

IMPLEMENTAZIONE DI UN ALGORITMO DI BSS IN FREQUENZA PER LA SEPARAZIONE DI SORGENTI ACUSTICHE

Francesco Nesta^{1,2}, Maurizio Omologo¹, Piergiorgio Svaizer¹

1) Fondazione Bruno Kessler - IRST, 2) Università di Trento

{*nesta,omologo,svaizer*}@fbk.eu

1. INTRODUZIONE

Negli ultimi dieci anni notevole interesse ha ottenuto un emergente ramo di ricerca dell'elaborazione numerica di segnali, la Blind Source Separation (BSS)[13]. Le tecniche di BSS, infatti, sono diventate in questi ultimi anni molto robuste e in un prossimo futuro potranno incrementare sostanzialmente le prestazioni di sistemi reali di riconoscimento automatico del parlato. Lo scopo della BSS è di separare diverse sorgenti avendo come un'unica informazione le misture dei segnali osservate mediante dei sensori. Con il termine "Blind" s'intende la capacità del sistema di effettuare la separazione senza avere nessuna conoscenza a priori sul processo di mixing o sulla natura dei segnali. Le tecniche di BSS sono applicate con successo da diversi anni nel campo della neurologia [7][8] mentre solo recentemente gli algoritmi di separazione per segnali acustici sono diventati sufficientemente robusti per essere usati in ambienti reali.

Nel caso delle applicazioni per segnali neurologici le tecniche di BSS sono basate su algoritmi di Independent Component Analysis (ICA) usati per separare segnali elettrici dell'attività cerebrale ottenuti mediante encefalogramma (EEG). Gli algoritmi di ICA si basano sull'assunzione che le sorgenti siano statisticamente indipendenti e sfruttano quest'ulteriore informazione per stimare un processo di separazione che permetta di ottenere le componenti di segnale indipendenti.

Nelle applicazioni audio un sistema di BSS molto efficace, utilizzato spesso come termine di paragone, è l'apparato uditivo periferico umano e per questo motivo tali tecniche sono inizialmente nate con lo scopo risolvere la separazione nel campo dei segnali acustici. L'uomo, infatti, è capace di discriminare più sorgenti sonore presenti in un ambiente, anche in condizioni molto avverse. Questa capacità è spesso referenziata con l'espressione "Cocktail Party Effect".

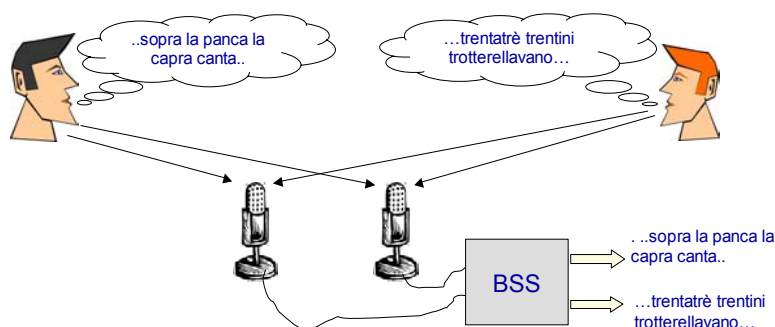


Figura 1: BSS per separazione di sorgenti acustiche

Un sistema di BSS può essere implementato attraverso differenti approcci. Principalmente è possibile individuare due categorie: 1) BSS ad un canale, 2) BSS multicanale.

Nel primo caso le sorgenti sono separate utilizzando come informazione le osservazioni di una sola miscela di segnali ottenuta mediante un sensore. Diverse metodologie sono state esplorate come l'Independent Subspace Analysis (ISA)[5] e la Non-Negative Matrix Factorization (NMF), e combinazioni ibride di differenti metodi per ottenere risultati più robusti. Non tutti questi metodi in realtà sono completamente "blind" poiché le diverse tecniche formulano ipotesi sulla struttura dei segnali di partenza assumendo ad esempio ridondanza di informazione nello spettro. La BSS mono canale può essere considerata allo stato attuale il ramo di ricerca più ambizioso della BSS e nonostante i recenti progressi l'affidabilità e la robustezza di questi metodi sono ancora insufficienti perché possano essere adottati in applicazioni reali.

Nel caso di BSS multicanale le sorgenti di segnale sono osservate utilizzando una schiera di sensori. Tale approccio permette di introdurre ulteriori osservazioni utili alla stima del processo di separazione. In particolare nel caso di BSS multicanale per applicazioni audio si utilizzano implicitamente informazioni di indipendenza spaziale delle sorgenti. Nonostante la BSS multicanale richieda ancora un elevato costo computazionale, i recenti progressi hanno portato a metodi molto affidabili e sufficientemente veloci perché possano essere implementati in real-time.

