

IL PROCESSO DI AURALIZZAZIONE: METODOLOGIA ED ESEMPLIFICAZIONE

A. Farina, P. Galaverna, M. Giabbani

Dipartimento di Ingegneria Industriale, Università di Parma, via delle Scienze – 43100 PARMA

SOMMARIO

In questo lavoro viene descritto il procedimento necessario ad ottenere l'auralizzazione di segnali sonori anecoici, facendo impiego delle risposte all'impulso rilevate sperimentalmente in un teatro. Tale procedimento consente di ascoltare l'effetto prodotto dal teatro in esame, che può così venire facilmente confrontato con quello di altri teatri.

Dopo la spiegazione della teoria matematica impiegata (convoluzione), vengono presentati alcuni esempi di applicazione della tecnica dell'auralizzazione ad alcuni famosi teatri italiani.

INTRODUZIONE

Storicamente la tecnica dell'auralizzazione deriva da quella, usata ormai da oltre 30 anni, di registrazione stereofonica mediante microfoni binaurali collocati su una testa artificiale [1,2]. Tale tecnica consisteva nella registrazione di rappresentazioni musicali nei teatri, che potevano venire ascoltate con grande fedeltà mediante riproduzione in cuffia.

Ad un certo punto ci si rese conto che era possibile separare la registrazione dell'evento artistico dalla sua riproduzione in un particolare spazio: se infatti la musica ed il canto vengono preventivamente registrati in una camera anecoica, è possibile aggiungere in seguito l'effetto di un particolare spazio acustico, in questo caso un teatro, mediante un opportuno filtraggio numerico di tali segnali anecoici. In tal modo può venire ricreata con grande realismo la sensazione di ascolto che si avrebbe nel teatro stesso, e questo per qualsiasi segnale anecoico si voglia riprodurre, senza la necessità di doverlo registrare di nuovo nel teatro stesso [3]. Tale procedimento è stato denominato Auralizzazione, in analogia al procedimento di visualizzazione che consente di riprodurre immagini anziché suoni.

Partendo dalla originale implementazione del metodo, che prevedeva l'utilizzo di risposte all'impulso binaurali come filtri numerici per l'auralizzazione, e che dunque era idoneo solo alla riproduzione in cuffia, si sono avuti, in questi ultimi due anni, significativi progressi: oggi è possibile realizzare filtri numerici idonei alla riproduzione del segnale auralizzato su una normale coppia di altoparlanti stereo, meglio se posti a breve distanza fra loro (stereo dipole [4]); inoltre, se nel teatro erano state misurate risposte all'impulso B-Format, è possibile utilizzare le stesse per la realizzazione di una completa riproduzione del segnale auralizzato secondo la metodica Ambisonics, che prevede un elevato numero di altoparlanti posti intorno alla zona di ascolto, con corretta riproduzione dell'informazione tridimensionale (compresi i suoni che provengono

dall'alto o dal basso).

Alla base di tutti questi effetti, comunque, sta una unica, semplice formulazione matematica, basata sulla **convoluzione** fra i segnali anecoici ed i filtri numerici (o risposte all'impulso, che dir si voglia). Nel seguito viene presentata la teoria matematica della auralizzazione, nelle sue varie implementazioni.

TEORIA: CONVOLUZIONE

La convoluzione di un segnale continuo $x(\tau)$ con un filtro lineare caratterizzato dalla risposta all'impulso $h(\tau)$ produce un segnale d'uscita $y(\tau)$ dato dalla ben nota relazione:

$$y(\tau) = x(\tau) * h(\tau) = \int_0^{\infty} x(\tau - t) \cdot h(t) dt \quad (1)$$

Quando il segnale in ingresso e la risposta all'impulso sono segnali digitali, e quindi campionati ($\tau = i \cdot \Delta\tau$), e la risposta all'impulso ha una lunghezza finita N , tale integrale si riduce ad una sommatoria:

$$y(i) = \sum_{j=0}^{N-1} x(i - j) \cdot h(j) \quad (2)$$

Bisogna effettuare la somma di N prodotti per ciascun dato campionato, e ciò produce un numero molto alto di somme e moltiplicazioni. Questo calcolo avviene in virgola mobile per evitare overflow ed un rumore numerico eccessivo. Per tali ragioni la "direct true convolution" in tempo reale, attualmente si può effettuare solo nei casi di risposte all'impulso lunghe poche centinaia di punti, mentre per descrivere in modo soddisfacente la risposta all'impulso di una tipica sala da concerto ci vogliono almeno 100000 punti (alla frequenza di campionamento di 44.1 kHz).

Comunque il calcolo della convoluzione può essere significativamente semplificato sfruttando la *FFT* (Fast Fourier Transform) e la *IFFT* (Inverse Fast Fourier Transform), poiché una convoluzione nel dominio del tempo si riduce ad una semplice moltiplicazione nel dominio della frequenza tramite lo spettro complesso di Fourier del segnale di ingresso e della risposta all'impulso. Poiché l'algoritmo di *FFT* suppone che il segmento di segnale analizzato sia periodico, la semplice implementazione del calcolo nel dominio della frequenza produce risultati insoddisfacenti: bisogna rimuovere dal segnale d'uscita la periodicità causata dall'*FFT*.

Questo può essere fatto mediante due algoritmi, chiamati "overlap-add" e "select-save". Si è qui utilizzato il secondo. La seguente figura mostra il diagramma di flusso del metodo:

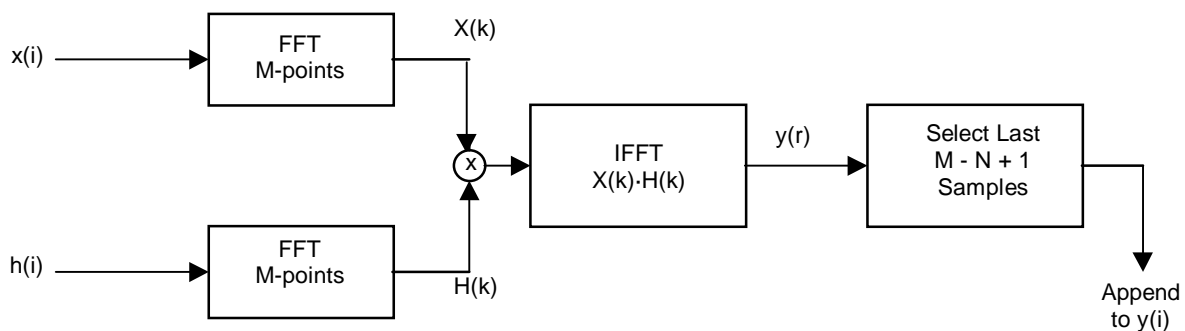


Fig. 1 - Flow Chart dell'algoritmo di Select-Save.

