

PRINCIPI DI ANALISI NUMERICA DEI SEGNALI*

Vengono presentati i fondamenti del campionamento dei segnali e le principali tecniche di elaborazione numerica normalmente impiegate per il trattamento digitale di segnali acustici, quali il calcolo del livello sonoro in decibel, con diverse costanti di tempo, l'analisi in frequenza tramite FFT, il filtraggio per bande di frequenza, la convoluzione dei segnali. Sono poi illustrate le principali applicazioni di tali tecniche per l'esecuzione di misure fonometriche, per la determinazione della risposta all'impulso, per il calcolo del tempo di riverberazione e di altri parametri utilizzati per la valutazione della qualità acustica degli ambienti. Un esempio pratico indica i passi che andrebbero seguiti operativamente nel corso di una misurazione.

16.1 Dai segnali analogici ai segnali digitali

La digitalizzazione di un segnale elettrico consiste nella misurazione della sua tensione in modo ripetuto nel tempo e nella memorizzazione della lunga sequenza di valori numerici che derivano da tali misurazioni. Nella grandissima parte dei casi, la tensione viene misurata ad intervalli temporali costanti. Trattando di segnali acustici, o segnali audio in generale, convertiti in segnali elettrici, il rateo di campionamento è sempre pari a diverse decine di migliaia di campioni al secondo.

Affinché il procedimento di digitalizzazione conservi senza artefatti l'intera informazione posseduta dal segnale analogico originario, occorre rispettare il teorema di Nyquist/Shannon (http://en.wikipedia.org/wiki/Nyquist-Shannon_sampling_theorem): il rateo di campionamento deve sempre essere almeno doppio della frequenza massima contenuta nel segnale in esame. Nelle normali applicazioni audio, essendo presenti segnali che possono anche superare leggermente 20 kHz, occorre impiegare una frequenza di campionamento di almeno 40 kHz, e il valore più usato si è di fatto attestato a 44100 Hz (CD audio). Tuttavia, al fine di ottenere una sempre maggior fedeltà, sono entrate nell'uso anche frequenze di campionamento standard più elevate: 48 kHz, 96 kHz, 192 kHz.

In pratica, dunque, a valle del campionamento si determina il valore numerico della tensione del segnale analogico originario ad istanti ben fissati, e si perde ogni informazione sul valore effettivo del segnale fra un istante di campionamento ed il successivo.

Anche il valore numerico misurato è discretizzato a sua volta, in quanto il dispositivo che effettua la lettura numerica del valore di tensione (un convertitore analogico-digitale, A/D o ADC) opera con precisione finita, solitamente rappresentando il valore di tensione con un numero intero e relativo segno, compreso in un intervallo di valori che va da 0 ad un valore massimo, determinato dal «numero di bit» impiegato per rappresentare tale valore numerico in codice binario. Ad esempio, un convertitore A/D a 16 bit produce un valore numerico intero compreso nel campo che va da -32768 e $+32768$ (comprendente 2^{16} valori interi).

* Angelo Farina.

Attualmente gran parte dei convertitori A/D impiegati opera con formato numerico a 24 bit, e dunque nel campo compreso fra $- 8388608$ e $+ 8388608$. A tali valori estremi corrispondono la tensione minima e massima che il convertitore è in grado di accettare in ingresso, normalmente pari a qualche volt.

Ovviamente per campionare segnali di piccola ampiezza per esempio provenienti dai microfoni, occorre amplificare e condizionare opportunamente il segnale analogico prima di applicarlo al convertitore, e variando il guadagno di tale sistema si può adattare il campo di valori numerici prodotti dal convertitore al campionamento di segnali acustici molto deboli o più elevati. In ogni caso, anche in termini di ampiezza il segnale viene approssimato con il valore intero più prossimo al vero valore del segnale analogico nell'istante di campionamento.

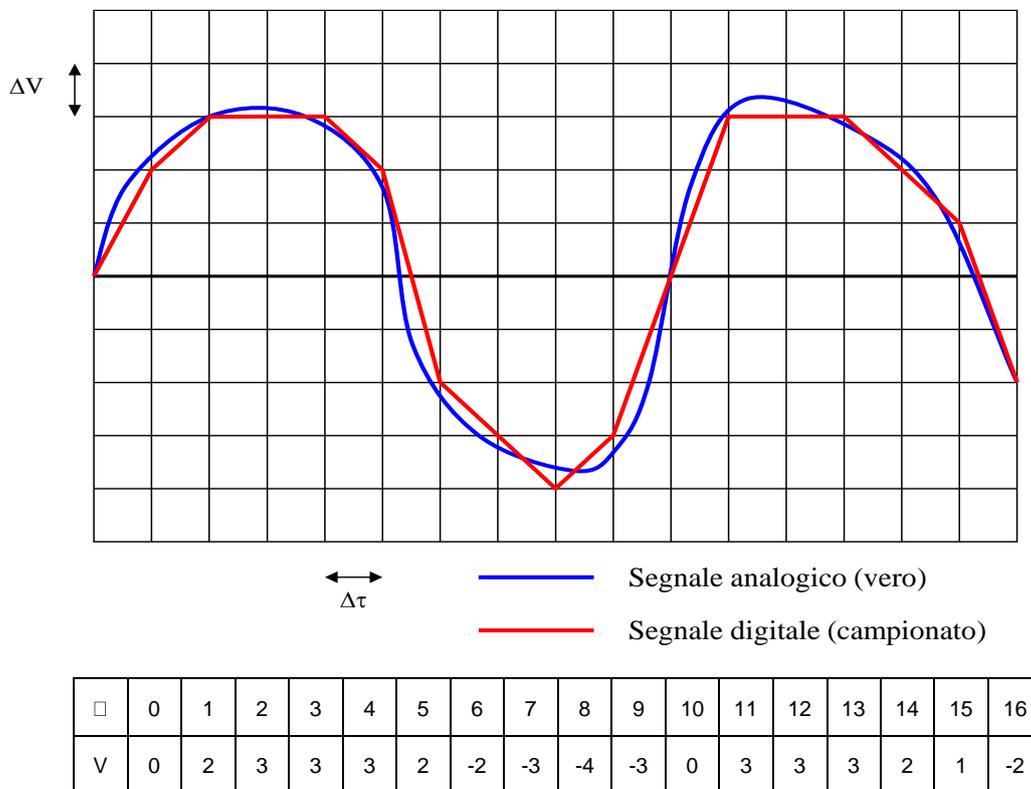


Fig. 16.1.1 Digitalizzazione di un segnale analogico (discretizzazione sia nel tempo, sia in ampiezza).

In sostanza, come mostrato in Fig. 16.1.1, al segnale analogico che varia con continuità sia nel tempo, sia in ampiezza, si sostituisce un segnale digitale che cambia per passi discreti anch'esso sia nel tempo, sia in ampiezza, risultando pertanto vincolato a giacere sui nodi di una griglia. Come mostrato in figura, la forma d'onda digitale risulta, in generale, non perfettamente aderente alla forma d'onda analogica, e pertanto il segnale digitalizzato presenta sempre un errore sistematico, chiamato «errore di quantizzazione» (o «errore di

campionamento»). L'entità di tale errore si riduce all'aumentare della frequenza di campionamento e della *precisione* del convertitore. Questa però non arriva mai a limitare l'errore al valore minimo teorico; per esempio per un convertitore a 24 bit questo è $1/8388607$, ma la maggior parte dei convertitori a 24 bit attualmente disponibili ha una precisione corrispondente a soli 20 o 21 bit. Infatti il «rumore termico» affligge inevitabilmente i 2 o 3 bit meno significativi del numero binario prodotto dal convertitore.

La precisione del convertitore influenza in modo diretto la possibilità di campionare un segnale con dinamica elevata, intesa come variazione fra suoni molto intensi e molto deboli. È chiaro infatti che il suono con il massimo valore istantaneo di pressione sonora deve comunque risultare compreso, con un adeguato margine di sicurezza, entro il valore di tensione massima gestito dal convertitore: se questo limite non viene rispettato, la forma d'onda risulta afflitta da una sistematica limitazione dei valori estremi, fenomeno noto come *clipping*, che va assolutamente evitato. Per tale ragione, considerato che certi tipi di rumore hanno un elevato valore del fattore di cresta, occorre solitamente regolare la preamplificazione del segnale, prima del convertitore, in modo che resti un sufficiente margine (*headroom*) fra il valore RMS del segnale sonoro e il limite di clipping, margine solitamente assunto pari o superiore a 24 dB.

Segnali molto deboli possono invece venire afflitti dal rumore di fondo termico che, come accennato, contamina solitamente i 2-3 bit meno significativi del valore numerico misurato. Volendo anche in questo caso mantenere un adeguato margine di sicurezza, tipicamente pari ad almeno 15 dB, si scopre che, anche utilizzando un convertitore A/D dell'ultima generazione e di eccellente qualità, il range dinamico disponibile non può di fatto superare 110 dB. Si tratta comunque di una escursione sufficientemente estesa, almeno per la maggior parte delle applicazioni audio.

Tornando agli errori causati dal campionamento nel tempo, occorre tenere conto della relazione fra la massima frequenza teoricamente raggiungibile sulla base del teorema di Shannon, «frequenza di Nyquist», pari a metà della frequenza di campionamento, e la massima frequenza effettiva del segnale da campionare. Avvicinandosi alla frequenza di Nyquist, l'errore di quantizzazione cresce, e inoltre si entra in una regione dove l'elettronica analogica di condizionamento del segnale sta operando un pesante filtraggio passa-basso. Occorre evitare di superare il limite posto dal teorema di Shannon: se il campionatore si trovasse ad operare con un segnale a frequenza superiore alla frequenza di Nyquist, si verificherebbe un fenomeno di grave distorsione del segnale, noto come *aliasing*. Il segnale di frequenza superiore a quella di Nyquist, per esempio di 6 kHz, verrebbe trasformato in un segnale spurio, nel dominio del segnale digitalizzato, ad una frequenza *specchiata*, sotto la frequenza di Nyquist della stessa entità. Se, ad esempio, si opera con una frequenza di campionamento di 48 kHz, la frequenza di Nyquist è 24 kHz. Se il segnale contiene una frequenza di 30 kHz, nel segnale digitalizzato comparirà una frequenza spuria di 18 kHz (di 6 kHz inferiore alla frequenza di Nyquist).

Il problema dell'*aliasing* viene evitato prevedendo, prima del convertitore, un circuito analogico di filtraggio passa-basso, che attenui progressivamente l'ampiezza del segnale, partendo da alcuni kilohertz prima della frequenza di Nyquist. In teoria, al di sotto della frequenza di taglio del filtro il segnale non dovrebbe subire alcuna alterazione, ma in pratica, avvicinandosi alla frequenza di taglio, cominciano a manifestarsi inevitabili distorsioni e

rotazioni di fase. Questo è il principale motivo che ha spinto ad un sistematico incremento della frequenza di campionamento: in linea di principio, la frequenza di 44100 Hz è già perfettamente sufficiente per la digitalizzazione dei segnali audio, considerato che la percezione uditiva umana non supera il valore di 20 kHz. Ma in pratica questo costringeva a realizzare un filtraggio anti-aliasing molto ripido, avendo a disposizione un intervallo di frequenza di poco più di 2 kHz (da 20000 a 22050 Hz) per variare il guadagno del filtro da 1 a 0. Un filtro così ripido manifesta gravi ripercussioni anche a frequenze perfettamente udibili, e questa è la principale causa di criticità per il suono digitale del CD-audio, fin da quando venne introdotto all'inizio degli anni Settanta.

Già il passaggio alla frequenza di campionamento di 48 kHz (registratori DAT, 1987), che potrebbe sembrare un incremento limitato, porta il filtro anti-aliasing a lavorare in condizioni molto meno critiche, perché dispone di un intervallo di frequenza praticamente doppio (da 20000 a 24000 Hz) per operare il *taglio*. È comunque ovvio che solo col passaggio alla frequenza di campionamento di 96 kHz è stato eliminato completamente il problema della distorsione del segnale a frequenza elevata, prossima alla frequenza di Nyquist, in quanto essa è divenuta uguale a ben 48000 Hz, mentre il segnale audio non supera 20 kHz. L'uso della frequenza di campionamento di 192 kHz, invece, non è solitamente utile per il campionamento di normali segnali audio, e diventa necessario solo nel caso di studio di fenomeni ultrasonici, sebbene venga impiegato per alcune registrazioni per audiofili, dove si punta alla massima fedeltà tecnicamente possibile.

In conclusione, allo stato attuale, i più comuni formati di registrazione digitale dei suono sono i seguenti:

- CD-audio (44 kHz, 16 bit): formato obsoleto e di qualità appena accettabile, anche se molto diffuso;
- DAT (48 kHz, 16 bit): formato obsoleto, qualità accettabile per applicazioni con dinamica limitata a circa 70 dB;
- DVD-video (48 kHz, 24 bit): formato molto diffuso per applicazioni multimediali, qualità accettabile per applicazioni ad ampia dinamica (sino a 110 dB);
- DVD-audio (96 kHz, 24 bit): formato poco diffuso per applicazioni multimediali, ottimale in termini di minimizzazione dell'errore e fedeltà anche per applicazioni ad ampia dinamica (sino a 110 dB) e frequenza elevata (sino a 30 kHz);
- Blue-Ray-Disc (192 kHz, 24 bit): formato finora poco diffuso (la gran parte dei Blue-Ray sono tuttora a 48 kHz), ma utilizzabile anche per lo studio di fenomeni ultrasonici (sino a circa 70 kHz).

Infine, va segnalato che per l'elaborazione del segnale acustico su personal computer ha preso piede l'utilizzo della rappresentazione numerica in virgola mobile (solitamente in formato IEEE a 32 bit). Allorché infatti al segnale registrato digitalmente si applicano alcune elaborazioni specifiche, la sua dinamica può risultare espansa: un formato in virgola mobile diventa l'unico modo per evitare il fenomeno del clipping, pur mantenendo integro l'intero campo dinamico disponibile. Sono pertanto comuni registrazioni digitali, alle varie frequenze di campionamento sopra indicate, in cui i valori numerici sono memorizzati in virgola mobile nel cosiddetto formato «float-32».

16.2 Tecniche di filtraggio digitale applicate a segnali audio digitali

Nel dominio dei segnali digitalizzati, la maggior parte delle operazioni di filtraggio può essere operata con facilità; in pratica, si opera su una lunga sequenza di numeri con segno. Particolarmente semplice risulta poi l'esecuzione di filtri numerici se si tratta di emulare sistemi fisici lineari e tempo-invarianti.

In particolare, si sono affermate due tecniche di filtraggio numerico molto semplici, note come filtraggio FIR (*Finite Impulse Response*) e filtri IIR (*Infinite Impulse Response*) (Oppenheim, Schaffer, 1975). Il filtraggio FIR è il più semplice dei due: ogni campione filtrato in uscita $y[i]$ è ottenuto come somma dei prodotti di N campioni del segnale originario in ingresso $x[j]$, moltiplicati per N coefficienti di filtraggio $h[j]$, a partire dal campione in ingresso corrente $x[i]$ e andando *indietro* nel tempo di N campioni:

$$y[i] = \sum_{j=0}^{N-1} x[i-j] \cdot h[j] . \quad (16.2.1)$$

Con questa tecnica è possibile emulare tutti i classici filtri analogici (passa-alto, passa-basso, passa-banda ecc.), nonché la risposta di sistemi fisici reali, come la riverberazione in un ambiente, la propagazione del suono attraverso un divisorio, la schermatura di una barriera ecc. L'accuratezza ottenuta è elevatissima, e la determinazione dei coefficienti di filtraggio $h[j]$ è relativamente facile, sia progettando il filtro nel dominio della frequenza, sia misurando un sistema fisico reale.

La sequenza dei coefficienti di filtraggio $h[j]$ è la «risposta all'impulso» del sistema, ed è anch'essa un suono che può essere ascoltato e processato. Di fatto, è il suono che si sente allorché all'ingresso del sistema viene applicato un impulso di brevissima durata (un solo campione del segnale digitalizzato), di ampiezza pari al fondo scala del sistema digitale, segnale detto « δ (delta) di Dirac» (benché sia solo un'approssimazione numerica della teorica distribuzione δ di Dirac).

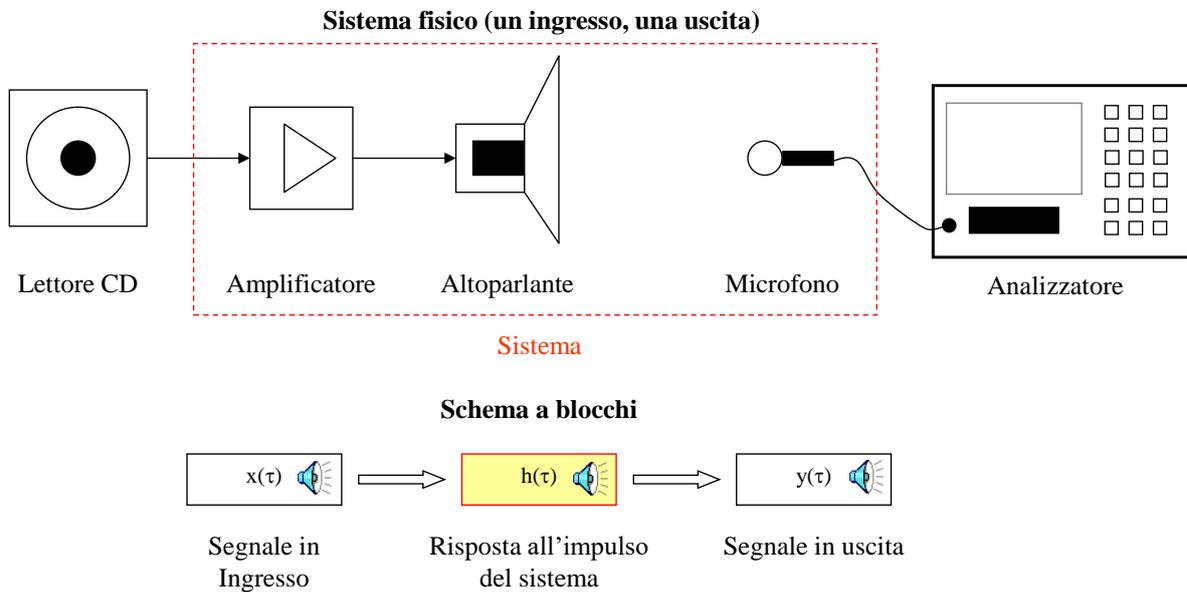


Fig. 16.2.1 Sistema lineare tempo-invariante ad un ingresso e una uscita.

Una volta determinata la risposta all'impulso di un sistema lineare e tempo-invariante, il suo effetto acustico può essere emulato fedelmente con l'operazione di filtraggio FIR descritta dalla (16.2.1), operazione che viene definita matematicamente col termine di «convoluzione», e che può venire applicata indifferentemente a qualunque tipo di segnale in ingresso, sia esso musica, parlato, rumore di un macchinario ecc. La linearità del sistema assicura che il filtraggio applicato resterà lo stesso qualunque sia il segnale che attraversa il sistema. Il concetto di sistema lineare tempo-invariante, con un ingresso e una uscita, è schematizzato in Fig. 16.2.1.

Il filtraggio FIR (convoluzione) è di uso così frequente che si è consolidata una rappresentazione compatta della relazione (16.2.1), rappresentando l'operatore convoluzione col simbolo asterisco (*):

$$y[i] = x[i] * h[j] \quad . \quad (16.2.2)$$

Gran parte dei processori digitali sono attualmente ottimizzati per l'esecuzione di questa operazione, che pertanto può essere considerata una sorta di mattone elementare di qualsiasi elaborazione numerica del suono.

Il limite del filtraggio digitale con tecnica FIR consiste nel fatto che per sistemi con estesa memoria, per esempio la risposta acustica di una cattedrale, occorre utilizzare un filtro con una risposta all'impulso molto lunga: il valore di N può arrivare a parecchie centinaia di migliaia di campioni. Questo richiede una elevata occupazione di memoria, che su processori di vecchia generazione assorbiva una quota cospicua della potenza di calcolo disponibile.

