

IL TRATTAMENTO DIGITALE DEGLI SPAZI PER LA MUSICA: UNA VERIFICA SPERIMENTALE

Angelo Farina (1), Lamberto Tronchin (1),(2)

(1) Dipartimento di Ingegneria Industriale, Università degli Studi di Parma

(2) D.I.E.N.C.A. – C.I.A.R.M., Università degli Studi di Bologna

E-MAILs: farina@pcfarina.eng.unipr.it - <http://pcfarina.eng.unipr.it>

tronchin@ciarm.ing.unibo.it - <http://ciarm.ing.unibo.it>

SOMMARIO

Nel presente lavoro viene analizzato la tecnica dell'auralizzazione degli spazi per la musica, sotto il punto di vista dell'analisi psicoacustica mirata alla determinazione della qualità acustica degli ambienti.

Il lavoro è stato condotto in collaborazione con l'Institute of Sound and Vibration Research - University of Southampton, UK.

Sono state effettuate registrazioni di brani musicali riprodotti all'interno delle camere acustiche disponibili presso l'ISVR; successivamente sono state misurate le risposte all'impulso nelle medesime posizioni di registrazione e convolute con lo stesso segnale anecoico diffuso negli ambienti di registrazione. Tali brani sono stati poi riproposti in camera anecoica a un campione di ascoltatori inglesi e confrontati. Da tale confronto soggettivo è emerso che la tecnica dell'auralizzazione è in grado di creare simulazioni pressochè indistinguibili dalle registrazioni dal vivo. Infine, è stato sviluppato un algoritmo di calcolo per l'inversione delle risposte all'impulso sperimentali, che è stato applicato alle risposte all'impulso determinate nell'ambiente usato per le prove di ascolto, rendendo possibile la rimozione della riverberazione ambientale dalle registrazioni.

TEORIA

Nello studio della correlazione fra parametri acustici oggettivi e sensazioni soggettive legate all'ascolto musicale, si sono sovente impiegate riproduzioni binaurali di registrazioni di brani musicali effettuate in diversi ambienti, allo scopo di agevolarne la comparazione. Sebbene tale tecnica sia di per sè estremamente più pratica e proficua rispetto al trasferimento degli ascoltatori da una sala all'altra, tuttavia presenta alcuni problemi, legati alla necessità di effettuare registrazioni binaurali di durata adeguata, immuni dal rumore di fondo, e con un vasto assortimento di tipi di segnale.

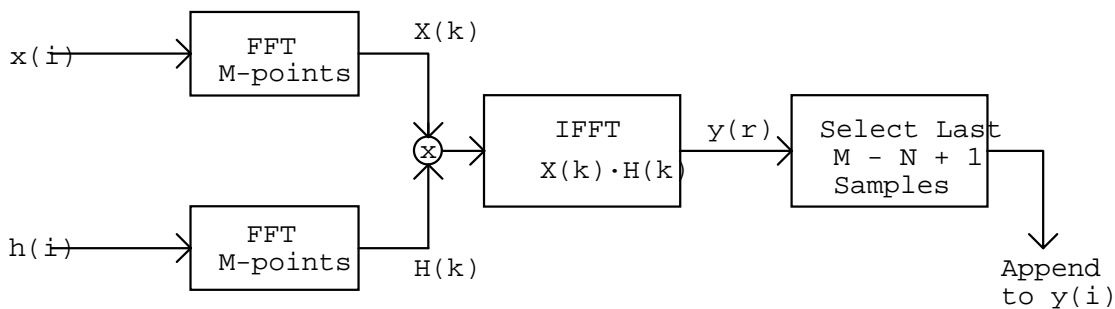
Gli autori hanno dunque pensato di impiegare la tecnica dell'auralizzazione per sintetizzare i segnali di prova per i test soggettivi di qualità acustica. Nella definizione corrente di "auralizzazione" si indica la possibilità di conferire ad un segnale sonoro inizialmente

anechoico tutti gli effetti temporali, spettrali e spaziali causati da un ambiente di ascolto reale o virtuale. Pertanto, con tale termine, spesso si indica il processo combinato di riverberazione artificiale e ricostruzione della sensazione spaziale prodotte da una sala sul segnale sonoro [1]. Con la definizione, invece, “*equalizzazione digitale*” degli spazi si indica la possibilità di creare un filtro digitale in grado di correggere la risposta di un canale di trasmissione, includendo in esso sia i trasduttori impiegati per la diffusione sonora, sia la stessa risposta dell’ambiente. Pertanto si tratta, in sostanza, del processo inverso dell’auralizzazione.

