

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

Elenco delle figure	2
Abstract	3
1 Dai canali audio agli Oggetti sonori	4
1.1 Standard di riproduzione	5
1.1.1 audio in una dimensione	5
1.1.2 Audio in due dimensioni	6
1.1.3 audio in tre dimensioni	7
1.2 Stream audio object-oriented	9
2 Spazializzatore 3D e Format MDA	11
2.1 Format MDA	12
2.1.1 format MDA per riproduzione audio	15
2.2 Spazializzatore 3D	15
3 Decodifica format MDA per adattamento a varie configurazioni	18
3.1 tecnica VBAP	18
3.1.1 vbap 2D	19
3.1.2 VBAP 3D	21
3.2 integrazione del formato MDA con la tecnica VBAP	23
3.2.1 utilizzo MDA in geometrie tridimensionali	23
3.2.2 utilizzo MDA in geometrie bidimensionali	24
3.2.3 esempio di integrazione con sistema planare (quadrifonico e 5.1) .	25
4 Metodo Wave Field Syntesis	27
4.1 Principio fisico alla base e algoritmo di implementazione	27
Bibliografia	30

Elenco delle figure

1.1	Configurazione stereo	6
1.2	Configurazione 5.1 e 7.1	7
1.3	Configurazione quadrifonica e ottofonica	8
1.4	Configurazione ottofonica tridimensionale	8
2.1	Coordinate spaziali	12
2.2	parametri di apertura e divergenza	13
2.3	Schema esplicativo funzionamento format MDA	14
2.4	azione di richiamo di un object-fragment	14
2.5	MDA Creator, DTS technology	16
3.1	composizione vettoriale delle sorgenti reali e virtuale	19
3.2	angoli delle sorgenti reali e virtuale	20
3.3	composizione vettoriale delle sorgenti reali e virtuale in tre dimensioni . .	21
3.4	configurazione a più altoparlanti per la tecnica VBAP in tre dimensioni .	22
4.1	Principio di Huygens-Fresnel	28
4.2	Salti concettuali della WFS	28

Abstract

Nel corso di questa tesi si cercherà di rispondere ad una semplice domanda: posso svincolarmi dal fatto che un format audio possa essere riprodotto solo dalla configurazione di riproduzione adatta a quel format?

La risposta è sì, in quanto esistono già tecniche che trasformano un format in un'altro adatto ad un'altra configurazione di riproduzione ma;

cercherò di introdurre, nella prima parte di questa tesi, un **format** generale che non ha bisogno di dover passare da questo stadio di trasformazione (in quanto esso stesso non è inizialmente adattato per nessuna configurazione di riproduzione, sto parlando del format **Multi-Dimesional Audio MDA** orientato verso l'**audio ad oggetti** (object-oriented)).

In una seconda parte invece cercherò di implementare questo format in alcune delle più comuni configurazioni di riproduzione servendomi della tecnica di spazializzazione **Vector-Base Amplitude Panning VBAP** di cui spiegherò anche il funzionamento.

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 ¹ 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.

¹con 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 ² quindi di conseguenza nello spazio sonoro ³ 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 ⁴ 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**⁵(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

²per canale audio si intende un supporto in cui scorre solo un'informazione sonora

³per 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

⁴per 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

⁵qui 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

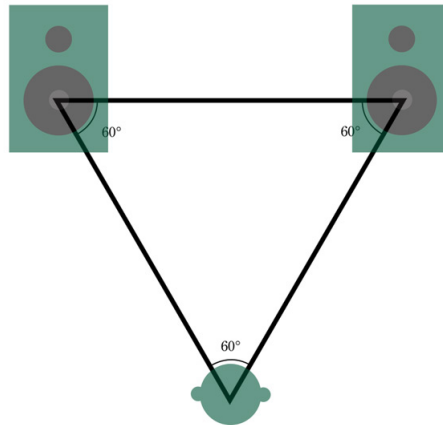


Figura 1.1: Configurazione stereo

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**⁶ 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° , $\pm 30^\circ$ e $\pm 110/120^\circ$

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

⁶dove 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)

