UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

BAKALÁŘSKÁ PRÁCE

Aplikace Fourierovy analýzy na rozpoznávání kvality samohlásek podle jejich formantů

Vedoucí diplomové práce:
RNDr. Tomáš Fürst, Ph.D.
Rok odevzdání: 2013

Vypracoval:
Jan Walach
ME, III. ročník

Prohlášení

Prohlašuji, že jsem vytvořil tuto bakalářskou práci samostatně za vedení RNDr. Tomáše Fürsta, Ph.D. a že jsem v seznamu použité literatury uvedl všechny zdroje použité při zpracování práce.

V Olomouci dne 13. dubna 2013

Poděkování

Rád bych na tomto místě poděkoval vedoucímu bakalářské práce RNDr. Tomáši Fürstovi, Ph.D. za obětavou spolupráci, za velké množství praktických připomínek, rad a podnětů, které mi poskytl během vypracovávání této práce. Dále děkuji všem těm, kteří mi ochotně poskytli nahrávky svých samohlásek. Také děkuji rodině a přátelům, kteří mě během mého studia a této práce podporovali.

Obsah

1 Úvod

Lidská řeč je jedním ze základních dorozumívacích prostředků. Z pohledu neznalého se může zdát, že řeč a matematika spolu moc nesouvisí, ale opak je pravdou. Tato práce se zabývá analyzováním a klasifikací řeči - konkrétně samohlásek.

Dle teoretických základů je každá samohláska určena určitými charakteristikami ve spektrální oblasti - formanty. V rámci této práce si kladu za cíl pořídit nahrávky samohlásek a ukázat jak formanty jednotlivé samohlásky charakterizují. Dalším cílem této práce je vytvořit program v prostředí MATLAB, který by byl schopen zvuk nahraný do mikrofonu analyzovat a určit o jakou samohlásku se jedná, a jak je hezká, respektive kvalitní. Takovýto program by mohl být užitečný pro lidi, kteří z nějakého důvodu, například mrtvice, ztratili schopnost mluvit a snaží se naučit mluvit znovu.

První část práce je věnována zvuku, konkrétně jeho šířením ve vzduchu, vnímáním zvuku člověkem a také tvorbou zvuku, zejména v lidském těle. Popisuji také vlastnosti zvuku a českou mluvenou řeč.

V dalších dvou kapitolách se zabývám spektrální analýzou zvuku dvěma metodami, konkrétně Fourierovou transformací a metodou Linear predictive coding. Dále nastíním procesy, které pomáhají při analýze řeči. V praktické části pak rozebírám konkrétní postupy, které jsem použil pro analýzu samohlásek a pro určení její kvality.

2 Zvuk

V této kapitole se budu zabývat tvorbou, šířením zvuku, a také vnímáním zvuku lidským uchem. Čerpal jsem zejména z literatury [1] a [4].

2.1 Šíření zvuku ve vzduchu

Pokud jakýkoli předmět začne vibrovat, poruší tak tlak vzduchu, čímž vzniká zvuk. Zvuk tedy můžeme definovat jako mechanické vlnění v látkovém prostředí, které je schopno vyvolat sluchový vjem.[11]

Musíme rozlišovat pojmy zvuk a tón. Zvuk je jakýkoli sluchový vjem, zatímco tón je periodický zvuk.

Pro jednodušší popis a práci s tóny byly zavedeny určité znaky, které jej popisují. Mezi hlavní charakteristiky patří výška, intenzita, hlasitost, barva a délka.

- *Výška* je dána kmitočtem a udává se v hertzích [Hz]. 1 Hz odpovídá tomu, že 1 kmit trvá 1 sekundu.
- *Intenzita zvuku* I je rovna výkonu zvukového zdroje P dopadající na jednotku plochy S za jednotku času. Jednotkou je Watt na metr čtverečný $[Wm^{-2}]$.
- *Hlasitost* H je subjektivní vnímání závisející zejména na intenzitě, a také na frekvenci zvuku.
- *Barva zvuku* je vlastnost, díky které můžeme například rozeznat různé hudební nástroje, hrající ve stejné výšce. V mluvené řeči, je pak barva zvuku určena existencí harmonických složek, které jsou ve zvuku zdůrazněny.
- *Délka* popisuje, jak dlouho zvuk trvá.

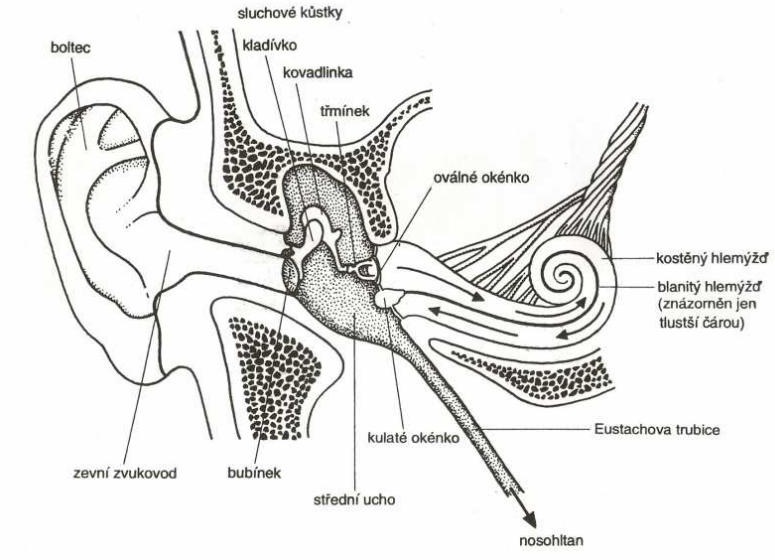
2.2 Vnímání zvuku lidským uchem

Lidské ucho se skládá ze tří částí. První částí je vnější ucho, které se dále dělí na ušní boltec, jehož úkolem je směřovat akustické vlny do další části vnějšího

ucha - do zvukovodu. Zvukovod začíná membránou nazývanou bubínek. Bubínek je velmi pružný a zvukové vlny, které do něj naráží ho rozkmitají. Tyto kmity jsou přeneseny na tři sluchové kůstky zvané kladívko, kovadlinka a třmínek. Tímto převodem se mění původní akustická vlna na vlnu mechanickou, která postupuje dále do vnitřního ucha. Ve vnitřním uchu nalezneme tzv. hlemýždě, který je vyplněný kapalinou. Mechanické vlny narážející do hlemýždě způsobují změnu tlaku, což zachycuje množství nervových zakončení, které tedy převádí mechanické vlny na nervové signály, které jsou poté posílány do mozku.

Lidské ucho není dokonalé, a tak dokáže zpracovat pouze některé zvukové signály. Mladý člověk, je schopen vnímat zvuk v rozsahu přibližně 20 Hz - 20 kHz a v intenzitě do 130 dB. Postupně s věkem se pásmo slyšitelnosti zmenšuje.

Běžná řeč se vyskytuje v mnohem menším pásmu, přibližně mezi 180 Hz - 6 kHz a 30dB až 80 dB.[4]



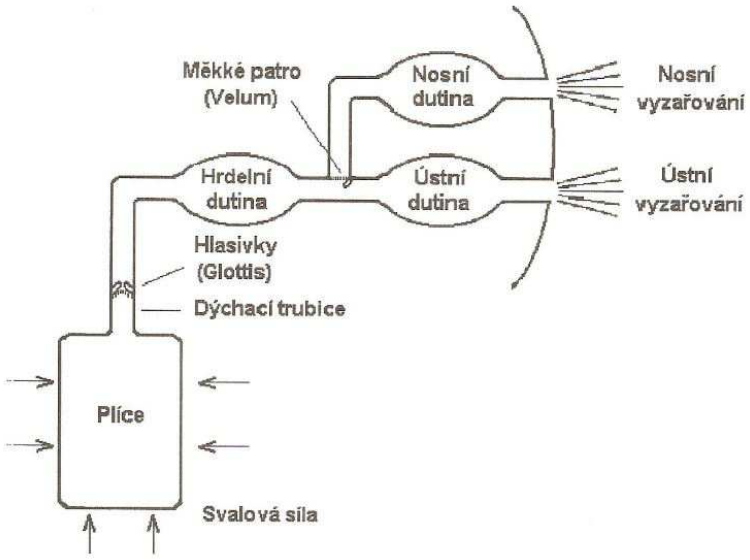
Obr. 1: Lidské ucho, zdroj [2]

2.3 Tvorba zvuku v lidském těle

Následující tři podkapitoly jsem z části převzal a upravil z literatury [4].

Vznik řeči v lidském těle je velice komplikovaná záležitost. Lidský hlas pracuje

na bázi periodického zhušťování vzduchu. Plíce pumpují vzduch do hlasivkové štěrbiny, obklopené hlasivkami. Hlasivky, popřípadě samotný proud vzduchu, otevírá a zavírá hlasivkovou štěrbinu, čímž umožňuje nebo neumožňuje proudění vzduchu dále do vokálního traktu, a tím proměňují proud vzduchu na signál. Následně je tento signál zpracován pomocí vokálního traktu. Vokální trakt se skládá z dutiny hrdelní, dutiny ústní a dutiny nosní. Tyto dutiny mění svůj tvar podle polohy artikulačních orgánů, tedy jazyku, rtů a zubů. Existuje několik modelů tvorby řeči, například elektronický model nebo akustický válcový model. Pro potřeby této práce je však zásadní LPC model tvorby řeči, který je popsán v kapitole 4.



Obr. 2: Model tvorby zvuku (převzato z [4])

2.4 Fonetický popis českého jazyka

Česká mluvená řeč i většina světových jazyků mají podobnou fonetickou strukturu. Věty se skládají ze slov, ty se skládají ze slabik. V mluvené řeči se dále slabiky skládají z fonému, přičemž za foném považujeme každou hlásku, která má schopnost měnit význam slova. Foném je tedy nejmenší jednotka řeči. Český jazyk má 39 fonémů.[8] Fonémy můžeme dále dělit na dvě kategorie - samohlásky a souhlásky.