Poiché l'algoritmo produce in uscita come risultato solo $M-N+1$ dati convoluti, la finestra iniziale di M punti deve essere shiftata verso destra sulla sequenza di ingresso esattamente di $M+N-1$ punti, prima di effettuare la convoluzione del segmento successivo.

La contropartita è che è richiesta una *FFT* di lunghezza $M > N$. Tipicamente, il fattore 4 ($M = 4 \cdot N$) è quello che consente la maggior efficienza all'algoritmo "select save": se N è 65536 (2^{16}), è necessario calcolare la *FFT* su sequenze di dati di lunghezza $M = 65536 \cdot 4 = 262144$ punti. Viene quindi richiesta un'allocazione di memoria molto grande, tipicamente 1 Mbyte, per memorizzare la sequenza di ingresso o il segnale d'uscita. Il fabbisogno totale di memoria per l'intero algoritmo di "select save" è quindi di qualche Mbytes. Questa è la ragione per la quale questo non può essere facilmente implementato sulle diffusissime schede DSP di basso costo, che non sono dotate, però, di tali quantità di memoria.

Principalmente, l'algoritmo di "select save" riduce di molto il numero di moltiplicazioni in virgola mobile necessarie per la convoluzione. Ciascuna *FFT* o *IFFT* richiede $M \cdot \log_2(M)$ moltiplicazioni: una coppia di *FFT* e *IFFT* produce comunque in uscita $\frac{3}{4} M$ nuovi dati, e così il numero di moltiplicazioni per ciascun dato in uscita è intorno a 50, invece di 65536.

Sino ad un anno fa circa il processo di convoluzione poteva venire realizzato in tempo reale solo tramite sofisticate e costosissime workstation [5], ma oggi sono disponibili appositi moduli addizionali (in gergo "plug-ins") per i principali softwares di manipolazione delle forme d'onda, che girano su qualsiasi PC di fascia alta [6,7,8]. Alcuni di tali softwares, inoltre, includono già un modulo di convoluzione [9,10]. L'unica limitazione nell'impiego di questi strumenti software è che, in generale, essi non consentono di gestire correttamente risposte all'impulso multicanale, cosa che rende più macchinosa l'implementazione delle tecniche di riproduzione descritte nei seguenti paragrafi.

AURALIZZAZIONE BINAURALE IN CUFFIA

Per l'effettuazione dell'auralizzazione con ascolto in cuffia è sufficiente disporre di risposte all'impulso binaurali, ottenute piazzando la testa artificiale in una idonea posizione nel teatro, e collocando l'altoparlante omnidirezionale in una o più posizioni significative. Per quanto riguarda l'opera lirica, è necessario disporre di almeno due risposte all'impulso binaurali, una con sorgente sul palcoscenico e l'altra con sorgente in fossa. La prima verrà utilizzata come filtro per la traccia audio contenente la voce dei cantanti, la seconda per la base orchestrale.

Ovviamente si ottengono risultati molto migliori, soprattutto in termini di effetto spaziale della riproduzione, se l'emissione dell'orchestra è divisa in alcune sezioni e se ciascuna di esse viene convoluta con una risposta all'impulso diversa, ottenuta dalla corrispondente posizione della sorgente sonora omnidirezionale. Analogamente, se sul palcoscenico sono presenti più cantanti, la traccia audio di ciascuno andrebbe convoluta con la risposta all'impulso binaurale corrispondente alla posizione da esso occupata.

La seguente figura illustra lo schema di tale multi-convoluzione:

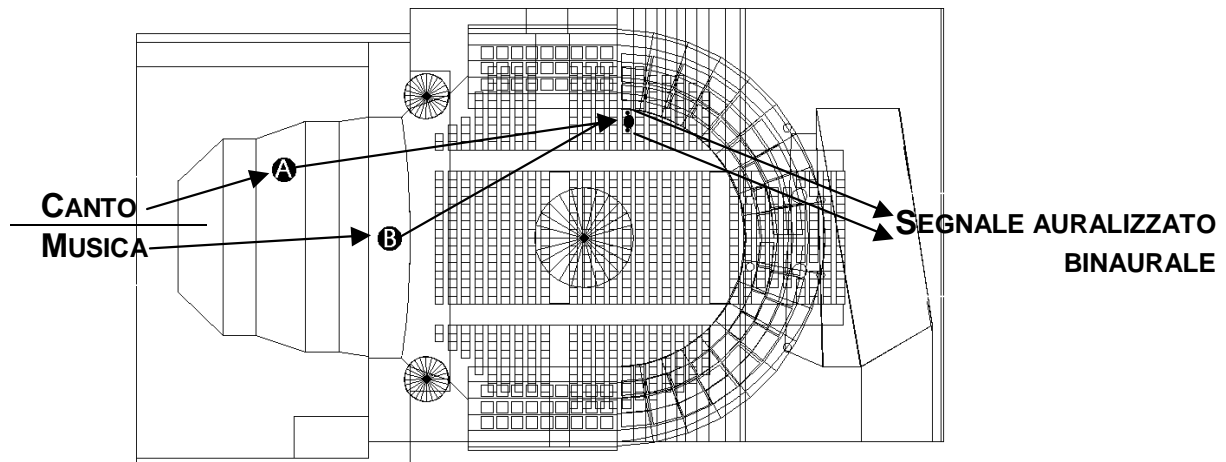


Fig. 2 – Schema di convoluzione multipla.

