

## Lezione 1 - Oltre la stereofonia: tecnologie multicanale per la spazializzazione

### Cenni storici e configurazioni: l'audio nel cinema

#### I primi esperimenti

Nello studio delle tecnologie volte alla realizzazione di spazi d'ascolto che superino la stereofonia occorre distinguere due percorsi: il primo è quello svolto nell'industria cinematografica, mentre il secondo è quello che si è snodato, in modo più tortuoso, nell'ambiente della ricerca audio destinata alla musica. Vedremo in quali circostanze questi due percorsi abbiano avuto o potrebbero essere destinati ad avere un approdo comune. Tenendo presente le differenze peculiari dei due campi d'utilizzo, che si concretizzano nella presenza o meno delle immagini, iniziamo a tracciare un percorso storico dal cinema. È interessante notare come la spinta ad esplorare le tecniche di ripresa finalizzate alla stereofonia (malgrado all'epoca non esistesse ancora questo termine, ma si parlasse di "audio binaurale") venne nel 1931 ad Alan Dower Blumlein, l'inventore di queste tecniche, attraverso il cinema: egli riteneva inconcepibile che il dialogo non "seguisse" gli spostamenti degli attori sullo schermo, e che talvolta il suono di una voce venisse da un punto diverso rispetto all'immagine sullo schermo ("...I've got a way to make it follow the person")<sup>1</sup>. Possiamo affermare che il fattore "localizzazione" è stato il punto di partenza per tutte le ricerche finalizzate alla spazializzazione del suono, cui hanno fatto seguito altri fattori, come la definizione timbrica, la profondità, la spaziosità e, infine, il "realismo". Per tutto ciò si è dovuto aspettare circa 25 anni, dalle prime ricerche di Blumlein, perché le sue teorie trovassero un'applicazione pratica e commerciale. Nel frattempo, il cinema continuava nel suo cammino di sperimentazione.

Il film "Fantasia" del 1939 è il primo esempio di riproduzione in surround: La colonna sonora venne registrata con ben 44 microfoni su otto piste audio mediante registratori ottici, e da qui missata su tre tracce, alle quali si aggiunse una quarta traccia con funzione di "control-track", in grado di gestire la spazializzazione delle tre tracce audio verso un totale di sei punti di diffusione, tre frontali dietro lo schermo e tre alle spalle del pubblico, per un totale di 54 altoparlanti. L'intero sistema di diffusione venne battezzato "Fantasound", richiedeva una complessa installazione e messa a punto nella sala cinematografica dove veniva proiettata la pellicola, ed ebbe vita brevissima, a causa anche delle spese belliche a cui gli Stati Uniti dovettero andare incontro.

---

<sup>1</sup> R. C. Alexander: *The inventor of stereo: The life and works of Alan Dower Blumlein*; Focal Press

## Il Dolby Surround

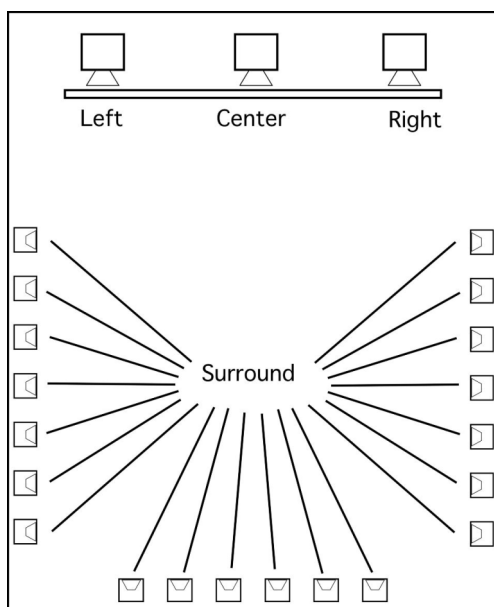


fig. 1: Il Dolby Surround nel cinema

Il passo successivo verso la spazializzazione audio nel cinema avviene circa 30 anni dopo con l'introduzione del sistema Dolby Surround (o "Dolby Stereo"), la cui configurazione di ascolto è raffigurata in fig. 1, consistente in tre canali frontali (left/center/right) e di un canale posteriore (surround). La particolarità del sistema consiste nel fatto che i quattro canali vengono codificati mediante una matrice analogica la cui uscita è un segnale stereo left/right (Lt/Rt), in modo da risultare compatibile con i sistemi di lettura stereofonici. Lo

schema della matrice di codifica (encoder) è riportato in fig. 2.

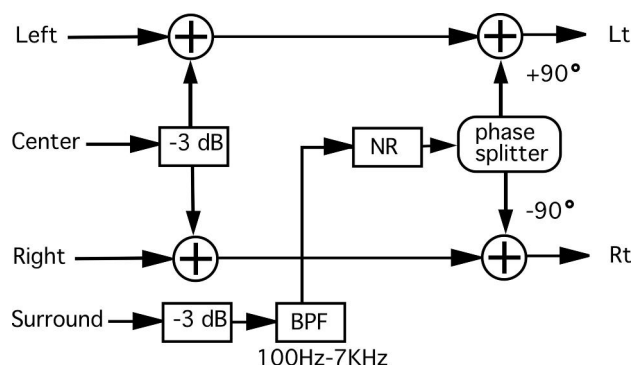


fig. 2: la matrice Dolby Surround

In presenza di un opportuno decoder, la matrice viene decodificata e distribuita alle quattro uscite. Per tale particolarità, questa tecnica è nota anche come 4-2-4, mentre il tipo di messaggio a quattro canali di uscita è noto come LCRS. La fig. 3 esemplifica il processo.

