

POLITECNICO DI TORINO

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



TESI DI LAUREA MAGISTRALE

L'Immagine Sonora, tra il sensoriale e il narrativo il delicato
equilibrio di immagini e suoni al cinema.

Analisi delle più recenti innovazioni tecniche e delle loro
implicazioni estetiche e produttive.

Relatori

Prof. Enrico Maria Verra
Prof. Marco Carlo Masoero

Candidata

Andrea Maria Vittoria Belotti

Dicembre 2020

*Al mio Faro nella Notte
e ai pesci rossi che danzano contro correnti di tonni..*

Indice

Introduzione	1
1. L'Immagine Sonora, equilibrio dinamico tra sensi e narrazione	9
1.1 STORIA DELLA CINEMATOGRAFIA SONORA	12
1.1.1 "You ain't heard nothing yet"	13
1.1.2 Il Cinema Verbocentrico	15
1.1.3 Verso la Stereofonia – i Registratori Magnetici	17
1.1.4 Verso la Stereofonia – L'Arrivo della Televisione	19
1.1.5 Lavalier e Registrazione Multitraccia	21
1.1.6 La Rivoluzione Silenziosa del Dolby Stereo	22
1.1.7 Apocalypse Now	25
1.1.8 L'Era Digitale	26
1.2 LA VOCAZIONE "SENSUALE" DEL CINEMA	28
1.2.1 Il Film-Concerto di Jullier e il Ritorno al Cinema Muto di Chion	28
1.2.2 Lo Spettatore Cinestesico e il "Profumo" del Suono	33
1.2.2.1 Le dita che sentivano sé stesse	35
1.2.2.2 Sfocare i confini tra pubblico e schermo	40
1.3 TRA IMMERSIONE E NARRAZIONE QUAL È IL FUTURO DELLE TECNOLOGIE IMMERSIVE?	48
1.3.1 Da Gravity a Roma, la deriva intimista dell'Atmos	52
1.3.1.1 Gravity tra spazio e spazializzazione	52
1.3.1.2 Roma e i profumi sonori che avvolgono il silenzio di Cleo	57
1.3.2 Dalla realtà virtuale di Jullier alle "protesi del presente" qual è il futuro dell'immersività?	63
2. Sistemi audio immersivi dal multicanale all'approccio object-based	72
2.1 CONFIGURAZIONI MULTICANALE TRADIZIONALI	73
2.1.1 Monofonia (C)	75
2.1.2 Stereofonia (L, R)	75
2.1.3 (L, C, R) 3.0 – (3 - 0 stereo)	77
2.1.4 Surround (L, C, R, S) - (3 - 1 stereo)	78
2.1.5 Surround 5.1 (L, C, R, Ls, Rs, LFE) – (3 - 2 stereo)	78
2.1.6 Surround 7.1 (L, C, R, Lss, Lrs, Rss, Rrs, LFE)	79
2.2 NOZIONI DI BASE DI UN SISTEMA AUDIO IMMERSIVO	80
2.2.1 Configurazione di una sala Immersiva – Height/Top Speakers	83
2.2.1.1 AURO 3D	87
2.2.1.2 NHK 22.2	91
2.2.1.3 DOLBY ATMOS	93
2.2.2 Gli elementi Bed	97
2.2.3 Gli Oggetti	98
2.3 SMPTE 2098-1/2098-2	100
2.3.1 Comitato Tecnico 25 CSS Immersive Audio Suite	102
2.3.1.1 La Suite di Documenti	106
2.4 PRINCIPI BASE DI UN SISTEMA SONORO OBJECT-BASED	108

2.4.1	<i>Rappresentazione nello Spazio e Sistema di Coordinate di un Oggetto</i>	109
2.4.2	<i>Renderer</i>	115
2.4.2.1	Directional Pairwise Panning (DIR)	119
2.4.2.2	'Dual Balance' Panning (DB)	122
2.4.2.3	Distance-Based Panning (DIST)	123
2.4.3	<i>Caratteristiche degli Immersive Audio Metadata ST 2098-1</i>	123
2.4.3.1	Bed Metadata ST 2098-1	124
2.4.3.2	<i>Object Metadata ST 2098-1</i>	126
2.5	FORMATO BITSTREAM 2098 - 2	134
2.5.2	<i>Immersive Audio (IA) Frame</i>	137
2.5.3	<i>Struttura logica di un elemento Bed</i>	141
2.5.3.1	Bed Remap	143
2.5.4	<i>Struttura Logica di un Oggetto</i>	145
2.5.5	<i>Esempi Applicativi</i>	147
2.5.5.1	Esempio di una struttura tipica di oggetti e bed attivi simultaneamente	149
2.5.5.2	Esempio di una struttura contenente un <i>BedRemap</i>	150
2.5.5.3	Esempio di una struttura contenente un <i>ConditionalBed</i>	152
2.6	GLI ALTRI FORMATI	153
2.6.1	<i>Home Theatre</i>	154
2.6.2	<i>Il Binaurale</i>	157
2.6.2.1	Registrazioni Binaurali	159
2.6.2.2	Sintesi Binaurale	161
3. L'Immagine Sonora dalla sceneggiatura alla sua realizzazione missata in atmos per un ascolto binaurale, analisi e sperimentazione del workflow e delle procedure applicate ad un vero progetto audiovisivo _____		165
3.1	IL PROGETTO	166
3.1.1	<i>Soggetto Della Narrazione</i>	168
3.2	LA REALIZZAZIONE	172
3.2.1	<i>La Scena Sonora</i>	172
3.2.2	<i>Il Suono</i>	173
3.2.3	<i>La Location</i>	175
3.2.4	<i>I personaggi</i>	177
3.2.5	<i>La fotografia</i>	178
3.2.6	<i>L'immagine finale</i>	180
3.3	WORKFLOW IN ATMOS	182
3.3.2	<i>Strumenti di postproduzione</i>	189
3.3.3	<i>Confronto tra workflow multicanale e Atmos</i>	193
3.3.3.1	WORKFLOW MULTICANALE	194
3.3.3.2	WORKFLOW ATMOS	198
3.3.4	<i>Editing e mix audio del progetto</i>	202
3.4	CONSIDERAZIONI, VANTAGGI E SVANTAGGI DI UN WORKFLOW IMMERSIVO	209
Conclusioni _____		216
Bibliografia e Sitografia _____		222
Ringraziamenti _____		226

Indice delle figure

Figura 1 You ain't heard nothing yet.....	13
Figura 2 Registratore Nagra	18
Figura 3 L'incredibile successo di pubblico all'uscita di Star Wars in sala.....	23
Figura 4 Walter Murch durante la lavorazione di Apocalypse Now.	25
Figura 5 Copertina del libro di Vivian Sobchack	35
Figura 6 Immagini di Ben Burtt tratte dal documentario Making Waves	42
Figura 7 La scena tratta dal film Lisbon Story (1995) di Wim Wenders [10]	46
Figura 8 Immagine d'apertura del film Gravity (2013) di Alfonso Cuarón	53
Figura 9. Scena tratta dal film Roma (2018) di Alfonso Cuarón	62
Figura 10 Grafico dell'utilizzo di Netflix via smartphone. Fonte Ofcom	68
Figura 11 Grafici crescita piattaforme streaming nel mondo. Fonte Statista	69
Figura 12 Sezione laterale di una sala cinematografica in Dolby Atmos [27]	74
Figura 13 Soundfiled Configuration di un Sistema Surround 5.1 [22]	78
Figura 14 Soundfiled Configuration di un Sistema Surround 7.1 [22]	79
Figura 15 Tabella riassuntiva dei Canali Audio Immersivi [22]	84
Figura 16 Tabella raffigurante alcune configurazioni Audio Immersive [22].....	85
Figura 17 Modello 3D di una configurazione di sala 11.1 Auro 3D [24].....	87
Figura 18 Configurazione Home Theatre sistema Auro 3D [24].....	88
Figura 19 Disposizione degli altoparlanti di un sistema AutoMax 26.1. [24].....	90
Figura 20 Configurazione di un sistema NHK 22.2 [25].....	91
Figura 21 Elenco canali del sistema NHK 22.2 suddivisi in "U+ M + B" [25]	92
Figura 22 Configurazione degli altoparlanti di un sistema Dolby Atmos. [27]	93
Figura 23 Sezione di una sala Atmos suddivisa per Layer di altoparlanti. [22]	94
Figura 24 sistema object-based (OBA) puro [24]	100
Figura 25 Diagramma a Blocchi workflow sistemi immersivi prima. [22].....	103
Figura 26 Diagramma a Blocchi workflow sistema immersivo dopo [22].....	105
Figura 27 Sistema di Coordinate [22].....	109
Figura 28 Rappresentazione ideale di un cinema come cubo unitario. [28]	110
Figura 29 Posizione nello spazio in un sistema allocentrico. [28]	113
Figura 30 Accresciuta risoluzione spaziale di un sistema object-based [27].....	115
Figura 31 Illustrazione di diversi algoritmi di panning [30]	118
Figura 32 Algoritmo di Panning VBAP [31]	120
Figura 33 Elemento bitstream ObjectZoneDefinition19 [28].....	132
Figura 34 . Gli altoparlanti inclusi nelle 19 zone di sala [28]	133
Figura 35 Struttura di un Immersive Audio Bitstream [33].....	135
Figura 36 Vista logica di un IA Frame c. [28]	137
Figura 37 Grafico UML di un Immersive Audio Frame [22]	139
Figura 38 Vista logica di un elemento Bed [28].....	141
Figura 39 Vista logica di un sotto-elemento BedRemap [28].	143
Figura 40 Vista logica di un Oggetto [28]	145

Figura 41 IA Frame contenente Bed e Oggetto attivi contemporaneamente [28]	149
Figura 42 Vista logica di un IA Frame contenente un BedRemap. [28]	150
Figura 43 Vista logica di un IA Frame contenente due ConditionalBed. [28]	152
Figura 44 Configurazione home theatre Auro 3D. [34]	154
Figura 45 Configurazione Home Theatre Dolby Atmos 7.1.4. [35]	155
Figura 46 Altoparlanti che abilitano verticalità mediante ceiling reflections [35]	156
Figura 47 Dummy head [36]	159
Figura 48 Omini Binaural Microphone [33]	161
Figura 49 Processo di convoluzione per sintesi binaurale sorgenti statiche. [33]	163
Figura 50 Immagine tratta dal teaser prospettiva del Pala Olimpico di Torino.	175
Figura 51 Immagine tratta dalle riprese effettuate al Pala Olimpico di Torino..	177
Figura 52 Immagine tratta dal set presso il Pala Olimpico di Torino.	179
Figura 53 Immagine tratta dal teaser del documentario oggetto di tesi.	181
Figura 54 Immagine tratta dal teaser del documentario oggetto di tesi.	181
Figura 55 Immagine tratta dal teaser del documentario oggetto di tesi.	182
Figura 56 Immagine tratta dal documentario “Making Waves” [4].	184
Figura 57 Immagini tratte dal set presso il Pala Olimpico di Torino.	188
Figura 58 Consols di missaggio fisiche [38].	190
Figura 59 Dolby Atmos Theatrical RMU (Rendering and Mastering Unit).	191
Figura 60 Dolby Atmos Rendering Application Version 3. [40]	192
Figura 61 Sessione ProTools in 5.1.	195
Figura 62 Panning in un sistema multicanale.	196
Figura 63 Registrazione del master	197
Figura 64 Master 5.1 master stereo a confronto	197
Figura 65 Sessione ProTools in Atmos.	199
Figura 66 Interfaccia software dell’unità di rendering Dolby Atmos	199
Figura 67 Plugin di pan di uno spostamento verticale e spread di un oggetto.	200
Figura 68 Workflow di mix in Dolby Atmos. [38]	201
Figura 69 Immagine tripartita della fase di mix.	205
Figura 70 Interfaccia di panning in fase di mix Atmos.	206
Figura 71 Percezione parametri di Immersione, Dinamica e Localizzazione [41]	212
Figura 72 Grafico finale che mostra i gradi di preferenza [41]	213

Introduzione

Il primo titolo assunto da questa tesi nelle sue versioni ancora embrionali, dai connotati certamente meno tecnici ma più figurativi, era:

L'Immagine Sonora. Il controverso e affascinante rapporto tra Immagine e Suono – reagenti verso il perfetto equilibrio chimico, il film.

Si partiva infatti dalla riflessione, più avanti nel corso degli studi scoprendo essere sostenuta anche da Massimo Mariani¹, che immagini e suoni sono come reagenti che a contatto danno vita ad una reazione chimica nella quale gli elementi primari disperdono le loro peculiarità originarie in virtù della formazione di una nuova sostanza, o in altri termini, di un nuovo mondo, diverso, straniero e distante dai due mondi primari, la cui esplorazione è stata chiamata, appunto, Cinema.

D'altronde l'interpretazione stessa della parola “immagine sonora” presenta due accezioni, quella tecnica, ovvero di come il suono mediante tecnologie di riproduzione viene rappresentato nello spazio, e quella estetica, ovvero di come il suono coniugato con le immagini sia in grado di dare vita a nuove dimensioni dell'esperienza cinematografica.

Dal cinema muto al sonoro, dalla stereofonia al surround, fino alla nuova frontiera digitale dei sistemi sonori immersivi, lo sviluppo tecnologico cinematografico continua ad andare nella direzione che Jullier aveva indicato come il capovolgimento della gerarchia immagine-suono ² a favore di ciò che Giulio Latini, Docente di Comunicazione Multimediale presso l'Università

¹ Nota [10]

² Nota [5]

degli Studi di Tor Vergata, definisce “un’ideologia aprospettica di sala”³ in cui la centralità storica dello schermo perde il suo protagonismo.

L’immagine sonora trova così nuova forma nello spazio, ove però la dimensione uditiva corre più in fretta di quella visiva. Sembra che il futuro, prossimo o anteriore, dell’immagine bidimensionale sia la Virtual Reality, quando e come però è ancora ignoto, dato il nostro radicato legame cognitivo con l’immagine bidimensionale. Ma la rappresentazione nello spazio dell’audio pare vada diffondendosi molto più speditamente rispetto a quella dell’immagine, in una sorta di legge del contrappasso laddove agli albori del cinema il sonoro nacque molto dopo rispetto all’immagine.

Con l’introduzione sul mercato dei relativamente nuovi sistemi di riproduzione tridimensionale, che proseguono l’eredità tecnologica dei *film-concerto* di Jullier, l’immagine sonora, intesa come rappresentazione sonora spaziale, raggiunge nuovi livelli di precisione ed immersività nella scala del *bagno di sensazioni*.

Se l’ambiente di ascolto *reale* infatti è lo spazio acustico che occupiamo durante le vite di tutti i giorni, in cui i suoni arrivano da tutte le direzioni possibili avvolgendoci in quella che è la realtà da noi percepita, l’obiettivo di un sistema audio immersivo è quello di restituire uno spazio acustico in cui riproporre questa percezione di totale immersione durante l’esperienza di ascolto. Si tratta dunque di una modalità di riproduzione sonora capace di ricreare le proprietà spaziali della scena acustica considerata (posizione e orientamento delle sorgenti, forma e caratteristiche dell’ambiente, introduzione della verticalizzazione sonora).

Per fare questo i principali sistemi di riproduzione sonora

³ Nota [11]

tridimensionale non solo aggiungono la terza dimensione mancante ai vecchi sistemi multicanale, ovvero l'altezza, ma si avvalgono anche di un approccio ibrido, introdotto sul mercato dall'Atmos della Dolby, tra un sistema basato sui classici canali (*Channel-Based-Approach*), i cosiddetti Bed, ed uno basato su *Oggetti (Object-Based-Approach)*.

Un simile sistema, oltre ad introdurre la verticalizzazione menzionata, *The Voice of God*, adotta singoli elementi audio (*oggetti*), accompagnati da set di metadati, in grado di fornire le informazioni necessarie per renderizzare la posizione di un suono all'altezza di precisi diffusori all'interno di una *Soundfield Configuration* e di descriverne cambiamenti e caratteristiche che variano rapidamente anche nel corso dello stesso frame, fornendo così un ambiente acustico in cui far muovere i suoni quasi illimitato.

Tale approccio vede nell'unità di rendering il suo maggior punto di forza, quella che Brian Vessa, *Founding Chair* del Comitato Tecnico 25 Cinema Sound System, ha definito una *magic box*, poiché attraverso di essa il mix sonoro viene letto ed interpretato a seconda della configurazione cui quel renderer è associato, mentre lo standard di workflow attualmente è quello di realizzare un mix per ogni sistema di riproduzione di riferimento, pena un drastico taglio di quelle sonorità deputate a canali non presenti nella configurazione finale. I set di metadati che accompagnano invece gli oggetti consentono, sulla base di specifiche e procedimenti che si analizzeranno in seguito, di interpretare la posizione dei suoni in maniera relativa secondo un sistema di coordinate allocentrico.

In questo modo un fonico di mix realizzando un unico master ottiene un lavoro che verrà automaticamente ottimizzato a seconda del playback environment dell'utente finale. Lavorando dunque alla massima qualità si ha la garanzia che il mix risultante sia equamente bilanciato tanto nell'ascolto in cuffia quanto in un sistema 7.1. Un fattore questo che oltre a produrre

enormi vantaggi in termini di risparmio di tempi e lavoro tiene conto anche di quella che è la tendenza odierna al consumo di contenuti audiovisivi secondo cui la sala cinematografica non è che una delle tappe del viaggio distributivo di un film.

Tali sistemi dunque sembrano aver capitalizzato l'esperienza storica dell'evoluzione tecnologica del sonoro cinematografico rispondendo alle problematiche di flessibilità, retrocompatibilità e preservazione dell'intento artistico nel passaggio da sala mix a sala di riproduzione

Tuttavia come fa notare Giulio Latini in una conferenza presso l'Associazione Italiana Tecnici del Suono, che si approfondirà nel corso del primo capitolo, *non si arriva a una trasformazione tecnologica soltanto perché lo vogliono, lo pretendono i detentori di quella tecnologia*, una tecnologia si innesta all'interno di una più ampia *istanza spettatoriale*, senza dimenticare che si parla di tecnologie che si fanno veicolo di una forma d'arte o se non altro di una forma d'espressione umana, per cui il semplice apporto di natura tecnologico-procedurale non è sufficiente per designare l'efficacia della sua diffusione.

In passato ad esempio i sistemi di riproduzione stereofonica avevano visto la luce già nel 1940 con il FANTASOUND di *Fantasia*, ma la vera consacrazione dello stereo quale frontiera del sonoro cinematografico avvenne solo grazie a *Star Wars* nel 1977. Questo come si vedrà può essere interpretato alla luce del fatto che tali tecnologie insieme all'apporto tecnico devono saper interpretare un momento storico e al contempo incontrare creativi illuminati che sappiano far esplodere il potenziale della tecnologia stessa.

Ecco perché l'obiettivo principale di questa tesi è dunque quella di illustrare la struttura, natura e funzionamento delle tecnologie immersive non solo da un punto di vista tecnico ma anche attraverso l'osservazione di

uno scenario estetico-culturale più ampio, con la finalità una volta comprese le sue peculiarità, di interrogarsi e prospettare come esse possano inserirsi nel dialogo con l'identità narrativa di un film, al fine di porsi non quale mero amplificatore sensoriale ma come reale strumento grammatico-emozionale.

Sulla base di tali riflessioni i capitoli di questa tesi saranno divisi come segue:

Capitolo 1 – Approfondimento storico-culturale e dell'estetica cinematografica all'interno di cui si inseriscono le tecnologie immersive.

Nel primo capitolo come sopra menzionato si illustrerà l'evoluzione storica del sonoro filmico sino ad approdare all'eredità dei sistemi multicanale, ponendo a confronto una delle posizioni più note nella storia della critica cinematografica, ovvero il concetto di film-concerto di Laurent Jullier con alcune riflessioni contemporanee in merito al concetto di vocazione sensoriale del cinema.

In fine facendo tesoro delle riflessioni sulla sensorialità dell'esperienza filmica in relazione al concetto di “bolle immersive” e partendo da un'analisi del sonoro filmico delle ultime due opere del regista Alfonso Cuarón, si analizzerà quella che all'interno della tesi è stata chiamata la “deriva intimista dell'Atmos” che come si vedrà non lo è solo da un punto di vista narrativo ma anche, grazie alla codifica binaurale e alle nuove piattaforme streaming, una reale nonché probabile prospettiva del cinema di domani.

Capitolo 2 – Approfondimento tecnico-strutturale delle tecnologie immersive, dalla struttura di sala all'Immersive Audio Bitstream (IAB) secondo lo standard SMPTE 2098

Nel secondo capitolo si illustreranno le peculiarità di un Sistema Sonoro

Immersivo percorrendo i principali documenti che costituiscono lo standard SMPTE 2098 in materia di *Immersive Audio Metadata and Bitstream Specification* quale unico e nuovo formato di distribuzione per tutti i sistemi Sonori Immersivi in circolazione.

L'intento infatti sarà quello in prima istanza di comprendere quali sono le implementazioni tecnologiche che caratterizzano un Sistema Immerisvo rispetto ai tradizionali sistemi di diffusione surround e in seconda istanza approfondirne il formato bitstream, preziosa chiave di volta che il digitale introduce nel workflow cinematografico, per la comprensione del potenziale racchiuso nei nuovi Sistemi Immersivi oltre che di innovatività per le esigenze del mercato audiovisivo.

Al di là infatti di comprendere quali sono i passaggi in essere per favorire la diffusione di tale tecnologia e, chissà, magari un giorno vederla affermarsi quale standard, l'obiettivo di approfondire la parte di questi documenti dedicata al nuovo formato IA Bitstream risiede anche nella possibilità di osservare come la discretizzazione operata dal digitale consenta effettivamente di ampliare le prospettive tanto tecniche quanto creative di un prodotto audiovisivo.

Capitolo 3 – Approfondimento di natura applicativa delle prospettive narrative messe a disposizione dagli strumenti di un sistema immersivo mediante un reale progetto audiovisivo, ponendo a confronto workflow multicanale classico ed immersivo.

Dai sistemi multicanale a quelli immersivi come cambia il workflow operativo di un professionista del sonoro? Come, per citare nuovamente il Prof. Giulio Latini, è possibile che l'Atmos o più in generale i sistemi immersivi siano strumento di *amplificazione e riconversione di alcune istanze ligustiche del film e non solo di restituzione denotativa?*

Nel tentativo di riflettere su tali quesiti ed affinare il mindset necessario per affrontare un workflow immersivo, nel terzo capitolo quale parte conclusiva della tesi viene proposto un approfondimento sulla dialettica audio-visiva del teaser di un documentario espressamente realizzato nel periodo tra Dicembre 2019 e Settembre 2020, come parte integrante del progetto di ricerca, sul quale è stato condotto un mix sonoro in Dolby Atmos finalizzato all'ascolto in binaurale.

E' stato dunque condotto da colei che scrive un progetto documentaristico sulla figura di Gianlorenzo Blengini, il Commissario Tecnico della Nazionale Maschile di Pallavolo Italiana, uscente argento olimpico a Rio 2016. Il progetto nasceva per raccontare il suo viaggio verso le Olimpiadi di Tokyo 2020, un viaggio nel frattempo interrotto e stravolto dagli accadimenti a livello mondiale della pandemia da Covid-19.

Dalla pre-produzione, alla fase di recording fino alla post-produzione e mix sono state dunque approfondite le distinzioni tra workflow immersivo e multicanale ed implementate sul teaser oggetto di tesi grazie alla preziosa guida tanto teorico-culturale quanto applicativa, ricevuta nel corso dell'intero progetto, di Mirko Perri, pluripremiato sound designer italiano ai David di Donatello unico ad aver realizzato il primo film Atmos rilasciato in Italia, *Youth* di Paolo Sorrentino, coinvolto nel progetto di tesi grazie al Prof. Enrico Verra.

Pertanto con il materiale raccolto fino alla primavera-estate 2020 è stato strutturato un teaser promozionale del lavoro per il quale è stata scritta e pensata una "scena sonora" che raccontasse di quell'Olimpiade *che doveva essere e non è stata*, emblema dello stravolgimento mondiale di cui tutti noi siamo protagonisti.

Grazie alla possibilità di confrontarsi con un professionista di simile esperienza quale Mirko Perri è stato dunque cercato di progettare una scena

che potesse rispondere alle necessità d'ascolto di un sistema sonoro tridimensionale, curandone la rilevanza all'interno dell'impianto narrativo del progetto, ed infine attraverso il suo studio di postproduzione a Roma realizzare il mix sonoro in Atmos oggetto di tesi.

Immagini Sonore, frutto dell'incontro tra dimensioni pertinenti a diverse sfere sensoriali, che evocano un sogno, quello cinematografico, che sfuma e si dilata, ove la tecnologia diventa imprescindibile strumento ed opportunità per ampliare i confini della narrazione in un momento storico in cui i confini sono più che mai definiti.

Questa è dunque la struttura di contestualizzazione che è stata scelta per l'approfondimento di uno strumento che prima ancora che tecnologico deve essere espressione del medium cinematografico, all'interno della cornice di pensiero secondo cui la tecnica ridisegna i confini dell'arte e l'arte espande i confini della tecnica, in questo caso, *alla ricerca del controverso ed affascinante rapporto tra immagine e suono*.

1 L'Immagine Sonora, Equilibrio Dinamico tra Sensi e Narrazione

Nel corso di questo primo capitolo si faranno delle riflessioni ed approfondimenti in merito al conteso socio-estetico-culturale all'interno del quale le tecnologie sonore immersive si sono recentemente sviluppate.

L'obiettivo della tesi infatti è indagare il funzionamento di tali sistemi di riproduzione, in rapporto anche alla loro controparte visiva, per comprenderne le peculiarità tanto tecnologico-applicative quanto estetico-culturali, cui sono dedicati appunto i seguenti paragrafi.

Si partirà dunque da un'analisi di quella che è stata l'evoluzione del sonoro nella storia del cinema, suoni ed immagini hanno infatti avuto storicamente un rapporto controverso che tra logiche di mercato, esigenze tecnologiche e predisposizioni culturali, ha mutato nel tempo la grammatica narrativa e l'estetica cinematografica.

Sulla base dell'eredità, di cui ancora oggi raccogliamo i frutti, dei sistemi multicanale e degli illuminati pionieri del sonoro degli anni '70 della cinematografia americana, si porrà a confronto una delle posizioni più note nella storia della critica cinematografica, ovvero il concetto di film-concerto di Laurent Jullier con alcune riflessioni contemporanee in merito al concetto di vocazione sensoriale del cinema, nel tentativo di comprendere ove si collochi il precario baricentro dell'equilibrio dinamico tra sensi e narrazione.

In questo senso verranno in aiuto alcune teorie tratte dall'approccio fenomenologico di Vivian Sobchack, la quale concepisce il cinema come *conoscenza carnale* e lo spettatore quale *soggetto cinestesico* in grado di comprendere il film innanzitutto come esperienza sensoriale e cross-modale. Da tali riflessioni in merito alla vocazione "sensuale" del cinema in generale,

con l'ausilio di un documento di ricerca di Lucy Fife Donaldson, che interpreta il lavoro stesso dei professionisti del suono in chiave sinestesica, i quali per giungere alla corretta interpretazione sonora di un film affrontano processi che rivelano la natura chiaramente connessa con sensi ed emozioni di questa professione, si tratteggeranno mediante esempi reali le varie fasi di lavorazione che conducono al concepimento della dimensione sonora di un film.

In fine facendo tesoro delle riflessioni sulla sensorialità del cinema, dominio di cui si avvalgono le tecnologie immersive, per meglio contestualizzare le prospettive attuali di tale strumento si riproporranno alcuni utili quesiti del Prof Giulio Latini⁴, che nel corso di un seminario sul primo film Atmos rilasciato in Italia, parafrasandolo, si domanda, quali sono gli strumenti che l'Atmos mette realmente a disposizione per superare il concetto di bolla immersiva e porsi come strumento in grado di sostenere il cinema quale promotore di “un pensiero aumentato ancorché di una sensazione aumentata? E ancora, qual è l'istanza spettatoriale” più ampia cui appartengono le tecnologie immersive?

Nel tentativo di fornire alcuni ipotetici scenari in merito, attraverso un approccio induttivo per così dire, si proporrà un'analisi del sonoro filmico delle ultime due opere del regista Alfonso Cuarón, che al momento è il solo⁵

⁴ Come anticipato in introduzione, docente di Comunicazione Multimediale presso l'Università degli Studi di Tor Vergata. Il seminario cui si fa riferimento sopra verrà approfondito a seguire nel paragrafo 1.3.

⁵ Il fatto che ci si spinga a definire che il regista sia *il solo* ad aver fatto esperienza degna di nota dell'Atmos nasce non solo dal parere di alcuni professionisti del sonoro cinematografico come più avanti si introdurrà, ma anche da un'affermazione tratta da un'intervista nel 2019 di Glenn Kiser, direttore del Dolby Institute che durante lo *speech* dedicato al *making of* sonoro di *Roma*, ultimo film del regista, afferma: “Alfonoso, oh my gosh! So before I joined Dolby, the creative Dolby Institute, I ran *Skywalker Sound* for 11 years, so I heard a few tracks over the course of the years, but I'm really not exaggerating

ad aver restituito l'utilizzo cinematografico più virtuoso dell'Atmos, per poi concludere con alcune riflessioni relativamente alle effettive prospettive future che le tecnologie immersive hanno coerentemente alla società odierna.

L'interdisciplinarietà di cui si caratterizza questo capitolo, al fine di proporre uno scenario possibile alla contestualizzazione estetico-culturale delle recenti tecnologie immersive, prende spunto dalla definizione di *infrastrutturalismo* data da Adriano d'Aloia e Ruggero Eugeni in merito al moderno status della teoria cinematografica secondo la quale, dovendo fare i conti con un'inesorabile "dispersione" e "frammentazione" (specchio del periodo attraversato dal cinema stesso), ha dovuto rivisitare i suoi confini evolvendo in: "teoria dunque non come prodotto, ma piuttosto come condizione attiva, come habitat persistente e come rete di connessioni non sempre immediatamente visibili, che rende comunque possibile e fruttuosi la riflessione, il discorso, il confronto con altre discipline." [1]

Secondo dunque gli autori tale confronto negli ultimi anni è stato condotto nella direzione della filosofia, delle scienze sperimentali e della mediologia, "Per ciascuna di queste aree un concetto si è imposto quale baricentro della discussione: rispettivamente l'esperienza, l'organismo, il dispositivo."

Dalle teorie fenomenologiche della Sobchack, ad alcuni riferimenti in merito alla cross-modalità dei sensi cara alle neuroscienze, fino al confronto tra la realtà virtuale di Jullier e quella dell'infosfera di Luciano Floridi questo è l'approccio dunque che ci cercherà di offrire nei seguenti paragrafi con la finalità di introdurre al meglio lo studio delle tecnologie sonore immersive.

when I say that the work that you and the team did on Roma is the more complex, rich and detailed sound mix I've ever heard. So congratulations to everyone on the team!"

1.1 Storia della Cinematografia Sonora

“Oggi la tecnologia di elaborazione e di riproduzione del suono è capace di realizzare cose straordinarie. Si riesce a lavorare su livelli di dettaglio molto sofisticati. A voler semplificare, si potrebbe immaginare un contenitore in cui poter inserire le informazioni sonore di una scena: più è capiente il contenitore, più si potranno riprodurre in modo limpido e differenziato molti eventi sonori simultanei. Se oggi il contenitore è molto capiente, nel passato non lo era affatto; questo significa che i pionieri del cinema sonoro si trovavano costantemente di fronte a scelte restrittive, costretti a scegliere se privilegiare l’ascolto dei dialoghi, degli effetti sonori, delle ambientazioni o della musica per non ‘stropicciare’ il risultato finale. Bisognava dunque scegliere a quali degli elementi citati sopra affidare il compito di sostenere la narrazione.” [2]

In virtù della panoramica storica che seguirà risulta interessante introdurre, secondo quanto appena suggerito dalle parole di Mirko Perri, il concetto di sonoro cinematografico come una scatola in costante espansione, più la scatola si ingrandisce più i rapporti narrativi tra suoni ed immagine si modificano. Tale rapporto infatti non è esattamente espressione di una funzione lineare, come agli inizi è stato per il binomio formati panoramici e sistemi stereofonici (immagine più grande → suoni più sensazionali).

Più la scatola si allarga più la grammatica sonora muta quella visiva, come se della primaria reazione chimica, di cui si parlava in introduzione, si necessitasse rivedere i reciproci rapporti ponderali dei reagenti in gioco.

Qui di seguito si riporterà una breve storia, principalmente da un punto di vista tecnologico, dell’evoluzione del sonoro filmico, con l’obiettivo di focalizzare come gradualmente i rapporti immagini-suoni siano stati modificati in termini narrativi fino al momento cruciale che ancora oggi

rimane la chiave di volta all'interno degli equilibri estetico-narrativi della cinematografia contemporanea, l'arrivo del Dolby Stereo e dei formati surround.

Momento storico-culturale e tecnologico che definitivamente consolida “il travagliato passaggio da un cinema “verbocentrico” “vococentrico” ad un cinema pienamente audiovisivo nel quale si sviluppa progressivamente un'autonoma ricerca estetica sul suono”. [3]

1.1.1 “You ain't heard nothing yet”

Il sonoro sincronizzato con le immagini fa per la prima volta capolino nel cinema con il film *Don Juan* (1926), grazie ad un'invenzione dei fratelli Warner che trovarono il modo di sincronizzare meccanicamente un grammofono con un proiettore, il *Vitaphone*. Tuttavia l'ingresso della parola sincronizzata con le immagini verrà ascoltata per la prima volta in un film nel 1927 con *The jazz singer*, laddove la frase iniziale dell'attore Al Jolson “(..) You ain't heard nothing yet” non poteva essere per il cinema sonoro battesimo più profetico.



Figura 1. L'attore Al Jolson in *The jazz singer* di Alan Crosland nella scena iniziale in cui pronuncia la storica frase *You ain't heard nothing yet*, film che ha segnato la nascita del cinema sonoro.

Consolidata la reazione entusiastica del pubblico di fronte all'esperienza di poter ascoltare le voci degli attori durante la visione di una pellicola, il film sonoro si afferma ben presto quale standard del cinema. Proseguono pertanto sviluppi tecnologici in questo senso per ovviare ai limiti strutturali che la sincronizzazione meccanica del grammofono causava e in virtù di tali sforzi alla fine degli anni Venti si affermano i primi formati ottici più stabili ed affidabili in merito al precario sistema precedente.

Nei formati ottici infatti “i suoni dei microfoni venivano convertiti in oscillazioni elettriche a loro volta trasformati in flussi luminosi e impressi su pellicola attraverso un procedimento fotografico. In riproduzione invece la banda ottica veniva “letta” da una cellula fotoelettrica che convertiva le modulazioni luminose in impulsi elettrici a loro volta inviati ai sistemi di amplificazione.” [3]

Tuttavia l'introduzione della presa diretta sul set, per quanto da un lato donasse ai film una nuova magia per il potere comunicativo trasferito dalle voci degli attori, dall'altro generò notevoli sconvolgimenti nel workflow fino a quel tempo conosciuto per la realizzazione di un film. In primo luogo infatti, come fa notare Di Donato [3], le cineprese producevano un rumore che interferiva in maniera drastica sul sonoro, al punto che fu necessario relegare il loro posizionamento in gabbioni. Inoltre l'impossibilità di intervenire in postproduzione sul suono richiedeva che le scene venissero realizzate in continuità con più camere che riprendessero in contemporanea.

Questi due fattori influirono notevolmente sulla grammatica visiva la quale fu caratterizzata da posizioni di camera molto standard, onde evitare che le cineprese fossero visibili nel reciproco campo visivo, da focali strette, che riducevano notevolmente la profondità di campo, e privata praticamente di tutti i movimenti di macchina, salvo qualche morbido pan su cavalletto.

Inoltre essendo i microfoni dell'epoca omnidirezionali c'era la necessità di posizionarli il più possibile vicino agli attori imponendo ulteriore fissità sia dei loro movimenti che delle posizioni di camera. In fine tutti i teatri di posa dovettero essere ricostruiti con materiali fonoassorbenti, così come evitate le scene in esterno per rendere il più possibile insonorizzato il set.

Tutto questo pertanto comportò una certa regressione della dialettica cinematografica dei primi film sonori conquistata fino a quel periodo nel corso dello sviluppo del cinema muto.

1.1.2 Il Cinema Verbocentrico

Nel corso degli anni Trenta parte dei problemi sopra riportati vennero risolti e le tecniche di registrazione e riproduzione sonora trovarono piena standardizzazione, infatti figure come il sound engineer divennero parte integrante del set. E' in questi anni che si consolida l'approccio al suono *Verbocentrico*⁶, secondo il quale l'intelligibilità dei dialoghi era di primaria importanza ed i suoni (musica e pochi rumori) erano esclusivamente a

⁶ Questo termine è stato coniato da Michel Chion nel suo famoso libro "L'audiovisione suono e immagine nel cinema" volendo indicare attraverso di esso quel periodo della cinematografia in cui il suono cinematografico privilegia in primo luogo la voce, non a caso durante la presa diretta è la voce che viene favorita nelle registrazioni del suono ed è sempre la voce che viene isolata nel mix sonoro come uno "strumento solista", per la quale gli altri suoni, musica e rumori, sono semplicemente d'accompagnamento. "Il suono nei film è *voco-* e *verbocentrico*, soprattutto, perché lo sono anche gli esseri umani nel loro comportamento abituale. Quando in un dato ambiente sonoro senti delle voci, quelle voci catturano e focalizzano la tua attenzione prima di qualsiasi altro suono (vento che soffia, musica, traffico). Solo dopo, se sai molto bene chi sta parlando e di cosa sta parlando, potresti rivolgere la tua attenzione dalle voci al resto dei suoni che senti. Quindi se queste voci parlano in un linguaggio accessibile, cercherai prima il significato delle parole, passando ad interpretare gli altri suoni solo quando il tuo interesse per il significato sarà stato soddisfatto." [3]

servizio del contenuto visivo e dialogico, senza avere una propria funzione narrativa. Uno dei motivi fondamentali era dovuto al fatto che il sistema ottico adoperato per fissare il suono consentiva la riproduzione sonora entro la soglia dei 10 000 Hz, contenendo ampiamente la banda in frequenza delle voci, che non supera i 2000 Hz, ma tagliando via un'elevata porzione di frequenze percepibili (la soglia di udibilità umana infatti è tra i 20 Hz e i 20 000 Hz).

Nonostante questo fu lo standard adottato nella cinematografia per circa 30 anni, già nel corso di questo periodo i filmmakers iniziarono a realizzare la necessità dell'inserimento degli effetti sonori, poiché non tutti i suoni sono registrabili grazie alla presa diretta. Iniziò così lentamente ad evolvere la figura del *sound editor* il quale aggiungeva effetti quali galoppate, pugni, tuoni, in postproduzione, pur continuando a concepire l'intellegibilità dei dialoghi quale obiettivo nonché motore principale dell'intera narrazione.

Nel documentario *Making Waves: The Art of Cinematic Sound* [4] si cita che il primissimo sound designer, secondo i canoni con i quali si contraddistinguerà questa figura tra i futuri anni '60 e '70, è Murray Spivack che nel 1933 con *King Kong* fece un incredibile lavoro di registrazione e manipolazione dei suoni (rallentandoli o invertendone la velocità) per animare realisticamente i personaggi di King Kong e il dinosauro.

Tuttavia questo fu un esperimento assolutamente pionieristico che prima di tramutarsi in standard dovette attendere svariati anni. Per molto tempo gli studios adatteranno stock di effetti in cui i rumori delle cavalcate, dei colpi di pistola e così via facevano parte di un campionario. Tale pratica infatti costituiva un costo enormemente inferiore, ma al contempo standardizzava il linguaggio sonoro di tutti i film, un problema che pochi si ponevano poiché al contributo del sound design veniva attribuita ben poca importanza nell'identità narrativa di un racconto.

1.1.3 Verso la Stereofonia – i Registratori Magnetici

Siamo negli anni '40 e per la prima volta viene sperimentata all'interno di un film la stereofonia quale nuova tecnica di riproduzione attraverso un sistema, il FANTASOUND, progettato apposta per la proiezione del film Walt Disney *Fantasia*.

Sino a questo momento gli unici metodi di registrazione e riproduzione erano monofonici, ovvero i suoni venivano registrati su un singolo canale riproducibile attraverso un singolo altoparlante. Le tecniche stereo invece permettono di dislocare i suoni a livello spaziale anche sul piano orizzontale, dunque se la monofonia restituisce percezione di prossimità attraverso la regolazione dei volumi, grazie alla stereofonia è possibile offrire anche la percezione di spazializzazione dell'ascolto. Tuttavia il FANTASOUND venne ben presto abbandonato, giudicato come inutile costoso virtuosismo, soprattutto dagli esercenti poco propensi ad affrontare elevati costi per modificare i loro impianti di riproduzione monofonici.

Bisognerà infatti attendere la diffusione della registrazione su traccia magnetica all'interno del mondo musicale prima di riprendere in mano il discorso sulla stereofonia e non solo.

Con le registrazioni musicali su traccia magnetica, si possono vedere i primi effetti di quello che in introduzione è stato citato come il concetto di una *scatola che va via via allargandosi*. Attraverso questa tecnologia infatti in primo luogo i nastri incisi vantavano la possibilità di registrare e riprodurre un spettro in frequenza più ampio (dai 100 ai 19 000 Hz), consentendo finalmente, grazie anche alle registrazioni multitraccia che si vedranno più avanti, di stratificare maggiormente la colonna sonora e attraverso le nuove tecniche di missaggio non doversi più limitare ad

accompagnare i dialoghi con solo l'aggiunta di musica e poco altro ancora per non compromettere l'intelligibilità [3].

Qualche tempo dopo anche il cinema decise di assumere questa tecnica di registrazione qualitativamente migliore e più maneggevole in termini di gestione sia in produzione che in postproduzione, tuttavia il problema maggiore continuò ad esser rappresentato nella fase di riproduzione. Gli esercenti opposero, come ai tempi del FANTASOUND, notevoli resistenze alla conversione dei propri sistemi ottici monofonici, pertanto i film registrati attraverso recorder magnetici per essere riprodotti dovevano venir convertiti su banda ottica con conseguente notevole perdita dell'incremento di dinamica e impoverimento della qualità nel corso della traduzione dai due formati.



Figura 2 Nagra il primo prototipo di registratore analogico audio a nastro in bobine, sinonimo di registratore portatile ad alta qualità che agevolò moltissimo le procedure di presa diretta sul set.

Alla fine degli anni '50 dunque le tecniche di registrazione per la produzione dei contenuti avevano raggiunto un livello superiore rispetto a quelle di riproduzione. Grande ingresso in questo senso fu proprio il maneggevole registratore magnetico *Nagra*, ben presto assimilato dai filmmakers indipendenti del cinema europeo e nel corso degli anni '60 assunto quale standard di registrazione fino all'arrivo del digitale.

1.1.4 Verso la Stereofonia – L'Arrivo della Televisione

La diffusione della stereofonia inizialmente non fu limitata dalla mancanza di sistemi tecnologici o dotazioni tecniche, come dimostrano l'invenzione del FANTASOUND negli anni '40 e l'introduzione del Nagra nel '50, ma da un generale disinteresse nella sua adozione sia da parte degli autori che, soprattutto, da parte degli esercenti. Ciò che diede dunque un primo vero impulso alla sua diffusione fu l'arrivo della televisione nelle case delle persone. La concorrenza che intercorreva con la TV spinse il cinema a cercare nuove forme d'intrattenimento spettacolare per attrarre il pubblico nelle sale e la prima risposta in questo senso fu quella di ampliare le dimensioni dello schermo.

Dal cosiddetto formato Academy 1.33:1, proporzionale a quello televisivo, vennero proposti nuovi formati quale il Cinerama (4:1), il Cinemascope (2.55:1) e il ToddAO (2.20:1). Fino a quel momento il cinema era stato "unicentrico" ovvero il centro dello schermo era il punto focale dell'immagine e della sala e similmente anche l'audio proveniente dal canale monofonico centrale era coerente con questo tipo di narrazione.

L'ampliamento dei confini dello schermo richiese che anche il sistema di diffusione fosse in grado di restituire questa spazialità, ecco che ai formati panoramici venne associato il suono stereofonico. Il Cinerama fu corredato di 4 canali su pellicola, di cui 2 surround, il Cinemascope sempre di 4 canali di cui uno surround mentre il ToddAO addirittura di 6 canali.

Per la prima proiezione di Cinerama venne creata appositamente una demo chiamata "This is Cinerama!" a cui ad un breve documentario seguiva una parte più sensazionale di un otto volante narrato in soggettiva al fine di far emergere il sensazionalismo veicolato dalla combinazione tra suono stereofonico e formato panoramico [3]. Sulla falsa riga di "This is Cinerama!" nel corso degli anni '50 vengono realizzati più o meno una decina di film allo scopo di suscitare un forte coinvolgimento nello spettatore senza grandi sviluppi narrativi.

Schermi più ampi da indagare in cui il centro non rappresenta più il punto focale dell'immagine, suoni che mirano a trasferire sensazioni forti e ad avvolgere i nuovi formati visivi, questo è, come si è detto in introduzione, il "travagliato passaggio da un cinema "verbocentrico" "vococentrico" ad un cinema pienamente audiovisivo nel quale si sviluppa progressivamente un'autonoma ricerca estetica sul suono." [3]

Tuttavia le continue resistenze degli esercenti rallenteranno questo processo ancora una volta, costringendo gli *studios* a produrre pellicole in Cinemascope stampate anche con traccia ottica monofonica, poiché la maggior parte di essi non voleva convertire per ragioni di costi i propri sistemi di riproduzione, relegando l'utilizzo del formato stereofonico solo alle produzioni più costose ed alle sale più prestigiose.

A causa di tale arretratezza tecnologica e contenutistica negli anni '60 la televisione stava uccidendo il cinema, la tipologia di contenuti offerti da quest'ultima sembrava essere un po' "fuori sync" con il contesto generale

mentre i grandi concerti musicali dei Beatles erano divenuta la nuova attrazione dei giovani.

Nel corso di questi anni registi e sound designers che si riveleranno pionieri di una tipologia di cinema di cui ancora oggi ereditiamo gli effetti, stanno maturando un loro linguaggio ed una loro estetica. Walter Murch nel 1963 trascorrerà un anno di studi in Francia in cui verrà a contatto con l'approccio innovativo della Nouvelle Vague. Al suo ritorno a scuola di cinema diventerà amico di registi dall'altrettanto fertile sensibilità quali Francis Ford Coppola e George Lucas ibridando le loro visioni di cinema sulla base delle quali fonderanno *American Zoetrope* una giovane realtà produttiva tra le cui finalità vi era quella di abbattere le barriere tra montaggio video, montaggio audio e mix sonoro.

1.1.5 Lavalier e Registrazione Multitraccia

Nel frattempo nell'ambito della produzione due nuovi fondamentali elementi si diffondono a supporto della sopra citata "scatola in espansione", si parla dei radiomicrofoni e delle registrazioni multicanale. I radiomicrofoni, piccole capsule microfoniche agganciate all'abbigliamento dell'interprete che trasmettono il segnale a un ricevitore, svincolarono definitivamente i movimenti degli attori e delle cineprese dalle limitazioni imposte dalle registrazioni sonore precedenti. I lavalier infatti consentivano registrazioni pulite dalle interferenze circostanti ed unitamente ai sistemi multitraccia era possibile riprendere in contemporanea più attori simultaneamente nella realizzazione di scene corali, liberandosi dalla costruzione tipicamente lineare del cinema classico. Una tecnica di cui Robert Altman con il suo *sound engineer* Jim Webb fece tesoro nel film *Nashville*, dalla costruzione sonora e

narrativa memorabili, come ricorda nel documentario *Making Waves* Lora Hirschberg, re-recording mixer di film quali *The Dark Knight* di Christopher Nolan, “It was amazing, how the story was driven by the sound in a way that I don’t think had happened before then in American films”.

1.1.6 La Rivoluzione Silenziosa del Dolby Stereo

Nel 1972 i film venivano riprodotti con lo stesso sistema di diffusione monofonico del 1939, è in questo panorama dunque che si affaccia finalmente la rivoluzione, quella “silenziosa”, di Ray Dolby.

Ray Dolby, dopo aver inventato il Dolby A un metodo per liberare dal costante fruscio di fondo le registrazioni analogiche, si dedica a trovare un modo per diffondere i sistemi stereofonici all’interno delle sale cinematografiche, fronteggiando il noto campo minato costituito dalle storiche resistenze degli esercenti cinematografici.

Egli dunque inventò un metodo che sulla pellicola Cinemascope fosse in grado di ricavare 4 canali da due sole tracce ottiche, in questo modo propose un modo che non imponesse agli esercenti di stravolgere il proprio equipaggiamento ma con la semplice aggiunta di un processore al lettore ottico consentì di adattare i sistemi di diffusione mono con la riproduzione del formato stereo da lui ideato. Questa tipologia di stampa della pellicola consentì inoltre una sorta di retrocompatibilità, poiché di fronte ad una configurazione monofonica il sistema avrebbe riprodotto solo il canale mono, in questo caso però perdendo tutte le informazioni indirizzate agli altri canali.

All'interno di tale panorama ancora una volta la musica⁷ ebbe il merito di fare da traino per il cinema, poiché fu Barabara Streisand, nel 1976 con il suo film *A star is born*, che spinse fortemente affinché la pellicola venisse realizzata e riprodotta in stereo, per via dei concerti in esso rappresentati che la Streisand desiderava una volta al cinema potessero avvolgere completamente l'audience.

Tuttavia il vero successo ed affermazione di questo formato, nonché rivoluzione nell'intero modo di concepire l'intrattenimento cinematografico, si ebbe grazie a *Guerre Stellari* attraverso il lavoro fatto dal sound designer Ben Burtt insieme al regista George Lucas.



Figura 3 L'incredibile successo di pubblico all'uscita di Star Wars in sala.

“Star Wars was the revolution, it was the soundtrack that change everything” poiché instillò l'idea che il suono fosse “cool” ed ebbe il merito di diffondere in molti cineasti dell'epoca la diffusa consapevolezza che la

⁷ Non a caso l'industria musicale aveva da tempo adottato i sistemi stereofonici già negli anni '50 con i dischi in vinile.

dimensione sonora oltre che coadiuvare i processi narrativi dovesse fornire un “esperienza sensoriale coinvolgente”. [3]

Da Guerre Stellari in poi la lotta tra standardizzazione dei sistemi di riproduzione tra creativi ed esercenti venne combattuta in campo aperto.

Nonostante il primo episodio di Star Wars fosse stato un successo, tutte le informazioni sonore inserite nei canali surround vennero perse nelle sale mono, pertanto Lucas con l’uscita del secondo episodio *The return of the Jedi* insoddisfatto della qualità di riproduzione del precedente fonda la società THX (Thomlison Holman Experiment). Attraverso di essa lo scopo fu quello di definire dei parametri precisi tecnici e di strumentazione a cui gli esercenti avrebbero dovuto attenersi al fine di ricevere la certificazione di qualità degli impianti, un’operazione questa che aumentò notevolmente l’attenzione, storicamente trascurata, alle condizioni di proiezione, affinché la sala si tramutasse sempre più in un ambiente capace di tradurre al meglio le volontà di chi ha realizzato il film.

Ben Burt con *The return of the Jedi* definì la gestione del sonoro multicanale che ancora oggi rappresenta lo standard più diffuso, informazioni dialogiche che mandano avanti il filo narrativo attraverso il canale centrale ed effetti speciali e “vibrazionali” rispettivamente attraverso canali surround e subwoofer, un’esperienza cinematografica nuova da cui emergerà la famosa trattazione sui film-concerto e il cinema post-moderno di Laurent Jullier.

Ioan Allen senior vice president Dolby nel corso del documentario *Making Waves* afferma, in merito al lavoro di George Lucas in Star Wars “It is the imaginative director who will say, let’s take the next step in the sound story. George Lucas and Gary Kurtz, Barbara Streisand on *A star is born*, Francis Coppola, Stanley Kubrick. Those are the key players who will say *Yes I’ll do this!*”

E così con l'introduzione nel 1977 dei subwoofers necessari a riprodurre le componenti a basse frequenze (20-200 Hz), in grado appunto di trasferire la componente vibrazionale del suono che muta l'esperienza di visione dello spettatore in diretto coinvolgimento fisico, il suono viene definitivamente sganciato dallo storico rapporto di subordinazione alle immagini [3].