Il presente lavoro costituisce uno studio preliminare, ed è teso a verificare che la tecnica di auralizzazione consenta di ottenere segnali binaurali sintetici sufficientemente simili alle registrazioni binaurali effettuate dal vivo, e contemporaneamente ad analizzare la possibilità di operare una equalizzazione digitale dell’ambiente.

Da diversi anni sono disponibili sistemi hardware basati su DSP in grado di effettuare convoluzioni in tempo reale, ma essi sono a tutt’oggi molto costosi, e non sempre consentono facilitazioni quali la trasportabilità d’uso [2]. Parallelamente, i sistemi basati su software consentono l’utilizzo di un qualsiasi PC, ma non sempre riescono ad effettuare il calcolo in tempi ragionevoli. Il sistema usato, invece, consente l’utilizzo dell’hardware di un qualsiasi PC e contemporaneamente un tempo di calcolo molto vicino al tempo reale.

Nel caso analizzato, tale processo è stato effettuato tramite convoluzione tra le risposte all’impulso sperimentali ed il segnale di ingresso, mediante l’algoritmo “*select save*”, avente funzionamento riassumibile nel diagramma che segue [3]:



I brani così convoluti sono stati successivamente riproposti nelle camere di prova, come descritto nel paragrafo seguente.

La riproduzione di sequenze musicali all’interno di ambienti sonori, comunque, comporta in ogni caso una modificazione o distorsione del segnale originario, causato da una risposta in frequenza non piatta del sistema costituito da trasduttori acustici ed ambiente di ascolto. Per ovviare a tale modificazione è necessario, pertanto, determinare un filtro digitale inverso dell’intera catena, che consenta una ottimale condizione di ascolto del segnale. La possibilità di creazione di filtri inversi *finiti* e *causali* è stata recentemente studiata da diversi autori [4,5], i quali hanno fornito alcune formulazioni matematiche per tale problema.

Secondo alcuni autori [5], la risposta all’impulso originaria $h(t)$ può essere separata nelle due componenti “*a fase minima*” e “*a fase massima*”, nel modo seguente:

$$h(t) = h_{\min}(t) * h_{\max}(t) \quad (1)$$

Tale decomposizione viene resa possibile partendo dalla z-trasformata, supposta razionale, $H(z)$ del segnale discreto nel tempo e a fase miscelata, che può essere espressa, pertanto:

$$H(z) = \frac{Az^{-r} \cdot \prod_{k=1}^{m_i} (1 - a_k \cdot z^{-1}) \cdot \prod_{k=1}^{m_o} (1 - b_k \cdot z)}{\prod_{k=1}^{p_i} (1 - c_k \cdot z^{-1}) \cdot \prod_{k=2}^{p_o} (1 - d_k \cdot z)} \quad (2)$$

in cui $|a_k|, |b_k|, |c_k|, |d_k| < 1$; A rappresenta un fattore di scala, ed è reale e costante, mentre z^{-r} rappresenta un termine di ritardo di fase lineare. a_k e b_k rappresentano gli zeri di $H(z)$, mentre c_k e d_k sono i poli di $H(z)$. Poiché le sequenze rappresentanti le risposte all'impulso sperimentali $h(t)$ hanno durata finita, la funzione $H(z)$ può essere considerata consistente soltanto dei termini che contribuiscono gli zeri, ossia:

$$H(z) = Az^{-r} \cdot \prod_{k=1}^{m_i} (1 - a_k \cdot z^{-1}) \cdot \prod_{k=1}^{m_o} (1 - b_k \cdot z) \quad (3)$$

dalla quale si può quindi ottenere:

$$\ln[H(z)] = \hat{H}(z) = \ln[A] + \ln[z^{-r}] + \sum_{k=1}^{m_i} \ln[1 - a_k \cdot z^{-1}] + \sum_{k=1}^{m_o} \ln[1 - b_k \cdot z] \quad (4)$$