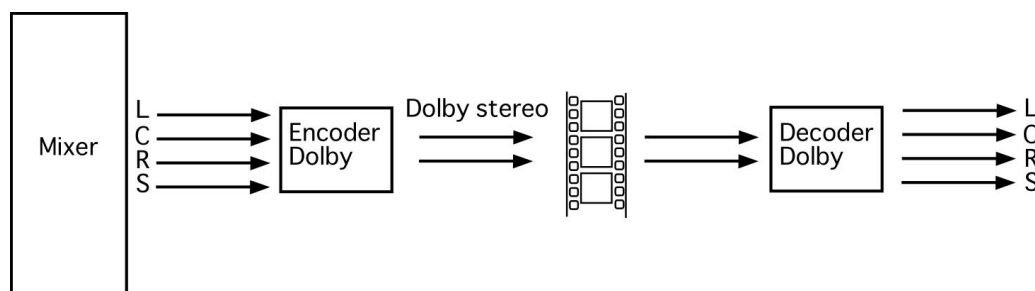


fig. 3: la catena Dolby

Questo sistema è tuttora usato nei cinema nelle tracce analogiche della colonna sonora delle pellicole ("ottico"), nonché in trasmissioni radio e televisive, nelle videocassette VHS, in alcune collane di CD, nei videogames, ecc. Quando verso la fine degli anni '80 il sistema è stato introdotto in queste numerose applicazioni "consumer" prese il nome di "Dolby Pro-Logic", con poche modifiche circuitali. Nello stesso periodo il sistema si è evoluto, in ambito cinematografico, nel "Dolby SR" (Dolby Spectral Recording), aggiungendo all'encoder un sistema di riduzione del rumore in grado di elevare la dinamica. Il processo conta una serie di limitazioni, tra le quali le più evidenti sono l'unicità del canale surround e la sua banda passante (100 Hz - 7Khz), ma è intuitivo, osservando quello che avviene nella matrice di codifica, che il contenuto dei canali stereo in ingresso può essere abbastanza stravolto, a misura della presenza e della quantità di suoni in fase e fuori fase: il decoder tenderà a interpretare i suoni fuori fase come appartenenti al surround, anche se questi sono il prodotto di altri processi, come la riverberazione, o come un tipo di ripresa stereofonica basata sulla differenza di fase (coppia omnidirezionale spaziata). È a causa di queste limitazioni che la Dolby prescrive per un corretto messaggio LCRS che il segnale da inviare ai monitor in regia sia preso all'uscita di una catena completa di encoder/decoder.

## Il Surround 5.1 e oltre

Al film *Star Wars*, verso la fine degli anni '70, spetta il primato dell'introduzione di un canale dedicato esclusivamente alle basse frequenze, apportando un enorme passo avanti nello sviluppo della dinamica del suono, in quanto rendeva disponibile agli altri canali una grande quantità di energia sonora. Altre pietre miliari sono stati i film *Superman* e *Apocalypse Now*, con il raddoppio dei canali surround. Successivamente, l'industria ha trovato più conveniente rivolgersi verso un sistema a canali discreti, migliore rispetto al sistema matrix usato fino ad allora, dando vita così al sistema 5.1, dove ai 5 canali (tre frontali, due surround), si aggiunge un canale di esaltazione delle basse frequenze (LFE, Low Frequency Enhancer), cui, essendo a banda ridotta (120 Hz), viene attribuita convenzionalmente valenza 0.1. Nella fig. 4 è esemplificata la configurazione.

