

Lezione 11 - La spazializzazione

La storia degli esperimenti di spazializzazione dell'audio senza immagini è una storia di pura ricerca, condotta al di fuori dell'industria cinematografica, senza le possibilità economiche di cui questa poteva disporre a fronte di concrete possibilità di rientro degli investimenti. Esistono due percorsi paralleli, a volte collegati, a volte indipendenti, che riguardano rispettivamente la ripresa e l'ascolto dell'audio spazializzato.

La quadrifonia

Verso la fine degli anni '60 si iniziò a pensare ad una diffusione sonora che andasse oltre la stereofonia e che prevedesse una diffusione del suono proveniente alle spalle dell'ascoltatore. La configurazione iniziale è quella che si può osservare in fig. 1, con i quattro diffusori posti agli angoli di un quadrato.

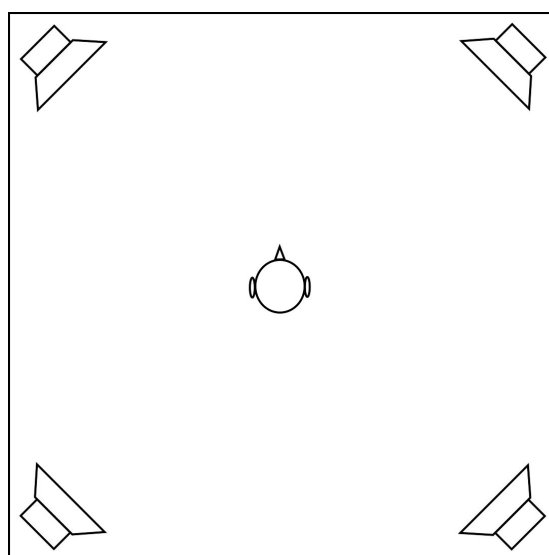


fig. 1: lo schema classico della quadrifonia

L'impulso a questi studi ebbe una natura prevalentemente commerciale, in quanto la rapidità con cui avvenne il passaggio dall'ascolto monofonico a quello stereofonico dieci anni prima indusse molti a pensare che il passaggio successivo sarebbe stato altrettanto rapido, e che quindi occorresse introdurre nuovi standard per l'ascolto domestico e per la pubblicazione di realizzazioni discografiche. Furono introdotti numerosi sistemi di codifica in grado di produrre registrazioni che potessero essere incise su dischi al vinile e potessero fornire una buona compatibilità stereo, tra i quali possiamo citare il sistema SQ (Stereo Quadraphonic) introdotto dalla casa discografica CBS, che si avvaleva di un sistema a matrice espresso dalle formule:

$$L_t = LF + 0.707 RB - j0.707 LB$$

$$R_t = RF - 0.707 LB + j0.707 RB$$

dove:

LF, RF, LB e RB = i quattro segnali quadrifonici

$+j$ = rotazione di fase $+90^\circ$

$-j$ = rotazione di fase -90°

Questo sistema ebbe una diffusione maggiore rispetto ai concorrenti in quanto aveva alle spalle una casa discografica in grado di fornire un catalogo di registrazioni effettuato con tecnica quadrifonica, ma esso comunque ebbe a competere con numerosi concorrenti, i quali si avvalevano sia di sistemi a matrice simili all'SQ, sia anche di sistemi a modulazione, come il CD-4 (Compatible Discrete 4) della JVC (vedi fig. 2), che sfruttando una frequenza portante per differenziare i canali posteriori dagli anteriori miravano a superare il limite fisico dei due canali contenuti nel solco di un disco in vinile.



fig. 2: alcuni standard quadrifonici

I problemi incontrati dalla quadrifonia furono di diverso tipo:

- a) alla consueta incapacità di creazione di immagini virtuali sonore ai lati e alle spalle dell'ascoltatore si aggiunse la problematica creazione di immagini virtuali anche sul fronte, dovuta all'angolazione dei diffusori di 45° rispetto all'asse;
- b) l'eccessiva quantità di standard ed il costo delle apparecchiature aggiuntive necessarie alla decodifica scoraggiò i potenziali fruitori delle registrazioni quadrifoniche e di conseguenza la produzione stessa di tali registrazioni;
- c) la mancanza di una ricerca avanzata in merito alla ripresa di audio spazializzato, almeno fino all'entrata in scena di Michael Gerzon.

