

ELEMENTI DI SPAZIALIZZAZIONE DEL SUONO

E. Giordani

Lo spazio reale e sue virtualizzazioni

La *spazializzazione virtuale* si avvale di tecniche di DSP e di concetti di acustica e psicoacustica per simulare uno spazio immaginario attraverso la normale tecnica di diffusione elettroacustica (riproduzione bicanale). La riproduzione deve avvenire in luoghi acusticamente neutri (non riverberanti e senza colorazioni).

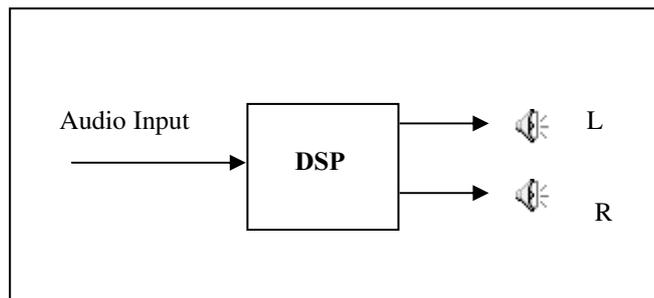


Fig. 1

Questa tecnica è oggi applicata prevalentemente al vasto campo della Virtual Reality, dai simulatori di volo alle applicazioni multimedia (3D sound) in genere; risulta più appropriata per applicazioni individualizzate per la forte dipendenza dalla posizione d'ascolto.

La *spazializzazione fisica* del suono consiste invece nella riproduzione del suono attraverso sistemi di multi-canale. La riproduzione può essere lineare a N canali o prevedere una qualche forma di codifica (come la tecnica Dolby Surround o la tecnica Ambisonica o l'Olofonia). Questa tecnica è prevalentemente utilizzata nella riproduzione tridimensionale del cinema, nella musica elettronica ed acustica, nelle installazioni sonore.

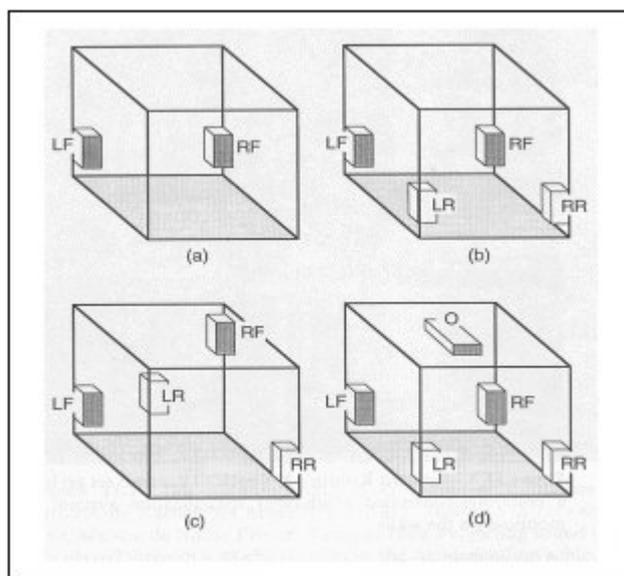


Fig. 2

In figura 2 sono mostrate alcune configurazioni per la spazializzazione. A) Stereo Base, LF=left front, RF=right front . B) Quadrifonica RR= right rear, LR= left rear, C) Quadrifonia Perifonica .RF e LR sono montati come in C ma con una quota di elevazione. Quando il suono muta la sua posizione orizzontale si muove anche verticalmente. D) Configurazione a 5 altoparlanti con il quinto direzionato dall'alto verso il basso

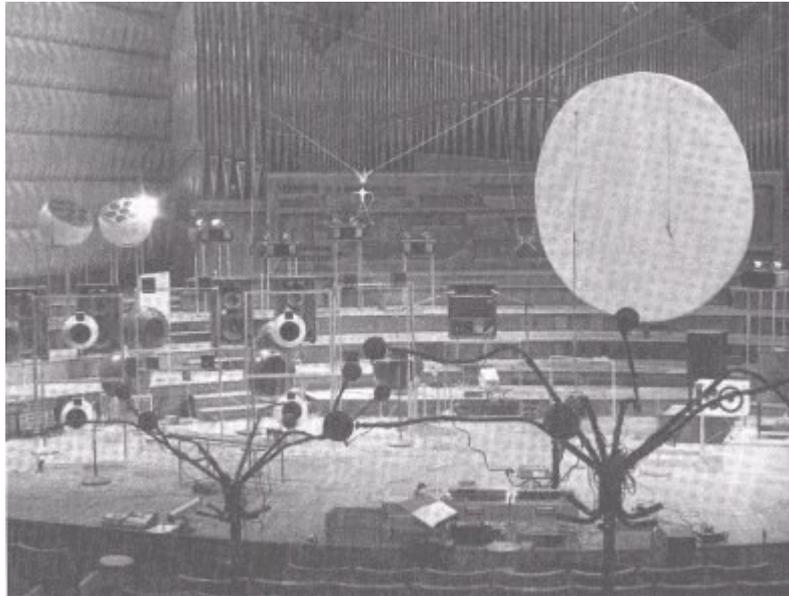


Fig. 3

La figura 3 mostra un Acusmonium, un sistema di spazializzazione multicanale progettato dal GRM (Gruppo per la Ricerca Musicale) installato nell'auditorium O. Messiaen a Parigi. Il suono viene diffuso attraverso circa 80 altoparlanti pilotati da una console mixer a 48 canali. L'immagine sonora che si ottiene è molto ampia e può competere con quella di una grande orchestra. L'ascoltatore può porsi frontalmente od eventualmente all'interno della struttura. L'idea base è comunque quella di "riorchestrare spazialmente e timbricamente" le composizioni elettroniche.

Benché la spazializzazione del suono abbia notevolmente tratto beneficio dalle recenti tecniche digitali, vi sono stati molti tentativi in passato di utilizzare il parametro spazio come variabile estetica, espressiva e compositiva.

Può essere utile ricordare l'uso di altoparlanti rotanti da parte di K. Stockhausen, impiegati in almeno due composizioni, *Kontakte* del 1960 e *Hymnen* del 1967 (fig.4) e un particolare amplificatore a doppio rotore (Leslie) per l'organo elettromagnetico Hammond (figg.5/6), divenuto poi famoso nell'ambito della musica jazz, blues e pop-rock.

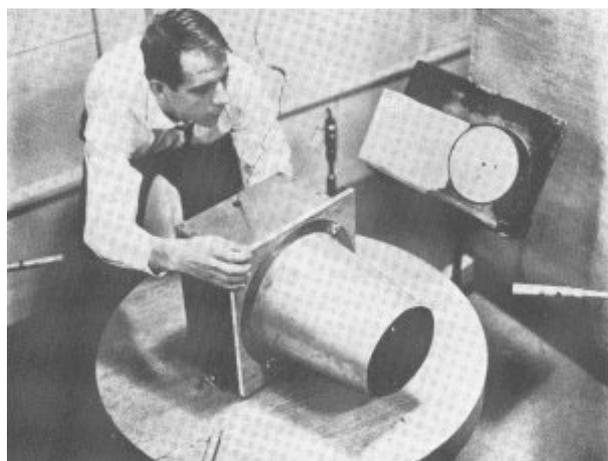


Fig.4

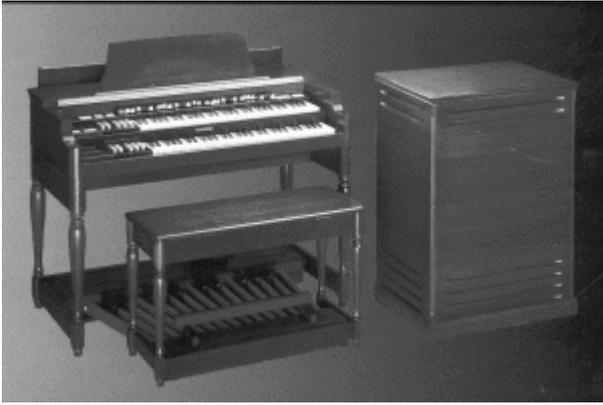


Fig.5

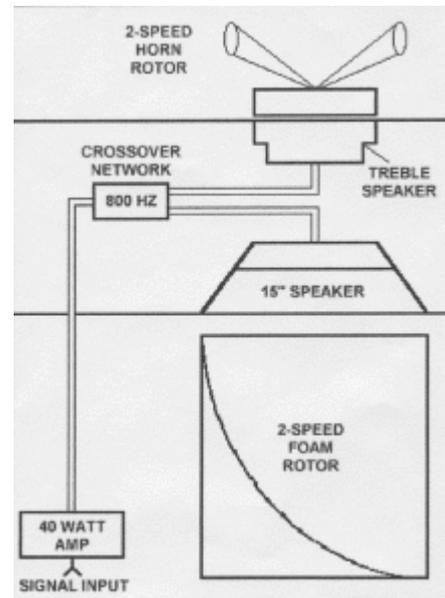


Fig. 6

Naturalmente, attraverso le attuali tecniche di DSP si è arrivati a simulare digitalmente il suono in movimento con un discreto grado di approssimazione. Occorre tenere presente che il modello di simulazione è molto complesso poiché si deve tenere conto di molti fenomeni concomitanti che interagiscono tra loro: effetto Doppler, filtri tempo-varianti, variazioni di fase, distorsioni, riflessioni, caratteristiche degli amplificatori e degli altoparlanti.

