

Sonagramy

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

Kľúčové slová: rečový signál, spektrogram, sonagram, úzkopásmový filter, širokopásmový filter

Keywords: speech signal, spectrogram, sonagram, narrow-band filter, wide-band filter

1. Úvod

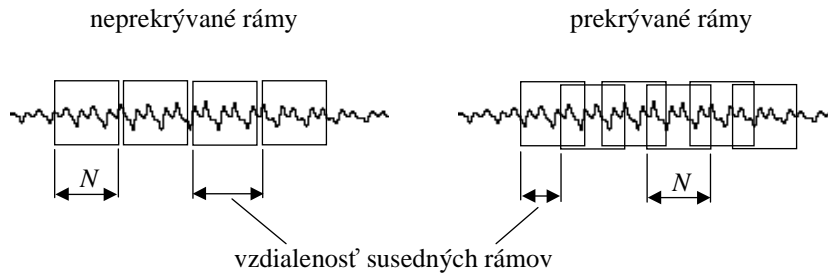
V štúdiu sa budeme zaoberať analýzou zmien spektra v čase. Predpokladáme, že čitateľovi je pojem spektrum dobre známy (bližšie pozri napr. v Zimmermann, 2010). Pretože rečový signál je sekvenciou hlások a slabík a trvanie jednej slabiky môže byť v rozsahu 50 – 500 milisekúnd, pričom sa spektrum mení aj v samotnom sonantickom jadre slabiky, statické spektrum frekvenčného rezu prináša veľmi málo informácií o súvislej reči. História vývoja metód na meranie zmien spektra začala v r. 1934, keď J. C. Steinberg publikoval spôsob „ručného“ výpočtu spektra z každého hlasivkového kmitu, samozrejme bez kalkulátora. V štúdiu *The Sound Spectrograph* (Koenig et al., 1946) autori opisujú prvý elektromechanický prístroj, využívajúci elektronickú filtráciu rečového signálu, ktorý kreslil na otáčajúci sa valec trojrozmerný graf zmien spektra – spektrogram. V úvode autori píše: „The sound spectrograph is a wave analyzer which produces a permanent visual record showing the distribution of energy in both frequency and time.“ Išlo o trojdimenzionálny graf, zachytávajúci závislosť frekvencie a akustickej energie na čase. Elektromechanický zvukový spektrogram sa v druhej polovici minulého storočia vyrábala v americkej firme Kay Elemetrics Company pod označením Sona-Graph, niesol aj označenie Visible Speech a uvedená firma vytvorila pre trojrozmerné spektrogramy osobitné – zákonom chránené označenie sonagram. Dnes je táto technológia nahradená algoritmami rýchlej Fourierovej transformácie, aplikačné programové prostredie je napísané pre personálne počítače a sonagramy sa študujú na monitore, resp. dajú sa vytlačiť na tlačiarňami.

2. Princíp spektrogramov

Pri skúmaní spektra dlhších úsekov rečového signálu, napr. dvojhlások, slabík, slov a pod., sa hľadajú informácie o zmene spektra ako funkcie času. Generovanie reči v procese artikulácie je spojitý dej, spektrum sa mení kontinuálne a ak má byť zmena frekvenčnej skladby vyhodnocovaná z jej číselných hodnôt, musí byť signál analyzovaný v diskretných časových okamihoch. Períodu, z ktorej vyplýva frekvencia snímania spektrálnych rezov, môžeme odvodiť z frekvencie hlasivkových kmitov. Výška základného tónu pre mužský a ženský hlas sa pohybuje v rozpätí približne od 100 do 300 Hz. Jeden spektrálny rez po každom glotálnom kmite je dostatočnou vzorkou na získanie informácie o zmene spektra. Tento odhad vyjadrený dĺžkou periódy cyklického snímania rezov predstavuje jednu analýzu za 10–30 milisekúnd. Keď kvantita priemerného slova je približne 500 ms, na jeho analýzu je potrebných 16 až 50 rezov. Každý spektrálny rez je dvojrozmerným grafom (frekvencia/stredný výkon). Spojité zmeny spektra analyzovaného slova sú pozorovateľné vtedy, ak tieto spektrálne rezy znázorníme v ich pôvodnom poradí s malými medzerami vedľa seba tak, aby vytvorili trojrozmerný graf série spektrálnych rezov (obr. 2). K *x*-ovej osi (frekvencia) a *y*-ovej osi (výkonové spektrum) tak pribudne *z*-ová os (čas) s diskretnými časovými okamihmi, v ktorých sú vytvorené spektrálne rezy (grafy výkonového spektra).

