

Spektrálna skladba segmentov rečového signálu

Július Zimmermann, Filozofická fakulta PU, zimmer@unipo.sk

Kľúčové slová: rečový signál, Fourier, spektrum, spektrálne čiary, váhová funkcia.

Keywords: speech signal, Fourier, spectrum, spectral bin, weighting function.

1. Úvod

V štúdiu sa budeme zaoberať určením spektrálnej skladby rečového signálu. Akusticko-auditívna komunikácia prebieha prenosom akustickej energie, ktorá nie je stála, ale mení sa v čase tak, aby reprezentovala informáciu. Teda informácia sa premieňa na zmenu fyzikálnej veličiny v čase. Najjednoduchšou zmenou – signálom v tomto zmysle je periodický sínusový, resp. kosínusový signál. Zložený signál sa dá analyzovať jeho transformáciou na uvedené najjednoduchšie – harmonické signály, čím sa z originálu – pôvodného signálu získa jeho obraz – spektrum.

V spektrálnej skladbe rečového signálu možno nájsť formanty, kľúčové parametre samohlások a sonórnych spoluhlások. Pozorovaním kontinuálnych zmien spektra súvislej reči sa dajú detegovať celkové akustické vlastnosti signálu, určovať hranice hlások a do istej miery poznať aj moduláciu hlasu v nadsegmentálnej rovine.

Prax ukázala, že mnoho problémov v experimentálnej fonetike je spôsobených neznalosťou princípov, na báze ktorých sa konštruuje spektrum. Aby výklad týchto princípov nestratil exaktnosť, výpočet a grafické zobrazenie spektra budeme vysvetľovať pomocou matematických konštrukcií. Pretože matematický výklad pre spoločenskovedné odbory nie je jednoduchý, oprelí sme sa o overené metodické postupy autorov, uvedených v zozname literatúry, predovšetkým o dielo J. Prchala, F. Kroutla, V. Čížeka, J. Chmúrneho a F. Žideka.

2. Spektrum spojitých signálov

Francúzsky matematik Joseph J. B. Fourier (Jean Baptiste Joseph Fourier, 1768–1830) publikoval v r. 1822 svoje dielo *Analytická teória tepla* (*Théorie analytique de la chaleur*). Túto teóriu neskoršie prepracoval Dirichlet, Riemann, Lebesgue a iní a poznáme ju ako Fourierov rad, resp. Fourierov integrál. Podľa nej každý periodický signál $x(t)$ s periódou T môže byť rozvinutý na jednoduché harmonické signály. Fourierov rad je súčet jednosmernej zložky a nekonečného radu jednoduchých sínusových a kosínusových priebehov, ktorých frekvencie sú dané celistvými násobkami základnej frekvencie s periódou T . Signál $x(t)$ môže byť nahradený trigonometrickým tvarom Fourierovho radu:

$$x(t) = A_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{2\pi n t}{T} + b_n \sin \frac{2\pi n t}{T} \right) \quad \text{pre } n = 1, 2, 3, \dots$$

A_0 je jednosmerná zložka (stredná hodnota signálu $x(t)$). Parametre a_n , b_n sú Fourierove koeficienty. Po výpočte koeficientov a_n , b_n možno vytvoriť reálne amplitúdové spektrum, graf, zobrazujúci inventár harmonických zložiek, daných svojimi frekvenciami a amplitúdami.

Fourierov rad môže byť vyjadrený okrem trigonometrického tvaru aj v exponenciálnom tvare použitím komplexných čísel. V zložitejších výpočtoch je exponenciálny tvar vhodnejší a kompaktnější, než trigonometrický tvar. Periodickú funkciu $x(t)$, ktorá reprezentuje časové zmeny signálu, možno rozvinúť na intervale $t_0 \leq t \leq t_0 + T_0$ do Fourierovho radu v exponenciálnom tvare:

$$x(t) = \sum_{n=-\infty}^{\infty} C_n \cdot e^{jn\omega_0 t}$$

pričom komplexné Fourierove koeficienty sú:

$$C_n = \frac{1}{T_0} \int_{t_0}^{t_0+T_0} x(t) \cdot e^{-jn\omega_0 t} dt \quad \text{pre } n = 0, \pm 1, \pm 2, \pm 3, \dots$$

kde ω_0 je uhlová frekvencia základnej harmonickej zložky signálu $x(t)$. V zmysle rozkladu periodického signálu na jeho harmonické zložky je koeficient C_n vektorom amplitúdy n -tej harmonickej zložky, $|C_n|$ má význam jej fyzikálnej amplitúdy. Jednosmerná zložka $A_0 = C_0$. Dvojstranné spektrum periodických signálov je diskrétné (čiarové), skladá sa z harmonických zložiek. Najnižšia harmonická zložka C_1 má uhlovú frekvenciu ω_0 , ostatné (vyššie) harmonické zložky sa nachádzajú na uhlových frekvenciách $n\omega_0$, resp. $-n\omega_0$.

Dva posledné vzťahy vyjadrujú diskrétné spektrum periodického signálu, v ktorom amplitúdy harmonických zložiek nadobúdajú nenulové hodnoty iba v celočíselných násobkoch základnej frekvencie signálu. Vyriešime ešte otázku, aké vlastnosti bude mať spektrum neperiodického signálu.

Periodický signál sa zmení na neperiodický, ak periódu T_0 budeme zväčšovať do nekonečna. Z toho vyplýva, že $\Delta\omega$ medzi susednými harmonickými frekvenciami sa bude blížiť k nule, diskrétné frekvencie $n\omega_0$ sa zmenia na spojitú frekvenciu ω a diskrétné spektrum sa zmení na spojitú. Dve posledné rovnice vyjadrujúce vzťah medzi originálom a obrazom nadobudnú tvar:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega t} dt$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \cdot e^{j\omega t} d\omega$$

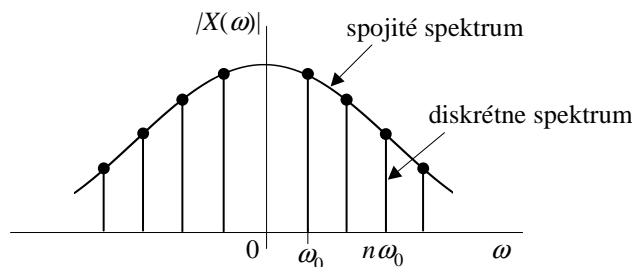
Funkciu $X(\omega)$ skrátene nazývame spektrom, Fourierovým obrazom časovej funkcie $x(t)$. Je známa ako Fourierova transformácia signálu z časovej oblasti t do spojitaj frekvenčnej oblasti ω . Funkciu $x(t)$ nazývame vzorom, (niekedy tiež originálom), je to Fourierov integrál, ktorý je spätnou Fourierovou transformáciou, vyjadruje signál nekonečným súčtom harmonických zložiek v intervale $\omega \in (-\infty, \infty)$.

