

POLITECNICO DI TORINO

Corso di Laurea in Ingegneria del Cinema e dei Mezzi di  
Comunicazione

Tesi di Laurea Magistrale

**Tecnologie di spazializzazione  
audio per la ripresa e la  
riproduzione multicanale**



**Relatore**  
prof. Marco Masoero

**Candidato**  
Andrea Criniti

Luglio 2019

# Indice

<b>1</b>	<b>Introduzione</b>	<b>3</b>
<b>2</b>	<b>Il sistema Holophonics</b>	<b>5</b>
2.1	Teoria Duplex . . . . .	5
2.2	HRTF . . . . .	8
2.2.1	Misura delle HRTF . . . . .	8
2.2.2	Implementazione delle HRTF . . . . .	10
2.2.3	Rappresentazione delle HRTF . . . . .	11
2.3	Rising Sun Holophone H2-Pro . . . . .	14
2.3.1	Caratteristiche del microfono . . . . .	14
2.3.2	Sperimentazioni . . . . .	16
<b>3</b>	<b>Olofonia</b>	<b>19</b>
3.1	Il principio di Huygens . . . . .	19
3.2	Il sistema olofonico . . . . .	20
3.3	Registrazione e riproduzione . . . . .	21
<b>4</b>	<b>Wavefield Synthesis(WFS)</b>	<b>25</b>
4.1	Principi fisici . . . . .	25
4.2	Auralizzazione . . . . .	27
4.3	Il sistema WFS . . . . .	28
4.4	Aree di applicazione . . . . .	28
4.5	Vantaggi e svantaggi . . . . .	29
<b>5</b>	<b>Ambisonics</b>	<b>31</b>
5.1	Introduzione . . . . .	31
5.2	La teoria Ambisonic . . . . .	34
5.3	Codifica di Ambisonic . . . . .	36

5.4	Decodifica di Ambisonic . . . . .	37
5.5	Array microfonici . . . . .	38
5.5.1	SoundField ST450 . . . . .	39
5.5.2	Sennheiser Ambeo VR 3D . . . . .	42
5.5.3	mh acoustics em32 Eigenmike . . . . .	44
5.6	Vantaggi e svantaggi . . . . .	49
<b>6</b>	<b>Il 3D-VMS</b>	<b>51</b>
6.1	Descrizione del sistema . . . . .	51
6.1.1	Calcolo dei coefficienti del filtro . . . . .	52
6.1.2	Sintesi dei microfoni virtuali . . . . .	55
6.2	L'interfaccia grafica per l'utente . . . . .	57
6.3	Performance del sistema . . . . .	59
6.3.1	Latenza . . . . .	59
6.3.2	Risposta in frequenza . . . . .	59
<b>7</b>	<b>Casi applicativi con 3D-VMS</b>	<b>61</b>
7.1	Utilizzo del 3D-VMS in realtime: ripresa di un'opera lirica al Teatro Regio di Torino . . . . .	62
7.1.1	Panoramica . . . . .	62
7.1.2	Microfonazione dell'orchestra . . . . .	62
7.1.3	Microfonazione del palcoscenico . . . . .	65
7.2	Utilizzo del 3D-VMS in playback: post-produzione di un concerto di banda musicale . . . . .	70
7.2.1	Panoramica . . . . .	70
7.2.2	Strumentazione di ripresa . . . . .	70
7.2.3	Post-produzione . . . . .	71
7.2.4	Risultati . . . . .	79
7.3	Vantaggi del 3D-VMS rispetto alle tradizionali tecniche multi-microfoniche . . . . .	79
<b>8</b>	<b>Conclusione</b>	<b>81</b>
	<b>Bibliografia</b>	<b>85</b>

# Capitolo 1

## Introduzione

Nelle nostre attività quotidiane siamo costantemente immersi in un campo sonoro, che si tratti dei rumori del traffico, del suono di un ambiente naturale o della riverberazione di una stanza. La nostra percezione del suono è tridimensionale, dal momento che grazie alle nostre due orecchie siamo in grado di identificare la posizione di sorgenti sonore che ci circondano, distinguendo suoni che arrivano da destra o da sinistra, dall'alto o dal basso, da dietro o da davanti. Per questo motivo negli ultimi anni si sono sviluppate sempre più tecnologie audio in grado di ricreare il realismo nell'ascolto di un evento sonoro. Aumentare il realismo e l'immersione in un sistema virtuale e nel cinema permette all'utente di raccogliere un maggior numero di informazioni e di migliorare la percezione uditiva, quindi di provare un'esperienza nel complesso più piacevole. Basti osservare i recenti sviluppi e la grande diffusione della realtà virtuale (VR) nei contenuti visuali per capire quanto sia essenziale un costante perfezionamento dell'audio 3D.

Per spazializzazione audio o audio 3D si intende un insieme di tecnologie atte a fornire un'abilità di ascolto con caratteristiche spaziali elaborate, assimilabili quanto più possibile all'ascolto nella vita reale. Per fare ciò queste tecnologie si basano sulle informazioni uditive usate dal cervello per localizzare sorgenti sonore, per percepire la loro dimensione e l'ampiezza dell'ambiente riverberante che le contiene. Un esempio di tecnologia di ripresa basata su queste informazioni psicoacustiche è il sistema Holophonics di Zuccarelli, che combinando la teoria Duplex e le HRTF definisce un microfono in grado di simulare il modo in cui le nostre orecchie e la nostra testa registrano il campo sonoro.

La registrazione e la riproduzione audio sono utilizzati prevalentemente a scopo di intrattenimento, per cui è importante che gli eventi sonori vengano ripresi nella maniera più esteticamente adeguata. A tal proposito, le tecnologie di spazializzazione audio moderne offrono un enorme potenziale e un grande ventaglio di possibilità creative per il sound engineer che intende ricreare un'esperienza d'ascolto vicina alla vita reale. Di contro, in contrasto con le tradizionali tecniche stereofoniche, le tecnologie per l'audio 3D introducono nuovi livelli di complessità per quanto riguarda la registrazione, la codifica, la compressione, la trasmissione e il rendering del contenuto audio.

In generale, la ripresa di un campo sonoro richiede l'utilizzo di microfoni molto direttivi, in grado di separare le sorgenti situate in punti diversi dello spazio, laddove non sia pratico o possibile registrare individualmente ciascuna sorgente. Una volta che il suono è stato registrato, in fase di post-produzione i segnali vengono processati, mixati e predisposti per la loro rappresentazione nel formato finale. Uno dei problemi più rilevanti riguarda il sempre più grande numero di canali necessari alla riproduzione di un suono 3D e la difficoltà di gestione e automatizzazione delle posizioni delle sorgenti sonore. A tal proposito la principale svolta è stata l'adozione di una rappresentazione del suono indipendente dai canali di riproduzione (anche detto *channel-free audio*). In questo caso ogni sorgente/segnale può essere gestito esclusivamente in base alle proprie caratteristiche di spazializzazione, senza alcun vincolo con lo specifico formato di riproduzione in uso. Tra le tecnologie che operano secondo questo paradigma si trovano la *Wavefield Synthesis (WFS)*, che fa uso di segnali audio *object-based* (ossia elementi sonori a cui è associata una posizione virtuale che viene utilizzata per localizzarli opportunamente nello spazio una volta definito il formato di *playback*) e l'*Ambisonics*, dove le informazioni spaziali sono codificate indipendentemente dai canali di output.

Lo scopo di questa tesi è quello di presentare alcune delle principali tecnologie di registrazione e riproduzione audio 3D multicanale, descrivendone i principi di funzionamento base, gli strumenti (*hardware* e *software*) utilizzati, i vantaggi e gli svantaggi. La parte finale della tesi è invece una relazione su due casi applicativi del sistema 3D-VMS, utilizzato sia in fase di registrazione che di post-produzione, a cui ho assistito affiancando il dott. Leonardo Scopece durante il mio tirocinio alla Rai di Torino, tra Dicembre 2018 e Febbraio 2019.

# Capitolo 2

## Il sistema Holophonics

Il sistema Holophonics venne proposto nei primi anni '80 dal ricercatore italo-argentino Hugo Zuccarelli, il quale applicò il modello olografico ai fenomeni acustici. Questo sistema si basa essenzialmente su due principi psicoacustici:

- Teoria Duplex
- HRTF

### 2.1 Teoria Duplex

Il sistema uditivo umano è in grado di localizzare le sorgenti sonore con una buona precisione in certe direzioni. La teoria Duplex è un modello proposto da Lord Rayleigh [22] per approssimare il meccanismo con cui opera il sistema uditivo umano per localizzare un suono nel piano azimutale. Tale modello si basa su due grandezze:

- Inter-aural level differences (ILD), ossia la differenza di intensità (in decibel) con cui un suono giunge alle due orecchie.
- Inter-aural time differences (ITD), ossia la differenza di tempo con cui un suono giunge alle due orecchie.

Questi due meccanismi sono utili per definire la direzione di un suono nel piano azimutale (o orizzontale), cioè per distinguere la provenienza spaziale tra destra e sinistra. Un suono viene percepito più vicino ad un orecchio quando arriva prima a tale orecchio o quando vi arriva con intensità maggiore.

Se si considera una sinusoide pura, la percezione nel piano orizzontale è legata alla differenza temporale con cui essa giunge alle due orecchie.

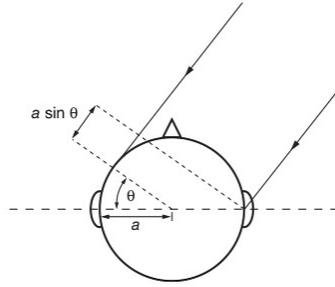


Figura 2.1: Schematizzazione della testa umana per il calcolo dell'ITD.

Approssimando la testa umana con una sfera rigida di raggio 8,5 cm, con le orecchie diametralmente opposte e data un'onda incidente piana a frequenze basse (minori di 1,5 kHz) (Figura 2.1), l'ITD può essere calcolata tramite semplici relazioni geometriche [29]

$$ITD = \frac{a}{c}(\sin \theta - \theta) \cos \phi; \quad -90^\circ \leq \theta \leq 90^\circ \quad (2.1)$$

con

$c$ : velocità del suono

$\theta$ : angolo azimutale

$\phi$ : angolo di elevazione (si considera  $\phi = 0$ )

Per frequenze prossime ai 1500 Hz la lunghezza dell'onda è circa pari a 22,5 cm, quindi paragonabile alla distanza inter-aurale. L'ITD fornisce una localizzazione sempre più ambigua all'aumentare della frequenza, a causa della variazione di fase che avviene nel momento in cui un suono deve girare intorno alla testa dell'ascoltatore prima di arrivare all'orecchio più lontano dalla sorgente (Figura 2.2). A queste frequenze il sistema uditivo per aliasing tende a percepire un ITD più corto, fornendo un risultato ambiguo, in cui ad un ITD corrispondono più punti sul piano azimutale. La ITD è dunque più efficace alle basse frequenze, mentre per le alte frequenze la IID diventa determinante per localizzare il suono.

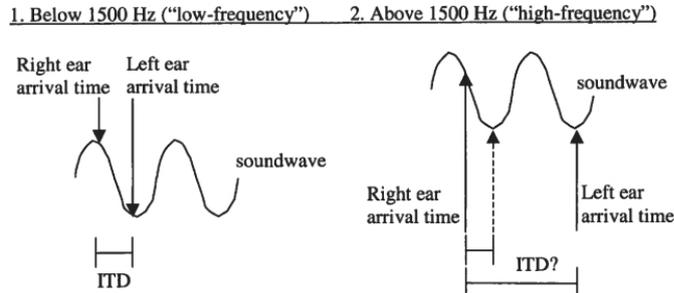


Figura 2.2: L'ambiguità del ITD nel determinare la posizione azimutale alle alte frequenze.

Quando la sorgente non è posizionata nel piano mediano (ossia il piano verticale immaginario che passa attraverso il centro degli occhi) la testa diffrange e riflette il suono facendo sì che il segnale giunga alle due orecchie con ampiezze diverse. Questo meccanismo è molto più evidente alle alte frequenze, le cui lunghezze d'onda sono paragonabili alle dimensioni della testa. La soglia minima per percepire un ILD è di circa 1 dB [19], mentre la differenza massima è di circa 20 dB [9] a 6 kHz.

La teoria Duplex descrive come avviene la stima della localizzazione spaziale soltanto sul piano azimutale senza però fornire indicazioni sull'angolo di elevazione e sulla distanza della sorgente dall'ascoltatore, caratteristiche di un suono nello spazio libero. La descrizione di una sorgente in un punto dello spazio, tramite la teoria Duplex, fornisce un set infinito di punti di localizzazione lungo curve di uguale distanza dalle orecchie, detto "cono di confusione" (Figura 2.3). Tutti i punti del cono sono dunque indistinguibili perché hanno uguali ITD e ILD.

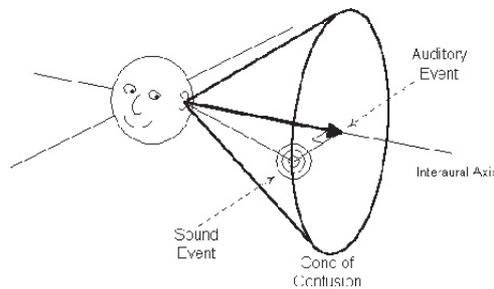


Figura 2.3: Cono di confusione.

## 2.2 HRTF

La teoria Duplex non è in grado di definire in modo completo i meccanismi di localizzazione spaziale del sistema uditivo umano. Per una corretta stima della posizione è necessario considerare anche le interferenze dell'orecchio esterno e del corpo (riflessione, diffrazione e distorsione lineare). Tali interferenze provocano nella struttura spettrale del suono delle micro variazioni, che dipendono dalla direzione di arrivo [3]. È stato verificato che nel caso di suoni complessi le variazioni di ampiezza e fase permettono al sistema uditivo una stima, più o meno corretta, della sorgente acustica in tutte le dimensioni spaziali [18].

Le HRTF (head related transfer function) sono funzioni di trasferimento tra la sorgente e un punto del canale uditivo, per varie direzioni di arrivo, definite per caratterizzare correttamente i meccanismi di percezione spaziale con cui opera il sistema uditivo. Le HRTF catturano i cambiamenti spettrali spaziodipendenti che avvengono quando un'onda acustica si propaga dalla sorgente al timpano dell'ascoltatore. Formalmente, una HRTF è definita come la risposta in frequenza specifica dell'orecchio sinistro/destro, misurata dalla posizione della sorgente a quella dell'orecchio. Quando un suono viene filtrato tramite opportune HRTF e inviato alle orecchie dell'ascoltatore (per esempio attraverso delle cuffie), il suono sintetizzato viene percepito come proveniente da una sorgente virtuale in una prescelta posizione nello spazio.

Definizione di HRTF [27]

*Per una sorgente acustica locata nello spazio di elevazione nel punto relativo all'ascoltatore di coordinate:  $(r, \theta, \phi)$  (distanza  $r$ , azimuth  $\theta$ , elevazione  $\phi$ ), si definisce HRTF, indicata con  $H_{r,\theta,\phi}(\omega)$ , il rapporto (complesso) tra il livello di pressione sonora (SPL) in un punto del canale uditivo in prossimità del timpano e l'SPL alla locazione della sorgente in campo libero (ovvero misurata in camera anecoica).*

### 2.2.1 Misura delle HRTF

In pratica, si inseriscono delle sonde microfoniche all'interno di entrambi i padiglioni auricolari di un manichino o di un ascoltatore reale. Sebbene le

misure su soggetti reali, in teoria, portano a risultati più attendibili, è più comune l'utilizzo di un manichino, dal momento che le misure richiedono lunghi tempi in cui il soggetto deve mantenere una determinata postura e dal momento che ogni piccola modifica nel posizionamento dei microfoni comporta variazioni nella risposta misurata. Nel caso di un soggetto reale si può, in alternativa, utilizzare dei microfoni esterni al canale uditivo, minimizzando i problemi sopra descritti. Questa misura però porterebbe a eliminare il contributo spettrale dovuto alla colorazione che il canale uditivo opera sulla HRTF.

Quindi si stabilisce un angolo di riferimento per la posizione della testa e si emette uno stimolo<sup>1</sup> tramite un altoparlante in una specifica posizione, caratterizzata da un azimut  $\theta$  e da un'elevazione  $\phi$ . Come mostrato in Figura 2.4 gli altoparlanti vengono disposti su una sfera posta virtualmente attorno al punto di ascolto, così da misurare funzioni di trasferimento per direzioni diverse.

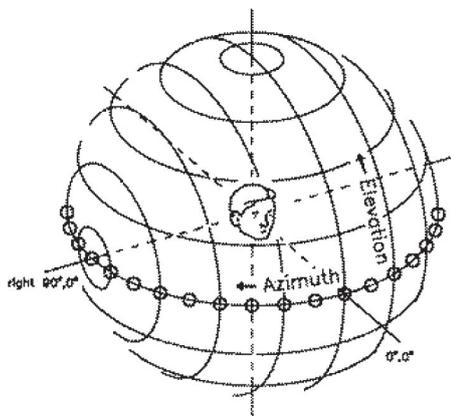


Figura 2.4: Posizionamento degli altoparlanti in coordinate sferiche.

Dal momento che la misura avviene generalmente in campo lontano (far field), si può ricondurre la funzione di trasferimento a sole tre variabili, che perciò diventa  $H_{\theta,\phi}(\omega)$  [27].

<sup>1</sup>Generalmente si utilizza come stimolo una sequenza binaria pseudo random (PBRS)

### 2.2.2 Implementazione delle HRTF

Quando una HRTF è misurata, i risultati sono ottenuti sotto forma di due HRIR (head related impulse responses), una per orecchio. Le versioni campionate delle HRIR sono generalmente implementate come filtri digitali FIR che descrivono il modo in cui un suono viene modificato dalla testa e dal corpo dell'ascoltatore prima di raggiungere l'orecchio corrispondente. In Figura 2.5 è mostrato un esempio di HRIR.

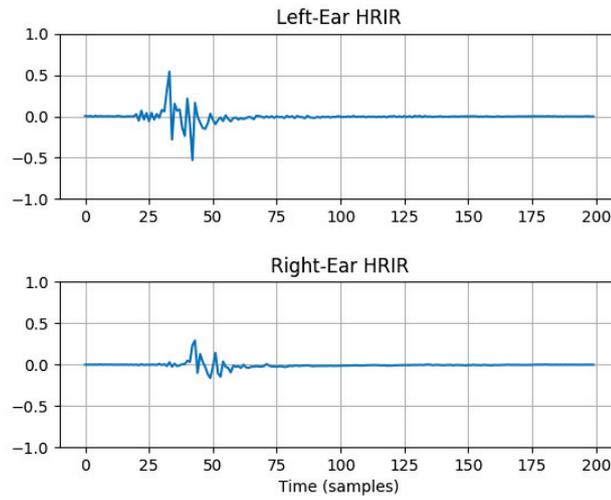


Figura 2.5: Esempio di HRIR per entrambe le orecchie.

È possibile ricavare facilmente le informazioni su ITD e ILD tramite la differenza tra la funzione dell'orecchio destro  $H_{\theta,\phi}^R(\omega)$  e quella dell'orecchio sinistro  $H_{\theta,\phi}^L(\omega)$ : sia il rapporto

$$H(\omega) = \frac{H_{\theta,\phi}^R(\omega)}{H_{\theta,\phi}^L(\omega)}$$

definito come funzione di trasferimento interaurale, il cui modulo è pari a

$$A = 20 \log_{10} |H(\omega)| = A_R - A_L [dB].$$

Nel caso di azimuth  $\theta$  e elevazione  $\phi$  costanti, la grandezza  $A$  rappresenta l'ILD.

L'ITD può invece essere definito come

$$\frac{d\Phi(\omega)}{d\omega} = \frac{d[\Phi_R(\omega) - \Phi_L(\omega)]}{d\omega}.$$

### 2.2.3 Rappresentazione delle HRTF

W. Gardner and K. Martin, due ricercatori del MIT, hanno effettuato una serie di misurazioni con un manichino KEMAR [10]: sono state misurate le risposte all'impulso di entrambe le orecchie al suono emesso da un altoparlante Realistic Optimus Pro 7, posizionato a 1,4 metri dal KEMAR. È stato utilizzato un MLS (maximum length sequence), una sequenza binaria pseudocasuale, come stimolo per ottenere la risposta all'impulso ad una frequenza di campionamento di 44,1 kHz. Da tale misurazione sono stati ottenuti 710 diversi campioni, per elevazioni che variano tra i  $-40^\circ$  e  $+90^\circ$ ; i risultati sono stati raccolti in un database e resi disponibili sul loro sito web [1].

La Figura 2.6a e la Figura 2.6b mostrano la rappresentazione di un HRIR dell'orecchio destro ad uno stimolo impulsivo sul piano orizzontale ( $\phi = 0$  e  $\theta$  variabile) e sul piano mediano ( $\phi$  variabile e  $\theta = 0$ ). Le aree ad alta luminosità rappresentano valori più alti della risposta. Nella Figura 2.6a la risposta ha un picco positivo per suoni che arrivano da destra (azimut di  $90^\circ$ ), mentre è più debole per i suoni che arrivano da sinistra (azimut di  $270^\circ$ ); si può inoltre notare che il ritardo temporale con cui il suono giunge all'orecchio varia con l'azimut approssimando un andamento sinusoidale, conforme con l'andamento teorico dell'ITD formulato in (2.1). La Figura 2.6b mostra invece che quando la sorgente si muove attorno alla testa, sul piano mediano, i cambiamenti della risposta all'impulso sono meno evidenti. Il tempo con cui il suono arriva all'orecchio è all'incirca uguale per le diverse elevazioni; le principali differenze sono relative agli arrivi di ampiezza minore, posti ad un certo ritardo e dovuti alle riflessioni della pinna dell'orecchio. Questi cambiamenti sono però più evidenti nel dominio della frequenza, dove si manifestano come un filtro notch la cui posizione varia con l'elevazione.