Ak má byť trojrozmerný graf série rezov objektívny, musia byť parametre všetkých rezov rovnaké. Zásadný vplyv na čitateľnosť trojrozmerného grafu má dĺžka okna a stupeň

vyhladenia krivky výkonového spektra. Z podstaty trojrozmerného grafu vyplýva, že séria spektrálnych rezov je sériou analyzovaných rámov, definovaných konštantnou dĺžkou okna (N vzorkami). Rámy sa môžu v časovej osi prekrývať, alebo môžu nasledovať tesne za sebou (obr. 1). K voliteľným parametrom teda pribúda stupeň prekrývania rámov, daný vzájomnou vzdialenosťou dvoch susedných rámov. Ak je táto vzdialenosť menšia, než dĺžka okna, rámy sa prekrývajú. Stupeň prekrývania rámov môže byť daný počtom vzoriek (podobne ako dĺžka okna N), alebo časovým údajom v milisekundách.



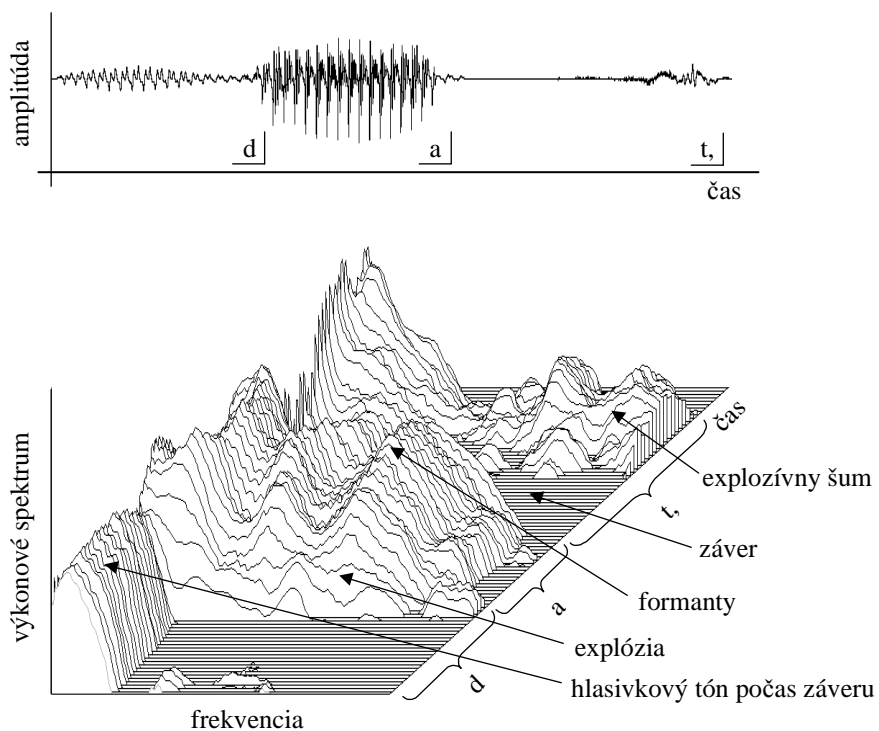
Obr. 1. Segmentovanie analyzovaného signálu na rámy a stupeň prekrývania rámov vyjadrený vzdialenosťou susedných rámov.

Prekrývaním rámov sa kompenzuje negatívny účinok váhovej funkcie na segment signálu. Keď sú na okrajoch segmentu (definovaného oknom) dôležité informácie, váhová funkcia ich zoslabí. Prekrývaním rámov sa do stredu nasledujúceho okna dostane tá časť signálu, ktorá bola okrajom predchádzajúceho okna.