Va osservato che il segnale prodotto dalla convoluzione di un segnale anecoico con una risposta all'impulso binaurale è idoneo solo all'ascolto tramite cuffia. Ciononostante, occorre comunque applicare una idonea equalizzazione in frequenza, per tener conto del particolare modello di testa artificiale impiegata e di cuffia d'ascolto utilizzata, cosa che può essere fatta misurando dapprima la risposta della testa binaurale con la cuffia applicata su di essa, indi filtrando opportunamente i segnali convoluti in modo da compensare per tale risposta in frequenza. La seguente figura mostra la fase di misurazione della risposta all'impulso di una cuffia da riproduzione installata sulla testa binaurale.



Fig. 3 – Equalizzazione del complesso testa binaurale / cuffia di riproduzione.

Il risultato di tale misurazione è visibile nella seguente figura:

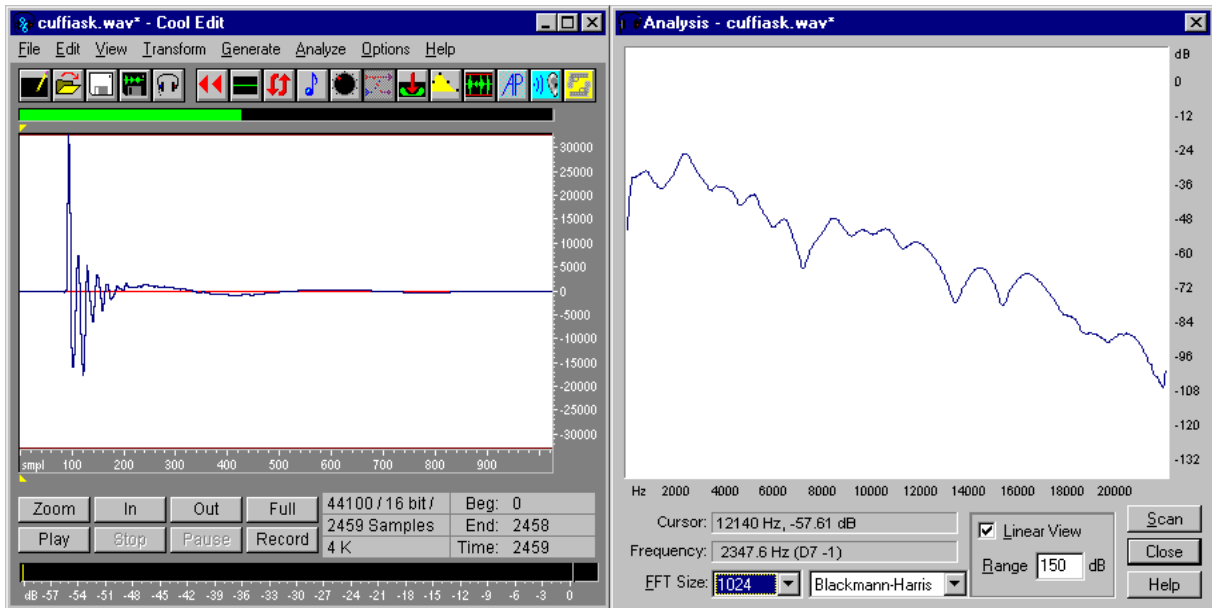


Fig. 4 – Risposta all'impulso ed in frequenza del complesso cuffia-testa.

L'equalizzazione della risposta in frequenza può venire ottenuta creando un opportuno filtro inverso a fase minima, secondo la ben nota procedura di Nelly e Allen [11]. La seguente figura illustra il filtro inverso così ottenuto, che può venire applicato al segnale binauralizzato mediante una ulteriore convoluzione.

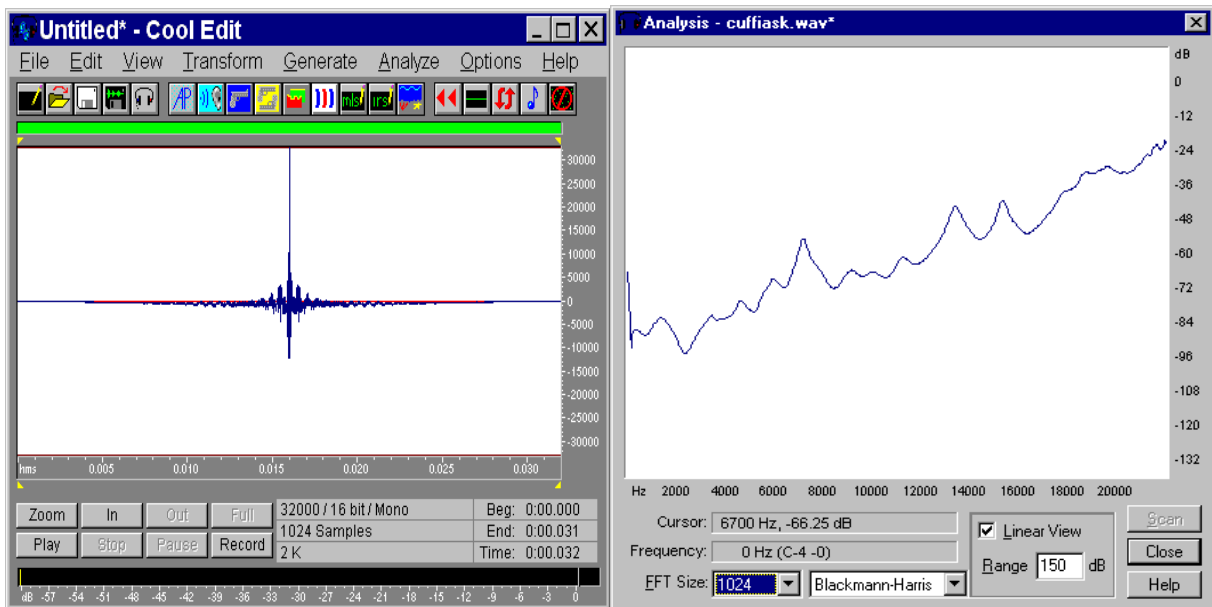


Fig. 5 – Filtro inverso per l'equalizzazione.