Va tuttavia sottolineato che l'evoluzione dei processori, da un lato, e lo sviluppo di particolari algoritmi di *convoluzione veloce*, descritti più avanti nel par. 16.4, hanno consentito già da alcuni anni il superamento di questi limiti (Torger, Farina, 2001). Viceversa, quando tali capacità di elaborazione veloce non esistevano, era stata sviluppata una tecnica di filtraggio alternativa, il filtraggio IIR, che consente di ottenere un filtraggio quasi altrettanto fedele del FIR, utilizzando però un numero di coefficienti di filtraggio molto inferiore. L'accorgimento consiste nel reimpiegare per il calcolo del campione filtrato corrente $y[i]$, non solo i campioni in ingresso $x[i]$, ma anche i precedenti campioni filtrati $y[i-1]$, $y[i-2]$ ecc. In pratica, una formula ricorsiva di filtraggio, in cui il campione *nuovo* è calcolato come combinazione lineare di un certo numero N di campioni in ingresso e di un altro numero M di campioni in uscita precedenti:

$$y[i] = \sum_{j=0}^{N-1} x[i-j] \cdot b[j] - \sum_{j=1}^{M-1} y[i-j] \cdot a[j] . \quad (16.2.3)$$

Si noti che il contributo dei termini $y[i]$ è preso, convenzionalmente, col segno negativo, e che i coefficienti di filtraggio dei campioni in uscita e in ingresso sono denominati, rispettivamente, $a[j]$ e $b[j]$. In tal caso, le sequenze di coefficienti di filtraggio non sono più suoni udibili, e scompare la possibilità di determinarli applicando all'ingresso del sistema uno speciale segnale di test, come invece accadeva con il segnale delta di Dirac per la determinazione dei coefficienti di filtraggio FIR, $h[j]$ (risposta all'impulso).

In generale, la determinazione delle sequenze dei coefficienti di filtraggio IIR, $a[j]$ e $b[j]$, è assai complessa e per nulla intuitiva, ed esistono interi volumi dedicati al problema della loro determinazione (Smith III, 2007). Va anche precisato che il filtraggio IIR è una rappresentazione non perfettamente accurata del comportamento dei sistemi fisici, e dà spesso luogo a errori non trascurabili, specie in termini di errori di fase. Tuttavia, in passato il filtraggio IIR è stato molto utilizzato, grazie al fatto che richiede un minor numero di coefficienti di filtraggio. Quando la potenza di calcolo necessaria alla computazione delle formule (16.2.1) e (16.2.3) risultava sostanzialmente proporzionale al numero di moltiplicazioni da eseguire, si riteneva che effettuare il filtraggio con meno coefficienti fosse più efficiente.

La successiva evoluzione delle architetture di elaborazione digitale, da un lato, e lo sviluppo di avanzate tecniche di filtraggio FIR veloce, dall'altro, hanno spostato i termini del problema, al punto che oggi è in generale possibile effettuare filtri FIR con centinaia di migliaia di coefficienti, utilizzando una potenza di calcolo inferiore a quella richiesta da un filtraggio IIR, con poche decine di coefficienti.

Infine, vanno segnalati due ulteriori problemi tipici del filtraggio IIR: il rischio di instabilità e il *rumore di calcolo*. Il primo deriva dal fatto che, se i coefficienti $a[j]$ sono troppo grandi, il sistema rischia il fenomeno del *feedback*. In pratica, il segnale tende ad aumentare di ampiezza in modo incontrollato, e poiché esso rientra nel sistema, lo porta a divergere progressivamente fino al raggiungimento di valori fuori scala (*overflow*). Per evitare il feedback, occorre condizionare opportunamente i valori dei coefficienti di filtraggio, in modo da garantire la stabilità del sistema di filtraggio, anche se stimolato da segnali in ingresso molto grandi.

Il secondo problema è invece tipico dell'implementazione su processori operanti con matematica intera (a *virgola fissa*): stante la natura ricorsiva del filtraggio IIR, il rumore di quantizzazione causato dall'elaborazione con precisione finita tende ad essere amplificato, e il segnale elaborato presenta una *base* di rumore maggiore di quello del segnale in ingresso, che diventa particolarmente udibile in assenza di segnale in ingresso. Ciò si verifica, ma in minore misura, anche con il filtraggio FIR, se si utilizza un processore in virgola fissa; tuttavia, di solito i processori di basso costo, essendo ottimizzati proprio per il filtraggio FIR, presentano accorgimenti circuitali tali da limitare l'accumulo del rumore di quantizzazione, e il filtraggio risulta assai meno rumoroso.

16.3 Dai segnali campionati ai livelli sonori in decibel

Si consideri il caso pratico di dover elaborare un segnale audio campionato al fine di emulare il comportamento di un fonometro, un comune misuratore di livello sonoro. Se il microfono dello strumento ha una sensibilità S di 50 mV/Pa, un segnale di calibrazione di 94 dB a 1 kHz, cui corrisponde la pressione sonora efficace di 1 Pa (rms), darà luogo a un segnale elettrico sinusoidale di ampiezza $50 \cdot \sqrt{2}$ mV, uguale quindi a 70,7 mV.

Il segnale viene poi campionato ad una frequenza di 48 kHz, tramite un convertitore A/D a 16 bit, utilizzando uno stadio di preamplificazione configurato in modo che al fondo scala digitale (limite di clipping) corrisponda un livello sonoro di picco di 100 dB, quindi una pressione di 2 Pa. In pratica, il valore di fondo scala corrisponde ad una tensione $V_{\max} = 100$ mV, che il convertitore trasforma nel numero intero 32768. Catturando uno spezzone di 512 punti si ottiene il segnale digitalizzato di Fig. 16.3.1, con sulle ascisse il numero del campione (da 0 a 511), e in ordinata il corrispondente valore di ampiezza come numero intero a 16 bit, compreso, dunque, tra ± 32768 .

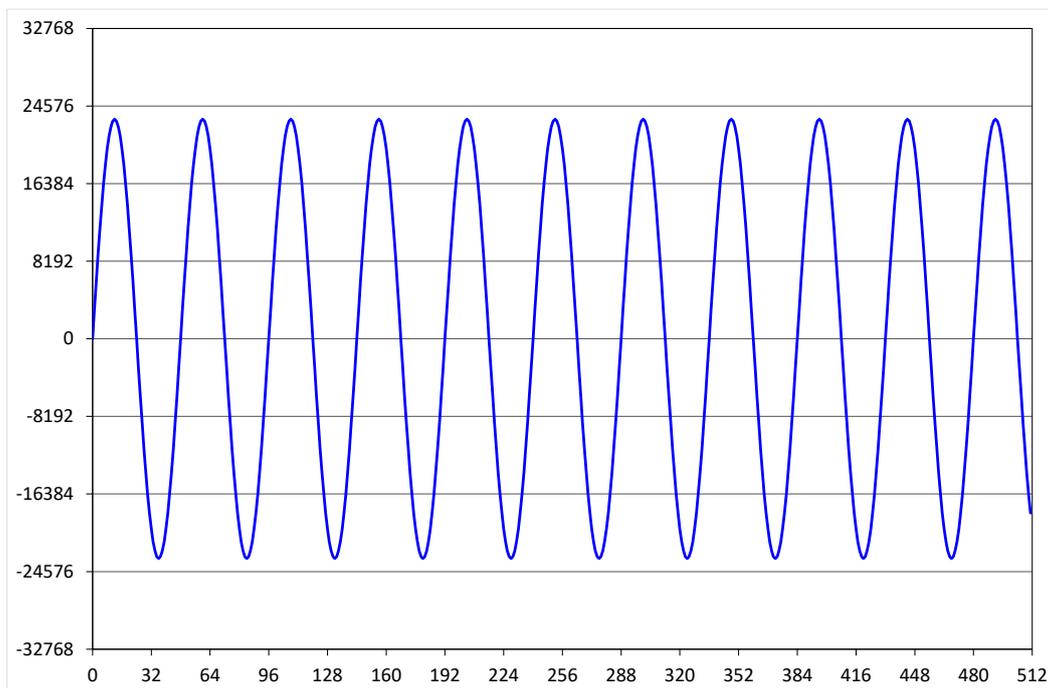


Fig. 16.3.1 Campionamento digitale di una sinusoide a 1 kHz.

L'esempio rende evidente il legame fra le ampiezze del segnale di pressione in Pa, le corrispondenti ampiezze del segnale elettrico in mV, e la loro rappresentazione numerica in numeri interi.

Le impostazioni elencate in precedenza consentono di analizzare segnali con livello massimo di picco di 100 dB, che è il limite di clipping, in un range dinamico che si estende sicuramente per almeno 80 dB, cioè a partire da un livello di pressione sonora di 20 dB. Ma ai segnali reali è associato un fattore di cresta, cioè un rapporto fra valore di picco e valore efficace, ben superiore ai 3 dB del segnale sinusoidale, e il campo dinamico effettivo risulterà più limitato. Se, ad esempio, il fattore di cresta è di 20 dB, il campo dinamico effettivo andrà da 20 a 80 dB.

Sotto questo punto di vista, l'analisi del segnale digitale risulta differente da quella analogica: passando attraverso un convertitore analogico-digitale, il campo dinamico dello strumento viene ridotto dalla presenza di segnali con fattore di cresta elevato. Pertanto, anche se strumenti digitali con convertitori a 16 bit vengono accreditati di campi dinamici di oltre 80 dB (con segnale sinusoidale, in accordo alle norme IEC), il campo dinamico effettivo con segnali sonori aventi spiccato carattere impulsivo può essere ridotto anche a meno di 50 dB. Tuttavia i convertitori più recenti, a 24 bit, minimizzano il problema, consentendo di operare in un campo dinamico utile sino a 110 dB.

Effettuare la conversione in decibel dei valori contenuti in un segnale digitalizzato risulta molto semplice se ci si limita al calcolo del livello di pressione sonora istantanea; ciascun valore numerico rappresentante un punto della forma d'onda $x[i]$ va anzitutto convertito in un valore di pressione istantanea, $p_{ist}[i]$,

$$p_{ist}[i] = \frac{x[i]}{x_{max}} \cdot \frac{V_{max}}{S} = \frac{x[i]}{32768} \cdot \frac{100}{50} \frac{(\text{mV})}{\left(\frac{\text{mV}}{\text{Pa}}\right)} \text{ Pa}, \quad (16.3.1)$$

quindi si calcola il valore del livello, in dB, con la solita relazione:

$$L_{ist}(i) = 20 \cdot \log_{10} \left(\frac{p_{ist}[i]}{p_0} \right) = 20 \cdot \log_{10} \left(\frac{p_{ist}[i]}{20 \cdot 10^{-6}} \right) \text{ dB}. \quad (16.3.2)$$

Applicando le operazioni elencate al tratto di segnale sinusoidale precedente, si ottiene la rappresentazione di Fig. 16.3.2. Si osserva che il valore massimo (picco) raggiunge 97 dB: negli altri istanti, invece, il livello è sempre inferiore, arrivando a valori molto bassi in corrispondenza dei punti in cui la forma d'onda va verso lo zero.

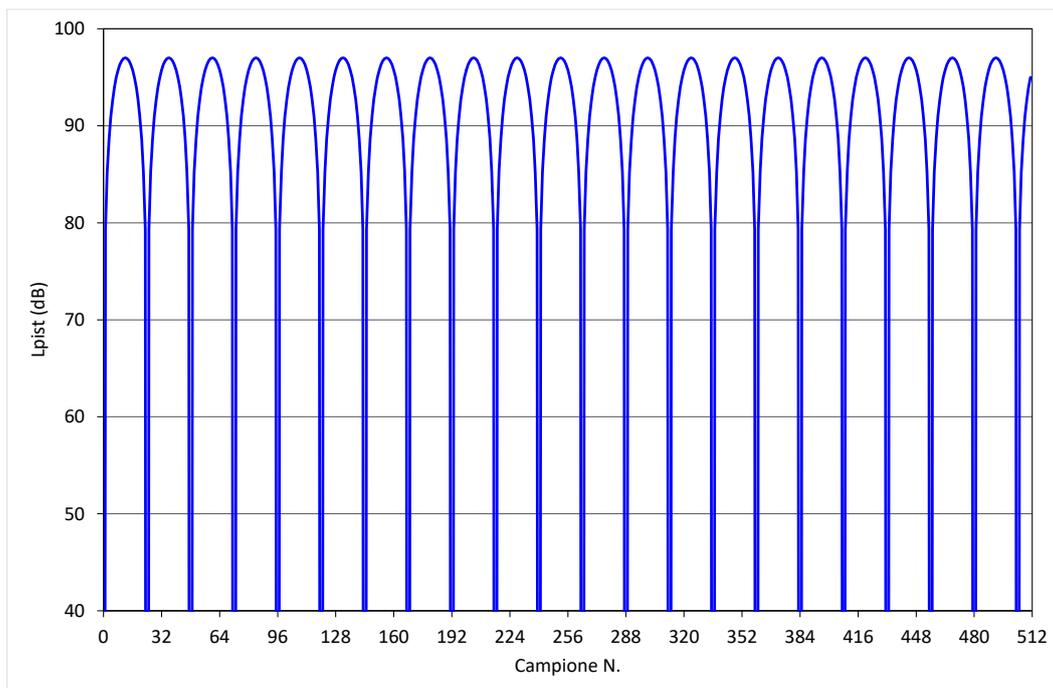


Fig. 16.3.2 Livello di pressione sonora istantanea di un segnale sinusoidale.

Questa forma di rappresentazione del livello di pressione sonora istantanea nel tempo non è molto usata: viene quasi solo impiegata per studiare le riflessioni del suono in un ambiente confinato. Emettendo un suono impulsivo nel punto occupato dalla sorgente sonora, e campionando il segnale di pressione sonora nel punto di ascolto, si ottiene la risposta all'impulso dell'ambiente, in questo caso chiamata anche *reflectogramma* o

ecogramma: essa mette in luce facilmente il decadimento del suono, cioè la coda sonora, e l'eventuale presenza di echi, come mostrato in Fig. 16.3.3.

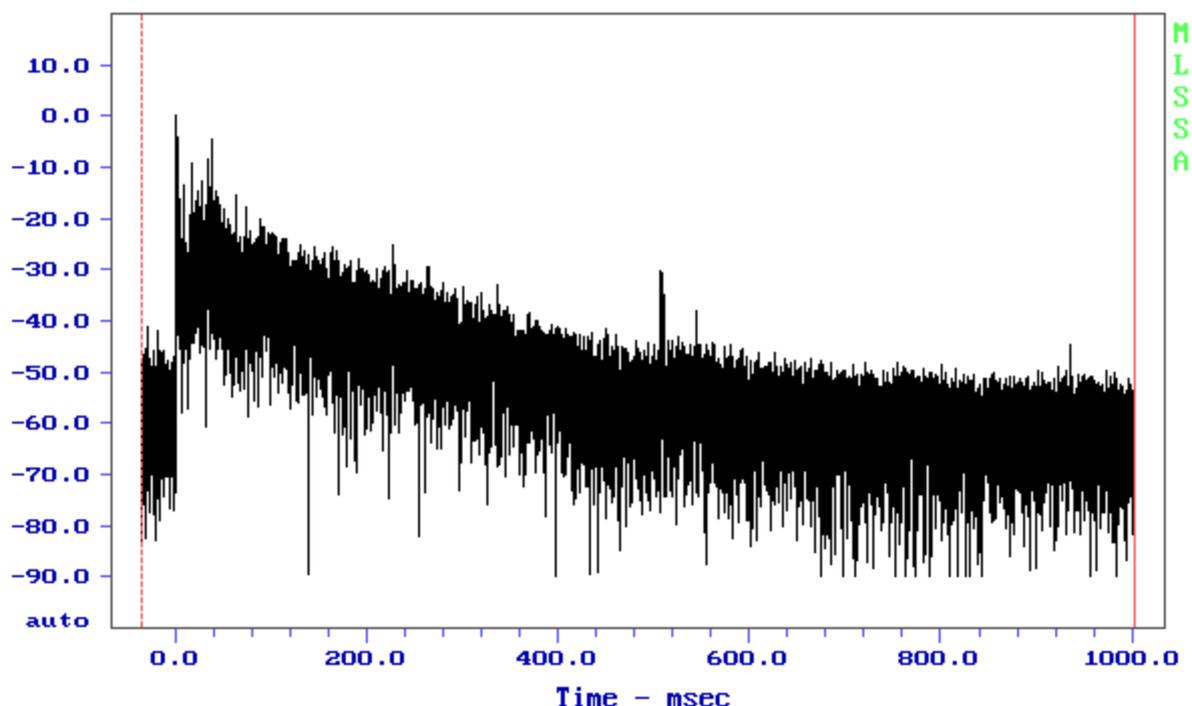


Fig. 16.3.3 Reflectogramma di una sala con eco.

Al di là dell'applicazione proposta, in generale risulta necessario calcolare non livelli istantanei, riferiti ai singoli campioni della forma d'onda digitalizzata, bensì valori medi efficaci su un certo intervallo di tempo. In particolare, operando con segnali digitalizzati, è molto più semplice calcolare il livello sonoro equivalente su un segmento di segnale, piuttosto che valutare l'evoluzione temporale dei livelli istantanei con costante di tempo *Fast*, *Slow* o *Impulse*.

La definizione di Livello sonoro equivalente, nell'intervallo di integrazione T , di un segnale continuo $p_{ist}(t)$, è, come noto,

$$L_{eq,T} = 10 \cdot \log_{10} \left(\frac{\frac{1}{T} \cdot \int_0^T [p_{ist}(t)]^2 \cdot dt}{p_0^2} \right) \text{ dB.} \quad (16.3.3)$$

Applicando la definizione al segnale digitalizzato, con N punti campionati nell'intervallo di tempo T , si ottiene:

$$L_{eq,T} = 10 \cdot \log_{10} \left(\frac{\frac{1}{N} \cdot \sum_{i=0}^N [p_{ist}(i)]^2}{p_0^2} \right) \text{ dB}, \quad (16.3.4)$$

da cui risulta che è sufficiente elevare al quadrato i singoli valori digitalizzati, e poi calcolare il valore medio della sequenza di valori elevati al quadrato.

Applicando la procedura al solito tratto temporale di segnale sinusoidale di esempio, si ricava un valore di livello sonoro equivalente di 93,98 dB. Non si ottiene il valore *esatto* di 94 dB, perché un spezzone di 512 punti non contiene un numero esatto di periodi della sinusoide. Inoltre, c'è sempre un errore, anche se molto piccolo, dovuto alla discretizzazione in ampiezza a 16 bit e al campionamento temporale con passo finito.

In presenza di un segnale digitalizzato molto lungo, si possono anche calcolare diversi valori parziali di L_{eq} in successione, ad esempio uno per ogni minuto: in tal modo si ottiene una rappresentazione complessiva della storia temporale del livello sonoro, che può venire memorizzata con una occupazione di memoria infinitamente più piccola di quella del segnale digitale originario. In Fig. 16.3.4 è rappresentata la storia temporale di 24 ore di un rumore, costituita da $24 \times 60 = 1440$ valori di $L_{eq,1min}$.

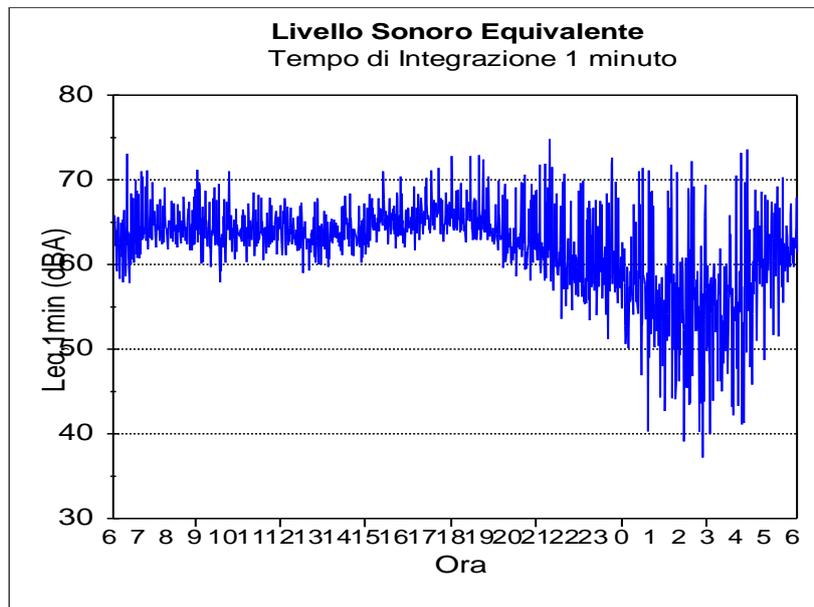


Fig. 16.3.4 Sequenza di valori di $L_{eq,1min}$ per un intero periodo di 24 h.

Viceversa, risulta molto più complicato calcolare il valore istantaneo con una data costante di tempo esponenziale, perché in tal caso si deve emulare la risposta del circuito analogico corrispondente.