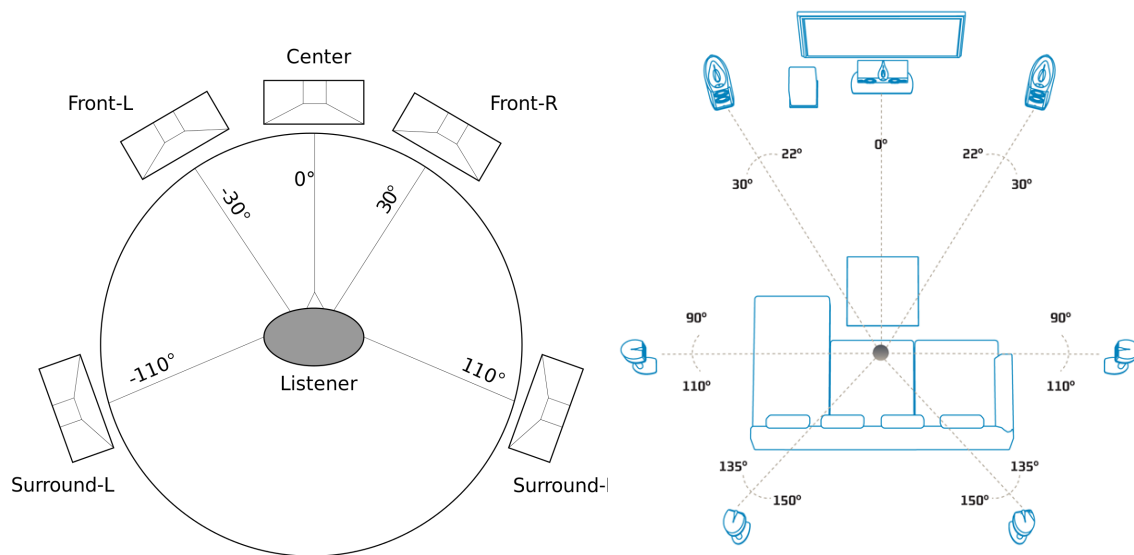


Figura 1.2: Configurazione 5.1 e 7.1

In parallelo una configurazione che trova il suo utilizzo nel campo della ricerca musicale e in minor misura nell'ascolto di musica è la **quadrifonia**; questa tecnica prevede l'utilizzo di 4 speaker posti sui vertici di un quadrato (angoli di $\pm 45^\circ$ e $\pm 135^\circ$) con al centro l'ascoltatore, anche questa permette una ricostruzione totale sul piano orizzontale ma ha il vantaggio di avere gli angoli tra gli speaker uguali (90°) il che rende più semplice un'ipotetica elaborazione di segnale per pilotarla

Un successivo utilizzo di questa tecnica è stata sviluppata implementando otto altoparlanti ad angoli rispettivi di 45° il che pone questi ultimi a $\pm 22,5^\circ$, $\pm 67,5^\circ$, $\pm 112,5^\circ$, $\pm 157,5^\circ$ come in figura 1.3

1.1.3 audio in tre dimensioni

Per finire possiamo introdurre anche l'altezza nelle nostre configurazioni, infatti sempre più spesso anche nell'uso commerciale (soprattutto nei cinema e ora anche negli home cinema) si cominciano ad utilizzare anche speaker posti sopra l'ascoltatore in modo da riprodurre oggetti sonori provenienti da quella direzione, un'esempio commerciale che da bene l'idea è la tecnologia Dolby Atmos (si veda nel paragrafo 2 la spiegazione di quest'ultimo) ma noi per scopi diciamo accademici prendiamo come esempio la disposizione ottofonica tridimensionale composta da 2 piani orizzontali quadrifonici con speaker ad angoli posti al di sopra e al di sotto dell'ascoltatore ad angoli di $\pm 45^\circ$ e $\pm 135^\circ$ con elevazione di angoli $\pm 45^\circ$ come in figura 1.4

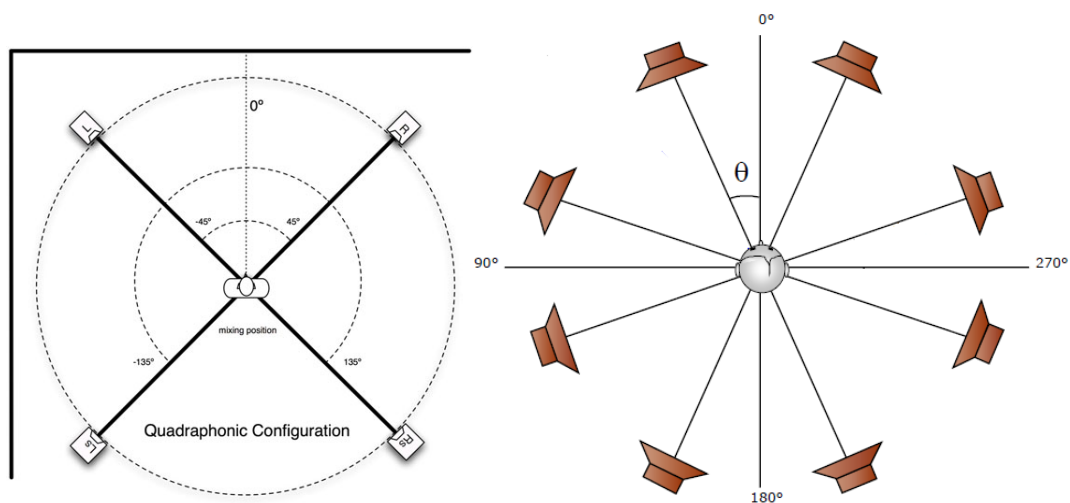


Figura 1.3: Configurazione quadrifonica e ottofonica

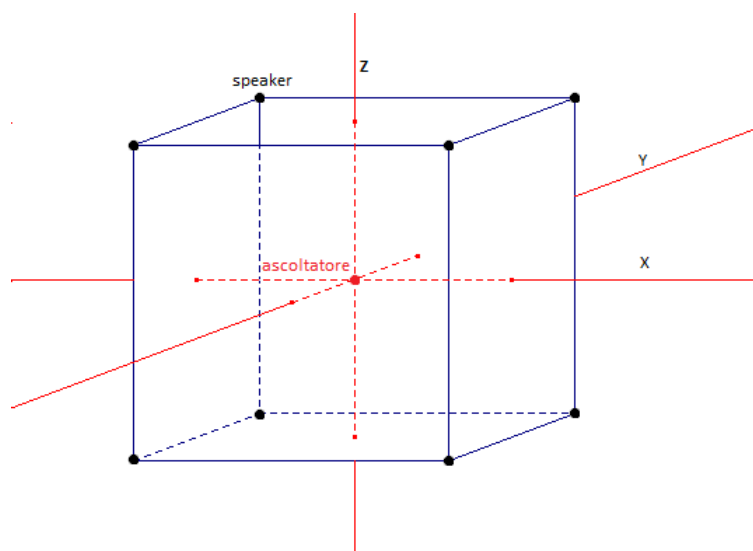


Figura 1.4: Configurazione ottofonica tridimensionale

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 scarlo 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 due⁷.

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.

⁷esiste 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 format¹ 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**.