Funkcia $X(\omega)$ je komplexnou funkciou reálnej premennej ω

$$X(\omega) = |X(\omega)| \cdot e^{j\varphi(\omega)}$$

Modul $|X(\omega)|$ nevyjadruje amplitúdu jednotlivých harmonických zložiek, ale iba spektrálnu hustotu amplitúd (niekedy skrátene amplitúdové spektrum) neperiodického signálu. Rozmer spektrálnej hustoty amplitúd je [amplitúda/Hz]. Spektrum neperiodického signálu je spojitú.

Modul spektrálnej funkcie $|X(\omega)|$ neperiodického signálu – spojitú spektrum, je rovnicou obalovej krivky amplitúdového spektra periodického signálu – obalovou krivkou diskrétného spektra (obr. 1).



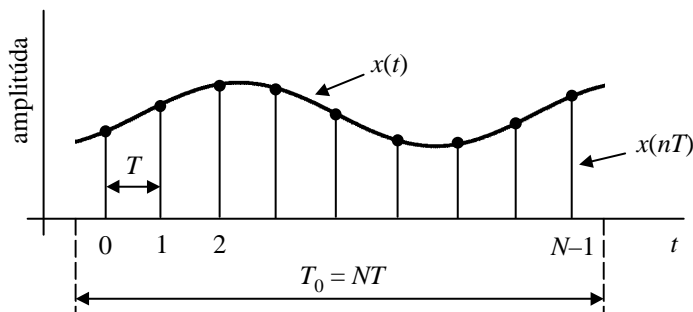
Obr. 1. Spektrum neperiodického a periodického signálu.

3. Diskrétna Fourierova transformácia

Diskrétna Fourierova transformácia (DFT) slúži na konštrukciu spektra digitalizovaných rečových signálov, pozostávajúcich zo sekvencie diskretných vzoriek. Proces DFT pozostáva z nasledovných krokov:

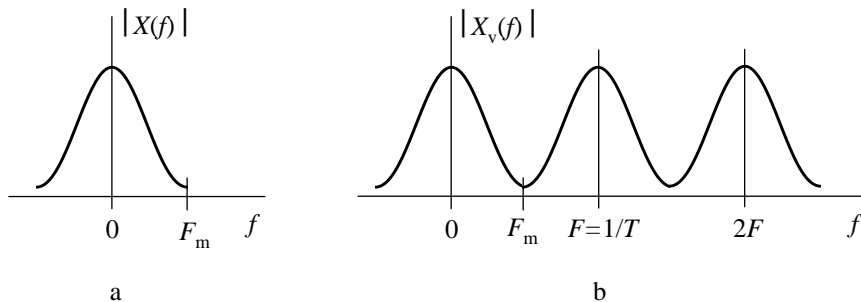
1) Spojitý signál $x(t)$ sa najprv vzorkuje vzorkovacou frekvenciou F , teda nahradí časovou postupnosťou vzoriek $x(nT)$ v diskretných periodických okamihoch $t = nT$, $n = 0, 1, 2, \dots$; T je vzorkovacia perióda. Vznikne tak diskretný signál.

2) Vzorkovaný signál sa obmedzí časovým intervalom T_0 tak, že sa vynásobí časovo obmedzenou funkciou, napr. pravouhlým impulzom. Počet vzoriek na intervale T_0 bude konečný, poradové číslo vzoriek $n = 0, 1, 2, \dots, N-1$ (obr. 2).



Obr. 2. Vzorkovaný, časovo obmedzený signál.

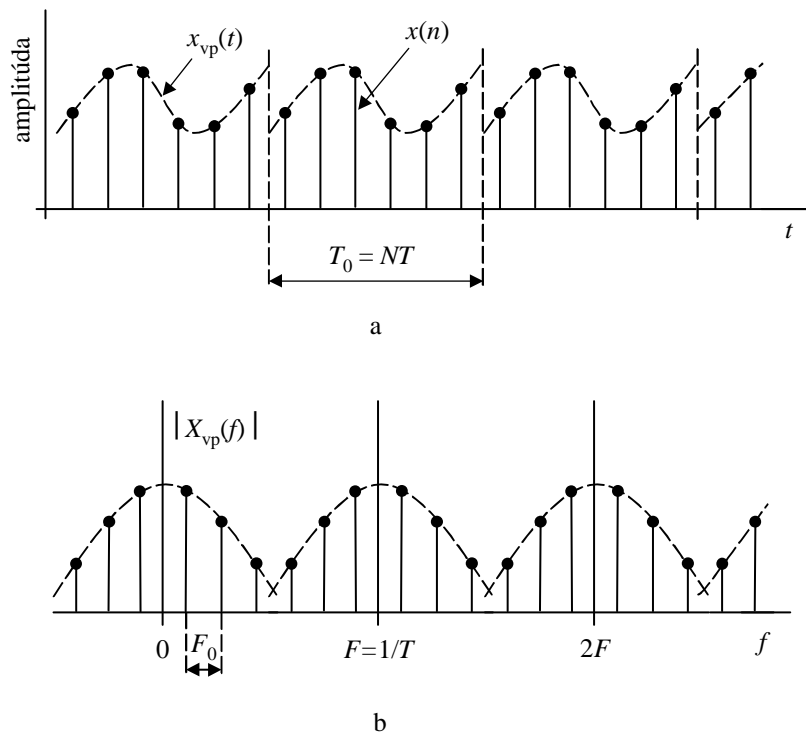
Pretože diskretný signál, získaný vzorkovaním spojitého signálu $x(t)$ ešte nie je periodický, jeho spektrum $X_v(f)$ nie je diskrétné, ale spojité a periodické s periódou $F = 1/T$ (obr. 3). Obmedzenie signálu $x_v(t)$ pravouhlým impulzom naruší pôvodný priebeh spojitého spektra, tvar periódy $X_v(f)$ nie je zhodný s tvarom spektra $X(f)$. Pre maximálnu frekvenčnú zložku F_m signálu $x(t)$ na základe vzorkovacej teóremy platí: $F_m \leq F_{vz}/2$.



Obr. 3. a – spektrum spojitého signálu $x(t)$, b – spektrum vzorkovaného signálu $x_v(t)$.

3) Aby mohol byť výsledok DFT – spektrum analyzovaného signálu $x(t)$ – spracovaný na počítači, musí mať diskretnú formu. Periodicitu neperiodického signálu zaistíme tak, že jeho segment pokladáme za jednu periódu, ktorú opakujeme, čím získame príznak periodicity.