Percezione Spaziale del Suono

La *localizzazione* delle sorgenti sonore nello spazio dipende primariamente da due fattori denominati **ITD** (*interaural time difference*) e **IID** (*interaural intensity differences*) la cui conoscenza risale a circa il 1907. Sebbene il fenomeno della localizzazione sia in generale più complesso, sono stati compiuti molti studi e ricerche su ciò che potremmo definire un caso particolare di localizzazione, definito comunemente *lateralizzazione*.

Infatti il modello impiegato per lo studio della lateralizzazione prevede che un ipotetico ascoltatore (avente una testa perfettamente rotonda e privo di orecchi esterno) sia posto in un ambiente anecoico (privo di qualsiasi riflessione apprezzabile) davanti ad una sorgente sonora posta a livello dei suoi occhi (fig. 7).

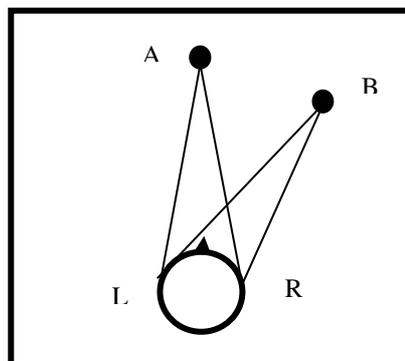


Fig. 7

Questo modello implica il calcolo dei due cammini che rappresentano il fronte dell'onda sonora dal suo centro d'origine ai due punti che rappresentano l'ingresso del canale uditivo. Si suppone inoltre che i punti A e B siano posti nel piano centrale della sfera, agli estremi dell'asse interaurale.

In queste condizioni, con la sorgente in posizione A, i due cammini sono esattamente identici ed i rispettivi fronti sonori raggiungono i timpani contemporaneamente e con la stessa intensità. Viceversa, nella posizione B, la sorgente è a circa 60 gradi di azimut a destra dell'ascoltatore quindi i due cammini sono ora diversi: il cammino BL è maggiore del cammino BR quindi il fronte d'onda arriverà un po' in ritardo al punto L rispetto ad R e con un'intensità minore.

La minore intensità è giustificata dal fatto che il percorso $BL > BR$ e che quindi vi è una maggiore perdita energetica (che si ricorda essere direttamente proporzionale all'inverso del quadrato della distanza). Inoltre la testa determina un'effetto schermo che attenua ulteriormente l'energia acustica entrante al punto L.

Considerando la sorgente B ad una distanza sufficientemente grande, si può ipotizzare che i due fronti d'onda siano praticamente paralleli e posto che la distanza LR sia mediamente 20/21 cm il modello può essere ulteriormente semplificato (fig. 8).

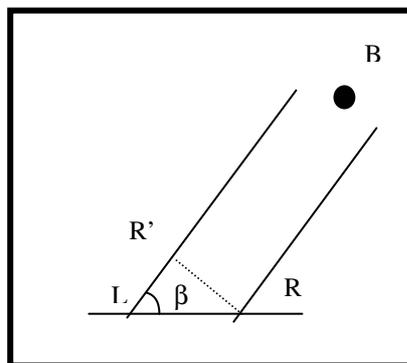


Fig. 8

Da semplici considerazioni geometriche si può dimostrare che vale la relazione:

$$LR' = LR * \cos \beta$$

E quindi :

$$Td = 341 / LR'$$

Dove 341 = velocità media del suono in m/s

Td = ritardo temporale del cammino BL rispetto a BR.

Il valore massimo del ritardo si ottiene quando $\beta = 0^\circ$ o $\beta = 180^\circ$ (sorgente B tutta a destra o tutta a sinistra) e vale all'incirca **640 μsec** (massimo valore di ITD).

Per frequenze non superiori ad 1 KHz, il nostro sistema uditivo è in grado di apprezzare le differenze di fasi relative. Con la sorgente in posizione B, come già anticipato, si produce anche un apprezzabile differenza interaurale di intensità (IID), ma solo per quelle componenti di frequenza la cui lunghezza d'onda è più piccola del diametro della testa, ovvero per frequenze maggiori di circa 1.5 kHz. Inoltre le frequenze superiori saranno attenuate in misura tanto maggiore tanto più grande è la frequenza.

Per esempio, un'onda sinusoidale di frequenza 3 kHz ad un valore di azimut di 90° sarà attenuata di circa 10 dB, a 6 kHz di circa 20 dB e a 10 kHz di circa 35 dB.

Ma al di sotto di circa 1 kHz la caratteristica di IID diviene ininfluente poiché il fronte d'onda produce diffrazione (cioè curva attorno all'ostacolo) e quindi il contributo alla localizzazione è minimo.

Generalmente però, si è portati ad interpretare differenze di intensità a livello timpanico come cambiamenti nella posizione della sorgente. Tutte le console di missaggio sono infatti dotate di un dispositivo di tipo IID denominato *panpot* (**panoramic-potentiometer**). Specialmente con ascolto in cuffia, il panpot crea IID indipendentemente dalla frequenza, cioè viene impiegato per separare le sorgenti in molti casi. Ciò perché il contenuto frequenziale di suoni tipici contiene frequenze al di sopra e al di sotto dell'ipotetico punto di cross-over per l'IID e l'ITD, e così l'ascoltatore è sensibile alla caratteristica IID per gran parte della gamma udibile (con grado variabile fino a circa 200 hz).

Se una sorgente sonora monoaurale (*diotica*) viene convogliata attraverso una cuffia, l'immagine virtuale uditiva è unica ed appare al centro della testa dell'ascoltatore. Una situazione simile si verifica con un sistema di riproduzione a due canali; per trovare la posizione ideale di ascolto (*sweet spot*), è sufficiente ascoltare la riproduzione di una sorgente monoaurale e aggiustare la propria posizione fino alla percezione di un'immagine centrale.

All'aumentare indiscriminato dei valori di IID e ITD al di sopra di determinate soglie, la posizione dell'immagine virtuale del suono scivolerà verso l'orecchio in anticipo o di maggiore intensità. Raggiunti tali limiti, il suono smetterà di muoversi lungo l'asse interaurale e si posizionerà nelle posizioni estreme. Nella figura 9 viene mostrata un diagramma comparativo di localizzazione in funzione di IID e ITD. In ordinata una scala di valori soggettiva (da 1 a 5) misura il grado di lateralizzazione dal centro. Relativamente a due diversi tipi di stimoli (impulsi brevi o rumore a larga banda), il campo effettivo di ITD è circa da 0.005 a 1.5 ms mentre il campo effettivo di IID va da 1 a 10 dB.

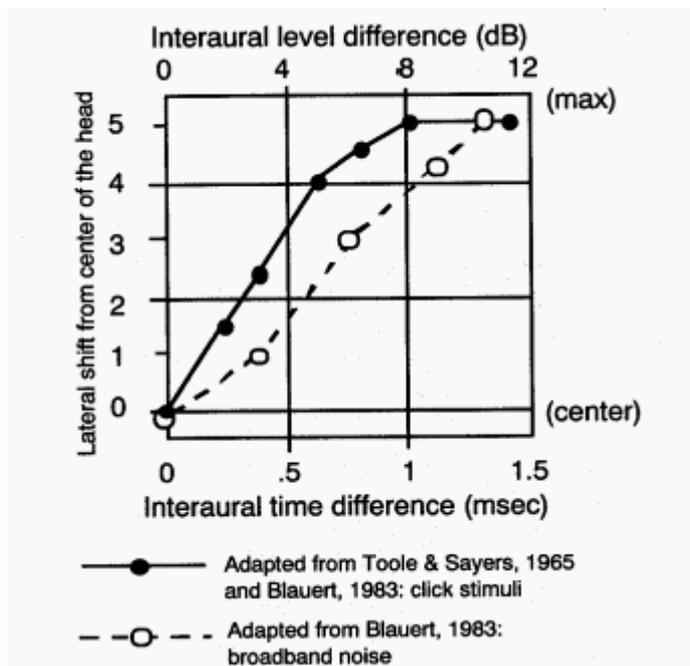


Fig. 9

Benchè il processo di lateralizzazione venga identificato lungo l'asse interaurale, occorre ricordare che esperimenti condotti su vari ascoltatori mostrano l'esistenza di immagini indotte di tipo *verticale* ed inoltre può verificarsi un tipo di percezione che viene attribuita ad una sorgente posta sul lato *anteriore* o *posteriore* della testa.

Quando la differenza interaurale (ITD) raggiunge valori molto elevati si producono effetti percettivi piuttosto confusi sino al limite estremo dei 40 ms, oltre il quale il sistema uditivo

percepisce il suono in forma di *eco*. Inoltre all'interno di quest'ambito allargato può essere inserito il fenomeno dell'**effetto di precedenza** o **effetto Haas** per mezzo del quale è possibile spiegare un'importante meccanismo inibitorio del sistema uditivo che consente la localizzazione del suono anche in presenza di forte riverberazione.

In figura 10 viene rappresentato in forma molto schematica l'immagine uditiva in funzione di diversi ambiti di ritardo temporale, fino ad un massimo di 40 ms.