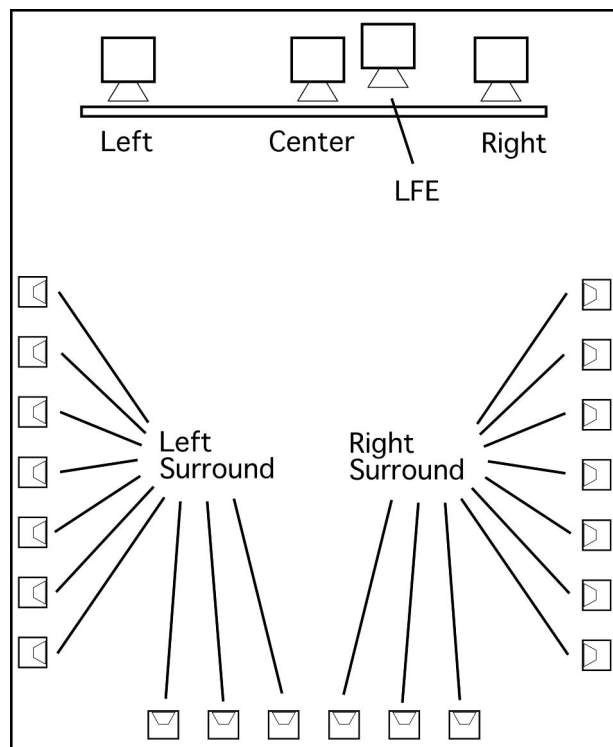


fig. 4: Il surround 5.1

Nel 1983, in occasione dell'uscita del film *Return of the Jedi* debutta per opera della casa di produzione *Lucasfilm* la certificazione THX, che prevede una serie di norme riguardanti la costruzione delle sale cinematografiche, tra cui: il rumore interno ed esterno, l'angolo di visione, la distanza dallo schermo, e una serie di parametri concernenti l'audio: distorsione, pressione sonora, dislocazione delle casse acustiche, ecc. Verso la fine degli anni '80 la transizione è completata con l'avvento del suono digitale, con una serie di standard, di cui i più significativi sono: Dolby Digital, DTS e Sony SDDS.

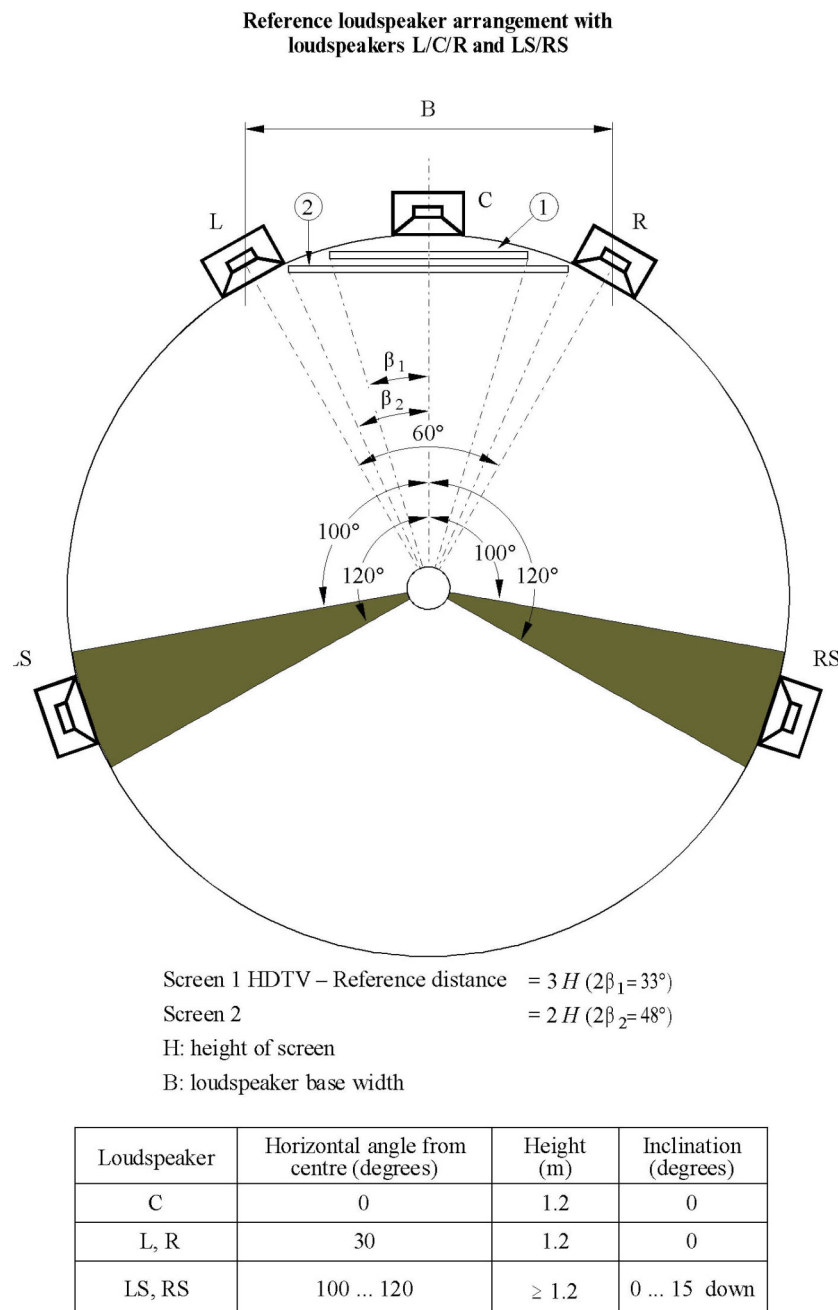
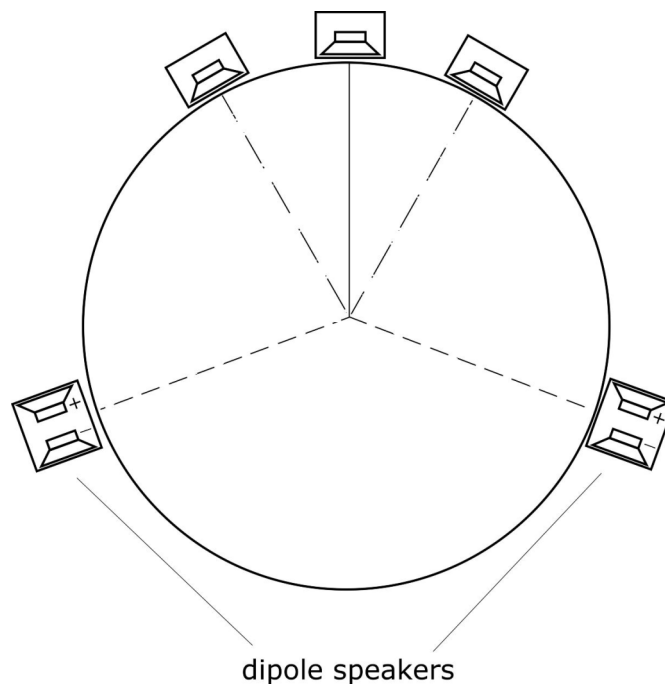


fig. 5: Lo standard per l'audio 5.1 nell'home-video (ITU R-BS 775)

