Portál pre odborné publikovanie ISSN 1338-0087

Dereverberácia reči s využitím lineárnej predikcie

András Imrich · Informačné technológie

27.02.2017

Predmetom tejto práce je algoritmus na skvalitnenie reči znehodnotenej dozvukom s využitím lineárnej predikcie. V úvode práce sú popísané základné mechanizmy vzniku dozvuku a ich vplyv na rečový signál. Nasleduje popis algoritmu s úpravou a spriemerňovaním hlasivkových cyklov, ktorý je určený na potlačenie dozvuku v rečovom signáli snímanom mikrofónovým poľom. Výhodou algoritmu je výpočtová nenáročnosť a úspešnosť pri vysokých úrovniach šumu a dozvuku. Záver práce je venovaný popisu jednotlivých evaluačných metód a prezentácii úspešnosti algoritmu pri rôznych podmienkach.

1. Účel dereverberácie

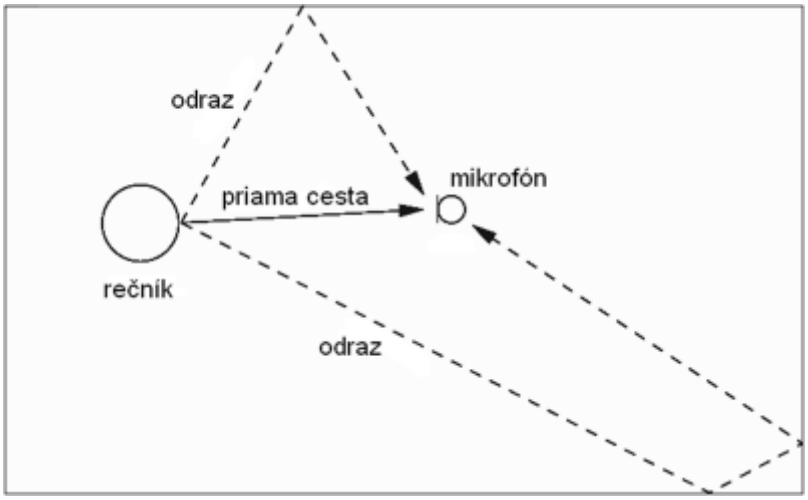
Spracovanie rečových signálov je predmetom výskumu už niekoľko dekád, o čom svedčí nepreberné množstvo publikácií v tejto oblasti. Pri skvalitňovaní audiosignálov sa vo väčšine prípadov jednalo o redukciu šumu a subpriestorové algoritmy sa donedávna využívali iba na vytváranie priestorových efektov (napr. stereo). Potlačenie nežiaducich priestorových efektov zo snímaných signálov – odstránenie dozvuku sa stalo predmetom vedeckých výskumov a článkov až v priebehu posledných rokov. Vo významnej miere to bolo vyvolané znížením ceny a masovým rozšírením prenosných zariadení ako sú mobily, laptopy, PDA, používateľské rozhrania v automobilovom priemysle atď. so súčasným zvýšením ich výpočtového výkonu. Vznikol tak priestor na služby ako hlasové ovládanie, konverzia reči na text, identifikácia hovoriaceho a pod., z čoho vyplynuli požiadavky na vývoj dereverberačných algoritmov.

Účinné dereverberačné metódy sú vo všeobecnosti výpočtovo náročné, a ich potenciálne využitie je prevažne v mobilných aplikáciách, takže si ťažko predstaviť komerčné využitie dereverberácie pred dvadsiatimi rokmi. Neustále zvyšovanie ako štandardov na trhu osobných elektronických zariadení, tak ich výkonu, však otvára cestu aj tejto oblasti skvalitňovania audiosignálov. Algoritmus s úpravou a spriemerňovaním hlasivkových cyklov (ÚSHC), ktorý je predmetom tejto práce, má byť kompromisom medzi čo najlepším výsledkom a čo najnižšími požiadavkami na výpočtový systém.

2. Vznik dozvuku

Ak je rečový signál snímaný v uzavretom priestore jedným alebo viacerými mikrofónmi umiestnenými v určitej vzdialenosti od rečníka, pozorovaný signál sa skladá z mnohých

superponovaných kópií rečového signálu. Tieto sú oneskorené a utlmené kvôli odrazom od stien a ďalších objektov v priestore, ako je znázornené na Obr. 1. Je zrejmé, že odrazy dorazia k mikrofónu s oneskorením, pretože cesta každého odrazu je dlhšia ako priama, a tiež utlmené, kvôli frekvenčne závislej absorpcii odrazových povrchov [1].

