dvě pole pro zobrazení vstupního signálu a závislosti zkoumaného parametru. Vstupní signál je načten pomocí tlačítka *Načíst audio*; součástí je možnost zvukové reprodukce tlačítkem *Přehrát*, využitelné pro déle trvající signály. Rozbalovací okna vpravo dole nabízí výběr parametru a metody. Specifikace délky segmentu pro krátkodobou analýzu a prahu pro výběr spojených úseků k uložení na disk je možné provést v textových polích *Délka okna* a *Práh*. Samotná analýza je zahájena tlačítkem *Start*. Po zobrazení výstupních dat lze ihned změnit parametry a provádět další analýzy.

Modul shlukování úseků signálu podle kritéria periodicity představují funkce *LOKACE\_pitch*, *LOKACE\_podobnost* a *LOKACE\_znelost*. Nejsou tedy přímou součástí jednotlivých algoritmů a jsou implementovány pouze v rámci GUI. Úseky jsou ukládány do příslušných složek ve formátu *.mat*.



Obr. 4-3 – Grafické uživatelské rozhraní (analýza podobnosti /a/, alg. NCC(lin), okno o 40 ms, práh 90% podobnost)

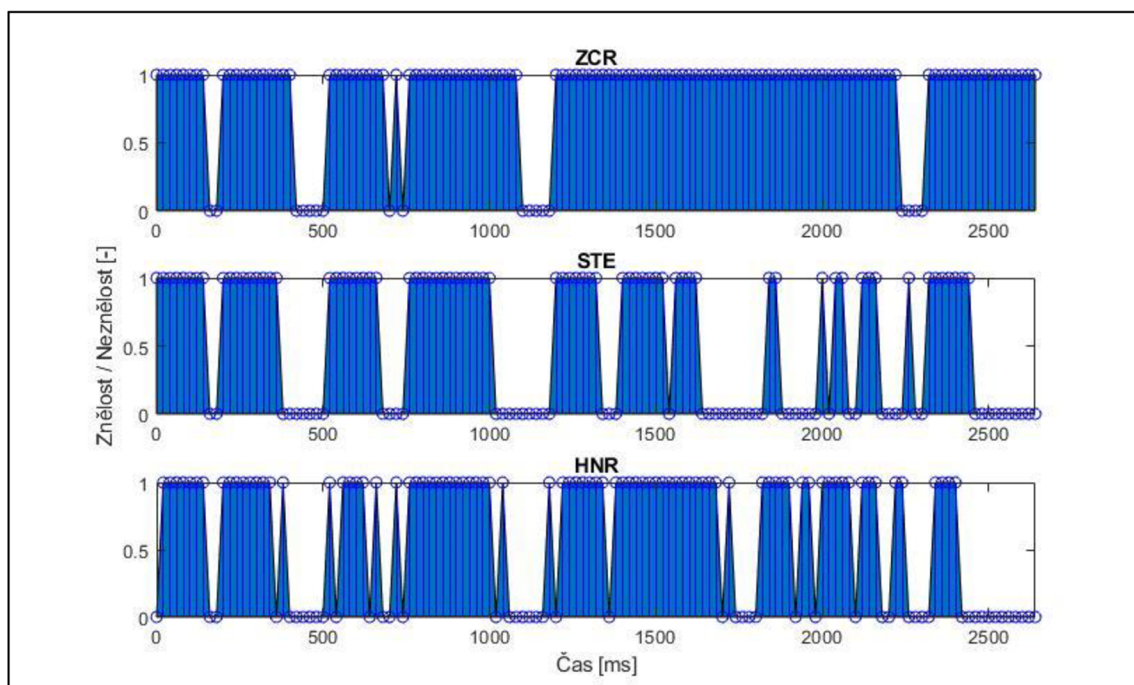
## 4.2 Testování

Testování algoritmů bylo pro lepší zobrazitelnost prováděno mimo vytvořené GUI v souboru *testy.m*.

### 4.2.1 Znělost/Neznělost

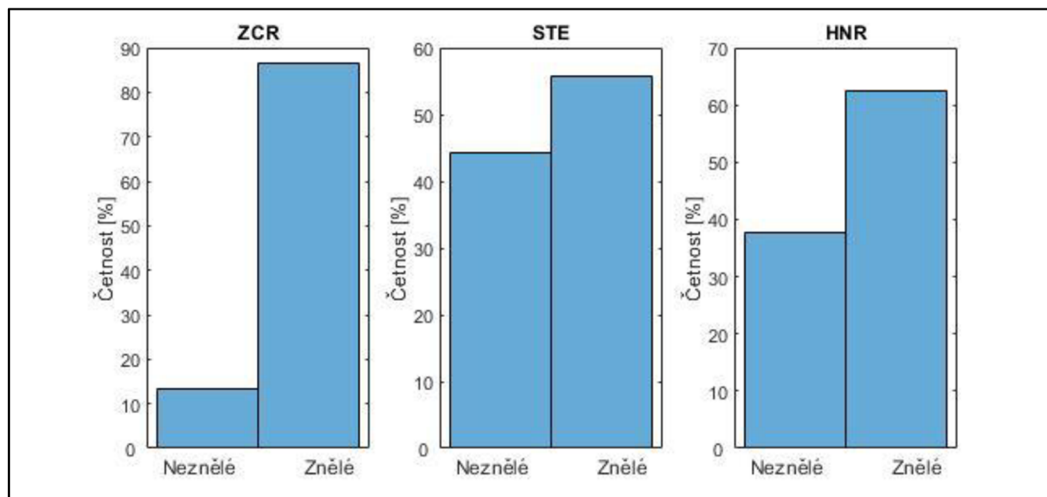
Algoritmy budou testovány na nahrávce věty „*Vážení přátelé, přeji vám dobrý večer.*“, namluvenou hercem Zdeňkem Svěrákem. Nahrávka má délku 2646 ms a při vzorkovací frekvenci 44100 sa/s má 116692 vzorků. Délka segmentu byla zvolena o optimální hodnotě 20 ms.

Obr. 4-4 zobrazuje výsledky určování znělosti/neznělosti algoritmy ZCR, STE a HNR. Hodnota 1 představuje znělý segment, 0 neznělý. Podle obr. 4-7, zobrazujícího manuálně rozčleněný vstupní signál a průběhy tohoto signálu po vynulování neznělých částí, můžeme blíže prozkoumat účinnost algoritmů. Všechny algoritmy takřka spolehlivě odstranily výrazné neznělé souhlásky /ž/ a /ř/, pouze HNR v tomto vykazuje menší chybovost. Souhláska /č/ ve slově *večer* nebyla zcela správně detekována algoritmem STE kvůli závislosti na velikosti amplitudy, která mohla být při sejmutí mikrofonom zkreslena. Algoritmus STE však společně s algoritmem HNR filtrovaly častěji než ZCR. Lze vidět, že narozdíl od ZCR detekovaly exploziva /p/, /b/, /d/ a /t/ i hlásky /v/ a /r/ díky jejich nízké energii v poměru k energii znělých hlásek.