Prima di esaminare in dettaglio le differenze tra questi formati, occorre parlare dell'enorme impatto sulla diffusione dell'audio multicanale generato

del capillare diffondersi dei sistemi home-theater: mentre con la tecnologia a nastro (VHS) l'esperienza d'ascolto nelle sale cinematografiche restava comunque una esperienza peculiare e molto diversa rispetto all'ascolto domestico, l'avvento di sistemi a costo relativamente contenuto in grado di portare nelle case il suono dei cinema, e di supporti innovativi (DVD) ha portato una crescente richiesta di audio multicanale, richiesta che in seguito dal cinema si è anche spostata verso la musica senza immagini. Una delle conseguenze è stata la necessità di proporre un sistema standard d'ascolto multicanale relativo all'home-video. La risposta sta nella pubblicazione della International Telecommunication Union del 1993 nota come Recommendation ITU-R BS 775 di cui vediamo in fig. 5 riprodotto il disegno originale.

Come si può osservare, alla tradizionale disposizione dei diffusori Left/Right a  $30^\circ$  ciascuno rispetto all'asse d'ascolto, si aggiunge il canale centrale e i due surround, con la particolarità che sono tutti posizionati ad eguale distanza dal punto di ascolto (detto anche "sweet spot"). I diffusori surround sono angolati di circa  $110^\circ$  rispetto all'asse (fra  $100^\circ$  e  $120^\circ$ ), sebbene esista una configurazione alternativa che prevede l'uso di quattro diffusori per i due canali surround, con angolazioni intorno a  $100^\circ$  e  $135^\circ$ , per aumentare la sensazione di campo diffuso. Esiste a questo proposito una variante di questo schema, in cui il campo sonoro del surround diffuso viene reso mediante diffusori dipolari, come illustrato in fig. 6.



*fig. 6: surround con diffusori dipolari*

Da notare infine, nella tabella, l'indicazione dell'altezza dei diffusori: mt 1,2 (corrispondente approssimativamente all'altezza delle orecchie di una persona seduta) per i tre frontali, stessa altezza o superiore per i surround, con possibilità di inclinare questi ultimi verso il basso. Questo standard, che come si può osservare stabilisce anche criteri per le dimensioni e la posizione

dello schermo, è rivolto principalmente alle regie audio dedicate alla post-produzione e al missaggio per l'home-video, ma è naturalmente anche un suggerimento per l'ascolto ottimale in ambiente domestico.

### **Gli standard digitali multicanale**

Con l'avvento del suono digitale, e con l'adozione universale dell'audio multicanale, è sorto un problema di come alloggiare i dati digitali sulla pellicola e alla velocità di trasferimento ("transfer-rate") dei dati stessi in lettura: sei canali campionati a 48 KHz alla risoluzione di 20 bit richiedono un transfer-rate di 5,76 Mbits/sec, che, considerando la velocità di scorrimento della pellicola di 24 fotogrammi/sec significa 240.000 bit da alloggiare in ogni fotogramma, cosa fisicamente impossibile. Sono sorti allora una serie di standard digitali, tutti basati sul principio della compressione dei dati, ognuno alla ricerca del miglior compromesso tra la qualità e la velocità dei dati.

Vediamo quali sono state le risposte tecnologiche a questa sfida esaminando più in dettaglio i formati più significativi:

#### **1. Dolby Digital (AC-3)**

Questo sistema, adottato principalmente nel cinema e nei supporti DVD video (DVD-V) basa il suo principio di funzionamento su sofisticati algoritmi basati sulla psicoacustica e sul mascheramento uditivo, riuscendo a limitare la quantità di informazione da memorizzare al tipo di informazione più significativo all'ascolto umano. Il risultato è un singolo flusso di dati (bitstream) con un bit-rate di circa 320-384 Kbit/sec, contenente tutte le informazioni di un audio 5.1, che fa di questo formato il più efficiente, soprattutto in relazione allo spazio occupato nei DVD.

