Optymalizacja procesu redukcji szumu w kontrolerze akustycznego echa i szumu.

Beghdad Ayad, Gérard Faucon, i Régine Le Bouquin – Jeannès

Streszczenie

W urządzeniach głośnomówiących transmitowany sygnał mowy jest zniekształcany przez dookolny szum i echo akustyczne. Należy więc dokonać przetworzenia sygnału przed jego transmisją, aby zredukować te zniekształcenia. Klasycznymi rozwiązaniami są kaskadowe struktury, w których system usuwania echa poprzedza lub występuje za systemem redukcji szumu. Ostatnio zaproponowaliśmy nowy, połączony system, w którym pre-przetwarzanie szumu pozwala na poprawienie wydajności systemu usuwania echa. Pre-przetwarzanie redukuje szum, ale zniekształca oryginalne echo. Artykuł ten przedstawia optymalizację pre-przetwarzania. Przedstawione są rezultaty w postaci wzmocnienia i Echo Return Loss Enchancment.

1. Wprowadzenie.

W niektórych aplikacjach, takich jak zestawy telekonferencyjne czy telefoniczne zestawy głośnomówiące, transmitowany sygnał mowy jest zakłócany przez występujący w środowisku szum i echo spowodowane sprzężeniem między głośnikiem a mikrofonem. Rozproszenie głośnomówiących zestawów komunikacyjnych wywołuje konieczność zapewnienia użytkownikom komfortu użytkowania. Tak więc w celu uzyskania wysokiej jakości transmitowanego sygnału mowy, należy rozwiązać oba problemy (szumu i echa). Chociaż poświęcono wiele uwagi osobno zagadnieniom usuwania echa i redukcji szumu, to stosunkowo niewiele badań dotyczyło połączonych systemów, dokonujących obu operacji. Naszym celem jest optymalizacja takiej połączonej struktury tak, aby uzyskać sygnał mowy jedynie nieznacznie zniekształcony, o niskim poziomie echa i szumu.

Sygnał przechwytywany przez mikrofon x(t) składa się z sygnału mowy s(t), echa e(t) oraz szumu n(t), natomiast głośniki emitują sygnał z(t) skorelowany z e(t). Optymalna, w sensie minimalizacji błędu średniokwadratowego, struktura do przetwarzania sygnałów x(t) i z(t) jest bardzo prosta do wyprowadzenia. Proces jej wyprowadzania składa się z dwóch etapów. W pierwszym z nich estymujemy echo przez zastosowanie filtracji sygnału z(t). Funkcja przenoszenia filtru dana jest wzorem

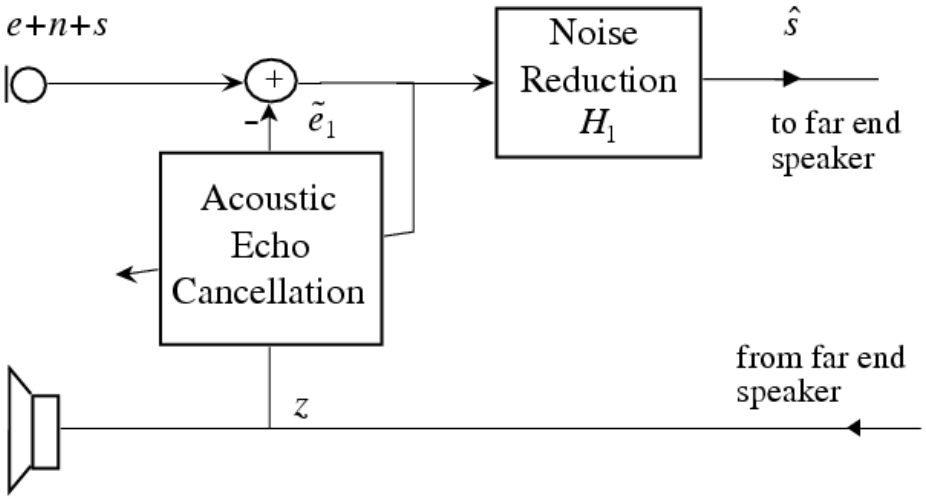
$$\frac{\gamma_{xz}(f)}{\gamma_{zz}(f)}$$