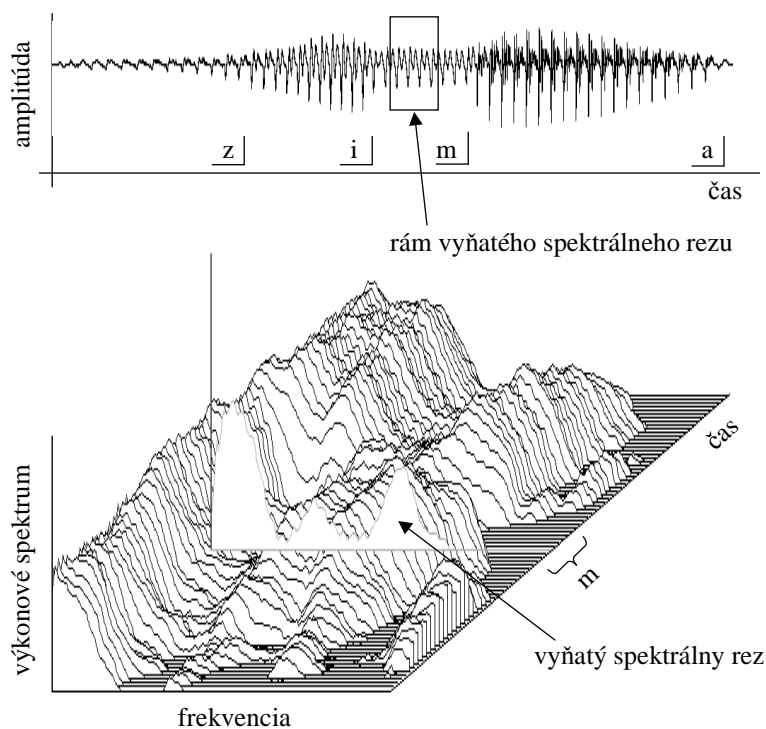
Druhým prínosom prekrývaných rámov je väčší počet spektrálnych rezov v analyzovanom úseku reči. Vyššie frekvenčné rozlíšenie v spektrálnom reze sa dosahuje dlhším oknom, z čoho vyplýva, že jeden rám pri vyššom frekvenčnom rozlíšení analyzuje dlhší segment signálu na časovej osi. Keď sú dlhšie rámy neprekrývané, vtedy je v analyzovanom úseku reči málo spektrálnych rezov a séria rezov nezachytí rýchlejšie zmeny spektra.

Na obr. 2 je oscilogram slova [dat,] a trojrozmerný graf série jeho spektrálnych rezov. Signál bol segmentovaný oknom dĺžky $N = 512$ vzoriek, medzi susednými rámami je vzdialenosť 64 vzoriek. V grafe možno dobre pozorovať artikulačné fázy: hlasivkový tón počas záveru znejšej spoluhlásky [d] a nasledujúcu explóziu, formantovú štruktúru samohlásky [a], záverovú fázu znejšej explózie [t,] a explóziívny šum po uvoľnení záveru.

Spektrum, zobrazené trojrozmerným grafom série spektrálnych rezov možno vyhodnocovať dvojakým spôsobom. Po prvé, v trojrozmernom grafe (obr. 2) sú dobre badateľné spojité zmeny spektra v procese artikulácie. Po druhé, spektrálne rezy si zachovávajú všetky informácie o výkonovom spektre aj napriek tomu, že sú čiastočne zakryté v trojrozmernom grafe. Možno ich teda vyňať zo spoločného grafu a vyhodnocovať osobitne. Na obr. 3 je trojrozmerný graf slova [zima]. Pretože signál spoluhlásky [m] má malé amplitúdy, akustická energia jej spektrálnych zložiek je menšia než akustická energia predchádzajúcej samohlásky [i]. Spektrálny rez spoluhlásky [m] je zakrytý, dá sa vyhodnocovať až po vyňatí zo série rezov.



Obr. 2. Oscilogram a trojrozmerný graf série spektrálnych rezov slova *dat'*.