Un fonometro analogico, impostato per esempio su *Slow*, mostra un valore di livello sonoro che è dato dalla media dei livelli degli ultimi secondi trascorsi; tale media, però, non viene effettuata dando uguale peso a ciò che è avvenuto molto tempo prima e a ciò che invece è appena successo: i valori di livello istantaneo più recenti hanno maggior peso, mentre i valori più lontani nel tempo perdono via a via d'importanza, fino a venire completamente dimenticati. Questo comportamento è ottenuto con un circuito analogico resistenza-condensatore (*RC*): il prodotto fra resistenza e capacità ha le dimensioni di un tempo, e infatti rappresenta la «costante di tempo» del circuito *RC*, in secondi. Valori convenzionali del prodotto *RC* sono: 1 s (*slow*); 125 ms (*fast*); 35 ms (*impulse*, in presenza di livelli crescenti nel tempo); 1,5 s (*impulse*, per i livelli decrescenti).

Si noti che, per la costante di tempo *Impulse*, il prodotto *RC* assume valori diversi a seconda del fatto che il livello sonoro istantaneo stia salendo o scendendo. Questo è un retaggio del passato, a quando gli strumenti erano a *lancetta*, e consentiva di leggere il valore massimo raggiunto in presenza di un evento sonoro impulsivo, cosa che sarebbe risultata impossibile se anche la costante di tempo *in discesa* fosse stata di 35 ms.

Matematicamente, la risposta di un fonometro con una certa costante di tempo *RC* può essere espressa come:

$$L_{ist,RC}(t) = 10 \cdot \log_{10} \left(\frac{\frac{1}{RC} \cdot \int_{-\infty}^t e^{-\frac{\tau}{RC}} \cdot [p_{ist}(t-\tau)]^2 \cdot d\tau}{p_0^2} \right) \text{ dB.} \quad (16.3.5)$$

Ovviamente sarebbe possibile utilizzare questa espressione anche per il calcolo numerico del livello sonoro a partire dal segnale digitalizzato, ma la mole di calcoli necessaria risulterebbe enorme, perché in corrispondenza di ciascun punto della forma d'onda campionata si dovrebbe operare di nuovo il calcolo della sommatoria di tutti i dati precedenti, elevati al quadrato e moltiplicati per gli opportuni coefficienti di peso dati dal termine esponenziale.

Fortunatamente esiste un metodo assai semplice per ottenere in successione l'intera storia temporale con costante di tempo esponenziale: ad ogni istante, è sufficiente calcolare il nuovo valore della pressione sonora al quadrato, come media pesata dei quadrati del valore all'istante attuale della pressione sonora e del valore all'istante precedente della pressione sonora mediata:

$$(p_{RC}[i])^2 = B \cdot (p_{ist}[i])^2 + A \cdot (p_{RC}[i-1])^2, \quad (16.3.6)$$

cui corrisponde un filtraggio IIR *minimale*, con un solo coefficiente *A* e un solo coefficiente *B*, applicato però non al segnale vero e proprio, ma ad un segnale ottenuto elevando al quadrato ciascun campione. Relativamente a tale segnale, il filtro IIR implementa un filtro passa-basso del prim'ordine.

La somma dei fattori di peso *A* e *B* deve essere uguale a 1, e per ciascun valore della costante di tempo esiste un opportuno valore del fattore di peso *B* e, di conseguenza, di $A = 1-B$, che consente di calcolare il valore corrente della media esponenziale. Se,

esemplificando, si campiona a 48 kHz, dunque con periodo di campionamento $\Delta t = 1/48000 = 0,028333$ ms, si ottengono i seguenti valori del fattore di peso B :

$$\text{Slow} - B = \frac{e^{-\frac{\Delta t}{2 \cdot RC}} \cdot \Delta t}{RC} = 0,0000208331$$

$$\text{Fast} - B = \frac{e^{-\frac{\Delta t}{2 \cdot RC}} \cdot \Delta t}{RC} = 0,0001666528$$

$$\text{Impulse (salita)} - B = \frac{e^{-\frac{\Delta t}{2 \cdot RC}} \cdot \Delta t}{RC} = 0,0005950610$$

$$\text{Impulse (discesa)} - B = \frac{e^{-\frac{\Delta t}{2 \cdot RC}} \cdot \Delta t}{RC} = 0,0000138888.$$

Adottando una frequenza di campionamento differente, naturalmente si ottengono valori dei fattori di peso diversi. Se, per esempio, si applica la ponderazione temporale *Fast* al solito segnale sinusoidale di prova, si ottiene il risultato rappresentato in Fig. 16.3.5. Poiché il tratto considerato è temporalmente molto breve (circa 10,67 ms), il livello *Fast* non fa in tempo a raggiungere il valore stabile di circa 94 dB, ma termina con un valore di circa 83,08 dB: è necessario un tempo pari a almeno 5-8 costanti di tempo, cioè quasi 1 s, affinché il valore *Fast* si assesti correttamente. Inoltre, all'inizio si intravedono ancora i gradini prodotti dalle singole semionde della sinusoide.

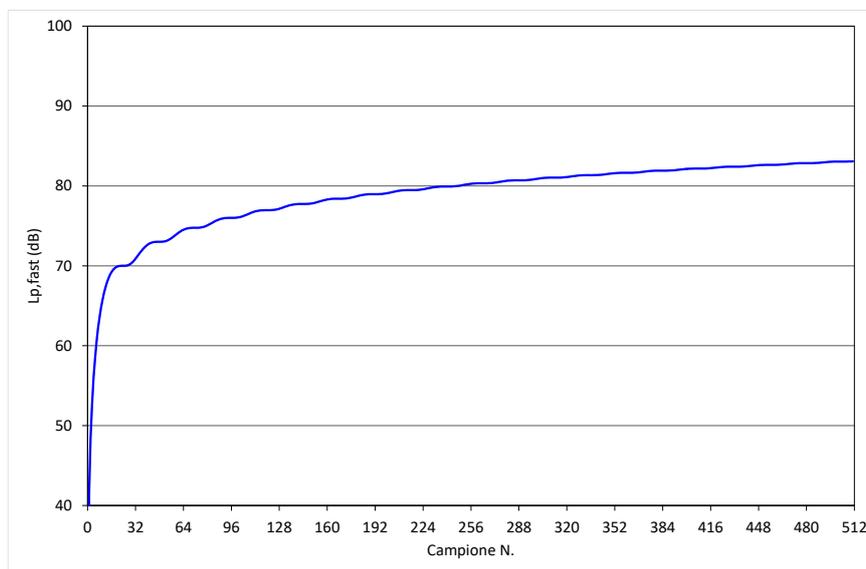


Fig. 16.3.5 Livello sonoro con costante di tempo *Fast*.

Poiché non c'è alcun motivo di rappresentare un segnale con un campionamento temporale fitto come la forma d'onda originaria, è possibile ricampionare la storia del segnale, mediato temporalmente, con un intervallo di campionamento pari, ad esempio, alla metà della costante di tempo, ottenendo ugualmente una corretta descrizione dell'evoluzione temporale del livello sonoro.

Uno dei pregi dell'analisi digitale è la possibilità di calcolare contemporaneamente la storia temporale del livello sonoro con varie costanti di tempo, sia lineari sia esponenziali, e apprezzarne le differenze. In Fig. 16.3.6 viene messa a confronto l'analisi temporale di uno stesso evento sonoro campionato digitalmente con diverse costanti di tempo, *Slow*, *Fast*, e con integrazioni lineari, *Short Leq*, di diversa estensione: 2 s, 1 s, 250 ms e 125 ms. In Tab. 16.3.1 si può osservare come i valori massimi raggiunti cambino con la scelta del tempo di integrazione, mentre i valori complessivi di L_{eq} e di *SEL* (*Single Event Level*) siano sempre gli stessi.

Tab. 16.3.1 Risultati dell'analisi con diversi tempi di integrazione.

Tempo di integrazione	$L_{A,max}$ (dB)	L_{eq} (dB)	<i>SEL</i> (dB)
Slow	97,0	118,9	101,9
Fast	99,3	118,9	101,9
Lin (2 s)	95,9	118,9	101,9
Lin (1 s)	96,9	118,9	101,9
Lin (250 ms)	99,0	118,9	101,9
Lin (125 ms)	99,6	118,9	101,9

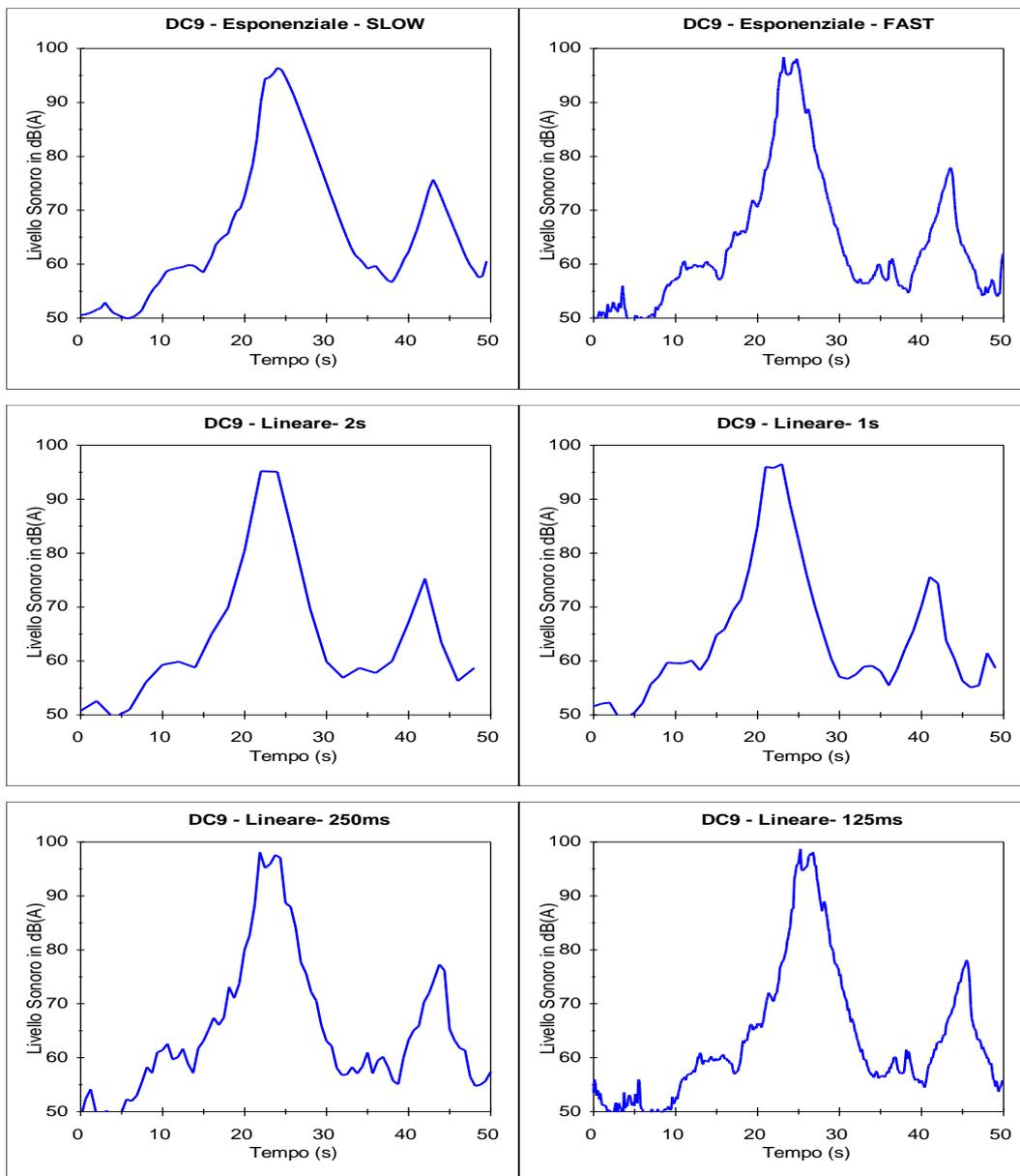


Fig. 16.3.6 Profilo del livello sonoro nel tempo con diversi tempi di integrazione esponenziale e lineare.

6.4 La trasformata veloce di Fourier (FFT)

A partire da un segnale digitalizzato, è facile effettuare un'analisi in frequenza a banda costante (tutte le bande hanno la stessa larghezza), mediante una trasformazione di Fourier di un campione temporale di segnale.

La trasformazione discreta di Fourier (DFT, *Discrete Fourier Transform*) è una semplice operazione di conversione dal dominio del tempo al dominio della frequenza, che produce una analisi spettrale a bande di larghezza costante, e in numero finito, pari alla metà più 1 del numero di punti campionati N su cui si opera la trasformazione, come schematizzato in Fig. 16.4.1. Per ciascuna banda di frequenza (eccetto la prima e l'ultima), si ottiene un valore complesso, ovvero un modulo e una fase della pressione sonora, che rappresentano modulo e fase di una onda sinusoidale.

I valori delle frequenze di centro-banda si ottengono partendo da 0 Hz, e incrementando la frequenza a passi costanti Δf , sino alla frequenza di Nyquist, con Δf data da:

$$\Delta f = \frac{f_{Nyquist}}{N/2} \quad (16.4.1)$$

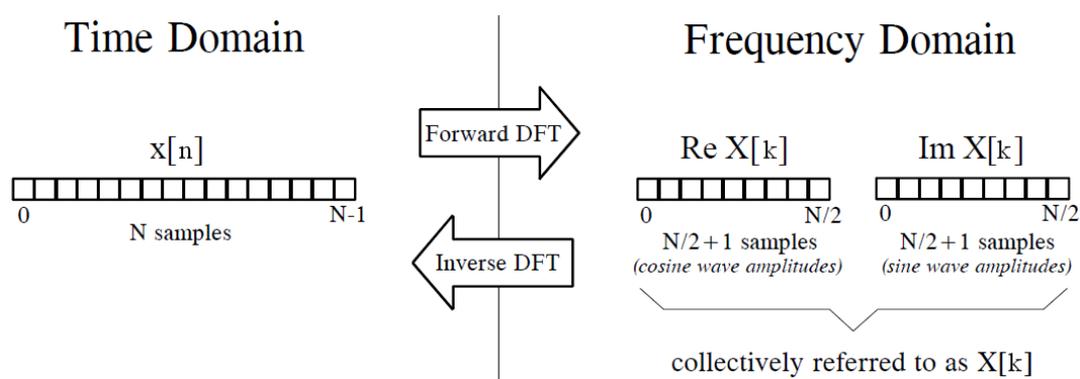


Fig. 16.4.1 – DFT: conversione fra dominio del tempo e della frequenza

Nella (16.4.1) la determinazione degli $N/2+1$ coefficienti spettrali complessi (direttamente come numero complesso, oppure separatamente come parte reale e immaginaria di $X[k]$) è possibile con le seguenti relazioni:

$$X[k] = \frac{\sqrt{2}}{N} \cdot \sum_{n=0}^{N-1} x[n] \cdot \exp(-j \cdot 2 \cdot \pi \cdot k \cdot n / N)$$

$$X_r[k] = \frac{\sqrt{2}}{N} \cdot \sum_{n=0}^{N-1} x[n] \cdot \cos(2 \cdot \pi \cdot k \cdot n / N) \quad k = 0, \dots, N/2 \quad (16.4.2)$$

$$X_i[k] = -\frac{\sqrt{2}}{N} \cdot \sum_{n=0}^{N-1} x[n] \cdot \sin(2 \cdot \pi \cdot k \cdot n / N).$$

Si noti che a rigore le (16.4.2) producono soltanto $N/2-1$ risultati non nulli, in quanto per $k = 0$ si ottiene un valore proporzionale alla componente continua (DC) della forma d'onda, che però solitamente viene assunta uguale a zero, allorché il segnale campionato corrisponde ad una pressione sonora sinusoidale (il cui valore medio nel tempo è, per definizione, nullo). Analogamente, anche il valore alla frequenza di Nyquist ($k = N/2$) è in generale sempre nullo, grazie alla presenza del filtro anti-aliasing. Dunque $X[0] = X[N/2] = 0$; pertanto le componenti spettrali significative sono in tutto $N/2-1$.

La frequenza, f , di ciascuna componente spettrale di indice k si calcola come:

$$f = k \cdot \frac{f_{Nyq}}{N/2} = k \cdot \frac{f_{samp}}{N} \quad (\text{Hz}) \quad (16.4.3)$$

in cui f_{Nyq} è la frequenza di Nyquist e f_{samp} la frequenza di campionamento.

Ovviamente è possibile ricavare, per ciascun coefficiente spettrale, il modulo e la fase a partire dalle componenti reale ed immaginaria $X_r[k]$ e $X_i[k]$:

$$X_{mod}[k] = \sqrt{X_r[k]^2 + X_i[k]^2}$$

$$X_{phase}[k] = \text{atan2}(X_i[k], X_r[k]), \quad (16.4.4)$$

dove le componenti spettrali $X_{mod}[k]$ e $X_{phase}[k]$ rappresentano i valori dell'ampiezza e della fase di $N/2-1$ onde cosinusoidali, che sommate riproducono il segnale originario $x[n]$.

Si noti inoltre che nelle (16.4.2) compare un fattore di normalizzazione uguale a $\sqrt{2}/N$, necessario affinché venga rispettato il *teorema di Parseval*, espresso dalla successiva relazione (16.4.5), che garantisce che l'energia contenuta nel segnale originario, nel dominio del tempo, sia pari all'energia dello spettro corrispondente nel dominio della frequenza; esso fa anche sì che il modulo di ciascuna componente spettrale $X_{mod}[k]$ rappresenti l'ampiezza efficace (rms) di ciascuna onda cosinusoidale:

$$\frac{1}{N} \sum_{n=0}^{N-1} (x[n])^2 = \sum_{k=1}^{N/2-1} (X_{\text{mod}}[k])^2 . \quad (16.4.5)$$

Come previsto dal teorema di Fourier, esistono formule inverse delle (16.4.2), che consentono di ricostruire la forma d'onda campionata $x[n]$ come somma di $N/2-1$ funzioni sinusoidali, a partire dai coefficienti spettrali espressi in modulo e fase, oppure in parte reale e immaginaria:

$$\begin{aligned} x[n] &= \sqrt{2} \cdot \sum_{k=1}^{N/2-1} X_{\text{mod}}[k] \cdot \cos(2 \cdot \pi \cdot k \cdot n / N + X_{\text{phase}}[k]) \\ x[n] &= \sqrt{2} \cdot \sum_{k=1}^{N/2-1} \{X_r[k] \cdot \cos(2 \cdot \pi \cdot k \cdot n / N) + X_i[k] \cdot \sin(2 \cdot \pi \cdot k \cdot n / N)\} \end{aligned} \quad n = 0, \dots, (N-1) \quad (16.4.6)$$

Si osservi che è stato assunto, come enunciato in precedenza, che le componenti di frequenza 0 e di Nyquist siano nulle. Le relazioni (16.4.6) sono note come *Trasformazione di Fourier inversa* (IDFT, *Inverse Discrete Fourier Transform*). Anche in questo caso, la presenza del fattore $\sqrt{2}$ necessario a trasformare il valore efficace dell'ampiezza di ciascuna onda cosinusoidale nel corrispondente valore di ampiezza o di picco.

Normalmente il calcolo della DFT operato secondo le (16.4.2) è piuttosto lento, e si preferisce impiegare in alternativa uno dei numerosi algoritmi che producono esattamente lo stesso risultato ma in modo molto più veloce. Complessivamente, questi algoritmi sono appunto denominati FFT (*Fast Fourier Transform*) e hanno ormai raggiunto una tale complessità da non poter essere illustrati in dettaglio in quest'ambito; per ragioni storiche si ricorda il lavoro di Cooley e Tukey (1965), che ha aperto la strada allo sviluppo degli algoritmi FFT. Inoltre, sulle moderne architetture di elaborazione digitale l'algoritmo è strettamente connesso con i dettagli strutturali del processore su cui esso viene implementato e ottimizzato, e la trattazione non potrebbe essere generale.

Con algoritmi e processori altamente ottimizzati (<http://software.intel.com/en-us/intel-ipp>; <http://www.fftw.org/>), il calcolo della FFT richiede tempi di elaborazione che possono arrivare alla centomillesima parte del tempo richiesto dal calcolo della DFT standard, rendendo questo tipo di elaborazione estremamente efficiente, soprattutto per valori elevati della lunghezza dello spezzone di segnale analizzato.

Ovviamente è stata sviluppata anche la versione *inversa* dell'algoritmo FFT (IFFT, *Inverse Fourier Transform*), che implementa la ricostruzione della forma d'onda nel tempo a partire dalle componenti spettrali, e costituisce dunque la versione *veloce* delle (16.4.6).

L'algoritmo FFT opera in modo ottimale su un numero di campioni N pari a un esponenziale di 2, per cui normalmente l'analisi FFT di un segnale acustico viene operata a partire da spezzoni di segnale di lunghezza uguale a 4096, 8192 o 16384 punti. Un moderno processore consente tuttavia l'analisi in tempo reale anche di segmenti temporali arbitrariamente lunghi, incrementando la risoluzione in frequenza, ma inevitabilmente perdendo la capacità di seguire l'evoluzione temporale dei fenomeni sonori, in quanto, ad

esempio con $N = 65536$, ogni spezzone analizzato risulta lungo ben più di 1 s, operando con una frequenza di campionamento di 48 kHz.