Ricordando che la costante A rappresenta un fattore di scala reale, che può essere supposto positivo, tralasciando inoltre il termine di fase lineare (semplice ritardo di propagazione), ed infine espandendo i restanti termini logaritmici in serie di potenze, valendo il teorema della convoluzione, e grazie alle proprietà del *complex cepstrum*, si ottiene:

$$\hat{h}(n) = \hat{h}_{\min}(n) + \hat{h}_{\max}(n) \quad (5)$$

dove:

$$\begin{aligned} \hat{h}(n) &= \ln|A| \text{ per } n = 0 \\ \hat{h}(n) &= -\sum_{k=1}^{m_i} \frac{a_k^n}{n} = \hat{h}_{\min}(n) \text{ per } n > 0 \\ \hat{h}(n) &= \sum_{k=1}^{m_o} \frac{b_k^n}{n} = \hat{h}_{\max}(n) \text{ per } n < 0 \end{aligned} \quad (6)$$

Oppenheim dimostrò però che per essere correttamente calcolate le componenti a fase minima e a fase massima della risposta all'impulso di input, quest'ultima deve essere preventivamente "pesata" mediante una curva esponenziale di ponderazione, ossia

$$h^{\text{exp}}(n) = h(n) \cdot e^{-\lambda n} \quad (7)$$

La risposta all'impulso originaria $h(t)$ può anche venire scomposta [5] in due parti, definite “*equivalent minimum phase*” e “*all-pass component*”, ossia:

$$h(t) = h_{eq}(t) * h_{ap}(t) \quad \leftrightarrow \quad H(z) = H_{mix}(z) = H_{eq}(z) \cdot H_{ap}(z) \quad (8)$$

In tale decomposizione la componente a fase minima risulta avere la peculiare caratteristica di possedere lo stesso spettro d'ampiezza di $h(t)$. Dal punto di vista della percezione sonora, la *all-pass component* è risultata esprimere l'impressione soggettiva della riverberazione del segnale, mentre la *equivalent minimum phase* è risultata essere correlata agli aspetti timbrici del sistema.

In questo caso, la funzione a fase mixata $H(z)$, ossia $H_{mix}(z)$, può venire scomposta in altre due funzioni, poiché si può scrivere:

$$H(z) = H_{mix}(z) = \prod_{k=1}^{m_i} (1 - a_k \cdot z^{-1}) \cdot \prod_{k=m_i+1}^{m_i+m_o} (z^{-1} - a_k) \cdot \frac{\prod_{k=m_i+1}^{m_i+m_o} (1 - a_k \cdot z^{-1})}{\prod_{k=m_i+1}^{m_i+m_o} (z^{-1} - a_k)} \quad (9)$$

ed in questo caso la funzione a fase mixata $H_{mix}(z)$ può essere vista come la (8), laddove si sono definite le espressioni:

$$H_{eq}(z) = \prod_{k=1}^{m_i} (1 - a_k \cdot z^{-1}) \cdot \prod_{k=m_i+1}^{m_i+m_o} (z^{-1} - a_k) \quad H_{ap}(z) = \frac{\prod_{k=m_i+1}^{m_i+m_o} (1 - a_k \cdot z^{-1})}{\prod_{k=m_i+1}^{m_i+m_o} (z^{-1} - a_k)} \quad (10)$$