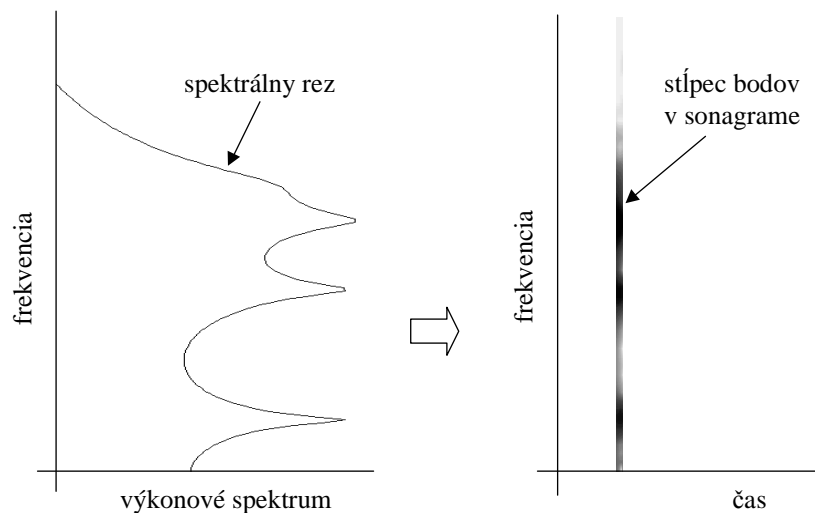


Obr. 3. Osobitné vyhodnocovanie spektrálneho rezu spoluhlásky *m*, vyňatého zo série spektrálnych rezov slova *zima*.

3. Sonagram

Trojrozmerný graf série spektrálnych rezov poskytuje základné informácie o zmene spektra v čase. Môžeme ho pokladať za jednoduchú formu dlhodobej Fourierovej transformácie. Trojrozmerný graf rezov má dva nedostatky. Po prvé, zakrytý rez sa dá pozorovať len po jeho vyňatí zo série rezov; v konečnom dôsledku ani vyňatie nerieši otázku synchronnej viditeľnosti všetkých rezov. Po druhé, pri zvyšovaní počtu rezov v trojrozmernom grafe sa strácajú kontúry jednotlivých rezov a graf sa stáva neprehľadným.

Zvyšovanie počtu spektrálnych rezov pri zachovaní ich viditeľnosti sa dosahuje zvláštnym typom trojrozmerného grafu – spektrogramom. Ak je spektrogram analýzou zvukových (teda aj rečových) signálov, nazýva sa sonagram. Problém vzájomného zakrývania rezov je riešený osobitnou úpravou tretieho rozmeru grafu: x -ovou osou je čas, y -ovou osou je frekvencia. Tretí rozmer, akustická energia (vyplývajúca zo stredného výkonu), nie je zobrazovaná z -ovou osou, ale odtieňom sivej farby bodov grafu. Presnejšie povedané, každý spektrálny rez je zobrazený ako samostatný stĺpec bodov (pixelov na obrazovke monitora, alebo bodov na vytlačennom grafe), pričom každý bod predstavuje pôdorys jednej spektrálnej čiary a odtieň bodu predstavuje jej stredný výkon (obr. 4).



Obr. 4. Princíp zobrazenia spektrálneho rezu v sonagrame.

Maximálny počet spektrálnych rezov v sonagrame je daný rozlišovacou schopnosťou (rozkladom) monitora v horizontálnom smere. Keď vodorovné rozlíšenie monitora je 1024 bodov na celej šírke obrazovky, pri využití 80% z tohto rozsahu môžeme v sonagrame zobrazit' až 800 rezov, teda 10-násobne viac, než v trojrozmernom grafe série spektrálnych rezov. Odstup susedných rezov, z ktorých vyplýva stupeň prekrývania rámov, je vzdialenosťou dvoch susedných rámov. Pre vzdialenosť v v dvoch susedných rezov platí:

$$v = \frac{s}{p} \tag{1}$$

kde s je rozsah analyzovaného signálu, p je počet horizontálnych pixelov, použitých na kreslenie sonagramu. Premenné v a s môžu byť dané časovou mierou, alebo počtom vzoriek, vzťah týchto dvoch premenných vyplýva zo vzorkovacej frekvencie.

Vzťah (1) hovorí, že stupeň prekrývania rámov je daný rozsahom signálu, analyzovaného sonagramom. Keď vezmeme do úvahy rámy s dĺžkou 64 – 1024 vzoriek a vzorkovacie frekvencie 10 kHz – 25 kHz, môžeme povedať, že jedným sonagramom sa dá analyzovať rečový signál v rozsahu do 2,5 sekundy. V tomto prípade bude hustota

spektrálnych rezov dostatočná na to, aby sonagram zobrazil spojité zmeny spektra. Ak by sme sonagramom analyzovali dlhšie úseky signálu, rámy by sa neprekrývali, resp. mohli by byť medzi nimi medzery a sonagram by nezachytil všetky informácie o priebehu spektrálnej skladby v časovej oblasti.