Michael Gerzon

A cavallo tra gli anni '60 e gli anni '70, in coincidenza col dibattito venutosi a creare intorno alla quadrifonia, una serie di studi, di cui fu pioniere principalmente Michael Gerzon¹, furono condotti in Inghilterra allo scopo di stabilire una serie di principi, concretizzatisi in seguito in una serie di applicazioni differenti, aventi come obiettivo una riproduzione del campo sonoro quanto più possibile vicino alla realtà percettiva umana. È interessante

¹ Michael Gerzon (1945 - 1996) ricercatore dell'Oxford Mathematical Institute.

notare che le teorie sviluppate da Gerzon e che culminarono con l'Ambisonic prendono il via dal punto in cui si era fermato Alan Blumlein negli anni '30.

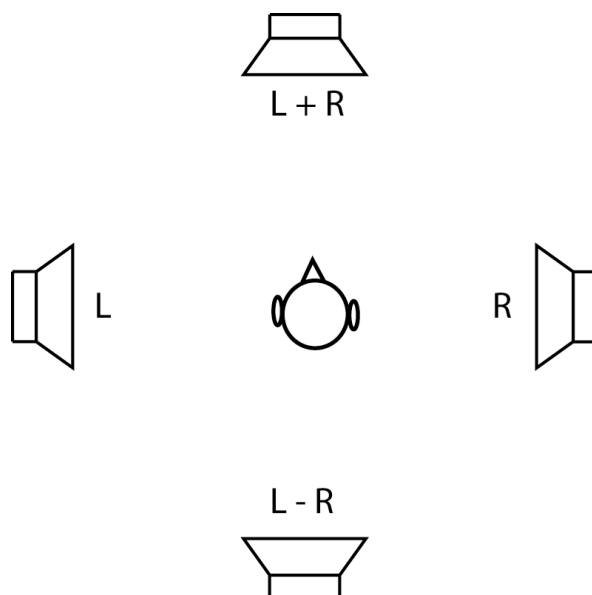


fig. 3: esperimento di spazializzazione condotto da Michael Gerzon

I primi esperimenti condotti da Gerzon furono rivolti all'ascolto spazializzato di audio registrato in stereofonia, con l'intento di estrarre dal materiale sonoro le informazioni mancanti alla ricostruzione di un campo sonoro completo sul piano orizzontale. La fig. 3 rappresenta uno degli esperimenti da lui condotti, con una inusuale disposizione di quattro punti di diffusione.

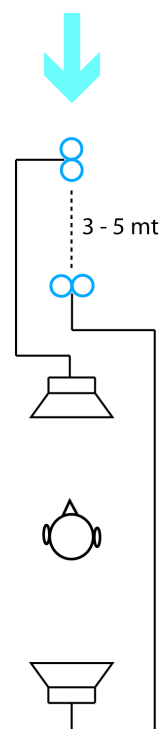
Un altro esperimento condotto da Gerzon (vedi fig. 4), la cosiddetta "Ambiophonic Technique"², riguarda invece un tentativo di ripresa audio con l'intento di fornire una sensazione di audio spazializzato attraverso un campo sonoro posteriore diffuso.

fig. 4: Ambiophonic Technique

Questi esperimenti, per quanto interessanti ed indicativi della volontà di superare la stereofonia per arrivare alla codificazione e alla realizzazione di sistemi di ripresa e di ascolto spazializzato, in realtà rivestono più che altro un ruolo di curiosità storica. Il vero apporto di Gerzon alla costruzione di un audio spazializzato è costituito dalla teoria Ambisonic.