Attuando, per esempio, l'analisi su 16384 punti di un segnale campionato a 48 kHz, si analizza un segmento avente una lunghezza temporale di $\frac{16384}{48000} = 0,341333\text{s}$; si ottiene in

questo modo uno spettro con $\frac{16384}{2} = 8192$ righe, equamente spaziate dalla frequenza 0 alla frequenza pari a metà di quella di campionamento, cioè 24000 Hz. La risoluzione in frequenza, cioè la larghezza di ogni banda, è uguale a $\frac{24000}{8192} = 2,9296875$ Hz. In questo

caso, si può dunque correttamente parlare di analisi di frequenza in banda stretta.

Un analizzatore FFT si definisce *in tempo reale* se è in grado di calcolare e visualizzare uno spettro con un tempo inferiore o uguale alla lunghezza temporale del segnale analizzato. Nel caso dell'esempio precedente, è necessario che venga calcolato e visualizzato un nuovo spettro in un tempo inferiore a 0,341 s. Tale prestazione di calcolo è ormai da parecchi anni alla portata anche di un PC a basso costo, grazie a librerie di calcolo FFT particolarmente efficienti e ottimizzate (<http://software.intel.com/en-us/intel-ipp>, <http://www.fftw.org/>). In precedenza, era invece necessario impiegare speciali circuiti, denominati DSP (*Digital Signal Processor*), che operavano la trasformazione FFT via hardware in tempi sufficientemente brevi.

Per rappresentare graficamente la composizione in frequenza di un segnale acustico, si costruisce un grafico (con asse delle frequenze lineare, o, più sovente, logaritmico) del valore di $X_{\text{mod}}[k]$ espresso in decibel; grafico che viene comunemente chiamato «spettro FFT» del segnale. Solo in casi molto particolari, risulta utile visualizzare anche il grafico della fase del segnale. In Fig. 16.4.3 è riportato un tipico spettro FFT del livello di pressione sonora in decibel.

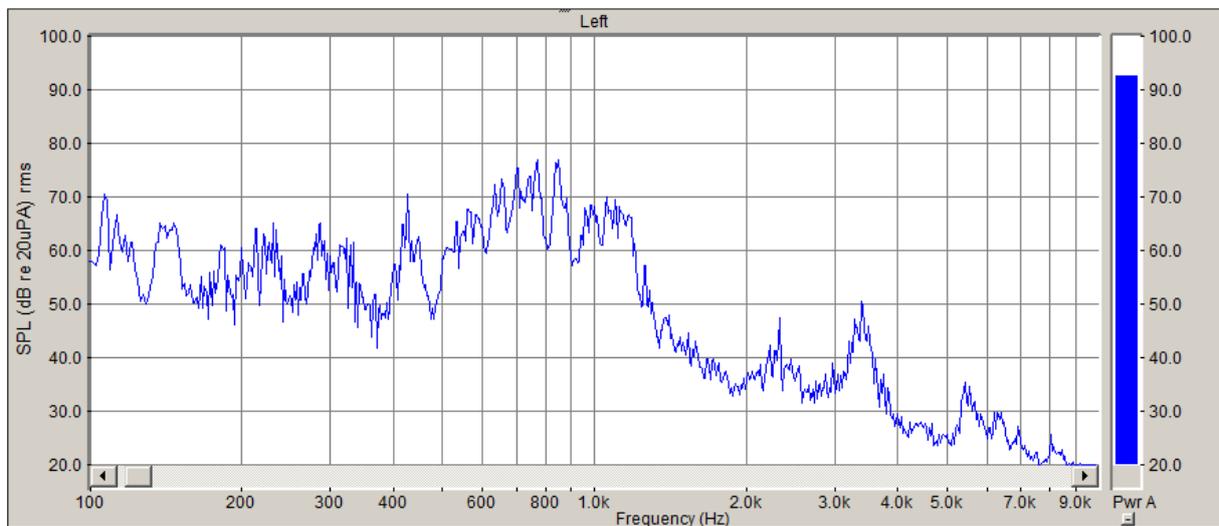


Fig. 16.4.2 Esempio di spettro FFT di un segnale acustico.

Con un analizzatore in tempo reale si può anche evidenziare l'evoluzione nel tempo dello spettro del segnale, rappresentando graficamente un *sonogramma* o un *Waterfall*, come mostrato nelle Figg. 16.4.3 e 16.4.4.

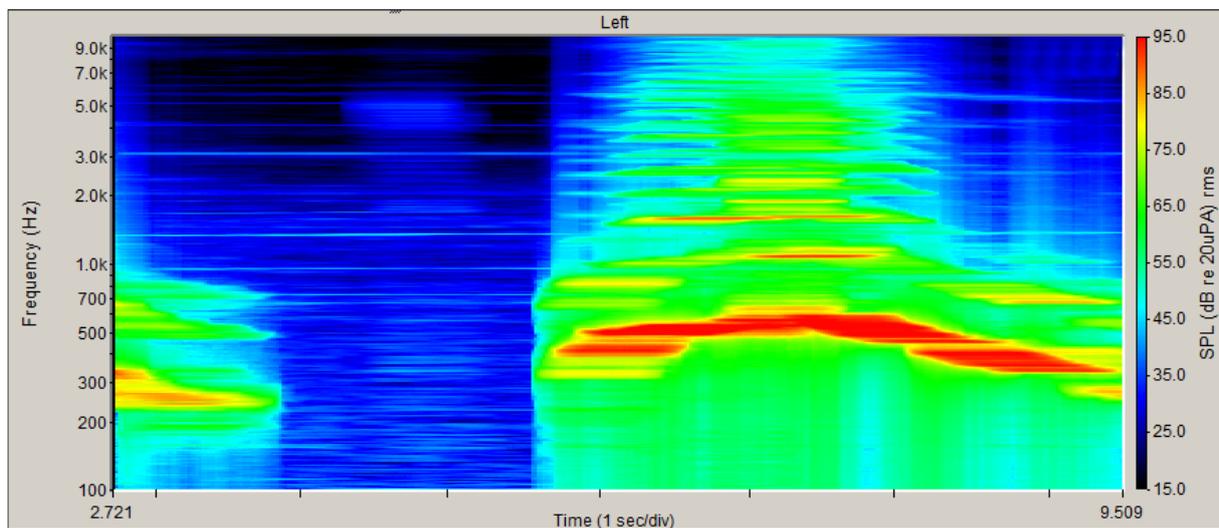


Fig. 16.4.3 Sonogramma.

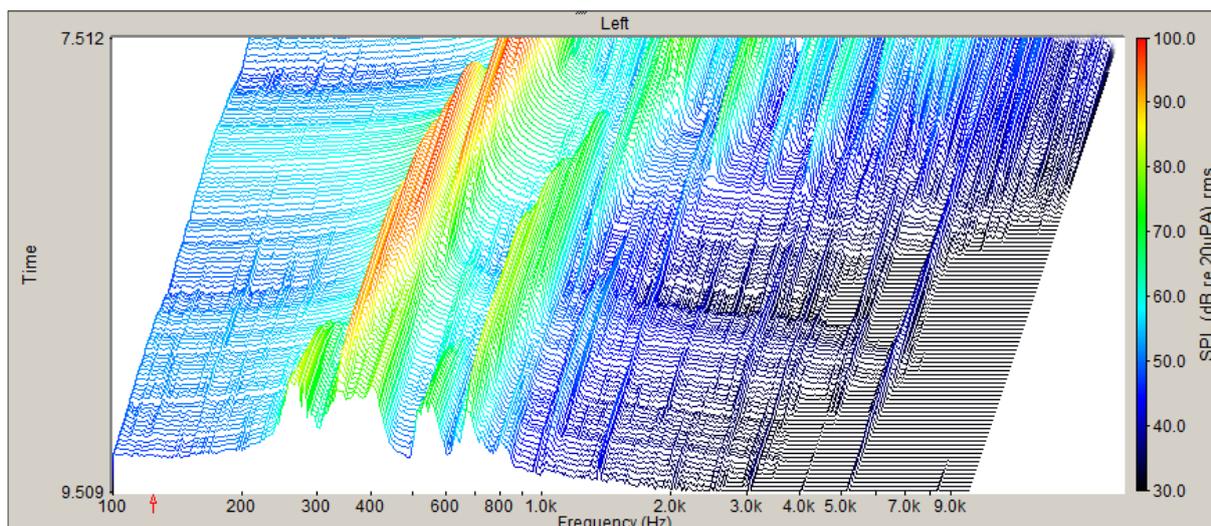


Fig. 16.4.4 Waterfall.

Infine, resta aperto il problema della *finestra temporale* da applicare al segmento temporale di segnale esaminato: è infatti noto che se non si applica alcuna finestra, cioè se si usa la cosiddetta *finestra rettangolare*, il fatto che all'inizio e alla fine del blocco di segnale analizzato, la forma d'onda non assuma in generale lo stesso valore, causa un errore di calcolo, che diviene molto evidente analizzando una sinusoida, come nell'esemplificazione di Fig. 16.4.5. Si osserva che mentre lo spettro teorico prevederebbe un'unica riga spettrale esattamente a 1 kHz, e nessuna componente ad altre frequenze, viceversa, lo spettro ottenuto realmente con la finestra rettangolare è disperso su tutte le frequenze, causa appunto del suddetto *effetto di bordo*. Si parla quindi di *leakage*, o dispersione.

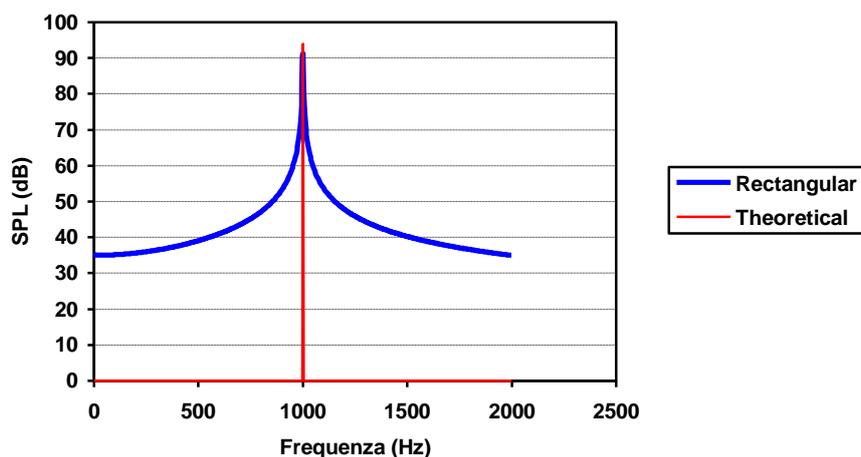


Fig. 16.4.5 Esempio di leakage.

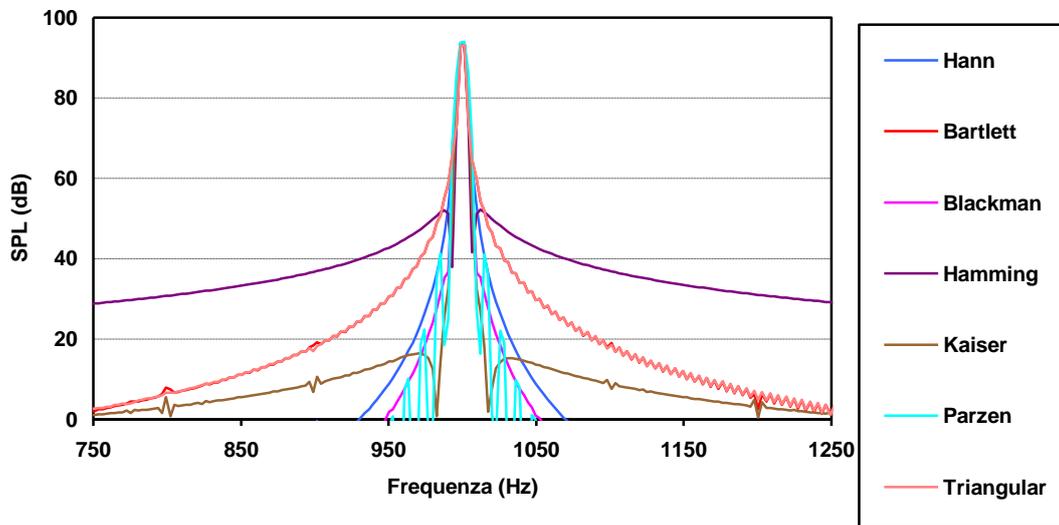


Fig. 16.4.6 Effetto di diverse finestre temporali.

Applicando invece finestre temporali di forma diversa, dallo stesso spezzone di segnale sinusoidale si ottengono le analisi mostrate in Fig. 16.4.6. Inoltre dalla figura si constata che, fra le finestre considerate, la *Hann* e la *Blackman* sono quelle che riducono maggiormente il fenomeno della dispersione.

Un difetto delle *finestre*, tuttavia, è che portano a pesare meno i dati in prossimità degli estremi del segmento temporale analizzato: ad esempio, la finestra *Hann* ha la forma mostrata in Fig. 16.4.7.

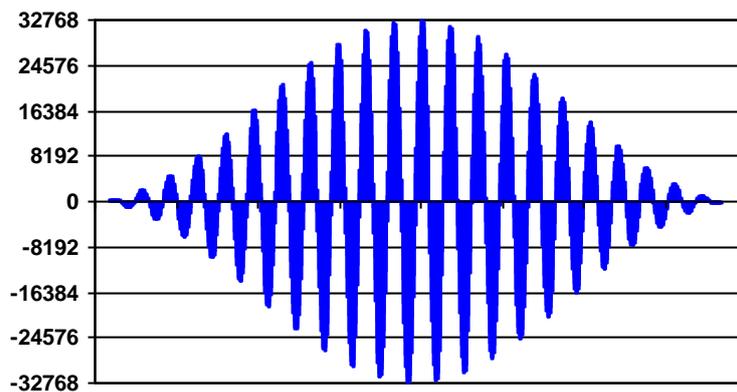


Fig. 16.4.7 Finestra Hann applicata al segnale sinusoidale.

Pertanto, le porzioni all'inizio e alla fine del segmento analizzato hanno ricevuto un *peso* molto ridotto. Per ovviare a questo inconveniente, disponendo di un analizzatore FFT

sufficientemente veloce, si opera un'analisi con finestre sovrapposte (*overlap*), in modo che dopo aver analizzato i primi 16384 punti, non si passa direttamente all'analisi dei successivi 16384, ma si sposta la finestra di analisi verso destra di 8192 punti (overlap del 50%). In tal modo, l'ultimo punto campionato dalla prima finestra, che ha ricevuto peso 0, si trova nella seconda finestra esattamente al centro, dove invece ha peso 1. Per ciascun punto, ovunque si trovi, la somma dei pesi di due analisi consecutive è sempre pari ad 1. In Fig. 16.4.8 è illustrata questa tecnica con finestre Hanning sovrapposte al 50%

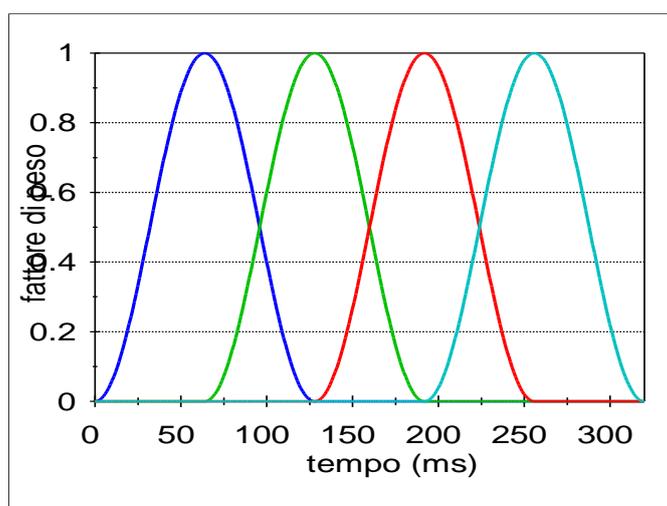


Fig. 16.4.8 Finestre Hann sovrapposte al 50%.

Soltanto con la tecnica di overlap un analizzatore FFT può per esempio calcolare correttamente il Livello sonoro equivalente: tuttavia è richiesta una velocità dell'analizzatore almeno doppia di quella necessaria a ottenere il real-time, perché lo stesso possa venire impiegato in sostituzione di un fonometro per misurazioni con effettivo valore legale. Volendo essere ancora più dettagliati, per mantenere un *ripple* inferiore ai limiti previsti dalle norme IEC sui misuratori di livello sonoro, occorre una sovrapposizione ancora maggiore fra due finestre successive, che debbono sovrapporsi almeno per il 75% della loro lunghezza.

Merita infine richiamare un utilizzo particolare della trasformazione FFT: tramite di essa sono stati realizzati algoritmi di convoluzione (filtraggio FIR) particolarmente veloci, grazie ai quali è possibile utilizzare filtri FIR estremamente lunghi (anche 1 milione di campioni), senza latenza di processamento e con poco carico computazionale sul processore. La base di questi metodi deriva dal fatto che la convoluzione fra due segnali nel dominio nel tempo (che in generale richiede l'effettuazione di N moltiplicazioni ed N somme, se N è la lunghezza del filtro, per il calcolo di ogni singolo campione filtrato), si riduce al semplice prodotto fra gli spettri FFT dei due segnali.

In caso di filtri significativamente lunghi, diventa pertanto molto più efficiente operare il passaggio dal dominio del tempo al dominio della frequenza impiegando l'algoritmo FFT, quindi effettuare la moltiplicazione nel dominio spettrale, e infine tornare nel dominio del tempo grazie ad una trasformazione FFT inversa, ottenendo infine il segnale filtrato. Il caso

più comune è quello di un segnale da filtrare di lunghezza cospicua, in teoria di durata pressoché infinita, che viene ad essere filtrato attraverso un filtro FIR di lunghezza ancora rilevante, ma comunque molto inferiore a quella del segnale da filtrare. In questo caso, non conviene operare un'unica trasformata FFT dell'intera lunghezza del segnale da filtrare, anche perché occorrerebbe attendere di averlo registrato per intero prima di poter iniziare l'elaborazione. Si procede allora ad una elaborazione *a blocchi*, come illustrato in Fig. 16.4.9, secondo un algoritmo noto come *Select and Save* (Oppenheim, Schaffer, 1975). Algoritmo anche denominato FFTFILT, poiché così viene chiamato nel noto programma di elaborazione Matlab™.

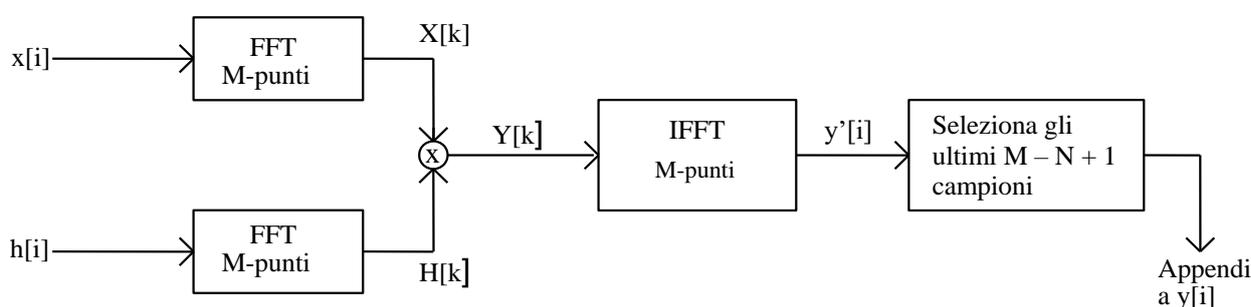


Fig. 16.4.9 L'algoritmo *Select and Save*

In pratica, si sceglie solitamente una lunghezza del blocco di elaborazione M , uguale a circa il doppio della lunghezza N del filtro, avendo cura di scegliere un valore di M uguale a una potenza di 2. Dalla lunga sequenza di ingresso viene dunque prelevato ed elaborato un blocco di M campioni, a cui si applica la FFT, producendo uno spettro di $M/2$ valori spettrali complessi $X[k]$. Preventivamente, si era operata la FFT della risposta all'impulso $h[i]$ del filtro, la cui lunghezza viene estesa ad M campioni, aggiungendo dei campioni di valore zero alla fine.

