



**SAPIENZA**  
UNIVERSITÀ DI ROMA

# **La cognizione dello spazio sonoro filmico: un approccio neurofilmologico**

**Facoltà di Lettere e Filosofia**

**Dipartimento di Storia dell'arte e spettacolo**

**Dottorato in Musica e spettacolo, XXX ciclo**

Curriculum Studi di teatro, arti performative, cinema e tecnologie per  
lo spettacolo digitale

**Candidato**

**Valerio Sbravatti**

**1642274**

Tutor

Prof. Mauro Di Donato

A/A 2016/2017

*Alla memoria di mio padre,*

*perché tutto è perdonato e ha vinto l'amore*

## Indice

Introduzione.....	5
Ringraziamenti.....	7
1 – Profilo storico della stereofonia: tecnologia e prassi.....	8
1.1 – Musica.....	8
1.1.1 – Il suono spaziale.....	8
1.1.2 – La nascita della stereofonia.....	8
1.1.3 – La diffusione della stereofonia.....	10
1.1.4 – La quadrifonia.....	11
1.1.5 – <i>Ambisonics</i> .....	12
1.1.6 – Il sonoro multicanale.....	12
1.2 – Cinema.....	13
1.2.1 – L’epoca mono.....	13
1.2.2 – L’avvento della stereofonia.....	14
1.2.3 – L’epoca del Dolby Stereo.....	21
1.2.4 – L’avvento del sonoro digitale: lo standard 5.1.....	27
1.2.5 – Nuove frontiere del cinema digitale: <i>Immersive Sound</i> .....	30
1.2.6 – Questioni filologiche.....	36
1.3 – Videogiochi.....	39
1.4 – Realtà virtuale.....	41
2 – Teoriche sullo spazio sonoro cinematografico.....	43
2.1 – Dalle origini del sonoro agli anni ’40.....	43
2.1.1 – Ingegneria.....	43
2.1.2 – Teoria del cinema.....	49
2.2 – Anni ’50-’60.....	50
2.2.1 – Ingegneria.....	50
2.2.3 – Teoria del cinema.....	53
2.3 – Anni ’70-’80.....	54
2.3.1 – Prassi.....	54
2.3.2 – Teoria del cinema.....	54
2.4 – Anni ’90: Stati Uniti.....	61
2.5 – Anni 1990-2000: Francia.....	69
2.6 – Dal 2000 a oggi.....	73
2.6.2 – Studi anglofoni.....	73
2.6.3 – Italia.....	93
2.7 – La collocazione dei suoni.....	97
3 – Studi cognitivisti sul cinema.....	108
3.1 – Profilo generale.....	108
3.1.1 – Teoria del cinema: dal cognitivismo classico alla Neurofilmologia.....	108
3.1.2 – Scienze cognitive.....	115
3.2 – La cognizione del sonoro filmico.....	117
4 – Teoriche sull’audio spaziale.....	125
5 – Aspetti tecnici e cognitivi dello spazio sonoro filmico: teoria e analisi.....	137
5.1 – Suono e spazio.....	137
5.2 – Cognizione incarnata e simulazione incarnata.....	138
5.3 – Integrazione multisensoriale e interazioni crossmodali.....	141
5.4 – La firma spaziale.....	147
5.5 – La spazializzazione acustica.....	154
5.5.1 – Terminologia e prassi.....	154

5.5.2 – La direzione.....	158
5.6 – Movimenti sonori.....	161
5.7 – Campo visivo e campo sonoro.....	165
5.7.1 – Questioni generali.....	165
5.7.2 – Il punto d’ascolto.....	169
5.7.3 – Campo e fuori campo.....	172
5.7.4 – Sintassi sonora: raccordi e ponti.....	173
5.7.5 – Spazio immersivo e spazio virtuale.....	177
5.8 – Spazio della storia e spazio del discorso.....	188
6 – Studi di caso.....	200
6.1 – <i>All’ovest niente di nuovo (All Quiet on the Western Front, USA 1930)</i> .....	200
6.2 – <i>La finestra sul cortile (Rear Window, USA 1954)</i> .....	202
6.3 – <i>Persona (Id., Svezia 1966)</i> .....	204
6.4 – <i>2001: Odissea nello spazio (2001: A Space Odyssey, USA/Regno Unito 1968)</i> .....	205
6.5 – <i>Lezioni di storia (Geschichtsunterricht, Italia/Germania Ovest 1972)</i> .....	210
6.6 – <i>Guerre stellari (Star Wars, USA 1977)</i> .....	212
6.7 – <i>Prénom Carmen (Id., Francia 1983)</i> .....	215
6.8 – <i>Strange Days (Id., USA 1995)</i> .....	218
6.9 – <i>Timecode (USA 2000)</i> .....	221
6.10 – <i>Elephant (Id., USA 2003)</i> .....	222
6.11 – <i>Gravity (Id., Regno Unito/USA 2013)</i> .....	224
6.12 – <i>La casa (Evil Dead, USA 2013)</i> .....	225
6.13 – <i>Salvo (Italia 2013)</i> .....	228
Conclusioni, limiti e futuri sviluppi.....	231
Bibliografia e sitografia.....	232
Filmografia.....	270

## Introduzione

Questa tesi è dedicata all'aspetto spaziale del sonoro filmico, indagato nei suoi fattori storici, tecnologici e cognitivi. Normalmente distingo i termini *suono* e *sonoro* (salvo usare sporadicamente il primo nel senso del secondo per evitare troppe ripetizioni), poiché suono è un iperonimo mentre sonoro si riferisce più strettamente a un insieme di suoni registrati.

Il mio ricorso alle scienze cognitive muove dalla volontà di rintracciare possibili risposte, chiavi di analisi, strade di ricerca, senza la pretesa di gestire con piena competenza ogni nozione – la verifica sarà eventualmente responsabilità degli/delle scienziati/e delle discipline in questione. A questo proposito avverto il bisogno di palesare una riflessione, considerando il clima culturale in cui mi sento collocato sin dall'inizio del dottorato. Il mio approccio riduzionistico, tipico di teorici e teoriche cognitivisti/e, non intende essere riduttivo: il mio interesse è quello di avviare una ricerca sui fenomeni neurocognitivi coinvolti nella percezione spaziale del sonoro filmico, che sono la base psicofisiologica su cui si sviluppano fenomeni culturali, sociali, estetici. Sono perfettamente consapevole che la materia da me affrontata, come qualsiasi altra questione di ambito umano, non debba essere presa come il risultato diretto di una predisposizione biologica. È mia opinione tuttavia che qualsiasi studio umanistico debba accettare il fatto che siamo anzitutto degli organismi con delle caratteristiche innate e universali, che sottendono ogni processo culturale. Il rapporto tra natura e cultura non dovrebbe essere inteso in modo manicheo, bensì in modo integrato. La disamina che qui propongo ambisce pertanto a essere discussa, verificata e approfondita con gli strumenti teorici e pratici delle scienze cognitive, nonché criticata, integrata e sviluppata in altri studi che, al contrario del mio, privilegiano questioni prettamente culturali e ideologiche. Per tale ragione mi ispiro all'approccio di studiosi quali Grodal (2015), Bordwell (2010), Gallese e Guerra (2015) e Smith (2017), che in termini filosofici generali può essere definito bio-culturalista o naturalista. Più in particolare sono debitore della proposta metodologica di D'Aloia e Eugeni (2014) definita Neurofilmologia, che si basa su un concetto chiave: lo/la spettatore/spettatrice – che io preferisco denominare *percipiente*, per ridurre l'“oculocentrismo” ed enfatizzare l'integrazione multisensoriale – è concepito/a non solo come un *cervello* (sede dei processi mentali) né solo come un *corpo* bensì come *organismo*, ovvero per dirla con Gallese e Guerra (2015) come un sistema cervello-corpo.

I primi quattro capitoli presentano le conoscenze storiche e teoriche necessarie alla disamina che segue. In particolare il capitolo 1 contiene un profilo storico della stereofonia cinematografica, con qualche cenno su applicazioni di diverso tipo poiché ritengo doveroso conoscere per sommi

capi anche l'uso della stereofonia in altre forme artistiche e comunicative al fine di comprendere le possibili analogie, differenze, intersezioni e influenze reciproche. Ciò dipende sia dal fatto che l'audio è sempre audio, ossia che la tecnologia è in senso ingegneristico fondamentalmente la stessa in tutte le applicazioni, sia dallo scenario sempre più crossmediale in cui il cinema è inserito. In termini nazionali la cinematografia maggiormente rappresentata è quella statunitense, sia perché è in essa che è avvenuta la quasi totalità di investimenti su nuovi sistemi sonori, sia perché è la cinematografia che conosco meglio, insieme a quella italiana sulla quale purtroppo le informazioni e la letteratura sono carenti.

Il capitolo 2 consiste in una sistemazione di riflessioni teoriche sullo spazio sonoro filmico, per l'elaborazione della quale ho attinto a una letteratura in italiano, in inglese e marginalmente in francese. I capitoli 3 e 4 contengono una succinta esposizione di altri utili spunti teorici, pertinenti nel primo caso agli studi sul cinema, propri della teoria del cinema o delle scienze cognitive, e nel secondo caso all'audio spaziale; diversamente dal capitolo precedente, qui non ho alcuna pretesa di esaustività, poiché mi limito ai testi che hanno influenzato la mia riflessione, con la consapevolezza di escludere molte altre ricerche. Lo studio di questa letteratura mi ha permesso di elaborare la parte originale della mia ricerca, sviluppata negli ultimi due capitoli: il capitolo 5 consiste nell'esposizione di una teoria e di un metodo di analisi in cui faccio appunto ricorso alle conoscenze degli ambiti scientifici di cui nei capitoli precedenti, con ulteriori approfondimenti, mentre il capitolo 6 presenta degli studi di caso in cui applico detto metodo. Sebbene il mio oggetto di studio sia il cinema, il metodo si propone di essere idoneo all'analisi di qualsiasi prodotto audiovisivo.

## Ringraziamenti

Ringrazio la mia compagna Ilaria, per l'amore, la stima e il sostegno costanti che mi esprime. Ringrazio le persone che hanno dato un contributo più o meno diretto a questo lavoro: Federico Savina, per avermi dedicato il 12/10/2016 un bel pomeriggio di conversazione a casa sua, con passione e cordialità (ricordo la tenerezza delle tazze per il tè già predisposte sul tavolo); Gianfranco Callarello, per i suoi preziosi materiali e la gratificante corrispondenza; Jay Beck, Mark Kerins, Peter P. Lennox, Bo N. Sherkman, Ana Tajadura-Jiménez, per avermi fornito o indicato testi; Lucia Cosi della Biblioteca di Storia della Musica della Sapienza, per la celerità nella consegna dei numerosi documenti da me richiesti; il mio collega di dottorato Sergio Lo Gatto, per avermi fatto conoscere il software Zotero che mi ha enormemente aiutato nella stesura della bibliografia; Luca Bandirali, cui continuo a esprimere riconoscenza per avermi stimolato a occuparmi di questa materia, nel 2012; infine Sergio Bassetti, Raymond Bellour, Daniele Buttafava, Enrico Carocci, Cristiano Ciccone, Simone Corelli, Vittorio Gallese, Michele Guerra, Niccolò Palazzetti, Filippo Ulivieri, e tutti/e gli/le appassionati/e di sonoro che animano blog e forum con interessanti e appassionanti riflessioni e condividendo materiali.

# 1 – Profilo storico della stereofonia: tecnologia e prassi

## 1.1 – Musica

### 1.1.1 – Il suono spaziale

Come osserva Holman (2008: 20-21), la stereofonia ha i suoi precursori nella musica eseguita dal vivo. Nel XVI secolo, nella basilica di San Marco a Venezia, si esaltò la tecnica denominata “a coro/i spezzato/i”, che aveva avuto degli antecedenti, con «la divisione dei cantori in semicori che s’alternano». Si arrivò alla «dispersione dei gruppi corali in luoghi diversi della chiesa, con effetto stereofonico, attivando forme di sonorità policentrica, che indicano il gusto barocco arrivante» (Gallico 1991: 75-76). Altri casi rilevanti si riscontrano sporadicamente anche nei secoli seguenti, aumentando sempre più fino al XX secolo in cui, soprattutto con il ricorso all’elettroacustica, lo spazio divenne sempre più un principio compositivo. Si considerino alcuni esempi importanti. La seconda orchestra prevista da Gluck nel secondo atto di *Orfeo e Euridice* (Vienna 1762) doveva essere collocata “derrière le théâtre”. Il *Notturmo* (o *Serenata n. 8*) in Re maggiore di Mozart K. 286 (1776-1777) è strumentato per quattro orchestre (archi e due corni) con effetti di eco, benché il compositore non specifichi la loro collocazione spaziale. Per la *Sinfonia fantastica* (1830), Berlioz richiese che alcuni strumenti fossero posizionati “derrière la scène”, mentre nel *Requiem* (1837) egli inserì quattro gruppi di ottoni posizionati ai quattro angoli del resto dell’organico strumentale e vocale; Mahler si avvale di strumenti dislocati in alcune sue sinfonie (tra fine Ottocento e inizio Novecento); Respighi indicò nella partitura di *I pini di Roma* (1924) che le buccine nell’ultimo movimento suonassero isolate (cfr. Solomon 2007).

### 1.1.2 – La nascita della stereofonia

Le prime registrazioni sonore erano monofoniche, cioè costituite da un solo canale sonoro, pertanto «The only “spatial” cues possible [...] were hints at distance and depth provided by reverberation» (Rumsey 2001: par. 1.4.1). Come osservano Théberge, Devine e Everrett, le origini culturali della stereofonia sono rintracciabili nel concetto di ascolto binaurale, sorto nel XIX secolo nel contesto della nascita delle scienze acustiche<sup>1</sup> che comportarono tra l’altro «a paradigm shift. [W]e note a transition from sound and listening as non-spatial phenomena to fundamentally spatial phenomena».

<sup>1</sup> Cfr. la riflessione di Sterne (2003: 114-115) sullo stetoscopio binaurale.

È su questa base concettuale che si sviluppano le ricerche sul suono spaziale. Il primo esempio di sonoro a due canali pare essere costituito dall'esperimento di Théâtrophone di Clement Ader all'esibizione di Parigi del 1881, con cui egli trasmise musica eseguita al Teatro dell'Opera della città usando più microfoni e inviando il segnale a coppie di ricevitori telefonici, per un ascolto binaurale (2015a: 5-6)<sup>2</sup>. Circa la radio, si ritiene che la prima trasmissione a due canali sia quella di un concerto a Manchester da parte della stazione 5XX della BBC, nel 1925 (Théberge, Devine e Everett 2015b: 268).

Il concetto di “auditory perspective” nacque ai Bell Laboratories negli anni '30 nel contesto di esperimenti di Harvery Fletcher e colleghi (cui partecipò il celebre direttore d'orchestra Leopold Stokowski): con tale espressione si designava «a mode of sound reproduction and a mode of listening». Per questa ragione venne ideato Oscar, un manichino con microfoni al posto delle orecchie atti a registrare suoni simulando l'ascolto umano; tali suoni venivano poi ascoltati in cuffia (Théberge, Devine e Everett 2015a: 15-17). Oltre a simili esperimenti, detti binaurali, gli ingegneri Steinberg e Snow si dedicarono anche a «less ideal arrangements» (Steinberg e Snow 1934: 12), cioè combinazioni di due microfoni e due altoparlanti, apprezzati da critici musicali (cfr. Théberge, Devine e Everett 2015a: 17). Essi in particolare sperimentarono con tre canali, ritenendo i risultati convincenti, mentre con due canali «central sources appeared to recede towards the rear of the stage and that the width of the reproduced sound stage appeared to be increased». D'altronde il loro lavoro era inteso per l'ampio spazio della sala cinematografica, dove in effetti il canale centrale è sempre stato usato perché «has the effect of stabilising the important central image for off-center listeners», mentre in ambito domestico la norma è sempre stata costituita da due canali (Rumsey 2001: par. 1.4.3). Il primo disco registrato in stereofonia pare essere quello contenente il *Poema del fuoco* di Scriabin diretto da Stokowski (1932) (cfr. Schoenherr 2001).

Nel frattempo in Inghilterra Alan Blumlein, personaggio accreditato come inventore della stereofonia grazie al suo brevetto del 1931, si dedicava a esperimenti di riproduzione sonora rifiutando il modello binaurale e preferendo l'uso di due microfoni per registrare con ciascuno dei quali un segnale separato da diffondere con due altoparlanti, sfruttando l'effetto psicoacustico poi chiamato *phantom image* ossia l'impressione della presenza di una fonte in un punto qualsiasi nello spazio intermedio tra due altoparlanti (cfr. Rumsey 2001: par. 1.4.4, 3.1.1; Théberge, Devine e Everett 2015a: 17). Questo è un principio cardine della stereofonia, da distinguere rispetto al modello binaurale (approfondirò la questione nel cap. 4). In conseguenza al lavoro di Blumlein ci furono esperimenti di stereofonia da parte di EMI nel 1933, senza seguito; poi, «nel 1956, la EMI e la RCA Victor proposero senza successo le loro registrazioni “stereosoniche” su nastro magnetico».

<sup>2</sup> Un'immagine del prospetto del dispositivo è reperibile in du Moncel (1887), qui tratto da Lange (2002).

Finalmente, nel 1958, lo stereo si impose grazie a dischi distribuiti da Audio Fidelity negli USA e da Pye e Decca in Inghilterra; «la stereofonia inaugurava una nuova maniera analitica di ascoltare la musica, grazie alla possibilità di distinguere i profili melodici e la tessitura armonica, non escluso il caso di uno strumento solista come il pianoforte» (Hains 2001: 803)<sup>3</sup>.

### 1.1.3 – La diffusione della stereofonia

La stereofonia domestica si diffuse quindi negli anni '60, periodo in cui la stereofonia divenne lo standard dell'industria musicale, grazie agli LP e poi alla radio FM stereo (cfr. Rumsey 2001: par. 1.4.5) e alle musicassette stereo (introdotte da Philips nel 1966 [Philips 2003]). Inizialmente, oltre a dischi registrati in stereofonia, «quite a lot of recordings were issued as “monophonic recording artificially reprocessed to give stereophonic effect on stereophonic equipment”». I primi album pop stereo enfatizzavano la direzionalità, tanto che si parlava di «pingpong stereo» (Rumsey 2001: par. 1.4.5). Come osserva Hains

la musica leggera [*ossia la popular music, n.d.a.*] [...] ha subito sfruttato le potenzialità espressive della ripartizione spaziale del suono creando ambienti particolari e a volte irreali [...]. Nella musica classica [*ossia la musica colta, n.d.a.*] la stereofonia è generalmente utilizzata per riprodurre la disposizione ideale dei musicisti in orchestra [...] ma essa può anche riprodurre prospettive virtuali più elaborate, come la messa in scena di un'opera con mezzi puramente sonori (2001: 803).

Similmente, Rumsey sostiene che taluni distinguono tra l'approccio “you are there” e quello “they are here”: nel primo caso l'enfasi è sulla fedeltà spaziale, il che ci dà l'impressione di essere nello spazio in cui i suoni hanno preso vita originariamente; nel secondo l'impressione dominante è quella di un ascolto diverso da quello del mondo reale, come se i musicisti si siano posizionati appena davanti a noi in modo da favorire la spazialità (2001: par. 1.3.1) (tornerò sulla questione nei capitoli seguenti). Un esempio celebre di spazializzazione pop è quello di alcune canzoni dei Beatles (benché originariamente missate in mono). Per esempio in *Please, please me* (1963) le voci, la chitarra acustica e l'armonica a bocca sono a destra mentre la chitarra elettrica, il basso e la batteria sono a sinistra; in *A Day in the Life* (1967) le voci e gli strumenti si muovono tra i due canali. Nei dischi di musica colta vige il principio di fedeltà all'ascolto in sala da concerto tipico di questa musica (diversamente da quella pop), che inoltre prima della seconda metà del Novecento

<sup>3</sup> Per un documentato resoconto del periodo riassunto in questo paragrafo, cfr. Beck (2003: 60-67).

tranne rari casi non prevede una netta separazione acustica. Ci sono comunque esempi di forte spazializzazione, riscontrabili per esempio in certi dischi Mercury Living Presence (dal 1955) (cfr. Bruil 2001), o Decca con la tecnica Phase 4 Stereo (dal 1961). Tra gli altri si considerino per la prima etichetta *La sagra della primavera* di Stravinskij diretta da Antal Doráti (1954) e il celebre disco con l'ouverture *1812* di Čaikovskij diretta sempre da Doráti (1958), mentre per la seconda etichetta il disco col *Bolero* di Ravel e le *Danze polovesiane* di Borodin diretti da Stanley Black (1964) e quello della *Sinfonia n. 9* di Beethoven diretta da Stokowski (1970): una separazione così netta è difficilmente percepibile in sala da concerto. Lo stesso vale per dischi della *Vittoria di Wellington* di Beethoven, data la sua qualità da messa in scena prevista dall'autore: fin dall'incisione diretta da Doráti nel disco Mercury Living Presence (1955) si suole porre rispettivamente a sinistra e a destra le due frasi di tamburo e le due marce che aprono l'opera a rappresentare i due eserciti che si preparano alla battaglia. Come accennato, nella seconda metà del Novecento i compositori ricorsero più frequentemente alla spazializzazione, sia nella disposizione degli strumenti (cfr. Lanza 1991: 138-139) sia con il ricorso ad apparecchi elettroacustici. Ives ha prestato particolare attenzione al dato spaziale della musica, separando nello spazio diversi ensemble, come in *The Unanswered Question* (1908) (cfr. Solomon 2007: 42-45). Un approccio simile è quello di Stockhausen, interessato anche alla resa del movimento nello spazio, come dimostra *Gruppen* per tre orchestre (1955-1957) (cfr. Solomon 2007 49-53). Nel *Kanon* (1962) di Krzysztof Penderecki l'orchestra d'archi viene registrata su due nastri che vengono poi diffusi da due altoparlanti posti alle spalle del pubblico (cfr. Avis 2001: 4). Questo espediente è reso sul disco separando nettamente i due canali (in questo caso un formato quadrifonico, su cui cfr. il cap 1.1.4, permetterebbe di restituire con maggiore fedeltà l'intenzione del compositore). Molte fonti indicano come primo concerto con suono surround quello eseguito dai Pink Floyd durante i Games For May presso il Queen Elizabeth Hall a Londra (1967) (cfr. Théberge, Devine e Everett 2015b; Calore 2009); tuttavia forse esiste almeno un precedente, poiché il 15/9/1964 al Teatro La Fenice di Venezia era stato eseguito *La fabbrica illuminata* di Luigi Nono, per soprano e nastro magnetico a quattro piste, per il quale il compositore ha indicato una diffusione in quadrifonia (Archivio Luigi Nono s.d.; Vidolin 2009: 48).

### **1.1.4 – La quadrifonia**

La quadrifonia è un tipo di stereofonia in cui gli altoparlanti sono quattro, disposti a costituire gli angoli di un quadrato circondando chi ascolta. Sistemi quadrifonici vennero realizzati a partire dal 1969 fino agli anni '70. Alcuni di essi avevano quattro canali discreti, mentre altri ricorrevano a

sistemi matriciali denominati 4-2-4: le informazioni di quattro canali venivano inserite all'interno dei soli due canali dei normali LP, cosicché un decodificatore potesse estrarre le informazioni supplementari (cfr. Curtis s.d.) (questo principio diverrà comune al cinema grazie al Dolby Stereo, stavolta però con tre canali frontali e uno surround). Gli altoparlanti frontali dovevano essere angolati a 90° piuttosto che a 30° come il normale stereo, compromettendo la compatibilità e l'immagine fantasma; inoltre secondo alcuni la riproduzione di dischi quadrifonici con sistemi a due canali, teoricamente compatibile, presentava effetti collaterali sgradevoli (cfr. Rumsey 2001: par. 1.4.9). La quadrifonia «fu accolta con entusiasmo da intenditori e melomani; ma per imporsi doveva ancora sedurre la grande platea dei consumatori», per la quale però era sufficiente la stereofonia classica (Hains 2001: 811). È esistito un sistema quadrifonico anche per la radio FM, il Quadraplex inventato da Luois Dorren (Dorren 1973).

### **1.1.5 – *Ambisonics***

Di rilevanza tecnica è anche l'*ambisonic sound*, «intended as a comprehensive approach to directional sound reproduction, involving any number of reproduction channels, based partly on an extension of the Blumlein principle to a larger number of channels» (Rumsey 2001: par. 1.4.10). Esso tuttavia è di rara applicazione. Alcuni esempi di incisioni *ambisonic* sono reperibili sul web (cfr. p.es. Ambisonic 2012).

### **1.1.6 – Il sonoro multicanale**

Il sonoro multicanale è diventato la norma nell'industria cinematografica a partire dalla fine degli anni '70 grazie al Dolby Stereo. L'attuale standard 5.1 si è imposto invece nei primi anni '90 grazie ai sistemi digitali. Nel corso di quel decennio i supporti domestici audiovisivi in grado di contenere audio in 5.1 sono stati il Laserdisc e il DVD. Il primo supporto audio digitale multicanale domestico (che esiste anche in versione stereo, più rara) è invece il Super Audio Compact Disc (SA-CD o SACD), introdotto da Sony e Philips nel 1999 e compatibile, nella sua versione ibrida, con i lettori CD: esso può contenere fino a 6 canali senza configurazioni predefinite. Nel 2000 fu introdotto il DVD-Audio (DVD-A), in grado di contenere audio in 5.1 (cfr. Massey 2004: par. 5.1.2, 5.1.3; Geutskens 2011). Il segnale in questi due supporti ha una definizione superiore rispetto a quella del CD, quindi possono essere a esso preferiti anche nella stereofonia classica. Tuttavia, come già accaduto con la quadrifonia, questi supporti sono rimasti molto minoritari; ad ogni modo il commercio di SACD di musica colta è tuttora vivo in una nicchia di audiofili, come dimostra il

catalogo della prestigiosa etichetta Channel Classics. Vigendo il principio “you are there”, la spazializzazione è tendenzialmente fedele alla sala da concerto, quindi i canali surround si limitano a contenere la riverberazione dei suoni provenienti dal versante frontale. Per gli audiofili esiste anche il Blu-ray con solo audio ad altissima definizione (cfr. Guttenberg 2016).

Le ragioni per cui la musica multicanale ha un mercato molto ristretto possono essere svariate. Kerins (2011: 3799-3834) ipotizza tre motivi convincenti: il primo è che i supporti multicanale rispetto al CD offrivano poche funzioni in più e un miglioramento qualitativo difficilmente percepibile dall’utente medio (diversamente da quanto accadde col DVD rispetto al VHS); il secondo è che per godere dell’audio 5.1 si è costretti a stare fermi nel *sweet spot*, cosa che l’utente medio non è disposto a fare essendo abituato ad ascoltare musica in automobile o con auricolari in vari luoghi (da fermo o muovendosi); il terzo consiste nell’assenza di stimolo visivo che valorizzerebbe questo modo di fruizione (cfr. anche Rumsey 2001: par. 7.4; Holman 2008: 14-16).

## 1.2 – Cinema

### 1.2.1 – L’epoca mono

Nei primi anni del sonoro si riscontra un dibattito tra i fonici circa quale fosse il modo più corretto di riprendere il suono. Si riscontrano due posizioni: quella di chi riteneva che si dovesse rispettare l’evento profilmico e la coerenza prospettica con le immagini, e quella di chi considerava più importante porre il dialogo – principale veicolo semantico – in primo piano, garantendone l’intelligibilità a prescindere dalla prospettiva visiva e riducendo la firma spaziale. Come vedremo, Lastra indica queste due posizioni rispettivamente come “modello fedeltà” e “modello telefono” (2000: 181-215). Il primo fu dominante nei primi anni, per essere poi presto rimpiazzato dall’altro. Il modello fedeltà è riscontrabile in casi sporadici in film che si distanziano dalle norme tecnico-stilistiche del cinema commerciale e che ricorrono al sonoro mono in presa diretta coerente con la prospettiva visiva e quindi enfatizzando la firma spaziale: si pensi a certi documentari “diretti”, o a certi film diretti da John Cassavetes, Jean-Luc Godard, Jean-Marie Straub e Danièle Huillet.

Nel periodo della conversione al sonoro si usavano due altoparlanti, uno sul palcoscenico e l’altro in corrispondenza della buca d’orchestra, ormai inutilizzata, ma presto si stabilì la regola di ricorrere a un altoparlante dietro lo schermo (cfr. Altman 1999: 41). Ovviamente non si trattava di stereofonia ma di una duplicazione del segnale mono, dettata da un principio di continuità con la prassi precedente di diffusione sonora dal vivo.

### 1.2.2 – L'avvento della stereofonia

Beck riferisce che la prima proiezione cinematografica stereofonica di cui si ha notizia avvenne presso i Bell Laboratories nell'autunno del 1937, supervisionata da Joseph P. Maxfield. Si ricorse a tre canali, per evitare l'effetto di buco al centro rilevato con due soli altoparlanti nei già citati studi precedenti. Il film «contained three scenes that essentially formulated three basic methodologies for considering spatialization in multi-channel systems». I modelli che scaturiscono sono quindi: primo, quello “binaurale” (musicale), «where the fidelity was related to the fixed position of the auditor and a stationary sound source»; secondo, quello stereofonico (spettacolare), «where the intent was to imitate motion even if it exaggerated the scale of the sound event»; terzo, quello “costruito”, più complesso, «where stationary and moving objects were depicted in spatial accord with the image» (Beck 2003: 68).

Il primo sistema commerciale cinematografico ad anticipare la stereofonia fu il Vitasound della Warner, usato per *I pascoli dell'odio* (*Santa Fe Trail*, USA 1940) e un numero imprecisato di altri film presumibilmente fino al 1941. Esso infatti ricorreva a più di un altoparlante, ma solo per incrementare l'intensità dei suoni e l'avvolgimento, e non per ottenere la direzionalità: la colonna sonora monofonica in alcuni momenti «of heightened action or effects» (Beck 2003: 69) (costituiti di musica ed effetti, senza dialoghi) era automaticamente inviata anche ad altoparlanti supplementari «located at or near the sides of the proscenium arch in a line with the center horns» (Levinson e Goldsmith 1941: 151; cfr. McGee 2001: 262; Sherlock 2004: 45-46; Corelli, Felici e Martinelli 2006: 68).

Il primo film commerciale a ricorrere alla stereofonia fu *Fantasia* (*Id.*, USA 1940), per il quale la Walt Disney Pictures e Leopold Stokowski avevano concepito il Fantasound (Garity e Hawkins 1942), prodotto dalla RCA, un sistema all'avanguardia, «radically unconventional» (Stull 1941: 58) e mai più usato. Già nel 1937 per il film Universal *Cento uomini e una ragazza* (*One Hundred Men and a Girl*, USA) durante l'incisione della musica era stata usata una pratica innovativa, ossia la registrazione su più tracce, per cui fu coinvolto Stokowski che diresse la musica per il film stesso. Le tracce erano sei, ciascuna per un insieme di strumenti dell'orchestra (violini, legni, ottoni, violoncelli e bassi, arpa, percussioni) (Anonimo 1938: 225): Stokowski esaltò la possibilità di missare separatamente strumenti diversi e di migliorare il realismo nella resa del timbro (Anonimo 1937: 8). Lo stesso Stokowski diresse la musica di *Fantasia*, per la registrazione della quale furono utilizzate otto tracce. La colonna sonora Fantasound era separata dal positivo con le immagini – che includeva comunque la normale colonna ottica mono, per sicurezza – e conteneva quattro canali ottici: tre erano dedicati al sonoro vero e proprio, con il segnale per i tre altoparlanti

frontali (sinistra, centro, destra) (Stull 1941: 58-59), mentre il quarto era costituito da una «special multi-tone control track for loudspeaker switching and for securing added dynamic range with a total absence of film noise on quite passages» (Anonimo 1940: 3). Inoltre era possibile

to control the sound manually in such a way that any given sound may be cut completely from the downstage speakers and reproduced at a higher level by only the auditorium speakers at the side, rear, or top of the house, making it possible to make the sound apparently come from any desired part of the auditorium.

Questa tecnica fu impiegata per l'*Ave Maria* di Schubert alla fine del film:

the orchestra is played on the downstage speakers as is the vocal choir. When the solo voice enters, it is first played only on the rear speakers, then faded to the ceiling speakers, and finally to the downstage speakers on one side only (Stull 1941: 59, 80).

Al Broadway Theater di New York furono installati almeno 80 altoparlanti, mentre non è chiaro se per l'anteprima a Los Angeles ne siano stati installati 96 per l'effetto surround oppure di meno<sup>4</sup>. Il Fantasound è insomma il primo, eccezionale tentativo di immersione sonora, che anticipa le attuali nuove frontiere<sup>5</sup>.

All'inizio degli anni '50 le case cinematografiche accusavano una decisa riduzione degli introiti, a causa di cambiamenti nell'assetto distributivo, degli effetti del conflitto e soprattutto della nascita della televisione (cfr. Valentini 2006: 71): per questa ragione il periodo fu segnato da «the most pervasive technological innovations in Hollywood since the late 1920s» (Bordwell 1985c: 359). Si intensificarono le sperimentazioni sia sul piano dell'immagine – gli schermi panoramici – sia sul piano del suono – la colonna magnetica e la stereofonia – anche se in entrambi i sensi la ricerca era iniziata già molti anni prima<sup>6</sup>.

La registrazione magnetica permetteva «maggiore maneggevolezza delle apparecchiature», la «risorsa finora unica di poter riascoltare immediatamente la registrazione [e] la possibilità [...] di cancellare i suoni precedenti e predisporre l'apparecchio immediatamente per una nuova registrazione» (Valentini 2006: 72); garantiva inoltre «better fidelity than optical sound», permetteva la copia «with less quality loss» (Handzo 1985: 391), aveva una «dramatically expanded frequency

<sup>4</sup> Handzo (1985: 419) parla di 96 altoparlanti senza specificare il nome della sala, mentre Stull (1941: 59) riferisce che al Carthay Circle c'erano meno altoparlanti rispetto al Broadway di New York.

<sup>5</sup> Per un dettagliato resoconto dello sviluppo del Fantasound cfr. Garity e Hawkins (1941: 12-14), da cui si evince che la prima versione del sistema (*Mark I*), prevedeva cinque altoparlanti, di cui tre frontali e due agli angoli posteriori.

<sup>6</sup> «Widescreen cinema was technically feasible at least two decades before its acceptance[.] Stereophonic sound was also anticipated during the 1930s» (Bordwell 1985c: 358).

range, a significantly improved signal-to-noise ratio, and a larger dynamic volume range» (Belton 1992: 155). Il sonoro magnetico aveva esordito con successo già nel 1945 per la registrazione post-sincronizzata, e nel 1951 era usato a Hollywood per il 75% della produzione e post-produzione sonora dei film (Handzo 1985: 391); tuttavia si doveva operare poi una conversione nella colonna ottica per le copie da distribuire, riducendo la qualità sonora (cfr. Belton 1992: 155).

Una prima svolta ci fu nel 1952 quando fu presentato *This is Cinerama* (USA), film che inaugurò un nuovo sistema audiovisivo annunciato fin dal titolo, il Cinerama appunto. Lo schermo era concavo, e su di esso proiettavano contemporaneamente tre proiettori allineati (35mm): i canali sonori, registrati su una pellicola separata, erano sette, di cui cinque frontali (sinistra, centro-sinistra, centro, centro-destra, destra) e due surround. Handzo, come altri, spiega che

The five speakers behind the screen corresponded to five microphones placed in accordance with the visual image. One to three additional microphones recorded off-screen sound. [...] Five of the channels were dedicated to the speakers behind the screen; the sixth “surround” channel was switched to the left or right electronically and the rear speaker was cued manually to pick up sounds as needed from the other tracks (1985: 419).

In realtà il Cinerama è sempre ricorso a sette canali, ma con una particolarità relativa al sesto e al settimo, inviati all'area surround:

Sound was controlled by the “Cinerama Engineer”, who, based on cue sheets for the show, could throw a switch that put channel six on both walls and shifted channel seven to speakers in the rear of the theatre. Over the years it became common practice in many theatres to do away with the theatre engineer and leave the controls for channels six and seven set in one of the two positions (in Hart s.d.; cfr. Sherlock 2004: 8-9, 11; Corelli, Felici e Martinelli 2006: 69; Holman 2008: 5).

Dunque il sistema anticipa contemporaneamente sia il principio dello “split-surround”, una possibile configurazione del Dolby 70mm raramente utilizzata (dal 1978) e diventata poi lo standard col 5.1 digitale (dal 1992), sia il principio del Dolby Digital Surround EX (1999) in cui un canale è emanato dall'area centrale posteriore. Nel Cinerama «Five, six, or seven channels of sound were recorded live on location, rather than being created from sound effects libraries in a studio» (C. A. Productions 2000), come dimostrano alcune immagini della produzione di *La conquista del West* (*How the West Was Won*, USA 1962) in cui si vedono talora cinque e talaltra sette microfoni (cfr. C. A. Productions 2000; Hart 2000). Secondo una testimonianza del fonico Hazard E. Reeves

(1953), «five microphones are placed about the scene to cover all the action within the range of the camera. Other mikes may be placed off to the sides, to the rear – anywhere – to record offstage effects». Dunque la colonna sonora rispecchiava in gran parte la collocazione profilmica delle fonti rappresentate riproponendo la configurazione dei canali della ripresa. Il Cinerama (nella sua versione originale) fu riservato a poche sale specializzate a pochi film, per lo più non narrativi (dunque con una dinamica simile a quella che si verificherà poi con l'IMAX), e fu usato l'ultima volta nel 1963 (cfr. Blake 1984: 2). Ci sono stati due sistemi affini al Cinerama, cioè il Cinemiracle, che fu usato solo per il film *Windjammer: The Voyage of the Christian Radich* (USA 1958) (cfr. McGee 2001: 60), e il Kinopanorama. Quest'ultimo fu introdotto in Unione Sovietica nel 1958, e non fu quasi mai installato in altri paesi; tuttavia film realizzati con questo sistema poterono essere proiettati con qualche accorgimento con gli apparecchi Cinerama o Cinemiracle. Il Kinopanorama prevedeva nove canali magnetici: i primi cinque erano come di consueto relativi all'area frontale, mentre gli altri quattro erano sinistro laterale, destro laterale, posteriore e superiore (dal soffitto) (cfr. Sherlock 2004: 12; Dorange 2015; Corelli, Felici e Martinelli 2006: 70-71). Nella versione Circular Kinopanorama al posto del surround posteriore c'era un canale per le basse frequenze diffuso da subwoofer (cfr. Sherlock 2004: 12), anticipando una consuetudine successiva.

Nell'aprile 1953 la Warner distribuì *La maschera di cera* (*House of Wax*, USA), un film che offriva immagini in 3D e suono stereofonico, con l'auspicio di riportare il pubblico in sala. Il formato sonoro era il WarnerPhonic Sound, un sistema appositamente ideato: in un nastro separato c'erano tre canali magnetici (sinistra, centro, destra), mentre nella banda ottica sulla pellicola c'erano due canali, uno dei quali contenente alcuni effetti destinati ai vari altoparlanti surround, laddove l'altro presentava l'intero sonoro del film come traccia di emergenza. Il New York Paramount installò per l'occasione 25 altoparlanti intorno alla sala (cfr. Handzo 1985: 420; Sherlock 2004: 46). Secondo alcuni resoconti la spazializzazione frontale e gli effetti surround erano di grande impatto, ma è interessante osservare il parere di Crowther sul *New York Times* (19/4/1953) che invece ritenne il risultato distraente e quasi comico (in Furmanek e Kintz s.d.). Il sistema fu usato per pochissimi film (cfr. McGee 2001: 263; Furmanek s.d.).

La svolta decisiva ci fu grazie al CinemaScope, il sistema con sonoro magnetico a maggior diffusione prima del Panavision in 70mm: esso fu realizzato da Twentieth Century Fox e presentato per la prima volta nel settembre del 1953 con *La tunica* (*The Robe*, USA). Il CinemaScope introdusse il processo di anamorfosi (già sperimentato in passato; cfr. Belton 1988: 25-26), e inoltre prevedeva 4 canali di sonoro magnetico: tre canali erano diffusi anteriormente (sinistra, centro, destra), il quarto posteriormente per ottenere l'effetto surround (comunque poco usato)<sup>7</sup>. Bordwell

<sup>7</sup> «Studios other than Fox never bothered mixing anything into the surround channel at all» (Handzo 1985: 420).

nota che «unlike most widescreen processes, which reproduced stereophonically from a monaural studio recording, CinemaScope productions were also recorded in stereo. Three microphones, spread across a sound boom, yielded the sound to be played through the three theater speakers» (1985c: 360; cfr. Valentini 2006: 75), dunque con una prassi affine a quella che abbiamo rilevato nel Cinerama. Sembra che a esclusione della Twentieth Century Fox gli *studios* non utilizzarono la stereofonia per i dialoghi, limitandosi a collocarli in tutti i canali frontali: si pensava infatti che essa fosse efficiente solo per la musica e taluni effetti, essendo distraente e fastidiosa nel caso dei dialoghi, come tuttavia accade in diversi film dell'epoca e come continuerà ad accadere in molti film stereofonici a seguire fino all'avvento del Dolby Stereo. Inoltre «[f]or a time, stereo recording actually dictated onscreen composition», in base alla posizione dei vari microfoni (Belton 1992: 162-163). L'industria era ottimista: «Darryl Zanuck annunciò che la 20th Century Fox avrebbe prodotto tutti i futuri film con questo sistema e fu presto seguita [da] tutte le principali case» (Valentini 2006: 75), tranne la Paramount che preferì a investire su un altro sistema, più economico (VistaVision). Il CinemaScope prevedeva però costi notevoli per gli esercenti, che dovevano necessariamente rinnovare le sale: ecco perché già nel 1954 «Fox CinemaScope films could be ordered in four-track magnetic, single-track magnetic, or conventional optical versions» (Handzo 1985: 420). Inoltre dal 1957 tutti i film in CinemaScope della Twentieth Century Fox ricorsero al sistema “magoptical” sviluppato dal Motion Picture Research Council, che permetteva di aggiungere alle piste magnetiche una colonna ottica mono, al fine della retrocompatibilità (cfr. Anonimo 1957: 21; Sherlock 2004: 42). Ciò comportò anche una modifica del rapporto d'aspetto: da 2,55:1 si passò a 2,35:1, che divenne uno standard<sup>8</sup>, poiché per ospitare anche la traccia ottica occorre ridurre lo spazio dedicato alle immagini. In ogni caso, la norma rimase la colonna ottica monofonica, mentre quella magnetica fu limitata a sale più grandi e a film più costosi e spettacolari (cfr. Bordwell 1985c: 360; Valentini 2006: 76).

Il Todd-AO, inaugurato con *Oklahoma!* (*Id.*, USA 1955), richiedeva una macchina da presa e proiettori appositi (con una velocità di scorrimento di 30 fps, un unicum), stavolta utilizzando una singola pellicola a 70mm<sup>9</sup>. I canali magnetici erano sei, di cui cinque frontali come nel Cinerama e uno solo per il surround (la pellicola a 70mm può infatti ospitare solo sei canali magnetici; v. figura 1). Handzo spiega che «in Todd-AO, only music was recorded in six-channel stereo; dialogue and effects were recorded monophonically and “panned”» (1985: 419-420). Questo sistema, molto costoso e complesso, ebbe scarsa diffusione essendo destinato alle poche sale attrezzate.

<sup>8</sup> Per continuità si usa indicare così anche il rapporto d'aspetto in Panavision che da esso deriva, in verità 2,39:1 e talvolta indicato piuttosto così o nell'approssimazione 2,40:1.

<sup>9</sup> Il negativo era a 65mm; il positivo era a 70mm per poter ospitare i 6 canali di sonoro magnetico; cfr. Blake (1984: 2).

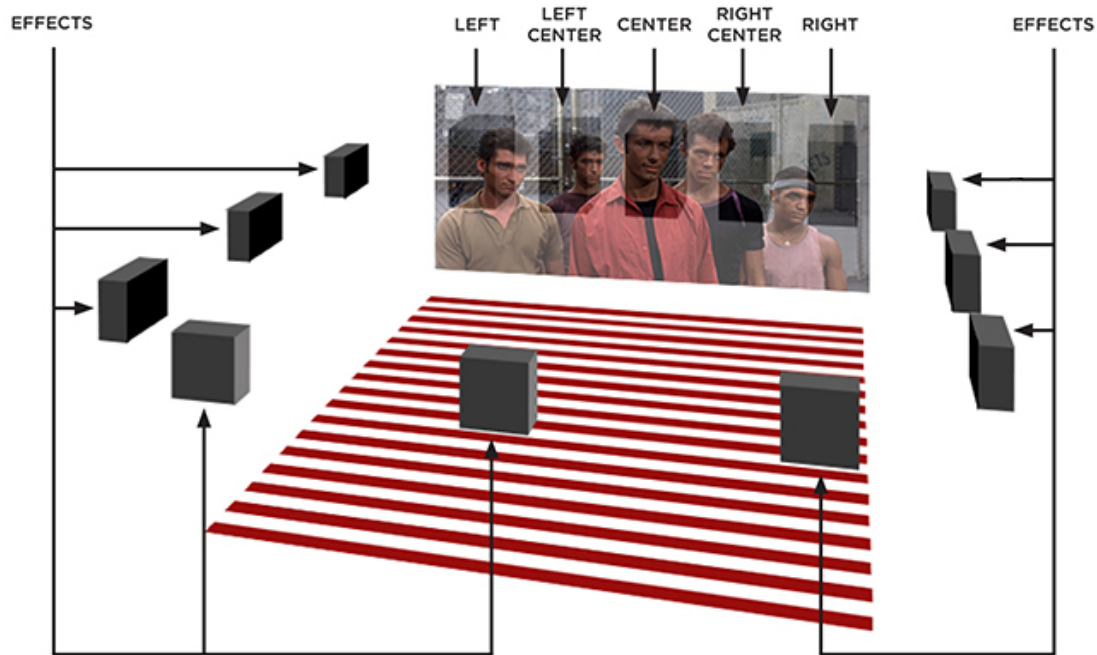


Figura 1: La configurazione dei canali nel 70mm, dal Todd-AO (1955) al 1976, dopo di che divenne rara per l'avvento del Dolby "baby boom". Il fotogramma è tratto da *West Side Story* (Id., USA 1961).

I film realizzati in 70mm in seguito fecero ricorso ad altri sistemi (p.es. Super Panavision 70, Ultra Panavision 70). Dopo il 1970 furono pochissimi i film realizzati in 70mm (URSS a parte<sup>10</sup>), mentre furono numerosi i casi di "blow-up", ossia di stampa in 70mm da negativo 35mm, il che permetteva di sfruttare la colonna sonora a sei canali magnetici. Tuttavia come spiega Blake «from 1971 to 1977 almost all stereo films were recorded four-track for release in the four-track 35mm, and/or six-track 70mm magnetic formats»: in quest'ultimo caso i canali frontali mediani – centro-sinistra e centro-destra – erano creati «by combining the center channel 50/50 with left and right tracks, respectively». Tale scelta era dovuta alla volontà di risparmio, mentre i fonici ritenevano questa prassi inferiore a quella in cui i sei canali sono tutti diversi (Blake 1984: 2).

Il Perspecta era un sistema di pseudo-stereofonia in cui la colonna sonora mono era inviata a tre amplificatori collegati a tre altoparlanti dietro lo schermo, passando attraverso il Perspecta Stereophonic Sound Integrator che recepiva dei segnali registrati nella colonna a frequenze inaudibili durante la proiezione: «the studio sound mixer had the ability, using one or several frequencies at different levels, to direct monophonic sound toward any part of the screen. He could also, using all the frequencies at maximum level, give greater sound volume to the presentation» (Verscheure 1995: 268-269<sup>11</sup>). Questo sistema fu lanciato dalla Paramount con *Bianco Natale*

<sup>10</sup> In URSS vennero prodotti molti film col sistema 70mm Sovscope 70; cfr. Silyanov (2015).

<sup>11</sup> Questo articolo contiene anche un approfondimento su *L'uomo che sapeva troppo* (*The Man Who Knew Too Much*, USA 1956), realizzato in Perspecta, come caso di studio sul restauro di una colonna sonora Perspecta mantenendo il medesimo effetto di spazializzazione (Verscheure 1995: 269-275).

(*White Christmas*, USA 1954), che è inoltre il primo film in VistaVision. Anche altri studios – incluso Toho in Giappone – ricorsero al Perspecta, che cadde in disuso nel 1962 (cfr. Kellogg 1955 431; Verscheure 1995: 269; IMDb s.d.)<sup>12</sup>.

Un altro sistema sonoro che fornisce una prova ulteriore della volontà dell'industria di offrire al pubblico un'esperienza viscerale è il Sensurround (Stumpf, Watson e Leonard 1974), sviluppato dalla MCA (Universal) e inaugurato da *Terremoto* (*Earthquake*, USA 1974), la cui pubblicità più significativa è a mio avviso quella che recita «you will feel as well as see and hear realistic effects such as might be experienced in an actual earthquake» (in Hauerslev 2011). Esso fu utilizzato in copie magnetiche a 4 canali e in copie ottiche mono, nonché in copie a 70mm ma solo fuori dagli USA. La colonna sonora era mono in entrambi i formati del 35mm, mentre nel 70mm musica ed effetti erano stereo (cfr. Beck 2003: 179); nella colonna ottica e nei canali 2 e 4 di quella magnetica in 70mm c'erano dei *control tones* inaudibili la cui presenza o assenza originava quattro combinazioni possibili relative all'attivazione di rumori bassi, all'invio di sonoro agli altoparlanti surround e all'intensità di ciascuno di questi due insiemi di suoni. Quindi era possibile attivare un generatore di rumore per produrre suoni estremamente bassi, fino a 16 Hz, riprodotti da potenti altoparlanti installati nella sala (davanti, dietro e ai lati) in un numero tra quattro e otto e avvertiti non solo come suoni ma anche come forti vibrazioni (tant'è che esistono racconti di soffitti crepati e altri effetti collaterali) (cfr. Hauerslev 2011). Lo stimolo sonoro diventava dunque stimolo tattile, anticipando la prassi del Dolby "baby boom" e quindi del futuro standard 5.1 che però non sono concepiti per l'"effetto terremoto". Inoltre era possibile inviare il sonoro mono anche agli altoparlanti supplementari. *Terremoto* fu l'unico film in cui il Sensurround venne utilizzato in questa versione: presto MCA realizzò una versione diversa del sistema, chiamata Mod-II e utilizzata per *La battaglia di Midway* (*Midway*, USA 1976), e infine un'ultima versione, chiamata Mod-III e utilizzata per *Rollercoaster – Il grande brivido* (*Rollercoaster*, USA 1977), *Battaglie nella galassia* (*Battlestar Galactica*, USA 1978) e, probabilmente, per copie giapponesi di *Galactica: l'attacco dei Cylon* (*Mission Galactica: The Cylon Attack*, USA 1979). In queste due versioni i suoni bassi erano inclusi nella colonna ottica mono grazie al sistema di riduzione del rumore dbx Type-II, e il resto del sonoro poteva essere inviato separatamente agli altoparlanti supplementari frontali o a quelli posteriori (cfr. Beck 2003: 172-187; Anonimo 1974; Sherlock 2004: 44-45)<sup>13</sup>. Il Sensurround ebbe vita breve forse soprattutto a causa dell'avvento del Dolby "baby boom", ma ebbe anche un certo successo (con un revival una tantum nel 2010; cfr. Fuchs 2014) – quantomeno sul piano concettuale, come osserva acutamente Beck (2003: 187).

<sup>12</sup> Secondo Handzo (1985: 421), «[Perspecta was] used mainly by MGM and Loew's Theaters».

<sup>13</sup> *Zoot Suit* (USA 1981) in 70mm fu presentato con l'equivoca dicitura "Sensurround Plus": in realtà presentava solo il dbx e null'altro del Sensurround (cfr. Hauerslev 2011).

Negli anni '50 iniziarono anche gli esperimenti con la televisione stereo, grazie al *simulcast* con la radio. Inizialmente la stereofonia era ottenibile trasmettendo un canale in televisione insieme alle immagini e l'altro canale via radio; poi, dagli anni '60 agli anni '80, si diffuse direttamente il segnale stereo via radio (cfr. Sedman 2015: 186-192).

### 1.2.3 – L'epoca del Dolby Stereo

Handzo rileva che «the third major sound revolution after the “talkies” (1927-1930) and magnetic stereo (1953-1955) began in 1975 with the advent of optical stereophonic sound-on-film pioneered by Dolby Laboratories» (1985: 422), siti a Londra e fondati nel 1965 dall'americano Ray Dolby (cfr. Valentini 2006: 105). Nel 1966 fu introdotto il sistema di riduzione del rumore Dolby A (Dolby A-type Noise Reduction), destinato ai riproduttori di nastri magnetici professionali, e l'anno seguente nacque il Dolby B destinato invece agli apparecchi domestici per audiocassette (cfr. Handzo 1985: 422; Valentini 2006: 105). Il successo di questi sistemi ebbe influenza sull'industria cinematografica: *Arancia meccanica* (*A Clockwork Orange*, Regno Unito/USA 1971) fu il primo film a usare il Dolby A in tutta la post-produzione (cfr. Blake 1984: 2). Durante gli anni '70 nacque il Dolby Stereo Variable Area (SVA), un sistema ottico ad area variabile che, nei due canali inclusi nello spazio della pellicola prima riservato alla traccia mono, conteneva informazioni per quattro canali nella configurazione LCRS (*left, center, right, surround*). I due canali, cui è applicato il Dolby A, sono detti Lt-Rt (*left total – right total*), proprio a denotare il fatto che, grazie all'applicazione di una matrice, riproducono informazioni supplementari rispetto ai due canali della stereofonia classica (L-R). Tale matrice agiva in termini di fase e livello: un segnale posto solo in Lt veniva riprodotto in L e, a -15 dB, in C; per Rt vale lo stesso, viceversa; un segnale posto in entrambi i canali a -3 dB veniva riprodotto in C se in fase e nel surround se fuori fase e inoltre, a -15 dB, anche in L e R. Si tratta dunque di un sistema 4:2:4 (o 4-2-4), espressione che indica il fatto che si missano 4 canali discreti dentro 2 matriciali per riottenere i 4 canali in riproduzione, ed è chiaro che la separazione non può essere netta come in un sistema discreto (cfr. Blake 1984: 3-4; Holman 2008: 11). I vantaggi del Dolby SVA rispetto ai sistemi magnetici precedenti risiedeva nel costo inferiore e nella facilità di conversione delle sale: «Dolby wisely designed its system to be compatible with the bulk of existing components» (Handzo 1985: 422). Il primo film a usare il Dolby SVA fu *Lisztomania* (*Id.*, Regno Unito 1975), al quale però mancava il canale surround, incluso invece in *È nata una stella* (*A Star is Born*, USA 1976) (cfr. Blake 1984: 2-5; Dolby Laboratories 1988: 3; Valentini 2006: 107; Kerins 2011: Appendix A). Il Dolby A fu applicato anche ai quattro canali magnetici di alcuni film distribuiti in 35mm – *Il piccolo principe* (*The Little*

*Prince*, Regno Unito/USA 1974) e *Nashville* (*Id.*, USA 1975) – e soprattutto ai sei canali magnetici di film distribuiti in 70mm, a partire da alcune copie di *La fuga di Logan* (*Logan's Run*, USA 1976) (cfr. Blake 1984: 27), in cui si può notare la persistenza del dialogo direzionale che era una caratteristica ricorrente nei film con sonoro magnetico, fin da quelli in CinemaScope e passando per quelli realizzati in altri formati negli anni seguenti, per esempio *Quelli della San Pablo* (*The Sand Pebbles*, USA 1966) e *2001: Odissea nello spazio* (*2001: A Space Odyssey*, USA/Regno Unito 1968)<sup>14</sup>. Il film che segnò il trionfo del Dolby Stereo imponendolo sul mercato, in entrambe le versioni, fu tuttavia *Guerre stellari* (*Star Wars*, USA 1977)<sup>15</sup>. Le copie in 70mm avevano un mercato più ristretto rispetto a quelle in 35mm, di conseguenza la versione più diffusa del Dolby Stereo nel mondo è stata quella ottica, malgrado fosse qualitativamente inferiore a quella magnetica. Inizialmente il Dolby SVA ricorreva alla matrice Sansui QS, ma nel 1979, a partire dai film *Hair* (*Id.*, USA) e *Uragano* (*Hurricane*, USA), essa fu sostituita da una nuova matrice che permetteva una maggiore separazione dei canali (cfr. Blake 1984: 3). Inoltre nel 1980 fu introdotta la Optical Bass Extension, per incrementare la resa delle frequenze sotto i 100 Hz anche nel formato ottico (cfr. Blake 1984: 33; Dolby Laboratories 1988: 4; v. figura 2).

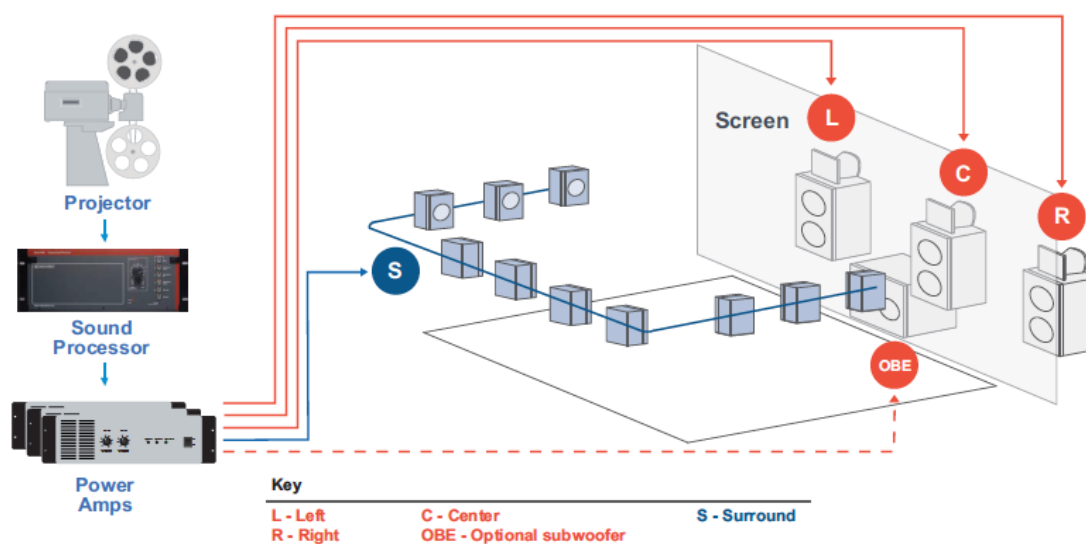


Figura 2: il sistema Dolby SVA (Dolby Laboratories 2014: 2).

La configurazione dei canali del Dolby Stereo magnetico era variabile: nel primo film in cui fu utilizzato, *La fuga di Logan* – e in pochi altri poi, tra cui *I cancelli del cielo* (*Heaven's Gate*, USA

<sup>14</sup> Faccio riferimento a edizioni home video che, presumibilmente, presentano il mix originale dei film, al netto di alcune differenze dovute al supporto (cfr. par. 1.2.6).

<sup>15</sup> *Guerre stellari* è significativamente indicato da Jullier (2006: 148-150) come il primo film ascrivibile al postmodernismo, in cui l'enfasi sull'effetto viscerale del sonoro è connotato.

1980), e *Tron* (*Id.*, USA 1982) – essa seguiva il modello tradizionale dei film in 70mm, ossia quello del Todd-AO con cinque canali frontali e uno surround, detto “discrete six-track” (cfr. Blake 1984: 32). Tuttavia per *È nata una stella* ci si rese conto che sarebbe stato più efficace inserire nei due canali frontali mediani lo stesso segnale dei due canali laterali (come d’altronde usava fare già dalla fine degli anni ’60), a un volume più basso, un formato detto “six-track spread” (cfr. Blake 1984: 2, 27) (non ho trovato fonti che chiariscano quali altri film sono stati missati così). Per *Guerre stellari* George Lucas e il fonico Ben Burtt ritennero invece che i due canali frontali mediani – e quindi i due altoparlanti a essi relativi già installati nelle sale adibite al 70mm – si potessero sfruttare piuttosto per incrementare le basse frequenze, normalmente carenti. Nacque così la configurazione di Dolby Stereo magnetico più diffusa, cosiddetta “baby boom”: i sei canali erano sinistra, centro, destra, surround, più due (tra loro identici) per le basse frequenze (sotto i 200 Hz). *Incontri ravvicinati del terzo tipo* (*Close Encounters of the Third Kind*, USA 1977) fu il secondo film a ricorrere a questa configurazione, ma stavolta la Dolby decise di far sostituire i vecchi altoparlanti con i subwoofer, adibiti esclusivamente alle basse frequenze; inoltre, diversamente da *Guerre stellari*, da questo momento in poi solo alcuni suoni a bassa frequenza furono inseriti nei canali “boom”, a seconda dell’intenzione espressiva, e non tutte le basse frequenze dell’intera colonna sonora. Si stabilì così lo standard del “baby boom”, il formato sonoro di quasi tutti i film distribuiti in 70mm in seguito (cfr. Blake 1984: 27-28; Handzo 1985: 423; Coate e Hauerslev 2015; v. figura 3).

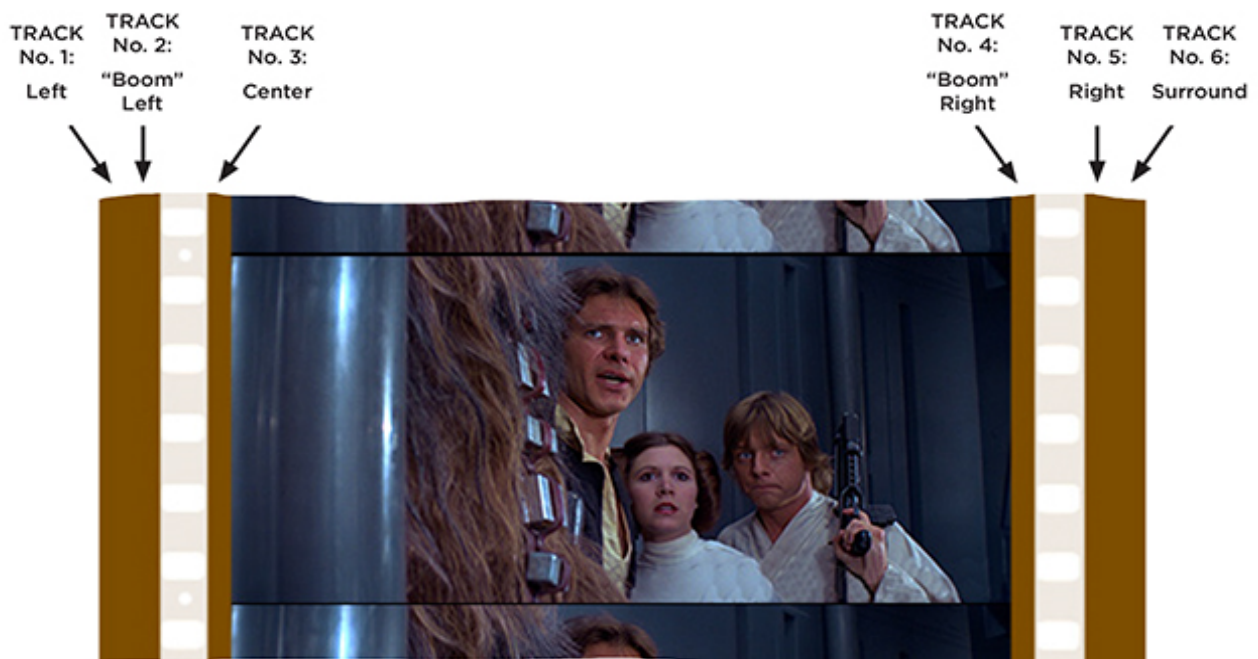


Figura 3: I sei canali magnetici del 70mm, qui nella configurazione Dolby “baby boom” (il fotogramma è tratto da *Guerre stellari* [1977], il primo film a ricorrere a questa configurazione).

Un'altra configurazione possibile è la cosiddetta *split-surround*, concepita da Walter Murch per *Apocalypse Now* (*Id.*, USA 1979) su richiesta di Francis Ford Coppola che voleva un sonoro quadrifonico (cfr. Beck 2003: 52; Beck e Ament 2015: 125). Il film fu il primo a presentare in tutte le copie in 70mm questa configurazione, che però fu sperimentata la prima volta già in alcune copie di *Superman* (*Id.*, USA 1978). L'area surround era divisa in sinistra e destra, e un canale era riservato alle basse frequenze: si tratta insomma di un 5.1 ante litteram (cfr. Blake 1984: 29-30; Valentini 2006: 108). Altri film in Dolby Stereo *split-surround* sono *La febbre del successo – Jazz Singer* (*The Jazz Singer*, USA 1980), *Pink Floyd The Wall* (*Id.*, Regno Unito 1982), e *Indiana Jones e l'ultima crociata* (*Indiana Jones and the Last Crusade*, USA 1989) (Blake 1984: 9.10; Beck e Ament 2015: 125-127; Dienstfrey 2016: 170-174).

È interessante notare, come accennato, che la maggior parte dei film distribuiti in 70mm dagli anni '70 in poi era realizzata in 35mm. Si ricorreva alla stampa in 70mm dal 35mm, ossia al cosiddetto "blow-up", proprio per poter sfruttare il sonoro magnetico, dato che il vantaggio in termini visivi era irrilevante. Girare in 65mm avrebbe permesso di utilizzare al meglio il positivo in 70mm, ma con costi decisamente superiori. La maggiorazione del costo della stampa in 70mm, invece, era evidentemente ritenuta un valido investimento, enfatizzando il miglioramento della resa acustica (cfr. Blake 1984: 34).

I formati magnetici migliorano la separazione dei suoni, e quindi la spazializzazione, poiché tutti i canali riprodotti sono discreti. Il Dolby SVA ha permesso di diffondere la stereofonia al cinema come mai prima di allora, ma resta soggetto a evidenti limiti poiché le informazioni per i quattro canali devono essere ricavate dai soli due materialmente presenti. Si consideri quanto osservato da Beck:

This regularly required the dialogue to be mixed into the central channel both to ensure comprehension and to avoid phasing. Effects could be positioned anywhere in the left-to-right space of the screen, but moving effects would be carefully monitored to ensure that their acoustic motion matched the motion on-screen. Music, however, provided few problems because it was rarely anchored to an on-screen image. But when it was, care was given to make sure that sounds did not drift in the stereoscope. While music and effects were occasionally deployed to the surround speakers, dialogue was strictly avoided, As a byproduct of the matrixed surround channels being 180° out of phase between the left and right channels, any downmixing to mono would result in the entire loss of the surround information, as each channel would cancel out the other. To ensure backward compatibility with monophonic optical projectors, any information exclusive to the surround channels would have to be expendable.

Therefore only “surround effects” were sent to the rear speakers so that no narrative information would be lost if the film was not played in its desired format (2003: 46-7).

Inoltre era impossibile applicare la divergenza nel Dolby Stereo ottico, una tecnica che nei formati magnetici permetteva di rendere più accurata la localizzazione in vari punti della sala. La sua assenza comportava che «traveling sounds, or sounds in the side channels, were much more noticeable than in a magnetic print with divergence. Therefore effects tended to either remain centered on the screen or to drift off into a side channel and the surrounds» (2003: 49).

Un'altra conseguenza problematica è quella del “magic surround”, ossia dei suoni che finiscono accidentalmente nel surround perché in controfase senza il proposito dei fonici. Per qualunque sistema contenente surround, ma soprattutto per il Dolby SVA, vale inoltre il rischio che la sala – per trascuratezza dell' esercente – non riproduca adeguatamente (o affatto) il canale surround. Ecco perché i fonici sono stati quasi sempre restii a sfruttarlo appieno (un'eccezione memorabile è ovviamente la sequenza iniziale di *Apocalypse Now*), tanto che Burt, durante il missaggio di *Il ritorno dello Jedi* (*The Return of the Jedi*, USA 1983), palesò forse più di chiunque altro (e forse con maggiore influenza, dato il prestigio suo e della saga di *Star Wars*) la regola non scritta di usare il surround solo per effetti non essenziali, ovvero come «an *enhancement* to what is already there [nei canali frontali]» (in Blake 1984: 45). Inoltre, forse anche per l'influenza del limite tecnico della matrice del Dolby SVA, in questo periodo si affermò ancor più fortemente la prassi di centrare il dialogo, evitando di renderlo direzionale. Insomma forse per una combinazione tra aspetti tecnologici e cognitivi il Dolby Stereo mantenne saldo il «‘one-wall’ narrative principle» (Sergi 2004: 20). Questo sistema è stato insomma storicamente di indiscutibile importanza, ma come raccomanda Dienstfrey (2016) bisogna curarsi di non mitizzarlo come pure è stato fatto.

Sotto l'influenza del buon esito del del Sensurround prima e del trionfo del Dolby Stereo poi, negli anni '70 si tentarono altre sperimentazioni sonore, destinate anch'esse a vita breve come il Sensurround. La Twentieth Century Fox realizzò il Sound 360, inaugurato da *L'ultima odissea* (*Damnation Alley*, USA 1977), la cui colonna sonora era quella magnetica del 35mm a quattro canali ma la cui distribuzione di altoparlanti prevedeva un quadrilatero: i canali sinistra e destra erano diffusi da altoparlanti posti sulle pareti rispettive, mentre il surround era diffuso da un altoparlante sulla parete posteriore. Secondo alcune fonti l'unico altro film a usare questo sistema fu *La maledizione di Damien* (*Damien: Omen II*, USA 1978); secondo altre invece il sistema non fu mai più usato (cfr. Village Cinema City s.d.; Disclord 2011). La Warner produsse il MegaSound, lanciato da *Stati di allucinazione* (*Altered States*, USA 1980), film la cui programmazione al Westwood Village Theater di Los Angeles divenne leggendaria tra i cinefili audiofili. Il sistema

ricorreva ai sei canali magnetici del 70mm con due peculiarità: si estesero le basse frequenze al di sotto della soglia dei 40 Hz e se ne incrementò l'intensità (ricordando il Sensurround); nei due canali per le basse frequenze si codificarono suoni da inviare rispettivamente al surround sinistro e a quello destro, in aggiunta al segnale surround mono presente sul canale apposito, con un principio simile a quello del Dolby Stereo 70mm *split-surround* (cfr. Blake 1984: 32; Sherlock 2004: 43). Il MegaSound fu utilizzato solo per altri tre film, tutti usciti negli USA nel 1981: *Atmosfera zero (Outland)*, *Superman II (Id.)* e *Wolfen – La belva immortale (Wolfen)*.

Dato il successo del Dolby SVA altre aziende svilupparono sistemi ottici. La Paramount realizzò il Vistasonic Sound, che permetteva di ospitare quattro canali discreti nella colonna ottica del 35mm, riducendo le dimensioni di ciascun canale. Esso fu usato in *Popeye – Braccio di ferro (Popeye, USA 1980)* e poi in *Il drago del lago di fuoco (Dragonslayer, USA 1981)*, ma non ebbe seguito (cfr. Blake 1984: 3; Sherlock 2004: 45). L'Ultra-Stereo seguiva lo stesso principio del Dolby SVA ed era con esso compatibile; fu introdotto nel 1984 ma rimase minoritario rispetto all'affermato concorrente (cfr. Corelli, Felici e Martinelli 2006: 73), malgrado un premio tecnico assegnato all'ideatore Jack Cashin da parte dell'AMPAS (cfr. Francis 1984: 110).

L'ultima tappa dell'evoluzione del Dolby Stereo analogico consiste in un nuovo e migliore sistema di riduzione del rumore chiamato Dolby SR (Spectral Recording) in sostituzione del Dolby A, sia per il sonoro ottico più diffuso sia per quello magnetico. Il primo film con cui fu sperimentato il Dolby SR fu *Rotta verso la terra (Star Trek IV – The Voyage Home, USA 1986)*, ma solo per alcune copie in 70mm. Il primo film distribuito in Dolby SR, in copie sia magnetiche sia ottiche, fu *Salto nel buio (Innerspace, USA 1987)*, seguito pochi giorni dopo da *RoboCop (Id., USA 1987)* (cfr. Coate e Hauerslev 2015)<sup>16</sup>.

È doveroso dedicare spazio a un caso unico quale quello di *Tommy (Id., Regno Unito 1975)*, perché indicativo di uno scambio tra l'industria cinematografica e quella musicale. Il film è basato sull'album dell'omonima opera rock di The Who. In quel periodo c'erano diversi sistemi hi-fi quadrifonici, che ricorrendo a matrici estraevano quattro canali dai due concretamente presenti sugli LP così concepiti. Ispirandosi a questi, l'ingegnere John Mosely concepì il Quintaphonic Sound. Il sonoro magnetico a quattro canali venne sfruttato per ottenere quattro canali matriciali più un quinto discreto. I canali sinistro e destro contenevano pertanto informazioni anche per due singoli altoparlanti posti nell'area surround – non insieme di altoparlanti, come da prassi cinematografica, dunque esattamente come gli LP quadrifonici; il canale centrale era usato come sempre per il relativo altoparlante frontale; il quarto canale, per il surround, rimase inutilizzato. Le poche sale che avevano le apparecchiature per riprodurre l'ormai desueto sonoro magnetico a quattro canali

<sup>16</sup> Per un resoconto sul “fenomeno Dolby” cfr. Sergi (2004: 11-35).

dovettero essere preparate appositamente per una tale configurazione. Questo sistema non fu mai più usato (cfr. Beck 2003: 128-130; Bray s.d.; Liebman 2010)<sup>17</sup>.

Occorre infine menzionare il THX: non si tratta di un sistema sonoro, bensì di un «protocollo a cui il suono filmico e la sala devono rispondere[,] un insieme di norme di qualità che si occupano di definire tutti i parametri tecnici dei diversi impianti e [...] di certificare la qualità delle sale in cui è operativo e il livello qualitativo della proiezione» (Valentini 2006: 116). Il THX fu inaugurato con *Il ritorno dello Jedi* (cfr. The THX Story s.d.). L'azienda è interessata a curare la spazializzazione anche nelle installazioni domestiche, come dimostra la tecnologia Advanced Speaker Array, la quale riconfigura digitalmente i canali surround a seconda del tipo di prodotto (cinema, gioco, musica) (cfr. THX Surround Sound Modes s.d.).

In ambito domestico nel 1982 si introdusse il Dolby Surround, ossia la versione del Dolby SVA per VHS, Betamax, Laserdisc e trasmissioni televisive (cfr. McGee 2001: 78). Il principio di funzionamento è il medesimo, ma il nome fu cambiato per distinguerlo dallo stereo domestico che presentava solo due canali frontali, mentre attraverso il decodificatore Dolby Surround si potevano ottenere tre canali dai due discreti, ossia sinistro e destro e surround. Nel 1987 fu introdotto il Dolby Pro Logic, che permetteva di ottenere anche il canale centrale. Da questo momento in poi la dicitura Dolby Surround rimase in uso per indicare la codifica, ossia il fatto che il segnale stereo contiene più canali da decodificare. Del Dolby Pro Logic esistono versioni successive, ciascuna delle quali aumenta il numero di canali riproducibili: il Dolby Pro Logic II (2000) offre un *upmix* in 5.1 nelle modalità Movie, Music e Game (cfr. Dolby Pro Logic II s.d.); il Dolby Pro Logic Ix (2002) offre un *upmix* da stereo o 5.1 a 6.1 o 7.1, ancora nelle tre suddette modalità (cfr. Dolby Pro Logic Ix s.d.); infine il Dolby Pro Logic Iiz (2009) offre un *upmix* da stereo, 5.1 o 7.1 a 9.1, dunque con due canali in alto (cfr. Dolby Pro Logic Iiz s.d.).

Negli anni '80 ci fu anche l'avvento della televisione con sonoro multicanale, con uno standard denominato appunto Multichannel Television Sound (MTS), che venne valorizzato soprattutto negli anni '90 (Sedman 2015: 192-200).

#### **1.2.4 – L'avvento del sonoro digitale: lo standard 5.1**

Il primo ricorso alla riproduzione di suono filmico digitale avvenne nel contesto dell'IMAX (70mm), un sistema che ha un circuito commerciale a sé stante, similmente al Cinerama che però ebbe un successo meno duraturo. Tra il 1971 e il 1988 l'IMAX ricorreva a sei canali magnetici su

---

<sup>17</sup> Per ulteriori documentazioni sui formati panoramici e stereofonici cfr. Wysotsky (1971), che dedica attenzione sia agli Stati Uniti sia all'Unione Sovietica.

pellicola 35mm, mentre nel 1988 venne introdotto il Digital Disc Playback (DDP), in cui il sonoro non compresso era inciso su tre CD per un totale di sei canali: sinistra, centro, destra, sinistra posteriore, destra posteriore, in cima allo schermo (in seguito il DDP fu sostituito dal Digital Theatre Audio Control [DTAC], in cui la colonna non compressa è contenuta in un DVD o in un hard disk). Le basse frequenze di tutti questi canali sono inviate a un poderoso subwoofer dietro lo schermo (cfr. Sørensen 1997; Anonimo 2012; IMAX Corporation 1999: 4, 24; Valentini 2006: 112). Nel 1993 nacque invece l'IMAX 3D Sound System, associabile o meno alla versione stereoscopica dell'IMAX. Le informazioni su questo sistema sono oscure: alcune fonti parlano di dieci canali e altre di otto, ma poiché i CD erano quattro suppongo che essi fossero in realtà otto. In ogni caso i canali principali erano diffusi da altoparlanti nella sala (se davvero erano otto resta da capire la collocazione dei due nuovi rispetto ai sei precedenti); due ulteriori canali erano relativi al Personal Sound Environment e diffusi da piccoli altoparlanti posti, in prossimità delle orecchie, nel dispositivo personale simile a cuffie (cui si poteva montare anche la coppia di lenti per proiezioni stereoscopiche). Questi due canali contenevano suoni che potevano incrementare l'effetto a 360° del campo uditivo, o forse che si intendevano generati alle spalle dell'utente (cfr. Berger in Schoenherr 2000; Scott 1994; Empire 1997; Valentini 2006: 112). Il primo film a usare questo sistema fu *Wings of Courage* (USA/Francia 1995), che è anche il primo film narrativo realizzato in IMAX (cfr. Blair 1995). Grazie alla tecnologia Proportional Point Source Sound, l'IMAX vanta di estendere il *sweet spot* per includere l'intera sala (cfr. Rayner 2013).

Il primo sistema sonoro digitale su pellicola è stato il CDS (Cinema Digital Sound), con configurazione 5.1 (ma c'è almeno un caso in cui i canali erano solo quattro, LCRS, ossia *Edward mani di forbice* [*Edward Scissorhands*, USA 1990]; cfr. Coate 2015). Il CDS fu usato inizialmente solo su 70mm, col film *Dick Tracy* (*Id.*, USA 1990) e poi *Giorni di tuono* (*Days of Thunder*, USA 1990); in seguito fu inserito anche in copie a 35mm, a partire da *Linea mortale* (*Flatliners*, USA 1990, ma solo in Europa) e *Edward mani di forbice* (anche negli USA) (cfr. Cinema Products Division 1991; Valentini 2006: 195, n. 3; Belton 2002: 101; Hauerlev 2003).

L'avvento del digitale su larga scala risale al 1992, con l'uscita di *Batman – Il ritorno* (*Batman Returns*, USA), che faceva ricorso al nuovo sistema Dolby Stereo Digital (poi rinominato semplicemente Dolby Digital; v. figura 4). I sei canali presentano la configurazione 5.1, ossia cinque canali normali (sinistra, centro, destra, surround sinistro, surround destro) e un canale, denominato LFE (Low Frequency Effects), riservato alle frequenze basse e diffuso dal subwoofer (cfr. Dolby Digital 5.1 s.d.; Dolby Laboratories 1997; Valentini 2006: 113)<sup>18</sup>. Nel 1993, *Jurassic*

<sup>18</sup> Secondo Di Donato (2004: 39) il sistema era stato testato l'anno prima in alcune copie di *Rotta verso l'ignoto* (*Star Trek VI: The Undiscovered Country*, USA 1991).

*Park (Id., USA)* inaugura invece il DTS (che prende il nome dall'azienda produttrice, Digital Theater Systems), i cui sei canali sono configurati come quelli del Dolby Digital ma sono contenuti in CD sincronizzati con il proiettore<sup>19</sup>. Nello stesso anno *Last Action Hero – L'ultimo grande eroe (Last Action Hero, USA)* lancia il SDDS (Sony Digital Dynamic Sound), a 7.1 canali: i due canali aggiuntivi sono quello di centro-sinistra e quello di centro-destra, come nel Todd-AO; tutti sono contenuti sulla pellicola (cfr. Allen 1997 38-41; Di Donato 2004: 40-41). Questo sistema ha avuto una diffusione modesta, «largely because of the added expense of executing a 7.1 sound mix in addition to the standard 5.1 one»; la maggior parte dei film in SDDS presenta un mix a sei canali, e dal 2007 non si è più fatto ricorso agli otto canali (cfr. Smith 2013)<sup>20</sup>. Mentre all'inizio «most studios were releasing exclusively in one format or another [...], during the summer of 1995, more and more studios began releasing their films in multiple digital formats»: il suono digitale è diventato così «a triple-system standard» (Belton 2002: 102). In effetti dalla metà degli anni '90 era comune avere le tracce di tutti e tre i sistemi sulla medesima copia, oltre a quella Dolby SVA obbligatoria (per le sale prive di lettore digitale e, in ogni caso, come supporto di backup in caso di problemi anche momentanei dell'audio digitale). Tutti e tre questi sistemi presentano suono compresso, ossia il bit rate della colonna sonora master viene ridotto affinché l'informazione possa essere contenuta sulla pellicola o sul CD. Il Dolby Digital ha il maggior rapporto di compressione, seguito dal DTS e poi dall'SDDS (e infine dal defunto CDS). Dunque se da un lato il sonoro digitale ha apportato miglioramenti rispetto a quello analogico ottico, anche in termini di spazializzazione, dall'altro per circa venti anni non è stato possibile sfruttarlo al meglio – ossia non compresso, come nella fase della post-produzione.

Una ulteriore novità fu introdotta nel 1999 da *Star Wars: Episodio I – La minaccia fantasma (Star Wars: Episode I – The Phantom Menace, USA)*, film che inaugurò il Dolby Digital Surround EX (cioè *extended*, 6.1): il nuovo canale è diffuso in posizione centrale posteriore, ed è ricavato dai due surround con la stessa matrice del Dolby SVA (cfr. Valentini 2006: 118; Dolby Digital Surround EX s.d.). Dal momento che questa matrice permette di ottenere 4 canali – e non solo tre – dai due discreti, si pensò in seguito di sviluppare il “Sonic Whole Overhead Sound”, ossia di ricavare un quarto canale da inviare a un altoparlante posto in alto sopra lo schermo. Tale canale fu usato solo in alcuni momenti di *We Were Soldiers (Id., USA 2002)* in proiezioni limitate in Arizona, ma l'idea denota la volontà di ricorrere alla spazializzazione verticale oltre a quella orizzontale (cfr. Wright

<sup>19</sup> Per la precisione i canali incisi sul CD sono 5, mancando quello LFE il quale viene estratto dal decodificatore a partire da informazioni presenti nei due canali surround (cfr. DTS 1999: 2). Valentini afferma che il DTS può «accogliere fino a dieci canali» (2006: 114): ciò si riferisce a una versione successiva del sistema chiamata DTS XD10 (cfr. Digital Theater Systems 2003: viii).

<sup>20</sup> Cfr. la lista completa di film in SDDS 8 (SDDS s.d.), da cui emerge che l'ultimo film così realizzato è *Surf's Up – I re delle onde (Surf's Up, USA 2007)*.

2015: 246; Fishman in Robinson 2003). Sempre nel 1999, con *Haunting – Presenze* (*The Haunting*, USA), la Digital Theater Systems lanciò il formato DTS-ES (Matrix 6.1), in cui il nuovo canale è analogo a quello del Dolby EX; l'anno seguente si introdusse il DTS-ES Discrete 6.1, in cui detto canale è invece discreto (cfr. Buettner 2000; Kerins 2011: Appendix A).

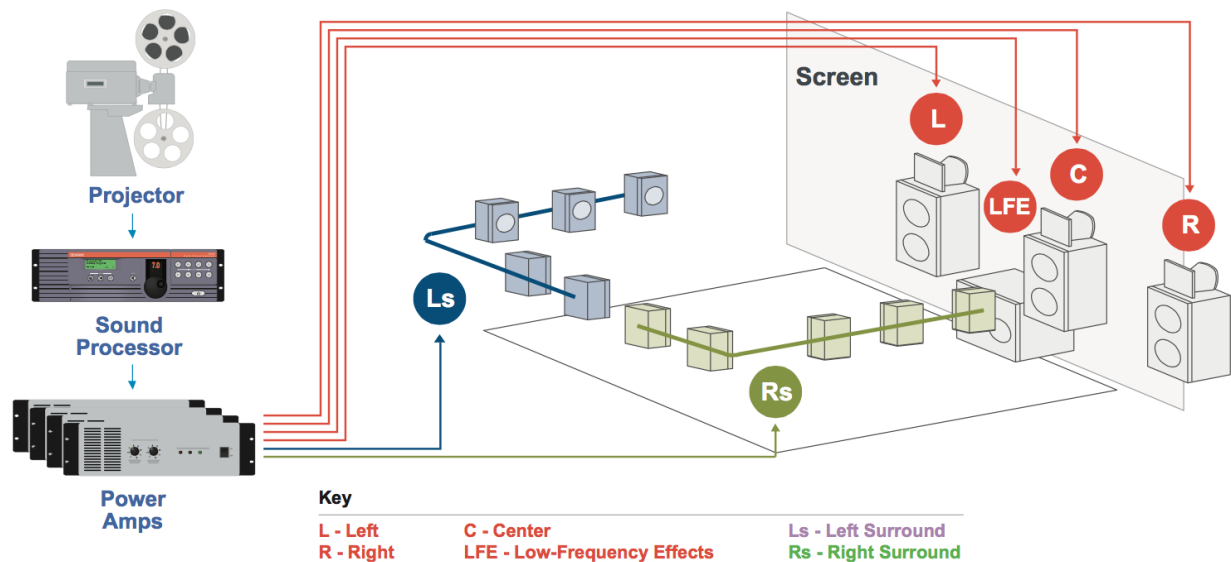


Figura 4: il sistema Dolby Digital 5.1 (Dolby Laboratories 2014: 3).

### 1.2.5 – Nuove frontiere del cinema digitale: *Immersive Sound*

All'inizio del secondo decennio dell'attuale millennio il cinema diventa normalmente digitale, nel formato DCP (Digital Cinema Package) che è il nuovo standard internazionale in sostituzione della pellicola. Il sonoro del DCP è nel formato PCM (Pulse Code Modulation) (cfr. DCI 2012: 27), che già dagli anni '80 costituisce lo standard dell'audio digitale, con frequenza di campionamento di 48 (o 96) kHz, profondità di 24 bit, e configurazione di canali 5.1. Ciò significa che il segnale non è più compresso, e anzi supera la qualità (oltre al numero di canali) del CD. Oltre al PCM 5.1 il DCP può ospitare altre tracce sonore (16 in totale) (cfr. DCI 2012: 68).

La Dolby abbandonò momentaneamente l'idea della spazializzazione verticale, preferendo sfruttare le nuove possibilità del cinema digitale per incrementare la spazializzazione orizzontale. Nacque dunque il Dolby Surround 7.1, inaugurato con *Toy Story 3 – La grande fuga* (*Toy Story 3*, USA 2010): ai 3 canali frontali si aggiunsero 4 canali surround, ossia sinistro e destro laterali e sinistro e destro posteriori (cfr. Dolby Surround 7.1 s.d.). In breve però Dolby decise di investire in un altro sistema che si pone come una novità ancor più radicale: il Dolby Atmos, inaugurato da

*Ribelle – The Brave (Brave, USA 2012)*<sup>21</sup>. Esso è un sistema ibrido, essendo contemporaneamente *channel-based* – come i sistemi precedenti – e *object-based*. Di base contiene dieci canali nel senso tradizionale del termine, nella configurazione 9.1: i nove canali normali sono quelli del 7.1 – ossia sinistra, centro, destra, sinistra laterale, destra laterale, sinistra posteriore, destra posteriore, LFE – cui si sommano due canali in alto. Ai dieci canali si possono aggiungere fino a 118 oggetti audio: si tratta di file che non sono inseriti in canali, bensì contengono metadati che indicano al decodificatore a quale altoparlante deve essere inviato il dato suono. Gli altoparlanti a disposizione nella sala sono fino a 64. Al subwoofer tradizionale dietro lo schermo si aggiungono due subwoofer surround, per incrementare l’impatto delle basse frequenze e renderlo più omogeneo nella sala. Dolby si è preoccupata di rendere il sistema retrocompatibile, consapevole di questa esigenza alla luce di quanto accaduto in passato. Il pacchetto Dolby Atmos contiene dunque anche mix in 7.1, 5.1; da quest’ultimo si può ricavare quello stereo matriciale Lt-Rt (cfr. Dolby Laboratories 2014).

Il Dolby Atmos permette dunque di incrementare il realismo percettivo, anzitutto ampliando l’area di provenienza dei suoni. Inoltre la localizzazione spaziale è più accurata, poiché gli oggetti audio possono essere collocati in maniera puntiforme in un singolo degli altoparlanti a disposizione nelle sei aree surround, e non solo come un «aural ambience emanating from a broader area of the theatre» (Smith 2012). Lo stesso vale per l’impressione di movimento, allorché il suono non passa da un canale (o *bed*) all’altro – ossia da un insieme di altoparlanti all’altro – bensì percorre ogni singolo altoparlante a disposizione. Allo stesso tempo il Dolby Atmos impone che tutti gli altoparlanti surround siano equivalenti ai tre frontali in termini di risposta in frequenza: ciò comporta che le aree surround non siano più svantaggiate rispetto a quella frontale per quanto concerne la fedeltà. Infine Dolby dichiara che l’Atmos abbandona il consueto principio del *sweet spot* a favore di un approccio *room-centric* che permette di fruire il suono in modo adeguato da qualunque poltrona nella sala (cfr. Dolby Laboratories 2014; Sergi 2013; v. figure 5-8).

---

<sup>21</sup> Il primo film italiano missato in Dolby Atmos in assoluto è il cortometraggio documentario *San Siro* (2014) (cfr. Facebook 2015), mentre il primo lungometraggio a normale distribuzione è *Youth – La giovinezza (Youth, Italia/Francia/Regno Unito/Svizzera 2015)*, ma solo nella versione originale in inglese (ringrazio Gianfranco Callarello per questa informazione) (cfr. Perri 2016).

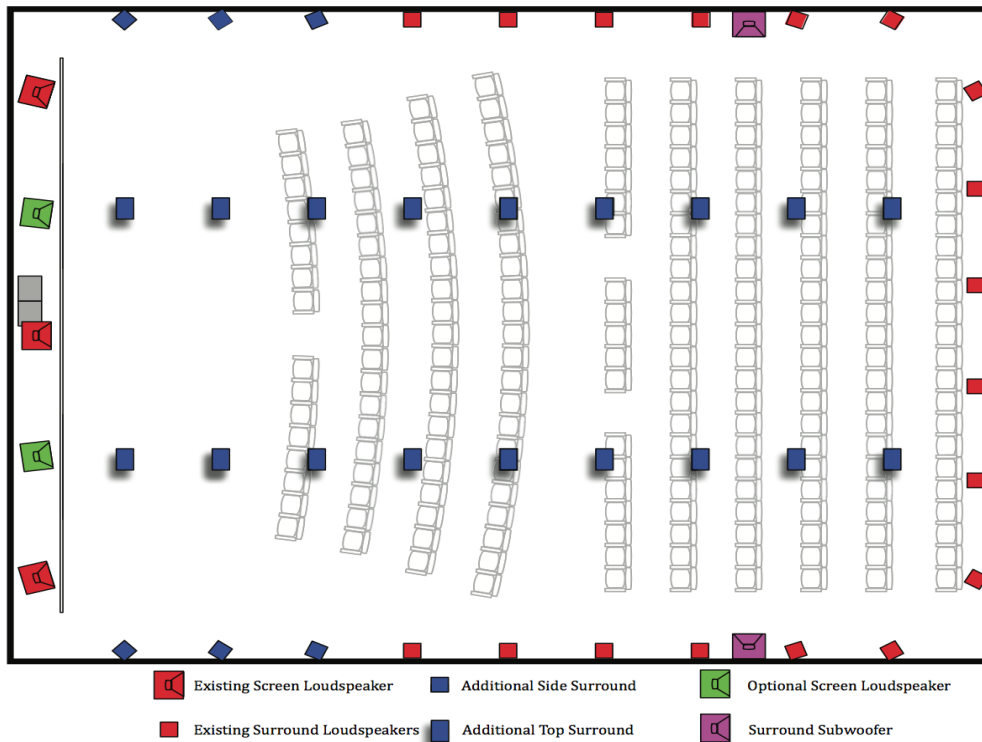


Figura 5: La collocazione degli altoparlanti in una sala con il Dolby Atmos (Dolby Laboratories 2014: 13).

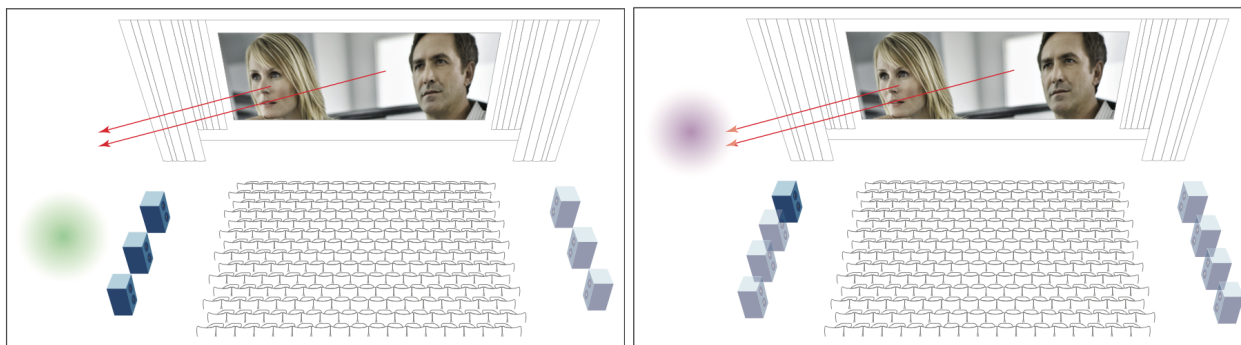


Figura 6: La diffusione di un suono inserito in un canale (a sx, cerchio verde) o di un oggetto sonoro (a dx, cerchio viola) (Dolby Laboratories 2014: 6).



Figura 7: La distinzione tra beds (canali) e oggetti nel Dolby Atmos (Dolby Laboratories 2014: 7).

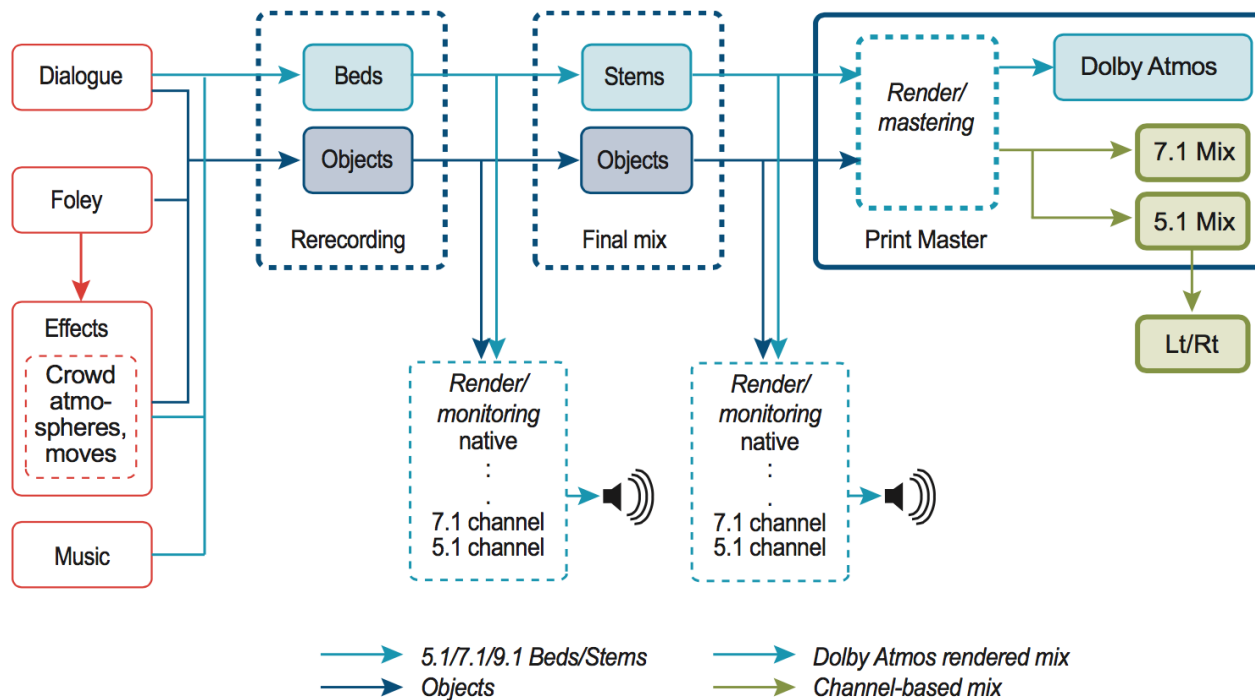


Figura 8: Le fasi del missaggio in Dolby Atmos (Dolby Laboratories 2014: 11).

