

SOLUȚII ROBUSTE PENTRU SUPRIMAREA ADAPTIVĂ A INTERFERENȚELOR

Proiect IDEI Cod 1027 - Faza Unică 2009

Raport privind impactul cercetării în domeniul compensatoarelor adaptive de ecou de linie sau de ecou acustic și analiza performanțelor. (Sinteza)

Similar algoritmului NLMS, performanțele algoritmului APA clasic (din punctul de vedere al vitezei de convergență, stabilității și dezadaptării) sunt determinate de mărimea pasului de reactualizare al algoritmului. Acest parametru trebuie ales pe baza compromisului dintre viteza mare de convergență și dezadaptarea scăzută. Pentru a satisface aceste cerințe contradictorii au fost dezvoltate o serie de algoritmi NLMS, respectiv APA, cu pas variabil (VSS-NLMS, respectiv VSS-APA). Totuși, majoritatea algoritmilor de tip VSS-APA existenți necesită setarea unor parametri care nu sunt disponibili a priori sau care trebuie estimați (de exemplu: puterea zgomotului de fond). În plus, chiar dacă algoritmi de tip VSS-APA ar oferi o soluție adecvată având în vedere cerințele de performanță enunțate anterior, rămâne de observat comportamentul acestora pe durata vorbirii simultane. Procedura standard constă în folosirea unui detector de vorbire simultană (DTD = “*Double Talk Detector*”) pentru a încetini sau a opri complet procesul de adaptare pe durata fenomenului. Totuși apare o întârziere în procesul de decizie a unui DTD. Pe durata unei perioade scurte de timp, câteva eșantioane de amplitudine mare pot trece nedectate și pot perturba puternic estimatul căii de ecou. În consecință, este de dorit ca robustețea algoritmului adaptiv să fie îmbunătățită astfel încât acesta să suporte o perioadă determinată de timp fenomenul de vorbire simultană a interlocutorilor fără să diveargă.

Algoritmul VSS-APA propus [1] este proiectat special pentru aplicațiile de compensare a ecoului. Prin folosirea pasului variabil de reactualizare se ia în considerare faptul că microfonul captează atât zgomotul de fond cât și o secvență de vorbire; aceste semnale trebuie recuperate din semnalul eroare al filtrului adaptiv. Relația rezultată depinde doar de semnalele disponibile în cadrul configurației de compensare a ecoului: semnalul captat de microfon și semnalul de la ieșirea filtrului adaptiv. Astfel, informațiile a priori referitoare la mediu nu mai sunt necesare în acest caz. Din acest punct de vedere algoritmul poate fi numit “non-parametric”. Un alt beneficiu îl reprezintă robustețea algoritmului propus în fața variațiilor semnalului “near-end”. Performanțele din timpul vorbirii simultane pot fi îmbunătățite prin utilizarea unui simplu mecanism Geigel DTD. Acest aspect favorizează utilizarea algoritmului propus în cazul aplicațiilor practice de compensare a ecoului.

În continuare va fi realizată o descriere succintă a algoritmului VSS-APA propus. În Fig. 1 este prezentată schema generală folosită în aplicațiile de compensare a ecoului. Scopul configurației este identificarea sistemului necunoscut (calea de ecou), prin utilizarea unui filtru adaptiv. Ambele sisteme au răspuns la impuls finit, definit de vectorii $\mathbf{h} = [h_0 \ h_1 \ \dots \ h_{N-1}]^T$ și $\hat{\mathbf{h}}(n) = [\hat{h}_0(n) \ \hat{h}_1(n) \ \dots \ \hat{h}_{L-1}(n)]^T$, unde notația T semnifică operația de transpunere iar n reprezintă momentul de timp; N este lungimea căii de ecou, iar L reprezintă lungimea filtrului adaptiv.