1.1.7 Apocalypse Now

Walter Murch in *Making Waves* racconta di come Francis Ford Coppola prima di girare il suo film sentì di un sistema di riproduzione musicale chiamato *Tomita Planets* che attraverso 4 altoparlanti agli angoli della stanza avvolgeva completamente l'ascoltatore nel suono e affermò di voler realizzare *Apocalypse Now* con la finalità di riprodurlo in questo sistema: "And the exhibitors are going to play the picture on our terms, with our sound, the way we want them to show it", ovvero 6 canali surround di cui 3 centrali, 1 surround e 2 subwoofer, l'antenato del formato che non molto tempo dopo divenne lo standard Dolby Digital 5.1.

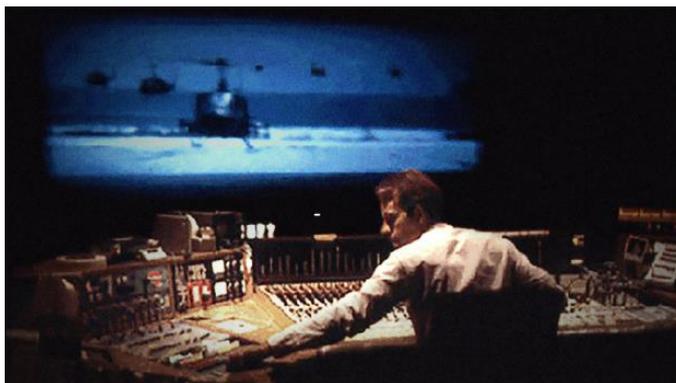


Figura 4 Walter Murch durante la lavorazione di *Apocalypse Now*. [4]

Coppola, Murch e il loro team inventarono qualcosa che non esisteva, nessuno di loro aveva neanche mai lavorato in stereo, pertanto da un workflow mono si trovarono a dover realizzare un intero campo sonoro, non a caso la realizzazione del suono di *Apocalypse Now*, come l'intero progetto filmico, fu un lavoro mastodontico in cui spesero un anno e mezzo a fare l'editing del film e nove mesi a fare il mix. Il lavoro fu gestito da Murch come una partitura musicale, ciascuno dei tecnici gestiva l'editing degli elementi che costituivano la storia, chi i vascelli, chi gli elicotteri, chi le folle, al fine di mantenere la consistenza sonora lungo il corso di tutta la storia [4].

Randy Thom all'epoca assistente di Murch nel documentario *Making Waves* riferendosi all'intera categoria dei sound designer afferma: "In a way, Walter Murch is the father of us all in this modern era of film sound". E a seguire Gary Rydstrom dice: "Apocalypse Now marked the culmination of over 50 years of film sound development. And its repercussions can still be felt today."

E' il periodo in cui accanto a un grande regista inizia a fare capolino il nome di un grande sound designer, un segnale che finalmente lo statuto del sonoro filmico ivi compreso chi vi lavora ha cambiato aura.

1.1.8 L'Era Digitale

La discretizzazione dei segnali analogici in sequenze di bit genererà uno stravolgimento nell'intero workflow cinematografico, i flussi di dati digitali infatti hanno un contenuto informativo duplicabile all'infinito, estremamente facile da manipolare a livelli di dettaglio infinitesimale e soprattutto senza il rischio di perdere qualità.

Con *Batman Returns* nel 1992 viene ufficialmente lanciato sul mercato il Dolby Digital a 5.1 canali (centrale, destro, sinistro, surround destro/sinistro, canale LFE). Dopodiché sul mercato iniziano ad affacciarsi numerosi concorrenti proponendo vari sistemi di decodifica digitale.

Uno dei grandi vantaggi del digitale per quanto riguarda l'audio inizialmente fu la capacità di inserire grandi quantità di dati in spazi molto ridotti, questo soprattutto fino a che i film rimasero in pellicola. La traccia audio digitale infatti veniva impressa sotto forma di pacchetti di dati collocati tra le perforazioni, dato il loro volume contenuto questo consentì il moltiplicarsi del numero di canali, fino a quando anche il video non si trasformò anch'esso in digitale.

Il digitale rappresenta dunque il presente. Un campo che consente ancora enormi possibilità d'esplorazione di cui le tecnologie immersive, oggetto di questa tesi, sono figlie. La possibilità infatti di indirizzare con precisione puntuale i suoni all'interno delle configurazioni di sala, l'opportunità di poter gestire mediante metadati e unità di rendering l'ottimizzazione del mix sonoro coerentemente ai sistemi di riproduzione dell'utente finale, sono tutte funzionalità le cui esigenze, osservate nei loro obiettivi originari, ovvero migliorare l'esperienza dello spettatore, garantire equivalenza di risultato tra sala mix e sala di proiezione, non nascono adesso ma già all'interno della storia della cinematografia (si pensi al THX ad esempio). Esigenze a cui oggi il digitale riesce a dare risposte mediante processi sempre più flessibili ed economici come si vedrà nel secondo capitolo.

1.2 La Vocazione “Sensuale” del Cinema

Gli studiosi che scrivono delle proprietà sensoriali del cinema hanno più volte indicato il suono come agente chiave nella sfocatura dei confini tra pubblico e film.

La capacità di recepire la dimensione acustica di un film anche attraverso il nostro corpo è stata amplificata attraverso l’evoluzione delle tecnologie in ambito sonoro che lavorano per creare una relazione sempre più “densa” tra corpo e suono. Il suono ha bucato lo schermo ed ha valicato i suoi confini strutturali vincendo la storica centralità dell’immagine e riempiendo a tutti gli effetti lo spazio. E’ in quest’ottica che le tecnologie dal Dolby Stereo fino al Dolby Atmos hanno intensificato la misura in cui “sentimento ed ascolto si trovano imbrigliati” [7].

Tuttavia questa proprietà di amplificazione sensoriale in costante espansione apportata alla dimensione visiva non è sempre stata vista di buon occhio dagli studiosi di cinematografia. In questo paragrafo ci si occuperà dunque di alcune riflessioni teoriche a confronto in merito al concetto di sensorialità filmica per capire meglio in che framework estetico-culturale si inseriscono le tecnologie immersive.

1.2.1 Il Film-Concerto di Jullier e il Ritorno al Cinema Muto di Chion

Con la definitiva affermazione delle tecnologie di diffusione sonora stereo a tre canali e poi surround, avviene quello che è stato definito da Laurent Jullier il capovolgimento della gerarchia immagine-suono. Per lungo tempo, come analizzato precedentemente, le immagini hanno guidato da protagoniste la narrazione filmica e anche quando il sonoro ha fatto il suo

ingresso, fu comunque l'aspetto dialogico ancorato alla centralità dello schermo che continuò a rappresentare l'ideologia cinematografica condivisa, ciò che Michel Chion chiama appunto *verbocentrismo*.

Limitazioni tecniche delle sale di proiezione ed un generale scarso interesse da parte dei creativi rese l'introduzione stessa del formato stereo in accoppiamento ai formati panoramici un ambito di scarso approfondimento per cui, come si è visto in precedenza, ben presto si ritornò alla stampa ottica del canale mono su pellicola Cinemascope.

Ioan Allen, vice presidente storico della Dolby Laboratories nel documentario *Making Waves* fa riferimento all'esclamazione negli anni '70 di un film executive che irritato del fermento che si stava creando attorno al nuovo formato affermò, in merito all'identità del cinema, "Santo Cielo! Si tratta di buone immagini su sedie confortevoli, non di suoni!".

Questa era la radicata cultura con la quale per lungo tempo ci si dovette confrontare nell'ambito cinematografico. Molti sono stati gli influssi e i processi che hanno portato ad un radicale cambiamento nella concezione del sonoro filmico, come descritto in precedenza, tra questi però senz'altro l'incontro tra George Lucas, Ben Burtt e la tecnologia Dolby Stereo si potrebbe definire un *turning point* storico.

Turning point che non a caso Laurent Jullier, con l'uscita di *Guerre Stellari* nel 1977, primo film in Dolby Stereo, considera l'atto di nascita del cinema postmoderno.

Egli attribuisce infatti al cinema post-moderno l'attributo di "film-concerto" [5] che diventerà molto celebre nella storia della teoria cinematografica.

I "film-concerto" sono un prodotto audiovisivo concepito per essere diffuso su un dispositivo tecnologico realizzato ad hoc, per la cui fruizione individuale è richiesto un sistema home theatre appropriato, in cui l'idea di

spettacolo è concepita per cancellare “l’effetto di rimando a favore di una dimostrazione al presente” ove i suoni prendono definitivamente il sopravvento sulle immagini.

Nella riproduzione di un film di questo genere lo spettatore si trova immerso in un “bagno di sensazioni” riflesso, da un punto di vista strutturale, nella posizione degli altoparlanti di sala che invadono per intero la cosiddetta “quarta parete”, favorendo più che un coinvolgimento una fusione tra immagine e spettatore “che mira letteralmente a far entrare il secondo nel primo”.

In questa “visione a prospettica di sala”⁸, cioè in cui il peso dei contributi informativi si sbilancia ben oltre i confini dello schermo, vengono bruciate le distanze tra film e spettatore che smettono di essere due unità distinte in una vera e propria comunione, verso la quale lo spettatore perde il controllo:

“Il cinema mi invita, o mi obbliga, a rimanere nell’orbita dei sensi. Sono confrontato e assalito da un flusso di sensazioni che non posso né attribuire a presenze fisiche né tradurre in astrazioni sistematizzate. Sono violentemente, visceralmente colpito da questa immagine e questo suono, senza poter ricorrere a nessun quadro di riferimento, alcuna forma di riflessione trascendentale, o alcun ordine simbolico.”⁹

In questo senso sono in molti ad essersi chiesti delle implicazioni controverse che questo comporta, come asserisce il Prof. Giulio Latini nel corso della conferenza presso l’AITS menzionata in introduzione, il fatto di

⁸ Concetto espresso dal Prof. Giulio Latini nell’intervento citato in introduzione e che verrà approfondito nel paragrafo 1.3

⁹ Shaviro, Steven (1993), *The Cinematic Body*, Minneapolis: The University of Minnesota Press. [Cit. in L.F. Donaldson *Feeling and Filmmaking: The Design and Affect of Film Sound*]

continuare ad aggiungere strumenti di immersione sonora deve diventare un modo per aggiungere significati al film e ridisegnare il linguaggio stesso del film e non un mero elemento per solleticare gli “istinti più bassi”, una semplice amplificazione del “bagno di sensazioni” che non consente allo spettatore, tempestato dalle sensazioni uditive, di fare un passo all’esterno del film e assestare la sua percezione cognitiva, poiché esclusivamente catturato sul piano emozionale. Egli, citando Gilles Deleuze, afferma che se “un’immagine non vale che per i pensieri che essa crea” lui aggiungerebbe, coerentemente all’ambito di dibattito, che “un’immagine non vale per i suoni che essa crea”.

Nel passare da un canale monofonico fisso al centro dello schermo ad una vera e propria esondazione di informazioni acustiche anche all’immagine viene imposto di mutare il proprio ruolo narrativo e distribuire le proprie informazioni in maniera differente.

E’ in altri termini definito il *supercampo* di Michel Chion “(..) lo spazio creato nei film multitraccia dagli ambienti naturali, dai rumori di città, musica e da ogni sorta di fruscio che circondano lo spazio visivo e che possono essere emessi dagli altoparlanti che si trovano al di fuori dei confini fisici dello schermo” [6]

Ciò che ha comportato il *supercampo* secondo Chion è quello di rendere quasi indipendente il suono dalle immagini, imponendo di rivedere completamente il modo con cui si lavora al montaggio ed alla costruzione di una scena. Ad esempio mutano d’importanza i campi lunghi di contestualizzazione poiché ci pensa il sonoro a traghettare e contestualizzare lo spettatore in un *eterno presente*.

Altrettanto interessante è la riflessione di Chion secondo cui le immagini paradossalmente di fronte a questa espansione del campo sonoro profusa dall’estensione degli altoparlanti di sala si contraggono in un sguardo

più intimista in cui l'immagine diventa quasi "voyeuristica". Le immagini intermedie dalla camera come se fosse un binocolo ci portano a spiare in situazioni a cui normalmente non potremmo accedere.

E in merito a questo "continuum sensoriale" creato dal suono che avvolge l'intimità dell'immagine Chion aggiunge:

"Questa ricerca di sensazioni (di peso, velocità, resistenza, materia e consistenza) potrebbe essere uno degli aspetti più nuovi e più forti del cinema attuale. A scapito, come oggetto, della delicatezza dei sentimenti, dell'intelligenza della sceneggiatura o del rigore narrativo? Probabilmente."

Però al contempo prosegue spezzando una lancia rispetto alla sensorialità proposta da questo nuovo cinema:

"Ma i tanto ammirati film dei vecchi tempi non hanno raggiunto la loro forza emotiva e la loro drammatica purezza a scapito di qualcos'altro, della "sensazione" per esempio, quando riproducendo i rumori ci davano una sensualità inferiore e stereo-tipizzata?"

Tutto questo sempre secondo Chion dà luogo ad un ulteriore paradosso "apparente". Il fatto che rumori, effetti, ambienti un tempo visti come elementi di disturbo in merito all'intelligibilità dei dialoghi abbiano ora conquistato un'importanza preminente, al punto da rendere i dialoghi meramente parte (non più principale) del flusso informativo, riporta indietro di 60 anni le lancette dell'orologio:

"I rumori stanno reintroducendo una sensazione acuta della materialità delle cose e degli esseri, e annunciano un cinema sensoriale che si ricongiunge a una tendenza di base del ... cinema muto. Il paradosso è solo apparente. Con il nuovo posto che occupano i rumori, la parola non è più centrale nei film."

1.2.2 Lo Spettatore Cinestesico e il “Profumo” del Suono

L’espansione tecnologico-strutturale dell’apparato cinematografico ha dunque inevitabilmente condotto ad un’espansione degli effetti sensoriali procurati nell’esperienza dello spettatore. Schermi più grandi, fitti perimetri di altoparlanti, uno o più canali a basse frequenze per la trasmissione di vibrazioni, si articolano nella creazione di ambienti di totale avvolgimento. Pertanto ogni qual volta si fronteggia l’arrivo di nuove tecnologie che puntano ad intensificare l’esperienza immersiva del cinema tornano gli interrogativi poc’anzi espressi. Dove si colloca il precario baricentro dell’equilibrio dinamico tra sensi e narrazione? Nei paragrafi che seguono si porterà sul piano della discussione un interessante approccio al concetto dei sensi in relazione al cinema, utile a rivedere alcune posizioni in merito all’apporto che un certo stimolo sensoriale può dare alla nostra comprensione del film.

In questo senso infatti sono state interessanti alcune letture di Vivian Sobchack e di Lucy Fife Donaldson che consentono di parlare di sensorialità del prodotto filmico sia da un punto di vista della ricezione dello spettatore che da un punto di vista di chi realizza il film, il cui rapporto “esperienziale” con le sonorità filmiche gli consente di interpretare con più accuratezza i significati intrinseci della narrazione.

Parafrasando alcune riflessioni della Donaldson infatti ci sono molti modi in cui è possibile interpretare il suono come “elemento cruciale dell’indirizzo sensoriale del cinema”.

In primo luogo il suono è *funzionale*, “[r]inforz[a] determinati gesti con suoni che non verrebbero uditi in un contesto reale”, come dice Mirko Perri nell’ articolo sull’ *Estetica sonora cinematografica* [2], in cui l’esempio più

comune è quello dei pugni nei film che vengono accompagnati da suoni che alludono ad uno spostamento d'aria consistente, con finalità che possono variare dalla necessità di dare consistenza al gesto oppure esasperarne la sensazione fino al paradosso, talvolta in chiave anche ironica.

In secondo luogo il suono è *trasformativo*, lavora per ampliare i confini dello schermo ricreando un contesto tridimensionale che espande e al contempo radica la vita delle immagini, dando loro profondità e prospettiva e descrivendone la qualità dei materiali e delle superfici. L'obiettivo infatti è che la materialità descritta visivamente prende corpo grazie ai suoni che trasferiscono informazioni “sulle forze, il peso, le impressioni e l'impatto del movimento nello spazio.”

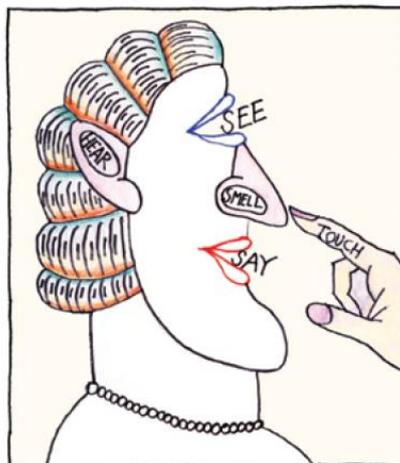
Infine, il suono è un fenomeno fisico e tattile che sollecita i sensi in maniera cross-modale restituendo un impatto sui corpi degli spettatori.

In tutti questi modi che sono ovviamente correlati, il suono fa appello alle percezioni sensoriali del pubblico, creando un'atmosfera a cui siamo chiamati a rispondere emozionalmente e cognitivamente. “Questa stretta integrazione tra film e spettatore si basa sulla nozione di intersoggettività fenomenologicamente informata del film e del pubblico.” [7]

Nei paragrafi seguenti si partirà da un approfondimento di quelli che sono gli effetti sul pubblico di un tale modo di concepire la visione di un film per poi, sulla base di alcune riflessioni della Donaldson, ragionare come questo stesso rapporto si rifletta in un rimando di meta-significati sul lavoro compiuto dai cosiddetti sound designers nell'interpretazione sonora di un film.

1.2.2.1 Le dita che sentivano sé stesse

Carnal Thoughts
EMBODIMENT AND MOVING IMAGE CULTURE



VIVIAN SOBCHACK

Figura 5 Copertina del libro di Vivian Sobchack da cui è tratto il saggio in lingua originale “Quello che le mie dita sapevano. Il soggetto cinestesico o della visione incarnata.”

Il film-concerto di Laurent Jullier che al “mostrare” predilige bagnare, avvolgere, trasportare lo spettatore ha connotato il concetto di dimensione sonora immersiva in maniera non del tutto positiva nella storia del cinema. Il bagno di sensazioni non lascia spazio al pensiero che nobilita il processo stesso di visione di un film, il bagno di sensazioni è un processo senza soluzione di continuità tra l’esperienza filmica e le sensazioni dello spettatore, il quale non può fare un passo indietro dalla proiezione cui sta assistendo poiché stordito dal “giro di giostra” che il film concerto gli sta proponendo.

Assodato che una certa costruzione sonora del film (quale quella postmoderna descritta da Jullier) associata ad impianti tecnologici di diffusione avvolgenti favorisce una ricezione più, se non esclusivamente, viscerale di quest'ultimo, in questo paragrafo ci si soffermerà in maniera più generale sul ruolo dei sensi all'interno dell'esperienza cinematografica, domandandosi se un approccio più sensoriale al cinema, ovvero che sovverte in un qualche modo la gerarchia cognizione – sensazione, sia davvero considerabile un *male*.

L'autrice Vivian Sobchack, considerata la fondatrice degli studi in merito ad un approccio fenomenologico al cinema, offre un interessante spunto di riflessione in questo senso. Più precisamente nel saggio "Quello che le mie dita sapevano. Il soggetto cinestesico o della visione incarnata" si oppone fermamente al trend storico di una certa critica cinematografica che ha declassato il concetto di esperienza filmica "sensoriale", ovvero produttrice di significati in quanto radicata attraverso le sensazioni corporee dei suoi spettatori.

La Sobchack fatta eccezione per alcuni studiosi contemporanei che cita, afferma "Fino a poco tempo fa tuttavia la teoria contemporanea del cinema ha generalmente ignorato o espunto sia la vocazione sensuale del cinema sia "l'essenza corporea materiale " dello spettatore." [8]

Ove per teoria contemporanea del cinema si riferisce a quel periodo i cui "la semiotica, lo strutturalismo e la psicoanalisi erano considerati antidoti metodologici ad una critica cinematografica blanda (..)”. Una formula questa che nonostante le evoluzioni teoriche rispetto alla cosiddetta passata *Grand Theory*, fa sentire ancora adesso i suoi effetti. Francesco Casetti osserva infatti, all'interno di un altro contesto argomentativo, che la teoria cinematografica è composta di fasi "che hanno lasciato dietro di sé delle scie le cui onde continuano a infrangersi sulla nostra spiaggia". [9]

E secondo la Sobchack gli effetti di una certo modo di pensare al cinema in rapporto ai sensi continua a produrre i suoi effetti: “In generale, tuttavia, la maggior parte dei teorici sembra ancora imbarazzato sconcertato da corpi che spesso di fronte a un film si comportano in modo rozzo e sguaiato, contraddicendo involontariamente la sensibilità raffinata, il discernimento intellettuale e il vocabolario della riflessione critica.”

Questo spesso si è tradotto in un’elaborazione di tale fenomeno quale mero affiancamento ad alte forme “cinetiche” d’intrattenimento come i parchi divertimento, in questo caso riferendosi a T. Gunning che ha definito il “cinema delle attrazioni” e i suoi effetti tipico della “triade Lucas-Spielberg-Coppola”.¹⁰

Questo rifiuto del presupposto del corpo in quanto produttore di significato è radicato nell’ideologia che vede l’esperienza umana come originariamente cognitiva il che comporta un altro, per l’autrice, (inaccettabile) “diffuso presupposto culturale” ovvero “che l’immagine filmica sia legata a una geometria meramente bidimensionale.”

Guardare al film privilegiando esclusivamente il riflesso cognitivo che produce escludendo la mediazione dei sensi, significa ridurre la sfera d’influenza del film ai confini dello schermo quale “modalità oggettiva di rappresentazione simbolica” e “disincarnando la visione soggettiva dello spettatore” che nei fatti amplia l’ecosistema del film stesso. Questo pertanto ha comportato per la teoria contemporanea del cinema difficoltà nel “comprendere come sia possibile per un corpo umano essere effettivamente “toccato” e “mosso” da un film”.

¹⁰ T. Gunning (1986), *The Cinema of attractions: Early films, its spectators and the avant-garde*. [Cit. in V. Sobchack Quello che le mie dita sapevano. Il soggetto cinestesico o della visione incarnata.]

Secondo il suo punto di vista infatti noi non vediamo meramente un film, noi ne facciamo esperienza poiché “Comprendiamo e sentiamo il film con tutto il nostro essere corporeo, modellato dalla storia e dalla conoscenza incarnata accumulate dalla nostra sfera sensoriale culturalmente connotata.”

Questo mina nel profondo i confini sin qui delineatisi di esperienza cinematografica come fenomeno scisso tra “sullo schermo” e “al di fuori dello schermo”. L’incontro infatti tra la natura espressiva del film e la natura percettiva dello spettatore avviene attraverso una terza entità di mediazione secondo la Sobchack il “corpo del film”.

“Inoltre questi corpi” (riferendosi a quelli al di fuori dello schermo, gli spettatori, e quelli sullo schermo, personaggi, immagini..) “sovvertono dall'interno la loro stessa fissità, amalgamando carne coscienza interscambiando il sensorio umano e quello tecnologico, così che il significato, e il luogo in cui è prodotto, non abbia un'origine distinta nei corpi degli spettatori o nella rappresentazione filmica, ma emerga dalla loro congiunzione. Potremmo definire questo corpo che agisce in maniera sovversiva nell' esperienza cinematografica *soggetto cinestesico*.”

Il soggetto cinestesico è un neologismo coniato dalla Sobchack che deriva dall’incontro tra le parole cinema, sinestesia e cenestesia. La sinestesia, quale percezione cross-modale, e la cenestesia, in quanto senso del sé proveniente dagli organi del nostro corpo, delineano gli aspetti che concorrono alla definizione di spettatore che non è surrogato né del corpo, né della coscienza.

I confini tra schermo e spettatore si sfumano così come l’interconnessione tra sensi che “si traducono vicendevolmente senza aver bisogno di un interprete, si comprendono vicendevolmente senza dover passare attraverso l'idea ”.

“Dunque, il soggetto cinestesico tocca lo schermo ed è toccato dello schermo - capace di commutare la visione nel tatto e viceversa senza pensarci e attraverso un'attività sensuale e cross-modale, di vivere il film come al contempo qui e là, anziché collocando chiaramente la sede dell'esperienza filmica sullo schermo o al di fuori di esso.”

Secondo tale visione dunque risulta emblematico osservare come lo sviluppo dell'impianto sonoro di un cinema sembri già alludere a questa rottura dei confini tra schermo e spettatore abbracciando quest'ultimo nella visione, riflettendo da un punto di vista strutturale l'ideologia di *conoscenza carnale* del film.

Nell'parole “Quello che le mie dita sapevano”, titolo del saggio da cui sono tratte le riflessioni sopra riportate, è racchiuso il senso di tutto il suo pensiero. Esse infatti si riferiscono al momento iniziale del film *Lezioni di piano* di Jane Campion in cui le mani della protagonista poste a ventaglio davanti agli occhi propongono una visione in soggettiva in cui, in virtù della sequenza di montaggio, in un primo momento non è possibile distinguere il fatto che quelle siano effettivamente le sue mani. L'autrice a tal proposito riporta la descrizione di C. Jacobs in *Modern Language Notes* di questo momento del film: “Lunghi, irregolari fasci di luce di un rosa tendente al rosso si aprono a ventaglio sullo schermo, sfocati come una fotografia a colori non riuscita di traslucidi vasi sanguigni”.¹¹

Queste immagini misteriose ed attraenti ma a primo impatto “cieche” poiché frutto non di un punto di vista oggettivo, scaturiscono nell'autrice una *conoscenza carnale*, come se le sua dita si sentissero sfiorate da

¹¹ C. Jacobs (1994), *Playing Jane Campion's Piano: Politically*. [Cit. in V. Sobchack *Quello che le mie dita sapevano. Il soggetto cinestesico o della visione incarnata.*”]

quell'immagine, prima ancora che l'informazione oggettivamente decodificabile dal suo cervello in merito all'identità di quella immagine venisse fornita dal film. L'autrice spiega che era come se le sue dita sapessero *cosa stavano guardando*, come se “sentissero sé stesse”.

Per quanto la Sobchack non si soffermi mai nell'approfondire come i suoni influiscano in questo processo, ma si limiti a sottolineare l'attività cross-modale dei sensi attivata dalla vista in primis e a seguire dall'udito, trattati come strutture individuali, io sottolineerei che è la specifica combinazione dei loro contributi, la cosiddetta immagine-sonora, che dà avvio a questo scambio d'informazioni tra sensi benché non direttamente stimolati.

Il suono come in una reazione chimica produce effetti sulla visione mutandone gli aspetti originari ed ampliando lo spettro sensoriale attraverso il quale noi ne percepiamo l'essenza contribuendo al nostro significato emozionale del film.

1.2.2.2 Sfocare i confini tra pubblico e schermo

Dalla visione incarnata di Vivian Sobchack, il cui corpo del film amalgama “sensorio umano e tecnologico”, ad alcune riflessioni in merito alla professione di chi concretamente si occupa di *sfocare i confini tra pubblico e schermo*. Come anticipato in precedenza infatti, ciò che emerge essere interessante nell'analisi della sensorialità veicolata dalla dimensione sonora del film è il processo stesso alla base di chi la interpreta e realizza, ovvero i processi che seguono i sound designers, che come propone Lucy Fife Donaldson, hanno essi stessi a che fare con una propensione verso la “fusione sinestesica” del lavoro.

Michel Chion in merito alla realizzazione degli effetti sonori di un film introduce il concetto di “interpretazione/traduzione” processo secondo il quale il suono registrato comunica un'impressione di quella dimensione acustica piuttosto che la sua rigorosa realtà:

"Lo spettatore del film riconosce che i suoni sono veritieri, efficaci e adatti non tanto se riproducono ciò che sarebbe udito nella stessa situazione nella realtà, ma se riproducono (trasmettono, esprimono) i sentimenti associati alla situazione " [6]

Proprio nelle maglie di questa descrizione, nel concetto di “sentimenti associati alla situazione” si intessono i procedimenti che seguono i sound designers per realizzare il materiale sonoro necessario a ricostruire acusticamente l'universo narrativo di cui sono responsabili.

Come racconta Mirko Perri, in una video-intervista condotta nel merito degli argomenti di questa tesi, oltre al prezioso confronto con il regista, il lavoro parte dalla sceneggiatura, nel saper leggere tra le righe della storia ove tutto è contenuto.

I suoni di un film, salvo precise annotazioni in sceneggiatura, albergano nelle maglie del racconto e chi si ritrova a supervisionare l'aspetto sonoro cerca già nelle prime fasi di avvicinarsi il più possibile a tutto quello che possa condurlo a contatto con l'assimilazione delle ambientazioni o delle caratteristiche della storia.

Ad esempio Steven Spielberg, come riporta la Donaldson, per le riprese di “Incontri Ravvicinati del Terzo Tipo” volle che il sound designer e sound effects editor Frank Warner andasse con lui sul set in Alabama prima delle riprese, benché l'accesso non fosse consentito a nessuno, affinché sentisse l'esperienza (“feel the experience”). Il fatto solo di vivere, respirare il luogo, infatti, lo avrebbe portato a calarsi nella percezione fisica degli spazi e dell'atmosfera che sarebbe stato necessario ricreare.

A volte invece succede il contrario che i suoni abitino la storia e i personaggi prima ancora che la storia abbia preso vita. Nel documentario *Making Waves* Ben Burtt racconta come da giovane studente l'inizio della sua incredibile carriera sia nata proprio perché un anno prima delle riprese George Lucas gli diede il compito di trovare la voce di un *Wookiee*. Il regista per poter dare corpo ai personaggi aveva bisogno di figurarsi in anticipo quali sarebbero state le loro voci, quale impronta ed espressione della loro identità, e Burtt iniziò dunque delle sedute di registrazione in vari zoo con altrettante manipolazioni successive in post-produzione tornando qualche tempo dopo con la voce richiesta. Il resto, come si suol dire, è storia.



Figura 6 Immagini tratte dal documentario *Making Waves*, sulla sinistra Ben Burtt presso uno zoo alla ricerca delle sonorità adatte “per dar una voce” al personaggio di Star Wars sulla destra, un Wookiee. [4]

Spesso in mancanza della possibilità di registrare i fenomeni che scaturiscono un certo evento è necessario ricostruire con l’immaginazione i suoni ad esso correlati e dunque porsi la domanda come riferisce R. King “qual è il suono analogo per questo, quale suono ci farebbe sentire così?”. L’autrice Lucy Donaldson parafrasando alcune sue conversazioni con il sound designer osserva che nella ricerca della risposta a questa domanda vi è la necessità di ragionare non solo su come il suono possa abbinare appropriatamente le immagini ma soprattutto come possa trasmettere “la

forza, l'agire, l'atmosfera" di quel suono, la cui funzione, come dice lui deve essere da un punto di vista subliminale immediata: "è più una reazione istintiva che stai cercando, non tanto una reazione riflessiva intellettuale."¹²

Dalla sceneggiatura, alla preparazione fino alla presa diretta e le sessioni di *field recording* la necessità di porre attenzione alle diverse caratteristiche cross-modali del suono si fa sempre più evidente.

A tal proposito risulta molto affascinante la dichiarazione di Walter Murch che nel merito del rapporto tra suono ed ambiente afferma di preferire le registrazioni vere dei suoni in virtù del loro speciale rapporto con l'aria e l'atmosfera circostante: "È una predisposizione generale a pensare sempre all'aria che circonda qualcosa. Per me è incredibilmente emozionante. L'aria ha molto a che fare con questo; è una specie di profumo del suono: il suono senza aria non ha odore. Quando hai aria intorno, ti colpisce all'improvviso, come un odore che non senti da quindici anni – e i ricordi tornano."¹³

In questo senso Murch parla non solo di come la ricostruzione di un suono o la sua riproduzione connotino la sua materialità ma anche di come il suo rapporto con l'ambiente, e dunque relativa registrazione, trasformi le sue qualità evocando i sensi di chi lo ascolta.

E a proposito di Murch, In merito alla titanica impresa *di Apocalypse Now*, Randy Thom, ai tempi suo assistente, nel documentario *Making Waves* dice "penso che la più grande lezione imparata da *Apocalypse Now* sia stata, stando lì seduto, cercare di immaginare di momento in momento quali suoni usare e quali suoni NON usare. Questo genere di decisioni rappresentano l'essenza di un film".

¹² R. King (2015), *Interview with author, interview, Los Angeles: Sony Studios, 29 July 2015* [Cit. in V. Sobchack *Feeling and Filmmaking: The Design and Affect of Film Sound*

¹³ V. Lo Brutto (1994), *Sound-On-Film: Interviews with Creators of Film Sound* [Cit. in V. Sobchack *Feeling and Filmmaking: The Design and Affect of Film Sound*

Dalla sceneggiatura, alle riprese fino al momento cruciale di tutta la lavorazione, la post-produzione.

In questa fase infatti, che eredita tutte le scelte e le considerazioni fatte in precedenza, avviene quello che Mirko Perri definisce il processo necessario a “(..)creare la corretta tessitura sonora in base alle specifiche qualità estetiche di ciò che vedo sullo schermo”. Nel medesimo articolo [2] per approfondire meglio il concetto avanza un’associazione interessante attraverso un esempio che mostra la palette colori di *The Revenant* graduale, armonica ed omogenea e di *Django Unchained* molto più contrastata e vivace, in merito alle quali si chiede se sia possibile parlare analogamente di *palette sonora*.

Egli illustra come infatti dall’ascolto del piano sequenza iniziale del film *The Revenant* tutti i suoni risultino, pur in un’azione movimentata, fluidi ed amalgamati nel contesto, mentre come è presagibile già dai colori della scena del secondo film citato, essa richiama sonorità che ricalcano la tipica estetica Tarantiniana “I dettagli, anche di elementi molto piccoli, sono esplosi sullo schermo e a volte esasperati fino all’inverosimile rafforzando l’estetica pop che l’immagine restituisce” .

A conclusione di tale riflessione Perri riporta nell’articolo il riferimento, quale estrema ma incisiva semplificazione del concetto, di due rumori di sparo molto diversi tra di loro che potrebbero restituire in chiave acustica due sottotesti narrativi complementari di una scena, il primo più minimale e naturalistico, il secondo più strutturato e “muscoloso” e a tal proposito prosegue: “Bisogna immaginare la palette audio come la possibilità di tendere, più o meno, verso l’uno o l’altro estremo di quanto ascoltato sopra cercando di trovare il gradiente di colore sonoro adatto all’immagine che si ha davanti. Bisogna considerare tutti i fattori possibili, il momento narrativo, il ritmo del montaggio, il colore fotografico, le sembianze dei personaggi,

l'ambientazione; si deve tenere conto insomma di tutte le caratteristiche generali del film per evitare di usare sonorità che generano involontarie dissonanze cognitive e non fanno aderire il suono all'immagine.”

Dunque parte del mestiere del sound designer è quello di sapersi calare all'interno della narrazione affrontando un processo di “immedesimazione” per arrivare a restituire una dimensione sonora che sappia cogliere l'essenza della narrazione nascosta tra le righe della storia. A proposito di *conoscenza incarnata*.

Un processo questo che Lucy Donaldson, riferendosi ad alcune dichiarazioni del sound designer Hillman in linea con tale modus operandi, lo definisce un “ricorso a un'interiorità ispirata a Stanislavskij”, in sostanza una maniera di immedesimarsi nella situazione al punto di poter diventare vero tramite della traduzione acustica del film. Pratica che, fuori da qualsiasi accezione figurata, i *foley artists* incarnano pienamente, nel loro ricalcare i movimenti di un attore sullo schermo (Figura 1.7 [10])¹⁴.

¹⁴ La scena mostra il protagonista intento ad intrattenere un gruppo di giovani amici nell'altra stanza simulando con dei sacchi di riso e un panno la sensazione di un cowboy che accende un fuoco. Come dice Massimo Mariani nel suo libro “Il suono per il cinema” questa scena “benché evidentemente estranea all'ambito del rumorismo “professionale” quello che, per intenderci, si svolge all'interno di studi altamente specializzati mediante le prestazioni di esperti” è interessante da citare nell'ambito del foley perché in grado “di portarne alla luce la primaria e sostanziale componente ludica. Componente senza la quale forse una simile tecnica non avrebbe nemmeno avuto origine, che è stata presente nello spirito di tutti i grandi creatori di suoni del cinema delle origini e classico – se ne veda lo straordinario lavoro e la prodigiosa carica inventiva già nelle prime animazioni disneyane.”



Figura 7 La scena tratta dal film *Lisbon Story* (1995) di Wim Wenders [10]

Va sottolineato tuttavia che questo modo di interpretare la dimensione sonora di un film è figlia di approcci illuminati che richiedono registi con una certa propensione e sensibilità ad accogliere il valore del sonoro nel proprio film, oltre che film che effettivamente richiedano una lavorazione di questo genere.

In questo senso in un'intervista rilasciata alla Dolby, Skip Livesay, tecnico del suono che lavora con registi come i fratelli Cohen o Alfonso Cuarón, racconta come un giorno un regista per il quale stava lavorando insistette molto sul fatto che in fase di sound design venissero risolti alcuni passaggi narrativi del film. Venne pertanto lasciato spazio a Livesay per lavorare sulla scena, il quale investì molte energie nell'elaborazione di suoni e ricostruzione di ambienti che potessero sottendere alla narrazione che gli era stata presentata.

Nel momento in cui dovette poi mostrare il suo lavoro, alla fine della scena, alle sue spalle racconta ironicamente di sentir piombare il tipico glaciale silenzio di disappunto. Regista e produttore infatti non erano per nulla soddisfatti del lavoro, poiché Livesay spiega che per lui sound design significava "creare" qualcosa che prima non esisteva e, per citare testualmente le sue parole emerse dal confronto con il regista, "to me sound design means sex and to you clearly sound design means something else". Livesay condivide infatti che in quel momento capì che ogni volta che l'autore parlava di sound

design in realtà lui si riferiva all'inserimento di suoni quale la chiusura di una porta, l'inserimento di una musica e un po' di mix generale. Mentre per lui sound design significa creare qualcosa che normalmente non sarebbe possibile registrare, ricostruire una nuova realtà.

Livesay ovviamente, una volta comprese le reali volontà dell'autore, cancellò tutto e ricominciò da capo. Questo perché, per citare Walter Murch, il regista è e deve rimanere *il sistema immunitario di un film*, e dunque nessuno meglio di lui può sapere quale dev'essere l'identità del progetto. Pertanto si riportano tali aneddoti e riflessioni, non per sovvertire il 'sistema del set', ma al fine di evidenziare che simili trattazioni di natura sonora si ispirano ad un certo modo d'intendere l'impiego di questo strumento nel film.

Ma è stato detto che registrazione, immaginazione, manipolazione e stratificazione sonora sono solo parte degli strumenti attraverso cui il sound designer cerca di sfumare sempre più i confini tra schermo e platea. E' giunti alla fase di mix infatti che egli cerca di suggellare definitivamente questo rapporto tra ambiente di riproduzione e spettatore.

Interessante in questo senso è il racconto di Mirko Perri in merito al suo lavoro su *Dogman* di Matteo Garrone, per il quale lui e la sua squadra hanno ricevuto il David di Donatello per il miglior suono:

“Nel film “il contesto” assume un ruolo fondamentale tanto da divenire un vero e proprio personaggio. Il non-luogo in cui è calata la vicenda di Marcello, il protagonista, sembra un'astrazione distopica, una cornice che racchiude all'interno un grande senso di umanità e una violenza feroce e disumana. Come tutti gli altri personaggi aveva bisogno di una voce propria. Marcello, con l'avanzare del racconto, entra in una spirale sempre più violenta. L'ambientazione sonora, parallelamente al protagonista, è rassicurante in prima battuta, poi comincia gradualmente a scemare, come

se il contesto in cui Marcello era perfettamente integrato lo abbandonasse, non gli parlasse più, fino ad arrivare ad un silenzio irrealistico sul finale, ovvero momento più drammatico del film. Appena lo schermo va a nero e si chiude lo sguardo su questa storia, il suono riemerge sui titoli di coda, torna a diventare un elemento naturalistico del tutto indifferente ai sentimenti dell'uomo, descrive un luogo che continua a vivere incurante del dramma appena mostrato nel tentativo di scuotere emotivamente lo spettatore di fronte all'assurdo della tragedia ormai consumata.” [2]

E' dunque proprio attraverso queste scelte, che rendono i professionisti del suono figure invisibili poiché efficaci se nascoste nelle trame del risultato finale, che “il corpo del pubblico viene intrecciato al corpo del film”. [7]

1.3 Tra immersione e narrazione qual è il futuro delle tecnologie immersive?

L'aggiunta di altoparlanti al soffitto, la discrezionalità con cui i suoni ora possono essere rivolti ad ogni singolo altoparlante in sala amplificando la diffusione e la profondità di campo del sonoro, i sistemi immersivi non possono che dare nuova tridimensionalità alla vocazione sensoriale del cinema poc'anzi raccontata, rendendo sempre più vivida e avvolgente l'esperienza che facciamo del *presente sullo schermo*. Tuttavia le perplessità in merito a tali sistemi non si esauriscono al tema della cosiddetta “mera amplificazione del bagno di sensazioni” di Jullier, qui di seguito se ne

riporteranno infatti delle altre attraverso cui si cercherà di strutturare una visione di più ampio raggio in merito all'utilizzo di tali tecnologie.¹⁵

Nel corso di un interessante intervento presso l'AITIS (Associazione Italiana Tecnici del Suono) [11] tenuto nel 2016 per parlare del *making of* sonoro di *Youth*, primo film rilasciato in Italia in Atmos, Giulio Latini, come anticipato precedentemente, Docente di Comunicazione Multimediale presso l'Università degli Studi di Tor Vergata, esprime una serie di riflessioni utili coerentemente all'ambito di approfondimento di questa tesi al fine di delineare le effettive prospettive di contestualizzazione di tale sistema. Partendo infatti da alcune considerazioni in merito al “ricatto emozionale” esercitato da quelle *bolle acustiche* “titillabili degli istinti più bassi”, egli prosegue ponendo alcuni interrogativi in merito alla necessità di guardare alle nuove tecnologie con un discernimento prospettico:

“C'è un'ulteriore domanda che a me personalmente impegna cioè come è sempre accaduto nella storia degli sviluppi tecnologici di qualunque tipo applicati al cinema e non, non è mai solo in questione una tecnologia applicata a un sistema di riproduzione, di espressione di immagine e suoni, ci si è sempre in fondo chiesti su queste tecnologie **di quale mondo più ampio stanno partecipando?**”

A tal proposito egli cita che non a caso quando il Cinemascope ha fatto la sua comparsa con il sistema stereo integrato il terreno nel quale muoversi era fertile, poiché il cinema doveva competere con la televisione, proponendo nuove forme di spettacolo, e poteva contare sul fatto che gli spettatori

¹⁵ N.d.R. Di seguito quando si parlerà delle tecnologie immersive ci si riferirà sempre all'utilizzo in particolare del Dolby Atmos, preso quale modello di riferimento, non solo perché primo ad aver stabilmente inserito sul mercato un sistema per canali ibridato con uno object-based, ma soprattutto perché i principali film cui si fa riferimento nella tesi, e degni di nota nell'ambito di questa ricerca, sono stati realizzati con questa tecnologia, al momento la più diffusa nel workflow cinematografico mondiale.

arrivavano da un'educazione sonora per via radiofonica e visiva per via fumettistica, il che li rendeva pronti ad un tale salto. Pertanto riprendendo il discorso in merito alle tecnologie immersive prosegue:

“Che spettatore è quello odierno che chiede l'Atmos piuttosto che altri sistemi. Che uomo è? Che donna è? Che bambino è? (..) Cioè non si arriva a una trasformazione tecnologica soltanto perché lo vogliono, lo pretendono i detentori di quella tecnologia. Le tecnologie si inscrivono dentro un più complesso contesto, cornice culturale e soprattutto una cornice che ha che vedere con una più complessa istanza spettatoriale”.

Qual è dunque lo spettatore di un film immersivo o più specificamente di un film in Atmos? Qual è il contesto tecnologico-culturale nel quale si configura questo nuovo sistema?

I sistemi sonori immersivi odierni oltre ad un ampliamento dei confini sonori si avvalgono di un workflow object-based che, come si spiegherà in dettaglio in seguito, consente un'estrema economia operativa e temporale al fine di restituire attraverso un unico master un prodotto facilmente fruibile non solo in diverse configurazioni di sala, ma anche in elevata qualità mediante fruizione individuale. L'escamotage tecnologico-commerciale evidenzia che le bolle immersive sonore vengono predisposte non solo più per la grande sala cinematografica ma anche per favorire il consumo mediale individuale, quello con un paio di cuffie davanti allo schermo del pc o dello smartphone.

Sulla base di tale scenario torna dunque pertinente la conclusione da parte del Prof. Giulio Latini del sopracitato intervento:

“Questo avvolgimento, questa bolla immersiva che ci costringe da molti anni al suo dominio, alla sua dominanza è qualche cosa che non appartiene solo al cinema, né solo all'Atmos evidentemente. (..) Ma essere sempre immersi dentro una percezione rappresentativa, illusoria, verosimile

della tridimensionalità è l'unica cosa che occorre? O forse invece il sistema stesso, l'Atmos, in linea di principio e di fatto in mano a delle sensibili figure, non solo quelle registiche, possono portare a qualcosa di altro. Che ci fa anche dimenticare questo fatto di essere immersi ma perché (..) **ci fa aumentare il sapere di un film**. Se un film non produce un sapere aumentato, ancorché una sensazione aumentata, francamente io non so tanto, troppo che cosa farmene.”

Sulla base di questi spunti e grazie al generoso confronto in merito con il sound designer Mirko Perri, nei prossimi due paragrafi si mirerà coerentemente ai sopracitati interrogativi, se non certo a dare risposte, ad ipotizzare scenari all'interno dei quali l'Atmos, e più in generale le tecnologie immersive, possono trovare reale e benefica contestualizzazione.

L'approccio adottato sarà del tipo induttivo, per così dire, ovvero si baserà sull'analisi degli ultimi due film del regista che al momento è stato tra i pochi, se non il solo, ad utilizzare l'Atmos con mirata efficacia narrativa, Alfonso Cuarón. Il regista infatti sperimentò questa tecnologia nel 2013, agli albori della sua presenza sul mercato, all'interno del film *Gravity*, ambientato nello spazio, in cui la tensione emotiva del film e l'assenza di gravità nella quale si muovevano Sandra Bullock e George Clooney parevano essere i presupposti ideali per l'utilizzo di una simile tecnologia sonora. Compresi e sperimentati a fondo gli strumenti offerti da tale sistema, nel 2018 egli trasporta il bagaglio esperienziale all'interno del film *Roma*, una “pellicola” dai connotati diametralmente opposti rispetto alla precedente, una storia molto più intima, in bianco e nero, interpretata da non attori e in lingua originale spagnola, dal taglio quasi autoriale.

Cuarón dunque, assimilate le logiche del sistema per renderle strumento e non soggetto delle finalità narrative, detta la rotta, quella che Perri suggerisce essere la *deriva intimista dell'Atmos*.

Come infatti condivide nel corso di un'intervista per questa tesi, oltre al sensibile ed inaspettato uso che ne fa il regista messicano da un punto di vista narrativo, è altamente probabile che l'Atmos trovi la sua vera fortuna grazie all'ascolto binaurale in cuffia, che come vedremo è il "downmix" che meglio interpreta e valorizza i connotati di altezza e localizzazione offerti dalle tecnologie sonore tridimensionali.

A proposito di bolle immersive.

1.3.1 Da *Gravity* a *Roma*, la deriva intimista dell'Atmos

1.3.1.1 *Gravity* tra spazio e spazializzazione

Nel 2013 Alfonso Cuarón stava realizzando *Gravity* ideato e scritto dal regista specificamente per un sistema surround 7.1. Mentre aveva praticamente concluso la fase di post-produzione, a poco tempo di distanza dalla proiezione in sala, viene contattato dalla Dolby per andare ad ascoltare una demo Atmos. Accolto in una sala completamente buia viene avvolto dal suono di una pioggia battente finalizzata a trasferire la sensazione del suono proveniente dall'alto, oltre che a trasferire la precisione spaziale puntiforme con cui questo era in grado di circondare l'ascoltatore. Il regista comprendendo al volo la potenzialità del sistema accoglie la proposta di riaprire il progetto di mix in 7.1 per realizzarlo nel nuovo sistema Dolby.

Cuarón, riferendosi a *Gravity*, dichiara in un'intervista di aver pensato ad un film in Atmos prima ancora di averlo mai utilizzato.

Lunghissimo piano sequenza iniziale del film, dalla prospettiva di un satellite vediamo la superficie terrestre e mentre la camera si muove in un morbido e lento pan laterale, che inizialmente è difficile distinguere se sia il movimento della crosta terrestre o della camera, veniamo subito introdotti nella chiave narrativa di tutto il racconto, due flebili voci radiofoniche una solo a destra e l'altra solo a sinistra provengono da qualche luogo sconosciuto. Sono voci soggettive? Ricordi?



Figura 8 Immagine d'apertura del film Gravity (2013) di Alfonso Cuarón

A mano a mano che si avvicina all'orizzonte uno *space shuttle* comprendiamo che le voci sono reali ed appartengono agli astronauti di una missione intenti a riparare un guasto. La camera sempre nello stesso lungo piano sequenza si muove morbidamente come se fosse essa stessa un personaggio che agile può fluttuare nello spazio mostrandoci i vari astronauti della spedizione, non li vediamo quasi mai in volto se non distrattamente dietro la sfocatura dei caschi spaziali, a guidarci e portarci dentro al contesto della storia sono le loro voci filtrate dalla radio (dunque dal loro punto di vista perché altrimenti nello spazio non potremmo udirle).

Estremamente precise nella loro collocazione spaziale esse guidano il racconto, le indicazioni del *Controllo di Missione Huston*, la musica diegetica messa dal responsabile della missione *Kolawsky* mentre gli astronauti si scambiano delle battute tra loro e poi l'ingegner *Stone* (Sandra Bullock), che come ad anticipare il presagio di quello che accadrà, chiede di spegnere la musica mentre non riesce a risolvere una parte del guasto. A un certo punto in radio viene annunciato di interrompere immediatamente l'operazione perché dei detriti relativi all'esplosione di alcuni satelliti stanno viaggiando ad una velocità impressionante e noi come i protagonisti rimaniamo ancora per poco ciechi al pericolo. Nulla sembra cambiare attorno nel silenzioso e quieto spazio ciò che guida e muove la nostra tensione è questo rimpallarsi di voci che descrivono quanto sta avvenendo, con una musica che ora inizia a crescere acuendo il senso di drammaticità fino al culmine della situazione in cui i detriti si scaraventano mediante forti vibrazioni, uccidendo gli altri membri dell'equipaggio, compromettendo irrimediabilmente lo shuttle e costringono *Kolawsky* e *Stone* alla deriva nello spazio tagliando completamente i ponti radio con *Huston*.

Da qui in poi prosegue la storia di *Gravity*.

Cuarón dunque investe il film di una grande sensibilità sonora cui trasferisce il compito di diventare un vero e proprio elemento narrativo. Essendo i suoni onde di pressione che viaggiano nella materia, nel vuoto dello spazio dove si trovano i protagonisti essi non sono direttamente udibili, pertanto la scelta del regista è stata quella di trasmetterne la percezione mediante il contatto con gli oggetti. Nell'afferrare e toccare gli elementi o nell'impatto degli urti dei detriti attorno a sé i suoni vengono "propagati" come vibrazioni trasmesse al loro orecchio, e di rimando allo spettatore. L'utilizzo del canale delle basse frequenze per questo tipo di interazione

restituisce la sensazione di una forte immedesimazione con il punto di vista del personaggio.

Già da questa scelta dunque si può osservare come il canale LFE solitamente adoperato per eventi di grande impatto visivo in questo caso è stato adottato non esclusivamente come strumento d'amplificazione sensoriale ma quale elemento funzionale ad un'esigenza propriamente narrativa.

Per raccontare gli effetti della mancanza di gravità, protagonista dell'intero impianto registico-narrativo della vicenda, che causa vorticosi spostamenti dei personaggi nella scena, il regista adotta un ulteriore espediente sonoro. Cuarón espande letteralmente i confini dello schermo avvalendosi del suono per ricostruire uno "spazio cosmico in sala", contravvenendo ad una delle più note regole del mix sonoro, disancora ovvero i dialoghi dagli altoparlanti centrali. Per avvolgere e coinvolgere lo spettatore a partecipare del disorientamento di cui sono vittime i protagonisti dispersi e al contempo per renderci costantemente consapevoli della reciproca posizione nello spazio, Cuarón si avvale di un'ampia spazializzazione dei dialoghi.