Molto meno diffuso del Dolby Atmos ma di precedente invenzione, l'Auro-3D è stato il primo sistema a ricorrere a una ricca spazializzazione verticale, dopo gli sporadici tentativi del passato. Esso è esclusivamente basato su canali, che in termini discreti sono i sei del 5.1 e talvolta gli otto del 7.1. L'idea dell'azienda Barco è stata quella di concepire un sistema davvero tridimensionale di suono che in più potesse sfruttare gli attuali standard per ricavarne ulteriori canali superiori, così da garantire massima qualità sonora e compatibilità. L'audio dell'Auro-3D necessita di essere descritto usando non solo due assi ortogonali ( $x, y$ ) bensì tre ( $x, y, z$ ), denotando così uno spazio propriamente tridimensionale. Tale spazio è suddiviso in tre strati: quello surround tradizionale, quello in altezza, quello in cima. Il PCM del DCP ha una profondità di 24 bit, tuttavia secondo alcuni studi 20 bit sono sufficienti per rendere appieno la gamma dinamica percepibile dal soggetto umano. Perciò l'Auro-3D ricorre a soli 20 bit di profondità sfruttando i 4 rimanenti per contenere informazioni che il decodificatore invia ai canali supplementari, corrispondenti a insiemi di altoparlanti disposti in alto e in cima; in particolare, in cima alla sala il segnale è monofonico (cfr. Van Baelen et al. 2012; Van Daele e Van Baelen 2012). La configurazione più comune dell'Auro-3D è quella Auro 11.1, con cui il sistema è stato inaugurato al cinema con il film *Red Tails* (USA 2012) (cfr. Goldberg 2012). L'Auro-3D segue dunque un principio per un verso affine a quello del vecchio Dolby SVA: la colonna sonora è presentata nello standard internazionale PCM, riproducibile in ogni sala senza perdita di informazioni; se la sala è dotata di decodificatore Auro, talune di queste informazioni (quelle relative ai canali superiori) vengono riprodotte come da

intenzioni dei fonici, stavolta senza il pericolo di errori come quello del “magic surround”; altrimenti esse vengono ignorate, permettendo comunque una riproduzione in alta definizione aderente alle intenzioni degli/delle artisti/e (cfr. Van Daele e Van Baelen 2012; v. figure 9-10).

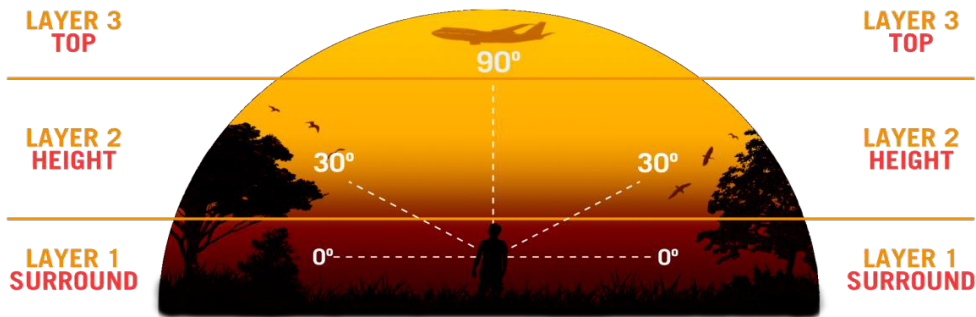


Figura 9: «Auro-3D®'s unique three-layer concept» (Barco 2015: 7).

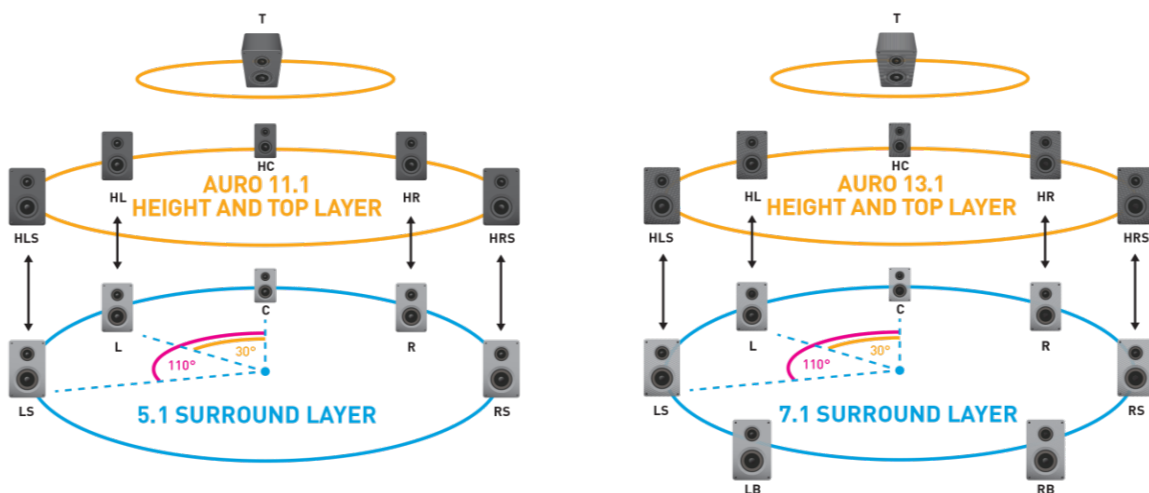


Figura 10: Le configurazioni Auro 11.1 e Auro 13.1 (Barco 2015: 16).

L'AuroMax, evoluzione dell'Auro-3D, è anch'esso un sistema ibrido che combina canali e oggetti. La sua configurazione ideale è 26.1, ossia 27 zone in cui collocare altoparlanti nella sala. Mentre l'Auro-3D vantava di usare solo canali per risultati immersivi migliori (cfr. Claypool, Van Baelen e Van Daele 2012), l'AuroMax integra gli oggetti sonori: ciò denota il successo del principio degli oggetti stessi, probabilmente grazie soprattutto alla diffusione del Dolby Atmos, usato sempre più spesso per grandi produzioni spettacolari. Come il suo predecessore, l'AuroMax si presenta come un sistema a tre strati di altoparlanti: quello surround, quello in altezza e quello in cima. L'azienda enfatizza la capacità di questa configurazione di riprodurre le riflessioni acustiche dei suoni rendendoli più naturali, migliorando il timbro e la localizzazione (cfr. Barco 2015). La Barco enfatizza il fatto che l'AuroMax massimizza le dimensioni del *sweet spot* (cfr. 2015: 17; v. figura 11).

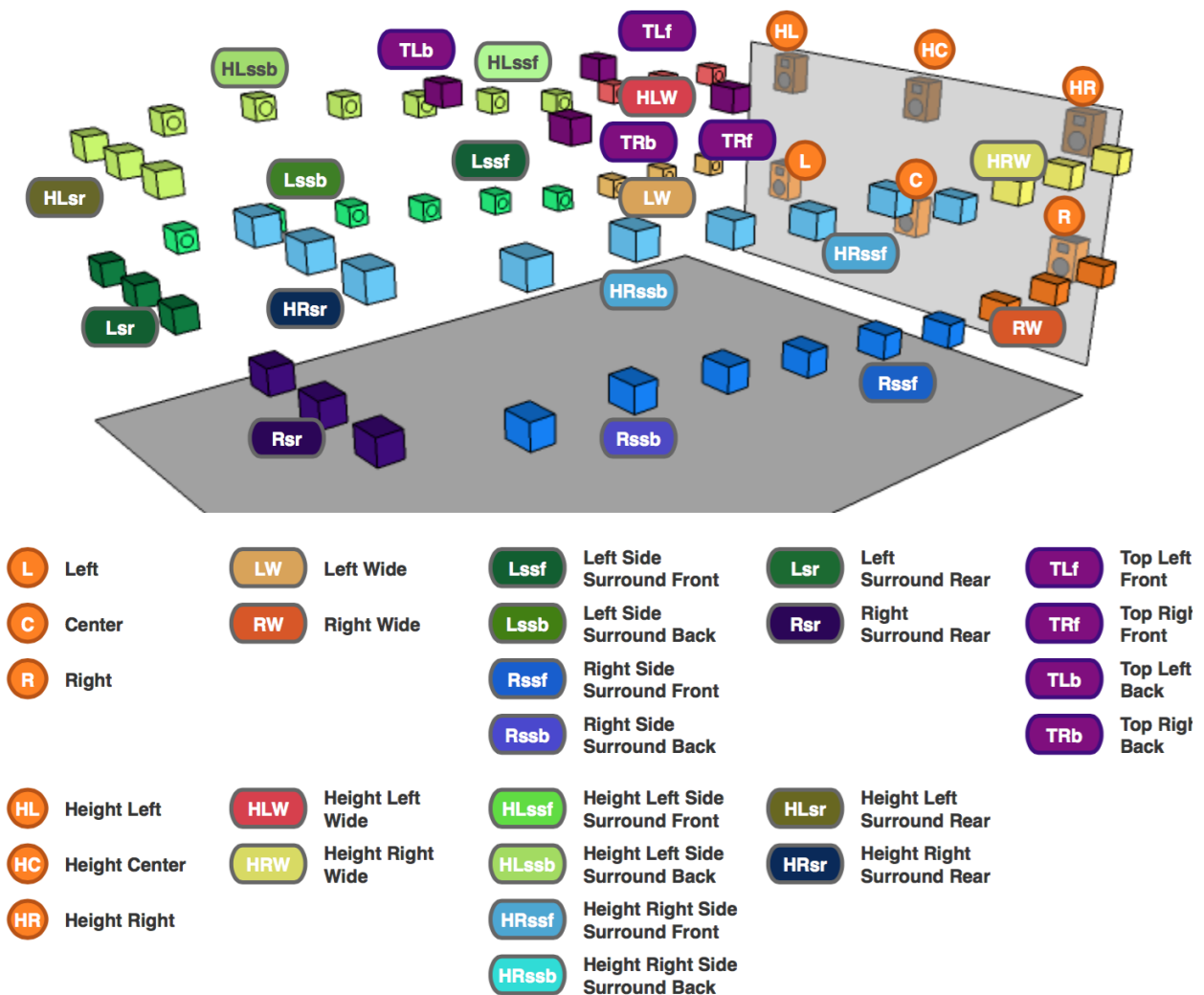


Figura 11: La configurazione AuroMax 26.1 (Barco 2015: 22).

Il funzionamento del DTS:X è meno chiaro dal materiale informativo reperibile pubblicamente, tuttavia anch'esso propone un'ibridazione tra canali e oggetti, e vanta immersività, realismo, flessibilità (poiché il sistema si adatta a qualsiasi configurazione di altoparlanti), fedeltà alle intenzioni degli artisti, retrocompatibilità. È interessante notare che l'azienda parla di prodotti audiovisivi – includendo oltre ai film anche i videogiochi, la radio, la musica – e considera svariati ambienti e modalità di fruizione quali la sala cinematografica, l'home theater, le cuffie, i dispositivi portatili, l'autoradio (cfr. DTS 2014, 2015a, b, c, 2016a, b).

Ciò che le nuove tecnologie di suono immersivo offrono indiscutibilmente è un potenziale aumento di realismo percettivo. Come utilizzarle è una facoltà dei fonici, che continuano a ispirarsi alle norme di spazializzazione già note, benché vi siano eccezioni memorabili che lasciano intendere possibili, allettanti sviluppi. Come osserva Latini, *Gravity* (*Id.*, Regno Unito/USA 2013) si è avvalso del Dolby Atmos con risultati straordinari, dimostrando che il sistema può essere usato

non necessariamente per mero scopo sensazionale bensì innervando «consustanzialmente e in notevole profondità di registri il disegno concettuale e formale, la calibratura drammaturgico-narrativa del film» (2015: 112-113). Sergi ritiene che il Dolby Atmos e le altre tecnologie di sonoro immersivo «[are] crossing into the uncharted territory of immersive narrative within conventional cinema» (2013: 114). Nel caso dei FPS e dei video VR l'estrema spazializzazione, quindi l'enfasi sull'area fuori dal campo visivo, è invece consueta perché questi prodotti offrono un'esperienza diversa da quella dei film anzitutto in considerazione della sua interattività (cfr. par. 1.3 e 1.4). Non si tratta perciò di incrementare grazie all'area surround quanto è già presente nell'area frontale, bensì di espandere lo spazio della rappresentazione portandolo a essere di 360° cosicché si riduce il senso di predominanza di un'area visiva schermica cui il sonoro è ancillare. Questa opzione favorisce il realismo percettivo, ma ancora una volta credo che ciò che conta maggiormente sia l'effetto che si vuole ottenere in termini di esperienza audiovisiva (ossia godimento estetico, emozione, intrattenimento, ecc.). La mimesi è dunque una possibilità al pari di altre, così come la celebrazione del dispositivo, ossia l'enfasi sulla capacità della tecnologia in sé e per sé – con un tono spregiativo si direbbe “fine a se stessa”.

Credo quindi che per i nuovi sistemi di sonoro immersivo cinematografico al momento valga per lo più il principio dell'equivalenza funzionale: essi sono utilizzati di solito come un 5.1 incrementato in ossequio a norme tradizionali di resa spaziale (cfr. Di Donato 2014b: 58). L'idea dominante non è né quella del realismo, né quella dell'audio delle attrazioni, bensì quella dell'efficacia drammatica, in cui l'intenzione narrativa conta più della mimesi e in cui lo spazio sonoro è costruito con naturalezza ma senza corrispondere scrupolosamente a quello denotato dalle immagini. Si tratta, citando quanto Kerins dice circa lo stile del surround digitale a confronto col Dolby Stereo precedente, di «a *new implementation of well-established principles*» (2011: 1407). Alla luce dell'attuale scenario crossmediale è lecito supporre che la spazializzazione acustica dei film somiglierà sempre più spesso a quella degli audiovisivi più radicalmente immersivi; tuttavia la varietà delle modalità di fruizione e la rapidità dei cambiamenti tecnologici rende qualsiasi previsione molto approssimativa (cfr. Sergi 2013: 114-119; Kerins 2015: 155<sup>22</sup>).

### 1.2.6 – Questioni filologiche

Uno degli aspetti più frustranti nello studio del sonoro filmico è quello della ricerca di copie filologicamente corrette, per poter comprendere come un film “suonava” all'epoca della sua prima

<sup>22</sup> Il saggio di Kerins contiene anche un sintetico ma valido resoconto dello stile sonoro dell'attuale millennio (2015: 144-155).

uscita, come i cineasti lo avevano concepito. Ciò può risultare difficile per più motivi: l'esistenza di più versioni "originali" (cfr. Kerins 2015: 148), la carenza di informazioni, l'incompatibilità tecnologica, la presenza di manipolazioni successive operate sulla colonna sonora (cfr. Beck 2003: 18). Quest'ultimo aspetto è ancora più grave perché tali manipolazioni benché perfettamente evitabili sono alquanto frequenti, soprattutto a opera di editori privi di scrupolo filologico quali le major.

La grande diffusione dei DVD e poi dei Blu-ray Disc (BD), con la loro capacità di contenere audio multicanale (fino a 5.1 per il DVD, fino a 8 per il BD) da ascoltare con impianti home theater anche di basso prezzo, sembra aver spinto gli editori a remissare le colonne sonore dei film per portarli alla configurazione 5.1, e meno spesso a quelle 7.1 o ai formati Dolby Atmos e Auro-3D, anche se in origine concepite diversamente. Ciò vale sia per film mono ma anche per film stereofonici nati con una configurazione di canali diversa (CinemaScope, Todd-AO, Dolby SVA, ecc.). Capita persino che un film già esistente in 5.1 venga manipolato per adattarlo agli impianti domestici (cfr. Kerins 2011: 193-214; 2015: 154), o forse per ripensamenti da parte di cineasti/e. Occorre dunque prestare attenzione alla validità filologica della copia che si utilizza per lo studio, confrontando le indicazioni a essa relative (p.es., la fascetta del supporto) con specifiche tecniche reperibili altrove: una fonte eccellente in tal senso è il sito [www.imdb.com](http://www.imdb.com). È utile anche consultare siti di recensioni e forum di appassionati in cui si possono rintracciare informazioni utili nonché confrontarsi<sup>23</sup>.

Si tenga presente che il formato 5.1 non garantisce che i vari canali, specificamente quelli surround, presentino un contenuto diverso l'uno dall'altro. Talvolta i due canali surround sono il duplicato l'uno dell'altro, ossia si tratta di "dual mono"; talaltra presentano solo una leggerissima mancanza di sincronia reciproca, al fine di dare l'impressione di una maggiore spaziosità ed evitare la percezione di un'immagine fantasma posteriore, cosa che in una sala cinematografica non avverrebbe poiché il canale surround è diffuso da svariati altoparlanti su tre pareti. Il canale LFE può limitarsi a contenere le frequenze più gravi dei canali originali, senza aggiunte o alterazioni. Si prenda come esempio *2001: Odissea nello spazio*: l'unica edizione digitale contenente il sonoro originale è quella in DVD del 1998 della MGM, poi ripubblicata dalla Warner nel 1999 (con il reinserimento di una frase di dialogo accidentalmente scomparsa). L'audio è 5.1, ma i due surround sono identici, salvo un leggero spostamento temporale durante la prima occorrenza del valzer *Sul bel Danubio blu*. Il canale LFE contiene presumibilmente solo le frequenze più gravi del sonoro originale, per una migliore resa.

La colonna sonora Fantasound di *Fantasia* è purtroppo scomparsa dalla circolazione.

<sup>23</sup> Cfr. p.es. [www.dvdbeaver.com](http://www.dvdbeaver.com), [www.blu-ray.com](http://www.blu-ray.com), [www.hometheaterforum.com](http://www.hometheaterforum.com), [www.originaltrilogy.com/forum](http://www.originaltrilogy.com/forum).

Il Quintaphonic Sound di *Tommy* è stato mantenuto nelle edizioni in DVD e BD, ma in una forma diversa da quello delle copie in 35mm: in queste ultime i canali discreti erano tre da cui ricavare i cinque canali in riproduzione, mentre nell'edizione digitale i cinque canali sono discreti, perché si è attinto al master precedente alla codifica matriciale per le copie positive. Ovviamente non c'era altra possibilità, poiché non esiste un sistema digitale che possa operare come quello del Quintaphonic Sound nelle sale. Ciò permette di ascoltare il sonoro realizzato in studio senza i limiti della codifica matriciale, ma quindi non *esattamente* ciò che il pubblico ascoltava nelle sale attrezzate con Quintaphonic Sound, creando una situazione paradossale che non va sottovalutata sul piano concettuale. Lo stesso discorso si potrebbe fare con film in Dolby SVA, come vedremo tra poco.

I film in CinemaScope dovrebbero avere sonoro 4.0, ossia tre canali frontali e uno surround. Twentieth Century Fox di solito offre il sonoro originale – insieme semmai alla versione remissata in 5.1 – soprattutto nelle edizioni statunitensi ma talvolta anche in quelle europee.

Per i film in Cinerama e quelli con la configurazione Todd-AO il problema si complica, perché non esiste una tecnologia domestica che permetta di avere cinque canali frontali. In questo caso si può ricorrere a due espedienti. Il primo consiste nel missare i canali mediani (centro-sinistra e centro-destra) rispettivamente insieme a quelli sinistro e destro, alterando la spazializzazione senza remissare; il formato dovrebbe essere quindi 4.0, 5.0, 4.1 o 5.1 a seconda che il surround sia mono o *dual mono* e che il canale LFE sia inserito o meno per contenere le basse frequenze originariamente incluse nei vari canali. Il secondo consiste nel ricorrere al formato 7.1, usando i tre canali frontali rispettivamente per i canali centro-sinistra, centro e centro-destra e quelli sinistro laterale e destro laterale per i canali sinistro e destro; ovviamente a meno che l'utente non possa spostare gli altoparlanti per l'occasione l'effetto è comunque alterato, perché suoni che dovrebbero emanare dall'area frontale provengono così dall'area laterale. Tuttavia si può presumibilmente approssimare l'effetto usando una codifica come quelle DTS-HD Master Audio e DTS-HD High Resolution, a 7.1 canali: esse hanno

a speaker re-mapping feature, which allows the user to select from seven different speaker layouts. Speaker remapping ensures that if the sound mixer is using a different speaker layout during audio creation and mixing, the consumer gets the best possible sound quality and spatial characteristics based on their home speaker layout. Speaker re-mapping uses sophisticated algorithms to electronically reposition speakers in the listening room (DTS 2006).

Una di dette configurazioni è quella Front Wide, in cui si aggiungono due canali frontali ciascuno a 60° di angolazione rispetto all'utente<sup>24</sup>. Il problema non si pone per i film che presentano questa configurazione ma i cui canali di centro-sinistra e di centro-destra sono solamente un mix dei canali adiacenti: in questo caso i canali mediani si possono eliminare senza perdita di informazioni.

I film in Dolby SVA e Ultra-Stereo devono presentare un sonoro stereo a due canali. Attivando il decodificatore Dolby Pro Logic si ottiene dai due canali discreti i quattro canali matriciali previsti. Teoricamente si potrebbe in aggiunta offrire al pubblico il sonoro a quattro canali del master precedente alla codifica matriciale, usando la configurazione 4.0, per far ascoltare il sonoro realizzato in studio dal fonico di missaggio prima dell'applicazione della limitante matrice Dolby. Tuttavia non è chiaro se ciò è sia stato mai fatto<sup>25</sup>; sembrerebbe che nella quasi totalità dei casi in cui il sonoro non venga remissato esso sia offerto nella versione a due canali, come quello contenuto nelle copie in 35mm.

I film in Dolby "baby boom" devono presentare un sonoro 4.1 o 5.1, a seconda che il surround sia mono o *dual mono*. Non è possibile avere due canali LFE, ma ciò non costituisce un problema dal momento che i due canali "boom" sono identici. La configurazione *split-surround* è identica al 5.1.

Tra i vantaggi del BD rispetto al DVD c'è il fatto che i film possono essere riprodotti a 24 fotogrammi al secondo, come da standard cinematografico. In particolare se il DVD è realizzato col sistema PAL (lo standard italiano) la velocità è di 25 fps, con conseguente innalzamento del diapason di circa un semitono (cfr. Miceli 2009: 28), mentre se è in NTSC (30 fps, di cui 24 originali e 6 doppianti) la velocità di riproduzione della colonna sonora non subisce alterazioni.

### 1.3 – Videogiochi

Come spiega Collins, il sonoro surround è stato un'importante novità per i videogiochi a partire dagli anni '90, inizialmente per giochi di corsa e di simulazione. La prima scheda audio per PC col supporto per il surround è Diamond Monster Sound (1997): «parameters such as a room's size and acoustic properties could be programmed into a game, which would initiate filters and effects to simulate the space». Il sonoro surround sarebbe diventato rilevante nel decennio seguente, permettendo al giocatore di localizzare il suono in un'area esclusa dal campo visivo; indizi su tale

<sup>24</sup> Si tratta della Configuration 6 nel documento DTS (2006: 12), dove tuttavia l'immagine è errata; quella corretta è reperibile in [https://sunmaiblog.files.wordpress.com/2010/10/8ecc8ac3b8c84fb5b41cda386f2f85a5-as\\_thumb.jpg?w=796&h=508](https://sunmaiblog.files.wordpress.com/2010/10/8ecc8ac3b8c84fb5b41cda386f2f85a5-as_thumb.jpg?w=796&h=508) (29/10/2016).

<sup>25</sup> Si può supporre che sia il caso di *Crocevia della morte* (Miller's Crossing, USA 1990) e *Poliziotto sadico* (Maniac Cop, USA 1988), ma non sono in possesso di notizie certe.

area possono essere forniti anche da stimoli visivi quali mappe, ma il suono rende lo scopo più immediato e più agevole perché contemporaneo alla visione del campo visivo. Di particolare interesse è il cosiddetto *positional audio*, «which used signal processing to locate a single sound in a specific location in a 3D space», cosicché «sound objects in a virtual space could maintain their location or path of motion while the gamer's character moved about [...] Volume was then adjusted depending on the distance to the player's character». Il surround si è sviluppato in particolare con l'avvento dei "first-person shooter" (FPS, ossia soprattutto in soggettiva) a partire da *Wolfstein 3D* (1992) e grazie soprattutto a *Doom* (1993), poiché gli effetti sonori possono comunicare fruttuosamente la collocazione dei nemici in uno spazio a 360°, anche al di fuori del campo visivo. Qualcosa di analogo avviene in giochi *stealth* a partire da *Thief: The Dark Project* (1998) (Collins 2008: cap. 4). Il 5.1 è dunque divenuto uno standard anche per i videogiochi: «all next-gen consoles are compatible with Dolby Digital's discreet surround 5.1». Sempre più spesso il sonoro surround «is used to help create a more realistic and immersive environment», nonché più utilmente «as an audio cue» (Collins 2008: 1393-1396, 1400).

Dockwray e Collins notano giustamente che i videogiochi sono «a significant part of continuing industry-led movement toward greater numbers of speakers and new technological standards». Si riscontrano anche investimenti sull'audio binaurale e, soprattutto in giochi di corsa, sull'*ambisonics*. La collocazione dei suoni nei giochi deve essere «constantly updated as the game is played», in base alla posizione virtuale del giocatore nello spazio denotato: ciò in fase di produzione avviene «using a three-dimensional graphical representation of the audio listener (player position)», e considerando due parametri per la prospettiva uditiva quali l'attenuazione (per dare il senso di distanza) e la spazializzazione (per dare il senso di direzione) (2015: 256-258).

Un altro dato fondamentale rilevato da Dockwray e Collins è che di solito i sistemi surround sono concepiti in riferimento alle convenzioni filmiche, le quali talora non sono ottimali per i giochi; tuttavia gli sviluppatori devono tenerne conto e adattarvisi (cfr. 2015: 258). Similmente Grimshaw sostiene che «the unknown factor of the player's use of stereo loudspeakers, surround sound systems or headphones presents problems for designing sound localization cues in FPS games» (2007: 3). Ciò evoca la complessa relazione tra l'ambiente dell'utente (*user environment*) e l'ambiente del gioco (*game environment*) analizzata da Stockburger (2003: 2), che vale anche per il cinema ma che nel caso dei giochi è più delicata perché la cognizione dell'ambiente sonoro può essere cruciale per la dinamica del gioco e quindi per l'esito della partita, e non solo un arricchimento estetico. Ad ogni modo l'*exit door effect* non costituisce un problema nei giochi, anzi la collocazione di suoni puntiformi nell'area surround può essere utile «to give the player a full 360-degree sonic space». Da questo punto di vista secondo Dockwray e Collins il sonoro surround è

«more important in video games than in film (depending of course on the genre of the game)», perché aumenta non solo il godimento o il senso di avvolgimento, ma anche l'abilità nel gioco offerta dalla localizzazione, rilevante soprattutto nel ricorso alla soggettiva. Il fatto che il gioco, diversamente dal film, sia interattivo, e che quindi il giocatore possa deacusmatizzare lo spazio fuori campo, favorisce il ricorso a una spazializzazione a 360°: la presenza di suoni nell'area surround non costituisce una distrazione, essendo parte fondamentale della dinamica del gioco (2015: 258-259; cfr. Goodwin 2009: 2). Un esempio è fornito da una clip di *Doom 3* in AstoundSound pubblicata dall'azienda stessa su YouTube<sup>26</sup>.

Kerins aveva già proposto simili riflessioni, distinguendo almeno tre approcci al 5.1 nei videogiochi: il primo è quello «immersive first-person» (gli esempi citati sono *Dead Space* [2008], *Halo 3* [2007], *Brothers-in-Arms: Hell's Highway* [2007]), che abbiamo già descritto; il secondo è l'approccio «third-person», «where the soundtrack suggests action beyond the frame, but the player occupies an outside perspective not identical to that of an onscreen character hearing» (come avviene in *Lego Batman* [2008]); il terzo approccio «uses surround to create a three-dimensional soundscape not oriented to specific onscreen action», come avviene in *Rock Band 2* (2008) in cui, a prescindere dalla prospettiva visiva, la spazializzazione resta immutata presentando gli strumenti nei tre canali frontali e i suoni del pubblico nei due canali surround (questo approccio è tipico dei filmati di concerti). Kerins ha anche ipotizzato le ragioni per cui «the video game industry has been so quick to embrace the possibilities of media that *requires* 5.1 surround»: la prima è che nell'ambito dei giochi si è più inclini ad accogliere nuove tecnologie; la seconda è ovviamente che «the immersive nature of 5.1 is a natural fit for the interactivity of video games»; la terza consiste nell'assunto che chi spende molti soldi per un sistema di gioco è probabile che possa permettersi anche un home theater (Kerins 2011: 3771-3799).

## 1.4 – Realtà virtuale

Su YouTube e altre applicazioni mobili ci sono video di realtà virtuale (“virtual reality”, VR) che è possibile fruire con un comune smartphone inserito in economici dispositivi come il *cardboard*, una scatoletta che si appoggia sul volto dotata di due lenti che restituiscono un'immagine stereoscopica. Spostando lo smartphone ci si può orientare a piacere in uno spazio a 360°, restando sul posto. Alcuni di questi video hanno audio spaziale (“spatial sound”), fruibile con auricolari, in cui il segnale sonoro è processato in modo tale da rendere accuratamente la spazializzazione delle fonti:

<sup>26</sup> “Doom 3 in full-blown AstoundSound® (Headphone Demo)”, <https://www.youtube.com/watch?v=mb57qpc0RD0> (24/10/2016).

modificando il punto di vista si modifica verosimilmente il punto di ascolto. Consideriamo alcuni esempi reperibili su YouTube. In *Philharmonic 360 video with spatial audio - Manuel de Falla*<sup>27</sup> abbiamo l'impressione di essere concretamente presenti in mezzo a una vera orchestra che sta provando un pezzo; lo stesso vale per *3D Spatial Audio : Trevi Fountain sound (360 Video)*<sup>28</sup> e *Gorillas in the Congo: VR with Spatial Audio*<sup>29</sup>, in cui ci sembra di essere davvero, rispettivamente, in mezzo alla folla presso la Fontana di Trevi e nella foresta davanti a un gorilla. Il caso di *The Resurgence of Jazz in New Orleans*<sup>30</sup> e *Sexteto Fantasma VR 360 - Canaro en Paris - Audio 3D*<sup>31</sup> è un po' diverso, perché gli artisti sono disposti in maniera inconsueta a nostro beneficio, palesando così l'artificialità della rappresentazione che implica una certa autoreferenzialità nell'uso della tecnologia, presumibilmente a discapito dell'assorbimento, benché i punti di vista e di ascolto siano ancora realistici, nel senso che una tale esperienza è mimetica verso quanto accadrebbe se fossimo concretamente presenti in quello spazio. L'utente ha come unico scopo di esplorazione dello spazio quello di godere del realismo percettiva dell'esperienza virtuale. Altri video si avvicinano alla narrazione: *360° Camera In Haunted House [4K Quality With 3D Surround Sound]*<sup>32</sup> e *360° Horror: The Ring VR Experience*<sup>33</sup> offrono un'esperienza vicina al film horror, o meglio alla casa stregata dei parchi divertimenti: la tecnologia è al servizio di eventi inverosimili, ancorché realistici in termini percettivi perché costruiti per attivare appieno la *presenza*, ossia «the feeling of being there» (McMahan 2003: 68). Stavolta l'assorbimento è essenziale per godere appieno dell'esperienza, che intende regalarci brividi: le fonti sonore catturano l'attenzione dell'utente in maniera dinamica, spingendo a modificare l'orientamento per comprendere la pur rudimentale successione di eventi.

<sup>27</sup> <https://www.youtube.com/watch?v=8ESEI0bqrJ4> (24/10/2016).

<sup>28</sup> <https://www.youtube.com/watch?v=oB1up7vRN3Y> (24/10/2016).

<sup>29</sup> <https://www.youtube.com/watch?v=Db-ug08ydI4> (24/10/2016).

<sup>30</sup> [https://www.youtube.com/watch?v=dFq9wbrV-v0&list=PLU8wpH\\_LfhmvCvcBGGui3LHC8DufjgvxNn&index=6](https://www.youtube.com/watch?v=dFq9wbrV-v0&list=PLU8wpH_LfhmvCvcBGGui3LHC8DufjgvxNn&index=6) (24/10/2016).

<sup>31</sup> [https://www.youtube.com/watch?v=uayp7VRoL3Y&list=PLU8wpH\\_LfhmvCvcBGGui3LHC8DufjgvxNn&index=1](https://www.youtube.com/watch?v=uayp7VRoL3Y&list=PLU8wpH_LfhmvCvcBGGui3LHC8DufjgvxNn&index=1) (24/10/2016).

<sup>32</sup> [https://www.youtube.com/watch?v=vQeN\\_YxstTo](https://www.youtube.com/watch?v=vQeN_YxstTo) (24/10/2016).

<sup>33</sup> <https://www.youtube.com/watch?v=SsFZjS19Zu8> (24/10/2016).

## 2 – Teoriche sullo spazio sonoro cinematografico

### 2.1 – Dalle origini del sonoro agli anni '40

#### 2.1.1 – Ingegneria

Nei primi anni del sonoro il problema dello spazio fu ampiamente trattato da fonici e ingegneri data la necessità di comprendere come poter utilizzare la nuova tecnologia. Circa la prospettiva sonora si riscontrano due posizioni teoriche: una secondo cui essa doveva essere coerente con quella visiva; un'altra secondo cui si doveva privilegiare l'intelligibilità del dialogo e, per un altro verso, utilizzare i suoni anzitutto secondo la necessità espressiva. La prima posizione pare essere maggioritaria all'epoca, mostrando il presupposto paradigmatico della mimesi, ma più in teoria nella letteratura che in pratica nei film; essa poi fu abbandonata a favore della seconda che resta tuttora un principio cardine della costruzione spaziale del sonoro.

Fin dai primi anni del sonoro si nota la consapevolezza della rilevanza della riverberazione e quindi dell'acustica del set (cfr. Dreher 1931a; Hopper 1931; Maxfield 1931; Olson e Massa 1934; Loye 1941) e – insieme – della sala (cfr. Hopper 1931; Townsend 1931; Baker e Smith 1934; Wolf 1934; Miller 1938; Seeley 1941). Maxfield e Harrison osservano che un sistema mono perde la direzionalità, ma controllando l'acustica della sala esso permette comunque di «simulate to a considerable degree in the reproduced music the effective space relationships of the original», proprio grazie al tempo di riverberazione che contribuisce alla percezione di profondità (1926: 495). Shea osserva che con l'avvento del sonoro i set hanno dovuto subire modifiche per necessità di acustica, che sono maggiori di quelle dei teatri ove il suono non è mediato dalla tecnologia. Infatti «the stage problems are at least doubled by the fact that two stages, the recording and the reproducing, take additive part in the sound picture between the original playing and the final hearing»; inoltre il microfono non ci permette di distinguere la direzione; infine spesso la costruzione delle scene «differs greatly from the visual and acoustic illusions they are intended to create» (1930: 357). Shea sembra asserire la necessità di mantenere la coerenza prospettica, allorché parlando delle difficoltà nel posizionamento microfonicò allude alla «correlation with the viewpoint of the camera as eyes and ears are correlated» (1930: 358). Bonneau asserisce che un uso intelligente della riverberazione permette di creare «a more or less accentuated sensation which may be called “sound perspective”» nonché di produrre «physiologically a kind of depth to the projection screen» (1931: 412).

Cass enfatizza il concetto di «illusion of reality» incrementata dal sonoro, ma manifesta anche l'idea di film come «form of dramatic expression», cui il sonoro giova (1930: 323-324). Per mantenere l'illusione occorre l'intelligibilità, che ormai è garantita, ma anche la coerenza prospettica che invece è trascurata. La prassi è quella di usare più macchine, con una distanza che va dal campo lungo al primissimo piano, mentre per il sonoro si usa un «single microphone system» che non presenta «any given point of audition, but is the sound which would be heard by a man with five or six very long ears, said ears extending in various directions». Ne risulta un ascolto da una «indefinite position» che danneggia l'illusione, il che per Cass costituisce il difetto maggiore del cinema sonoro. Egli suggerisce dunque di ricorrere a più canali, ognuno alimentato da uno o due microfoni e corrispondente a una posizione delle macchine (1930: 324-326).

Dreher si pone il problema di come riprendere al meglio il dialogo durante inquadrature lontane, in cui il microfono non può essere vicino alla fonte quanto dovrebbe per una buona qualità. La soluzione proposta è quella dei «microphone concentrating devices». Tuttavia anche lui assume che le prospettive debbano essere coerenti, asserendo che per inquadrature in campo lungo occorre che il sonoro sembri distante (1931a: 24). Infatti l'autore segue l'usuale paradigma dell'epoca, ritenendo che le principali caratteristiche di un buon sonoro siano l'intelligibilità del dialogo e la naturalezza, ossia la fedeltà acustica all'originale. La naturalezza è addirittura definita «the sole aim of sound technic»: tuttavia «since the reproduction of sound is an artificial process, it is necessary to use artificial devices in order to obtain the most desirable effects». Un esempio è quello del dialogo, normalmente riprodotto a un volume maggiore rispetto all'originale, comportando «a compromise between intelligibility and strict fidelity» (Dreher 1931b: 756). Dreher a tal proposito riferisce una disparità di opinioni sul tema della prospettiva sonora: mentre si conviene sul fatto che «a close-up picture requires close-up sound», non c'è unanimità circa il trattamento del suono in inquadrature distanti. Alcuni ritengono che bisogna mantenere il primo piano sonoro, mentre «more discerning observers» (l'espressione svela condivisione da parte di Dreher) sostengono la coerenza prospettica, sebbene il sonoro debba restare intelligibile qualora la narrazione lo richieda (1931b: 757): in simili frequenti casi dunque è meglio compiere una scelta innaturale, il che si può fare in maniera sottile senza attirare attenzione (1931b: 764). Peraltro l'autore afferma che «In some cases, the special audible background can be secured *only* by re-recording» (1931b: 759).

Una riflessione esemplare segnata dal realismo è quella di Maxfield, il quale rileva la natura binoculare e binaurale della percezione visiva e uditiva umana, che la tecnica cinematografica deve rispettare (1931: 69). L'illusione di profondità visiva è resa possibile «by the proper choice of the focal lengths of the lens used [and] by artificial methods of illumination»; quella acustica invece è ottenibile «by control of the ratio of direct to reflected sound» (1931: 70). A questo si aggiunge un

effetto psicologico per cui quando si crea la giusta illusione di profondità audiovisiva «the voice appears to follow the image not only forward and backward in the scene but also across the scene» (1931: 71). Malgrado il campo lungo sonoro sia meno intelligibile del primo piano, se associato a un campo lungo visivo è percepito come più intelligibile del primo piano, poiché mantiene l'illusione realistica. Con le tecniche summenzionate è dunque possibile rafforzare l'illusione di realtà per cui il suono sembra provenire dalla fonte visibile sullo schermo, avendo l'accortezza di mantenere la stessa distanza visiva e sonora. La riverberazione influenza la percezione di distanza, come fa inoltre l'effettiva distanza tra microfono e fonte, la quale dovrebbe corrispondere a quella della macchina, ma poiché spesso si usano focali lunghe occorre avvicinarlo (1931: 71-72). Maxfield rileva 3 classi di ambiente in relazione alla quantità di riverberazione: «live», «average», e «dead» (quest'ultima sembrerebbe rendere l'illusione più difficile) (1931: 75). Un suono in campo lungo associato a un primo piano visivo sembra venire da lontano, anche alzando volume: dunque si rompe l'illusione. Tuttavia secondo l'autore essa sarebbe compromessa anche usando un suono in primo piano con un campo lungo visivo (benché l'effetto sia meno fastidioso), perché il suono sembrerebbe provenire da uno specifico punto dello schermo, ossia quello ove è sito l'altoparlante (1931: 77).

La posizione di Nathan Levinson, direttore di registrazione per gli studi Warner Bros.-First National, è più vicina alla prassi che si stabilirà presto, poiché egli enfatizza il realismo ma solo in quanto uno tra vari aspetti fondamentali – peraltro esso è connotato come naturalezza piuttosto che come mimesi, ed è strettamente connesso agli effetti sonori. L'autore spiega che lo scopo finale dei fonici è quello di «provide *pleasing entertainment*»; inoltre egli dichiara che «the importance of having the dialogue always clearly understandable goes without saying», e che insieme al realismo si deve produrre «a *pleasing composite sound track*» (Levinson 1933: 6-7, corsivi miei). Similmente Mueller rileva tra i maggiori problemi del sonoro quello di mantenere il bilanciamento tra dialogo, musica e rumore: se le ultime due componenti sono troppo forti rendono inintelligibile il dialogo, ma se sono troppo piano non sono «lifelike» quindi si perde sia il realismo sia «much dramatic value» (1935: 79).

Fletcher (1934a; 1934b) riferisce dimostrazioni di sonoro in «auditory perspective», ossia in stereofonia (benché questo termine non sia usato dall'autore, che vi farà però ricorso in Fletcher 1940 e 1941). Egli spiega che quando si ha una singola fonte si può creare l'illusione di movimento anche solo manipolando la registrazione; tuttavia il movimento all'indietro è reso più difficile perché oltre a ridurre l'intensità occorre aggiungere riverberazione (1934b: 321). Questi esperimenti sembra che furono apprezzati, stando alle note dell'anonimo curatore dell'articolo (cfr. p.es. 1934b:

317), per l'aggiunta di realismo e l'incremento di possibilità artistiche (cfr. Anonimo 1934: 213, 263).

Garity e Hawkins rilevano alcune deficienze nei sistemi sonori tradizionali, di cui tre sono connessi allo spazio. Il primo concerne la «point-source of sound», che ha vantaggi per il dialogo quando l'azione è sita al centro ma deteriora musica ed effetti. Il secondo è la «fixed localization of the sound-source at screen center»: ciò ha comportato una tecnica centripeta di ripresa e montaggio, che può essere superata con la flessibilità offerta da tre altoparlanti. Il terzo consiste nella «fixed source of sound», mentre «by artificially causing the source of sound to move rapidly in space the result can be highly dramatic and desirable». Gli autori evidenziano quindi i vantaggi del Fantasound per sopperire a questi limiti, tra cui la possibilità di ottenere «directional perspective» (1941: 16-17).

Il realismo è un argomento cardine anche per Olson e Massa, secondo i quali uno dei fattori imprescindibili per ottenerlo è il mantenimento di «reverberation and binaural characteristics of the original sound» (1934: 63). Gli autori rilevano l'esistenza di due tipi fondamentali di scene, sul piano acustico: quelle in cui «action and sound are intimately related»; quelle in cui «sound effects are incidental». Benché non sia palesato, si può desumere che questa distinzione sia correlata alla costruzione spaziale, poiché nel primo caso (che «includes the great majority of motion pictures») i suoni prendono vita nello spazio dell'azione, mentre nel secondo i suoni – specificamente l'accompagnamento orchestrale – «has no direct relation to the scene, but serves only to create a suitable mood in the audience» (1938: 63-64). Gli autori descrivono possibili sistemi in considerazione del loro realismo spaziale. Quello ideale è binaurale e permette di trasferire la riverberazione del set a chi ascolta; tuttavia ci sono due svantaggi, ossia il fatto che ognuno dovrebbe tenere cuffie e essere nella stessa posizione dei microfoni, a meno di sacrificare la coerenza prospettica (ancora una volta assunta come necessaria) (1938: 64-65). Un altro è un'approssimazione al sistema ideale ed è multicanale: esso richiede un gran numero di microfoni per coprire l'intera area dell'azione, ciascuno corrispondente a un altoparlante sito nella stessa posizione; tuttavia un tale numero di canali è problematico, infatti si era già tentato di ridurli a tre, come ha fatto Fletcher (1934a; 1934b) presso i laboratori Bell. Altri problemi consistono nel fatto che «the acoustical characteristics of the theater is superimposed upon that of the set» e nel rischio di allargare troppo i suoni (per ragioni di efficacia acustica) così da farli sembrare provenienti fuori dallo schermo (Olson e Massa 1934: 65-66). La terza soluzione, la più pratica, consiste nel ricorso a un singolo canale in cui, grazie alla registrazione con microfoni direzionali, si può comunque valorizzare la qualità spaziale del sonoro che può essere proficuamente elaborata da chi fruisce grazie a un ascolto binaurale. Detto ascolto permette non solo di valutare la direzione e la distanza

ma anche di focalizzarsi sulla fonte principale discriminandola dai suoni incidentali, migliorando l'intelligibilità. Riproducendo un tale sonoro con altoparlanti direzionali si può evitare che le caratteristiche acustiche della sala mascherino quelle della registrazione (1938: 66-67). A proposito della ripresa direzionale gli autori osservano acutamente che «a “center of gravity” of the recorded sound can be established comparable with the “center of gravity” of the action» (1938: 71). Anche Olson e Massa rilevano l'importanza della coerenza prospettica nella distanza, e inoltre il fatto che per il realismo occorre che i suoni riflessi siano inferiori al suono diretto (1934: 73).

In un intervento di Maxfield sulla stereofonia egli rileva che, non essendoci al cinema un'illusione sonora laterale, «our eyes have been trained to “pull” the sound the necessary distance sidewise, to make it appear to come from the visual image of its source». Grazie alla stereofonia invece «this mental strain is relieved». Essa dunque permette la localizzazione «in both a sidewise and a fore-and-aft direction», migliorando la qualità e il senso di realtà (1938: 131). Quest'ultimo in particolare costituisce un fattore importante per il pubblico, e Maxfield vi insiste arrivando a dire che «to obtain the full, ultimate illusion of reality it will be necessary to combine with a colored stereoscopic picture, stereophonic sound» (1938: 131-132). L'autore è comunque consapevole dell'utilità di una motivazione drammaturgica dell'uso della stereofonia, come emerge nell'esempio della sua capacità di attirare l'attenzione su determinate aree dello schermo in assenza di uno stimolo visivo (1938: 134).

L'articolo di Maxfield, Colledge e Friebus (1938) è un testo fondamentale nella discussione sulla stereofonia dell'epoca. Gli autori si soffermano sull'acustica dello spazio intorno alla fonte e sulla tecnica del posizionamento microfonico. A proposito del primo aspetto, essi distinguono due tipi di «acoustic control». Il primo è quello dei «production set», ove si ambisce a creare l'illusione che il suono sia esistito davvero nel luogo mostrato, per il cui scopo è utile ridurre il più possibile la riverberazione. Il secondo è quello degli «scoring stages», in cui si manifestano due problemi, ossia quelli di produrre da un lato musica di accompagnamento e dall'altro musica da sincronizzare con una fonte visibile; in ogni caso per ottenere «the expected dramatic effect» e un'alta qualità sonora occorre curare la riverberazione, stavolta necessaria (1938: 666-670). Circa la «acoustic perspective», gli autori osservano che con il mono l'immagine sofferisce alla deficienza del sonoro nella localizzazione laterale mentre il sonoro contribuisce alla prospettiva nella profondità (ossia nella distanza) (1938: 670). Anche Maxfield, Colledge e Friebus sostengono che mentre i primi piani sono sonorizzati correttamente c'è la tendenza a posizionare i microfoni troppo vicini alla fonte in inquadrature più distanti. La soluzione proposta dagli autori denota un'attenzione significativa al problema, perciò conviene riportare un estratto esteso:

Where the acoustic perspective is *of no dramatic importance*, a single close-up track can be used for all the picture takes, the sound being dubbed at slightly lower level for the long-shot scenes. If, however, *the perspective contributes materially to the dramatic effect*, it is possible to obtain full acoustic perspective by the use of two simultaneous sound-tracks [*sic*]. The first track has a microphone position corresponding to the closest close-up, while the second track has a position corresponding to the longest long-shot. By mixing these two tracks in the proper proportions in the dubbing process, the sound can be made to appear to come from any intermediate distance necessary to fit the picture (1938: 672, corsivi miei).

Un altro concetto rilevante è quello di «liveness», ossia una costante che include la distanza microfónica e l'acustica dello spazio, determinando la percezione della firma spaziale (1938: 672). Secondo gli autori «the greater the liveness of the reproduced sound, the farther does the source of sound appear to be from the immediate foreground». Si tratta del fattore principale nel determinare l'impressione di distanza; a esso segue «the loudness at which the sound is recorded and reproduced» (1938: 673).

Gli autori, in base a un'indagine sperimentale, hanno rilevato che l'accuratezza nella localizzazione con la stereofonia era la stessa rispetto all'utilizzo di una fonte reale su un palcoscenico; inoltre «the average error, laterally, is somewhat less than half the average error fore and aft, for both stereophonic and the real source of sound», a conferma della nostra destrezza nella localizzazione laterale nonché dell'efficacia della stereofonia (Maxfield, Colledge e Friebus 1938: 675). Da questo consegue che l'incoerenza tra la posizione della voce e quella dell'immagine del parlante sarebbe inaccettabile (1938: 676). Gli autori affrontano quindi il noto problema tecnico della ripresa in coerenza prospettica, notando la difficoltà di posizionare il microfono in un campo lungo, ma soprattutto il problema della debolezza dell'area centrale di una ripresa stereofonica, che era stato risolto ricorrendo a tre microfoni (cfr. Fletcher 1934a; 1934b). Questa soluzione è mutuata dagli autori, che tuttavia ritengono sufficiente utilizzare due soli altoparlanti in riproduzione ai quali inviare il medesimo segnale del microfono centrale, rendendo così efficace il primo piano sonoro al centro. Effetti stereofonici si possono anche sintetizzare, laddove è presente un solo suono alla volta (Maxfield, Colledge e Friebus 1938: 677-678).

Reiskind osserva l'esistenza di due metodi di «multiple-speaker reproduction»:

one, the stereophonic method, uses two or three channels to produce motion of the sound source and thus allows the sound to follow the picture within the confines of the screen and in some cases to produce "off-screen" effects. The other method makes no attempt to provide sound motion within the screen area. Instead, the sound source for music and sound effects, which

generally are not localized on the screen, is broadened beyond the screen area by the use of multiplied groups of loud speakers.

Questo fatto è per lui importante poiché, come è noto fin dagli inizi del sonoro, il dialogo è una questione diversa da musica, cori ed effetti, essendo il primo prodotto «by approximately a point-source» mentre gli altri esistono in un'area ampia che inoltre è spesso fuori campo, da cui deriva l'interesse di ottenere un effetto spaziale piuttosto che una localizzazione della fonte. A tale scopo Reiskind ipotizza persino di diffondere il suono da tutta l'area schermica e talora addirittura di circondare il pubblico col suono (1941: 154-156).

Fletcher discute lo «stereophonic sound-silm system» (rimpiazzando l'espressione “auditory perspective” in uso prima) (1941: 331), osservando – come già in Fletcher (1934a: 10) – che, mentre in teoria per la stereofonia occorre un infinito numero di canali, in pratica ne bastano tre per una spazializzazione orizzontale frontale, mentre qualora si volesse aggiungere la dimensione verticale essi dovrebbero essere nove. Il sistema stereofonico è opposto a quello binaurale, in cui bastano due canali per una resa perfetta (1941: 333-334).

### 2.1.2 – Teoria del cinema

In questo periodo gli studiosi più famosi che hanno riflettuto sulla dimensione spaziale sono Arnheim e soprattutto Balázs. Secondo Arnheim il suono ha la capacità di trasformare «l'immagine bidimensionale [...] in un'apertura su una realtà tridimensionale» (Elsaesser e Hagener 2009: 150), il che però costituisce un limite, come emerge chiaramente da questo significativo passaggio:

L'acustica completa perfettamente l'illusione. Il bordo dello schermo non è più una cornice, bensì il margine di un tutto, di uno spazio teatrale: il suono trasforma lo schermo in un palcoscenico nello spazio. Una delle principali e speciali attrattive del cinema è il fatto che ogni scena pone una sfida: frammentazione di immagini e movimento sulla superficie *versus* corpi plastici e moto nello spazio. Il film sonoro sospende quasi del tutto questo doppio gioco esteticamente significativo (in Elsaesser e Hagener 2009: 151; per altre riflessioni dell'autore sul cinema sonoro cfr. Arnheim 1985).

Balázs sottolinea l'importanza dei primi piani sonori, che avrebbero permesso al film sonoro di diventare una nuova arte qualora fossero stati gestiti con i principi del montaggio («sound-montage») (1985: 116-117). Balázs ritiene che, diversamente dal primo piano visivo, quello sonoro non può venire isolato dal contesto perché l'udito non è selettivo come la vista (1985: 120-121). Egli sostiene che «sound alone is not space-creating», ma esso fornisce realismo allo spazio grazie

al suo senso di profondità (1985: 117-119). L'apparente inferiorità dell'udito rispetto alla vista è dovuta per Balázs in gran parte a una minore educazione del primo rispetto alla seconda, e questa deficienza può essere sfruttata per ottenere effetti sorprendenti; tuttavia l'autore rileva comunque che ci sono dei limiti nell'udito, poiché i suoni non producono forme nello spazio, non possono trasmetterci il senso di dimensione o di direzione, e sono difficilmente localizzabili (1985: 122-124). Queste osservazioni devono ovviamente essere riconsiderate alla luce dello sviluppo del sonoro spaziale. Balázs sostiene che i suoni prodotti su un palcoscenico assumono sempre un colore diverso rispetto al loro contesto naturale, dovuto allo spazio artificiale in cui sono presentati, mentre il microfono li offre con il loro "timbro locale", cioè così come li si ascolterebbe nel luogo che è loro proprio: «we hear the sounds as the microphone originally heard them [...]. [W]e are transferred from our seats to the space in which the events depicted on the screen are taking place» (1985: 124-125).

Ernest Irving, compositore e direttore d'orchestra inglese che ha lavorato nel cinema, commenta le convenzioni nell'uso della musica nei film avanzando una riflessione insolita sulla tecnologia di registrazione sonora, che pare presagire Balázs (1985: 124-125). Secondo l'autore «[i]f the sound-track were made and projected at, say, twice [the normal speed, 90ft.,] it would be possible to hear in a kinema [*sic*] a more perfect performance of a musical work than any listener can purchase in the Albert Hall»: con maggiore spazio fisico si potrebbero includere gli ipertoni, che normalmente spariscono, arricchendo la resa timbrica. Il risultato sarebbe persino migliore di un ascolto in sala da concerto, dove «according to the position of [the auditor's] seat he is subject to echoes, reverberations, maskings and distortion»: con il dispositivo cinematografico invece «[t]he sound could be scientifically collected for him by properly arranged microphones and delivered to him as a correctly balanced whole. [...] The recorded film can give the listener the advantages without the disadvantages» (Irving 1943: 226-227).

## 2.2 – Anni '50-'60

### 2.2.1 – Ingegneria

Gli anni '50 furono un periodo di innovazione in quanto nell'industria cinematografica degli Stati Uniti venne introdotta la stereofonia magnetica. Il numero del *Journal of the SMPTE* del settembre 1953 contiene importanti testi relativi alla questione. Fletcher ribadisce una tesi già sostenuta (1934a: 10; 1941: 333) per cui in teoria occorrerebbe un numero infinito di microfoni e diffusori,

ma in pratica ne bastano tre per una spazializzazione orizzontale frontale, mentre qualora si volesse aggiungere la dimensione verticale essi dovrebbero essere nove; stavolta però l'autore ipotizza che usare il canale centrale dall'alto permette di spazializzare in verticale senza ulteriori canali. D'altronde i sistemi stereofonici non hanno un numero fisso di canali, bensì quanto basta per dare l'illusione che siano infiniti (1953: 356-357). Secondo Fletcher un sistema a tre canali migliora anche la resa del volume (1953: 363).

Grignon parte dal medesimo presupposto di Fletcher, e inoltre afferma che la stereofonia si distingue dal mono e dai sistemi binaurali perché trasporta «*the original performance to the location of the ultimate audience*». L'autore spiega che la Twentieth Century Fox e la Western Electric Co. hanno collaborato nello sviluppo di metodi per «*recording dialogue and music for use in motion pictures, without basically changing accepted fundamental forms which include the use of long, medium and close shots and intercutting techniques*»; tuttavia «*some new approach is needed*» (1953: 364-365). Grignon sostiene che per usare lo stereo occorre considerare il montaggio visivo, il che richiede attenzione da parte di registi, direttori della fotografia e montatori: «*picture may have to concede somewhat more importance to sound*» (1953: 374). L'autore afferma che la stereofonia migliora la qualità sonora e rende meglio l'acustica del set anche trascurando «*subjective sound placement*»; inoltre essa permette di creare più illusioni dunque di aprire nuove strade espressive, ed è valorizzata soprattutto dagli schermi panoramici. I tre aspetti principali della ripresa stereofonica sono «(1) *sound placement matching visible or desired implied action*; (2) *smooth sound placement transition* and (3) *some pickup in all microphones at all times*». Infine Grignon considera «*Surround, or auditorium, loudspeakers*» i quali «*are gaining favor for special effects*»; tuttavia essi «*are frequently incompatible with the screen picture. It seems that surround effects must be used with discretion*» (1953: 375-377).

In un articolo successivo di tre anni, Grignon asserisce che quattro canali sonori sono «*lifelike*» e costituiscono un buon compromesso, ribadendo che idealmente occorrerebbe un numero infinito di canali per riprodurre spazialmente un suono (1956: 22). Parlando nello specifico del film *Carousel* (*Id.*, USA 1956), egli ritiene che si sia ottenuto un buon risultato privo dell'impressione della provenienza di un suono da un punto specifico dietro lo schermo. Inoltre l'uso curato del surround «*further enhanced the sound effectiveness on the big choral and musical numbers*», ed è forse grazie alla spazializzazione acustica che Grignon rileva come la musica «*did not intrude*» malgrado il volume molto alto rispetto alle voci (1956: 36).

Giova considerare alcune riflessioni interessanti in ambito musicale concomitanti alla diffusione della stereofonia, che propongono in ambito audiofilo questioni già incontrate in ambito

ingegneristico. Bohn argomenta su come ricreare sul disco il punto d'ascolto della sala da concerto, contestando l'espressione «concert-hall realism» usata dalle compagnie discografiche. Infatti secondo l'autore «most music sounds best when heard at a reasonable distance from the musicians»; inoltre «the real goal in home music reproduction should be the duplication, as closely as possible, of the sound heard in a good not-too-close seat in a concert hall». A tal scopo non si deve esagerare il volume bensì rispettare fedelmente l'acustica dell'ambiente originale, rendendo il più possibile incospicua quella del salotto (1954: 37). Le compagnie invece usano riprese in primo piano, posizionando i microfoni in alto (1954: 37), piuttosto che presso il sedile centrale come sarebbe opportuno (salvo adattarsi all'acustica della sala) (1954: 94). Circa la stereofonia, Bohn spiega che essa non è necessaria: per il direttore potrebbe essere utile perché rende il suo punto d'ascolto, ma il pubblico di solito è a 2/3 di distanza dall'orchestra, da cui risulta un angolo di ampiezza scarsa (1954: 92).

Plass analizza i tre fattori caratteristici dell'ascolto binaurale, ossia le differenze di intensità, fase e tempo (1954: 78-80); poi osserva che nell'uso di due canali «the aim is to bring you the same intensity and phase differences as you would have heard at the actual recording session». L'autore spiega quindi i vantaggi rispettivi della riproduzione in cuffia e con altoparlanti. Le prime sono eccellenti per l'illusione di ascoltare il suono originale e la direzionalità. Un'idea curiosa di Plass è che il sistema binaurale ideale sarebbe quello che permette di ruotare i microfoni durante la registrazione e poi conformemente, attraverso un sistema meccanico, la testa dell'ascoltatore. Gli altoparlanti sono più usati a causa del disagio delle cuffie, ma ci sono degli svantaggi. Anzitutto accade che a ciascun orecchio giunge anche ciò che dovrebbe ascoltare solo l'altro; poi le differenze di intensità e fase sono diverse dall'originale, «however, there appears to be a considerable improvement in the sound when there are any such differences present from a double source, even if they are not identical with those in the original source»; infine se si ruota la testa le differenze interaurali di intensità e fase cambiano (diversamente da quanto accade con la cuffia), ma non allo stesso modo di come avverrebbe dal vivo, fermo restando l'aggiunta di un «feeling of presence». Plass conclude seguendo – senza palesarlo – la posizione di ingegneri degli anni '30 per cui sebbene in teoria serva un numero infinito di altoparlanti in pratica buoni risultati sono ottenibili con tre, ma spesso anche soli due possono offrire «surprisingly realistic sound» (1954: 80).

Cohen si concentra sul fatto che abbiamo due orecchie, da cui deriva la potenziale utilità della riproduzione binaurale, per lui sinonimo di stereofonia e descritta anche come «third dimensional music». Tuttavia l'autore non considera la stereofonia come buona o cattiva di per sé, bensì come una possibilità in più che per i musicisti è opportuno esplorare per evitare il rischio – sussistente – che essa sparisca «as “just another gadgeteering novelty”». Cohen porta l'esempio di

brani musicali in cui la collocazione spaziale delle fonti è rilevante (p.es. la *Sinfonia n. 4* di Nielsen e l'ouverture *Leonore n. 3* di Beethoven) (1954: 28-30). Un'osservazione pregnante dell'autore concerne la cognizione crossmodale della musica: egli parla di «musical kinetics» per denotare «the movement of the musical themes *physically* before our eyes as well as tonally for our ears». Ciò deriva dalla considerazione che la maggior parte del pubblico segue con lo sguardo il movimento della musica tra i vari strumenti: quindi «the concertgoer is immersed in more than just musical sounds. He is immersed in musical motion as well», un fattore che è opportuno mantenere a casa e che la stereofonia permette. Cohen sostiene che questa potenzialità spaziale della musica, dal vivo e incisa, dovrebbe essere sfruttata dai compositori intrinsecamente ai pezzi (1954: 78, 80). A proposito del movimento concreto delle fonti, l'autore si ispira al cinema ipotizzando di usare «two mobile boom-microphones, separated according to binaural requirements, and move them in sequence past the performers who are to be represented in motion» (1954: 83).

### 2.2.3 – Teoria del cinema

Nel 1969, Burch dedica un importante libro al cinema (*Praxis du cinéma*) che include l'analisi di alcune possibili relazioni dialettiche determinate dal suono. Una è interna alla colonna sonora, e si stabilisce tra due diversi tipi di materiale, cioè «il suono diretto e il suono ricomposto dal missaggio». Altre due risultano invece dal rapporto con l'immagine: la prima riguarda il contrasto tra la distanza apparente del microfono e quella della macchina da presa rispetto alla fonte (p.es. primo piano sonoro e campo lungo visivo); la seconda concerne la creazione di un'analogia tra le due componenti, soprattutto in riferimento a quella che la musica che stabilire con oggetti presenti nell'immagine (p.es. *wood-block* e canne di bambù) (2000: 92-95).

Circa lo spazio filmico Burch identifica tre tipi di rapporti spaziali tra inquadrature: il primo consiste nella «continuità spaziale, *con o senza continuità temporale*», in cui la seconda inquadratura contiene in tutto o in parte il frammento di spazio di quella precedente; gli altri due presentano discontinuità spaziale, e più specificamente il secondo presenta un'inquadratura «*manifestamente vicina* al frammento di spazio visto in precedenza», mentre il terzo comporta una «discontinuità totale e radicale» (2000: 22). L'autore divide lo spazio fuori campo in sei aree: quattro sono attigue ai rispettivi bordi del quadro, la quinta è dietro la macchina da presa, la sesta «comprende tutto ciò che si trova dietro la scena (o dietro un elemento della scena)» (2000: 29); questi spazi entrano in gioco, tra l'altro, grazie al suono off (2000: 38).

## 2.3 – Anni '70-'80

### 2.3.1 – Prassi

Gli anni '70, soprattutto grazie alla diffusione dell'audiofilia e al successo del Dolby Stereo, sono negli Stati Uniti un periodo di "ossessione per il suono", come scrive Schreger. Egli sostiene inoltre che il cinema di Hollywood è più che mai al servizio del realismo, e il sonoro vi contribuisce dando l'impressione di trovarsi dentro un dato spazio (1978: 34). Tuttavia l'autore sembra rifiutare la direzionalità del suono, che in una scena di *Il cacciatore* (*The Deer Hunter*, USA 1978) sarebbe «needlessly disorienting» (1978: 36).

Circa la firma spaziale e i diversi approcci di presa diretta, le riflessioni dell'influente fonico Jim Webb sono particolarmente interessanti. Egli spiega che la sua tecnica di registrazione multitraccia con radiomicrofoni, da lui sperimentata per i film di Robert Altman, è contrapposta all'approccio tradizionale in cui si usano una o due tracce in cui finiscono tutti i suoni: in questo caso raccogliere le informazioni necessarie è più difficile, mentre mantenere la prospettiva è spesso più facile. Per Webb l'ideale sarebbe quello di garantire entrambe le cose. Egli parla significativamente di «philosophical shortcomings» (in Webb e Ketteler 1980: 116) oltre a quelli concreti, riferendosi al fatto che la sua tecnica uccide la prospettiva, intesa anche come «mood» di un ambiente. La soluzione è quella di ridurre la complessità del dialogo per lasciare più spazio ad ambienti ed effetti (1980: 117).

Barnett torna sulla questione coinvolgendo Webb e il suo *boom man* Chris McLaughlin. L'uso del singolo boom è definito «perspective miking» (Barnett 1980: 54) ed è concepito come alternativa o come integrazione alla registrazione multitraccia con radiomicrofoni: in questo secondo caso si tratta di ricorrere al *backmiking*, ovvero si riprende contemporaneamente la fonte principale in primo piano col radiomicrofono e l'ambiente col boom (1980: 60).

### 2.3.2 – Teoria del cinema

Negli anni '80, nel pieno della teoria ortodossa o Grand Theory (Currie 1999; Bordwell e Carroll 1996), si riscontra un incremento di studi sul sonoro che segna l'avvio del processo che porterà alla nascita dei *sound studies*. I testi più rilevanti sono il numero 60 di *Yale French Studies* intitolato *Cinema/Sound*, a cura di Rick Altman (1980), il libro *La Voix au cinéma* di Chion (1982) – con cui l'autore inizia ad affrontare il sonoro filmico, diventandone presto tra i più celebri studiosi

–, e l'antologia *Film Sound: Theory and Practice* curata da Weis e Belton (1985), contenente testi già editi.

Metz usa un approccio semiologico per interrogarsi sul modo in cui percepiamo il mondo dei suoni. Egli ritiene che l'identificazione di un suono comporti la domanda «un suono di cosa?»: la descrizione completa prevede un termine per il suono e uno per la fonte, ma dovendo sopprimere uno dei due termini si favorirebbe il secondo, che garantirebbe la maggiore riconoscibilità (1985: 155). Perciò sul piano ideale «the aural source is an object, the sound itself a “characteristic” [...] linked to the object, and that is why identification of the latter suffices to evoke the sound, whereas the inverse is not true» (1985: 156). Secondo Metz nella cultura occidentale le qualità primarie di un oggetto sono solitamente di tipo visivo e tattile; quelle sonore, olfattive, ma anche alcune tra quelle visive (p.es. il colore), sono percepite come secondarie. In questo senso l'autore parla di “sostanzialismo primitivo”, secondo il quale le qualità primarie determinano la sostanza dell'oggetto, mentre quelle secondarie indicano attributi del medesimo (1985: 156-157). Questo approccio influenza anche la teoria del cinema, laddove si parla di suoni «*off-screen*»: Metz osserva che il suono non è mai fuori campo perché si trova contemporaneamente dentro lo schermo e dentro la sala, cioè si diffonde nello spazio, non ha confini; perciò quando si dice che il suono è fuori campo, ci si riferisce in realtà all'immagine della fonte (su questo cfr. Kerins 2011: 3086-3092). Metz ipotizza che la ragione di queste tesi risieda in un dato fisico, poiché «spatial anchoring of aural events is much more vague and uncertain than that of visual events»; tale approccio subordina indebitamente il suono all'immagine, interpretandolo come un attributo, un non-oggetto (1985: 157-158). Una delle tesi più discutibili di Metz, criticata come vedremo da altri studiosi, è quella per cui non ci sarebbe alcuna differenza sensibile tra un suono reale e la sua riproduzione (1985: 161, n. 7).

Doane è rappresentativa dell'approccio di critica ideologica tipica del periodo. L'autrice vuole dimostrare che le tecniche, il linguaggio e le discussioni relativi alla colonna sonora sono sintomatici di specifici scopi ideologici, cioè l'affermazione dell'unità del film e quindi la negazione della sua eterogeneità materiale. Secondo l'ideologia borghese del visibile l'immagine consiste in una rappresentazione veritiera della realtà e perciò va fruita in modo razionale. Il suono invece, essendo intangibile e privo di concretezza, è connesso all'emozione e all'intuitività, cioè a un altro ordine di realtà (1985a: 54-55). L'assunto ideologico dell'unità rischia di essere contraddetto dal mezzo filmico: ecco perché nella prassi «synchronization and totality are fetishized and the inseparability of sound and image is posited as a goal» (1985a: 56). Doane cita la montatrice Helena Van Dongen, che spiega: «there is no separation of *I see* in the image and *I hear* on the track. Instead, there is the *I feel, I experience*, through the grand-total of picture and track combined» (in Reisz 1964: 155, qui tratto da Doane 1985a: 56). Un'evidente conseguenza di quanto

spiegato è rilevabile nella prassi di celare il lavoro di costruzione della colonna sonora: ecco perché, per esempio, si evitano tagli bruschi oppure il silenzio, sempre sostituito da un suono d'ambiente (Doane 1985a: 57-58).

Doane spiega che il dialogo è l'elemento sonoro più importante, quello che determina il livello di musica ed effetti – i quali non devono mai comprometterne l'intelligibilità – e inoltre è l'unico a restare legato all'immagine durante le fasi produttive (1985a: 58). Relativamente al suono, il concetto di realismo collide con quello di intelligibilità: secondo l'autrice «the Hollywood sound film operates within an oscillation between two poles of realism: that of the psychological (or the interior) and that of the visible (or the exterior)». Il trattamento del dialogo prova la maggiore importanza del realismo psicologico, come dimostrano i compromessi che favoriscono l'intelligibilità a discapito del realismo. Insomma «it is the “talkie” which appears in 1927 and not the sound film» (1985a: 59). Il principio dell'autenticità è applicato più alla voce che a musica ed effetti, i quali possono essere usati in modo non naturalistico se lo scopo primario è quello di ottenere un effetto drammatico (1985a: 61).

La studiosa spiega che nei primi anni del cinema sonoro le voci erano registrate con un approccio proprio della radio e del telefono, cioè si trattavano tutte le voci come se ciascuna fonte fosse alla medesima distanza. In seguito, grazie all'evoluzione tecnologica, questa prassi sarebbe cambiata perché era intollerabile che la prospettiva visiva e quella sonora non coincidessero. Il corpo di spettatore e spettatrice è dunque ancora una volta concepito come unificato e non frammentato (Doane 1985a: 59-60). In realtà nel cinema tradizionale, specialmente a partire dagli anni '60, si riscontra anche una prassi diversa, quella per cui la prospettiva visiva e quella sonora non coincidono: l'esempio tipico si ha quando un campo lungo è montato con un primo piano sonoro (cfr. Belton 1985: 68-69; Altman 1992b: 47-55).

Doane ha anche studiato i problemi del suono in chiave psicoanalitica. La tesi principale è che «the body reconstituted by the technology and practices of the cinema is a *fantasmatic* body»: i suoi attributi sono l'unità e l'impressione di presenza reale. Circa il primo, si confronti il saggio di Doane trattato poc'anzi. Circa il secondo, l'autrice spiega che gli sviluppi tecnologici mirano a ridurre il rumore della registrazione così da nascondere il lavoro del dispositivo e quindi diminuire la distanza tra oggetto e rappresentazione, incrementando l'impressione della presenza del suono. La voce, ancorata al corpo che a sua volta è ancorato allo spazio, è un supporto al riconoscimento e all'identificazione della e con la star; la post-sincronizzazione presenta il rischio di dissociare la voce dal proprio luogo di origine, alterando la percezione dello spazio rappresentato. La tecnica del suono insomma conferma «the cinema's function as a *mise-en-scène* of bodies». La necessità di

offrire un'impressione di spazio omogeneo è resa evidente anche dal posizionamento dell'altoparlante dietro lo schermo (Doane 1985b: 162-165).

Doane segue Metz nell'affermare che il visibile è posizionato gerarchicamente sopra l'udibile: ciò è confermato dall'espressione "voce off", sebbene a essere "off" sia non la voce bensì l'immagine della sua fonte. Il suono viene discusso di solito in relazione all'immagine, ma questo non deve comportare una subordinazione (Doane 1985b: 165-166).

Secondo Doane esistono tre tipi di spazi in un film:

1- lo spazio della diegesi, che «has no physical limits, it is not contained or measurable», (è quello su cui il film narrativo classico fonda la credibilità, a discapito degli altri due che vengono negati);

2- lo spazio visibile sullo schermo, che «is measurable and "contains" the visible signifiers of the film»;

3- lo spazio acustico della sala, in cui il suono si espande confermando che «sound is not "framed" in the same way as the image<sup>34</sup>. In a sense, it *envelops* the spectator» (1985b: 166).

Anche Camper osserva che il suono riempie lo spazio della sala, mentre l'immagine resta confinata sullo schermo (Camper 1985: 371; cfr. anche Buckland 2000: 48; Chion 2009: 226-227). Tale fenomeno offre secondo lui due possibilità estetiche: primo, quella di conferire una presenza spaziale all'azione mostrata, ovvero rendere le immagini più palpabili (idea condivisa da altri teorici tra cui Balázs 1985: 119); secondo, quella di enfatizzare i limiti spaziali dell'immagine (Camper 1985: 371, 380).

Doane sostiene che la voce off amplia la diegesi, essendo al servizio della costruzione dello spazio più che dell'immagine. Il rischio è quello di palesare l'eterogeneità materiale del cinema: tuttavia nel film classico l'espedito è usato in modo disturbante solo qualora si intenda produrre un preciso effetto drammatico, altrimenti si tratta di una dissociazione solo momentanea tra voce e corpo. Il piacere dell'ascolto ha origine nella prima infanzia, periodo in cui lo spazio è definito per il bambino in termini sonori e non visivi. La voce della madre «is a major component of the "sonorous envelope" which surrounds the child and is the first model of auditory pleasure» (1985b: 170; cfr. Kerins 2011: 3394-3395); l'intervento della voce del padre impone una separazione, e di conseguenza la voce della madre si configura come l'oggetto perduto (Doane 1985: 166-171). Al cinema l'involucro sonoro è costruito in modo tale da sostenere il piacere narcisistico derivato dalla percezione di unità e identità stabilita da spettatore e spettatrice in relazione al proprio corpo: «the aural illusion of position [...] guarantees the singularity and stability of a point of audition, thus

<sup>34</sup> Chion, come vedremo, enfatizza così fortemente questo dato da affermare conseguentemente che la colonna sonora (*bande-son*) non esiste (cfr. 1999: 3; 2001: 45; 2009: 227-230).

holding at bay the potential trauma of dispersal, dismemberment, difference» (1985: 171). Per ottenere questo effetto occorre attenuare l'aggressività della voce, la quale è dotata di una profondità che contrasta con la piattezza dell'immagine. In sostanza il cinema classico mira a ottenere effetti omogeneizzanti, avendo «a stake in perpetuating the image of unity and identity sustained by [the spectator's] body and in staving off the fear of fragmentation. The different sensory elements work in collusion and this work denies the material heterogeneity of the "body" of the film» (1985: 171-172).

Belton segue la riflessione di Jean-Louis Baudry secondo cui il dispositivo cela le tracce della trasformazione che opera sulla realtà; questa prassi, come già osservato da Doane, riflette l'ideologia borghese, che presenta come naturale ciò che è invece appunto prodotto di un'ideologia, mascherando le operazioni svolte. Tuttavia per Belton «even the work that seeks to efface itself, can never disappear» (1985: 63). Egli sostiene che il suono è spesso percepito «*through or in terms of the image*», perciò si manifesta come secondario. Inoltre nel suono si perde l'effetto psicologico di autenticità della riproduzione offerto dall'immagine nonché il suo senso di integrità – infatti non si può scomporre l'immagine in frammenti, cosa che invece è consueta col suono, tradizionalmente diviso nelle tre categorie di dialogo, rumore e musica. Belton concorda con Metz circa la percezione del suono come attributo dell'immagine, da cui deriva l'assenza di oggettività e autenticità; il disturbo arrecato dal doppiaggio, specialmente se in altra lingua, avrebbe origine proprio da questo dato (1985: 64) (in verità è difficile concordare con Belton su quest'ultimo punto, considerando la longevità di tale pratica quantomeno in diversi paesi europei). Secondo l'autore dunque «The sound track corresponds not, like the image track, directly to "objective reality" but rather to a secondary representation of it, i.e., to the images that, in turn, guarantee the objectivity of the sounds» (1985: 66). Il sonoro non duplica la realtà profilmica, bensì crea un mondo immaginario; di conseguenza, duplica il suono dell'immagine, non quello della realtà – se prima la colonna sonora veniva registrata, ora viene costruita. Il principio di fondo del suono filmico è che ogni suono deve significare: infatti la tecnologia si è imposta di eliminare il rumore, che intralcia la trasmissione di suoni portatori di senso. Tuttavia Belton ritiene che il sistema Dolby, pur costituendo all'epoca della stesura del saggio l'ultima tappa del progresso tecnologico in direzione dell'"inaudibilità", rende innaturale il suono, a causa dell'eccessiva pulizia dovuta all'eliminazione di alcune frequenze, quindi a una calcolata alterazione del segnale. Lo studioso rileva anche nella stereofonia in generale un simile problema di artificialità conseguente alla ricerca di realismo. La posizione del microfono è sempre stata subordinata a quella della macchina da presa, perciò l'ampliamento delle proporzioni del quadro ha reso necessario aumentare i canali di diffusione del suono: ma per Belton i sistemi multicanali, «rather than creating a more perfect illusion of depth on the screen, call attention to the

arbitrariness of their choice of sources» (1985: 66-68, 70). Nei film a partire dagli anni '60 Belton nota un uso nuovo del suono, influenzato in parte dalla televisione: spesso si ha un primo piano sonoro associato a un campo lungo visivo (1985: 68-69).

In un articolo che si rivelerà fondamentale nelle riflessioni successive sul suono, Alan Williams si interroga sul problema della concezione del cinema come linguaggio, osservando che i sostenitori di questa idea hanno fatto riferimento esclusivo all'immagine. Secondo l'autore la tesi corrente sarebbe così riassumibile: «Whereas the image track *represents* the space-time accessible to the camera during shooting, the sound track *reproduces* the sonic material selected for inclusion in the film» (1980: 51) (Belton, come abbiamo visto, fa eccezione affermando che il suono è una rappresentazione secondaria all'immagine; 1985: 66). Jean-Louis Baudry è uno degli studiosi che cade in questo equivoco, perché avvalorava la tesi per cui il dispositivo registra il suono "così com'è", pertanto trascurando «the material circumstances of production and reception of the sound» (Williams 1980: 52). Il problema per Williams sta nel sottovalutare il lavoro del dispositivo: un suono è costituito da «the entire volume of air that vibrates during a sonic event», e il fatto che un medesimo suono viene percepito come identico in ogni sua occorrenza a prescindere dall'ambiente è dovuto alla nostra tendenza a ignorare determinate variazioni «in favor of an identity posited as *necessary* to them». In realtà

if sound is a three-dimensional, material event, if the notion of a sound-in-itself independent of its environment is not real but a construction of the subject that listens, then it follows that sound recording can not by definition reproduce "the sounds themselves" – since it is obligated by its nature to render a sound [...] as recorded *from one point* of the space in which and through which the sound exists. [Thus] it is never the literal, original "sound" that is reproduced in recording, but one perspective on it, a *sample*, a *reading* of it (1980: 53).

A questo punto è chiaro che il dispositivo sonoro, così come quello visivo, opera «a significant perceptual work *for us*», restituendoci non l'esatta esperienza fisica ma solo *i segni* della stessa (1980: 58).

Levin affronta lo stesso problema ontologico ma con una prospettiva diversa. Contrariamente a quanto asserito da diversi studiosi (Balázs, Mast, Cavell, Metz, Baudry), Levin contesta la tesi della trasparenza o della neutralità della tecnologia sonora, secondo cui il suono registrato equivarrebbe a quello originale. Malgrado sul piano fenomenologico i due suoni possano sembrare equivalenti, occorre considerare gli effetti ideologici della tecnologia (1984: 56-58). Levin osserva che il suono esiste solo nel tempo e nello spazio, pertanto pur assumendo che la

registrazione sia perfetta la sua riproduzione in un altro spazio darebbe comunque vita a un suono diverso: «In its three-dimensionality sound thus seems more directly tied to the space of its occurrence, its “here and now”, even when the sound is mechanically reproduced». Seguendo Adorno e Eisler, l'autore ritiene che questa specificità spaziale del suono, da cui deriva che «the reproduced sound is *always different* from the original», gli conferisce l'effetto di straniamento (Levin 1984: 62). Ciò comporta inoltre che il suono è percepito come più presente dell'immagine, un fatto che dovrebbe essere sfruttato per palesare «the *mediated* character of the visual register». La mancata perdita dimensionale del sonoro potrebbe essere usata dunque per sottolineare le trasformazioni operate dal processo fotografico: insomma «The acoustic supplement threatens to discredit and even displace the visual domain it was meant to support and strengthen!». Il dominio del dialogo conferma invece la volontà di mantenere l'egemonia del visivo, poiché la fonte è collocata nell'immagine o nella diegesi, mentre nel caso della musica extradiegetica non è così – pertanto essa costituisce una minaccia:

The attempt to unite sound and image necessarily denies the specificity of the acoustic and the visual domains and thereby obscures the incompatibility of these media. [...] By diegeticising [*sic*] the acoustic, especially through the use of synchronous sound, cinema attempts to contain and even repress the supplement which it at the same time requires (1984: 63-64).

Levin ritiene che, seppure in modo diverso dall'immagine, il suono sia comunque «framed» in qualche modo: in effetti il sonoro a lui contemporaneo era normalmente ottico, quindi letteralmente fotografico. Inoltre egli nota che «Because the recording process can only capture sound waves from one point (the site of the microphone) in the space which constitutes that sound, recording is inherently perspectival» (1984: 65-66).

Wood studia, con un approccio più sistematico, come l'introduzione del sonoro ha contribuito alla riorganizzazione del tempo e dello spazio nei primi film sonori hollywoodiani. Elementi fondamentali sono il dominio del campo/controcampo per le scene di dialogo e soprattutto l'introduzione di

a “sonorous landscape” [...]. Thus a *wholly new* mode of spatial delineation dependent upon auditory “indicators” was henceforth activated. The inscription of a scenographic perspective was no longer confined to visually-determined means, but was also reinforced by the creation of an auditory perspective.

Un indicatore fondamentale è il suono d'ambiente o di sfondo; allo stesso tempo grazie a relazioni di volume e di riverberazione era possibile contribuire, insieme col visivo, alla creazione di profondità, anche simulando condizioni di ascolto binaurale (1984: 20-21). Wood rileva poi la presenza di due posizioni teoriche nei discorsi dell'epoca: la prima prevede il mantenimento di una «*position of audition*» coerente con la scala visiva, la seconda – prevalente – che invece privilegia l'intelligibilità della voce a prescindere, malgrado la conseguente riduzione di realismo che contraddice l'elogio dell'incremento di realismo permesso dal sonoro. Dunque il livello della voce solitamente restava quello del «master sound take», a meno che non si volesse enfatizzare «the voice's dramatic effect». Ciò vale pure per la riverberazione ma in misura minore, perché essa poteva aiutare a distinguere ambienti diversi. Insomma il livello e la riverberazione non erano costantemente coerenti con la prospettiva visiva ma «their subtle presence within every scene did create a more precise auditory delimitation of the scenographic space» (1984: 22-24).

*La Voix au cinéma* di Chion (ed. or. 1982) è come accennato un testo molto influente nell'ambito degli studi sul suono in cui l'autore analizza l'uso della voce nel cinema per lo più con l'approccio psicoanalitico all'epoca consueto. Innanzitutto Chion propone per la prima volta la sua idea provocatoria: «there is no soundtrack» (1999: 3; cfr. 2001: 45; 2009: 227-230). Egli spiega che «A film's aural elements are not received as an autonomous unit. They are immediately analyzed and distributed in the spectator's perceptual apparatus according to the relation each bears to what the spectator sees at the time»; il/la percipiente opera dunque un «*instantaneous perceptual triage*» governato dall'immagine, nel senso che i suoni sono interpretati in base alla loro relazione col visivo (1999: 3). Secondo Chion grazie alla stereofonia «the voice may be contained in a space that is no longer defined solely in visual terms, but also auditory ones», poiché è possibile costruire uno spazio uditivo tridimensionale. Il Dolby permette anche una migliore resa del silenzio e quindi una maggiore separazione dei singoli suoni, ora più palpabili (1999: 165-168). Come vedremo, Chion tornerà su questi argomenti in testi futuri.

## 2.4 – Anni '90: Stati Uniti

Il libro *Sound Theory/Sound Practice* curato da Rick Altman (1992), in cui si riscontra un approccio da *cultural studies*, è fondamentale nello sviluppo degli studi sul suono contemporanei ma non considera ancora il sonoro digitale essendo stato pubblicato proprio lo stesso anno dell'avvento del Dolby Digital, ossia dell'introduzione su larga scala dell'audio digitale nel cinema. Il libro contiene delle pagine rilevanti anche circa lo spazio sonoro, su cui incidono le tesi di Williams (1980): infatti un concetto fondamentale proposto da Altman è quello di *spatial signature*

(firma spaziale), nel senso che ogni incisione è «signed [...] with the mark of the particular circumstances in which it was heard» (1992a: 24). Una descrizione più dettagliata della firma spaziale fornita dallo stesso Altman è la seguente:

The testimony provided by every sound as to the spatial circumstances of its production. In order to recognize a given sound's spatial signature we must be sensitive to multiple aspects of the sound's physical nature, including its ratio of direct to indirect sound, its frequency range (as compared to culturally conditioned expectations), and its volume (with respect to perceived distance). Spatial signature can of course be manipulated by postproduction techniques such as gating, equalizing, filtering, or adding reverb (1992a: 252).

In un altro intervento Altman discute di una questione annosa per Hollywood e dei numerosi problemi che comporta: «what relation should obtain between image scale and sound scale?» (1992b: 46). Ci sono stati nella storia tre approcci principali alla localizzazione del suono: 1- manipolazione nel contesto dell'esibizione (1927-1931); 2- manipolazione durante la produzione, specialmente nella scelta e nel posizionamento dei microfoni e nel controllo dei livelli durante il montaggio (1929-oggi); 3- sviluppo della tecnologia multicanale, da cui poi la stereofonia (1930-oggi) (1992b: 47). L'esistenza del primo approccio è comprensibile in quanto si tratta di una continuità con la prassi precedente, relativa ai film muti, in cui c'erano interventi dal vivo durante la proiezione. Quando si impose il parlato fu aggiunto un secondo altoparlante posizionato sopra lo schermo e adibito al dialogo, mentre l'altro – nella buca d'orchestra – era destinato alla musica; l'invio del segnale all'uno o all'altro altoparlante era gestito manualmente dal proiezionista durante lo spettacolo. Questo diede vita a ipotesi di stereofonia, non realizzate. Tra gli ingegneri del suono, come abbiamo visto, si riscontra molta attenzione verso la prospettiva sonora: di norma essi prendevano come modello la natura e l'unità del corpo umano, perciò era necessario che ci fosse corrispondenza tra la prospettiva visiva e quella sonora – il che poteva essere facilmente garantito dall'uso di un singolo microfono posizionato vicino alla linea di visione della macchina da presa. In pratica comunque almeno fino al 1945 la norma fu un'altra (come dimostra già *Applause* [USA 1929]), e cioè la discrepanza tra il sonoro – dalla prospettiva uniforme – e il visivo – dalla prospettiva composita. Ciò è stato favorito dall'utilizzo, a partire dagli anni '30, di un singolo microfono mobile che veniva spostato dal microfonista secondo necessità mantenendo la stessa distanza da ciascun parlante. Questa prassi, insieme all'uso di altre tecniche (per aggiungere riverberazione, per equalizzare la voce, ecc.) «assured Hollywood both the economic benefits and the requisite control associated with a system permitting the *construction* of a sound track rather

than the direct *recording* of already constructed sounds»; lo scopo era quello di fornire «a clean, clear, continuous sound record, oblivious to image scale but attuned to dialogue intelligibility, story continuity, and freedom of action». Alla fine degli anni '30, sia nella teoria sia nella pratica questa sensibilità divenne dominante: i tecnici puntavano ormai alla produzione di una colonna dal volume omogeneo e con una prospettiva da primo piano, e infatti erano encomiati «for the clarity of their sound tracks [...] rather than for their spatial realism» (1992b: 47-54).

Altman tiene a notare che i fonici cinematografici provenivano per la maggior parte dalla radio, perciò erano condizionati dalla prassi relativa a questo mezzo, in cui non si poneva il problema dell'adesione a una prospettiva visiva. Inoltre c'erano anche altri condizionamenti: per esempio, sebbene nel caso del dialogo si evitasse la riverberazione in nome dell'intelligibilità, nel caso della musica esso era desiderato per ottenere conformità con le ampie sale da concerto (1992b: 54; sul problema della riverberazione cfr. Kerins 2011: 1483-1497).

Secondo l'autore ci sono complessivamente tre fattori che determinano le scelte in termini di livelli del sonoro nella Hollywood degli anni '30. Il primo è il criterio dell'intelligibilità del dialogo: questo poneva problemi teorici, perché collideva col desiderio di rispettare il punto di visione, ma la questione come abbiamo notato fu presto superata, a conferma della maggiore importanza rivolta all'esigenza narrativa piuttosto che al realismo percettivo. Il secondo, diretta conseguenza del primo, è proprio il criterio del realismo, che aveva acquisito una connotazione diversa rispetto ai primi anni del sonoro. Laddove inizialmente «the reality standard constantly held up to the cinema sound track was daily life in the real world», più avanti il realismo è stato assimilato alla teatralità: «To call for intelligibility [...] is thus to call for adherence to the theater as code of reality». Il terzo fattore concerne «the subject placement implied by the dominant sound model». Altman distingue in merito tra *ascoltatore esterno* e *ascoltatore interno*: mentre all'inizio si favoriva il primo, in seguito il favore passò al secondo. Infatti «we are asked not to hear, but to identify with someone who will hear for us. [...] [T]his technique locates us in a very specific place – the body of the character who hears for us»; la conseguenza fondamentale è che il «point-of-audition» (qui inteso come soggettiva sonora, sul modello di «point-of-view shot») «constitutes the perfect interpellation, for it inserts us into the narrative at the very intersection of two spaces which the image alone is incapable of linking, thus giving us the sensation of controlling the relationship between those spaces» (Altman 1992b: 58-61).

Altman spiega che, come «perspective images are always made for us», così il suono in primo piano senza riverberazione è «sound spoken *toward* me rather than away from me [...], sound that is produced *for me*». Ciò dimostrerebbe che «while the image is carefully avoiding signs of discursivity in order better to disguise Hollywood's underlying discourse, the sound track overtly

adopts the discursive approach of low-reverb sound in order better to draw us into a fabricated nature» (1992b: 61-62; cfr. il microfono simbolico di Chion 2009: 292-293). Il ragionamento di Altman è difficilmente applicabile ad altri aspetti del lavoro sonoro, per esempio il montaggio: in questo caso la prassi classica è assimilabile a quella del piano visivo, nel senso che si fa il possibile per nascondere i numerosi artifici affinché l'effetto sia fluido e "naturale". Questo discorso, come vedremo, si complica relativamente alla spazializzazione, un aspetto che porta tuttora con sé l'annosa questione del rapporto tra realismo e artificio, ovvero tra occultamento e celebrazione del dispositivo.

Un altro intervento importante di Altman è dedicato a *Quarto potere* (*Citizen Kane*, USA 1940) e alla questione del *deep-focus sound*, ossia il correlativo sonoro della *deep-focus photography*, un concetto utile perché pertiene alla costruzione dello spazio sonoro. L'autore analizza il *sound design* del film evidenziando la discrepanza tra la profondità della focale visiva – ingente – e quella della "focale sonora" – esigua: infatti *Quarto potere* presenta in prevalenza il consueto trattamento del dialogo, in primo piano a prescindere dalla distanza visiva (Altman 1996: 96-97). D'altronde il panfocus palesa l'artificialità delle convenzioni dello spazio sonoro (1996: 107). Inoltre il film attingerebbe allo stile radiofonico coevo, che comprende il rifiuto di una costruzione realistica dello spazio sonoro a favore di una riduzione del medesimo «to a foreground/background dialectic, with spatial references developed only to the extent that they can serve a discursive purpose» (1996: 103). Secondo Altman si è concordi nel ritenere il panfocus problematico per scene di dialogo, perché ci si ritroverebbe a mantenere lo stesso livello di dialogo per attori che sono contemporaneamente a distanze diverse dalla macchina. Un simile problema è evitato «by cutting to a closer shot of a distant speaker, by restricting deep-focus shots to a single speaker, or by using point-of-audition sound» (1996: 104).

Lastra fonda la sua riflessione su questa premessa: «the relationship of original sound to represented sound is a (perhaps *the*) fundamental issue» (1992: 66). Egli condivide le posizioni di Williams (1980) secondo cui il suono risulta dal volume di aria in vibrazione nella sua interezza nello spazio della produzione. Tra gli studiosi, secondo Lastra, ci sono stati invece equivoci proprio sulla definizione di "suono". Per esempio Metz usa il termine non tanto per riferirsi al dato sensoriale, quanto a quello semiologico o sociale: egli quindi non descrive un *evento* (irripetibile), bensì una *struttura* (ripetibile e intelligibile) (Lastra 1992: 67-68)<sup>35</sup>. Lastra si sofferma poi sul problema del suono originale a partire dalle affermazioni di Levin (1984), il quale parla di *inganno* a proposito dell'impressione per cui un suono è recepito allo stesso modo a prescindere

<sup>35</sup> «I wanted to show that the perceptual object is a constructed unity, *socially constructed*, and also (to some extent) a linguistic unity» (Metz 1985: 159; cfr. anche *ibid.*: 161, n. 7).

dall'ambiente e di *trasformazione* operata dalla registrazione sonora. Come si chiede Lastra, se la distinzione originale/copia si basa su una trasformazione, che senso ha parlare di inganno? Inoltre se l'identità di un suono è un inganno dovuto alla prassi d'ascolto costruita socialmente, deve esserci «some non-constructed experience, a pure perception, against which to measure this *as a deception*». Il problema per l'autore è proprio che tutte le prassi d'ascolto sono costruite, perciò non c'è ragione di prenderne una come riferimento assoluto per le altre. La relazione causale tra originale e registrazione non dovrebbe essere privilegiata sul piano teoretico, anche perché persino il concetto di “originale” è problematico a prescindere dall'atto della registrazione. Il punto nodale del pensiero di Lastra è il seguente:

the original sound can be posited as a theoretical absolute only through the dogmatic assertion that one socially sanctioned experience of the sound, its “best” presentation, is somehow “logically” the essential nature of the sound in question. In fact, the primary ideological effect of sound recording might be the creation of the *effect* that there is an “original” independent of its representation (1992: 69-70).

Poiché «[a]ll recordings do transform [...] the sounds they *represent* [*corsivo mio, n.d.a.*]», sebbene non sempre nello stesso modo, ne consegue che «*all* sound recordings are *staged* representations (i.e. socially constructed). [...] The act of representation, by selecting only certain objects or objects in a certain form, or from a certain point of view, *pre-structures* its objects *for the device*» (1992: 72-73, 75).

Lastra rileva due modelli conflittuali nella prassi filmica di Hollywood: quello della “fedeltà” e quello del “telefono”, nella definizione dei quali la dimensione spaziale riveste un ruolo determinante. Il primo modello «stresses the importance of representing a spatio-temporally specific production of sound», il secondo invece «stresses intelligibility or legibility at the expense of specificity, if necessary»: insomma l'opposizione è «uniqueness versus recognizability or event versus structure». Nel secondo modello abbiamo quindi scarsa riverberazione e rumore e un forte senso di direzione, frontalità e soprattutto del concetto chiave di intelligibilità: tutto ciò concorre alla ricezione del dialogo «not simply as an abstract sound, but as a sound with a specific function – a narrative function». Ecco perché si è imposta l'idea che una registrazione è buona se presenta «recognizable speech sounds» a prescindere dalla fedeltà. Nel primo caso invece abbiamo il «point-of-audition (POA) sound [which] represents the experience of hearing within the diegesis, normally the hearing of a character»<sup>36</sup>: insieme al *POV shot* esso comporta «a representation of a space taken

<sup>36</sup> Chion usa il termine con due accezioni, quella soggettiva e quella spaziale, quest'ultima analoga a ciò che Bordwell e Thompson chiamano «prospettiva sonora»; cfr. Chion (2009: 485-486); Bordwell e Thompson (2003: 408-410).

to be “real” within the diegesis» che quindi dà rilievo alla firma spaziale dei suoni (cfr. anche Altman 1992b: 58-62). Infatti una tipica inquadratura hollywoodiana presenta uno spazio enfatizzando «*narratively* important information rather than *perceptually* specific information», da cui consegue un maggior grado di sostituibilità di dette inquadrature rispetto ai *POV shot*. Lastra nota che nel dibattito dell’epoca l’idea della specificità spaziale del suono – non solo del dialogo – era correlata al problema dell’inintelligibilità: detto altrimenti

[a] recording with a high degree of reflected sound, or some other indicator of spatial signature, is linked to sound considered as an *event*, while closely-miked sound, with a relatively “contextless” spatial signature, is linked to sound considered as an intelligibile *structure* – as a signifying element with a larger structure (1992: 76-78).