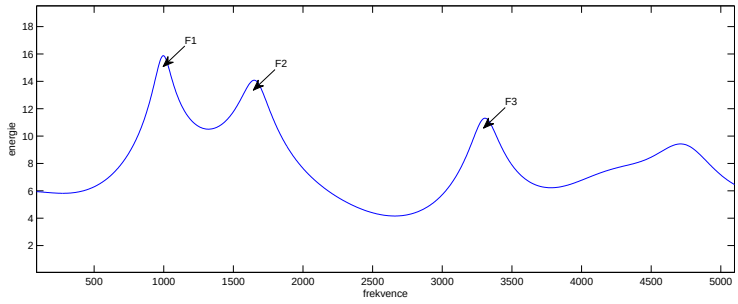
Souhlásky mají poměrně krátkou dobu trvání, a v jejich spektru[1] se vyskytuje šum neboli nepravidelné kmity. Proto se poměrně špatně rozeznávají. Souhlásky se dále rozlišují na znělé a neznělé. Za znělou hlásku považujeme takovou hlásku, při jejímž vyslovení hlasivky téměř periodicky kmitají.

Samohlásek najdeme v češtině celkem deset - a, e, i ,o ,u, á, é ,í ,ó, ú. Všechny samohlásky jsou znělé a jsou charakteristické tím, že mají lokálně přibližně periodický průběh, a také patrné rezonanční kmitočty - formanty.

2.5 Formanty

Vokální trakt mění zvuk vytvořený hlasivkami tak, že je ovlivněno spektrum signálu. To znamená, že některé frekvence v signálu jsou utlumeny a některé jsou naopak posíleny. Frekvence, které jsou nejvíce posíleny, označil Gunnar Fant jako *formanty*. Ve spektrální oblasti jsou tedy formanty reprezentovány lokálními maximy. Formanty také představují rezonanční frekvence jednotlivých dutin. První tři formanty jsou někdy ztotožňovány s největšími dutinami tedy:

$$F_1 = \text{dutina hrdelní}$$
$$F_2 = \text{dutina ústní}$$
$$F_3 = \text{dutina nosní}$$



Obr. 3: Logaritmické LPC spektrum samohlásky „a“ s vyznačenými formanty

Formantů ve spektru můžeme najít 5 i více. Pro účely této práce jsou však stěžejní zejména první dva formanty, podle kterých je možno odhadnout, o jakou samohlásku se jedná.

[1]Spektrum je podrobně vysvětleno v kapitole 3 a 4

3 Fourierova transformance

Obr. 4: Jean-Babtiste Joseph Fourier, zdroj [5]

Fourierova transformace dostala svůj název po francouzském matematikovi Jean-Babtiste Joseph Fourierovi, žijícím na přelomu 18. století. Základní myšlenkou předcházející Fourierově transformaci bylo to, že lze libovolnou periodickou funkci reprezentovat trigonometrickou řadou. To znamená, že jí lze rozložit na součty sinusoid s různými periodami. K takovémuto rozkladu se používají Fourierovy řady. Podmínku periodicity pak zjemňuje Fourierův integrál, který předpokládá, že každá funkce je periodická, s tím, že některé funkce jsou periodické s nekonečnou periodou. V této práci se nebudu zabývat konkrétním rozborem Fourierovy řady, ani Fourierovým integrálem. Tento rozbor je popsán v [1].

Fourierova transformace je v dnešní době velice využívaná. Používá se například při řešení diferenciálních rovnic, při rychlém řešení násobení velkých čísel, v oblasti zpracování obrazu a podobně. Je také základem zpracování a analyzování signálu. Fourierova transformace převádí signály z časové oblasti do oblasti frekvenční. V různé literatuře, se definice Fourierovy transformace mírně liší, v této

práci ale použiji definici tuto:

Definice 3.1. Nechť $f(t)$ je reálná nebo komplexní funkce s s reálnou proměnnou t. Pak její Fourierova fransformace $\widehat{f}(\nu)$ s reálnou proměnnou ν je definovaná jako

$$\widehat{f}(\nu) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i\nu t}dt, \tag{1}$$

kde t reprezentuje čas a ν reprezentuje frekvenci. $\left|\widehat{f}(\nu)\right|$ nazývá se *spektrem* a udává, kolik energie obsahuje funkce $f(t)$ ve frekvenci ν. *Inverzní Fourierova transformace* je potom dána vztahem:

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(\nu)e^{2\pi i\nu t}d\nu \tag{2}$$

Inverzní Fourierova transformace je „opakem“ Fourierovy transformace. Převádí totiž signál z frekvenční domény do domény časové.

Poznámka 3.1. Fourierova transformace má několik značení například námi použité $\widehat{f}(\nu)$ nebo Ff, $Ff(x)(\nu)$.

3.1 Diskrétní Fourierova transformace

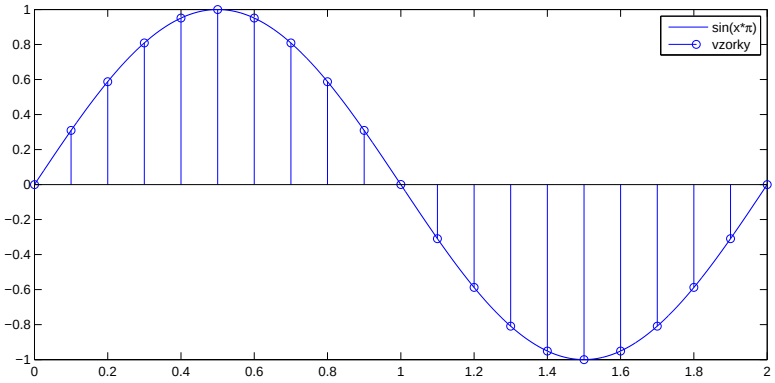
Fourierova transformace je vhodná pro použití při zpracování spojitých signálů, ale i tak někdy neexistuje analytické řešení a Fourierova transformace se musí spočítat numericky. Problém také nastane, když chceme zpracovat digitální, tedy diskrétní, signály. U těchto signálů nemáme spojitou funkci, a tedy Fourierovu transformaci nemůžeme použít. V dnešní době je však většina signálů zpracovávána pomocí počítačů a tedy digitálně. Proto byla zavedena diskrétní Fourierova transformace, zkráceně DFT, která pracuje právě s diskretizovanými funkcemi.

3.2 Vzorkování a kvantování

Ještě před tím, než si popíšeme DFT, si ukážeme, jak signály digitálně zaznamenáváme.

V praxi se zvukové signály nahrávají tím, že z jejich původně spojitého signálu pořídíme mnoho vzorků (samplů) a tím ho převedeme na signál digitální. Takovýto převod obstarává A/D (analogově/digitální) převodník, který nejprve provede vzorkování a poté kvantování.

Vzorkování se provádí tak, že převodník v periodických intervalech odebere vzorek. Počet vzorků, zaznamenaných za jednu sekundu se nazývá vzorkovací frekvence [Hz].



Obr. 5: Příklad vzorkování spojité funkce s frekvencí 10 Hz

Při převodu ze spojité funkce na diskrétní může dojít k jevu zvanému aliasing neboli česky falšování. K aliasingu dochází při nedodržení Shannonova teorému (někdy nazývaný Nyquistův teorém), který říká, že vzorkovací frekvence f_v musí být alespoň dvakrát větší než nejvyšší frekvence v signálu f_m, aby byl původní signál beze ztráty informace rekonstruovatelný ze signálu původního.

$$f_m < \frac{f_n}{2}. \tag{3}$$

Jako příklad aliasingu si můžeme připomenout filmy se záběry na rychle se točící kola, kdy se může zdát, že se kola točí opačným směrem, než by měly nebo si můžeme představit hodiny, na které se díváme každých 50 minut, a proto se může zdát, že velká ručička hodin jde pozpátku. Matematicky je aliasing popsán například v literatuře [1], odkud je převzat následující příklad.

Příklad 3.1. Nechť $f(t)$ je libovolná funkce cos s frekvencí ν, například

$$f(t) = A\cos(2\pi\nu t).$$

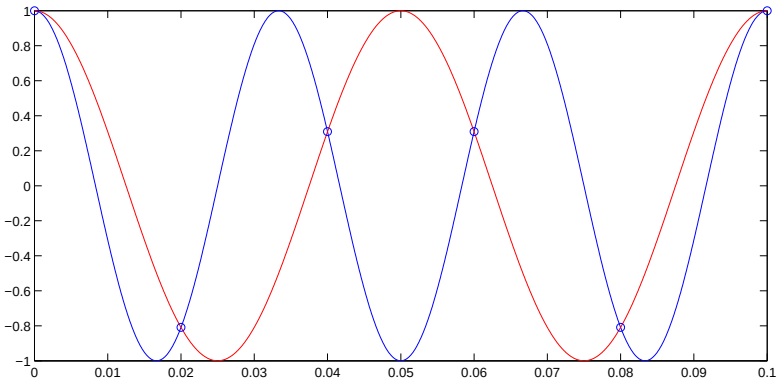
Mějme vzorkovací frekvenci $N = 1/\Delta t$ vzorků za vteřinu a hodnotu funkce v M-tém vzorku danou jako

$$f(M/N) = A\cos(2\pi\nu M/N).$$

Pokud ν je větší než $N/2$, například $N/2 + \alpha$ pak

$$f(M/N) = A\cos(2(N/2+\alpha)M\pi/n)$$
$$f(M/N) = A\cos(M\pi + 2\alpha M\pi/n)$$
$$f(M/N) = (-1)^M A\cos(2\alpha M\pi/n).$$

Změna znaménka α nic nemění na těchto rovnicích, dostaneme tedy stejný výsledek jako u vlny $\nu = N/2 - \alpha$ místo $\nu = N/2 + \alpha$. Na obrázku 6 pak vidíme, že vzorky jsou v tomto výpočtu ty samé body, kde se grafy funkcí $A\cos(2(N/2+\alpha)\pi t)$ a $A\cos(2(N/2-\alpha)\pi t)$ kříží. V podstatě se tedy může frekvence, která je větší než polovina vzorkovací frekvence zdát jako menší frekvence, což znehodnocuje výsledky Fourierovy transformace.