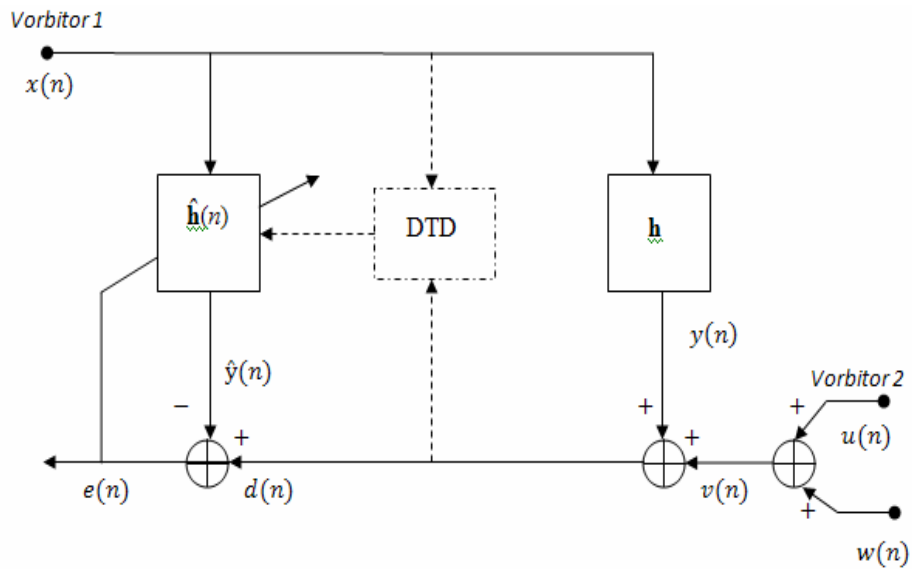


Fig. 1. Configurație adaptivă de compensare a ecoului.

Relațiile ce definesc algoritmul APA clasic sunt următoarele:

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{X}^T(n) \hat{\mathbf{h}}(n-1) \quad (1)$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu \mathbf{X}(n) [\mathbf{X}^T(n) \mathbf{X}(n)]^{-1} \mathbf{e}(n) \quad (2)$$

unde $\mathbf{d}(n) = [d(n), d(n-1), \dots, d(n-p+1)]^T$ este vectorul semnalului dorit de lungime p (p fiind ordinul de proiecție). Matricea $\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \dots, \mathbf{x}(n-p+1)]$ este matricea semnalului de intrare, unde $\mathbf{x}(n-l) = [x(n-l), x(n-l-1), \dots, x(n-l-L+1)]^T$ (cu $l=0,1,\dots,p-1$) reprezintă vectorii semnalului de intrare. Constanta μ reprezintă pasul de reactualizare al coeficienților în cadrul algoritmului. Relația (2) se poate rescrie astfel:

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{X}(n) [\mathbf{X}^T(n) \mathbf{X}(n)]^{-1} \boldsymbol{\mu}(n) \mathbf{e}(n) \quad (3)$$

unde

$$\boldsymbol{\mu}(n) = \text{diag}\{\mu_0(n), \mu_1(n), \dots, \mu_{p-1}(n)\} \quad (4)$$

este o matrice diagonală de dimensiune $p \times p$. În mod evident, (2) se obține în cazul particular $\mu_0(n) = \mu_1(n) = \dots = \mu_{p-1}(n) = \mu$. Folosind coeficienții filtrului adaptiv la momentul n , vectorul eroare a posteriori poate fi definit astfel

$$\boldsymbol{\varepsilon}(n) = \mathbf{d}(n) - \mathbf{X}^T(n) \hat{\mathbf{h}}(n) \quad (5)$$

Observăm ca relația (1) definește vectorul eroare a priori $\mathbf{e}(n)$. Înlocuind în (5) cu relația (3) și ținând cont de (1), rezultă

$$\boldsymbol{\varepsilon}(n) = [\mathbf{I}_p - \boldsymbol{\mu}(n)] \mathbf{e}(n) \quad (6)$$

unde \mathbf{I}_p reprezintă matricea identitate de dimensiune $p \times p$.