La Figura 2.7 mostra la risposta in frequenza dell'orecchio destro del KEMAR per una sorgente che si muove sul piano orizzontale. Anche in questo grafico si può osservare che per ogni frequenza, al variare dell'azimut, l'andamento della risposta è circa sinusoidale e che i suoi valori sono più alti quando la

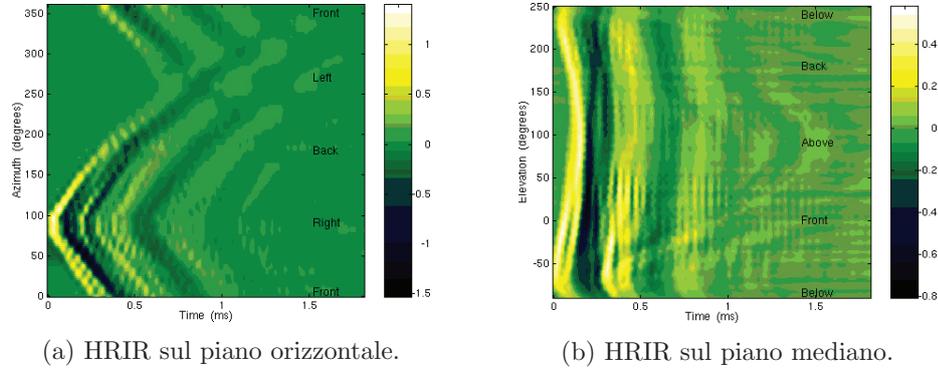


Figura 2.6: HRIR ottenute delle sperimentazioni con KEMAR.

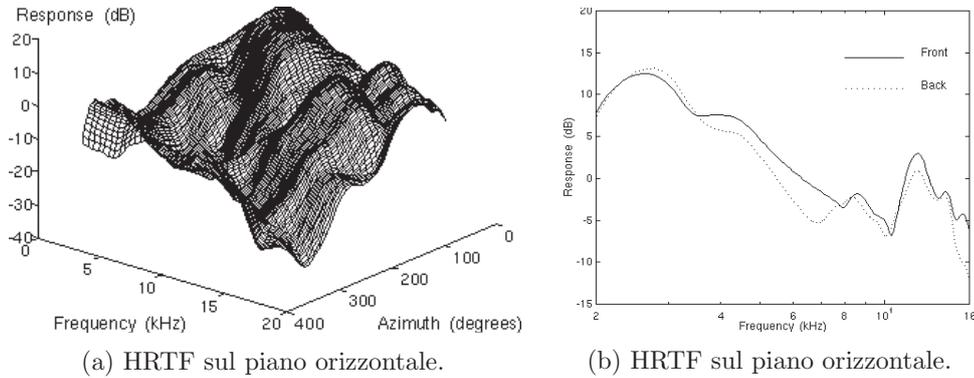


Figura 2.7: HRTF ottenute delle sperimentazioni con KEMAR.

sorgente è a  $90^\circ$  di azimuth (quindi diretta verso l'orecchio destro) e più bassi quando la sorgente è a  $270^\circ$  di azimuth (quindi dal lato opposto della testa). Il grafico di destra mostra che tra i 4 kHz e 7 kHz la risposta ad un suono frontale è di pochi dB più alta rispetto a quella ad un suono posteriore. Il picco attorno ai 4 kHz è dovuto alla risonanza del canale uditivo. Il notch attorno ai 10 kHz è dovuto alle riflessioni della pinna, nominate sopra.

La Figura 2.8 mostra una serie di risposte in frequenza delle HRTF valutate per una sorgente che si muove nel piano mediano, al variare dell'elevazione. Anche per queste curve si osserva il massimo intorno ai 4 kHz, dovuto alla risonanza del canale uditivo; in questo caso però la frequenza del notch dovuto alla pinna cambia significativamente, a partire da 6 kHz per basse

elevazioni e fino a 10 kHz per alte elevazioni. Inoltre, quando la sorgente è posta frontalmente, il notch non è presente e la risposta risulta praticamente piatta.

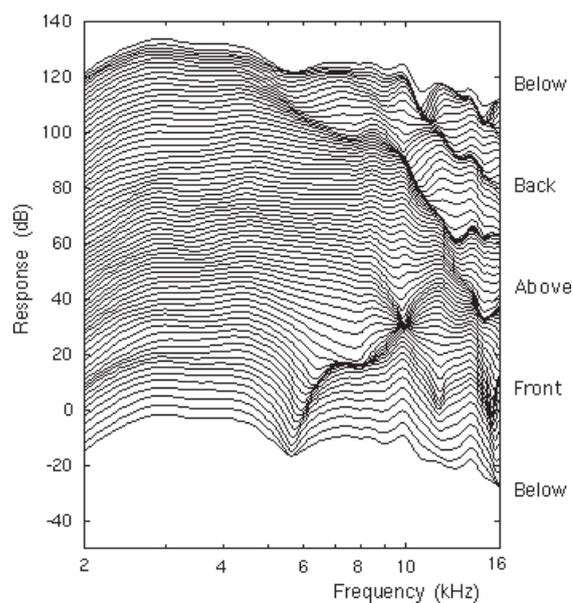


Figura 2.8: HRTF sul piano mediano.

Per la piena comprensione delle HRTF non è sufficiente conoscere le loro risposte sul piano orizzontale e su quello mediano, dal momento che le HRTF sono funzioni complesse dell'azimut e dell'elevazione. Inoltre, sebbene le HRTF contengano le tre grandezze necessarie alla caratterizzazione della percezione spaziale (risposta dell'orecchio sinistro, risposta dell'orecchio destro e ritardo interaurale), esse risultano dipendenti dalla soggettività dell'ascoltatore. Infatti quando una persona ascolta una sorgente attraverso le HRTF di un'altra persona possono verificarsi errori di localizzazione significativi.

## 2.3 Rising Sun Holophone H2-Pro

La tecnica Holophonics si basa sull'Holophone H2-PRO<sup>2</sup> (Figura 2.9), della ditta canadese Rising Sun,, un moderno microfono professionale che si basa sui principi sopra riportati per catturare la spazialità di un suono e creare campi sonoro per configurazioni a 5.1, 6.1 o 7.1 canali discreti, per broadcast televisivo (standard TV, DTV, and HDTV), broadcast radiofonico, lungometraggi e altre applicazioni professionali. Le registrazioni del H2-PRO producono segnali discreti in realtime, facilmente adattabili a formati surround professionali e consumer (es.: Dolby, DTS, e Circle Surround). Tra gli utilizzatori di questo microfono: ESPN, CBS, FOX HDTV, RAI, Mediaset, Sky Italy e molti altri.



Figura 2.9: Il microfono Holophone H2-Pro.

### 2.3.1 Caratteristiche del microfono

L'H2-PRO è un sistema composto da 8 capsule omnidirezionali DPA 4060, di cui 7 disposte sul perimetro esterno di un supporto ellittico, che simula le caratteristiche di una testa umana, e una interna. Come mostrato in Figura 2.10, tre capsule sono destinate alla ripresa anteriore, tre alla ripresa posteriore, una alla ripresa azimutale e una (quella interna) riceve il suono filtrato acusticamente dalla testa e agisce da filtro passa-basso naturale,

---

<sup>2</sup>[www.holophone.com](http://www.holophone.com)

generando informazioni LFE<sup>3</sup>. Alcune specifiche dichiarate dal costruttore [13]:

- Range di frequenze: 7 x 20 Hz-20 kHz  $\pm 2$  dB, 1 x 20 Hz-110 Hz  $\pm 2$  dB
- Sensibilità: nominale 20 mV/Pa  $\pm 3$  dB a 1 kHz
- Livello equivalente rumore pesato A: tipicamente 23 dB-A a 20  $\mu$ Pa (max. 26 dB-A)
- Max SPL: 134 dB SPL prima del clipping
- Distorsione armonica totale: 123 dB SPL peak ( $< 0\%$  THD)
- Impedenza d'uscita: 30-40  $\Omega$

Le 7 capsule esterne catturano un range di frequenze da 20 Hz a 20 kHz, mentre quella interna cattura un range di frequenze da 20 Hz a 110 Hz. Gli 8 connettori XLR consentono di ottenere 8 segnali discreti bilanciati (Left, Right, Center, Left Surround, Right Surround, Top, Center Rear e Low Frequency), il che rende possibile ottenere un suono surround anche inviando le uscite del microfono direttamente agli altoparlanti.

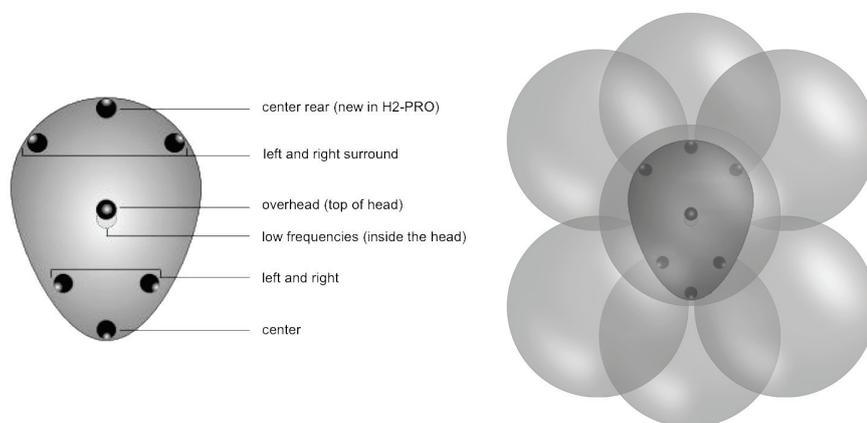


Figura 2.10: Posizione delle capsule e diagramma polare del H2-PRO.

<sup>3</sup>LFE: low frequency effects; il canale LFE contiene soltanto basse frequenze ( $< 120$  Hz) e non ha effetti diretti sulla direzionalità del suono riprodotto. La traccia LFE può essere riprodotta su qualunque altoparlante, ma quando il sistema d'ascolto lo permette, si tende ad associarla ad un subwoofer.

L'Holophone è compatibile con i seguenti formati audio:

- 6.1 IMAX
- Dolby Digital EX
- DTS ES
- Circle Surround II (CS II)
- 5.1 Dolby Digital
- Dolby ProLogic II
- DTS
- Circle Surround
- AAC (Advanced Audio Coding – MPEG 4)
- WMA (Windows Media Audio)
- 7.1 SDDS
- 4-2-4 Dolby Stereo
- Stereo

A seconda dell'applicazione, sono possibili diversi setup per la registrazione con l'H2-PRO. Nello scenario più semplice, l'H2-PRO si collega direttamente con console o interfaccia di registrazione: tramite cavi XLR si inseriscono gli output del microfono negli input di una console con controllo di gain e phantom power (Figura 2.11). Gli 8 segnali possono quindi essere trattati come 8 microfoni indipendenti.

### 2.3.2 Sperimentazioni

Il centro ricerche della RAI [23], in collaborazione con l'Università di Parma [7], ha sperimentato e valutato l'H2-PRO. I seguenti diagrammi polari (Figura 2.12) sono stati misurati per ognuna delle 8 capsule nella camera anecoica dell'Università di Ferrara.

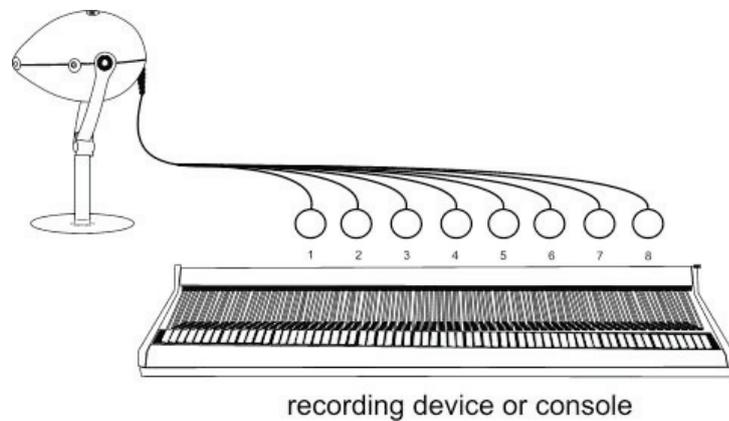


Figura 2.11: Schema di collegamento base per l'Holophone H2-PRO.

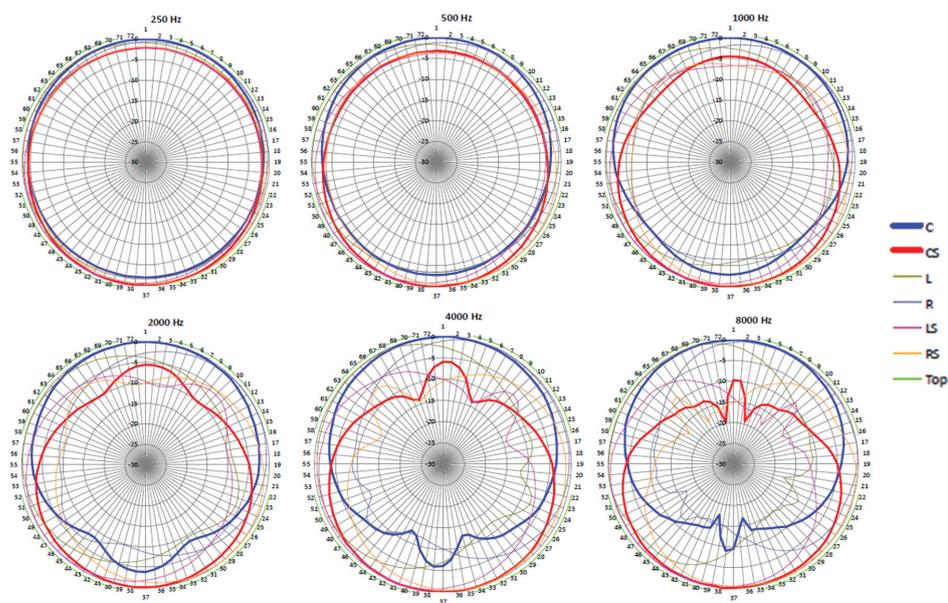


Figura 2.12: Diagrammi polari delle 8 capsule misurati in camera anecoica.

Tra le problematiche emerse:

- La direttività e gli angoli tra le singole capsule non possono essere modificati in post-produzione

- Non c'è sufficiente separazione tra le sorgenti a causa della scarsa direttività della capsule.
- L'immagine sonora catturata è tendenzialmente inaccurata sebbene la registrazione risulti molto spaziosa e con una buona risposta in frequenza
- Un posizionamento non ideale del microfono può compromettere la qualità della ripresa, dal momento che non c'è possibilità di codificare i segnali e effettuare un riposizionamento virtuale in post-produzione.

Tuttavia dalla ripresa di spettacoli di diversa natura con l'H2-Pro sono emersi anche grandi vantaggi, relativi alla sua semplice operabilità e immediatezza. Tra questi:

- Praticità di utilizzo: le dimensioni fisiche del microfono sono tali da posizionarlo senza eccessivo ingombro all'interno del campo sonoro da riprendere. La strumentazione richiesta si limita ad un interfaccia audio con un sufficiente numero di input e in grado di alimentare le capsule e ad un computer o altro dispositivo di registrazione
- Nessuna codifica di segnale richiesta: i segnali ripresi dalle 8 capsule possono essere utilizzati direttamente come output per una delle configurazioni viste sopra, senza richiedere ulteriori processi di codifica e decodifica. Se da una parte questo rende il microfono meno versatile, dall'altra offre la possibilità di creare facilmente e velocemente prodotti professionali in formati d'ascolto standard.

# Capitolo 3

## Olofonia

L'idea alla base dell'olofonia è quella di ricostruire il campo sonoro a partire da un suo campionamento spaziale. In altre parole, è possibile ricreare un campo sonoro all'interno di un volume, partendo dalla registrazione sulla superficie che lo contiene [15].

### 3.1 Il principio di Huygens

L'olofonia si basa sul principio di Huygens, che afferma che il campo sonoro prodotto da una sorgente è equivalente al campo generato da una distribuzione continua di sorgenti disposte sul fronte d'onda della sorgente originale.

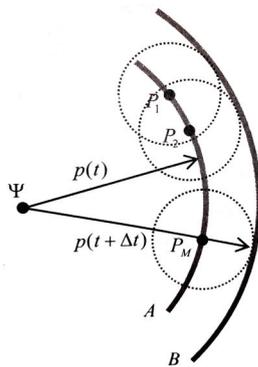


Figura 3.1: Rappresentazione del campo sonoro e delle sorgenti secondarie.

Facendo riferimento alla Figura 3.1 un campo sonoro emesso dalla sorgente  $\Psi$  può essere ricostruito attraverso una distribuzione di sorgenti secondarie definita da  $M$  altoparlanti  $P_1, P_2, \dots, P_M$ . Il principio di Huygens è descritto matematicamente dall'integrale di Kirchhoff-Helmholtz:

$$P_R(\mathbf{r}_R) = \frac{1}{4\pi} \iint_S [P_S(\mathbf{r}_S) \cdot \nabla G(\mathbf{r}_R, \mathbf{r}_S) - G(\mathbf{r}_R, \mathbf{r}_S) - \nabla P_S(\mathbf{r}_S)] \cdot \mathbf{n} dS \quad (3.1)$$

con

- $P_R(\mathbf{r}_R)$ <sup>1</sup>: pressione in un punto qualunque di coordinate  $\mathbf{r}_R$  contenuto nel volume  $V$
- $P_S(\mathbf{r}_S)$ : la pressione, causata dalla sorgente secondaria, in un punto di coordinate  $\mathbf{r}_S$  sulla superficie  $S$
- $S$ : superficie che racchiude il volume  $V$  (Figura 3.2)

e dove  $G(\mathbf{r}_R, \mathbf{r}_S)$  è la funzione di Green, definita come

$$G(\mathbf{r}_R, \mathbf{r}_S) = \frac{e^{-jk|\mathbf{r}_R - \mathbf{r}_S|}}{|\mathbf{r}_R - \mathbf{r}_S|}$$

Esplicitando la funzione di Green e ponendo  $\mathbf{R} = \mathbf{r}_R - \mathbf{r}_S$  e  $R = |\mathbf{r}_R - \mathbf{r}_S|$  la formula (3.1) diventa

$$P_R(\mathbf{r}_R) = \frac{1}{4\pi} \iint_S \left[ \nabla P_S(\mathbf{r}_S) \cdot \mathbf{n} - \frac{\mathbf{R}}{R} \cdot \mathbf{n} (1 + jkR) \frac{P_S(\mathbf{r}_S)}{R} \right] \cdot \frac{e^{-jkR}}{R} dS \quad (3.2)$$

ossia conoscendo la pressione sulla superficie  $S$  che racchiude il volume  $V$  e il gradiente normale alla superficie  $S$  è possibile descrivere il campo acustico in un volume  $V$ .

## 3.2 Il sistema olofonico

L'integrale della formula (3.2) va inteso come un insieme secondario di sorgenti, dette sorgenti di Kirchhoff, distribuite sulla superficie  $S$ . Tale insieme

---

<sup>1</sup> $\mathbf{r}$  è il vettore in coordinate polari  $(r, \theta, \phi)$

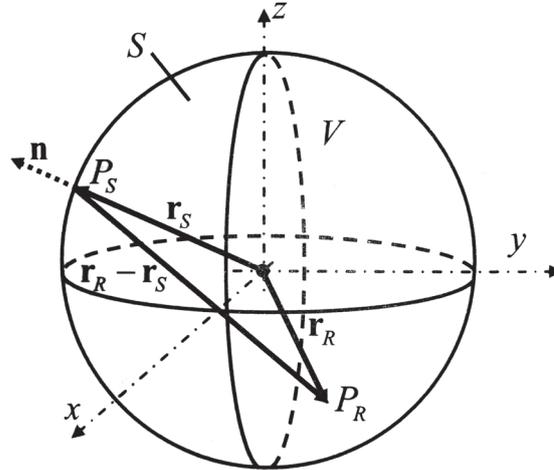


Figura 3.2: Formalismo dell'integrale di Kirchhoff-Helmoltz.

costituisce un metodo per riprodurre il campo di pressione in un punto arbitrario contenuto nel volume, a partire dalle caratteristiche note del campo sulla superficie che lo comprende. Perciò è necessario conoscere la pressione  $P_S(\mathbf{r}_S)$  e il suo gradiente  $\nabla P_S(\mathbf{r}_S)$  al fine di ricostruire perfettamente il campo acustico  $P_R(\mathbf{r}_R)$ . Per fare questo occorre campionare la pressione in un opportuno numero di punti sulla superficie  $S$ , con coppie di opportuni trasduttori, capaci di registrare sia la pressione che il suo gradiente.

### 3.3 Registrazione e riproduzione

Il sistema di registrazione è formato da coppie di microfoni, uno omnidirezionale (in grado di catturare la pressione del campo) e l'altro direzionale con figura a otto (in grado di catturare il gradiente di pressione del campo). Il sistema di riproduzione è analogamente formato da coppie di sorgenti a monopolo (altoparlanti a cassa chiusa), e sorgenti a dipolo (altoparlanti a cassa aperta): i primi generano segnali di pressione, i secondi sfruttano l'emissione posteriore per generare gradienti di pressione. In Figura 3.3 è mostrato il collegamento ideale tra sistema di ripresa e sistema di riproduzione. I segnali registrati dai microfoni direzionali alimentano gli altoparlanti a dipolo, mentre quelli registrati dai microfoni omnidirezionali alimentano gli altoparlanti chiusi.

