UNIVERSITA' DEGLI STUDI DI PARMA

FACOLTA' DI INGEGNERIA

CORSO DI DOTTORATO DI RICERCA IN INGENERIA INDUSTRIALE

XX CICLO

Settore Scientifico Disciplinare ING-IND/10

STUDI DI ELETTROACUSTICA MUSICALE ED ACUSTICA ARCHITETTONICA INERENTI A TEATRI, SALE DA CONCERTO E SALE D'ASCOLTO

Tutore:

Prof. Ing. ANGELO FARINA

Coordinatore:

Prof. Ing. MARCO SPIGA

Tesi di Dottorato di:

Dott. Ing. ANDREA CAPRA

INDICE

INDICE DELLE FIGURE

1 La teoria

1.1 Psicoacustica

Per parlare di *spazializzazione* del suono occorre capire quali meccanismi percettivi mette in atto il nostro sistema uditivo per localizzare le sorgenti sonore che ci circondano. Conoscere come la nostra mente reagisce a stimoli sonori, capire come vengono discriminate le loro direzioni di provenienza, fornisce uno strumento per l'ottimizzazione dei sistemi di riproduzione tridimensionali che come obbiettivo si prefiggono la ricostruzione fedele di un campo sonoro.

1.1.1 Lateralizzazione

Un'osservazione ovvia, ma con effetti non scontati sulla percezione sonora, è che il nostro sistema uditivo capta suoni attraverso due ricettori ben distinti e separati fra loro, le orecchie. Esse presentano una distanza media di circa 18 cm ed un suono proveniente da una sorgente posta lateralmente come in Figura 1.1 arriverà non simultaneamente alle due orecchie.

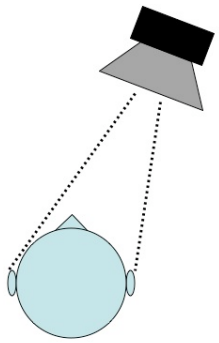


Figura 1.1 – Percezione binaurale di un suono

La differenza di cammino, in termini di tempo, del suono alle due orecchie è chiamata *Interaural Time Difference* (I.T.D) ed è uno dei metodi con cui valutiamo la posizione delle sorgenti sonore. La differenza di livello sonoro, *Interaural Level Difference* (I.L.D), tra le due orecchie non dipende da una differenza di cammino,

quanto piuttosto dall'ombra acustica che la testa produce (Figura 1.2) sull'orecchio meno esposto o vicino alla sorgente.

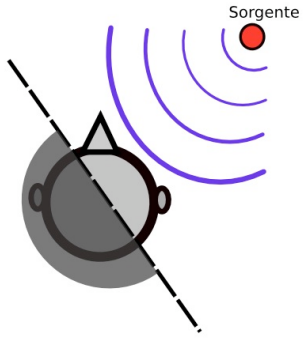


Figura 1.2 – Percezione binaurale di un suono

E' facile quindi intuire come l'I.T.D. non dipenda dalla frequenza, sebbene la fase non sia costante, mentre l'I.L.D. dipende fortemente dalla frequenza del suono emesso dalla sorgente a causa della diffrazione che subisce nel contatto con la testa. Semplificando si può dire che se la lunghezza d'onda risulta essere maggiore del diametro della testa essa verrà diffratta mentre, se risulta inferiore, verrà attenuta con un effetto risultante di filtraggio passa-basso.

Dai tanti studi effettuati si è visto che tali meccanismi, da soli, non bastano per determinare in modo corretto la posizione di una sorgente sonora, questo per la presenza dei "coni di confusione" (Begault 2000). Ogni suono che arriva da un cono di direzioni avrà lo stesso livello, la stessa fase e le stesse differenze di tempo degli altri e la posizione della sorgente diverrebbe, per questo, ambigua. Per risolvere questo ambiguità il nostro sistema uditivo utilizza altri due meccanismi percettivi: i movimenti della testa ed un filtraggio che dipende dall'angolazione.

Prendiamo quindi la sorgente in Figura 1.2. Se la testa ruota verso sinistra il livello all'orecchio destro aumenta e quello a sinistra diminuisce; se la testa, vice versa, ruota verso destra il livello all'orecchio sinistro aumenta gradualmente e quello a destra diminuisce. Questi movimenti si sono rivelati importantissimi per evitare di confondere un suono frontale con uno posteriore.

Il filtraggio precedentemente citato avviene ad opera della pinna, che assegna una fase e uno spettro ben precisi all'onda sonora che arriva alla testa in relazione

all'angolo di provenienza. Questo permette la localizzazione di un suono anche utilizzando un solo orecchio.

1.1.2 Head Related Transfer Function

Onde sonore che arrivano da differenti direzioni nella spazio subiscono uno scattering a contatto con l'orecchio, con l a testa, con i pantaloni e il busto dell'ascoltatore. Questo scattering produce un filtraggio acustico dei segnali che arrivano all'orecchio destro e a quello sinistro, filtro che può essere descritto da una funzione di trasferimento complessa chiamata *Head Related Transfer Function* (H.R.T.F.).

La HTRF descrive tutti i cambiamenti che occorrono alle nostre orecchie rispetto alla forma d'onda, alla fase e all'ampiezza, mentre la sorgente sonora si muove rispetto all'ascoltatore. Spesso tale funzione viene misurata con microfoni posti nei timpani di una testa costruita con materiale fonoassorbente per tenere conto del filtraggio operato dai padiglioni auricolari e dal canale uditivo.

Vediamo di capire meglio cosa sia la HTRF. Per trovare la pressione sonora che una sorgente arbitraria *x(t)* produce al timpano dell'orecchio, tutto quello di cui si ha bisogno è la risposta all'impulso h(t) dalla sorgente al timpano. Questa è chiamata *Head Related Impulse Response (HRIR)* e la sua trasformata di Fourier *H(f)* è chiamata *Head Related Transfer Function (HRTF)*.

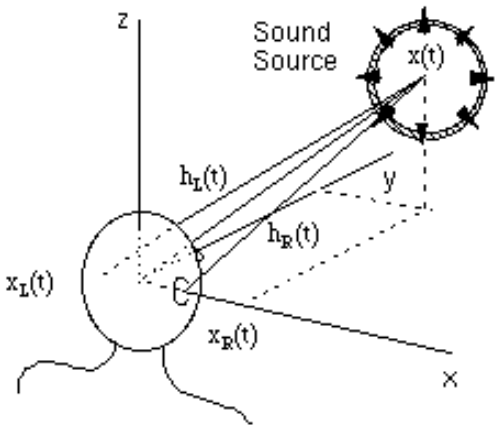


Figura 1.3 – Sorgente e risposte all'impulso in relazione alla testa

Le seguenti espressioni rappresentano la pressione sonora che arriva all'orecchio destro ($x_R(t)$) e all'orecchio sinistro ($x_L(t)$).

$$x_L(t) = \int h_L(\tau) x(t-\tau) d\tau \qquad (1.1)$$

$$x_R(t) = \int h_R(\tau) x(t-\tau) d\tau \qquad (1.2)$$

La funzione HTRF è una complicata funzione a quattro variabili: tre coordinate spaziali e la frequenza. In coordinate sferiche, per distanze maggiori di un metro, la sorgente è detta "in campo lontano" e l'HTRF decade con l'inverso della distanza.

1.1.3 Localizzazione del suono in una stanza

La maggior parte della nostra vita si svolge non in ambienti anecoici ma in presenza di mura, soffitti, pavimenti e oggetti abbastanza grandi da riflettere o modificare la propagazione dell'onda sonora. E' impensabile credere che queste riflessioni non comportino modificazioni nella percezione e nella localizzazione del suono. Si prenda, ad esempio, l'I.T.D.: esso dipende dalla coerenza dei segnali nelle due orecchie. Un suono riverberato contiene informazioni incoerenti e in una stanza grande, dove il suono riverberato prevale su quello diretto, l'I.T.D. diventa inaffidabile.

Non si può dire la stessa cosa per l'I.L.D. perchè, come mostrato da numerosi esperimenti, con esso non importa se il suono sia binaurlamente coerente oppure no. Ovviamente l'I.L.D. è negativamente influenzato dalla presenza di onde stazionarie ma presenta un vantaggio da non sottovalutare, l'assorbimento delle pareti. La maggioranza dei materiali riflettenti, anche in minima parte, presenta assorbimento che, per semplicità, qui assumiamo aumentare con l'aumento della frequenza. Nella maggior parte dei casi, quindi, il suono riflesso risulta meno potente del suono diretto.

Il cervello umano mette in atto anche un'altra strategia per localizzare, in una stanza, la sorgente sonora, una strategia basata sulle onde che per prime giungono al sistema uditivo. Questa strategia è conosciuta come *effetto precedenza* perchè la prima onda

sonora che arriva, il suono diretto con informazioni accurate per la localizzazione, ha la precedenza su quelle che arrivano in ritardo, le riflessioni, contenenti informazioni che potrebbero trarre in inganno. Senza tale effetto, la localizzazione in un ambiente chiuso sarebbe virtualmente impossibile: non ci sarebbe nessuna informazione di I.T.D. utilizzabile e, a causa delle onde stazionarie, il livello del suono risulterebbe non correlato alla vicinanza della sorgente.

1.2 Il suono "surround"

Con l'avvento di nuove e avanzate tecnologie tutto il mondo dell'intrattenimento (cinema, musica, videogiochi, arte) spinge verso la totale immersione nell'evento, sia esso visivo o sonoro. Per questo, nel campo della riproduzione audio, si stanno sviluppando sistemi in grado di conferire all'ascoltatore la sensazione di essere presente alla scena sonora. Come vedremo in questo capitolo diverse sono le tecniche verso cui la ricerca investe energie, ciascuna con le proprie approssimazioni e con i propri aspetti positivi: presso il Dipartimento di Ingegneria Industriale, da anni l'attenzione si è concentrata su due particolari tecniche surround (Ambisonics e Stereo Dipolo) che puntano alla ricostruzione fedele del campo sonoro partendo però da approcci differenti, come verrà mostrato in questo capitolo.

1.2.1 La storia

La ricerca della riproduzione fedele e spaziale di un evento sonoro ebbe origine in Francia alla fine del diciannovesimo secolo durante un concerto di musica classica, quando, usando alcuni telefoni posti in linea frontalmente al palco, avvenne la trasmissione dell'esecuzione musicale su altrettanti ricevitori posti in una stanza adiacente. La qualità del suono, ovviamente era molto bassa ma per la prima volta si riuscì a conferire spazialità all'evento sonoro riprodotto.

Alla fine degli anni '20 e nei primi anni '30 venne messa a punto e formalizzata una tecnica per rendere, in fase di riproduzione, una certa spazialità al suono. La tecnica che sviluppata da Alan Blumlein (U.K.) e dalla RCA (U.S.A) era stata pensata per sistemi che usano solo un ristretto numero di canali d'informazione per la riproduzione attraverso un paio di loudspeakers.

La tecnica sviluppata da Blumlein consiste in un paio di microfoni con direttività a Figura di otto, con le capsule poste il più vicino possibile, con il lobo frontale di uno dei due posto a 45° a sinistra rispetto alla normale ed il lobo frontale dell'altro posto a -45° rispetto alla normale (Figura 1.4).

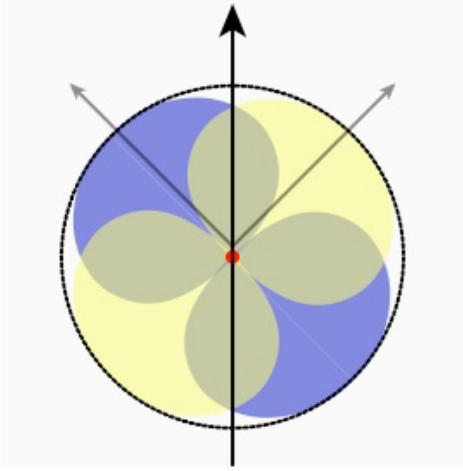


Figura 1.4 – Stereo Blumlein

Nonostante questa tecnica produca eccellenti immagini stereofoniche presenta un macroscopico difetto: per la direttività a "figura di otto" dei microfoni impiegati, il suono catturato dai lobi posteriori viene anch'esso proiettato frontalmente; per questo, molte delle registrazioni stereofoniche, per tanti anni, hanno presentano una eccessiva riverberazione.

Gli amanti di questo tipo di tecnica stereofonica, per ovviare al problema appena descritto, hanno sostituito i "figura di otto" con capsule a "cardioide" cambiando di conseguenza anche l'apertura dei due microfoni. Questa modifica consente quindi di arginare il problema della riverberazione, pagando però in termini di una minore accuratezza dell'immagine sonora. Le tante tecniche stereofoniche che nel corso degli anni sono state sperimentate (MS stereo, AB stereo, ORTF, Decca, Dolby) sono ancora estremamente diffuse in quanto la maggior parte dei sistemi di riproduzione audio sfruttano ancora due soli loudspeaker (coppia stereo) posti, rispetto all'ascoltatore, a ±30°.

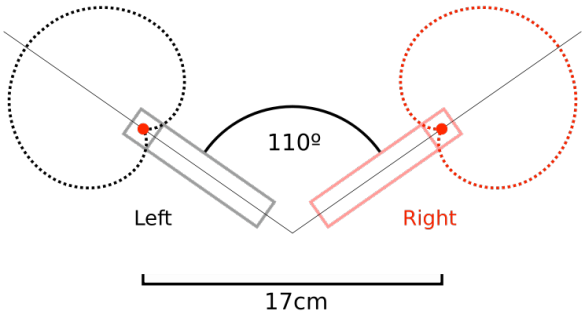


Figura 1.5 – Stereo ORTF

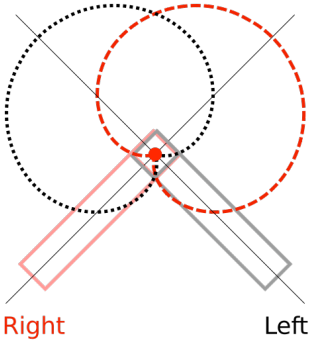


Figura 1.6 – Stereo XY

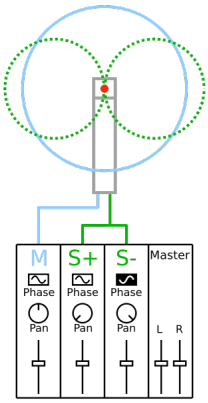


Figura 1.7 – Stereo MS

La prima tecnica di riproduzione audio "surround" sperimentata è quella chiamata "*quadrifonia*". Si tratta di 4 altoparlanti posizionati ai vertici di un quadrato: il suono, in fase di codifica, è semplicemente distribuito attraverso guadagni alle casse. Nonostante gruppi rock di successo, negli anni '70, abbiano provato a produrre

album in quadrifonia, questo tipo di registrazioni non ebbe un grande successo commerciale.

Con la grande diffusione di strumenti tecnologici estremamente evoluti, supporti digitali in grado di immagazzinare dati multicanale, console per videogiochi, schermi sempre più coinvolgenti, in questi anni è cresciuta esponenzialmente la richiesta di un suono sempre più avvolgente, che possa fare da naturale complemento alle immagini di televisioni e schermi dall'altissima risoluzione. Per questo motivo la ricerca si spinge verso sistemi in grado di rendere l'ascoltatore, in modo accurato, partecipe dell'evento musicale che sta ascoltando, immerso nell'ambiente sonoro del videogioco che sta utilizzando, trasportato all'interno del film che sta guardando. Per questo lo standard ITU 5.1 (5 altoparlanti e un sub-woofer disposti come in Figura 1.8) ebbe velocemente una grande diffusione, molto maggiore rispetto alla sua predecessora "quadrifonia".

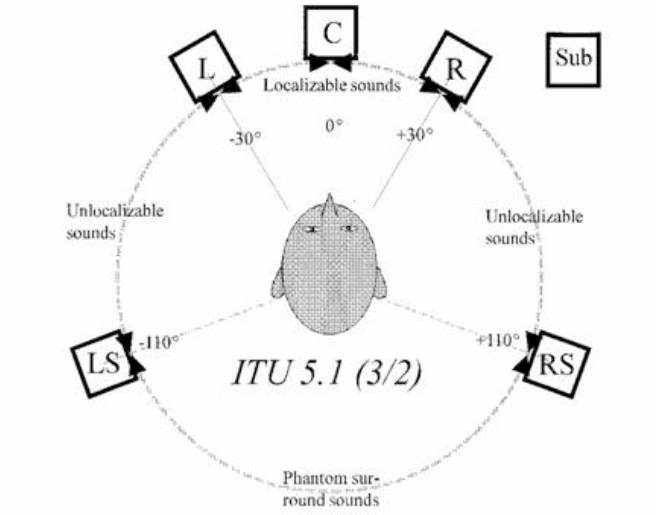
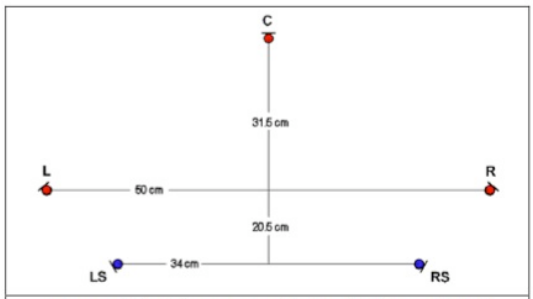


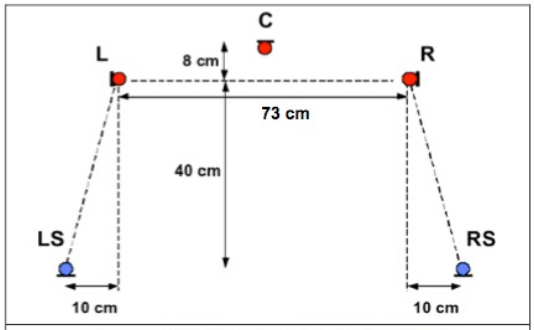
Figura 1.8 – Disposizione dei diffusori acustici secondo lo standard ITU 5.1

Anche le tecniche di ripresa sonora si sono adeguate a questo nuovo modo di riproduzione: tante sono le tecniche sviluppate ed ogni *sound engineer* le utilizza in base al suo gusto personale. Nelle figure sottostanti vengono mostrate tre delle tecniche microfoniche surround più diffuse.



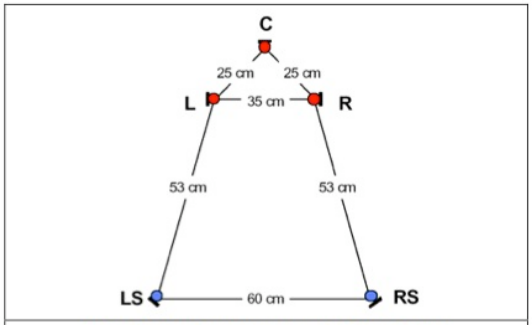
Schema del sistema microfonico Williams MMA
C : Cardioide, 0°
L, R : Cardioide, ± 40°
LS, RS : Cardioide, ± 120°

Figura 1.9 – Schema microfonico Williams MMA



Schema del sistema microfonico OCT
C : Cardioide, 0°
L, R : Super Cardioide, ± 90°
LS, RS : Cardioide, ± 180°

Figura 1.10 – Schema microfonico OCT



Schema del sistema microfonico INA-5
C : Cardioide, 0°
L, R : Cardioide, ± 90°
LS, RS : Cardioide, ± 150°

Figura 1.11 – Schema microfonico INA-5

I sistemi attualmente in uso per la riproduzione surround sono essenzialmente divisibili in due distinte categorie:

- Sistemi che diefiniscono in modo non del tutto restrittivo un layout degli speaker, ma senza alcun riferimento a come i segnali siano stati registrati e quindi distribuiti sui vari canali. I due sistemi più famosi appartenenti alla categoria sono senz'altro il Dolby Digital - AC3(Dolby Labs) e il DTS (Kramer).
- Sistemi che forniscono indicazioni su come il suono riprodotto sia stato registrato o sintetizzato, avendo come scopo una ricostruzione fedele del campo sonoro. Tra le principali tecniche troviamo l'Ambisonics, la Wave Field Synthesis e lo Stereodipolo.

In questa tesi, visto l'interesse nella ricostruzione realistica di un campo sonoro, l'utilizzo si concentra solo sulla seconda tipologia.

1.2.2 Ambisonics

L'*Ambisonics* è un metodo per registrare informazioni su un determinato campo sonoro per poi riprodurlo attraverso un array di loudspeakers: in questo modo è possibile dare all'ascoltatore l'impressione di trovarsi immerso in un reale campo sonoro tridimensionale. Si parla di "impressione" perchè, come verrà mostrato successivamente, la perfetta ricostruzione può avvenire solo in presenza di un numero elevatissimo di casse in una zona d'ascolto ("sweet spot") molto ristretta.

La teoria si basa sulla decomposizione in armoniche sferiche del campo sonoro. Per far questo occorre descrivere il campo in coordinate sferiche dove un punto r (vettore) è descritto dal raggio r, un azimuth θ e un'elevation δ.

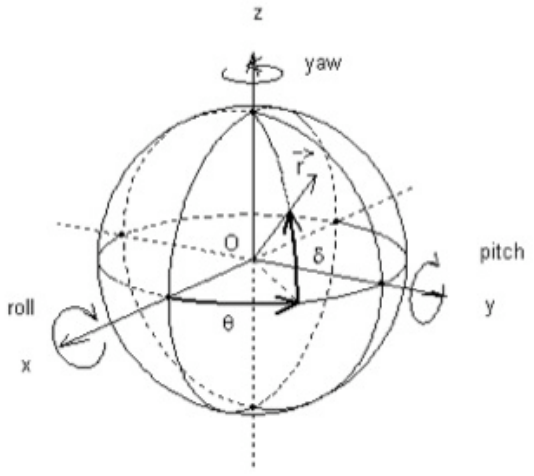


Figura 1.12 – Rappresentazione delle coordinate polari

Il campo di pressione in un determinato punto attorno all'origine degli assi può essere descritto dalla serie di Fourier-Bessel, i cui termini sono prodotti pesati di funzioni direzionali $Y_{mn}^{\sigma}(\theta,\delta)$, dette **armoniche sferiche**, e funzioni radiali:

$$p(\vec{r}) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{\substack{0\le n<m \\ \sigma=\pm 1}} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta) + \sum_{m=0}^{\infty} j^m h^-{}_m(kr) \sum_{\substack{0\le n<m \\ \sigma=\pm 1}} A_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta) \qquad (1.3)$$

Dove:
j^m = numero immaginario elevato alla "m"
$j_m(kr)$ = funzioni sferiche di Bessel (componenti radiali)
$Y_{mn}^{\sigma}(\theta,\delta)$ = armoniche sferiche
$h_m^-(kr)$ = funzioni divergenti sferiche di Henkel = $j_m(kr)$-$jn_m(kr)$
$k = 2\pi f/c$
m = ordine dell'ambisonic variabile
n = indice che varia da –m a +m
B_{mn}^{σ} = segnale Ambisonic

Questa equazione descrive la situazione di una sfera (R1< r < R2) libera da sorgenti. La prima parte dell'espressione descrive il campo tra R1 e R2 dovuto alle sorgenti esterne ("through-going" field) mentre la seconda parte descrive il campo tra R1 e R2 dovuto alle sorgenti interne ("inside sources").

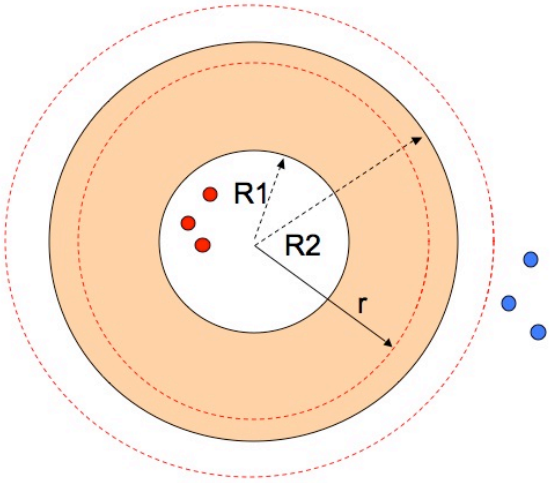


Figura 1.13 – Rappresentazione di sorgenti esterne ed interne all'area di riproduzione audio

Poichè nei sistemi Ambisonic l'approccio è quello di considerare solo sorgenti esterne all'anello di loudspeaker utilizzati per la riproduzione, inizialmente possiamo

trascurare la seconda parte della formula 1.3 e concentrare l'attenzione solo sulla prima parte, ottenendo:

$$p(\vec{r}) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{\substack{0 \le n < m \\ \sigma = \pm 1}} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta) \qquad (1.4)$$

dove B_{mn}^{σ} è l'espressione nel dominio della frequenza di ciò che chiameremo "segnali Ambisonic".

Vediamo quindi come sono definite le armoniche sferiche $Y_{mn}^{\sigma}(\theta,\delta)$:

$$Y_{mn}^{\sigma\,(\mathrm{N3D})}(\theta,\delta) = \sqrt{2m+1}\sqrt{(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}}P_{mn}(\sin\delta) \times \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \quad (\text{ignored if } n = 0) \end{cases} \qquad (1.5)$$

dove:
$P_{mn}(\xi)$ = funzioni di Legendre di grado m e ordine n
$\delta_{p,q}$ = delta di Kronecker = {1 se p=q, 0 altrrimenti}

Occorre quindi definire cosa siano le *funzioni di Legendre*: sono soluzioni dell'equazione differenziale di Legendre; la si trova nella soluzione in coordinate sferiche dell'equazione di Laplace. Le soluzioni, al variare del grado m, formano la successione polinomiale di Legendre:

$$P_n(x) = (2^n n!)^{-1} \frac{d^n}{dx^n}\left[(x^2-1)^n\right] \qquad (1.6)$$

I primi polinomi di Legendre, ottenuti troncando la formula appena descritta, sono i seguenti:

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3) \end{aligned}$$

Utilizzando questi polinomi ed inserendoli nell'espressione delle armoniche sferiche (1.5) è possibile ottenere la rappresentazione di tali armoniche nello spazio.

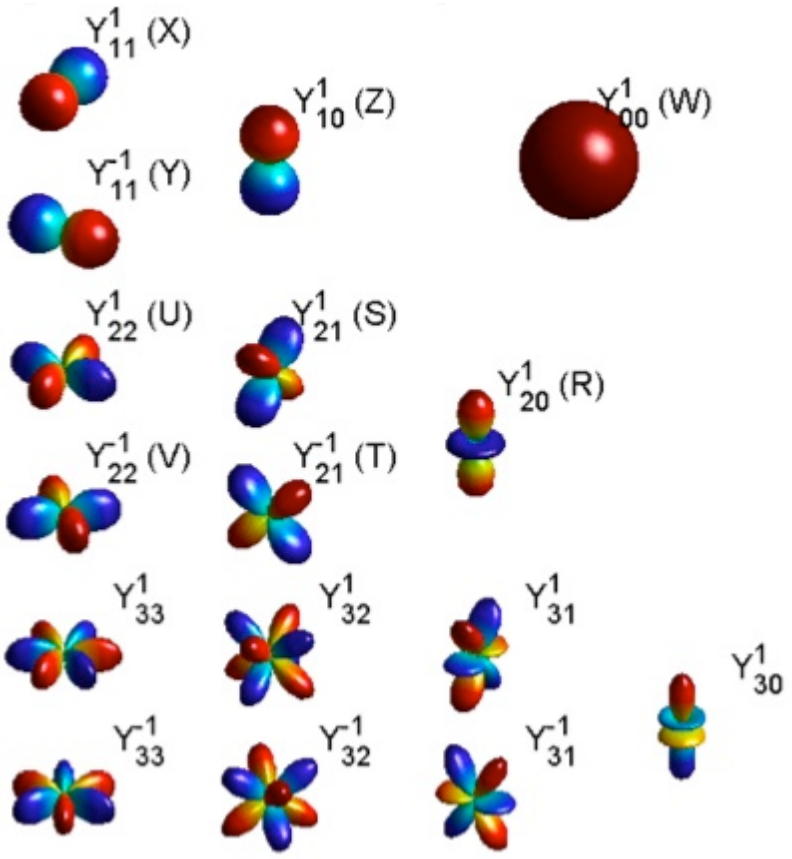


Figura 1.14 – Armoniche sferiche dall'Ordine 0 all'Ordine 3.

Come si può facilmente notare ci sono (2m+1) componenti per ogni grado m (includendo 2 componenti orizzontali che si hanno per m=n, con m>=1).

Per addentrarsi nella teoria Ambisonic occorre capire cosa rappresentino i segnali B_{mn}^{σ}. Essi sono strettamente legati alla pressione e alle sue derivate dei successivi ordini attorno all'origine O. Restringendo l'analisi alle prime 4 componenti possiamo dire che B_{00}^{+1}(= W) rappresenta la pressione nel punto di misura mentre B_{11}^{+1} (= X), B_{11}^{-1} (= Y), B_{10}^{+1} (= Z) sono legati al gradiente della pressione, cioè alla sua derivata prima e quindi rappresentano la velocità delle particelle d'aria lungo i tre assi.

Ogni gruppo di componenti di ordine più elevato in aggiunta dà un'approssimazione del campo sonoro su una zona sempre più ampia attorno all'origine. In pratica, solo un numero finito di componenti (su fino ad un dato ordine M) possono essere trasmessi e sfruttati, e quindi stimati.

Quindi il campo rappresentato è tipicamente approssimato dalla serie troncata:

$$p(\vec{r}) = \sum_{m=0}^{M} j^m j_m(kr) \sum_{\substack{0 \le n < m \\ \sigma = \pm 1}} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \qquad (1.7)$$

dove M rappresenta l'ordine dell'Ambisonic considerato.

Osservando l'andamento delle armoniche sferiche in relazione alle funzioni sferiche di Bessel *j(kr)* [4] si nota come le armoniche con un'alta variabilità angolare vengano associate a funzioni radiali che presentano il primo massimo ad una distanza elevata *kr*. Ciò vuol dire che una grande risoluzione direzionale corrisponde ad una grande espansione radiale. Per avere quindi uno "sweet spot" più ampio devo salire con l'ordine: il massimo lo si ha infatti per valori sempre più ampi. Inoltre, salendo con la frequenza, lo sweet spot si restringe complicandone maggiormente la ricostruzione.

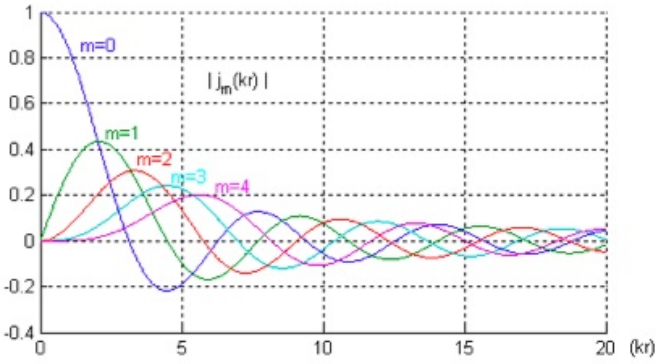


Figura 1.15 – Funzioni di Bessel in funzione di kr *(Daniel,Nicol,Moreau, 2003)*

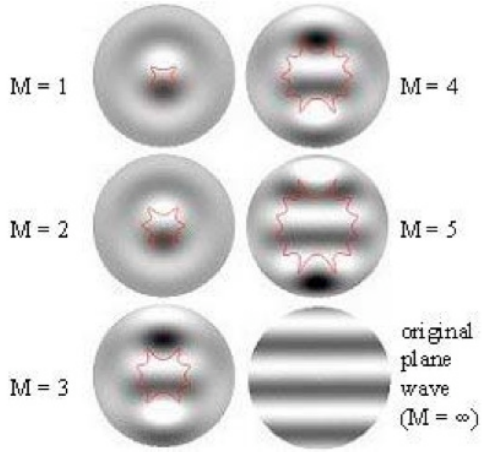


Figura 1.16 – Aumento dell'area di buona ricostruzione del campo in funzione dell'ordine M (Daniel,Nicol,Moreau, 2003)

Si capisce quindi come nella teoria tutto sembri funzionare bene ma come la sua trasposizione nel reale diventi complessa a causa di limitazioni tecnologiche e fisiche, condizioni al contorno che sono praticamente impossibili da raggiungere. Nel corso della lettura si capiranno meglio i limiti che l'applicazione di questa tecnica incontra.

1.2.2.1 Riproduzione Ambisonic

Lo scopo è quello di ricostruire un campo sonoro assemblando acusticamente i componenti ambisonic (campo di pressione e sue derivate) $\tilde{B}_{mn}^{\sigma}$ al centro dell'array di altoparlanti. Supponendo che le casse siano sufficientemente lontane dal punto d'ascolto, i segnali S_i sono codificati come onde piane di coefficienti c_i:

$$c_i = \begin{bmatrix} Y_{00}^{+1}(\theta_i,\delta_i) \\ Y_{11}^{+1}(\theta_i,\delta_i) \\ Y_{11}^{-1}(\theta_i,\delta_i) \\ \dots \\ Y_{mn}^{\sigma}(\theta_i,\delta_i) \end{bmatrix} \quad \tilde{B} = \begin{bmatrix} \tilde{B}_{00}^{+1} \\ \tilde{B}_{11}^{+1} \\ \tilde{B}_{11}^{-1} \\ \dots \\ \tilde{B}_{mn}^{\sigma} \end{bmatrix} \quad S = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \dots \\ S_N \end{bmatrix} \qquad (1.8)$$

Il cosìdetto principio di *re-encoding* può essere scritto nella forma:

$$\tilde{B} = C*S \qquad \text{dove } C = [c_1 \dots c_N] \qquad (1.9)$$

Si ottiene quindi un vettore ci per ogni speaker. L'operazione di decodifica cerca di derivare i segnali S mantenendo gli originali segnali Ambisonic B:

$$S = D*B \qquad (1.10)$$

Per assicurare che $\tilde{B} = B$ allora devo ottenere una matrice di decodifica D uguale a

$$D = pinv(C) = C^T \cdot (C \cdot C^T)^{-1} \qquad (1.11)$$

Il caso di layout regolari semplifica l'espressione della matrice di decodifica:

$$D = \frac{1}{N} C^T \qquad (1.12)$$

Con i layout regolari si preserva la condizione di ortonormalità propria delle armoniche sferiche (esse sono una base ortonormale perchè soddisfano i requisiti necessari). Ricordiamo infatti che il prodotto sferico scalare è:

$$\langle F \mid G \rangle_{4\pi} = \frac{1}{4\pi} \oiint F(\theta,\delta) G(\theta,\delta) d\Omega \qquad (1.13)$$

ma nel nostro caso non abbiamo un continuo su una sfera ma solo N componenti. Inoltre, nelle basi ortonormali, $A*A^T=1$.

La ricostruzione del campo è consentita solo fino ad una certa frequenza che dipende dalla grandezza dell'area. Occorre applicare quindi criteri e soluzioni diversificate in relazione alla frequenza considerata ("*max* r_E" e "*inphase*"). Questa tipo di decodifica ottimizzata è fatta semplicemente applicando guadagni g_m nelle appropriate bande di frequenza ai componenti B_{mn}^{σ} prima di processare la decodifica base:

$$D = \frac{1}{N} C^T \cdot Diag\left(\left[\cdots g_m \cdots\right]\right) \qquad (1.14)$$

1.2.2.1.1 Caso dell'onda piana

La decomposizione in armoniche sferiche di un'onda piana di incidenza (θs,δs) che trasporta un segnale S porta ad una semplice espressione del componente Ambisonic:

$$B_{mn}^{\sigma} = S * Y_{mn}^{\sigma}(\theta_s,\delta_s) \qquad (1.15)$$

Ciò vuol dire che una sorgente in campo lontano è codificata semplicemente applicando *guadagni reali* (dati dalle armoniche sferiche nella direzione d'arrivo) al

segnale di pressione S ricevuto. Per esempio, prendendo in considerazione il primo ordine, otteniamo:

$$\begin{bmatrix} W \\ X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \cos\theta\cos\delta \\ \sin\theta\cos\delta \\ \sin\delta \end{bmatrix} \bullet S \quad (1.16)$$

E' importante notare che il canale W è stato ridotto di 3 dB per portare all'equiparazione dei livelli sui 4 canali.

1.2.2.1.2 Direttività e funzioni di panning

Combinando le equazioni di codifica (1.15) e decodifica (1.14) si possono derivare funzioni di "panning" G(γ):

$$G(\gamma) = \frac{1}{N}\left(g_0 + 2\sum_{m=1}^{M} g_m \cos(m\gamma) \right) \quad \text{for 2D,}$$

$$G(\gamma) = \frac{1}{N}\sum_{m=0}^{M} (2m+1) g_m P_m(\cos\gamma) \quad \text{for 3D,}$$

Così che l'i-esimo speaker è alimentato dal segnale

$$S_i = S * G(\gamma) \qquad (1.17)$$

con $\gamma = \arccos(\vec{u}_i, \vec{u}_s)$ angolo tra la direzione dello speaker "u_i" e la direzione della sorgente "u_s".

