

# **UNIVERSITÀ DEGLI STUDI DI GENOVA**

Dottorato in Digital Humanities - Arte, Spettacolo e Tecnologie Multimediali  
Ciclo XXXIV



Tesi di Dottorato

## **MUSICA IMMERSIVA IN FORMATO BINAURALE**

RELATORE

Chiar.mo Prof.re Gianni Viardo Vercelli

Chiar.mo Prof.re Mauro Coccoli

CANDIDATO Andrea De Sotgiu

ANNO ACCADEMICO 2021/2022

## Abstract

La tesi si propone di studiare la percezione del senso di immersività nella musica in formato binaurale, con l'obiettivo di migliorare l'esperienza d'ascolto e metterla a confronto con la musica ascoltata in formato stereofonico.

Attraverso un'analisi delle basi della fisica del suono e della psicoacustica, il saggio vuole illustrare quali sono le tecnologie che permettono alla musica immersiva di essere prodotta. A tal proposito verranno presi in esame sia la storia delle tecniche di riproduzione sonora, che gli strumenti attualmente disponibili, per permettere alla comunità scientifica di avere un quadro completo sulle risorse odierne.

Il binomio tra musica e spatial audio è in continua crescita con il passare del tempo, per questo il testo si propone anche di evidenziare quali sono i principali problemi di ricerca e gli standard necessari sui quali concentrare eventuali prossimi lavori.

Il saggio si conclude con un esperimento che mette a confronto due differenti tipologie di mix di una canzone: una versione stereofonica e una versione in formato binaurale, processata tramite la Digital Audio Workstation Logic Pro X e la Dolby Atmos Production Suite.

I risultati porteranno delle evidenze interessanti a favore della versione in formato binaurale e diverse suggestioni sugli sviluppi futuri.

The thesis aims to study the perception of the sense of immersion in music in binaural format, with the target of improving the listening experience and comparing it with the music listened to in stereophonic format.

Through an analysis of the basics of the physics of sound and psychoacoustics, the essay wants to illustrate which technologies allow immersive music to be produced. In this regard, both the history of sound reproduction techniques and the tools currently available will be examined to allow the scientific community to have a complete picture of today's resources.

The combination of music and spatial audio is constantly growing with the passage of the years, which is why the text also aims to highlight the main research problems and the necessary standards on which to focus any future works.

The essay ends with an experiment that compares two different types of song mixes: a stereo version and a binaural format version, processed through the Digital Audio Workstation Logic Pro X and the Dolby Atmos Production Suite.

The results will bring interesting evidence in favor of the binaural format version and various suggestions on future developments.

*A Ilaria.  
Al mio orecchio destro.*

# INDICE

<b>ABSTRACT</b>	<b>2</b>
<b>INTRODUZIONE</b>	<b>5</b>
PUBBLICAZIONI	7
STRUTTURA DELLA TESI	8
<b>CAPITOLO 1 – IL SUONO</b>	<b>9</b>
1.1 LE PROPRIETÀ DEL SUONO	9
1.2 LA FREQUENZA E IL PERIODO	10
1.3 LA LUNGHEZZA D’ONDA	13
1.4 AMPIEZZA	14
1.5 FASE	14
1.6 LA VELOCITÀ DEL SUONO	15
1.7 INVILUPPO ADSR	16
1.8 FORME D’ONDA ELEMENTARI	17
1.9 DECIBEL E LUFS	18
1.10 IL SUONO DIGITALE	20
1.11 LA FREQUENZA DI CAMPIONAMENTO E LA QUANTIZZAZIONE	20
1.12 STRUMENTI DI ANALISI DELLE FREQUENZE: ANALIZZATORE DI SPETTRO	22
1.13 PROCESSORI DI SEGNALE: L’EQUALIZZATORE	23
1.14 PROCESSORI DI SEGNALE: IL COMPRESSORE	26
<b>CAPITOLO 2 - L’ASCOLTO BINAURALE</b>	<b>33</b>
2.1 L’ORECCHIO UMANO	33
2.2 LE CURVE ISOFONICHE	35
2.2 HEAD RELATED TRANSFER FUNCTION	36
2.3 EFFETTI E FENOMENI PSICOACUSTICI	39
2.3.1 <i>Il fenomeno dei battimenti</i>	39
2.3.2 <i>Effetto Haas - Legge del primo fronte d’onda</i>	39
2.3.3 <i>Mascheramento</i>	39
2.3.4 <i>Ricostruzione della fondamentale</i>	40
2.3.5 <i>Effetto “cocktail party”</i>	40
2.3.6 <i>Illusioni acustiche</i>	40
2.4 LA PSICOLOGIA DELLA MUSICA	41
2.5 STORIA DELLE PRINCIPALI TECNICHE DI RIPRODUZIONE SONORA	42
2.5.1 <i>Gli albori dell’industria discografica e la monofonia</i>	42
2.5.2 <i>Le sperimentazioni e la celebrazione della stereofonia</i>	46
2.5.3 <i>La rovina della quadrifonia e l’Ambisonics</i>	52
2.5.4 <i>L’avvento del surround e la Dolby-mania</i>	57
2.5.5 <i>Il nuovo millennio, la musica liquida e l’estensione massima del surround</i>	61
2.5.6 <i>Dolby Atmos: l’audio spaziale diventa una realtà commerciale</i>	63
2.6. AUDIO SPAZIALE IN CUFFIA	67
<b>CAPITOLO 3 - IMMERSIVE MUSIC PRODUCTION</b>	<b>76</b>
3.1 DAW - DIGITAL AUDIO WORKSTATION	77
3.2 SPAZIALIZZATORI VIRTUALI PER LA MUSICA IMMERSIVA	78
3.2.1 <i>Formati</i>	79
3.2.2 <i>Monitoring</i>	80
3.2.3 <i>HRTF Database Sets</i>	81
3.2.4 <i>Software e virtual spatializers</i>	82

3.3. LE FASI DELLA PRODUZIONE NELLA MUSICA IMMERSIVA	96
3.3.1 <i>Pre-produzione</i>	96
3.3.2 <i>Produzione</i>	99
3.3.3 <i>Mixing</i>	100
3.3.4 <i>Mastering</i>	103
3.4 CONSEGUENZE COMMERCIALI	105
<b>CAPITOLO 4 - CONFRONTO TRA STEREOFONIA E FORMATO BINAURALE NELLA MUSICA IMMERSIVA ASCOLTATA IN CUFFIA</b>	<b>108</b>
4.1 RESEARCH QUESTIONS	109
4.2 METODOLOGIA: MIXING STEREO E MIXING BINAURALE	110
4.3 METODOLOGIA: IL QUESTIONARIO	114
4.4 RISULTATI	117
4.5 CONCLUSIONI E SVILUPPI FUTURI	122
<b>CONCLUSIONI</b>	<b>123</b>
<b>APPENDICE - REPOSITORY</b>	<b>124</b>
<b>BIBLIOGRAFIA</b>	<b>125</b>

## **Introduzione**

*Non c'è udito non spaziale.*

Jens Peter Blauert

New York, ottobre 2019. Entro all'interno dello Jacob K. Javits Convention Center consapevole di essere, per la prima volta nella mia vita, da solo dall'altra parte del mondo. Tre giorni all'insegna dell'edizione numero 147 dell'Audio Engineering Society Pro Audio Convention. Una delle prime session alla quale partecipo è il panel diretto dagli Abbey Road Studios di Londra in merito all'audio spaziale all'interno del contesto musicale, esattamente il mio argomento di ricerca. In una delle slide viene mostrata un'immagine di uno studio di ripresa sonora in cui erano stati disposti una miriade di microfoni con differenti tecniche di microfonaione. L'intento era capire quale fosse il miglior modo per riprendere la spazialità di una stanza.

Gli speaker continuano a raccontare gli studi che stanno conducendo, io nel frattempo prendo appunti ascoltando ciò che stanno dicendo: "Non esistono ancora standard in questo nuovo settore. Stiamo cercando di arrivare ad un punto comune, ma ci vorrà ancora tempo". Appoggio la penna e mi tranquillizzo. Se anche gli Abbey Road Studios stanno cercando di capire come orientarsi in questo marasma di informazioni significa che sono sulla strada giusta.



*Figura 1 - AES 147th Convention, New York, 2019.*

Mi sono chiesto spesso quale fosse la forma migliore per scrivere questa tesi di dottorato. La mia formazione è un ibrido in costante mutamento: esperienza pluriennale nel giornalismo e nella radiofonia; laureato magistrale in Digital Humanities - Comunicazione e Nuovi Media presso l'Università di Genova, con una tesi di laurea sulla ri-mediazione dei media sonori (e annesso approfondimento sullo spatial audio e sul Dolby Atmos); Corso di specializzazione in Urban Music Production conseguito

presso SAE Institute di Milano. Musicista. Adesso mi trovo alla fine di questo lungo capitolo, concludendo il ciclo di studi di dottorato in Digital Humanities - Arte, Spettacolo e Tecnologie Multimediali, con una ricerca interamente dedicata alla spazializzazione sonora nel formato binaurale, inserita nel contesto musicale. Essere un “umanista digitale”, una figura nata da pochi anni in campo professionale e accademico, non significa essere un ingegnere o un tecnico, neppure essere un letterato o un filosofo, bensì avere tutte quelle nozioni utili per poter mediare tra queste figure. Qualora foste in una di queste classi di insegnamento, chiedo perdono in anticipo se doveste trovare alcune nozioni non particolarmente folte di tecnicismi, ma se c’è qualcosa che ho imparato facendo ricerca, e analizzando gli articoli che ho letto in questi tre anni, è che talvolta è molto difficile comprendere il duro lavoro di alcuni ricercatori solo perché la forma e il contenuto sono stati espressi con particolare rigore e freddezza. Per questo motivo ho deciso di rendere questo saggio il più scorrevole possibile alla lettura di chi se ne interesserà, nella consapevolezza di scrivere un testo accademico e non un romanzo.

Nell’introduzione di uno dei libri fondamentali per il mio argomento di ricerca, Spatial Audio di Francis Rumsey, vi è una frase che rappresenta alla perfezione le motivazioni alla base di questa materia:

*Everyday life is full of three-dimensional sound experiences.*  
Francis Rumsey - Spatial Audio, 2001.

Ho utilizzato queste parole per iniziare ogni lezione che ho condotto in merito alle mie ricerche e continuerò a usarla fino a quando farò questo lavoro. Ogni momento delle nostre giornate è scandito da suoni che ci circondano e questi suoni hanno delle caratteristiche spaziali. Il mio proposito è utilizzare le tecniche e la tecnologia a nostra disposizione, attraverso i software di manipolazione e spazializzazione sonora, per dare alla musica quelle caratteristiche spaziali che non può avere con la stereofonia. Sono consapevole del fatto che questo mio contributo sia una goccia in un oceano di informazioni che cresce ogni anno in maniera smisurata, ma al contempo ho capito anche che racconta un presente volto ad un futuro in cui gli ascoltatori avranno un nuovo modo di fruire e ascoltare la musica. Utenti che guarderanno al passato con “tenerezza”, sapendo che noi, per molti decenni, abbiamo ascoltato dalle nostre cuffie soltanto 135° di apertura spaziale limitata al piano orizzontale.

## **Pubblicazioni**

Nel corso del triennio di dottorato sono stati pubblicati tre articoli scientifici, scritti insieme ai professori che hanno curato questo lavoro e a stimati colleghi, che hanno contribuito alla creazione di questo saggio:

Bellanti, E., Corsi, A., De Sotgiu, A., & Vercelli, G. (2018, June 1). “Changes”: *An immersive spatial audio project based on low-cost open tools*. [https://doi.org/10.1007/978-3-319-95270-3\\_15](https://doi.org/10.1007/978-3-319-95270-3_15)

De Sotgiu, A., Coccoli, M., & Vercelli, G. (2020a). Tools for Immersive Music in Binaural Format. *Ambient Intelligence – Software and Applications*, 54–60. [https://doi.org/10.1007/978-3-030-58356-9\\_6](https://doi.org/10.1007/978-3-030-58356-9_6)

De Sotgiu, A., Coccoli, M., & Vercelli, G. (2020b). *Comparing the perception of “sense of presence” between a stereo mix and a binaural mix in immersive music*. Audio Engineering Society Convention 148. [http://www.aes.org/e-lib/browse.cfm?elib=20826&fbclid=IwAR3OBHYJGv1y6TlXrWaLx3qYwLxM7X4BTzQsAdTVUtBSTlFD8xra5CK\\_cs](http://www.aes.org/e-lib/browse.cfm?elib=20826&fbclid=IwAR3OBHYJGv1y6TlXrWaLx3qYwLxM7X4BTzQsAdTVUtBSTlFD8xra5CK_cs)

È stato fondamentale, come già descritto, anche il contributo dell’Audio Engineering Society e delle occasioni proposte da questa istituzione, la quale non solo mi ha dato la possibilità di incontrare esperti del settore ogni anno, creando dei confronti oltremodo utili, ma ha permesso anche di accrescere le mie conoscenze su questo ambito di ricerca grazie alla moltitudine di letteratura messa a disposizione.

## **Struttura della tesi**

La tesi verrà aperta da un capitolo interamente dedicato alle fondamenta della fisica del suono analogico e un approfondimento sulle basi inerenti al suono digitale. Tutte queste nozioni saranno centrali come supporto per tutte le trattazioni che verranno compiute nel corso del testo.

Il secondo capitolo tratterà i capisaldi dell’anatomia dell’ascolto, della percezione e della psicoacustica, per poi concentrarsi sulla storia delle tecniche di riproduzione sonora, dalla nascita del fonografo alle tecnologie odierne. Il legame tra anatomia dell’orecchio e tecnologia d’ascolto è un punto nevralgico per la musica immersiva e per l’ambito di ricerca dello spatial audio, motivo per cui in questa fase della tesi verrà anche definita la scelta di procedere approfondendo unicamente l’ascolto in cuffia.

Il terzo capitolo entrerà nel merito dell’immersività sonora e analizzerà tutti gli strumenti utili per la produzione di musica immersiva: dagli spazializzatori virtuali alle Digital Audio Workstation, che permettono la lavorazione di audio multicanale.

Verranno anche approfonditi i formati in uso e le tecniche di produzione (pre-produzione, produzione, mix e mastering) applicati alla musica immersiva in formato binaurale.

Infine, il quarto ed ultimo capitolo sarà a riguardo di un esperimento volto a dimostrare la possibilità, tra ascoltatori che non hanno un training specifico d’ascolto di contenuti in audio spaziale, di preferire la versione binaurale (in Dolby Atmos) a una versione stereofonica di una traccia musicale ascoltata in cuffia.



# CAPITOLO 1 – Il suono

Considerando il suono come un fenomeno generativo, questo capitolo approfondisce tutte le tematiche di base necessarie a comprendere i fenomeni fisici inerenti alle onde sonore. Essere a conoscenza dei concetti che seguiranno sarà centrale per poter operare all'interno di tutti gli ambiti legati al mondo dell'audio (registrazione, riproduzione, manipolazione e spazializzazione sonora).

## 1.1 Le proprietà del suono

Il suono è una variazione della pressione dell'aria ripetuta un certo numero di volte in un intervallo di tempo. L'aria è fondamentale per il suono: la collisione tra le sue particelle permette (tra loro) la trasmissione di una vibrazione attraverso una sorgente sonora, che genera la propagazione del suono in uno spazio (Marco Sacco, 2014).

Prendiamo un altoparlante, la classica sorgente sonora. Il movimento dei coni all'interno di un altoparlante spinge le particelle d'aria. A loro volta, le particelle adiacenti verranno colpite e spostate in avanti. Queste, muovendosi, creano una continua oscillazione propagandosi nello spazio e generando il suono.

Vedremo più avanti in che modo il nostro cervello percepisce questi cambiamenti di pressione. Prima di tutto è necessario approfondire quali siano fisicamente le proprietà del suono.

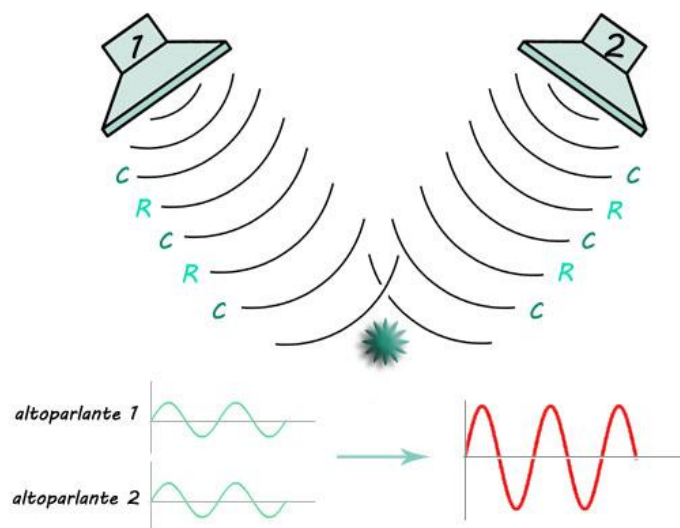


Figura 2 - Trasmissione delle onde sonore nell'aria (OpenProf.com)

Sono diverse le proprietà fisiche fondamentali del suono. In merito alle particelle d'aria citate precedentemente, questa oscillazione si può dire che avvenga a forma d'onda. Per questo motivo, il suono viene rappresentato fisicamente con diversi tipi di onde, più o meno complesse. La forma più semplice è la sinusoidale, la cui equazione è  $y = \sin x$ .

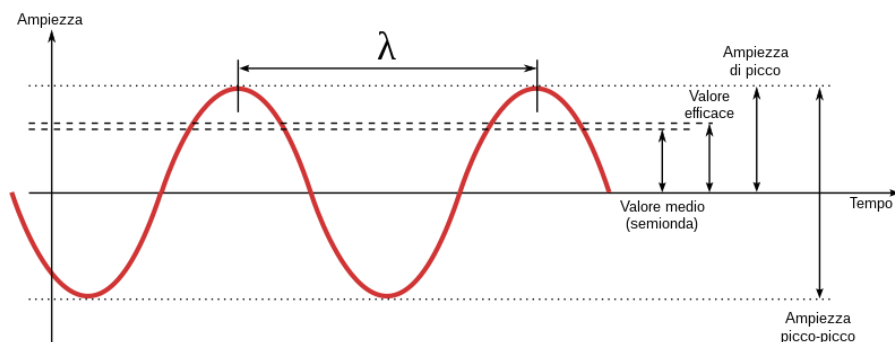


Figura 3 - Proprietà dell'onda sinusoidale (Wikipedia.com)

Questa sinusoide ha diverse proprietà: la frequenza ( $f$ ), il periodo ( $T$ ), la lunghezza ( $\lambda$ ), l'ampiezza ( $A$ ), la fase ( $\varphi$ ) e la velocità.

La forma d'onda di una sinusoide ha un suono molto morbido e profondo. Si può ascoltare, ad esempio, da strumenti come i sintetizzatori, creati appositamente dall'uomo per riprodurre e manipolare diverse forme d'onda. Un celebre esempio è il Minimoog inventato da Robert Moog negli anni '70, famoso ancora oggi per la profondità delle sue basse frequenze.



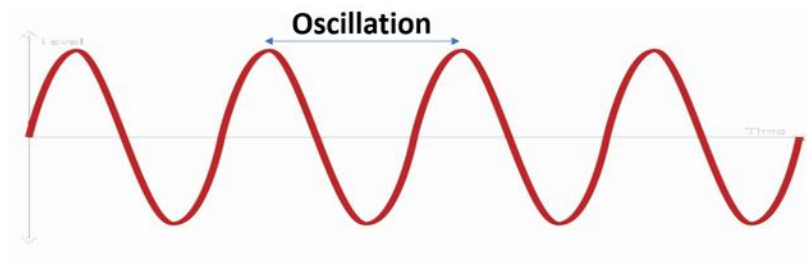
Figura 4 - Minimoog

Siccome la sinusoide viene anche identificata come “forma d'onda pura”, un altro suo utilizzo di rilievo è in campo medico, nella conduzione di esami acustici.

## 1.2 La frequenza e il periodo

La *frequenza* può essere descritta come il numero di cicli compiuti da un'onda in un secondo. Un ciclo è una semionda positiva più una semionda negativa. Un'onda viene misurata in Hertz (Hz). Un'onda di frequenza di 1 Hz compierà un ciclo completo in un secondo.

La sinusoide nella figura successiva rappresenta 5 Hz.



**Time Interval = 1 sec**  
**Number of Oscillation = 5**  
**Frequency = 5 Hz**

*Figura 5 - Onda con una frequenza di 5 Hz (Medium.com)*

Il nostro orecchio non percepisce tutte le frequenze che vengono emesse da qualsiasi sorgente, esiste un **campo di udibilità**, che va da 20 Hz a 20 kHz (20 000 Hz). Sotto i 20 Hz abbiamo le *sub frequenze*, che vengono percepite dal nostro corpo attraverso le vibrazioni, mentre sopra i 20 kHz ci sono gli *ultrasuoni*. L'invecchiamento, i traumi semplici (come ascoltare musica ad alto volume ripetutamente) o i traumi clinici abbassano la soglia di udibilità. A 30 anni, in base alle esperienze vissute, non verranno percepiti più né i 20 Hz né i 20 kHz, anche se nella normalità nessuno se ne accorge, spesso perché la musica che ascoltiamo viene filtrata da sorgenti non in grado di emettere queste frequenze. Le vecchie casse di un'auto decennale, le cuffiette in dotazione con il proprio cellulare di media fascia o le casse del nostro computer emettono musica che non arriva agli estremi della soglia di udibilità, eppure la percepiamo lo stesso.

# Instrument Frequency Chart

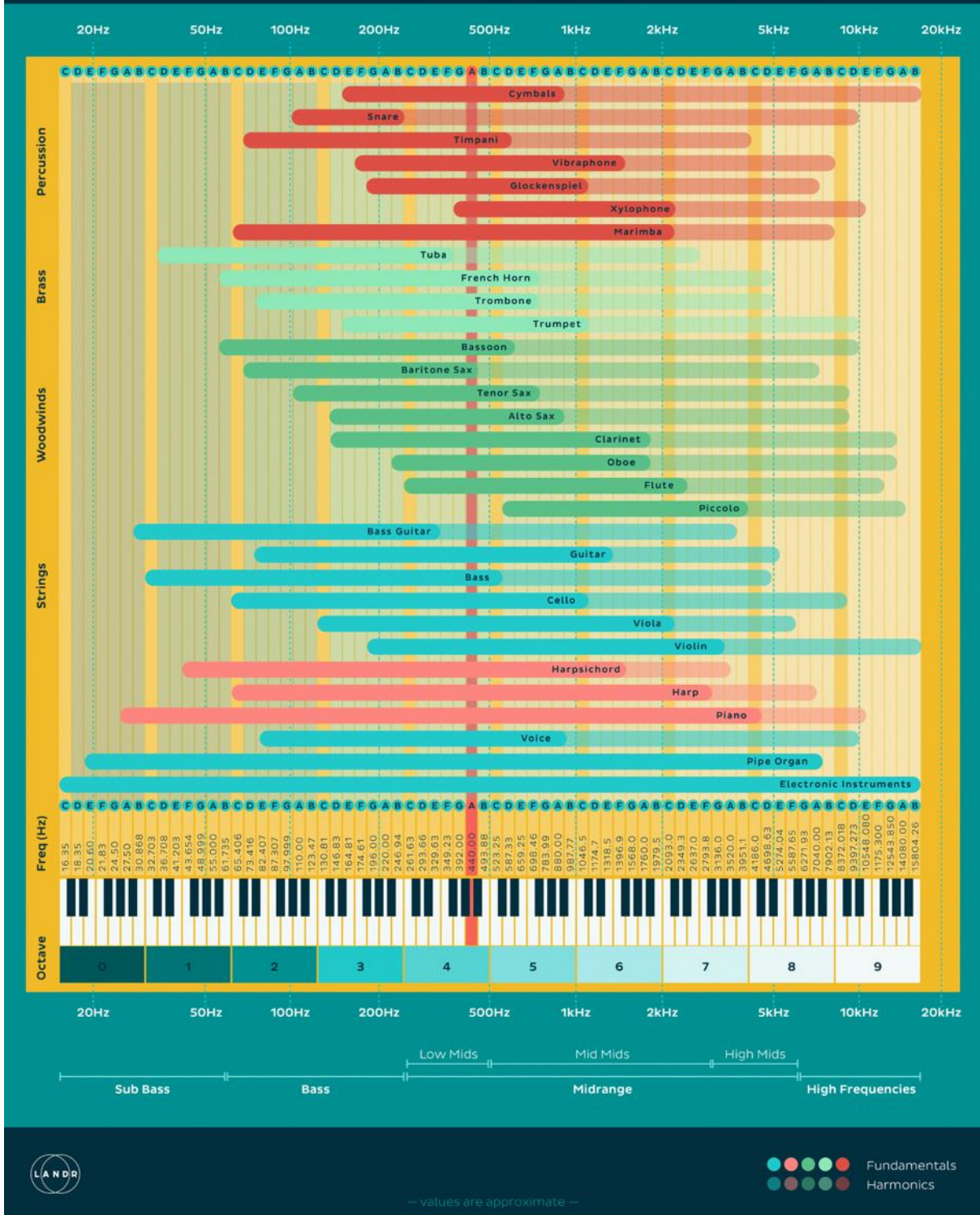


Figura 6 - Classificazione degli strumenti musicali in base alla soglia di udibilità (Landr.com)

Come descritto in questa immagine, ogni strumento musicale ha un range di frequenze che definisce un tono (grave, acuto, medioso). Tutto ciò che riguarda l'ambito musicale è un continuo lavoro con le frequenze. Le note stesse non sono altro che frequenze distinte all'interno di una scala musicale.

Prendiamo per esempio la frequenza 440 Hz, identificata come nota La, emessa dai diapason e dagli accordatori per dare un riferimento durante l'accordatura di una chitarra. A 440 Hz troveremo la frequenza *fondamentale* della nota, ovvero quella che dà il corpo a ciò che stiamo ascoltando. Successivamente, raddoppiando continuamente il valore della fondamentale (880 Hz - 1760 Hz - 3520 Hz ecc.) troveremo le sue *armoniche*: tutte quelle frequenze che caratterizzano e definiscono un suono.

Con il termine *periodo* si intende il tempo impiegato da un'onda per compiere un ciclo completo. La relazione corrisponde alla formula  $T = \frac{1}{f}$ .

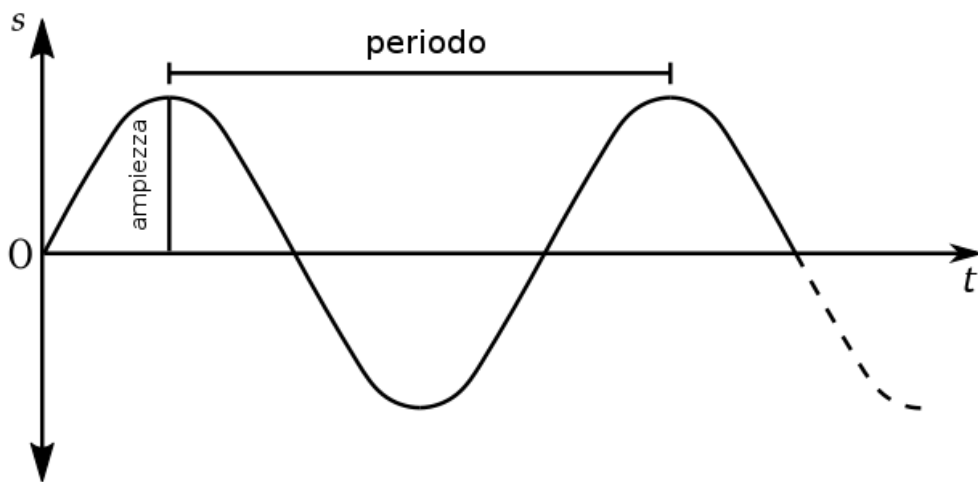
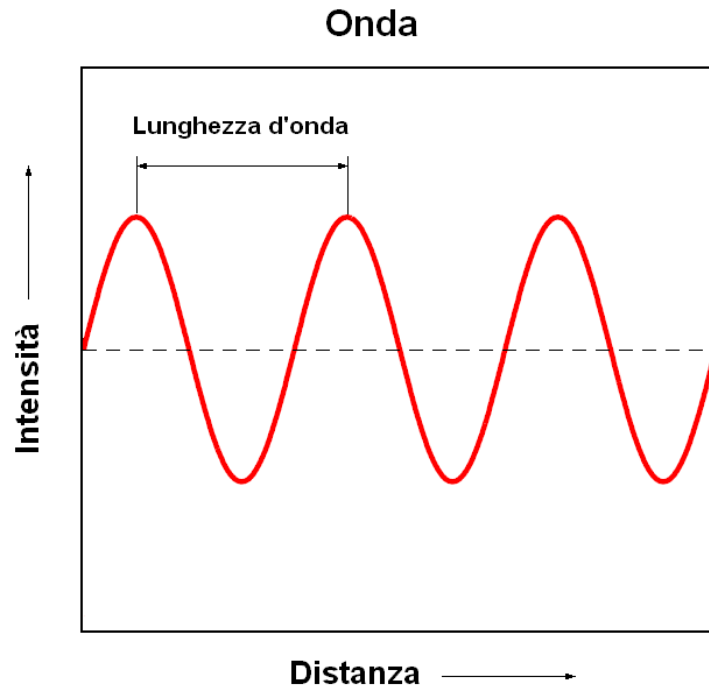


Figura 7 - Periodo dell'onda (Wikipedia.com)

### 1.3 La lunghezza d'onda

La distanza tra punti corrispondenti sull'onda (come gli apici delle semionde positive) è definita *lunghezza d'onda*, riassumibile con la formula  $\lambda = \frac{c}{f}$ , dove c sta per la velocità del suono nel mezzo che si sta considerando (ad esempio nell'aria è circa 344 m/sec). È importante considerare questa misura in un contesto inerente allo spazio/tempo. Le basse frequenze hanno una lunghezza d'onda maggiore delle alte, l'orecchio umano comincia a percepire suoni di frequenza superiore a 20 Hz a una distanza di 15-18 metri quando questi sono emessi all'aria aperta.



*Figura 8 - La lunghezza d'onda (Wikipedia)*

Questi concetti sulle lunghezze d'onda hanno un'applicazione nella nostra vita comune nei grandi eventi di musica dal vivo (stadi, parchi all'aperto o aeroporti), in cui gli ingegneri del suono devono calcolare dove posizionare le casse per permettere al pubblico di ascoltare un concerto senza alcun problema. Se questi calcoli dovessero essere errati le conseguenze potrebbero essere particolarmente fastidiose, come guardare un concerto dal terzo anello di uno stadio e ascoltare il proprio artista preferito con un ritardo di qualche secondo tra la voce e i suoi movimenti.

### 1.4 Ampiezza

Il termine *ampiezza* in ambito fisico è ciò che noi comunemente riconosciamo come volume.

Esistono due differenti tipologie di ampiezza: l'ampiezza di picco e l'ampiezza efficace. L'ampiezza di picco è una misura assoluta che definisce il punto più alto di un'onda; l'ampiezza efficace è quella percepita dall'orecchio (detta anche *valore quadratico medio* o RMS – Root Mean Square). Nel corso dei prossimi capitoli tutte le tematiche inerenti al volume sonoro verranno trattate ampiamente nel dettaglio.

### 1.5 Fase

Un ciclo completo di una forma d'onda sinusoidale coincide con un cerchio che compie 360°. La fase individua i punti dove passa la circonferenza, i quali corrispondono a determinati punti della sinusoide.

Una differenza di fase (o sfasamento) può essere vista come la distanza tra due punti, che ruotano alla stessa velocità (e alla stessa frequenza), ma che partono da posizioni

diverse. Lo sfasamento ha senso solo tra due punti che ruotano alla stessa velocità angolare. Due suoni sovrapposti, con la stessa frequenza, ma contrapposti di fase, generano il silenzio.

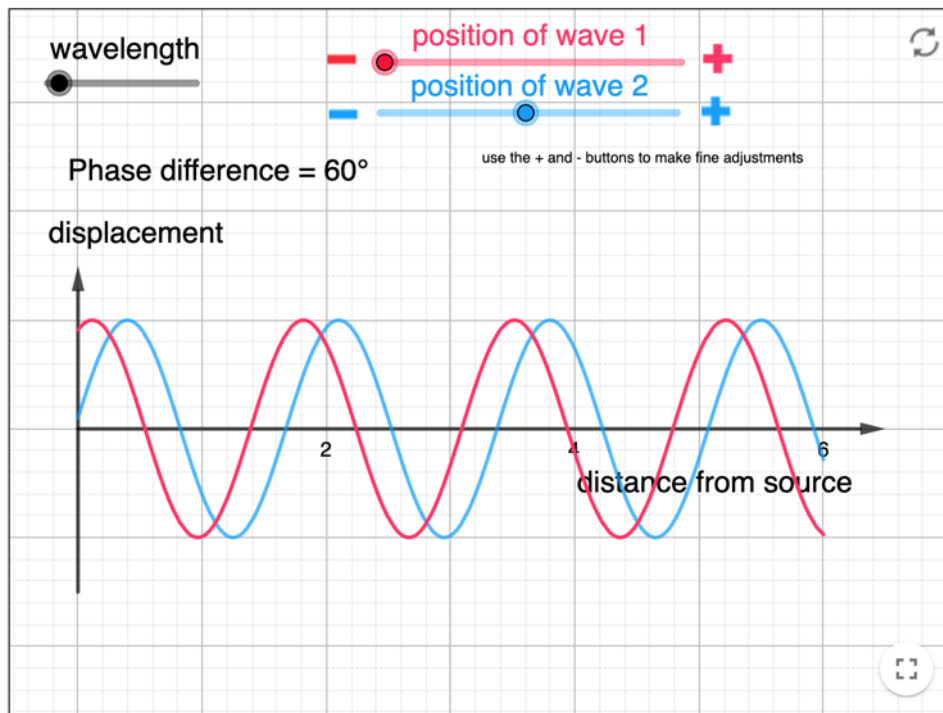


Figura 9 - Differenze di fase di  $60^\circ$  tra onda 1 (rossa) e onda 2 (blu) - (Geogebra.org<sup>1</sup>)

La fase è un concetto complesso, molto spesso uno scoglio da dover apprendere, ma fondamentale in molti ambiti di questo settore. Lo ritroviamo, ad esempio, come punto centrale delle tecniche di microfona e di registrazione. Per poter registrare strumenti musicali con più microfoni (ad esempio una batteria) sarà necessario assicurarsi che tutti i segnali ripresi simultaneamente creino meno problematiche possibili a livello di fase, per evitare un lungo lavoro di editing e di problem solving. Riprenderemo molto spesso la fase nel corso dei capitoli, con il passare degli esempi questo concetto sarà più semplice da assorbire.

## 1.6 La velocità del suono

Come già accennato la *velocità* del suono nell'aria è di 344 m/s. Più il mezzo ha una superficie rigida, più il suono si propaga velocemente. Ecco perché è difficile il trattamento acustico delle stanze, dove la rifrazione e la riflessione del suono sulle pareti deve essere calibrata per creare un'armonia con le frequenze. Ogni mezzo di trasmissione ha una sua velocità del suono considerando una temperatura costante di

<sup>1</sup> Utilizzando il seguente link è possibile interagire con le due onde e testare le differenze di fase: <https://www.geogebra.org/m/ymyxaejr>

23,24 °C. Al variare della temperatura, cambiano le caratteristiche del mezzo, quindi anche la velocità del suono. Si riscontra un aumento di velocità di 0,6 m/s per ogni incremento di 1 °C della temperatura del mezzo.

## 1.7 Inviluppo ADSR

Il suono ha una sua evoluzione di dinamiche in base all'intensità con cui vengono suonate le frequenze. Questa intensità viene chiamata anche ampiezza del suono e il suo andamento ha diverse connotazioni in base all'*inviluppo ADSR* (*Attack, Decay, Sustain, Release*).

*Attack*: quel momento che va dall'inizio del suono (attacco) al suo momento di ampiezza massima.

*Decay*: subito dopo l'apice, l'ampiezza incomincia a diminuire, con piccolo momento di decadenza.

*Sustain*: la decadenza del suono si stabilizza fino a raggiungere una soglia costante per un certo tempo.

*Release*: l'ampiezza diminuisce fino all'estinzione del suono.

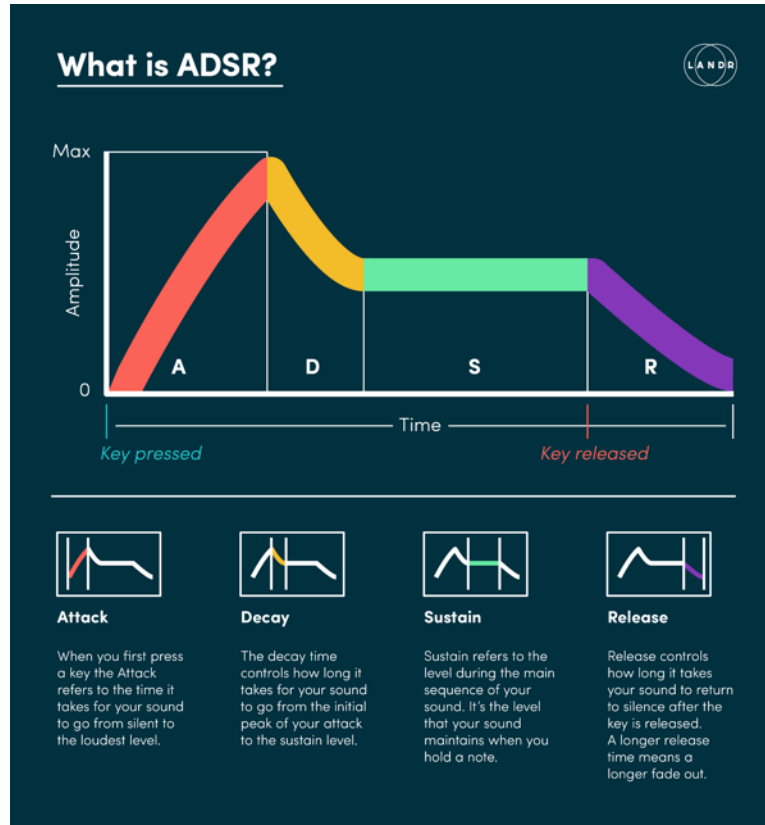


Figura 10 - Inviluppo ADSR



Immaginiamo di pizzicare la corda di una chitarra: le vibrazioni faranno esprimere alla corda una nota al suo massimo volume (rispetto alla forza utilizzata), poi, dopo una lieve diminuzione, la nota continuerà a suonare, fino alla sua decadenza. Questo è un esempio fisico di involuppo ADSR.

Un rullante avrà tempi di Attack e Decay molto brevi, un violino li avrà nettamente più lunghi. In fase di sound design, utilizzando sintetizzatori analogici o virtuali, è possibile agire sui parametri del sintetizzatore stesso per decidere quali saranno le tempistiche delle quattro fasi descritte, in modo da generare una dinamica ben precisa.

Per andare a modificare l'involuppo ADSR di un suono si ricorre talvolta all'utilizzo di effetti, come riverberi o echo, che possono cambiare totalmente l'atteggiamento di un suono. Si definisce *dry* un suono fedele a come è stato registrato, senza nessun effetto applicato; si definisce *wet* un suono effettato.

## 1.8 Forme d'onda elementari

Fino ad ora l'unica forma d'onda che “sembra essere” stata citata è la sinusoidale. In realtà, qualsiasi strumento musicale ha una sua forma d'onda molto complessa. Ne esistono diverse che vanno a complicarsi man mano per ogni strumento e diventano ancora più fitte nel momento in cui più strumenti vanno a sovrapporsi (parlando di musica, anche la voce è da considerare uno strumento musicale). Nella comunità scientifica sono diversi i ricercatori che lavorano con le forme d'onda e con l'intelligenza artificiale nel tentativo di ricreare fedelmente il suono di alcuni strumenti musicali tramite la sintesi. Per completezza, quindi, oltre alla sinusoidale, è necessario citare le altre forme d'onda elementari, alla base di tantissimi processi di ricostruzione sonora e di sound design (Cipriani, Giri, 2019):

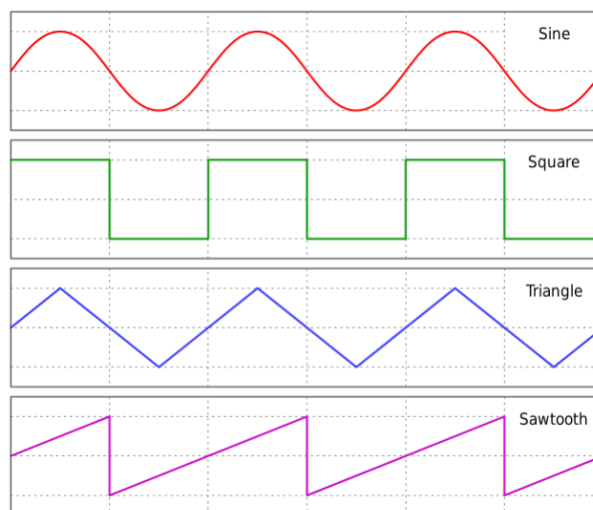


Figura 11 - Forme d'onda elementari (Wikipedia.com)

*Onda quadra (square wave)*: ripropone un suono più “squadrato”, maggiormente evidenziato dalla frequenza e dal periodo dell’onda. Questa caratteristica è dovuta agli angoli a 90°, che fanno emergere le continue differenze d’onda da positivo a negativo. Una nota importante è che spesso la square wave viene modulata tramite il parametro *pulse width* che permette di modificarne il suono.

*Onda a dente di sega (sawtooth wave)*: questo tipo di onda è particolarmente spigolosa, a tratti può ricordare un suono distorto, ma controllato. Molto celebre nell’uso dei bassi nella musica elettronica, un chiaro esempio potrebbe essere la celebre *Satisfaction* di Benny Benassi.

*Onda triangolare (triangle wave)*: è un’onda più regolare della sawtooth, meno spigolosa, un suono che ricorda quello di un’onda sinusoidale, anche se diverso.

## 1.9 Decibel e LUFS

Il concetto di decibel è alla base della teoria del suono, coincide con l’ampiezza sonora e le grandezze logaritmiche, spesso connesse ai concetti sul suono.

Una *scala logaritmica* diverge da una *scala lineare* poiché la proporzionalità tra le due grandezze non è costante, ma, appunto, logaritmica.

Scala lineare: 1-1, 2-2, 3-3 ...

Scala logaritmica: 10-1, 100-2, 1000-3 ...

In parole povere, la funzione logaritmica è inversa rispetto alla funzione esponenziale.

Il *Bel* è il logaritmo del rapporto tra una grandezza e il suo valore di riferimento (nel caso del suono serve ad indicare la grandezza e il valore dell’ampiezza di una forma d’onda). 1 Decibel equivale a 1/10 di *Bel*.