Le formule suddette possono essere anche scritte utilizzando un formalismo leggermente differente, evidenziando la relazione intercorrente tra le due diverse scomposizioni, ossia:

$$H_{eq}(z) = H_{min}(z) \cdot \prod_{k=1}^{m_o} (1 - b_k \cdot z^{-1}) \quad \text{e} \quad H_{ap}(z) = \frac{\prod_{k=1}^{m_o} (1 - b_k \cdot z)}{\prod_{k=1}^{m_o} (1 - b_k \cdot z^{-1})} \quad (11)$$

Tale decomposizione rende agevole il calcolo del complex cepstrum di queste sequenze separandone le componenti a fase minima equivalente a passa tutto: il prodotto di convoluzione viene separato in somma semplice, ed in questo caso, sarà:

$$\hat{h}(n) = \hat{h}_{min}(n) + \hat{h}_{max}(n) = \hat{h}_{eq}(n) + \hat{h}_{ap}(n) \quad (12)$$

Purtroppo, sebbene possa essere quindi possibile, mediante il complex cepstrum, separare le componenti di fase equivalente minima da quella passa-tutto, non appare altrettanto agevole effettuare le reciproche inversioni. Inoltre, ad ulteriore svantaggio della tecnica

illustrata, la componente “passa-tutto” risulta essere particolarmente rilevante per i casi di acustica architettonica, e quindi nelle sale in particolare, e in sostanza ineliminabile in ambienti caratterizzati da alta componente riverberante, come ad esempio le chiese.

Parallelamente alle tecniche evidenziate, vi è la possibilità di determinare in maniera *approssimata* l’inversione delle risposte all’impulso. La base di tale metodologia consiste nella determinazione della risposta all’impulso inversa mediante un tipico problema ai minimi quadrati:

$$I = \sum_{n=-\infty}^{+\infty} e^2(n) = \sum_{n=-\infty}^{+\infty} [\delta(n) - \hat{\delta}(n)]^2 = \sum_{n=-\infty}^{+\infty} [\delta(n) - h_{inv}(n) * h(n)]^2 \quad (13)$$

E’ stato dimostrato [6] che il procedimento appena discusso [“zero delay”] produce una versione approssimata della inversione a “fase minima equivalente” della risposta all’impulso, quindi senza analizzare le proprietà di fase del segnale originario. Per questa ragione risulta preferibile la inversione “optimum delay”, che contempla anche l’inversione della componente *passa tutto*, sebbene fornisca come risultato del processo di deconvoluzione il segnale originario con un ritardo temporale (e pertanto non in tempo reale). A tale scopo si rendono necessarie le operazioni di “zero padding”, la cui lunghezza ottimale è di N-1 punti, e di *time reversal* da operarsi prima e dopo l’inversione, in grado di rimuovere la acausalità del filtro inverso ottenuto, che risulta così stabile, acausale e finito.

La minimizzazione del valore discretizzato di I, dato da

$$I = \sum_{K=-N}^N [\delta_k - \hat{\delta}_k]^2 \quad (14)$$

produce una matrice particolare, semi-definita simmetrica, detta matrice di *Toeplitz*, che consente una soluzione numerica semplificata del problema, mediante un algoritmo suggerito da Wiener-Levinson.

MATERIALI E METODI

La procedura analizzata è stata applicata nelle camere acustiche in dotazione dell’ISVR di Southampton, UK. Per le registrazioni sono state utilizzate le seguenti apparecchiature (gentilmente fornite dall’ISVR Consultant Service):

- PC dotato di scheda audio SB AWE 32
- Amplificatore YAMAHA P 2160
- 4 altoparlanti omnidirezionali CANON S-50 disposti nei 4 vertici del perimetro delimitato della sala
- Testa artificiale (KEMAR).
- Coppia di microfoni, adattatori ad angolo e amplificatori microfonici B&K 5134.
- DAT SONY TCD D10;

L’acquisizione delle risposte all’impulso è stata effettuata mediante scheda MLSSA, utilizzando la stessa strumentazione per la diffusione del segnale pseudorandom all’interno degli ambienti di prova.

