

### Lezione 2 - Ambisonic

#### Cenni storici

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 Gerzon<sup>1</sup>, 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 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.

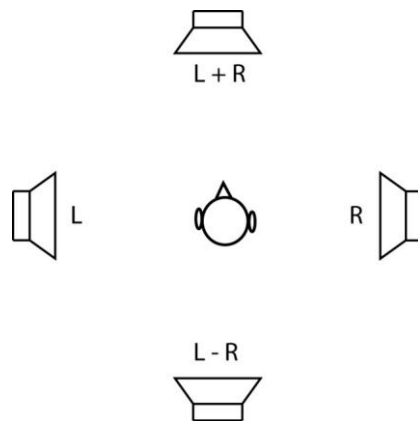
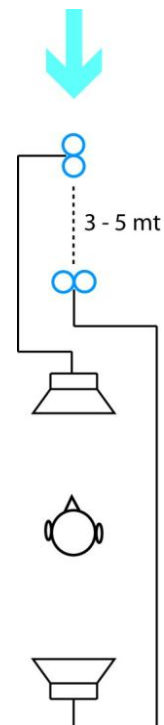


fig. 1: esperimento di spazializzazione condotto da Michael Gerzon

I primi esperimenti condotti da Gerzon furono rivolti all'ascolto spazializzato di audio registrato in stereofonia, mirando ad estrarre dal materiale sonoro le informazioni mancanti alla ricostruzione di un campo sonoro completo sul piano orizzontale. La fig. 1 rappresenta uno degli esperimenti da lui condotti, con una inusuale disposizione di quattro punti di diffusione.



<sup>1</sup> Michael Gerzon (1945 - 1996) ricercatore dell'Oxford Mathematical Institute.

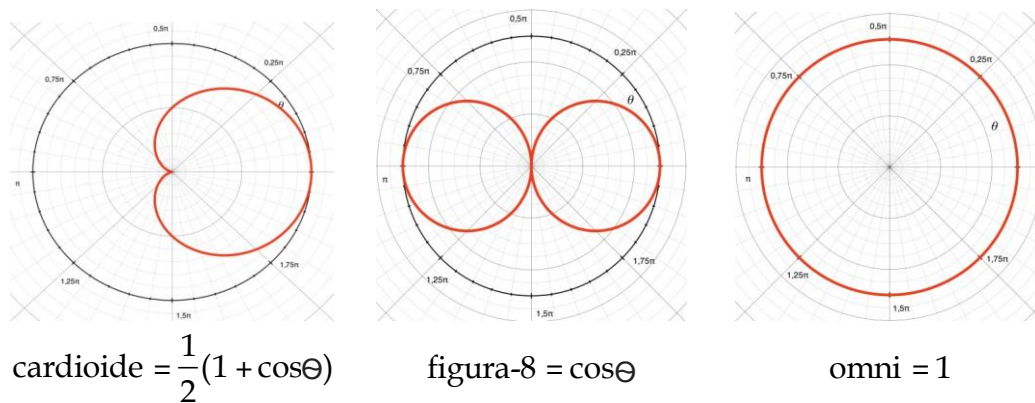
Un altro esperimento condotto da Gerzon (vedi fig. 2), la cosiddetta “Ambiophonic Technique”<sup>2</sup>, riguarda invece un tentativo di ripresa audio con l’intento di fornire una sensazione di audio spazializzato attraverso un campo sonoro posteriore diffuso. *fig. 2: Ambiophonic Technique*

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.

### Impianto teorico dell’Ambisonic

La teoria Ambisonic prevede la possibilità di registrare e riprodurre informazioni sonore provenienti da tutte le direzioni dello spazio, compresa l’elevazione verticale. A 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. 3).



*fig. 3: 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)

---

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

## 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):

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

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

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

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

dove:

$\theta$  = angolo di scostamento orizzontale

$\varphi$  = 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. 4. 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 "Bformat del primo ordine".