Comunque anche in questo caso esistono ancora dei problemi, quali ad esempio la separazione di sorgenti spazialmente molto vicine o quando il numero di sorgenti è superiore a quello dei sensori (caso "underdetermined"). Inoltre il tempo di riverberazione è ancora molto influente sulla qualità complessiva del segnale separato.

Oltre alla sopracitata esiste un'ulteriore possibile classificazione: 1) algoritmi per misture istantanee 2) algoritmi per misture convolutive.

Nel primo caso i segnali osservati ai microfoni (nel caso audio) possono essere visti come una combinazione lineare dei segnali originali delle sorgenti. In queste condizioni gli algoritmi di ICA sono sufficienti a separare efficacemente le sorgenti. Nel secondo caso invece i segnali osservati ai microfoni sono una versione dei segnali originali convoluta con la risposta impulsiva dell'ambiente. Questo è il caso di un sistema che cerchi di separare sorgenti acustiche in un ambiente riverberante.

In questo lavoro ci occupiamo di sviluppare un algoritmo di BSS utilizzabile in ambienti reali anche in condizioni di riverberazione molto avverse. Un approccio molto comune, lo stesso da noi adottato, è di eseguire la separazione nel dominio della frequenza. In questo caso, infatti, è possibile ricondurre il problema della separazione di misture convolutive a misture istantanee, permettendo quindi l'utilizzo delle efficienti tecniche di ICA. La separazione dei segnali viene effettuata in modo indipendente per ogni frequenza e i segnali sono successivamente ricostruiti nel dominio del tempo. In questo lavoro proponiamo in particolare un nuovo approccio per risolvere il noto problema delle "permutazioni", tipico dei metodi basati sul dominio della frequenza.

Prima di entrare nei dettagli della tecnica di BSS proposta, nella sezione 2 e 3 formuliamo il problema di separazione rispettivamente per il caso di misture istantanee e convolutive. Nella sezione 4 viene affrontato il problema dell'ambiguità della soluzione di ICA e di seguito un nuovo schema per la soluzione delle permutazioni è presentato nella sezione 5. Infine risultati sperimentali ottenuti su dati reali sono presentati nell'ultima sezione e confermano la validità dell'approccio proposto.

2. BSS PER MISTURE ISTANTANEE E ICA

Consideriamo un vettore $\mathbf{x}(t)$ di N elementi le cui componenti $x_i(t)$ rappresentano il valore assunto dal segnale i -esimo all'istante t . Se supponiamo che i segnali siano miscelati istantaneamente possiamo indicare con $\mathbf{H}(t)$ la matrice di mixing $M \times N$ e con $\mathbf{y}(t)$ il vettore dei segnali miscelati captati da M microfoni. Se il sistema di mixing è stazionario, possiamo considerare la matrice di mixing costante e i segnali miscelati $y_i(t)$ sono rappresentabili come una combinazione lineare dei segnali originali delle sorgenti. Per semplificare la trattazione supponiamo che $N=M=2$:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad (1)$$

Il problema di separazione è ricondotto ad una semplice soluzione del sistema lineare attraverso il calcolo della matrice di separazione $\mathbf{W}$, inversa di quella di mixing $\mathbf{H}$. Si può subito osservare che perché sia possibile l'inversione, $\mathbf{H}$ non deve essere singolare. Nel caso semplice in cui $M = N$ l'inversione è matematicamente possibile e quindi il calcolo di si traduce nell'effettuare una stima della matrice $\mathbf{H}$. In ogni modo nel caso "blind" i segnali originali non sono conosciuti a priori e quindi non esiste una soluzione analitica del problema. Infatti si può osservare che il numero di incognite è superiore a quello delle equazioni. La matrice $\mathbf{W}$ quindi può essere solo stimata utilizzando un processo iterativo che usi un'informazione aggiuntiva.

L'approccio più comune per la separazione di misture istantanee è quello di supporre che le sorgenti siano statisticamente indipendenti. In questo caso la matrice $\mathbf{W}$ è stimata in modo da massimizzare l'indipendenza statistica tra segnali di output associati alle sorgenti. Questo principio è sfruttato dai metodi di ICA ormai largamente usati nel campo della neurologia per l'analisi, per la decomposizione di segnali elettrici associati all'attività cerebrale.