Ambisonic

Il campo teorico dell'Ambisonic riguarda la possibilità di registrare e riprodurre informazioni sonore provenienti da tutte le direzioni dello spazio, compresa l'elevazione verticale. A



² Il termine Ambiophonic verrà in seguito sfruttato per le ricerche di Ralph Gasgal (vedi più oltre)

seconda del numero di altoparlanti impiegati, che, al contrario della quadrifonia, non è di una quantità prefissata, è possibile rappresentare un numero più o meno grande di dimensioni spaziali sonore con diversi gradi di definizione.

Per una comprensione profonda delle regole matematiche alla base della teoria Ambisonic occorre tornare alla definizione di alcune curve polari inserendo per ciascuna di esse la corrispondente equazione trigonometrica (vedi fig. 5).

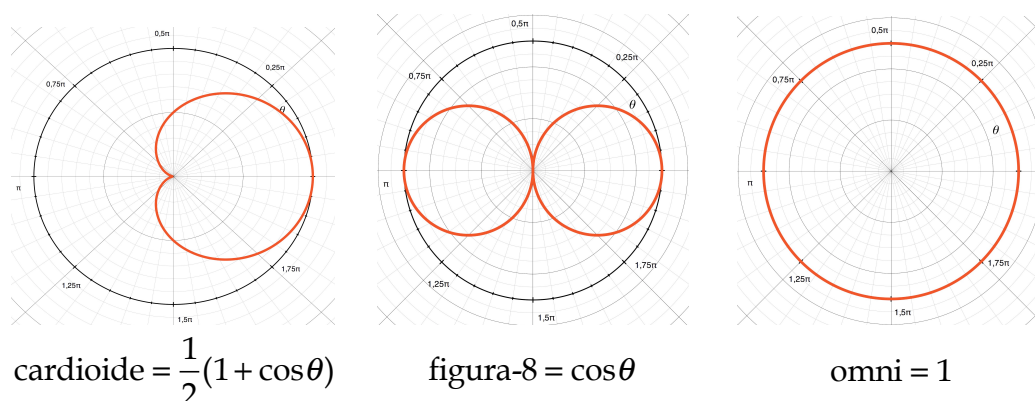


fig. 5: le equazioni delle curve polari

Dalle equazioni evidenziate in figura risulta evidente come la curva cardioide si debba considerare come un derivato delle altre due, che rappresentano le due fondamentali famiglie di microfoni:

- 1) omnidirezionali (pressure microphones)
- 2) direzionali (pressure gradient o velocity microphones)

L'Ambisonic stabilisce che il campo sonoro tridimensionale, rappresentato in un diagramma cartesiano, può essere descritto mediante una serie di componenti che prendono il nome di "armoniche sferiche" (spherical harmonics). La quantità minima necessaria alla descrizione di un campo sonoro tridimensionale è costituita da quattro componenti e precisamente da tre componenti direzionali (velocity) ed una componente omnidirezionale (pressure):

$$W = \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}} = 0.707 = \text{omnidirezionale (pressure)}$$

$$X = \cos(\theta)\cos(\varphi) = \text{direzione avanti-dietro (velocity front-back)}$$

$$Y = \sin(\theta)\cos(\varphi) = \text{direzione sinistra-destra (velocity left-right)}$$

$$Z = \sin(\varphi) = \text{direzione alto-basso (velocity up-down)}$$

dove:

θ = angolo di scostamento orizzontale

φ = angolo di scostamento verticale (elevazione)

Nella rappresentazione polare le componenti X, Y e Z corrispondono alla risposta di tre microfoni figura-8 perfettamente coincidenti ed orientati sulle direttrici dello spazio tridimensionale, più un segnale W rappresentante una componente omnidirezionale sferica, come illustrato in fig. 6. Il segnale W avrà per convenzione un'attenuazione di -3 dB (quantificata nel coefficiente 0.707) in confronto alle altre tre componenti per compensare il livello totale nel momento in cui questo segnale si combina con gli altri al fine di ottenere dei pattern polari e mantenere la corretta informazione direzionale. All'interno del sistema Ambisonic l'insieme di queste quattro componenti, per ragioni che vedremo più oltre, prende il nome di "B-format del primo ordine".