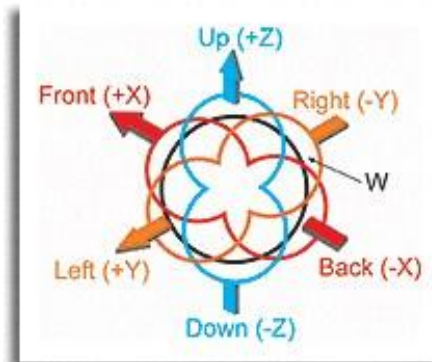


fig. 4: rappresentazione tridimensionale del B-format

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:

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

$$S = Y$$

Mentre una coppia coincidente Blumlein da queste altre:

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

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

### Codifica e Decodifica Ambisonic

Abbiamo visto come è possibile conoscendo i due angoli di scostamento orizzontale  $\theta$  e verticale  $\varphi$ , descrivere la direzione nello spazio di una sorgente sonora, tramite 4 componenti  $W, X, Y$  e  $Z$  ricavate tramite le formule:

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

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

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

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

*Codifica Ambisonic di primo ordine*

Questo processo costituisce la Codifica Ambisonic, più precisamente la Codifica Ambisonic B-Format. A tal proposito è doveroso sottolineare che il sistema Ambisonic descrive la direzione angolare di una sorgente sonora ma non contempla informazioni riguardo la distanza della sorgente dall'ascoltatore.

La peculiarità del sistema Ambisonic è che la decodifica non prevede una determinata configurazione di speaker in riproduzione, ma si adatta elasticamente ad un sistema con un numero variabile di speaker posizionati a qualsiasi angolazione dal punto di ascolto.

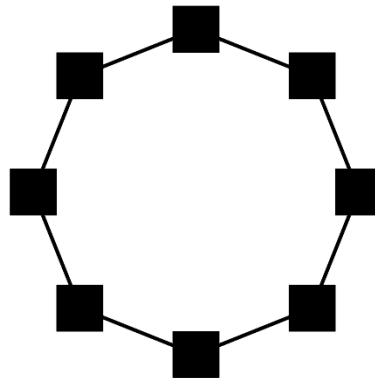
La riproduzione deve essere processata tramite una decodifica del B-Format secondo la seguente formula:

$$P_n = \frac{1}{N} [\sqrt{2W} + kX\cos(\theta) \cos(\varphi) + kY\sin(\theta)\cos(\varphi) + k\sin(\varphi)]$$

N = numero di speaker  
n = ennesimo speaker  
k = costante del peso delle componenti ambisoniche  
θ = angolo scostamento orizzontale dello speaker  
Φ = angolo scostamento verticale dello speaker

*Decodifica Ambisonic di primo ordine*

Un esempio pratico se dovessimo riprodurre una sorgente sonora codificata Ambisonic su un sistema di ascolto con 8 speaker disposti a diamante come il seguente:



Gli speaker sono posizionati a partire dal centrale a distanza di 45° gradi l'uno dall'altro, quindi 0°, 45°, .....315°. Il sistema è planare quindi l'angolo di elevazione φ sarà 0° per ogni speaker.

Applicando la formula di decodifica e ponendo k=1, per ogni speaker si avrà:

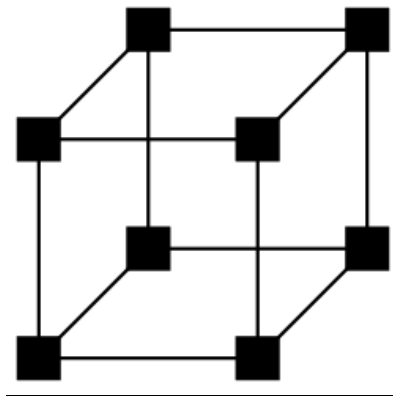
$$P_1 = \frac{1}{8} [\sqrt{2W} + X\cos(0^\circ) \cos(0^\circ) + Y\sin(0^\circ)\cos(0^\circ) + \sin(0^\circ)]$$

$$P_2 = \frac{1}{8} [\sqrt{2W} + X\cos(45^\circ) \cos(0^\circ) + Y\sin(45^\circ)\cos(0^\circ) + \sin(0^\circ)]$$