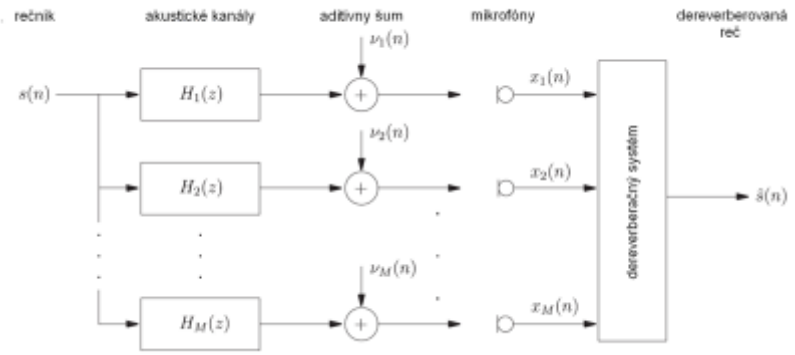


Obr. 1 Vznik dozvuku v uzavretom priestore

Rečový signál s dozvukom sa javí ako ten istý signál prichádzajúci z viacerých zdrojov rôzne rozmiestnených v priestore a tým v rozdielnych časoch a s rozdielnymi intenzitami. To dodáva priestorovosť do nasnímanej reči a percepčne sa javí, akoby rečník hovoril cez prekážku alebo akoby bol veľmi vzdialený od mikrofónu. Mierny dozvuk v reči ešte nezhoršuje zrozumiteľnosť a je vnímaný ako prirodzený a pri posluchu príjemný [2]. Návrhom akustického prostredia je ho možné kontrolovať, čo sa praktizuje pri ozvučovaní a prizvučovaní, dozvuk však zhoršuje spracovateľnosť reči systémami automatického rozpoznávania reči.

Problémy spojené s dozvukom často možno obísť umiestnením mikrofónu blízko k ústam rečníka, napr. použitím náhlavnej súpravy. Tým sa výrazne zvýši intenzita signálu z priamej cesty oproti odrazom a šumu, čo ich učiní prakticky zanedbateľnými. Takéto riešenie však nemusí byť vždy praktické, čo otvára cestu dereverberačným algoritmom.

Nech čistý rečový signál s(n) od rečníka prechádza akustickými kanálmi $H_m(z)$,m = 1,...,M . Výstupy týchto kanálov sú merané M mikrofónmi ako signály $x_m(n)$. Aditívny šum je reprezentovaný $v_m(n)$ a nech je to jediný druh šumu v modeli reverberácie a dereverberácie. Signál $x_m(n)$ meraný mikrofónom m je superpozíciou signálu priamej cesty, s príslušným oneskorením a útlmom, a teoreticky nekonečného počtu odrazov prichádzajúcich s vlastnými oneskoreniami a útlmami. Tento model je znázornený na Obr. 2.



Obr. 2 Model reverberácie a dereverberácie

Ak útlm a oneskorenie pre každý kanál m=1,...,M a odrazovú cestu i=0,1,...,∞ opíšeme akustickou impulzovou odozvou $h_{m,i}$, signál $x_m(n)$ sa dá vyjadriť ako

$$x_m(n) = \sum_{i=0}^{\infty} h_{m,i}(n)s(n-1) \tag{1}$$

Cieľom dereverberácie je nájsť systém, ktorý zo vstupov $x_m(n)$, m=1,...,M vráti výstup $\hat{s}(n)$, ktorý je dostatočne presným odhadom čistého rečového signálu s(n). Význam výrazu „dostatočne presný" je závislý od aplikácie, kritériom môže byť stredná kvadratická chyba priebehu, percepčná kvalita alebo iné. Akustické kanály $H_m(z)$ sú pritom neznáme.

3. Lineárna predikcia reči

Základný princíp lineárnej predikčnej (LP) analýzy je založený na predpoklade, že n-tá vzorka rečového signálu s(n) môže byť vyjadrená lineárnou kombináciou predchádzajúcich P vzoriek a budiacej postupnosti u(n), čo môže byť vyjadrené vzťahom

$$s(n) = -\sum_{i=1}^{P} a_i.s(n-1) + G.u(n) \tag{2}$$

kde G je zosilnenie, P je rád modelu a a_i sú lineárne predikčné koeficienty, o ktorých predpokladáme, že sú krátkodobo konštantné. Vyjadrením rovnice (2) v Z-oblasti a úpravou dostaneme prenosovú funkciu sústavy v tvare