¹per 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 Sferiche:** 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 sistema di 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 definiti i valori x, y e z su queste rette possiamo ricavare un angolo di azimuth θ , un'angolo di elavazione ϕ e un raggio ρ posso trasformare le coordinate cartesiane in sferiche in questo modo:

$$\begin{cases} x = \rho \sin(\theta) \cos(\phi) \\ y = \rho \cos(\theta) \cos(\phi) \\ z = \rho \sin(\phi) \end{cases} \Rightarrow \begin{cases} \theta = \arctg\left(\frac{x}{y}\right) \\ \phi = \arctg\left(\frac{\sqrt{y^2 + x^2}}{z}\right) \\ \rho = \sqrt{x^2 + y^2 + z^2} \end{cases} \quad (2.1)$$

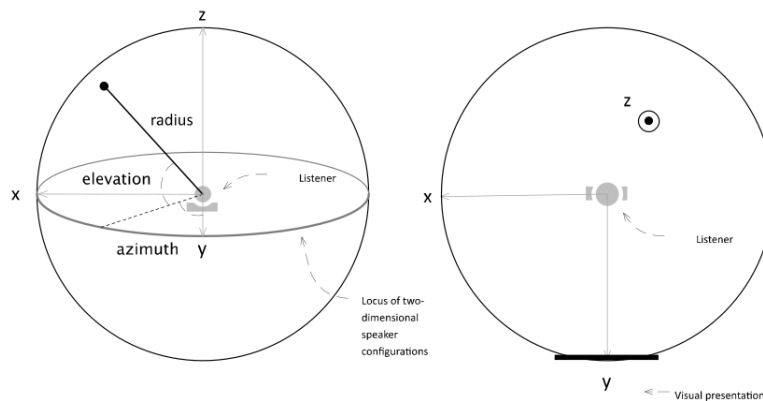


Figura 2.1: Coordinate spaziali

- **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° 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.

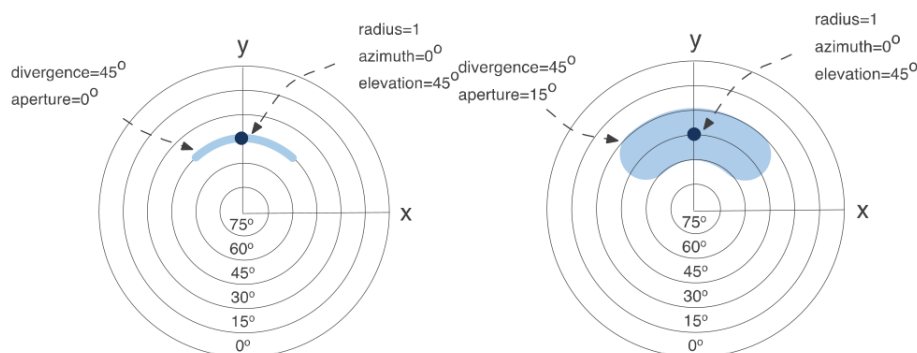


Figura 2.2: parametri di apertura e divergenza

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.