---

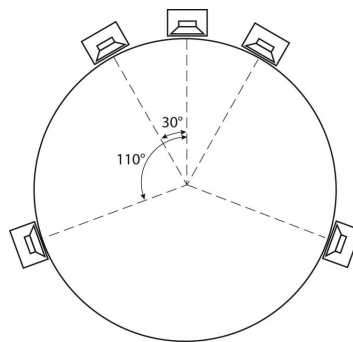

$$P_8 = \frac{1}{8} [\sqrt{2W} + X\cos(315^\circ) \cos(0^\circ) + Y\sin(315^\circ)\cos(0^\circ) + \sin(0^\circ)]$$

Stesso discorso se avessimo un sistema di ascolto con 8 speaker ai vertici di un cubo:



Basterà sostituire nelle otto formule di decodifica per ogni speaker i valori delle posizioni angolari, in questo caso comprendente diversi angoli di elevazione  $\varphi$  per gli speaker del quadrato inferiore e superiore.

La stessa sorgente può essere riprodotta su un sistema d'ascolto con un numero ridotto di speaker, come un sistema stereo, trifonico oppure un sistema multicanale 5.1 ITU-775, lo standard tipico dell'home-theatre, vedremo avanti che quest'ultimo assume uno specifico formato Ambisonic.



*fig. 12: la disposizione ITU-775*

## 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 quelli illustrati in fig. 5, 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-backdown (*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. 5: microfoni tetraedrici ( 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. 6 è riportata la risposta in frequenza rilevata dallo stesso Gerzon<sup>3</sup> 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.

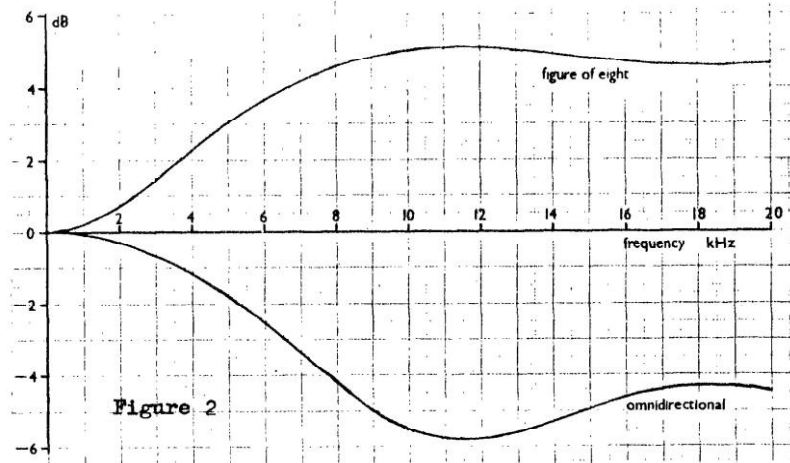


fig. 6: errori di risposta in frequenza nel microfono tetraedrico

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

---

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



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

fig. 7: array Ambisonic orizzontale



Parimenti è evidente che ove non interessasse la componente di elevazione il microfono tetraedrico pu 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. 7.

### Ambisonic di ordine superiore

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  $V$ . Per il 3° ordine si aggiungooe le componenti  $K, L, M, N, O, P$  e  $Q$ .

