

Cancellazione dell'eco acustica mediante filtraggio
adattativo LMS e decomposizione del segnale in
sottobande

Paolo Annibale

13 febbraio 2007

Sommario

Il problema dell'eco acustica è un aspetto non trascurabile nella progettazione di un sistema di telefonia o di comunicazione mobile; la presenza di eco infatti, non solo può compromettere la stabilità dei sistemi operanti in anello chiuso (*full-duplex*), ma genera fastidio e diminuisce l'intelligibilità delle conversazioni; è pertanto auspicabile che tali sistemi prevedano un dispositivo adibito alla cancellazione dell'eco acustica (*Acoustic Echo Cancellor, AEC*). Poichè il comportamento acustico del sistema è in generale di tipo *tempo-variante*, le tecniche numeriche utilizzate per realizzare un AEC si basano sul filtraggio adattativo, tipicamente, per motivi di semplicità ed efficienza, viene impiegato l'algoritmo LMS (*Least-Mean-Square*). A fronte di queste considerazioni, le prestazioni di un AEC sono determinate, a parità massima attenuazione ottenibile, dalla **velocità di convergenza** e dal **costo computazionale**. Essendo la voce umana un segnale colorato, per ottenere prestazioni soddisfacenti in termini di velocità di convergenza, l'algoritmo LMS deve essere impiegato all'interno di schemi di filtraggio sofisticati. Un possibile approccio prevede la decomposizione del segnale in ingresso in sottobande che, presentando uno spettro più uniforme rispetto al segnale originario, possono essere filtrate con prestazioni migliori in termini di velocità di convergenza; ciascuna sottobanda inoltre può essere sottocampionata incrementando l'efficienza dell'AEC. L'articolo presenta una panoramica sulla realizzazione di un AEC basato su LMS e decomposizione spettrale; si tratteranno i fondamenti teorici della cancellazione dell'eco acustica, le tecniche per la decomposizione del segnale in sottobande e, facendo riferimento ad una simulazione in linguaggio MATLAB, si confronteranno le prestazioni con quelle del filtraggio LMS sul segnale a banda completa e del il filtraggio adattativo in frequenza (*Frequency Domain Adaptive Filtering, FDAF*) .

0.1 Introduzione alla cancellazione dell'eco

Lo studio approfondito e completo di un sistema di telecomunicazione presenta numerose difficoltà ed esula dagli scopi di questo lavoro, pertanto si analizzerà un modello semplificato, comunque sufficiente per comprendere gli aspetti principali della cancellazione dell'eco acustica. In un sistema di telefonia reale esiste accoppiamento fra i due terminali, in questa sede invece, per semplicità, si considererà il problema della cancellazione dell'eco solo nel terminale ricevente come se si trattasse di un sistema isolato; verrà inoltre trascurato il ritardo introdotto dalla rete, al fine di concentrarsi solo su quello acustico e quello introdotto dalle operazioni numeriche di filtraggio. Sotto queste ipotesi è possibile rappresentare il problema della cancellazione dell'eco mediante un semplice modello come quello in figura 1. Il sistema è costituito da un **parlatore lontano**, che in assenza dell'AEC è infastidito dall'ascoltare l'eco della proprio voce, un **parlatore vicino**, che si trova nel terminale ricevente caratterizzato da un determinato **comportamento acustico** e da una coppia **microfono-altoparlante**; il compito dell'AEC è rimuovere dal segnale catturato dal microfono la voce del parlatore lontano.

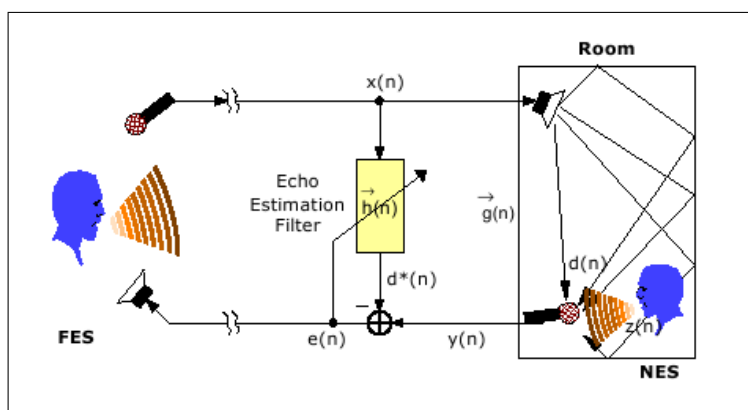


Figura 1: Modello per la cancellazione dell'eco acustica mediante filtraggio adattativo