Figura 1.17 – Aumento della direttività in funzione dell'ordine Ambisonic (Daniel,Nicol,Moreau, 2003)

1.2.2.1.3 Situazione di campo vicino

In genere, lavorando con un sistema Ambisonic, ci si trova davanti ad un anello più o meno regolare di loudspeakers. Se devo rappresentare sorgenti in campo lontano,

cioè più distanti del raggio R dell'anello di speaker, voglio che l'onda che giunge all'ascoltatore (posto nell'origine) sia piana; ho però un limite dovuto alla vicinanza degli speaker all'ascoltatore che naturalmente curvano il fronte d'onda per effetto di prossimità. Se poi devo rappresentare sorgenti all'interno dell'anello, poiché situate virtualmente vicino all'ascoltatore, avrebbero bisogno di un effetto di prossimità che conferisca loro realismo e che quindi, in fase di riproduzione, deve essere ricreato attraverso appositi filtri che emulino l'effetto di "*near-field*". Essi, per un ordine Ambisonic elevato, sono necessari quindi in tre situazioni:

- Compensazione degli effetti *near-field*. Ad un distanza, ad esempio, di 2 metri dagli speaker, l'effetto su componenti del secondo ordine (o di un ordine più elevato) non può più essere trascurato.
- Creazione di sorgenti sonore virtuali poste vicino all'ascoltatore. Nel contesto di musica elettro-acustica codificata in HOA e di realtà virtuale è fondamentale riuscire a creare sorgenti in prossimità dell'utente per conferire realismo alla simulaizone
- Sintesi di segnali HOA (High Order Ambisonic) catturati mediante microfono.

Caso 1

Supponiamo di dover effettuare la riproduzione di un sorgente lontana: il segnale, come già anticipato, non riesce a presentarsi all'ascoltatore come un'onda piana in quanto l'anello di altoparlanti è in prossimità dell'ascoltatore; i fronti d'onda sono ancora curvi perchè non hanno lo spazio per formarsi correttamente. Si rende necessario, in fase di codifica, applicare un filtro inverso che abbassi il guadagno elevato alle basse frequenze:

$$S = D \cdot Diag\left(\left[\cdots \frac{1}{F_m^{(R/c)}(\omega)} \cdots\right]\right) \cdot B \qquad (1.18)$$

Questa codifica è praticabile finché i filtri inversi sono stabili e permette di mantenere i fronti d'onda nella forma originale. Infatti, senza la compensazione di campo vicino, un'onda piana codificata viene ricostruita come un'onda sferica che arriva dall'array di speaker.

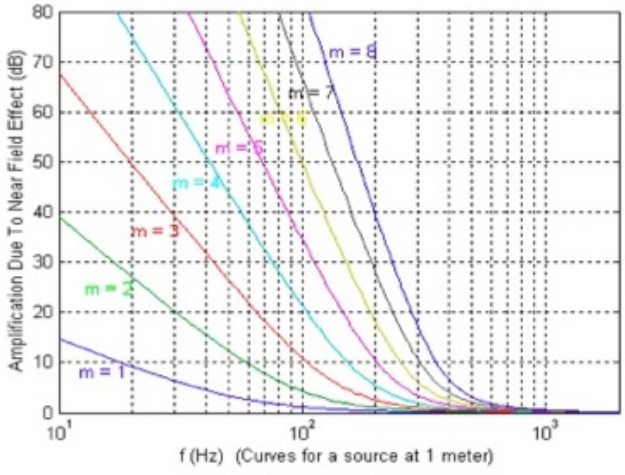


Figura 1.18 – Boost infinito alle basse frequenze per i componenti ambisonic causato dall'effetto di campo vicino (Daniel,Nicol,Moreau, 2003)

Caso 2

Supponiamo di voler riprodurre una sorgente posta tra l'array di speaker e l'ascoltatore: man mano che questa si avvicina al punto d'ascolto, deve aumentare per effetto di prossimità il guadagno alle basse frequenze. Tale effetto (*Near-field effect*) in un ordine *m* corrisponde semplicemente ad un filtro analogico che è l'inverso di un filtro passa-alto di ordine *m*. Ciò implica un guadagno DC infinito alle basse frequenze. Ogni offset di voltaggio (in una realizzazione analogica) o un bias dovuto all'errore di arrotondamento (per una realizzazione digitale) rischia di tradursi in un componente DC fuori controllo. Da qui la necessità di avere un meccanismo di controllo, un feedback, che riduca la componente infinita in un guadagno finito (*finite gain shelf-filter*) o con un filtro inverso per campo vicino che corrisponda ad una sorgente più lontana.

Jerome Daniel [6], per descrivere il comportamento di un'onda sferica, utilizza le precedenti equazioni per ottenere la codifica di una sorgente ad una distanza finita ρ:

$$B_{mn}^{\sigma} = S \bullet F_{m}^{(\rho/c)}(\omega) \bullet Y_{mn}^{\sigma}(\theta,\delta) \qquad (1.19)$$

$$F_{m}^{(\rho/c)}(\omega) = \sum_{n=0}^{m} \frac{(m+n)!}{(m-n)!n!2^{n}} \left(\frac{-jc}{\omega\rho} \right)^{n} \quad \text{, with } \omega=2\pi f \qquad (1.20)$$

dove *S* rappresenta la pressione nell'origine mentre 1/ρ un'attenuazione e ρ/c un ritardo (invertito perchè siamo in frequenza) dovuti alla distanza finita.

Da notare la presenza di 2^n: nella formula di Jerome Daniel non è presente a causa di un errore tipografico confermato dallo stesso autore.

Poichè i filtri $F_m^{(\rho/c)}$ sono integratori e quindi instabili per loro natura alle basse frequenze, la codifica HOA (High Order Ambisonic) non è in grado di rappresentare fisicamente (per mezzo, ad esempio, di segnali ad ampiezza finita) campi sonori naturali, finché questi includono sorgenti più o meno in campo vicino: si ha un aumento di guadagno infinito alle basse frequenze dei componenti Ambisonic, dovuto all'effetto di campo vicino, che deve essere reso con un guadagno finito.

Chiamiamo, dalla precedente formula (1.20):

$$a_{m,n} = \frac{(m+n)!}{(m-n)!n!2^n} \quad (1.21)$$

$X = \frac{c}{sr}$, dove $s = j\omega = j2\pi f$ e *r* la distanza dalla sorgente.

I valori di $a_{m,n}$ per ordini che vanno da 1 a 4 sono:

m	**$a_{m,n}$**
1	1, 1
2	1, 3, 3
3	1, 6, 15, 15
4	1, 10, 45, 105, 105

Tabella 1.1 – Valori di $a_{m,n}$ in funzione dell'ordine m

Per la realizzazione pratica dei filtri occorre fattorizzare i polinomiali in *s* nel primo e secondo ordine. Considerandoli essere polinomiali in X occorre fare tale operazione solo una volta:

m	**1**	**X**	**X^2**
1	1	1	
2	1	3	3
3	1	3.6778	6.4595
	1	2.3222	
4	1	4.2076	11.4877
	1	5.7924	9.1401

Tabella 1.2 – Fattorizzazione dei polinomiali del primo e secondo ordine

Date queste fattorizzazioni, si possono ora esprimere, per ogni ordine *n* e per ogni valore di *c*, come un prodotto di sezioni:

$$\begin{aligned} H_1(s) &= 1 + b_{1,1}s^{-1} \\ H_2(s) &= 1 + b_{2,1}s^{-1} + b_{2,2}s^{-2} \end{aligned}$$

La via standard per trasformare un filtro analogico in uno digitale nel dominio *z* è di usare la *traformata bilineare* che consiste nella sostituizone:

$$s = 2F_s \frac{1 - z^{-1}}{1 + z^{-1}}$$

dove F_s è la frequenza di campionamento.

Andando ad effettuare le sostituzioni si ottengono i filtri. Il problema è che per valori realistici di *c*, *r* e F_s l'effetto dipende fortemente da piccole variazioni di valori: la precisione in campo digitale potrebbe non essere sufficiente per avere un effetto dei filtri soddisfacente perchè tali filtri producono in uscita un guadagno DC infinito. La versione di filtri sviluppata da F. Adriaensen [11] permette di eliminare il problema introducendo un piccolo ritardo: i nuovi filtri danno ancora un guadagno infinito alle basse frequenze ma poiché la loro espressione contiene un ritardo si possono usare termini di feedback per introdurre un filtro inverso di *near field*. Ciò significa che nel caso intervenga un boost pressoché infinito possa subentrare il filtro inverso di near field che limita il guadagno della componente DC.

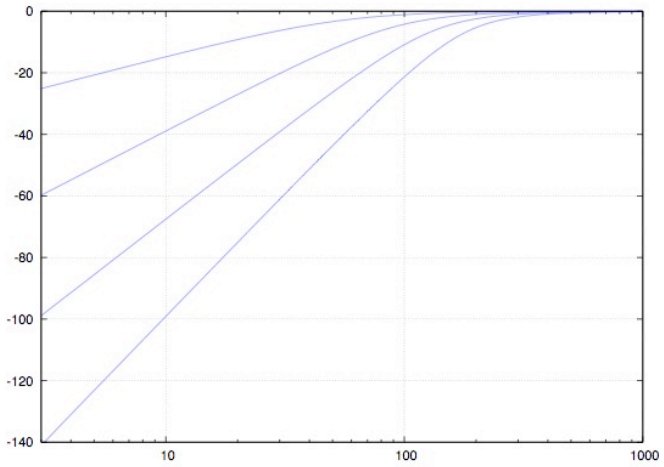


Figura 1.19 - Filtro inverso per una sorgente posta ad 1m

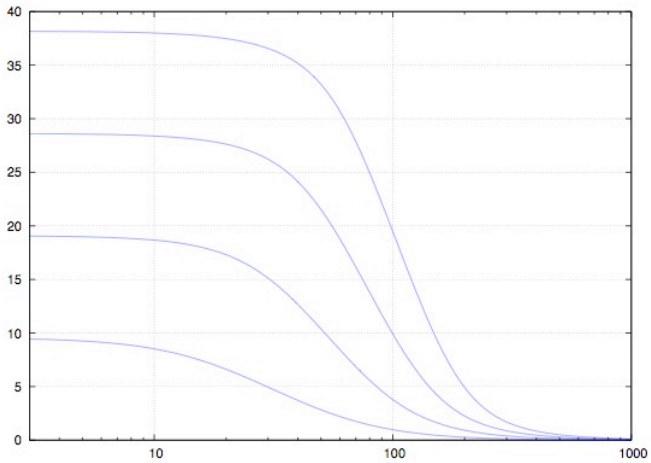


Figura 1.20 - Filtri per una sorgente virtuale ad 1m e a 3m di distanza

Caso 1 e 2

Si può pensare di introdurre la compensazione per campo vicino già nello stadio di codifica e non più in quello di decodifica. Si giunge quindi ai "*Distance (or Near Field) Coding Filters*":

$$H_m^{NFC(\rho/c,R/c)}(\omega) = \frac{F_m^{\rho/c}(\omega)}{F_m^{R/c}(\omega)} \qquad (1.22)$$

Essi sono caratterizzati da un'amplificazione finita alle basse frequenze [m x 20log(R/ρ)] in dB che risulta positiva (aggiunta di un boost) per sorgenti dentro l'anello (ρ<R) e negativa (eliminazione di un boost) per sorgenti fuori dall'anello (ρ>R).

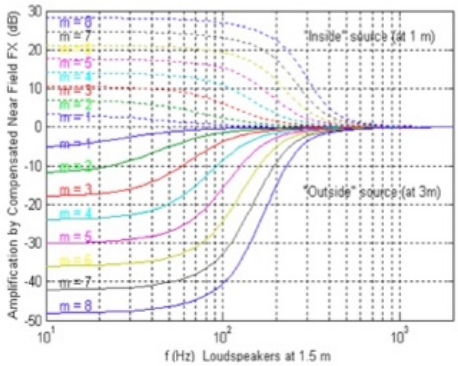


Figura 1.21 – Amplificazione finita di componenti ambisonic (Daniel,Nicol,Moreau, 2003)

Tali filtri vanno così inseriti nella formula:

$$B_{mn}^{\sigma NFC(R/c)} = S \cdot H_m^{NFC(\rho/c,R/c)}(\omega) Y_{mn}^{\sigma}(\theta,\delta) \qquad (1.23)$$

Si giunge quindi ad un nuovo formato denominato "*Near Field Compensated High Order Ambisonic (NFC-HOA)*":

$$B_{mn}^{\sigma NFC(R/c)} = \frac{1}{F_m^{(R/c)}(\omega)} B_{mn}^{\sigma} \qquad (1.24)$$

dove R rappresenta la distanza dagli speakers.

Questo formato può rappresentare qualsiasi campo sonoro realistico attraverso segnali ad ampiezza finita e richiede solo la decodifica ($\mathbf{D} = \frac{1}{N}\mathbf{C}^T$). Tutto ciò che è stato descritto in precedenza sul "panning" può essere affrontato sostituendo le espressioni e i filtri qui appena descritti.

1.2.2.1.4 Energia e velocità

Nel momento in cui si sviluppa un decoder in grado di passare il giusto segnale alle casse per una corretta ricostruzione del campo, non basta considerare la forma dell'array di casse e la loro distanza dall'ascoltatore, ma occorre considerare la percezione psicofisica di chi ascolta. Gerzon fu il primo a studiare e sviluppare approcci efficienti per la progettazione e il disegno di decoder tenendo conto di tali meccanismi percettivi psicoacustici.

Sviluppando una *Metatheory* [9] [10], evidenziò due criteri principali che sono stati usati in modo estensivo per disegnare decoder Ambisonic. Tali criteri sono basati sui vettori di velocità ed energia, rispettivamente per basse ed alte frequenze.

<u>Velocità</u>

Questo criterio di localizzazione è utilizzato per la predizione della direzione del suono per le basse frequenze (<700 Hz), dove la differenza di fase interaurale è la

principale informazione per la localizzazione e dove la diffrazione della testa è trascurabile.

Data una combinazione di onde piane (P_n, ϕ_n) al punto r=0, il vettore velocità è definito come somma di u_n vettori pesati dai loro rispettivi guadagni di ampiezza G_n.

$$V = \frac{\sum_{n=0}^{N-1} G_n u_n}{\sum_{n=0}^{N-1} G_n} = r_V \cdot u_V \qquad (1.25)$$

$$\text{dove } r_v \geq 0 \text{ e } u_V = \begin{pmatrix} \cos\theta_V \\ \sin\theta_V \end{pmatrix}$$

E' importante ricordare che:

$$V = -\frac{j}{k}\frac{\vec{\nabla}p}{p} = \frac{1}{k}\vec{\nabla}\varphi_p - j\frac{1}{2k}\vec{\nabla}|p|^2$$

dove la parte reale rappresenta la propagazione della fase e la parte immaginaria la propagazione dell'energia.

Nel caso di un singola sorgente fantasma, V è relativa alla descrizione al primo ordine del campo d'onda ricostituito da

$$V = \left(\frac{X}{\sqrt{2}W}, \frac{Y}{\sqrt{2}W}\right) \qquad (1.26)$$

Ciò significa che le soluzioni di decodifica pseudo-inversa, automaticamente verificano r_v=1 e θ_v=Ψ, dove φ rappresenta l'angolo di incidenza rispetto a x:

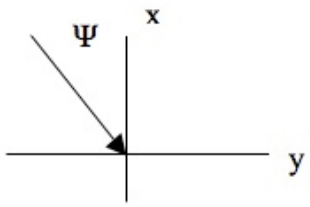


Figura 1.22 – Angolo d'incidenza

Quando le distanze dal punto centrale sono piccole rispetto alla lunghezza d'onda è possibile considerare il campo d'onda sintetizzato come un'onda piana. E' possibile

provare tramite formule matematiche che r_v è in relazione con la velocità di propagazione lungo la direzione $-u_v$.

Ne risulta che il campo d'onda ricostruito è localmente un'onda piana che viene dall'azimuth θ_v ed ha una velocità di propagazione pari a $c'=c/r_v$ lungo quella direzione, dove le onde piane "naturali" hanno una velocità pari a c.

Ciò significa che per basse frequenze la ITD (Interaural Time Difference) è differente da quella che si avrebbe naturalmente quando $r_v \neq 1$

$$\left.\begin{aligned} ITD_{natural} &= \frac{d}{c}\sin(\theta_V - \gamma) \\ ITD_{synthetic} &= \frac{d}{c'}\sin(\theta_V - \gamma) \end{aligned}\right\} \Rightarrow ITD_{synthetic} = r_V ITD_{natural} \qquad (1.27)$$

Se θ_p è la direzione percepita dalla sorgente e θ_v l'angolo predetto:

$$\frac{d}{c'}\sin(\theta_V - \gamma) = \frac{d}{c}\sin(\theta_p - \gamma) \Rightarrow \sin(\theta_p - \gamma) = r_V \sin(\theta_V - \gamma) \qquad (1.28)$$

- Quando $r_v=1$ l'angolo $\theta_p = \theta_v$ e l'immagine è stabile anche se la testa ruota.
- Quando $r_v<1$ la direzione percepita diventa più vicina all'asse x (asse mediano interaurale) rispetto a quella predetta.
- Quando $r_v>1$ c'è una lateralizzazione accentuata.

In generale si può dire che $r_v<1$ genera mancanza di precisione nella localizzazione. Avere un buon effetto di lateralizzazione è importante non solo per la localizzazione ma anche per preservare impressioni spaziali buone, come ad esempio l'avvolgimento.

Energia

Nel dominio di quelle frequenze per cui la ricostruzione acustica non è più valida per rendere la diffrazione dovuta alla testa e dove all'orecchio arriva una somma statistica di pressione fuori fase, sembra molto più sensato considerare una somma

della distribuzione energetica spaziale. Ecco quindi che Gerzon introdusse il vettore E (energia).

$$E = \frac{\sum_{n=0}^{N-1} G_n{}^2 u_n}{\sum_{n=0}^{N-1} G_n{}^2} = r_E \cdot u_E \qquad (1.29)$$

$$\text{dove } 0 \le r_V \le 1 \text{ e } u_E = \begin{pmatrix} \cos\theta_E \\ \sin\theta_E \end{pmatrix}$$

θ_E rappresenta l'angolo della direzione prodotta per le alte frequenze; r_E rappresenta la concentrazione di energia nella direzione θ_E ed è <1 in presenza di una sola sorgente.

Avere r_E <1 per alte frequenze implica lo stesso tipo di comportamento dell'immagine che si ha quando r_V <1 per basse frequenze (che vuol dire una mancanza di lateralizzazione ed un effetto di elevazione dell'immagine sonora). Più in generale aumenta la sfuocatura della localizzazione.

Per entrambe le componenti la lunghezza del vettore rappresenta una misura della "qualità" della localizzazione (se ha valore pari a 1 indica un buon effetto di localizzazione), l'angolo del vettore rappresenta la direzione da cui si percepisce che il suono provenga.

Supponiamo di avere un array di speaker. Il decoder Ambisonic non fa altro che assegnare ad ogni speaker il segnale di un microfono virtuale puntato in quella direzione. Cambiando il *polar pattern* di tale microfono cambia quindi il suono prodotto dallo speaker corrispondente. Per array di speakers regolari, la progettazione e l'ottimizzazione di un decoder Ambisonics è solo questione di utilizzare la risposta di un microfono virtuale per basse frequenze ed un altro microfono leggermente diverso per le alte frequenze utilizzando *shelving filters*. Essi sono filtri che discriminano il guadagno da applicare in base alla frequenza.

Finché il *polar pattern* del microfono virtuale è lo stesso per tutti gli speakers, l'angolo di localizzazione supposto è sempre lo stesso di quello della sorgente codificata, con solo la qualità della localizzazione (lunghezza del vettore) affetta dal cambiamento dei polar pattern.

Di seguito vengono riportate le equazioni dei due vettori:

$$P = \sum_{i=1}^{n} g_i \qquad E = \sum_{i=1}^{n} g_i^2$$

$$Vx = \sum_{i=0}^{n} g_i \cos(\theta_i)/P \qquad Ex = \sum_{i=0}^{n} g_i^2 \cos(\theta_i)/E$$

$$Vy = \sum_{i=0}^{n} g_i \sin(\theta_i)/P \qquad Ey = \sum_{i=0}^{n} g_i^2 \sin(\theta_i)/E$$

Dove g_i = rappresenta il guadagno di uno speaker (assunto reale per semplicità), n è il numero degli speakers, θ_i è la posizione angolare del i^{th} speaker.

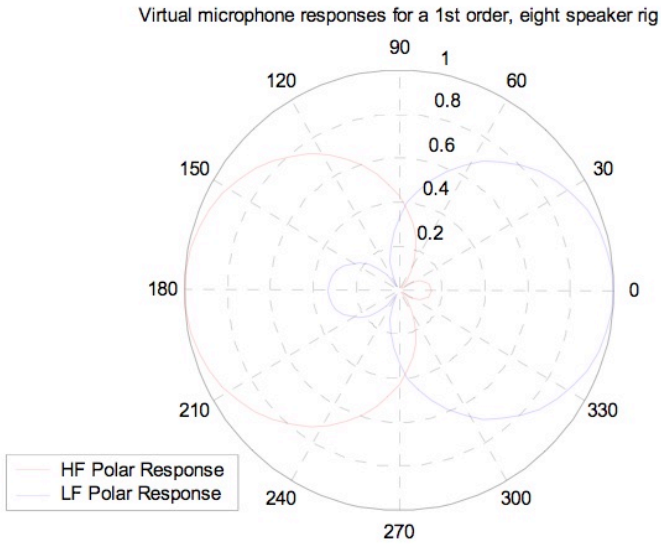


Figura 1.23 – Polar pattern dei microfoni virtuali per anello regolare, in alta e bassa frequenza

Quando vengono utilizzati layout irregolari, non solo la *magnitude* dei vettori necessita una compensazione ma occorre tenere conto anche del volume complessivo del suono prodotto, altrimenti si può incappare in artefatti di decodifica.

A questo proposito si prenda in considerazione la configurazione non regolare ITU a 5 speaker. Se tutti gli speaker hanno lo stesso polar pattern allora il suono che arriva dalla parte frontale sarà senz'altro più "sonoro" di quello che arriva dal retro. Inoltre anche la direzione percepita del suono risulterà distorta.

Per valutare la performance di un certo layout, regolare o irregolare) occorre utilizzare le equazioni sottostanti. Le condizioni che devono essere soddisfatte sono:

- la lunghezza del vettore (R_e o R_v) deve essere il più vicino possibile a 1

- $\theta = \theta_v = \theta_e$ per tutti i valori di θ (dove θ è l'angolo della sorgente codificata)
- Pv=Pe e devono essere costanti per tutti i valori di θ

$$Vx = \sum_{i=1}^{N} g_i \times \cos(SPos_i)/P_V$$

$$Vy = \sum_{i=1}^{N} g_i \times \sin(SPos_i)/P_V$$

$$Ex = \sum_{i=1}^{N} g_i^2 \times \cos(SPos_i)/P_E$$

$$Ey = \sum_{i=1}^{N} g_i^2 \times \sin(SPos_i)/P_E$$

$$R_E = \sqrt{E_x^2 + E_y^2} \qquad \theta_E = \tan^{-1}\left(E_y/E_x\right)$$

$$R_V = \sqrt{V_x^2 + V_y^2} \qquad \theta_V = \tan^{-1}\left(V_y/V_x\right)$$

$$P_V = \sum_{i=1}^{n} g_i \qquad P_E = \sum_{i=1}^{n} g_i^2$$

where:
g_i = Gain of the i^{th} speaker
$SPos_i$ = Angular position of the I^{th} speaker.

In pratica le condizioni qui definite sono difficili da soddisfare perchè il miglior risultato deve essere trovato su tutti i 360° di posizioni della sorgente codificata. Per un sistema a 5 speaker è un risultato difficile, se non impossibile, da raggiungere.
Non esiste comunque una sola possibilità di decodifica ottima ma occorre cercare e valutare mediante test i set più efficaci.

1.2.2.2 Registrazione Ambisonic

Il primo esempio di microfono Ambisonic nacque negli anni '70 dalle teorie di *Michael Gerzon*, sfruttando una tecnica non lontana da quella Blumlein. Sfruttando tre microfoni a *figura di otto* e un microfono omnidirezionale, teoricamente coincidenti, è possibile campionare le armoniche sferiche di ordine 0 e ordine 1: il segnale omnidirezionale dà le informazioni sulla pressione (W), i tre segnali rimanenti, orientati secondo gli assi cartesiani, sono proporzionali alla velocità delle particelle d'aria attorno al punto d'origine. Non avendo a disposizione 4 microfoni coincidenti si pensò di utilizzare 4 capsule con direttività a *cardioide* poste su quattro degli otto vertici di un cubo dalle dimensioni estremamente ridotte (Figura 1.25).

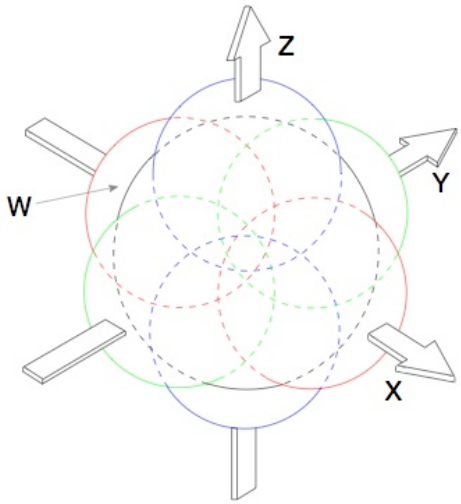


Figura 1.24 – Diagramma polare dei segnali B-format

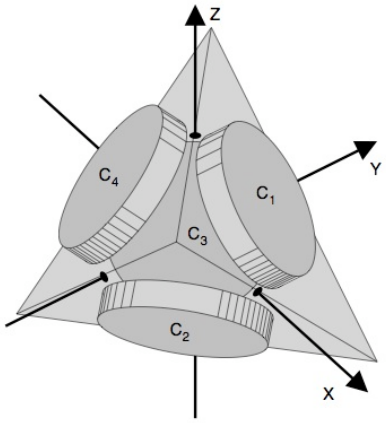


Figura 1.25 – Disposizione delle capsule in un microfono della stessa tipologia del Soundfield

La conversione tra i segnali delle capsule (A-format) ed i segnali Ambisonic del 1° ordine (B-format) avviene tramite una serie di somme e sottrazioni:

$$W = C1+C2+C3+C4 \quad (1.30)$$

$$X = C1+C2-C3-C4 \quad (1.31)$$

$$Y = C1-C2+C3-C4 \quad (1.32)$$

$$Z = C1-C2-C3+C4 \quad (1.33)$$

Anche in questo caso le capsule non sono coincidenti: l'approssimazione introdotta si traduce nel disallineamento tra le capsule in termini di tempo/fase che porta alla

"colorazione" (filtraggio) dello spettro dei segnali B-format. Esso però non influisce sull'ampiezza del segnale in quanto le capsule sono molto vicine l'una all'altra.

Un microfono di questa tipologia presenta indubbi vantaggi, primo fra tutti il condensare 4 microfoni in una sola unità. L'altro grosso vantaggio è quello di poter manipolare i siegnali B-format per ruotare ed inclinare virtualmente il microfono, incrementarne la direttività, sintetizzare un infinito numero di microfoni con direttività variabile.

Consideriamo le seguenti formule:

$$g_W = \sqrt{2}$$
$$g_X = \cos(\theta)\cos(\alpha)$$
$$g_Y = \sin(\theta)\cos(\alpha)$$
$$g_Z = \sin(\alpha)$$

$$S = 0.5 \times \left[(2-d)g_W W + d(g_X X + g_Y Y + g_Z Z)\right] \qquad (1.34)$$

dove:
S = segnale del microfono sintetizzato
θ = azimuth del microfono sintetizzato
α = elevazione del microfono sintetizzato
d = fattore diretivo ($0 \le d \le 2$)

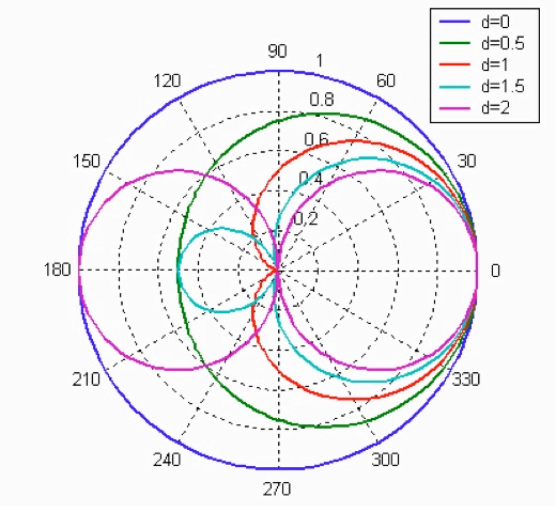


Figura 1.26 – Direttività variabile ottenuta con segnali B-format (Wiggins, 2004)

Assegnando un valore al fattore *d* è possibile sintetizzare diversi tipi di microfono:

Fattore *d*	**Polar pattern**
0	omnidirezionale
0.5	subcardioide
1	cardioide
1.5	ipercardioide
2	Figura di otto

Tabella 1.3 – Polar pattern del microfono sintetizzato in funzione del coefficiente d

Per avere la rotazione attorno all'asse Z, prendendo θ come angolo di rotazione, i segnali B-format si trasformano come nelle formule seguenti:

$$\begin{aligned} W' &= W \\ X' &= X \cdot \cos(\theta) + Y \cdot sen(\theta) \\ Y' &= X \cdot \cos(\theta) - Y \cdot sen(\theta) \\ Z' &= Z \end{aligned} \qquad (1.35)$$

Per avere la rotazione attorno all'asse X:

$$\begin{aligned} W' &= W \\ X' &= X \\ Y' &= Y \cdot \cos(\theta) - Z \cdot sen(\theta) \\ Z' &= Z \cdot \cos(\theta) + Y \cdot sen(\theta) \end{aligned} \qquad (1.36)$$

Poiché proprietà come queste possono avere un vasto numero di applicazioni, in rete è presente un VST Plug-in, Visual Virtual Microphone che permette di sintetizzare un arbitrario numero di microfoni partendo dal segnale B-format.

Figura 1.27 – Interfaccia del Visual Virtual Microphone

Come si può vedere dall'immagine dell'interfaccia grafica (Figura 1.27), il Visual Virtual Microphone consente di cambiare l'azimut e l'elevazione del microfono, creare in uscita configurazioni di microfoni con varia direttività, il tutto anche in tempo reale. Ad esempio, nel caso si disponga di un microfono B-format, è possibile ottenere, utilizzando tale plug-in VST, i 5 canali dello standard ITU 5.1: al posto dei cinque microfoni richiesti se ne può utilizzare uno solo.

Figura 1.28 – Soundfield ST250, Soundfield SPS200, DPA-4 e TetraMic

Il primo microfono di questo tipo ad essere commercializzato fu il Soundfield, tutt'ora il più diffuso, ma oggi è possibile trovare altre microfoni con disposizione A-format di dimensioni minori e pari qualità, come quelli mostrati in Figura 1.28.

1.2.2.2.1 Microfoni HOA

Se nel caso di riproduzione ambisonic si vuole ottenere segnali che pilotino gli altoparlanti a partire dai segnali ambisonic, nel caso di un microfono che "campioni" le armoniche spaziali il discorso si ribalta: da segnali registrati si vuole quindi ricostruire i segnali ambisonic.
La seguente formula di tali segnali

$$B_{mn}^{\sigma} = \frac{1}{i^m j_m(kr)} \iint_s p(r,\theta,\delta) Y_{mn}^{\sigma}(\theta,\delta) dS \quad (1.37)$$

suggerisce un metodo per stimare segnali HOA dalla registrazione della pressione acustica su una superficie sferica continua di raggio R.

Un microfono teorico può essere rappresentato come nella figura sottostante (Figura 1.29).

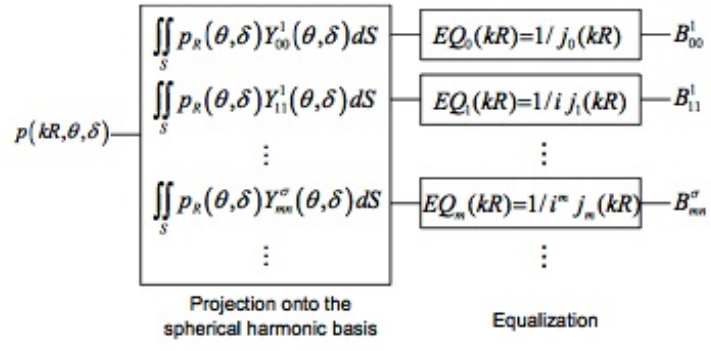


Figura 1.29 – rappresentazione di un microfono HOA

E' importante notare che questo metodo è valido solo se $j_m(kR) \neq 0$. I grafici delle funzioni di Bessel mostrano come, fissato un raggio R ed un ordine ben preciso, all'interno del range kR, la funzione j(kR) assume regolarmente il valore 0.
In più, quando $EQ_m(kR)$ è definito, porta una grossa amplificazione dei segnali per kR attorno agli zeri di $j(kR)$.
Tutto ciò porta ad una indeterminazione del campo misurato per alcune frequenze: tali frequenze sono funzione del raggio e dell'ordine e corrispondono agli zeri delle funzioni di Bessel.
Prendiamo, ad esempio, il caso del segnale B_{00}^1: esso non può essere stimato con un array microfonico sferico alle frequenze $f_a=ac/2R$ dove a=intero positivo, c=velocità del suono, R=raggio della sfera, k=2R numero d'onda.
Una soluzione per evitare frequenze che presentino indeterminazioni è quella di usare microfoni direzionali [5].

1.2.2.2.2 Formulazione discreta del problema

Ovviamente, non avendo a disposizione una sfera con superficie continua, occorre distribuire una serie di sensori su una superficie sferica in modo da "campionare" il suono. Prendiamo allora, su una superficie di raggio R, un set di Q sensori posti nel punto con coordinate (θ_q,δ_q), $1 \leq q \leq Q$. Si pensino come condizioni al contorno a due ipotesi:

- si considera una superficie sferica di una sfera piena ed i sensori sono omnidirezionali
- si considerano microfoni direzionali

Il q-esimo sensore rileva il seguente segnale acustico:

$$S_q = \sum_{m=0}^{+\infty} W_m(kR) \sum_{n=0}^{m} \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta_q, \delta_q) \qquad (1.38)$$

dove

$$W_m(kR) = \begin{cases} \alpha j_m(kR) + i(\alpha - 1) j_m'(kR) \\ i^{-m+1} (kR)^2 h_m^{-'}(kR) \end{cases} \qquad (1.39)$$

in cui la prima parte è per microfoni direzionali e la seconda per la sfera rigida. Si stimino segnali HOA fino ad un ristretto ordine M così che il numero totale $K=(m+1)^2$ di componenti non superi il numero Q di sensori.

Il sistema risolvente le equazioni può essere scritto così:

$$T \bullet b = s \qquad (1.40)$$

dove

$$\mathbf{T} = \begin{pmatrix} Y_{00}^1(\theta_1, \delta_1) & \cdots & Y_{M0}^1(\theta_1, \delta_1) \\ \vdots & \vdots & \vdots \\ Y_{00}^1(\theta_Q, \delta_Q) & \cdots & Y_{M0}^1(\theta_Q, \delta_Q) \end{pmatrix} \cdot \mathrm{diag}[W_m(kR)]$$

$$= \mathbf{Y} \cdot \mathrm{diag}[W_m(kR)], \quad 0 \le m \le M$$

$$\mathbf{s} = \left(S_1, S_2, \ldots, S_Q\right)^t, \ \mathbf{b} = \left(B_{00}^1, B_{11}^1, B_{11}^{-1}, B_{10}^1, \ldots, B_{M0}^1\right)^t.$$

Per una geometria sferica dell'array si ha che:

$$T = Y \bullet (\text{matrice diagonale}) \qquad (1.41)$$

dove:

Y = matrice reale che ha per colonne y_{mn}^{σ}, le componenti armoniche sferiche
matrice diagonale = filtri dipendenti dal raggio

Il vettore *S* contiene i segnali rilevati dai Q sensori e *b* rappresenta il vettore i cui elementi sono i segnali HOA sconosciuti fino all'ordine finito M. Finché la matrice *T* è una (Q x k) con più righe che colonne (Q≥k), il sistema è sovra determinato. Tramite passaggi che qui non vengono riportati per semplicità si arriva al seguente schema:

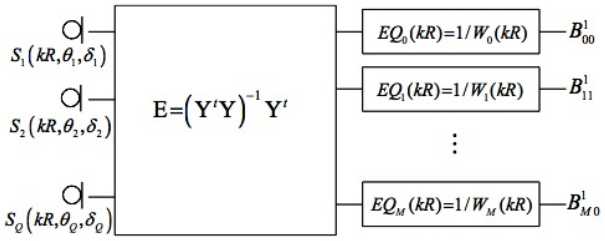


Figura 1.30 – Schema di conversione dei segnali microfonici in segnali ambisonic

1.2.2.2.3 Aliasing spaziale e rumorosità

In accordo con il criterio di Shannon, la più ampia distanza *d* tra i sensori definisce una frequenza limite sopra la quale interviene l'aliasing spaziale:

$$f_{al} = \frac{c}{2d} = \frac{c}{2R\gamma} \qquad (1.42)$$

dove :
c = velocità del suono
R = raggio array microfonico
γ = massimo angolo tra i sensori

L'aliasing spaziale comporta un sottocampionamento delle armoniche sferiche. Il numero di armoniche distinguibili dall'array dipende sia dal numero di sensori sia dalla loro posizione sulla sfera. Ciò vuol dire che per stimare *k* differenti armoniche spaziali fino ad un ordine massimo *M,* come minimo servono *k* sensori.
In più, la posizione del sensore deve soddisfare le condizioni di "ortonormalità discreta" per le armoniche sferiche campionate:

$$\frac{1}{Q} y_{mn}^{\sigma\ t} \cdot y_{m'n'}^{\sigma'} = \delta_{mm'}\delta_{nn'}\delta_{\sigma\sigma'} \qquad (1.43)$$

Se le armoniche sferiche di ordine maggiore di M sottocampionate sono presenti nel campo sonoro, esse ricadono sulle componenti di ordine inferiore e possono seriamente compromettere la stima dei segnali HOA.