$$H(Z) = \frac{G}{1+\sum_{i=1}^{P} a_i.z^{-1}} = \frac{G}{A(z)} \tag{3}$$

kde

$$A(z) = 1 + \sum_{i=1}^{P} a_i.z^{-1} \tag{4}$$

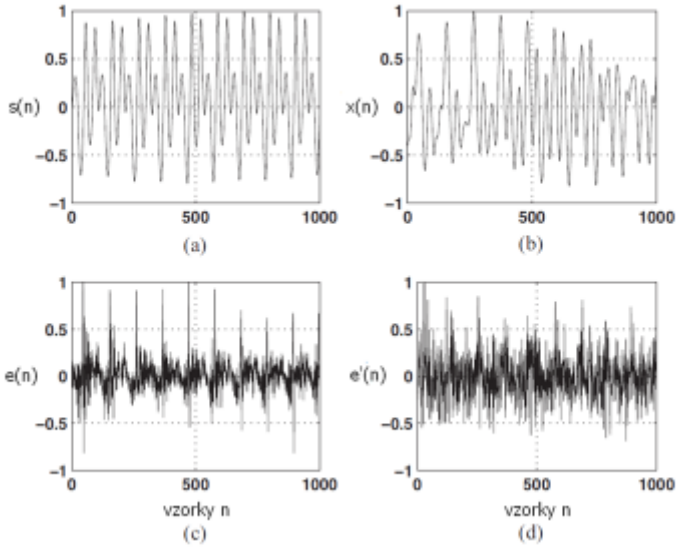
je tzv. inverzný filter. Na výpočet LP koeficientov a_i a koeficientu zosilnenia G sa používa metóda najmenších štvorcov. Pri analýze rečového signálu nie je známa budiaca funkcia u(n), vychádzame preto z odhadu rečového signálu, ktorý je lineárnou kombináciou iba predchádzajúcich P vzoriek:

$$\hat{s}(n) = -\sum_{i=1}^{P} a_i.s(n-1) \tag{5}$$

Definujeme predikčnú chybu e(n) ako rozdiel medzi skutočným signálom a jeho odhadom

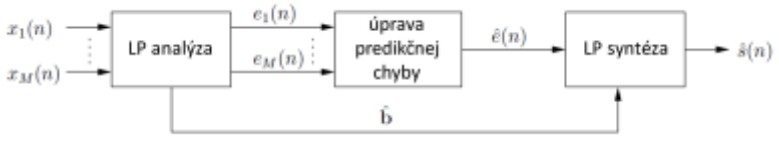
$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i=1}^{P} a_i . s(n-1) \quad [3] \ (6)$$

Spätná syntéza reči z LP koeficientov a predikčnej chyby je proces imitujúci činnosť vokálneho traktu, pričom model hlasového traktu je krátkodobo popísaný predikčnými koeficientmi a_i. Hlasový trakt reprezentovaný prenosovou funkciou H(z) je budený predikčnou chybou e(n) a výstupom je rekonštruovaný signál. Dôležitou vlastnosťou LP analýzy je, že LP koeficienty a_i sa pri znehodnotení reči dozvukom menia iba v malej miere. Podstatou dereverberačných algoritmov s lineárnou predikciou je spracovanie predikčnej chyby, ktorá je dozvukom ovplyvňovaná podstatne. Príklady priebehov predikčných chýb čistého a reverberovaného rečového signálu sú na Obr. 3.



Obr. 3 Priebeh znelého segmentu (a) čistého rečového signálu, (b) rečového signálu s dozvukom, (c) predikčnej chyby čistého rečového signálu, (d) predikčnej chyby rečového signálu s dozvukom

Charakteristickým znakom predikčnej chyby čistého rečového signálu v znelých segmentoch sú kváziperiodické excitačné špičky, ktoré majú tvar úzkych impulzov. Ich polohy v čase zodpovedajú hlasivkovým impulzom. Ak je rečový signál znehodnotený dozvukom, excitačné špičky sú buď do určitej miery rozprestreté v čase, alebo nasleduje viacero excitačných špičiek (a tým zdanlivo hlasivkových impulzov) za sebou. Poloha týchto dozvukových excitačných špičiek je zdanlivo náhodná, avšak majú tendenciu korelovať so skutočným hlasivkovým impulzom. Utlmením chybových dozvukových špičiek pri zachovaní resp. dotvarovaní skutočnej excitačnej špičky sa dá realizovať dereverberácia rekonštruovaného rečového signálu v znelých segmentoch.



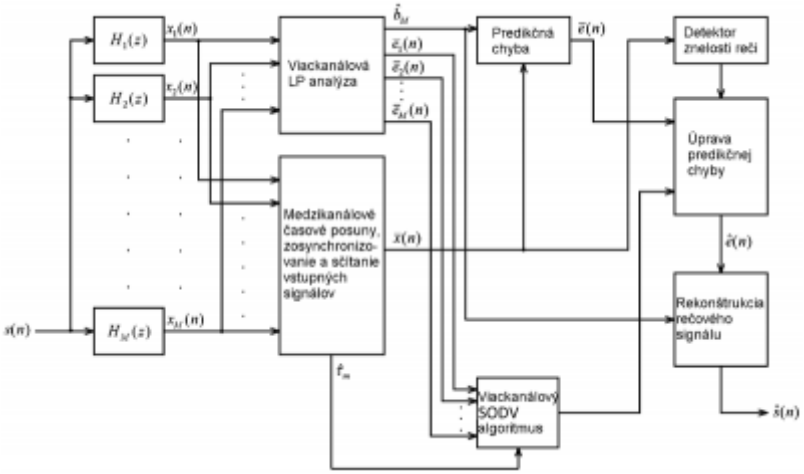
Obr. 4 Bloková schéma dereverberačného algoritmu založeného na LP