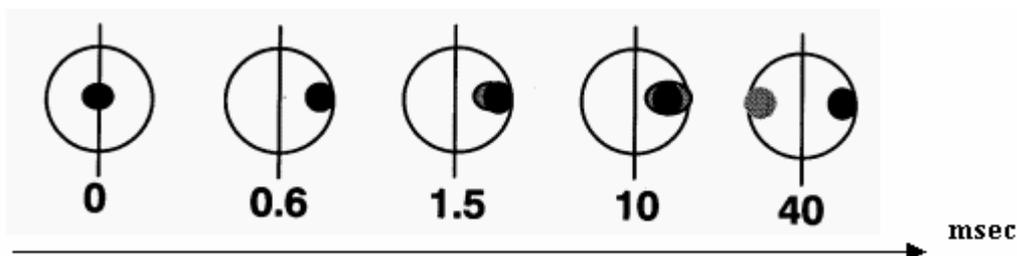


Fig. 10

Tra 0 e 0.6 ms, l'immagine scivola lungo l'asse interaurale, come predetto dalla teoria dell'ITD. Quando il ritardo aumenta fino a circa 1.5 ms la posizione rimane generalmente invariata, ma lievemente allargata (vedi area grigia). All'aumentare del ritardo l'area si allarga ulteriormente e sopra i 10 ms il centro di gravità tende a regredire verso il centro. Attorno al valore medio di 40 ms, dipendentemente dal tipo di sorgente, l'area si separa improvvisamente in due immagini distinte fino a degenerare nella percezione di un'eco.

Percezione della Distanza della Sorgente Sonora

La percezione della distanza, in assenza di altre caratteristiche oggettive e soggettive può essere messa in relazione diretta con l'intensità della sorgente. Occorre però distinguere tra senso della distanza **assoluta** e **relativa**. Nel primo caso, ci si riferisce all'abilità di stimare la distanza di una sorgente dopo una certa esposizione ma senza alcuna informazione di familiarità cognitiva. Nel secondo caso si deve intendere la capacità di stimare il senso di distanza per diverse stimolazioni a distanze diverse in un contesto ambientale.

Relativamente all'impressione di distanza, l'intensità (soggettivamente corrispondente al **loudness**) gioca un ruolo diverso a seconda del contesto. Normalmente il ruolo dell'intensità diventa più importante quando si è in presenza di suoni non familiari. Per esempio, il ticchettio di un orologio all'interno e il suono di un treno che transita all'esterno della nostra abitazione produce una stima della distanza che è l'esatto contrario rispetto all'informazione d'intensità. Ma in un contesto diverso, lontano dal contesto familiare, la distanza dei rumori e dei suoni provenienti dall'ambiente è prevalentemente stimata in base all'intensità.

In condizioni di totale assenza di riverberazione (condizioni anecoiche), è possibile descrivere la dipendenza della distanza rispetto all'intensità per mezzo della **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.*

Come si vede in figura 11, la superficie quadrata più interna rappresenta l'intensità di riferimento della sorgente (1 watt/m^2). Poiché la superficie della sfera è proporzionale al quadrato del raggio, l'energia iniziale si distribuisce su una superficie che è quattro volte maggiore (vedi quadrato sulla sfera più esterna) rispetto a quella iniziale per un raddoppio della distanza. Allora l'intensità sarà nel nuovo punto $\frac{1}{4} \text{ watt/m}^2$ cioè 6 dB minore.

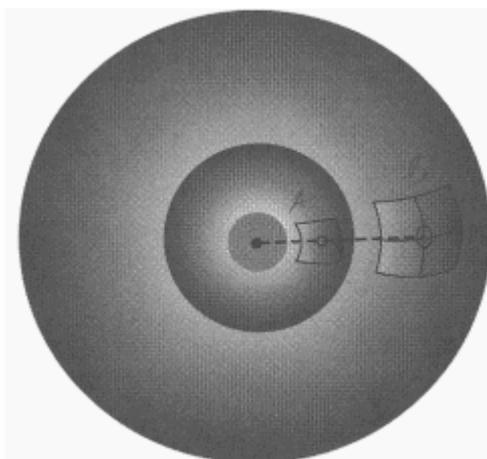


Fig. 11

In figura 12 sono riportati i valori in dB SPL (Sound Pressure Level) della stessa sorgente posta a 1, 2, 4 e 8 metri dall'orecchio destro di un ascoltatore. Data un'intensità di riferimento di 85 dB SPL a 1 metro di distanza, l'intensità diminuisce di 6 dB ad ogni raddoppio della distanza, in accordo con la legge inversa del quadrato.

Poiché la variabile psicoacustica che si correla meglio con l'intensità sonora è espresso in **sones**, la dipendenza dalla distanza viene valutata meglio secondo le curve isofone. Per fare un esempio, il range dinamico di un'orchestra può variare da 40 a 100 dB SPL, corrispondente a a 1 e 50 sones rispettivamente che rappresentano circa 5 o 6 volte il raddoppio della sensazione. Si può affermare allora che la variazione media corrispondente a un raddoppio della distanza è esprimibile attraverso una riduzione di circa 10 dB (cfr. fig 12)

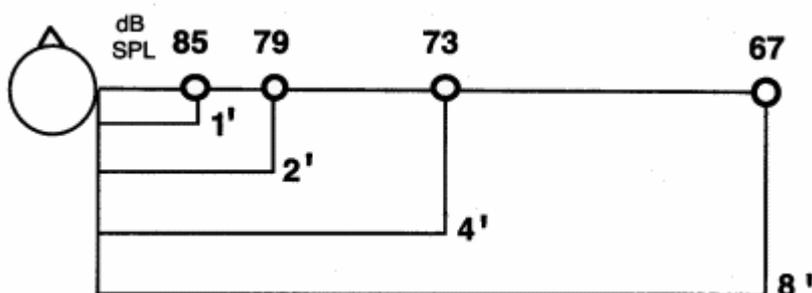


Fig. 12

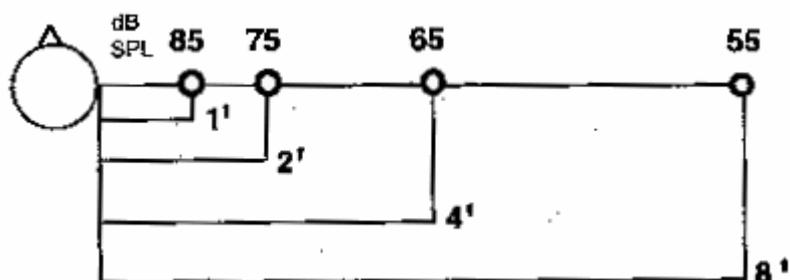


Fig.13

E' importante sottolineare che la variazione d'intensità può rappresentare efficacemente il fattore principale nella valutazione della distanza solo se non sono presenti altri fattori come per esempio la riverberazione e la presenza simultanea di altre sorgenti sonore. Poiché le condizioni normali d'ascolto sono più prossime a questa condizione, la valutazione della distanza ed eventualmente una sua percezione simulata non può essere basata unicamente sulla variazione dell'intensità sonora, sia che si usi la legge dell'inverso del quadrato (cfr. fig12) o la scala del loudness (cfr. fig. 13).

Attraverso misurazioni sul campo o facendo uso di programmi di simulazione è possibile calcolare l'intensità sonora che raggiunge un ipotetico ascoltatore posto in un luogo chiuso di cui si conoscano le proprietà geometriche e costitutive, nelle diverse condizioni: a) condizione anecoica, b) in presenza di riflessioni primarie (*early reflections o ER*) nei primi 80 ms, c) in presenza di riverberazione (*late reverberation o LR*) da 80 ms a diversi secondi. Generalmente in un contesto di tipo riverberante, la variazione della proporzione tra l'energia diretta e quella riflessa rappresenta un riferimento molto più forte dell'intensità nel giudicare la distanza della sorgente sonora.

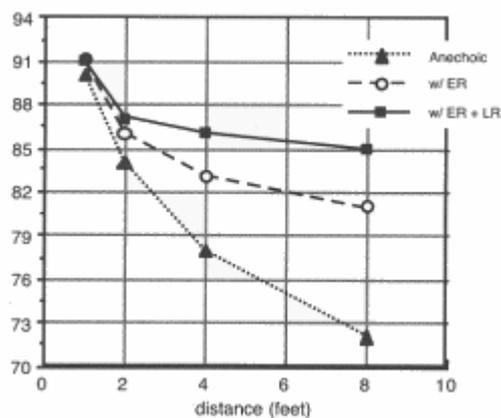


Fig. 14

In figura 14 sono riportati 3 diagrammi di riduzione dell'intensità misurata per le tre diverse condizioni descritte sopra. Si può allora notare la grande differenza di tra le varie condizioni, soprattutto quando la distanza è relativamente grande e che in presenza di riverberazione la diminuzione dell'energia sonora con la distanza è molto contenuta, come mostrato dall'andamento meno ripido della curva a tratto continuo (condizione ER + LR).

Quindi diverse condizioni ambientali determinano una percezione della distanza non univocamente. A questi fattori deve essere aggiunto anche il grado di aspettativa e familiarità della sorgente sonora. Ad esempio, in camera anecoica si può dimostrare da esperimenti fatti, che la valutazione della distanza di uno stimolo vocale riprodotto da un altoparlante posto a varie distanze è sempre funzione della pressione sonora. Se lo stimolo è invece rappresentato da una persona che parla a varie distanze e in 3 diversi modi (parlato normale, sussurrato e gridato) la distanza viene valutata in base alla modalità di emissione. In generale si osserva una sovra-estimazione per il gridato e una sotto-estimazione per il sussurrato. In figura 15 è riassunto questo tipo di esperimento.