0.2 Identificazione di un sistema incognito

Come anticipato le tecniche di cancellazione dell'eco acustica si basano sull'utilizzo del filtro numerico adattativo LMS (*Least Mean Square*) in particolare, data la natura non stazionaria della voce, si utilizza la variante normalizzata (*Normalized LMS, NLMS*); le proprietà dell'algoritmo NLMS possono essere apprese facilmente dalla letteratura [4] [5], brevemente l'idea è quella di utilizzare il blocco NLMS per **identificare un sistema incognito** (*Unknown System o Plant*), che nella fattispecie corrisponde al cammino acustico tra l'altoparlante e il microfono nel terminale di ricezione. Più precisamente il compito dell'AEC nel terminale ricevente è rimuovere dal segnale catturato dal microfono quella porzione di segnale che presenta correlazione con il segnale prodotto dal parlatore lontano, ovvero il segnale $x(n)$ (*Reference Signal*) convoluto con la risposta

impulsiva $g(t)$ corrispondente al cammino acustico fra l'altoparlante e il microfono; così facendo si eviterà che il parlatore lontano riascolti, insieme al segnale utile prodotto dal parlatore vicino $z(n)$, la propria voce ritardata e filtrata. Il cancellatore deve pertanto *emulare* il comportamento filtrante della stanza, generando a partire da $x(n)$ e dal segnale errore $e(n)$, una *stima* di $d(n)$ (*Desired Signal*), quindi sottrarla dal segnale microfonico realizzando di fatto la cancellazione dell'eco (figura 2). Affinchè il filtro adattativo possa riprodurre la risposta impulsiva del cammino acustico, l'ordine del filtro deve essere comparabile con quello della risposta impulsiva desiderata; La conseguenza principale di questo fatto è che la lunghezza del filtro adattativo cresce proporzionalmente al massimo ritardo che si desidera coprire, pertanto in applicazioni caratterizzate da ambienti con tempi di riverbero significativi la lunghezza richiesta al filtro potrebbe essere proibitiva dal punto di vista del costo computazionale.

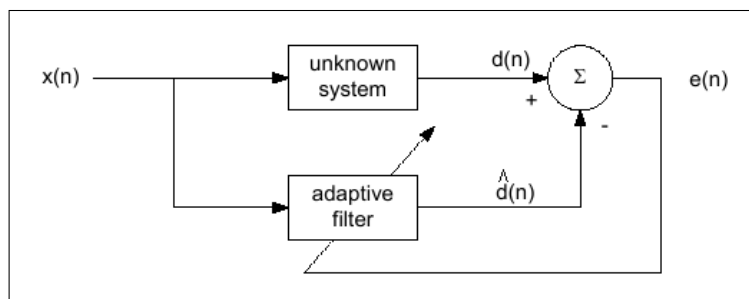


Figura 2: Identificazione di un sistema incognito mediante filtraggio adattativo

0.3 Decomposizione del segnale in sottobande

L'idea base legata alla scelta di decomporre il segnale in sottobande, è quella di migliorare le caratteristiche di convergenza del filtro NLMS eccitandolo con segnali caratterizzati da **spettro il più possibile uniforme**; è noto infatti dalla teoria che la velocità di convergenza dell'NLMS, sotto l'ipotesi di stazionarietà, è massima quando il rapporto fra l'autovalore massimo e quello minimo della matrice di autocorrelazione R_{xx} del segnale in ingresso $x(n)$ è pari ad uno, il che si verifica con segnali a spettro uniforme. La decomposizione dello spettro del segnale in sottobande con estensione in frequenza minore consente di ottenere il desiderato effetto *sbiancante*, ciascuna sottobanda infatti presenterà variazioni di ampiezza minori in frequenza. Lo schema a blocchi del filtraggio adattativo per sottobande (*Subband Adaptive Filtering, SAF*) osservabile in figura 3, prevede pertanto due blocchi di analisi, per decomporre rispettivamente $x(n)$ e $d(n)$, un banco di filtri NLMS, uno per ogni sottobanda, e un blocco di sintesi, per ricostruire il segnale errore $e(n)$ a partire dalle uscite dei filtri NLMS. A causa della banda ristretta, le singole sottobande possono essere **sottocampionate**, di conseguenza anche la lunghezza dei filtri NLMS può essere ridotta, migliorando l'efficienza del sistema.