Na Obr. 4 je všeobecná bloková schéma dereverberačného algoritmu so spracovaním

predikčnej chyby M pozorovaných signálov. Lineárne predikčné koeficienty $\hat{b}$ musia byť odhadnuté dostatočne presne vzhľadom k čistému rečovému signálu, načo je predikčná chyba $\hat{e}(n)$ je upravená tak, aby rekonštruovaný signál $\hat{s}(n)$ bol odhadom čistého rečového signálu s(n). Teraz popíšeme algoritmus ÚSHC, ktorý kombinuje vybrané metódy formovania smerovej charakteristiky mikrofónového poľa, modifikácie rečového signálu úpravou predikčnej chyby a identifikácie vlastností prostredia naslepo s následnou ekvalizáciou [1].

4. Algoritmus úpravy a spriemerňovania hlasivkových cyklov

Predikčná chyba určuje charakter rekonštruovaného signálu jednak polohou a tvarom skutočnej excitačnej špičky, ale aj informáciou medzi dvoma hlasivkovými impulzami. Problémom u algoritmov s úpravou predikčnej chyby je, že modifikácia skutočných excitačných špičiek alebo priebehu predikčnej chyby medzi nimi spôsobuje skreslenie rekonštruovaného signálu, ktorý potom znie neprirodzene. Väčšina metód navyše do dereverberačného procesu nezahŕňa neznelé a tiché segmenty reči, ktoré sa ponechávajú dozvukom znehodnotené. Popisovaný algoritmus, ktorého bloková schéma je na Obr. 5, má ošetriť práve tieto problémy.



Obr. 5 Bloková schéma algoritmu s ÚSHC

4.1 Viackanálová LP analýza

Rečový signál s dozvukom je snímaný mikrofónovým poľom s M mikrofónmi ako $x_m(n)$n m=1,...,M. Predmetom 1. kroku je určenie LP koeficientov $\hat{b}_M$ z M kanálov, ktoré sú ekvivalentné koeficientom pre jeden kanál s čistým rečovým signálom. Experimenty ukázali, že toto je splnené pre LP koeficienty počítané po 20ms segmentoch ako

$$\hat{b}_M = \frac{1}{M}\sum_{m=1}^{M} b_{m,i},\ \ i = 1, 2, \ldots, P \qquad (7)$$

kde $b_{m,i}$ je i-tý predikčný koeficient v m-tom kanáli určený zo vstupného signálu $x_m(n)$ a P je rád LP analýzy. Tieto predikčné koeficienty budú neskôr využité k určeniu predikčnej chyby a k rekonštrukcii dereverberovanej reči.

4.2 Medzikanálové časové posuny a formovanie smerovej charakteristiky

Ku korektnému sčítaniu M kanálov do jedného potrebujeme poznať časové

oneskorenia jednotlivých kanálov oproti referenčnému. Tieto posuny budú dôležité aj pre ďalšie časti algoritmu. V [1] bola na ich výpočet navrhnutá metóda krížových korelácií popísaná v [4] ako pomerne jednoduchá a dostatočne presná pre mierny dozvuk. Nech $x_{ref}(n)$ je signál referenčného a $x_m(n)$ je signál porovnávaného kanálu. Odhad posunu signálu m-tého kanálu oproti referenčnému $\hat{\tau}_m$ je určený polohou maxima krížovej korelácie medzi týmito signálmi

$$\hat{\tau}_m = argmax_\tau r_{x_{ref}x_m}(\tau) \tag{8}$$

pričom $r_{x_{ref}x_m}(\tau)$ je spätná Fourierova transformácia krížovej korelácie Fourierovych obrazov porovnávaných signálov;

$$r_{x_{ref}x_m}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{X_{ref}(e^{j\omega})X_m^*(e^{j\omega})}{|X_{ref}(e^{j\omega})||X_m^*(e^{j\omega})|} e^{j\omega\tau} d\omega \tag{9}$$

kde X^* je číslo komplexne združené k X. Vypočíta sa M-1 posunov medzi referenčným signálom $x_1(n)$ a porovnávanými signálmi $x_2(n)$ až $x_M(n)$, tak ako v predchádzajúcom kroku po 20ms segmentoch. Vzájomným posunom a sčítaním signálov $x_1(n)$ až $x_M(n)$ do výsledného signálu $\bar{x}(n)$ podľa

$$\bar{x}(n) = \frac{1}{M}\sum_{m-1}^{M} x_m(n - \hat{\tau}_m) \tag{10}$$

nielenže dostaneme jeden signál s potlačeným nekorelovaným šumom, ale zároveň formujeme smerovú charakteristiku mikrofónového poľa. Táto je pritom vďaka počítaniu (8) po segmentoch dynamická, t. z. hlavný lalok smerovej charakteristiky sleduje rečníka, ktorý pritom môže byť v pohybe vzhľadom k mikrofónovému poľu. Formovanie smerovej charakteristiky samotné znižuje úroveň dozvuku, keďže odrazy prichádzajúce zo smerov mimo hlavný lalok smerovej charakteristiky sú utlmené.

