

Akvizícia merateľných dát v experimentálnej fonetike

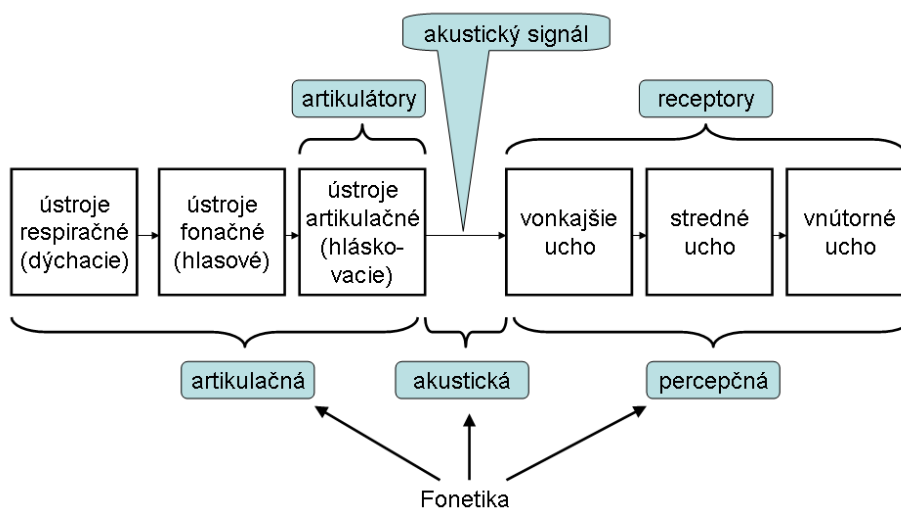
Július Zimmermann, Filozofická fakulta PU, zimmer@unipo.sk

Kľúčové slová: experimentálna fonetika, digitalizácia, vzorkovanie, dynamika signálu, šum, aliasing.

Keywords: experimental phonetics, digitalisation, sampling, dynamics of signal, noise, aliasing.

1. Predmetom záujmu experimentálnej fonetiky je exaktné skúmanie procesu akusticko-auditívnej komunikácie, v ktorej rečník kóduje informáciu do akustického signálu. Poslucháč dekóduje zvlnený akustický signál tak, aby v konečnom dôsledku detegoval informáciu, expedovanú rečníkom. Zvukové jednotky pritom môžeme abstrahovať dvojako – odlišujeme jednotlivé/variantné – fónu a všeobecné/invariantné – fonému (Sabol – Zimmermann, 2002, s. 37 – 40). Rýdzo fonetická analýza zvukov ľudskej reči vlastne neexistuje. Akákoľvek fónická analýza je užitočná vtedy, keď skúma vlastnosti rečových segmentov a suprasegmentov v procese jazykovej komunikácie. Inými slovami povedané, fonetické experimentovanie je nemysliteľné bez akustickej percepcie, bez posluchovej testov; prinajmenšom fonetik porovnáva namerané s tým, čo počuje, aby fonologicky určoval význam nameraných parametrov.

2. Ak neberieme do úvahy procesy kódovania a dekódovania informácie, akt rečovej komunikácie sa dá rozdeliť na tri časti: na organogézu reči, na postupné pozdĺžne vlnenie akustického signálu a na percepciu zmien akustického tlaku. Na základe tohto rozdelenia jestvujú tri odlišné fonetiky – fonetika artikulačná, fonetika akustická a fonetika perцепčná (obr. 1).



Obr. 1. Podstata určenia troch oblastí procesu rečovej komunikácie, skúmaných experimentálnou fonetikou.

Experimentálne najzložitejšou je artikulačná fonetika. Skúma dynamickú činnosť svalstva tváre počas artikulácie (bližšie pozri Krňoul, 2004), aj samotnú dynamickú činnosť artikulačného traktu – ústrojenstvo respiračné, fonačné a artikulačné – zavedením sofistikovaných senzorových prvkov do artikulačných dutín. Signály z týchto snímačov sa matematicky a graficky spracúvajú personálnym počítačom. Kritickým momentom je skúmaná osoba z dvoch dôvodov. Po prvé, každý rečník má inak usporiadaný artikulačný

trakt, z čoho vyplýva nutnosť prispôbovať senzory pre konkrétneho rečníka. Po druhé, skúmaná osoba musí prejsť ochotu spolupracovať na výskume a prekonať isté hygienické a estetické predsudky. Na obr. 2 uvádzame záber na aplikáciu zariadenia Artikulograph.