In conclusione, per minimizzare gli effetti dell'aliasing spaziale, in generale i sensori devono essere di numero maggiore delle armoniche sferiche da campionare: $Q>k$.

Man mano che si sale con l'ordine delle armoniche sferiche si hanno problemi sempre maggiori di rumorosità, una rumorosità insita nel sistema stesso di calcolo della matrice di codifica. Si rendono quindi necessari filtri che reiettino questo rumore tenendo alto il segnale utile; sono filtri che interessano le bande a bassa frequenza degli ordini elevati. Se si assume che la matrice Y di armoniche sferiche campionate soddisfi le condizioni di ortonormalità, si può scrivere:

$$b_{REG} = diag\left[F_m(kR) \cdot \frac{1}{W_m(kR)}\right] Es \qquad (1.44)$$

dove F(kR) rappresenta i filtri di regolarizzazione mentre b_{REG} contiene i segnali HOA regolarizzati. Il metodo di regolarizzazione più utilizzato è quello di Tikhonov, metodo che porta a filtri di forma

$$F_m(kR) = \frac{|W_m(kR)|^2}{|W_m(kR)|^2 + \lambda^2} \qquad (1.45)$$

dove λ rappresenta il parametro di regolarizzazione. Se $\lambda = 0$ allora non ho regolarizzazione mentre se aumento arrivo ad avere una sovra-regolarizzazione. Per capire quale debba essere il valore del parametro di regolarizzazione risulta utile le seguente formula che lega il filtro di regolarizzazione ad un solo fattore di amplificazione *a*:

$$\lambda = \frac{1 - \sqrt{1 - \frac{1}{a^2}}}{1 + \sqrt{1 - \frac{1}{a^2}}} \qquad (1.46)$$

dove a è in stretta relazione con a_s (fattore di amplificazione del rumore sulla singola capsula espresso in dB).

$$a = \sqrt{Q}10^{\frac{a_s}{20}} \qquad (1.47)$$

1.2.2.2.4 Approccio alternativo

E' possibile affrontare il problema non partendo da un approccio "ambisonic" che tiene conto di una ripartizione della sorgente sonora attorno all'origine. Questo approccio alternativo infatti permette di ottenere Ambisonic di ordine elevato da un array formato da capsule di qualsiasi tipo (omni, caridoide, otto,ecc...) e poste in un qualunque modo nello spazio (non necessariamente equispaziate).
Lo scopo è quello di determinare una rappresentazione del campo il più vicino possibile al modello di campo sconosciuto, conoscendo soltanto i segnali delle capsule e le loro caratteristiche.
Partiamo dalla descrizione del campo attraverso la formula:

$$P(r,\theta,\phi,f) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} P_{l,m}(f) j^l j_l(kr) y_l^m(\theta,\phi) \qquad (1.48)$$

dove k=2πf/c e c=velocità del suono.
L'Ambisonic considera soltanto la parte angolare di questa decomposizione (le armoniche sferiche). Per descrivere pienamente il campo 3D occorre utilizzare anche le funzioni di Bessel.

Campionamento spaziale

La registrazione di un campo acustico è fattibile utilizzando un set di capsule. L'input è quindi il campo acustico mentre l'output è formato dai segnali in uscita dal microfono: le capsule effettuano un campionamento del campo in cui sono immerse. In analogia con il campionamento temporale, è possibile trovare un equivalente del teorema di Shannon, anche se in questo caso la situazione è leggermente più

complicata perchè ci troviamo con un layout irregolare. Per limitare l'aliasing spaziale, la decomposizione di Fourier-Bessel è troncata ad un certo ordine L. Quest'ordine dipende soprattutto dal numero N di capsule utilizzate. In generale si può dire che L debba essere preso in modo da soddisfare la seguente regola

$$(L+1)^2 \leq N \qquad (1.49)$$

Lo scopo di tutto questo è di ottenere la relazione di campionamento, la relazione che rende possibile il calcolo teorico dei segnali prodotti dall'array microfonico a partire da un campo sonoro noto. Questa relazione dipende dalle caratteristiche dell'array (posizione, orientazione e tipo di capsule).

Per esempio, se le capsule sono omnidirezionali il campionamento che eseguono non è altro che il valore della pressione calcolata in quel punto. In questo caso, se si conosce il campo acustico in cui è inserito l'array, lo si conosce anche in quei determinati punti in cui sono posizionate le capsule, così che è possibile determinare il segnale teorico che quelle capsule dovrebbero restituire. Il campo acustico può essere conosciuto in modi differenti, incluso il modello precedentemente descritto, così che risulta possibile determinare i segnali che le capsule restituiranno se sono noti i coefficienti di Fourier-Bessel. Con capsule comuni (più precisamente capsule lineari), c'è una relazione lineare di matrice tra i segnali delle capsule e i coefficienti di Fourier-Bessel (coeff. nel dominio della frequenza).

Se i segnali delle capsule sono contenuti in un vettore $\boldsymbol{c}$ e se i coefficienti di Fourier-Bessel sono contenuti in un vettore $\boldsymbol{p}$, la relazione è data da

$$c = Bp$$

La matrice B è chiamata la matrice di campionamento e dipende solo dalle caratteristiche dell'array.

Essa rappresenta la via attraverso cui l'array microfonico campiona ed estrae informazioni dal campo acustico.

Codifica spaziale

La codifica spaziale ha lo scopo di investigare il campo acustico tridimensionale usando i segnali provenienti da poche capsule. L'obbiettivo è quello di calcolare i

coefficienti di Fourier-Bessel a partire ai segnali delle capsule invertendo la relazione di campionamento (c=Bp).

Intervengono purtroppo alcuni problemi:

- La matrice B non è in generale una matrice quadrata: ciò vuol dire che spesso B^{-1} non esiste.
- Le capsule non sono mai posizionate esattamente dove dovrebbero essere; un array microfonico reale ha sempre errori di posizionamento.
- La modellazione non è perfetta. Ad esempio, se utilizziamo capsule dicharate come omnidirezionali, dobbiamo considerar che esse non sono mai perfettamente omnidirezionali; misurano la pressione su una piccola superficie e non in un punto; la sensibilità varia da una capsula all'altra.
- Le capsule, anche quelle più belle, hanno una certa rumorosità.

Il primo problema è facilmente risolvibile attuando un'inversione anche per le matrici non quadrate. Anche ignorando gli altri punti il risutato a cui si giunge è decisamente insoddisfacente. Per esempio, a basse frequenze, la lunghezza d'onda raggiunge diversi metri e in questo modo, in un array relativamente piccolo, il segnale non differisce molto da capsula a capsula. Finchè le informazioni spaziali sono ottenute tramite la differenza di segnale da capsula a capsula, il rumore di queste capsule diventa via via sempre più importante con la diminuzione della frequenza.

Se viene effettuata una inversione diretta, il rumore avrà un impatto molto forte sui segnali derivati. Per risolvere il problema si introduce il parametro μ (varia tra 0 e 1) che specifica il compromesso desiderato tra la risoluzione spaziale e l'amplificazione del rumore. Un valore pari a 1 significa che non si vuole prendere in considerazione il rumore; più piccoli sono i valori di μ meno è il rumore amplificato ma più bassa è la risoluzione spaziale.

Il processo di codifica è determinato attraverso una matrice di codifica, E, che dà il campo stimato in funzione dei segnali delle capsule.

Se il segnale c contiene i segnali e il vettore p contiene i coefficienti di Bessel-Fourier desiderati del campo stimato, la relazione è data da

$$\hat{p} = Ec \qquad (1.50)$$

La matrice E è calcolata a partire da B (matrice di campionamento) e da μ seguendo la relazione:

$$E = \mu B^T \left(\mu B B^T + (1-\mu) I\right)^{-1} \qquad (1.51)$$

Dove I è la matrice identità e B^T la trasposta coniugata della matrice B.
Codificare un campo con un array microfonico richiede di specificare la matrice B e da essa ricavarsi la matrice E. Ogni coefficiente di E, dipendendo dalla frequenza, rappresenta un filtro che può essere parametrizzato. Per un array con N capsule e una codifica di ordine L, ci sono

$$N\,x\,(L+1)^2 \text{ filtri.}$$

La codifica è quindi fatta applicando questi filtri ai segnali che arrivano dalle capsule. Il tutto può essere schematizzato con lo schema seguente:

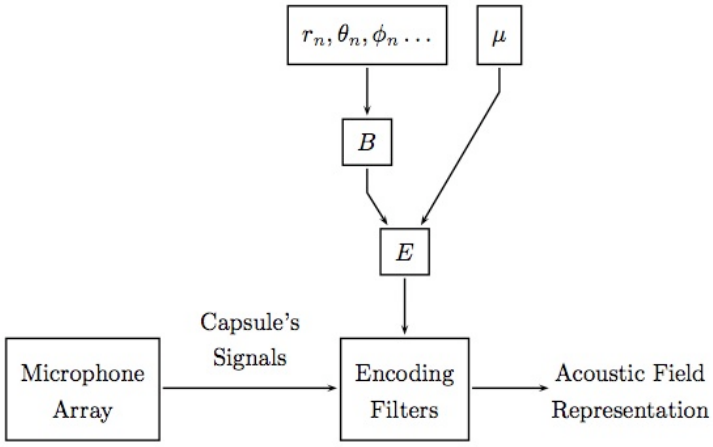


Figura 1.31 – Schema di codifica

Per capire meglio l'approccio da seguire occorre considerare gli N segnali delle capsule che, processati, devono dar luogo a L segnali (quelli ambisonic):

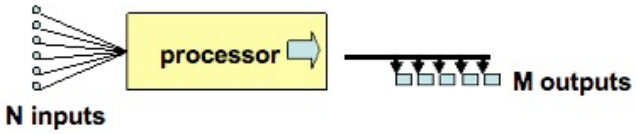


Figura 1.32 – Schema della conversione degli N-inputs in M-segnali ambisonic

Ciò che esegue il processing è un banco di filtri. Chiamiamo allora $\boldsymbol{x_i}$ i segnali degli N microfoni e $\mathbf{y_i}$ gli M segnali Ambisonic, ottenendo la seguente formula:

$$y_j = \sum_{i=1}^{N} h_{ij} \otimes X_i \qquad (1.52)$$

dove $\boldsymbol{h_{ij}}$ sono i filtri pre-calcolati. $\boldsymbol{y_j(t)}$ rappresenta la forma d'onda campionata nel dominio del tempo di un'onda con caratteristiche spaziali ben definite:

- un'onda sferica centrata in un punto ben preciso P_{source} di emissione
- un'onda piana con una certa direzione
- un'armonica sferica riferita ad un ben preciso punto P_{rec}

I filtri sono generalmente calcolati con procedimenti computazionali pesanti (come abbiamo visto in precedenza) e facendo affidamento sull'idealità ed uguaglianza di tutte le capsule che formano l'array. In alcune implementazioni, ogni singolo microfono è corretto mediante un suo filtro, aumentando la complessità di calcolo.

a) inversione del singolo microfono

L'approccio fin qui descritto mette da parte il calcolo puramente teorico dei filtri, derivandoli invece da un set di risposte all'impulso. In pratica viene creata una matrice di coefficienti che deve essere numericamente invertita (generalmente utilizzando parametri di regolarizzazione). In questo modo i segnali di uscita dei microfoni sono il più possibile vicino al segnale ideale. In questo modo, oltre a rendere i microfoni molto simili tra loro, si correggono tutte le imperfezioni e gli artefatti acustici di ognuno (riflessioni, diffrazioni, ecc...).

b) calcolo della matrice di codifica

Si faccia riferimento alla seguente immagine:

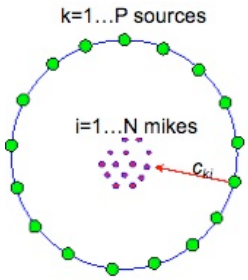


Figura 1.33 – Campionamento di suono proveniente da un numero finito di direzioni

Le risposte all'impulso dell'array microfonico $c_{k,j}$ sono misurate per un numero P di direzioni entranti, ottenendo la matrice C:

$$C=\begin{bmatrix} c_{1,1} & c_{2,1} & c_{k,1} & c_{P,1} \\ c_{1,2} & c_{2,2} & c_{k,2} & c_{P,2} \\ c_{1,i} & c_{2,i} & c_{k,i} & c_{P,i} \\ c_{1,N} & c_{2,N} & c_{k,N} & c_{P,N} \end{bmatrix}$$

Processando questa matrice C possiamo ottenere la matrice di filtri da applicare ai segnali microfonici.

Occorre disegnare i filtri in modo tale che la risposta del sistema sia la funzione prescritta teorica V_k per la k-esima funzione (una Delta di Dirac di ampiezza unitaria nel caso dell'esempio, come la funzione di ordine 0 omnidirezionale). Si definisca Vk il vettore di risultati conosciuti (saranno ovviamente differenti per diagrammi polari di ordini maggiori).

$$\begin{aligned} &\sum_{i=1}^{N} h_{i,0} \otimes c_{1,i} = \delta \\ &\sum_{i=1}^{N} h_{i,0} \otimes c_{2,i} = \delta \\ &\text{.....} \\ &\sum_{i=1}^{N} h_{i,0} \otimes c_{k,i} = \delta \\ &\text{.....} \\ &\sum_{i=1}^{N} h_{i,0} \otimes c_{p,i} = \delta \end{aligned} \qquad (1.53)$$

Una volta che questa matrice di N filtri è calcolata (per esempio utilizzando il metodo di Kirkeby), l'output dell'array microfonico, sintetizzante il prescritto diagramma polare di ordine 0, sarà dato da:

$$y_0 = \sum_{i=1}^{N} x_i \otimes h_{i,0} \qquad (1.54)$$

Per calcolare la matrice degli N coefficienti di filtraggio h_{kj} viene usato il metodo dei minimi quadrati. Seguendo tale metodo si calcola "l'errore quadratico totale" definito come:

$$\varepsilon_{tot} = \sum_{k=1}^{P}\left[\sum_{i=1}^{N}\left(h_{i0} \otimes c_{ki}\right) - v_k\right]^2 \qquad (1.55)$$

Si forma quindi un sistema di N equazioni lineari per minimizzare l'errore totale, imponendo che

$$\frac{\partial \varepsilon_{tot}}{\partial h_{i0}} = 0 \quad \text{(i=1...N)} \qquad (1.56)$$

Durante il calcolo del filtro inverso, di solito calcolato nel dominio della frequenza, ci si trova davanti ad espressioni che richiedono il calcolo del rapporto tra due spettri complessi:

$$H = \frac{A}{D} \qquad (1.57)$$

Calcolare il reciproco del denominatore D non è affatto semplice poichè l'inverso di un segnale complesso a fase mista è generalmente instabile. La regolarizzazione di Nelson/Kirkeby è generalmente utilizzata per questo tipo di problematiche.

$$InvD(\omega) = \frac{Conj\left[D(\omega)\right]}{Conj\left[D(\omega)\right] \cdot D(\omega) + \varepsilon(\omega)} \qquad (1.58)$$

$$H = A \cdot InvD \qquad (1.59)$$

Come mostra l'immagine seguente, occorre tenere valori alti del parametro di regolarizzazione per frequenze molto basse e molto alte.

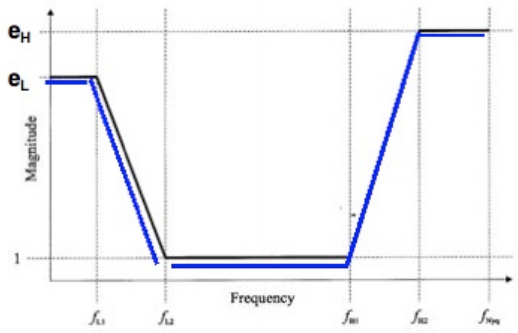


Figura 1.34 – Andamento sullo spettro del parametro di regolarizzazione

Nelle immagini che seguono sono riportati alcuni esempi di microfoni Ambisonic del 3° ordine e 4° ordine. La prima rappresenta un microfono del 3° ordine costruito dalla Trinnov: 24 capsule poste in una configurazione pseudo-casuale e da cui si possono ottenere i 16 segnali ambisonic del terzo ordine.

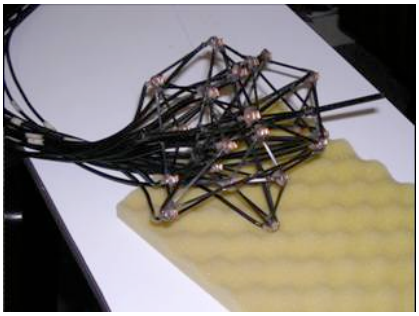


Figura 1.35 – Esempio di microfono HOA con disposizione delle capsule casuale

Nelle seguenti immagini si possono vedere due prototipi di microfoni del 4° ordine, formati entrambi da 32 capsule omnidirezionali, poste sulla combinazione di un dodecaedro ed un icosaedro.

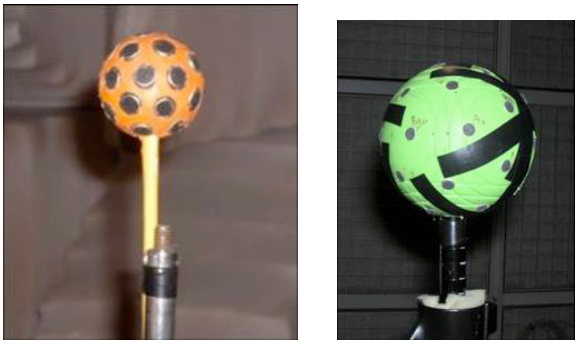


Figura 1.36 – Esempi di microfoni HOA con disposizione delle capsule regolare (Moreau, 2006)

1.2.2.2.5 Basse frequenze

Se si analizza in via del tutto teorica lo spettro del suono registrato con un microfono di ordine via via crescente, ci si accorge di un *boost* consistente alle basse frequenze che porta con se molta rumorosità.

In teoria, se sintetizzando un microfono del 1° ordine con una coppia di microfoni omnidirezionali (ordine 0) spaziati di d, si usa la relazione di Eulero:

$$\frac{dv}{dt} = -\frac{1}{\rho} * grad(p) \qquad (1.60)$$

Il gradiente di p viene approssimato da (pi-p2)/d (d = distanza fra i microfoni) ottenendo:

$$v = \int -\frac{1}{\rho * d}(p_1 - p_2)dt \qquad (1.61)$$

dove l'integrale è ovviamente un integrale nel tempo, e quindi corrisponde ad un filtraggio nel dominio della frequenza con pendenza costante pari a 6dB/ottava.

Non si tratta quindi un semplice *boost* alle basse frequenze, ma una vera e propria integrazione, estesa su tutto lo spettro. Questo filtraggio conferisce al segnale una risposta piatta ma aumenta il guadagno sul rumore a bassa frequenza.

Calcolando le armoniche sferiche di secondo ordine, occorre un integrale doppio (sempre partendo da microfoni a pressione, ottenendo una pendenza di 12 dB/ottava. Occorre però specificare che, se per fare un 1° ordine bastano due microfoni, per fare un 2° ordine ne servono 4 e che, raddoppiando il numero di microfoni, il rapporto segnale/rumore migliora di 3 dB. Quindi, a fronte di una curva più ripida, la base da cui si parte è un miglior rapporto segnale-rumore. Se poi non si utilizzano microfoni omnidirezionali, ma microfoni già direttivi, ad esempio cardioidi, la necessità di un filtraggio si riduce. Sempre per un 1° ordine, partendo da due cardioidi basta un filtraggio con pendenza di 3dB/ottava anziché 6 dB/ottava.

Nel caso del microfono implementato da France Telecom, pur essendo le capsule quasi perfettamente omnidirezionali l'effetto direttivo è dato dall'effetto di

schermatura della sfera rigida, che comunque riduce la necessità di filtraggio.
Per un array microfonico ben congegnato, si raggiungono solitamente prima i limiti di distorsione spaziale dati dalla spaziatura finita dei microfoni che non i limiti di rapporto S/N. Questo ovviamente se si utilizzano microfoni a basso rumore, ma se usando capsule ad elettrete di basso costo e qualità, il rumore non è assolutamente trascurabile. Ma anche con microfoni di buona qualità occorre trascurare ordini armonici elevati a frequenze molto basse e molto alte, causa l'inaccettabile distorsione spaziale delle *figure tridimensionali di direttività* relativi, ben prima che il rapporto S/N diventi critico.

1.2.3 Stereo dipolo

Questa tecnica di riproduzione audio mediante due canali, nata nel 1960 [16] [17], ebbe grande diffusione solo negli ani '70 per la ricerca sull'acustica degli Auditorium ad opera di Schoroeder et al. [18]. Si basa sulla *cross talk cancellation*, metodo alternativo alle cuffie per riprodurre, tramite diffusori acustici, registrazioni binaurali: questa tipologia di registrazione è effettuata utilizzando una coppia di microfoni nel padiglione auricolare di un ascoltatore o di un manichino. In questo modo si registra la pressione sonora, frutto di filtraggi operati dall'ombra acustica e dalla assorbimento del capo, presente all'interno dell'orecchio. Se si riproduce poi tale registrazione mediante un sistema in grado di consegnare ad ogni orecchio il segnale corrispondente, l'impressione di trovarsi presenti all'evento sonoro registrato è pressoché reale. Il sistema inizialmente usato era quello delle cuffie: il segnale del canale sinistro arriva all'orecchio sinistro senza interferire con quello che deve arrivare all'orecchio destro. La tecnica si rivelò efficace ma l'occlusione dell'orecchio rendeva, da un punto di vista psicoacustico, poco confortevole la ricostruzione sonora. Ecco quindi che si sperimentò lo *stereodipolo*, cercando una via per riprodurre registrazioni binaurali con due diffusori acustici al posto di cuffie o auricolari.
Ad una specifica posizione della testa dell'ascoltatore, il *cross-talk* che porta parte del segnale dell'altoparlante dietro all'orecchio sinistro e viceversa, viene cancellato attraverso segnali imposti dall'altoparlante complementare. Questo sistema ha un

limite alle basse frequenze a causa della piccolissima differenza di I.L.D (Paragrafo 1.1.1) e di fase. Viceversa, alle altissime frequenze, la corta lunghezza d'onda rende critica la posizione della testa dell'ascoltatore causando quindi inefficacia nella cancellazione dei cammini incrociati. Per applicare questa tecnica di cancellazione occorre inoltre che l'ambiente di ascolto sia acusticamente trattato: gli effetti delle riflessioni non risultano controllabili e quindi non risultano cancellabili oltre un certo ordine.

Lo stereo dipolo prevede di applicare la cancellazione di cross-talk a due altoparlanti posti molto vicini l'uno all'altro al fine di creare un dipolo acustico vicino ad un mono-polo, in cui gli effetti della cancellazione siano maggiormente controllabili.

Kirkeby et al [19] scoprirono che questa configurazione (con 10° di apertura rispetto all'asse dell'ascoltatore) minimizza gli effetti di *ringing* nei filtri di cross-talk e accresce l'area in cui la cancellazione ha effetto, permettendo maggiori movimenti del capo. Il costo di localizzazione di altoparlanti così vicini è che le basse frequenze richiedono un grosso guadagno in quanto si ottiene in tale zona una notevole cancellazione con conseguente inefficienza della cross-talk. La soluzione è quella di circoscrivere il filtraggio solo in una certa banda di frequenze fissando una *frequenza di cut-off* sotto la quale la cancellazione viene abbandonata e le frequenze vengono riprodotte senza ulteriori manipolazioni. Altro possibile accorgimento è quello di separare molto di più i woofer dei due diffusori rispetto ai tweeter.

Uno dei principali vantaggi dello stereo dipolo rispetto alle cuffie è la sua capacità di generare un'immagine sonora localizzata frontalmente senza perdere informazioni sui contributi laterali (contrariamente allo stereo normale). Questo permette di ottenere una sensazione molto più naturale durante l'ascolto rispetto a quella delle cuffie in cui il suono viene percepito come se stesse provenendo dal centro della testa, o rispetto al classico stereo dove il suono risulta molto frontale. Altro fondamentale vantaggio è la tolleranza del sistema a piccoli movimenti del capo, essendo il campo sonoro non solidale con la testa. Questo permette di avere un beneficio dai piccoli movimenti della testa nella localizzazione delle sorgenti, ottenendo durante l'ascolto una naturale sensazione di spazialità.

Nello specifico la tecnica impiega una matrice 2x2 di 4 filtri, calcolati in modo tale che il sistema cancelli il contributo dell'altoparlante sinistro sul destro e viceversa.

Questa matrice H è ottenuta invertendo la matrice C originale delle funzioni di trasferimento dai diffusori alle orecchie, misurate in via preliminare.

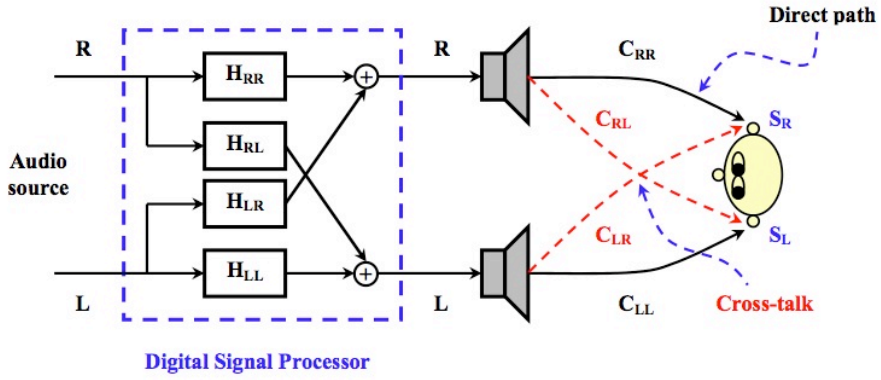


Figura 1.37 – Schema della tecnica dello stereo dipolo

1.2.3.1 Calcolo dei filtri inversi

Ogni sistema di registrazione/riproduzione richiede sempre l'inserimento di un set di filtri numerici, tipicamente filtri FIR preferibili a filtri IIR e WFIR per la semplicità di sintesi. Le tecniche di elaborazione di filtri inversi FIR sono numerose ma le più utilizzate risultano essere:

- filtri a fase minima
- filtri di Kirkeby
- filtri di N.Allen
- filtri di Mourjpoulos

Ciascuna di queste tecniche di filtraggio presenta vantaggi e svantaggi che esulano da questa trattazione. In questa sede la tecnica di inversione maggiormente utilizzata risulta quella di Kirkeby in quanto fornisce i migliori risultati.

1.2.3.1.1 Inversione di Kirkeby: caso singolo ingresso e singola uscita

La prima fase nell'elaborazione di un filtro inverso prevede la misura della funzione di trasferimento del sistema ottenuta utilizzando come sorgente l'impianto di riproduzione e come trasduttore un sistema microfonico. In un sistema a cuffie questo si ottiene ponendole sopra una testa acustica (dummy head) e misurando separatamente la risposta da canale destro a canale destro e viceversa.

Figura 1.38 – Dummy head con cuffie per la misura delle loro funzioni di trasferimento

Solo in questo modo nella misura non è presente alcun effetto di cross-talk acustico. Il risultato della misura è una singola risposta all'impulso stereofonica h. L'obbiettivo è trovare una seconda risposta all'impulso f che filtrata, attraverso l'operazione di convoluzione, con la risposta h misurata restituisca una Delta di Dirac δ.

$$f \otimes h = \delta \Rightarrow FFT(f) \cdot FFT(h) = FFT(\delta) \quad (1.62)$$

Dal momento che nel dominio della frequenza la convoluzione diventa una moltiplicazione che può essere condotta separatamente per ogni riga spettrale, il filtro f cercato si ottiene calcolando il reciproco della funzione spettrale complessa di h misurata (2.2).

$$FFT(f) = \frac{1}{FFT(h)} \quad (1.63)$$

Sfortunatamente questo semplice approccio non risulta ottimale. Infatti tipicamente h non risulta essere a fase minima e pertanto non ammette inversione diretta. Solo un'inversione approssimata può restituire un filtro inverso stabile, causale e a lunghezza finita. La tecnica elaborata da Kirkeby prevede di calcolare il reciproco nel dominio della frequenza aggiungendo un parametro di regolarizzazione al denominatore, funzione esso stesso della frequenza. Tale parametro ε è calcolato

attraverso un approccio a tentativi e definito da un compromesso tra la lunghezza dle filtro inverso che si vuole ottenere e l'accurata inversione di picchi spettrali e buchi. Tipicamente è difficile ottenere un valore ottimo per tutto il range di frequenze ed il risultato non è mai ottimale. Per tale motivo viene utilizzato un ε dipendente dalla frequenza:

$$FFT(f) = \frac{FFT'(h)}{FFT'(h) \cdot FFT(h) + \varepsilon} \qquad (1.64)$$

1.2.3.1.2 Inversione di Kirkeby: caso cross-talk stereo

Un filtro inverso f rappresentato da una matrice 2x2 dev'essere convoluto con la registrazione originale della funzione di trasferimento 2x2 la quale include anche i termini di cross-talk, al fine di ottenere alle orecchie dell'ascoltatore un segnale acustico identico al segnale originale registrato e quindi privo dei filtri di cross-talk.

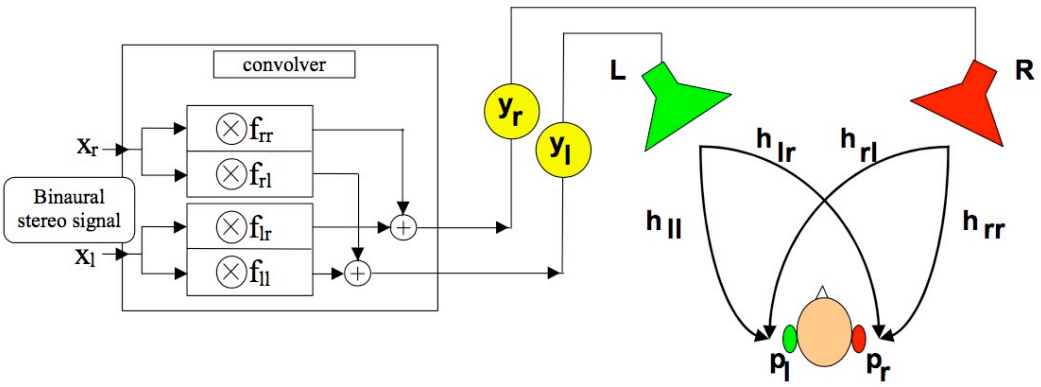


Figura 1.39 - Schema di cancellazione del cross-talk

Si prenda come riferimento la Figura 1.39. Imponendo che il segnale pl sia uguale a xl e che il segnale pr sia uguale a xr si ottiene:

$$\begin{cases} f_{ll} = (h_{rr}) \otimes InvDen \\ f_{lr} = (-h_{lr}) \otimes InvDen \\ f_{rl} = (-h_{rl}) \otimes InvDen \\ f_{rr} = (h_{ll}) \otimes InvDen \\ InvDen = InvFilter(h_{ll} \otimes h_{rr} - h_{lr} \otimes h_{rl}) \end{cases} \qquad (1.65)$$

Il problema, ancora una volta, risulta quello del calcolo del filtro inverso *InvFilter* in quanto il suo argomento risulta mixed-phase. Nel passato sono state effettuate prove di inversione con i filtri di Neely&Allen [20] e di Mourjopoulos [21] ma il metodo di Kirkeby-Nelson è risultato molto più robusto grazie al parametro di regolarizzazione, come prima dipendente dalla frequenza. Il denominatore della funzione reciproca è calcolato direttamente nel dominio della frequenza dove la convoluzione risulta essere una semplice moltiplicazione:

$$C(\omega) = FFT(h_{ll}) \cdot FFT(h_{rr}) - FFT(h_{lr}) \cdot FFT(h_{rl}) \qquad (1.66)$$

Quindi l'inverso complesso è preso aggiungendo il parametro di regolarizzazione:

$$InvDen(\omega) = \frac{Conj[C(\omega)]}{Conj[C(\omega)] \cdot C(\omega) + \varepsilon(\omega)} \qquad (1.67)$$

Si nota inoltre, dal sistema di equazioni (1.65) come nel momento in cui le 4 risposte fra altoparlante e punto d'ascolto h_{ll}, h_{rr}, h_{lr}, h_{rl} diventano praticamente identiche, si debba calcolare l'inverso di un numero molto prossimo a zero, ovvero tendente all'infinito. Questa condizione si ottiene in bassa frequenza quando la lunghezza d'onda risulta molto superiore alle dimensioni della testa e sono presenti piccolissime differenze di fase tra una risposta e l'altra. Da un punto di vista pratico questo si traduce nella richiesta di grossa potenza all'amplificatore (tendente all'infinito), nell'aumento della distorsione e quindi in una cancellazione totalmente inefficace (basso livello sonoro a bassa frequenza. I filtri di cancellazione del *cross-talk* tendono infatti a cancellare anche il segnale diretto.

1.2.3.2 Doppio stereo dipolo

Il doppio stereo dipolo, rispetto al singolo stereo dipolo precedentemente descritto, aggiunge una coppia di diffusori dietro all'ascoltatore (Figura 1.40). Il processing del segnale è formato da 2 matrici (*H-front* e *H-rear*) ottenute invertendo in modo indipendente le due matrici dirette *C-front* e *C-rear*.

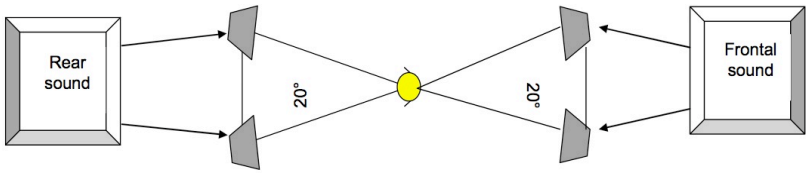


Figura 1.40 – Configurazione "doppio stereo dipolo"

In questo modo la pressione all'orecchio indotta dai due stereo dipoli (front e rear), in condizioni ideali (la testa dell'ascoltatore posizionata nello stesso punto della testa sui cui sono state eseguite le misure, l'ascoltatore perfettamente fermo, strumentazione di riproduzione ideale), dovrebbe essere esattamente la stessa; nelle reali condizioni lo stereo dipolo doppio conferisce al suono che arriva dal retro gli stesi vantaggi, precedentemente descritti, che il singolo stereo dipolo dà per il suono frontale. Ciò significa che sarà più realistico in situazioni in cui il suono posteriore è particolarmente importante, come ad esempio in un teatro (forti riflessioni posteriori o applausi).

2 Realizzazione di una sala d'ascolto

La sala d'ascolto utilizzata in questi tre anni di dottorato è situata all'interno della *Casa della Musica* (Parma), istituzione cittadina con cui il *Dipartimento di Ingegneria Industriale* (Università di Parma) dal 2004 ha firmato una convenzione. Scopo di questo accordo è quello di trovare un punto d'incontro tra la scienza e la musica, due mondi troppo spesso divisi, attraverso l'unione di competenze in campo acustico ed elettroacustico con competenze musicali.

Due sono gli spazi a nostra disposizione: uno è stato adibito ad ufficio ed uno a sala d'ascolto in cui sperimentare tecnologie di audio surround e in cui effettuare test.

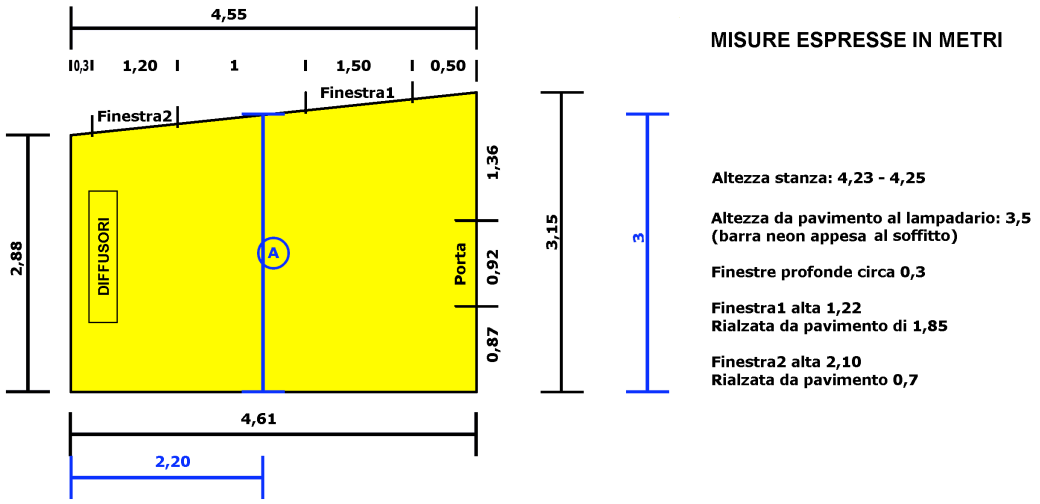


Figura 2.1 – Struttura della sala d'ascolto

Le dimensioni sono decisamente non ottimali da un punto di vista acustico (Figura 2.1). Due delle tre, infatti, risultano praticamente coincidenti, provocando la sovrapposizione delle frequenze di risonanza dovute alle coppie di pareti frontale-retrostante ed soffitto-pavimento. A causa delle proprietà assorbenti e riflettenti delle pareti e del soffitto, la risposta della sala non è piatta ed il sistema di riproduzione in essa introdotto ha un comportamento distorto nel tempo e nella frequenza dalle riflessioni e risonanze della cavità. L'articolazione del suono originale risulta quindi ridotta e la sua dinamica compromessa per cui l'utilizzo di altoparlanti di elevata qualità diventa inutile in un ambiente con una pessima acustica.