Lastra sostiene che i criteri di realismo e autenticità cambiano nel tempo, e che ogni nuova modalità di rappresentazione si pone in relazione con le precedenti: egli si chiede quindi quali siano i codici di realtà ai quali il film sonoro fa riferimento, ipotizzando che si tratti di quelli propri del cinema muto. Fin dai primi anni del sonoro ciò che sembrava costituire dei generici esperimenti sul suono era in realtà una specifica ricerca dedicata al dialogo, ovvero alla categoria sonora caratterizzata dall’essere un «*intelligible* conveyor of information». Nel film muto il dialogo era restituito nella forma di didascalia, la quale non è affetta «neither by the space of production nor by the space of reproduction»: anzi, essendo scritta è *leggibile* per definizione, e quindi intelligibile. Lo stile di registrazione che sarebbe diventato normale a Hollywood si pone in continuità con quell’approccio, poiché prevede che il suono sia abbastanza privo di contesto spaziale: «such sound is spatio-temporally *unspecific* and therefore in some sense more like writing than like speech» (1992: 79-82).

Anche Wurtzler si inserisce nel dibattito sulla prospettiva uditiva, osservando che nel cinema sonoro esistono due possibili processi: in uno, il film «posits a single, original, unified audio-visual event that precedes its representation (here, sound and image roughly coincide)»; nell’altro

the disjunction between sound and image perspective is “healed” through the very *discursive* nature of a hierarchical sound [...] practice that privileges intelligibility of the voice». Secondo l’autore sul piano teorico le due strade non sono in opposizione, perché entrambe «are capable of positing a similar “presence” of audio-visual event (1992: 98-99).

Wurtzler riflette poi, con un approccio psicoanalitico, sulla musica extradiegetica nel cinema hollywoodiano per comprendere come la sua presenza apparentemente destabilizzante possa essere

percepita come naturale. La prima possibile spiegazione, a livello storico, è che questa musica si pone in continuità con pratiche antecedenti (film muti con accompagnamento dal vivo), perciò già note e consolidate. La seconda è relativa al livello psicoanalitico: la musica potrebbe offrire un ritorno illusorio «to the prelinguistic plenitude of the Imaginary», così da compensare la frammentarietà delle posizioni immaginarie assunte dal soggetto durante l'esperienza filmica (p.es., il passaggio da una prospettiva sonora coerente con quella visiva a una incoerente) (per riflessioni analoghe di stampo psicoanalitico cfr. Kerins 2011: 3519-3562). La terza, più a monte, concerne l'inabilità del cinema e di altre forme artistiche «to “speak” an autonomous, centered spectating-auditing subject»: piuttosto i film «“speak” a fragmented subject ever in the process of formation, multiply hailed by different discourses, representations and the events they posit». Nel cinema sonoro di Hollywood si riscontra la coesistenza di approcci che nel caso della *popular music* si sono esclusi progressivamente, e cioè la *documentazione* di un evento, la *costruzione* di uno pseudo-evento e lo *smantellamento* della nozione di evento originale. Il primo approccio è riscontrabile nella soggettiva sonora, il secondo nell'alternanza tra la prassi dell'intelligibilità e quella della fedeltà, il terzo nella musica extradiegetica, la quale «*posits no event prior to representation*». Pertanto risulta chiaro che il film hollywoodiano è una *tecnologia della rappresentazione*, che non offre garanzia di un evento originale preesistente e che “emette” solo problematicamente un soggetto autonomo centrato: questi due concetti sono effetti ideologici risultanti da «discourses and practices surrounding representations, rather than by the representations themselves» (Wurtzler 1922: 100-102).

Belton analizza l'avvento del sonoro magnetico negli anni '50, notando che la stereofonia avvolgeva il pubblico passando, insieme alle immagini panoramiche, dalla netta segregazione spaziale della tradizione a una integrazione illusoria in cui si riduce il confine tra spazio reale e spazio finzionale. L'autore osserva comunque come si sia trattato di una «frozen revolution»: se è vero che esso ha apportato miglioramenti significativi nella post-produzione non si può dire altrettanto nel contesto della fruizione, poiché normalmente si è continuato a ricorrere al supporto ottico. Il CinemaScope è stato l'unico sistema con sonoro magnetico stereofonico ad avere un certo successo, tuttavia i costi ingenti per la conversione lo resero un fenomeno limitato a una minoranza di sale; questo non fu comunque un problema per i produttori, perché si trattava di grandi sale che da sole contribuivano in buona parte agli incassi dei film. In ogni caso, dovendo distribuire copie anche col sonoro monofonico ottico si faceva un uso limitato del quarto canale, quello surround, perdendo un elemento di novità. Il successivo avvento di sistemi a 70mm, benché anch'essi a modesta diffusione, contribuì ulteriormente ad associare la stereofonia magnetica con grandi produzioni spettacolari, ovvero «not so much with greater realism as with greater artifice» (1992:

154-158). Questo è un punto cruciale: sebbene alcuni teorici come Bazin e Comolli abbiano ritenuto che l'evoluzione tecnologica del cinema si sia tradotta in maggiore realismo, per i fruitori normalmente l'incremento del realismo viene percepito contemporaneamente come aumento dello spettacolo. Ciò è avvalorato dalla consuetudine di dichiarare palesemente il nuovo artificio, fattore che in teoria dovrebbe collidere con la ricerca di realismo. Insomma si riscontra una costante tensione tra i due poli, quello del realismo – soprattutto nel senso tecnico di fedeltà acustica, ossia di definizione sonora – e quello dello spettacolo e dell'artificio – in relazione all'applicazione artistica del sonoro (Belton 1992: 158-161).

Belton si sofferma sui canali surround del Cinerama, le cui dinamiche di presentazione e di ricezione sono sintomatiche della tensione sopracitata. Sebbene questi canali fossero sfruttati in ossequio al realismo posizionando il pubblico dentro l'azione grazie alla coerenza delle prospettive visiva e sonora, allo stesso tempo essi attiravano l'attenzione su di sé contraddicendo il tradizionale "realismo" hollywoodiano fondato sull'invisibilità («This self-consciousness remained consistent with Cinerama's overall marketing campaign»; 1992: 162). L'approccio del CinemaScope – tranne che nei film della 20th Century Fox – era più tradizionale, con la colonna dialoghi riprodotta da tutti e tre gli altoparlanti anteriori (1992: 161-163).

In conclusione Belton ribadisce che l'idea di realismo nel cinema è fortemente condizionata dalle prassi, il che vale non solo per la stereofonia ma anche per aspetti quali il colore. L'identificazione della stereofonia magnetica con lo spettacolo è dovuta a una differenziazione sia diacronica sia sincronica: sul piano diacronico perché, prima di quel momento, per anni «mono had come to be associated by audiences with realistic representation»; su quello sincronico perché, contemporaneamente al ricorso alla stereofonia nelle sale più prestigiose, in quelle economiche si continuava a usare la monofonia con il tradizionale "realismo". Dunque il sonoro stereo rispetto a quello mono «was not only different but deviant. Mono remained the dominant form of sound reproduction [...], functioning as a norm or "background set" against which stereo emerged as a violation of that norm» (1992: 164-166).

Klimek riflette sullo stile sonoro di trasposizioni cinematografiche di opere di Shakespeare, individuando la presenza di due modelli. Il primo è quello di «Shakespeare-as-playwright», che prevede un'adesione alla prassi teatrale e quindi, sul piano uditivo, una corrispondenza tra la prospettiva visiva e quella sonora – e perciò anche una potenziale presenza di riverberazione; si tratta di un approccio diverso da quello comune nel cinema classico, in cui la forte enfasi sull'intelligibilità del dialogo lo priva della sua firma spaziale (1992: 204-206). Questo modello è seguito – pur se con qualche eccezione – nei film *Enrico V* (*Henry V*, Regno Unito 1945) e *Riccardo III* (*Richard III*, Regno Unito 1955), entrambi diretti da Laurence Olivier, i quali

dimostrano «that cinema can convey a strong sense of the theatrical sound experience while at the same time using peculiarly cinematic potentials to enhance that experience» (Klimek 1992: 206-209). Il secondo modello è quello di «Shakespeare-as-poet», che prevede una maggiore manipolazione del suono dato che si ispira non alla fruizione dello spettatore teatrale bensì a quella del lettore, il quale «is less likely to experience a theatrical “voice in space” than an internalized “voice as poetry”». Questo approccio può prevedere l’eliminazione della firma spaziale in funzione della piena intelligibilità, come da prassi hollywoodiana, ma anche più in generale un trattamento non naturalistico del suono, a seconda delle esigenze espressive. Il modello è seguito in *Macbeth* (*Id.*, USA 1948) e *Falstaff* (*Falstaff (Chimes at Midnight)*, Spagna/Svizzera 1965), entrambi diretti da Orson Welles, in *Macbeth* (*Id.*, Regno Unito/USA 1971, diretto da Roman Polanski) e in *Romeo e Giulietta* (*Romeo and Juliet*, Italia/USA 1968, diretto da Franco Zeffirelli) (Klimek 1992: 210-215).

## 2.5 – Anni 1990-2000: Francia

Jullier enfatizza la rilevanza del sonoro nel cinema postmoderno, a partire significativamente da *Guerre stellari* il quale, benché non sia stato il primo film «con il sistema sonoro Dolby» come asserisce l’autore (2006: 53), è come abbiamo visto quello che ha imposto il Dolby SVA e ha lanciato il Dolby “baby boom”. Jullier non approfondisce l’aspetto spaziale del sonoro ma fa riferimento all’«*effetto bagno* procurato dai film-concerto, che dà allo spettatore la sensazione di fluttuare al centro di un magma, i cui suoni, soprattutto quelli gravi a grande dinamica, toccano direttamente, come l’acqua del bagno – persino in modo più *intrusivo* – tutto il suo corpo» (2006: 56). I vari altoparlanti comportano

una sovra-determinazione della direzione del fuori campo corrispondente alla “quarta parete” (lato dello spettatore), proprio là dove il film classico evita di ancorare i suoni. Questa evoluzione del dispositivo sonoro anticipa dunque una certa evoluzione nella modalità del *bagno di sensazioni* della relazione tra l’immagine e lo spettatore, che mira a far *entrare* più o meno letteralmente il secondo nella prima (2006: 58).

L’autore allude dunque alla riduzione della segregazione spaziale tra spettatore e film, sebbene il riferimento sia alla mera sollecitazione acustica pluridirezionale conforme all’approccio immersivo (che per Jullier concerne anche l’immagine; 2006: 65) piuttosto che alla costruzione dello spazio.

Jullier (2007: 333-334) parla di *miraggio audiovisivo* in riferimento al fatto che la separazione tra *sorgenti reali* (gli altoparlanti) e *sorgenti della finzione* non è percepita come tale dal/dalla percipiente. Anche lui riscontra (basandosi su Branigan 1984) due strategie di costruzione dello spazio sonoro (ciascuna delle quali ha tuttora i propri sostenitori), ossia la *prospettiva naturalistica*, in cui c'è coerenza dei punti di vista e d'ascolto, e il *realismo psicologico*, che «tradisce l'influenza della radio, per esempio nelle pellicole di Woody Allen, e dà la priorità all'intelligibilità – sentiremo sempre i suoni interessanti» (Jullier 2007: 35). A proposito del sonoro multicanale Jullier osserva che gli effetti «eccessivamente laterali» sono usati con parsimonia – almeno nei film-concerto – per evitare che il pubblico sposti l'attenzione dalla finzione agli altoparlanti, e sono riservati ai suoni d'ambiente, «dai contorni sfumati, difficili da identificare ma utili a comunicare delle sensazioni» (2007: 38-39). Questa tesi è conforme a quanto argomentato in Jullier (2006), infatti egli dà più importanza alle “sensazioni” che a una costruzione spaziale.

Il libro *Film, a Sound Art* (ed. or. 2003) è il testo più completo di Chion dedicato al sonoro nel cinema, in cui confluiscono scritti precedenti dell'autore. Secondo Chion nel film sonoro il dispositivo non è più «*the same as the work*»: la banda sonora non è il suono bensì un codice per riprodurlo, mentre la banda visiva è l'immagine. Inoltre il montaggio del suono non produce “inquadrature sonore”: vi è pertanto «a gulf that cannot be crossed between what how something was done and what is perceived», perché nel suono, diversamente dall'immagine, «*the fabrication unit* (the fragment of magnetic tape or virtual digital medium) does not automatically constitute a *perceptual unit* in itself» (2009: 223, 225). È a questo proposito che Chion afferma che non esiste colonna sonora (*bande-son*), ovvero un luogo «where sounds gather and make a unified front» (2009: 227): infatti «the visual shot is a *container* [...]. Sound is first of all *content* or “containable”, with no actual frame. What is designated by the word *image* in cinema is not the contained but the container: the frame» (2009: 226; cfr. 2001: 45, 70-71). Per indicare il dato tecnico dell'insieme dei suoni di un film l'autore preferisce parlare di *pista sonora* (in francese *piste sonore* in luogo di *bande-son*) (2009: 228). Questa tesi permetterebbe di contrastare le teorie che ritengono che immagine e suono siano elementi complementari e di sostenere piuttosto un modello «where a radical break is produced between the technical level and the levels of perception, discourse, effects, and theory», e in cui inoltre sia sostenuta l'idea di “suoni inauditi”, cioè quelli suggeriti dall'immagine ma assenti *de facto* – fenomeno possibile grazie all'“effetto-palinsteso” (2009: 229-230). A proposito dello spazio dei suoni Chion afferma anche che «[t]he notion of an auditory field in cinema is itself ambiguous [as it is] in everyday life since the location of a sound is never really clear. [We can] describe the place of [a] sound as a sphere with a central core and fuzzy outer contours» (2009: 248).

Chion sostiene che l'immagine possa attuare una *calamitazione spaziale* del suono, il che però avviene solo se il suono «conserva [...] una certa fissità spaziale» (2001: 73) e non è posto in un altoparlante lontano dallo schermo, altrimenti l'impressione di realtà si tramuta in eccesso di realismo, il quale può comportare l'*effetto quinta* (“in-the-wings effect”): esso «si produce quando un suono legato a una causa suscettibile di entrare in campo, o che ne è appena uscita, si attarda su uno degli altoparlanti laterali»; ne consegue una sensazione «perturbante rispetto alle nostre attitudini» ossia che la fonte sia presente realmente nello spazio (2001: 85, 87; cfr. 2009: 133-135). L'autore distingue inoltre la *profondità* dalla *distanza*, asserendo che nel primo caso ci si riferisce alle caratteristiche acustiche derivanti dalla lontananza della fonte dal microfono, mentre il secondo denota una «localizzazione mentale» basata su inferenze visive o determinate dal racconto (2001: 73-74).

Circa il luogo di provenienza dei suoni Chion propone un “tri-cerchio”, ossia un modello che presenta tre zone: una visualizzata, quella «in»; due acusmatiche, quelle «hors-champ» e «off» (rispettivamente “off” e “over” nella terminologia corrente italiana e inglese) (2009: 249-251). Il termine “acusmatico”, mutuato da Schaeffer, è opposto a “visualizzato” (2001: 74; cfr. 2009: 465-466). Le tre zone danno origine a tre confini, l'attraversamento dei quali viene sfruttato al cinema: il confine *in-fuori campo* è quello usato più di frequente (fonti sonore che entrano ed escono dal campo visivo); quello *in-off* è comune solo con la musica (la quale «nel cinema è l'“attraversamuri” per eccellenza», nonché «un ammorbidente di spazio e tempo»; 2001: 83-84); quello *off-fuori campo* è il più misterioso perché è il meno tangibile (2009: 259-260; cfr. 2001: 76-88). L'autore parla anche di *suono d'ambiente*, in riferimento a quello «that envelopes a scene, inhabits the space without raising the question of the location of its specific source(s) in the image» (2009: 467); di *suono interno*, ovvero percepito solo dal personaggio perché fisiologico (*oggettivo*) oppure mentale (*soggettivo*) (2009: 479; cfr. Bordwell e Thompson 2003: 406); infine di *suono «on the air»*, «when [its] diegetic source is shown to be some electrical mode of transmission [...] that allows [it] to go beyond the so-called natural mechanical laws of sound propagation and travel freely in space yet still remain anchored in the real time of the scene» (Chion 2009: 482; cfr. 2001: 76-79). Si noti che Chion definisce “extradiegetico” il «son off»: in realtà egli si riferisce genericamente a un suono che proviene da un altro spazio-tempo rispetto all'azione mostrata e non solamente a suoni esterni alla diegesi, ovvero al mondo narrativo (si tratta dunque di suono over, sebbene si possa usare questo termine anche per indicare la voce interna contestuale alla scena; torneremo sulla questione nel cap. 4). A proposito del suono fuori campo, Chion ne distingue uno *relativo* (o *mentale*), cioè determinato in relazione alla visibilità della fonte sullo schermo, e uno *assoluto* (o *reale*), permesso dai canali surround; un caso di questo secondo tipo è il *fuori campo spazzatura*, in cui si inserisce

nel surround – come prolungamento dei canali frontali – dei rumori che derivano da un evento in campo quale una catastrofe, una caduta, un'esplosione *et sim.* (2001: 84-86; cfr. 2009: 481-482). Inoltre l'autore differenzia il fuori campo *attivo* da quello *passivo*: nel primo caso «il suono acusmatico pone delle domande», creando attenzione e curiosità; nel secondo

il suono crea un ambiente che avvolge l'immagine e la stabilizza, senza [...] suscitare il desiderio di andare a vedere altrove o di anticipare la visione della sua sorgente, dunque di cambiare il punto di vista. [...] Il fuori campo passivo è costituito essenzialmente da *suoni-territorio* ed *elementi di scenografia sonora* (2001: 87; cfr. 2009: 481-482, 476).

Il suono multicanale, specificamente il Dolby Stereo, ha sviluppato il fuori campo passivo piuttosto che quello attivo, forse per timore dell'eccesso di realismo e dell'effetto quinta (2001: 87). In questo contesto Chion propone anche il concetto di *estensione*, in riferimento allo «spazio concreto più o meno largo e aperto che i suoni evocano e fanno sentire intorno al campo, e anche, all'interno del campo, intorno ai personaggi»; i due estremi dell'estensione sono quella *nulla*, «quando l'universo sonoro è limitato ai rumori sentiti [solo] da un dato personaggio», e quella *vasta*; «Il suono Dolby Stereo [...] incoraggia il gioco dell'estensione» (2001: 89; cfr. 2009: 477).

Un'altra idea peculiare di Chion è quella dell'assenza di un *microfono simbolico* assimilabile alla *macchina da presa simbolica*: detto altrimenti «[t]here is no symbolic mediator of our hearing embodied in the cinematic apparatus», da cui consegue l'impossibilità dell'identificazione primaria attraverso lo stimolo uditivo (2009: 292-293; cfr. 2001: 94-95). L'autore inoltre descrive due accezioni del concetto di *punto di ascolto*: quella *spaziale*, cioè «da dove sento» (2001: 91; «[t]he spatial position from which we can say we hear a sound at a given distance from us»; 2009: 485); quella *soggettiva*, in relazione all'entità «through whose ears it is suggested that we are hearing a sound». L'entità in questione può essere un personaggio, ma anche un oggetto o qualcosa di astratto e metafisico: Chion a questo proposito elenca dodici possibili orecchi, tra i quali quello tecnico (che è, ancora una volta, «not localizable or symbolizable»), quello mediato da un dispositivo diegetico, quello di un morto o di un nascituro, ecc. (2009: 299-307; cfr. 2001: 92). Una simile distinzione tra i due tipi di punto d'ascolto era stata già operata da Branigan in termini di *aural POV* – «where we hear from the character's point in space» – e di *aural perception* – «where we hear something only the character hears» (1999: 94). Circa la dimensione spaziale della musica, per solito meno considerata rispetto a quella temporale, Chion propone un'idea che ricorda quella di Gorbman (1987: 25): «[t]hrough its sonorities and orchestral range, and the horizon suggested by its melodies,

music often helps restore the impression of “vastness” that can become lost through the fragmentation of editing and that realist sound normally cannot achieve» (Chion 2009: 409).

## 2.6 – Dal 2000 a oggi

Nell’attuale millennio c’è stata una proliferazione degli studi sul suono, nel cinema e in generale tra le scienze umane, con una maggiore varietà di approcci. Si riscontra pertanto anche una maggiore quantità di studi che considerano la dimensione spaziale del sonoro filmico, sia in termini teoretici generali sia più in particolare analizzando determinati insiemi di film.

### 2.6.2 – Studi anglofoni

Altman, Jones e Tatroe (2000) analizzano l’«Hollywood multiplane sound system» rilevando come dopo un iniziale periodo di discontinuità sonora – dovuta in sostanza alla corrispondenza tra stacchi visivi e stacchi sonori – si è poi stabilita la prassi di rendere il sonoro continuo «thus unifying space and time rather than reinforcing the film’s image segmentaiton» (2000: 347). Si riscontra

the virtual omnipresence of low-level atmospheric sound effects» che hanno una «semi-sync nature [...]. Whereas synchronized effects [...] separate performance space from audience space, atmospheric effects [...] envelop the audience in a generalized sound ambience, with the sound seemingly coming from all around, thus sonically “enrolling” spectators in the film’s space (2000: 352).

Mentre gli “ambienti” creano un generico spazio tridimensionale ancorando la diegesi e fornendo continuità, il dialogo colloca il/la percipiente in una posizione frontale, portando il parlante in primo piano e facendo avanzare la narrazione. La musica favorisce questo approccio duale perché come gli ambienti essa rimane continuamente presente: pertanto in assenza di dialogo essa può godere di volume alto, mentre «In its low-volume mode (when it plays under dialogue or narratively important effects), music joins atmospheric effects in assuring continuity and a generalized sense of space» (2000: 352-353). Significativamente gli autori parlano di «“surround sound” impact» a proposito dell’intrinseca facoltà dei suoni d’ambiente di avviluppare il/la percipiente (2000: 358, n. 8).

Lastra (2000) approfondisce la fondamentale questione dei due modelli di ripresa sonora (*evento* e *struttura*) (1992), sia sul piano teoretico sia su quello pratico. Il primo modello, «heir to

metaphors of human simulation and described in terms of perceptual fidelity, emphasizes the literal duplication of a real and embodied (but invisible) auditor's experience of an acoustic event», da cui deriva un senso di presenza e immediatezza poiché si concepisce «a fully textured and self-sufficient original sound completely autonomous of the act of recording»; dunque si tratta di *ri-presentare* un evento originale “pro-phonographic” (2000: 181, 88, 241 n. 70). Il secondo modello ha radici in «the metaphor of writing, and emphasizes the mediacy, constructedness, and derived character of representation», da cui derivano i concetti di «repetition, legibility, and nonpresence» poiché non si rispetta la natura dell'originale; lo scopo non è tanto la *duplicazione* quanto la *produzione*. Pertanto nel primo modello suono e spazio sono omogenei e privi di gerarchie interne, unendo gli spazi di ricezione e di rappresentazione ossia collocando chi ascolta nello spazio profilmico dando l'impressione che non vi sia mediazione; la percezione è concepita come «ideally sensitive and “passive”». All'opposto, il secondo modello assume che suono e spazio siano eterogenei e strutturati secondo delle gerarchie, e la percezione è concepita come «directed and goal-oriented». Lo spazio sonoro è pertanto costruito «to produce a particular effect (often “realism”), and to be plausibly, rather than literally, coherent» (2000: 181-182). Lastra esamina quindi come l'argomento è stato affrontato nella letteratura del periodo di transizione al sonoro, confermando che la percezione e quindi la rappresentazione sono concepiti in modo variabile; comunque si impose il secondo modello che può essere paragonato a quanto Lindsley Lane chiama «the camera's “omniscient eye”» (Lastra 2000: 194). Il primo modello, riscontrabile nei film Vitaphone, fu invece abbandonato al cinema mentre rimase tipico delle incisioni orchestrali: in questo caso si favoriva una «embodied location offering a perceptually specific (if conventionally chosen) “vantage point”». Questa “disc aesthetic” è dedicata «not to represented *sounds*, but to represented *hearing*» (2000: 194-196).

Un'altra coppia concettuale imprescindibile è quella di *foreground* e *background*: Lastra rileva che i fonici separavano questi due livelli per ragioni ideologiche e pratiche, assumendo che il dialogo dovesse essere in primo piano e intelligibile mentre nello sfondo si potevano aggiungere «“smoothness” and realism», per favorire un senso di naturalezza e continuità. Perciò i suoni di sfondo dovevano essere *riconoscibili* piuttosto che strettamente realistici. Il risultato non somiglia tanto alla prospettiva rinascimentale quanto al cubismo o al fotomontaggio, pur mantenendo l'impressione di unità: «an apparently “monocular” (monaural) space is actually composed of a multiplicity of sound “shots” – different points of audition – combined on a homogeneizing (or at least shared) substrate or background», rispondendo a «hierarchies of “dramatic” relationship» (Lastra 2000: 206-207). Il concetto di fedeltà comunque non è mai sparito dal cinema classico, ma è stato ridotto a «the *effect* of a pure listening, that is, of a wholly prior event upon which we simply

“eavesdrop”». Ciò comporta il rispetto della prospettiva e un uso meno ovvio delle gerarchie per dare l'impressione di un punto d'ascolto situato nello spazio diegetico (2000: 212-3).

Secondo Lastra (2008) l'affermazione del *sound design* e l'imposizione del surround negli anni '70 ha denotato il rafforzamento «of cinema's recurring aspiration to provide a total artwork of full sensory immersion». Si poteva creare «complex and three-dimensional sonic environments», più ricchi e realistici, in cui immergere il pubblico come se fosse parte di quel mondo oppure con cui rendere meglio la soggettività dei personaggi: ciò «reopened the question of whether the cinema was a reproduction of real perceptions or rather a more obviously conventional system of rhetorically structured experiences designed to produce a particular narrative effect». Come in periodi precedenti, l'avvento di questa nuova “tecnologia sensoriale” ha comportato una riduzione dei confini «between human and machine, and between perception and representation». Lastra si propone di analizzare l'intreccio tra la prassi del *sound design*, ovvero «what was actually done with those sensory technologies», e la speculazione teorica «about the implications of particular forms of sensory manipulation». A proposito del secondo aspetto l'autore chiarisce che occorre definire i concetti di suono e di ascolto «in specific cultural and social terms» (2008: 124-125).

L'immersione sensoriale filmica è paragonata da Lastra al teatro di Wagner, compositore secondo Adorno dell'intossicazione, della fantasmagoria, e il cui stile drammatico «responds to the shattering of the human senses into discrete and disconnected zones, by providing an arena within which they appear to reconnect» (sul rapporto tra il cinema e Wagner cfr. Bordwell 1980). Queste caratteristiche sono riscontrabili palesemente in *Apocalypse Now* (Lastra 2008: 130-131; per un commento più specifico su questo film cfr. *ibid.*: 136-137). Relativamente invece alla dicotomia presentazione/rappresentazione (2008: 132-135), lo studioso ribadisce le tesi già esposte più ampiamente in testi precedenti (1992; 2000).

Un nuovo e originale metodo per l'analisi del punto d'ascolto negli audiovisivi è proposto da Maasø, benché egli si soffermi esclusivamente sulla voce, in quanto elemento sonoro principale nella comunicazione umana. Ci sono tre fattori acustici che permettono di valutare gli *aspetti spaziali* della voce: l'intensità, le caratteristiche della frequenza, la relazione tra suono diretto e riflesso. L'intensità costituisce il dato principale per valutare la distanza della voce nella vita quotidiana; le caratteristiche della frequenza e la riflessione del suono influiscono perché più aumenta la distanza, più si riducono le frequenze basse e aumenta il suono riflesso, condizionato dall'ambiente in cui avviene la propagazione del suono. Maasø lamenta la carenza di categorie analitiche relative al suono negli audiovisivi, ancor più grave considerando «the wealth of analytical tools, concepts, and theories for dealing with visual space and the mise-en-scène of the human body» (2008: 37-38).

L'autore utilizza perciò la nozione di prossemica, elaborata dall'antropologo Edward T. Hall negli anni '60 per enfatizzare «the importance of spatial proximity and physical relationships between subjects in interaction and communication»; Hall aveva individuato quattro aree prossemiche: intima (entro i 45 cm), personale (45-120 cm), sociale (1,2-3,7 m) e pubblica (oltre i 3,7 m) (Maasø 2008: 40). Analizzando il suono mediato dalla tecnologia occorre aggiungere altri aspetti relativi al dispositivo, perciò Maasø propone tre livelli di analisi: la distanza vocale, la distanza uditiva voluta («intended earshot»), la prospettiva del microfono. Il primo livello si riferisce a «the way the voice in itself signifies a proxemic relationship, as it would be in interpersonal communication», indipendentemente dal fatto di essere mediata; il secondo è relativo al volume della voce così come è gestito dal dispositivo, a prescindere dal volume del suono originale; il terzo comprende tutti i fattori determinati dalla scelta del microfono e del suo posizionamento (2008: 41). Partendo dal ben noto modello di analisi visiva, Maasø elabora una scala che comprende *primitissimo piano*, *primo piano*, *mezzo primo piano*, *campo medio*, *campo lungo* e *campo lunghissimo* (2008: 43). Maasø introduce infine un altro concetto:

the experience of the vocal proxemics of mediated voices is a result of what may be called a *schizophonic average* [*corsivo mio, n.d.a.*] of all three levels [...]. In the media these three levels will both interact in providing a *blend* of the three and, at the same time, allow each level to play an *independent* role.

Per esempio, la diffusione negli anni '80 del radiomicrofono a discapito del *boom* comporta in termini di media schizofonica una maggiore intimità nella prossemica vocale degli audiovisivi (2008: 47-48).

Tra gli studiosi di sonoro filmico più significativi dell'attuale millennio vi sono Sergi, Beck, Wright, Whittington e soprattutto Kerins. Secondo Sergi (1999) esistono nel contesto della proiezione due colonne sonore parallele a quella del film: la colonna sonora "strutturale", ovvero determinata dalla struttura fisica dell'ambiente, e quella del pubblico, che scaturisce dai suoni emessi dagli spettatori. Esse influenzano inevitabilmente la fruizione del film, e la seconda in particolare è difficilmente controllabile. Ne consegue un'ulteriore conferma del coinvolgimento attivo del pubblico, nonché un paradosso: se da una parte i cineasti vogliono offrire un "parco giochi" coinvolgente, dall'altra si preoccupano di standardizzare le condizioni di fruizione.

Circa quello che chiama «aesthetics of surround» Sergi (2004) spiega che «the "balance" of sound between front and rear does not *have* to be heavily weighted towards the front», come conferma l'allora recente introduzione del Dolby Digital Surround EX (e oggi i nuovi sistemi

immersivi). Si tratta perciò di «a choice, not a necessity», corrispondente a quanto l'autore definisce «one-wall narrative approach» (2004: 146).

Beck dedica la sua tesi di dottorato al sonoro del cinema statunitense degli anni 1967-1979, offrendo numerosi spunti di riflessione teorica. Come ho già accennato nel capitolo 1, Beck osserva che i modi di rappresentazione della stereofonia degli anni '50 hanno origine nell'esperimento presso i laboratori Bell del 1937 (cfr. Maxfield 1938), e si possono così descrivere: 1) il modello *binaurale* (tipico del Cinerama): riproduce un *punto d'ascolto* specifico, coerente col punto di vista, costruendo «an idealized spectatorial position»; 2) il modello *stereofonico* (tipico del CinemaScope): riproduce un *evento* profilmico, con accuratezza spaziale ancorché non necessariamente con coerenza prospettica; 3) il modello *costruito* (tipico del Perspecta): ricostruisce un *soundscape simulato* in post-produzione; quest'ultimo fu quello che prevalse (Beck 2003: 78-79). A proposito del Dolby SVA, Beck ritiene che esso fu «a giant step *backwards* in the history of film sound evolution» a causa sia dei suoi limiti tecnici sia del contesto produttivo in cui emerse, mantenendo in termini di spazializzazione il modello rappresentativo monofonico (2003: 132, 497). In particolare Beck nota che il Dolby SVA impose che il dialogo fosse missato

into the central channel both to ensure comprehension and to avoid phasing. Effects could be positioned anywhere in the left-to-right space of the screen, but moving effects would be carefully monitored to ensure that their acoustic motion matched the motion on screen. Music, however, provided few problems because it was rarely anchored to an on-screen image. But when it was, care was given to make sure that sounds did not drift in the stereoscape.

Inoltre il Dolby SVA impediva di applicare divergenza al segnale come si soleva fare con la stereofonia magnetica per omogeneizzare il più possibile la resa spaziale tra le diverse posizioni del pubblico nella sala; di conseguenza col Dolby SVA «traveling sounds, or sounds in the side channels, were much more noticeable than in a magnetic print with divergence. Therefore effects tended to either remain centered on the screen or to drift off into a side channel and the surrounds»; Beck sostiene che simili effetti si dimostrarono inopportuni tranne che nel genere fantascientifico, dati i suoi «flexible codes of realism» (2003: 46-49). I canali con le basse frequenze del “baby boom” secondo l'autore, similmente al surround del Dolby SVA, erano usati per ottenere un effetto spettacolare piuttosto che per una resa accurata dell'intero spettro acustico (2003: 51). Il Dolby Stereo in generale impose la prassi di «filling the soundtrack»: in sintesi

Few filmmakers have recognized the power of allowing the multi-channel soundtrack to expand and contract according to the demands of the story. Instead most prefer to fill the soundscape with sounds and to constantly surround the audience with acoustic information even if it has no relationship to the story at hand. With the transition to 5.1 digital sound in the mid 1990s, many filmmakers felt the need to utilize all of the speakers all of the time (2003: 498).

In un testo successivo, Beck riscontra la problematicità degli studi che riguardano film in Dolby Stereo, anzitutto perché occorre distinguere la versione ottica da quella magnetica. L'autore contesta alcune affermazioni di Chion: questi nota giustamente che nel Dolby SVA la separazione tra i canali è minore rispetto ai sistemi discreti, ma tralascia il fatto che la matrice comporta l'aumento di segnale surround e un movimento peculiare dei suoni. Secondo Beck «Dolby Stereo has the unique ability to construct acoustic *layers* on a soundtrack, and to render the sonic and spatial details of each layer with extreme accuracy». Eppure inizialmente queste potenzialità non sono state inizialmente sfruttate a causa dell'incompatibilità con convenzioni visive nella costruzione dello spazio (il controcampo, la regola dei 180°) (Beck 2008: 71-73). Secondo Beck il Dolby Stereo «expands the two-and-a-half-dimensional space of the screen (left-right, up-down, but only depth “into” the screen) to a full three dimensions by extending the diegetic space into the space of the theater». Ne consegue che (1) «the listener is surrounded by the sound field while several discrete layers of sound are heard simultaneously», e (2) i suoni possono attirare l'attenzione «through their spatial specificity», da cui risulta l'effetto di un *establishing shot*, un ambiente sonoro stabile mentre si destabilizzano la centralità narrativa del quadro e volendo anche delle regole di montaggio. Ne potrebbe risultare anche una riduzione del confine tra le componenti del sonoro – cioè voce, rumore e musica – tanto da rendere necessario il concepimento di una quarta componente, la «spatial ambience» (2008: 76-78). Beck analizza la sequenza di *Il silenzio degli innocenti* (*The Silence of the Lambs*, USA 1991) in cui Clarice scova Buffalo Bill, rilevando un uso eccellente delle potenzialità spaziali del Dolby SVA (2008: 78-83), tesi che sarà in parte confutata da Kerins (2011: 1176-1190).

Wright si è occupato di spazio sonoro e stereofonia a partire dalla sua tesi di laurea 2005 e poi di dottorato 2011. Nel primo testo egli affronta specificamente la transizione al Dolby Stereo trattando l'aspetto spaziale. Egli nota la ricorrenza del concetto di immersione connessa tradizionalmente al suono in sé, fin da quello monofonico (2005: 2), ma che è cruciale nel Dolby Stereo: infatti il sistema «reconfigures the relationship between audience member and diegesis. Whereas monaural film sound was directed *at* the audience, Dolby sound *surrounds* the audience by utilizing not only frontal loudspeakers, but also a U-shaped array of loudspeakers beside and behind

them», da cui un effetto tridimensionale (2005: 3; Wright però sembra non considerare che questa non è una novità del Dolby Stereo). L'autore, seguendo Lastra (2000) nota la divisione degli ingegneri negli anni '30 tra due modelli di stereofonia: quello di *fedeltà percettiva* (il sonoro come facsimile di un evento reale) e quello dell'*intelligibilità*. In ogni caso si può notare una nozione di realismo idealizzata anche nei sostenitori della prima posizione, che ammettevano la possibilità di manipolare la riproduzione (Wright 2005: 15-17). Il Fantasound segnò una tappa importante perché attinse a entrambi i modelli: da un lato fu enfatizzato il realismo, dall'altro si incise con più di tre microfoni dissociati l'uno dall'altro per poi missare i suoni in base alla funzione drammatica (2005: 23). La stereofonia degli anni '50 recupera il principio del facsimile proponendo un nuovo tipo di «cinematic realism», con una corrispondenza spaziale tra immagine e suono che durò poco e restando marginale (2005: 25-28). Il Dolby Stereo si attiene invece al modello dell'intelligibilità, però in una forma incrementata (2005: 52), perciò è necessario con diversi strumenti teorici (2005: 55): esso continua a seguire delle convenzioni stilistiche (specificamente una certa "centralità") ma allo stesso tempo dà luogo a un «representational metaspace»: «Dolby sound depends on the unification or "blending" of diegetic spaces, resulting in a representational model that presents an illusion of a three-dimensional, architectural sound space» (2005: 56). Da questo deriva la natura *delocalizzata* del Dolby Stereo, in quanto offre uno spazio sonoro *figurato* piuttosto che *letterale*; dunque piuttosto che un *approccio scientifico* allo spazio se ne presenta una *nozione astratta* (2005: 50-60). L'effetto è quello di uno spazio sonoro *invisibile* che simula uno spazio profilmico originale. Pertanto

The "hidden" nature of the loudspeakers as dissimulated objects or delocalized ones furthers the idea that the audience member is sutured into the aural environment of the film. [...] She is, in essence, a participant-eavesdropper, privileged to a sonorous event that is constructed in order to bring her into the space of the fiction (2005: 61).

Nella tesi di dottorato Wright affronta invece la questione della firma spaziale in presa diretta, osservando che la tradizione di riprendere col microfono vicino non è così rigida e che il fonico non trascura, come si pensa, la resa dello spazio originale: si predilige un bilanciamento tra l'integrità della voce e il rispetto della prospettiva (2011: 159-60) – però il radiomicrofono elimina la firma spaziale (2011: 176). In altri termini, l'opposizione tra fedeltà e intelligibilità non è netta (2011: 185). Tuttavia queste affermazioni di Wright sono forse influenzate troppo dall'approccio del fonico di presa diretta Ed Novick nella sua collaborazione con Christopher Nolan, che è a detta di

Wright stesso «a small miracle according to current production practices» (2011: 184; cfr. *ibid.*, 172-186).

Whittington afferma che seppure il sonoro multicanale possa circondare il/la percipiente come nella realtà, nella prassi esso è «far from realistic. Sounds that have been placed in the surround channels are formal construction and highly encoded with recording and presentation methods formulated over time and across media» (2007: 1357-1358). Egli arriva a sostenere, seguendo la tesi di Belton (1992) sulla prima stereofonia, che «Dolby Stereo surround sound is not meant to heighten realism at all, but just the opposite, to denounce it» (2007: 1362): ecco perché i film di fantascienza hanno fatto ricorso con grande efficacia a tale sistema a partire da *Guerre stellari*, dato che questo genere riduce il confine tra realismo e artificio (2007: 1366-1373). Secondo Whittington il Dolby Stereo «offered three specific components to the new science fiction spectacles: the localization of effects, elimination or avoidance of masking, and sonic enhancement through spatial placement» (2007: 1386-1387). L'allargamento dello spazio filmico grazie al sonoro permette di posizionare il pubblico nella diegesi, di immergerlo in nuovi mondi grazie a un «total sonic environment, which masks the real environment of the theater space to create a sonic space with no entry and no exit», avvicinando il cinema all'attrazione da parco di divertimenti (2007: 1401-1405). Whittington descrive una prassi di spazializzazione acustica inverosimile, considerando che i canali frontali sono favoriti a prescindere e che si riduce la collocazione puntiforme di suoni nell'area surround a meno che non si voglia enfatizzarne alcuni particolarmente significativi. Il sonoro surround diffonde intorno all'utente gli effetti d'ambiente creando «spatial anchors or cues», ovvero «a kind of sonic peripheral vision not provided by the screen directly ahead», talora acquisendo più importanza dell'immagine. Pur senza citare Oudart né sviluppare la tesi, Whittington osserva che la sutura non dipende solo dal campo/controcampo visivo ma anche nel sonoro surround (2007: 1422-1436).

Whittington enfatizza l'importanza del sonoro multicanale per film spettacolari tipicamente fantastici destinati a un ampio pubblico, i quali «attempt to create not just a narrative experience, but a cinematic experience that feels like an amusement park “ride” with immersive images and sounds». Il suono multicanale ha qui un triplice effetto: fornire una maggiore localizzazione di effetti e musica, eliminare il mascheramento e ampliare lo spazio (2011: 51).

Un testo fondamentale sulla spazializzazione acustica è *Beyond Dolby (Stereo)* di Kerins. L'autore nota giustamente che malgrado ci sia un buon numero di studiosi attenti al suono filmico manca tuttora l'inclusione di questioni relative all'uso del sonoro spaziale – lacuna grave dato che la maggior parte dei film da due decenni a oggi presenta una colonna multicanale (tipicamente 5.1) (2011: 70-71). Kerins dedica la prima parte del libro alla storia, alla produzione e allo stile del

suono stereofonico. Su quest'ultimo punto lo studioso esamina i modi in cui le potenzialità dei sistemi digitali surround (maggiore dettaglio e ampiezza dinamica, assenza di rumore, autonomia dei canali) hanno permesso – se non incoraggiato – nuove scelte nell'uso dei suoni, cioè l'espansione verso entrambi gli estremi di intensità, il ricorso a forti contrasti, l'aumento del numero dei suoni (la cui percettibilità è incrementata), il posizionamento preciso di qualsiasi tipo di suono nello spazio della sala e lo spostamento fluido dei suoni tra i vari canali (2011: 644-957). Emergono quindi le differenze tra il Dolby SVA, il più diffuso sistema surround analogico, e i sistemi indicati da Kerins con l'acronimo DSS (*Digital Surround Sound*), cioè il Dolby Digital, il DTS e l'SDDS, con almeno 6 canali discreti. Tuttavia occorre notare che

Dolby Stereo-era mixing practices have not disappeared entirely. Most dialogue continues to reside in the front center channel, and many filmmakers still have some reservations about using the surround channels *too* aggressively. The difference is that the way a given movie uses surround can now be dictated by what is appropriate for the *story* rather than by technology (2011: 859-861).

Le deviazioni dalle norme consolidate sono sempre più frequenti benché ancora marginali, forse per il timore che il pubblico sgradisca o si disorienti. A proposito del dialogo, Kerins rileva tre approcci alternativi a quello tradizionale che lo confina al canale centrale frontale: il primo prevede il posizionamento del dialogo nell'area in cui si intende emesso (p.es. *Strange Days* [*Id.*, USA 1995]); il secondo consiste invece nel dissociare palesemente il luogo di emissione del suono da quello della sua fonte (p.es. *The Ring* [*Id.*, USA 2002]); il terzo è un approccio intermedio in cui si suggerisce una corrispondenza tra suono e immagine che però viene poi negata (p.es. in *Ragazze interrotte* [*Girl, Interrupted*, USA 1999]) (2011: 926-950). Nel canale LFE vengono spesso missate come rinforzo le frequenze basse dei suoni degli altri canali, dato che non sono localizzabili dall'udito umano; altre volte presentano invece suoni realizzati appositamente (2011: 975-976).

Kerins esamina poi i modi in cui il DSS è connesso con modifiche nello stile visivo dei film, seguendo il concetto di *supercampo* di Chion (2001: 146-148) e proponendo quello di *ultracampo*; inoltre analizza le modalità con cui il DSS viene utilizzato al fine di *immergere* lo spettatore e di posizionarlo in un punto di ascolto soggettivo. Per Chion il supercampo, reso possibile dal Dolby Stereo, è l'insieme di suoni «che circondano lo spazio visivo e possono provenire da altoparlanti situati al di fuori degli stretti limiti dello schermo». Questo insieme di suoni conquista «sul campo visivo una sorta di esistenza autonoma, che non dipende momento per momento da ciò che si vede, ma che non acquista [...] un'autonomia e una pregnanza di rapporti dei suoni tra loro». Il

supercampo ha avuto tra le conseguenze quella di «sottrarre al piano totale il suo ruolo narrativo, poiché esso installa in permanenza [...] la coscienza [...] dell'insieme della scena intorno a ciò che si svolge all'interno del campo» (2001: 147). I canali surround sono un'estensione approssimativa dello spazio filmico denotato dal campo visivo, senza un'autentica coerenza prospettica; inoltre il surround produce un fuori campo passivo piuttosto che attivo, cioè contribuisce a creare «un ambiente che avvolge l'immagine e la stabilizza, senza [...] suscitare il desiderio di andare a vedere altrove o di anticipare la visione della sua sorgente, dunque di cambiare il punto di vista» (2001: 87). Dunque, come spiega Kerins, esso permette alla colonna sonora di fornire il contesto all'immagine liberando quest'ultima da tale responsabilità; di conseguenza «the sonic world becomes *bigger* than the image, which reveals only a small fraction of the world created by the superfield» (2011: 1024-1030). Questa nuova capacità di definire lo spazio attraverso il sonoro ha comportato una riduzione d'importanza del campo totale (2011: 1044-1046) e una maggiore frammentazione delle immagini (2011: 1065). Quindi questo tipo di colonna sonora per certi versi è passiva (il surround è destinato solo a creare un punto d'ascolto stabile) e continua (per bilanciare la frammentazione della colonna visiva; gli stacchi sonori sono perciò inaudibili o non sincronici con quelli visivi) (2011: 1067-1086). Il supercampo è in uso ancora oggi, ma il DSS offre delle potenzialità in più rispetto al Dolby Stereo che se applicate portano a un risultato diverso. Kerins parla quindi di *ultracampo*, un approccio che «sacrifices the “invisibility” of sound editing and mixing» (2011: 1102), con un continuo cambio mimetico di prospettiva, tanto che a ogni modifica del punto di vista c'è una coerente modifica del punto d'ascolto; perciò si hanno spostamenti costanti tra i canali, l'inserimento nel surround di tutti i tipi di suoni (anche i dialoghi), e il condizionamento dei tagli visivi su quelli sonori, come accade nella prima sequenza di *Salvate il soldato Ryan* (*Saving Private Ryan*, USA 1998) e in *Cars – Motori ruggenti* (*Cars*, Usa 2000) (2011: cap. 3, 5). Per la precisione

the ultrafield seeks not to provide a *continuous* aural environment, but rather to *continuously* provide an *accurate* spatial environment where aural and visual space match. In short, the ultrafield is the three-dimensional sonic environment of the diegetic world, continuously reoriented to match the camera's visual perspective (2011: 1101-1109).

La continuità dunque non appartiene più alla colonna sonora bensì alla costruzione dello spazio (2011: 1166), e alla colonna visiva è offerto un ben più alto grado di libertà (nel posizionamento della macchina e nel tipo di inquadratura, nella variazione della prospettiva visiva, nella velocità del montaggio) (2011: 1172, 1203, 1217, 1244, 1251). Specificamente a proposito della prospettiva

visiva, Kerins osserva che «where the superfield places the audience at a *distance* from the onscreen world, the ultrafield situates the audience *in the middle of the action*» (2011: 1276). In questo contesto Kerins, come già accennato, confuta in parte la tesi di Beck (2008) sulle qualità spaziali di una sequenza di *Il silenzio degli innocenti*: egli osserva che le infrazioni della continuità classica sono minime, e che il Dolby SVA qui come altrove – proprio per i suoi limiti – «can only give information *about* the space – it cannot provide any orientation *to* that space» (Kerins 2011: 1190).

Per Kerins il concetto di *immersione* è centrale nel DSS: esso infatti vuole

*immerse the audience in the diegetic world of the film. [...] [T]he audience is literally placed in the dramatic space of the movie, shifting the conception of cinema from something “to be watched from the outside” – with audience members taking in a scene in front of them – to something “to be personally experienced – with audience members literally placed in the middle of the diegetic environment and action* (2011: 1547, 1550-1551).

Questa possibilità è ingigantita dalla natura stessa del suono, caratterizzato da maggiore intimità e presenza rispetto all'immagine, sia grazie all'ontogenesi dell'udito (che si sviluppa nell'individuo prima della vista: il feto non vede ma ascolta) sia perché il suono comporta uno spostamento d'aria (2011: 1559-1561). A proposito dell'ultimo punto, Kerins nota che «the bass effects provided by the LFE create sounds that the audience *feels* rather than *hears*; this physical power creates the very physical sensation of being *at* an earth-shaking event, not just *watching* one» (2011: 1595). I fattori narrativi che favoriscono l'uso del surround sono non solo la creazione di un ambiente o di più ambienti contemporaneamente (2011: 1992-2082), ma soprattutto la definizione di punti di vista e di ascolto soggettivi. Kerins nota che le sequenze POV/POA

offer filmmakers the perceived freedom to be more experimental, particularly in the aural realm. [...] [I]f “interesting” sounds are those that violate standard practices, then logically POV scenes should have “interesting” sound since point of view is not the “normal” mode of filmmaking [...]: for a POV scene to be read as such, it *must* use a style distinct from the rest of a movie (2011: 2189-2195).

Con le sue potenzialità incrementate rispetto alla monofonia e al Dolby Stereo, il DSS permette perciò di definire non solo lo spazio ma anche la posizione dei personaggi e la loro psicologia, portando l'immersione a livelli ancora superiori (2011: 2233-2236).

Il DSS secondo Kerins «attempts [...] to create its own artificial space – to *replace* the acoustics of the venue with those of its own design» (2011: 1483). Ciò è sintomatico della

differenza sul piano acustico tra spazi di *produzione* del suono – tipicamente la sala per concerti – e uno spazio di *riproduzione* del suono come la sala cinematografica: «a concert hall is meant to be *heard* through the way it colors a musical performance while a movie theater is expected to *hide* its own presence and reproduce the soundtrack as “neutrally” as possible» (2011: 1488). Questo è un approccio recente, infatti in epoca monofonica la riverberazione era necessaria per dare un senso di avvolgimento; l’ampia diffusione del Dolby Stereo ha invece richiesto sale meno riverberanti (2011: 1493-1497).

Kerins discute poi la metodologia di analisi della colonna sonora, in riferimento specifico alla sua componente surround. Il film va visto una volta integralmente senza interruzioni «to determine its overall surround aesthetic», poi una seconda volta spostandosi al suo interno per individuare momenti chiave, che includono «both “representative” scenes that follow the movie’s primary mixing style and those scenes where the film deviates from its general surround style» (2011: 2488-2492). Tali momenti devono essere presi nell’interezza dei loro 6 canali e poi isolando ciascun canale o combinazioni dei medesimi, a seconda della necessità (2011: 2535-2536); questo metodo, per ammissione dello stesso Kerins, somiglia al «metodo delle mascherature» di Chion (2001: 178-179).

Kerins infine commenta alcune teorie del cinema, a partire da quelle sulla voce. Egli osserva che la voce è stata considerata a lungo l’elemento fondamentale del suono filmico a causa soprattutto del suo valore significante, trascurando invece la sua qualità prettamente sensoriale e la sua individualità (si parla d’altronde comunemente di *dialoghi* e non di *voci*) (2011: 3031, 3045-3048). La “sincronia temporale” è perseguita da sempre, ma la “sincronia spaziale” è possibile con una certa precisione solo grazie al DSS (2011: 3049-3070). A proposito del collocamento della fonte sonora, lo studioso spiega che i termini *onscreen* e *offscreen* sono impropri perché si riferiscono in realtà alla fonte visiva e non al suono (come già in Metz 1985: 157-158): peraltro nel cinema monofonico tutti i suoni sono *onscreen* nel senso che emanano dallo schermo (ossia dall’altoparlante da esso celato), ma grazie alla stereofonia essi possono essere *onscreen* o *offscreen* in termini sia visivi sia sonori, nelle quattro combinazioni cui si aggiungono svariati casi intermedi (Kerins 2011: 3086-3092). L’autore affronta la teoria dell’apparato di Jean-Louis Baudry e anche la rielaborazione della stessa da parte di Noël Carroll: Kerins ritiene che l’analogia posta da Baudry tra il cinema e la caverna di Platone «may be *more* accurate in the immersive digital surround era», in cui aumenta l’impressione di realtà e dunque la confusione tra *rappresentazioni* e *percezioni* (Kerins 2011: 3330-3334); inoltre proprio grazie al suo potere immersivo il DSS avvicina maggiormente l’esperienza filmica alla vita intrauterina e neonatale rispetto ai sistemi precedenti (2011: 3394-3395). Similmente l’autore trova che «Carroll’s image-centric, monophonic-based

perspective» sia inadeguata per spiegare il cinema stereofonico, poiché il fatto che i suoni provengono da vari punti della sala intorno al pubblico è cruciale per suscitare l'effetto di immersione nella diegesi. L'ultima teoria discussa è quella di Jacques Lacan – e poi di Slavoj Žižek – sulla relazione tra Reale e Simbolico: secondo Kerins il supercampo, poiché consiste di rumori, «foregrounds the *Real* at the expense of the structured Symbolic that *dialogue* and the *image track* represent» (2011: 3528-3529); l'ultracampo porta a possibilità ancora più radicali perché permette al Reale «to expand into what had been the last safe haven of the Symbolic» (2011: 3556), cioè l'immagine (perché la colonna visiva presenta talvolta disordine e frammentazione come il sonoro) (2011: 3551). Il musical è il genere in cui l'intervento del Reale (ovvero dei numeri musicali, di canto e danza) nel Simbolico è strutturale (2011: 3613-3616).

Kerins torna sulla questione del DSS e dell'immersione nello spazio diegetico all'interno di un recente volume collettaneo sul sonoro nel cinema, osservando che «while opinions differ on exactly how dramatic this shift has been and to what degree it represents a break from earlier practices, academics agree that modern soundtracks more actively impact and immerse audiences than their predecessors» (Kerins 2015: 151). Nello stesso libro sono rari i riferimenti allo spazio sonoro. Circa la stereofonia degli anni '50 Platte spiega che la Twentieth Century Fox e la MGM inserirono in apertura di loro film dei “mini-concerti” con lo scopo, tra l'altro, di dimostrare la stereofonia e di ricordare al pubblico «of the spatial configuration of the instrument sections, information that was key to appreciating the stereo mixing of the orchestra that continued once the feature began, with violins placed in the stereo field on stage left, woodwinds center, and brass and low strings on stage right» (Platte 2015: 68). L'autore rileva inoltre l'accoglienza mista del “traveling dialogue”, percepito da alcuni come una distrazione (2015: 72). Beck e Ament, pur non approfondendo la questione, notano che negli anni '80 domina un approccio conservatore nella spazializzazione, che prevede dialoghi centrati e pochi effetti nel surround. Gli autori elogiano il mix in *split-surround* di *Indiana Jones e l'ultima crociata* (*Indiana Jones and the Last Crusade*, USA 1989), «using the surround channels to provide enhanced realism as well as greater spectacle» (Beck e Ament 2015: 126).

Sergi e Wright si sono dedicati specificamente anche al Dolby Atmos. Sergi inquadra questo sistema nell'attuale contesto commerciale, insinuando che esso potrebbe denotare un tentativo dell'azienda Dolby di tornare a essere dominante in un mercato in cui si sono affacciati sistemi sonori tridimensionali (2013: 108-110). L'autore riferisce la sua prima esperienza col Dolby Atmos, basata su materiale promozionale, descrivendolo come un gigantesco passo avanti ed elogiandone specificamente l'omogeneità di timbro e risposta in frequenza che creano «a soundscape that is no longer easily identifiable as emanating from a number of channels». Sergi ipotizza quindi di dover

riconsiderare la sua critica al concetto di trasparenza della tecnologia sonora (cfr. Sergi 1999): nel Dolby Atmos

the goal of transparent sound (where the technology is truly self-effacing and filmmakers do not draw attention to sound directly) is not so much achieved as leaped over into something altogether different. Unlike previous systems, it is extremely difficult to detect the system boundaries [...]. The feeling is that of being in a simulated world rather than before the representation of one, as *Atmos* blurs in many ways the border between artifice and simulation (2013: 113).

A proposito dei termini simulazione e rappresentazione, Sergi osserva che essi «have been used by commentators and scholar to describe, respectively, full immersion in a real or virtual environment and being before one [...]. Simulation immerses audiences fully in the narrative world of a movie, video game, etc. Representation offers it for viewing/listening rather than immersion» (2013: 113, n. 17). Dolby Atmos e simili sistemi, nel loro «knocking on (and possibly breaking through) the door of artifice and crossing into the uncharted territory of immersive narrative within conventional cinema», pongono delle questioni che restano aperte, sancendo un potenziale «cultural shift» (2013: 114-115). Al momento lo schermo continua a essere il fulcro di gran parte del racconto, e il cinema è ancora preso come «a visual medium, with sound providing important but limited support to image», perciò il Dolby Atmos è una possibile minaccia, come dimostra il caso di *Ribelle – The Brave* (*Brave*, USA 2012): il regista Andrews era inizialmente perplesso sul Dolby Atmos in considerazione del fatto di avere così tanta attività lontana dallo schermo, ma ha cambiato idea quando ha capito che il sistema «is more 3D than 3D visuals because it puts you in the environment and it shoves you into that screen» (Richards<sup>37</sup> in Sergi 2013: 116, corsivo mio). Sergi osserva che «*Atmos* and other immersive narrative technologies are transforming auditoria into “environments” to be experienced, where screens are but one dimension of a complex articulation of narrative technologies and techniques» (2013: 117).

Considerazioni affini emergono dal saggio di Wright, che rileva tre aspetti specifici con cui il Dolby Atmos collide, relativi all'ideologia del messaggio e alle convenzioni lavorative: 1- la presunta intuitività che permette di rendere un unico mix ottimale in qualsiasi ambiente; 2- il downmix automatico; 3- la questione del realismo (2015: 228-229). Circa i primi due punti, il problema è che il Dolby Atmos applica algoritmi che adattano automaticamente il mix a una determinata sala, a seconda della sua conformazione, e che effettuano il downmix automatico in 7.1 o 5.1 qualora la sala non sia attrezzata per l'Atmos, in un processo dunque che sfugge al controllo

<sup>37</sup> La citazione è trascritta dal video *The Sound of Brave* (2012): <http://vimeo.com/44754809> (16/6/2017).

degli artisti e ne riduce la libertà creativa ossia – paradossalmente – uno dei fattori che l'azienda enfatizza maggiormente come potenzialità del nuovo sistema. Peraltro i fonici hanno tutt'ora un atteggiamento conservatore verso i canali surround, mentre Dolby sembra suggerire attraverso l'Atmos un nuovo approccio più flessibile (2015: 231-237). Inoltre i fonici statunitensi contemporanei «partecipate in a freelance marketplace» e «Their creative work depends on a shared set of conventions, inherited from the conversion era, that emphasize a cinematic realism based on clarity, intelligibility and the demands of the narrative» (2015: 243), perciò possiamo supporre che ciò restringe il margine di creatività laddove fosse in contraddizione con dette convenzioni e con le aspettative dei committenti (cfr. 2015: 228-229; Sergi 2013). Circa il realismo, Wright osserva che nei *white paper* del Dolby Atmos emerge una concezione diversa da quella tradizionale dei fonici, tutt'ora corrente, che non prevede un trattamento scrupolosamente realistico dei suoni in termini di coerenza prospettiva, ovvero di distanza e di direzione: «To suggest that mixers can obtain a “heightened sense of realism”<sup>38</sup> by panning sound to particular speakers at the expense of narrative transparency is at odds with Hollywood mixing conventions» (Wright 2015: 238-241). Significativamente, i fonici di missaggio di *L'uomo d'acciaio* (*Man of Steel*, USA 2013) hanno sfruttato le aree surround per collocarvi la musica, cosicché «What you end up with is like a cathedral» (Jenkins<sup>39</sup> in Wright 2015: 242), lasciando più spazio nei canali frontali per le informazioni narrative principali (Wright 2015: 241-242).

Riflessioni sullo spazio sonoro filmico emergono anche da svariati altri articoli accademici degli scorsi anni. Manolas e Pauletto notano che il sonoro funge tutt'ora da complemento per le immagini, mentre per sfruttare al meglio il sonoro surround (da loro indicato come «3D sound») bisogna ricorrere a una prassi produttiva diversa (2009: 41). Attingendo alla letteratura gli autori rilevano un probabile limite intrinseco nella sfruttabilità del sonoro surround al cinema, in considerazione di fattori quali la presenza di uno spazio visivo bidimensionale frontale e il problema dell'«in-the-wings effect» di Chion («effetto quinta» nella traduzione italiana; Chion 2001: 85; cfr. 2009: 134), da cui derivano le convenzioni irrealistiche di spazializzazione: in ogni caso essi sostengono che con una accurata progettazione sia possibile espandere le capacità del surround per (a) spazializzare fonti mobili e statiche, (b) la creazione di ambienti sonori di carattere narrativo, (c) la separazione di elementi sonori diversi (Manolas e Pauletto 2009: 41-42). Nel filmato realizzato appositamente dagli autori, la forte spazializzazione a 360° ha funzionato secondo un gruppo di fruitori col suo contributo alla comprensione degli eventi, segnatamente nel missaggio

<sup>38</sup> La citazione è tratta dalla versione del 2013 del *white paper* “Dolby Atmos: Next Generation Audio”, ma è rintracciabile anche nella versione 2014 del testo (Dolby Laboratories 2014).

<sup>39</sup> La citazione è trascritta dal video *The Sound of Man of Steel*, <http://soundworkscollection.com/videos/the-sound-of-man-of-steel> (16/6/2017).

della voce del protagonista che era sita nel surround laddove denotasse pensieri e nei canali frontali laddove denotasse la fonazione (2009: 45, 50). Si sono comunque rilevate alcune difficoltà nella distinzione mentale tra suoni ambientali “passivi” e suoni “attivi” e più in generale nella focalizzazione sui numerosi diversi eventi sonori simultanei (2009: 50-52).

Høier analizza l’uso del surround del cinema mainstream statunitense (2000-2012) prendendo come campione dieci film nominati a uno dei due Oscar per il sonoro (*sound editing* e *sound mixing*). Egli spiega che la «sonic velocity» («velocità sonora»; Chion 2001: 133), che Sobchack (2005: 12) descrive come equivalente sonoro del «visual spectacle», è incrementata dal sonoro surround il quale «also added what can be described as “ear candy” – desirable sounds that could heighten an audience’s emotional involvement by being intensive, surprising, exaggerated, pleasant, shocking and more». Allo stesso tempo però tale tecnica ha comportato dei limiti quali lo *sweet spot* e l’*exit door effect* (quest’ultimo evitabile se suoni direzionali nel surround vengono usati in momenti utili del racconto; Høier 2014: 256), da cui risulta «a practice that involves a healthy pragmatism: [...] *avoiding* dysfunctional results – and presenting a 360-degree sound experience only when it is fruitful to do so» (2014: 253). Perciò il dialogo, regolarmente intelligibile, è usualmente al centro, presentando talora una riverberazione a sinistra e a destra o anche nei surround. Se la voce è collocata nei surround si tratta per lo più di *walla*, guarda caso associato agli ambienti. Inoltre Høier nota acutamente che

diegetic (onscreen) space is mainly experienced as having a depth that is localized *behind* the screen, while the sound speakers are largely limited to presenting the illusion of a sonic space placed *in front* of the screen. In the pragmatic use of frontal voices and the presentation of small amounts of reverb, this will result in voices that are localized at ‘screen depth’, rather than behind or in front of the screen (2014: 254).

Gli effetti invece sono talora collocati con dinamismo per seguire fonti visive in movimento, contribuendo alla “sonic velocity” e creando talvolta «an audience experience of ‘a ride’», oppure si muovono per denotare dei movimenti fuori campo mentre l’immagine resta statica; più spesso si usa lo «stereo panning» frontale coerentemente col movimento visivo della fonte (2014: 255). Gli ambienti sono regolarmente a 360°, ed essendo passivi e di solito percepiti come distanti non rischiano di causare l’*exit door effect*. Per giunta essi possono avere qualità subliminali ed essere “percepiti” («felt») pur se non ascoltati coscientemente, contribuendo ad avvolgere e a fornire continuità in maniera discreta. Insomma l’ambiente rispetto agli effetti ha un legame più debole con le immagini, la connessione è più associativa, essendo l’intento quello di creare «a sonic

background» (2014: 257). Circa la musica, il suo trattamento stereofonico può influire sulla nostra cognizione della sua collocazione dentro o fuori la diegesi; in ogni caso «most film music is highly frontally weighted and the surround channels are most often used for adding space – and spice – rather than for spreading ‘the band all around’» (2014: 258-259). Høier affronta anche il problema del downmix, considerando che spesso al di fuori delle sale si fruiscono i film in stereo a due canali piuttosto che nella configurazione originaria. Egli rileva un impoverimento dell’esperienza spaziale generale, perché si perde la possibilità di avvolgere il pubblico in un ambiente sonoro; ma allo stesso tempo la prassi del surround enfatizza l’area frontale, perciò è solo con gli effetti direzionali surround che l’impoverimento è significativo (2014: 256, 258, 261).

Elvemo parte dall’idea che lo spazio della sala è divisibile in due parti: «one is *visual* and in front of us, the other *out of vision* and behind us»; pertanto col sonoro surround il «cinematic space» può essere rappresentato anche dietro di noi (2013: 34). L’autore ipotizza che

the experience of real visual space in front of us seems to disallow a simultaneous on-going processing of any imaginative visual space, since the process of active mental construction of imaginative visual space will use resources that normally are involved in the processing of information within the visual field. [I]t is likely that it would be difficult, if not impossible, to simultaneously have the two types of space processing connected with the space behind us, in the theatre (2013: 35).

Il fatto che l’*exit door effect* sia tuttora problematico sembrerebbe confermare questa ipotesi (2013: 35-36). Elvemo segue ipotesi scientifiche per cui lo spazio dietro di noi sarebbe esperito come «amodal spatial structural sensing», ossia «a sensation of a space containing relationships and ideas» in cui gli oggetti sono più concettuali che concreti (2013: 37). Inoltre ci sarebbero due modi in cui gli umani elaborano informazioni visive: il «scene-based allocentric frame of reference» che ci permette di elaborare una rappresentazione relativa, essenziale, priva di molti dettagli, relativa allo spazio «out of reach»; quello «egocentric frame of reference» che invece deve preparare il corpo all’azione e quindi richiede dati assoluti sullo spazio «within our reach» (2013: 36-38). Lo spazio visivo filmico è anzitutto allocentrico, ma le parti occluse di un’immagine possono favorire anche la creazione di «an illusive egocentric encoding of the cinematic space», visivamente ristretto all’area schermica e a quella immediatamente confinante. Lo spazio filmico extra-schermico, denotato esclusivamente dal sonoro, è in prevalenza egocentrico perché in assenza di stimolo visivo esso viene percepito più fortemente come corrispondente allo spazio reale (in ogni caso sono i suoni avvertiti come preparatori all’azione che si integrano più facilmente nella

percezione egocentrica). Elvemo suppone quindi che le informazioni sonore – e indirettamente visive – relative all’area extra-schermica siano insufficienti «to bridge the gap between our experience of a diegetic egocentric space behind us, and the restricted diegetic on- and offscreen space in front of us», rendendo forse impossibile la costruzione di uno spazio diegetico avvolgente (2013: 38-41). Secondo l’autore la soluzione consiste nel realizzare un sonoro che pur circondandoci non ci richiede di creare «an active concretisation or imagination of the fictive space behind, but which nevertheless represents that space», evitando così di compromettere la diegesi. I suoni d’ambiente sono quelli che si prestano meglio a questo scopo, ma restano i rischi intrinseci alla tecnologia (p.es. l’esistenza dello *sweet spot*, la qualità dell’impianto nella sala) (2013: 41-42).

Jones rileva che nel cinema tradizionalmente deduciamo che gli elementi importanti sono contenuti nel campo visivo, tuttavia nell’attuale millennio i videogiochi hanno condizionato la composizione dello spazio filmico tanto che siamo più propensi a concepire il mondo finzionale come «in-space» piuttosto che «in-frame»: si tratta di una «*macro mise-en-scene* [*sic*] containing the entire imaginary world» (2005: 1), che ha un effetto immersivo, e non più di una «screen-based mise en scene» in cui il pubblico è separato (2005: 5) (il precursore è *La tunica*; 2005: 2-3). In questo secondo caso lo spazio è rappresentativo, e i suoni sono di solito trattati come “significanti rappresentativi” senza necessariamente conferire loro una fattualità spaziale. Nei videogiochi lo spazio è invece fattuale: per giunta gli autori compongono “lo spazio” piuttosto che “il quadro”, dunque questo è solo una piccola parte di uno spazio più grande costruito contestualmente dal suono (2005: 5-6). Inoltre mentre «the framed image is a completely dynamic and constantly (unpredictably) shifting element, it is the surround-sound design that becomes the linear driver of the viewer/players perception of cinematic space» (2005: 8); d’altronde l’area surround è sfruttata molto più nei videogiochi che nel cinema (2005: 15). Né il formalismo di Ejzenštejn, fondato sul montaggio, né il realismo di Bazin, fondato sul long take e sulla profondità di campo, permettono di comprendere il paradigma della *macro-mise en scène* (2005: 11-15).

Collins ha offerto un’interessante analisi del punto d’ascolto in *Ratatouille* (*Id.*, USA 2007), film rappresentativo del cinema d’animazione ma che suggerisce spunti teorici più generali. Secondo l’autrice gli/le studiosi/e trascurano il fatto che al cinema è possibile creare identificazione pure grazie alla prospettiva uditiva, intesa in senso spaziale e in senso soggettivo (esterno o interno); essa viene realizzata rafforzando o contraddicendo la prospettiva visiva attraverso il posizionamento del microfono e negli altoparlanti e la manipolazione digitale del segnale. Il posizionamento del microfono incide su intensità e timbro (influenzato dalla riverberazione), che dipendono anzitutto dalla sua distanza: quando essa è entro un piede dalla fonte si parla di *close miking*; quando essa è maggiore ma il suono riflesso non supera quello diretto si parla di *distant*

*miking*; quando infine la distanza è tale da rendere il suono riflesso maggiore di quello diretto si parla di *ambient miking* (2013: 288). Questo aspetto non concerne solo lo spazio di per sé ma anche la componente relazionale ed emotiva, il che pertiene alla prossemica uditiva (2013: 290). Circa il posizionamento negli altoparlanti, Collins asserisce che tranne eccezioni in produzioni più astratte e artistiche esso è coerente con la prospettiva visiva, benché in realtà sia vero il contrario relativamente al dialogo (2013: 291-292). La manipolazione digitale del segnale può servire per aggiungere riverberazione e altri filtri. Ciò permette ad esempio di creare tre condizioni spaziali: quella in cui il suono è *occluso*, per cui un ostacolo filtra il passaggio del suono diretto e riflesso; quella in cui il suono è *ostruito*, in cui un ostacolo compromette il passaggio del solo suono diretto; infine quella in cui il suono è *escluso*, in cui un ostacolo filtra il passaggio del solo suono riflesso (2013: 292-293). Collins affronta poi il senso soggettivo della prospettiva uditiva, dividendo tra i casi di esteriorità (ascoltare soggettivamente suoni reali) e quelli di interiorità (ascoltare suoni mentali) (2013: 294-296).

Grajeda nota che l'avvento della stereofonia non è stato un semplice passo verso una maggiore fedeltà bensì un processo complesso che coinvolge anche questioni teoriche, commerciali, estetiche e sociali (2015: 37-39). Il dibattito dell'epoca manifesta una costante tensione tra il realismo e la sua negazione, con un risultato paradossale. Essendo stato il mono la norma fino ad allora, lo stereo era percepito come «audio of attractions», ossia come uno spettacolo che «openly celebrated rather than concealed the apparatus»: tuttavia la sua potenzialità era regolarmente dimostrata attraverso la riproduzione di suoni naturali o dell'esperienza quotidiana (2015: 40-46, 56). Il problema della resa stereofonica dei dialoghi al cinema era dovuto al fatto che infrangeva una norma consolidata, perciò lo stereo, benché in sé teoricamente più realistico, divenne «the unrealistic other to monaural film sound» (2015: 58).

Malsky commenta l'avvento del sonoro magnetico del CinemaScope in base alla letteratura tecnica dei fonici dell'epoca. Egli nota che l'ambizione era quella di creare «a commercial system that implemented aspects of Fletcher's auditory perspective, but one that was also eminently pragmatic and commercially viable»: ne risultò un «negotiated compromise» tra le potenzialità della nuova tecnologia stereofonica e la prassi preesistente. Malsky conclude dunque che «CinemaScope's sound sought a more convincing illusion over a more accurate replication of "reality." CinemaScope sought to reinscribe – though in an expanded form – the established stylistic norms of Hollywood sound» (2015: 223).

Un contributo più recente sul punto d'ascolto, sebbene incentrato sul silenzio, è quello di Wierzbicki il quale riconosce – rifacendosi a Bordwell e Thompson – l'esistenza di suoni interni ed esterni, nonché di una prospettiva sonora oggettiva e una soggettiva e, utilizzando altri termini, di

un'auricularizzazione zero e una interna (2016: 138-139). Circa il “quasi silenzio”, egli ne individua tre categorie: *environmental silence*, allorché il silenzio dipende dallo spazio in cui il personaggio sta ascoltando; *psychological silence*, laddove il silenzio dipende da uno stato mentale; *neurological silence*, qualora il silenzio dipenda da una condizione fisica momentanea o permanente (2016: 139-146).

Dienstfrey ha offerto uno studio che potrebbe aprire una nuova strada nella ricerca, ponendosi in modo critico verso la letteratura accademica. Secondo lui sarebbe nato un “Dolby myth” secondo cui si esagera il ruolo di questa azienda nella storia del sonoro cinematografico (un esempio verrebbe da Sergi 2004) (Dienstfrey 2016: 169). L'autore si propone piuttosto di svelare che il Dolby 70mm «was not a revolution but a continuation of stereo practices that Hollywood codified during earlier decades», e che «the Dolby myth was formed in reaction to a period of economic uncertainty for Hollywood's sound community, with editors and mixers compelled to promote themselves by crafting a flattering, yet untenable, history of their trade as a matter of professional necessity» (2016: 170). Dienstfrey ridimensiona l'importanza data ad *Apocalypse Now* per due motivi: anzitutto le sale dotate di *split-surround* all'epoca erano meno di 20 in tutto il mondo; poi, molti effetti surround descritti dagli/dalle studiosi/e non erano possibili dati i limiti del sistema; peraltro le versioni del film in 5.1 successive al restauro del 1997 presentano differenze rispetto al missaggio originale (2016: 171-171). In seguito l'autore afferma che il Dolby Stereo ha utilizzato tecnologie preesistenti: «The innovations regularly attributed to Dolby Labs and Walter Murch were developed by an older generation of sound engineers» (2016: 174, 177). Inoltre lo stile sonoro dell'era Dolby si poneva in continuità con i decenni precedenti, come dimostra un confronto tra film in Dolby 70mm quali *The Wiz* (*Id.*, USA 1978), *Alien* (*Id.*, USA 1979) e *Tron* (*Id.*, USA 1982), e film stereofonici antecedenti il Dolby Stereo quali *Viaggio al centro della terra* (*Journey to the Center of the Earth*, USA 1959), *Quelli della San Pablo* (*The Sand Pebbles*, USA 1966), e *Tora! Tora! Tora!* (*Id.*, Giappone/USA 1970): capitava già che nel surround fosse inserita della musica, nonché degli effetti sonori per «spectacular settings» diversamente da «ordinary settings» (Dienstfrey 2016: 177-180). Per giunta la preminenza dello schermo cinematografico dissuadeva i fonici dall'usare il surround con la stessa libertà che in teoria era permessa dai coevi sistemi multicanale domestici (il quali però in pratica soffrivano sovente anch'essi di limiti tecnici di spazializzazione): infatti il timore per l'*exit sign effect* influenza il missaggio dei film in Dolby 70mm seppure siano «commonly credited with surround-sound innovation» (2016: 184-185). In conclusione l'autore afferma che

Dolby did not dramatically transform Hollywood surround sound. Rather, the company inherited technologies and popularized mixing styles that were established by an earlier generation of motion picture sound engineers. What made Dolby distinct was the extent to which filmmakers exploited its brand for personal gain (2016: 186).

Ho l'impressione che le tesi di Dienstfrey siano valide ma che soffrano comunque di eccessivo revisionismo. Il fatto stesso che il Dolby Stereo abbia sancito l'imposizione della stereofonia come normale – ovvero che abbia permesso la diffusione di determinate tecnologie, ancorché preesistenti – è un traguardo rilevante, sebbene lo stile di spazializzazione non sia sostanzialmente mutato. È altresì rilevante il fatto che nessun sistema prima del Dolby 70mm aveva inglobato diverse possibilità tecniche disponibili separatamente, ossia l'uso del surround (un canale, o due seppure non esattamente discreti come nel Cinerama), del potenziamento dei bassi (come il Sensurround, che era mono), della riduzione del rumore.

### **2.6.3 – Italia**

In Italia lo studio del sonoro è decisamente marginale. Qualche riflessione sullo spazio sonoro è riscontrabile nel manuale di Valentini (2006) dedicato al suono nel cinema. Secondo l'autrice il sonoro «dà alla bidimensionalità dello spazio filmico una inedita profondità» e può attivare il fuori campo sia frontale sia posteriore alla macchina da presa. Quest'ultima area è particolarmente problematica «perché le parole proferite arrivano letteralmente da noi, dal nostro spazio, forse dalla nostra persona determinando un forte senso di straniamento». Ecco perché «questa risorsa è ancora poco sfruttata sul piano della voce» fin dall'epoca del Dolby Stereo con la diffusione del surround («il simbolo di questo sfondamento»), che comunque è utilizzato per ottenere «una totalità sonora che ingloba l'individuo e lo avvolge da tutti i lati». Infatti

Il cinema contemporaneo, dopo il mito tardo-rinascimentale portato avanti dal cinema classico del dominio prospettico dei suoni, tende a fare assaporare allo spettatore anche questa condizione primigenia di sottomissione ai suoni e di perdita delle coordinate spaziali rimettendo completamente in discussione la consistenza stessa del fuori campo (2006: 144-146).

Con tale riflessione Valentini coglie un aspetto del sonoro multicanale che richiama il bagno di suoni di Jullier (2006: 56), ossia il fatto che si abbandoni «la dimensione frontale dello spettacolo per gettare lo spettatore in una dimensione a prospettica, in cui come in una bolla di suoni egli galleggi immerso nel quadro dell'immagine» (Valentini 2006: 119). In un successivo testo l'autrice

ribadisce una simile posizione, notando che l'ambiente sonoro del film contemporaneo non è qualcosa in mezzo a cui il pubblico "cammina", ovvero «qualcosa di cui, in una visione umanistica, lo spettatore è al centro e che riorganizza intorno a sé», bensì «un'illusione aggettante verso lo spettatore [...] che di prospettico non ha quasi più nulla, che si muove verso lo spettatore, lo cattura, lo tocca» (2011: 16).

Lo studioso italiano che ha affrontato maggiormente la dimensione spaziale del sonoro filmico è Di Donato. Il primo saggio (Di Donato 2004) è integrato in un libro successivo (Di Donato 2006) dedicato esclusivamente alla spazializzazione acustica. Qui l'autore descrive due tendenze relative alla spazializzazione, le quali emergono già dagli esperimenti degli anni '30: la prima è caratterizzata dalla «concordanza spaziale», perciò è improntata «all'aderenza tra coordinate percettive e finzionale», e la si riscontra soprattutto nella realtà virtuale; la seconda, normale al cinema, è caratterizzata dalla «chiarezza espositiva», ossia «un criterio di distribuzione degli eventi acustici in base alla funzione che essi ricoprono all'interno del dispositivo filmico» (2006: 28-29, 50). Di Donato ricorre al concetto di immersione nell'accezione sensoriale, cioè intendendola come presenza la quale si può attuare in un contesto di sonoro tridimensionale come quello offerto dai sistemi surround (2006: 46-50); questa terminologia sarà riveduta dall'autore in un suo saggio successivo (2014b). Le funzioni dei vari canali in relazione al secondo modello sono così riassunte:

- Canale centrale (ascolto monoaurale) – dialoghi e principali funzioni narrative
- Canali destro/sinistro (ascolto stereofonico) – musica, effetti speciali e ambiente
- Canali surround (ascolto stereofonico) – musica, effetti speciali e ambiente
- Canale per le basse frequenze – effetti speciali e musica (2006: 52-3).

Questo modello è correlato alla prassi di costruire una scena sonora stabile in cui i punti di vista e d'ascolto sono dissociati (2006: 56). Ad ogni modo nel cinema contemporaneo si attua di solito una mediazione tra i due modelli, da cui scaturisce un'*immersione interrotta*: i suoni in campo sono diffusi prevalentemente dal canale centrale a prescindere dalla prospettiva visiva, ma determinati suoni possono comunque essere collocati nel surround con coerenza prospettica. Resta il fatto che il "verbocentrismo" di Chion sembra una costante anche in senso spaziale, poiché la voce è centrata (Di Donato 2006: 58-59).

Di Donato si concentra sul problema del surround, da cui può scaturire l'"effetto quinta" (cfr. Chion 2001: 85) qualora lo si usi in modo troppo realistico, portando a un indebolimento della diegesi e a uno spaesamento (il che non si pone invece nella realtà virtuale, in cui si perde la

centralità dello schermo) (Di Donato 2006: 62-64). Nel «processo di costante negoziazione tra piano visivo e sonoro» attuato dal cinema contemporaneo possono scaturire due problemi: «1) la mancanza di coincidenza spaziale tra stimoli uditivi e stimoli visivi, che può ingenerare nello spettatore un senso di disorientamento; 2) il possibile conflitto nelle procedure di messa in scena tra la qualità tridimensionale del suono surround e quella bidimensionale dell'immagine schermica». Ciò deriva inoltre dal fatto che il formato 5.1 «agisce su uno spazio a 360°» e non a 180° come per solito avviene con le immagini (2006: 65-66). Secondo l'autore la direzionalità del surround può essere sfruttata per produrre quattro tipi di effetto: 1) la determinazione di «uno spazio “aumentato”», ossia di coerente denotazione dello spazio diegetico fuori campo (2006: 91); 2) l'*effetto fondale*, allorché «i rumori di ambiente vengono diffusi dal surround indipendentemente dal contenuto visivo dell'inquadratura» (2006: 57); 3) «una dialettica tra “centralità” sonora e “multidirezionalità” (in funzione drammatica o simbolica)»; 4) l'attivazione di «una dimensione altra, che si aggiunge o addirittura oltrepassa quella del visibile» (2006: 92). La potenzialità connotativa del surround è evidente in particolare nel caso di acusmetri (2006: 101-108).

Di Donato, seguendo Altman (1995: 68), identifica ne *Il ritorno dello Jedi* l'iniziatore del nuovo corso del *sound design*, poiché Burt Truitt attribuisce al surround «una funzione puramente “attrazionale”» (Di Donato 2006: 71), che quindi attiva un fuori campo passivo (2006: 73). Infatti la tendenza è quella di comprimere gli effetti di spazializzazione acustica, come emerge dalla ricognizione statistica operata dall'autore su un campione di 150 film statunitensi del periodo 1996-2006, i cui risultati sono inseriti in una tabella (2006: 74). In ogni caso l'autore riscontra un incremento degli effetti di spazializzazione, un fatto che egli riconduce da un lato alla diffusione degli home theatre e dall'altra a una maggiore familiarità con ambienti acustici immersivi quali quelli offerti dai *First Person Shooters* (FPS) (2006: 76-77). Cionondimeno i suoni molto direzionali hanno di solito servono a «sottolineare passaggi salienti dell'intreccio, a determinare nello spettatore condizioni percettive anomale e a fornire indizi sulla porzione di spazio diegetico non visualizzata» (2006: 79). A proposito del movimento sonoro, Di Donato spiega che esso può denotare due tipi di spostamento: 1) quello del punto d'ascolto, come se si trattasse di una «carrellata sonora»; 2) quello della fonte rispetto a un punto d'ascolto fisso. Tuttavia è possibile operare una distinzione solo grazie all'immagine (2006: 85-86).

Lo studioso riconsidera infine il concetto di auricularizzazione teorizzato da Jost, notando come il suono surround permette processi di «soggettivizzazione dell'ascolto che non [hanno] alcun corrispettivo nel cinema classico monofonico» (Di Donato 2006: 120). Si dovrebbe dunque prevedere anche un'auricularizzazione «avatariana», laddove il punto d'ascolto sia mediato da tecnologie di realtà virtuale (2006: 125).

Di Donato offre dunque molte nozioni utili per la disamina che offrirò più avanti, che includono concetti quali il raccordo sonoro di posizione (2006: 69), quello di direzione acustica (2006: 85), e la “bilocazione audiovisiva” (2006: 101). Inoltre egli propone una riflessione cruciale sullo spazio sonoro, a partire dal rapporto tra montaggio visivo e montaggio sonoro. Se è vero che il secondo non definisce unità specifiche sonore come fa il primo con le inquadrature, è comunque possibile individuare segmenti acustici laddove in una scena si alternino punti d’ascolto marcati che denotino un cambio di posizione (2006: 117, n. 4).

I due saggi di Di Donato contenuti in una recente antologia (Di Donato e Valente 2014) affrontano questioni in cui la dimensione spaziale è determinante. Nel primo lo studioso parte dalla considerazione che «l’assemblaggio orizzontale della colonna sonora non dà luogo ad unità tecniche o linguistiche specifiche, come nel caso dell’inquadratura» (Di Donato 2014a: 13), il che già può incidere sulla costruzione spaziale: il sonoro permette una continuità spaziale. Il sonoro ha una «natura essenzialmente simulacrale» (2014a: 14). Il punto di ascolto non è condizionato quasi mai dalla posizione originari del microfono poiché la scena acustica viene costruita a posteriori in fase di missaggio». Circa il punto d’ascolto spaziale Di Donato spiega che esso «pesa assai meno del punto di vista [...]. D’altra parte, pur con un minore grado di precisione, anche il suono è in grado di evocare condizioni acustiche caratterizzate dalla localizzazione da cui si percepisce una scena». Di solito però i suoni non vengono trattati al fine di denotare «un punto di ascolto precisamente localizzato» né coerente col punto di vista. Ciò dipende sia dall’esigenza di garantire l’intelligibilità del dialogo sia da un altro aspetto:

la condizione ubiquitaria della visione cinematografica, che tramite il montaggio consente di trasportare istantaneamente lo sguardo da un punto ad un altro della scena, determina, sul piano acustico, una esperienza percettiva innaturale. Infatti una coincidenza assoluta tra punto di vista e punto di ascolto produrrebbe una linea sonora caotica e frammentata per via delle continue variazioni di volume e di colorazione ambientale (2014a: 21-23).

Questa osservazione è in buona parte condivisibile attenendosi ai dati delle scienze cognitive e alla fisiologia umana, ma allo stesso tempo resta il fatto che si potrebbe garantire la coerenza prospettica senza che ciò risulti caotico – se non per la mera infrazione di una convenzione stilistica, piuttosto che di una norma percettiva. Il confine tra oggettività e soggettività «è labile e ambiguo» nel caso dell’ascolto, diversamente dalla visione (2014a: 24). Nel caso della condizione di oggettività, Di Donato opera una distinzione persuasiva tra il punto d’ascolto *onnisciente* e quello *antropomorfico*: il primo concerne «una stilizzazione dell’ambiente acustico finalizzata a privilegiarne

l'intelligibilità», da cui scaturisce «una condizione di “ascolto assoluto”, ipertrofico e infallibile»; il secondo è invece «verosimilmente commisurato alla sfera percettiva di un udente situato all'interno della diegesi», offrendo «una condizione percettiva umanizzata» (2014a: 30-31).

Nel secondo saggio l'autore affronta il sonoro multicanale partendo da riflessioni di Chion e Jullier rispettivamente sul supercampo e sull'effetto bagno. Questi due studiosi descrivono due possibili strategie di ricordo alla stereofonia: «una che va nella direzione di una organizzazione analitica dello spazio acustico della sala, in grado di attivare una scena sonora virtuale attraverso la valorizzazione della direzionalità (il supercampo di Chion); l'altra che si manifesta in un assalto sensoriale dall'effetto ottundente e ipnogeno, una sovra-stimolazione» che, come dice Jullier (2006: 56) «dà allo spettatore la sensazione di fluttuare al centro di un magma» (Di Donato 2014b: 52). Rispetto al suo libro precedente (Di Donato 2006: 46-50), qui l'autore definisce questi due modelli rispettivamente coi termini di *tridimensionalità* e *immersione*, asserendo che il primo era normale nei primi anni della stereofonia (almeno relativamente ai canali frontali; non si fa riferimento all'area surround) mentre il secondo si è imposto negli anni '80 e resta dominante, e prevede un uso appunto prevalentemente “immersivo” (nell'accezione summenzionata) del surround (cui sono destinati ambienti, musiche e taluni effetti) mentre i canali frontali (e soprattutto quello centrale, nel caso del dialogo) contengono le informazioni narrative principali (2014b: 53-56). Di Donato osserva infine che, malgrado «l'allineamento tecnologico» tra i diversi media audiovisivi, permane questa separazione tra strategie di spazializzazione, poiché videogiochi e realtà virtuali tendono a presentare ambienti sonori tridimensionali (ancorché non necessariamente realistici) mentre i film continuano a usare le tecnologie con le funzioni stilistiche tradizionali (2014b: 56-58).

## 2.7 – La collocazione dei suoni

Un aspetto molto dibattuto da teorici e dalle teoriche che hanno affrontato il sonoro è quello concernente il “luogo” di provenienza dei suoni. Si tratta di rispondere a domande quali: la fonte è mostrata nell'immagine o è occultata? È presente nella scena o è esterna a essa? Fa parte del mondo narrativo oppure è un puro artificio stilistico? Ho dedicato a questo problema un articolo (Sbravatti 2016) in cui mi sono concentrato sulla musica, ma le mie osservazioni sono estensibili a qualunque tipo di suono. Qui riporto uno *status quaestionis* basato sul suddetto articolo, mentre il mio personale contributo a questo argomento sarà inserito, in una versione riveduta, nel capitolo 5.

Expression	vs	Inzidenz	(Erdmann-Becce-Brav)
Musica sincronica	vs	Musica asincronica	(Ejzenštejn, Pudovkin)
Musica realista	vs	Musica funzionale	(Manvell-Huntley)
Musica reale	vs	Musica funzionale	(G. Van Parys)
Musica effettiva	vs	Musica di commento	(Karel Reizš [sic])
Musica giustificata dall'immagine	vs	Musica non giustificata	(J.-R. Julien)
Livello interno	vs	Livello esterno	(S. Miceli)
Musica oggettiva	vs	Musica soggettiva	(Odile Larère)
Musica dello schermo	vs	Musica della fossa orchestrale	(M. Chion)
Musica diegetica	vs	Musica non diegetica	
Suono <i>in</i>	vs	Suono <i>off</i>	

Tabella 1: collocazione della musica (Simeon 1995: 47<sup>40</sup>).

Nella tabella 1 notiamo una varietà di termini che denota l'attenzione di vari studiosi verso svariati aspetti della presenza della musica. Il caso Erdmann-Becce-Brav è un'eccezione perché loro non si riferiscono affatto alla collocazione ma solo alla funzione della musica, che può essere espressiva (*Expression*) o informativa (*Inzidenz*). Spesso il contrasto è tra l'esigenza di realismo e la funzione espressiva (Ejzenštejn, Pudovkin, Manvell-Huntley, Van Parys, Reisz); altrove è tra oggettività e soggettività (Larère). Nel complesso sembra che gli studiosi si riferiscono alla connessione della musica col racconto (è certamente il caso di Miceli), ma talvolta essi alludono piuttosto alla connessione tra musica e immagine (Ejzenštejn, Pudovkin, Julien, Chion).

Per descrivere la collocazione di un suono in relazione alle immagini si suole ricorrere ai termini *in* (in inglese *on*), *off* e *over*; inoltre esso va distinto in *esterno* (se è concreto) e *interno* (se è mentale) (cfr. Casetti e di Chio 1990: 89-95; Bordwell e Thompson 2003: 404-407). Per chiarire invece la questione della collocazione dei suoni rispetto alla storia si è fatto affidamento al concetto di diegesi. Questo termine, proposto da Platone e usato con diverse accezioni da diversi studiosi (p.es. Souriau, Genette, Metz), di solito designa la sfera della storia, il mondo narrativo; pertanto alcuni suoni sono *diegetici* (*diegetic*) e altri no, ossia sono *extradiegetici* (*nondiegetic*). Questi termini hanno ottenuto successo soprattutto grazie al libro *Unheard Melodies* di Gorbman (1987), nonché al noto *Film Art: An Introduction* di Bordwell e Thompson (1<sup>a</sup> edizione 1979) i quali affrontano la questione associandola all'aspetto spaziale del suono. Tuttavia erano stati già usati e saranno ripetutamente discussi da vari teorici.

<sup>40</sup> Simeon prende questa tabella, aggiungendo proprie integrazioni, da J.-R. Julien (1987: 28), qui tratto da Simeon (1995: 47)., "Défense et illustration des fonctions de la musique de film", in *Les musiques des films*, a cura di Antoine Hennion, Privat, Toulouse 1987, p. 28.

Percheron, in un articolo del 1973, è tra i primi ad aver studiato la relazione del suono con l'immagine e la diegesi, elaborando un modello attraverso il quale sarebbe possibile comprendere la posizione e la funzione nel contesto narrativo delle canoniche tre categorie di suono (dialogo, musica, effetti). Nel modello confluiscono coppie di termini – alcune delle quali normalmente in uso in sede teorica – inserite in una tassonomia: suono on/suono off, una dicotomia che tuttavia «testifies to the image's primacy» ed è per certi versi problematica; diegetico/extradiegetico, ancor più problematica a causa della difficoltà di definizione del concetto di diegesi, ma comunque utile; sincronico/asincronico; marcato/non-marcato, in relazione alla modalità comunicativa (1980: 16-19).

Relativamente alla voce over Percheron opera una distinzione in 4 tipi: commento, voce over su flashback, monologo interiore, voce over “off”<sup>41</sup> non-marcata (cioè che non agisce in modo palese, e che può influenzare o meno «what happens and/or what is said in the frame»; 1980: 22). Ciascuno di questi tipi può essere specificato in relazione alla diegesi, alla sincronia e alla modalità comunicativa, come emerge dallo schema in figura 12.

	Commentary	Voice-over on flashback	Interior monologue	Unmarked “off” voice-over
1) diegetic	–	+	+	+
2) synchronized with image		–	+	+
3) directly communicated	+	+	+	–

Figura 12: Tipi di voce over secondo Percheron (1980: 21).

Lo studioso riassume la sua teoria nello schema in figura 13.

<sup>41</sup> Percheron in questo caso intende off come presente esclusivamente nel piano off; la voce over in flashback invece è tipicamente una voce che diventa off solo quando accompagna il flashback, mentre in un primo momento – e anche in seguito – è in.

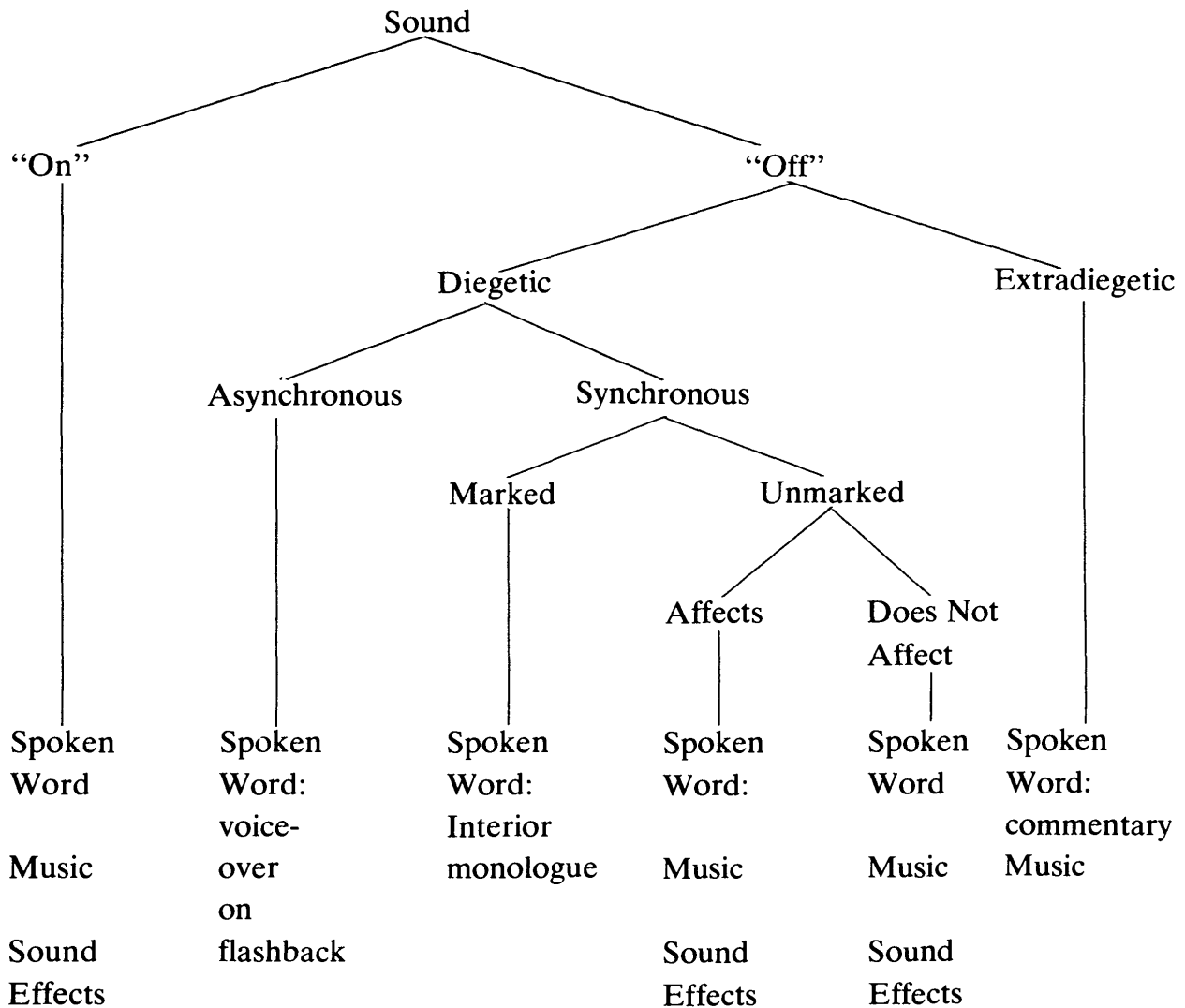


Figura 13: La teoria di Percheron (1980: 23).