Essi sono perfettamente solidali con la posizione dei personaggi nella scena e, poiché si trovano a vorticare senza ausilio della gravità, i dialoghi vengono gestiti per restituire esattamente questa sensazione, accurati ed oggettivi come se i personaggi sorvolassero sulle nostre teste per poi trovarsi alle nostre spalle chiedendo aiuto. Il regista definisce questa scelta di utilizzo del suono "geograficamente letterale". [12]

Skip Lievsay, re-recording mixer di Gravity, spiega che per fare questo è stato molto utile lavorare ad un secondo mix in Atmos, poiché ciascun dialogo che era in formato monofonico è stato convertito in oggetto al fine di poterne eseguire un panning sonoro di estrema precisione spaziale. (Si

rimanda per una migliore comprensione di questo argomento alla lettura del terzo capitolo).

A tal proposito Lievsay puntualizza che secondo lui sono semplicemente due le cose più importanti da fare lavorando in Atmos. La prima, utilizzarlo in maniera accurata proprio nelle scene che ti consentono di massimizzare l'esperienza immersiva, la seconda sapere dove non utilizzarlo. Un procedimento possibile grazie al fatto che l'Atmos, come si vedrà più avanti, integra il workflow multicanale con quello ad oggetti per consentire di lavorare in parallelo a seconda di quanto richiesto dalla scena.

Nel descrivere da un punto di vista sonoro le tipicità dello spazio e della storia Cuarón e Lievsay rompono un'altra regola tipica del mix sonoro, ovvero accompagnano gli unici suoni adottati nel racconto, le voci dei protagonisti e le vibrazioni prodotte dagli oggetti, con un'ampia spazializzazione dell'intero impianto musicale del film. Essi dunque trasgrediscono il principio secondo cui le musiche, se extradiegetiche, rappresentano dei commenti alla scena che vanno tipicamente riprodotti attraverso gli altoparlanti centrali. In un contesto narrativo classico infatti non vi è alcun motivo oggettivo per cui un commento musicale si muova in una sala, esso partecipa alla scena sottolineandone i passaggi narrativi, ma non partecipa alla scena nel senso che si muove attraverso di essa. L'accoppiamento di questa ampia spazializzazione musicale coniugata con le vibrazioni trasmesse dal contatto degli oggetti e il vuoto dello spazio, generano dunque nello spettatore un forte impatto di tensione emozionale, in cui i suoni si comportano come dei personaggi nella storia che mirano ad avvolgere lo spettatore in questo senso di presenza.

Cuarón inoltre utilizza il lungo piano sequenza iniziale, scelta estetico-narrativa già esplorata con il DOP Emanuel Lubezky in altri suoi film in passato, per restituire lo stesso clima che è possibile respirare dai

documentari nello spazio, caratterizzati da queste lunghe inquadrature uniche in cui non è ovviamente possibile avvalersi del montaggio. Inoltre questo consente di introdurre lo spettatore gradualmente nel clima del racconto in modo tale che quando si arriva al culmine del climax iniziale, in cui i detriti colpiscono lo shuttle, egli si sia tramutato in un terzo personaggio della storia, vivendo in prima persona gli eventi sullo schermo. Questo concetto dell'immedesimazione degli spettatori è un tema molto caro a Cuarón per il quale, come si vedrà anche in *Roma*, egli si avvale del suono in questa formula spazializzata e puntiforme. L'utilizzo infatti di lunghe inquadrature è un ottimo espediente attraverso il quale deputare all'accurato utilizzo del suono il compito di guidare il ritmo del racconto e richiamare lo spettatore nelle sue trame.

Una scelta questa che ovviamente richiede molto più lavoro di preparazione, poiché escludendo l'ausilio del montaggio è necessario preparare e ripetere le scene fino al punto che tutte le dinamiche interne alla storia si sviluppino senza intoppi. Una procedura che in *Gravity* ha richiesto molto lavoro di progettazione dei movimenti di camera e di simulatori virtuali in virtù degli effetti speciali del film e in *Roma* viceversa ha richiesto grande lavoro di coordinamento a volte di centinaia di comparse, a volte dei soli attori protagonisti, al punto di ripetere le scene anche una settantina di volte (in un'intervista emerge che sono stati impiegati anche tre giorni per "portare a casa" una sola scena). [13]

1.3.1.2 *Roma* e i profumi sonori che avvolgono il silenzio di Cleo

Si potrebbe dunque pensare che *Gravity*, quale film di genere dai grandi effetti speciali, interpretato da famosi attori hollywoodiani e ambientato

nella vastità dello spazio possa essere il più ovvio, se non scontato, campo d'applicazione di un sistema sonoro dalle altrettanto grandi promesse immersive che a rigor di logica si sposa perfettamente con quei film di genere dalle facili tensioni emozionali, quali sci-fi, horror e così via.

Eppure, come anticipato, Cuarón ci dimostra qualche anno dopo che l'aver fatto propria l'identità di tale tecnologia in *Gravity* l'ha condotto a declinare l'utilizzo di questo sistema in un genere totalmente in controtendenza, traducendo le logiche dell'Atmos in un film dal taglio molto più autoriale ed intimo, senza ingombranti effetti speciali o stars che vorticano nello spazio.

Roma racconta la storia di una giovane, silenziosa domestica messicana, Cleo, sullo sfondo altrettanto protagonista delle vicende della famiglia per la quale lavora, composta da una coppia ormai in crisi e da quattro giovani ragazzini, a cavallo tra il 1970 e il 1971 presso il quartiere Roma di Città del Messico. Un giorno Cleo esce con un ragazzo del quale rimarrà incinta che quando verrà a sapere della notizia la lascerà e successivamente intimerà aggressivamente di non cercarlo più. Cleo silenziosa porta avanti la sua gravidanza lavorando per la famiglia, Sofia, la padrona di casa, la sostiene in questo difficile momento mentre nel frattempo si consuma la sua di sofferenza per via del marito che ha trovato un'altra donna e si sta allontanando dal nucleo familiare. Maturata ormai la prospettiva di diventare ragazza-madre, un giorno Cleo accompagnata ad acquistare una culla per il nuovo bebè rimane invischiata in una violenta rivolta studentesca in cui si troverà faccia a faccia con una pistola puntata contro da Fermín, il fidanzato che l'ha abbandonata, coinvolto negli episodi rivoltosi. Per lo shock emotivo alla docile Cleo si rompono le acque, ma arrivata in ospedale solo ore dopo, a causa del traffico scaturito dagli eventi appena occorsi, dà alla luce una bimba senza vita. Il film volge lentamente verso una conclusione

catartica sia per Cleo che per Sofia, attraverso questo sguardo intimo ma pur sempre distaccato della camera, assistiamo attraverso un semplice viaggio al mare ad un processo di purificazione delle due donne che riescono a liberarsi in modi diversi del fardello che le opprimeva interiormente.

Il progetto sonoro di *Roma* ruota tutto attorno all'idea di offrire uno sguardo, come lo definisce Skip Lievsay, da "observational documentary". L'approccio è quello di trasportare gradualmente lo spettatore a contatto con l'intimità della storia raccontata. Ecco perché l'utilizzo del sonoro è stato anche qui pensato affinché restituisse una fedele spazializzazione di ciascun elemento presente sulla scena, tutto si muove in conformità all'immagine. L'idea è che il suono rappresenti solidamente ed oggettivamente l'ambientazione all'interno della quale lo spettatore venga gradualmente portato ad immergersi nel frame visivo. Questo espediente torna ad essere ampiamente utilizzato anche per i dialoghi che seguono attraverso il panning di missaggio la posizione esatta che i rispettivi personaggi occupano sulla scena.

Cuarón a tal proposito racconta, in un intervento in merito al suono di *Roma*, di un sound mixer che dopo la visione del film *Gravity* gli significò di trovarsi in disaccordo con questa estrema spazializzazione dialogica che conduce lo spettatore a distrarsi voltandosi ogni volta che le voci compaiono. Il regista riporta di aver risposto, un po' provocatoriamente, che comprendeva il punto di vista e che poteva anche capitare la prima volta, ma che lui crede ("trust") nelle "contemporary audiences" e che francamente chi si volta per le voci spazializzate gli pare uno spettatore alla stregua di chi alla fine del film va a sollevare lo schermo per controllare se ci sono gli attori dietro. [13]

Attraverso tale aneddoto viene dunque ribadito l'interesse dell'autore ad interpretare il cinema come la sperimentazione di un linguaggio di cui il

regista è responsabile, così come è responsabile del fatto che il pubblico possa rilevarne le intenzioni, una premura che Skip Lievsay racconta essere stata molto forte in lui nel corso della realizzazione di questo particolare film. Essendo infatti un racconto autobiografico (pare che uno dei ragazzini del quadro familiare sullo sfondo rappresenti proprio Alfonso Cuarón) egli aveva quale obiettivo principale quello di essere in grado di ricreare attraverso il suono esattamente tutte le suggestioni rimaste impresse nella sua memoria di bambino di quegli anni, affinché lo spettatore potesse vivere tali suggestioni fatte di ambienti, voci, musiche diegetiche, chiacchiericci, traffico e folle di Città del Messico.

Non a caso la magia di questo film è rappresentata dalla nitida e avvolgente densità sonora delle sue ambientazioni che realizzare ha richiesto un grande lavoro di presa diretta, un processo verificatosi essere altamente articolato, ricordando su più ampia scala quello che Altman, quale pioniere nell'utilizzo del multitraccia, fece agli inizi nel film *Nashville*. Il responsabile del mix di presa diretta infatti afferma di non aver mai gestito in vita sua una tale mole di informazioni, ciascuno era dotato di microfono in scena, i quali venivano sparsi ovunque poiché l'obiettivo era riprendere separatamente qualsiasi elemento che in un qualche modo emergesse e quando non era possibile farlo sono state organizzate delle sessioni di doppiaggio addirittura delle comparse. [13]

Tutto nella scena a livello sonoro è stato ricostruito puntualmente e tutto a seconda del movimento di camera è stato opportunamente spazializzato generando la sensazione di un'estrema nitidezza delle atmosfere cara a Cuarón per poter avvolgere lo spettatore del profumo di quei ricordi.

Skip Livesay, Sound Editor Supervisor e Re-Recording Mixer del progetto, definisce l'utilizzo della camera in questo film "Voyeristico" il che ci ricorda la riflessione precedentemene riportata di Michel Chion, immagini

intime cui solitamente non avremmo accesso, accompagnate da un sonoro avvolgente (nel caso di Chion tuttavia ci si riferisce ad una dimensione sonora ben più invasiva associata alla pervasività dei concerti rock). Non ci sono tagli veloci di campi e controcampi, questi morbidi pan a 360 gradi della camera consentono di guardare senza essere visti, mentre i suoni della scena si articolano nello spazio radicando lo spettatore nell'ambiente. Suoni ed immagini si fondono mantenendo le loro specificità in un compromesso tra sguardo voyeristico ed esperienza immersiva e quando le immagini dai morbidi e lunghi pan che indugiano sulla scena cambiano, gli stacchi sono volutamente nitidi per sottolineare il nuovo punto di vista.

A ribadire questo impianto narrativo vi è la scelta di non inserire musiche extradiegetiche nel film che commentano ed inclinano la percezione dello spettatore, il racconto sonoro rimane costantemente solidale con gli eventi e gli unici contributi musicali sono quelli direttamente presenti sulla scena, ogni stanza infatti è come se emettesse il proprio suono, tra feste, radio e ambienti. Lo spettatore è così libero di entrare ed indagare il racconto a proprio modo e la narrazione ne guadagna in una suspense generale poiché nessun commento musicale anticipa le prossime mosse del film. Inoltre questa scelta ci sottolinea la propensione del regista a sperimentare il linguaggio audiovisivo passando da un film come *Gravity*, in cui per raccontare gli eventi dei personaggi si alternano voci radiofoniche, silenzio e musica, ad un film in cui sono solo le sonorità delle strade, delle persone, degli ambienti, i cosiddetti effetti, che costruiscono l'atmosfera della storia.

Un'atmosfera sonora volutamente ricca di definizione prospettica che si articola tutta intorno alla protagonista per contro docile e silenziosa, le cui emozioni rimangono per lo più sempre sottaciute, ma è attraverso la purezza di quel silenzio avvolto dall'umanità dei profumi sonori evocati dal film che lo spettatore arriva direttamente a connettersi con le sue emozioni.



Figura 9. Scena tratta dal film *Roma* (2018) di Alfonso Cuarón in cui il fidanzato di Cleo è appena sparito dopo l'annuncio della sua gravidanza. L'immagine è riportata quale esempio di come il regista spesso mantenga la protagonista in primo piano la quale, silenziosa, cela allo spettatore il fragore delle sue emozioni mentre attorno a sé si articolano suoni di cui è possibile percepire l'esatta distanza prospettica e percezione di prossimità, dal venditore ambulante accanto a lei ad i passi delle signore in background. Suoni, seppur diegetici, spesso riflesso per similitudine o contrasto della dinamica dei suoi sentimenti.

Tutto questo ci evidenzia un passaggio fondamentale da *Gravity* a *Roma* lo sfruttamento dell'Atmos grazie alla sensibilità del regista si è tramutato da imponente strumento per trasferire tensione emotiva a mezzo per raccontare una storia più intima.

Durante un'intervista condotta dal Glenn Kiser in merito al *making of* sonoro di *Roma*, Alfonso Cuarón dichiara che lui comprese subito, a dispetto di quello che si potrebbe pensare, che l'Atmos era uno strumento che si prestava non tanto, o per lo meno non solo, per rinforzare l'impatto delle immagini quanto per restituire uno sguardo più intimo. Inoltre specifica che a suo parere questo strumento è utilizzato "molto male" nella maggior parte dei film poiché ne viene sfruttato essenzialmente il volume, "loudness". Benché infatti un sistema come questo sia efficace nell'incremento de volumi, secondo il suo punto di vista la vera potenzialità offerta è rappresentata dallo "sperad", ovvero la possibilità di lavorare sulla diffusività dei suoni. [13]

Il regista in un'altra intervista condotta da *Soundworks Collection* afferma che, per la sua capacità di separare perfettamente profondità, volumi e direzionare i contributi sonori a ciascun altoparlante di sala, "Dolby Atmos is the system I always dreamd of". In merito a questa osservazione che potrebbe sembrare molto tecnica e quasi forse propagandistica, Cuarón prosegue con una conclusione che non solo contestualizza la precedente affermazione ma a mio avviso racchiude un po' il senso dell'intero scopo di questa tesi:

"What inspires to me is that cinematic experience is a language itself. A language that use tools and sound is one of this tools. When you have the possibility to have sound coming out from every single different place, replicates the experience of just being alive, **used to be there in the moment, in the cinematic moment that you are presenting on the screen**". [12]

Il cinema inteso come linguaggio di cui il regista è maestro di sintassi, grammatico-emozionale, e in cui l'utilizzo di una tecnologia diventa campo d'esplorazione per espandere i termini creativi di un progetto, i confini emozionali dei personaggi e in questo caso parafrasando "l'eterno presente" di Jullier farlo divenire una maniera in cui autori e spettatori partecipino insieme dello stesso momento, dello stesso "cinematic moment".

1.3.2 Dalla realtà virtuale di Jullier alle "protesi del presente" qual è il futuro dell'immersività?

L'Atmos nasce inizialmente per diventare supporto sonoro del cinema 3D che ha avuto vita storicamente controversa e che ora pare un sistema di visione morente. Ma mentre il 3D con i suoi occhialini fa la sua definitiva (o

apparente) uscita di scena, i sistemi tridimensionali sonori, anche se più giovani d'età, lentamente resistono al trascorrere del tempo. (Citando uno fra tutti i sistemi sul mercato, le nuove sale in Dolby Atmos da 300 nel 2013 sono passate a 4000 nel 2019.)

L'interesse principale di questa tesi è quello di analizzare le caratteristiche di questa tecnologia quale proiezione sonora di immagini bidimensionali, ovvero coerentemente a quella che è la concezione comunemente condivisa di esperienza cinematografica. Tuttavia è più che pertinente pensare che sistemi sonori di questo genere non possano che essere il futuro, se non già il presente, di sistemi d'intrattenimento visivo ben più immerisvi ed interattivi di un "semplice" film, uno fra tutti la realtà virtuale con i suoi visori a 360°. In questo caso infatti, l'idea che il suono con il suo essere avvolgente guidi le "immagini" che il soggetto deve esplorare è intrinseco al concetto di immersività ed utilità di un contesto virtuale.

Jullier aveva predetto che i film-concerto con la loro propensione a non consentire allo spettatore di poter sfuggire dai suoni del film, poiché anche tappandosi le orecchie non vi sarebbe stato scampo dai subwoofers, sarebbero stati preludio di "quell'inferno virtuale":

"Nel vero universo virtuale non ci sono schermi; si sale di un'altra tacca nell'individualismo, perché solo chi è fornito di protesi prova qualcosa (nel finale di *Guerre Stellari*, certo gli spettatori "pilotavano" un'astronave individuale, ma la pilotavano tutti insieme, avevano tutti nello stesso tempo la medesima immagine davanti agli occhi). Poiché tutte queste protesi sopprimono ogni informazione percettiva proveniente dalla realtà del soggetto, si può dire che questo soggetto è totalmente immerso in questo "nuovo mondo": grado supplementare, di nuovo, del bagno di sensazioni, poiché è impossibile sfuggirvi (..)"

Ragionando sulle proprietà immersive dei sistemi sonori tridimensionali che mirano a catapultare lo spettatore nell'intreccio della storia al fine di recepire con sempre più verosimiglianza le sensazioni dei personaggi, risulta altrettanto attuale una considerazione fatta da Jullier in merito ad alcune esperienze ludiche presenti nei parchi divertimento Disneyland, che spesso egli associa alla tipologia d'intrattenimento offerta dai film-concerto "Il corpo dello spettatore e del personaggio sono soggetti alle stesse forze di quelle del personaggio: sempre più dunque la distanza tra i due si assottiglia e lo stadio successivo non può che consacrare la scomparsa del personaggio."

Ma se a scomparire saranno i personaggi, il prossimo passaggio sarà la scomparsa del cinema? A favore di uno spettatore che si tramuta in un esploratore di un mondo virtuale in uno scenario in cui: "Realtà virtuale e immagini di sintesi suonano dunque le campane a morto dell'immagine in termini di opposizione tra campo e fuori-campo."?

E' questa una prospettiva che i sistemi sonori surround nella loro evoluzione immersiva da un punto di vista strutturale preconizzano perfettamente. La dimensione sonora si libera dai confini dello schermo, si muove indipendentemente e costruisce uno spazio quasi illimitato verso il quale l'immagine contrae debito nel concetto che ben sintetizza Giulio Latini quale "ideologia a prospettica del suono", ove lo schermo perde la sua centralità storica. [11]

Non è un caso se la teoria del cinema odierna secondo l'*infrastrutturalismo* precedentemente citato, ibrida alla filosofia e alle neuroscienze anche il campo della mediologia, in una concezione di cinema che per essere compreso necessita di interrogarsi sul senso stesso di schermo e di dispositivo.

Laurent Jullier in questo senso era stato ancora una volta quasi profetico e, appoggiandosi al lavoro di Jonahan Steuer, dà la seguente

definizione di Realtà Virtuale “(..) quando la percezione è mediata da una tecnica di comunicazione, afferma Steuer, si percepiscono due ambienti, quello che ci circonda fisicamente e l’ambiente presentato tramite i media, che si può esprimere sottoforma di tracce (Steuer usa un’altra parola) o di segni di sintesi; in queste condizioni la realtà virtuale è un ambiente in seno al quale il soggetto sperimenta la teletrasparenza, che qualifica la predominanza del secondo ambiente sul primo.” Proseguendo inoltre dà la seguente definizione di engagement: “Più in un dispositivo è elevata la teletrasparenza, più il soggetto ha la sensazione di esistere in un mondo diverso”.

Letta da un altro punto di vista questa potrebbe essere una definizione di quella che è divenuta a tratti protesi e a tratti estensione del nostro corpo e che in quest’epoca storica racchiude in sé molteplici fenomeni dell’esperienza umana che ci portano letteralmente a vivere in un “mondo diverso”, mi riferisco allo smartphone.

lo smartphone si può in effetti considerare come un strumento che attraverso sé stesso filtra la nostra percezione dell’ambiente, sia esso forma di socialità o di intrattenimento, che estende ma spesso prevale su quello reale che ci circonda. In un qualche modo esso pur non richiedendoci vere e proprie protesi per il suo utilizzo, per la sua capacità di proiettarci in altre realtà virtuali si connota di una certa teletrasparenza.

“Infatti il segno più rilevante del mutamento in corso riguardo agli equilibri tra contenuto e comunicazione – mutamento che il sociologo americano John Perry Barlow anni fa ha comparato alla scoperta del fuoco nella storia della civiltà – è nel fatto che ora la comunicazione non è più un medium simile a una protesi che aumenta la funzionalità dei nostri sensi permettendoci di vedere o di sentire più lontano, ma è divenuto un ambiente

totale, globale, collettivo, un'atmosfera che non si può non respirare, neanche da parte di chi si illude di sottrarvisi, appunto un'infosfera¹⁶.” [14]

Che sia preludio o palestra del cyerspazio verso il quale ci muoviamo è difficile, o facile secondo i punti di vista dirlo, ma in questo ambito di studio risulta interessante sottoporre tali riflessioni perché con l'ampia diffusione delle piattaforme di streaming video, gli smartphone sono diventati anche strumento di accesso all'esperienza cinematografica, se non quella propriamente detta senz'altro quella di Netflix.

Dal Cinemascope, al Superfield di Chion fino ai 13x7 pollici di schermo antigraffio.

Pertanto non è assolutamente un caso se un colosso come la Dolby che lavora nel settore del cinema, che oltre che arte va sempre ricordato essere parte di un meccanismo industriale, abbia ideato una funzione per certi modelli di smartphone che abilita l'ascolto dell'Atmos sul proprio telefono. Anche i PC di ultima generazione sono stati studiati attraverso “magie di fasi e controfasi” in riproduzione per restituire questo effetto di percezione

¹⁶ Per infosfera si intende “L'insieme dei mezzi di comunicazione e delle informazioni che da tali mezzi vengono prodotte.” (Treccani) Un termine recentemente ripreso da Luciano Floridi, filosofo “digitale” contemporaneo, nella sua opera *Quarta rivoluzione. Come l'infosfera sta trasformando il mondo* (Cortina, Milano 2017). In cui amplia i confini delle accezioni assunte da questo termine, “Dopo le precedenti grandi svolte della storia e della scienza moderna, ossia le tre rivoluzioni antropologiche, la copernicana, la darwiniana e la psicanalitica, entra in scena una rivoluzione informatica che riesce a mutare le coordinate globali della stessa democrazia, oltre che della cultura.” Contenuto e comunicazione vedono i loro confini sfumarsi così come si dissolvono le barriere tra vita online e offline, la comunicazione non necessita più di un medium quale protesi, la comunicazione è diventata l'ambiente che abitiamo senza distinzione tra realtà e realtà virtuale. “Si delinea, così, nell'odierna comunicazione non più un'«estensione di noi stessi», come intendeva McLuhan” (The Extension of Man era il sottotitolo del suo saggio *Understanding Media* del 1964) “ma il trapasso a una nuova «condizione umana», a un inedito modello antropologico i cui tratti sono comandati da questa realtà onnicomprensiva della quale Internet ne è il vessillo imperante.” [14]

immersiva e recentemente Netflix ha introdotto sulla sua piattaforma, per gli utenti possessori di abbonamento premium, la fruizione dell'Atmos per i titoli presenti in catalogo girati in tale formato. [15]

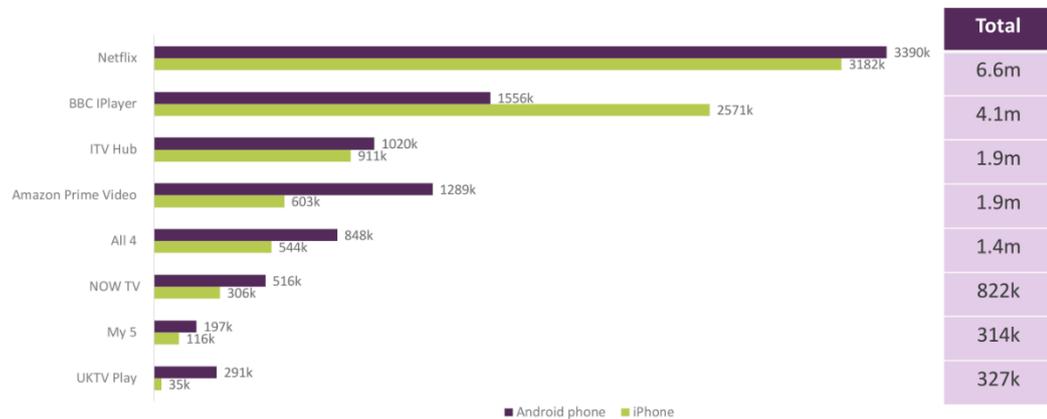


Figura 10 Fonte Ofcom. Il grafico illustra l'accesso degli utenti unici maggiorenni in UK mediante smartphone alle principali piattaforme di video streaming, datato Marzo 2019. Considerando che Netflix nel 2019 ha totalizzato più di 12 milioni di iscritti in UK, secondo questo grafico si registra circa un 50% di accessi al mese mediante telefono. [15]

Le grandi aziende del sonoro cinematografico infatti, in coda alla Dolby, consce che la distribuzione in sala di un film oggi non è che una delle brevi tappe che lo attende, sfruttando il sistema object-based (di cui si rimanda più avanti per la trattazione) riesumano la tecnica del binaurale, vecchio metodo di registrazione e codifica dell'audio oggi più che mai attuale.

Il binaurale è infatti un sistema di registrazione o sintesi del suono che consente di restituire le caratteristiche spaziali di quest'ultimo replicando la rispettiva facoltà umana di ricezione. Dunque oggi da un solo mix si può facilmente estrarre anche la versione binaurale di quest'ultimo e per ascoltarlo sono sufficienti un paio di cuffie, non sarà come avere i subwoofer a filo pelle e una semisfera di altoparlanti sulla testa, ma nel buio delle proprie stanze, seduti viaggiando in treno, sorvolando in aereo, aspettando

il pullman alla fermata, la tecnologia trova il modo di portare le bolle immersive sonore ad un altro gradino della “scala evolutiva”.

Il cinema non combatte più per detenere il monopolio dell’esperienza immersiva ma capisce che oggi l’obiettivo è quello di moltiplicare la portabilità dell’esperienza, un passo ancora oltre rispetto all’home theatre.

Il che se uno riflette non è neanche troppo un capriccio della tecnologia. In un mondo in cui circa 152 milioni di persone hanno un abbonamento Netflix e 75 milioni (non è dato sapere dal grafico se utenti singoli o sovrapposti [16]) hanno

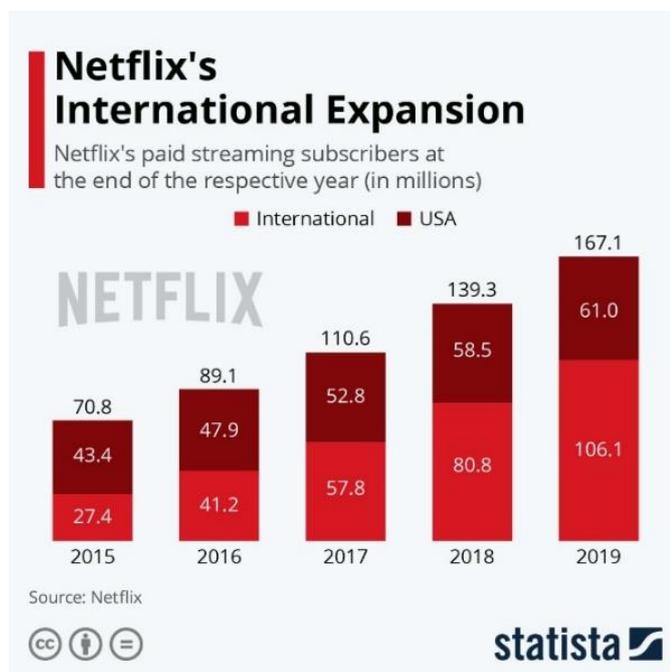
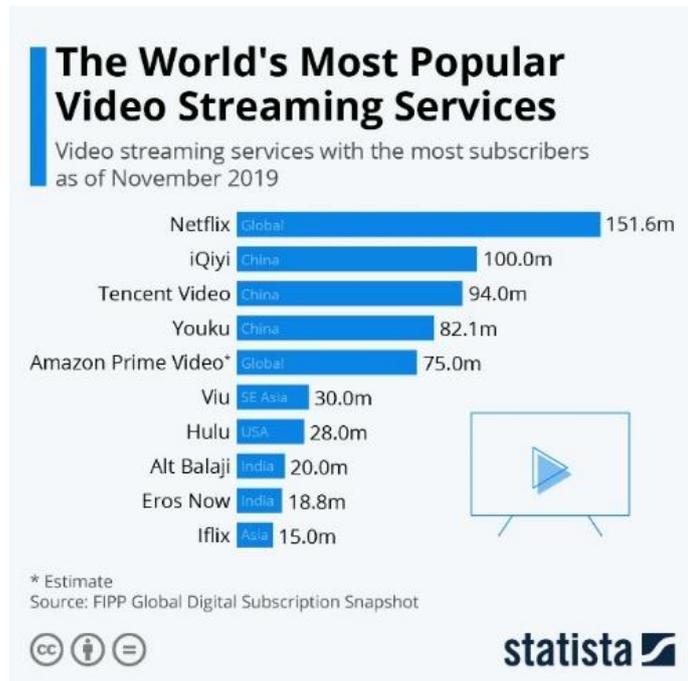


Figura 11 In azzurro vengono illustrate le piattaforme di video streaming per ordine crescente di iscritti a pagamento al mese di Novembre 2019. (N.B. Solo i dati riferiti a Netflix e Amazon Prime sono globali). In rosso la crescita costante degli utenti globali Netflix dal 2015 al 2019. Fonte Statista [16]

un abbonamento Amazon Prime Video e ancora Netflix stesso, detentore del mercato al momento, registra un aumento costante delle iscrizioni annuali, l'idea di espandere i confini dello schermo, non quello cinematografico, ma quello ben più ridotto dello smartphone, attraverso la dimensione sonora potrebbe divenire supporto essenziale di quella che è divenuta la generale propensione di consumo di contenuti audiovisivi.

Il fatto che si stia vivendo un periodo storico che vada verso un maggior "isolazionismo" è abbastanza evidente e certamente la società che si riconfigurerà dopo i recenti avvenimenti causati dal Covid non farà che accentuare tale propensione. Già solamente il fatto che per via della pandemia in corso i cinema, quale luogo di aggregazione sociale, abbiano dovuto chiudere, ha favorito una certa crescita del consumo mediale mediante piattaforme streaming e la stima è che questa tendenza vada aumentando. I produttori cinematografici stessi non potendo in questo periodo storico far uscire i propri film nelle sale si avvalgono dei servizi online per gestire una minima distribuzione dei loro contenuti. Pare infatti che Netflix abbia guadagnato qualcosa come 16 milioni di nuove iscrizioni nei primi tre mesi del 2020 e altri circa 10 milioni tra Aprile e Giugno a causa del lockdown. [17] Il mondo tristemente o realisticamente va verso una contrazione della socialità e un incremento di esperienze d'intrattenimento individuali, il pubblico è già abituato a trascorrere ore sul proprio smartphone o laptop, trasferire attraverso di essi anche l'esperienza cinematografica è qualcosa di ormai connaturato agli eventi.

Tuttavia questi dati potrebbero confortare in merito al fatto che pare che il pubblico continui ad avere "fame di film", infatti piattaforme come Netflix si trovano ad affrontare il tema che maggiore è il numero di iscritti maggiore è la pressione nella costante produzione di contenuti audiovisivi di

qualità che preservino la fedeltà degli utenti nel tempo. In merito a questo Netflix non a caso da iniziale collettore di formati si è trasformato anche in produttore stesso di prodotti originali.

Dunque riprendendo le parole del Prof. Giulio Latini in introduzione: “queste tecnologie **di quale il mondo più ampio stanno partecipando?** (..) Che spettatore è quello odierno che chiede l’Atmos piuttosto che altri sistemi. Che uomo è? Che donna è? Che bambino è?”

Le riflessioni sopra riportate non sono certamente sufficienti a dare una risposta esaustiva a queste domande, né tantomeno è l’intento di questa tesi, ma a mio avviso in questo panorama resta vivida la considerazione di Cuarón, il cinema è un linguaggio e gli autori devono servirsi dei suoi strumenti per poter declinare al meglio la narrazione. Il regista lo dice “I trust contemporary audiences” il pubblico ha dimostrato nel tempo di saper riconoscere la qualità dalla sua forma al contenuto e così le tecnologie immersive, come qualsiasi altra tecnologia, devono continuare ad essere terreno di sperimentazione per l’ampliamento dei confini narrativi, interpretazione della contemporaneità e certo, pur attraverso i suoni connotati, fonte di “un pensiero aumentato ancorché di una sensazione aumentata”. [11]

Tuttavia ad oggi produrre film in Atmos risulta ancora fortemente costoso e laborioso, specie per il mercato audiovisivo italiano, e il riscontro di pubblico non è stato ancora sufficiente per elevarlo, se mai verrà elevato, a standard di realizzazione del sonoro filmico. Forse farà la fine del cinema 3D, forse troverà il modo di essere assimilato in maniera sensibile ed intelligente così come ha fatto Cuarón nei processi linguistici di realizzazione del film o forse è solo il trampolino di lancio, il preludio verso la tanto preannunciata realtà virtuale verso la quale ci stiamo muovendo.

2. Sistemi Audio Immersivi dal multicanale all'approccio object-based

In questo capitolo verranno analizzate le peculiarità di un Sistema Sonoro Immersivo percorrendo i principali documenti che costituiscono lo standard SMPTE 2098 in materia di *Immersive Audio Metadata and Bitstream Specification*.

L'intento infatti sarà quello in prima istanza di comprendere quali sono le implementazioni tecnologiche che caratterizzano un Sistema Immerisvo rispetto ai tradizionali sistemi di diffusione surround e in seconda istanza approfondirne il formato bitstream, preziosa chiave di volta per la comprensione del potenziale racchiuso nei nuovi Sistemi Immersivi oltre che di innovatività per le esigenze del mercato audiovisivo.

Il merito della SMPTE, sulla base di alcune richieste avanzate da organizzazioni rappresentanti gli esercenti e le major di Hollywood come la NATO (National Organisation of Theater Owners) UNIC e DCI (Digital Cinema Initiative), è stato quello di dare vita ad un unico standard aperto per il suono immersivo che potesse essere poi utilizzato come formato di distribuzione per tutti i sistemi Sonori Immersivi in circolazione.

L'obiettivo era infatti quello di creare un formato audio fruibile indipendentemente dalle difformità di dotazione tecnologia, a patto che in linea con le specifiche, al fine di favorirne l'adozione e dunque abbattere le ritrosie, specie degli esercenti, verso questo nuovo sistema sonoro.

Il capitolo partirà dunque dall'analisi delle tradizionali configurazioni di sala, all'esposizione dei concetti di base di un sistema audio immersivo, fino all'analisi del suo Immersive Audio Bitstream (IAB) come standardizzato dai documenti SMPTE 2098 per poi vedere i potenziali

distributivi abilitati dal formato stesso di questa nuova tecnologia.

E' infatti parere di chi scrive che affrontare lo studio dei sistemi immersivi partendo dalle sue "molecole" costitutive quali gli Oggetti, i Bed e i Metadati sia il modo migliore per poter avvicinare i creativi alla comprensione delle intrinseche opportunità dello strumento ed al contempo comprendere le funzionalità abilitate da questa tecnologia che ha saputo declinare il potenziale offertogli dall'ingresso del digitale nel cinema.

2.1 Configurazioni Multicanale Tradizionali

Si definisce **Soundfield**, campo sonoro, lo spazio acustico creato riproducendo contemporaneamente una o più sorgenti audio.

Quando si è spettatori di un film al cinema si è letteralmente immersi in un "campo sonoro" generato dai diversi altoparlanti che costituiscono la sala cinematografica e che occupano un perimetro di sala stabilito dalla rispettiva **Soundfield Configuration** ovvero dalla peculiare disposizione degli altoparlanti per quella sala.

Negli standard internazionali che descrivono le configurazioni degli altoparlanti stereo come ITU-R BS.775-3 (2012) [18], la nomenclatura di configurazione è spesso nella forma "n-m stereo", dove n è il numero di canali anteriori e m è il numero di canali posteriori o laterali (quest'ultimo si incontra solo nei sistemi surround). Questa distinzione infatti è utile perché sottolinea il ruolo diverso dei canali surround rispetto a quelli frontali.

Sempre in termini di nomenclatura un'altra formula comunemente utilizzata è quella che indica il numero totale di canali di una data configurazione seguiti dal numero di canali dedicati agli effetti a bassa frequenza (LFE) (esempio 5.1 surround).

Tuttavia rispetto ai sistemi immersivi nelle configurazioni surround classiche l'obiettivo non era quello di restituire un'immagine sonora a 360° con la stessa precisione in tutte le direzioni, questo perché i canali frontali sinistro e destro si trovano spesso in posizioni coerenti con le configurazioni stereo a due canali, il che rende gli altoparlanti dello schermo abbastanza vicini tra loro, lasciando un gap rispetto agli altoparlanti surround laterali.

Non a caso la nuova configurazione immersiva Dolby Atmos ha inserito anche due “semi array” di altoparlanti che congiungessero i diffusori di una configurazione surround “classica” con lo schermo.

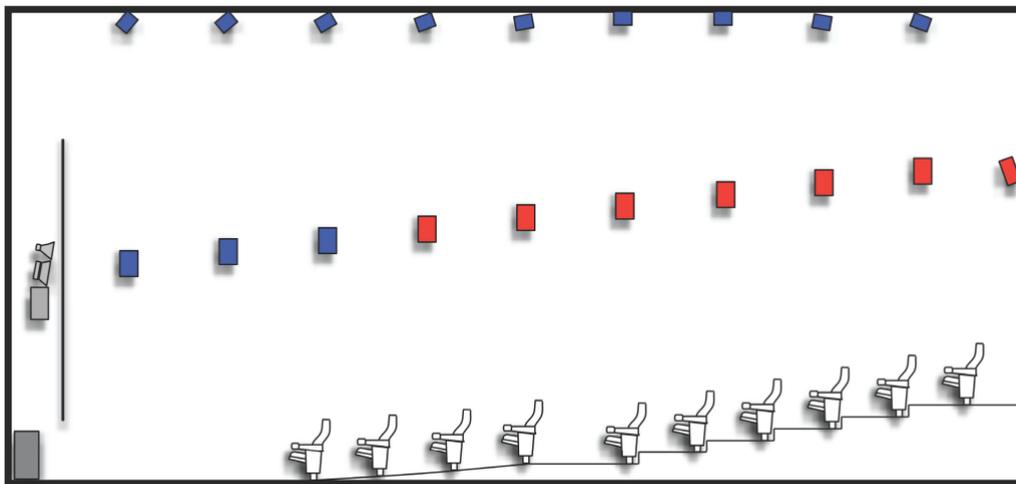


Figura 12 Sezione laterale di una sala cinematografica in Dolby Atmos in cui in rosso sono indicati gli altoparlanti di una configurazione classica surround, mentre in blu gli altoparlanti aggiunti per rendere immersiva la configurazione di sala. I tre altoparlanti laterali blu in continuazione a quelli rossi rappresentano un'implementazione dei diffusori al fine di consentire un panning dei suoni più fluido dallo schermo alla sala e viceversa. [27]

Ad oggi la maggior parte delle configurazioni sonore che caratterizzano le sale sono del tipo surround 5.1 e 7.1, il cui suono generato dalla rispettiva configurazione bidimensionale di altoparlanti si attesta su un piano orizzontale chiamato "Base Layer"¹⁷.

¹⁷ Si riferisce al piano orizzontale nominale degli altoparlanti utilizzati nei sistemi surround 5.1 e 7.1, definito nello standard SMPTE 2098-5.

Tali configurazioni si trovano ad essere la struttura di base su cui vengono poi implementati i sistemi immersivi al fine di ricostruire uno spazio sonoro tridimensionale. Ecco perché a seguire verrà proposta una panoramica delle più tradizionali "Soundfield Configuration" il cui campo sonoro, come è stato detto, si attesta esclusivamente sul piano orizzontale.

2.1.1 Monofonia (C)

Nei sistemi monofonici tutti gli impulsi vengono registrati su un'unica traccia e successivamente riprodotti attraverso un unico altoparlante solitamente posizionato centralmente (Central) dietro lo schermo. Attraverso un sistema monofonico è molto difficile restituire tutte le informazioni direzionali, infatti durante la registrazione mediante singolo microfono, i suoni provenienti da diverse direzioni vengono catturati e fusi insieme in un unico segnale. Tuttavia un sistema monofonico è in grado di restituire un discreto senso della profondità. Variando infatti i livelli d'intensità di un suono è possibile replicare la sensazione che esso si trovi più o meno prossimo all'ascoltatore.

In ogni caso i suoni riprodotti da un sistema monofonico creano un'immagine sonora percettivamente ancorata all'altoparlante che li diffonde e dunque restituiscono un effetto *speaker-centric*.

2.1.2 Stereofonia (L, R)

Un primo antesignano esperimento di suono stereofonico fu sperimentato nel 1940 con il FANTASOUND progettato appositamente per la riproduzione del film Disney *Fantasia*¹⁸, tuttavia esso fu considerato un esperimento

¹⁸ Per il film *Fantasia* fu progettato per la prima volta un sistema stereo surround chiamato FANTASOUND che utilizzava tre canali a banda ottica stampati su pellicola dedicata

troppo costoso e poco redditizio per cui si dovrà aspettare fino agli anni '50 prima che i sistemi stereofonici inizino ad essere adottati su larga scala.

Un sistema stereofonico è dunque caratterizzato da due altoparlanti posizionati dietro lo schermo rispettivamente alla destra ed alla sinistra (Left, Right) dell'ascoltatore in grado di organizzare i suoni non solo su un asse verticale ma di dislocarli anche su un asse orizzontale.

Ciascuno dei due altoparlanti di un sistema stereofonico, diversamente dai sistemi monofonici, sono alimentati da due canali distinti l'uno dall'altro. Dunque due flussi informativi diversi destinati ai due diffusori, destro e sinistro, con l'intento di ricostruire nella percezione dell'ascoltatore l'esatta distribuzione nello spazio delle sorgenti che contribuiscono a creare un evento sonoro. All'interno del *range* descritto dai due altoparlanti un sound designer può localizzare qualsiasi suono nello spazio, poiché la stereofonia aggiunge una nuova dimensione alla riproduzione, permettendo di posizionare le sorgenti nelle varie direzioni planari (destra, sinistra, centro).

Se un ascoltatore si trova nel cosiddetto *sweet spot*¹⁹ e se gli altoparlanti riproducono la stessa traccia alla medesima intensità la sensazione recepita sarà che il suono si trovi esattamente nel centro dello schermo. Se per caso vi è invece uno sbilanciamento nel volume tra i due canali è possibile dar luogo ad un effetto di dinamizzazione spaziale fino al

esclusivamente al suono più una quarta traccia separata per eseguire il panning dei canali a ciascuno dei 10 gruppi di altoparlanti sparsi, 9 surround e in alcune sale addirittura 1 al soffitto. Si potrebbe definire il primo sistema sonoro 3D della storia. Tuttavia a causa degli elevati costi comportati dall'allestimento delle sale, questo genere di sistema verrà accantonato fino ai successivi sviluppi della stereofonia. [3] [21]

¹⁹ Anche definito in maniera più formale "punto di ascolto di riferimento", è l'area o punto ideale di ascolto in sala. Nel caso del suono surround è il punto focale tra quattro o più altoparlanti, cioè la posizione in cui tutti i fronti d'onda arrivano simultaneamente.

punto che se uno solo degli altoparlanti emette suono la localizzazione della fonte sarà completamente sbilanciata in quella direzione.

2.1.3 (L, C, R) 3.0 – (3 - 0 stereo)

Questa è la configurazione di altoparlanti da schermo tipicamente usata al cinema in cui, oltre ai due canali stereofonici ne viene aggiunto un terzo centrale monofonico, equidistante dagli altri due (Left, Center, Right).

L'aggiunta di un terzo canale comporta svariati vantaggi [19]. Primo fra tutti, consente di avere un suono fronte schermo più ampio perché il canale monofonico aggiuntivo “ancora” i suoni all'immagine centrale e consente così di allontanare la distanza relativa tra il canale destro e sinistro spostandoli maggiormente verso i lati dello schermo. Secondo, restituisce una più ampia gamma di posizioni d'ascolto in sala

Infine, l'immagine sonora emanata dal canale centrale non soffre di alterazioni timbriche come quella riprodotta dai due canali stereo, un fattore questo necessario quando devono essere riprodotti suoni ben *ancorati* con il centro dell'immagine. Si ricorda infatti che in stereofonia essendo il contenuto dei due canali uno diverso dall'altro, poiché tesi a ricostruire nella percezione dell'ascoltatore la distribuzione spaziale, il riconoscimento della provenienza di un suono esattamente nel centro dello schermo avviene solo se c'è una precisa simmetria d'ascolto rispetto alle casse, cosa che sarebbe impossibile da garantire contemporaneamente a tutti gli spettatori di un cinema.

2.1.4 Surround (L, C, R, S) - (3 – 1 stereo)

Nell'approccio *3-1* un ulteriore canale *Surround* (o canale degli *effetti*) viene aggiunto ai tre canali frontali, ovvero un singolo canale monofonico indirizzato a tutti gli altoparlanti che avvolgono dietro e lateralmente l'area degli spettatori.

L'intenzione fu quella di restituire agli spettatori un'esperienza avvolgente nel rapporto immagine-effetti e fu proposta dalla 20th Century Fox negli anni '50 per competere con i nuovi sistemi d'intrattenimento televisivo.

Tale canale surround (o canale degli effetti) non fu inteso per abilitare la spazializzazione dell'immagine a 360°, fattore questo per altro tecnicamente impossibile perché il sistema surround *3-1* è costituito di un singolo canale indirizzato ad un'elevata quantità di altoparlanti monofonici.

2.1.5 Surround 5.1 (L, C, R, Ls, Rs, LFE) – (3 – 2 stereo)

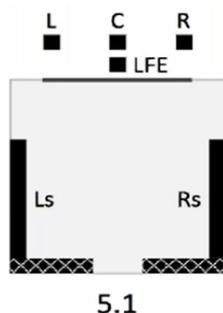


Figura 13 Soundfiled Configuration di un Sistema Surround 5.1 (SMPTE ST 428-12) [22]

La nomenclatura internazionale standard per un sistema 5.1 è “3-2-1” anche definito “3-2 stereo” volendo sottolineare che il sistema oltre i tre canali frontali è composto di due canali surround stereo rispetto al sistema

precedente ed un canale dedicato alle basse frequenze LFE. Quest'ultimo è un canale aggiunto non per le componenti a bassa frequenza dei canali principali ma un vero e proprio canale dedicato per accompagnare l'impatto di certi effetti visivi sullo schermo (esplosioni, ecc.).

La sala viene dunque divisa in due con i nuovi canali stereo Left surround e Right surround, con maggiori possibilità di restituire dinamizzazione spaziale.

2.1.6 Surround 7.1 (L, C, R, Lss, Lrs, Rss, Rrs, LFE)

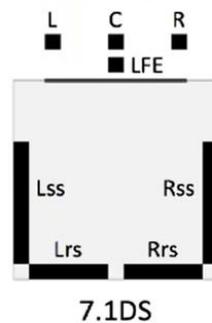


Figura 14 Soundfiled Configuration di un Sistema Surround 7.1 (SMPTE ST 428-12) [22]

Il sistema 7.1 DS è rappresentazione di una delle ultime configurazioni surround introdotte quale evoluzione del sistema 5.1. Questo formato prosegue con l'approccio di aumentare il numero dei canali surround dividendo quelli esistenti Left Surround, Right Surround in quattro "zone" (ovvero suddividendoli in Rear e Side Surround rispettivamente per il lato destro e sinistro), consentendo così di restituire sempre più dinamismo agli elementi sonori da spostare attraverso la sala.

Negli anni '90 venne messo a punto un precedente formato denominato 7.1 SDS, la cui configurazione tuttavia non si occupò di ampliare l'area surround ma semplicemente di aggiungere dietro lo schermo i canali LC (Left Center) e RC (Right Center), cioè due canali frontali destro e sinistro ma

leggermente più spostati verso il centro dello schermo.

Tutte queste configurazioni dunque rappresentano la base strutturale su cui viene implementato l'Audio Immersivo, che, come si vedrà in seguito, si avvale di tali canali classici per riprodurre gli elementi *Bed*.

2.2 Nozioni di base di un Sistema Audio Immersivo

L'ambiente di ascolto *reale* è lo spazio acustico che occupiamo durante le vita di tutti i giorni, in cui i suoni arrivano da tutte le direzioni possibili avvolgendoci in quella che è la realtà da noi percepita. L'obiettivo di un sistema audio immersivo è dunque quello di restituire uno spazio acustico in cui riproporre questa percezione di totale immersione durante l'esperienza di ascolto. Si tratta dunque di una modalità di riproduzione sonora capace di ricreare le proprietà spaziali della scena acustica considerata (posizione e orientamento delle sorgenti, forma e caratteristiche dell'ambiente, ecc...).

Mentre il suono surround infatti mira a circondare l'ascoltatore di sorgenti sonore posizionate solo sul piano orizzontale ed è strutturato sulla base della conoscenza a priori della posizione degli altoparlanti e della reciproca relazione rispetto all'ascoltatore; un sistema audio immersivo tridimensionale invece ha l'obiettivo di restituire all'ascoltatore la percezione di essere all'interno della scena:

This results in a “you are there”, first-person perspective, in contrast to the loudspeaker “they are here. [19]

Per fare questo, un sistema sonoro immersivo aggiunge la terza dimensione mancante ai sistemi di diffusione multicanale classici, quella

dell'altezza, emulando così la peculiarità del mondo reale in cui i suoni possono arrivare da tutte le direzioni e adotta una serie di tecniche in grado di restituire il dinamismo spaziale di una vera e propria esperienza d'ascolto 3D.

Generalmente i sistemi che consentono questa tipologia d'esperienza sono i seguenti [20]:

- Sistema Binaurale.
- Sistema di *Sintesi di Campi Sonori*.
- Sistema basato su Canali (discreti).

Il primo mira a replicare le caratteristiche dei segnali che vengono raccolti ed elaborati, nella fase d'ascolto, dal sistema uditivo umano. Si riferisce al suono a due canali che entra direttamente nell'orecchio destro e sinistro dell'ascoltatore e verrà analizzato nel paragrafo 2.6 di questo capitolo.

Il secondo è una rappresentazione fisica delle onde sonore che cerca di codificare un intero campo sonoro indipendentemente dalla disposizione degli altoparlanti prevista per una certa configurazione (*non-speaker-centric, channel-free*). Il termine *Campo Sonoro* si riferisce alla cattura, riproduzione e descrizione delle onde sonore, in contrasto con le tecniche binaurali e surround il cui obiettivo è ricreare eventi uditivi o oggetti sonori mediante fenomeni percettivi, creando “immagini virtuali” la cui percezione dipende dalla posizione occupata rispetto al sistema di diffusione. Attraverso l'approccio di sintesi di *Campi Sonori* per creare o riprodurre dei suoni vengono invece controllate le proprietà fisiche delle onde sonore

Il sistema oggetto d'interesse di questa tesi tuttavia è quello su cui si basano la maggior parte delle configurazioni di sala odierne, ovvero il terzo.

Più precisamente ci si occuperà di una forma ibrida tra un approccio basato su canali (*Channel-Based-Approach*), i cosiddetti Bed, ed uno basato su *Oggetti (Object-Based-Approach)*, di cui nel terzo capitolo si vedrà anche un approfondimento di natura applicativa. Una formula questa adottata dalla Dolby con la sua tecnologia Atmos seguita da una serie di altre importanti aziende sul mercato nell'ambito dei sistemi sonori immersivi.

Un simile sistema dunque, oltre ad introdurre la verticalizzazione del suono mediante altoparlanti in altezza, adotta singoli elementi audio (*oggetti*) accompagnati da set di metadati in grado di fornire le informazioni necessarie per renderizzare la posizione di un suono all'altezza di precisi diffusori all'interno di una *Soundfield Configuration* e di descriverne cambiamenti e caratteristiche che variano rapidamente anche nel corso dello stesso frame, fornendo così un ambiente acustico in cui far muovere i suoni quasi illimitato. Gli oggetti da cui vengono generati i segnali degli altoparlanti (o delle cuffie) sono principalmente le sorgenti sonore virtuali di cui è composta una scena, il che grazie ad un sistema di rendering adeguato e ai rispettivi set di metadati consente di riprodurre il mix sonoro con una certa flessibilità rispetto al *playback environment* di riferimento, preservando l'intento creativo del prodotto.