Obr. 6: Aliasing - křížení grafů funkcí

Aby se zamezilo aliasingu, ještě před použitím převodníku se používá tzv. antialiasing filtr, který odstraní všechny frekvence tak, aby byl splněn Shannonův teorém, resp. odstraní všechny frekvence větší než polovina vzorkovací frekvence.

Druhým úkolem A/D převodníku je kvantování neboli přiřazení hodnot na ose y. Počítačově lze čísla vyjádřit pouze s určitou přesností neboli v určitém

kvantilu, od toho název kvantování. Hodnota každého vzorku se pak vyjadřuje jako N-tá mocnina čísla 2. [1]

3.3 DFT

Jak je popsáno výše, diskrétní Fourierova transformace je v praxi mnohem více využívaná než její spojitá verze. Proto si ji popíšeme.

Pokud zobecníme rovnici (1) pro diskrétní případ, kdy platí, že vstupní signál $f(t)$ je nyní roven diskrétní posloupnosti $f(t_k) = \{f_0, f_1, ..., f_{N-1}\}$ a označením $f_k \equiv f(t_k)$, kde $t_k \equiv \Delta k$ pro k=0,1,...,N-1, pak Diskrétní Fourierova transformace f_n je definována jako

$$F_k = \sum_{n=0}^{N-1} f_n e^{-2\pi i \frac{nk}{N}}. \tag{4}$$

Inverzní diskrétní Fourierovu transformaci (IDFT) pak dostaneme úpravou rovnice (4), kterou vynásobíme $e^{2\pi i \frac{kl}{N}}$ a vytvoříme součty podle k.

$$\sum_{k=0}^{N-1} F_k e^{2\pi i \frac{kl}{N}} = \sum_{n=0}^{N-1} f_n \left[\sum_{k=0}^{N-1} e^{2\pi i \frac{k(l-n)}{N}} \right] \tag{5}$$

Pokud $l \neq n$ pak suma v hranatých závorkách je rovna N a pokud $l = n$ pak je tato suma rovna nule. Dostaneme tedy:

$$f_l = \frac{1}{N} \sum_{k=0}^{N-1} F_k e^{2\pi i \frac{kl}{N}}, \tag{6}$$

kde $l = 0, 1, 2, ..., N-1$.

Podobně jako u spojité verze Fourierovy transformace f_k označuje signál, v tomto případě diskrétní, F_n označuje diskrétní Fourierovu transformaci, jejíž množina hodnot se nazývá spektrum.

Poznámka 3.2. Rozlišujeme několik druhů spekter.

1. Komplexní spektrum:

$$X(k) = \widehat{f}(\nu) \tag{7}$$

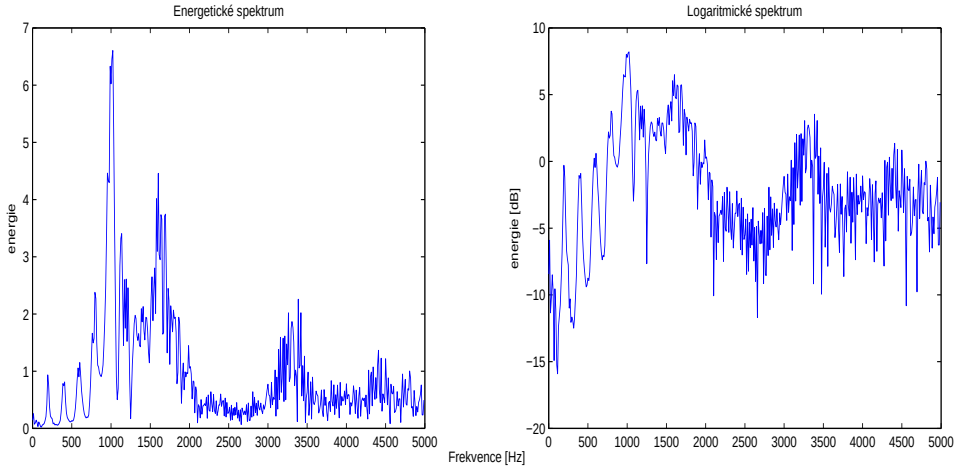
2. Energetické spektrum

$$E_X(k) = \left|\widehat{f}(\nu)\right| \tag{8}$$

3. Logaritmické spektrum

$$LS_X(k) = 10\log_{10}\left|\widehat{f}(\nu)\right|^2 \tag{9}$$

při použití logaritmického spektra získáme na ose y energii v decibelech [dB]. [4]



Obr. 7: Různé podoby spekter stejného signálu získaného pomocí DFT

Poznámka 3.3. Na obrázku 7. je zobrazena pouze polovina spekter. Druhá polovina je pak zrcadlově stejná jako první.

Uveďme si příklad výpočtu DFT převzatý z literatury [1].

Příklad 3.2. Nechť N=4. Potom čísla $e^{\frac{2\pi ik}{N}}$ jsou čísla komplexní rozmístěná na jednotkové kružnici. Platí tedy:

$$\begin{aligned} e^0 &= 1 \qquad (k=0) \\ e^{\frac{2\pi i}{4}} &= i \qquad (k=1) \\ e^{\frac{4\pi i}{4}} &= 1 \qquad (k=2) \\ e^{\frac{6\pi i}{4}} &= -i \qquad (k=3) \end{aligned}$$

Rovnice DFT tedy vypadají takto:

$$F_0 = f_0 + f_1 + f_2 + f_3$$
$$F_1 = f_0 - if_1 - f_2 + if_3$$
$$F_2 = f_0 + f_1 + f_2 + f_3$$
$$F_3 = f_0 + if_1 + f_2 + if_3$$

A rovnice IDTF takto:

$$f_0 = \tfrac{1}{4}(F_0 + F_1 + F_2 + F_3)$$
$$f_1 = \tfrac{1}{4}(F_0 - iF_1 - F_2 + iF_3)$$
$$f_2 = \tfrac{1}{4}(F_0 + F_1 + F_2 + F_3)$$
$$f_3 = \tfrac{1}{4}(F_0 + iF_1 + F_2 + iF_3)$$

3.4 Rychlá Fourierova transformace

Rychlá Fourierova transformace (neboli FFT z anglického Fast Fourier Transform, nebo Cooley-Tukeyho algoritmus) je efektivní algoritmus pro výpočet DFT. FFT se v dnes používá prakticky ve všech programech, které s Fourierovou transformací pracují. Výjimkou není ani MATLAB, Maple nebo Mathematica.

Vysvětlení rychlé Fourierovy transformace jsem převzal z literatury [1].

Předpokládejme, že M je sudé. Pak můžeme pravou stranu rovnice (4) rozdělit na dvě sumy - na sudé a liché části:

$$F_n = \sum_{n=0}^{\frac{N}{2}-1} f_{2k} e^{-2\pi i \frac{(2n)k}{N}} + \sum_{n=0}^{\frac{N}{2}-1} f_{2k+1} e^{-2\pi i \frac{(2n+1)k}{N}} \tag{10}$$

Můžeme si všimnout, že hodnota $F_{k+\frac{N}{2}}$ je podobná hodnotě F_k. Dále víme, že $e^{-\pi i(2n)k} = 1$ a $e^{-\pi i(2n+1)k} = (1)^k$. Dosazením dostaneme:

$$F_{k+\frac{N}{2}} = \sum_{n=0}^{\frac{N}{2}-1} f_{2k} e^{-2\pi i \frac{(2n)k}{N}} + (-1^k) \sum_{n=0}^{\frac{N}{2}-1} f_{2k+1} e^{-2\pi i \frac{(2n+1)k}{N}} \tag{11}$$

Takže pro výpočet hodnoty $F_{k+\frac{N}{2}}$ nám stačí polovina práce oproti klasické DFT plus čas potřebný na sčítání a odčítání výsledků. Dva součty ve výše uvedené

rovnici jsou také diskrétními Fourierovými transformacemi (kdy pravou stranu násobíme $e^{\frac{-2\pi ik}{N}}$) pro $N/2$ bodů místo N bodů, takže když je $\frac{N}{2}$ sudé, můžeme celý proces znovu aplikovat.

Příklad 3.3. Aplikace FFT na Příklad 2.2.

$$F_0 = (f_0 + f_2) + (f_1 + f_3)$$
$$F_2 = (f_0 - f_2) - (f_1 + f_3)$$
$$F_2 = (f_0 - f_2) - i(f_1 - f_3)$$
$$F_3 = (f_0 - f_2) + i(f_1 - f_3)$$

Algoritmus FFT je nejsilnější, pokud je N mocnina čísla 2. Pak je potřeba spočítat přibližně $2N \log_2(N)$ operací namísto N^2 při použití DFT.

Poznámka 3.4. Uvědomme si, že obvyklá vzorkovací frekvence je 44100 vzorku za sekundu. Při tomto množství vzorků bychom pomocí DFT museli spočítat $44100^2 \cong 2{\cdot}10^9$ operací. Naproti tomu pomocí FFT stačí spočítat $2{\cdot}\log_2(44100) = 1,4 * 10^6$.

4 Lineární predikce

Lineární predikce, zkráceně LPC (*Linear predictive coding*), je velmi využívaná metoda v oblasti zpracování signálů. LPC slouží k odhadnutí vzorku signálu z minulých p vzorků. Matematicky jej vystihuje rovnice:

$$\widehat{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{12}$$

kde $\hat{s}(n)$ je tedy odhad n-tého členu v posloupnosti, který je nalezen jako lineární kombinace p předchozích členů, p je počet LPC koeficientů neboli řád prediktoru a α_k jsou predikční koeficienty. Hledáme tedy optimální α_k tak, aby se každá hodnota signálu byla co nejblíže lineární kombinaci předchozích hodnot. Pomocí koeficientů α_k,tak jednoznačně určíme charakteristiky periodického signálu.

Poznámka 4.1. $\hat{s}(n)$ v dalším textu značí pouze odhad n-tého členu v posloupnosti, nikoliv Fourierovu transformací.