Percheron afferma che nello schema il dialogo «remains the major structural element, the dominant signifying material, the primary component of sound in the cinema. [...] Sound effects and music are rarely conveyors of meaning» (1980: 22-23).

Branigan è stato tra i primi studiosi a riflettere sullo status di suono diegetico all'interno di un discorso sullo spazio filmico. Egli propone di usare il termine "intra-diegetico" per indicare un suono udibile da un solo personaggio. Egli porta l'esempio di un rumore apparentemente extradiegetico di cui poi si rivela la natura diegetica: usando una error theory, si stabilisce che il suono è sempre stato diegetico; usando invece una hypothesis theory si stabilisce che il suono era inizialmente extradiegetico, poiché «it is *our* relationship to character and sound which has changed and that is what is being measured». Ciò dimostra che «the "same" sound may have different *functions* at different times» (1981: 62). Insomma «the error theory seeks to assing an absolute

division between diegetic and non-diegetic without taking into account [...] the hypotheses the viewer must make about the story» (1981: 63).

Chion, come abbiamo visto, ha trattato approfonditamente il rapporto tra spazio sonoro e spazio visivo (2001: 74-88; 2009: 249-260). Egli nota inoltre che la distinzione tra musica di commento e musica della scena (*source music*) è infelice perché una musica può commentare anche quando appartiene alla seconda categoria (fattore colto già da Gorbman 1980: 198-200; 1987: 23-25); Chion usa pertanto termini più neutrali per denotare i due concetti, ovvero rispettivamente *musica dello schermo* e *musica della buca* (2001: 82; 2009: 411-412), e precisa che «there is no reason an “actual” music cue could not function as “commentative” at the same time» (2009: 412; inoltre Chion asserisce che l’empatia si genera solitamente – ma non necessariamente – dalla musica della buca, mentre l’anempatia si genera solitamente – ma non necessariamente – dalla musica dello schermo; 2009: 430-431). Una posizione analoga è condivisa da studiosi quali Donnelly (2005: 12), Norden (2007), Heldt (2013: II.iv.d). Inoltre la distinzione tra diegetico ed extradiegetico può essere ambigua (o gestita con ironia), come affermano Bordwell e Thompson (2003: 403-408) nonché svariati/e altri/e studiosi/e che hanno proposto termini specifici per denotare simili circostanze. Brown (1992: 886 e *passim*) parla di “quasi-diegetic music”, ossia musica extradiegetica che in qualche modo sembra emanare dalla diegesi. Kassabian ritiene troppo rigida la dicotomia diegetico/extradiegetico perciò usa l’espressione «source scoring», mutuata da Earle Hagen, per indicare la musica che «falls between diegetic and nondiegetic music», nel senso che è musica diegetica (*source*) che però commenta come se fosse uno score (2001: 587-670). Anche Holbrook pone enfasi sulla capacità della musica diegetica di commentare, il che a sua detta non è una circostanza rara perciò non può essere considerata un’eccezione alla regola – dato che il presupposto comune è che la musica diegetica, diversamente dallo score, non commenta. L’autore propone un nuovo modello che mutua termini da Kassabian e da Chion: in aggiunta ai tre consueti tipi di musica – in diegetica, off diegetica (“in-between music”), off extradiegetica – egli descrive una musica “ambi-diegetica”, ossia musica diegetica che contribuisce allo sviluppo drammatico del film (Holbrook 2003: 207-230; cfr. 2004: 171-176; 2005a: 154-155; 2005b: 48-49; 2009: 279). Taylor, rifacendosi a una fonte indeterminata dei primi anni ’90, usa il termine «transdiegetic sound [...] referring to sound’s propensity to cross the border of the diegetic to the non-diegetic and remaining unspecified» (2007: 3). Slowik parla di «diegetic withdrawal», un fenomeno per cui un brano musicale passa da dentro a fuori la diegesi per fungere da commento (2013: 1-25).

In quanto alle relazioni tra i suoni e il racconto, Neumeyer trova ambigua e insufficiente la terminologia corrente e propone quindi un proprio modello tripartito *ancoraggio* → *diegesi* → *narrazione*. I suoni dell’esperienza cinematografica possono essere inseriti all’interno di uno dei tre

stadi. Il primo stadio prevede due livelli, ovvero il *mondo reale* (quello dell'ambiente acustico in cui si svolge la proiezione) e il *mondo filmico* (quello che emana dagli altoparlanti), e due modalità, quella minacciosa e quella non minacciosa. Il secondo stadio si riferisce all'appartenenza o meno dei suoni filmici alla diegesi (e quindi essi possono essere *diegetici* o *extradiegetici*). Il terzo stadio concerne «the functions that arise from the treatment of the opposition [diegetic/nondiegetic] in the narrating instance of a particular film and shift the balance toward interpretation». La dicotomia diegetico/extradiegetico è appunto concepita di solito come una semplice *opposizione*; essa è altrimenti interpretabile come un *continuum* «with a potentially infinite number of intermediate stages, each one a distinct function» (è opinione di Kassabian 2001: 587-670). Neumeyer sostiene di preferire piuttosto una terza opzione, quella di Stilwell, che mantiene l'opposizione ma include anche «a distinct and complex space inbetween, the “fantastical gap”» (Neumeyer 2009: 29-30; Stilwell 2007: 184-202), cioè

the border region [that] is a transformative space, a superposition, a transition between stable states. [M]ovement through the gap between diegetic and nondiegetic [...] takes on great narrative and experiential import. These [...] are important moments of revelation, of symbolism, and of emotional engagement within the film and without (Stilwell 2007: 200; cfr. le “transizioni tra livelli” di Miceli 2009: 657-633).

Il primo scopo della diegesi è quello di preparare l'ancoraggio al livello del film (che è secondario rispetto al semplice sedersi in una sala), il secondo è quello di permettere l'interpretazione. Nel passaggio alla diegesi «the diegetic has the advantage of simplicity, whereas the nondiegetic is always more complex (because its narrative level has to be constructed – that is to say, the sound has to be interpreted)». Nel passaggio alla narrazione invece

the roles are reversed: the nondiegetic has the advantage of (relative) simplicity (since it is already narrative in function), whereas the diegetic is always more complex (because it provokes the question, Is this environmental sound or does it have [is it meant to have] narrative significance?) (Neumeyer 2009: 31).

Neumeyer tocca degli aspetti in parte rintracciabili anche nella riflessione di Smith – sempre concernente il problema del rapporto tra musica e diegesi (ma il discorso, come di consueto, è estensibile a tutti i suoni) – che si colloca su un livello meno astratto e più analitico. Lo studioso elabora un modello (v. figura 14) in cui la musica diegetica è intesa, diversamente dalla norma,

come appartenente alla fabula ma non all'intreccio, mentre la musica che si colloca in entrambe le sfere è indicata come *source music* (musica della scena)

Nondiegetic Music	Source Music	Diegetic Music (implied)
Scores/"pit music"	Musical performances	Referenced or inferred pieces
<i>Syuzhet</i>	<i>Syuzhet/fabula</i>	<i>Fabula</i>

Figura 14: La teoria di Smith (2009: 2-3).

Circa gli sconfinamenti tra il piano diegetico (nell'accezione comune) e quello extradiegetico, Smith spiega che occorre separare il concetto di diegesi da quello di realismo: «Realistic sound levels are, thus, a sufficient condition for diegetic music on an integrated sound track, but they are hardly a necessary condition» (2009: 4). Sebbene esista una zona destabilizzata e ambigua tra i due piani, ossia il già citato «fantastical gap» di Stilwell (2007), essa agisce meno di quanto si pensi: si tende infatti a confondere la manipolazione del suono (il volume, il riverbero) della registrazione musicale con un suo scivolamento dal diegetico all'extradiegetico e viceversa (Smith 2009: 6-7). Similmente, «[b]ecause the use of nondiegetic score is so common in fiction filmmaking, one usually assumes that any music not clearly located in the story world is likely to be nondiegetic»: in realtà si potrebbe trattare di uno *spostamento spaziale* – ossia di un fenomeno, trascurato dai teorici, simile allo *spostamento temporale* di Bordwell e Thompson (2003: 411; «differimento», nell'edizione italiana) – il che comporta che la musica provenga da uno spazio storico diverso da quello mostrato al momento (e non da uno spazio estraneo alla storia) (Smith 2009: 14). Lo studio della musica nei film necessita in conclusione, per evitare simili fraintendimenti, di una comprensione più precisa «of the difference between film *narrative* and film *narration* [*corsivi miei, n.d.a.*]» (Smith 2009: 23).

Biancorosso segue la proposta di Smith, studiando casi di ambiguità percettiva dentro al modello diegesi. Egli nota che quando l'ambiguità è risolta, per esempio svelando che una musica apparentemente extradiegetica è in realtà diegetica, c'è «multistability» piuttosto che ambiguità in senso proprio. L'assenza di una firma spaziale realistica dalla musica è uno degli elementi che più spesso portano a interpretare tale musica come extradiegetica, in aggiunta a inferenze basate sulla plausibilità interna al testo o su convenzioni acquisite. In *Il dittatore dello stato libero di Bananas* (*Bananas*, USA 1971) la presenza dell'arpista nell'armadio è così improbabile che è imprevedibile, perché (1) il suono dell'arpa è privo di una firma spaziale realistica, (2) nulla nel contesto della

storia suggerisce che questo evento sarebbe plausibile (coerenza interna), e (3) la musica è usata con la consueta funzione di commento extradiegetico (il che è esattamente l'oggetto dell'ironia di Allen). Biancorosso afferma che le configurazioni ambigue non sono vaghe quanto multivalenti: la vera ambiguità si avrebbe qualora un pezzo di musica fosse interpretabile come diegetico ed extradiegetico allo stesso tempo (2009: 11-16).

Una delle revisioni più radicali del modello diegesi è offerta da Winters, secondo il quale la musica che è solitamente indicata come extradiegetica è tale solo a volte, per esempio durante le sequenze dei titoli. Nei frequenti casi in cui la musica non è diegetica ma sembra comunque emanare dal mondo narrativo (specificamente dalle azioni e dai sentimenti dei personaggi), dovremmo parlare di «intradiegetic music», che è parte della fabula e non solo dell'intreccio come quella extradiegetica. Questa distinzione è plausibile se abbandoniamo l'idea che il film sia un mezzo realistico: il fatto che il mondo narrativo ricordi quello reale non comporta che il primo condivida tutte le regole del secondo. Anzi siamo inclini a interpretare come diegetici suoni che non potrebbero esserlo da un punto di vista realistico (p.es. suoni nello spazio cosmico), solo perché essi sembrano prodotti da oggetti diegetici; allo stesso modo, se la musica in un dato film sembra emanare dalla diegesi anche se non c'è alcuna fonte concreta e realistica, è legittimo inferire che secondo le regole del mondo narrativo di detto film la musica può emanare direttamente da un personaggio. Winters avanza anche l'idea che quando è difficile ripensare a una scena senza la musica (il che accade spesso, diversamente dal caso della visualizzazione mentale della stessa con differenti inquadrature), è perché la musica è parte della fabula e dunque è intradiegetica (2010: 224-244).

Yacavone asserisce che «on many [...] models of the diegetic/nondiegetic distinction, there is a denoted or narrated world and a space outside of it», che «is the site of various narrating and denoting processes». L'autore suggerisce che «there is, instead, a denoted world-reality, conveying and bound up with a story or stories taking place within it», il quale «constitutes but one circumscribed area, or level, within a greater experiential whole» che lui chiama «the world of a film work, as distinct from the represented world within it» il quale include «many other levels – and spaces or gaps between them – besides represented (denoted) and narrative ones». Perciò Yacavone distingue tra «a film's more distinctly aesthetic, presentational features» e «its representational ones» (2012: 35-36). Questi aspetti interagiscono costantemente, e «the possible representational (denotative), narrative (story), and expressive (affective) uses of music are not mutually exclusive»; inoltre «we may attend to aspects of both the how and the what of a representation, and of narration, simultaneously, while still distinguishing between the two» (2012: 29, 33).

Heldt sostiene che «the problem with [the diegetic/nondiegetic] distinction is not that it is simplistic, but (a) that it has sometimes been used simplistically, and (b) that it is but one element in a bigger toolbox». Perciò l'autore tenta di studiare il modello diegesi con maggiore profondità è di «expand the range of narratological concepts [most importantly focalization] applied to film music» (2013: II.i). Le idee di Heldt più rilevanti per la disamina che proporrò sono le seguenti: 1) «The search for a terminological system accounting for all the cases is pointless. [...] Concepts need to point out salient features of a phenomenon, but that does not preclude those features from becoming the stuff of creative play»; 2) «anything can happen to the music at any point in a film; the borderline [between diegetic and nondiegetic] is only conceptual and offers no resistance to being crossed»; 3) è necessario enfatizzare la funzione della musica – assieme alla sua collocazione – che nel caso di Heldt è saminata da un punto di vista strettamente narratologico.

Anche alcuni/e studiosi/e italiani hanno affrontato la questione. Cano e Cremonini colgono le dinamiche spaziali del suono, proponendo la distinzione in campo/fuori campo e intradiegetico/extradiegetico (il secondo tipo sempre fuori campo), ma sostengono che la necessità della vista per determinare la collocazione della fonte «chiarisce la imprescindibilità delle immagini nella definizione di una prospettiva sonora» (1995: 25-26). Essi, seguendo Jost, considerano anche l'esistenza dell'auricularizzazione (o “focalizzazione auditiva”) e del punto di ascolto (1995: 27-29).

Ramaglia propone tre approcci alla relazione audiovisiva:

**diegetico** (analisi del linguaggio audiovisivo in rapporto al racconto), **sincronico** (analisi del comportamento del suono rispetto all'immagine nell'ambito temporale ristretto di una determinata sequenza), **diacronico** (analisi delle ricorrenze sonore e della struttura audiovisiva di un intero film) (2004: 5).

In base al primo approccio Ramaglia ripropone i consueti termini di «in-off-over» e «diegetico-extradiegetico» con l'aggiunta di alcune categorie secondarie, interessanti perché definiscono alcuni casi intermedi: 1) *irrealtà diegetica*: vi appartengono suoni non fisici, che non hanno una fonte visibile, perché provenienti dalla mente del personaggio (in questo caso definiti da altri «suoni diegetici interni» o «interni soggettivi», come abbiamo visto poc'anzi) oppure perché paranormali; 2) *ambiguità diegetica*: i suoni non sono riconducibili chiaramente alla diegesi né al di fuori di essa; l'ambiguità può essere infine sciolta o meno, producendo a seconda dei casi tensione, ilarità, straniamento; 3) *assurdo diegetico*: si infrange la logica che governa la diegesi (p.es. la reazione di un personaggio a un suono extradiegetico il quale comunque non cessa di essere percepito dal

pubblico come tale) (2004: 14-19). Nell'interazione del suono con le immagini, Ramaglia distingue tra parallelismo, contrappunto e indifferenza. Nel parallelismo il sonoro «si amalgama meticolosamente sul [visivo], incarnando una funzione di accompagnamento [...] a sostegno delle immagini» (il Mickey Mousing è l'esempio più estremo) (2004: 23). Il contrappunto ha sempre un ruolo attivo e può adempiere diverse funzioni (referenziale, interpretativa, straniante, ironico-grottesca, estetica, provocatoria) (2004: 31-34). Sia il parallelismo sia il contrappunto possono essere dinamico-ritmici o convenzionali: il secondo termine indica i casi in cui essi sono basati su convenzioni radicate (concernenti la geografia, la storia, il timbro, ecc.) e non su aspetti meramente percettivi (2004: 23-27). L'indifferenza indica una combinazione audiovisiva che è percepita come imprevedibile e casuale, da cui l'effetto anempatico (in questo caso il suono è solitamente – ma non necessariamente – diegetico) (2004: 30, 50; cfr. Chion 2001: 18; 2009: 430-433).

Bandirali, ispirandosi a Metz e a Sergio Bassetti, distingue tra due funzioni del suono, la *denotazione* e la *connotazione*: «un suono “denota” qualcosa quando rimanda a una “presenza” del campo narrativo, ossia dell'insieme degli esistenti [...]. Per quanto attiene alla connotazione, ossia al senso espressivo-simbolico, ciascun suono denotato può connotare o essere connotato» (2006: 44-45). Da qui l'autore ricava una seconda distinzione, quella tra «un suono del mondo rappresentato [ovvero] della storia narrata, dunque *storico*» e «un suono che appartiene esclusivamente al piano dello stile, ossia della messa in forma di questo mondo rappresentato [...] dunque *discorsivo*»; sui suoni storici è possibile fare delle “operazioni discorsive” più o meno forti (2006: 47-48). Da questa tesi Bandirali produce la tabella 2.

	Parola	Rumore	Musica
Storico	dialogo dei personaggi, voce interiore	rumore reale di scena o rumore mentale (pensato o ricordato)	musica eseguita <i>live</i> nella scena o riprodotta elettronicamente (radio, lettore cd)
Discorsivo	voce over	rumore iper-reale non prodotto nella scena (ad esempio rumore che accompagna il movimento di macchina), o rumore inserito in partitura	commento musicale e repertorio over

Tabella 2: la teoria di Bandirali (2006: 47-48).

A proposito della «“collocazione” presunta della *fonte di emanazione*» della musica, Miceli distingue tre livelli, che costituiscono le funzioni drammaturgiche primarie della musica:

1- *livello interno*: «[l']evento musicale [è] prodotto nel contesto narrativo della scena/sequenza. La sua presenza può essere *manifesta* oppure *dedotta dal contesto*» (2009: 643);

2- *livello esterno*: l'evento musicale «*non [è] prodotto all'interno della narrazione [pertanto] non [è] condiviso da personaggi e spettatori, bensì indirizzato esclusivamente a questi ultimi*»; questo livello «*prescinde del tutto dalla verosimiglianza [e] si manifesta come puro artificio*» (2009: 649); esso può essere a sua volta di due tipi:

a- *acritico*, quando la musica è «in perfetta sintonia con le istanze narrative e con le relative atmosfere[;] l'intervento musicale, benché atto interpretativo, si limita a confermare e rafforzare l'espressività dell'episodio» (2009: 650-651);

b- *critico*, quando la musica «commenta l'episodio filmico con soluzioni discordanti generando così un corto circuito semantico che nega le aspettative dello spettatore, spingendolo a esercitare un ruolo attivo, ovvero *interpretativo*» (2009: 652);

3- *livello mediato*: «si ha il privilegio di accedere alle sensazioni musicali che il personaggio “ascolta” dentro di sé, sulla spinta di un ricordo o di un'emozione» (2009: 655; questo livello è paragonabile alla musica metadiegetica di Gorbman 1987: 197-198).

Tra questi livelli si possono avere delle transizioni e casi di ubiquità (2009: 657-633).

Cecchi propone una revisione critica della distinzione diegetico/extradiegetico rintracciandone l'aspetto problematico: la frequente incertezza circa la collocazione di dati suoni in relazione alla diegesi è sufficiente a dimostrare che il modello è malfermo (2010: 6). Cecchi non nega l'interesse nella determinazione della collocazione della fonte sonora – che è un aspetto «raffigurativo» – ma tuttavia ne limita l'importanza a favore dell'aspetto «funzionale», ovvero del «momento costruttivo» che dovrebbe essere centrale in una teoria dell'audiovisione (2010: 7-8).

## 3 – Studi cognitivisti sul cinema

### 3.1 – Profilo generale

Intendo ora sintetizzare alcuni studi cognitivisti rivolti al cinema, a opera di studiosi/e di cinema o di scienze cognitive. Nel primo paragrafo vi sono quelli utili per inquadrare l'approccio neurofilmologico, mentre nel secondo affronto quelli dedicati alla relazione tra suoni (quasi esclusivamente musicali) e immagini.

#### 3.1.1 – Teoria del cinema: dal cognitivismo classico alla Neurofilmologia

L'apripista degli studi cognitivisti sul cinema è Bordwell (1985a), il quale però parla allora di costruttivismo più che di cognitivismo, come ancora in Bordwell (1989b). L'autore fa riferimento sia ai processi *top-down* sia a quelli *bottom-up*, ma i primi sono predominanti nella sua disamina. Le tesi fondamentali sono due: secondo la prima, la narrazione è un processo che prevede l'elaborazione mentale da parte di fruitori e fruitrici di una fabula a partire da un intreccio plasmato concretamente dallo stile, ossia dall'organizzazione di tecniche cinematografiche. Questo processo avviene grazie a indizi forniti dal film sui quali chi fruisce applica degli schemi mentali al fine di produrre inferenze utili alla comprensione del film (Bordwell 1985a). La seconda tesi è che i significati (referenziali, espliciti, impliciti, sintomatici) non sono trovati nel film ma sono costruiti mentalmente da chi fruisce (Bordwell 1989b).

Le pagine di Bordwell (1985a) dedicate allo spazio filmico restano a oggi un punto di riferimento imprescindibile e virtuoso. Egli non considera chi fruisce come «the apex of a literal or metaphorical pyramid of vision», bensì come un soggetto che costruisce lo spazio dinamicamente (Bordwell 1985a: 99). Bordwell si distanzia sia dalla teoria “prospettivista”, rappresentata principalmente da Gibson, sia dalla corrente della Gestalt associata con Arnheim: egli propone piuttosto un approccio costruttivista, in cui si rileva comunque la compresenza di processi *bottom-up* e *top-down*. L'assunto principale è che lo stile fornisce indizi sullo spazio, i quali sono processati dal/dalla percipiente: infatti «any pictorial representation is inherently incomplete and potentially ambiguous» (Bordwell 1985a: 100-101).

Bordwell sostiene che il suo oggetto d'esame è lo spazio *scenografico*, costruito in base a tre insiemi di indizi: lo spazio *inquadrato*, lo spazio del *montaggio* e lo spazio *sonoro* (1985a: 113). Circa quest'ultimo, l'autore rileva la presenza della dicotomia figura/sfondo e della prospettiva

sonora, la quale prevede tradizionalmente che il dialogo sia in primo piano e dunque che non ci siano i drastici cambi di prospettiva rilevabili visivamente. Ciò funziona perché «for spatial information, sight outranks hearing in the human sensory system. While auditory frequency, amplitude, and timbre can approximately locate a sound source, determining exact and distance is more difficult than with vision». Di solito quindi ci affidiamo prevalentemente alla vista, sicché è questa a fornire l'informazione spaziale principale, sebbene a volte la prospettiva sonora venga enfatizzata. In ogni caso Bordwell chiarisce che «the “spatiality” of sound on film is as manipulated as the image» (1985a: 118-119). Lo studioso considera lo spazio sonoro anche nel suo capitolo di *The Classical Hollywood Cinema* dedicato allo spazio, asserendo che «in the Hollywood cinema the space constructed by the soundtrack is no less artificial than that of the image» (1985b: 54).

Le zone di spazio off sono di due tipi: quelle extradiegetiche e quelle diegetiche. Circa lo spazio off diegetico, Bordwell spiega che «it would be extremely inefficient [...] to project and recall all areas of offscreen space at every moment. Therefore the viewer bets that only certain offscreen areas will become narratively significant and she or he attends to cues that reinforce or dispute that». Al sonoro è attribuita una forte capacità di fornire indizi sullo spazio off, a partire dal consueto ricorso al suono d'ambiente «to suggest a vague but consistent world offscreen». Inoltre ci può essere un gioco sottile «between the fabula world and an indefinite spot we call “sound over”, from which nondiegetic music and commentary issue», il cui risultato è talvolta ambiguo (1985a: 120).

Bordwell getta esplicitamente le basi per un approccio cognitivista al cinema in un articolo del 1989 (1989a; cfr. Bordwell 1990). *Post-Theory* (Bordwell e Carroll 1996) sancisce l'esistenza di una corrente cognitivista di studi sul cinema, in opposizione alla “Grand Theory” (o “the orthodox view”; cfr. Currie 1999) dominante nel ventennio precedente e dipendente da tesi derivate dal marxismo (specialmente quello di Althusser), dalla psicoanalisi (specialmente quella di Lacan) e dalla semiotica (specialmente quella di Barthes; cfr. la precedente critica di Carroll 1988). Il cognitivismo non è tanto una nuova teoria quanto un nuovo approccio, influenzato dalla filosofia analitica (di cui Carroll è un illustre esponente) e che attinge a discipline quali la psicologia e la fisiologia; esso intende esaminare soprattutto i processi mentali e fisiologici attivati durante la visione dei film e più in generale si propone di studiare problemi specifici e delimitati col rigore del metodo scientifico, l'affidamento a dati empirici e sperimentali, e liberandosi dal fardello di una grande teoria onnicomprensiva (cfr. Bordwell 1989a; Bordwell 1996; Carroll 1996; Currie 1999; Plantinga 2002). Si riscontra comunque anche la posizione leggermente diversa di Buckland (2000), che si inserisce nel solco di quanto lui chiama “cognitive film semiotics”, una corrente cognitivista

non pura come quella statunitense bensì sviluppatasi in Europa dalla semiotica del cinema di Metz e rappresentata da Casetti, Odin, Colin e Chateau (Buckland 2000: 3).

In tempi più recenti, Bordwell (2010) rileva l'esistenza di tre famiglie di teorie filmiche. La prima è basata sul concetto di *comunicazione*, per cui la rappresentazione cinematografica è «a process whereby a message is passed from a sender to a receiver» (Bordwell 2010: 2). La seconda intende la rappresentazione come una *significazione*, ossia «a public, impersonal set of systems or structures», e si è sviluppata nell'ambito di strutturalismo e post-strutturalismo (2010: 3-4). Secondo ambedue le famiglie, «spectatorial engagement is a matter of meaning». Bordwell individua una terza via, per lui maggiormente proficua, definita *empirica-esperienziale*, in cui si rilevano prospettive cognitive e naturalistiche e in cui si assume che il risultato della fruizione del film, intesa come processo che attiva spettatori e spettatrici, non abbia come risultato «a message or a passage among signifiers, but an experience» (2010: 5). L'autore sottolinea l'esistenza di «a large dose of “bottom-up” processes» di cui occorre tener conto, e rifiuta il relativismo culturale proprio del post-strutturalismo a favore di «the assumption that a great deal of the film experience draws on non-filmic capacities and skills, many of which are acquired in ordinary experience» (2010: 6-7; cfr. Anderson 1996; Levin e Simons 2000; Smith, Levin e Cutting 2012; Zacks 2015).

Questa riflessione di Bordwell mi sembra consequenziale a un aggiornamento degli studi cognitivisti sul cinema, rispetto al cognitivismo classico, in cui hanno giocato un ruolo determinante sia l'attenzione ai processi percettivi e a quelli emotivi sia le neuroscienze cognitive (cfr. Currie 1995; Smith 1995; Anderson 1996; Tan 1996; Plantinga e Smith 1999; Smith 2003; Grodal 1999, 2015; Plantinga 2009; De Vincenti e Carocci 2012; Carocci 2012). Inoltre gli anni Novanta vedono sia la scoperta dei neuroni specchio sia lo sviluppo della risonanza magnetica funzionale (fMRI), due avvenimenti che hanno segnato la ricerca neurofisiologica, mentre nell'attuale millennio le nozioni di neuroscienze hanno influenzato la riflessione sulle arti e sull'estetica, e quindi anche sul cinema (cfr. Grabowski 2015), e le tecniche neuroscientifiche sono state applicate alla ricerca sulla fruizione filmica. Il primo esperimento neuroscientifico di visione di immagini filmate di cui sono a conoscenza è quello di Stoléro et al. (1999), che hanno usato la tomografia a emissione di positroni per studiare le aree del cervello attivate in uomini durante la visione di filmati pornografici. In questo caso tuttavia l'intento era quello di analizzare la risposta a uno stimolo sessuale, a prescindere dal fatto che esso fosse filmato; insomma il filmato era un sostituto di uno stimolo sessuale non mediato, il che è riscontrabile in diversi esperimenti di tema affine (cfr. Cohen, Rosen e Goldstein 1985; Redouté et al. 2000; Mouras et al. 2003; Ferretti et al. 2005; Moulrier et al. 2006; Mouras et al. 2008; Kühn e Gallinat 2011). Il lavoro di Hasson et al. (2004) si è invece rivelato pionieristico perché lascia trasparire l'interesse dei/delle neuroscienziati/e nella realizzazione di

esperimenti per comprendere le risposte neurofisiologiche di chi guarda un film in senso più generale. Qui l'intento è quello di rispondere alla seguente domanda: «To what extent do all brains work alike during natural conditions?» (2004: 1634). Il risultato della fMRI applicata a soggetti durante la visione di una porzione di *Il buono, il brutto, il cattivo* (Italia/Germania Ovest/Spagna/USA 1966) suggerisce una sincronizzazione intersoggettiva dell'attività corticale, poiché «brains of different individuals show a highly significant tendency to act in unison during free viewing of a complex scene such as a movie sequence» (2004: 1638). Il film sembra dunque utilizzato nell'esperimento come un sostituto di un qualunque stimolo visivo, ma ne consegue una ricaduta esplicita sulla teoria del cinema: infatti Hasson et al. (2008b) intervengono sulle pagine di *Projections* per proporre il paradigma della *inter-subject correlation of brain activity* e il nuovo approccio chiamato “neurocinematics” (cfr. Hasson et al. 2008a; cfr. anche l'esperimento psicologico sulla memoria in Furman et al. 2007). Questo approccio neuroscientifico è condiviso da ricercatori di tale ambito: si confrontino la “brain theory of cinema” di Dudai (2008), lo studio di “evolutionary cinematics” sui primati a opera di Ghazanfar e Shepherd (2011), e gli studi di *affective neuroscience* applicati al cinema da Raz e colleghi (cfr. Gilroy et al. 2013; Raz, Hagin e Hendler 2013; Raz et al. 2014; Raz e Hendler 2014; Raz et al. 2016). Shimamura (2013) parla invece di “psychocinematics”, e si riferisce a un approccio scientifico interdisciplinare all'esperienza filmica considerando anzitutto aspetti psicologici e biologici (percettivi, cognitivi, emotivi) e quindi la risposta estetica di chi fruisce, basandosi su ricerche empiriche e sperimentali (p.es., anche qui, con la fMRI). L'autore fa anche qualche cenno al suono, in relazione alla forma e alle emozioni, tuttavia ricorre a una terminologia impropria (il termine «nondiegetic» significherebbe “esterno alla scena” piuttosto che al mondo narrativo, e «musical soundtrack» significherebbe “musica extradiegetica”) (Shimamura 2013: 8).

D'Aloia e Eugeni hanno coniato invece il termine Neurofilmologia, che trovo particolarmente convincente. Per filmologia s'intende «the research program that during the 1950s and the 1960s intersected for the first time theoretical and empirical approaches within a systematic investigation of film viewing experience» (2014: 8), il che dunque non esclude a priori l'apporto di alcuna disciplina. Il prefisso “neuro-” indica il nuovo contributo delle neuroscienze, sia teoretico sia sperimentale. In questo modo gli autori propongono una soluzione dell'opposizione tra i due modelli di spettatore rilevabili nelle teorie del cinema: da un lato il *viewer-as-mind* dell'approccio cognitivista/analitico, in cui l'enfasi è sui processi mentali, e dall'altro il *viewer-as-body* dell'approccio fenomenologico/continentale, in cui l'enfasi è sui processi sensitivi, affettivi e motori (cfr. p.es. Sobchack 2004). Il modello proposto dagli autori è quindi quello del *viewer-as-organism*, sottolineando l'interazione tra mente e corpo: in questo modo si tiene conto di «many

processes of different nature (sensory, perceptual, cognitive, emotional, motor, active, mnemonic)», e si conferisce la medesima attenzione sia ai processi *bottom-up* sia a quelli *top-down* (2014: 7, 16-17; un interessante esempio di un simile approccio è quello di Hogan 2009). In ogni caso tutto ciò non comporta una svalutazione del ruolo della cultura, come Dudai dimostra esemplarmente: «brain made cognition, cognition made culture, culture made film, film makes cognition» (2008: 23).

Il modello dello/della spettatore/trice come organismo è coerente con la teoria sviluppata in un recente libro (2015) dallo studioso di cinema Michele Guerra insieme a Vittorio Gallese, neuroscienziato tra gli scopritori dei neuroni specchio che si dedica anche allo studio della fruizione delle opere artistiche e in particolare del cinema (cfr. Freedberg e Gallese 2007; Gallese e Guerra 2012; 2013; 2014). Gli autori parlano di *sistema cervello-corpo* (Gallese e Guerra 2015: 36), collocandosi tra i sostenitori della teoria della cognizione incarnata e rifiutando sia il cognitivismo classico – che suppone il passaggio unidirezionale dalla percezione alla cognizione all’azione – sia la tesi, proposta dalla psicologia evoluzionistica, dell’esistenza di moduli cognitivi dotati di specifica funzione (2015: 37-45). In particolare Gallese e Guerra sviluppano la tesi della *simulazione incarnata*,

un meccanismo funzionale di base del nostro cervello grazie al quale riusciamo parte delle risorse neurali che normalmente utilizziamo per interagire col mondo, modellando i rapporti e le relazioni che con esso stabiliamo, mettendole al servizio della sua percezione e immaginazione [...]. Comprendiamo il senso di molti dei comportamenti e delle esperienze altrui grazie al *riuso* degli stessi circuiti neurali su cui si fondano le nostre esperienze agentive, emozionali e sensoriali in prima persona (2015: 24).

Gli autori descrivono quindi alcune forme di simulazione incarnata. La prima concerne la mappatura dello spazio peripersonale (a portata di mano, *zuhanden*), la seconda e la terza riguardano l’attivazione rispettivamente dei neuroni canonici e dei neuroni specchio, la quarta pertiene ai meccanismi di rispecchiamento di sensazioni ed emozioni. La simulazione incarnata si inserisce nell’ipotesi della *shared manifold* (“sistema della molteplicità condivisa”), già avanzata in passato da Gallese. Questo sistema ha tre livelli operazionali: il *livello fenomenologico*, che concerne l’empatia intesa in senso lato, come comprensione di azioni, emozioni e sensazioni altrui permessa dalla condivisione; il *livello funzionale*, che concerne la simulazione incarnata; il *livello sub-personale*, che risulta dall’attività di circuiti neuronali (2015: 292; cfr. Gallese 2001). Tutto questo porta a una nuova visione dell’intersoggettività che ha come elemento cardine la *cognizione motoria* prima che i processi deduttivi e inferenziali considerati dalla Teoria della Mente, portando a

una nuova definizione dell'empatia (Gallese e Guerra 2015: 26-7, 287-288). Gli autori parlano anche di *simulazione liberata*, specifica della persona umana poiché deriva dalla commistione tra simbolico e corporeo (2015: 75). Nel libro Gallese e Guerra affrontano alcuni aspetti dello stile cinematografico, ossia la costruzione di sguardi, i movimenti di macchina, il montaggio, il primo piano, avvalendosi anche di dati sperimentali (cfr. Heimann et al. 2014; 2016). A questo proposito gli autori avevano già affermato che, malgrado il cinema fosse stato usato dalle neuroscienze «in modo soprattutto strumentale», ossia «più come un mero stimolo che non come un oggetto di analisi in senso stretto», il film è anche «un valido strumento di indagine per la ricerca neuroscientifica», la quale può «aprire orizzonti importanti sui temi più filosoficamente dibattuti» (Gallese e Guerra 2013: 12; cfr. Gallese e Guerra 2015: 87-88). Nel capitolo 5 integrerò nella loro teoria la mia disamina sul sonoro, pertanto approfondirò svariate delle questioni qui accennate.

Il recente libro di Smith (2017), *Film, Art, and the Third Culture*, è a mio parere un testo di riferimento filosofico obbligato per chi come me intende ricorrere a un approccio cognitivista, perciò intendo dedicarvi particolare attenzione. L'autore affronta l'incontro tra quelle che sono ritenute le due grandi famiglie per eccellenza della cultura umana: le scienze umane e le scienze naturali. L'autore si ispira in particolare a C. P. Snow, che parlava proprio di «two cultures» proponendo un superamento di questa rigida dicotomia. Smith a sua volta sostiene la necessità di elaborare una *terza cultura* che egli denomina *naturalismo*, in cui la speculazione sui fatti culturali dell'umanità si pone in costante integrazione con le rilevazioni empiriche e con il rigore metodologico delle scienze naturali. L'autore individua due tendenze associabili al naturalismo, quella del *replacement naturalism* e quella del *cooperative naturalism*. La prima tendenza, più radicale, mira a studiare i fenomeni culturali rimpiazzando l'approccio esistente e ricorrendo solo alle scienze naturali. La seconda tendenza, promossa da Smith, intende invece integrare i vari saperi e i vari metodi, ma ricorrendo sempre a un atteggiamento propriamente scientifico, cioè basato sul rigore metodologico e su una maggiore formalizzazione. Il naturalismo si contrappone in questo senso alle due forme di *autonomismo* che costituirebbero tuttora l'approccio dominante nelle scienze umane: quello più rigoroso, cioè la totale esclusione delle scienze naturali, e quello meno rigoroso ma basato su un ricorso furbo o superficiale ad alcune nozioni di dette scienze.

L'oggetto di indagine di Smith è costituito dagli audiovisivi narrativi, cioè film tradizionalmente intesi e serie televisive. La fruizione di film è concepita dall'autore come un'esperienza estetica multimodale (essendo anzitutto visiva e sonora, ed eventualmente tattile) che espande quella ordinaria e che noi non ci limitiamo a vivere ma che *assaporiamo* in modo più o meno consapevole. Il metodo naturalistico di Smith è elaborato in modo *piecemeal*, cioè affrontando di volta in volta problemi specifici della fruizione filmica, in ossequio implicito alla

proposta di Bordwell e Carroll (1996). L'autore mostra un'attenzione specifica verso la teoria evuzionistica e le neuroscienze, da cui deriva il principio cardine della *triangolazione* dell'analisi: si deve tenere conto dell'interrelazione tra dati *fenomenologici*, *psicologici* e *neuronal*, in considerazione della natura incarnata della cognizione, dei limiti fisiologici della razionalità e dei meccanismi sub-personali (ne è prova il fatto che l'architettura cognitiva è plasmata dall'architettura neuronale). Smith dimostra che è possibile continuare a costruire teorie e a fornire spiegazioni con densità di argomentazioni conformemente alle aspettative delle scienze umane, e che è possibile analizzare i casi particolari pur partendo da alcuni assunti generali e universali. Insomma Smith si tiene lontano non solo dal culturalismo e dal relativismo radicale ma anche da posizioni che ritiene ugualmente difettose come quella del behaviorismo neuronale e più in generale dello scientismo.

A questo proposito Smith è molto vicino a Carroll (1996), poiché pur con tutta la sua fiducia nella scienza ne riconosce ugualmente la fallibilità. I saperi sono sempre soggetti a revisione, e per giunta la scienza è inevitabilmente contaminata dalla cultura e dai valori; cionondimeno ci si deve sforzare di ricercare la verità, col rigore del metodo e la forza dei dati e delle prove sperimentali. Non si escludono dunque questioni sociali e di analisi critica, anche perché Smith ritiene che i fattori socioculturali siano intrinseci alla *natura* umana. Il conflitto tra scientismo e culturalismo è perciò risolto, in chiave naturalista, con un approccio *bioculturale*, grazie al quale si indaga il substrato biologico dei fenomeni culturali ricorrendo a nozioni e metodi delle scienze naturali: la cultura è studiata a partire dalla biologia, come suo *byproduct*. Smith previene eventuali accuse di riduzionismo, determinismo e reazionismo (come già Carroll 1996), mostrando che porsi il problema della potenzialità biologiche umane sottese alla cultura non implica un determinismo biologico tale da minare il libero arbitrio, anzi può capitare che natura e cultura siano in conflitto. Insomma i singoli fenomeno culturali – e dunque, in questo contesto, i singoli film – possono essere studiati partendo dal presupposto di una *metacultura*, ovvero di costanti universali a base biologica; ad ogni modo, oltre a riconoscere le specificità culturali, il naturalismo cooperativo ammette che vi sono fatti che travalicano le competenze delle scienze naturali, diversamente dai naturalisti intransigenti.

Smith applica il suo metodo a diversi aspetti pertinenti all'esperienza filmica quali il riflesso di trasalimento, la coscienza, l'effetto Kulešov, e soprattutto l'emozione e l'empatia, confutando potenziali contestazioni di universalismo acritico. Ne è un esempio eclatante il fatto che i film selezionati sono non soltanto hollywoodiani – cioè opere di massa che pertanto si prefiggono di essere universali, avallando più facilmente le tesi bioculturali – ma provengono anche da altri paesi e quindi da altre culture, delle quali si possono valorizzare le dinamiche peculiari derivanti da quelle

generali in quanto biologiche. Tuttavia per convincere i culturalisti più rigidi a una maggiore apertura alle scienze naturali avrebbe forse giovato includere maggiori riferimenti alla letteratura ortodossa, che io mi sono impegnato a presentare qui oggettivamente in relazione al tema da me trattato.

### 3.1.2 – Scienze cognitive

Nell'attuale millennio si riscontra un incremento anche degli studi psicologici teorici e sperimentali sul cinema che non pertengono solo alla percezione ma anche alla cognizione narrativa, come dimostrano per esempio Tim J. Smith (Smith 2005, 2012; Smith e Henderson 2008; Rayner et al. 2009; Võ et al. 2012; Smith e Mital 2013), Jeffrey Zacks (Zacks et al. 2006; Zacks, Speer e Reynolds 2009; Zacks 2015), e soprattutto James E. Cutting, Daniel T. Levin e Joseph P. Magliano<sup>42</sup>. Cutting si era già occupato di cinema negli anni '80 affrontando un problema percettivo, ossia la visione da angolazioni non ottimali (Cutting 1986; 1987). Negli ultimi anni Cutting e colleghi/e si sono occupati/e ripetutamente di cinema (soprattutto statunitense), affrontando varie questioni sia percettive sia cognitive attraverso rilevazioni quantitative empiriche e sperimentali: i meccanismi attentivi (Cutting, DeLong e Nothelfer 2010), la luminosità (Cutting et al. 2011; Cutting 2014b), il movimento (Cutting, DeLong e Brunick 2011; Cutting 2014b), la dissolvenza (Cutting, Brunick e DeLong 2011), il modo di inquadrare i personaggi (Cutting 2015), la durata delle inquadrature e il ritmo del montaggio (Nothelfer, DeLong e Cutting 2009; Cutting et al. 2011; Cutting e Candan 2015; Cutting 2016b), la memoria visiva (Candan, Cutting e DeLong 2015), la percezione delle espressioni facciali (Cutting e Armstrong 2016), l'interazione tra racconto e stile (Cutting, Brunick e DeLong 2011; Cutting 2016a), la segmentazione degli eventi narrativi (Cutting, Brunick e Candan 2012; Cutting 2014a), le applicazioni terapeutiche del cinema (Armstrong e Cutting 2017), la cognizione temporale (Armstrong e Cutting 2016) e quella spaziale (Cutting, Iricinschi e Brunick 2013; Cutting e Iricinschi 2015). L'indagine è sullo stimolo visivo (cfr. anche Smith, Levin e Cutting 2012; DeLong, Brunick e Cutting 2014; Brunick, Cutting e DeLong 2013), ancorché quello sonoro non sia del tutto ignorato: per esempio Cutting e Candan (2013: 28) alludono al ruolo della colonna sonora nella resa della continuità.

Cutting, Iricinschi e Brunick propongono un metodo di mappatura puramente astratta dello spazio filmico concernente le co-occorrenze dei personaggi nelle scene (o nelle inquadrature). Dall'indagine gli autori traggono quattro punti. Il primo è che questo metodo, diversamente da

---

<sup>42</sup> Si confronti anche la tesi di dottorato di Hale, contenente esperimenti atti a «investigate whether schema theory has value for researching and designing film experiences» (2009: 2).

ricerche precedenti, permette di sganciarsi dallo spazio fisico pur mantenendo l'oggettività. Il secondo consiste nella rilevazione che la presentazione dei personaggi in un film è in accordo con le esigenze del racconto. Il terzo è che il metodo contribuisce all'analisi visiva dei film. Il quarto è che, in considerazione del secondo punto e dell'apprendimento statistico proprio degli individui, «these maps should bear some recognizable similarity to the relations among characters in the minds of viewers, and by themselves could guide some coarse understanding of the film without a necessary comprehension of verbal events across the film» (Cutting, Iricinschi e Brunick 2013: 100-103). Cutting e Iricinschi si sono invece occupati del *re-establishing shot* e il suo ruolo nella segmentazione, all'interno di una riflessione sull'elaborazione cognitiva delle informazioni narrative (ma solo sul piano visivo). Le conclusioni cui giungono gli autori sono quattro: «(a) re-establishing shots include less background than establishing shots; but (b) more background than the average of remaining shots of a scene or subscene; and (c) re-establishing shots are shorter in duration than establishing shots, but (d) longer than the remaining shots». Tutto questo favorisce in chi guarda la rilevazione e la memorizzazione dello spazio in cui agiscono i personaggi (Cutting e Iricinschi 2014: 450).

Levin si è occupato sin dagli anni '90 di *inattentional blindness*, in particolare a partire dall'articolo di Simons e Levin (1997; cfr. Mack et al. 1992), facendo riferimento al cinema e usando filmati per esperimenti (cfr. Levin e Simons 1997; cfr. anche Simons e Chabris 1999). Ciò pertiene anche alla cognizione spaziale, nel cinema e nella realtà, come dimostra lo studio di Levin e Wang (2009) dedicato specificamente allo spazio filmico. Gli autori riprendono la tesi di Levin e Simons che confermano l'affinità tra i processi percettivi di eventi reali e mediati soffermandosi su «a common tendency to perceive events as continuous in the face of large view-to-view inconsistencies» (2000: 357). Dunque sia nella realtà sia nel cinema siamo spesso incapaci di individuare cambiamenti visivi, e la percezione di continuità si basa su inferenze. Levin e Wang sostengono che rappresentazioni spaziali, determinate dalla direzione dello sguardo e dalla geometria spaziale, sono create automaticamente, e si soffermano anzitutto sulla percezione dello sguardo, cruciale nello stile cinematografico (2009: 28-29). Un dato significativo è che gli spettatori decodificano informazioni spaziali a un livello di astrazione tale che alcune informazioni ingannevoli possono passare inosservate; anche qualora le si noti, ciò che prevale è «consistency in narrative and emotional logic» (2009: 42). Dai dati sperimentali risulta che «viewers often create online representations of the layout of important objects in a scene, and possibly the geometry of the environment containing them». Di solito i sottospazi vengono integrati in casi di *consistent-view sets* ma non di *multiple-view sets*, a meno che ciò non sia rilevante per la comprensione narrativa (2009: 47-48). Gli autori concludono con tre ipotesi. Primo, la decodifica di determinati indizi

spaziali è influenzata da «both by long-term experience and by short-term priorities and cue validities». Secondo, «some spatial processes relate things at specific locations in the environment to people's internal thoughts in a relatively automatic way, but other more complex thought-object relationships require deliberation». Terzo, le rappresentazioni spaziali a lungo termine avvengono solo talvolta, a seconda di quanto i punti di vista siano stabili – perciò «representations of sets where any view was possible appeared quite weak and resistant to experience and memory» (2009: 48-9).

Magliano studia i meccanismi di comprensione narrativa letteraria e visiva sin dagli anni '90. Egli si è dedicato ad argomenti quali le inferenze predittive generate durante la visione di film (Magliano, Dijkstra e Zwann 1996), l'indicizzazione spaziotemporale nella comprensione del film (Magliano, Miller e Zwaan 2001), l'impatto del montaggio in continuità sulla segmentazione di eventi nei film narrativi (Magliano e Zacks 2011; cfr. Zacks, Speer e Reynolds 2009) nonché la segmentazione di eventi continui in prima persona (utilizzando un videogioco; Magliano et al. 2014), il confronto tra i processi cognitivi di racconti scritti e quelli di racconti audiovisivi (Magliano et al. 2013; cfr. anche Zacks, Speer e Reynolds 2009; Magliano et al. 2015), il rapporto tra movimenti oculari e comprensione narrativa del film (Hutson et al. 2014; Hutson et al. 2015; Loschky et al. 2015; Hutson et al. 2016). È palese l'interesse di Magliano nel diffondere simili nozioni anche nel campo degli studi di cinema, come dimostra l'articolo pubblicato su *Projections* (Magliano e Clinton 2016) in cui gli autori riflettono – con encomio – su come *Narration in the Fiction Film* di Bordwell (1985a) ha influenzato le successive teorie di comprensione narrativa basate anche su dati sperimentali.

### **3.2 – La cognizione del sonoro filmico**

La componente sonora dei film che è stata maggiormente studiata è la musica, a mio avviso per due ragioni. La prima è che la musica di commento emerge come un elemento palesemente creativo in quanto puramente espressivo, dunque meritevole di attenzione critica, mentre il trattamento delle voci e il montaggio degli effetti sembrava forse una questione per lo più tecnica e non artistica. La seconda ragione è che la musica è una forma artistica riconosciuta da millenni e che è studiata da una specifica disciplina, la musicologia. I testi di psicologia cognitiva dediti alla relazione audiovisiva nel cinema sono stati anch'essi dedicati soprattutto alla musica. Tra i primi studi si rileva quello di Tannenbaum (1956) – che analizza sperimentalmente il ruolo della musica nella connotazione di un dramma teatrale o televisivo – ma l'autrice più significativa è Annabel J. Cohen, psicologa cognitiva che si occupa della musica nei film sin dagli anni '80. In particolare, Marshall e

Cohen hanno condotto un esperimento che conferma come la musica possa alterare il significato di un dato aspetto di un film riguardo specificamente alla dimensione temporale, qualora lo stimolo musicale sia congruente con quello visivo; in assenza di congruenza, lo stimolo visivo sembra essere preponderante nella costruzione di significati (1988: 108-111).

Cohen ha affrontato spesso l'interazione crossmodale tra la musica e le immagini di un film (Cohen 1990; 1993; 1998; 2002; 2005), fino ai due più recenti articoli (Cohen 2013; 2015) in cui è presente la formulazione più aggiornata di quello che l'autrice chiama *congruence-association model with working narrative* (CAM-WN), ove «*Congruence* focuses on the structure of music, which can overlap with structure in other sensory domains» mentre «*association* focuses on the meanings that music can bring to mind» (2013: 176; cfr. 2015: 10). Cohen propone un'altra utile coppia concettuale: con *internal semantics* intende «the emotional perspective of the perceiver: how the audience member feels while watching a film, and how the audience member feels about the characters and events depicted in the film»; con *external semantics* intende invece «the regularities of nature and cultural conventions of human interaction and human-environment interaction» (2013: 183). Dall'integrazione di congruenza e associazione con i due tipi di semantica risultano quattro insiemi di fattori. Per fare qualche esempio la congruenza può manifestarsi sul piano della semantica interna nella produzione di una risposta estetica, oppure nel veicolare l'attenzione su qualcosa, anche senza che la musica sia ascoltata coscientemente, mentre sul piano della semantica esterna la congruenza può manifestarsi nell'interazione tra le attività di un personaggio e una musica diegetica; l'associazione può invece realizzarsi sul piano della semantica interna allorché la musica stimoli una particolare reazione emotiva, mentre sul piano della semantica esterna può risultare dalla riconoscimento di convenzioni musicali che evocano specifici contesti o umori (2013: 185).

Il CAM-WN è sintetizzato nella seguente figura, in cui il *working narrative* è inteso come il risultato di due fonti di informazioni: «lower-order multimodal physical/sensory input» di tipo uditivo, visivo e tattile, veicolato dal film nonché in piccola parte risultante dalla propriocezione, e «higher-order knowledge, based on [the audience's] own remembered or stored experience» ovvero la memoria a lungo termine. C'è dunque un'integrazione di processi *bottom-up*, specificamente l'analisi percettiva esogena, e processi *top-down*, ossia le ipotesi endogene sul significato della suddetta (Cohen 2013: 192). Cohen ritiene che la struttura della musica può influenzare direttamente la cognizione dello stimolo visivo, il quale è però preponderante (2013: 176). Questa formulazione del CAM-WN (v. figura 15), il quale ha origine in Marshall e Cohen (1988), ha il vantaggio di dare il giusto peso a ciascuna componente del film e a includere aspetti strutturali della musica che non sono solo temporali.

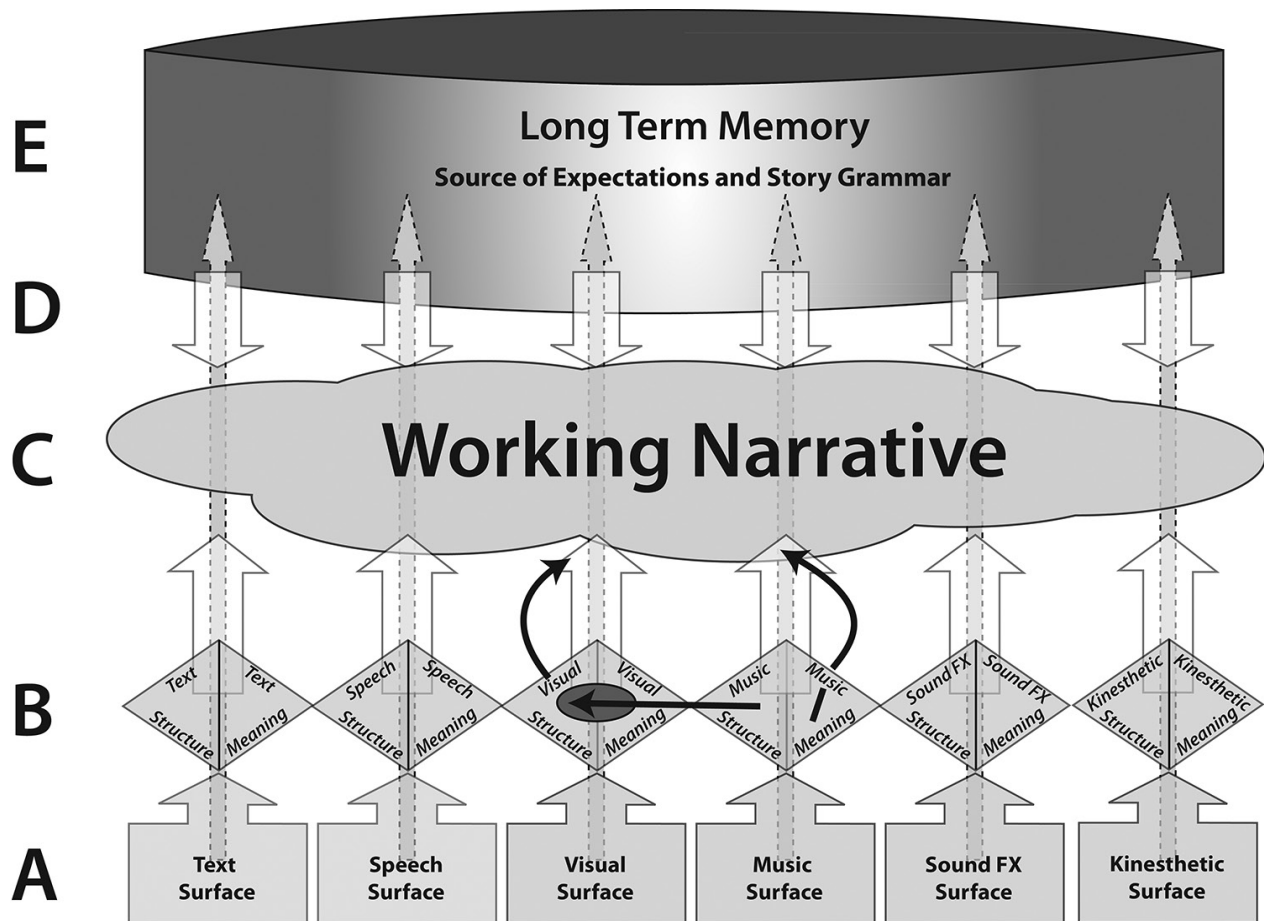


Figura 15: Il CAM-WN di Cohen (2013: 195).

La coppia congruenza/incongruenza tra musica e racconto, relativamente sia alla struttura sia al significato, è cruciale in diversi studi sperimentali. Ciò è riscontrabile, oltre che come accennato in Marshall e Cohen (1988), anche in Boltz, Schulkind e Kantra (1991), che dimostrano il contributo della musica nella memorizzazione degli eventi narrativi e la sua capacità di *foreshadowing* (confermata da Tan, Spackman e Bezdeck [2007], che dimostrano inoltre come la musica influenzi l'interpretazione delle emozioni dei personaggi anche con differimento temporale), e in Bolivar, Cohen e Fentress che rilevano la presenza di una *semantic salience*, poiché non tutte le informazioni catturano l'attenzione allo stesso modo: ancora una volta, sembra esserci il dominio del visivo rispetto al sonoro. In ogni caso gli autori ritengono che «one can effectively predict a priori the effect of music on interpretation of video» (1994: 50). Boltz (2001) ha dimostrato in via sperimentale che la musica può non solo fornire una connotazione e un contributo al coinvolgimento emotivo di un film, ma anche condizionare l'interpretazione degli eventi narrati da parte del/della percipiente, veicolare la sua attenzione su determinati dettagli e favorire la

memorizzazione. Boltz (2004) lo ribadisce e avalla sperimentalmente il *congruence-associationist model* di Cohen (2001), rilevando l'importanza della congruenza del mood nella comprensione e nella memorizzazione delle informazioni audiovisive. Inoltre ella rileva «a visual dominance effect that emerged consistently», ovvero che «film information is more likely to be encoded and subsequently represented within the cognitive system than is music» (Boltz 2004: 1199). In questi studi si ragiona sull'influenza della musica sulle immagini filmiche, assumendo le seconde come lo stimolo principale. Boltz, Ebendorf e Field (2009) analizzano invece come le immagini filmiche possano influenzare la cognizione della musica, ispirandosi a studi di ambito prettamente musicale come quello di Geringer, Cassidy e Byo (1996), mentre Cohen, MacMillan e Drew (2006) indagano l'assorbimento psicologico nel film esaminando finalmente anche le altre due componenti sonore filmiche, ossia effetti e dialogo, e rilevando come la musica svolga un ruolo dominante nell'assorbimento.

Altri studi sperimentali hanno affrontato le emozioni e il ruolo narrativo dello stimolo musicale. Il contributo emotivo fornito dalla musica è ovviamente un dato empirico costantemente rilevato; tra i primi esperimenti scientifici c'è quello di Thayer e Levenson, che dimostra come «musical scores can alter electrodermal responses to a stressful film, thus providing preliminary experimental validation for the efficacy of using music in this manner» (1983: 50). Studi successivi hanno confermato la rilevanza emozionale della musica anche in correlazione al suo contributo alla cognizione narrativa (cfr. Parke, Chew e Kyriakakis 2007; Eschrich, Münte e Altenmüller 2008; Blumstein, Davitian e Kaye 2010; Hoeckner et al. 2011; Costabile e Terman 2013). Una rilevazione particolarmente interessante è quella di Pavlović e Marković, secondo cui «the emotional quality of visual information [...] has a stronger effect than the emotional quality of auditory information (music)» nel caso di qualunque delle emozioni considerate tranne la paura (2011: 84; cfr. Parker e Heerema 2007: 2). Altri studi dedicati al ruolo narrativo della musica hanno esaminato aspetti differenti, per esempio il senso di risoluzione (Thompson, Russo e Sinclair 1994) e la memoria (Furman et al. 2007).

Anche il musicologo Scott D. Lipscomb si è dedicato allo studio della correlazione musica-immagine anche col ricorso a esperimenti. Lipscomb e Kendall (1994) concettualizzano la musica come referenziale o meno e il visivo come concreto o astratto, e dimostrano come la colonna musicale può modificare il significato di un film (v. figura 16). Gli autori conferiscono un valore predominante alla dimensione temporale della musica, ragionando in termini di accenti (cfr. anche Lipscomb 1995, Lipscomb 2005); in ogni caso il valore associativo, ossia di significato, precede l'elaborazione della relazione strutturale tra musica e immagini.



Bordwell 1989b). Il discorso è diverso per i dialoghi e per i rumori, poiché entrambe queste categorie sono normalmente diegetiche e referenziali.

Tre autori che hanno affrontato il sonoro filmico considerando la teoria della simulazione incarnata sono Fahlenbrach, Chattah e Ward (nell'ambito più generale dei *sound studies* cfr. Roddy e Furlong 2014). Fahlenbrach ricorre alla teoria delle metafore concettuali e degli *image schemas* (seguendo Mark Johnson e George Lakoff; un simile approccio alla cognizione filmica sarà sviluppato da Coegnarts e Kravanja 2012; 2014; 2015; 2016; Coegnarts 2015; Coegnarts et al. 2016) per studiare le metafore audiovisive in una prospettiva incarnata, tenendo conto di fattori neurologici, schemi associativi e significati simbolici (Fahlenbrach 2008: 86-87). Ella descrive l'interazione crossmodale di suoni e immagini, ribadendo il predominio della vista e dell'udito nella cognizione rispettivamente di spazio e tempo, da cui deriva la *sincretisi* (Chion 2001: 65-67); inoltre Fahlenbrach individuando in particolare tre qualità nell'integrazione percettiva audiovisiva, ossia «duration (e.g., long/short), intensity (strong/weak), position (above/below, central/ peripheral, close/distant)» (Fahlenbrach 2008: 96).

Chattah si dedica esclusivamente alla musica, asserendo che oltre al significato simbolico di cui può essere portatrice la musica, basato sull'associazione, c'è anche la possibilità dell'esistenza di «a phenomenological score, employing a wide range of techniques to elicit embodied meanings» (2015: 112). Una tesi fondamentale di Chattah è che «Music's potential to engender affective experiences stems from our inhabiting comparable environments, from our innate capacity for embodiment, and most importantly, from our proclivity toward metaphoric reasoning» (2015: 81). Anche Chattah, come Fahlenbrach (2008), ragiona sugli *image schemas*, ossia «abstract-level mental structures that emerge from our perception of objects, events, or concepts» (Chattah 2015: 82). L'autore individua specificamente tre schemi: il *linearity schema*, analizzato in termini di congruenza strutturale e semantica; i *path and expectational schema*, da cui scaturisce congruenza sintattica; il *container schema*, da cui scaturisce la congruenza formale o strutturale. In tutti questi casi l'autore tiene conto dei processi cognitivi incarnati, grazie alla considerazione del sistema umano dei neuroni specchio (2015: 106), mostrandosi in conclusione vicino agli ultimi studi di Cohen.

Ward riconosce che il «visual narrativism» è dominante nei Film Studies, mentre occorre considerare che il suono modifica la percezione visiva e viceversa in un'interazione crossmodale (2015: 158). Ward ribadisce che la vista ha un peso maggiore nella strutturazione dello spazio mentre l'udito ha un peso maggiore nella strutturazione del tempo, ma ad ogni modo entrambi concorrono in modo interrelato alla manipolazione di spazio, tempo e movimento: il cinema è dunque un mezzo multimodale, ma soprattutto è inteso come «an *affective* spatio-temporal system»

(2015: 159). L'autore distingue due tipi di immersione offerta dal film, determinata l'una dal *perceptual design* e l'altra dal *narrative design*. L'immersione percettiva, che incrementa con l'evoluzione tecnologica dei sistemi surround (2015: 163), «inducts core mood and feeling-states within the body of the audience through (mostly) pre-attentional, nonconscious mechanisms which process multimodal information», mentre l'immersione narrativa «structures emotion and meta-emotion through (mostly) attentional, conscious mechanisms». Pertanto la prima è «the abstraction and simulation of physical experience», mentre la seconda è «the abstraction and simulation of social experience». Il risultato della prima è quello che Wojciehowski e Gallese (2011) chiamano “Feeling of Body”, generato dalla simulazione liberata, laddove il risultato della seconda è la “Theory of Mind”. Ward spiega che l'immersione percettiva dà vita all'immersione narrativa come conseguenza della simulazione incarnata, e che «The primary function of these forms of simulation is to produce affective experience» (Ward 2015: 164). L'autore sostiene pertanto che esiste un *cinematic proto-narrative*, «an affective linking of the body's experience of mediated space through time» (2015: 166), così come esiste un realismo percettivo inteso in termini di «manipulation of basic rules of human perception to create coherent units of meaning during sound design» (2015: 157). Ward propone di usare la teoria della simulazione incarnata per integrare altri tre modelli: il BRECVEMA (cfr. Juslin 2013), il *mood-cue approach* di Greg Smith e l'*emotional communication model* di Bartsch e Hübner. In tutti questi modelli la cognizione incarnata delle emozioni ha un ruolo cruciale (Ward 2015: 173-177).

Collins si dedica a videogiochi e realtà virtuali più che al cinema, offrendo validi spunti di analisi per la materia di interesse di questa tesi. Ella riflette sull'esperienza ludica audiovisiva nel quadro della cognizione incarnata, considerando il ruolo dei neuroni specchio e l'empatia verso i personaggi del gioco. L'autrice suggerisce che «encouraging the use of surround sound speakers would likely increase the effects described here, since we are then even more immersed in the sound, our own body in the middle of the peripersonal auditory space» (2011: 7). Collins ha partecipato anche a tre studi sperimentali. Rojas et al. sostengono che «the perception of visual quality of virtual models can be influenced by background sound» (2012: 391). Cullen et al. dimostrano che, grazie a effetti crossmodali, l'audio può avere un impatto sulla percezione della profondità in giochi stereoscopici 3D (S3D), giungendo ad asserire che «if S3D depth contributes to feelings of motion sickness in gamers, sound could be used to increase the sense of depth without implementing extreme S3D settings. In this way, we may be able to reduce the actual S3D depth and compensate for this reduction through the use of sound». (2012: 39). Cullen et al. confermano queste tesi in un esperimento che include l'audio surround 5.1, il quale «appears to increase the

apparent depth within an S3D environment»; pertanto gli autori raccomandano «that game developers consider designing 5.1 surround sound audio for games with S3D in mind» (2013: 12).

## 4 – Teoriche sull'audio spaziale

Un impulso decisivo nella riflessione sullo spazio sonoro viene da R. Murray Schafer con la pubblicazione nel 1977 del libro *The Soundscape: Our Sonic Environment and the Tuning of the World*, la cui seconda e ultima edizione è del 1994 (Schafer 1994). In questo tipo di approccio il *soundscape*, che in quanto paesaggio è uno spazio, non è indagato specificamente in termini di cognizione spaziale quanto di cultura sonora, ma include un utile sistema di analisi, come vedremo nel capitolo 5. Un autore che si è dedicato al *soundscape* e allo spazio acustico sia in termini teorici che sul piano artistico è Truax (1996; 1998; 2006; 2012; Truax e Barrett 2011), il quale ha notato che

Film soundtracks are designed to create sonic environments according to a set of conventions understood by the audience (and only loosely related to the real world), but arguably the largest amount of capital for auditory environment design today is invested in the digital game industry where complex and detailed soundscapes are the norm, both for realistic and fantasy worlds (Truax 2008: 106).

Inoltre, come abbiamo accennato nel capitolo 1, la componente spaziale è stata inclusa nella prassi compositiva soprattutto nella musica contemporanea, con una varietà di approcci (cfr. Harley 1994; Barreiro 2010: 290).

Nell'ambito delle scienze applicate, il paradigma storicamente dominante in relazione alla dimensione spaziale del suono è quello della stereofonia, così come fu sperimentata da Blumlein (1933) nel Regno Unito e presso i Bells Laboratories negli Stati Uniti a opera di Fletcher e colleghi (Fletcher 1934a; 1934b), su cui argomenterò nel capitolo 5. Ricerche e sperimentazioni sullo spazio sonoro si riscontrano ovviamente sia tra tecnici/che sia tra artisti/e del suono, al fine di una messa in pratica. I miei autori di riferimento sono Lennox e Myatt, Rumsey, Holman. Anche sulla rivista *Organised Sound* ci sono stati numerosi interventi relativi a questioni spaziali del sonoro e della musica (cfr. Truax 1999; Worrall 1999; Barrett 2002; Westerkamp 2002; Rebelo 2003; Schäfer e Krebbs 2003; Schütze 2003; Bayle 2007; Smalley 2007; Otondo 2008; Baalman 2010; Trębacz 2012; Van Nort 2013; Nelson 2015; Schroeder 2016). La rivista ha inoltre dedicato due numeri al rapporto tra suono e spazio, uno nel 1999 e l'altro nel 2010. Nell'introduzione al numero del 1999, Myatt asserisce che «There has been an extraordinary amount of interest in sound spatialisation in recent years». Tuttavia l'autore manifesta scetticismo verso l'effettivo risultato dei sistemi di suono presumibilmente tridimensionale, il che secondo lui si spiega con due ipotesi: «(i) that our approach

to spatialised sound reproduction is not always appropriate, and (ii) that our understanding of audio spatial perception is limited». Secondo Myatt si dovrebbe dunque usare un nuovo approccio olistico, «perception-oriented», che includa «active hypothesis testing» e che attinga all'ecologia della percezione di Gibson. In effetti «We do not possess the ability to identify distances absolutely from aural information alone; there are no Euclidean units of measurement that can be communicated via our auditory perception mechanisms» (Myatt 1999: 91). Queste tesi sono sviluppate altrove dall'autore, come vedremo. Nell'introduzione al numero del 2010 Harrison e Wilson affermano che nei 12 anni precedenti ci sono state innovazioni tecniche radicali che hanno inciso anche sull'aumento del numero dei canali a disposizione. Secondo gli autori si rischia di assumere erroneamente che il numero dei canali corrisponda al numero degli altoparlanti. Inoltre essi rilevano la problematicità dell'applicazione indiscriminata dell'approccio *assolutista*, che prevede una posizione fissa da parte di chi ascolta e da cui conseguono alcune illusioni acustiche; a questo approccio se ne può sostituire uno *relativista* in cui «the important image for the listener to perceive is not that the sound moves from *here* to *there*, but *that it moves*». Harrison e Wilson notano che i due approcci sembrano convergere (2010: 183).

Confrontarsi con questa letteratura è utile per un approccio interdisciplinare alle questioni spaziali del suono nonché per verificare analogie e differenze teoretiche e pratiche. Si noti per esempio come secondo Normandeau sono due le idee principali che spingono a realizzare musica multicanale: «The first argues that if a speaker is sent a less complex sound, it is able to represent that sound with better accuracy and clarity. [...] The second concept [...] arises out of the ability of human ears (like all mammals' ears) to localise sound in space with great accuracy» (2009: 278). Relativamente a questa seconda idea, Barreiro riscontra due approcci principali nella musica elettroacustica: il primo «concerns the precise position of the sounds at specific locations in the space and the depiction of clearly identifiable trajectories within that space»; il secondo «is based on a more generalised (or diffused) distribution of the sounds, not concerned with their precise localisation», da cui può scaturire la sensazione che «sounds *are* the space» (2010: 291). Macedo propone una tipologia di cinque spazi sonori, di cui il primo è *metaforico* mentre gli altri sono letterali. Lo spazio *acustico* è determinato da riflessione, diffrazione e risonanza del suono; lo spazio come *spazializzazione sonora*, con effetti di distanza, direzione, movimento, avvolgimento e segregazione di flussi; spazio come *referenza*; spazio come *luogo* (2015: 243-247). Chapman distingue invece tre «acoustic environments» che noi fruiamo contemporaneamente durante l'ascolto: quello dell'ambiente registrato, quello “virtuale” creato in post-produzione, quello dello spazio di esibizione (2017: 47).

Jordan dimostra esemplarmente la proficuità dell'incontro tra lo studio dell'audiovisione (ispirato a Chion) e l'ambito musicale, specificamente i *soundscape*. In questo tipo di composizione c'è un'attenzione alla rappresentazione spaziale che può essere applicata al sonoro sincronizzato con immagini, poiché questo è costituito in buona parte da una rappresentazione uditiva dello spazio. Inoltre «The practice of film sound design is at least as important a precursor to soundscape composition as *musique concrète*». Jordan parla pertanto di *reflective audioviewing* – sul modello del *reflective listening* di Katharine Norman – il quale racchiude tutte le quattro accezioni con cui il termine *reflection* è stato usato: (1) la firma spaziale; (2) «the way a recording reflects the sound that was recorded»; (3) «the way that representational recordings reflect within the space of a composition»; (4) il modo in cui chi ascolta riflette sul suono (Jordan 2012: 248-50). Grazie all'analisi di *Elephant* (*Id.*, USA 2003), Jordan conclude che «if soundscape composition uses creative practice to extend acoustic ecology's goal of fostering enhanced environmental awareness, then the cinema is fertile ground for pushing this goal even further into the realm of the audiovisual» (2012: 256).

Gli studi di Lennox e i suoi colleghi sono riconducibili all'ambito dell'audio spaziale. Gli autori hanno prodotto ricerche sia teoretiche sia sperimentali finalizzate alla realizzazione di opere di sonoro spaziale, ma lo spessore speculativo di queste ricerche è tale da meritare attenzione in questo contesto. Lennox, Myatt e Vaughan (1999) è il primo esempio importante. Gli autori ritengono che il passaggio «from surround sound to true 3-D requires an updating of the psychoacoustical theories which underlie current technologies». I sistemi ambisonic permettono di ottenere illusioni tridimensionali ma mantenendo la difficoltà di localizzazione delle immagini sonore. 1. Subentra quindi il problema del realismo, che per loro deve essere inteso anzitutto come credibilità; inoltre l'audio 3D deve avere tra i requisiti minimi «the potential for exploration, either through movement or selective attention» (1999: 2). Lennox, Vaughan e Myatt introducono il concetto di *externalisation*, ossia «what makes the world and its things “out there”»: questa *out-there-ness* è la principale differenza tra campi sonori (*sound fields*) – che ne sono privi – e ambienti sonori (*sound environments*) – che invece la presentano (2001a: 2). Di conseguenza i primi

are simplified, quantifiable representations of the latter, and are necessarily “task-specific”. Thus, examples of perception in sound fields are special cases. It follows that there may be classes of information (or information-yielding properties) available in sound *environments* which may not be incorporated in particular sound *fields*. It further follows that theories of spatial perception cannot adequately be founded on theories of sound *field* perception (2001b: 1).

Gli autori suggeriscono di riconsiderare il concetto di spazio superando la teoria euclidea per affidarsi a Gibson: egli propone «a move away from perception as signal processing towards perception as information processing» (Lennox, Myatt e Vaughan 1999: 3). Essi concepiscono quindi la percezione «as a form of hypothesis-testing, with predicted/actual comparison and concomitant memory storage and retrieval mechanisms»; ne consegue la cruciale distinzione tra *percezione* e *sensazione*: «Perception attempts to “know” the present, whilst sensation occurs in the very recent past». Inoltre occorre notare che «more cognitive processing is devoted to novelty and a correspondingly lower order of resolution suffices for non-novelty» (1999: 3). Insomma la percezione spaziale è intesa come «the cognitive individuation of features of our environment» che sarebbe «a pattern-recognition process using a wide variety of informational and signal qualities» (1999: 5). La selezione delle informazioni è proporzionata all'obiettivo e dipende da un determinato quadro di riferimento, quindi è in accordo con esigenze interne e informazioni esterne (1999: 4). La percezione spaziale uditiva è inoltre considerata «a time-domain process» (1999: 5), poiché con l'ascolto

we have no need, or mechanism, to delineate ‘edges’ in order to detect organisms. We make quite different discernments about objects in the auditory world, whose perceptual edges are temporal rather than spatial. Auditory sensation depends on the existence of energetic processes in a way that vision does not - one may see an environment in which nothing is moving, one would not hear it» (1999: 6).

Gli autori contestano la tesi per cui la percezione spaziale uditiva dipenderebbe dall'ascolto binaurale e dunque da fenomeni inter-aurali. Infatti essi rilevano un problema intrinseco nella stereofonia la quale, fondandosi sul principio dell'immagine fantasma, comporta «the constraint of a single specifically-located perceiver [...] equidistant from the speakers», da cui conseguono altri limiti. Di conseguenza gli autori promuovono un approccio complementare alla «“perceiver-centered” philosophy», ossia una «“object-centered”, “sound environment”-depiction philosophy» ispirata alla teoria delle *affordances* di Gibson: si assume che l'oggetto, l'ambiente e il/la percipiente sono parti integranti, e lo scopo è quello della ricostruzione delle informazioni percettivamente importanti (Lennox, Myatt e Vaughan 1999: 5; Lennox, Vaughan e Myatt 2001b: 2).

Dato che noi ascoltiamo *oggetti* in *luoghi* (Lennox, Myatt e Vaughan 1999: 4), gli autori separano – con consapevole semplificazione (Lennox, Vaughan e Myatt 2001a: 4) – le percezioni

spaziali in “cosa” e in “dove” (Lennox, Myatt e Vaughan 1999: 6). La rappresentazione uditiva del “cosa”, che è in *foreground*, determina *perceptual significance*, incorporata nei/lle percipienti (Lennox, Vaughan e Myatt 2001b: 2); essa non è accurata quanto quella visiva in termini di forma, mentre può esserlo in termini di composizione materiale, struttura, tipo e dimensioni (Lennox, Myatt e Vaughan 1999: 6). Inoltre l’udito può fornire, a prescindere dalle differenze inter-aurali, informazioni sul “dove” – il *background* (Lennox, Vaughan e Myatt 2001a: 2) – non ottenibili dalla vista e che possono invitare all’azione (Lennox, Myatt e Vaughan 1999: 6): esse sono definite *ambience labelling information*, e sono intese come proprietà intrinseche, oggettive e misurabili di un contesto (o una *texture*), poiché dipendono dall’interazione con gli oggetti sonori (Lennox, Vaughan e Myatt 2001a: 2-4). Gli autori soggiungono che il requisito minimo per la credibilità di un audio 3D è proprio «a stable and consistent audible “background” (context), against which sounding objects “make sense”. The most important application for ambience labelling information is to achieve a tangible sense of envelopment», per il quale non basta circondare l’utente con fonti sonore (anzi, ne è sufficiente una sola), bensì occorre «filling in the background/texture, by placing a sound source coherently within a space» (2001a: 9). Inoltre

the essential elements of this context are coherence and consistency, more important even than complexity or accuracy. It is necessary to present a “believable” background to the auditory display that reacts consistently to sounding objects within it in order to allow perceptual significance and hence selective attention to foreground objects, or rather it is the coherence over time of the background material that enables selective inattention that in turn allows objects within the audio space to be perceived as features with relevance and perhaps importance or even urgency (2001b: 4).

La *perceptual significance* e l’*ambient labelling information* sono due aspetti cruciali nel *perceptual space*, distinto dallo spazio fisico astratto (Lennox, Myatt e Vaughan 1999: 8; Lennox, Vaughan e Myatt 2001b: 2) e che può appunto essere considerato «in the light of a “foreground/background axis”, from maximum significance to completely ignored» (2001b: 3). Mentre la seconda è esterocentrica, la prima consiste nella sua elaborazione da parte del/della percipiente (2001b: 2) ossia in «the ways in which cognitive functions select for attention those information-yielding properties which have the potential for facilitating the most useful predictions». Da ciò consegue tra l’altro «che many of an organism’s judgements of distances are made with reference to the time taken to close them» (Lennox, Myatt e Vaughan 1999: 7). L’ipotesi è dunque che noi comprendiamo gli spazi non tanto in base ai postulati euclidei quanto a delle

«units of urgency» – ovvero delle «“relevance hierarchies” that pertain to individual survival» – tra cui segnatamente la *comingness*: è più importante accorgersi che qualcosa sta arrivando piuttosto che capire nello specifico di cosa si tratta o mappare lo spazio (Lennox, Vaughan e Myatt 2001a: 6; 2001b: 2-3). Ovviamente «the most important “what” that *could come* is a predator, hence “*behaviour*” that may convey “*intention*” will require the most rapid processing» (2001a: 6). Un criterio utile a questo proposito è quello dell’attenzione selettiva, che deve essere considerata insieme alla *disattenzione* selettiva, poiché noi tendiamo a trascurare ciò che ci risulta di scarsa rilevanza percettiva: perciò un disturbo inaspettato in un dato contesto di solito cattura l’attenzione (un esempio è il «*change-of-facingness*», che tende a essere valutato meglio della *facingness* in sé) (2001a: 7). Insomma si stabiliscono delle «“relevance hierarchies” that pertain to individual survival», che sono centrali nella *perceptual significance*: «such hierarchies [...] do not merely facilitate a “bottom-up” unidirectional information flow from the periphery to the higher functions, but [...] “top down” and “sideways” information flows are quite as important» (2001b: 3).

Gli autori conferiscono grande importanza ai giudizi di distanza (che nel caso dell’ascolto possono essere persino superiori alla vista allorché riguardino oggetti celati), un fattore ritenuto cruciale ancorché non adeguatamente spiegato dalla teoria inter-aurale (Lennox, Vaughan e Myatt 2001a: 4). Pertanto lo spazio percettivo sonoro viene suddiviso in questi tipi principali:

here and urgent  
 here but not urgent  
 near and coming  
 near and could come  
 near and going  
 adjacent space and coming  
 adjacent space and moving (could come)  
 distant and coming  
 distant and mildly interesting (Lennox, Myatt e Vaughan 1999: 7).

Gli autori propongono di classificare «the spatial characteristics of sounds in terms of the way in which our perceptual systems grade stimuli for action», il che permette di offrire indizi psicoacustici. Questo processo di astrazione, inteso come «philosophical tool», è denominato *cartoonification* (1999: 8): essa consiste non solo nell’eliminazione di informazioni superflue «but also in exaggerating particular information features to facilitate particular intended perceptions, unambiguously. This involves *distortion* of information to engender *accuracy* of perception», il che corrisponde alla nostra esagerazione cognitiva di aspetti rilevanti di quanto percepiamo (Lennox,

Vaughan e Myatt 2001a: 5). Ne consegue che «immersion in a believable soundfield that engenders a sense of “out-thereeness” requires a process of removing perceptually spurious, and exaggerating perceptually relevant ambience labelling information» (2001a: 6).

Gli autori individuano quattro aree nella *cartoonification: wheres, whats, whats into wheres, events* (Lennox, Vaughan e Myatt 2001b: 4-6). Circa la *where cartoonification*, gli autori individuano tre tipi di spazio, in ordine decrescente di importanza percettiva: «“my space”, or what we might term the “immediate zone”, and, to a slightly lesser extent, any adjacent space» (2001a: 6-7); «Events and things in adjacent space are less interesting than those in my space, but nevertheless inherently possess the potential for impinging on my space» (2001b: 4). Infine il «“Distant” space is inherently very non-threatening, generally, and is almost never confused with either of the other two» (2001b: 4).

Relativamente alla *what cartoonification*, gli autori sostengono che ciascuno dei suddetti spazi può essere scomposto in «“things” (that instantiate sound) and “place-features” (that do not “make” sound). The *things* are the perceptual foreground items, to which we pay selective attention, whilst the place-features are *heard*, but not *attended to*» (Lennox, Vaughan e Myatt 2001a: 6-7). Per individuare una “cosa” attraverso l’ascolto occorrono dei parametri di *thingness*, segnatamente quelli di *behaviour* e di *body* (2001a: 8). Le “cose” si possono ulteriormente distinguere tra *oggetti* e *organismi*, i quali si distinguono in base al comportamento: «an “object” is just that which makes a sound [while] an “organism” is that which makes a sound that betrays sentient behaviour, either vocal or locomotory» (2001b: 5).

Circa la *what-into-where cartoonification*, gli autori precisano che «“things” (the object of “what” analysis) are not “point sources” but have very definite shape and size. This has particular implications for our perceptual ability to locate them in any environment». Come accennato la *comingness* è ritenuta la classe di informazione più urgente, ma si discute anche della *movingness* e della *facingness* (2001b: 5-6).

La *cartoonification of events* consiste di eventi statici, ossia di modifiche nell’output sonoro che non prevedono movimento. Gli eventi chiave sono *startiness, ongoingness, changingness, e finishing* (2001b: 6).

In conclusione gli autori asseriscono che produrre musica 3D significa produrre un ambiente di informazioni, comprensibili in termini di “cosa” e “dove”, mentre il surround – essendo debitore dello “stereo” – è un sistema basato sul “cosa” poiché il “dove” è limitato alla separazione angolare con scarsa profondità (Lennox, Myatt e Vaughan 1999: 7). L’ipotesi è che «Full audio bandwidth may be reserved for ‘what’ information, whilst “where” information may be stored and transmitted as control data for utilisation by local signal processors in accordance with local diffusion and

listening conditions» (1999: 9). Questo principio, come abbiamo visto, è proprio dei sistemi di audio *object-based*.

Tutti questi argomenti sono riproposti in maniera approfondita nella tesi di dottorato di Lennox, in cui l'autore chiarisce il suo approccio interdisciplinare – da me fortemente condiviso – mostrando l'interrelazione tra diversi ambiti teorici (v. figura 17).

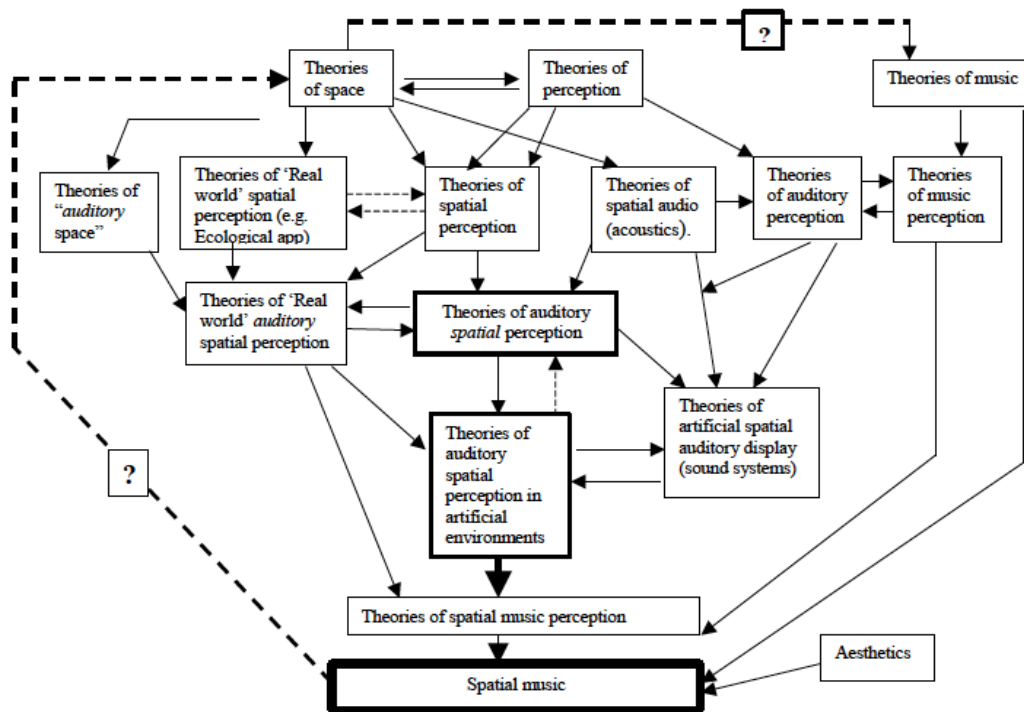


Figura 17: «“Ask theory” flowchart» (Lennox 2004: 14).

Lennox propone anche un utile schema con le componenti dei *causal contexts*, ossia di ciò con cui la percezione si rapporta e che sono inserite le une nelle altre (v. figura 18).

Alcuni anni dopo, Lennox e Myatt ribadiscono la constatazione che l'audio spaziale è concepito nel paradigma della stereofonia: di conseguenza «surround sound should accurately furnish *sensation*, from which *spatial perception* is supposed to naturally follow. Implicitly, this is a *bottom-up, event-driven* model of perception» (2007a: 1), in cui il pubblico è essenzialmente statico e passivo (2007b: 1). Ancora una volta gli autori intendono superare questo paradigma ampliando l'apparato teorico e concentrandosi sugli elementi *top-down* dei processi percettivi, cosicché si rimuove la primazia delle *sensazioni* spaziali e si promuove il concetto di *ipotesi* spaziali:

## **Entities, non-entities, events and relationships,**

### **1. Entities** are either *corporeal* or *ethereal*.

**Ethereal:** *'an idea', 'a sound,' 'an event' 'a collection' or 'association'.*

**Corporeal:** *'things' or 'features'.*

**Things:** *organisms or objects*

**Organisms:** *Prey, predator, ally, competitor or crowd.*

**Objects:** *Tools, weapons, food*

**Features ('entities of potential facility'):**

*Obstacle, trap (self or other), shelter (hide behind/under), way (gap, doorway path, escape, gain access), vantage point (safe, remote viewing)*

### **2. Non-entities**

**Formless Substance** *e.g.: air*

**(Perceptual) Background:** *collection of uninteresting/non-urgent items (e.g.-grains of sand)*

**Place:** *container in which action takes place*

### **3. Events**

**Bounded:** *sequence of changes involving cause, process and effect.*

**Ongoing processes:** *causal sequence without discernable boundary.*

### **4. Relationships:**

**Near / far:** *Within reach/ can reach 'me', or not. Affects 'perceptual significance' of items.*

**Nearer /further than** *in comparison with other salient item(s)*

**Moving:** *Signifies entity, possibly animate.*

**Change of movement** *–acceleration/ direction; good signifier of organism-hood. Predictions require frequent updates.*

**Coming / going:** *Threat or reward, imminent or receding*

**Passing:** *Salient change of significance, from 'coming' to 'going'.*

**Facing-ness:** *Characteristic of entities that facilitates prediction; can be a good signifier of organisms' intentions.*

Figura 18: *i causal context di Lennox (2004: 219).*

These are not limited to the kinds of *ideas* we can cognitively hold, they can be embedded in the *structure* of perceptual mechanisms that support those ideas. [...] If we assume that well-being in the future is a primary aim of perception, we also assume that the future is cognitively represented and that received sensations are rarely totally surprising. They are always embedded in an ongoing causal context with which perceptual systems must be aligned (2007a: 1-2).

Stavolta gli autori si affidano a teorie *modulari* della percezione, le quali suppongono che la percezione spaziale coinvolga sistemi neuronali specializzati. Anche la distinzione summenzionata tra “cosa” e “dove” è pertinente: il fatto che «all spatial environments will support *items* in *locations*» è correlato all’ipotesi che i due componenti, e forse persino diversi aspetti dei medesimi, siano processati da sistemi neuronali distinti (2007a: 2). Circa la fase creativa, gli autori promuovono la modellazione di diversi aspetti del campo sonoro in ordine di complessità: «Modeling the physics of a sound field proximate to an ideal percipient [...]; modeling the physics of a delineated volume of space [...]. [M]odeling of the physics of dynamic objects and their interactions with passive physical bodies. Finally, the modeling of intentionality [...]» (2007a: 3). In conclusione gli autori si pongono il problema del *significato spaziale*, che è lo scopo finale di un audio spaziale oltrepassando le più semplici componenti di direzione e distanza:

The central contention here is that meaning is not simply a feature of the opinions that a percipient might form after an analysis of the physical signals available to sensation. Rather, perception can access meanings directly via dedicated modules that have evolved during the phylogenetic history of a species’ interaction with the causal regularities of the world. The output of these modules form the “bound together” holistic perceptions of place which we are aware of, and the contents of our spatial perception are largely determined by our interactions with these meanings (2007a: 3).

Lennox e Myatt distinguono tra due tipi di «reference frames», uno *egocentrico* («perceiver-centred») e l’altro *eterocentrico* («non perceiver-centred»):

In the former, the position of items is referenced from the perceiver, whereas in the latter, external landmarks can be used as reference, so that an environment can have shape irrespective of the position of the particular perceiver. This second type of frame, that supports perception of an externalized world, is vital for perceiver mobility (2007b: 2).

Gli autori riservano grande attenzione al movimento, che si può categorizzare in base all’urgenza («coming, going, passing, fast, slow, accelerating, changing direction and so on») (2007b: 2). Ancora una volta la causalità è cruciale: considerando la distinzione tra vicino e lontano, gli autori notano che «perceptual and physical “near” may differ. For example, a physically-near item in an adjacent room is perceptually further than items in this room. An item that can move fast is perceptually closer than a slow item. In this instance, we could say that such an item is causally

near». Tornando sulla cartonificazione, gli autori ribadiscono che «The effectiveness of an artificial environment rests on an appropriate simplification of elements, with respect to the task in hand. ... [O]ne can signify a spatial attribute rather than fully specify it», ovvero si dovrebbe favorire «plausibility rather than absolute physical accuracy» (2007b: 3-4; cfr. 2011: 1). Infatti

perception cartoonifies sense-data in terms of what things are and where things are. We use an initial coarse classification of thing (organism or object), place (the shared environment in which perceiver and things perceived have location), feature (that might afford access, climbing, hiding, hefting or obstacle), relationship (nearness, direction, hidden behind, on top of, in front of... a feature), dynamic behaviour (coming, fast, passing, departing, bouncing scraping, accelerating) and finally, intentionality. Cognitive spatial frames of reference are cartoons of spatial layout with respect to potential interaction (2011: 3).

Affinché sia esplorabile, un campo sonoro (*sound field*) – inteso come «a cartoon of the audible part of an environment» (2007b: 4) – dovrebbe avere le seguenti caratteristiche:

- contain perceptually significant sound events
- provide more information to the listener through exploration/movement than that provided to passive/static listeners
- provide audio information relating to the environment that surrounds the listener and the items in the illusion
- provide information about the relationships between sounding objects
- contain information relating to the interaction of objects and environments
- consistently display objects, features and relationships in a stable, non-egocentric reference frame (2007b: 3).

A conferma della loro scarsa fiducia verso il paradigma stereofonico, gli autori osservano che «thing- but not place- character should seem tangible. Unfortunately, phantom images are elusive and lose their tangibility as soon as a perceiver moves within a local sound field, whilst the surround (ambience reproduction) loudspeakers can be too source-like» (2011: 4).

Lennox affronta simili questioni anche in riferimento alla componente emozionale. Egli ripropone un cambio di paradigma da “music-in-a-place” a “music-as-a-place” ispirandosi all’ecologia, asserendo che così si può trattare in modo olistico «the emotional consequences of spatio-musical arrangement» (2009: 2). Una musica così concepita

doesn't just immerse [listeners] in a thick blanket of sound, nor should it rely simply on arranging sources around a listener. This artificial environment should be able to engage the occupants' attention, inattention, anticipation and feelings. We must accept that the feelings we have about places are not merely due to subjective responses brought about by operant conditioning; places do have intrinsic character and emotional engagement is the proper response to place (2009: 18).

Ho dedicato ampio spazio alla teoria di questi autori perché mi ha fornito molti stimoli e concetti utili per la ricerca che esporrò nel capitolo 5, ancorché sia limitata ai processi *top-down*. In quanto segue attingerò alle scienze cognitive per una indispensabile comprensione dei processi *bottom-up*, da integrare in una teoria sulla cognizione dello spazio di opere sonore e audiovisive.

## 5 – Aspetti tecnici e cognitivi dello spazio sonoro filmico: teoria e analisi

### 5.1 – Suono e spazio

Spesso si è asserito che la dimensione peculiare del suono fosse quella temporale, e che nel caso dell'immagine si trattasse di quella spaziale (cfr. la “litania audiovisiva” in Sterne 2003: 30-31). Come spiega Casati «Un oggetto visto è visto innanzitutto occupare una porzione dello spazio» (1991: 15); inoltre «l'immagine è un oggetto tale che per sua natura, oltre che ad essere visto come oggetto materiale, presenta alla visione un oggetto materiale – ovvero l'oggetto che noi vediamo tramite essa, e di cui essa è l'immagine» (1991: 37). Il suono è stato spesso considerato una qualità soggettiva o secondaria degli oggetti (cfr. Metz 1985), ma è più appropriato asserire che il suono è un *evento*, dato che esso è definito come la sensazione percepita dall'udito dovuta alla propagazione di onde meccaniche. Pur asserendo, come fa Yost, che «sound, itself, has no spatial properties» (2001: 440), e che la corteccia uditiva sembra essere priva di una rappresentazione spaziale topografica diversamente da quella visiva (Kong et al. 2012), lo spazio è ovviamente coinvolto in qualunque discorso sul suono, innanzitutto a livello fisico-acustico trattandosi appunto di un fenomeno consistente nelle vibrazioni di un corpo elastico trasmesse nell'ambiente (cfr. Pierce 1988: 35-47). Casati e Dokic sostengono infatti che i suoni sono «*événements dans l'objet résonnant*», e che interessano la *regione spaziale* occupata da esso (1994: 44). Come abbiamo notato, l'attenzione alla componente spaziale dei suoni emerge in occidente già nel contesto musicale precedente all'invenzione della registrazione sonora, fin dal XVI secolo. Con l'incremento dello studio scientifico dell'acustica, sul piano fisico e psicologico, è possibile una comprensione più rigorosa dei fenomeni spaziali del suono dal vivo e, più avanti, di quello registrato.