Da notare che non bisogna utilizzare oggetti per creare un mix immersivo e non è necessario avere canali per realizzare un audio immersivo, è possibile avere l'uno o l'altro o entrambi. L'Audio Immersivo combina la materia sonora con i metadati e consente agli oggetti audio e canali audio di essere *renderizzati* in una configurazione multipla di altoparlanti [21].

Nei prossimi paragrafi dunque si procederà con la spiegazione di ciascuno di questi elementi che rendono peculiare un Programma Audio Immersivo secondo l'approccio sopra citato.

2.2.1 Configurazione di una sala Immersiva – *Height/Top Speakers*

Con i nuovi sistemi immersivi al cinema arriva "la voce di Dio", così qualche tempo fa veniva pubblicizzata la peculiarità di diffondere il suono attraverso altoparlanti al soffitto, come nel caso di Auro 3D e Dolby Atmos.

I Sistemi Audio Immersivi infatti verticalizzano la dimensione del suono rispetto allo spettatore. Più precisamente posizionano le sorgenti sonore sia in corrispondenza che oltre il cosiddetto *base layer* cioè, come descritto in precedenza, il piano di diffusione ad altezza ascolto degli spettatori, formato tipicamente dagli altoparlanti di un sistema surround classico.

Pertanto tale verticalizzazione tradizionalmente viene restituita dai sistemi sul mercato attraverso due versioni aggiuntive di altoparlanti, definiti nello standard SMPTE ST 2098-5 [22].

- ***Height Layer***: si riferisce al livello degli altoparlanti posizionati subito sopra gli speaker del *base layer* di un sistema surround.
- ***Top Layer***: si riferisce al livello di altoparlanti posizionati sul soffitto che riproducono i suoni esattamente sopra le teste degli ascoltatori.

Come si può vedere dalla tabella seguente tratta dallo Standard SMPTE 2098-5 in materia di *D-Cinema Immersive Audio Channels and Soundfields Groups* le configurazioni in altezza, aggiuntive rispetto ad un *Soundfield* tradizione, per quanto riguarda il cosiddetto *Height Layer*, sono le stesse di un sistema surround classico 5.1 o 7.1 ma che godono di elevazione rispetto al piano d'ascolto classico (Lh, Rh, Ch, Lsh, Rsh, Lssh, Rssh, Lrsh, Rrsh). Oppure per quanto riguarda il Top Layer possono essere

di tre tipologie, due array da soffitto rispettivamente per il canale destro e quello sinistro (Lts, Rts) e un altoparlante sempre da soffitto centrale (Ts).

Name	Abbreviation	Description
Left Height	Lh	Intended to drive the Left Height Loudspeaker.
Right Height	Rh	Intended to drive the Right Height Loudspeaker.
Center Height	Ch	Intended to drive the Center Height Loudspeaker.
Left Surround Height	Lsh	Intended to drive the Left Surround Height Loudspeaker(s).
Right Surround Height	Rsh	Intended to drive the Right Surround Height Loudspeaker(s).
Left Side Surround Height	Lssh	Intended to drive the Left Side Surround Height Loudspeaker(s).
Right Side Surround Height	Rssh	Intended to drive the Right Side Surround Height Loudspeaker(s).
Left Rear Surround Height	Lrsh	Intended to drive the Left Rear Surround Height Loudspeaker(s).
Right Rear Surround Height	Rrsh	Intended to drive the Right Rear Surround Height Loudspeaker(s).
Left Top Surround	Lts	Intended to drive the Left Top Surround Loudspeaker(s).
Right Top Surround	Rts	Intended to drive the Right Top Surround Loudspeaker(s).
Top Surround	Ts	Intended to drive the Top Surround Loudspeaker(s).

Figura 15 Tabella riassuntiva dei Canali Audio Immersivi tratti dallo standard SMPTE ST 2098-5 [22]

Come poc'anzi espresso sono essenzialmente configurazioni surround di *base layer speakers* ai quali vengono poi aggiunti *height layer speakers* e/o *top layer speakers*:

- I *Left/Right/Center Height*: sono tre canali aggiuntivi posizionati dietro lo schermo sopra il livello del Base Layer.
- I *Left/Right Surround Height*: sono due canali aggiuntivi posizionati sopra il livello *Base Layer* in corrispondenza degli altoparlanti di una configurazione 5.1 surround.
- I *Left/Right Side Surround Height* e i *Left/Right Rear Surround Height*: sono quattro canali aggiuntivi posizionati sopra il livello *Base Layer* in corrispondenza degli altoparlanti di una configurazione 7.1

surround.

- I *Left/Right Top Surround* e il *Top Surround* sono invece le configurazioni che prevedono la diffusione del suono esattamente sopra le teste degli spettatori.

Al fine di poter illustrare come un sistema sonoro immersivo coniughi le passate configurazioni surround con i nuovi altoparlanti in altezza, nella figura 5 tratta dal webinar SMPTE in merito alla presentazione da parte di Brian Vessa del nuovo standard 2098 sull'Immersive Audio [22], è possibile osservare diverse configurazioni in cui a sinistra è riportata la piantina di una sala cinematografica relativamente alla posizione degli altoparlanti al livello del Base Layer, mentre a destra la medesima piantina coerentemente però agli altoparlanti al livello del Top ed Height Layer.

Name	Base Layer Loudspeakers	Height/Top Layer Loudspeakers
9.1OH		
9.1HT		
11.1HT		

Figura 16 Tabella raffigurante la sezione di un cinema con alcune possibili configurazioni Audio Immersive suddivisa illustrando a sinistra gli altoparlanti a livello Base Layer e a destra gli altoparlanti a livello Top/Height Layer. [22]

- L'esempio **9.1 OH** rappresenta una configurazione 7.1 per quanto riguarda il *Base Layer* che aggiunge poi due canali a soffitto restituendo così un totale di 9 canali più il canale per le basse frequenze LFE. E' questa la configurazione adottata proprio dal **Dolby Atmos**.
- La configurazione **9.1 HT** adotta invece un sistema 5.1 per quanto riguarda il Base Layer ed aggiunge 4 canali in altezza al livello del *Height Layer* per i due canali surround e per i due altoparlanti da schermo destro e sinistro.
- L' **11.1 HT** unisce ad una base 5.1 sei canali in altezza, di cui due *Height Surround*, un *Top Surround* e ancora i tre altoparlanti da schermo. E' questa, come vedremo più avanti, la configurazione adottata da **Auro 3D** (6+5+0).

ITU-R BS.2051–0 (ITU, 2014) [23]

Qui di seguito verranno analizzati alcuni dei principali sistemi sonori immersivi presenti sul mercato oggi. Per alcuni di essi potrebbe essere più immediato riferirsi ad un'altra nomenclatura rispetto a quella sopra proposta, ovvero quella della specifica ITU-R BS.2051–0 (ITU, 2014) per la configurazione di altoparlanti per la riproduzione audio multicanale.

La raccomandazione illustra infatti sia la configurazione di posizione che quella direzionale degli altoparlanti utilizzando tre livelli: superiore, medio e inferiore. Lo strato intermedio indica il piano orizzontale (vicino al livello delle orecchie degli ascoltatori) mentre gli strati superiore ed inferiore indicano rispettivamente il piano in altezza ed il piano del pavimento. La

raccomandazione denota ogni configurazione utilizzando il numero di altoparlanti in ogni strato, una configurazione "U + M + B" indica un layout di altoparlanti che utilizza U altoparlanti nello strato superiore, M altoparlanti nello strato intermedio e B altoparlanti nello strato inferiore. Tale rappresentazione tuttavia non distingue tra i due diversi livelli d'altezza visti in precedenza *Height Layer* e *Top Layer* definiti nello Standard SMPTE 2098-5.

Qui di seguito verranno dunque descritte le principali soluzioni in uso sul mercato attraverso i sopra citati metodi di classificazione di disposizione degli altoparlanti.

2.2.1.1 AURO 3D

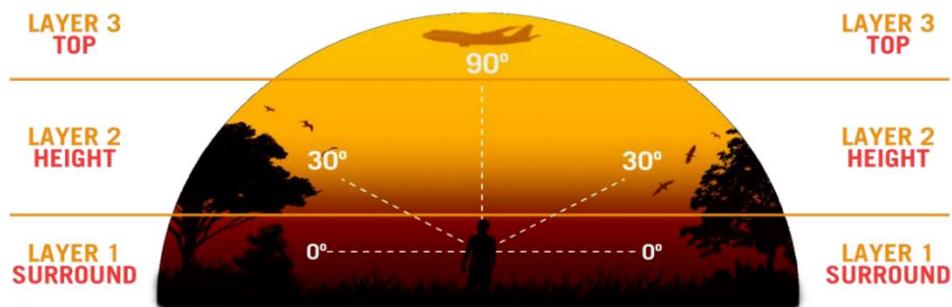


Figura 17 Modello 3D di una configurazione di sala 11.1 Auro 3D [24]

Auro 3D è stato il primo sistema sul mercato ad introdurre la dimensione della verticalità sonora. E' un sistema basato solo su canali che divide lo spazio in tre strati sintetizzabili come quelli visti in precedenza *Base*, *Height* e *Top Layers* [24].

Auro 3D infatti crea un campo sonoro spaziale aggiungendo uno strato di altezza attorno al pubblico sopra il tradizionale sistema audio surround bidimensionale. Come è possibile vedere dalle figura 18, la configurazione spaziale si innesta sulle configurazioni surround di base 5.1 e 7.1 tramutandole rispettivamente nei sistemi 11.1 e 13.1 attraverso l'aggiunta di quelli che sono i canali frontali in altezza *Left/Right/Center Height* (HL, HR, HC), i canali surround in altezza *Left/Right Surround Height* (HLS, HRS) o in alternativa i canali surround laterali e posteriori in altezza *Left/Right Side Surround Height* e i *Left/Right Rear Surround Height* (HLSS, HRSS, HLRS, HRRS) e in fine la cosiddetta *Voice of God* il canale *Top Surround* (T).

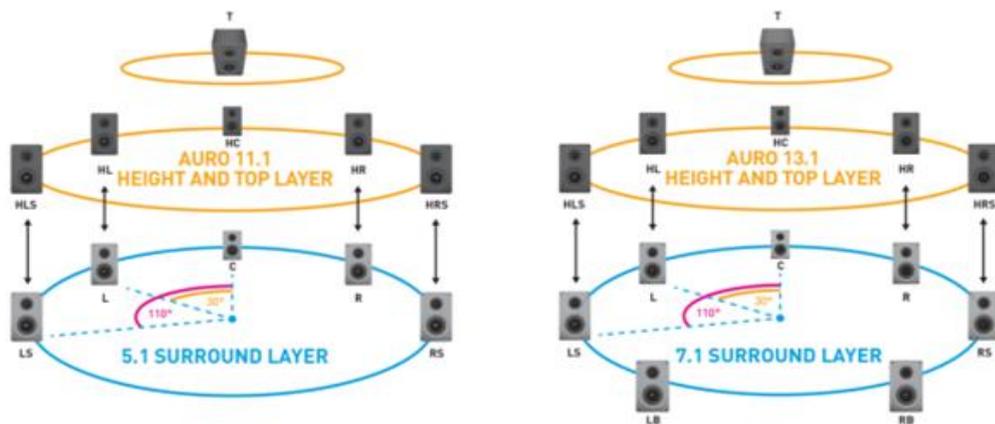


Figura 18 Configurazione Home Theatre dei Bed (canali) di un sistema Auro 3D rispettivamente 11.1 e 13.1. [24]

Come introdotto all'inizio di questo capitolo, sulla base del lavoro di standardizzazione promosso dalla SMPTE con l'obiettivo di creare uno standard aperto per l'*Immersive Sound* utilizzabile come formato di distribuzione per tutti i sistemi Immersivi, Auro 3D, che insieme alla Dolby e ad altri interlocutori ha contribuito attivamente alla scrittura di tali documenti, ha in risposta a tale richiesta dato alla luce il suo nuovo sistema **AuroMax**.

AuroMax infatti, rispetto al precedente sistema Auro 3D esclusivamente basato su canali, implementa le sue versioni ibridando l'approccio per canali con quello ad oggetti, approccio introdotto precedentemente dal Dolby Atmos.

Secondo la Auro 3D tuttavia la loro configurazione strutturata su tre livelli di altoparlanti sarebbe più accurata nella localizzazione dei suoni sull'asse verticale rispetto ad altre configurazioni *object-based* presenti sul mercato che si avvalgono di due soli livelli di altoparlanti. Questo perché il sistema uditivo umano è orientato orizzontalmente e dunque non può distinguere sull'asse verticale suoni di sorgenti *phantom*²⁰ tra due altoparlanti nella stessa maniera in cui lo farebbe se posizionati nell'asse orizzontale.

Dunque sperimentazioni interne ad Auro Technologies avrebbero provato che diffondere i suoni attraverso tre livelli di altoparlanti sia la maniera più eloquente di ricreare la diffusione naturale di un suono nell'emisfero all'altezza ed oltre la testa di uno spettatore.

²⁰ L'utilizzo di sorgenti fantasma, *phantom sources*, è alla base del metodo di posizionamento di suoni all'interno di una scena. I suoni possono essere posizionati apparentemente tra due altoparlanti, restituendo la percezione della localizzazione di un suono anche là dove in realtà non c'è alcun altoparlante. Ciò si ottiene variando l'ampiezza e il ritardo dei segnali degli altoparlanti sinistro e destro.

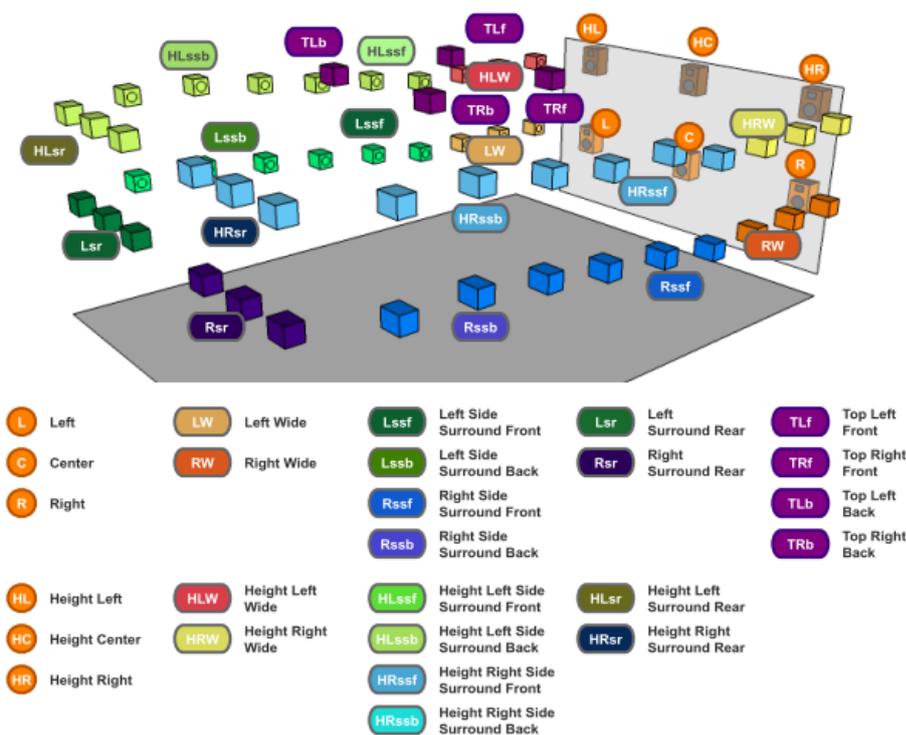


Figura 19 Disposizione degli altoparlanti divisi per zone di un sistema AutoMax 26.1. [24]

Al fine di fornire un più alto grado di compatibilità con i propri sistemi Auro 11.1 ed una maggior flessibilità tra i vari *playback environment* sono state definite più configurazioni di riproduzione AuroMax®.

Queste ultime fanno uso di "zone", dividendo i tradizionali array di altoparlanti surround nella stanza in gruppi più piccoli di altoparlanti in cui poter posizionare discretamente i suoni trattati mediante oggetti.

Le zone sono indirizzabili individualmente e forniscono una maggiore risoluzione per l'audio basato su oggetti. Possono arrivare nella versione di AuroMax più completa fino a 27, passando per 23 nella versione 22.1, fino a 21 nella 20.1. In quest'ultima versione oltre ad eliminare il "proscenio", ovvero quegli array di altoparlanti che congiungono il sistema surround con lo schermo consentendo un panning di oggetti più fluido, gli altoparlanti a soffitto vengono divisi in due zone anziché quattro.

2.2.1.2 NHK 22.2

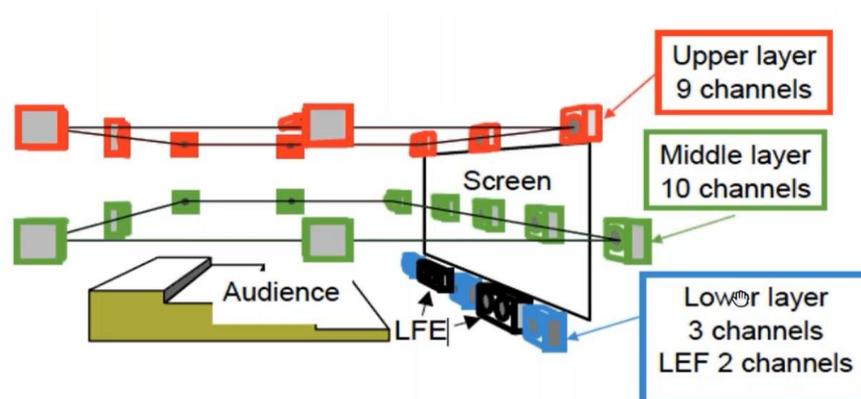


Figura 20 Configurazione di un sistema NHK 22.2 [25]

Il sistema multicanale 22.2 utilizza nella sua configurazione 9 altoparlanti in corrispondenza del *Height Layer*, 10 altoparlanti in corrispondenza del *Base Layer* e tre altoparlanti in corrispondenza del pavimento. Un sistema che secondo la specifica ITU-R BS.2051-0 (ITU, 2014) verrebbe così descritto: $9 + 10 + 3$.

Il sistema NHK è stato progettato per fornire un sistema multicanale in grado di completare l'esperienza visiva offerta dalle televisioni di Ultra Alta Definizione (UHDTV) utilizzati nel sistema di *digital broadcasting* del Giappone. Tali sistemi televisivi infatti hanno una definizione pari a 4320×7680 px e un angolo di visione orizzontale di 100° , fornendo una risoluzione visiva pari al doppio della qualità di una pellicola in 70mm.

Per un simile apparato visivo il sistema surround 5.1 forniva un'esperienza limitata, ecco perché la società giapponese di broadcasting NHK ha sviluppato il sistema multicanale 22.2.

Secondo quanto affermato da Kimio Hamasaki [25] un sistema sonoro NHK al fine di supportare correttamente l'estrema qualità visiva di un sistema UHDTV deve essere in grado di:

- Localizzare stabilmente i suoni frontali sull'intera area dello schermo.
- Riprodurre l'immagine sonora in tutte le direzioni attorno allo spettatore, inclusa l'elevazione.
- Riprodurre un'impressione spaziale tridimensionale che aumenti il senso della realtà.
- Creare un'ampia area di ascolto con un'eccezionale qualità del suono.
- Essere compatibile con i sistemi audio multicanale esistenti.

Il sistema multicanale 22.2 risulta pertanto così suddiviso:

Channel No.	Channel name (Shortened form)	Channel name
1	M-F-L	Middle layer - Front - Left
2	M-F-L/C	Middle layer - Front - Left/Center
3	M-F-C	Middle layer - Front - Center
4	M-F-C/R	Middle layer - Front - Center/Right
5	M-F-R	Middle layer - Front - Right
6	M-S-L	Middle layer - Side - Left
7	M-S-R	Middle layer - Side - Right
8	M-B-L	Middle layer - Back - Left
9	M-B-C	Middle layer - Back - Center
10	M-B-R	Middle layer - Back - Right
11	U-F-L	Upper layer - Front - Left
12	U-F-C	Upper layer - Front - Center
13	U-F-R	Upper layer - Front - Right
14	U-S-L	Upper layer - Side - Left
15	U-S-C	Upper layer - Side - Center
16	U-S-R	Upper layer - Side - Right
17	U-B-L	Upper layer - Back - Left
18	U-B-C	Upper layer - Back - Center
19	U-B-R	Upper layer - Back - Right
20	L-F-L	Lower layer - Front - Left
21	L-F-C	Lower layer - Front - Center
22	L-F-R	Lower layer - Front - Right
23	LFE-L	LFE - Left
24	LFE-R	LFE - Right

Figura 21 Elenco dei canali del sistema NHK 22.2 suddivisi nei tre strati "U+ M + B".

[25]

2.2.1.3 DOLBY ATMOS

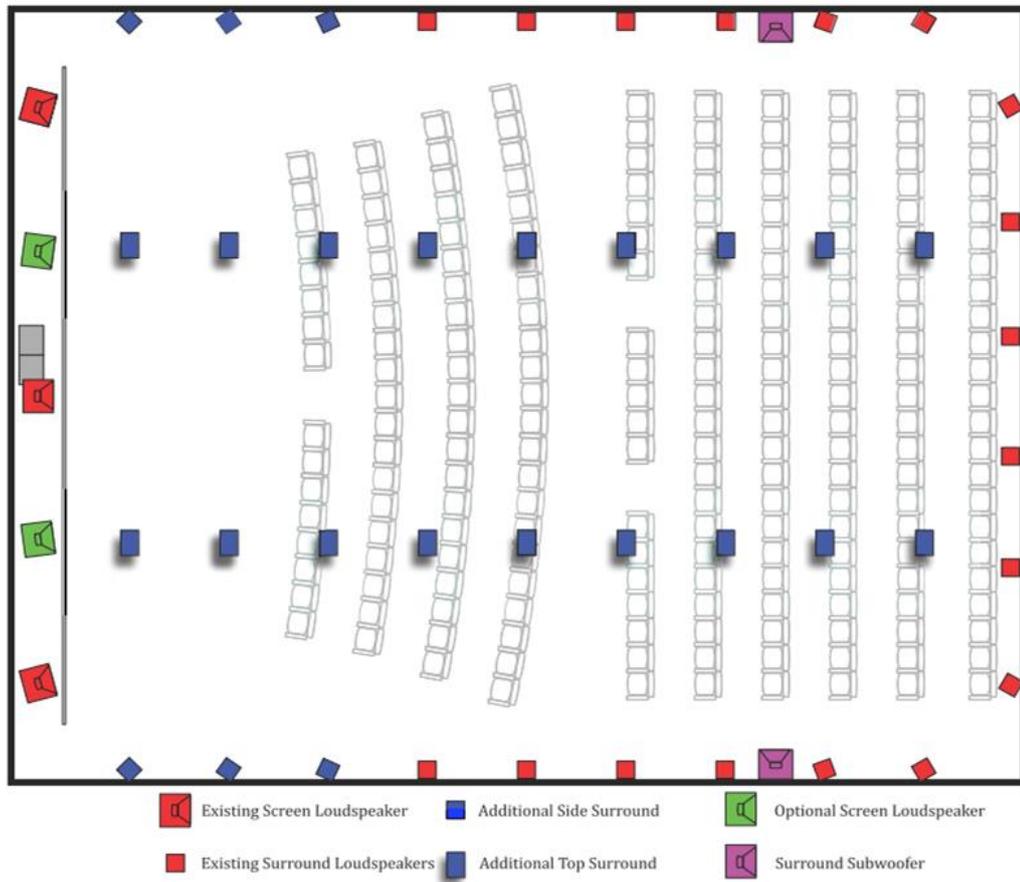


Figura 22 Configurazione degli altoparlanti di un sistema Dolby Atmos. [27]

L'ultimo sistema analizzato è quello del Dolby Atmos che al sistema multicanale 7.1 aggiunge due array di altoparlanti al soffitto tramutandosi di fatto in un sistema 9.3 (considerando anche i due eventuali canali LFE supplementari). Se vengono aggiunti anche i canali facoltativi dello schermo allora si tramuta in un sistema 11.1. (7.2.3, 7 canali frontali e surround, 2 canali sopraelevati, 3 subwoofer).

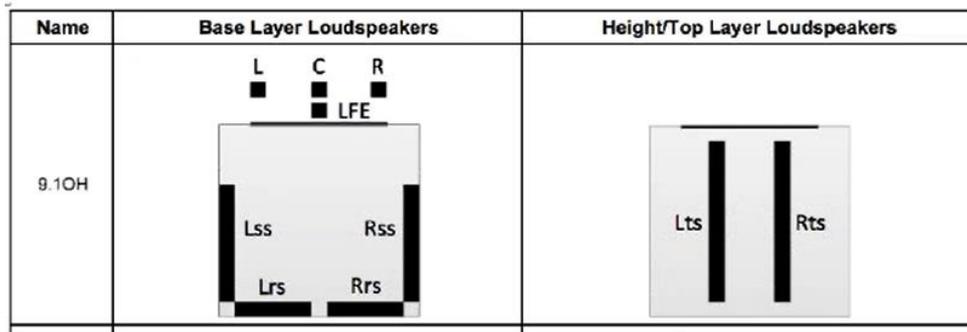


Figura 23 Sezione di una sala Atmos suddivisa per Layer di altoparlanti. [22]

Come si può vedere dalla Figura 22, in rosso sono segnalati gli altoparlanti ereditati dalle configurazioni surround precedenti, in blu invece i diffusori aggiunti. Oltre alle due file a soffitto, come accennato in precedenza, vengono anche aggiunti alcuni altoparlanti a congiunzione tra la sala e lo schermo al fine di consentire panning più fluidi tra lo schermo e il resto della disposizione surround.

Rispetto alle configurazioni viste in precedenza l'Atmos correda il suo sistema di un numero inferiore di altoparlanti, escludendo del tutto il contributo del cosiddetto *Height Layer*. Il vero grande passo avanti compiuto da parte della Dolby nell'ambito della configurazione immersiva di sala è stato dunque quello di proporre ed affermare sul mercato un'ibridazione tra l'approccio *channel-based* e l'approccio *object-based*.

Nel 2012 infatti Stuart Bowling, Dolby worldwide Technical Marketing Mgr. Cinema, nell'ambito di una video [26] presentazione del nuovo Dolby Atmos, ripercorre quelle che sono state le fasi procedurali nello sviluppo di una tecnologia che rappresentasse un vero passo avanti rispetto alle configurazioni multicanale esistenti.

Le evoluzioni fino al sistema surround 7.1 hanno portato a dividere la sala in un numero sempre crescente di aree al fine di favorire la dinamizzazione spaziale dell'esperienza sonora. Pertanto egli afferma che da un punto di vista logico le progressioni più plausibili a loro disposizione

andavano nella direzione di aggiungere ulteriori canali tramutando il sistema in una configurazione 9.1, 11.1 o addirittura 13.1.

Tuttavia la scelta fatta dalla Dolby è stata quella di confrontarsi con i *content creators* ridisegnando del tutto l'esperienza multicanale fino a quel momento veicolata. Stuart Bowling infatti domanda "what if we can go back and completely redefine how we bring the surround sound to the audience?" e successivamente aggiunge "what Dolby is bringing now is the next generation surround sound format for the industry".

In queste due affermazioni viene dunque suggerito il vero cambiamento d'approccio introdotto dalla configurazione Atmos rispetto alle precedenti esistenti sul mercato. L'approccio della Dolby infatti non è stato tanto quello di intensificare il numero di canali attraverso cui circondare lo spettatore, quanto quello di reinventare il formato surround emancipandosi da quello che era l'approccio per canali ibridandolo con quello che viene definito appunto *object-based*, in grado di aumentare la definizione e precisione spaziale grazie al corredo di metadati che ne descrivono le caratteristiche tempo-varianti.

Pertanto ai sopracitati 10 canali del sistema 9.1, di cui due in altezza, vengono aggiunti fino a 118 oggetti audio da poter distribuire fino a 64 altoparlanti di sala.

L'aumento della risoluzione spaziale dei suoni grazie agli oggetti consente, secondo quanto afferma la Dolby, di considerare l'ascolto non più favorito solo nel cosiddetto *sweet spot*, ma ovunque ci si trovi seduti in sala, promuovendo una soluzione *room-centric*.

Le potenzialità offerte da un approccio *object-based* non sono solo di natura estetica. Infatti una delle caratteristiche fondamentali introdotte, memore la Dolby delle esperienze passate, è stata quella di impostare un formato in grado di garantire **retrocompatibilità** con i sistemi precedenti

sulla base del principio "author once optimize everywhere" [27].

Il mix audio, grazie all'impiego dei metadati ed agli algoritmi di rendering, in fase di riproduzione viene automaticamente ottimizzato al fine di restituire la miglior resa in virtù del *playback environment* di riferimento. Consentendo in questo modo di risparmiare una notevole quantità di tempo nella creazione di mix differenti per ogni destinazione del prodotto.

Un fattore questo di estrema importanza per garantire da un lato la preservazione degli intenti creativi laddove i sistemi di riproduzione sono i più disparati e non è quasi mai possibile realizzare un mix ad hoc per ogni singolo sistema di diffusione e dall'altro agevolare in termini economici e procedurali l'intero processo.

Non è dunque un caso se tecnologie in ambito di diffusione sonora immersiva, anche di precedente invenzione rispetto all'Atmos, sulla base del lavoro di standardizzazione promosso dalla SMPTE al fine di dare vita ad un unico standard aperto per il Suono Immersivo, hanno in seguito implementato i rispettivi sistemi ibridandoli con l'approccio *object-based*, come AuroMax nel 2015 e il sistema DTS:X sempre nel 2015.

Pertanto oggi la maggior parte dei Formati Audio Immersivi indipendentemente dalla configurazione di sala adottata si basano su l'ibridazione di un approccio *channel-based* e *object-based* di cui analizzeremo i principi di base nei prossimi paragrafi per poi ripercorrerli più approfonditamente da un punto di vista del formato attraverso il lavoro di standardizzazione del Comitato Tecnico *SMPTE 25 Cinema Sound System*.

2.2.2 Gli elementi Bed

Un elemento Bed è composto da un gruppo di canali audio²¹ intesi per essere riprodotti contemporaneamente attraverso una data configurazione di sala, quale il sistema 5.1, il 7.1 o il 9.1, che funge da base per il mix immersivo della colonna sonora.

I Beds (letti) formano dunque la base per il mix, il loro funzionamento è assimilabile alle pratiche surround classiche poiché essi non considerano gli altoparlanti in maniera indipendente. Al contrario essi si avvalgono di una matrice di altoparlanti nello spazio che riproduce il campo sonoro stereofonico di fondo, consentendo ai fonici di mix di partire da un modus operandi a loro familiare su cui poter poi muovere indipendentemente gli oggetti sonori, gli unici ad avere la facoltà di essere renderizzati a casse indipendenti nello spazio.

I Beds sono utilizzati solitamente per contenuti che non richiedono flessibilità o un incremento di risoluzione spaziale durante la riproduzione, come ad esempio musiche, dialoghi, ambienti.

Rispetto al passato i Bed oggi si possono avvalere non solo del campo sonoro bidimensionale sul piano orizzontale generato dalle tradizionali configurazioni surround, ma anche della terza dimensione, l'altezza, introdotta proprio dalle configurazioni sonore immersive.

²¹ Raccolta di campioni audio sequenziali destinata ad essere consegnata ad una serie di altoparlanti o ad uno singolo nel caso degli altoparlanti da schermo.

2.2.3 Gli Oggetti

Si definisce un Oggetto Sonoro (**DEF - SMPTE ST 2098-1**):

- Un elemento audio (forme d'onda audio) con metadati associati che descrivono proprietà posizionali e altre proprietà che possono variare nel tempo.
- Insieme di campioni audio e metadati associati destinati alla riproduzione in base alla posizione nello spazio e ad altre proprietà indicate dai metadati. La posizione può essere associata ad uno o più altoparlanti.

Gli oggetti sono dunque l'elemento chiave essenziale nella creazione di un ambiente sonoro immersivo che insieme ai rispettivi metadati incarnano l'intento artistico da tradurre di volta in volta nei diversi sistemi di riproduzione.

Nell'ambito infatti della capacità di restituire dinamismo e precisione spaziale ai suoni all'interno di un ambiente di riproduzione, il più grande passo in avanti fatto rispetto alle odierne configurazioni surround è rappresentato dalla possibilità di indirizzare un suono non più ad un esclusivo canale, dunque limitandosi a farlo muovere attraverso aree predefinite della sala, ma di poterlo indirizzare direttamente ad ogni singolo speaker. E' ciò che viene chiamato il "panning through array" in linea con quanto definito da Stuart Bowling la *migrazione dalla logica per canale* [26], cioè la possibilità di far scorrere un suono lungo una serie di altoparlanti restituendo quella precisione di movimento e di localizzazione ad oggi impossibile da rendere con le attuali configurazioni surround.

Un canale infatti è generalmente definito attraverso la sua assegnazione

ad uno o più altoparlanti. All'interno di un canale è come se ci fosse un'etichetta che informasse il *payout* che quando riproduce quel canale tutto il suo contenuto audio deve essere indirizzato allo stesso array di altoparlanti. Un oggetto invece non è assegnato ad un altoparlante esso è assegnato ad una posizione che il renderer leggerà e valuterà quale miglior localizzazione restituirgli a seconda del sistema di riproduzione di riferimento (*target environment*).

Questo rappresenta dunque un notevole passo avanti non solo per le sopracitate caratteristiche di miglioramento della precisione spaziale che un sistema *object-based* consente di restituire, ma anche per le opportunità di adattabilità offerte da un simile approccio. La possibilità infatti di poter renderizzare la posizione di un oggetto sonoro coerentemente alle indicazioni fornite dai metadati associati consente di garantire, come vedremo in seguito, che il contenuto in questione sia ottimizzato su una più ampia gamma di dispositivi di riproduzione (cinema, home theatre, dispositivi mobili).

Questo processo di posizionamento degli elementi sonori nello spazio è stato utilizzato, come accennato precedentemente, già all'inizio degli anni '40 con l'introduzione del sistema FANTASOUND e successivamente evoluto negli ormai comuni sistemi audio surround 5.1 e 7.1.

Fino a poco tempo fa però, a causa delle limitazioni tecniche dei vari supporti di consegna, questi oggetti venivano pre-mixati in un piccolo numero di altoparlanti o canali che potevano essere riprodotti esclusivamente sui layout di altoparlanti corrispondenti, senza richiedere ulteriore elaborazione.

Solo recentemente, con il passaggio al cinema digitale, sono stati favoriti questi nuovi approcci per trasportare una serie di oggetti indipendenti da renderizzare all'interno dell'ambiente di riproduzione.

Al contrario infatti di una sequenza pre-mixata destinata a una singola configurazione di riproduzione, la trasmissione audio basata su oggetti conserva una risoluzione spaziale e l'intento artistico originale con maggior accuratezza lungo tutta la riproduzione del prodotto. Ciò fornisce maggiore adattabilità e l'opportunità di offrire prodotti più ricchi ed esperienze audio coinvolgenti all'interno di ogni ambiente.

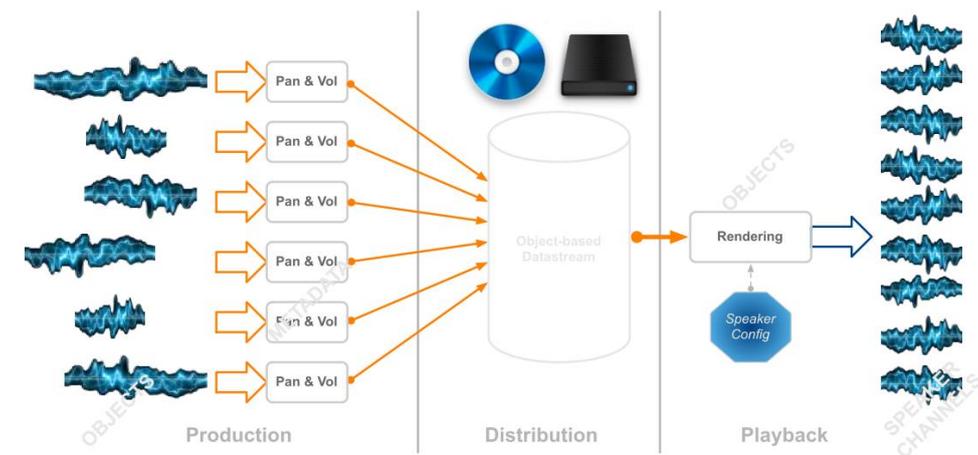


Figura 24 Immagine tratta dal white paper della AuroMax raffigurante una sistema object-based (OBA) puro in cui gli elementi audio ed i metadati non sono renderizzati a dei canali di destinazione stabiliti a priori, ma immagazzinati come oggetti audio, e relativi metadati, in un formato bitstream. In questo modo il mix risultante è speaker agnostic e pertanto la sua posizione verrà renderizzata in maniera flessibile a seconda del sistema di riproduzione di destinazione. [24]

2.3 SMPTE 2098-1/2098-2

Uno dei temi di principale interesse di questa tesi è approfondire le peculiarità che rendono un sistema audio immersivo un sistema sia da un punto di vista artistico che da un punto di vista tecnologico un vero e proprio passo in avanti rispetto ai sistemi pre-esistenti. E' infatti noto, come approfondito nel primo capitolo, che nella storia del cinema, fino all'affermarsi delle configurazioni surround, sono state molteplici ed annose

le resistenze alle evoluzioni dei sistemi di registrazione e diffusione sonora.

Pertanto è parere di alcuni operatori del settore che anche queste nuove configurazioni immersive facciano la stessa fine di quella che è stata la controversa storia del cinema 3D²².

Tuttavia uno dei fattori di vera innovatività di un programma audio immersivo risiede anche nel suo formato in grado di sfruttare ampiamente il potenziale offerto dal digitale. La possibilità di ottimizzare l'intento di mix a seconda della configurazione di destinazione e di garantire retrocompatibilità con i sistemi precedenti sono peculiarità rese possibili in ampia misura dalla struttura del formato audio immersivo. Questo è il motivo per cui nei paragrafi a seguire verrà presentato il funzionamento di un programma audio immersivo mediante l'approccio proposto nei primi due documenti della suite di standardizzazione in materia di *Immersive Audio Bitstream* proposti dalla SMPTE, come descritti nelle *SMPTE Standard Webcast Series* dedicate agli standard 2098-1 e 2098-2 [22] [28].

La SMPTE, Society of Motion Picture and Television Engineers, organismo che stabilisce standard internazionali dal 1916, nel 2014 ha infatti iniziato a lavorare ad un set di documenti chiamati *25 CSS Immersive Audio Suite*, ove l'acronimo CSS si riferisce a *Cinema Sound System*.

L'obiettivo del Comitato Tecnico è stato quello di creare degli standard

²² Il cinema 3D è comparso nelle sale all'inizio del XX secolo ed ha trovato successo la prima volta sul mercato alla metà di quest'ultimo quale strumento per fare concorrenza all'invenzione della televisione che stava portando via pubblico alle sale cinematografiche. Negli anni si sono susseguite alternanze di apparente morte e rinascita di questa tecnologia fino al 2009 che con l'uscita del film *Avatar* sembrava impossibile immaginare un futuro senza 3D. Tuttavia a causa di svariati fattori tra cui il fatto che gli spettatori non hanno mai assimilato l'uso degli occhiali 3D, spesso anche di scarsa qualità, il fatto che il 3D non restituisse una resa ottimale nelle televisioni di casa e la nascita dei televisori Ultra HD, sembra nuovamente che questa tecnologia abbia perso del tutto appeal sul pubblico.

che consentissero l'interoperabilità nell'ambito dello scambio di file audio immersivi per la distribuzione cinematografica.

Come è stato esposto precedentemente infatti esistono numerose configurazioni sonore immersive possibili sul mercato (Auro3D, Dolby Atmos, DTS:X..) e ciascuna delle quali presenta un proprio workflow operativo e strumentale, dalla creazione dei contenuti alla loro diffusione, rispettivamente incompatibile con gli altri Programmi Audio Immersivi.

L'obiettivo della SMPTE è stato dunque quello di uniformare con degli standard la creazione del bitstream audio immersivo e di altri processi intermedi al fine di rendere la creazione, distribuzione, riproduzione di un prodotto immersivo indipendente dalla tecnologia di riferimento e dunque interoperabile tra i vari sistemi al fine di favorire la diffusione di tale tecnologia.

Nell'ambito di questa tesi dunque seguire i passi dei primi documenti dello standard sarà necessario per comprendere meglio gli elementi costitutivi, i rispettivi rapporti, attributi e relazioni di un formato audio immersivo.

Inoltre essendo stato tale lavoro della SMPTE basato in buona misura su quanto fatto dalla Dolby per la tecnologia Atmos, premurandosi di rendere il nuovo formato rispettivamente compatibile, questo sarà implicitamente un modo per comprendere alcune delle peculiarità di questa tecnologia di cui ci si avvarrà da un punto di vista applicativo nel terzo capitolo.

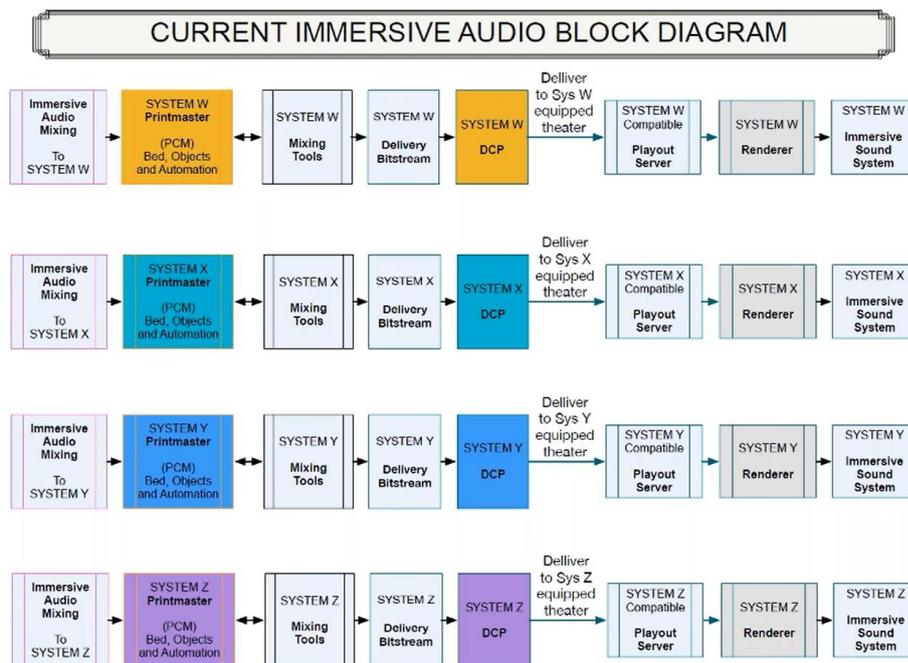
2.3.1 Comitato Tecnico 25 CSS Immersive Audio Suite

Come detto in introduzione, sulla base di alcune richieste avanzate da organizzazioni rappresentanti gli esercenti e le major di Hollywood come la

NATO (National Organisation of Theater Owners) UNIC e DCI (Digital Cinema Initiative), la SMPTE nel 2018 ha dato vita ad un unico standard aperto per il Suono Immersivo che potesse essere utilizzato come formato di distribuzione per tutti i sistemi sonori immersivi in circolazione.

Nell'ambito infatti della produzione di contenuti audio immersivi, lo scenario nel quale si è mossa la SMPTE nel 2013 circa, prima del lavoro di standardizzazione, era molto frastagliato come ben rappresentato nel diagramma a blocchi in figura 25.

Quello che accadeva è che per ogni sistema audio presente sul mercato al fine di produrre un lavoro di missaggio compatibile era necessario procedere attraverso un workflow ad hoc per ciascun sistema.



© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)

Figura 25 Diagramma a Blocchi del workflow operativo dei sistemi audio immersivi prima del lavoro di standardizzazione della SMPTE. [22]

Supponendo infatti quattro generici sistemi immersivi differenti X, Y, W, Z per ciascuno di essi era necessario realizzare un mix e un master appositi, adottare strumenti di mix specifici ed un formato di bitstream e DCP compatibili, affinché il lavoro fosse riproducibile dallo specifico renderer del sistema di riferimento.

Un *content creator* dunque qualora avesse voluto creare un lavoro riproducibile sui sistemi X, Y, W, Z avrebbe dovuto seguire quattro workflow differenti per ciascuna configurazione di destinazione. Questo perché ogni sistema accettava esclusivamente input frutto di un workflow compatibile con la tecnologia nativa, altrimenti tutto il lavoro di personalizzazione del mix sonoro, pensato per una configurazione e riprodotto su un'altra, veniva pesantemente impoverito per via di mancanza di compatibilità di formato.

Un processo questo ovviamente troppo oneroso rispetto al ROI derivante e che ha spesso disincentivato gli esercenti a convertirsi alle nuove configurazioni sonore immersive.

Ciò che dunque è stato fatto dalla SMPTE e in particolare dal Comitato Tecnico-25 CSS è stato quello di lavorare ad un sistema di standardizzazione del lato front-end di questo processo, producendo un set di documenti in grado di uniformare con degli standard la creazione di un unico input, affinché non fosse più necessario creare un mix per ogni possibile *layout* (figura 26).

Pertanto l'obiettivo è stato quello di portare in sala un mix conforme con gli standard SMPTE e che i vari sistemi di produzione e diffusione fossero in grado di gestirlo e riprodurlo.

Ogni broadcaster infatti solitamente adoperava un formato proprietario, ecco perché la suite di documenti ha puntato ad ottenere un solo bitstream, chiamato anche *mezzanino*, da cui transcodificare e derivare tutti i formati necessari richiesti dal mercato, come si può osservare dalla

figura 26 in cui il *back-end* è rimasto sostanzialmente invariato.

In questo modo un *content creator* deve solo più occuparsi di creare un unico mix ed un unico master, adottando in sala mix lo strumento di monitoring preferito, a patto che compatibile con lo standard (ST 2098-1/2098-2), ed impacchettarlo in un DCP, anch'esso conforme (ST 429-18/429-19).

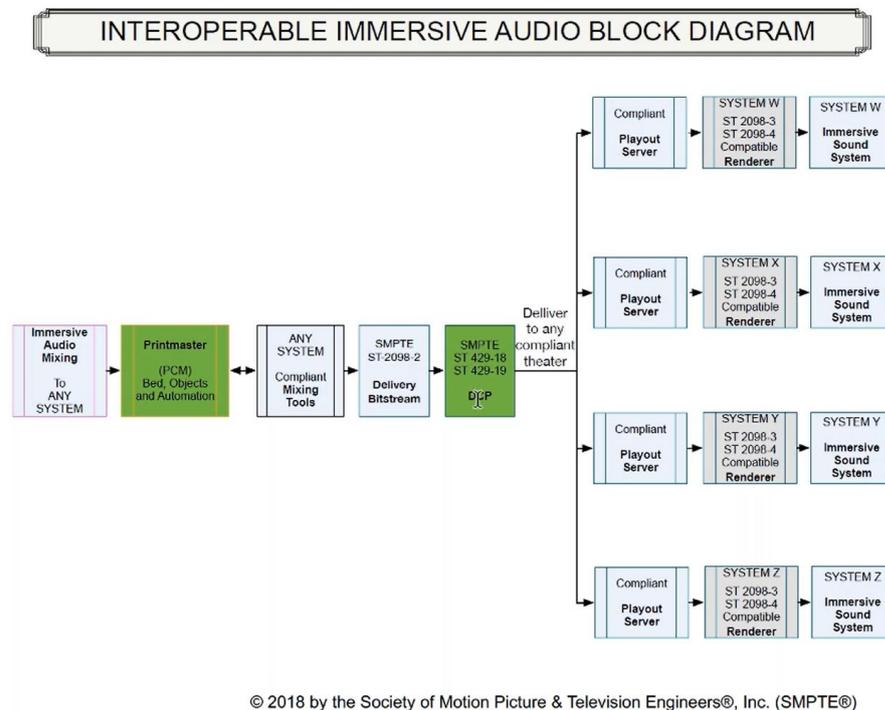


Figura 26 Diagramma a Blocchi del workflow operativo di un sistema audio immersivo reso interoperabile dopo il lavoro di standardizzazione della SMPTE. [22]

Nella fase di creazione di questi standards al tavolo di discussione vi erano seduti anche i vari *manufacturers* coinvolti nel settore cinematografico, al fine di prendere in esame i vari punti di vista e modus operandi, uno fra tutti si ricorda nuovamente essere stata la Dolby che ha ampiamente contribuito nello strutturare questo formato in virtù del lavoro fatto in passato per la propria tecnologia Atmos.

Qui di seguito si riporta un breve dettaglio dei vari documenti [22] [29] che compongono questo processo di standardizzazione.

2.3.1.1 La Suite di Documenti

ST 2098 *Immersive Audio Suite* a cura del Comitato Tecnico 25

CSS:

- **ST 2098-1 Immersive Audio Metadata**

Il "documento genitore" nella definizione dei metadati audio immersivi.

- **ST 2098-2 Immersive Audio Bitstream Specification**

Il documento principale nella definizione delle specifiche di un Immersive Audio Bitstream (IAB), cioè della rappresentazione codificata (bitstream) di un programma audio immersivo.

- **EG 2098-3 Immersive Audio Expectations**

Guida di natura ingegneristica, non uno standard dunque, che fornisce linee guida sull'implementazione. In questo caso dà delle linee guida in merito a quelle che sono le aspettative in fase di rendering, dunque essenzialmente riferendosi ai metadati definiti in 2098-1 illustra come questi verranno gestiti dal renderer.

- **RP 2098-4 Immersive Audio Renderer Interoperability Testing Procedure**

Illustra quelle che sono le pratiche raccomandate in fase di rendering sulla base di quanto specificato nelle linee guida precedenti, ovvero come utilizzare il materiale di prova ed allestire un *playback environment* per testare l'interoperabilità del renderer.

- **ST 2098-5 D-Cinema Immersive Audio Channels and Soundfiels Groups**

E' lo standard che definisce i canali D-Cinema Immersive Audio, essenzialmente riporta le specifiche che standardizzano i canali che vengono trasportati nel bitstream e le correnti configurazioni di altoparlanti per un sistema immersivo.

Comitato Tecnico SMPTE per il Cinema Digitale, TC-21 DC

Parallelamente alla suite 2098, il Comitato Tecnico SMPTE per il Cinema Digitale, TC-21 DC, ha lavorato per standardizzare i protocolli necessari a garantire che i dispositivi audio immersivi potessero comunicare in modo interoperabile con il confezionamento e la distribuzione di DCP, lavorando dunque ad una suite di documenti così composta:

- **ST 429-18 D-Cinema Packaging - Immersive Audio Track File**

Spiega la struttura del file MXF in cui viene inserito l'Immersive Audio Bitstream (IAB).

- **ST 429-19 D-Cinema Packaging - DCP Operational Constraints for Immersive Audio**

Spiega i vincoli operazionali di un DCP.

- **ST 430-14 D-Cinema Operations- Digital Sync Signal and Aux Data Transfer Protocol**

Spiega la creazione di un segnale di sincronizzazione digitale e di un protocollo di trasferimento dati ausiliario.

I primi due documenti della suite 2098 descrivono le caratteristiche dell'Immersive Audio Bitstream ed annessi metadati. La quasi totalità dei restanti documenti che caratterizzano lo standard sopra riportati sono volti invece ad uniformare alcuni processi del workflow di produzione e distribuzione al fine di renderli compatibili con la gestione e riproduzione di tale formato (IAB).

Dunque nei capitoli a seguire verranno essenzialmente seguite le linee guida dei primi due documenti della suite 2098, perno dell'intero processo di standardizzazione, così come illustrato nei Webinar SMPTE di presentazione dello standard 2098 da Brian Vessa founding Chair del comitato TC-25CSS e Stan Cosette, Senior Staff Engineer Dolby Laboratories.

2.4 Principi base di un sistema sonoro *object-based*

Un sistema sonoro *object-based* è caratterizzato essenzialmente da un renderer, come si vedrà basato su un proprio sistema di codifica dello spazio acustico, tarato su un sistema di riferimento le cui coordinate vengono utilizzate per descrivere movimenti e proprietà di un oggetto nello spazio attraverso set di metadati. Nelle tecnologie ibride basate sia su un sistema *object-based* che *channel-based*, il mix sonoro è composto di *oggetti* e di *bed* per i quali esistono altrettanti gruppi di metadati, in questo caso non tempo varianti.