AURALIZZAZIONE BINAURALE TRAMITE ALTOPARLANTI STEREO

Una registrazione binaurale o una auralizzazione ad essa equivalente non può venire riprodotta come tale su una normale coppia di altoparlanti stereo, per due motivi:

- 1) il segnale proveniente dall'altoparlante sinistro finisce anche sull'orecchio destro, e viceversa.
- 2) il segnale binaurale contiene già le risposte in frequenza dell'orecchio esterno, che verrebbe in tal modo ri-applicata una seconda volta.

La figura 6 mostra ciò che accade quando una coppia di altoparlanti è posta di fronte ad un ascoltatore: il segnale proveniente da ciascun altoparlante raggiunge entrambe le orecchie, cosicché all'ingresso del canale uditivo i canali destro e sinistro arrivano mixati.

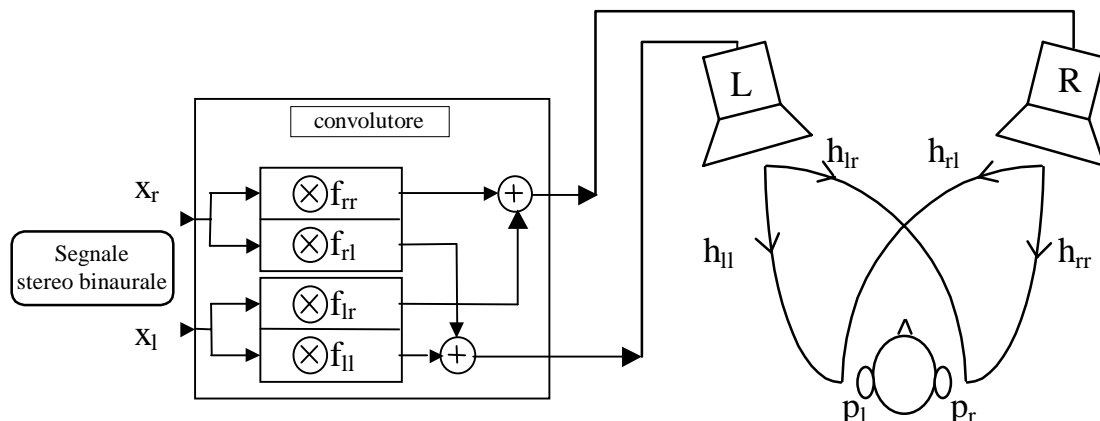


Fig. 6 - Percorso dei segnali che giungono alle orecchie attraverso gli altoparlanti.

In più la funzione di trasferimento relativa alla testa *HRTF* (Head Related Transfer Functions) è già compresa nelle risposte all'impulso usate per la convoluzione binaurale, ma il segnale che arriva dagli altoparlanti interferisce nuovamente con la testa dell'ascoltatore, e in questo modo il filtraggio dovuto alla testa viene operato due volte.

Per queste ragioni, è necessario usare un algoritmo che ritrasformi i segnali che arrivano alle orecchie, in modo che sull'orecchio sinistro arrivi solo x_l e sull'orecchio destro arrivi solo x_r .

Se sono state misurate le 4 risposte all'impulso fra gli altoparlanti e le orecchie (denominate rispettivamente h_{ll} , h_{lr} , h_{rl} ed h_{rr}), è possibile calcolare facilmente le 4 risposte all'impulso "inverse", denominate f_{ll} , f_{lr} , f_{rl} e f_{rr} , capaci di operare l'opportuno pre-filtraggio. In particolare, dopo semplici passaggi matematici (soluzione di un sistema di 4 equazioni in 4 incognite), si ottiene:

$$\begin{cases} f_{ll} = (h_{rr}) \otimes \text{InvDen} \\ f_{lr} = (-h_{lr}) \otimes \text{InvDen} \\ f_{rl} = (-h_{rl}) \otimes \text{InvDen} \\ f_{rr} = (h_{ll}) \otimes \text{InvDen} \\ \text{InvDen} = \text{InvFilter}(h_{ll} \otimes h_{rr} - h_{lr} \otimes h_{rl}) \end{cases} \quad (3)$$

Bisogna quindi misurare le quattro risposte all'impulso relative agli altoparlanti e alla testa dell'ascoltatore, il cui effetto è così completamente eliminato dal segnale audio da riprodurre. Ciò significa che, così facendo, viene eliminato non solo il crosstalk, ma anche qualsiasi altro filtraggio indesiderato dovuto alla risposta in frequenza degli altoparlanti. In più, viene eliminata

anche la funzione di trasferimento relativa alla testa, evitando il doppio filtraggio esaminato precedentemente. E' opportuno, a tal proposito, che venga utilizzata la stessa testa artificiale già impiegata per la misura delle risposte all'impulso nel teatro, in modo che anche la particolare scelta della testa binaurale venga resa ininfluente sul risultato.

In teoria questo post-processing dovrebbe anche eliminare le riflessioni dell'ambiente d'ascolto, ed in questo modo non sarebbe più necessario trattare acusticamente la sala. Comunque, l'esperienza mostra che avviene un fenomeno sfortunato: la presenza delle riflessioni dovute all'ambiente d'ascolto causa un effetto di *comb filter* nel dominio della frequenza, producendo difficoltà numeriche nel calcolo e risposte all'impulso modificate con ampi picchi che distruggono l'originale equilibrio della risposta in frequenza.

Esistono due soluzioni per risolvere tale limitazione: se l'ascolto non può svolgersi in una sala appositamente trattata, tramite i calcoli illustrati precedentemente, si fa in modo di estrarre solo la parte anecoica delle quattro risposte all'impulso sperimentali. In questo modo la compensazione digitale del *crossstalk* non tiene conto della presenza delle riflessioni dell'ambiente, ma esse sono ancora presenti, e causano ancora alterazioni spettrali e una certa quantità di *crossstalk*.

In alternativa si può usare una piccola camera anecoica: anche se non è perfettamente anecoica (in particolare alle basse frequenze), è sufficiente per minimizzare le instabilità numeriche nel filtraggio. In questo caso le quattro risposte all'impulso non sono troncate, ma la loro lunghezza è ridotta applicando una finestra semi-gaussiana, per tagliare fuori il rumore dalla coda riverberante.