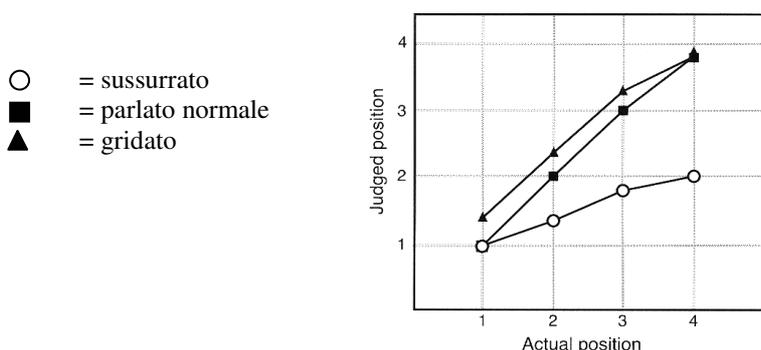


Fig. 15

La percezione della sensazione di distanza è dipendente anche dalla perdita selettiva dell'energia acustica. Generalmente, infatti, il fronte d'onda che raggiunge l'ascoltatore è quasi piano (il raggio di curvatura aumenta progressivamente con l'allontanarsi dalla sorgente) mentre è curvo in prossimità della sorgente. Questo è probabilmente il motivo per cui si nota una certa enfaticizzazione delle basse frequenze rispetto a quelle acute in tale condizione (*effetto di prossimità*).

Nell'ambito di un metro dalla sorgente questo fenomeno risulta molto evidente, ma a circa 2 metri non viene praticamente più rilevato poiché il fronte è meno incurvato. D'altra parte la cosa è spiegabile dal punto di vista psicoacustico. A brevissima distanza dalla sorgente l'intensità sale rapidamente e quindi il sistema percettivo è in una zona in cui le curve a eguale intensità (curve di Fletcher) mostrano una maggiore sensibilità alle basse frequenze.

Nonostante questo fenomeno sia facilmente sperimentabile, molto probabilmente nell'ascolto quotidiano è di marginale importanza.

In effetti, il fenomeno di gran lunga più appariscente è quello per cui, all'aumentare della distanza si percepisce una progressiva perdita del contenuto di alte frequenze. Ciò è dovuta principalmente all'effetto dell'umidità dell'aria e della sua temperatura e si può valutare attraverso il **coefficiente di assorbimento** dell'aria che rappresenta l'attenuazione del suono prodotta dalla viscosità e dal calore durante un singolo periodo di pressione sonora.

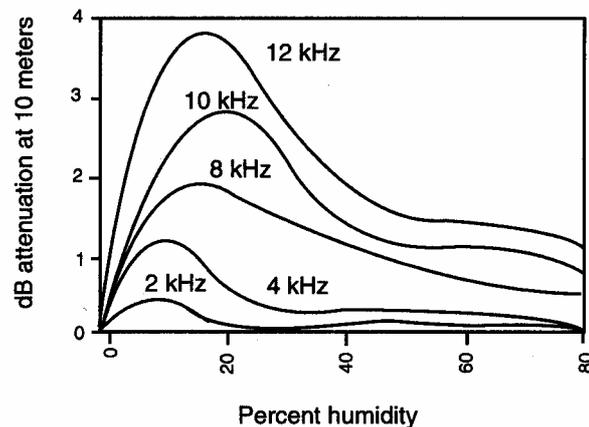


Fig. 16

La figura 16 mostra l'effetto di assorbimento del suono nell'aria in funzione dell'umidità relativa per diversi valori di frequenza.

Altri fattori che nelle condizioni di ascolto reale possono eventualmente contribuire all'attenuazione dell'energia acustica sono:

1. l'attenuazione dovuta alla presenza di ostacoli;
2. l'attenuazione dovuta alla pioggia, alla neve, alla nebbia;
3. l'attenuazione dovuta alla presenza di vento, turbolenza atmosferica e differenze di temperatura;
4. l'attenuazione dovuta alla presenza di vegetazione.

Nel caso 1) si avrà riflessione o diffrazione a seconda delle dimensioni rispetto alla lunghezza d'onda del suono. Quando la lunghezza d'onda (λ) è molto più piccola delle dimensioni dell'ostacolo quest'ultimo deflette il fronte d'onda secondo la legge della riflessione e quindi l'attenuazione sarà massima. Nel caso in cui λ sia invece confrontabile con le dimensioni dell'ostacolo, si forma una zona d'ombra acustica dietro l'ostacolo stesso., come mostrato in figura 17.

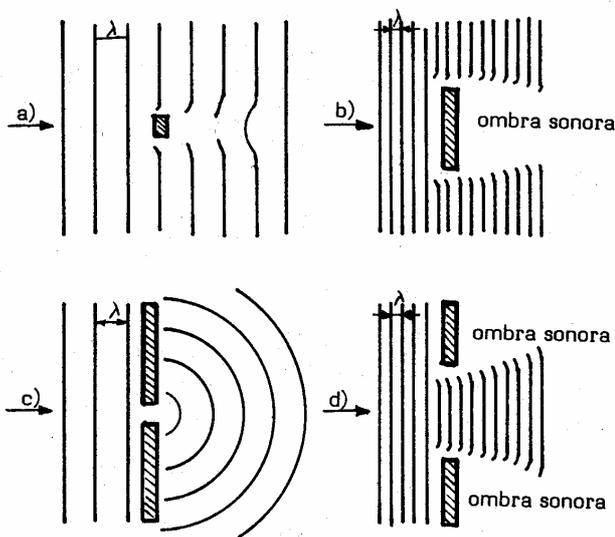


Fig. 17

Alla luce di queste considerazioni è quindi possibile definire una strategia di base per la simulazione della variazione di distanza apparente di una sorgente elettroacustica. Per produrre quindi un suono la cui distanza apparente diminuisce progressivamente occorre in primo luogo diminuire la sua intensità (ampiezza), applicare un filtro di tipo passa-basso ed eventualmente aggiungere energia di tipo riflessa (echi, riverberazione). In particolare è importante che l'energia diretta e quella di tipo riflessa si combinino tra loro secondo proporzioni diverse, come schematizzato in figura 18.

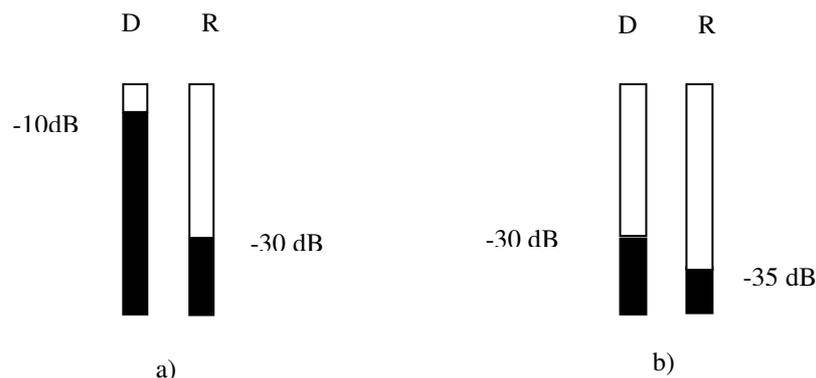


Fig. 18

Nella stessa figura sono infatti mostrati gli indicatori di livello per le due componenti del suono, diretto (D) e riverberato (R). In a) la sorgente è vicina: il suono diretto è predominante rispetto alla componente riverberata (-10 dB e -30 dB rispettivamente). In b) la sorgente è lontana: l'ampiezza complessiva è diminuita ed il rapporto tra l'energia diretta e quella riflessa si è ridotto sensibilmente (-30 dB e -35 dB rispettivamente).

In generale si può dire che per la simulazione di una variazione della sorgente in grandi spazi aperti può essere sufficiente l'impiego del parametro intensità e filtraggio mentre per la simulazioni di luoghi chiusi o genericamente delimitati può essere utile considerare solo l'intensità e la riverberazione come nell'esempio appena visto.

Per migliorare ulteriormente la simulazione della variazione di distanza occorre considerare un'ulteriore informazione che viene definita come **rapporto di riverberazione locale e globale**. Questo variabile ha senso quando la riproduzione avviene attraverso un sistema multiplo di altoparlanti. La riverberazione globale è equamente distribuita tra tutti gli altoparlanti mentre la riverberazione locale è quella di una particolare coppia di altoparlanti adiacenti. Quindi un suono può avere una debole riverberazione globale ma una forte riverberazione locale: ciò simula il caso della presenza di un'apertura in un largo spazio tra i due diffusori.

Il considerare questa variabile aiuta inoltre a ridurre l'effetto di mascheramento che si ha quando l'energia diretta e quella riverberante (unica) si equivalgono. In questo caso il mascheramento è tale da cancellare quasi completamente la localizzazione (azimut). Allora dividendo in due parti (locale e globale) la riverberazione e facendo in modo che la prima cresca con l'aumentare della distanza secondo la regola

$$\text{riverberazione locale} = 1 - 1/\text{distanza}$$

si migliora il senso di percezione della distanza e della localizzazione. Infatti man mano che la distanza aumenta la relazione tende al valore unitario. Cioè quando la sorgente è prossima all'ascoltatore, la riverberazione è equamente distribuita in tutti i canali. Come la sorgente si allontana, la riverberazione si concentra nella direzione della sorgente.

Percezione della Velocità della Sorgente Sonora (Effetto Doppler)

Nella realtà, se una sorgente e ascoltatore sono in moto reciproco nello spazio, la percezione dello spostamento da parte dell'ascoltatore è ulteriormente enfatizzato dall'**Effetto Doppler** descritto per la prima volta nel 1842 dall'astronomo C. Doppler e utilizzato per la prima volta nel campo della computer music da J. Chowning.