gdzie $\gamma_{xz}(f)$ to skrośna widmowa gęstość mocy między sygnałami x i z, a $\gamma_{zz}(f)$ to widmowa gęstość mocy sygnału z. Sygnał wychodzący z filtru jest odejmowany od sygnały przechwytywanego przez mikrofon. W przypadku idealnego układu usuwającego echo, sygnał mowy i szum są transmitowane bez zmian, natomiast echo jest całkowicie usuwane. W drugim etapie następuje redukcja szumu przez filtr Wienera, którego wzmocnienie dane jest wzorem

$$\frac{\gamma_{ss}(f)}{\gamma_{ss}(f)+\gamma_{nn}(f)}$$

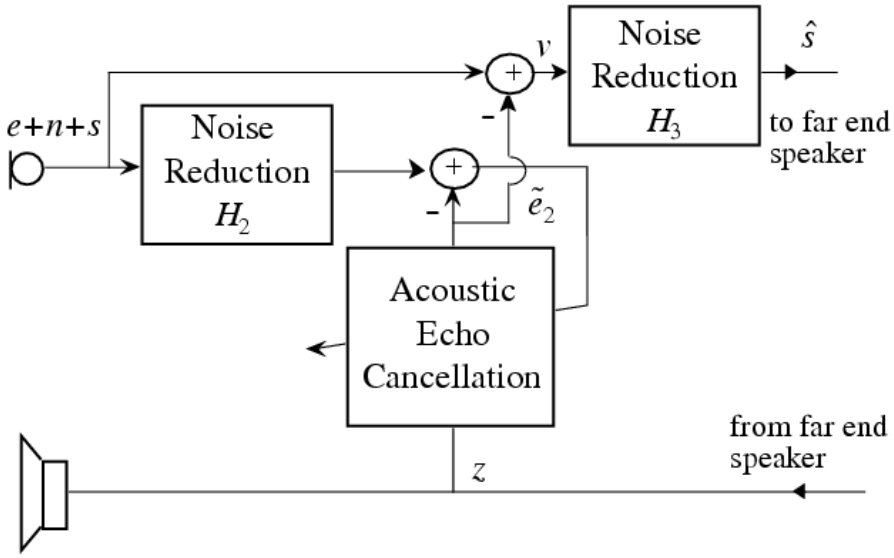
Tak więc optymalna struktura składa się z dwóch kaskadowo połączonych optymalnych filtrów, przy czym układ usuwania echa (AEC – Acoustic Echo Cancellation) poprzedza układ redukcji szumu (NR - Noise Reduction). Struktura taka nazywana jest AEC+NR (Rysunek 1.).

W praktyce system AEC jest systemem adaptatywnym. Współczynniki filtru AEC są zniekształcane przez wszechobecny szum otoczenia i okazuje się że skomplikowanym jest zatrzymanie procesu adaptacji w momencie pojawienia się sygnału mowy. W celu redukcji wpływu szumu na system AEC, można zamienić system AEC i system redukcji szumu miejscami tak, że adaptacja może zostać zatrzymana w trybie równoległego nadawania (Double Talk – DT; obecny zarówno sygnał mowy jak i echo). Niestety, zakłócający szum jest wtedy redukowany w mniejszym stopniu, a implementacja systemu odbiega od optymalnej struktury. Jednak mimo to echo estymowane przez układ AEC jest bliższe oryginalnemu echu, gdy system usuwania szumów poprzedza filtr AEC. W [3] eksperymenty dowiodły, że pomimo zniekształcenia echa przez system redukcji szumów, korzystnie jest najpierw przeprowadzić



Rysunek 1: Struktura AEC+NR

redukcję szumów aby otrzymać dokładniejszą estymatę echa. Tak więc zaproponowano [3,4] nową strukturę, nazywaną AEC+2NR (Rysunek 2.).



Rysunek 2: Struktura AEC+2NR

Wpływ szumu na system AEC jest redukowany przez zastosowanie filtru redukującego szum H_2 w mikrofonie. Następnie wykonywane jest usuwanie echa akustycznego AEC. Echo e_2 estymowane przez system AEC jest odejmowane od sygnału odbieranego przez mikrofon x(t) aby otrzymać sygnał v(t) = s(t) + n(t) + e(t) - e_2(t). Następnie stosuje się na sygnale v(t) drugi filtr redukcji szumów w celu uzyskania ostatecznej estymaty. Tym sposobem adaptacja AEC może zostać zatrzymana w trybie równoległego nadawania (DT mode) a v(t) zawiera niezmieniony sygnał mowy.