Per ciascun blocco che viene elaborato, si passa dunque alla moltiplicazione, frequenza per frequenza, dei valori spettrali derivanti dall'analisi FFT del blocco di dati corrente con i valori spettrali precalcolati del filtro $H[k]$. Allo spettro risultante $Y[k]$ si applica poi la trasformata inversa (IFFT), che restituisce un segmento temporale di M campioni. Di questi si prelevano solo gli ultimi $M - N + 1$ campioni, scartando i rimanenti. Si procede quindi all'elaborazione del successivo blocco di M punti in ingresso, avendo spostato la finestra di analisi in avanti nel tempo, di $M - N + 1$ campioni (essa risulta pertanto sovrapposta circa per metà con la precedente finestra di analisi). Di fatto, dunque, il sistema opera fornendo in uscita solo metà dei campioni elaborati, i restanti vanno scartati in quanto contaminati dalla *coda circolare* prodotta dalla FFT. In tal modo, si evita anche di dover utilizzare una finestra (Hann, Blackman ecc.)

Grazie all'algoritmo *Select and Save*, il filtraggio FIR ha conosciuto un grande successo, e ha rimpiazzato in moltissime applicazioni il meno accurato filtraggio IIR, giacché

il principale vantaggio di quest'ultimo, consistente nella riduzione del necessario numero di coefficienti e dunque del numero di moltiplicazioni, viene più che compensato dalla notevole efficienza computazionale della FFT.

Tuttavia, anche l'algoritmo *Select and Save* presenta un problema significativo, quello della latenza: il segnale processato, infatti, risulta ritardato rispetto al segnale in ingresso di una latenza pari tipicamente a M campioni. In caso di filtri FIR di lunghezza significativa, per esempio la risposta all'impulso di un grande spazio come una chiesa, ciò si traduce in un ritardo di parecchi secondi, che è ovviamente del tutto inaccettabile per applicazioni dal vivo, o in cui si debba mantenere il sincronismo tra il segnale audio e l'immagine video.

Per ridurre la latenza, e conseguire una ulteriore ottimizzazione della velocità di elaborazione, è stato sviluppato un algoritmo di convoluzione veloce basata su FFT, ancora più complesso, in cui si opera la partizioni in blocchi non solo del segnale da filtrare, ma anche della stessa risposta all'impulso del filtro. In pratica, la sequenza dei coefficienti di filtraggio viene partizionata secondo diverse *ricette*, la più semplice e prestazionale delle quali è la partizione in blocchi tutti della medesima lunghezza, che viene solitamente scelta uguale a 4096 o 8192 campioni, fornendo un buon compromesso tra riduzione della latenza (meno di 100 ms) ed efficienza dell'algoritmo di FFT (che non rende su blocchi troppo piccoli).

Per applicazioni in cui anche una latenza di 100 ms è eccessiva, si adottano tecniche di partizione della risposta all'impulso del filtro più sofisticate, come la tecnica di partizione in blocchi di lunghezza via a via crescente (Gardner, 1995), oppure la convoluzione ibrida, in cui il primo blocco è processato in modo tradizionale nel dominio del tempo, e i blocchi successivi elaborati mediante FFT (Barry, 1988). In questo caso, la latenza viene ridotta ad un solo campione, esattamente come con il filtraggio FIR tradizionale, pur mantenendo la possibilità di utilizzare filtri FIR arbitrariamente lunghi, con potenza di calcolo comunque compatibile con un processore moderno.

Gli algoritmi di convoluzione partizionata veloce sono tuttavia alquanto complessi da implementare: la Fig. 16.4.10 illustra il diagramma a blocchi del più semplice di questi algoritmi (partizionamento a blocchi di ugual lunghezza), ma già si può osservare che tradurre il complesso diagramma a blocchi in un codice di elaborazione ottimizzato ed efficiente richiede capacità di programmazione non comuni. Fortunatamente un codice altamente ottimizzato che opera la convoluzione partizionata, chiamato *BRUTEFIR*, è disponibile in open-source (<http://www.ludd.luth.se/~torger/brutefir.html>).

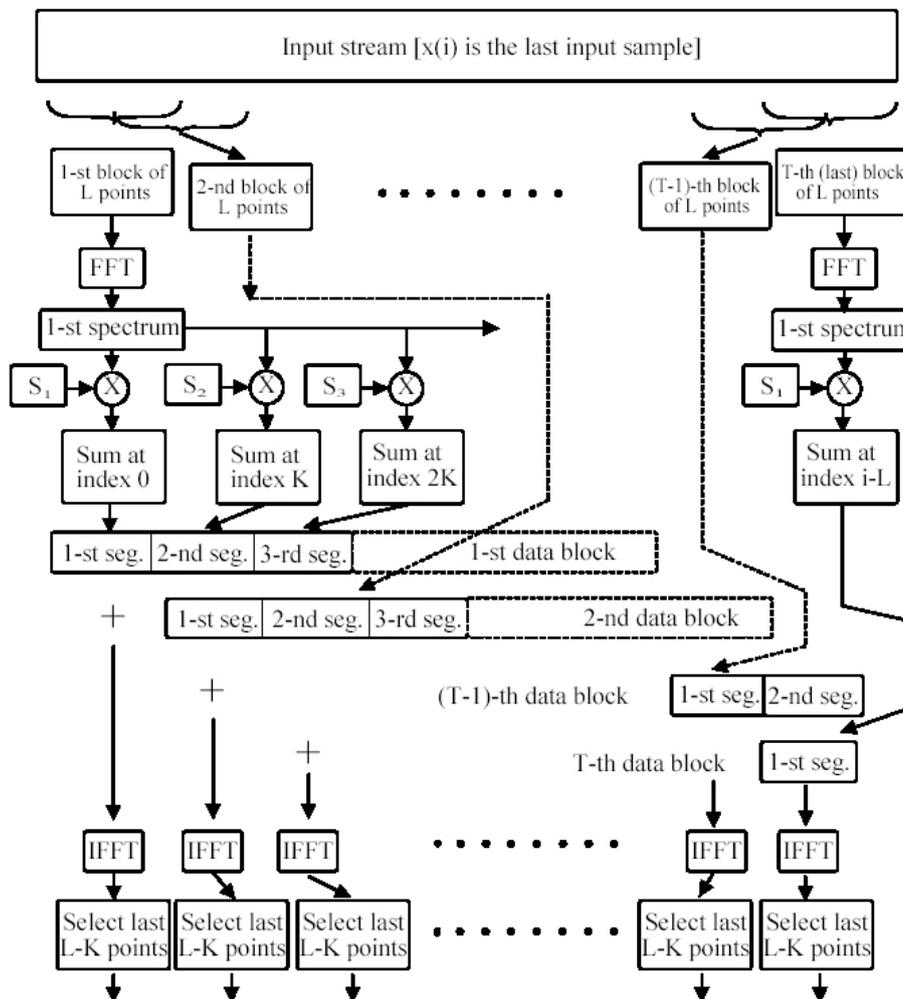


Fig. 16.4.10 L'algoritmo *Uniformly Partitioned Convolution*.

16.5 Tecniche di analisi spettrale mediante filtri digitali IIR

In questo paragrafo vengono esaminate le tecniche di elaborazione numerica del segnale impiegate per operare i due tipici filtraggi in frequenza tipicamente richiesti nelle applicazioni acustiche:

- filtraggio in bande d'ottava e terzo d'ottava;
- ponderazione in banda larga con curva A o C.

Come illustrato nel paragrafo precedente, è possibile attualmente operare tali filtraggi utilizzando filtri FIR di adeguata lunghezza, e mantenendo una modesta richiesta di potenza di calcolo e una bassa latenza di processamento grazie all'algoritmo di convoluzione

partizionata. Ciò consente di ottenere un andamento spettrale del guadagno del filtro perfettamente aderente all'andamento teorico descritto nelle norme per i filtri d'ottava e frazionari (IEC 61260) e per le ponderazioni in banda larga, A e C (IEC 60651). Allo stesso tempo, il filtraggio FIR consente di non avere alcuna distorsione di fase, e di evitare così problemi nel calcolo dei tempi di riverberazione o di altri parametri acustici, come descritto più avanti nel par. 16.7.

Nonostante gli indubbi vantaggi, il filtraggio FIR è tuttora poco utilizzato per applicazioni acustiche: ciò deriva dal fatto che le norme IEC menzionate guardano ancora a reti di filtraggio realizzate con circuiti analogici, e dall'esistenza di opportune implementazioni di filtraggio IIR che emulano da vicino il comportamento di tali circuiti analogici, presentando di fatto le stesse distorsioni di fase e conseguentemente la stessa risposta all'impulso del filtro. In particolare, la norma IEC 60651 è stata scritta ipotizzando che i filtri siano ottenuti dalla successione di un certo numero di stadi di filtraggio elementare, costituiti da filtri del primo ordine, di tipo passa-alto o passa-basso, corrispondenti agli schemi circuitali riportati in Fig. 16.5.1.

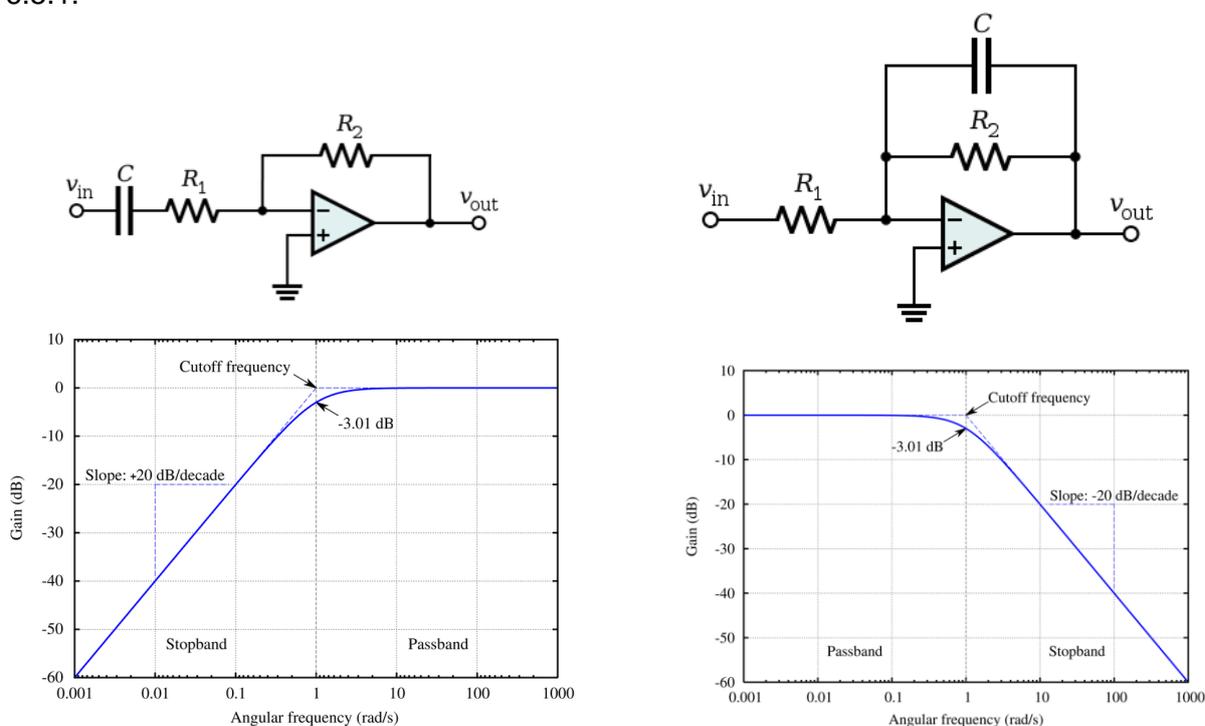


Fig. 16.5.1 Circuito analogico di filtraggio attivo: (a) passa alto – (b) passa basso

Tali circuiti danno luogo ad una frequenza di taglio del filtro passa-alto o passa-basso:

$$f_c = \frac{1}{2 \cdot \pi \cdot R_1 \cdot C} \quad [passa-alto] \tag{16.5.1}$$

$$f_c = \frac{1}{2 \cdot \pi \cdot R_2 \cdot C} \quad [passa-basso]$$

Utilizzando il tradizionale filtraggio IIR, i due schemi di filtraggio si traducono in formule di calcolo ricorsive estremamente semplici,

$$\begin{aligned}
 y_{LP}[n] &= b \cdot x[n] + (1-b) \cdot y_{LP}[n-1] && [passa-basso] \\
 y_{HP}[n] &= x[n] - y_{LP}[n] && [passa-alto],
 \end{aligned}
 \tag{16.5.2}$$

dove l'unico coefficiente di filtraggio, b , si calcola in funzione della frequenza di taglio f_c :

$$b = 1 - \exp\left(\frac{2\pi f_c}{f_{samp}}\right)
 \tag{16.5.3}$$

In pratica, una volta determinato il coefficiente di filtraggio b con la (16.5.3), la relazione (16.5.2) consente di calcolare anzitutto il segnale filtrato passa-basso $y_{LP}[n]$ con la classica formula ricorsiva simile alla (16.2.3), con un solo elemento x e un solo elemento y , quindi si ricava il segnale filtrato passa-alto semplicemente sottraendo dal segnale originale $x[n]$ il segnale filtrato passa-basso $y_{LP}[n]$.

Disponendo di tali elementi di base, risulta dunque facile costruire filtri più complessi, semplicemente come successione di filtri del primo ordine di tipo passa-alto o passa-basso. Ad esempio, per ottenere un filtro che effettua la ponderazione del segnale secondo la consueta curva A, occorre implementare in cascata 6 filtri del primo ordine (Tab. 16.5.1).

Tab. 16.5.1 Stadi di filtraggio per la curva di ponderazione A.

N.	Tipo di filtro	Frequenza di taglio
filtro 1	passa-alto	20,6 Hz
filtro 2	passa-alto	20,6 Hz
filtro 3	passa-alto	107,7 Hz
filtro 4	passa-alto	737,9 Hz
filtro 5	passa-basso	12200 Hz
filtro 6	passa-basso	12200 Hz

Per effettuare invece la ponderazione secondo la curva C, sono sufficienti 4 stadi di filtraggio (Tab. 16.5.2).

Tab. 16.5.2 Stadi di filtraggio per la curva di ponderazione C.

N.	Tipo di filtro	Frequenza di taglio
filtro 1	passa-alto	20.6 Hz
filtro 2	passa-alto	20.6 Hz
filtro 3	passa-basso	12200 Hz
filtro 4	passa-basso	12200 Hz

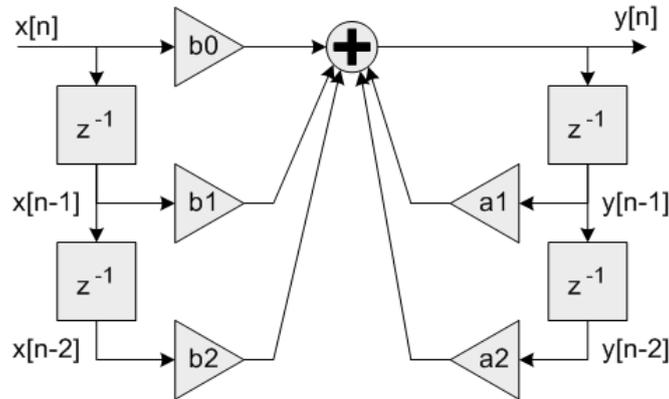


Fig. 16.5.2 Filtraggio IIR con schema *biquad* (Direct Form 1).

Ovviamente si potrebbe operare allo stesso modo anche per realizzare i filtri a banda d'ottava e di 1/3 d'ottava. Siccome però in tal caso occorrerebbe un numero decisamente più elevato di filtri del primo ordine, si preferisce, per ragioni di efficienza computazionale, utilizzare come elemento di base un operatore di filtraggio di secondo ordine molto efficiente, denominato *biquad*, il cui schema computazionale è mostrato in Fig. 16.5.2 e la cui formula di calcolo ricorsiva è la seguente:

$$y[n] = b_0 \cdot x[n] + b_1 \cdot x[n-1] + b_2 \cdot x[n-2] - a_1 \cdot y[n-1] - a_2 \cdot y[n-2] . \quad (16.5.4)$$

Un filtro biquad è in grado di operare filtraggi passa-alto (HP) e passa-basso (LP) del secondo ordine, ma anche un filtraggio passa-banda (BP) del primo ordine, equivalendo di fatto alla cascata di una coppia di filtri passa-alto e passa-basso del primo ordine, con frequenze di taglio leggermente diverse.

Il calcolo dei cinque coefficienti di filtraggio di un biquad $[b_0, b_1, b_2, a_1, a_2]$ è possibile conoscendo il tipo di filtro (HP, LP, BP), la frequenza di taglio f_{cut} e il *fattore di merito* Q (che vale ad esempio 0,707, se si vuole implementare un filtro di tipo *Butterworth*, con risposta in frequenza priva di ondulazioni), facendo impiego delle relazioni seguenti (e in Tab. 16.5.3):

$$K = \tan\left(\frac{\pi \cdot f_{cut}}{f_{samp}}\right); \quad D = 1 + \frac{K}{Q} + K^2 . \quad (16.5.5)$$

Tab. 16.5.3 Calcolo dei coefficienti filtraggio *biquad*

Tipo	b_0	b_1	b_2	a_1	a_2
LP	$\frac{K^2}{D}$	$2 \cdot \frac{K^2}{D}$	$\frac{K^2}{D}$	$2 \cdot \frac{K^2 - 1}{D}$	$\frac{1 - K/Q + K^2}{D}$
HP	$\frac{1}{D}$	$-\frac{2}{D}$	$\frac{1}{D}$	$2 \cdot \frac{K^2 - 1}{D}$	$\frac{1 - K/Q + K^2}{D}$
BP	$\frac{K/Q}{D}$	0	$-\frac{K/Q}{D}$	$2 \cdot \frac{K^2 - 1}{D}$	$\frac{1 - K/Q + K^2}{D}$

Come nel caso del filtraggio A e C, anche per il filtraggio per bande di ottava o di 1/3 di ottava occorre in generale applicare in cascata un adeguato numero di filtri biquad, ciascuno dotato di appropriati valori di tipo, frequenza di taglio e fattore Q. In tal caso, però, la norma IEC non fornisce direttamente la *ricetta* per realizzare un filtro normalizzato: vengono invece specificate, per punti, dettagliate maschere di accettabilità per la risposta in frequenza dei filtri, come mostrato in Fig. 16.5.3.

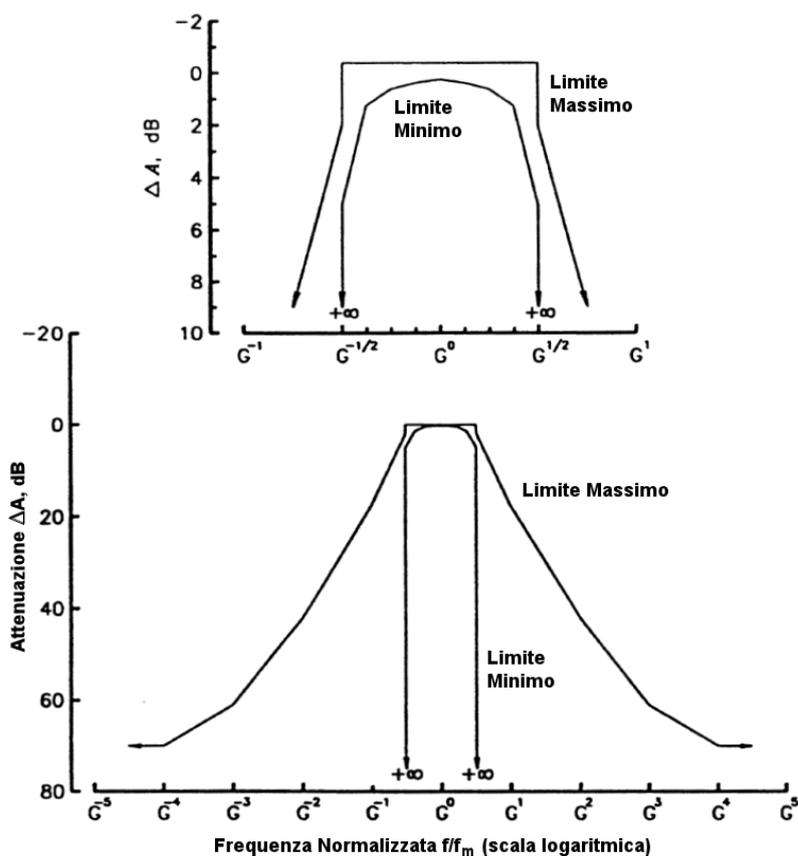


Fig. 16.5.3 Maschere di accettabilità dei filtri d'ottava in classe I secondo la norma IEC 61260.

