

UNIVERSITÀ DEGLI STUDI DI ROMA · TOR VERGATA

---

Macroarea di Lettere e Filosofia  
Master in Sonic Arts

## TITOLO

**Relatore:**

**Prof. Giuseppe Silvi**

**Presentata da:**

**Lorenzo Ferri**

Anno Accademico 2016/17



# Indice

<b>Elenco delle figure</b>	<b>2</b>
<b>Abstract</b>	<b>3</b>
<b>1 Dai canali audio agli Oggetti sonori</b>	<b>4</b>
1.1 Standard di riproduzione . . . . .	5
1.1.1 audio in una dimensione . . . . .	5
1.1.2 Audio in due dimensioni . . . . .	6
1.2 Stream audio object-oriented . . . . .	7
<b>2 Spazilizzatore 3D e Format MDA</b>	<b>9</b>
2.1 Format MDA . . . . .	10
2.1.1 format MDA per riproduzione audio . . . . .	13
<b>3 Metodo Wave Field Syntesis</b>	<b>14</b>
3.1 Principio fisico alla base e algoritmo di implementazione . . . . .	14
<b>Bibliografia</b>	<b>16</b>

# Elenco delle figure

1.1	Configurazione stereo . . . . .	6
1.2	Configurazione 5.1 e 7.1 . . . . .	7
2.1	Coordinate spaziali . . . . .	10
2.2	parametri di apertura e divergenza . . . . .	11
2.3	Schema esplicativo funzionamento format MDA . . . . .	12
2.4	azione di richiamo di un object-fragment . . . . .	12
3.1	Principio di Huygens-Fresnel . . . . .	15
3.2	Salti concettuali della WFS . . . . .	15

# Abstract

# Capitolo 1

## Dai canali audio agli Oggetti sonori

Siamo all'inizio dello sviluppo di questa tesi, l'abstract ci ha permesso di capire qual'è il nostro scopo e i passaggi che percorrerà questa stesura, ora non rimane che cominciare.

Nel mondo odierno l'avanzamento tecnologico ha permesso a tutti coloro che ne hanno voglia la possibilità di poter ascoltare un contenuto, basti pensare a chi ha un hifi in casa, chi un'impianto per home theater o chi direttamente va al cinema; in tutte queste situazioni l'ascoltatore medio ha accesso a questi contenuti diciamo in modo smart cioè senza preoccuparsi tanto di tutto il lavoro e la progettazione che sta dietro alla realizzazione del contenuto e sulla sua successiva riproduzione, questo è possibile grazie a delle regole che stanno alla base di tutta questa catena; un esempio chiarirà meglio l'uso improprio della parola regola.

Consideriamo per esempio un cinema, sappiamo dall'acustica che il modo in cui sono progettati i diffusori e la sala di proiezione coloreranno <sup>1</sup> in qualche maniera il suono e quindi non faranno ascoltare lo stesso contenuto sonoro che è stato creato per quel film, è quindi obbligo per il cinema (almeno lo spero) tarare l'ascolto in modo che non ci sia questa colorazione, qui per esempio il fatto di rendere non colorato il suono costituisce una regola, uno standard che è necessario affinché si abbia un giusto ascolto del contenuto.

Come quest'ultima ci sono diverse regole, diversi standard da rispettare affinché sia garantito che il produttore di contenuti sonori e l'ascoltatore abbiano accesso alla stessa informazione, nel nostro caso siccome la tesi è mirata ad un'aspetto dell'audio faremo riferimento solo a quelle regole che stanno a base della spazializzazione quindi quegli standard che hanno a che fare in qualche modo con la geometria e sui supporti di riproduzione.

---

<sup>1</sup>con il termine colorare nell'audio e in acustica si intende la tendenza di un sistema a modificare lo spettro in frequenza di un segnale audio che vive al suo interno

Ora è d'obbligo fermarsi un attimo e ripercorrergli brevemente (per quanto riguarda la spazializzazione) in modo da capire in che punto vogliamo intraprendere una strada concettualmente diversa, anche perchè alla fine del ragionamento dovremo ricollegarci ad essi.

## 1.1 Standard di riproduzione

### 1.1.1 audio in una dimensione

Il modo più semplice e basilare con cui riusciamo a riprodurre del materiale audio è la **MONOFONIA**; essa è una tecnica attuabile solo con l'ausilio di una cassa acustica e sfrutta un solo canale audio <sup>2</sup> quindi di conseguenza nello spazio sonoro <sup>3</sup> vive una sola informazione sonora.