V treťom kroku sa zo segmentu vzorkovaného signálu $x_v(t)$ vytvorí periodický signál $x_{vp}(t)$ s periódou $T_0 = NT$. Jeho spektrum $X_{vp}(f)$ bude diskrétné; obalová krivka tohto spektra bude spojité spektrum $X(f)$ (obr. 4).



Obr. 4. a – vzorkovaný periodický signál, b – spektrum vzorkovaného periodického signálu.

4) Realizuje sa diskretná Fourierova transformácia podľa vzťahu:

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi nk}{N}}$$

kde: $k = 0, 1, 2, \dots, N-1$ je poradové číslo spektrálnej čiary,
 $n = 0, 1, 2, \dots, N-1$ je poradové číslo vzorky.

$X(k)$ je k -ta spektrálna čiara, $x(n)$ je n -tá vzorka signálu $x(t)$. N vzoriek $x(n)$ v časovej doméne, ktoré sa vyskytujú v periodických okamihoch nT , je transformovaných na N spektrálnych čiar $X(k)$ vo frekvenčnej doméne. Vzďalenessi medzi spektrálnymi čiarami sú dané periódou $F_0 = 1/T_0$. N spektrálnych čiar sa opakuje s periódou $F = 1/T = NF_0$. Vstupnými hodnotami DFT sú hodnoty vzoriek $x(n)$ a ich počet N .

Postupnosť $X(k)$ tvorí diskretné spektrum signálu. Táto postupnosť je postupnosťou komplexných čísel, teda postupnosťou dvojíc čísel; prvé číslo je reálnou, druhé imaginárnou zložkou komplexného čísla. Diskretné amplitúdové dvojstranné spektrum získame z týchto dvojíc výpočtom podľa vzťahu:

$$|X(k)| = \sqrt{\text{Re}^2[X(k)] + \text{Im}^2[X(k)]}$$

kde $\text{Re}[X(k)]$ je reálna, $\text{Im}[X(k)]$ je imaginárna zložka komplexného čísla. Amplitúda dvojstranného spektra $|X(k)|$ má podobný význam ako amplitúda $|C_n|$.

Frekvenčný odstup medzi spektrálnymi čiarami výstupného spektra DFT udáva hustotu spektrálnych čiar. Hustejšie spektrálne čiary zvyšujú frekvenčné rozlíšenie. Frekvenčný odstup je určený výrazom:

$$\Delta\omega = \frac{2\pi}{N} \Rightarrow \Delta f = \frac{F}{N}$$

Pre frekvenciu spektrálnej čiary – spektrálnej zložky platí:

$$W_k = k \frac{F}{N}$$

Výsledkom procesu DFT sú číselné hodnoty amplitúd spektrálnych čiar, z ktorých sa kreslí v rôznych úpravách graf spektra vstupného signálu. Frekvenčné rozlíšenie je dané hustotou spektrálnych čiar v rozpätí od jednosmernej zložky až po Nyquistovu frekvenciu. DFT rozdelí frekvenčnú škálu od jednosmernej zložky až po vzorkovaciu frekvenciu F na N spektrálnych čiar, pričom je pre výsledné spektrum užitočná len ich prvá polovica, $N/2$.

Algoritmus DFT spracúva konečný počet N vzoriek signálu, teda z kontinuálneho signálu vyrezáva segment $T_0 = NT$. Ostré odrezanie signálu môžeme vyjadriť v časovej oblasti jeho násobením obdĺžnikovým oknom – váhovou funkciou $w(n)$:

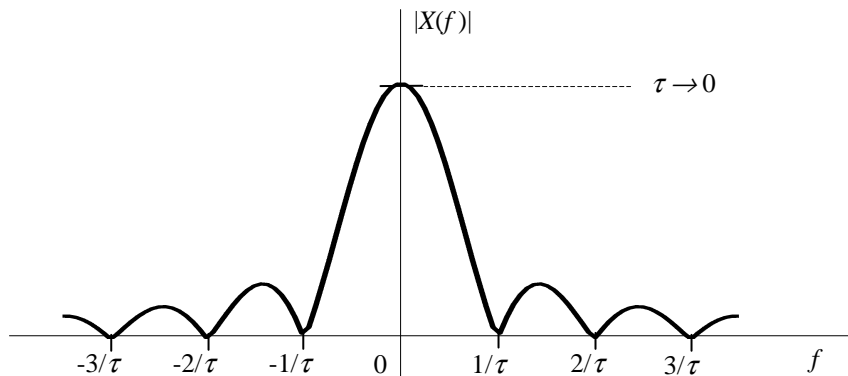
$$\begin{aligned} w(n) &= 1 && \text{pre } 0 \leq n \leq N - 1 \\ w(n) &= 0 && \text{pre } 0 > n > N - 1 \end{aligned}$$

Z toho vyplýva, že segment signálu $x_w(n)$ určený N vzorkami bude:

$$x_w(n) = x(n) \cdot w(n)$$

Z tohto vzťahu vyplýva, že spektrum signálu je ovplyvnené spektrom váhového okna. Spektrálnym obrazom súčiny signálov $x(n)$ a $w(n)$ je konvolúcia ich spektier $X(k)$ a $W(k)$. Analogicky možno odvodiť, že súčin spektier dvoch signálov je konvolúciou týchto signálov.

Časový priebeh obdĺžnikového okna je totožný s pravouhlým signálom. Dvojstranné amplitúdové spektrum pravouhlého impulzu znázorňuje obr. 5:



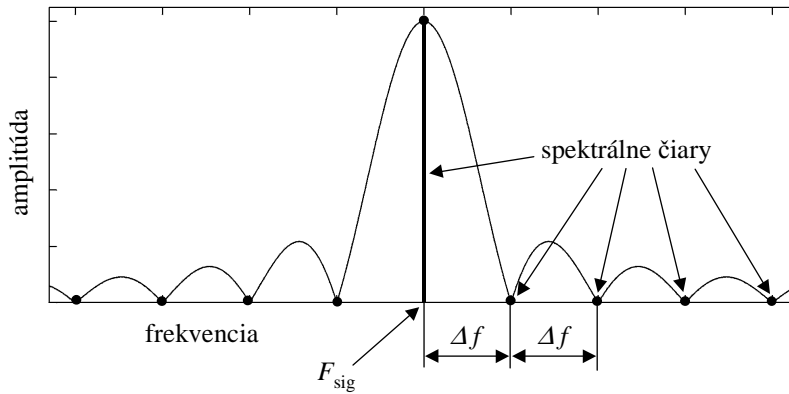
Obr. 5. Dvojstranné spektrum osamelého pravouhlého impulzu šírky τ .