Il fenomeno culturale della “litania audiovisiva”, e specificamente il paradigma “immagine : spazio = suono : tempo”, ha con ogni probabilità una base biologica, ammettendo che «In the spatial domain, vision generally captures audition, whereas in the temporal domain audition tends to capture vision» (Noesselt et al. 2008). Infatti sia nella realtà (cfr. Shams, Kamitani e Shimojo 2001; Kitagawa e Ichihara 2002; Yost 2001; Middlebrooks 2015) sia nel cinema la vista è il senso privilegiato per la cognizione spaziale – detta altrimenti, l'immagine è il veicolo principale di informazioni spaziali<sup>43</sup>. Occorre inoltre constatare che il predominio dell'immagine concerne anche

<sup>43</sup> Ci sono alcune differenze nella percezione spaziale, visiva e sonora, in base al sesso (cfr. p.es. Neuhoff, Planisek e Seifritz 2009), che meritano di essere considerate in future ricerche.

la realizzazione del film, e non solo la sua fruizione. Nella prassi cinematografica la norma è quella di produrre i suoni in riferimento a delle immagini, piuttosto che viceversa. Mentre la riproduzione comporta spesso operazioni quali lo storyboard e il *production design* (che non è semplicemente scenografia), lo stesso non vale così sovente per il *sound design*. La presa diretta di solito privilegia il dialogo, mentre il resto delle operazioni di *messa in banda*<sup>44</sup> – che include la manipolazione dei dati spaziali – è concentrato nella post-produzione. Ciò non stupisce sia per la consuetudine, sia per le diverse caratteristiche dell'immagine e del suono, sia per la neurofisiologia umana.

## 5.2 – Cognizione incarnata e simulazione incarnata

Negli anni '90 del Novecento inizia una svolta negli studi sul cinema grazie tra l'altro a un'inedita attenzione al corpo inteso non solo nel senso politico-ideologico ma anche come mezzo di percezione e cognizione di stimoli sensoriali. Grazie alla proliferazione delle scienze cognitive si è potuto integrare il cognitivismo classico, tutto centrato sulla mente, con un interesse verso i processi corporei e neurali. Da questo contesto teorico e sperimentale nasce il concetto di cognizione incarnata (*embodied cognition*), con cui ci si riferisce proprio al fatto che i processi cognitivi, ovvero di elaborazione di informazioni, non avvengono solo nella mente o nel cervello ma anche nel resto del corpo.

Notoriamente il sonoro è stato trascurato negli studi sul cinema, fino a iniziare a ricevere la giusta attenzione negli anni '80. Ciò può essere dipeso da due delle fallacie descritte da Altman (1980; 1992) ossia quella storica e soprattutto quella ontologica. Si tratta dell'idea che il cinema è un mezzo essenzialmente visivo, poiché inizialmente era solo visivo (fallacia storica) e poiché in ogni caso l'immagine senza suono è comunque cinema mentre il suono senza immagine non lo è (fallacia ontologica). Un altro motivo tuttavia potrebbe consistere nel fatto che la cultura occidentale dà più importanza alla visione piuttosto che all'ascolto (cfr. Sterne 2003), perciò nel caso di un prodotto audiovisivo c'è la tendenza a pensare che si tratti di immagini con aggiunta di suoni piuttosto che di suoni con aggiunta di immagini. D'altronde ciò ha una motivazione basata sulla prassi, come accennato, poiché nel cinema normalmente si lavora ai suoni in dipendenza dalle immagini, piuttosto che viceversa o in maniera costantemente integrata in tutte le fasi produttive. Ma a prescindere da queste considerazioni resta il fatto che per l'essere umano la vista è un senso privilegiato rispetto all'udito in termini di cognizione. Insomma non c'è da stupirsi che gli studi

---

<sup>44</sup> Traduco così il termine *mise-en-bande* proposto utilmente da Altman, Jones e Tatroe (2000).

cognitivisti abbiano dato un peso decisamente maggiore alle immagini, e che essi siano in ritardo rispetto ad altri approcci nella considerazione del sonoro.

Le consuete formulazioni del concetto di cognizione incarnata infatti tendono tuttora a trascurare il sonoro. Questa dissertazione si prefigge anche lo scopo di proporre un'integrazione dello stimolo sonoro nella teoria della cognizione incarnata nonché in quella della simulazione incarnata proposta da Gallese e recentemente sviluppata in maniera estesa da Gallese e Guerra (2015). Ciò è a mio avviso doveroso sia perché l'interazione crossmodale nel cinema è la norma, sia perché lo sviluppo tecnologico denota la volontà di un sempre maggiore coinvolgimento fisico di chi fruisce grazie al sonoro, il che è testimoniato oggi dal ricorso a sistemi sonori immersivi quali, soprattutto, il Dolby Atmos.

Una delle scoperte più importanti per la comprensione dei processi cognitivi è quella dei neuroni specchio, a partire dalla ricerca di Rizzolatti et al. (1992). Gli studi sul ruolo dei neuroni specchio nella cognizione visiva sono proliferati, ma gli/scienziati/e hanno valutato anche il ruolo dell'udito. Kohler et al. (2002) hanno scoperto che nelle scimmie l'ascolto della rottura del guscio di arachide è fondamentale per l'attivazione dei neuroni specchio, e infatti questo suono è una conferma del successo dell'operazione. Keysers et al. descrivono questi neuroni specchio audiovisivi come neuroni che «respond with a similar intensity of discharge whether the action is only heard, only seen or both heard and seen»; ciò suggerisce che «the neurons code the action in an abstract way, which does not depend on the source of information (auditory or visual)». Gli autori giungono alla conclusione teorica che «multimodal integration can be used for the meaningful representation and recognition of ecologically relevant actions» (2003: 635; cfr. anche Gazzola, Aziz-Zadeh e Keysers 2006). Una simile tesi è stimolante per lo studio della fruizione filmica, come mostrano Gallese e Guerra i quali, partendo da un nuovo modello neuro-cognitivo «in cui azione, percezione e cognizione sono strettamente integrate», parlano di «integrazione multimodale» intesa anche in senso audio-motorio: infatti dal momento che «Alcune azioni sono caratteristicamente accompagnate da sonorità», certi suoni «ci fanno comprendere, quasi vedere, il comportamento di qualcuno che non vediamo» (2015: 218, 62). Ricciardi et al. hanno evidenziato la rilevanza dell'ascolto sullo sviluppo e sull'attivazione del sistema specchio: il suono può attivare la rappresentazione mentale di azioni imparate non solo mediante la visione ma anche mediante l'ascolto. Gli autori concludono che il sistema specchio non andrebbe definito *visuomotor* perché non dipende strettamente dalla vista (2009: 9722-9723).

Incrociando gli studi di cinema con le neuroscienze cognitive, Gallese e Guerra (2015) offrono una proposta alquanto complessa che può essere collocata nel novero della Neurofilmologia, poiché essi studiano i processi neuro-cognitivi in atto durante l'esperienza filmica,

considerando pertanto chi fruisce come un sistema cervello-corpo, similmente al modello *viewer-as-organism* di D'Aloia e Eugeni (2014). Per quanto mi riguarda preferisco usare il termine *percipiente* al posto di spettatore/trice, per ridurre l'“oculocentrismo” ed enfatizzare l'integrazione multisensoriale. Gallese e Guerra si limitano a fare qualche accenno all'ascolto; proviamo dunque a integrarlo nella loro teoria. Anzitutto gli autori spiegano che la cognizione motoria dello spazio peri-personale avviene con stimoli visivi, uditivi e tattili; perciò anche un suono può attivare i neuroni che controllano il movimento delle parti corporee vicine alla regione di provenienza del suono stesso (2015: 49-53); tuttavia la spazialità del film è da loro intesa pressoché solo in senso visivo, mentre dovrebbe esserlo in senso audiovisivo – invece nel cinema normale il tatto è coinvolto solo indirettamente attraverso certe vibrazioni acustiche, il che comunque non è un dato trascurabile. Gallese e Guerra sostengono che quanto accade in un film viene dispiegato «in uno spazio bidimensionale che ci appare illusoriamente nelle tre dimensioni e nel quale sentiamo di poterci orientare e addirittura muovere» malgrado la segregazione spaziale rispetto allo schermo (2015: 79). Ciò che sfugge tuttavia è che lo spazio filmico è denotato non solo dalle immagini sullo schermo ma anche dal sonoro, e quindi che la segregazione va intesa in una maniera diversa. Ciò vale sia nel cinema monofonico che in quello stereofonico surround: in quest'ultimo caso la questione è ancor più stringente perché le fonti sonore si estendono letteralmente al di fuori dello schermo e circondano il pubblico, aumentando le possibilità di conferire profondità alle immagini sia da una o più fonti frontali sia letteralmente grazie a fonti posizionate nella lunghezza offrire l'impressione della presenza di entità nello spazio peri-personale. La segregazione spaziale tra il mondo reale in cui siamo siti e il mondo illusorio del film (cfr. 2015: 111) grazie al sonoro è quindi meno estrema anzitutto allorché incrementa l'impressione di realtà grazie al sincronismo ossia all'esperienza crossmodale, che suggerisce un maggiore senso di concretezza spaziale ovvero di profondità – e quindi in un certo senso di tridimensionalità. Inoltre la possibilità di collocare suoni intorno al/alla percipiente riduce l'impressione di segregazione, perché i suoni possono apparire più vicini, *zuhanden* – cioè a portata di mano, siti nello spazio peri-personale (2015: 72): ciò è permesso anche dal fatto che il suono colma la distanza meglio dell'immagine perché viaggia nell'aria e raggiunge le nostre orecchie senza la delimitazione spaziale dell'immagine inquadrata. La questione andrebbe esaminata anche in relazione alle immagini stereoscopiche, tema escluso da questa dissertazione (cfr. Cullen et al. 2012; 2013).

### 5.3 – Integrazione multisensoriale e interazioni crossmodali

Per le scienze cognitive è un fatto acquisito che la percezione è un'esperienza multisensoriale, in cui cioè sono coinvolti i vari sensi (Morein-Zamir, Soto-Faraco e Kingstone 2003): come spiega Chandrasekaran «Combining information from multiple senses creates robust percepts, speeds up responses, enhances learning, and improves detection, discrimination, and recognition» (2017: 25). Ci sono sempre più prove che dimostrano come nella percezione di un evento si può verificare l'influenza di una modalità sensoriale su un'altra (Sanabria et al. 2004: 540): esiste dunque una *integrazione multisensoriale* da cui derivano delle specifiche *interazioni crossmodali* (cfr. p.es. Caclin et al. 2002: 616)<sup>45</sup> che secondo Shimojo e Shams (2001: 508) sono la regola piuttosto che l'eccezione e di cui si è iniziato a studiare anche i correlativi neuronali con la fMRI (Bushara et al. 2003). Esistono inoltre delle corrispondenze crossmodali, alcune delle quali paiono essere universali, ossia indipendenti dalla cultura e dal linguaggio, per esempio quelle tra l'altezza del suono e l'elevazione e tra l'intensità e la dimensione (Parkinson et al. 2012). A questo proposito si consideri che, come rilevano McBeath e Neuhoff (2002), un incremento dinamico dell'intensità può far percepire un suono come gradualmente crescente di altezza, in conformità con quanto accade di solito in natura; perciò l'intensità può condizionare la percezione della frequenza. Spence (2011: 988-989) differenzia le corrispondenze crossmodali in tre tipi (strutturali, statistiche e semantiche), e ne riassume alcune possibili tra ascolto e visione in una tabella (2011: 979) (v. figura 19).

Auditory Dimension	Visual Dimension	Crossmodal Correspondence	High-Pitch/Loud Sound Corresponds to:
Pitch	Elevation	Yes	High elevation
	Brightness	Yes	Brighter stimulus
	Lightness	Yes	Lighter stimulus
	Shape/angularity	Yes	More angular shape
	Size	Yes	Smaller object
	Spatial frequency	Yes	High spatial frequency
	Direction of movement	Yes	Upward movement
	Contrast	No	N/A
	Hue	No	N/A
	Loudness	Brightness	Yes
Lightness		No	N/A

Figura 19: Alcune corrispondenze crossmodali (Spence 2011: 979).

<sup>45</sup> Si ipotizza che ciò avvenga anche in situazioni di sostituzione sensoriale quali l'ecolocazione; cfr. p.es. Buckingham et al. (2015).

Da questo risulta chiaro che alcune qualità di un suono possono suggerire delle caratteristiche spaziali estrinseche che associamo alla percezione visiva. Come suggerisce lo stesso termine italiano, l'altezza – direttamente proporzionale alla frequenza delle vibrazioni, e descrivibile in una scala dal grave all'acuto – evoca in misura direttamente proporzionale una posizione più o meno alta in una dimensione verticale, mentre un suo graduale incremento o decremento evoca un movimento rispettivamente verso l'alto o verso il basso, dunque una salita o una discesa. Inoltre maggiore è l'altezza maggiore è l'impressione di piccole dimensioni (al netto dell'intensità) e di forma angolosa. L'intensità – direttamente proporzionale all'ampiezza delle vibrazioni, e descrivibile in una scala dal debole al forte – evoca invece in misura direttamente proporzionale una maggiore o una minore estensione, mentre un suo graduale incremento o decremento evoca rispettivamente un aumento o una riduzione di estensione – dunque un ingrandimento o un rimpiccolimento – o di brillantezza. Il timbro (dipendente dalla forma dell'onda, ovvero dai suoni armonici che accompagnano un suono fondamentale) è detto anche il “colore” di un suono, ed è usualmente descritto in una scala dallo scuro al chiaro.

Shams, Kamitani e Shimojo (2001: 147) come molti altri prima di loro spiegano che la visione è la modalità dominante tanto da portare spesso a trascurare segnali di altre modalità; ne sono esempi l'effetto ventriloquo e la *visual capture* (cfr. Kitagawa e Ichihara 2002). Il primo è un effetto crossmodale ampiamente trattato dalle scienze cognitive, ed è così chiamato perché è quello che permette l'impressione che il suono prodotto dal corpo dei ventriloqui provenga dal fantoccio che essi animano: si tratta dunque di «The mislocalization of sounds toward the position of co-occurring visual stimuli» (Bonath et al. 2007: 1697), e si manifesta sia online – ossia nell'immediato – sia offline come *after-effect* (cfr. Radeau e Bertelson 1974; Kitagawa e Ichihara 2002; Zwiers, Van Opstal e Paige 2003; Bertelson et al. 2006). Affinché l'effetto ventriloquo si realizzi occorrono due requisiti: 1) la propagazione del suono deve essere sincronica con l'immagine dell'evento che lo provoca; 2) la distanza tra l'immagine della fonte presunta e la concreta fonte del suono deve rientrare in un angolo di 30° (cfr. Jack e Thurlow 1973; Bregman 1990: 307). Per *visual capture* si intende la capacità della vista di dominare su altri stimoli sensoriali nella produzione di una percezione, non solo sull'udito ma anche per esempio sul tatto (cfr. Hay, Pick e Ikeda 1965; Singer e Day 1969; Pavani, Spence e Driver 2000; Ernst e Banks 2002; Holmes, Crozier e Spence 2004) e sulla propriocezione (Pick, Warren e Hay 1969; Pavani, Spence e Driver 2000); l'effetto ventriloquo è dunque spesso inteso come un esempio della cattura visiva dell'udito nella cognizione spaziale.

Tuttavia la teoria della predominanza generale della visione o della cattura sensoriale è stata già da tempo riconsiderata poiché ci si è accorti/e che esse non valgono a prescindere dal tipo di

compito. Sicuramente ciò è consueto nel compito qui trattato, ossia la percezione spaziale: Chandrasekaran asserisce che «vision plays a large role in the perception of space and subjects find it difficult to ignore visual information» (2017: 25). Perrott et al. (1990: 225) hanno ipotizzato che la percezione spaziale uditiva si sia evoluta per guidare l'orientamento dello sguardo così da verificare visivamente l'eventuale presenza di pericolo (cfr. Heffner e Heffner 1992: 228; Hall e Moore 2003: R91; Asutay e Västfjäll 2015a, 2015b). La vista sembra essere necessaria anzitutto per un normale sviluppo della capacità di rappresentazione uditiva spaziale (Gori et al. 2013: 5; Tonelli, Brayda e Gori 2015: 1); inoltre essa è più accurata nella percezione spaziale tanto da dominare sull'udito nella localizzazione e aiutarlo in questo (cfr. Williams 2012: 904) anche qualora lo stimolo visivo sia inutile a un dato compito (Jackson 1953), poiché come spiegano Tonelli, Brayda e Gori può calibrare i singoli processi di localizzazione uditiva anche dopo una breve osservazione dell'ambiente (2015: 1, 7). Gli autori deducono pertanto che esiste una *calibrazione crossmodale* la quale contribuisce a costruire *rappresentazioni mentali sopramodali* dello spazio (Tonelli, Gori e Brayda 2016: 4-5). L'ipotesi di processi sopramodali è condivisa da altri studi: per esempio Green et al. rilevano che esiste un sistema fronto-parietale che controlla sia l'attenzione visuospatiale sia quella audiospatiale, da cui suppongono che esso sia indipendente da una singola modalità sensoriale; inoltre lo spostamento dell'attenzione verso il luogo in cui ci si aspetta un suono «leads to anticipatory pretarget biasing within auditory cortices» (2011: 3563-3564; cfr. anche Frissen et al. 2005; Green, Conder e McDonald 2008; Ricciardi et al. 2009: 9722). In effetti già da una revisione della letteratura neuroscientifica del 2001 Shimojo e Shams avevano concluso che «cortical pathways previously thought to be sensory-specific are modulated by signals from other modalities» (2001: 505).

In altre ricerche si propone un'ipotesi diversa. Battaglia, Jacobs e Aslin (2003: 1391) rilevano l'esistenza di due modelli per spiegare fenomeni di influenza crossmodale: il modello “winner-take-all”, come nel caso della *visual capture*, in cui si assume che una modalità abbia la meglio sull'altra; il modello ottimale di integrazione sensoriale basato sulla *maximum-likelihood estimation*. Gli autori suggeriscono che entrambi i modelli hanno pregi e difetti e che probabilmente un approccio ibrido sarebbe il più adeguato (2003: 1395). Alais e Burr dubitano che il ventriloquismo risulti da una cattura visiva del suono, che accade solo quando la localizzazione visiva è buona: quando lo stimolo visivo è indistinto è il suono a catturare la visione – con un *effetto ventriloquo inverso* – mentre quando sono entrambi poco distinti nessuno dei due pare dominare. Inoltre la localizzazione bimodale è più precisa di quella unimodale, perciò essi deducono che non vi sia tanto una cattura di un senso da parte dell'altro bensì un combinazione ottimale di informazioni visive e uditive (2004b: 260-261; cfr. Alais e Burr 2004a). Inoltre alcuni

studi (Alais, Morrone e Burr 2006; Burr e Alais 2006) suggeriscono che esistono meccanismi attentivi indipendenti per ascolto e visione; l'attenzione potrebbe essere intramodale o sopramodale a seconda di quale soluzione sia più idonea a un dato scopo, in una maniera statisticamente ottimale in cui il peso di un senso dipende dalla sua affidabilità (cfr. Ernst e Banks 2002; Gori et al. 2008). In ogni caso fenomeni che si potrebbero definire di cattura o di ricalibrazione sensoriale sono stati osservati anche da parte dall'olfatto sulla visione (Kuang e Zhang 2014) e da parte del tatto sulla visione (Gori et al. 2012) e sull'ascolto (Caclin et al. 2002; Tajadura-Jiménez et al. 2010a; Tonelli, Gori e Brayda 2016). In particolare Bruns, Spence e Röder rilevano una ricalibrazione tattile di rappresentazioni spaziali uditive (parlando di «audio-tactile ventriloquism aftereffect»), concludendo che «the brain uses not only visual but also non-visual information as a reference for constant recalibration of auditory spatial representations» (2011: 343; cfr. anche Gondan et al. 2011).

L'effetto ventriloquo ha ricevuto molta attenzione da Bertelson e Radeau, che lo hanno studiato sia come effetto immediato sia come *after-effect* (Radeau e Bertelson 1974), osservando che «adaptation to auditory-visual discordance is little affected by cognitive variables» (Radeau e Bertelson 1978: 342): il bias crossmodale avviene anzitutto a livello preconsco – quindi in modo automatico e involontario, ossia pre-attentivo – da cui scaturisce una fusione percettiva in cui possono eventualmente subentrare processi attentivi esogeni o endogeni (cfr. anche Radeau e Bertelson 1977; Bertelson e Radeau 1981; Bertelson e Aschersleben 1998; Bertelson et al. 2000a, 2000b; Vroomen, Bertelson e de Gelder 2001a, 2001b; sull'attenzione visiva esogena ed endogena cfr. Theeuwes 1991). In un importante studio, Driver (1996) analizza la complessità del rapporto tra stimolo visivo e stimolo sonoro verbale in relazione all'effetto ventriloquo. La parola è un evento percepito in maniera multimodale, cui contribuisce il movimento labiale, generando un'integrazione audiovisiva (cfr. anche Summerfield 1992) che talora dà luogo a errori come nel caso dell'effetto McGurk (McGurk e MacDonald 1976; Shimojo e Shams 2001: 506; Neuhoff 2011: 71; Williams 2012: 909-910). Ciò è un esempio di interazioni crossmodali automatiche e involontarie, che quindi precedono l'elaborazione cognitiva e l'attenzione (cfr. Spence e Driver 2000: 2061; Kobayashi et al. 2012) ma che sono utili nel mondo reale perché la precisione della localizzazione visiva compensa la limitatezza di quella uditiva. Da parte loro Kafaligonul e Oluk (2015) spiegano che alcune associazioni audiovisive tra immagini in movimento e suoni statici influiscono sulla percezione visiva del movimento a basso livello, quindi prima dell'attivazione dell'attenzione motoria di alto livello. In seguito intervengono comunque anche meccanismi cognitivi, in base all'attenzione selettiva. Insomma in simili processi l'influenza dell'attenzione sembrerebbe essere nulla o comunque marginale (cfr. Bonath et al 2007; Grassi e Casco 2009). L'effetto ventriloquo è stato

studiato anche con tecniche neuroscientifiche: Bonath et al. (2007) ricorrono all'elettroencefalografia e alla fMRI per studiarne le basi neuronali (su cui si era espressa anche Radeau 1998), concludendo che esso «provides a classic example of how sensory inputs from the different modalities are integrated in the brain to achieve unified and coherent internal representations of environmental events» (Bonath et al. 2007: 1700).

Pur ammettendo che la visione sia il mezzo privilegiato per la cognizione spaziale ci sono fenomeni di calibrazione crossmodale in cui la percezione visiva è condizionata dal suono, anche solo in termini di incremento di sensibilità (cfr. Pérez-Bellido, Soto-Faraco e López-Moliner 2013: 1074). Radeau e Bertelson (1976, 1987) sostengono che esiste una cattura uditiva (*auditory capture*), ovvero un bias uditivo della localizzazione visiva, ancorché minore rispetto a quello contrario. Alcune possibili conferme di questo vengono per esempio da Stein et al. (1996), i quali rilevano che un suono può far percepire un LED come più intenso, e più in generale che il suono influenza la risposta dell'attività neuronale visiva, mentre in un importante studio Sekuler, Sekuler e Lau (1997) dimostrano come il suono può alterare la percezione visiva del movimento in un *bounce-inducing effect* (cfr. Watanabe e Shimojo 2001; Sanabria et al. 2004; Grassi e Casco 2009). Soto-Faraco et al. osservano che anche la percezione della direzione di un movimento sonoro apparente è plasmata dal movimento visivo apparente, concludendo che nell'interazione crossmodale sono significativi fattori sia spaziali sia temporali (2002: 139; cfr. anche Teramoto et al. 2012). Hidaka et al. si dedicano specificamente al movimento visivo indotto dal suono spiegando che «the alternation of sound location induces illusory visual motion when vision cannot provide accurate spatial information», da cui si può dedurre che «the neural representations of auditory and visual motion processing can bias each other, which yields the best estimates of external events in a complementary manner» (2009: 1; cfr. anche Teramoto, Hidaka e Sugita 2010). Kobayashi et al. (2012) rilevano che il suono può creare l'illusione di un movimento visivo non solo per processi *top-down* ma anche quando la differenza tra i suoni è indistinguibile. Inoltre sembra che alcune caratteristiche musicali (tempo, modo) possano influenzare non solo l'eccitazione e l'umore di chi ascolta ma anche le sue abilità spaziali (cfr. Rauscher, Shaw e Ky 1993, 1995; Husain, Thompson e Schellenberg 2002; Perham e Withey 2012) o di altro tipo (p.es. la guida di un veicolo, reale [Brodsky 2002] o virtuale [Cassidy e MacDonald 2010]). Il suono ha notoriamente una funzione di induzione sensorimotoria determinata soprattutto dalla sua componente ritmica, che pertiene alla dimensione temporale (cfr. Cano e Cremonini 1995: 87-99; Cano 2002: 148-156; Aschersleben e Bertelson 2003).

Alcuni esperimenti confermano simili interazioni crossmodali anche relativamente alla fruizione di tecnologie audiovisive (cfr. Perrott et al. 1991; Pierno et al. 2005). Perrott et al. (1996) e

Nelson et al. (1998) dimostrano l'efficacia di informazioni spaziali uditive virtuali per il rilevamento di oggetti visivi; questo miglioramento della resa è confermato da Flanagan et al. (1998: 466) e da Bolia, D'Angelo e McKinley (1999: 668) in termini di riduzione del tempo di ricerca visiva in un ambiente virtuale, e da Larsson, Västfjäll e Kleiner (2001: 4) in riferimento alla performance e alla memoria. Gröhn, Lokki e Takala dimostrano che la navigazione di uno spazio virtuale migliora quando gli stimoli sono audiovisivi, benché la navigazione esclusivamente visiva sia migliore di quella esclusivamente sonora. In particolare si nota che «rough location was based on auditory perception while the final approach was dominated by vision» (2003: 3-4). Ciò confermerebbe che la visione è più precisa nella percezione spaziale ma che l'ascolto dà un utile contributo allorché palesa la collocazione di una fonte esclusa dal campo visivo. Zhou et al. (2007) sostengono che l'audio 3D nella realtà aumentata aiuta la percezione della profondità, l'esecuzione di compiti e la collaborazione tra utenti, oltre a incrementare il realismo (cfr. Moeck et al. 2007) e il senso di immersione. Turner, Berry e Holliman (2011) sperimentano l'associazione a immagini stereoscopiche di “audio 3D”, ovvero di più altoparlanti allineati nella lunghezza tra chi ascolta e lo schermo: essi dimostrano che così facendo si percepisce un oggetto come più vicino qualora il suono sia più vicino, malgrado la profondità visiva resti immutata. Le influenze crossmodali tra audio e video possono concernere anche altri aspetti: per esempio la percezione generale della qualità del rendering di immagini e suoni dipende non solo da valori assoluti ma anche dall'influenza reciproca dei due stimoli (Begault 1999: 7; Hulusic, Aranha e Chalmers 2008; Rojas et al. 2012; cfr. Zielinski, Rumsey e Bech 2008); inoltre il *beat rate* di uno stimolo sonoro influenza la percezione del *frame rate* (Hulusic et al. 2009); infine il suono può migliorare la percezione della fluidità di un movimento visivo (Mastoropoulou et al. 2005; Hulusic et al. 2010).

Molta della letteratura sulla percezione crossmodale tra visione e ascolto si è incentrata su interazioni spaziali, come appunto l'effetto ventriloquo (cfr. Bertelson et al. 2006: 428), poiché come spiegano Morein-Zamir, Soto-Faraco e Kingstone «In the spatial domain, vision can bias the perceived location of sounds whereas sounds rarely influence visual localization». Tuttavia «Outside the spatial domain [...] audition can bias vision»: gli autori si dedicano alla dimensione temporale, avallando studi precedenti secondo cui l'ascolto cattura la percezione temporale visiva, fenomeno chiamato *auditory driving*. Ne consegue che il ventriloquismo *tout court* è più precisamente un ventriloquismo spaziale, poiché esiste anche un *ventriloquismo temporale* (*temporal ventriloquism*) (2003: 154-155; cfr. Aschersleben e Bertelson 2003; Bertelson e Aschersleben 2003). Uno degli studi in questione è quello già citato di Shams, Kamitani e Shimojo (2001), che dimostra come un singolo flash accompagnato da beep multipli è percepito come flash multipli. In effetti, sebbene la questione del sincronismo pertenga alla dimensione temporale,

essendo cruciale per l'effetto ventriloquo diventa indirettamente anche un problema spaziale (cfr. Radeau e Bertelson 1987). Anderson (1996: 81-84) pone la sincronia come il fattore decisivo nell'illusione audiovisiva: se non c'è sonoro, le immagini sembrano lacunose e dunque “meno realistiche”; se il sonoro non è perfettamente sincronizzato, si palesa l'artificialità e subentra la distrazione o si riduce l'interesse. Anderson tuttavia omette l'effetto ventriloquo, per il quale è indispensabile anche una prossimità spaziale. Infatti se il sonoro sincronico viene da un punto che dista dalla fonte presunta più di 30° di angolazione, l'impressione di unità si perde (cfr. Jack e Thurlow 1973; Bregman 1990: 307). Radeau e Bertelson (1977) dimostrano che il sincronismo conta più del realismo per l'adattamento alla discordanza crossmodale; inoltre può verificarsi un fenomeno di salienza in cui un segnale con una particolare configurazione temporale (p.es. l'interruzione di un pattern) attrae su di sé la localizzazione (Radeau e Bertelson 1987).

Il primo studioso di mia conoscenza ad aver parlato di ventriloquismo a proposito del cinema è Rick Altman (1980), pur senza ricorrere alle scienze cognitive. Quello che Chion chiama *calamitazione spaziale* è proprio l'effetto ventriloquo, ma egli omette di precisare che un simile risultato non si ottiene qualora vi sia una distanza tra fonte visiva e fonte sonora superiore di 30° (cfr. Jack e Thurlow 1973). Per giunta anche il concetto di *sincretismo* proposto dall'autore coinvolge implicitamente lo spazio: con questo neologismo Chion fonde “sincronismo” e “sintesi”, indicando «A spontaneous and reflexive psychophysiological phenomenon that is universal and [...] not from cultural conditioning» che consiste «in perceiving the concomitance of a discrete sound event and a discrete visual event as a single phenomenon». Si crea involontariamente l'impressione che vi sia «a common cause to sound and image, even if their nature and source are completely different and even if they have little or no relation to each other in reality» (2009: 492). Il film sonoro fa dell'effetto ventriloquo un principio cardine, dato che il suono proviene da un altoparlante e non dalla fonte denotata né dall'area dello schermo in cui essa è eventualmente visibile, senza che l'artificio sia palese o distraente. Le stesse scienze cognitive usano gli audiovisivi come esempio consueto di circostanza in cui si manifesta l'effetto ventriloquo (cfr. p.es. Warren, Welch e McCarthy 1981: 557; Spence e Driver 2000: 2057; Shimojo e Shams 2001: 506; Battaglia, Jacobs e Aslin 2003: 1391; Alais e Burr 2004b: 257; Burr e Alais 2006: 243; Hulusic et al. 2012: 119).

## 5.4 – La firma spaziale

Spazializzare in senso lato significa conferire un carattere spaziale, rappresentare qualcosa come collocato nello spazio. Pertanto si può spazializzare un suono anche in un singolo canale, dal momento che la collocazione nello spazio non è suscitata solamente dalla distribuzione di più fonti

sonore nello spazio stesso. Tuttavia la consuetudine è quella di intendere la spazializzazione proprio in questo senso ristretto, da cui l'idea che un singolo canale sia puntiforme, "adimensionale". Perciò è utile descrivere le caratteristiche spaziali del sonoro mono e in generale di un suono unidirezionale con l'espressione *firma spaziale*, la quale si riferisce ai caratteri spaziali di un suono, mediato dalla tecnologia o meno. Essa è determinata dal rapporto tra suono diretto e suono riflesso e dalle caratteristiche dello spettro in termini di frequenza e intensità (Altman 1992d: 252; Maasø 2008: 37). Per semplicità mi sembra quindi proficuo considerare la firma spaziale tralasciando l'ascolto binaurale: le caratteristiche acustiche poc'anzi citate non dipendono dalle differenze interaurali, e sono quindi rilevabili anche in un ascolto monoaurale o da un singolo altoparlante. Normalmente le persone odono allo stesso modo da entrambe le orecchie, quindi in modo binaurale, ma il termine "monoaurale" si usa anche per indicare fenomeni che dipendono da come ciascuna delle due orecchie percepisce il segnale indipendentemente dall'altra (cfr. Middlebrooks 2015: 103). Inoltre l'analisi della percezione del sonoro propagato da un solo altoparlante è di rilevanza nel nostro contesto poiché è proprio di qualsiasi sistema, non solo di quelli mono, perché concerne anzitutto la propagazione sonora da un singolo altoparlante a prescindere da eventuali altri. La componente direzionale è intrinseca nei suoni dal vivo, mentre nel caso della registrazione sonora essa dipende dall'uso di almeno due diffusori. Infatti il problema della direzione è subentrato in un secondo momento, e ha comportato lo sviluppo della stereofonia. Perciò, relativamente al sonoro filmico, è utile considerare la firma spaziale a prescindere dalla direzionalità la quale è pertinente alla spazializzazione acustica essendo assente da una fonte puntiforme. In questo senso la firma spaziale è il primo fattore – o meglio, insieme di fattori – che contribuisce alla cognizione della prospettiva sonora. I fattori in questione sono la *distanza*, la *frontalità* e la *profondità*.

Gli spazi coinvolti nel sonoro filmico, e in generale nel suono registrato, sono tre. Il primo è quello, reale, in cui i suoni si propagano al momento dell'incisione (set, location, studio) a partire da una fonte reale, il quale può non esistere qualora i suoni siano generati sinteticamente ovvero in assenza di una fonte reale. Il secondo è quello, immaginario, che risulta da operazioni di ripresa, missaggio e montaggio sonoro, da cui risulta un'ambiente sonoro (cfr. Lennox, Vaughan e Myatt 2001a, b) costituito da "immagini sonore" (*sound* o *auditory image*; cfr. p.es. Begault 2000: 70; Yost 2001; Nelson, Donovan e Takahashi 2015) ovvero "oggetti sonori" (*sound object*; cfr. p.es. Kallinen e Ravaja 2007) o "oggetti uditivi" (*auditory object*; cfr. p.es. Peterson 2001: 195; Breebaart e Faller 2007: 33)<sup>46</sup>, insomma di fonti denotate (quelle cui di solito mi riferisco,

<sup>46</sup> La locuzione *oggetto sonoro* (*sound object*) in questa accezione può essere confusa con il concetto tecnico che denota un file audio trattato come un oggetto piuttosto che inserito in un canale, proprio dei sistemi audio *object-based* (cfr. cap. 1). A scanso di equivoci sarebbe opportuno indicare la questione percettiva con l'espressione *oggetto uditivo* (*auditory object*) e riservare *oggetto sonoro* a quella tecnica; meglio ancora si potrebbe indicare quest'ultima con la locuzione *oggetto audio* (*audio object*), come nei *white paper* Dolby, poiché "audio" fa

indicandole ellitticamente come “fonti”, distinte dalle fonti materiali ovvero gli altoparlanti): è qui che nasce lo *spazio sonoro filmico*. Il terzo è quello, reale, in cui il sonoro si propaga per giungere alle orecchie del pubblico a partire da fonti concrete ovvero dagli altoparlanti, e che nel caso di ascolto in cuffia è ridottissimo (cfr. Théberge, Devine e Everett 2015: 11; Zacharov e Koivuniemi 2001: 1). Perciò il sonoro registrato aggiunge altri fattori alla firma spaziale reale. Prendendo in esame la voce, Maasø propone come abbiamo visto tre livelli di analisi: la distanza vocale, propria della voce di per sé, come accade nella comunicazione interpersonale; la distanza uditiva voluta («intended earshot»), che scaturisce da operazioni tecnologiche sull’intensità del segnale; la prospettiva del microfono, a seconda del tipo del medesimo e della sua posizione. È dalla *media schizofonica* tra questi tre livelli che scaturisce l’effetto finale, ossia la firma spaziale e quindi il senso di una distanza e di una profondità. In verità la firma spaziale dipende anzitutto dalle caratteristiche dell’ambiente in cui il suono si propaga, e poi dal tipo e dalla posizione del microfono che lo registra.

La percezione della distanza dipende dall’intensità (per lo più in assoluto, e solo in parte in termini differenziali tra le due orecchie) e dal rapporto tra suono diretto e suono riflesso, infatti all’aumentare della distanza si riduce l’intensità in particolare del suono diretto. La percezione della distanza è meno accurata rispetto alla localizzazione orizzontale o verticale, ed è molto influenzata dalla conoscenza pregressa del suono in questione (cfr. Coleman 1963; Begault 2000: 2; Yost 2001: 444-445; Hall e Moore 2003: R91; Pulkki, Lokki e Rocchesso 2011: 153; Middlebrooks 2015: 104-105). Oltre al suono *diretto*, che raggiunge direttamente il/la percipiente, vi sono le *riflessioni o diffrazioni di primo ordine*, onde che arrivano entro un tempo tale che i singoli percorsi restino distinguibili, e infine le *riflessioni o diffrazioni tarde*, da cui scaturisce la *riverberazione* (cfr. Begault 2000: 82-83; Pulkki, Lokki e Rocchesso 2011: 164; Hulusic et al. 2012: 105). Collins spiega che la posizione del microfono a seconda della distanza plasma non solo l’intensità e ma anche il timbro del suono, a causa della relazione tra suono diretto e suono riflesso. Secondo l’autrice entro i 30cm si parla di *close miking*; poco oltre si ha *distant miking*, in cui le riflessioni si equivalgono al suono diretto; oltre si parla di *ambient miking*, poiché le riflessioni dominano sul suono diretto (2013: 288). Tuttavia non va considerata solo la distanza del microfono ma anche la sua figura polare e il suo orientamento rispetto alla fonte. Un microfono può infatti avere una diversa sensibilità direzionale, registrando “a fuoco” (Corelli, Felici e Martinelli 2006: 136-137) – ossia con una prevalenza del suono diretto rispetto a quello riflesso – una porzione di spazio dall’angolo di incidenza più o meno ampio, come mostra la figura 20.

---

riferimento a un fatto tecnologico, salvo se usato come prefisso per termini che si riferiscono all’udito.

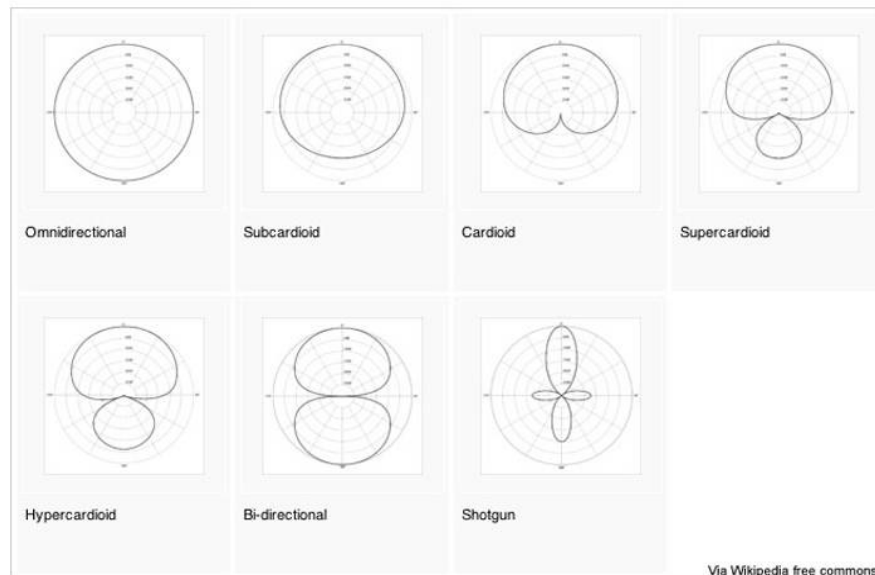


Figura 20: figure polari dei microfoni.

Basandomi sul concetto di *facingness* proposto da Lennox, Vaughan e Myatt (2001a: 7; 2001b: 5-6) definisco la *frontalità* come il grado di impressione di trovarci di fronte alla fonte, da cui scaturisce la sensazione della nostra posizione relativa alla fonte sonora in termini di angolazione. Infatti orientando le orecchie altrove – o riducendo il fuoco del suono, col posizionamento microfónico o nel missaggio – si riduce la quantità di suono diretto e si altera la percezione dello spettro. Un chiaro esempio si ha dalla vita reale: se una persona ci parla di spalle lo capiamo anche a occhi chiusi, poiché il suo stesso corpo pone un ostacolo alla propagazione del suono (cfr. Chion 2001: 93). Collins rileva tre tipi di circostanze in cui qualche ostacolo influenza la percezione uditiva. Il suono è *occluso* quando è ostacolato integralmente, il che vale più le frequenze sono alte: ne deriva un suono di intensità ridotta e ovattato, che si può ottenere nel missaggio con un filtro passa basso il quale lascia appunto passare le frequenze basse eliminando gradualmente quelle alte. Il suono è *ostruito* quando è ostacolato il suono diretto ma non quello riflesso; la riduzione d'intensità è minore. Il suono è infine *escluso* quando solo le riflessioni sono ostacolate, dunque il segnale diretto è pulito (2013: 292-293).

Le prime due zone prossemiche del modello di Maasø sono quella intima e quella personale: la prima corrisponde al primissimo piano sonoro (PPP; *extreme close up*, ECU), ossia a come udiremmo se la fonte fosse a una distanza tra i 5 e i 45 cm da noi: si ascolta solo il suono diretto, asciutto, con bassi corposi, e si percepiscono decisamente i rumori corporali (soprattutto quelli prodotti dalla bocca). La seconda corrisponde al primo piano sonoro (PP; *close up*, CU), ossia a come udiremmo se la fonte fosse a una distanza tra i 45 e i 120 cm da noi: le riflessioni sono appena

percepibili e il suono è meno asciutto, si ha quindi un minimo senso dell'ambiente. In questi casi dunque c'è l'impressione della presenza della fonte nel nostro spazio *peri-personale* (a portata di mano, *zuhanden*), mentre dalla zona successiva in poi le fonti sono nello spazio *extra-personale* (fuori dalla portata della mano, *vorhanden*) (su queste definizioni cfr. Gallese 2007; Gallese e Guerra 2015: 72; Vignemont e Iannetti 2015). La zona sociale vicina corrisponde al mezzo primo piano sonoro (MPP; *medium close up*, MCU), ovvero a una distanza tra 1,2 e 2,2 m: il suono diretto continua a prevalere ma con una evidente presenza di riflessioni e una minore presenza di frequenze basse (e alte: cfr. Rumsey 2001: par. 2.2), e si può ancora percepire rumori corporali. La zona sociale lontana corrisponde al campo medio sonoro (CM; *medium shot*, MS), ossia a una distanza dai 2,2 ai 3,7 m: suono diretto e riflesso si equivalgono approssimativamente. La zona pubblica vicina corrisponde al campo lungo sonoro (CL; *long shot*, LS), ovvero a una distanza tra i 3,7 e i 7,6 m: qui prevale il suono riflesso pur senza perdere l'intelligibilità, e il suono si affina. La zona più distante, quella pubblica lontana, corrisponde al campo lunghissimo sonoro (CLL; *extreme long shot*, ELS), ossia a una distanza che supera i 7,6 m; oltre i 12-13 m si può avvertire l'eco (Maasø 2008: 43). Considerando l'attivazione dei neuroni specchio attraverso l'ascolto è lecito supporre che le osservazioni di Gallese e Guerra sulla simulazione incarnata attivata dal PP visivo valgano anche per il PP sonoro. Se lo stimolo è multimodale il PP sonoro può rinforzare l'impatto del PP visivo, così come il rumore della rottura del guscio d'arachide è per la scimmia una conferma crossmodale del compimento dell'azione di rompere il guscio. Se il suono sembra provenire da un punto che oltrepassa lo spazio peri-personale probabilmente l'effetto è affine a quello di un'immagine alla medesima distanza: i meccanismi di simulazione incarnata hanno luogo ma con un'intensità minore dato che l'entità in questione non è *zuhanden*.

La resa della distanza, come accennato, dipende anzitutto dalla distanza del microfono dalla fonte, ma può essere mutata manipolando l'intensità ovvero aumentando o riducendo il volume del segnale; tuttavia così facendo non si otterrebbe necessariamente lo stesso suono che si udirebbe a una minore o maggiore distanza, perché non cambierebbero le altre qualità acustiche (Hall e Moore 2003: R91). Si potrebbe avere insomma l'impressione non tanto di una diversa firma spaziale bensì della modifica del volume di una incisione e quindi solo in parte di una modifica della distanza, il che è in ogni caso diverso dalla naturale riduzione di intensità di un evento sonoro. L'effetto di riflessione può comunque essere aggiunto nel messaggio ricorrendo in maniera più o meno ingente ai riverberi, che sono congegni i quali applicano una riverberazione artificiale. Essi si dividono in due categorie: quelli *a simulazione* trattano il segnale per simulare una riverberazione reale o ipotetica; quelli *a convoluzione* applicano al segnale il campionamento della riverberazione di ambienti reali, per ottenere un suono che somigli il più possibile a quello che si avrebbe

nell'ambiente in questione (cfr. Corelli, Felici e Martinelli 2006: 308-310; Rose 2015: 466-469). Un suono può essere statico o dinamico, quindi la sua firma spaziale può gradualmente cambiare dando l'impressione di avvicinamento o allontanamento, come vedremo. È interessante notare anche che l'acustica dell'ambiente e quindi la firma spaziale possono condizionare la risposta emotiva all'ascolto di certi suoni, come dimostrano studi classificabili nell'ambito dell'*emoacustica* (*emoacoustics*) (cfr. Asutay et al. 2012). Per esempio suoni propagati in stanze ampie – perciò con maggiori riflessioni – sono percepiti come più sgradevoli e minacciosi rispetto al caso di stanze piccole, sia in termini soggettivi sia a livello fisiologico (Tajadura-Jiménez et al. 2010c; Hagman 2010: 41). Inoltre un suono può essere percepito come più intenso se presentato in un contesto emotivamente negativo (Asutay e Västfjäll 2012; cfr. anche Asutay e Västfjäll 2015b).

A prescindere dalla scala dei piani è doveroso stabilire se un suono (così come un'immagine) è *in primo piano* oppure *sullo sfondo*, ovvero in termini gestaltici se è una *figura* o se è lo *sfondo* (ciò è di rilevanza anche per l'analisi musicale; cfr. Bent e Drabkin 1990: 47; Lennox, Myatt e Vaughan 1999; Lennox, Vaughan e Myatt 2001a, b; cfr. Bregman 1990: 18-21). Questa distinzione è fondamentale per ragionare sui processi cognitivi, come dimostra l'effetto *cocktail party* (su cui torneremo): la nostra attenzione è catturata da o concentrata su un certo suono, la figura, rispetto al quale gli altri suoni sono uno sfondo, cioè vengono mascherati (cfr. Hulusic et al. 2012: 116). Il sonoro puntiforme permette di percepire anche la *larghezza apparente* (*apparent o auditory source width*, ASW) occupata da una fonte, che Rumsey (2001: par. 2.3) descrive come “spaziosità della fonte” (*source spaciousness*) (cfr. Carlile 2011: 55; Trębacz 2012); a ogni modo essa vale soprattutto nella stereofonia, in cui c'è una larghezza concreta (cfr. Begault 1999: 4-5). Tutte le caratteristiche sin qui trattate possono essere sia statiche sia dinamiche.

Rumsey ha elaborato un metodo di valutazione della qualità spaziale del sonoro usando un paradigma “*scene-based*” ove «the elements of the reproduced scene [are] grouped according to their function within the scene, at levels appropriate to the task» (2002: 657). Egli distingue *attributi dimensionali* da *attributi immersivi* (questi saranno trattati più avanti). I primi sono la larghezza (*width*), la profondità (*depth*) e la distanza, e si riferiscono – andando da un livello “micro” a un livello “macro” – alla singola fonte, a un insieme di fonti (*ensemble*) e – tranne che per la distanza – all'ambiente ed eventualmente a una scena.

La distanza e la profondità sono state spesso sovrapposte (cfr. p.es. Toole 1985), il che è in parte ragionevole. Infatti un senso naturale di distanza è possibile solo in un ambiente che ha una “profondità”, ossia una concretezza spaziale ancorché circoscritta a un singolo punto (ossia un singolo altoparlante). Si può insomma parlare di profondità del sonoro così come si parla dell'impressione di profondità di una superficie piana come quella di un dipinto o dell'immagine

bidimensionale cinematografica. In altri termini, anche se non c'è vera profondità – intesa come distanza concreta da una fonte – ce n'è una somiglianza, che è data appunto da quei fattori riconducibili alla firma spaziale (nonché all'impressione di movimento), la quale non sarebbe concepibile in assenza di “profondità”. Inoltre, come abbiamo accennato, aumentare o ridurre il volume di un suono può suggerire rispettivamente una riduzione o un aumento della distanza, incidendo sulla firma spaziale, ma ciò accadrebbe in modo poco naturale a meno di una coerente modifica della firma spaziale nel suo complesso, ovvero di una coerenza nella profondità, che è data dall'intensità insieme alla riverberazione. Inoltre il concetto di profondità al cinema, nella relazione audiovisiva, può efficacemente alludere allo spazio fuori campo, intendendo cioè che lo spazio filmico grazie al sonoro può essere denotato in modo più esteso rispetto a quanto permesso dall'immagine. Perciò il sonoro rende lo spazio “più profondo”, come se avesse una concretezza tridimensionale che si estende oltre la bidimensionalità dello schermo. Significativamente già tra i teorici del passato si rileva l'idea che il suono possa fornire all'immagine una terza dimensione, un senso di profondità (cfr. Balázs 1985: 119; Doane 1985b: 166; Camper 1985: 371). Ciò potrebbe dipendere anzitutto dal semplice fatto che il sonoro rende l'esperienza cinematografica (e in generale di immagini in movimento) più vicina alla realtà (cfr. Anderson 1996: 80-81; Larsson, Västfjäll e Kleiner 2003: 1) – la quale è tridimensionale – poiché aggiunge altri due stimoli sensoriali, ossia quello acustico e in misura indiretta e marginale quello tattile (cfr. Shilling, Zyda e Wardynski 2002: 3). Di tutto questo si ha almeno una conferma sperimentale: stando all'esperimento di Cullen et al. (2012) l'associazione di un sonoro mono a immagini stereoscopiche ne aumenta la percezione di profondità. La firma spaziale, come pure la spazializzazione acustica, può comunque presentarsi con diversi approcci, non necessariamente realistici, sia di per sé sia nell'interazione audiovisiva. La firma spaziale del suono può incrementare l'informazione spaziale dell'immagine, rafforzando il senso di profondità eventualmente già presente sullo schermo. Per giunta il suono off può denotare uno spazio più ampio di quello mostrato. La prospettiva sonora può essere elaborata cognitivamente a prescindere da un contestuale stimolo visivo, ma ovviamente nelle interazioni crossmodali con le immagini filmiche la configurazione dello spazio si complica: essa può infatti essere coerente o incoerente con quella visiva, in tutto o in parte. Questi aspetti audiovisivi sono trattati nel paragrafo 5.7.

## 5.5 – La spazializzazione acustica

### 5.5.1 – Terminologia e prassi

Per *stereofonia* (*stereophony*) in senso lato si intende una tecnica di registrazione e/o di riproduzione del suono che dà l'impressione di una localizzazione di diverse fonti sonore ascoltando da un'area denominata *sweet spot*; per far ciò si ricorre a due o più canali di informazioni. Questo principio deriva dalle ricerche di Blumlein (1933) e in quelle statunitensi coeve (cfr. Fletcher 1934a; 1934b). Viceversa nella *monofonia* (*monophony*, o *monaural sound*) c'è un solo canale di informazioni. Come abbiamo visto è comune un uso più ristretto del termine stereofonia, soprattutto nella forma abbreviata *stereo* (cfr. p.es. Rumsey 2001: par. 3.1; Breebaart e Faller 2007: 168), con il quale si denota la tecnica a due soli canali, che è la forma di stereofonia più diffusa al di fuori del cinema. Ci sono inoltre casi che potremmo chiamare di *pseudo-stereofonia*, «a family of techniques that allow the derivation of two channels of sound from a monaurally recorded, one-channel source» (Begault 2000: 181; cfr. Breebaart e Faller 2007: 56-57). Questa definizione descrive esattamente tecniche che editori di dischi dal 1959 chiamavano “*duophonic* (o *electronically reprocessed*, o *rechanneled*) *stereo*” (Barry 2015: 132); tuttavia ritengo che il termine pseudo-stereofonia possa denotare più genericamente anche casi di premeditata spazializzazione acustica di un segnale mono che viene inviato in più punti dell'ambiente d'ascolto, come il Perspecta o il Sensurround. Esistono anche casi che Holman definisce *multi-point mono*, in cui suoni diversi sono diffusi in più punti dell'ambiente ma, diversamente dallo stereo, «each of the component parts lacks a recorded space for it to “live” in»; infatti lo stereo vero è «“solid”, that is, each sound source is meant to produce a sensation that it exists in a real 3-D space» (2008: 109). Il comune caso di solito chiamato *dual mono* (benché il termine possa essere usato in ambito professionale nel senso di *split stereo*; Avid Technology 2014: 312; Dolby Laboratories 2000: 4-14) è ulteriormente diverso: si tratta della diffusione del segnale mono su due canali, senza rielaborazione stereofonica, ma non a scopo di spazializzazione bensì solo perché gli attuali sistemi di riproduzione audio sono tutti almeno a due canali. Nel caso di un film mono in digitale è quasi certo che il sonoro sia nel formato a due canali, quindi stereo, ma i due canali sono identici; se fosse in formato mono sarebbe comunque diffuso da ambedue gli altoparlanti di un sistema stereo classico (come un normale televisore), senza per questo diventare stereofonico.

Si parla più specificamente di *quadrifonia* (*quadraphonics*) per indicare il ricorso a quattro canali (nell'industria musicale), e più in generale di *suono multicanale* (*multichannel sound*) per denotare l'uso di tre o più canali. Si suole inoltre opporre il suono stereo al *suono surround*

(*surround sound*), intendendo col primo ancora una volta l'assetto a due canali e con il secondo il ricorso ad almeno tre canali di cui almeno uno diffuso posteriormente o lateralmente. Si ricorre anche alla nomenclatura “m/n”, ove m è il numero di canali frontali e n quello di canali laterali o posteriori (quindi una indicazione più precisa dello standard 5.1 sarebbe 3/2.1) (Dolby Laboratories 2000: 4-14)<sup>47</sup>. In ogni caso si evita di parlare di stereofonia in riferimento all'ascolto in cuffia, preferendo in questo caso il termine *binaurale* (*binaural*), in cui il segnale può essere “stereo” (cioè diverso tra sinistra e destra) ma è inviato separatamente alle due orecchie, mentre nella stereofonia *stricto sensu* entrambi i canali giungono a entrambe le orecchie, sebbene con differenze interaurali di tempo e di intensità. È bene ricordare che in senso percettivo “binaurale” indica una relazione tra ambedue le orecchie, quindi per evitare equivoci sarebbe opportuno riferirsi al senso tecnologico del termine come *audio* o *sistema binaurale* (*binaural audio* o *system*), oppure come *head-related audio rendering systems* contrapposto a *field-based methods* (Nordahl e Nilsson 2014: 221). Inoltre nelle scienze cognitive la stimolazione di un singolo orecchio è detta *monotic*, mentre quella di ambedue le orecchie è detta *diotic* qualora il segnale sia il medesimo e *dichotic* qualora i segnali siano differenti (cfr. Kaplan e Pickett 1981; Bregman 1990).

Oggi giorno anche l'espressione “sonoro multicanale” è diventata problematica, poiché esistono sistemi che oltre (o in alternativa) ai canali usano gli oggetti audio; dunque il termine canale è ora usato in senso strettamente tecnico, nella sua accezione tradizionale, e non più come sinonimo *de facto* di flusso di informazioni. Inoltre, soprattutto nell'ambito di videogiochi e realtà virtuale, si stanno diffondendo espressioni quali *3D audio/sound* o *spatial audio/sound* per indicare una spazializzazione forte che dà l'impressione di denotare uno spazio a 360° in cui è possibile orientarsi. Nell'ambito delle scienze applicate il concetto di tridimensionalità – quindi l'espressione audio 3D – non è appropriato al sonoro privo di dimensione verticale concreta o simulata, assumendo che un'autentica tridimensionalità prevede tre assi ortogonali (x, y, z) e non solo due (x, y). Da questo punto di vista il sonoro mono è puntiforme ovvero “adimensionale” (una sola sorgente), quello stereo solo frontale è unidimensionale (più sorgenti nella retta y), infine quello multicanale solo orizzontale è bidimensionale (più sorgenti nelle rette ortogonali y e x) (Guastavino e Katz 2004; Van Baelen et al. 2012: 3; Henein 2013). Per giunta la configurazione surround tuttora più consueta, il 5.1, non permette un'ottimale spazializzazione a 360° (cfr. Rumsey 2001: par. 4.3.1; Breebaart e Faller 2007: 12). Il termine *periphony* (Gerzon 1973) per indicare la riproduzione verticale oltre che orizzontale è poco usato.

Nella discussione relativa all'intervento di Grignon (1953), Richard H. Ranger asserisce che secondo lui si ha stereofonia laddove si offra al pubblico «an apparent location of the sound by the

<sup>47</sup> Rumsey usa l'indicazione “n-m stereo”, invertendo il significato delle due lettere (2001: par. 3.1).

direction and intensity of what it receives regardless of the means by which this is accomplished» (citato in Grignon 1953: 378). A mio avviso questo principio è cruciale nella definizione di stereofonia, sebbene fino ad allora – e ancora per qualche anno a seguire – si presupponesse un evento originale ripreso con una fedeltà spaziale, come dimostra la nomenclatura di Haynes (1954). Quello che suggerisce buona parte della letteratura tecnica, come pure l'etimologia, è che sia ha stereofonia ove si abbia un audio solido, il che è permesso dalla propagazione di stimoli sonori in più punti dello spazio in relazione al *sweet spot*; in questo contesto si può produrre l'effetto di immagine fantasma ovvero di *summing localization* (cfr. Blauert 1997: 204; Rumsey 2001: par. 3.1.1; Holman 2008: 184-187; Breebart e Faller 2007: 32-34; Middlebrooks 2015: 107; Nelson, Donovan e Takahashi 2015: 3). Ciò costituisce l'essenza della stereofonia, a prescindere non solo dal numero di canali (almeno due, ovviamente) ma anche dal modo in cui tale effetto spaziale è realizzato. Se si prendesse la stereofonia come un sistema di fedeltà spaziale – al modo dei primi ricercatori, che partivano dal presupposto del rispetto di una performance unitaria, tipica della musica colta nonché di una messa in scena teatrale – si dovrebbero escludere dal suo ambito tutte le musiche e le colonne sonore in cui l'effetto stereofonico è determinato esclusivamente dal missaggio piuttosto che dalla ripresa. Il termine stereofonia è stato usato anche per applicazioni cinematografiche multicanali – p.es. il CinemaScope, il Dolby Stereo – a prescindere dalla fedeltà a un evento originario, invero per solito inesistente (cfr. testi tecnici quali Wysotsky 1971 e Alkin 1973). Per quanto sia ovviamente utile distinguere tra diverse tecniche di spazializzazione, resta fermo che la modalità di fruizione sia immutato: si tratta di ascoltare una registrazione da un'area specifica, il *sweet spot*, il che permette eventualmente di avere immagini fantasma nello spazio tra gli altoparlanti. In altri termini i processi cognitivi restano i medesimi, come pure la modalità di fruizione concreta. Occorre considerare che ci può essere una registrazione che preveda il *sweet spot* evitando qualsiasi immagine fantasma, qualora i segnali degli altoparlanti siano assolutamente diversi (cfr. Breebaart e Faller 2007: 34-35); tuttavia permane l'effetto di solidità, ossia di un ambiente sonoro definito presente davanti o intorno a noi. Il caso del multi-mono è più problematico, poiché se il segnale è diffuso da più altoparlanti frontali entro una certa ampiezza l'effetto spaziale è irrisorio, mentre qualora detto segnale fosse diffuso, attraverso comandi automatici o manuali, da altoparlanti intorno a un pubblico posizionato in un'area prestabilita si avrebbe un effetto di immersione e quindi in un certo senso di solidità, sebbene non possa produrre immagini fantasma a guisa della stereofonia. In questo caso il *sweet spot* avrebbe quindi minore rilevanza, eppure non sarebbe indifferente sedersi al centro della sala piuttosto che in prossimità di una parete, il che renderebbe sbilanciata la percezione del suono. In ogni caso si tratta di un allontanamento dal paradigma stereofonico tale da permetterci di concettualizzare questa

circostanza come qualcosa di diverso dalla stereofonia, dunque una pseudo-stereofonia, o una monofonia spazializzata, o un mono multicanale (Hirvonen, Tikander e Pulkki 2004: 3) – nel senso che il singolo segnale viene veicolato su più canali di un impianto, indi su più altoparlanti. Casi del genere sono quelli del Vitasound, del Perspecta e del Sensurround, che infatti sono indicati come mono, nonché in parte del Fantasound. Col Vitasound era possibile riprodurre la colonna mono anche da altoparlanti esterni allo schermo, mentre il Perspecta permetteva di riprodurla dal centro, da sinistra, da destra o contemporaneamente da tutti e tre gli altoparlanti dietro lo schermo. Il Fantasound era a tre canali ma in più permetteva di inviare manualmente il segnale desiderato ad altoparlanti ai lati, dietro e in cima alla sala. Ovviamente tutti gli effetti pseudo-stereofonici descritti sinora sono ottenibili anche con un sistema stereofonico. I casi di rielaborazione di un segnale mono in stereo sono stereofonici a valle, ancorché non a monte: pertanto dovrebbero rientrare nella pseudo-stereofonia ma non sono descrivibili come multi-mono, perché ci sono effettivamente segnali diversi. Il termine più consueto per questo tipo di operazione – desueta – è probabilmente *duophonic sound*. Le applicazioni multicanali che si avvalgono di diversi principi per la resa spaziale, abbandonando il *sweet spot* a favore di un approccio policentrico, presentano un allontanamento ancor più radicale dal paradigma stereofonico, ma non sono di nostro interesse in quanto estranee al cinema. La stereofonia per antonomasia, ovvero lo stereo, è quella a due canali, tanto che per scansare gli equivoci si predilige il termine multicanale laddove i canali siano più di due, ma come abbiamo visto ci può essere un sonoro multicanale non stereofonico. Perciò per evitare fraintendimenti si dovrebbe parlare di stereofonia a x canali (*x-channel stereo*), ove  $x \geq 2$ .

La ricerca sul sonoro spaziale parte dal presupposto di fonti anzitutto frontali a chi ascolta, ed eventualmente anche surround. In effetti la capacità umana di localizzazione uditiva è più precisa di fronte, mentre diminuisce posteriormente e ancor più lateralmente (Breebaart e Faller 2007: 27): ecco perché è più complesso produrre immagini fantasma posteriormente e soprattutto lateralmente (Massey 2004: par. 4.3; Juge, Pras e Frissen 2016: 2). Perciò in linea di principio anche un sistema solo laterale o posteriore – con al massimo una fonte frontale – può produrre un effetto di solidità, ma non mi risulta che esistano configurazioni stereofoniche con meno di due canali frontali.

Una delle presunte novità dei sistemi immersivi è come abbiamo visto una riconsiderazione del *sweet spot*. In un *white paper* del Dolby Atmos si dichiara che al posto del *sweet spot* c'è un approccio *room-centric* che permetterebbe di fruire il suono in modo adeguato da qualunque poltrona nella sala (Dolby Laboratories 2014: 6; cfr. Sergi 2013: 110-111, 113): pur ammettendo che l'ascolto sia ottimale da qualunque poltrona – ipotesi dubbia – resta il fatto che si offre un campo sonoro univoco, monocentrico. Si può supporre che accade qualcosa di analogo a ciò che altre aziende descrivono mantenendo il termine *sweet spot*: secondo la Barco l'AuroMax massimizza le

dimensioni del suddetto (Barco 2015: 17), mentre a proposito del DTS Sound si parla di «Broad sweet spot with elevated sound image» (DTS 2016). Lo scopo è insomma quello di garantire una resa ottimale da qualunque sedile della sala, piuttosto che quello di superare il paradigma stereofonico. La tecnica degli oggetti audio a sua volta non compromette questo paradigma: sebbene il segnale e il canale non si possano più considerare sinonimi come da tradizione – l’oggetto è un segnale sonoro pur non trattandosi di un canale – il processo cognitivo resta immutato. Peraltro si consideri che sul piano percettivo non c’è alcuna differenza tra canale e oggetto, trattandosi di una differenza prettamente tecnica (ancorché non trascurabile): la percezione sarebbe la medesima di un ipotetico sistema che avesse tanti canali quanti sarebbero i canali sommati agli oggetti.

### 5.5.2 – La direzione

In un sonoro monofonico, o comunque in una fonte singola, non si può ottenere direzionalità, poiché la direzione è una sola – si tratta infatti di un sonoro puntiforme, “adimensionale”. Affinché si abbia direzionalità occorre che vi sia più di una direzione, ovvero la disposizione di più fonti in almeno una dimensione. Le tre dimensioni sono la *larghezza*, la *lunghezza* e l’*altezza*. Quindi la disposizione di suoni in queste dimensioni è possibile con una spazializzazione acustica relativa alle medesime. Perciò il sonoro può essere unidimensionale, bidimensionale o tridimensionale, a seconda di quante di queste dimensioni ne caratterizzano la spazializzazione. Il sonoro unidimensionale viene spazializzato nella larghezza, come nello stereo classico, e quello bidimensionale nella larghezza e nella lunghezza, come nella stereofonia cinematografica tradizionale. Il concetto di tridimensionalità applicato al sonoro è stato usato spesso in modo improprio per riferirsi o all’effetto di profondità o a una precisione direzionale nelle due dimensioni di un sistema surround; in senso stretto si riferisce invece ai suoni spazializzati in tutte e tre le dimensioni, come accade negli attuali sistemi immersivi.

La localizzazione uditiva normalmente utilizzata nel mondo reale è quella orizzontale (relativamente all’azimut), che dipende da indizi binaurali, specificamente dalle differenze interaurali di tempo (*interaural time differences*, ITDs) e dalle differenze interaurali di intensità o di livello (*interaural intensity or level differences*, IIDs o ILDs). I due concetti si riferiscono rispettivamente al fatto che il suono arriva alle due orecchie in tempi diversi e con intensità diversa a seconda dell’orientamento orizzontale della fonte rispetto alle medesime. Questi fenomeni rientrano nell’insieme delle *head-related transfer functions* (HRTF), perché dipendono appunto dalla posizione e dalla conformazione del capo, il quale ostruisce il passaggio delle frequenze al di

sopra di 800 Hz (cfr. Blauert 1997: 36-200; Pulkki, Lokki e Rocchesso 2011: 140; Breebaart e Faller 2007: 27-28; Yost 2001: 443-444; Middlebrooks 2015: 100). Ciò permette di comprendere la direzione di provenienza del suono, anche a occhi chiusi, al di fuori del campo visivo o comunque in punti a maggiore angolazione in cui la dominanza visiva è minore (Begault 1999: 7-8). Le differenze inter-aurali sono note fin dalle ricerche di Rayleigh (1876; 1907), fondamentali per l'elaborazione della cosiddetta teoria *duplex*. Essa afferma che l'ITD e l'ILD dominano la localizzazione rispettivamente a frequenze basse – soprattutto < 1 kHz – e a frequenze alte – soprattutto > 2 kHz – con una maggiore difficoltà di localizzazione tra 1 e 2 kHz. Le differenze sono nulle solo quando la fonte è equidistante dalle orecchie, frontalmente o posteriormente, o quando due suoni sono nel “cono di confusione”, da cui risulta ambiguità nella distinzione tra davanti e dietro (soprattutto nell'ascolto in cuffia: Gutierrez-Parera e Lopez 2016) e nell'elevazione; questa circostanza nella realtà si elimina con un minimo movimento del capo così da avvalersi delle HRTF (Middlebrooks 2015: 101, 103-104; Zhang e Hartmann 2010: 39; Hulusic et al. 2012: 115; Pulkki, Lokki e Rocchesso 2011: 140; Begault 2000: 41; Yost 2001: 440-442; Letowski e Letowski 2012: 8; Rumsey 2001: par. 2.1; Holman 2008: 178-180). La localizzazione orizzontale è più accurata di quella verticale (cfr. Morrongiello 1988; Perrott e Saberi 1990; Barbour 2003; Yost 2001: 444; Weger, Marentakis e Höldrich 2016), il che è comprensibile in termini evolucionistici perché gli esseri umani e più in generale i mammiferi hanno necessità di percepire la presenza di altri animali a livello del terreno (Heffner e Heffner 1992: 228; Begault 2000: 37), ma in ogni caso occorre distinguere tra il piano verticale mediano e quello laterale: nel primo caso ciò che muta è solo l'angolo di incidenza del suono rispetto alle orecchie, senza alcuna differenza interaurale, mentre nel secondo si presentano differenze interaurali di intensità e di tempo (cfr. Butler e Humanski 1992). In assenza di indizi binaurali restano quelli spettrali, definiti monoaurali in quanto non necessitano di due orecchie; perciò i padiglioni auricolari sono particolarmente importanti per la localizzazione sul piano verticale mediano, dato che a seconda dell'angolazione verticale la percezione dello spettro si modifica (cfr. Begault 2000: 50; Pulkki, Lokki e Rocchesso 2011: 141; Wallis e Lee 2015: 767; Middlebrooks 2015: 103). Wallis e Lee (2015) rilevano che la localizzazione sul piano verticale mediano è governata dall'effetto *pitch-height*, ove le frequenze più alte sono percepite come localizzate più in alto nello spazio rispetto alle frequenze basse; inoltre non risulta esserci l'effetto precedenza (cfr. più avanti). In ogni caso modifiche degli indizi di localizzazione dovute a un movimento relativamente lento della fonte migliorano la localizzazione (Breebaart e Faller 2007: 46). Le frequenze più basse tendono a propagarsi in maniera omnidirezionale (Rumsey 2001: par. 1.2.1), quindi sono meno puntiformi: ecco perché si può usare

un singolo subwoofer, come nel 5.1, anche per rafforzare le frequenze basse di suoni provenienti da altri canali (Holman 2008: 180; cfr. Rumsey 2001: par. 5.3.5; Massey 2004: par. 3.3.2).

Il *precedence effect*, così chiamato a partire dallo studio di Wallach, Newman e Rosenzweig (1949) ma noto anche come *law of the first wavefront* (cfr. Brown, Stecker e Tollin 2015: 1), è un effetto psicoacustico che influenza la percezione della direzione: infatti noi percepiamo il suono diretto come se fosse fuso con le sue riflessioni senza percepire queste come suoni distinti provenienti da altre direzioni, perciò ne consegue un fenomeno di *localization dominance* da parte delle onde dirette. Ciò avviene quando il ritardo è entro i 5 ms, oltre il quale invece si ottiene l'eco e dunque più suoni distinti. L'effetto precedenza permette anche la *summing localization*, ovvero l'impressione che un suono sia localizzato tra due altoparlanti di un sistema stereofonico – quindi le due direzioni si sommano in un'unica direzione (cfr. Blauert 1997: 204; Yost 2001: 445; Begault 2000: 36-37; Breebaart e Faller 2007: 33-34; Holman 2008: 111; Middlebrooks 2015: 107; Nelson, Donovan e Takahashi 2015: 3).

Un altro fenomeno percettivo di interesse relativamente alla cognizione spaziale è quello noto come effetto *cocktail party* (cfr. Cherry 1953; Cherry e Taylor 1954; Pollack e Pickett 1957; Williams 2012), che consiste nel miglioramento della comprensione di parole cui si rivolge l'attenzione in un ambiente in cui ci sono altre voci o rumori. È quindi un caso di *auditory stream segregation* (cfr. Bregman 1990; Begault 2000: 184; Yost 2001: 452), ovvero di una forma di mascheramento uditivo (Yost 2001: 451; Hulusic et al. 2012: 116). Questo effetto può essere sfruttato in un messaggio stereofonico, come già avevano asserito Pollack e Pickett (1958). Begault e Erbe (1994) dimostrano come un segnale spazializzato favorisce l'intelligibilità: la comprensione delle parole pronunciate da una voce ascoltata in cuffia insieme a un rumore diotico migliora quando il segnale con la voce è dicotico piuttosto che diotico, soprattutto a certe angolazioni. Bronkhorst (2000: 123) afferma che l'intelligibilità delle parole è determinata dalla loro distribuzione spaziale, oltre che dal numero di fonti e dal profilo temporale dei suoni interferenti<sup>48</sup>. Insomma la separazione spaziale ottenibile grazie alla stereofonia può aiutare l'intelligibilità piuttosto che denotare una distribuzione spaziale. Ciò può essere utile per missare una musica corposa e molti effetti insieme al dialogo: questo può mantenere una maggiore intelligibilità se viene missato in un canale mentre le altre due colonne propagano da canali diversi. È il principio attuato per *L'uomo d'acciaio* (*Man of Steel*, USA 2013), stando a quanto riporta Wright (2015: 241-242), in cui i fonici di missaggio hanno sfruttato le sei aree surround del Dolby Atmos per propagare abbondante musica mentre i canali frontali erano riservati a dialoghi ed effetti principali. Un simile suggerimento è riscontrabile anche in un *white paper* del Dolby Atmos: «moving music

<sup>48</sup> Sui processi neurali delle informazioni sonore spaziali cfr. Yost (2001: 447-449).

offscreen frees up the screen loudspeakers for audio effects and can help clarify dialogue tracks» (Dolby Laboratories 2014: 10).

## 5.6 – Movimenti sonori

Gallese e Guerra sostengono che ci sono quattro movimenti fondamentali nel cinema atti a provocare risonanza nei/lle percipienti: «il movimento dei soggetti e degli oggetti presenti nell'inquadratura; il movimento *tra* le immagini, cioè il montaggio; il movimento ottico prodotto da dispositivi come lo zoom; il movimento fisico della camera» (2015: 147). Mi pare indispensabile aggiungere anche altri movimenti, scaturenti dal sonoro. Una prima distinzione facile da stabilire, poiché fondamentalmente tecnologica, è quella tra il movimento unidirezionale reso dai suoni puntiformi, proprio potenzialmente di qualunque film sonoro, e il movimento pluridirezionale reso dai suoni che si spostano tra altoparlanti, reso possibile dalla stereofonia. In entrambi i casi occorre distinguere ulteriormente se il movimento scaturisce dalla fonte denotata, dal punto d'ascolto o da entrambi; ciò è possibile sia col solo ascolto sia tanto più nella relazione audiovisiva. La disamina che segue vale sia per il sonoro unidirezionale sia per quello pluridirezionale, ma in quest'ultimo caso si può avere anche un movimento concreto di un oggetto audio qualora si sposti da un canale a un altro, il che è possibile in fase di registrazione con un assetto microfonico stereofonico e soprattutto in fase di missaggio usando il *panning*, ossia la panoramica sonora (cfr. Holman 2008: 110-119).

Neuhoff (2004: 95-97) spiega che ci sono due teorie per spiegare la percezione del movimento uditivo: la prima sostiene che ciò avviene comparando delle “*snapshot*” statiche di una fonte sonora; la seconda afferma che il sistema uditivo risponde direttamente al movimento sonoro, come dimostra una crescente quantità di dati psicofisici. Si può comunque ipotizzare che i due meccanismi interagiscano.

Il movimento sonoro può essere a mio parere inteso con quattro sfumature che pertengono a due coppie concettuali: la prima concerne la distinzione tra un movimento esplicito e uno implicito, mentre la seconda come accennato concerne la distinzione tra il movimento della fonte e quello del punto d'ascolto. Il movimento è *esplicito* quando *il suono si muove*, ossia la fonte o il punto d'ascolto si spostano nello spazio, mentre è *implicito* quando *il suono non si muove* ma capiamo che scaturisce dall'azione di un ente concreto, quindi presumibilmente da qualcosa che si muove pur senza spostarsi rispetto a un dato punto d'ascolto. Inoltre il movimento sonoro può scaturire dal movimento della fonte rispetto a un punto d'ascolto fisso, dal movimento del punto d'ascolto

rispetto a una fonte fissa, o dall'interazione tra i movimenti di entrambi, con due possibilità: il punto d'ascolto può mantenere la stessa relazione spaziale con la fonte oppure no.

In altri termini il movimento sonoro della fonte rispetto a punto d'ascolto fisso può essere esplicito o implicito. Il movimento è esplicito quando c'è la percezione di spostamento della fonte nello spazio denotato rispetto al punto d'ascolto, il che può accadere grazie a modifiche relative a intensità, frequenza e riverberazione (quindi di timbro) che offrono l'impressione di avvicinamento, di allontanamento, di entrambi (quindi di passaggio), o di angolazione (cfr. Lennox e Myatt 2007b: 3). Il movimento è invece implicito allorché vi sia la percezione di udire da un punto d'ascolto fisso un evento spazialmente statico (cioè in cui la fonte è in un determinato punto), il quale però scaturisce necessariamente dal movimento di un esistente, sia se si tratta di un'azione (p.es. un colpo alla porta) sia nel caso di un avvenimento privo di agente (p.es. lo scorrimento di un ruscello). Ci può essere altrimenti un movimento del punto d'ascolto, che anche in questo caso può essere esplicito o implicito. Il movimento del punto d'ascolto è esplicito quando questo si muove rispetto a una fonte fissa o rispetto a una fonte mobile con la quale il rapporto spaziale è variabile; il movimento è invece implicito quando il punto d'ascolto mantiene lo stesso rapporto spaziale in termini di distanza e di angolazione rispetto alla fonte mobile, perciò in assenza di una sensazione puramente uditiva dello spostamento.

Di Donato (2006: 85-86) sostiene che in assenza di indizi visivi la distinzione tra il movimento della fonte e il movimento del punto d'ascolto non può avvenire. Ciò tuttavia non è del tutto corretto. È vero che la nostra cognizione visiva è più accurata di quella sonora in termini spaziali, tanto che possiamo facilmente accorgerci se il movimento che percepiamo dipende dallo spostamento di un'attrice o da un movimento di macchina, mentre non c'è alcuna differenza lampante tra lo spostamento di un attore dal microfono e lo spostamento del microfono dall'attore (lo stesso dicasi per espedienti di missaggio che rendono fenomeni spaziali). Tuttavia bisogna tener conto della rilevanza del *contesto*, il quale ci permette di operare una simile distinzione. Infatti ci possono essere dei fattori che ci chiariscono qual è la natura del movimento anche senza riferimento visivo, sia in termini *bottom-up* (risposta percettiva) sia in termini *top-down* (inferenze cognitive).

Un esempio è utile per chiarire le caratteristiche del movimento sonoro. Ipotizziamo anzitutto di avere uno stimolo esclusivamente uditivo. È possibile percepire il movimento di una persona che si avvicina o si allontana grazie al rumore dei passi udito da un punto d'ascolto statico: pertanto percepiamo uno spostamento in relazione a esso grazie alla modifica della firma spaziale, il che è compreso come un movimento esplicito poiché la fonte denotata – la persona – si sposta nello spazio compiendo un tragitto. Cionondimeno il rumore di passi sul posto, ovvero uno scalpitio, può suggerire un movimento poiché comprendiamo che la persona si sta *muovendo*, ancorché nella

staticità della sua collocazione poiché la fonte denotata permane in un punto. La medesima considerazione vale per il nostro punto d'ascolto qualora la firma spaziale resti immutata. In altri termini è un movimento per così dire *senza spostamento*. Tuttavia se continuiamo a percepire lo scalpito ma la firma spaziale cambia possiamo sospettare che a muoversi non sia la persona ma il nostro punto d'ascolto. Se ci sono informazioni pregresse sullo spazio in questione o una descrizione verbale di quanto accade possiamo fare ulteriori inferenze; in più potrebbero esserci altre fonti che ci aiutino a comprendere la natura del movimento. Qualora vi sia una persona che parla presso il mare e a cambiare è solo la firma spaziale della voce vuol dire che il punto d'ascolto rispetto al mare, ossia nello spazio contestuale, resta fisso; il movimento della voce è esplicito. Se invece la firma spaziale dei due suoni cambia similmente si tratta di un movimento del punto d'ascolto; il movimento suggerito da entrambe le fonti è esplicito. Se infine a cambiare è solo la firma spaziale del mare capiamo che c'è uno spostamento del punto d'ascolto che mantiene il medesimo rapporto spaziale con la persona che parla ma non col mare; il movimento della fonte "mare" – o meglio il nostro movimento rispetto a essa – è esplicito, mentre quello della persona è implicito perché lo spostamento è inferito ma non percepito come tale rispetto al punto d'ascolto.

Per riassumere in un sonoro unidirezionale nel caso del movimento esplicito si ha uno *spostamento* della fonte sonora o del punto d'ascolto nello spazio denotato, mentre nel caso del movimento implicito si ha o una permanenza della fonte sonora e del punto d'ascolto nel medesimo punto dello spazio denotato oppure uno spostamento sia della fonte sonora sia del punto d'ascolto mantenendo la medesima relazione spaziale (v. tabella 3). In un sonoro pluridirezionale il movimento esplicito può inoltre essere appunto *direzionale*, allorché il suono si sposti tra canali. Ovviamente la presenza di immagini cui i suoni sono sincroni fornisce un'integrazione crossmodale che permette di comprendere meglio la rappresentazione nonché di incrementare la cognizione incarnata. Ho voluto scomporre nel modo più dettagliato possibile la questione, ma in ogni caso durante la fruizione il processo cognitivo tende a essere più semplice sia per la presenza di immagini sia perché ciò che conta è la comprensione di quanto accade, la quale nel cinema narrativo è condizionata dalla narrazione. A questo livello di analisi la natura del punto d'ascolto – oggettiva o soggettiva – è irrilevante, benché ovviamente non lo sia affatto in assoluto, come vedremo.

<i>Movimento sonoro</i>	Fonte fissa	Fonte mobile
Punto d'ascolto fisso	Implicito (assenza di spostamento assoluto)	Esplicito
Punto d'ascolto mobile	Esplicito	Implicito se la relazione spaziale è costante (assenza di spostamento relativo) Esplicito se la relazione spaziale è variabile

Tabella 3: movimenti sonori unidirezionali.

L'effetto Doppler è un fenomeno acustico che, poiché si basa su indizi monoaurali, denota movimento anche con una propagazione puntiforme. Si tratta dell'aumento o della diminuzione della frequenza di un suono rispettivamente all'avvicinarsi o all'allontanarsi della sua fonte rispetto all'ascoltatore (Begault 2000: 40; McBeath e Neuhoﬀ 2002; Pulkki, Lokki e Rocchesso 2011: 178-179; Letowski e Letowski 2012: 78-79; Middlebrooks 2015: 105).

Quando il movimento è di avvicinamento il suono attiva una risposta peculiare che è stata studiata dalle neuroscienze (anche con *brain imaging*) sotto la denominazione di *auditory looming bias* (corrispettivo sonoro del *visual looming bias*). Infatti i suoni che si avvicinano sono più salienti biologicamente di quelli che si allontanano perché possono rappresentare l'avvicinamento di un predatore, mentre i secondi possono segnalare una fuga riuscita o una caccia fallita (Cappe et al. 2008: 1045). Di conseguenza si crea un bias percettivo poiché tendiamo a sovrastimare l'incremento d'intensità e quindi il movimento di un suono verso di noi, rispetto alla riduzione d'intensità di un suono che si allontana, nonché a sottostimare il tempo necessario al contatto. Ciò accade significativamente più con suoni vocali che col rumore bianco, ovvero con suoni affini a quelli naturali (cfr. Neuhoﬀ 2001; Seifritz et al. 2002; Hall e Moore 2003; Maier et al. 2004; Maier, Chandrasekaran e Ghazanfar 2008; Neuhoﬀ, Planisek e Seifritz 2009; Tyll et al. 2012). Per giunta Neuhoﬀ (2016) ha dimostrato che, rispetto ai suoni in allontanamento, quelli in avvicinamento sono percepiti come più veloci, soprattutto all'aumentare della loro vicinanza. Il bias può essere ulteriormente incrementato nel caso di suoni che hanno una connotazione negativa (Tajadura-Jiménez et al. 2010b) e in condizioni di ansia (Riskind et al. 2014), il che conferma la rilevanza dello stato emotivo del/della percipiente per la cognizione audiospaziale.

## 5.7 – Campo visivo e campo sonoro

### 5.7.1 – Questioni generali

In medicina si definisce *campo visivo* (*visual field*, o *field of vision*) lo spazio che può essere abbracciato dallo sguardo tenendo fisso un solo occhio o entrambi in un punto. Per estensione nel cinema si può intendere per campo visivo la porzione di spazio visibile sullo schermo. Si definisce invece *campo sonoro* (*sound field*) la regione dove hanno sede le vibrazioni sonore. In questa definizione, diversamente dalla precedente, non si fa riferimento al soggetto umano. Il *campo uditivo* (o *di udibilità*) non è connesso allo spazio ma all'insieme di frequenze acustiche udibili da un individuo tra la soglia di udibilità minima e quella di udibilità massima. D'altronde l'udito non è selettivo come la vista, potendo percepire contemporaneamente da qualunque direzione, ancorché con una diversa sensibilità. Come abbiamo visto Lennox e colleghi intendono il *sound field* come una rappresentazione semplificata, "task-specific", di un ambiente sonoro (*sound environment*) (Lennox, Vaughan e Myatt 2001b: 1), ovvero come «a cartoon of the audible part of an environment» (Lennox e Myatt 2007b: 4). Tuttavia questo contraddice la definizione riconosciuta della locuzione. Rumsey (2002) usa *environment* in un senso che ricorda più quello di *sound field* secondo Lennox e colleghi, benché il termine porti con sé un'altra connotazione, mentre parla di *scene* per intendere il complesso dello spazio sonoro presentato. In ogni caso si può definire questo tipo di spazio *task-specific* come l'insieme dei suoni udibili da un dato *punto d'ascolto* (*point of audition*), in analogia col punto di vista (che essendo selettivo coincide col campo visivo). Nel cinema quindi il campo sonoro è quello che scaturisce dalla propagazione acustica della colonna sonora a denotare un certo spazio più o meno concreto, mentre il punto d'ascolto è un concetto più ristretto che si riferisce alla cognizione prospettica del suono.

È giusto distinguere tra processi di percezione visiva e uditiva, con la relativa terminologia, senza per questo trascurare i correlativi sonori di certe dinamiche visive. Ma più a monte il problema non dovrebbe essere quello di rintracciare costantemente gli analoghi sonori dei processi visivi, bensì di verificare il contributo del sonoro alla percezione e alla cognizione, nel nostro caso relativamente alla dimensione spaziale.

Come abbiamo visto Chion sostiene che il montaggio del suono non produce "inquadrature sonore" (2009: 223), e che non esiste colonna sonora (*bande-son*), ovvero un luogo «where sounds gather and make a unified front» (2009: 227). L'autore afferma anche che «[t]he notion of an auditory field in cinema is itself ambiguous [as it is] in everyday life since the location of a sound is never really clear. [We can] describe the place of [a] sound as a sphere with a central core and fuzzy

outer contours» (2009: 248). Altrove Chion spiega che i vari suoni filmici «do not constitute and integrated, homogeneous entity in themselves for the simple fact of belonging to the world of sound» (2009: 491), poiché si attua il processo chiamato *audio-division*, ossia il senso di divisione che i suoni provocano alle immagini – perché solo alcuni elementi visivi sono così rafforzati – e che l'immagine provoca al suono – dentro e fuori campo, «thereby preventing the constitution of a soundtrack as a unified ensemble» (2009: 467-468). Ne consegue che in presenza di immagini le relazioni che i vari suoni stabiliscono tra loro sono molto più deboli rispetto a quelle che ciascuno di essi costituisce con un dato elemento visivo; ancora una volta, ciò è possibile per l'assenza di un quadro uditivo che possa racchiudere i vari suoni (2009: 491). Chion però non analizza adeguatamente la questione, poiché tiene conto solo di una diversa fenomenologia dello stimolo acustico rispetto a quello visivo senza indagare i processi percettivo-cognitivi. Il primo problema è infatti prettamente terminologico, dato che per colonna sonora si dovrebbe intendere l'insieme dei suoni in un film, ovvero come un dato tecnico, che ovviamente esiste. La distinzione di Chion tra *bande-son* – termine rifiutato – e *piste sonore* – termine accettato, poiché proviene dal gergo della tecnologia del suono – è anodina poiché entrambe le espressioni denotano un fatto tecnico<sup>49</sup>. Il disagio dell'autore nasce dal fatto che *bande* si usa anche per la colonna visiva (*bande-images*), che è delimitata da un quadro, ma questo non dovrebbe comportare automaticamente l'attribuzione al sonoro di qualità proprie solo del visivo. È vero che l'immagine è contenuta in un'inquadratura che ne è dunque il contenitore, mentre il suono non ha un contenitore percettivo delimitato, tuttavia ciò dipende dalla diversa natura del suono rispetto all'immagine, a prescindere dalla mediazione tecnologica. Infatti anche nella realtà le immagini, che non sono necessariamente contenute in qualcosa, sono percepite come delimitate spazialmente, perché la vista è molto selettiva e il campo visivo ha dei contorni così come gran parte degli oggetti della percezione. I suoni sono invece delle vibrazioni acustiche che possono giungere alle nostre orecchie da qualunque direzione contemporaneamente, e poiché viaggiano nell'aria senza permetterci di percepirne dei contorni sfuggono a un ipotetico contenitore. Inoltre il suono è meno soggetto a ostacoli, diversamente dalla vista: la delimitazione spaziale del campo visivo si dà nelle tre dimensioni, mentre il campo sonoro dipende solo dalla distanza e da ostacoli isolanti. Pertanto non occorre prendere il concetto di

<sup>49</sup> Nella traduzione inglese del libro di Chion, Gorbman decide di tradurre *piste sonore* con *sound channel*, pur dichiarando la problematicità del termine, dal momento che non ne trova uno diverso da *track* per tradurre letteralmente *piste* (cfr. Chion 2009: 234). In effetti *channel* può essere confuso col canale inteso come un flusso audio da riprodurre, e in una colonna sonora possono esserci più canali. D'altronde in inglese si usa anche il termine *track* per indicare il canale (p.es. "Dolby 6-track"). L'italiano potrebbe essere avvantaggiato perché in ambito cinematografico sono riconosciute le locuzioni sinonime colonna sonora, banda sonora e pista sonora; tuttavia non vi identifico alcuna connotazione che possa giustificare l'esigenza di una differenziazione affine a quella attuata in francese da Chion. Dette locuzioni nascono per indicare un oggetto concreto, ossia la porzione di film su cui è inciso il suono, ma si usano per estensione in riferimento al sonoro del film a prescindere dal suo formato.

colonna sonora come un equivalente sonoro della colonna visiva nel senso della sua qualità di contenitore, di delimitatore spaziale di stimoli. Proprio perché l'udito funziona diversamente dalla vista, è lecito parlare in senso prettamente tecnico di colonna sonora per riferirsi all'insieme dei suoni di un film, così come la colonna visiva costituisce l'insieme delle immagini. Inoltre è lecito parlare, stavolta in senso percettivo e cognitivo, di punto di vista e punto di ascolto per indicare rispettivamente ciò che vediamo e ciò che ascoltiamo in un dato momento. L'ambiente sonoro filmico può offrirci delle informazioni spaziali similmente a quanto accade nella realtà, grazie alle dinamiche proprie dell'udito, che non includono la delimitazione spaziale analoga a quella del campo visivo. Per giunta non è necessario negare l'esistenza di una colonna sonora affinché si possano considerare i "suoni inauditi", come invece asserisce Chion: ammettendo ciò si dovrebbe parimenti rinnegare il concetto di colonna visiva, dato che esistono anche delle "immagini invisibili" allorché si cela allo spettatore ciò che è sotto gli occhi del personaggio, e soprattutto perché qualsiasi inquadratura presenta solo una porzione di spazio estrapolata da uno spazio più ampio (cfr. Aumont et al. 2009: 17-23; Elsaesser e Hagener 2007: 3-30).

La tesi dell'audio-divisione è altrettanto discutibile, sia sul piano empirico sia integrando i dati delle scienze cognitive. Chion ne fa una regola, ma essa sembra più un'eccezione proprio perché il cinema ricorre usualmente al sincronismo, da cui la sincretisi teorizzata dall'autore stesso. Quindi normalmente percepiamo eventi visivi ed eventi sonori sincroni come un'entità omogenea, per nulla divisa. Quando vi sono suoni extradiegetici – su tutti la musica di commento – essi non comportano necessariamente un senso di divisione quanto di integrazione, di aggiunta, di sostegno *et sim.* Asserire che le immagini dividano i suoni è un'approssimazione eccessiva, perché ciò cui allude Chion non è tanto una divisione – nel senso di scomposizione, separazione, ripartizione – di qualcosa che altrimenti sarebbe unito, quanto una costruzione spaziale crossmodale. Detto altrimenti, ciò che il visivo fa al sonoro è influenzarne la cognizione spaziale più di quanto non accada viceversa, come abbiamo verificato grazie alle scienze cognitive.

La tesi dell'inesistenza di un microfono simbolico – equiparabile alla macchina da presa simbolica, che invece esiste (Chion 2001: 94-95; 2009: 292-293) – ha un fondamento benché necessiti di essere formulata diversamente. È vero che durante la visione di un film si tende a non porsi problemi circa la ripresa audio o il trattamento spaziale dei suoni, non solo per pigrizia ma anche per i limiti della cognizione audiospaziale. Nel caso di un'inquadratura basta un colpo d'occhio per avere un'idea del posizionamento spaziale della macchina da presa, mentre nel caso del sonoro è decisamente meno immediato e più complesso ragionare sul posizionamento microfonic. Inoltre nel caso del sonoro c'è da sempre molto più lavoro di costruzione in post-produzione; per giunta nel montaggio sonoro e nel missaggio non si ragiona tanto in base a un

ipotetica posizione del microfono quanto alla resa spaziale, mentre allorché si aggiungono elementi visivi a un'inquadratura – o la si elabori interamente – c'è giocoforza un problema di obbedienza all'originaria – o all'immaginaria – posizione della macchina da presa. Ma l'idea di un microfono simbolico è legittima proprio perché si tratta di qualcosa di *simbolico*, il che esclude la pretesa di rilevare la *reale* posizione di un microfono durante le riprese. Se ci sembra di vedere e ascoltare da vicino un personaggio significa che esistono una macchina da presa e un microfono quantomeno *simbolici* che rendono possibile ciò. A ogni modo convengo che il punto nodale non è tanto questo bensì quello del punto d'ascolto (cfr. par. 5.7.2).

Chion sostiene che noi non percepiamo delle unità del montaggio sonoro paragonabili alle inquadrature (il che rafforzerebbe l'idea che non esista una *bande-son*). Ciò è senza dubbio la convenzione, ma ci possono essere circostanze differenti. Si può infatti creare con il solo stimolo uditivo un effetto affine a quello che scaturisce dallo stacco visivo, in almeno due circostanze. La prima, più generale, è quella in cui un frammento sonoro viene tagliato – all'inizio o alla fine – in modo repentino o comunque evidente, tanto da permetterci di comprendere che c'è stato appunto un taglio di montaggio. La seconda, più particolare, è quella trattata da Di Donato in cui vi è un cambiamento evidente di prospettiva sonora, affine a quello che scaturisce nel campo visivo. In una scena si possono presentare diversi punti d'ascolto, ben distinti l'uno dall'altro, da cui risulta la percezione di «segmenti acustici [...] chiaramente identificabili» (2006: 117, n. 4). Queste opzioni non sono state utilizzate regolarmente nel cinema, il che può motivare il fatto che non siano state considerate nemmeno in teoria, salvo osservare casi di montaggio sonoro frammentario come quello di alcuni film di Godard o del cinema più sperimentale. Insomma si può concepire un'unità costituente la colonna sonora pur senza ragionare in termini di “inquadrature sonore”, poiché il suono di per sé non è “inquadrabile”. La questione meriterebbe di essere approfondita anche in relazione alla segmentazione cognitiva degli eventi narrativi, che noi sogliamo eseguire in base a indizi visivi (cfr. D'Aloia e Eugeni 2016): infatti come vedremo il sonoro può denotare uno spazio diverso da quello visibile.

Anderson (1996: 80) nota che una significativa differenza tra visione e ascolto consiste nel fatto che la prima è meno sensibile alla discontinuità rispetto alla seconda. Ciò potrebbe dipendere dal fatto che siamo abituati a interrompere la visione attraverso i battiti di ciglia e le saccadi (cfr. Smith e Henderson 2008; Smith 2015), che non hanno un corrispettivo uditivo. Questa potrebbe essere una ragione biologica della prassi di rendere il montaggio sonoro letteralmente inaudibile mentre il montaggio visivo è invisibile solo per traslato; d'altronde la consuetudine della continuità sonora potrebbe essere uno dei fattori di riuscita del montaggio in continuità (cfr. Cutting e Candan 2013: 28).