Pro periodický signál s periodou L je zřejmé, že $s(n) = s(n-L)$. LPC ale odhaduje samply z počtu koeficientů mnohem menších než N.

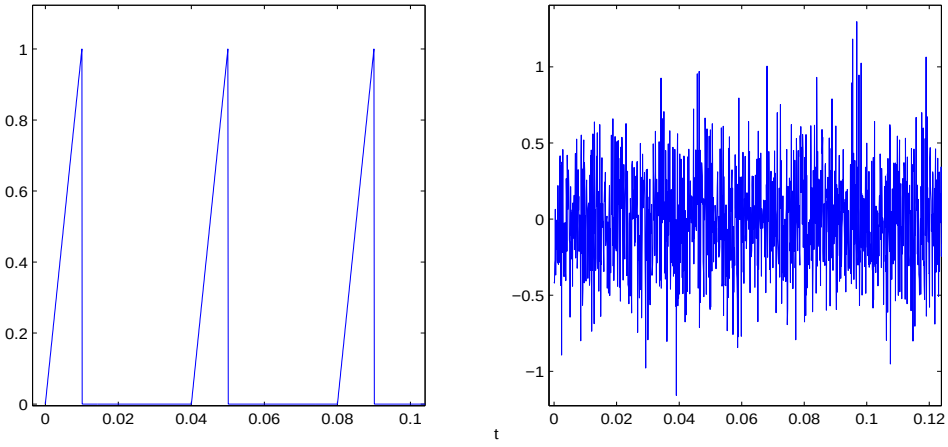
4.1 LPC model řeči

LPC model řeči se asi nejvíce hodí pro oblast zpracování řeči. Tímto modelem se dá fyzikálně popsat tvorba řeči, a navíc se dá matematicky velmi dobře zapsat.

Signál, vytvořený hlasivkami se nazývá budící signál, který je, jak je popsáno v kapitole 2, přeměněn vokálním traktem. Za budící signál považujeme buď bílý šum, nebo jednotkový vlak pulsů, v závislosti na tom, zda modelujeme souhlásky nebo samohlásky, resp. znělé či neznělé zvuky. Bílý šum je náhodný signál, v němž jsou rovnoměrně zastoupeny všechny frekvence. Jednotkový vlak pulsů jsou stále stejné, opakující se pulsy, vystihuje jej rovnice:

$$v(n) = \sum_{k=0}^{p} \delta(n - kL), \tag{13}$$

kde $\delta(n)$ je jednotkový impuls a N je požadovaná perioda výstupního signálu.



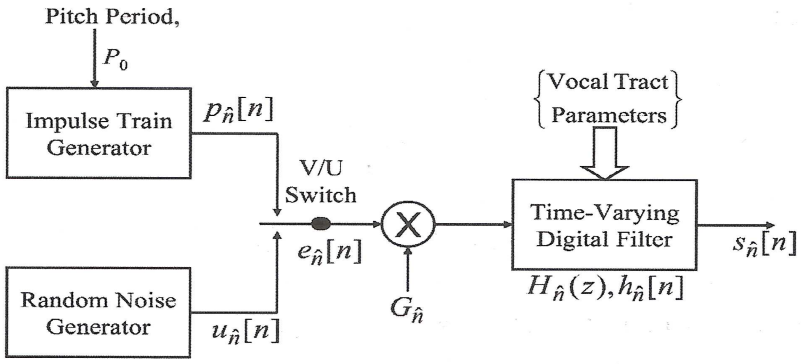
Obr. 8: Jednotkový vlak pulsů (vpravo) a bílý šum (vlevo)

Nyní převedeme výše popsaný model převést do matematického vyjádření. Musíme ovšem rozlišovat mezi modelem produkce řeči, který ji skutečně vystihuje a LPC modelem, kterým řeč aproximujeme.

Nejprve si ukážeme, jak vypadá skutečný model řeči. Vystihuje jej rovnice:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \tag{14}$$

kde G je zesílení, a $u(n)$ modeluje přísun vzduchu přes hlasivky neboli budící signál.



Obr. 9: Tvorba řeči, metoda LPC, zdroj [2]

V dalším textu použijeme *z-transformaci*, proto si ji popíšeme.

Z-transformace se používá zejména pro spektrální analýzu signálu. Z-transformace převádí vzorky x(n) na komplexní funkce komplexní proměnné X(z).

$$X(z) = \sum_{0}^{\infty} x(n) z^{-n} \tag{15}$$

Použijeme-li z-transformaci na rovnici (15) dostaneme:

$$S(z) = \sum_{k=1}^{p} a_k S(z) z^{-k} + GU(z) \tag{16}$$

Rovnici dále upravíme:

$$S(z) - \sum_{k=1}^{p} a_k S(z) z^{-k} = GU(z) \tag{17}$$

$$S(z)(1 - \sum_{k=1}^{p} a_k z^{-k}) = GU(z) \tag{18}$$

A označíme $H(z)$ jako:

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{19}$$

Platí tedy:

$$S(z) = H(z) \cdot GU(z) \tag{20}$$

Poznámka 4.2. $H(z)$ se označuje jako přenosová funkce s koeficienty a_k.

LPC model vystihuje již zníněná rovnice:

$$\widehat{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{21}$$

S využitím z-transformace:

$$\widehat{S}(z) = \sum_{k=1}^{p} \alpha_k z^{-k} S(z) \tag{22}$$

A označíme $P(z) = \sum_{k=1}^{p} \alpha_k z^{-k}$. Pak platí:

$$P(z) = \frac{\widehat{S}(z)}{S(z)} \tag{23}$$

Dále definujeme predikční chybu n-tého vzorku jako rozdíl skutečné a odhadované hodnoty

$$e(n) = \widehat{s}(n) - s(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \tag{24}$$

$$E(z) = S(z) - \sum_{k=1}^{p} \alpha_k z^{-k} S(z) \tag{25}$$

Úpravou dostaneme:

$$\frac{E(z)}{S(z)} = \frac{S(z)(1 - \sum_{k=1}^{p} \alpha_k z^{-k})}{S(z)} = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \tag{26}$$

Poslední část rovnice označíme jako $A(z)$. Platí tedy:

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \tag{27}$$

Na chybový signál $e(n)$ se tedy můžeme dívat jako na filtraci signálu $s(n)$ filtrem $A(z)$ s koeficienty α_k. Platí tedy:

$$E(z) = S(z) \cdot A(z) \tag{28}$$

Takovémuto filtru ($A(z)$) se říká analyzující filtr.

Pokud LPC model odhadne skutečný model správně, tedy $\alpha_k = a_k$, pro $k = 1, 2, ...p$, pak

$$e(n) = Gu(n), \tag{29}$$

tedy chyba predikce je rovna budícímu signálu a navíc $A(z)$ je inverzním filter pro $H(z)$ tedy:

$$H(z) = \frac{1}{A(z)} \tag{30}$$

Touto inverzní filtrací naopak můžeme získat analyzující signál tím, že chybový signál filtrujeme syntetizujícím filtrem s přenosovou funkcí $H(z)$. Platí tedy:

$$S(z) = E(z) \cdot H(z) \tag{31}$$

protože $E(z) = GU(z)$ platí také:

$$S(z) = GU(z) \cdot H(z) \tag{32}$$

Připomeňme, že $S(z)$ je výstupní signál, $U(z)$ budicí signál a $H(z)$ reprezentuje vokální trakt. Zvhledem k tomů, že G je konstanta, můžeme pro jednoduchost definovat $G(z) = GU(z)$.

4.2 Výpočet LPC koeficientů

Výpočet koeficientů α_k je základní úlohou pro analýzu signálů pomocí metody LPC.

Nyní připomeňme predikční chybu n-tého vzorku

$$e(n) = s(n) - \widehat{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k), \tag{33}$$

kterou se snažíme minimalizovat přes celý segment signálu. Minimalizujeme tedy účelovou funkci

$$J = \sum_{n=1}^{N} e^2(n) \tag{34}$$

K výpočtu LPC koeficientů se používá zejména autokorelační metoda. Další používanou metodou je kovariační metoda, která je popsaná například v [3]

4.2.1 Autokorelační metoda

Jak již bylo naznačeno, autokorelační metoda je nejpoužívanější, zejména pro svou robustnost a stabilu. Tato metoda vychází ze dvou předpokladů:
1. Signál $s(n)$ má konečný počet prvků N
2. Mimo tento signál jsou vzorky nulové

Účelovou funkci můžeme rozepsat jako

$$J = \sum_{n=1}^{N} e^2(n) = \sum_{n=1}^{N} [s(n) - \widehat{s}(n)]^2 = \sum_{n=1}^{N} \left[s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \right]^2 \tag{35}$$

Pro minimalizaci položíme parciální derivace funkce (36) podle α_k rovny nule

$$\frac{\partial J}{\partial \alpha_k} = 0; \tag{36}$$

což vede na systém p rovnic s p neznámými koeficienty α_k.

Po výpočtu a úpravě těchto rovnic, uvedeno například v literatuře [3], dostaneme:

$$\sum_{k=1}^{p} \alpha_k R(|k-j|) = R(j) \qquad pro\ j = 1, 2, ...p \tag{37}$$

kde

$$R(j) = \frac{1}{N} \sum_{n=1}^{N-j-1} s(n)s(n-j). \tag{38}$$

Tyto rovnice se nazývají *Yuleovy-Walkerovy rovnice* a lze je přepsat do maticového tvaru, aby se s ním lépe pracovalo. Takováto matice se pak nazývá *autokorelační matice*.

$$\begin{pmatrix} R(0) & R(1) & R(2) & \dots & R(M-1) \\ R(1) & R(0) & R(1) & \dots & R(M-2) \\ \vdots & \vdots & \vdots & & \\ R(M-1) & R(M-2) & R(M-3) & \dots & R(0) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ R(M) \end{pmatrix} \tag{39}$$

neboli

$$\mathbf{R}\alpha = \mathbf{R_p} \tag{40}$$

Vzhledem k tomu, že známe autokorelační koeficienty $R(0), R(1), ..., R(M)$ můžeme rovnici vyřešit pomocí násobení inverzní maticí R zleva. Rychlejší a používanější metodou řešení rovnice je ovšem *Levinson-Durbinův algoritmus*, který můžeme použít díky tomu, že matice je *ekvidiagonální*, tedy že prvky na její hlavní diagonále jsou si rovny. Takováto matice se nazývá *Toeplitzova*.