Výsledným produktom diskretnej Fourierovej transformácie je graf konvolúcie spektra pravouhlého okna $W(k)$ a spektra signálu $X(k)$, je čiarovým spektrom s obalovou krivkou zhodnou s krivkou na obr. 5. Tento graf môže nadobudnúť dve podoby:

a) Ak je splnená podmienka:

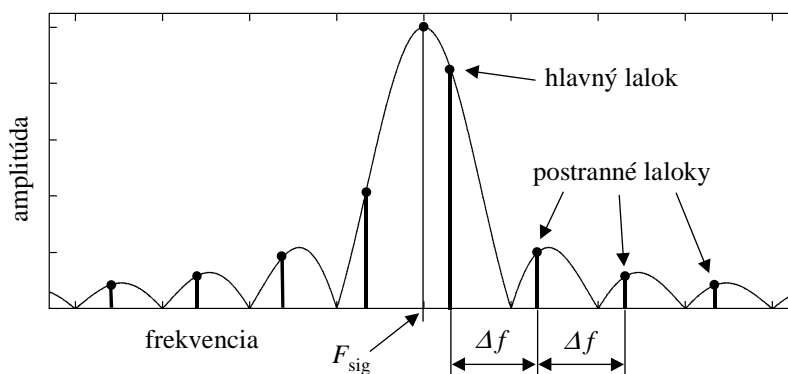
$$\frac{F_{\text{sig}}}{F_{\text{vz}}} N = m \quad \text{kde } m \text{ je prirodzené číslo } (1, 2, 3, \dots).$$

nedôjde k presakovaniu spektra; os symetrie grafu bude pretínať frekvenčnú os na frekvencii signálu, amplitúdy susedných spektrálnych čiar budú nulové (obr. 6). Vzdialenosť spektrálnych čiar $\Delta f = F/N$, spektrálne čiary sú umiestnené na frekvenciách F_{sig} , $(F_{\text{sig}} \pm 1)F/N$, $(F_{\text{sig}} \pm 2)F/N$, $(F_{\text{sig}} \pm 3)F/N$, atď.



Obr. 6. DFT sínusového signálu, keď nedochádza k presakovaniu spektra.

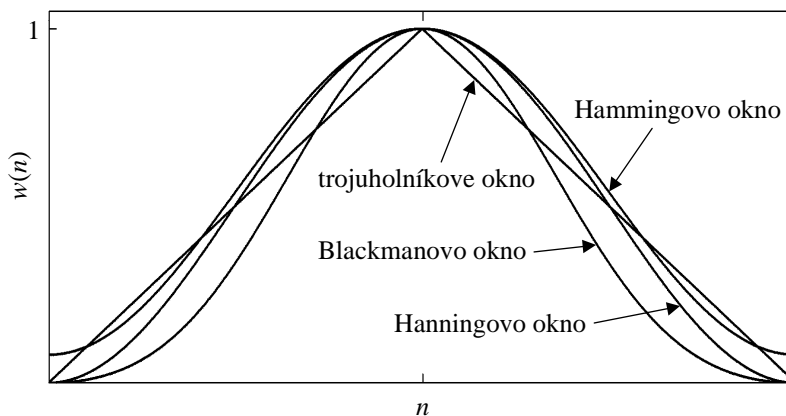
b) Ak táto podmienka nie je splnená, vtedy sa obalová krivka posunie po frekvenčnej osi tak, aby sa jej najvyšší bod nachádzal nad frekvenciou signálu. V tomto prípade F_{sig} sa nezhoduje s frekvenciou žiadnej spektrálnej čiary, ani susedné spektrálne čiary nebudú ležať na nulových bodoch obalovej krivky spektra pravouhlého impulzu. V okolí frekvencie signálu sa objavia susedné spektrálne čiary s nenulovou amplitúdou, dôjde k presakovaniu spektra (obr. 7). Amplitúdy spektrálnych čiar budú určované obalovou krivkou spektra, jej hlavným a vedľajšími oblúkmi – hlavným lalokom a postrannými lalokmi. Príčinou vzniku lalokov vo frekvenčnej oblasti je Gibbsov jav, následok strmých prechodov na okrajoch pravouhlého okna. Amplitúda lalokov určuje, v akej miere presakuje spektrum do postranných frekvencií, ktoré sa v signáli nenachádzajú. Najväčšie presakovanie sa prejaví vtedy, keď sa stred lalokov bude nachádzať v strede medzi susednými spektrálnymi čiarami. Tu je podstata vzťahu medzi diskontinuitou signálu, zapríčinenou ostrými hranami časového okna a presakovaním spektra, graficky znázorneným postrannými lalokmi.



Obr. 7. DFT sínusového signálu, keď dochádza k presakovaniu spektra.

Zmenšením výšky postranných lalokov možno minimalizovať presakovanie spektra do susedných spektrálnych čiar. Tento efekt sa dá dosiahnuť vyhladením ostrých okrajových hrán pravouhlého časového okna, teda iným, než pravouhlým oknom, vhodnou matematicky definovanou váhovou funkciou. Zmenou tvaru okna sa teda dosahuje zníženie postranných lalokov (potlačenie presakovania), vedľajším účinkom je rozšírenie hlavného laloku (zhoršenie frekvenčného rozlíšenia). Voľba tvaru okna je kompromisom medzi potlačením presakovania a zhoršením frekvenčného rozlíšenia. Inými slovami, široký hlavný lalok, zapríčínujúci nepriaznivé frekvenčné rozlíšenie je spojený s malou amplitúdou postranných lalokov, pri ktorých sa presakovanie znižuje. Opačne, úzky hlavný lalok, umožňujúci presnejšie odčítať frekvenciu signálu, je spojený s väčšou amplitúdou postranných lalokov, keď je presakovanie spektra väčšie.

Pre analýzu rečového signálu sú vhodné predovšetkým okná Hanningovo, Hammingovo, Blackmanovo a trojuholníkové (obr. 8). Okná Hanningovo, Hammingovo a Blackmanovo rozširujú hlavný lalok a znižujú amplitúdu postranných lalokov.



Obr. 8. Grafy váhových funkcií.

Efektívnym spôsobom výpočtu diskretnej Fourierovej transformácie je rýchla Fourierova transformácia (Fast Fourier Transform – FFT). Podstatou jej algoritmu je zníženie počtu potrebných komplexných násobení z hodnoty N^2 na hodnotu $M \log_2(N)$ predovšetkým preto, aby sa mohla analýza realizovať v reálnom čase. V praxi jestvuje viacero algoritmov, urýchľujúcich výpočet DFT. Pri analýze rečových signálov sa používajú postupy vychádzajúce z podmienky $N = 2^m$, m je prirodzené číslo. Jeden z týchto algoritmov je algoritmus decimácie v čase. Iným postupom, urýchľujúcim výpočet DFT, je algoritmus decimácie vo frekvencii.