4.3 Výpočet predikčnej chyby

Jednotnú predikčnú chybu $\bar{e}(n)$ dostaneme aplikáciou filtra so strednými LP koeficientmi $\hat{b}_M$ (7) na signál $\bar{x}(n)$ (10) z predchádzajúceho bodu:

$$\bar{e}(n) = \hat{b}_M^T \bar{x}(n) \tag{11}$$

Táto predikčná chyba bude neskôr upravovaná tak, aby pri LP syntéze došlo k rekonštrukcii skvalitneného rečového signálu.

4.4 Identifikácia skutočných hlasivkových impulzov

V predchádzajúcich krokoch sme sa dostali od signálov z M mikrofónov k jedinej množine LP koeficientov $\hat{b}_M$ a k jednej predikčnej chybe $\bar{e}(n)$. Táto predikčná chyba obsahuje skutočné excitačné špičky aj ich kópie vyvolané dozvukom. Aby bolo možné dozvukové špičky eliminovať, je nutné presne identifikovať pravé excitačné špičky odpovedajúce hlasivkovým impulzom. V [5] bol popísaný algoritmus, ktorý má byť vhodný na tento účel a na hľadanie hlasivkových impulzov využíva skupinové oneskorenie a dynamický výber (SODV) z nájdených kandidátov. Pre náš účel sme tento algoritmus modifikovali na viackanálový, t. z. algoritmus je aplikovaný priamo na

vstupné rečové signály, zvlášť pre každý kanál. Výstupom je množina možných hlasivkových impulzov (kandidátov) pre každý kanál, ktoré sa posunú v čase podľa $\hat{\tau}_m$ (8).

Výsledkom je jedna množina kandidátov odpovedajúca signálu $\bar{x}(n)$. Táto obsahuje takmer všetky skutočné hlasivkové impulzy, ale aj značnú časť chybných detekcií. Výber najpravdepodobnejších skutočných hlasivkových impulzov z množiny všetkých kandidátov je realizovaný dynamickým programovaním. Nakoniec sa vstupná predikčná chyba rozdelí na segmenty zodpovedajúce hlasivkovým cyklom, pričom odhady skutočných excitačných špičiek sú na hraniciach segmentov.

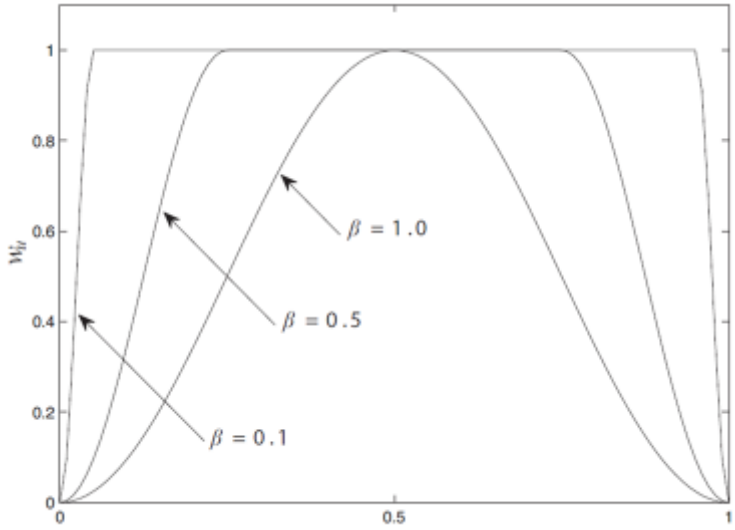
4.5 Detekcia znelých a neznelých segmentov rečového signálu

Presná detekcia znelých a neznelých častí reči je pre dereverberačné algoritmy založené na LP kritická. Samotný algoritmus SODV nie je dostačujúci, pretože hoci má dobré výsledky pri znelej reči, v neznelých a tichých segmentoch deteguje neexistujúce hlasivkové impulzy. Zvolili sme preto detektor popísaný v [6], u ktorého bola prezentovaná prijateľná presnosť pri minimálnej výpočtovej a implementačnej náročnosti. Tento detektor je nutné najprv natrénovať manuálnym segmentovaním trénovacích nahrávok reči, po prvotnom natrénovaní sa dokáže sám adaptovať na zmenu rečníka či prostredia.