Obr. 2. Práca so zariadením Artikulograph v artikulačnej fonetike.

Predmetom záujmu percepčnej fonetiky je vnem segmentov a suprasegmentov a zisťovanie rozdielov medzi tým, čo je výsledkom fyzikálneho merania a tým, čo počuje ľudské sluchové ústrojenstvo. Môžeme tu hovoriť aj o fyziologických jednotkách vnemu, ktorými sa korigujú výsledky bohato rozvinutých analytických nástrojov, prevzatých z prírodných vied. Percepčná fonetika sa nezaobíde bez dômyselne pripravených posluchových testov, ktoré sa po splnení podmienok štatistickej významnosti spracúvajú pravdepodobnostnými a stochastickými prístupmi.

Akustická fonetika skúma akustický rečový signál, ktorým sa informácia dostane od expedienta k percipientovi. Delí sa na metódy analýzy, metódy syntézy a metódy resyntézy rečového signálu. Výskum zvyčajne začína akvizíciou signálu, po ktorej nasleduje analýza v časovej a frekvenčnej doméne. Hlavným nástrojom okrem dôsledne premysleného projektu je aplikácia diskretných matematických metód s výstupmi v číselnej aj grafickej podobe. Pretože je najviac rozšírená, v ďalšom texte sa jej budeme podrobne venovať.

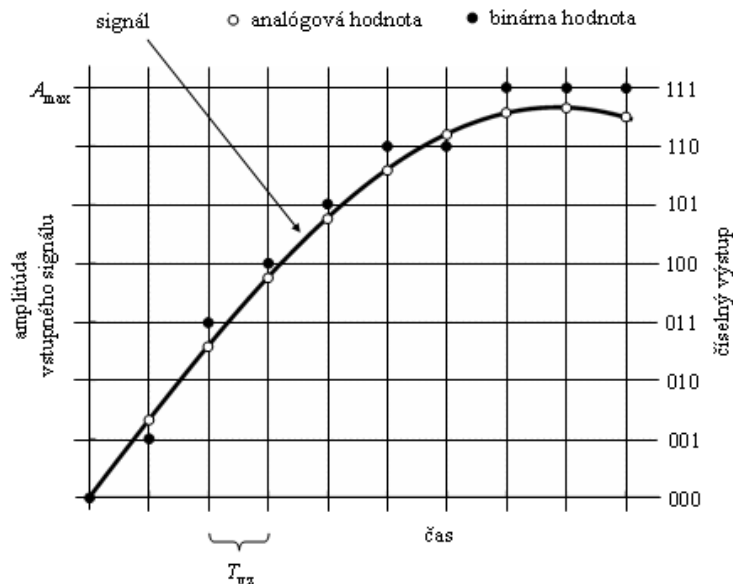
3. Rečový signál je stochastický, nestacionárny, v niektorých úsekoch kváziperiodický, s relatívne veľkou šírkou frekvenčného pásma. Je prirodzené, že tak zložitý jav sa dá dôsledne analyzovať len v jeho digitálnej podobe. Druhou príčinou nevyhnutnosti digitalizácie je potreba vykonávať prípravné činnosti – prehliadku osciloskopickéj krivky signálu na monitore počítača, používanie časovej lupy, segmentovanie signálu, editovanie signálu, zmenu amplitúdy, resp. vzorkovacej frekvencie signálu atď.