La sensazione che abbiamo ad ascoltare questo tipo di riproduzione è di sentire una sorgente puntiforme collocata nel punto in cui è messa la cassa.

Molto simile a questo metodo è il **DUALMONO** che utilizza sempre un solo canale audio ma questo viene sdoppiato e ripartito su due altoparlanti, questo fa sì che si crei una sorgente fantasma <sup>4</sup> esattamente al centro tra la linea congiungente i due altoparlanti.

In questo caso non ho utilizzato la parola casse acustiche in quanto possiamo utilizzare questo metodo sia in queste ultime che in cuffia avendo lo stesso risultato di percezione.

Il prossimo metodo descritto è lo **STEREO**<sup>5</sup>(configurazione 2.0) in cui avendo a disposizione due speaker la sensazione di spazialità viene data dalla differenza di potenza del segnale inviata alle casse infatti se il segnale risulta più forte in una delle due sorgenti acustiche, la sorgente fantasma risulterà più spostata verso quest'ultima.

Questo avviene sia per casse in aria libera che per cuffie.

Parlando dello stereo cominciamo a parlare di standard in quanto la configurazione prevede di creare un triangolo equilatero con ai vertici i due altoparlanti e l'ascoltatore, se non si segue questa direttiva non si avrà la giusta spazialità e collocazione spaziale dei

---

<sup>2</sup>per canale audio si intende un supporto in cui scorre solo un'informazione sonora

<sup>3</sup>per spazio sonoro si intende lo spazio acustico dove si generano e si propagano le onde sonore, nei nostri casi sarà sempre uno spazio chiuso quindi di conseguenza le leggi fisiche vigenti sono quelle degli spazi chiusi

<sup>4</sup>per sorgente fantasma si intende il fenomeno per cui se due sorgenti distanti tra di loro riproducono lo stesso segnale sonoro, la nostra percezione ci porta a pensare che la so

<sup>5</sup>qui c'è da fare una precisazione: utilizzato la parola stereo perchè è comune utilizzarlo in questo caso, in verità formalmente lo stereo è una modalità di riproduzione e di registrazione che oltre ad usare la IID utilizza anche la ITD ed è come se i due canali componenti questo standard sentissero esattamente quello che sentirebbero le nostre orecchie

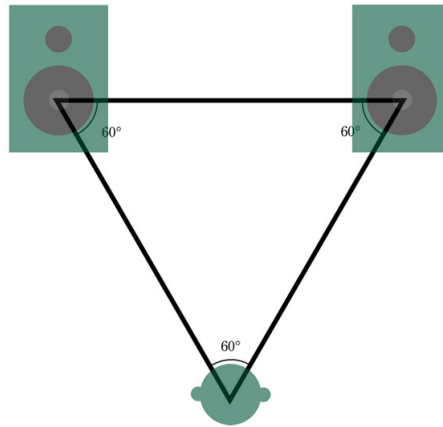


Figura 1.1: Configurazione stereo

suoni.

Questi descritti sono i metodi principali di riproduzione in cui si comincia ad intravedere un primo approccio di spazializzazione sonora, ora l'ascolto di musica si ferma generalmente qui alla riproduzione in una dimensione ma l'avvento dei film e dei cinema ha portato al concetto di audio in due dimensioni in particolare al surround.

### 1.1.2 Audio in due dimensioni

Per audio in due dimensioni si intende una riproduzione che pone l'ascoltatore al centro di un piano di diffusione in modo che si possano sentire suoni provenienti anche dai lati e da dietro; diverse sono le configurazioni possibili ma quelle che interessano a noi sono principalmente tre.

La più diffusa configurazione surround è sicuramente la **5.1**<sup>6</sup> usata principalmente dalla Dolby e dalla DTS per riproduzione di audio per film.