În conformitate cu ideea de bază a APA, se poate impune anularea unui număr de p erori a posteriori, adică $\boldsymbol{\varepsilon}(n) = \mathbf{0}_{p \times 1}$, unde $\mathbf{0}_{p \times 1}$ reprezintă un vector coloană cu toate cele p elemente nule. Presupunând ca $\mathbf{e}(n) \neq \mathbf{0}_{p \times 1}$, va rezulta din (6) că $\boldsymbol{\mu}(n) = \mathbf{I}_p$. Aceasta corespunde cu relația (2) de modificare a coeficienților algoritmului APA clasic cu pasul $\mu = 1$. În cazul absenței semnalului local, $\mathbf{v}(n) = \mathbf{0}$, schema din Fig. 1 se reduce la configurația ideală a “identificării unui sistem”. În acest caz alegerea parametrului $\mu = 1$ are sens, deoarece oferă cele mai bune rezultate.

Se poate observa faptul că schema din Fig. 1 poate fi interpretată ca o combinație între două clase de configurații adaptive. În primul rând, reprezintă o configurație pentru identificarea sistemului, deoarece scopul este identificarea unui sistem necunoscut (calea de ecou) cu semnalul de ieșirea lui perturbat de un semnal aparent nedorit (semnalul vocal al interlocutorului local). În același timp, ea poate fi privită din punctul de vedere al suprimării interferenței, ținând recuperarea semnalului util (semnalul vocal local) perturbat de ecoul nedorit. În consecință, semnalul util trebuie recuperat din semnalul eroare al filtrului adaptiv. Deoarece existența semnalului local nu poate fi neglijată în aplicație, trebuie impusă o nouă condiție: $\boldsymbol{\varepsilon}(n) = \mathbf{v}(n)$, unde $\mathbf{v}(n) = [v(n), v(n-1), \dots, v(n-p+1)]^T$ este un vector coloană și reprezintă vectorul de lungime p a semnalului vocal local. Considerând relația (6) rezultă

$$\varepsilon_{l+1}(n) = [1 - \mu_l(n)] e_{l+1}(n) = v(n-l) \quad (7)$$

unde variabilele $\varepsilon_{l+1}(n)$ și $e_{l+1}(n)$ sunt elementele cu indexul $(l+1)$ ai vectorilor $\boldsymbol{\varepsilon}(n)$, respectiv $\mathbf{e}(n)$, cu $l = 0, 1, \dots, p-1$. Se dorește găsirea unei expresii pentru parametrul $\mu_l(n)$ astfel încât

$$E\{\varepsilon_{l+1}^2(n)\} = E\{v^2(n-l)\} \quad (8)$$

În relația (8), $E\{\bullet\}$ definește media statistică. Dacă vom ridica la pătrat relația (7) și vom lua în considerare mediile statistice, vom obține:

$$[1 - \mu_l(n)]^2 E\{e_{l+1}^2(n)\} = E\{v^2(n-l)\} \quad (9)$$

Prin rezolvarea ecuației pătratice definite de (9) se obțin două soluții pentru parametrul $\mu_l(n)$, și anume

$$\mu_l(n) = 1 \pm \sqrt{\frac{E\{v^2(n-l)\}}{E\{e_{l+1}^2(n)\}}} \quad (10)$$

Urmând un model de analiză realizat în literatură, în care se menționează că se preferă o valoare a pasului de reactualizare cuprinsă în intervalul (0;1) decât în intervalul (1;2) (deși ambele soluții sunt stabile, prima prezintă eroare pătratică medie mai stabilă în aceleași condiții de viteză de convergență), din cele două soluții se va alege

$$\mu_l(n) = 1 - \sqrt{\frac{E\{v^2(n-l)\}}{E\{e_{l+1}^2(n)\}}} \quad (11)$$

Din punctul de vedere practic este mai util ca relația (11) să fie rescrisă folosind estimatele puterilor celor două semnale

$$\mu_l(n) = 1 - \frac{\hat{\sigma}_v(n-l)}{\hat{\sigma}_{e_{l+1}}(n)} \quad (12)$$