2.1 Primo Allestimento

Il primo trattamento acustico nell'ambiente d'ascolto venne effettuato nella primavera del 2004 [31] in occasione di una sessione di test in cui i soggetti eseguivano ascolti in cuffia. Si scelse per questo di creare una sorta di cabina acustica in cui l'ascoltatore fosse isolato dal resto della stanza (Figura 2.2).

Figura 2.2 - Primo allestimento nella sala d'ascolto

Per questo allestimento si utilizzarono pannelli di metallo coperti da materiale fonoassorbente (poliuretano espanso), lo stesso materiale che poi è stato posto a coprire porzioni di parete.

Per un semplice ascolto in cuffia si è rivelato un buon trattamento, confortevole, isolante da stimoli sonori e uditivi.

2.2 Secondo allestimento

Come già espresso in precedenza, la corrispondenza fra due delle tre dimensioni provoca una coincidenza delle frequenze di risonanza accentuandone gli effetti.
A tal fine nella sala sono stati applicati a parete una serie di pannelli di lana di vetro e poliuretano di circa 2m di altezza (Figura 2.3). Tali elementi sono in grado di assorbire energia nelle zone medio-alte dello spettro. Per quanto riguarda l'assorbimento in bassa frequenza, il suo controllo risulta molto più difficoltoso a causa delle dimensioni regolari della sala. Sono quindi stati applicati alcuni box auto-costruiti, alcuni diffusori e pannelli vibranti posti a distanza dalla parete con il fine di creare delle cavità risonanti; tali elementi risonanti sono stati posizionati in corrispondenza dei punti di massima ampiezza delle risonanze, ovvero in corrispondenza delle pareti.
Con questa tipologia di intervento il tempo di riverbero ottenuto all'interno della saletta è di circa 0.2 secondi nelle medio-alte frequenze ma cresce considerevolmente in bassa frequenza a causa degli appena citati effetti di risonanza.

Figura 2.3 – Pannelli di poliuretano espanso alle pareti

2.3 Terzo allestimento

Il trattamento acustico in precedenza mostrato non poteva essere definitivo. C'era, infatti, bisogno di un trattamento permanente, che risolvesse alcuni problemi (risonanze in bassa frequenza) ancora presenti.
Obiettivo di un trattamento acustico di questo è quello di ottenere un tempo di riverbero e una risposta in frequenza ottimali per l'ascolto. Le soluzioni attuabili

erano l'eliminazione dei parallelismi fra le pareti, l'adozione di forme geometriche con dimensioni non identiche fra loro e l'adozione di assorbenti attivi (trappole acustiche) in grado di limitare l'effetto delle risonanze fra cui, risonatori di Helmoltz, pannelli vibranti, tube traps.

In teoria ogni riproduzione sonora dovrebbe avere luogo in una camera anecoica. In pratica questo non restituisce i risultati migliori in termini di realismo dell'ascolto. Infatti la sensazione che si ottiene all'interno di un ambiente eccessivamente assorbente è di poco realismo sonoro provocando sull'ascoltatore sensazioni sgradevoli che influiscono sul giudizio. Quindi è necessario mantenere un piccolo riverbero all'interno dell'ambiente, senza che questo influisca sul timbro e sulla dinamica del suono riprodotto. A tal fine l'andamento in frequenza del tempo di riverbero deve essere il più costante possibile.

Le sensazioni psicoacustiche legate al tempo di riverbero risultano differenti in bassa ed in alta frequenza. In alta frequenza la finestra d'integrazione dell'orecchio umano e corta e la variabilità spaziale della risposta molto alta a causa delle corte lunghezze d'onda; la risposta in frequenza percepita quindi dipende solamente dalle riflessioni precoci e può cambiare con piccoli movimenti dell'ascoltatore. L'ultima parte della risposta all'impulso viene quindi percepita come coda sonora riverberante. Per le basse frequenze invece il tempo d'integrazione dell'orecchio umano risulta molto più lungo, quindi le prime riflessioni contribuiscono unicamente ad accrescere la sensazione di livello associata al fronte diretto, modificandone anche la percezione della dinamica e del timbro.

2.3.1 Scelta del punto d'ascolto

La stanza, considerate le applicazioni e comunque le ridotte dimensioni complessive, può ospitare un unico ascoltatore, accomodato su di una sedia posta sopra una pedana lignea rialzata (ricoperta di moquette). Il punto d'ascolto (in Figura 2.4 indicato con la lettera A) è stato scelto garantendo una superficie sferica utile di raggio 150cm attorno all'ascoltatore, sulla quale sono installati gli altoparlanti dei vari sistemi di riproduzione.

La posizione prestabilita tiene conto della distribuzione del campo sonoro nelle basse frequenze ed è collocata in una posizione intermedia, dove nodi e ventri delle onde stazionarie proprie della sala sono tenuti il più possibile lontani. Figura 2.4 spiega graficamente quanto appena esposto.

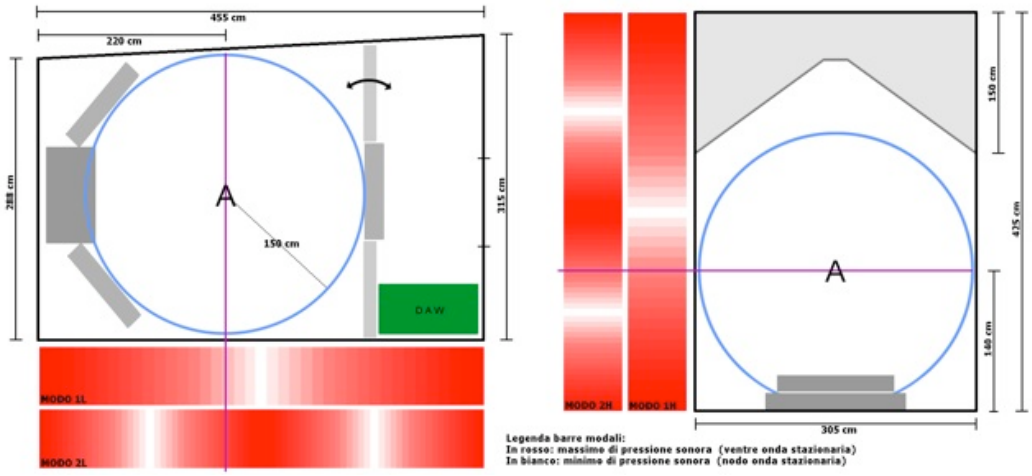


Figura 2.4 - Pianta della stanza, evidenziando punto d'ascolto scelto (A) in funzione dei primi modi assiali nelle coordinate L, W.

2.3.2 Analisi modale

L'acustica passiva rappresenta il primo passo per ottenere un ambiente di riproduzione affidabile ed il più possibile neutro. Si procede attraverso un'analisi teorica modale, supponendo in prima approssimazione l'ambiente parallelepipedo. Lo studio modale evidenzia le principali problematiche della stanza in oggetto: come conseguenza delle piccole dimensioni, i modi dei primi ordini cadono nelle frequenze udibili, questi quindi saranno eccitati dal campo sonoro irradiato dalle sorgenti e sono in grado di conservare per un lungo tempo l'energia. Si generano quindi marcati fenomeni di riverberazione per alcune frequenze ed in generale una conformazione spaziale del campo sonoro tutt'altro che omogenea [22][23][24].

Dall'analisi modale delle dimensioni della stanza (Tabella 2.1) si evidenzia la stretta vicinanza tra due dei modi a 40 Hz.

Ordine nelle 3 coordinate			Frequenza naturale	Ordine nelle 3 coordinate			Frequenza naturale	Ordine nelle 3 coordinate			Frequenza naturale
Nl	***Nh***	***Nw***	***f_n (Hz)***	***Nl***	***Nh***	***Nw***	***f_n (Hz)***	***Nl***	***Nh***	***Nw***	***f_n (Hz)***
1	0	0	**37.80**	2	0	0	**75.6**	3	0	0	**113.41**
0	1	0	**40.47**	0	2	0	**80.94**	0	3	0	**121.41**
0	0	1	**57.33**	0	0	2	**114.67**	0	0	3	**172.00**

Tabella 2.1 – Prime risonanze assiali di un ambiente parallelepipedo di 455x300x425 cm. N_l N_h N_w *esprimono l'ordine del modo nelle tre coordinate lunghezza, altezza e profondità;* f_n *la frequenza di risonanza naturale*

2.3.3 Controsoffitto

Il primo trattamento acustico qui descritto consiste in una controsoffittatura di pannelli dallo spessore di 6cm in *Fiberform*, materiale poroso formato da fibre di poliestere. Il materiale sfruttato, se collocato direttamente sulle pareti, conta di un coefficiente d'assorbimento α pressoché costante a partire dai 500Hz, ed estende l'azione fonoassorbente sulle basse frequenze se installato a distanza dalle superfici rigide perimetrali. Nella particolare configurazione mostrata in Figura 2.5 è stato possibile:

I. creare una cavità (il volume d'aria interno, dai pannelli alla superficie rigida esterna di muratura) dalle massime dimensioni permesse nella stanza
II. un trattamento fonoassorbente a larga banda considerata la distanza variabile dalla superficie rigida esterna, capace di coprire anche le basse frequenze
III. si è cercato di modificare le geometrie dell'ambiente, principali responsabili nella creazione delle risonanze naturali [22].

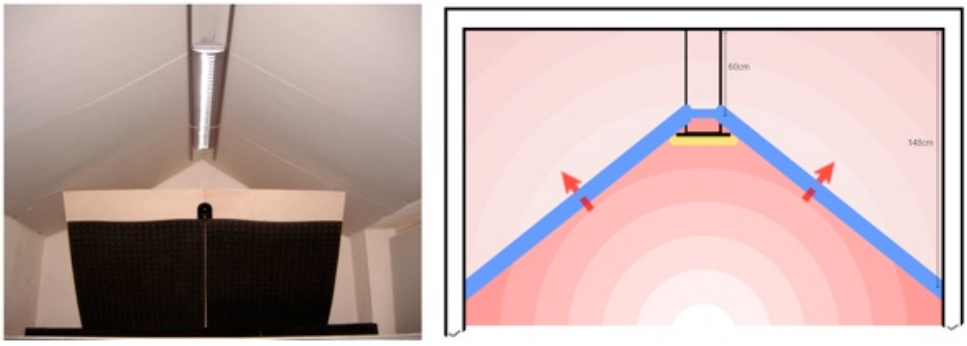


Figura 2.5 – Controsoffitto/Trappola per bassi in Fiberform

2.3.4 Risonatori di Helmholtz

Per contrastare il primo modo assiale longitudinale si è optato per una schiera di *risonatori di Helmholtz*. L'elemento base è un volume d'aria delimitato da pareti rigide e collegato all'esterno da un'apertura detta "collo". Il suono incidente fa vibrare l'aria contenuta nel collo del risonatore, che si comporta come una massa vibrante collegata a una molla costituita dall'aria contenuta nella cavità. Tale risonatore è in grado di dissipare energia acustica in calore, per effetto dell'attrito viscoso che si verifica a causa delle oscillazioni dell'aria contenuta nel collo, in corrispondenza della frequenza fondamentale di risonanza del sistema massa-molla appena descritto [23][24].

Le cavità sono state ricavate da contenitori di plastica rigida da 18.9 litri; per adattare meglio la frequenza d'azione si è aggiunto un sottile strato di moquette nella parte interna del collo, di lunghezza e spessore variabile fino ad ottenere l'accoppiamento con il modo longitudinale, raggiungendo la frequenza centrale di circa 38Hz ed una adeguata dispersione (si confronti Tabella 2.1).

La parte sinistra di Figura 2.6 mostra l'analisi spettrale estratta durante le fasi di accordatura [22], la curva piena in blu rappresenta la configurazione scelta, dove inoltre è stato aggiunto del materiale poroso all'interno della cavità, al fine di attenuare i massimi secondari caratteristici dei contenitori, dovuti da onde stazionarie interne alla cavità. L'aggiunta di materiale fonoassorbente nel volume interno ha permesso di ottenere un trattamento acustico pressoché trasparente alle medie frequenze, allargando inoltre ulteriormente la campana di risonanza.

Una schiera dei risonatori appena descritti è stata installata nel fondo della sala d'ascolto, posizione ottimale, in corrispondenza del massimo di pressione sonora alla frequenza di 37.8Hz, conseguente al modo longitudinale che si vuole attenuare (Figura 2.6 al centro).

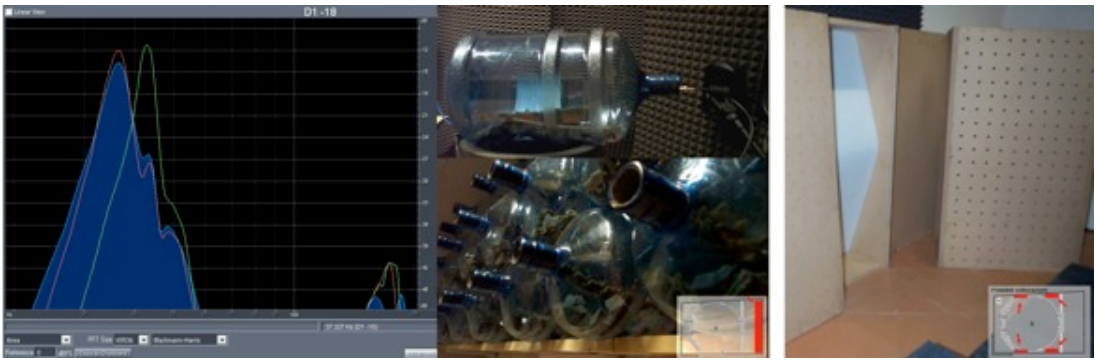


Figura 2.6 - A Sinistra: paragone FFT risuonatori nelle differenti conFigurazioni durante accordatura; Al centro fasi di accordatura risuonatori (in alto) ed assetto finale (sotto); A destra: pannelli forati, evidenziata la posizione scelta.

La parte destra di Figura 2.6 mostra i pannelli forati, anche questo trattamento si basa su di una cavità risonante ed è l'estensione del risonatore di Helmholtz. Ogni foro corrisponde ad un collo avente la sua area superficiale e una lunghezza determinata dallo spessore del pannello, ed il volume d'aria retrostante è in comune per tutti i risonatori. In presenza di più fori la mutua iterazione tra essi determina la comparsa di fenomeni dissipativi a frequenze diverse dalla frequenza di risonanza, ottenendo uno spettro d'assorbimento più ampio rispetto a quello che si ottiene con risonatori singoli [23][24]. Questi fenomeni, unitamente all'effetto dei materiali porosi inseriti all'interno della cavità (lana di roccia e fiberform), hanno permesso di realizzare pannelli fonoassorbenti che assorbono una banda estesa approssimativamente dai 100 ai 300Hz. Una stima della frequenza fondamentale del trattamento descritto è stata ottenuta dalla formula:

$$f_0 = \frac{c_0}{2\pi}\sqrt{\frac{P}{d\cdot(l + r\pi/2)}} \qquad [\text{Hz}]$$

dove: c_o è la velocità di propagazione del suono nel mezzo [m/s]; P è la percentuale di foratura del pannello; d è lo spessore interno della cavità a forma di parallelepipedo; r il raggio del foro (per tener conto del fattore correttivo di bocca); l lo spessore del pannello, ovvero la lunghezza del collo.

Spessore cavità (mm, interno)	Volume Cavità (m cubi)	Spessore pannello (lunghezza collo)	Diametro fori (mm)	Passo fori (mm)	Percenuale foratura	Frequenza di risonanza
220	0.106	18	10	50	2.50%	**114.8**
140	0.067	18	5	35	1.33%	**113.9**

Tabella 2.2 – Specifiche di dimensionamento pannelli forati

2.4 Risultati della bonifica acustica

Per ogni fase intermedia si sono acquisite le risposte all'impulso dell'ambiente, nel punto d'ascolto scelto, tramite la tecnica dello sweep seno-logaritmico: si riproduce un tono sinusoidale crescente in frequenza nel tempo, contemporaneamente si registra mediante microfono omnidirezionale da misura (Figura 2.7 in fondo), infine in post-processing viene effettuata una convoluzione con il segnale inverso, ottenendo la risposta impulsiva cercata, da intendere come descrittore acustico dell'ambiente [22][25].

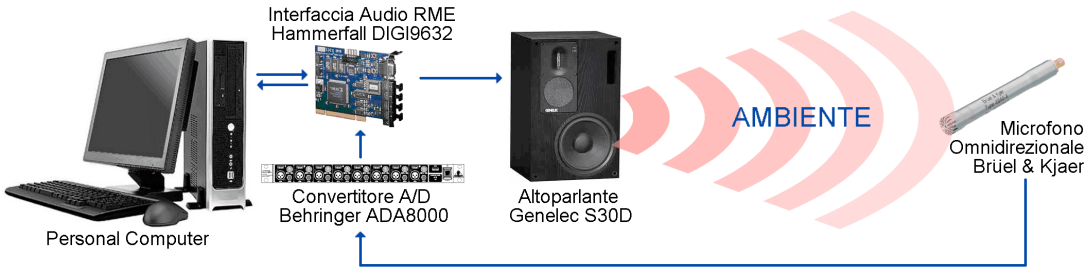


Figura 2.7 – Routing e strumentazione adottata, con relativo schema di acquisizione misure, tramite tecnica dello sweep seno logaritmico.

Analizzando le risposte all'impulso, ottenute mediante software *Adobe Audition* e plug-in *Aurora* [27][28], si ottiene un confronto oggettivo prima/dopo l'azione dei vari trattamenti acustici installati.

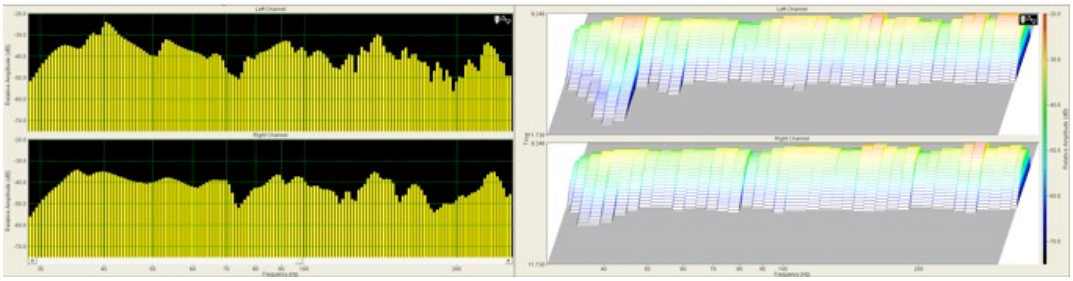


Figura 2.8 – Analisi spettrale, stazionaria (a sinistra) e mediante grafico waterfall (a destra).

La Figura 2.8 mostra un confronto, in termini di spettro frequenziale, fra il precedente allestimento "ante operam", descritto nella parte alta dei grafici, e dopo l'azione dei trattamenti acustici installati, in basso. Nell'insieme la conformazione del campo sonoro risulta molto più omogenea alle basse frequenze, le prime risonanze intorno ai 40Hz ora presentano un massimo attenuato di oltre 15dB, ed un andamento in tutta la regione modale più morbido, come effetto dello smorzamento attuato dai trattamenti acustici (parte sinistra di Figura 2.8). Le lunghe code riververberanti, generate dall'energia accumulata per effetto delle onde stazionarie, si sono ridotte per oltre il 25% (parte destra di Figura 2.8).

Parallelamente si è condotta un'analisi mediante Audio Quality Test (AQT) [26][29], basata su *burst monotonali* a frequenze crescenti ben precise con intervalli stretti a piacere, più adatta per l'analisi di transitori, tipici dei segnali musicali. Essa permette una risoluzione spettrale maggiore alle basse frequenze (più vicina alle capacità del sistema uditivo) e considera anche fenomeni psicoacustici, in particolare il *mascheramento acustico* e l'*effetto Haas*. AQT fornisce due parametri: il *Bilanciamento Tonale*, e l'*Articolazione*; il secondo esprime la dinamica sonora ottenibile all'interno di un ambiente. Figura 2.9 mostra in forma sintetica l'esito del paragone mediante test AQT. Come si vede l'Articolazione è mediamente migliorata, soprattutto nella prima ottava e nelle medio-basse frequenze. E' invece diminuita, seppure marginalmente, l'articolazione nell'intorno degli 80Hz, difatti non sono stati collocati nella sala trattamenti efficaci per contrastare i modi 2L e 2H, onde stazionarie che degradano la dinamica del suono nel punto d'ascolto, prossimo ai nodi di pressione sonora relativi alle risonanze appena menzionate.

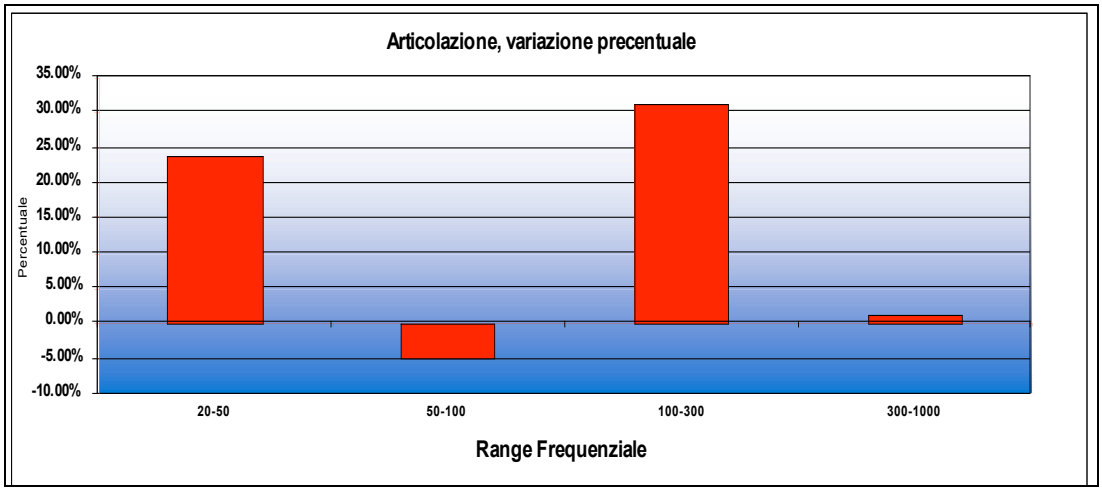


Figura 2.9 - Variazione percentuale del parametro AQT "Articolazione", dopo l'installazione dei trattamenti acustici

Infine, l'immagine sottostante (Figura 2.10) mostra il tempo di riverbero all'interno della stanza in configurazione finale.

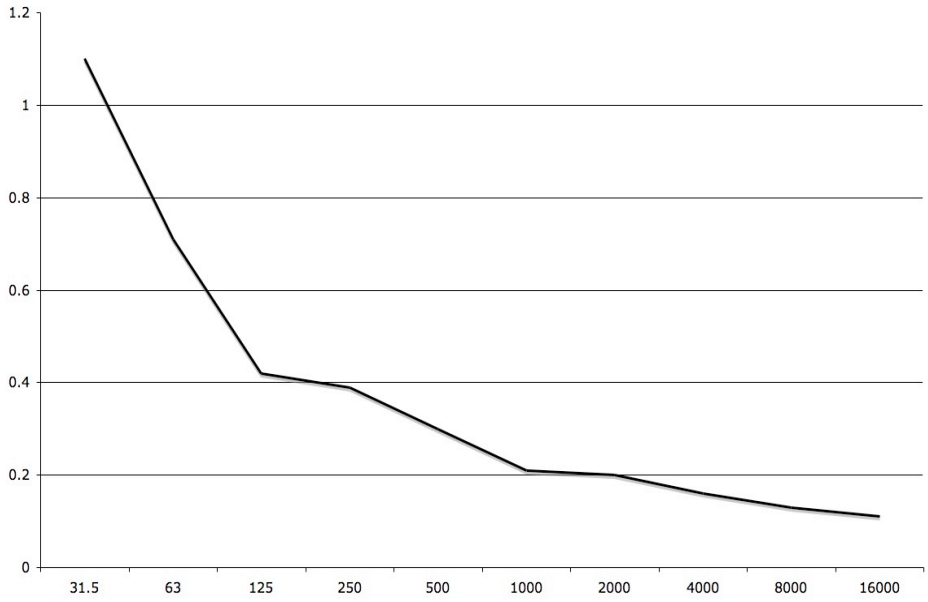


Figura 2.10 – Misura del tempo di riverbero all'interno della saletta d'ascolto dopo l'ultimo trattamento acustico

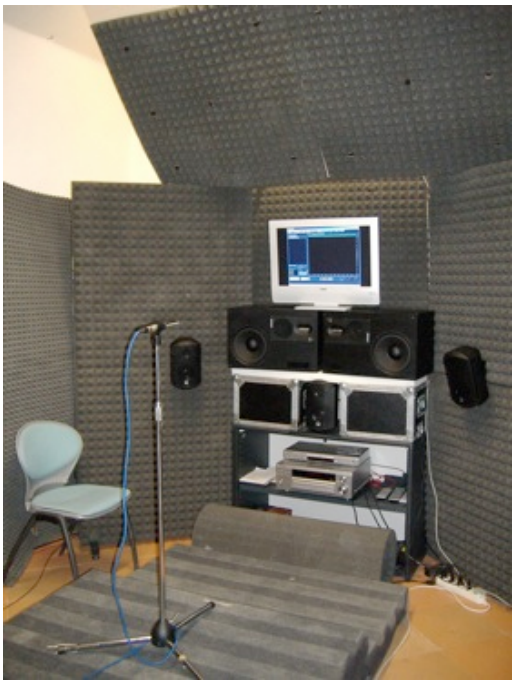


Figura 2.11 – Misura dell'acustica della sala

2.5 L'hardware

Il cuore del sistema è un PC Asus S-Presso (Pentium IV, 2GHz), silenzioso perchè raffreddato a liquido e quindi ottimo per applicazioni in cui è richiesta una dose molto bassa di rumorosità. All'interno vi è stata alloggiata una scheda RME HDSP Madi, che sfrutta un protocollo (MADI) in grado di gestire un ampio numero di canali; in questo caso ne sono stati utilizzati solamente 24. La scheda è connessa ad un'unità esterna, che converte i segnali MADI in segnali ADAT; questi ultimi vengono convertiti in segnali analogici tramite due convertitori (Behringer ADA 8000 e Apogee 16 DA). Da qui i segnali vengono consegnati a tre amplificatori multicananle QSC.

Per il sistema Ambisonic sono stati scelte, come diffusori,16 Turbosound Impact 50; per il Triplo Stereo Dipolo sono stati utilizzate 2 QSC AD82S (stereo dipolo posteriore), due Turbosound Impact 50 (stereo dipolo superiore) e 2 Genelc S30D (stereo dipolo frontale). Tutti gli altoparlanti sono stati posizionati come in Figura 2.12 e Figura 2.13.

La posizione d'ascolto è situata nel centro: l'anello di speaker planare ha un raggio di 1.42 m, l'anello superiore ha una distanza media di 1.60 m, l'anello più basso ha una

distanza media di 1.80 m; lo stereo dipolo frontale dista 1.44 m dall'ascoltatore, quello posteriore e quello superiore 1.30 m.

Per avere un livello uguale per tutti gli speaker, è stata fatta una calibrazione usando rumore rosa e aggiustando i guadagni sugli amplificatori in modo da avere lo stesso livello nel punto d'ascolto.

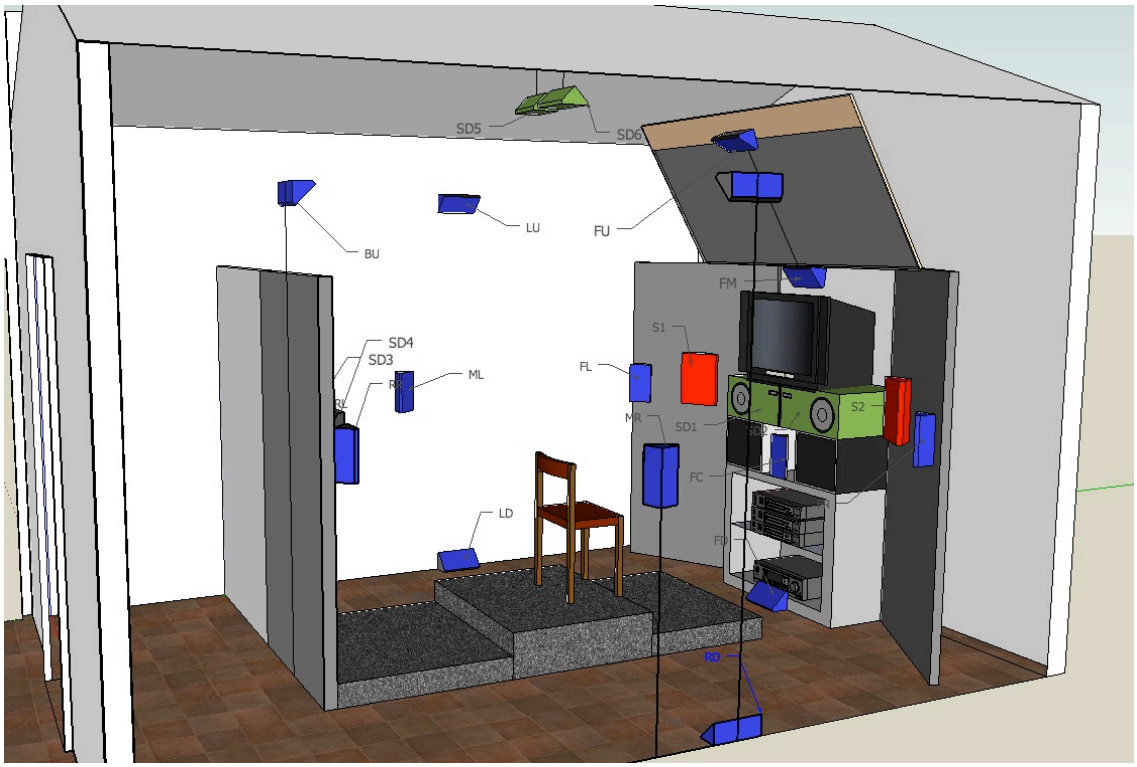


Figura 2.12 – Visione frontale della stanza d'ascolto. In Blu gli speaker Ambisonic, in Verde gli speaker Stereo Dipolo, in Rosso lo stereo normale

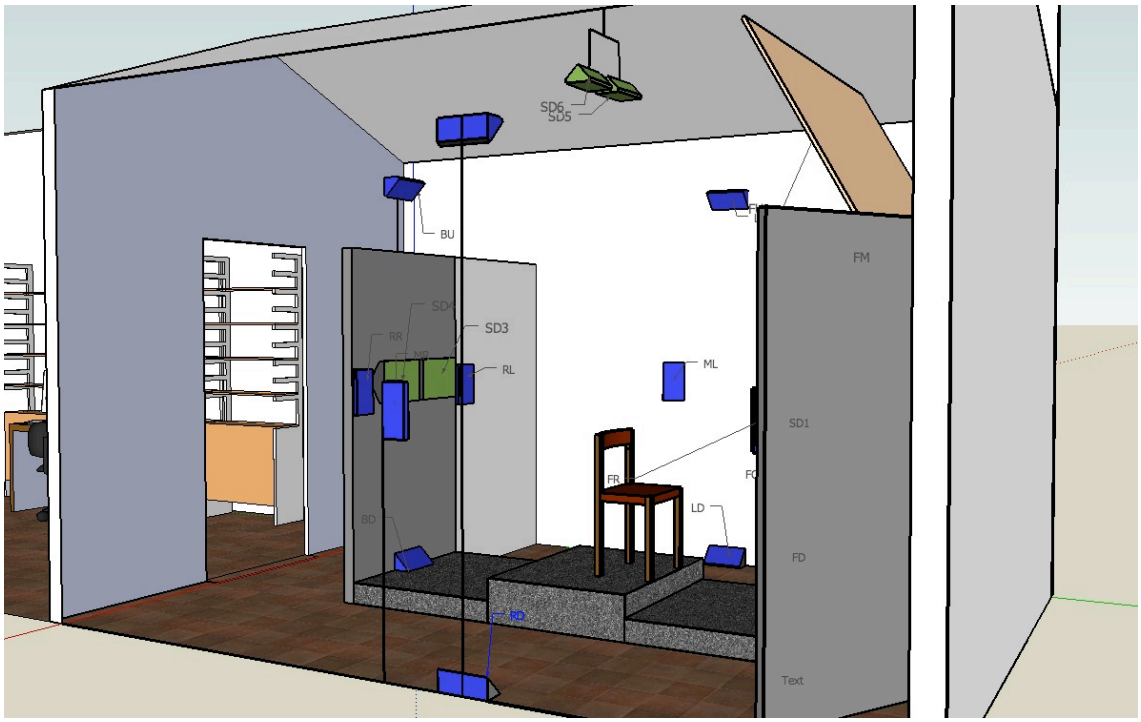


Figura 2.13 – Visione posteriore della stanza d'ascolto. In Blu gli speaker Ambisonic, in Verde gli speaker Stereo Dipolo, in Rosso lo stereo normale

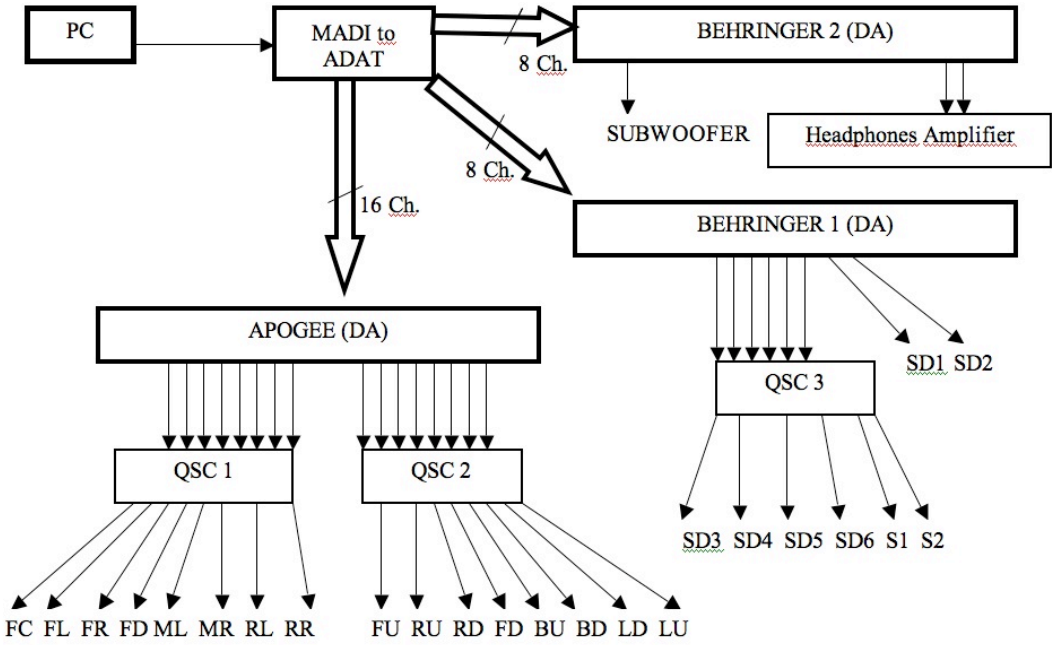


Figura 2.14 – Layout dell'hardware

3 Test d'ascolto

In questo capitolo vengono presentati i numerosi test d'ascolto effettuati per valutare l'efficacia dei sistemi di riproduzione utilizzati, l'acustica delle sale e la capacità uditiva degli ascoltatori.

3.1 Realismo dei sistemi di riproduzione sonora

I test presentati in questo paragrafo sono stati eseguiti nella stanza d'ascolto della Casa della Musica, ancora con il secondo tipo di trattamento acustico (Paragrafo 2.2).

Il routing dei segnali audio è stato progettato ad hoc per questo tipo di test. Poiché una delle attività condotte all'interno di questa sala riguarda il confronto tra sistemi di riproduzione/acquisizione di alcune proprietà del campo acustico, sono state utilizzate risposte all'impulso di spazi destinati alla musica convolute con registrazioni anecoiche. In particolare, risposte all'impulso binaurali e stereofoniche (O.R.T.F.), registrate in cinque auditorium e sale da concerto, sono state scelte per testare la qualità spaziale dei quattro sistemi di riproduzione: sistema stereo normale, cuffie, sistema stereo dipolo, sistema stereo dipolo doppio. Una registrazione musicale anecoica, convoluta con le risposte all'impulso è stata riprodotta attraverso i quattro sistemi. Mediante filtraggi ed un'attenta calibrazione i 4 sistemi sono stati resi il più possibile comparabili fra loro, ripristinando inoltre, all'interno della sala d'ascolto, i livelli sonori originari presente in fase di misura. Attraverso una serie di test soggettivi sono state valutate le dimensioni dell'ambiente, la distanza delle sorgenti e il realismo delle riproduzioni.

