

Spazializzare il suono

Le orecchie sono posizionate alla stessa altezza, agli antipodi della testa (alta impedenza acustica). La testa agisce sulla localizzazione delle sorgenti sonore, producendo due principali effetti: *Interaural Time Difference* e *Interaural Level Difference*.

Interaural Time Difference ITD (Differenze Interaurali di Tempo): sono determinate dalla distanza fisica delle due orecchie. Le onde sonore sono costrette a percorrere una maggiore distanza per raggiungere l'orecchio più lontano. Si supponga la testa sferica e la sorgente lontana: è possibile esprimere lo spazio ulteriore Δx che un raggio sonoro deve percorrere per raggiungere l'orecchio più lontano, come:

$$ITD \sim \frac{a}{v}(\theta + \sin(\theta))$$

v = velocità di propagazione del suono

a = raggio della testa

θ = azimuth

Quando la sorgente è frontale $\theta = 0$ l'ITD è minimo, viceversa quando la sorgente si trova ai lati $\theta = \frac{\pi}{2}$ è massimo.

Interaural Level Difference ILD (Differenze Interaurali di Livello): La differenza d'intensità ai due orecchi, nell'ipotesi che la testa funzioni come una sfera di raggio r , per un suono di lunghezza d'onda λ , che provenga da destra o da sinistra, è data da:

$$\Delta = \frac{3}{4} \left(\frac{2\pi r}{\lambda} \right)^4$$

L'ILD interessa fortemente la percezione delle alte frequenze. Di fatto, il nostro sistema uditivo è in grado di apprezzare differenze di fasi relative per frequenze non superiori a $1KHz$. Un'onda sinusoidale di frequenza $f = 3KHz$ ad un angolo di azimuth di 90° , ad esempio, sarà attenuata $\approx 10dB$, a $6KHz$ di circa $20dB$ e, a $10KHz$ di circa $35dB$ e, così via. Al di sotto di $1KHz$ l'ILD è ininfluente: il fronte d'onda produce diffrazione e quindi il contributo alla localizzazione è minimo.

Percezione del suono nello spazio

Fissata la posizione dell'ascoltatore, la precisa localizzazione della sorgente è completamente determinata dalla conoscenza di tre grandezze: *angolo sul piano orizzontale*, *angolo sul piano verticale* e *distanza*.

La localizzazione sul piano orizzontale dipende dal ritardo interaurale. L'orecchio che per primo riceve il segnale fornisce le informazioni utili alla localizzazione della sorgente.

La percezione della distanza, in condizioni di totale assenza di riverberazione, è in diretta relazione con l'intensità. Secondo la legge inversa del quadrato: data un'intensità di riferimento e distanza, *l'intensità di una sorgente sonora omnidirezionale si riduce di circa 6dB per ogni raddoppio della distanza dalla sorgente* ($g = \frac{1}{d^2}$).

Molto più complessi sono i meccanismi che interessano la localizzazione sul piano verticale. Essa, dipende dalla superficie esterna delle orecchie, dal torso e, infine, dalle spalle. Le componenti del suono al di sopra di $\approx 6KHz$ vengono riflesse e filtrate dalla superficie esterna delle orecchie. Il suono prima di raggiungere il timpano subisce delle micro-variazioni spettrali e temporali a causa di micro-riflessioni indotte dalle pieghe del padiglione uditivo e dall'angolo di incidenza. Spalle e torso sono argomento assai complicato dal punto di vista geometrico. È possibile, comunque, una descrizione semplificata considerando lo *snouman model* (considera il torso ellissoidale): misurando la risposta impulsiva all'orecchio, si nota che all'impulso iniziale seguono una serie di altri impulsi con ritardo che varia al variare dell'elevazione, come se il torso si comportasse alla pari di un filtro a pettine introducendo notch periodici nello spettro.

Sistemi di coordinate

Un sistema di coordinate, altro non è che un *sistema di riferimento basato su coordinate (o etichette) per l'individuazione della posizione di un oggetto in uno spazio definito*.

Si parla di:

- sistema di riferimento **monodimensionale** (x)
- sistema di riferimento **bidimensionale** (x, y)
- sistema di riferimento **tridimensionale** (x, y, z)

Dato un piano di riferimento, la posizione di un punto al suo interno solitamente la si indica con una coppia di numeri reali, coordinate cartesiane (x, y). Si parla, invece, di sistema polare quando la il punto lo si indica con la coppia (ρ, ϕ), con ρ che definisce la distanza del punto dal punto di origine e, ϕ l'angolo. Se consideriamo il vettore \vec{v} , $\rho = ||\vec{v}||$ e, ϕ l'angolo tra il vettore e il verso positivo dell'asse x .

Posto il triangolo rettangolo $\overset{\Delta}{OPS}$, con

\overline{PS} = lato opposto

\overline{OP} = ipotenusa

\overline{OS} = lato adiacente

allora:

$$\begin{cases} \sin(\phi) = \frac{\overline{PS}}{\overline{OP}} \\ \cos(\phi) = \frac{\overline{OS}}{\overline{OP}} \\ \tan(\phi) = \frac{\overline{PS}}{\overline{OS}} \end{cases}$$

$$\begin{cases} \sin(\phi) = \frac{y}{\rho} \\ \cos(\phi) = \frac{x}{\rho} \end{cases}$$

$$\begin{cases} \rho = \sqrt{x^2 + y^2} = \sqrt{\rho^2(\sin^2(\phi) + \cos^2(\phi))} \\ \phi = \arctan\left(\frac{y}{x}\right) = \arccos\left(\frac{x}{\rho}\right) = \arcsin\left(\frac{y}{\rho}\right) \end{cases}$$