In pratica dunque occorre aggiustare manualmente, con grande attenzione, le proprietà di ciascun filtro, in modo da farne rientrare la risposta in frequenza all'interno delle maschere per la classe di precisione scelta. Potendo ad esempio limitarsi alla classe II (la meno accurata), la ricetta per la sintesi del filtro di banda d'ottava di 1 kHz, operando ad una frequenza di campionamento di 48 kHz, è riassunta in Tab 16.5.4. Dovendo invece rispettare la classe 0 (la più accurata), per la sintesi di un filtro di 1/3 d'ottava occorrono tipicamente almeno otto biquad di diverso tipo.

Tab. 16.5.4 Stadi di filtraggio per un filtro di banda d'ottava a 1 kHz.

N.	Tipo di filtro	Frequenza di taglio	Q
biquad 1	passa-banda	755 Hz	6
biquad 2	passa-banda	1000 Hz	3
biquad 3	passa-banda	1330 Hz	6

La realizzazione di filtri dalla risposta in frequenza molto ripida e larghezza di banda limitata (si ricordi che per i filtri d'ottava e di 1/3 d'ottava la larghezza di banda si riduce al diminuire della frequenza di centro banda), pur essendo possibile impiegare un numero adeguato di sezioni di filtraggio IIR di tipo biquad, e un'accorta scelta dei parametri di filtraggio di ciascuno stadio, presenta però una notevole contropartita negativa: il filtraggio determina un notevole ritardo nella risposta calcolata, che tende a estendersi per un tempo significativo rispetto al segnale in ingresso. Di fatto, il filtro così configurato è caratterizzato da una risposta all'impulso di lunghezza non trascurabile, che può portare ad una serie di conseguenze, ad esempio nel calcolo del tempo di riverberazione o di altri parametri che caratterizzano la qualità acustica di una sala (si veda, a questo proposito, il par. 16.7).

La Fig. 16.5.4 mostra un esempio di risposta all'impulso di un filtro di 1/3 d'ottava con frequenza centrale 100 Hz, realizzato utilizzando 8 sezioni di filtraggio biquad. Si può osservare che la risposta all'impulso inizia con un cospicuo ritardo (il picco massimo ha un ritardo di circa 45 ms), ma soprattutto che il segnale decade poi molto lentamente, con una serie di oscillazioni periodiche (*ringing*), determinando un valore del *tempo di riverberazione* intrinseco del filtro pari a ben 0,43 s.

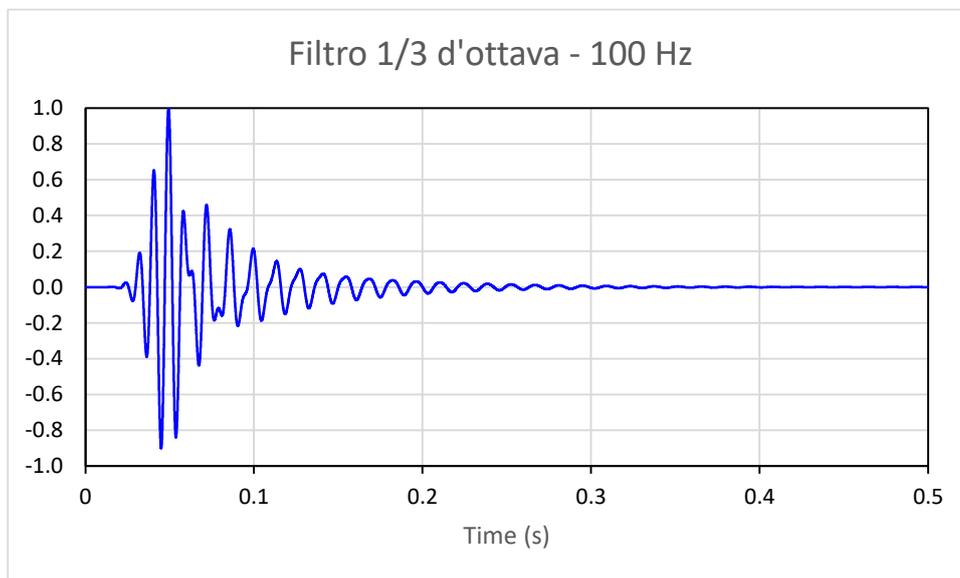


Fig. 16.5.4 Risposta all'impulso di un filtro IIR per l'analisi in banda di 1/3 d'ottava in classe 0.

Risulta dunque ovvio che un filtraggio IIR così pesante comporta una sorta di eccessivo prolungamento nel tempo del segnale filtrato; se si sceglie di utilizzare filtri IIR tradizionali per l'analisi in frequenza per bande di 1/3 d'ottava, occorre dunque accontentarsi di una risposta in frequenza meno selettiva, corrispondente a filtri di classe I, o addirittura, di classe II. Questi ultimi essendo appunto raccomandati per la misura di grandezze acustiche che dipendono dalla variazione del livello sonoro nel tempo, come lo stesso tempo di riverberazione. Esistono tuttavia alcuni accorgimenti per limitare gli effetti di questo fenomeno nella misura del tempo di riverberazione e di altri parametri acustici, che vengono illustrati nel par. 16.7.

16.6 Misura della risposta all'impulso di un sistema elettroacustico.

Le tecniche MLS e ESS.

La tecnica di misura dei coefficienti h della risposta all'impulso è in linea di principio molto semplice: è sufficiente immettere nel sistema, come segnale in ingresso $x[i]$, la funzione impulsiva ideale, la δ di Dirac, ovvero un segnale costituito da un campione avente il valore massimo consentito dalla rappresentazione digitale impiegata (ad esempio +1, nei moderni sistemi a virgola mobile) seguito da una lunga sequenza di zeri (*silenzio digitale*) (Tab. 16.6.1).

1	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---

Fig. 16.6.1 - Codifica binaria di una δ di Dirac

Convolvere il δ di Dirac con la risposta $h[l]$ del sistema, risulta

$$\begin{aligned} y_1 &= 1 \cdot h_1 + 0 \cdot h_2 + 0 \cdot h_3 + \dots = h_1 \\ y_2 &= 0 \cdot h_1 + 1 \cdot h_2 + 0 \cdot h_3 + \dots = h_2 \\ y_3 &= 0 \cdot h_1 + 0 \cdot h_2 + 1 \cdot h_3 + \dots = h_3 \end{aligned} \quad , \quad (16.6.1)$$

e ad ogni passo di campionamento si ottiene quindi un coefficiente $h[l]$.

Questo tipo di misurazione, benché teoricamente semplicissima, è in pratica molto complicato da realizzare: infatti, alimentando, per esempio, un altoparlante con il segnale \square , esso dovrebbe essere in grado di produrre un impulso molto breve e di elevatissima potenza (per una buona misurazione occorre superare il rumore di fondo di almeno 60 dB). Solo con un segnale di una certa durata diventa possibile erogare un adeguato quantitativo di energia sonora mantenendo la potenza a valori accettabili.

Di conseguenza, è entrato nell'uso corrente l'impiego di segnali di test di elevata durata temporale, che vengono ricondotti all'impulso ideale con procedimenti matematici. In particolare, i due metodi che hanno avuto maggiore successo sono stati, in ordine cronologico, il metodo MLS (*Maximum Length Sequence*) (Rife, Vanderkooy, 1989) e il metodo ESS (*Exponential Sine Sweep*) (Farina, 2000).

Per entrambi i metodi, l'operazione di trasformazione del segnale emesso in un impulso ideale è particolarmente semplice, in quanto risulta matematicamente determinato un segnale *inverso* $z[l]$, tale che la sua convoluzione con il segnale di test $x[l]$ produce la funzione impulso ideale δ di Dirac

$$x[l] * z[l] \Rightarrow \delta[l]. \quad (16.6.2)$$

Tale processo di *re-impacchettamento* del segnale di test è una *deconvoluzione*. Applicando il processo di deconvoluzione, non al segnale di test $x[l]$, ma al segnale ricevuto dal microfono $y[l]$, si ricava la risposta all'impulso $h[l]$ del sistema:

$$\begin{aligned} y[l] &= x[l] * h[l] \\ y[l] * z[l] &= x[l] * z[l] * h[l] = \delta[l] * h[l] = h[l]. \end{aligned} \quad (16.6.3)$$

16.6.1 I segnali MLS

Un segnale particolarmente efficace per la misurazione della risposta all'impulso proviene dalla teoria dei numeri e, come accennato in precedenza, prende il nome di MLS. Il segnale, percepito in modo molto simile a un rumore bianco stazionario, è una sequenza

binaria costruita tramite uno *shift register* (registro a scorrimento), la cui struttura è schematizzata in Fig. 16.6.2.

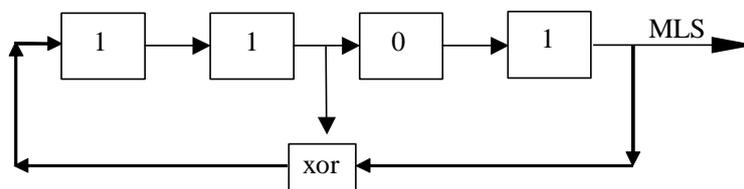


Fig. 16.6.2 Esempio di *shift register* a 4 bit; i numeri binari scorrono verso destra ad ogni passo temporale.

Con opportuni valori iniziali e un corretto posizionamento dello XOR, si ottiene un segnale MLS di lunghezza massima, pari a $L = 2^n - 1$ campioni, dove n è il numero di celle dello shift register (Borish, Angell, 1983). Le sequenze MLS godono di alcune proprietà notevoli:

a. poiché la sequenza è nota in partenza in maniera deterministica, si trova matematicamente il segnale inverso $z[i]$ come sequenza inversa nel tempo della sequenza originale: $z[i] = x[L-1-i]$;

b. essendo inoltre $z[i]$ un segnale binario fatto solo di 1 e 0, il prodotto di convoluzione risulta particolarmente semplice: le moltiplicazioni diventano solamente somme;

c. grazie alla semplicità di questo segnale è possibile elaborare la convoluzione con il segnale inverso $z[n]$ direttamente nel dominio del tempo, grazie ad un particolare algoritmo chiamato *Hadamard Fast Transform*;

d. essendo lo spettro sonoro dell'MLS *piatto* come quello del rumore bianco, è possibile calcolare anche l'analisi in frequenza senza necessità di applicare alcuna correzione spettrale (equalizzazione).

In acustica si usano normalmente shift register con un numero n di celle compreso fra 16 e 20. Nel caso di 16 celle, la sequenza MLS risulta di 65535 campioni. Spesso le sequenze vengono poi ripetute nel tempo, allo scopo di effettuare medie su più cicli, aumentando il rapporto segnale/rumore.

La tecnica MLS esiste dal 1975, ed è stata sicuramente la più usata fino a circa il 2000. Ciò è anche da attribuire alla diffusione di una specifica scheda di acquisizione, che, a partire dal 1989, implementava per via hardware lo shift register, corredata da un software chiamato MLSSA («melissa») particolarmente potente ed efficace (<http://www.mlssa.com/>).

Il problema principale della misura con tecnica MLS è dovuto alla ~~estrema~~ suscettibilità a fenomeni di distorsione non lineare eventualmente presenti nella catena elettroacustica, in particolare nei trasduttori (microfoni e, soprattutto, altoparlanti). Tali effetti non lineari causano, a valle del processo di deconvoluzione, la comparsa di *picchi spuri* nel segnale misurato, che possono apparire come echi tardivi, e che alterano irrimediabilmente la risposta all'impulso misurata, sia al fine della determinazione di grandezze acustiche oggettive, sia rispetto alla possibilità di utilizzarla come filtro numerico nei processi di auralizzazione (vedi Capitolo 20). Inoltre, l'utilizzo di medie sincrone per migliorare il rapporto

segnale/rumore, può comportare la cancellazione delle componenti ad alta frequenza, nel caso il sistema sia leggermente tempo-variante (come in realtà lo sono tutti i sistemi acustici, causa il movimento dell'aria negli ambienti e le impercettibili variazioni di temperatura). Conseguentemente, il metodo MLS può essere usato solo con altoparlanti molto fedeli, fatti funzionare ben al di sotto del loro limite di distorsione, e per un tempo non troppo lungo, all'interno di ambienti in cui le condizioni si mantengano molto stabili.

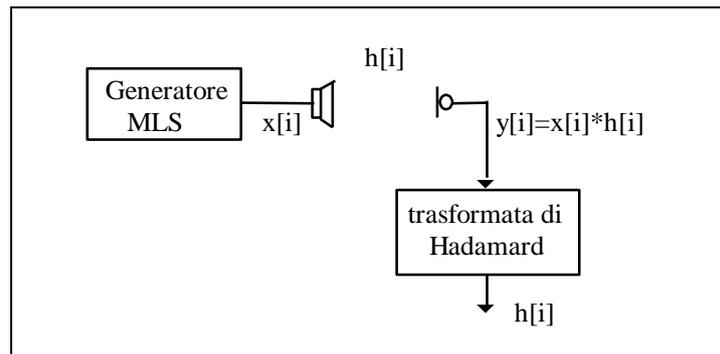


Fig. 16.6.3 - Schema di generazione, acquisizione e calcolo della risposta all'impulso con tecnica del rumore *pseudo random MLS*.

Le limitazioni della tecnica MLS sono evidenziate nell'esempio, intenzionalmente estremizzato per evidenziare le problematiche connesse, riportato in Fig. 16.6.4, dove si possono notare numerosi picchi (falsi echi) causati dalle distorsioni armoniche riscontrate durante il processo di misurazione, e provocati da fenomeni non lineari nell'elettronica di microfoni e altoparlante. Come si vedrà nel paragrafo successivo, tali fenomeni possono essere evitati utilizzando altre tecniche di deconvoluzione che consentono di separare le non linearità della catena di misura dalla risposta lineare del sistema acustico in esame.

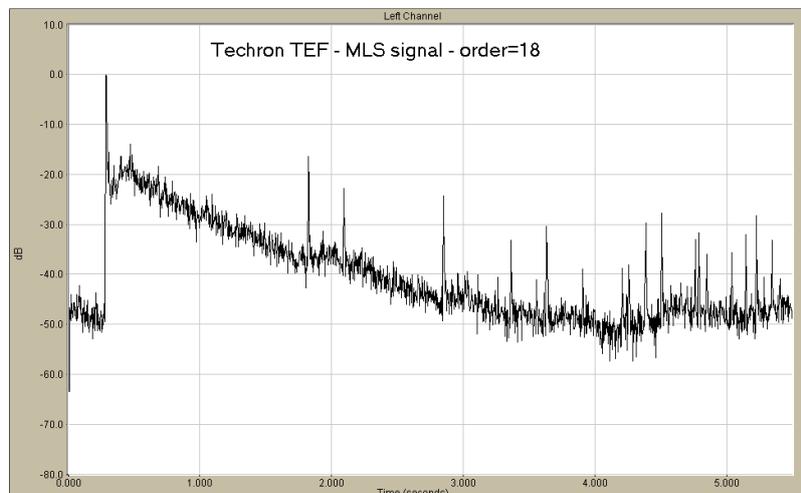


Fig. 16.6.4 Misura di risposta all'impulso con segnale MLS. Sono evidenti i fenomeni di non linearità dovuti ai trasduttori impiegati

16.6.2 Segnali sine sweep

La tecnica del segnale MLS, come accennato, non consente la separazione del contributo del sistema acustico dalle distorsioni armoniche del sistema elettronico, inoltre opera male in presenza di sistemi tempo-varianti. La separazione è invece ottenibile con la tecnica del segnale *sweep*, un segnale composto da una forma d'onda di tipo sinusoidale puro che inizia dalle frequenze basse e sale progressivamente verso le alte, al trascorrere del tempo (Müller, Massarani, 2001). Esso presenta un particolare vantaggio, simile a quello del segnale MLS: il suo filtro inverso $z[i]$ non è altro che il segnale di partenza ribaltato sull'asse dei tempi. Cioè

$$\begin{aligned} x[i] &= \text{segnale di sweep} \\ z[i] &= x \text{ rovesciato nell'asse dei tempi.} \end{aligned} \quad (16.6.4)$$

Pertanto la determinazione numerica di $z[i]$ è immediata e precisa. Purtroppo alla facilità della determinazione di $z[i]$ non sono associati gli altri vantaggi numerici posseduti dal segnale MLS mediante la trasformata di Hadamard. Tuttavia, se negli anni Novanta le possibilità di calcolo numerico, assai limitate, costringevano a sacrificare tecniche analitiche raffinate a vantaggio di metodologie meno precise, ma richiedenti limitata potenza di calcolo, lo sviluppo dei processori e la messa a punto di efficienti algoritmi di convoluzione hanno successivamente consentito l'implementazione di procedimenti computazionali molto più pesanti. I segnali *sweep*, pur necessitando di una vera convoluzione per fornire la risposta all'impulso del sistema secondo la (16.6.3), non soffrono delle limitazioni tipiche del segnale MLS, soprattutto in presenza di sistemi non lineari e non tempo-invarianti.

Tra i vari segnali *sweep* utilizzati, particolare attenzione merita lo *sweep* esponenziale (ESS, *Exponential Sine Sweep*), il cui sonogramma è rappresentato in Fig. 16.6.5: la

frequenza *istantanea* del segnale sinusoidale cresce esponenzialmente con il trascorrere del tempo, e la velocità con cui aumenta risulta costante su un asse logaritmico della frequenza. In altre parole, la frequenza raddoppia (incrementa di una ottava) ad intervalli temporali costanti.

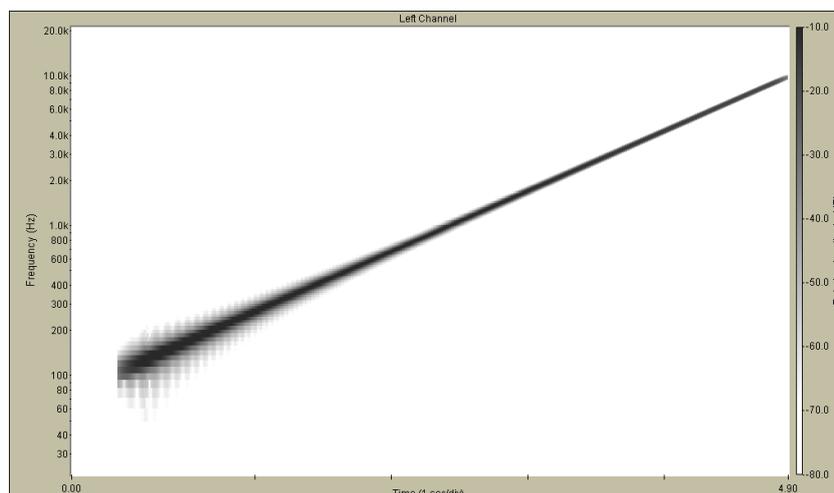


Fig. 16.6.5 Rappresentazione nel dominio tempo-frequenza (sonogramma) di un segnale sweep esponenziale.

La relazione che fornisce la rappresentazione analitica del segnale sweep esponenziale di lunghezza L campioni, nel campo di frequenza compreso tra i limiti di frequenza f_{\min} e f_{\max} , è la seguente:

$$x(t) = \sin \left[\frac{2 \cdot \pi \cdot f_{\min} \cdot L}{f_{\text{samp}} \cdot \ln \left(\frac{f_{\max}}{f_{\min}} \right)} \cdot \left(\left(\frac{f_{\max}}{f_{\min}} \right)^{\frac{i}{L}} - 1 \right) \right]. \quad (16.6.5)$$

Applicando il segnale ESS ad un sistema elettroacustico con distorsione non lineare, si sviluppa il processo della *distorsione armonica*, consistente nella creazione di numerose armoniche del segnale di ingresso: in particolare, ad ogni istante, oltre al segnale alla frequenza istantanea di ingresso f_1 , compaiono anche componenti multiple di frequenza f_2 , f_3 , f_4 ecc. In Fig. 16.6.6 è riportato il sonogramma della risposta del sistema elettroacustico, ossia il segnale $y[l]$: si osserva che le distorsioni armoniche prodotte dalle non linearità dei trasduttori producono delle bande diagonali, aventi la stessa pendenza sul piano tempo-frequenza (quest'ultima su scala logaritmica) del segnale di ingresso.

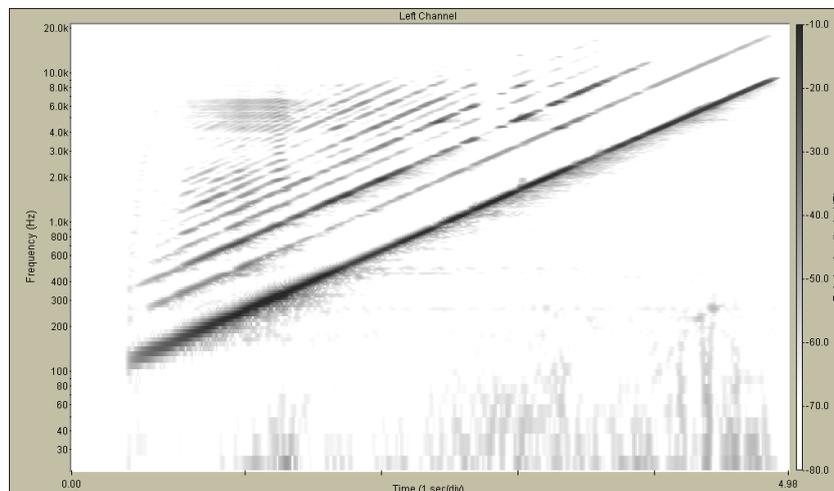


Fig 16.6.6 Segnale *sweep* esponenziale in uscita dal sistema elettroacustico.