Druhým voliteľným parametrom pri kreslení sonagramu je dĺžka okna, určujúca frekvenčné rozlíšenie, resp. šírku hlavného laloku váhovej funkcie. Váhová funkcia má rovnaký účinok na signál ako pásmový priepust: hlavný lalok svojou šírkou určuje prepúšťané frekvenčné pásmo; postranné laloky utlmujú ostatné frekvenčné zložky, potláčajú presakovanie spektra (podrobnejšie v Zimmermann, 2002, s. 74 – 82, 96 – 98). Teda dĺžka okna rozhoduje o tom, či filter bude širokopásmový alebo úzkopásmový. Táto vlastnosť okna (rámu) má zásadný vplyv na podobu sonagramu.

Šírka hlavného laloku sa meria vo vzdialenosti -3 dB od úrovne špičky hlavného laloku (obr. 5). Dá sa vypočítať z frekvenčného odstuhu Δf vynásobením koeficientom ξ podľa vzťahu:

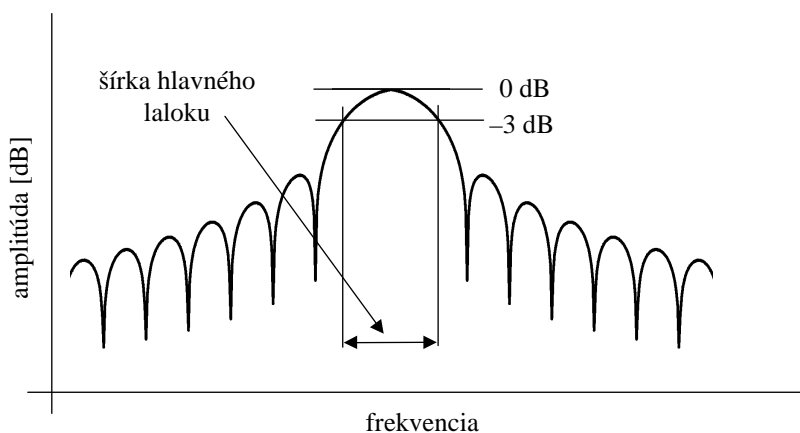
$$\text{šírka pásma} = \xi \Delta f = \xi \frac{F}{N} \quad (2)$$

kde F je vzorkovacia frekvencia, N je dĺžka daná počtom vzoriek. Koeficient ξ je charakteristikou zvoleného okna, jeho veľkosti sú uvedené v tabuľke:

okno		
Hanningovo	Hammingovo	Blackmanovo
1,44	1,30	1,68

Vypočítané šírky pásma Hanningovho okna pre rôzne vzorkovacie frekvencie a dĺžky okien uvádzame v nasledovnej tabuľke:

F	N								
	50	75	100	125	200	256	512	600	1024
10000	288	192	144	115	72	56	28	24	14
12500	360	240	180	144	90	70	35	30	18
16000	461	307	230	184	115	90	45	38	22
20000	576	384	288	230	144	112	56	48	28
25000	720	480	360	288	180	141	70	60	35



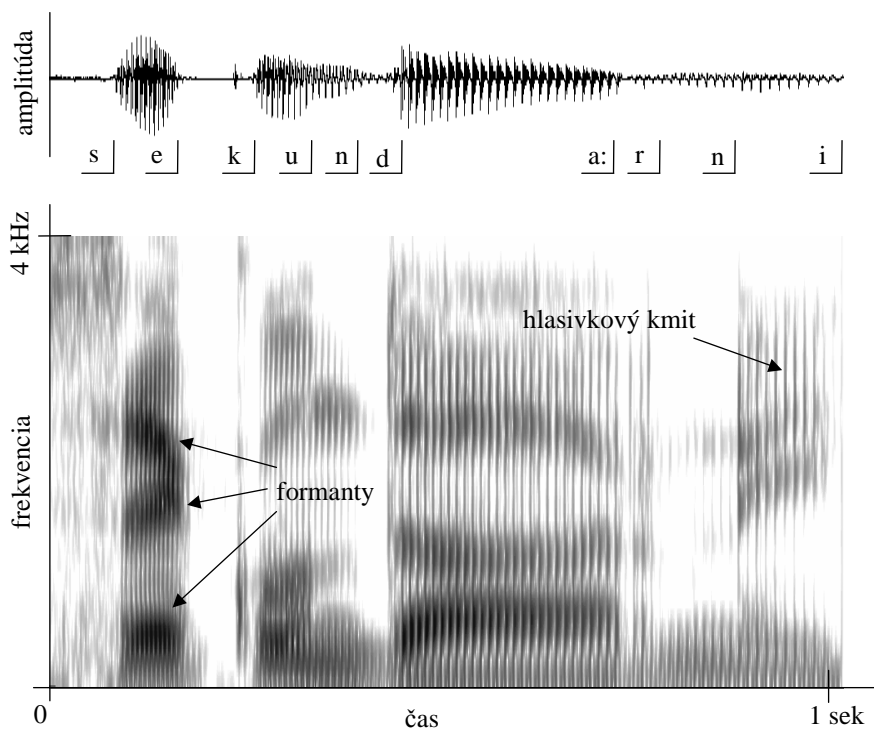
Obr. 5. Určenie šírky hlavného laloku, z ktorej vyplýva šírka pásma filtra.