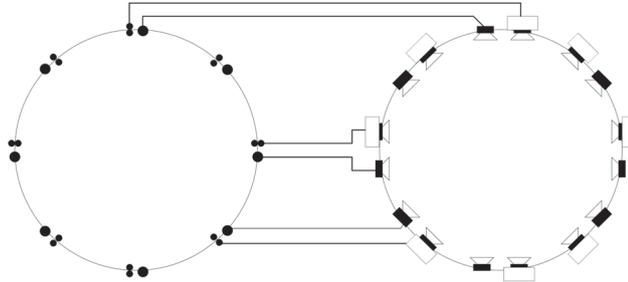


Figura 3.3: Collegamento ideale di sistema di registrazione e sistema di riproduzione. La distribuzione di microfoni e altoparlanti sulle due superfici è continua.

Per questo approccio si utilizzano degli array di microfoni che realizzano il campionamento del campo acustico sulla superficie  $S$ , con un adeguato sistema di coordinate spaziali. La minima lunghezza d'onda registrabile (quindi riproducibile) è determinata dalla distanza effettiva tra la coppia di microfoni in ripresa, in accordo col teorema del campionamento di Shannon. Al fine di scegliere un adeguato numero di trasduttori sarà quindi necessario valutare le dimensioni di  $S$  e la frequenza minima che si desidera riprodurre. Nella pratica, servirebbe un numero irrealistico di trasduttori per soddisfare il teorema di Shannon. Da quanto emerge dagli esperimenti di Uncini [26], l'utilizzo di un numero non sufficiente di microfoni durante il campionamento causa una distorsione d'aliasing spaziale, che per frequenze maggiori di 1,5 kHz non degradano percettivamente il segnale riprodotto. Un'approssimazione, detta formulazione di Rayleigh [21], permette di usare un numero finito di trasduttori, disposti lungo punti discreti su una curva del piano d'interesse. Inoltre, dal momento che i contributi delle coppie di altoparlanti risulta ridondante [21] in pratica è possibile utilizzare solo altoparlanti dipolari in fase riproduzione e solo microfoni a gradiente di pressione in fase di registrazione.

Nel caso in cui la riproduzione sia limitata ad un piano, invece che ad un volume, le sorgenti di Kirchoff possono essere disposte su una curva; ad esempio, in Figura 3.4, è mostrato lo schema di ripresa a 8 canali per la spazializzazione 2D su una superficie circolare [4].

Per quanto riguarda l'ambiente di riproduzione, ricostruire in modo fedele un campo sonoro con questa tecnica richiederebbe una condizione perfettamen-

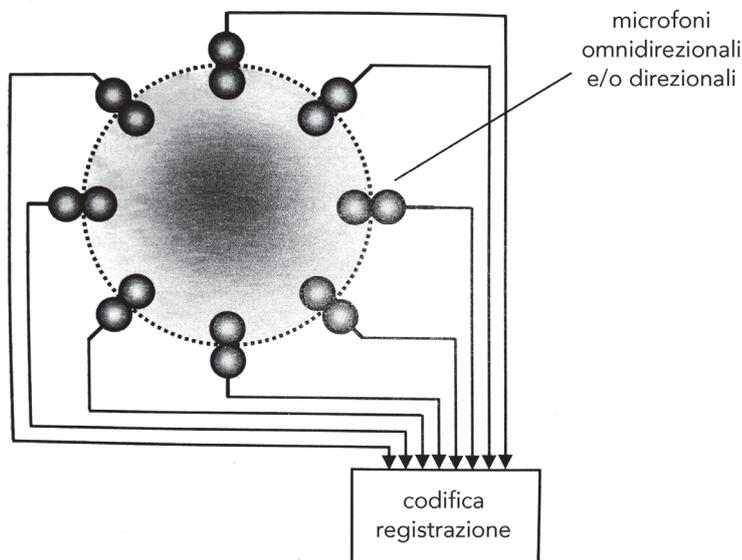


Figura 3.4: Ripresa con sistema olofonico a 8 canali.

te anecoica, per evitare che l'ascolto venga influenzato dalla riverberazione intrinseca della sala. Per ricostruire un campo in ambiente non anecoico si può utilizzare lo schema in Figura 3.5, dove la matrice di funzioni di trasferimento  $G(z)$  è definita in modo da eliminare o quantomeno minimizzare l'effetto della sala (descritto da una funzione di trasferimento  $C(z)$ ).

È possibile determinare la matrice  $G(z)$  tramite una procedura di calibrazione durante la riproduzione: sia  $e[n]$  l'errore tra il segnale ottenuto dai microfoni  $y[n]$  e un segnale di riferimento  $d[n]$  (per esempio, una sequenza PRBS), tale per cui

$$e[n] = d[n] - yn$$

$G(z)$  è la matrice filtri FIR tali per cui  $e[n]$  viene minimizzato. Questa calibrazione permette di eliminare l'influenza della sala  $C(z)$  e nell'ipotesi in cui i filtri siano sufficientemente lunghi e la  $C(z)$  sia a fase minima, si ottiene  $G(z) = C^{-1}(z)$ .

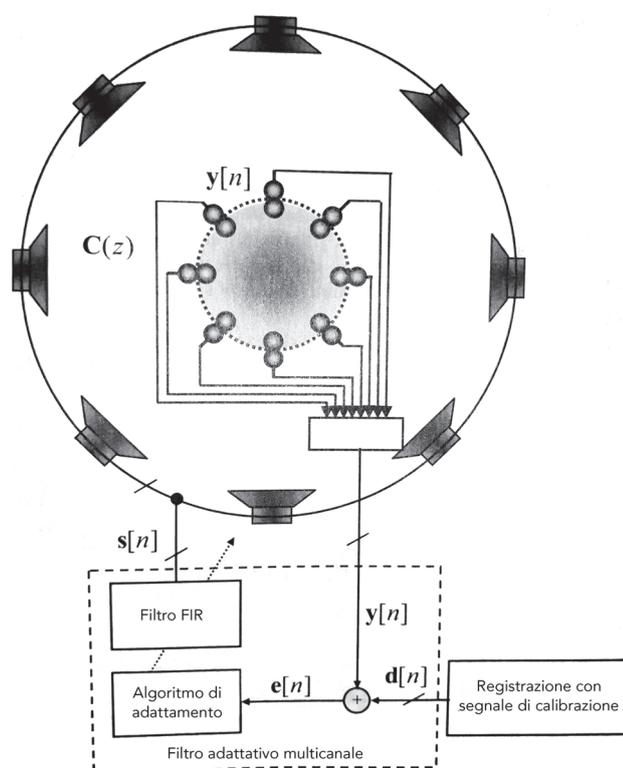


Figura 3.5: Calibrazione di un sistema di riproduzione olofonico.

## Capitolo 4

# Wavefield Synthesis(WFS)

La Wavefield Synthesis (WFS), introdotta negli anni '80 [2], è una tecnologia per la sintesi di campi sonori con proprietà spaziali e temporali naturali, all'interno di un'area definita dal sistema di riproduzione.

### 4.1 Principi fisici

Questa tecnica è un caso speciale di olofonia e fa riferimento alla teoria del fronte d'onda introdotta dal principio di Huygens, secondo cui la propagazione di un'onda in un mezzo può essere descritta attraverso il contributo di opportune sorgenti secondarie: sia  $S$  una sorgente sonora (Figura 4.1a), detta primaria, che emette fronti d'onda sferici; in accordo col principio di Huygens ogni fronte può essere pensato come la somma delle onde sferiche emesse da sorgenti secondarie sulla superficie  $O$ . Nella WFS le sorgenti secondarie sono rappresentate da degli altoparlanti, in grado di ricreare virtualmente la sorgente  $S$  (Figura 4.1b)). L'ascoltatore nel punto arbitrario  $P$  del campo sonoro percepisce lo stesso fronte d'onda emesso dalla sorgente  $S$ .

Nonostante la WFS si rifaccia al sistema olofonico, quindi ad un approccio utilizzabile sia in registrazione che in riproduzione, questa tecnologia si è diffusa prevalentemente come sistema di riproduzione. Il principio di funzionamento della WFS si basa sull'utilizzo di array di altoparlanti per ricreare un campo sonoro che sia identico ad un appropriato evento sonoro reale. I campi acustici naturali sono in genere formati dalla combinazione dei campi emessi dalle singole sorgenti e dalla risposta all'impulso dell'ambiente. Nella

WFS le sorgenti virtuali vengono rappresentate nello spazio per mezzo di sorgenti puntiformi che emettono tracce mono, a cui sono associati dei metadati contenenti informazioni di spazializzazione (solitamente azimuth e elevazione). Il contributo dell'ambiente alla scena riprodotta può essere quello del reale ambiente di registrazione oppure quello di una diversa stanza, le cui proprietà acustiche siano state pre-registrate o ricavate matematicamente.

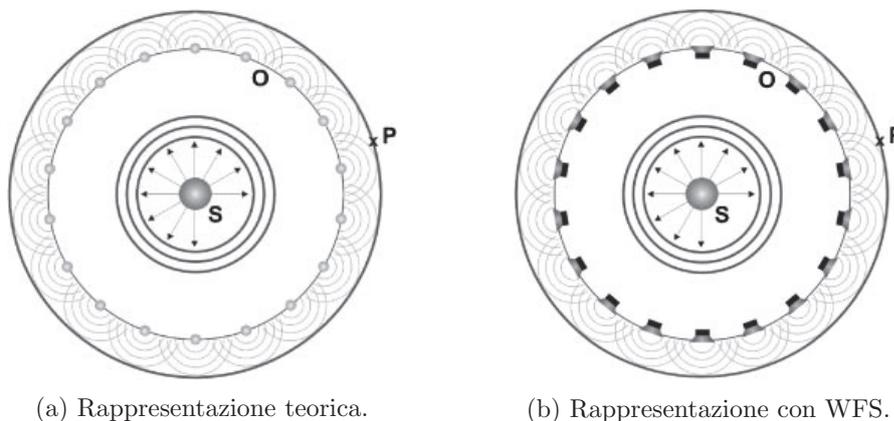


Figura 4.1: Il principio di Huygens.

La WFS è in grado di:

- riprodurre sorgenti virtuali localizzate nella stessa posizione sull'intera area di ascolto (freccie tratteggiate in Figura 4.2)
- riprodurre onde piane localizzate nella stessa direzione sull'intera area di ascolto (freccie continue in Figura 4.2)
- migliorare la localizzazione di una sorgente virtuale riproducendo una realistica distribuzione della sua ampiezza (es. quando un ascoltatore si avvicina alla sorgente, l'ampiezza del suono aumenta in modo realistico).

Inoltre, la capacità di riprodurre fronti d'onda piani permette di ottenere un livello di pressione sonora più o omogeneo sull'area di ascolto, dal momento che per le onde piane questo si mantiene approssimativamente uguale su tutte le direzioni di propagazione. L'insieme di queste caratteristiche consente la sintesi di campi sonori complessi che rimangono coerenti anche muovendosi all'interno dell'area di ascolto.

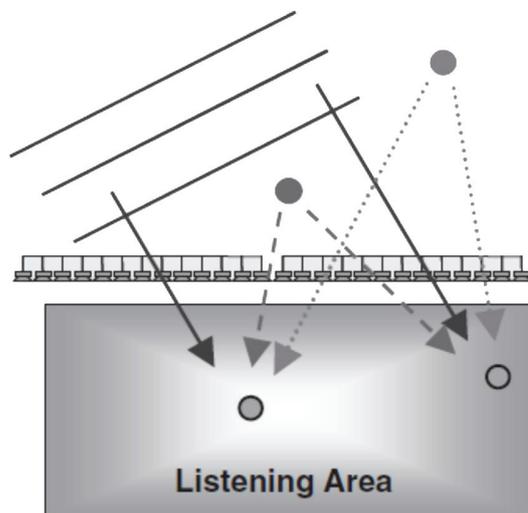


Figura 4.2: Tipi di propagazione dell'onda acustica con WFS.

## 4.2 Auralizzazione

Come già accennato, una delle peculiarità della tecnica WFS è la possibilità di ricostruire in modo realistico una certa caratteristica acustica in un certo ambiente d'ascolto perfettamente anecoico [27]. Per fare ciò bisogna dapprima ricavare la risposta all'impulso della sala che si vuole simulare e ottenere la matrice  $\mathbf{h}[n]$ . In seguito occorre filtrare le sorgenti di segnale  $\mathbf{x}'[n]$  con la matrice di risposte impulsive dell'ambiente  $\mathbf{h}[n]$ :

$$\mathbf{x}[n] = \mathbf{h}[n] * \mathbf{x}'[n]$$

Tuttavia in applicazioni reali la sala d'ascolto non potrà mai essere completamente anecoica: una corretta auralizzazione richiede l'eliminazione delle risposte all'impulso della sala d'ascolto  $\mathbf{c}[n]$  tra l'array di altoparlanti e i punti d'ascolto desiderati. Per fare ciò occorre utilizzare un banco di filtri in grado di equalizzare la sala di ascolto e ridurre l'influenza sul campo acustico ricostruito, con un processo di calibrazione analogo a quello descritto nel capitolo precedente.

### 4.3 Il sistema WFS

Il sistema WFS si suddivide essenzialmente in 3 fasi: creazione, elaborazione e riproduzione delle sorgenti sonore. Tali fasi sono integrate in una Spatial Audio Workstation (SAW), un software generalmente composto dai seguenti moduli:

- Server per l'audio: registra e memorizza una serie di dati audio, fino a 64 canali d'input per le sorgenti virtuali. È possibile collegare questo modulo con delle convenzionali console per il mixaggio e pre-mixaggio dei segnali e delle sorgenti virtuali.
- Elaborazione e riproduzione: sezione che riceve l'audio in ingresso e calcola i segnali da inviare agli altoparlanti.
- Simulazione dell'ambiente: permette di creare stanze "reali" basate su misure acustiche o stanze artificiali basate su parametri percettivi.
- Interfaccia utente: mostra graficamente le posizioni delle sorgenti virtuali, i segnali in uso ed eventuali errori nel sistema

Tra i software commerciali attualmente reperibili si trovano SonicEmotion e IOSONO.

### 4.4 Aree di applicazione

Gli algoritmi della WFS comportano degli intrinseci tempi di ritardo che sono sufficientemente brevi per l'utilizzo in performance live. I contesti in cui questa tecnica può essere applicata sono svariati:

- Sale da concerto: se le condizioni acustiche della sala sono buone e non richiedono filtri di equalizzazione, l'utilizzo della WFS è semplice e rende più naturale un suono elettronicamente amplificato.
- Eventi all'aperto: per tali situazioni sono richieste una distribuzione omogenea del livello di pressione sonora lungo l'intera area di ascolto e la coerenza spaziale tra suono e elementi visivi sul palco. I comuni array lineari di altoparlanti possono controllare la pressione acustica in zone specifiche, ma presentano problemi nelle regione in cui le onde dei

diversi array si sovrappongono; inoltre non possono garantire la coerenza spaziale. La WFS è in grado di soddisfare entrambe le condizioni, perché si basa sulla generazione di un campo sonoro continuo, quindi senza sovrapposizioni di onde differenti. Inoltre, la WFS consente di ricreare attorno all'area di ascolto un'ambiente artificiale, con proprietà acustiche di una stanza chiusa (adatta per esempio ad un concerto di musica classica). Questa versatilità è utile per adeguare il contesto al tipo di performance in atto.

- Cinema: oltre ad un'accurata rappresentazione del campo acustico nella sala, la WFS consente di riprodurre sorgenti sonore nella loro reale profondità spaziale. Questo strumento risulta decisamente suggestivo se combinato con immagini in live action.
- Sistemi home theatre: sebbene si tratti di una soluzione ancora troppo dispendiosa, i benefici della WFS possono essere applicati anche a piccole stanze. Gli attuali problemi di questa tecnologia in ambienti domestici riguardano l'installazione degli array di altoparlanti in uno spazio ristretto (a cui in futuro potrebbero ovviare alternative più pratiche come sistemi di pannelli DML [14] o sorgenti nascoste, integrabili all'interno di mobili e dietro la carta da parati) e l'acustica dell'ambiente (gestibile tramite trattamento acustico e applicazione di tecniche di equalizzazione della stanza).

## 4.5 Vantaggi e svantaggi

I più grandi vantaggi della WFS rispetto alla riproduzione multicanale tradizionale sono:

- migliorata qualità sonora nell'audio riprodotto: WFS crea un campo sonoro continuo, quindi idealmente privo di problematiche legate alla fase dei segnali riprodotti.
- Migliorata realistica nella localizzazione delle sorgenti virtuali.
- Per questa tecnica invece che di sweet spot (inteso come unico punto di ascolto ottimale) si può quindi parlare di "sweet zone", intesa come zona relativamente ampia in un cui l'ascolto è corretto per più ascoltatori contemporaneamente. L'ampiezza della zona, quindi il numero

di ascoltatori coinvolti, dipendono dalle caratteristiche dell'impianto di riproduzione. Per esempio, quando si utilizza questa tecnologia per riprodurre un'onda piana proveniente da una certa direzione, tutti gli ascoltatori percepiranno il suono come proveniente dalla medesima direzione, al di là della loro posizione nell'area di ascolto. Con i sistemi di riproduzione discreti questo non è possibile e la direzione di arrivo di un suono dipende dalla posizione relativa tra ascoltatore e altoparlante.

- Memorizzazione dell'audio basata sulla sorgente (object-based): con un multicanale tradizionale il formato dell'audio dipende dal numero di canali di riproduzione (quindi dal numero di altoparlanti), con la WFS, invece, ogni elemento audio viene registrato e memorizzato individualmente e rappresentato con la migliore qualità audio permessa dal sistema di riproduzione usato.

Gli svantaggi di questa tecnica sono legati a requisiti tecnologici che attualmente la rendono difficilmente applicabile in ottica consumer, ad esempio:

- Necessità di un elevato numero di altoparlanti: l'approssimazione della formulazione di Rayleigh [21] per i sistemi olofonici è infatti valida solo se la distanza tra gli altoparlanti è piccola o paragonabile alla lunghezza d'onda del campo da riprodurre. Questo comporta che gli altoparlanti debbano essere molto vicini tra loro, per riprodurre correttamente anche le frequenze più alte. Per esempio, per ottenere una corretta rappresentazione a 4 kHz è necessario un altoparlante ogni 9 cm. Considerando di dover coprire un'area sufficientemente ampia da accogliere un piccolo gruppo di ascoltatori, questa condizione comporta un numero di altoparlanti dell'ordine delle centinaia solo per rappresentare un'onda piana.
- Necessità di opportune interfacce multicanale, non comunemente diffuse in commercio.
- Alto costo computazionale richiesto per l'elaborazione dell'audio e la sintesi del campo da parte del SAW.

# Capitolo 5

## Ambisonics

L'Ambisonics è una tecnica di sonorizzazione surround introdotta negli anni '70 da Gerzon in seguito a una serie di studi sulla spazializzazione del suono e sulla percezione uditiva [11]. Questa teoria nasce come tecnica di registrazione e riproduzione multicanale per ricreare nell'ascoltatore l'esperienza di un'immersione completa nell'immagine sonora.

### 5.1 Introduzione

La tecnica Ambisonics tratta la scena audio come una sfera di suono proveniente da diverse direzioni attorno a un punto centrale, che coincide con la posizione del microfono in fase di registrazione o con lo sweet spot dell'ascoltatore in fase di ascolto. Mentre nel tradizionale audio multicanale (es. stereo, 5.1, 7.1 surround) ad ogni canale è associato il corrispondente altoparlante, in Ambisonics ogni canale contiene informazioni sul campo acustico.

I segnali Ambisonics sono suddivisi per ordini:

- Ordine 0: segnali che contengono informazioni sulla pressione del campo sonoro registrato.
- Ordine 1 (FOA, first-order Ambisonic): segnali che contengono informazioni sul gradiente di pressione del campo registrato.

- Ordine 2 e superiori (HOA, higher-order Ambisonic): segnali che contengono informazioni sulle derivate della pressione del campo acustico registrato. Permettono alti gradi di direttività.

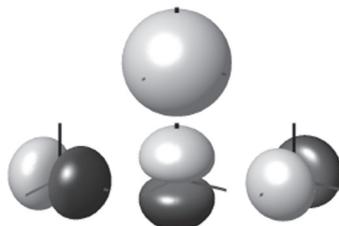


Figura 5.1: Segnale monopolo di un microfono omnidirezionale (in alto) e segnali dipoli di tre microfoni con caratteristica a otto (in basso).

Nei suoi esperimenti [12] utilizzò quattro microfoni cardioidi posti ai vertici di un tetraedro, così da rappresentare un evento sonoro in un punto dello spazio tramite quattro segnali (Figura 5.1):

- 1 segnale monopolo ( $W$ ): componente di ordine 0, rappresenta la componente omnidirezionale del campo.
- 3 segnali dipoli ( $X, Y, Z$ ): componenti di ordine 1, presentano dei diagrammi polari di tipo bidirezionale.

Le caratteristiche direzionali del campo acustico vengono ricostruite sommando le componenti armoniche sferiche del campo stesso, ognuna delle quali dev'essere ripresa da un microfono con le stesse caratteristiche di direttività. Per capire come le armoniche sferiche funzionano e come un campo acustico può essere rappresentato nel dominio delle armoniche sferiche, si possono osservare i pattern dei primi ordini di armoniche: la Figura 5.2 mostra l'ampiezza di armoniche sferiche fino all'ordine 3. In seguito alla registrazione delle armoniche sferiche, i segnali sono combinati tra di loro con opportuni livelli di gain per sintetizzare dei microfoni virtuali con orientamento e direttività desiderati.