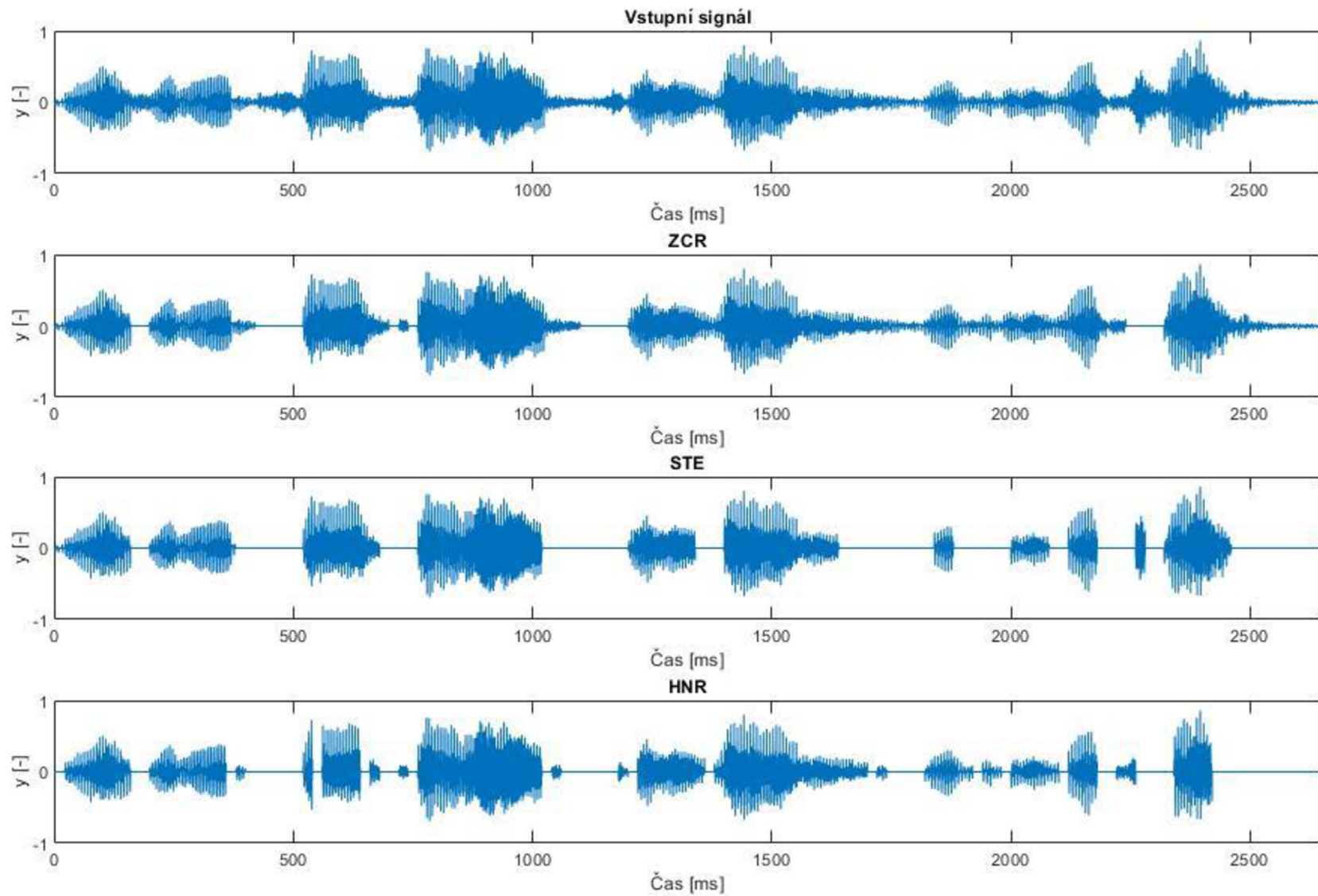
Obr. 4-4 - Grafické znázornění analýzy znělosti/neznělosti algoritmy ZCR, STE a HNR

Obr. 4-5 dává informaci o četnosti znělých a neznělých segmentů ve zpracovaném signálu formou histogramů. Z předchozích závěrů může být algoritmus ZCR považován za nejméně přesný - na histogramu se to projevuje velmi malým poměrem neznělosti/znělosti, který zbylé dva algoritmy vyhodnotily jako podstatně větší – četnost neznělých hlásek udávají kolem 40%.

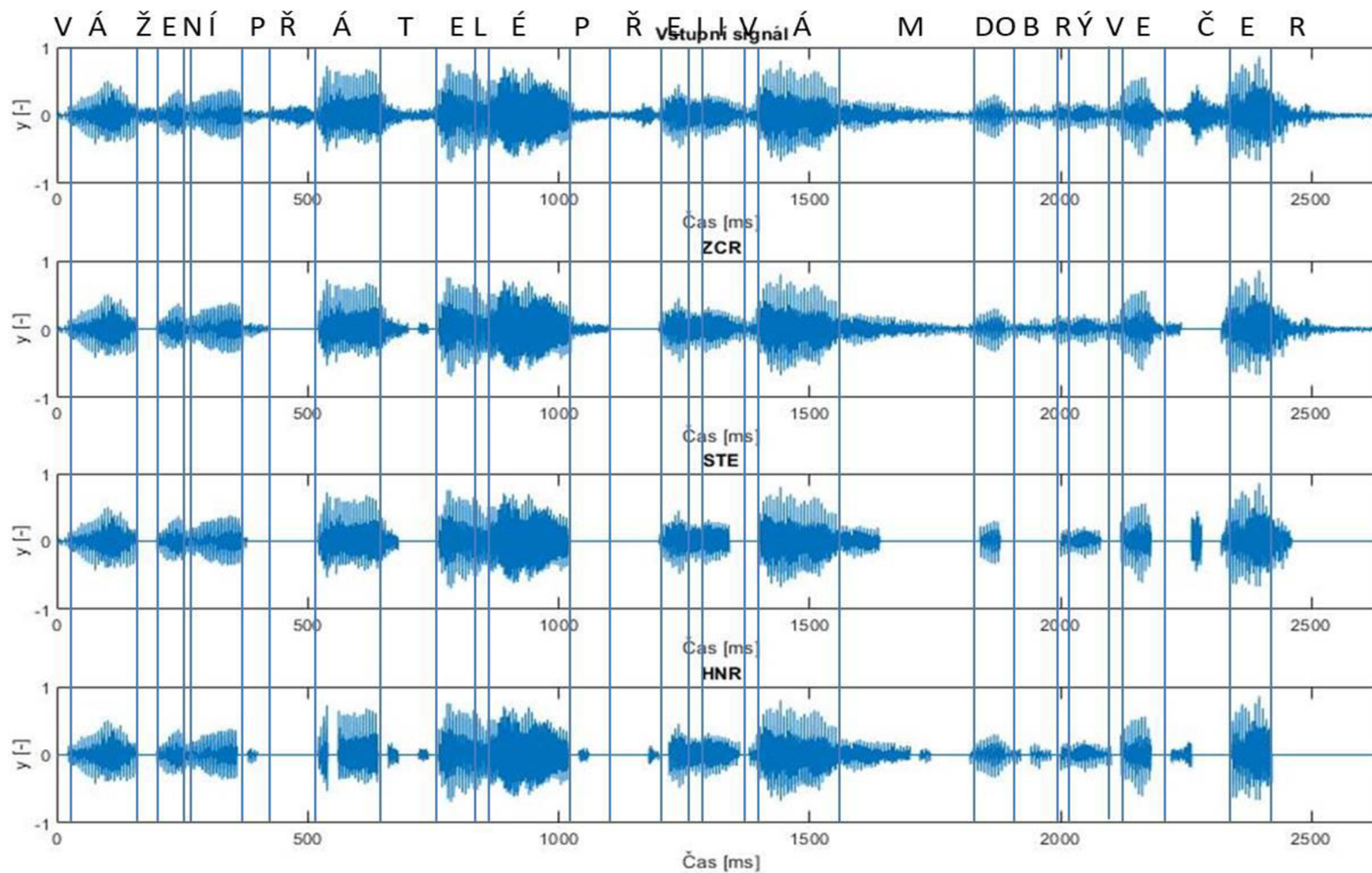


**Obr. 4-5 - Histogramy vyhodnocených segmentů**

Obr. 4-6 zobrazuje průběh vstupního signálu a jeho verze s vynulovanými částmi představujícími segmenty, jež byly algoritmy vyhodnoceny jako neznělé.