Le misure sono state ripetute nelle seguenti camere acustiche:

- Camera anecoica grande;

- Camera anecoica piccola;
- Camera riverberante;
- IEC room (in due differenti configurazioni).

Nel caso della IEC Chamber, le misure e le registrazioni sono state ripetute sia nella situazione di camera sprovvista di alcun arredo, sia nella situazione di disposizione di pannelli diffondenti nell'intera superficie delle due pareti laterali. Sono state, inoltre, utilizzati anche una coppia di altoparlanti molto direttivi, (SONY QUAD PRO 63) per testare l'eventuale influenza della direttività degli altoparlanti nella fase di riascolto.

E' stata successivamente invertita, secondo la procedura indicata, la risposta all'impulso della camera anecoica minore (sala utilizzata per le prove di ascolto), e sia i brani registrati, sia i brani sintetizzati mediante auralizzazione sono stati convoluti con il filtro inverso così ottenuto.

E' stata anche ottenuta una seconda serie di brani musicali, nella quale il brano anecoico è stato convoluto con risposte all'impulso di vari noti teatri italiani, ed ovviamente sempre equalizzato digitalmente mediante convoluzione con l'inverso della risposta all'impulso del sistema di riproduzione. I risultati provenienti da questa seconda serie di tests sono in corso di elaborazione e saranno pubblicati in un secondo momento.

Sono state prodotte due differenti cassette DAT, nelle quali le sequenze dei brani erano riproposte in ordine diverso. La successione così ottenuta è stata pertanto riproposta all'interno della camera anecoica minore ad un gruppo di ricercatori dell'ISVR, utilizzati quali soggetti particolarmente qualificati, dopo essere stati sottoposti ad audiogramma, ai quali era stato chiesto di confrontare le varie coppie di brani, ponendo particolare attenzione alle eventuali differenze che emergessero tra di esse.

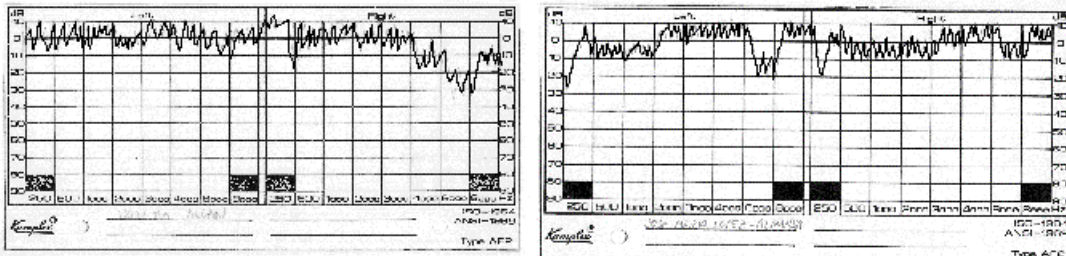


Figura 1 audiogrammi raccolti dagli intervistati

RISULTATI

Fra le coppie di brani presentate erano anche state inserite coppie "di controllo", costituite da un lato dalla mera ripetizione dello stesso brano (coppie veramente uguali), e dall'altro da coppie in cui venivano volutamente presentati due brani sicuramente diversi, perchè provenienti da registrazioni e/o simulazioni eseguite in sale diverse.

Come si può notare, la percentuale di riscontro tra le coppie veramente uguali e le coppie provenienti dalle operazioni di convoluzione, hanno fornito un risultato sicuramente soddisfacente. La minor percentuale di somiglianza tra le coppie registrazioni-convoluzioni, inoltre, va quasi esclusivamente posta in relazione alle misure effettuate nella camera riverberante, per la quale la misurazione della risposta all'impulso di soli 32768 punti, dovuta alla limitazione della scheda di acquisizione in possesso dell'ISVR,

ha comportato un inevitabile troncamento della coda sonora, in fase di convoluzione, e pertanto ha abbassato il livello di similarità tra i due brani.