Variabila prezentă la numitor în relația (12) poate fi calculată recursiv astfel

$$\hat{\sigma}_{e_{l+1}}^2(n) = \lambda \hat{\sigma}_{e_{l+1}}^2(n-1) + (1-\lambda) e_{l+1}^2(n) \quad (13)$$

unde λ reprezintă un factor de ponderare calculat astfel $\lambda = 1 - 1/(KL)$, cu $K > 1$; valoarea inițială este $\hat{\sigma}_{e_{l+1}}^2(0) = 0$. Estimarea variabilei $\hat{\sigma}_v(n-l)$ nu poate fi realizată în mod direct în cazul aplicațiilor de timp real precum cele de compensare a ecoului.

În continuare se va presupune $L = N$. În cazul single-talk semnalul local este format doar din zgomotul de fond, $w(n)$. Puterea sa poate fi estimată pe durata perioadelor de liniște și poate fi presupusă constantă. Astfel, relația (12) devine

$$\mu_l(n) = 1 - \frac{\hat{\sigma}_w}{\hat{\sigma}_{e_{l+1}}(n)} \quad (14)$$

Pentru ordinul de proiecție $p = 1$, se obține algoritmul non-parametric VSS-NLMS (NPVSS-NLMS) [2]. Totuși, zgomotul de fond poate varia în timp. Acest fapt conduce la necesitatea de a estima puterea acestuia periodic. În plus, atunci când zgomotul de fond variază în intervalul de timp dintre două estimări consecutive sau pe durata vorbirii simultane, noua

valoarea estimată nu va fi disponibilă imediat; prin urmare, până la următoarea estimare a puterii zgomotului, funcționarea algoritmului va fi perturbată.

În cazul vorbirii simultane (fenomenul “double-talk”), semnalul “near-end” $v(n)$ este compus din zgomotul de fond, $w(n)$, și semnalul vocal al interlocutorului local, $u(n)$. Putem scrie $v(n) = w(n) + u(n)$. Estimarea corectă a puterii acestui semnal compus este dificil de realizat din cauza caracterului nestaționar al semnalului vocal. În consecință, relația (12) este inutilă și prezența blocului DTD devine obligatorie pentru controlul procesului de adaptare în aceste perioade. Atunci când este prezent și semnalul vocal local în semnalul captat de microfon, estimatul puterii semnalului local poate fi exprimat ca în relația

$$\sigma_v^2(n) = \sigma_w^2(n) + \sigma_u^2(n), \quad (15)$$

unde $\sigma_u^2(n)$ reprezintă estimatul puterii semnalului vocal local. Se presupune că semnalul vocal și zgomotul de fond nu sunt corelate. Atunci $E\{d^2(n)\} = E\{y^2(n)\} + E\{v^2(n)\}$, rezultând că $E\{d^2(n)\} - E\{y^2(n)\} = E\{v^2(n)\}$. Presupunând că filtrul adaptiv a converș într-o anumită măsură se poate scrie că $E\{y^2(n)\} \cong E\{\hat{y}^2(n)\}$. Prin urmare, $E\{v^2(n)\} \cong E\{d^2(n)\} - E\{\hat{y}^2(n)\}$ și

$$\hat{\sigma}_v^2(n) \cong \hat{\sigma}_d^2(n) - \hat{\sigma}_y^2(n). \quad (16)$$

În mod corespunzător, membrul drept din relația (16) oferă un estimat al puterii semnalului local. Esențial este faptul că estimatul depinde doar de semnalele disponibile în cadrul aplicației de compensare a ecoului, semnalul captat de microfon, $d(n)$, și ieșirea filtrului adaptiv, $\hat{y}(n)$. Ținând cont de aspectele menționate anterior, relația (12) se poate rescrie astfel

$$\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n-l) - \hat{\sigma}_y^2(n-l)}}{\hat{\sigma}_{e_{l+1}}(n)} \quad (17)$$

pentru $l = 0, 1, \dots, p-1$. Comparativ cu (12), relația (17) este mai potrivită aplicațiilor reale.