Obr. 4-6 - Vstupní signál a jeho verze po zpracování algoritmy ZCR, STE a HNR

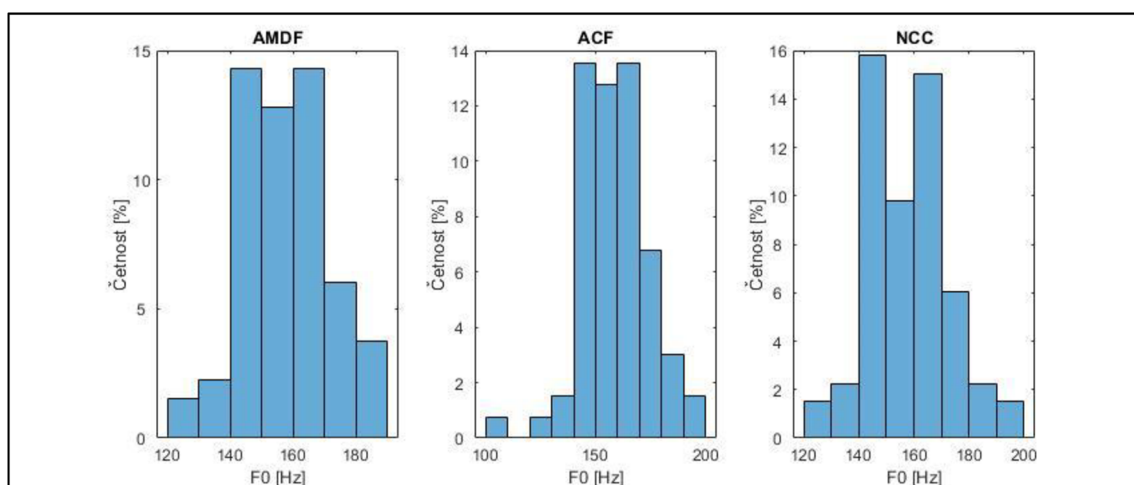


Obr. 4-7 - Vstupní signál a jeho verze po zpracování algoritmy ZCR, STE a HNR + manuální rozčlenění věty

## 4.2.2 Intonace

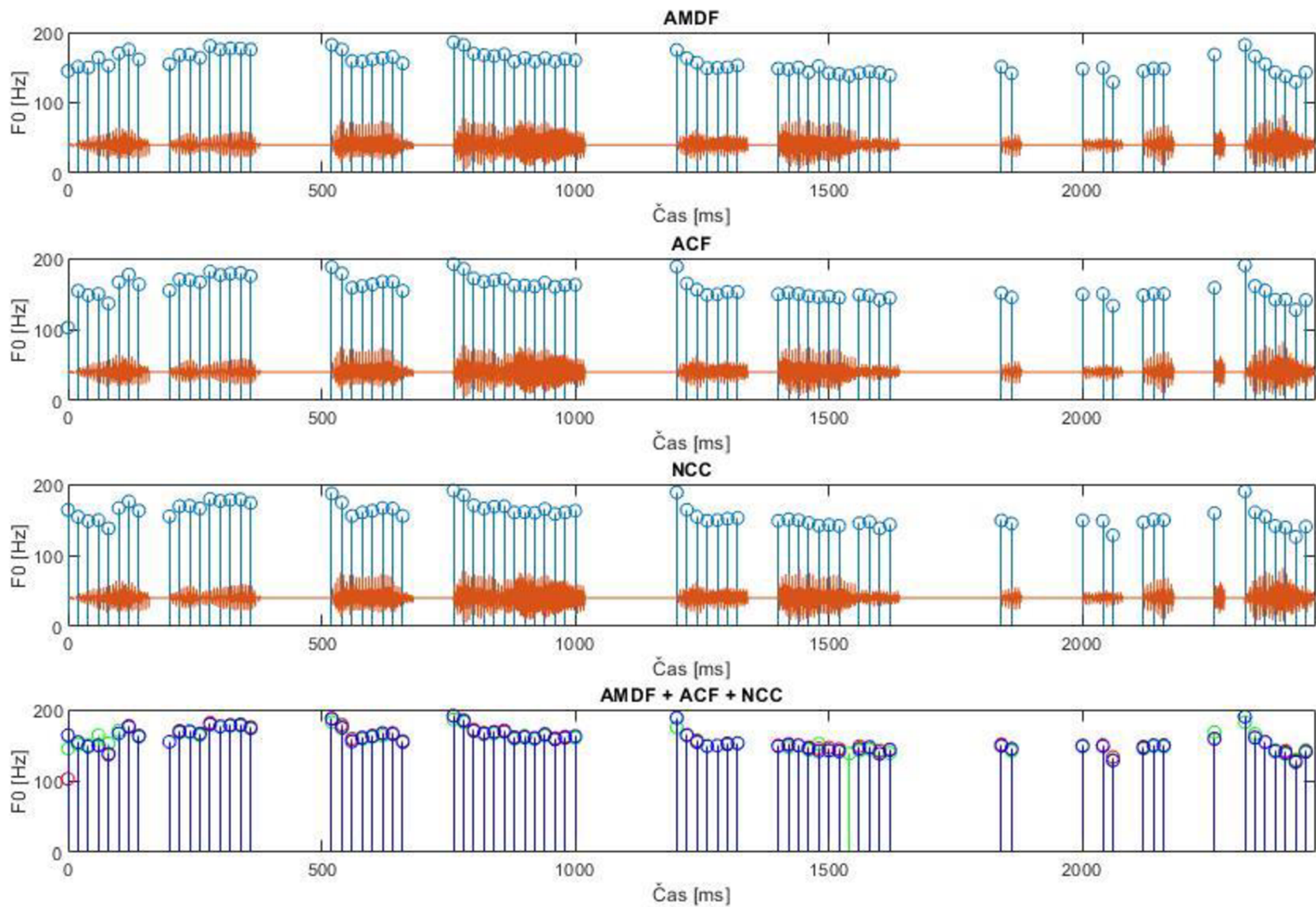
Intonace byla opět testována na předešlé větě. Zvolená délka okna měla pro oba algoritmy hodnotu 20 ms, jako předzpracující algoritmus pro určení znělosti/neznělosti je defaultně nastaven algoritmus STE.

Obr. 4-8 udává statistickou informaci o rozložení segmentů na frekvenční ose podle základní frekvence. Hlas se pohybuje v mezích 60 – 400 Hz, typických pro běžnou řeč; v analyzované větě základní frekvence nestoupá nad 200 Hz. V histogramech můžeme vidět poněkud rozdílnou četnost výskytu detekovaných frekvencí. Tyto rozdíly jsou způsobeny odchylkami jednotlivých metod při určování pozic začátků period.