Per la stima di $\mathbf{W}$ è quindi necessaria una metrica che valuti la dipendenza statistica tra i segnali. Se consideriamo le osservazioni dei segnali $y_i(t)$ come realizzazioni di una variabile aleatoria possiamo considerare i segnali solo da un punto di vista statistico, perdendo l'informazione temporale. I segnali osservati possono essere considerati come una combinazione lineare di variabili aleatorie associate ai segnali delle sorgenti.

$$\mathbf{y} = \mathbf{H}\mathbf{x} \quad (2)$$

Supponendo che le sorgenti non siano gaussiane possiamo usare come misura di dipendenza la gaussianità delle componenti ottenute, poiché secondo il teorema del limite centrale la somma di variabili aleatorie indipendenti tende asintoticamente ad assumere una densità di probabilità gaussiana. La matrice $\mathbf{W}$ può quindi essere stimata attraverso un approccio a gradiente seguendo la direzione che minimizza la gaussianità dei segnali delle sorgenti. La gaussianità può essere valutata mediante misure di kurtosis, negentropy e in generale attraverso statistiche di ordine superiore.

Un altro approccio analogo, proveniente dalla teoria dell'informazione, è quello di utilizzare la distanza statistica tra le sorgenti attraverso il calcolo della mutua informazione.

Algoritmi strettamente legati a questa misura sono l'InfoMax[3] e le sue recenti varianti come lo Scaled Natural Gradient[4]. Ad ogni iterazione (*iter*) un algoritmo di ICA basato sul gradiente naturale stima la matrice di separazione secondo la seguente ricorsione:

$$\mathbf{W}_{iter} \longleftarrow \mathbf{W}_{iter-1} + \eta [\mathbf{I} - \langle \alpha(\bar{\mathbf{x}}) \bar{\mathbf{x}}^H \rangle_{\tau}] \mathbf{W}_{iter-1} \quad (3)$$

dove $\langle \cdot \rangle_{\tau}$ è l'operatore di media su τ , $\mathbf{I}$ è la matrice unitaria, $\bar{\mathbf{x}}$ rappresenta il segnale *i-esimo* ottenuto separando le misture $\mathbf{y}$ con la matrice $\mathbf{W}$ ottenuta al passo di iterazione precedente:

$$\bar{\mathbf{x}} = \mathbf{W}_{iter-1} \mathbf{y} \quad (4)$$

ed $\alpha(\bar{\mathbf{x}})$ è una funzione non lineare definita da:

$$\alpha(\bar{x}_i) = -\frac{\delta}{\delta \bar{x}_i} \log p(\bar{x}_i) \quad (5)$$

in cui $p(\bar{x}_i)$ rappresenta la distribuzione di probabilità di $\bar{x}_i$.

3. BSS PER MISTURE CONVOLUTIVE

Nel caso reale di segnali acustici osservati mediante microfoni, il modello basato su misture istantanee non è più appropriato. Infatti l'acustica dell'ambiente modifica temporalmente i segnali osservati ai microfoni secondo una propria caratteristica risposta convolutiva.

Supponiamo di avere nuovamente 2 sorgenti e 2 microfoni e indichiamo con $x_i(t)$ le sorgenti di segnale. Nel caso convolutivo possiamo rappresentare la matrice di mixing come una matrice $\mathbf{H}$ di elementi $h_{ij}(l)$ che rappresentano le risposte impulsive (a supporto limitato) tra il microfono *i-esimo* e la sorgente *j-esima*:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} h_{11}(l) & h_{12}(l) \\ h_{21}(l) & h_{22}(l) \end{bmatrix} * \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad (6)$$

dove abbiamo indicato con * l'operazione di convoluzione. I segnali osservati dai microfoni possono essere rappresentati come una combinazione lineare delle sorgenti convolute con le risposte impulsive $h_{ij}(l)$.

$$y_i(t) = \sum_{j=1}^N \sum_{l=0}^{L-1} h_{ij}(l) x_j(t-l), \quad i = 1, \dots, N \quad (7)$$

dove L è la lunghezza delle risposte impulsive (per ipotesi a supporto limitato). Per ottenere una stima dei segnali di partenza $\bar{\mathbf{x}}$ è necessario determinare una matrice di deconvoluzione in modo che i segnali nel tempo siano statisticamente indipendenti. A differenza del caso di misture istantanee nel caso convolutivo è necessario stimare un numero elevato di parametri dipendente della lunghezza dei filtri di deconvoluzione L. Nonostante la formulazione nel dominio del tempo sia quella più corretta, l'eccessivo costo computazionale dovuto alla congiunta stima dei filtri $h_{ij}(l)$ rende impraticabile la separazione in tempi di convergenza ragionevoli. Per semplificare la stima dei parametri $h_{ij}(l)$ un metodo comunemente utilizzato è quello di effettuare la separazione nel dominio delle frequenze. Come ben noto l'operazione di convoluzione nel tempo corrisponde ad un prodotto nel dominio delle frequenze, quindi applicando la trasformata di Fourier alla (7) otteniamo:

$$Y_i(f) = \sum_{j=1}^N h_{ij}(f) X_j(f) \quad (8)$$

E' evidente quindi che la separazione per segnali convolutivi possa essere ricondotta alla separazione di misture istantanee trattando ogni involuppo di frequenza in modo indipendente. Essendo i segnali $\mathbf{x}(t)$ non stazionari, applicando la Short-Time Fourier Transform (STFT) si ottengono degli involuppi temporali per ogni frequenza. Il segnale osservato ai microfoni può essere rappresentato come:

$$y_i(f, \tau) = \sum_{j=1}^N h_{ij}(f) x_j(f, \tau) \quad (9)$$

Si può quindi notare che le componenti di frequenza osservate ai microfoni sono una combinazione lineare delle componenti associate ai segnali originali. E' evidente quindi che nel dominio delle frequenze si possano applicare gli algoritmi per misture istantanee basati su ICA, effettuando la separazione in modo indipendente lungo le frequenze.

Applicando la separazione agli involuppi si determina una matrice di separazione $\mathbf{W}(f)$ che separa le componenti di frequenza indipendenti in $\bar{\mathbf{x}}(f, \tau)$:

$$\bar{\mathbf{x}}(f, \tau) = \mathbf{W}(f) \mathbf{y}(f, \tau) \quad (10)$$

I coefficienti dei filtri di deconvoluzione $w_{ij}(l)$ saranno infine ottenuti mediante antitrasformazione degli elementi $W_{ij}(f)$ in modo da ottenere la stima di $\mathbf{x}(t)$ come:

$$\bar{\mathbf{x}}(t) = \mathbf{W}(l) * \mathbf{y}(t) \quad (11)$$

dove $\mathbf{W}(l)$ è una matrice di filtri di deconvoluzione $N \times N$ parametrizzata in l .

Sebbene l'approccio in frequenza permetta l'utilizzo di filtri di deconvoluzione molto lunghi ad un basso costo computazionale ci sono diversi problemi che necessitano di essere risolti legati all'ambiguità della soluzione calcolata attraverso l'ICA.

4. AMBIGUITA' NELLA BSS IN FREQUENZA

Nella sezione 2 abbiamo mostrato come la separazione possa essere effettuata massimizzando l'indipendenza statistica delle sorgenti. La soluzione per la matrice $\mathbf{W}$ in ogni caso non è unica. Infatti definita una matrice diagonale di scalari $\mathbf{C}$:

$$\mathbf{C} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} \quad (12)$$

Ogni matrice $\mathbf{CW}$ è sempre una soluzione secondo ICA in quanto l'indipendenza statistica dei segnali è invariante rispetto al prodotto con degli scalari. Nel dominio delle frequenze gli scalari sono dei numeri complessi e per ogni frequenza le matrici di separazione possono essere scalate in modo non coerente. La risoluzione dell'ambiguità del modulo dei coefficienti c_i permette di non alterare il contenuto spettrale dei segnali originali mentre la corretta stima della fase dei coefficienti c_i può essere interpretata come una corretta ricostruzione del segnale privo di riverberazione. Un esatto calcolo dei coefficienti non è possibile ma è possibile determinare una matrice di scaling che minimizzi la distorsione sui segnali. Un approccio comunemente adottato è quello del Minimal Distortion Principle (MDP)[9].

Analogamente se supponiamo di avere una matrice di permutazione $\mathbf{\Lambda}$ definita ad esempio come:

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (13)$$

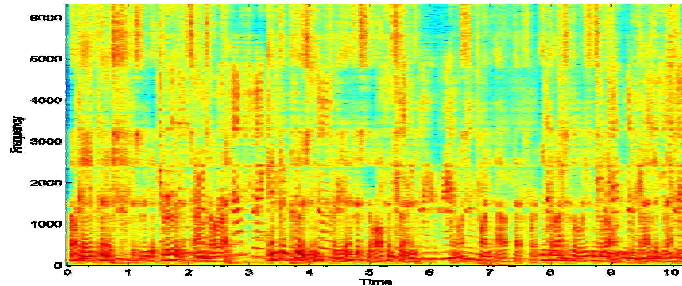


Figura 2: Tipica situazione di disallineamento a blocchi

anche in questo caso ogni matrice $\mathbf{A}\mathbf{W}$ è sempre soluzione ICA in quanto la permutazione corrisponde solo ad una differente associazione tra le righe di input e di output ma l'indipendenza statistica è sempre verificata. Ricordiamo che nella BSS nel dominio del frequenze la separazione è effettuata in modo indipendente e quindi l'allineamento degli involuppi di frequenza di uscita non è necessariamente costante. In altre parole se non è applicato nessun metodo di allineamento che risolva il problema delle permutazioni, non c'è nessuna certezza che tutte le frequenze del segnale separato in uscita $\bar{x}_i(t)$ appartengano alla stessa sorgente $x_i(t)$. Un esempio tipico di permutazione è mostrato in figura 2. Si può notare che differenti bande dello spettro sono associate evidentemente a segnali differenti.