In tutti i casi, comunque, il problema nell'impiego della teoria suddetta consiste nel calcolo del filtro inverso *InvDen*, poichè il segnale da invertire è un segnale causale, tempo-finito, a fase non minima (o, come si dice, *mixed-phase*). In generale non esiste una soluzione esatta al problema del filtraggio inverso di un sistema la cui risposta all'impulso è *mixed-phase*, ed è necessario utilizzare una tecnica di inversione numerica approssimata. Fra tutte quelle disponibili (Neely e Allen [11], Mourjopoulos [12], Nelson e Kirkeby [13]), l'ultima indicata si presenta come la più interessante. In pratica, il filtraggio inverso viene operato nel dominio della frequenza, prendendo a ciascuna frequenza il reciproco del numero complesso che esprime la funzione di trasferimento del sistema. Poichè può capitare che tale numero complesso sia molto piccolo, il suo reciproco potrebbe essere molto grande, causando instabilità nel filtro numerico corrispondente una volta ritrasformato nel dominio del tempo.

Per evitare tale comportamento, Nelson e Kirkeby suggeriscono di impiegare una tecnica di regolarizzazione, consistente nell'aggiungere una piccola quantità, positiva e reale, al denominatore in modo che il valore del reciproco resti sempre inferiore ad un certo massimo. La scelta del valore di tale piccola quantità, chiamata parametro di regolarizzazione ϵ , può essere operata "per tentativi" finchè si ottiene un filtro inverso ragionevole, visto che il calcolo dello stesso è rapidissimo.

In pratica, la relazione che fornisce il filtro inverso di Nelson e Kirkeby è la seguente:

$$K(f) = \frac{\text{Conj}[C(f)]}{\text{Conj}[C(f)] \cdot C(f) + \epsilon} \quad (4)$$

Una ulteriore estensione di tale metodo, tuttora in corso di sviluppo, prevede di utilizzare un parametro di regolarizzazione variabile con la frequenza $\epsilon(f)$, in modo da ottimizzare l'inversione nella parte centrale della banda audio, e "filtrando via" progressivamente le frequenze estreme (molto basse e molto alte) [14]. In tal modo, si riesce ad ottenere un filtro inverso con maggior capacità di filtraggio alle frequenze che servono di più, con risultati decisamente migliori in termini di qualità e naturalezza del filtraggio.

Anche se in base alla teoria su esposta la posizione dei due altoparlanti è ininfluenza sul risultato, in pratica si verifica che esiste una posizione ottimale. Essa però non coincide con il classico “triangolo stereo” suggerito per l’ascolto delle normali incisioni, in cui gli altoparlanti sono collocati in posizioni situate a $\pm 30^\circ$ rispetto all’asse longitudinale della testa dell’ascoltatore. La posizione ottimale per ottenere la migliore *cross-talk cancellation* si ha invece allorché i due altoparlanti sono molto vicini tra loro, con una apertura angolare di $\pm 5^\circ$. La seguente figura pone a confronto la disposizione standard degli altoparlanti di un sistema stereo tradizionale con lo “stereo dipole”.

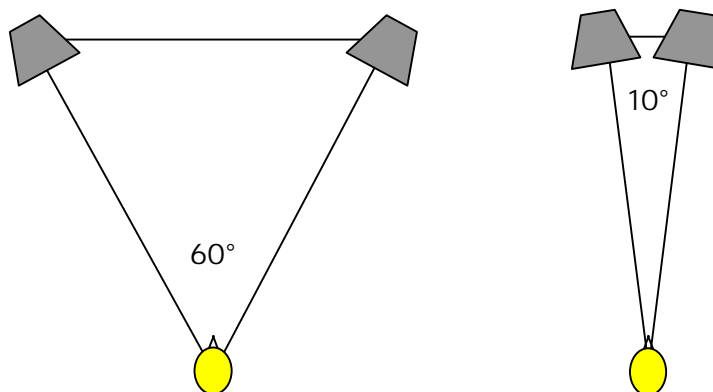


Fig. 7 – Configurazione stereo standard (sinistra) e “stereo dipole” (destra).

Va osservato infine che è sostanzialmente conveniente distribuire i segnali auralizzati direttamente nel formato “stereo dipole” anziché in quello binaurale: infatti in tal modo essi sono resi indipendenti dalla particolare testa artificiale impiegata (diventano in certo modo “universali”), e sono già pronti per l’ascolto su un economico sistema stereo senza la necessità di ulteriori filtri digitali. All’occorrenza, comunque, è estremamente facile riportare un segnale filtrato per la riproduzione su “stereo dipole” al formato binaurale originario, riconvolvendolo con il set di 4 risposte all’impulso (h_{ll} , h_{lr} , h_{rl} ed h_{rr}) che erano servite come base per il calcolo dei filtri inversi (e che dovrebbero sempre venire fornite a corredo del segnale “stereo dipole”, ad esempio come ultimo brano del CD).

AURALIZZAZIONE CON TECNICA B-FORMAT (Ambisonics)

Se il rilevamento della risposta all’impulso nel teatro era stato effettuato mediante un microfono Soundfield (o equivalente) [15], si hanno a disposizione risposte all’impulso di tipo B-format, costituite da un segnale omnidirezionale (W) corrispondente alla pressione acustica, e dai tre segnali provenienti dai microfoni a figura di 8 (XYZ), corrispondenti alle componenti cartesiane della velocità di oscillazione delle particelle d’aria. La seguente figura illustra il significato fisico del segnale B-format ottenuto dal microfono tridimensionale Soundfield.