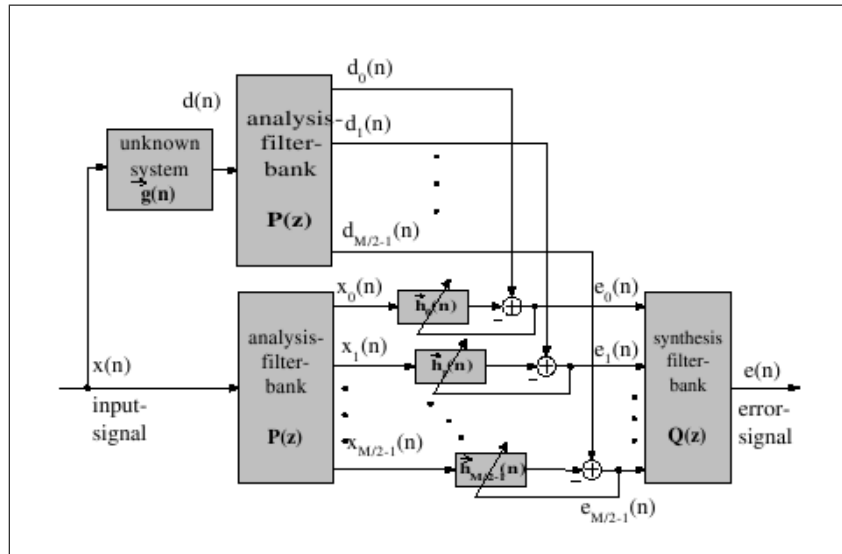


Figura 3: Schema a blocchi del filtraggio adattativo mediante decomposizione in sottobande

0.3.1 Caratteristiche della decomposizione spettrale

E' importante a questo punto concentrarsi sulle caratteristiche che deve soddisfare l'operazione di decomposizione e ricomposizione del segnale poichè essa riveste un ruolo critico nel raggiungimento degli obiettivi di convergenza ed efficienza del sistema. Anzitutto è opportuno scegliere un numero di sottobande M sufficiente per apprezzare l'effetto decorrelante, per contro un numero elevato di sottobande non solo grava sul costo computazionale ma degrada il processo di ricostruzione del segnale; nella pratica si sceglie M pari ad una potenza di due come 8, 16 o 32. La decomposizione del segnale può essere ottenuta mediante modulazione di un opportuno filtro passa basso (*Prototype filter*) sulla frequenza centrale di ciascuna sottobanda (figura 4).

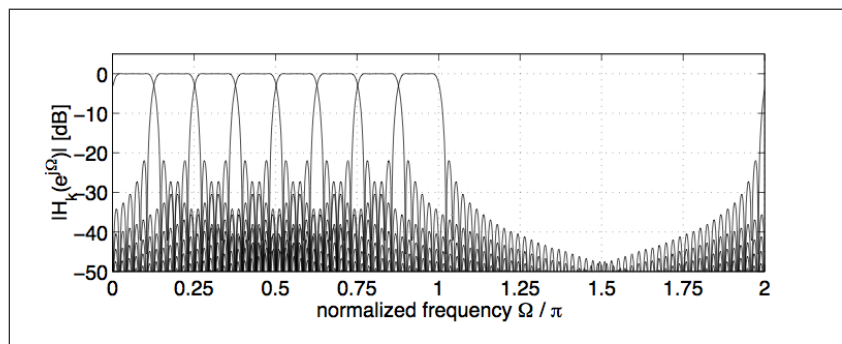


Figura 4: Decomposizione in sottobande mediante modulazione di un filtro prototipo

Le caratteristiche richieste al filtro prototipo, osservabili in figura 5, sono

ovviamente stringenti in termini di **selettività in frequenza** e **attenuazione in banda oscura**; essendo vincolati ad utilizzare filtri FIR per avere **sfasamento lineare** il raggiungimento di tali caratteristiche si paga con l'aumento della lunghezza della risposta impulsiva del filtro. La scelta della lunghezza del filtro di analisi e del suo complementare di sintesi è quindi estremamente critica, infatti, se da una parte la qualità della separazione garantisce un miglior effetto decorrelante e un più alto fattore di decimazione delle sottobande, dall'altra la lunghezza del filtro non può crescere eccessivamente in quanto introdurrebbe nel sistema un ritardo e un costo computazionale inaccettabili. Nella pratica si cercherà un *trade-off* fra queste due esigenze contrapposte.