Obr. 4-8 - Histogramy základní frekvence segmentů

Na obr. 4-9 můžeme sledovat průběhy F0, tedy intonaci nebo melodii řeči. Grafy vykazují velmi podobný průběh, jak ilustruje poslední z grafů, vyhodnocení jitteru tedy u všech algoritmů můžeme považovat za poměrně přesné.

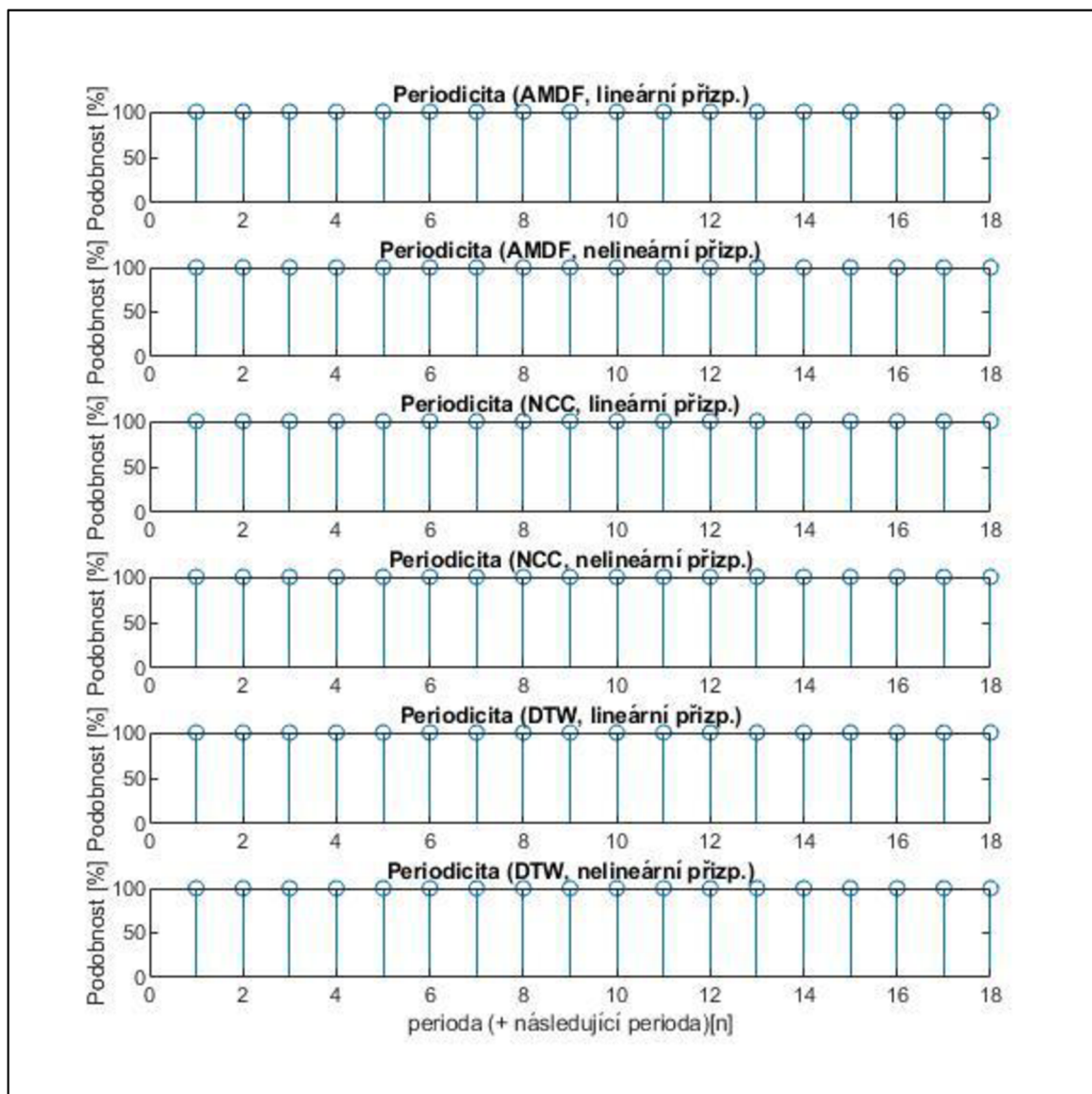


Obr. 4-9 - Grafické zobrazení průběhů F0, resp. intonace



### 4.2.3 Podobnost

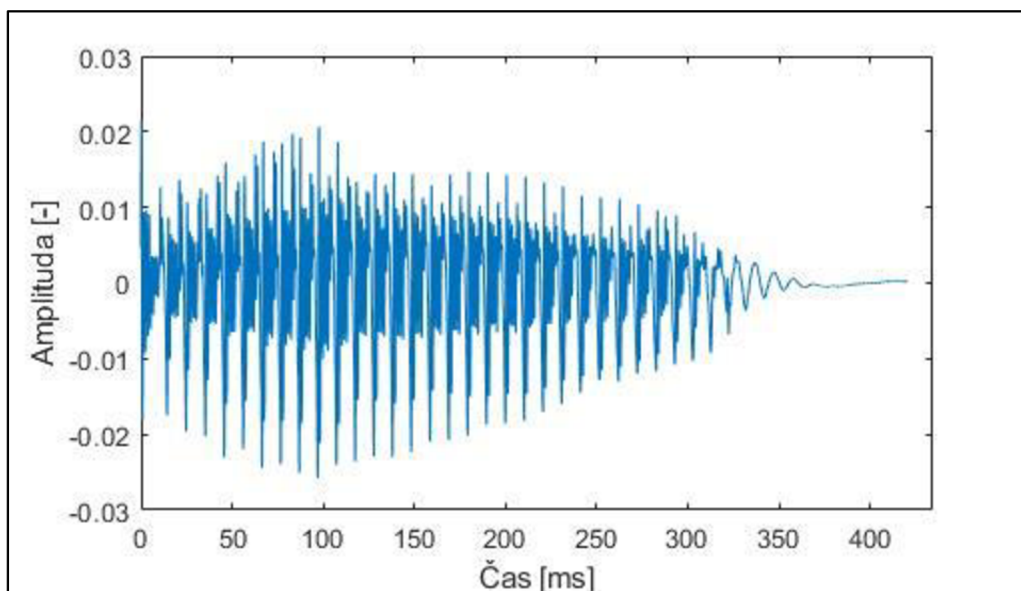
Algoritmy byly nejprve testovány na dokonale periodickém signálu složeném z dvaceti totožných period, extrahovaných z nahrávky hlásky /a/. Při délce okna 30 ms byl pro všechny algoritmy získán soubor 18 porovnání o hodnotách 100%. Program nepokračuje pro poslední dvě periody – v situaci, kdy momentálně analyzované okno přesahuje celý signál, je algoritmus ukončen. Výsledek je zobrazen na obr. 4-10.