2. Systemy AEC i NR.

System usuwania echa akustycznego realizuje algorytm uogólnionego wielo-opóźnieniowego filtru (Generalized Multi-Delay Filter, GMDF). Bazuje on na blokowej, adaptacyjnej procedurze filtrowania w dziedzinie częstotliwości. Dwie różnice między nim a standardowym schematem to: a) podział odpowiedzi impulsowej na przedziały, co pozwala na kontrolę całkowitego opóźnienia przetwarzania oraz b) wprowadzenie parametru kontrolującego pokrywanie się kolejnych bloków wejściowych w celu modyfikacji tempa aktualizacji współczynników filtru.

Algorytm redukcji szumów jest wyprowadzony z estymatora minimalno średniokwadratowego błędu z krótko czasową amplitudą widmową, zaproponowanego przez Ephraima i Malaha [6]. Bazuje on na modelowaniu składników widmowych sygnałów mowy i szumu jako niezależnych gaussowskich zmiennych

losowych. Algorytm ten użyty jako pre-przetwarzania (filtr H_2) w strukturze AEC+2NR zostanie zoptymalizowany i poniżej przedstawiamy bardziej szczegółowy opis tej techniki, aby zrozumieć w którym miejscu pojawiają się optymalizacje. Niech Y(f) będzie widmem sygnału wejściowego systemu redukcji szumów y(t), który złożony jest z sygnału w(t) i szumu n(t). Estymata sygnału dana jest wzorem

$$W(f)=G_1(f)\cdot G_2(f)\cdot Y(f)$$

gdzie $G_1(f)$ jest filtrem Wienerowskim a $G_2(f)$ reprezentuje funkcję wzmocnienia biorącą pod uwagę niepewność sygnału mowy [7]. Estymator ten zależy od wartości *a priori* stosunku sygnał – szum (SNR), R_{prio}, wartości *a posteriori* SNR, R_{post}, zdefiniowanych odpowiednio jako

$$R_{prio}(f)=\frac{E[|W(f)|^2]}{E[|N(f)|^2]} \qquad R_{post}(f)=\frac{|Y(f)|^2}{E[|N(f)|^2]}$$

oraz prawdopodobieństwa nieobecności sygnału q(f). $E[|N(f)|^2]$ jest mocą szumów uzyskaną w czasie przerw w mówieniu. Ephraim i Malah zaproponowali aby estymować wartość a priori stosunku sygnał – szum na podstawie podejścia decyzyjnego:

$$R_{prio}(f,n)=\lambda\,\frac{A^2(f,n-1)}{E[|N(f)|^2]}+(1-\lambda)Q(R_{post}(f,n)-1)$$

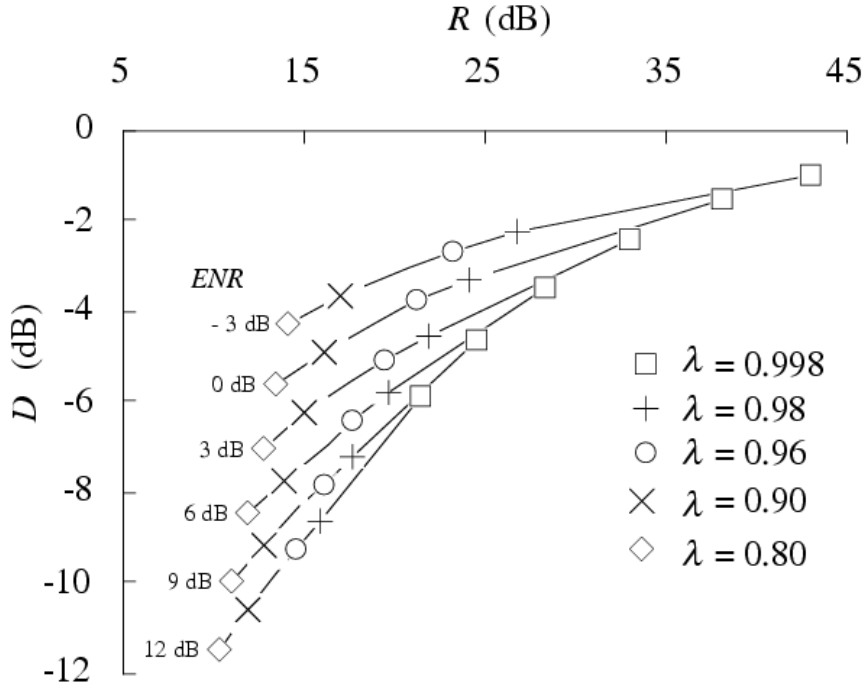
gdzie n jest numerem aktualnego bloku, A(f, n -1) jest amplitudą sygnału estymowanego w bloku (n – 1), λ jest współczynnikiem wagowym, Q(u) jest operatorem zdefiniowanym przez Max(u, 0). $R_{post}(f, n)$ jest otrzymywane bezpośrednio jako stosunek kwadratu wielkości sygnału w bloku n do mocy szumów $E[|N(f)|^2]$.