Úzkopásmový filter so šírkou pásma približne 45 Hz (Isačenko, 1968, s. 86) zobrazí v sonagrame v spektre hlások s periodickou štruktúrou harmonické zložky ako násobky základného tónu, čím umožní skúmať výškové modulácie hlasu.

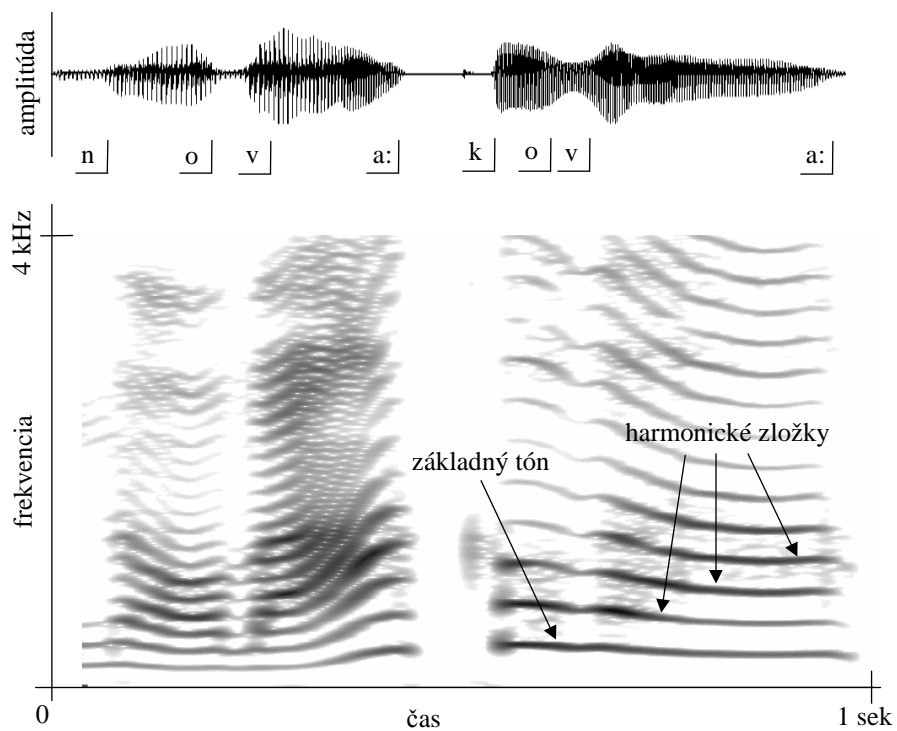
Širokopásmový filter so šírkou pásma približne 300 Hz je vhodný na sledovanie formantov. Pretože je dôsledkom kratšieho okna, je schopný zaznamenať aj krátkodobé zmeny spektra, na sonagrame sa teda ukážu hlasivkové kmity ako jemná vertikálna štruktúra v kresbe grafu.

Pri vytváraní sonagramov sa dá experimentovať s ďalšími voliteľnými parametrami, teda typom okna (pravouhlé, Hanningovo, Hammingovo, Blackmanovo a trojuholníkové), preemfázou a určením rozsahu časovej, resp. frekvenčnej osi. Logaritmický priebeh frekvenčnej osi poskytuje podrobnejší pohľad na výškovú moduláciu hlasu. Vyhladenie krivky spektrálneho rezu sa nastavuje šírkou pásma. Užitočnou voľbou je nastavenie tretieho rozmeru sonagramu (stredného výkonu) odtieňom sivej farby, alebo farebnou škálou. Táto voľba súvisí s grafickými schopnosťami monitora a tlačiarne.