Table 1 Ambisonic B Format Channels to 3<sup>rd</sup>. Order

Order	m,n, $\zeta$	Channel	SN3D definition	FuMa weight
0	0,0,1	W	1	$1/\sqrt{2}$
1	1,1,1	X	$\cos\theta\cos\phi$	1
	1,1,-1	Y	$\sin\theta\cos\phi$	1
	1,0,1	Z	$\sin\phi$	1
2	2,0,1	R	$(3\sin^2\phi - 1)/2$	1
	2,1,1	S	$(\sqrt{3}/2) \cos\theta\sin(2\phi)$	$2/\sqrt{3}$
	2,1,-1	T	$(\sqrt{3}/2) \sin\theta\sin(2\phi)$	$2/\sqrt{3}$
	2,2,1	U	$(\sqrt{3}/2) \cos(2\theta)\cos^2\phi$	$2/\sqrt{3}$
	2,2,-1	V	$(\sqrt{3}/2) \sin(2\theta)\cos^2\phi$	$2/\sqrt{3}$
3	3,0,1	K	$\sin\phi(5\sin^2\phi - 3)/2$	1
	3,1,1	L	$(\sqrt{3}/8) \cos\theta\cos\phi(5\sin^2\phi - 1)$	$\sqrt{45}/32$
	3,1,-1	M	$(\sqrt{3}/8) \sin\theta\cos\phi(5\sin^2\phi - 1)$	$\sqrt{45}/32$
	3,2,1	N	$(\sqrt{15}/2) \cos(2\theta)\sin\phi\cos^2\phi$	$3/\sqrt{5}$
	3,2,-1	O	$(\sqrt{15}/2) \sin(2\theta)\sin\phi\cos^2\phi$	$3/\sqrt{5}$
	3,3,1	P	$(\sqrt{5}/8) \cos(3\theta)\cos^3\phi$	$\sqrt{8}/5$
	3,3,-1	Q	$(\sqrt{5}/8) \sin(3\theta)\cos^3\phi$	$\sqrt{8}/5$

In fig. 8 si possono osservare le rappresentazioni polari tridimensionali delle componenti fino al 3° ordine, con la differenziazione tra parte positiva (in grigio) e parte negativa (in nero).