Nel caso delle permutazioni non esiste una soluzione ampiamente riconosciuta e in questi ultimi anni numerosi approcci sono stati proposti in quanto la strategia di risoluzione delle permutazioni è quella che maggiormente determina le prestazioni totali del sistema. Un ulteriore problema della BSS in frequenza è la presenza di spike dovuti alla circolarità dell'FFT. Essendo infatti i filtri in frequenza a supporto limitato i filtri di deconvoluzione nel tempo dovrebbero essere teoricamente di lunghezza infinita con andamento periodico mentre nella pratica sono troncati ad L . Per risolvere tale problema generalmente si effettua un filtraggio delle risposte in frequenza basate su finestrazione di hanning[18].

5. RISOLUZIONE DELLE PERMUTAZIONI

Mentre i problemi dello scaling e della circolarità sono risolti mediante tecniche largamente riconosciute, per le permutazioni non c'è una soluzione ufficialmente utilizzata e efficace per tutte le situazioni. In generale esistono diversi approcci per risolvere tale ambiguità che possono essere classificati in 3 categorie principali:

- Allineamento basato sulla DOA (direction of arrivals). Gli involuppi di frequenza sono raggruppati secondo informazioni di fase. Infatti la disposizione delle sorgenti in punti differenti dello spazio definisce differenti DOA che possono essere stimate dagli involuppi di frequenza separati o direttamente dalle matrici di separazione $\mathbf{W}(f)$ [16].

- Allineamento basato su correlazione di ampiezza. Tale metodo utilizza informazioni di correlazione statistica tra l'energia di involuppi adiacenti, nell'ipotesi quindi che i segnali siano correlati in frequenza (come nel caso del parlato) [15][16].

- Coerenza delle matrici di separazione. Suppone che la risposta impulsiva dell'ambiente non presenti discontinuità evidenti e quindi tale continuità si propaghi anche sulle matrici di separazione[14]. Tale metodo è strettamente correlato con quello basato sulla DOA ma non effettua una stima esplicita delle DOA delle sorgenti.

Nel nostro algoritmo di BSS abbiamo sviluppato due tecniche complementari basate sugli ultimi due approcci. In [10] abbiamo mostrato come le matrici di separazione $\mathbf{W}(f)$ varino con continuità lungo le frequenze. La fase del rapporto degli elementi di una riga di $\mathbf{W}(f)$ è infatti strettamente connessa con il modello di propagazione dell'onda acustica delle sorgenti e teoricamente varia linearmente lungo le frequenze. In [11] abbiamo mostrato che se ad una determinata frequenza l'ICA è inizializzato con la matrice di separazione $\mathbf{W}(f-I)$ ottenuta per una frequenza adiacente, ci sarà un'alta probabilità che l'ICA converga verso

una soluzione che ha una permutazione coerente lungo tutte le frequenze. Per evitare le permutazioni l'approccio da seguire potrebbe essere quello di effettuare la separazione a partire dalle frequenze più alte verso quelle più basse operando ad ogni frequenza la ricorsione. In ogni modo le matrici di separazione $\mathbf{W}(f)$ ottenute da ICA possono essere considerate come un'osservazione rumorosa delle matrici reali:

$$\overline{\mathbf{W}}(f) = \mathbf{W}(f) + \mathbf{N}(f) \quad (14)$$

Tale rumore è principalmente introdotto dalla riverberazione che introduce delle non linearità della fase anche lungo frequenze adiacenti. Per questo motivo metodi di inizializzazione ricorsiva come quello in [14] falliscono in presenza di alta riverberazione o di semplici errori locali di stima della $\mathbf{W}(f)$. Per migliorare questo approccio in [10] abbiamo mostrato come la $\mathbf{W}(f)$ possa essere stimata ricorsivamente mediante un filtraggio Least Mean Square che minimizzi le fluttuazioni del determinante delle $\mathbf{W}(f)$. In questo modo l'ICA lungo le frequenze è ricorsivamente inizializzato con una stima "smooth" di $\mathbf{W}(f)$ evitando quindi che discontinuità locali possano generare errori di allineamento che si propagano lungo le frequenze successive.