#### **2. DTS**

Il sistema della Digital Theater Systems (DTS) data il suo esordio nelle sale nel 1993 con il film *Jurassic Park*, e presenta rispetto al Dolby Digital alcune evoluzioni, che riguardano la variabilità del numero dei canali, della frequenza di campionamento (fino a 192 KHz) e anche della risoluzione dinamica, che lo rendono per diversi aspetti superiore. Ma la caratteristica forse più significativa è la particolarità, per il cinema, di avere l'audio non sulla pellicola, ma su un CD-ROM collegato alla pellicola tramite una traccia di time-code. Anche il DTS è adoperato nei DVD-V, sebbene con algoritmi differenti, oltre che in un piccolo catalogo di CD audio rimasterizzati con questo sistema. L'algoritmo di codifica del DTS (noto come "Coherent Acoustics") può operare su una varietà di bit-rate, di frequenze di campionamento e di risoluzione, a seconda della banda di frequenza in cui viene suddiviso lo spettro sonoro. Questa particolarità consente al sistema di ottimizzare i parametri di campionamento e, naturalmente, il bit-rate, che può variare da un minimo di 754 Kbit/sec a un massimo di 1.5 Mbit/sec.

#### **3. SDDS**

Il Sony Dynamic Digital Sound si basa anch'esso su algoritmi di compressione dei dati fondati su principi di psicoacustica e in parte ripresi dalla Dolby, in particolare sull'Adaptive Transform Acoustic Coding (ATRAC), sul quale la Sony ha basato la tecnologia mini-disk. È un formato che nei cinema si avvale di un sistema di diffusione 7.1, illustrato in fig. 7, ed è pensato per sale cinematografiche dotate di uno schermo di grandi dimensioni. Con 8 canali campionati a 44.1 KHz e 20 bit (LFE a parte), è in grado di offrire un bit-rate tra 1060 e 1411 Kbit/sec.

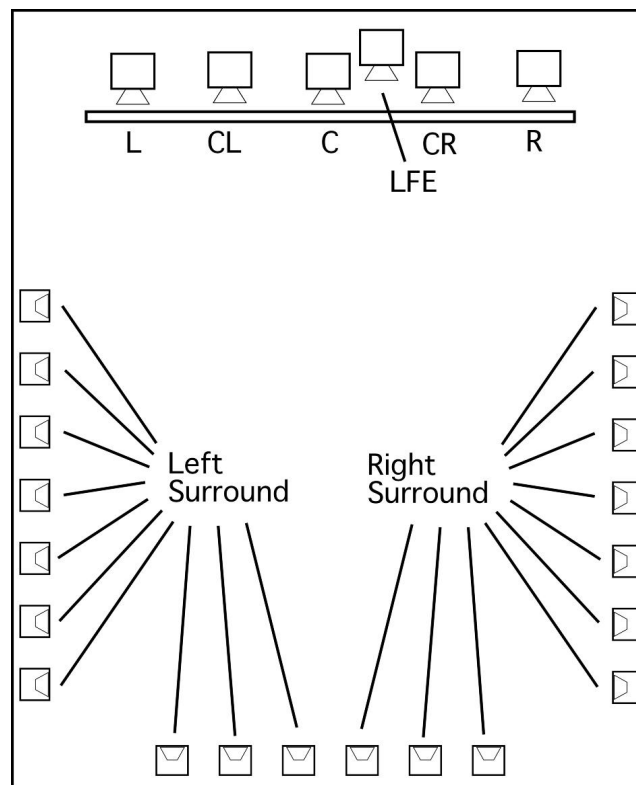


fig. 7: il surround 7.1

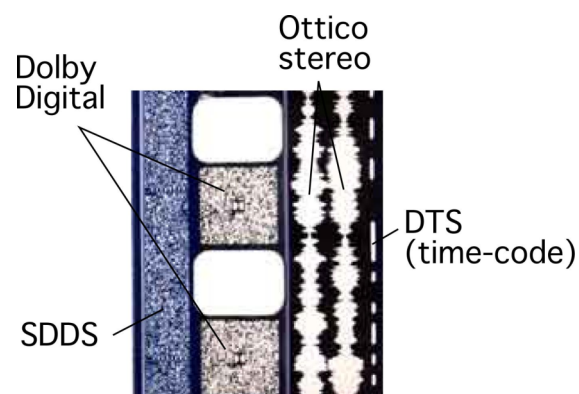
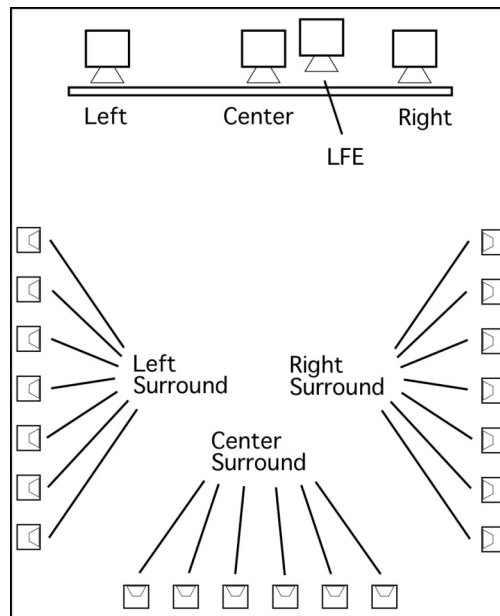


fig. 8: le tracce audio su pellicola

Come si può osservare nella fig. 8, la tecnologia ha permesso di sfruttare tutte le zone della pellicola cinematografica non occupate dall'immagine e dai fori

necessari al trascinamento per ospitare l'audio nei vari standard, per cui ogni copia di film può essere consegnata in una grande varietà di sale. Le macchine di proiezione delle sale cinematografiche sono dotate di un sistema di emergenza in virtù del quale, nel momento in cui venga a mancare l'informazione da uno dei formati digitali, passano automaticamente la lettura sulla traccia ottica stereo, che deve essere sempre presente.