Obr. 4-10 - Podobnost sousedních period dokonale periodického signálu pro algoritmy AMDF, NCC a DTW

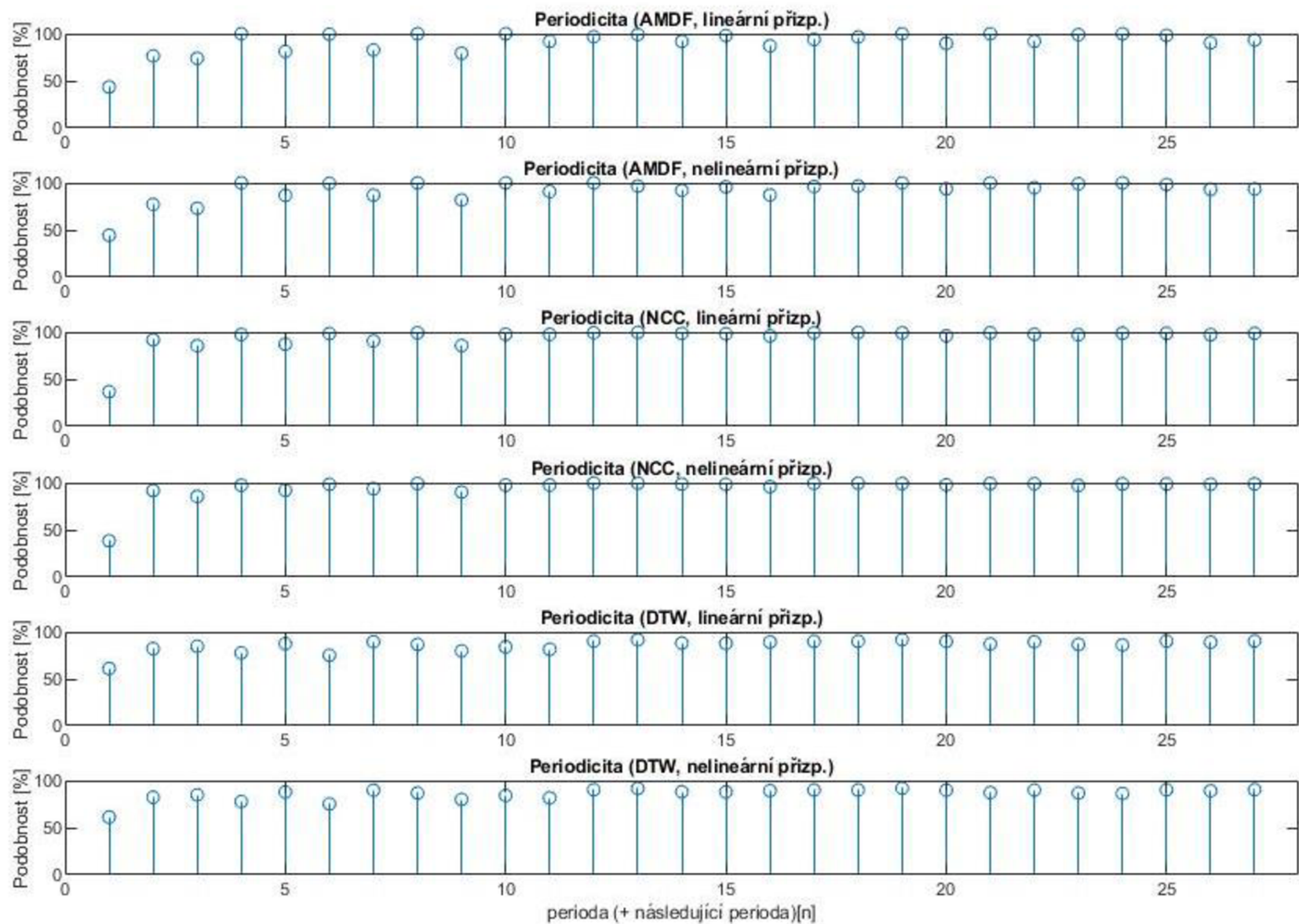
Tímto testem bylo ověřeno, že program vyhodnotí schimmer sousedních period dokonale periodického signálu jako nulový a vzájemnou podobnost jako stoprocentní, tzn. periodičita signálu je maximální.

Následující test byl proveden na reálném řečovém signálu, a to na nahrávce české hlásky /a/, viz obr. 2-1. Nahrávka má délku 551,4739 ms a při vzorkovací frekvenci 44100 sa/s má 24320 vzorků. Jak bylo zmíněno v kapitole 4.1.4, podobnost sousedních period lze určovat pouze u periodického, resp. kvaziperiodického signálu. Zmíněná nahrávka však obsahuje hlasivkový náběh, u něhož vlivem postupného buzení hlasivek nejsou první periody zcela zřetelné a při zpracování algoritmy se objevují chyby vlivem nenalezení periodicity v segmentu, což ústí v ukončení algoritmu. Signál je proto nutné předzpracovat, což v tomto případě zahrnuje manuální ořez počáteční části signálu. Experimentálně byla nalezena nová ideální počáteční hodnota ve vzorku č. 5786. Ořezaný signál je zobrazen na obr. 4-11.



**Obr. 4-11 - Předzpracovaný signál hlásky /a/**

V tomto stavu je již možné signál podrobit analýze algoritmy. Délka segmentu byla zvolena v 40 ms. Výsledky jsou zobrazeny na obr. 4-12. Výsledné grafy vykazují velmi podobný průběh, u DTW je patrné mírný pokles hodnot. Několik prvních porovnání sousedních period (cca do porovnání 3. a 4. periody signálu) vždy vykazuje nižší hodnoty podobnosti. To koresponduje s průběhem analyzovaného signálu, jehož amplituda nejprve s rozkmitáváním hlasivek nabývá na velikosti a jakosti, než se kolem 100. milisekundy ustálí. Zde graf periodicity s kolísáním ustaluje svou hodnotu mezi 80-100%. Následné odeznívání signálu /a/ není zahrnuto v měření, protože algoritmy porovnávání ukončily po přesáhnutí signálu oknem.



Obr. 4-12 - Podobnost sousedních period hlásky /a/ pro algoritmy AMDF, NCC a DTW

Pro ověření výhod lineárního přizpůsobení jsou v tab. 2 vypsány výsledné hodnoty podobnosti mezi jednotlivými porovnávanými periodami, vypočtené algoritmem na bázi AMDF. Jak bylo očekáváno, algoritmus s nelineárním přizpůsobením časové osy dosáhl průměrně vyšších hodnot podobnosti mezi periodami díky odstranění vhodných bodů.

**Tab. 2 – Porovnání podobnosti sousedních period pro algoritmus AMDF**

Porovnávané periody	Podobnost [%]	
	Lineární př.	Nelineární př.
1; 2	43,57	44,42
2; 3	76,50	77,18
3; 4	73,86	73,09
4; 5	100,00	100,00
5; 6	81,09	86,84
6; 7	99,47	99,47
7; 8	82,77	87,05
8; 9	100,00	100,00
9; 10	79,30	81,83
10; 11	100,00	100,00
11; 12	92,08	90,88
12; 13	96,84	100,00
13; 14	99,17	96,66
14; 15	91,86	92,05
15; 16	98,11	95,86
16; 17	87,14	87,14
17; 18	93,90	96,16
18; 19	96,63	96,63
19; 20	100,00	100,00
20; 21	89,76	93,73
21; 22	100,00	100,00
22; 23	92,12	94,84
23; 24	99,07	99,07
24; 25	100,00	100,00
25; 26	98,32	98,32
26; 27	90,51	93,10
27; 28	93,24	93,53
Průměrná hodnota:	90,94	91,77
Směrodatná odchylka:	12,33	11,93