Cohen (2001: 253-254) ipotizza che durante la fruizione di un film si possa attuare un processo di *sordità disattentiva* (*inattentional deafness*), così come del suo corrispettivo visivo, la *cecità disattentiva* (*inattentional blindness*): si tratta della mancata percezione rispettivamente di un evento sonoro o di un elemento visivo. Alla cecità disattentiva è correlata la ben nota *cecità al cambiamento* (*change blindness*), ossia la mancata percezione di una modifica pur visibile in un'area. Simili fenomeni hanno ricevuto attenzione dalle scienze cognitive a partire dagli anni '90 (cfr. Rock et al. 1992; Rensink, O'Regan e Clark 1997; Simons e Levin 1997; Simons e Chabris 1999; Rensink 2002; Beck, Angelone e Levin 2004; Simons e Rensink 2005; Smith, Lamont e Henderson 2012; Smith 2015), e si è studiato il loro ruolo nella fruizione filmica (cfr. Levin e Simons 1997; Levin e Simons 2000; Smith 2005; Smith e Henderson 2008; Smith 2017: 109-110). Vi è anche una *sordità al cambiamento* (*change deafness*) (cfr. Asutay e Västfjäll 2014), meno trattata ma che sarebbe anch'essa interessante da studiare nel contesto della fruizione filmica.

### 5.7.2 – Il punto d'ascolto

Sono stati ipotizzati vari termini per analizzare l'ontologia e la fenomenologia del sonoro relativamente al suo dato spaziale, usando per solito delle opposizioni concettuali. Ciò vale sia per la firma spaziale sia per la spazializzazione. La più nota – poiché forse trattasi della prima – è quella tra *fedeltà* e *intelligibilità*: nel primo caso ci si riferisce alla resa di un evento sonoro fedele a come lo si sarebbe udito realmente, quindi con una evidente firma spaziale; nel secondo ci si riferisce alla comprensibilità di un oggetto percettivo, alludendo a una pulizia del suono tale da celare la firma spaziale. Dunque nella fedeltà si privilegia la *concretezza* spaziale di un suono, mentre l'intelligibilità tende all'*astrazione* spaziale. Lastra parla di “modello fedeltà” (il cui principio è quello della *duplicazione*) e di “modello telefono” (il cui principio è quello della *produzione*), in riferimento al fatto che nel telefono l'esigenza è quella di comprendere al meglio le parole di chi interloquisce – infatti si parla estremamente vicino alla capsula ricevente, riducendo il più possibile la distanza e quindi lo spazio percorso dal suono. Si noti che il termine fedeltà è stato usato anche in riferimento alla qualità tecnica dell'audio (frequenza, rapporto segnale/rumore), come nella locuzione “alta fedeltà” (*high fidelity*, o *hi-fi*), il che non implica alcun riferimento alla componente spaziale poiché fa riferimento solo alla definizione dell'audio (cfr. Rumsey 2001: par. 1.3.1). Altri hanno parlato di suono inteso come *evento* opposto al suono inteso come *struttura* ovvero come *linguaggio*, dunque una concretezza rispetto a un'astrazione. Similmente si è contrapposto la *riproduzione* o la *registrazione* alla *costruzione*, nonché la *trascrizione* alla *significazione*. Con un'ulteriore sfumatura si è proposto di distinguere un *ascolto rappresentato* (*represented hearing*) –

quindi la verosimiglianza di un punto d'ascolto – da un *suono rappresentato* (*represented sounds*) in uno spazio poco definito. Con l'introduzione della stereofonia si è posto un simile problema anche relativamente alla spazializzazione, come abbiamo visto.

Occorre tuttavia considerare due aspetti della questione. Da una parte c'è il rapporto con un evento pro-fonografico (dunque la sua registrazione e la sua resa), dall'altra c'è la costruzione di un evento puramente fonografico. In altri termini c'è anzitutto un *cosa*, ovvero un evento da rappresentare, uno spazio da denotare; poi c'è un *come*, ovvero il tipo di resa che si sceglie per un dato evento e lo spazio in cui si svolge. Un evento pro-fonografico può essere reso senza rispettare la fedeltà spaziale; allo stesso tempo un evento fonografico senza referente reale può essere reso in modo mimetico rispetto al suo ipotetico svolgimento reale. Nel cinema gli eventi sonori sono solitamente costruiti con operazioni di montaggio; il messaggio può rendere lo spazio in maniera realistica o meno. Rumsey categorizza le applicazioni di audio spaziale proprio in base alla resa, a prescindere dalla natura dell'evento rappresentato: le due categorie sono *creative/artistic illusion*, in cui rientrano buona parte della musica e il cinema, e *accurate 3D rendering* (inteso come spazio a 360° non necessariamente tridimensionale), in cui rientrano simulazioni e realtà virtuale (2001: par. 1.5); in ogni caso non è detto che una registrazione implichi un ambiente acustico naturale, poiché potrebbe trattarsi di una creazione artificiale ovvero di "*acoustic fiction*" (2001: par. 6.1.3).

Nella relazione audiovisiva può esserci coerenza o meno tra la prospettiva visiva e quella sonora, sia in termini di firma spaziale sia relativamente alla direzione. Inoltre la prospettiva uditiva può risultare anomala a prescindere da quella visiva basandosi su inferenze circa il mondo rappresentato. Insieme al punto di vista bisogna analizzare il *punto d'ascolto*, ossia il punto dello spazio da cui ci sembra di percepire il suono. Chion (2001: 90-93) conferisce a questo concetto due accezioni: quella *spaziale*, oggettiva, e quella *soggettiva*. Gaudreault e Jost (1990: 134-136) fanno un'ulteriore differenziazione grazie al concetto di *auricularizzazione* (il correlativo uditivo della ocularizzazione): essa è *zero* quando il punto d'ascolto è oggettivo ed è *interna* quando esso è soggettivo; quest'ultima si divide inoltre in *primaria* quando il punto d'ascolto coincide esattamente con quello del personaggio, mentre è *secondaria* quando ci rendiamo conto di ascoltare lo stesso suono del personaggio pur se non attraverso le sue orecchie. Questa distinzione è utile perché capita spesso di condividere l'ascolto di un evento sonoro con un personaggio pur senza condividere esattamente il suo punto d'ascolto. Di Donato offre un'ulteriore proposta teorica distinguendo tra due tipi di punto d'ascolto oggettivo: quello *onnisciente*, il più diffuso al cinema, in cui i suoni sono trattati privilegiando la funzione espressiva piuttosto che la mimesi, ed è quindi inverosimile (si tratta della "*god's ear*" *perspective* di Begault 1999: 5); quello *antropomorfico*, il quale è invece mimetico (Di Donato 2014a: 30-31). Il punto d'ascolto può essere soggettivo anche quando il punto

di vista è oggettivo (e viceversa): nell'esempio di *Parigi di Notte* (*Paris vu par*, Francia 1965) citato da Di Donato il punto d'ascolto è soggettivo (auricularizzazione interna primaria), perché ascoltiamo attraverso le orecchie del ragazzo, ma il punto di vista è oggettivo (ocularizzazione zero). In future ricerche sarebbe opportuno anche considerare la differenza tra una prospettiva egocentrica e una prospettiva allocentrica (Tversky e Hard 2009).

La coerenza prospettica nella distanza non è misurabile, perché non c'è modo di verificare con precisione il rapporto tra la distanza della fonte, le caratteristiche ambientali e la qualità sonora. Inoltre occorre considerare che l'utile scala dei piani sonori proposta da Maasø non prevede una corrispondenza biunivoca con la scala dei piani visivi. La differenza tra mezzo primo piano, primo piano e primissimo piano nello stimolo visivo è più esigua rispetto a quella tra i tre termini riferiti allo stimolo sonoro. Si può comunque approssimare la valutazione in un modo chiaro. Inoltre talvolta una parvenza di coerenza prospettica è suggerita da un cambio contestuale di prospettiva senza che una delle due circostanze sia propriamente coerente, per esempio quando a un campo lunghissimo visivo è associato un campo medio sonoro per poi passare a un campo ravvicinato con sonoro in primo piano.

I dispositivi elettronici di trasmissione di immagini e suoni sono un elemento critico nell'analisi spaziale visiva e sonora, come vedremo più avanti, ma ciò vale in parte anche per gli specchi. In una scena di *Le verità nascoste* (*What Lies Beneath*, USA 2000) un computer, visibile alla sinistra del quadro, si accende da solo, e il suono dell'accensione emana coerentemente dal canale sinistro; ma la sua immagine è il riflesso in uno specchio, perciò esso è in realtà situato a destra nello spazio. Nel messaggio di questo suono è stata favorita l'immagine rispetto allo spazio denotato, forse nel timore che il pubblico avvertisse come distraente la presenza del suono nel canale destro.

Alla luce dei dati sperimentali possiamo ipotizzare che la scarsa accuratezza che si riscontra di solito nella costruzione dello spazio sonoro filmico sia permessa dall'interazione crossmodale con lo stimolo visivo, che fornisce le principali informazioni spaziali. Rifacendoci alla teoria della cattura visiva possiamo dire che le immagini catturano i suoni; anche escludendo un fenomeno di cattura visiva possiamo asserire che le immagini ancorano i suoni a una rappresentazione spaziale visiva che è il veicolo principale di informazioni spaziali per il/la percipiente. Dato che lo stimolo visivo è preponderante nella percezione spaziale si può supporre che un cambio di prospettiva visiva lasci sullo sfondo la dinamica spaziale del sonoro, oppure che la plasmi.

### 5.7.3 – Campo e fuori campo

La terminologia che descrive la relazione spaziale tra un suono e la sua fonte rispetto al campo visivo è costituita dai tre termini *in*, *off* e *over*. Un suono è *in* se la sua fonte è visibile in campo; è *off* se la sua fonte non è visibile perché fuori campo ma comunque sita in uno spazio contiguo con quello visibile; è *over* se la sua fonte non è visibile perché estranea allo spazio (e spesso anche al tempo) della scena. Il suono mentale rientra nel caso *over*, ma può essere anche descritto come suono interno o interiore. Chion nota giustamente che ci sono certi tipi di suono i quali sfuggono a una classificazione del genere. Certi suoni d'ambiente (o “suoni-territorio”) non sono chiaramente *in* o *off*, perché avvolgono la scena «senza sollevare la questione ossessionante della localizzazione e della visualizzazione» (2001: 78; cfr. 2009: 467); d'altra parte la loro fonte non è sempre ben individuabile nello spazio, come nel caso del vento o della pioggia. Inoltre la collocazione dei suoni *on the air*, cioè mediati dalla tecnologia, è viziata dalla “schizofonia”, poiché la fonte è moltiplicata (cfr. 1990: 78-79; 2003: 482-483). La voce di una radiocronista è *off* se non vediamo la sua fonte – ossia la persona – ma se la ascoltiamo attraverso una radio potremmo dire che la fonte è la radio stessa. Perciò se la radio fosse inquadrata avremmo un suono in relativamente al concetto di “suono diffuso dalla radio” mentre sarebbe *off* relativamente al concetto di “suono prodotto dalla persona”.

Il suono *off* può essere attivo o passivo a seconda di quanto attiri l'attenzione su di sé e quindi su una certa area di spazio fuori campo. A questo proposito consideriamo la teoria di Burch: egli asserisce che le sei aree di fuori campo possono essere attivate dal suono *off*, ma escludendo la direzionalità del sonoro l'attivazione specifica di un'area avviene su base di inferenze determinate dall'osservazione dell'inquadratura e dalla cognizione del mondo filmico. Col sonoro direzionale è possibile attivare un'area specifica dello spazio *in* e di quello *off* proprio grazie alla propagazione direzionale del suono. Ipotizziamo di numerare le aree di Burch come in figura 21. I canali sinistro e destro possono attivare l'area relativa dello schermo, e approssimare rispettivamente la denotazione delle aree 1 e 3. Nel contesto dell'home theater si deve tener conto di alcune differenze, poiché i tre canali frontali non si propagano da dietro lo schermo bensì da zone prossime a esso (cfr. Martinelli e Corelli 2006: 5). L'area 2 può essere attivata coi nuovi sistemi immersivi, mentre non mi è noto a oggi un sistema che diffonda il suono da sotto lo schermo, pertanto l'area 4 è esclusa da un'attivazione direzionale. L'area 6 coincide in termini direzionali con quella visibile in campo, perciò la differenza tra un suono *in* e uno *off* proveniente dalla sesta area può essere determinata solo dalla firma spaziale e da indizi visivi. L'area 5 può essere attivata in modo direzionale laddove sia possibile utilizzare esclusivamente gli altoparlanti posteriori, ma l'uso dei canali surround del 5.1 può approssimare l'effetto. In un sistema 5.1 inoltre le aree 1 e 3 possono essere espanse nella

lunghezza grazie ai due canali surround, e in un sistema audio tridimensionale ciò vale anche per l'area 2.

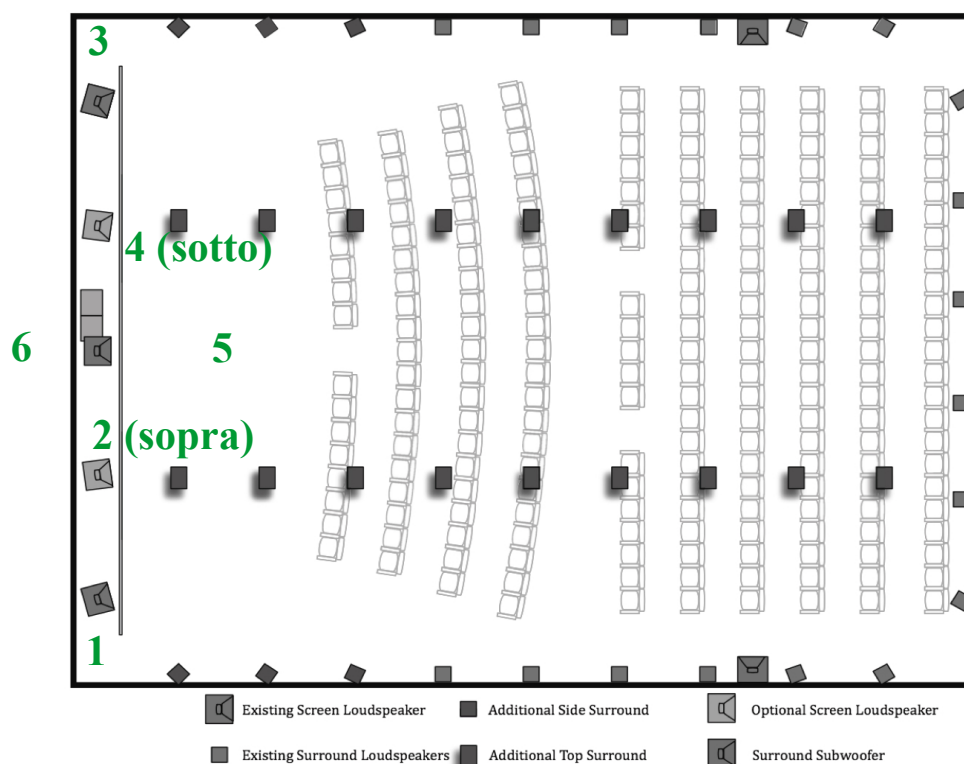


Figura 21: Le sei aree fuori campo (cfr. Burch 2000). Mia rielaborazione della figura dei Dolby Laboratories (2014: 13).

### 5.7.4 – Sintassi sonora: raccordi e ponti

Come notano Troscianko, Messe e Hinde, «The soundtrack usually “glues” the visually disparate shots, together forming a whole, which keeps the perceiver engaged» (2012: 416). Come abbiamo accennato il sonoro ha una capacità di creare continuità più spiccata rispetto a quella delle immagini, il che risulta evidente ragionando in termini di raccordi e di ponti. I raccordi ben noti nella teoria e nella prassi del film sono tutti visivi tranne uno, il raccordo sonoro. Con questo termine si fa riferimento alla capacità del suono di legare due inquadrature stabilendo una continuità spaziale tra le due. Occorre però anzitutto distinguere tra due diversi modi con cui un suono può mettere in continuità le inquadrature. Il primo si riscontra laddove un suono off è sincronico con la prima inquadratura e la seconda ne mostra la fonte o, pur tenendola fuori campo, ce ne palesa l'esistenza. Qui la continuità è data da un nesso causale sebbene il suono non sia sincronico anche con la seconda inquadratura. La nostra inferenza è che il suono che abbiamo udito un istante prima sia stato prodotto da qualcosa presente nello spazio che ci viene mostrato in seguito allo stacco di

montaggio. Il secondo tipo di raccordo sonoro si ha quando il suono è sincronico con ciascuna inquadratura. Qui insomma il suono funge da legante, il che è un caso comunissimo rappresentato emblematicamente dai suoni d'ambiente: essi forniscono uno sfondo acustico comune a quanto le inquadrature mostrano. Tuttavia proprio perché questi suoni sono percepiti come uno sfondo realistico tendono a non essere concepiti come materia di raccordo sonoro, il che accade più facilmente con dei suoni che costituiscono una figura. In ogni caso potremmo definire il primo tipo *raccordo sonoro interrotto*, conseguente a una *inferenza causale*, e il secondo *raccordo sonoro continuo*, conseguente alla capacità di *legante* propria del suono. Un raccordo di entrambi i tipi può comunque ingannare, ossia essere un *falso raccordo*.

È opportuno estendere il concetto di raccordo sonoro in affinità a quelli visivi, perché la messa in banda può seguire dei raccordi spaziali come le immagini, oppure può ometterli o ancora può deliberatamente infrangerli. Una simile operazione può risultare cavillosa qualora si volesse identificare un corrispettivo sonoro di ciascun raccordo visivo, per almeno due ragioni: una concerne la prassi produttiva, l'altra consiste nella differenza tra la natura dei due stimoli. La colonna sonora è solitamente costruita dopo il montaggio, il quale invece è condizionato dalla messa in scena e dalla messa in quadro, tanto che in fase di ripresa si tiene conto di come si intende in seguito montare le inquadrature. Perciò, tendenzialmente, i raccordi visivi vanno pensati già durante le riprese mentre quelli sonori scaturiscono dal montaggio del suono applicato a una colonna visiva già montata. Inoltre proprio perché non esiste un corrispettivo sonoro *stricto sensu* dell'inquadratura risulta spesso improprio ragionare sui rapporti tra suoni come se si trattasse di unità ben definite; per giunta, come abbiamo notato, per noi esseri umani il suono non ha delle coordinate spaziali rigorose come quelle visive. Occorre comunque considerare dei processi di cognizione spaziale che sono affini a quelli che scaturiscono dai raccordi, dunque dipendenti dalla relazione tra inquadrature piuttosto che esclusivamente dai rapporti verticali tra le colonne nel corso di una singola inquadratura. Anzitutto il raccordo di *posizione* e quello di *direzione* possono sussistere sul versante del sonoro (cfr. Di Donato 2006: 65-85): il primo prevede che il suono mantenga nel controcampo la posizione nello spazio stereofonico coerente con quella della fonte nello spazio filmico, mentre il secondo richiede che il suono venga spostato nello spazio stereofonico coerentemente con lo spostamento della fonte nello spazio filmico dopo lo stacco. Per semplicità concettuale si potrebbero unificare i due raccordi secondo il criterio della direzionalità: infatti, diversamente dalle immagini, il posizionamento di un suono nello spazio di fruizione è già espresso in termini di direzione, così come il suo eventuale spostamento. Tuttavia la distinzione tra posizione e direzione in riferimento ai raccordi chiarisce immediatamente se il suono presenta un movimento pluridirezionale o meno. Inoltre si può parlare di panoramica sonora (*panning*) nel caso

di qualunque spostamento direzionale di un suono, dato che solitamente non è possibile distinguere quando il movimento in questione è stato realizzato in fase di registrazione (movimento della fonte rispetto ai microfoni) o in fase di missaggio (il *panning* vero e proprio) (cfr. Holman 2008: 110-119).

Il raccordo di posizione sonora può essere infranto qualora la direzione sia diversa da quella che ci si aspetterebbe sulla base del realismo. Tuttavia la prassi di favorire l'area frontale e in particolare il canale centrale può comportare che alcune scelte di spazializzazione non siano percepite come un'infrazione del raccordo di posizione bensì come la sua omissione; la valutazione dipende dal contesto. In molti film può capitare che alcuni dialoghi, soprattutto off, siano direzionali nei canali frontali laterali, mentre la maggior parte di essi, tanto più se in, sia collocata nel canale centrale. Nell'ipotesi in cui si raccordino in base alla posizione dei suoni e non altri nel contesto della medesima scena, si produrrebbe l'impressione di un'infrazione. Insomma grazie alla stereofonia l'organizzazione dello spazio a 180° è possibile anche nel sonoro, ma tutt'oggi questa opzione viene applicata sporadicamente: si lascia per lo più che lo spazio visivo ancori quello sonoro il quale nelle sue informazioni principali (anzitutto il dialogo) resta centrato e senza gli stessi numerosi cambi di prospettiva (cfr. Martinelli e Corelli 2006). Ciò probabilmente è dovuto al timore dell'effetto ping-pong, presumibilmente distraente perché l'udito tollera meno la discontinuità rispetto alla vista. Guarda caso la direzionalità del suono si presenta più facilmente nel corso di una singola inquadratura piuttosto che di un campo/controcampo.

Ci può essere un correlativo sonoro anche del raccordo sull'asse, qualora si abbia il passaggio da un piano ravvicinato a uno distante – o viceversa – dello stesso suono collocato nel medesimo punto; così come con le immagini, questo ci fornisce l'impressione di esserci avvicinati o allontanati rispetto alla fonte. In questo caso si potrebbe parlare di raccordo *sulla distanza*.

L'idea di un corrispettivo sonoro del raccordo di sguardo è bislacca dal momento che non esiste un correlativo uditivo dello sguardo – come scrisse Marcel Duchamp, si può veder vedere ma non si può sentire udire. Eppure talvolta è possibile veder udire, qualora qualcuno tenda l'orecchio, oppure si può inferire da una reazione che qualcuno abbia udito qualcosa. In ogni caso ci possono essere varie relazioni tra un soggetto ascoltatore e una fonte sonora e più in generale tra diversi punti di ascolto, oggettivi o soggettivi, tanto da riportare la questione al di fuori di una tipologia di raccordi e piuttosto in una disamina basata sugli altri concetti esposti in questo capitolo.

Un altro modo di montare sequenze tra loro è quello di usare il *ponte sonoro* (*sound bridge*), che consiste in una momentanea asincronia tra sonoro e visivo che avviene nel contesto di una transizione tra sequenze o tra scene (cfr. Bordwell e Thompson 2003: 412-414). Esso può essere di due tipi: *a precedere*, quando ascoltiamo suoni della sequenza successiva già prima dello stacco

sulla prima inquadratura di questa; *a seguire*, quando continuiamo ad ascoltare suoni della sequenza precedente mentre le immagini ce ne mostrano una successiva. Nel caso dello *split edit* invece gli stacchi visivi e sonori alla nuova inquadratura di una stessa scena non coincidono temporalmente, e quindi si ha un *overlapping sound* – diverso dall'*overlapping dialogue*, che indica scene di dialogo in cui più battute si accavallano. Questo tipo di montaggio è particolarmente evidente proprio con il dialogo: la prosecuzione di una battuta dopo lo stacco visivo è chiamata *J-edit* (o *J-cut*), mentre quando lo stacco segue l'inizio del dialogo si parla di *L-edit* (o *L-cut*) (cfr. Salt 2015).

Il ponte sonoro di solito denota uno spazio diverso da quello visibile, poiché normalmente le due sequenze avvengono in spazi diversi e in tempi diversi. Tuttavia potrebbe anche trattarsi di una transizione a uno spazio diverso nello stesso tempo, oppure allo stesso spazio in un tempo diverso. L'aspetto temporale non pertiene a questa tesi, tuttavia è doveroso a questo proposito fare una precisazione, in conformità a quanto sostenuto da Smith (2009) circa il differimento spaziale. La coincidenza temporale di un evento sonoro con uno visivo, o anche con un evento non mostrato ma comunque appartenente al mondo rappresentato, è detta *sincronia* – da cui *sincrono* – mentre l'assenza della medesima è detta *asincronia* – da cui *asincrono*. La sincronia non richiede necessariamente il realismo: il suono può anche essere incongruente con l'immagine cui è associato; essa può inoltre avvenire tra eventi della storia (con o senza nesso causale o indexicale), tra elementi estranei alla storia e tra combinazioni dei due (p.es. un accordo del commento musicale può essere sincronico con un evento della storia o anche con uno stacco di montaggio). Il termine *sincronismo*, che in senso stretto è sinonimo di sincronia, ispirandosi a Ejzenštejn, Pudovkin e Aleksandrov ([1928] 1964) e a Pudovkin (1929) può essere usato in senso più astratto per indicare l'associazione costante a fini realistici dei suoni alla loro fonte, il che richiede ovviamente anche l'esatta coincidenza temporale. I film normalmente presentano sincronismo, più raramente asincronismo. In altri termini si può asserire che a livello della narrazione un suono storico può essere *simultaneo* o meno rispetto a quanto mostrano le immagini storiche; qualora non lo fosse potrebbe svolgersi in un tempo *anteriore* o *posteriore*. Durante il ponte sonoro quindi c'è asincronia, ovvero assenza di simultaneità: nel ponte a precedere il suono è posteriore, nel ponte a seguire esso è anteriore. Tuttavia questo accade solo nel contesto di una transizione usualmente alquanto breve, perciò non ne risulta tanto un'impressione di asincronismo, bensì la percezione di una scelta stilistica destinata solo a contribuire a una resa esteticamente valida un passaggio alla sequenza successiva.

### 5.7.5 – Spazio immersivo e spazio virtuale

La stereofonia è stata introdotta al cinema come un sistema a tre canali frontali e altoparlanti intorno al pubblico, e i canali frontali non sono stati mai meno di tre. La scelta di usare un canale centrale è dovuta al fatto che la spazializzazione migliora, e specificamente si migliora la resa delle immagini fantasma, soprattutto considerando la notevole distanza tra i due altoparlanti laterali in una sala cinematografica. Il segnale surround nasce dalla volontà di avvolgere il pubblico in un'esperienza che oltrepassa i confini dello schermo. Tuttavia l'area surround è stata tradizionalmente riservata a certi tipi di suoni, lasciando che le informazioni principali restassero ancorate all'area frontale in prossimità dello schermo, ossia quella che cattura l'attenzione del/della percipiente. Nella prima fase della stereofonia (1952-1976) era consueto spazializzare dialogo ed effetti nei canali frontali in coerenza con la posizione della fonte nello spazio rappresentato. Nei primi film stereofonici si riprendeva il suono col principio della fedeltà usando più microfoni, ma presto si tornò al microfono singolo per ragioni di praticità; eppure nel missaggio si seguiva ancora la fedeltà spaziale, soprattutto nella direzione. Con l'avvento del Dolby Stereo si è stabilita invece la prassi di centrare il dialogo, collocandolo solo al centro o anche ai canali laterali contemporaneamente. Il surround era inizialmente riservato alla musica di commento – ma solo in alcuni momenti – e ad alcuni effetti particolarmente significativi. Col passare del tempo si è fatto un uso più costante del surround, ma sempre riservandolo a musica, ambienti e alcuni *hard effects* – cioè rumori che denotano un singolo evento ben definito sul piano spazio-temporale, usualmente percepiti come figura. Il Dolby SVA ha permesso alla stereofonia di diventare normale, ma allo stesso tempo ha limitato le possibilità di spazializzazione. Anzitutto, trattandosi di un sistema matriciale, non era possibile una totale separazione tra tutti i canali; inoltre il sistema era retrocompatibile solo in parte, poiché la riproduzione di una simile colonna con un impianto mono – possibile – portava alla cancellazione del canale surround, il quale pertanto non poteva contenere informazioni fondamentali ma solo rafforzare l'effetto percettivo. Il Dolby 70mm non aveva questi limiti ma la prassi era quella di fare un unico missaggio a quattro canali, per esigenze di tempo e di costo. I sistemi digitali hanno gradualmente portato a un uso più ingente dei canali surround, anche con suoni direzionali (soprattutto effetti, più raramente dialoghi).

Fin dal periodo di diffusione della stereofonia, gli anni '50, si ragiona su modelli diversi di utilizzo della spazializzazione acustica, tanto nel cinema quanto nella discografia. La dicotomia più nota è quella costituita da *you are there* e *they are here*: nel primo caso ci si riferisce all'impressione di trovarsi nel luogo in cui si è svolto l'evento rappresentato, col principio della fedeltà; nel secondo si intende l'impressione che la fonte denotata sia presente nello spazio di ascolto, alludendo dunque

a un evento in uno spazio costruito appositamente per chi fruisce l'incisione (cfr. Bartlett 1983; Rumsey 2001: par. 1.3.1). La prima locuzione è stata dominante nell'alta fedeltà, come dimostra anche la pubblicità dell'epoca (cfr. p.es. Stephens 1954; International Electronics Corporation 1955; Munston 1955), sebbene spesso la stereofonia dell'epoca enfatizzasse la spazializzazione allontanandosi dalla fruizione in sala da concerto. Il principio *they are here* si presta maggiormente alla *popular music* in cui l'evento registrato è concepito appositamente per l'incisione, piuttosto che essere la versione incisa di un evento altrimenti fruibile dal vivo.

Al cinema la questione è più complessa per almeno due ordini di fattori: da un lato non ci sono solo suoni musicali, bensì un insieme di voci, rumori e musiche atto a denotare per solito degli eventi narrativi; dall'altro c'è l'interazione crossmodale attivata con le immagini. L'approccio tradizionale nel messaggio cinematografico comporta che i suoni vengono posti in una gerarchia per cui certe loro caratteristiche sono gestite in ossequio all'efficacia espressiva piuttosto che al realismo. Perciò i dialoghi sono usualmente in primo piano e frontali anche quando visivamente chi parla è più distante o di spalle, infrangendo la coerenza prospettica nella distanza e nella frontalità. Inoltre i dialoghi sono abitualmente collocati nel canale centrale a prescindere dalla posizione di chi parla, infrangendo la coerenza prospettica nella direzione. I canali surround sono costantemente utilizzati, assieme a quelli frontali, per i suoni d'ambiente (non associabili a un singolo evento, tendenzialmente percepiti come sfondo), mentre gli "hard effects" sono in prevalenza collocati nei canali frontali a prescindere dalla loro posizione nello spazio filmico, per evitare di attirare l'attenzione su di sé – ovvero in un'area distante da quella dello schermo<sup>50</sup>. Quando gli effetti vengono collocati nei canali surround (talora abbondantemente, p.es. in scene d'azione) la loro posizione è di solito prospetticamente coerente, ma comunque non come conseguenza di un approccio realistico bensì della ricerca di un risultato stilisticamente valido. La musica diegetica ha spesso una firma spaziale piuttosto pertinente, tuttavia per solito in termini direzionali è missata con la prassi discografica e resta statica a prescindere dalle immagini. Per un esempio lampante si consideri la sequenza della festa nel primo atto di *Eyes Wide Shut* (*Id.*, Regno Unito/USA 1999), in cui la musica suonata dall'ensemble mantiene la stessa spazializzazione a prescindere dalla prospettiva visiva continuamente cangiante. Come nota Høier, la pratica del surround del cinema mainstream statunitense (2000-2012) – a mio avviso tuttora dominante – «involves a healthy pragmatism [...] avoiding dysfunctional results – and presenting a 360-degree sound experience only when it is fruitful to do so» (Høier 2014: 253). Insomma l'immersione sonora offerta dal cinema è di norma atta a favorire principalmente l'impressione di prossimità allo spazio filmico e a

<sup>50</sup> La scelta di limitare i suoni nell'area surround è dipesa in passato anche da motivi pragmatici, come ho spiegato nel par. 1.2.3. Questo può aver contribuito a stabilire una norma persistente anche dopo il superamento di tali limiti tecnologici.

incrementare l'impatto emotivo o l'«*effetto bagno*» (Jullier 2006: 56), piuttosto che a denotare realisticamente uno spazio, continuo con quello visibile sullo schermo, in cui possiamo orientarci razionalmente.

Questa seconda opzione è ovviamente possibile, e costituisce un altro modello di spazializzazione acustica. In svariati momenti di *mother!* (*madre!*, USA 2017, 5.1) ascoltiamo da tutti i canali normali quanto accade nello spazio sonoro; il surround ci permette di localizzare i suoni off, anche i dialoghi, rafforzando la nostra capacità di *simularci presenti* (Gallese e Guerra 2015) come testimoni immateriali in uno spazio filmico a 360° come avviene in modo più o meno accentuato in *Toy Story 3 – La grande fuga* (*Toy Story 3*, USA 2010, Dolby Surround 7.1), *Salvo* (Italia/Francia 2013, 5.1), *Gravity* (*Id.*, Regno Unito/USA 2013, Dolby Atmos), *Maze Runner – Il labirinto* (*The Maze Runner*, USA 2014, Dolby Atmos e Auro 11.1), *Veloce come il vento* (Italia 2016, Dolby Surround 7.1). Questo modello di spazializzazione è consueto in altre forme audiovisive. Nei videogiochi First-person shooter (FPS) – p.es. *Call of Duty: Advanced Warfare* (2014) – e anche in alcuni giochi d'avventura in terza persona come *Tomb Raider* (2013) e *Resident Evil 2: Revelations* (2014) i suoni vengono spazializzati in conformità alla prospettiva visiva: essendo l'esperienza del gioco interattiva, ciò non incide solo sull'assorbimento ma anche sulla dinamica del gioco, poiché il sonoro offre a chi gioca indizi sulla collocazione di fonti che possono risultare decisivi per l'esito della partita (cfr. Collins 2008: cap. 4), tanto che Goodwin ritiene il surround cruciale nei videogiochi e solo «a nice but dispensible decoration» nel cinema (2009: 2). In questo caso la mimesi spaziale è una strategia espressiva mirata a incrementare l'efficacia dell'esperienza ludica aumentando le possibilità di interazione, o quantomeno l'impatto sensoriale (quest'ultimo vale anche per giochi di diverso genere; cfr. Dockwray e Collins 2015: 260-261). I video VR solitamente hanno audio a 360°, in conformità al tipo di esperienza sensoriale che intendono offrire, più mimetica che diegetica, diversamente dal cinema in cui vale il contrario.

Come abbiamo visto nel cap. 1 i nuovi sistemi di sonoro immersivo permettono di incrementare il realismo percettivo anzitutto ampliando l'area di provenienza dei suoni. La localizzazione è più accurata, poiché gli oggetti audio possono essere collocati in maniera puntiforme – ossia in un punto circoscritto, cioè in un singolo altoparlante tra quelli a disposizione nelle aree surround. Lo stesso vale per l'impressione di movimento, allorché l'oggetto non passa da un canale all'altro – ossia da un insieme di altoparlanti all'altro – bensì da un singolo altoparlante a un altro. Inoltre il Dolby Atmos impone che gli altoparlanti surround siano equivalenti ai tre frontali in termini di risposta in frequenza, acciocché le aree surround non abbiano una definizione inferiore. Infine, come abbiamo visto nel Dolby dichiara che l'Atmos abbandona il principio del *sweet spot*, ossia il punto ideale da cui ascoltare per godere appieno della stereofonia, a favore di un

approccio *room-centric* che permette di fruire opportunamente il suono da qualunque poltrona, mentre Barco sottolinea che l'AuroMax massimizza le dimensioni del *sweet spot*. Da ciò deriva presumibilmente un aumento del senso di immersione percettiva.

A fronte di questi vantaggi – alcuni indubbi, altri da verificare – c'è da considerare un potenziale limite del Dolby Atmos che concerne la possibilità di controllo per i fonici. Da un lato il sistema è molto accurato circa i parametri tecnici da rispettare e la collocazione dei suoni, dall'altro occorre considerare che il decodificatore interviene con algoritmi che adattano il mix alla determinata sala. Di conseguenza lo slogan “Author Once, Optimize Everywhere” è meno entusiasmante di quanto possa sembrare, perché l'ottimizzazione è un processo automatico del decodificatore che sfugge al controllo degli artisti (cfr. Wright 2015: 235-238).

Alla luce delle suddette considerazioni si può supporre che i fonici tendano a mantenere un approccio conservatore per ridurre il rischio che l'opera sia alterata durante la proiezione (cfr. Sergi 2013: 115), così non sfruttando al meglio le potenzialità spaziali offerte dalla tecnologia, come ho potuto riscontrare nei film che ho visto in Dolby Atmos<sup>51</sup>. Un illustre precedente di conservatorismo per motivi tecnici si riscontra, come già detto, nella testimonianza di Ben Burtt il quale, relativamente al messaggio di *Il ritorno dello Jedi* (*The Return of the Jedi*, USA 1983), dichiarò di concepire il surround come «an *enhancement* to what is already there [nei canali frontali]» (citato in Blake 1984: 45) per evitare il rischio di cancellazione di suoni che fossero destinati esclusivamente agli altoparlanti surround, spesso trascurati in sale mediocri. D'altra parte tale approccio è forse privilegiato a prescindere, per ragioni di consuetudine o motivazioni estetiche e commerciali: è improbabile che i cineasti in generale siano propensi a stravolgere le convenzioni stilistiche consolidate da decenni a cui essi e il pubblico sono abituati e di cui è garantita l'accoglienza positiva. Inoltre, come spiega Wright, i fonici statunitensi «participate in a freelance marketplace» e «Their creative work depends on a shared set of conventions», il che restringe il margine di creatività laddove fosse in contraddizione con dette convenzioni e con le aspettative dei committenti (2015: 243-245; cfr. Sergi 2013).

Nel riflettere sulla ricaduta di sistemi di suono immersivo sull'estetica filmica sorgono alcune domande. Essi comportano la nascita di un nuovo stile oppure si limitano a incrementare

<sup>51</sup> Si tratta di *Avengers: Age of Ultron* (*Id.*, USA 2015), *Star Wars – Il risveglio della forza* (*Star Wars: Episode VII – The Force Awakens*, USA 2015), *Independence Day – Rigenerazione* (*Independence Day: Resurgence*, USA 2016), *Inside Out* (*Id.*, USA 2015), nella sala iSens del cinema UCI Parco Leonardo di Fiumicino (Roma); *Alla ricerca di Dory* (*Finding Dory*, USA 2016), nella sala iSens del cinema UCI Roma Est; *Kong: Skull Island* (*Id.*, USA 2017), presso il cinema Arcadia (Melzo, MI); *La mummia* (*The Mummy*, USA 2017) e *Blade Runner 2049* (*Id.*, USA 2017), presso il cinema Andromeda (Roma). I due film Pixar qui nominati, *Inside Out* e *Alla ricerca di Dory*, sono una parziale eccezione perché c'è la spazializzazione frontale dei dialoghi come da abitudine della casa di produzione. Purtroppo ho visto tutti questi film nella versione doppiata italiana, ma poiché dalle testimonianze di addetti ai lavori mi risulta che attualmente il doppiaggio di film di major rispetta la spazializzazione originale ritengo che la mia osservazione resti valida.

l'efficacia di pratiche già consolidate? Aumentano il realismo? Sono distraenti? Simili domande sono sorte anche in passato con l'introduzione di nuove tecnologie come nel caso del Cinerama e del CinemaScope. Questi sistemi intendevano incrementare l'impatto percettivo di immagini e suoni, rendendo l'esperienza in sala cinematografica più intensa rispetto a quella televisiva; l'ascolto iniziò a essere immersivo, poiché i suoni circondavano il pubblico. Come spiega Belton (1992) si riscontra una costante tensione tra il polo del realismo, poiché l'avanzamento tecnologico rende i suoni sempre più vicini a quelli reali sul piano percettivo, e quello dello spettacolo, perché l'applicazione artistica di tali tecnologie tende a trascurare la rappresentazione realistica. Sebbene i canali surround del Cinerama fossero sfruttati in ossequio al realismo, grazie alla coerenza delle prospettive visiva e sonora, allo stesso tempo essi attiravano l'attenzione su di sé contraddicendo il tradizionale "realismo" hollywoodiano, fondato sull'invisibilità. L'approccio dei primi film in CinemaScope prevedeva la collocazione dei dialoghi nei tre canali frontali coerentemente con la posizione del parlante sullo schermo, con un effetto realistico che tuttavia rimase sporadico perché avvertito come distraente (cfr. Belton 1992: 158-163; Malsky 2015).

Per indagare ulteriormente la questione giova considerare l'*exit door* (o *exit sign*) *effect* (cfr. Kerins 2011: cap. 4; Elvemo 2013: 32), che Chion – con una sfumatura diversa – chiama *in-the-wings effect* ("effetto quinta" nella traduzione italiana; 2001: 85). Si presume che un suono circoscritto nell'area surround attiri su di sé l'attenzione distraendo il pubblico, che per un istante lo interpreta come un suono proveniente non dal film ma dal mondo reale, come se qualcuno avesse sbattuto la porta di uscita della sala. Di conseguenza i fonici tendono a evitare di collocare dei suoni nel surround senza una precisa intenzione espressiva, anche se da una prospettiva realistica tali suoni dovrebbero essere collocati proprio lì. Un discorso analogo vale per i dialoghi, che normalmente sono collocati nel canale centrale a prescindere dalla posizione del parlante nello spazio diegetico; si presume che il pubblico sarebbe distratto dal continuo movimento del dialogo tra gli altoparlanti che scaturirebbe qualora lo si collocasse in coerenza con la prospettiva visiva. Dato il loro enorme potenziale, i nuovi sistemi immersivi si pongono come una sfida all'*exit door effect*, poiché una quantità ingente di suoni può essere collocata e spostata "puntualmente" nella sala. Un esercente, riferendo di un'esperienza col Dolby Atmos, dice: «you'd occasionally think someone in the middle of the auditorium stood up and started talking. Before you realize that it is actually a character in the movie that is just a little bit off-screen. I found myself laughing, but that's where that character would actually be standing» (Rudyard Coltman in Fuchs 2013). Insomma detti sistemi possono favorire il realismo spaziale, incrementando l'impressione di presenza, ma rischiano di essere percepiti come maggiormente artificiali, non tanto perché offrono uno spettacolo irrealistico quanto perché attirano l'attenzione su di sé distraendoci dalla storia. Mutuando gli

efficaci termini di Grajeda si può asserire che la tensione è tra l'“occultamento del dispositivo”, per ridurre il più possibile l'impressione che gli stimoli sonori siano mediati così da garantire il realismo o l'assorbimento, e la “celebrazione del dispositivo”, in cui il lavoro del medesimo è mostrato palesemente per enfatizzare il godimento che l'artificio può apportare. Secondo Grajeda i primi prodotti stereofonici possono essere descritti come «audio of attractions», o «stereo as spectacle», poiché l'intento era quello di enfatizzare la nuova capacità tecnologica prescindendo dal realismo o dalla trasparenza (2015: 45-46, 59). Dunque poiché le convenzioni consolidate sono percepite come la norma – quindi *naturali* ancorché non realistiche – qualsiasi allontanamento da esse, in direzione di un maggiore realismo o dell'enfasi su una nuova spettacolarità, può essere percepito come un'aberrazione catturando su di sé l'attenzione e paradossalmente riducendo l'immersione. Inoltre sembra che «total photo- and audio-realism is not necessary for a virtual reality environment to produce in the viewer a sense of immersion» (McMahan 2003: 68). In forme come certi videogiochi d'avventura o i video VR la convenzione è invece quella di denotare uno spazio audiovisivo a 360°, in cui la collocazione di suoni circoscritti nell'area surround non solo non è distraente bensì è un principio fondamentale dell'esperienza offerta, non ponendo più il problema dell'*exit door effect* (Collins 2008: cap. 4).

Come abbiamo visto Kerins individua due modelli di costruzione dello spazio sonoro permessi dai sistemi multicanale: uno è quello che Chion chiama *supercampo*, riconducibile all'avvento del Dolby Stereo negli anni '70, in cui il surround è usato per un ampliamento approssimativo dello spazio rappresentato prevalentemente dalle immagini; l'altro è quello che Kerins chiama *ultracampo*, reso possibile dai sistemi digitali, in cui c'è un'accurata direzionalità degli eventi sonori. Ancora oggi il ricorso all'ultracampo avviene per lo più in momenti tipici della narrazione, come un'aberrazione rispetto alla norma e quindi qualcosa che enfatizza la spettacolarità o la rilevanza drammaturgica di una sequenza. I nuovi sistemi immersivi possono fornire una vera tridimensionalità sonora e quindi aumentare il realismo della spazializzazione (ossia un incremento di credibilità rispetto al surround tradizionale) (cfr. Lennox, Myatt e Vaughan 1999: 2), ma ciò non significa che i fonici siano costretti a farlo o lo desiderino e che il pubblico lo pretenda. Nei *white paper* del Dolby Atmos c'è un'enfasi sul realismo non rintracciabile nelle dichiarazioni e nella prassi dei fonici (cfr. Wright 2015: 228-229), come si nota da queste citazioni:

unprecedented realism [...]. [Y]ou'll get more precise, realistic sounds as an object [...] passes overhead (Dolby Laboratories 2015: 2, 7); recreate reality [...]. In real life, sounds move in specific and sometimes complicated ways [...]. And when you lose those details, it detracts

from the brain's sense that what it's watching is real. [...] [T]he most realistic and immersive entertainment experience ever available for home theaters (Dolby Laboratories 2016: 2, 15).

Il concetto di realismo è stato usato come strategia di marketing della Dolby già alla fine degli anni '70, non per un autentico interesse verso il realismo ma per la volontà di presentare il Dolby Stereo come qualcosa di diverso da ciò cui il nome Dolby era associato principalmente, ossia il sistema di riduzione del rumore per impianti hi-fi domestici. Allo stesso tempo, il concetto di realismo permetteva di enfatizzare come dice Grainge «the *use value*» che il Dolby Stereo aveva nel contesto cinematografico, che non si limitava a un aumento del volume (un luogo comune nei discorsi dell'epoca sul sonoro multicanale) ma che comportava nuove possibilità artistiche di cui il pubblico poteva godere coscientemente (2008: 254-256).

A tutt'oggi per i/le cineasti/e il realismo sembra essere secondario rispetto all'efficacia drammaturgica e all'intenzione espressiva (cfr. Sergi 1999). L'ingegnere Ron Lagerlof a proposito dei nuovi sistemi immersivi spiega: «The goal of all this technology is to envelop the audience and transport them into the story, so the viewer really shouldn't be thinking whether a sound is coming from an overhead loudspeaker or from the sidewall» (in Cinema Technology 2014: 85). Similmente, secondo gli ingegneri Cribbs e McCrigger «immersive sound allows audiences to be immersed in the sound field and, more importantly, into the story» (2014: 72). Simili posizioni sembrano essere rappresentate meglio da un'altra frase di un *white paper* del Dolby Atmos, già citata a proposito di *L'uomo d'acciaio*: «moving music offscreen frees up the screen loudspeakers for audio effects and can help clarify dialogue tracks» (Dolby Laboratories 2014: 10); un'opzione non realistica, bensì rispondente a un criterio di efficacia espressiva. Kerins ha insomma ragione quando sostiene che il suono surround nel cinema enfatizza la storia ed è guidato da essa, ossia asseconda le esigenze drammaturgiche piuttosto che il realismo o l'attrazione, garantendo così che il pubblico non si trovi di fronte ma in mezzo allo «literally *placed* in the dramatic space of the movie» (2011: 1550-1551). Similmente Sergi afferma che col Dolby Atmos «The feeling is that of being in a simulated world rather than before the representation of one, as *Atmos* blurs in many ways the border between artifice and simulation» (2013: 113). D'altra parte la spazializzazione acustica del film molto di rado presenta l'effettivo dato spaziale dei suoni in presa diretta: si tratta di un'operazione di missaggio, parte della messa in banda, gestibile liberamente dai fonici senza alcun vincolo di rispetto per un evento originale profilmico. È dunque utile sostituire l'opposizione “realismo vs irrealismo/spettacolo/artificio” col concetto di *resa spaziale*, condizionata da fattori oggettivi – la fedeltà acustica spaziale – e soprattutto soggettivi – la qualità percepita dell'audio spaziale – i quali includono l'effetto di *naturalzza*, correlato alla familiarità, alle convenzioni, alle esigenze

stilistico-narrative (in questa terminologia mi ispiro particolarmente a Chion 2001: 109; Rumsey 2002; Begault 2006). Circa la distrazione, si può dedurre che essa subentri non tanto per la ricerca del realismo o per la violazione di una norma intrinseca o estrinseca al film, quanto qualora una data scelta stilistica – tanto più se viola una norma – sia priva di una motivazione drammaturgica.

Il problema della spazializzazione acustica è molto affrontato da chi si occupa di ambienti virtuali, in relazione ai concetti di immersione e presenza. È noto che questi due termini molto ricorrenti sono usati con varie accezioni e talvolta come sinonimi, in relazione sia a tecnologie audiovisive sia in generale a mezzi artistici e comunicativi. In ogni caso la presenza è comunemente intesa come l'impressione di *essere lì* (*being there*), ovvero in uno spazio denotato dal mezzo piuttosto che in quello concreto in cui ci si trova davvero, insomma una *illusione percettiva di non-mediazione* (Lombard e Ditton 1997), mentre l'immersione viene intesa anzitutto in senso tecnologico ed eventualmente in senso percettivo o psicologico. Una delle prime e più importanti sistemazioni della questione è quella offerta da Lombard e Ditton (1997) i quali individuano sei accezioni di presenza: (1) ricchezza sociale; (2) realismo; (3) trasporto; (4) immersione; (5) attore/attrice sociale in un medium; (6) medium come attore sociale. In questa disamina i significati più rilevanti sono (2), (3) e (4). Circa il realismo, gli autori osservano che bisogna distinguere tra un *realismo sociale* – la verosimiglianza – e un *realismo percettivo*. Una distinzione per certi versi simile è quella di Slater tra *place illusion* – ossia la presenza, che pertiene al “come” – e *plausibility illusion* – ossia «the illusion that what is apparently happening is really happening» (2009: 3553). Il trasporto può avere tre forme: *you are there, it is here* (cioè *they are here*) e *we are together* (ovvero «condividiamo lo stesso spazio»). Infine l'immersione può essere intesa con due significati: quello *percettivo*, cioè quanto un ambiente virtuale sommerge il sistema percettivo dell'utente, il che è oggettivo; quello *psicologico*, cioè quanto l'utente si sente presente nell'ambiente virtuale ed eventualmente coinvolto/a o assorbito/a, il che è soggettivo (cfr. anche Sheridan 1996; Regenbrecht, Schubert e Friedmann 1998; Freeman e Lessiter 2001; McMahan 2003; Slater 2003, 2009; Västfjäll 2003; Gooskens 2014; Loomis 2016; Kim, Jeon e Kim 2017; Wang, Petrina e Feng 2017). In effetti nell'ambito della tecnologia audio *immersive sound* fa riferimento non a uno stile ma a una potenzialità, tuttavia in questo caso implicando la terza dimensione – quella verticale – che non è un requisito necessario nel resto della letteratura per usare il termine “immersivo”, quindi è bene curarsi di evitare la confusione.

Nilsson, Nordahl e Serafin (2016) propongono un'eccellente sistemazione delle definizioni di immersione, individuandovi quattro accezioni: (1) proprietà di un sistema; (2) risposta percettiva; (3) risposta a racconti; (4) risposta a sfide. L'accezione (1) si può indicare con la locuzione *System Immersion* (cfr. anche Slater 1999; Nordahl e Nilsson 2014: 218). Il significato (2) si riferisce

all'impressione di presenza in uno spazio immersivo, dunque la *spatial presence* (Larsson, Västfjäll e Kleiner 2003: 3). Poiché questa è strettamente correlata alla capacità immersiva del sistema, gli autori semplificano il quadro elaborando il modello tridimensionale di immersione in figura 22.

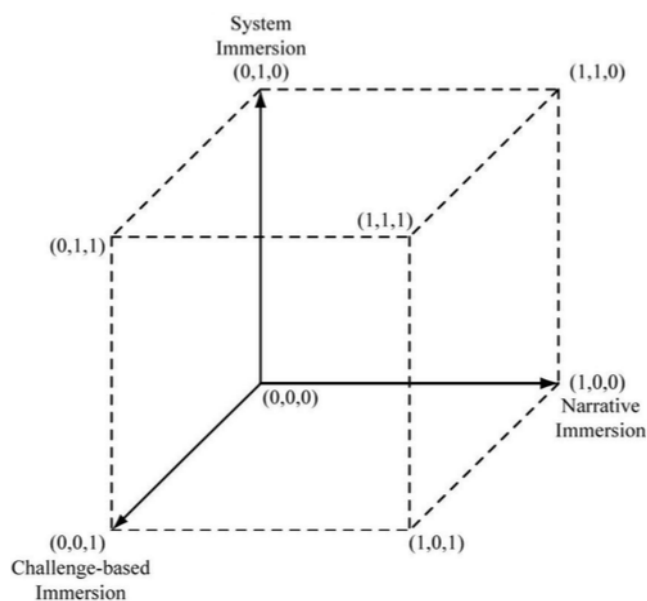


Figura 22: il modello tridimensionale di immersione di Nilsson, Nordahl e Serafin (2016: 118).

I valori 0 e 1 sono gli estremi di una scala, ove 0 rappresenta l'assenza di immersione e 1 il suo massimo livello possibile.

Si riscontrano anche studi sperimentali in merito alla questione che possono permetterci di approfondire la conoscenza dei meccanismi cognitivi. Il sonoro, soprattutto se è coerente col visivo, migliora l'effetto percettivo degli ambienti immersivi (tanto che Nordahl e Nilsson lo giudicano necessario più che utile; 2014: 228) benché il rapporto tra presenza e realismo sia complesso (cfr. Hendrix e Barfield 1996; Larsson et al. 2007), a conferma della problematicità del concetto di realismo il quale non si basa solo su un dato percettivo. Freeman e Lessiter (2001) dimostrano che le basse frequenze di un canale LFE aumentano la presenza, a conferma della rilevanza della componente vibrotattile dello stimolo sonoro, che è un'ulteriore sollecitazione sensoriale del corpo del/della percipiente nel cinema (cfr. Kerins 2011: cap. 4; Takahashi et al. 1999; Schust 2004; Takahashi 2011; Whittington 2014: 183, n. 2). Västfjäll (2003) rileva che l'ascolto a sei canali surround risulta essere quello che permette la maggiore presenza e di realismo emotivo, ma spiega che il nesso causale tra i due fatti non è chiaro. Lipscomb e Kerins (2004) hanno ottenuto tre risultati particolarmente interessanti. Confrontando l'impatto di brani musicali e sequenze filmiche in due modalità audio – due canali stereo e 5.1 – essi dimostrano che (a) in entrambi i casi le reazioni sono state differenti in base alle categorie valutative proposte (p.es. immersivo, artificiale,

naturale, intimo), (b) nella musica il surround è meno significativo, (c) l'educazione visiva è più significativa di quella musicale al fine di apprezzare l'apporto del surround. Ciò può far supporre che la costruzione di uno spazio – tipica del cinema piuttosto che della musica – renda il surround più motivato, soprattutto in relazione al genere di film, e che anche quando la visione non è coinvolta l'interazione crossmodale con le capacità visuospatiali è cruciale poiché migliorano la cognizione spaziale più di quanto non faccia l'ascolto (cfr. anche Kerins, Lipscomb e Rendahl, s.d.). Il miglioramento della resa della profondità permessa dal surround è stato dimostrato da Cullen et al. (2013) associando un sonoro 5.1 a immagini stereoscopiche.

L'esperimento di Ekman e Kajastila (2009) usa i videogiochi ma offre dati validi anche per il cinema. Essi sostengono che indizi uditivi di localizzazione influiscono sulla risposta emotiva: i suoni espansi e quelli propagati da dietro sono più spaventosi di quelli rispettivamente puntiformi e frontali. Ciò è coerente con altri dati scientifici. Asutay e Västfjäll (2015a, 2015b) dimostrano che esiste un bias verso l'area posteriore a noi relativamente sia all'emozione sia all'attenzione, due livelli correlati che rendono più salienti biologicamente i suoni provenienti da fuori del campo visivo, come se il sistema uditivo avesse una funzione di allarme: in questo modo siamo alquanto pronti a voltarci per verificare visivamente la natura della fonte soprattutto quando è percepita come negativa, il che aumenta la nostra sicurezza (cfr. anche Juslin e Västfjäll 2008; Tajadura-Jiménez et al. 2010c). L'esperimento di Kobayashi, Ueno e Ise (2015) incrocia dati psicologici e fisiologici, concludendo che un ambiente sonoro virtuale spazializzato in 3D incrementa la presenza e l'attivazione del sistema nervoso simpatico, similmente a quanto accadrebbe con un'intrusione reale nello spazio personale. Tutto questo aiuta a comprendere le ragioni della spazializzazione aggressiva anche nel surround usuale nel cinema horror (cfr. il mio studio di caso su *La casa* [*Evil Dead*, USA 2013], par. 6.12). In ogni caso diversi studi hanno posto il problema della preferenza da parte di chi fruisce nei confronti di una configurazione piuttosto che di un'altra. Guastavino e Katz (2004) riscontrano che l'audio 2D è giudicato come più naturale e realistico di quello 3D, il che potrebbe confermare il rischio di veicolare troppe informazioni. Allo stesso tempo però essi affermano che ciò dipende dal tipo di applicazione e dal contesto in cui la tecnologia viene utilizzata. I partecipanti all'esperimento di Song, Ellermeier e Hald (2011) sull'ascolto musicale hanno preferito un'ampio stereo con poco surround, ma comunque la valutazione cambiava in base al genere musicale.

Questi dati sono particolarmente significativi per la comprensione della cognizione dello spazio sonoro filmico. Possiamo supporre che il sonoro è per sua natura più immersivo delle immagini, anche quando non è "immersivo" nel senso di una concreta tridimensionalità. Cionondimeno occorre distinguere l'immersione nell'accezione sensoriale o percettiva e

l'immersione psicologica, come quella cui alludono Gallese e Guerra parlando di «relazione immersiva coi personaggi della storia» (2015: 78; cfr. Wojciehowski e Gallese 2011). A mio avviso è più utile usare il termine *assorbimento* per denotare la seconda circostanza, per praticità terminologica, come per giunta fanno Cohen, MacMillan e Drew (2006). Inoltre possiamo ritenere che la problematicità dell'*exit door effect* non sia dovuta esclusivamente all'abitudine bensì alla nostra biologia. Un suono direzionale nel surround cattura l'attenzione esogena in un modo peculiare rispetto a un suono frontale per motivi intrinseci alla funzionalità del sistema uditivo. Gallese e Guerra hanno dedicato giustamente molta attenzione alla cognizione del montaggio in continuità a 180° (2015: 192-207), il quale come si è detto non è accompagnato di solito da una coerente organizzazione dello spazio sonoro. In future ricerche neuro-cognitive sarebbe interessante studiare la relazione crossmodale di un sonoro raccordato nella posizione e nella direzione in conformità con le inquadrature, per verificare quanto è distraente e quanto invece contribuisce proficuamente alla costruzione dello spazio.

Di solito una sequenza filmica denota un singolo spazio con immagini e suoni. Tuttavia il sonoro potrebbe denotare anche uno spazio e un tempo diversi. In *La fiamma del peccato* (*Double Indemnity*, USA 1944) le parole di Walter che parla al dittafono avvengono in uno spazio e un tempo posteriori a quelli denotati nel flashback. Ancora, ci potrebbe essere una denotazione visiva di più spazi contemporaneamente, come accade con lo *split screen* (nonché di tempi differenti). Nelle famose inquadrature in *split screen* di film stereofonici diretti da Brian De Palma non sempre si usa lo stesso stile di spazializzazione: in *Il fantasma del palcoscenico* (*Phantom of the Paradise*, USA 1974) e *Blow Out* (*Id.*, USA 1981) le voci provengono coerentemente da una direzione diversa a seconda dell'area dello schermo in cui è mostrata la fonte; in *Passion* (Francia/Germania/Spagna/Regno Unito 2012) essi provengono dal centro a prescindere, mentre in *Femme Fatale* (*Id.*, Francia 2002) si presentano entrambe le opzioni.

La resa spaziale può presentare una varietà di scelte stilistiche. Anzitutto alcuni tipi di firma spaziale sono convenzionali. Un classico esempio è quello della riverberazione che evoca suoni solo mentali oppure grandi distanze a prescindere dalla conformazione dello spazio. In *Psycho* (*Psycho*, USA 1960) c'è un esempio di ambedue i casi: le voci che Marion ascolta nella sua mente sono riverberate; la voce di Sam che chiama Arbogast udito da Norman mentre questi è presso lo stagno produce un'inverosimile eco che nondimeno rafforza l'idea del lungo tragitto compiuto dalla voce. Circa la direzionalità, in generale i suoni del mondo denotato possono essere spazializzati in maniera coerente o incoerente con la prospettiva visiva o la conformazione spaziale. Eppure nello specifico questa spazializzazione può avere un grado più o meno alto di precisione o approssimazione (cfr. il concetto di *cartoonification* di Lennox, Myatt e Vaughan 1999: 8; Lennox,

Vaughan e Myatt 2001b: 4-6), e in alcuni casi essa può essere puramente connotativa. Prendiamo ad esempio alcuni momenti di *The Bourne Supremacy* (*Id.*, USA 2004). Spesso i canali surround presentano effetti d'ambiente che non sono tanto incoerenti con un dato punto d'ascolto quanto approssimativi: l'intento è quello di dare un'idea dello spazio e un effetto di immersione in esso senza per questo denotarlo scrupolosamente. Qualche volta invece il rumore di veicoli nel surround è coerente ed è ricordato in base alla posizione, dunque spazializzato in modo preciso. Le voci udite dal protagonista nella propria mente, fin dalla prima sequenza, si muovono tra i cinque canali senza per questo denotare alcun movimento in un certo spazio, ma solo per suggerire la confusione mentale del personaggio: in questo caso la scelta di spazializzazione è puramente connotativa, benché sia riferita a suoni della storia (sebbene mentali e non concreti). Qualcosa di affine avviene in *Il colore del crimine* (*Freedomland*, USA 2006) in cui le voci dei ricetrasmittitori sono diffuse nei cinque canali suggerendo il caos in corso nel mondo narrativo piuttosto che un certo punto d'ascolto.

## 5.8 – Spazio della storia e spazio del discorso

I termini più diffusi per distinguere i suoni interni al mondo narrativo rispetto a quelli esterni a esso sono *diegetico* (*diegetic*) ed *extradiegetico* (*nondiegetic*). Essi conseguono a una teoria diegetica della narrazione (cfr. Bordwell 1985a) in cui cioè si ritiene che il mondo rappresentato nel film sia la diegesi, e che qualcosa può essere interno o esterno a essa; ciò che è esterno costituisce una scelta narrativa che non denota un evento della storia narrata. Bordwell e Thompson (2003) classificano giustamente la questione come un aspetto spaziale del suono: infatti l'operazione cognitiva è quella di chiedersi «da dove viene questo suono?». Siamo infatti consapevoli che un elemento presente nel film non proviene necessariamente dal mondo narrativo. Ciò vale anche per le immagini, come dimostrano in maniera palese i titoli di testa anche quando sono sovrapposti su immagini diegetiche: sappiamo benissimo che quelle scritte sono estranee al mondo della storia. Nel caso del sonoro la questione è decisamente più rilevante, probabilmente per due motivi. Il primo – estrinseco alla cognizione uditiva – è che la musica di commento è una convenzione radicata e diffusissima, rappresentando una comunissima presenza extradiegetica nel cinema. Il secondo potrebbe essere invece intrinseco alla cognizione uditiva, la quale è più dipendente dalla cognizione visiva di quanto sia viceversa: questo potrebbe portarci a dubitare più facilmente della reale collocazione di un suono acusmatico, laddove la visione della sua fonte lo colloca con maggiore forza nello spazio in questione.

In un mio articolo (Sbravatti 2016) mi sono basato sulla teoria di Bandirali (2006) il quale, seguendo Chatman (1980), distingue la storia dal discorso e quindi parla di suono storico e suono discorsivo piuttosto che di suono diegetico ed extradiegetico. Bandirali arriva a questa proposta dopo aver argomentato l'utilità di distinguere la denotazione dalla connotazione, ossia la funzione di un suono a prescindere dalla sua collocazione. Né Bandirali né io avevamo considerato la teoria di Bordwell (1985a), cosa che invece qui intendo argomentare.

La distinzione diegetico/extradiegetico è valida ma limitata, perché col suo riferimento esclusivo alla collocazione trascurando la funzione rende problematica l'analisi di alcune circostanze. Risulta ben più utile ragionare insieme nei termini di collocazione e funzione, oppure di denotazione e connotazione, o ancora di contenuto ed espressione. In questo senso la teoria di Chatman risulta particolarmente utile. Nel suo importante libro del 1978 egli ha aggiornato i concetti strutturalisti di storia e discorso, ove la prima è il *cosa* di una narrazione e il secondo è il *come*. In altri termini la storia è il *contenuto* e il discorso è l'*espressione*; della storia fanno parte gli *esistenti* e gli *eventi*. Di conseguenza la storia e il discorso definiscono in un certo senso uno spazio, perché ci sono enti e fatti che sono dentro la storia e altri che le sono esterni; allo stesso tempo però il discorso designa come detta storia viene narrata, in riferimento alle scelte narrative. Perciò la *discorsività* si dà sia come una collocazione – ciò che è presente solo nel discorso – sia soprattutto come una funzione – come il discorso lavora. Perciò possiamo asserire che alcuni suoni sono storici (diegetici) e altri sono discorsivi *tout court* (extradiegetici), ma allo stesso tempo possiamo analizzare il trattamento discorsivo dei suoni storici. Nella lingua inglese c'è una comodità in più: si può parlare di *discourse-sound* intendendo un suono (o più specificamente una voce, un rumore, una musica) collocato solo nel discorso, mentre si può dire *discursive sound* per specificarne la funzione. Se un suono è nel discorso è per definizione discorsivo, ma se un suono è storico non significa che non abbia alcuna valenza espressiva. Bordwell (1985a) rifiuta la distinzione storia/discorso perché la associa alle teorie comunicative in cui si intende la storia come un messaggio prestabilito da veicolare attraverso il discorso. La sua teoria invece asserisce che la storia è ricostruita mentalmente da chi fruisce sulla base di indizi offerti dall'intreccio e plasmati dallo stile, il che avviene con un processo di inferenze svincolate dal linguaggio, tranne ovviamente per quanto concerne il linguaggio usato in un film (testo scritto, dialoghi in una lingua nota). L'autore inoltre precisa che il concetto di discorso non chiarisce la narrazione cinematografica perché include l'intreccio e lo stile, i quali devono essere distinti. A mio avviso si può usare la coppia storia/discorso in conformità alla teoria di Bordwell, laddove si intenda la storia esattamente come la fabula e il discorso come l'interazione tra intreccio e stile, senza implicare un quadro

comunicativo linguistico. In questo modo si mantiene la pertinente distinzione di Bordwell ma allo stesso tempo si può semplificare la terminologia di analisi, come intendo dimostrare.

Ci sono due problemi principali nelle teorie che affrontano la collocazione dei suoni. Il primo concerne la visibilità della fonte, ed è estrinseco alla distinzione diegetico/extradiegetico; il secondo è invece a essa intrinseco perché, come accennato, pertiene al valore espressivo dei suoni. Nella teoria si asserisce che il suono in campo è sempre diegetico, ovvero che solo il suono fuori campo va distinto tra diegetico (off o over) ed extradiegetico (over). In realtà non è così, perché anche elementi visibili possono essere extradiegetici, come nel caso già accennato dei titoli. Il fatto che il suono extradiegetico in non è contemplato può dipendere da due ragioni. Quella più eclatante è che si tratta di un caso raro, mentre la musica di commento over è comunissima. L'altra ragione ha probabilmente una causa percettiva: la visibilità della fonte è una conferma della tangibilità, della concretezza del suono nel mondo denotato, a causa della nostra inclinazione a conferire maggiore importanza allo stimolo visivo piuttosto che a quello sonoro relativamente alla presenza, allo status di realtà. In un film il suono invisibile è segregato dallo spazio visivo, che normalmente coincide con lo spazio storico, tanto che può essere più facilmente interpretato come un suono escluso dallo spazio storico, mentre la sua visibilità nel medesimo spazio storico confermerebbe la sua presenza concreta in quel mondo.

In questa disamina risulta efficace un paragone con l'opera lirica. Qui l'accompagnamento orchestrale al canto è extradiegetico, e l'orchestra è segregata rispetto allo spazio della rappresentazione. Nel caso di Wagner la buca orchestrale è stata addirittura coperta, per occultare gli strumenti così da rafforzare l'assorbimento in un'esperienza d'arte totale, come se appunto la visibilità dell'orchestra riportasse inopportuna alla realtà. Quando invece vediamo uno strumento musicale nella scena siamo automaticamente portati a inferire che si tratti di una presenza storica, proprio perché è visibile nello spazio della rappresentazione e non è da esso segregata come l'orchestra. Infatti quando Figaro canta nella storia di *Il barbiere di Siviglia* (1816, Rossini/Sterbini) c'è una chitarra in scena con lui.

Un suono e un'immagine possono essere puramente discorsivi anche quando si tratta di inserti audiovisivi estranei alla storia, o di entità collocate nello spazio storico pur non facendone parte, o ancora di esistenti nella storia che compiono azioni non interpretabili come eventi storici. Si consideri un esempio per ciascuna di queste circostanze. La prima sequenza di *L'uomo che sapeva troppo* (*The Man Who Knew Too Much*, USA 1956) presenta un'orchestra in campo intenta a suonare. La musica è dunque in, ma non è diegetica dal momento che nulla lascia intendere che questo evento sia parte della storia. Si tratta evidentemente di un'ouverture, di un modo per introdurci il racconto nel quale in effetti la medesima orchestra svolgerà un ruolo rilevante,

specificamente in relazione a un colpo di piatti – strumenti i quali occupano buona parte dell'inquadratura, alla fine della sequenza, mentre una didascalia ci fornisce alcune informazioni. In *Tutti pazzi per Mary* (*There's Something About Mary*, USA 1998) il cantastorie e il suo collega agiscono nello spazio storico ma non nel tempo storico, non trattandosi di esistenti nel mondo narrativo. Sono dunque una presenza extradiegetica, ma i suoni che emettono sono in allorché i due musicisti sono in campo. Il fatto che uno dei due venga colpito alla fine del film da un personaggio della storia cui segue la fuga dell'altro è solo una gag, irrelata al racconto. Infine, in alcuni film il protagonista si rivolge al pubblico parlando, come in *The Wolf of Wall Street* (*Id.*, USA 2013): in simili momenti il parlante è diegetico in quanto esistente nella storia, ma quanto dice e fa è extradiegetico nel senso che non si tratta di un evento della storia bensì di un artificio espressivo che ricorda il monologo interiore o gli "a parte" teatrali.

Un secondo problema che emerge negli studi sulla collocazione del suono, specialmente della musica, è come accennato nell'accavallamento di due aspetti: la sua *collocazione* – dentro o fuori dalla diegesi – e la sua *funzione* – denotare un evento della storia oppure commentare. Il nostro processo cognitivo è anzitutto il seguente: se un suono denota è presente nella storia, se un suono non denota è assente dalla storia. Nel caso della musica siamo così abituati a concepire l'esistenza di un brano inserito nel film al solo scopo di commentare che inferiamo automaticamente che si tratta di una presenza esterna alla storia. Tuttavia sebbene il suono che non denota serva a connotare, ciò non significa che un suono che denota non possa anche connotare. Da qui scaturisce una carenza analitica, poiché si ritiene sovente che la musica diegetica sia presente solo per denotare un evento mentre quella extradiegetica serve ovviamente ad aggiungere significati o a enfatizzarli, a incrementare l'emozione, ad arricchire la forma, ecc. Ne consegue la difficoltà di descrivere alcuni casi in cui la musica diegetica sembra commentare così da mettere presumibilmente in crisi l'interpretazione della sua collocazione.

Il suono può essere manipolato in una grande varietà di modi, e come spiega Cecchi (2010) è importante indagare la funzione dei suoni senza sterilirsi sulla distinzione tra i due tipi di collocazione, che comunque resta un problema rilevante non solo per chi fruisce ma anche per i cineasti. Questo principio risulta particolarmente utile in quei casi in cui la collocazione non è minimamente discutibile ma in cui c'è comunque una differenza tra come noi percepiamo i suoni diegetici e come essi sarebbero in teoria percepibili nella diegesi stessa. L'esempio più eclatante viene da quei film in cui, per così dire, gli attori parlano una lingua diversa da quella dei personaggi. In *Vincitori e vinti* (*Judgement at Nuremberg*, USA 1961), i tedeschi parlano tedesco solo per qualche istante; in seguito li sentiamo parlare inglese come gli anglofoni. Sappiamo tuttavia che ciò non sta accadendo nella diegesi, poiché si tratta di un artificio stilistico atto a evitare

di dover ricorrere ai sottotitoli, o di far recitare gli interpreti in una lingua non madre, o di scegliere interpreti diversi da quelli desiderati. Perciò le voci in questione non sono propriamente diegetiche, nel senso che non prendono vita nella diegesi nel modo in cui le ascoltiamo, ma non sono extradiegetiche poiché denotano un dialogo diegetico. Una simile situazione accade anche in *Ratatouille* (*Id.*, USA 2007), come nota Collins (2013): di solito sentiamo i topi parlare inglese, come gli esseri umani con cui convivono, ma sappiamo che nella storia non è così, infatti qui topi e persone non sono in grado di parlare gli uni con le altre; i topi parlano piuttosto una loro lingua peculiare, dal suono di squittio, che noi udiamo solo una volta – dal punto d’ascolto dell’anziana – per alcuni momenti nel primo atto del film.

Anche in questo caso il paragone con l’opera lirica è fruttuoso. Il fatto che i personaggi cantino non ci lascia inferire che essi stanno cantando nella storia, bensì che loro si esprimono parlando ma che questo evento ci viene presentato in forma di canto accompagnato, così come il fatto che essi parlano in italiano malgrado siano in un paese straniero – p.es. la Cina di *Turandot* (1926, Puccini/Adami/Simoni) – è interpretato come un artificio discorsivo. La distinzione tra recitar cantando e cantar recitando, e più tardi tra recitativo e aria, si riferisce a due modi espressivi diversi, ma anche il primo dei due non deve essere interpretato come la realtà denotata, ancorché sia maggiormente naturale del canto più impostato. Insomma in un simile contesto il canto non è diegetico, sebbene il contenuto del medesimo lo sia, trattandosi di comunicazione o della concretizzazione di pensieri che costituiscono eventi storici. In alcuni casi invece c’è bisogno di denotare il canto nella storia, il che non può avvenire semplicemente con l’espedito di far cantare i personaggi, dato che i/le cantanti li interpretano normalmente ricorrendo al canto. Si suole quindi usare degli stratagemmi, tra cui come accennato quello di collocare strumenti sulla scena.

Il concetto di discorso può includere le considerazioni di Bordwell (1985a) sul grado di *self-consciousness* della narrazione. Infatti alcuni eventi sembrano denotati in una maniera talmente fedele che pare non scaturire da una narrazione, mentre altri sono estremamente artificiosi. La narrazione oscilla quindi tra i due estremi di *covert* e *overt*. Ciò che cambia non è l’evento denotato ma il modo in cui detto evento è denotato: in altri termini, cambia il come e non il cosa, ancorché ovviamente il come influenzi inevitabilmente il cosa. Si tratta insomma del grado di discorsività, ossia di quanto l’evento storico risulti plasmato dal discorso – dall’estremo dell’irrelevanza a quello dell’ostentazione.

Smith, debitore di Bordwell, ha proposto una distinzione tra tre tipi di musica – dunque di suono – ovvero *nondiegetic*, *source* e *diegetic*: nei primi due casi il suono è tale *stricto sensu*, cioè è uno stimolo acustico, è udibile, è concreto; nel terzo caso esso è solo inferito, dunque è immaginario, astratto, pertanto non è di nostro interesse in questo contesto ancorché lo sia

indubbiamente sul piano narratologico. Tuttavia trovo due problemi in questo modello. Il primo, marginale, è che l'espressione «diegetic music» rischia di indurre in equivoco, dato che è usata diversamente dalla prassi. La musica implicata è effettivamente diegetica, poiché è parte della fabula, ma essendo inaudibile da noi non è musica sul piano percettivo, mentre la musica diegetica comunemente intesa è musica udibile. Il secondo problema è più rilevante perché concerne la teoria: Smith asserisce che gli altri due tipi di musica sono inclusi nell'intreccio, in un caso come indizio di un evento della fabula, nell'altro come presenza esclusa dalla fabula. Tuttavia questo contraddice il concetto di intreccio per come è inteso da Bordwell. Esso infatti è un concetto astratto che scaturisce da come gli eventi e gli esistenti sono elaborati cognitivamente da chi fruisce. Perciò l'intreccio non contiene alcun elemento che è estraneo alla storia, ma fa riferimento a come gli eventi sono organizzati temporalmente. È lo stile che include degli elementi che possono essere estranei alla storia, ossia extradiegetici, come per esempio i movimenti di macchina “primari”. Ovviamente lo stile influenza anche gli elementi diegetici, per esempio i costumi o l'illuminazione. Lo *score* – la musica di commento – è una scelta puramente stilistica, esclusa dall'intreccio e quindi dalla fabula, mentre la *source music* è un indizio dell'intreccio che denota un evento della fabula; inoltre su essa incide lo stile, ossia la scelta di un brano piuttosto che di un altro e le scelte di montaggio sonoro e di missaggio.

Diversi/e autori/trici hanno fornito validi spunti di riflessione sulla questione, che è mia intenzione integrare e riformulare in un impianto cognitivista. Kassabian (2001) ha il merito di essersi avvalsa di termini di Earle Hagen, ossia di un artista piuttosto che di uno studioso, parlando non solo – com'è consueto – di *score* e *source music* ma anche di *source scoring*, in riferimento alla musica storica che commenta. Kassabian inoltre manifesta disagio verso l'opposizione diegetico/extradiegetico preferendo concettualizzare un continuum tra i due estremi in cui la musica è variamente collocabile. La tesi è condivisibile, e può essere rielaborata con la coppia storia/discorso. Infatti un suono è udibile se è incluso nel discorso, ma la sua collocazione può essere esclusivamente discorsiva oppure anche storica. Tuttavia non esiste un suono esclusivamente storico dal momento che è udibile, perché qualcuno lo ha manipolato, per quanto discretamente. A seconda di quanto tale manipolazione sia palese si va appunto dal suono diegetico “ovvio” (*source music*, nel caso della musica) al suono totalmente extradiegetico (come lo *score*), passando per una scala intermedia (nella quale si può parlare in diversa misura di *source scoring*, in riferimento alla musica). Il concetto di *fantastical gap* proposto da Stilwell (2007) è in taluni casi particolarmente pertinente, per esempio nel musical. Tuttavia in generale è un concetto superfluo, poiché la questione si può risolvere più semplicemente in riferimento al grado di espressività di un elemento storico, senza il bisogno di rilevare l'esistenza di un terzo luogo. La musica metadiegetica descritta

da Gorbman (1987), ossia il livello mediato secondo Miceli (2009), possono essere utilmente descritti come dei casi stimolanti di collocazione sospesa tra la storia e il puro discorso. Quando la collocazione è senza dubbio nella mente di un personaggio si tratta di suono storico interiore, che non è suono in senso stretto poiché è un evento puramente mentale ma che ci viene presentato in forma di suono concreto. Non c'è bisogno quindi di parlare di un terzo "livello", benché come al solito la circostanza possa essere complicata da altri interventi discorsivi. In *Un amore splendido* (*An Affair to Remember*, USA 1958) Nickie torna a casa della nonna dove aveva passato un giorno intenso con l'amata Terry. Durante la scena si ode la stessa musica che la nonna aveva suonato al pianoforte in quella occasione quando Terry si era unita a lei cantando. Il contesto suggerisce che la musica emani dalla mente di Nickie, sia perché egli è assorbito nel ricordare l'evento sia perché il suono del pianoforte ha una firma spaziale realistica. Si tratta quindi di una musica storica interiore, che denota l'attività mentale di Nickie, ma ovviamente essa fornisce una connotazione, trattandosi di un brano languido e romantico – il che vale anche per la prima scena in cui lo si ascolta, eseguito concretamente. Per giunta oltre al pianoforte ascoltiamo la melodia al violino, il che è presumibilmente un intervento esclusivo del discorso, atto a rafforzare l'effetto percettivo. Qualora invece la collocazione sia ambigua è doveroso notarlo, asserendo quindi che il suono che il discorso ci presenta ha un appiglio storico, potendo trattarsi dell'emanazione dei pensieri di un personaggio.

Pur sapendo che un certo suono non è frutto di un evento del mondo narrativo esso è pur sempre uno stimolo sonoro che può evocare delle qualità spaziali o un movimento, suscitando una risposta psicofisiologica affine a quella che si avrebbe con suoni storici. Un suono extra-storico può evocare un esistente o un evento storico denotandolo grazie a qualche sorta di somiglianza. Cano e Cremonini parlano di *simbolismo spazio-temporale* della musica da cui dipendono – nella dimensione spaziale – gli effetti di alto e basso o di salita e discesa; inoltre termini quali "volume" e "crescendo" alludono alla dimensione spaziale (1995: 132-133). A questo proposito si consideri l'analisi dei primi cartoni animati della Warner da parte di Curtis, il quale spiega che si dovrebbe abbandonare la coppia diegetico/extradiegetico parlando piuttosto di uso *isomorfo* e *iconico* del suono: nel primo caso «the sound and image have the same "shape"», cioè la loro relazione è governata dal ritmo e dal movimento; nel secondo la relazione audiovisiva «is not one of fidelity, but of analogy» tra qualità visive e «the timbre, volume, pitch, and tone of accompanying sound». Queste forme del suono si possono presentare anche con la tipica relazione indexicale del sincronismo cinematografico (1992: 200-203). Formulando meglio questa interessante tesi potremmo ribadire che esistono delle corrispondenze crossmodali tra suono e immagini, cosicché un suono può essere isomorfo se ha la stessa "forma" di un'immagine – sulla base di altezza, intensità, timbro, ritmo, ecc. – e può essere iconico se presenta un'analogia con il suono

palesamente diverso che esso intende evocare. Perciò, prendendo come esempio *The Wacky Wabbit* (USA 1942), la musica che accompagna in rigorosa sincronia i movimenti dei personaggi è un caso di isomorfismo, mentre l'uso onomatopeico del colpo di tamburo e di piatti che evoca una caduta è un caso di iconicità. Da parte sua Cohen spiega che la struttura della musica di commento può interagire con la struttura delle immagini contribuendo alla cognizione: una melodia ascendente o discendente attiva una risposta neuronale che evoca l'ascesa o la discesa, quindi può rafforzare la cognizione incarnata di un movimento ascendente o discendente visibile, come succede in *Shining* (*The Shining*, USA/Regno Unito 1980) quando Danny sfiora la lama del coltello (Cohen 2013: 188). Gorbman (1987: 25), Chion (2009: 409) e Lipscomb e Tolchinsky (2005: 393) affermano che una musica di commento può veicolare qualche informazione sullo spazio rappresentato pur non facendone parte. Green osserva che, diversamente da quanto asserito nella tradizione, «music is inherently spatial as it is temporal», poiché occorre distinguere il concetto di spazio da quello di «motion» (2011: 62). Infatti Green rileva in alcuni score hollywoodiani classici di Erich Wolfgang Korngold e Max Steiner delle caratteristiche che aiutano a veicolare, segnatamente grazie al contrasto e ai poli *thickness* e *thinness*, delle informazioni spaziali relative a tutte e tre le dimensioni: la distanza (o profondità, *depth*), grazie a effetti di eco nella struttura musicale, alla qualità timbrica e alla riverberazione; l'altezza («the vertical: what is high and low, above and beneath»), grazie al fattore acustico che in italiano si designa proprio con lo stesso termine; la larghezza, che scaturisce da fenomeni di concentrazione ed espansione (2011: 61-66). Per esempio in una celebre sequenza di *Lawrence d'Arabia* (*Lawrence of Arabia*, Regno Unito/USA 1964) – la cui musica, originale, è di Maurice Jarre – il crescendo all'accordo fortissimo di tutta l'orchestra che accompagna il primo viaggio del protagonista nel deserto evoca la vastità dell'ambiente, oltre a stimolare una certa risposta emotiva. Un quartetto d'archi che avesse suonato la medesima musica avrebbe sortito un effetto ben più moderato. Ci può essere una considerazione culturale, poiché sappiamo che la grande orchestra ha bisogno di uno spazio ampio mentre il quartetto è una formazione da camera, ma a prescindere da questo la massa sonora sprigionata dall'orchestra è percepita come più “grande”, sia per la maggiore intensità, sia per la varietà timbrica, sia eventualmente per certi caratteri della firma spaziale. Quest'ultima infatti se è asciutta o con scarsa profondità evoca spazi piccoli – poiché non offre chiari indizi spaziali – mentre se è bagnata e profonda evoca spazi grandi – perché si ha una maggiore sensazione del viaggio delle onde sonore nello spazio. Il punto non è la coerenza tra la firma spaziale esclusiva del discorso e quella – tangibile o inferita – dello spazio storico, bensì come le caratteristiche del sonoro stimolano determinate risposte cognitive. La firma spaziale dell'orchestra di *Lawrence d'Arabia*, con la sua profondità e il suo senso di distanza, è percepita come traccia di uno spazio ampio, quindi di vastità,

la quale è propria anche del deserto visibile sullo schermo, creando una corrispondenza crossmodale; non ci si pone il problema del fatto che essa riverbera come non potrebbe accadere se l'orchestra fosse nel deserto. In *It Follows* (*Id.*, USA 2013) la musica di commento talvolta panoramica in tutti i canali: non si tratta di denotare un evento della storia ma il movimento fornisce comunque la percezione di un ente che si sposta in ogni direzione, perciò alludendo percettivamente alla minaccia invisibile che perseguita i personaggi insorgendo da direzioni imprevedibili.

Ci può insomma essere un appiglio storico laddove si sfrutti qualche elemento storico per motivare determinate scelte sonore. Per esempio in *Complesso di colpa* (*Obsession*, USA 1976) quando il protagonista entra in chiesa si ascolta un brano per organo, il quale sembra collocato nella storia sia per la firma spaziale sia per il tipo di strumento. Tuttavia il brano, senza soluzione di continuità, si trasforma in un tipico *score*, e all'organo si aggiunge l'orchestra. In *Non aprite quella porta* (*The Texas Chain Saw Massacre*, USA 1974) la musica di commento è chiaramente una presenza solo discorsiva, ma essendo "rumoristica" non si può fare a meno di notare come essa evochi gli esistenti della storia: gli oggetti profilmici forniscono un appiglio alle scelte timbriche, dando l'impressione che per realizzare la musica siano stati utilizzati gli oggetti stessi.

Secondo Terrone «il senso di un suono dipende da tre componenti», che intendo chiamare *componenti interpretative*: «la causa materiale, il significato veicolato e l'effetto percettivo. Nel rumore prevale la causa materiale, nella parola prevale il significato veicolato, nella musica prevale l'effetto percettivo» (2004: 27). Piuttosto che di parola preferisco parlare di voce, dato che nell'uso comune parola è intesa anche come vocabolo scritto, mentre la voce è anzitutto suono fonico. Questa ha un forte potere significante, dato che normalmente – sia nella comunicazione quotidiana sia nel cinema – essa è usata per parlare, ossia appunto per produrre parole, quindi come strumento linguistico. Ecco perché in essa prevale il significato veicolato, e si suole parlare più spesso di dialogo, il che pone enfasi sul contenuto semantico veicolato dalla voce a scapito dei suoi aspetti fonici. I rumori sono in genere destinati a scopi realistici, ovvero al fine di denotare in senso indexicale la presenza di un esistente o un evento nella storia, ecco perché in essi prevale la causa materiale. La musica, quantomeno secondo la tradizione occidentale post-illuministica (cfr. Kassabian 2001), è di per sé asemantica e priva di referente, perciò in essa prevale l'effetto percettivo, ovvero la connotazione. La divisione dei suoni filmici nelle tre categorie di voce, rumore e musica è certamente utile, e risponde in parte alla prassi che da sempre si riscontra nel cinema sonoro tradizionale, in parte all'attitudine umana a realizzare questa distinzione anche nella vita quotidiana. Tuttavia occorre riflettere sui limiti di queste categorie, e sugli sconfinamenti dell'una nelle altre, ovvero sulle «forme ibride» come scrive Terrone:

Si ha una *musica-parola* (il leit-motiv, la canzone) e un *rumore-parola* (sirene, suonerie, applausi) quando si accentua il significato veicolato da musica e rumore. Si ha una *parola-musica* (p. es. le cantilene di Renée Zellweger) e un *rumore-musica* (i suoni iper-realistici) quando si accentuano gli effetti percettivi prodotti da parola e rumore. Si ha, infine, una *parola-rumore* (p. es. in *Stalker* [Id., URSS 1979, n.d.a.]) e una *musica-rumore* (i sospiri di Jane Birkin, il sassofono di Benny Golson in *The Terminal* [Id., USA 2004, n.d.a.]) quando si accentuano le cause materiali della parola e della musica» (2004: 27).

I suoni appartenenti a una delle tre categorie possono insomma subire un *trattamento differenziale*, cioè assumere una qualità che li riconduce per certi versi a un'altra categoria. Consideriamo altri casi. Quando la voce è usata per il *walla*, cioè il vociare indistinto in scene affollate, essa è più vicina al rumore perché diventa un suono ambientale, privo di significato linguistico. Il caso della musica che diventa voce è facilmente individuabile qualora nella partitura dei pezzi inseriti siano presenti appunto delle voci: per esempio in *Shining* il coro di *Jutrznia* di Krzysztof Penderecki sembra costituire una manifestazione sonora degli spettri che abitano l'albergo. Il confine tra musica e rumore può essere molto sottile, sia perché i rumori d'ambiente possono essere usati in senso musicale, sia perché la musica del discorso può incorporare dei rumori. In altre scene di *Shining* la musica contiene dei suoni elettronici originali che, soprattutto in un primo momento, potrebbero sembrare più un'esasperazione del battito cardiaco dei personaggi che non dei suoni discorsivi. Nella prima sequenza di *Shutter Island* (Id., USA 2010) è inserita parte del III movimento della *Sinfonia n. 3* di Penderecki: poiché essa si svolge su una nave, la nota grave ribattuta nel pezzo evoca con forza il suono di una sirena navale, quasi fosse un'onomatopea musicale (è dunque un uso iconico del suono). Nel primo episodio di *Kwaidan* (*Kaidan*, Giappone 1964) la musica di Tōru Takemitsu si confonde con i rumori di scena provocati dal personaggio mentre fugge dal fantasma (è dunque un uso isomorfo del suono). *Gli uccelli* (*The Birds*, USA 1963) è un caso eccellente di film senza *score*, dove tuttavia i rumori – specificamente i versi e il battiti d'ali degli uccelli – sono trattati per certi versi con modalità simili a quelle della musica, anzitutto assumendo sovente una funzione fortemente discorsiva: in effetti è stato proprio un compositore, Bernard Herrmann, a fare da consulente alla realizzazione elettronica di questi effetti. Una strategia simile caratterizza l'uso dei suoni elettronici dei dispositivi tecnologici in film quali *Il pianeta proibito* (*Forbidden Planet*, USA 1956) e *Abbandonati nello spazio* (*Marooned*, USA 1969).

<b><i>SUONO</i></b>	<b>Collocazione</b>	<b>Funzione</b>
<b><i>Astratto</i></b>	<p>Storia (esclusivamente); storico inaudibile; fuori colonna.</p> <p>Fa parte della storia, ovvero appartiene al <i>contenuto</i>; è escluso dall'espressione, dunque dal sonoro.</p>	<p>Denotazione indiretta: esistente o evento non denotato direttamente dal sonoro, puramente inferenziale.</p>
<b><i>Concreto</i></b>	<p>Storia (tramite Discorso): udibile, in colonna.</p> <p>Fa parte della storia, ovvero appartiene al <i>contenuto</i>; è incluso nell'espressione, dunque nel sonoro.</p> <p>Discorso (esclusivamente): non fa parte della storia, ovvero appartiene solamente all'<i>espressione</i>.</p>	<p>Denotazione indexicale: evento denotato dal sonoro.</p> <p>Connotazione: essendo mediato dal discorso può essere trattato in maniera discorsiva, in modo più o meno evidente.</p> <p>Connotazione: "commenta", aggiunge una sfumatura di senso, contribuisce all'effetto percettivo.</p> <p>Denotazione (di tipo simbolico e/o iconico): referente storico denotato dal sonoro in maniera non indexicale.</p>

Tabella 4: la collocazione dei suoni rispetto alla storia.

Tabella 5: riassunto.

<i>Indizi audiospaziali</i>				
Unidirezionali (monaurali)	Firma spaziale (intensità, suono diretto e riflesso, timbro)	Distanza	CLL CL CM MPP PP PPP	
		Frontalità		
		Profondità		
	Movimento	Implicito Esplicito		
		4 combinazioni tra fonte e punto d'ascolto		
Pluridirezionali (monaurali e binaurali)	Direzione (ITD e ILD, timbro)	Orizzontale Verticale		
		Movimento	Panoramica	
Punto d'ascolto	Oggettivo Soggettivo			
	Verosimile ("fedeltà") Inverosimile ("intelligibilità")			
<i>Indizi spaziali audiovisivi</i>				
Collocazione	Rispetto all'immagine	In Off Over	6 aree	
	Rispetto al racconto	Storia e discorso		
Punto d'ascolto e punto di vista	Coerenza Incoerenza			
Raccordi	Distanza Posizione Direzione			
Funzione di legante				
Ponte	A precedere A seguire			
Corrispondenze crossmodali				
<i>Indizi temporali audiovisivi</i>				
Rispetto all'immagine	Suono sincrono Suono asincrono			
Rispetto alla storia	Suono simultaneo Suono non simultaneo	Anteriore Posteriore		

## 6 – Studi di caso

Ogni film è stato fruito nella sua interezza usando un televisore e un impianto audio. In seguito ho rivisto e riascoltato alcuni momenti selezionati al computer, ricorrendo all'ascolto in cuffia; nel caso dei film stereofonici ho estratto i singoli canali per visualizzarne la forma d'onda e ascoltarli separatamente. I nomi degli/delle artisti/e sono ricavati dai titoli di testa o di coda dei film e integrati con i dati reperibili su [www.imdb.com](http://www.imdb.com).

### 6.1 – *All'ovest niente di nuovo (All Quiet on the Western Front, USA 1930)*

Prodotto da Carl Laemmle, Jr.

Distribuzione: Universal.

Diretto da Lewis Milestone.

Sincronizzazione e *score*: David Broekman.

Supervisore alla registrazione sonora: C. Roy Hunter.

Colonna sonora: mono (Western Electric System).

Copia per l'analisi: telecinema da positivo britannico non restaurato, reperibile in <https://www.viki.com/videos/219512v-all-quiet-on-the-western-front>.

Il film è un esempio dello stile sonoro del periodo di transizione. Si avvertono i limiti dell'assenza di messaggio, allora impossibile, ciononostante le sequenze di combattimenti colpiscono per la ricchezza della scena sonora, con numerosi rumori intensi dovuti alla guerra (esplosioni, spari, sibili). Nel complesso la costruzione dello spazio sonoro risulta oggi grezza, perché siamo abituati a ben altre raffinatezze. Si ha l'impressione di frammentazione, di un'approssimazione del punto d'ascolto maggiore rispetto a quella che si sarebbe affermata di lì a qualche anno, con il compimento della transizione al sonoro. Si può notare come alcuni stacchi sonori siano evidenti, repentini, nient'affatto inaudibili, risultando dal mero accostamento di diverse prese dirette, oppure dall'interruzione di una ripresa per inserire un effetto (p.es., i sibili prima delle esplosioni, oppure il fischietto alla fine della sequenza degli stivali e il motore dell'aereo che verso la fine colpisce Kat): in questi ultimi casi la scena sonora sembra interrompersi per lasciare spazio all'effetto in questione. Quando si ha l'impressione di continuità della ripresa sonora malgrado il montaggio analitico si può indurre che siano state usate più macchine da presa contemporaneamente (come nel

caso di alcune inquadrature nella caserma durante l'insediamento delle reclute). In altri casi invece il sonoro montato con un'inquadratura non ne costituisce la presa diretta: qui capita dunque che l'inquadratura risulti "muta" mentre ascoltiamo senza soluzione di continuità quanto avviene nello spazio adiacente (come nel caso dei primi piani dei ragazzi urlanti nell'aula scolastica).

Nella maggior parte delle scene la prospettiva sonora è coerente con quella visiva, denotando che la collocazione del microfono è prossima a quella della macchina da presa. Ciò favorisce l'approccio realistico, in nome della fedeltà piuttosto che dell'intelligibilità. Ne è un esempio l'inquadratura di Himmelstoss che tratta sprezzante i soldati nella caserma. Tuttavia qui talora la sua voce perde il fuoco pur quando l'attore non dà le spalle alla macchina, denotando che il microfono non segue la sua bocca. Ciò comporta una parziale perdita di realismo, che però evidentemente non è della stessa natura di quella che si ha col modello dell'intelligibilità, ove lo scopo principale è quello di mantenere sempre in PP e a fuoco la voce. Un chiaro esempio di coerenza prospettica viene anche dalla scena in cui i commilitoni parlano della voglia di tornare a casa dalle mogli (in particolare si nota un raccordo sulla distanza in corrispondenza del raccordo sull'asse che inquadra il soldato presso la porta) e nella scena con Paul e Kat in cui questi viene colpito a morte (la sua voce è fuori fuoco quando lui è di spalle).