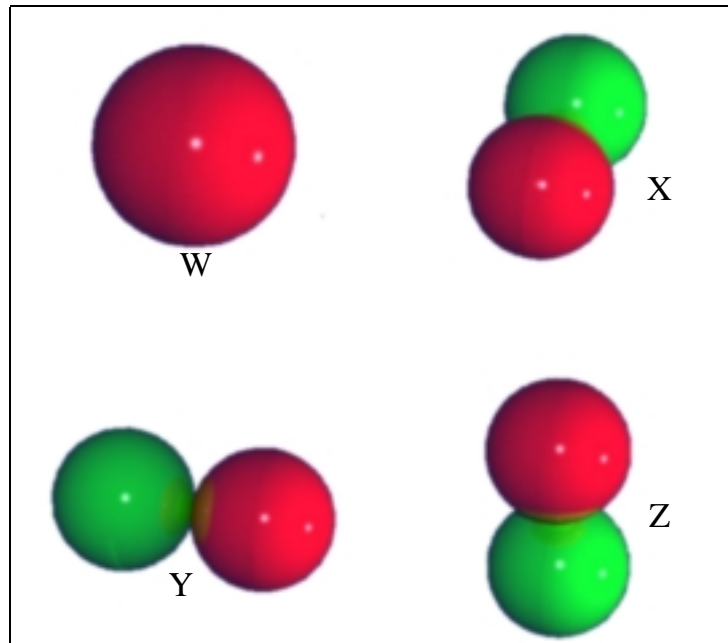


Fig. 8 – Diagrammi di direttività delle 4 componenti del segnale B-format.

Come nei casi precedenti, è possibile ricostruire una “registrazione virtuale” in formato B-format convolvendo opportuni segnali anecoici (musica e canto) con le risposte all’impulso B-format ottenute con diverse posizioni della sorgente, e mixando fra loro i risultati di tali convoluzioni. Il segnale B-format così ottenuto è sostanzialmente indistinguibile da una registrazione effettuata dal vivo nello stesso posto d’ascolto dello stesso teatro.

L’aspetto interessante di questa metodica è la possibilità di impiegare, per la riproduzione del segnale B-format, un sistema di altoparlanti configurato secondo la metodica Ambisonics. Si tratta di circondare un idoneo volume di ascolto con un adeguato numero di altoparlanti disposti uniformemente: la configurazione più semplice è quella di un cubo, con 8 altoparlanti identici posti nei vertici dello stesso.

Dal segnale B-format è agevole ricalcolare i segnali da inviare agli 8 altoparlanti. Utilizzando la numerazione degli stessi riportata nella seguente figura n. 9, e le tradizionali convenzioni sulla orientazione degli assi X, Y e Z, nonché il fatto che il microfono Soundfield produce un segnale W con un guadagno ridotto di 3 dB rispetto a quello degli altri 3 canali, si ottiene questo insieme di relazioni per ricostruire i segnali di alimentazione degli 8 altoparlanti:

$$\begin{aligned}
 F_1 &= W + X + Y + Z & F_5 &= W + X + Y - Z \\
 F_2 &= W - X + Y + Z & F_6 &= W - X + Y - Z \\
 F_3 &= W - X - Y + Z & F_7 &= W - X - Y - Z \\
 F_4 &= W + X - Y + Z & F_8 &= W + X - Y - Z
 \end{aligned}
 \tag{5}$$

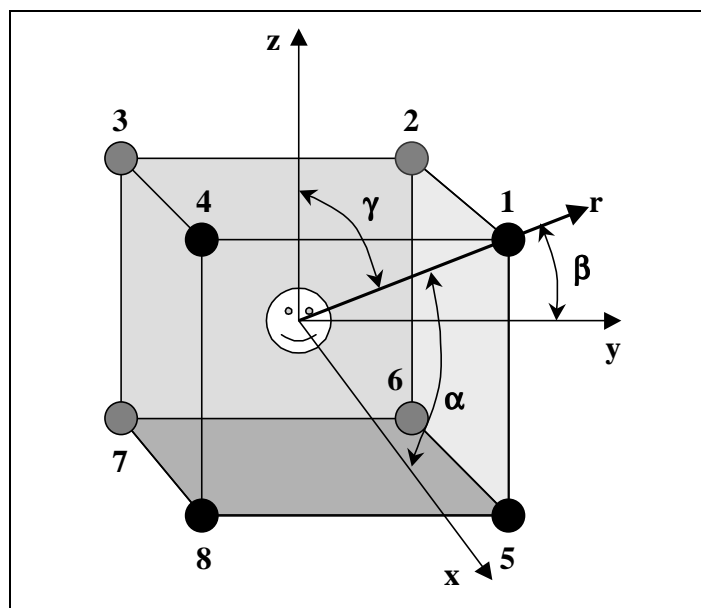


Fig. 9 – Geometria del sistema di riproduzione Ambisonics 3D ad 8 altoparlanti.

L'elaborazione è dunque molto semplice, qualunque software di manipolazione di segnali audio multitraccia (ad esempio, Cool Edit Pro oppure GoldWave) è in grado di processare i 4 segnali del B-format per produrre gli 8 segnali richiesti dal sistema di riproduzione Ambisonics. E' ovviamente possibile fare di meglio, utilizzando relazioni più complicate, se il sistema di altoparlanti è più complesso, compensando per la non uniforme posizione degli stessi o addirittura per l'impiego di altoparlanti non identici, come mostrato in [16].

Il principale vantaggio della riproduzione Ambisonics risiede nel fatto che l'ascolto tridimensionale diviene possibile per gruppi di ascoltatori anche numerosi (dipendentemente dal numero di altoparlanti e dalle caratteristiche del locale di ascolto), pur mantenendo sempre una elevata compattezza dell'informazione (sono necessari solo 4 canali anziché 2, indipendentemente dal numero di altoparlanti usati per la riproduzione).

E' inoltre possibile riconvertire il segnale B-format in un segnale binaurale, e da questo in un segnale "stereo-dipole", operando una decodifica "virtuale" del B-format su un sistema Ambisonics con un elevato numero di altoparlanti, e convolvendo il segnale indirizzato a ciascuno di essi con la risposta all'impulso binaurale corrispondente alla posizione angolare dello stesso rispetto ad un ascoltatore posto al centro del sistema di riproduzione.

Purtroppo invece non è possibile la conversione opposta: da una risposta all'impulso binaurale non può essere ricostruita facilmente la risposta all'impulso B-format. Questo è uno dei motivi per cui la caratterizzazione oggettiva dell'acustica di un teatro va fatta preferibilmente mediante la misura di risposte all'impulso B-format, anziché mediante risposte all'impulso binaurale.