Coeficienții filtrului adaptiv trebuie reactualizați folosind relația (3), parametrii necesari fiind calculați conform (17). Totuși, în practică expresia (3) de actualizare trebuie rescrisă astfel

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{X}(n) [\delta \mathbf{I}_p + \mathbf{X}^T(n) \mathbf{X}(n)]^{-1} \boldsymbol{\mu}(n) \mathbf{e}(n) \quad (18)$$

unde δ este o constantă pozitivă numit *factor de regularizare*.

Pentru un ordin de proiecție $p=1$, se obține algoritmul VSS-NLMS propus în [3]. Algoritmul VSS-NLMS s-a dovedit a fi robust în fața variațiilor semnalului local, în special în cazul vorbirii simultane (double-talk). Dezavantajul, comparativ cu algoritmul NLMS clasic, îl constituie reducerea vitezei de convergență. Este important de subliniat faptul că singurii parametri necesari funcționării algoritmului sunt cei disponibili în mod direct din

configurația filtrului adaptiv. Cu alte cuvinte, nu sunt necesare informații a priori asupra mediului acustic, ceea ce înseamnă ușurință în controlul algoritmului în practică.

În vederea testării performanțelor algoritmului studiat în aplicații de AEC reale, s-au efectuat o serie de simulări în precizie finită, pentru a determina modalitatea de lucru cea mai potrivită în vederea implementării pe FPGA. Rezultatele au fost prezentate în detaliu [4], în continuare fiind subliniate doar câteva elemente importante ale implementării.

Formula care dă pasul pentru algoritmul VSS-NLMS [3] este:

$$\mu(n) = \mu_{NLMS}(n) \quad \text{pentru } n \leq L$$

$$\mu(n) = \frac{1}{\delta + x_L^T(n)x_L(n)} \left[1 - \frac{\sqrt{\hat{\sigma}_d^2(n) - \hat{\sigma}_y^2(n)}}{\xi + \hat{\sigma}_e(n)} \right] \quad \text{pentru } n > L$$

Pentru aproximarea radicalului, ne vom folosi de următoarea proprietate: fiind dat un număr pozitiv, se cunoaște că șirul $x_n = \frac{1}{2}(x_{n-1} + \frac{a}{x_{n-1}})$ converge la \sqrt{a} . Această proprietate poate fi folosită pentru a calcula numărul \sqrt{a} pentru un a fixat.

De asemenea, o altă problemă apare în momentul în care estimatul varianței semnalului dorit este mai mic decât estimatul varianței semnalului de estimat al ecoului. În cazul reprezentării în precizie finită, acest lucru se traduce prin apariția unei depășiri la calculul radicalului. Pentru a evita această situație, vom considera că, atunci când se prezintă un astfel de inconvenient, valoarea pasului nu se modifică (rămâne egală cu valoarea pasului de la iterația anterioară).

Alegerea numărului de biți în aplicații este o problemă foarte delicată. Performanțele algoritmului depind în mod direct de reprezentarea cât mai fidelă a semnalelor în discuție; în același timp, implementarea pe FPGA presupune, în mod evident, resurse limitate, iar creșterea nejustificată a numărului de biți duce la folosirea în exces a resurselor și la scăderea vitezei de prelucrare a datelor. Trebuie astfel realizat un compromis în ceea ce privește alegerea numărului de biți necesari, care să nu altereze semnificativ performanțele algoritmului și să asigure în același timp o folosire economică a resurselor disponibile.