4. Aplikačné poznámky o využití DFT a FFT pri analýze rečového signálu

Výpočet diskretného spektra rýchlou Fourierovou transformáciou je spracovaním časového okna, ktoré obsahuje N vzoriek digitalizovaného signálu. Keď predpokladáme, že vzorkovanie reči prebieha s frekvenciou 10–40 kHz a hodnota $N = 2^m$, pričom m je prirodzené číslo z intervalu $\langle 5, 13 \rangle$, potom FFT môže produkovať jeden spektrálny obraz určeného časového segmentu signálu v rozsahu od jednej milisekundy do stoviek milisekúnd. Z úvahy vyplýva, že jednému výpočtu FFT by sme mohli podrobiť zlomok hlasivkového kmitu, v krajnom prípade aj niekoľko slabík; z tohto rozsahu sa s ohľadom na reč môžeme pohybovať v rozpätí od jedného hlasivkového kmitu po celú hlásku. Analýza rozsiahlejšieho zvuku si vyžiada opakovaný výpočet, spojený signál sa transformuje do série svojich spektrálnych rezov, upravených na dlhodobé priemerné spektrum, alebo na sonagram.

Diskrétné amplitúdové jednostranné spektrum zobrazuje inventár harmonických zložiek signálu; amplitúdy spektrálnych čiar sú ich reálnymi amplitúdami A_n . Pretože vždy sa prejaví presakovanie amplitúd zložiek signálu do susedných spektrálnych čiar, frekvenčné rozlíšenie sa zvyšuje ich väčšou hustotou. Informáciu o amplitúde spektrálnej čiary obsahuje výška spektrálnej čiary – y -ová súradnica jej vrcholu; x -ová súradnica spektrálnej čiary určuje frekvenciu zložky. Z toho vyplýva, že harmonická zložka je definovaná len súradnicami bodu, ktorý je vrcholom spektrálnej čiary. Z tohto dôvodu sa spektrálne čiar v grafoch spektra nezobrazujú, kreslia sa len ich vrcholy pospájané do kontinuálnej lomenej čiary. Táto spojnica pri vyššom počte spektrálnych čiar nadobudne podobu hladkej krivky, napodobňujúcej kontúru spojeného spektra.

Aby graf spektra náležite ukázal, aké vlastnosti má spektrum, obidve osi grafu sa modifikujú. Frekvenčná os sa upravuje tak, aby zobrazovala normalizovanú Nyquistovu

frekvenciu. V tomto prípade vodorovná os zobrazí frekvencie v rozsahu od nuly až po polovicu vzorkovacej frekvencie (Nyquistovu frekvenciu), pričom je Nyquistova frekvencia pokladaná za celok, rovný jednotke. Inou úpravou frekvenčnej osi je jej premena na melovu stupnicu, keď sa berú do úvahy vlastnosti vnemu sluchového orgánu z hľadiska jeho nelinearity vzťahu medzi fyzikálne definovanou frekvenciou a vnímanou tónovou výškou.

Druhá, y -ová os pôvodného amplitúdového spektra rečového signálu sa zmení tak, aby ukazovala hodnotu výkonového spektra v decibelovej stupnici. Výkonové spektrum, štatistická charakteristika náhodných javov, predstavuje rozloženie výkonu analyzovaného signálu na jednotlivé zložky amplitúdového spektra. Pretože rečový signál má charakter náhodného procesu, štatistické metódy sú účinným prostriedkom na jeho rozbor.

Stredný výkon periodického signálu vyjadreného v exponenciálnom tvare bude:

$$P = C_0^2 + \sum_{n=1}^{\infty} 2|C_n|^2$$

kde C_n je komplexná amplitúda n -tej harmonickej zložky dvojstranného spektra, C_0 je jednosmerná zložka signálu. Keď berieme do úvahy, že reálna amplitúda A_n zložky jednostranného spektra je dvojnásobkom absolútnej hodnoty komplexnej amplitúdy C_n zložky dvojstranného spektra, teda:

$$A_n = 2|C_n|$$

potom pre stredný výkon ľubovoľného periodického signálu platí:

$$P = A_0^2 + \sum_{n=1}^{\infty} 2\left(\frac{A_n}{2}\right)^2 = A_0^2 + \sum_{n=1}^{\infty} \frac{1}{2} A_n^2$$

Pretože $A_0 = C_0$, jednosmerné zložky jednostranného a dvojstranného spektra sú zhodné. Stredný výkon spektrálnej zložky je:

$$2|C_n|^2 = \frac{1}{2} A_n^2.$$

Krátkodobá Fourierova analýza je rýchla Fourierova transformácia aplikovaná na krátke časové okno N vzoriek, produkujúca spektrálne čiary – harmonické zložky vstupného rečového signálu, pre ktoré sa vypočítajú hodnoty stredného výkonu. Toto spektrum, skrátene uvádzané ako výkonové spektrum, je množinou reálnych hodnôt stredných výkonov zložiek. Vyhodnocovať súbor týchto hodnôt je nepraktické. Zvyčajne sa zobrazujú dvojrozmerným grafom, v ktorom nezávisle premennou (x -ovou osou) je frekvencia (frekvenčná oblasť – doména), závisle premennou (y -ovou osou) je stredný výkon. Ak vrcholy spektrálnych čiar, ktorých amplitúdy sú prepočítané na stredné výkony, spojíme obalovou krivkou, získame graf výkonového spektra – periodogram. Vo výkonovom spektre má väčšina spektrálnych čiar malú amplitúdu. Dominantné frekvenčné zložky, ukazujúce na akustické ohniská rečového signálu, sa v tomto spektre vykresľujú ako výrazné špičky. Pre zväčšenie dynamického rozsahu stredného výkonu, teda vertikálnej osi grafu, sa hodnoty stredného výkonu transformujú na úroveň L v decibeloch podľa vzťahu:

$$L_{dB} = 10 \log_{10} \frac{P}{P_r}$$

kde P je stredný výkon spektrálnej čiary a P_r je vzťažný, referenčný výkon. Takto transformovaný stredný výkon má logaritmický priebeh, čím sa zobrazovanie zložiek signálu s nízkymi hodnotami stredného výkonu zosilní.

Graf výkonového spektra môže nadobudnúť rôzne podoby. Optimalizuje sa tak, aby ukázal tie charakteristiky, ktoré sú zaujímavé v analyzovanom segmente rečového signálu. Z toho, čo sme doteraz uviedli o algoritme výpočtu FFT vyplýva, že s podobou spektra sa dá experimentovať voľbou počtu vzoriek N , vstupujúcich do algoritmu FFT a voľbou váhovej

funkcie. K týmto voliteľným parametrom treba pridať ešte stupeň vyhladenia krivky spektra, preemfázu a rozsahy x -ovej a y -ovej osi.