Questa configurazione, come spiega da specifica [2], è composta da un canale LFE per il subwoofer e di 5 canali distribuiti in 5 satelliti disposti rispettivamente a  $0^\circ$ ,  $\pm 30^\circ$  e  $\pm 110/120^\circ$

Una successiva configurazione è la **7.1** che riporta angoli di  $0^\circ$ ,  $\pm 30^\circ$ ,  $\pm 90/110^\circ$  e  $\pm 135/150^\circ$

---

<sup>6</sup>dove la prima cifra sta per il numero di diffusori nel piano e la seconda cifra sta per il numero di LFE (canale audio dedicato al subwoofer)



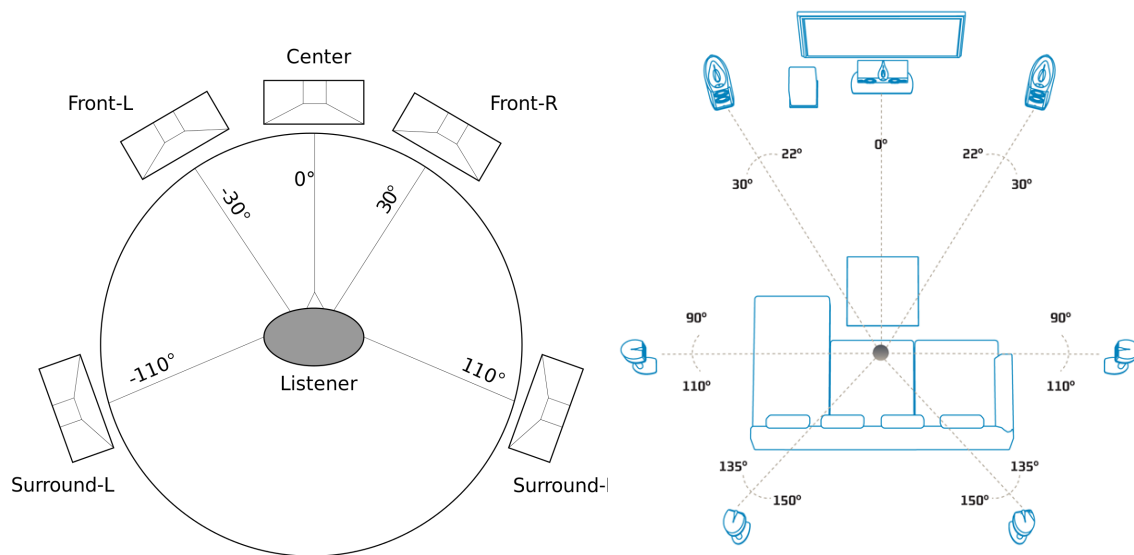


Figura 1.2: Configurazione 5.1 e 7.1

## 1.2 Stream audio object-oriented

Ora capito quali sono i principali standard di riproduzione andremo a spiegare come come queste configurazioni andranno pilotate, non ci interessa tanto il contenuto di ogni canale ma il fatto che:

- nel più semplice dei casi ad ogni speaker sarà assegnato un canale audio differente (esempio Dolby Digital 5.1 dove ad ogni altoparlante è pilotato da un segnale discreto indipendente), si veda [3].
- nei casi più complessi ad ogni altoparlante sarà assegnato un mix di canali dato da una matrice di decodifica proprietaria (esempio si guardi il funzionamento del decoder Dolby Pro Logic), si veda [4].

Un'altra importante considerazione è il fatto che lo stesso stream che utilizzo in un sistema di grado maggiore per esempio un 5.1 posso scalarlo in un sistema di grado minore esempio uno stereo, questo perchè essendo tutti questi degli standard è possibile fare un downgrade dello stream oppure un upgrade (solo però sotto alcuni artifici per essere adattato al sistema di riproduzione; allora viene da chiedersi perchè cambiare concezione e approdare in uno stream object-oriented?

La domanda trova facile risposta nel fatto che questi giochi di upgrade o downgrade si possono fare solo su sistemi che hanno una certa grado di compatibilità (ad esempio i sistemi Dolby) e se questo avvenisse dovrei comunque prendere uno stream, codificarlo

in uno stream diciamo generale e riadattarlo decodificandolo nello stream che piloterà la mia riproduzione.

Allora la domanda sorge spontanea, perchè non creare uno stream diciamo generale?