Coppie registrazioni-convoluzioni	Coppie veramente uguali	Coppie veramente diverse
74.6 %	84.5 %	0 %

Tab 1- Risultato complessivo dei questionari alla prima domanda

Questo risultato, comunque, è in perfetta armonia con quanto verificato recentemente in studi analoghi, relativi alla efficacia della operazione di convoluzione di risposte all'impulso simulate energetiche, [7,8] oltre ad uno studio compiuto sulle possibilità di effettuare copie "virtuali" di strumenti musicali [9].

QUESTIONNAIRE - summer 1995

CONVOLUTED - RECORDED MUSIC in anechoic chamber - SOUTHAMPTON

name: *Delma Allen*

position: *Research Fellow, Rm 2035*

Pair n. *1*

Are the piece 'A' and 'B' the same? Yes No

If the pieces are not the same, please compile the following questionnaire, referred to the different pairs of music ('A' is the first music):

	very lot (-3)	a lot (-2)	slightly (-1)	equal (0)	slightly (+1)	a lot (+2)	very lot (+3)	
A is pleasant	0	0	0	0	0	0	0	B is pleasant
A is unclear	0	0	0	0	0	0	0	B is unclear
A is soft - Harp	0	0	0	0	0	0	0	B is soft - Harp
A is vibrant	0	0	0	0	0	0	0	B is vibrant
A is rounded	0	0	0	0	0	0	0	B is rounded
A is vigorous	0	0	0	0	0	0	0	B is vigorous
A is diffuse	0	0	0	0	0	0	0	B is diffuse
A is overbearing	0	0	0	0	0	0	0	B is overbearing
A is dry	0	0	0	0	0	0	0	B is dry
A is light	0	0	0	0	0	0	0	B is light
A is weak	0	0	0	0	0	0	0	B is weak
A has more pronounced bass	0	0	0	0	0	0	0	B has more pronounced bass
A has more pronounced treble	0	0	0	0	0	0	0	B has more pronounced treble
A is soft	0	0	0	0	0	0	0	B is soft

DAT CASSETTE N.1: EXAMPLES FOR SUBJECTIVE TESTS

After the recorded data already explained, the following sequence (random distribution) of piece has been utilised for validation; the recorded pieces with bass and treble at the same level has been utilised

1. Anechoic piece		00.30
2. Rev. - An. small	(recorded)	Truly diff. 01.23
3. An. large - An. large	(conv. - rec.)	Equal to 14 02.38
4. Iec1 - Iec1	(recorded)	Truly eq. 03.46
5. Iec2 - Iec2	(rec. - conv.)	Equal to 11 04.56
6. Iec1 - Iec1	(conv. - rec.)	Equal to 8 06.05
7. An. large An. large	(recorded)	Truly eq. 07.11
8. Rev. - Iec2	(recorded)	Truly diff. 08.13
9. Iec1 - Iec1	(rec. - conv.)	Equal to 5
10. Rev. - Iec1	(recorded)	Truly diff. 10.18
11. An. small - An. small	(rec. - conv.)	Equal to 16 11.24
12. Iec2 - Iec2	(conv. - rec.)	Equal to 4 12.26
13. An. small - An. small	(recorded)	Truly eq. 13.30
14. Iec2 - Iec2	(recorded)	Truly eq. 14.40
15. An. large - An. large	(rec. - conv.)	Equal to 2 15.53
16. Rev. - An. large	(recorded)	Truly diff. 16.56
17. An. small - An. small	(conv. - rec.)	Equal to 10 18.04

In a second step, the following sequences of pieces, coming from convolution between the same anechoic music and IR of many Italian Concert Halls, has been utilised:

18. Church "S. Lucia",	Bologna
19. Theatre "alla Scala",	Milan
20. Hall "Europa",	Bologna
21. Hall "Poggio Imperiale",	Florence
22. Theatre "Pergola",	Florence
23. Theatre "Comunale",	Ferrara
24. Theatre "Bellini",	Catania
25. Theatre Comunale,	Gradisca
26. Church "S. Domenico",	Foligno
27. Theatre "Comunale",	Florence
28. Little Theatre "Comunale",	Florence
29. Theatre "Comunale",	Bologna
30. Theatre "Verdi",	Florence

figura 2 esempio di questionario e insieme delle coppie presentate agli intervistati