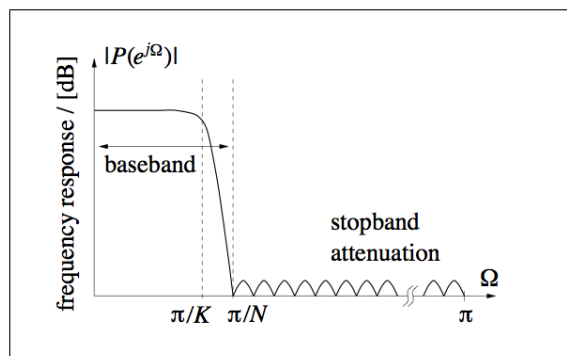


Figura 5: Caratteristiche del filtro prototipo

0.3.2 Banco di filtri per la decomposizione del segnale

A questo punto è necessario ricordare che le operazioni di analisi e sintesi di un segnale per sottobande possono essere realizzate in modo estremamente efficiente mediante un **banco di filtri polifase**. Sfruttando le cosiddette *identità nobili* [3] è possibile spostare l'operazione di decimazione a monte dei filtri passa banda ottenuti per modulazione che quindi lavoreranno ad un *rate* inferiore; infine sostituendo i filtri passa banda con le sequenze polifase ottenute dal prototipo passa basso, è possibile ottenere le singole sottobande del segnale ricombinando le uscite dei filtri polifase mediante una matrice di modulazione (ad esempio la matrice *GDFT*, *Generalized Discrete Fourier Transform* [2]). Se la banda del filtro passa basso è π/K allora le K sequenze corrispondenti alle sottobande possono essere decimate di un fattore $N = K$ (*sequenze massimamente decimate*). Tuttavia nella pratica si utilizzerà un valore N inferiore a K , infatti, poiché le sottobande vengono modificate dal filtraggio NLMS, una buona ricostruzione non sarebbe possibile senza introdurre una certa ridondanza spettrale, il che equivale a scegliere un fattore di decimazione $N < K$. Si ricorda inoltre che poiché il segnale in ingresso è reale sarà necessario processare solo $M = K/2$ sequenze complesse. L'operazione di ricomposizione delle sottobande in un segnale a banda completa viene effettuata in modo altrettanto efficiente da un banco di filtri complementare; per approfondire questi argomenti si rimanda a [2] e [3].