ESEMPI APPLICATIVI

Viene qui presentato un confronto fra tre famosi teatri italiani: il teatro alla Scala di Milano, il teatro La Fenice di Venezia (ormai distrutto, come noto), ed il teatro Farnese di Parma. In tutti e tre i casi è stata effettuata la misurazione della risposta all'impulso binaurale, con un ricevitore collocato in platea, sulla sinistra e nella metà posteriore della stessa, mentre la sorgente sonora omnidirezionale era collocata in due posizioni: al centro del palcoscenico ed al centro della fossa d'orchestra.

La fig. 10 riporta il confronto fra le risposte all'impulso binaurali dei tre teatri, relativamente alla posizione della sorgente sonora sul palcoscenico.

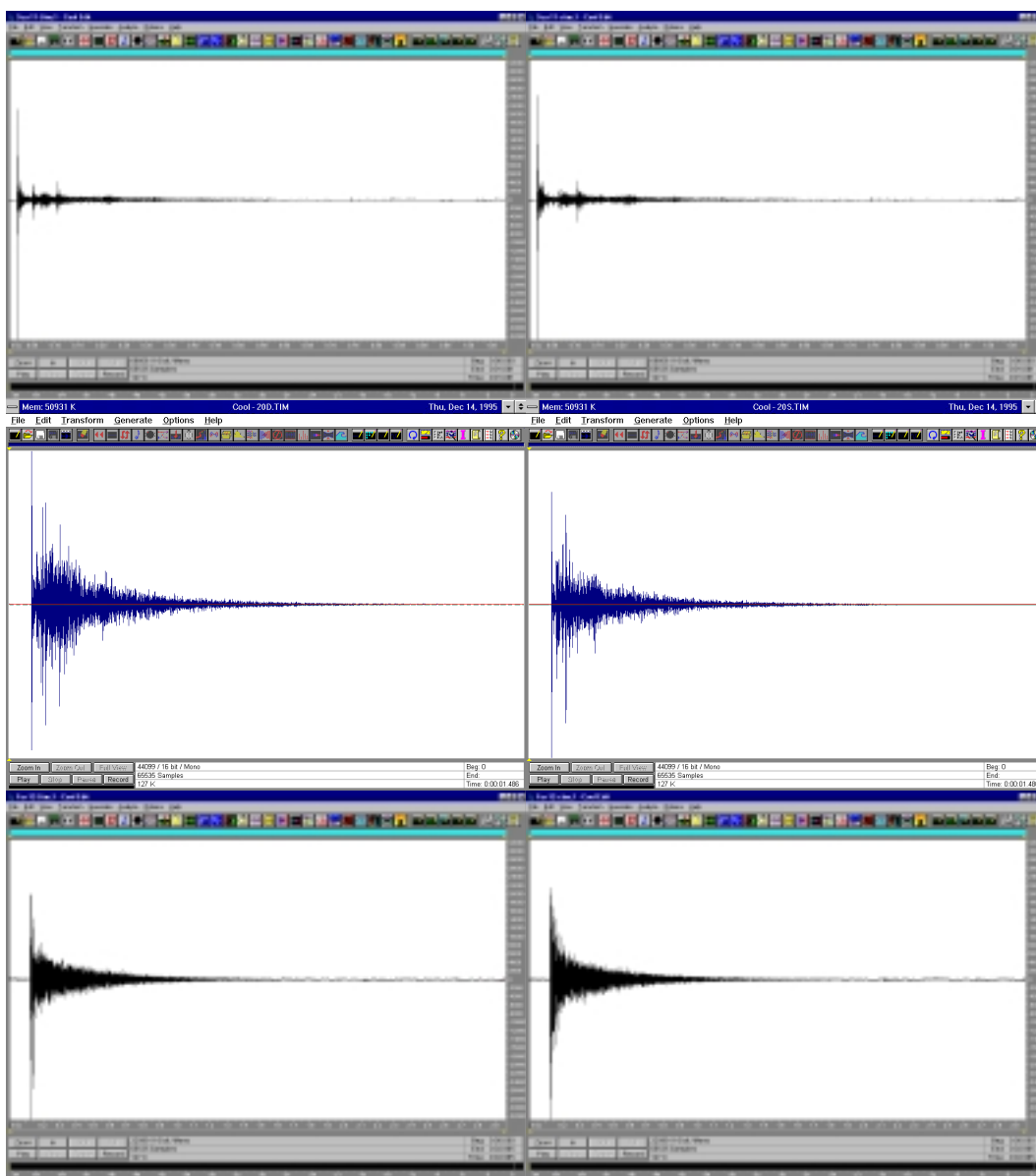


Fig. 10 – risposta all'impulso del teatro alla Scala (sopra), del teatro La Fenice (in mezzo) e del teatro Farnese (sotto).

Il segnale da convolvere è costituito da una registrazione stereofonica anecoica, in cui su una traccia è stata registrata la base musicale (Pianoforte, Lamberto Tronchin) e sulla seconda traccia il canto di una soprano solista (Loredana Bigi). Ovviamente la prima traccia è stata convoluta con la risposta all'impulso binaurale ottenuta con la sorgente in fossa, e la seconda con quella ottenuta con la sorgente sul palcoscenico. I risultati delle due convoluzioni vengono poi mixati in una unica traccia stereo binaurale, idonea per l'ascolto in cuffia. Il filtraggio "stereo-dipole" rende infine i segnali idonei all'ascolto tramite altoparlanti.

La seguente figura mostra l'interfaccia grafica del software di convoluzione impiegato per ottenere l'auralizzazione.