La grandezza che si incontra più comunemente è quella del dB<sub>spl</sub> (SPL: sound pressure level, livello di pressione sonora). Il valore di riferimento è la pressione atmosferica, misurata in Pascal, pari a 0.00002 Pa.

Di seguito un elenco alcuni suoni che possiamo ascoltare e la loro intensità in dB<sub>spl</sub> (Tratto da *Imparare la tecnica del suono* - Marco Sacco, 2014):

Soglia del dolore

<b>Suoni</b>	<b>dB<sub>spl</sub></b>
Soglia di udibilità/Camera anecoica	<b>0</b>
Stormire di foglie	<b>10</b>
Studio di registrazione estremamente silenzioso	<b>20</b>
Fischiettare leggero a 1,5 metri	<b>30</b>
Studio di registrazione vuoto/Auditorium silenzioso	<b>40</b>
Casa (Notte)	<b>50</b>
Conversazione (a un metro)	<b>60/70</b>
Cabina di un jet (Cruise)	<b>80</b>
Traffico stradale intenso (a 1,5 metri)	<b>90</b>
Fortissimo di musica classica	<b>100</b>
Musica da film (a 6 metri)	<b>110</b>
Musica rock (a 3 metri)	<b>120</b>
Sirena 50HP	<b>130</b>
Colpo di pistola Colt 45 (a 8 metri)	<b>140</b>

Non esistono solo i dB<sub>spl</sub>, ne esistono di diversi tipi, anche in funzione dell'elettronica legata al suono. Il valore **Root Means Square (RMS)**, già citato, permette di percepire qual è il valore medio di decibel a cui suona una traccia ed è stato utilizzato, per alcuni decenni, come parametro di riferimento in fase di mastering (finalizzazione delle tracce).

L'unione tra dB<sub>spl</sub> e inviluppo ADSR permette di percepire la gamma dinamica (Dynamic Range) di un suono, ovvero l'intervallo tra il valore massimo e il valore minimo di decibel che esso può assumere.

Un altro parametro utile per la misurazione del volume sono i LUFS (Loudness Units relative to Full Scale), ovvero una misura standardizzata del volume audio che considera sia la percezione umana che l'intensità del segnale elettrico (LANDR, LUFS). Questa misurazione è utile per definire gli obiettivi di normalizzazione sonora nei diversi media (es. cinema, TV, radio, streaming), è diversa dalle altre ed è ormai diventata uno standard a livello professionale. A differenza dei dB Full Scale (FS), che ci permettono di monitorare il livello più alto di picco (arrivando alla fastidiosa distorsione digitale), i LUFS ci permettono di conoscere il volume complessivo di una traccia, calcolando il volume integrato che prende in considerazione l'intera durata di una canzone. Gli LU (Loudness Unit) aiutano a valutare quanto "suona forte" un mix. Con diverse tipologie di strumenti è possibile andare ad analizzare i LUFS a breve termine (Short Term), ovvero il volume percepito negli ultimi tre secondi di ascolto,

oppure i LUFS momentanei (Momentary), molto simili ai dB Full Scale, ma misurati negli ultimi 400ms di ascolto. Tutto ciò verrà maggiormente approfondito nel terzo capitolo, all'interno del paragrafo dedicato al mastering.

## 1.10 Il suono digitale

L'introduzione della tecnologia digitale ha portato dei cambiamenti radicali nel mondo della musica e in tutto il settore audio. La praticità della manipolazione dei bit ha permesso trasformazioni difficili da pensare da un punto di vista analogico, ha introdotto nuovi generi e nuovi strumenti di lavoro.

La rivoluzione digitale sta continuando ancora oggi, da quando è nata non si è mai fermata. Dai primi effetti digitali come riverberi, echo e distorsori si è arrivati alla produzione musicale condensata in un computer. L'audio immersivo, nonché il fulcro di questa tesi, è la massima espressione odierna del progresso portato dalla digitalizzazione sonora.

Tutto ciò ha cambiato profondamente il mercato e i suoi standard, perché i costi di produzione si sono abbassati. Gli home studio dei giovani produttori musicali sono attuali e alcune delle canzoni suonate nelle radio di tutto il mondo sono state registrate in una camera da letto, non in un costoso studio (come è accaduto, ad esempio, per l'album di Billie Eilish "When We All Fall Asleep, Where Do We Go?", vincitore del premio Best Album ai Grammy Award del 2019) (Mamo, 2020).

Inoltre, tutto questo cambiamento si è fatto strada in parallelo alla crescita esponenziale dei tutorial e della didattica online, che ha velocizzato l'apprendimento di tutti gli strumenti del settore.

Ma come è possibile trasformare un segnale analogico in un suono digitale?

## 1.11 La frequenza di campionamento e la quantizzazione

Il suono, come abbiamo visto, è una variazione ciclica nella pressione atmosferica. Un microfono, ad esempio, capta questa variazione tramite le lamelle presenti nella capsula al suo interno, le quali traducono il proprio movimento in un segnale elettrico. La trasformazione di un segnale da analogico (continuo) a digitale (discontinuo) prende il nome di *campionamento*. Il termine deriva dal costante prelievo di piccoli campioni dal segnale analogico in entrata. Questa costanza viene identificata come una frequenza, detta frequenza di campionamento.

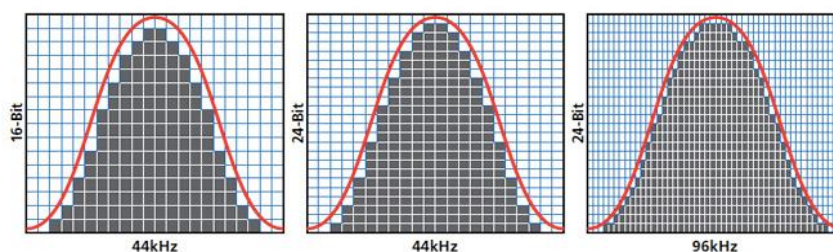


Figura 12 - Frequenza di campionamento (hifidagostini.it)

Quindi, preso un segnale analogico, da questo verranno estratti piccoli pezzetti, che porteranno ad una necessaria perdita di dati, dovuta allo spazio tra un campionamento e un altro. La domanda da porsi a questo punto è: ma i valori istantanei persi erano davvero necessari?

Un segnale audio può essere scomposto nella somma di singole sinusoidi, ognuna con la propria frequenza. Il teorema di Nyquist sancisce che, se il campionamento viene eseguito ad una frequenza pari almeno al doppio della banda del segnale che si sta campionando, il passaggio dall'analogico al digitale potrà avvenire senza perdita di informazione. Ciò vuol dire che c'è la possibilità di campionare un suono analogico e di renderlo a tutti gli effetti reale senza perdita di qualità. Ecco perché, considerando l'escursione della soglia di udibilità (20 Hz – 20 kHz), la frequenza di campionamento dei CD è pari a 44.1 kHz, più del doppio della banda della soglia.

E questo non è l'unico valore che entra in gioco nella trasformazione analogico/digitale. I piccoli valori istantanei che vengono campionati devono essere arrotondati al fronte di non considerare i valori istantanei mancanti dal campionamento. Questa operazione prende il nome di *quantizzazione*, ovvero quando i campioni vengono prelevati ad intervalli costanti e per ogni intervallo viene individuato un punto centrale che viene quantizzato ad un valore che ne approssima il valore reale. È qui che entra effettivamente in funzione il passaggio tra analogico e digitale. Un segnale composto da una sequenza finita di intervalli quantizzati verrà convertito in forma binaria. Maggiore è il numero di intervalli, minore è l'errore (Marco Sacco, 2014). La quantizzazione impiegata per la realizzazione dei CD è a 16 bit, che corrispondono a  $2^{16}$  (65.536) intervalli al secondo. Negli studi di registrazione la quantizzazione arriva a 24 bit (1.677.216 intervalli), o più alta, con frequenze di campionamento che possono arrivare fino a 192 kHz.

Riassumendo, campionare un segnale ad una certa frequenza (es. 44.1 kHz) significa estrarne altrettanti campioni al secondo. Ciascun campione verrà diviso in ulteriori intervalli quanti i bit dedicati al progetto (se fossero 16 bit sarebbero 65.536). Ogni secondo di un CD stereo equivale a:

$$2 \text{ (canali stereo)} \times 16 \text{ (bit della quantizzazione)} \times 44100 \text{ (campioni estratti)} = 1411200 \text{ bit al secondo}$$

Che corrispondono a:

$$1411200 / 8 \text{ (bit)}^2 = 176400 \text{ Byte/s}$$

Per trasformarli in KB è necessario eseguire la seguente proporzione:

---

<sup>2</sup> Secondo la regola per cui 1 Byte = 8 bit.

$$1 : 1024 = x : 176400$$

$$x = 172,26 \text{ KB/s}$$

Dunque, campionare un secondo in qualità CD di un segnale stereo occupa 172,26 KB/s. Ipotizzando la durata di 74 minuti di un normale CD, lo spazio necessario è di:

$$172,26 \text{ (KB/s)} \times 60 \text{ (sec)} \times 74 \text{ (minuti)} = 764 \text{ MB}$$

La dinamica (intesa come Dynamic Range) all'interno dei dispositivi digitali viene calcolata con una formula che parte dagli intervalli di quantizzazione:

$$\text{Dynamic Range} = 20 \log (\text{intervalli di quantizzazione}/1)$$

Che corrisponde ai seguenti risultati:

<b>Numero di bit necessari</b>	<b>Intervalli di quantizzazione</b>	<b>Dynamic Range</b>
8	256	48,2 dB
16	65536	96,33 dB
20	1048576	120 dB
24	16777216	144,5 dB

## 1.12 Strumenti di analisi delle frequenze: analizzatore di spettro

Per analizzare l'andamento delle frequenze all'interno di una traccia in esecuzione viene utilizzato l'*analizzatore di spettro* detto *spettrometro*. Attraverso la rappresentazione decibel/frequenza verranno visualizzate in orizzontale (ascisse) tutte le frequenze udibili all'orecchio umano, mentre in verticale (ordinate) il livello di decibel a cui stanno suonando. Uno spettrometro viene visualizzato anche negli equalizzatori grafici, che permettono di diminuire o aumentare una certa banda di frequenza.



Figura 13 - Analizzatore di spettro SPAN Plus di Voxengo

Lavorare con questo strumento ci permette di visualizzare ciò che stiamo ascoltando. Il concetto di visualizzazione è fondamentale in ambito sonoro, soprattutto perché, come verrà ampiamente descritto nel prossimo capitolo, il nostro udito ha delle caratteristiche estremamente personali, che possono estremizzare o nascondere alcuni dettagli. Poter visualizzare l'andamento delle frequenze grazie all'analizzatore di spettro è un grande aiuto, soprattutto quando si vanno a comparare due canzoni che vogliono suonare in maniera simile. Imparare a saper riconoscere pregi e difetti di una canzone tramite la visualizzazione è una skill degli odierni sound engineer.

Il mondo della musica è famoso per l'attenzione alla fedeltà d'ascolto. Gli strumenti che oggi ci permettono di lavorare su un file audio danno la possibilità di agire in modo estremamente preciso, quindi l'ascoltatore medio, che non è un audiofilo, richiede una qualità tecnicamente più alta anno dopo anno.

### 1.13 Processori di segnale: l'equalizzatore

Un *equalizzatore* è un circuito (o un algoritmo) in grado di amplificare o attenuare una determinata banda di frequenza, lasciando le altre inalterate. In forma di algoritmo, gli equalizzatori vengono integrati nei mixer di ultima generazione o nei programmi di produzione/editing audio.



Figura 14 - Fab Filter Pro Q 3

I controlli di cui sono dotati gli equalizzatori sono i seguenti:

*Guadagno*: regola l'aumento o la diminuzione della frequenza selezionata

*Q*: regola l'ampiezza angolare della curva che modifica le frequenze. Questo valore può essere relativo anche alla pendenza (slope rate), che stabilisce quanto rapidamente un'ampiezza decada. Essa si misura in dB per ottava.

*Intervalli*: possono essere fissati diversi intervalli per modellare come si desidera la forma dell'equalizzazione.

Equalizzazioni più comuni (alcune di queste chiamate filtri perché eliminano direttamente alcune frequenze) sono le seguenti:



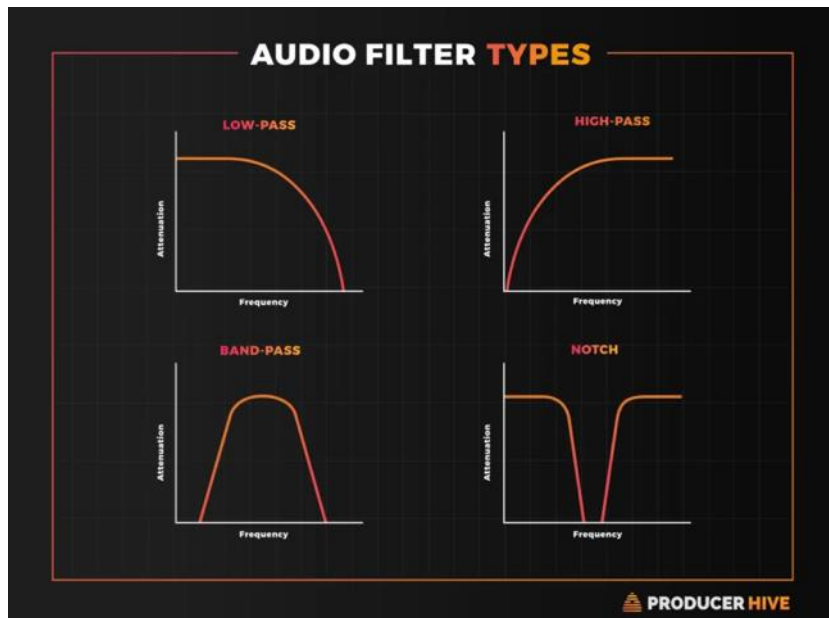


Figura 15 - Tipologie di equalizzazioni (ProducerHive.com)

*Passa basso (Low pass filter - LPF):* elimina la parte delle alte frequenze, andando a togliere brillantezza a suoni destinati ad essere più scuri.

*Passa alto (High pass filter - HPF):* **elimina** le basse frequenze. Molto utilizzato, ad esempio, sulla voce: andando a rimuovere tutte quelle informazioni sotto gli 80 Hz, si elimineranno tutte quelle frequenze fastidiose nel cantato e in collisione con altri strumenti.

*Passa banda (Band pass filter):* equalizzazione utile per confinare un suono all'interno di un perimetro di frequenze, facendo pulizia agli estremi di banda.

*A Reiezione di banda (Notch):* questa tipologia di equalizzazione è molto estrema ed utilizzata specificatamente per eliminare un certo numero di frequenze. Si utilizza con un Q molto stretto per evitare che vada ad impattare troppo evidentemente sulla resa sonora.

*A campana (Peak Bell):* una delle equalizzazioni più utilizzate. Enfattizza o diminuisce un determinato range di frequenze. Utilizzando un Q molto stretto si può operare con molta precisione per realizzare equalizzazioni estremamente dettagliate.

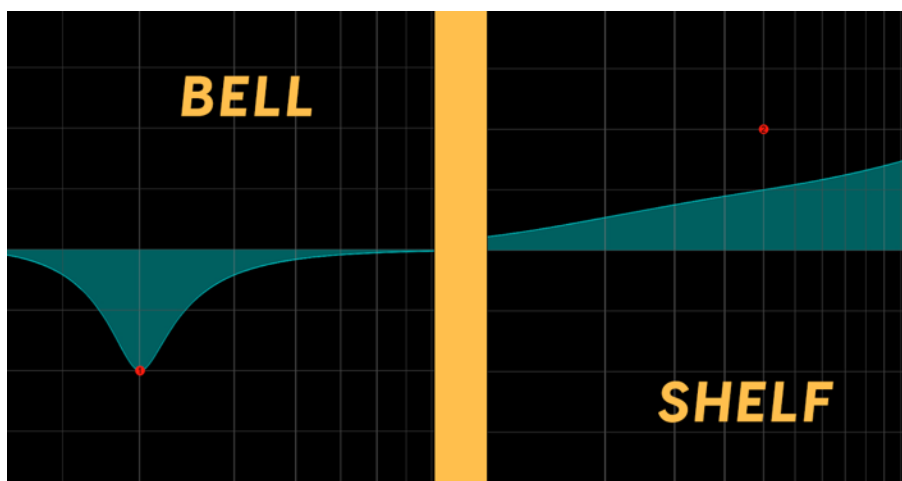


Figura 16 - Equalizzazioni Peak Bell e Shelf

A scaffale (*Shelving - Shelf*): equalizzazione molto morbida, tendenzialmente applicata agli estremi di banda nei casi in cui essi si vogliono attutire o evidenziare.

Negli ultimi anni è molto in voga l'utilizzo di equalizzatori che possiedano una componente dinamica (come il software mostrato in figura 14). Una volta evidenziate le frequenze da trattare, è possibile settare una soglia oltre la quale la frequenza in eccesso verrà dominata da una compressione.

### 1.14 Processori di segnale: il compressore

Un *compressore* è un processore di dinamica (hardware o software). Dopo l'ingresso di un segnale audio, questo verrà ridotto (o incrementato, nel caso degli *expander*<sup>3</sup>) in base ad un *rapporto di compressione*, per poi essere restituito in uscita. Questo può avvenire, ad esempio, durante la registrazione di segnali audio con volumi non costanti, come la voce o una chitarra, dove ci sia bisogno di applicare una riduzione di volume sui picchi per poter avere un flusso più costante. Il compressore serve principalmente a questo.

Immaginiamo un misuratore di decibel che va da meno infinito a 0: dopo aver settato una soglia (es. -10 dB), tutto ciò che sarà al di sotto della soglia verrà lasciato intatto, mentre tutto ciò che sarà sopra la soglia verrà compresso secondo un rapporto di compressione (2:1, 3:1, 4:1, etc.). I picchi di volume verranno attutiti e avranno maggior risalto i suoni rimasti sotto la soglia. Questo procedimento cambia completamente l'escursione di un suono, detta anche dinamica.

---

<sup>3</sup> Expander: processori di dinamica che hanno un funzionamento opposto ai compressori.

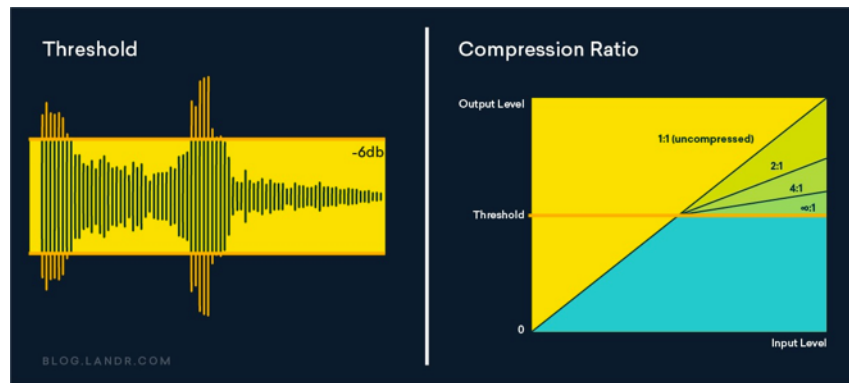


Figura 17 - La compressione (Landr.com)

Riassumendo, queste sono le caratteristiche principali di un compressore:

*Threshold (soglia)*: valore espresso in dB per definire il gap sonoro tra lo 0 e la soglia che subirà la compressione.

*Ratio (rapporto)*: definisce la quantità di compressione. Esistono diversi rapporti:

1:1 – Nessuna riduzione.

2:1 – Segnale dimezzato una volta superata la soglia.

3:1 fino a 8:1 – Alti valori di compressione, la dinamica risulta particolarmente schiacciata.

10:1 Il compressore inizia a comportarsi come un limiter, uno strumento analogo che non permette al segnale di superare una certa soglia applicando una fortissima compressione (molto usato nel mastering).

*Gain (guadagno)*: per ristabilire una coerenza tra il suono non processato e il suono compresso, è necessario riportare entrambi i suoni allo stesso livello di decibel, con la differenza che, applicando il guadagno di dB ad un suono compresso, emergeranno molto di più tutte le frequenze che sono rimaste sotto la soglia. Pareggiare i decibel sottratti dalla compressione si chiama anche *make-up gain*.

*Attack time (tempo di attacco)*: in base all'involuppo ADSR del suono in entrata, settare un tempo di attacco in funzione del compressore (più o meno veloce) risulterà fondamentale per la resa sonora in uscita. Un colpo di rullante avrà un tempo di attacco molto veloce perché raggiungerà l'apice della sua dinamica in un tempo brevissimo.

*Release time (tempo di rilascio)*: allo stesso modo, seguendo l'involuppo ADSR, settare un giusto tempo di rilascio in base al suono che si sta processando è altrettanto importante per la resa che si vuole ottenere.

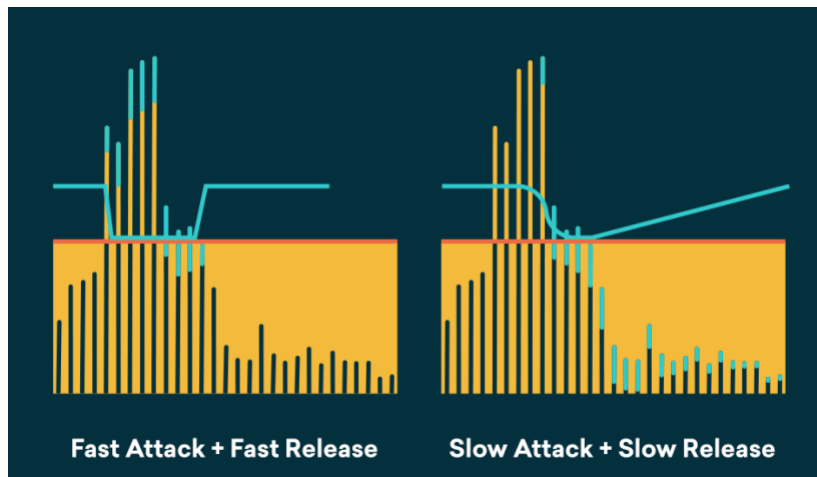


Figura 18 - Gestione Attack e Release nella compressione (Landr.com)

*Hold time (tempo di tenuta)*: permette di ritardare l’inizio del tempo di rilascio dopo che il segnale è sceso sotto la soglia, mantenendo il compressore attivo più a lungo.

I compressori analogici sono sempre stati un punto forte dell’*outboard* di ogni sound engineer. I loro circuiti elettrici permettono di dare un suono caldo e caratteristico. Da quasi vent’anni le software house del settore musicale (Acustica Audio, Universal Audio, Softube, Brainworx, Slate Digital, Waves, ecc.) stanno cercando di ricreare una sensazione di “analog-feeling” nelle emulazioni e nei plugin, raggiungendo nel presente ottimi risultati.

Una visione odierna è che la musica non sia solamente In-the-Box (processata tramite computer). Molti studi prediligono l’ibrido tra analogico e digitale, integrando con semplicità entrambi i flussi di segnale nel workflow.

Ecco che risulta importante conoscere le differenze sonore apportate dalle diverse tipologie di compressori (Kinarecords, 2020):

*Compressore Variable-Mu (VariMu)*: i primi compressori introdotti nel mercato. Come tutti i prodotti valvolari con un suono unico e ancora oggi molto apprezzato. Attacco e release abbastanza lenti e saturazione intensa. Una caratteristica interessante è l’assenza di Ratio. La compressione, morbida e musicale, viene regolata tramite il volume di ingresso (Input gain). Il compressore VariMu è ideale per enfatizzare l’attacco di qualsiasi suono (es. pianoforte o chitarre), ma anche fare “glue-compression” sui channel bus o in fase di mastering. Non è adatto a comprimere picchi di volume su transienti veloci (es. rullanti). Il compressore valvolare VariMu più celebre è il Fairchild 670, nato intorno agli anni ‘50. Esistono molti plugin che emulano questo hardware, tra cui l’*Ultramarine* di Acustica Audio, il *Fairchild Tube Limiter* di Universal Audio. Altri famosi modelli sono l’*RCA BA-6A* e il *Manley Variable Mu Stereo Compressor*.



Figura 19 - emulazione UAD del FairChild 670

**Compressori a Transistor (FET):** gli anni passano e le valvole vengono man mano sostituite dai transistor (l'acronimo FET indica proprio Field Effect Transistor). Attacco e release molto più veloce e saturazione aggressiva rispetto ai compressori valvolari. Una nota particolare è che i parametri e i knob di attack e release funzionano al contrario (in senso antiorario). Possibilità di controllare la Ratio. Sono molto adatti a comprimere batterie e suoni percussivi. Ritrovare il suono caratteristico della saturazione e della distorsione armonica di questi compressori è molto difficile nei plugin odierni. L'hardware più celebre è l'UREI 1176, riprodotto in diverse emulazioni software (es. 1176 di Universal Audio o CLA-76 di Waves), ma anche come revisioni di altri marchi di outboard (es. Wa76 di Warm Audio).



Figura 20 - Urei 1176 replica Universal Audio

**Compressori ottici (Opto):** i compressori ottici sono chiamati "Opto" per il funzionamento del loro fotoresistore, un componente elettronico sensibile alla luce. Una piccola lampada si illumina al passaggio di un segnale audio, l'intensità della luce determina quanto il segnale verrà compresso. Attacco e Release molto lenti, di conseguenza hanno una compressione molto dolce. Ottimi su voci, bassi e suoni con attacchi lenti, da evitare se si cerca una compressione precisa, veloce e lineare. Anzi, è proprio la loro "imprecisione" a dare ai compressori ottici il loro suono caratteristico. In questa categoria è assolutamente da citare il Teletronix LA-2A, considerato da molti professionisti uno dei migliori hardware mai prodotti. Introdotto negli anni '60 è stato utilizzato in moltissimi dischi, in particolar modo

sulle voci. Un altro compressore ottico da menzionare è il Tube-Tech CL 1B, nato negli anni '80. Software degni di nota sono London Acoustics di Acustica Audio, Teletronix LA-2A di Universal Audio e CLA-2A di Waves.



Figura 21 - Teletronix La-2a replica Universal Audio

*Compressori a Stato Solido (VCA):* i compressori a stato solido, detti Voltage Controlled Amplifier, sono i più versatili e precisi nel mondo analogico. Attack e Release velocissimi, completo controllo dei parametri, alti livelli di compressione e risposta in frequenza lineare. Perfetti per comprimere transienti pronunciati e picchi di volume irregolari (es. batterie e percussioni), utili anche per un utilizzo creativo della compressione. Essi sono frutto del progresso, aumentando controllo e possibilità hanno perso carattere e sonorità, a favore di una maggiore costanza e linearità. Il G Compressor di Solid State Logic è probabilmente il più famoso tra i compressori VCA sul mercato, insieme all'API 2500, il DBX 160A e l'Empirical Labs Distressor.

Di tutti i modelli appena citati ci sono le conseguenti emulazioni di Universal Audio, Native Instruments, Waves, Plugin Alliance e molte altre software house.



Figura 22 - SSL G Compressor serie 500



Figura 23 - Empirical Labs Distressor

*Compressori digitali:* all'interno di tutti i software di produzione musicale (DAW - Digital Audio Workstation) si possono trovare diversi compressori digitali nativi oppure si possono acquistare dai migliori marchi di plugin. Utilizzare un compressore digitale, che non sia un'emulazione, permette di avere attacco e rilascio velocissimi, completo controllo dei parametri, massima linearità e alti livelli di compressione. Ottima scelta per avere uno strumento che agisca minuziosamente senza dare ulteriore carattere al suono. Tra i vari software presenti sul mercato è da citare Fab Filter Pro C2, un compressore preciso e affidabile, con l'opzione dry/wet, utile per la tecnica della compressione parallela.



Figura 24 - Fab Filter Pro C2

Ora che sono stati esposti tutti i concetti fondamentali sulla fisica del suono e sono stati introdotti i principali strumenti per processare i segnali, la tesi procederà con un capitolo

sulla psicoacustica, trattando elementi legati alla percezione sonora e alla spazializzazione.



## CAPITOLO 2 - L'ascolto binaurale

Il modo in cui l'essere umano ascolta un suono è strettamente connesso alla fisica dello stesso. La percezione dello spazio, e l'annessa localizzazione dei suoni, è determinata dall'apparato uditivo, quindi dall'ascolto. Non solo, tutte le tecnologie legate al mondo della musica sono state create analizzando il funzionamento tecnico dell'apparato uditivo. Un microfono, ad esempio, contiene alcune lamelle che vibrano e traducono la loro energia in un segnale elettrico, esattamente come accade nei nostri timpani. Ecco, quindi, che la sensazione di immersione, ricercata in questa tesi e in molti altri ambiti scientifici, dipende direttamente dal modo in cui il nostro cervello viene stimolato, eluso, incuriosito, a scoprire cosa c'è oltre a un'abitudine (ad esempio la stereofonia, seguendo il fil rouge di questo saggio). In questo contesto entra a far parte anche la forma ludica dell'arte sonora e quali siano i suoi ruoli percettivi, definiti chiaramente dalla psicoacustica, la disciplina che studia i suoni in relazione alle sensazioni uditive da un punto di vista psicologico e fisiologico (Dizionario Treccani, Psicoacustica).

Queste saranno le tematiche principali del secondo capitolo, che andrà a fondo sulla definizione di ascolto binaurale, sull'anatomia del nostro apparato uditivo, sulle dinamiche della percezione spaziale e sulle tecnologie di riproduzione sonora.

### 2.1 L'orecchio umano

L'anatomia umana ha ispirato tutti gli strumenti utilizzati per la registrazione di un suono, che sono, in sostanza, una replica dell'orecchio. Dopo aver captato lo spostamento delle particelle dell'aria (energia acustica), questo viene trasformato prima in energia meccanica, successivamente tradotta in energia elettrica. Gli impulsi elettrici vengono inviati al cervello grazie alle terminazioni nervose e vengono elaborati permettendo la percezione e l'ascolto. Il nostro apparato uditivo è diviso in tre sezioni:

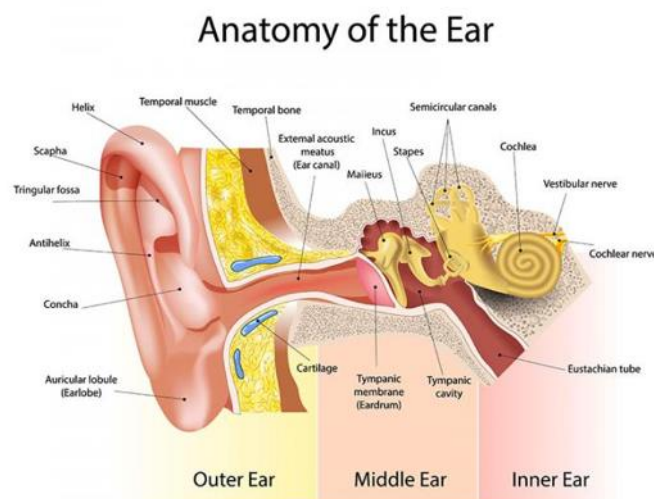


Figura 25 - Anatomia dell'orecchio umano ([beltonetrystate.com](http://beltonetrystate.com))

*Orecchio esterno (Outer Ear)*: appena arrivato in prossimità delle orecchie, il suono incontra il *padiglione auricolare*, che grazie alla sua vasta superficie riflette il suono verso il *condotto uditivo* (lungo mediamente 3 cm). Come descritto in Roginska, A., & Geluso, P, 2017: “Il canale uditivo umano permette di aumentare selettivamente la pressione sonora da 30 a 100 volte per frequenze intorno a 3 kHz tramite un effetto di risonanza passiva dovuto alla lunghezza del condotto uditivo. Questa amplificazione rende gli esseri umani particolarmente sensibili alle frequenze nell'intervallo 2-5 kHz, che sembra essere direttamente correlato alla percezione del parlato. Le convoluzioni verticalmente asimmetriche del padiglione auricolare sono sagomate in modo tale che l'orecchio esterno possa trasmettere più componenti ad alta frequenza, da una sorgente elevata, che dalla stessa sorgente a livello dell'orecchio. Allo stesso modo, le alte frequenze tendono ad essere più attenuate per le sorgenti posteriori rispetto a quelle anteriori, come conseguenza dell'orientamento e della struttura del padiglione (Blauert, 1997)”.

*Orecchio medio (Middle Ear)*: il condotto uditivo finisce con una membrana, il *timpano*, che vibra in base alla frequenza che ha raggiunto l'orecchio. Opposti al timpano, sono connessi tre ossicini: *martello*, *incudine* e *staffa*. Questi amplificano la vibrazione del timpano e la trasmettono alla *coclea* (orecchio interno), un ulteriore osso riempito di un fluido denso, che necessita dell'amplificazione del suono per vibrare. I tre ossicini citati, inoltre, impediscono all'orecchio di rimanere danneggiato a causa di una pressione sonora troppo elevata, grazie ad una serie di piccoli legamenti. All'interno dell'orecchio medio risiede un'apertura, la *tuba di Eustachio*, un canale che conduce alla cavità orale per dare uno sfogo verso l'esterno in modo da equilibrare la pressione atmosferica ai lati del timpano. La sua funzione principale è quella di far corrispondere l'impedenza relativamente bassa dei suoni nell'aria rispetto all'alta impedenza del fluido nell'orecchio interno (l'impedenza in questo contesto si riferisce alla resistenza di un mezzo al movimento). Senza questa azione, ci sarebbe una perdita di trasmissione di 1000:1 corrispondente a una perdita di sensibilità di 30 dB (Roginska & Geluso, 2017).

*Orecchio interno (Inner Ear)*: questa è la parte che permette la trasmissione al cervello del suono grazie alla conversione dell'energia meccanica in impulsi elettrici. La staffa è in contatto con la coclea attraverso una membrana (*finestra ovale*). La coclea, come già accennato, è un osso a forma di chiocciola che contiene del fluido, ha tre piccoli canali orientati secondo le tre dimensioni spaziali, servono al cervello per la percezione dell'equilibrio. Il fluido riceve la vibrazione amplificata dalla staffa attraverso la finestra ovale e permette di trasferirla all'organo incaricato alla conversione dell'energia meccanica in elettrica: l'*organo del Corti*. Al suo interno vi è la *membrana basilare* che possiede una miriade di *cellule ciliate* (circa 16000 tra interne ed esterne), che si mettono a vibrare in base alla vibrazione del fluido. Ogni gruppo di cellule è direttamente connesso ad una terminazione nervosa, che permette la conversione della loro vibrazione in impulsi elettrici, i quali arriveranno al cervello dove verranno elaborati e percepiti come

suoni. Si parla di gruppi di ciglia perché una singola frequenza non ecciterà soltanto una ciglia. Il fenomeno descritto viene chiamato *banda critica* ed è alla base dei diversi fenomeni di psicoacustica.

## 2.2 Le curve isofoniche

Le *curve isofoniche* rappresentano graficamente la reazione dell'orecchio umano alle diverse frequenze ascoltate. Il rapporto sui grafici è lo stesso dello spettro frequenza (frequenza per dB). Esse sono state ricavate in una *camera anecoica*, progettata per eliminare (o ridurre al minimo) le riflessioni sulle pareti, in modo da far pervenire all'ascoltatore solo un segnale diretto.

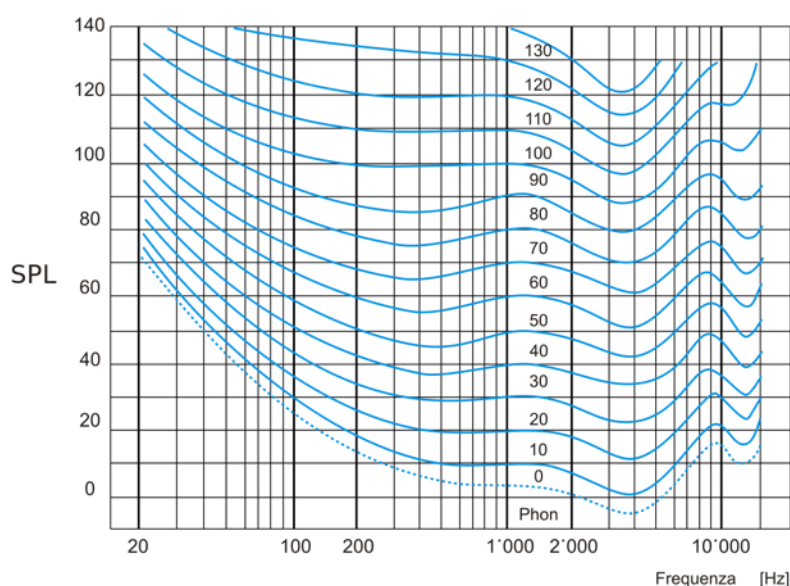


Figura 26 - Curve isofoniche (Wikipedia.com)

Leggendo il grafico è possibile capire quanto le basse frequenze debbano essere maggiormente amplificate per essere ascoltate. Prima di fare un esempio, seguendo le linee tracciate tra ascisse o ordinate, è necessario specificare che la frequenza di riferimento per ogni curva è 1 kHz; il valore che identifica una curva si chiama *phon* e corrisponde al valore dei dB alla frequenza di riferimento. La curva a 90 phon necessita di circa 110 dB per poter ascoltare una frequenza a 30 Hz alla stessa intensità di una frequenza a 1 kHz. In tutte le curve c'è un abbassamento di intensità tra i 2 kHz e i 6 kHz, questo perché, come precedentemente citato, in questo range ci sono le frequenze del parlato e il nostro orecchio è stato anatomicamente creato per percepirle. Un esempio dell'applicazione delle curve isofoniche nella tecnologia di tutti i giorni è l'impostazione "Loudness" presente negli amplificatori e negli stereo, che tende a rendere percepibili le basse frequenze allo stesso livello delle altre frequenze, cambiando l'equalizzazione e la percezione di una traccia.

Il concetto illustrato in questo paragrafo verrà ripreso nel confronto tra l'ascolto in cuffia e l'ascolto con le casse, una differenza in cui le curve isofoniche hanno particolare importanza in termini di percezione.

## 2.2 Head Related Transfer Function

Quotidianamente siamo abituati a percepire segnali differenti che arrivano alle nostre orecchie. La nostra percezione è quella di ascoltare uno o più segnali provenienti dalle loro rispettive direzioni, ma in realtà ciò che ascolta l'orecchio destro è un segnale diverso rispetto da ciò che ascolta il nostro orecchio sinistro. La *fusione binaurale* è la facoltà del cervello che ci permette di unire questi due continui stimoli uditivi in uno solo e che ci fornisce tutte le informazioni utili alla localizzazione di un suono. Per definire la localizzazione di un suono si utilizza molto spesso un sistema di coordinate che ragiona in termini di *azimuth*, *distance* ed *elevation*. Adattando questi termini a un sistema di riferimento cartesiano ortogonale, dove per l'origine passano tre assi orientati (X, Y e Z), azimuth rappresenterebbe l'asse X, distance l'asse Y ed elevation l'asse Z.

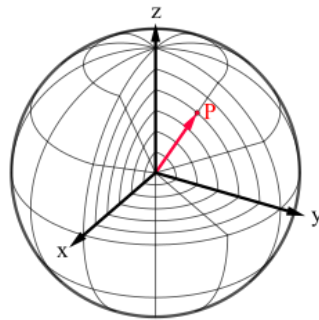


Figura 26 - Coordinate polari sferiche X, Y, Z (Wikipedia)

La localizzazione di un suono posizionata di fronte a un ascoltatore si troverà a  $0^\circ$  azimuth e  $0^\circ$  elevation. Mentre la quota della distance rappresenterà i metri che separano l'ascoltatore della sorgente sonora.

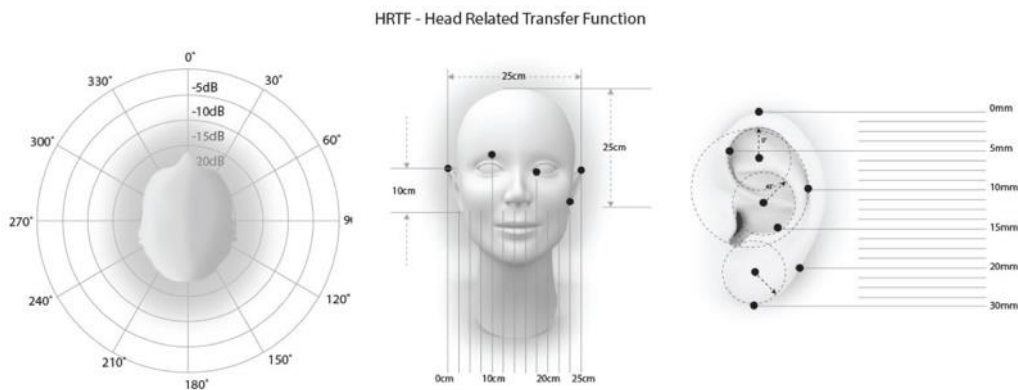


Figura 27 - Head Related Transfer Function

Nell'azimuth i valori positivi espressi in gradi indicheranno un suono che si trova a destra (es. 45°) sul piano orizzontale. Mentre i valori negativi indicheranno un suono che si trova a sinistra (es. -45°). A 180° un suono si troverà alle spalle dell'ascoltatore. Allo stesso modo lavorano i valori espressi in gradi per indicare l'elevation (asse Z): i valori positivi indicheranno l'altezza verso l'alto sul piano verticale, i valori negativi l'altezza verso il basso.

Il parametro distance, invece, possiede solo valori positivi che riferiscono al sistema internazionale di unità di misura della lunghezza (Roginska & Geluso, 2017).

Ogni persona percepisce lo spazio circostante attraverso l'analisi cerebrale delle differenze di ascolto che intercorrono tra le proprie orecchie. Siccome non abbiamo occhi dietro alle nostre spalle e sopra la nostra testa, tendiamo ad utilizzare la vista come strumento per la percezione di ciò che abbiamo di fronte a noi e lasciamo alle nostre orecchie il compito di controllare ciò che accade sopra e dietro di noi.

Ipotizziamo di essere all'aperto e che un suono deciso arrivi da destra rispetto alla posizione dell'ascoltatore (30° azimuth, 0° elevation e 1 mt distance):

*ITD (Interaural Time Delay)*: ci sarà un ritardo di percezione tra l'orecchio destro e l'orecchio sinistro (a favore dell'orecchio destro) che darà un primo indizio del suono proveniente alla destra dell'ascoltatore.

*ILD (Interaural Level Difference)*: a questo ritardo bisognerà anche aggiungere anche una differenza di volume (o di ampiezza), perché il suono proveniente da destra verrà percepito più forte dall'orecchio destro rispetto all'orecchio sinistro.

*Pinna Filter*: mentre l'orecchio destro avrà una posizione favorevole ad accogliere il segnale emesso, all'orecchio sinistro arriverà un segnale filtrato sia a causa della forma del padiglione auricolare dell'ascoltatore, sia della testa, che sarà di ostacolo alle onde prodotte dalla sorgente sonora.

Questa descrizione è possibile dal momento in cui l'ascoltatore risulta normoudente. In casi di ipoacusia monolaterale (percezione di segnali soltanto da un orecchio), le difficoltà di localizzazione sonora saranno maggiori, proprio per l'impossibilità di avere un corretto funzionamento dell'Interaural Time Delay e dell'Interaural Level Difference.