3.1.1 Hardware

La disposizione degli altoparlanti è stata tenuta asimmetrica rispetto all'asse della sala in modo da ridurre l'eccitazione delle risonanze proprie di tale ambiente. Gli altoparlanti sono stati tenuti ad una distanza 1.5m dall'ascoltatore. Gli altoparlanti utilizzati sono i seguenti:

- prototipo Audiolink AL105 per lo stereo convenzionale, disposti a ±30° dalla linea mediana di simmetria

- Genelec S30D per lo stereo dipolo frontale. Tali monitor sono stati posizionati sul lato mantenendo internamente i tweeter (22cm), in posizione esterna i mid-range (43cm) e i woofer (83cm). Queste distanze corrispondono ad un'apertura, rispetto alla linea mediana, di 4°, 8° e 16°. Questa disposizione risulta in accordo con quanto indicato dall'Università di Southampton nella teoria dell'OSC. Mantenendo i woofer a maggiore distanza fra di loro si minimizza lo stress richiesto dagli altoparlanti nella realizzazione della cancellazione di cross-talk, diminuendo la distorsione e il carico sugli stessi ma migliorandone in questo modo la qualità audio.
- QSC AD-S82H per lo stereo dipolo posteriore. Le casse sono state posizionate mettendo i driver centrali separati di 45cm (9° rispetto alla linea mediana).
- Sennheiser HD560, cuffie acustiche "aperte".

Nonostante siano stati utilizzati diversi tipi d'altoparlanti, la loro risposta in frequenza è stata resa identica attraverso l'utilizzo di filtri inversi a 4096 tap fra i 100 Hz e i 20 KHz (sotto i 100 Hz il contributo acustico richiesto è limitato e quindi non influente nel corso dei test soggettivi.

Al fine di rendere invisibile all'utente la configurazione di degli altoparlanti, le finestre sono state coperte con pannelli opachi e i test condotti unicamente alla luce del monitor del pc (con un piccolo contributo di luce ambientale). Questo ha permesso agli ascoltatori di non individuare il posizionamento degli altoparlanti, evitando quindi di esserne influenzati.

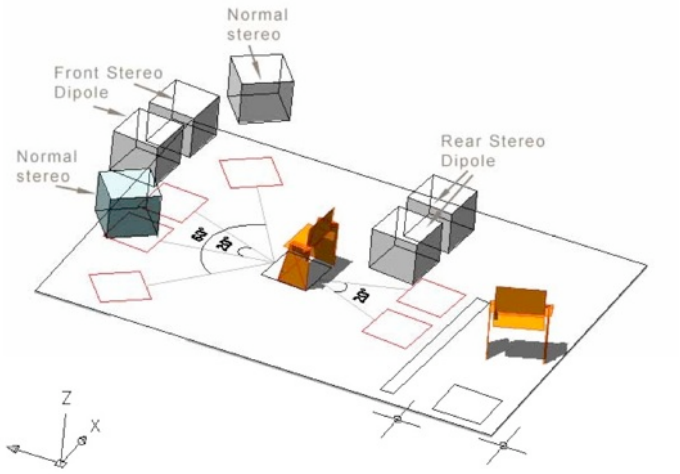


Figura 3.1 – Schema della sala d'ascolto

3.1.2 Tecnologia e layout

Al fine di comparare tutti e quattro i sistemi contemporaneamente, è stata sviluppata una speciale rete di segnali audio. Concettualmente la sala è stata concepita con un'interfaccia utente, un gruppo di elaborazione e la parte di trasduzione del segnale. Un computer portatile, posizionato nel punto di ascolto, è stato connesso ad una scheda audio Edirol F-101 e quattro delle sue uscite analogiche sono state usate per pilotare i sistemi audio precedentemente descritti. Utilizzando un software scritto per l'occasione e descritto in seguito, è stato possibile selezionare in tempo reale uno dei 4 sistemi e alimentarlo con la registrazione appropriata (ORTF per lo stereo normale, binaurale per gli altri sistemi). Tali registrazioni sono state ottenute mediante la convoluzione off-line del segnale anecoico con le risposte all'impulso rispettivamente misurate con la microfonazione ORTF e con la testa binaurale Neumann KU70.

Gli 8 canali in uscita (4 stereo) sono stati connessi con un pc separato destinato all'elaborazione audio, equipaggiato con una scheda *Aardvark Q10*, dotata di 4 canali in ingresso e 4 in uscita. Tutte le operazioni di I/O sono state implementate a 24 bits, 96KHz con processing digitale in floating-point a 32 bit di precisione.

I 4 canali d'ingresso della scheda Aardvaark ricevono i 4 canali stereo d'uscita provenienti dal notebook e dalla scheda Edirol. I 4 canali stereo vengono processati attraverso il software *AudioMulch*, un host VST multicanale. Al suo interno sono state inizializzate due istanze del plug-in *Voxengo* per i filtraggi necessari ai 4 sistemi di riproduzione. I segnali così processati vengono inviati attraverso le uscite della scheda audio ai 4 sistemi.

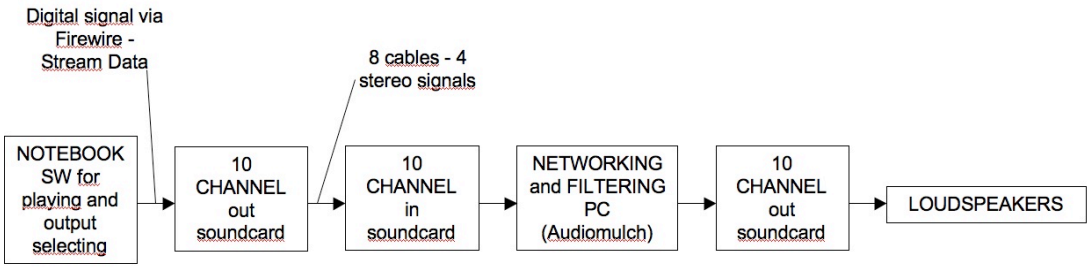


Figura 3.2 – Rete audio allestita presso il laboratorio di Casa della Musica

I filtri implementati all'interno di *Voxengo* sono riportati di seguito:

- Filtri di cancellazione del cross-talk (matrice 2 x 2) per lo stereo dipolo frontale/posteriore e per l'equalizzazione delle casse.
- Filtri semplici stereo per le cuffie e lo stereo normale.

In tutte queste inversioni le risposte all'impulso misurate fra gli altoparlanti del sistema e il punto di ascolto sono state troncate subito dopo l'arrivo del fronte diretto. Questo è dovuto al fatto che l'inversione di riflessioni di ordine eccessivo richiede l'utilizzo di filtri digitali molto lunghi e, a causa della stretta dipendenza della fase di tali riflessioni dalla posizione di ascolto, il filtraggio diventa instabile. Invertendo solamente il suono diretto si hanno quindi diversi vantaggi:

- I filtri inversi risultano corti, con conseguente bassa potenza di calcolo richiesta in fase di processing.
- La latenza del convolutore è più corta e la reazione ai comandi dell'ascoltatore immediata.

3.1.3 Collezione delle tracce audio e registrazione del segnale di test

Al fine di investigare la capacità dei 4 sistemi installati nella sala di riproduzione è stata allestita una sessione di test soggettivi. Le risposte di 5 diversi auditorium sono state selezionate all'interno di una vasta banca dati per coerenza della strumentazione impiegata e presenza di tutte le informazioni necessarie a ricostruire la corretta calibrazione dei sistemi. Le misure di tali risposte sono state effettuate utilizzando un dodecaedro unitamente ad un subwoofer come sorgente (preventivamente equalizzato per ottenere uno spettro di emissione piatto) con la tecnica dello sweep seno-logaritmico. La strumentazione di misura prevede invece l'utilizza di una testa Neumann KU70 su cui sono stati applicati due microfoni Neumann a cardioide AK40 in configurazione O.R.T.F. Ad essi viene aggiunto un microfono Soundfield per le misure in B-format [32].

I cinque *auditoria* riprodotti sono:

- Sale piccola, grande e media presenti nel Parco della Musica (Roma-Italy)
- Auditorium Paganini (Parma-Italy)
- Kirishima's Miyama Conseru (Kirishima-Japan)

Per ciascun auditorium sono state selezionate due diverse posizioni d'ascolto del ricevitore. In tutti i casi la posizione del ricettore era sull'asse di simmetria dell'auditorium e la sorgente ad un metro fuori asse sul palcoscenico in corrispondenza del proscenio.

Tutte e dieci le risposte all'impulso sono state convolute con una traccia anecoica calibrata registrata per l'occasione. Come strumento si è scelta la fisarmonica, soprattutto perchè poco direttiva; è stata registrata tramite microfono da misura ad una distanza di 2.5m. Tommasi Dradi (Figura 3.4), musicista esperto della scena musicale parmigiana, ha eseguito "La ballata di Michè" di Fabrizio De Andrè, un valzer con dei legati è un accompagnamento articolato. I livelli di pressione sonora della sorgente, in bande d'ottava e normalizzate a 1m, sono mostrate in Figura 3.3. Il livello equivalente pesato A (L_{eq}) della fisarmonica, normalizzato ad un metro, è di 80 dB(A); la registrazione era di circa 45 sec.

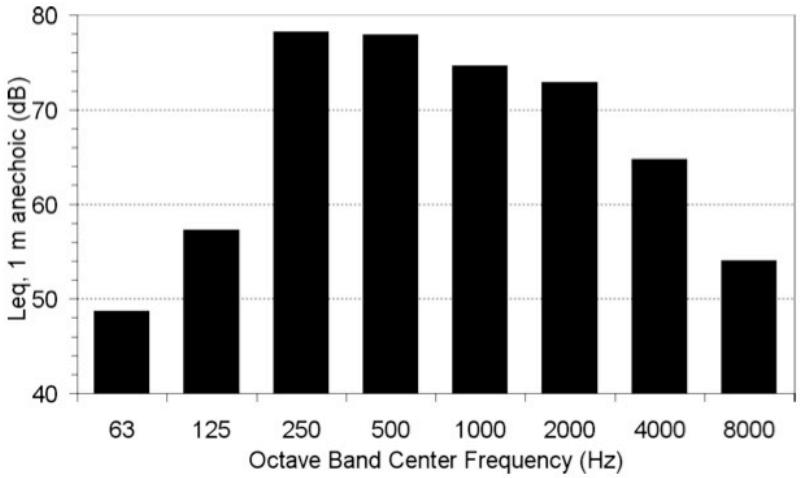


Figura 3.3 – L_{eq} della traccia audio normalizzato ad 1m

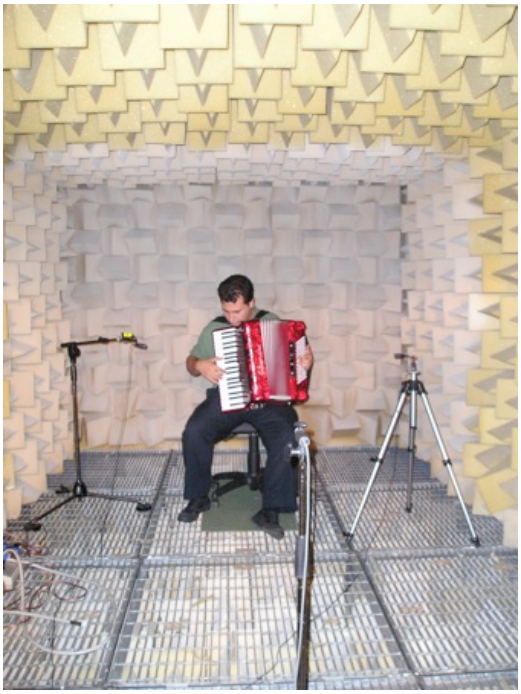


Figura 3.4 – Registrazione de "La ballata di Miché"

3.1.4 Test soggettivi

Nella costruzione del DOA, ovvero la matrice statistica delle tracce audio da presentare ai soggetti, si è notato come, avendo 10 diverse risposte all'impulso (5 auditoria x 2 posizioni), 4 sistemi di riascolto e 3 domande, risulti impossibile presentare la lista completa delle possibili conFigurazioni a ciascun ascoltatore. Pertanto si è deciso di sottoporre a ciascun soggetto cinque diverse configurazioni dell'auditorium con due sistemi d'ascolto diversi (per un totale di 10 configurazioni da ascoltare). La copertura statistica di tutte le configurazioni è stata fatta utilizzando una sequenza bilanciata di conteggio fra i soggetti.

Un software ad hoc è stato scritto per la gestione del test, la raccolta dei dati e la gestione del routing delle tracce audio. Selezionando uno dei bottoni si attiva il play della traccia selezionata, senza tempi di silenzio o latenza fra una traccia selezionata e la successiva. Il soggetto può quindi saltare fra le tracce seguendo il criterio e l'ordine a lui più congeniali. Le tre domande vengono visualizzate ma l'accesso alla successive è concesso solo nel caso in cui si sia risposto completamente alla domanda attiva (un punteggio per ogni configurazione). Anche l'ordine delle domande è casuale e deciso ogni volta dal software. Le tre domande presentate sono:

- Quale è la dimensione dell'ambiente che stai ascoltando?
- Quanto è realistico il suono?
- Quanto lontano senti l'artista in metri?

Al soggetto testato non sono state date informazioni aggiuntive sulla configurazione degli altoparlanti e le condizioni di scarsa luminosità hanno permesso di rendere difficoltosa l'individuazione del loro layout. L'unico sistema ovviamente identificabile è stato quello delle cuffie: il software stesso fornisce istruzioni su quando indossarle e quando toglierle. Trenta soggetti, con esperienza musicale, hanno partecipato all'esperimento.

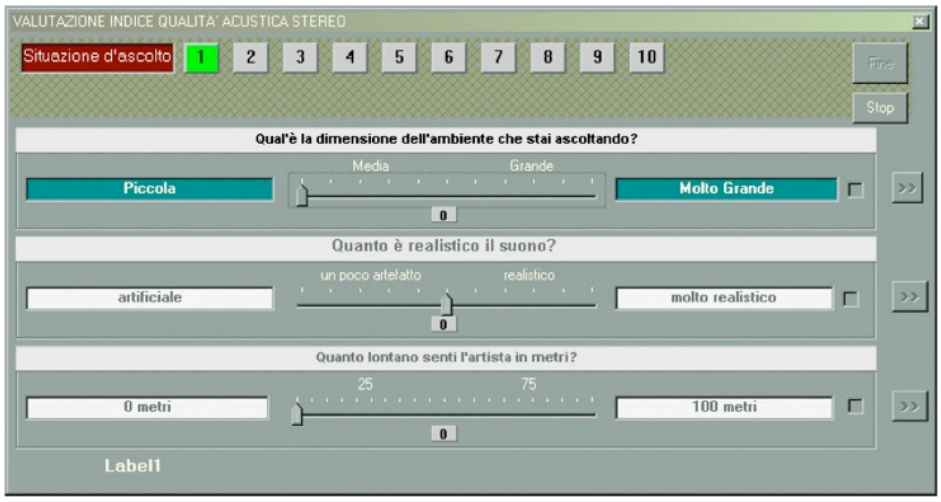


Figura 3.5 – Software sviluppato per la sessione di test

3.1.5 Risultati

Fra tutti gli aspetti legati alla spazialità dei sistemi di riproduzione, tre aspetti sono stati investigati: il realismo della riproduzione, la percezione delle dimensioni dell'ambiente, la percezione della distanza tra sorgente e ascoltatore. I risultati restituiti dall'analisi statistica sono di seguito rappresentati.

3.1.5.1 Stima della distanza

L'analisi ANOVA mostra un poco significativo effetto sui giudizi in dipendenza dal sistema audio ($f=3.86$, $p=0.0099$, $df=3$) e un grandissimo effetto da parte della situazione scelta ($f=11.45$, $p<0.0001$, $df=9$). La capacità dei sistemi di riproduzione

di restituire correttamente la distanza sorgente-ricevitore è stata investigata mediante il calcolo di tre parametri:

- Correlazione tra i risultati del test e le reali distanze
- Valore RMS dell'errore della distanza (m) fra il caso reale e i casi valutati nella sala d'ascolto
- Valore RMS dell'errore del logaritmo della distanza tra i casi suddetti

Mentre la migliore correlazione è risultata essere quella del sistema "doppio stereo dipolo", il valore *rms* dell'errore fra le distanze è risultato inferiore nel caso dello stereo standard; utilizzando una distanza logaritmica il sistema migliore è risultato il singolo stereo dipolo. Va fatto notare come la correlazione non tenga in considerazione il valore assoluto dei voti ma semplicemente ne valuti il comportamento relativo e la bontà di fit lungo la linea di regressione, mentre l'errore RMS è sensibile alle deviazioni in senso assoluto.

In tutti e tre i casi il sistema rivelatosi peggiore nella valutazione della distanza sorgente-ricevitore è il sistema cuffie.

	Correlazione (r^2)	Errore RMS	Errore RMS (log(m))
O.R.T.F	0.39	9.3	0.23
Cuffie	0.34	14.9	0.28
Stereo Dipolo	0.59	9.6	0.19
Doppio Stereo Dipolo	0.63	10.3	0.22

Tabella 3.1 – Correlazione tra i risultati

3.1.5.2 Stima della dimensione dell'ambiente d'ascolto

L'analisi ANOVA mostra come la valutazione delle dimensioni della sala siano affette dalla situazione ($f=6.89$, $p<0.0001$, $df=9$) e dell'auditorium ($f=8.47$, $p<0.0001$, $df=4$) ma non significativamente dal sistema audio ($f=2.4$, $p=0.066$, $df=3$). Inoltre va fatto notare come esista una correlazione elevata fra le distanze sorgente-ricevitore percepite e le dimensioni dell'ambiente percepite (logicamente, in quanto in una sala piccola non possono esserci grandi distanze sorgente-ricevitore). Dai risultati ottenuti si può dire che la miglior correlazione tra votazioni e dimensioni lineari della sala si sia ottenuta nel caso del doppio stereo dipolo; gli altri tre sistemi si sono rivelati poco accurati nella simulazione di questo aspetto. Ciò è un risultato

da ritenersi attendibile, in quanto il doppio stereo dipolo risulta il sistema in grado di restituire la miglior ambienza durante l'ascolto, ovvero di simulare parte del fronte posteriore.

	Physical Distance	Estimated Distance
O.R.T.F.	0.46	0.95
Headphones	0.44	0.85
Stereo Dipole	0.33	0.58
Double Stereo Dipole	0.56	0.86

Tabella 3.2 – Coefficienti di correlazione (r) tra le stime delle dimensioni dell'auditorium e la distanza sorgente-ricevitore (fisica e stimata)

3.1.5.3 Valutazione del realismo

L'analisi ANOVA mostra come sostanzialmente la situazione non influisca sul giudizio (e questo è sicuramente un risultato che conferma l'affidabilità dei sistemi di riproduzione), mentre dipenda molto dal sistema audio utilizzato (f=4.15, p=0.0068, df=3). Le cuffie, ancora una volta, hanno dato i peggiori risultati in termini di realismo, mentre i migliori risultati sono stati ottenuti con l'utilizzo di stereo dipolo singolo e stereo standard.

Questo dimostra ulteriormente come la necessità di condurre test d'ascolto con sistemi di riproduzione avanzati sia elevata e debba prendere il posto del tradizionale ascolto in cuffia.

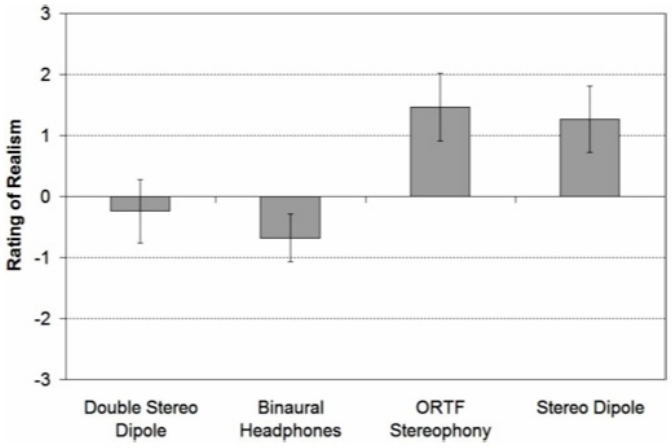


Figura 3.6 – Giudizio sul realismo dei 4 sistemi audio

Non ci è dato di sapere come un suono naturale (in situazioni reali) venga giudicato tale; nonostante questo possiamo supporre che i soggetti, aventi una notevole esperienza in campo musicale, abbiano dato i loro giudizi in base al loro personale ricordo di performance musicali dentro a sale da concerto. Ai soggetti era, infatti, chiesto di immaginarsi dentro ad una sala da concerto, non di vedersi immersi in una piccola stanza d'ascolto.

3.1.5.4 Valutazioni sui risultati

Per questo tipo di studio si è dovuto operare alcuni compromessi. Per esempio la stanza d'ascolto scelta non è anecoica (l'assenza di riflessioni l'avrebbe resa ideale per la cancellazione perfetta del cross-talk), gli speaker scelti sono di differenti modelli (nonostante la risposta in frequenza sia stata resa pressoché identica) ed il numero di sale da concerto testate era probabilmente piccolo. Nonostante ciò lo studio ha portato ad importanti considerazioni:

- La riproduzione binaurale mediante cuffia non è efficace come le alternative proposte. I punteggi sul realismo della riproduzione sono bassi, così come è cattiva la percezione della distanza sorgente-ricevitore. Questo rappresenta un importante risultato, in quanto quasi tutte le campagne di indagine, al momento, vengono effettuate ricorrendo a riproduzione mediante cuffia.
- Il doppio stereo dipolo sembra un po' inefficiente in termini di realismo. Una possibile spiegazione di ciò potrebbe essere trovata nel fatto che la testa dei soggetti non era fissa in una sola posizione: la qualità del suono e la stabilità dell'immagine sonora alle alte frequenze potrebbe essere stata degradata da movimenti accidentali.
- Il singolo stereo dipolo risulta molto efficiente in termini di realismo e percezione della distanza. Dei tre sistemi binaurali testati questo appare certamente come il migliore. Non avendo il dipolo posteriore non ci sono problemi d'interferenza davanti-dietro come succedeva per il doppio stereo dipolo.

Il sistema stereofonico presenta punteggi alti per quanto riguarda il realismo di riproduzione e appare come l'unico sistema in cui le valutazioni delle dimensioni

della stanza possono essere legate ad un parametro fisico (lunghezza). Nonostante ciò, per la valutazione della distanza, lo stereo è risultato meno efficace rispetto allo stereo dipolo.

3.2 Valutazione di teatri e sale da concerto in soggetti normudenti e ipoacusici.

Con l'aumento dell'età, il nostro sistema uditivo perde la sua funzionalità: diminuiscono il livello e l'intervallo di frequenze dei suoni percepiti. Se si pensa che la maggior parte dei frequentatori abituali di teatri e sale da concerto ha un'età superiore ai 40 anni è quindi da tenere presente che queste persone possano avere una perdita uditiva alle alte frequenze, oltre ad altri possibili tipologie di disturbi dell'apparato uditivo. Nella maggior parte dei casi, chi è consapevole del proprio deficit uditivo indossa un apparecchio acustico che, però, è spesso ottimizzato per l'ascolto del parlato, non per l'ascolto musicale. Per tutti questi motivi, una larga parte di chi normalmente siede in teatro non percepisce correttamente il suono prodotto dall'esecutore (cantanti, orchestra, singoli musicisti, ecc...).

Abbiamo più volte sottolineato l'importanza della ricerca di una correlazione tra parametri oggettivi e descrittori soggettivi dell'acustica di una sala: per effettuarla in modo significativo occorre indagare quale sia la percezione dell'utente "medio" di teatro e tenerla in considerazione in fase di test.

I test d'ascolto, eseguiti quindi con tale obbiettivo e qui diffusamente descritti, sono stati svolti in collaborazione con la *Facoltà di Medicina e Chirurgia* (Audiometria) dell'Università di Parma, oggetto di tesi dell'Ing. Marco Binelli e della Dott.sa Daniela Marmiroli sotto la mia supervisione.

3.2.1 L'hardware

Il sistema di riproduzione è costituito da due diffusori *Genelec S30D Studio Monitors* a tre vie disposti frontalmente rispetto alla postazione d'ascolto ad una distanza di 1.5 m. Tali diffusori sono stati disposti sul lato ponendo in posizione interna i tweeter (distanza 18 cm) e in posizione esterna i mid-range (40 cm) e i woofer (76 cm). Gli

angoli sono rispettivamente di 7°, 14° e 30° dalla linea mediana di simmetria, secondo i principi enunciati nel Paragrafo 1.1.3.

Per quanto riguarda il routing hardware del segnale, il sistema utilizzato è illustrato in Figura 3.7. Tutti i segnali circolanti sono in formato digitale S/PDIF, a 16 bit e con frequenza di campionamento pari a 96 kHz.

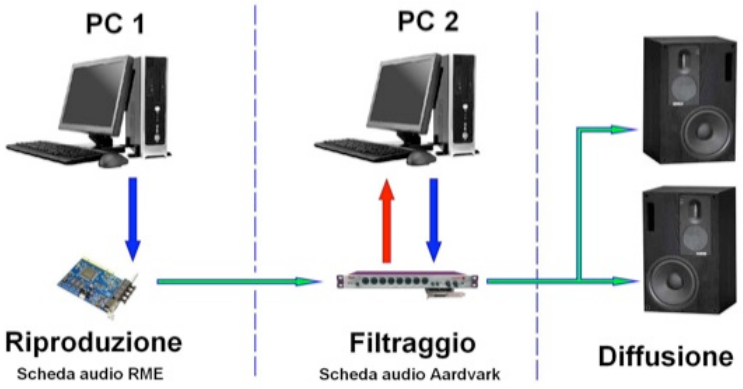


Figura 3.7 – Routing del segnale

Come è possibile notare, uno dei due PC utilizzati (il quale monta una scheda audio *RME DIGI 9636/52*) funge da *player*, mentre l'altro (dotato di una scheda *Aardvark Direct Pro Q10*) è utilizzato come elaboratore di segnale in tempo reale, al fine di eseguire i filtraggi di cross talk per lo stereo dipolo e l'equalizzazione.

Nel PC 1, che fa da riproduttore, viene utilizzato un software scritto in *Visual Basic* in grado di eseguire uno dei 5 file in formato *WAV* utilizzati per il test soggettivo. Questo programma invia i dati alla scheda audio *RME*, la quale a sua volta li instrada sull'uscita coassiale. Da questa uscita il segnale raggiunge l'ingresso digitale della scheda *Aardvark*, collegata al PC 2 che a sua volta elabora i dati in tempo reale mediante l'host VST *Plogue Bibule*. Una volta elaborato, il segnale è indirizzato ai diffusori acustici sempre mediante cavi coassiali.

L'elaborazione di segnale all'interno del PC 2 permette sia la cross-cancellazione caratteristica dello stereo dipolo sia l'equalizzazione.

I diffusori utilizzati sono dotati di amplificatori di segnale interni ed accettano in ingresso segnali bilanciati sia in formato analogico che digitale (protocollo S/PDIF). Si è scelta la soluzione digitale in modo da evitare la dipendenza della qualità di riproduzione dal tipo di scheda audio utilizzata. Ciò consente una buona flessibilità del sistema.

3.2.2 Filtraggio delle stereo dipolo

I filtri di stereo dipolo sono stati sintetizzati impiegando il PC dotato di scheda audio *Aardvark* ed un manichino dotato di microfoni auricolari (Figura 3.8).

Figura 3.8 - Testa Neumann KU70

La testa *KU70* costituisce un sistema di acquisizione binaurale. Essa modella la forma di una comune testa umana. È dotata di un padiglione auricolare modellato sempre sul modello umano e presenta una risposta in frequenza da 20Hz a 20KHz. In Figura 3.9 è mostrato il set sperimentale.

Figura 3.9 – Misura dello stereo dipolo mediante dummy-head

Per ottenere le risposte all'impulso necessarie alla generazione dei filtri inversi, si è utilizzata la tecnica dello *sweep* seno logaritmico. I segnali sono stati generati tramite il *plug-in Aurora* del software *Adobe Audition 1.5*

I filtri inversi descritti nel Paragrafo 1.1.3 sono stati generati ancora una volta mediante *Aurora*, scegliendo di troncare le risposte impulsive dopo il fronte d'onda diretto e comunque appena prima delle prime riflessioni (Figura 3.10).

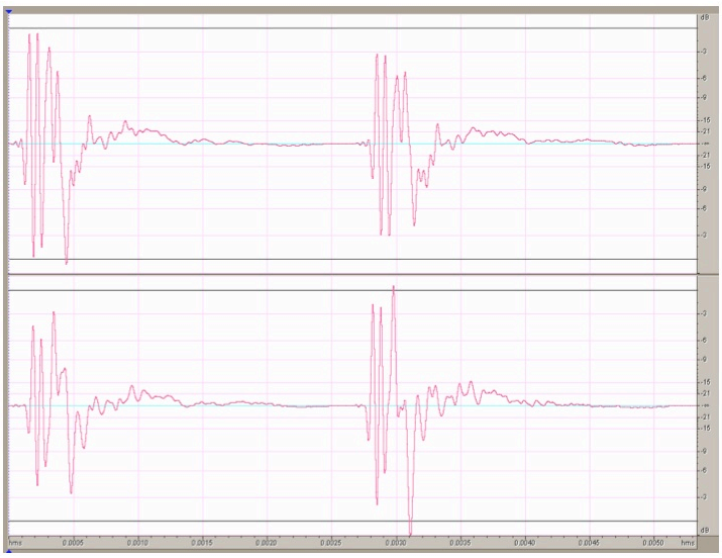


Figura 3.10 – Risposte all'impulso del solo suono diretto

Il troncamento della risposta è preceduto da una dissolvenza in uscita per cercare di non introdurre artefatti nelle risposte stesse. Le risposte sono lunghe 256 campioni ed il fading è sugli ultimi 86. E' da notare che l'inversione del solo fronte diretto esclude la possibilità di correggere le riflessioni della stanza e di qui la necessità di avere una buona acustica della sala.

Una volta realizzati i filtri inversi si è verificata la loro efficacia mediante una nuova risposta impulsiva, con il segnale di test processato da tali filtri (Figura 3.10).

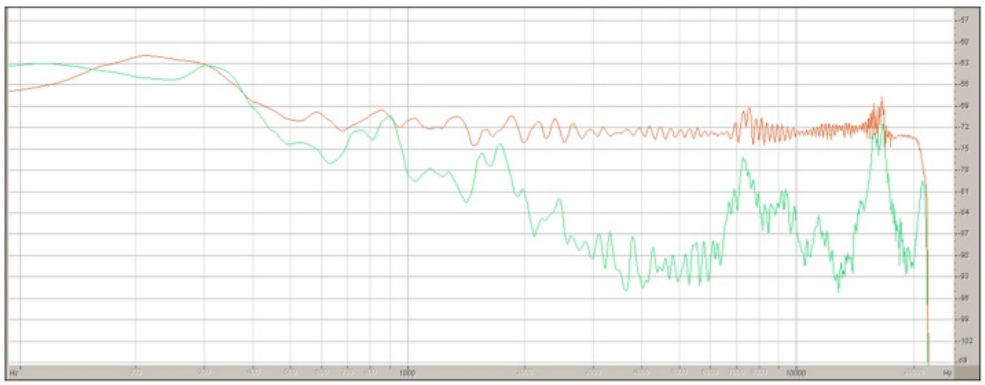


Figura 3.11 – Cancellazione post-filtering

Si può notare come la curva verde, rappresentante lo spettro del percorso trasversale (ad es. diffusore destro – orecchio sinistro) sia parecchi dB inferiore allo spettro di colore rosso, indicante il percorso diretto (diffusore sinistro – orecchio sinistro) nella banda d'interesse (200-10000 Hz).

3.2.2.1 Equalizzazione

Il passo successivo alla creazione dei filtri è l'equalizzazione. In questo caso l'operazione si è resa necessaria a causa delle condizioni non ottimali dell'acustica della sala e del campo d'azione limitato dei filtri inversi generati da *Aurora*.
Dopo aver generato un rumore rosa sul PC 1 mediante *Audition*, il segnale, elaborato in tempo reale con i filtri di stereo dipolo nel PC 2, è stato inviato ai diffusori e catturato mediante i microfoni della dummy-head; da questi è stato poi inviato alla scheda audio *Aardvark* ed analizzato mediante un analizzatore di spettro in tempo reale in bande di mezze ottave. Quest'ultima operazione è stata eseguita con un *VST plug-in* Voxengo SPAN VST. L'allineamento delle bande per ottenere la risposta in frequenza ottimale si è potuta realizzare mediante un altro *VST plug-in* (equalizzatore grafico '*MarvelEq*') inserito a valle dei filtri di stereo dipolo.

3.2.3 Filtraggio dei brani del test d'ascolto

Il brano adottato nel test d'ascolto è stato scelto tra un'insieme di brani presenti sul CD '*Anechoic Orchestral Music Recording*' di marca *Denon*, registrati in camera anecoica. Ciò si è reso necessario al fine rispettare l'acustica virtuale dei luoghi d'ascolto sottoposti al test. Grazie a registrazioni di risposte impulsive (*Impulse Response*, IR) binaurali delle sale d'ascolto in esame, si è potuto riprodurre il brano scelto come se fosse eseguito all'interno delle varie sale.
Il filtraggio è stato eseguito mediante *plug-in Aurora* in *Audition*, quindi *off-line*, secondo lo schema di Figura 3.12. Con *IR palco sinistro* (*IR palco destro*) è indicata la risposta impulsiva binaurale delle sale d'ascolto misurata con sorgente sul lato sinistro (destro) del palco e posto d'ascolto in platea centrale intorno alla decima fila. Con questo metodo si è cercato di tener conto del fatto che al microfono sinistro

delle registrazioni di partenza, arriva un segnale sonoro dalla parte di sinistra dell'orchestra, ma anche dalla destra e viceversa. Il segnale proveniente da sinistra va filtrato con la risposta "*palco sinistro - orecchio sinistro*", mentre il segnale proveniente da destra viene filtrato con "palco destra - orecchio sinistro". Stessa cosa per l'orecchio destro.

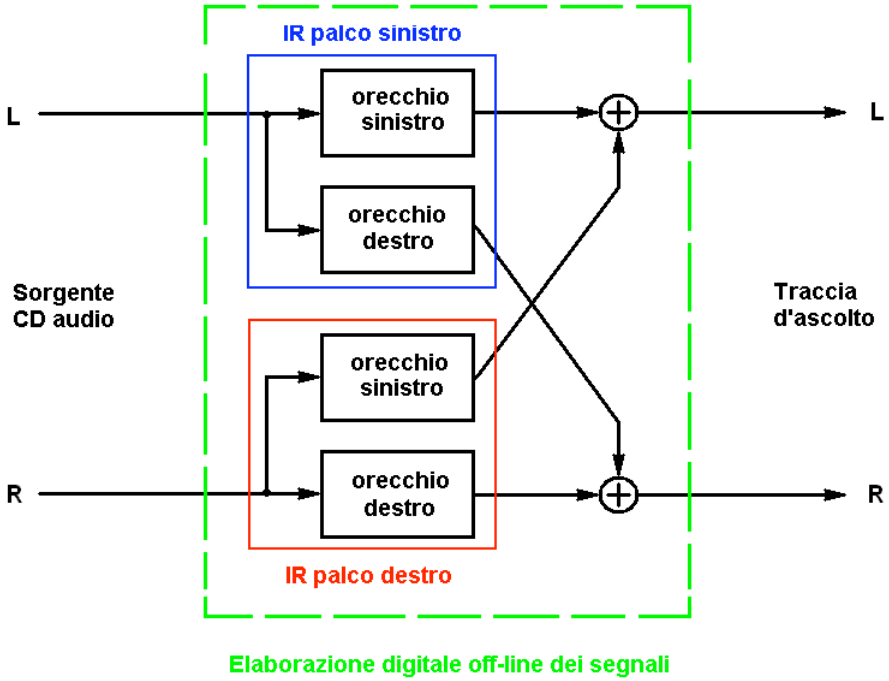


Figura 3.12 – Schema del filtraggio off-line delle tracce

Con IR palco sinistro (o IR palco destro) è indicata la risposta impulsiva binaurale delle sale d'ascolto misurata con sorgente sul lato sinistro (o destro) del palco e posto d'ascolto in platea centrale intorno alla decima fila. Con questo metodo si è cercato di tener conto del fatto che al microfono sinistro delle registrazioni di partenza, arriva un segnale sonoro dalla parte di sinistra dell'orchestra, ma anche dalla destra e viceversa. Il segnale proveniente da sinistra va filtrato con la risposta "*palco sinistro - orecchio sinistro*", mentre il segnale proveniente da destra viene filtrato con "*palco destro - orecchio sinistro*". Stessa cosa per l'orecchio destro.

3.2.4 Il test

Nel seguente capitolo si illustreranno le varie fasi di costruzione del test d'ascolto, dalla redazione e distribuzione del questionario, alla raccolta ed elaborazione dei dati statistici, alla scelta del campione, fino alla struttura del test vero e proprio.