Vstupný signál detektora, $\bar{x}(n)$ z (10), je najprv filtrovaný hornopriepustným filtrom s deliacou frekvenciou 200Hz. Následne je signál delený na mikrosegmenty s dĺžkou 10ms bez prekrývania. V týchto segmentoch sa rozhodne, či ide o znelú/neznelú reč alebo ticho podľa počtu prechodov signálu nulou, podľa energie signálu, autokorelácie s posunom jednej vzorky a LP koeficientov. Na základe detegovaných segmentov reči sú priradené aj hlasivkovým cyklom z bodu 4.4 odpovedajúce triedy.

4.6 Spriemerňovanie hlasivkových cyklov

Tento krok je jadrom celého ÚSHC algoritmu a predstavuje vlastnú dereverberáciu. Súčtový signál $\bar{x}(n)$ je rozdelený do hlasivkových cyklov $\bar{e}(n_l),\ l = 1, 2, \ldots$, podľa výsledkov SODV algoritmu z bodu 4.4. Aby došlo k zanechaniu skutočného hlasivkového impulzu na hraniciach hlasivkového cyklu a k útlmu dozvukových impulzov v jeho okolí, na každý hlasivkový cyklus je aplikovaná vážiaca funkcia. Reálny hlasivkový impulz (aj v čistom rečovom signáli) však nie je skutočným impulzom, má určité časové trvanie, a navyše je identifikovaný s neistotou asi 1ms a pri procese by nemal byť modifikovaný. Vyhovujúcou vážiacou funkciou je Tukeyovo okno, znázornené na Obr. 6. Činiteľ tvaru $0 \leq \beta \leq 1$ je laditeľný parameter, ktorým môžeme regulovať, aká časť hlasivkového cyklu bude zahrnutá v spriemerňovaní a ako budú tvarované excitačné špičky. Odstránenie dozvuku mimo hlasivkový impulz je realizované spriemerňovaním predikčnej chyby viacerých susedných hlasivkových cyklov.



Obr. 6 Vážiaca funkcia pre rôzne činitele tvaru

Konečné vyjadrenie l-tého skvalitneného hlasivkového cyklu v znelých segmentoch reči je

$$\hat{e}(n_l) = (I - W)\bar{e}(n_l) + \frac{1}{2\chi+1}\sum_{i=-\chi}^{\chi} W\bar{e}(n_{l+i}) \tag{12}$$

kde W je diagonálna matica s vážiacou funkciou w_u

$$W = diag\{w_0, w_1, \ldots, w_{L-1}\} \tag{13}$$

a I je jednotková matica. Skvalitnený hlasivkový cyklus teda vznikne spriemernením hlasivkového cyklu $\bar{e}(n_l)$ a jeho χ susedných cyklov vynásobených vážiacou funkciou w_u (skvalitnenie informácie medzi hlasivkovými impulzami), a prenásobením $\bar{e}(n_l)$ inverznou vážiacou funkciou 1-w_u (vytvarovanie samotných hlasivkových impulzov). Dereverberácia neznelých a tichých segmentov reči sa rieši časovo variantným dekonvolučným filtrom $\hat{g}(n_l)$, aktualizovaným iteračne iba počas znelých segmentov ako

$$\hat{g}(n_l) = \gamma\hat{g}(n_{l-1}) + (1-\gamma)\hat{g} \tag{14}$$

pričom $0 \leq \gamma \leq 1$ je pamätací faktor s typickými hodnotami 0,1 až 0,3 a iteračný postup začína s $\hat{g}(0) = [1, 0, 0, \ldots]^T$. Koeficienty dekonvolučného filtra $\hat{g}$ sa dajú nájsť ako

$$\hat{g} = R_{\bar{e}\bar{e}}^{-1} r_{\bar{e}\hat{e}} \tag{15}$$

$R_{\bar{e}\bar{e}}$ je autokorelačná matica $r_{\bar{e}\hat{e}}$ a je vektor krížovej korelácie medzi $\bar{e}(n_l)$ a $\hat{e}(n_l)$. Vyjadrenie l-tého dereverberovaného hlasivkového cyklu v neznelých a tichých segmentoch reči je potom

$$\hat{e}(n_l) = \hat{g}^T(n_l)\bar{e}(n_l) \tag{16}$$

Rovnice (12) až (16) sú exaktné pre periodické hlasivkové impulzy, a tým pre hlasivkové cykly s rovnakou dĺžkou. Reálne sú hlasivkové impulzy iba kváziperiodické

a cykly sa líšia dĺžkou o niekoľko vzoriek. Toto bolo pri implementácii ošetrené symetrickým doplnením nulami, resp. orezaním spriemerňovaných hlasivkových cyklov.