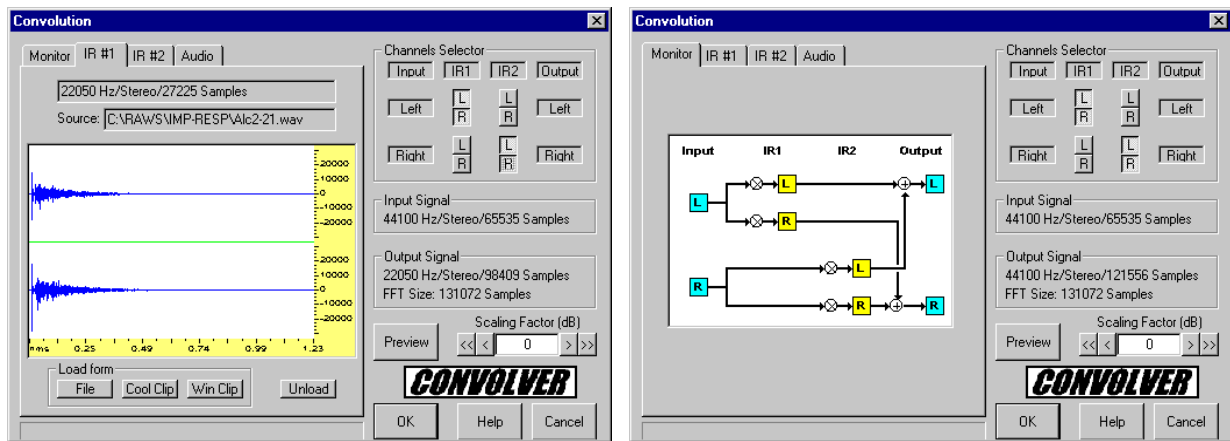


Fig. 11 – Software di convoluzione in tempo reale.

Ovviamente l’ascolto andrebbe eseguito in camera anecoica, o comunque in un ambiente trattato in modo che il suo tempo di riverbero sia almeno $\frac{1}{4}$ di quello del teatro che viene riprodotto; l’ascolto in un ambiente non trattato risulta come la sovrapposizione della sua coda sonora con quello del teatro da riprodurre, con conseguente esagerazione della stessa.

CONCLUSIONI

La tecnica dell’auralizzazione, implementata mediante convoluzione di segnali anecoici con risposte all’impulso sperimentali, consente di ricreare la sonorità degli spazi in cui queste ultime sono state acquisite. Le capacità di calcolo dei moderni computers rendono possibile il procedimento di convoluzione in tempo reale senza l’impiego di costosi sistemi dedicati, e quindi la tecnica dell’auralizzazione è oggi disponibile per tutti a costo molto basso.

Quello che ancora manca, per una sua diffusione generalizzata, è da un lato la disponibilità di un vasto repertorio di segnali anecoici, possibilmente in formato multi-traccia (con sezioni orchestrali e singole voci cantanti su tracce separate). Dall’altro lato, sebbene negli ultimi anni siano state misurate centinaia di risposte all’impulso in teatri e sale varie, esse sono state ottenute con metodiche non standardizzate: quindi le auralizzazioni ottenute sono difficilmente confrontabili fra loro.

Nel presente lavoro sono state comunque presentate due metodiche di auralizzazione standard, basate la prima su un formato stereofonico (a 2 canali) recentemente messo a punto, e denominato “stereo dipole”, la seconda sul classico B-format (Soundfield, Ambisonics) (4 canali). In entrambi i casi la riproduzione avviene mediante altoparlanti, ed i segnali in ciascuno dei due formati risultano indipendenti dalle periferiche impiegate nell’acquisizione delle risposte all’impulso, in particolare dallo specifico sistema microfonic impiegato.

La tecnica di produzione di segnali “stereo dipole” consente di “riciclare” il vasto repertorio di risposte all’impulso binaurali già esistente, purché sia ancora disponibile la particolare testa artificiale impiegata per la loro registrazione. Viceversa, per il futuro appare conveniente orientare l’acquisizione di dati sperimentali sulla metodica B-format, che promette una maggiore oggettività della misura stessa e nello stesso tempo è più idonea alla riproduzione per gruppi di ascolto numerosi.

BIBLIOGRAFIA

- [1] J. Baluert – “Spatial Hearing” – The MIT Press, Cambridge, MA, 1983.
- [2] H. Moller – “Fundamentals of Binaural Technology” – Applied Acoustics vol. 36 (1992) pp. 171-218.
- [3] Kleiner M., Dalenback B., Svensson P. - “Auralization - An Overview” - *JAES*, vol. 41 N.11, p.861-875, November 1993.
- [4] O. Kirkeby, P. A. Nelson, and H. Hamada – “Virtual Source Imaging Using the Stereo Dipole’”, Pre-prints of the 103rd AES Convention, New York, 26-29 September 1997.
- [5] D. McGrath, A. Reilly – “Huron - A Digital Convolution Workstation” - Audio Engineering Society - Sydney - The 5th Australian Regional Convention (1995).
- [6] Aurora Software v. 2.0 – [HTTP://www.ramsete.com/aurora](http://www.ramsete.com/aurora)
- [7] Acoustic Modeler – The Sonic Foundry – [HTTP://www.sonicfoundry.com](http://www.sonicfoundry.com)
- [8] Acoustec Software v. 1.0 – [HTTP://www.acoustec.com](http://www.acoustec.com)
- [9] Cool Edit Pro v. 1.1 – [HTTP://www.syntrillium.com](http://www.syntrillium.com)
- [10] Samplitude Studio 2496 – SEK'D – [HTTP://www.sekd.com](http://www.sekd.com)
- [11] S.T. Neely, J.B. Allen, “Invertibility of a room impulse response”, *J.A.S.A.*, vol.66, pp.165-169 (1979).
- [12] J.N. Mourjopoulos, “Digital Equalization of Room Acoustics”, *JAES* vol. 42, n. 11, 1994 November, pp. 884-900.
- [13] O. Kirkeby, P. A. Nelson, H. Hamada – “The "Stereo Dipole"-A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers” – *JAES* vol. 46, n. 5, 1998 May, pp. 387-395.
- [14] O. Kirkeby and P. A. Nelson – “Digital Filter Design for Virtual Source Imaging Systems”, *Pre-prints of the 104rd AES Convention*, Amsterdam, 15 - 20 May, 1998.
- [15] Farina A., Tronchin L., “3D Impulse Response measurements on S.Maria del Fiore Church, Florence, Italy” – Proc. of ICA98, International Conference on Acoustics, Seattle (WA) 20-26 June 1998.
- [16] Farina A., Ugolotti E. – “Software Implementation of B-Format Encoding and Decoding”, *Pre-prints of the 104rd AES Convention*, Amsterdam, 15 - 20 May, 1998.