La Dolby laboratories ha già implementato una tecnologia che risponde in questa domanda, infatti nel suo ultimo brevetto **Dolby Atmos** (si veda documentazione [5]) ha implementato nello stream di riproduzione anche una parte dedicata agli oggetti sonori, infatti oltre ad avere un tappeto sonoro dato dalla configurazione 7.1, la configurazione dolby atmos permette la creazione di 128 oggetti indipendenti nello spazio sonoro 3D che vengono riprodotti grazie anche all'ausilio di speaker posti al di sopra dell'ascoltatore, ma non ci soffermeremo troppo su questa configurazione, piuttosto spenderemo tempo a parlare di come Dolby é riuscita incanalare gli oggetti e la relativa informazione spaziale dentro questo stream.

Normalmente quando si crea un prodotto in fase di post-produzione si fa un down-mix del prodotto per passare dal contenuto dei singoli canali allo stream di riproduzione, per esempio chiudendo un mix in uno studio di registrazione classico miscelo i vari canali nello stream audio stereo con peso costituito dalla funzione di panpot deciso per ogni canale, questo succede anche in configurazioni più complicate come il 5.1 l'unica differenza è che panpot non sarà più in una dimensione ma in due<sup>7</sup>.

L'idea della Dolby invece è stata quella di fermarsi prima della fase di downmix (almeno per quanto riguarda il contenuto degli oggetti sonori) e di intraprendere una strada diversa:

Prima di tutto si devono preparare gli oggetti sonori esattamente come si vuole siano riprodotti, in secondo luogo si dispongono questi oggetti in uno spazio di riproduzione virtuale creato appositamente da un qualche software o plugin nel quale a ogni oggetto verrà assegnato un punto preciso nello spazio attraverso delle coordinate spaziali.

Questo sostanzialmente sta alla base dell'audio ad oggetti, nel prossimo capitolo vedremo come concretamente implementare questo nuovo procedimento.

---

<sup>7</sup>esiste una funzione di panpot standard per ogni configurazione di riproduzione

## Capitolo 2

# Spazializzatore 3D e Format MDA

In questo capitolo andremo a spiegare concretamente come possiamo creare questo stream audio ad oggetti e come in fase di post produzione posso spazializzare il mio contenuto sonoro.

Come accennato prima in fase di post-produzione non usiamo più il solito panner installato in banchi mixer o in software DAW in quanto essi sia che siano stereo o 5.1 o quello che vogliamo, non hanno una spazializzazione a oggetti quindi la cosa da fare è implementare un software apposito che, sotto forma di stand-alone o di plug-in, dia in uscita gli oggetti sonori con la loro relativa posizione (in realtà oltre che alla posizione bisognerà tener conto anche di altri paramentri come grandezza, volume ecc...).

Parlando fino ad ora di Dolby è plausibile che prenderemo come panner 3D e come format<sup>1</sup> quello proprietario di Dolby Atmos ma il fatto è che essendo questa tecnica di spazializzazione soggetta a brevetto, essa é diciamo chiusa a sole applicazioni che riguardano il mondo Dolby, altri marchi hanno prodotto format proprietari simili ma sempre vincolati al brevetto, format quali Auro3D per citarne uno.

Noi invece ricerchiamo qualcosa che sia un format open e libero in grado di adattarsi e a diventare uno standard; la **Digital Theater System DTS** (storica rivale di Dolby) anche lei interessata all'audio ad oggetti ha creato un format open che fa al caso nostro: questo si chiama **Multi-Dimesional Audio MDA**.

---

<sup>1</sup>per format intendo lo stesso concetto che in precedenza ho usato con la parola stream

## 2.1 Format MDA

Si veda [6] per la documentazione riguardante questo paragrafo.

Il gioco che sta alla base di questo format è la scrittura di **metadati** contenenti gli attributi dell'oggetto sonoro da spazializzare, questo format essendo principalmente stato creato per il cinema avrà attributi che per la sola riproduzione musicale sono non necessari ma di questo ne parleremo dopo.

Questi attributi sono:

- **Coordinate Spaziali:** sono un set di 3 valori che indicano la posizione dell'oggetto sonoro in relazione con la posizione dell'ascoltatore che avrà come valori la terna (0,0,0).