Come spiegato in Roginska & Geluso (2017): "Gran parte della ricerca sulla localizzazione del suono umano in azimuth è derivata dalla "duplex theory" di Lord Rayleigh (1907) che enfatizza il ruolo di due segnali primari: differenze di tempo interaurale (ITD) e differenze di intensità interaurale (IID), detti anche Interaural Level Differences, (ILD), in particolare se specificati in dB<sub>SPL</sub>. Originariamente vennero condotti esperimenti su un suono con una singola frequenza (sinusoidale - sine wave) volti a definire che gli ILD fossero dovuti all'head-shadowing (il filtro di percezione sonora dovuto alla forma della testa) che determina la localizzazione delle alte frequenze (circa 2 kHz per persone con una testa grossa), mentre le ITD erano

importanti solo sulle basse frequenze. Le ricerche di Bernestein (2001) e Brughera, Dunai e Hartmann (2013) hanno invece dimostrato che le ITD contribuiscono alla localizzazione delle alte frequenze (sopra i 4 kHz)”. Sempre in Immersive Sound di Roginska & Geluso (2017) si evidenzia che: “La ricerca binaurale degli ultimi decenni, tuttavia, indica seri limiti della duplex theory. Ad esempio, per le sorgenti vicine, l'IID è disponibile anche alle basse frequenze (Shinn-Cunningham, 2000). Allo stesso modo, è noto che i segnali ITD basati sulla tempistica relativa degli involucri di ampiezza ("envelope ITD") dei suoni ad alta frequenza possono essere utilizzati da un meccanismo come il rilevamento delle coincidenze interaurali (Henning, 1974, 1980; van de Par & Kohlrausch, 1997; Bernstein e Trahiotis, 2010)”.

In situazioni in cui alcuni, ma non tutti, i segnali spaziali risultano alterati all'ascolto, una forma alternativa di ri-mappatura dei segnali consiste nel ridurre le informazioni spaziali fornite dai segnali alterati e, invece, fare più affidamento sui segnali che rimangono intatti (Roginska & Geluso, 2017). Nel caso specifico dell'ipoacusia monofonica, numerosi studi hanno dimostrato che il comportamento di localizzazione del suono nei mammiferi si adatta sia durante lo sviluppo che nell'età adulta attribuendo maggior peso ai segnali spaziali monofonici invariati forniti dall'orecchio con udito normale (Agterberg et al., 2014; Kumpik, Kacelnik & King, 2010; Keating et al., 2013).

Una sorgente sonora situata fuori dall'asse 0° (centro anteriore) darà origine ad una differenza di tempo tra i segnali in arrivo alle orecchie dell'ascoltatore che è correlata al suo angolo di incidenza. Il ritardo di tempo massimo tra le orecchie è dell'ordine di 650 µs o 0,65 ms ed è chiamato *ritardo binaurale*.

L'orecchio è sensibile alle differenze di fase interaurali solo alle basse frequenze e la sensibilità alla fase inizia a deteriorarsi al di sopra di circa 1 kHz (non a caso l'esatta frequenza analizzata nelle curve isofoniche). Alla presenza di basse frequenze, le cellule ciliate dell'orecchio interno si attivano regolarmente in punti specifici della fase del ciclo sonoro, ma alle alte frequenze questo schema diventa più casuale.

Vi sono informazioni ambigue al di sopra di circa 700 Hz, dove la distanza tra le orecchie è pari a mezza lunghezza d'onda del suono, perché è impossibile dire quale orecchio è in ritardo e quale è in anticipo. Inoltre, sorgono frequenze in cui la differenza di fase è zero. Le differenze di fase possono anche creare confusione in ambienti riflettenti in cui gli effetti dei riflessi possono modificare i segnali di fase presenti nelle orecchie.

Blauert ha trovato prove delle cosiddette "bande direzionali" che sono regioni dello spettro di frequenza che appaiono potenziate o attenuate per particolari posizioni della sorgente sul piano mediano. Una regione intorno a 8 kHz sembra corrispondere abbastanza strettamente alla percezione dall'alto, mentre le regioni da 300–600 e 3000–6000 Hz sembrano essere correlate abbastanza strettamente alla percezione frontale. Le regioni centrate su circa 1200 Hz e 12000 Hz sembrano essere strettamente correlate alla percezione posteriore.

Infine, nel contesto della valutazione spaziale di un suono è molto importante considerare anche il parametro *distance*. Quando avviene un'emissione molto vicino a un ascoltatore ci sono anche dei notevoli cambiamenti negli spettri HRTF, come recensito da Huopaniemi (1999). Sembra esserci un grado di aumento delle basse e delle alte frequenze nella differenza di livello interaurale tra le orecchie quando esse si devono rapportare con sorgenti poste a distanze molto ravvicinate.

## **2.3 Effetti e fenomeni psicoacustici**

### **2.3.1 Il fenomeno dei battimenti**

Quando due frequenze sono leggermente differenti (es. 300Hz e 305Hz), si rischia di non essere in grado di percepire distintamente due suoni, ma uno unico simile ad un battito, il cui ritmo è dato dalla differenza di fase tra le due frequenze. Le bande critiche dell'orecchio interno vengono eccitate in maniera particolarmente ravvicinata, talvolta sovrapposta, dando vita al *fenomeno dei battimenti*. Nella modulazione sonora dei sintetizzatori, un effetto simile è dato dagli oscillatori LFO (Low Frequency Oscillator) che possono creare un tremolo analogo a quello del fenomeno dei battimenti.

### **2.3.2 Effetto Haas - Legge del primo fronte d'onda**

Quando siamo in una stanza, il suono che percepiamo arriva alle nostre orecchie da diverse angolazioni. In primis dalla sorgente, successivamente, con un leggero ritardo, dalle riflessioni del segnale sulle pareti della stanza. Questo ritardo è dovuto al percorso più lungo che compie il suono per raggiungere l'orecchio. L'onda che verrà percepita per prima dal nostro cervello sarà quella che identificherà la direzione di provenienza del suono, questo è l'*effetto Haas* detto anche *effetto di precedenza* o *legge del primo fronte d'onda*. La direzione viene identificata anche nel caso di differenti intensità tra le onde percepite. L'effetto si verifica quando il ritardo è minore di 30-35 ms, intervallo definito come *zona di Haas*, che varia in base all'involuppo ADSR del suono. Nel caso in cui il ritardo uscisse dalla zona, i due suoni verrebbero percepiti come distinti, quindi relativi all'effetto eco.

Per il parlato, l'effetto di precedenza scompare per ritardi superiori a 50 ms, ma per la musica l'effetto può apparire anche per ritardi di circa 100 ms.

### **2.3.3 Mascheramento**

Una frequenza che possiede un'ampiezza più alta maschera le frequenze vicine con ampiezze più basse, perché già decodificate dalle ciglia dell'orecchio interno appartenenti alla stessa banda critica. Questa teoria è alla base degli algoritmi di compressione del formato .mp3.

### 2.3.4 Ricostruzione della fondamentale

La *fondamentale* è la nota principale che dà il nome a un accordo. Il nostro cervello è in grado di ricostruire la fondamentale dalle armoniche presenti in un suono, nonostante gli venga presentato un range ridotto di frequenze. Per questo motivo quando ascoltiamo la musica dalle cuffie di bassa qualità ci sembra lo stesso di sentire “bene” tutte le frequenze, perché percepiamo le fondamentali e la percezione ci sembra gradevole. Ciò che in realtà non sentiamo sono la totalità delle frequenze multiple della fondamentale che permettono di cogliere di maggiori sfumature nella musica, donando profondità e chiarezza nell’ascolto.

### 2.3.5 Effetto “cocktail party”

Il nostro cervello è capace di captare un suono all’interno di diversi suoni sovrapposti, grazie anche all’aiuto del principio dell’attenzione. Ad esempio, riusciamo ad ascoltare un discorso di un nostro amico durante una festa, nonostante il brusio e la musica di sottofondo.

### 2.3.6 Illusioni acustiche

Proprio come l’apparato visivo, anche quello uditivo è soggetto a illusioni. Questo a causa dei limiti fisici dell’apparato, prima tra questi la soglia di udibilità. Una famosa illusione acustica è la *Scala di Shepard*, una scala musicale ascendente o discendente che non ha mai fine. Poste tre scale che eseguono le 7 note nello stesso momento, al salire di tono, il volume delle note più alte scende, mentre quello delle note più basse sale. Il cervello non riesce ad accorgersi che la frequenza più alta è susseguita da un’altra ottava rispetto a quella di partenza, quindi quando il giro dovrebbe finire (dopo 7 note), l’ottava nota sembra far continuare la scala in progressione o in discesa all’infinito. Questa illusione acustica è stata usata recentemente da Hans Zimmer per la colonna sonora del film *Dunkirk* (2017), diretto da Christopher Nolan. Su un articolo<sup>4</sup> de “Il Post” viene riportata la seguente dichiarazione del regista:

*“C’è un’illusione acustica, se vogliamo chiamarla così, che si chiama “scala Shepard” e che avevo già usato in The Prestige con il compositore David Julyan. È un’illusione che fa credere che ci sia sempre un tono ascendente. È un effetto cavatappi. [...] Ho scritto la sceneggiatura secondo questo principio, con tre linee temporali che danno una costante idea di intensità crescente. Volevo costruire una musica con dei simili principi matematici.”*

---

<sup>4</sup> I rumori di Dunkirk – Il Post: <http://www.ilpost.it/2017/09/02/dunkirk-suoni-colonna-sonora/>



## 2.4 La psicologia della musica<sup>5</sup>

Inizia tutto da una nota, alla quale se ne aggiungono altre suonate da diversi strumenti, fino alla creazione di una melodia. Questa è la musica, un gruppo armonico di sonorità che arriva al nostro cervello passando dalle nostre orecchie. Dalla nascita dei primi suoni sono passati “soltanto” 2,5 milioni di anni, quando nel Paleolitico alcuni ominidi iniziarono a creare la prima tecnologia. Si può quindi affermare che per l’epoca la musica fu un nuovo media, un nuovo modo di comunicare, di aggregare i popoli, fattore che non è cambiato negli anni. Infatti, a pensarci bene, la musica è sempre stata un nuovo media, perché è sempre stata in evoluzione, come tutta la tecnologia d’altronde, e dal Paleolitico è arrivata ad essere uno dei mercati più influenti al mondo. Lo spirito di aggregazione è rimasto tutt’oggi, in particolar modo, nell’evento del concerto, un ritrovo sacro tra chi emette il sonoro e chi ne è fruitore, quindi ascoltatore. Ma perché, da una singola nota, tante persone si sono raggruppate per 2,5 milioni di anni? Una risposta scientifica può essere data dal cognitivismo musicale. La psicologia della musica studia come il cervello percepisca ed elabori l’informazione musicale, come la pratica musicale influenzi il funzionamento cerebrale e che tipo di applicazioni siano possibili in campo terapeutico. Grazie ad essa sappiamo che la percezione musicale inizia dall’età prenatale, poiché il feto comincia a rispondere a suoni e rumori a partire dal terzo trimestre della gravidanza. La prima melodia ascoltata è la vocalità materna, insieme a tutte le onde sonore che riescono ad attraversare il grembo materno e arrivare al feto. Una volta neonati, i piccoli preferiscono la ninnananna che ascoltavano in fase prenatale rispetto ad una nuova. Allo stesso modo, seguendo il meccanismo della riproduzione, che ripropone al cervello lo stesso tipo di informazione generando sicurezza e conoscenza, lo psicologo Peter Hepper, nel 1991, ha condotto un esperimento su alcune madri che guardavano la stessa serie televisiva durante la gravidanza. I neonati, tra i due e quattro giorni dalla nascita, i quali avevano ascoltato decine di volte la sigla del telefilm in grembo, mostrarono una diminuzione del battito cardiaco durante l’ascolto di quella canzone. Quindi, la musica è strettamente connessa alle nostre emozioni. Nel caso descritto la sensazione di qualcosa di conosciuto ha prodotto calma e tranquillità.

Le emozioni vengono considerate come sistemi complessi che svolgono una parte importante nel rapporto tra individuo e ambiente, si distinguono in due grandi categorie: le emozioni primarie e le emozioni complesse. Tutte le emozioni che riguardano il processo evolutivistico e la sopravvivenza sono considerate *primarie* (es. la paura), mentre tutte quelle che sono il risultato di un’esperienza sono definite come *complesse* (es. la malinconia). Il mercato musicale si fonda proprio sull’esperienza che l’ascoltatore ha con il sonoro: ascoltare un disco genererà diversi tipi di emozioni in base agli stati d’animo durante la fruizione, mentre andare ad un concerto sarà

---

<sup>5</sup> Molte delle informazioni di questo capitolo sono state tratte da Schön, D., Akiva-Kabiri, L., & Vecchi, T. (2007). *Psicologia della musica*. Carocci.

un'esperienza irripetibile e istantaneamente emozionante. Sono molti gli aspetti che ci portano a sentire un'emozione nella musica. In primis *la struttura di un brano*, con diversi andamenti, toni e dinamiche. Tra i fattori di carattere strutturale a risaltare è sicuramente il *tempo*. Un ritmo veloce ha una connotazione eccitante, mentre uno lento dovrebbe calmare e rilassare. Quando si va in discoteca, la musica commerciale e l'EDM si avvicinano ai 128 battiti per minuto, sincronizzando il battito cardiaco con la musica, cercando di generare eccitazione e fermento, quindi divertimento. Nei reggae-party, invece, capita spesso di vedere movimenti lenti, ritmi rilassanti, non a caso i battiti per minuto sono tra i 60 e i 90.

Un altro fattore importante è il *modo*, che definisce il carattere di un brano e la conseguente emozione. Già nell'antica Grecia venivano utilizzati diversi modi che prendevano il nome dai popoli (lidio, dorico, frigio), mentre nel sistema tonale occidentale i modi *maggiore e minore* sono i più conosciuti, ai quali i bambini dai sette anni sono sensibili e ne riconoscono la connotazione positiva (per il maggiore) e negativa (per il minore). Un esempio molto semplice è provare a suonare l'accordo La in modo maggiore e in modo minore: il primo scaturirà un sorriso, l'altro malinconia. Altri fattori sono *la timbrica* e il *tempo d'attacco* degli strumenti, ma soprattutto la *complessità armonica e ritmica* di un brano. Musiche troppo dissonanti e con ritmi non regolari hanno una comprensione difficile. Il jazz non viene compreso da tutti a causa della ritmica complessa, che richiede attenzione, mentre i generi pop e rock, con un tempo spesso ricorrente, sono alla base del mercato musicale.

Al fronte di tutte queste caratteristiche, anche se può sembrare scontato, il lato soggettivo ed esperienziale di ogni ascoltatore è fondamentale per la percezione delle emozioni nella musica. Ci sono sicuramente dei fattori comuni che mettono d'accordo diverse personalità e permettono ad alcune tracce di diventare dei successi, ma è difficile riuscire a capire come arrivare a suscitare determinate sensazioni con sicurezza.

## **2.5 Storia delle principali tecniche di riproduzione sonora**

Una linea temporale più lunga di un secolo, collega i punti cardini della storia della riproduzione sonora, dalle prime registrazioni fino all'audio spaziale. In questo paragrafo, oltre ad approfondire i dettagli storici, verranno descritte tutte le principali tecnologie. Si scoprirà l'importanza centrale dei mezzi di comunicazione (in particolar modo del cinema) nell'evoluzione tecnologica in ambito sonoro.

### **2.5.1 Gli albori dell'industria discografica e la monofonia**

**1877** Il primo passo significativo che aprì le porte all'industria musicale avvenne grazie a Thomas Edison, il pioniere per eccezione della comunicazione sonora, il quale inventò il *fonografo*, uno strumento che permetteva di incidere su una stagnola gli impulsi percepiti da una membrana. In sostanza, il fonografo registrava e riproduceva. La

scoperta non era assolutamente rivolta alla musica. Inizialmente, Edison pensò al fonografo come sostegno burocratico della firma personale sui documenti, registrando la voce del firmatario (migliorando il fonautografo inventato da Édouard-Léon Scott de Martinville nel 1857). Sebbene questa idea, intelligente e innovativa, non venne mai sfruttata, la sua invenzione ha permesso di poter incidere un suono e di riprodurlo, regalando al mondo lo strumento principale per la nascita della discografia, del sonoro nel cinema e di tutte le relative conseguenze.

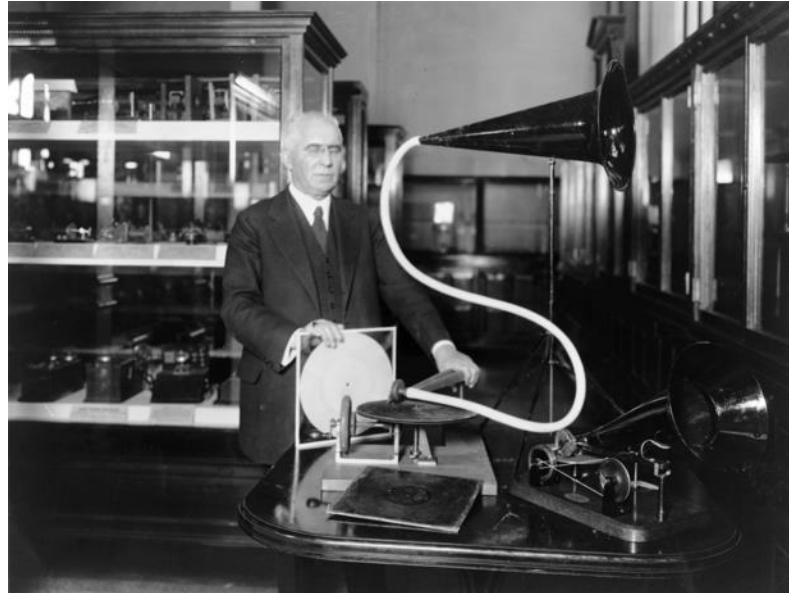


*Figura 28 - Il fonografo di Edison (Wikipedia.com)*

Alla mostra di Parigi del 1881, l'inventore Clement Ader fu un precursore assoluto dei tempi e fece un esperimento spesso ricordato come primo esempio conosciuto di trasmissione stereofonica della musica (Hertz, 1981). Egli posizionò dei pickup telefonici nelle luci della ribalta dell'Opera di Parigi (distanziati rispetto la lunghezza del palco) e trasmise i segnali a coppie di auricolari per ricevitori telefonici durante la mostra. I visitatori felici ascoltarono l'opera dal vivo e con un certo realismo spaziale. Purtroppo, era tutto estremamente in anticipo con i tempi e perciò l'ottimo risultato di questo esperimento non divenne una scoperta.

1887

Dieci anni dopo il fonografo di Edison, Emile Berliner sfruttò le competenze tratte dall'invenzione per brevettare il *grammofono*, che poteva registrare e incidere segnali, per poi sfruttare le incisioni presenti su un disco (detto disco fonografico) e riprodurle da una grande tromba in ottone. Questa invenzione ebbe un grande successo a livello commerciale, surclassando il fonografo.



*Figura 29 - Il grammofono di Berliner (Stradeejay.it)*

Nel 1890 la società britannica Electrophone dà vita al primo dispositivo che consentiva ai propri clienti di connettersi agli ascolti dal vivo degli spettacoli nei teatri dell'opera di Londra: in sostanza, il primo esempio di *cuffia*. Gli abbonati al servizio potevano ascoltare l'esibizione attraverso un paio di enormi auricolari collegati sotto il mento, tenuti da una lunga asta. Inizia così anche il percorso dell'ascolto in cuffia.

1895

Dalle esperienze tratte nei precedenti vent'anni, Edison brevettò il *Cinetofono*, che permetteva di incidere una registrazione sonora sulle brevi pellicole usate per le riprese cinematografiche, le quali, sfruttando un supporto per la riproduzione (ad es. il grammofono), permettevano di ascoltare i suoni registrati durante le proiezioni delle immagini. Il problema di questo strumento era l'assenza di sincronizzazione tra suono e immagini, motivo per il quale l'invenzione non fece fortuna. Si può dire, quindi, che il cinema sia nato con il tentativo di avere il sonoro durante le proiezioni, soltanto che gli studios, ormai attrezzati per i film muti, non giudicarono interessanti le scoperte del loro secolo, perciò decisero di rimandare questa eventualità al futuro. Le tecnologie erano ormai tali per poter dare vita alle prime case discografiche, che spesso

coincidevano con le aziende produttrici di fonografi e grammofoni, tanto che lo stesso Emile Berliner fondò, insieme a Eldrige R. Johnson, la Victor Talking Machine Company nel 1901 a Camden, negli Stati Uniti. Prima di loro vennero la Columbia Records, fondata negli Stati Uniti da Edward Easton nel 1888 (la più antica casa discografica al mondo) e la Gramophone Company di William Barry Owen e Edmund Trevor Lloyd Wynne Williams nel 1898 (Regno Unito).

**1920** Quando la Warner Bros rischiò la bancarotta, decise di giocare l'asso nella manica prima del definitivo fallimento, lanciando il primo film sonoro, che divenne un cult. "Il cantante di jazz" di Alan Crosland fu un successo, imponendo alle case di produzione americane la rincorsa alle nuove tecnologie. In pochi anni la tecnica venne perfezionata, sancendo la nascita di una nuova vita per il cinema. Nacquero lavori come la sonorizzazione e il doppiaggio.

Analizzando le tecnologie del periodo 1877-1920 possiamo dedurre che la prima tecnica di registrazione e di riproduzione di un suono mai adottata sia stata la *monofonia*, ovvero tutto ciò che viene registrato e riprodotto da un singolo canale. Un'altra peculiarità della monofonia è la mancanza degli elementi atti ad individuare i caratteri spaziali dei suoni stessi (Monofonia, Treccani).

Ciò che, invece, consideriamo quotidianamente come monofonia è l'ascolto di un segnale audio che abbia lo stesso livello di dB sia nell'ascolto dall'orecchio sinistro (prossimamente inteso anche come Left - L), che nell'ascolto dall'orecchio destro (prossimamente inteso anche come Right - R). Quando ascoltiamo un suono monofonico dalle cuffie abbiamo la percezione di ascoltarlo al centro della nostra testa, esattamente come se avessimo una fonte sonora che emette musica a  $0^\circ$  rispetto alla nostra posizione.

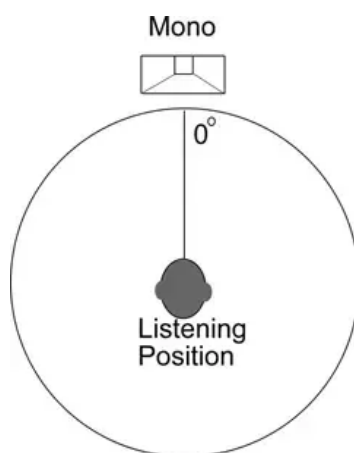


Figura 29 - Posizione d'ascolto da una sorgente monofonica

## 2.5.2 Le sperimentazioni e la celebrazione della stereofonia

**1920** L'ingegnere radiofonico Franklin Doolittle ha depositato alcuni brevetti per la registrazione a 2 canali (nel 1921) e per la trasmissione (nel 1924). La stazione radio di sua proprietà, WPAJ, ha iniziato a trasmettere nel Connecticut il suono catturato con due microfoni, emettendolo su due frequenze radio separate (Paul, 2009).

In quel periodo storico iniziarono ad essere utilizzate un nuovo tipo di cuffie che utilizzavano al loro interno dei piccoli altoparlanti a ferro mobile (il primo tipo di altoparlante elettrico). Tra le più celebri e sensibili c'erano quelle dell'azienda britannica Kolster-Brandes Ltd, produttrice di apparecchi radiofonici.

**1930** Nel 1931, l'ingegnere britannico Alan Blumlein (1903–1942) depositò un brevetto ampiamente considerato come segno della nascita della stereofonia (Blumlein, 1931). Ha definito la sua invenzione "stereo", un sistema di trasmissione binaurale, e ha spiegato che è possibile creare un'impressione sonora realistica e multidirezionale utilizzando due percorsi acustici (left - right channel) (Roginska & Geluso, 2017). Inoltre, ha affermato che le informazioni di fase e ampiezza acustica catturate da due microfoni direzionali possono essere ricostruite utilizzando solo due altoparlanti (Blumlein, 1931). Sulla base di questi principi, Blumlein ha spiegato che un sistema a 2 canali era in grado di produrre una replica simile a un'immagine sonora direzionale originale. Ai suoi tempi furono sviluppati concetti di sistema audio direzionale più elaborati che utilizzavano un gran numero di altoparlanti, ma alla fine lo stereo a 2 canali divenne lo standard mondiale. Va notato che le idee di Blumlein si estendevano oltre la riproduzione stereo a 2 canali. Particolarmente innovativa fu anche la sua creazione della tecnica di microfonação chiamata "coppia Blumlein" (Rumsey, 2001).

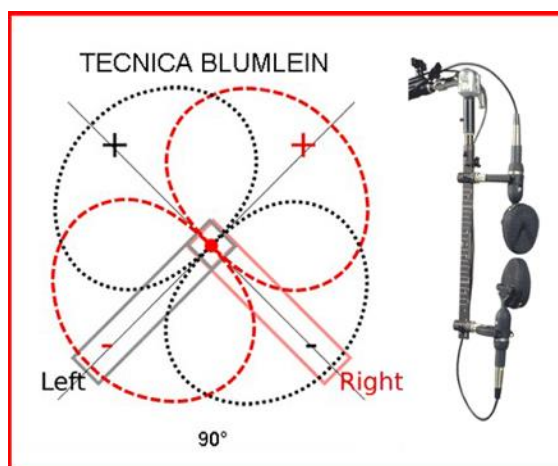


Figura 30 - Coppia Blumlein ([thesoundmaster.it](http://thesoundmaster.it))

Il lavoro di Blumlein era così in anticipo sui tempi che, quando morì in un incidente aereo all'età di 38 anni, lui e il suo lavoro erano ancora in gran parte sconosciuti.

Per comprendere a pieno le scelte della ricerca in questo periodo storico è necessario considerare il contesto socioculturale: perfino una radio monofonica era un grande investimento per una famiglia media e gli investimenti sulla ricerca provenivano da società cinematografiche, con l'intento di avere un risvolto commerciale sulla riproduzione audio multicanale, per attirare gli ascoltatori verso un'esperienza multimediale che superasse qualsiasi cosa potessero sentire a casa. Tuttavia, molti dei primi tentativi furono prototipi costosi, spesso finiti per essere abbandonati.

**1940** Affascinato dalla possibilità di sperimentare l'ambiente sonoro, Walt Disney realizza il film *Fantasia*, primo film nella storia del cinema ad avere un suono a più canali (detto anche multicanale) completamente realizzato ad hoc per la proiezione, rinominato *Fantasound*. Per riuscire al meglio in questo intento fu fondamentale la collaborazione con il direttore d'orchestra Leopold Stokowski, che conosceva diverse tecniche sulla psicoacustica e sulla percezione stereofonica, grazie agli esperimenti condotti sia nei laboratori Bell negli anni '30, che con Harvey Fletcher nel 1933. La Philadelphia Orchestra suonò la colonna sonora. Durante la registrazione, in sincrono con le immagini, vennero utilizzate per la prima volta una traccia click (con il tempo del metronomo da seguire in cuffia) e otto registratori (di tipologia ottica<sup>6</sup>), i quali permettevano la separazione delle sezioni orchestrali. Sei dei canali utilizzati vennero dedicati all'orchestra, un canale all'ambiente e l'ultimo per un mix balance. L'avanguardia delle tecniche utilizzate è stata tale da anticipare le sovraincisioni (dette anche overdub), realizzate per la prima volta durante questo contesto. La durata intera delle registrazioni fu di circa 40 giorni.

---

<sup>6</sup> Sistema basato sulla registrazione del sonoro su cellula fotoelettrica, divenne un sistema standard audio per le pellicole negli anni '30, prima dell'avvento di nuove tecnologie come la registrazione magnetica o altro.

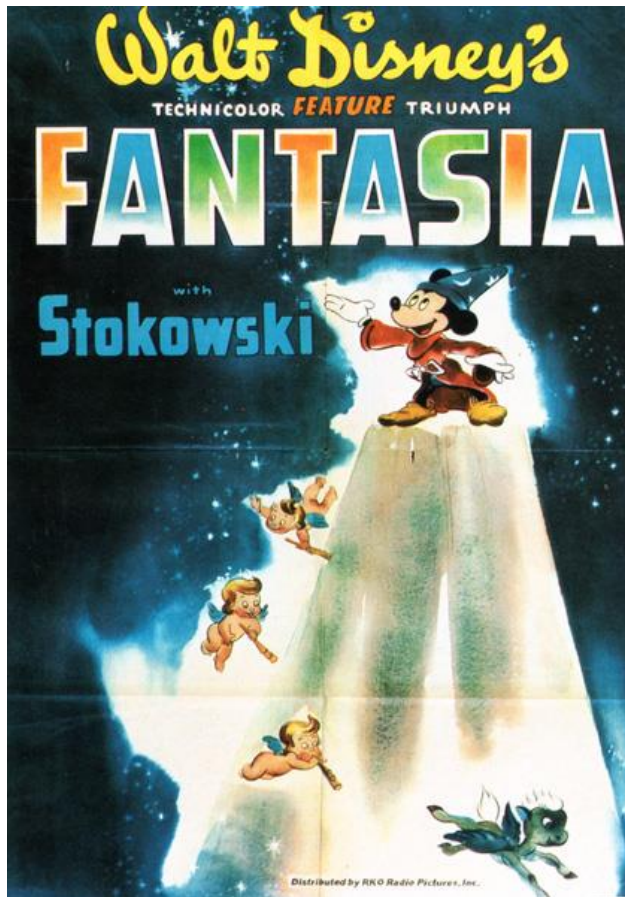


Figura 31 - Fantasia di Walt Disney (1940) - (Mymovies.it)

Restava, però, il problema della riproduzione durante la proiezione del film: il Fantasound necessitava di 54 altoparlanti per ogni teatro, un costo complessivo di circa 85.000 dollari. Fu questa la motivazione per cui solo due sale vennero dotate di questo sistema: il Broadway Theater di New York e il Carthay Circle Theater di Los Angeles. In questi teatri venivano utilizzati due proiettori: il primo proiettava l'immagine e serviva anche da backup audio in caso di guasto della colonna sonora (perché possedeva un mix mono di tutto il sottofondo sonoro); il secondo, sincronizzato all'immagine proiettata dal primo, aveva quattro tracce audio mono: la traccia di controllo (che forniva eventuali cambiamenti di ampiezza e di toni da apportare), il left channel, il right channel e il canale centrale. Un'altra invenzione legata al Fantasound è il "panpot", un filtro che tuttora si utilizza per regolare sui mixer il balance della stereofonia, ideato dall'ingegnere a capo della Disney, William Garity, con la volontà di realizzare il movimento del suono dal fronte al retro della sala.





Figura 32 - Panpot per il bilanciamento left/right (Soundonsound.com)

Fantasia debuttò il 13 novembre del 1940 al Broadway Theater di New York. Tutto l'audio venne mixato live da sei ingegneri del suono sotto la supervisione del direttore d'orchestra.

## 1950

I film stereo divennero di moda: fornivano diversi canali frontali, più almeno un canale dedicato ai diffusori nella zona posteriore della sala, ai quali erano dedicati gli effetti. Questi divennero noti in futuro come canali surround.

Il formato sonoro multicanale raggiunse un successo commerciale negli anni '50. La registrazione passò da ottica a magnetica, successivamente incollata sulla pellicola. Il suono stereo cinematografico (a quattro canali) era ben diverso dal suono stereo (a due canali), che arrivò nelle case nel 1958, poiché dipendeva ancora dalla tecnologia del fonografo. Fu proprio in quell'anno a nascere la Koss Corporation nel Milwaukee per mano di John C. Koss, un audiofilo e musicista jazz, che iniziò a produrre le prime cuffie stereo, promettendo un ascolto ad alta fedeltà. Commercialmente fu facile implementare due canali stereo rispetto al canale mono, per questo motivo divenne una norma che ancora oggi è presente nella quasi totalità dei dispositivi di uso comune.

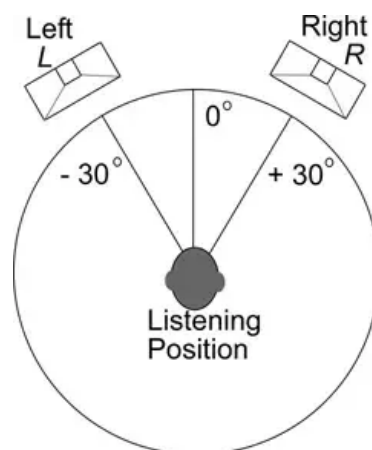
A rendere definitivamente efficace l'avvento della stereofonia fu l'introduzione della radio a transistor FM a due canali stereo, che servì da impatto diretto nei confronti dei consumatori.

## 1960

Gli anni '60 sono riconosciuti come gli anni della rivoluzione. Per quanto riguarda gli audiofili, in questo periodo storico furono due le introduzioni che ancora oggi persistono. La radio Sony EFM-117J, rilasciata nel 1964, è stato uno dei primi dispositivi ad avere la

predisposizione per il connettore jack 3,5 mm, che verrà utilizzato in futuro su tutti i dispositivi portatili. Inoltre, intorno alla metà degli anni '60, Ray Dolby fonda i Dolby Laboratories (1965). Il primo prodotto commerciale che viene dato alla luce è uno strumento per la riduzione del rumore nell'audio analogico chiamato Type A - Dolby Noise Reduction. Il congegno era destinato alle società discografiche. Riuscire ad avere una registrazione pulita senza interferenze, a quei tempi, era davvero come trovare la gallina dalle uova d'oro. Il sistema funzionò a tal punto che Dolby venne incaricato di produrre una versione per i consumatori, che uscì nel 1968 col nome di Type B - Dolby Noise Reduction, destinata all'elettronica di consumo.

Ricapitolando, la caratteristica principale dell'ascolto binaurale è la possibilità di localizzare un suono. Il termine *stereofonia* si utilizza nelle tecniche di registrazione e riproduzione sonora a due canali (L-R), in cui i segnali hanno valori differenti tra il canale Left e il canale Right, permettendo di dare maggiore sensazione di spazialità ai suoni. È molto importante, per non dire fondamentale, ricordarsi che quando si parla di "spazialità" in termini stereofonici ci si riferisce ad uno spazio limitato a circa 135° sul piano orizzontale. Lo stesso spazio che i teorici di acustica definirebbero nella posizione ideale di ascolto, con due casse anteriori all'ascoltatore poste a +30° Right e a -30° Left.



*Figura 33 - Posizione d'ascolto della stereofonia*

All'interno di un mix stereo è di uso frequente "giocare" creativamente con le posizioni degli strumenti, utilizzando automazioni sui panpot dei canali per dare maggiore volume a ciò che si trova Left o Right. Maggiori informazioni sugli standard del mix stereo verranno approfonditi nel terzo capitolo, nel paragrafo dedicato al mixing.

La differenza tra la percezione binaurale di un singolo suono (in cui un unico fronte d'onda della sorgente viene ascoltato separatamente dalle due orecchie, come nell'ascolto naturale) e quelle che Snow (Steinberg & Snow, 1934) chiamava "situazioni stereofoniche" (in cui più segnali di altoparlanti vengono utilizzati per creare "immagini

fantasma ') è stata riconosciuta anche da Alan Blumlein, il cui brevetto del 1931 (Blumlein, 1931), ha dimostrato che introducendo differenze di ampiezza tra una coppia di altoparlanti sarebbe possibile creare differenze di fase tra le orecchie, simili a quelle dell'ascolto naturale.

I dischi stereo divennero ampiamente disponibili al pubblico negli anni '60. Molte registrazioni sono state pubblicate come registrazioni monofoniche rielaborate artificialmente per dare un effetto stereofonico. L'album dei Beatles, Sergeant Pepper, della metà degli anni '60, è stata la prima registrazione pop multitraccia (quattro tracce) ed è stata pubblicata sia in stereo, che in mono. Sempre negli anni '60 fu lanciata la radio FM con capacità stereo, fornendo un livello di qualità mai sperimentato prima dalla maggior parte dei consumatori.

La riproduzione stereofonica a due canali (in termini standard internazionali "2-0 stereo", ovvero due canali anteriori e nessun canale surround) è spesso chiamata semplicemente "stereo" ed è il modo più comune che la maggior parte delle persone conosce per trasmettere alcuni contenuti spaziali in registrazione e riproduzione del suono. Negli standard internazionali che descrivono le configurazioni degli altoparlanti stereo, la nomenclatura per la configurazione è spesso nella forma 'nm stereo', dove n è il numero di canali anteriori e m è il numero di canali posteriori o laterali (quest'ultimo si incontra solo nei sistemi surround). Detto questo, è stata condotta una grande quantità di ricerche sull'"effetto stereofonico" di base e sulla sua ottimizzazione, in particolare per i sistemi a due canali, come esemplificato nell'Antologia AES sulle tecniche stereofoniche (Eargle, 1986).

È necessario un ritardo compreso tra 0,5 e 1,5 ms affinché un segnale appaia completamente a sinistra o completamente a destra a  $\pm 30^\circ$ , a seconda della natura del segnale stesso. A volte è conveniente lavorare con segnali stereo nel cosiddetto formato "somma e differenza", poiché consente il controllo della larghezza dell'immagine e del bilanciamento del segnale ambientale. Il segnale somma o principale è indicato con 'M' o Mid (informazioni provenienti dal centro), e si basa sull'aggiunta di segnali L e R, mentre il segnale differenziale o laterale è indicato con 'S' oppure Side, e si basa sulla sottrazione di R da L per ottenere un segnale che rappresenta la differenza tra i due canali.

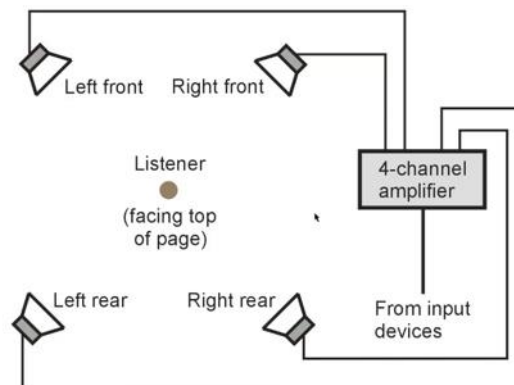
Il sistema stereofonico sviluppato da Steinberg e Snow negli anni '30 utilizzava tre canali. Due canali sono diventati la norma nei sistemi consumer solo per motivi di economia e convenienza, e in particolare perché era molto più semplice registrare due canali su un disco analogico rispetto a tre.

Ci sono vari vantaggi dello stereo a tre canali, motivo per il quale questo sistema è ancora utilizzato negli screen channel dei cinema. In primo luogo consente un palcoscenico sonoro anteriore leggermente più ampio rispetto allo stereo a due canali, se lo si desidera, perché il canale centrale agisce per "ancorare" l'immagine centrale, mentre gli altoparlanti sinistro e destro possono essere posizionati maggiormente ai lati

(circa  $\pm 45^\circ$ ). Si noti, tuttavia, che, nell'attuale standard audio surround a 5 canali, gli altoparlanti L e R sono infatti posizionati a  $\pm 30^\circ$ , per compatibilità con lo stereo a due canali.

### 2.5.3 La rovina della quadrifonia e l'Ambisonics

**1970** Appena una scoperta entra nelle case della gente, nasce subito l'esigenza di portare all'estremo quello che si ottiene da questa tecnologia. Siccome la monofonia venne soppiantata dalla stereofonia, negli anni '70 prese piede l'idea, tra gli audiofili, che ci fosse la possibilità di andare oltre ai due canali stereofonici. Iniziarono le sperimentazioni e i tentativi per costituire un sistema quadrifonico ("quad") all'interno delle proprie abitazioni. L'idea fu quella di posizionare, all'interno di una stanza, quattro altoparlanti: due anteriori all'ascoltatore (come avviene nella stereofonia) e due posteriori in posizione speculare. Nonostante la buona volontà, questo sistema non ebbe successo, a causa dei problemi di cancellazione di fase che sorgevano ponendo le coppie di casse poste l'una di fronte all'altra.



*Figura 34 - Posizione d'ascolto nella quadrifonia  
(Youtube.com - Stan Gibilisco)*

Il primo sistema "quad" fu proposto nel 1968, in realtà, da Peter Scheiber, che sviluppò un sistema per comprimere quattro canali analogici in soli due a scopo di archiviazione, ricostruendo i quattro canali originali sotto determinati vincoli di separazione dei canali e artefatti di fase (Torick, 1998; Davis, 2003). In ogni caso, nonostante l'aggressiva commercializzazione dei formati quad negli anni '70, il formato non riuscì a raggiungere, come descritto, il successo commerciale, provocando una diffusa perdita di fiducia sul futuro della tecnologia del suono 3D.

La quadrifonia non fu, però, un totale insuccesso, perché l'idea di utilizzare più di due canali fu di grande ispirazione nella musica, nel cinema (come vedremo tra poco con la nascita del Dolby Surround) e nella ricerca.

**1973** È proprio dalle rovine della quadrifonia che un matematico (e appassionato di audio) inventò uno schema di codifiche che si è distinto nel tempo. Michael Gerzon inventò nel '73 lo schema di Gerzon, detto anche Ambisonics, il quale prevedeva l'uso di funzioni di base armoniche e sferiche che potessero codificare le porzioni di un campo sonoro originato attorno alla posizione di un ascoltatore (Gerzon, 1973). Ha le sue basi teoriche anche nei lavori di Barton e Fellgett, di cui si possono trovare buoni riassunti negli articoli scientifici di Gerzon (1973, 1974, 1977). Ha anche la sua origine nel lavoro intrapreso in precedenza da Cooper e Shiga (1972).

L'intelligenza di questo sistema è quella di sfruttare le figure polari dei microfoni per ricreare una tridimensionalità in fase di registrazione. Ambisonics mira a offrire un approccio gerarchico completo (diviso per ordini) per il rilevamento, l'archiviazione, la trasmissione e la riproduzione del suono. Il primo ordine, nonché la vera eredità degli studi di Gerzon, è stato nominato *FOA - First Order Ambisonics* ed è composto da 4 canali.