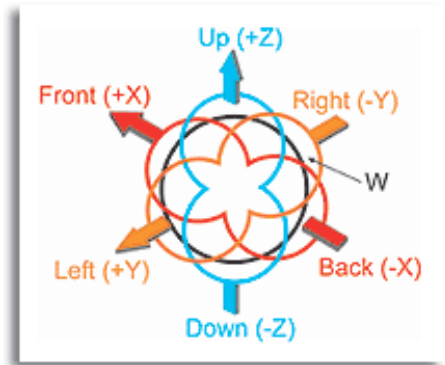


fig. 6: rappresentazione tridimensionale del B-format

Una delle caratteristiche che fanno dell'Ambisonic un sistema estremamente potente ed aperto è che dai segnali originati con l'applicazione della matrice B-format è possibile estrarre le curve polari corrispondenti a qualsiasi combinazione di coppia coincidente. Così una configurazione Mid-Side può essere ricavata dalle formule:

$$M = \sqrt{2}W + X$$

$$S = Y$$

Mentre una coppia coincidente Blumlein da queste altre:

$$L = \frac{X + Y}{\sqrt{2}}$$

$$R = \frac{X - Y}{\sqrt{2}}$$

Formati Ambisonic

Il sistema Ambisonic si compone in realtà di una serie di formati, che riguardano lo stadio d'elaborazione dei segnali in una catena di registrazione e/o riproduzione. Qui di seguito riassumiamo i principali:

- 1) A-format
- 2) B-format
- 3) C-format
- 4) D-format
- 5) G-format.

L'A-format consiste nel segnale d'uscita di un microfono, come quello illustrato in fig. 7, costituito da quattro capsule montate in modo estremamente ravvicinato (idealmente coincidente) ed orientate nello spazio come le facce di un tetraedro, e denominate: left-front-up (LFU), right-front-down (RFD), left-back-down (LBD) e right-back-up (RBU), tenendo presente che tale orientamento non è su di un piano lineare, ma che ognuna di queste coppie ha una divergenza sul piano verticale.



fig. 7: microfono tetraedrico (A-format)

Il B-format precedentemente descritto è una prima elaborazione del segnale A-format ottenuta mediante una matrice di trasformazione. La matrice per il passaggio dall'A-format al B-format è data da:

$$W = 0.5(LFU + LBD + RFD + RBU)$$

$$X = 0.5(LFU - LBD + RFD - RBU)$$

$$Y = 0.5(LFU - RBU - RFD + LBD)$$

$$Z = 0.5(LFU - LBD + RBU - RFD)$$

Questa matrice di trasferimento è in realtà solo teorica, in quanto occorre tenere presente che la non-coincidenza fisica della quattro capsule provoca un errore nella risposta in frequenza nel momento in cui i quattro segnali si combinano nella matrice, in particolare nelle frequenze la cui lunghezza d'onda sia inferiore alla devianza di ogni capsula dal centro virtuale dell'array. A titolo esemplificativo, in fig. 8 è riportata la risposta in frequenza rilevata dallo stesso Gerzon³ per un array microfonico tetraedrico con devianza pari a 1,47 cm. La risposta in frequenza relativa a "figure of eight" si intende comune ad ognuna delle tre componenti X, Y e Z. Risulta evidente come tali errori possano indurre imprecise ricostruzioni del campo sonoro.