Nel prossimo capitolo analizzeremo più in profondità tutti gli elementi sopracitati costitutivi di un sistema sonoro *object-based*, descrivendo alcuni dei principali set di metadati secondo quanto riportato nel documento SMPTE 2098-1.

2.4.1 Rappresentazione nello Spazio e Sistema di Coordinate di un Oggetto

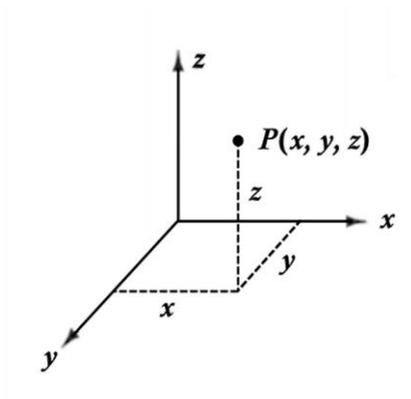


Figura 27 Sistema di Coordinate [22]

Al fine di poter posizionare gli Oggetti di un mix sonoro nello spazio di riproduzione è necessario definire un sistema di riferimento.

Una considerazione fondamentale che viene fatta nella scelta di un sistema di riferimento è la distinzione tra sistema egocentrico ed allocentrico. Il primo codifica la posizione di un oggetto rispetto alla posizione di un osservatore. Il secondo codifica la posizione di un oggetto utilizzando posizioni di riferimento e direzioni relative ad altri oggetti nell'ambiente.

Quando si sceglie un sistema di riferimento spaziale per il missaggio, secondo Riedmiller J.C. and Tsingos N. [30] devono venir presi in considerazione i seguenti fattori:

1. Come catturare al meglio l'intento artistico.
2. Come preservare e riprodurre al meglio l'intento artistico in una varietà di ambienti di ascolto (ciò che viene definita *traduzione*).
3. Gli strumenti e le interfacce utilizzate per catturare l'intento artistico.

4. Un comportamento consistente nei diversi *playback environment*.

Queste dunque sono state proprio le considerazioni fatte nell'ambito del documento di standardizzazione ST 2098-1 in merito alla definizione di un sistema di coordinate [28] che consentisse di renderizzare con accuratezza un oggetto nello spazio e che qui di seguito vengono riportate.

- I metadati posizionali degli oggetti audio utilizzano un sistema di coordinate cartesiane. La coordinata x rappresenta la dimensione sinistra-destra, la coordinata y rappresenta la dimensione fronte-retro e la coordinata z rappresenta la dimensione basso-alto.
- I valori delle coordinate cartesiane utilizzati per la posizione di un oggetto audio devono essere normalizzate rispetto ai punti di riferimento di un cubo unitario che rappresenta un modello cinematografico ideale. Il cubo ha dunque una dimensione pari ad 1 su tutti i lati, in modo che tutti gli assi vadano da 0 ad 1 con 0 che rappresenta un muro ed 1 che rappresenta l'altro muro diametralmente opposto.

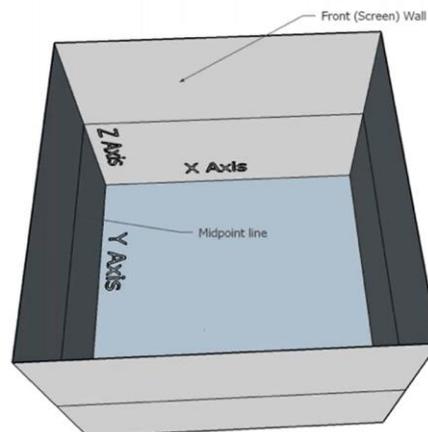


Figura 28 Rappresentazione ideale di un cinema come cubo unitario. [28]

- Il sistema di riferimento e i metadati posizionali dunque sono di natura **allocentrica** e ciò significa che i punti cardinali si basano su **posizioni relative**, coerentemente alle pareti del cubo unitario:
 1. La faccia **anteriore** del cubo deve corrispondere alla parete anteriore nominale del cinema. Inoltre, la superficie frontale dove sono installati gli altoparlanti sarà considerata la parete anteriore nominale del cinema.
 2. La faccia **sinistra** del cubo deve corrispondere alla parete sinistra nominale del cinema. La superficie dove sono installati gli altoparlanti più a sinistra deve essere considerata la parete sinistra nominale del cinema.
 3. la faccia **destra** del cubo deve corrispondere alla parete destra nominale del cinema. La superficie dove sono installati gli altoparlanti più a destra deve essere considerata la parete destra nominale del cinema.
 4. La faccia **posteriore** del cubo deve corrispondere alla parete posteriore nominale del cinema. La superficie posteriore dove sono installati gli altoparlanti deve essere considerata la parete posteriore nominale del cinema.
 5. Il piano a mezza altezza del cubo deve corrispondere all'altezza del sistema di altoparlanti bidimensionale esistente (configurazione 5.1, 7.1).
 6. La faccia **superiore** del cubo deve corrispondere al soffitto del cinema. La superficie sopraelevata dove sono installati gli

altoparlanti deve essere considerata il soffitto nominale del cinema.

- L'origine del sistema di riferimento è definita dall'intersezione del muro frontale, dove si trova lo schermo, con il muro sinistro all'altezza del *base layer*.
- Le posizioni all'interno del cubo devono essere rappresentate utilizzando l'interpolazione lineare tra i punti di riferimento adiacenti su ciascun asse. Le posizioni al di fuori del cubo, se supportate, devono essere rappresentate in base ad un'estrapolazione lineare dal punto di riferimento sul cubo.
- Il piano frontale corrisponde alla posizione dello schermo; "sinistra" è relativa ad un osservatore nel cubo, rivolto verso lo schermo. I metadati possono supportare le posizioni degli oggetti audio all'interno ed all'esterno del cubo. Come minimo, i metadati di posizione devono supportare le posizioni sul ed all'interno del cubo dal punto medio dell'asse z alla parte superiore del cubo.

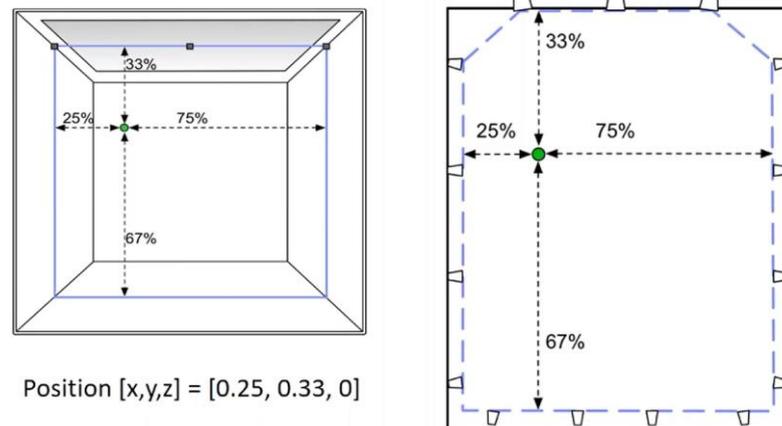


Figura 29 L'immagine rappresenta come in un sistema allocentrico la posizione degli oggetti viene preservata rispetto alla forma e alle dimensioni dell'ambiente di riproduzione grazie all'utilizzo di coordinate relative. A sinistra è rappresentato un generico interfaccia di panning e a destra la sua traslazione all'interno di una sala con dimensioni reali. La linea azzurra a sinistra rappresenta il perimetro di una stanza quadrata nominale. La linea tratteggiata blu a destra mostra il perimetro della stanza definita dalla posizione degli altoparlanti. [28]

MAPPATURA

La mappatura della posizione di un oggetto all'interno del cubo rispetto agli altoparlanti del cinema (o della sala di missaggio) è il risultato di una funzione del renderer. Tuttavia, il punto di riferimento all'interno del cubo deve avere un significato definito indipendentemente dalla forma della stanza in cui sono mappate le posizioni.

Utilizzando dunque un sistema di riferimento allocentrico l'intento non è quello di preservare la posizione degli oggetti rispetto ad un osservatore, **ma rispetto alla forma e alle dimensioni dell'ambiente di riproduzione.**

Come si può osservare dalla figura 29, l'immagine a sinistra mostra un possibile interfaccia di panning di uno strumento di *mastering*, l'immagine a destra mostra la vista dall'alto semplificata di un cinema. Nell'interfaccia di

panning, la sorgente sonora (punto verde) è stata posizionata ad $1/3$ del percorso fronte-retro e ad $1/4$ del percorso sinistra-destra.

La pianta del cinema dall'alto mostra invece come la posizione dell'oggetto verrebbe mappata in uno spazio esteso più in lunghezza che in larghezza e che dunque non presenta dimensioni quadrate ideali.

Da questa rappresentazione si evince come il renderer grazie ai metadati posizionali ed alla natura allocentrica del sistema di riferimento prende le coordinate assegnate in sala di mix e posiziona gli oggetti in relazione a quella che è la struttura della sala di riproduzione.

Come si può vedere dall'esempio **ciò che viene preservata è la distanza relativa**, in questo caso, dai muri frontale e posteriore e da quelli laterali. Il renderer andrà a posizionare quel suono in una posizione distante il 33% ($1/3$) dal muro frontale e il 25% ($1/4$) dal muro di sinistra, all'altezza del Base Layer, rispettando le medesime distanze proporzionali impostate in sala mix.

Ecco perché si parla di preservare l'**intento posizionale** o l'intento di mappatura, i layout delle sale cinematografiche variano gli uni dagli altri il che interferisce con la possibilità di restituire la posizione precisa di un suono così come assegnata in sala mix, ma la possibilità di basarsi su coordinate relative consente di ottimizzare la posizione di quel suono qualsiasi sia il layout del sistema di *playback* in cui viene riprodotto.

Uno degli effetti dell'utilizzo di un sistema di riferimento allocentrico viene illustrato nel White Paper di presentazione del Dolby Atmos in cui grazie alla possibilità di posizionare precisamente un oggetto nello spazio, coerentemente alle caratteristiche di sala, i suoni *off-screen* rispettano la linea dello sguardo degli attori e la rispettiva localizzazione, pur percepita proveniente da diverse direzioni a seconda della seduta in sala, verrà

riconosciuta come coerente anche dagli spettatori cui sembrerà di condividere lo stesso spazio in cui avvengono le azioni sullo schermo.

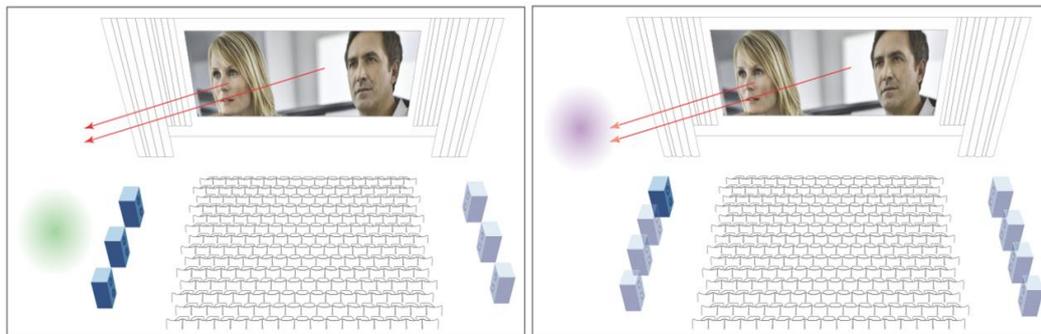


Figura 30 La possibilità di posizionare oggetti sonori nello spazio con precisione puntiforme (immagine a destra) anziché indirizzarli ad un intero canale (immagine a sinistra) consentirà alle immagini sullo schermo di seguire con estrema precisione i suoni fuori campo. Questa accresciuta risoluzione spaziale, unitamente all'utilizzo di un sistema di coordinate allocentrico consente agli spettatori di avere la percezione di condividere lo stesso spazio in cui avvengono le azioni sullo schermo. [27]

Dunque tutti i metadati posizionali vengono generati e decodificati utilizzando questo sistema di riferimento, applicato sia agli oggetti che ai Bed, solo infatti utilizzando lo stesso sistema di riferimento è possibile garantire che la relazione spaziale tra oggetti e canali venga preservata.

2.4.2 Renderer

Il *renderer* è un dispositivo o algoritmo che legge le tracce audio e i metadati associati e li converte in un altro set di segnali audio destinati a dispositivi di riproduzione.

Gli algoritmi di rendering audio, nel caso degli oggetti, mappano un segnale audio generalmente monofonico a una serie di altoparlanti per generare la percezione di un evento uditivo in una posizione sorgente prevista

nello spazio.

In un Sistema Audio Immersivo dunque, il rendering è quel processo in cui l'audio immersivo viene preso come input e viene trasformato in un output di riproduzione ottimizzato per la specifica configurazione del campo sonoro immersivo di riferimento. E' dunque un elemento chiave per codificare lo spazio acustico di riproduzione e conseguentemente posizionare nella maniera più precisa e flessibile possibile i suoni nello spazio.

Brian Vessa, presidente fondatore della SMPTE TC 25 CSS, definisce ironicamente il renderer come una *scatola magica* che prende l'intento di mix convogliato dall'audio immersivo e dai suoi metadati e decide come dovrebbe essere riprodotto nel sistema sonoro a cui è collegato, in modo tale che l'intento creativo del mix sia preservato nel miglior modo possibile.

In generale qualsiasi algoritmo di rendering di oggetti utilizzato in applicazioni di riproduzione professionali o *consumer* deve soddisfare (come minimo) i seguenti requisiti funzionali di alto livello [30]:

1. Supportare un panning ⁵ fluido per oggetti renderizzati nella stanza ed attraverso di essa, sia nel piano bidimensionale di un sistema surround che nel cubo tridimensionale di un sistema immersivo.
2. Supportare l'operazione di rendering utilizzando un layout bidimensionale tradizionale di diffusori surround anche quando sono presenti diffusori a soffitto. Ad esempio, se nei parametri di un oggetto è riportato $z = 0$, esso verrà renderizzato attraverso il layout degli altoparlanti 2D, mentre se viene riportato $z = 1$, deve venir renderizzato solo tramite altoparlanti in altezza.
3. Implementare un approccio che riduca al minimo il numero di altoparlanti utilizzati per riprodurre un oggetto, questo al fine di

preservare la fedeltà timbrica.

4. Supportare un controllo creativo aggiuntivo (tramite metadati) che includa la possibilità di mascherare dinamicamente alcuni degli altoparlanti (o zone di altoparlanti) quando un oggetto viene renderizzato attraverso specifiche configurazioni di diffusori.
5. Conservare la potenza all'interno delle configurazioni di altoparlanti pari a: $\sum_i G_i^2 = 1$.

Esistono diversi approcci di rendering, algoritmi quali Wave Field Synthesis (WFS) o Higher Order Ambisonics (HOA) mirano a codificare un intero campo sonoro indipendentemente dalla disposizione degli altoparlanti di riproduzione [31].

Altri algoritmi, ampiamente utilizzati dai professionisti del settore audio, tentano di ricreare i segnali appropriati tramite tecniche di panning di ampiezza [31] in cui viene calcolato un vettore guadagno²³ normalizzato $[G_i]$ ($1 < i < n$), dove $\sum_i G_i^2 = 1$, ed assegnato al segnale sorgente per

²³ Il *panning* è l'operazione di distribuzione di un segnale sonoro (sia esso mono o stereo) in un nuovo campo sonoro (stereo o multicanale). Il pan o pan pot (abbreviazione di "potenziometro di panning") è un controllo analogico con un indicatore di posizione che può variare da completamente a sinistra a completamente a destra. Il software di mixaggio audio sostituisce i pan pot con manopole o cursori virtuali su schermo che funzionano come le loro controparti fisiche. Il panning determina dunque la quantità di segnale sorgente inviata ai bus sinistro e destro. I potenziometri dividono i segnali audio in canali sinistro e destro, ciascuno dotato di un proprio controllo di guadagno.

⁶ Il guadagno corrisponde alla componente di amplificazione del segnale in ingresso sul canale in uscita, ovvero il rapporto tra l'ampiezza del segnale in uscita e quello in ingresso. Viene spesso espresso utilizzando le unità logaritmiche di decibel (dB) ("guadagno dB"). Un guadagno pari a 0 dB corrisponde ad un rapporto unitario e quindi a non avere amplificazione.

ciascuno degli n altoparlanti in uso.

Il segnale dell'Oggetto $s(t)$ viene quindi riprodotto da ciascun altoparlante come: $\mathbf{G}_i(\mathbf{x}, \mathbf{y}, \mathbf{z}) \cdot \mathbf{s}(t)$ in modo da ricreare un'opportuna spazializzazione dei segnali come indicato dalle coordinate dell'oggetto in questione $\langle x, y, z \rangle$.

La struttura di un algoritmo di panning deve equilibrare diversi *tradeoffs* che si presentano nella riproduzione di un segnale quali, garantire fedeltà timbrica e accuratezza spaziale, fluidità e sensibilità rispetto alla posizione dello spettatore, fattori questi che, se valorizzati uno rispetto all'altro, possono influenzare il modo in cui un oggetto in una data posizione nello spazio viene percepito dagli spettatori. Alcuni dei parametri necessari a gestire questi *tradeoffs* verranno analizzati più avanti (paragrafo 2.4.5) andando nel merito dei metadati che descrivono le proprietà di un oggetto.

Di seguito verranno analizzati alcuni algoritmi di panning di ampiezza come riassunti da Jeffrey C. Riedmiller and Nicolas Tsingos [30] (figura 31) tra cui il VBAP adottato dal Dolby Atmos. Si ricorda infine che un sistema object-based può avvalersi anche di più approcci di rendering a seconda della scena sonora che deve essere ricreata.

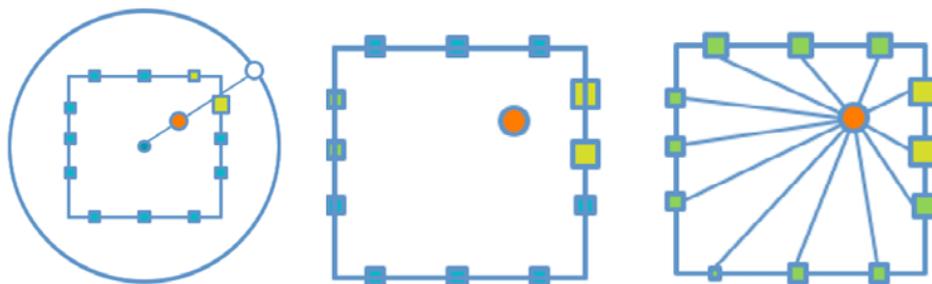


Figura 31 Illustrazione di diversi algoritmi di panning: (a) Directional Pairwise Panning, (b) Dual-Balance Panning, (c) Distance Based Panning. I quadratini verdi rappresentano gli altoparlanti attivi utilizzati per renderizzare l'oggetto P (il pallino arancione), i quadratini azzurri invece sono gli altoparlanti non attivi per questa operazione di rendering. La dimensione di ciascun quadratino rappresenta indicativamente l'ampiezza del segnale inviato a ciascun altoparlante. [30]

2.4.2.1 Directional Pairwise Panning (DIR)

L'algoritmo di Panning a coppie direzionali (DIR) è una strategia comunemente utilizzata che si basa esclusivamente sul vettore direzionale descritto dalla posizione di riferimento del sistema (generalmente rappresentata dal cosiddetto *sweet spot* di sala o dal centro della stanza come indicato in figura) e dalla posizione P dell'oggetto da renderizzare.

La coppia di altoparlanti che racchiude il vettore direzionale viene di conseguenza utilizzata per posizionare (renderizzare) la posizione di quell'oggetto nello spazio durante la riproduzione.

Un'estensione del panning a coppie direzionali per supportare i layout degli altoparlanti 3D è il panning di ampiezza basato su vettori (VBAP) che utilizza in questo caso triplette di altoparlanti per renderizzare un suono con la desiderata direzione di incidenza tridimensionale rispetto all'ascoltatore.

Una tripletta di altoparlanti può essere ottenuta attraverso la triangolazione²⁴ dello scafo convesso²⁵ o involuppo convesso di una array di altoparlanti [32].

Pertanto per una data posizione dell'oggetto P e dello *sweet spot* O, una singola tripletta di altoparlanti è ottenuta intersecando il corrispondente vettore di direzione $\mathbf{d} = \mathbf{p} - \mathbf{O} / \|\mathbf{p} - \mathbf{O}\|$ con l'involuppo convesso triangolare.

La direzione l può essere espressa in funzione delle direzioni l_1, l_2, l_3 dei 3 altoparlanti corrispondenti come: $\mathbf{d} = \mathbf{g}_1 \mathbf{l}_1 + \mathbf{g}_2 \mathbf{l}_2 + \mathbf{g}_3 \mathbf{l}_3$, in cui il

²⁴ La triangolazione è una tecnica che permette di calcolare distanze fra punti sfruttando le proprietà dei triangoli.

²⁵ Lo scafo convesso per un insieme di punti X in uno spazio vettoriale reale V è l'insieme convesso minimo contenente X.

vettore di guadagni per ogni altoparlante $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3]$ può essere quindi ottenuto come:

$$\mathbf{G} = \mathbf{d}^T \mathbf{L}^{-1},$$

dove \mathbf{L} è la matrice 3×3 dei vettori di direzione degli altoparlanti.

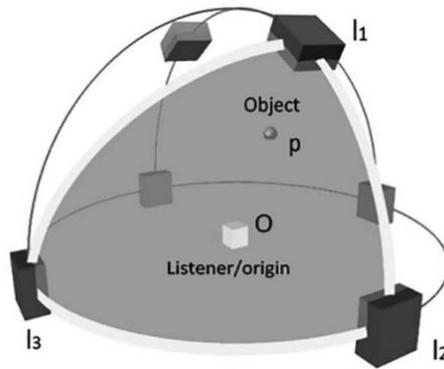


Figura 32 Algoritmo di Panning VBAP che utilizza triplette di altoparlanti per renderizzare un Oggetto sonoro P con la desiderata direzione di incidenza tridimensionale rispetto all'ascoltatore O. La tripletta di altoparlanti l1, l2, l3 racchiude la desiderata direzione di incidenza tridimensionale rispetto all'ascoltatore O. La tripletta di altoparlanti l1, l2, l3 racchiude la desiderata direzione d'incidenza (OP). [31]

L'algoritmo VBAP richiede anche che lo scafo convesso descritto dagli altoparlanti racchiuda l'intera sfera (o emisfero) delle direzioni. Al fine di prevenire movimenti irregolari della sorgente ed evitare la necessità di tagliare le coordinate spaziali degli oggetti, alcuni sistemi includono altoparlanti virtuali in aree della configurazione di destinazione non coperte da altoparlanti fisici.

Pertanto durante il rendering, l'algoritmo VBAP viene applicato ad una configurazione di altoparlanti estesa da quelli virtuali.

I segnali ottenuti per gli altoparlanti virtuali vengono elaborati in un

downmix²⁶ per gli altoparlanti fisici effettivi. I guadagni di downmix per la mappatura degli altoparlanti virtuali su quelli fisici sono derivati distribuendo l'energia degli altoparlanti virtuali equamente agli altoparlanti vicini.

E' interessante sottolineare che un caso d'uso importante per quanto riguarda l'aggiunta di altoparlanti virtuali sono i layout di riproduzione che consistono solo di altoparlanti nel piano orizzontale, ovvero all'altezza del *base layer*: in questo caso, un altoparlante virtuale viene dunque aggiunto allo zenit, cioè sopra il centro dell'area di ascolto, restituendo la percezione di movimenti fluidi, ad esempio, per oggetti sonori che devono sorvolare le teste degli spettatori.

Nicolas Tsingos nel libro *Immersive Sound* segnala che questo algoritmo di panning presenta alcune criticità che si riporta qui di seguito.

La prima è legata al fatto che l'algoritmo, utilizzando solo la direzione della sorgente rispetto ad una posizione di riferimento, non può distinguere tra sorgenti di oggetti in posizioni diverse lungo lo stesso vettore di direzione.

L'altra riguarda la possibilità che venga introdotta instabilità quando gli oggetti vengono spostati vicino al centro della stanza. Infatti viene fatta una nota a margine in merito al fatto che le soluzioni di panning direzionale possono anche creare transizioni nette tra altoparlanti quando gli oggetti si avvicinano al centro della stanza. Dove c'è il rischio che un piccolo movimento della posizione di un oggetto non sempre venga tradotto in una piccola variazione nei guadagni degli altoparlanti.

²⁶ Il processo di mixdown o downmix converte un programma con una configurazione multicanale in un programma che ha a disposizione meno canali. Un esempio tipico è il downmix da audio surround 5.1 a stereo o da stereo a mono.

2.4.2.2 'Dual Balance' Panning (DB)

L'algoritmo di panning "Dual-Balance" è ad oggi l'approccio di *pan pot*²⁷ sinistra/destra, fronte/retro più comunemente utilizzato nei sistemi surround 5.1, 7.1.

Di conseguenza, il panning a *doppio bilanciamento* funziona generalmente su un set di quattro altoparlanti che racchiudono la posizione desiderata dell'oggetto nel campo 2D.

Estendendo questo approccio a sistemi di diffusione tridimensionali caratterizzati anche dalla verticalizzazione del suono l'algoritmo produce un panning "triple balance" in cui vengono generate tre serie di guadagni unidimensionali corrispondenti a valori di equilibrio sinistra/destra, fronte/retro, sopra/sotto.

Questi valori possono poi essere moltiplicati per ottenere i guadagni finali degli altoparlanti:

$$G_i(x,y,z) = G_{xi}(x) \cdot G_{yi}(y) \cdot G_{zi}(z)$$

Questo approccio fornisce completa fluidità per gli oggetti di cui è necessario fare un panning attraverso la stanza sia esso in 2D o 3D e rende inoltre più semplice controllare con precisione quando devono essere utilizzati gli altoparlanti del *base layer* o dei *top/height layer*.

²⁷ *Pan pot*, o manopola di pan, è l'abbreviazione di "potenziometro panoramico" che consente di posizionare il segnale in qualsiasi parte del campo, nel caso surround, da sinistra all'estrema destra e dallo schermo fino al fondo della sala.

2.4.2.3 Distance-Based Panning (DIST)

In contrapposizione con i primi due, questo algoritmo per determinare i guadagni di panning utilizza la distanza relativa tra la posizione (2D o 3D) desiderata dell'oggetto e ciascun altoparlante in uso.

Di conseguenza, questo approccio generalmente utilizza tutti gli altoparlanti disponibili in uso piuttosto che un sottoinsieme limitato il che consente panning di oggetti più fluidi ma con il compromesso d'essere incline ad artefatti timbrici.

Il *Distance-Based Panning* e il *Dual-Balance Panning* in generale dunque supportano operazioni di panning più fluide, ovvero che a piccole variazioni di posizione di un oggetto corrispondono piccole variazioni dei guadagni degli altoparlanti, rispetto al *Directional Pairwise Panning*.

2.4.3 Caratteristiche degli *Immersive Audio Metadata ST 2098-1*

Come ampiamente citato in precedenza i metadati sono un elemento fondamentale nel processo atto ad interpretare e preservare l'intento creativo di un programma audio immersivo.

Nei paragrafi che seguiranno verranno analizzati nel dettaglio alcuni dei principali set di metadati che caratterizzano il formato immersivo presentati all'interno dei documenti di standardizzazione SMPTE 2098.

Verranno suddivisi in due categorie “interattività e personalizzazione nella fase di broadcasting” e “controlli artistici per il rendering di oggetti nella riproduzione cinematografica”, di cui la seconda inerente solo agli oggetti, al fine di illustrarne con immediatezza le rispettive finalità. I primi

appunto volti a garantire l'interoperabilità del nuovo formato, i secondi a fornire strumenti in grado di preservare con maggior accuratezza l'intento artistico dei sound designer alla luce delle nuove configurazioni di sala.

Tale suddivisione è intesa solo ai fini di una miglior delucidazione della loro funzione all'interno di un programma audio immersivo, ma non è un termine di catalogazione presentato nei documenti SMPTE. Nel paragrafo 2.5 invece verranno illustrati questi stessi parametri ed alcuni altri aggiuntivi da un punto di vista costitutivo dell'Immersive Audio Bitstream così come riportato nel documento di standardizzazione SMPTE 2098-2.

2.4.3.1 Bed Metadata ST 2098-1

I beds sono una raccolta di canali, tipicamente presenti per l'intero programma e che fungono da base per il mix. I beds sono elementi in continuità con i workflow multicanale precedenti. Dunque da un punto di vista creativo sono gli oggetti che portano con sé nuovi parametri di controllo rispetto al passato.

Quanto è stato fatto nel processo di ibridazione dell'approccio *channel-based* con l'approccio *object-based* è stato quello di corredare i beds di una serie di metadati in grado di favorire l'adattabilità del contenuto di ciascun bed in relazione al sistema di riproduzione [22].

Considerando la moltitudine di configurazioni di sala attraverso le quali i prodotti cinematografici possono essere riprodotti, i parametri che seguono sono strumenti abilitati dallo standard per consentire ai *content creators* di gestire già in fase di mix delle operazioni di downmix o di ottimizzazione preventiva del proprio lavoro, al fine di preservare l'intento creativo originale attraverso i più disparati *playback environment*.

Interattività e personalizzazione nella trasmissione

Remap Coefficients - Coefficienti di Remapping

I coefficienti di remapping sono coefficienti che specificano come mappare i canali originali del mix su una diversa configurazione di destinazione. Sono dunque un insieme di valori che indicano quanto guadagno deve essere applicato a ciascun canale audio di un bed per generare ciascuna uscita della configurazione di destinazione.

Se all'interno di un bed vengono forniti i coefficienti di remapping, come si vedrà in seguito, un set di metadati aggiuntivi dovrà identificare le condizioni in cui tale remapping deve essere attuato, ovvero il *target environment* di riferimento. **ST 2098-1**

Dunque il *content creator* è nella possibilità di specificare direttamente quali sono i valori di remapping attraverso cui vuole che i canali vengano configurati in un diverso target environment. Una sorta di remix in caso di diverso sistema di riproduzione.

Conditional Bed - Bed Condizionali

Un fonico di mix potrebbe voler includere uno o più bed alternativi all'interno del programma audio immersivo per le diverse configurazioni del campo sonoro di destinazione. I metadati che accompagnano un Conditional Bed devono dunque supportare l'identificazione di bed alternativi e le condizioni (*target environment*) in cui dovranno essere utilizzati. (Secondo lo standard le configurazioni di destinazione supportate devono essere almeno per i sistemi 5.1 e 7.1 DS - **ST 2098-1**).

A differenza dei Coefficienti di Remapping che indicano nuovi parametri di mix che si attivano in determinati *playback environment* i Bed Condizionali aggiungono veri e propri nuovi dati audio al bitstream.

Dunque se un determinato Bed è stato realizzato per essere riprodotto in una sala 7.1, ma si prevede che possa essere riprodotto anche in sale con diverse configurazioni è possibile inserire diversi bed già mixati per la rispettiva destinazione che vengono attivati a seconda di dove verranno riprodotti.

Dunque un set di metadati aggiuntivi verrà inserito al fine di indicare non solo se sono presenti dei Bed Condizionali o meno, ma anche in quali casi essi andranno attivati e i relativi parametri che consentano di recuperare gli elementi audio mixati ad hoc.

Pertanto le soluzioni attivate dallo standard al fine di consentire ad un *content creator* di preservare il proprio intento creativo attraverso i diversi *playback environment* sono le seguenti:

- Specificare dei coefficienti di remix per configurazioni note.
- Utilizzare dei Bed alternativi preparati ad hoc per determinate configurazioni spaziali.
- Affidarsi alle impostazioni automatiche con cui sono configurati i renderer.

2.4.3.2 Object Metadata ST 2098-1

Gli oggetti di un programma audio immersivo sono la vera novità in termini di opportunità creative rispetto al passato.

Quando si parla di *object metadata* ci si riferisce ai *Dynamic Metadata*

ovvero parametri in grado di cambiare nel tempo, salvo diverse indicazioni. Un oggetto senza il suo relativo set di metadati sarebbe privo di significato poiché è attraverso questi parametri che un renderer interpreta le volontà di mix e le ottimizza a seconda della configurazione a disposizione.

Qui di seguito verranno analizzati i principali set di metadati di un Oggetto attivati dallo standard coerentemente alla loro funzione [22].

Interattività e personalizzazione nella fase di Broadcasting

Conditional Object

Similmente ai *Conditional Beds* se la posizione specificata per un Oggetto non dovesse essere supportata da una certa configurazione, un fonico di mix potrebbe voler includere uno o più Oggetti alternativi in sostituzione. I metadati di un Oggetto Condizionale devono dunque supportare l'identificazione di oggetti alternativi e le condizioni (*target environment*) in cui devono essere utilizzati. **ST 2098-1**

Object Description

Descrive una caratteristica dell'Oggetto audio che può essere utilizzata per indirizzarlo durante l'elaborazione. I tipi supportati includono: dialogo, musica ed effetti. **ST 2098-1**

Object Lifetime

Specifica gli intervalli di tempo (inizio e durata per un singolo intervallo) in cui un oggetto audio è attivo (o presente), ovvero gli intervalli temporali in cui i metadati dell'oggetto audio possono influire

sul rendering di quell'Oggetto. ST 2098-1

E' un parametro opzionale che però consente di dare delle informazioni al renderer in merito alla durata degli oggetti in questione, consentendogli di "sapere cosa aspettarsi" da quella porzione di bitstream.

Controlli artistici per il rendering di oggetti nell'audio cinematografico

Object Gain - Guadagno di un Oggetto

Definisce il guadagno di un oggetto, parametro chiave nell'operazione di panning, il quale può variare nel corso della durata di un frame e deve consentire che possa venir applicato un guadagno anche pari a 0 dB.

Object Position - Posizione di un oggetto

Definisce l'**intento posizionale** di un oggetto nello spazio, come precedentemente espresso, si parla di intento perché non è possibile definire con precisione la posizione di un oggetto, poiché quest'ultima dipende dalle operazioni di rendering e dalla relativa configurazione di sala. Ecco perché le coordinate spaziali di un oggetto sono specificate all'interno di un sistema di riferimento allocentrico.

Object Snap - Snap Tolerance

Indica il grado secondo cui la conservazione del timbro dell'oggetto ha priorità sulla conservazione della sua posizione. Questa proprietà indica attraverso valori agli estremi rispettivamente che "preservare il timbro

dell'oggetto ha la massima priorità" e "preservare la posizione dell'oggetto ha la massima priorità". **ST 2098-1**

Il timbro è quella caratteristica che consente di distinguere due suoni con la stessa durata, intensità ed altezza, ma che non fornisce informazioni in merito alle dimensioni o alla posizione della sorgente.

Dunque se in un set di metadati ObjectSnap è settato ad 1, significa che per quell'oggetto o gruppo di oggetti restituire fedelmente la percezione del timbro è di primaria importanza. Pertanto per preservare il timbro di quel suono lo si "aggancerà" (=to snap) all'altoparlante più vicino.

Altrimenti se è l'intento posizionale di quel suono che si vuole preservare, ObjectSnap verrà settato a zero e quel suono verrà allocato al maggior numero di altoparlanti possibile.

Questo parametro è particolarmente importante anche perché dà delle linee guida al renderer in situazioni di criticità. Lo scenario infatti nel quale ci si ritrova spesso a lavorare quando si parla di *Immersive Sound* è che non si è mai esattamente certi di quale sarà il *target environment* di destinazione. Pertanto se per caso il renderer si trovasse nella condizione di dover posizionare un suono in un punto dello spazio in cui non è previsto nessuno speaker, i metadati ObjectSnap gli danno un importante metro di valutazione in merito a quale scelta operare in tale situazione. Ovvero se per allocare quel determinato suono sarà più importante preservare il timbro, la scelta ricadrà sull'altoparlante più vicino, viceversa, se invece è stato segnalato che di quel suono era importante la sua localizzazione.

Questo è quello che Brian Vessa nel webinar SMPTE di presentazione dello standard 2098-1 definisce:

"(..) uno dei trade-offs che si ritrova ad affrontare quando si cerca di riprodurre il suono immersivo nel mondo reale."

Object Spread - Diffusione di un Oggetto

Definisce la dimensione in termini sonori di quell'Oggetto e la sua forma all'interno dello spazio tridimensionale, se si vuole che la dimensione di quel suono sia puntiforme ObjectSpread viene settato a 0, se invece si vuole che quel suono occupi l'intero cinema o che comunque il suono precettivamente occupi un'area più che una precisa posizione, allora il parametro sarà settato ad 1.

Object Decor - Decorrelazione di un Oggetto

E' il parametro che elabora i segnali sorgente, utilizzati per riprodurre un evento uditivo, al fine di alterare la loro relazione preservando il suono originale per ogni singolo segnale. Il valore minimo indica che non si intende applicare l'effetto di decorrelazione e il valore massimo indica che si intende applicare l'effetto di decorrelazione massimo. **ST**

2098-1

L'immagine sonora durante la riproduzione di un oggetto audio attraverso due o più altoparlanti in un sistema sonoro immersivo può essere percepita come localizzata o diffusa a seconda che i segnali sorgente che rappresentano l'oggetto audio siano rispettivamente correlati o decorrelati. La riproduzione di un suono con più segnali correlati produce un'immagine sonora facilmente individuabile. La riproduzione di un suono attraverso più sorgenti connesse ma non

correlate produce un'immagine sonora più ampia e diffusa. Questo può essere definito un effetto di decorrelazione. Descrizione Generale - **ST 2098-1**

Questo parametro viene dunque utilizzato piuttosto sovente nell'ambito dell'audio immersivo poiché per via della configurazione tridimensionale a disposizione è facile che si voglia lavorare sulla possibilità di localizzare un suono puntualmente (ObjectDecor=0) o viceversa che si voglia restituire la percezione di diffusione dello stesso (ObjectDecor=1).

ObjectDecor viene inoltre adoperato quando un determinato suono è renderizzato utilizzando più altoparlanti contemporaneamente al fine di evitare che nella riproduzione possa incorrere in distorsioni quali il "comb-effect", ove il suono diretto e riflesso prodotto da più sorgenti all'interno di uno spazio chiuso finisce per combinarsi in fase o in controfase generando ritardo o totale cancellazione del suono risultante.

Zone Metadata

Obect Zone Control: quando presente, indica il grado di esclusione dal rendering di un determinato insieme di altoparlanti - zone - (ST 2098-1).

- L'effettiva mappatura delle zone rispetto agli altoparlanti sarà definita, per un cinema, al momento della configurazione del renderer.
- Le zone devono essere configurate come una partizione non sovrapposta dell'insieme di tutti gli altoparlanti disponibili.

Dunque i metadati Object Zone Control sono utilizzati per escludere dal panning degli oggetti alcune zone (set di altoparlanti) della sala cinematografica. Un'applicazione tipica di questo parametro si presenta nell'operazione di panning di un oggetto attraverso la sala, ovvero quando si vuole che un suono vada dallo schermo, al fondo del cinema passando per il soffitto ma escludendo gli altoparlanti dei muri laterali; oppure ancora, viceversa, quando si vuole che tale suono durante il panning includa solo gli altoparlanti laterali ed escluda la zona del soffitto. Facendo infatti tutti gli altoparlanti di una configurazione parte del cubo unitario di riferimento per le coordinate nello spazio, è necessario specificare quali zone si vogliono escludere, altrimenti in fase di rendering potrebbero essere incluse automaticamente.

ObjectZoneDefinition19

ObjectZoneDefinition19
ElementID ElementSize ZoneInfo (per block) ZoneGain (per zone)

Figura 33 L'immagine anticipa gli approfondimenti del prossimo paragrafo poiché è una rappresentazione logica dell'elemento bitstream ObjectZoneDefinition19 che indica come per ciascuna zona sia indicato un intervallo di valori in merito al grado d'inclusione di ogni zona. [28]

E' un parametro aggiuntivo rispetto all'esistente formato Dolby Atmos, che come già introdotto è stato preso a modello per una parte della stesura degli standard. ObjectZoneDefinition19 è un *sub element*, che verrà approfondito più avanti, di un oggetto ed è

utilizzabile solo nella misura in cui il renderer riconosce che il *playback environment* di riferimento supporta le 19 zone di sala cui si riferisce questo parametro (configurazione quest'ultima appartenente ad Auro3D). Se dunque il renderer riconosce la configurazione di sala corrispondente, rimpiazza il parametro Object Zone Control visto in precedenza con i parametri contenuti in ObjectZoneDefinition19.

Zone Gain: rappresenta il grado di inclusione di una zona nella riproduzione di un suono. Questa proprietà deve supportare un intervallo di valori. L'intervallo deve includere valori estremi che indicano "completamente abilitato" e "completamente disabilitato". Un Oggetto può avere un valore di guadagno separato per ciascuna zona di base. **ST 2098-1**

Number	Description
1	Base layer screen Loudspeakers left of center
2	Base layer center screen Loudspeakers
3	Base layer screen Loudspeakers right of center
4	Height layer screen Loudspeakers left of center
5	Height layer center screen Loudspeakers
6	Height layer screen Loudspeakers right of center
7	Base layer rear wall Loudspeakers left of center
8	Base layer center rear wall Loudspeakers
9	Base layer rear wall Loudspeakers right of center
10	Height layer rear wall Loudspeakers left of center
11	Height layer center rear wall Loudspeakers
12	Height layer rear wall Loudspeakers right of center
13	Base layer left wall Loudspeakers
14	Height layer left wall Loudspeakers
15	Base layer right wall Loudspeakers
16	Height layer right wall Loudspeakers
17	Top layer Loudspeakers left of center
18	Top layer center ceiling Loudspeakers
19	Top layer Loudspeakers right of center

Zones

Figura 34 . La tabella tratta dal Webinar SMPTE di presentazione del nuovo standard ST 2098-1 illustra gli altoparlanti inclusi nelle 19 zone di sala cui si riferisce il parametro ObjectZoneDefinition19. [28]

2.5 Formato Bitstream 2098 - 2

2.5.1 Struttura di un Immersive Audio Bitstream

STRUTTURA

All'interno dello standard ST 2098-2 viene descritta la struttura di un Bistream Audio Immersivo comprensiva di quelli che sono i parametri necessari alla standardizzazione del formato [28]. E' stato già infatti approfondito il fatto che questi standard sono stati redatti al fine di uniformare la produzione di un unico input in grado di essere facilmente codificato e riprodotto attraverso i vari dispositivi sul mercato.

Un Bistream Audio Immersivo è composto di una sequenza di **Bitstream Frames** i quali a loro volta sono caratterizzati di un **Preambolo** e di un **IA Frame**.

Il preambolo è un elemento ereditato dall'implementazione Dolby dell'audio immersivo che era alla base degli standard SMPTE e non ha alcun ruolo oltre al fatto che doveva essere ammesso per rendere il contenuto esistente compatibile con lo standard. E' stato lasciato indefinito, salvo per il fatto che, se è presente, dovrebbe essere vuoto e lo standard gli consente anche di non essere presente.

E' invece l'IA Frame a contenere tutti i metadati e i dati audio che definiscono il programma audio immersivo. Come si può osservare dalla figura 35 al suo interno è possibile trovare più di un oggetto e più di un elemento bed i cui rispettivi dati audio son trasportati separatamente dai metadati.

I set di metadati che compongono un IA Frame possono essere a loro volta scomposti in **SubBlocks**. I SubBlocks di un elemento bed se presenti

contengono i coefficienti di remix per quel determinato Bed. I subBlocks di un oggetto sono invece suddivisioni temporali dei metadati di quell'oggetto che ne descrivono eventuali cambiamenti repentini nel corso della durata di un frame, come ad esempio i metadati di posizione di un oggetto che può variare la sua posizione nello spazio più di una volta nel corso di durata di un singolo frame. [33]

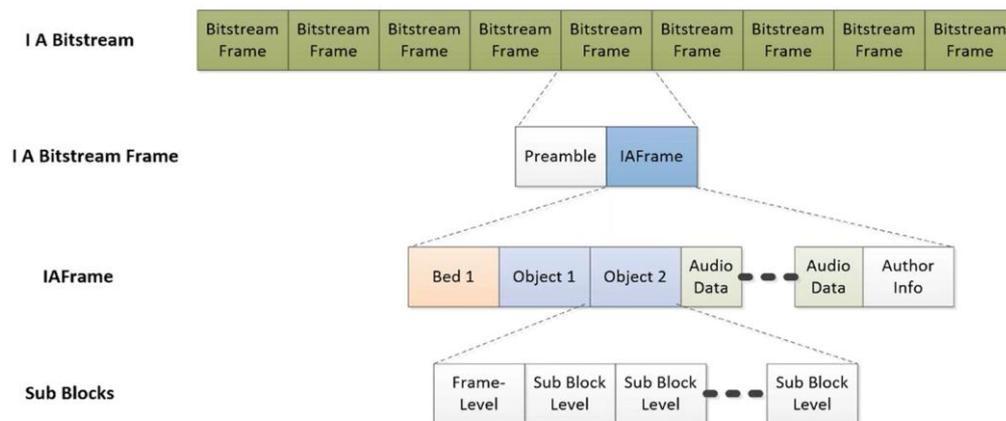


Figura 35 Struttura di un Immersive Audio Bitstream secondo i criteri di standardizzazione SMPTE 2098-2. [33]

RETROCOMPATIBILITÀ

Analizzando la precedente struttura da un punto di vista logico²⁸ si può osservare che il massimo livello di un IA Bitstream è rappresentato da un IA Frame e tutti gli elementi che lo compongono derivano dal seguente modello, ovvero tutti i suoi elementi presentano questo pattern:

- **ElementID:** identifica mediante un ID univoco l'elemento e la sua

²⁸ Lo scopo della Vista Logica è specificare i requisiti funzionali di un sistema, descrivendone gli oggetti che lo compongono e le rispettive relazioni.

sintassi.

- **ElementSize**: fornisce la lunghezza di quel byte, indicando dove finisce quell'elemento.
- Contents (seguono i rispettivi contenuti per quell'elemento).

La combinazione di ElementID e ElementSize consentono al formato una delle sue qualità principali, la **retrocompatibilità**, di cui qui di seguito si dettagliano le caratteristiche:

- Molti degli elementi presenti nello standard erano già alla base del bitstream Dolby Atmos esistente, pertanto il lavoro fatto sul formato è stato ideato al fine di rendere i cambiamenti compatibili con quest'ultimo.
- Al fine di garantire compatibilità tra i vari dispositivi di riproduzione la combinazione di Element ID ed ElementSize permette il *Key-Length Decoding*. Tale meccanismo di decodifica si basa sulla lettura dell' ID univoco (Key) di un elemento da parte del decoder che stabilisce se è in grado di riconoscere o meno quell'elemento, se la risposta è affermativa procede con la decodifica, altrimenti passa alla lettura della lunghezza in byte (Length) di quell'elemento al fine di poter scorrere il bitstream fino all'elemento successivo. Proprio come avviene per il funzionamento di un file MXF. Per esempio se un decoder di un sistema surround non riconosce l'Element ID di un elemento riproducibile solo in un sistema immersivo esso semplicemente verrà saltato nel corso della lettura, senza inficiare l'intero processo di decodifica.
- La logica strutturale sopra esposta si coniuga con una delle regole

principali seguite nella creazione del formato, ovvero che *per rendere i contenuti di un bitstream compatibili con l'aggiunta di nuovi elementi, non sono gli elementi pre-esistenti a venir cambiati bensì quelli nuovi ad esser resi opzionali.* (2018 - Stan Cosette – Webinar SMPTE). Dunque non ci sono nuovi elementi **richiesti** all'interno dello standard, gli elementi nuovi inseriti sono tutti **opzionali**.

- Non sono stati applicati cambi di sintassi agli elementi esistenti.

2.5.2 Immersive Audio (IA) Frame

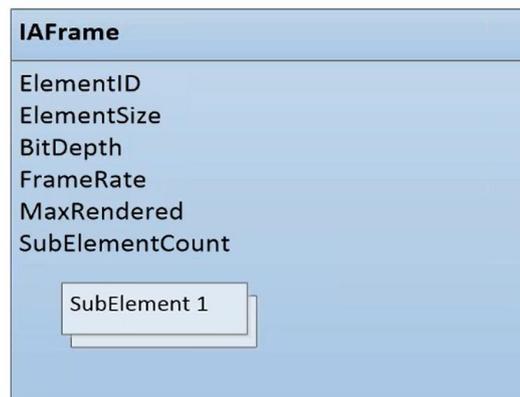


Figura 36 Vista logica di un IA Frame con un insieme ristretto dei parametri che lo caratterizzano. [28]

Un IA Frame presenta sempre i due parametri identificativi **ElementID** ed **ElementSize** e pertanto tutti gli elementi in esso contenuti ereditano lo stesso modello presentando a loro volta questi due parametri.

Un IA Frame è caratterizzato inoltre da una serie di parametri opzionali quali **BitDepth**, profondità di bit, **FrameRate**, **MaxRendered**, che dà un'indicazione al renderer di quanti elementi dovrà renderizzare per

quel frame, e **SubElementCount** il quale determina quanti SubElement, oggetti o bed, sono presenti consentendo al *parser* di trovarli all'interno del frame.

Le specifiche dello standard ST 2098-2 si riferiscono sempre ad un singolo IA Frame senza porre condizioni sul contenuto dei frames adiacenti nel bitstream, in questo modo ciascun IA Frame è decodificabile indipendentemente.

Alcune applicazioni possono inoltre porre dei vincoli sul bitstream affinché gli IA Frames di cui è composto contengano parametri consistenti. Ad esempio può essere imposto che *profondità di bit* e *framerate* rimangano uguali nel corso dello stesso bitstream, altrimenti diventerebbe difficile il processo di decodifica.

Qui di seguito si riporta un diagramma UML²⁹ che illustra di quali elementi sia composto un IA Frame e quali siano i rapporti che intercorrono tra i vari elementi.

²⁹ I diagrammi UML permettono la modellazione della struttura statica e del comportamento dinamico di un sistema. Il sistema è rappresentato come un insieme di oggetti di cui vengono riportati attributi, operazioni e reciproche relazioni.

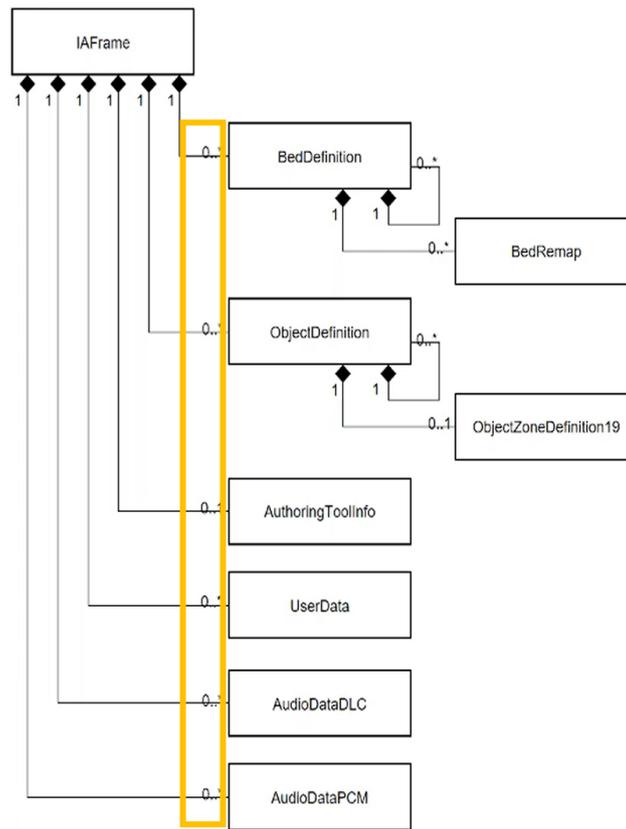


Figura 37 Grafico UML di un Immersive Audio Frame che illustra i rapporti che intercorrono tra i vari elementi che lo compongono. [22]

Osservazioni

- La prima cosa che si può osservare dalla rappresentazione è che un IA Frame contenente tutti gli altri elementi ricopre il livello più alto nella gerarchia.
- Gli zeri segnalati nel rettangolo arancione in figura stanno ad indicare che tutti gli elementi contenuti in un IA Frame sono opzionali, dunque secondo la specifica nessuno di essi è obbligatoriamente richiesto all'interno del bitstream, questo al fine di consentire il principio di estensibilità poc'anzi espresso.