*fig. 9: il surround 6.1*

Esiste infine lo standard 6.1, supportato sia dalla Dolby (Dolby EX) che dalla DTS (DTS ES), rappresentato in fig. 9, che prevede un canale centrale posteriore, allo scopo di migliorare la localizzazione dei suoni alle spalle del pubblico.



## La percezione del suono nell'audio multicanale

Occorre a questo punto fare qualche considerazione sulle problematiche relative alla percezione acustica dell'audio multicanale, riprendendo il discorso iniziato nella lezione dedicata alle coppie stereofoniche, in cui avevamo evidenziato come la localizzazione della direzione di provenienza del suono avvenisse attraverso l'elaborazione effettuata dal nostro cervello delle informazioni relative alla differenza di tempo e alla differenza di intensità (ITD e IID). Questa asserzione, in realtà, è valida solo per i suoni di provenienza frontale.

### Il cono di confusione e le HRTF

Se si pensasse che le due grandezze di cui abbiamo parlato, le ITD e le IID, siano sufficienti per l'individuazione precisa della provenienza di un suono in tutte le dimensioni dello spazio, saremmo in errore. Infatti, come si può vedere dalla fig. 10, esiste una zona di provenienza del suono, che giace sulla superficie di un cono immaginario con vertice nell'orecchio, dove le ITD e le IID sono assolutamente equivalenti, sebbene la posizione fisica nello spazio di queste sorgenti sia sostanzialmente diversa.

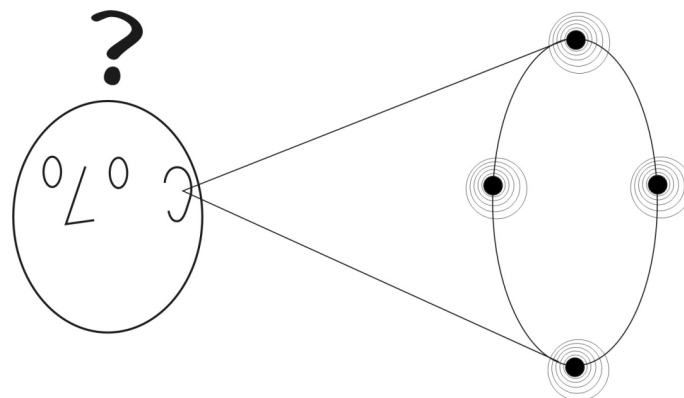


fig. 10: il cono di confusione

A causa della possibilità d'errore nella percezione della direzione del suono, questo fenomeno prende il nome di "cono di confusione" (Cone of confusion). Naturalmente, è importante sottolineare che, per quel che riguarda la registrazione e più in generale la riproduzione elettronica dei suoni, questo problema è irrilevante finché si rimane nell'ambito della stereofonia, poiché nella stereofonia vengono trattati esclusivamente i suoni "davanti" all'ascoltatore, ossia quei suoni che contribuiscono alla formazione di un fronte sonoro. Il problema sorge nel momento in cui si prenda in considerazione una spazializzazione dei suoni "a 360 gradi", come nel caso delle diffusioni dove l'ascoltatore è previsto sia immerso in un sistema di ascolto a quattro e più canali. Infatti, la capacità del cervello di creare un'immagine sonora virtuale situata in un certo punto collocato tra l'altoparlante sinistro e quello destro decade completamente nel momento in

cui si voglia applicare lo stesso criterio, ad es. in un sistema quadrifonico, tra l'altoparlante sinistro-frontale e l'altoparlante sinistro-posteriore.

La presenza del cono di confusione ci spinge a fare qualche considerazione sulla differenza di percezione dei suoni avanti/dietro e in elevazione verticale. Il nostro apparato di percezione uditiva è stato progettato, e si è affinato in anni d'evoluzione, come un sistema atto a lavorare su un piano essenzialmente orizzontale-frontale (sinistra/destra), ma è comunque, anche se in modo meno raffinato, capace di distinguere un suono di provenienza frontale da uno di provenienza posteriore, come pure un suono che venga dall'alto o dal basso. Questo è reso possibile sia dalla complessità dell'orecchio esterno, la cui conformazione crea una serie di piccole linee di ritardo e quindi di comb-filter che danno la possibilità di decodificare queste informazioni, sia dalla capacità, anche con un movimento impercettibile della testa, di mettere in relazione le informazioni ricevute nel tempo, e quindi di ricostruire un'immagine virtuale della collocazione spaziale della fonte sonora.

L'insieme delle informazioni ricavate dalle ITD, dalle IID e dalla conformazione dell'orecchio esterno ("pinna"), della testa e del busto ("torso") generano delle funzioni dette HRTF (head related transfer functions), le quali descrivono le differenze tra le due orecchie nell'elaborazione e nelle trasformazioni dei suoni che avvengono prima dell'ingresso del suono nel canale uditivo. Tali funzioni non hanno un valore assoluto, bensì statistico, in quanto ogni individuo ha le "sue" HRTF, dovute alla sua propria conformazione fisica.