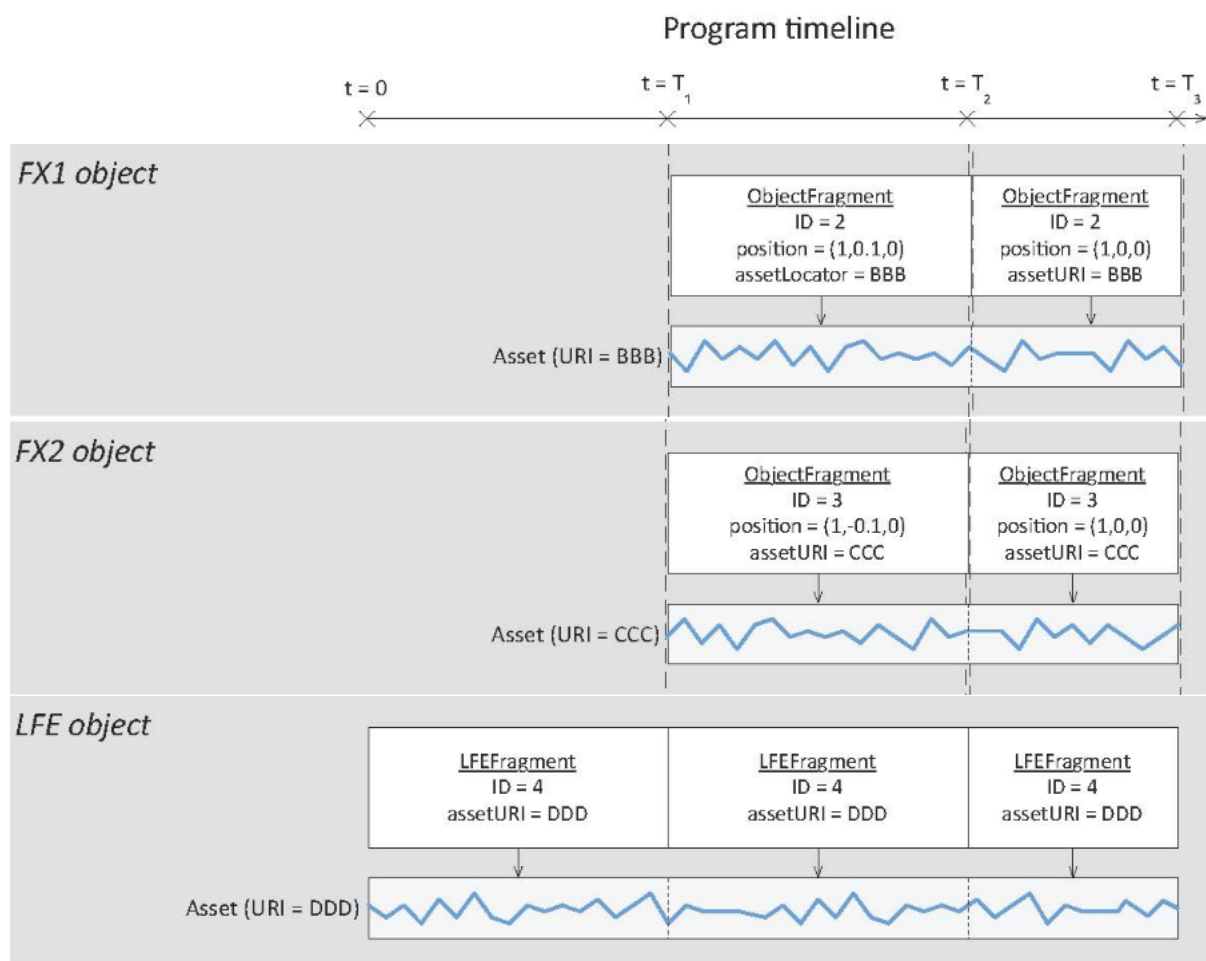


Figura 2.3: Schema esplicativo funzionamento format MDA

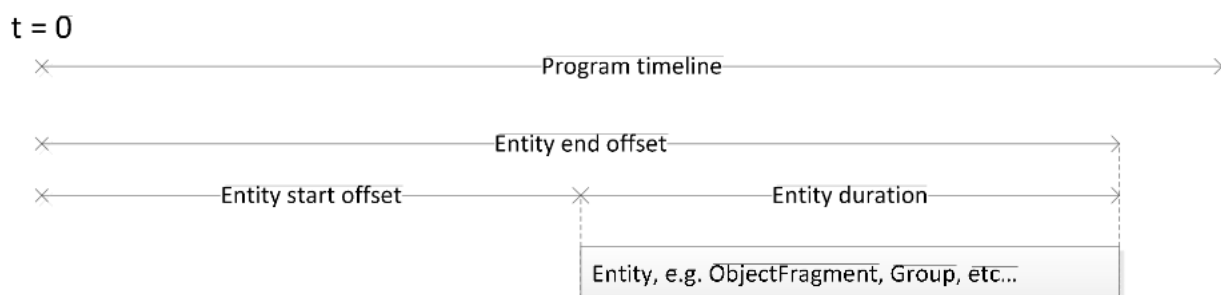


Figura 2.4: azione di richiamo di un object-fragment

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.

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 sottodivisioni.

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).

2.2 Spazializzatore 3D

Precedentemente abbiamo visto come è costituito il format MDA ora non ci rimane di piegare produrre materiale compatibile con questo format.

Nel workflow di produzione di materiale audio 3D bisognerà fermarsi quindi prima del mixdown (sia che sia stereo, 5.1 ecc...) e bisognerà prendere un qualche software che sia in grado di sostituire il mio mixer e sostituirlo con uno 3D, diverse aziende stanno facendo software di questo tipo in quanto vogliono interfacciarsi a questo format ed in qualche modo creare un collegamento tra il loro format proprietario e l'MDA; per esempio la Dolby ha un panner 3D per Dolby Atmos mka che si interfaccia con l'MDA, anche la Auro Technologies ha adottato la stessa politica o in alternativa la casa Fairlight ha creato uno spazializzatore di nome 3DAW (secondo me molto interessante) che supporta anche l'MDA.

Detto questo ho ricercato uno spazializzatore adatto e la mia scelta è ricaduta proprio sull'**MDA creator** proprietaria della DTS in quanto è la più intuitiva e la miglior scelta per integrazione (in quanto essa ha creato sia il panner che il format).



Figura 2.5: MDA Creator, DTS technology

Tutte le informazioni su MDA creator verranno prese dal suo manuale [7].

questo plugin disponibile solo per Protools è di semplice comprensione, come da manuale bisogna mettere in insert in ogni traccia questo plugin (condizione necessaria è che tutte le tracce devono avere la stessa lunghezza), poi aprendo l'interfaccia del panner bisognerà posizionare ogni oggetto nel punto in cui si desidera e con divergenza e apertura voluta (si veda il paragrafo 2.1 per sapere i parametri del format), logicamente questi parametri possono essere automatizzati per spostare e modificare l'oggetto nel tempo.

Inoltre si possono creare dei Bed Object che sarebbero oggetti i quali vengono direttamente codificati nel sistema di riproduzione scelto (esempio classico è la colonna sonora in 5.1 nei film) ai quali verranno sommati gli oggetti veri e propri del format e l'LFE-Object per gli effetti a bassa frequenza.

Una volta fatto questo l'MDA creator metterà a disposizione tre scelte di output, quella che interessa a noi è la modalità di esportazione con mda file: questa scelta ci porta alla creazione dei file **.map**, **.mix** e **.mda**, quest'ultimo è proprio il file contenente i metadati e gli oggetti sonori proposti in questo format.

Capitolo 3

Decodifica format MDA per adattamento a varie configurazioni

In questo capitolo parleremo di come invece da un format MDA posso passare a una configurazione di riproduzione usuale, mediante alcune tecniche quali il VBAP, il BINAURALE e la WFS.

3.1 tecnica VBAP

Il **Vector Base Amplitude Panning VBAP** è una tecnica di spazializzazione che permette di localizzare una sorgente attraverso la differenza di intensità, mi spiego meglio:

uno dei due parametri con cui il nostro cervello discrimina la direzione di una sorgente è la IID (si veda [8]) che indica la differenza di intensità che le nostre orecchie devono percepire per far sì che si percepiamo una sorgente fantasma in un punto dello spazio, quindi se riproduciamo un segnale coerente (cioè in fase fra gli altoparlanti) con minimo di due di essi e gli riproduciamo con intensità diverse, allora ci sembrerà di sentire il segnale proveniente da un punto previsto dalla IID.