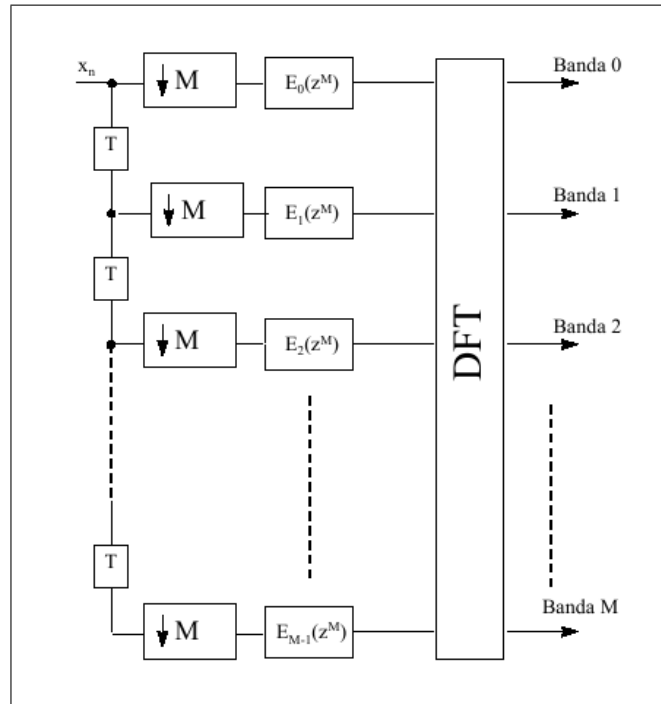


Figura 6: Banco di filtri di analisi realizzato mediante struttura polifase

0.4 Simulazione in MATLAB e conclusioni

Nella simulazione si è scelto di generare la risposta impulsiva $g(h)$ del terminale ricevente con il **metodo delle sorgenti immagine**; la propagazione delle onde acustiche a raggi è infatti una buona approssimazione per segnali con frequenza superiore ai $300Hz$, come appunto la voce umana; in pratica si assume che la risposta impulsiva dell'ambiente sia costituita da una serie di ritardi puri a breve distanza e ampiezza elevata dovuti alle riflessioni primarie, i cosiddetti *echi principali*, seguiti da un coda di ritardi più deboli e fitti corrispondenti alle riflessioni secondarie detta *riverbero*. In figura 7 si può osservare la risposta impulsiva $g(n)$ per una ambiente di medie dimensioni caratterizzato da un coefficiente di riflessione $r = 0.6$ e un tempo di riverberazione di circa $125ms$.

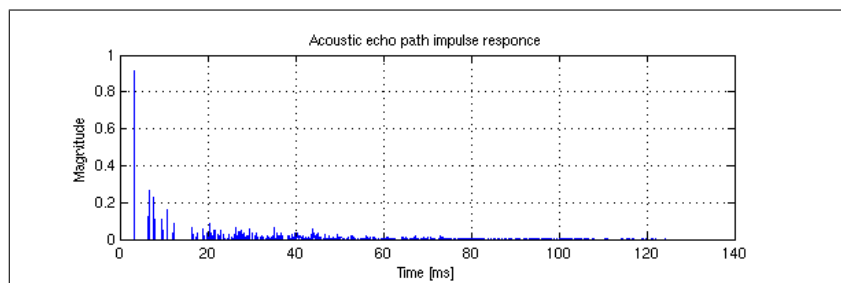


Figura 7: Risposta impulsiva dell'ambiente $g(n)$

Come segnale di riferimento $x(n)$ si è scelto di generare un segnale a spettro uniforme per poi colorarlo mediante un filtro ricorsivo del primo ordine, è possibile comunque anche caricare un file audio contenente del parlato. Al segnale desiderato $d(n)$ è stato aggiunto del rumore bianco tale da ottenere un SNR pari a $30dB$. Nella figura 8 è osservabile il filtro passa basso con banda $\pi/16$ impiegato come prototipo per l'estrazione di $M = 8$ sottobande complesse. La lunghezza del filtro è pari a 448 campioni, il fattore di decimazione scelto è $D = 7$ e la lunghezza dei filtri NLMS è di $L = 150$ campioni contro $L = 1024$ del sistema a banca completa e di quello FDAF, tale lunghezza consente in entrambi i casi di coprire un ritardo massimo di circa $T_s * L$ comparabile con il massimo ritardo significativo introdotto dall'acustica dell'ambiente.