Nella sequenza presso la scuola nel primo atto del film si ha un effetto curioso: le prime frasi del docente sono incomprensibili perché la colonna è occupata dall'intensa musica della banda che passa fuori dall'edificio. Sembra che il punto d'ascolto sia esterno all'aula mentre il punto di vista è interno. Quando la banda si allontana la voce del docente diventa comprensibile, in un CL coerente con quello visivo. Al raccordo sull'asse che mostra l'uomo in mezza figura corrisponde un raccordo sonoro sulla distanza che porta la voce in MPP. Nel controcampo sui ragazzi la voce del docente è in CM, per poi tornare in MPP allo stacco su di lui. Durante l'inquadratura di un ragazzo, cui seguono le sue immagini mentali, si ascolta di nuovo la voce del docente in CM, denotando un punto d'ascolto soggettivo del ragazzo, il che accade subito dopo anche con un altro giovane. In altri controcampi invece la firma spaziale della voce del docente resta immutata. La frammentazione del sonoro è evidente nelle inquadrature che seguono le urla dei ragazzi: mentre la colonna mantiene la continuità durante il serrato montaggio di diversi volti, in seguito si avvertono distintamente quattro stacchi sonori corrispondenti a quelli visivi (a eccezione del primissimo piano del docente, inserito sulla coda della presa diretta dell'inquadratura precedente).

Una situazione analoga è rilevabile nella sequenza in cui i soldati incontrano le donne francesi. Quando l'uomo che per primo le nota urla «Women!» la firma spaziale della sua voce è già cambiata, denotando che il microfono è stato spostato verso le donne che pochi istanti dopo entrano in campo a seguito di una panoramica verso sinistra. L'inquadratura successiva degli uomini che

gioiscono risulta “muta” perché non c’è suono sincrono con quanto mostrato bensì la prosecuzione del sonoro proprio dello spazio in controcampo, che mostra le donne che cantano. Il sonoro degli uomini che si rivolgono alle donne è in MPP sia quando sono inquadrati loro sia nel controcampo delle donne denotando la permanenza del medesimo punto d’ascolto più vicino agli uni che alle altre (a prescindere dal cambio di punto di vista); invece nell’inquadratura in cui le donne parlano le loro voci sono distanziate come lo erano quelle dei soldati ma entrano nella scena sonora anche quelle off degli uomini, in CL – coerentemente con la prospettiva. Inoltre allo stacco sui soldati mentre le donne si allontanano si può ascoltare un taglio brusco nella colonna sonora, il che avviene anche allo stacco successivo sulla carrellata laterale, in cui i punti di vista e di ascolto sono coerentemente lontani.

Un caso interessante di incoerenza prospettica, con un approccio tipico del modello dell’intelligibilità che diverrà dominante, si ha nella sequenza presso l’ospedale. Durante la carrellata tra i letti verso l’ingresso del salone ascoltiamo in MPP il dialogo tra l’uomo e la suora a prescindere dal punto di vista. Risulta chiaro che questo punto d’ascolto è determinato dalla posizione del microfono in prossimità del letto dove avviene il dialogo, infatti la voce della suora è in CM prima che ella si avvicini al letto.

La musica di armonica a bocca che si ascolta nell’ultima sequenza è l’unica presenza musicale extra-storica (escludendo quella che apre il film, la quale però non commenta immagini storiche). Essa viene interrotta dallo sparo, che è l’ultimo suono del film. Quanto segue è un finale dalla valenza simbolica in cui il silenzio rafforza la denotazione di uno spazio irreali: infatti, sovrimpressa all’immagine di un cimitero dall’alto, c’è l’inquadratura di soldati che marciano di spalle e che si voltano uno per uno, come per guardare un’ultima volta la vita prima di incontrare la morte.

## **6.2 – *La finestra sul cortile (Rear Window, USA 1954)***

Prodotto e diretto da Alfred Hitchcock.

Distribuzione originale: Paramount.

Suono: John Cope, Harry Lindgren.

Montaggio sonoro: Howard Beals (non accreditato).

Missaggio: Loren L. Ryder (non accreditato).

Musica: Franz Waxman.

Colonna sonora: mono (Western Electric System).

Copie per l'analisi: DVD Universal Studios (USA 2005), audio Dolby Digital 2.0 *dual mono*; BD Universal Studios (Italia 2013), audio DTS-HD Master Audio 2.0 *dual mono*.

Il sonoro di questo film è un valido esempio sia di una resa classica dello spazio sonoro sia di un approccio peculiare al medesimo. I suoni nell'appartamento di Jeff sono trattati con naturalezza: le voci sono normalmente frontali e in PP, con qualche oscillazione verso il MPP e più raramente il CM o il PPP, e quelle provenienti dal telefono sono udibili per lo più in soggettiva; i rumori rispettano approssimativamente la prospettiva visiva. Come tutte le inquadrature sul cortile e sulle finestre sono da un punto di vista collocabile nell'appartamento di Jeff (salvo alcune eccezioni), così tutti i suoni provenienti dall'esterno sono da un punto d'ascolto ivi collocabile. Perciò il sonoro lavora con le immagini per farci simulare presenti nell'appartamento e condividere la posizione dei personaggi che in esso si trovano. La scena sonora è costituita dai suoni potenzialmente udibili nell'appartamento e nel cortile, provenienti dalle varie finestre e dai dintorni della città; di volta in volta possiamo individuare diversi punti d'ascolto, pressoché sempre ad auricularizzazione interna, a seconda di quanto avviene in determinate aree spaziali e dell'enfasi narrativa a ciò conferita, con una focalizzazione pressoché costantemente interna.

Si consideri la sequenza in cui Tom visita Jeff prima di conoscere Lisa. Il vociare e la musica provenienti dall'appartamento del musicista decrescono quando i personaggi iniziano a parlare, con una scelta stilistica atta a favorire l'intelligibilità del loro dialogo piuttosto che a denotare un evento storico (verosimilmente le persone nell'altro appartamento non hanno ridotto il volume di voci e musica). Inoltre quando Jeff risponde al telefono e lo passa a Tom la voce dell'interlocutore è udibile in soggettiva, ma poco dopo essa si riduce a un livello così debole da non essere più intelligibile. Si tratta evidentemente di una scelta narrativa funzionale al mantenimento della suspense: l'informazione ci viene celata per essere svelata qualche minuto dopo dalla voce di Tom. Anche in seguito il volume dei suoni dell'appartamento del musicista viene aumentato o ridotto a seconda se Jeff e Lisa parlano o meno.

Similmente si noti come i rumori off della città (p.es. sirene e motori) sono chiaramente udibili – diventando figura – nei momenti privi di dialogo piuttosto che quando i personaggi parlano – ove sono sfondo. Ciò avviene palesemente nella sequenza in cui Jeff osserva Thorwald e la moglie litigare. Le voci di questi due personaggi sono in CLL, dal punto d'ascolto di Jeff, sebbene il punto di vista sia solo in parte soggettivo (ocularizzazione interna secondaria): vediamo quanto vede Jeff, dalla sua direzione, ma a distanze più ravvicinate di quanto possa fare lui a occhio nudo. Questo momento fornisce anche una prova di come suoni storici possano acquisire una valenza espressiva senza per questo far dubitare della loro collocazione nello spazio storico: i suoni più

preminenti sono una sirena urbana, segnale di allarme, e una sinistra sirena navale, che paiono commentare il rapporto problematico tra i due coniugi e anticiparne il macabro sviluppo<sup>52</sup>.

Un discorso analogo vale per la musica. L'unico brano decisamente extra-storico che ascoltiamo è l'ouverture sui titoli di testa e la coda alla fine del film. Per il resto la collocazione spaziale della musica è storica. Nella prima sequenza la musica di apertura prosegue senza soluzione di continuità tuttavia modificando la firma spaziale, il che la sposta da fuori a dentro il mondo narrativo. Poco dopo si intuisce che essa proveniva dalla radio nell'appartamento del musicista, il quale alcuni istanti dopo il termine del pezzo – cui segue un annuncio – cambia stazione introducendo un altro brano. Il personaggio del musicista permette nel corso del film di avere una fonte storica di musica dal vivo. Ciò ha una grande rilevanza drammaturgica e formale, poiché la canzone che egli sta componendo dissuade Miss Lonelyhearts dal suicidio (portandola poi ad avere un incontro con lui) e soprattutto funge da commento alla costruzione della relazione amorosa tra Jeff e Lisa – infatti si intitola *Lisa*, che è anche il nome ripetuto nel testo della canzone.

### **6.3 – *Persona* (Id., Svezia 1966)**

Scritto, prodotto e diretto da Ingmar Bergman.

Distribuzione: Svensk Filmindustri.

Suono: P. O. Pettersson, Lennart Engholm.

Missaggio: Olle Jakobsson.

Effetti sonori: Evald Andersson.

Musica: Lars Johan Werle.

Colonna sonora: mono.

Copia per l'analisi: DVD Bim (Italia), audio 2.0 *dual mono*.

Lo spazio filmico di *Persona* tende alla rarefazione, sia sul piano visivo sia sul piano sonoro. Ciò è dimostrato fin dal prologo: il primo ambiente narrativo visibile è l'esterno di un edificio, ma le inquadrature ravvicinate non danno un'idea chiara dello spazio, e il sonoro si limita a presentare dei rintocchi di campana in CL. Lo stesso vale per i piani all'interno della sala, che non permettono di comprendere la posizione relativa dei corpi delle persone anziane, mentre il bambino è sdraiato a ridosso di una parete bianca. Qui il sonoro, oltre allo squillo di un telefono e ad altri piccoli rumori in e off, presenta il suono di una goccia fuori campo in PP che non riceve alcuna motivazione. Nelle

<sup>52</sup> Devo questa riflessione a Gary Rydstrom, che la pone nel video *The Sound of Hitchcock* (USA 2008), incluso nelle edizioni DVD e BD di *La finestra sul cortile*.

scene successive presso la clinica le voci sono costantemente in PP o PPP a prescindere dal punto di vista, come pure i pochi rumori storici. Non c'è un fondo percepibile né un senso di firma spaziale, il che rafforza l'effetto di astrazione dello spazio. Ciò vale anche nelle scene all'interno della casa prospiciente il mare, in cui l'ambiente è più realistico ma è inquadrato come al solito prevalentemente con piani ristretti per dare una grande importanza alle figure umane piuttosto che alla cognizione degli spazi della casa. Nelle scene in esterno c'è un paesaggio sonoro coerente, tuttavia esso si mantiene per lo più distante sullo sfondo (tranne che in assenza di dialogo) mentre le voci continuano a essere pulite in PP, in una guisa irrealistica. Laddove vi siano inquadrature più ampie la distanza del punto d'ascolto è comunque più ridotta, come dimostra il campo lungo di Elisabeth inseguita da Alma sulla spiaggia in cui la voce di costei è in MPP. L'impressione generale è insomma quella di una colonna priva della naturalezza (altrettanto artificiale) del sonoro classico, ricordando in questo senso quelle interamente post-sincronizzate di film quali *L'eclisse* (Italia 1962) e *8 ½* (Italia/Francia 1963).

La goccia d'acqua udibile mentre Alma legge la lettera in auto evoca quella del prologo e non sembra essere sincronistica. Si tratta di un tipico caso di ambiguità nella collocazione, poiché pur ammettendo che il suono non proviene dallo spazio della scena non è chiaro se si tratta di un elemento del solo discorso oppure se è un modo per denotare col sonoro lo spazio del prologo. In ogni caso tale rumore contribuisce a costruire un significato implicito, ovvero il fatto che da questo momento il rapporto tra le due donne cambierà andando incontro alla crisi evocata dal prologo in cui i volti delle due donne erano proiettati su uno schermo accarezzato dal bambino.

## **6.4 – 2001: Odissea nello spazio (2001: A Space Odyssey, USA/Regno Unito 1968)**

Prodotto e diretto da Stanley Kubrick.

Distribuzione originale: MGM.

Montaggio sonoro: Winston Ryder.

Supervisore al sonoro: A. W. Watkins.

Suono in presa diretta: H. L. Bird.

*Chief dubbing mixer*: J. B. Smith.

Musica preesistente di Richard Strauss, György Ligeti, Johann Strauss figlio, Aram Khachaturian.

Colonna sonora: 6 canali magnetici (Super Panavision 70, Cinerama).

Copia per l'analisi: DVD MGM (USA, 1998), audio Dolby Digital 5.1.

La versione originale del film in 70mm ha una colonna sonora magnetica a sei canali con la configurazione Todd-AO, ovvero sinistra, centro-sinistra, centro, centro-destra, destra e surround. È stata inoltre distribuita una versione in 35mm con colonna magnetica a quattro canali. Stando alle informazioni rilasciate dall'archivio del British Film Institute (Londra), la copia ivi conservata è risalente al 1990 ed è stata approvata da Kubrick, ma la colonna è stampata col protocollo Dolby 41, ovvero “baby boom” col sistema Dolby A e una configurazione diversa rispetto all'originale. Qualora si tratti solo di un adattamento alla consuetudine del 1990 relativamente alle proiezioni in 70mm, appunto il “baby boom”, si presume che le basse frequenze inserite nei canali dedicati siano ricavate dal missaggio originale. Circa i canali mediani eliminati non è dato sapere se si è semplicemente eliminato parte del sonoro o se i suddetti sono stati missati nei canali attigui. La copia italiana in 70mm posseduta da e proiettata presso il cinema Arcadia presso Melzo (MI) ha dei canali mediani diversi rispetto a quelli esterni, dimostrando che il missaggio originale prevedeva cinque segnali diversi<sup>53</sup>. Il primo DVD del film (1998) contiene una colonna 5.1 che sembrerebbe essere originale in quanto i due canali surround sono identici – salvo una leggera differenza di fase in alcuni momenti – e i dialoghi sono direzionali, come usava nel 1968. Si dovrebbe capire se questa colonna è ricavata direttamente dalla versione “baby boom”, oppure se si tratta di un nuovo adattamento dei sei canali originali. La versione restaurata distribuita in 70mm con sonoro magnetico contiene i canali originali salvo l'applicazione del Dolby SR, ma non è da escludere un remissaggio conservativo per attingere alle incisioni musicali originali e alle colonne separate piuttosto che a una versione di successiva generazione quale il master filmico<sup>54</sup>. Tuttavia esistono anche copie in DTS, il cui mix è diverso, corrispondendo a quello presente nelle edizioni home video – DVD e BD – distribuite dal 2001 in poi, pressoché privo di direzionalità dei dialoghi. Esso pertanto è filologicamente inadeguato e altera gravemente la resa spaziale originaria<sup>55</sup>.

Questo film è rappresentativo della stereofonia degli anni '60 e '70 prima dell'avvento del Dolby Stereo ma segna un'evoluzione rispetto alla stereofonia dei primi anni: se la resa spaziale dei canali frontali segue il medesimo approccio, quella del surround è più immersiva poiché esso è usato per tutta la musica di commento, per gli ambienti e per alcuni effetti, e non solo in momenti sporadici come in altri film (e talvolta il canale surround era persino assente, come dimostrano film molto diversi quali *Lilli e il vagabondo* [*Lady and the Tramp*, USA 1955], *Quando la moglie è in*

<sup>53</sup> Ho assistito alla proiezione del film l'11/3/2017 e ho avuto questa sensazione (condivisa da Filippo Ulivieri, studioso del cinema di Kubrick), che mi è stata confermata dal proiezionista Daniele Buttafava il quale ha proiettato la copia in passato.

<sup>54</sup> Ciò stando a Cumbow (2001) è avvenuto per la rimasterizzazione in digitale del 2001.

<sup>55</sup> La colonna sonora italiana della copia di Arcadia trascura anch'essa la direzionalità originaria dei dialoghi, sembrando identica alla versione home video digitale.

vacanza [*The Seven Year Itch*, USA 1956], *Playtime* [*Id.*, Francia 1967], nonché il più tardo *Zardoz* [*Id.*, Irlanda/USA 1973]). Basti confrontarlo con *La tunica* (*The Robe*, USA 1953) e *Come sposare un milionario* (*How to Marry a Millionaire*, USA 1953): nel primo caso il surround interviene per momenti topici della narrazione, per suggerire una dimensione soprannaturale, mentre nel secondo caso esso è addirittura usato una sola volta per un minuto durante il coro all'inizio del racconto. Nell'epoca del Dolby 70mm l'immersività del surround diverrà un poco più accentuata, come dimostrano per esempio *Alien* (*Id.*, USA 1979) e *Apocalypse Now* (*Id.*, USA 1979).

Il film si apre con un'ouverture costituita da un frammento del brano *Atmosphères* di György Ligeti, su schermo nero. In seguito appare il logo MGM, insieme al quale inizia l'introduzione del poema sinfonico *Così parlò Zarathustra* di Richard Strauss. Questo incipit fornisce un'impostazione sonora – relativamente alla musica del discorso, unico suono presente – che caratterizza l'intero film, in cui la musica ha per lo più una firma spaziale affine ed è inserita in tutti i canali. La massa sonora di questo brano è molto corposa, considerando l'impiego di una grande orchestra che culmina in un “tutti” fortissimo. Il senso di ampiezza è dato dalla firma spaziale, grazie a un certo grado di riverberazione e all'uso del surround. La vastità è suggerita anche dalle immagini che mostrano l'allineamento di tre astri nello spazio cosmico. Si noti inoltre che il brano presenta la ripetizione per tre volte di un inciso di tre note ascendenti – Do-Sol-Do – che pare alludere ai tre oggetti celesti; inoltre il moto ascendente ha un correlativo visivo poiché gli astri si muovono come a suggerire una progressiva elevazione del punto di vista.

La prima sequenza narrativa, priva di musica, adempie alla stessa funzione relativamente a rumori e voci della storia. La resa spaziale è realistica sia in termini di distanza sia soprattutto in termini di direzione. C'è quindi coerenza nelle prospettive, come accadrà pressoché in ogni momento del film. Le voci e i rumori sono propagati da tutti i canali frontali con una diversa direzione a seconda della posizione della fonte sullo schermo. I rumori d'ambiente e le urla delle scimmie sono inseriti leggermente anche nel canale surround. Nel suo complesso quindi il film offre uno spazio sonoro molto mimetico, che favorisce un senso di presenza ancorché quasi solo frontale, data la scarsa immersione nello spazio storico offerta dal surround.

La prima apparizione del monolito (come pure le due successive) è accompagnata da un frammento del *Kyrie* dal *Requiem* di Ligeti. La collocazione di questa musica è presumibilmente nel discorso, trattandosi di un “commento”. Eppure dato il suo stile e il contesto del racconto essa potrebbe essere interpretata come un'emanazione sonora del misterioso oggetto. La sua interruzione è repentina, sincrona con lo stacco a una scena successiva ovvero con la sparizione del monolito dallo schermo, il che è una scelta inconsueta anche all'interno del film (cfr. Paulus 2009). A ogni modo non c'è una firma spaziale che aiuti l'interpretazione.

La resa spaziale è confermata chiaramente nella sequenza dell'arrivo di Floyd sulla stazione orbitante, in cui si hanno le prime battute di dialogo del film, in cui la firma spaziale e la direzione delle voci sono coerenti con la prospettiva visiva. Nell'inquadratura di Floyd e Miller che camminano c'è una parziale infrazione della prospettiva, poiché la direzione delle loro voci e dei loro passi muta in maniera esigua (il maggior livello del canale sinistro, in coerenza prospettica, è pressoché impercettibile). Si noti come la voce femminile *on the air* dell'annuncio è inserita anche nel surround, proprio perché propagata nell'intero ambiente, ma non nel canale sinistro: probabilmente si tratta di una separazione spaziale atta a favorire il mascheramento, al fine di mantenere la piena intelligibilità del dialogo dei due uomini le cui voci paiono avere il medesimo livello di quella della donna. La distanza è resa con maggiore approssimazione del solito, poiché le voci sono troppo vicine (in effetti in un coerente CL si sarebbe rischiato di affaticare la comprensione), eppure c'è un senso di avvicinamento che è relativamente proporzionale a quello visivo. L'ambiente della cabina telefonica è diverso rispetto a quello del resto dell'astronave, il che permette di denotare un cambio di ambiente sonoro.

La sequenza di dialogo tra Floyd e gli scienziati russi presenta inizialmente una distanza sonora, MPP, forse troppo vicina rispetto a quella visiva (campo medio). Tuttavia anche in questo caso si mantiene il senso di proporzione poiché il raccordo sull'asse in avanti è accompagnato da un raccordo sull'asse sonoro, poiché il dialogo è ora in PP. L'inquadratura dello scienziato russo non presenta un accertabile cambio di distanza sonora, il che non stupisce essendo la voce già in PP; inoltre nemmeno il raccordo sull'asse all'indietro finale comporta un cambio di distanza sonora, salvo un'aumento della firma spaziale negli ultimi secondi, come a condurre gradualmente alla dissolvenza in chiusura. In ogni caso l'impressione generale è ancora una volta quella di una coerenza prospettica.

La resa spaziale è alquanto rigorosa anche nella sequenza della conferenza. Quando Floyd inizia a parlare, inquadrato in un totale frontale, la sua voce è centrata e in CM, mentre dopo lo stacco sulla sua mezza figura viene raccordata sia sulla distanza (a un PP) sia nella posizione poiché è spostata sulla destra ove lui è visibile. Nel totale in controcampo la voce mantiene la stessa direzione ma è in MPP poiché è aumentata la distanza e risulta anche un po' meno a fuoco, ovvero meno frontale, come si nota dopo lo stacco successivo su un campo medio frontale in cui la voce è coerentemente anch'essa frontale.

La parte della "missione Giove" presenta un *sound design* particolarmente suggestivo, a partire dal fatto che la voce di HAL 9000 è inserita in tutti i canali in modo uniforme (tranne quando si ode *en abîme* dalla trasmissione televisiva), rafforzando la sua onnipresenza nell'ambiente, mentre quella dei personaggi è direzionale col solito principio. Consideriamo due sequenze di

grande interesse, l'una il *set up* dell'altra, cioè quelle in cui rispettivamente Bowman e Poole escono dall'astronave Discovery per intervenire sul pezzo presumibilmente difettoso. Il sibilo dell'ossigeno e il respiro hanno una firma spaziale che suggerisce la trasmissione radio, come se ascoltassimo dal punto d'ascolto dell'astronauta dentro al Discovery. Tuttavia ascoltiamo lo stesso suono anche durante le inquadrature nello spazio aperto, con la differenza che qui c'è sullo sfondo silenzio assoluto mentre sul Discovery si ascolta il fondo dell'astronave, un debole suono grave e ovattato. Se si aggiunge il fatto che sibilo e respiro sono inseriti in tutti i canali in modo non direzionale ne risulta l'effetto di una soggettiva uditiva che ci permette di simularci nel corpo di Bowman prima e di Poole poi. Si noti la scelta sottile di aumentare il livello del sibilo pochi istanti prima che HAL stacchi l'ossigeno a Poole, ovvero prima di un repentino silenzio assoluto: questa scelta, insieme ai rapidi raccordi sull'asse a stringere sull'occhio di HAL presente sulla capsula, si avvale dell'effetto pauroso dell'*auditory looming* per incrementare il senso di pericolo imminente. Un'alternanza affine a quella tra l'assenza di fondo nello spazio aperto e la sua presenza dentro il Discovery si ripete quando Bowman esce con la capsula per raccogliere il cadavere di Poole: le inquadrature dentro alla capsula presentano un fondo con rumori elettronici per lo più privi di salienza, mentre quelle esterne presentano silenzio assoluto. Ciò costituisce un chiaro esempio di come col solo sonoro si possa denotare l'alternanza di spazi similmente a quanto avviene per solito in una sequenza a montaggio alternato. Un'altra eccezione al rigore spaziale del film si ha ancora una volta con suoni *on the air*: le voci radiotrasmesse (con coerente distorsione) di Bowman e di HAL sono udibili durante le inquadrature nello spazio aperto (senza fondo) oltre che nei rispettivi ambienti in cui sono ricevute. Inoltre si rileva la scelta sottile ma curiosa di collocare dette voci sulla destra in un totale che mostra il Discovery a sinistra e la capsula a destra, come a voler segnalare la direzione o il punto d'ascolto di Bowman: tuttavia si tratta di una trasmissione radio priva della firma spaziale di un ambiente in cui si propagherebbe e inaudibile dalla prospettiva suggerita dall'immagine. Perciò la posizione scelta non comporta una denotazione indexicale di una specifica direzione nello spazio filmico ma solo una corrispondenza nello spazio visivo, ovvero nel quadro.

La sequenza del viaggio oltre l'infinito è quella più immersiva, grazie alla sensazione di movimento in avanti in soggettiva e all'avviluppante massa sonora. In particolare le basse frequenze, per loro natura le più immersive, sono piuttosto aggressive e producono un effetto vibrotattile particolarmente efficace. Nell'ultima sequenza del film il sonoro rafforza il senso di vuoto della stanza in cui si è ritrovato il protagonista. I rumori dell'anziano Bowman a tavola hanno una marcata riverberazione, come pure la musica di Ligeti, tratta da *Aventures* e appositamente trattata con riverbero. La firma spaziale che questa musica acquisisce, sommata alla sua natura

“rumoristica” (voci che emettono versi non intonati e strani rumori), è trattata in modo differenziale suggerendo un evento storico, come se si trattasse di suoni extraterrestri che si propagano nella stanza e quindi udibili da Bowman (cfr. Paulus 2009).

## **6.5 – Lezioni di storia (*Geschichtsunterricht*, Italia/Germania Ovest 1972)**

Prodotto e diretto da Jean-Marie Straub e Danièle Huillet.

Sonoro: Jety Grigioni.

Musica: Johann Sebastian Bach.

Colonna sonora: mono.

Copia per l'analisi: videoregistrazione della trasmissione in telecinema sul canale Rai Tre, audio mono.

Il film presenta delle caratteristiche emblematiche dello stile dei due registi che lo pongono perciò a distanza radicale dalle convenzioni del cinema narrativo. Esso è costituito da un adattamento dal frammento del romanzo *Die Geschäfte des Herrn Julius Caesar* (1937-1939) di Bertolt Brecht. Il testo è presentato da un uomo di epoca attuale e dai personaggi d'epoca romana che egli incontra. La forma filmica è severa: la recitazione ha uno stile rigorosamente impostato, l'immagine ha il desueto rapporto d'aspetto 1,37:1 e le riprese sono in buona parte statiche. L'ambiente sonoro è essenziale, presentando esclusivamente la registrazione in presa diretta delle voci e dei rumori d'ambiente. L'unica eccezione si riscontra alla fine del film, ove è inserito un frammento della *Passione secondo Matteo* di Bach, senza motivazione indexicale.

Il punto d'ascolto è sempre coerente col punto di vista: la firma spaziale delle voci cambia a ogni inquadratura, risultando a seconda dei casi più vicine o più lontane. Una piccola eccezione è nell'inquadratura di un paesaggio mentre il banchiere è fuori campo: la firma spaziale della sua voce off denota un differente punto d'ascolto ma non una presenza di spalle. Ciascuna ripresa sonora è sincronica con le immagini, e il montaggio sonoro si limita a mettere in serie le singole riprese, perciò sono udibili gli stacchi grazie alle interruzioni dei rumori di fondo. Ciò rispetta la prassi che dei due cineasti, in aperta antitesi con quella del cinema classico che prevede una diffusa post-sincronizzazione e un montaggio sonoro inaudibile. Come risulta da un'intervista (Anonimo 1985), Straub e Huillet manifestano un'idiosincrasia verso il doppiaggio nel modo in cui viene normalmente utilizzato, segnatamente in Italia. Huillet spiega che la presa diretta di voci e rumori

insieme, diversamente dal doppiaggio, richiede che «the locations that you choose have to be right not only in terms of the images but also in terms of the sound» (in *ivi*: 151). Straub aggiunge che «[f]ilming in direct sound you can't fool with the space, you have to respect it, and in doing so offer the viewer the chance to reconstruct it, because the film is made up of "extracts" of time and space». Come osservato da Straub l'uso esclusivo della presa diretta "integrale" condiziona il montaggio dell'immagine, perché «you can't allow yourself to play with the images: you have blocks of a certain length and you can't use the scissors any way you want, for pleasure, for effect» (in *ibidem*). Huillet conferma l'intransigenza di tale approccio asserendo che «each image has a sound and you're forced to respect it» (in *ivi*: 252).

Da tutto questo consegue una decisa materialità del suono, ovvero un rilievo particolare della causa materiale dei suoni che nel caso delle voci, le quali si fanno portatrici di un significato preciso (il testo di Brecht), viene dissociata dall'effetto percettivo, privo di qualsivoglia risonanza emotiva diversamente dagli stili attoriali consueti. Anche Doane riscontra nei film di Huillet e Straub (e in quelli di Godard) una deliberata disgiunzione tra suono e significato per enfatizzare ciò che Barthes chiama la «grana della voce» contro il suo potere di rappresentazione (Doane 1985b: 173). Allo stesso tempo si rileva un rispetto quasi sacrale per la dimensione spazio-temporale dell'evento registrato dal dispositivo, con un approccio in questo senso vicino a quello del cinema documentario. Come spiegato da Ruoff nei documentari "osservativi" si ha un'enfasi sulla presa diretta del suono che favorisce la fedeltà rispetto all'intelligibilità: «the sounds resurface in their *phenomenological materiality* [*corsivo mio, n.d.a.*]», cui consegue la riduzione della possibilità di dividere i suoni nelle tre categorie canoniche, e la mancanza di fluidità nel montaggio sonoro (1992: 217-218). In *Lezioni di storia* la piena intelligibilità è sempre garantita – perché quanto viene presentato è ovviamente una messa in scena – ma essa è comunque inscindibile dalla fedeltà.

Uno stile così insolito intacca un poco l'idea di Chion dell'assenza di un contenitore sonoro assimilabile a quello visivo, come lui stesso riconosce commentando un'altra opera di Huillet e Straub, *Les yeux ne veulent pas en tout temps se fermer ou Peut-être qu'un jour Rome se permettra de choisir à son tour* (Germania Ovest/Francia 1969). Questo film «mostrava [...] ciò che avrebbe potuto essere una scena sonora, in altri termini un contenente sonoro dei suoni, a condizione di ammettere che i suoni fossero le voci degli attori e il loro testo, e il contenente il rumore urbano nel quale voce e testo erano calati». Quando gli attori parlano a lungo fuori campo, «la loro voce non risuona come quella del fuori campo tradizionale, interamente determinato dall'immagine: essa è "nello stesso luogo" di quella degli attori che si vedono, luogo disegnato dal rumore d'ambiente». Nel caso di *Lezioni di storia* a seconda dell'ambientazione si ascoltano sullo sfondo rumori urbani o

suoni della natura, i quali svolgono nei confronti delle voci il medesimo «ruolo di rivestimento e di omogeneizzazione, di inserimento in un luogo» di cui parla Chion (2001: 71; cfr. anche 2009: 227).

Oltre alle sequenze parlate e ad altre brevi inquadrature vi sono riprese di 10 minuti ciascuna che mostrano viaggi in auto dell'uomo contemporaneo, con la macchina da presa che lo inquadra posizionata sui sedili posteriori. Le immagini ci mostrano un viaggio per le strade di Roma, lasciandoci osservare a piacere il paesaggio urbano; allo stesso modo il sonoro è in pratica un'opera di *soundscape* documentario dall'interno dell'auto (di cui pertanto ascoltiamo costantemente il motore), una traccia sonora oggettiva dell'epoca allora presente cui Straub e Huillet hanno contrapposto quella passata evocata da Brecht e dai costumi degli attori.

## **6.6 – *Guerre stellari (Star Wars, USA 1977)***

Produzione: Lucasfilm Limited.

Distribuzione: Twentieth Century Fox.

Regia: George Lucas.

Suono in presa diretta: Derek Ball.

Supervisore al montaggio del suono: Sam Shaw.

Effetti speciali sonori e di dialogo: Ben Burtt.

Montaggio sonoro: Robert R. Rutledge, Gordon Davidson, Gene Corso.

Missaggio: Don MacDougall, Bob Minkler, Ray West, Robert Litt, Mike Minkler, Lester Fresholtz, Richard Portman.

Musica: John Williams.

Colonna sonora: Dolby SVA; Dolby 70mm; mono.

Copie per l'analisi: Laserdisc (Giappone 1986), audio Dolby Surround.

Il film è stato distribuito originariamente in tre versioni: quella in Dolby 70mm “baby boom”, quella in Dolby SVA e infine quella mono. Negli anni la colonna sonora è stata più volte modificata, tanto che oggi è difficile reperire quella originale, anche perché il film nella sua interezza è oggi distribuito nella sua “edizione speciale” del 1997 con ulteriori modifiche fino all'edizione in BD del 2011. Mentre la versione “baby boom” è inaccessibile, quella SVA è reperibile grazie alle edizioni giapponesi in Laserdisc e costituisce l'oggetto della presente analisi. L'ascolto è avvenuto utilizzando il processore Dolby Pro Logic per ottenere i quattro canali codificati nei due canali matriciali. Non ho potuto visualizzare i quattro canali in assenza di un software adatto.

Il celebre incipit è un emblema del sonoro denso, spettacolare e immersivo di questo film che ha segnato la storia del *sound design* (cfr. Jullier 2006). La prima sequenza del film stabilisce il tono della narrazione grazie anzitutto al sonoro: un'ouverture per grande orchestra sinfonica che evoca avventura, eroismo e gloria. Sul piano spaziale la grande massa sonora ("tutti") si espande in un ampio ambiente stereofonico con una firma spaziale mediamente riverberante, suggerendo un vasto spazio, come in effetti mostrato dalle immagini – lo spazio cosmico disseminato di stelle. Dopo il tema musicale principale e una frase di transizione, quasi al termine della panoramica verso il basso c'è un accordo grave fortissimo, in cui spiccano i timpani e il tam-tam, che veicola un effetto di solidità e di sostegno, creando una corrispondenza crossmodale sia con la conclusione del movimento verso il basso, sia con l'apparizione nella parte inferiore dell'inquadratura della porzione di un gigantesco astro, sia infine con l'apparizione della grande astronave nemica che segue di qualche secondo. Circa quest'ultimo punto si noti come le vibrazioni del tam-tam si fondono con il sibilo della prima astronave, creando un legame uditivo tra lo spazio astratto della musica discorsiva e quello concreto della storia. Inoltre il rombo della grande astronave sembra la prosecuzione dei suoni più gravi del summenzionato accordo, il quale pare quindi costituirne l'anticipazione sonora. I rumori provocati dalle astronavi sono molto riverberanti, a conferma dell'ampiezza spaziale che si intende trasmettere col sonoro. L'immersione è prodotta dall'abbondanza di frequenze basse, col loro effetto vibrotattile, nonché dall'uso aggressivo del surround sia per la musica sia per i numerosi rumori (si ricordi che il canale surround non è discreto ma deriva da una decodificazione dei canali matriciali Lt-Rt, perciò può contenere anche suoni non intenzionali derivanti dai canali frontali).

Lo stacco sulla scena all'interno dell'astronave piccola è raccordato dal sonoro, privo di soluzione di continuità. L'ambiente sonoro è ancora rumoroso, considerando gli effetti storici della battaglia e la musica. Si noti come la sirena (che si espande anche nel surround) cambia firma spaziale e direzione pur senza una chiara localizzazione, a conferma della consueta approssimazione della resa spaziale del sonoro. Mentre rumori e musica sono stereofonici i dialoghi sono costantemente centrati (nonché frontali e in piani ravvicinati) come da nuova prassi dell'epoca del Dolby Stereo, ma mantengono una firma spaziale che conferisce naturalezza. Una significativa eccezione alla regola dei dialoghi mono si ha nel caso dei Jawa, le cui voci sono diffuse nei canali frontali, ma d'altronde si tratta di un vociare incomprensibile piuttosto che di normale dialogo. Lo stesso vale per le voci dei clienti della taverna: si noti in particolare la risata della grossa creatura verso cui Luke si gira, che viene raccordata nella posizione (proviene da sinistra quando è off e dal centro quando è in, malgrado il raccordo presenti un piccolo ritardo; l'effetto è comunque poco percepibile a causa della matrice Dolby SVA). Nei momenti più silenziosi sono percepibili dei fondi

che denotano gli ambienti interni ed esterni, ad esclusione delle scene nello spazio aperto. Ambienti, musica e molti effetti si espandono nel surround, creando il supercampo; proprio perché questo canale non è discreto i suoni che contiene non sono separati chiaramente dai canali frontali, il che si percepisce soprattutto con certi effetti come i rumori delle spade laser nel duello tra Obi-Wan Kenobi e Darth Vader e soprattutto con la musica di commento – quasi onnipresente, seguendo l'approccio *wall-to-wall* dello *score* classico (cfr. Kalinak 1992). Ciò che si ascolta nel surround offre per lo più un senso di riverberazione, un ampliamento della dimensione frontale, con un risultato molto spettacolare benché la musica presenti talvolta una profondità un poco confusa.

Talvolta la distanza delle voci approssima la distanza del punto di vista, come nella scena dell'arrivo di C3PO e R2D2 nel deserto: il dialogo del primo è in MPP nella prima inquadratura (un campo lungo che stringe su un campo medio con uno zoom), mentre nei momenti successivi resta in PP a prescindere dal tipo di piano, salvo tornare in MPP quando C3PO alla conclusione della scena è in campo lungo. Un esempio relativo alla frontalità è riscontrabile nella scena in cui Obi-Wan Kenobi propone l'addestramento a Luke: le loro voci perdono la frontalità quando si voltano di spalle rispetto al punto di vista.

Anche il trattamento della musica storica nella taverna è canonico: essa ha una forte firma spaziale ed è per lo più in CM salvo passare a un MPP quando vengono inquadrati i musicisti da vicino; inoltre la sua spazializzazione resta immutata a prescindere dalla prospettiva spaziale.

La sequenza dello scontro finale dimostra come il livello degli effetti sia gestito a scopo prettamente espressivo, acciocché i dialoghi siano facilmente intelligibili, piuttosto che di fedele denotazione spaziale rispetto a un dato punto d'ascolto. Tuttavia, diversamente dallo scontro precedente tra i protagonisti sul Millennium Falcon e le astronavi da caccia nemiche, qui il sonoro contribuisce a identificare i diversi spazi inquadrati: c'è differenza negli sfondi tra lo spazio aperto, gli interni della base e della Morte Nera, e gli interni delle varie navicelle, talvolta distinguibili anche a occhi chiusi data la loro differenza. La voce soprannaturale di Obi-Wan Kenobi udita da Luke ha ovviamente una qualità diversa rispetto alle altre, essendo in PPP e leggermente riverberante.

La scena finale presenta un classico esempio di collocazione ambigua della musica: la marcia ha chiaramente una funzione discorsiva, per connotare il trionfo dei protagonisti; nondimeno essa potrebbe essere un evento storico, perché la presenza di musica in una cerimonia del genere è affatto plausibile (Kassabian 2001: 654-670).

## 6.7 – *Prénom Carmen* (Id., Francia 1983)

Produzione: Alain Sarde per Sara Films, JLG Films, Films A2.

Regia: Jean-Luc Godard.

Suono: François Musy.

Colonna sonora: mono.

Copia per l'analisi: DVD RaroVideo (Italia, 2006), audio 2.0 *dual mono*.

*Prénom Carmen* presenta uno stile sonoro decisamente anti-classico, colmo di soluzioni “aberranti” per le quali Godard è famoso e che sono state studiate da Alan Williams in un articolo del 1982. In generale la colonna sonora è dominata dalla frammentarietà, poiché i suoni sono spesso montati con netta indipendenza dalle immagini, sottoposti a interruzioni brusche e all'intromissione arbitraria di frammenti, e prevede inoltre l'abolizione della gerarchia tra le categorie di suono. Come nota Williams, nell'«American studio style» c'è «a clear and reassuring *hierarchy* of sonic importance, and this is reinforced by a kind of step-system of sonic presence, sounds being made to seem either very close (important) or distant (“atmosphere”)); lo stile di Godard si pone invece in antitesi con questo, poiché il cineasta «somehow “deconstructs” classical narrative». Se ne ha prova lampante nel trattamento dei suoni d'ambiente insieme al dialogo: quest'ultimo è registrato in presa diretta, rispettando il punto di vista, con una firma spaziale aggressiva che lo radica nell'ambiente in cui viene prodotto ed è perciò anche contaminato da rumori invadenti i quali dunque «refuse to go away when “more important” information appears. In other words, at least in his location sound recording, [Godard] refuses to do the spectator's work». Mentre «[c]lassical narrative sound recording is “transparent” – or *inaudible* as recording – in the same way that classical visual practices are», le scelte di Godard «result in a sonic texture that is *continuously audible*, both by the fact of implicit comparison with “normal” sound and by the work required in listening». Williams fa un riferimento esplicito allo spazio sonoro asserendo che «[w]ether realistic or not [...], Godard's refusal to edit and mix transparently results in a close aural equivalent of long takes with great depth of field»: come la profondità di campo di William Wyler è «liberale e democratica» secondo un celebre saggio di Bazin ([1948] 1999), così «Godard's location sound is [...] egalitarian in that it does not pick and choose for the auditor» (Williams 1985: 336-337). Nel cinema di questo regista insomma «the impact of “raw materials” is preserved», provocando una compresenza di formalismo (la separazione nega l'unità enfatizzando il lavoro del dispositivo) e realismo (la registrazione “così com'è”) (1985: 344). Relativamente alla resa spaziale c'è una tensione tra il realismo della presa

diretta e la sua coerenza prospettica da un lato, e la denotazione frammentaria e innaturale di campi sonori diversi a prescindere dai campi visivi.

Giova esaminare la sequenza in cui Carmen visita lo zio in ospedale, perché è emblematica di svariati aspetti del campo sonoro del film nel suo complesso. Il suono della scena all'esterno dello stabile è introdotto con un ponte sonoro a precedere durante la scena che mostra il quartetto d'archi; le voci sono in CM, coerentemente con la prospettiva. Nella seconda inquadratura, in cui la giovane si dirige verso l'ingresso dell'ospedale, il suono di presa diretta tace per pochi istanti e viene sostituito da un frammento di un quartetto, per poi tornare udibile come prima. Frammenti musicali come questo tornano ancora nel corso della sequenza, missati insieme al dialogo, in modo programmaticamente disordinato. All'interno della stanza d'ospedale le voci dei due personaggi sono molto riverberanti e sporcate dallo strepito dei locali adiacente, con un risultato iperrealistico antitetico all'artificialità di altri aspetti (lo stile attoriale e dialogico). Una frazione di secondo dopo che Carmen scoppia a ridere il sonoro si silenzia totalmente per tornare a essere udibile nell'inquadratura successiva, pochi istanti dopo: il suono della risata è quindi spezzato da questo momento di silenzio, soluzione presente anche altrove nel film in relazione ad altri suoni. Un simile lavoro sui tagli sonori rende quasi possibile isolare mentalmente delle unità sonore di montaggio; tuttavia ciò che prevale è un effetto di interruzione e di cesura che palesa l'artificio del prodotto filmico, piuttosto che l'individuazione di diversi ambienti sonori.

Nella sequenza appena analizzata il dialogo ha tendenzialmente una rilevanza maggiore rispetto alle altre due categorie sonore. Invece il segmento filmico che segue la rapina, che segna l'inizio dell'avventura tra Carmen e Joseph, è centrato su rumori e musica, ancora pesantemente sottoposti all'estetica del frammento. Durante l'inquadratura sulle mani dei due protagonisti al dialogo e ai rumori storici off si aggiunge la musica over, un estratto dal II movimento ("Adagio ma non troppo") del *Quartetto per archi* n. 10 op. 74 di Beethoven (dalla collocazione ambigua, come vedremo); poi i suoni della scena scompaiono e resta la musica insieme ai gridi dei gabbiani, versi che anticipano le inquadrature della riva marina. In questo contesto sonoro avviene uno stacco su una strada trafficata di notte, e poco dopo la musica esce bruscamente dalla colonna; segue l'entrata del rumore del mare, che sul piano uditivo ha affinità con quello del traffico automobilistico, creando un cortocircuito percettivo (anticipato dalla prima sequenza del film). Con una dissolvenza in chiusura il sonoro ammutolisce, per poi ospitare ancora la musica che entra con una dissolvenza, pochi istanti prima dello stacco su un'inquadratura della riva cui segue l'ingresso del suono d'ambiente in. Il gioco di entrata e uscita *a ondate* della musica prosegue per alcuni secondi, interrompendosi un istante prima dello stacco sulla scena successiva, in cui i due protagonisti giungono in un'autostazione. Il rumore del mare prosegue immutato su questa inquadratura, e si

interrompe a ridosso del rombo di un camion, dando luogo ancora una volta a un'analogia tra i due elementi sonori, quello naturale e quello artificiale. In questa circostanza si può parlare di ponte sonoro a seguire, perché l'immagine mostra la scena successiva in anticipo rispetto al suono, ma un sistema estetico come quello di *Prénom Carmen* mette spesso in crisi definizioni di questo tipo: per esempio i suoni dei gabbiani e del mare anticipano l'immagine della riva con la modalità di un ponte sonoro, ma tali suoni vengono estromessi prima dello stacco visivo e reintrodotti un attimo dopo, provocando perciò una discontinuità. Inoltre la frammentarietà del montaggio visivo e sonoro comporta all'interno del film alcuni casi di disgiunzione temporale tra immagine e suono, tanto che risulta difficile stabilire la loro relazione cronologica. Nel caso di questa sequenza la collocazione temporale nella storia di immagini e suoni della strada resta indefinita, quindi non si può chiarire se i rumori del mare sono anteriori o posteriori a essi bensì solo che non c'è simultaneità.

La musica è sempre storica: la canzone di Tom Waits – udibile una sola volta – proviene apparentemente da un dispositivo audio nella scena (stando al realismo della riverberazione), mentre i quartetti di Beethoven sono suonati da personaggi del racconto. Questi ultimi pezzi, che ricorrono molto di frequente, offrono uno spunto interessante per riflettere sul confine tra storia e discorso. Talvolta l'inserimento della musica è associato alle immagini degli strumentisti, talaltra no, e non tutti i frammenti dei quartetti che ricorrono nel film sono simultanei alle scene: perciò se da una parte è indubbio che questa musica sia genericamente implicata nel racconto, dall'altra non si può stabilire in modo inequivoco l'appartenenza alla storia di ciascuno dei frammenti. In ogni caso la distinzione è alquanto cavillosa, restando fermo che questi pezzi svolgono una funzione discorsiva, soprattutto quando intervengono in modo asincrono, come accade nella sequenza analizzata poc'anzi. Si noti che quando durante le suonate viene inquadrata Claire, la violista, i suoi movimenti sullo strumento sono palesemente incongruenti con la musica che si ascolta. D'altronde la donna che interpreta Claire, Myriem Roussel, è un'attrice estranea al quartetto Prat, mentre gli altri musicisti sono autentici – lo si nota chiaramente grazie alla verosimiglianza dei loro movimenti: è evidente come l'interesse sia nell'enfatizzare la finzione, e quindi anche le dinamiche discorsive. La scelta di ricorrere a quartetti crea una corrispondenza crossmodale perché replica sul piano dell'organico strumentale la quadripartizione del mondo narrativo, abitato dalle storie di Carmen e Joseph, dello zio regista, del film da realizzare e dei musicisti che concertano i quartetti medesimi (cfr. Ghighi 1985).

## 6.8 – *Strange Days* (*Id.*, USA 1995)

Prodotto da James Cameron e Steven-Charles Jaffe per Lightstorm Entertainment.

Distribuzione: Twentieth Century Fox.

Regia: Kathryn Bigelow.

*Sound design*: Gary Rydstrom; assistente: Lora Hirschberg.

Suono in presa diretta: David Ronne.

Supervisione al montaggio sonoro: Gloria S. Borders.

Missaggio: Richard Beggs, Lora Hirschberg, Tom Johnson.

Musica: Graeme Revell.

Colonna sonora: Dolby Digital 5.1.

Copia per l'analisi: DVD Twentieth Century Fox (Italia 2002), audio Dolby Digital 5.1.

Questo film ha un sonoro rappresentativo del cinema spettacolare contemporaneo, salvo la peculiarità delle riprese in soggettiva visiva e sonora (apparentemente senza stacchi) mediate nel racconto dal congegno denominato SQUID. Nei primi secondi del film ascoltiamo dei rumori elettronici e poi due voci. Questi suoni sono inseriti nel canale centrale, mentre gli altri presentano un rumore appena percepibile, perciò l'effetto è quello di una singola direzione. Quello che vediamo e udiamo in seguito è esattamente l'esperienza visiva e uditiva di un personaggio, registrata con lo SQUID e rivissuta dal protagonista del film, Lenny. Il sonoro ha qui una resa spaziale scrupolosamente mimetica in ultracampo, con coerenza nella firma spaziale e anche nella direzione: i suoni provengono dal canale che rispetta la loro posizione nello spazio filmico, con un mirabile effetto di presenza in soggettiva in cui piombiamo repentinamente nel passaggio da un sonoro pressoché mono. La voce del personaggio di cui riviviamo l'esperienza (per il tramite di Lenny) è diffusa dal canale centrale e da quelli surround, mentre il rumore provocato dalla calza mentre egli la indossa sul capo – come pure quando in seguito la toglie – emana da tutti i canali normali: questa scelta conferisce una preminenza ai suoni nello spazio corporeo rispetto agli altri, rendendo persino la stimolazione tattile della calza passata sopra alle orecchie.

Quando Lenny toglie lo SQUID la resa spaziale cambia drasticamente, passando a uno stile convenzionale – il consueto supercampo – che costituisce la norma intrinseca del film per le scene normali (cioè non mediate dallo SQUID) così come la norma estrinseca propria del cinema contemporaneo. I dialoghi sono costantemente frontali in PP con qualche marginale cambio di frontalità e di distanza approssimativamente coerente con la prospettiva visiva, mentre la direzione non muta essendo queste voci nel canale centrale (il *walla* è tuttavia spesso negli altri canali,

proprio perché è più “ambiente” che voce significante). Gli effetti sono invece stereofonici (come la musica), senza precisione prospettica soprattutto quando off, e possono espandersi anche nel surround per denotare un punto d’ascolto con naturalezza, mentre il loro volume è sovente gestito in ossequio alle esigenze narrative piuttosto che al realismo. Lo si nota chiaramente quando Lenny torna a casa per la prima volta nel racconto: il rumore dell’elicottero si sposta nei canali, anche surround, per enfatizzare l’ampiezza del supercampo, ma si tratta di un suono di scarsa salienza (questa scelta è ripetuta più tardi nella scena in strada in cui Lenny, Mace e Max si confrontano). Quando Lenny si sveglia il giorno seguente si ascolta l’audio del televisore (unidirezionale, con riverberazione negli altri canali) ricordato più volte nella posizione e nella distanza tra il canale centrale e quello destro, per suggerire l’orientamento reciproco del televisore rispetto all’uomo, tuttavia senza scrupolo realistico: in questa evenienza infatti si sarebbe dovuto inserire il suono nel surround destro piuttosto che nel destro frontale quando il televisore è nell’altra stanza, e nel canale destro piuttosto che al centro quando esso è alla destra frontale del punto d’ascolto. Tale scelta è dovuta probabilmente al timore di un effetto troppo distraente data la salienza e la durata prolungata della trasmissione televisiva che cattura l’attenzione di Lenny. L’approssimazione si manifesta in maniera eclatante verso la fine della sequenza, quando Lenny sceglie un abito in camera sua: l’audio del televisore resta a destra malgrado i due stacchi del campo/controcampo. Anche al risveglio di Lenny il giorno successivo c’è un suono *on the air*, la telefonata di Max. In questo caso però la voce è trattata come un normale dialogo, in MPP centrato, salvo la distorsione telefonica dal punto d’ascolto di Lenny. Questo è un ulteriore esempio di come la voce che costituisce dialogo abbia una resa spaziale più astratta rispetto ad altri suoni pur vocali e di rilevanza narrativa come la trasmissione televisiva, con cui Lenny non interagisce diversamente da quanto accade con la telefonata. Un’altra eccezione è riscontrabile quando Lenny e Mace vanno da Tick in auto: mentre sono fermi a un posto di blocco si ascolta la battuta «Let it go» dal fuori campo sinistro diffusa in CL dal canale sinistro. La differenza nella resa spaziale è palese anche nelle scene d’azione non mediate dallo SQUID: rispetto per esempio alla summenzionata sequenza iniziale, nell’inseguimento in auto tra i poliziotti e Lenny e Mace gli effetti sono stereofonici e si espandono anche nel surround ma sempre nel modello supercampo, e il campo visivo è indispensabile per comprendere la localizzazione delle fonti. La sequenza dell’incontro tra Lenny e Faith in discoteca è invece emblematica di una resa spaziale convenzionale della musica storica. Il suono della canzone cantata da Faith mantiene un punto d’ascolto statico malgrado i cambi di punto di vista. Si potrebbe asserire comunque che detto punto d’ascolto sia quello di Lenny, che resta fermo.

Lo spazio sonoro del film è molto fitto, come pure quello visivo. Ciò è favorito anche dall’abbondante uso di musica la cui collocazione è spesso ambigua, poiché ha una firma spaziale e

una plausibilità che la fanno sembrare sovente storica malgrado la funzione sia alquanto discorsiva. A ogni modo una simile densità si affida allo stimolo visivo per la localizzazione. Ciò emerge chiaramente nella sequenza in cui Lenny torna a casa in auto dopo l'affare con Tick. La voce dell'uomo è sempre frontale e in PP a prescindere dal punto di vista (dentro o fuori dall'auto, di fronte o di lato o di dietro). I rumori e le voci da fuori sono presenti o assenti in modo arbitrario, e la musica dell'ultima parte della sequenza ha una parvenza più extra-storica non tanto per la firma spaziale – potrebbe provenire tuttora dall'autoradio – quanto per la maggiore qualità discorsiva. A questo approccio contribuiscono altre scelte di sconfinamento dei suoni. Ci sono infatti diversi ponti a precedere: un esempio lo si ha mentre termina la clip erotica fruita da Lenny nel bagno di un locale, allorché il suo dialogo della scena successiva si ascolta qualche istante prima dello stacco visivo, mentre la musica (presumibilmente storica, data l'ambientazione e la firma spaziale) fa da legante. Inoltre quando Lenny e Mace vanno a casa di lei dopo lo scontro coi poliziotti c'è una musica udibile in PP che poi viene ridotta di volume sullo stacco all'inquadratura dell'interno dell'auto; la modifica di firma spaziale potrebbe denotare il fatto che la musica proviene dall'autoradio, benché all'inizio della sequenza avesse la firma spaziale della consueta musica discorsiva. Il brano prosegue senza soluzione di continuità nella scena seguente all'interno della casa cambiando ancora firma spaziale, venendo ancorata al nuovo spazio, finché Mace non spegne il televisore interrompendola. In un certo senso quindi si è sfruttata la musica per anticipare un evento nello spazio successivo a quello denotato inizialmente. In sostanza comunque questo rafforza la funzione di legante che la musica adempie regolarmente nel film pur senza la giustificazione storica che riceve qui. Il sonoro è anche usato per denotare due spazi contemporaneamente: durante il racconto di Faith a Lenny e Mace le immagini alternano il presente e il passato dei ricordi di Faith, mentre il sonoro a volte presenta allo stesso tempo i suoni delle due unità spaziotemporali. Si noti inoltre che la riverberazione in diverse scene che costituiscono ricordi è un caso di firma spaziale puramente espressiva, poiché è un modo per denotare il fatto che i suoni sono mentali come le immagini – trattate anch'esse in modo anomalo poiché leggermente rallentate.

La gestione espressiva dei volumi raggiunge l'apice nell'ultima sequenza, quando l'agente Steckler cade a terra esanime: i suoni storici (e, gradualmente, la musica di commento) vengono silenziati per isolarci mentalmente dallo spazio come presumibilmente stanno facendo i personaggi, per poi essere gradualmente reintrodotti. In questo modo il sonoro offre un contributo vigoroso alla conclusione del climax e quindi, di lì a poco, del racconto.

## 6.9 – *Timecode* (USA 2000)

Prodotto da Mike Figgis e Annie Stewart per Red Mullet Productions.

Distribuzione: Screen Gems, Sony.

Scritto e diretto da Mike Figgis.

Suono in presa diretta: Eddie Bydalek, Robert Janiger.

Missaggio: Chris David, Leslie Shatz.

Musica: Mike Figgis e Anthony Marinelli.

Colonna sonora: 5.1.

Copia per l'analisi: DVD Sony Pictures Home Entertainment (USA 2000), audio Dolby Digital 5.1.

Questo film è sperimentale sotto vari aspetti: è improvvisato a partire da un soggetto; è girato contemporaneamente con quattro videocamere senza stacchi; le quattro riprese sono visibili allo stesso tempo sullo schermo quadripartito (indico lo spazio visivo di ciascun quadro dall'alto in senso orario con le lettere da A a D). Il sonoro 5.1 è utilizzato in ossequio a un simile approccio, perciò in maniera affatto peculiare. I vari canali non denotano lo spazio sonoro in maniera unitaria bensì frammentata, tuttavia con due rilevanti differenze rispetto allo spazio visivo. In primo luogo, mentre lo schermo presenta contemporaneamente quattro spazi o comunque quattro campi visivi per tutto il tempo, il sonoro spesso privilegia un ambiente piuttosto che un altro, che diventa figura, mentre gli altri se udibili restano sullo sfondo. In secondo luogo, mentre lo schermo mantiene la stessa segregazione tra i quattro campi visivi nella quadripartizione dello schermo, il sonoro non riserva ciascun canale in maniera costante a un singolo punto d'ascolto (cfr. Kerins 2011: 2075-2083). La separazione tra i canali favorisce certamente l'intelligibilità tuttavia non ci si può affidare alla direzione per comprendere in maniera univoca quale dei quattro è il campo visivo corrispondente, bensì occorre basarsi sul sincronismo (*bottom-up*) o su inferenze (*top-down*). Per esempio nei primi minuti del film ascoltiamo il suono di A dai due canali destri e quello di B dai due canali sinistri, ma in seguito le direzioni cambiano: durante l'alterco tra le due donne in auto in A il sonoro è prevalentemente a destra e al centro, per poi essere di nuovo spostato via dal centro e nel surround destro. Il sonoro di D all'arrivo della donna nell'edificio è anch'esso propagato dai canali destri come quello di A ma con l'aggiunta del canale centrale, il che favorisce la segregazione delle voci. Inoltre la nostra possibilità di scegliere il flusso sonoro cui dedicare attenzione endogena è manipolata dallo stile poiché i vari flussi non sono sempre tutti ugualmente a disposizione, dal momento che i livelli cambiano in base alle intenzioni narrative: per esempio la voce della paziente e della psicoterapeuta durante la seduta (B) non sono sempre udibili malgrado non vi siano

apparentemente micro-eventi che motivino tale scelta. Perciò la nostra attenzione esogena può essere catturata non solo dalla salienza di qualche suono nel campo sonoro ma anche dal rientro nel campo sonoro di un dato flusso: tutto ciò ci spinge presumibilmente a orientare lo sguardo alla ricerca del campo visivo corrispondente per comprendere meglio quanto sta accadendo. La musica di commento nel corso del film è missata invece come da prassi in tutti i canali (in quello centrale a livello più basso), anche con panoramiche (si noti il sassofono contralto della prima sequenza, che ruota intorno al *sweet spot*). Insomma *Timecode* è un chiaro esempio di film con sonoro altamente immersivo che però non permette di simularci presenti in uno spazio a 360° poiché le varie fonti che ci circondano non costituiscono uno spazio unitario.

## 6.10 – *Elephant (Id., USA 2003)*

Prodotto da Dany Wolf.

Distribuzione: HBO Films.

Regia: Gus Van Sant.

*Sound design*: Leslie Shatz.

Suono in presa diretta: Felix Andrew.

Microfonista: Neil Riha.

Montaggio del dialogo: David A. Cohen.

Musica preesistente di Ludwig van Beethoven, Hildegard Westerkam, ecc.

Colonna sonora: Dolby Digital 5.1.

Copia per l'analisi: BD Optimum Releasing (Regno Unito 2004), audio DTS-HD Master Audio 5.1.

Il sonoro di questo film sembra esser stato registrato interamente in presa diretta (salvo la musica di commento), con boom e radiomicrofoni: si riscontra sia un'accentuata firma spaziale, sia la compresenza e la frequente mobilità di voci e rumori nei tre canali frontali, sia il fatto che i canali surround, appena percepibili, presentano lo stesso contenuto degli omologhi frontali (sinistro/surround sinistro, destro/surround destro). L'impressione complessiva è quella di una fedeltà agli eventi rappresentati e una costante mobilità del punto d'ascolto, il che è affine con lo stile visivo. Tuttavia la prospettiva uditiva non sempre coincide con quella visiva: ciò in termini formali potrebbe essere discutibile, nondimeno ne risulta un effetto perturbante coerente col tono della narrazione e di alcune scelte recitative. Questo effetto è invero favorito anche da altri fattori sonori ovvero le scelte musicali (i due brani di Beethoven, eseguiti nella storia, e quelli extra-

storici) e più in generale dal fatto che le immagini e i suoni paiono spesso influenzate dagli stati psicologici dei personaggi piuttosto che da un approccio documentario (cfr. Jordan 2012).

I suoni off della prima scena offrono già una grande concretezza allo spazio sonoro, data la firma spaziale e il movimento laterale di voci e rumori. La stessa concretezza è propria delle due sequenze successive, pur senza coerenza prospettica. Le voci di John e del padre non sono ostacolate dal parabrezza malgrado il punto di vista sia dall'esterno dell'auto, e sono centrate. Le voci dei tre personaggi della sequenza seguente mantengono un piano ravvicinato (soprattutto quella del personaggio principale, Elias, sempre in PP) malgrado si allontanino dal punto di vista, e restano anch'esse centrate. Il rumore dei passi tra le foglie secche è un chiaro esempio di suono che denota un movimento in modo implicito, poiché non ci sono indizi acustici di uno spostamento né della fonte né del punto d'ascolto, ai quali sopperisce ovviamente il campo visivo.

Diversamente dalle suddette sequenze, in altri momenti del film il sonoro offre anche indizi direzionali, sebbene spesso senza coerenza prospettica. Ciò risulta evidente nelle tre presentazioni dell'incontro nel corridoio tra John e Elias in cui questi scatta una fotografia all'altro: in tutti e tre i casi le loro voci sono ben separate, senza che vi sia conformità al punto di vista né denotazione di uno specifico punto d'ascolto. Nel primo e nel secondo caso le loro voci provengono in buona parte dalla direzione relativa alla loro posizione sullo schermo (sinistra/destra e destra/sinistra), ma comunque con una spazializzazione eccessiva; nel terzo caso, focalizzato su Michelle che passa loro vicino, c'è la stessa spazializzazione del primo benché la prospettiva visiva sia analoga al secondo e si riscontra pure una incoerenza nella distanza, con un risultato particolarmente spaesante. Una circostanza di coerenza direzionale si ha invece nella scena della consegna del fucile ad Alex e Eric, in cui si ha anche una panoramica sonora insieme a quella visiva e un raccordo di posizione sonora tra la prima e la seconda inquadratura: si noti come l'audio del televisore è al centro, poi panoramica verso sinistra durante la panoramica visiva verso destra e infine viene ricordato al centro quando lo stacco ci riporta frontalmente al televisore.

Un punto d'ascolto eccentrico è riscontrabile anche nella sequenza del dibattito sull'omosessualità, preceduta dalla didascalia "Acadia". I/le partecipanti sono disposti/e in cerchio, e la macchina compie una panoramica verso destra inquadrando ciascuna persona. Il sonoro è fatto di dialoghi che si accavallano e che non permettono di orientarci, poiché le direzioni sono a volte chiare (la voce del professore da destra si sposta al centro e poi a sinistra), altre volte confuse; la firma spaziale cambia come a segnalare modifiche del punto d'ascolto che però non sono chiaramente riconducibili a uno specifico spostamento nell'ambiente. Ancora una volta lo spazio sonoro è alquanto concreto pur rendendo difficoltosa la localizzazione.

L'uso della musica merita attenzione perché pone un interessante problema spaziale. Nella sequenza presso il campo di football si ascolta l'adagio della sonata *Al chiaro di luna* di Beethoven, come se si trattasse di una tipica musica di commento, senza una firma spaziale coerente con lo spazio rappresentato. Questo evento è puramente discorsivo, ancorché si tratti dell'anticipazione di un evento storico (più avanti è Alex a suonare il pezzo in scena). All'adagio si aggiunge tuttavia un altro brano musicale, *Türen der Wahrnehmung* (Porte della percezione) di Hildegard Westerkamp. Si tratta di un'opera di *soundscape*, il che non la rende percepibile come musica nel senso tradizionale: i suoni che la costituiscono sembrano infatti denotare uno spazio che è solo in parte coerente con quello del racconto, con un risultato inquietante che pare alludere al malessere esistenziale dei personaggi (cfr. Jordan 2007, 2012).

## 6.11 – *Gravity* (*Id.*, Regno Unito/USA 2013)

Prodotto da Alfonso Cuarón e David Heyman per Esperanto Filmoj e Heyday Films.

Distribuzione: Warner Bros.

Regia: Alfonso Cuarón.

*Sound design*: Glenn Fremantle.

Suono in presa diretta: Chris Munro.

Microfonista: Steve Finn.

Supervisione al montaggio sonoro: Glenn Fremantle.

Missaggio: Skip Lievsay, Niv Adiri, Christopher Benstead.

Musica: Steven Price.

Colonna sonora: Dolby Atmos.

Copia per l'analisi: BD Warner Bros. (Italia 2014).

*Gravity* è stato realizzato in Dolby Atmos, ma purtroppo non ho avuto modo di ascoltarlo in questa versione. Pertanto mi limiterò qui a rilevare sinteticamente ciò che resta nel *downmix* in 5.1 degli aspetti più rilevanti dell'estrema spazializzazione acustica che caratterizza il film. L'impianto formale di *Gravity* mira evidentemente a una profonda immersione del/della percipiente in ambienti privi di forza di gravità quali lo spazio cosmico e le astronavi, riducendo l'importanza delle coordinate spaziali cui siamo abituati. Il montaggio analitico è ridotto, poiché si preferisce ricorrere a inquadrature lunghe che esplorano lo spazio. Questo permette ai suoni di mantenere la direzionalità evitando l'effetto ping-pong, presumibilmente molesto. In effetti i continui

spostamenti dei personaggi nello spazio sono denotati dallo spostamento direzionale delle loro voci, in modo costante e senza distinzione gerarchica tra i canali frontali e quelli surround (la resa della distanza è invece più approssimativa). Un aspetto curioso consiste nel fatto che nelle numerose scene in cui le voci sono radiotrasmesse la collocazione delle medesime rispetta la posizione del corpo del/della parlante nel campo visivo, come se il suono provenisse direttamente da lì, sebbene non possa essere così poiché le voci non si propagano nell'aria bensì attraverso apparecchi tecnologici. Perciò abbiamo a che fare con un punto d'ascolto inverosimile il quale però garantisce un memorabile effetto di presenza poiché si simula l'integrazione multisensoriale che avrebbe luogo qualora le voci non fossero radiotrasmesse, ossia la corrispondenza appunto tra la direzione delle voci e la posizione dei corpi.

L'inquadratura che passa dall'esterno all'interno e di nuovo all'esterno del casco della protagonista dà l'impressione di un passaggio da un punto di vista oggettivo a uno soggettivo (e ritorno). La sua voce è come al solito distorta dalla trasmissione radio all'esterno, mentre all'interno è priva di mediazione; inoltre qui proviene dai canali surround, il che è un segnale della sua presenza alle nostre spalle oppure della nostra condivisione del suo punto d'ascolto.

Ogni volta che una persona urta un oggetto possiamo ascoltare il suono dell'impatto, soprattutto attraverso il canale LFE. Ciò non costituisce un'infrazione del realismo del film che elimina qualsiasi suono dallo spazio cosmico bensì una resa delle vibrazioni che i personaggi avvertono a causa dell'urto. Infatti i suoni che noi udiamo sono talmente gravi da enfatizzare enormemente la loro qualità vibrotattile piuttosto che uditiva. Il sonoro permette dunque di simulare la percezione soggettiva corporale dei personaggi, mentre il punto d'ascolto resta quello inverosimile già descritto.

## **6.12 – *La casa (Evil Dead, USA 2013)***

Prodotto da Bruce Campbell, Sam Raimi, Robert G. Tapert per TriStar Pictures, FilmDistrict, Ghost House Pictures.

Distribuzione: Sony Pictures, TriStar Pictures.

Regia: Fede Alvarez.

*Sound design*: Jonathan Miller, Stephen Hunter Flick, Stuart Provine.

Suono in presa diretta: Myk Farmer.

Microfonista: Nikora Edwards, Sandy Wakefield.

Supervisione al montaggio sonoro: Jonathan Miller.

Missaggio: Jonathan Wales, Richard Kitting.

Musica: Roque Baños.

Colonna sonora: 5.1.

Copia per l'analisi: DVD Sony Pictures Home Entertainment (Italia 2013), audio Dolby Digital 5.1.

Questo film è rappresentativo delle norme di resa spaziale sonora del cinema di paura (thriller-horror), le quali costituiscono l'oggetto della presente analisi. Gli espedienti sonori per la paura cui il film ricorre sono dei veri e propri topoi del genere: in termini tipologici si tratta di rumori ambigui o perturbanti, urla, voci bestiali e aberranti e musica dissonante e "rumoristica"; in termini parametrici si tratta di contrasti di intensità (e specialmente il fragore improvviso, da cui scaturisce il riflesso di trasalimento acustico), l'uso di frequenze estreme (gravi o acute) e di timbri ruvidi. Seguendo Blumstein, Davitian e Kaye (2010) si può asserire che i suoni tipici di questo genere cinematografico sono analoghi non lineari di quelli che nel regno animale sono associati alla paura, confermando indirettamente la tesi di Linda Williams (1991) circa la non-linearità dell'horror in quanto "genere del corpo". La vitalità e la materialità dei corpi sottoposti a mutilazioni e trasformazioni repellenti come ci si aspetta dal genere è incrementata dal sonoro, che offre una conferma crossmodale della loro concretezza (suppongo inoltre che per realizzare detti rumori siano stati usati dei vegetali, come da prassi dei/delle rumoristi/e, il che spiegherebbe la loro efficace qualità organica). L'uso di piani ravvicinati nel messaggio di questi suoni, come pure delle frequenti urla dei personaggi, ci dà l'impressione di prossimità al nostro spazio personale, il che in un contesto emotivo negativo è un ulteriore motivo di paura. Un contributo decisivo all'impatto sensoriale è fornito dalla musica del discorso, eseguita da una grande orchestra sinfonica cui nel terzo atto si aggiunge il coro, e propagata ovviamente da tutti i 5.1 canali. Ci sono svariati momenti di dissonanza e di stilemi ansiogeni (uso "rumoristico" degli strumenti musicali, ritmi concitati, tremoli d'archi, ecc.), con qualche interessante corrispondenza crossmodale: per esempio mentre Mia fugge dal mostro verso la fine del film il ronzio della motosega è evocato da ottoni gravi mentre il sibilo della lama del coltello che taglia l'aria è reso con note acutissime dei flauti.

Come abbiamo visto gli indizi uditivi di localizzazione influiscono sulla risposta emotiva al film, poiché i suoni espansi e quelli propagati da dietro sono più spaventosi di quelli rispettivamente puntiformi e frontali (cfr. Ekman e Kajastila 2009). Infatti i suoni provenienti da fuori del campo visivo sono biologicamente più salienti, proprio per la funzione di allarme del sistema uditivo che diversamente dalla vista agisce a 360° (cfr. Asutay e Västfjäll 2015a). Inoltre la spazializzazione acustica intorno al pubblico incrementa l'attivazione del sistema nervoso simpatico, come se avvenisse una reale intrusione nel nostro spazio personale (cfr. Kobayashi, Ueno e Ise 2015). È interessante notare anche che l'acustica dell'ambiente e quindi i caratteri spaziali dei suoni possono

condizionare la risposta emotiva suscitata dall'ascolto, come dimostra l'*emoacustica* (cfr. Asutay et al. 2012). Tutte queste dinamiche hanno luogo nel *sound design* di questo film, caratterizzato per lo più dal supercampo. Infatti il 5.1 è usato regolarmente per un'espansione approssimativa ma efficace dell'ambiente, popolato spesso di suoni aberranti: l'ampliamento dello spazio permette di far percepire i suoni come più sgradevoli e minacciosi (Tajadura-Jiménez et al. 2010c). Quando invece i canali presentano una collocazione direzionale dei suoni, quindi una denotazione spaziale più precisa, il risultato è ancor più intenso perché abbiamo l'impressione della posizione o del movimento di un ente minaccioso in un'area esclusa dal nostro campo visivo. In questo modo si può rendere l'assedio dei personaggi nella baita in maniera letterale sui/sulle percipienti, rafforzando la simulazione incarnata della propria presenza nello spazio filmico. Inoltre il fatto che i canali surround vengono attivati maggiormente nelle sequenze paurose (come in questo film così nella norma del genere) è un indizio cognitivo che favorisce la nostra ansia. La presenza di un canale dedicato alle basse frequenze è anch'essa motivo di ulteriore sollecitazione sensoriale: le frequenze al di sotto dei 120 Hz sono meno direzionali di quelle più alte, perciò sono intrinsecamente immersive, e possono avere un impatto non solo uditivo ma anche vibrotattile, allorché sono avvertite dal corpo del/della percipiente (cfr. Kerins 2011; Takahashi 2011; Whittington 2014: 183), il che presumibilmente incrementa la paura dando l'impressione di maggiore prossimità allo spazio personale (cfr. Branje et al. 2014); per giunta i suoni gravi e profondi sono più allarmanti di quelli acuti, evocando il verso di un predatore (cfr. Parker e Heerema 2008: 2).

Significativamente il film ci introduce nel mondo narrativo e nella sua atmosfera negativa attraverso la colonna sonora. Le prime immagini narrative sono precedute da un suono profondo e altri rumori sottili e inquietanti (e persino un fragore sincronico con il logo di Ghost House Pictures). Poi una mosca (che cita il film originale, *La casa* [*The Evil Dead*, USA 1981]) si ode ronzare nei canali frontali e surround: un'enfasi così insolita su una fonte off apparentemente insignificante è perturbante. Inoltre l'uso di suoni non sincronistici è rilevante poiché un suono di cui non vediamo la fonte attira su di sé l'attenzione esogena spingendoci a orientare lo sguardo alla ricerca della fonte medesima, per permetterci di comprenderne visivamente la natura così da accertarci della sua innocuità oppure reagire qualora costituisse un pericolo. Dunque l'esclusione della fonte stessa dal nostro campo visivo ci mette automaticamente in uno stato di reattività, incrementando la nostra ansia.

La scena che segue è già indicativa dello stile sonoro pauroso del film. La prima apparizione fugace di un personaggio minaccioso è accompagnata da uno *stinger*<sup>56</sup> moderato. Il rumore dei passi del personaggio è udibile fuori campo nei canali surround, mettendoci in allarme come la ragazza. Nel momento in cui l'uomo la aggredisce c'è un forte *stinger* seguito da musica intensa, finché un altro *stinger* non conclude la scena quando la ragazza viene colpita sulla testa; perciò abbiamo il primo caso sia del ruolo dei canali surround sia della ricorrente interazione tra silenzio e suoni forti che provoca il riflesso di trasalimento acustico (cfr. Xu, Chia e Jin 2005). Qualcosa di affine avviene nella scena in cui Natalie viene catturata da Mia nella cantina: c'è un'alternanza tra l'addensamento e la rarefazione del sonoro, usando i suoni della storia (che fanno trasalire Natalie e noi con lei) e la musica discorsiva; alcuni suoni aberranti si muovono nei canali surround senza una chiara ragione di denotazione spaziale, incrementando l'angoscia.

### 6.13 – *Salvo* (Italia 2013)

Prodotto da Massimo Cristaldi e Fabrizio Mosca.

Distribuzione originale: Good Films.

Scritto e diretto da Fabio Grassadonia e Antonio Piazza.

Suono in presa diretta: Guillaume Sciamia.

Montaggio sonoro: Emanuela Di Giunta.

Effetti sonori: Studio 16 Sound, Italo Cameracanna, Riccardo Cameracanna.

Missaggio: Fabio Chiossi.

Colonna sonora: Dolby Digital 5.1.

Copia per l'analisi: DVD Koch Films (Italia 2014), audio DTS 5.1.

Se è vero che il sonoro surround evoca il cinema spettacolare e artificioso per cui è nato, *Salvo* dimostra mirabilmente come l'ultracampo possa essere usato in un contesto stilistico e drammaturgico improntato al realismo e alla sobrietà. Il sonoro è infatti molto direzionale come dimostra già la seconda sequenza, che presenta una sparatoria in strada: i rumori provengono da tutti i canali permettendo una localizzazione mimetica (anche nella distanza) piuttosto che limitandosi a una generica immersione. Ciò risulta ancor più evidente nella lunga sequenza presso la casa di Rita. La percezione audiospaziale è qui cruciale anche nella storia, poiché il suono è una guida alla localizzazione per Salvo e soprattutto per Rita la quale è ipovedente; l'ambientazione

---

<sup>56</sup> Si tratta di un suono sforzato della musica di commento, uno degli espedienti uditivi per eccellenza del cinema thriller-horror (cfr. Tompkins 2014).

della casa, nella penombra, e la scelta di piani ravvicinati rende noi stessi inclini a un ascolto più attento. Le inquadrature sono per lo più oggettive, ma il punto d'ascolto ha una forte parvenza di soggettività laddove i personaggi sono inquadrati di spalle (poiché condividiamo abbastanza la loro prospettiva). Un'eccezione alla direzionalità si ha con la musica storica (la canzone *Arriverà* proveniente dallo stereo): quando la fonte è fuori campo a destra mentre Salvo sale la scala la musica continua per qualche istante a provenire dal centro come quando la fonte era centrale. Anche poco prima nel controcampo su Salvo che osserva Rita la canzone resta centrale, ma qui l'impressione non è tanto di incoerenza prospettica quanto di soggettività, come se il punto d'ascolto fosse quello dell'uomo, o comunque di permanenza del punto d'ascolto precedente malgrado il cambio di punto di vista. Per il resto i suoni sono inseriti, panoramici e raccordati nei vari canali nel rispetto quasi assoluto del realismo spaziale. Si noti il movimento della canzone mentre Salvo è al piano superiore, e il modo in cui è raccordata nella posizione (insieme alla voce del giovane che porta la spesa a Rita) allo stacco in soggettiva di Salvo che guarda di sotto, oltre al verosimile cambio di firma spaziale. L'uso del suono off e la sua rilevanza per la cognizione spaziale è eclatante nel momento della lotta tra i due uomini. Dopo che Salvo spinge Rita a terra egli sale le scale correndo, seguito dall'altro uomo; i suoni si spostano nella direzione in cui essi si trovano, e con lo stacco all'inquadratura di Rita i suoni vengono raccordati nella posizione. Da questo momento seguiamo il percorso di Rita, inquadrata frontalmente, che si dirige verso il luogo di provenienza dei suoni, che noi continuiamo a sentire in coerenza col punto di vista. Il lungo piano ravvicinato sul volto della donna è un modo consueto di favorire l'empatia in chi guarda (cfr. Gallese e Guerra 2015: 216-217), mentre gli indizi veicolati dal sonoro ci permettono al contempo di percepire cosa accade nell'altro ambiente e di pensare per questo anche noi.

Una circostanza simile si verifica nella sequenza in cui Rita si sveglia nel nascondiglio dove Salvo l'ha rinchiusa: il rumore da lui provocato con la pala è raccordato nei vari canali a seconda della posizione relativa al punto di vista, cosicché mentre osserviamo da vicino il volto della donna che brancola nel buio possiamo localizzare nella direzione e nella distanza Salvo e quindi presumibilmente la via d'uscita verso cui Rita tenta di dirigersi. La stessa prassi è riscontrabile nella scena che precede la fuga di Salvo e Rita, mentre lei aspetta che lui uccida i persecutori.

La scrupolosità della resa sonora vale comunque più per i suoni off che per quelli in, come dimostrano i seguenti esempi. Durante l'inquadratura di Salvo e di un altro uomo per strada mentre il primo si dirige a casa la voce dell'altro uomo è in PP malgrado la distanza visiva non lo giustifichi. Inoltre nel dialogo tra Enzo e la donna nella stanza di Salvo le voci sono al centro malgrado i personaggi siano ai lati dell'inquadratura, alquanto distanziati (il rapporto d'aspetto è 2,39:1). Invece nella scena in cui Salvo discute col boss sullo sterrato le voci sono a una distanza

piuttosto coerente col punto di vista (MPP, PP, PPP), nondimeno esse mantengono il centro a prescindere dalla posizione laterale del boss.

## Conclusioni, limiti e futuri sviluppi

Questo lavoro si pone in termini innovativi piuttosto che conclusivi. Data l'originalità dell'approccio che mi sono impegnato a perseguire, sono consapevole che alcune delle questioni trattate ricevono meno attenzione e minore sviluppo rispetto ad altre. Mi sono preoccupato di stabilire una base di partenza per una doverosa integrazione dell'analisi del sonoro negli studi di stampo cognitivista. Perciò intendo proseguire la mia ricerca per arricchire quanto ho iniziato a definire, partendo dai numerosi materiali già esistenti che non ho potuto includere in questa tesi. Gli scopi che mi prefiggo sono principalmente due: in primo luogo quello di favorire la comprensione dei processi di cognizione audiospaziale per includerli negli studi di scienze umane dedicati alla rappresentazione e alla cognizione dello spazio anzitutto nel cinema e più in generale negli audiovisivi e in altre forme artistiche sonore (musica, radio, installazioni, ecc.); in secondo luogo quello di proporre alcuni temi di ricerca per le scienze sperimentali. A quest'ultimo proposito uno sviluppo immediato del presente lavoro potrebbe essere quello per il quale nel luglio del 2016 mi sono consultato coi proff. Vittorio Gallese e Michele Guerra presso l'Università di Parma, ovvero un esperimento di elettroencefalografia ad alta densità su soggetti esposti a uno stimolo audiovisivo prima mono o stereo e poi 5.1, per verificare la differenza nella risposta fisiologica.

Il limite maggiore di questa dissertazione potrebbe consistere in una incompleta padronanza dei dati delle scienze sperimentali cui ho attinto per formulare domande, dare forma ad alcune rilevazioni empiriche e trovare delle risposte. Perciò una prima verifica teorica e metodologica potrebbe venire proprio da chi si occupa di scienze cognitive ed è sensibile all'esperienza estetica. Miei colleghi e mie colleghe di scienze umane potrebbero d'altronde suggerire un modo diverso di concettualizzare le questioni che mi pongo a prescindere dalla loro eventuale conferma sperimentale, nonché ampliare la letteratura di riferimento includendo altre lingue rispetto a quelle da me considerate (italiano, inglese e marginalmente francese). Un altro limite concerne l'impossibilità di analizzare una colonna sonora realizzata con uno dei nuovi sistemi immersivi: per far ciò occorre una stazione di lavoro idonea e una copia adeguata del film, che non è detto che sia quella per l'home video (che può contenere audio tridimensionale), il che non mi è stato possibile. Malgrado i miei riferimenti a diversi modi di stile non ho sviluppato approfonditamente la questione, cosa che si potrebbe fare nell'immediato a partire dal metodo che ho elaborato e dalle mie rilevazioni, per giunta espandendo il campione di analisi rispetto a quello qui proposto in termini quantitativi nonché qualitativi, dato che sono esclusi per esempio film di culture diverse da quelle occidentali.

## Bibliografia e sitografia

- Allen, J. F. (1997). The Eight-Channel Advantage. *Boxoffice*, 133(1), 38–41.
- Altman, R. (1992a). Afterword: A Baker's Dozen Terms for Sound Analysis. In R. Altman (Ed.), *Sound Theory/Sound Practice* (pp. 249–253). New York: Routledge.
- Altman, R. (1992c). The Material Heterogeneity of Recorded Sound. In R. Altman (Ed.), *Sound theory, sound practice* (pp. 15–31). New York: Routledge.
- Altman, R. (1995). The Sound of Sound: A Brief History of the Reproduction of Sound in Movie Theatres. *Cineaste*, 21 (January).
- Altman, R. (1996). Deep-focus sound: Citizen Kane and the radio aesthetic. In R. Gottesman (Ed.), *Perspectives on Citizen Kane* (pp. 94–121). New York: G. K. Hall & Co.
- Altman, R. (Ed.). (1980). Cinema/Sound. *Yale French Studies*, (60).
- Altman, R. (Ed.). (1992b). *Sound Theory/Sound Practice*. New York: Routledge.
- Altman, R., & Handzo, S.. (1995). The Sound of Sound: A Brief History of Reproduction of Sound in Movie Theaters. *Cinéaste*, 21(1–2), 68–71.
- Altman, R., Jones, M., & Tatroe, S. (2000). Inventing the Cinema Soundtrack: Hollywood's Multiplane Sound System. In D. Neumeyer, C. Flinn, & J. Buhler (Eds.), *Music and Cinema* (pp. 339–359). Hanover & London: Wesleyan University Press.
- Altman, R. (1999). Film Sound – All Of It. *Iris*, (27), 31–48.
- Ambisonics. (2012, June 4). Contributed Audio. Retrieved October 25, 2016, from <https://ambisonic.info/audio.html>
- Anderson, J. D. (1996). *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Carbondale and Edwardsville: Southern Illinois University Press.
- Anonimo. (1985). Direct Sound: An Interview with Jean-Marie Straub and Danièle Huillet. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 150–153). New York: Columbia University Press.
- Anonimo. (1934). “Solid Music”: Three-Dimensional Sounds Created. *Everyday Science and Mechanics*, 213, 263.
- Anonimo. (1937). Stokowski Explains New Recording Device. *Cinema Progress*, 2(4), 8.
- Anonimo. (1938). Separate Tracks for Six Orchestra Units. *The Cine-Technician*, 3(14), 225.
- Anonimo. (1940). Fantasound by RCA as presented in Walt Disney's Fantasia. *International Projectionist*, 15(11), 3.

- Anonimo. (1957). 20th-Fox Adopts Small-Sprocket Magoptical. *International Projectionist*, 32(3), 21.
- Anonimo. (2012, February 15). Big Enough for Everest: Inside IMAX Sound. Retrieved October 24, 2016, from <http://www.creativeplanetnetwork.com/news/news-articles/big-enough-everest-inside-imax-sound/382615>
- Anonimo. (2013, May). Exhibitors sound off on Dolby.Atmos. *Film Journal International*, 116(5), 66–68.
- Archivio Luigi Nono. (n.d.). La fabbrica illuminata, per voce femminile e nastro magnetico. Retrieved October 23, 2016, from <http://www.luiginono.it/it/luigi-nono/opere/la-fabbrica-illuminata>
- Armstrong, K. L., & Cutting, J. E. (2016). Viewer Comprehension and Temporal Shifts in Hollywood Film. *Elephant & Castle*, (14), 5–25.
- Armstrong, K. L., & Cutting, J. E. (2017). Events, Movies, and Aging. *Journal of Applied Research in Memory and Cognition*, 6(2), 137–140.
- Arnheim, R. (1985). A New Lacoön: Artistic Composites and the Talking Film [1938]. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 112–115). New York: Columbia University Press.
- Asutay, E., & Västfjäll, D. (2015a). Attentional and emotional prioritization of the sounds occurring outside the visual field. *Emotion*, 15(3), 281–286. <https://doi.org/10.1037/emo0000045>
- [Asutay, E., & Västfjäll, D. \(2015\). Negative emotion provides cues for orienting auditory spatial attention. \*Frontiers in Psychology\*, 6\(article 618\). https://doi.org/10.3389/fpsyg.2015.00618](https://doi.org/10.3389/fpsyg.2015.00618)
- Asutay, E., Västfjäll, D., Tajadura-Jiménez, A., Genell, A., Bergman, P., & Kleiner, M. (2012). Emoacoustics: A study of the psychoacoustical and psychological dimensions of emotional sound design. *Journal of the Audio Engineering Society*, 60(1/2), 21–28.
- Atkinson, S. (2011). Surrounded by sound: The aesthetics of multichannel and hypersonic soundscapes and aural architectures. *The Soundtrack*, 4(1), 5–21. [https://doi.org/10.1386/st.4.1.5\\_1](https://doi.org/10.1386/st.4.1.5_1)
- Aumont, J., Bergala, A., Marie, M., & Vernet, M. (2009). *Estetica del film*. (D. Buzzolan, Trans.). Torino: Lindau.
- Avid Technology. (2014). Pro Tools® Reference Guide, Version 11.3. Retrieved from [http://akmedia.digidesign.com/support/docs/Pro\\_Tools\\_Reference\\_Guide\\_v11\\_3\\_83709.pdf](http://akmedia.digidesign.com/support/docs/Pro_Tools_Reference_Guide_v11_3_83709.pdf)
- Avis, P. (2001). Penderecki. *Orchestral and Choral Works*. In CD *Krzysztof Penderecki, Orchestral & Choral Works*. EMI.

- Baalman, M. A. J. (2010). Spatial Composition Techniques and Sound Spatialisation Technologies. *Organised Sound*, 15(03), 209–218. <https://doi.org/10.1017/S1355771810000245>
- Baker, G. W., & Smith, M. A. (1934). Some Practical Applications of Acoustic in Theaters. *Journal of the Society of Motion Pictures Engineers*, 22(2), 148–152.
- Balázs, B. (1985). Theory of the Film: Sound [1949]. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 116–125). New York: Columbia University Press.
- Bandirali, L. (2006). Benvenuti nel concerto del reale. *Close up. Storie Della Visione*, (18), 41–55.
- Barco. (2015, November 24). AuroMax®: Next generation Immersive Sound system (Rev. 2.0). Retrieved from [http://www.auro-3d.com/wp-content/uploads/documents/AuroMax\\_White\\_Paper\\_24112015.pdf](http://www.auro-3d.com/wp-content/uploads/documents/AuroMax_White_Paper_24112015.pdf)
- Barnett, S. (1980). Film Sound Techniques for Preserving Sound Perspective in Film Production Recording. *Recording Engineer/Producer*, 11(4), 54–61.
- Barreiro, D. L. (2010). Considerations on the Handling of Space in Multichannel Electroacoustic Works. *Organised Sound*, 15(03), 290–296. <https://doi.org/10.1017/S1355771810000294>
- Barrett, N. (2002). Spatio-musical composition strategies. *Organised Sound*, 7(3), 313–323. <https://doi.org/10.1017/S1355771802003114>
- Bartlett, B. (1983). Microphone techniques for realistic reproduction. *Modern Recording*, 11(9), 8–15.
- Bayle, F. (2007). Space, and more. *Organised Sound*, 12(03). <https://doi.org/10.1017/S1355771807001872>
- Beck, J. (2003). *A Quiet Revolution: Changes in American Film Sound Practices, 1967-1979* (PhD dissertation). University of Iowa, Iowa City.
- Beck, J., & Ament, V. T. (2015). The New Hollywood, 1981-1999. In Kalinak, Kathryn (Ed.), *Sound: Dialogue, Music, and Effects* (pp. 107–132). London & New York: I. B. Tauris.
- Beck, J., & Grajeda, T. (2008). *Lowering the Boom: Critical Studies in Film Sound*. Urbana and Chicago: University of Illinois Press.
- Begault, D. R. (2000). *3-D sound for virtual reality and multimedia*. Moffett Field: NASA. Retrieved from <http://ntrs.nasa.gov/search.jsp?R=20010044352>
- Begault, D. R. (2006). Preference versus Reference: Listeners as participants in sound reproduction. *Spatial Audio and Sensory Evaluation Techniques*, Guildford, UK. Retrieved from <https://iosr.uk/projects/ias/papers/Begault.pdf>
- Begault, D. R., & Erbe, T. (1994). Multichannel spatial auditory display for speech communications. *Journal of the Audio Engineering Society*, 42(10), 819–826.

- Belton, J. (1985). Technology and Aesthetics of Film Sound. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 63–72). New York: Columbia University Press.
- Belton, J. (1988). CinemaScope and Historical Methodology. *Cinema Journal*, 28(1), 22–44.  
<https://doi.org/10.2307/1225015>
- Belton, J. (1992). 1950s Magnetic Sound: The Frozen Revolution. In R. Altman (Ed.), *Sound Theory/Sound Practice* (pp. 154–167). New York: Routledge.
- Belton, J. (2002). Digital Cinema: A False Revolution. *October*, 100, 98–114.
- Belton, J. (2004). The Curved Screen. *Film History*, 16(3), 277–285.
- Benshoff, H. M. (Ed.). (2014). *A Companion to the Horror Film*. Chichester: John Wiley & Sons.
- Bent, I. (1990). *Analisi musicale*. (C. Annibaldi, Ed.). Torino: EDT.
- Berg, J., & Rumsey, F. (2000). Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction. In *Audio Engineering Society Convention 109*. Audio Engineering Society. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=9132>
- Biancorosso, G. (2009). The Harpist in the closet: Film music as epistemological joke. *Music and the Moving Image*, 2(3), 11–33.
- Blair, I. (1995, January 27). They Said It Couldn't Be Done, But "Wings" Proved 'em Wrong. *Chicago Tribune*. Retrieved from [http://articles.chicagotribune.com/1995-01-27/entertainment/9501270334\\_1\\_henri-guillaumet-french-director-jean-jacques-annaud-imax-3-d](http://articles.chicagotribune.com/1995-01-27/entertainment/9501270334_1_henri-guillaumet-french-director-jean-jacques-annaud-imax-3-d)
- Blake, L. (1984). *Film Sound Today: An Anthology of Articles from Recording Engineer/Producer*. Hollywood: Reveille Press.
- Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. Cambridge: MIT press.
- Blu-ray.com. [www.blu-ray.com](http://www.blu-ray.com)
- Blumlein, A. (1933, June 14). Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems. GB394325.
- Blumstein, D., Davitian, R., & Kaye, P. D. (2010). Do film soundtracks contain nonlinear analogues to influence emotion? *Biology Letters*, 6(6), 751–754.
- Bohn, W. C. (1954). Front Row Center and Thirty Feet Up: Let's be sane about "concert-hall realism." *High Fidelity Magazine*, 4(2), 37, 92–94.
- Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology: A Journal of Research in Music Cognition*, 13(1–2), 28–59.

- Boltz, M. G. (2001). Musical Soundtracks as a Schematic Influence on the Cognitive Processing of Filmed Events. *Music Perception*, 18(4), 427–454.
- Boltz, M. G. (2004). The cognitive processing of film and musical soundtracks. *Memory & Cognition*, 32(7), 1194–1205.
- Boltz, M. G., Ebendorf, B., & Field, B. (2009). Audiovisual interactions: The impact of visual information on music perception and memory. *Music Perception: An Interdisciplinary Journal*, 27(1), 43–59.
- Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory & Cognition*, 19(6), 593–606.
- Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory & Cognition*, 19(6), 593–606.
- Bonneau, P. (1931). The Talking Film. *Journal of the Society of Motion Pictures Engineers*, 16(4), 399–426.
- Bordwell, D. (1980). The musical analogy. *Yale French Studies*, (60), 141–156.
- Bordwell, D. (1985a). *Narration in the Fiction Film*. Madison: University of Wisconsin Press.
- Bordwell, D. (1985b). Space in the classical film. In D. Bordwell, J. Staiger, & K. Thompson, *The Classical Hollywood Cinema: Film Style & Mode of Production to 1960* (pp. 50–59). London: Routledge.
- Bordwell, D. (1985c). Widescreen processes and stereophonic sound. In D. Bordwell, J. Staiger, & K. Thompson, *The Classical Hollywood Cinema: Film Style & Mode of Production to 1960* (pp. 358–364). London: Routledge.
- Bordwell, D. (1989a). A case for cognitivism. *Iris*, (9), 11–40.
- Bordwell, D. (1989b). *Making Meaning: Inference and Rhetoric in the Interpretation of Cinema*. Cambridge: Harvard University Press.
- Bordwell, D. (1990). A Case for Cognitivism: Further Reflections. *Iris*, (11), 107–112.
- Bordwell, D. (2010). The Part-Time Cognitivist: A View from Film Studies. *Projections*, 4(2), 1–18. <https://doi.org/10.3167/proj.2010.040202>
- Bordwell, D., & Carroll, N. (Eds.). (1996). *Post-theory: reconstructing film studies*. Madison: University of Wisconsin Press.
- Bordwell, D., & Thompson, K. (2003). *Cinema come arte. Teoria e prassi del film*. (P. Bonini, Trans.). Milano: Il Castoro.
- Bordwell, D., Staiger, J., & Thompson, K. (1985). *The Classical Hollywood Cinema: Film Style & Mode of Production to 1960*. London: Routledge.
- Bragg, H. E., & Belton, J. (1988). The Development of CinemaScope. *Film History*, 359–371.