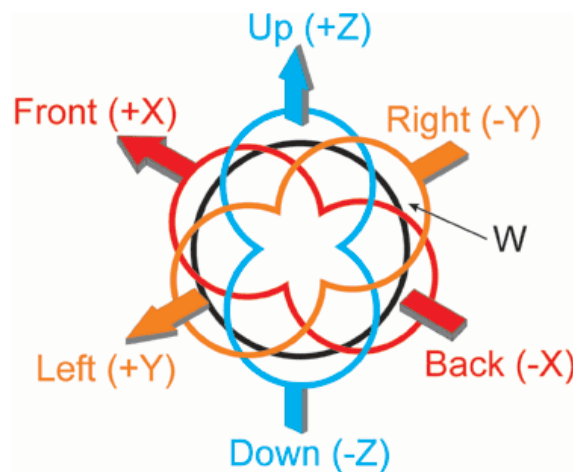


Figura 35 - First Order Ambisonics - FOA (Soundonsound.com)

I segnali microfonicici fanno riferimento a quattro canali e alle seguenti forme polari:

- 1 microfono: figura omnidirezionale ( O ).
- 3 microfoni: figura-a-otto ( 8 )

Le tre figure-a-otto simulano la posizione delle coordinate spaziali del piano cartesiano (X, Y, Z). La figura omnidirezionale (W) vuole andare a colmare i gap rimanenti, con l'idea che, all'unione di tutte le figure polari in questione, si vada a creare una nuova figura polare simile a una sfera (Gerzon, 1975). Per identificare meglio X e Y, esse si considerino rispettabilmente Mid e Side nella tecnica Mid/Side precedentemente descritta nella stereofonia. È molto importante sottolineare come da questi studi si inizi a considerare la componente verticale nella registrazione e nella riproduzione sonora.

Esistono numerosi formati nel sistema Ambisonics:

*A-Format*: per la ripresa microfonica.

*B-Format*: per l'elaborazione dei segnali in A-Format.

*C-Format*: per il broadcasting.

*D-Format*: per il decoding, l'encoding e la riproduzione.

Un formato noto come UHJ<sup>7</sup>, descritto originariamente da Gaskell del Dipartimento di ricerca della BBC (Gaskell, 1979), viene utilizzato anche per codificare informazioni ambisonic multicanale in due o tre canali pur mantenendo una buona compatibilità mono e stereo per gli ascoltatori "non surround". Ambisonics potrebbe quindi essere considerato il successore teorico dello stereo coincidente su due altoparlanti, poiché è la logica estensione dei principi di Blumlein al suono surround (Rumsey, 2001).

Come viene spiegato dettagliatamente nel caposaldo "Spatial Audio" di Francis Rumsey (2001): "A-Format (FOA) è costituito da quattro segnali provenienti da un microfono con quattro capsule sub-cardioidi orientate. Queste sono capsule montate su quattro facce di un tetraedro e corrispondono ad anteriore sinistro (LF), anteriore destro (RF), posteriore sinistro (LB) e posteriore destro (RB), sebbene due delle capsule puntino verso l'alto e due puntano verso il basso.



Figura 36 - Sennheiser Ambeo VR Mic - 4 Channel A-Format FOA

---

<sup>7</sup> UHJ è stato sviluppato dal team Ambisonics, incorporando il lavoro svolto dalla BBC (sul loro sistema quadrifonico, Matrix H), da Duane H. Cooper (sul sistema quadrifonico UD-4/UMX della Nippon Columbia) e da altri ricercatori, basandosi sull'allora attuale versione di Ambisonics, il System 45J. Le iniziali indicavano alcune delle fonti incorporate nel sistema: U da Universal (UD-4); H dalla matrice H; e J dal sistema 45J.

B-format (FOA) è costituito da quattro segnali che tra loro rappresentano le componenti di pressione e velocità del campo sonoro in qualsiasi direzione. Un segnale in B-Format può essere derivato da un microfono in A-Format. Per derivare segnali in B-Format da questi segnali a coppia di capsule si utilizza la tecnica della somma e della differenza. Pertanto:

$$\begin{aligned} X &= 0,5 ((LF - LB) + (RF - RB)) \\ Y &= 0,5 ((LF - RB) - (RF - LB)) \\ Z &= 0,5 ((LF - LB) + (RB - RF)) \end{aligned}$$

Infine, W, essendo una componente di pressione omnidirezionale, si ricava semplicemente sommando le uscite delle quattro capsule in fase, quindi:

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

(W ha un'equalizzazione che prevede un boost alle frequenze molto basse, poiché deriva da capsule microfoniche che non hanno una risposta in frequenza particolarmente estesa in quel range).

Prendendo  $\vartheta$  come angolo di incidenza nel piano orizzontale (azimut), e  $\eta$  come angolo di elevazione sopra l'orizzontale (elevation), allora in B-Format i modelli polari dei diversi segnali possono essere rappresentati come segue:

$$\begin{aligned} X &= \sqrt{2} \cos\vartheta \cos\eta \\ Y &= \sqrt{2} \sin\vartheta \cos\eta \\ Z &= \sqrt{2} \sin\eta \\ W &= 1 \end{aligned}$$

C-Format è costituito da quattro segnali L, R, T e Q, conformi alla gerarchia UHJ, e sono i segnali utilizzati per la trasmissione, compatibili con mono o stereo:

$$\begin{aligned} L &= \text{canale sinistro} \\ R &= \text{canale destro} \\ T &= \text{canale che consente una decodifica orizzontale più accurata} \\ Q &= \text{canale contenente informazioni sull'altezza} \end{aligned}$$

Se  $L + R$  è definito come  $\Sigma$  (simile a M, oppure Mid, in MS stereo) e  $L - R$  è definito come  $\Delta$  (simile a Side, in MS stereo), allora:

$$\begin{aligned} \Sigma &= 0,9397W + 0,1856X \\ \Delta &= j (-0,3420W + 0,5099X) + 0,6555Y \\ T &= j (-0,1432W + 0,6512X) - 0,7071Y \\ Q &= 0,9772Z \end{aligned}$$

La gerarchia in C-Format o UHJ consente una varietà di forme di codifica a matrice per segnali stereo, a seconda della quantità di informazioni spaziali da trasmettere e del numero di canali disponibili.

Per concludere, i segnali in D-Format sono quelli distribuiti agli altoparlanti per la riproduzione e vengono regolati in base alla disposizione scelta nella stanza”.

Il lavoro di Gerzon ha ispirato molti ricercatori che hanno sviluppato nuovi ordini e aggiunto canali al suo schema. Successivamente al FOA sono nati il *Second Order Ambisonics - SOA* a 8 canali e gli *High Order Ambisonics - HOA* dai 16 canali in poi.

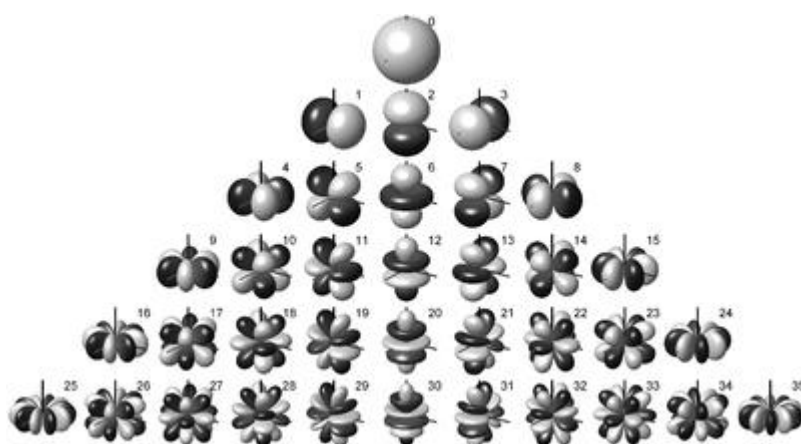


Figura 37 - Gerarchia ordini Ambisonics (Wikipedia.com)

La crescente miniaturizzazione della tecnologia audio ha consentito lo sviluppo di microfoni Ambisonic a 32 canali (Manola, Genovese & Farina, 2012) e 64 canali (O'Donovan & Duraiswami, 2010), oltre a strumenti software per la spazializzazione utilizzando HOA, che vedremo nel prossimo capitolo della tesi. Poiché la codifica Ambisonic è indipendente da qualsiasi sistema di riproduzione specifico, il rinnovato interesse odierno per il suono 3D ha portato alla richiesta di utilizzare Ambisonics come formato flessibile di produzione e distribuzione per contenuti audio 3D (Frank, Zotter & Sontacchi, 2015).





Figura 38- Microfoni HOA em32 Eigenmike® (mhacoustics.com)

#### 2.5.4 L'avvento del surround e la Dolby-mania<sup>8</sup>

**1975** Parallelamente al flop commerciale della quadrifonia e alla nascita dell'Ambisonics nella ricerca, l'audio cinematografico prosegue con la sua crescita. Iniziano a prendere piede le invenzioni dei Dolby Laboratories: il film *Arancia Meccanica* (*A Clockwork Orange*, 1971) di Stanley Kubrick utilizzò nelle fasi di pre-missaggio e masterizzazione il Dolby A, mentre *Callan di Don Sharp* (1974) lo utilizzò anche sulla pellicola cinematografica. La storia dei laboratori Dolby inizia, quindi, con i circuiti elettronici di attenuazione del rapporto segnale-rumore dovuto al fruscio dei nastri magnetici in fase di registrazione. Fu nel 1975 che Dolby presentò una nuova tecnologia sonora per le pellicole a 35 mm, chiamata inizialmente Dolby Stereo, la quale aggiungeva ulteriori canali audio rispetto alla stereofonia, utilizzando la tecnologia ottica che veniva usata dagli anni '30. Le esperienze maturate dalla quadrifonia e dall'audio multicanale, unite ai successi derivati da il Dolby A, permisero di codificare sulle pellicole cinematografiche due tracce fisiche con quattro canali di informazioni: sinistro, centrale,

---

<sup>8</sup> Molte informazioni di questo paragrafo sono state estrapolate da *Dolby® Atmos® Next-Generation Audio for Cinema - White Paper*.

destro e surround (posteriore). Il primo film in formato Dolby Surround fu Star Wars di George Lucas (1977), che segnò la nascita dell'audio multicanale nel cinema e vinse il premio Oscar per il sonoro.

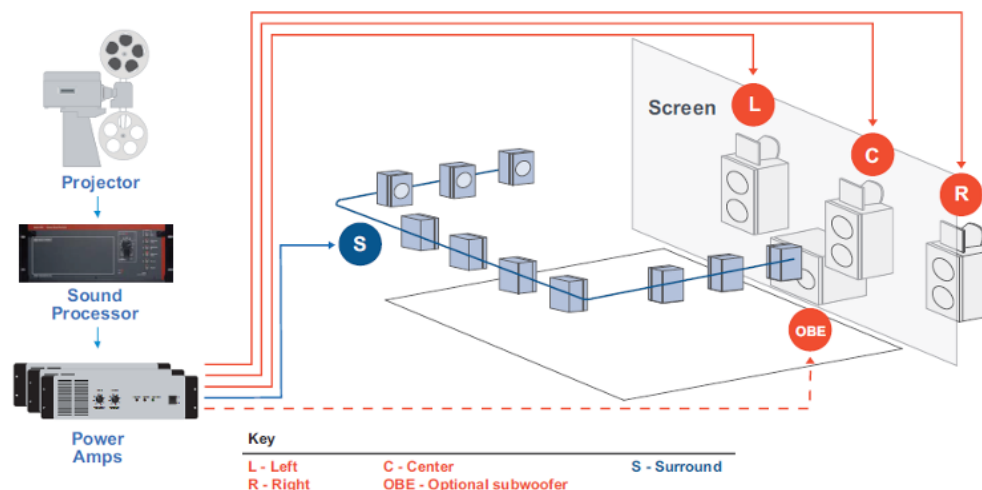


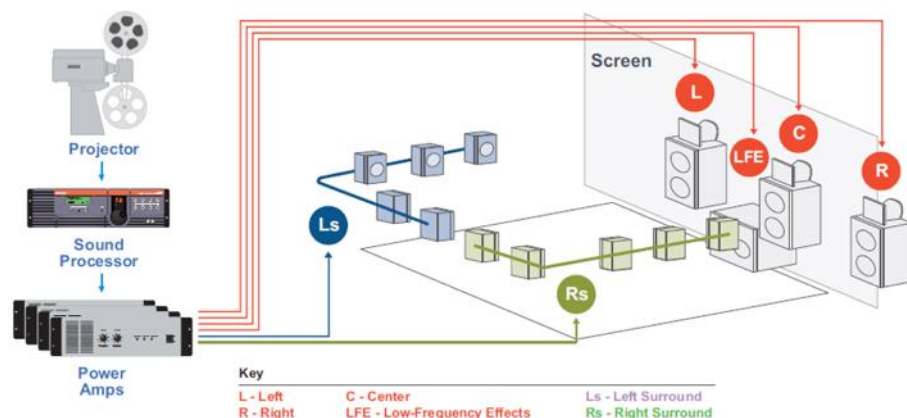
Figura 39 - Dolby Stereo / Dolby Surround (Dolby® Atmos® Next-Generation Audio for Cinema - White Paper)

Gli anni '80 sono alle porte e nel 1979 un oggetto segna il mercato dei consumatori, il Walkman (Sony TPS-L2). Questo oggetto è una vera rivoluzione: comodo e portatile, forniva in dotazione un paio di cuffie stereo per ascoltare ovunque i propri tapes preferiti; era presente uno sdoppiatore che permetteva di collegare due paia di cuffie, rendendo l'ascolto un momento di condivisione e portando la musica fuori dalle mura domestiche; il tasto "hotline", una volta premuto, permetteva alle persone che stavano ascoltando contemporaneamente di poter dialogare utilizzando un piccolo microfono interno. I costi di questi primi modelli erano molto alti e quindi poco adatti ad un pubblico di giovani. Sony fece fortuna riducendo i costi di produzione, eliminando la funzione del tasto hotline e l'ascolto condiviso (seppur molto innovative dal punto di vista dell'usabilità), a favore di prezzi bassissimi e molto competitivi.

**1980**

Esplode la Dolby-mania, tanto da essere ancora in voga ad oggi. Dopo l'exploit di Guerre Stellari arrivò la certificazione di qualità sonora THX (Tomlison Holman Experiment) introdotta da George Lucas con "The Return of the Jedi" (Il Ritorno dello Jedi) del 1983. THX garantiva qualità tecniche e acustiche per la diffusione sonora nei cinema, venivano analizzati diversi parametri tra i quali la posizione degli speakers in relazione allo schermo o l'arredamento (dalle pareti, alle poltrone, fino, come ovvio, al materiale fonoassorbente). Con la venuta

degli impianti casalinghi Dolby Surround Pro Logic, anche THX con Home THX entrò a far parte dell'ambiente domestico. L'idea di Home Theater venne introdotta all'inizio degli anni '80 con l'intento di trasferire l'esperienza cinematografica a casa. L'ambizione era importante, perché era difficile riuscire a dare una resa di qualità senza conoscere gli ambienti dei consumatori. Grazie alle videocassette stereo o ai laser disc, i quattro canali originali codificati Dolby venivano decodificati da dei decoder realizzati per l'uso domestico (Dolby Surround Pro Logic). Verso la fine degli anni '80 Dolby si dedicò all'audio digitale, introducendo il Dolby Digital per le pellicole 35 mm. I canali a disposizione divennero sei, con la dicitura 5.1, che ne garantiva i seguenti: destro, sinistro, centrale, sinistro surround, destro surround, il tutto con a disposizione più di tutta la banda di frequenza udibile all'uomo (20-22050 Hz). Il sesto canale, invece, era dedicato agli effetti potenti a bassa frequenza (Low Frequency Emission – LFE) che venivano percepiti perlopiù sulla pelle grazie alle vibrazioni. Questi suoni, detti infrasuoni, poiché sentiti sul corpo, furono sviluppati con l'introduzione del Sensurround.



*Figura 39 - Dolby Digital 5.1 (Dolby® Atmos® Next-Generation Audio for Cinema - White Paper)*

Questo periodo storico è da segnalare anche per i numerosi sistemi proprietari proposti nel corso degli anni, basandosi in un modo o nell'altro su principi binaurali. A volte la natura dei sistemi proprietari era avvolta nel mistero, con l'affermazione che questa o quella scatola nera poteva magicamente trasformare immagini sonore noiose e piatte in splendidi spazi sonori tridimensionali, ma con poche informazioni su come ottenere ciò. Come descrive perfettamente Francis Rumsey in *Spatial Audio* (2001) per fortuna la maggior parte di questi dispositivi è stata brevettata e i brevetti sono di pubblico dominio se si ha la tenacia di cercarli. L'autore racconta anche alcuni dettagli interessanti sul

sistema olofonico brevettato da Hugo Zuccarelli nel 1980: “L'indagine sul brevetto (Brevetto europeo 0 050 100) rivela che il sistema è poco più di una *dummy head* accuratamente costruita, in cui i microfoni si trovano all'estremità del timpano nel condotto uditivo e una certa equalizzazione viene proposta per rimuovere l'effetto della risonanza del condotto. Viene proposta una parrucca per rendere la testa più realistica e viene simulata la cavità orale. Le dimostrazioni all'epoca erano considerate molto realistiche, sebbene fossero spesso gestite con cura sul palco e utilizzassero segnali abbastanza vicini alla testa o in movimento (che in genere danno segnali spaziali troppo esasperati). Gli effetti olofonici sono stati incorporati in numerosi album. Un interessante articolo di Barry Fox su Studio Sound (Fox, 1983) esamina parte della storia e del dibattito che circonda l'olofonia. Un sistema chiamato *QSound* ha ricevuto molta attenzione alla fine degli anni '80 e all'inizio degli anni '90, essendo un altro sistema che affermava di offrire un posizionamento completo della sorgente sonora da due altoparlanti. All'epoca era concesso in licenza per l'uso su numerosi album. Un'indagine sul brevetto europeo per QSound (n. 0 357 402) rivela che in realtà non si tratta di un sistema che si basa su tecniche di sintesi binaurale o cancellazione del crosstalk, ma si basa su differenze di fase e ampiezza dipendenti dalle frequenze derivate empiricamente tra i due canali di una coppia stereo, che danno l'impressione di essere sorgenti in posizioni diverse. Sebbene sia probabile che questi producano segnali binaurali alle orecchie dell'ascoltatore, corrispondenti strettamente agli HRTF delle sorgenti naturali per le posizioni delle sorgenti virtuali interessate, non viene fornita tale analisi. Più recentemente sono stati sviluppati numerosi sistemi proprietari binaurali/transaurali che si basano fortemente sull'elaborazione del segnale digitale per la sintesi di sorgenti virtuali utilizzando il filtraggio HRTF. La proprietà intellettuale specifica di ogni prodotto sembra generalmente risiedere nell'ottimizzazione dei dettagli HRTF per diverse situazioni, nell'efficienza computazionale e negli algoritmi di annullamento del crosstalk. Un esempio di tale sistema è Sensaura (ormai non più in sviluppo, nda), originariamente sviluppato dai Central Research Labs nel Regno Unito. Questo è il risultato di molte ricerche sulla conservazione dell'accuratezza timbrica nella catena del segnale dalla testa fittizia o da sorgenti sintetizzate, attraverso la cancellazione del crosstalk all'ascoltatore umano”.

**1990** Il primo film ad utilizzare Dolby Digital è datato 1992 ed è *Batman Returns* di Tim Burton.

Con il processore Dolby Digital Surround EX, brevettato da Lucasfilm per Star Wars Episode I – The Phantom Menace (La minaccia fantasma, George Lucas, 1999), venne aggiunto un canale centrale posteriore ai canali surround, per dare al regista la possibilità di effetti di “oltrepassamento”, facendo scorrere il suono da fronte a retro sala senza attraversare i lati. Dal punto di vista commerciale, il Dolby Digital è stato il punto di riferimento dei formati venduti ai consumatori, dal laser disc ai DVD, dalla trasmissione della TV digitale alle applicazioni multimediali.

In parallelo, il mondo della ricerca vede al centro l'Italia, in particolare il centro di ricerca CSELT di Torino, che inizia le sperimentazioni sul formato MPEG in questo periodo storico, che verrà riconosciuto internazionalmente come formato .mp3 nel 1997, cambiando radicalmente il mondo dell'audio.

Il primo lettore MP3 commercializzato fu l'MPMan F10 presentato nel marzo 1998 alla fiera CeBIT dalla compagnia sud-coreana "Saehan Information Systems", venduto dalla Either Labs nell'estate seguente per 250 dollari con 32 MB di memoria flash.

Il suono surround si espande in una scena sonora virtuale panoramica che include l'ambiente di ascolto o posizionando l'ascoltatore sulla scena sonora circondato da sorgenti sonore virtuali.

La disposizione degli altoparlanti e la configurazione dei canali sono specificate in ITU-R BS.775 (ITU-R, 1993).

Quando si trasferiscono programmi originariamente prodotti per il cinema su supporti televisivi (ad es. DVD), potrebbe essere necessario mixare nuovamente parte del contenuto del canale subwoofer nei canali principali a larghezza di banda completa. È importante che qualsiasi audio a bassa frequenza molto significativo per l'integrità del contenuto del programma non venga inserito nel canale LFE. I principali limiti del formato surround 5.1 sono in primo luogo che non era inteso per un'accurata capacità di imaging fantasma a 360° (poiché non comprende la componente verticale).

### **2.5.5 Il nuovo millennio, la musica liquida e l'estensione massima del surround**

**2001** È passato solo un anno dal millenium bug, quando la Apple di Steve Jobs presenta il primo iPod e, con lui, iTunes. Un dispositivo e un software che segneranno uno spartiacque per diversi settori nel mondo dell'audio. La musica diventa un file; il CD, arrivato al suo apice, inizia

un lento e inconsapevole declino; la discografia si affaccia a uno dei momenti peggiori dalla sua nascita.

Statisticamente ogni innovazione è strettamente connessa a una grande crisi, questa si sarebbe chiamata *pirateria*. Nel 2002 nasce eMule, uno dei software peer-to-peer più celebri, che permetterà l'agile scambio di musica, film e contenuti multimediali, distruggendo il mercato musicale e i suoi meccanismi. Ci vorranno almeno 10 anni per ristabilire i giusti equilibri dopo un cambiamento così grande.

**2007** Il nuovo millennio verrà ricordato per la rapida avanzata della tecnologia, riassunta anche dalla prima legge di Moore (ipotizzata per la prima volta nel 1965):

*«La complessità di un microcircuito, misurata ad esempio tramite il numero di transistor per chip, raddoppia ogni 18 mesi (e quadruplica quindi ogni 3 anni).»*

Dall'iPod si passa all'iPhone. Sulla linea temporale è quasi la frazione di un millisecondo, per l'essere umano è un'altra rivoluzione nel giro di 6 anni. Nasce, così, lo smartphone, e con esso le applicazioni. Un mondo permeato da telefonini con i tasti e Nokia 3310 diventa, in pochi anni, un mondo iperconnesso e social. Spotify viene rilasciato nel 2008, Facebook (già sviluppato nel lontano 2004) cresce vertiginosamente, lo stesso vale per YouTube.

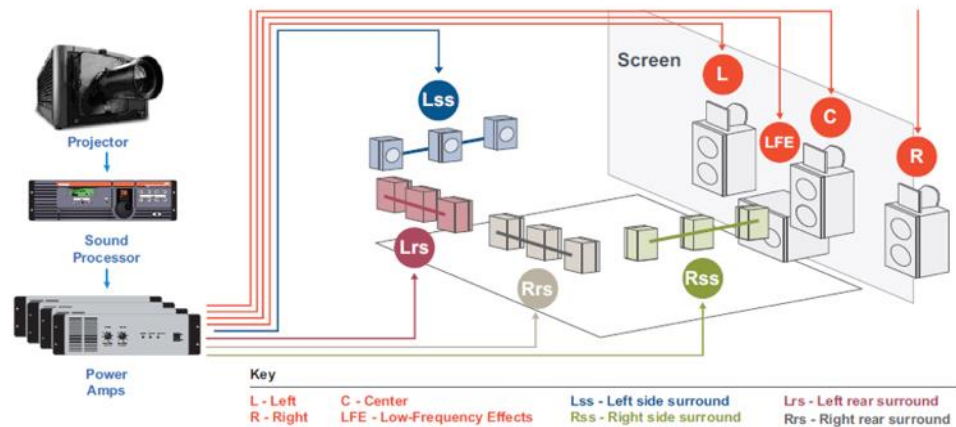
La fruizione dei contenuti cambia, l'attenzione si sposta e quello che un tempo era uno strumento per telefonare a parenti, amici e colleghi diventa il centro nevralgico della vita quotidiana. Insieme a tutto ciò, un altro strumento da non dimenticare a casa diventano gli auricolari (ormai con microfono integrato), utili per la musica, per telefonare e per lo svago (sport, pause, trasferimenti, ecc.).

**2010** Sono dovuti passare vent'anni prima dell'evoluzione del Dolby Digital in Digital Plus<sup>9</sup> (Dolby Digital 7.1), sintomo del fatto che la tecnologia introdotta negli anni '90 fu davvero innovativa. Questo formato, utilizzato anche nelle trasmissioni digitali HDTV e nei Blu-ray Disc, aggiunse due nuovi canali surround, dividendo il controllo del left surround e right surround in quattro parti. Il primo film ad utilizzare la

---

<sup>9</sup> “Maggiore efficienza di codifica, l'estensione del bitrate fino a 6,144 Mbps e l'aumento del numero di canali discreti a 7.1, tutto mantenendo una buona compatibilità con il Dolby Digital convenzionale. L'algoritmo dei Dolby Laboratories implementato per il presente formato, può prevedere, in realtà fino a 13 canali audio, suddividendo il left e right surround in sei, otto o dieci zone (9.1, 11.1, 13.1)” – Dolby Atmos: il futuro del cinema, Matteo Vaccaro, 2015.

tecnologia 7.1 è stato Toy Story 3 – La grande fuga di Lee Unkrich nel 2010.



*Figura 40 - Dolby Digital Plus 7.1 (Dolby® Atmos® Next-Generation Audio for Cinema - White Paper)*

Più di 4000 sale sono state adattate a questo formato. Con la digitalizzazione del sonoro è stato possibile applicare il Dolby Digital anche nell'Home Theater. Un altro sistema innovativo fu il Digital Theatre System (DTS) introdotto con Jurassic Park (1993, Steven Spielberg), ora diffuso su DVD e Blu-Ray. Il DTS era un sistema di codifica audio con perdita di informazione (lossy), ma anche senza perdita di informazione (lossless), nella versione DTS-HD Master Audio. Utilizzando 1,4 Mbps anziché 640 Kbps, il DTS possedeva una migliore fedeltà e dinamica rispetto al Dolby Digital. I dati venivano separati dalla pellicola in un CD dedicato alla colonna sonora.

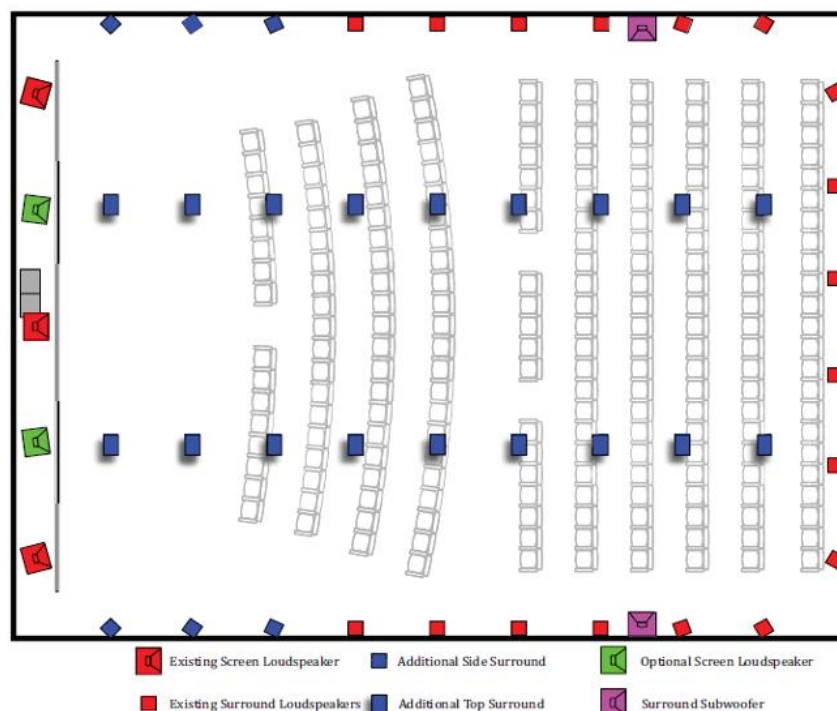
### 2.5.6 Dolby Atmos: l'audio spaziale diventa una realtà commerciale

**2012** Dall'avvento del 7.1 gli ingegneri Dolby hanno continuamente lavorato per migliorare il senso di immersione che lo spettatore vuole provare al cinema. Questa ricerca è culminata con l'annuncio, nel 2012, del nuovo sistema Dolby Atmos, divenuto lo standard della nuova generazione sonora. Il primo film in questo formato fu proiettato nel giugno dello stesso anno al Dolby Theatre di Hollywood, Los Angeles (California) e fu Brave (Ribelle) di Mark Andrews e Brenda Chapman, prodotto da Pixar. La prima sala europea ad adottare questo sistema fu la Cinesa Diagonal Mar di Barcellona, sono susseguite 25 sale in tutto il mondo, divenute mille nel 2013 e successivamente in continua espansione (Vaccaro, 2015).

Ci sono delle differenze di base che distinguono i sistemi 5.1 e 7.1 dal Dolby Atmos: il suono ha origine sopra la testa dell'ascoltatore (questo porta l'audio ad avere una nuova dimensione, l'altezza) e vi è un miglioramento della qualità sonora. La volontà alla base di questa tecnologia è di arrivare a replicare sempre più fedelmente l'esperienza reale, dove il suono ha origine in tutte le direzioni e non su un piano orizzontale. La tridimensionalità permette questo senso di realismo.

Quando si guarda un film, sono diversi i suoni che possono provenire dall'alto: un aeroplano, i cinguettii degli uccelli sui rami alti di un albero, lo sparo di un cecchino da un tetto. C'è anche un merito nella gestione dell'ampiezza dei volumi e delle equalizzazioni in questa nuova tecnologia, che permette di sfruttare array (gruppi) di speakers ai quali destinare alcuni suoni senza togliere importanza ad altri, rendendo il mix ancora più efficace.

Risulta fondamentale introdurre il concetto di *spazializzazione*. I suoni nella vita reale hanno delle fonti diverse nello stesso momento in cui arrivano alle nostre orecchie. Con il Dolby Atmos è possibile fissare il punto di origine di un suono e dargli un indirizzo da uno speaker (o da un array di speaker) ad altri nella sala. Come espresso nel White Paper dei Dolby Laboratories (2013) legato alla tecnologia Atmos, se un personaggio pone lo sguardo verso una sorgente fuori dallo schermo, il suono seguirà l'effetto espresso dall'immagine, e sarà determinante anche la posizione dell'ascoltatore per questo. Un'altra differenza con i sistemi precedenti di Dolby è che gli screen channel (a centro sala, dietro lo schermo) emettono solo l'audio presente durante l'azione che si vede sullo schermo (saranno quindi prediletti i dialoghi in campo), gli speakers che circondano la sala sono totalmente dedicati all'immersione della scena.





*Figura 41 -Posizione altoparlanti in una sala cinematografica con sistema Dolby Atmos  
(Dolby ® Atmos ® Next-Generation Audio for Cinema - White Paper)*

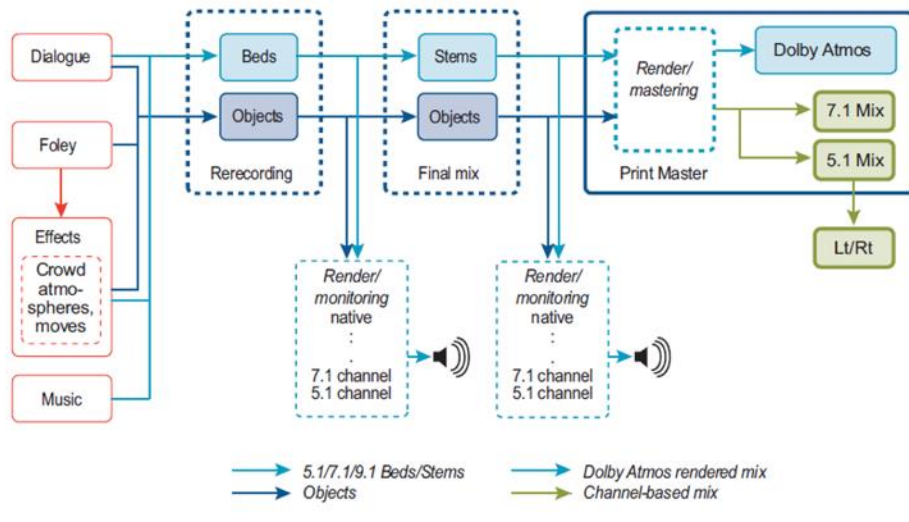
Il formato Atmos ha completamente cambiato il modo di indirizzare il suono alle casse, per questo sono stati introdotti gli *audio objects*, dotati di *positional data*. Gli audio objects sono un gruppo di elementi sonori con la stessa posizione fisica nella scena (e nell'auditorium).

Possono essere statici o possono muoversi. Essi sono controllati da alcuni metadati, detti *positional data*, che restituiscono la posizione del suono in un punto nel tempo. Quando il suono è in fase di pre-missaggio, i sound designer assegnano i metadati agli audio objects, i quali potranno muoversi tra le coordinate spaziali XYZ ed essere destinati agli speakers della sala grazie al processore, configurato da dei tecnici specializzati che indicheranno il numero di speakers e la loro posizione. Oltre agli audio object, Dolby Atmos utilizza i *beds*, suoni che non avranno spostamenti nella stanza. Questi sono submix di canale, conservati separatamente dal missaggio e successivamente combinati in un singolo bed come parte finale del master. Gli audio object e i metadati vengono portati alla fase di missaggio finale nei teatri di doppiaggio, i metadati vengono eventualmente editati e definiti, per poi essere monitorati con l'hardware Dolby Rendering Master Unit (RMU).

Le configurazioni dei canali possono essere di diverso tipo: 5.1, 7.1 o altre come il 9.1. Possono essere utilizzate 128 tracce, 7.1 beds channels più 120 audio objects per esempio. Per fare chiarezza, i beds sono gruppi di suoni destinati a singoli altoparlanti o gruppi di altoparlanti (come gli screen channels) che non avranno necessariamente caratteristiche spaziali, mentre gli audio objects sono posizionati nella stanza, con la possibilità di muoversi, e vengono renderizzati in real time dal processore in base alla locazione fisica degli altoparlanti. Il processore consente il rendering di beds e objects fino a 64 uscite audio.

In tutti i film, in fase di missaggio, i suoni vengono divisi tendenzialmente in tre categorie: dialoghi, musica ed effetti. Un sottogruppo di effetti viene chiamato *foley* (rumore dei passi, porta che sbatte, ecc.), creato dai foley artist o, detto in modo anacronistico, rumoristi. Dal set arrivano a questa fase centinaia di file registrati in presa diretta, che nelle sessioni vengono scremati e scelti. Vengono creati alcuni premix destinati ai canali bed contenenti i dialoghi in modalità mono e alcuni gruppi in loop. Successivamente i dialoghi sono fissati nell'ambiente sonoro come object. I Foley subiscono lo stesso procedimento. Effetti e foley vengono divisi in gruppi tra i quali l'atmosfera, le folle, i movimenti. La musica subisce un passaggio diverso, che porterà ad un premix musicale destinato ai beds. La musica definitiva, invece, verrà destinata agli speakers surround come object fuori dallo schermo, lasciando spazio agli screen channel per i dialoghi. Come preannunciato, la fase finale di missaggio avviene nei dubbing theatre, in cui il tecnico del suono usa i premix dei beds insieme agli objects e ai positional data per creare dei gruppi, detti stems, contenenti ad esempio dialoghi, musica, effetti, foley e sottofondo. Vengono tenuti come base musica ed effetti da

destinare al doppiaggio in lingua straniera. Ogni stem, quindi, è come un nuovo gruppo di suoni che consiste in canali bed e audio objects corredati di positional data.



*Figura 43 -Beds e Audio Objects  
(Dolby ® Atmos ® Next-Generation Audio for Cinema - White Paper)*

Durante la fase di mastering i metadati sono portati al formato Dolby Atmos. Il master è salvato nel formato MXF, Material Exchange Format, un contenitore di flussi audio/video con diversi formati al suo interno e con metadati inclusi. Successivamente avviene il passaggio da MXF a DCP (Digital Cinema Package), il formato digitale per le opere cinematografiche, attraverso il sistema Dolby Secure Content Creator (SCC200), che può includere anche i sottotitoli se necessario. Il DCP verrà crittografato tramite la specifica SMPTE, un timecode standard per i fotogrammi sui video e sulla pellicola cinematografica, per la perfetta sincronizzazione. Una volta consegnato il DCP alle sale che non saranno equipaggiate di Dolby Atmos, siccome il file MXF è inserito come traccia ausiliaria, qualora il cinema non possedesse il processore CP850 dedicato alla tecnologia Atmos, il file audio verrà riprodotto con le specifiche audio del processore adottato (5.1, 7.1). Il formato supporta audio a 24 bit PCM (Pulse Code Modulation, metodo di rappresentazione di un segnale analogico-digitale), con possibili frequenze di campionamento di 96 kHz o 48 kHz e la possibilità di contenere fino a 128 tracce. Il software di compressione genera quindi due differenti pacchetti di tracce, uno low-budget a 48 kHz e uno ad alta risoluzione da 96 kHz.



*Figura 44 -Processore CP850  
(Dolby ® Atmos ® Next-Generation Audio for Cinema - White Paper)*

Il processore Atmos CP850 necessita di una catena di lavoro digitale, a partire dal proiettore. Supporta fino a 64 altoparlanti. Tra le varie entrate e uscite al suo interno, è connesso un server digitale con il protocollo AES/EBU che utilizza 16 canali in entrata e in uscita. I dati sono trasferiti via connessione Ethernet (1000BaseT) dal server principale al processore per la decodificazione, la renderizzazione e la sincronizzazione. Per poter creare dei contenuti audio in Dolby Atmos sono necessari alcuni strumenti, che includono la Dolby Atmos Monitor Application e il Dolby Atmos Panner Plug-in (che verranno approfonditi nel capitolo 3 e nel capitolo 4).

La Dolby Atmos Monitor Application è utilizzata per monitorare visivamente l'audio renderizzato nella Dolby RMU e in più permette cambiare la tipologia di rendering da Atmos a 7.1 o 5.1, registrare e riprodurre una copia del master. Lo strumento per il panning Dolby Atmos Panner Plug-in consente di muovere il suono nella scena sui piani XYZ.

## **2.6. Audio spaziale in cuffia**

*Il suono coinvolgente prodotto da altoparlanti o cuffie ha la capacità di fornire un'illusione senza soluzione di continuità di realtà alternativa e cambiare il modo in cui ci relazioniamo e ci comportiamo con il suono (Roginska & Gehuso, 2017).*

Dopo i primi passi nell'ambito di ricerca dell'audio spaziale, prima di scegliere l'argomento su cui specializzarsi, ci si può trovare di fronte ad un bivio difficile: la scelta tra l'ascolto in cuffia o l'ascolto con altoparlanti.

L'utilizzo degli altoparlanti richiede un budget molto alto e una stanza trattata dove disporre meticolosamente le casse. Inoltre, la sperimentazione dovrà avvenire necessariamente all'interno della stanza o del laboratorio. Non basta, quindi, la buona

volontà del ricercatore o del dottorando per poter procedere ad esperimenti, qualora il proprio istituto non fosse dotato degli strumenti necessari per svolgere i test. Operare con gli altoparlanti, per un'università o un centro di ricerca, significa scegliere di dare un focus specifico a quel tipo di ricerca, considerando i costi che andranno a ricadere direttamente sul bilancio dei dipartimenti (e riducendo, quindi, le possibilità di altri laboratori).

Lavorare con le cuffie, invece, è molto meno costoso, escludendo i rari casi di test di altissimo livello con macchine estremamente sperimentali. Oltre ad avere un budget ridotto, la portabilità è estrema e si possono progettare esperimenti anche al di fuori del proprio laboratorio.

Cuffie e altoparlanti sono due mondi paralleli: il primo è diretto a una ricerca focalizzata sull'ascolto del singolo utente, con uno strumento che utilizziamo tutti i giorni, volto a tutti quei settori in cui l'utilizzo delle cuffie è fondamentale per la percezione acustica (musica, gaming, simulazione, sport, virtual reality); il secondo, invece, predilige un ascolto all'interno di uno spazio fisico, raggiunto volontariamente dall'utente per vivere un'esperienza attraverso gli stimoli acustici (cinema, concerti, eventi, installazioni museali, automotive).

Questa tesi ha lo scopo di soffermarsi sul formato binaurale e, quindi, di approfondire l'ascolto in cuffia come strumento per la riproduzione sonora e l'ascolto di audio spaziale.

Le cuffie sono nate dalla necessità di liberare le mani di una persona quando si utilizza un telefono.

Con la crescita esponenziale della tecnologia a partire dagli anni 2000, l'utilizzo degli auricolari è definibile come un vero e proprio fenomeno socio-culturale, poiché questi strumenti sono diventati quasi uno scudo di protezione dalla frenesia del mondo, una possibilità, in qualsiasi momento, per fruire dei nostri contenuti multimediali preferiti. Se le applicazioni sono state ideate anche con lo scopo di combattere le attese, le cuffie sono diventate un alleato in questa battaglia.

L'ambiente naturale di ascolto può essere definito come lo spazio acustico che occupiamo durante la nostra vita quotidiana. Uno spazio uditivo virtuale è un ambiente acustico creato attraverso l'uso di altoparlanti o cuffie progettate per sostituire o aumentare l'ambiente di ascolto naturale.

Le tecniche di riproduzione audio binaurale sfruttano i segnali uditivi spaziali naturali dell'uomo per ricreare un ambiente uditivo virtuale attraverso l'utilizzo delle cuffie (Roginska & Geluso, 2017).

Come descritto nei paragrafi precedenti, le nostre orecchie hanno una forma anatomica che differenzia l'ascolto da persona a persona. Le cuffie si frappongono nel processo d'ascolto della realtà che ci circonda, modificando il sistema HRTF in base alle tipologie di cuffie utilizzate. In sostanza, non solo ognuno di noi possiede diverse

caratteristiche di ascolto della realtà, ma anche diverse caratteristiche di ascolto della stessa cuffia.

Sono stati condotti alcuni esperimenti in cui venne somministrata a un tester l'HRTF di un'altra persona, bloccando i padiglioni auricolari e alimentando i segnali direttamente al condotto uditivo. Ciò che è stato scoperto è che la capacità di localizzazione di un segnale è notevolmente ridotta. Dopo poco tempo, però, le orecchie sembrano adattarsi alle nuove informazioni (Rumsey, 2001).

Si ritiene che la cosiddetta risonanza della conca (quella creata dalla cavità principale al centro del padiglione auricolare) sia responsabile della creazione di un senso di esternalizzazione, che permette alle persone di ascoltare un suono come se fosse all'esterno della propria testa.