Come coordinate abbiamo l'asse  $X$  posto di fianco all'ascoltatore, l'asse  $Y$  posto di fronte e l'asse  $Z$  posto verso l'alto, quindi definito un angolo di azimuth  $\theta$ , un'angolo di elavazione  $\phi$  e un raggio  $\rho$  posso definire le coordinate polari come:

$$\begin{cases} x = \rho \sin(\theta) \cos(\phi) \\ y = \rho \cos(\theta) \cos(\phi) \\ z = \rho \sin(\phi) \end{cases} \quad (2.1)$$

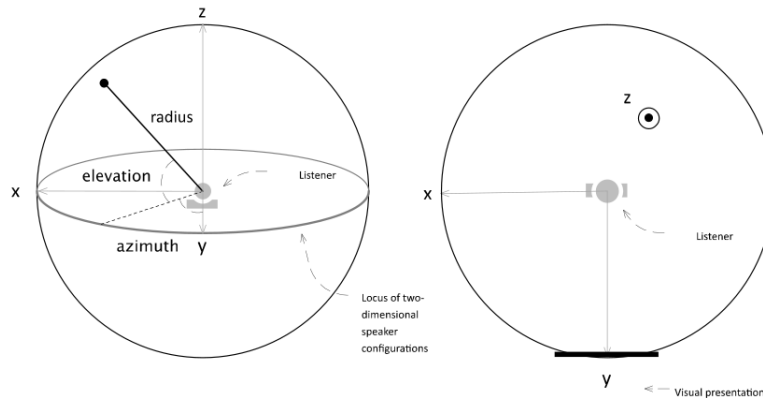


Figura 2.1: Coordinate spaziali

- **Identificativo oggetto sonoro:** scritto come URI é un numero (int) che identifica l'oggetto sonoro, sarebbe come un indirizzo.
- **Gain:** è un valore che indica la quantità di segnale che deve avere l'oggetto sonoro.

- **apertura:** scritta in gradi indicherebbe la grandezza dell'oggetto sonoro da ricreare (più l'oggetto è grande più gli altoparlanti posti marginalmente della sorgente fantasma si attivano, esempio se il valore è  $180^\circ$  allora si attiverebbero tutti i diffusori)
- **divergenza:** anch'essa indicata in gradi quantifica la grandezza che deve avere l'oggetto sonoro ma solo sul piano orizzontale, un esempio più esplicativo si ha in figura 2.2.

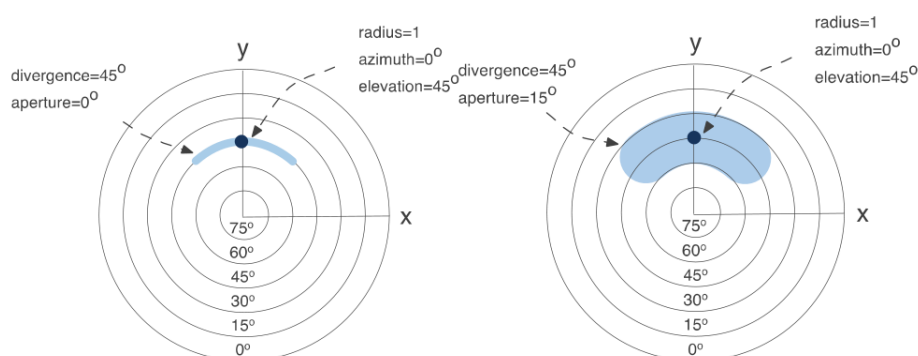


Figura 2.2: parametri di apertura e divergenza

queste sono i principali parametri scritti nei metadati che ci servono per la spazializzazione, abbiamo quindi un oggetto sonoro e i suoi metadati ma è facile pensare che se l'oggetto si muove nello spazio o cambia dimensioni il contenuto dei suoi metadati cambia nel tempo, quindi il format racchiude in un nuovo oggetto chiamato **Object-Fragment** dove saranno segnati tutti metadati descritti sopra (che per tutto lo svolgimento temporale dell'object fragment rimarranno inalterati) con più l'aggiunta di alcuni parametri come l'**ID** (identificativo oggetto), e i sample dell'oggetto sonoro da riprodurre.

Sopra tutto poi c'è una timeline (figura 2.4) che ha la funzione di richiamare gli object-fragment quando servono, essa è suddivisa in segmenti temporali che sono dati da  $\frac{1}{f_c}$  dove  $f_c$  è la frequenza di campionamento impostata per la totalità degli oggetti sonori, e dove a ogni segmento è associato una lista di ID che richiama oggetti da riprodurre, un'esempio esplicativo è dato dallo schema 2.3.