Digitalizácia analógového rečového signálu (obr. 3, podrobnejšie v Zimmermann, 2002, s. 28 – 37) je jeho transformácia do diskretnéj – číslicovej podoby. Analógovo-digitálny (A/D) prevod prebieha tak, že sa v pravidelných časových okamihoch meria amplitúda vstupného analógového signálu, čím sa získa postupnosť číselných hodnôt. Na výstupe A/D prevodu musia byť binárne čísla, aby mohli byť spracované počítačom. m -bitovým binárnym číslom sa dá vyjadriť 2^m čísel, napr. 3-bitové číslo môže nadobudnúť $2^3 = 8$ stavov (na obr. 3 sú to stavy 000, 001, 010, 011, 100, 101, 110, 111), 4-bitovým číslom $2^4 = 16$ čísel, 8-bitovým číslom 256 čísel, atď. Počet čísel binárneho čísla je konečný, ale počet odlišných hodnôt, ktoré získame meraním analógového signálu, je nekonečný. Nutne musíme nameranú okamžitú analógovú hodnotu zaokrúhliť na najbližšiu binárnu hodnotu, čo spôsobuje istú nepresnosť.

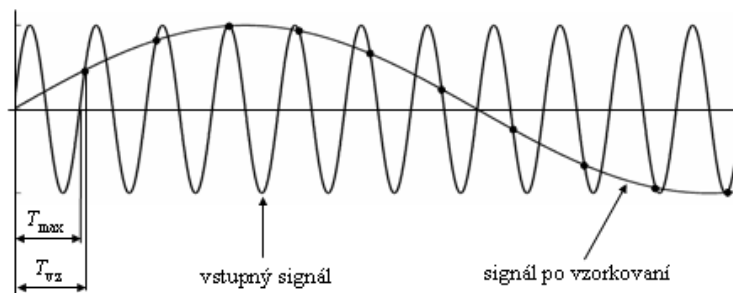
4. Kvalitu procesu digitalizácie analógového signálu určujú viaceré parametre, predovšetkým vzorkovacia frekvencia a šírka prevodu. Snímanie veľkosti amplitúdy – vzorkovanie sa realizuje vzorkovacou frekvenciou f_{vz} , pričom vzorkovacia perióda $T_{vz} = 1/f_{vz}$. Hodnota vzorkovacej frekvencie je kľúčovým parametrom A/D prevodu. Pri jej malých hodnotách vzorkovanie nezachytáva skutočný priebeh analógového signálu, príliš veľké hodnoty f_{vz} zbytočne zaťažujú prevod. Pre vzorkovaciu frekvenciu platí vzorkovacia teoréma, definovali ju nezávisle na sebe Shannon, Kotelnikov a Nyquist. Vzorkovaciu teorému môžeme formulovať takto: vzorkovacia frekvencia f_{vz} musí byť aspoň 2-násobkom frekvencie najvyššej sínusovej zložky f_{max} , ktorá sa v zloženom tóne vstupného signálu nachádza. Platí:

$$f_{vz} \geq 2 f_{max} \quad (1)$$

Negatívny jav, ktorý vzniká pri nedodržaní nerovnosti (1), sa nazýva aliasing, jeho následkom môže byť úplná zmena vstupného signálu. Vznik aliasingu vysvetľuje obr. 4. Zo vzorkovacej teorémy vyplýva, že na prevod jednej vlny sínusového tónu sú potrebné aspoň dve vzorky. Inými slovami: vstupný sínusový signál s frekvenciou vyššou ako $f_{vz}/2$, vzorkovaný frekvenciou f_{vz} , negatívne ovplyvní výsledok prevodu.



Obr. 3. Princíp A/D prevodu signálu.



Obr. 4. Vznik aliasingu pri vzorkovaní s príliš nízkou vzorkovacou frekvenciou, $f_{vz} < f_{max}$.

Pokúsme sa analyzovať vznik aliasingu. Do úvahy budeme brať frekvenciu vstupného signálu f a vzorkovaciu frekvenciu f_{vz} . Vzorkovaním vzniká následnosť impulzov, výška

každého impulzu je rovná amplitúde analógového signálu v okamihu snímania vzorky. Fourierovou analýzou sa dá dokázať, že spektrum následnosti impulzov je diskrétna a má periodickú podobu (Prchal, 1987, s. 191). Pre nadbytočné frekvencie f_i tohto spektra, ktoré vznikli vzorkovaním, platí nerovnosť:

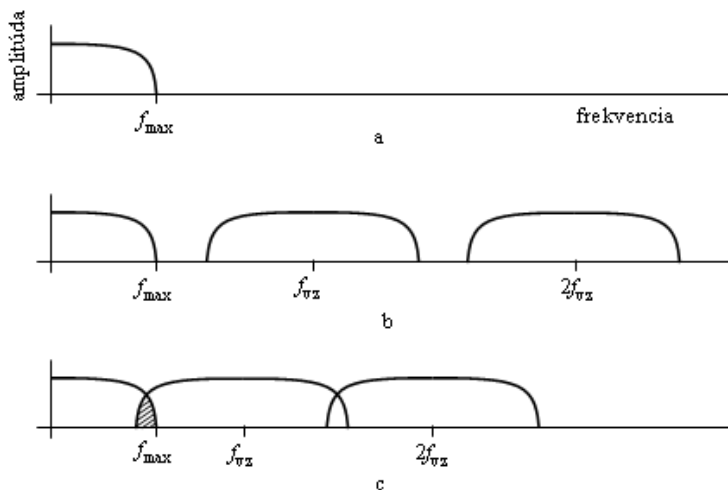
$$\frac{k \cdot f_{vz}}{2} \leq f_i \leq \frac{(k+2) \cdot f_{vz}}{2} \quad (2)$$

pričom k sú nepárne celé čísla, $k = 1, 3, 5 \dots$

Spektrum následnosti impulzov získané Fourierovou analýzou je diskrétna, jeho obalová krivka má tvar znázornený na obr. 5b, resp. 5c. Zvláštnu pozornosť musíme venovať pásmu frekvencií, označených na obr. 5c šrafovaním. Tieto frekvencie sú nežiadúce, priamo zasahujú do vstupného signálu a menia jeho vlastnosti v oblasti vysokých kmitočtov. Nazývajú sa aliasingové kmitočty s frekvenciou f_a . Z obr. 5b a 5c vyplýva, že aliasingové tóny nezasiahnu do frekvenčného pásma vstupného signálu vtedy, ak $f_{vz} \geq 2 \cdot f_{max}$. Vstupná sínusová zložka s frekvenciou f vyvolá vznik aliasingového tónu f_a vtedy, ak platí:

$$f_a = \left| f - \frac{(k+1) \cdot f_{vz}}{2} \right| \quad (3)$$

kde k sú nepárne celé čísla, $k = 1, 3, 5 \dots$



Obr. 5. a – spektrum vstupného analógového signálu, b – spektrum vzorkovaného signálu ak $f_{vz} > 2 \cdot f_{max}$, c – spektrum vzorkovaného signálu ak $f_{vz} < 2 \cdot f_{max}$.

Následky aliasingu si môžeme overiť na príklade. Nech vzorkovacia frekvencia $f_{vz} = 30$ kHz a vstupný signál má frekvenciu $f = 16,5$ kHz. Vyšetříme, akú frekvenciu má aliasingový tón f_a . Podľa vzťahu (3) pre $k = 1$ bude:

$$f_a = \left| f - \frac{(k+1) \cdot f_{vz}}{2} \right| = \left| 16,5 - \frac{(1+1) \cdot 30}{2} \right| = 13,5 \text{ kHz}$$

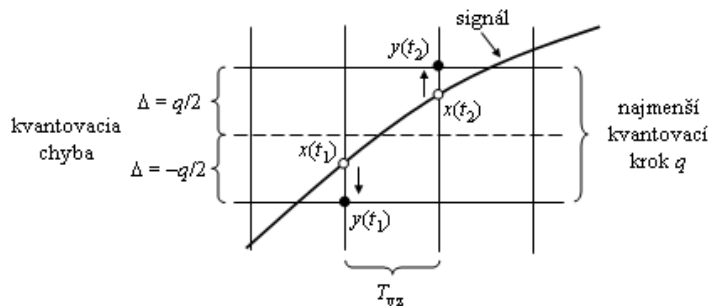
Podľa vzorkovacej teóremy (1) najvyšší tón, ktorý môže byť korektné spracovaný vzorkovaním pri $f_{vz} = 30$ kHz, je $f_{vz}/2 = 15$ kHz. Ak by v našom príklade išlo o zložený tón s maximálnou frekvenciou $f_{max} = 16,5$ kHz, aliasingom negatívne ovplyvnené pásmo by malo rozpätie $f_{vz}/2 - f_a = 1,5$ kHz.