Uno dei problemi principali di ricerca derivato dall'ascolto in cuffia, sul quale molti ricercatori stanno proponendo soluzioni hardware e software, è proprio la *In-Head Localization (IHL)*, la sensazione di percezione dei suoni all'interno della propria testa. Nel precedente paragrafo sono stati evidenziati i dubbi emersi nel corso degli anni sulla duplex theory di Lord Rayleigh's nel 1907, che hanno un riscontro anche con questo tema come ripreso da Roginska & Geluso nel 2017 nel caposaldo Immersive Sound: "La duplex theory non può spiegare la capacità dei soggetti di localizzare i suoni sul piano mediano verticale (direttamente davanti all'ascoltatore), dove i segnali interaurali sono minimi. Allo stesso modo, quando i soggetti ascoltano gli stimoli in cuffia, i suoni vengono percepiti come lateralizzati all'interno della testa anche se sono presenti differenze temporali interaurali e di intensità inerenti alla posizione di una sorgente esterna. I risultati di molti studi suggeriscono ora che queste carenze della duplex theory riflettono l'importante contributo della localizzazione del filtraggio dipendente dalla direzione che si verifica quando le onde sonore in ingresso interagiscono con le orecchie esterne, i padiglioni auricolari o altre strutture corporee come le spalle e il busto. Il segnale principale che il sistema uditivo umano utilizza per determinare l'elevazione di una sorgente sonora è lo spettro mono determinato dall'interazione del suono con i padiglioni auricolari (Wightman & Kistler, 1997). Tuttavia, piccole asimmetrie della testa possono fornire un debole segnale di elevazione binaurale. In particolare, c'è un range spettrale che si sposta in un range di frequenze che va da circa 5 kHz a 10 kHz, mentre la sorgente si sposta da 0° (direttamente davanti all'ascoltatore) a 90° (sopra la testa dell'ascoltatore), che è considerato il principale segnale di elevazione (Musicant & Butler, 1985; Moore, Oldfield & Dooley, 1989). Quando il suono si propaga da una sorgente alle orecchie di un ascoltatore, gli effetti di riflessione e rifrazione tendono ad alterare il suono in modi sottili e l'effetto dipende dalla frequenza. Ad esempio, per una posizione particolare, un gruppo di componenti ad alta frequenza centrati a 8 kHz può essere attenuato più di una banda diversa di componenti centrati a 6 kHz. Tali effetti o filtri dipendenti dalla frequenza variano anche notevolmente con la direzione della sorgente sonora. Pertanto, per una diversa posizione della sorgente, la banda a 6 kHz può essere più attenuata della banda di frequenza più alta a 8 kHz. È chiaro che gli ascoltatori utilizzano questo tipo di effetti dipendenti dalla frequenza per discriminare una posizione da un'altra. Esperimenti hanno dimostrato che i segnali provenienti nel

padiglione auricolare sono parzialmente responsabili dell'esternalizzazione o della sensazione "out of the head" (Gardner & Gardner, 1973; Oldfield & Parker, 1984a, b; Plenge, 1974; Shaw, 1974)".

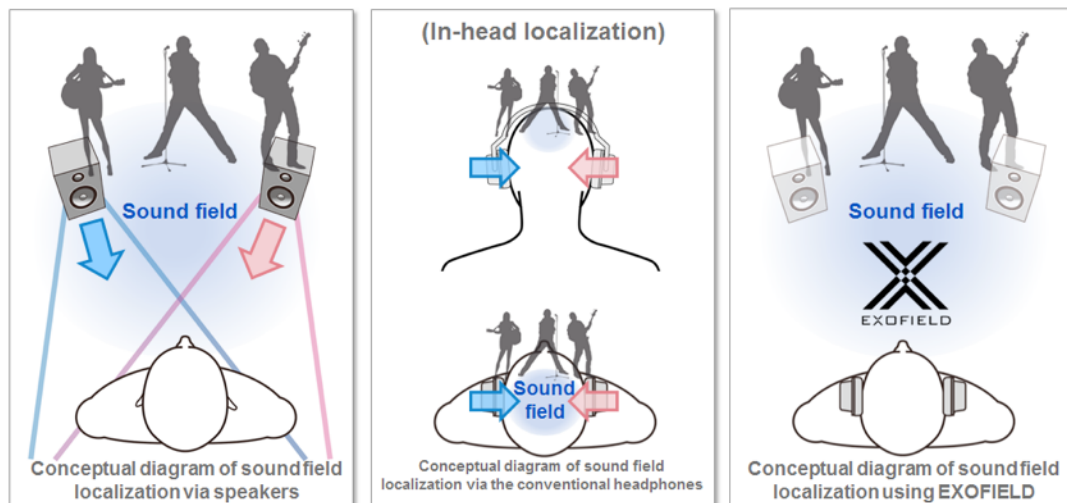


Figura 45 -In-Head Localization, diagramma concettuale  
(JVC Pro - Sistema Exofield) (JVC, 2018)

La localizzazione interna alla testa è una caratteristica molto comune e non è per nulla un problema nel momento in cui utilizziamo gli auricolari per parlare al cellulare o per ascoltare la radio. È ben diverso se si vuole progettare un'esperienza sonora a tre dimensioni. Tra le soluzioni provate vi è stata la progettazione di cuffie che stimolano la risonanza della conca dalla direzione anteriore, al di fuori dell'ingresso del condotto uditivo, in modo da sovrapporre determinate caratteristiche uditive del singolo ascoltatore ai segnali binaurali riprodotti, migliorando così la percezione "out of the head" dell'ascolto in cuffia (Tan e Gan, 2000).

Inoltre, come descritto nell'articolo di JVC Pro (2018), e ripreso dalla Figura 45, un'altra soluzione è il tracciamento di un profilo personale dell'HRTF, che renderebbe più semplice il controllo delle frequenze in fase d'ascolto.

Trovare una soluzione diminuirebbe il gap con l'ascolto rispetto agli altoparlanti, che permettono l'esternalizzazione dei suoni tramite la distanza dall'ascoltatore e attraverso la sensazione fisica delle basse frequenze (percepita sulla nostra pelle).

Un altro problema di ricerca ricorrente è la distinzione corretta della localizzazione sonora durante l'ascolto in cuffia. La percezione delle differenze di ascolto sull'asse Y - Front/Back e sull'asse Z - Up/Down è molto complesso. Nell'errata percezione della localizzazione da parte dei tester sono diversi gli interrogativi che possono interrompere il processo di analisi, tra i quali pensare di aver scelto la cuffia sbagliata, aver progettato male un esperimento, aver utilizzato un software di spazializzazione (che vedremo nel prossimo capitolo) con un algoritmo acerbo. Le persone, durante l'ascolto tramite altoparlanti, riescono a percepire chiaramente dove localizzare un suono, nonostante nella fruizione di un film, ad esempio, i suoni provengano da molte sorgenti differenti.

Al contrario, nell'ascolto in cuffia, la In-Head Localization, sommata ai problemi di Front/Back e Up/Down localization, non permette di trovare facilmente una soluzione (o uno standard) per realizzare contenuti sonori spaziali che siano percepiti chiaramente.

Per lavorare ad algoritmi che migliorino le sensazioni descritte è necessario focalizzarsi sui ritardi di fase percepiti dall'apparato uditivo durante l'esternalizzazione sonora. Prendendo nuovamente come esempio lo spettatore al cinema, in questo caso entrambe le orecchie sentono tutti gli altoparlanti e il cervello tende a localizzare i suoni in base al ritardo interaurale derivante dal primo fronte d'onda in arrivo, identificato come la sorgente sonora protagonista della scena.

Geoff Martin, e i suoi colleghi della McGill University, hanno scoperto che i ritardi di tempo e le differenze di ampiezza richieste per far sembrare che le sorgenti si trovino in determinati luoghi dipendevano da quale coppia di altoparlanti dei cinque (nel sistema surround) era coinvolta (Martin et al., 1999). Tra la coppia di altoparlanti posteriori era richiesto un ritardo di solo circa 0,6 ms affinché un segnale appaia completamente da un lato, circa la metà di quello richiesto per i canali anteriori. Ciò è probabilmente dovuto all'angolo più ampio sotteso dagli altoparlanti posteriori in tale configurazione (circa 120° invece di 60°). Per i segnali ritardati tra gli altoparlanti anteriore sinistro e posteriore sinistro i risultati non sono stati convincenti a causa della difficoltà di localizzare i segnali sulla base di ritardi tra due sorgenti allo stesso lato della testa. La loro conclusione è stata che le differenze di ampiezza tra i canali fornivano informazioni più stabili rispetto alle differenze di tempo.

Un altro punto importante per la progettazione dell'ascolto in cuffia, utilizzando le emulazioni di spazializzazione sonora, è conoscere le corrette informazioni spaziali e il comportamento tipico di un determinato segnale all'interno di una stanza. Nella spazializzazione la giusta componente di rifrazioni e riflessioni, ricreate virtualmente, accentuano la sensazione di immersione e di esternalizzazione.

Come riportato in Rumsey, 2001: “Ad una certa distanza, nota come *distanza critica* o raggio della stanza, le componenti del suono diretto e riflesso sono di livello uguale. La distanza dipende dal livello del suono riflesso nella stanza, che è a sua volta correlato al tempo di riverbero della stanza (tempo impiegato dal riverbero di un suono per decadere di 60 dB al di sotto del livello originale della sorgente). La distanza critica può essere calcolata abbastanza facilmente se si conoscono alcuni dati sulla stanza e sulla sorgente:

$$Distanza\ critica = 0,141\sqrt{RD}$$

dove R è la cosiddetta 'costante della stanza' e D è il fattore di direttività della sorgente (uguale a 1 per una sorgente omnidirezionale). Il fattore di direttività è il rapporto tra il livello di pressione sonora sull'asse di radiazione, anteriore o normale, e il livello che si osserverebbe con una sorgente omnidirezionale. Le riflessioni nel primo periodo di tempo dopo il suono diretto (fino a 50–80 ms) hanno tipicamente l'effetto di ampliare o approfondire gli attributi spaziali di una sorgente.

Nel periodo fino a circa 20 ms possono causare una grave colorazione timbrica se sono a livelli elevati. Dopo 80 ms tendono a contribuire maggiormente al senso di avvolgimento o ampiezza dell'ambiente. David Griesinger ha affermato più volte che le riflessioni comprese tra 50 e 150 ms sono problematiche nella riproduzione di un suono, servendo principalmente a sminuire l'intelligibilità e la chiarezza.

Begault ha esaminato una serie di sfide che sono state affrontate da coloro che hanno tentato di implementare sistemi audio 3D di successo basati su segnali binaurali (Begault, 1991), e questo è un documento utile per coloro che desiderano approfondire la questione. Ha riassunto le principali sfide per i progettisti di sistemi come (a) eliminare le inversioni fronte-retro e il suono intracranico (suoni all'interno della testa); (b) ridurre la quantità di dati necessari per rappresentare le caratteristiche percettivamente più salienti delle misurazioni HRTF; (c) risolvere i conflitti tra la frequenza desiderata e le caratteristiche di risposta di fase e gli HRTF misurati”.

In ultima analisi, dopo aver analizzato il bivio comune tra headphones e loudspeakers, e aver evidenziato i principali problemi di ricerca sui quali porre l'attenzione nel momento in cui si decide di procedere utilizzando le cuffie, ecco quali sono le tipologie di cuffia e le specifiche da considerare:

*Cuffie intra-auricolari o auricolari (in-ear):* tipologia molto leggera e portatile di uso quotidiano, viene inserita all'interno del condotto uditivo per condurre più facilmente il segnale al timpano. Se utilizzate con funzione *noise canceling*, le frequenze intercettate dall'anatomia del padiglione auricolare non ostacoleranno più le frequenze emesse dalla cuffia.

*Cuffie esterne:* sono più grandi e massicce, con la possibilità di avere dei sistemi di emissione più potenti. Vengono divise in *over-ear* quando ricoprono completamente il padiglione auricolare e in *on-ear* quando ci si appoggiano sopra. In questa tipologia vi è un ulteriore sottoinsieme: le cuffie *aperte*, con la parte posteriore del padiglione aperta, che permettono l'ingresso di altre frequenze provenienti dall'esterno; le cuffie *semiaperte*, un intermezzo tra le aperte e le chiuse; le cuffie *chiuse*, che vogliono isolare l'ascoltatore da ciò che accade al di fuori (molto utilizzate per le registrazioni delle batterie o per suonare ai concerti).

Le cuffie possono avere una trasmissione aerea dei suoni, come accade anche per le casse, oppure una *conduzione ossea*, quando viene bypassato l'orecchio esterno, a favore delle ossa craniche, per inviare un segnale all'orecchio interno.

La presenza del cavo come connessione a un dispositivo può influenzare il segnale in fase di ascolto: in particolar modo, se si utilizzano cuffie senza fili la latenza di trasmissione del segnale può pregiudicare la qualità dell'ascolto.

Esistono altre caratteristiche molto importanti da valutare nella scelta di una determinata cuffia:



*Impedenza*: all'aumentare dell'impedenza di un paio di cuffie, è necessaria più tensione (più ampiezza) per poter ascoltare ad un giusto volume. Le cuffie a bassa impedenza sono comprese tra 16 e 32 ohm, mentre le cuffie ad alta impedenza sono circa 100-600 ohm.

Nel caso delle cuffie ad alta impedenza, molto spesso serviranno degli amplificatori per poterle ascoltare, che aumentino la potenza del segnale in entrata. Inserire una cuffia ad altissima impedenza nel jack di un computer potrebbe farci sentire il segnale molto basso in fase di ascolto. Negli studi di registrazione e nell'hifi si utilizzano solitamente cuffie ad alta impedenza, perché sono più fedeli e hanno meno interferenze, mentre gli auricolari che spesso utilizziamo per un uso quotidiano sono a bassa impedenza.

*Sensibilità (elettrica)*: misura dell'efficacia con cui un auricolare converte un segnale elettrico in ingresso in un suono udibile. Indica quindi quanto sono rumorose le cuffie per un determinato livello di azionamento elettrico. Può essere misurato in decibel di livello di pressione sonora per milliwatt (dB<sub>SPL</sub>/mW) o decibel di livello di pressione sonora per volt (dB<sub>SPL</sub>/V). L'abbinamento di cuffie ad alta sensibilità con amplificatori di potenza può produrre volumi pericolosamente alti e danneggiare le cuffie. La sensibilità delle cuffie è generalmente compresa tra circa 80 e 125 dB/mW.

*Risposta in frequenza*: le aziende produttrici più serie e professionali indicano una reale banda di frequenza d'ascolto prodotta dalle proprie cuffie. Considerando che il piccolo altoparlante al loro interno potrà raramente riprodurre delle basse frequenze fedelmente, pensare che tutte le cuffie riproducano esattamente dai 20 Hz ai 20 kHz è un'ingenuità. Dichiarazioni come "risposta in frequenza da 4 Hz a 20 kHz" sono generalmente sopravvalutazioni; la risposta del prodotto a frequenze inferiori a 20 Hz è in genere molto ridotta. È anche vero che, nei casi delle cuffie in-ear, il segnale è così vicino al timpano che basta molta meno potenza per permettere al nostro orecchio di ascoltare bene le basse frequenze.

*Flat vs Hi-Fi*: ogni cuffia ha una sua equalizzazione e ogni orecchio ha il suo HRTF. Questo significa che trovare la giusta accoppiata cuffia-orecchie non è semplice. Prima di scegliere qualsiasi cuffia, bisogna capire quale sia la destinazione d'uso. Le cuffie Hi-Fi sono caratterizzate da un'equalizzazione enfatizzata, sono progettate per stimolare gli audiofili. Le cuffie da studio di registrazione hanno un'equalizzazione flat, per essere il più coerenti possibili nel monitoraggio delle scelte da applicare ad un suono durante una produzione o un mix.

Per convertire i segnali elettrici in suoni, le cuffie utilizzano diverse tipologie di trasduttori:

*A bobina mobile:* trasduttore costituito da un elemento magnetico fissato al telaio delle cuffie che crea un campo magnetico statico. Il magnete è generalmente composto da ferrite o neodimio. La bobina mobile è sospesa nel campo magnetico del magnete, attaccata a un diaframma (in cellulosa, polimero, carbonio, carta o simili) con un elevato rapporto rigidità-massa. Quando la corrente passa attraverso la bobina, si crea un campo magnetico variabile che reagisce contro il campo magnetico statico, esercitando una forza variabile sulla bobina e facendo vibrare la stessa. Con essa vibra anche il diaframma, che spinge l'aria producendo onde sonore.

*Elettrostatici:* questi trasduttori sono composti da un sottile diaframma (di solito una membrana rivestita con pellicola in PET sospesa tra due elettrodi) caricato elettricamente. Il segnale elettrico sonoro viene applicato agli elettrodi creando un campo magnetico; in base alla polarità del campo, il diaframma viene attirato verso le piastre, forzando l'aria attraverso le perforazioni. Combinando l'aria a un segnale elettrico, che guida la membrana, viene generata un'onda sonora. Le cuffie elettrostatiche sono più costose e rare, necessitano di particolari amplificatori per funzionare. La risposta in frequenza può superare i 20 kHz.

*Electret:* un trasduttore Electret funziona con gli stessi mezzi di un trasduttore elettrostatico. Tuttavia, possiede una carica permanente incorporata. Anch'esse sono rare e costose.

*Magnetico:* trasduttore costituito da una membrana discretamente grande che contiene un modello di filo incorporato. Questa rimane sospesa tra due serie di magneti permanenti, allineati in modo opposto. La corrente passa attraverso i fili nella membrana, producendo un campo magnetico che reagisce con il campo dei magneti per indurre il movimento, il quale dà origine al suono.

*Balanced armature:* utilizzato molto negli in-ear dei cantanti per i concerti, questi trasduttori sono costituiti da un'armatura magnetica mobile impernata in modo da potersi muovere nel campo magnetico del magnete permanente. Il termine "bilanciato" deriva dal momento in cui nel campo magnetico non c'è forza netta sull'armatura, perché tutto risulta esattamente centrato. La corrente elettrica attraversa la bobina, magnetizza l'armatura e sposta il diaframma per emettere il suono. La banda di frequenza delle cuffie con questi trasduttori è limitata agli estremi dello spettro uditivo. Queste tecnologie sono state rese popolari negli anni '20 dalle cuffie radio Baldwin Mic Diaphragm e vennero utilizzate nei telefoni militari della Seconda Guerra Mondiale.

*Tecnologie termoacustiche:* l'effetto termoacustico genera suono dalla frequenza audio al Joule riscaldando il conduttore, un sistema completamente diverso che non è magnetico e non fa vibrare altoparlanti. Nel 2013 un gruppo di ricerca dell'Università di Tsinghua ha dimostrato un auricolare a filo sottile di nanotubi di

carbonio basato sul meccanismo termoacustico (Wei, Y., Lin, X., Jiang, K., Liu, P., Li, Q., & Fan, S., 2013).

*Magnetostrizione*: questa tecnologia è strettamente connessa alle *Bonephones*, le cuffie a conduzione ossea già citate, che funzionano vibrando contro un lato delle ossa del cranio per trasmettere le onde all'orecchio interno. Ciò è particolarmente utile per le persone che soffrono di problemi di perdita dell'udito che non interessano la trasmissione dei segnali all'apparato nervoso.

## CAPITOLO 3 - Immersive Music Production

Questo capitolo ha come obiettivo quello di illustrare quali sono i processi e gli strumenti che permettono la produzione di musica immersiva.

Prima di riflettere su cosa significhi manipolare digitalmente un suono e quali siano gli strumenti utili a riprendere o spazializzare una fonte sonora, è necessario andare al fulcro del concetto del termine *immersivo*, quale significato, paradigma e binario di tutte le applicazioni che hanno a che fare con la sensazione di immersione che l'utente prova nel momento in cui vive un'esperienza interattiva. Anche se questo può sembrare scontato, non lo è per niente, dal momento in cui, in senso paradossale, molti editor testuali cercano ancora di correggere il termine italiano "immersività" non riconoscendolo come idoneo al contesto (cosa che non avviene con la declinazione "immersive" inglese).

Conducendo diverse ricerche per trovare la fonte del termine "immersivo" si deduce che esso entri a far parte della nostra società negli anni '80 grazie alla letteratura fantascientifica cyberpunk e a scrittori come Bruce Bethkem, William Gibson, Bruce Sterling, Pat Cadigan. Nei loro scritti essi narravano e ipotizzavano un futuro, a tratti distopico, in cui la tecnologia era protagonista del tessuto sociale. Perché è necessario definire questi umanisti dei pionieri? Poiché erano avanguardisti di una corrente e di una comunità pop-underground che aveva inteso, con netto anticipo, le possibili conseguenze socio-culturali dell'arrivo di una rete che potesse connettere le persone. Questi potrebbero essere definiti come primi esempi di "umanisti digitali", per la loro vicinanza con le competenze di cibernetica e robotica, con i professionisti dell'ingegneria e dell'informatica, e per l'ampia formazione in lettere. Non è da considerarsi un caso che tutta la dialettica inerente al *cyberspazio*, quale un universo digitale in cui le persone si ritrovano immerse, vada pari passo con il tema attuale del *metaverso*. A quarant'anni di distanza sembra quasi che, poiché la visione di alcuni intellettuali si avverasse, essa aveva bisogno soltanto che la tecnologia riuscisse a colmare il gap con la loro proiezione. Eppure, il mondo, le persone comuni, non erano pronte e non sono pronte tutt'oggi, considerando che le tematiche inerenti al metaverso non sono per nulla chiare all'uomo comune.

Dalla nascita del protocollo HTTP nel 1991, grazie a Tim Berners Lee, ci siamo ritrovati immersi in un mondo digitale parallelo, una sovrastruttura le cui fondamenta sono basate attraverso i collegamenti tra i computer di tutto il mondo. Dalle università alle Big Tech, l'obiettivo comune è sempre stato il tentativo di far vivere all'utente una sensazione di immersione sempre più accurata. Se si presta attenzione, anche la bolla del *dot com*, che si è successivamente tramutata nel nostro attuale presente, non è altro che dare la possibilità a un utente di sentirsi immerso in una realtà che accorcia le distanze e permette di poter acquistare beni e servizi tramite un account personale, un'identità digitale. A tal proposito, nella comunità scientifica, e in ambiti di ricerca

come la sociologia della devianza, la psicologia dei gruppi virtuali o la filosofia, si continua a discutere molto sul significato e sulla semiotica dell'identità digitale. Ed il gaming è stato uno dei territori cardine per avvicinare le persone all'immersività tramite un avatar. Videogiochi come *Second Life* (2003), che replicavano la vita all'interno di una comunità umana, ma in un contesto virtuale on-line, hanno molto in comune su quanto viene descritto come "futuro" dalle aziende che stanno investendo sul metaverso (es. Meta, Decentraland, The Sand Box). Seguendo queste stime, in cui si sta già aprendo un orizzonte immobiliare digitale e una rincorsa ad accaparrarsi anticipatamente delle posizioni di spicco, nel momento in cui i visori avranno lo spazio che meritano all'interno della società in cui viviamo, avremo la possibilità di accedere ad uno step successivo nella conoscenza del cyberspazio, in cui verrà meno l'immaginazione, a favore della possibilità di stimolare attivamente i cinque sensi, tramite la partecipazione attiva dell'utente all'interno di una realtà virtuale, che sarà possibile calcare letteralmente.

In tutto questo scenario, qual è il ruolo dell'audio? Il suono ha da sempre accompagnato tutti i media multimediali, dalla radio al gaming. Come è stato descritto nello scorso capitolo, la psicoacustica ha un ruolo centrale nella sensazione di immersione, agendo direttamente sulla percezione cerebrale e stimolando non solo l'udito, ma anche la localizzazione e l'equilibrio. Questo ci dimostra come un videogioco o una simulazione, giocati con un visore, ma percepiti solo con la vista, non potrebbero dare quel senso di realtà che ci si aspetta da un'esperienza immersiva. Perciò le grandi case produttrici di game engine quali Unity, Unreal Engine, o case produttrici di visori come Meta, Google o HTC, hanno investito molto tempo e risorse nello sviluppo di sistemi che integrino la possibilità di manipolare i suoni rendendoli spaziali, immersivi, attraverso la spazializzazione virtuale.

La musica, in tutto questo vortice di cambiamento, è un settore in cui c'è ancora moltissimo margine di crescita e non esistono ancora degli standard definiti inerenti all'ambito immersivo. La stereofonia regna ancora da padrona nel mercato consumer e nella discografia, l'audio spaziale non è un fenomeno popolare, ma potrebbe diventarlo, come vedremo nel paragrafo dedicato alle conseguenze commerciali. È questo settore il vero ago della bilancia, perché la musica arriva a tutti tramite diversi media e su diversi dispositivi. Ciò che sta emergendo dal progresso degli strumenti tecnologici a nostra disposizione è che la stereofonia è ancora una certezza, ma non è più un muro invalicabile. La strada verso la musica immersiva è stata aperta.

### **3. 1 DAW - Digital Audio Workstation**

Le Digital Audio Workstation (DAW) sono sistemi elettronici utilizzati per le registrazioni, l'editing audio, la manipolazione e la riproduzione sonora. Sono software che si appoggiano a schede audio hardware e sfruttano i loro convertitori audio per permettere il flusso di segnale da analogico a digitale (e viceversa). Nate all'inizio degli anni '70, le Digital Audio Workstation hanno dovuto aspettare circa vent'anni per

consolidarsi in questo settore, periodo in cui negli studi di registrazione vi è stata la transizione dalle registrazioni a nastro alle registrazioni digitali.

Le DAW posso essere di due tipi:

*Su computer*: sono software che possono essere acquistati e installati sui propri laptop o desktop computer. Questi programmi richiedono un buon comparto hardware per avere prestazioni professionali. Tra i più utilizzati vi sono Pro Tools (PC/MAC), Ableton Live (PC/MAC), Logic Pro X (Mac), FL Studio (PC/MAC), Steinberg Cubase (PC/MAC) e molti altri. Il concetto alla base dell'interfaccia è di contenere tutto ciò che era presente in uno studio di registrazione, per dare la possibilità di averne uno a portata di mano in qualsiasi momento. Tra i diversi marchi cambiano alcune caratteristiche grafiche e funzionali, ma è tutto ridotto alla gestione di più tracce (detta anche multitraccia), alle quali è possibile aggiungere effetti, processori di dinamica, plugin (nativi o esterni) e VST (Virtual Instruments).

*Integrate*: esistono alcune DAW utilizzate in ambito professionale (ad esempio Pro Tools) che possono essere integrate in mixer digitali di ultima generazione. Queste macchine sono presenti negli studi di registrazione di alto livello, poiché molto costose, a causa della componentistica hardware particolarmente sofisticata.



Figura 46 - Digital Audio Workstation (Landr.com)

### 3.2 Spazializzatori virtuali per la musica immersiva

Nel corso degli anni, le università e i centri di ricerca si sono dedicati allo sviluppo di algoritmi che permettano a suoni monofonici o stereofonici di essere spazializzati. Spazializzare un suono, come è stato già trattato, vuol dire dare ad esso dei connotati che rientrano nei parametri dell'HRTF, caratteristiche successivamente manipolabili attraverso le scelte dell'utente. In sostanza, significa poter dare all'utente la possibilità di posizionare un suono in un ambiente virtuale, che prenderà le caratteristiche psicoacustiche di quella posizione nello spazio.

L'applicazione è molto simile a ciò che avviene in un software di modellazione 3D: all'interno di un ambiente virtuale neutro, definite le grandezze dell'ambiente, si procede ad attribuire valori (positivi o negativi) in gradi all'asse X, in metri all'asse Y e in gradi all'asse Z. Tali valori daranno una posizione, che sarà la posizione di quel suono nell'ambiente.

Immaginiamo di avere un suono che abbia come posizione le seguenti coordinate polari:

Azimuth: + 30°

Elevation: + 50°

Distance: ?

Senza la definizione di una grandezza metrica sarà difficile capire a che distanza il suono verrà riprodotto. Motivo per il quale la distance è il primo valore da dover attribuire. Inoltre, se non viene definita anticipatamente la grandezza ideale dell'ambiente in cui verrà inserito il suono sarà difficile restituire un attributo realistico al suono (specialmente se all'interno di una stanza, perché allora sarà necessario adottare plugin che emulano il comportamento delle rifrazioni e delle riflessioni).

In questo capitolo verranno presi in esame tutti i tool che possono essere utilizzati per la produzione musicale immersiva, lasciando non elencati tutti quegli strumenti, come i game engine, utili ad altri tipi di sviluppo. Un interessante confronto tra differenti plugin di spazializzazione audio, non inerenti al mondo musicale, che evidenzia alcune delle fondamenta da considerare per valutare questi strumenti è l'articolo *Let's Test: 3D Audio Spatialization Plugins* del sito Designing Sound (Gould, 2018).

### 3.2.1 Formati

Posti come acquisiti i concetti del capitolo precedente, è necessario fare un approfondimento su alcuni formati e su alcune sigle che si possono incontrare nell'utilizzo dei software di spazializzazione.

Per quanto riguarda il comparto Ambisonics, vi sono diverse interpretazioni degli ordini in base ai formati che si scelgono. L'azienda Audiokinetic, all'interno delle proprie risorse didattiche presenti sul web<sup>10</sup>, mette a disposizione una spiegazione semplice e completa dei formati FuMa e ACN: "I due formati di ordinamento dei componenti principali per Ambisonics sono Furse-Malham, comunemente chiamato FuMa, e Ambisonics Channel Number, comunemente chiamato ACN. Come si vede nell'immagine seguente, il primo utilizza una notazione con lettere che - seguendo l'ordine alfabetico per raggruppamento - inizia con il canale W (omni), si sposta in basso a destra (X), quindi in basso a sinistra (Y) e quindi al centro in basso (Z); poi passa all'ordine successivo e parte dalla R, si sposta alla sua destra (S), poi alla sua sinistra (T), poi ancora più a destra (U), poi ancora più a sinistra(V); quindi passa

---

<sup>10</sup> [https://www.audiokinetic.com/library/edge/?source=Help&id=ambisonics\\_channel\\_ordering](https://www.audiokinetic.com/library/edge/?source=Help&id=ambisonics_channel_ordering)

all'ordine successivo e segue uno schema simile. D'altra parte, nell'ordine ACN, lo schema è numerato in un ordine da sinistra verso destra, molto più facile da seguire”.

Dal blocco superiore mono alle basi del 1°, 2° e 3° ordine, l'immagine illustra anche che ogni ordine Ambisonics può essere visto come una piramide con rispettivamente 4, 9 e 16 canali.”

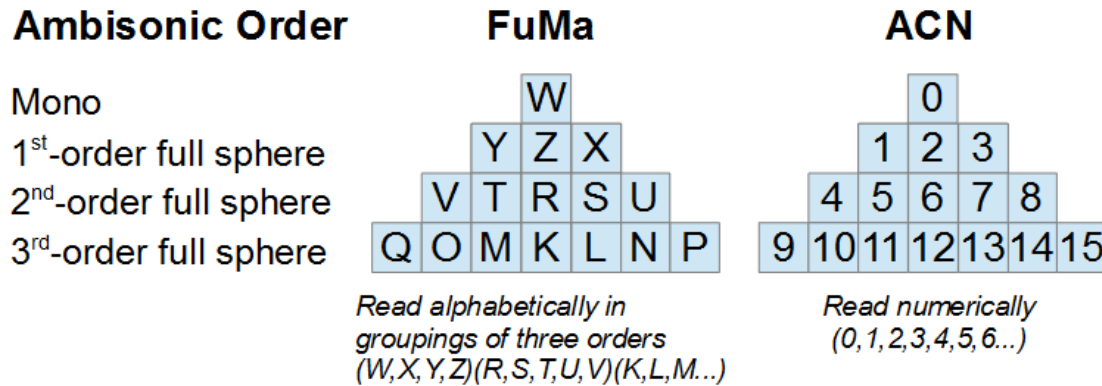


Figura 47 - Formati FuMa e ACN (Audiokinetic.com)

Molti plugin e DAW preferiscono lavorare in un altro formato analogo chiamato AmbiX, una specifica flessibile per le registrazioni Ambisonics che evita i problemi di incompatibilità.

L'argomento è estremamente più ampio e tocca informazioni dettagliate e ingegneristiche. Si consiglia la lettura dell'articolo *AmbiX - A Suggested Ambisonics Format* di Nachbar, Zotter, Deleflie e Sontacchi (2001) per approfondire meglio l'argomento.

### 3.2.2 Monitoring

Sebbene il lavoro di chi si occupa di musica immersiva (o di audio spaziale) sia legato al mondo dell'ascolto, è altrettanto importante avere degli strumenti di monitoring che prevedano la visualizzazione del posizionamento nello spazio. Nel corso di questo paragrafo verranno descritti diversi software, molti con caratteristiche simili, ma sicuramente con alcune differenze, che rendono importante la loro scelta, soprattutto per la fase di progettazione e realizzazione. Lavorare con uno strumento che non permetta di visualizzare (tramite uno *scope*) l'effettivo posizionamento delle frequenze in uno spazio a 360° è da considerarsi limitante, soprattutto se non di facile consultazione o con un'interfaccia complessa. È vero che l'immersività sonora è una sensazione che va progettata basandosi in primo luogo sull'apparato uditivo, ma considerando il gap attuale tra i vari algoritmi di spazializzazione presenti nel settore, il monitoraggio grafico delle frequenze posizionate nello spazio fa la differenza.



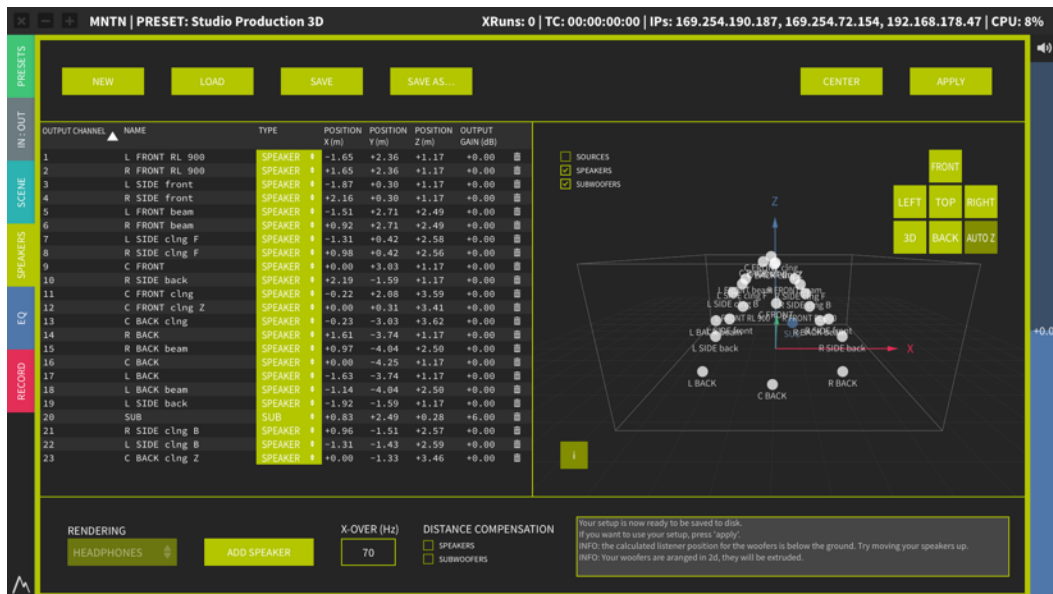


Figura 48 - Monitoring visivo dell'audio nello spazio (mntn.rocks)

Un secondo aspetto da citare è la possibilità, in alcuni strumenti hardware e software, di ascoltare musica utilizzando l'*head tracking*, e quindi facendo percepire uno spostamento interattivo tramite i movimenti della testa dell'ascoltatore. Questo tipo di scelta non viene approfondito, poiché lo si reputa poco produttivo in termini di musica immersiva, rispetto ad esempio al gaming dove questo aspetto è fondamentale. L'ascolto in cuffia della musica, nel quotidiano, è qualcosa che si fa molto spesso in movimento e mentre si stanno svolgendo altre azioni. Immaginare di far provare all'ascoltare un cambiamento del posizionamento d'ascolto mentre si sta camminando per la strada è da ritenere confusionario e probabilmente pericoloso. Diversa è la situazione circoscritta in cui l'ascoltatore, magari inserito in un contesto museale, è spinto a concentrarsi sui movimenti della propria testa e andare a ricercare volontariamente una determinata sensazione. Questo, però, non è il focus principale della tesi, motivo per il quale non vi è un approfondimento sull'utilizzo dell'*head tracking*.

### 3.2.3 HRTF Database Sets

Alcuni software mettono a disposizione la scelta di set di HRTF raccolti dalle università e dalle aziende. Questo nel tentativo, più che giusto, di progettare un'esperienza che sia il più compatibile possibile con le diverse tipologie di orecchio. In questo campo è stato identificato, nel 2015, un formato standard chiamato Spatially Oriented Format for Acoustics (SOFA)<sup>11</sup> per la memorizzazione di dati acustici orientati nello spazio, come le funzioni di trasferimento relative alla testa (HRTF) e le risposte all'impulso binaurale o spaziale (BRIR, SRIR). SOFA è stato standardizzato dall'Audio Engineering Society

<sup>11</sup>[https://www.sofaconventions.org/mediawiki/index.php/SOFA\\_\(Spatially\\_Oriented\\_Format\\_for\\_Acoustics\)](https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics))

(AES) come AES69-2015. Uno dei database più vasti è quello sviluppato dall'IRCAM di Parigi, uno degli istituti di ricerca più importanti del settore.

### 3.2.4 Software e virtual spatializers

Come descritto a inizio capitolo, uno dei principali focus della tesi è quello di evidenziare gli strumenti utili alla produzione di musica immersiva in formato binaurale. Negli ultimi tre anni vi è stata una crescita sostanziale dei tools prodotti dalle università e dalle aziende interessate a questo mercato. Di seguito sono state elencate, e descritte, tutte le risorse disponibili in ordine alfabetico. Un naturale proseguimento dell'articolo scientifico *Tools for Immersive Music in binaural format* (De Sotgiu, Vercelli & Coccoli, 2020):

**Ableton - Live  
(DAW - a  
pagamento)**

Questa DAW è una delle più utilizzate per la produzione musicale. Originariamente dedicata alla musica elettronica, oggi è a tutti gli effetti uno dei software più affidabili. Al momento non ha la possibilità di gestire l'audio multicanale, ma sfruttando il routing e la sua compatibilità con *Max MSP* e i sistemi *Max4Live*, questo diventa uno strumento potentissimo con cui lavorare in audio spaziale in formato binaurale.

**Apple - Logic  
Pro X  
(DAW - a  
pagamento)**

Logic Pro X è la DAW ammiraglia di Apple. Nata nei primi anni '90, il suo sviluppo è stato tale da diventare oggi uno strumento utilizzato in molti studi professionali in tutte le fasi della produzione musicale. Negli ultimi tre anni Apple ha dimostrato un grande interesse per tutti i contenuti Dolby Atmos, rendendoli fruibili su Apple Music. Nel 2021 è stato rilasciato un importante aggiornamento di Logic Pro X, che ha introdotto la possibilità di creare e gestire sessioni di produzione musicale in Dolby Atmos, fornendo i template della canzone *MONTERO (Call Me By Your Name)* di Lil Nas X. Il plugin Dolby Production Suite, ormai nativo su Logic Pro, verrà descritto nel corso di questo paragrafo e del prossimo capitolo.

**Artysin - RL360  
(Software -  
Licenza free per  
7 giorni)**

Questo è un plugin molto semplice che contiene gli elementi di base per il posizionamento di un suono in un ambiente virtuale. Interessante la possibilità di scegliere tra differenti *room presets* che definiscono l'emulazione della grandezza di una stanza. Può essere utilizzato per le produzioni musicali, ma anche per la realizzazione di podcast e video.

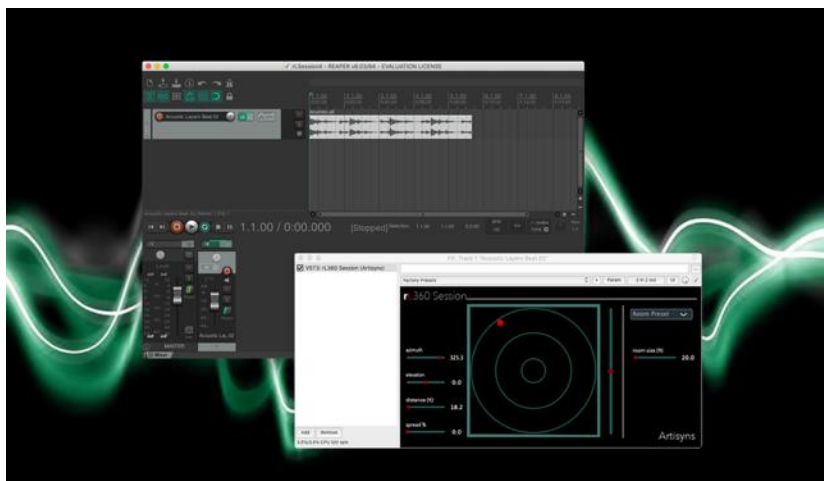


Figura 49 - Artysin - RL360

**Auro Technologies (Piattaforme software a pagamento)**

Auro 3D / Auro Technologies è una realtà nata nel 2005 che ha fatto dell'esperienza immersiva il suo cavallo di battaglia, tanto da diventare oggi una presenza attiva su diversi settori: cinema, gaming, AV System, automobili e devices. Tramite i suoi servizi e le sue suite è diventata una delle principali aziende concorrenti in questo settore.

**Avid Pro Tools (DAW - a pagamento)**

Il software che è stato l'industry standard per alcune decadi. Ancora oggi è una delle DAW più utilizzate negli studi di registrazione. È stata tra le prime ad avere un contratto con i Dolby Laboratories per associare la Dolby Production Suite al suo interno e poter lavorare in Dolby Atmos.

**binaural\_spat (kasper fangel skov) (Pacchetto di software in free download)**

Set di software sviluppati con Max MSP e utilizzabili sia con Max che con Ableton Live, per dare possibilità di spazializzazione lavorando con sistemi multicanali. Sul sito, è possibile scaricare i codici sorgenti su GitHub.

Prendendo come DAW di riferimento Ableton live, per utilizzare il software è necessario creare una nuova traccia audio o midi e aggiungere il plugin *binaural\_spat source*. Quando si carica un dispositivo sorgente, viene chiesto di assegnare al dispositivo un numero (compreso tra 1 e 16). Assicurarsi di non caricare due dispositivi sorgente su una traccia e di fornire sempre i numeri nell'ordine corretto.



Figura 50 - *binaural\_spat source*

Come raffigurato nell'immagine, oltre ai parametri inerenti alle coordinate sono presenti altri parametri di mapping legati alla distanza e alla velocità. Il dispositivo sorgente prende un ingresso stereo e lo converte in mono, non appena il plugin è stato applicato sulla traccia. Il segnale mono viene automaticamente indirizzato all'ingresso corrispondente nel dispositivo di ascolto. È possibile aggiungere fino a 16 dispositivi sorgente (sebbene 16 non siano consigliati a causa dell'elevato carico della CPU).

Per quanto riguarda la traccia utile al monitoring: creare una nuova traccia audio in Ableton e aggiungere il dispositivo *binaural\_spat listener*. Questa traccia funzionerà da ricevitore dell'audio proveniente da *binaural\_spat source* per un massimo di 16 dispositivi. Da qui sarà possibile anche monitorare l'input dei dispositivi sorgente.