In un sistema a tre dimensioni, identifichiamo la posizione del punto nel piano con un insieme di tre valori (x, y, z) o (ρ, ϕ, ψ)

$$\begin{cases} x = \rho \cdot \cos(\psi) \cos(\phi) \\ y = \rho \cdot \cos(\psi) \sin(\phi) \\ z = \rho \cdot \sin(\psi) \end{cases}$$

$$\begin{cases} \rho = \sqrt{x^2 + y^2 + z^2} & \rightarrow \in [0, +\infty) \\ \phi = \arctan\left(\frac{z}{\sqrt{x^2 + y^2}}\right) & \rightarrow \text{azimuth} \in [0, 2\pi) \\ \psi = \arctan\left(\frac{y}{x}\right) & \rightarrow \text{elevazione} \in [0, \pi) \end{cases}$$

Amplitude panning

Progettare e gestire la posizione di una sorgente audio virtuale diffusa da due o più altoparlanti mediante il controllo dell'ampiezza (amplitude panning).

L'ascoltatore percepisce l'illusione di un singolo segnale, una sorgente codidetta *virtuale* localizzata in una porzione di spazio definita dalla posizione degli altoparlanti, del ricevente e dalla quantità di ampiezza del segnale su ciascun altoparlante.

Dato un segnale $x = g \cdot s$

$$\begin{cases} x_{left} = g_1 \cdot s \\ x_{right} = g_2 \cdot s \end{cases}$$

Posti ϕ_1 e ϕ_2 l'azimuth di una coppia di altoparlanti adiacenti, l'angolo di panning è definito da:

$$\begin{cases} g_1 = \sqrt{\frac{\phi_2 - \phi_{pan}}{\phi_2 - \phi_1}} \\ g_2 = \sqrt{\frac{\phi_{pan} - \phi_1}{\phi_2 - \phi_1}} \end{cases}$$

con $\phi_1 \leq \phi_{pan} \leq \phi_2$

Linear and costant power panning function: la funzione lineare di diffusione genera un controllo dell'ampiezza lineare mediante correlazione inversa.

$$\begin{aligned}\text{left} &= g \\ \text{right} &= 1 - g\end{aligned}$$

con $g_1 + g_2 = 1$

Posto $\phi \in [0, \frac{\pi}{2}]$

$$\begin{cases} L(\phi) = (\frac{\pi}{2} - \phi) \cdot 2\pi \\ R(\phi) = \phi \cdot 2\pi \end{cases}$$

Quando $\phi = \frac{\pi}{4}$, la sorgente dovrebbe trovarsi al centro $L(\frac{\pi}{4}) + R(\frac{\pi}{4}) = 1$. Una funzione di controllo del genere, con il segnale posizionato al centro del panorama stereo, produce, però, una forte sensazione di attenuazione del segnale, cosiddetta *hole in the middle*: il segnale appare più forte agli estremi che al centro. Per ovviare a ciò, si ricorre all'utilizzo di un'altra funzione che segue curve logaritmiche, la *constant power panning function*.

$$\begin{cases} L(\phi) = \cos(\phi) \\ R(\phi) = \sin(\phi) \end{cases}$$

$$\cos\left(\frac{\pi}{4}\right) = \sin\left(\frac{\pi}{4}\right) = 0.71$$

con un attenuazione di

$$-20 \log_{10}(0.71) = -3dB$$

VBAP (Vector Base Amplitude Panning)

Il vettore p che punta alla sorgente virtuale può essere espresso come la combinazione lineare dei vettori associati agli speaker

$$p = g_1 l_1 + g_2 l_2$$

in forma matriciale

$$p^T = g L_{12}$$

$$g = [g_1 \ g_2]$$

$$L_{12} = [l_1 \ l_2]^T$$

Ciò è possibile solo se L_{12}^{-1} esiste.

$$g = p^T L_{12}^{-1} = [p_1 \ p_2] \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1}$$

Si ragiona per coppie di speakers (base), generalmente adiacenti (ciascun altoparlante appartiene a due diverse coppie; lo speaker 4, ad esempio, appartiene sia alla coppia 3-4 che a 4-5). La localizzazione della sorgente virtuale è possibile, allora, attivando di volta in volta la coppia interessata (*arco attivo*), proprio come avviene in un sistema con soli due altoparlanti. Ciò che varia è il numero di basi, nello specifico, il campo sonoro che può essere riprodotto con il metodo VBAP è dato dall'unione degli archi attivi delle basi degli altoparlanti disponibili.

Nota: Il riferimento per approfondire questo argomento è: Ville Pullki, *Virtual Sound Source Positioning Using Vector Base Amplitude Panning*, Audio Engineering Society, Inc. 1997

DBAP (Distance-Based Amplitude Panning)

Il modello DBAP estende quanto detto fino ad ora da una coppia di altoparlanti ad un array di un numero indefinito di altoparlanti.

Poniamo le coordinate cartesiane della nostra sorgente (x_s, y_s) ; la posizione dell' i -esimo speaker (x_i, y_i) . La distanza della sorgente dagli speakers è data da

$$d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2}$$

Gli speakers sono sempre tutti attivi, con ampiezza relativa

$$g_i = \frac{k}{2d_i a}$$

con k coefficiente che dipende dalla posizione della sorgente e di tutti gli altoparlanti

$$k = \frac{2a}{\sqrt{\sum_{i=1}^N \frac{1}{d_i^2}}} \quad a = 10^{\frac{-R}{20}}$$

con $R = 6dB$ a rappresentare l'attenuazione al raddoppio della distanza.