3. Optymalizacja pre – przetwarzania redukcji szumów.

Jaki jest najlepszy filtr redukcji szumów H_2 w strukturze AEC+2NR do zastosowania na sygnale mikrofonu? Sposobem modyfikacji H_2 jest zmiana wartości współczynnika wagowego λ w estymacie wartości a priori stosunku sygnał - szum. Obliczamy wartość czynnika redukcji szumów R i czynnika zniekształceń D, wprowadzane przez filtr H_2:

$$R=\frac{1}{M}\sum_{k=1}^{M}10\,\log\frac{P_n(k)}{P_{nf}(k)}, \qquad D=\frac{1}{M}\sum_{k=1}^{M}10\,\log\frac{P_{e-e_f}(k)}{P_e(k)};$$

e_f i n_f reprezentują echo i szum filtrowane przez filtr redukcji szumów H_2, $P_u(k)$ jest mocą u obliczoną w k-tym bloku złożonym z 256 próbek, M jest liczbą bloków, w których występują razem sygnały szumu i echa (tryb pojedynczego nadawania; Single Talk – ST). Rysunek 3. przedstawia zniekształcenia D w



Rysunek 3: R w funkcji D dla różnych wartości ENR

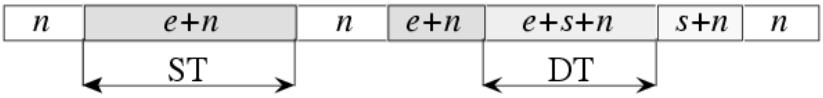
zależności od współczynnika redukcji szumów R dla różnych wartości λ i różnych stosunków echo – szum (Echo-to-Noise Ratio; ENR). ENR jest definiowany przez stosunek mocy echa do mocy szumów w M blokach. Jeśli λ zmierza do 1, to zwiększa się redukcja szumów i zniekształcenie echa, co odpowiada mniejszemu wzmocnieniu filtru H_2.

4. Rezultaty.

Wpływ filtru redukującego szumy H_2 na wydajność struktury AEC+2NR szacowana jest na podstawie pomiarów. Dla porównania prowadzone są też badania struktury AEC+NR.

a) metodologia oszacowania

Baza danych jest otrzymywana przez rejestrowanie osobno sygnału mowy, echa i zniekształcającego szumu tak, aby wziąć pod uwagę różne wartości SNR i ENR. Sygnały te rejestrowane są w samochodzie a szum jest wywoływany prędkością poruszania się samochodu (130 km/h). Na podstawie tych zapisów tworzymy pliki sygnałów złożonych (rysunek 4.), ich pierwszą częścią jest echo (tryb ST), a druga odpowiada mowie nałożonej na szum (tryb DT).