- Branigan, E. (1981). The Spectator and Film Space – Two Theories. *Screen*, 22(1), 55–78.
- Branigan, E. (2010). Soundtrack in Mind. *Projections*, 4(1), 41–67.  
<https://doi.org/10.3167/proj.2010.040104>
- Branje, C., Nespoil, G., Russo, F., & Fels, D. I. (2014). The Effect of Vibrotactile Stimulation on the Emotional Response to Horror Films. *ACM Computers in Entertainment*, 11(1), 1–13.  
<https://doi.org/10.1145/2543698.2543703>
- Bray, J. (n.d.). “Tommy” on Blu-ray – “Who’s” Your Daddy? Retrieved October 23, 2016, from  
<http://www.technofile.com/bluray/tommy.asp>
- Breebaart, J., & Faller, C. (2007). *Spatial Audio Processing: MPEG Surround and Other Applications*. Chichester: John Wiley & Sons.
- Bridgett, R. (2007). Post-production sound: a new production model for interactive media. *The Soundtrack*, 1(1), 29–39.
- Brown, R. S. (1994). *Overtones and undertones: Reading film music*. Berkeley and Los Angeles: University of California Press.
- Bruil, R. A. (2001). Wilma Cozart Fine is Mercury. Retrieved October 23, 2016, from  
<http://www.soundfountain.com/amb/mercury.html>
- Buckland, W. (2000). *The Cognitive Semiotics of Film*. Cambridge: Cambridge University Press.
- Buckland, W. (2003). Orientation in Film Space. A Cognitive Semiotic Approach. *Recherches En Communication*, 19(19), 87–102.
- Buettner, S. (2000, September). DTS-ES Discrete 6.1: The Ultimate Backward Compatible Surround Format. *Widescreen Review®: The Essential Home Theatre Resource™*, (41).
- Buhler, J., Flinn, C., & Neumeyer, D. (Eds.). (2000). *Music and cinema*. Hanover & London: Wesleyan University Press.
- Burch, N. (2000). *Prassi del cinema [1969]*. (C. Bragaglia & R. Provenzano, Trans.). Milano: Il Castoro.
- C. A. Productions. (2000, May 4). So Unique in Sound Recording. Retrieved October 25, 2016, from  
<http://www.cineraadventure.com/theworks-sound.htm>
- Calore, M. (2009, May 12). May 12, 1967: Pink Floyd Astounds with ‘Sound in the Round.’ Retrieved October 23, 2016, from  
[https://www.wired.com/2009/05/dayintech\\_0512/](https://www.wired.com/2009/05/dayintech_0512/)
- Camper, F. (1985). Sound and Silence in Narrative and Nonnarrative Cinema. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 369–381). New York: Columbia University Press.

- Candan, A., Cutting, J. E., & DeLong, J. E. (2015). RSVP at the movies: dynamic images are remembered better than static images when resources are limited. *Visual Cognition*, 23(9–10), 1205–1216.
- Cano, C. (2002). *La musica nel cinema: musica, immagine, racconto*. Roma: Gremese Editore.
- Cano, C., & Cremonini, G. (1995). *Cinema e musica. Il racconto per sovrapposizioni* (2nd ed.). Firenze: Vallecchi.
- Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia*, 47(4), 1045–1052. <https://doi.org/10.1016/j.neuropsychologia.2008.11.003>
- Carocci, E. (2012). *Attraverso le immagini. Tre saggi sull'emozione cinematografica*. Roma: Bulzoni.
- Carr, R. E., & Hayes, R. M. (1988). *Wide Screen Movies: A History and Filmography of Wide Gauge Filmmaking*. Jefferson (NC): McFarland.
- Carroll, N. (1988). *Mystifying movies: fads & fallacies in contemporary film theory*. New York: Columbia University Press.
- Carroll, N. (1996). Prospects for Film Theory: A Personal Assessment. In D. Bordwell & N. Carroll (Eds.), *Post-Theory: Reconstructing Film Studies* (pp. 37–68). Madison: University of Wisconsin Press.
- Casetti, F., & di Chio, F. (2009). *L'analisi del film*. Milano: Bompiani.
- Cass, J. L. (1930). The Illusion of Sound and Picture. *Journal of the Society of Motion Pictures Engineers*, 14(3), 323–326.
- Cecchi, A. (2010). Diegetico vs. extradiegetico: revisione critica di un'opposizione concettuale in vista di una teoria dell'audiovisione. *Worlds of Audiovision*. Retrieved from [http://www-5.unipv.it/wav/pdf/WAV\\_Cecchi\\_2010\\_ita.pdf](http://www-5.unipv.it/wav/pdf/WAV_Cecchi_2010_ita.pdf)
- Chandrasekaran, C. (2017). Computational principles and models of multisensory integration. *Current Opinion in Neurobiology*, 43, 25–34. <https://doi.org/10.1016/j.conb.2016.11.002>
- Chapman, D. (2017). Context-based Sound and the Ecological Theory of Perception. *Organised Sound*, 22(01), 42–50. <https://doi.org/10.1017/S1355771816000327>
- Chatman, S. (2010). *Storia e discorso. La struttura narrativa nel romanzo e nel film* [1978]. (E. Graziosi, Trans.). Milano: Il Saggiatore.
- Chattah, J. (2008). Conceptual Integration and Film Music Analysis. *Semiotics*, 21, 772–783.
- Chattah, J. (2015). Film music as embodiment. In M. Coëgnarts & P. Kravanja (Eds.), *Embodied cognition and cinema* (pp. 81–114). Leuven: Leuven University Press.

- Cherry, E. C., & Taylor, W. K. (1954). Some Further Experiments upon the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, 26(4), 554–559.
- Chion, M. (1999). *The voice in cinema* [1982]. (C. Gorbman, Trans.). New York: Columbia University Press.
- Chion, M. (2001). *L'audiovisione: suono e immagine nel cinema* [1994]. (D. Buzzolan, Trans.). Torino: Lindau.
- Chion, M. (2009). *Film, a Sound Art* [2003]. (C. Gorbman, Trans.). New York: Columbia University Press.
- Choisel, S., & Wickelmaier, F. (2007). Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. *The Journal of the Acoustical Society of America*, 121(1), 388. <https://doi.org/10.1121/1.2385043>
- Cinema Products Division (Optical Radiation Corporation). (1991, September). Opuscolo senza titolo. Retrieved from <https://math.berkeley.edu/~berlek/pubs/cds.pdf>
- Cinema Technology. (2014). Loudspeaker technologies for immersive sound in cinemas. *Cinema Technology*, 27(3).
- Claypool, B. (2014). Setting the Record Straight. *Film Journal International*, 117(4), 82–86.
- Claypool, B., Van Baelen, W., & Van Daele, B. (2012). Auro 11.1 versus object-based sound in 3D: All aspects compared. Barco. Retrieved from [http://www.barco.com/securedownloads/cd/MarketingKits/3d-sound/White%20papers/Auro%2011.1\\_versus\\_objectbased\\_sound\\_in\\_3D.pdf](http://www.barco.com/securedownloads/cd/MarketingKits/3d-sound/White%20papers/Auro%2011.1_versus_objectbased_sound_in_3D.pdf)
- Coate, M. (2015, December 7). 'Edward Scissorhands': The 70mm and CDS Presentations. Retrieved October 24, 2016, from <http://www.in70mm.com/news/2015/hands/index.htm>
- Coate, M., & Hauerslev, T. (Eds.). (2015, December 31). Presented in 70mm and six track magnetic Dolby Stereo. Retrieved from <http://in70mm.com/library/process/dolby/index.htm>
- Coëgnarts, M. (2015). *Embodied Cognition and Cinema: The Sensory-Motor Grounding of Abstract Meaning in Film*. Universiteit Antwerpen (Belgium).
- Coëgnarts, M., & Kravanja, P. (2012). Embodied visual meaning: Image schemas in film. *Projections*, 6(2), 84–101.
- Coëgnarts, M., & Kravanja, P. (2014). A study in cinematic subjectivity: Metaphors of perception in film. *Metaphor and the Social World*, 4(2), 149–173.
- Coëgnarts, M., & Kravanja, P. (2015). With the past in front of the character: Evidence for spatial-temporal metaphors in cinema. *Metaphor and Symbol*, 30(3), 218–239.

- Coëgnarts, M., & Kravanja, P. (2016). Perceiving emotional causality in film: a conceptual and formal analysis. *New Review of Film and Television Studies*, *14*(4), 440–466.
- Coëgnarts, M., Kiss, M., Kravanja, P., & Willemsen, S. (2016). Seeing Yourself in the Past: The Role of Situational (Dis)Continuity and Conceptual Metaphor in the Understanding of Complex Cases of Character Perception. *Projections*, *10*(1), 114–138.
- Cohen, A. B. (1954). Reflections on Having Two Ears. *High Fidelity Magazine*, *4*(6), 28–31, 78–83.
- Cohen, A. J. (1990). Understanding musical soundtracks. *Empirical Studies of the Arts*, *8*(2), 111–124.
- Cohen, A. J. (1993). Associationism and musical soundtrack phenomena. *Contemporary Music Review*, *9*(1–2), 163–178. <https://doi.org/10.1080/07494469300640421>
- Cohen, A. J. (1998). The functions of music in multimedia: A cognitive approach (pp. 13–20). Presented at the Proceedings of the Fifth International Conference on Music Perception and Cognition, Seoul.
- Cohen, A. J. (2001). Music as a source of emotion in film. In *Music and emotion: Theory and research* (pp. 249–272). New York: Oxford University Press.
- Cohen, A. J. (2002). Music cognition and the cognitive psychology of film structure. *Canadian Psychology/Psychologie Canadienne*, *43*(4), 215–232.
- Cohen, A. J. (2005). How music influences the interpretation of film and video. In *Perspectives in Systematic Musicology* (pp. 15–36). Los Angeles: Dept. of Ethnomusicology, University of California.
- Cohen, A. J. (2013). Film Music and the Unfolding Narrative. In M. A. Arbib (Ed.), *Language, Music, and the Brain. A Mysterious Relationship* (pp. 173–201). Cambridge: MIT Press.
- Cohen, A. J. (2015). Congruence-Association Model and Experiments in Film Music: Toward Interdisciplinary Collaboration. *Music and the Moving Image*, *8*(2), 5–24.
- Cohen, A. S., Rosen, R. C., & Goldstein, L. (1985). EEG hemispheric asymmetry during sexual arousal: psychophysiological patterns in responsive, unresponsive, and dysfunctional men. *Journal of Abnormal Psychology*, *94*(4), 580.
- Coleman, D., Romao, T., Villamin, C., Sinnett, S., Jakobsen, T., & Kingstone, A. (2013). Finding Meaning in All the Right Places: A Novel Measurement of Dramatic Structure in Film and Television Narratives. *Projections*, *7*(2). <https://doi.org/10.3167/proj.2013.070205>
- Collins, K. (2008). *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*. Cambridge: MIT Press.

- Collins, K. (2011). Making gamers cry: mirror neurons and embodied interaction with game sound. In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound* (pp. 39–46). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2095673>
- Collins, K. (2013). Sonic Subjectivity and Auditory Perspective in *Ratatouille*. *Animation*, 8(3), 283–299.
- Copeland, P. (2008). *Manual of analogue sound restoration techniques*. London: British Library.
- Corelli, S., Felici, F., & Martinelli, G. (2006). *Elementi di cinematografia sonora*. Roma: Lambda.
- Costabile, K. A., & Terman, A. W. (2013). Effects of Film Music on Psychological Transportation and Narrative Persuasion. *Basic and Applied Social Psychology*, 35(3), 316–324. <https://doi.org/10.1080/01973533.2013.785398>
- Cribbs, B., & McCrigler, L. (2014, September). The spectrum of immersive sound. *Film Journal International*, 117(9), 62–65.
- Cullen, B., Galperin, D., Collins, K., Hogue, A., & Kapralos, B. (2013). The effects of 5.1 sound presentations on the perception of stereoscopic imagery in video games. In *Stereoscopic Displays and Applications XXIV* (Vol. 8648, pp. 1–13). Burlingame, CA. <https://doi.org/10.1117/12.2003136>
- Cullen, B., Galperin, D., Collins, K., Kapralos, B., & Hogue, A. (2012). The effects of audio on depth perception in S3D games. In *Proceedings of the 7th audio mostly conference: A conference on interaction with sound – Corfu, Greece — September 26 - 28, 2012* (pp. 32–39). New York: ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2371461>
- Cumbow, R. C. (2001). In *2001: A Space Odyssey*, limited edition. Warner Home Video, USA.
- Currie, G. (1995). *Image and mind: Film, philosophy and cognitive science*. Cambridge: Cambridge University Press.
- Currie, G. (1999). Cognitivism. In T. Miller (Ed.), *A Companion to Film Theory* (pp. 105–122). Oxford: Blackwell Publishers.
- Curtis, J. (n.d.). Quadraphonic Formats. Retrieved October 23, 2016, from <http://www.obsoletemedia.org/audio/quadraphonic-formats/>
- Cutting, J. E. (1986). The shape and psychophysics of cinematic space. *Behavior Research Methods*, 18(6), 551–558.
- Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3), 323.
- Cutting, J. E. (2014). Event segmentation and seven types of narrative discontinuity in popular movies. *Acta Psychologica*, 149, 69–77.

- Cutting, J. E. (2015). The framing of characters in popular movies. *Art & Perception*, 3(2), 191–212.
- Cutting, J. E. (2016a). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review*, 23(6), 1713–1743.
- Cutting, J. E. (2016b). The evolution of pace in popular movies. *Cognitive Research: Principles and Implications*, 1(1), 30.
- Cutting, J. E., & Armstrong, K. L. (2016). Facial expression, size, and clutter: Inferences from movie structure to emotion judgments and back. *Attention, Perception, & Psychophysics*, 78(3), 891–901.
- Cutting, J. E., & Candan, A. (2013). Movies, evolution, and mind: Fragmentation and continuity. *The Evolutionary Review*, 4(3), 25–35.
- Cutting, J. E., & Candan, A. (2015). Shot durations, shot classes, and the increased pace of popular movies. *Projections*, 9(2), 40–62.
- Cutting, J. E., Brunick, K. L., & Candan, A. (2012). Perceiving event dynamics and parsing Hollywood films. *Journal of Experimental Psychology: Human Perception and Performance*, 38(6), 1476.
- Cutting, J. E., Brunick, K. L., & DeLong, J. E. (2011). How act structure sculpts shot lengths and shot transitions in Hollywood film. *Projections*, 5(1), 1–16.
- Cutting, J. E., Brunick, K. L., & DeLong, J. E. (2011). The changing poetics of the dissolve in Hollywood film. *Empirical Studies of the Arts*, 29(2), 149–169.
- Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, faster, darker: Changes in Hollywood film over 75 years. *I-Perception*, 2(6), 569–576.
- Cutting, J. E., DeLong, J. E., & Brunick, K. L. (2011). Visual activity in Hollywood film: 1935 to 2005 and beyond. *Psychology of Aesthetics, Creativity, and the Arts*, 5(2), 115.
- Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film. *Psychological Science*, 21(3), 432–439.
- Cutting, J. E., Iricinschi, C., & Brunick, K. L. (2013). Mapping narrative space in Hollywood film. *Projections*, 7(2), 80–107.
- Cutting, J., & Iricinschi, C. (2015). Re-Presentations of Space in Hollywood Movies: An Event-Indexing Analysis. *Cognitive Science*, 39(2), 434–456.
- Czyzewski, A., Kostek, B., Ody, P., & Zielinski, S. (2000). Determining influence of visual cues on the perception of surround sound using soft computing. In *International Conference on Rough Sets and Current Trends in Computing* (pp. 545–552). Berlin: Springer. Retrieved from [http://link.springer.com/chapter/10.1007/3-540-45554-X\\_68](http://link.springer.com/chapter/10.1007/3-540-45554-X_68)

- D'Aloia, A. (2013). *La vertigine e il volo. L'esperienza filmica fra estetica e neuroscienze cognitive*. Roma: Fondazione Ente dello Spettacolo.
- D'Aloia, A., & Eugeni, R. (2016). Connecting (and dis-connecting) events. A neurofilmological approach to narrative time. In *in/between: cultures of connectivity, The NECS 2016 Conference*. Potsdam, Germany. Retrieved from [www.academia.edu/27806086/Connecting\\_and\\_dis-connecting\\_events.\\_A\\_neurofilmological\\_approach\\_to\\_narrative\\_time](http://www.academia.edu/27806086/Connecting_and_dis-connecting_events._A_neurofilmological_approach_to_narrative_time)
- D'Aloia, A., & Eugeni, R. (Eds.). (2014). Neurofilmology. Audiovisual Studies and the Challenge of Neurosciences. *Cinéma & Cie*, 14(22/23).
- De Vincenti, G., & Carocci, E. (Eds.). (2012). *Il cinema e le emozioni. Estetica, espressione, esperienza*. Roma: Fondazione Ente dello Spettacolo.
- DeLong, J. E., Brunick, K. L., & Cutting, J. E. (2014). Film Through the Human Visual System: Finding Patterns and Limits. In J. C. Kaufman & D. K. Simonton (Eds.), *The Social Science of Cinema* (pp. 123–137). New York: Oxford University Press.
- Di Donato, M. (2004a). Lo sviluppo tecnologico della cinematografia sonora dalle origini all'avvento del digitale. In M. Di Donato (Ed.), *L'occhio che ascolta: studi sui rapporti suono-immagine nella forma cinematografica* (pp. 11–43). Roma: Lithos.
- Di Donato, M. (2006). *La spazializzazione acustica nel cinema contemporaneo. Tecnica, linguaggio, modelli di analisi*. Roma: Onyx.
- Di Donato, M. (2014a). Messa in scena sonora e punto di ascolto. In M. Di Donato & V. Valente (Eds.), *Ascoltare il cinema. Studi sul suono nel film* (pp. 13–31). Roma: Bulzoni.
- Di Donato, M. (2014b). Tridimensionalità e immersione: il suono filmico all'epoca del surround. In *Ascoltare il cinema. Studi sul suono nel film* (pp. 47–58). Roma: Bulzoni.
- Di Donato, M., & Valente, V. (Eds.). (2014). *Ascoltare il cinema. Studi sul suono nel film*. Roma: Bulzoni.
- Dienstfrey, E. (2016). The Myth of the Speakers: A Critical Reexamination of Dolby History. *Film History*, 28(1), 167–193. <https://doi.org/10.2979/filmhistory.28.1.06>
- Dietz, M. J., Friston, K. J., Mattingley, J. B., Roepstorff, A., & Garrido, M. I. (2014). Effective Connectivity Reveals Right-Hemisphere Dominance in Audiospatial Perception: Implications for Models of Spatial Neglect. *Journal of Neuroscience*, 34(14), 5003–5011. <https://doi.org/10.1523/JNEUROSCI.3765-13.2014>
- Digital Cinema Initiatives, LLC. (2012, October 10). Digital Cinema System Specification: Version 1.2 with Errata as of 30 August 2012 Incorporated. Retrieved from [http://dcimovies.com/specification/DCI\\_DCSS\\_v12\\_with\\_errata\\_2012-1010.pdf](http://dcimovies.com/specification/DCI_DCSS_v12_with_errata_2012-1010.pdf)

- Digital Theater Systems, Inc. (1999a, August). DTS-6 /-6D System Testing and Troubleshooting.
- Digital Theater Systems, Inc. (1999b, November 12). DTS Post Production Guide. (white paper TM-E229). Retrieved from <http://www.film-tech.com/warehouse/wareview.php?id=605&category=2>
- Digital Theater Systems, Inc. (2003a). XD10 Cinema Media Player: Installation and Operating Guide (Version 1.0).
- Digital Theater Systems, Inc. (2003b, June). DTS XD10 Cinema Media Player: Installation and Operating Guide (v. 1.0). Retrieved from <http://www.film-tech.com/warehouse/manuals/DTSXD10.pdf>
- Disclord. (2011, April 17). post #2 in “DAMNATION ALLEY: Sound 360.” Retrieved October 26, 2016, from <http://www.quadrasonicquad.com/forums/showthread.php?14402-DAMNATION-ALLEY-Sound-360>
- Doane, M. A. (1985a). Ideology and the Practice of Sound Editing and Mixing. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 54–62). New York: Columbia University Press.
- Doane, M. A. (1985b). The Voice in the Cinema: The Articulation of Body and Space. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 172–176). New York: Columbia University Press.
- Dockwray, R., & Collins, K. (2015). A Symphony of Sound: Surround Sound in Formula One Racing Games. In P. Théberge, K. Devine, & T. Everett (Eds.), *Living Stereo: Histories and Cultures of Multichannel Sound* (pp. 247–266). New York: Bloomsbury.
- Dolby Digital 5.1. (n.d.). Retrieved October 24, 2016, from <http://www.dolby.com/us/en/professional/technology/cinema/dolby-digital.html>
- Dolby Laboratories Inc. (1997). Dolby Digital: The sound of the future – here today. Retrieved from <http://www.film-tech.com/warehouse/wareview.php?id=1013&category=2>
- Dolby Laboratories Inc. (1998). A Chronology of Dolby Laboratories: May 1968 – May 1998. Retrieved from <http://www.film-tech.com/warehouse/wareview.php?id=1011&category=2>
- Dolby Laboratories Inc. (2000). Dolby Digital Professional Encoding Guidelines [Issue 1]. white paper S00/12972. Retrieved from <http://www.beussery.com/pdf/beussery.dolby.pdf>
- Dolby Laboratories Inc. (2014). Dolby® Atmos® Next-Generation Audio for Cinema (Issue 3). Retrieved from <http://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-next-generation-audio-for-cinema-white-paper.pdf>
- Dolby Laboratories Inc. (2015). Dolby Atmos® Enabled Speaker Technology.
- Dolby Laboratories Inc. (2016). Dolby Atmos® for the Home Theater.

- Dolby Pro Logic II. (n.d.). Retrieved October 24, 2016, from <http://www.dolby.com/us/en/technologies/dolby-pro-logic-ii.html>
- Dolby Pro Logic IIX. (n.d.). Retrieved October 24, 2016, from <http://www.dolby.com/us/en/technologies/dolby-pro-logic-iix.html>
- Dolby Pro Logic IIZ. (n.d.). Retrieved October 24, 2016, from <http://www.dolby.com/us/en/technologies/dolby-pro-logic-iiz.html>
- Dolby Surround 7.1. (n.d.). Retrieved August 11, 2016, from <http://www.dolby.com/us/en/technologies/dolby-surround-7-1.html>
- Donnelly, K. (2005). *The Spectre of Sound: Music in Film and Television*. London: BFI.
- Dorange, A. (2015, January 16). Kinopanorama around the World. Retrieved October 23, 2016, from <http://www.in70mm.com/news/2015/kinopanorama/index.htm>
- Dorren, L. (1973, January 2). Compatible four channel FM system. *US 3708623*. Retrieved from <https://www.google.com/patents/US3708623>
- Dreher, C. (1931a). Microphone Concentrators in Picture Production. *Journal of the Society of Motion Pictures Engineers*, 16(1), 23–30.
- Dreher, C. (1931b). Recording, Re-Recording, and Editing of Sound. *Journal of the Society of Motion Pictures Engineers*, 16(6), 756–765.
- Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, 381(6577), 66–68.
- DTS, Inc. (2006, November). DTS-HD Audio: Consumer White Paper for Blu-ray Disc and HD DVD Applications. Retrieved from [http://www.opusproductions.com/pdfs/DTS\\_HD\\_WhitePaper.pdf](http://www.opusproductions.com/pdfs/DTS_HD_WhitePaper.pdf)
- DTS, Inc. (2014). Building Mobile Games with DTS Headphone:X: Mobile Game Developer Overview. Retrieved from <https://resources.dts.com/bp/#/assets>
- DTS, Inc. (2015a). DTS Audio Solutions for Home A/V Products. Retrieved from <https://resources.dts.com/bp/#/assets>
- DTS, Inc. (2015b). Play-Fi. Retrieved from <https://resources.dts.com/bp/#/assets>
- DTS, Inc. (2015c). DTS Technology for Automotive. Retrieved from <https://resources.dts.com/bp/#/assets>
- DTS, Inc. (2016a). Premium Cinema Experience With DTS:X. Retrieved from <https://resources.dts.com/bp/#/assets>
- DTS, Inc. (2016b). DTS Sound™ Audio Enhancement Suite. Retrieved from <https://resources.dts.com/bp/#/assets>
- du Moncel, T. (1887). *Le téléphone* [qui tratto da Lange 2002]. Librairie Hachette.

- Dudai, Y. (2008). Enslaving central executives: toward a brain theory of cinema. *Projections*, 2(2), 21–42.
- DVDBeaver. [www.dvdbeaver.com](http://www.dvdbeaver.com)
- Ejzenštejn, S. M., Pudovkin, V., & Aleksandrov, G. (1964). Il futuro del sonoro. Dichiarazione [1928]. In P. Gobetti (Ed.), *Forma e tecnica del film e lezioni di regia* (pp. 523–524). Torino: Einaudi.
- Ekman, I., & Kajastila, R. (2009). Localization Cues Affect Emotional Judgments – Results from a User Study on Scary Sound. In *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=15177>
- Elsaesser, T., & Hagener, M. (2009). *Teoria del film. Un'introduzione* [2007]. (F. De Colle & R. Censi, Trans.). Torino: Einaudi.
- Elvemo, J.-M. (2013). Spatial perception and diegesis in multi-channel surround cinema. *The New Soundtrack*, 3(1), 31–44. <https://doi.org/10.3366/sound.2013.0034>
- Empire. (1997, December 19). IMAX Cinema Opens in London. *Empire*. Retrieved from <http://www.empireonline.com/movies/news/imax-cinema-opens-london/>
- Eschrich, S., Münte, T. F., & Altenmüller, E. O. (2008). Unforgettable film music: The role of emotion in episodic long-term memory for music. *BMC Neuroscience*, 9(1), 48. <https://doi.org/10.1186/1471-2202-9-48>
- Facebook. (2015). “San Siro” di Yuri Ancarani [event]. Retrieved October 31, 2016, from <https://www.facebook.com/events/108824202782340/>.
- Fahlenbrach, K. (2008). Emotions in sound: Audiovisual metaphors in the sound design of narrative films. *Projections*, 2(2).
- Ferretti, A., Caulo, M., Del Gratta, C., Di Matteo, R., Merla, A., Montorsi, F., Pizzella, V., Pompa, P., Rigatti, P., & Rossini, P. M. (2005). Dynamics of male sexual arousal: distinct components of brain activation revealed by fMRI. *Neuroimage*, 26(4), 1086–1096.
- Fletcher, H. (1934a). Auditory Perspective — Basic Requirements. *Electrical Engineering*, 53(1), 9–11.
- Fletcher, H. (1934b). Transmission and Reproduction of Speech and Music in Auditory Perspective. *Journal of the Society of Motion Pictures Engineers*, 22(5), 314–329.
- Fletcher, H. (1940). Stereophonic Reproduction from Film. *Journal of the Society of Motion Pictures Engineers*, 34(6), 606–613.
- Fletcher, H. (1941). The Stereophonic Sound-Film System—General Theory. *Journal of the Society of Motion Pictures Engineers*, 37(10), 331–352.

- Francis, T. (1984). A Toast to Don Quixote. *Box Office*, 120(12), 109–111.
- Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, 11(5), 197–203. <https://doi.org/10.1016/j.tics.2007.02.003>
- Freeman, E., & Driver, J. (2008). Direction of Visual Apparent Motion Driven Solely by Timing of a Static Sound. *Current Biology*, 18(16), 1262–1266. <https://doi.org/10.1016/j.cub.2008.07.066>
- Fuchs, A. (2014, August 15). Earth-shattering: FJI salutes the 40th anniversary of Sensurround's quakes and battles. *Film Journal International*. Retrieved from <http://www.filmjournal.com/node/9780>
- Furman, O., Dorfman, N., Hasson, U., Davachi, L., & Dudai, Y. (2007). They saw a movie: long-term memory for an extended audiovisual narrative. *Learning & Memory*, 14(6), 457–467.
- Furmanek, B. (n.d.). The first year of stereophonic motion pictures. Retrieved October 31, 2016, from <http://www.3dfilmarchive.com/the-first-year-of-stereophonic-sound>
- Furmanek, B., & Kintz, G. (n.d.). Cacophony in 3-D; 'House of Wax' Puts Stereoscopy in a Noisy Chamber of Horrors. Retrieved October 23, 2016, from <http://www.3dfilmarchive.com/House-of-Wax>
- Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Fear influences perceived reaching to targets in audition, but not vision. *Evolution and Human Behavior*, 34(1), 49–54. <https://doi.org/10.1016/j.evolhumbehav.2012.09.002>
- Gallese, V. (2001). The “shared manifold” hypothesis. From mirror neurons to empathy. *Journal of Consciousness Studies*, 8(5–6), 33–50.
- Gallese, V., & Guerra, M. (2012). Embodying movies: Embodied simulation and film studies. *Cinema: Journal of Philosophy and the Moving Image*, 3, 183–210.
- Gallese, V., & Guerra, M. (2013). Film, corpo, cervello: prospettive naturalistiche per la teoria del film. *Fata Morgana*, 20, 77–91.
- Gallese, V., & Guerra, M. (2014). Corpo a corpo. Simulazione incarnata e naturalizzazione dell'esperienza filmica. *Psicobiettivo*, 34(1), 156–177.
- Gallese, V., & Guerra, M. (2015). *Lo schermo empatico. Cinema e neuroscienze*. Milano: Raffaello Cortina Editore.
- Gallico, C. (1991). *L'età dell'Umanesimo e del Rinascimento*. Torino: EDT.
- Garity, W. E., & Hawkins, J. N. A. (1941). “Fantasound”: a Technologic Epoch. *International Projectionist*, 16(9), 16–18.
- Garity, W. E., & Hawkins, J. N. A. (1942, October 13). Sound Reproducing System. *US2298618*.

- Garner, T., Grimshaw, M., & Nabi, D. A. (2010). A preliminary experiment to assess the fear value of preselected sound parameters in a survival horror game. In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound* (p. 10). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1859809>
- Gaudreault, A., & Jost, F. (1990). *Le Récit cinématographique*. Paris: Éditions Nathan.
- George, S., Zielinski, S., & Rumsey, F. (2006). Feature Extraction for the Prediction of Multichannel Spatial Audio Fidelity. *IEEE Transactions on Audio, Speech and Language Processing*, 14(6), 1994–2005. <https://doi.org/10.1109/TASL.2006.883248>
- Geringer, J. M., Cassidy, J. W., & Byo, J. L. (1996). Effects of music with video on responses of nonmusic majors: An exploratory study. *Journal of Research in Music Education*, 44(3), 240–251.
- Geutskens, Y. (2011, October 12). SA-CD FAQ. Retrieved October 23, 2016, from [http://www.ps3sacd.com/sacdfaq.html#\\_Toc168933215](http://www.ps3sacd.com/sacdfaq.html#_Toc168933215)
- Ghazanfar, A. A., & Shepherd, S. V. (2011). Monkeys at the Movies: What Evolutionary Cinematics Tells Us about Film. *Projections*, 5(2), 1–25.
- Ghighi, G. (1985). Rassegna stampa cinematografica. In E. Ghezzi & M. Causio (Eds.), libretto allegato al DVD “Prénom Carmen”, RaroVideo, s.d. Italia.
- Gilroy, S. W., Porteous, J., Charles, F., Cavazza, M., Soreq, E., Raz, G., Ikar, L., Or-Borichov, A., Ben-Arie, U., Klovatch, I. (2013). A Brain-Computer Interface to a Plan-Based Narrative. In *IJCAI* (pp. 1997–2005).
- Goldberg, M. (2012, January 12). RED TAILS to Play in Auro-3D 11.1 Surround Sound in Select Locations. Retrieved October 25, 2016, from <http://collider.com/red-tails-auro-3d-11-1-surround-sound/>
- Goldmark, D., Kramer, L., & Leppert, R. (2007). *Beyond the soundtrack: representing music in cinema*. Berkeley: Univ of California Press.
- Gondan, M., Blurton, S. P., Hughes, F., & Greenlee, M. W. (2011). Effects of spatial and selective attention on basic multisensory integration. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1887–1897. <https://doi.org/10.1037/a0025635>
- Goodwin, S. N. (2009). 3-D sound for 3-D games-beyond 5.1. In *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Audio Engineering Society.
- Gooskens, G. (2014). Immersion. *Proceedings of the European Society for Aesthetics*, 6, 181–189.
- Gorbman, C. (1980). Narrative film music. *Yale French Studies*, (60), 183–203.
- Gorbman, C. (1987). *Unheard Melodies: Narrative Film Music*. Indiana University Press.

- Grabowski, M. (Ed.). (2015). *Neuroscience and Media: New Understandings and Representations*. New York and London: Routledge.
- Grainge, P. (2008). Selling Spectacular Sound: Dolby and the Unheard History of Technical Trademarks. In *Lowering the Boom: Critical Studies in Film Sound* (pp. 251–268). Urbana and Chicago: University of Illinois Press.
- Grajeda, T. (2015). The “Sweet Spot”: The Technology of Stereo and the Field of Auditorship. In P. Théberge, Devine, Kyle, & Everett, Tom (Eds.), *Living Stereo: Histories and Cultures of Multichannel Sound* (pp. 37–64). New York: Bloomsbury.
- Green, E. (2011a). Steiner, Korngold and the Musical Expression of Physical Space – A Preliminary Note. *International Review of Aesthetics and Sociology of Music*, 42(1), 59–78.
- Green, E. (2011b). Steiner, Korngold and the Musical Expression of Physical Space—A Preliminary Note. *International Review of the Aesthetics and Sociology of Music*, 42(1), 59–78.
- Green, J. J., Doesburg, S. M., Ward, L. M., & McDonald, J. J. (2011). Electrical Neuroimaging of Voluntary Audiospatial Attention: Evidence for a Supramodal Attention Control Network. *Journal of Neuroscience*, 31(10), 3560–3564. <https://doi.org/10.1523/JNEUROSCI.5758-10.2011>
- Grignon, L. D. (1953). Stereophonic Recording and Reproducing System. *Journal of the Society of Motion Pictures and Television Engineers*, 61(3), 355–363.
- Grignon, L. D. (1956). Roadshow Sound System for “Carousel.” *International Projectionist*, 31(6), 22, 35–36.
- Grimshaw, M. (2007). The Resonating Spaces of First-Person Shooter Games. In *Games Computing and Creative Technologies: Conference Papers (Peer-Reviewed). Paper 4*. Liverpool: School of Games Computing and Creative Technologies.
- Grimshaw, M. (2009). The audio Uncanny Valley: Sound, fear and the horror game. Retrieved from <http://ubir.bolton.ac.uk/246/>
- Grodal, T. (1999). *Moving pictures: A new theory of film genres, feelings, and cognition*. Oxford: Oxford University Press.
- Grodal, T. (2015). *Immagini-corpo. Cinema, natura, emozioni* [2009]. (M. Guerra, Ed., K. McManus, Trans.). Parma: Diabasis.
- Guastavino, C., & Katz, B. F. G. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *The Journal of the Acoustical Society of America*, 116(2), 1105. <https://doi.org/10.1121/1.1763973>

- Guttenberg, S. (2016, July 10). Deja vu: Yet another high-resolution audio “format.” Retrieved October 29, 2016, from <https://www.cnet.com/news/deja-vu-yet-another-high-resolution-audio-format/>
- Hagman, F. (2010). *Emotional response to sound. Influence of spatial determinants*. Chalmers University of Technology, Göteborg, Sweden. Retrieved from <http://publications.lib.chalmers.se/records/fulltext/133873.pdf>
- Hains, J. (2001). Dal rullo di cera al CD. In *Enciclopedia della musica* (Vol. I. Il Novecento, pp. 783–819). Torino: Einaudi.
- Hale, G. G. (2009). *A cognitive schematic analysis of film* (PhD dissertation). University of York.
- Hall, D. A., & Moore, D. R. (2003). Auditory neuroscience: The salience of looming sounds. *Current Biology*, 13(3), R91–R93.
- Handzo, S. (1985). Appendix: A Narrative Glossary of Film Sound Technology. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice*. New York: Columbia University Press.
- Harley, M. A. (1994). *Space and Spatialization in Contemporary Music: History and Analysis, Ideas and Implementations* (PhD dissertation). McGill University, Faculty of Music, Montreal.
- Harrison, J., & Wilson, S. (2010). Editorial. *Organised Sound*, 15(03), 183–184. <https://doi.org/10.1017/S1355771810000233>
- Hart, M. (2000). Cinerama Sound Recording for *How the West Was Won*. Retrieved October 23, 2016, from <http://www.widescreenmuseum.com/widescreen/htwwwsound.htm>
- Hart, M. (2010). Cinerama. Retrieved October 25, 2016, from <http://www.widescreenmuseum.com/widescreen/wingcr1.htm>
- Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections*, 2(1), 1–26.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664), 1634–1640.
- Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, 28(10), 2539–2550.
- Hauerslev, T. (2003, September). 70mm & 35mm CDS. Retrieved October 24, 2016, from <http://www.in70mm.com/library/process/cds/index.htm>
- Hauerslev, T. (2011, May 25). ...in Sensurround: ‘Earth-Shattering’ – A new dimension to the motion picture. Retrieved October 23, 2016, from [http://www.in70mm.com/news/2011/in\\_sensurround/english/index.htm](http://www.in70mm.com/news/2011/in_sensurround/english/index.htm)
- Haynes, N. M. (1954). Stereophonic Nomenclature. *Audio Engineering*, 38(1), 19, 54.

- Heffner, R. S., & Heffner, H. E. (1992). Visual Factors in Sound Localization in Mammals. *The Journal of Comparative Neurology*, 317(3), 219–232.
- Heimann, K. S., Uithol, S., Calbi, M., Umiltà, M. A., Guerra, M., & Gallese, V. (2016). “Cuts in Action”: A High-Density EEG Study Investigating the Neural Correlates of Different Editing Techniques in Film. *Cognitive Science*, 41(6), 1555–1588.
- Heimann, K., Umiltà, M. A., Guerra, M., & Gallese, V. (2014). Moving mirrors: a high-density EEG study investigating the effect of camera movements on motor cortex activation during action observation. *Journal of Cognitive Neuroscience*, 26(9), 2087–2101.
- Heldt, G. (2013). *Music and Levels of Narration in Film*. Bristol: Intellect.
- Hendrix, C., & Barfield, W. (1996). The sense of presence within auditory virtual environments. *Presence: Teleoperators & Virtual Environments*, 5(3), 290–301.
- Hermann, T., Hunt, A., & Neuhoff, J. G. (Eds.). (2011). *The Sonification Handbook*. Berlin: Logos Verlag.
- Hoeckner, B., Wyatt, E. W., Decety, J., & Nusbaum, H. (2011). Film music influences how viewers relate to movie characters. *Psychology of Aesthetics, Creativity, and the Arts*, 5(2), 146.
- Hogan, P. C. (2009). On the Meaning of Style: Cognition, Culture, and Visual Technique in Bimal Roy’s *Sujata*. *Projections*, 3(2), 71–90.
- Høier, S. (2014). Surrounded by Ear Candy? The Use of Surround Sound in Oscar-nominated Movies 2000-2012. *Nordicom Review*, 35.
- Holbrook, M. (2003). A Book-review Essay on the Role of Ambi-diegetic Film Music in the Product Design of Hollywood Movies: Macromarketing in La-La-Land. *Consumption, Markets and Culture*, 6(3), 207–230.
- Holbrook, M. B. (2004). Ambi-Diegetic Music in Films as a Product Design and -Placement Strategy: The Sweet Smell of Success. *Marketing Theory*, 4(3), 171–185. <https://doi.org/10.1177/1470593104045535>
- Holbrook, M. B. (2005a). Ambi-diegetic Music in the Movies: The Crosby Duets in *High Society*. *Consumption Markets & Culture*, 8(2), 153–182. <https://doi.org/10.1080/10253860500112859>
- Holbrook, M. B. (2005b). The Ambi-Diegesis Of ‘My Funny Valentine’. In S. Lannin & M. Caley (Eds.), *Pop Fiction: the Song in Cinema*. Bristol: Intellect.
- Holbrook, M. B. (2009). A cinemusicaliterary analysis of the American dream as represented by biographical jazz comedepictions in the golden age of Hollywood biopics: Blow, Horatio, Blow; O, Jakie, O; Go, Tommy, Go; No, Artie, No. *Marketing Theory*, 9(3), 259–313. <https://doi.org/10.1177/1470593109338141>

- Holman, T. (2008). *Surround Sound: Up and Running*. Burlington, MA: Focal Press.
- Home Theater Forum. [www.hometheaterforum.com](http://www.hometheaterforum.com)
- Hopper, F. L. (1931). The Measurement of Reverberation Time and its Application to Acoustic Problems in Sound Pictures. *Journal of the Acoustical Society*, 499–505.
- Horning, S. S. (2004). Engineering the Performance: Recording Engineers, Tacit Knowledge and the Art of Controlling Sound. *Social Studies of Science*, 34(5), 703–731. <https://doi.org/10.1177/0306312704047536>
- Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception: An Interdisciplinary Journal*, 20(2), 151–171.
- Hutson, J., Loschky, L. C., Smith, T., & Magliano, J. P. (2014). The look of evil: How are eye movements influenced by film comprehension? *Journal of Vision*, 14(10), 760.
- Hutson, J., Smith, T., Magliano, J., Heidebrecht, G., Hinkel, T., Tang, J. L., & Loschky, L. (2015). Eye movements while watching narrative film: a dissociation of eye movements and comprehension. *Journal of Vision*, 15(12), 116.
- Hutson, J., Smith, T., Magliano, J., Hinkel, T., Boberg, C., Caldera, M., Menzies, C., Tran, K., & Loschky, L. (2016). Attentional synchrony during narrative film viewing: Turning off the “tyranny of film” through a task manipulation at odds with narrative comprehension. *Journal of Vision*, 16(12), 123–123.
- IMAX Corporation. (1999). IMAX: The 15/70 Filmmaker’s Manual. Retrieved from <https://kinotechnologies.files.wordpress.com/2011/02/filmmaker1.pdf>
- IMDb. (n.d.). Perspecta Stereo Titles (Sort by Release Date, asc). Retrieved October 26, 2016, from [http://www.imdb.com/search/title?sound\\_mixes=perspecta\\_stereo&sort=release\\_date,asc&ref\\_=tt\\_dt\\_spec](http://www.imdb.com/search/title?sound_mixes=perspecta_stereo&sort=release_date,asc&ref_=tt_dt_spec)
- International Electronics Corporation. (1955). Insetto pubblicitario. *High Fidelity Magazine*, 5(2), 102.
- Irving, E. (1943). Music in Films. *Music & Letters*, 24(4), 223–235.
- Jones, M. (2005). Composing Space: Cinema and Computer Gaming-The Macro-Mise En Scene and Spatial Composition. In *Imaginary Worlds Symposium*. Citeseer. Retrieved from <https://pdfs.semanticscholar.org/0068/39ebc48510ffe7f43a98a4d19e3358f40a72.pdf>
- Jordan, R. (2007). The Work of Hildegard Westerkamp in the Films of Gus Van Sant: An Interview with the Soundscape Composer (and some added thoughts of my own). *Offscreen*, 11(8–9).

- Jordan, R. (2012). The Ecology of Listening while Looking in the Cinema: Reflective audioviewing in Gus Van Sant's *Elephant*. *Organised Sound*, 17(03), 248–256. <https://doi.org/10.1017/S1355771811000458>
- Julien, J.-R. (1987). Défense et illustration des fonctions de la musique de film [qui tratto da Simeon 1995: 47]. In A. Hennion (Ed.), *Les musiques des films*. Toulouse: Privat.
- Jullier, L. (2006). *Il cinema postmoderno* [1997]. (C. Capetta, Trans.). Torino: Kaplan.
- Jullier, L. (2007). *Il suono nel cinema. Storia, regole, mestieri* [2006]. (E. Mugellini, Trans.). Torino: Lindau.
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(05). <https://doi.org/10.1017/S0140525X08005293>
- Kalinak, K. (1992). *Settling the score: music and the classical Hollywood film*. University of Wisconsin Press.
- Kalinak, K. (Ed.). (2015). *Sound: Dialogue, Music, and Effects*. London & New York: I. B. Tauris.
- Kassabian, A. (2001). *Hearing Film: Tracking Identifications in Contemporary Hollywood Film Music*. New York and London: Routledge.
- Kellogg, E. W. (1955). History of Sound Motion Pictures – Final Installment. *Journal of the Society of Motion Pictures and Television Engineers*, 64(8), 422–437.
- Kellogg, J. (2014, April). Multi-Dimensional Sound and the Role of MDA. *Film Journal International*, 117(4), 90–91.
- Kerins, M. (2011). *Beyond Dolby (Stereo): Cinema in the Digital Sound Age*. Bloomington: Indiana University Press.
- Kerins, M., Lipscomb, S. D., & Rendahl, A. (2010). The influence of 5.1 surround sound presentation on the cinematic and music listening experiences. Unpublished paper.
- Kim, M., Jeon, C., & Kim, J. (n.d.). A Study on Immersion and Presence of a Portable Hand Haptic System for Immersive Virtual Reality. *Sensors*, 17(5), 1141.
- Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, 416(6877), 172–174.
- Klimek, M. P. (1992). Imagining the Sound(s) of Shakespeare: Film Sound and Adaptation. In R. Altman (Ed.), *Sound Theory/Sound Practice* (pp. 204–216). New York: Routledge.
- Kobayashi, M., Teramoto, W., Hidaka, S., & Sugita, Y. (2012). Indiscriminable sounds determine the direction of visual motion. *Scientific Reports*, 2. <https://doi.org/10.1038/srep00365>
- Kobayashi, M., Ueno, K., & Ise, S. (2015). The Effects of Spatialized Sounds on the Sense of Presence in Auditory Virtual Environments: A Psychological and Physiological Study.

*Presence: Teleoperators and Virtual Environments*, 24(2), 163–174.  
[https://doi.org/10.1162/PRES\\_a\\_00226](https://doi.org/10.1162/PRES_a_00226)

- Kobrak, H. G. (1951). Auditory Perspective A Study of the Biological Factors Related to Directional Hearing. *Journal of the Society of Motion Pictures and Television Engineers*, 57(4), 328–335.
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297(5582), 846–848.
- Kühn, S., & Gallinat, J. (2011). A quantitative meta-analysis on cue-induced male sexual arousal. *The Journal of Sexual Medicine*, 8(8), 2269–2275.
- Lange, A. (2002, March 31). Le premier medium électrique de diffusion culturelle: le Theatrophone de Clement Ader (1881). Retrieved October 25, 2016, from <http://histv2.free.fr/theatrophone/theatrophone.htm>
- Lanza, A. (1991). *Il secondo Novecento*. Torino: EDT.
- Larsson, P., Västfjäll, D., Olsson, P., & Kleiner, M. (2007). When what you hear is what you see: Presence and auditory-visual integration in virtual environments. In *Proceedings of the 10th annual international workshop on presence* (pp. 11–18).
- Lastra, J. (1992). Reading, Writing, and Representing Sound. In R. Altman (Ed.), *Sound Theory/Sound Practice* (pp. 65–86). New York: Routledge.
- Lastra, J. (2000). *Sound Technology and the American Cinema: Perception, Representation, Modernity*. New York: Columbia University Press.
- Lastra, J. (2008). Film and the Wagnerian Aspiration: Thoughts on Sound Design and the History of the Senses. In J. Beck & T. Grajeda (Eds.), *Lowering the Boom: Critical Studies in Film Sound* (pp. 123–138). Urbana and Chicago: University of Illinois Press.
- Latini, G. (2006). *L'immagine sonora: caratteri essenziali del suono cinematografico*. Roma: Artemide.
- Latini, G. (2015). Dolby Atmos: verso il corpo assoluto del suono cinematografico. *Imago*, 12(Oltre il corpo del cinema. Reti, virtualità, apparati), 107–116.
- Lennox, P. (2004). *The Philosophy of Perception in Artificial Auditory Environments: Spatial Sound and Music*. University of York.
- Lennox, P. (2009). The emotional contents of the 'space' in spatial music. International Conference on Music and Emotion, Durham, UK.

- Lennox, P., & Myatt, T. (2007a). A Perceptual Approach to the Composition of Meaning in Artificial Spatial Audio. In *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*. Audio Engineering Society.
- Lennox, P., & Myatt, T. (2007b). Concepts of perceptual significance for composition and reproduction of explorable sound fields. In *Proceedings of the 13th International Conference on Auditory Display*. Schulich School of Music, McGill University.
- Lennox, P., & Myatt, T. (2011). Perceptual cartoonification in multi-spatial sound systems. In *Proceedings of the 11th International Conference on Auditory Display*. Retrieved from <http://derby.openrepository.com/derby/handle/10545/347158>
- Lennox, P., Myatt, T., & Vaughan, J. (2001). 3D Audio as an Information Environment. In *Audio Engineering Society Conference: 19th International Conference: Surround Sound-Techniques, Technology, and Perception*. Audio Engineering Society. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=10117>
- Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, 4(4), 501–506.
- Levin, D. T., & Simons, D. J. (2000). Perceiving Stability in a Changing World: Combining Shots and Intergrating Views in Motion Pictures and the Real World. *Media Psychology*, 2(4), 357–380. [https://doi.org/10.1207/S1532785XMEP0204\\_03](https://doi.org/10.1207/S1532785XMEP0204_03)
- Levin, D. T., & Wang, C. (2009). Spatial representation in cognitive science and film. *Projections*, 3(1), 24–52.
- Levin, T. (1984). The acoustic dimension. *Screen*, 25(3), 55–68.
- Levinson, N. (1933). Re-Recording, Dubbing, or Duping. *American Cinematographer*, 13(11), 6–7.
- Levinson, N., & Goldsmith, L. T. (1941). Vitasound. *Journal of the Society of Motion Pictures Engineers*, 37(8), 147–153.
- Liebman, M. (2010, September 3). Tommy Blu-ray Review – I’ve got “Tommy;” who’ll bring the baked beans, the pinball machine, and the hallucinogens? Retrieved October 23, 2016, from <http://www.blu-ray.com/movies/Tommy-Blu-ray/11805/#Review>
- Lipscomb, S. D. (1995). *Cognition of musical and visual accent structure alignment in film and animation*. University of California, Los Angeles.
- Lipscomb, S. D. (2005). The perception of audio-visual composites: Accent structure alignment of simple stimuli. *Selected Reports in Ethnomusicology*, 12, 37–67.
- Lipscomb, S. D., & Kendall, R. A. (1994). Perceptual judgement of the relationship between musical and visual components in film. *Psychomusicology: A Journal of Research in Music Cognition*, 13(1–2), 60.

- Lipscomb, S. D., & Kerins, M. (2004). An empirical investigation into the effect of presentation mode in the cinematic and music listening experience. In *Proceedings of the 8th International Conference on Music Perception & Cognition, Evanston, IL*. Retrieved from [http://www.lipscomb.umn.edu/docs/ICMPC8proceedings\\_APR10\\_LipscombKerins.pdf](http://www.lipscomb.umn.edu/docs/ICMPC8proceedings_APR10_LipscombKerins.pdf)
- Lipscomb, S. D., & Kim, E. M. (2004). Perceived match between visual parameters and auditory correlates: an experimental multimedia investigation. In *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 72–75). Retrieved from <http://www.icmpc8.umn.edu/proceedings/ICMPC8/PDF/AUTHOR/MP040304.PDF>
- Lipscomb, S. D., & Tolchinsky, D. E. (2005). The role of music communication in cinema. In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication* (pp. 383–404). New York: Oxford University Press.
- Lipscomb, S. D., & Zehnder, S. M. (2004). Immersion in the virtual environment: The effect of a musical score on the video gaming experience. *Journal of Physiological Anthropology and Applied Human Science*, 23(6), 337–343.
- LoBrutto, V. (1994). *Sound-on-film: Interviews with creators of film sound*. Westport, CT: Praeger.
- Loomis, J. M. (2016). Presence in Virtual Reality and Everyday Life: Immersion within a World of Representation. *Presence*, 25(2), 169–174.
- Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PloS One*, 10(11), e0142474.
- Loye, D. P. (1941). Acoustic Design Features of Studio Stages, Monitor Rooms, and Review Rooms. *Journal of the Society of Motion Pictures Engineers*, 36(6), 593–621.
- Maasø, A. (2008). The Proxemics of the Mediated Voice. In J. Beck & T. Grajeda (Eds.), *Lowering the Boom: Critical Studies in Film Sound* (pp. 36–50). Urbana and Chicago: University of Illinois Press.
- Macedo, F. (2015). Investigating Sound in Space: Five meanings of space in music and sound art. *Organised Sound*, 20(02), 241–248. <https://doi.org/10.1017/S1355771815000126>
- Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992). Perceptual organization and attention. *Cognitive Psychology*, 24(4), 475–501.
- Magliano, J. P., & Clinton, J. A. (2016). A Vision of the Viewer: Situating Narration in the Fiction Film in the Context of Theories of Narrative Comprehension. *Projections*, 10(1), 3–11.
- Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, 35(8), 1489–1517.

- Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes*, 22(3), 199–224.
- Magliano, J. P., Larson, A. M., Higgs, K., & Loschky, L. C. (2016). The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Memory & Cognition*, 44(2), 207–219.
- Magliano, J. P., Loschky, L. C., Clinton, J. A., & Larson, A. M. (2013). Is Reading the Same as Viewing? B. Miller, LC and McCardle, P., Editors, *Unraveling the Behavioral, Neurobiological and Genetic Components of Reading Comprehension*, Pages, 78–90.
- Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, 15(5), 533–545. <https://doi.org/10.1002/acp.724>
- Magliano, J. P., Radvansky, G. A., Forsythe, J. C., & Copeland, D. E. (2014). Event segmentation during first-person continuous events. *Journal of Cognitive Psychology*, 26(6), 649–661.
- Manolas, C., & Pauletto, S. (2009). Enlarging the Diegetic Space: Uses of the Multi-channel Soundtrack in Cinematic Narrative. *The Soundtrack*, 2(1), 39–55.
- Marshall, S. K., & Cohen, A. J. (1988). Effects of Musical Soundtracks on Attitudes toward Animated Geometric Figures. *Music Perception: An Interdisciplinary Journal*, 6(1), 95–112. <https://doi.org/10.2307/40285417>
- Martinelli, G. (2006). *L'ascolto professionale*. Roma: Lambda.
- Martinelli, G., & Corelli, S. (2006). La costruzione del suono per l'immagine cinematografica: dalla presa diretta al mixage, ed il rapporto con l'ambiente home-theater. Presented at the Multichannel Audio: per l'immagine e per il suono, Rimini.
- Massey, H. (Ed.). (2004). The Recording Academy's Producers & Engineers Wing Recommendations For Surround Sound Production. The National Academy of Recording Arts & Sciences, Inc.
- Maxfield, J. P. (1931). Some Physical Factors Affecting the Illusion in Sound Motion Pictures. *Journal of the Acoustical Society*, 16(7).
- Maxfield, J. P. (1938). Demonstration of Stereophonic Recording with Motion Pictures. *Journal of the Society of Motion Pictures Engineers*, 30(2), 131–135.
- Maxfield, J. P., & Harrison, H. C. (1926). Methods of High Quality Recording and Reproducing of Music and Speech based on Telephone Research. *Bell System Technical Journal*, 5(3), 493–523.
- Maxfield, J. P., Colledge, A. W., & Friebus, R. T. (n.d.). Pick-up for Sound Motion Pictures (Including Stereophonic). *Journal of the Society of Motion Pictures Engineers*, 30(6), 666–679.

- McGee, M. (2001). *Encyclopedia of Motion Picture Sound*. Jefferson (NC): McFarland.
- McKenzie, I., Lennox, P., & Wiggins, B. (2015). Inside-Outside: 3-D music through Tissue conduction. In *Proceedings of the International Conference on the Multimodal Experience of Music 2015, Sheffield, UK*. Retrieved from <http://derby.openrepository.com/derby/handle/10545/347177>
- McMahan, A. (2003). Immersion, Engagement, and Presence: A Method for Analyzing 3-D Video Games. In *The Video Game Theory Reader* (Mark J. P. Wolf and Bernard Perron). New York: Routledge.
- McMahan, A., & Buckland, W. (2005). The Cognitive Semiotics of Virtual Reality. In *Virtual Reality International, an Aqliated Conference of the Human Computer Interaction (HCI) International Conference*. Las Vegas. Retrieved from [http://intelligentagent.com/archive/Vol5\\_No1\\_ave\\_mcmahanbuckland.htm](http://intelligentagent.com/archive/Vol5_No1_ave_mcmahanbuckland.htm)
- Meandri, I., & Valle, A. (Eds.). (2011a). *Suono/Immagine*. Torino: Kaplan.
- Meandri, I., & Valle, A. (Eds.). (2011b). *Suono/Immagine/Genere*. Torino: Kaplan.
- Metz, C. (1985). Aural Objects. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 154–161). New York: Columbia University Press.
- Miceli, S. (2009). *Musica per film: Storia, Estetica-Analisi, Tipologie*. San Giuliano Milanese (MI): Ricordi-LIM.
- Miceli, S. (2010). *Musica e cinema nella cultura del Novecento*. Roma: Bulzoni.
- Miller, W. C. (1938). Basis of Motion Picture Sound. In Research Council of the Academy of Motion Picture Arts and Sciences (Ed.), *Motion Picture Sound Engineering* (pp. 1–10). New York: D. Van Nostrand Company.
- Moeck, T., Bonneel, N., Tsingos, N., Drettakis, G., Viaud-Delmon, I., & Alloza, D. (2007). Progressive perceptual audio rendering of complex scenes. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games* (pp. 189–196). ACM.
- Moulier, V., Mouras, H., Péligrini-Issac, M., Glutron, D., Rouxel, R., Grandjean, B., Bittoun, J., & Stoléru, S. (2006). Neuroanatomical correlates of penile erection evoked by photographic stimuli in human males. *Neuroimage*, 33(2), 689–699.
- Mouras, H., Stoléru, S., Bittoun, J., Glutron, D., Péligrini-Issac, M., Paradis, A.-L., & Burnod, Y. (2003). Brain processing of visual sexual stimuli in healthy men: a functional magnetic resonance imaging study. *Neuroimage*, 20(2), 855–869.
- Mueller, W. A. (1935). A Device for Automatically Controlling the Balance Between Recorded Sound. *Journal of the Society of Motion Pictures Engineers*, 25(1), 79–86.

- Muston Manufacturing Inc. (1955). Insetto pubblicitario Munston Maestro. *High Fidelity Magazine*, 5(9), 47.
- Myatt, T. (1998). Sound in space. *Organised Sound*, 3(2), 91–92.
- Nelson, P. (2015). The Materiality of Space. *Organised Sound*, 20(03), 323–330. <https://doi.org/10.1017/S1355771815000254>
- Neuhoff, J. G. (2004). Auditory Motion and Localization. In J. G. Neuhoff (Ed.), *Echological Psychoacoustics* (pp. 87–111). New York: Academic Press.
- Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Immersion Revisited: A Review of Existing Definitions of Immersion and Their Relation to Different Theories of Presence. *Human Technology*, 12.
- Norden, M. F. (2007). Diegetic Commentaries. *Offscreen*, 11(8–9). Retrieved from [http://offscreen.com/view/soundforum\\_2](http://offscreen.com/view/soundforum_2)
- Normandeau, R. (2009). Timbre Spatialisation: The medium is the space. *Organised Sound*, 14(03), 277. <https://doi.org/10.1017/S1355771809990094>
- Nothelfer, C. E., DeLong, J. E., & Cutting, J. E. (2009). Shot structure in Hollywood film. *Indiana University Undergraduate Journal of Cognitive Science*, 4, 103–113.
- Olson, H. F., & Massa, F. (1934). On the Realistic Reproduction of Sound with Particular Reference to Sound Motion Pictures. *Journal of the Society of Motion Pictures Engineers*, 23(2), 63–81.
- Originaltrilogy. [www.originaltrilogy.com/forum](http://www.originaltrilogy.com/forum)
- Otondo, F. (2008). Contemporary trends in the use of space in electroacoustic music. *Organised Sound*, 13(01). <https://doi.org/10.1017/S1355771808000095>
- Parke, R., Chew, E., & Kyriakakis, C. (2007). Quantitative and visual analysis of the impact of music on perceived emotion of film. *Computers in Entertainment (CIE)*, 5(3), 5.
- Parker, J. R., & Heerema, J. (2008). Audio Interaction in Computer Mediated Games. *International Journal of Computer Games Technology*, 2008, 1–8. <https://doi.org/10.1155/2008/178923>
- Pavlović, I., & Marković, S. (2011). The effect of music background on the emotional appraisal of film sequences. *Psihologija*, 44(1), 71–91. <https://doi.org/10.2298/PSI1101071P>
- Percheron, D. (1980). Sound in Cinema and its Relationship to Image and Diegesis. *Yale French Studies*, (60), 16. <https://doi.org/10.2307/2930001>
- Perri, M. (2016, January 30). La mia prima volta in Dolby Atmos. Retrieved October 24, 2016, from <http://www.mirkoperri.com/?p=2944>
- Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Attention, Perception, & Psychophysics*, 48(3), 214–226.

- Pierce, J. (2014, September). Immersive sound. *Boxoffice*, 150(9).
- Pinn, J. (2014, June). Is immersive sound the savior of digital cinema sound? *Boxoffice*, 150(6), 28–32.
- Plantinga, C. (2002). Cognitive film theory: An insider's appraisal. *Cinemas: Revue d'études Cinématographiques/Cinemas: Journal of Film Studies*, 12(2), 15–37.
- Plantinga, C. (2009). *Moving viewers: American film and the spectator's experience*. Berkeley and Los Angeles: University of California Press.
- Plantinga, C. R., & Smith, G. M. (1999). *Passionate views: Film, cognition, and emotion*. Baltimore and London: Johns Hopkins University Press.
- Plass, G. (1954). Stereophony from the Outside In: Have you ever thought of your head as a breakwater? *High Fidelity Magazine*, 4(2), 78–80.
- Pollack, I., & Pickett, J. M. (1957). Cocktail party effect. *The Journal of the Acoustical Society of America*, 29(11), 1262–1262.
- Pollack, I., & Pickett, J. M. (1958). Stereophonic listening and speech intelligibility against voice babble. *The Journal of the Acoustical Society of America*, 30(2), 131–133.
- Ramaglia, V. (2004). *Il suono e l'immagine. Musica, voce, rumore e silenzio nel film*. Roma: Dino Audino.
- Rayner, A. (2013, August 10). The Sound of IMAX. Retrieved October 24, 2016, from <http://www.homecinemachoice.com/news/article/the-sound-of-imax/16250>
- Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, 20(1), 6–10.
- Raz, G., & Hendler, T. (2014). *Forking cinematic paths to the self: neurocinematically informed model of empathy in motion pictures*. Berghahn Journals.
- Raz, G., Hagin, B., & Hendler, T. (2013). E-motion pictures of the brain: Recursive paths between affective neuroscience and film studies. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring cognition at the movies* (pp. 285–336). New York: Oxford University Press.
- Raz, G., Jacob, Y., Gonen, T., Winetraub, Y., Flash, T., Soreq, E., & Hendler, T. (2014). Cry for her or cry with her: context-dependent dissociation of two modes of cinematic empathy reflected in network cohesion dynamics. *Social Cognitive and Affective Neuroscience*, 9(1), 30–38.
- Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., Jacob, Y., Atzil, S., Admon, R., Bleich-Cohen, M., Maron-Katz, A., Hendler, T., & Barrett, L. F. (2016). Functional connectivity dynamics during film viewing reveal common networks for

- different emotional experiences. *Cognitive, Affective, & Behavioral Neuroscience*, 16(4), 709–723. <https://doi.org/10.3758/s13415-016-0425-4>
- Rebelo, P. (2003). Performing space. *Organised Sound*, 8(02). <https://doi.org/10.1017/S1355771803000086>
- Recuber, T. (2007). Immersion Cinema: The Rationalization and Reenchantment of Cinematic Space. *Space and Culture*, 10(3), 315–330. <https://doi.org/10.1177/1206331207304352>
- Redouté, J., Stoléru, S., Grégoire, M.-C., Costes, N., Cinotti, L., Lavenne, F., Le Bars, D., Forest, M.G., & Pujol, J.-F. (2000). Brain processing of visual sexual stimuli in human males. *Human Brain Mapping*, 11(3), 162–177.
- Reeves, H. E. (1953). Adding the Sound to Cinerama. Retrieved October 23, 2016, from <http://www.in70mm.com/cinerama/archive/sound/index.htm>
- Reiskind, H. I. (1941). Multiple-Speaker Reproducing Systems for Motion Pictures. *Journal of the Society of Motion Pictures Engineers*, 37(8), 154–163.
- Reisz, K. (1964). *The Technique of Film Editing* [qui tratto da Doane 1985a: 56]. London & New York: Focal Press.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368–373.
- Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of Music*, 32(4), 371–388. <https://doi.org/10.1177/0305735604046096>
- Robinson, S. M. (2003, March 2). post #1, in thread “Decoding a Height Channel.” Retrieved October 24, 2016, from <https://www.avforums.com/threads/decoding-a-height-channel.70372/>
- Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without Attention: Results of a New Method. *Cognitive Psychology*, (24), 502–534.
- Roddy, S., & Furlong, D. (2014). Embodied aesthetics in auditory display. *Organised Sound*, 19(1), 70–77.
- Rojas, D., Kapralos, B., Cristancho, S., Collins, K., Hogue, A., Conati, C., & Dubrowski, A. (2012). Developing Effective Serious Games: The Effect of Background Sound on Visual Fidelity Perception with Varying Texture Resolution. In *Medicine Meets Virtual Reality 19* (pp. 386–392).
- Rose, J. (2015). *Producing Great Sound for Film & Video: Expert Tips from Preproduction to Final Mix*. New York and London: Focal Press.
- Rumsey, F. (2001). *Spatial Audio*. Oxford: Focal Press.

- Rumsey, F. (2002). Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm. *Journal of the Audio Engineering Society*, 50(9), 651–666.
- Salt, B. (1992). *Film style and technology: History and analysis*. Starword.
- Salt, B. (n.d.). How They Cut Dialogue Scenes. Retrieved from <http://www.cinematics.lv/dev/CutDial.pdf>
- Sbravatti, V. (2016). Story-Music / Discourse-Music: Analyzing the Relationship Between Placement and Function of Music in Films. *Music and the Moving Image*, 9(3), 19–37. <https://doi.org/doi:10.5406/musimoviimag.9.3.0019>
- Schafer, R. M. (1994). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester: Destiny Books.
- Schäfer, S., & Krebs, J. (2003). Sound – Time – Space – Movement: the Space-soundInstallations of the artist-couple <sabine schäfer // joachim krebs>. *Organised Sound*, 8(02). <https://doi.org/10.1017/S1355771803000128>
- Schiavio, A., Menin, D., & Matyja, J. (2014). Music in the flesh: Embodied simulation in musical understanding. *Psychomusicology: Music, Mind, and Brain*, 24(4), 340–343. <https://doi.org/10.1037/pmu0000052>
- Schoenherr, S. (n.d.). IMAX Film Format. Retrieved October 24, 2016, from <http://sunnycv.com/steve/filmnotes/imax.html>
- Schoenherr, S. E. (2001). Stereo. Retrieved October 23, 2016, from <http://www.aes.org/aeshc/docs/recording.technology.history/stereo.html>
- Schreger, C. (1978). The Second Coming of Sound. *Film Comment*, 14(5), 34–37.
- Schroeder, F. (2016). Museum City: Improvisation and the narratives of space. *Organised Sound*, 21(03), 249–259. <https://doi.org/10.1017/S1355771816000224>
- Schust, M. (2004). Effects of low frequency noise up to 100 Hz. *Noise and Health*, 6(23), 73–85.
- Schütze, S. (2003). The creation of an audio environment as part of a computer game world: the design for Jurassic Park – Operation Genesis on the XBOX™ as a broad concept for surround installation creation. *Organised Sound*, 8(02). <https://doi.org/10.1017/S1355771803000074>
- Scott, V. (1994, January 28). What 3-D glasses do for dimensional optics a new...individual sound system does to deliver revolutionary acoustic enhancement for movies. *United Press International*. Retrieved from <http://upi.com/4348874>
- SDDS. (n.d.). SDDS 8 Movies. Retrieved October 24, 2016, from [https://web.archive.org/web/20120912042229/http://www.sdds.com/news\\_movies.cfm?showfilms=8](https://web.archive.org/web/20120912042229/http://www.sdds.com/news_movies.cfm?showfilms=8)

- Sedman, D. (2015). Television: Now with Two Channels of Audio. In P. Théberge, Devine, Kyle, & Everett, Tom (Eds.), *Living Stereo: Histories and Cultures of Multichannel Sound* (pp. 185–206). New York: Bloomsbury.
- Seeley, E. S. (1941). A Compact Direct-Reading Reverberation Meter. *Journal of the Society of Motion Pictures Engineers*, 37(12), 557–568.
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, 385(6614), 308. <https://doi.org/10.1038/385308a0>
- Sergi, G. (1998). Tales of the Silent Blast: *Star Wars* and Sound. *Journal of Popular Film and Television*, 26(1), 12–22. <https://doi.org/10.1080/01956059809602769>
- Sergi, G. (1999). The Sonic Playground: Hollywood Cinema and its Listeners. Retrieved January 26, 2014, from <http://www.filmsound.org/articles/sergi/index.htm>
- Sergi, G. (2004). *The Dolby era: Film sound in contemporary Hollywood*. Manchester: Manchester University Press.
- Sergi, G. (2006). In Defence of Vulgarity: The Place of Sound Effects in the Cinema. *Scope: An Online Journal of Film Studies*, 5. Retrieved from <https://www.nottingham.ac.uk/scope/documents/2006/june-2006/sergi.pdf>
- Sergi, G. (2013). Knocking at the door of cinematic artifice: Dolby *Atmos*, challenges and opportunities. *The New Soundtrack*, 3(2), 107–121. <https://doi.org/10.3366/sound.2013.0041>
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1), 147–152.
- Shea, T. E. (1930). Recording the Sound Picture. *Bell Laboratories Record*, 8(8), 356–362.
- Sherlock, D. J. (2004). Wide Screen Movies Corrections. Retrieved from [www.film-tech.com/warehouse/tips/WSMC20.pdf](http://www.film-tech.com/warehouse/tips/WSMC20.pdf)
- Shimamura, A. P. (2013). *Psychocinematics: Exploring cognition at the movies*. New York: Oxford University Press.
- Silyanov, D. (2015, December 31). Sovscope 70 – 70mm Films in USSR. Retrieved October 23, 2016, from <http://www.in70mm.com/library/process/sovscope/index.htm>
- Simeon, E. (1995). *Manuale di storia della musica nel cinema*. Milano: Rugginenti.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28(9), 1059–1074.
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1(7), 261–267.
- Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5), 560–565.

- Slowik, M. (2013). Diegetic Withdrawal and Other Worlds: Film Music Strategies before King Kong, 1927–1933. *Cinema Journal*, 53(1), 1–25.
- Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound*, 12(01), 35. <https://doi.org/10.1017/S1355771807001665>
- Smith, G. M. (2003). *Film structure and the emotion system*. Cambridge: Cambridge University Press.
- Smith, J. (1996). Unheard melodies? A critique of psychoanalytic theories of film music. In *Post-Theory: Reconstructing Film Studies*. Madison: University of Wisconsin Press.
- Smith, J. (2009). Bridging the gap: reconsidering the border between diegetic and nondiegetic music. *Music and the Moving Image*, 2(1), 1–25.
- Smith, J. (2013). Atmos, all around: A guest post by Jeff Smith. Retrieved from <http://www.davidbordwell.net/blog/2013/05/01/atmos-all-around>
- Smith, M. (1995). *Engaging characters: Fiction, emotion, and the cinema*. Oxford: Clarendon Press.
- Smith, M. (2017). *Film, Art, and the Third Culture: A Naturalized Aesthetics of Film*. Oxford: Oxford University Press.
- Smith, T. J. (2005). *An attentional theory of continuity editing* (PhD dissertation). Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh.
- Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, 6(1), 1–27.
- Smith, T. J., & Henderson, J. M. (2008). Edit Blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research*, 2(2).
- Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8), 16–16.
- Smith, T. J., Levin, D., & Cutting, J. E. (2012). A Window on Reality: Perceiving Edited Moving Images. *Current Directions in Psychological Science*, 21(2), 107–113. <https://doi.org/10.1177/0963721412437407>
- Sobchack, V. (2004). *Carnal thoughts: Embodiment and moving image culture*. Berkeley and Los Angeles: University of California Press.
- Sobchack, V. (2005). When the Ear Dreams: Dolby Digital and the Imagination of Sound. *Film Quarterly*, 58(4), 2–15. <https://doi.org/10.1525/fq.2005.58.4.2>
- Solomon, J. W. (2007). *Spatialization in Music: The Analysis and Interpretation of Spatial Gestures* (PhD dissertation). University of Georgia, Athens.

- Song, W., Ellermeier, W., & Hald, J. (2011). Psychoacoustic evaluation of multichannel reproduced sounds using binaural synthesis and spherical beamforming. *The Journal of the Acoustical Society of America*, *130*(4), 2063–2075.
- Sørensen, R. (1997, September). Imax Sound System: IMAX Dome Luxury, Copenhagen, Denmark. Retrieved October 24, 2016, from [http://in70mm.com/newsletter/1997/50/imax\\_sound/index.htm](http://in70mm.com/newsletter/1997/50/imax_sound/index.htm)
- Steinberg, J. C., & Snow, W. B. (1934). Auditory Perspective - Physical Factors. *Electrical Engineering*, *53*(1), 12–17.
- Stephens. (1954). Inserto pubblicitario Stephens Manufacturing Corporation. *High Fidelity Magazine*, *4*(2), 16.
- Sterne, J. (2003). *The Audible Past: Cultural Origins of Sound Reproduction*. Durham: Duke University Press.
- Stichting tot Behoud van Historische Philips Producten [Fondazione per la preservazione dei prodotti storici Philips]. (2003). Philips Compact Cassette. Retrieved October 25, 2016, from <http://www.philips-historische-producten.nl/cassette-uk.html>
- Stockburger, A. (2003). The Game Environment from an Auditive Perspective. Presented at the Level Up, Digital Games Research Conference, Utrecht. Retrieved from [http://www.stockburger.at/files/2010/04/gameenvironment\\_stockburger1.pdf](http://www.stockburger.at/files/2010/04/gameenvironment_stockburger1.pdf)
- Stoléru, S., Gregoire, M.-C., Gerard, D., Decety, J., Lafarge, E., Cinotti, L., Lavenne, F., Le Bars, D., Vernet-Maury, E., & Rada, H. (1999). Neuroanatomical correlates of visually evoked sexual arousal in human males. *Archives of Sexual Behavior*, *28*(1), 1–21.
- Stull, W. (1941). Fantasound – Disney’s New Sound System. *American Cinematographer*, *22*, 58–59, 80, 82.
- Stumpf, R. J., Watson, W. O., & Leonard, R. J. (n.d.). Special effects generation and control system for motion pictures. *US 3973839 A*. Retrieved from <http://www.google.com/patents/US3973839>
- Tajadura-Jiménez, A., Larsson, P., Väljamäe, A., Västfjäll, D., & Kleiner, M. (2010c). When room size matters: Acoustic influences on emotional responses to sounds. *Emotion*, *10*(3), 416–422. <https://doi.org/10.1037/a0018423>
- Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., & Västfjäll, D. (2010b). Embodied auditory perception: the emotional impact of approaching and receding sound sources. *Emotion*, *10*(2), 216.
- Takahashi, Y. (2011). A study on the contribution of body vibrations to the vibratory sensation induced by high-level, complex low-frequency noise. *Noise and Health*, *13*(50), 2–8.

- Takahashi, Y., Yonekawa, Y., Kanada, K., & Maeda, S. (1999). A pilot study on the human body vibration induced by low frequency noise. *Industrial Health*, 37(1), 28–35.
- Tan, E. S. (1996). *Emotion and the structure of narrative film: Film as an emotion machine*. New York and London: Routledge.
- Tan, S.-L., Spackman, M. P., & Bezdek, M. A. (2007). Viewers' Interpretations of Film Characters' Emotions: Effects of Presenting Film Music Before or After a Character is Shown. *Music Perception: An Interdisciplinary Journal*, 25(2), 135–152.
- Tannenbaum, P. H. (1956). Music background in the judgment of stage and television drama. *Educational Technology Research and Development*, 4(2), 92–101.
- Taylor, H. M. (2007). The Success Story of a Misnomer. *Offscreen*, 11(8–9). Retrieved from [http://offscreen.com/view/soundforum\\_2](http://offscreen.com/view/soundforum_2)
- Teramoto, W., Hidaka, S., & Sugita, Y. (2010). Sounds move a static visual object. *PLoS One*, 5(8), e12255.
- Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y., & Suzuki, Y. (2012). Sounds can alter the perceived direction of a moving visual object. *Journal of Vision*, 12(3), 11–11.
- Terrone, E. (2004). Il rumore nel cinema: una domanda senza risposte. *Segnocinema*, (130), 23–27.
- Thayer, J. F., & Levenson, R. W. (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology: A Journal of Research in Music Cognition*, 3(1), 44.
- The THX Story. (n.d.). Retrieved October 24, 2016, from <http://www.thx.com/about-us/the-thx-story/>
- Théberge, P., Devine, K., & Everett, T. (Eds.). (2015a). *Living Stereo: Histories and Cultures of Multichannel Sound*. New York: Bloomsbury.
- Théberge, P., Devine, K., & Everett, T. (Eds.). (2015b). Stereo Timeline. In *Introduction: Living Stereo*. In *Living Stereo: Histories and Cultures of Multichannel Sound* (pp. 267–272). New York: Bloomsbury.
- Thompson, W. F., Russo, F. A., & Sinclair, D. (1994). Effects of underscoring on the perception of closure in filmed events. *Psychomusicology: A Journal of Research in Music Cognition*, 13(1–2), 9.
- THX Surround Sound Modes. (n.d.). Retrieved October 24, 2016, from <http://www.thx.com/consumer/thx-technology/thx-surround-sound-modes/>
- Tompkins, J. (2014). Mellifluous Terror: The Discourse of Music and Horror Films. In Ha. M. Benshoff (Ed.), *A Companion to the Horror Film* (pp. 185–204). Chichester: John Wiley & Sons.

- Townsend, L. M. (1931). Some Experiences in Adapting Theaters for Sound Reproduction. *Journal of the Society of Motion Pictures Engineers*, 16(5), 600–602.
- Trębacz, E. (2012). Depth Modulation: Composing motion in immersive audiovisual spaces. *Organised Sound*, 17(02), 156–162. <https://doi.org/10.1017/S1355771812000088>
- Truax, B. (1996). Soundscape, acoustic communication and environmental sound composition. *Contemporary Music Review*, 15(1–2), 49–65.
- Truax, B. (1998a). Composition and diffusion: space in sound in space. *Organised Sound*, 3(2), 141–146.
- Truax, B. (1998b). Models and strategies for acoustic design. Presented at the International Conference on Acoustic Ecology “Stockholm, Hey Listen!,” Stockholm.
- Truax, B. (2006). Acoustic space, architecture and acoustic ecology. In *Proceedings of International Cross-Disciplinary Conference Ryerson University*. Toronto.
- Truax, B. (2012). Music, Soundscape and Acoustic Sustainability. *Moebius Journal*, 1(1).
- Truax, B., & Barrett, G. W. (2011). Soundscape in a context of acoustic and landscape ecology. *Landscape Ecology*, 26(9), 1201–1207. <https://doi.org/10.1007/s10980-011-9644-9>
- Tversky, B. (2005). Functional significance of visuospatial representations. In P. Shah & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 1–34). Cambridge: Cambridge University Press.
- Valentini, P. (2006). *Il suono nel cinema: storia, teoria e tecniche*. Venezia: Marsilio.
- Van Baelen, W., Bert, T., Claypool, B., & Sinnaeve, T. (2012). Auro 11.1: A new dimension in cinema sound. Barco. Retrieved from [http://www.barco.com/securedownloads/cd/MarketingKits/3dsound/White%20papers/Auro%2011.1\\_A%20new%20dimension%20in%20cinema%20sound.pdf](http://www.barco.com/securedownloads/cd/MarketingKits/3dsound/White%20papers/Auro%2011.1_A%20new%20dimension%20in%20cinema%20sound.pdf)
- Van Daele, B., & Van Baelen, W. (2012, February 28). Productions in Auro-3D: Professional workflow and costs (Rev. 0.6). Barco. Retrieved from <http://www.auro-3d.com/wp-content/uploads/documents/Professional-Workflow-White-Paper-v0-6-20120228.pdf>
- Van Nort, D. (2013). A Collaborative Approach to Teaching Sound Sculpting, Embodied Listening and the Materiality of Sound. *Organised Sound*, 18(02), 207–217. <https://doi.org/10.1017/S1355771813000125>
- Verscheure, J.-P. (1995). The Challenge of Sound Restoration from 1927 to Digital. *Film History*, 7(3), 264–276.
- Vidolin, A. (2009). Conservazione e restauro del patrimonio culturale della musica elettronica. L'esempio degli Archivi Luigi Nono. In *Atti della giornata di studi Patrimoine musical du*

- XXe siècle* (pp. 45–56). Cité de la Musique, Paris. Retrieved from [http://www.citedelamusique.fr/pdf/insti/recherche/patrimoinexx/vidolin\\_it.pdf](http://www.citedelamusique.fr/pdf/insti/recherche/patrimoinexx/vidolin_it.pdf)
- Village Cinema City. (n.d.). Sydney in 70mm – Part 3: The Multiplex. Pitt Centre, Village Cinema City. Retrieved from [http://in70mm.com/library/70mm\\_in/australia/sydney/pdf/village\\_cinema.pdf](http://in70mm.com/library/70mm_in/australia/sydney/pdf/village_cinema.pdf)
- Võ, M. L.-H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, *12*(13), 3–3.
- Vroomen, J., & de Gelder, B. (2004). Temporal Ventriloquism: Sound Modulates the Flash-Lag Effect. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 513–518. <https://doi.org/10.1037/0096-1523.30.3.513>
- Wang, Y. F., Petrina, S., & Feng, F. (2017). VILLAGE—Virtual Immersive Language Learning and Gaming Environment: Immersion and presence. *British Journal of Educational Technology*, *48*(2), 431–450.
- Ward, M. (2010). Voice, Videogames, and the Technologies of Immersion. In *Voice: Vocal Aesthetics in Digital Arts and Media* (pp. 267–279). Cambridge: MIT Press. Retrieved from [http://www.academia.edu/download/39481564/Voice\\_\\_videogames\\_\\_and\\_\\_the\\_\\_technologies\\_\\_of\\_\\_immersion\\_\\_Mark\\_\\_S\\_\\_Ward\\_\\_2010\\_\\_penultimate\\_version\\_.pdf](http://www.academia.edu/download/39481564/Voice__videogames__and__the__technologies__of__immersion__Mark__S__Ward__2010__penultimate_version_.pdf)
- Ward, M. (2015). Art in Noise: An Embodied Simulation Account of Cinematic Sound Design. In M. Coëgnarts & P. Kravanja (Eds.), *Embodied Cognition and Cinema* (pp. 155–186). Leuven: Leuven University Press.
- Webb, J., & Ketteler, D. (1980). Using the Multi-Track Format for Production Film Recording. *Recording Engineer/Producer*, *11*(2), 110–117.
- Weis, E., & Belton, J. (1985). *Film sound: Theory and practice*. New York: Columbia University Press.
- Westerkamp, H. (2002). Linking soundscape composition and acoustic ecology. *Organised Sound*, *7*(01). <https://doi.org/10.1017/S1355771802001085>
- Whittington, W. (2007). *Sound design and science fiction*. University of Texas Press.
- Whittington, W. (2011). Contemporary Film Technology. In J. Nelmes (Ed.), *Introduction to Film Studies* (5th ed., pp. 42–58). London: Routledge.
- Wierzbicki, J. (2016). Narrative cinema’s “sounds of silence”: variations on the POA. *Sound Studies*, *2*(2), 137–150.
- Wierzbicki, J. (Ed.). (2012). *Music, sound and filmmakers: sonic style in cinema*. New York: Routledge.
- Williams, A. (1980). Is sound recording like a language? *Yale French Studies*, (60), 51–66.

- Winters, B. (2010). The non-diegetic fallacy: Film, music, and narrative space. *Music and Letters*, 91(2), 224–244.
- Wolf, S. K. (1934). Acoustical Requirements for Wide-Range Reproduction of Sound. *Journal of the Society of Motion Pictures Engineers*, 22(4), 242–253.
- Wood, N. (1984). Towards a semiotics of the transition to sound: spatial and temporal codes. *Screen*, 25(3), 16–25.
- Wright, B. (2005). *Making Films Sound Better: The Transition to Dolby Sound in Hollywood Cinema* (MA thesis). Carleton University, Ottawa.
- Wright, B. (2011). *Sound from Start to Finish: Professional Style and Practice in Modern Hollywood Sound Production* (PhD thesis). Carleton University, Ottawa.
- Wright, B. (2015). Atmos Now: Dolby Laboratories, Mixing Ideology and Hollywood Sound Production. In P. Théberge, Devine, Kyle, & Everett, Tom (Eds.), *Living Stereo: Histories and Cultures of Multichannel Sound* (pp. 227–246). New York: Bloomsbury.
- Wurtzler, S. (1992). She sang live, but the microphone was turned off: the live, the recorded, and the subject of representation. In R. Altman (Ed.), *Sound Theory/Sound Practice* (pp. 87–103). New York: Routledge.
- Wysotsky, M. Z. (1971). *Wide-Screen Cinema and Stereophonic Sound*. (R. Spottiswoode, Ed., A. E. C. York, Trans.). New York: Hastings House.
- Yacavone, D. (2012). Spaces, Gaps, and Levels: From the Diegetic to the Aesthetic in Film Theory. *Music, Sound, and the Moving Image*, 6(1), 21–37.
- Zacks, J. M. (2015). *Flicker: Your brain on movies*. Oxford: Oxford University Press.
- Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, 138(2), 307.
- Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, 1076(1), 150–162.
- Zhou, Z., Cheok, A. D., Qiu, Y., & Yang, X. (2007). The role of 3-D sound in human reaction and performance in augmented reality environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(2), 262–272.

## Filmografia

N.B.: i film sono elencati automaticamente in ordine alfanumerico, in base al primo carattere del titolo italiano o di quello originale in assenza di distribuzione italiana.

*2001: Odissea nello spazio (2001: A Space Odyssey, USA/Regno Unito 1968)*

*360° Camera In Haunted House [4K Quality With 3D Surround Sound],*  
[https://www.youtube.com/watch?v=vQeN\\_YxstTo](https://www.youtube.com/watch?v=vQeN_YxstTo) (24/10/2016)

*360° Horror: The Ring VR Experience,* <https://www.youtube.com/watch?v=SsFZjS19Zu8>  
 (24/10/2016)

*3D Spatial Audio : Trevi Fountain sound (360 Video),* <https://www.youtube.com/watch?v=oB1up7vRN3Y> (24/10/2016)

*8 ½ (Italia/Francia 1963)*

*Abbandonati nello spazio (Marooned, USA 1969)*

*Alien (Id., USA 1979)*

*All'ovest niente di nuovo (All Quiet on the Western Front, USA 1930)*

*Alla ricerca di Dory (Finding Dory, USA 2016)*

*Apocalypse Now (Id., USA 1979)*

*Applause (USA 1929)*

*Arancia meccanica (A Clockwork Orange, Regno Unito/USA 1971)*

*Atmosfera zero (Outland, USA 1981)*

*Avengers: Age of Ultron (Id., USA 2015)*

*Batman – Il ritorno (Batman Returns, USA 1992)*

*Battaglie nella galassia (Battlestar Galactica, USA 1978)*

*Bianco Natale (White Christmas, USA 1954)*

*Blade Runner 2049 (Id., USA 2017)*

*Blow Out (Id., USA 1981)*

*Carousel (Id., USA 1956)*

*Cars – Motori ruggenti (Cars, Usa 2000)*

*Cento uomini e una ragazza (One Hundred Men and a Girl, USA 1937)*

*Come sposare un milionario (How to Marry a Millionaire, USA 1953)*

*Complesso di colpa (Obsession, USA 1976)*

*Crocevia della morte (Miller's Crossing, USA 1990)*

*Dick Tracy* (*Id.*, USA 1990)

*Doom 3 in full-blown AstoundSound® (Headphone Demo)*, <https://www.youtube.com/watch?v=mb57qpc0RD0> (24/10/2016)

*È nata una stella* (*A Star is Born*, USA 1976)

*Edward mani di forbice* (*Edward Scissorhands*, USA 1990)

*Elephant* (*Id.*, USA 2003)

*Enrico V* (*Henry V*, Regno Unito 1944)

*Eyes Wide Shut* (*Id.*, Regno Unito/USA 1999)

*Falstaff* (*Falstaff (Chimes at Midnight)*, Spagna/Svizzera 1965)

*Fantasia* (*Id.*, USA 1940)

*Femme Fatale* (*Id.*, Francia 2002)

*Galactica: l'attacco dei Cylon* (*Mission Galactica: The Cylon Attack*, USA 1979)

*Giorni di tuono* (*Days of Thunder*, USA 1990)

*Gli uccelli* (*The Birds*, USA 1963)

*Gorillas in the Congo: VR with Spatial Audio*, <https://www.youtube.com/watch?v=Db-ug08ydI4> (24/10/2016)

*Gravity* (*Id.*, Regno Unito/USA 2013)

*Guerre stellari* (*Star Wars*, USA 1977)

*Hair* (*Id.*, USA 1979)

*Haunting – Presenze* (*The Haunting*, USA 1999)

*I cancelli del cielo* (*Heaven's Gate*, USA 1980)

*I pascoli dell'odio* (*Santa Fe Trail*, USA 1940)

*Il buono, il brutto, il cattivo* (Italia/Germania Ovest/Spagna/USA 1966)

*Il cacciatore* (*The Deer Hunter*, USA 1978)

*Il colore del crimine* (*Freedomland*, USA 2006)

*Il dittatore dello stato libero di Bananas* (*Bananas*, USA 1971)

*Il drago del lago di fuoco* (*Dragonslayer*, USA 1981)

*Il fantasma del palcoscenico* (*Phantom of the Paradise*, USA 1974)

*Il pianeta proibito* (*Forbidden Planet*, USA 1956)

*Il piccolo principe* (*The Little Prince*, Regno Unito/USA 1974)

*Il ritorno dello Jedi* (*The Return of the Jedi*, USA 1983)

*Il silenzio degli innocenti* (*The Silence of the Lambs*, USA 1991)

*Incontri ravvicinati del terzo tipo* (*Close Encounters of the Third Kind*, USA 1977)

*Independence Day – Rigenerazione* (*Independence Day: Resurgence*, USA 2016)

- Indiana Jones e l'ultima crociata (Indiana Jones and the Last Crusade, USA 1989)*
- Inside Out (Id., USA 2015)*
- It Follows (Id., USA 2013)*
- Jurassic Park (Id., USA 1993)*
- Kong: Skull Island (Id., USA 2017)*
- Kwaidan (Kaidan, Giappone 1964)*
- L'eclisse (Italia 1962)*
- L'ultima odissea (Damnation Alley, USA 1977)*
- L'uomo che sapeva troppo (The Man Who Knew Too Much, USA 1956)*
- L'uomo d'acciaio (Man of Steel, USA 2013)*
- La battaglia di Midway (Midway, USA 1976)*
- La casa (Evil Dead, USA 2013)*
- La casa (The Evil Dead, USA 1981)*
- La conquista del West (How the West Was Won, USA 1962)*
- La febbre del successo – Jazz Singer (The Jazz Singer, USA 1980)*
- La fiamma del peccato (Double Indemnity, USA 1946)*
- La finestra sul cortile (Rear Window, USA 1954)*
- La fuga di Logan (Logan's Run, USA 1976)*
- La maledizione di Damien (Damien: Omen II, USA 1978)*
- La maschera di cera (House of Wax, USA 1953)*
- La mummia (The Mummy, USA 2017)*
- La tunica (The Robe, USA 1953)*
- Last Action Hero – L'ultimo grande eroe (Last Action Hero, USA 1993)*
- Lawrence d'Arabia (Lawrence of Arabia, Regno Unito/USA 1964)*
- Le verità nascoste (What Lies Beneath, USA 2000)*
- Les yeux ne veulent pas en tout temps se fermer ou Peut-être qu'un jour Rome se permettra de choisir à son tour (Germania Ovest/Francia 1969)*
- Lezioni di storia (Geschichtsunterricht, Italia/Germania Ovest 1972)*
- Lilli e il vagabondo (Lady and the Tramp, USA 1955)*
- Linea mortale (Flatliners, USA 1990)*
- Lisztomania (Id., Regno Unito 1975)*
- Macbeth (Id., Regno Unito/USA 1971)*
- Macbeth (Id., USA 1948)*
- Maze Runner – Il labirinto (The Maze Runner, USA 2014)*

*mother! (madre!, USA 2017)*  
*Nashville (Id., USA 1975)*  
*Non aprite quella porta (The Texas Chain Saw Massacre, USA 1974)*  
*Oklahoma! (Id., USA 1955)*  
*Parigi di Notte (Paris vu par, Francia 1965)*  
*Passion (Francia/Germania/Spagna/Regno Unito 2012)*  
*Persona (Id., Svezia 1966)*  
*Philharmonic 360 video with spatial audio - Manuel de Falla, <https://www.youtube.com/watch?v=8ESEI0bqrJ4> (24/10/2016)*  
*Pink Floyd The Wall (Id., Regno Unito 1982)*  
*Playtime (Id., Francia 1967)*  
*Poliziotto sadico (Maniac Cop, USA 1988)*  
*Popeye – Braccio di ferro (Popeye, USA 1980)*  
*Prénom Carmen (Id., Francia 1983)*  
*Psyco (Psycho, USA 1960)*  
*Quando la moglie è in vacanza (The Seven Year Itch, USA 1955)*  
*Quarto potere (Citizen Kane, USA 1940)*  
*Quelli della San Pablo (The Sand Pebbles, USA 1966)*  
*Ragazze interrotte (Girl, Interrupted, USA 1999)*  
*Ratatouille (Id., USA 2007)*  
*Red Tails (USA 2012)*  
*Ribelle – The Brave (Brave, USA 2012)*  
*Riccardo III (Richard III, Regno Unito 1955)*  
*RoboCop (Id., USA 1987)*  
*Rollercoaster – Il grande brivido (Rollercoaster, USA 1977)*  
*Romeo e Giulietta (Romeo and Juliet, Italia/USA 1968)*  
*Rotta verso l'ignoto (Star Trek VI: The Undiscovered Country, USA 1991)*  
*Rotta verso la terra (Star Trek IV – The Voyage Home, USA 1986)*  
*Salto nel buio (Innerspace, USA 1987)*  
*Salvate il soldato Ryan (Saving Private Ryan, USA 1998)*  
*Salvo (Italia/Francia 2013)*  
*San Siro (Italia 2014)*

- Sexteto Fantasma VR 360 - Canaro en París - Audio 3D*, [https://www.youtube.com/watch?v=uayp7VRoL3Y&list=PLU8wpH\\_LfhmvCvcBGui3LHC8DufjgvxNn&index=1](https://www.youtube.com/watch?v=uayp7VRoL3Y&list=PLU8wpH_LfhmvCvcBGui3LHC8DufjgvxNn&index=1)  
(24/10/2016)
- Shining (The Shining)*, USA/Regno Unito 1980)
- Shutter Island (Id.)*, USA 2010)
- Stalker (Id.)*, URSS 1979)
- Star Wars – Il risveglio della forza (Star Wars: Episode VII – The Force Awakens)*, USA 2015)
- Star Wars: Episodio I – La minaccia fantasma (Star Wars: Episode I – The Phantom Menace)*, USA 1999)
- Stati di allucinazione (Altered States)*, USA 1980)
- Strange Days (Id.)*, USA 1995)
- Superman (Id.)*, USA 1978)
- Superman II (Id.)*, USA 1981)
- Surf's Up – I re delle onde (Surf's Up)*, USA 2007)
- Terremoto (Earthquake)*, USA 1974)
- The Bourne Supremacy (Id.)*, USA 2004)
- The Resurgence of Jazz in New Orleans*, [https://www.youtube.com/watch?v=dFq9wbrV-v0&list=PLU8wpH\\_LfhmvCvcBGui3LHC8DufjgvxNn&index=6](https://www.youtube.com/watch?v=dFq9wbrV-v0&list=PLU8wpH_LfhmvCvcBGui3LHC8DufjgvxNn&index=6) (24/10/2016)
- The Ring (Id.)*, USA 2002)
- The Sound of Brave* (2012), <http://vimeo.com/44754809> (16/6/2017)
- The Sound of Hitchcock* (USA 2008)
- The Sound of Man of Steel*, <http://soundworkscollection.com/videos/the-sound-of-man-of-steel>  
(16/6/2017)
- The Terminal (Id.)*, USA 2004)
- The Wacky Wabbit* (USA 1942)
- The Wiz (Id.)*, USA 1978)
- The Wolf of Wall Street (Id.)*, USA 2013)
- This is Cinerama* (USA 1952)
- Timecode* (USA 2000)
- Tommy (Id.)*, Regno Unito 1975)
- Tora! Tora! Tora! (Id.)*, Giappone/USA 1970)
- Toy Story 3 – La grande fuga (Toy Story 3)*, USA 2010)
- Tron (Id.)*, USA 1982)
- Tutti pazzi per Mary (There's Something About Mary)*, USA 1998)

*Un amore splendido (An Affair to Remember, USA 1958)*

*Uragano (Hurricane, USA 1979)*

*Veloce come il vento (Italia 2016)*

*Viaggio al centro della terra (Journey to the Center of the Earth, USA 1959)*

*Vincitori e vinti (Judgement at Nuremberg, USA 1961)*

*We Were Soldiers (Id., USA 2002)*

*West Side Story (Id., USA 1961)*

*Windjammer: The Voyage of the Christian Radich (USA 1958)*

*Wings of Courage (USA/Francia 1995)*

*Wolfen – La belva immortale (Wolfen, USA 1981)*

*Youth – La giovinezza (Youth, Italia/Francia/Regno Unito/Svizzera 2015)*

*Zardoz (Id., Irlanda/USA 1974)*

*Zoot Suit (USA, 1981)*