Na prvý pohľad sa môže zdať, že vzorkovacia frekvencia, spĺňajúca podmienku vzorkovacej teóremy, je dostatočnou obranou proti aliasingu. Okrem rešpektovania vzťahu (1) musíme brať do úvahy ešte ďalšie okolnosti. Po prvé, maximálna frekvenčná zložka

vstupného signálu nie je vždy známa. Po druhé, signál môže obsahovať aj také vyššie harmonické tóny, ktoré nás pri analýze nezaujímajú. Po tretie, mnohé algoritmy matematickej analýzy pracujú racionálnejšie pri nižšej vzorkovacej frekvencii. Na základe týchto skutočností sa vždy pred vzorkovaním frekvenčné zložky vyššie než $f_{vz}/2$ odstránia, čím sa dosiahne rovnosť $f_{max} = f_{vz}/2$. Hodnota $f_{vz}/2$ sa stanoví podľa toho, k čomu má digitalizácia slúžiť. Ak sa má A/D prevodom spracovať zvukový signál tak, aby sa zachovala jeho verná podoba, musí byť hodnota $f_{vz}/2$ vyššia, nanajvýš rovná s najvyšším tónom, ktorý je vnímateľný sluchom. Keď sa rečový signál digitalizuje za účelom analýzy základného – hlasivkového tónu, môže byť horná hraničná frekvencia nastavená na hodnotu $f_{vz}/2 = 5$ kHz. Polovičná hodnota vzorkovacej frekvencie, $f_{vz}/2$ sa v literatúre označuje výrazom Nyquistova frekvencia.

Frekvenčné zložky vstupného signálu vyššie než $f_{vz}/2$ sa odstraňujú vstupným antialiasingovým filtrom. Tento filter má charakter dolného priepustu, teda prepúšťa na svoj výstup len nižšie frekvenčné zložky, nepresahujúce hodnotu $f_{vz}/2$.

5. V odseku 3., kde sme naznačili princíp digitalizácie, sme uviedli, že odmeraná okamžitá amplitúda vstupného analógového signálu sa musí vyjadriť najbližšou binárnou hodnotou výstupného dvojkového čísla, musí sa kvantovať, z čoho vyplýva nepresnosť prevodu. Pri kvantovaní vzniká kvantovacia chyba, ktorá je zdrojom kvantovacieho šumu. Analýzu kvantovacieho šumu uskutočníme pomocou obr. 6.



Obr. 6. Princíp kvantovania signálu a pôvod kvantovacieho šumu.

Výstupné dvojkové číslo, pozostávajúce z m bitov, má 2^m stavov. Rozlišovacia schopnosť prevodníka je daná šírkou prevodu, počtom bitov m výstupného dvojkového čísla. Maximálna amplitúda vstupného signálu A_{max} môže byť kvantovaná len do 2^m možných stavov. Hodnota $1/2^m$ je najmenším kvantovacím krokom q , je zhodná s najmenej významným bitom číslicového obrazu hodnoty analógového vstupu.

Podľa obr. 6 kvantovaním analógovej hodnoty $x(t_1)$ vytvoríme číslicový obraz $y(t_1)$, kvantovaním hodnoty $x(t_2)$ vytvoríme obraz $y(t_2)$, atď., pričom kvantujeme na najbližšiu hodnotu binárneho čísla. Najväčšia chyba, ktorá v tomto procese môže vzniknúť, je kvantovacia chyba, ktorú označíme symbolom Δ . Kvantovacia chyba môže byť kladná alebo záporná, platí nerovnosť:

$$|\Delta| \leq \frac{q}{2} \quad (4)$$

Keď predpokladáme, že hodnota amplitúdy vstupného signálu kolíše okolo nulovej hodnoty, teda že môže mať kladné aj záporné hodnoty, potom pre maximálnu kladnú amplitúdu signálu platí:

$$A_{max} = \frac{2^m}{2} \cdot q = 2^{m-1} \cdot q \quad (5)$$

Efektívna hodnota signálu bude:

$$A_{ef} = \frac{2^{m-1} \cdot q}{\sqrt{2}} \quad (6)$$

Podobne môžeme zhodnotiť aj kvantovací šum výpočtom jeho stredného výkonu. Stredný výkon P harmonického signálu $x(t)$ vyčíslený za dobu jednej periódy T (Prchal, 1987, s. 35) bude:

$$P = \frac{1}{T} \int_0^T x^2(t) dt \quad (7)$$

Aplikujme tento vzťah na kvantovací šum. Stredná hodnota štvorca kvantovacej chyby Δ za predpokladu rovnakej pravdepodobnosti chyby v intervale $-q/2$ až $q/2$ bude:

$$P = \frac{1}{q} \int_{-q/2}^{+q/2} \Delta^2 d\Delta = \frac{1}{q} \left[\frac{\Delta^3}{3} \right]_{-q/2}^{+q/2} = \frac{1}{q} \left(\frac{q^3}{24} + \frac{q^3}{24} \right) = \frac{1}{q} \cdot \frac{2q^3}{24} = \frac{q^2}{12} \quad (8)$$

Hodnotu $P = q^2/12$ môžeme pokladať za energiu šumu. Na základe vysloveného predpokladu o konštantnom rozdelení pravdepodobnosti výskytu chyby v intervale $-q/2$ až $+q/2$ má tento šum charakter bieleho šumu. Efektívna hodnota šumu je:

$$N_{ef} = \sqrt{P} = \frac{q}{\sqrt{12}} \quad (9)$$

Efektívna hodnota kvantovacieho šumu N_{ef} je priamo úmerná hodnote kvantovacieho kroku q . Vieme ju zhodnotiť až vtedy, keď ju porovnáme s efektívnou hodnotou analógového signálu. V praxi je dôležitou charakteristikou pomer signálu k šumu, v našom prípade platí pre tento pomer na základe vzťahov (6) a (9):

$$\frac{A_{ef}}{N_{ef}} = \sqrt{1,5 \cdot 2^m} \quad (10)$$

Pretože výsledkami výpočtov podľa vzťahu (10) sú číselné hodnoty v príliš veľkom rozsahu, pomer signálu k šumu sa vyjadruje v decibeloch. Takto vyjadrený parameter je označovaný výrazom Signal to Noise Ratio (SNR), počíta sa podľa vzorca:

$$\left[\frac{A_{ef}}{N_{ef}} \right]_{dB} = 20 \cdot \log(\sqrt{1,5 \cdot 2^m}) = 1,76 + 6,02m \quad (11)$$

V tabuľke 1 uvádzame hodnoty odstupe signál-šum pre prakticky najpoužívanéjšie šírky prevodu.

Tabuľka 1. Pomer signál-šum vypočítaný zo šírky prevodu.

šírka prevodu (bit)	8	9	10	11	12	13	14	15	16
SNR (dB)	50	56	62	68	74	80	86	92	98

Z tabuľky 1 vyplýva, že následkom predĺženia šírky prevodu o každý bit je zmenšenie kvantovacieho šumu o 6 dB.

6. Dynamickým rozsahom (DR) v analógovej technike je pomer maximálnej a minimálnej amplitúdy signálu, ktorá môže byť zariadením spracovaná bez skreslenia. Udáva sa v decibeloch. Pri A/D prevode je tento pomer jednoznačne daný šírkou prevodu, teda počtom 2^m stavov. Najmenšie dvojkové číslo výstupnej binárnej hodnoty patrí najnižšej amplitúde vstupného signálu a najväčšie dvojkové číslo najvyššej amplitúde. Výpočet dynamického rozsahu v decibeloch môžeme vykonať na základe vzťahu:

$$DR = 20 \cdot \log \frac{A_{\max}}{\frac{A_{\max}}{2^m}} = 20 \cdot \log 2^m \quad (12)$$

7. Pri digitálnom zázname sa nedá ideálne nastaviť veľkosť amplitúdy signálu, aby sa naplno využila rozlišovacia schopnosť šírky prevodu. Preto musíme jeho vzorky normovať. Normovanie je transformácia vektora vzoriek rečového signálu tak, aby maximálna kladná (resp. záporná) amplitúda signálu pokryla celý rozsah 16-bitového rozlíšenia, t.j. 65536 odlišných úrovní. V tomto procese sa škáluje a posunie hodnota každej vzorky smerom ku kladnému, resp. zápornému maximu, zároveň sa odstraňuje jednosmerná zložka signálu, ktorá sa pokladá hneď po nedostatočnej amplitúde za druhý negatívny vplyv na kvalitu digitalizovania. Normovanie zvyšuje dynamický rozsah signálu, znižuje kvantovaciu chybu prevodu; matematická analýza je po normovaní optimalizovaná.