Ora la cosa diventa abbastanza semplice in quanto se vogliamo ricreare una spazializzazione 2D basteranno due speaker, se invece vogliamo fare un 3D allora serviranno un minimo di 3 speaker posti a triangolo, configurazioni del genere esistono già e sono state ampiamente testate ed affinate, per tutta la documentazione riguardante il VBAP si faccia riferimento a [9].

E' d'obbligo affrontare un po di matematica per riuscire poi ad adattare il VBAP alle configurazioni esposte in 1.1, partiamo subito con il caso bidimensionale con 2 speaker

affrontando la cosa direttamente con calcolo matriciale (all'inizio più difficile ma meglio generalizzabile per dopo).

3.1.1 vbap 2D

Consideriamo un sistema fatto da due monitor, definiamo ora arco attivo la porzione di arco compresa tra i due speaker, ora conoscendo la posizione della nostra sorgente virtuale collocata in punto dell'arco attivo (cosa che si adatta esattamente al nostro caso) l'unica cosa che non sappiamo sono i gain g_1 e g_2 da applicare ad ogni segnale che vanno rispettivamente alle due casse per ricostruire esattamente lo spazio sonoro, quindi definendo un sistema di riferimento cartesiano \mathbf{X}, \mathbf{Y} definirò come $\mathbf{l}_1 = [l_{11} \ l_{12}]^T$ $\mathbf{l}_2 = [l_{21} \ l_{22}]^T$ i vettori unitari che puntano verso i due altoparlanti e $\mathbf{p} = [p_1 \ p_2]^T$ il vettore unitario che punta verso la sorgente virtuale.

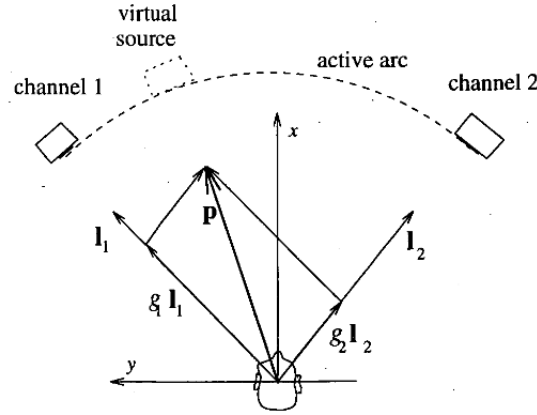


Figura 3.1: composizione vettoriale delle sorgenti reali e virtuale

Il trucco ora sta nel fatto che posso riscrivere il vettore \mathbf{p} in funzione del vettore contenente i due gain $\mathbf{g} = [g_1 \ g_2]$ come:

$$\mathbf{p} = \mathbf{g} \cdot \mathbf{l} = [g_1 \ g_2] [\mathbf{l}_1 \ \mathbf{l}_2]^T = g_1 \mathbf{l}_1 + g_2 \mathbf{l}_2 \quad (3.1)$$

In questo caso però possiamo ulteriormente compattare la scrittura in quanto:

$$\mathbf{p} = g_1 [l_{11} \ l_{12}]^T + g_2 [l_{21} \ l_{22}]^T = [g_1 \ g_2] \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \quad (3.2)$$

ora trovando tutto in funzione \mathbf{g} abbiamo che:

$$[g_1 \ g_2] = [p_1 \ p_2] \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1} \Rightarrow \mathbf{g} = \mathbf{p}^T \mathbf{L}_{12}^{-1} \quad (3.3)$$

Logicamente assumo che la matrice \mathbf{L}_{12} ammetta l'inversa, ora parametrizzo tutto in funzione di ϕ e ϕ_0 nel modo seguente:

$$p_1 = \cos(\phi), \ p_2 = \sin(\phi), \ l_{11} = \cos(\phi_0), \ l_{12} = \sin(\phi_0), \ l_{21} = \cos(-\phi_0), \ l_{22} = \sin(-\phi_0)$$

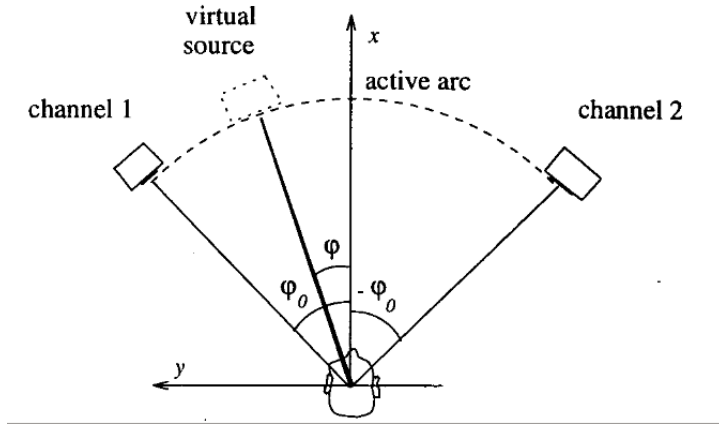


Figura 3.2: angoli delle sorgenti reali e virtuale

fatto questo, essendo il nostro spazio ortogonale, posso calcolare direttamente i coefficienti dei due gain invertendo prima la matrice, il risultato è il seguente:

$$g_1 = \frac{\cos\phi \sin\phi_0 + \sin\phi \cos\phi_0}{2\cos\phi_0 \sin\phi_0} \quad (3.4)$$

$$g_2 = \frac{\cos\phi \sin\phi_0 - \sin\phi \cos\phi_0}{2\cos\phi_0 \sin\phi_0}$$

Logicamente introduciamo un coefficiente C che indica il gain complessivo (ci servirà per definire la distanza dell'oggetto) e che scalerà i nostri due gain in questo modo:

$$[g_1 \ g_2]_{scaled} = \frac{\sqrt{C} [g_1 \ g_2]}{\sqrt{g_1^2 + g_2^2}} \quad (3.5)$$