³ M. Gerzon: The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound - AES 50th Convention - Londra 1975

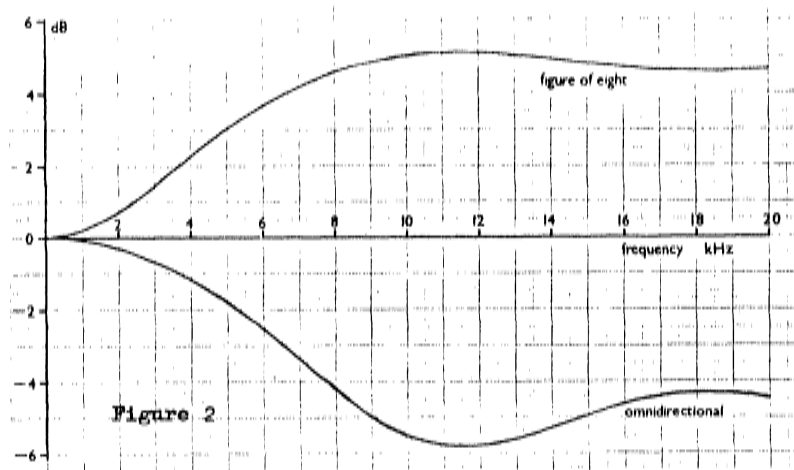


fig. 8: errori di risposta in frequenza nel microfono tetraedrico

La matrice reale di trasferimento dovrà quindi tenere conto della geometria reale delle capsule formanti il microfono tetraedrico, in quanto i segnali risultanti dovranno subire una compensazione di frequenza e di fase:

$$X = F_1(\omega)(LFU - LBD + RFD - RBU)$$

$$Y = F_1(\omega)(LFU - RBU - RFD + LBD)$$

$$Z = F_1(\omega)(LFU - LBD + RBU - RFD)$$

$$W = F_0(\omega)(LFU + LBD + RFD + RBU)$$

dove:

$F_1(\omega)$ = compensazione in equalizzazione e in fase per le componenti X, Y e Z

$F_0(\omega)$ = compensazione in equalizzazione e in fase per la componente W

Da notare che, se nel prosieguo del trattamento di questi segnali la componente verticale non dovesse interessare, il parametro Z può semplicemente essere omissso.

fig. 9: array Ambisonic orizzontale



Parimenti è evidente che ove non interessasse la componente di elevazione il microfono tetraedrico può essere egregiamente sostituito da un array coincidente composto da un microfono omnidirezionale e due microfoni figura-8 che forniscono le componenti X e Y, come schematizzato in fig. 9.

Ambisonic di ordine superiore (High Order Ambisonics - HOA)

Il sistema Ambisonic, per come l'abbiamo fin qui esaminato, appartiene al 1° ordine, ma le armoniche sferiche di un campo sonoro nello spazio, e quindi le componenti direzionali ad esse associate, possono in realtà essere descritte in

modo ancora più dettagliato, aumentandone il numero. Un sistema Ambisonic del 2° ordine è costituito da nove componenti, e precisamente dalle quattro componenti del 1° ordine, più altre cinque, e precisamente R , S , T , U e V :

$$W = \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}} = 0.707$$

$$X = \cos(\theta)\cos(\varphi)$$

$$Y = \sin(\theta)\cos(\varphi)$$

$$Z = \sin(\varphi)$$

$$R = 1.5\sin^2(\varphi) - 0.5$$

$$S = \cos(\theta)\sin(2\varphi)$$

$$T = \sin(\theta)\sin(2\varphi)$$

$$U = \cos(2\theta)\cos^2(\varphi)$$

$$V = \sin(2\theta)\cos^2(\varphi)$$

In fig. 10 si possono osservare le rappresentazioni polari tridimensionali delle componenti del 2° ordine, con la differenziazione tra parte positiva (in rosso) e parte negativa (in verde).