L'Effetto Doppler è una variazione di altezza del suono (pitch shift) causata dal moto relativo della sorgente e dell'ascoltatore. E' possibile sperimentare ogni giorno questo tipo di effetto al passaggio di un'ambulanza a sirene spiegate o in prossimità di un tracciato ferroviario. Mentre la sorgente si avvicina la frequenza del suono percepito è maggiore di quella emessa mentre quando si allontana la frequenza è minore. Questo è dovuto al fatto che gli intervalli dei fronti d'onda emessi dalla sorgente si ravvicinano tra loro quando la sorgente si approssima, mentre si diradano quando se ne allontana. Ciò porta quindi alla percezione di una lunghezza d'onda apparente che non coincide con quella originale causando quindi un'alterazione della frequenza percepita. (Cfr fig. 19). I punti S1, S2, S3 in figura 19 b rappresentano posizioni successive della sorgente in movimento. In un certo istante l'effetto Doppler sposta tutte le frequenze dello stesso intervallo logaritmico poiché la compressione/rarefazione dei fronti d'onda è una quantità percentuale e quindi vengono preservati i rapporti inter-frequenziali tra le componenti.

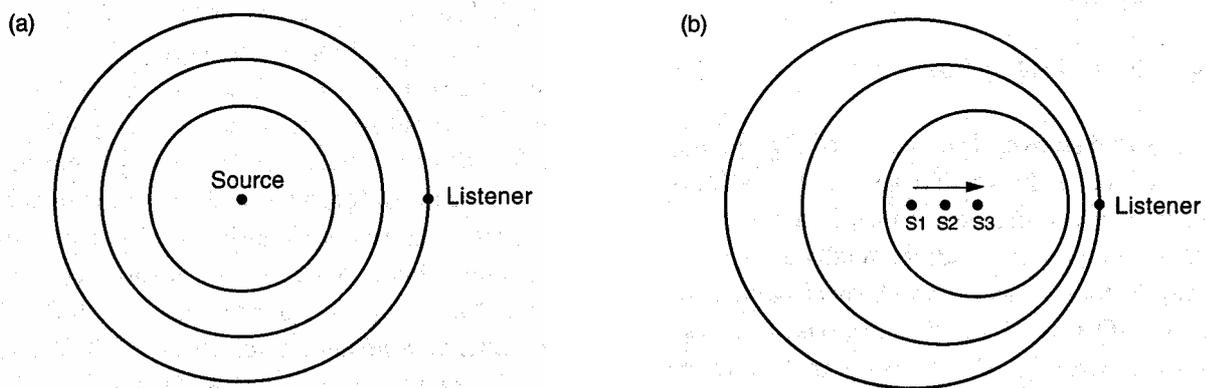


Fig. 19

La frequenza apparente è quindi funzione della velocità relativa tra sorgente e ascoltatore secondo la formula seguente:

$$F_a = F_o \times (v / v - V_s)$$

dove:

F_a = frequenza apparente

F_o = frequenza originale

v = velocità del suono (344 m/s)

V_s = velocità della sorgente

Per esempio una sorgente che si avvicina a 100 Km/h (\cong 28 m/s) si avrà:

$$F_a = F_o \times (344 / 344 - 28) = F_o \times 1.0886$$

Se $F_o = 440$ Hz si avrà $F_a \cong 479$ Hz, cioè un intervallo frequenziale intermedio tra un semitono temperato e una seconda maggiore.

Nel caso in cui la sorgente sia immobile e l'ascoltatore sia in movimento si ha:

$$F_a = F_o \times (v - V_a) / v$$

dove V_a = velocità dell'ascoltatore.

Se invece l'ascoltatore e l'osservatore sono entrambi in moto lungo la stessa retta si ha:

$$F_a = F_o \times (v - V_a) / (v - V_s)$$

E' opportuno ricordare che queste considerazioni si riferiscono alla *velocità radiale* relativa tra sorgente e ascoltatore il che significa considerare per esempio il moto della sorgente rispetto ad un centro (cfr. figura 20 a). In tal caso si ha un'effettiva variazione della distanza istantanea tra i due.

Al contrario, se una sorgente si muove circolarmente attorno ad un ascoltatore, poiché la distanza rimane costante (raggio della circonferenza), la velocità radiale è nulla e quindi l'effetto Doppler non si verifica (cfr. figura 20 b).

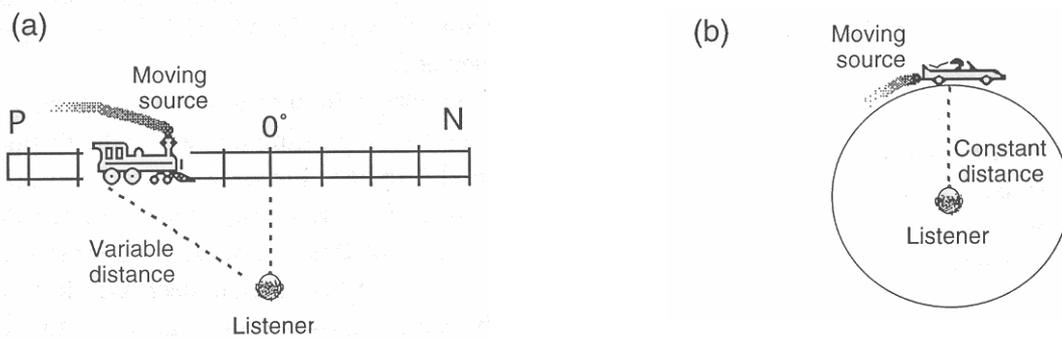


Fig. 20

Quando la velocità della sorgente $V_s = v$ (velocità del suono), quando l'ascoltatore è fermo, applicando la $F_a = F_o \times (v / v - V_s)$ si ha $F_a = \infty$. Allora tutte le onde viaggiano come un pacchetto compatto e arrivano assieme all'ascoltatore. Se $V_s > v$, le onde sonore saranno ricevute in ordine inverso rispetto all'ordine d'emissione. Se $V_s \gg v$, l'Effetto Doppler si traduce in quello che è comunemente noto come **onda balistica** o **sonic bang**, generalmente udibile a terra quando un aereo supera il muro del suono ad una distanza relativamente breve da questa.

Virtualizzazione della percezione spaziale del suono

Come già detto in precedenza, applicando un certo numero di principi di acustica e di psicoacustica, è oggi possibile attraverso le tecniche digitali di DSP, ricreare situazioni d'ascolto virtuali che tengano conto della localizzazione, del movimento nello spazio dell'ambientazione.

Per la sola localizzazione su piano si è già riferito della possibilità di avvalersi dei parametri IID e ITD.

Sebbene l'utilizzazione del modello ITD sia più aderente alla realtà, poiché il nostro sistema percettivo è in grado di apprezzare i ritardi generalmente sempre presenti nell'ascolto reale, nella pratica molto spesso la localizzazione viene solo affidata al modello IID.

Nella prassi della riproduzione stereofonica, è consuetudine consolidata utilizzare la tecnica del *panning* (ripartizione dell'energia acustica sui due canali L e R) per posizionare le varie sorgenti in punti diversi del panorama. La tecnica più brutale è quella in cui la distribuzione di energia è lineare. Per esempio, se si applica una funzione lineare decrescente al canale L e contemporaneamente la sua funzione complementare, si ottiene un'immagine acustica del tipo mostrata in figura 21.

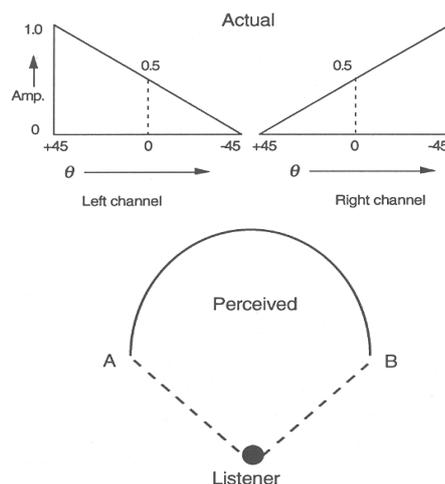


Fig. 21

Il criterio lineare viene applicato definendo i valori di ampiezza di A e B attraverso la semplice relazione :

$$A = \theta/\theta_{\max}$$

$$B = 1 - A$$

dove

θ_{\max} = angolo formato dalla congiungente tra l'ascoltatore e il punto A (es. canale L)

θ = angolo generico tra il punto medio tra A e B ed A.

Poiché l'intensità percepita secondo la legge di percezione dell'intensità si ha che:

$$I = \sqrt{A^2 + B^2}$$

Ciò significa che nel punto intermedio tra A e B i due valori d'intensità valgono entrambi 0.5. Allora per la relazione quadratica si avrà:

$$\sqrt{0.5^2 + 0.5^2} = \sqrt{0.25 + 0.25} = \sqrt{0.5} = 0.707$$

cioè una differenza di circa 3 dB.

In queste condizioni viene così a crearsi un buco al centro dal momento che il nostro sistema di percezione tende ad ascoltare il segnale in maniera più forte alle due estremità (punti A e B), dove in effetti il valore è unitario (0 dB).