Logicamente sarebbe molto limitativo usare questa tecnica per ricreare una riproduzione stereo, infatti con qualche accorgimento nulla ci vieta di poter estendere ad n

diffusori questa tecnica, l'unica accortezza è che il nostro procedimento dovrà selezionare solo due speaker alla quale attribuire la realizzazione della sorgente virtuale, questa scelta si basa sulla posizione dell'oggetto infatti quest'ultimo potrà essere messo solo in un'arco attivo (l'intero arco giro è suddiviso in archi attivi che non si sovrappongono) quindi le due casse che delimitano questo arco, saranno la coppia imputata a svolgere la riproduzione; faccio un piccolo esempio:

consideriamo come numero n arbitrario di casse consecutive a_n disposte ad un angolo $\phi_{0,n}$ tali che $0^\circ \leq \phi_{0,1}, \phi_{0,2}, \dots, \phi_{0,n} < 360^\circ$, posizionando un qualsiasi oggetto virtuale in un qualsiasi angolo ϕ' dovrà verificarsi che $\phi_{0,m} \leq \phi' \leq \phi_{0,m+1}$ dove m sta per il numero dello speaker adiacente alla sorgente virtuale ma con angolazione minore, quindi in questo caso la coppia di casse da selezionare saranno a_m e a_{m+1} , quindi per rendere effettivo il metodo descritto nel caso di due speaker dovremo fare una piccola modifica in quanto dovremo assumere:

$$\phi_0 = \frac{\phi_{0,m+1} - \phi_{0,m}}{2} \quad \phi = \phi' - \frac{\phi_{0,m+1} + \phi_{0,m}}{2} \quad (3.6)$$

3.1.2 VBAP 3D

Capito qual'è il modello e l'algoritmo di implementazione del VBAP in due dimensioni estenderemo il concetto in tre dimensioni.

la più piccola configurazione di tre altoparlanti consiste nel disporli ai vertici di un triangolo equilatero come in figura 3.3, quindi ora basta prendere la formule usate sopra ed estenderle con la dimensione Z introducendo quindi l'angolo θ che indica l'elevazione dal piano orizzontale.

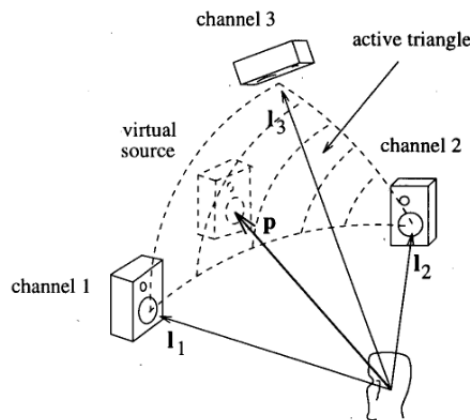


Figura 3.3: composizione vettoriale delle sorgenti reali e virtuale in tre dimensioni

E' logico anche che la sorgente virtuale potrà essere collocata soltanto nella calotta¹ delimitata dalle rette congiungenti gli altoparlanti a due a due, quindi esattamente come sopra se definiamo i vettori unitari che puntano alle tre casse come $\mathbf{l}_1 = [l_{11} \ l_{12} \ l_{13}]^T$ $\mathbf{l}_2 = [l_{21} \ l_{22} \ l_{23}]^T$ $\mathbf{l}_3 = [l_{31} \ l_{32} \ l_{33}]^T$ e il vettore della sorgente virtuale $\mathbf{p} = [p_1 \ p_2 \ p_3]^T$ risulta che:

$$\mathbf{p} = \mathbf{g} \cdot \mathbf{l} = [g_1 \ g_2 \ g_3] [\mathbf{l}_1 \ \mathbf{l}_2 \ \mathbf{l}_3]^T \quad (3.7)$$

quindi ribaltando l'equazione ci rimane che:

$$[g_1 \ g_2 \ g_3] = [p_1 \ p_2 \ p_3] \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}^{-1} \Rightarrow \mathbf{g} = \mathbf{p}^T \mathbf{L}_{123}^{-1} \quad (3.8)$$

anche in questo caso possiamo calcolare i tre coefficienti scalati in questa maniera:

$$[g_1 \ g_2 \ g_3]_{scaled} = \frac{\sqrt{C} [g_1 \ g_2 \ g_3]}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \quad (3.9)$$

in questo caso fare un'esempio non è possibile in quanto il VBAP in tre dimensioni è molto legato alla geometria di implementazione e la trigonometria sulla sfera è analiticamente pesante anche se possibile, quindi si preferisce lasciare al calcolatore il compito del calcolo vettoriale senza lasciare a noi il compito di tradurre il tutto in funzioni angolari, si veda per esempio il passaggio dalla formula 3.3 alla formula 3.4.

Ora, come sopra, possiamo arrivare ad un numero n di altoparlanti per coprire tutto l'angolo solido, anche qui il nostro algoritmo dovrà selezionare gli altoparlanti che saranno imputati alla riproduzione del nostro oggetto sonoro, gli stessi che delimiteranno la calotta attiva che racchiude la sorgente virtuale.