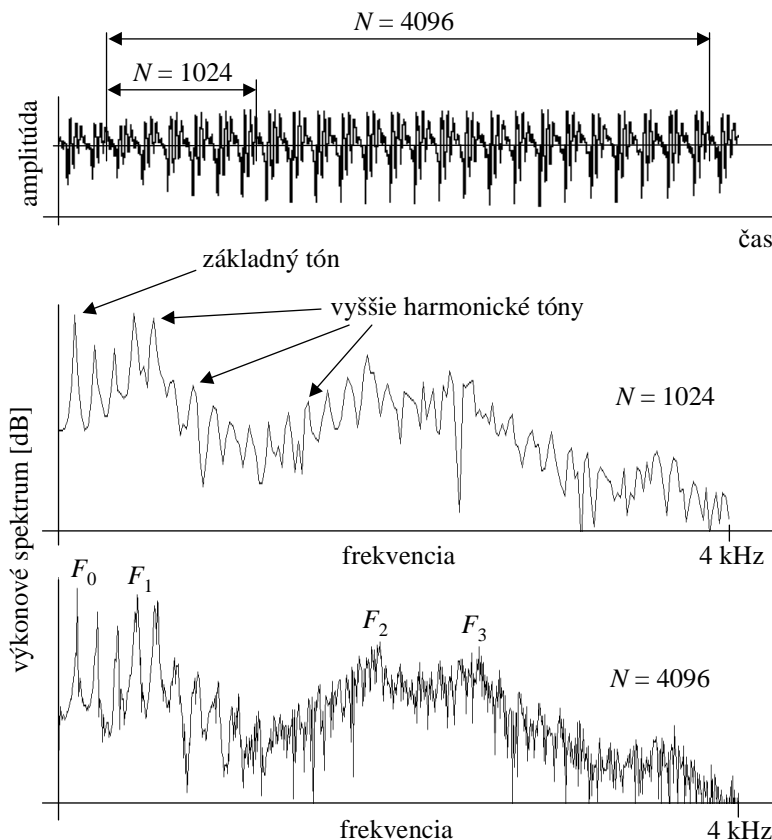
Prvým krokom pred spustením algoritmu FFT je výber segmentu rečového signálu, ktorý má byť analyzovaný. Periodická vlna sonantického jadra slabiky má odlišné spektrum a trvanie v porovnaní so zvukom frikatív alebo explozív, teda východiskom pri nastavovaní parametrov účinného výpočtu FFT budú akustické vlastnosti hlások. Segment je jednoznačne určený dĺžkou okna váhovej funkcie a polohou okna na časovej osi.

Dĺžka okna je segment vstupného rečového signálu, daný počtom vzoriek N , ktorý má byť spracovaný algoritmom FFT. Dĺžka okna určuje dĺžku váhovej funkcie v časovej doméne. Z hodnoty N vyplýva počet spektrálnych čiar medzi nulovou a vzorkovacou frekvenciou, resp. počet Fourierových koeficientov. Dĺžka okna $N = 2^m$, m je prirodzené číslo. Vhodnými hodnotami N sú čísla 32, 64, 128, 256, 512, 1024, 2048 a 8192. S voľbou N súvisí aj určenie polohy okna na časovej osi signálu. Poloha okna sa volí na oscilograme určením vzorky (časového okamihu), ku ktorej sa pridá ďalších $N - 1$ nasledujúcich vzoriek.

Vyššie hodnoty N určujú väčší počet spektrálnych čiar, z čoho vyplýva vyššie frekvenčné rozlíšenie vo výslednom spektre. Na druhej strane, rečový signál môže v niektorých úsekoch rýchle meniť svoju frekvenčnú skladbu; tieto zmeny možno zachytiť postupnosťou krátkodobých FFT s menšou dĺžkou okien. Z povedaného vyplýva, že voľba dĺžky okna musí byť kompromisom medzi frekvenčným rozlíšením a rozlíšením časových zmien spektra analyzovaného signálu. Na dĺžku okna vplyva aj vzorkovacia frekvencia, jej voľba súvisí s maximálnou frekvenčnou zložkou, ktorú hodláme vzorkovaním zachytiť a analyzovať.

Príkladom využitia krátkodobej FFT pri analýze spektra dlhej samohlásky [e:] je obr. 9. Prvý graf zhora ukazuje oscilogram jadra samohlásky, v grafe sú označené dva segmenty, začínajúce v rovnakom okamihu. Pri vzorkovacej frekvencii 20 kHz dĺžka kratšieho segmentu je 1024, dlhšieho 4096 vzoriek. Obidva segmenty obsahujú stacionárne, kvaziperiodické vlny, vyňatie segmentov z kontinuálneho signálu bolo uskutočnené pravouhlým oknom.

Druhý a tretí graf obr. 9 sú grafy výkonových spektier, vypočítaných pre rozdielne hodnoty N . V prostrednom grafe vidíme hlavné laloky harmonických tónov, postranné laloky susedných harmonických zložiek sa vzájomne prekrývajú. Hlavný lalok s najnižšou frekvenciou je základný tón F_0 , ostatné laloky, vyššie harmonické zložky, sú celistvými násobkami jeho frekvencie. Dominantné harmonické tóny s prevládajúcimi lalokmi stredného výkonu sú formanty, produkty výrazných rezonancií artikulačných dutín pri generovaní vokálov a sonórnych konsonantov. Základný tón je nulovým formantom F_0 , jeho zmeny vnímame ako melódiu rečového signálu, patrí do skupiny suprasegmentálnych vlastností reči. Vyššie formanty F_1 , F_2 , F_3 , atď., charakterizujú segment, sonantickú hlásku, ich konfigurácia na frekvenčnej osi definuje konkrétnu hlásku reči (formant F_1 , F_2 , niekedy aj F_3), resp. farbu hlasu (formant F_3 a vyššie formanty).



Obr. 9. Oscilogram jadra dlhej samohlásky [e:] a výkonové spektrá získané krátkodobou FFT s rozdielnymi hodnotami N , pri vzorkovaní frekvenciou 20 kHz.