Una valida alternativa a questo difetto è quella di applicare due curve complementari a *potenza costante*, tali che nel punto centrale producano un valore di intensità unitario (cfr. fig.22) Tali curve sono:

$$A = \sqrt{2}/2 \times [\cos(\theta) + \sin(\theta)] = 0.707 \times [\cos(\theta) + \sin(\theta)]$$

$$B = \sqrt{2}/2 \times [\cos(\theta) - \sin(\theta)] = 0.707 \times [\cos(\theta) - \sin(\theta)]$$

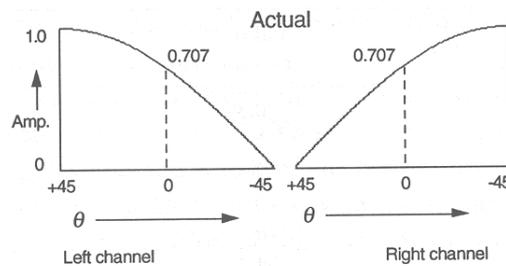
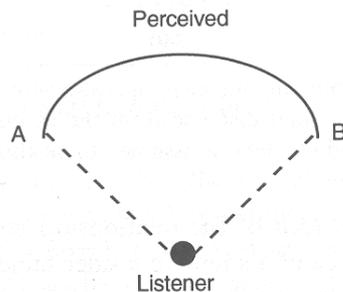


Fig. 22



In questo modo, l'ascoltatore viene a trovarsi al centro di un'ipotetica circonferenza e le traiettorie virtuali di spostamento distano costantemente il valore del suo raggio (Cfr. Fig. 22)

Percezione della componente verticale del suono

Come si è già visto in precedenza, la localizzazione del suono nel piano interaurale avviene principalmente per effetto di due meccanismi legati rispettivamente alla differenza di intensità e di tempo. Ma come si può intuire, essendo la propagazione di tipo sferico, esistono condizioni di rilevamento di tali grandezze anche per posizioni non planari. In altre parole, i valori di IID e ITD relativi ad un dato punto sul piano, possono coincidere con stessi valori riferiti ad un punto posto al di sopra, al di sotto o addirittura dietro l'ascoltatore. Questo fatto induce a pensare che devono esistere altri elementi di informazione per la percezione dell'elevazione in primo luogo e della localizzazione in senso generico.

La ricerca ha da tempo dimostrato che le componenti del suono al di sopra di circa 6 KHz vengono riflesse e filtrate in modo molto particolare dalla superficie esterna delle orecchie e che questa particolare conformazione è responsabile della percezione della componente di elevazione del suono. A tale informazione contribuiscono in modo sostanziale anche le spalle.

La caratteristica filtrante di tale complesso anatomico può essere riprodotta come risposta in frequenza del filtro digitale che rappresenta tale modello e viene comunemente detta "*Head-Related Transfer Function*" (*Funzione di trasferimento relazionata alla testa*) o più semplicemente **HRTF**. Naturalmente, esistono un numero infinito di HRTF, una per ogni punto nello spazio e più o meno diverse a seconda dell'ascoltatore.

In pratica, il suono della sorgente, prima di raggiungere il timpano viene sottoposto ad una modificazione spettrale e temporale che sono legati ai fenomeni di micro-riflessione prodotte dalle pieghe del padiglione uditivo. L'esperienza costante del fenomeno ha come conseguenza la capacità di associare a tali variazioni spettrali-temporali l'informazione posizionale della sorgente.

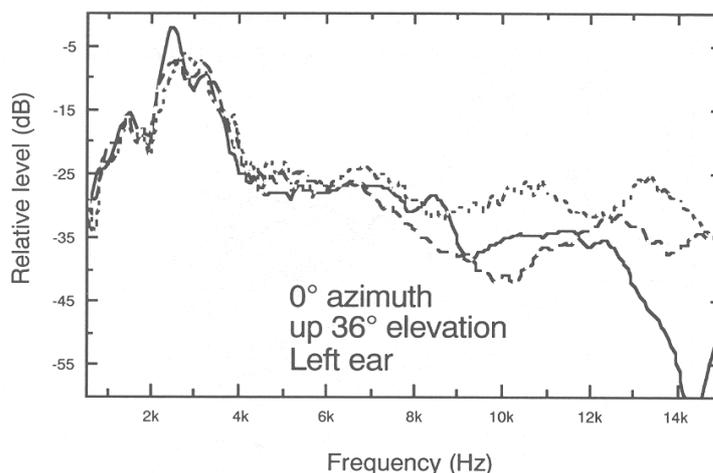


Fig. 23

La figura 23 mostra l'andamento di una funzione HRTF per tre diversi soggetti rispetto ad una sorgente posizionata a 0° gradi azimut, 36° di elevazione rilevata dall'orecchio sinistro.

Nonostante una certa omogeneità generale nell'andamento di tale risposta, si possono notare alcune marcate differenze che denotano la grande difficoltà di produrre misurazioni obiettive.

La conoscenza delle funzioni HRTF consente di riprodurre, con un certo grado di approssimazione, situazioni d'ascolto tridimensionali (3D-Sound) attraverso l'uso di due soli sistemi di diffusione del suono. In particolare, l'applicazione delle funzioni HRTF fornisce risultati soddisfacenti solo se il sistema di riproduzione approssima la condizione di *binauralità*. Tale condizione viene raggiunta

completamente quando la riproduzione avviene in cuffia o ancora meglio in una condizione per la quale il soggetto è posto in una camera anecoica, con la testa fissa in un'unica posizione e con sonde di riproduzione acustica inserite nei canali uditivi.

Questa particolare condizione ha dato luogo ad un certo sviluppo della tecnica di registrazione binaurale che consiste nella ripresa del suono attraverso l'uso di due microfoni all'interno di una testa artificiale.

L'uso della tecnica HRTF per la riproduzione di un suono 3D dipende dalla possibilità di disporre di un'adeguata e dettagliata descrizione delle funzioni di filtraggio. Alcuni centri di ricerca hanno prodotto *data-base* di funzioni HRTF con diversi gradi di risoluzione. In pratica tale funzioni vengono desunte campionando lo spazio attraverso la registrazione di impulsi temporali lungo una griglia di valori di azimut e di elevazione. Il suono viene ripreso da una testa artificiale ed elaborato matematicamente per fornire un numero finito di *risposte impulsive*. Tali risposte vengono poi utilizzate dal sistema di elaborazione (DSP) e riproduzione del suono per ricostruire un qualsiasi posizionamento attraverso la tecnica dell'interpolazione.

In figura 24 sono mostrati due manichini per la rilevazione di HRTF e per registrazioni binaurali rispettivamente KEMAR di Knowles Electronics e AACHEN Mod. HMS II di Head Acoustics

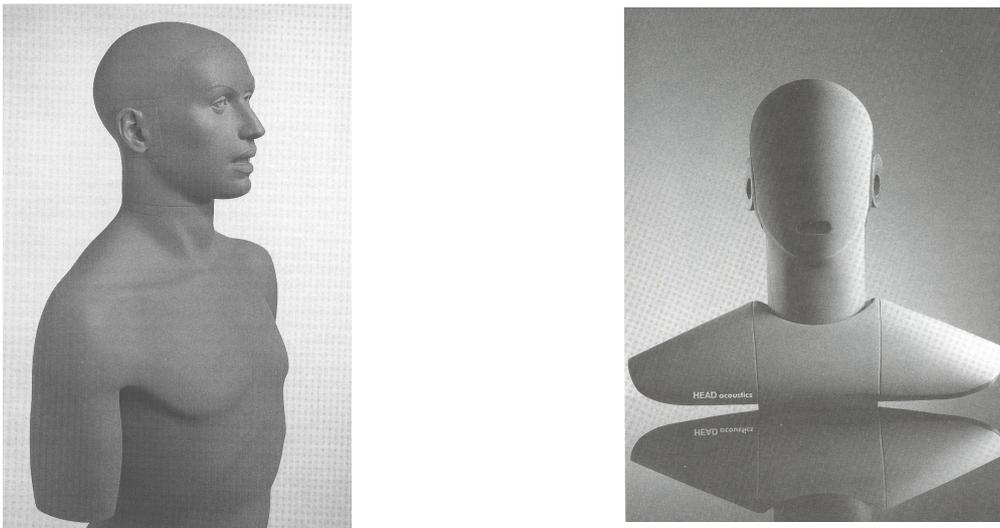


Fig. 24

La creazione di un data-base di HRTF può essere ottenuto attraverso sonde microfoniche poste nel canale uditivo di un individuo oppure impiegando una testa artificiale (*dummy-head*) come quelle mostrate in figura 24. In entrambi i casi viene registrato in camera anecoica un segnale che rappresenta - in prima approssimazione- la risposta del padiglione dell'orecchio esterno stimolato con un segnale impulsivo (segnale analitico) riprodotto da un altoparlante posto a circa 1 metro di distanza. La rilevazione comprende generalmente la registrazione simultanea della risposta destra (*risposta impulsiva binaurale*) e sinistra e viene ripetuta tante volte quante sono le posizioni relative tra sorgente e rilevatore che si desidera memorizzare.

Nella figura 25 viene schematizzato l'intero processo di acquisizione delle HRTF e del loro eventuale impiego.