- Solo gli elementi bed e gli oggetti contenuti in un IA Frame possono essere composti di altri elementi, i cosiddetti SubElements. Un BedElement può essere composto dunque da zero o più BedElement, ovvero i Bed Condizionali, e da zero o più BedRemap. Mentre un oggetto può essere composto da zero o più oggetti, Conditional Objects, e da zero fino ad un ObjectZoneDefinition19, un parametro che può sostituire alcune informazioni di controllo.
- Come si osserva dal diagramma gli altri elementi AuthoringToolInfo, UserData, AudioDataDLC, AudioDataPCM non possono contenere subElement.
- Gli elementi che si trovano ai livelli superiori hanno una relazione "genitore-figlio" con gli altri elementi, dunque tutti gli elementi sono figli dell'IA Frame che si trova al livello più alto. In generale ogni elemento contenuto in un altro elemento è figlio di quest'ultimo.
- Un elemento figlio, se presente, può sostituire una porzione o la totalità di un elemento genitore (ad esempio: BedRamp, attraverso i suoi coefficienti di remapping sostituisce una parte dell'elemento genitore, Conditional Bed invece sostituisce per intero il Bed originario).

Nei paragrafi successivi si procederà con l'analisi da un punto di vista logico della struttura degli elementi principali di un IA Bitstream, ovvero i bed e gli oggetti.

2.5.3 Struttura logica di un elemento Bed

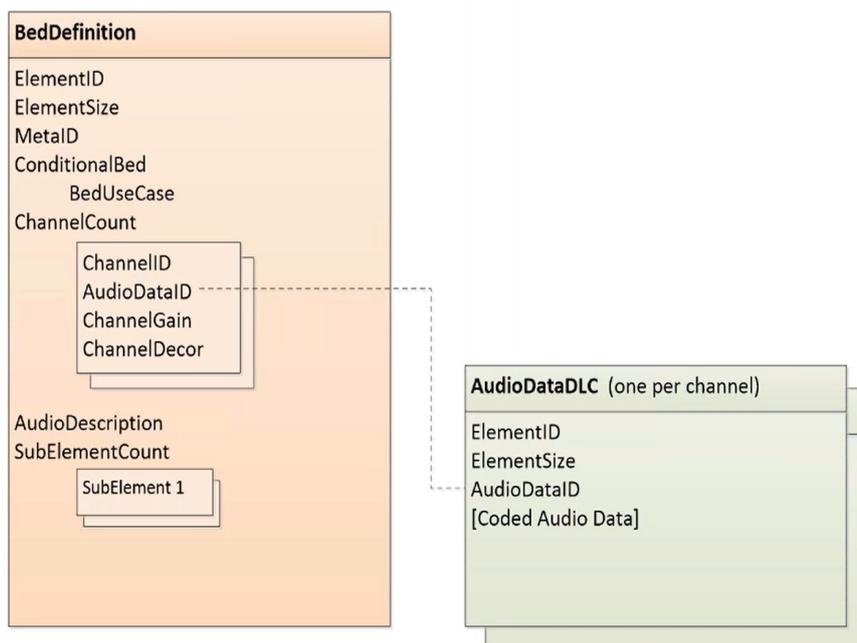


Figura 38 Vista logica di un elemento Bed con un insieme ristretto dei parametri che lo caratterizzano. [28]

Qui di seguito verranno analizzati i principali attributi che costituiscono la struttura di un elemento Bed [28].

- **ElementID:** è il *Bed Identifier*, cioè un ID univoco che identifica l'elemento. Due bed non devono avere lo stesso identificatore in nessun istante di tempo. Ciò consente di fare riferimento a bed multipli se lo si desidera. ST 2098-1
- **ElementSize:** indica la dimensione in byte dell'elemento bed, al fine di agevolare il processo di decodifica.
- **MetaID:** parametro utilizzato per identificare un elemento particolare all'interno di un gruppo di ElementID, per esempio se ci sono dei bed multipli all'interno del bitstream, MetaID consente di

identificare un bed rispetto ad un altro.

- **ConditionalBed:** parametro ampiamente analizzato in precedenza che fornisce ai *content creator* strumenti di adattabilità del mix. Esso dunque identifica se un bed è condizionale o meno. Qualora esso lo sia deve recare con sé un altro parametro.

BedUseCase: indica in quale condizione (*target environment*) tale Bed condizionale deve essere attivato.

- **ChannelCount:** parametro che indica quanti canali ci sono all'interno dell'elemento bed (uno o più di uno, ma mai più di quanti ChannelID presenti). Per ogni canale viene associato un altro set di metadati così composto:

ChannelID: indica la destinazione del canale (*routing destination*), ovvero dove dovrà essere riprodotto il suo contenuto.

AudioDataID: indica qual è l'elemento che trasporta i dati audio associati al presente canale.

ChannelGain: indica i parametri di guadagno per questo canale.

ChannelDecor: indica il valore di decorrelazione per questo canale.

Come si può osservare dal riquadro verde in figura, per ogni canale è presente il parametro **AudioDataID** che punta all'elemento che trasporta i dati audio corrispondenti. Ci saranno pertanto tanti elementi (Audio Data DLC nell'immagine) quanti sono i canali. Ogni

elemento che trasporta dati audio contiene dunque un ID identificativo, un parametro che indica le dimensioni del pacchetto e un parametro AudioDataID cui è connesso il corrispettivo canale.

- **Audio Description:** è un parametro che fornisce informazioni in merito alla natura del contenuto di quell'elemento, può essere sia un campo di testo libero che invece presentare categorie classiche come "dialoghi", "musiche", "effetti"(ecc..).
- **SubElementCount:** è un contatore che indica quanti subElement sono presenti nella struttura bed. Nel caso dei bed un subElement può essere costituito da Bed Condizionali o Coefficienti di Remapping.

Al fine di proseguire nella comprensione della struttura gerarchica che caratterizza gli elementi appartenenti al formato ed al fine di comprendere i nuovi strumenti forniti ai creativi dallo standard per l'ottimizzazione dei propri lavori segue l'analisi degli attributi del subElement *BedRemap*.

2.5.3.1 Bed Remap

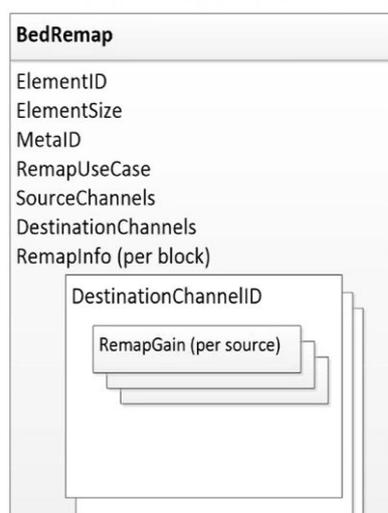


Figura 39 Vista logica di un sotto-elemento BedRemap con l'insieme dei parametri che lo caratterizzano [28].

Come visto in precedenza all'interno degli standard è previsto che i *content creator* abbiano la facoltà di specificare dei coefficienti di remapping, una specie di remix di alcuni Bed, per adattare il mix alle esigenze di diversi playback environment.

Se pertanto nel corso della decodifica di un file all'interno di un elemento bed è presente un SubElement il cui parametro RemapUseCase è verificato, ovvero il parametro che indica in quali condizioni devono essere attivati i coefficienti di remapping, al contenuto del bed originale viene applicato il downmix indicato.

Pertanto gli attributi in esso contenuti saranno i seguenti:

- **ElementID** ed **ElementSize**: parametri ereditati dall'elemento genitore secondo quanto illustrato prima.
- **RemapUseCase**: specifica qual è la configurazione a cui deve essere applicato il remix
- **SourceChannels**: parametro che indica i canali cui è destinato l'elemento bed originale.
- **DestinationChannels**: indica i canali di ridistribuzione del bed originale secondo i coefficienti di remapping riportati nel parametro seguente.
- **RemapInfo**: parametro costituito dai contatori per i canali *Source* e *Destination* attraverso cui viene creato un loop. Per ogni canale destinazione è assegnato un canale sorgente cui è applicato un certo valore di guadagno (RemapGain(per source)). L'output risultante sarà il nuovo canale destinazione cui è stato applicato il richiesto set di valori di remapping, reindirizzato a destinazione grazie al

parametro **DestinationChannelID**. Dunque ciò che viene creato è essenzialmente una matrice attraverso cui è possibile applicare un "remix" al bed originale.

2.5.4 Struttura Logica di un Oggetto

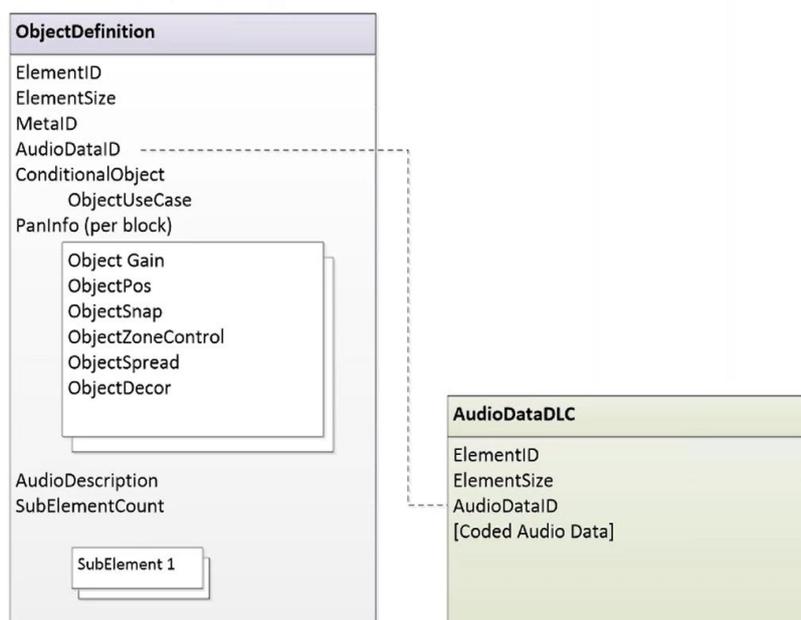


Figura 40 Vista logica di un Oggetto con un insieme ristretto dei parametri che lo caratterizzano. [28]

Un oggetto senza il suo set di metadati è impossibile da decodificare ecco perché gli attributi presenti nella sua struttura sono molto importanti nell'analisi del formato Audio Immersivo. Ogni oggetto infatti non è indirizzato in maniera statica ad una destinazione esso è al contrario renderizzato ad una posizione nello spazio che varia a seconda del sistema di riproduzione. I metadati associati dunque, come analizzato nei paragrafi precedenti, forniscono le informazioni necessarie a questo tipo di operazione.

Inoltre una delle maggiori distinzioni che è possibile segnalare in merito

alla struttura di un oggetto è che ad essi è associabile un solo elemento che trasporta dati audio contrariamente all'elemento bed che può averne tanti quanti sono i canali da esso trasportati.

Struttura di un oggetto:

- **ElementID:** fornisce un'identificazione univoca per un oggetto. Due oggetti non possono avere lo stesso identificatore in nessun istante di tempo. Un identificatore di oggetto è statico per la durata dello stesso. La cardinalità dello spazio dell'identificatore deve essere sufficientemente grande da supportare il numero più alto di oggetti audio simultanei esistenti in qualsiasi momento durante la presentazione. **ST 2098-1**
- **ElementSize:** dimensione in byte dell'oggetto.
- **MetaID:** parametro utilizzato per identificare un elemento particolare all'interno di un gruppo di ElementID, per esempio in un gruppo di oggetti un metaID consente di identificare uno specifico oggetto.
- **AudioDataID:** (*Waveform Reference*) parametro che fa riferimento all'elemento che trasporta i Dati Audio per quell' oggetto. Il riferimento deve consentire un'identificazione univoca.
- **ConditionalObject:** parametro che indica se per l'oggetto in questione è presente un Oggetto Condizionale, ovvero un oggetto sostitutivo creato apposta per determinate configurazioni (*target environment*). Pertanto il rapporto tra AudioDataID e AudioDataDLC è condizionale ed è espresso da questo parametro.

ObjectUseCase: indica quali configurazioni richiedono l'attivazione del ConditionalObject.

- **PanInfo:** è il blocco che trasporta tutti i Dynamic Metadata, analizzati in precedenza, renderizzati per ciascun oggetto:

**ObjectGain - ObjectPosition - ObjectSnap -
ZoneControl - ObjectSpread - ObjectDecor**

- **AudioDescription:** è il parametro che fornisce informazioni in merito alla natura del contenuto di quell'oggetto (effetti ecc..).
- **SubElementCount:** è il parametro che segnala quanti SubBlock sono contenuti, ovvero quante suddivisioni interne al frame ci sono per quell'oggetto. Si tratta infatti di suddivisioni temporali dei metadati che descrivono cambiamenti repentini di un oggetto all'interno dello stesso frame. Tipicamente se siamo a 24 fps la durata di un subBlock può essere di 5 ms, circa 8 subBlock per frame.

2.5.5 Esempi Applicativi

Per meglio comprendere i temi sin qui esposti, qui di seguito si riportano alcuni esempi illustrati da Stan Cosette nel corso del Webinar di presentazione del nuovo standard SMPTE 2098-2 [28] al fine di comprendere le relazioni che intercorrono tra gli elementi di un IA Frame in alcuni casi tipici.

I diagrammi che seguono sono la rappresentazione logica di una versione semplificata degli elementi che costituiscono il bitstream, in alto è

riportato il nome dell'elemento a cui seguono un sottogruppo dei parametri contenuti, associati al valore cui sono impostati. La linea verticale che li unisce indica invece i rapporti gerarchici che intercorrono tra i vari elementi, quelli più in alto contengono gli elementi sottostanti secondo un rapporto padre-figlio.

L'elemento che si trova sempre in cima alla gerarchia è l'IA Frame il cui *ElementID* è settato sempre ad 8 perché questo è l'ID riportato nella specifica per gli Immersive Audio Frame, così come *Version* è sempre impostato ad 1 poiché quello è il valore predisposto nella specifica per questo formato bitstream.

Il valore *MaxRendered* invece indica il massimo numero di elementi, bed e oggetti, che dovranno essere renderizzati contemporaneamente per quel frame ad un dato istante temporale. Questo valore è dunque un indice per il renderer di quanti elementi dovrà gestire per quel frame. Per evitare di andare in contro ad un eventuale sovraccarico tale parametro è comunque impostato in modo tale da non poter eccedere il valore di 128 elementi, se la frequenza di campionamento dei dati audio è impostata a 48 Khz, oppure di 64 elementi se la frequenza di campionamento è impostata a 96KHz.

All'interno dell'IA Frame, così come negli altri elementi, si trova il parametro *SubElementCount* che indica quanti elementi "figlio" sono contenuti in quell'elemento "padre". Tale conteggio vale solo per i diretti sotto-elementi di quell'elemento, dunque se un bed contenuto in un IA Frame contiene dei sotto-elementi questi non saranno riportati nel conteggio dei sotto-elementi dell'IA Frame in cima alla gerarchia.

All'interno dello standard inoltre ci sono alcune limitazioni in merito al possibile numero di elementi "figlio", non è consentito infatti avere “figli dei figli” ovvero elementi contenuti in eventuali sotto-elementi di un oggetto o di un bed.

2.5.5.1 Esempio di una struttura tipica di oggetti e bed attivi simultaneamente

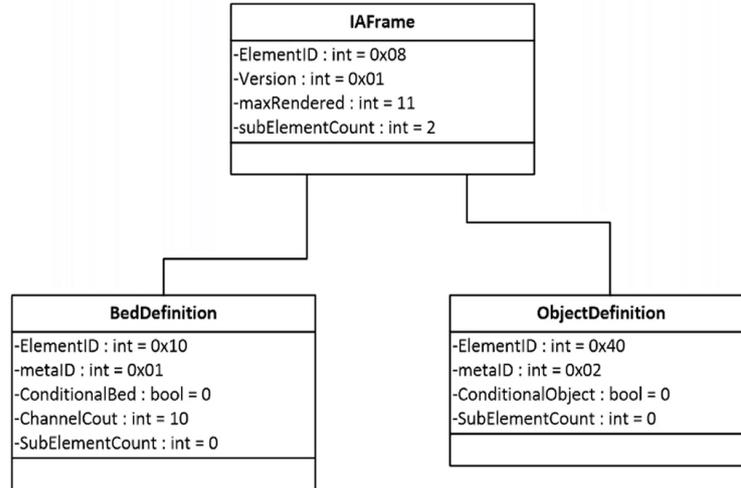


Figura 41 Struttura logica di un IA Frame contenente un elemento Bed ed un Oggetto attivi contemporaneamente nel frame. [28]

Questo è uno degli esempi che illustra una delle strutture più tipiche in cui elementi Bed e Oggetti sono coinvolti contemporaneamente.

All'interno dell'IA Frame, come possiamo osservare anche dalla struttura del diagramma, `SubElementCount` è impostato a 2, dunque esso contiene due elementi figlio.

Il fatto che nessuno dei due sotto-elementi contenga elementi condizionali, come si può osservare dal fatto che `ConditionalBed` e `ConditionalObject` sono settati a zero, li rende entrambi attivi e pertanto renderizzabili contemporaneamente. Motivo per il quale la voce `MaxRendered` nell'IA Frame è impostata su 11, perché il valore massimo di elementi da renderizzare contemporaneamente all'interno di questo frame è pari ad un oggetto insieme ai 10 canali dell'elemento bed.

2.5.5.2 Esempio di una struttura contenente un *BedRemap*

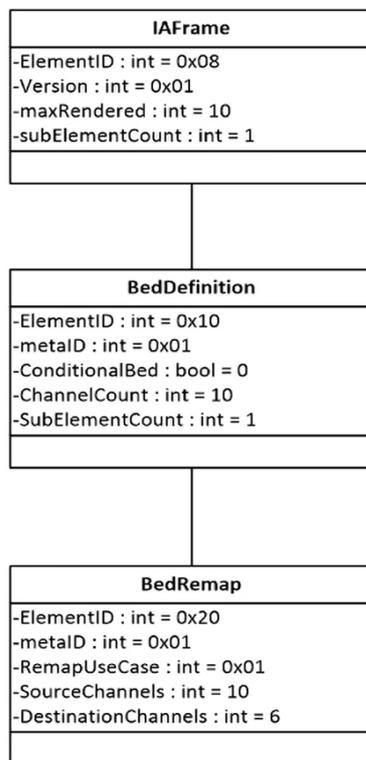


Figura 42 Vista logica di un IA Frame che esemplifica una possibile struttura contenente un *BedRemap*. [28]

Nell'esempio sopra riportato è possibile osservare sin da subito mediante la rappresentazione grafica delle gerarchie che l'elemento IA Frame contiene un sotto-elemento *BedDefintion* a sua volta contenente un sotto-elemento *BedRemap*.

Infatti la voce *subElementCount* sia per l'elemento Bed che per l'IA Frame riporta il valore 1, questo perché come spiegato all'inizio del paragrafo, tale contatore si riferisce solo ai sotto-elementi diretti del blocco genitore.

La peculiarità di questo esempio è legata al fatto che il sotto-elemento dell'elemento *BedDefinition* è rappresentato da un remapping del Bed in

questione. Pertanto per poter dar luogo a tale remapping è necessario valutare se il parametro `RemapUseCase`, che indica qual è il *playback environment* a cui è rivolto il remapping, è verificato. In questo caso il fatto che sia settato ad un 1, coerentemente a quanto riportato nella specifica, significa che la configurazione di sala corrispondente è un sistema surround 5.1. Pertanto il remapping avrà luogo **se e solo se** la configurazione di sala di destinazione è del tipo surround 5.1.

Seguono poi i parametri `SourceChannels` e `Destination Channels` per impostare la matrice di remapping. Come si può notare il valore di *Source* è settato a 10, corrispondente al numero di canali da ri-mappare del bed originale. Mentre il parametro *Destination* è impostato a 6, valore esattamente corrispondente al numero di canali presente in un sistema 5.1.

Ritornando dunque al blocco rappresentante l'IA Frame possiamo ora comprendere che il valore `MaxRendered` è settato a 10 perché nell'eventualità che venga renderizzato o il bed originale (10 canali) o il sotto-elemento `BedRemap` (6 canali), 10 sarà il numero massimo di canali che il renderer dovrà gestire contemporaneamente.

2.5.5.3 Esempio di una struttura contenente un *ConditionalBed*

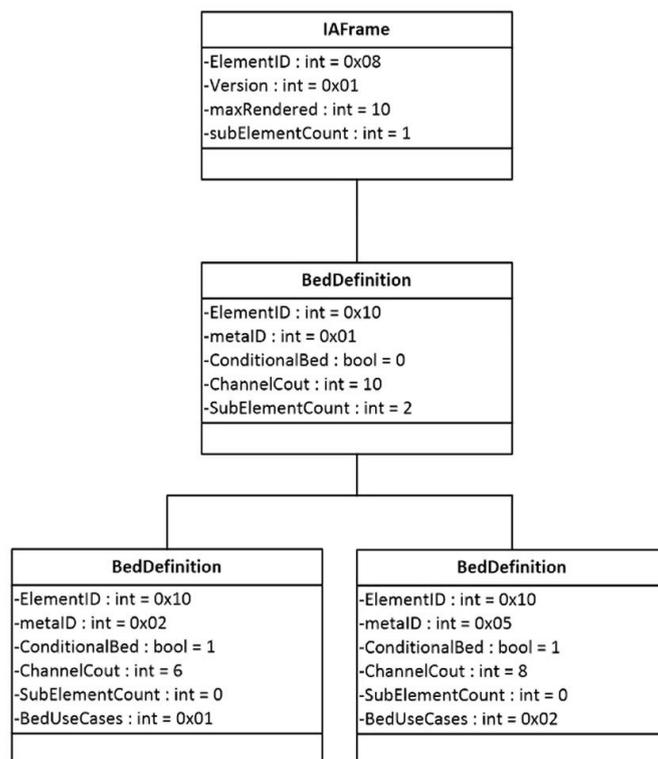


Figura 43 Vista logica di un IA Frame che esemplifica una possibile struttura contenente due *ConditionalBed*. [28]

Il Terzo ed ultimo esempio illustra invece una struttura contenente dei Bed Condizionali.

Come detto in precedenza un elemento condizionale rimpiazza completamente l'elemento originale solo se le condizioni di utilizzo lo richiedono, ovvero se il renderer segnala che la sua configurazione di sala corrisponde a quanto riportato nel parametro *BedUseCase* (o nel caso di un oggetto, *ObjectUseCase*).

Come è possibile osservare l'elemento IA Frame possiede un solo elemento figlio, l'elemento *BedDefinition*. Quest'ultimo a sua volta presenta due sotto-elementi i quali dichiarano di essere dei Bed Condizionali, come si

può osservare dal parametro *ConditionalBed* impostato ad 1. Questo significa che tali sotto-elementi potranno sostituire completamente il bed originale di cui sono figli se si verifica la rispettiva condizione d'uso. In questo caso il primo sotto-elemento *BedUseCase* è settato ad 1, secondo la specifica corrispondente ad un sistema 5.1, mentre il secondo sotto-elemento è settato a 2, corrispondente ad un sistema 7.1, come è possibile verificare dal rispettivo numero di canali in essi contenuti.

Qualora però i due casi d'utilizzo non si verificassero viene confermato l'utilizzo dell'elemento *BedDefinition* originale e anche qui dal confronto emerge che il numero massimo di canali che il renderer si potrebbe trovare a dover gestire contemporaneamente è pari a 10.

2.6 Gli altri formati

L'immersività delle esperienze sonore veicolata dai sistemi fin qui analizzati non nascono per essere mero appannaggio delle sale cinematografiche.

Diversamente da quello che si potrebbe pensare giunti fino a questo punto, le esperienze audio immersive non sono state pensate per essere fruibili esclusivamente in contesti in cui un equipaggiamento tecnico di alto livello è condizione indispensabile.

In un contesto socio-culturale come quello odierno in cui la fruizione indipendente di contenuti audiovisivi grazie ai servizi streaming (Netflix, Amazon Prime) è diventata pratica di ampia diffusione tra i consumatori, estendere la fruibilità dei prodotti anche al di fuori della sala cinematografica è di fondamentale importanza per la longevità di una tecnologia.

Qui di seguito verranno pertanto analizzati due contesti attraverso cui un'esperienza audio immersiva può essere sperimentata anche all'interno

delle mura di casa propria. Questi approcci come si vedrà non sono di recente invenzione, ma ancora una volta a fare la differenza in tutto questo processo è la struttura del nuovo formato digitale che rende la creazione di diverse versioni di distribuzione sempre più facile ed economica in termini di tempo per i fonici di mix.

Ecco perché, al di là di ovvie motivazioni commerciali, Stuart Bowling, *Dolby worldwide Technical Marketing Mgr.* affermò in merito al Dolby Atmos che fosse "(...) the next generation of surround format for the industry" [27].

Non è infatti solo questione di intensificazione dell'esperienza ma è anche la scalabilità con cui la creazione e distribuzione di questa esperienza è stata progettata.

2.6.1 Home Theatre

Le versioni Home Theatre dei sistemi sonori multicanale non sono una novità. Da sempre aziende come la Dolby accompagnano le loro tecnologie inserendo sul mercato anche le rispettive configurazioni per un ascolto ottimale a casa propria.

Anche dunque i sistemi immersivi non fanno eccezione e così ciascuna azienda ha inserito sul mercato le proprie configurazioni home theatre [34].



Figura 44 Configurazione home theatre Auro 3D. [34]

Auro 3D, una delle primissime aziende a immettere sul mercato configurazioni sonore tridimensionali nello spazio, ovvero aggiungendo la terza dimensione in altezza, propone configurazioni 11.1 e 13.1 che aggiungono alla base surround, oltre al diffusore da soffitto, quella che è la chiave di distinzione secondo Auro 3D rispetto ai competitors, ovvero il livello di altoparlanti all'altezza del cosiddetto *height layer*.

Secondo Auro 3D, in virtù della tecnologia da loro adottata, l'utilizzo di un sistema *object-based* all'interno di una sala di dimensioni ridotte sarebbe inutile, poiché solo con configurazioni di almeno 20 altoparlanti sarebbe possibile restituire gli effetti immersivi desiderati. Effetti che, date le dimensioni ridotte di una stanza di casa, stando a quanto da loro affermato, sarebbero già ampiamente forniti dallo strato sonoro creato dagli *height speakers*.



Figura 45 Configurazione Home Theatre Dolby Atmos 7.1.4. [35]

Per quanto riguarda invece il Dolby Atmos alle configurazioni 5.1 o 7.1 vengono aggiunti gli altoparlanti da soffitto o, poiché aggiungere questi ultimi può essere causa di installazioni poco gradite e non sempre possibili, in alternativa la famosa *Voice of God* viene ricreata attraverso *ceiling reflections* [35]. Dolby dunque abilita degli altoparlanti in grado di riprodurre

sia il suono sul piano orizzontale che, attraverso dei moduli aggiuntivi, il suono verso l'alto, il quale sfruttando le riflessioni del soffitto ricrea percettivamente una bolla sonora immersiva attorno all'ascoltatore. Questo tipo di combinazione può essere anche ricreata attraverso soundbar che in un'unica soluzione, stando a quanto dichiarato, dovrebbero consentire gli effetti sopra descritti.

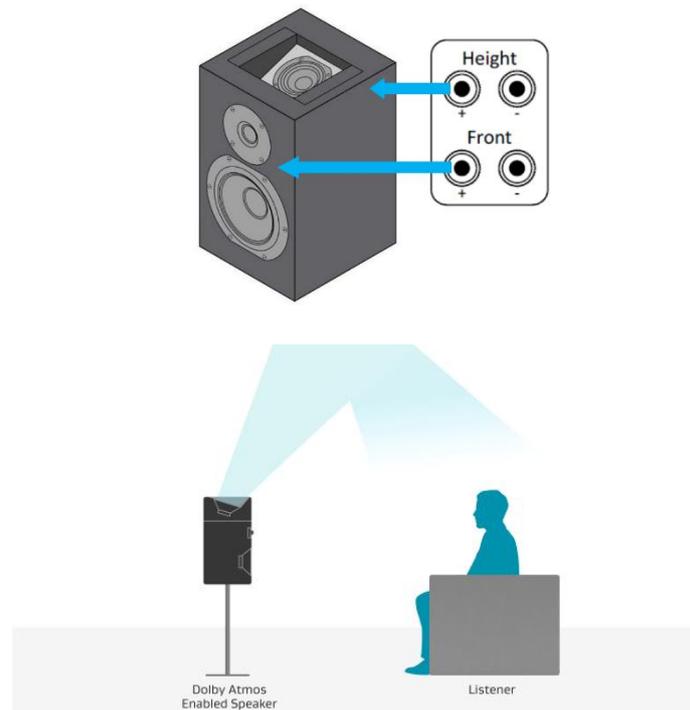


Figura 46 Altoparlanti di una configurazione Home Theatre Dolby Atmos che abilitano la verticalità del suono mediante *ceiling reflections*. [35]

In generale i restanti principali brand sul mercato seguono queste linee guida di massima.

Ciò che è interessante analizzare però è che grazie all'impressionante avanzamento tecnologico degli ultimi anni nella realizzazione di dispositivi mobili, i grandi brand dei sistemi sonori immersivi hanno puntato ad estendere la loro distribuzione attraverso i dispositivi mobili, validando

alcuni di questi come idonei a trasferire un'esperienza sonora immersiva. Certamente i risultati non possono essere paragonabili a quelli di una sala cinematografica o un salone di casa ben calibrato, ma è curioso analizzare che la rappresentazione spaziale adottata per consentire l'ascolto individuale attraverso questi nuovi dispositivi sia una tecnologia in realtà vecchia più di un secolo, stiamo parlando del binaurale.

Nel prossimo paragrafo e nel prossimo capitolo, verrà presentata la tecnica binaurale, le sue caratteristiche e come all'interno di un workflow di mix immersivo generare un master inteso per la riproduzione in binaurale risulti più semplice e personalizzato grazie alla traduzione che viene fatta dai metadati di tutte le sorgenti sonore introdotte.

2.6.2 Il Binaurale

I sistemi di riproduzione immersivi mirano a restituire un ambiente di ascolto *reale* in cui i suoni arrivano da tutte le direzioni possibili avvolgendoci in quello che è il contesto di realtà da noi percepito quotidianamente.

E' proprio in quest' ambito che si inserisce il binaurale, una tecnica, la cui prima invenzione risale già verso la fine del XIX secolo, che replica la facoltà umana dell'ascolto, nell'intento di restituire un ambiente sonoro tridimensionale.

Il suono nell'ambiente circostante infatti non arriva "puro" alla nostra percezione uditiva, esso a seconda della sua direzione d'incidenza viene alterato in fase e in frequenza dall'interazione con le spalle, la testa e la forma del padiglione auricolare (pinna). La somma di tali contributi (riverberazioni) e dei contributi diretti, una volta giunti all'orecchio, vengono inviati al cervello ed interpretati, consentendoci di comprendere la

localizzazione nello spazio di quella sorgente.

Questo processo avviene grazie alla differenza d'intensità (ILD) e alla differenza di tempo di arrivo – fase – (ITD) dei suoni che investono congiuntamente le due orecchie, le quali definiscono i meccanismi alla base della nostra percezione della localizzazione di suoni.

L'*Interaural Time Difference* (ITD), si verifica quando l'orecchio controlaterale (più lontano dalla sorgente sonora) riceve lo stimolo sonoro in ritardo rispetto all'altro e viene rilevata quando una sorgente non si trova nel piano mediano.

L'ILD, *Interaural Level Difference*, rappresenta la differenza spettrale dei due segnali che raggiungono le orecchie, tale differenza è causata dal fatto che la testa getta un'ombra acustica per l'orecchio controlaterale e comporta dunque una differenza interaurale nel livello del suono rilevato da ciascun orecchio.

La ITD e la ILD sono le fondamenta su cui si basa la *Duplex Theory* di Lord Rayleigh's (1907) teoria base della localizzazione dei suoni sul piano azimutale a cui successivamente sono stati aggiunti i contributi spettrali e le variazioni in funzione della posizione, necessari a ricostruire la spazializzazione percepita dall'uomo delle sorgenti sonore.

La combinazione di ITD e ILD insieme alle informazioni in merito alla colorazione spettrale sono catturate nella Head Related Transfer Function (HRTF).

Il suono binaurale dunque si riferisce al suono a due canali che entra direttamente nell'orecchio destro e sinistro dell'ascoltatore, il cui ascolto richiede necessariamente un paio di cuffie. Esso si distingue dal suono stereo, benché entrambi a due canali, per via del contenuto informativo in essi trasportato. Nel binaurale infatti il suono viene filtrato dai **segnali sopra descritti** intesi ad imitare la percezione spaziale umana.

Questi segnali possono essere ottenuti in modo naturale (ad esempio attraverso le registrazioni binaurali) o attraverso l'elaborazione successiva del segnale (sintesi binaurale).

2.6.2.1 Registrazioni Binaurali

La maniera più efficace sperimentata per registrare direttamente l'audio in formato binaurale è attraverso le *dummy heads* ovvero teste di manichini. Una *dummy head* [36] è infatti la rappresentazione fisica di una testa umana, con tutte le relative caratteristiche anatomiche, dimensioni, forma, altezza, localizzazione delle orecchie e forma del padiglione auricolare al cui interno vengono posizionati due microfoni stereofonici.

Alcuni manichini sono dotati anche di torso, in questo modo vengono registrati i suoni così come verrebbero recepiti da un vero orecchio umano, i cui segnali acustici arrivano come somma dei contributi diretti e riflessi dalle caratteristiche anatomiche di ciascun individuo (spalle, testa ed orecchie). Il risultato è dunque una registrazione a due canali il cui contenuto informativo trasporta tutti i segnali uditivi spaziali che sarebbero presenti se ad ascoltare ci fosse stato un vero essere umano.



Figura 47 A sinistra una dummy head, a destra un manichino di simile struttura ma che comprende l'ingombro spaziale anche delle spalle. [36]

REGISTRAZIONI BINAURALI INTERATTIVE

Per quanto accurata, la metodologia sopra proposta presenta un limite, ovvero che l'immagine sonora rilevata dai microfoni rappresenta una prospettiva fissa dell'ascoltatore non tenendo conto di eventuali movimenti del soggetto, il quale in fase di ascolto per mantenere una prospettiva ottimale non potrebbe ruotare la testa o interagire con l'ambiente circostante.

Una soluzione è stata ideata con il metodo Motion-Tracked Binaural (MTB) [33] un metodo di registrazione che permette di catturare e riprodurre suoni binaurali consentendo all'ascoltatore di muovere la testa al fine di ottenere un'immagine sonora la cui prospettiva è "listen-centric".

Basato sul metodo delle *dummy heads*, il metodo MTB usa microfoni distribuiti su una superficie sferica che approssima la testa di un ascoltatore. Un array di microfoni è distribuito sul piano orizzontale della superficie dove vengono catturati i segnali audio. Questo consente la registrazione dei suoni indipendentemente dall'orientamento della testa. In questo modo, durante la riproduzione, la prospettiva dell'ascoltatore all'interno della scena sonora può essere aggiustata attraverso l'utilizzo di informazioni di monitoraggio della testa (*head tracking information*). Una volta ottenuta l'informazione spaziale (*head tracker*), viene scelto il suono registrato sulla sfera che si trovava nella posizione più vicina a quella dell'ascoltatore in quel momento. Se il suono cade nell'intervallo tra due microfoni, il segnale risultante sarà frutto dell'interpolazione tra i due segnali.



Figura 48 Omini Binaural Microphone, [33], simile al metodo MTB, in cui l'array di microfoni anziché su una superficie sferica è posizionato all'interno di 4 coppie di "padiglioni auricolari" lungo le facce di un cubo.

2.6.2.2 Sintesi Binaurale

La sintesi binaurale può essere usata per simulare un segnale binaurale che rappresenta una sorgente sonora così come apparirebbe in un punto nello spazio ad un ascoltatore.

Per fare questo è necessaria la HRTF (Head Related Transfer Function), una funzione che riassume il filtraggio acustico operato dalla testa, dal busto e dal padiglione auricolare coerentemente ad una data posizione nello spazio della sorgente sonora.

Le HRTF sono infatti frutto di misurazioni dei segnali rilevati all'altezza del canale uditivo di un ascoltatore per posizioni diverse nello spazio di una data sorgente sonora.

Per ogni segnale in una data posizione esiste una funzione HRTF per l'orecchio destro ed una per l'orecchio sinistro, diverse l'una dall'altra, le quali contengono i segnali spaziali necessari per un ascoltatore al fine di percepire una sorgente sonora in quella posizione spaziale. Quando questi segnali contenuti nelle HRTF vengono sovrapposti sul segnale sonoro che si

intende sintetizzare, viene creata una rappresentazione dei segnali dell'orecchio destro e sinistro così come questi apparirebbero alle soglie del canale uditivo umano in un'esperienza d'ascolto reale, creando così l'illusione che il suono sembri provenire dalla posizione rappresentata nelle funzioni HRTF.

SORGENTI STATICHE

La sintesi binaurale è ottenuta partendo da un segnale monofonico di una sorgente sonora e facendone la convoluzione con i filtri contenenti la HRIR destra e sinistra che rappresentano la posizione in cui il suono binaurale deve essere percepito.

La HRIR è definita come la risposta impulsiva ad un suono da parte di un singolo orecchio (destro o sinistro), misurata da uno specifico punto nello spazio. Dunque, la convoluzione del segnale $x(t)$ emesso da una sorgente con la HRIR della specifica posizione nello spazio, converte il suono della sorgente in quello $y(t)$ percepito da un dato ascoltatore:

$$y(t) = x(t) * \text{HRIR}$$

La trasformata di Fourier del segnale risultante restituisce la HRTF, che contiene tutte le informazioni spaziali ed anatomiche del soggetto che comportano la trasformazione dell'onda sonora nel percorso dalla sorgente alle soglie del sistema uditivo.

Le HRIR usate per la sintesi binaurale possono essere generiche, misurate indipendentemente, selezionate da un database e/o personalizzate. Siccome vengono misurate in una camera anecoica, è possibile aggiungere artificialmente del riverbero alla sorgente sonora sintetizzata per ricreare gli effetti ambientali.

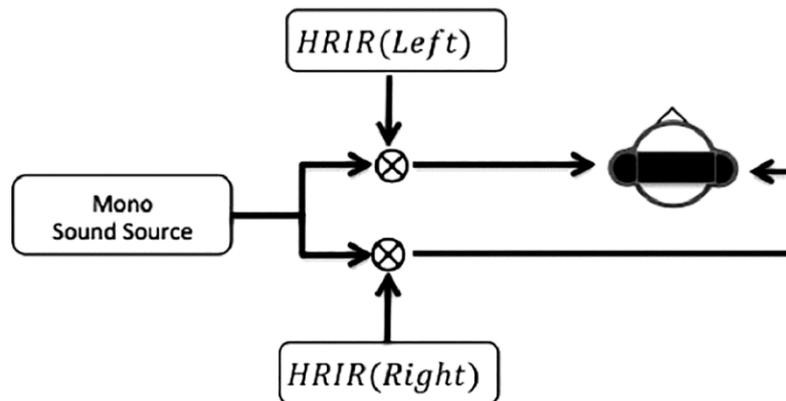


Figura 49 Diagramma riportato in Immersive Sound, the art and science of binaural and multi-channel audio che rappresenta il processo di convoluzione per la sintesi binaurale di sorgenti statiche. [33]

SORGENTI IN MOVIMENTO

Il processo di sintesi sopra descritto è ottimale quando sia la sorgente sonora che l'ascoltatore sono fissi nello spazio. Tuttavia, nell'ambiente sia le sorgenti sonore che i soggetti ascoltatori si muovono costantemente. Rappresentare dunque fedelmente la realtà implica essere in grado di replicarne il costante dinamismo sonoro in relazione ai movimenti di un ascoltatore, il quale spesso per poter meglio definire la localizzazione nello spazio di una sorgente muove la testa proprio al fine di acquisire maggiori informazioni su di essa.

Pertanto al fine di restituire una percezione sonora fluida e realistica, sia che la sorgente rimanga fissa e l'ascoltatore sia in movimento, sia viceversa che la sorgente sia in movimento e l'ascoltatore fermo, le HRTF devono essere costantemente adattate per restituire l'impressione di movimento mentre la posizione delle sorgenti cambia dinamicamente rispetto all'ascoltatore.

Per fare questo sono necessarie delle tecniche [33] che permettono il tracciamento (tracking) della posizione del soggetto ascoltatore rispetto ad

una data posizione. I principali strumenti utilizzati possono essere sensori inerziali, *tracker* magnetici, acustici od ottici.

Le informazioni di posizione del soggetto ottenute dai trackers sono utilizzate per adattare le HRTF da applicare per sintetizzare un ambiente binaurale in real-time. Per creare transizioni uniformi nello spazio le HRTF vengono interpolate continuamente tra due o più posizioni delle HRTF misurate nello spazio.

Rispetto ad un modello d'ascolto statico dunque, un sistema interattivo che utilizza il tracciamento della posizione della testa genera significativi benefici non solo nel creare un ambiente immersivo più realistico ma anche nel migliorare l'accuratezza spaziale, diminuendo le aree di confusione che un sistema statico potrebbe generare nella distinzione tra suoni derivanti da avanti/dietro o sopra/sotto.

3 L'Immagine sonora dalla sceneggiatura alla sua realizzazione missata in Atmos per un ascolto binaurale, analisi e sperimentazione del workflow e delle procedure applicate ad un vero progetto audiovisivo.

Nel terzo capitolo quale parte conclusiva della tesi viene proposto un approfondimento sulla dialettica audio-visiva del teaser di un documentario espressamente realizzato nel periodo tra Dicembre 2019 e Settembre 2020 come parte integrante del progetto di ricerca, sul quale è stato condotto un mix sonoro in Dolby Atmos finalizzato all'ascolto in binaurale.

Tale lavoro come verrà approfondito di seguito rappresenta l'espressione applicativa sintesi degli approfondimenti del primo e del secondo capitolo. Esso infatti è stato oggetto di riflessioni sia in merito a quello che è il contributo offerto alla narrazione da una dimensione sonora spazializzata, sia in merito alla possibilità di sperimentare la flessibilità messa a disposizione da questa tecnologia nella creazione di diversi formati di distribuzione.

Il capitolo sarà inoltre oggetto di approfondimenti pratici e riflessioni sul flusso di lavoro condotto in un sistema tridimensionale per il quale è stata scelta la tecnologia Dolby Atmos, in quanto maggiormente diffusa sul mercato cinematografico con più di 4000 sale nel mondo e circa 1000 film realizzati (dati del 2018).

Si specifica infine che tale workflow di sperimentazione è stato condotto mediante la supervisione e guida di Mirko Perri, noto sound designer del nostro Paese vincitore di tre David di Donatello per il suono nonché

professionista ad avere realizzato il primo film rilasciato in Italia in Dolby Atmos, *Youth* di Paolo Sorrentino (2016).

3.1 Il progetto

Il cinema è composto d'immagini e suoni in movimento che si intessono in trame registiche diverse a servizio del racconto di una storia, per lo più finalizzata a restituire la magia più grande operata dal medium cinematografico, un'emozione.

Motivo per il quale per sperimentare il workflow di un mix in Atmos è stato scelto di lavorare non su delle tracce campione, ma su una storia vera e propria che consentisse di far emergere non solo le peculiarità procedurali di un lavoro condotto con questa tecnologia, ma anche una riflessione di come essa si inserisca in un'architettura narrativa con la finalità di porsi a servizio di quest'ultima.

Nel corso dei mesi di realizzazione di questa tesi è stato dunque condotto da colei che scrive un progetto documentaristico sulla figura di Gianlorenzo Blengini, l'allenatore della Nazionale Maschile di Pallavolo Italiana, uscente argento olimpico a Rio 2016. Il progetto nasceva per raccontare il suo viaggio verso le Olimpiadi di Tokyo 2020, un viaggio nel frattempo interrotto e stravolto dagli accadimenti a livello mondiale della pandemia da Covid-19.

Pertanto con il materiale raccolto fino alla primavera-estate 2020 è stato strutturato un teaser promozionale del lavoro per il quale è stata scritta e pensata una "scena sonora" che raccontasse di quell'Olimpiade *che doveva essere e non è stata*, emblema dello stravolgimento mondiale di cui tutti noi siamo protagonisti.

La scelta della realizzazione di un teaser ha, da un lato, la valenza di essere un prodotto audiovisivo compiuto e di breve durata attraverso il quale poter sperimentare nell'interezza del formato la tecnologia in oggetto e, dall'altro lato, la valenza di richiedere una particolare attenzione per quelli che sono gli aspetti sonori e dunque emozionali del lavoro, per la sua intrinseca finalità promozionale.

Chiaramente ad oggi lavorare ad un film pensato per essere riprodotto in un sistema immersivo vuol dire lavorare progettando e sviluppando la massima qualità possibile. Un investimento questo di costi ed expertise non preventivabile per il progetto oggetto di questa tesi. Tuttavia l'obiettivo è stato quello di portare a confronto un workflow ideale con quello che poi è stato effettivamente possibile sviluppare con l'obiettivo e l'intento di affinare il mindset necessario allo sviluppo ed utilizzo di questa tecnologia.

Quali sono le scene più adatte per essere ascoltate in un sistema immersivo? Le dinamiche interne? Le necessità tecniche? I limiti?

Grazie alla possibilità di confrontarsi con un professionista di grande esperienza quale Mirko Perri è stato cercato di progettare una scena che potesse rispondere alle necessità d'ascolto di un sistema sonoro tridimensionale.

Il lavoro risultante e che si proseguirà a dettagliare non ha la pretesa di replicare le stesse prerogative estetico-narrative di un film da milioni di dollari come Gravity, non a caso firmato da un regista del calibro di Alfonso Cuarón. Tuttavia si cercherà illustrando *best and worst practices* di sviscerare lo sviluppo di un workflow al momento ancora così scarsamente adottato dall'Industria Cinematografica Italiana.

Nella convinzione infatti che, anche se il cinema italiano per via del nostro mercato e della tipologia di contenuti che produciamo non sia ancora incline alla conversione ai nuovi sistemi immerisivi, la tecnologia in un

ambito di studi come Ingegneria del Cinema vada esplorata per comprenderne peculiarità tecniche ed estetiche al fine di proporre spunti di ricerca, tesi a favorirne l'assimilazione nella grammatica dei processi creativi tipicamente europei.

3.1.1 Soggetto Della Narrazione

Il soggetto verte sugli eventi che hanno condotto l'attuale Commissario Tecnico della Nazionale di Pallavolo Maschile a raccogliere i maggiori successi della sua carriera in ambito olimpico, fino ad affrontare lo straniante attuale periodo storico in cui il mondo dello sport fronteggia il primo rinvio olimpico della storia, evento emblematico quale specchio dei tempi che stiamo vivendo. Per ragioni di privacy nel presente elaborato il soggetto della narrazione non è riportato nella sua interezza.

OMISSIS

OMISSIS

OMISSIS

OMISSIS

3.2 La realizzazione

Per la realizzazione del soggetto sopra descritto bisognava coordinare la creazione dei contributi visivi e sonori necessari al racconto ed alla sperimentazione Atmos. Ho scelto che la seconda parte del teaser, relativa agli effetti del Covid-19, fosse oggetto delle riflessioni di natura sonora di questo capitolo. Qui di seguito pertanto si proseguirà nel dettagliare il lavoro compiuto per la sua realizzazione, dalla sceneggiatura, alla progettazione, fino al set. Nel paragrafo successivo (3.3) invece si dettaglierà il lavoro fatto in post produzione.

3.2.1 La Scena Sonora

Nel soggetto sopra riportato il climax narrativo è rappresentato dall'annuncio del rinvio delle Olimpiadi, dopo il quale gli eventi si disgregano e allentano la tensione in linea con lo smarrimento da essi generato. Per arrivare a questo momento e raccontare i suoi effetti si è partiti da una sceneggiatura e da alcune riflessioni di regia in merito alla funzione sonora all'interno della scena di cui si riporta un estratto qui di seguito.

“Davanti a quell'enorme palazzetto, l'attesa di Gianlorenzo Blengini si è tramutata in nostalgia, forse, di qualcosa mai avvenuto. Di quell'Olimpiade nata all'insegna di quella bella data pseudo palindroma, 2020. Sembrava di buon auspicio e invece attorno a sé solo il rumore lontano della gomma delle suole a terra, dei colpi sordi sul pallone, dei fiati di fatica, dei fischi arbitrali, delle grida d'esultanza in un crescendo che diventa una partita sonora di un gioco mai giocato. Di un'Olimpiade mai avvenuta.

Si racconterà di quella partita di pallavolo mai giocata, di quell'Olimpiade attesa e rinviata per la prima volta nella storia, la si racconterà mediante una partita sonora di un gioco mai avvenuto in attesa che la realtà torni "reale". L'audio diventa la voce narrante degli effetti disorientanti di questo virus invisibile che avendo privato agli eventi la possibilità di accadere ha anche privato alle immagini la possibilità di raccontare. Si potrebbe parlare dunque non tanto di suoni che raccontano immagini, quanto di suoni che in assenza di immagini concorrono all'evocazione di un immaginario. Immagini Sonore appunto che evocano un sogno che sfuma e si dilata, ove la tecnologia diventa imprescindibile strumento ed opportunità per ampliare i confini della narrazione in un momento storico in cui i confini sono più che mai univocamente definiti."

Delineato pertanto il racconto dal quale si voleva partire l'obiettivo è stato in primo luogo quello di trasformare in immagini, coerentemente all'utilizzo del sonoro spazializzato che si voleva sperimentare, quanto sopra riportato attraverso i seguenti passaggi.

3.2.2 Il Suono

Attraverso il confronto con Mirko Perri è stato possibile condurre nell'ambito della progettazione e ricostruzione dello spazio sonoro tridimensionale una serie di scelte consapevoli.

Nel corso infatti di un'intervista relativa alle peculiarità dei sistemi sonori immersivi, ho potuto chiedergli coerentemente al progetto in oggetto, quale fosse la dinamica visiva, secondo la sua conoscenza nel merito dell'audio spazializzato, che meglio potesse sostenere un sistema sonoro

immersivo in Dolby Atmos. Dal confronto sono emerse le seguenti indicazioni:

- La grande potenzialità di un sistema tridimensionale è quella di dare allo spettatore la possibilità di ricoprire una posizione d'ascolto privilegiata, come se la soggettiva della scena avvenisse nella sua testa. Un tempo infatti se ad un campo lunghissimo con due persone molte piccole in lontananza veniva associato un audio della conversazione nitido e ad alto volume, quello che si voleva trasferire era la percezione di una conversazione che avveniva nella testa del personaggio, dunque un espediente soggettivo in un contesto oggettivo, ovvero in un contesto di riproduzione in cui le uniche casse utilizzate sono quelle dello schermo, senza poter trasferire un coinvolgimento totale dell'ascoltatore. Si necessitava dunque di una tecnica per trasferire un espediente narrativo. Se invece la tecnologia consente un totale coinvolgimento sonoro al punto di mettere l'ascoltatore nella condizione di recepire i suoni ambientali come se accadessero tutti intorno a lui con estrema precisione, allora le scene diventano impattanti se il punto di vista attraverso il quale vengono raccontate è quello di chi vive l'emozione più forte.
- Una delle maniere migliori per sfruttarlo pertanto è quella di non descrivere in maniera oggettiva la scena sonorizzando solo gli elementi presenti sullo schermo, bensì, come poc'anzi anticipato, assumere il punto di vista di colui che all'interno della narrazione in quel momento vive le sensazioni più rilevanti. Come se lo spettatore fosse un musicista all'interno dell'orchestra e non colui che ha pagato il biglietto per sedere solo di fronte allo spettacolo.

- La chiave inoltre per rendere la percezione di questi suoni che avvolgono e guidano la narrazione è quella di costruire la scena con un certo dinamismo. Secondo Mirko Perri infatti l'Atmos funziona per scene che prevedono spostamenti veloci e trasversali all'interno dello spazio e non scene lente in cui il rapporto tra i vari oggetti sonori rimane statico nello spazio e nel tempo.
- Di qui il suggerimento finale, al fine di massimizzare un'eventuale esperienza sonora immersiva, ricreare un ambiente che assumesse il punto di vista di chi su un campo da pallavolo vive l'emozione più forte, ovvero il giocatore. In questo contesto, restituendo i contributi sonori più significativi che circondano l'atleta, si avvolgerà lo spettatore in fase di ascolto delle medesime sonorità (ricezione, attacco, difesa, respiri, fiati, fischi di pallone che vorticano attraverso il campo..) proiettandolo all'interno dell'azione.

Partendo pertanto anche da questi suoi preziosi suggerimenti sono stati progettati e realizzati i passaggi che si prosegue a dettagliare.