4.2.2 Levinson-Durbinův algoritmus

Jak již bylo naznačeno Levinson-Durbinův algoritmus je využívaný zejména kvůli jeho rychlosti, jednoduchosti a poměrně malým výpočetním nárokům. Tento algoritmus je založen na rekurzivním výpočtu, který jsem převzal z literatury [3]:

$$P_0 = R(0) \tag{41}$$

$$\alpha_1^{(1)} = k_1 = \frac{R(1)}{R(0)} \tag{42}$$

$$P_1 = P_0 \cdot (1 - k_1^2) \tag{43}$$

Pro $m = 2, 3, ..., p$:

$$a_m^{(m)} = k_m = -\frac{R(m) + \sum_{j=1}^{m-1} a_j^{(m-1)} R(m-j)}{P_{m-1}} \tag{44}$$

$$a_j^{(m)} = \alpha_j^{(m-1)} - k_m \alpha_{m-j}^{(m-1)},$$

kde j=1,2,...,m-1

(45)

$$P_m = P_{m-1}(1 - k_m^2) \tag{46}$$

$$\alpha_i = \alpha_i^{(p)},$$

kde i=1,2,...,p

(47)

Výsledné α_i jsou hledané autokorelační koeficienty.

Ještě nám zbývá odhadnout parametr G. Předpokládáme, že energie [2] skutečného signálu a signálu odhadnutého pomocí LPC jsou stejné.

$$G^2 \sum_{n=0}^{N-1} u^2(n) = \sum_{n=0}^{N-1} e^2(n) \tag{48}$$

Pokud má budící signál normalizovanou energii rovnu 1, pak je parametr G roven [7]:

$$G^2 = R(0) - \sum_{p}^{k=1} R(k). \tag{49}$$

4.3 LPC spektrum

Vztah (33) se dá vyjádřit ve frekvenční doméně vyčíslením nad jednotkovou kružnicí substitucí $z = e^{i\Theta}$.

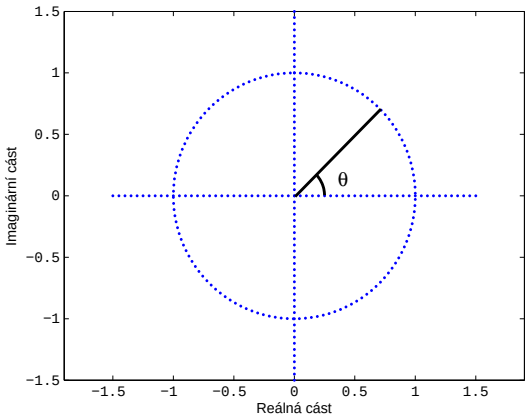
$$S(e^{i\Theta}) = H(e^{i\Theta}) \cdot G(e^{i\Theta}) \tag{50}$$

kde Θ je normovaná frekvence definovaná vztahem

[2]Enegrie signálu s konečné délky N je rovna : $E = \sum_{N-1}^{n=0} x^2(n)$

$$\Theta = 2\pi \frac{f}{f_s} = 2\pi f T = \omega T \tag{51}$$

kde f je skutečná frekvence v Hz, f_s je vzorkovací frekvence, T je vzorkovací krok v sekundách a ω je úhlová frekvence v radiánech za sekundu.



Obr. 10: Grafické vyjádření Θ na jednotkové kružnici

Víme, že budící signál je realizován bílým šumem a periodickým vlakem pulsů, jejichž spektrální hustota je konstantní. Proto platí, že spektrum signálu je dané pouze spektrem syntetizujícího filtru a parametrem zesílení G.

$$S(e^{i\Theta}) = H(e^{i\Theta}) \cdot G \tag{52}$$

Pro energické spektrum pak platí:

$$\left|H(e^{i\Theta})\right|^2 = \left|\frac{G}{A(e^{i\Theta})}\right|^2 \tag{53}$$

Často se uvažuje normované LPC spektrum s jednotkovým zesílením [4]. Díky z-transformaci vyjádřené jako:

$$S(f) = \left|\frac{1}{1 - \sum_m \alpha_m z^{-m}}\right|^2, \tag{54}$$

kde $f = z = e^{j\Theta}$, a f je frekvence. Neboli:

$$S(f) = \left| \frac{1}{1 - (\alpha_1 e^{-2i\Theta} + \alpha_2 e^{-i\Theta} + ... + \alpha_M e^{-Mi\Theta})} \right|^2 . \tag{55}$$

Obvykle se používají spektra v logaritmické podobě:

$$LS(f) = 10 \log_{10} [S(f)] \tag{56}$$

Tento tvar se nazývá logaritmické spektrum, a díky této úpravě má na ose x frekvenci v Hz a na ose y dB [4].

4.4 Určení formantů pomocí LPC analýzy

Existuje několik metod jak určit frekvence formantů díky datům získaných z LPC analýzy. My si popíšeme dvě asi nejpoužívanější metody.

První metoda, která již byla naznačena, chápe lokální maxima LPC logaritmického spektra jako formanty.

Další metodou, jak určit formanty je tzv. metoda kořenů, nebo také pólů. Vycházíme z toho, že $H(z)$ charakterizuje vokální trakt, a proto obsahuje informaci o spektru. V rovnici (31) se zaměříme na jmenovatele $H(z)$. $A(z)$ pak můžeme přepsat do tvaru:

$$A(z) = z^p - \sum_{p}^{k=0} a_k z^{p-k} \tag{57}$$

Tato rovnice je vlastně polynom p-tého stupně, a můžeme u něj určit reálné i komplexní kořeny. Pro ukázku určení komplexních kořenů zvolme p=2. Pak platí:

$$A(z) = z^2 - a_1 z + a_2 \tag{58}$$

$$z^2 - a_1 z + a_2 = (z - z_1)(z + z_2), \tag{59}$$

kde z_1, z_2 jsou komplexní kořeny. Po roznásobení a porovnání koeficientů u stejných mocnin dostaneme:

$$a_1 = z_1 + z_2 \tag{60}$$

$$a_2 = -z_1 z_2 \tag{61}$$

Komplexní kořeny musí být pro reálné koeficienty komplexně sdružené, platí tedy:

$$z_1 = z_0 = r_0 e^{i\Theta} \tag{62}$$

$$z_2 = z_0^{\star} e^{-i\Theta}, \tag{63}$$

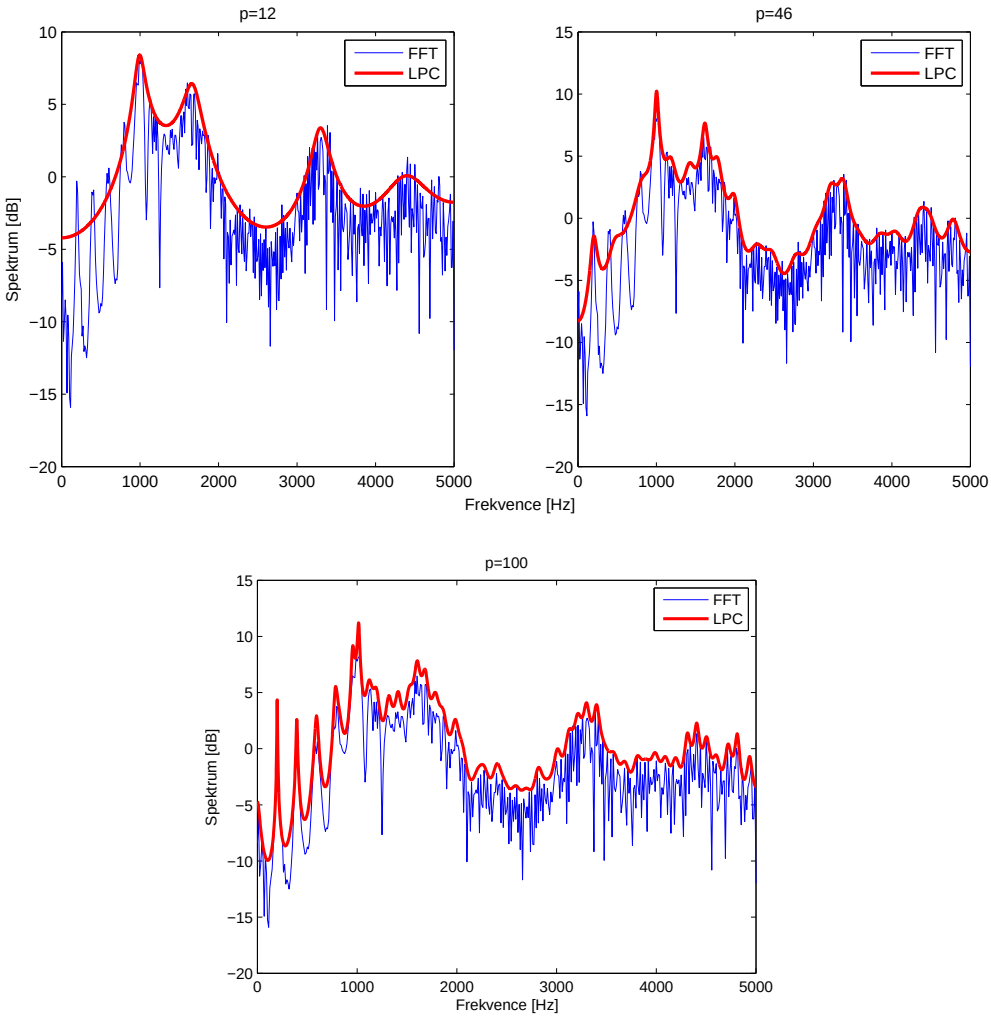
kde r_0 je absolutní hodnota komplexně sdružených pólů (někdy též nazývaný modul) a Θ je normovaná frekvence. Znovu připomeňme, že platí $H(z) = 1/A(z)$, a proto komplexně sdružené kořeny $A(z)$ jsou póly $H(z)$, a proto určují maxima přenosové funkce $H(e^{i\Theta})$. Pokud tedy zvolíme správně počet predikčních koeficientů, obvykle se řídíme pravidlem [3],[4]

$$p \approx f_s/1000 + 2, \tag{64}$$

pak budou tyto kořeny velice blízko formantovým frekvencím.