### 4.3 Srovnání s programem Praat

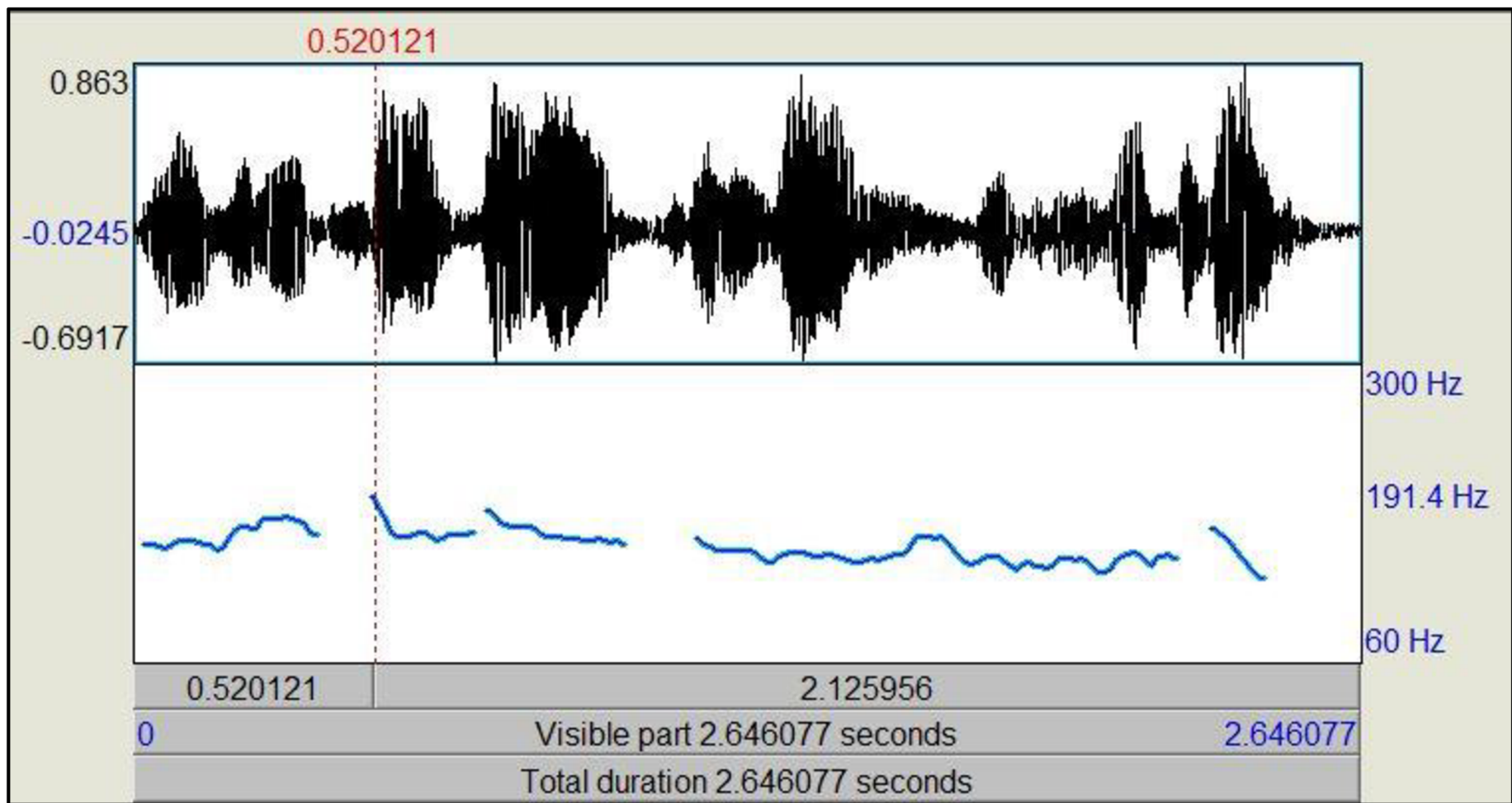
Pro srovnání byly provedeny testy určování znělosti/neznělosti a intonace ve volně dostupném softwarovém nástroji Praat, který vyvinul prof. Boersma z Amsterdamské univerzity a který je určen k analýze řečových signálů. Předmětem analýzy je opět věta popsaná v kap. 4.2.1.

Určování znělosti/neznělosti a intonace patří mezi základní funkce programu Praat. K těmto využívá autokorelační funkci, jenž narozdíl od jiných programů, např. MDVP, zpracovává originální signál bez normalizace jeho amplitudy. Znělost je určena z hodnoty extrému autokorelační funkce, která je srovnávána s hodnotou *voicing threshold* o defaultní hodnotě 0,45. Pokud je práh nižší než hodnota extrému, je segment vyhodnocen jako znělý. Základní frekvence je určena ze vzdáleností extrémů v autokorelační funkci.

Jak lze vidět na obr. 4-13, určení znělosti/neznělosti proběhlo s mírnými odchylkami proti dosaženým výsledkům s metodami ZCR, STE a HNR. K přiblížení se výsledkům bylo nutné ladění hodnoty *voicing threshold*, která byla experimentálně stanovena na optimální hodnotu 0,6. Druhá z řídicích veličin, *voiced/unvoiced cost*, spojená s délkou analyzovaného segmentu a citlivějším měřením, nebyla měněna.

Praat oproti algoritmům zcela přesně nevyhodnotil hlásku /p/, dále nepřesně hlásky /ž/, /v/ a celé slovo „dobrý“, které kompletně určil jako znělé. Správně jako neznělé určil /ž/, /č/, /t/ a /r/. Změnou hodnoty *voicing threshold* lze přesnost vyhodnocování zvýšit, ovšem za cenu špatného vyhodnocení poměrně dlouhého úseku znělé hlásky /e/ (druhé ve slově večer). Je tedy možné předpokládat, že při hlubším výzkumu metod ZCR, STE a HNR a při určení jejich citlivostí na konkrétní hlásky by bylo možné jejich kombinací do jediného algoritmu vytvořit poměrně spolehlivý nástroj k vyhodnocování znělosti/neznělosti.

Křivka F0, získaná analýzou Praatem, se svým průběhem podobá výsledným křivkám získaným naprogramovanými algoritmy. Melodie tedy byla určena velmi podobně, určené hodnoty F0 se však mírně o liší o jednotky Hz. Např. v čase 520 ms je základní frekvence Praatem určena o hodnotě 191,4 Hz, zatímco naprogramovaný algoritmus pracující s metodou ACF v tomto čase F0 vyhodnotil jako 187,7 Hz. Je nutné poznamenat, že Praat se velmi liší v hustotě vyhodnocených frekvencí; zatímco algoritmy vyhodnocují frekvenci v segmentu o nastavené délce 20 ms z velikosti první krátkodobé periody, Praat zřejmě vyhodnocuje daleko kratší úseky.