Rysunek 4: Sygnał złożony

Przedstawione są tylko trzy pomiary [3,8]:

- indeks podobieństwa SIM w trybie ST

$$SIM = \frac{1}{N}\sum_{k=1}^{N} 10\ \log\frac{P_e(k)}{P_{e-e_i}(k)}, \quad i = 1,2$$

- Echo Return Loss Enchancement ERLE w obu trybach

$$ERLE = \frac{1}{N}\sum_{k=1}^{N} 10\ \log\frac{P_e(k)}{P_{e_r}(k)}$$

- wzmocnienie G w trybie DT

$$G = \frac{1}{N}\sum_{k=1}^{N} 10\ \log\frac{P_{e+n}(k)}{P_{s-s_f}(k)}$$

k jest indeksem blokowym a N jest liczbą bloków odpowiadającą estymacji przeprowadzonej w trybach ST i DT, s_f jest ostateczną estymatą lokalnego sygnału mowy, e_r reprezentuje resztkowe echo obliczone w następujący sposób:

- w strukturze AEC+NR, e_r jest otrzymywane przez filtrację różnicy $e - e_1$ przy użyciu H_1,
- w strukturze AEC+2NR, e_r otrzymywane jest przez filtrację różnicy $e - e_2$ przy użyciu H_3.

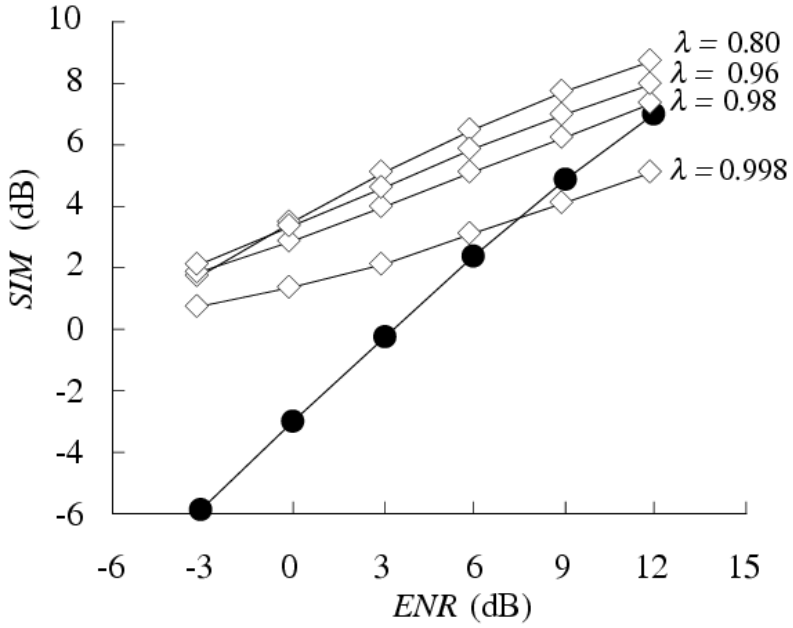
Optymalizacja dotyczy wy łącznie struktury AEC+2NR. Parametry systemu AEC i filtrów H_1 oraz H_3 są ustalone. Wybraliśmy następujące wartości parametrów: dla algorytmu GMDF długość odpowiedzi impulsowej wynosi 25, jest ona podzielona na L = 2 segmenty, a nakładanie się miedzy kolejnymi blokami wynosi (265 – 32) próbki, krok adaptacji równy jest 0.33; dla filtrów redukcji szumów H_1 i H_3, dobrano współczynnik wagowy λ równy 0.98, prawdopodobieństwo nieobecności sygnału q(f) równe 0.5. Moc szumów określana jest na podstawie dziesięciu bloków po 256 próbek, w których obecny jest tylko szum.

Rysunki 5 do 10 przedstawiają wykonane pomiary, uśrednione na zestawie dziesięciu plików, gdzie -●- odpowiada strukturze AEC+NR (kontynuowana adaptacja), a -◇- odpowiada strukturze AEC+2NR.

W trybie ST, stosunek ENR zmienia się od -3 dB do 12 dB a w trybie DT ENR i SNR są identyczne i zmieniają się w zakresie -3 dB do 12 dB.