Tale metodo si è dimostrato molto efficace e permette di ottenere notevoli risultati anche in condizioni di riverberazione molto critiche e con piccole porzioni di segnale osservato. Ad ogni modo se le sorgenti sono molto vicine e se i microfoni sono distanziati in modo da ottenere aliasing spaziale la stima delle $\mathbf{W}(f)$ può non essere accurata e può quindi generare disallineamenti a blocchi su bande dello spettro (come quelli mostrati in figura 2). È necessario quindi risolvere l'ambiguità con una tecnica complementare.

Come tecnica alternativa è possibile utilizzare un approccio basato sulla correlazione dell'energia dello spettro lungo frequenze adiacenti. Infatti per la maggior parte dei segnali acustici osservati e in particolare per il parlato esiste un'alta correlazione interfrequenziale. Un approccio molto comune è quello di calcolare il coefficiente di correlazione tra gli involucri di frequenza adiacenti e di determinare la permutazione che massimizza la correlazione totale. Anche in questo caso però un errore locale può ripercuotersi negativamente verso tutte le frequenze successive. Una soluzione è di effettuare una stima ricorsiva dell'involuppo in frequenza e di utilizzare tale involuppo come base di riferimento per l'allineamento. Tale tecnica presentata in [12] si è mostrata efficace anche per la separazione di sorgenti posizionate lungo una stessa direzione angolare la quale impedisce ovviamente l'utilizzo di tecniche basate sulla stima del modello di propagazione.

Anche in questo caso però tale tecnica può diventare instabile se in una porzione dello spettro ripetuti errori modificano la base di riferimento generando di conseguenza errori di allineamento nelle frequenze successive.

Per quanto siano efficaci entrambi i metodi, essendo basati su di un approccio ricorsivo sono sensibili ad errori locali che possono propagarsi lungo le frequenze successive. È necessario quindi minimizzare la probabilità di propagazione di un errore locale. Per i succitati motivi proponiamo un nuovo metodo di risoluzione delle permutazioni che combina entrambe le tecniche minimizzando gli errori di decisione. Tale metodo è basato su di uno schema di due "decisori interlacciati" in cui gli errori di decisione sono corretti reciprocamente. In tale schema supponiamo che gli errori di decisione siano incorrelati. In figura 3 mostriamo lo schema interlacciato di decisione delle permutazioni:

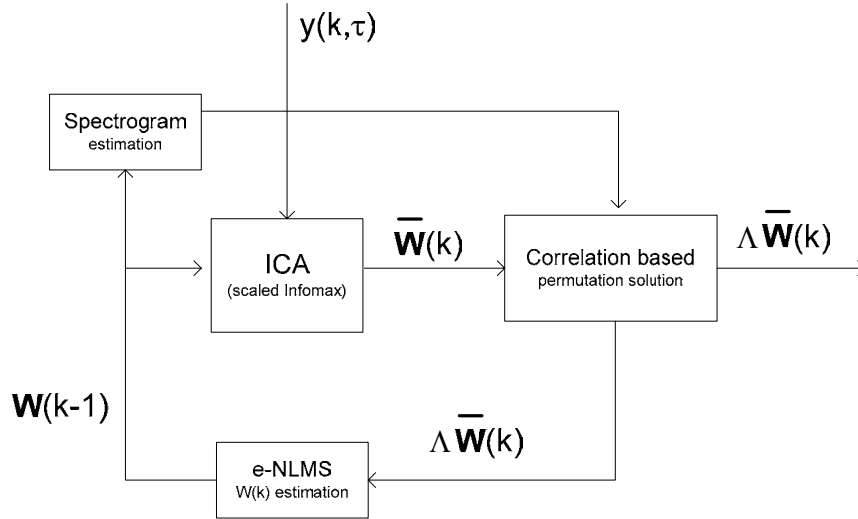


Figura 3: Schema interlacciato di correzione delle permutazioni

Ad ogni bin di frequenza k l'ICA genera una matrice di separazione $\bar{\mathbf{W}}(k)$. Il metodo basato sulla correlazione delle energia tra gli involucri determina la permutazione ottimale Λ . La matrice $\Lambda \bar{\mathbf{W}}(k)$ è usata come osservazione per l'algoritmo che stima la matrice $\mathbf{W}(k)$. In tal modo un errore sulla permutazione è corretto dal blocco basato sulla stima LMS di $\mathbf{W}(k)$. La matrice risultante, indicata con $\mathbf{W}(k-1)$ è usata per inizializzare l'ICA e per ristimare lo spettro. Analogamente se il blocco "e-NLMS" genera un errore, al passo successivo il blocco di decisione "Correlation based" riallineerà le matrici evitando che la permutazione errata si propaghi lungo tutte le frequenze. Di conseguenza è facile comprendere che la probabilità che ci sia una permutazione errata sia:

$$p(e_{\text{permutazione}}) = p(e_1, e_2) = p(e_1) p(e_2) \quad (15)$$

dove $p(e_1)$ e $p(e_2)$ sono le probabilità di errore dovute ai singoli metodi di decisione in cui abbiamo supposto che i due metodi generino degli errori incorrelati.