Da notare che è presente anche un LFE-Object in cui non è segnata la posizione, questo perchè data la natura omnidirezionale delle basse frequenze non avrebbe senso collocarle spazialmente e anche perchè solo il/i subwoofer sono in grado di riprodurre tali frequenze.

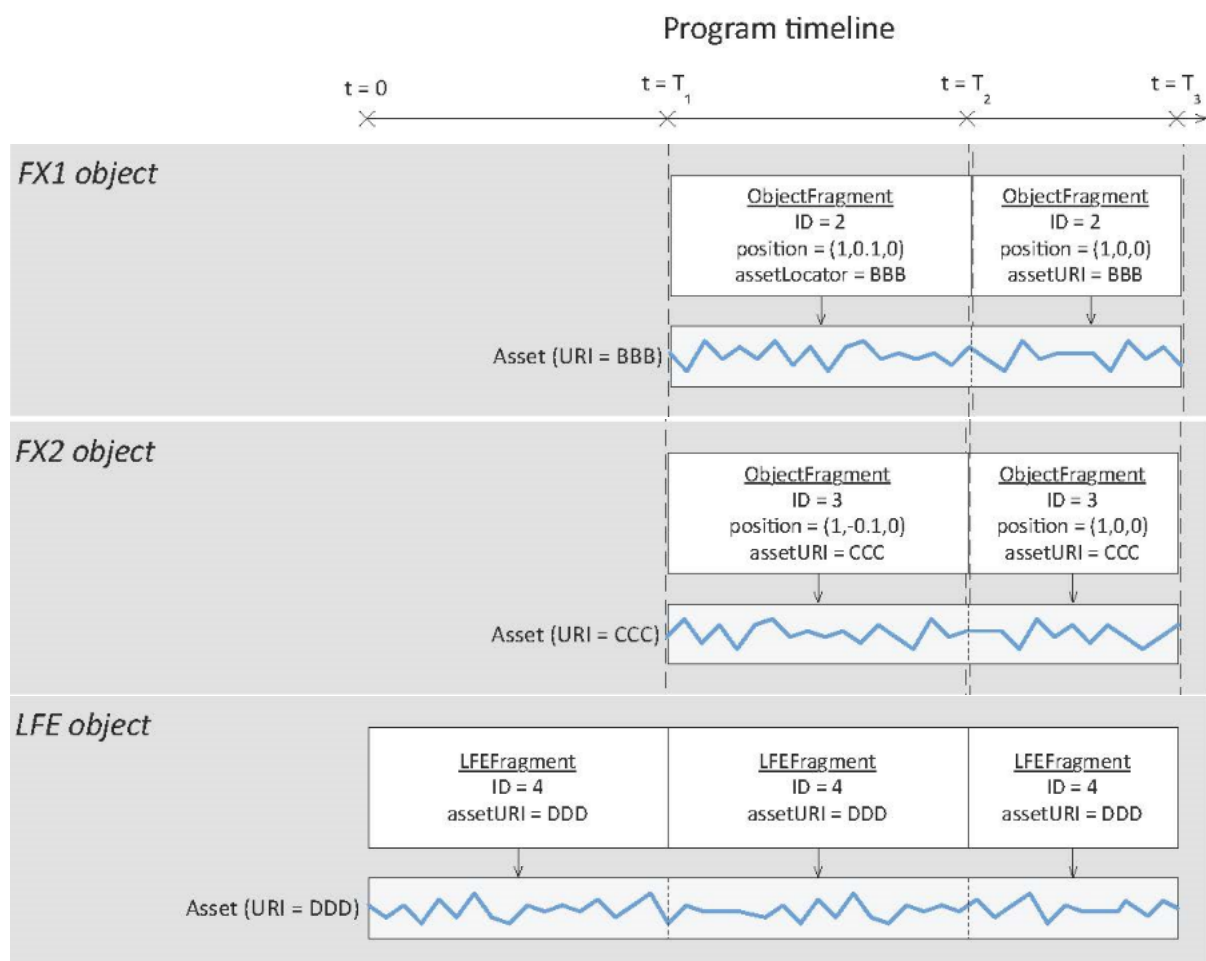


Figura 2.3: Schema esplicativo funzionamento format MDA

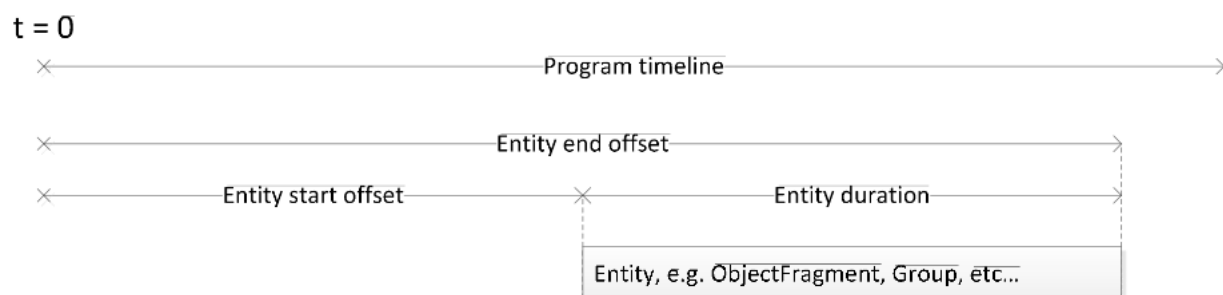


Figura 2.4: azione di richiamo di un object-fragment

### 2.1.1 format MDA per riproduzione audio

Qui faccio una piccola annotazione personale che mi tornerà utile più avanti, nella maggior parte dei casi quando si fa musica in studio di registrazione è abbastanza difficile che abbiamo oggetti sonori in movimento e di dimensioni che variano nel tempo, quindi si potrebbe fare una piccola modifica al format in modo da alleggerire il successivo rendering per creare fisicamente lo spazio sonoro.

Sostanzialmente abbiamo tre tipi di oggetti sonori in musica: oggetti fermi, oggetti che balzano tra due punti spaziali ed oggetti costituiti da effetti vari (il più importante per la spazializzazione è il riverbero e per questo ci occuperemo solo di questo).

tutti questi oggetti possono essere pensati come oggetti statici quindi che hanno coordinate spaziali e dimensioni fisse, in questo caso si potrebbe alleggerire il format in quanto non avrebbe più senso parlare di object-fragment (ricordiamo che gli attributi di spazializzazione cambiano allora si avranno diversi object-fragment, uno per ogni configurazione spaziale) in quanto per l'intera esecuzione del brano si avrebbe solo un object-fragment che fissa la posizione e la grandezza dell'oggetto, quindi si possono direttamente assegnare queste due ultime all'oggetto sonoro senza passare per ulteriori suddivisioni.