3.2.3 La Location

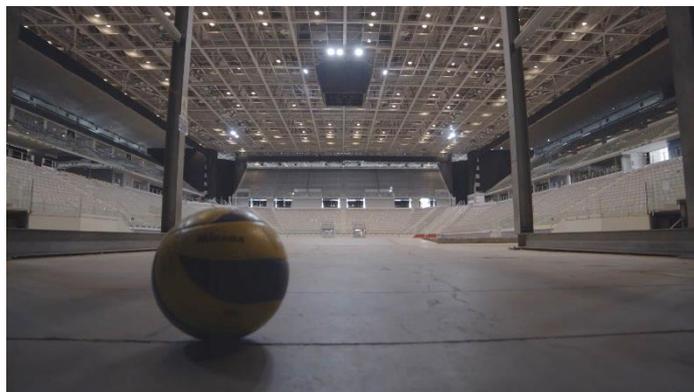


Figura 50 Immagine tratta dal teaser oggetto di tesi raffigurante una prospettiva del Pala Olimpico di Torino.

Al fine di poter dare un volto a questa narrazione è stata scelta quale location il Pala Olimpico di Torino, in connessione con il protagonista perché tra Giugno e Luglio 2020, proprio nei giorni in cui noi effettuavamo le riprese, doveva essere sede delle finali di una competizione pallavolistica internazionale da disputarsi alla vigilia delle Olimpiadi.

Era infatti necessario girare sia la prima intervista post quarantena a Chicco Blengini che le immagini per la ricostruzione del sopracitato racconto, in un luogo che rappresentasse l'emblema delle grandi sedi dello sport svuotate della loro vita che da templi del dinamismo si sono tramutati in arene silenziose.

Al di là dei vari aspetti logistici, il primo grande ostacolo che si è dovuto affrontare, una volta ottenuta la possibilità di girare presso il palazzetto, è stato nel breve tempo concesso durante il sopralluogo capire come “coprire” fotograficamente uno spazio di tali dimensioni.

Le nostre riprese si sarebbero divise in due mezze giornate, la prima dedicata all'intervista a Gianlorenzo Blengini più alcune sue immagini di contesto e la seconda dedicata a riprendere il palazzetto nella sua interezza.

L'obiettivo era quello che lo stadio fosse protagonista abitato solo più da suoni di eventi mai accaduti.

Tuttavia ben presto mi sono ritrovata ad affrontare un problema non trascurabile, ovvero che il Pala Alpitour essendo una struttura polifunzionale chiusa al pubblico a causa del Covid da mesi, non era allestito da arena pallavolistica. Al contrario esso, sia in termini di illuminazione che di “scenografia”, era completamente spoglio. Il che da un lato era un bene per la percezione della totale solitudine all'interno della quale avrei voluto inserire il protagonista, ma dall'altro lato non forniva allo spettatore nessun elemento che lo ricollegasse all'immaginario sportivo d'interesse.

L'evocatività dell'immagine sonora risultante infatti necessitava di elementi visivi che simbolicamente riconducessero ad un immaginario sportivo collettivo all'interno del quale i suoni avrebbero potuto popolare la vita che mancava.

Da queste riflessioni congiuntamente a quanto condiviso con Mirko Perri nell'intervista sopra riportata sono state prese le scelte che seguirò a dettagliare e che hanno pian piano plasmato l'idea originale, come spesso accade nel mondo dell'audiovisivo quando la realtà irrompe e bisogna fare i conti con la resistenza che la materia oppone alle idee.

3.2.4 I personaggi



Figura 51 Immagine tratta dalle riprese effettuate al Pala Olimpico di Torino.

L'emblema riconoscibile affinché l'ambiente scelto da “non luogo” potesse acquisire significato pertinente con la narrazione è stato dunque introdotto attraverso due giocatori di pallavolo. Essi sono stati vestiti con la stessa divisa recante i colori della Nazionale ed avrebbero dovuto rappresentare simbolicamente l'idea della pallavolo in generale.

E' stato detto infatti che l'esperienza in Atmos è massimizzata se il punto di vista è quello di una soggettiva che catapulti l'ascoltatore nello stesso mondo del protagonista della scena. Il protagonista del lavoro è Gianlorenzo Blengini e i suoni che dovevano abitare lo stadio dovevano

rappresentare una proiezione del suo stato emotivo frutto di una commistione tra i suoi ricordi e le sue aspirazioni.

Pertanto l'introduzione di un elemento visivo oltre che sonoro quale quello dei giocatori di pallavolo è stato pensato con la finalità di dare una forma ai pensieri dell'allenatore. Pensieri che fluiscono alla sua mente nel presente mentre si uniscono agli eventi del recente passato in una commistione che finirà per dissiparsi al suo ingresso in un stadio vuoto, eco del silenzio del presente.

In questa chiave dunque è stato cercato di pensare come la spazializzazione sonora potesse restituire una chiave più intimista del racconto in linea con le necessità narrative, volendo pensare ai sistemi immersivi come ad una formula tecnologicamente più avanzata per restituire un'espansione dei confini dello schermo.

3.2.5 La fotografia

Per gli stessi sopracitati motivi anche la scelta della fotografia è stata pensata con la finalità d'essere in grado di contestualizzare l'ingresso del nuovo elemento all'interno della narrazione, ovvero i giocatori di pallavolo, di favorire il dinamismo richiesto dall'utilizzo degli oggetti sonori e infine di armonizzare i problemi di location.

Se infatti in un primo momento protagonista del lavoro doveva essere un campo da pallavolo all'interno di un palazzetto da migliaia di persone vuoto, l'ingresso dei pallavolisti quale elemento narrativo ha richiamato all'attenzione due nuove tematiche con le quali fare i conti. In primo luogo i giocatori in fase di ripresa non potevano essere riconoscibili, perché mentre

un campo da pallavolo può significare tutti i campi da pallavolo del mondo, un volto richiamerà sempre ad un'identità riconoscibile, che in questo caso non aveva nulla a che fare con la narrazione del mondo interiore del personaggio.



Figura 52 Immagine tratta dal set presso il Pala Olimpico di Torino durante la realizzazione di parte del teaser e della scena sonora oggetto di tesi. In foreground il direttore della fotografia, in background la sottoscritta mentre lavoravamo congiuntamente alla composizione della scena.

Pertanto è stato scelto di girare immagini poco nitide, come fossero proiezioni oniriche che solamente dalla dinamica dei movimenti e delle silhouette restituissero l'identità della rappresentazione, alternando inquadrature descrittive con immagini dai movimenti rapidi e veloci il cui punto di vista si potrebbe definire solidale con il pallone da pallavolo, unico elemento mantenuto quando possibile a fuoco nella scena.

In secondo luogo sempre per via della mancanza di un vero e proprio contesto sportivo all'interno del quale inserirli è stato scelto che il pattern visivo ricorrente rimanesse quello delle luci dello stadio costituendo una vera e propria matrice visiva all'interno della quale le sagome dei giocatori si muovono.

Per quanto riguarda la dotazione tecnica è stato scelto di garantire retrocompatibilità visiva con le immagini del documentario già girate nei mesi precedenti.

E' stata quindi usata una Blackmagic URSA Mini Pro 4K della stessa linea della Black Magic Cinema Pocket, con cui appunto sono state girate le prime scene, ed un kit ottiche di diverse lunghezze fisse al fine di poter coprire secondo le necessità le varie angolazioni di ripresa.

3.2.6 L'immagine finale

Dal confronto tecnico a quello estetico, fronteggiando anche alcuni limiti comportati dalla messa in scena, la narrazione ha iniziato a prendere piede ed è stata realizzata come segue.

Il protagonista Gianlorenzo Blengini, in un certo senso emblema di tutti noi, è protagonista di un succedersi di eventi a cascata che lo condurranno a veder dissolvere una certezza monolitica quale quella rappresentata dal calendario di un'Olimpiade. Ripercorriamo insieme a lui in una giostra senza tregua gli eventi che lo hanno visto raggiungere il successo e quelli che lo avrebbero condotto forse a godere di un altro momento luminoso della carriera, ma il mondo inverte il suo asse e tutto inizia a precipitare velocemente.

Gianlorenzo Blengini cammina verso una meta, non sappiamo ancora quale, ma vediamo e sentiamo che cosa accompagna questo percorso.



Figura 53 Immagine tratta dal teaser del documentario oggetto di tesi.

Come fugaci interferenze si intromettono immagini sfocate, veloci e dinamiche che tra loro sembrano ricostruire un gioco pallavolistico scomposto. Sono richiami evocativi di giocatori di pallavo, ma non identificano nessun giocatore preciso, sono scie blu, come gli azzurri della nazionale, che si muovono nello spazio mantenendo il ritmo degli eventi.

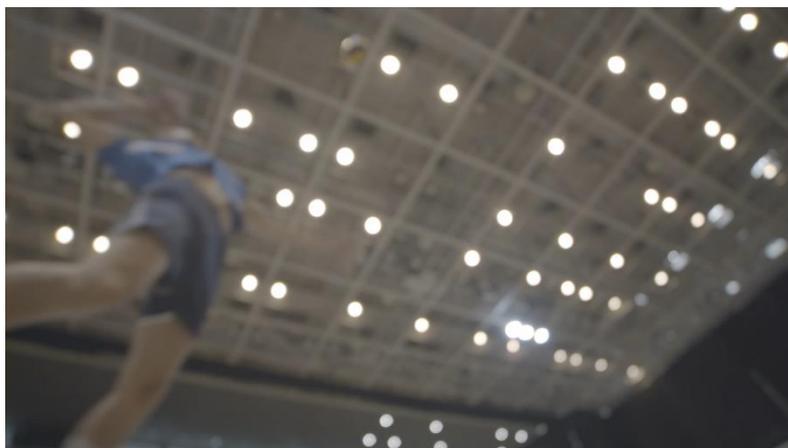


Figura 54 Immagine tratta dal teaser del documentario oggetto di tesi.

Sono l'evocazione di qualcosa che non c'è e doveva essere? Sono l'emblema dello sport? Sono i pensieri convulsi di un uomo abituato a vivere nel gioco ed ora come molti costretto all'immobilità? Fino a che tutto questo in un crescendo sonoro e di velocità prorompe in un caos visivo in cui la sagoma indistinta di un giocatore si allontana per ricongiungersi con quella

di Blengini che è arrivato alla sua meta, apre una porta. Silenzio. Chicco Blengini al centro di uno stadio vuoto.

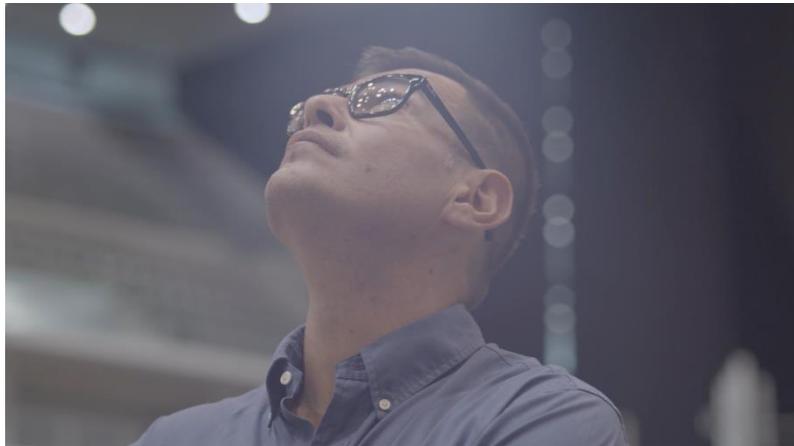


Figura 55 Immagine tratta dal teaser del documentario oggetto di tesi.

Le sagome visive che abitavano quello spazio non ci sono più, emblemi, ricordi, pensieri, sogni o fantasmi...

ora solo più il silenzio ed un auspicio. Tokyo 2021.

3.3 Workflow in Atmos

La lavorazione del suono in un film parte dalla lettura della sceneggiatura, nel difficile esercizio di immaginare un suono già nella fase di scrittura all'interno della quale il sound designer³⁰, così come i fonici e gli altri professionisti coinvolti nel processo, estrapola le informazioni necessarie al proprio lavoro. Essa prosegue con la presa diretta di dialoghi e ambienti sul set per poi giungere alle stratificate e complesse fasi di montaggio, composizione delle musiche e mix finale.

³⁰ Per sound designer si intendo “colui che si occupa della supervisione del disegno sonoro del film”, il quale “riferendo e confrontandosi con le altre figura chiave della filiera, presiede alla struttura complessiva del concetto sonoro del film” [10]

Nel corso del montaggio, momento aureo della realizzazione del sonoro filmico, vengono “puliti”³¹ e sistemati i dialoghi, se necessario doppiati, collezionati tutti i suoni che occorrono alla ricostruzione della dimensione del film attraverso registrazioni ambientali particolari, ricreazione di effetti sonori ad hoc (si pensi ai foley artists³²) e infine realizzato il montaggio scena vero e proprio di tutti questi elementi.

Musiche, dialoghi, ambienti ed effetti vengono spesso considerati secondo una comune metafora come gli elementi musicali di un’orchestra attraverso cui viene creata la sinfonia della colonna sonora e, come in ogni concerto che si rispetti, alla fine di questa lavorazione si passa al tassello conclusivo di tutto il film, ovvero il momento in cui il direttore d’orchestra dosa il contributo di ciascun musicista per armonizzare lo spettacolo finale, il “mix”.

Durante il missaggio si stabilisce infatti come lo spettatore sentirà quel prodotto cinematografico, tutti i “materiali sonori” [10] che lo compongono verranno dosati o modificati mediante effetti, quali rallenti o riverberazione, al fine di definire l’ascolto finale del film quale passaggio conclusivo per interpretare le volontà registiche e le necessità narrative della storia.

³¹ Verifica e controllo della qualità di registrazione. Nell’ambito dei dialoghi tra le operazioni di “pulizia del suono” vi è quella di mettere un fondo sonoro di base, spesso registrato direttamente sul set, in modo da non far percepire il vuoto che c’è tra gli stacchi di registrazioni.

³² Figura professionale già approfondita nel primo capitolo, in italiano rumorista, definito formalmente come “(..) un esperto in grado di ricreare sonorità realistiche – nel caso più semplice soprattutto passi e rumori degli attori – permettendo la registrazione insieme allo scorrere delle immagini e, tra le altre cose, garantendo un enorme risparmio dei tempi di sincronizzazione.” [10]

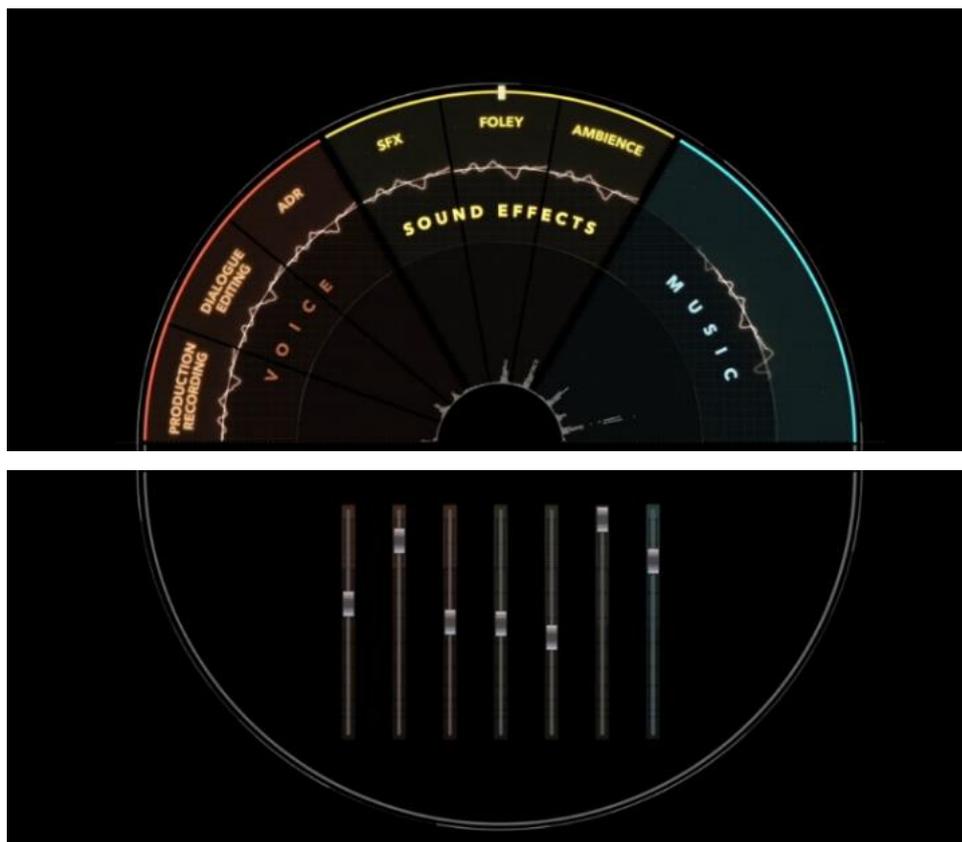


Figura 56 Immagine tratta dal documentario “Making Waves: The Art of Cinematic Sound” [4], film documentario sulla storia del sound design nel cinema degli USA, che illustra nella semi-circonferenza superiore il “circolo dei talenti” che compongono il workflow di composizione di una colonna sonora cinematografica e nella semicirconferenza inferiore il fatto che tutti questi materiali verranno dosati in sala mix al fine di restituire il risultato finale, il film.

All’interno di questo stratificato e affascinante mondo a metà tra l’artistico, il tecnico e l’artigianale l’introduzione dell’Atmos va da un punto di vista operativo ad incidere essenzialmente sull’ultimo tassello di tutta questa catena, ovvero il messaggio sonoro. Tale tecnologia dunque non porta particolari stravolgimenti rispetto al flusso di lavoro fino ad oggi adottato, tuttavia l’introduzione dei canali in altezza e del sistema sonoro object-based producono comunque delle riverberazioni anche su altri fronti della catena sonora.

Una di queste, forse la principale, è il modo in cui un autore che vuole utilizzare l'Atmos nel proprio film deve pensare alla realizzazione di una scena al fine di massimizzare l'esperienza sonora immerisva coerentemente alle proprie esigenze narrative. Questo infatti è proprio quanto si è cercato di dettagliare nel paragrafo 3.2 precedente provando ad ipotizzare un mindset adatto rispetto alla realizzazione del progetto in oggetto.

Qui di seguito, sempre grazie alle conversazioni condotte con l'esperto nell'ambito del sonoro cinematografico Mirko Perri, si proseguirà con l'affrontare quelli che sono i cambiamenti introdotti anche nelle fasi successive di lavorazione, dal set, al montaggio scena fino alla fase di mix, approfondendo come e dove è stato possibile rispettare o meno tale workflow nel corso della realizzazione del progetto di teaser oggetto di tesi.

3.3.1 Registrazione sonora e strumentazione

I “materiali sonori” che compongono la colonna audio di un film, come brevemente anticipato in introduzione, sono i dialoghi, i rumori (di cui fanno parte gli effetti, gli ambienti, il silenzio) e le musiche.

I formati di registrazione per i sopracitati elementi, nel passaggio da multicanale ad Atmos, rimangono pressoché invariati. Qui di seguito se ne descriveranno brevemente i fattori d'interesse per la realizzazione di una colonna spazializzata, fatta eccezione per le musiche la cui composizione ovviamente segue processi totalmente diversi che non sono oggetto di studio di questa tesi.

Partendo dai dialoghi dunque, il formato di registrazione continua a rimanere generalmente monofonico con la possibilità in sala mix di essere

riverberati su altri canali, disponendo nel caso di un sistema immersivo anche dei due canali in altezza.

Gli effetti e gli *hard effects*³³, continuano a rimanere anch'essi in formato monofonico oppure, a seconda delle esigenze artistiche, a poter essere ripresi con configurazioni microfoniche che replichino la disposizione del sistema di riproduzione. Ad esempio nel caso di un'esplosione può essere utile utilizzare un contributo già registrato in 5.1 che non dover ricostruire in un secondo momento la diffusione attraverso i vari canali. Tuttavia in Atmos è generalmente preferibile lavorare con il maggior numero possibile di effetti monofonici da poter localizzare indipendentemente nello spazio e ricostruire così un'intera ambientazione sonora spazializzata e dinamica. La categoria degli effetti infatti appartiene proprio agli oggetti sonori ampiamente descritti in precedenza.

Gli ambienti, i cosiddetti beds, sono gli elementi della colonna sonora tipicamente riprodotti per canali poiché intesi per ricostruire le connotazioni spaziali di un ambiente, ecco perché in fase di registrazione andrebbero ripresi sul più alto numero possibile di canali, coerentemente a quella che dovrebbe essere la destinazione di riproduzione finale. Dunque per un sistema sonoro immersivo sarebbe preferibile registrarli direttamente con configurazioni microfoniche che replichino in termini direzionali la disposizione delle sorgenti del sistema di riproduzione di riferimento. Nel caso di un sistema Atmos devono essere in grado di restituire una spazialità su 9 canali (3 centrali, 4 surround e 2 in altezza).

³³ Gli *hard effects* sono suoni visibilmente sincronizzati con l'immagine. Gli effetti in generale invece sono suoni chiaramente distinti dal fondo ma potrebbero anche non essere riconducibili ad una causa visibile su schermo

Siccome non è pratica comune, oltre che essere poco maneggevole, riprendere i suoni con 9 microfoni in contemporanea quello che viene fatto è utilizzare delle tecniche che con un numero inferiore di microfoni e una decodifica successiva danno la possibilità di restituire lo spazio sonoro desiderato.

Ad esempio, come spiega Mirko Perri nel suo articolo in merito alla lavorazione in Atmos del film *Youth* di Sorrentino [37], al fine di poter realizzare delle ambientazioni naturali e di mantenere gli spettatori costantemente avvolti nello spazio sonoro anche attraverso gli altoparlanti in altezza ed “evitare di accendere e spegnere all’occorrenze l’audio del soffitto”, ha adottato un Double MS³⁴ per le registrazioni sul piano orizzontale, più due microfoni omnidirezionali per gli altoparlanti *overhead*.

Successivamente le registrazioni in Double MS sono state convertite in B-Format, una volta ottenuto il segnale B-Format esso è stato convertito in 7.0 e successivamente aggiunto alle registrazioni in altezza per comporre un ambiente immersivo [37].

Nel caso specifico del teaser realizzato non è stato purtroppo possibile lavorare con tale formato di registrazione, questo perché in fase di recording ho avuto l’opportunità di avvalermi dell’expertise di un valido professionista torinese abituato a lavorare in presa diretta con un massimo di due canali, dunque in stereo. Questo a riprova del fatto che registrare con le tecniche

³⁴ E’ una tecnica microfonica utilizzata per registrare i suoni ambientali surround usando tre microfoni in un’unica soluzione compatta. E’ composto da un cardioide frontale, un microfono direzionale rivolto all’indietro e un microfono laterale a figura di 8, da cui poi è possibile estrarre cinque canali (centrale, destro, sinistro, sinistro surround e destro surround). E’ interessante notare che le codifiche di uscita saranno proprio “fisiche”, come se fosse stata impostata una simile microfonaione e non si tratterà di una elaborazione virtuale (che è ciò che fanno gli algoritmi di conversione ad esempio da stereo a 5.1).

microfoniche sopra descritte non è raro ma al contempo non rappresenta lo standard di produzione sonora in Italia, questo perché solitamente per la tipologia di film realizzati è sufficiente registrare gli ambienti in stereo e poi in postproduzione ampliarli per la riproduzione sui sistemi multicanale. Pertanto i background sono stati registrati in stereo. Tuttavia essendo che gran parte del lavoro finale del teaser sarebbe stato composto di contributi di interviste alternati su una base musicale ed essendo che l'obiettivo sarebbe stato quello di consentire un ascolto binaurale del prodotto, è stato infine valutato che questo non avrebbe rappresentato un vero e proprio ostacolo alla sperimentazione progettuale



Figura 57 Immagini tratte dal set presso il Pala Olimpico di Torino durante la realizzazione di parte del teaser e della scena sonora oggetto di tesi. Sulla sinistra il fonico, il direttore della fotografia e la sottoscritta in quanto regista, sulla destra i giocatori di pallavolo intenti a seguire le indicazioni al fine di poter riprendere i dettagli visivi e sonori del gioco pallavolistico necessario alla scena.

Ai fini invece del sistema object-based che si sarebbe dovuto andare a sfruttare in post-produzione è stato cercato di registrare il maggior numero possibile di contributi sonori monofonici relativi alle ricostruzioni del gioco sportivo con i pallavolisti (fiati, palleggi, bagher, cigolii delle scarpe di gomma, sospiri) con la finalità di avere tutti i contributi sonori necessari ad operare una buona spazializzazione della scena, evitando il più possibile l'adozione di librerie.

3.3.2 Strumenti di postproduzione

Il fulcro distintivo di un workflow sonoro Atmos rispetto alle procedure di missaggio per canali, come già ampiamente spiegato, è rappresentato dalla possibilità di disegnare con precisione nello spazio la posizione degli oggetti, avendo così un'ambientazione sonora praticamente illimitata nella quale inserire lo spettatore.

Per poter generare i metadati posizionali di un oggetto, Dolby Atmos mette a disposizione diversi strumenti [38]:

- *Dolby Atmos Panel Plugin*, un plugin di pan che introduce l'asse z che comunica con l'unità di rendering attraverso l'applicazione software *Dolby Monitor*.
- Le workstation audio digitali abilitate, quali Avid Pro Tools Ultimate e Steinberg Nuendo, in grado di generare nativamente oggetti audio e metadati e comunicare così direttamente con l'unità di rendering.

- Le *Mixing Consols* di grande formato, quali Avid System 5, AMS Neve DFC Gemini, Harrison MPC, in grado anch'esse di generare metadati comunicando direttamente con il renderer Dolby Atmos.

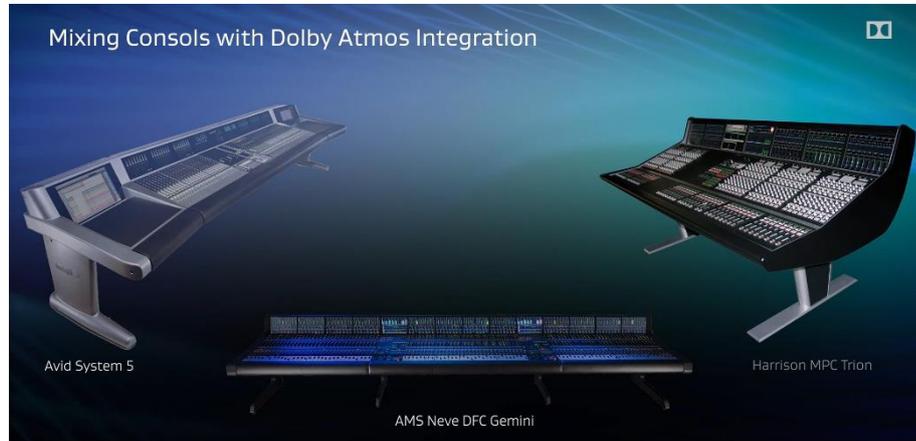


Figura 58 Consols di missaggio fisiche in grado di generare nativamente metadati in fase di missaggio ad oggetti [38].

I metadati generati dal processo di panning e manipolazione degli oggetti attraverso questi strumenti vengono direttamente inviati ad un sistema di rendering. La strumentazione e le modalità con cui essi comunicano con il renderer variano a seconda della distribuzione cui è destinato il prodotto per il quale si lavora. Si possono distinguere pertanto tre diversi workflow [39]:

- Missaggio per il Cinema.
Se il contesto di lavorazione sono ampie sale mix in cui vengono missati film per la riproduzione cinematografica, allora la dotazione necessaria è quella che viene chiamata **Dolby Atmos for Cinema**. Quest'ultima è composta da un'unità di rendering RMU, una piattaforma hardware preconfigurata che contiene il Renderer Dolby Atmos, nonché gli strumenti di progettazione Atmos per la calibrazione della stanza e l'applicazione di Monitoring Dolby Atmos.

La RMU contiene lo stesso identico renderer che si trova nei processori Dolby Cinema, ciò significa che quando si lavora in una sala mix approvata, si ha la certezza, stando a quanto affermato dalla Dolby, che quello che si sente in sala mix si tradurrà in un'identica esperienza d'ascolto al cinema.



Figura 59 Dolby Atmos Theatrical RMU (Rendering and Mastering Unit) server, unità fisica di rendering.

- Lavorazione per prodotti audiovisivi di alto profilo.

Per questa tipologia di workflow la Dolby propone **Dolby Atmos Mastering Suite**, una dotazione hardware e software, composta da un'unità di rendering fisica che comunica attraverso un'applicazione software con una workstation digitale (una tra le due viste in precedenza). Attraverso l'interfaccia di rendering si possono monitorare gli oggetti sonori la cui creazione nonché manipolazione mediante programmi quali ProTools genera delle coordinate 3D inviate automaticamente al renderer sotto forma di metadati sincronizzati. L'unità di rendering fisica e la workstation digitale sono connesse via MADI (o via Dante a seconda della tipologia di unità fisica adottata).

- Lavori editoriali e di premixing per alimentare un flusso di lavoro più ampio.

Per questo tipo di lavorazioni viene proposta invece la soluzione **Dolby Atmos Production Suite**, è un pacchetto software solo per sistemi Mac che lavora sullo stesso computer dove si trova la workstation digitale, per far comunicare workstation e unità di rendering vengono utilizzati dei plugin a seconda della tipologia di workstation adottata. Questo comporta che alcune funzionalità non siano disponibili ecco perché per prodotti audiovisivi complessi e di alta qualità è necessario usare le prime due soluzioni.

E' stato detto dunque che per comunicare con l'unità de rendering Dolby Atmos si avvale di un'interfaccia le cui caratteristiche verranno dettagliate brevemente qui di seguito [40].

Quella rappresentata in figura rappresenta la terza ed ultima versione al momento rilasciata dell'interfaccia di rendering Atmos, la quale fornisce tutte le informazioni necessarie per monitorare che cosa accade nel mix minuto per minuto. Attraverso di essa inoltre è possibile impostare tutte le configurazioni necessarie, oltre che realizzare il master finale.

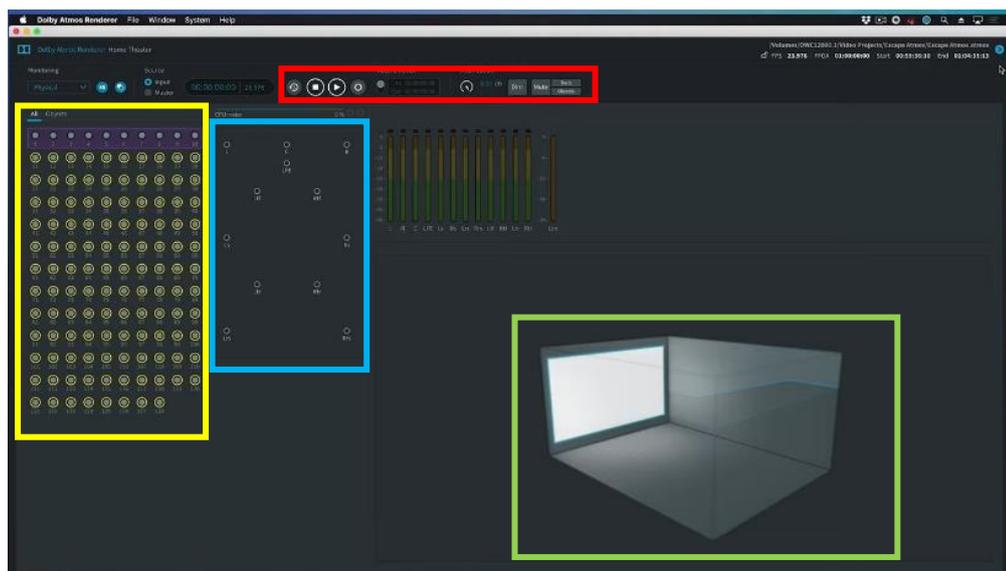


Figura 60 Dolby Atmos Rendering Application Version 3. [40]

In giallo è segnalata la griglia di input in cui vengono indicati in ogni momento quali sono gli input del mix attivi contemporaneamente (i primi dieci sono canali, i restanti 118 elementi possono essere trattati come canali o oggetti monofonici), in azzurro è segnalata la configurazione di sala, mentre nel riquadro verde è segnalato il modello tridimensionale virtuale della sala di riproduzione. Al suo interno man mano che vengono convertiti canali in oggetti questi appariranno in verde, le cui dimensioni indicano il volume, l'alone indica la diffusione se puntiforme o spazialmente più diffusa, mentre i colori diversi segnalano il livello relativo tra oggetti.

In rosso invece sono segnalati i comandi di mastering attraverso i quali, prendendo come sorgente il mix realizzato attraverso workstation, viene registrato quanto finalizzato in sala mix.

Si segnala infine che attraverso il menù a tendina "window" è possibile accedere alla finestra di rendering binaurale per settare le impostazioni di un'eventuale codifica binaurale. Quale nota a margine si fa presente che una delle mancanze del sistema Atmos segnalate sui vari forum di utilizzatori pare che sia il fatto che esso non consenta di importare le librerie di binaurale preferite ma che imposti automaticamente le proprie caratteristiche di filtraggio, senza dare dunque margini di personalizzazione.

3.3.3 Confronto tra workflow multicanale e Atmos

Montaggio e missaggio sonoro di un film sono complesse e stratificate procedure sulle quali esiste ampia letteratura tecnica ed estetica. La finalità del presente paragrafo non è quella di analizzarne le peculiarità artistico-strutturali, quanto quella di illustrare attraverso un esempio le principali differenze che intercorrono in postproduzione tra un workflow multicanale

classico ed uno in Atmos, secondo quanto analizzato insieme a Mirko Perri nel corso di un approfondimento pratico attraverso il suo studio attrezzato e casa di post-produzione a Roma, svolto in streaming purtroppo per via del Covid.

Tale panoramica sarà anche propedeutica all'approfondimento operativo fatto sulla struttura sonora del teaser analizzato nel paragrafo successivo (3.3.4).

3.3.3.1 WORKFLOW MULTICANALE

Per poter affrontare un lavoro di editing e mix sonoro multicanale classico sono necessarie una scheda audio, un software di editing, una workstation digitale (integrata nello stesso software di editing) o fisica e una configurazione di casse per la riproduzione, adeguata alla tipologia di mix che si desidera realizzare.

Una sessione di lavoro multicanale rappresenta un “sistema chiuso” poiché ogni traccia audio inserita corrisponde ad una precisa destinazione nel sistema di riproduzione. Questo favorisce anche il fatto che spesso in fase di lavorazione i confini tra editing e mix siano più sfumati, ovvero è pratica comune già in fase di montaggio audio lavorare ad un pre-mix della scena. Un confine questo che l'Atmos tornerà a rendere un po' più rigido perché se in montaggio le operazioni sono assimilabili al flusso di lavoro precedente è in fase di mix che viene sperimentata la vera e propria scalabilità di questa tecnologia.

Nell'immagine che si riporta qui di seguito è possibile osservare un esempio di semplice stratificazione sonora di una scena pensata per un sistema 5.1. In alto si trova il suono monofonico di una roccia, in centro il

verso di un'aquila in stereo, in basso la traccia ambientale di un contesto naturale registrata in 5.1. Ogni traccia che si vede sarà automaticamente indirizzata ad un preciso set di altoparlanti, la prima verso la cassa centrale, la seconda alle coppie di casse stereo dietro lo schermo e la terza spalmata all'interno del sistema surround 5.0.

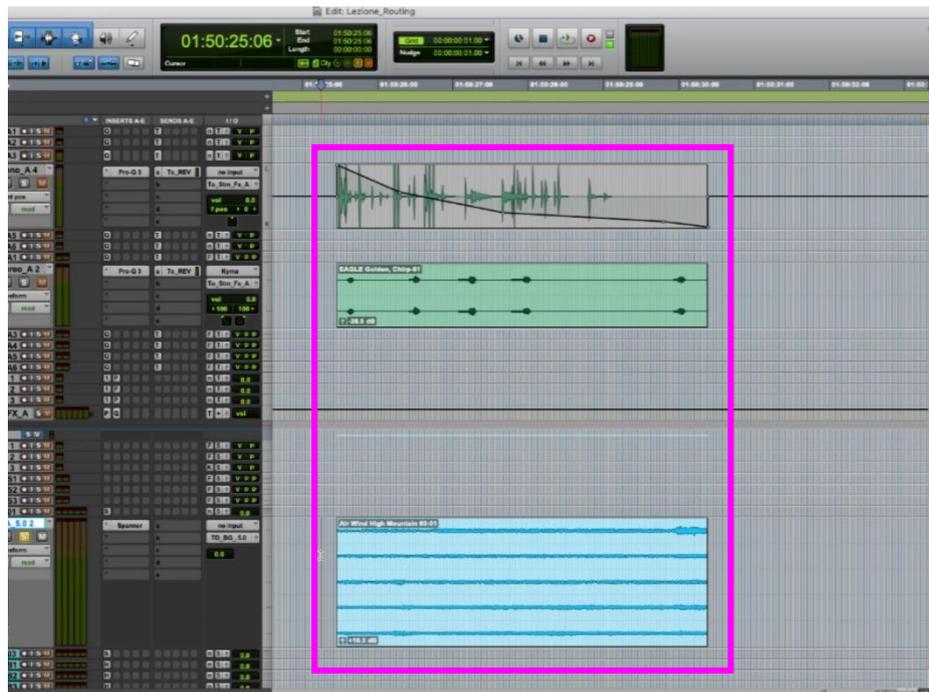


Figura 61 Sessione ProTools in 5.1 di una semplice scena sonora (in fucsia) composta da tre materiali sonori di diverso formato.

Una volta definito con esattezza il montaggio scena si passa all'applicazione di effetti, quali riverbero o equalizzazione, dei contributi sonori che lo richiedono. Contestualmente si procede con la calibrazione di questi suoni all'interno del sistema di riproduzione scelto, in questo caso 5.1. All'interno della composizione vengono fatte delle lavorazioni di panning, spostando i suoni solo sul piano orizzontale, per dosarli sulle varie casse della configurazione bidimensionale.

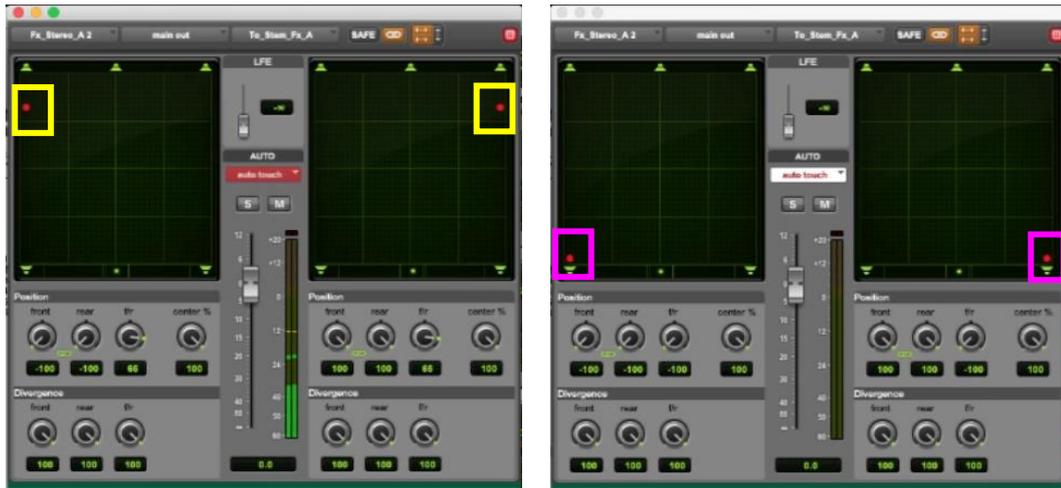


Figura 62 A sinistra il suono stereofonico dell'aquila in prossimità delle casse dx e sx dello schermo, a destra la risultante finale del movimento di panning del suono fatto dallo schermo fino agli altoparlanti disposti a fondo sala.

A lavorazione terminata, un processo lungo ed accurato in cui tra le varie operazioni vengono calibrati i volumi generali e scelti i dosaggi di riverbero anch'essi se necessario spazializzati, viene avviata la registrazione della composizione con i relativi panning sonori sulle 6 piste (se fossimo in un sistema 7.1, 8 piste), in questo modo si ottiene il master finale in cui vengono memorizzate le rispettive spazializzazioni sul piano orizzontale. Essendo un sistema chiuso per ascoltare questo master realizzato su 6 piste saranno necessarie altrettante piste (i 3 canali dello schermo, i 2 canali surround ed uno per le basse frequenze).

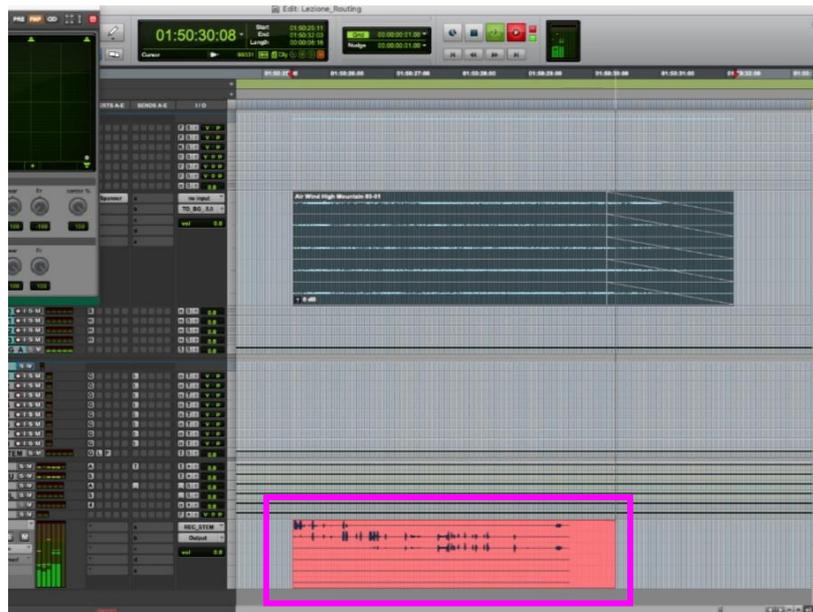


Figura 63 Registrazione del master della sessione di montaggio e mix della breve scena sonora sopra descritta in cui ogni pista viene registrata su un canale diverso.

Qualora si necessitasse di avere un master anche per un ascolto stereo, dunque finalizzato per un ascolto a due soli canali, per farlo bisognerebbe lavorare facendo alcune ottimizzazioni finalizzate a non sovraccaricare in termini di volumi i due soli canali, perché sei piste verrebbero concentrate in due. Operate tali ottimizzazioni è possibile realizzare il secondo master. In un sistema pre-mappato dunque, per ogni tipologia d'ascolto devo realizzare un master opportuno a sua volta frutto di aggiustamenti di montaggio e scelte di mix create ad hoc per ciascuna configurazione. Operazioni che, data la stratificazione sonora e la complessità su cui si articola un film, richiedono numerose lavorazioni tutt'altro che banali in termini temporali.

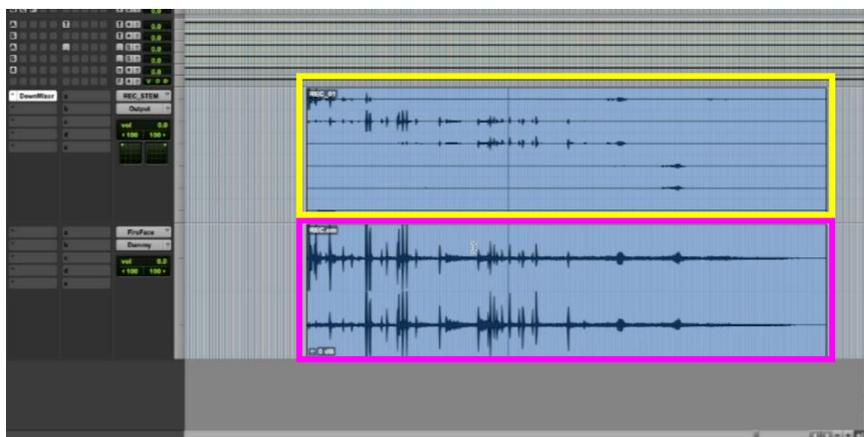


Figura 64 Nella figura vengono riportati in giallo il master 5.1 e in fucsia il master in stereo, risultanti dal semplice montaggio scena visto in figura 61. Questo evidenzia che per ogni tipologia d'ascolto un sound designer deve realizzare master differenti contrariamente a quanto accade per l'Atmos.

3.3.3.2 WORKFLOW ATMOS

All'interno di un workflow Atmos come si vedrà qui di seguito le variazioni di lavorazione per un sound designer sono minime ma sostanziali.

All'interno del software di montaggio sono stati adottati gli stessi “materiali sonori” negli stessi formati visti in precedenza e si può scegliere di lavorare all'editing similmente a quanto visto prima, ovvero avvalendosi solo dei bed in cui i materiali sonori vengono trattati come dei canali pre-mappati. In questo modo le operazioni di panning fatte su un bed continuano a venir codificate all'interno di un sistema pre-mappato, ove rispetto al workflow precedente non cambia nulla salvo il fatto che è possibile sfruttare anche i canali in altezza.

Le differenze subentrano qualora si decidesse di trattare i materiali sonori anziché come dei canali come degli oggetti, la cui operazione richiede semplicemente l'attivazione di un bottone (Figura 3.12). Questo farà sì che nell'interfaccia di rendering del sistema Atmos si materializzino dei puntini verdi all'interno dello spazio virtuale, raffigurante l'ambiente di riproduzione, attraverso il quale è possibile osservare dove e come le manipolazioni spaziali dell'oggetto vengono renderizzate e dunque ascoltate in sala.

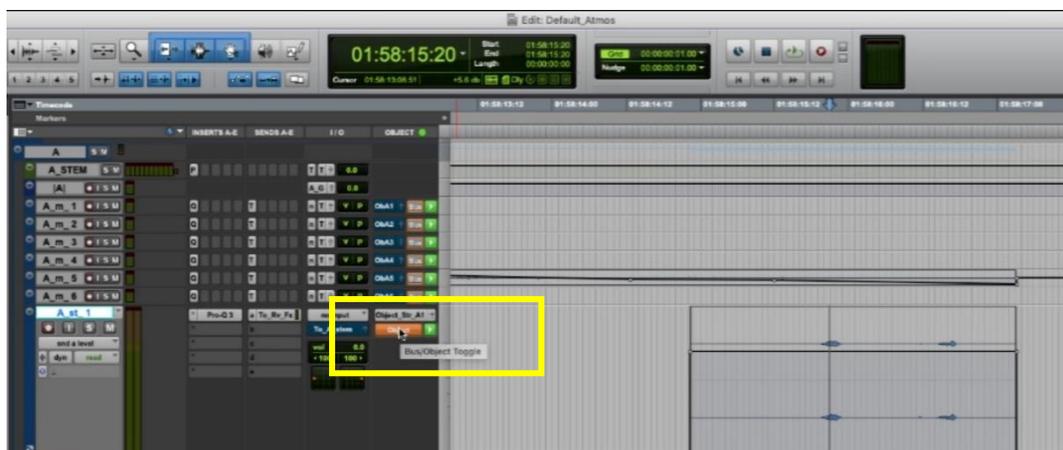


Figura 65 Sessione ProTools in Atmos della stessa scena sonora vista in precedenza in cui il canale del suono stereofonico associato al verso dell'aquila viene convertito in oggetto.

Qui di seguito si elencano sinteticamente le variazioni scaturite da tale operazione:

- Mediante l'operazione sopra descritta di conversione di un canale in oggetto tutte le lavorazioni di pan e relative manipolazioni fatte su quell'oggetto verranno scritte come metadati ed inviate al renderer, motivo per il quale adesso nell'interfaccia di monitoring è apparso il suono trasformato in oggetto all'interno della configurazione virtuale.

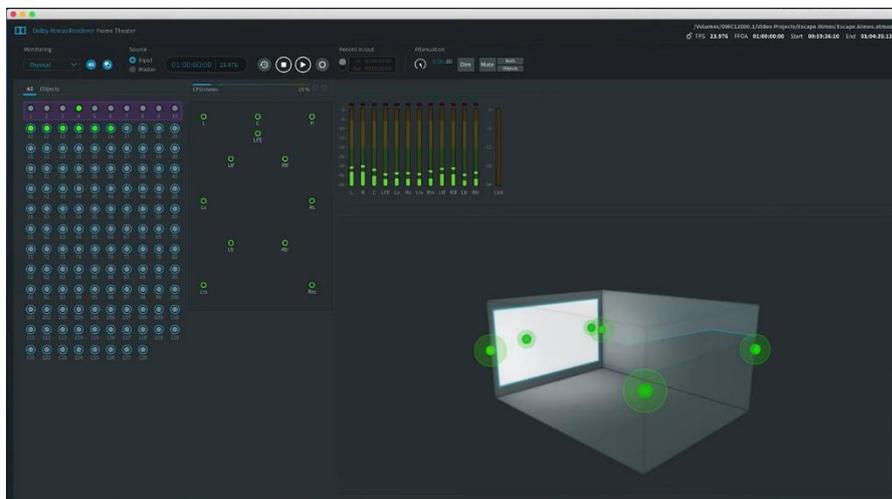


Figura 66 Interfaccia software dell'unità di rendering Dolby Atmos (in questo caso non raffigurante il mix sopra descritto, come si può notare dalla presenza di più oggetti sonori attivi in contemporanea) in cui i pallini verdi all'interno della configurazione di sala virtuale rappresentano le caratteristiche spaziali degli oggetti sonori istante per istante.

- All'interno dell'interfaccia di pan (Figura 67) inoltre non si avrà più a disposizione solo il piano orizzontale in cui far muovere i suoni bensì anche la dimensione in altezza (una funzione questa disponibile anche per i beds). Attraverso l'interfaccia di rendering è possibile osservare come i movimenti di panning vengono tradotti dal renderer all'interno della sala.

- Oltre alle proprietà posizionali il fonico di mix ha la possibilità di specificare vari parametri che vengono tradotti automaticamente come metadati man mano che lavora alla definizione delle caratteristiche spaziali di un oggetto sonoro, quali quelli osservati nel capitolo precedente (ObjectGain, ObjectPosition, ObjectSnap, ObjectDecor..). Uno fra tutti si ricorda quello relativo alla diffusione di un oggetto, ovvero se si vuole che esso abbia una precisione puntiforme o che diffonda il suono a più altoparlanti adiacenti (objectSpread).



Figura 67 A sinistra esempio all'interno del plugin di pan di uno spostamento verticale di un oggetto, a destra dello spread di un oggetto.

- Ogni oggetto spostato e manipolato all'interno di una configurazione sonora viene automaticamente estratto dal bed e scritto come un file a parte, questo consente la scalabilità del sistema. Se infatti un oggetto continuasse a rimanere all'interno del

canale non sarebbe possibile descrivere le sue proprietà spaziali in maniera indipendente, mentre il fatto che venga separato dal resto fa sì che quando il renderer legge il suono contenuto e i suoi metadati è in grado di trattarlo in maniera indipendente ed ottimizzarne la posizione nello spazio che non è fissa come quella già impostata per i canali.

Quelle sopra descritte sono in sintesi le caratteristiche di distinzione principali che un professionista della post-produzione audio si troverà ad affrontare all'interno di un flusso di lavoro Atmos, ovviamente man mano che la complessità della scena aumenta, aumentano anche le criticità e le peculiarità sonore da gestire all'interno dell'intera lavorazione. Qui di seguito si riporta un'immagine che sintetizza l'intero processo.