Tale tecnica combina in modo ottimale i due metodi di allineamento i quali essendo complementari utilizzano tutta l'informazione a disposizione.

6. RISULTATI SPERIMENTALI

Il sistema è stato implementato per funzionare con un numero generico di sorgenti. In questo lavoro mostriamo i risultati numerici ottenuti nella separazione di due sorgenti in differenti condizioni di riverberazione ($T_{60} = 160$ ms e $T_{60} = 700$ ms). La disposizione dei microfoni e delle sorgenti è indicata in figura 4.

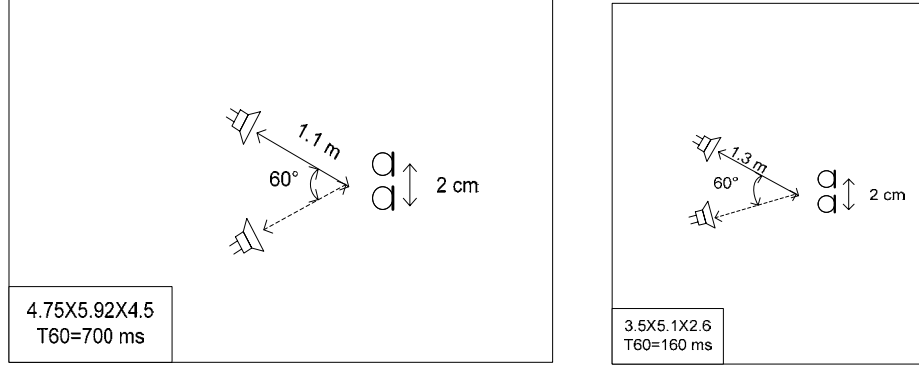


Figura 4: Configurazione dei test reali effettuati in camere con differenti T_{60}

Inoltre l'algoritmo è stato applicato a registrazioni audio di altri ricercatori e disponibili liberamente on-line. Per tali registrazioni non abbiamo valutazioni numeriche dato che i segnali parziali di riferimento non sono disponibili. Ad ogni modo in tutti i casi la separazione è stata efficace e il metodo si è dimostrato robusto anche in casi in cui è presente aliasing spaziale o quando le sorgenti sono posizionate lungo una stessa direzione angolare. Esempi di separazione sono disponibili on-line in [19].

Le prestazioni del sistema sono state calcolate misurando il Signal Interference Ratio (SIR) dell' i -esima sorgente di segnale come:

$$SIR_i : 10 \log_{10} \frac{\langle \left| \sum_l u_{ii}(l) x_i(t-l) \right|^2 \rangle_t}{\langle \left| \sum_{k \neq i} \sum_l u_{ik}(l) x_k(t-l) \right|^2 \rangle_t} \text{ (dB)} \quad (15)$$

$$u_{ik}(l) = \sum_{j=1}^N \sum_{\tau=0}^{L-1} w_{ij}(\tau) h_{jk}(l-\tau) \quad (16)$$

dove indichiamo con $u_{ik}(l)$ la risposta impulsiva tra la sorgente $x_k(t)$ e il segnale separato $\bar{x}_i(t)$. Il SIR riportato in tabella 1 è quello mediato su tutte le sorgenti.

Configurazione	SIR medio dopo la BSS
2 sorgenti (T60=160 ms)	18.1 dB
2 sorgenti (T60=700 ms)	12.5 dB

Tabella 1 – Prestazioni in termine di Signal Interference Ratio

I segnali utilizzati hanno una durata di 9 secondi e sono stati campionati con $f_s=16$ kHz. Nel test con $T_{60} = 160$ ms è stata applicata una la STFT con finestre di hanning di 4096 punti. Nel test con $T_{60} = 700$ ms si sono utilizzate finestre lunghe 8192 punti per ottenere filtri di deconvoluzione più lunghi in grado di coprire un tempo di riverberazione maggiore. I valori indicati in tabella 1 mostrano come i livelli di separazione siano molto elevati anche nel caso di condizioni di riverberazione molto avverse. Questo indica che, sebbene la BSS in frequenza sia intrinsecamente e asintoticamente limitata dalla riverberazione, è possibile comunque ottenere performance accettabili se il problema delle permutazioni è risolto correttamente.