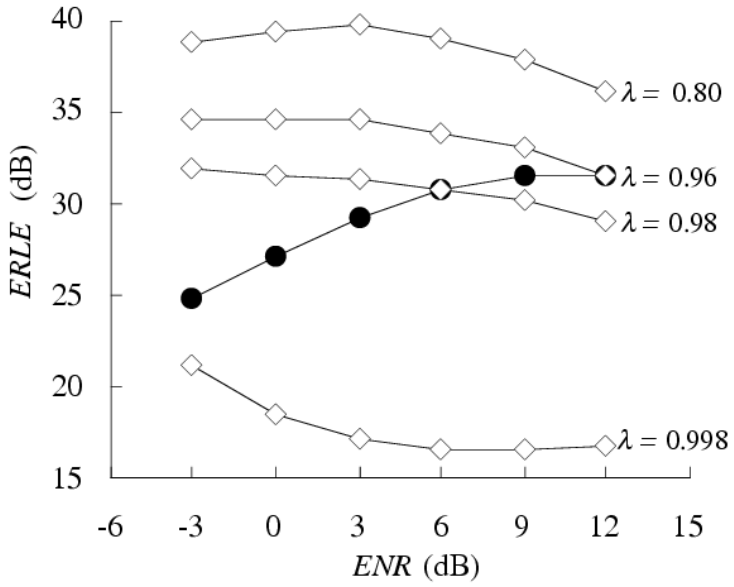
b) Wpływ współczynnika wagowego λ

W trybie ST filtr AEC podlega tylko zniekształceniom wywołanym przez szum. Rysunki 5 i 6 przedstawiają wartości parametrów SIM i ERLE dla różnych wartości λ.



-●- *AEC+NR*, -◇- *AEC+2NR*

Rysunek 5: SIM w trybie ST



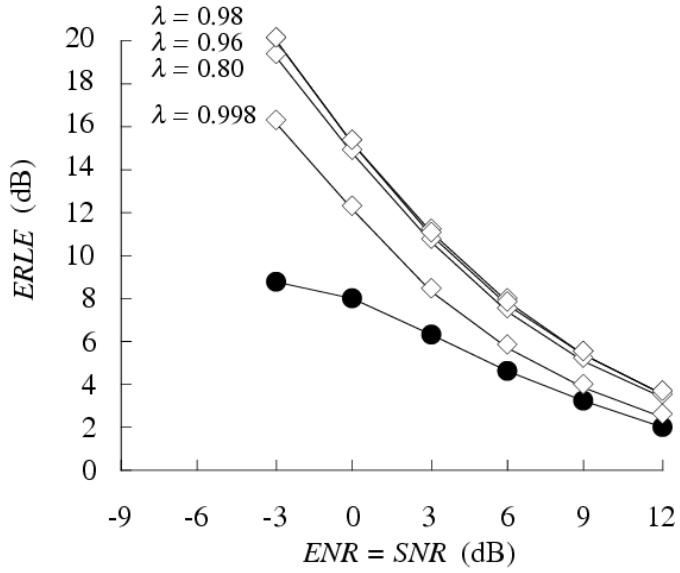
-●- *AEC+NR*, -◇- *AEC+2NR*

Rysunek 6: ERLE w trybie ST

Wartości λ zmieniające się w przedziale [0.5;0.8] prowadzą do porównywalnych wyników dla wysokich wartości ENR; dla niskich wartości EN, wydajność spada ze względu na obecność niezależnej redukcji szumów, więc te wyniki nie są przedstawiane. λ = 0.80 daje najlepsze wartości parametrów SIM i ERLE, co odpowiada redukcji szumów przez filtr H_2 na poziomie około 10 dB (Rysunek 3.). Kiedy λ zmierza do 1, echo jest bardziej zniekształcone i filtracja jest mniej wydajna.