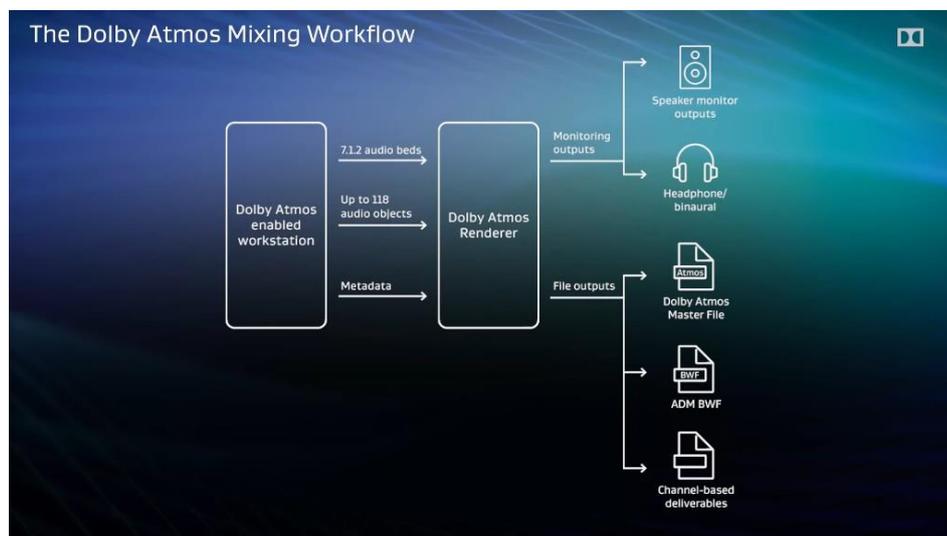


Figura 68 Workflow di mix in Dolby Atmos. [38]

In figura 3.15 si può infatti osservare come attraverso una workstation abilitata, sia essa fisica o digitale, è possibile lavorare sui vari materiali sonori trattabili sia come canali che come oggetti. Il sound

designer ha dunque a disposizione sino ad un massimo di 10 canali e 118 oggetti sonori che possono “suonare” in contemporanea. Man mano che viene manipolato ciascun oggetto vengono scritti set di metadati che insieme ad oggetti e canali vengono mandati al renderer il quale genererà file diversi che descrivono da una lato le caratteristiche dei canali e dall’altro le proprietà tempo varianti definite per gli oggetti, questo al fine di consentirne una riproduzione ottimizzata per le configurazioni di destinazione.

I files finali generati dunque sono i seguenti: il *Master File* (che contiene tutto l’audio del mix, gli oggetti sonori e in un file a parte anche i metadati per ciascun oggetto), un file contenente solo i canali ed un *ADM file* contenente canali, oggetti e metadati tutti insieme. All’interno di questi files sono contenute le informazioni necessarie per consentire un ascolto ottimizzato sia per i sistemi multicanale (9.1/7.1/5.1/2.0/1.0) che per l’ascolto binaurale in cuffia.

3.3.4 Editing e mix audio del progetto

Tornando dunque alla finalizzazione del teaser oggetto di tesi, qui di seguito verrà approfondito come attraverso l’intreccio dei tre seguenti “materiali sonori” è stata impostata la costruzione narrativa del montaggio sonoro di *Torino Tokyo 2020*:

1. La musica

Il teaser in oggetto nasce in una formula atipica, ovvero quale prequel di una storia ancora da raccontare, anziché quale montaggio ritmato di momenti significativi di un racconto già interamente avvenuto. Questo al fine di presentare in maniera accattivante il personaggio principale e

ingenerare curiosità nei confronti di quello che sarà la prosecuzione della storia, nonché oggetto di narrazione del documentario. In questo scenario dunque, come è tipico anche per i trailer promozionali, la musica assume un ruolo fondamentale nella gestione dei ritmi narrativi. Ecco perché il montaggio video è stato letteralmente cucito addosso alla musica. Come, a fronte dell'ascolto del teaser, si può rilevare nella scena iniziale in cui i tasti di pianoforte accompagnano esattamente l'accensione dei fari all'inizio e al contempo in chiusura circolare scandiscono l'apertura della porta dello stadio e il primo ingresso di Blengini al suo interno.

I temi musicali adottati sono due al fine di scandire proprio in due fasi la narrazione, senza soluzione di continuità, per cui dalla prima si viene vorticosamente proiettati nella seconda, adottando la leva dei Giochi Olimpici quale *trait d'union*.

La prima fase imprime un ritmo energetico, coinvolgente ed adrenalinico al fine di sottolineare i successi sportivi che lo hanno condotto alla fama, nonché alla conquista dell'argento alle Olimpiadi di Rio 2016. Passando così in meno di un anno da secondo allenatore a Commissario Tecnico della Nazionale che porta la squadra a vincere ciò che solo i più grandi prima di lui sono riusciti a conquistare.

La seconda fase invece è scandita dai ritmi di un pezzo molto diverso. Si subentra nel presente in cui i successi sportivi sono accompagnati dalla sensazione di un presagio di qualcosa che di lì a poco verrà svelato, un eco silenzioso apparentemente ancora innocuo poiché lontano, il Covid.

La musica di lì incalza aumentando la concitazione degli eventi accompagnati dalla frenesia delle immagini pallavolistiche che subentrano come incursioni fulminee. Tuttavia il climax della tensione non viene

affidato all'apice musicale e questo ci introduce al secondo elemento sonoro narrativo, gli effetti sonori.

2. Gli effetti sonori

Sin dalla fase di progettazione gli effetti sonori rappresentavano uno degli elementi principali con cui si puntava a ricostruire il clima di attesa e malinconia evocato dall'idea decritta nel paragrafo 3.1.1 in merito alle visioni di un gioco che *doveva essere e non è stato*. Per tale motivo nel corso delle riprese presso il Pala Olimpico si sono collezionati tutti i contributi sonori dei pallavolisti in maniera separata, come i palleggi, i respiri, i fiati, i cigolii delle scarpe ecc..

In fase di montaggio, la scelta di voler privilegiare il racconto attraverso un crescendo di adrenalina che si trasformasse in un clima più tetro, come se gli eventi fossero scanditi da un ritmo che non accenna mai a calmarsi fino al culmine della tensione, ha generato un cambiamento nell'accezione narrativa inizialmente preventivata per gli effetti pallavolistici.

Tutte le incursioni legate alle sagome di questi giocatori sono state accompagnate da una stratificazione dalle sonorità proprie della scena (il rumore della rincorsa per schiacciare che produce una serie di cigolii nel tempo di rincorsa, dei gemiti di fatica nell'atto di attacco, il colpo inferto al pallone, il rumore dell'impatto a terra e ancora i fiati una volta terminato lo sforzo) al fine di popolare lo stadio di una moltitudine immaginaria di atleti. Queste sonorità, sovrapposte allo stadio, in chiaro contrasto con lo sua desolazione e all'inquadratura di spalle dell'allenatore hanno la finalità di ricreare da un lato la sensazione che lo stadio prenda vita come se fosse un carattere a sé e dall'altro lato manifestarsi come proiezione del pensiero dell'allenatore.

In fase di mix è stato poi successivamente lavorato affinché questi suoni vorticassero nello spazio attorno all'ascoltatore con la finalità di movimentare la percezione di dinamismo di questi “atleti immaginari”, aumentando la convulsione emotiva scaturita dagli eventi.

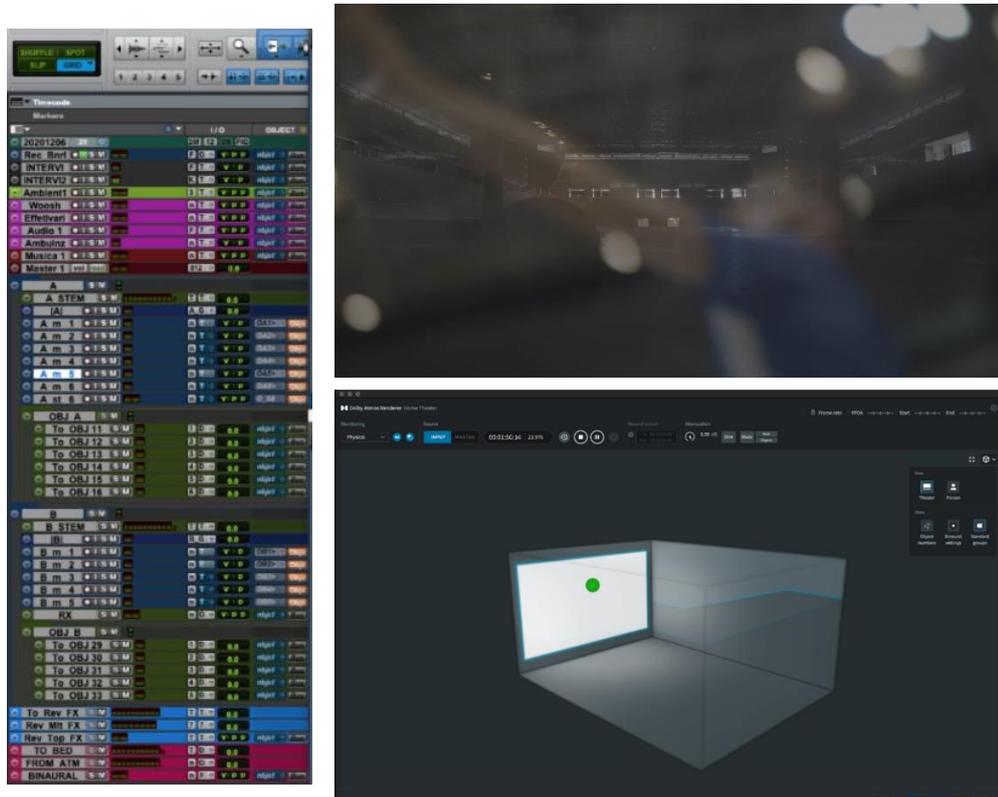


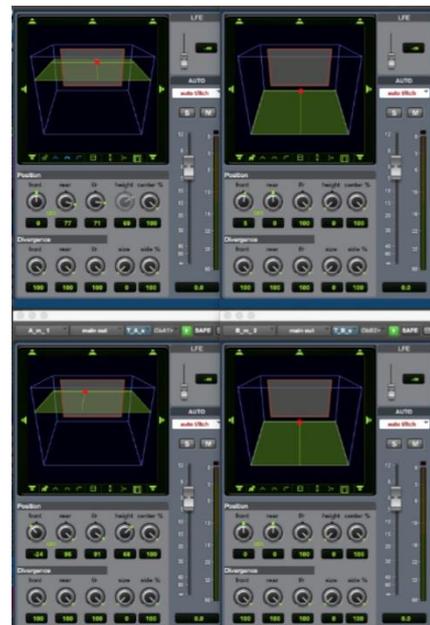
Figura 69 Questa immagine tripartita è tratta dal lavoro di mix in Atmos operato sul teaser. A sinistra è riportata l'interfaccia ProTools attraverso cui gli oggetti (aree blu) in timeline vengono mandati al renderer (immagine in basso a dx) ove monitorare istante per istante le coordinate spaziali impresse all'oggetto sonoro mediante l'operazione di pan pot. In alto a dx invece è raffigurato un frame cui è associato il suono che viene fatto muovere nello spazio.

Sempre attraverso l'utilizzo degli effetti sonori, dopo l'annuncio del rinvio delle Olimpiadi da parte del presidente del CIO, un pesante rimbalzo in attacco di un giocatore chiude definitivamente le sonorità della partita, introducendo lo spettatore in una sorta di ambientazione sonora composta da echi di ambulanze, sottofondi di pioggia e respiri affannosi di un giocatore che ora lentamente diventa nebbiosa sagoma lontana che

svanisce nella schiena di Blengini qualche istante prima che entri nello stadio. In questo modo si dissolve lentamente la dimensione sportiva “onirica” in un richiamo sonoro che, a mio giudizio, ha caratterizzato la quarantena della primavera 2020. L’eco delle ambulanze che, se solitamente si confondono nel traffico cittadino, durante il lockdown sembrava essere divenuto vivido ricordo per chiunque del drammatico qui ed ora che ciascuno di noi stava condividendo. Per chi sarà quell’ambulanza? Verrà dalle nostre parti? Sarà Covid? Inoltre l’idea di mescolare questi suoni alla pioggia nasce in primo luogo per evocare a livello subliminale un senso di malinconia e in secondo luogo perché anche la pioggia in una città senza traffico, vuota dei suoi soliti rumori, ha acquisito una magia ed una malinconia diverse dal solito, proprio come quella pioggia battente che in Piazza San Pietro accompagnava le parole del Papa in una benedizione *Urbi et Otbi* di un mondo sottosopra.

Infine grazie ai canali in altezza dell’Atmos la pioggia è stata fatta provenire dall’alto, una tridimensionalità sonora che è stato possibile restituire anche in binaurale.

Figura 70 Interfaccia di panning relativa ad alcuni oggetti della scena iniziale del teaser, in fase di mix Atmos (come si può vedere dalla dimensione altezza presente).



3. I dialoghi

Ultimo elemento narrativo che ha in un secondo momento assunto un inaspettato sottotesto all'interno del racconto è rappresentato dalle voci delle interviste che intessono il filo logico dell'intera storia. Infatti al di là della naturale funzione descrittiva, grazie ad un suggerimento di Mirko Perri, in fase di montaggio video e conseguentemente anche audio, è stato dato il compito ad una concitata sovrapposizione di voci televisive, a tratti nitide a tratti impastate con i rumori sportivi, il compito di costruire il punto massimo nel climax della tensione.

Nell'evoluzione narrativa del teaser infatti il momento di rottura della tensione della storia è rappresentato dall'annuncio del rinvio olimpico, frutto di una serie di eventi a cascata che nell'arco di una manciata di settimane (trasposti nel video, secondi) sono andati ad acquisire connotati via via sempre più drastici e globali. Giocando dunque sulla leva di una memoria collettiva, poiché ciascuno di noi nel periodo di Marzo 2020 ha visto scorrere davanti ai suoi occhi l'immagine del Presidente del Consiglio così come di tanti altri premieres internazionali, si è fatto emergere il momento storico che a domino non ha risparmiato nessuno. Ottimizzando dunque questa chiave narrativa è stato condotto un montaggio sonoro che mantenesse quasi sempre in primo piano la voce intellegibile del presidente Conte, poiché parlando in italiano avrebbe mantenuto, seppur volutamente a pezzi, il filo del discorso chiaro per l'ascoltatore, inoltre avendo scandito i vari DPCM di questi tempi di Covid, il suo tono avrebbe assunto un certo sottotesto simbolico/subliminale. Dopodiché alla sua voce gradualmente sono state prima inframezzate e poi sovrapposte le voci degli altri interlocutori a loro volta scandite da rumori pallavolistici concitati. Questa stratificazione ha consentito di riuscire a fare da un lato perno su una memoria collettiva per

cui, essendo ricordo recente di tutti, non necessitava d'essere un discorso intellegibile e dall'altro lato ha costruito la sensazione del crescendo della gravità e della convulsione del caos degli eventi che ha condotto in fine, e ora le voci tutte si zittiscono tranne una, al rinvio ufficiale delle Olimpiadi di Tokyo 2020 a Tokyo 2021. Ora sì che le voci tacciono per lasciare spazio solo più a suoni confusi ... fino al completo silenzio, il silenzio di un stadio vuoto, di un gioco fermo, di una realtà che stenta a tornare reale.

Questo crescendo di musiche, effetti sonori, dialoghi per poi cedere il passo al silenzio hanno dunque costruito il crescendo emotivo del teaser.

Da un punto di vista di mix specie in questa fase le voci sono dunque state trattate come oggetti puntiformi che vorticavano attorno (All'allenatore? All'ascoltatore?). Grazie infatti alla guida di Perri, si è preventivato un lavoro di spazializzazione dialogica non poiché il suono seguisse la posizione dei personaggi in scena, i quali al contrario assumevano tutti una posizione fissa centrale, ma poiché si è lavorato sull'amplificazione di un disorientamento generale. In cui il suono rappresentava il punto di vista interno, che poteva essere quello di Blengini, dei giocatori di Pallavolo, ma in realtà di ciascuno di noi poiché tutti in questo tempo storico siamo stati inermi in balia del domino di tali eventi.

3.4 Considerazioni, vantaggi e svantaggi di un workflow immersivo

In generale dunque secondo quanto analizzato sin qui si potrebbero riassumere i pregi di una lavorazione Atmos nei seguenti punti:

- Da un punto di vista estetico è possibile lavorare attraverso un'incredibile definizione spaziale del suono, ampliando i confini dello spazio e dell'immagine bidimensionale inserendo lo spettatore in un contesto davvero immersivo.
- Da un punto di vista operativo i parametri di lavorazione rimangono praticamente identici rispetto ai workflow precedenti, non ci sono nuove competenze particolarmente complesse che un tecnico del suono esperto debba assimilare rispetto a quelle utili in un workflow multicanale.
- Da un punto di vista economico consente ai professionisti del sonoro cinematografico di lavorare ad un solo master che in sé contiene già le caratteristiche necessarie per essere adattato indipendentemente dal sistema di riproduzione, un risparmio questo di tempo e di conseguenza di denaro notevole.
- Da un punto di vista del mercato, grazie agli standard approfonditi nel secondo capitolo, nati anche sulla base del formato Atmos esistente, viene garantita flessibilità di riproduzione a prescindere dalla tipologia di sistema immersivo adottato del consumatore finale.

- Da un punto di vista distributivo lavorare in Atmos non migliora solo la qualità d'ascolto cinematografica ma favorisce anche sistemi di riproduzione individuali quali smartphone e piattaforme streaming, perché grazie alla possibilità di creare da un unico master diversi formati d'ascolto, viene favorita la riproduzione binaurale per l'ascolto in cuffia. Quest'ultima inoltre beneficia particolarmente della conversione Atmos -> binaurale poiché nell'ascolto in cuffia può venir sfruttata la precisione spaziale anche in altezza realizzata appunto in Atmos.

Tuttavia lavorare seguendo workflow immersivi, e nel caso particolare in Atmos, non presenta solo vantaggi, in quanto nuova tecnologia vi sono infatti criticità che rendono ancora lungo, e per certi versi anche a rischio, il processo di totale assimilazione di questo formato all'interno del mercato audiovisivo. Qui di seguito si elencano alcune delle principali motivazioni rilevate:

- Da un punto di vista estetico è vero che il potenziale offerto dall'Atmos può essere d'impatto, ma resta comunque una tecnologia che migliora ma non rivoluziona i sistemi surround già ampiamente esplorati nel corso della storia del cinema. Ciò significa che per valorizzare una colonna sonora Atmos bisogna realizzare film che si prestino in termini visivi, narrativi e di ritmo a questa tipologia di grammatica sonora. Il che richiede a sua volta registi che conoscano lo strumento lo comprendano e che se ne sappiano avvalere. Altrimenti il rischio è di un grande investimento in termini realizzativi per un moderato riscontro in termini estetico-narrativi.

Ad esempio Mirko Perri in occasione della sua sperimentazione del formato Atmos per il film Youth afferma che la sua volontà era

quella di ricreare una vera e propria installazione sonora, in cui le sonorità di un bosco fossero frutto di una ricostruzione puntuale elemento per elemento (uccellini, vento ecc..). Le scene avevano un ritmo lento per cui le immagini non erano movimentate e dinamiche e quindi nella ricostruzione di rapporti fissi tra gli oggetti sonori la delusione una volta arrivato in sala è stata rilevare che lo spettatore a seconda di dove si trovasse seduto “sentiva un film diverso”. In questo senso l’Atmos “non restituisce un’ambientazione tridimensionale solida quando il suono è statico, interviene molto l’acustica della sala, è una cosa che si vince nel momento in cui c’è un film che si muove molto. Se la camera indaga e tu muovi il suono insieme alla camera hai un’impressione più piacevole che trovarti in una posizione che non ha la proporzione adatta a quella che avevi pensato facendolo in un’altra posizione (N.d.R. sala mix)”.

A proposito di difformità d’ascolto rilevate dagli spettatori in sala, è interessante analizzare una ricerca condotta da Tomáš Oramus e Petr Neubauer, pubblicata su un *Convention Paper* della AES (Audio Engineering Society) [41] la quale aveva come obiettivo rilevare attraverso un focus group di ascoltatori un eventuale percepito miglioramento d’ascolto tra una diffusione sonora in Atmos ed una in 5.1. La ricerca è stata condotta sia su un gruppo di ascoltatori esperti (sound designers, studenti della materia) che su un gruppo di persone non esperte, con l’obiettivo di verificare un confronto percettivo tra i seguenti fattori: “Immersione Spaziale (Coinvolgimento)”, “Localizzazione”, “Dinamica”, “Qualità Audio” e “Preferenza” tra i due formati. I gruppi di persone venivano sottoposte alla proiezione sia di porzioni di film che di solo ascolto di campioni sonori nei due diversi formati (5.1 e Atmos), in ordine sparso al fine

di non influenzare il giudizio. Alla fine di ogni ascolto dovevano rispondere a delle domande in merito ai parametri sopra citati su una scala di gradimento successivamente convertita da -4 a +4, ove 0 indicava che il soggetto non aveva rilevato distinzioni tra i due formati, - 4 indicava forte preferenza per il formato 5.1 e +4 per l'Atmos. Le aspettative dei ricercatori erano di rilevare per tutti i parametri indicati un indice di gradimento che propendesse nettamente verso il formato Atmos specie nel gruppo d'ascolto dei professionisti del settore.

Il risultato finale invece è stato che gli spettatori hanno percepito solo lieve differenza tra i due formati, con una debole preferenza dei parametri ascoltati in Atmos salvo per la "localizzazione". In generale attraverso il focus group non è stato possibile provare che gli ascoltatori esperti abbiano potuto rilevare nette differenze rispetto ai non esperti.

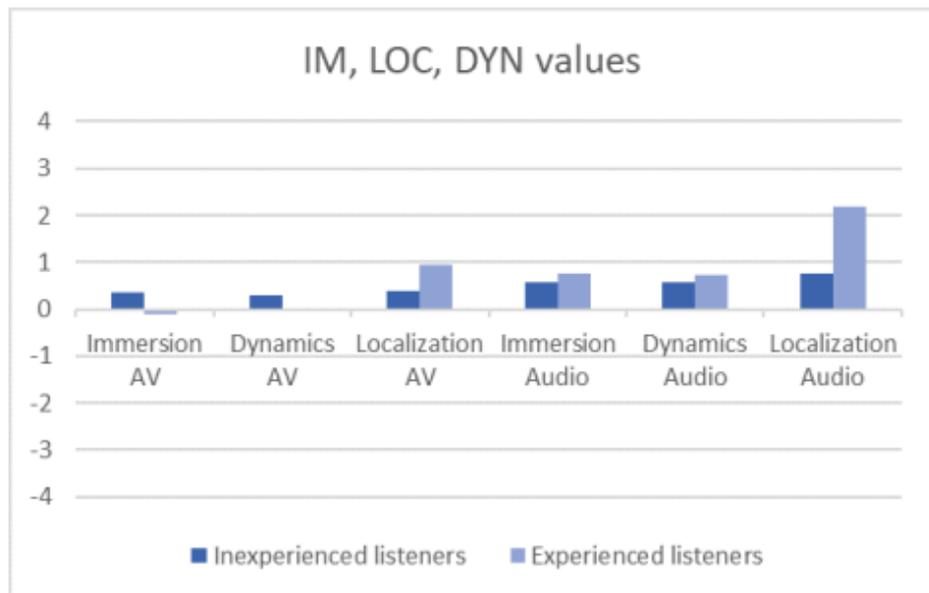


Figura 71 Percezione dei parametri di Immersione, Dinamica e Localizzazione a confronto per i campioni Audiovisivi (AV) e solo Audio. [41]

Dal questionario sulla preferenza i ricercatori concludono che non si può parlare di migliore o peggiore esperienza poiché il focus

group ha espresso preferenze più o meno bilanciate sia per il 5.1 che per l'Atmos, (da considerare che chi ha significato forte preferenza per l'Atmos, 28%, è quasi bilanciato dalla somma di chi ha espresso forte e debole preferenza per il 5.1, 30%) quindi più che esperienza d'ascolto migliore, bisognerebbe dire 'differente'.

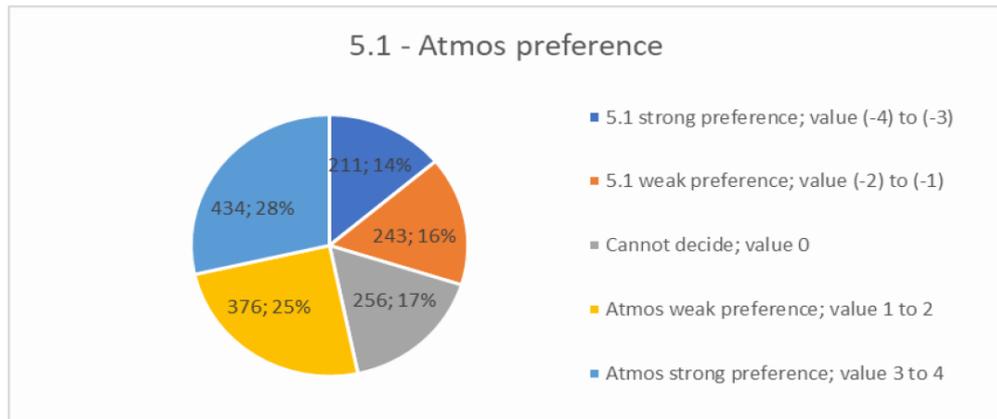


Figura 72 Grafico finale che mostra i gradi di preferenza: debole, forte o indecisione, tra i due formati. [41]

Per un approfondimento in merito alle condizioni certificate in cui sono stati effettuati questi esperimenti e la discussione dei dati si rimanda alla lettura dell'articolo di ricerca. Da i dati sopra riportati possiamo tuttavia evincere che in linea generale l'area di miglioramento dell'Atmos non rappresenta una "rivoluzione" nell'ascolto rispetto ai passati formati. Il fatto che però il parametro localizzazione risulti in generale l'unico con lo scarto maggiore rispetto agli altri fa presagire che uno sforzo congiunto da parte dei creativi a fronte della conoscenza della tecnologia può produrre degli effettivi miglioramenti se restituiti all'interno di un contesto narrativo significativo.

- Da un punto di vista progettuale l'Atmos pur non stravolgendo i workflow classici richiede comunque una maggior lavorazione in fase di realizzazione di un film. Come si accennava in precedenza, ad esempio, non esistono ricche librerie già realizzate di ambienti registrati su 9 canali, spesso questi devono essere creati ad hoc attraverso sessioni di *field recording*, aumentando la mole lavorativa richiesta. Stesso dicasi per la post-produzione. La maggior libertà di gestione dei suoni offerta dai sistemi object-based dall'altro lato richiede più tempo nella realizzazione di ambientazioni sonore adeguate. Come emerso dalle interviste citate in precedenza, un conto è per restituire un'atmosfera urbana registrare la scena di un traffico cittadino. Un conto è invece ricostruire oggetto per oggetto ogni elemento che si muove nello spazio, differenti macchine, di differenti cilindrata che si incrociano, voci di passanti, rumori di lavori stradali, urla, clacson, ciascuno di questi elementi deve essere ricostruito puntualmente e fatto muovere nello spazio in maniera coerente. Un lavoro questo che richiede più tempo ed esperti del settore in grado di ricostruire con consapevolezza paesaggi sonori coerenti.
- Al punto appena espresso in merito alle tempistiche si ricollega la tematica economica. Lavorare in Atmos costa di più che non lavorare in un classico sistema multicanale. La variabile costo è essenzialmente legata ai fattori appena analizzati. Più tempo necessario per realizzare un lavoro, più professionisti impiegati per la sua finalizzazione, più sofisticata è l'expertise richiesta per la creazione del prodotto finale e maggiori sono ovviamente i costi risultanti per una produzione. Tutto questo solo per la fase di progettazione ed editing sonoro, senza considerare il mix finale. A titolo di esempio si cita infatti che una

giornata in sala di missaggio può arrivare a costare ad una produzione italiana fino a 1000 euro al giorno, questo significa che aumentare ulteriormente la mole di lavoro, rendendo dunque più sofisticato il tipo di spazializzazione richiesta in sala mix non può che essere una variabile ulteriore che si aggiunge ai costi generali. In una fase tra l'altro notoriamente molto delicata, quale momento di chiusura del film, in cui solitamente i nervi dei creativi sono più tesi perché resta poco margine per variazioni di sorta. Un problema questo degno di nota se si valuta che rispetto ai canoni americani le tempistiche per la post-produzione sonora di un film in Italia sono piuttosto castigati. Visto e considerato che non è tradizione nostrana pensare al sonoro filmico come un campo all'interno del quale sonorizzare e ricostruire tutto da capo. *Last but not the least* un master in Dolby Atmos ha anch'esso un costo a sé di alcune migliaia di euro.

- Da un punto di vista di mercato infine, è vero che, mettendo a confronto workflow multicanale con workflow immersivo, quest'ultimo richiede più tempo per essere finalizzato ma restituisce una completezza ed ottimizzazione del lavoro finito che il primo invece non garantisce, imponendo che il flusso di lavoro venga ripetuto o comunque modificato per ciascun master realizzato. Tuttavia bisogna considerare che, specie in Italia, l'adozione di questa tecnologia, così come la relativa produzione di contenuti, è così poco diffusa che per una produzione, e di riflesso per un sound designer, sarebbe inutile lavorare alla massima qualità possibile (sistema 9.1) se poi alla fine il film verrà riprodotto al massimo in 5.1. Si moltiplicherebbero tempi e costi e si richiederebbe un carico computazionale non necessario alle rispettive macchine di post-produzione.

Conclusioni

Sulla base degli approfondimenti e delle considerazioni sin qui riportati, l'Atmos e più in generale le tecnologie immersive sono la nuova rivoluzione del sonoro cinematografico?

Stando alle riflessioni raccolte nel corso di questo progetto di tesi sia di natura teorica grazie al confronto con esperti del settore che di natura applicativa come illustrato, la risposta, sperando di non deludere nessuno giunti a questo punto della trattazione, sembrerebbe essere no.

Le tecnologie immersive in termini di cambiamento dell'esperienza filmica, benché apportino certamente delle nuove e più accurate esperienze d'ascolto, non sono paragonabili a quello che fu il terremoto nel passaggio dal mono allo stereo e ancora al pionieristico esperimento surround di *Apocalypse Now*. Walter Murch, Coppola, Burtt, Lucas sono stati i fondatori di un modo di pensare e realizzare il suono di cui ancora oggi raccogliamo i frutti e l'eredità. L'Atmos potrebbe essere forse considerabile se non una rivoluzione senz'altro una frontiera futura per quanto riguarda il workflow di postproduzione. L'idea infatti che attraverso un unico lavoro di mix si possa preservare l'intento creativo di tale prodotto su una pluralità di asset distributivi, rappresenta una svolta rispetto ai limiti tecnici del passato ed una peculiarità che fortemente interpreta il momento storico di fruizione mediale, oltre ovviamente ad agevolare le dinamiche produttive favorendo un notevole risparmio di tempi e costi. Inoltre il moltiplicarsi delle sale di riproduzione Dolby Atmos nel mondo e il lavoro fatto dalla SMPTE nel creare un unico e nuovo formato di distribuzione per tutti i sistemi Sonori Immersivi in circolazione, sono indice di una certa propensione a preparare il terreno a queste configurazioni, benché ancora distanti dall'essere lo standard in uso.

Tuttavia seppur il workflow immersivo consenta un notevole risparmio in termini di realizzazione di “deriverables” i vantaggi da esso derivanti non sono ancora paragonabili ai costi per affrontare invece la produzione vera e propria di un contenuto che potenzi davvero lo strumento. Il rapporto dunque costi benefici non è attualmente equiparabile, per lo meno per quelli che sono gli standard di produzione italiani.

Lo stesso Cuarón pur avendo dato vita ad un film intimista dal taglio più autoriale, per affrontare la realizzazione della complessità delle scene sonore che voleva ricostruire, ha richiesto un lavoro di presa diretta, montaggio e mix sonoro ben più complessi ed onerosi degli standard dell’industria, in un film in cui il budget investito è stato quello di 15 milioni di dollari, cifra neanche lontanamente paragonabile ai budget europei associati al concetto di film d’autore. Il punto è che la qualità in questo senso richiede costi maggiori di realizzazione e laddove produrre un film in Atmos può avere senso solo nella misura in cui ve ne siano utilizzi virtuosi, e non solo per dare un’amplificazione dei volumi generali del film, per il quale possono essere più che sufficienti i sistemi multicanale classici, è legittimo domandarsi quali siano i rapporti costi-benefici, estetico-produttivi derivanti dall’utilizzo di una simile tecnologia.

E’ inoltre vero che la stereofonia stessa, anche dopo essersi stabilita come nuova frontiera successivamente all’uscita di *Star Wars*, incontrò comunque delle resistenze nel tempo da parte dei creativi nella sua adozione. Ogni tecnologia ha il suo tempo di assestamento durante il quale fronteggia destini incerti che ne possono vedere una totale estinzione per poi risorgere in nuove forme rivisitate oppure ancora vedere un lento assestamento del proprio status come un seme piantato in un terreno (e in questo processo fondamentali sono i creativi con cui la tecnologia viene a contatto).

Mirko Perri in merito al panorama italiano odierno cita che la sensibilità della nuova generazione di registi italiani, quali Matteo Rovere di *Veloce come il Vento* e il *Primo re* o Gabriele Mainetti di *Lo chiamavano Jeeg Robot*, sta portando in campo nuovi approcci al modo di intendere la dimensione sonora di un film, realizzando appunto storie che senza una controparte sonora altrettanto efficace, quale l'incisività delle immagini e l'articolazione del racconto, renderebbe l'intero film letteralmente amputato di una dimensione narrativa fondamentale. L'ingresso stesso in Italia di nuovi players quali Netflix e Amazon cominciano a portare una filosofia e dei capitali alternativi tesi a realizzare da un punto di vista qualitativo prodotti diversi rispetto allo standard del mercato cinematografico italiano del passato, in cui per un certo periodo drammi generazionali e commedie sono stati i prodotti principali della nostra cinematografia. Generi questi ultimi che per la tipologia narrativa più legata alla verosimiglianza con la quotidianità meno si prestano a particolari sperimentazioni sonore.

Ma forse il cinema non è la finalità ultima dell'immersività sonora ed è solo questione di tempo, i meeting olografici con cui si sono recentemente incontrati i potenti della terra, Elio Germano che porta la realtà virtuale a teatro, i recenti accadimenti pandemici sembrerebbero aver impresso un'accelerazione a tutte quelle tecnologie che ci consentono di essere "distanti ma uniti", uno slogan politico che sembra divenuto più che altro una (triste?) profezia. Lo smart working, i visori di VR della playstation, gli ologrammi transoceanici, quando incontrarsi di persona non si può più, allora sì che la tecnologia diventa compendio della realtà, prolungamento fisico di una società sempre più distante e che dunque dell'immersività necessita per replicare una sensazione di prossimità. Una risorsa in più? Senza dubbio. Eppure intonava proprio così *l'inferno virtuale* preconizzato da Jullier. Ma non è né oggetto, né obiettivo di questa tesi speculare sul futuro, vizi e virtù,

della deriva tecnologica immersiva, quanto piuttosto interrogarsi come esserne interpreti creativi e se soprattutto questo possa aver un senso in abbinamento ad un medium ereditato dal XX secolo come il cinema, che ha già molto se non tutto di sé visto esplorare e certamente anche senza l'invenzione dell'Atmos avrebbe potuto proseguire il suo corso.

Ci si avvia pertanto alla conclusione di questa tesi, nel tentativo di avanzare una risposta a tale interrogativo attraverso alcune riflessioni del francese Michel Chion, emblematico studioso di questo settore poiché è stato tra i primissimi, persino rispetto agli americani stessi come ammette Walter Murch, a dare al suono cinematografico dignità di trattazione formale [6]:

“It is also characteristic that this silence has been broken by a European rather than an American—even though sound for films was an American invention, and nearly all of the subsequent developments (including the most recent Dolby SR-D digital soundtrack) have been American or Anglo-American.”

L'autore nel 1994 faceva osservazioni che potrebbero essere tranquillamente ricontestualizzate al presente e che tra l'altro rappresentano anche il fulcro riflessivo e di approfondimento di questo intero progetto di ricerca. Egli infatti nel paragrafo *Verso un cinema sensoriale* osserva che nonostante le limitazioni tecniche alla fine degli anni Venti registi come Eisenstein, Epstein e Murnau avevano già avuto un interessante approccio fisico e sensoriale al film, mentre riferendosi ai registi contemporanei afferma: “Pochissime delle loro controparti oggi sono innovatori pronti ad affrontare le sfide di nuove possibilità tecniche, in particolare per quanto riguarda il suono Dolby. Sintomo, forse, di una nuova tappa nell'eterna "crisi" del cinema. Francamente, molti registi europei hanno semplicemente ignorato l'incredibile mutazione provocata dalla standardizzazione Dolby.”

Chion prosegue infatti osservando come registi quali Fellini, Kubrik e Godard abbiano fatto un uso a suo giudizio poco fantasioso e non degno di nota del nuovo strumento Dolby, osservando che nessuno di loro ha offerto spunti originali e che la *Nouvelle Vague* stessa tornò addirittura all'utilizzo della tecnica monofonica.

“Potremmo continuare l'elenco e notare che dal più vecchio (Bresson) al più giovane (Carax), sembra esserci una gara di chi può mostrare il minimo entusiasmo per le nuove risorse sonore: quasi tutti le trascurano o le usano senza inventare qualcosa di nuovo.”

Un paradigma questo che chi lavora nel mondo del sonoro conosce bene, da sempre l'audio combatte la sua battaglia dell'invisibilità rispetto alla sua controparte visiva. L'immagine stimolando il prediletto senso della vista è innegabile che rappresenti un campo di ricerca che gerarchicamente occupa una posizione privilegiata rispetto al suo alleato sonoro. Ma tornando alle riflessioni di Chion egli, visto il generale scarso interesse degli autori, avanza un parallelismo interessante tra regia e compositori di musica classica:

“Cosa offre Dolby stereo a un regista? Niente di meno che l'equivalente di un pianoforte a coda a otto ottave, quando ciò che si aveva prima era un montante di sole cinque ottave, meno potente e meno capace di sfumature. In breve, Dolby offre un guadagno di risorse a livello di spazio sonoro e dinamica del suono che, ovviamente, nessuno è obbligato a utilizzare tutto il tempo ma che è comunque disponibile.”

“Ricordiamo che Beethoven ha scritto le sue sonate per pianoforte per uno strumento più piccolo del pianoforte di oggi: dove lui ha raggiunto i limiti della sua tastiera, oggi abbiamo altre due o tre ottave. Da questo punto di vista sarebbe forse più corretto suonare Beethoven con un pianoforte della sua epoca. Ma ci sarebbe qualcosa di assurdo nel vedere i compositori di oggi scrivere brani per pianoforti moderni con gli stessi limiti di quelli che

costringevano l'autore della *Pathetique*. Lo chiameremmo lavorare con i paraocchi; e questo è precisamente quello che stanno facendo molti registi in questi giorni, indipendentemente da qualsiasi questione finanziaria.”

A questo punto l'autore si avvia ad una conclusione avanzando un'analisi sul rapporto tra densità sonora e silenzi, che ritengo possa essere chiave e spunto altrettanto interessante da assimilare nelle odierne prospettive immersive, che ricorda, aggiungerei, la regola percettiva del confronto secondo cui un bianco è ancora più bianco se accostato ad un nero:

“Scrivere in grande non significa necessariamente riempire tutto lo spazio disponibile. Significa che anche quando scrivi solo una singola nota o linea melodica, lo spazio vuoto intorno alla nota è più grande. Dolby stereo aumenta la possibilità di vuoto nell'audio del film e allo stesso tempo amplia lo spazio che può essere riempito. È questa capacità di vuoto e non solo di pienezza che offre possibilità ancora da esplorare. Kurosawa ha sfruttato magnificamente questa dimensione in *Dreams*: a volte l'universo sonoro si riduce a un unico punto: il suono della pioggia, un'eco che scompare, una semplice voce.”

La “scatola si ingrandisce”, benché il confronto *Atmos-Surround* si assottigli rispetto allo *Stereo-Mono*, gli strumenti aumentano e lo spazio filmico potenzialmente muta le sue prospettive, ma si torna a ricordare, solo se *sensibili figure registiche* raccolgono la sfida e si interrogano sulla natura degli strumenti a disposizione quali nuovi chiavi linguistiche del film, le sinestesie delle *immagini sonore* incontrano l'ossimoro della *rivoluzione silenziosa* Dolby di cui ancora oggi raccogliamo l'eredità degli evoluti frutti tecnologici, in una costante ricerca di ridefinizione dei confini del film, nella tutela della sua identità ma al contempo espandendone gli orizzonti, alla ricerca del controverso ed affascinante equilibrio tra immagini e suoni.

Bibliografia e Sitografia

- [1] A. D'Aloia e R. Eugeni, «Philo-, Neuro-, Post. Che cos'è e cosa sarà la teoria del cinema.,» in *Teorie del Cinema il dibattito contemporaneo*, Raffaello Cortina Editore, 2017.
- [2] M. Perri, «Estetica Sonora Cinematografica,» Novembre 2020. [Online]. Available: <http://www.mirkoperri.com>.
- [3] M. Di Donato, «Lo sviluppo tecnologico della cinematografia sonora dalle origini all'avvento del digitale.» in *L'occhio che ascolta: studi sui rapporti suono-immagine nella forma cinematografica*, Roma, Lithos, 2006, pp. 11-43.
- [4] M. Costin, Regia, *Making Waves: The Art of Cinematic Sound*. [Film]. Stati Uniti: Good Movies Entertainment, 2019.
- [5] L. Jullier, *Il cinema postmoderno*, Torino: Edizioni Kaplan, 2006.
- [6] M. Chion, *Audio-Vision Sound on Screen*, New York: Columbia University Press, 1994.
- [7] L. F. Donaldson, *Feeling and Filmmaking: The Design and Affect of Film Sound.*, Londra: University of West London, 2015.
- [8] V. Sobchack, «Quello che le mie dita sapevano. Il soggetto cinestesico o della visione incarnata,» in *Teorie del Cinema il dibattito contemporaneo.*, Raffaello Cortina Editore, 2017.
- [9] F. Casetti, «Post-, Grand, classica, o “travirgolette”. Cos'è e cos'è stata la teoria del cinema.» in *Teorie del Cinema il dibattito contemporaneo*, Raffaello Cortina Editore, 2017, pp. 373-385.
- [10] M. Mariani, *Il suono per il cinema. Elementi per l'analisi e la progettazione della materia sonora nell'audiovisivo.*, Novara: UTET Università, 2020.
- [11] G. Latini e AITS, *Youth e il Dolby Atmos. Valenza del suono e Workflow Audio nel film di Paolo Sorrentino.*, [Online] Associazione Italiana Tecnici del Suono, 2016.

- [12] A. Cuarón, Interview, *The Sound of Gravity*. [Intervista]. 2013. Available: <https://soundworkscollection.com/>
- [13] S. Livesay e A. Cuarón, Interview, *The Sound of Roma with Skip Lievsay*. [Intervista]. 2019. Available: <https://soundworkscollection.com/>
- [14] G. Ravasi, «Interrogativi antropologici contemporanei,» *Munera Rivista Europea di Cultura.*, 2019.
- [15] Ofcom, «Interactive Report of Media Nations UK,» 2019.
- [16] M. Iqbal, «Business of Apps,» Netflix Revenue and Usage Statistics, 30 Ottobre 2020. [Online]. Available: <https://www.businessofapps.com/data/netflix-statistics/>.
- [17] J. Walsh, «Forbes,» 20 Ottobre 2020. [Online]. Available: <https://www.forbes.com/sites/joewalsh/2020/10/20/netflix-subscriber-growth-slows-after-surgings-during-pandemic/> .
- [18] I.-R. B. 775-3, «Multichannel stereophonic sound system with and without accompanying picture.,» 2012.
- [19] F. Rumsey, «Surround Sound,» in *Immersive, Sound, the art and science of binaural and multi-channel audio*, New York, Routledge, 2018, pp. pp. 180-217.
- [20] K. Sungyoung, «Height Channels,» in *Immersive Sound, the art and science of binaural and multi-channel audio*, New York, Routledge, 2018, pp. 40 - 56.
- [21] B. Braxton, «History of 3D Sound,» in *Immersive Sound, the art and science of binaural and multi-channel audio*, New York, Routledge, 2018, pp. 40 - 56.
- [22] B. Vessa, Interview, *What is Immersive Audio and why is so cool?*. [Webcast] [smpteconnect](https://www.smpteconnect.com/). 14 Agosto 2018.
- [23] I. Unit, «Advanced sound system for programme production. Recommendation ITU-R BS. 2051-0,» 2014.

- [24] B. A. Technologies, *AUROMAX®: Next-Generation Immersive Sound System. White Paper.*, 2015.
- [25] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama e A. Ando, «22.2 Multichannel Sound System for Ultra High-Definition TV,» in *SMPTE Technical Conference Publication ·October 2007*, 2007.
- [26] S. Bowling, Interview, *About Dolby Atmos*. [Intervista]. 2012.
- [27] D. L. Inc, *Dolby® Atmos® Next-Generation Audio for Cinema White Paper (Issue 3)*, 2014.
- [28] S. Cosette, Interview, *The Ins and Outs of ST 2098- 2 Immersive Audio Bitstream*. [Webcast] [smpteconnect](https://www.smpteconnect.com). 27 Agosto 2018.
- [29] M. Goldamn, «[smpte.org/blog/](https://www.smpte.org/blog/),» settembre 2019. [Online]. Available: <https://www.smpte.org/blog/immersive-audio-rollout-looms>.
- [30] J. Riedmiller e N. Tsingos, *Recent Advancements in Audio – How a Paradigm Shift in Audio Spatial Representation & Delivery Will Change the Future of Consumer Audio Experiences*, 2015.
- [31] N. Tsingos, «Object Based Audio,» in *Immersive Sound, the art and science of binaural and multi-channel audio*, New York, Routledge, 2018, pp. 244 - 272.
- [32] H. Choi, «An alternative implementation of VBAP with graphical interface for sound motion design,» in *Proceedings of the 18th International Conference on Auditory Display, Atlanta, (GA) USA*, 2012.
- [33] A. Roginska, «Binaural Audio Through Headphones,» in *Roginska A., Immersive Sound, the art and science of binaural and multi-channel audio*, New York, Routledge, 2018, pp. 88-122.
- [34] B. Technologies, *AURO-3D® HOME THEATER SETUP - Installation Guidelines*, 2015.
- [35] I. D. Technologies, *Dolby Atmos® Home Theater Installation Guidelines*, 2018.

- [36] Zhang, Samarasinghe, Chen e Abhayapala, *Surround by Sound: A Review of Spatial Audio Recording and Reproduction*, mdpi.com/journal/applsci, 2017.
- [37] M. Perri, *La mia prima volta in Dolby Atmos*, 2016. [Online]. Available: <http://www.mirkoperri.com>.
- [38] D. Gould, *Creating in Dolby Atmos Part 2: Atmos post-production and mixing solutions*, Dolby Institute Tutorial Series.
- [39] D. Gould, *Part 1: What is Dolby Atmos and object-based mixing?*, Dolby Institute Tutorial Series.
- [40] D. Gould, *David Gould. Part 3: Atmos Renderer application versions*, Dolby Institute Tutorial Series.
- [41] T. Oramus e P. Neubauer, «COMPARISON STUDY OF LISTENERS' PERCEPTION OF 5.1 AND DOLBY ATMOS.» in *Convention Paper 10278 Presented at the 147 th Convention*, New York, 2019.
- [42] A. Roginska e P. Geluso, «Introduction,» in *Immersive, Sound, the art and science of binaural and multi-channel audio*, New York, Routledge, 2018, pp. 1 - 5.

Ringraziamenti

Al Prof. Enrico Maria Verra, Professore e Maestro, per la meticolosa concretezza investita nella realizzazione dell'intero percorso di tesi, per la fiducia attribuitami, per il lungimirante realismo che non ha mai smesso d'essere generosamente formativo e "in ascolto" e per aver condiviso, con la Sua distintiva sobrietà, la saggezza della Sua professione, per me guida indispensabile per la concretizzazione di questo progetto.

Al Prof. Marco Carlo Masoero, per il prezioso sostegno accademico offerto alla mia persona nel corso di questo tragitto di tesi e per l'accurata struttura conferita al Suo corso di formazione magistrale *Ingegneria del Suono*, occasione formativa importante per gli studenti da cui in parte è nato anche questo stesso progetto.

A Mirko Perri, guida acuta, sensibile e concreta, per essersi generosamente speso nel non far mai mancare la Sua disponibilità, senza il cui competente e illuminante discernimento tante fondamentali riflessioni teoriche ed applicative contenute in questa tesi non sarebbero state possibili. E.... per l'onore ricevuto nell'aver messo direttamente mano al suono del mio progetto.

A Gianlorenzo Blengini, che a 10 anni, quando vidi la Nazionale di Pallavolo in TV salire sul podio olimpico, mai avrei pensato che il Suo allenatore di qualche tempo dopo un giorno mi avrebbe concesso l'onore d'essere il protagonista del lavoro documentaristico che caratterizza la mia tesi di laurea. Grazie Chicco, per la franchezza, la disponibilità e la serietà che hai sempre riservato alla mia persona!

Alla Pallavolo che un tempo mi ha rapito e poi spezzato il cuore, ma che grazie a quella ferita oggi ho trovato la mia strada.

Al Politecnico di Torino, ricordando ancora quei computer dei test di ammissione che a 19 anni mi sembravano immensi tanto da potermi inghiottire e che invece si sono rivelati nel tempo insospettabili amici. A quelle ampie aule di Analisi I, in cui 300 studenti riuniti sotto il cappello di tre semplici iniziali di un cognome si guardavano attorno chiedendosi se ce l'avrebbero fatta a uscirne vivi e....

Wow! Che incredibile avventura che è stata!

A tutti coloro che sono stati linfa e sostegno quali preziosi tecnici, strategici consiglieri, fondamentali supporter, di *Torino Tokyo 2020*, per cui questo progetto di tesi ha combattuto strenuamente al fine di porsi come sviluppo tanto teorico quanto applicativo. Alla Dott.sa Rachele Venco e il Dott. Daniele Donati per averci aperto le porte del Pala Olimpico, suggestiva dimora di musica e sport della nostra città, emblema di resistenza di questo periodo storico, in cui loro ci hanno riservato accoglienza e disponibilità oltre che condiviso i difficili momenti di cui sono, e noi tutti ai vari livelli siamo, protagonisti. A Mirko Guerra, professionista d'esperienza che ha generosamente sostenuto il profilo di ricerca accademico oltre che di qualità sonora di questo lavoro. A Domenico Bruzzese, alias Dominique (e ai tuoi ragazzi!), per lo spirito imprenditoriale e il guizzo intelligente del tuo occhio, strumento indispensabile al tuo tocco fotografico fondamento prezioso per la realizzazione di questo progetto. A Jacopo Capurso, per l'intelligenza acuta ed una non convenzionale concezione del lavoro che lo rendono collega ed amico, e un giorno gli auguro grande produttore, poliedrico e competente, senza il quale le riprese al Pala Olimpico non sarebbero state quello che sono state. Ad Eugenia Ravezzani amica e legale brillante che generosamente nel

momento del bisogno mi ha aiutata con la sua competente professionalità. E ancora... ad Agostino e i ragazzi di Lacumbia, a Letizia Salerno Pittalis, a Stefano Capello, Francesca Piovano, Gianfranco Cumino, Enrico Agosti, Alberto Vittone e infine a Marco Furlani che, grazie al Prof. Enrico Verra, è stato ponte prezioso e di cara disponibilità consentendo il contatto con Mirko Perri. E a tutti coloro che sono stati puzzle infinito o infinitesimale ma ugualmente vitale nella realizzazione di questo progetto!

E...passando ai ringraziamenti *emotional*...

Ai miei amici, che se mi guardo dentro non riesco più a distinguere la mia forma se non grazie al contributo che ciascuno di essi ha generosamente dato alla mia persona. [...] A tutti Voi amici miei dedico e auguro in queste emblematiche pagine che concludono per me un percorso di studi importante che la vita possa donarvi sulla base di tutto ciò che più desiderate quanto preziosamente meritate da questo Viaggio che ha voluto che i nostri destini si intrecciassero!

A Michael Jordan e Russel Crowe, a Christopher Nolan e *The Last Dance*, ad Hans Zimmer e *Il Gladiatore*, a tutti quei film, quei registi, quelle colonne sonore, quelle icone che hanno nutrito i miei sogni e ristorato la realtà del mio presente e i piani del mio futuro.

A Sandro, Luciano, Shakespeare e tutti gli Invisibili, *Uno* fra tutti in particolare, che da *Dio Onnipotente o Incompetente* (e ancor prima) ad oggi sono esercito segreto dei miei "Sogni".

E infine ma anche e soprattutto in principio...

Al mio Faro nella Notte dei sentieri più bui quelli della Vita e quelli dell'Anima. Maestro, Guida, Casa. A Te che da sempre insegni ai pesciolini rossi a danzare contro correnti di tonni. A Te, che nulla è cambiato da quel

nastro segnaletico bianco e rosso nei freddi inverni del Ruffini, agli “shampoo” del *tutto e per tutto* tanto tre giorni prima dell’esame, quanto a questo famoso mese prima della Laurea. A Te sola che hai visto. A Te sola che hai sempre visto e saputo. A Te dedico ogni sforzo, sospiro, dubbio, dolore, gioia che solo Tu puoi sapere esserci realmente stato dietro questo viaggio. E che senza di Te nulla sin dal *principio* sarebbe stato.

Grazie Mamma!