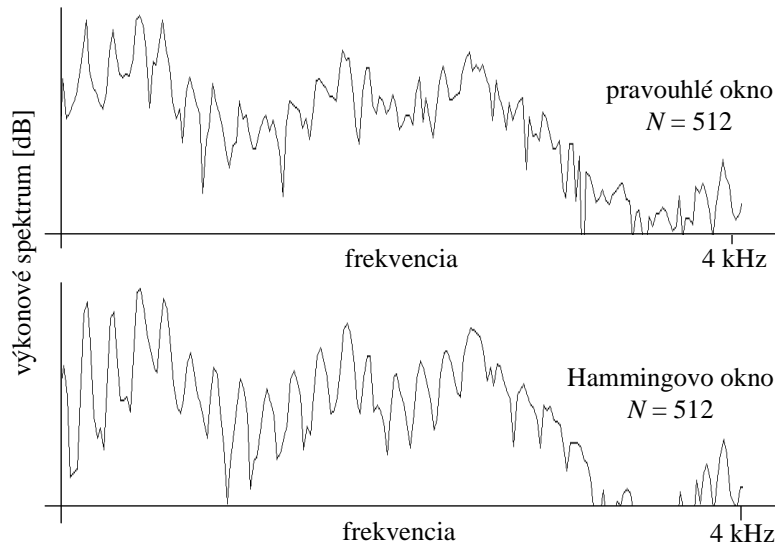
Spodný graf na obr. 9 sa líši od prostredného grafu vyššou hodnotou N , hustota spektrálnych čiar je štvornásobná. Spojnica ich vrcholov je zbytočne „rozkolísaná“, zvýšené frekvenčné rozlíšenie neprospieva identifikácii harmonických tónov. Ak by sme hodnotu $N = 1024$ (pri vzorkovaní 20 kHz) v prostrednom grafe znížili na polovicu, resp. na štvrtinu, kresba hlavných lalokov by bola neostrá, frekvenčné rozlíšenie by nebolo priaznivé.

Ďalej ukážeme, ako sa dá optimalizovať krivka spektra ďalšími parametrami krátkodobej FFT.

Druhým voliteľným parametrom krátkodobej FFT je okno – váhová funkcia, ktorá z kontinuálneho signálu vyreže segment dĺžky N vzoriek. Okná majú odlišné charakteristiky v časovej, resp. frekvenčnej doméne, ich úlohou je znížiť presakovanie spektra, spôsobené diskontinuitou signálu pri vyrezaní segmentu N vzoriek. V strede frekvenčnej odozvy každého okna sa nachádza hlavný lalok, okolo ktorého sa na obidve strany rozkladajú postranné laloky. Užší hlavný lalok zvyšuje frekvenčné rozlíšenie, nižšie postranné laloky potláčajú presakovanie spektra. Voľba okna je kompromisom medzi zvýšením frekvenčného rozlíšenia a potlačením presakovania.

Na obr. 10 uvádzame príklad krátkodobej analýzy FFT vokálu [e] zo slabiky [se]. Signál bol vzorkovaný frekvenciou 10 kHz. Horný graf ukazuje výkonové spektrum vokálu pri $N = 512$, váhovou funkciou v tomto prípade je pravouhlé okno. V spodnom grafe je výkonové spektrum toho istého segmentu, získaného Hammingovým oknom. Hammingovo okno má širší hlavný lalok než pravouhlé okno a postranné laloky Hammingovho okna sú viac potlačené, než postranné laloky pravouhlého okna. Z dvojice grafov na obr. 10 vyplýva, že Hammingovo okno v porovnaní s pravouhlým oknom viac potláča presakovanie spektra, čo sa prejavuje výraznejšími hlavnými lalokmi harmonických zložiek rečového signálu. Širší

hlavný lalok Hammingovho okna v tejto analýze neznižuje frekvenčné rozlíšenie, hodnoty frekvencií formantov sa dajú odčítať s približne rovnakou presnosťou. Hammingovo okno produkuje hladšie spektrum, než pravouhlé okno. Vyššie postranné laloky pravouhlého okna sa prejavujú ostrejšími zmenami spektrálnej krivky, charakteristickými pre šumové spektrum.

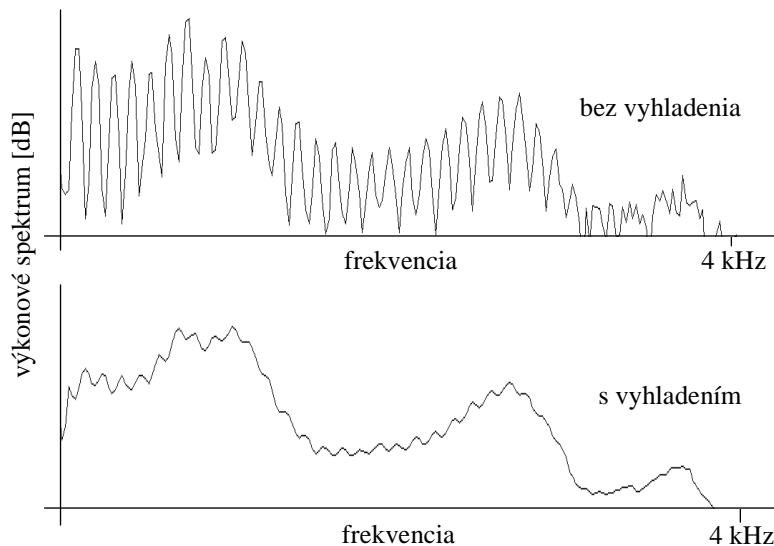


Obr. 10. Výkonové spektrá vokálu [e], zvýraznenie harmonických zložiek v spodnom grafe je dosiahnuté Hammingovým oknom.

Vyhľadanie krivky výkonového spektra potláča jej zvlnenie, spôsobené prítomnosťou harmonických zložiek v signáli. Vyhladená krivka zobrazuje dominantné rezonancie v spektre – frekvencie formantov, vedľajšie akustické aktivity harmonických frekvencií sú skryté v obalovej krivke spektra.

Stupeň vyhladenia určuje, do akej miery má byť potláčanie zvlnenia uplatnené. Udáva sa počtom Fourierových koeficientov, z ktorých sa vypočíta kľzavý priemer. Presnejšie povedané, z postupnosti stredných výkonov spektrálnych čiar (v dB) sa výpočtom kľzavých priemerov n hodnôt vytvorí nová postupnosť, ktorá sa po grafickom zobrazení stáva vyhladenou krivkou spektra. Na stupeň vyhladenia vplýva aj frekvenčné rozlíšenie pôvodného nevyhladeného spektra a výška hlavných lalokov harmonických zložiek signálu.