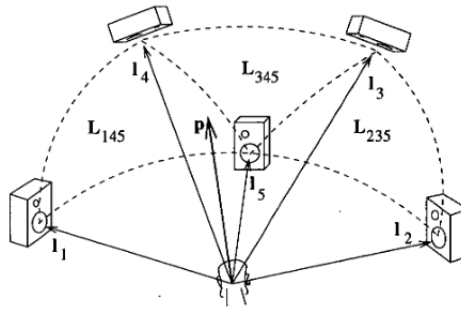


Figura 3.4: configurazione a più altoparlanti per la tecnica VBAP in tre dimensioni

¹con calotta si intende una porzione di superficie di una sfera

3.2 integrazione del formato MDA con la tecnica VBAP

Ricapitolando siamo arrivati al punto di sapere cos'è il format MDA e che cos'è la tecnica di spazializzazione VBAP, il passo successivo è riuscire ad integrare i due insieme, la cosa non è difficile ma procediamo per gradi.

ricordiamo che il format MDA contiene oggetti sonori a cui vengono associati metadati contenenti posizione nelle tre dimensioni e grandezza, essi devono essere spaccettati e adattati in modo da essere letti dall'algoritmo del VBAP.

Un concetto fondamentale da capire è che essendo le informazioni, nei metadati, relative a uno spazio tridimensionale, esse si adatteranno benissimo a una riproduzione VBAP tridimensionale con la somma delle calotte attive coprenti l'intera superficie sferica in cui è immerso l'ascoltatore, se il tipo di riproduzione non è questo allora bisognerà adattare i metadati in modo da poter essere congruenti alla configurazione che andranno a pilotare, logicamente queste trasformazioni dovranno il più possibile lasciare inalterata la percezione spaziale a quella che è stata decisa in fase di post-produzione.

andremo ora a vedere come integrare le due cose per le configurazioni del capitolo 1.1

3.2.1 utilizzo MDA in geometrie tridimensionali

Per prima cosa facciamolo per le geometrie tridimensionali infatti per queste non c'è bisogno di fare delle trasformazioni sulle coordinate in quante esse possono essere applicate direttamente, l'unica cosa che c'è da fare è un piccolo discorso sulle distanze e sui raggi.

Introduciamo due raggi, ρ_0 che sarebbe la lunghezza del vettore congiungente l'ascoltatore con uno degli speaker (non importa quale dato che tutti gli speaker sono posizionati sulla superficie di una sfera quindi si avranno tutti raggi uguali) e definiamo ρ_1 come la lunghezza del vettore congiungente l'ascoltatore e il più vicino degli oggetti virtuali che si vuole creare, premettendo che io posso solo riprodurre sorgenti virtuali a partire dalla superficie della sfera verso l'esterno, allora qualsiasi di questi che avranno $\rho_n \leq \rho_0$ non sarò in grado di riprodurli nella maniera corretta, quindi dovrò attuare una trasformazione in modo da lasciare inalterata la distanza relativa fra le sorgenti virtuali in questo modo:

$$\left\{ \begin{array}{ll} se \ \rho_1 \leq \rho_0 & \Rightarrow \ \rho'_n = \rho_n + (\rho_0 - \rho_1) \ \forall n \geq 1 \\ se \ \rho_1 \geq \rho_0 & \Rightarrow \ \rho'_n = \rho_n \ \forall n \geq 1 \end{array} \right. \quad (3.10)$$

fatto questo non ci rimane che intervenire sul fattore C introdotto nel paragrafo 3.1.1 in quanto esso contribuisce alla sensazione di lontananza, mi spiego meglio:

Principalmente sono quattro i fattori che intercorrono alla corretta ricostruzione in lontananza di un'oggetto, essi sono l'intensità sonora, la colorazione timbrica, il rapporto riverbero/segnale diretto e il delay temporale², l'informazione relativa agli ultimi tre parametri possono essere contenute direttamente nel segnale audio dell'oggetto sonoro, invece noi dobbiamo considerare l'intensità sonora in quanto legata alla distanza dall'oggetto all'ascoltatore e quindi al suo raggio³.

come detto sempre nel paragrafo 3.1.1 il fattore C è legato al gain complessivo dell'oggetto ma quest'ultimo posizionato sulla superficie della sfera, ora sappiamo dalla formula 3.10 che ogni oggetto è posizionato o sulla sfera o al suo esterno quindi possiamo prenderci la libertà di scalare il coefficiente C in base alla distanza dell'oggetto dalla sfera in questo modo:

$$\mathbf{g}_n \text{ scaled} = \frac{\sqrt{\frac{C_n}{r'_n - r_0}} \mathbf{g}_n}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \quad (3.11)$$

così in teoria con il fatto che l'intensità sonora scala come il quadrato della distanza dovremmo avere una giusto riposizionamento dell'oggetto.

Logicamente quanto detto qui sopra dovrà essere applicato in tutti casi di VBAP sia tridimensionale che non in quanto bisogna tener conto della distanza degli speaker ed è molto difficile che si abbiano tutti gli oggetti alla stessa distanza.

3.2.2 utilizzo MDA in geometrie bidimensionali

introduciamo ora un problema aggiuntivo: come faccio a riprodurre un programma sonoro tridimensionale in un'impianto bidimensionale? semplice, non si può; possiamo però introdurre un passaggio dal format MDA alla riproduzione in modo da togliere una dimensione, vediamo come.

Nei metadati del format MDA sono presenti gli angoli θ, ϕ e il raggio ρ che identificano la posizione, quello che ci da informazioni sull'altezza è l'angolo (appunto detto di

²la spiegazione del perchè di tutti questi parametri richiederebbe molto tempo quindi diamo per buone queste assunzioni

³in realtà ogni modifica del raggio si dovrebbe ripercuote anche sugli altri tre parametri cosa che nel nostro caso non avviene essendo questi ultimi tre contenuti nell'audio dell'oggetto sonoro, siamo quindi di fronte ad un'errore ma che possiamo considerare trascurabile visto l'entità ridotta delle trasformazioni sui raggi

elevazione) ϕ quindi ponendo semplicemente $\phi = 0^\circ$ sulla carta elimineremmo la dimensione verticale ma comprometteremmo anche la percezione degli oggetti originariamente posti sopra l'ascoltatore, per esempio:

Mettiamo il caso che un oggetto sia posto ad un angolo $\theta = 90^\circ$ e un'angolo $\phi = 89^\circ$, esso praticamente si trova quasi perfettamente al disopra della nostra testa, ponendo semplicemente quest'ultimo a 0 si avrebbe uno spostamento netto dell'oggetto alla completa sinistra dell'ascoltatore il che è incongruente con la scelta iniziale del produttore, una soluzione al caso potrebbe essere di creare un ascolto immersivo per quell'oggetto sfruttando tutti gli speaker del piano di ascolto e scalando il segnale da inviare a tutti questi ultimi in base alla posizione della proiettante dell'oggetto sul piano orizzontale, nel caso sopra la proiettante dell'oggetto sarebbe un punto molto vicino all'ascoltatore immediatamente alla sua sinistra, quindi i due speaker a sinistra dovrebbero riprodurre la metà più un poco di potenza sonora attribuita all'oggetto, invece gli speaker a destra dovrebbero riprodurre la metà meno un poco di potenza.

3.2.3 esempio di integrazione con sistema planare (quadrifonico e 5.1)

Per tutti gli esempi che farò vorrò riprodurre tre oggetti con seguenti coordinate sferiche per far vedere ogni caso:

oggetto 1	$\theta_1 = 15^\circ$	$\phi_1 = 0^\circ$	$\rho_1 = 1,5m$	$C_1 = 1,3$
oggetto 2	$\theta_2 = 275^\circ$	$\phi_2 = 0^\circ$	$\rho_2 = 2,5m$	$C_2 = 1,6$
oggetto 3	$\theta_3 = 180^\circ$	$\phi_3 = 45^\circ$	$\rho_3 = 3,0m$	$C_1 = 0,8$

Cominciamo con la quadrifonia, una possibile configurazione potrebbe essere questa:

speaker 1	$\theta_{0,1} = 45^\circ$	$\rho_0 = 2m$
speaker 2	$\theta_{0,2} = 45^\circ$	$\rho_0 = 2m$
speaker 3	$\theta_{0,3} = 45^\circ$	$\rho_0 = 2m$
speaker 4	$\theta_{0,4} = 45^\circ$	$\rho_0 = 2m$

quindi come prima cosa applichiamo la formula 3.10 in modo da scalare i raggi degli oggetti e rendergli

$$\rho'_1 = 2m \quad \rho'_2 = 3m \quad \rho'_3 = 3,5m$$

successivamente è facile trovare i gain e i gain scalati del primo e del secondo oggetto in quanto il primo è riprodotto dagli speaker 4 e 1 invece il secondo dagli speaker 3 e 4 in questo modo:

per entrambi gli oggetti, prima applico la formula 3.6 in modo da trovare i nuovi angoli da mettere nella formula 3.4 poi considerando i coefficienti C_n e i raggi ρ'_n ricavo dalla formula 3.11 i gain scalti per entrambi gli oggetti che in questo caso saranno:

$$\begin{aligned}
g_{(1,obj\ 1)} &= g_{(1,obj\ 1)scaled} = \\
g_{(4,obj\ 1)} &= g_{(4,obj\ 1)scaled} = \\
g_{(3,obj\ 2)} &= g_{(3,obj\ 2)scaled} = \\
g_{(4,obj\ 2)} &= g_{(4,obj\ 2)scaled} =
\end{aligned} \tag{3.12}$$

Capitolo 4

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**.

4.1 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

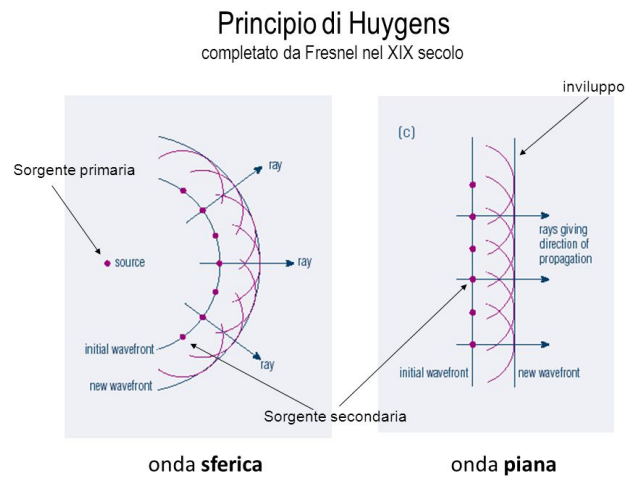


Figura 4.1: Principio di Huygens-Fresnel

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

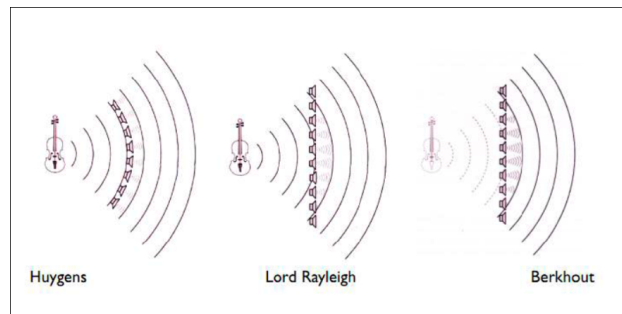


Figura 4.2: Salti concettuali della WFS

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.

conclusioni

Bibliografia

- [1] https://en.wikipedia.org/wiki/Huygens-Fresnel_principle
- [2] https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.775-3-201208-1!!PDF-E.pdf
- [3] https://it.wikipedia.org/wiki/Dolby_Digital
- [4] 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
- [7] <//digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1049&context=laessp>
- [8] https://en.wikipedia.org/wiki/Sound_localization#ITD_and_IID
- [9] <http://lib.tkk.fi/Diss/2001/isbn9512255324/article1.pdf>