Figura 2 arată performanțele comparative ale algoritmului VSS-NLMS în cazul reprezentării pe 16, 14, 12 și respectiv 10 biți față de cazul reprezentării în precizie infinită. Se observă că performanțele algoritmului se degradează succesiv pe măsură ce micșorăm numărul de biți alocați reprezentării semnalelor. Reprezentările pe 16 și 14 biți sunt cele mai apropiate de reprezentarea în precizie infinită. Totuși, studiind, pe de o parte, evoluția erorii reziduale și nivelul acesteia, și, pe de altă parte, gama dinamică a semnalelor în cauză, ajungem la concluzia că o reprezentare mai potrivită ar fi cea în care pentru semnalul vocal

de intrare se alocă 12 biți, iar pentru coeficienții ce definesc răspunsul la impuls al ecoului 16 biți. Figura 3 arată performanțele algoritmului VSS-NLMS-UM în situația în care $x(n)$ și $v(n)$ sunt cuantizați pe 12 biți, iar coeficienții h pe 16 biți.

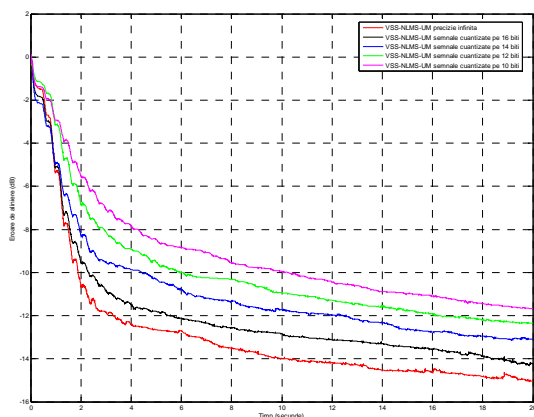


Fig. 2. Efectul cuantizării pe 16/14/12/10 biți a semnalelor de intrare (x , h , v) asupra performanțelor algoritmului VSS-NLMS.

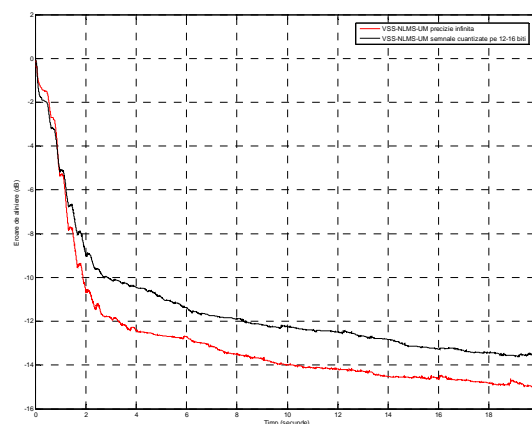


Fig. 3. Eroarea de aliniere obținută pentru algoritmul VSS-NLMS; semnalul vocal de intrare și zgomotul sunt cuantizați pe 12 biți, coeficienții h ai căii de ecou sunt cuantizați pe 16 biți.

Bibliografie

- [1] C. Paleologu, J. Benesty, S. Ciochină, “A Variable Step-Size Affine Projection Algorithm Designed for Acoustic Echo Cancellation”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1466-1478, Nov. 2008.
- [2] J. Benesty, H. Rey, L. Rey Vega, and S. Tressens, “A nonparametric VSS NLMS algorithm,” *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.
- [3] C. Paleologu, S. Ciochină, J. Benesty, “Variable Step-Size NLMS Algorithm for Under-Modeling Acoustic Echo Cancellation”, *IEEE Signal Processing Letters*, vol. 15, pp. 5-8, 2008.
- [4] C. Anghel, C. Paleologu, J. Benesty, and S. Ciochină, “FPGA Implementation of an Acoustic Echo Canceller Using a VSS-NLMS Algorithm”, *Proc. IEEE International Symposium on Signals, Circuits and Systems (ISSCS)*, Iași, România, 2009, pp. 369-372.

DIRECTOR PROIECT,
Prof. dr. ing. Silviu CIOCHINĂ