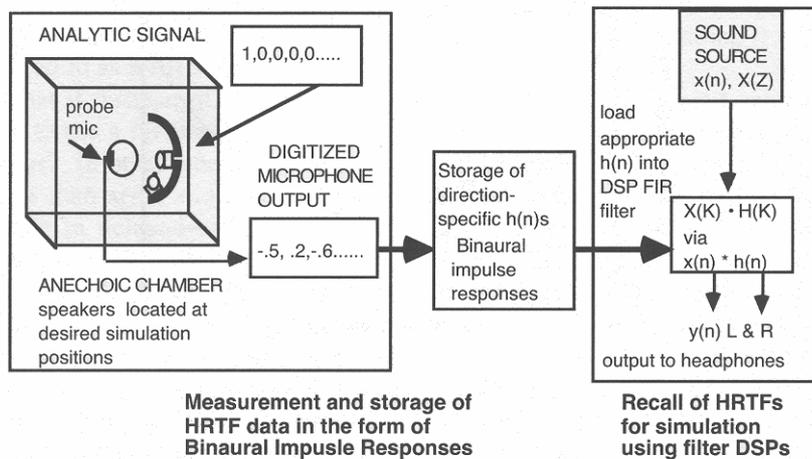


Fig. 25

In figura 26 è invece mostrato il grafico relativa ad una possibile risposta impulsiva registrata con tale metodo. La freccia indica il picco della risposta (valore massimo)

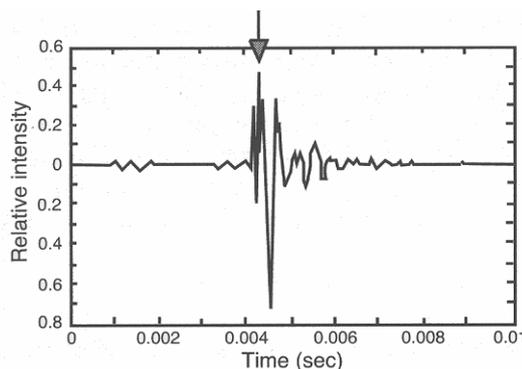


Fig. 26

La trasformata di Fourier (FFT) di tale segnale produce una risposta in frequenza del tipo visto in figura 23.

Poiché sono molti i fattori di non linearità che interessano tutto il sistema, la registrazione finale della risposta porta al suo interno elementi di errore che possono incidere sulla qualità della riproduzione del materiale audio al momento dell'utilizzazione da parte del DSP. Per questo motivo vengono impiegati metodi matematici al fine di ottimizzare la risposta attraverso la massimizzazione del rapporto segnale-rumore (S/N).

Un altro problema nella generazione dell'intero data-base è quello relativo all'equalizzazione di tutte le risposte, sia dal punto di vista temporale che frequenziale.

Dal punto di vista frequenziale una delle varie alterazioni spettrali da compensare è rappresentato dalla risonanza del canale uditivo che è presente nelle HRTF realizzate con sonde microfoniche ma che è assente quando vengono usate le teste artificiali. Nel primo caso la risonanza deve essere compensata una volta per tutte altrimenti in fase di utilizzazione (riproduzione binaurale) si produce una risonanza doppia (quella presente nella risposta HRTF e quella propria dell'utilizzatore della cuffia di riproduzione).

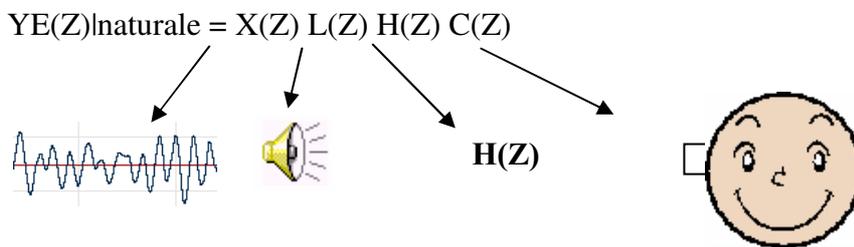
In generale appare chiaro che non è banale neutralizzare tutti gli effetti "distorcenti" che intervengono nella catena di acquisizione delle HRTF e nella catena di riproduzione virtuale.

Almeno teoricamente è possibile descrivere l'equalizzazione attraverso un processo di correzione della risposta per far coincidere la spazializzazione naturale e la sua copia virtuale.

Nel processo vengono coinvolti i seguenti elementi espressi nel dominio della variabile Z (frequenza complessa):

- A(Z)** : segnale analitico **M(Z)**: microfono sonda **C(Z)**: canale uditivo **L(Z)**: altoparlante
- HP(Z)** : cuffia **H(Z)** : HRTF naturale in campo libero
- RAW(Z)** : HRTF distorta **COR(Z)**: HRTF compensata per la riproduzione
- INV(Z)**: filtro inverso per la compensazione
- YE(Z)**: segnale in arrivo al timpano (in riproduzione in cuffia)
- YM(Z)**: segnale in arrivo al timpano (in fase di registrazione)
- X(Z)**: segnale da spazializzare

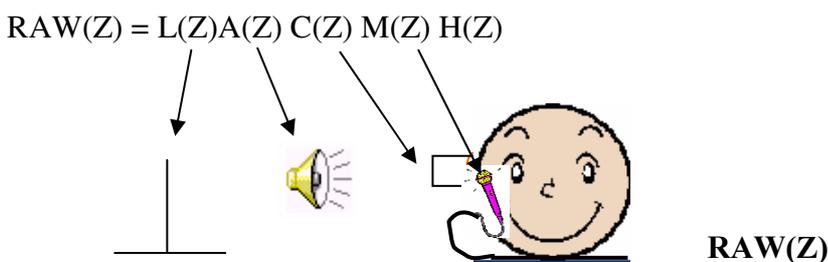
Ogni catena di oggetti coinvolti può essere rappresentata da un'equazione composta dai prodotti delle singole funzioni. Quindi nell'ascolto spazializzato naturale (reale) si avrà:



Nell'ascolto spazializzato virtuale si dovrà ottenere che:

$$YE(Z)|_{\text{virtuale}} = YE(Z)|_{\text{naturale}}$$

Allora nella fase di misura delle HRTF distorte si avrà:



Parallelamente deve essere misurata l'influenza della cuffia, del canale uditivo riproducendo il segnale analitico attraverso la cuffia stessa e captato attraverso il microfono sonda:

$$YM(Z) = A(Z) HP(Z) C(Z) M(Z)$$

Per ottenere la COR(Z) bisogna trovare prima il filtro inverso INV(Z) che la può equalizzare:

$$INV(Z) = \frac{1}{A(Z) HP(Z) C(Z) M(Z)}$$

A questo punto è possibile equalizzare e quindi correggere la RAW(Z) attraverso il prodotto:

$$COR(Z) = RAW(Z) INV(Z).$$

Sostituendo le espressioni di RAW(Z) e INV(Z) rispettivamente e semplificando i termini che compaiono contemporaneamente a numeratore e denominatore si ottiene:

$$COR(Z) = \frac{H(Z) L(Z)}{HP(Z)}$$

A questo punto per creare la sorgente virtuale è necessario eseguire il prodotto di **convoluzione** con il segnale che deve essere spazializzato X(Z) e successivamente esso transiterà attraverso la cuffia nel canale uditivo. Essendo la convoluzione un'operazione coincidente con il prodotto delle rispettive funzioni nella variabile Z si avrà allora (tenendo conto di tutti gli elementi coinvolti):

$$YE(Z)|virtuale = [X(Z) COR(Z)] [HP(Z) C(Z)]$$

Sostituendo il valore di COR(Z) trovato in precedenza si ottiene semplificando:

$$YE(Z)|virtuale = X(Z) L(Z) H(Z) C(Z) = YE(Z)|naturale.$$

Come detto in precedenza, questo processo è solo parzialmente riproducibile nella pratica per la difficoltà di realizzare il filtro inverso in maniera precisa. Alcuni processori DSP di commercio rendono disponibili pacchetti software per il calcolo di tali filtri, non senza evidenti difficoltà.

Virtualizzazione del movimento delle sorgenti e interpolazione

L'impiego della tecnica HRTF per la localizzazione virtuale del suono può essere estesa, non senza qualche difficoltà, alla simulazione del movimento della sorgente nello spazio.

Ciò significa che è possibile definire un percorso, una traiettoria qualunque nello spazio virtuale. La griglia di valori espressi dal numero totale di HRTF a disposizione non copre ovviamente tutti i possibili punti dello spazio ma può essere utilmente impiegata assieme ad una tecnica di **interpolazione lineare**. Questo significa che un opportuno software è in grado di individuare l'HRTF di un punto qualunque dello spazio mediando tra le posizioni più prossime disponibili nel data-base. Purtroppo, se la griglia di funzioni HRTF non è fitta, l'interpolazione lineare può portare molto spesso a risultati intermedi poco consistenti.

Ad esempio il programma di sintesi Csound (MIT-Ems) mette a disposizione un modulo (opcode) la cui sintassi è la seguente:

```
aleft, aright hrtfer asig, kAz, kElev, "HRTFcompact"
```

dove

aleft, aright = segnali sinistro e destro d'uscita

hrtfer = opcode

asig = segnale monoaurale da spazializzare

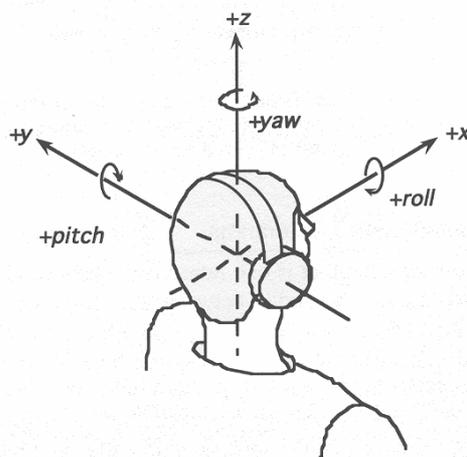
kAz = controllo dell'azimuth in gradi (valori positivi a destra, negativi a sinistra)

kElev = controllo dell'elevazione (valori positivi sopra il piano orizz., negativi sotto)

HRTFcompact = nome del data-base HRTF

L'uscita è quindi binaurale (in cuffia) ed è ottenuta attraverso un processo di convoluzione tra il segnale *asig* e le funzioni HRTF. Il controllo dell'azimuth e dell'elevazione può essere tempo-variante in modo da poter costruire una qualsiasi traiettoria spaziale virtuale impiegando la tecnica dell'interpolazione lineare.