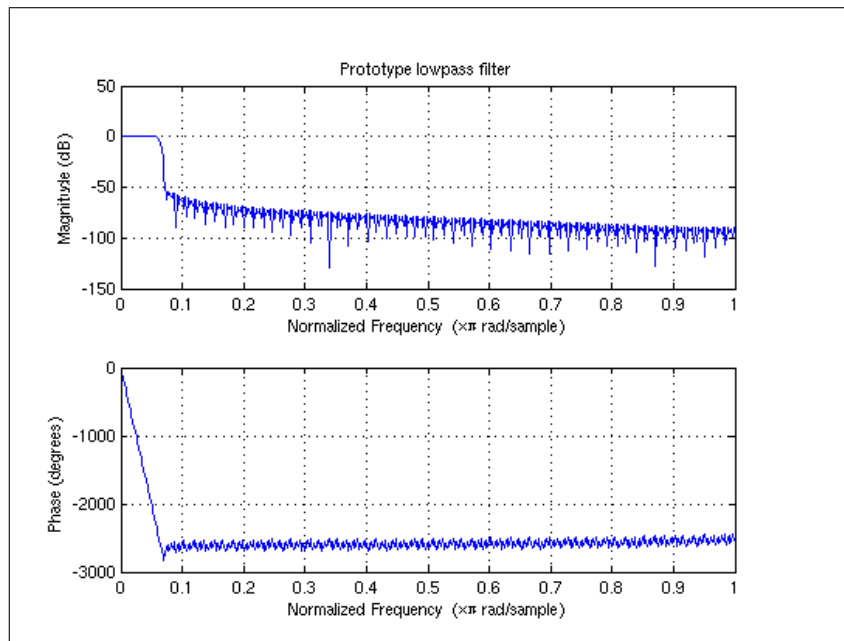


Figura 8: Filtro prototipo di banda $\pi/16$

Sono stati eseguite diverse simulazioni variando il fattore di correlazione b fra i campioni del segnale sintetico $x(n)$ in ingresso all'AEC; nelle figure 9, 10 e 11 sono visibili i grafici dell'*AERLE*, *Acoustic Echo Return Loss Enhancement* per differenti valori di b (0.7, 0.8, 0.9). I grafici mostrano il miglior comportamento dell'AEC basato su decomposizione spettrale sia rispetto all'NLMS semplice, sia rispetto all'FDAF, la variante dell'NLMS nel dominio della frequenza utilizzata spesso perchè implementabile in modo efficiente mediante FFT [6]; in tutti e tre i casi, infatti, l'AEC basato su SAF raggiunge un'attenuazione di circa $30dB$ dopo circa un secondo, dimostrando, come ci si aspettava, robustezza nei confronti anche dei segnali più colorati ($b = 0.9$); anche la variante FDAF non è significativamente influenzata dalla variazione del fattore di correlazione, tuttavia le caratteristiche attenuanti e di convergenza sono scadenti rispetto al sistema SAF; come era prevedibile, invece, le prestazioni del filtraggio NLMS sul segnale a banda completa degradano sensibilmente all'aumentare della correlazione. Queste considerazioni vengono infine confermate effettuando la simulazione con

in ingresso un segnale di parlato; i grafici mostrati nelle figure 12 e 13 evidenziano una maggiore velocità di convergenza, quindi un miglior comportamento attenuante dell'AEC basato su decomposizione spettrale. Si conclude pertanto che i risultati delle simulazioni sono in accordo con quanto previsto dalla teoria: a fronte di una maggiore complessità in fase di progettazione, la scomposizione del segnale in sottobande consente, in presenza di segnali colorati, prestazioni migliori in termini di velocità di convergenza. Non solo la struttura del banco di filtri per l'analisi e sintesi si presta ad una implementazione efficiente che consente di ridurre il costo computazionale rispetto al sistema a banda completa, per una trattazione dettagliata dei costi computazionali della struttura si rimanda a [1].

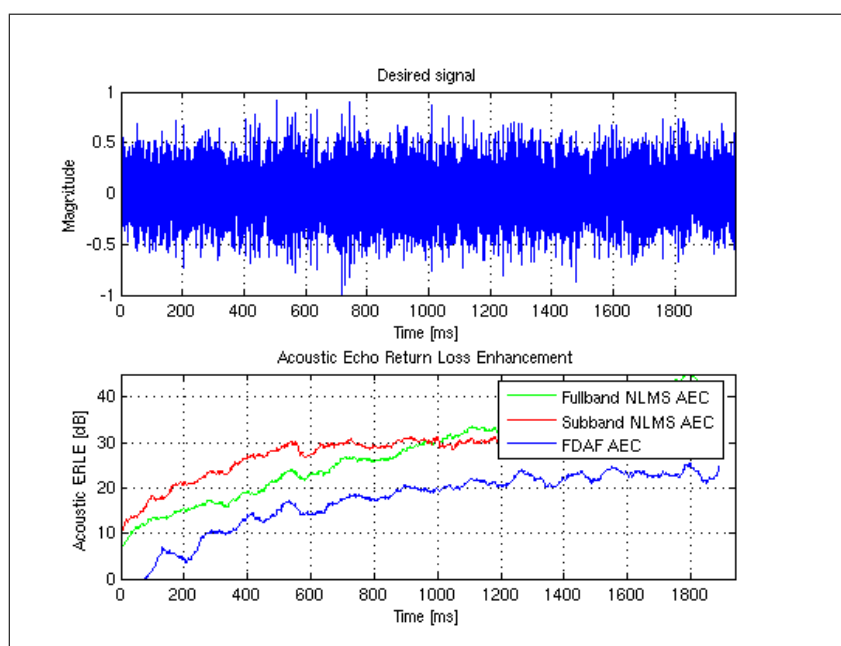


Figura 9: Grafico dell'AERLE ottenuto con segnale sintetico $x(n)$ e fattore di correlazione $b = 0.7$