Al fine di svolgere il test d'ascolto soggettivo su una popolazione in qualche modo allenata all'ascolto ed al giudizio, si è pensato di reclutare persone frequentanti il *Teatro Regio* e l'*Auditorium Paganini* di Parma mediante la distribuzione di un questionario.

3.2.4.1 Redazione e distribuzione del questionario

La preparazione del questionario (Appendice A.3) è stata effettuata in collaborazione con gli uffici della fondazione *Teatro Regio*. Le domande inserite sono di diversa natura: le prime (1 - 4) di carattere anagrafico volte a classificare la popolazione del teatro; altre (5 - 9), più personali, volte alla conoscenza delle abitudini dei soggetti e quindi al perfezionamento dei profili di ascoltatore; altre ancora (10 - 12) necessarie alla verifica del grado di informazione sulle problematiche legate all'udito. L'ultima domanda è la richiesta di partecipazione al test svolto presso la *Casa Della Musica*, con assenso alla trattazione dei dati personali.

I questionari consegnati all'ingresso sono stati restituiti al termine dello spettacolo ad un apposito banco allestito all'interno dei locali sopra citati. Il numero di questionari compilati raccolti è pari a 900. Di questi, oltre 300 presentano una compilazione totale e quindi un assenso alla partecipazione al test.

3.2.4.2 Raccolta ed elaborazione dei dati

Tutti i dati raccolti sono stati utilizzati al fine di trarne le statistiche.

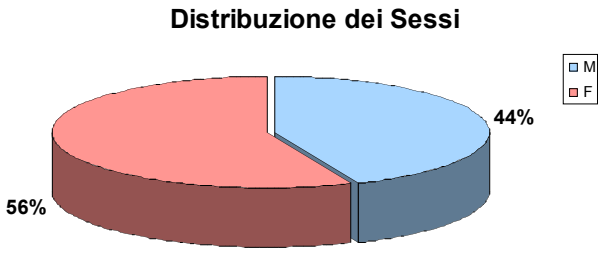


Figura 3.13 – Popolazione maschile e femminile

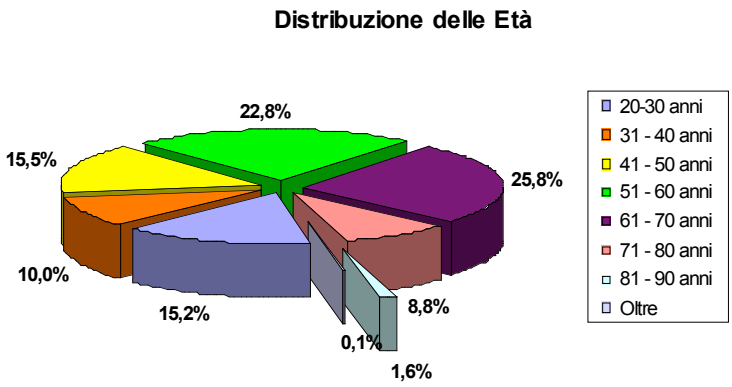


Figura 3.14 – Classi di età e percentuali sul totale

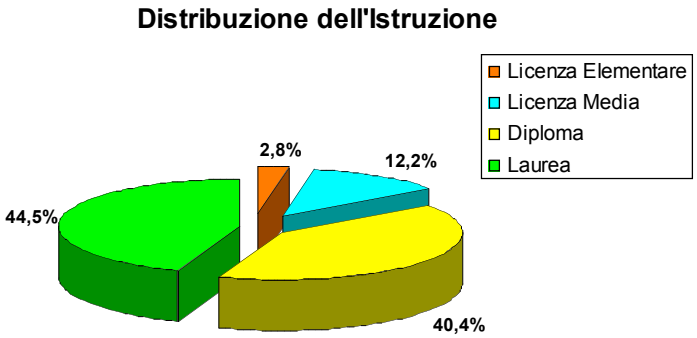


Figura 3.15 – Suddivisione in base al titolo di studio

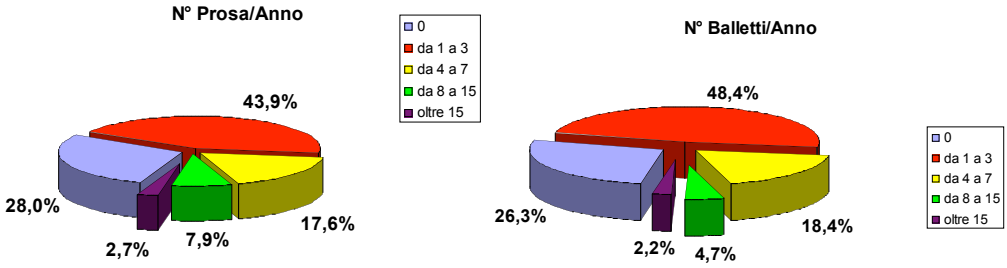


Figura 3.16 – Quantità di spettacoli di prosa e balletto frequentati in un anno

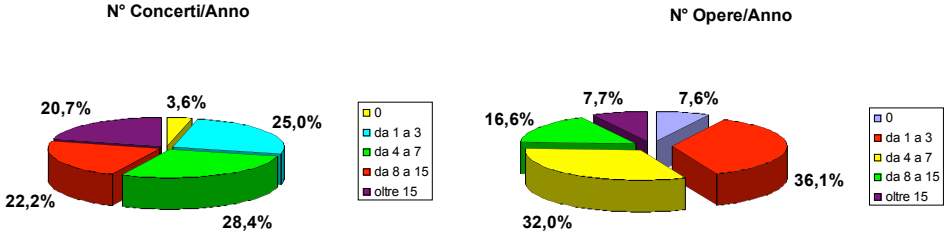


Figura 3.17 – Quantità di concerti e opere frequentati in un anno

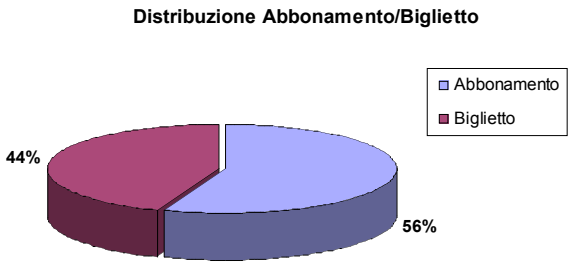


Figura 3.18 – Preferenza tra abbonamento e biglietto singolo

Basandosi sulle statistiche raccolte, è stato scelto un campione dei soggetti che hanno aderito al test. La selezione è avvenuta basandosi soprattutto sulla distribuzione delle età, visto che è, dal punto di vista medico, la più significativa per quanto riguarda i problemi di udito che affliggono la popolazione.

3.2.4.3 Il test d'ascolto

Il test d'ascolto eseguito presso la Casa Della Musica è strutturato nelle seguenti parti:

- Esame audiometrico;
- Ascolto in campo libero;
- Ascolto con protesi acustica.

L'esame audiometrico eseguito a cura di un tecnico audiometrico (nonché laureanda presso la facoltà di Medicina e Chirurgia dell'università di Parma, Daniela Marmiroli) è servito a raccogliere dati necessari per la programmazione degli apparecchi acustici utilizzati nel secondo ascolto oltre a creare statistiche utilizzate anche da noi per l'analisi dei risultati. Utilizzando un audiometro portatile e cuffie acustiche chiuse, ogni soggetto è stato sottoposto all'esame di valutazione dell'udito per una durata di 10 minuti: l'intervallo di frequenze testate va dai 125 Hz agli 8000 Hz.

La seconda fase del test consiste nell'ascolto del brano Overture de "*Le Nozze di Figaro*" (Mozart) tramite stereo dipolo (Capitolo 1). Per ottenere i giudizi sull'acustica delle cinque sale (A,B,C,D,E) secondo gli aggettivi proposti, si è ricorso ad un particolare software (Figura 3.19). Questo programma, sintetizzato tramite *Visual Basic*, consente di riprodurre una delle cinque tracce stereo a seconda della

selezione che avviene tramite mouse pigiando uno dei bottoni con le lettere associate alle sale. Se un tasto associato ad una sala diversa da quella in ascolto viene schiacciato mentre la traccia è in riproduzione, questa cessa immediatamente, mentre inizia la riproduzione dell'altra dal punto del brano a cui era arrivata la prima. L'effetto che si ha è quello di un cambiamento virtuale di sala d'ascolto mentre il brano è in esecuzione. Per ciò che riguarda i giudizi, sono espressi in via grafica mediante un click su uno dei pallini bianchi (che al click si annerisce) tra le coppie di aggettivi. Naturalmente, più il pallino annerito è vicino all'aggettivo, più il giudizio pende per l'aggettivo stesso.

Figura 3.19 – Software di compilazione per il test d'ascolto

Data la presenza di soggetti con esperienza nell'uso del mouse e del PC scarsa o addirittura nulla, sono stati preparati fogli per la raccolta dei giudizi.

La terza ed ultima fase del test consiste nella ripetizione della seconda fase ma inserendo nelle orecchie dei soggetti le protesi acustiche "*Oticon Tego*" (Figura 3.20) su entrambe le orecchie ed opportunamente regolate per compensare il deficit uditivo di ogni singola persona.

I soggetti analizzati in totale sono 30 e per ogni soggetto sono stati registrati due test, con e senza apparecchio acustico.

Figura 3.20 – Apparecchi acustici Oticon Tego

3.2.5 Risultati

3.2.5.1 Esame Audiometrico

Una prima classificazione può essere fatta studiando le perdite medie di udito secondo le classi di età dei soggetti.

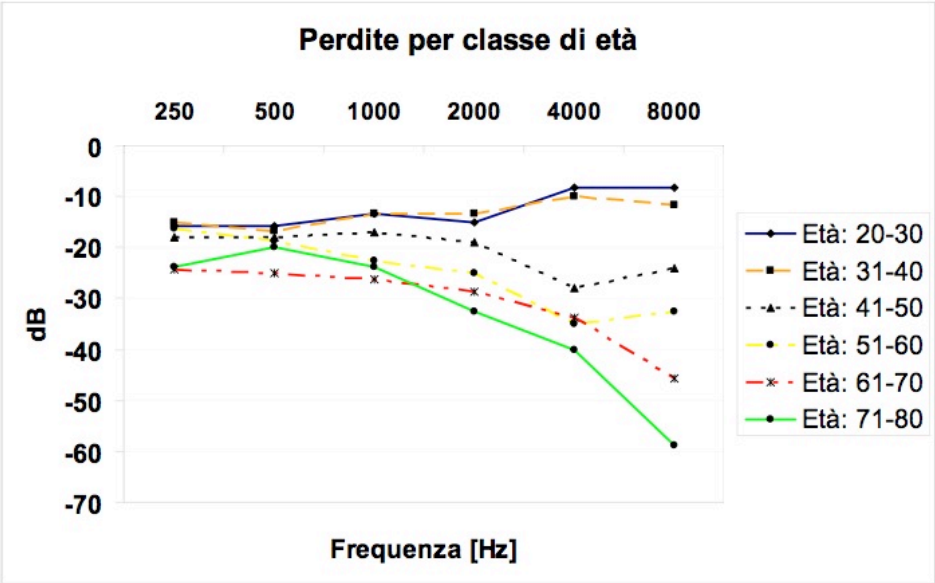


Figura 3.21 – Perdite uditive nei soggetti sotto test divise per classe d'età

Come si può osservare dal grafico, nemmeno i soggetti più giovani (da 20 a 40 anni) godono di un udito perfetto, anche se, secondo le statistiche mediche, dovrebbero risultare per la maggior parte normudenti. La perdita sulle basse frequenze è, con tutta probabilità, imputabile a sintomi di patologie da raffreddamento (come il comune raffreddore), le quali generano fluidi che intasano i condotti uditivi. Per quel che riguarda i soggetti da 41 a 60 anni, le curve cominciano ad evidenziare un progressivo calo verso le frequenze acute, ma con una risalita sull'ultima banda.

Questa risalita è tipica di chi ha subito un trauma acustico. Infine, si riscontrano, nelle curve relative ai soggetti più anziani (da 61 a 80 anni), le caratteristiche proprie della presbiacusia, ovvero un tracciato in discesa.

Una seconda analisi che può essere fatta, è il confronto tra la considerazione delle proprie capacità uditive da parte dei soggetti, ed i risultati del test audiometrico. Come si vede dal grafico di Figura 3.22, una larga maggioranza del pubblico giudica normale il proprio udito. Se paragoniamo questo risultato alle perdite medie sui 4 e 8 kHz (bande importanti nella percezione musicale), scopriamo che questa percentuale cala, mentre aumenta considerevolmente la zona di forte compromissione.

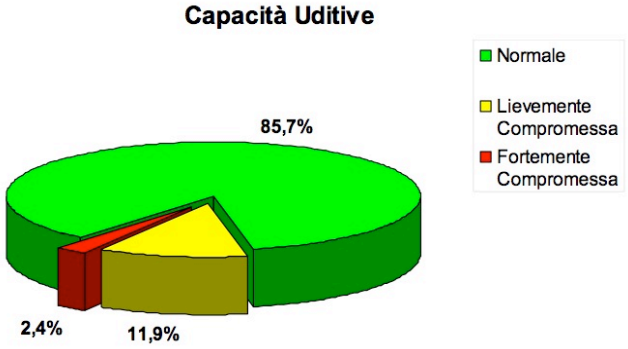


Figura 3.22 – Distribuzione delle risposte ai questionari sul giudizio delle proprie capacità uditive

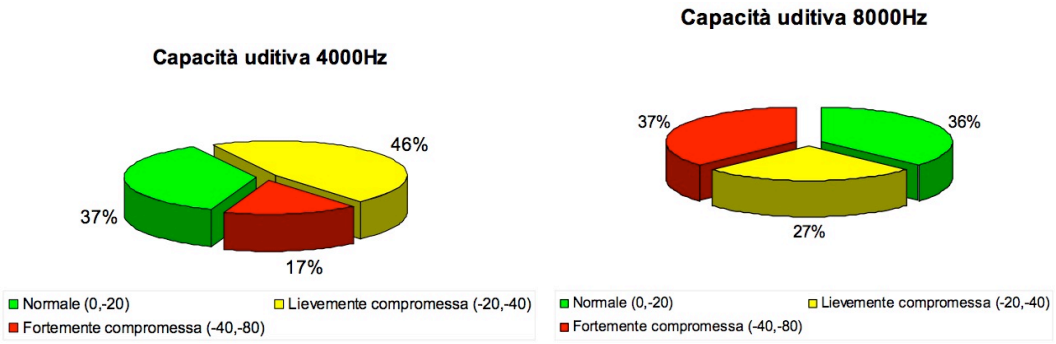


Figura 3.23 – Distribuzione delle perdite secondo gli esami audiometrici

Una considerazione su questi risultati può essere quella di una generale sottostima dei problemi uditivi da parte della popolazione. Questa affermazione è riscontrabile anche nella letteratura medica.

3.2.5.2 Test d'ascolto

L'analisi dei risultati è stata svolta attraverso due metodi: analisi multivariata e metodo dei minimi quadrati. L'analisi multivariata, realizzata con il software Simca P11, ci ha fornito dati interessanti in merito alle coppie di aggettivi [37] proposti che sono riportati di seguito:

Piacevole-Spiacevole
Rotondo-Spigoloso
Morbido-Duro
Diffuso-Localizzabile
Distaccato-Avvolgente
Secco-Rimbombante
Acuti accentuati-Acuti
Bassi accentuati-Bassi
Sommesso-Sonoro

Si è cercata, in questo caso, una relazione tra la prima coppia di aggettivi le altre. Questo dà un'indicazione dei parametri su cui si basa un giudizio di piacevolezza di un ascolto. Nel grafico sotto riportato (Figura 3.24), sono disposte n ordine decrescente di importanza le coppie che influenzano la prima (domanda 1) valutate nel test senza audioprotesi; in Figura 3.25 vi sono invece le coppie relative al test con apparecchio acustico.

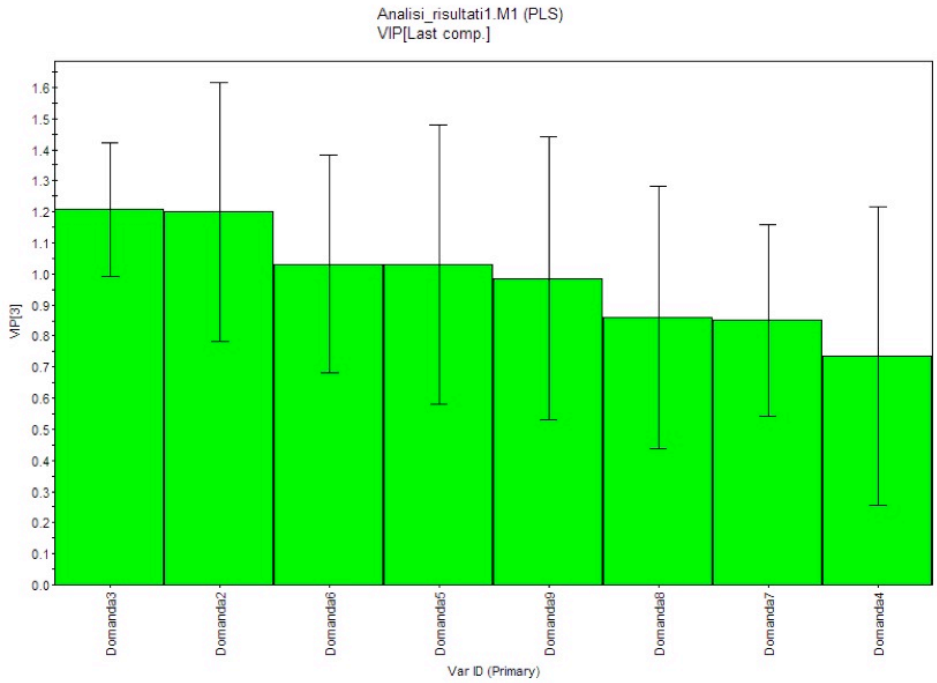


Figura 3.24 - Influenza dei giudizi 2-8 sul giudizio piacevole-spiacevole in assenza di audio protesi

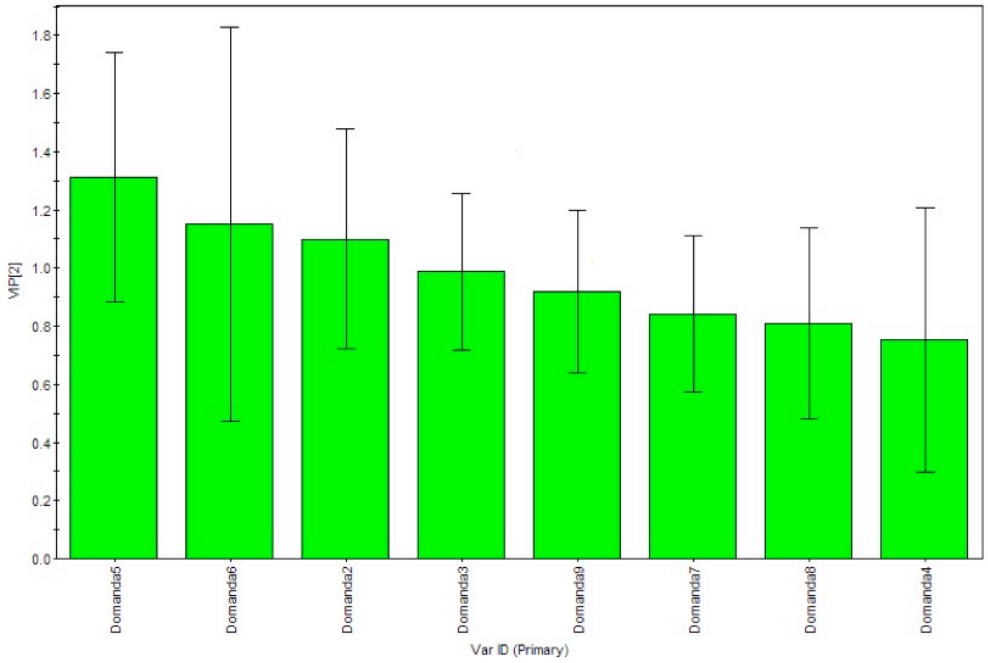


Figura 3.25 – Influenza dei giudizi 2-8 sul giudizio piacevole-spiacevole in presenza di audio protesi

È possibile notare che le prime quattro coppie (Rotondo-Spigoloso, Morbido-Duro, Distaccato-Avvolgente, Secco-Rimbombante) siano le stesse in entrambi i casi.

Questo significa che i giudizi associati a 'Piacevole Spiacevole non variano con o senza audioprotesi.

Il secondo metodo di analisi (minimi quadrati) pone l'attenzione sulla relazione 'parametri oggettivi –parametri soggettivi', dove per parametri oggettivi si intendono quei descrittori citati nella norma ISO 3382 per la descrizione acustica di una sala.

Di seguito si riportano i coefficienti di regressione ottenuti nel test senza audioprotesi (Tabella 3.3) e con audioprotesi (Tabella 3.4). In grassetto sono evidenziati i valori più elevati.

	C50 [dB]	C80 [dB]	D50 [%]	Ts [ms]	EDT [s]	T20 [s]	T30 [s]	LF	IACC	TB	BR
Piacevole-Spiacevole	0,01	-0,05	0,10	-0,06	0,07	0,15	0,24	-0,16	-0,06	**-0,33**	-0,24
Rotondo-Spigoloso	0,01	0,02	0,15	-0,18	-0,07	0,02	0,18	-0,23	0,01	**-0,36**	-0,22
Morbido-Duro	0,10	0,06	0,21	-0,18	-0,03	0,03	0,14	-0,27	0,02	**-0,36**	-0,19
Diffuso-Localizzabile	0,15	0,10	0,19	-0,14	-0,06	-0,07	-0,10	-0,19	0,04	-0,16	-0,02
Distaccato-Avvolgente	0,04	-0,03	-0,03	0,12	0,09	0,03	-0,11	0,09	-0,05	0,12	0,08
Secco-Rimbombante	-0,12	-0,23	-0,18	**0,29**	0,25	0,24	0,08	0,23	-0,22	0,05	-0,11
Acuti Accent.-Acuti Ridotti	-0,15	-0,24	**-0,30**	**0,41**	**0,29**	0,23	0,01	**0,39**	-0,20	**0,30**	0,04
Bassi Accent.-Bassi Ridotti	0,24	0,27	**0,36**	**-0,39**	-0,21	-0,17	0,01	**-0,44**	0,20	**-0,32**	-0,03
Sommesso-Sonoro	0,11	0,25	0,23	**-0,38**	**-0,34**	**-0,31**	-0,13	**-0,30**	0,22	-0,14	0,05

Tabella 3.3 -Matrice dei coefficienti di regressione lineare nel test senza audioprotesi

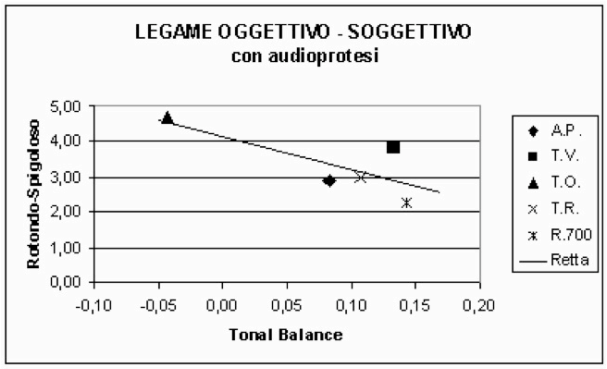
	C50 [dB]	C80 [dB]	D50 [%]	Ts [ms]	EDT [s]	T20 [s]	T30 [s]	LF	IACC	TB	BR
Piacevole-Spiacevole	-0,03	-0,04	0,06	-0,08	0,01	0,09	0,21	-0,13	-0,03	-0,27	-0,21
Rotondo-Spigoloso	0,13	0,11	0,26	-0,26	-0,09	-0,03	0,10	**-0,33**	0,06	**-0,39**	-0,17
Morbido-Duro	0,08	0,14	0,20	-0,29	-0,19	-0,12	0,08	-0,29	0,12	-0,27	-0,09
Diffuso-Localizzabile	0,11	0,12	0,14	-0,14	-0,06	-0,05	0,02	-0,16	0,10	-0,09	0,03
Distaccato-Avvolgente	-0,04	-0,02	-0,14	0,15	0,04	-0,02	-0,13	0,20	0,00	**0,30**	0,18
Secco-Rimbombante	-0,23	-0,28	**-0,30**	**0,34**	0,24	0,26	0,17	**0,33**	-0,21	0,12	-0,11
Acuti Accent.-Acuti Ridotti	-0,27	**-0,33**	**-0,39**	**0,45**	**0,29**	0,27	0,09	**0,47**	-0,25	0,28	-0,04
Bassi Accent.-Bassi Ridotti	0,12	0,22	0,23	**-0,34**	-0,25	-0,21	-0,01	**-0,30**	0,20	-0,18	0,02
Sommesso-Sonoro	0,22	**0,29**	**0,32**	**-0,38**	-0,27	-0,26	-0,11	**-0,37**	0,24	-0,18	0,08

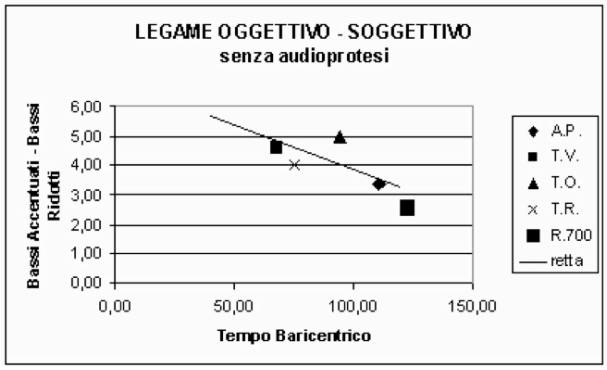
Tabella 3.4 - Matrice dei coefficienti di regressione lineare nel test con audioprotesi

È possibile quindi fare alcune osservazioni. Il parametro Tonal Balance è quello con maggiori correlazioni nel test senza apparecchi. Nel test con apparecchi alcune di

queste correlazioni scompaiono (in particolare quelle con acuti accentuati - acuti ridotti e bassi accentuati – bassi ridotti) forse a causa di compressioni del segnale interne alle protesi. Altri relazioni che permangono attraverso le due prove sono soprattutto quelle relative all'istante baricentrico dell'energia Ts. Questo si lega bene a coppie di aggettivi come Secco- Rimbombante, Acuti accentuati - Acuti ridotti, Bassi accentuati - Bassi ridotti, Sommesso – Sonoro. Altro parametro con buone correlazioni è il Lateral Fraction con le ultime tre coppie sopra citate. Infine è da notare la quasi totale assenza di relazioni significative con parametri come T20, T30 e IACC che in precedenti studi hanno invece rivelato alti valori di regressione.

Di seguito si riportano alcuni grafici che evidenziano i legami più importanti tra i parametri oggettivi – soggettivi. Sono disegnate le rette che meglio approssimano un legame lineare secondo il metodo dei minimi quadrati e i punti effettivamente trovati attraverso medie sul campione. Le abbreviazioni sui nomi dei teatri sono le seguenti:T.R.: Teatro Regio di Parma ,A.P.: Auditorium Paganini di Parma, T.V.: Teatro Valli di Reggio Emilia ,R. 700: Auditorium di Roma sala 700, T.O.: Teatro Olimpico di Vicenza.





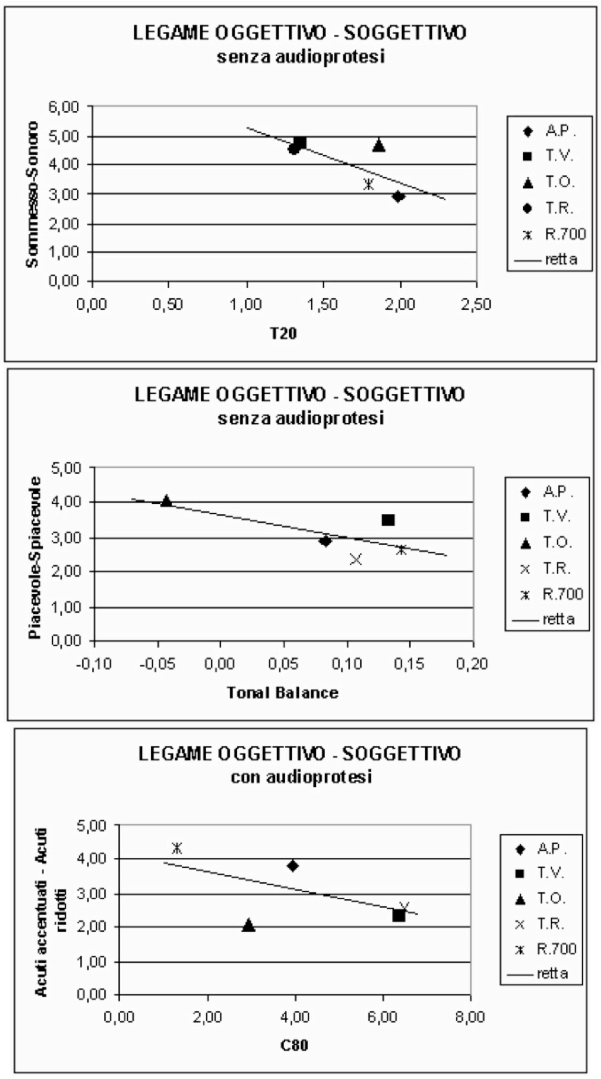


Figura 3.26 – Rette di regressione lineare per coppie di giudizi

In Figura 3.26 è possibile notare che al crescere del bilanciamento tra medi e bassi (parametro Tonal Balance) il suono è considerato più rotondo. Lo stesso parametro, valutato con audioprotesi, sembra influire sulla piacevolezza: un eccessiva enfatizzazione delle basse frequenza rende spiacevole il suono.

Si può osservare che se l'istante baricentrico dell'energia (Ts) aumenta si ha una sensazione di accentuazione delle basse frequenze dovuta proprio al maggior peso dell'energia dovuta alle riflessioni. Per quanto riguarda il parametro di chiarezza C80 si nota che se questo aumenta si può provare una sensazione di accentuazione delle frequenze acute. Ciò probabilmente si deve al fatto che aumentando la frazione di energia diretta, si possono distinguere meglio i suoni acuti.

3.3 Localizzazione in Stereo Dipolo e Ambisonic

Per riprodurre fedelmente un campo acustico, per dare all'ascoltatore l'impressione di trovarsi realmente presente all'evento sonoro è importante che il sistema audio scelto per la riproduzione permetta una corretta localizzazione delle sorgenti sonore. In poche parole, occorre che un suono che originariamente proveniva da una particolare direzione, in fase di riascolto venga percepito proprio in quella direzione.
Con i sistemi audio oggi diffusi (stereo, surround 5.1, ecc...), il suono è solo indirizzato alla cassa più vicina alla direzione originale ma esso, se ad esempio il suono arriva da una direzione supponiamo collocata tra due casse, non sarà mai percepito correttamente. Diverso è il discorso se si parla dei già citati sistemi "Stereo Dipolo" e "Ambisonic", in cui la ricostruzione del campo sonoro avviene il più possibile fedelmente.
Ovviamente, in quanto sistemi intrinsecamente differenti, i due permettono una localizzazione più o meno accurata e proprio per comprenderne le differenze si è scelto di effettuare test d'ascolto compartivi. Anche in questo caso è stata utilizzata la saletta d'ascolto alla Casa della Musica attrezzata con il trattamento acustico definitivo (Capitolo 2.3) ed equipaggiata con l'hardware descritto nel Capitolo 2.5.

3.3.1 Sistemi audio 3D

3.3.1.1 Ambisonic

Mentre per layout regolari è possibile trovare matrici di decodifica (Capitolo 1.1.2) tramite metodi analitici, una metodologia così schematica non sembra esistere per disposizioni irregolari degli speaker come quella utilizzata per questo test.
Per estrapolare le matrici corrette di decodifica è stato usato il software *Makedec*, che consente una simulazione virtuale ed interattiva delle performance dei decoder Ambisonic di primo e secondo ordine (Figura 3.27). Questo software calcola la matrice di decodifica in bassa frequenza operando la pseudo-inversione della matrice per ottenere i componenti ambisonic. Questa matrice può essere poi successivamente modificata, se necessario, ed è anche il punto di partenza per l'ottimizzazione manuale della matrice per le alte frequenze. Il primo step per trovare una buona matrice di conversione in alta frequenza è quello di aggiustare i guadagni "pre-order"

della matrice alle basse frequenze, aggiustando poi i polar pattern individuali corrispondenti all'output di ogni speaker.

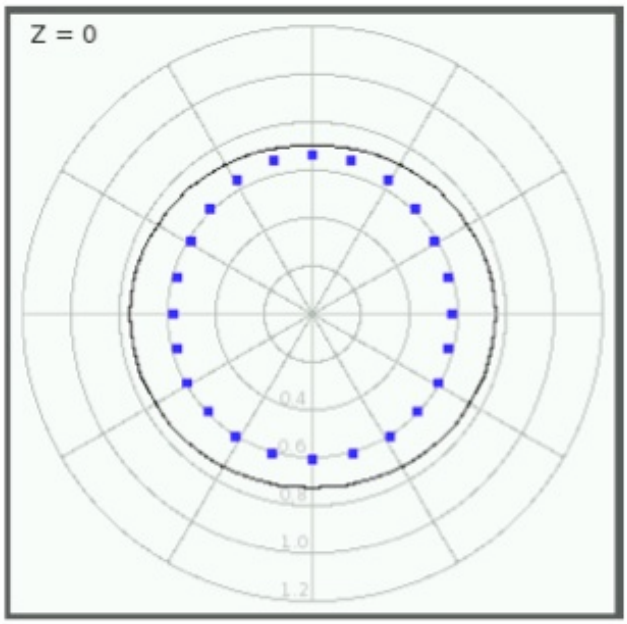


Figura 3.27 – Makedec: vettore energia sul piano orizzontale (Adriaensen 2007)

Una volta trovate le matrici ottimali è possibile salvarle in un formato compatibile con il software Ambdec. Oltre ad implementare i filtri di *cross-over* allineandone la fase, questo programma permette di avere la compensazione per campo vicino e un'equalizzazione dei ritardi per ogni singolo speaker.

3.3.1.2 Stereo dipolo

L'implementazione di questo sistema, ad opera dell'Ing. Simone Fontana, è avvenuta in accordo con i principi teorici espressi nel Capitolo 1.1.3, con l'aggiunta di uno stereo dipolo sul capo dell'ascoltatore per conferire maggiore energia al suono proveniente dall'alto.

I due stereo dipoli planari sono formati da due altoparlanti in posizione orizzontale, con una buona distanza tra i woofer ed una più piccola tra i tweeter, in accordo con la "Distribuzione Ottimale delle Sorgenti". Lo stereo dipolo posteriore presenta 10° di elevazione ed è quindi più alto di quello frontale che si trova nel piano azimutale: questa scelta è dovuta al fatto che la risposta nella parte posteriore ha alcuni

avvallamenti e generalmente è più debole. Dando un po' di elevazione ai trasduttori si ottiene una migliore linearità della risposta [40].

Per l'inversione è stata usata una regolarizzazione dipendente dalla frequenza, con un parametro di regolarizzazione pari a 0.05 e con un profilo, dipendente dalla frequenza, pari a 10 sotto i 200 Hz e 100 sopra i 16000 Hz. La funzione "target" è una delta per il suono diretto ed una "funzione-zero" per il cross-talk; i filtri sono stati progettati con lunghezza di 4096 campioni. L'algoritmo di inversione è stato usato per trovare i filtri di linearizzazione dei tre stereo dipoli e risolve i problemi di sincronizzazione tra le tre coppie di segnali (le tre delta, nel dominio del tempo) sono allineate.

La cancellazione si è rivelata ottimale nell'intervallo 700-10000 Hz, raggiungendo -80 dB con un valore medio di cancellazione di -40 dB (Figura 3.28).

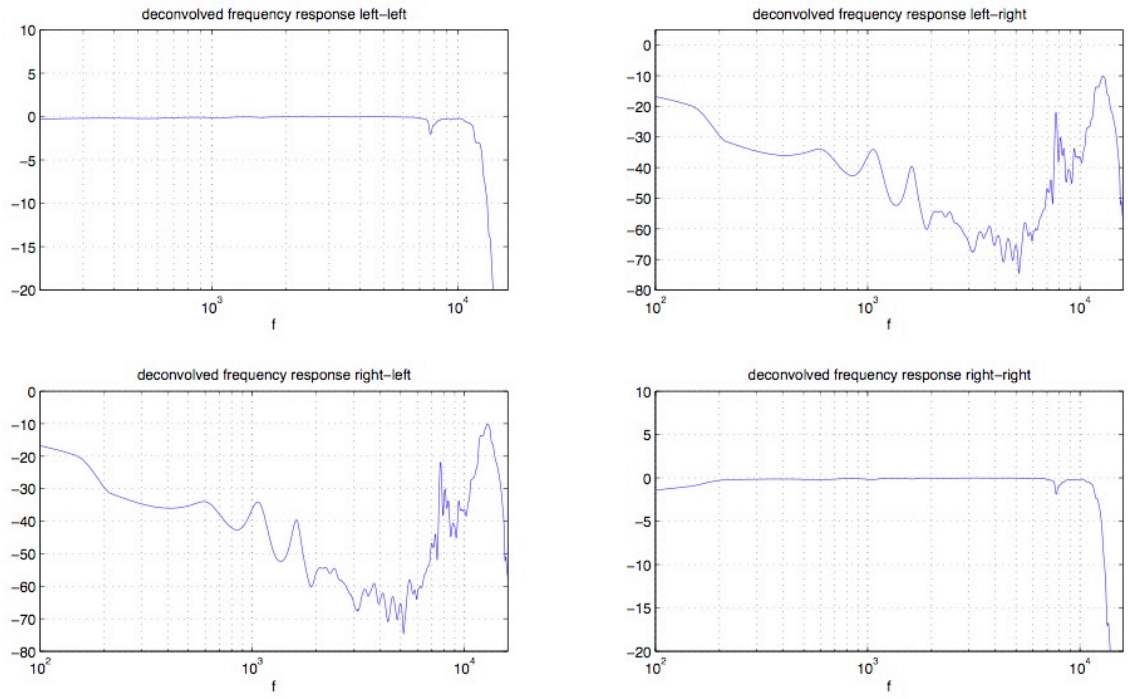


Figura 3.28 – Risultati dell'effettiva cancellazione del cross-tlak

I risultati reali, diversamente da quelli simulati, sono affetti da rumore di amplificazione e da spostamenti spettrali dovuti alla leggera varianza temporale del canale acustico [46].