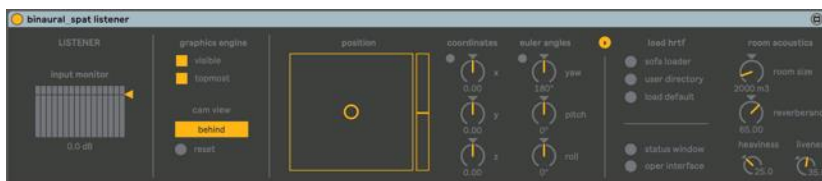


Figura 51 - *binaural\_spat listener*

Come impostazione predefinita, il plugin listener carica un database HRTF predefinito (KEMAR-HRTF). Facendo clic su "sofa loader" verrà visualizzata l'interfaccia Sofa Loader inclusa dalla libreria Spat5. Qui è possibile scegliere e scaricare i file HRTF dai database esistenti per ottenere un risultato migliore.

Tra le possibilità date da questo pacchetto di software non manca l'head-tracking, disponibile utilizzando il plugin *gyrosc2live*.



Figura 52 - gyros2live

Per utilizzarlo è necessario abilitare la comunicazione OSC (Open Sound Control) tra l'app iOS GyrOSC e il dispositivo di ascolto. Il dispositivo gyros2live riceve i dati del giroscopio dall'app (preferibilmente installata su un iPhone) e semplifica la mappatura degli angoli di Eulero (beccheggio, imbardata, rollio) ai parametri "beccheggio", "imbardata" e "rollio" nel plugin listener.

Quando l'app e il dispositivo sono configurati, posizionare l'iPhone sopra la testa e fare clic su "riorienta". Muovendo la testa si dovrebbe essere in grado di vedere/ascoltare il risultato di conseguenza. Per impostare sistemi di rilevamento della testa più complessi con posizione e orientamento della testa, creare un dispositivo Max for Live simile, mappare sia i dati in posizione (coordinate X, Y, Z) che gli angoli di Eulero all'interno del dispositivo di ascolto.

**Cycling '74 -  
Max MSP  
(Software a  
pagamento)**

Max è un software che unisce coding e produzione musicale in un unico ambiente di sviluppo grafico per la musica e la multimedialità. Utilizzato dai creativi del settore da più di quindici anni, è alla base di alcuni dei plugin più interessanti (ad. es. Envelop) grazie alla partnership con Ableton e all'estensione Max4Live, che permette di creare nuovi applicativi. È molto utilizzato anche nel sound design e nel sonic interaction design.

**Dear Reality -  
Dear VR PRO /  
Dear VR AMBI  
Micro  
(Plugin a  
pagamento)**

La software house Dear Reality produce diverse soluzioni inerenti al mondo dell'immersività. Il plugin Dear VR Pro è uno strumento estremamente utile anche nel contesto musicale. È uno spazializzatore con un algoritmo preciso, per questo è spesso utilizzato nelle convention per la realizzazione di test e demo. Presenta tutte le caratteristiche di base per il posizionamento sfruttando un sistema di coordinate. Molto importante la sezione Reverb, che permette di scegliere diversi algoritmi di acustica virtualizzata (es. riverbero di una cattedrale), manipolabile tramite il knob *damping* e i valori di *size*. Da sottolineare anche la sezione Reflections, con la funzione *real-time auralization*. Gestire le riflessioni di una stanza

virtuale permette di dare dei connotati estremamente reali a un suono, come è già stato descritto nel capitolo 2. La sezione *Output Format* permette di scegliere in che modo renderizzare il segnale audio che viene spazializzato (binaural, Ambisonics, surround, 7.1.4 e molti altri formati). È possibile sfruttare l'head tracking tramite l'applicazione Spatial Connect.



Figura 53 - Dear VR Pro

Lavorando in formato Ambisonics all'interno di una DAW che permetta l'audio multicanale (citato all'interno di questo paragrafo), sarà possibile sfruttare il plugin *Dear VR AMBI MICRO* per monitorare e renderizzare in formato binaurale la sessione spazializzata con il plugin Dear VR Pro.



*Figura 54 - Dear VR Ambi Micro*

Importando il plugin sul Master Channel e selezionando come Input lo stesso formato Ambisonics della sorgente, sarà possibile renderizzarlo in Output scegliendo tra i formati disponibili (tra i quali il formato binaural a due canali).

**Dolby  
Laboratories -  
Dolby Atmos  
Production  
Suite  
(Piattaforma  
fornita con  
alcune DAW)**

Dolby Atmos Production Suite è un insieme di software prodotto dai laboratori Dolby che permette di lavorare in Dolby Atmos e renderizzare il risultato finale direttamente dalla propria DAW. Le Digital Audio Workstation che, al momento, hanno a disposizione questa possibilità sono Pro Tools (Avid), Nuendo (Steinberg) e Logic Pro X (Apple). Questo software è stato scelto per l'esperimento condotto nel corso di questa tesi e verrà descritto ampiamente nel prossimo capitolo.

**Envelope - E4L  
(Bundle di  
plugin in free  
download)**

Envelop for Live (E4L) è una raccolta di plugin di produzione audio immersiva open source compatibili solo con Ableton Live 10/11 Suite. E4L consente ad artisti e designer di creare mix spaziali per gli eventi Envelop o per qualsiasi ambiente audio multicanale, nonché applicazioni VR/AR basate su cuffie. Come dichiarano gli sviluppatori di questo bundle di software: "Crediamo che il suono immersivo sia il futuro della produzione audio. L'ascolto dall'interno della musica amplifica l'impatto emotivo, creativo e collettivo del suono".



*Figura 55 - Envelop applicato su Ableton Live 11*

Utilizzare questi software è molto semplice. Essi sfruttano il formato Ambisonics per codificare la spazializzazione sonora di un suono mono o stereo e decodificarla in binaurale (o in altri formati). Accedendo al sito ufficiale, dopo aver fornito tutti i dati necessari alla profilazione per il free download, importare la cartella scaricata all'interno della sezione "Places" di Ableton.

Creare una nuova traccia audio.

→ Aggiungere il plugin E4L Master Bus ad essa.

Creare un nuovo strumento MIDI o una traccia audio che sia una sorgente sonora. Notare come l'audio venga codificato immediatamente in formato Ambisonics e venga automaticamente instradato attraverso il Master Bus E4L.

→ Aggiungere il plugin E4L Mono Panner / E4L Stereo Panner alla traccia.

→ Utilizzare i controlli presenti per posizionare il suono nello spazio 3D.

All'interno del pacchetto fornito in free download sono presenti, al momento, i seguenti applicativi, tra i quali strumenti utili al monitoring visuale e alcuni estremamente creativi come riverberi e delay in formato Ambisonics:

*E4L Auraverb*

*E4L B-Format Convolution Reverb*

*E4L B-Format Sampler*



*E4L Brownian Delay*  
*E4L Delay Boids*  
*E4L HOA Scope*  
*E4L Transform*  
*E4L LED MIDI Realy*  
*E4L LED Parameter Control*  
*E4L Master Bus*  
*E4L Meter*  
*E4L Mono Panner*  
*E4L Multi-Delay*  
*E4L Quad Panner*  
*E4L Source Panner*  
*E4L Spatial Slicer*  
*E4L Spinner*  
*E4L Stereo Panner*

**Facebook Audio  
360 - Spatial  
Workstation  
(Bundle di  
plugin in free  
download)**

Plugin gratuito nato come supporto all'audio nella cinematic VR, ma utilizzato anche nella realizzazione di progetti in audio spaziale (De Sotgiu, Coccoli, Vercelli, 2017). All'interno del sito ufficiale ci sono diversi tutorial che spiegano come integrare questo software con piattaforme multicanali come Reaper o Avid Pro Tools. Al giorno d'oggi, considerando i concorrenti, l'utilizzo di questo strumento nel settore musicale può presentare diverse limitazioni: può essere una soluzione per compiere alcuni esperimenti, ma è consigliabile utilizzarlo per ciò che è stato progettato, ovvero la gestione dell'audio nella Cinematic VR.

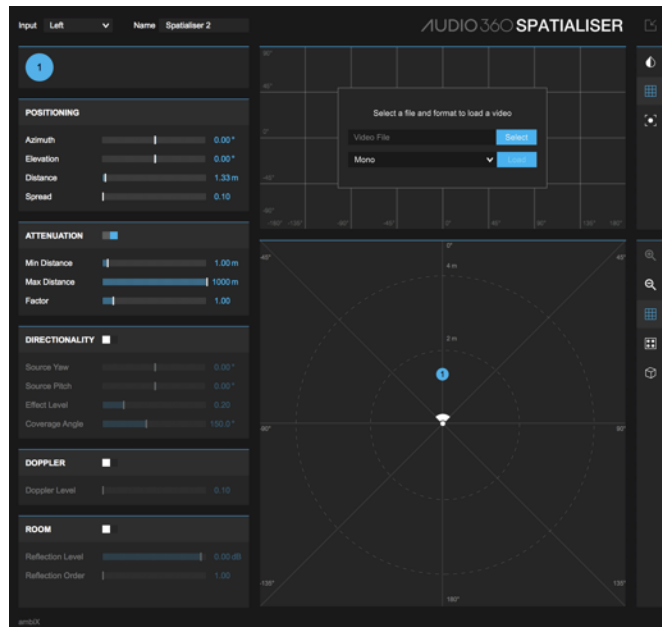


Figura 56 - Facebook Audio 360 Spatial Workstation

**Flux Audio -  
Spat Revolution  
(Piattaforma  
software a  
pagamento)**

Il grande lavoro svolto dall'istituto di ricerca francese IRCAM ha portato al risultato commerciale dell'azienda Flux Audio, con la piattaforma Spat Revolution, software completamente dedicato all'audio immersivo, con un ampio riferimento al reparto musicale. La grande qualità di questo software è l'importante lavoro grafico per il monitoring tramite visualizzazione e l'attenzione posta sull'algoritmo di renderizzazione binaurale.

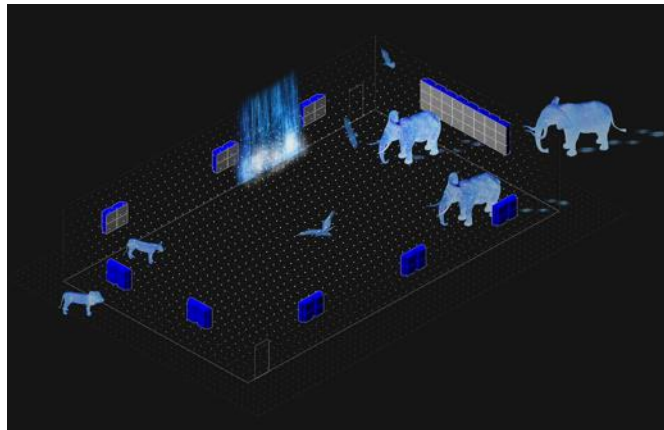
Spat Revolution viene anche utilizzato per le nuove tecniche di progettazione di riproduzione sonora legate alla Wave Field Synthesis.



*Figura 57 - Spat Revolution*

**Holopot  
(Sistema audio  
software e  
hardware a  
pagamento)**

La soluzione proposta da questa azienda berlinese è basata sul dialogo tra la suite software per la progettazione della diffusione spaziale da loro sviluppata e l'hardware Holopot Matrix Array, un sistema di diffusori. La mission è diretta all'ascolto immersivo in uno spazio ridotto come descritto sul sito ufficiale: "e se il suono potesse davvero essere controllato nel modo in cui possiamo controllare la luce? E se potessimo creare paesaggi sonori che sembrano più reali della realtà? E se ogni posto in un locale potesse diventare il posto migliore?"



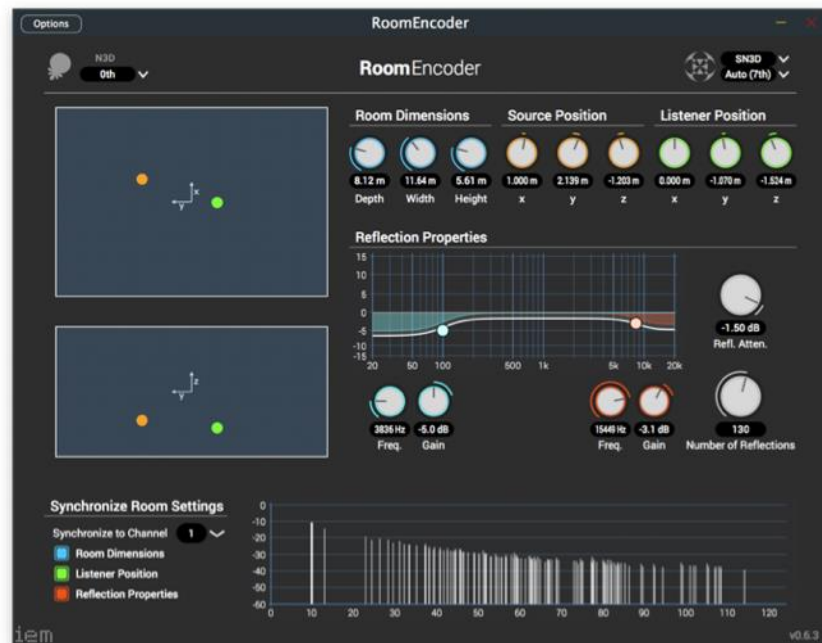
*Figura 58 - Sistema di diffusori Holopot*

**IEM - Plugin  
Suite  
(Bundle di  
plugin in free  
download)**

L'Istituto di Musica Elettronica e Acustica (IEM) di Graz, fondato nel 1965, intreccia il progresso delle arti con la ricerca e lo sviluppo sperimentale. L'IEM funge da interfaccia pionieristica tra scienza e arte, tra nuove tecnologie e pratica musicale. Tra i primi ad occuparsi di audio spaziale, hanno sviluppato una suite di software per la gestione e la manipolazione dell'audio spaziale multicanale e quindi compatibile con tutte quelle DAW che permettono questo tipo di routing (ormai sempre più frequente). Tra i software a disposizione vi sono da sottolineare compressori ed equalizzatori multicanale:

*AllRADecoder*  
*BinauralDecoder*  
*CoordinateConverter*  
*DirectionalCompressor*

*DirectivityShaper*  
*DistanceCompensator*  
*DualDelay1*  
*EnergyVisualizer*  
*FdnReverb*  
*MatrixMultiplier*  
*MultiBandCompressor*  
*MultiEncoder*  
*MultiEQ*  
*OmniCompressor*  
*ProbeDecoder*  
*RoomEncoder*  
*SceneRotator*  
*SimpleDecoder*  
*StereoEncoder*  
*ToolBox*



*Figura 59 - IEM Plug-in Suite - Room Encoder*

**IRCAM - Tools**  
**(Software a**  
**pagamento)**

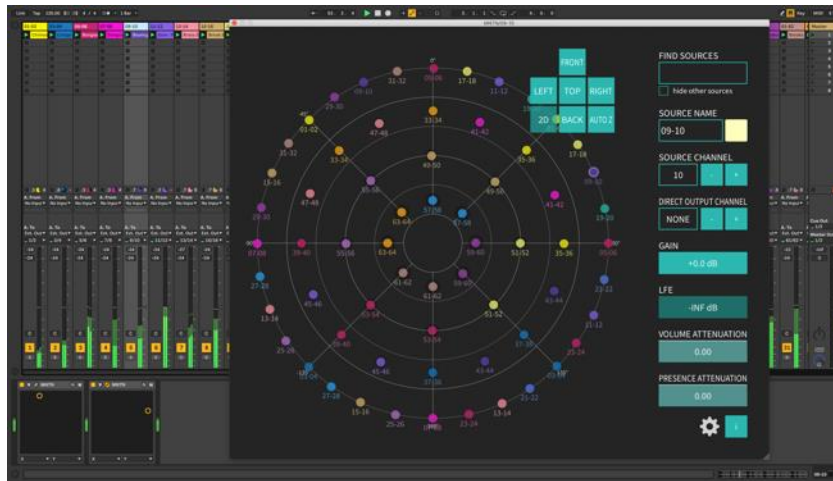
Nel corso di questa tesi sono già stati citati più volte i laboratori francesi dell'IRCAM. Sul sito dell'istituto di ricerca è possibile scaricare software e documentazione utile in merito agli strumenti da loro sviluppati riguardo l'audio immersivo. Tra questi riverberi e renderizzatori spaziali.

**MNTN**

MNTN - The sound of Mountain è un'azienda che ha sviluppato una production suite molto flessibile, anche con

**(Piattaforma software a pagamento)**

le DAW che non hanno la possibilità di gestire l'audio multicanale. L'aggiornamento e la crescita dei loro applicativi, nel corso degli anni, sono stati molto veloci. Oggi vantano la possibilità di gestire la renderizzazione del prodotto finale per ogni tipologia di supporto: dall'audio binaurale al Dolby Atmos (7.1.4), l'audio object based e l'Auro 3D. La suite ha un ottimo sistema di monitoring tramite visualizzazione.



*Figura 60 - Sistema MNTN*

**Reaper  
(DAW a pagamento, con licenza gratuita per 60 giorni)**

Digital Audio Workstation con possibilità di gestione di audio multicanale. Molto utilizzata per la sua licenza gratuita con tempistiche flessibili e quindi ottima per fare esperimenti. Esiste anche una nicchia di produttori musicali che è estremamente fidelizzato a questo software.

**RØDE  
(Software in free download)**

Rode SoundField, oltre ad essere un microfono Ambisonics, è anche un software ad esso associato per la manipolazione e la renderizzazione della ripresa B-Format. Può essere anche utilizzato per il posizionamento dei suoni nello spazio virtuale e la renderizzazione finale in mastering.

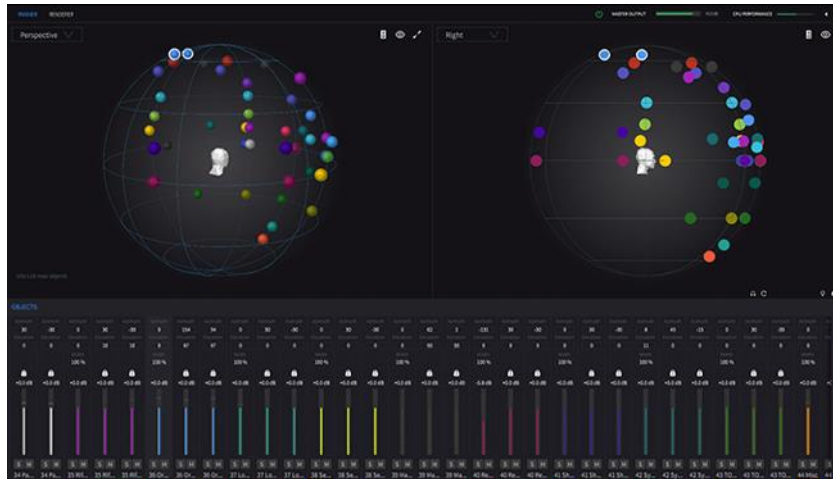
**Sennheiser  
Ambeo  
(Software in free download)**

Allo stesso modo, come è stato descritto per il precedente microfono Ambisonics, anche Sennheiser ha sviluppato sia software che hardware (Ambeo VR Mic) per la gestione della ripresa Ambisonics B-Format.

**Sony - 360  
Reality Sound  
(Piattaforma software a pagamento)**

Formato di object audio prodotto da Sony e competitor di Dolby Atmos e di Auro 3D. Il 360 Reality Audio nasce aperto, sia dal punto di vista delle codifiche (utilizza di fatto l'MPEG-H 3D Audio, utilizzato anche da Fraunhofer IIS) che dell'impacchettamento dei dati: sin da subito, eventuali

terze parti possono avere accesso a tutti i dati e i tool relativi allo standard proposto<sup>12</sup>. Ovviamente la potenza commerciale del brand ha permesso a questo standard di essere utilizzato anche tra grandi distributori di servizi mediatici, come streaming services e multimedia agency.



*Figura 61 - Sony 360 Reality Sound*

**Sound Particles  
(Piattaforma  
software a  
pagamento)**

Sound Particles è un software di immersive audio in grado di generare migliaia (anche milioni) di suoni in un mondo audio 3D virtuale. Viene utilizzato in particolar modo nella post-produzione audio cinematografica. Tramite i suoi algoritmi può creare delle sensazioni sonore estremamente avvolgenti. Al suo interno sono presenti diversi preset, applicabili ai singoli suoni, che permettono di creare un'esperienza spaziale autentica.

---

<sup>12</sup> <https://www.dday.it/redazione/29234/sony-360-reality-audio-come-funziona>



*Figura 62 - Sound Particles*

Essendo estremamente specifico sulla tridimensionalità sonora, è più difficile applicarlo concretamente in ambito musicale, ma resta uno strumento dalle potenzialità incredibili.

**New Sound  
Technology -  
Spatial Audio  
Designer  
(Software a  
pagamento)**

Lo Spatial Audio Designer (SAD) è uno strumento leader nel mercato per la creazione di contenuti e il monitoraggio in surround o in audio 3D. Indipendentemente dalle strutture dei bus delle DAW (gestione multicanale), SAD consente il monitoraggio e il missaggio di tutti i formati audio attuali e futuri (Stereo, 5.1, 7.1, Dolby Atmos 7.1.2 Bed, 11.1 e 22.2). Se non è disponibile un sistema di altoparlanti appropriato, SAD fornisce la virtualizzazione degli altoparlanti con l'ascolto binaurale 3D Headphone Surround 3D, flessibile e regolabile per l'utilizzo di cuffie standard.

**Waves  
(Software a  
pagamento)**

Basato sull'innovativa tecnologia Nx di Waves, Nx Virtual Mix Room è un plug-in di monitoraggio virtuale che offre, in cuffia, la profondità tridimensionale e l'immagine stereo panoramica che si dovrebbe ascoltare dagli altoparlanti in una stanza trattata acusticamente.

Nx include anche i componenti Nx Ambisonics, che consentono di monitorare l'audio in Ambisonics B-Format per progetti in audio a 360° e VR utilizzando le normali cuffie stereo. Per una nitidezza visiva ottimale, Nx Ambisonics è dotato di un misuratore spaziale che

rappresenta il contenuto della frequenza delle tue tracce in ogni direzione del campo sonoro sferico a 360°.

Infine, Nx include una funzione di calibrazione dell'equalizzazione delle cuffie, che consente di selezionare una curva di correzione dell'equalizzazione per modelli di cuffie specifici. Sulla base dei dati di misurazione di precisione delle cuffie forniti da [www.headphone.com](http://www.headphone.com), le curve Nx EQ sono progettate per bilanciare eventuali caratteristiche estreme nella risposta in frequenza, correggendole verso un bilanciamento di frequenza comune e fornendo un punto di partenza più trasparente per il monitoraggio e il messaggio.

**Audiokinetic -  
Wwise  
(Software a  
pagamento)**

Vero e proprio audio engine, si trova a metà strada tra una Digital Audio Workstation e i software di produzione audio che si trovano nei game engine. Questa è la sua forza, il fatto di essere potentissimo. Ha al suo interno software nativi utili per differenti scopi: dalla sintesi, all'audio spaziale, dalla musica interattiva al mixing dinamico. Software complesso e tecnico, dal carattere ingegneristico, ma al tempo stesso, sul suo sito, sono presenti diversi tutorial per imparare ad utilizzarlo.

### **3.3. Le fasi della produzione nella musica immersiva**

Ciò che determina l'inizio e la fine di una produzione musicale sono le diverse fasi che devono essere affrontate: pre-produzione, produzione, mixing e mastering. Nel corso degli anni intercorsi a studiare le tecnologie in ambito audio spaziale, ho fatto diversi paragoni tra il processo di produzione standard e il workflow utile a produrre audio immersivo. In questo paragrafo verrà definito, quindi, come introdurre elementi immersivi nelle fasi della produzione musicale.

#### **3.3.1 Pre-produzione**

Come accade per ogni progetto di carattere tecnologico o umanistico, anche nell'ambito musicale è molto importante studiare il contesto di riferimento prima di svolgere qualsiasi step. Un primo passo sarà, quindi, procedere con un ampio *brainstorming* che possa comprendere differenti ascolti, per individuare le references che saranno utili al progetto. Verranno esaminati i generi musicali e i sottogeneri, per poi andare a definire il range di BPM (Battiti Per Minuto) che avrà la produzione, un punto centrale a livello di psicologia della musica (come descritto nel secondo capitolo di questo saggio).



Fare musica significa comunicare attraverso le note e comunicare significa trasmettere un messaggio, perciò è molto importante chiedersi: quale messaggio voglio comunicare? E con quale mezzo?

Sappiamo che sfrutteremo armonie, melodie e arrangiamenti per arrivare al risultato finale, ma quali saranno i canali primari che verranno utilizzati per fruire della nostra produzione? Questo è molto importante, perché identificare se il nostro ascoltatore tipo utilizzerà maggiormente le casse, le cuffie, l'ascolto in macchina o l'ascolto al cinema influirà definitivamente sul nostro processo creativo.

Tutto ciò riconduce alle tecniche utilizzate nella User eXperience per l'individuazione dei modelli di target: creare *personas*, ovvero modelli di utenti che possiedano quante più caratteristiche possibili di una persona comune, ci aiuterà a familiarizzare con l'audience che avremo come obiettivo.

Raggiunti questi step avremo chiare le ambizioni del progetto. Solo a questo punto potremo definire quali saranno gli strumenti da scegliere e da impiegare nella fase della produzione.

Ad esempio, si potrebbe decidere di scegliere un determinato studio dove andare a registrare le tracce, perché fornito di tutte quelle attrezzature che possono valorizzare il nostro lavoro.

In questa fase, optando di scegliere di produrre musica immersiva, sarà necessario definire l'impatto ideale che vuole avere l'immersività nella nostra produzione. Il senso di immersione nella musica non può essere una costante, esattamente come nei film in 3D: le scene di azione, avendo molti movimenti, non possono spingere troppo sulla stereoscopia, perché rischiano di creare forti mal di testa agli spettatori. Nella musica il concetto è differente. Per non creare confusione all'ascoltatore e fargli percepire davvero la spazialità, è necessario farlo concentrare su alcuni movimenti precisi, che gli facciano capire cosa sta accadendo in una determinata sezione. Ma tutto ciò verrà preso meglio in analisi nel capitolo del mix.

Rimanendo ancorati, invece, sulla fase della pre-produzione, la scelta degli strumenti tecnici tratterà la strada e il destino del nostro progetto. Decidere di optare per registrazioni binaurali con la tecnica microfonica delle *dummy head* (es. Neumann KU100) può avere un buon impatto sul progetto. Come è possibile vedere nella prossima figura, le registrazioni binaurali prevedono una testa simile a un manichino, con tanto di padiglione auricolare, e due microfoni posizionati nei pressi dei timpani. Si può dire che ogni dummy head possiede un HRTF personalizzato.



Figura 63 - Neumann KU-100

Di contro il risultato della registrazione sarà una traccia a due canali (L/R) poco manipolabile in post-produzione. Sarà necessario, quindi, allestire uno scenario musicale in uno spazio che rimarrà tale anche nella registrazione. Ad esempio, immaginando di porre una o più dummy head all'interno di una stanza, o di uno spazio all'aperto, gli strumentisti potrebbero essere posizionati tutti intorno al microfono, lasciando il cantante nella parte anteriore, come stesse cantando di fronte all'ascoltatore.

Sempre riguardo la microfonazione, al posto della tecnica appena descritta si potrebbe procedere nell'utilizzo di microfoni Ambisonics B-Format di differenti ordini in base alla quantità di canali necessari. Ad esempio, l'azienda mh acoustics, come mostrato nella Figura 38 del capitolo 2, ha brevettato l'array microfonico em32 Eigenmike®, che lavora in Ambisonics 3rd order HOA a 32 canali. In questo caso la possibilità di manipolazione sonora in post-produzione potrebbe essere molto più viva, anche tramite l'utilizzo di software di decoding e di spazializzazione che agevolino la scelta della posizione delle tracce registrate all'interno dello spazio virtuale.

Come già descritto nel capitolo precedente, è possibile anche scegliere di non utilizzare microfonazioni multicanale e procedere con normali tecniche di registrazione monofoniche/stereofoniche, per lasciare il processo di spazializzazione virtuale agli algoritmi dei software. Questi sono comunque utili in quella che è l'ultima fase della pre-produzione, che è la definizione di un prototipo d'ascolto (o demo), volto a definire tutte le sezioni di una canzone, le parti da far suonare ai musicisti e ad avere un ascolto anticipato dell'ipotetico risultato finale.

Ricapitolando, ecco gli step fondamentali da seguire nella fase di pre-produzione:

1. Fare un *brainstorming* e raccogliere tutte le idee utili.
2. Individuare le *references* che funzioneranno da traino per il progetto.
3. Creare delle *personas* per individuare il target dell'ascoltatore tipo.
4. Scegliere quali *tecniche di microfonazioni* verranno utilizzate nella fase di produzione.

5. Scegliere quali *tools di spazializzazione* utilizzare durante le fasi di produzione e mixing.
6. Scegliere quali saranno i *media principali* dove verrà ascoltata la nostra produzione.
7. Realizzare una *demo* che permetterà la definizione della struttura della canzone e le parti suonate dagli strumentisti.

### 3.3.2 Produzione

Avendo svolto un ottimo lavoro di pre-produzione, la fase di produzione diventa un passaggio semplice, in cui controllare gli eventuali imprevisti. Si riparte dalle sessioni di registrazione e ripresa sonora, che andranno programmate in un calendario da rispettare entro la deadline. Le conseguenze di un'errata pre-produzione possono portare a riversare determinate scelte direttamente all'interno del processo di produzione.

Questa possibilità può spezzare la creatività, un punto centrale di questa fase. Avere delegato le scelte tecniche ad uno step precedente permette di dare massimo sfogo alle intuizioni e ai guizzi, compiere alcune piccole modifiche che possono valorizzare il risultato finale. Ma per fare questo è necessario avere la mente libera e affrontare con slancio l'eventuale problem solving.

Difficilmente una fase di produzione rispetta al 100% il programma della pre-produzione, soprattutto quando ci sono tante figure coinvolte tra sound engineer e artisti. Le responsabilità sono molte, l'errore è umano, ma è dagli errori talvolta che nascono le vere innovazioni. Ad esempio, è stato proprio durante il processo di produzione di Believe, il successo di Cher del 1998, che l'utilizzo errato di uno strumento per l'intonazione della voce ha creato un effetto che ha reso originale e riconoscibile quella canzone. Quell'effetto è oggi uno degli strumenti più utilizzati nel settore, l'Autotune.

Ciò che però è estremamente importante è scegliere definitivamente le registrazioni da mandare alla fase successiva: il mix. Dividere la creazione di una traccia in differenti fasi non è soltanto un modo di segmentare un processo di lavoro o di scegliere le persone che interverranno al suo interno. Fare ciò è necessario a non tornare più indietro sulle decisioni prese. Non avere controllo delle fasi, in questo settore, può diventare un circolo vizioso che rischia di non portare a fine progetto.

Dopo aver controllato di avere registrato tutte le parti e di aver attuato tutte le modifiche utili, è arrivato il momento di esportare tutte le tracce del progetto e inviarle al mix engineer per la realizzazione del mix. In sostanza, la fase di produzione è riassumibile in:

1. *Registrazione* delle parti definite.
2. *Problem solving* degli eventuali imprevisti sopraggiunti.
3. Aggiunta di nuovi guizzi definibili come *additional-production*.
4. *Post-produzione*, editing e rifinitura delle registrazioni effettuate.
5. Invio delle tracce separate al *mix engineer*.

### 3.3.3 Mixing

Se la produzione è paragonabile al cervello all'interno del processo di creazione di musica immersiva, la fase di mix è il cuore. In questo step non sarà solo necessario scegliere i volumi delle tracce, ma si agirà in maniera chirurgica per far suonare la traccia al meglio possibile, giocando con le dinamiche (in compressione) e i colori (in equalizzazione) degli strumenti registrati. Spesso, infatti, quando il budget lo consente, difficilmente chi svolge il lavoro di produzione si dedica anche al mixing.

Nel caso specifico della musica immersiva il discorso si fa ancora più delicato. Come è stato raccontato nell'introduzione a questo saggio, oggi non sono ancora presenti degli standard definiti per compiere scelte precise. Questo ha un valore molto importante. Il trattamento in mix della stereofonia è ormai delineato. Esso deriva dai sessant'anni di esperimenti e di produzioni a livello commerciale, viene insegnato e tramandato nelle migliaia di scuole per fonici in tutto il mondo. Il mix dell'audio immersivo, invece, è un nuovo mondo estremamente vasto e sconosciuto, che ha necessariamente bisogno di sperimentazioni per definire degli standard. L'audio percepito in cuffia ancor di più. Lunge da me quindi, dare delle indicazioni che siano definitive, bensì tutto quello che tratterò nel dettaglio sarà il risultato di esperimenti e valutazioni compiute grazie all'esperienza nel settore e ai pareri condivisi con i colleghi incontrati nel percorso.

Nel mix stereofonico la posizione standard degli strumenti nello spazio è definibile tramite il panning (left o right) e si può riassumere in questa immagine:

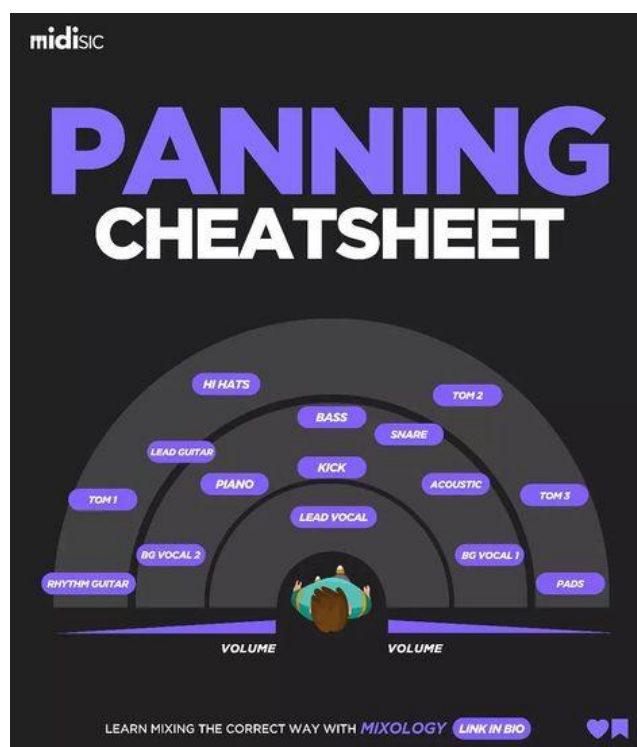


Figura 64 - Posizioni standard nel mix stereo (Midisic.com)

Al centro troviamo gli strumenti e le frequenze che devono avere caratteristiche monofoniche, come la cassa della batteria (Kick), il basso (Bass) e la voce principale (Lead Vocal). Possiamo notare una componente massiccia di basse frequenze in questa posizione, come accade anche con il surround, per dare potenza e profondità al coinvolgimento dell'ascoltatore. Questa sezione del mix può essere definita anche Mid, utilizzando le terminologie già spiegate nella tecnica Mid/Side, citata anche nel secondo capitolo.

Ai lati (o anche Side), invece, vi saranno tutti quegli strumenti che hanno usufruito del panning sul mixer (agendo sui panpot) e permetteranno a una canzone, mixata stereofonicamente, di fare la differenza rispetto a un'identica versione monofonica. Questa sensazione di spazialità stereo è cresciuta insieme allo sviluppo di questa tecnica, grazie anche a strumenti che hanno permesso di evidenziare la percezione laterale degli strumenti, come gli Imager, che permettono di definire quali bande di frequenze destinare alla sezione Mid o Side e con quanta enfasi.

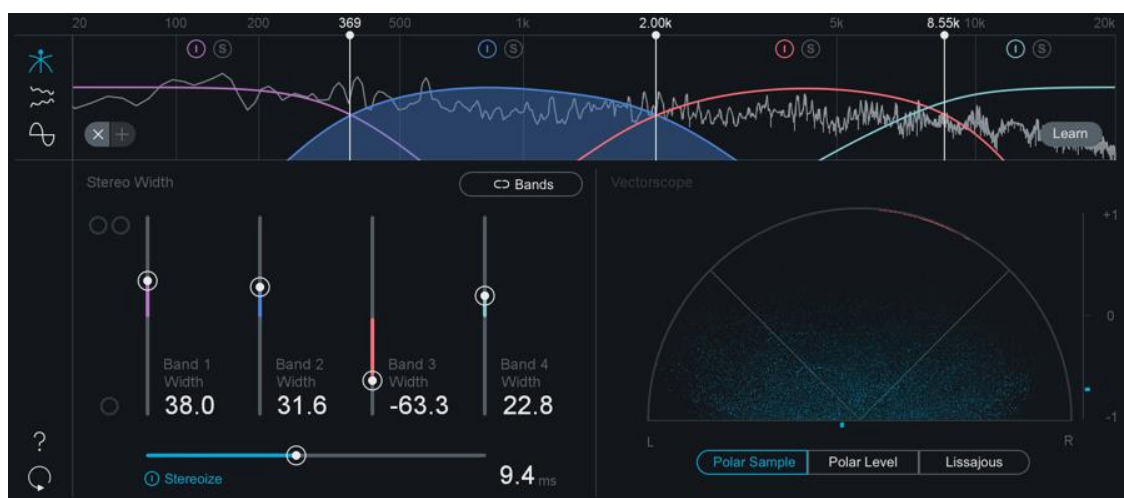


Figura 65 - Imager di Izotope Ozone

La storia è anche un cerchio che si chiude, o che talvolta ricomincia il giro. Così, stranamente, dopo aver raggiunto un ampio livello di apertura stereofonica nel momento in cui la musica è diventata liquida e digitale, recentemente la tendenza di alcuni generi ha riportato, invece, una chiusura dell'immagine stereofonica. Il ritorno del vinile (che per ragioni fisiche della puntina del giradischi non può avere un'estrema stereofonia) e l'utilizzo dei cellulari o delle piccole casse Bluetooth da parte degli adolescenti, hanno portato alcune produzioni (in particolare nell'hip hop) ad avere concentrazione di strumenti nelle sezioni monofoniche delle canzoni.

“Il medium è il messaggio” d'altronde, come diceva Marshall McLuhan nel 1964. Possiamo trarre da tutto ciò quanto sia importante monitorare la componente monofonica delle tracce stereo tramite strumenti come Levels di MasteringTheMix o molti altri mono checker hardware o software (persino nativi nelle DAW).

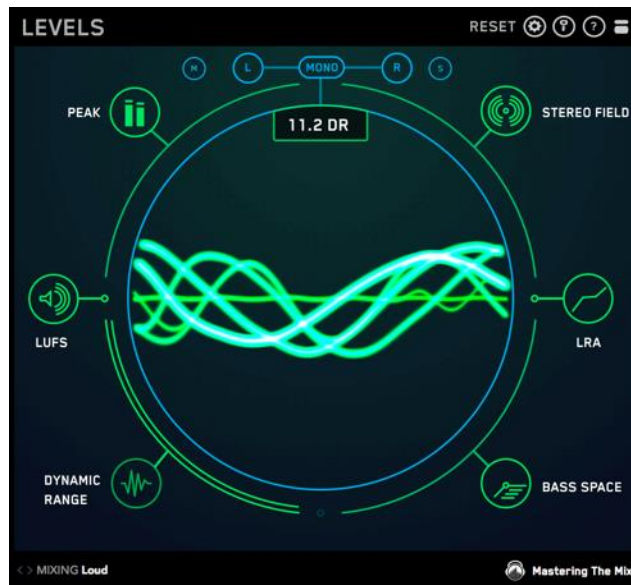


Figura 66 - Levels di MasteringTheMix

Le posizioni degli strumenti che sono state rappresentate in questo paragrafo, tipiche della stereofonia e dell'ascolto frontale, riprendono le tecniche di posizionamento orchestrale, determinate dall'acustica studiata all'interno dei teatri fin dall'antichità:

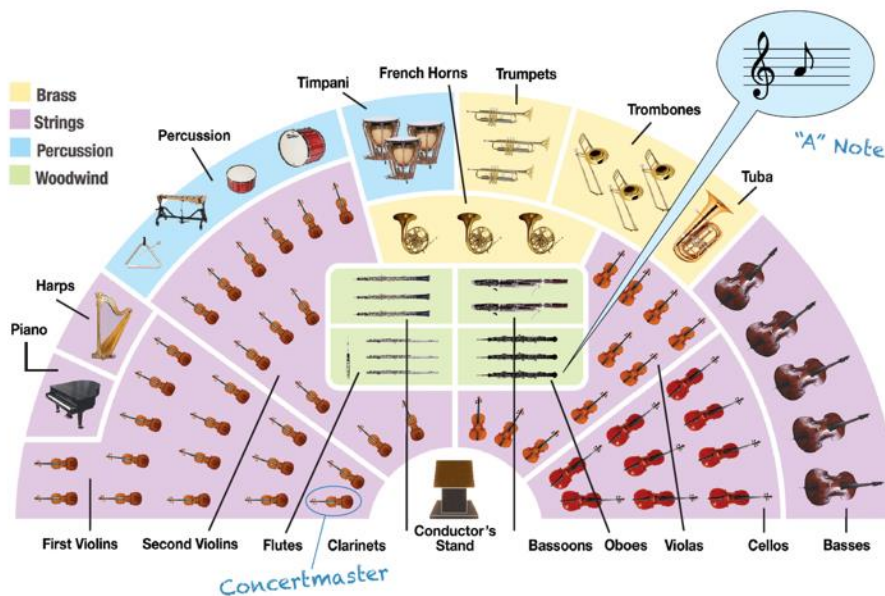


Figura 67 - Disposizione orchestrale (kennedy-center.org)

Come si svolge il mix di musica immersiva? La tesi entrerà nel dettaglio di questo argomento nel quarto e ultimo capitolo. La differenza di un mix immersivo è che invece di avere a disposizione un ascolto frontale stereofonico, si possiede un ascolto a 360 gradi, in cui ogni strumento può essere posizionato seguendo gli assi X, Y, Z. Cambierà molto l'impatto e il messaggio che arriverà all'ascoltatore.

Come riportato in Spatial Audio (Rumsey, 2001): "North e Hargreaves (1997) suggeriscono che c'è un grado ottimale di complessità nella musica che porta a risposte

positive dell'ascoltatore (ad. mi piace / non mi piace), e questo potrebbe benissimo estendersi alla complessità spaziale della riproduzione musicale. [...] Quelle registrazioni audio surround equilibrate della musica pop hanno certamente scoperto che è necessaria una grande attenzione nel grado in cui le sorgenti sono poste dietro l'ascoltatore, si muovono e generalmente confondono l'immagine complessiva creata".