4.7 Rekonštrukcia skvalitneného rečového signálu

Odhad čistého rečového signálu získame aplikáciou inverzného filtra s LP koeficientmi $\hat{b}_m$ (7) z bodu 4.1 na upravenú predikčnú chybu $\hat{e}(n)$ (16) z bodu 4.6:

$$\hat{s}(n) = [b_M^{-1}]^T \hat{e}(n) \tag{17}$$

5. Evaluácia výsledkov

Popísaný algoritmus a pomocný trénovací program boli implementované v Matlabe. Na vyhodnotenie úspešnosti algoritmov sme použili voľne dostupnú audio databázu nahranú Tomom Sullivanom na univerzite Carnegie Mellon. Všetky nahrávky boli navzorkované s frekvenciou 16kHz a s 16-bitovým lineárnym kvantovaním. Každá nahrávka obsahuje signály 15-prvkového mikrofónneho poľa a referenčný čistý rečový signál, snímaný mikrofónom v blízkosti úst rečníka (náhlavná súprava). Signál snímaný náhlavným mikrofónom bol pri vyhodnocovaní výsledkov použitý ako referenčný a považovaný za čistý.

Mikrofónne pole obsahuje tri 7-prvkové subpolia s ekvidištančne rozmiestnenými mikrofónmi, líšiace sa ich vzájomnou vzdialenosťou. Každá nahrávka je potom ešte charakterizovaná prostredím a vzdialenosťou medzi rečníkom a prostredným mikrofónom, pričom rečník sa nachádza na osi mikrofónneho poľa. Vo všetkých nahrávkach hovoria mužskí rečníci a ich dĺžka je asi 3s. Vybrané nahrávky boli spracované popísaným algoritmom a jeho úspešnosť bola vyhodnotená percepčnými objektívnymi metódami.

5.1 Miera PESQ

Je to objektívna miera používaná v telefónii. Čistý a znehodnotený signál sú najprv normalizované na rovnakú hlasitosť a filtrované; odozva filtra je podobná štandardnému telefónnemu slúchadlu. Signály sa potom časovo zosynchronizujú a transformujú na spektrum hlasitosti. Rozdiely v hlasitosti sa spriemernia cez čas a frekvenciu tak, aby výsledok predikoval subjektívne hodnotenie. Výsledok je z intervalu 1,0 až 4,5, kde vyššia hodnota znamená lepšie hodnotenie [7]. Prehľad hodnotení pre rôzne nahrávky podľa miery PESQ je v Tab. 1.

Tab. 1 Hodnotenia podľa miery PESQ

Prostredie	**Vzdialenosť medzi mikrofónmi (cm)**	**Skóre PESQ**	
		Pred spracovaním	**Po spracovaní**
Konferenčná miestnosť, rečník 1m od mikrofónneho poľa	4	2,17	2,18
	8	2,11	2,26

Konferenčná miestnosť, rečník 3m od mikrofónneho poľa	4	1,77	1,94
	8	1,7	1,99
Hlučná počítačová pracovňa, rečník 1m od mikrofónneho poľa	4	2,2	2,25
	8	2,2	2,25

Vo všetkých prípadoch došlo k miernemu zlepšeniu, najvýraznejšie v konferenčnej miestnosti pri vzdialenosti rečníka od stredu mikrofónneho poľa 3m – prípad s najväčším znehodnotením dozvukom. Výsledky indikujú, že algoritmus je úspešnejší pri väčšom rozstupe mikrofónov.

5.2 Pomer SNRseg

V čase segmentovaný pomer signál/šum, je vypočítaný ako

$$SNRseg = \frac{10}{M}\sum_{m=0}^{M-1} log\frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1}(x(n)-\hat{x}(n))^2} \quad (18)$$

kde x(n) je čistý signál, $\bar{x}(n)$ je posudzovaný signál, N je dĺžka rámca (30ms) a M je počet rámcov [7]. Výsledky sú v Tab. 2.

Tab. 2 Hodnotenia podľa pomeru SNRseg.

Prostredie	**Vzdialenosť medzi mikrofónmi (cm)**	**Pomer SNRseg (dB)**	
		Pred spracovaním	**Po spracovaní**
Konferenčná miestnosť, rečník 1m od mikrofónneho poľa	4	-7,89	-7,78
	8	-8,3	-8,06
Konferenčná miestnosť, rečník 3m od mikrofónneho poľa	4	-6,5	-5,3
	8	-6,08	-5,52
Hlučná počítačová pracovňa, rečník 1m od mikrofónneho poľa	4	-8,72	-5,85
	8	-8,53	-5,75

Vo všetkých prostrediach došlo k zlepšeniu odstupu signálu od šumu, avšak nevieme posúdiť, či došlo naozaj k redukcii šumu alebo dozvuku. Zlepšenie SNRseg v konferenčnej miestnosti je 0,1 až 1,2 dB, v hlučnej počítačovej pracovni až 2,9dB. Algoritmus nie je primárne určený na redukciu šumu, výsledky ale naznačujú, že pri vysokých úrovniach šumu je tento potláčaný.