Ak vektor vzoriek signálu je premennou x a vektor normovaných vzoriek signálu s odstránenou jednosmernou zložkou je premennou y , potom výpočet premennej y sa dá vykonať vzťahom

$$y = (x - \text{mean}(x)) / \max(\text{abs}((x - \text{mean}(x)))) \quad (13)$$

V tomto vzťahu „mean“ značí priemernú hodnotu vzoriek signálu, „max“ značí maximálnu hodnotu a „abs“ značí absolútnu hodnotu.

Na preukázanie vplyvu nedostatočnej (nenormovanej) amplitúdy signálu uvádzame nasledujúcu tabuľku:

Tabuľka 2: Vzťah medzi šírkou prevodu, SNR a DR.

Šírka prevodu m [bit]	2^m	DR [dB]	SNR [dB]	Amplitúda %
8	256	48,16	49,92	0,4
9	512	54,19	55,94	0,8
10	1024	60,21	61,96	1,56
11	2048	66,23	67,98	3,12
12	4096	72,25	74	6,25
13	8192	78,27	80,02	12,5
14	16384	84,29	86,04	25
15	32768	90,31	92,06	50
16	65536	96,33	98,08	100

Z tabuľky možno odčítať, ako sa zmení rozlišovacia schopnosť prevodníka, dynamický rozsah a odstup signál-šum pri nedostatočnej amplitúde. Napr. ak by mal signál len 12% z maximálneho rozsahu špička-špička, signál by bol kvantovaný na 8192 odlišných úrovní, dynamický rozsah by sa znížil o 18 dB a odstup signál-šum by poklesol o 18 dB.

Praktické dôsledky zníženej (nenormovanej) amplitúdy sa neprejavujú v spektre. V sonagrame sa objaví šum v miestach s nízkou energiou. V experimentoch, kde sa skúmajú lokálne maximá, určuje sa orezávacia hladina a porovnávajú sa analýzy viacerých nahrávok s nerovnakými amplitúdami, sa tento nedostatok, pravda, môže prejaviť.

8. Uviedli sme rozbor podmienok, nevyhnutných na to, aby mal digitalizovaný rečový signál vyhovujúcu kvalitu na fonetickú analýzu. Nekvalita signálu sa najčastejšie prejaví prítomnosťou parazitných šumov, skresleným (nepravým) spektrom a malou amplitúdou. Niektoré algoritmy pokladajú degradovaný signál za ruch, resp. za nulový signál, z čoho vyplýva zlyhanie procesu analýzy.

Literatúra

KRŇOUL, Z.: *Vizuální syntéza řeči – Mluvicí hlava*. Odborná práce ke státní doktorské zkoušce. Plzeň, Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra kybernetiky 2004.

PRCHAL, J.: *Signály a soustavy*. Praha, SNTL – Nakladatelství technické literatury, ALFA – Vydavatelství technické a ekonomické literatury 1987. 323 s.

SABOL, J. – ZIMMERMANN, J.: *Akustický signál – semióza – komunikácia*. Acta Facultatis Philosophicae Universitatis Prešoviensis. Monographia 51 (AFPh UP 99/181). Prešov, Prešovská univerzita v Prešove, Filozofická fakulta 2002. 144 s.

ZIMMERMANN, J.: *Spektrografická a škálografická analýza akustického rečového signálu*. Prešov, Náuka 2002.

Abstract

The paper deals with keeping of speech signal at its converting from continual to discrete form. First there are defined three fields of experimental phonetics – articulation, acoustics and perception phonetics. There is explained the principle of A/D conversion; by means of explanation of sampling theorem there is derivate sampling frequency. The next phenomena is negative aliasing, the author shows the relation between aliasing and Nyquist frequency. The paper contains also the explanation of quantization noise, relation between signal and noise and the explanation of dynamic signal range. In the end there is given the way of correction of insufficient amplitude signal by means of standardization.

„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.“