Ecco che seguendo queste tecniche è possibile ricostruire un'esperienza immersiva sia a livello sonoro che a livello musicale, giocando soprattutto con le automazioni dello spazio per mettere in evidenza lo spostamento di alcuni strumenti. Ed è proprio in questo momento che la mancanza di standard renderà il tutto estremamente creativo e sperimentale, ma allo stesso tempo porterà possibile confusione ed eventuali problemi di percezione, come verrà evidenziato nel prossimo capitolo. Riassumendo in pillole questo approfondimento sul mix immersivo:

1. Scegliere la *posizione degli strumenti dello spazio* in base al messaggio che si vuole inviare all'ascoltatore.
2. Scegliere uno o più *software di spazializzazione* che abbiano un algoritmo affidabile di virtualizzazione dell'HRTF (o che dispongano di differenti set di HRTF).
3. Iniziare a mixare lasciando invariate le *posizioni Mid* derivanti dalle tecniche stereo.
4. *Sperimentare* i movimenti e le posizioni di tutto ciò che non è stato posto nell'ascolto frontale.
5. Decidere se far compiere dei *leggerissimi movimenti*, in alcuni determinati punti, anche ad alcuni elementi presenti nell'ascolto frontale (sapendo che questa decisione spiazzerà l'ascoltatore).

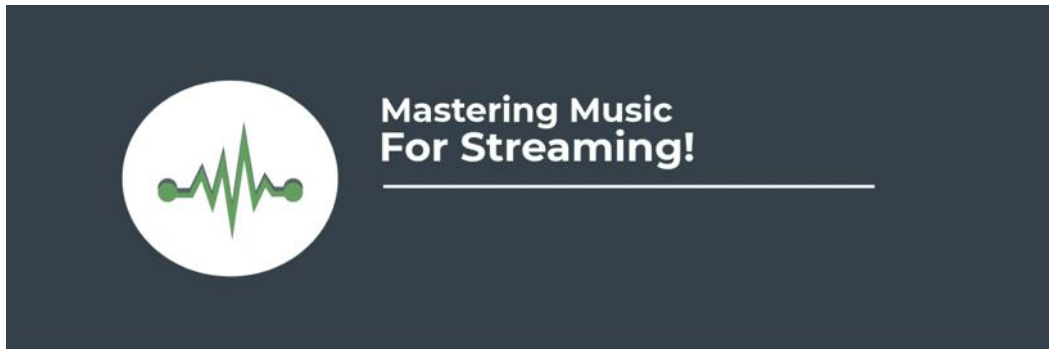
### 3.3.4 Mastering

Il mastering è una fase fondamentale, l'ultima nella catena di produzione. Il tecnico di mastering ha il compito di adattare il mixato alle sorgenti finali alle quali sarà destinato (cuffie, casse dell'auto, cinema, tv, radio, vinile, ecc.). Arrivati a questo punto, la scelta su quale media mettere maggior carico di responsabilità dovrebbe essere chiara, considerando che questa è una decisione da prendere fin dall'inizio, ovvero dalla pre-produzione.

Un problema sorto con la multimedialità è stato quello di rendere le canzoni il più versatili possibili in modo da essere riprodotte su diversi dispositivi. Qualcosa di molto complesso, che tendenzialmente non viene avvertito quando un mastering è ben svolto (se non dagli addetti ai lavori) grazie all'aiuto della compensazione del cervello e ai principi psicoacustici della ricostruzione della fondamentale.

Il valore medio dell'ampiezza sonora di una canzone viene misurato oggi in LUFS. Ogni piattaforma ha le sue specifiche. Ad esempio, iTunes, YouTube, Tidal e Spotify

richiedono parametri diversi di mastering. Questo è dovuto anche ai formati e alle compressioni digitali impiegate da questi servizi.



## LUFS For Streaming

(loudness units relative to full scale)

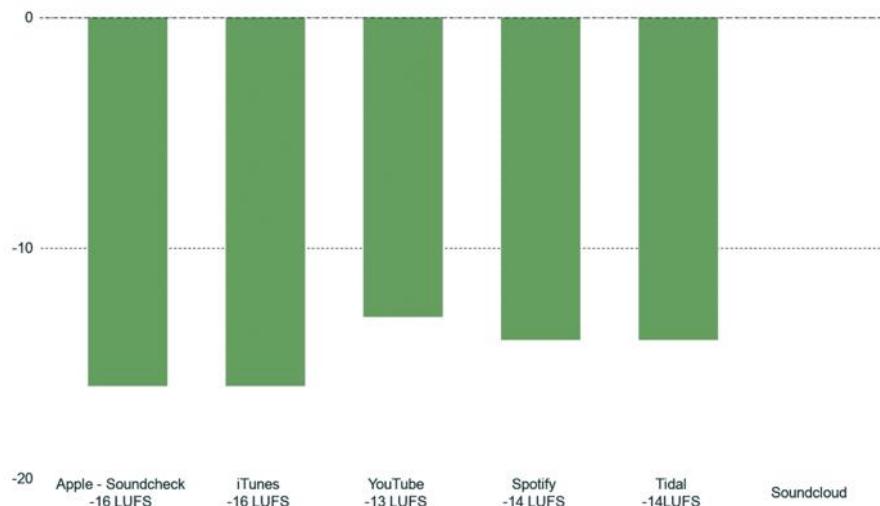
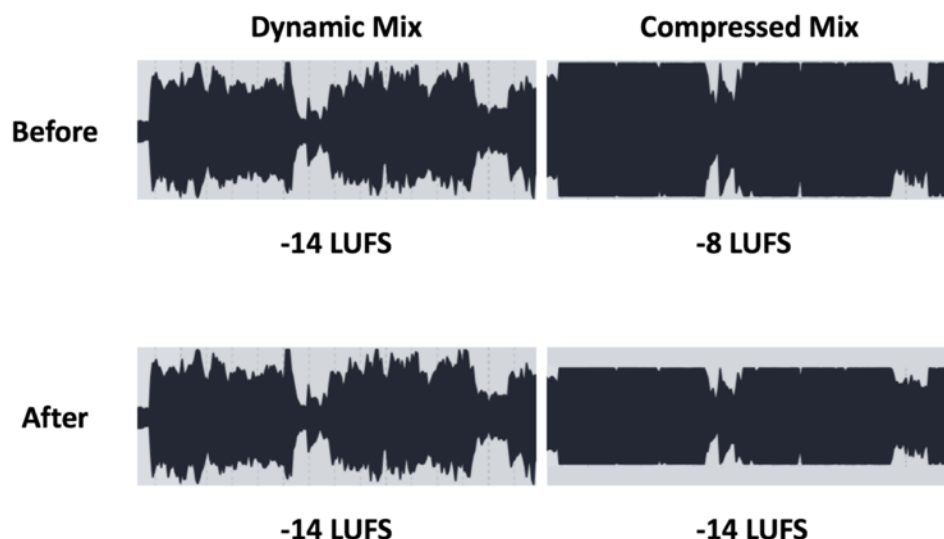


Figura 68 - Valori in LUFS per le diverse piattaforme di streaming  
([www.subaqueousmusic.com](http://www.subaqueousmusic.com))

Nonostante la scelta negli ultimi anni di virare le specifiche di mastering da una misurazione in  $db_{RMS}$  alla misurazione in LUFS, per cercare di preservare la dinamica nelle canzoni, l'ultima tendenza ha riportato in voga la rincorsa a cercare di arrivare a valori di LUFS sempre più alti per suonare più forti dei competitor (Spotify richiede tendenzialmente canzoni a -14 LUFS, ma molte vengono comunque caricate a -8). Un vero e proprio ritorno della *Loudness War*, iniziata negli anni 2000 con i CD. Uno degli esempi più famosi e riconosciuti di quell'epoca è l'album *Californication* dei *Red Hot Chili Peppers*, prodotto da Rick Rubin.



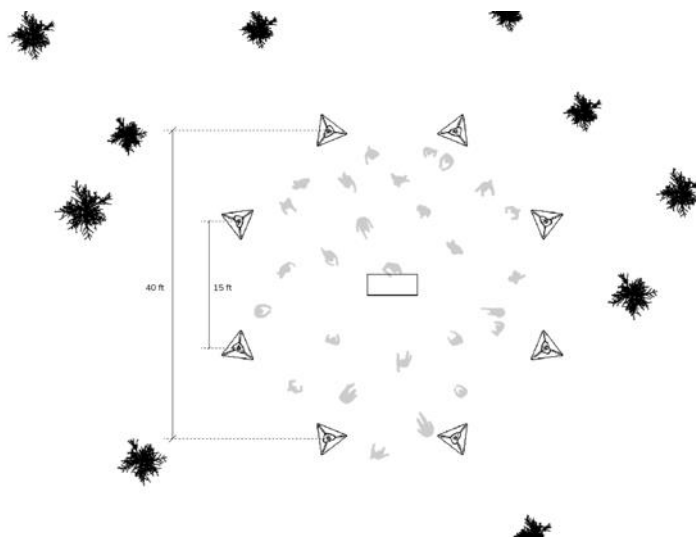


*Figura 69 - Differenza di dinamica nel caricamento su un servizio di streaming*

La fase di mastering, in formato binaurale, segue gli stessi processi del mastering normale, lavorando a due canali Left e Right. L'unica differenza che eviterei, personalmente, sono le applicazioni di equalizzazioni Mid/Side e l'evidenziamento dell'apertura "stereofonica" tramite Imager, poiché tutto questo dovrebbe essere avvenuto meticolosamente in fase di mix con gli strumenti di spazializzazione e la conseguente decodifica in formato binaurale.

### **3.4 Conseguenze commerciali**

Le applicazioni e le possibili conseguenze commerciali di tutti i concetti affrontati fino ad ora sono svariate. Immaginiamo di guardare nella sfera di cristallo e di poter prevedere quali saranno i prossimi passi in questo settore, sapendo che, rileggendo queste parole tra qualche anno, qualcosa sarà andato in porto, qualcosa non avrà ancora (o non vedrà mai) la luce. Iniziando dalla musica, che è il main focus di questo saggio, è probabile che il Dolby Atmos in cuffia (o la binaural music in generale) non troveranno uno spazio in breve tempo. Per quanto la possibilità effettivamente ci sia, è ancora troppo imponente l'importanza commerciale della stereofonia per farci dimenticare di lei. Soprattutto perché l'utente medio non è a conoscenza di cosa sia realmente l'audio spaziale e l'ascolto binaurale. La vera forza dell'esperienza immersiva è un avvenimento in presenza, come il cinema o i concerti. Ecco quale può essere il vero turning point di questo settore. L'utilizzo crescente di eventi immersivi in cui il suono viene proposto, sempre più di frequente, secondo le tecniche dell'audio spaziale. Già esistono situazioni di questo tipo: Envelop, della quale abbiamo parlato nel capitolo sui software di spazializzazione virtuale, permette di partecipare ad eventi di ascolto in audio spaziale, con otto speakers disposti ad ottagonato, tramite le tecnologie create dall'azienda.



*Figura 70 - Envelop Listening Experience*

Allo stesso modo, anche Dolby non vuole rimanere indietro, bensì tracciare una rotta parallela. Infatti, da differenti anni, il comparto Atmos è stato sviluppato anche per applicare il sistema agli eventi dal vivo.

Si è parlato molto delle tecnologie Atmos nel cinema, sono ormai dieci anni che questa tecnologia è stata rilasciata, eppure non è presente nella maggior parte dei cinema del mondo. Insomma, la strada da fare è molta, anche perché finché l'utente medio non avrà un training d'ascolto, è molto probabile che sarà difficile che esso colga davvero le proprietà spaziali e immersive da un prodotto musicale, soprattutto quando viene ascoltato in cuffia, senza l'esternalizzazione delle casse e la percezione delle basse frequenze sul proprio corpo.

Spostandoci leggermente su focus secondari, il settore automotive avrà un'importanza fondamentale nel training dell'ascoltatore, perché l'ambiente interno alle auto possiede caratteristiche acustiche spaziali da decenni. Il modo in cui le auto verranno progettate, soprattutto con l'arrivo del 5G e delle nuove tecnologie ad esso legate, potrà fare la differenza rispetto al modo in cui le persone ascoltano la musica. È anche per questo che, in questi anni, molti servizi di streaming come Tidal, Apple Music e Amazon Music hanno deciso di avere in catalogo delle playlist totalmente dedicate alla musica immersiva in Dolby Atmos. Un passo di rilievo in questa direzione sembra averlo mosso Apple, che tutto d'un tratto, nel 2021, ha fornito la Dolby Atmos Production Suite sulla DAW Logic Pro X. Allo stesso tempo, l'azienda di Cupertino ha fatto uscire sul mercato delle cuffie che permettano di fruire dell'audio spaziale, spingendo sull'acceleratore nel dare un'impronta marcata all'utente medio.

Anche il broadcasting si è munito per la trasmissione di audio multicanale, come dimostrato dalle ricerche del Fraunhofer Institute in collaborazione con Qualcomm (Businesswire, 2015). In questo caso viene posta particolare attenzione anche all'utilizzo interattivo del comparto audio nel broadcast, quindi il formato multicanale non viene usato soltanto per aumentare l'immersività, ma soprattutto per dare delle possibilità maggiori all'ascoltatore, che può decidere se cambiare i livelli dell'audio in

real-time, ad esempio disattivare o far emergere una telecronaca durante un evento sportivo.

Nel corso del saggio è stato poco citato il gaming, poiché ha già una sua ampia letteratura legata all'audio spaziale, molto più vasta della musica immersiva, la quale è tutt'oggi una branca minore. Riguardo ai videogiochi, però, un punto centrale che può coinvolgere tutti i rami dell'audio spaziale è l'utilizzo del visore, non ancora arrivato davvero nelle case delle persone. Una tecnologia che procede con la sua crescita, ma non è ancora esplosa davvero, lasciando ai nastri di partenza opportunità incredibili come la Cinematic VR, gli eventi in streaming visibili con il visore, la realtà aumentata su dispositivi indossabili (es. occhiali). Tutte situazioni già esistenti, sia chiaro, ma non ancora realmente sul mercato.

Anche le simulazioni avrebbero un possibile beneficio diretto dal miglioramento delle tecnologie per la produzione di musica immersiva, non solo a livello di percezione dell'immersione nella simulazione in sé, ma anche in applicativi collaterali, come l'auralizzazione nei contesti di valutazione architettonica.

Infine, un territorio dove la musica immersiva viene applicata e lo sarà sempre di più è il contesto delle installazioni museali. C'è da aspettarsi Biennali e Triennali sempre più ricche di esperienze di immersione sonora. Ma il vero interesse, in realtà, non è da porre sulle conseguenze che possiamo prevedere in un breve futuro, bensì su quelle che non possiamo minimamente immaginare.

## CAPITOLO 4 - Confronto tra stereofonia e formato binaurale nella musica immersiva ascoltata in cuffia

Questo saggio è stato scritto per introdurre tutte quelle che sono, ad oggi, le maggiori risorse riguardanti l'immersività musicale nel formato binaurale, ma non solo. Durante il percorso di ricerca che si è appena compiuto sono cambiate le tecnologie, si sono affinati gli algoritmi e sono aumentate le possibilità per testare determinate dinamiche immersive. Sarebbe stato molto più complesso quattro anni fa, nel 2018, pensare di mettere a confronto la stereofonia percepita in cuffia con l'immersività virtuale ricreata dalla binauralizzazione. Oggi esiste almeno la possibilità di tentare, anche a costo di sbagliare: l'errore è un valore che nella ricerca ha un peso positivo, perché permette di non commetterne ancora quando si percorre una strada già percorsa da un altro ricercatore.

Un primo passo è stata la pubblicazione, insieme ai professori Gianni Vercelli e Mauro Coccoli, dell'articolo "Comparing the perception of 'sense of presence' between a stereo mix and a binaural mix in immersive music" durante l'edizione 148 dell'Audio Engineering Society Convention di Vienna.

L'abstract dell'articolo recitava come di seguito:

*“Lo scopo di questo articolo è indagare se la musica binaurale coinvolge gli ascoltatori più della musica stereofonica. È stata condotta un'indagine sulle differenze di percezione durante l'ascolto della musica con le cuffie. È stato condotto un esperimento su un campione di otto persone per trarre alcuni risultati preliminari sulle preferenze degli ascoltatori tra un mix stereo e un mix binaurale. Per comprendere il grado di "sense of presence" dei soggetti testati è stato utilizzato un questionario basato sul modello ITC-SOPI, modificato per le specifiche esigenze dell'esperimento. La mancanza di standard sia nelle tecniche di mixaggio che nel software da utilizzare per creare un prodotto professionale in formato binaurale è stata una delle maggiori difficoltà incontrate.”*

Il vero problema emerso da questo esperimento era che non solo vi era una mancanza di standard in questo settore inerente alle tecniche di mixaggio, ma mancava anche un modello di questionario per capire davvero se la “sense of presence” dell'ascoltatore fosse maggiore nella stereofonia o nell'audio 3D binaurale. Non è bastato riadattare il modello ITC-SOPI, utilizzato in ambito video per la valutazione della “sense of presence”, nel tentativo di vagliare l'immersività musicale. Il risultato è stato accorgersi che durante l'esperimento la confusione principale derivava proprio dal questionario ereditato dal modello ITC-SOPI. Ma, come è stato anticipato nel corso di questa introduzione, è proprio dagli errori che si può cambiare rotta e trovare un nuovo modo per affrontare i problemi di ricerca.

Successivamente alla pubblicazione sono usciti altri interessanti articoli inerenti al contesto come Abehsera Morell, P. & Lee, H. (2021), Kim, S., & Howie, W. (2021),

Engel, I., Alon, D. L., Scheumann, K., Crukley, J., & Mehra, R. (2022), che hanno portato ad alcune valutazioni, e conferme, molto importanti, il cui risultato è stata la strutturazione di un nuovo esperimento volto alla comparazione tra stereofonia e audio 3D binaurale in ambito immersive music.

#### 4.1 Research questions

La principale domanda di ricerca che mi sono posto in partenza è la stessa posta da John Paul Titlow in un suo articolo (Titlow, J. P., 2017): “*L’audio spaziale 3D può reinventare la musica?*”.

Se la musica viene ascoltata ancora oggi utilizzando la stereofonia, ma le tecniche presenti attualmente possono sublimare questa tecnica, *perché questo non avviene?* Per riuscire ad avere delle risposte è stato necessario svolgere alcune operazioni che verranno descritte nel paragrafo sulle metodologie.

Quello che però è necessario analizzare prima è il contesto in cui siamo inseriti oggi, una situazione che permette a queste domande di ricerca di essere pensate.

Al fronte di decine di software e svariate tecnologie, descritte precedentemente lungo questo saggio, non si è ancora abbattuto il muro che divide i consumatori dall’utilizzo delle tecnologie immersive in larga scala. Questo ha un impatto diretto sul training dell’ascoltatore medio, che non sa riconoscere a fondo la differenza tra le due diverse fonti (stereo e 3D binaurale). Oggi la musica immersiva è disponibile solo su alcuni dei servizi di streaming, talvolta fruibile solo con alcune tipologie di cuffie: un’opzione non proprio aperta a tutti, considerando anche che il target medio dei fruitori dei servizi di streaming non è vastissimo:

*0-17 anni:* sotto i 18 anni si tende a non utilizzare i servizi di streaming semplicemente perché a pagamento. Questa fascia, però, è estremamente attiva con il gaming, un settore che investe molto sull’immersività.

*18-24 anni:* periodo della prima indipendenza economica. L’utilizzo dei servizi di streaming è presente e, insieme alla prossima fascia, rappresenta il core target.

*25 - 34 anni:* fascia che rappresenta i primi utenti dei servizi di streaming musicali al momento della loro creazione, per questo è la maggior utilizzatrice.

*35 - 44 anni:* da questa fascia in poi l’utilizzo degli streaming services musicali diventa un accessorio ad altre situazioni, come l’ascolto dallo stereo di casa o dai dispositivi con assistenza vocale.

*45 - 54 anni:* l’utilizzo di questi servizi non è tra le piattaforme più utilizzate per fascia di appartenenza.

*55 - 64 anni:* fascia d’età concentrata maggiormente sull’uso dei social network e sulla partecipazione attiva alle community online.

*Più di 65 anni:* utenti particolarmente distanti dalle tecnologie legate all’immersività.

L’impossibilità ad avere un ascoltatore medio che abbia un training d’ascolto legato all’immersività è allo stesso tempo una grande opportunità per capire se realmente la riproduzione di audio spaziale binaurale ha le carte per sovvertire la stereofonia.

## 4.2 Metodologia: mixing stereo e mixing binaurale

Si è deciso di procedere utilizzando la Digital Audio Workstation Logic Pro X di Apple, la quale, come discusso in precedenza, ha ricevuto un aggiornamento che ha permesso la fruizione e l’utilizzo della Dolby Atmos Production Suite (con annessa binauralizzazione virtuale tramite il sistema Atmos).

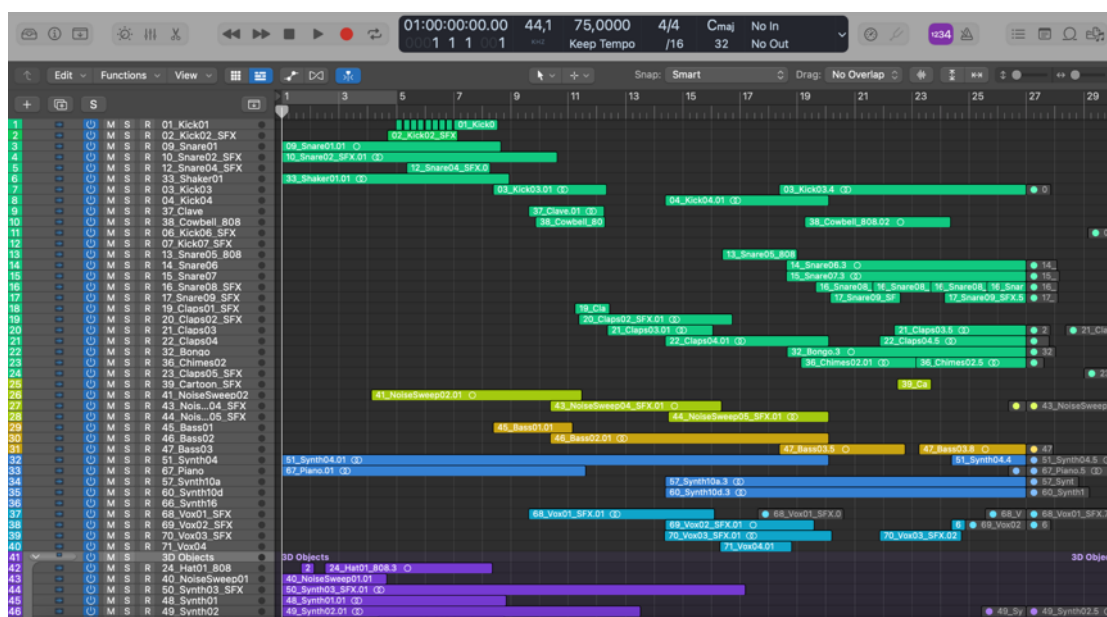


Figura 71 - Logic Pro X - Dolby Atmos mix session

Inoltre, questa serie di plugin permette un eccellente monitoring visuale della posizione dei suoni nello spazio. Come descritto in Tabry, Zatorre & Voss, 2013, quando le sorgenti sonore sono visibili, oltre che udibili, la disponibilità di informazioni visive può migliorare l’accuratezza delle stime di localizzazione del suono e quindi migliorare la realizzazione dell’esperimento.

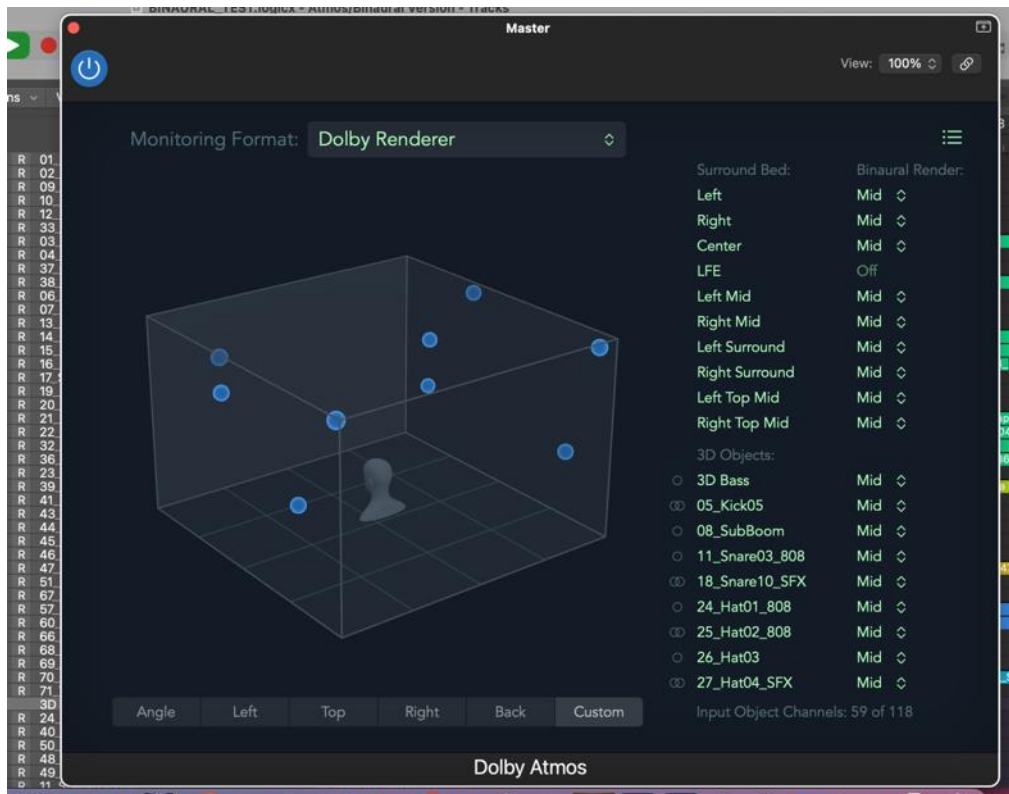


Figura 72 - Logic Pro X - Dolby Atmos Renderer

Non sono stati scelti sistemi open source, poiché, in questa fase, vista la mancanza di standard tecnici legati agli algoritmi di spazializzazione virtuale, i software open source tendono ad avere un grado di errore che renderebbe l'esperienza poco credibile. Inoltre, l'epidemia mondiale portata dal Covid-19 ha chiuso tutte le strutture universitarie, rendendo impossibile la creazione di un ambiente controllato per la realizzazione dell'esperimento. Per questo motivo, il test è stato pensato come applicabile a qualsiasi dispositivo con qualsiasi tipologia di cuffia. Da qui l'importanza di avere un sistema certificato che permettesse un'esportazione valida del file audio da ascoltare.

Ai tester non verrà richiesto alcun tipo di sistema di head-tracking, poiché non coerente con le dinamiche di questa ricerca. Ciò è stato dimostrato anche da Begault (2000), il quale ha determinato che il rilevamento della testa non è particolarmente importante per migliorare l'accuratezza della determinazione dell'azimut della sorgente o l'esternalizzazione dei segnali, migliorando principalmente la percezione front-back. Per la realizzazione dell'esperimento è stata utilizzata la repository gratuita Cambridge Music Technology dedicata agli esercizi di mixing, dalla quale sono stati scaricati gli audio files multitraccia della canzone Got Your Love di Karl Hungus (genere: future bass / tracce totali: 71).

È stato il genere indicato, perché nel mix stereofonico sono spesso presenti degli elementi di hard panning left/right, che potrebbero dare un elemento spaziale stereofonico utile nel confronto con l'audio spaziale 3D in formato binaurale.

Per mettere a confronto una traccia stereofonica (definita anche Traccia A) e la stessa che avesse connotati spaziali e immersivi in formato binaurale (definita Traccia B) è stato necessario realizzare i mix di entrambe le tracce, utilizzando sempre le stesse cuffie in fase di mixing. Sfruttando le potenzialità della DAW Logic Pro X si è proceduto, quindi, alla realizzazione del mix stereo della Traccia A, seguendo tutte quelle caratteristiche standard spiegate nel capitolo 3 di questo saggio e nei manuali di missaggio come “The Mixing Engineer’s Handbook di Owsinski (2017).

Per quanto riguarda il mix in audio binaurale, il procedimento è stato il seguente:

1. La sessione di mix della Traccia B è stata impostata ripartendo dalla sessione di mix della Traccia A (con annesse posizioni di panning nello spettro stereofonico).
2. Sono stati evidenziati e suddivisi gli elementi da spazializzare, come avviene per gli audio objects nel mixing del Dolby Atmos cinematografico. Guardando la prossima immagine, il gruppo di colore viola, denominato 3D Objects, rappresenta le tracce in discussione. Questo perché è molto importante, secondo l’interpretazione del sottoscritto, dare all’ascoltatore la possibilità di percepire alcuni movimenti specifici all’interno dello spazio a 360°, rispettando allo stesso tempo la percezione stereofonica di elementi (detti anche beds) quali la voce, il basso e alcune componenti ritmiche come kick o snare, che, resi mobili nello spazio, andrebbero a creare confusione e spaesamento.

Il gruppo 3D Objects contiene 32 tracce, dalla n°42 alla n°74.

Per studiare come è stato progettato il mix, la sessione è stata resa scaricabile dalla repository creata appositamente per il saggio e individuabile nella parte finale del testo.

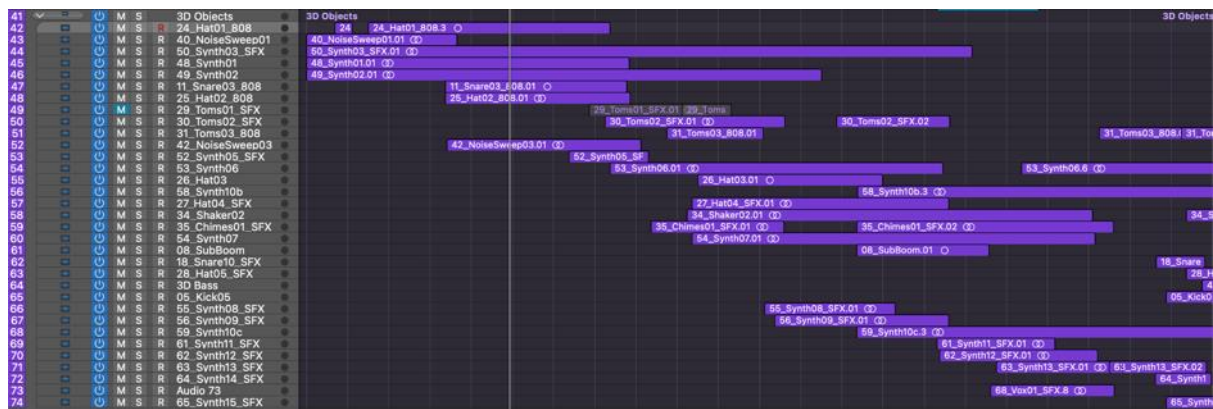


Figura 73 - Logic Pro X - Raggruppamento degli audio object da spazializzare

3. Utilizzando contemporaneamente il 3D Object Panner (da applicare sul canale che si intende spazializzare) e il Dolby Atmos Renderer (presente sul canale Master per la renderizzazione in real-time in formato binaurale), dare un equilibrio tra le posizioni nello spazio a 360° sfruttando le opzioni di visualizzazione degli audio objects.





Figura 74 - Logic Pro X - 3D Object Panner

4. Anche se la visualizzazione aiuta molto ad avere una corretta posizione nello spazio a 360°, non bisogna dimenticarsi che l'utente finale, nella musica immersiva, non ha riferimenti visivi delle posizioni degli strumenti nello spazio. Sarà quindi molto importante fare affidamento alle proprie orecchie e alle proprie sensazioni durante la realizzazione del mix.
5. Una volta definite tutte le posizioni, sarà necessario inserire un limiter sul canale Master e portare la traccia in formato binaurale allo stesso livello di LUFS della traccia stereofonica, per non influire negativamente sul test con un'escursione di dinamiche e di volumi troppo evidente.
6. Esportare la traccia in formato .wav - 16 bit - 44100 Hz. Il bouncing delle tracce non avrà nessuna differenza siccome il formato binaurale è a 2 canali (left/right) esattamente come il formato stereo.

In fase di esportazione si è deciso di esportare sia la Traccia A, sia la Traccia B, di durata massima 1.39 minuti (sezioni della canzone: intro, prima strofa, bridge e ritornello) per evitare una perdita di concentrazione dei tester in fase di ascolto. terminate le tracce, esse sono state caricate sulla piattaforma fidbak.audio, utilizzata molto dagli addetti ai lavori del settore, poiché permette di caricare file audio senza alcuna compressione.

### 4.3 Metodologia: il questionario

È stato realizzato un questionario, compilabile tramite Google Form, che avesse delle caratteristiche accessibili a tutti gli utenti e che non ripetesse gli errori commessi nell'articolo precedentemente pubblicato (De Sotgiu, Vercelli & Coccoli, 2020). Nella scrittura del questionario sono stati importanti i concetti emersi da Nunally e Bernstein (1994) sugli effetti soggettivi della riproduzione spaziale del suono e sulla distinzione tra giudizi e sentimenti. I giudizi sono risposte o percezioni umane essenzialmente prive di opinioni personali, che possono essere verificate esternamente (es. "quanto è lungo questo pezzo di corda?" o "qual è la posizione di questa sorgente sonora?"). I sentimenti, invece, sono legati alle preferenze e a una risposta emotiva, quindi non possono essere verificati esternamente (es. ciò che "mi piace/non mi piace" o ciò che è "buono/cattivo").

La prima pagina del questionario riporta tutte le informazioni utili al tester per capire il contesto in cui si trova e le ragioni dell'esperimento:

*TEST D'ASCOLTO CON CUFFIE/CUFFIETTE*

*Ciao! Mi chiamo Andrea De Sotgiu e sono un dottorando dell'Università degli Studi di Genova. Ciò che ti chiedo di fare è molto semplice: ascolta questi due file e rispondi alle domande. Ti consiglio di ascoltare prima la versione A e poi la versione B. Non esiste una risposta giusta, esiste solo la tua risposta!*

*LEGGI ATTENTAMENTE:*

*Ascolta le tracce con le cuffie. Qualsiasi esse siano vanno bene.  
Non proseguire il test fin quando non hai ascoltato tutte e due le tracce.*

*Ecco a te il link per effettuare il test:*

*<https://fidbak.audio/desso/player/db53c989a4e4/097ac8bebcbc>*

*Ascoltate? Adesso puoi proseguire!*

Non avendo controllo su ciò che i tester facessero, questo primo step non fornisce volontariamente alcuna informazione sul motivo per il quale i tester stanno ascoltando queste tracce, ovvero il confronto tra stereofonia e formato binaurale. Questo per evitare due atteggiamenti: un tester che non conosce nulla riguardo l'immersività potrebbe

decidere di andare a leggere alcune informazioni prima di procedere con l'ascolto, inquinando il test; un tester che conosce informazioni inerenti all'immersività, potrebbe procedere con un ascolto troppo attento che, allo stesso modo, inquinerebbe il test. Impartendo, invece, curiosità nel tester, questo sarà propenso a svolgere il questionario limitandosi a seguire le regole per conoscere maggiori informazioni su ciò che sta svolgendo.

Dopo aver ascoltato le due tracce, il secondo step si apre con una descrizione:

*RISPONDI ALLE DOMANDE*

*Ti stai chiedendo perché stai facendo questo test?*

*Ti svelerò tutto appena avrai finito di rispondere!*

*Alcune domande ti potranno sembrare ripetitive.*

*Non ti preoccupare, è solo per essere sicuro di ciò che hai provato. Vorrei capire che sensazioni hai sentito durante l'ascolto. Bada bene, non sono interessato a sapere se la canzone ti sia piaciuta, bensì concentrati su ciò che hai vissuto mentre stavi ascoltando le due versioni.*

Il questionario è costituito dalle seguenti domande:

1. *Quale tipologia di cuffie hai utilizzato per ascoltare le tracce? (se non hai usato le cuffie, ripeti il test prima di rispondere alle domande).*

Possibilità di scelta: Cuffiette (con cavo) / Cuffiette In-Ear (con cavo) / Cuffie (con cavo) / Cuffie Over Ear (con cavo) / Cuffiette classiche (Bluetooth) / Cuffiette In-Ear (Bluetooth) / Cuffiette AirPods (Bluetooth) / Cuffie (Bluetooth) / Cuffie Over Ear (Bluetooth).

La scelta di specificare così tante cuffie è perché, in primo luogo, ogni utente deve sentirsi libero di utilizzare il mezzo che vuole. Inoltre, non è detto che l'utente sappia realmente che cuffie stia utilizzando, per questo motivo sono state introdotte anche delle categorie meno specifiche e più generali. In questo modo, si potrà capire meglio quali siano state le scelte degli utenti, soprattutto se specifiche o meno.

2. *Hai sentito delle differenze tra le due versioni?*

Possibilità di scelta: sì / no.

3. *Quale versione hai preferito?*

Possibilità di scelta: versione A / versione B.

4. *Perché hai scelto proprio questa versione? Scrivilo qui sotto!*

5. *Quale delle due versioni ti ha fatto sentire più coinvolto emotivamente?*  
Possibilità di scelta: versione A / versione B.
6. *Ripensando alla Versione A, quale aggettivo sceglieresti per definire la sensazione che hai provato nell'ascolto?*  
Possibilità di scelta: disturbo / immersione / spaesamento / rassicurazione.
7. *Ripensando alla Versione B, quale aggettivo sceglieresti per definire la sensazione che hai provato nell'ascolto?*  
Possibilità di scelta: disturbo / immersione / spaesamento / rassicurazione.
8. *Quale delle due versioni ti ha dato un maggiore senso di immersività (come se fossi dentro alla canzone)?*  
Possibilità di scelta: versione A / versione B.
9. *In entrambe le versioni avrai probabilmente percepito che alcuni strumenti sono stati automatizzati per muoversi durante l'ascolto. Ma in una delle due questa sensazione era maggiore, secondo te quale?*  
Possibilità di scelta: versione A / versione B.
10. *A quale di queste fasce d'età appartieni?*  
Possibilità di scelta: ho meno di 18 anni / 18 - 24 anni / 25 - 34 anni / 35 - 44 anni / 45 - 54 anni / 55 - 64 anni / ho più di 65 anni.
11. *Ti lascio un box dove puoi scrivere ulteriori ed eventuali dettagli che avresti voluto aggiungere alle tue risposte. Hai concluso il tuo test, clicca "Avanti" per scoprire le ragioni di questo sondaggio e per inviare le risposte!*

Il questionario è particolarmente corto, veloce e diretto. Poche possibilità di scelta, minore possibilità di avere dei risultati difficili da consultare.

L'ultimo step è una descrizione finale legata alla musica immersiva prima dell'invio del questionario:

*IMMERSIVE AUDIO PRODUCTION*  
*Hai mai sentito parlare della "Musica Immersiva"?*  
*Intanto sfatiamo un mito: la musica 8D non esiste,*  
*non credere a chi dice il contrario! Esistono un*  
*sacco di termini che puoi approfondire per scoprire*  
*il mondo dell'immersività sonora, eccone alcuni:*  
*binaural format, ambisonic format, spatial audio,*  
*immersive music, immersive soundscapes.*

*Negli ultimi tre anni mi sono occupato di studiare come la musica possa superare la stereofonia utilizzando la virtualizzazione dello spazio sonoro applicata agli strumenti musicali. Ora che sono al termine di questo lungo viaggio ci sono molti più strumenti a disposizione per comprendere ciò di cui ti sto parlando. Basti pensare che in alcuni servizi di streaming musicale puoi già ascoltare musica in formato binaurale dalle tue cuffiette. Ma all'inizio del mio percorso di studi tutto questo era davvero poco comprensibile. È grazie a persone come te, che continuano a testare novità scientifiche, che la comunità di ricercatori fa nuove scoperte eclatanti ogni anno. Grazie ancora per il tuo supporto!*

**CLICCA "INVIA" PER INVIARE LE TUE RISPOSTE!**

L'approccio con i tester è tutt'altro che estremamente formale, perché vuole essere aperto a tutti. Avere a che fare con un test di carattere scientifico può apparire molto complesso, soprattutto quando si trattano argomenti di ricerca capillari come la musica immersiva.

#### 4.4 Risultati

I risultati raccolti dall'esperimento sono stati interessanti e inaspettati. All'esperimento hanno partecipato 53 persone provenienti da tutte le fasce d'età, esclusi coloro i quali hanno meno di 18 anni o hanno più di 65 anni.

A quale di queste fasce d'età appartieni?

53 risposte

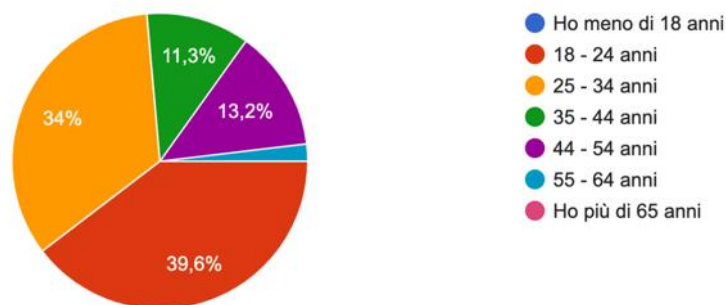


Figura 75 - Risultati del test: domanda 11

Sono state utilizzate quasi tutte le tipologie di cuffie: le tre più utilizzate sono state le generiche cuffiette con il cavo, seguite dalle EarPods / AirPods (Bluetooth) e le cuffiette in-ear con il cavo.

Quale tipologia di cuffie hai utilizzato per ascoltare le tracce? (se non hai usato le cuffie, ripeti il test prima di rispondere alle domande)

53 risposte

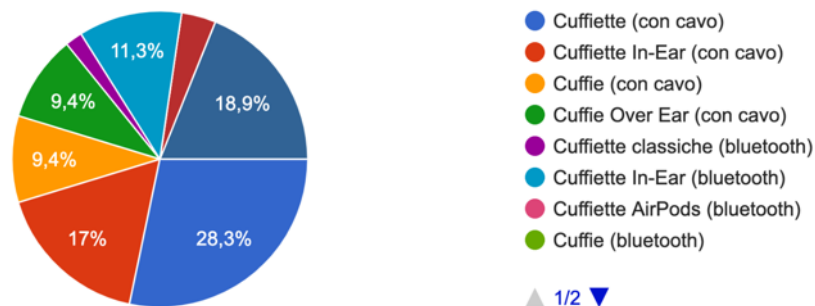


Figura 76 - Risultati del test: domanda 1

Il 64% degli utenti hanno utilizzato cuffie con il cavo per ascoltare l'esperienza. La versione che è stata preferita è la versione B, ovvero la traccia mixata in audio 3D binaurale:

Quale versione hai preferito?