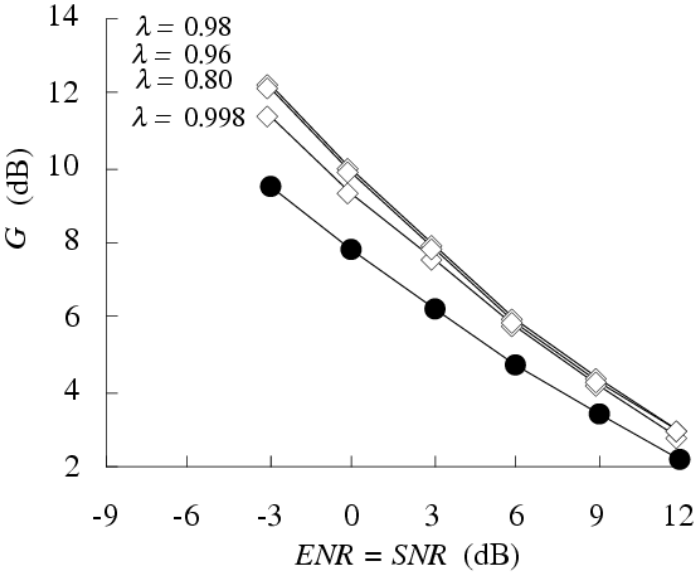
W trybie DT pytanie brzmi: czy jest jakakolwiek korzyść z zatrzymania adaptacji w strukturze AEC+2NR? Rozważmy dwa przypadki:

1. adaptacja jest kontynuowana. Widzimy, że dla $0.80 \leq \lambda \leq 0.90$, parametr ERLE (rysunek 7.) i wzmocnienie (rysunek 8.) są całkiem podobne; dla $\lambda < 0.80$, zaobserwowaliśmy pewną degradację i wyniki nie zostały przedstawione. Ponieważ otrzymujemy interesujące rezultaty w obu trybach dla $\lambda = 0.80$, to ta wartość może zostać zachowana w obu sytuacjach.