3.3.2 Il test

Il test è consistito nell'ascolto di un suono proveniente da differenti direzioni: i soggetti sotto test hanno dovuto indicare la direzione (elevazione e azimut) percepita. Per ogni ascoltatore sono state testate 25 direzioni (Figura 3.29) uguali per ogni soggetto, scelti su una griglia virtuale con passi di 45° per l'azimut e 30° per l'elevazione (quest'ultima è stata limitata a -30° sotto 0°). Sono state riprodotte, in modo pseudo-casuale 21 direzioni tramite i due sistemi audio, divise in due sequenze: nella prima 10 posizioni sono state riprodotte da un sistema e 11 dall'altro, viceversa nella seconda sequenza. Dei 20 soggetti scelti, 10 hanno ascoltato la prima sequenza, dieci la seconda. Sono stati inseriti nella valutazione anche dei suoni provenienti direttamente da altoparlanti posizionati in 8 differenti punti (4 nella prima sequenza e 4 nella seconda), con lo scopo di testare la capacità dei soggetti nel discriminare, in un campo sonoro reale e non simulato, l'esatta direzione d'arrivo. In questo modo si è valutata anche l'affidabilità dei soggetti scelti.

All'inizio della sessione il soggetto aveva la possibilità di aggiustare la sua posizione all'interno dello *sweet spot* usando un segnale di riferimento, posizionato a 0° di elevazione e 0° azimut (posizione frontale) e suonato da entrambi i sistemi.

Una piccola sfera di plastica, con gli angoli di azimut ed elevazione segnati sulla superficie, ha aiutato i soggetti nel trovare e scrivere le direzioni percepite.

Il segnale di test usato è rumore rosa filtrato nella banda tra i 300 Hz e i 16 kHz, con una durata di 130 ms (12 ripetizioni in ogni direzione virtuale); tra una direzione e la successiva c'erano 10 secondi di silenzio. Le direzioni sono state sintetizzate, per il sistema Ambisonic, usando il plug-in VST *Gerzonic Paorama* [41] inserito in nell'host VST *Plogue Bidule* [42], ottenendo 4 segnali (B-format) per ogni posizione; i segnali per lo Stereo Dipolo sono stati generati usando *Voxengo Pristine Space* [43], inserito anch'esso in Plogue Bidule, usandolo quindi per la cancellazione di cross-tlak e per la convoluzione con un set di HRTF scelte tra quelle della raccolta *Listen* [44].

Tutte le posizioni, in accordo con le sequenze prestabilite, sono state posizionata all'interno di due sessioni di *Ardour* [45], una "digital audio workstation". Il modo scelto per presentare i dati segue le indicazioni di H.Moller [46]. Sull'asse y vengono graficate le 21 posizioni virtuali con i loro nomi espliciti e, sull'asse x, le 33

possibili posizioni. L'ordine scelto per le etichette (Back Left, Back Left Low, ecc...) segue l'azimut delle posizioni, così che sia possibile dividere la griglia di presentazione in tre settori verticali e tre orizzontali, come mostrato in Figura 3.25. Tre sono i settori considerati: emisfero sinistro, emisfero destro e piano mediano.
Gli errori possono allora essere inizialmente classificati (Figura 3.29) in:

- Confusione Destra/Sinistra e Sinistra/Destra, che ci si aspetta essere assente
- Confusione Mediano/Destra, Destra/Mediano, Mediano/Sinistra e Sinistra/Mediano, che ci si aspetta essere scarsamente presente

Per la corretta comprensione dei risultati è importante notare che il termine XX nella confusione XX/YY è sempre riferito alla posizione simulata e YY a quella percepita.
Gli errori in ogni settore possono essere classificati come errori Front/Back e Back/Front (Figura 3.29).
Una differente classificazione è relativa alla possibilità di raggruppare alcune posizioni nei cosiddetti "coni di confusione", che possono essere definiti come "punti nello stesso spazio tridimensionale che producono gli stessi ITD (Interaural Time Difference) e ILD (Interaural Level Difference). Le sorgenti virtuali posizionate nello stesso cono possono essere difficili da distinguere perchè necessitano di un'accurata precisione spettrale in fase di riproduzione; l'altra faccia della stessa medaglia è che suoni in differenti coni sono più facili da distinguere. Un esempio di cono di confusione è il piano mediano, dove ITD e ILD sono uguali a zero. Gli altri coni sono indicati in Tabella 3.4.

3.3.3 I risultati

3.3.3.1 Presentazione

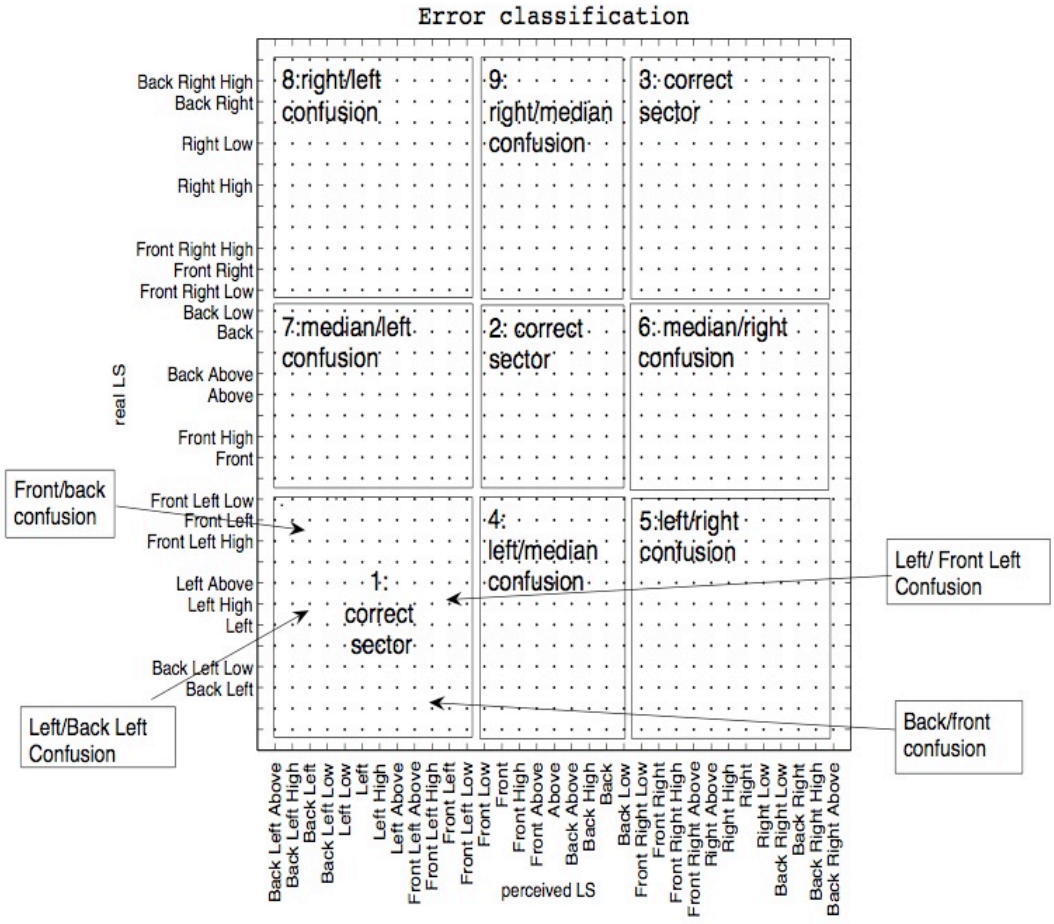


Figura 3.29 – Zone di classificazione degli errori

135	-30	Back Left Low	Left hemisphere	Cone 1
135	30	Back Left High	Left hemisphere	Cone 1
45	-30	Front Left Low	Left hemisphere	Cone 1
45	30	Front Left High	Left hemisphere	Cone 1
90	-30	Left Low	Left hemisphere	Cone 2
90	30	Left High	Left hemisphere	Cone 2
45	0	Front Left	Left hemisphere	Cone 3
135	0	Back Left	Left hemisphere	Cone 3
135	60	Beck Left Above	Left hemisphere	Cone 4

45	60	Front Left Above	Left hemisphere	Cone 4
0	-30	Front Low	Median plane	Cone 5
0	0	Front	Median plane	Cone 5
0	30	Front High	Median plane	Cone 5
0	60	Front Above	Median plane	Cone 5
0	90	Above	Median plane	Cone 5
180	60	Back Above	Median plane	Cone 5
180	30	Back High	Median plane	Cone 5
180	0	Back	Median plane	Cone 5
180	-30	Back Low	Median plane	Cone 5
315	60	Front Right Above	Right hemisphere	Cone 6
225	60	Back Right Above	Right hemisphere	Cone 6
225	0	Back Right	Right hemisphere	Cone 7
315	0	Front Right	Right hemisphere	Cone 7
270	30	Right High	Right hemisphere	Cone 8
270	-30	Right Low	Right hemisphere	Cone 8
315	30	Front Right High	Right hemisphere	Cone 9
315	-30	Front Right Low	Right hemisphere	Cone 9
225	30	Back Right High	Right hemisphere	Cone 9
225	-30	Back Right Low	Right hemisphere	Cone 9

Figura 3.30 – Posizioni simulate

Un possibile errore derivante dalla suddivisione in coni è quello "Dentro al cono – Fuori dal cono". Questa tipologia di errore non verrà però qui indagata.

3.3.3.2 I risultati del test

Nella Figura 3.31 vengono graficati i risultati della localizzazione dei diffusori reali. Si può osservare un buon accordo con la maggior parte delle posizioni: cinque degli speaker sono stati esattamente localizzati dall'80% dei soggetti e dal 100% se si considera un errore di 30° in elevazione e azimut. Si può pensare che questo piccolo errore residuo possa essere eliminato scegliendo un metodo più naturale di selezione

della direzione di arrivo, per esempio utilizzando un puntatore laser. La sorgente posteriore è male localizzata, in accordo però con le difficoltà che normalmente si hanno nella localizzazione nel piano mediano. La sorgente "*back low*" (dietro, in basso) presenta una stranissima localizzazione per uno dei soggetti che però non è stato scartato in quanto, nel caso delle altre sorgenti, ha mostrato ottime capacità di identificarne la loro posizione. La sorgente "*left high*" (in alto a sinistra) è stranamente male localizzata: questo effetto è stato attribuito a qualche possibile riflessione critica presente nella stanza non spiegato in modo chiaro.

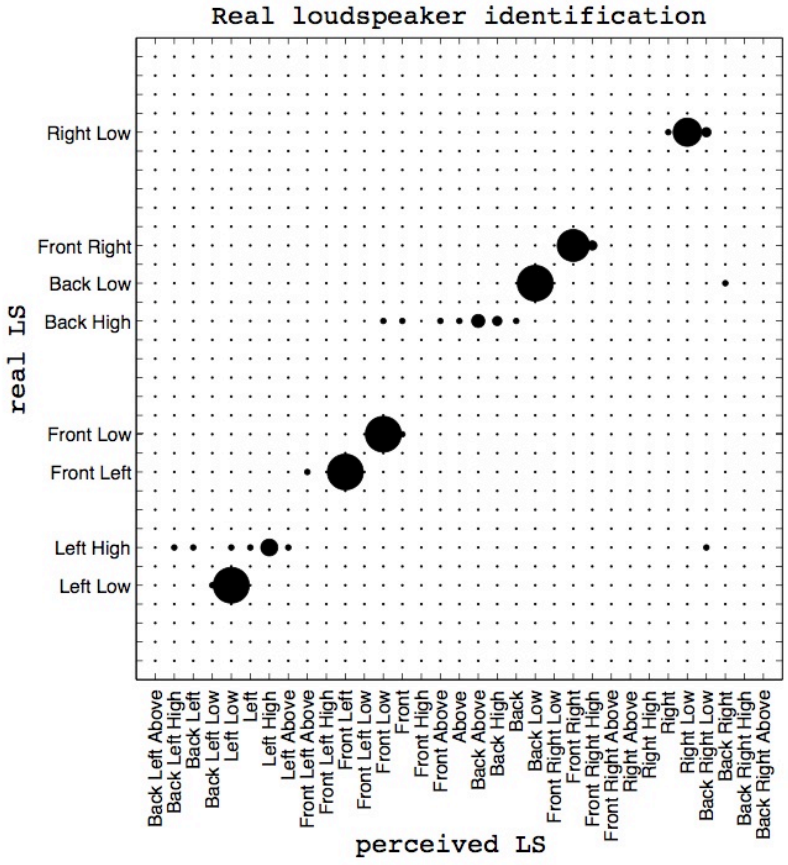


Figura 3.31 – Localizzazione delle sorgenti reali

Esaminiamo ora i risultati dei test sulla localizzazione, riportati in Figura 3.32 e in Figura 3.33. Come prima cosa si può notare che la localizzazione delle sorgenti virtuali nel settore corretto è avvenuta nel 93.23% dei casi considerando la simulazione con stereo dipolo e nell'89.05% dei casi utilizzando il sistema

Ambisonic. Come ci si aspettava non si sono riscontrati errori del tipo *Left/Right* o *Right/Left*. Altri risultati globali sono contenuti nella tabella sottostante (Tabella 3.5). Se si guarda al risultato della confusione *left/median*, esso sembra confermare, nel caso dello stereo dipolo, un'alta difficoltà nella localizzazione nell'emisfero sinistro. Nel caso Ambisonic questa peculiarità non è stata evidenziata: la quota di confusione *right/median* è molto simile a quella *left/median*, indicando una generalizzata (ma non dominante) tendenza a commettere errori di azimut.

Tipo di confusione	**Stereo dipolo**	**Ambisonic**
right/median (sector 9)	0.9%	2.9%
median/right (sector 6)	0.9%	0.4%
median/left (sector 7)	0%	0%
left/median (sector 4)	5.74%	2.85%

Tabella 3.5 – Risultati globali del test

Un altro risultato che può essere utile nell'evidenziare una polarizzazione nell'emisfero sinistro è contenuto in Figura 3.34, dove è stato graficato il numero di valutazioni esatte per le posizioni nei settori 1, 2 e 3. Si può notare come le sorgenti a sinistra tendano ad essere meglio localizzate. Si può quindi concludere che non ci siano effettivi problemi di polarizzazione artificiale nell'emisfero sinistro e che lo strano risultato nella localizzazione *left/high* per sorgenti reali sia dovuto al ristretto numero di soggetti (10) che hanno ascoltato quel particolare stimolo sonoro, oltre che alla ben nota difficoltà nel discriminare sorgenti non azimutali.

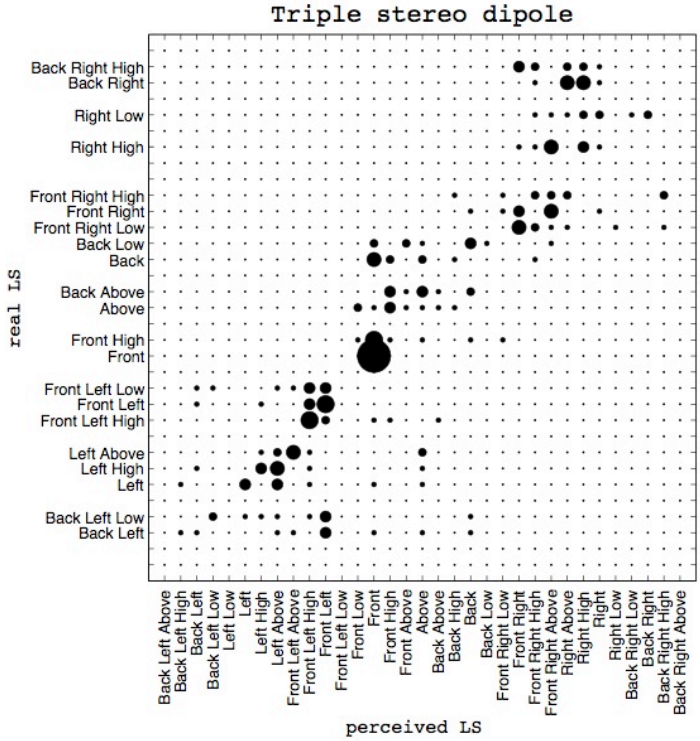


Figura 3.32 – Localizzazione Stereo Dipolo

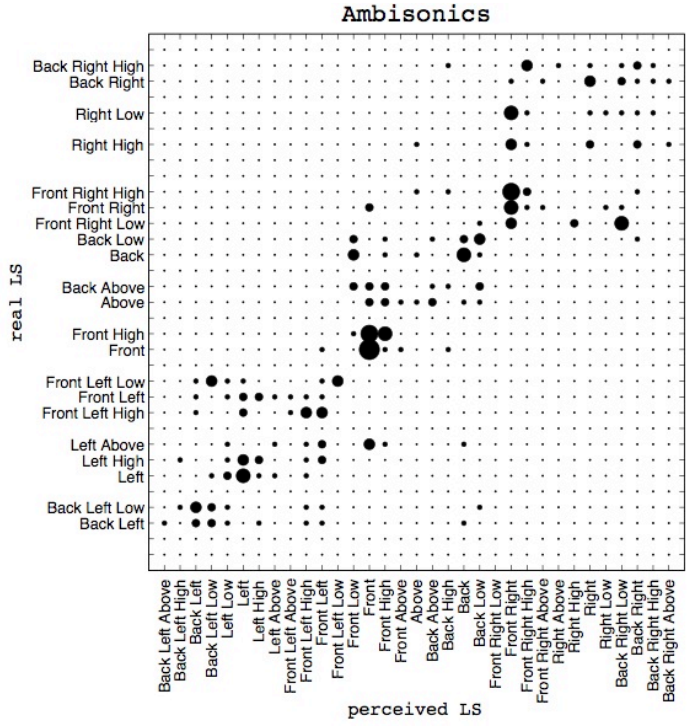


Figura 3.33 – Localizzazione Ambisonic

Nella Figura 3.35 sono state graficate le confusioni *up/down* e *down/up* per i due sistemi. Nella Figura 3.36 vengono mostrati gli errori *back/front* e *front/back*; in Figura 3.37 è stato graficato il numero di corrette valutazioni, su 10 giudizi complessivi, per ogni step di azimut e elevazione nulla; in Figura 3.38 si ha la stessa cosa ma variando l'elevazione.

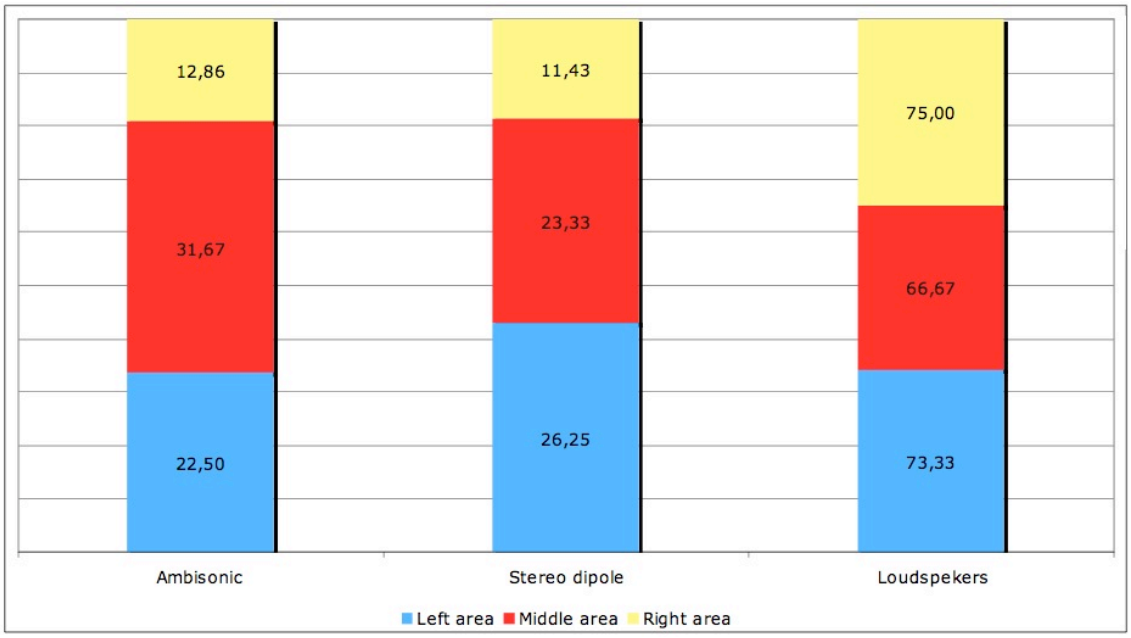


Figura 3.34 – Percentuale di posizioni correttamente localizzate per ogni settore ed ogni sistema

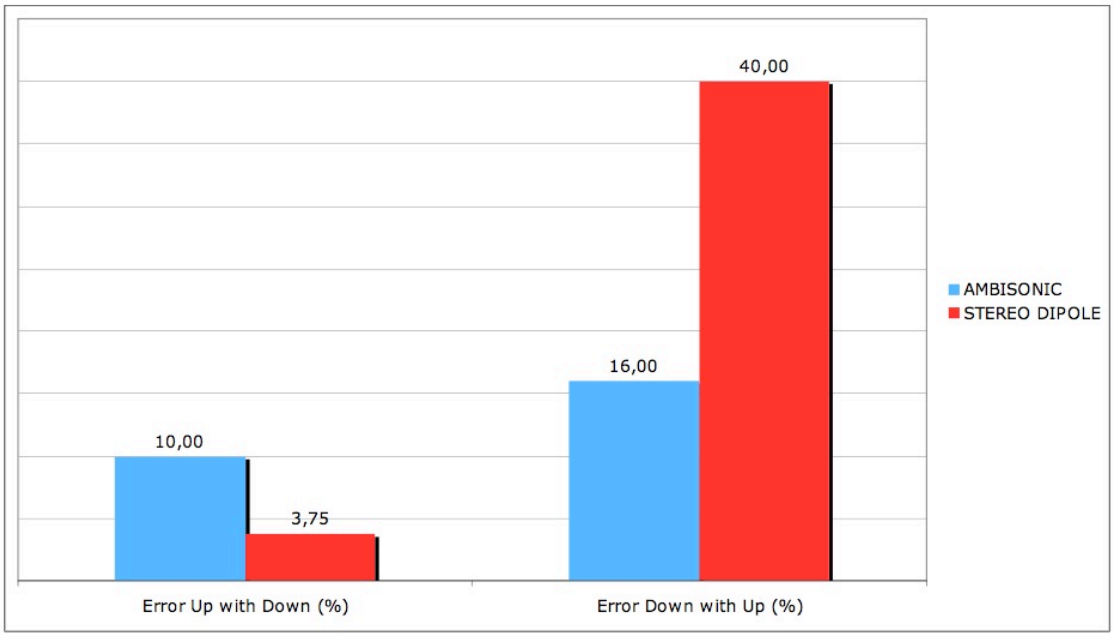


Figura 3.35 – Confusione tra l'emisfero superiore e quello inferiore

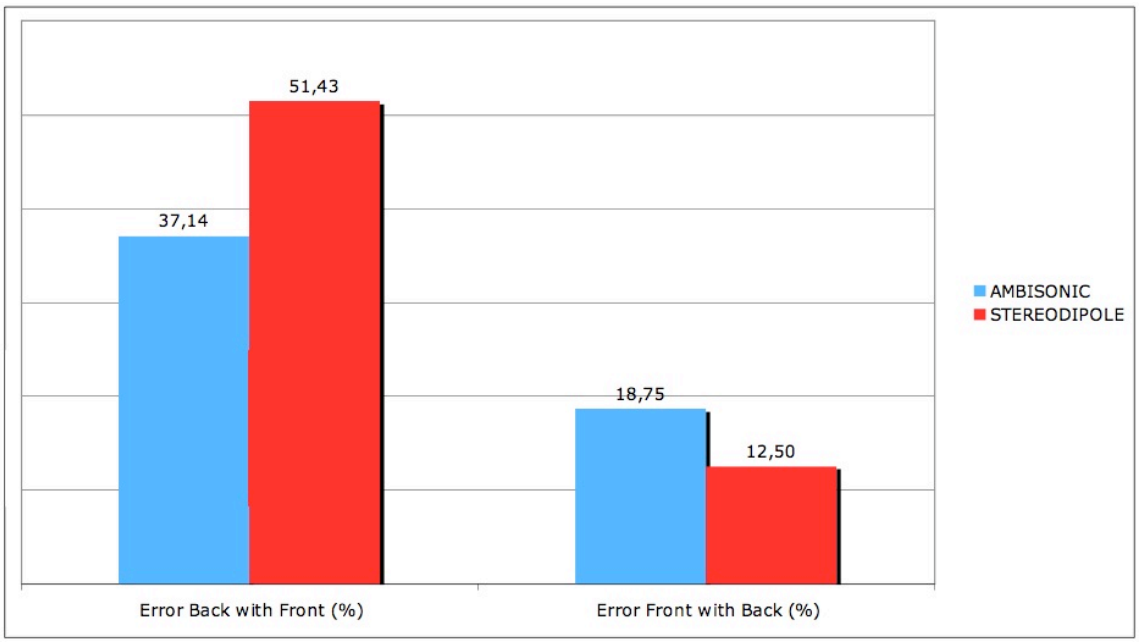


Figura 3.36 – Confusione tra l'emisfero frontale e quello posteriore

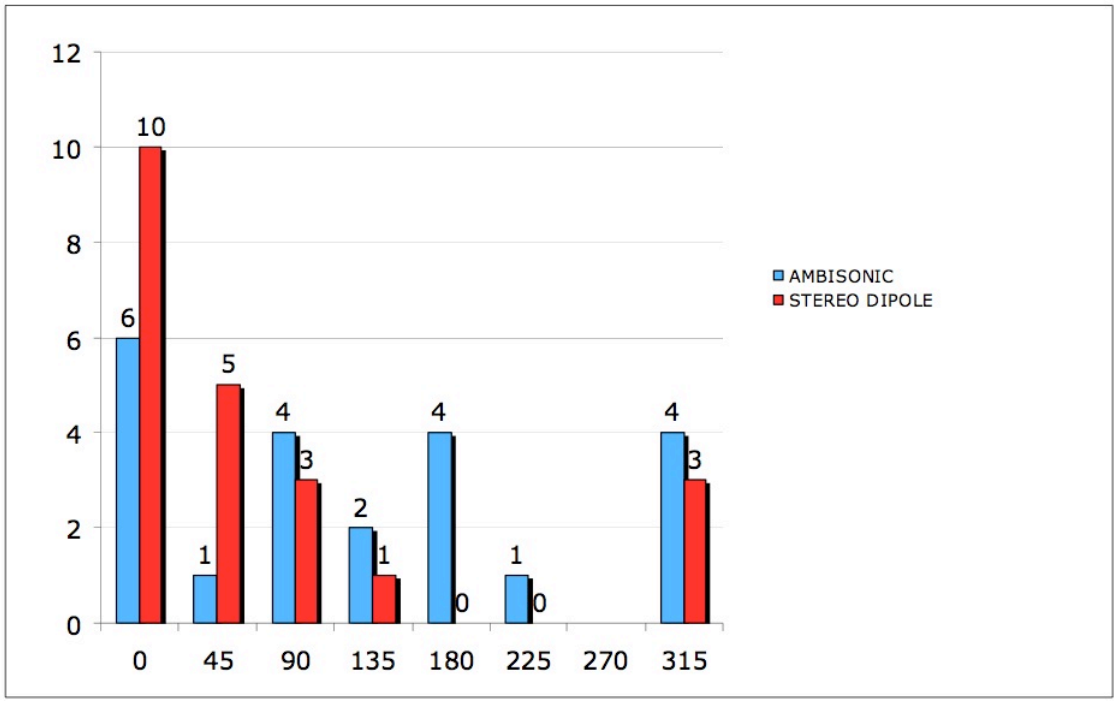


Figura 3.37 – Numero di esatte localizzazioni in funzione dell'azimut

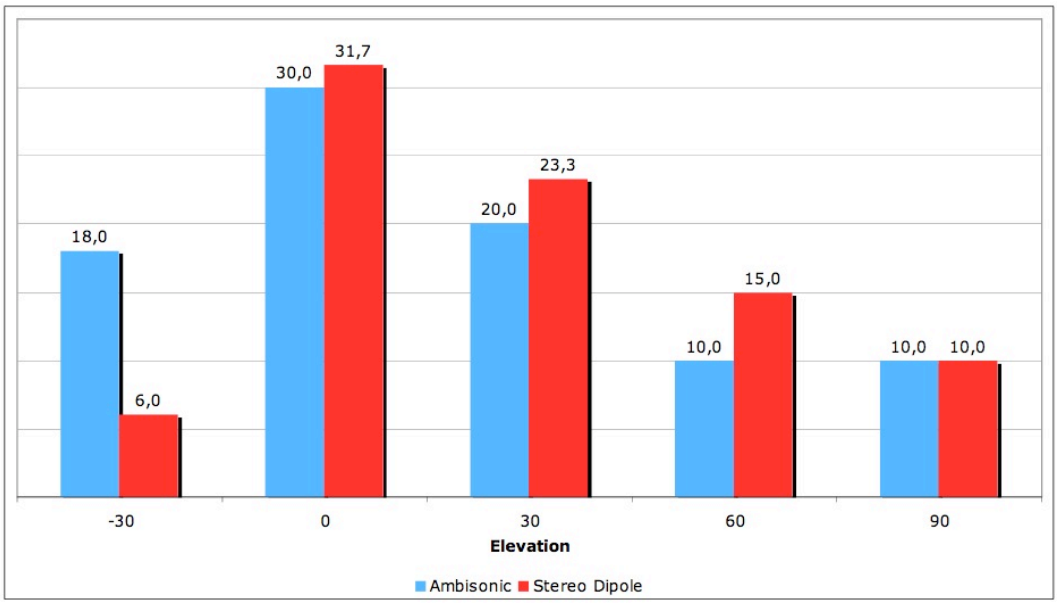


Figura 3.38 – Numero di localizzazioni esatte in funzione dell'elevazione (considerando tutti gli azimut)

3.3.3.3 Discussione dei risultati

Dai risultati riportati, la prima osservazione che si può fare è che una precisa ed accurata localizzazione, con entrambi i sistemi, non è possibile. I tre settori principali sono ben distinguibili ma purtroppo solo una media del 21% di sorgenti sono bene localizzabili (Figura 3.34).

La confusione *back/front* in Figura 3.36 (44% di media) è più alta che quella *front/back* (15% di media): questo significa che le sorgenti non ben localizzate nel settore posteriore vengono collocate frontalmente più che posteriormente. La confusione *up/down* (Figura 3.35) è in media il 6.5% e il 28% lo si ha per la confusione *up/down*: le sorgenti in basso sono spesso localizzate nell'emisfero alto.

Le sorgenti frontali in genere sono meglio localizzabili di quelle poste nella parte posteriore (Figura 3.37) e le sorgenti con elevazione nulla sono meglio identificabili che quelle con elevazione non nulla (Figura 3.38).

Il motivo per cui si sono ottenute prestazioni non così buone come ci si aspettava è certamente dovuto ai due sistemi ma è importante sottolineare come le sorgenti posteriori e non azimutali sono in generale più difficili da localizzare anche nella vita reale, ne è una prova il risultato della localizzazione degli speaker all'interno del test. In più, i risultati che sono stati riportati non consentono nessuna tolleranza e una sorgente è considerata "esattamente localizzata" solo se è stata correttamente

posizionata sulla griglia: ai soggetti era stato chiesto di forzare il loro giudizio indicando una delle posizioni della griglia, senza nessuna possibilità di scegliere un punto intermedio tra due possibili posizioni della griglia. Se si permettesse una certa tolleranza sicuramente si otterrebbero risultati migliori.

Facendo il confronto dei due sistemi possiamo osservare che le sorgenti basse sono male localizzate usando lo stereo dipolo, che tende a far percepire con un'elevazione maggiore di quella reale (40% di confusione *down/up* in confronto al 16% ottenuto con il sistema Ambisonic). questo risultato può essere almeno in parte giustificato dalla presenza dello stereo dipolo esattamente sopra la testa dell'ascoltatore. Anche nel caso di confusione *back/front* lo stereo dipolo presenta performance peggiori (51% di errore in confronto al 36% dell'Ambisonic).

Gli errori *up/down* e *front/back* sono pochi per entrambi i sistemi con migliori, anche se di poco, prestazioni dello stereo dipolo che permette una migliore localizzazione delle sorgenti frontali. Inoltre lo stereo dipolo presenta risultati migliori nella percezione dell'elevazione, fatta eccezione per l'emisfero inferiore, dove la sua affidabilità crolla drasticamente.

Da questi dati è alquanto difficile rispondere se sia meglio affidarsi, per avere una buona localizzazione, ad una riproduzione mediante stereo dipolo o sistema ambisonic. Confrontando i risultati ottenuti con quelli in [8] si osserva come, nel caso dell'Ambisonic, i risultati siano migliori con il nostro sistema, mentre è l'opposto nel caso dello stereo dipolo. Questa differenza può essere dovuta all'implementazione dei sistemi ma soprattutto al fatto che in questa sessione di test i soggetti non erano costretti a localizzare posizioni solo sul piano azimutale. Se restringessimo l'analisi al solo piano azimutale lo stereo dipolo presenterebbe migliori prestazioni, soprattutto nella zona frontale.

4 Conclusioni

Con questo lavoro di ricerca si è cercato di individuare la strada migliore per effettuare test d'ascolto affidabili e realistici, che permettano di valutare l'acustica di teatri o sale da concerto confrontandole, in tempo reale, tra loro. Per questo l'attenzione si è focalizzata sugli ascoltatori, valutandone le capacità uditive, e sui sistemi di riascolto tridimensionale, valutandone realismo e fedeltà di ricostruzione del campo sonoro.

La ricerca condotta sugli ascoltatori che abitualmente frequentano il Teatro Regio e l'Auditorium Paganini, entrambe strutture di Parma, ha evidenziato come la maggior parte di essi sia di età superiore ai 40 anni: con l'aumento dell'età, per il naturale invecchiamento dell'apparato uditivo, la capacità uditiva si riduce, soprattutto alle alte frequenze. Tale fenomeno è stato evidenziato con esami audiometrici sui soggetti testati ed è confermato da numerosi studi in ambito medico. Questa ricerca ha dimostrato la necessità, nella fase progettuale di spazi destinati alla musica, di tenere conto del deficit uditivo "naturale" che falsa, in parte, la percezione del suono.

Per quanto riguarda i sistemi di riascolto, in questa tesi si è concentrata l'attenzione soprattutto su *stereo dipolo* e *sistema ambisonic*. Essi si basano su principi differenti ma puntano entrambi alla ricostruzione, nel punto d'ascolto, della corretta percezione sonora "trasportando" virtualmente l'ascoltatore nello spazio desiderato.

Allo stato attuale è difficile stabilire quale dei due sistemi risulti il migliore perchè gli studi effettuati hanno evidenziato limiti e pregi di entrambi. Sono stati valutati realismo e performance nel fornire la corretta localizzazione delle sorgenti, sottoponendo gli ascoltatori a test d'ascolto mirati. Per continuare in modo proficuo questi studi è in fase di realizzazione, presso la Casa del Suono (Parma), una stanza che permetta l'ascolto ottimizzato di entrambi i sistemi e ne consenta la fruizione ad un largo numero di utenti.

APPENDICI

A.1 – Realizzazione di una sorgente omnidirezionale

Per la registrazione di risposte all'impulso da utilizzare poi in fase di test è fondamentale l'utilizzo di una buona sorgente omnidirezionale, che presenti un spettro in frequenza lineare. Quella presentata in queste righe è stata realizzata con l'Ing. Marcello Bia, di cui ero correlatore di laurea.

Si è scelto di creare una cassa a due vie con un *subwoofer* per le basse frequenze e un *dodecaedro* per le frequenze medio alte. In questo modo si è cercato di coprire tutte le frequenze con un livello di potenza soddisfacente. Essendo il dodecaedro un'unità già assemblata e funzionante, il problema di progetto fa riferimento al subwoofer e alla banda di frequenze di sua competenza. Si è pensato di montare il subwoofer in una cassa in configurazione passa banda a carico simmetrico. Il carico simmetrico si ha quando le frequenze di risonanza dei due volumi che contengono l'altoparlante sono uguali e il cono del woofer subisce un carico uguale su entrambi i lati.