4.5 Srovnání FFT a LPC

Popsali jsme si dvě metody, konkrétně metodu FFT a LPC, díky kterým můžeme převést signály z časové domény do domény frekvenční. Obě tyto metody se tedy pro spektrální analýzu signálu dají použít. Ukazuje se, že LPC spektrum má podobu vyhlazeného FFT spektra. Na obrázku můžeme vidět FFT logaritmické spektrum a LPC logaritmické spektra při různém počtu predikčních koeficientů p.



Obr. 11: Srovnání logaritmických spekter metodou FFT a LPC

Volba počtu predikčních koeficientů je tedy důležitá a závisí na této volbě to, jako moc vyhlazené LPC spektrum bude. Pro optimální výsledky bychom se měli řídit již zmíněným pravidlem $p \approx \frac{f_s}{1000} + 2$, které se v praxi ukázalo jako dostatečné pro modelování hlavních posílených oblastí, a zároveň ne příliš velké pro zbytečné rozkolísání spektra.

5 Úprava signálu

Před tím, než začneme signál zpracovávat je žádoucí, abychom ho upravili. Děláme to proto, že řečový signál je variabilní a také proto, že signál může obsahovat různé šumy, rušení a okolní zvuky, které komplikují jeho další zpracování a výsledky.

5.1 Preemfáze

Dlouhodobé spektrum řečového signálu klesá asi o 6dB/oktávu [3]. Velká část energie signálu leží v pásmu do 300Hz. Na druhou stranu nás zajímají hlavně ty informace, které leží v pásmu nad 300Hz. Víme, že šum má přibližně rovnoměrné spektrum. Z toho vyplývá, že negativní vlivy šumu působí více na energeticky slabší složky řečového signálu. U samohlásek navíc první formant obvykle převyšuje formanty ostatní. Tyto efekty zmírníme filtrací řečového signálu filtrem s horní propustí

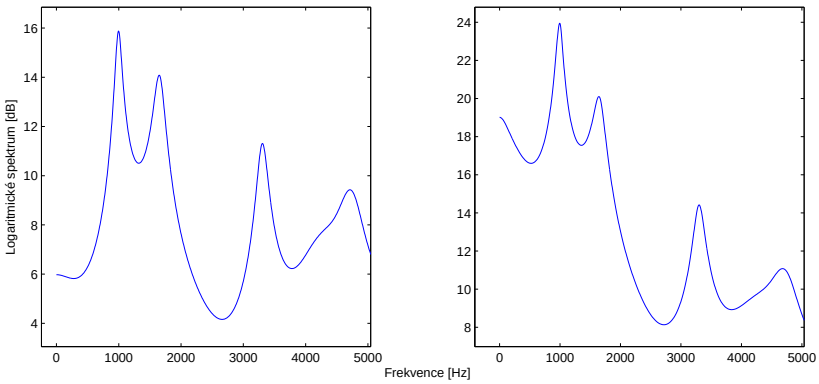
$$H(z) = 1 - \alpha \cdot z^{-1} \tag{65}$$

což v časové doméně naplňuje vztah

$$s_p(n) = s(n) - \alpha s(n-1) \tag{66}$$

Koeficient alfa se obvykle volí v intervalu $\langle 0.9 - 1 \rangle$. [3], [4]

Poznámka 5.1. Filtr s horní propustí je takový filtr, který odstraní frekvence nižší, než určitou mez, respektive propustí frekvence vyšší



Obr. 12: Logaritmické spektrum s (vlevo) a bez (vpravo) použití preemfáze

5.2 Segmentace a windowing

Jak již bylo naznačeno, signál lidské řeči se v čase mění, přičemž předpokládáme, že signál je kvazistacionární, tedy že se mění v čase „pomalu“. Proto se signál zpracovává pomocí krátkodobé analýzy. Signál proto rozdělíme na stejně dlouhé částí (segmenty), které se vzájemně překrývají, čímž jsou více stabilní a mají lépe odhadnutelné chování v čase a frekvenci. Délka segmentu musí být nejen dostatečně malá, aby byly parametry uvnitř segmentu přibližně konstantní, ale i dostatečně velká, aby bylo zachováno dostatečné množství period na přesné odhadnutí frekvence. Tomuto odpovídají segmenty o délce 10 až 30ms [4]. Jednotlivým vzorkům z těchto segmentů se přiřazují váhy pomocí tzv. oken.

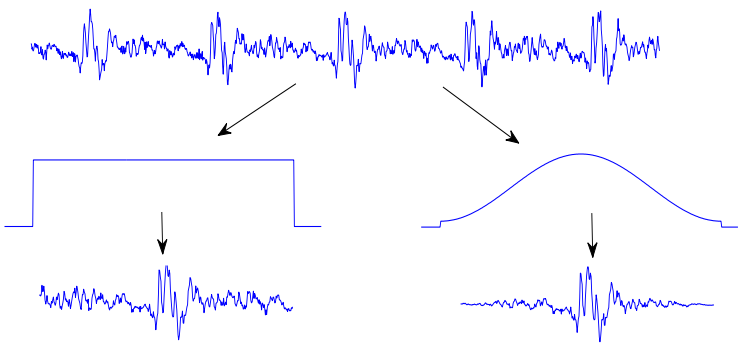
$$s_w(n) = s(n)w(n) \tag{67}$$

V praxi se obvykle využívají dva druhy oken - pravoúhlé a Hammingovo. Pravoúhlé okno je definováno jako:

$$w(n) = \begin{cases} 1 & n = 1, 2, ...N \\ 0 & \text{jinak} \end{cases} \tag{68}$$

Pro zpracování řečových signálů je vhodnější Hammingovo okno, které má maximum uprostřed a má hladké konce, což minimalizuje efekt rozkolísanosti na koncích segmentu. Hammingovo okno má tvar:

$$w(n) = \begin{cases} 0,54 - 0,46\cos(2\pi n/N) & n = 1,2,...N \\ 0 & \text{jinak} \end{cases} \tag{69}$$



Obr. 13: Aplikace Pravoúhlého (vlevo) a Hammingova (vpravo) okna na řečový signál

6 Praktická část

Cílem této práce je tedy vytvořit program, který je schopen rozeznat samohlásky a zhodnotit, jak je daná samohláska „hezká“ respektive kvalitní. Formantové struktury u mluvené řeči mužů a žen se mírně liší, proto jsem se zaměřil pouze na jednu skupinu - ženy. Aby program úspěšně fungoval, pořídil jsem 20 nahrávek ženských hlasů. V těchto nahrávkách ženy přečetly 5 samohlásek - a, e, i, o, u celkem třikrát, a to v různém pořadí. Konkrétně:

A, E, I, O, U,
E, O, A, I, U,
I, A, U, O, E.

Celkově tedy mám 60 nahrávek každé samohlásky, které jsem analyzoval. Všechny nahrávky jsou na přiloženém CD. Označení nahrávek je následující: 1a, 1_2a, 1_3a,...,20u, 20_2u, 20_3u, kdy 1a je první samohláska „a“ přečtená první osobou, 1_2a je druhá samohláska „a“ přečtená první osobou, apod.

Původním záměrem bylo, aby nahrávka každé samohlásky byla dlouhá alespoň jednu sekundu. Tento časový limit se však v praxi projevil jako příliš dlouhý, proto jsem se spokojil s nahrávkami dlouhými alespoň 0,1 sekundy, resp. 100ms, po upravení samohlásky, tedy odstřižení začátků a konců, kdy signál ještě nemusí být periodický.

Nahrávky byly pořízeny v softwaru Praat verze 5.3.43 (dostupné z http://www.fon.hum.uva.nl/praat) ve formátu .wav na mikrofon Genius MIC-01A. Stříhání nahrávek bylo provedeno v programu Audacity 2.0.1 (dostupné z http://audacity.sourceforge.net).

6.1 Analyzování samohlásek

V první části programu analyzuji všechny nahrané samohlásky pomocí m-filu *analyza_databaze.m*. Tento m-file zjistí charakteristiky všech samohlásek v databázi, podle metod popsaných níže.

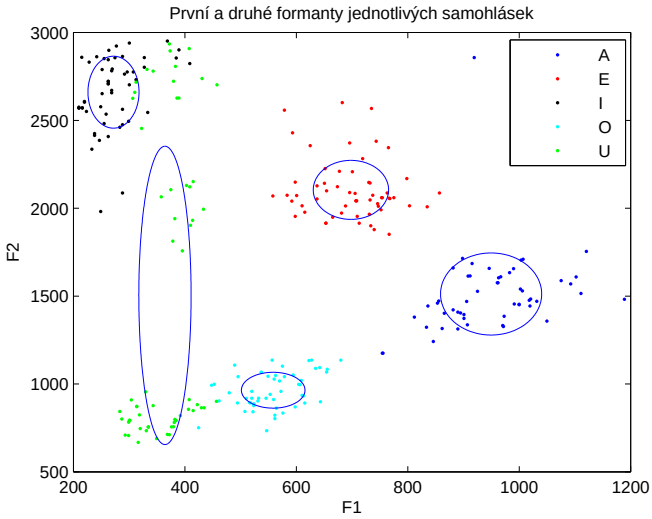
Při analýze samohlásek využívám preemfázi a segmentaci s okny trvajícími

30 ms, které se překrývají v 5 ms intervalech. Zároveň používám Hammingovské okno a také LPC analýzu s predikčním koeficientem $p = 12$.