### **Componenti Sferiche Ambisonic fino al 3° Ordine**

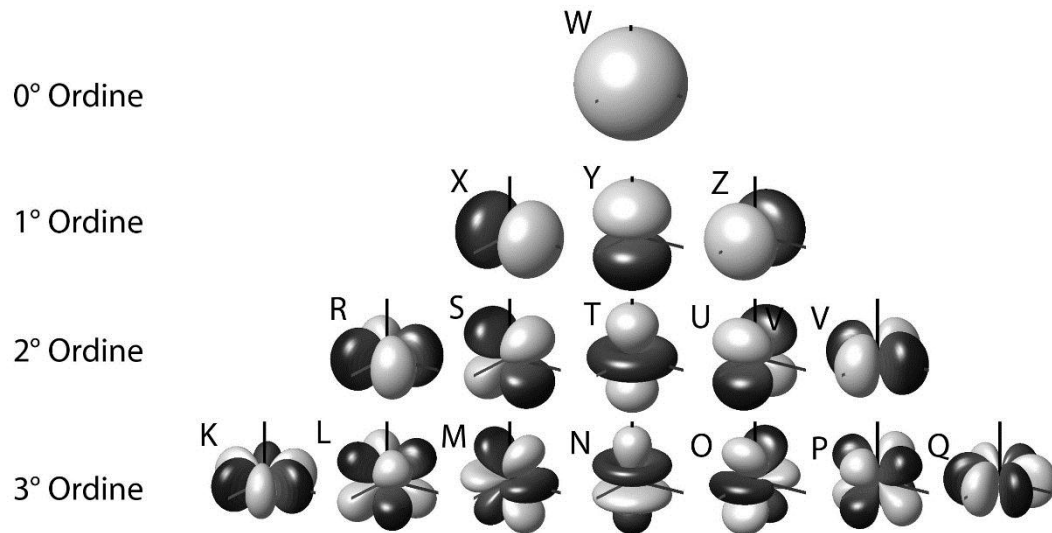


fig. 8: rappresentazioni polari delle componenti finio al 3° ordine

### **DECODIFICA AMBISONIC FINO AL TERZO ORDINE**

**I ordine** 
$$P_n = \frac{1}{N} [\sqrt{2}W + kX\cos(\Theta)\cos(\Phi) + kY\sin(\Theta)\cos(\Phi) + kZ\sin(\Phi)]$$

**II ordine** 
$$+ kR\left(\frac{3}{2}\sin^2(\Phi) - \frac{1}{2}\right) + kS\cos(\Theta)\sin(2\Phi) + kT\sin(\Theta)\sin(2\Phi) + kU\cos(2\Theta)\cos^2(\Phi) + kV\frac{1}{2}\sin(2\Theta)\cos^2(\Phi)$$

**III ordine** 
$$+ kN\cos(2\Theta)\sin^2(\Phi)\cos(\Phi) + kO\sin(2\Theta)\sin^2(\Phi)\cos(\Phi) + kP\cos(3\Theta)\cos^3(\Phi) + kQ(3\Theta)\cos^3(\Phi)]$$
  

$$+ kV\frac{1}{2}\sin(2\Theta)\cos^2(\Phi) + kK\sin(\Phi)(5\sin^2(\Phi) - 3) + kL\cos(\Theta)\cos(\Phi)(5\sin^2(\Phi) - 1) + kM\sin(\Theta)\cos(\Phi)(5\sin^2(\Phi) - 1)$$

N = numero di speaker

n = ennesimo speaker

k = costante del peso delle componenti ambi soniche

theta = angolo scostamento orizzontale

phi = angolo scostamento verticale

La costante K può dipendere dall'ordine, è quindi dalle componenti spaziali utilizzate e andrebbe riscalata per mantenere l'ampiezza finale coerente.

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. 9 la differenza tra un pattern direttivo del primo ordine (in azzurro) ed un pattern direttivo del secondo ordine (in rosso).

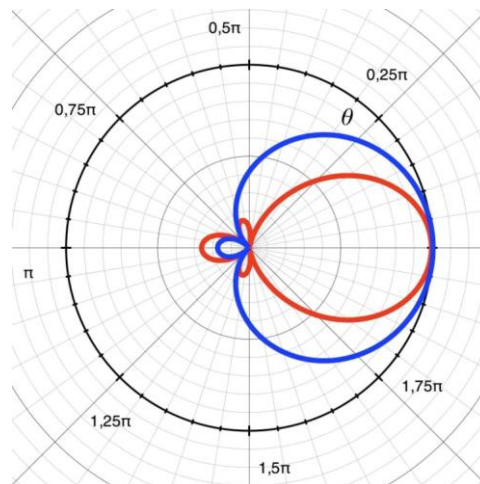


fig. 9: confronto tra i pattern polari del 1° e del 2° ordine

## C-Format

Il C-format 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"). 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.

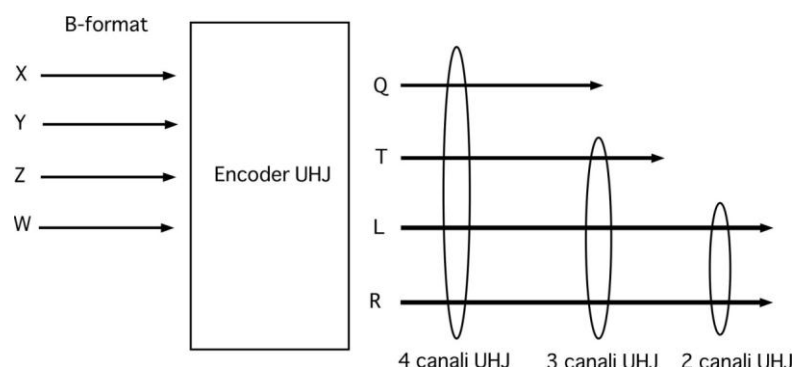


fig. 10: la gerarchia UHJ (C-format)

Il C-format, il cui schema è raffigurato in fig. 10, 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  $90^\circ$

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 Cformat, 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 pu 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 bidimensionale (senza la componente in elevazione) essa viene detta "pantofonica" (pantophonic), mentre una configurazione d'ascolto tridimensionale viene detta "perifonica" (periphonic). Nella fig. 11 vediamo l'esemplificazione di un sistema Ambisonic pantofonico a sei canali di uscita e quella di un sistema perifonico a otto canali.