Figura A.1.1 - Dodecaedro "omnidirezionale" e cassa in PVC in cui è visibile il woofer.

Dopo una prima fase di progettazione mediante software, per la realizzazione della cassa è stato scelto un tubo in PVC del diametro di 39 cm e altezza 96 cm. Esso è stato diviso in due alloggiando il woofer nel piano di divisione ed inserendo un reflex ottimizzato in lunghezza e diametro. Il cross-over è stato affidato al processing di una scheda DSP in cui sono stati programmati due filtri IIR, un passa-basso ed un

passa-alto, con frequenza di taglio attorno ai 100 Hz. Il segnale così trattato viene passato ad un amplificatore (Powermod 450W della Powersoft), anch'esso alloggiato nella cassa e alle sue uscite sono collegati woofer e dodecaedro.

Il sistema così costruito però presentava uno sbilanciamento di livelli tra basse e alte frequenza, nonché un andamento poco lineare in frequenza. Per equalizzare il sistema si è proceduto secondo la tecnica intensimetrica indicata nella norma ISO 9614: partendo dal calcolo dell'intensità sonora che fluisce attraverso una superficie nota, si ottiene la potenza della sorgente (formula sottostante).

$$P = I \cdot S$$

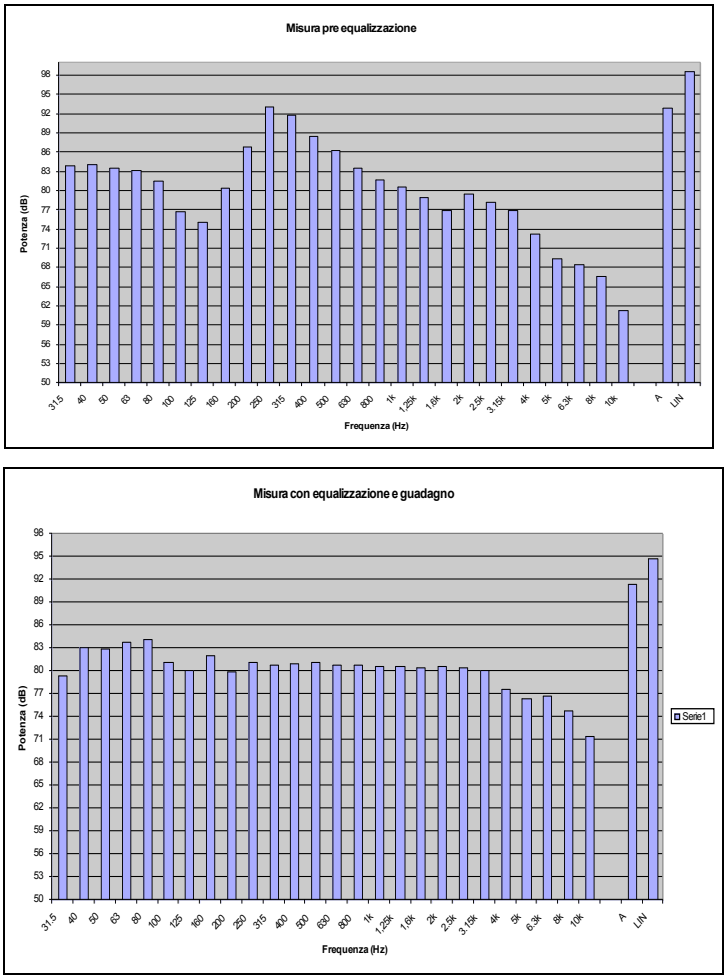


Figura A.1.2 - Prima e dopo l'equalizzazione

Figura A.1.3 - Aspetto finale dell'intero sistema

Come si può vedere dall'immagine l'oggetto realizzato ha dimensioni considerevoli: purtroppo, la scelta di richiudere, in caso di trasporto, il dodecaedro in una delle due camere del subwoofer ha portato all'incremento del diametro e dell'altezza. La linearità della risposta è molto buona ma la potenza ottenuta (96 dB lineari) si è rivelata troppo bassa in relazione alle dimensioni.

Il sistema è stato utilizzato a Roma, all'interno dell'Auditorium Parco della Musica (Sala Santa Cecilia) per la taratura di un modello CAD, utile poi per la simulazione acustica di tale ambiente.

A.2 – Microfono sferico

Si vuole descrivere in queste pagine il lavoro svolto per realizzare un microfono che "catturi" le armoniche sferiche del 4° ordine. Primo problema: quante capsule utilizzare? Dalla letteratura si evince una regola generale:

$$(L+1)^2 \leq N$$

dove L rappresenta l'ordine di troncamento delle funzioni di Fourier-Bessel mentre N è il numero di capsule necessarie. Volendo così arrivare fino al quarto ordine occorrono almeno 32 capsule.

Componenti del sistema microfonico

Capsule

Le capsule scelte sono delle Knowles Electronics FG-23329-P07. Presentano tre fili più il "terminale-calza" e richiedono un'alimentazione di circa 1.5volt. Le dimensioni della capsule sono di 2.5mm mentre la sezione del cavo è di 2mm.

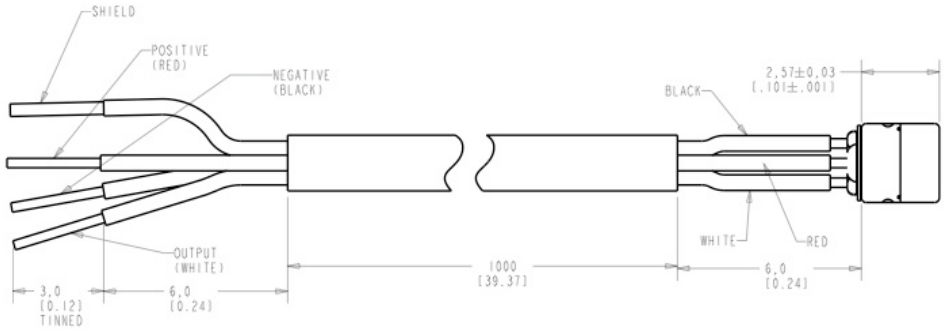


Figura A.2.1 – schema delle capsule microfoniche

Alimentatore

Le 32 capsule necessitano di alimentazione ad 1.5volt. che viene loro fornita attraverso un alimentatore autocostruito funzionante a batteria, con tasto di accensione e LED indicante il suo funzionamento.

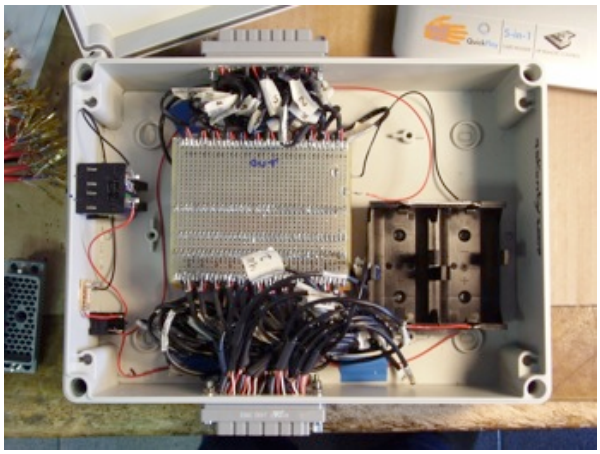


Figura A.2.2 – Fase di montaggio dell'alimentatore

Connessioni

Per trasportare il segnale dal microfono all'alimentatore e dall'alimentatore ai convertitori è stato usato un cavo multicore a 32 coppie schermate (Eurocable) CV LKSSS32C. La scelta per i connettori è caduta su Edac a 90 poli e 56 poli. In figura sono riportati gli schemi delle saldature.

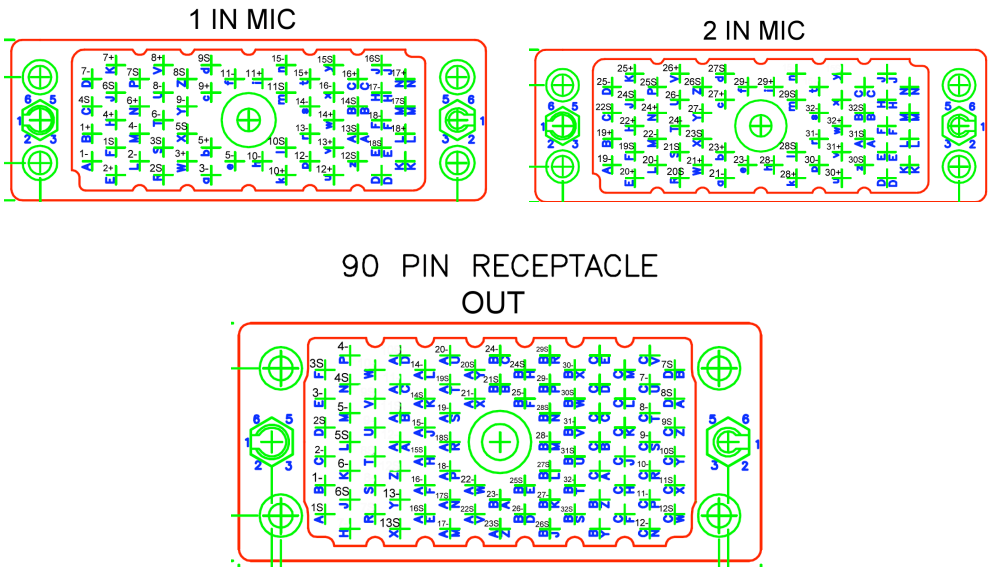


Figura A.2.3 – Schemi dei connettori

Progettazione e simulazione

Per posizionare correttamente le capsule nello spazio è stata realizzata una sfera in poliuretano espanso di diametro pari a 7cm. I 32 microfoni sono disposti sui vertici

di un icosaedro e un dodecaedro inscritti nella sfera, con le seguenti coordinate cartesiane:

Capsula n.	X	Y	Z
1	0.8507	0.5257	0
2	0.5774	0.5774	0.5774
3	0.5774	0.5774	-0.5774
4	0.3568	0.9342	0
5	0	0.3658	-0.9342
6	0	0.8507	-0.5257
7	-0.5257	0	-0.8507
8	-0.5774	0.5774	-0.5774
9	-0.3568	0.9342	0
10	-0.8507	0.5257	0
11	-0.9342	0	-0.3658
12	-0.9342	0	0.3658
13	-0.5257	0	0.8507
14	-0.5774	0.5774	0.5774
15	0	0.8507	0.5257
16	-0.3568	-0.9342	0
17	0.3568	-0.9342	0
18	0.8507	-0.5257	0
19	0	0.3658	0.9342
20	0.5774	-0.5774	0.5774
21	0	0.8507	0.5257
22	0	-0.8507	-0.5257
23	0.5774	-0.5774	-0.5774
24	0	-0.3658	-0.9342
25	0.5257	0	-0.8507
26	0.9342	0	-0.3658
27	0.9342	0	0.3658
28	0.5257	0	0.8507
29	0	-0.3658	0.9342
30	-0.5774	-0.5774	0.5774
31	-0.8507	-0.5257	0
32	-0.5774	-0.5774	-0.5774

Figura A.2.4 – Coordinate delle capsule sulla superficie sferica

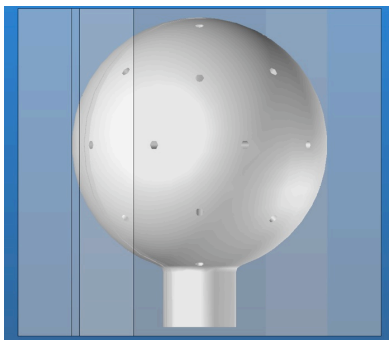


Figura A.2.5 – Disegno su Top Solid della sfera

Il raggio è un buon compromesso tra la riduzione dell'aliasing spaziale alle alte frequenze e l'affidabilità del campionamento delle basse frequenze per ordini elevati.

L'aliasing spaziale si verifica quando ho un sottocampionamento spaziale dell'onda sonora. Crea problemi nel capire da che direzione arriva l'onda. Per questo l'onda dev'essere campionata in almeno 2 punti per lunghezza d'onda; se ciò non avviene la direzione di provenienza dell'onda diventa ambigua. Da un punto di vista matematico si verifica aliasing se:

$$\frac{v}{2f} = \Delta x$$

Dove *v* rappresenta la velocità dell'onda, *f* la frequenza e *Δx* la distanza tra due punti di campionamento successivi.

In via del tutto teorica, sapendo la distanza di due capsule adiacenti, è possibile sapere la a che frequenza interviene l'aliasing. Nel nostro caso conosciamo l'angolo tra due capsule poste sui vertici del dodecaedro (0.73rad) e l'angolo tra la capsula sul dodecaedro e la capsula sull'icosaedro (0.65rad) [5].

Per calcolare la frequenza a cui potrebbe intervenire l'aliasing occorre calcolare la distanza tra le capsule *l* :

$$l = r * \alpha^{rad}$$

dove *r* rappresenta il raggio della sfera. Si ha quindi:

$$l_1 = 0.035 * 0.73 = 0.0255m$$

$$l_2 = 0.035 * 0.65 = 0.0227m$$

Applicando la (1) si ottengono le frequenze a cui, teoricamente, interviene aliasing spaziale:

$$f_{c1} = \frac{342}{2 * 0.255} = 6692Hz$$

$$f_{c1} = \frac{342}{2 * 0.227} = 7516Hz$$

La previsione risulterebbe senz'altro azzeccata nel caso di un array lineare. In questo caso, avendo le capsule microfoniche disposte in 3D attorno ad un centro, non è detto che la previsione risulti affidabile in modo così preciso. E' possibile comunque pensare, in via del tutto ottimistica, che l'aliasing spaziale possa intervenire per frequenze tra i 6000 Hz e i 7000 Hz.

Nell'articolo dello staff di France Telecom [2] viene suggerita una semplice regola per capire a quale ordine occorre arrivare per avere una descrizione accurata del campo (errore non superiore al 4%), tenendo presente il raggio della sfera su cui si sta campionando e la frequenza dell'onda sonora:

$$M = \lceil K * R \rceil$$

Dove K=2πf/c è il numero d'onda mentre R rappresenta il raggio della sfera. A questo punto è facile vedere come, supponendo un'onda piana che incida sulla sfera di raggio 3.5cm, con una frequenza di 1KHz basti un primo ordine per una descrizione accurata, mentre per un'onda con frequenza pari a 4 kHz serve almeno un 3° ordine per una corretta rappresentazione.

Freq. (Hz)	**500**	**1000**	**2000**	**4000**	**6000**	**8000**	**10000**	**16000**
Ordine	1	1	2	3	4	5	6	10

Grazie al contributo dell'Ing. Filippo Fazi (ISVR of Southampton) si sono effettuate simulazioni per capire l'influenza delle *funzioni di Bessel* sull'andamento del campo sonoro ad 1m di distanza dal centro del microfono, come mostrato in figura sottostante:

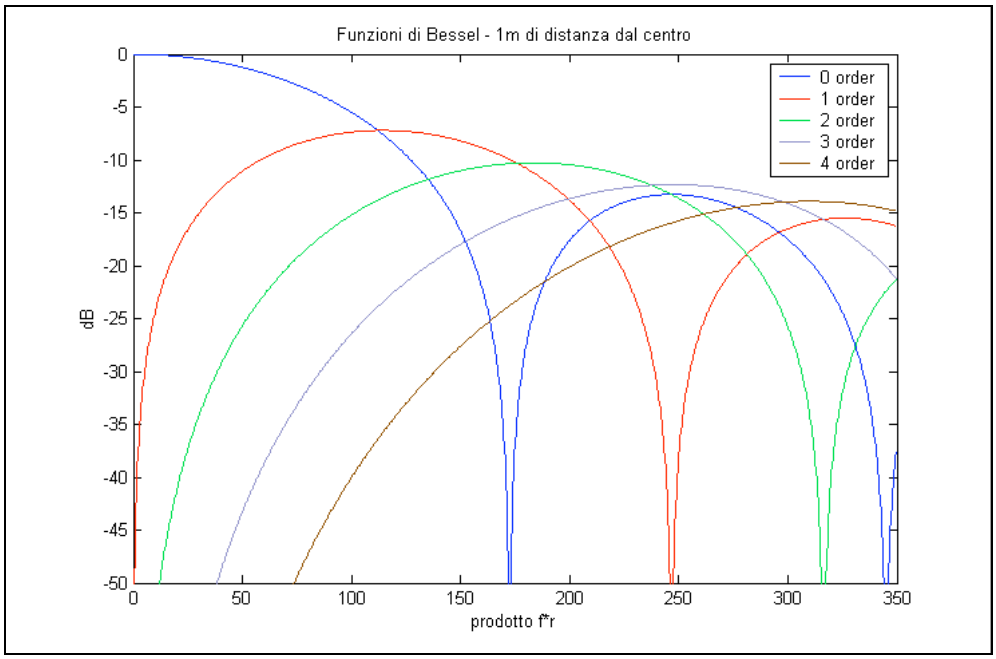


Figura A.2.6 – Funzioni di Bessel ad 1 m dal centro della sfera

Si può osservare come per una frequenza di 180Hz siano equivalenti gli apporti del primo e secondo ordine, mentre il contributo dell'ordine 0 è praticamente nullo. Le funzioni di Bessel (in valore assoluto), in dipendenza dal prodotto f*r (dove r è il raggio del microfono), sono rappresentate nella figura sottostante.

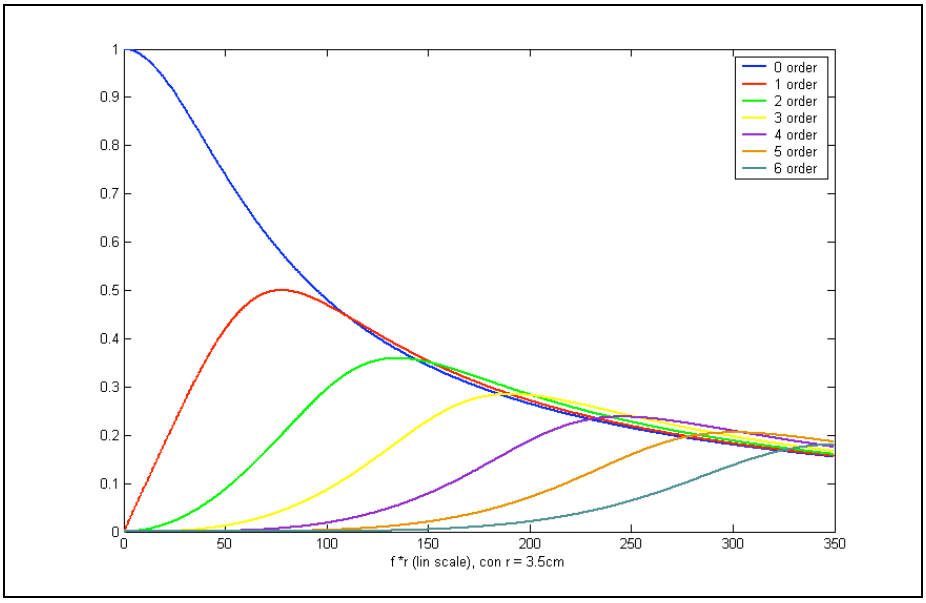


*Figura A.2.7 – Funzioni di Bessel in relazione al prodotto f*r*

Infine, le stesse funzioni di Bessel sono state rappresentate in scala logaritmica:

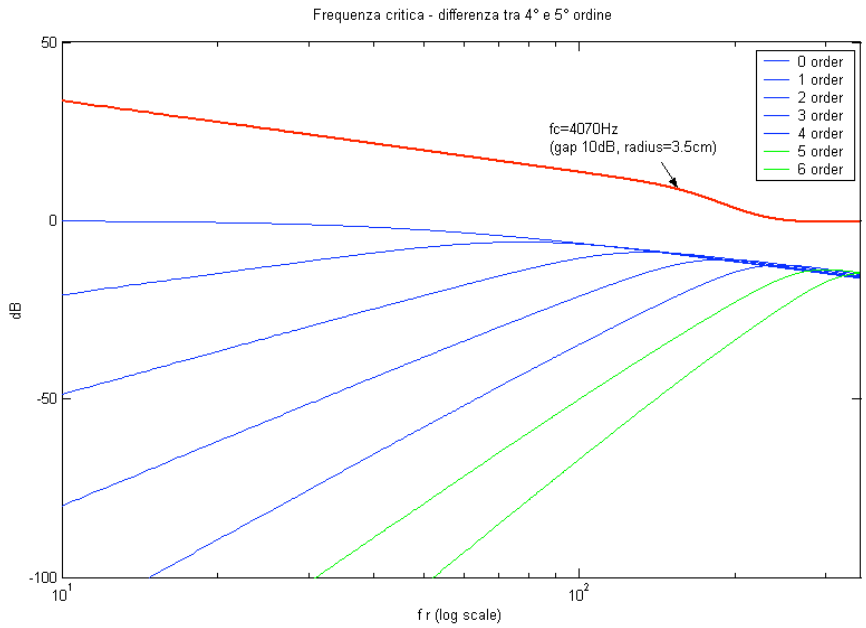


Figura A.2.8 –Funzioni di Bessel in scala logaritmica

In teoria, per evitare l'aliasing spaziale tra il 5° ordine e il 4° ordine, converrebbe massimizzare la differenza tra le due funzioni radiali: più il livello in decibel delle armoniche sferiche del 5° ordine è basso rispetto a quelle del 4° ordine, meno influiranno su di esse. Ma, come si può vedere dalla figura precedente, il massimo lo si ha per un rapporto *f*r* molto basso; prendendo come compromesso un gap di 10dB fra i due ordini si può osservare come, per un raggio di 3.5cm, la frequenza critica sia a circa 4070Hz.

In fase di filtraggio sarà quindi conveniente effettuare uno *smoothing* per tagliare le componenti del 4° ordine sopra i 4000Hz.

Realizzazione

La sfera è stata divisa in due parti in modo da agevolare l'inserimento delle capsule. Le due metà sono state quindi incollate tramite silicone (le colle contenenti diluente rischiano di "mangiare" l'espanso).

I microfoni che possono avere performance peggiori sono quelli in prossimità del supporto della sfera. Tale supporto ha un diametro di 1.5cm, necessario per far passare il fascio di fili delle capsule avente diametro di circa 1.1cm.

Figura A.2.9 – Inserimento delle capsule microfoniche nelle due semi-sfere

La sfera microfonica ha bisogno di un cavalletto, innestabile al piccolo supporto. Così si è pensato di saldare direttamente i fili delle capsule ad uno spezzone di 2m di cavo "multi-core", tenuto ancorato ad un tubo di alluminio. Tale tubo è stato scelto della lunghezza di 60 cm, diametro interno di 2.1cm e filettatura da un lato (1.6) per innestarlo su un'asta microfonica standard.

Routing del segnale

I segnali provenienti dai 32 microfoni vengono convertiti in digitale da 4 convertitori ADA 8000 della Behringer (48 kHz, 24 bit); essi, tramite il protocollo ADAT, comunicano con la scheda audio M-Audio Profire ed i segnali digitali vengono così gestiti da un laptop.

Considerazioni

Il microfono è perfettamente funzionante ma, per creare i filtri inversi che trasformino i 32 segnali in segnali Ambisonic, sono necessarie misure in camera anecoica. Purtroppo non è facile né economico trovare una camera delle dimensioni adatte e fornita di strutture mobili in grado di muovere il microfono con assoluta precisione: per questo motivo il microfono, realizzato nella seconda metà del 2007,

non è stato ancora utilizzato. Altro problema sono le dimensioni del cavo multi-core che rendono l'intero sistema non facilmente trasportabile.

Figura A.2.10 – Microfono a 32 capsule

A.3 – Questionario distribuito ai frequentatori del Teatro Regio e dell'Auditorium Paganini di Parma

COMUNE DI PARMA

UNIVERSITÀ DEGLI STUDI DI PARMA
FACOLTA' DI MEDICINA E CHIRURGIA
FACOLTA' DI INGEGNERIA

PER UN BUON ASCOLTO DELLA MUSICA A TEATRO

Chiediamo gentilmente il suo contributo per migliorare, ove possibile, l'ascolto della musica. La preghiamo di compilare questo modulo e di imbucarlo nell'apposito raccoglitore all'uscita del teatro.
La ringraziamo anticipatamente per la sua collaborazione.

1) Età: ☐20-30 ☐31-40 ☐41-50 ☐51-60 ☐61-70 ☐71-80 ☐81-90 ☐oltre

2) Sesso: ☐M ☐F

3) Istruzione: ☐lic. elementare ☐lic. media ☐diploma ☐laurea

4) Ha compiuto studi musicali? ☐Si ☐No

5) Quanti concerti all'anno va ad ascoltare? ☐1-3 ☐4-7 ☐8-15 ☐oltre

6) Quante opere liriche all'anno? ☐1-3 ☐4-7 ☐8-15 ☐oltre

7) Quanti balletti/operette/varietà? ☐1-3 ☐4-7 ☐8-15 ☐oltre

8) Quanti spettacoli di prosa/recital/cabaret? ☐1-3 ☐4-7 ☐8-15 ☐oltre

9) Lei acquista preferibilmente: ☐Abbonamento ☐Biglietti

10) Come giudica la sua capacità uditiva?
☐Normale ☐Lievemente compromessa ☐Fortemente compromessa

Segue sul retro…

11) Il Teatro Regio mette a disposizione speciali cuffie per migliorare l'ascolto della musica per persone con difficoltà uditive.
   a. Ne è a conoscenza? ☐ Sì ☐ No
   b. Le ha mai utilizzate? ☐ Sì ☐ No
   c. Se sì, con quale risultato?

☐ Insufficiente ☐ Sufficiente ☐ Buono ☐ Ottimo

12) Sa che esistono dispositivi che possono amplificare selettivamente alcune frequenze, risolvendo problemi uditivi in modo mirato? ☐ Sì ☐ No

Stiamo svolgendo uno studio presso la Casa della Musica che prevede un **test d'ascolto volto a migliorare la percezione soggettiva della musica**. Le interesserebbe partecipare? Se sì, compili anche il riquadro sottostante e sarà contattato per fissare un appuntamento.

Cognome_____Nome_____

Indirizzo_____Tel._____

Informativa legge sulla privacy (D. Lgs. 196 del 30/06/03).
Si garantisce la massima riservatezza nel trattamento dei dati, che verranno utilizzati esclusivamente per le finalità dichiarate e non verranno comunicati a terzi. Responsabile del trattamento dei dati è D. Marmiroli, che è possibile contattare al seguente indirizzo e-mail: labellamusica@libero.it

Dichiarazione di consenso.
Il sottoscritto presta il suo consenso al trattamento dei propri dati limitatamente ai fini e alle modalità sopra esposte.

Data_____ Firma_____

*Si ringrazia la **Fondazione Teatro Regio** di Parma per la cortese collaborazione.*

BIBLIOGRAFIA

[1] B. Wiggins "*An investigation into the real-time manipulation and the control of three-dimesional sound fields*", University of Derby, 2004.

[2] M.Hartmann, "*How we localize sounds*", Phys. Today n.52, 1999, pp.24-29.

[3] M. Sacco, "*Imparare la tecnica del suono*", Lambda Edizioni, pp. 37-39.

[4] B. Wiggins "*An investigation into the real-time manipulation and the control of three-dimesional sound fields*", University of Derby, 2004.

[5] S.Moreau, J.Daniel, S.Bertet, "*3D Sound Field Recording with Higher Oreder Ambisonics - Objective Measurements and Validation of a 4th Order Spherical Microphone*", 120th AES Convention, Paris (France), 20-23 May, 2006

[6] J. Daniel, "*Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable New Ambisonic Format*".

[7] J.Daniel, R.Nicol, S.Moreau, "*Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*", 114th AES Convention, Amsterdam (The Netherlands), 22-25 March 2003.

[8] E.Benjamin, R.Lee, A.Heller, "*Localization in Horizontal-Only Ambisonic Systems*", 121st AES Convention, 2006 October 5-8, San Francisco, CA, USA.

[9] M. A. Gerzon, "*General Metatheory of Auditory Localisation*", 92nd AES Convention, 1992 March 24-27, Vienna.

[10] M. A. Gerzon, "*Psychoacoustic Decoders for Multispeaker Stereo and Surround Sound*", 93rd AES Convention, 1992 October 1-4, San Francisco, CA, USA.

[11] Fons Adriaensen, "*Near Field filters for Higher Order Ambisonics*", www.kokkinizita.net/linuxaudio

[12] A.Farina, R.Glasgal, E.Armelloni, A. Torger, "*Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music*", 19th AES Conference.

[13] D. Malham, "*The Role of The Single Point Soundfield Microphone In Surround Sound Systems*", Microphones & Loudspeakers, AES UK Conference.

[14] A. Laborie, R. Bruno, S. Montoya, "*A new Comprehensive Approach of Surround Sound Recording*", 114th AES Convention, 2003 March 22-25, Amsterdam, The Netherlands.

[15] Z. Li, R. Duraiswami, N.A. Gumerov, "*Capture and recreation of higher order 3D sound fields via reciprocity*"

[16] B. B. Bauer, "*Stereophonic earphones and binaural loudspeakers*", J.Audio Eng.Soc., vol.9, pp 148-151,1961

[17] M. R. Schroeder and B. S. Atal, "*Computer simulation of sound trasmission in rooms*". IEEE Int. Conv. Rec. 7, pp 150-155, 1963

[18] M. R. Schroeder, D. Gottlob and G. F. Sierbrasse, "*Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters*", Journal of the Acoustical Society of America, vol. 56, no. 4,

pp. 1195-1201, 1974

[19] O. Kirkeby, P. A. Nelson and H. Hamada, "*The 'stereo dipole' – a virtual source imaging system using two closely spaced loudspeakers*", J. Audio Emg. Soc., vol 46, no. 5, pp. 387-395

[20] S. T. Neely and J. B. Allen, "*Invertibility of a room impulse response*", J.A.S.A., vol. 66, pp.165-169, 1979

[21] J. N. Mourjpoulos, "*Digital equalization of room acoustics*", JAES vol. 42, n. 11, 1994 November, pp.884-900

[22] Cammi A., "*Progetto e sviluppo di una sala d'ascolto di piccole dimensioni*", tesi di laurea anno accademico 2005-06, Università degli Studi di Parma, Facoltà di Ingegneria.

[23] Cigolani S., Spagnolo R., "*Acustica musicale e Architettonica*", UTET Libreria (2001).

[24] Kuttruff H., "*Room acoustics*", Fourth edition (1996).

[25] Farina A., "*L'acustica dei piccoli ambienti d'ascolto*", AES Italian Section, Milano (2001)

[26] Farina A., Cibelli G., Bellini A., "*AQT – A new objective mesaurement of the Acoustical Quality of sound reproduction in small compartments*", AES Amsterdam (2001)

[27] Software Adobe Audition, http://www.adobe.com

[28] Farina A., "*Aurora*", modulo Plug-in per Adobe Audition

[29] Azzali A., "*AQT Analisys*", modulo Plug-in per Adobe Audition

[30] Azzali A. "*Sviluppo e implementazione di modelli psicoacustici per l'analisi della qualità audio e della comunicazione in ambienti ridotti*", Tesi di Dottorato, Corso di Dottorato di Ricerca in Ingegneria Industriale, Università di Parma, XIII Ciclo.

[31] Capra A. "*Caratterzzazione quantitativa e valutazione attraverso test soggettivi di teatri e sale da concerto*", Tesi di Laurea, Relatore A.Farina, Univeristà di Parma, Anno Accademico 2002-2003

[32] A. Farina, R. Ayalon, "*Recording concert hall acoustics for posterity*", 24th International Audio Eng. Soc. Conf. on Multichannel Audio, Banff, Canada, paper no. 38, 2003

[33] International Organization for Standardization, *ISO 3382 (1997)*, Acoustics-Measurement of reverberation time of rooms with reference to other acoustical parameters

[34] D. Cabrera, D. Gilfillan, "*Auditory distance perception of speech in the presence of noise*", Proc. Int. Conf. on Auditory Display, Kyoto, Japan, pp. 431-439, 2002.

[35] D. Cabrera, D. Jeong, H. J. Kwak and J.-Y. Kim, "*Auditory room size perception for measured and modeled rooms*", Internoise, Rio de Janiero, 2005.

[36] A. Farina, L. Tronchin, "*Acoustic qualities of theatres: correlation between experimental measures and subjective evaluations*", Applied Acoustics 62 (2001) 889-912.

[37] A. Azzali, D. Cabrera, A. Capra, A. Farina, P. Martignon, "*Reproduction of Auditorium Spatial Impression with Binaural and Stereophonic Sound Systems*", 118th AES Convention, Barcelona, 28- 31 May 2005.

[38] S. G. Norcross, G. A. Soulodre, M.C. Lavoie, "*Evaluation of Inverse Filtering Techniques for Room/Speaker Equalzation*", 113th AES Convention, Los Angeles, 5-8 October 2002.

[39] E.Alajmo – "Otorinolaringoiatria" - Piccin Nuova Libreria, Padova, p.19

[40] Takashi Takeuchi, Philip A Nelson, and Martin Teschl, "*Elevated Control Transducers for virtual acoustic imaging*" , 112th Convention 2002 May 10–13 Munich, Germany

[41] http://www.gerzonic.net

[42] http://www.plogue.com

[43] http://www.voxengo.com

[44] http://recherche.ircam.fr/equipes/salles/listen/

[45] http://ardour.org

[46] P.D Hatziantoniou, J.N Mourjopoulos "*Errors in Real-Time Room Acoustics Dereverberation*" J. Audio Eng. Soc, Vol. 52, No. 9, pp. 883 -899, 2004.

[47] H. Moller et all, "*Binaural Technique: Do We Need Individual Recordings?*", J. Audio Eng. Soc.,Vol. 44, No.6, 1996 June

[48] S. Fontana, S. Campanini, A. Farina, "*New method for auralizing the results of room acoustics simulation*", ICA 2007, Madrid.

[49] C. Varani, E. Armelloni, A. Farina "*Implementation of a double StereoDipole system on a DSP board - Experimental validation and subjective evaluation inside a car cockpit*", 115th AES Convention, New York, 10-13 October 2003

Ringraziamenti

Se sono arrivato fino a qui in gran parte lo devo ad Angelo Farina: mi ha offerto la possibilità di approfondire l'acustica, un campo così affine alla mia principale passione, la musica, da non distinguerne, spesso, i confini. Con estrema disponibilità mi ha spronato nei momenti difficili, consigliato e segnato la via nel mondo del suono.

Un grazie sentito a tutto il LAE, soprattutto a Paolo e Stefano, che hanno creduto in me. E un grazie particolare a Fons per la fiducia che mi sta dando in questo difficoltoso passaggio tra mondo accademico e mondo lavorativo.

Grazie alle persone che hanno condiviso, tutto o in parte, con me questo cammino: Paolo Martignon, un amico a cui devo tanto del lavoro svolto e dei tre anni passati assieme, della mia crescita professionale, musicale e umana; Simone Fontana che in questo ultimo anno mi ha affiancato con un impagabile buonumore, il suo suono binaurale e una marea di neologismi; Andrea Azzali, il compagno di viaggio preferito, quello che per primo mi ha introdotto nel mondo del lavoro e che forse mi darà un appartamento nel suo grattacielo; tutto il glorioso "sottobosco" LAE nelle persone di Andrea Rosati con le sue perle musicali e il pane con il formaggio fuso, Carlo Chiari insostituibile compagno di Mac e di musica elettronica, Lorenzo Conti futuro sindaco di Besenzone (perchè è l'unico che ha fatto le scuole alte) e responsabile dei rapporti con il Vaticano, Lorenzo Rizzi compagno di viaggio e di misure notturne estreme: senza di loro non mi sarei divertito, non avrei visto il lavoro come un susseguirsi di ore piacevoli da passare in compagnia... grazie ragazzi; gli amici di AIDA (Fabio Bozzoli, Enrico Armelloni, Christian Varani, Erica Pincolini, Antonio Valdessalici, Giacomo Frassi) che mi sopportano e mi supportano; Simone Campanini, che pazientemente mi permette di sperimentare registrazioni sul suo coro; Alberto Amendola con il suo ottimismo e la sua disponibilità; gli amici che ho seguito per la tesi di laurea (Andrea Cammi, Marco Binelli, Marcello Bia, Daniela Marmiroli); tutta la combriccola Audiolink ed in particolare l'accoppiata Paolo&Guido, che con rara ed estrema disponibilità, senza

disperarsi, hanno assecondato le mie velleità di muratore falegname elettricista fonico magazziniere saldatore; Palazzina 5 e l'ufficio di Palazzina 8; Barbara Musi e la sua nuova piccola creatura; Carlo Augenti, fedele compagno di dottorato.

Grazie a tutti gli amici e parenti, mio nonno in primis, che mi hanno sempre spronato ad andare avanti per la mia strada, pur non capendo totalmente che tipo di lavoro io stia facendo.

Grazie a Gio, che continua a scommettere su di me, che non smette di entusiasmarsi per le mie scelte lavorative e le mie passioni musicali.

Grazie ai miei genitori: se sono arrivato in fondo ai tre anni di dottorato lo devo al loro instancabile supporto, alla loro pazienza, al loro desiderio di vedermi felice.

Grazie a Sara, alla sua gioia, al suo entusiasmo, ai suoi colori, alla sua speranza nel futuro.

"God only knows what I'd be without you..."
(B.Wilson)