V programu jsem použil celkem 6 metod pro rozpoznání samohlásek. Část z nich je zaměřena na analýzu formantů a část na porovnávání celých LPC logaritmických spekter, dále označovaných jako spektra.

6.1.1 Metoda 1 - Formanty pomocí peaků

V této části m-file najde formanty metodou peaků. Za formanty tedy považujeme lokální maxima na spektru. Dále zpracuje všechny první a druhé formanty získané v první části[3], a vytvoří elipsy v rovině, které charakterizují jednotlivé samohlásky. Střed elips položíme rovno středním hodnotám prvních a druhých formantů a za velikosti poloos považujeme směrodatné odchylky těchto formantů. To, o kterou samohlásku se jedná a jak je kvalitní, určuje vzdálenost k těmto elipsám.

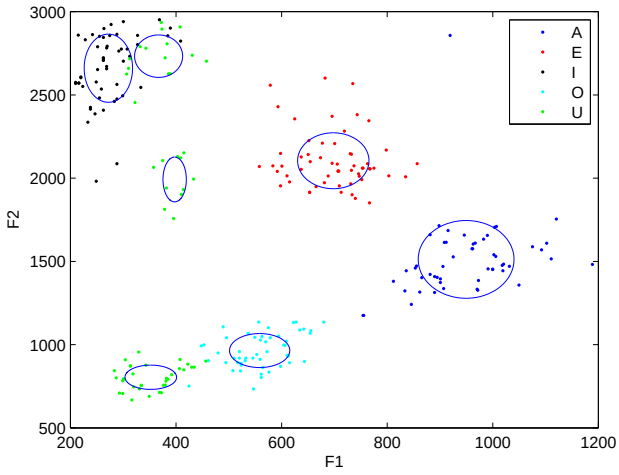


Obr. 14: Srovnání prvních a druhých formantů získaných metodou peaků a vytvořené formantové elipsy

Jak je z obrázku 14. patrné, formanty samohlásek A, E, I a O se nacházejí na přibližně stejných hodnotách. Naproti tomu se druhý formant samohlásky U

[3]Některé hodnoty, které byly extrémně daleko od všech ostatních formantů byly vypuštěny

soustřeďuje do tří oblastí. Důvodem je to, že se první a druhý formant spojil v jeden. Jinými slovy druhý formant nemá díky velikosti prvního formantu tak velkou energii, aby vytvořil lokální maximum. Pokud budeme předpokládat, že jsou druhé formanty samohlásky U spočteny správně, můžeme je rozdělit do tří oblastí.

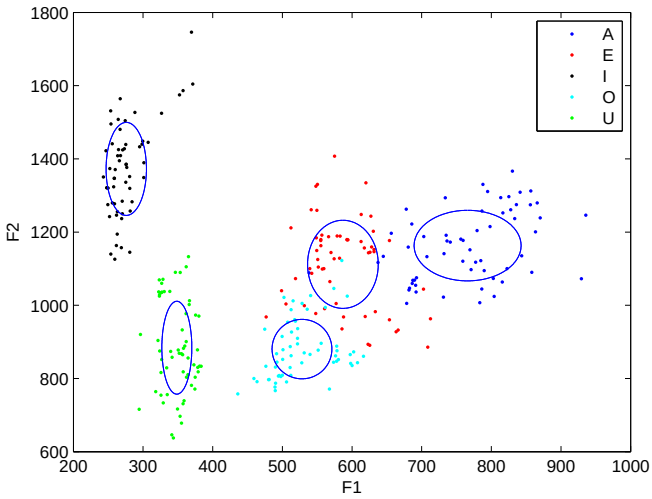


Obr. 15: Srovnání 1. a 2. formantů pomocí metody kořenů a formantové elipsy

Toto rozdělení pomohlo jen částečně. Z obrázku 15. můžeme vypozorovat, že jedna z elips samohlásky U a elipsa samohlásky I jsou hodně blízko u sebe. Proto touto metodou může dojít ke zkreslení výsledků, respektive může docházek k záměně samohlásek I a U. Proto v této metodě použijeme pouze jednu elipsu pro samohlásku U, ale konstatuji, že tato metoda není vhodná pro rozpoznání a klasifikaci samohlásky U. Dalším problémem je blízkost elips samohlásek O, U, které se při menším zkomolení samohlásky také mohou plést.

6.1.2 Metoda 2 - Formanty pomocí kořenů

Formanty můžeme získat různými metodami. V této části programu jsme použili metodu kořenů, popsanou v kapitole (4.4). V ostatních ohledech je metoda stejná jako Metoda 1.



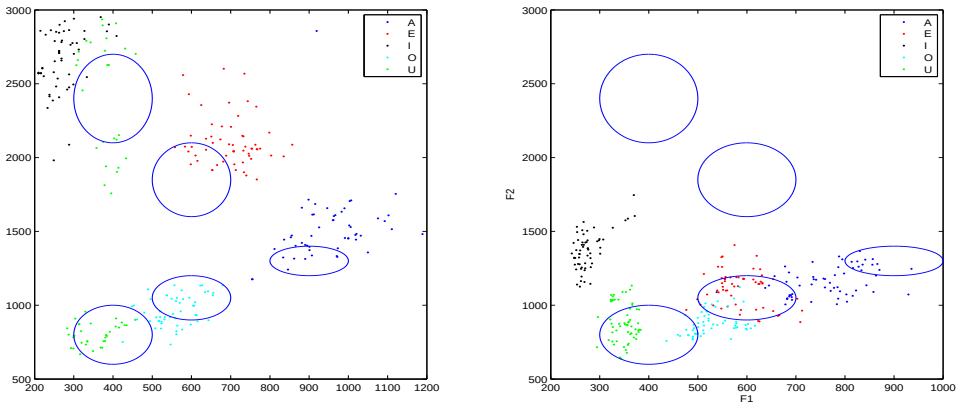
Obr. 16: Srovnání formantů získaných metodou kořenů a vytvořené elipsy

S využitím této metody se postavení formantů poměrně razantně mění. Pokud srovnáme hodnoty formantů získaných touto metodou s hodnotami, kde by se formanty měly vyskytovat, zjistíme, že první formanty jsou na podobných pozicích, ale druhé formanty, zejména u samohlásky I jsou menší, než jaké by měly podle literatury i podle předchozí metody být.

I přesto je tato metoda využitelná, a to zejména pro rozlišení samohlásky O a U.

6.1.3 Metoda 3 - Formanty podle literatury

Tato metoda je podobná dvěma předcházejícím. Rozdílem je, že elipsy sestavíme z hodnot nalezených v literatuře [4].



Obr. 17: Srovnání formantů s elipsami získanými z literatury

Na obrázku porovnáváme již zmiňované elipsy sestavené podle literatury a námi spočtené formanty pomocí metody peaků a metody kořenů. Můžeme si všimnout, že formanty jsou přibližně na stejných místech jako by se měli vyskytovat. Ale i tak jsou zde poměrně velké rozdíly a proto tuto metodu nemůžeme použít.

6.1.4 Metoda 4 - Nejbližší shoda se spektrem

Tato metoda je zcela odlišná od prvních tří metod. V této fázi nás nezajímá pouze poloha formantů, ale zkoumáme podobnost celých spekter. Porovnáváme tedy rozdíly každého spektra jednotlivých samohlásek s námi analyzovanou nahranou hláskou. Protože začátek spekter je důležitější (podobně jako jsme zkoumali pouze první dva formanty), použijeme váhy typu:

$$v(n) = e^{-f(n)/500},$$

kde f je frekvence. Při porovnávání využijeme L^2 normu. Řídíme se tedy rovnicí:

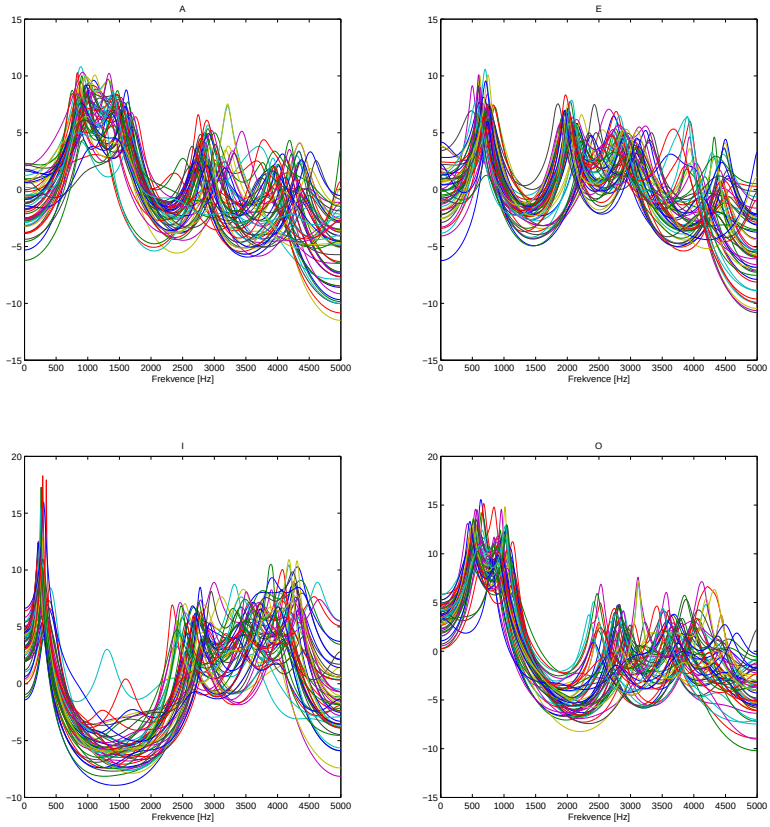
$$D_m = \sum_{n=0}^{N} \left[LS_m(n) - ls(n)\right]^2 \cdot v(n),$$

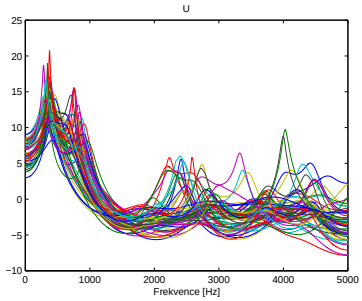
kde $m = 1, 2, ...M$ a M je počet spekter v databázi dané samohlásky, D_m je odchylka od m-tého spektra, LS_m je spektrum jednotlivých samohlásek v databázi, ls je spektrum námi zkoumané samohlásky a v jsou váhy. Hledáme takovou

odchylku, která je minimální. Analyzovanou samohlásku pak označíme za tu samohlásku, jejíž odchylka je minimální. Porovnání nejmenších odchylek jednotlivých samohlásek pak považujeme za kvalitu samohlásky.