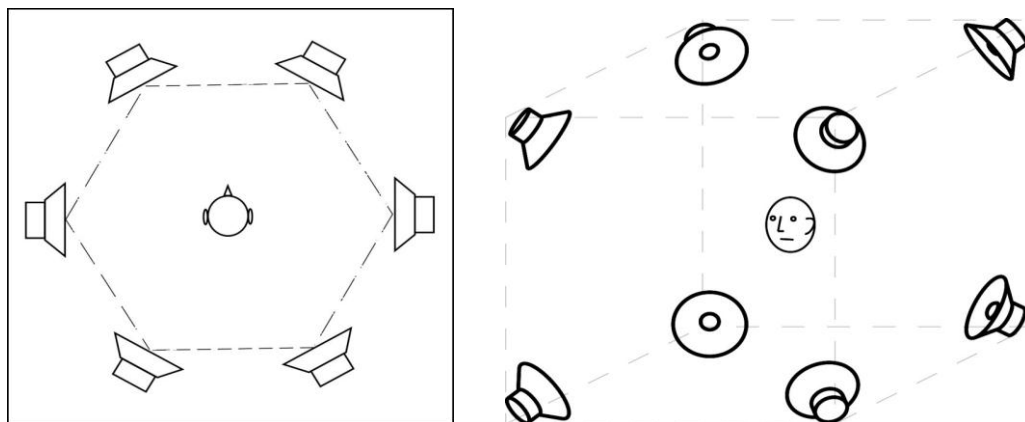


fig. 11: Ambisonic a sei canali pantofonico e otto canali perifonico (D-format)

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

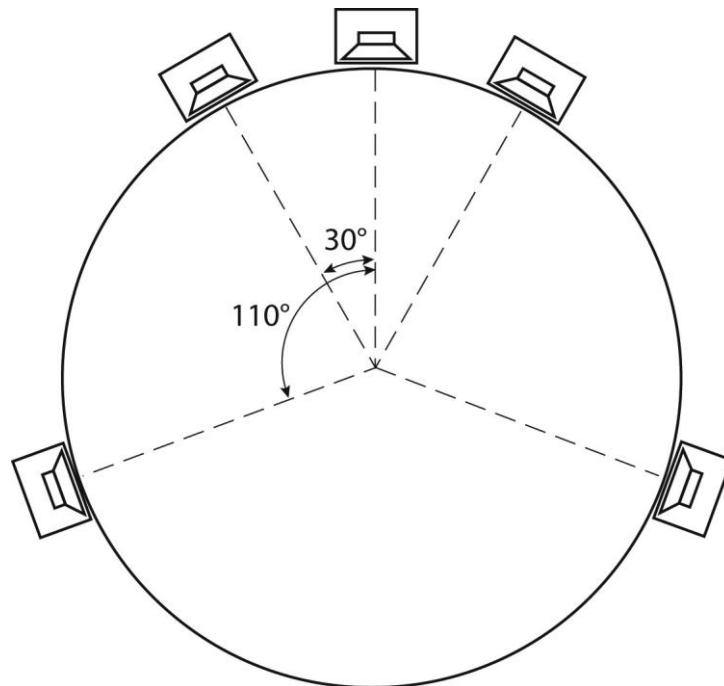


fig. 12: la disposizione ITU-775

Tenendo conto delle angolazioni previste nella disposizione ITU-775 (vedi fig. 12), 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.