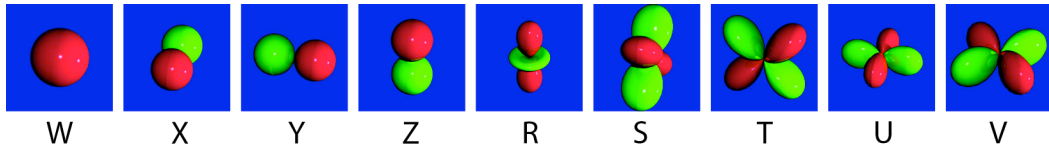


fig. 10: rappresentazioni polari delle componenti del 2° ordine

Recenti studi⁴ hanno portato alla creazione di coefficienti aggiuntivi per le armoniche del 2° ordine, al fine di preservare il guadagno unitario nella combinazione dei componenti. Così le ultime quattro armoniche vanno riscritte in questo modo:

$$S = \frac{\sqrt{3}}{2}\cos(\theta)\sin(2\varphi)$$

$$T = \frac{\sqrt{3}}{2}\sin(\theta)\sin(2\varphi)$$

$$U = \frac{\sqrt{3}}{2}\cos(2\theta)\cos^2(\varphi)$$

$$V = \frac{\sqrt{3}}{2}\sin(2\theta)\cos^2(\varphi)$$

⁴ Malham, D.G. *Higher order Ambisonic systems for the spatialisation of sound* Proceedings, ICMC99, Beijing, October 1999

Come risulta evidente anche dalle osservazioni delle curve polari tridimensionali, il beneficio apportato dall'introduzione di armoniche di ordine superiore al primo consiste nella possibilità di definire meglio i pattern polari risultanti dalla combinazione delle componenti. A titolo esemplificativo si può osservare in fig. 11 la differenza tra un pattern direttivo del primo ordine (in azzurro) ed un pattern direttivo del secondo ordine (in rosso).

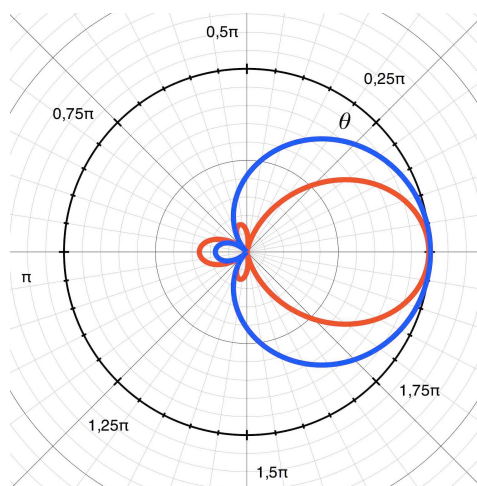


fig. 11: confronto tra i pattern polari del 1° e del 2° ordine

C-Format

Il C-format è una ulteriore derivazione del segnale B-format, ed è costituito da quattro segnali, L, R, T e Q, usati per una serie di formati di registrazioni e trasmissioni radio compatibili mono o stereo (la cosiddetta “gerarchia UHJ”, sviluppata dalla BBC e basata sull'utilizzo dei segnali Ambisonic). La differenza, rispetto ai formati precedenti, consiste nel fatto che alcune di queste componenti possono essere considerate opzionali, e che il risultato del processo di encoding è un segnale stereo, compatibile quindi con un formato commerciale o una trasmissione radiofonica.