-●- *AEC+NR*, -◇- *AEC+2NR (adaptation continued)*

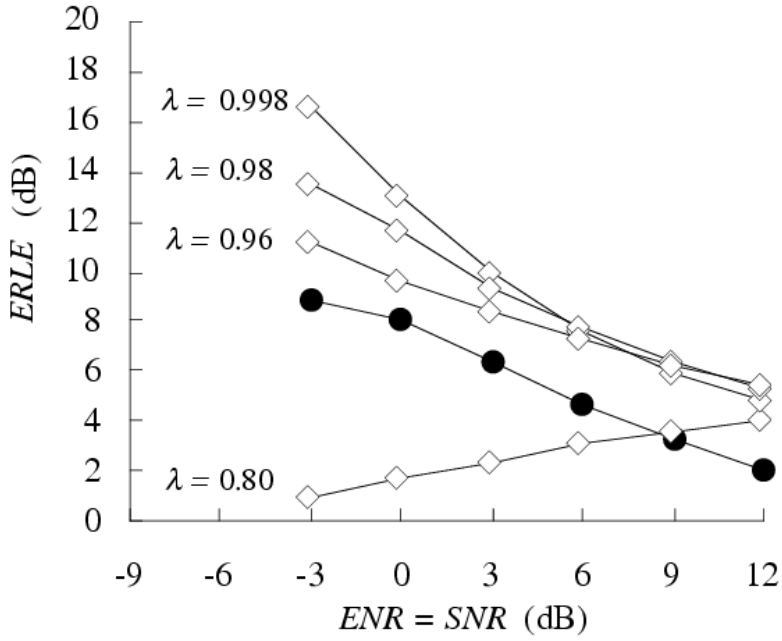
Rysunek 7: ERLE w trybie DT



-●- *AEC+NR*, -◇- *AEC+2NR (adaptation continued)*

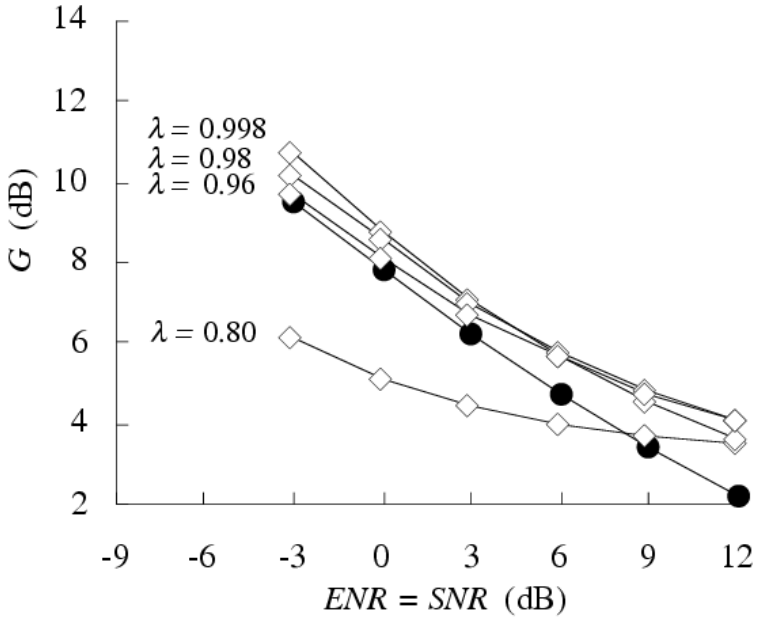
Rysunek 8: Wzmocnienie w trybie DT

2. adaptacja zatrzymana. Jesteśmy pewni, że sygnał mowy s, który ma zostać transmitowany, nie jest zmieniany przez układ AEC. Wartość parametru ERLE (rysunek 9.) pozostaje wysoka dla $0.96 \leq \lambda \leq 0.998$. Przypadki te odpowiadają redukcji szumów większej niż 20 dB dla ENR ≤ 0 dB. Dla tych wartości wzmocnienie struktury AEC+2NR jest większe niż struktury AEC+NR (rysunek 10.). Wartość $\lambda = 0.96$ wydaje się być dobrym wyborem zarówno dla trybu ST jak i DT.



-●- *AEC+NR*, -◇- *AEC+2NR (adaptation stopped)*

Rysunek 9: ERLE w trybie DT



-●- *AEC+NR*, -◇- *AEC+2NR (adaptation stopped)*

Rysunek 10: Wzmocnienie w trybie DT

Wnioskując z tych eksperymentów: dla niskich wartości ENR (i SNR), wydaje się że lepiej jest kontynuować adaptację oraz, dla wysokich ENR (i SNR) lepiej jest ją zatrzymać. Nieoficjalne testy odsłuchowe potwierdzają ten wniosek.

5. Wnioski.

Zaproponowano optymalizację pre–przetwarzania zawarta w nowym kontrolerze akustycznego echa i szumu. Wzięte pod uwagę zostały różne wartości współczynnika wagowego w pre – przetwarzaniu w trybach ST i DT, gdy adaptacja filtru AEC jest kontynuowana bądź zatrzymywana. W ostatnim przypadku, lokalny sygnał mowy jest mniej zniekształcany. Kompletne przedmiotowe oszacowanie musi zostać przeprowadzone w celu walidacji pomiarów.

Bibliografia

[1] R. MARTIN, J. ALTENHÖNER, "Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction", *ICASSP*, pp. 3043 – 3046, Maj 1995.

[2] H. YASAUKAWA, "Acoustic Echo Canceller with Sub-band Noise Cancelling", *Electronics Letters*, vol. 28, no 15, pp. 1403-1404, Lipiec 1992.

[3] G. FAUCON, R. LE BOUQUIN JEANNÈS, "Joint Systems for Acoustic Echo Cancellation and Noise Reduction", *EUROSPEECH*, pp. 1525-1528, Wrzesień 1995.

[4] R. LE BOUQUIN JEANNÈS, B. AYAD, "Systèmes Combinés d'Annulation d'Echo et de Réduction de Bruit", *GRETSI*, pp. 459-462, Wrzesień 1995.

[5] E. MOULINES et al., "The Generalized Multidealy Adaptive Filters: Structures and Convergences Analysis", *IEEE Trans. on Signal Processing*, vol 43, no1, pp 14-28, Styczeń 1995.

[6] Y. EPHRAIM, D. MALAH, "Speech Enhancment Using a Minimum Mean Square Error Short – Time Spectral Amplitude Estimator", *IEEE Trans. on ASSP*, vol ASSP-32, no6, pp.1109-1121, Grudzień 1984.

[7] A. AKBARI AZIRANI, "Rehaussement de la Parole en Ambiance Bruitée. Application aux Télécommunications Mains-Libres", *Thèse de l'Université de Rennes 1*, Listopad 1995.

[8] A. GILLOIRE, "Performance Evaluation of Acoustic Control: Required Values and Measurment Procedures", *Annals of Telecommunications*, 49 no7-8, pp. 368-372, Lipiec – Sierpień 1994.

Tłumaczenie

Michalina Pionke, gr. 1E, nr indeksu 97300