### Percezione timbrica e localizzazione

È interessante anche osservare alcuni fenomeni connessi alla differenza di percezione timbrica tra i canali frontali e posteriori.

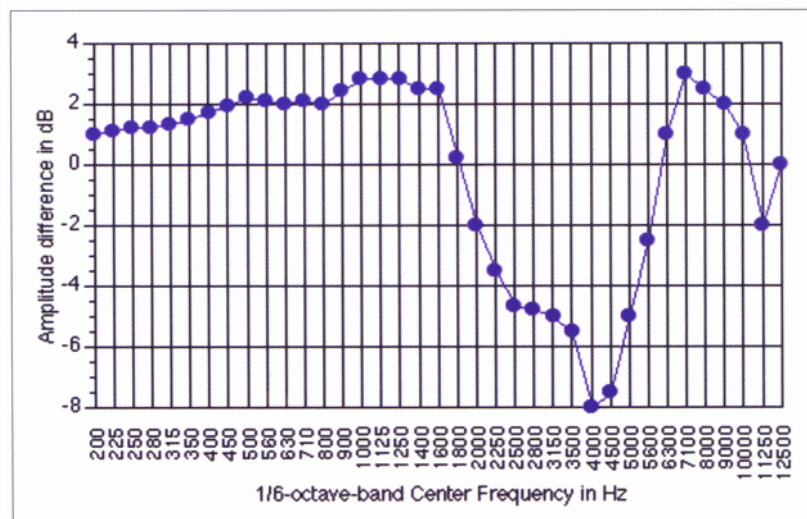


fig. 11: risposta percettiva sui suoni posteriori

Se osserviamo il grafico di fig. 11, che rappresenta la differenza di percezione tra suoni frontali (generati da un altoparlante a 30° dall'asse) e suoni posteriori

(generati da un altoparlante a 120° dall'asse) misurata nel canale uditivo, vediamo come tutta la zona di frequenze medio-alte, tra 2.000 e 5.000 Hz, la zona cioè dove la percezione uditiva è massima per l'identificazione timbrica, presenta una sensibile flessione nella curva di risposta<sup>2</sup>.

Si dovrebbe concludere che il timbro di uno strumento varia nella nostra percezione a seconda che sia suonato davanti a noi o dietro a noi. Ma, nella realtà, se un suonatore suona il suo strumento girando intorno ad un ascoltatore, la percezione timbrica dello strumento rimane invariata, e questo a causa delle HRTF che compensano gli spostamenti e le differenze timbriche. Lo stesso meccanismo non si ripete nell'audio spazializzato, ad es., con un sistema 5.1, ed il grafico illustrato sopra ritorna nella sua validità. Il fenomeno è dovuto principalmente alla scarsità di fonti sonore per il surround (che spiega anche le tendenze recenti all'incremento del numero di queste fonti), e alla conseguente necessità di ricorrere alle immagini sonore virtuali (phantom images) per ricostruire la provenienza del suono negli spazi intermedi tra gli altoparlanti.

### **Phantom images**

L'udito umano è in grado, ascoltando un suono in uscita da una coppia di altoparlanti, di costruire delle "immagini sonore virtuali" (phantom images) collocate nello spazio che separa i due altoparlanti, nel momento in cui il suono sia stato ripreso con una coppia stereofonica oppure sia stato regolato tramite il potenziometro pan-pot in fase di missaggio. Questo fenomeno, mentre è valido e facilmente riscontrabile nel caso di una coppia di altoparlanti posta di fronte all'ascoltatore, non è più valido nel momento in cui la coppia sia posta lateralmente al punto di ascolto. Osserviamo la fig. 12, che rappresenta il risultato sperimentale sulla percezione effettuato nel 1975 dalla BBC: il grafico rappresenta il risultato angolare percettivo in funzione della differenza di segnale di quattro canali disposti secondo la configurazione quadrifonica classica. I cerchi concentrici rappresentano la differenza di livello in dB; il grafico "ad ali di farfalla" rappresenta il risultato percettivo, ossia la direzione in cui si forma l'immagine sonora virtuale, rispetto al punto di ascolto. Per livelli uguali sui due canali frontali l'immagine compare nel centro esatto, per una differenza di -10 dB il suono viene percepito a circa 22° dall'asse, mentre a partire da una differenza di -25 dB la percezione punta direttamente verso il suono più forte. La stessa cosa avviene tra il canale sinistro e il destro sui canali posteriori, mentre lateralmente il risultato è decisamente differente: per livelli uguali l'immagine virtuale, anziché collocarsi a 90°, si colloca a circa 65°, ed esiste una zona tra -10 e -12 dB, dove il suono tende a "saltare" da un altoparlante all'altro, ossia ad apparire provenire maggiormente dall'uno o dall'altro degli altoparlanti a seconda di impercettibili movimenti della testa.