6.1.5 Metoda 5 - Porovnání se všemi spektry z databáze

Tato metoda je hodně podobná metodě předcházející. Srovnáváme tedy odchylky mezi spektry samohlásek v databázi s námi analyzovaným spektrem. Rozdílem je, že tentokrát sčítáme všechny odchylky spekter jednotlivých samohlásek tak, jak jsou popsány v Metodě 4. Analyzovanou hlásku pak určíme jako tu, jejíž součet všech odchylek je nejmenší a srovnání součtů odchylek považujeme za kvalitu samohlásky.





Obr. 18: Spektra nahrávek v databázi, rozdělené podle samohlásek

6.1.6 Metoda 6 - Srovnání derivací

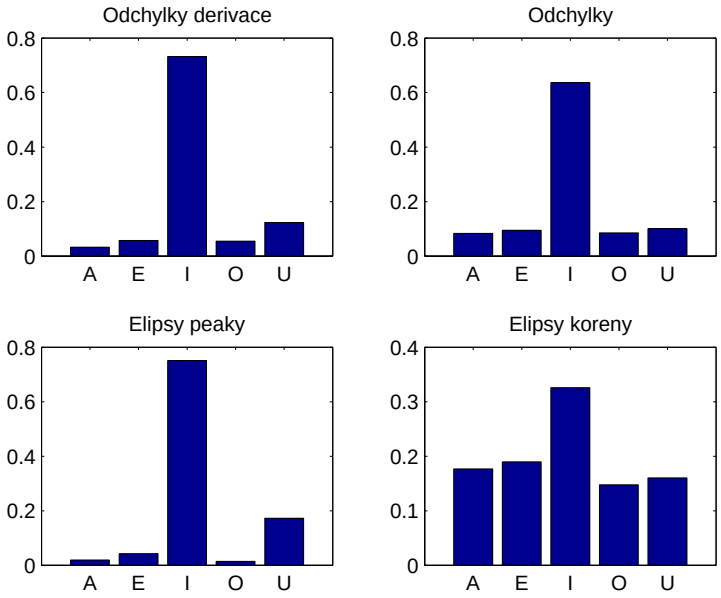
V této metodě znovu vycházíme ze spekter samohlásek v databázi. Nyní ale porovnáváme jejich derivace s derivací námi zkoumaného spektra. Derivaci neporovnáváme absolutně, ale zajímá nás pouze to, jestli je kladná nebo záporná, což odpovídá tomu, kdy je spektrum rostoucí respektive klesající. Vzhledem k tomu, že porovnáváme dvě diskrétní funkce, můžeme odchylku definovat jako počet bodů, kdy se znaménko derivace spektra z databáze liší od analyzovaného spektra, přičemž používáme stejné váhy jako v Metodě 4 a 5.

U metod 4, 5, 6 se ukázalo, že spektra samohlásek O, U si jsou podobná. Naopak se projevilo, že spektra samohlásek A, I, E jsou jedinečná. Lze tedy říci, že tyto metody jsou velice vhodné pro jejich rozpoznání.

6.2 Hodnocení kvality samohlásky

Vzdálenost prvního a druhého formantu v rovině k elipsám je již v podstatě hodnocení kvality samohlásky. Snažíme se co nejvíce se přiblížit k dané elipse nebo nejlépe přímo do ní zapadnout. Podobně se můžeme dívat na metody, kdy porovnáváme grafy jednotlivých spekter, nebo derivace jednotlivých spekter. Kvůli grafickému vyjádření kvality samohlásky ale chceme, aby větší hodnoty znamenaly kvalitnější samohlásku. Proto všechny vzdálenosti a odchylky vyjádříme jako převrácené hodnoty, tedy $1/V$, kde V jsou vzdálenosti nebo odchylky. Hodnoty poté normalizujeme.

Další metodou hodnocení kvality může být to, že budeme postupně rozšiřovat poloosy elips o násobek velikosti jejich poloos tak dlouho, dokud do ní formanty v rovině nepadnou, a přitom zaznamenáváme, kolikrát jsme museli elipsu zvětšit.



Obr. 19: Příklad hodnocení samohlásky I pomocí m-filu *poznej_hlasku.m*

Celkově se ukázalo, že pro rozpoznání samohlásek A, E, I dává lepší výsledky metoda peaků a metody srovnávání spekter a derivací spekter. Samohlásky O, U pak lépe klasifikuje metoda kořenů s přihlédnutím k metodám srovnávání spekter a derivací spekter.

7 Závěr

Při studiu dané problematiky se ukázalo, že využití Fourierovy transformace dává zbytečně mnoho informací o spektru signálu a bylo by velmi složité tyto informace protřídit tak, abychom z nich vytáhli pouze informace pro nás relevantní. V tomto ohledu je mnohem lepší využít LPC analýzu, kdy díky změně predikčních koeficientů můžeme jednoduše určit, kolik informací bude ve spektru obsaženo. Proto je v této práci pro zpracování samohlásek použita téměř výhradně LPC analýza.

Důležitým prvkem pro analýzu jednotlivých samohlásek je zisk dostatečného množství kvalitních samohlásek a jejich kvalitní zvukový záznam. Proto by bylo vhodné sesbírat databázi v rámci stovek nahrávek pro zpřesnění výsledků, a také by byla zajímavá možnost použití profesionálního nahrávacího zařízení.

Představil jsem celkem 6 metod pro trochu rozdílnou klasifikaci samohlásek. Metoda peaků i kořenů ukázala, že samohlásky jsou charakterizovány svými formanty. Podobně metody, které srovnávaly celé spektra samohlásek, ukázaly, že samohlásky se takto dají rozlišit. Pomocí metody peaků jsem ale zjistil, že samohláska U se soustřeďuje do tří různých oblastí. Nabízí se možné vysvětlení, jímž je spojení prvního a druhého formantu. Je možné, že oblast druhého formantu ve spektru je posílena, ale ne natolik, aby došlo k vytvoření lokálního maxima.

Dále se ukázalo, že existuje velká podobnost mezi samohláskami O a U, a to zejména co se týče průběhu jejich spekter. Také první a druhé formanty jsou si nedaleko od sebe, a proto není jednoduché samohlásky O a U rozlišit. Zjistil jsem také, že nejjednodušší je rozpoznat samohlásku I, a to díky velice specifickému průběhu spektra i umístění formantů. Samohlásky A, E jde také pomocí všech metod jednoduše rozlišit.

8 CD příloha

Na přiloženém CD najdete tři složky - M-file, Bakalářská práce a Obrázky.

- M-file - v této složce jsou obsaženy všechny m-fily, které byly v této práci použity. V každém m-filu je uveden popis toho, co dělá a jaké jsou vstupy a výstupy. Dále v této složce najdeme složku Data a soubor readme.txt
  - Data - v této složce jsou uloženy všechny nahrané samohlásky, které byly použity k analýze v této práci
  - Readme - zde jsou popsány konkrétní postupy toho, které m-fily byly v které části použity a pod jakými názvy byly výstupy ukládány
- Bakalářská práce - obsahuje tuto práci ve formátu PDF
- Obrázky - všechny obrázky použité v této práci ve formátu .eps

9 Seznam obrázků

Literatura

[1] Benson, D.: Music: A Mathematical Offering, Department of Mathematical Sciences, University of Aberdeen, Scotland, UK, 2006

[2] Rabiner L., Schafer R.: Introduction to Digital Speech Processing, Foundations and Trends® in Signal Processing, DOI: 10.1561/2000000001

[3] Sovka, P., Pollák P.: Vybrané metody číslicového zpracování signálů, Vyd. 2, Praha: ČVUT, 2003, 258 s, ISBN 80-010-2821-6

[4] Sigmund, M., Rozpoznávání řečových signálů: přednášky, 1. vyd, Brno: VUT FEKT, ústav radioelektroniky, 2007, 122 s. ISBN 978-80-214-3526-1.

[5] Fourierova transformace: http://cs.wikipedia.org/wiki/Fourierova_transformace, [citováno 18. 1. 2013]

[6] Aliasing [online], dostupné z: http://cs.wikipedia.org/wiki/Aliasing, [citováno 18. 1. 2013]

[7] Formanty [online], dostupné z: http://cs.wikipedia.org/wiki/Formant, [citováno 20. 2. 2013]

[8] Znělost [online], dostupné z: http://en.wikipedia.org/wiki/Voice_(phonetics), [citováno 20. 2. 2013]

[9] Černocký, H.: Zpracování řečových signálů - Studijní opora, SPEECH@FIT, Ústav počítačové grafiky a multimédií, Fakulta informačních technologií, VUT v Brně.

[10] Anatomie ucha [online], dostupné z: http://skolajecna.cz/biologie/Sources/Photogallery_Detail.php?intSource=1&intImageId=277

[11] Zvuk [online], dostupné z: http://cs.wikipedia.org/wiki/Zvuk [citováno 23. 2. 2013]

[12] Uhlíř J., Sovka P., Čmejla R., Úvod do číslicového zpracování signálu, Vydavatelství ČVUT, 2003

[13] Kafentzis G., Stylianou Y., Alku P.: Glottal inverse filtering using stabilised weighted linear prediction, Department of Signal Processing and Acoustics, Aalto University, Helsinki, Finland

[14] Magi C., a kol.: Stabilised Weighted Linear Prediction. Speech Communication, 51:401–411, 2009

[15] Walker J., Murphy P.: A Review of Glottal Waveform Analysis, Department of Electronic and Computer Engineering, University of Limerick, Limerick, Ireland