53 risposte

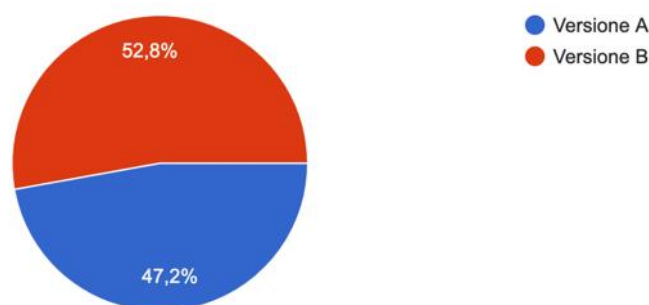


Figura 77 - Risultati del test: domanda 3

Questo è il vero risultato inaspettato: seppur di poco, la musica immersiva è stata preferita alla stereofonia, nonostante non ci siano degli standard di missaggio che aiutino i mix engineer a realizzare mix binaurali con lo scopo preciso di far percepire il senso di immersione all'utente. Nel momento in cui ci sarà una precisione sulle tecniche di mix da adottare, questo risultato potrà essere ancora più schiacciante.

Inoltre, entrando nel dettaglio di ciò che la versione B ha fatto percepire ai tester, il 45,3% ha definito come “immersione” la sensazione che ha provato nell’ascolto e il 24,5% ha indicato “rassicurazione”:

Ripensando alla Versione B, quale aggettivo sceglieresti per definire la sensazione che hai provato nell'ascolto?

53 risposte

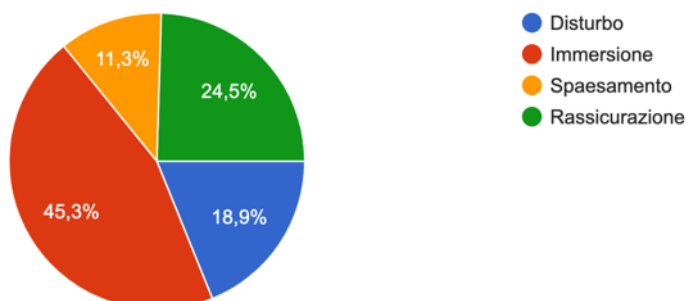


Figura 78 - Risultati del test: domanda 7

In totale il 69,8% ha espresso un parere positivo nei confronti dell’esperienza che ha vissuto. Anche per la versione A è accaduto lo stesso e il senso di immersione sembra essere stato percepito maggiormente:

Ripensando alla Versione A, quale aggettivo sceglieresti per definire la sensazione che hai provato nell'ascolto?

53 risposte

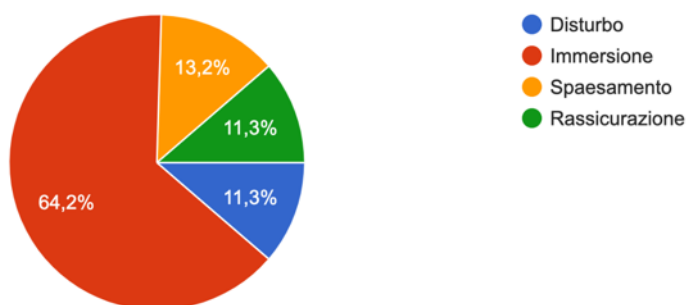
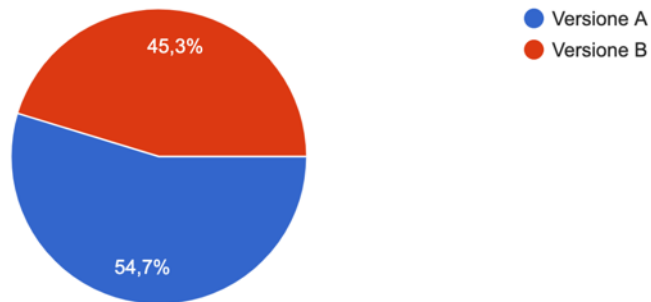


Figura 79 - Risultati del test: domanda 6

Minore è stato il senso di rassicurazione rispetto alla versione B. Uno dei possibili motivi per cui l’immersione è stata percepita maggiormente nella stereofonia è che nel mix della versione A erano presenti alcuni hard panning che giocavano molto con le sensazioni left e right. È probabile che nella versione B si potesse spingere molto di più nell’evidenziare questo aspetto. Risultato confermato anche da questi grafici:

Quale delle due versioni ti ha dato un maggiore senso di immersività (come se fossi dentro alla canzone)?

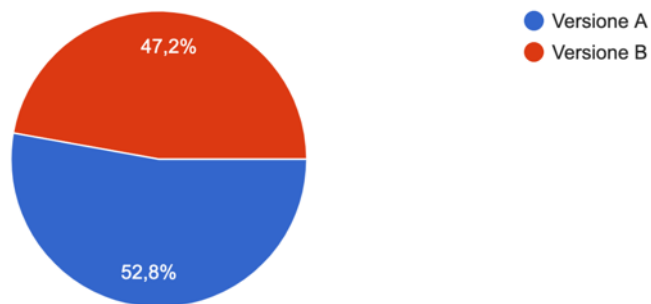
53 risposte



*Figura 80 - Risultati del test: domanda 8*

In entrambe le versioni avrai probabilmente percepito che alcuni strumenti sono stati automatizzati per muoversi durante l'ascolto. Ma in una delle d...questa sensazione era maggiore, secondo te quale?

53 risposte



*Figura 81 - Risultati del test: domanda 9*



Quale delle due versioni ti ha fatto sentire più coinvolto emotivamente?

53 risposte

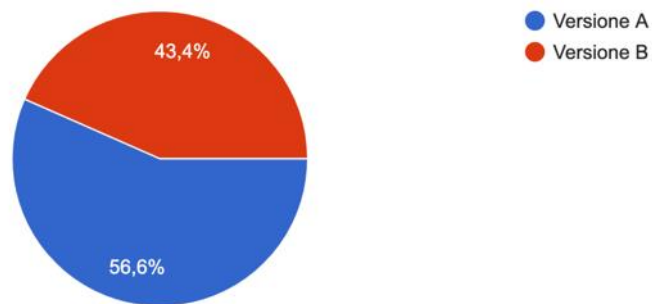


Figura 82 - Risultati del test: domanda 5

Hai sentito delle differenze tra le due versioni?

53 risposte

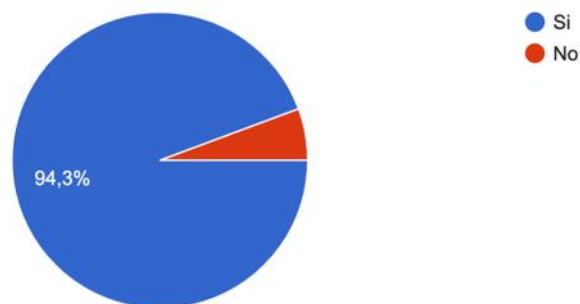


Figura 83 - Risultati del test: domanda 2

È nelle domande aperte che emergono alcuni risultati che possono chiarire maggiormente ciò che è stato indicato da questi grafici. Entrambe le versioni vengono definite “aperte” o con un “senso di immersione”. Il tester n°7 evidenzia ciò che gli ha fatto preferire una canzone rispetto a un’altra: “Ad un primo ascolto mi sembravano uguali, poi ho notato che nella prima traccia (la traccia che è stereofonica, nda) i suoni si muovono da destra a sinistra”. Il tester n°8, invece, sembra aver già avuto un training riguardo all’ascolto in audio spaziale. Egli preferisce la versione B perché: “L’audio sembrava più spazializzato”. Si deduce che l’importanza del training d’ascolto influenza il test e permette realmente all’utente di cogliere le differenze tra due versioni. Il tester n°9 indica che la versione B è: “Più ritmica”, quando in realtà contiene gli stessi strumenti. Probabilmente averli posti in determinate posizioni dello spazio a 360° ha permesso a questo ascoltatore di percepire meglio il groove.

Alcuni tester danno valutazioni su differenze di volume, altri addirittura di differenze armoniche. Il tutto evidenzia quanto sia soggettiva la percezione di un mix: dal contesto

d'ascolto (casa, autobus, all'aria aperta, in una stanza), all'attenzione posta; dal proprio HRTF alla cuffia che si sta utilizzando.

Nel prossimo paragrafo vi saranno ulteriori considerazioni in base ciò che è emerso da questo test, ma ciò che è più importante, è che, per quanta soggettività sia stata impiegata nelle decisioni dei tester, la maggioranza ha comunque preferito la versione B.

## **4.5 Conclusioni e sviluppi futuri**

La strada perché questo esperimento porti a nuovi risultati molto più eclatanti di quelli evidenziati è lunga e tortuosa, per questo avvincente. I dati analizzati sono un primo step che serve a dare alcune risposte utili sugli strumenti da utilizzare per creare questi test e come impostare futuri questionari. Sarebbe folle dire che questo esperimento abbia evidenziato totalmente una preferenza dell'audio binaurale alla stereofonia, ma ha chiarito che una preferenza può esistere.

Le tecnologie sono pronte all'uso, come è già stato ribadito, mancano solo gli standard teorici per permettere di fruire appieno di tutta la potenza che la musica immersiva porta con sé.

La scelta di lavorare a un mix stereofonico di un genere come la future bass è stata forse azzardata, probabilmente i risultati su un genere con meno transizioni strumentali sarebbero stati ancora migliori, ma questo è tutto da vedere. Il pensiero alla base della scelta di questo genere musicale è stata la consapevolezza che in esso vengono utilizzati molti suoni cinematografici, che sarebbero stati ottimi come 3D Audio Objects (cosa che, ad esempio, sarebbe stata molto meno presente in un genere come il rock).

A questo esperimento andrebbero fatti, quindi, almeno altri due miglioramenti: poter utilizzare sempre la stessa cuffia in un ambiente controllato e modificare il questionario per rendere più chiare alcune dinamiche (ad esempio una scissione ancora più chiara tra un giudizio oggettivo e una risposta emozionale o sentimentale).

Ma questo è solo un primo passo di molti altri. Nel momento in cui questa tesi è stata scritta, altre piattaforme e altri software di spazializzazione hanno portato nuovi aggiornamenti che dovrebbero essere testati. Più passerà il tempo e più sarà difficile scegliere quale algoritmo utilizzare per la spazializzazione e per la renderizzazione dei suoni, fin quando non si troverà una certificazione mondiale che attesterà gli algoritmi su cui basare le ricerche in ambito spatial audio.

La speranza è che questo scritto, una piccola goccia nell'oceano rappresentato dalla comunità scientifica, sia comunque utile per coloro i quali porteranno avanti le ricerche sulla musica immersiva.

## Conclusioni

Arrivati alla fine di questo percorso, le suggestioni sono molte, alcune di queste anticipate dal paragrafo 3.4 Conseguenze commerciali e dal paragrafo 4.5 Conclusioni e sviluppi futuri.

Il margine di crescita del settore della musica immersiva è lampante, sia per le possibilità già disponibili, che per quelle che la ricerca propone ogni mese agli eventi di settore.

Qualcosa che è emerso, ed è un punto cruciale dell'ascolto in cuffia, è la differenza d'ascolto di persona in persona. La struttura anatomica del nostro orecchio si potrebbe paragonare quasi ad un filtro personalizzato che permette ad ognuno di noi di ascoltare la sua realtà. In questo, dispositivi come le cuffie aumentano nettamente le differenze, proprio perché ogni cuffia ha una diversa tipologia e le combinazioni con le strutture anatomiche dell'essere umano possono essere milioni, se non di più.

Ecco che entra in gioco l'HRTF personalizzato, che molte università e aziende stanno già promuovendo in questo settore, per riuscire a tracciare un profilo d'ascolto diverso per ogni persona e quindi garantire una sensazione universale della percezione in cuffia. Non siamo così lontani: esattamente come i sistemi home-theatre possono scannerizzare una casa, per sfruttare riflessioni e rifrazioni a proprio vantaggio, anche le cuffie utilizzano già alcuni sensori per migliorare l'ascolto, ad esempio il noise-canceling. È molto probabile che in futuro qualsiasi tipo di cuffia profilerà il nostro orecchio dandoci il massimo dall'esperienza possibile.

In parallelo sarà necessario trovare gli standard mancanti di cui si è parlato in tutto il testo. Per fare ciò l'interesse nei confronti del settore dell'immersività audio/video deve necessariamente crescere e approdare finalmente a livello commerciale, ma con indotti molti più alti di adesso. Solo in tal caso queste tecnologie avranno davvero lo spazio che meritano. Non si può fare una previsione sicura di quando tutto ciò avverrà, potrebbe essere prestissimo, come tra dieci anni.

In conclusione, penso sia giusto ritagliarsi un momento per guardarsi indietro alla fine di questa maratona incredibile che è stato il dottorato di ricerca.

È la fine di un percorso molto lungo, con molte soddisfazioni, che è stato il percorso universitario. Al di là di ciò che verrà dopo, se il lavoro o la ricerca, penso che avere la possibilità di avere una borsa di studio per studiare un argomento che sia ama sia un privilegio. Per questo ringrazio il collegio docenti dell'Università di Genova, i professori Mauro Coccoli e Gianni Vercelli, per il supporto in questi anni.

È bello sapere che alla fine ha vinto la musica. Era giusto così.

## Appendice - Repository

È stata creata un repository in cui è stata inserita la sessione originale del progetto di Logic Pro X, ispezionabile ai fini di ricerca. Nello stesso progetto si trovano entrambe le sessioni: il mix della versione stereofonica e il mix della versione binaurale. Per accedervi è necessario andare in File → Project Alternatives → Scegliere tra “Atmos/Binaural Version” oppure “CUT Stereo Version”.

N.B. Per aprire il progetto è necessario avere l’aggiornamento 10.7.3 di Logic Pro X.



## Repository

<https://qrco.de/RepositoryPhDThesis>

## Bibliografia

- Abehsera Morell, P., & Lee, H. (2021, October 13). Binaural Mixing of Popular Music. Audio Engineering Society Convention 151. <https://www.aes.org/e-lib/browse.cfm?elib=21528>
- Agrawal, S., Simon, A., & Bech, S. (2019). Defining Immersion: Literature Review and Implications for Research on Immersive Audiovisual Experiences. New York, 14.
- Agterberg, M. J., Hol, M. K., Van Wanrooij, M. M., Van Opstal, A. J., & Snik, A. F. (2014). Single-sided deafness and directional hearing: Contribution of spectral cues and high frequency hearing loss in the hearing ear. *Frontiers in Neuroscience*, 8, 188.
- Aguilera, E., Lopez, J. J., Gutierrez, P., & Cobos, M. (2014). An Immersive Multi-Party Conferencing System For Mobile Devices Using Binaural Audio.
- ANSA, A. (2022, February 3). Metaverso, è corsa all'acquisto di immobili virtuali—Hi-tech. Agenzia ANSA. [https://www.ansa.it/sito/notizie/tecnologia/hitech/2022/02/03/metaverso-e-corsa-allacquisto-di-immobili-virtuali\\_eebf695a-db7d-4752-9399-39ec830b1319.html](https://www.ansa.it/sito/notizie/tecnologia/hitech/2022/02/03/metaverso-e-corsa-allacquisto-di-immobili-virtuali_eebf695a-db7d-4752-9399-39ec830b1319.html)
- Atmos è ovunque: La vittoria di Dolby è la sconfitta dell'home cinema. (n.d.). DDay.it. Retrieved April 20, 2022, from <https://www.dday.it/redazione/22069/atmos-e-ovunque-la-vittoria-di-dolby-e-la-sconfitta-dellhome-cinema>
- BBC - Radio 3 in immersive sound. (n.d.). BBC. Retrieved April 20, 2022, from <https://www.bbc.co.uk/programmes/articles/29L27gMX0x5YZxkSbHchstD/radio-3-in-immersive-sound>

- Begault, D. (1991). Challenges to the successful implementation of 3D sound. *J. Audio Eng. Soc.* 39, 11, pp. 864–870.
- Begault, D. (1994). *3D Sound for Virtual Reality and Multimedia*. Academic Press.
- Begault, D. (2000). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. Presented at AES 108th Convention, Paris, 19–22 February. Preprint 5134. Audio Engineering Society.
- Bellanti, E., Corsi, A., De Sotgiu, A., & Vercelli, G. (2018, June 1). “Changes”: An immersive spatial audio project based on low-cost open tools.  
[https://doi.org/10.1007/978-3-319-95270-3\\_15](https://doi.org/10.1007/978-3-319-95270-3_15)
- Bernstein, L. R. (2001). Auditory processing of interaural timing information: New insights. *Journal of Neuroscience Research*, 66(6), 1035–1046.
- binaural\_jit by Kasper Fangel Skov: Free Max for Live Binaural Spatialisation | Sonic Bloom. (2017, April 21). [https://sonicbloom.net/en/binaural\\_jit-by-kasper-fangel-skov-free-max-for-live-binaural-spatialisation/](https://sonicbloom.net/en/binaural_jit-by-kasper-fangel-skov-free-max-for-live-binaural-spatialisation/)
- Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization* (Revised Edition). MIT Press.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*, Rev. ed. (J. Allen, Trans.). Cambridge, MA: MIT Press.
- Blumlein, A. (1931). British Patent Specification 394325. Improvements in and relating to sound transmission, sound recording and sound reproducing systems.
- Bresler, Z. (2020). Music and Space: A case of live immersive music performance with the Norwegian post-rock band Spurv. *Music and Space*, 5.
- Bridge, T. B. (2018, January 25). G’Audio Lab Launches Live Streaming Audio for 360-Degree Video—The Broadcast Bridge—Connecting IT to Broadcast.

<https://www.thebroadcastbridge.com/content/entry/10360/gaudio-lab-launches-live-streaming-audio-for-360-degree-video>

- Brughera, A., Dunai, L., & Hartmann, W. M. (2013). Human interaural time difference thresholds for sine tones: The high-frequency limit. *Journal of the Acoustical Society of America*, 133, 2839–2855.
- Bruschi, V., Nobili, S., Cecchi, S., & Piazza, F. (2020). An Innovative Method for Binaural Room Impulse Responses Interpolation. 10.
- Cassorla, A. M., Kearney, G., Hunt, A., Riaz, H., Stiles, M., & Murphy, D. T. (2020). Augmented Reality for DAW-Based Spatial Audio Creation Using Smartphones. 10.
- Cipriani, A., & Giri, M. (2019a). *Musica elettronica e sound design. Teoria e pratica con Max 8—Volume 1 (4° edizione)*. ConTempoNet.
- Cipriani, A., & Giri, M. (2019b). *Musica elettronica e sound design. Teoria e pratica con Max 8—Volume 2 (3° edizione)*. ConTempoNet.
- Cipriani, A., & Giri, M. (2021). *Musica elettronica e sound design. Teoria e pratica con Max 8—Volume 3*. ConTempoNet.
- Compressore Audio: Quali tipi esistono e caratteristiche. (2020, May 8). Kinarecords. <https://www.kinarecords.com/compressore-audio-come-funziona-tipologie/>
- Conetta, R. & et al. (2015). Spatial Audio Quality Perception (Part 1): Impact of Commonly Encountered Processes. *Journal of the Audio Engineering Society*, 62(12), 831–846. <https://doi.org/10.17743/jaes.2014.0048>
- Cooper, D. and Shiga, T. (1972). Discrete matrix multichannel stereo. *J. Audio. Eng. Soc.*, 20, pp. 346–360.

Cosa sono i LUFS? Spiegazione della misurazione del volume. (2020, November 16).

LANDR Blog. <https://blog.landr.com/it/cosa-sono-i-lufs-spiegazione-della-misurazione-del-volume/>

Cosimi, E. (2017). Sintetizzatori virtuali. Teoria e tecnica. Curci.

Daubney, C. (1982). Ambisonics – an operational insight. *Studio Sound*, August, pp. 52–58.

Davis, M. F. (2003). History of spatial coding. *Journal of the Audio Engineering Society*, 51(6), 554–569.

De Sotgiu, A. D., & Vercelli, M. C. and G. (2020, May 28). Comparing the perception of “sense of presence” between a stereo mix and a binaural mix in immersive music. *Audio Engineering Society Convention 148*. [http://www.aes.org/e-lib/browse.cfm?elib=20826&fbclid=IwAR3OBHYJGv1y6TlXrWaLx3qYwLxM7X4BTzQsAdTVUtBSTlFD8xra5CK\\_cs](http://www.aes.org/e-lib/browse.cfm?elib=20826&fbclid=IwAR3OBHYJGv1y6TlXrWaLx3qYwLxM7X4BTzQsAdTVUtBSTlFD8xra5CK_cs)

De Sotgiu, A., Coccoli, M., & Vercelli, G. (2020). Tools for Immersive Music in Binaural Format. *Ambient Intelligence – Software and Applications*, 54–60. [https://doi.org/10.1007/978-3-030-58356-9\\_6](https://doi.org/10.1007/978-3-030-58356-9_6)

Dolby Laboratories (2013). *Dolby® Atmos® Next-Generation Audio for Cinema - White Paper*.

Dukes, B. (2013). *Audio Engineering Society*. 11.

Dwivedi, J. (n.d.). *Mixology—A Complete Mixing Ebook* • midisic. Midisic. Retrieved March 15, 2022.

Eargle, J. (ed.) (1986). *Stereophonic Technique: An Anthology of Reprinted Articles*. Audio Engineering Society.



- Elen, R. (1983). Ambisonic mixing – an introduction. Studio Sound, September, pp. 40–46.
- Engel, I., Alon, D. L., Scheumann, K., Crukley, J., & Mehra, R. (2022). On the Differences in Preferred Headphone Response for Spatial and Stereo Content. *Journal of the Audio Engineering Society*, 70(4), 271–283.
- Farnell, A. (2010). *Designing Sound*. MIT Press.
- Fonseca, N. (n.d.-b). *Sound Particles—3D Audio eBook*. Retrieved March 18, 2022, from <https://soundparticles.com/resources/ebooks/3daudio>
- Fox, B. (1983). Holophonics: an investigation. Studio Sound, July, pp. 90–96.
- Franinović, K., & Serafin, S. (2013). *Sonic Interaction Design*. MIT Press.
- Frank, M., Zotter, F., & Sontacchi, A. (2015). *Producing 3D audio in Ambisonics*.
- Fraunhofer IIS Demonstrates Real-Time MPEG-H Audio Encoder System for Broadcast Applications at IBC. (2014, September 10). <https://www.businesswire.com/news/home/20140910005837/en/Fraunhofer-IIS-Demonstrates-Real-Time-MPEG-H-Audio-Encoder-System-for-Broadcast-Applications-at-IBC>
- Fraunhofer IIS, Qualcomm and Technicolor to Demonstrate the World’s First Live Broadcast of MPEG-H Interactive and Immersive TV Audio. (2015, April 10). <https://www.businesswire.com/news/home/20150410005978/en/Fraunhofer-IIS-Qualcomm-and-Technicolor-to-Demonstrate-the-World%E2%80%99s-First-Live-Broadcast-of-MPEG-H-Interactive-and-Immersive-TV-Audio>
- Free Tools for Live Unlock 3D Spatial Audio, VR, AR | Ableton. (n.d.). Retrieved April 20, 2022, from <https://www.ableton.com/en/blog/free-tools-live-unlock-3d-spatial-audio-vr-ar/>

- Gardner, M. B., & Gardner, R. S. (1973). Problem of localization in the median plane: Effect of pinnae cavity occlusion. *Journal of the Acoustical Society of America*, 53, 400–408.
- Ge, Z., Li, L., & Qu, T. (2020). The Ambisonic Partially Matching Projection Decoding Method for Near-field Sound Sources. 8.
- Gerzon, M. (1973). Periphony: with-height sound reproduction. *J. Audio. Eng. Soc.*, 21, pp. 2–10.
- Gerzon, M. (1974). Surround sound psychoacoustics. *Wireless World*, 80, pp. 483–486.
- Gerzon, M. (1977). Criteria for evaluating surround sound systems. *J. Audio. Eng. Soc.*, 25, pp. 400–408.
- Gerzon, M. (1983) Ambisonics in multichannel broadcasting and video. Presented at 74th AES Convention, New York, October. Preprint 2034. Audio Engineering Society.
- Gerzon, M. (1983) Ambisonics in multichannel broadcasting and video. Presented at 74th AES Convention, New York, October. Preprint 2034. Audio Engineering Society.
- Gerzon, M. (1990). Three channels: the future of stereo? *Studio Sound*, June, pp. 112–125.
- Gerzon, M. (1992). Psychoacoustic decoders for multispeaker stereo and surround sound. Presented at 103rd AES Convention, San Francisco, 1–4 October. Preprint 3406. Audio Engineering Society.
- Gerzon, M. A. (1975). The design of precisely coincident microphone arrays for stereo and surround sound. *Proceedings of the 50th Audio Engineering Society Convention London, UK.*

- Gerzon, M. and Barton, G. (1992). Ambisonic decoders for HDTV. Presented at 92nd AES Convention, Vienna. Preprint 3345. Audio Engineering Society.
- Giliberti, D. A., Iseini, F., Pelagalli, N., Terenzi, A., & Cecchi, S. (2020). An Advanced Audio System for Stereo Reproduction Enhancement. 9.
- Gölles, L., Drack, V., Zotter, F., & Frank, M. (2020). Influence of horizontal loudspeaker layout geometry on sweet area shape for widened/diffuse frontal sound.
- Gould, R. (2018) Let's Test: 3D Audio Spatialization Plugins. Designing Sound.
- Guezenoc, C., & Séguier, R. (2020). Dataset Augmentation and Dimensionality Reduction of Pinna-Related Transfer Functions.
- GutierrezCastro, R. (n.d.). Dolby ® Atmos ® Next-Generation Audio for Cinema Issue 3. Retrieved March 26, 2022, from [https://www.academia.edu/42197048/Dolby\\_Atmos\\_Next\\_Generation\\_Audio\\_for\\_Cinema\\_Issue\\_3](https://www.academia.edu/42197048/Dolby_Atmos_Next_Generation_Audio_for_Cinema_Issue_3)
- Hear From The Masters. (n.d.). Retrieved April 20, 2022, from <https://professional.dolby.com/music/hear-from-the-masters/>
- Herre, J. & et al. (2015). MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding. Journal of the Audio Engineering Society, 62(12), 821–830. <https://doi.org/10.17743/jaes.2014.0049>
- Hertz, B. (1981). 100 years with stereo: the beginning. J. Audio Eng. Soc., 29, 5, pp. 368–372.
- Hollebon, J., & Fazi, F. M. (2020). Generalised Low Frequency 3D Audio Reproduction Over Loudspeakers. 11.
- HOLOPLOT - A new era of audio. (n.d.). Holoplot. Retrieved April 20, 2022, from <https://holoplot.com/>

MelodyVR - Live music that comes to you.

from <https://melodyvr.com/>

Howie, W. (2019). Pop and Rock music audio production for 22.2 Multichannel Sound: A Case Study. 10.

ITU-R (1993). Recommendation BS. 775: Multi-channel stereophonic sound system with or without accompanying picture. International Telecommunications Union.

Johansson, M. (n.d.). VR for Your Ears: Dynamic 3D Audio Is Coming Soon. IEEE Spectrum, 5.

Jorge, Davila-Chacon, Jindong, Liu, Stefan, & Wermter. (2019). Enhanced Robot Speech Recognition Using Biomimetic Binaural Sound Source Localization. IEEE Transactions on Neural Networks and Learning Systems, 30(1), 138–150.  
<https://doi.org/10.1109/TNNLS.2018.2830119>

JVC (2018): JVCKENWOOD “EXOFIELD” Headphone Technology. (n.d.).  
[http://pro.jvc.com/pro/pr/2018/ces/JVC\\_Exofield.html](http://pro.jvc.com/pro/pr/2018/ces/JVC_Exofield.html)

Kamekawa, T., & Marui, A. (2020). Are full-range loudspeakers necessary for the top layer of Three-Dimensional audio?

Kasper Fangel Skov. (n.d.). Retrieved April 20, 2022, from  
<https://www.kasperskov.dk/>

Keating, P., & King, A. J. (2013). Developmental plasticity of spatial hearing following asymmetric hearing loss: Context-dependent cue integration and its clinical implications. *Frontiers in Systems Neuroscience*, 7,  
<https://doi.org/10.3389/fnsys.2013.00123>.

Kim, S., & Howie, W. (2021). Influence of the Listening Environment on Recognition of Immersive Reproduction of Orchestral Music Sound Scenes. *Journal of the Audio Engineering Society*, 69(11), 834–848.

- Kim, S., Chon, S. H., Okumura, H., & Sakamoto, S. (2020). Comparing training effects associated with two sets of HRTF data on auditory localization performance. 6.
- Klein, F., Neidhardt, A., Seipel, M., & Sporer, T. (2017). Training on the acoustical identification of the listening position in a virtual environment. New York, 8.
- Krukowski, D. (2019). *Ways of Hearing*. MIT Press.
- Kumpik, D. P., Kacelnik, O., & King, A. J. (2010). Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans. *Journal of Neuroscience*, 30(14), 4883–4894.
- Laboratories, D. (2014). *Dolby Atmos Next Generation Audio for Cinema*. 18.
- Laboratories, D. (n.d.). *Authoring for Dolby Atmos Cinema Sound Manual*. 132.
- Lam, A., Lee, M.-L., & Philbert, S. (2019). *Measuring Speech Intelligibility Using Head-Oriented Binaural Room Impulse Responses*. New York, 5.
- Lau, F., & Meier, M. (2020). *Optimized binaural rendering of Next Generation Audio using virtual loudspeaker setups*.
- Lee, H., & Borzym, K. (2020). *Creating Virtual Height Loudspeakers for Dolby Atmos and Auro-3D Using VHAP*.
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments*, 10(3), 282–297.  
<https://doi.org/10.1162/105474601300343612>
- Let's Test: 3D Audio Spatialization Plugins. (n.d.). Retrieved April 20, 2022, from <https://designingsound.org/2018/03/29/lets-test-3d-audio-spatialization-plugins/>
- Live MPEG-H Audio System production chain for the 2019 Eurovision Song Contest showcased at IBC show – Fraunhofer Audio Blog. (n.d.). Retrieved November

- 11, 2019, from <https://www.audioblog.iis.fraunhofer.com/mpeg-h-production-chain-esc-ibc>
- Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., & Noisternig, M. (2013, May). Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions. Audio Engineering Society Convention 134. <http://www.aes.org/e-lib/browse.cfm?elib=16781>
- Mamo, H. (2020, January 31). Here's the \*Real\* Reason Finneas Recorded Billie Eilish's Grammy-Winning Album At Home. Billboard. <https://www.billboard.com/music/pop/finneas-explains-recorded-billie-eilish-album-bedroom-8549785/>
- Manola, F., Genovese, A., & Farina, A. (2012). A comparison of different surround sound recording and reproduction techniques based on the use of a 32 capsules microphone array, including the influence of panoramic video. Audio Engineering Society 25th UK Conference: Spatial Audio in Today's 3D World. York, UK.
- Martin, G., Woszczyk, W., Corey, J. and Quesnel, R. (1999). Sound source localisation in a five channel surround sound reproduction system. Presented at 107th AES Convention, New York, 24–27 September . Preprint 4994. Audio Engineering Society.
- Monofonia nell'Enciclopedia Treccani. (n.d.). Retrieved April 20, 2022, from <https://www.treccani.it/enciclopedia/monofonia>
- Moore, B. C. J., Oldfield, S. R., & Dooley, G. (1989). Detection and discrimination of spectral peaks and notches at 1 and 8 kHz. Journal of the Acoustical Society of America, 85, 820–836.

- Moorer, J.A. (2000) Audio in the New Millennium, *J. Audio Eng. Soc.*, vol. 48, no. 5, pp. 490, 492, 494-498.
- Murphy, M., & Anstey, R. (2020). Reimagining Robb: The Sound of the World's First Sample-based Electronic Musical Instrument circa 1927. 5.
- Musicant, A. D., & Butler, R. A. (1985). Influence of monaural spectral cues on binaural localization. *Journal of the Acoustical Society of America*, 77(1), 202–208.
- Nachbar, C., Zotter, F., Deleflie, E., & Sontacchi, A. (n.d.). Ambix - A Suggested Ambisonics Format.
- Neidhardt, A., & Reif, B. (2020). Minimum BRIR grid resolution for interactive position changes in dynamic binaural synthesis. 10.
- Nishiguchi, T., Nakayama, Y., Okumura, R., Sugimoto, T., Imai, A., Iwaki, M., Hamasaki, K., Ando, A., Kitajima, S., Otsuka, Y., & Shimaoka, S. (2007). Production and Live Transmission of 22.2 Multichannel Sound with Ultrahigh-definition TV. 11.
- North, A. and Hargreaves, D. (1997). Experimental aesthetics and everyday music listening. In *The Social Psychology of Music*, eds. North, A. and Hargreaves, D., Oxford University Press.
- November 16).
- Nunally, J., & Bernstein, I. (1994). *Psychometric Theory* (3rd ed.). New York and London: McGraw Hill.
- O'Donovan, A., & Duraiswami, R. (2010). Audio-visual panoramas and spherical audio analysis using the audio camera. *Proceedings of the 16th International Conference on Auditory Display (ICAD2010)* (pp. 167–168). Washington, DC.

- Oldfield, S. R., & Parker, S. P. (1984a). Acuity of sound localization: A topography of auditory space: I: Normal hearing conditions. *Perception*, 13, 581–600.
- Oldfield, S. R., & Parker, S. P. (1984b). Acuity of sound localization: A topography of auditory space: II: Pinna cues absent. *Perception*, 13, 601–617.
- Olson, A. F. (1976). A History of High-Quality Studio Microphones. *Journal of The Audio Engineering Society*, 24.
- Oramus, T., & Neubauer, P. (2020). Comparison of Perception of Spatial Localization Between Channel and Object Based Audio.
- Owsinski, B. (2017). *The Mixing Engineer’s Handbook 4th Edition*. Bobby Owsinski Media Group. <https://books.google.it/books?id=ks1DMQAACAAJ>
- Paisa, R. (n.d.). *Wavefield Synthesis for Max/MSP*.
- Pajunen, L., Politis, A., Vaalgamaa, M., Strömmer, S., & Pulkki, V. (2020). Effects of rigid spherical scatterer on spatial audio reproduction quality.
- Paul, S. (2009). Binaural recording technology: A historical review and possible future developments. *Acta Acustica United with Acustica*, 95, 767–788.
- Pike, C. W. (2019). *Evaluating the Perceived Quality of Binaural Technology* [Phd, University of York]. <https://etheses.whiterose.ac.uk/24022/>
- Pike, C. W. (n.d.). *Evaluating the Perceived Quality of Binaural Technology*. 605.
- Plenge, G. (1974). On the differences between localization and lateralization. *Journal of the Acoustical Society of America*, 56, 944–951.
- Poirier-Quinot, D., Hardoin, L., & Katz, B. F. (2020). Emily’s World: Behind the scenes of a binaural synthesis production. 5. *Proceedings of the 57th AES International Conference*. Hollywood, CA.
- Production Suite—MNTN | The Sound of the Mountain. (n.d.). MNTN | The Sound of the Mountain. Retrieved April 20, 2022, from <https://mntn.rocks/production-suite>



- Reardon, G., Roginska, A., Flanagan, P., Calle, J. S., Genovese, A., Zalles, G., Olko, M., & Jerez, C. (2017, October 8). Evaluation of Binaural Renderers: A Methodology. Audio Engineering Society Convention 143.  
<https://www.aes.org/e-lib/browse.cfm?elib=19310>
- Rentfrow, P. J., & Levitin, D. J. (Eds.). (2019). Foundations in Music Psychology: Theory and Research. MIT Press.
- Resonance Audio -. (n.d.). Retrieved April 20, 2022, from <https://resonance-audio.github.io/resonance-audio>
- Rikova, M. R., & Vermeir, G. (2009). Binaural Sound Source Localization in Real and Virtual Rooms. *J. Audio Eng. Soc.*, 57(4), 16.
- Roginska, A., & Geluso, P. (2017). Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio. Taylor & Francis.
- Rudrich, D. (n.d.). IEM Plug-in Suite. Retrieved April 20, 2022, from <https://plugins.iem.at/>
- Rumsey, F. (2001). Spatial Audio. Taylor & Francis.
- Rumsey, F. (2013). Upmix, downmix, shake it all about. *J. Audio Eng. Soc.*, 61(6), 5.
- Sacco, M. (2005). Imparare la tecnica del suono. LAMBDA s.r.l.
- Salvador, C. D., Sakamoto, S., Trevi, J., & Suzuki, Y. (2017). Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets. *Acoustical Science and Technology*, 38(2), 51–62.  
<https://doi.org/10.1250/ast.38.51>
- Schön, D., Akiva-Kabiri, L., & Vecchi, T. (2007). *Psicologia della musica*. Carocci.
- Shaw, E. A. G. (1974). The external ear. In W. D. Keidel & W. D. Neff (eds.), *Handbook of Sensory Physiology, Vol. 5/1, Auditory System* (pp. 455–490). New York: SpringerVerlag.

- Sheckells, M. (2019, January 11). Sony Unveils the ‘Future of Music’ With 360 Reality Audio at CES 2019. Billboard. <https://www.billboard.com/pro/sony-360-reality-audio-ces-2019-future-of-music/>
- Sheehan, M. (2020). Audio portraiture sound design and the development and creation of audio portraiture within immersive and binaural audio environments. 3.
- Sloma, U., Klein, F., Werner, S., & Kannookadan, T. P. (2019). Synthesis of binaural room impulse responses for different listening positions considering the source directivity. New York, 9.
- Smith, M. T. (2002). Manuale di ingegneria del suono. Hoepli.
- Snow, W. B. (1953). Basic Principles of Stereophonic Sound. Journal of the Society of Motion Picture and Television Engineers, 61(5), 567–589. <https://doi.org/10.5594/J00963>
- SOFA (Spatially Oriented Format for Acoustics)—Sofaconventions. (n.d.). Retrieved April 20, 2022, from [https://www.sofaconventions.org/mediawiki/index.php/SOFA\\_\(Spatially\\_Oriented\\_Format\\_for\\_Acoustics\)](https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics))
- Spat Revolution. (n.d.). FLUX: Immersive. Retrieved April 20, 2022, from <https://www.flux.audio/project/spat-revolution/>
- Spatial Audio Designer – Plug-ins | New Audio Technology. (n.d.). Retrieved April 20, 2022, from <https://newaudiotechnology.com/products/spatial-audio-designer/>
- Spatial Audio. (n.d.-a). Danca. Retrieved April 20, 2022, from <https://danca.tv/spatial-audio/>
- Spatial Audio. (n.d.-b). Abbey Road. Retrieved April 20, 2022, from <http://www.abbeyroad.com/spatial-audio>

- Steinberg, J. and Snow, W. (1934). Auditory perspectives– physical factors. In Stereophonic Techniques, Audio Engineering Society.
- Stereofonia nell'Enciclopedia Treccani. (n.d.). Retrieved April 20, 2022, from <https://www.treccani.it/enciclopedia/stereofonia>
- Subaqueous. (n.d.). Understanding LUFS Standards for Mastering. Subaqueous Music. Retrieved April 20, 2022, from <https://www.subaqueousmusic.com/mastering-for-streaming-online-lufs-standards>
- Tabry, V., Zatorre, R. J., & Voss, P. (2013). The influence of vision on sound localization abilities in both the horizontal and vertical planes. *Frontiers in Psychology*, 4, 932. doi: 10.3389/fpsyg.2013.00932
- Tan, C-J. and Gan, W-S. (2000). Direct concha excitation for the introduction of individualized hearing cues. *J. Audio Eng. Soc.*, 48, 7/8, pp.642–653.
- The “Mixing Secrets” Free Multitrack Download Library from Cambridge Music Technology. <https://www.cambridge-mt.com/ms/mtk/#ShortWaveRadioBand>
- The end of the Loudness Wars. (n.d.). AudioSEX - Professional Audio Forum. Retrieved April 20, 2022, from <https://audiosex.pro/threads/the-end-of-the-loudness-wars.31707/>
- TIDAL and Dolby are Bringing Dolby Atmos Music to TIDAL’S HiFi Members. (2019, December 12). Dolby Newsroom. <https://news.dolby.com/en-WW/184250-tidal-and-dolby-are-bringing-dolby-atmos-music-to-tidal-s-hifi-members>
- Titlow, J. P. (2017, June 5). Can Spatial 3D Audio Reinvent Live Music? *Fast Company*. <https://www.fastcompany.com/40425973/can-spatial-3d-audio-reinvent-live-music>

- Torick, E. (1998). Highlights in the history of multichannel sound. *Journal of the Audio Engineering Society*, 372, 368–372.
- Treccani. Psicoacustica in “Dizionario di Medicina.” (n.d.). Retrieved March 19, 2022, from [https://www.treccani.it/enciclopedia/psicoacustica\\_\(Dizionario-di-Medicina\)](https://www.treccani.it/enciclopedia/psicoacustica_(Dizionario-di-Medicina)).
- Tsingos, N., Govindaraju, P., Zhou, C., & Nadkarni, A. (2016). XY-stereo Capture and Up-conversion for Virtual Reality. Los Angeles, 8.
- Turner, K., & Pras, A. (2019). Is Binaural Spatialization the Future of Hip-Hop? New York, 9.
- Ueno, Y., Mizumachi, M., & Horiuchi, T. (2020). Perceptual evaluation of binaural rendering and stereo width control in headphone reproduction. 5.
- Vaccaro, M. (2015). Dolby Atmos: Il futuro del cinema.  
<https://docplayer.it/55381476-Dolby-atmos-il-futuro-del-suono-nel-cinema.html>
- Vintage Sony 1960’s EFM-117J radio | #367603127. (n.d.). Worthpoint. Retrieved March 26, 2022, from <https://www.worthpoint.com/worthopedia/vintage-sony-1960-efm-117j-radio-47603127>
- Voong, T. M., Reuter, C., & Oehler, M. (2020). Influence of individual HRTF preference on localization accuracy – a comparison between regular and bone conducting headphones. 9.
- Walton, T., & Evans, M. (2018). The role of human influence factors on overall listening experience. *Quality and User Experience*, 3(1), 1.  
<https://doi.org/10.1007/s41233-017-0015-4>
- Walton, T., & University, N. (n.d.). The Quality of Experience of Next Generation Audio: Exploring System, Context and Human Influence Factors. 223.

- Wei, Y., Lin, X., Jiang, K., Liu, P., Li, Q., & Fan, S. (2013). Thermoacoustic Chips with Carbon Nanotube Thin Yarn Arrays. *Nano Letters*, 13, 4795–4801.  
<https://doi.org/10.1021/nl402408j>
- Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening: II: Psychophysical validation. *Journal of the Acoustical Society of America*, 85, 868–878.
- Wightman, F. L., & Kistler, D. J. (1997). Monaural sound localization revisited. *Journal of the Acoustical Society of America*, 101(2), 1050–1063.
- Wightman, F. L., & Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America*, 105, 2841–2853.
- Wittek, H., & Theile, G. (2017). Development and application of a stereophonic multichannel recording technique for 3D Audio and VR. New York, 10.
- Zacharov, N., Pedersen, T., & Pike, C. (2016). A common lexicon for spatial sound quality assessment—Latest developments. 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), 1–6.  
<https://doi.org/10.1109/QoMEX.2016.7498967>
- Zotter, F., & Frank, M. (2019). *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-17207-7>