CONCLUSIONI

Al termine delle elaborazioni dei dati ottenuti presso presso l'ISVR di Southampton è risultata una particolare attendibilità delle convoluzioni di risposte all'impulso con brani di musica anecoica, anche qualora si voglia simulare il comportamento di ambienti piuttosto riverberanti. In tale caso, rimangono ancora alcune difficoltà per le acquisizioni di risposte all'impulso sufficientemente lunghe, poiché viene richiesto un numero minimo di campioni sufficienti almeno a coprire la coda della risposta all'impulso per la durata approssimativa del tempo di riverberazione. Queste difficoltà possono però venire superate impiegando un nuovo tipo di sistema di misura MLS, anch'esso interamente software, e pertanto in grado di deconvolvere risposte all'impulso di lunghezza arbitraria, nonchè rimuovere le componenti cross-talk che si generano nella fase di ascolto tramite altoparlanti: tale sistema è già stato sviluppato, ma non era ancora disponibile al momento della sperimentazione eseguita presso l'ISVR.

La procedura qui illustrata consente di effettuare ulteriori indagini psicoacustiche, evitando alcuni problemi particolarmente notevoli, quali la ripetibilità delle risposte fornite, l'influenza nell'espressione del giudizio di componenti non acustiche, la difficoltà di correlazione tra espressioni del giudizio e medie talvolta arbitrarie di alcuni parametri acustici. In questo caso, invece, è possibile correlare i valori forniti dai soggetti direttamente con i parametri acustici della particolare risposta all'impulso utilizzata per la convoluzione, senza dover effettuare medie assai discutibili su valori molto lontani tra loro, per ragioni acustiche, come nel caso dell'ITDG. Inoltre, l'operazione di deconvoluzione permette di operare il prefiltraggio delle risposte all'impulso da utilizzare per le convoluzioni, in funzione dell'ambiente e delle caratteristiche elettroacustiche (risposta in frequenza) del sistema di riproduzione, ottenendo pertanto una equalizzazione digitale del sistema di riproduzione.

RINGRAZIAMENTI

Si ringrazia il Dr. Ian Flindell per la preziosa collaborazione fornita durante la campagna di misure presso l'ISVR.

BIBLIOGRAFIA

- [1] KLEINER M, DALENBACK B., SVENSSON P., - "Auralization - an overview" – in *Journal of Audio Engineering Society* - vol. 41, n° 11, 1993
- [2] CONNOLLY B. - "A User Guide for the Lake FDP-1 plus" - Lake DSP Pty. Ltd, Maroubra (Australia) - 1992
- [3] OPPHENEIM, A.V., SCHAFER. R., "Digital Signal Processing" - Prentice Hall, Englewood Cliffs, New York, 1975
- [4] MOURJOUPOLOUS, J., "Digital equalization of room acoustics", in *Journal of Audio Engineering Society* n. 42, 1994;
- [5] NEELY S.T., ALLEN J.B., "Invertibility or a room impulse response" – *Journal of Acoustic Society of America*, **66**, 1979
- [6] MOURJOUPOLOUS, J., "On the variation and invertibility of room impulse response functions", *Journal of sound and vibration*, **102**, 1985
- [7] FARINA A., "An example of adding spatial impression to recorded music: signal convolution with binaural impulse responses" - Proceedings of *Acoustics and Recovery of Spaces for Music* - Ferrara, 27-28 October 1993
- [8] FARINA A., "Auralization software for the evaluation of the results obtained by a pyramid tracing code: results of subjective listening tests" - Proceedings of *ICA95*, Trondheim (Norway), 1995.
- [9] FARINA A., LANGHOFF A., TRONCHIN L., "Subjective Comparisons of "Virtual" Violins Obtained by Convolution" - Proceedings of *CIARM95*, Ferrara, 1995