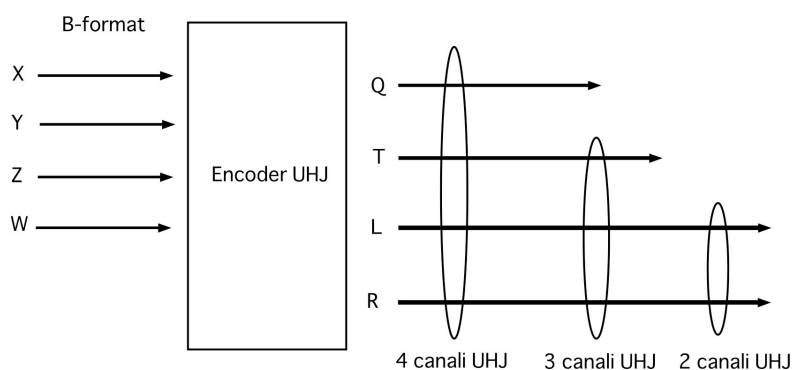


fig. 12: la gerarchia UHJ (C-format)

Il C-format, il cui schema è raffigurato in fig. 12, è prodotto ancora una volta da una matrice (Encoder UHJ), e fornisce un segnale L - R compatibile, un canale T opzionale per precisare la collocazione orizzontale, e un canale Q, anch'esso opzionale, contenente l'informazione in elevazione.

I canali supplementari possono essere codificati, ad es. in una trasmissione radiofonica, mediante la modulazione di una frequenza portante oltre la banda udibile. Per la codifica dei segnali Left-Right sono invece indicate le seguenti equazioni:

$$\text{Left} = (0.0928 + 0.255j)X + (0.4699 - 0.171j)W + 0.3277Y$$

$$\text{Right} = (0.0928 - 0.255j)X + (0.4699 + 0.171j)W - 0.3277Y$$

dove j rappresenta una rotazione di fase di 90°

L'adozione dell'encoder UHJ ha rappresentato un tentativo di commercializzazione delle registrazioni realizzate col sistema Ambisonic, per cui esistono tutt'ora cataloghi discografici nei quali tali registrazioni possono essere reperite e riascoltate con un decoder UHJ.

D-Format e oltre

Il D-format rappresenta una serie di implementazioni riguardanti disposizioni di altoparlanti per riprodurre segnali Ambisonic, sia in B-format che in C-format, del 1° ordine o di ordine superiore. La possibilità di poter determinare, attraverso le equazioni delle armoniche sferiche, la posizione degli altoparlanti nello spazio ne fanno un sistema completamente aperto, ed infatti esso può consistere in una serie di disposizioni di un numero imprecisato di fonti sonore: quattro, sei, o anche otto altoparlanti. Nel caso la disposizione degli altoparlanti sia effettuata per la riproduzione bi-dimensionale (senza la componente in elevazione) essa viene detta "pantofonica" (pantophonic), mentre una configurazione d'ascolto tri-dimensionale viene detta "perifonica" (periphonic). Nella fig. 13 vediamo l'esemplificazione di un sistema Ambisonic pantofonico a sei canali di uscita e quella di un sistema perifonico a otto canali.