---

<sup>2</sup> E. A. G. Shaw: *Transformation of sound pressure level from the free field to the eardrum in the horizontal plane*, AES Journal vol. 56 n. 6

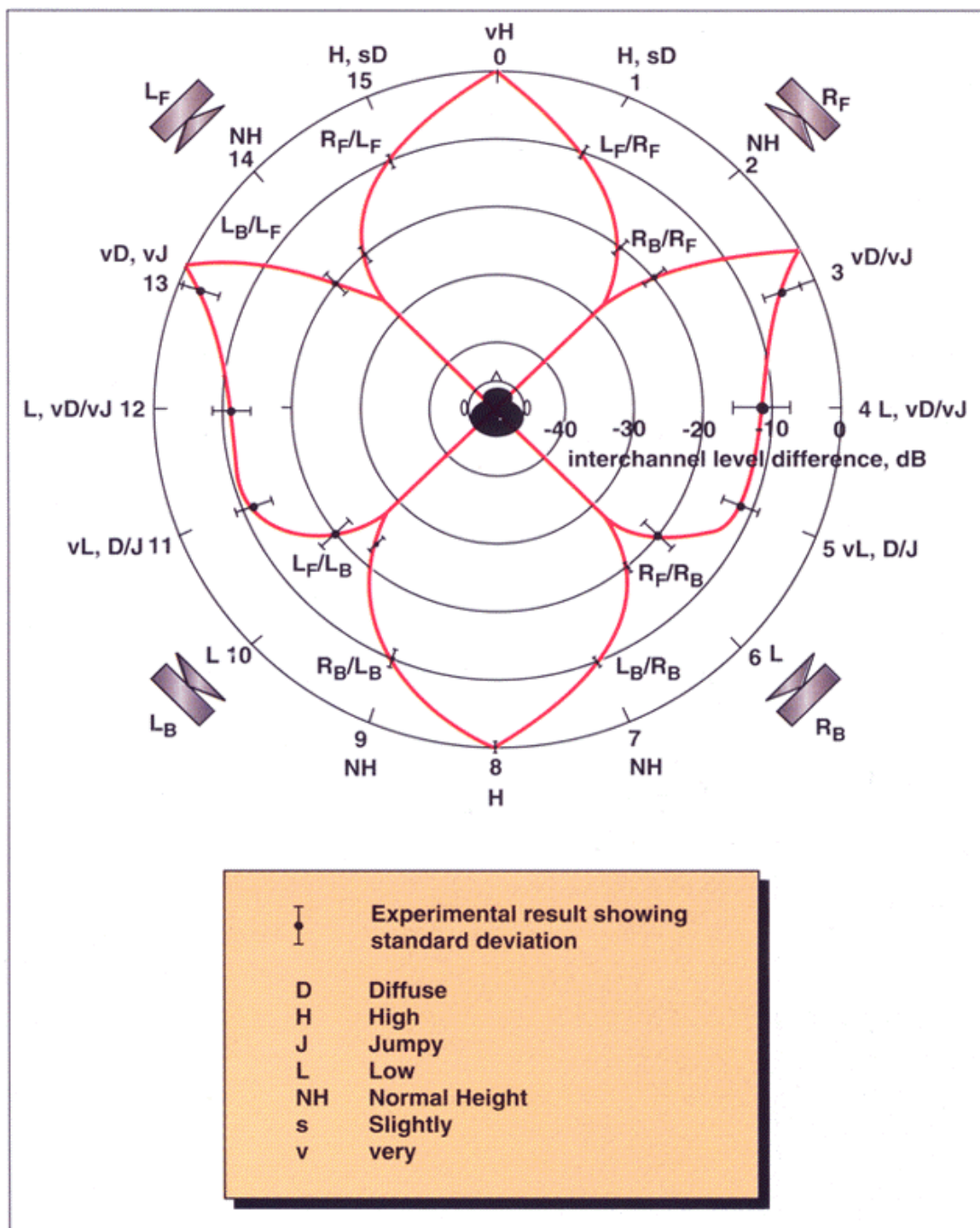


fig. 12: studio delle immagini sonore virtuali in quadrafonia (BBC, 1975)

## Il canale LFE

Da tutto quello che è stato esposto fin qui, in particolare riguardo alle configurazioni surround, si può notare quanta poca attenzione sia stata data al collocamento fisico dell'altoparlante dedicato alla gestione del canale LFE, ossia del canale relativo alle frequenze più gravi. Nelle configurazioni per il cinema generalmente il sub-woofer è sistemato "da qualche parte" nelle vicinanze dell'altoparlante frontale centrale; nella Recommendation ITU-R

BS775 relativa al 5.1 per l'home-theater non è data una chiara indicazione della migliore collocazione del sub-woofer. Esistono considerazioni di carattere psicoacustico di cui tener conto per spiegare i motivi di questa indeterminazione. È stato sperimentalmente dimostrato come l'udito umano abbia il suo massimo di percezione, riguardo alla direzionalità, nella regione delle frequenze "medio-alte" intorno ai 2.000-3.000 Hz, e come la percezione della direzionalità sia molto meno affinata man mano che i suoni scendono nelle regioni dei gravi, fino a diventare nulla al di sotto di 80-100 Hz. Un modo intuitivo per spiegare questo fenomeno è la considerazione che le frequenze gravi, per diffrazione, hanno molta più facilità a "girare" intorno alla testa: un suono alla frequenza di 20 Hz vibra con una lunghezza d'onda di 17 metri, contro i 17 millimetri di una frequenza di 20.000 Hz, e considerando le dimensioni fisiche di una testa appare evidente come questa sia considerata come ostacolo in modo sostanzialmente diverso a seconda della frequenza.