Nel caso del segnale *sweep* esponenziale, il *filtro inverso* $\hat{z}[j]$ da utilizzare per la deconvoluzione, costituito dal segnale di test originario ribaltato lungo l'asse dei tempi, deve essere modificato. Infatti, il solo ribaltamento nell'asse dei tempi comporterebbe un pesante filtraggio della risposta in frequenza misurata, in quanto già di per sé il segnale di test possiede uno spettro che non è piatto, ossia simile ad un rumore bianco; ha invece uno spettro rosa, cioè un spettro il cui livello sonoro decresce proporzionalmente con la frequenza, con pendenza negativa di 3 dB/ottava.

Il corrispondente filtro inverso, pertanto, deve contenere uno spettro in grado di compensare l'andamento del segnale di ingresso. Tale compensazione si ottiene applicando un filtraggio delle ampiezze con pendenza positiva di 6 dB/ottava (che equivale matematicamente ad una derivazione rispetto al tempo). In tal modo, il segnale *sweep* inverso avrà uno spettro con pendenza positiva di 3 dB/ottava, coerente con la pendenza dello spettro del segnale di test. A seguito del *time reversal* lungo l'asse dei tempi, come in Fig. 16.6.7, tale filtraggio appare visivamente come una progressiva diminuzione dell'ampiezza allo scorrere del tempo, procedendo verso le basse frequenze.

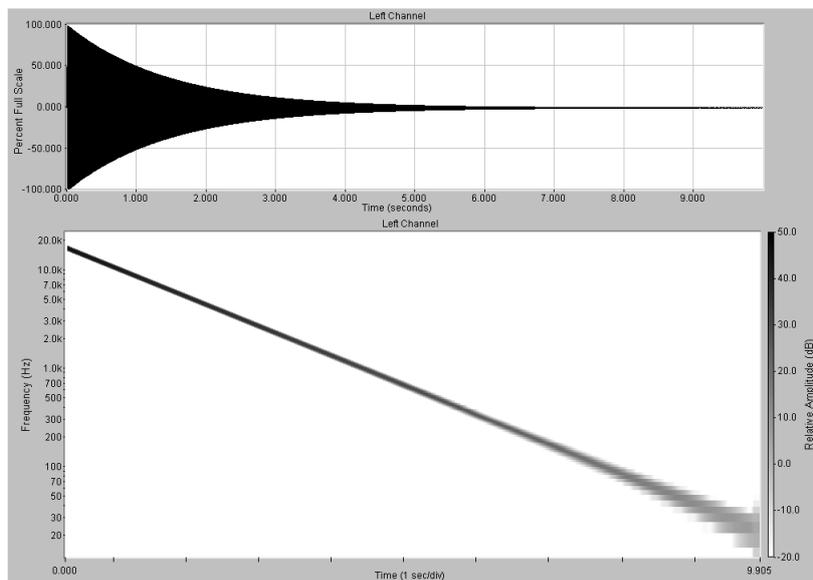


Fig. 16.6.7 *Filtro inverso* $z[i]$ del segnale *sweep* esponenziale:

diagramma nel dominio del tempo (sopra) e sonogramma (sotto)

Dal prodotto di convoluzione tra il segnale in uscita $y[i]$ e filtro *inverso* $z[i]$, si ottiene la risposta all'impulso del sistema elettroacustico $h[i]$. I vari contributi non lineari, ciascuno dei quali deriva dall'accumulazione della banda di risposta del corrispondente ordine armonico, anticipano la risposta all'impulso lineare del sistema. Tale separazione è chiaramente visibile Fig. 16.6.8, sia nel dominio del tempo, sia sul piano tempo-frequenza.

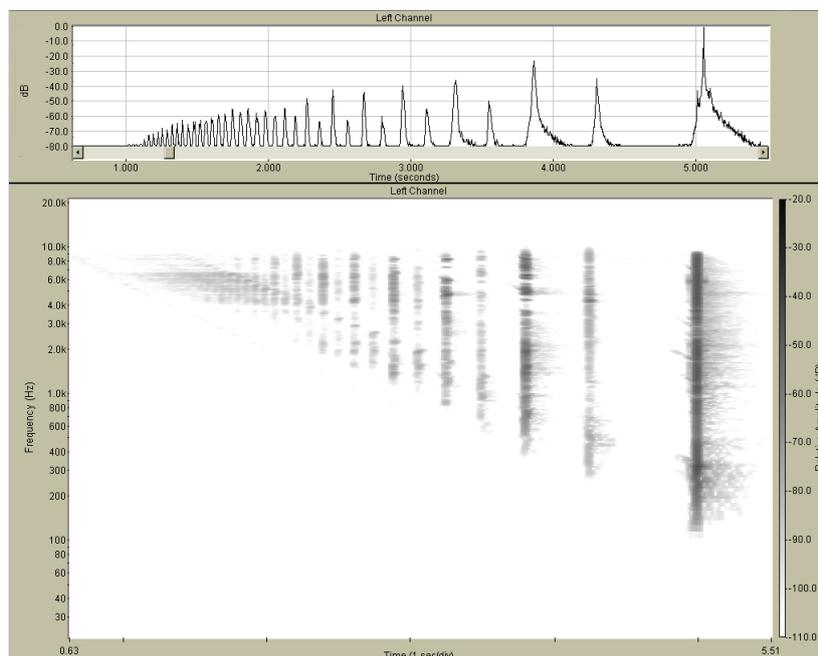


Fig. 16.6.8. Risposta all'impulso nei domini del tempo (sopra) e nel piano tempo-frequenza (sotto) ottenuta da un segnale *sweep* esponenziale.

La risposta all'impulso lineare, corrispondente al sistema acustico oggetto della misurazione, è collocata nella parte destra dei diagrammi, e preceduta dall'insieme delle distorsioni armoniche indotte dalla catena di misurazione. Nel diagramma corrispondente al dominio del tempo, rappresentato in Fig. 16.6.8 (in alto), essa è costituita dall'ultimo picco a destra, più elevato degli altri; gli impulsi precedenti, da destra verso sinistra, costituiscono le varie distorsioni armoniche, a partire dalla distorsione del 2° ordine. Nella Fig 16.6.8 (in basso) si può apprezzare il contributo in frequenza corrispondente sia alla risposta all'impulso del sistema acustico, sia alle relative distorsioni. Ovviamente, ciascun prodotto di distorsione inizia da una frequenza minima proporzionalmente più grande della frequenza f_{\min} del segnale di test ESS, mentre la frequenza massima rimane comunque limitata dall'estremo superiore del filtro inverso, coincidente con f_{\max} .

La misurazione effettuata con sweep di tipo esponenziale consente dunque non solo di estrarre la risposta all'impulso lineare, corrispondente al sistema acustico, priva di qualsiasi artefatto causato dalle eventuali distorsioni nei trasduttori, ma anche di misurare la risposta non lineare del sistema, quantificando la risposta all'impulso di ciascun ordine armonico e dunque misurando tali distorsioni, cosa particolarmente utile nello studio di altoparlanti.

Inoltre, la stessa misura della sola risposta all'impulso lineare avviene con miglior rapporto segnale/rumore e completa immunità da artefatti causati dalla non perfetta tempo-invarianza del sistema, poiché non vengono effettuate medie sincrone, ma si opera con un singola sweep di notevole lunghezza (anche 30 s, in alcuni casi).

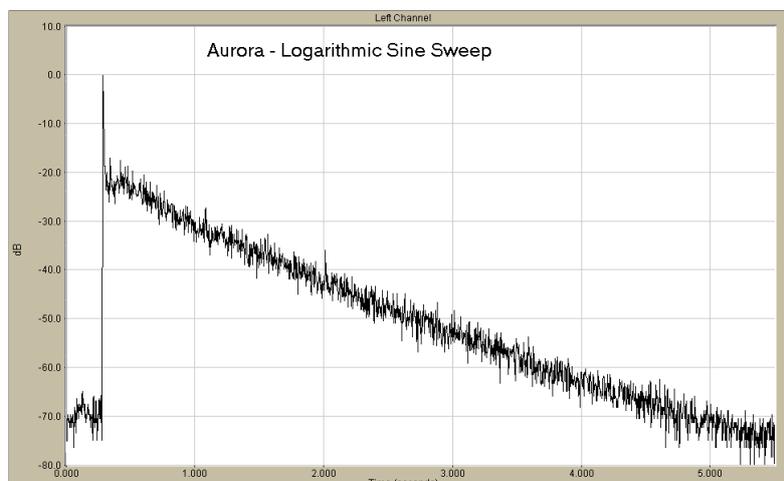


Fig. 16.6.9 Misura di risposta all'impulso con segnale ESS.

Confrontando il risultato di Fig. 16.6.9, che rappresenta la misurazione di risposta all'impulso effettuata in una cattedrale, con l'analogia misurazione effettuata con la tecnica MLS, riportata precedentemente in Fig. 16.6.4, si nota il maggior rapporto segnale/rumore ottenuto con il segnale ESS, e la totale assenza di distorsioni armoniche (*falsi echi*) presenti invece nella misura MLS.

La trattazione del segnale ESS sembra un po' incompleta: bisognerebbe citare i fondamentali lavori di [Toras-Rosell, Jacobsen, 2011] e [Kemp, Primak, 2011] e aggiungere un paio di figure opportune.

16.7 Calcolo numerico dei principali parametri acustici

In questo paragrafo vengono descritte le tecniche di elaborazione di una risposta all'impulso acquisita sotto forma di segnale audio campionato, con la finalità di calcolare i principali parametri acustici (tempo di riverberazione, chiarezza ecc.) utilizzati per la descrizione delle qualità acustiche degli ambienti, in relazione alle specifiche destinazioni d'uso. Per significato e utilizzo di tali parametri, peraltro descritti nelle norme ISO 3382, si rimanda ai Capitoli 19 e 20.

16.7.1 Ricostruzione del decadimento sonoro a partire dalla risposta all'impulso

Il tempo di riverberazione può essere ottenuto a partire dalla risposta all'impulso. In particolare viene ricostruita la curva di decadimento che sarebbe stata prodotta da una sorgente di rumore in funzionamento continuo, a seguito del suo improvviso spegnimento. Il legame tra risposta all'impulso e tempo di riverberazione è stato studiato da Schröder (1965).

Il decadimento del rumore contiene fluttuazioni casuali dovute alla storia del segnale immediatamente precedente allo spegnimento della sorgente. La media temporale $\langle n^2(\tau) \rangle$ di un numero molto grande di misure del decadimento quadrato $n^2(\tau)$, è collegata alla risposta all'impulso al quadrato $h^2(\tau)$ dalla relazione:

$$\langle n^2(\tau) \rangle = \int_{\tau}^{\infty} h^2(\tau') d\tau' = \int_0^{\infty} h^2(\tau') d\tau' - \int_0^{\tau} h^2(\tau') d\tau'. \quad (16.7.1)$$

Poiché si opera su segnali campionati, gli integrali nella (16.7.1) si trasformano in semplici sommatorie:

$$\langle n^2[i] \rangle = \sum_i^{N-1} h^2[i] = \sum_0^{N-1} h^2[i] - \sum_0^{i-1} h^2[i]. \quad (16.7.2)$$

In Fig. 16.7.1 è visibile la rappresentazione grafica della cosiddetta «integrazione di Schröder». La ricostruzione della curva di decadimento può avvenire mediante l'integrazione *all'indietro* della risposta all'impulso al quadrato (curva A), fino a raggiungere il livello stazionario; oppure, procedendo per tempi crescenti, per sottrazione dal livello stazionario, calcolato a parte (la prima delle due sommatorie nel secondo termine dell'espressione 16.7.2)), dei valori assunti progressivamente dall'integrale della risposta all'impulso al quadrato (seconda sommatoria nella (16.7.2); curva B in figura).

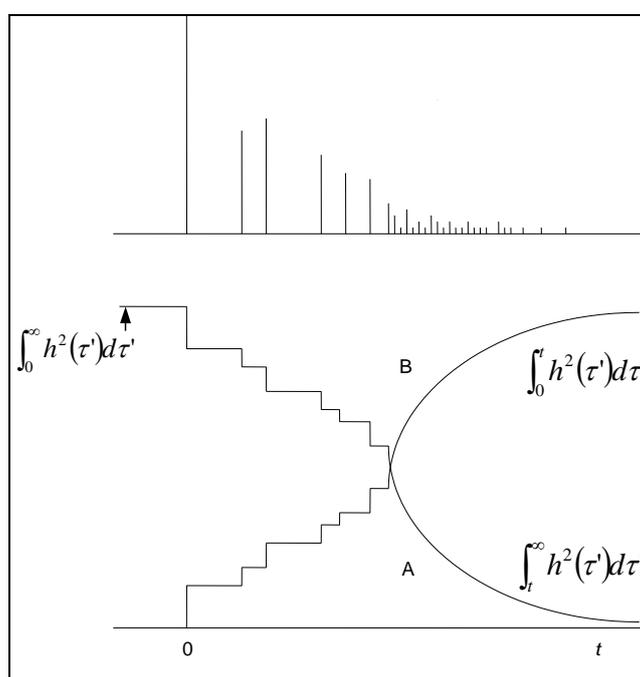


Fig. 16.7.1 Ricostruzione della curva di decadimento mediante integrazione *all'indietro*.

Trasformando il grafico in scala logaritmica, è possibile calcolare il tempo di riverberazione con gli stessi procedimenti utilizzati per le tradizionali misure del tempo di riverberazione con sorgente continua interrotta; in particolare, secondo la norma ISO 3382, il tempo di riverberazione T_{20} viene estrapolato dalla pendenza del tratto di decadimento di 20 dB a partire da -5 dB sino a -25 dB rispetto al livello stazionario di partenza. Per inciso, si ricorda che T_{20} è comunque rappresentativo di un decadimento di 60 dB, in perfetto accordo alla definizione originaria di Sabine (vedi Capitolo 19).

La norma ISO 3382 prevede esplicitamente che la misura del tempo di riverberazione venga effettuata mediante integrazione all'indietro di Schröder ogniqualvolta il segnale di eccitazione impiegato sia di tipo impulsivo o pseudo-impulsivo.

Attualmente sono disponibili strumenti di misura o programmi per PC capaci di effettuare automaticamente l'integrazione della risposta all'impulso, fornendo risultati accurati di tempo di riverberazione in funzione della frequenza, e dunque in genere perfettamente confrontabili con quelli ottenuti da misure classiche di rumore stazionario interrotto. In Fig. 16.7.2 è mostrato un esempio di curva di decadimento da rumore impulsivo e relativo integrale di Schröder.

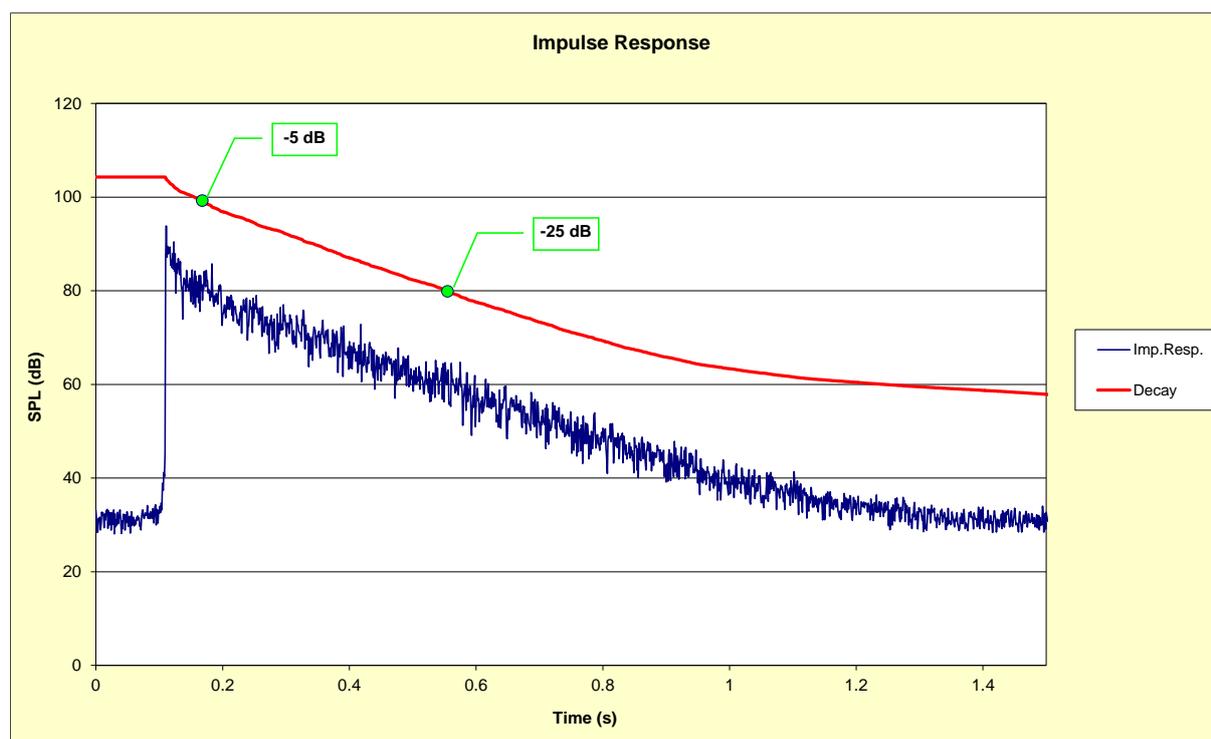


Fig. 16.7.2 Curva di decadimento da rumore impulsivo e integrale di Schröder.

In Fig. 16.7.2 si osserva che la pendenza della curva integrata cambi avvicinandosi al termine della risposta all'impulso; ciò è dovuto al fatto che viene integrato anche l'inevitabile rumore di fondo, non associate all'impulso. Questo si traduce in un secondo tratto di

decadimento, di pendenza molto inferiore a quella del primo tratto, sostanzialmente lineare, con la possibilità che possa essere confuso con un caso di decadimento a doppia pendenza (per esempio, dovuto ad ambienti acusticamente accoppiati). In realtà, la misura relativa alla figura si riferisce ad un ambiente acusticamente diffuso, in cui ci si aspetta un decadimento perfettamente rettilineo, e la doppia pendenza è dunque un tipico artefatto prodotto dall'integrazione di Schröder in presenza di un non trascurabile rumore di fondo. Il valore corretto del tempo di riverberazione si ottiene estrapolando la pendenza del primo tratto di curva, e tralasciando il secondo.

Purtroppo la norma ISO 3382 non illustra esplicitamente gli accorgimenti per evitare il problema dell'erronea integrazione del rumore di fondo, e suggerisce semplicemente un criterio empirico con cui definire l'intervallo ottimale di risposta all'impulso da integrare, in modo da limitare l'influenza del rumore di fondo. Tale criterio, tuttavia, non tiene conto del fatto che il rumore di fondo varia significativamente con la frequenza, e andrebbero quindi effettuate tante integrazioni separate alle varie frequenze, ciascuna su un segmento temporale di lunghezza diversa. In realtà, i moderni sistemi per la misura del tempo di riverberazione mediante risposta all'impulso integrata incorporano avanzate tecniche numeriche di rimozione degli effetti indesiderati del rumore di fondo, che consentono la sua eliminazione *per sottrazione* in fase di integrazione, e ottenere un decadimento molto più esteso e immune da doppie pendenze artificiali (www.aurora-plugins.com). In Fig. 16.7.3 è riportato un esempio dell'effetto della correzione dell'integrazione con sottrazione del rumore di fondo: si può osservare che il decadimento ricostruito risulta perfettamente lineare fino a oltre 50 dB al di sotto del livello stazionario, mentre in assenza della correzione il tratto di decadimento corretto si ferma a -35 dB. Le varie tecniche di correzione e ottimizzazione dell'integrale di Schröder sono presentate nel lavoro di Guski e Vörländer (2014).