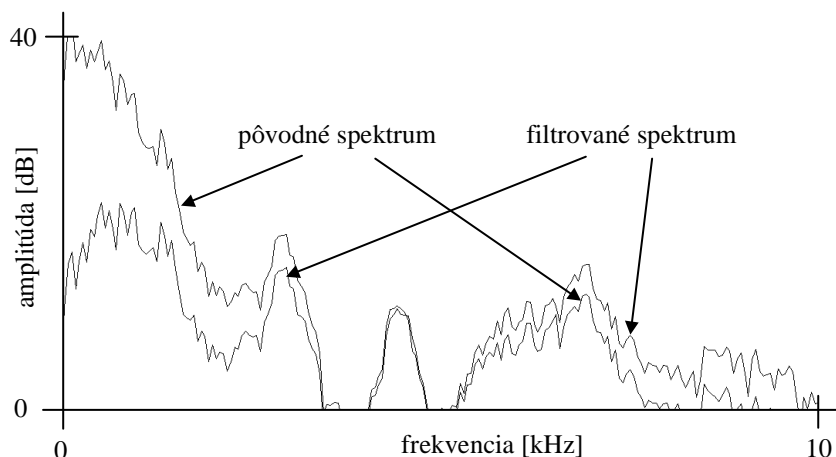
Na obr. 11 sú porovnané výkonové spektrá samohlásky [a] pred vyhladením a po vyhladení. Signál bol vzorkovaný frekvenciou 10 kHz, v oboch analýzach bolo zvolené Hanningovo okno pre $N = 512$.



Obr. 11. Výkonové spektrum vokálu [a], získané Hanningovým oknom pri $N = 512$ a jeho vyhladenie klzavým priemerom z 9 vzoriek.

Preemfáza je vyrovnanie spektra nižších a vyšších frekvenčných zložiek signálu. Akustická energia dominantných harmonických zložiek – formantov je rozložená tak, že klesá smerom k vyšším frekvenciám, z čoho vyplýva, že vyššie formanty sú málo výrazné v krivke spektra. Zdvihnutím úrovně vyšších frekvenčných zložiek vstupného signálu a znížením úrovně nižších frekvencií sa vyrovná spektrálny tvar jeho periodických zložiek, čím sa vyváži pozorovanie nízkych a vysokých frekvencií. Úlohu úpravy spektrálneho tvaru signálu v uvedenom zmysle môže plniť digitálny hornopriepustný filter.

Účinok hornopriepustného filtra pri úprave spektra slabiky [sa] vidíme na obr. 12. Z grafu je zrejmé, ako filtrovanie vyrovnalo pôvodné spektrum slabiky [sa] zosilnením vyšších frekvenčných zložiek a utlmením nízkych tónov. Na stredné frekvencie filtrovanie nemá podstatný vplyv.



Obr. 12. FFT analýza celkového spektra slabiky [sa] pred filtrovaním a po úprave hornopriepustným filtrom.

Rozsah x -ovej osi grafu spektra – periodogramu určuje, ktoré frekvenčné zložky analyzovaného signálu budú zobrazené v spektre. Maximálny rozsah frekvenčnej stupnice,

v ktorom sa môže nachádzať informácia o zložení spektra, leží v rozpätí od nuly po Nyquistovu frekvenciu ($F_{vz}/2$). Ak tento rozsah pokladáme za celok (100%), vtedy je zobrazený rozsah frekvencií definovaný relatívne, percentuálnym údajom minimálnej a maximálnej frekvencie. Keď Nyquistovu frekvenciu nepokladáme za celok, ale je daná svojou absolútnou hodnotou v Hertzoch, minimálna a maximálna frekvencia zobrazenej časti spektra sa udáva svojou absolútnou hodnotou (v Hz). Zvolený rozsah frekvenčnej osi pôsobí ako frekvenčná lupa pri prehliadke detailov, súvisiacich s frekvenčnou skladbou.

Rozsah y-ovej osi rozhoduje o tom, aké rozpätie stredného výkonu (v dB) bude zobrazovať graf spektra. Maximálny rozsah zvislej osi, v ktorom nájdeme informácie o rozložení akustickej energie harmonických zložiek, je 120 dB. Toto rozpätie je relatívne, môže sa posúvať okolo nuly do kladných alebo záporných hodnôt. Rozsah výkonov v grafe sa udáva decibelovou hodnotou minimálnej a maximálnej zobrazenej hodnoty, pôsobí ako lupa pri skúmaní detailov, súvisiacich s rozložením energie v zložkách spektra.

Literatúra

- ČÍŽEK, V.: *Diskrétní Fourierova transformace a její použití*. Praha, SNTL – Nakladatelství technické literatury 1981. 160 s.
- CHMÚRNÝ, J. – ŽIDEK, F.: *Signály a systavy*. Bratislava, Elektrotechnická fakulta Slovenskej vysokej školy technickej 1984. 272 s.
- ISAČENKO, A. V.: *Spektrografická analýza slovenských hlások*. Bratislava, Vydavateľstvo Slovenskej akadémie vied 1968. 263 s.
- KRÁL, Á. – SABOL, J.: *Fonetika a fonológia*. Bratislava, Slovenské pedagogické nakladateľstvo 1989, s. 108.
- KROUTL, F.: *Signály a hluky*. Praha, Nakladatelství dopravy a spojů 1964. 359 s.
- OPPENHEIM, A. V. – SCHAFER, R. W.: *Discrete-Time Signal Processing*. New Jersey, Prentice-Hall, Englewood Cliffs 1989.
- PRCHAL, J.: *Signály a soustavy*. Praha, SNTL – Nakladatelství technické literatury, ALFA – Vydavateľstvo technickej a ekonomickej literatúry 1987, 323 s.
- SABOL, J. – ZIMMERMANN, J.: *Základy akustickej fonetiky*. 3. vyd. Košice, Rektorát Univerzity P. J. Šafárika v Košiciach 1986a. 105 s.
- SABOL, J. – ZIMMERMANN, J.: *Dynamika formantovej štruktúry sonantického jadra slabiky*. *Govor/Speech (Zagreb)*, XIII (1996), č. 1–2, s. 91–104 (vyšlo 1998).
- ZIMMERMANN, J.: *Spektrografická a škálografická analýza akustického rečového signálu*. Prešov, Náuka 2002. 176 s.

Abstract

The study describes the principle of the short term spectrum calculation of a speech signal based on the decomposition of the periodical function by Fourier. The computation is determined for analog and discrete signal, for periodical and non-periodical segments. In the discrete domain there is an implied option of sampling frequency and the length of the weighting window. There are defined spectral bins and the consequences of spectral leakage. In the final practical part the author describes the principal of power spectrum, periodogram, curve smoothing, preemphasis and range of the frequency scale in regard to Nyquist frequency.

„Táto štúdia, bola vytvorená realizáciou projektu *Vybudovanie lingvokulturologického a prekladateľsko-tlmočnického centra*, na základe podpory operačného programu Výskum a vývoj financovaného z Európskeho fondu regionálneho rozvoja.“