Ancora più complesso è il caso di quando si desidera in tempo reale produrre una spazializzazione virtuale dinamica in base ai movimenti della testa. In questo caso occorre dotare il soggetto di un sistema di sensori xyz che rilevino in tempo reale la posizione istantanea come mostrato in figura 29.



yaw = imbardata
roll = rollio
pitch = beccheggio

Fig. 29

Una situazione analoga è data invece dalla necessità di mantenere invariata il posizionamento della sorgente virtuale, indipendentemente dal movimento della testa. In questo caso è ancora

indispensabile ricorrere allo stesso sistema di tracciamento e rilevamento della posizione mostrato in figura 29. La tecnologia attuale per questo tipo di tracciamento utilizza sensori di rilevamento posizionale di tipo elettromagnetico, ma possono anche essere di tipo ottico, meccanico ed acustico. In ogni caso l'elaborazione dei dati HRTF avviene sempre attraverso un meccanismo di interpolazione lineare tra le funzioni temporali (risposte impulsive) derivanti dal campionamento "spaziale".

Ad ogni risposta (o coppie di risposte) corrisponde una HRTF (in funzione della frequenza) che in sostanza rappresenta la maschera del filtro corrispondente da implementare.

I filtri sono quindi delle strutture FIR (Finite Impulse Response) entro cui deve transitare il segnale da spazializzare. In pratica i filtri si realizzano con una linea di ritardo con un numero di prese (*taps*), tante quanto è lunga la risposta in campioni. Per esempio, se ogni risposta impulsiva ha una lunghezza di 256 campioni, il filtro FIR dovrà avere la struttura mostrata in figura 30.

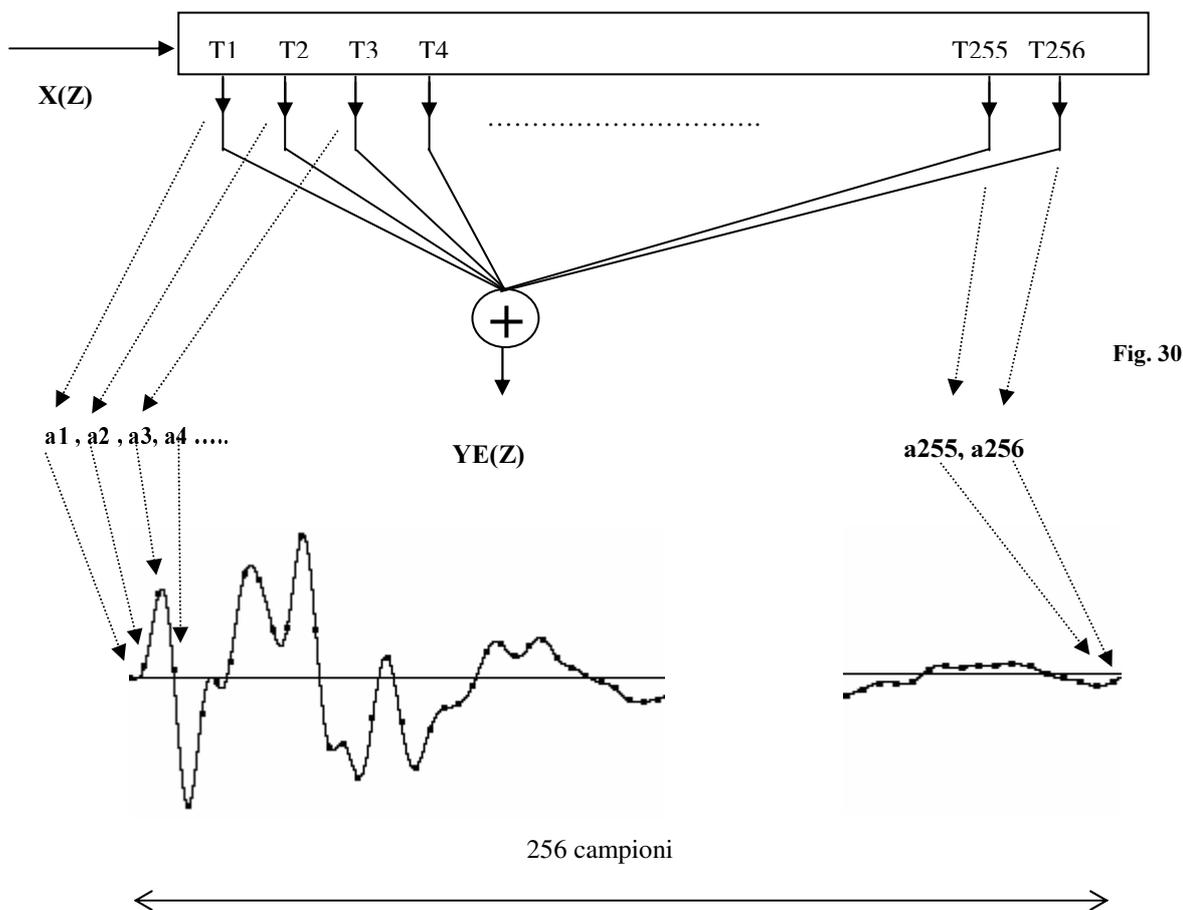


Fig. 30

Come si vede, ogni presa (tap) è ritardata di un valore pari all'inverso della frequenza di campionamento (in pratica T1, T2, ecc. coincidono temporalmente con la posizione dei campioni le cui ampiezze relative sono a1, a2, a3, ecc..). Nella figura 31 è riportato l'andamento dell'HRTF relativo alla risposta impulsiva di figura 29 di cui ne è la sua trasformata di Fourier.

Dovendo individuare un punto dello spazio o una traiettoria spaziale compresa tra due punti conosciuti, ciascuno rappresentato da una specifica HRTF, il meccanismo di interpolazione combina linearmente le due risposte temporali relative. Ovviamente ci si aspetta che la trasformata di Fourier del punto intermedio ottenuto sia coerentemente il punto intermedio tra le relative risposte in frequenza. In realtà, se le risposte sono molto diverse tra loro, i valori intermedi ottenuti attraverso la combinazione delle risposte temporali porta a funzioni HRTF errate.

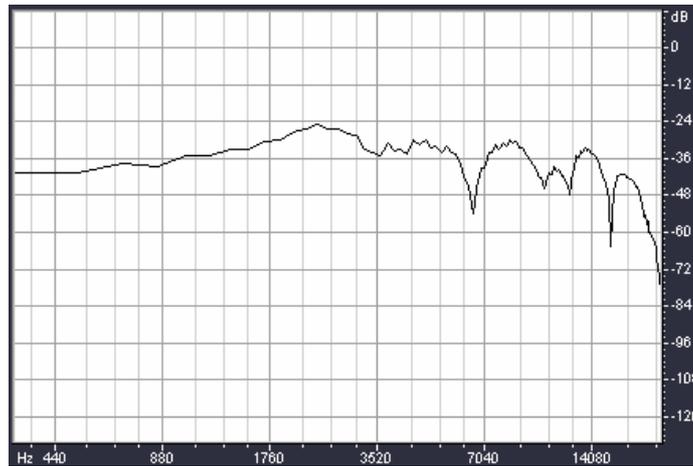
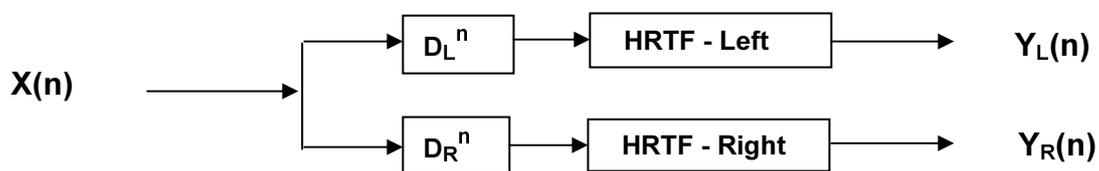


Fig. 31

Ciò porta a concludere che è importante disporre di un data-base di campionamento dello spazio molto dettagliato in modo tale da avere risposte di punti adiacenti non molto dissimili tra loro. La tecnica di spazializzazione 3D attraverso le funzioni HRTF può essere applicata con una variante che scinde l'informazione temporale da quella frequenziale. Ovviamente, occorre elaborare matematicamente le funzioni registrate ottenendo così un insieme di funzioni HRTF artificiali o sintetiche alle quali va aggiunta l'informazione specifica temporale. Lo schema di applicazione finale di quest'ultima tecnica è mostrato in figura 32.



Il segnale $X(n)$ da spazializzare è inviato a due rami di elaborazione posti in parallelo. $D_{L/R}$ sono i ritardi (sinistro e destro) per un particolare valore di ITD mentre HRTF -Left/Right sono HRTF sintetiche derivate da HRTF misurate dopo il procedimento matematico che impongono una modificazione spettrale per una certa posizione specificata.