7. CONCLUSIONI E ATTIVITA' FUTURE

In questo lavoro è stato presentato un nuovo metodo per la risoluzione del problema delle permutazioni che emerge nell'approccio in frequenza della blind source separation. E' stato formulato un nuovo schema interlacciato che permette di combinare i vantaggi di due metodi complementari di risoluzione delle permutazioni, presentati in lavori precedenti. Tale schema è stato utilizzato per implementare un algoritmo di BSS in grado di separare sorgenti acustiche in ambienti molto riverberanti e i risultati sperimentali confermano la validità dell'approccio. Avendo osservato che tale tecnica è molto robusta anche per la separazione di segnali acustici molto brevi, stiamo sviluppando un'estensione per il caso di sorgenti in movimento in grado di separare delle sorgenti acustiche in tempo reale.

BIBLIOGRAFIA

- [1] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.
- [2] S. Araki, S. Makino, R. Mukay, T. Nishikawa, H. Saruwatari "Fundamenta limitation of frequency domain blind source separation for convolved mixture of speech
- [3] Cardoso, J.-F. (1997). "Infomax and maximum likelihood for source separation." *IEEE Letters on Signal Processing*, 4:112–114.
- [4] Douglas, S.C. Gupta, M. "Scaled Natural Gradient Algorithms for Instantaneous and Convolutional Blind Source Separation", *ICASSP 2007*, vol II, pp 637-640, Apr. 2007
- [5] M. A. Casey and A. Westner. "Separation of mixed audio sources by independent subspace analysis." In *Proceedings of International Computer Music Conference*, Berlin, Germany, 2000.
- [6] Lee and H.S. Seung. "Algorithms for non-negative matrix factorization" In *Proceedings of Neural Information Processing System* pages 556-562, Denver, USA 2001

- [7] Makeig S, Bell AJ, Jung T-P, Sejnowski TJ, "Independent Component Analysis of Electroencephalographic data", *Advances in Neural Information Processing Systems*, 1996
- [8] McKeown MU, Makeig S, Brown GG, Jung PT, Kindermann SS, Bell AJ, Sejnowski TJ (1998): "Analysis of fMRI data by blind separation into independent component analysis", *Human Brain Mapping* 6: 160-188.
- [9] K.Matsuoka, S. Nakashima, "Minimal distortion principle for blind source separation" in: International Symposium on ICA and Blind Signal Separation, San Diego, CA, USA, December 2001
- [10] Francesco Nesta, Piergiorgio Svaizer, Maurizio Omologo, "Separating short signals in highly reverberant environment by a recursive frequency-domain BSS", in proceedings of HSCMA, Trento, May 2008.
- [11] Francesco Nesta, Piergiorgio Svaizer, Maurizio Omologo, "A BSS method for short utterances by a recursive solution to the permutation problem", in proceedings to SAM, Darmstadt, July 2008.
- [12] Francesco Nesta, Maurizio Omologo, "A recursive high order statistics approach to solve the permutation problem in frequency-domain BSS", in proceedings of Toni Mian Workshop, October 2007
- [13] M. S. Pedersen, J. Larsen, U. Kjems, L. C. Parra, "A Survey of Convolutional Blind Source Separation Methods", Springer Handbook of Speech, Nov. 2007.
- [14] D.-T. Pham, Ch. Serviere, and H. Boumaraf, "Blind separation of convolutional audio mixtures using nonstationarity," in Proceedings of the 4th International Conference on Independent Component Analysis and Blind Signal Separation (ICA '03), pp. 981-986, Nara, Japan, April 2003.
- [15] D.-T. Pham, Ch. Serviere, and H. Boumaraf, "Blind separation of speech mixtures based on nonstationarity," in Proceedings of 7th International Symposium on Signal Processing and Its Applications (ISSPA '03), vol. 2, pp. 73-76, Paris, France, July 2003.
- [16] H. Sawada, R. Mukai, S. Araki, S. Makino, "Frequency-Domain Blind source separation" in Speech Enhancement (J. Benesty, S. Makino J. Cheen), ch. 13, pp. 299-327, Springer, 2005
- [17] H. Sawada, R. Mukai, S. Araki, S. Makino. "A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation." IEEE transaction on speech and audio processing vol.12. no.5 september 2004
- [18] H. Sawada, R. Mukai, S. de la Kethulle de Ryhove, S. Araki, S. Makino "Spectral smoothing for frequency-domain blind source separation", IWAENC 2003, Sept. 2003, Kyoto, Japan
- [19] <http://shine.fbk.eu/people/nesta/>