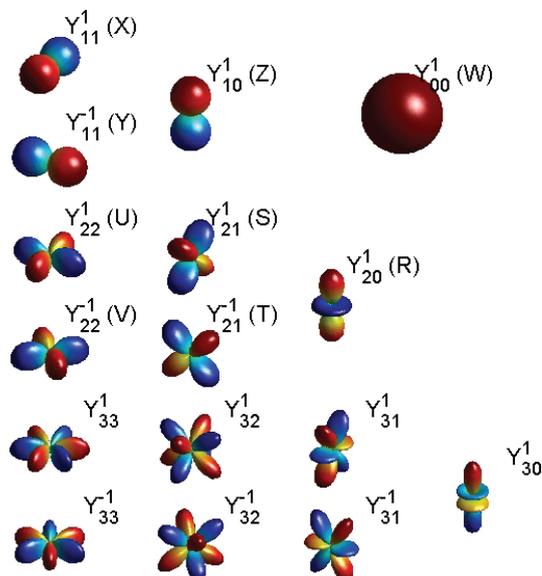


Figura 5.2: Pattern di direttività di armoniche sferiche reali di vari ordini. In rosso le ampiezze positive, in blu le ampiezze negative.

La codifica dei segnali prodotti dal microfono può essere effettuata in diversi modalità, ottenendo i seguenti formati:

- Formato-A: è il formato di base per la registrazione per l'audio Ambisonic. Si ottiene come output di 4 microfoni direttivi disposti ai vertici di un tetraedro, con lo stesso livello di gain. È la registrazione grezza ottenuta da ciascuna delle capsule di un microfono Ambisonic. Questo segnale è specifico per ogni modello di microfono, dal momento che microfoni diversi possono differire nel tipo di capsule e (in leggera misura) nel loro posizionamento.
- Formato-B: usato con microfoni con caratteristiche direttive simili alle armoniche sferiche descritte sopra. Può essere derivato dall'A-format, tramite un'opportuna decodifica, trasformando gli output dell'array tetraedrico in 4 segnali: il primo segnale, W, contiene informazioni sull'ampiezza del segnale, mentre gli altri, X, Y e Z, consentono di definire la direttività manipolando le relazioni di fase tra di loro. Due degli standard di B-format più diffusi sono il meno recente FuMa (Furse-Malham) e ambiX, adottato su piattaforme di distribuzione come Youtube)

- Formato-C: è una sorta di matrice progettata per codificare il contenuto B-format e renderlo adatto alla trasmissione su un numero minore di canali (es. la trasmissione di audio a 2 canali tipica dei sistemi broadcast). Naturalmente, a seconda dei canali disponibili saranno necessari più o meno compromessi sulla compatibilità e sulla risoluzione direzionale dell'audio Ambisonic. Ad oggi questo formato è praticamente in disuso, data la facilità con cui si riesce trasmettere un segnale a 4 canali a piena larghezza di banda.
- Formato-D: è il segnale B-format decodificato e pronto per la riproduzione in un determinato array di altoparlanti. Può trattarsi di un layout convenzionale (es. 5.1 o 7.1 surround), di rappresentazioni binaurali [28] oppure di configurazioni meno comuni.

Il formato Ambisonic più popolare, largamente diffuso in realtà virtuale e nei video a 360°, è il formato a 4 canali chiamato Ambisonics B-format, che sta alla base del cosiddetto First-order Ambisonics (FOA). Tale B-format è indipendente dal sistema usato per la riproduzione. Infatti, a partire dal B-Format è possibile ricavare i segnali corrispondenti ad una data configurazione di altoparlanti applicando delle matrici che ottimizzano la ricostruzione dell'esperienza d'ascolto reale basandosi su criteri psicoacustici. La tecnica FOA utilizza microfoni in grado di campionare le componenti direzionali di ordine 0 (microfoni a pressione, che catturano soltanto le componenti di pressione dell'onda sonora) e dell'ordine 1 (microfoni a gradiente di pressione, che catturano la velocità della particelle dell'onda sonora nel mezzo di propagazione).

## 5.2 La teoria Ambisonic

La teoria Ambisonic si basa sulla ricostruzione del campo sonoro attraverso onde piane, a partire dalle sue armoniche sferiche, il cui sviluppo permette di modellare la direttività e la spazializzazione di un suono. L'ipotesi di base è che le sorgenti siano considerate puntiformi e che emettano solo onde piane.

Per lo sviluppo che segue [24] si consideri un sistema in coordinate cilindriche  $(r, \phi, z)$  in cui per semplicità si ha  $z = 0$ . Il campo sonoro di riferimento  $P_{ref}$ , in un punto  $\mathbf{r} = (r, \phi)$ , generato da un'onda piana che arriva dalla direzione  $\Psi$  può essere descritto dall'equazione

$$P_{ref}(r, \phi) = P_{\Psi} \cdot e^{j\mathbf{k}_n \cdot \mathbf{r}} = P_{\Psi} \cdot e^{jkr \cos(\phi - \Psi)} \quad (5.1)$$

dove  $\mathbf{k}$  rappresenta un'onda piana con numero d'onda  $k = \frac{2\pi}{\lambda}$  e la direzione d'arrivo  $\Psi$ ,  $\mathbf{r}$  rappresenta il punto di ascolto. Durante la riproduzione,  $N$  altoparlanti sono posizionati simmetricamente su una circonferenza (Figura 5.3). È necessario che il raggio della circonferenza sia maggiore della minima lunghezza d'onda che si vuole riprodurre. Il generico altoparlante  $n$  emette un'onda piana  $P_n$  la cui equazione è

$$P_n(r, \phi) = P_n \cdot e^{j\mathbf{k}_n \cdot \mathbf{r}} = P_n \cdot e^{jkr \cdot \cos(\phi - \phi_n)}. \quad (5.2)$$

dove  $\phi_n$  è l'angolo da cui l'altoparlante sta emettendo.

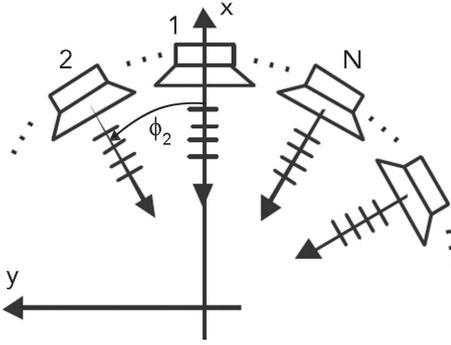


Figura 5.3: Disposizione simmetrica di  $N$  altoparlanti.

La sovrapposizione delle singole onde nel punto di ascolto è descritta dall'onda  $P_{Ambisonic}$  di equazione

$$P_{Ambisonic}(r, \phi) = \sum_{n=1}^N S_n(r, \phi) = \sum_{n=1}^N P_n \cdot e^{jkr \cdot \cos(\phi - \phi_n)}$$

Per ricreare il campo sonoro di partenza, è necessario imporre il seguente vincolo

$$P_{Ambisonic}(r, \phi) = S_{ref}(r, \phi). \quad (5.3)$$

Le onde piane sono caratterizzate da una pressione costante e da un termine di fase che può essere descritto come sviluppo della serie di Bessel-Fourier [20], per cui l'eq. (5.1) diventa

$$\begin{aligned}
S_{ref}(r, \phi) &= P_{\Psi} J_0(kr) + 2P_{\Psi} \sum_{m=1}^{\infty} i^m J_m(kr) \cos[m(\phi - \Psi)] \\
&= P_{\Psi} \left\{ J_0(kr) + 2 \sum_{m=1}^{\infty} i^m J_m(kr) [\cos(m\phi) \cos(m\Psi) + \sin(m\phi) \sin(m\Psi)] \right\}
\end{aligned}$$

e l'eq. (5.2) diventa

$$\begin{aligned}
S_{Ambisonic}(r, \phi) &= \\
&= \sum_{n=1}^N P_n \left\{ J_0(kr) + 2 \sum_{m=1}^{\infty} i^m J_m(kr) [\cos(m\phi) \cos(m\phi_n) + \sin(m\phi) \sin(m\phi_n)] \right\}
\end{aligned}$$

L'equazione di eguaglianza (5.3) è verificata per le seguenti condizioni, dette matching conditions

$$\begin{aligned}
P_{\Psi} &= \sum_{n=1}^N P_n \\
P_{\Psi} \cos(m\Psi) &= \sum_{n=1}^N P_n \cos(m\phi_n) \\
P_{\Psi} \sin(m\Psi) &= \sum_{n=1}^N P_n \sin(m\phi_n)
\end{aligned} \tag{5.4}$$

Il sistema teorico contiene un numero infinito di equazione; in pratica però il numero di equazioni è limitato dall'ordine di sistema  $m$  (valore finito). Più alto l'ordine di sistema, maggiore la precisione con cui il campo viene ricostruito e maggiore l'estensione dell'area di ascolto corretto.

### 5.3 Codifica di Ambisonic

Le eq. (5.4) contengono tutte le informazioni necessarie per la codifica. I segnali ottenuti, detti segnali Ambisonic, contengono le informazioni sull'ampiezza e sulla direzione della sorgente. Per esempio, i segnali Ambisonic per una sorgente  $S$  di direzione  $\Psi$  son dati da

$$\begin{aligned}
m = 0 \quad W &= S/\sqrt{2}; \\
m = 1 \quad X &= S \cdot \cos \Psi \quad Y = S \cdot \sin \Psi \\
m = 2 \quad U &= S \cdot \cos 2\Psi \quad X = S \cdot \sin 2\Psi \\
&\dots \quad \dots \quad \dots
\end{aligned} \tag{5.5}$$

La notazione  $(W, X, Y, \dots)$  è la convenzione comune con cui si definiscono i segnali di un sistema Ambisonic. Nel caso 2D le componenti dell'asse  $z$  sono nulle, per cui per  $m = 0$  si ha  $Z = 0$ , per  $m = 1$  si ha  $R = S = 0$  e così via.

## 5.4 Decodifica di Ambisonic

Per ottenere le equazioni di decodifica è utile riscrivere le eq. (5.4) in forma matriciale:

$$b = A \cdot x$$

dove  $b = [W, X, Y, U, V, \dots]^T$  rappresenta il segnale Ambisonic,

$$A = \begin{bmatrix}
1 & 1 & \dots & 1 \\
\cos \phi_1 & \cos \phi_2 & \dots & \cos \phi_N \\
\sin \phi_1 & \sin \phi_2 & \dots & \sin \phi_N \\
\vdots & \vdots & \ddots & \vdots \\
\cos m\phi_1 & \cos m\phi_2 & \dots & \cos m\phi_N \\
\sin m\phi_1 & \sin m\phi_2 & \dots & \sin m\phi_N
\end{bmatrix}$$

è la matrice definita dalla disposizione dell'array di altoparlanti e

$$x = [P_1, P_2, P_3, \dots, P_N]^T$$

è il vettore incognito che contiene le ampiezze  $P_n$  da assegnare ad ogni altoparlante. Per risolvere il sistema si ricorre all'equazione

$$x = A^T \cdot (A \cdot A^T) \cdot b$$

Le ampiezze  $P_n$  sono trovate pesando i segnali Ambisonic in base alla posizione del rispettivo altoparlante. Se l'array di altoparlanti è disposto simmetricamente su una circonferenza, si otterrà

$$P_i = \frac{1}{N} (W + 2X \cos \phi_i + 2Y \sin \phi_i + 2U \cos 2\phi_i + 2V \sin 2\phi_i + \dots)$$

Un vincolo da imporre nella decodifica è che il numero di altoparlanti sia maggiore del numero di segnali Ambisonic codificati e che l'array di riproduzione sia il più simmetrico possibile, altrimenti diventa necessaria una parametrizzazione delle ampiezze  $P_n$ . Un segnale HOA di ordine  $i$  ha infatti bisogno di almeno  $(i + 1)^2$  altoparlanti per la riproduzione. Quindi per un segnale di ordine 2 sono necessari almeno 9 altoparlanti, per un segnale di ordine 4 almeno 25 e così via.

## 5.5 Array microfonicici

Lo stato dell'arte per la cattura del campo sonoro tridimensionale prevede l'utilizzo di moderni array microfonico di tipo coincidente a simmetria sferica, basati sulla decomposizione del campo in armoniche sferiche. Nell'utilizzare configurazioni a capsule coincidenti (es. microfoni tetraedrici come il microfono SoundField) si osservano però determinati problemi di filtering, dovuti alla distanza tra le capsule (che nella realtà non possono essere perfettamente coincidenti) e alla direzione di arrivo del suono. A tale riguardo a questo tipo di microfoni viene applicata un'equalizzazione tale da appiattire la risposta all'incidenza del campo sonoro [12]. Rimane però il principale limite dell'aliasing spaziale che si verifica quando la lunghezza d'onda del suono è paragonabile alla distanza tra le capsule. Trattandosi in generale di distanze di pochi centimetri, i primi problemi di aliasing si osservano da circa 8 kHz in su e causano una deviazione dalle forme ideali delle caratteristiche polari delle capsule.

Più recentemente sono stati costruiti microfoni HOA dotati di un grande numero di trasduttori di pressione disposti sulla superficie di una sfera (es: il microfono Eigenmike). Trattandosi di sensori più piccoli di quelli usati nei microfoni tetraedrici, è possibile garantire una migliore coincidenza alle alte frequenze, una bassa diffrazione del suono e una minore interferenza col campo sonoro. Il risultato è una figura polare più consistente lungo tutto lo spettro di frequenze, rispetto ai microfoni tetraedrici. Questa caratteristica rende questo tipo di array microfonico adatti ad una ripresa spaziale più accurata.

### 5.5.1 SoundField ST450

Il microfono SoundField<sup>1</sup> è il diretto successore del microfono Ambisonic [8] ed è stato tra i primi strumenti fisici in grado di mettere in pratica questa tecnologia. Si tratta di un array microfonico composto da 4 capsule sub-cardioidi (quindi con una figura polare leggermente più omnidirezionale che cardioidi) disposte ai vertici di una superficie a tetraedro. (Figura 5.4), tale da ricreare al meglio la condizione di capsule coincidenti nello spazio. Le quattro capsule sono identificate dalla direzione verso cui puntano: left-front (LF), right-front (RF), left-back (LB) e right-back (RB). Questa convenzione (Figura 5.4a) è intesa col microfono in verticale, con l'asta di supporto che punta verso il pavimento e col logo *SoundField* rivolto verso l'osservatore. Tuttavia per un corretto funzionamento non è richiesto alcun orientamento specifico.

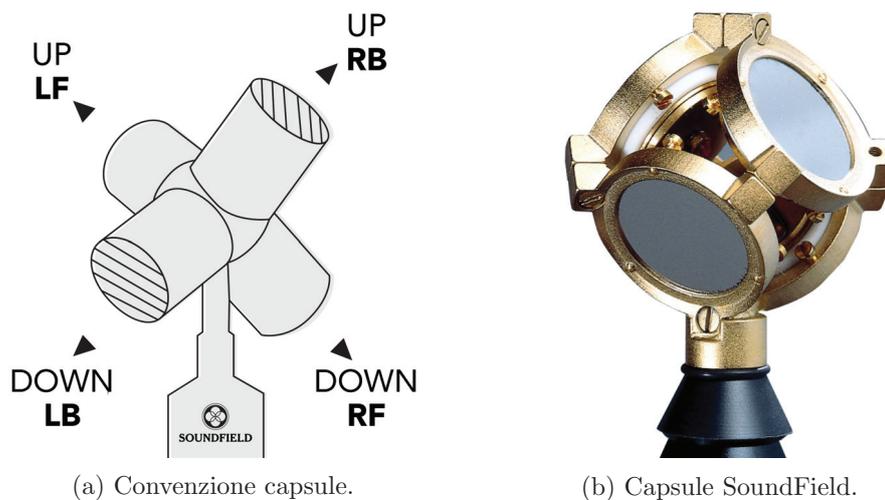


Figura 5.4: Disposizione delle capsule a struttura tetraedrica.

Questa configurazione consente la registrazione in A-format, da cui è possibile ottenere le componenti del B-format tramite una serie di semplici elaborazioni

---

<sup>1</sup>[www.soundfield.com](http://www.soundfield.com)

dei segnali delle 4 capsule:

$$X = 0,5[(LF - LB) + (RF - RB)]$$

$$Y = 0,5[(LF - RB) - (RF - LB)]$$

$$Z = 0,5[(LF - LB) + (RB - RF)]$$

$$W = 0,5(LF + LB + RF + RB)$$

Tuttavia i seguenti fattori fisici influenzano la ricostruzione dell'immagine sonora:

- Distanza tra le capsule non ideale: solitamente compresa tra 3 e 10 cm, dovrebbe però essere inferiore a 5 mm. Questo produce un fenomeno di aliasing dalle frequenze medio-alte in poi.
- Figura polare delle capsule non ideale: l'irregolarità dei diagrammi polari, soprattutto alle alte frequenze, può causare un notevole degrado nella ripresa di suoni acuti.

Per ovviare in parte a queste problematiche, i segnali d'output vengono opportunamente equalizzati in modo da simulare una migliore coincidenza delle capsule, almeno fino a una certa frequenza.

Il vantaggio di registrare in A-format e successivamente ottenere il B-format sta nella possibilità di modificare elettronicamente la posizione virtuale del microfono (es. capovolgerlo a testa in giù o modificare la direzione verso cui è puntato) o le sue caratteristiche polari, semplicemente alterando l'elaborazione dei segnali. In questo modo, nonostante il posizionamento fisico del microfono sia importante da un punto di vista acustico, è possibile ottimizzarne la posizione virtuale in post-produzione. I parametri spaziali, (Figura 5.5) su cui è possibile agire sono, come definiti dal costruttore:

- Azimuth: controlla la direzione orizzontale virtuale verso cui il microfono sta puntando. Può essere modificato processando le componenti  $X$  e  $Y$  del B-format con funzioni di controllo sinusoidali, quindi combinando linearmente i risultati in modo da ottenere i nuovi valori  $X'$  e  $Y'$ . Per esempio, sia  $\theta$  l'offset azimutale, allora

$$X' = X \cos \theta + Y \sin \theta$$

$$Y' = Y \cos \theta - X \sin \theta$$

- Elevation: controlla la direzione verticale virtuale verso cui il microfono sta puntando. Analogamente all'Azimuth, può essere modificato agendo sulle componenti  $Z$  e  $Y$  o  $Z$  e  $X$  del B-format.
- Dominance: determina la precedenza frontale o posteriore. Agisce come una sorta di zoom che permette di favorire la cattura di suoni provenienti dalla direzione frontale o dalla direzione posteriore.

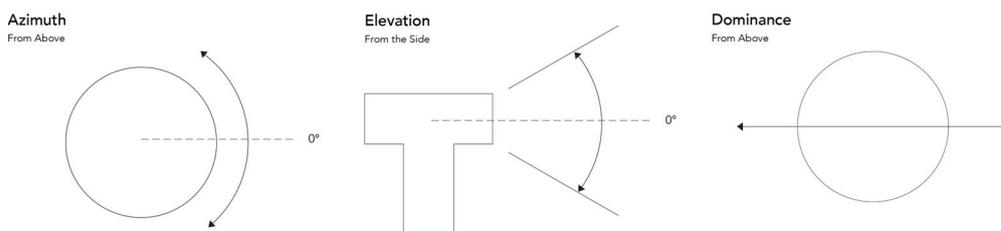


Figura 5.5: Parametri di puntamento.

Nel caso del microfono SoundField ST450 [25] la conversione dei segnali da A-format a B-format viene effettuata dall'unità di controllo abbinata ST450 MKII, a cui il microfono è collegato tramite un cavo Lemo a 12 pin. In output è possibile inviare i segnali in B-format ad un computer attraverso due cavi XLR a 5 pin, uno per le componenti  $X$  e  $W$  e l'altro per le componenti  $Y$  e  $Z$ . In Figura 5.6 è mostrato lo schema di configurazione base per l'utilizzo del ST450. Il software necessario per manipolare il campo acustico del B-format è il plug-in SoundField SurroundZone (Figura 5.7), integrabile nella propria DAW d'utilizzo. Questo strumento consente un'estrema versatilità nel riposizionamento dei microfoni virtuali in post-produzione e restituisce il prodotto finale per qualunque formato d'ascolto standard richiesto.



Figura 5.6: Setup base per l'utilizzo di SoundField ST450.

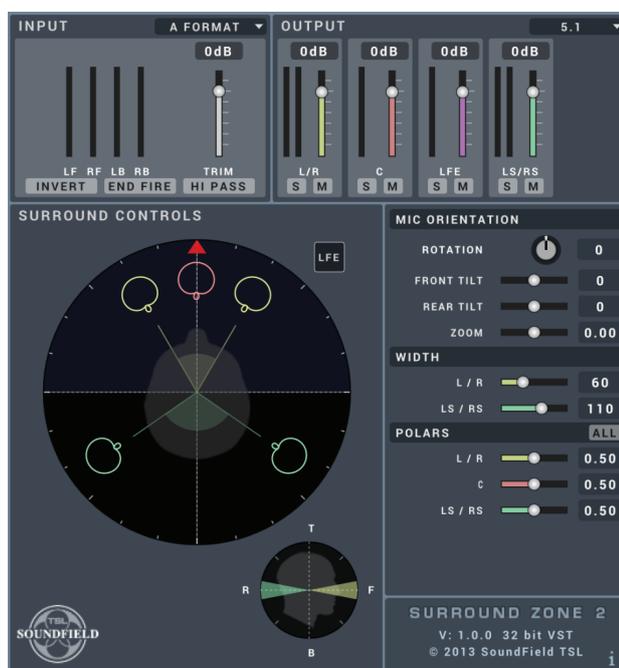


Figura 5.7: Interfaccia grafica del plug-in SoundField SurroundZone.