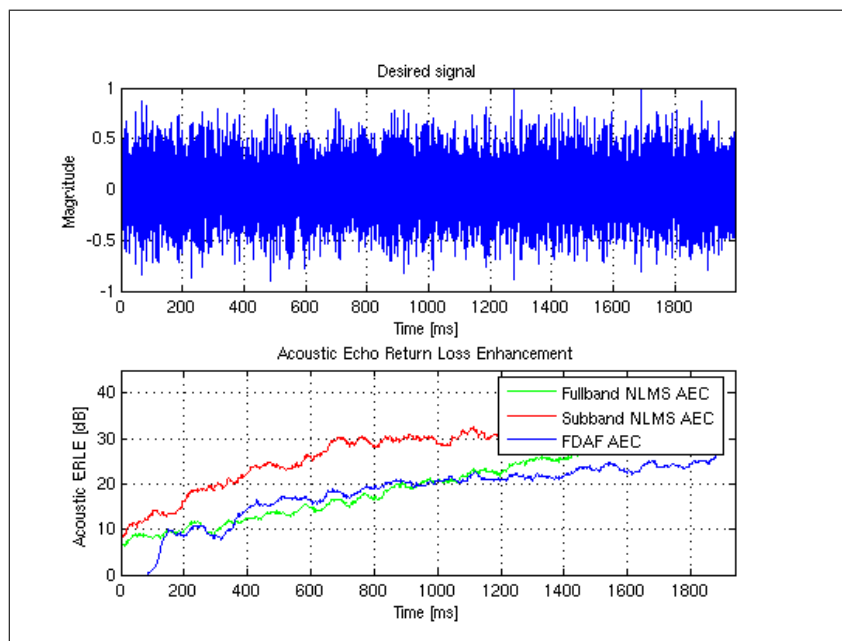


Figura 10: Grafico dell'AERLE ottenuto con segnale sintetico $x(n)$ e fattore di correlazione $b = 0.8$

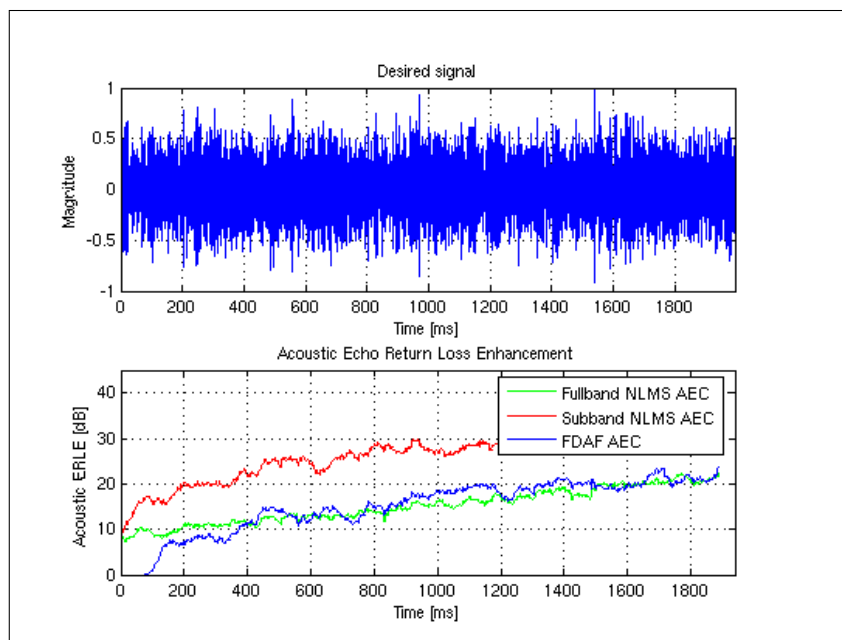


Figura 11: Grafico dell'AERLE ottenuto con segnale sintetico $x(n)$ e fattore di correlazione $b = 0.9$