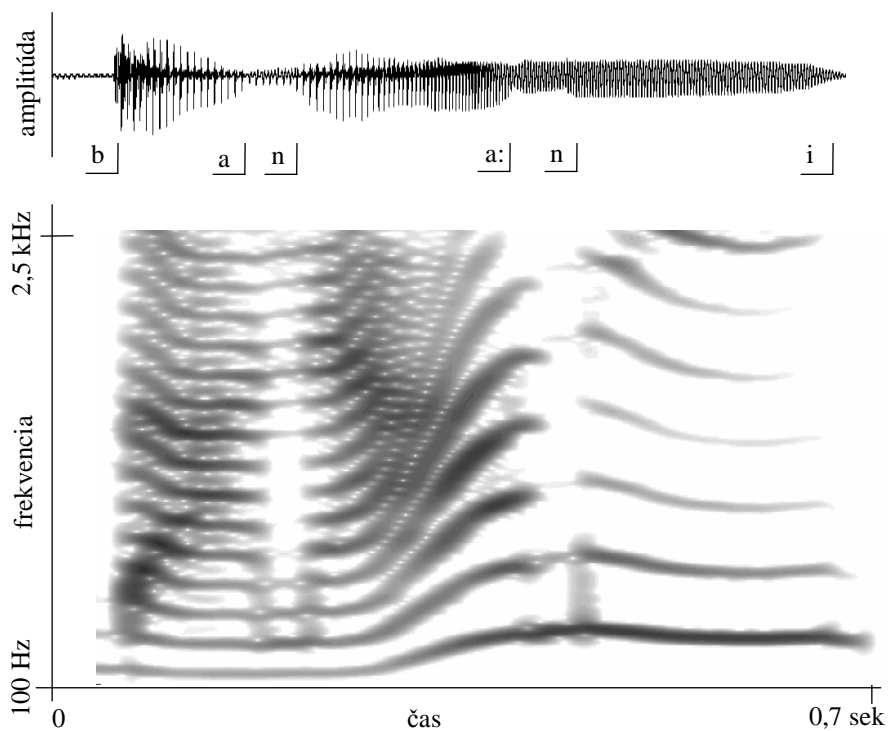
Na obr. 6, 7 a 8 uvádzame širokopásmový a úzkopásmový sonagram a sonagram so skráteným rozsahom frekvenčnej osi.



Obr. 6. Širokopásmový sonagram slova *sekundárny*. Šírka pásma filtra = 247 Hz.



Obr. 7. Úzkopásmový sonagram slova *Nováková?* Šírka pásma filtra = 31 Hz.



Obr. 8. Úzkopásmový sonagram slova *banány?* Šírka pásma filtra = 31 Hz.

Literatúra

- ISAČENKO, A. V.: *Spektrografická analýza slovenských hlások*. Bratislava, Vydavateľstvo Slovenskej akadémie vied 1968. 263 s.
- KOENIG, W.– DUNN, H. K. – LACY, L. Y.: *The Sound Spectrograph*. Bell Telephone Laboratories, Inc., New York, 1946.
- KRÁL, Á. – SABOL, J.: *Fonetika a fonológia*. Bratislava, Slovenské pedagogické nakladateľstvo 1989, s. 108.
- SABOL, J. – ZIMMERMANN, J.: *Základy akustickej fonetiky*. 3. vyd. Košice, Rektorát Univerzity P. J. Šafárika v Košiciach 1986a. 105 s.
- ZIMMERMANN, J.: *Spektrografická a škálografická analýza akustického rečového signálu*. Prešov, Náuka 2002. 176 s.
- ZIMMERMANN, J.: *Spektrálna skladba segmentov rečového signálu*. Jazyk a kultúra. Internetový časopis Lingvokulturologického a prekladateľsko-tlmočnického centra excelentnosti pri Filozofickej fakulte Prešovskej univerzity v Prešove. Ročník 1, číslo 2/2010.

Abstract

The paper deals with the method of exploration changes of speech spectrum in time. At first it is explained the way of connection of spectral section in such a way in order to be the sequence of spectrums enrich in time axis. Then it is discussed the principle of sonagrams, three dimensional spectral graph relation time, frequency and energy. There are explained the parameters by which is given time or frequency resolution of sonagrams. In the end the author explains the difference between the narrow-band and wide-band sonagrams.

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.