per quanto riguarda gli oggetti statici questo trucco calza alla perfezione, per il riverbero per esempio basterà avere come oggetto il solo contenuto del brano riverberato e assegnarli una giusta divergenza e apertura; per quanto riguarda gli oggetti balzanti (come potrebbe essere per esempio un tremolo fatto con un pan-pot in una tastiera) basterà creare in fase di post-produzione due oggetti diversi che hanno lo stesso contenuto sonoro e che differiscono soltanto che l'intensità sonora  $I' = \alpha I$  di un'oggetto corrisponderà a una intensità  $I'' = (1 - \alpha)I$  del secondo oggetto (dove  $I$  è l'intensità che dovrebbe avere originariamente l'oggetto).

## Capitolo 3

# Metodo Wave Field Syntesis

Il metodo Wave Field Syntesis è un metodo diverso dai precedenti illustrati in quanto non si avvale della psicoacustica per ingannare la nostra mente e farci credere che stiamo ascoltando qualcosa che realmente non c'è, ma questa tecnica permette di ricreare fisicamente il fronte d'onda e quindi l'informazione sonora distribuita nello spazio acustico come se la sorgente che si vuole creare sia realmente collocata nel punto che vogliamo, per questo, forse anche impropriamente, catalogherò questo metodo come **AUDIO 3D**.

### 3.1 Principio fisico alla base e algoritmo di implementazione

Per riuscire a capire in fondo cosa sta alla base di questa tecnica bisognerà spiegare due semplici principi di meccanica ondulatoria:

- **Principio di Huygens-Fresnel:** consideriamo una qualsiasi onda che abbia un fronte d'onda arbitrario, questa legge afferma che ogni punto del fronte d'onda in questione può essere visto come un'infinità di sorgenti secondarie puntiformi che generano un'infinità di fronti d'onda secondari in accordo in fase e in ampiezza e che sommando la totalità di questi ultimi si può ricostruire il fronte d'onda originale.
- **Principio di Rayleigh:** questo principio riguarda la diffrazione in quanto se un fronte d'onda colpisce una fenditura di dimensioni paragonabili alla sua lunghezza d'onda, esso verrà ritrasmesso al di là della fenditura come se fosse una sorgente puntiforme.