5.3 Vzdialenosť WSS

Predstavuje rozdiel v zmenách spektrálnej obálky percepčne vážený po frekvenčných pásmach. Vzdialenosť WSS je definovaná ako

$$d_{WSS} = \frac{1}{M}\sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} W_{WSS}(j,m)(S_c(j,m)-S_P(j,m))^2}{\sum_{j=1}^{K} W_{WSS}} \quad (19)$$

s K=25, počtom segmentov M a váhami $W_{WSS}(j,m)$. $S_C(j,m)$ a $S_P(j,m)$ sú spektrálne

zmeny v j-tom frekvenčnom pásme pre čistý a posudzovaný signál [7]. Touto mierou by principiálne bolo možné veľmi účinne zhodnotiť úroveň dozvuku, ak by v signáli s dozvukom nebol aditívny šum. Reálne je posudzovaný príspevok ako dozvuku, tak aj šumu. Prehľad výsledkov je v Tab. 3.

Tab. 3 Hodnotenia podľa vzdialenosti WSS

Prostredie	Vzdialenosť medzi mikrofónmi (cm)	Vzdialenosť WSS	
		Pred spracovaním	Po spracovaní
Konferenčná miestnosť, rečník 1m od mikrofónneho poľa	4	61,5	46,92
	8	55,27	45,92
Konferenčná miestnosť, rečník 3m od mikrofónneho poľa	4	69,54	58
	8	63,47	57,89
Hlučná počítačová pracovňa, rečník 1m od mikrofónneho poľa	4	75,47	51,15
	8	76	53,03

Podľa tejto miery sú lepšie výsledky dosiahnuté s menším rozostupom mikrofónov a pri menšej vzdialenosti rečníka od mikrofónneho poľa. Najvyššia účinnosť algoritmu vychádza v hlučnej počítačovej pracovni, vzdialenosť WSS je teda pravdepodobne pre evaluáciu dereverberačných algoritmov priveľmi citlivá na šum.

6. Záver

Popísaný algoritmus preukázateľne potláča dozvuk v rečových signáloch znehodnotených dozvukom a šumom. Je odolný voči vysokým hladinám šumu, ktoré dokonca do určitej miery potláča. Výpočtovo je pomerne nenáročný, a to aj napriek využitiu dynamického programovania (potrebný počet výpočtových operácií je závislý od charakteru signálu). Daňou za to je, že ani teoreticky neumožňuje dokonalú dereverberáciu a je vhodný skôr pre prostredia s výraznejším dozvukom. Veľkou výhodou ÚSHC algoritmu je, že pri jeho vhodnom nastavení nedochádza k vnímateľnému skresleniu rečového signálu ani pri výraznom dozvuku, a môže byť vhodnou metódou predspracovania signálu pre iné skvalitňujúce algoritmy.

Poďakovanie

Táto publikácia vznikla vďaka podpore v rámci operačného programu Výskum a vývoj pre projekt „(Centrum informačných a komunikačných technológií pre znalostné systémy) (kód ITMS:26220120020), spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja“.

Literatúra

1. NAYLOR, Patrick A. – GAUBITCH, Nikolay D.: Speech dereverberation. London: Springer, 2010. 388 s. ISBN 978-1-84996-056-4.
2. SMETANA, Ctirad: Ozvučování. Praha: SNTL, 1987. 216 s.
3. JUHÁR, Jozef: Rečové technológie. Košice: Equilibria, s.r.o., 2011. 517 s. ISBN 978-8--89284-75-7.

4. KNAPP, Charles H. – CARTER, Clifford G.: The generalized correlation method for estimation of time delay. In: IEEE Transactions on acoustics, speech, and signal processing. roč. 24, č. 4 (1976), s. 320-327.
5. NAYLOR, Patrick A. – KOUNOUDES, Anastasis – GUDNASON, Jon – BROOKES, Mike: Estimation of glottal closure instants in voiced speech using the DYPSA algorithm. In: IEEE Transactions on audio, speech, and language processing. roč. 15, č. 1 (2007), s. 34-43.
6. ATAL, Bishnu S. – RABINER, Lawrence R.: A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition. In: IEEE Transactions on acoustics, speech, and signal processing. roč. 24, č. 3 (1976), s. 201-212.
7. HU, Y. – Loizou, P.: Evaluation of objective quality measures for speech enhancement. In: IEEE Transactions on speech and audio processing. roč. 16, č. 1 (2008), s. 229-238.

Katedra elektroniky a multimediálnych telekomunikácií, Fakulta elektrotechniky a informatiky, Technická univerzita v Košiciach