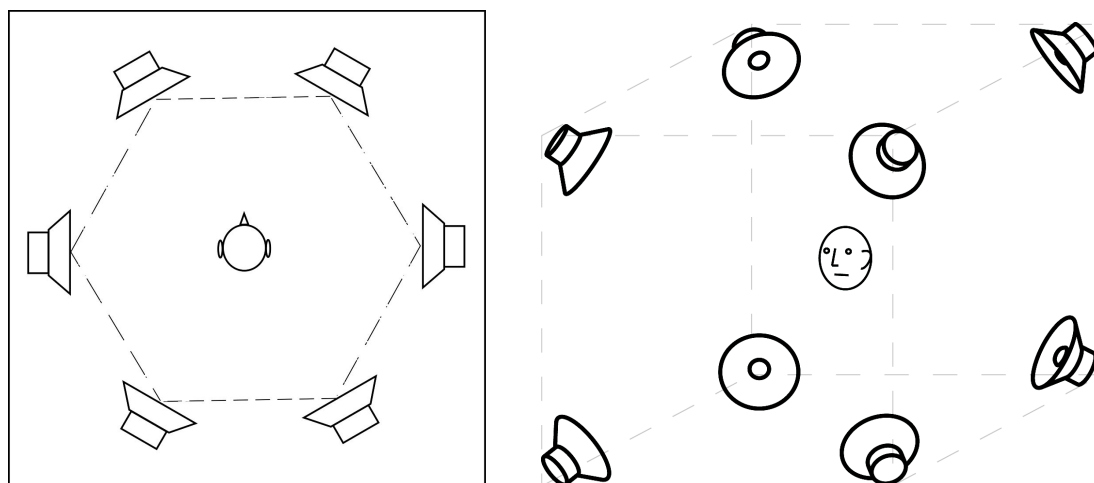


fig. 13: Ambisonic a sei canali pantofonico e otto canali perifonico (D-format)

Va infine menzionato il G-format, il quale consiste nell'implementazione nel sistema d'ascolto 5.1 della ripresa con uscita B-format.

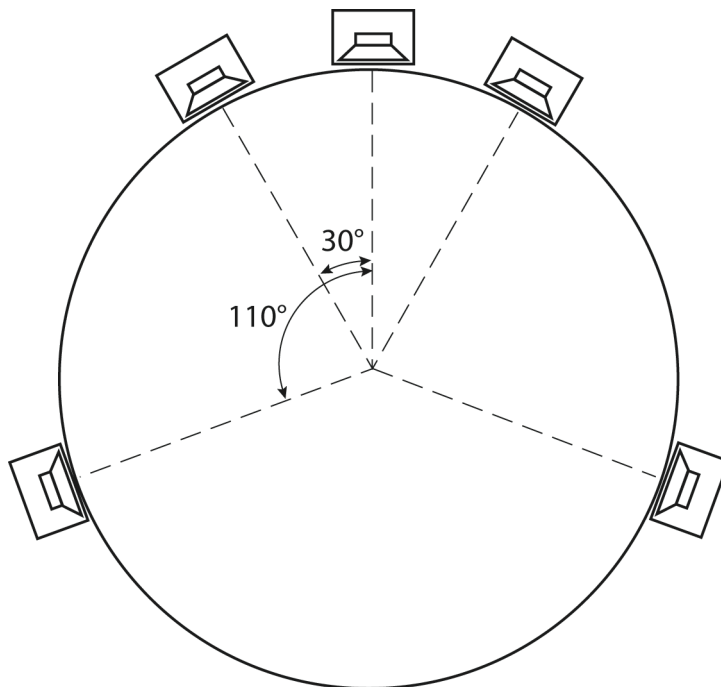


fig. 14: la disposizione ITU-775

Tenendo conto delle angolazioni previste nella disposizione ITU-775 (vedi fig. 14), l'applicazione delle formule esposte per un B-format del 1° ordine (escludendo la componente di elevazione) ci fornisce i seguenti coefficienti:

$$\text{Left} = W + X \cos(30^\circ) + Y \sin(30^\circ) = W + 0.87X + 0.5Y$$

$$\text{Right} = W + X \cos(30^\circ) - Y \sin(30^\circ) = W + 0.87X - 0.5Y$$

$$\text{Center} = W + X$$

$$\text{Left Surround} = W - X \cos(110^\circ) + Y \sin(110^\circ) = W - 0.34X + 0.94Y$$

$$\text{Right Surround} = W - X \cos(110^\circ) - Y \sin(110^\circ) = W - 0.34X - 0.94Y$$

È intuitivo comprendere come il sistema B-format possa racchiudere in sé la chiave per l'implementazione in qualsiasi configurazione spaziale delle sorgenti.

Il sistema Ambisonic non ha avuto fortuna commercialmente, probabilmente perché era da una parte in ritardo (la quadrifonia aveva scoraggiato altre avventure) e da una parte in anticipo sui tempi. L'avvento massiccio dell'audio multicanale a partire dagli anni '90 ha riportato d'attualità l'Ambisonic, e i suoi studi hanno dimostrato tutta la loro validità teorica aprendo nuove prospettive di ricerca.