### 5.5.2 Sennheiser Ambeo VR 3D

Un'alternativa più recente e più economica al SoundField è il microfono Sennheiser Ambeo VR 3D<sup>2</sup>, parte della tecnologia Ambeo sviluppata da Sen-

<sup>2</sup><https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic>

nheiser per la registrazione e la riproduzione di audio 3D immersivo. La struttura e il principio di funzionamento dell'Ambeo VR 3D è praticamente analogo a quello dei microfoni SoundField, con 4 capsule cardioidi disposte ai vertici di un tetraedro (Figura 5.8), quindi predisposte per una ripresa in A-format.



Figura 5.8: Sennheiser Ambeo VR 3D.

Anche in questo caso il microfono presenta un connettore d'output a 12 pin a cui va collegato il cavo in dotazione, composto a un estremo con connettore DIN12 e all'altro estremo da 4 connettori XLR a 3 pin corrispondenti alle 4 capsule. I 4 XLR sono quindi collegati ad un'opportuna interfaccia audio in grado di assicurare la linearità di fase tra i segnali, fornire +48V di alimentazione a ciascuna capsula e su cui è possibile selezionare l'orientamento del microfono (upright, upside down o end-fire). Per garantire una ripresa d'ambiente più realistica possibile bisogna inoltre assicurarsi che i livelli di gain dei 4 segnali siano perfettamente identici tra loro, dal momento che anche una piccola differenza può influenzare l'immagine spaziale. Una volta effettuata la registrazione del campo sonoro in A-format, i segnali vengono codificati in B-format tramite il software plug-in proprietario Ambeo (Figura 5.9), che consente di modificare la posizione e di ruotare virtualmente il microfono nello spazio, applicare filtri di equalizzazione ai segnali e selezionare il formato d'output (tra FuMa e ambiX).

Sulla carta l'Ambeo VR 3D e i microfoni SoundField presentano caratteristiche simili e il fatto di condividere la stessa tecnologia di base comporta anche lo stesso limite nella rappresentazione dell'altezza dei suoni ripresi, che non risulta ben definita quanto la rappresentazione dei suoni sul piano orizzontale. Il costo attuale dell'Ambeo è decisamente vantaggioso rispetto

a quello dei microfoni SoundField<sup>3</sup>, tuttavia trattandosi di un prodotto ancora molto recente non sono reperibili confronti obiettivi tra i due modelli tali da stabilire quale dei due abbia complessivamente un migliore rapporto qualità-prezzo.

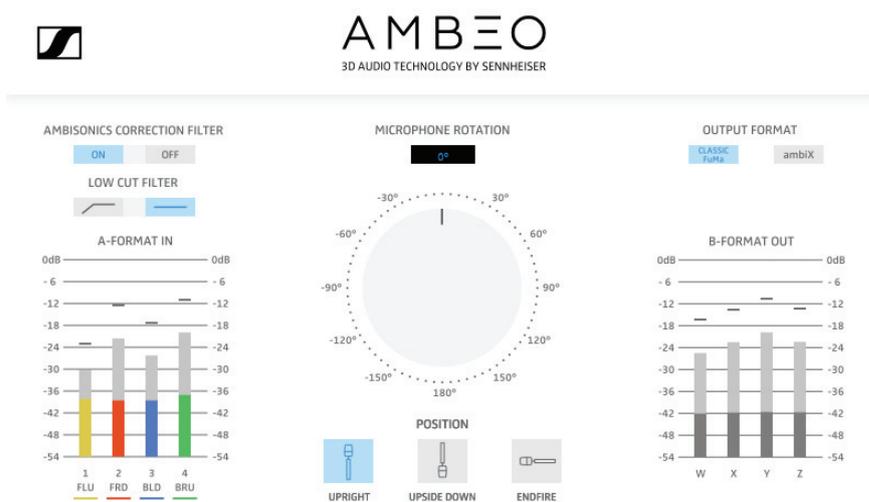


Figura 5.9: Interfaccia grafica del plug-in Ambeo.

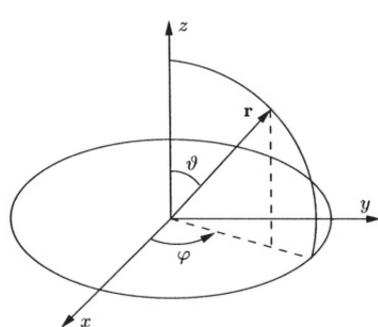
### 5.5.3 mh acoustics em32 Eigenmike

L'em32 Eigenmike di mh acoustics<sup>4</sup> è formato da una sfera rigida con un diametro di 8,4 cm che comprende 32 capsule microfoniche da  $\frac{1}{2}$ ", disposte omogeneamente sulla superficie. Con questo alto numero di sensori il microfono è in grado di registrare il campo sonoro con una grande precisione spaziale: i 32 segnali grezzi possono infatti essere elaborati per formare i cosiddetti eigenbeam, un sistema di microfoni virtuali a fascio (o beam pattern) fino all'ordine 4 Ambisonic, con 25 canali HOA, rappresentabili tramite le funzioni armoniche sferiche precedentemente descritte. Nell'utilizzo dell'array Eigenmike sono adottate le seguenti convenzioni per un sistema di coordinate sferiche (Figura 5.10a):

<sup>3</sup>SoundField ST450 (7561\$) vs Ambeo VR 3D (1649\$); [www.creativefieldrecording.com/2017/03/30/ambisonic-field-recording-resources/](http://www.creativefieldrecording.com/2017/03/30/ambisonic-field-recording-resources/)

<sup>4</sup>[www.mhacoustics.com](http://www.mhacoustics.com)

- Angolo verticale,  $\vartheta$ : è la direzione sul piano verticale; può variare tra  $0^\circ$  e  $180^\circ$ <sup>5</sup>. A  $0^\circ$  corrisponde la normale all'estremità della sfera opposta all'asta di supporto, coincidente con l'asse  $Z$  in Figura 5.10b (se si orienta il microfono con l'asta verso il pavimento, tale direzione punta al soffitto). A  $90^\circ$  corrisponde la direzione del piano orizzontale e a  $180^\circ$  la direzione dell'asta (tipicamente verso il pavimento).
- Angolo orizzontale (o azimuth),  $\phi$ : è la direzione sul piano orizzontale; può variare tra  $0^\circ$  e  $360^\circ$ . A  $0^\circ$  corrisponde la normale al logo *mh acoustics* sul supporto, coincidente con l'asse  $X$  in figura. Quest'angolo cresce in senso anti-orario guardando la sfera dall'alto.



(a) Sistema di coordinate sferiche.



(b) Eigenmike em-32 e i suoi assi.

Figura 5.10: Convenzione di riferimento.

A causa della geometria fissa dei sensori sulla sfera, esiste una frequenza spaziale di Nyquist oltre cui i beampattern mostrano un significativo aliasing, tale da provocare una degradazione delle caratteristiche polari desiderate. Per il modello em32, il limite di Nyquist è approssimativamente 8 kHz. Tutti gli eigenbeam sono calcolati in modo da avere una risposta in frequenza nominalmente piatta in ampiezza fino a questo limite. In seguito il comportamento spettrale diventa più complesso.

In Figura 5.11 è mostrato parte dello spettro per eigenbeam fino all'ordine 4, con  $\vartheta = 0^\circ$  e  $\phi = 0^\circ$ . Si osserva che man mano che aumenta l'ordine

<sup>5</sup>In certi casi l'angolo verticale può variare tra  $-90^\circ$  (pavimento) e  $90^\circ$  (soffitto), con  $0^\circ$  corrispondente al piano orizzontale.

degli eigenbeam codificati, cresce anche la frequenza di cutoff, ossia si riduce il range di frequenze su cui un certo ordine opera. Tale strategia è stata adottata dai costruttori per limitare il rumore generato dal microfono stesso durante il processo di codifica.

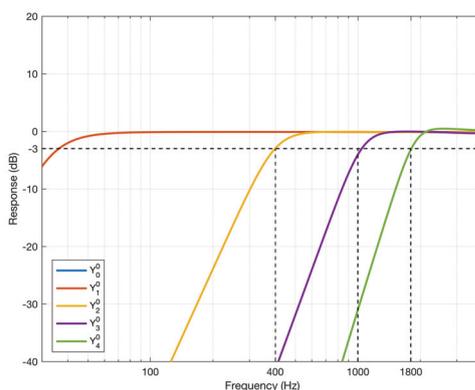


Figura 5.11: Risposta in frequenza degli Eigenbeam fino all'ordine 4, con  $\vartheta = 0^\circ$  e  $\phi = 0^\circ$ . È stata applicata una normalizzazione per facilitare l'interpretazione.

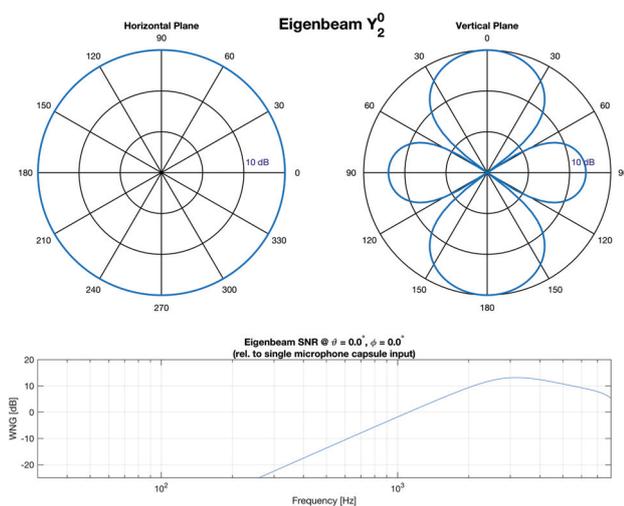


Figura 5.12: Figure polari e grafico SNR per gli eigenbeam di ordine 2.

In Figura 5.12 sono mostrati il grafico SNR e il diagramma polare [5] per gli eigenbeam di ordine 2, per  $\vartheta = 0^\circ$  e  $\phi = 0^\circ$ . In particolare, il grafico SNR mostra l'andamento in frequenza del SNR di codifica del rispettivo eigenbeam in relazione al SNR di un microfono con una singola capsula: il WNG (white-noise-gain), in ordinata, è un parametro utile per quantificare il SNR d'output del codificatore, quindi per stimarne la qualità. I grafici di tutti gli altri ordini sono disponibili sul sito web del costruttore.

La signal chain di un Eigenmike è dunque composta dai seguenti step (Figura 5.13):

1. Successivamente alla fase di registrazione, i 32 segnali microfonici grezzi sono soggetti ad una calibrazione tale da compensare le differenze d'ampiezza tra le varie capsule e migliorare la successiva codifica in eigenbeam. I valori di calibrazione sono ottenuti sperimentalmente e vengono memorizzati nel firmware interno di ciascun array.
2. I 32 segnali sono quindi inviati a una DAW o altro software in utilizzo, dove il Modal Beamformer effettua la codifica in eigenbeam e li combina linearmente per ottenere i pattern polari desiderati. Il beamformer consente inoltre di direzionare i beampattern nello spazio 3D. Le configurazioni disponibili sono le seguenti:
  - Omnidirezionale: uguale sensibilità in tutte le direzioni
  - Dipolo (ordine 1): sensibilità massima a  $0^\circ$  e  $180^\circ$  e nulla a  $90^\circ$  e  $270^\circ$
  - Cardioide (ordine 1-4): sensibilità nulla a  $180^\circ$
  - Supercardioide (ordine 1-4): massimo rapporto di potenza tra lobo frontale e lobo posteriore.
  - Ipercardioide (ordine 1-4): massima direttività
3. Parallelamente al processo di codifica, è implementato un filtro passa-alto (high-frequency extension) per trattare i segnali superiori agli 8 kHz e linearizzare la risposta in frequenza, che come visto in precedenza diventa problematica oltre il limite spaziale di Nyquist, intrinseco della sfera.
4. Infine, l'output del beamformer e del HF extension si combinano tramite un crossover e si ottiene così il segnale d'output a banda intera.

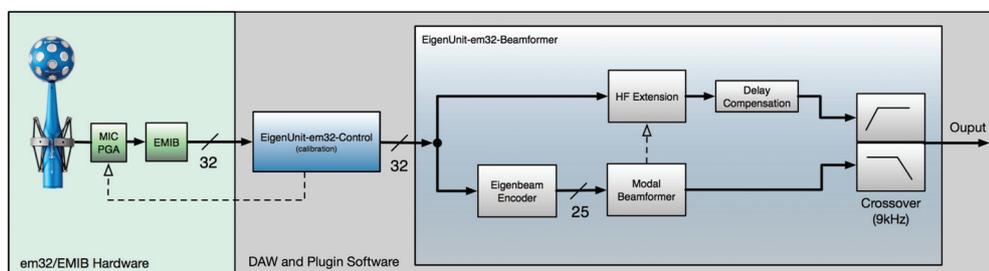


Figura 5.13: Signal chain dell'Eigenmike.

All'interno della sfera sono integrati preamplificatori microfonici e convertitori A/D a 24 bit. In questo modo il sistema fisico risulta estremamente pratico, non richiedendo preamplificatori e convertitori esterni; è infatti necessario soltanto un cavo standard CAT-5 (o superiore) per connettere l'array alla fonte di alimentazione esterna e all'interfaccia mh acoustics EMIB (Eigenmike Microphone Interface Box). L'EMIB, in Figura 5.14, converte i segnali microfonici proprietari in uno stream audio a protocollo Firewire, accessibile tramite computer come un driver a 32 canali. Tali output sono quindi gestiti tramite DAW o altri software, ad esempio mh acoustics EigenStudio o il sistema 3D-VMS, sviluppato da una collaborazione tra CRIT (Centro Ricerche Rai) e l'Università di Parma. Opzionalmente, una telecamera può essere piazzata sul microfono in direzione del suo asse  $X$  positivo, in modo da dare un riferimento visivo che aiuti a posizionare correttamente i microfoni virtuali su una scena estesa a  $360^\circ$ , durante la post-produzione.



Figura 5.14: Setup base per l'utilizzo dell'Eigenmike em-32.

Di recente la Rai ha sviluppato e costruito il proprio array microfonico per registrare uno stream Ambisonic, a partire dal principio dell'Eigenmike: il RAI-AIDA [23] (Figura 5.15). Le 32 capsule e l'elettronica dell'Eigenmike sono stati smontati e ridisposti in un cilindro alto circa 27 cm e con base di diametro 8 cm. Tale configurazione ha permesso di migliorare la ripresa di informazioni azimutali, dando una maggiore direttività (quindi una maggiore distinzione dei suoni) su sorgenti sonore poste l'una di fianco all'altra.



Figura 5.15: Microfono Rai-AIDA.

## 5.6 Vantaggi e svantaggi

Per riassumere, i vantaggi della tecnica Ambisonics sono:

- Possibilità di codifica di sorgenti arbitrarie.
- Facilità di rotazione dell'intero campo sonoro (adatto per esempio per applicazione VR a 360°)
- Totale indipendenza dal formato di riproduzione. I canali Ambisonic non hanno niente a che fare con i canali delle configurazioni d'ascolto tradizionale (es. stereo e 5.1 surround), ma sono invece basati su principi fisici del campo acustico.
- Numero fisso di canali: a differenza del WFS in cui il numero di canali di riproduzione dipende dal tipo di soggetto sonoro da rappresentare (object-based), Ambisonics usa un numero predefinito di canali che può eventualmente essere adatto ad un metodo object-based.

- Offre un'esperienza d'ascolto omogenea: i panning sono rappresentati in modo costante, senza evidenti “salti” e artifici. Questo garantisce una buona sensazione d'immersione nel campo sonoro.
- Richiede un numero moderato di altoparlanti, a differenza della WFS che invece può richiederne centinaia.

I principali svantaggi invece sono:

- Scarse proprietà direttive: specialmente per segnali di ordine 1 (cardioidi). Per segnali HOA il risultato è invece molto più accettabile.
- Scarsa distinzione dell'altezza dei suoni.
- Piccolo sweet spot: l'area d'ascolto ottimale ristretta rende questa tecnologia inadatta alla riproduzione in grandi sale. Questo problema può però essere ottimizzato con l'utilizzo di segnali d'ordine maggiore e opportune decodifiche psicoacustiche.
- Complessità di codifica: l'utilizzo di array microfonicici con capsula quasi coincidenti rende difficile il processo di codifica dei segnali, a causa di problematiche sopra citate (es. aliasing e comb-filtering alle alte frequenze); l'applicazione di opportuni equalizzazioni e filtri di correzioni permette di minimizzare questi difetti, aumentando però il costo computazionale.
- Complessità di decodifica: la decodifica dei segnali può risultare particolarmente complessa nel caso di configurazioni di destinazione non comuni. Inoltre la complessità aumenta con l'ordine del segnale da codificare.

# Capitolo 6

## Il 3D-VMS

Il 3D-VMS (3D Virtual Microphone System) è un sistema sviluppato da Scoppece e Farina [7] attorno all'Eigenmike, microfono sferico con 32 capsule ad alta qualità. I 32 segnali vengono filtrati tramite un processore di convoluzione, in grado di sintetizzare fino a 7 microfoni virtuali direttivi in realtime e fino a  $n \times 7$  in playback. Per mezzo di un'interfaccia grafica e di un'immagine panoramica è possibile variare il puntamento e l'angolo di cattura di ciascun microfono virtuale.

### 6.1 Descrizione del sistema

Considerando gli  $M$  segnali d'ingresso, provenienti dalle capsule microfoniche, da convertire in  $V$  segnali d'uscita, corrispondenti ai microfoni virtuali (ciascuno con propri direttività e puntamento), è necessario utilizzare una matrice di  $M \times V$  filtri FIR (Figura 6.1). Sia  $x_m$  il segnale d'input degli  $M$  microfoni,  $y_v$  il segnale d'output dei  $V$  microfoni virtuali e  $h_{m,v}$  la matrice di filtri, i segnali processati si possono esprimere come

$$y_v(t) = \sum_{m=1}^M x_m(t) * h_{m,v}(t)$$

Quindi, ogni microfono virtuale è ottenuto come somma delle convoluzioni tra gli  $M$  segnali d'input e un set di  $M$  opportuni filtri FIR, ottenuti empiricamente.

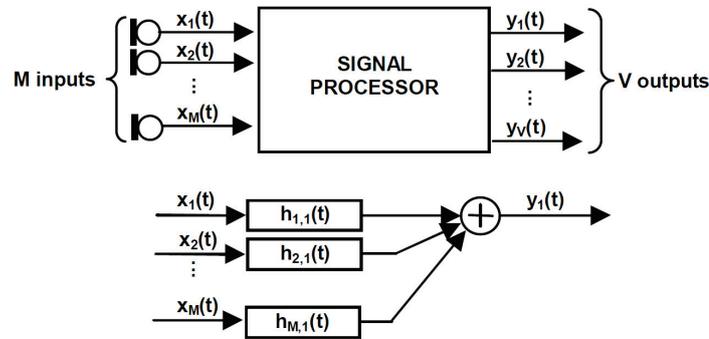


Figura 6.1: Schema di elaborazione dei segnali.

In generale, le tecnologie destinate a registrare e riprodurre la spazialità di un suono si basano su array di microfoni e di altoparlanti e su rappresentazioni matematiche del campo sonoro, che viene scomposto in onde piane, in armoniche sferiche (HOA) o in funzioni complesse di Hankel. L'approccio alla base del 3D-VMS invece definisce una matrice di filtraggio digitale non basata soltanto su analisi teoriche (ad esempio la teoria Ambisonics), ma anche su misure effettuate in camera anecoica. I segnali d'output sono ottenuti direttamente filtrando i segnali d'input, senza la codifica e decodifica tipica dell'Ambisonics.

### 6.1.1 Calcolo dei coefficienti del filtro

La misurazione avviene in camera anecoica, dove il microfono è posto al centro di un array di altoparlanti (Figura 6.2). Si misurano i coefficienti di risposta all'impulso di ciascuna capsula per ognuna delle sorgenti sonore poste nelle  $D$  posizioni e si ottiene una matrice  $M \times D$ .

Si esegue quindi un'ottimizzazione dei coefficienti<sup>1</sup> in modo da ottenere dei risultati il più vicino possibile alle risposte ideali e per correggere eventuali deviazioni e artefatti acustici avvenuti durante la ripresa microfonica (es: schermature, diffrazioni, riflessioni, ecc.). Questo approccio consente di sintetizzare microfoni virtuali a cardioidi di ordine superiore, indipendenti dalla frequenza, come mostra la Figura 6.3.

<sup>1</sup>Ad esempio, tramite il metodo dei minimi quadrati, che consente di trovare una

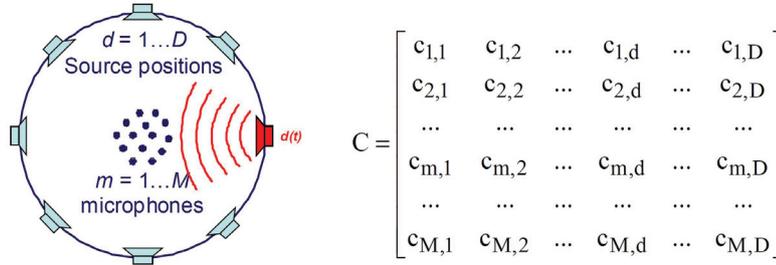


Figura 6.2: Matrice di coefficienti delle IR degli M microfoni.