Il salto concettuale ora è breve in quanto se una sorgente acustica reale emette un fronte d'onda ed esso impatta in una serie di fenditure disposte spazialmente in un modo preciso, il fronte d'onda passerà al di ognuna delle fenditure (principio di Rayleigh) e



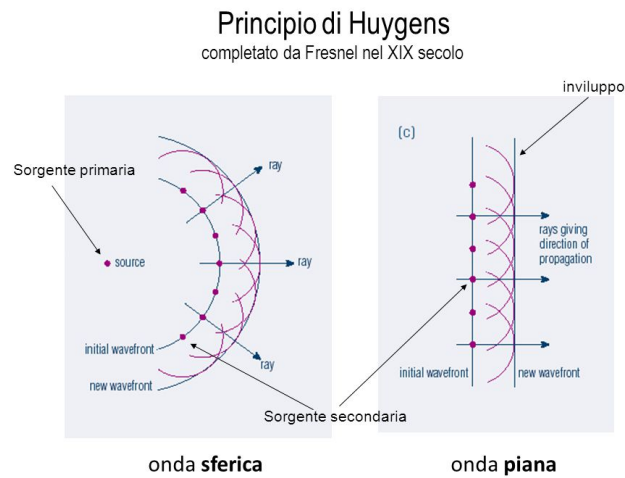


Figura 3.1: Principio di Huygens-Fresnel

la somma della totalità dei fronti d'onda secondari ricreerà esattamente il fronte d'onda originale.

Ora l'unica cosa che è rimasta da fare è sostituire ogni fenditura con un altoparlante e far riprodurre ad esso un segnale preciso che combinato con i segnali degli altri altoparlanti ricreerà fedelmente (almeno a livello concettuale) lo spazio sonoro che vogliamo ottenere

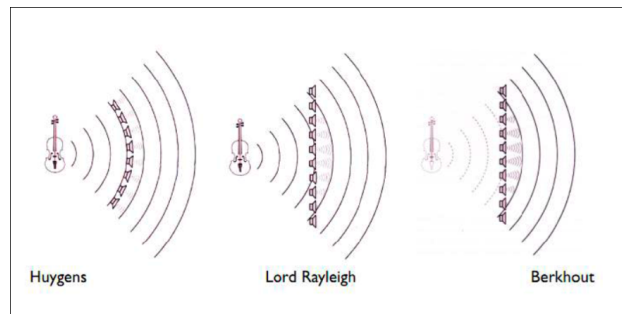


Figura 3.2: Salti concettuali della WFS

Ora le questioni che ci vengono naturali sono come e con quale segnale pilotare ogni altoparlante; le risposte possono essere molteplici ma tutte devono tener conto della geometria di progettazione del nostro sistema WFS, prendiamo uno dei casi più semplici.

# Bibliografia

- [1] [https://en.wikipedia.org/wiki/Huygens-Fresnel\\_principle](https://en.wikipedia.org/wiki/Huygens-Fresnel_principle)
- [2] [https://www.itu.int/dms\\_pubrec/itu-r/rec/bs/R-REC-BS.775-3-201208-I!!PDF-E.pdf](https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.775-3-201208-I!!PDF-E.pdf)
- [3] [https://it.wikipedia.org/wiki/Dolby\\_Digital](https://it.wikipedia.org/wiki/Dolby_Digital)
- [4] [https://it.wikipedia.org/wiki/Dolby\\_Surround\\_Pro\\_Logic](https://it.wikipedia.org/wiki/Dolby_Surround_Pro_Logic)
- [5] <https://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-next-generation-audio-for-cinema-white-paper.pdf>
- [6] [http://www.etsi.org/deliver/etsi\\_ts/103200\\_103299/103223/01.01.01\\_60/ts\\_103223v010101p.pdf](http://www.etsi.org/deliver/etsi_ts/103200_103299/103223/01.01.01_60/ts_103223v010101p.pdf)