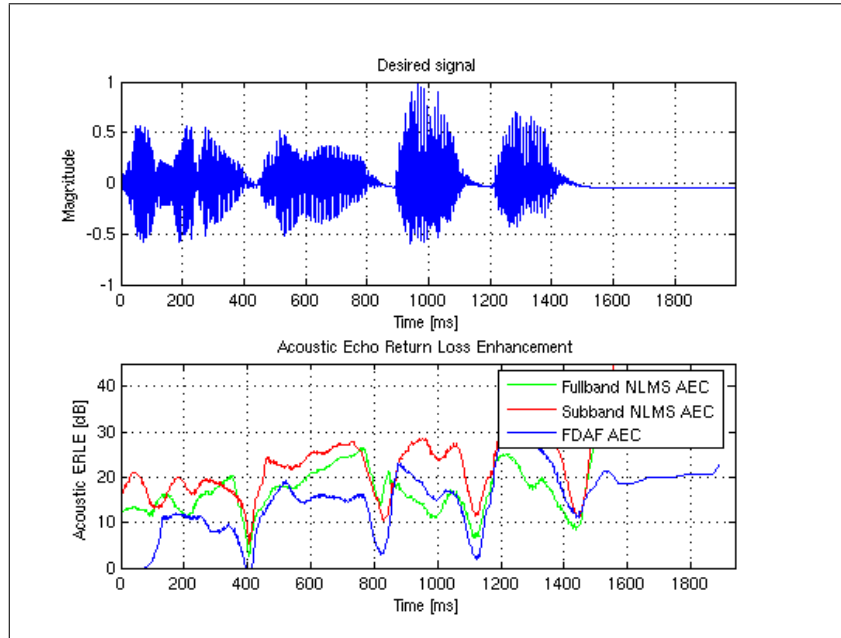


Figura 12: Grafico dell'AERLE ottenuto con un segnale di parlato $x(n)$

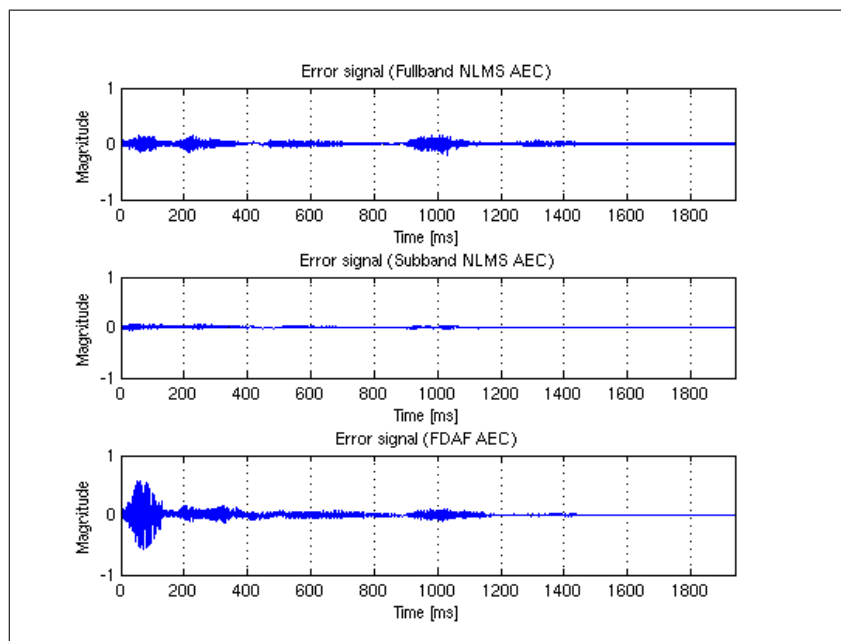


Figura 13: Segnale errore $e(n)$ in uscita dai tre AEC

Bibliografia

- [1] Sandrock,Schmitt, *Realization of an Adaptive Algorithm with Aubband Filtering Approach for Acoustic Echo Cancellation in Telecommunication Application*, DSPecialists GmbH, D-10254 Berlin, Germany.
- [2] S. Weiss, A. Stenger, R.W. Steewart, R. Rabenstein, *Steady-State Performance Limitations of Subband Adaptive Filters*, IEEE Transactions on Signal Processing, Vol. 49, Num. 9, Sep. 2001.
- [3] F. Rocca, *Fondamenti di elaborazione numerica dei segnali*, CUSL, pp. 410.
- [4] B. Widrow, S.D. Sterns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ,1985.
- [5] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall,Englewood Cliffs, 1968.
- [6] J.J. Shynk, *Frequency-Domain and Multirate Adaptive Filtering*, IEEE Signal Processing Magazine, vol. 9, no. 1, pp. 14-37, Jan. 1992
- [7] K. Murano, S Unagami, F. Amano.*Acoustic Echo Cancellation and Application*, IEEE communications magasine, pp 49-55, Jan. 1990.
- [8] V.Allen, M. E. Henderson, E. S. Wheeler, J. Williams *Acoustic Echo Cancellation in a Reverberatory Chamber Using an Adaptive Least Means Square Algorithm*, MS State DSP Conference, pp 100-111, 1995.