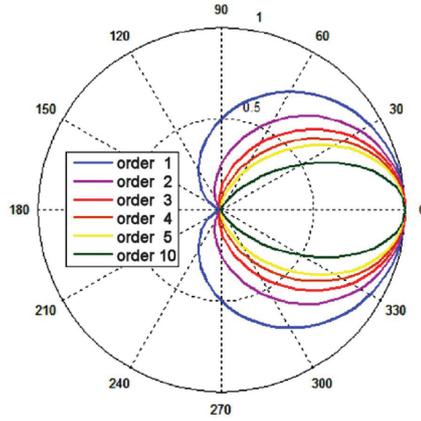


Figura 6.3: Diagrammi polari di microfoni cardioidi virtuali di ordini vari.

Il fattore di direttività  $Q$  di un microfono virtuale di ordine  $n$  può essere descritto in coordinate sferiche come segue

$$Q_n(\vartheta, \phi) = [Q_1(\vartheta, \phi)]^n$$

dove  $Q_1(\vartheta, \phi)$  è il fattore di direttività di un microfono cardioide del primo ordine, pari a

$$Q_1(\vartheta, \phi) = 0,5 + 0,5 \cdot \cos(\vartheta) \cos(\phi)$$

---

funzione che più si avvicina ad un insieme di dati [16].

I filtri di processamento  $h$  sono quelli che trasformano la risposta all'impulso misurata  $c$  nella risposta all'impulso teorica  $p$ :

$$\sum_{m=1}^M c_{m,d} * h_m \Rightarrow p_d \quad d = 1, \dots, D$$

dove la risposta all'impulso ideale  $p_d$  può essere ottenuta applicando il fattore di direttività  $Q$  ad una delta di Dirac  $\delta$  ritardata nel tempo,  $p_d = Q_d \delta$ . Passando nel dominio delle frequenze, la convoluzione diventa una semplice moltiplicazione, per cui si ha

$$\sum_{m=1}^M C_{m,d,k} \cdot H_{m,k} \Rightarrow P_d \quad \text{indice di frequenza, } k = 0, \dots, N/2$$

Raccogliendo i valori di  $C, H$  e  $P$  in opportune matrici, in cui si considerano tutti gli  $M$  segnali d'ingresso, tutte le  $D$  direzioni delle sorgenti e tutti i  $V$  segnali da creare, si ottiene

$$[H_k]_{M \times V} = \frac{[P]_{D \times V}}{[C_k]_{D \times M}}$$

Al fine di produrre filtri causali, la delta di Dirac  $\delta$  viene ritardata di  $n_0$  campioni, che nel dominio delle frequenze corrisponde a

$$\delta_k = e^{-j2\pi k \frac{n_0}{N}}$$

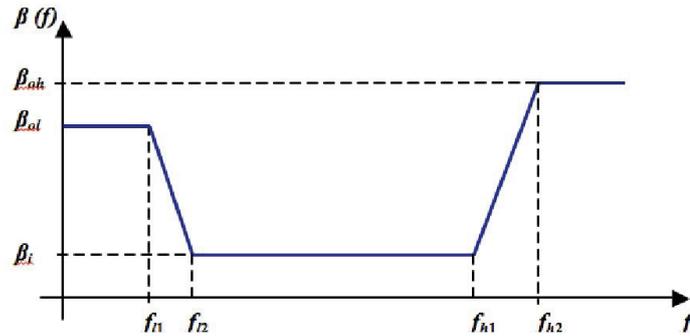


Figura 6.4: Parametro di regolarizzazione in funzione della frequenza.

Scegliendo  $n_0 = \frac{N}{2}$  e inserendo un parametro di regolarizzazione  $\beta_k$  al denominatore, per evitare un'eccessiva enfasi alle frequenze a cui il segnale è basso, si ottiene

$$[H_k]_{M \times V} = \frac{[C_k]_{M \times D}^* \cdot [Q]_{D \times V}}{[C_k]_{M \times D}^* \cdot [C_k]_{D \times M} + \beta k \cdot [I]_{M \times M}}$$

Il parametro  $\beta$  è dipendente dalla frequenza ed è stato scelto in modo da assumere valori bassi e costanti per le frequenze di maggiore utilizzo del microfono e valori più alti alle bassissime e altissime frequenze, dove sono più probabili delle instabilità numeriche (Figura 6.4).

### 6.1.2 Sintesi dei microfoni virtuali

Il calcolo della matrice di filtri viene effettuato tramite uno script di Matlab, a cui si passano il numero di microfoni virtuali da sintetizzare, la loro direttività, il loro azimut e la loro elevazione. Da qui è possibile invertire la matrice di risposte all'impulso e ottenere la matrice di filtri da applicare ai segnali dalle capsule. In Figura 6.5 e Figura 6.6 sono mostrati alcuni degli ordini sintetizzabili.

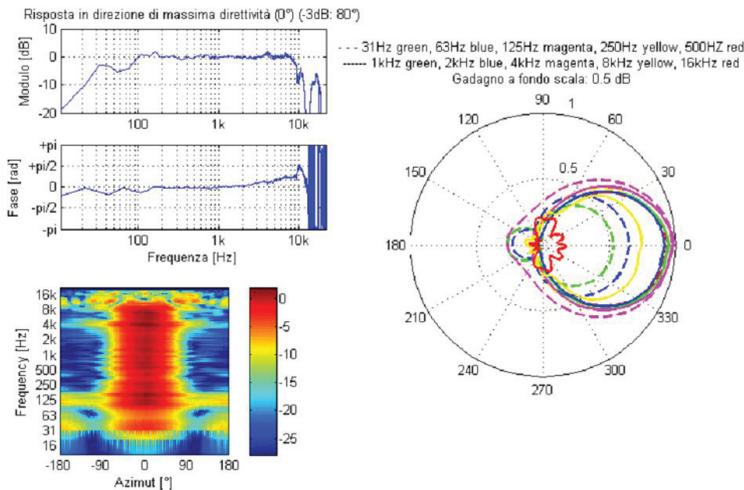


Figura 6.5: Cardioide di ordine 3.

Confrontando un microfono virtuale di ordine 6 con un reale microfono a fucile, il Sennheiser MKH 70-1 (diagramma polare in Figura 6.7), si può notare che la direttività del secondo dipende fortemente dalla frequenza, per cui un suono non in asse viene colorato molto, come se passasse attraverso

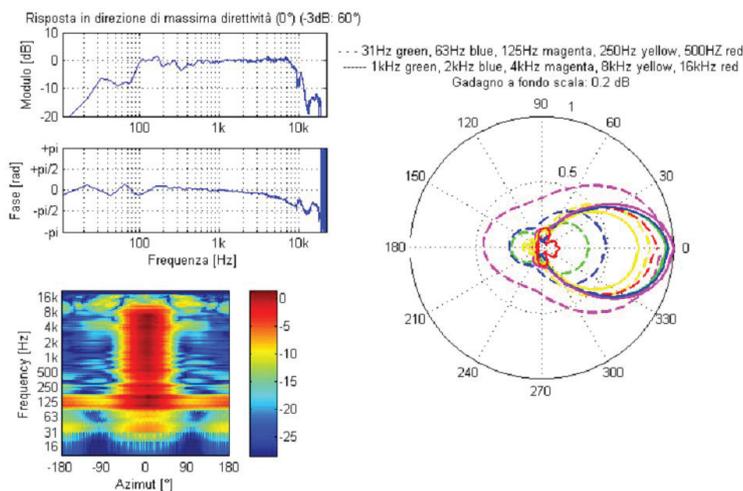


Figura 6.6: Ipercardioide di ordine 6.

un filtro passa-basso. Nel caso del microfono virtuale, invece, la direttività rimane pressoché costante da i 250 Hz in su; il risultato è un microfono più direttivo e più trasparente.

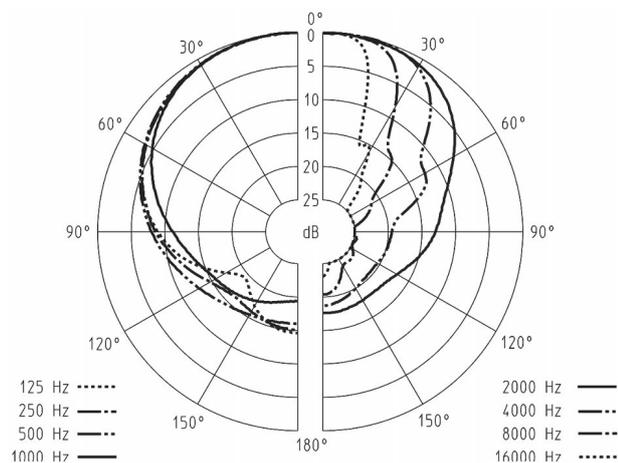


Figura 6.7: Diagramma polare del Sennheiser MKH 70-1

## 6.2 L'interfaccia grafica per l'utente

La schermata iniziale del software 3D-VMS richiede di selezionare la configurazione adottata in fase di registrazione, tra cui il modello del microfono utilizzato (quindi la matrice corrispondente), l'angolo su cui si vuole concentrare la ripresa e l'orientamento del microfono (dritto o a "testa in giù"). Sono inoltre selezionabili le interfacce da usare per la riproduzione e per la registrazione dell'audio. Il software può essere utilizzato in due modalità:

- **Playback:** permette di caricare un file audio a 32 canali, registrato in precedenza con un Eigenmike (o altro array microfonico) e un contenuto visivo opzionale. In questa modalità è possibile effettuare una nuova mappatura dei microfoni virtuali e migliorare eventuali configurazioni di ripresa non ottimali. Questa modalità consente di sintetizzare fino a  $n \times 7$  microfoni virtuali.
- **Realtime:** questa modalità è utilizzata per modificare in tempo reale il posizionamento di 7 microfoni virtuali, sintetizzati a partire da uno stream di 32 segnali in ingresso. Può essere utilizzata ad esempio, per il broadcasting o l'amplificazione<sup>2</sup> di un evento live.

L'opzione di utilizzo del timecode infine consente di sincronizzare un eventuale contenuto video ripreso contemporaneamente all'audio. Nella schermata principale di 3D-VMS è infatti possibile caricare un'immagine o video esteso fino a 360°, possibilmente ripresi dall'esatto "punto di vista" del microfono (Figura 6.9), in modo da rendere più intuitivo e rapido il puntamento dei microfoni virtuali. In questa vista ognuno dei microfoni virtuali viene rappresentato con un cerchio colorato, il cui diametro varia in base alla direttività selezionata (più grande il diametro, minore l'ordine del microfono) e la cui posizione, in coordinate  $x$  e  $y$ , rappresenta rispettivamente l'angolo di azimuth e l'angolo di elevazione. Il pannello sottostante comprende tutti i parametri di controllo dei 7 microfoni virtuali accessibili individualmente. Un'apposita sezione (Figura 6.10) permette inoltre di variare il livello di gain per ciascun microfono virtuale, sebbene si tenda a mantenere un valore uguale per tutti.

---

<sup>2</sup>Non tutte le condizioni sono ideali per l'amplificazione tramite 3D-VMS, dal momento che il sistema presenta un ritardo di processamento non sempre trascurabile.

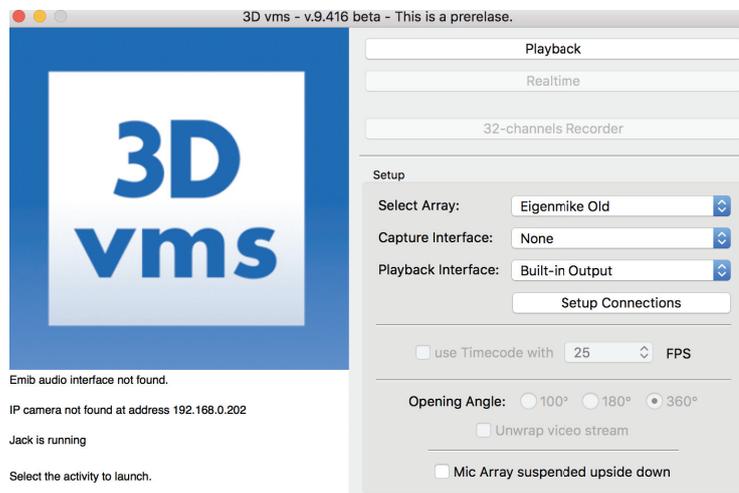


Figura 6.8: Schermata iniziale di 3D-VMS



Figura 6.9: Schermata principale di 3D-VMS

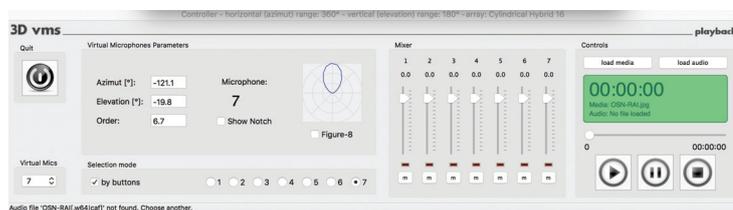


Figura 6.10: Parametri di controllo dei microfoni virtuali

## 6.3 Performance del sistema

### 6.3.1 Latenza

La latenza totale del processamento dipende da due fattori:

- Filtri FIR.
- Operazioni di convoluzione e buffering di sistema.

Il sistema è stabile quando usa filtri FIR della lunghezza di 1024 o 2048, partizionati in blocchi da 256 campioni (512 campioni in caso in cui si effettui una convoluzione veloce).

Lunghezza FIR	Campioni	Latenza (ms)
1024	256	23
2048	256	34
1024	512	34
2048	512	46

Tabella 6.1: Latenze di sistema

La latenza massima del sistema (46 ms) è sufficientemente contenuta da permettere l'uso del 3D-VMS in applicazioni di broadcasting. Nel caso di eventi live il 3D-VMS può essere utilizzato per amplificare le sorgenti sonore; se affiancato da tradizionali microfoni di supporto sarà necessario ritardare i segnali di questi in modo da compensare la latenza di processamento del sistema. Naturalmente l'output complessivo non sarà in perfetto sincronismo con la sorgente ripresa, ma tale sfasamento può considerarsi trascurabile per il pubblico. Tuttavia se tra gli scopi dell'amplificazione c'è anche il monitoraggio di una performance (ad esempio da parte di musicisti sul palcoscenico) questo ritardo non è accettabile.

### 6.3.2 Risposta in frequenza

Con l'utilizzo di un parametro  $\beta$  basso tra i 100 Hz e 10 kHz, si è misurata una risposta in frequenza praticamente piatta in questo range, anche con un microfono virtuale di ordine 6. Sotto i 100 Hz tale microfono fornisce ancora un segnale con un output significativo, ma la sua direttività risulta molto più

compromessa rispetto alle alte frequenze. Per estendere la risposta anche su altre frequenze sarebbe necessario aumentare la lunghezza dei filtri FIR, ma con filtri oltre i 2048 campioni il sistema sarebbe meno stabile. Inoltre il tentativo di aumentare il range di frequenze ha introdotto un aumento del rumore alle basse frequenze per i microfoni virtuali di ordine superiore [7].

## Capitolo 7

### Casi applicativi con 3D-VMS

Durante il mio tirocinio svolto presso Rai ho avuto l'occasione di assistere ad alcuni casi pratici di utilizzo del sistema 3D-VMS, frequentemente adottato nelle riprese di eventi live di varia natura e prevalentemente accoppiato ad una configurazione con 1 o 2 Eigenmike, più eventuali microfoni di supporto. È necessario infatti sottolineare che nonostante la sua grande innovatività e versatilità il 3D-VMS è una tecnologia ottima nella ripresa panoramica di un evento sonoro, ma non consente di catturare la stessa cura del dettaglio che si può ottenere con una tradizionale tecnica multi-microfonica. Il 3D-VMS permette di ricreare il realismo di una scena, nello stesso modo in cui questa è percepita, ad esempio, da uno spettatore: durante un evento dal vivo gli stimoli uditivi che giungono allo spettatore sono molteplici e frutto non soltanto delle principali sorgenti, ma anche della riverberazione della sala. Ciò che invece è proposto come standard "estetico" nell'industria musicale è la creazione di un'immagine sonora in cui la componente riverberante è meno presente e le principali sorgenti sonore sono riprese individualmente, così da consentire una riproduzione più intellegibile e "a fuoco" di ciascun suono. Una condizione lontana dal realismo, insomma.

Ad ogni modo è importante sottolineare che la microfonazione delle varie sezioni dello spettacolo è avvenuta esclusivamente a scopo di registrazione e successiva trasmissione su radio e altri supporti e non a scopo di amplificazione. Per amplificare un evento in realtime in queste condizioni bisogna considerare la latenza introdotta dall'elaborazione digitale del segnale da parte del 3D-VMS, che oltre un certo valore non può essere ridotta, nonché il problema più comune di effetto Larsen.

## 7.1 Utilizzo del 3D-VMS in realtime: ripresa di un'opera lirica al Teatro Regio di Torino

### 7.1.1 Panoramica

La ripresa audio della prova generale di un'opera lirica da parte di Rai Radio è avvenuta combinando due mix separati: uno dedicato all'orchestra sinfonica, disposta nella buca del teatro sottostante al palco, e uno dedicato agli attori sul palcoscenico e alle reazioni del pubblico in sala. I due mix sono stati gestiti tramite due mixer indipendenti, controllati da un tecnico Rai ciascuno, quindi combinati e registrati tramite un registratore digitale Tascam DA-98 HR.

### 7.1.2 Microfonazione dell'orchestra

L'orchestra sinfonica, come da convenzione, è composta dalle seguenti tipologie di strumenti:

- archi
- arpa
- legni
- ottoni
- percussioni

Per la ripresa della sezione orchestrale è stata adottata una tecnica multi-microfonica, con microfoni disposti in modo da suddividere i diversi strumenti per zone. In particolare, sono stati utilizzati degli Schoeps della serie modulare Colette (Figura 7.1), microfoni a condensatore composti da una capsula MK5, dedita alla conversione delle onde acustiche in tensione elettrica e alla direttività del diagramma polare, e da un amplificatore CMC5, necessario per l'alimentazione e la polarizzazione della capsula capacitiva.



Figura 7.1: Schoeps serie Colette. Sistema modulare

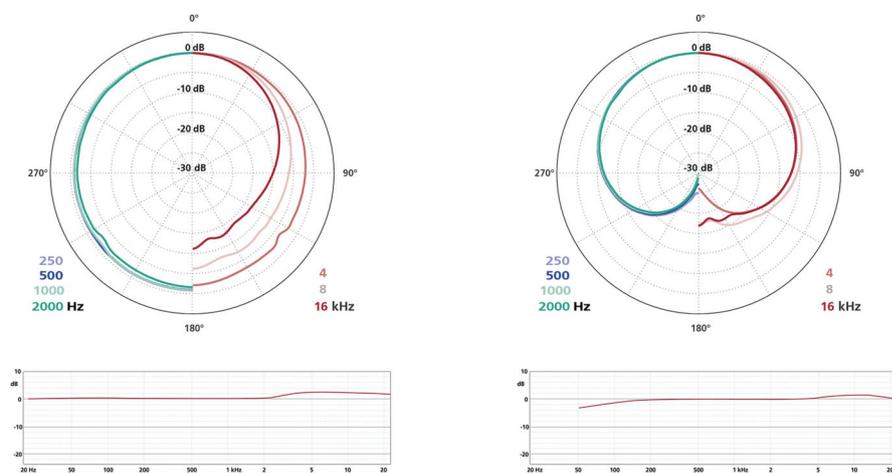
Questi microfoni sono stati posizionati a circa 2 metri di altezza e puntati in basso verso i musicisti, per mezzo di speciali aste attive Schoeps che permettono di distanziare ciascuna capsula dal suo amplificatore, quindi di gestire le capsule come microfoni in miniatura. Ciò consente un posizionamento dei microfoni poco intrusivo tra i vari orchestrali.

La capsula MK5 è composta da un singolo diaframma e presenta la possibilità di variare la figura polare tra cardioide e omnidirezionale. Osservando la Figura 7.2a e la Figura 7.2b<sup>1</sup>, si può notare la grande linearità della risposta della capsula in entrambe le configurazioni polari, fatta esclusione per una leggera enfasi sulle medio-alte e le alte frequenze che donano un suono brillante alla ripresa. Il diaframma piccolo invece consente una ripresa più fedele dei transienti, donando quindi un'immagine sonora più dettagliata, e presenta inoltre un diagramma molto regolare all'aumentare della frequenza, il che consente di riprendere definito non solo la sorgente sonora in asse con la capsula, ma anche le sorgenti adiacenti.

Sulle prime parti, occupate dagli archi, i microfoni sono stati configurati con caratteristica polare cardioide, per ottenere una maggiore direttività sugli strumenti solisti, di cui spesso si vogliono accentuare i passaggi rispetto al resto dell'orchestra. Inoltre, la presenza del lobo di attenuazione aiuta a ridurre eventuali rumori indesiderati provenienti dall'esterno dell'orchestra. Per gli strumenti nelle file, quindi senza parti solistiche, i microfoni sono stati configurati con caratteristica omnidirezionale. Il diagramma omnidirezionale permette di riprendere un maggiore numero di strumenti rispetto al cardioide, mantenendo una migliore risposta alle basse frequenze in tutte le direzioni, ma minori dettagli nel suono. Per i contrabbassi, gli strumenti più gravi dell'orchestra, è stato usato un diverso tipo di microfono a condensatore a diaframma largo: il Neumann TLM 170, con caratteristica polare cardioide

---

<sup>1</sup>I grafici fanno riferimento alla configurazione MK5 + CMC6, quest'ultimo successore del CMC5, ma con comportamento similare.



(a) Caratteristica omnidirezionale.

(b) Caratteristica cardioide.

Figura 7.2: Diagrammi polari e risposte in frequenza dello Schoeps MK5.

allargato (per consentire una ripresa più definita anche fuori asse), piazzato a circa 1 metro di altezza da terra.