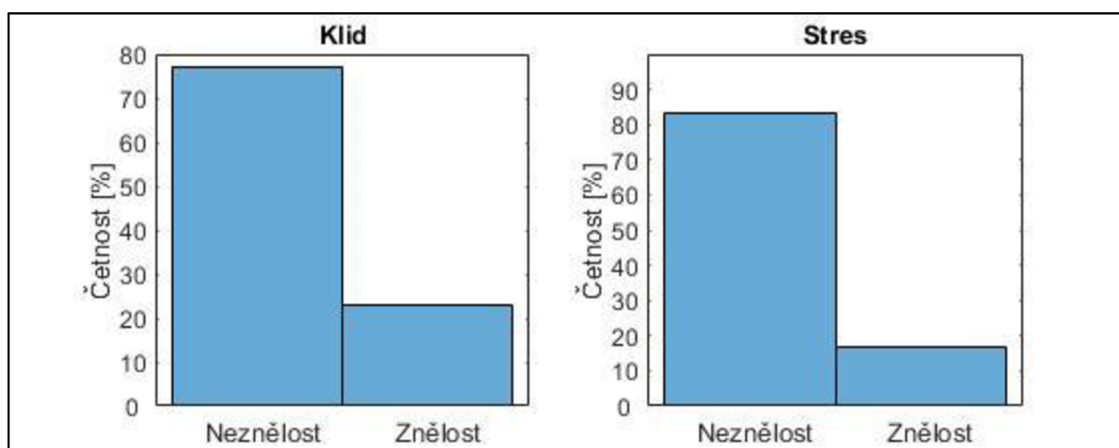
Obr. 4-13 - Analýza programem Praat

## 4.4 Praktické využití

Praktické využití práce bylo demonstrováno v příspěvku soutěže Student EEICT 2020.

Význam analýzy periodicity řečového signálu spočívá ve velmi širokém využití při rozpoznávání řeči. Příkladem může být zkoumání ovlivnění řečového signálu stresem. Řeč je vytvářena kontrolou napětí svalů řečového ústrojí, podle [17] stres do tohoto procesu vstupuje tím, že způsobuje zvýšené svalové napětí a zrychlení respirace, čímž znemožňuje plnou kontrolu nad tvorbou hlasu. To se projeví změnou základní frekvence, resp. intonace řeči, hlasovým třesem, rychlostí promluvy či počtem a délkou pauz v řeči.

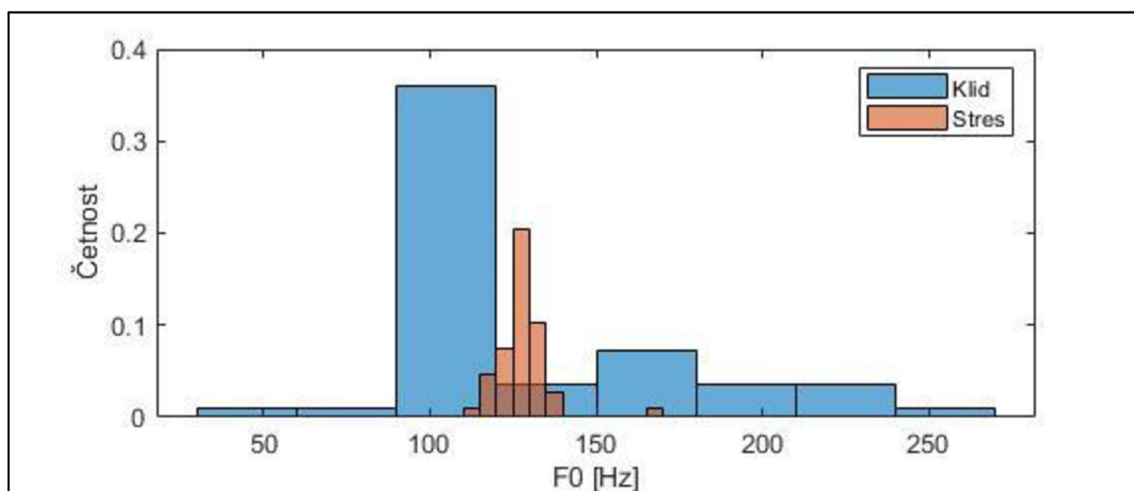
Tyto vlivy byly analyzovány v rámci dvou řečových signálů (22 kHz, 16 bitů, mono, formát wav), resp. nahrávek spontánně mluveného odborného textu v českém jazyce, a to stejnou osobou mužského pohlaví. Text na obou nahrávkách je totožný. V první nahrávce je mluvčí pod vlivem silného stresu vyvolaného důležitou zkouškou na univerzitě, druhá nahrávka je oproti tomu namluvena klidnou řečí. Vzhledem k zadržování mluvčího ve stresu a množství přechů stesový signál trvá 6:25 a je téměř dvakrát delší, než záznam klidné mluvy o trvání 3:27. Délka signálu klidné řeči je postačující pro spolehlivost testů ze statistického hlediska [18]. Prvním analyzovaným parametrem byla znělost. Předpokládaným výsledkem byla větší koncentrace neznělých úseků v řeči pod vlivem stresu. Signály byly rozděleny na úseky o délce 20 ms, ty byly rozděleny na znělé/neznělé algoritmem STE a z výsledných dat byly vytvořeny histogramy, viz obr. 4-13.



Obr. 4-14 - Histogramy vyhodnocených segmentů v klidu a ve stresu

Ačkoli má délka nahrávky stresované řeči téměř dvakrát delší dobu trvání než klidná řeč, je velmi zajímavým výsledkem, že i když poměr znělých úseků stres/klid je přibližně 1,38, tedy že stresovaná nahrávka je ve svém průběhu mírně častěji znělá, počet neznělých úseků je v ní přesně dvakrát větší. Můžeme tedy potvrdit, že v mluvě pod vlivem stresu se skutečně daleko častěji vyskytují pauzy a zasekávání.

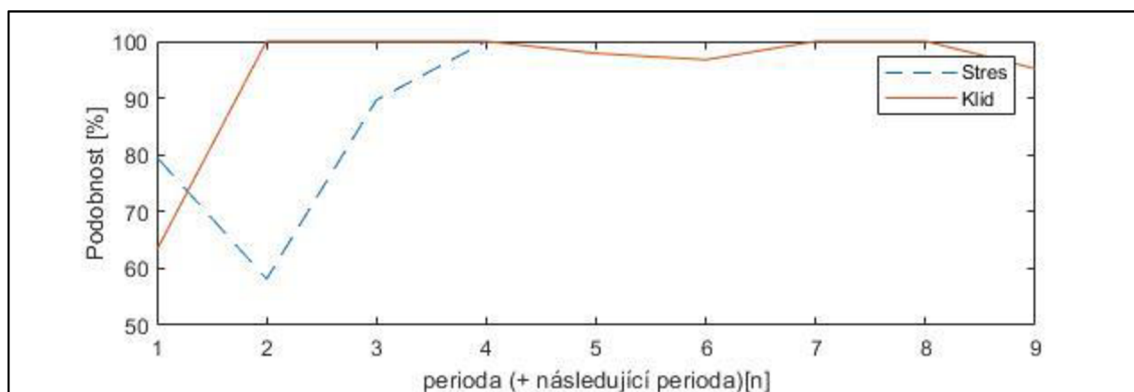
Následujícím analyzovaným parametrem byla intonace řeči. Je běžné, že vlivem stresu a s tím souvisejícího zvýšeného napětí hlasivek stoupne základní frekvence hlasu. Intonace byla určena algoritmem na bázi AMDF a z dat byly opět vytvořeny histogramy, viz obr. 4-14. Podle očekávání je z nich zřejmé, že zatímco intonace klidné řeči se koncentruje na nízkých frekvencích okolo 100 Hz a její vyšší frekvence dotvářejí pouze přízvuk, veškerá intonace řeči pod vlivem stresu se soustřeďuje výše okolo 127 Hz v rozpětí cca 70 Hz a zcela postrádá známky melodie. Řeč tedy pod vlivem stresu tónově stoupá a je monotónní. Výrazně větší soubor dat pro klidovou řeč i přes kratší dobu nahrávky lze vysvětlit větším poměrem znělých úseků k neznělým.