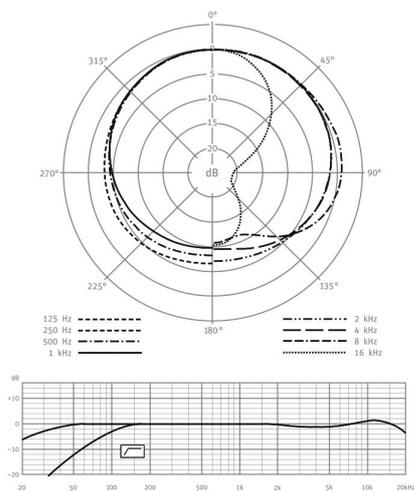


Figura 7.3: Diagramma polare e risposta in frequenza del Neumann TLM-170 con caratteristica wide-cardioid

Come si può osservare dai grafici in Figura 7.3, la risposta del Neumann è tanto lineare quanto quella dello Schoeps in configurazione cardioide. La differenza più sostanziale sta nella variazione del diagramma alle alte frequenze, dove il Neumann diventa molto più direttivo rispetto allo Schoeps. Tuttavia questo effetto risulta trascurabile se si considerano le frequenze emesse dai contrabbassi. La scelta di un microfono a diaframma largo in questo caso è dunque dovuta esclusivamente alla sua migliore capacità di ripresa di suoni fuori asse. Ciò permette un angolo di ripresa maggiore, anche se la ricchezza dei dettagli degli elementi registrati non sarà la stessa di quelli registrati dagli Schoeps. Ma trattandosi di strumenti delle seconde file questo compromesso può considerarsi accettabile.

### 7.1.3 Microfonazione del palcoscenico

Per la ripresa del palco è stata utilizzata una combinazione dei seguenti sistemi:

- 9 radiomicrofoni lavalier con capsule DPA 4060, piazzati individualmente sui 9 personaggi principali
- 1 mh Acoustics Eigenmike + 3D-VMS a bordo palco
- 2 Neumann KMR82i (boccascena)
- 4 Neumann KMR81i (boccascena)

I DPA 4060<sup>2</sup> sono microfoni a condensatore in miniatura (la capsula ha un diametro di soli 3 mm) che è possibile installare direttamente sugli attori in scena. Si tratta di microfoni omnidirezionali con una risposta in frequenza molto lineare (Figura 7.4), salvo per un leggero boost intorno ai 10 kHz che però può essere ridotto applicando una diversa griglia al microfono (detta appunto soft-boost grid). Il loro diretto piazzamento sulla sorgente permette di ottenere una ripresa molto ricca di dettagli; tale sistema infatti ricopre più del 50% del mix di questa sezione. Di contro, un posizionamento non ottimale può comportare una ripresa “ovattata”, in cui le frequenze alte e medio-alte vengono eccessivamente assorbite dal costume dell’attore su cui è montato. Inoltre, i radiomicrofoni per loro natura possono presentare problemi nella ricetrasmisione del segnale, il che li rende più imprevedibili rispetto a sistemi microfonici via cavo.

---

<sup>2</sup>[www.dpamicrophones.com/dscreet/4060-series-miniature-omnidirectional-microphone](http://www.dpamicrophones.com/dscreet/4060-series-miniature-omnidirectional-microphone)

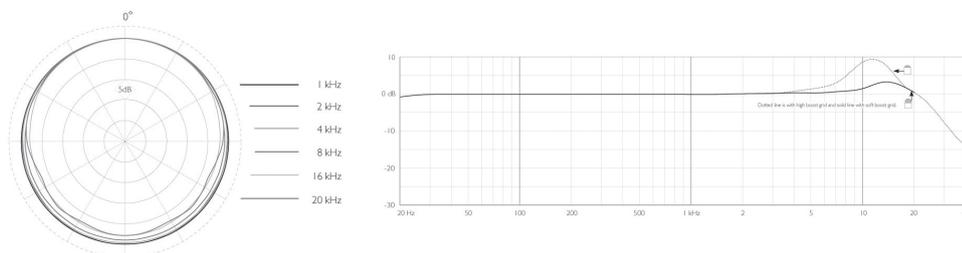


Figura 7.4: Diagramma polare e risposta in frequenza del DPA 4060

Una seconda criticità, non propriamente tecnica, di questi radiomicrofoni in miniatura è dovuta al fatto che gli attori che li indossano possano distrattamente dimenticarsi della loro presenza quando escono di scena. Sta al fonico addetto al mix di palcoscenico spegnere e accendere prontamente questi segnali quando l'attore esce o entra in scena.

L'Eigenmike, situato nella parte centrale del bordo palco, utilizza tutte le 32 capsule per coprire l'intera scena frontale. La sua risposta in frequenza è particolarmente efficace tra i 300 Hz e i 12 kHz e ciò lo rende uno strumento più adatto alle voci degli attori, piuttosto che alla sezione strumentale. Tramite il sistema 3D-VMS sono stati utilizzati tutti e 7 i microfoni virtuali disponibili nella modalità realtime, con la configurazione indicata in Tabella 7.1, mantenuta invariata per tutto lo spettacolo. I segnali sono stati memorizzati in formato W64: questo formato a 64-bit, da Sony, è stato progettato per superare il limite dei 4 GB posto dal formato Microsoft RIFF/WAV a 32-bit, permettendo di ottenere file audio di dimensione molto maggiore.

Mic	Azimuth (°)	Elevazione (°)	Ordine	Gain (dB)
1	80.0	10.0	5.0	0.0
2	40.0	10.0	4.5	0.0
3	0.0	10.0	4.0	0.0
4	-40.0	10.0	4.5	0.0
5	-80.0	10.0	5.0	0.0
6	45.0	5.0	5.5	0.0
7	-45.0	5.0	5.5	0.0

Tabella 7.1: Parametri di configurazione del 3D-VMS



Figura 7.5: Schermata 3D-VMS.

Dalla Tabella 7.1 e dalla Figura 7.5 si osserva che il posizionamento dei microfoni è simmetrico rispetto al centro del palcoscenico. Dai valori di azimut si nota che i microfoni virtuali sono stati impostati in modo da riprendere la scena sul palco con un angolo di circa  $160^\circ$  (da  $80^\circ$  a  $-80^\circ$ )<sup>3</sup>, riuscendo a coprire gran parte del palcoscenico, pur mantenendo l'attenzione sul centro, dov'è concentrato un maggior numero di microfoni.

Tenendo conto dell'altezza media degli attori e della loro posizione sul palco durante le scene, sono stati impostati diversi valori di elevazione a seconda che i microfoni fossero destinati a riprendere una scena a bordo palco (maggiore elevazione) o più indietro (minore elevazione). L'elevazione permette quindi

<sup>3</sup>Bisogna considerare che la direttività di microfono si riduce con la distanza da questo, per cui l'area reale ripresa con tale configurazione è presumibilmente maggiore, sebbene i suoni agli estremi saranno meno a fuoco.

di compensare parzialmente la distanza della sorgente dal microfono. Nel caso specifico i primi 5 microfoni (da 1 al 5) sono stati configurati in modo da riprendere gran parte del palcoscenico, a circa 2 m dal bordo palco. I microfoni 6 e 7 invece sono stati configurati per catturare azioni più lontane, a circa 4-5 m dal bordo palco e prevalentemente centrali nel palcoscenico.

Per quanto riguarda l'ordine di direttività, si osserva che i microfoni destinati all'area centrale del palco sono leggermente meno direttivi, pensati con una caratteristica supercardioide in modo da mantenere più ampio l'angolo di ripresa nelle zone in cui vi è una maggior concentrazione di attori o in cui questi si muovono di più. Per lo stesso ragionamento, i valori di direttività dei microfoni 6 e 7 sono stati impostati più alti<sup>4</sup> per compensare la distanza: il cono di ripresa si allarga fino a renderli dei cardioidi nella zona desiderata.

Il sistema 3D-VMS fornisce una panoramica sonora dell'intera scena, ma nel caso di scene a livelli d'intensità molto variabili (come lo sono le opere liriche, in genere) sarà difficile riprodurre correttamente a fuoco le diverse sorgenti. Per questo motivo tale sistema ricopre il 30%-40% del mix di questa sezione, lasciando ai radiomicrofoni e ai fucili nel boccascena il compito di catturare più dettagli. Inoltre, per compensare il ritardo dei segnali elaborati dal sistema 3D-VMS rispetto a quelli reali (circa 40 ms) è stato introdotto un ritardo analogo sui segnali provenienti dai microfoni convenzionali, tramite i mixer della regia audio.

Un totale di 6 microfoni a fucile sono stati posti lungo tutto il boccascena, puntando verso le parti principali del palco. Questi permettono di riprendere più nel dettaglio cori o insiemi di attori non microfonati individualmente. I due tipi di microfoni utilizzati, Neumann KMR-81i e Neumann KMR-82i, sono entrambi dei fucili ad interferenza (di gradiente di pressione) con caratteristica polare ipercardioide. In Figura 7.6 sono mostrati i grafici di ciascun modello. Le risposte in frequenza dei due microfoni sono praticamente identiche (se non si considera l'opzione di filtro passa-basso presente sul KMR-82i); dai diagrammi polari si osserva una maggiore direttività nel KMR-82i, la cui figura rimane più consistente anche dopo i 2 kHz. Il KMR-82i risulta quindi più indicato per distanze maggiori rispetto al KMR-81i. Considerando che la direttività di un microfono diminuisce con la distanza dalla sorgente, per

---

<sup>4</sup>Un ordine 5 o 6 corrisponde ad caratteristica polare ipercardioide.

dei fucili posti a 7-8 metri di altezza, il cono di ripresa tenderà ad allargarsi fino a renderli dei cardioidi in grado di riprendere un gruppo di attori.

Infine, due microfoni Schoeps MK5+CMC4 a diaframma piccolo sono stati posti in punti estremi e opposti del bordo palco e diretti verso la sala, con lo scopo di riprendere anche le reazioni del pubblico, che in fase di registrazione sono state opportunamente enfatizzate o silenziate dai tecnici addetti ai mixer.

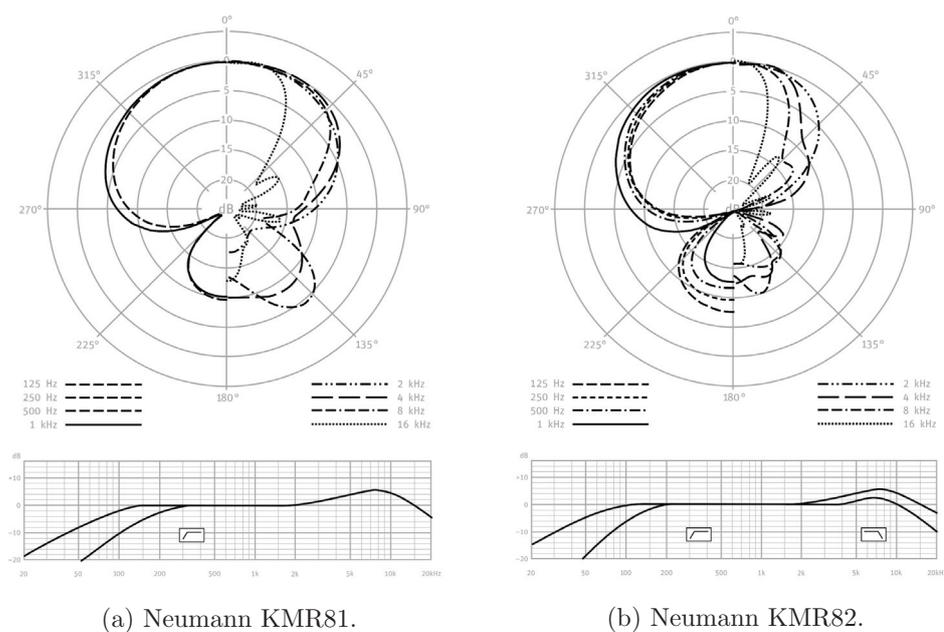


Figura 7.6: Diagrammi polari e risposte in frequenza di Neumann KMR81 e KMR82.

## 7.2 Utilizzo del 3D-VMS in playback: post-produzione di un concerto di banda musicale

### 7.2.1 Panoramica

Questo secondo evento ha rappresentato un caso critico dal momento che le riprese del concerto sono avvenute in condizioni non ottimali: la registrazione della banda musicale in questione, composta da 50 strumentisti, è avvenuta senza microfoni di supporto di alcun tipo ed esclusivamente per mezzo di una coppia di Eigenmike, appesi al soffitto della sala. Lo scopo di tale registrazione è stato la produzione di un mix in formato 5.1 surround e di un mix stereo dell'intero concerto.

### 7.2.2 Strumentazione di ripresa

Come anticipato, il setup di registrazione del concerto è stato minimale:

- 2 mh Eigenmike, posti a circa 4 metri sopra il direttore della banda; l'altezza ideale, data la disposizione dei musicisti su poche file, sarebbe stata a circa due metri dal direttore. Come mostrato in Figura 7.7, entrambi i microfoni sono stati appesi al soffitto della sala, a circa 5 metri l'uno dall'altro e a testa in giù; questo è stato fatto per evitare ostacoli frontali durante la ripresa e per raggiungere anche le file posteriori della banda. Il microfono di sinistra è stato correttamente puntato verso la banda, mentre il microfono di destra è stato erroneamente<sup>5</sup> puntato verso il pubblico.
- Una videocamera è stata utilizzata per una ripresa audio e video di riferimento.

---

<sup>5</sup>Trattandosi di un microfono le cui capsule sono uniformemente distribuite su una superficie sferica, l'orientamento del microfono non incide direttamente la qualità della ripresa. Tuttavia seguire la convenzione di orientamento agevola una corretta configurazione del software 3D-VMS.

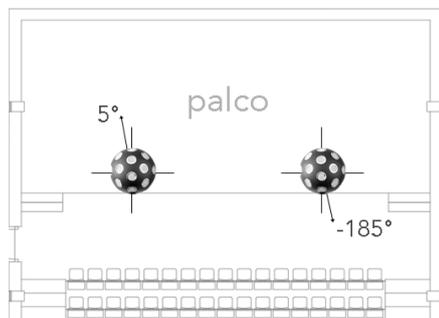


Figura 7.7: Posizione e orientamento degli Eigenmike nella sala.

### 7.2.3 Post-produzione

La post-produzione delle riprese è avvenuta nella regia audio del Centro Produzioni Rai di Torino, già predisposta per l'ascolto multicanale fino a 7+4.1<sup>6</sup>, con monitor Genelec. I software utilizzati in questa fase sono stati 3D-VMS, per la microfona virtuale della banda, e Pro Tools, su cui sono state in seguito riversate le tracce audio ottenute dal 3D-VMS. Nel software 3D-VMS sono stati utilizzati 7 microfoni virtuali a partire da ciascun Eigenmike, per un totale di 14 microfoni virtuali. Dopo aver selezionato il corretto modello di microfono e l'opzione "Mic array suspended upside down", dal pannello iniziale del software, si è effettuato il posizionamento dei microfoni virtuali sulla banda. I segnali derivanti dal primo Eigenmike sono stati elaborati in modo da coprire la parte sinistra della banda, con la seguente configurazione:

Mic	Azimuth (°)	Elevazione (°)	Ordine	Gain (dB)
1	135.0	5.0	4.0	0.0
2	45.0	-90.0	6.0	0.0
3	30.0	-90.0	6.0	0.0
4	10.0	-90.0	6.0	0.0
5	-10.0	-90.0	6.0	0.0
6	-30.0	-90.0	6.0	0.0
7	-45.0	-90.0	6.0	0.0

Tabella 7.2: Parametri del 3D-VMS per l'Eigenmike di sinistra.

<sup>6</sup>7 altoparlanti in piano, 4 altoparlanti in aria e 1 subwoofer.

I microfoni da 2 a 7 sono stati configurati per riprendere un totale di  $90^\circ$ , dal centro della banda al suo estremo sinistro, e disposti simmetricamente su questa sezione. L'angolo di elevazione così basso è dovuto al fatto che i due array sferici si trovavano molto più in alto e distanti rispetto ai musicisti. Per lo stesso motivo è stato impostato l'ordine di direttività massimo (caratteristica ipercardioidale), che ha fatto sì che questi microfoni agissero in maniera equivalente a dei cardioidi posti in prossimità della banda. Il microfono 1 è stato invece posizionato in modo riprendere le reazioni del pubblico, ma la sua elevazione è stata mantenuta alta per non puntare direttamente sulle persone e catturare l'ambiente della sala. Il gain di tutti i microfoni è rimasto invariato per evitare di enfatizzare certe sezioni della banda a discapito di altre e mantenere una ripresa più fedele possibile alla situazione reale. In modo speculare, i segnali derivanti dal secondo Eigenmike sono stati configurati per microfonare omogeneamente la parte destra della banda:

Mic	Azimuth ( $^\circ$ )	Elevazione ( $^\circ$ )	Ordine	Gain (dB)
1	135.0	-90.0	6.0	0.0
2	155.0	-90.0	6.0	0.0
3	170.0	-90.0	6.0	0.0
4	-170.0	-90.0	6.0	0.0
5	-150.0	-90.0	6.0	0.0
6	-135.0	-90.0	6.0	0.0
7	-45.0	5.0	4.0	0.0

Tabella 7.3: Parametri del 3D-VMS per l'Eigenmike di destra.

In questo caso si è dovuto tenere conto dell'errata orientazione del microfono, puntato accidentalmente con l'asse  $X$  verso il pubblico, sommando  $180^\circ$  all'angolo di azimut di ciascun microfono.

Una volta completati i settaggi del 3D-VMS si è potuto procedere col riversamento delle tracce dei 14 microfoni in Pro Tools, per effettuare un mix in surround 5.1. Nel caso di ascolto musicale, il setup per un playback in 5.1 è quella consigliato dallo standard ITU-R BS 775 [17], mostrata in Figura 7.8. Tale configurazione è la più utilizzata nel caso di mix multicanale e di ascolti critici a distanza ravvicinata (nearfield).

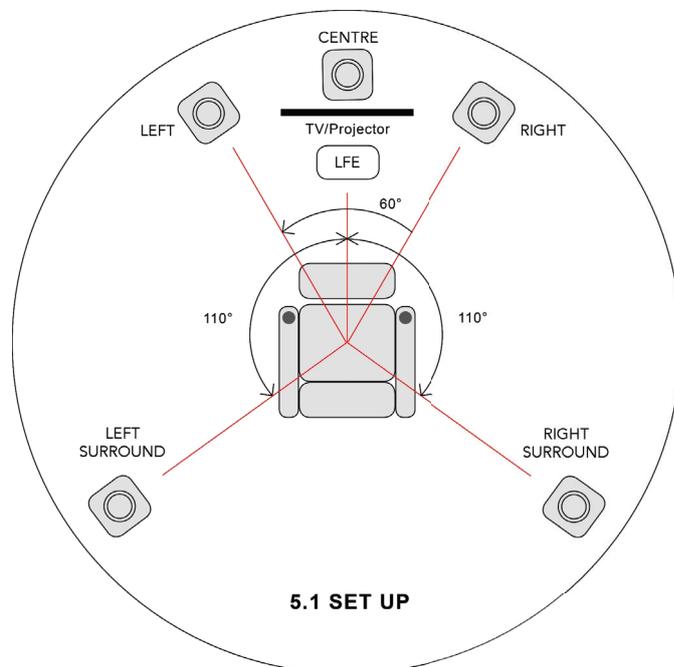


Figura 7.8: Setup standard per 5.1 surround.

Il 5.1 surround presenta 5 canali a larghezza di banda completa (generalmente tra 20 Hz e 20 kHz), posti idealmente alla stessa distanza dalla postazione di mixaggio, denominati (Front) Right, (Front) Left, Centre, Surround Right, Surround Left, e 1 canale LFE (low frequency effects), destinato alla riproduzione di basse frequenze tramite subwoofer<sup>7</sup>.

In aggiunta ai canali front destro e sinistri, tipici delle configurazioni stereo, il surround 5.1 comprende quindi

- un canale centrale: in genere destinato alla riproduzione di elementi principali di un mix, ad esempio i dialoghi in un film o la voce principale in una composizione musicale

<sup>7</sup>Se si considera un sample rate di 48 kHz la massima frequenza teoricamente ottenibile dai canali non LFE è 24 kHz. Il subwoofer in genere emette frequenze fino ad un massimo di 120 Hz, cioè 1/200 o 0.005 della larghezza di banda degli altri canali. Per questo motivo la configurazione 5.1 dovrebbe più correttamente chiamarsi 5.005, per indicare la banda di frequenze che il LFE emette in rapporto con gli altri canali.

- due canali surround: disposti alle spalle dell'ascoltatore, sono comunemente destinati alla diffusione dell'ambianza e a suoni non primari
- un canale LFE: destinato alla riproduzione di suoni a bassa frequenza tramite un subwoofer. Questo può essere pensato come un canale supplementare a cui inviare un segnale che benefici di un'enfatizzazione delle basse frequenze

Dopo aver opportunamente impostato gli I/O del progetto di Pro Tools, sono state create 14 tracce audio 5.1 e a ciascuna è stato assegnato uno dei microfoni virtuali configurati in 3D-VMS. Il panner surround di traccia 5.1, in Figura 7.9, consente di posizionare con semplicità il rispettivo segnale nello spazio.

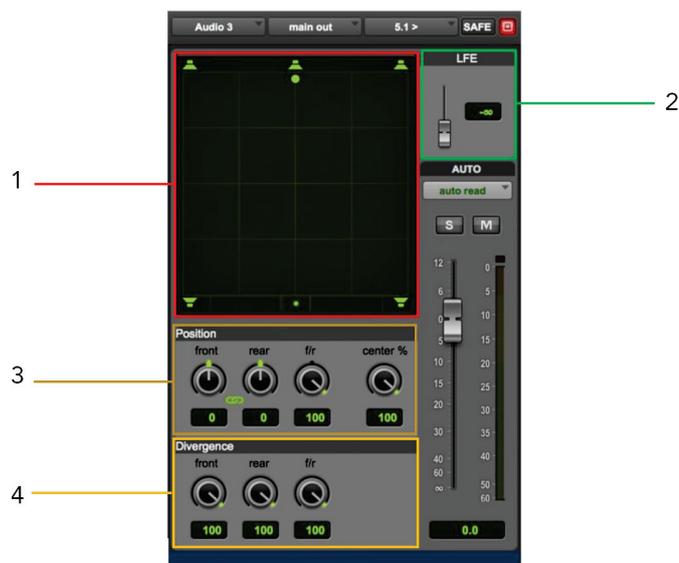


Figura 7.9: Panner surround di Pro Tools.

Il panner è composto dalle seguenti sezioni:

1. Griglia X/Y: la rappresentazione grafica del segnale nello spazio ricreato dai 5 altoparlanti a larghezza di banda completa. Il cursore verde indica la posizione corrente del segnale e può essere spostato arbitrariamente all'interno della griglia.

2. Fader LFE: permette di definire se e in che quantità il segnale contribuisce al canale LFE.
3. Controlli di posizione: permettono di impostare la posizione del panner (in alternativa all'uso della griglia X/Y). Front e Rear controllano la posizione corrente sull'asse X (destra/sinistra); i due controlli sono dipendenti tra loro. F/R (Front/Rear) controlla la posizione corrente sull'asse Y. Center % definisce quanto segnale viene diretto al canale centrale (con lo 0% si ottiene un mix con cosiddetto "phantom centre", ovvero "centro fantasma", perché il canale centrale viene omesso nella distribuzione dei segnali). Nel nostro caso, il Center % è stato mantenuto a 100% per tutti i microfoni virtuali.
4. Divergenza: determina la larghezza del segnale pan-pottato rispetto agli speaker adiacenti: un segnale divergente al 100% sarà pan-pottato completamente su un solo altoparlante e sarà udibile soltanto su quello. Al diminuire della divergenza le sorgenti sonore tendono progressivamente ad allargarsi, perché il segnale viene parzialmente diretto anche negli altoparlanti adiacenti. È uno strumento particolarmente utile per assicurare l'udibilità e la copertura del suono in grandi sale e per migliorare la chiarezza delle voci. Nel nostro caso, tutti i valori di divergenza sono stati mantenuti a 100, come di default.

Una volta riversati i 14 microfoni virtuali nelle rispettive tracce 5.1, i segnali sono stati pan-pottati nello spazio riferendosi alle posizioni reali della banda, come mostrato in Figura 7.10. Le tracce contenenti i segnali dei microfoni virtuali elaborati a partire dall'Eigenmike di sinistra sono state disposte in modo omogeneo a partire dal canale C (audio 7, corrispondente al microfono virtuale 7) fino al canale L (audio 2, corrispondente al microfono virtuale 2). La traccia corrispondente al microfono virtuale 1 è stata invece pan-pottata nel canale LS e le è stato applicato in insert un equalizzatore che agisce come filtro passa-alto, per eliminare brusio e rumori indesiderati provenienti dal pubblico e enfatizzare leggermente le frequenze sopra i 10 kHz, aggiungendo più aria al suono. Specularmente, le tracce corrispondenti ai microfoni virtuali elaborati a partire dall'Eigenmike di destra sono state disposte dal canale C fino al R, mentre quella corrispondente al microfono virtuale 7 è stata pan-pottata sul canale RS e le è stato applicato un filtro passa-alto analogo a quello usato per il canale di sinistra. Sono quindi stati creati due



o più in generale un bus risulta pratico quando c'è bisogno di applicare un plug-in su tutte le tracce presenti nel bus senza inserirlo individualmente in ogni traccia: in questo caso infatti sono stati utilizzati i seguenti plug-in sul master, in modo da agire su tutte le tracce contemporaneamente:

1. Avid Channel strip (Figura 7.11): è stato effettuato un taglio leggero e molto largo intorno ai 240 Hz per togliere un po' di "gonfiore" dal mix, quindi un boost leggero e largo intorno ai 2 kHz, per accentuare i dettagli.
2. Waves L360° Surround Limiter (Figura 7.12), usato come maximizer multibanda, ossia uno strumento in grado di prevedere i picchi di dinamica e di pre-attenuarli mantenendo la massima trasparenza nel suono.



Figura 7.11: Interfaccia di Avid Channel Strip.

Per ottenere un mix stereo a partire dal surround 5.1 è stato creato un canale stereo a cui è stato assegnato il canale Master 5.1 come input e il plug-in Avid Down Mixer (Figura 7.13) in insert, che permette una rapida conversione da 5.1 a 2.0.

Il downmix ha richiesto un'ulteriore modifica ai livelli dei 6 canali del surround dal momento che nel passaggio ad una configurazione stereo lo spazio



Figura 7.12: Interfaccia di Waves L360 Surround Limiter.

per il posizionamento dei segnali si riduce notevolmente e avvengono molte sovrapposizioni di suoni, quindi un diverso equilibrio tra le tracce. Infine, anche sul canale Master stereo è stato inserito un maximizer multibanda per agire sulle dinamiche del mix totale, il Waves L3 Multimaximizer (praticamente il corrispettivo stereo di L360). Anche in questo caso, un taglio intorno ai 240 Hz ha ripulito il mix da suoni troppo gonfi e un leggero incremento intorno ai 2 kHz ha aumentato i dettagli udibili.



Figura 7.13: Interfaccia di Avid Down Mixer.

#### 7.2.4 Risultati

Nonostante le condizioni di ripresa non fossero ottimali, il risultato finale è stato complessivamente soddisfacente: le caratteristiche timbriche degli strumenti della banda sono state rispettate e il prodotto 5.1 surround risulta molto avvolgente. Un eventuale impiego di microfoni di supporto disposti in mezzo agli strumentisti avrebbe sicuramente migliorato a percepire più dettagli, soprattutto per gli strumenti più gravi che sono meno in grado di bucare il mix e con la configurazione usata tendono a suonare leggermente fuori fuoco. Bisogna considerare però che, vista la natura dell'evento, la presenza di microfoni di supporto può ritenersi meno necessaria rispetto al caso dell'opera lirica. Infatti, le composizioni per banda sono generalmente brani con intensità elevate e relativamente statiche (si pensi a una fanfara), se confrontate con quelle di un'opera lirica, dove la grande varietà di dinamiche dell'orchestra va di pari passo con la recitazione degli attori.

### 7.3 Vantaggi del 3D-VMS rispetto alle tradizionali tecniche multi-microfoniche

Nel secondo caso applicativo, la post-produzione è stata effettuata da un personale diverso da quello che ha effettuato le riprese dell'evento. Questo dimostra come il sistema 3D-VMS+Eigenmike sia uno strumento estremamente versatile che permette, previa corretta comunicazione del setup in fase di ripresa, la suddivisione dei compiti della catena di produzione in modo molto efficiente. Certamente anche utilizzando una tradizionale tecnica multi-microfonica è possibile che ad occuparsi di ripresa e di post-produzione siano persone diverse, almeno in teoria. Nella pratica la gestione di un grande numero di microfoni può risultare difficoltosa per un tecnico di post-produzione non presente in fase di registrazione, quindi non direttamente a conoscenza di tutte le eventuali criticità e strategie adottate in quella specifica situazione.

Per riprendere questo concerto con una tecnica multi-microfonica tradizionale sarebbero necessari almeno 20 microfoni, con altrettante aste, un'interfaccia audio con un sufficiente numero di canali d'input e un adeguato numero di cavi. Il posizionamento dei microfoni in mezzo ad un organico così numeroso richiederebbe un significativo compromesso tra qualità e praticità, dal momento che si cerca sempre di minimizzare l'ingombro dovuto dagli stru-

menti di registrazione e dai loro supporti. I cavi quindi, disposti in modo da non essere accidentalmente tirati e da non ostacolare musicisti e addetti ai lavori, verrebbero portati in una sala di regia lontana dal palcoscenico e collegati ad un mixer, ad una DAW o ad un altro registratore. Inoltre non è così raro che, a microfonazione avvenuta, i microfoni vengano spostati (involontariamente o no), rendendo la ripresa inutilizzabile. Una squadra di produzione che debba gestire una situazione di questo tipo impiegherebbe almeno un giorno per preparare e montare adeguatamente la strumentazione necessaria e un giorno per smontarla. In seguito alla microfonazione, i livelli di gain verrebbero impostati individualmente per ciascun microfono ed eventualmente aggiustati in corso d'opera nel caso in cui i microfoni o i musicisti si spostino rispetto alle loro posizioni iniziali. Sebbene si cerchi sempre di utilizzare posizionamenti tali da evitare interferenze di fase distruttive tra i segnali dei vari microfoni, non è da escludere che durante la post-produzione sia necessario un allineamento tra le fasi di alcuni segnali.

La ripresa per com'è avvenuta, con l'utilizzo di due Eigenmike e 3D-VMS, sebbene non sia stata effettuata in condizioni ottimali ha fornito un prodotto finale in linea con gli standard professionali, dimezzando i tempi di produzione (rispetto ad una ripresa analogica con tecnica multi-microfonica) e minimizzando il numero di strumentazione necessaria. Si tratta quindi di una tecnica estremamente pratica e veloce, che se usata in modo corretto e in condizioni opportune consente di ottenere un risultato più che soddisfacente anche senza l'affiancamento di microfoni di supporto. Senza dubbio i suoni ripresi soltanto da un microfono lontano, per quanto direttivo sia, non saranno mai definiti e puliti quanto quelli ripresi da un microfono a pochi centimetri dalla sorgente. Per questo motivo una ripresa ottimale richiede sempre l'utilizzo di microfoni di supporto. In post-produzione la versatilità del 3D-VMS è indubbiamente superiore: poichè il sistema cattura l'intero ambiente, al di là dell'orientamento delle capsule, è possibile ottenere una mappatura dei microfoni differente da quella utilizzata durante la ripresa dal vivo. Infatti grazie ai  $n \times 7$  microfoni virtuali utilizzabili in playback è possibile ricostruire immagini sonore molto più dettagliate e di seguire letteralmente le sorgenti anche nel caso di scene particolarmente movimentate.

## Capitolo 8

### Conclusione

Da quanto emerso in questa tesi, una delle principali questioni riguardo alla registrazione di campi sonori è come ottenere un microfono semplice e compatto con un'alta direzionalità. Tra i microfoni descritti, ognuno si distingue per qualità diverse: l'Holophone è senza dubbio un microfono molto intuitivo e pratico se si vuole ottenere in poco tempo un prodotto in 5.1 o 7.1 surround, senza particolari manipolazioni del campo registrato. Di contro, i risultati delle sperimentazioni effettuate con questo microfono [23] dimostrano che le riprese non sono mai immersive quanto quelle ottenibili con altre tecnologie citate. Questo è dovuto principalmente alla scarsa direttività delle capsule utilizzate, che, nonostante la buona qualità costruttiva, non sono in grado di separare adeguatamente le sorgenti in fase di ripresa. Il risultato è dunque un'immagine sonora pulita e brillante, ma la cui spazializzazione non è sufficientemente realistica. Per questi motivi l'Holophone rimane più adatto alla realizzazione di formati audio standard, in cui sia sufficiente ricreare l'illusione di immersione uditiva, senza però una corretta (dal punto di vista del realismo) disposizione spaziale degli elementi sonori rappresentati. Per quanto riguarda la registrazione con tecnica Ambisonics, negli ultimi anni si è assistito ad una grande evoluzione nell'ambito dei microfoni a simmetria sferica, sia dal punto di vista della varietà di costruttori e modelli disponibili in commercio, che nello sviluppo di nuove tecniche per l'elaborazione dei segnali e l'ottenimento di immagini sonore desiderate [6]. I microfoni SoundField rappresentano storicamente un caposaldo nell'Ambisonics: nati verso la metà degli anni '70, sono tutt'ora largamente utilizzati. La tecnologia su cui si basano i SoundField (e analogamente il più economico Sennheiser Ambeo) permette una ripresa totale e realistica del campo sonoro, indipen-

dentemente dall'orientamento del microfono; la possibilità di ridirezionare i segnali anche in seguito alla registrazione dona un'ampia manovrabilità in post-produzione. I SoundField si sono affermati in particolar modo nelle riprese di eventi musicali tradizionali (es. musica da camera), grazie alla loro struttura a capsule coincidenti, che consente un'eccellente compatibilità nei downmix stereofonici e monofonici (i principali formati di ascolto musicale). Tuttavia anche per questi microfoni limiti fisici comportano alcuni difetti nella direttività delle capsule, soprattutto alle alte frequenze, compromettendo la corretta spazializzazione delle sorgenti del campo. Tra tutti gli strumenti descritti, l'Eigenmike sembra essere la soluzione più completa, in termini di spazializzazione e di versatilità. Le 32 capsule ad alta qualità garantiscono una ripresa del campo sonoro decisamente più esaustiva rispetto ai precedenti microfoni e l'abilità di processare elaborati segnali HOA (con alte direttività) e di direzionarli arbitrariamente nello spazio aprono la strada ad un'infinità di possibilità creative. Tuttavia le grandi complessità di codifica e decodifica di segnali HOA comportano grandi costi computazionali. A tal proposito il sistema 3D-VMS propone una soluzione più pratica e attuabile per l'utilizzo dell'Eigenmike, dal momento che evita le codifiche tipiche dell'Ambisonics e fornisce dei microfoni virtuali direzionabili applicando dei filtri digitali direttamente ai segnali grezzi delle 32 capsule, il tutto senza soluzione di continuità. In aggiunta, l'interfaccia grafica del software rende il suo utilizzo molto intuitivo e immediato, dando così all'utente l'opportunità di concentrarsi su aspetti della ripresa più artistici che tecnici. Grazie alla possibilità di utilizzo in realtime e in post-produzione, il sistema 3D-VMS rappresenta un nuovo approccio alla ripresa sonora per il broadcasting, la produzione musicale e cinematografica. È necessario sottolineare ancora una volta che tutti i microfoni per audio 3D qui riportati sono da considerare per tecniche di registrazione di base, ossia come strumenti in grado di catturare il campo sonoro in maniera "panoramica", senza una particolare cura del dettaglio. Se l'intento della registrazione è la ricostruzione di un campo acustico realistico un microfono di questo tipo può essere sufficiente. Se invece lo scopo è un prodotto che rispecchi certi standard estetici d'ascolto (ribadisco estetici, quindi non necessariamente realistici), allora è consigliabile affiancare una serie di microfoni di supporto, per la ripresa ravvicinata delle sorgenti.

Per quanto riguarda le tecniche di riproduzione, nonostante lo stato dell'arte avanzato dell'audio 3D i formati multicanale più diffusi ad oggi sono i

cosiddetti “surround”, in cui il suono circonda l’ascoltatore solo sul piano orizzontale. Tra questi lo stereo, il 5.1 e la sua estensione 7.1 sono le configurazioni più utilizzate nel cinema e nella musica, sia a livello consumer che professionale. La tecnica Holophonics di Zuccarelli e l’Ambisonics offrono un modo semplice per ottenere prodotti con questi formati, ma la tridimensionalità e la realistica del suono sono raramente soddisfacenti. L’Ambisonics è comunque in grado di decodificare i segnali in formati più complessi e questo permette di ricreare una spazializzazione più realistica. Tuttavia la gestione dei segnali Ambisonic è tecnologicamente complicata, soprattutto al crescere degli ordini dei segnali. A tal proposito la WFS sembra essere la tecnica di riproduzione più promettente, sia dal punto di vista di una corretta spazializzazione del campo che nell’offrire un ascolto ottimale a più persone contemporaneamente. Tuttavia le configurazioni di altoparlanti richieste da questa tecnica sono difficili da realizzare, se non in grandi spazi e con grandi risorse. Dunque l’utilizzo della WFS è attualmente limitato a installazioni artistiche e spettacoli audiovisivi, ma sono in corso ricerche per consentire di integrare questo sistema d’ascolto anche in ambienti più contenuti, come quelli domestici.

In conclusione, l’audio 3D si basa su principi fisici e psicoacustici già ben noti e oggi si sta largamente diffondendo con un nuovo e rinnovato approccio, che introduce miglioramenti e rende il prodotto finale sempre più realistico e accessibile. Le tecnologie a supporto dell’audio 3D stanno crescendo rapidamente per rispondere alle esigenze introdotte dall’evoluzione della VR, che si sta affermando non solo nell’ambito dell’intrattenimento, ma anche nel mondo dell’educazione, delle scienze sociali, del digital marketing e in molti altri campi. Inoltre l’audio, in modo analogo e complementare al video, è in grado di influenzare molti aspetti artistici nella rappresentazione di un prodotto, ma come osservato con l’avvento del video 3D, non è possibile adattare istantaneamente il modo in cui gli ascoltatori sono abituati a fruirne. Come in altri campi multimediali, infatti, l’introduzione di un nuovo paradigma d’ascolto apre la strada ad un nuovo linguaggio estetico che rende necessario anche un percorso parallelo alla crescita tecnologica, in cui “educare” l’ascoltatore alle nuove dimensioni e sfumature introdotte dall’audio 3D. Il potenziale creativo concesso dalla spazializzazione audio è virtualmente infinito e al momento ancora non del tutto esplorato ed espresso. Per questo motivo nel futuro si prevedono nuove e numerose ricerche in quest’ambito.



# Bibliografia

- [1] Hrtf measurements of a kemar dummy-head microphone. <https://sound.media.mit.edu/resources/KEMAR.html>. [Online; accessed 08-July-2019].
- [2] Augustinus J Berkhout. A holographic approach to acoustic control. *Journal of the audio engineering society*, 36(12):977–995, 1988.
- [3] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.
- [4] George H Dunteman. *Principal components analysis*. Number 69. Sage, 1989.
- [5] Eigenmike. *Beamformer Data Specification for Eigenmike Software-Beamformer*.
- [6] Angelo Farina, Marco Binelli, Andrea Capra, E Armelloni, S Campanini, and A Amendola. Recording, simulation and reproduction of spatial soundfields by spatial pcm sampling (sps). In *International Seminar on Virtual Acoustics, Valencia, Spain*, pages 24–25, 2011.
- [7] Angelo Farina, Andrea Capra, Lorenzo Chiesi, and Leonardo Scopece. A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in post production. In *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*. Audio Engineering Society, 2010.
- [8] Ken Farrar. Soundfield microphone. *Wireless World*, 85(1526):48–50, 1979.

- [9] WE Feddersen, TT Sandel, DC Teas, and LA Jeffress. Localization of high-frequency tones. *the Journal of the Acoustical Society of America*, 29(9):988–991, 1957.
- [10] William G Gardner and Keith D Martin. Hrtf measurements of a ke-mar. *The Journal of the Acoustical Society of America*, 97(6):3907–3908, 1995.
- [11] Michael A Gerzon. Periphony: With-height sound reproduction. *Journal of the Audio Engineering Society*, 21(1):2–10, 1973.
- [12] Michael A Gerzon. The design of precisely coincident microphone arrays for stereo and surround sound. In *Audio Engineering Society Convention 50*. Audio Engineering Society, 1975.
- [13] Holophone. *H2Pro Surround Sound Microphone, User Guide*.
- [14] Ulrich Horbach, Diemer de Vries, and Etienne Corteel. Spatial audio reproduction using distributed mode loudspeaker arrays. In *Audio Engineering Society Conference: 21st International Conference: Architectural Acoustics and Sound Reinforcement*. Audio Engineering Society, 2002.
- [15] Maurice Jessel. *Acoustique théorique*, volume 1. Masson, 1973.
- [16] Ole Kirkeby and Philip A Nelson. Digital filter design for inversion problems in sound reproduction. *Journal of the Audio Engineering Society*, 47(7/8):583–595, 1999.
- [17] Thomas Lund. Enhanced localization in 5.1 production. In *Audio Engineering Society Convention 109*. Audio Engineering Society, 2000.
- [18] John C Middlebrooks and David M Green. Sound localization by human listeners. *Annual review of psychology*, 42(1):135–159, 1991.
- [19] A William Mills. Lateralization of high-frequency tones. *The Journal of the Acoustical Society of America*, 32(1):132–134, 1960.
- [20] Philip McCord Morse and K Uno Ingard. *Theoretical acoustics*. Princeton university press, 1986.

- [21] Rozenn Nicol. *Restitution sonore spatialisée sur une zone étendue: application à la téléprésence*. PhD thesis, 1999.
- [22] John William Strutt Baron Rayleigh. *The theory of sound*, volume 2. Macmillan, 1896.
- [23] Leonardo Scopece. Riprese olofoniche e ambisoniche-il sistema 3d-vms.
- [24] Alois Sontacchi and Robert Höldrich. Enhanced 3d sound field synthesis and reproduction system by compensating interfering reflections. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, 2000.
- [25] Soundfield. *ST450 MKIIPortableMicrophone System, User Guide*.
- [26] Aurelio Uncini. Sintesi di circuiti adattabili con approccio statistico e informativo. *dispense del corso Teoria dei Circuiti II, Univ. Roma "La Sapienza"*, (<http://infocom.uniroma1.it/aurel/TdC2/lectures>), 2000.
- [27] Aurelio Uncini. *Audio digitale*. McGraw-Hill, 2006.
- [28] Jakob Vennerød. Binaural reproduction of higher order ambisonics-a real-time implementation and perceptual improvements. Master's thesis, NTNU, 2014.
- [29] Robert Sessions Woodworth and Harold Schlosberg. *Experimental psychology*. Oxford and IBH Publishing, 1954.