**Obr. 4-15 - Histogram F0 pro řeč v klidu a ve stresu**



Posledním analyzovaným jevem je schimmer popisovaný podobností sousedních period. U řeči pod vlivem stresu se předpokládá chvějící se hlas. Za tímto účelem byly z totožného místa na obou nahrávkách extrahovány hlásky /a/ o délce 10 krátkodobých period, jež byly následně zpracovány algoritmem na bázi AMDF s lineárním přizpůsobením délky. Obr. 4-15 ilustruje předpokládanou skutečnost, že periodičita znělé hlásky pronesené v klidu je vyšší, než periodičita hlásky vyřčené ve stresu.



**Obr. 4-16 - Průběh podobnosti sousedních period hlásky /a/ pro řeč v klidu a ve stresu**

## 5. ZÁVĚR

Práce shrnuje některé z metod v oblasti zpracování řečového signálu, resp. v oblasti vyhodnocování tří jeho stěžejních parametrů. Samotnému popisu těchto metod předchází teoretická část, jež rozebírá analyzované parametry ve vztahu k řečovému signálu. Toto obnáší jednak seznámení s podstatou těchto signálů a metodou krátkodobé analýzy, dále pak úvod do problematiky segmentace signálu a následný rozbor parametrů znělost/neznělost, intonace a periodicitu. Následující úsek práce se věnuje výpočetním metodám k analýze uvedených parametrů.

Praktická část práce sestává z algoritmů naprogramovaných v prostředí MATLAB. Čítá řadu funkcí, z nichž vždy tři jsou použitelné k analýze jednoho ze zkoumaných parametrů, přičemž některé z algoritmů používají výsledky jiných. Protože je vycházeno z metody krátkodobé analýzy, součástí práce jsou rovněž pomocné algoritmy na segmentaci nebo předzpracování signálu.

Poslední část práce popisuje dosažené výsledky při testování algoritmů na reálných řečových signálech, konkrétně na signálech znělé hlásky /a/ a české věty. Také je uveden příklad praktického použití při analýze klidné řeči a řeči ve stresu. Tyto výsledky byly prezentovány v rámci soutěžního příspěvku v soutěži EEICT, přičemž příspěvek se díky nim ve své kategorii umístil mezi výherními příspěvky na prvním místě.

Vzhledem k funkčnosti algoritmů a k dosažení očekávaných výsledků při testech považují vypracovaná řešení za uspokojivá, nicméně algoritmy kvůli své náchylnosti na některé nežádoucí jevy v signálech (např. neperiodické úseky glotálních pulsů na začátku signálu znělých hlásek) zdaleka nejsou dokonalé a související problémy musí být řešeny např. ručním ořezáním nevyhovujících částí vstupního signálu.

Přidanou hodnotou programu je unikátní algoritmus k analýze znělé hlásky a vykreslení grafu vzájemné podobnosti sousedních period. Svou použitelností nabízí široké možnosti k analýze řečových signálů a vlivů na něj působících; kromě toho je potenciálně možné ho použít i k analýze signálů odlišných od řečových, např. pro ověřování kvality hudebních nástrojů. I v tomto ohledu bude jistě mít smysl ve vývoji nadále pokračovat a rozšířit spektrum použitelnosti na další praktické aplikace, z nichž se nabízí např. výzkum vlivů nestacionárního šumu [19] a hlasového úsilí [20].

# Literatura

- [1] P. Mareš, Úvod do lingvistiky a lingvistické bohemistiky (vydání první). Karolinum, Praha, 2014.
- [2] *Mluvní orgány* [online]. Dostupné z: <https://www.czechency.org/>
- [3] J. Psutka, L. Müller, J. Matoušek, V. Radová, *Mluvíme s počítačem česky*. Academia, Prague, 2006.
- [4] Elizabeth C. Zsiga, *The Sounds of Language: An Introduction to Phonetics and Phonology*. Wiley-Blackwell, 2013.
- [5] J. P. Teixeira, C. Oliveira, and C. Lopes, Vocal acoustic analysis-jitter, shimmer and HNR parameters, *Procedia Technology*, 2013, vol. 9, pp. 1112-1122.
- [6] S. Pospíšil, *Shluková analýza signálu EKG*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Ústav biomedicínského inženýrství, 2013.
- [7] R. G. Bachu, S. Kopparthi, B. Adapa, and B. D. Barkana, Voiced/unvoiced decision for speech signals based on zerocrossing rate and energy. In *Proc. Advanced Techniques in Computing Sciences and Software Engineering*. Springer, Dordrecht, 2010, pp. 279-282.
- [8] M. Jalil, F. A. Butt and A. Malik, Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals, In *International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE)*, Konya, 2013, pp. 208-212.
- [9] J. Heranová, *Harmonicita jako možný indikátor hranic mezi segmenty v češtině*. Praha: Univerzita Karlova, Filozofická fakulta, Fonetický ústav, 2010.
- [10] *Distance and correlation* [online]. Dostupné z: <http://www.analytictech.com/>
- [11] L. R. Rabiner and R. W. Schafer, *Theory and Applications of Digital Speech Processing*. Prentice Hall, London, 2011.
- [12] *Time Domain: AMDF* [online]. Dostupné z: <http://mirlab.org/>
- [13] M. Ross, H. Shaffer, A. Cohen, R. Freudberg and H. Manley, Average magnitude difference function pitch extractor, in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1974, pp. 353-362.
- [14] *Performance evaluation of pitch detection algorithms* [online]. Dostupné z: <http://access.feld.cvut.cz/>
- [15] P. Motlíček, Estimation of fundamental frequency in speech. In *Proc. 1st Conference of Czech student AES*. 2000, FEEC BUT Brno, pp. 1-6.
- [16] L. Černohous, *Metoda dynamického borcení času*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2015.
- [17] E. S. Jackson, M. Tiede, D. Beal, and D. H. Whalen, The impact of social-cognitive stress on speech variability, determinism, and stability in adults who do

- and do not stutter. *Journal of Speech, Language, and Hearing Research*, 2016, vol. 59, no. 6, pp. 1295-1314.
- [18] M. Sigmund, Statistical analysis of fundamental frequency based features in speech under stress. *Information Technology and Control*, 2013, vol. 42, no. 3, pp. 286-291.
- [19] P. Zelinka and M. Sigmund, Hierarchical classification tree modeling of nonstationary noise for robust speech recognition. *Information Technology and Control*, 2010, vol. 39, no. 3, pp. 202-210.
- [20] P. Zelinka and M. Sigmund, Automatic vocal effort detection for reliable speech recognition. In *Proc. International Workshop on Machine Learning for Signal Processing*, 2010, Kittila, pp. 349-354.

# Přílohy

Součástí této práce je přiložené CD, jehož obsahem je:

- elektronická verze bakalářské práce
- soubor Readme s krátkým návodem k orientaci v algoritmech
- složka MATLAB obsahující všechny napsané algoritmy + tři adresáře pro ukládání shlukovaných segmentů řečového signálu, a to *Intonace\_useky*, *Podobnost\_useky* a *Znelost\_useky* + audio soubory *hlaska\_a.wav* a *veta.wav*