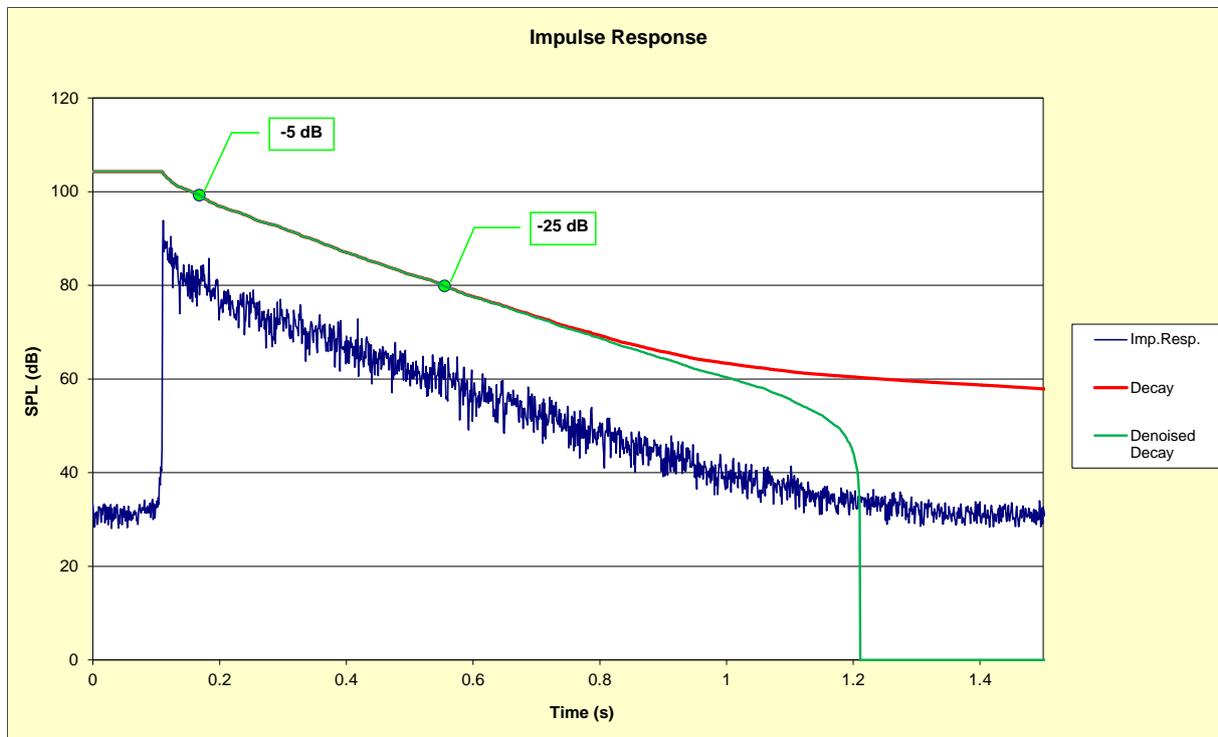


Fig. 16.7.3 Correzione dell'integrazione della risposta all'impulso con sottrazione del rumore di fondo.

L'ultimo problema da affrontare, nell'elaborazione numerica dei tempi di riverberazione, riguarda l'applicazione di un banco di filtri per ottenerne i valori in funzione della frequenza. L'uso di filtri di banda piuttosto stretta (ad esempio, bande di 1/3 d'ottava) e di ordine elevato può rendere problematica la misurazione di tempi di riverberazione relativamente brevi, tipicamente ad alta frequenza. La ragione risiede nel fatto che un filtro passa-banda di larghezza limitata presenta esso stesso, come accennato in precedenza, una *riverberazione intrinseca*, associata alla durata della propria risposta all'impulso. Ciò non costituisce un problema finché il *tempo di riverberazione* del filtro è limitato a 1/10 del tempo di riverberazione effettivo. Quando esso diventa invece confrontabile con il tempo di riverberazione da misurare, il risultato dell'elaborazione tende ad essere alterato da una sistematica sovrastima di quest'ultimo. Per limitare il problema sono note due tecniche:

- utilizzo di filtri IIR (la cui risposta all'impulso è fortemente asimmetrica, con brevissimo avvio e lunga coda sonora tardiva); conviene applicare il filtraggio alla risposta all'impulso dopo averla rovesciata lungo l'asse dei tempi, per poi raddrizzarla una volta filtrata (di fatto, filtrandola *all'indietro*);
- utilizzo di filtri FIR a fase lineare, massimamente *impaccati* nel dominio del tempo nell'intorno dell'impulso centrale.

In Fig. 16.7.4 è riportata la risposta all'impulso di un filtro di 1/3 d'ottava con frequenza centrale di 100 Hz, di tipo IIR *rovesciato nel tempo*; la Fig. 16.7.5 mostra quella di un filtro di tipo FIR a fase lineare. Il primo dà luogo ad un tempo di riverberazione del filtro di 0,033 s, il secondo invece di 0,047 s. Se si fosse impiegato il tradizionale filtro IIR *in avanti*, senza alcun accorgimento, il suo tempo di riverberazione sarebbe stato di 0,17 s, rendendo impossibile la misura accurata di valori del tempo di riverberazione inferiori a 1,7 s.

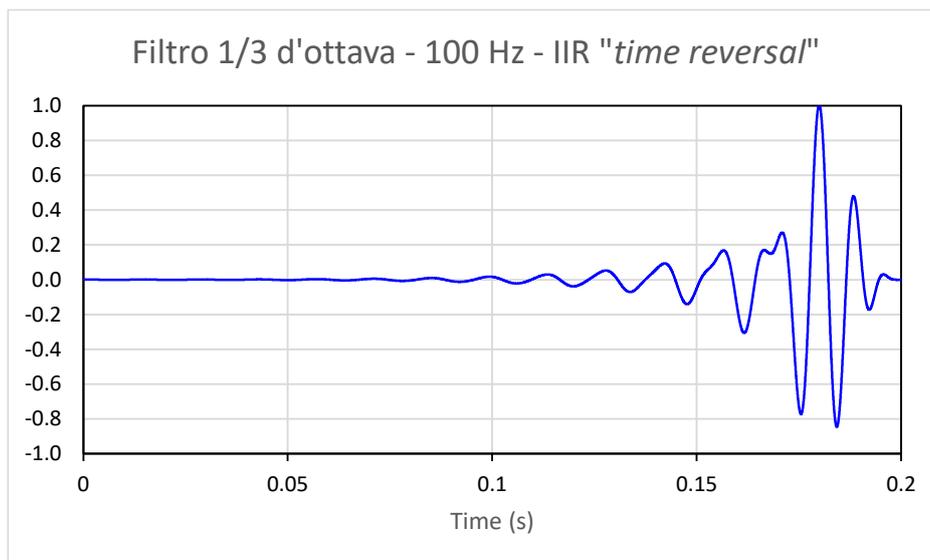


Fig. 16.7.4 Filtro IIR *rovesciato nel tempo*; 1/3 d'ottava a 100 Hz (vedi Fig. 16.5.4).

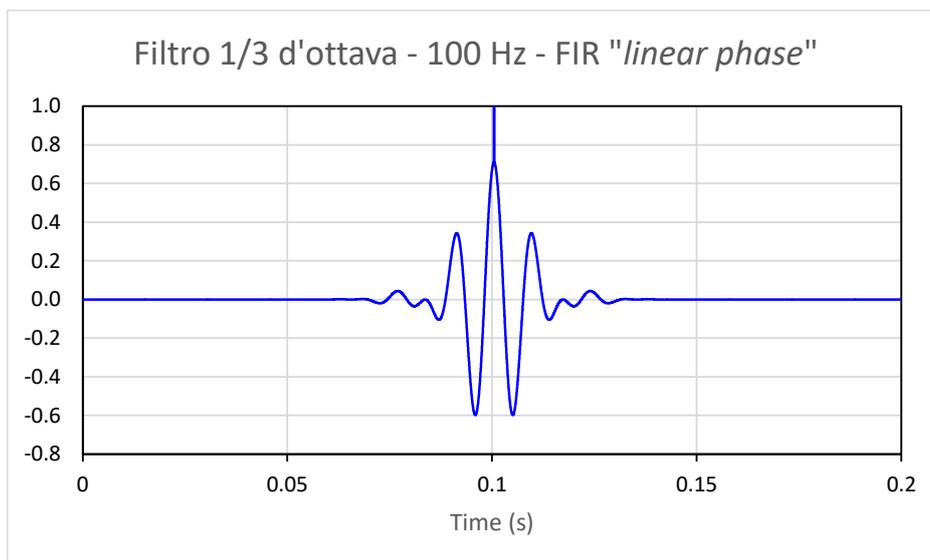


Fig. 16.7.5 Filtro FIR a fase lineare; 1/3 d'ottava a 100 Hz.

In assenza degli accorgimenti descritti, in generale non è affidabile eseguire la misura dei tempi di riverberazione per bande di 1/3 d'ottava: operando con tradizionali filtri IIR, ci si deve limitare ad un filtraggio per bande d'ottava prima di utilizzare la tecnica di integrazione all'indietro.

16.7.2 Calcolo dei parametri descrittivi della qualità acustica di ambienti

I parametri utilizzati per la descrizione della qualità acustica di spazi chiusi, sale per la musica o per altre destinazioni d'uso, come indicato dalla norma ISO 3382, sono definiti sulla base dell'integrazione dell'energia sonora, intesa (impropriamente) come somma dei quadrati dei valori dei coefficienti della risposta all'impulso $h[i]$, integrando su segmenti temporali ben definiti a partire dall'istante di arrivo del suono diretto.

Ad esempio, l'*Indice di chiarezza* C_{80} (vedi par. 19.2) è definito come il rapporto in scala logaritmica fra energia iniziale, che arriva nel punto di ricezione entro i primi 80 ms, ed *energia tardiva*, che giunge successivamente:

$$C_{80} = 10 \cdot \log_{10} \left[\frac{\int_0^{80ms} h^2(\tau) \cdot d\tau}{\int_{80ms}^{\infty} h^2(\tau) \cdot d\tau} \right]. \quad (16.7.2)$$

Operando con segnali campionati, parrebbe dunque del tutto ovvio applicare direttamente la definizione (16.7.2), sostituendo due sommatorie ai due integrali, avendo cura di definire gli indici di inizio e fine della somma in modo che corrispondano al campione di arrivo del fronte di salita del suono diretto (facilmente individuabile, rilevando il superamento di una soglia prefissata) e ad un successivo campione corrispondente ad un ritardo di 80 ms. Tale approccio funzionerebbe correttamente solo se si elaborasse il segnale a banda larga, mentre la norma ISO 3382 prevede che, come per i tempi di riverberazione, anche per i descrittori acustici si operi un filtraggio del segnale per bande d'ottava, e se ne calcoli dunque un valore a ciascuna frequenza.

Tuttavia, come discusso nel par. 16.5 (e mostrato in Fig. 16.5.4), un filtraggio passa-banda determina una sorta di *prolungamento* nel tempo del segnale, dovuto al fatto che il filtro di banda d'ottava avrà esso stesso una risposta all'impulso di una certa lunghezza, che cresce al diminuire della frequenza di centro banda. Pertanto, nell'elaborare i valori dei parametri acustici in bande d'ottava, soprattutto a bassa frequenza, se non si tenesse conto di questo effetto si andrebbe incontro ad un errore sistematico.

Per risolvere il problema, la norma prevede due possibili modalità:

- effettuare anzitutto la separazione fra porzione iniziale e tardiva della risposta all'impulso, quindi applicare a ciascuna di esse, separatamente, il filtraggio in banda d'ottava, avendo cura di integrare l'uscita del filtro per un tempo uguale alla lunghezza del segmento analizzato aumentata di almeno il doppio della lunghezza

La voce viene simulata dalla modulazione in ampiezza di un segnale *portante*, il rumore rosa filtrato, in modo da assumere un andamento spettrale rappresentativo appunto della voce umana maschile o femminile. Nel punto di ricezione, il segnale ha una modulazione meno profonda; le *valli* vengono infatti riempite in parte dalla coda riverberante dei massimi che le precedono, e in parte dal rumore acustico di fondo presente nell'ambiente.

Esiste una relazione matematica che consente di determinare, con ragionevole approssimazione, il valore della *MTF* in funzione del tempo di riverberazione, T_f , e del rapporto segnale/rumore, SN_f (Houtgast et al., 1980; Houtgast, Steeneken, 1985),

$$m(F) = \left[1 + \left(\frac{2 \cdot \pi \cdot F \cdot T_f}{13.8} \right)^2 \right]^{-\frac{1}{2}} \cdot \frac{1}{1 + 10^{\left(\frac{-SN_f}{10} \right)}} , \quad (16.8.1)$$

Questa formula è più propriamente valida assumendo l'ipotesi di campo riverberante perfettamente diffuso con decadimento perfettamente esponenziale e disregando il contributo del suono diretto dalla risposta all'impulso e dal rapporto segnale/rumore. (Houtgast et al., 1980). Ma nella frase in alto c'è scritto che è usabile con ragionevole approssimazione.

dove il primo fattore rappresenta l'effetto della coda sonora (riverberazione), e il secondo l'effetto del rapporto segnale/rumore, $SN_f = L_{signal,f} - L_{noise,f}$, nella banda di frequenza d'ottava, f , considerata. Va osservato che il secondo fattore, relativo al rapporto segnale/rumore, non dipende dalla frequenza di modulazione F , mentre il primo, legato alla riverberazione, fornisce valori via a via decrescenti al crescere della stessa frequenza F .

Oltre alla stima approssimata dei valori di *MTF*, in base alla (16.8.1), è anche possibile, almeno in linea di principio, la loro misurazione diretta con una sorgente di rumore filtrato e modulato. Al variare della frequenza di modulazione tra 0,63 Hz e 12,5 Hz e della frequenza di centro banda della portante (rumore rosa filtrato in banda d'ottava) fra 125 Hz e 8 kHz, si ottiene un insieme di 98 valori di *MTF*, corrispondenti a 14 valori per ciascuna delle 7 bande d'ottava. Successivamente, i valori vengono mediati per bande d'ottava, e infine, applicando opportuni coefficienti di pesatura spettrale (che dipendono dalla voce maschile o femminile del parlatore), si ottiene un solo valore, lo *STI* cercato

Qualora i calcoli vengano limitati alle sole due bande di ottava di 500 Hz e di 2 kHz, con soli 4 valori della frequenza di modulazione nella prima banda e 5 nella seconda, la media risultante viene chiamata *Rapid-STI*, ossia *RASTI*. Altre versioni ridotte sono lo *STITEL* (si usa una sola frequenza di modulazione per ciascuna delle 7 bande d'ottava) e lo *STIPA* (due frequenze di modulazione per ciascuna delle 7 bande d'ottava). Nelle Tab. 16.8.1, 16.8.2, 16.8.3, sono riportate le matrici ridotte *RASTI*, *STITEL* e *STIPA*.

La nuova norma IEC 60268-16 (2011) dice che il RASTI è diventato obsoleto e non dovrebbe più essere usato. Il metodo però è previsto in Aurora e dunque è presente nelle immagini che seguono che si riferiscono all'applicazione del programma. Bisognerebbe anche citare Aurora.

I metodi rapidi, *RASTI*, *STITEL* e *STIPA*, consentono una misurazione secondo la definizione, ovvero generando davvero il rumore rosa filtrato in bande, la cui intensità viene

modulata con legge sinusoidale, mentre la valutazione completa dell'intera matrice di 98 valori di MTF (e dunque del vero STI) richiede viceversa tempi proibitivi, pertanto in genere non viene praticata.

È tuttavia possibile effettuare il calcolo della MTF , e conseguentemente dello STI , a partire dalla risposta all'impulso $h_f(\tau)$, opportunamente filtrata per ciascuna banda d'ottava, secondo la formulazione di Schröder (1981):

$$m'(F) = \frac{\int_0^{\infty} h_f^2(\tau) \cdot e^{-j \cdot 2 \cdot \pi \cdot F \cdot \tau} \cdot d\tau}{\int_0^{\infty} h_f^2(\tau) \cdot d\tau} \quad (16.8.2)$$

In questa relazione e nella successiva mancano al numeratore, più correttamente, le due barre per indicare il modulo della trasformata di Fourier del quadrato della risposta all'impulso.

Il valore di $m'(F)$ non tiene però conto dell'effetto del rumore di fondo, ma solo della *coda riverberante* e del massimo precedente, e corrisponde dunque al primo dei due fattori nella (16.8.1). Volendo dunque ricavare il valore di MTF a partire da misure della risposta all'impulso, il valore di $m'(F)$ va corretto (ridotto) per tenere conto del rapporto segnale/rumore:

$$m(F) = \frac{\int_0^{\infty} h_f^2(\tau) \cdot e^{-j \cdot 2 \cdot \pi \cdot F \cdot \tau} \cdot d\tau}{\int_0^{\infty} h_f^2(\tau) \cdot d\tau} \cdot \frac{1}{1 + 10^{\left(\frac{L_{noise} - L_{signal}}{10}\right)}} \quad (16.8.3)$$

Se per la registrazione è possibile utilizzare un normale microfono omnidirezionale, come sorgente sonora occorre invece utilizzare una cosiddetta *bocca artificiale*, che riproduce le caratteristiche della voce umana in termini di direttività, andamento spettrale e potenza sonora; in Fig. 16.8.2 ne sono riportati tre diversi modelli.

Tab. 16.8.1 Matrice *MTF* ridotta (RASTI).

Modulation Frequency F [Hz]	Octave Band Frequency [Hz]	
	500	2k
0.7		•
1.0	•	
1.4		•
2.0	•	
2.8		•
4.0	•	
5.6		•
8.0	•	
11.2		•

Tab. 16.8.2 Matrice *MTF* ridotta (STITEL).

Modulation Frequency F [Hz]	Octave Band Frequency [Hz]						
	125	250	500	1k	2k	4k	8k
0.71			•				
1.12	•						
1.78						•	
2.83				•			
4.53							•
6.97					•		
11.33		•					

Tab. 16.8.3 Matrice *MTF* ridotta (STIPA).

Modulation Frequency F [Hz]	Octave Band Frequency [Hz]						
	125	250	500	1k	2k	4k	8k
0.63			•				
0.80						•	
1.00		•					
1.25					•		
1.60	•						
2.00				•			
2.50							•
3.15			•				
4.00						•	
5.00		•					
6.25					•		
8.00	•						
10.00				•			
12.50							•

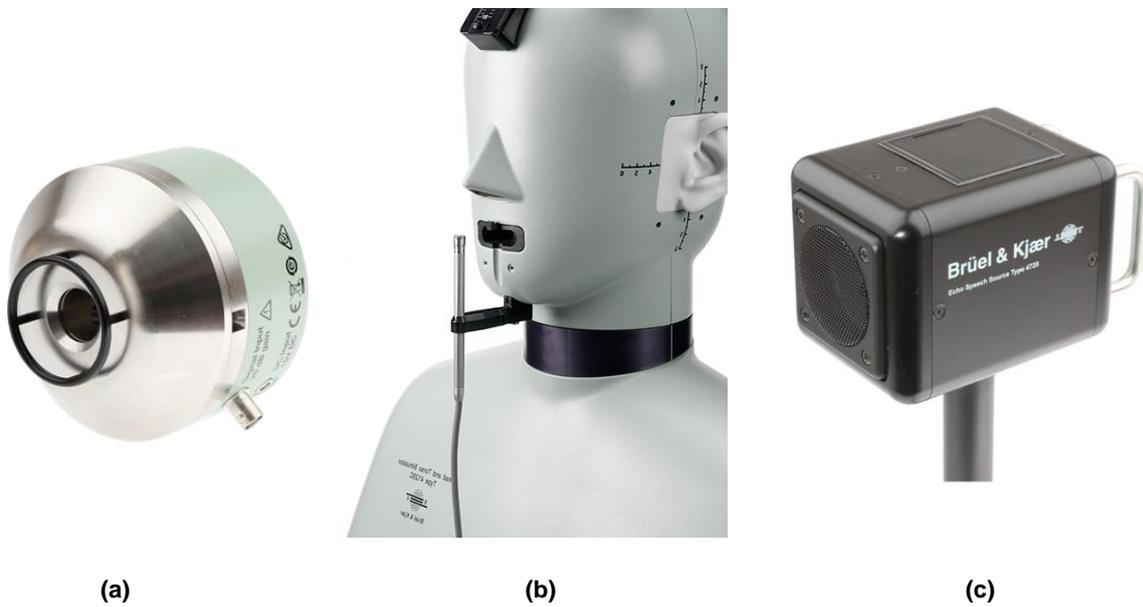


Figura 16.8.2 Sorgenti Brüel & Kjær: (a) bocca artificiale, mod. 4227, (b) manichino, mod. 4128 e (c) sorgente compatta, mod. 4720.

Solo utilizzando una sorgente sonora con direzionalità appropriata, ovviamente posizionata nella stessa postazione e orientazione del vero parlatore, si ottiene una realistica distribuzione del suono nell'ambiente in esame, e diventa possibile misurare il valore di STI in tutti i punti, per mapparne la distribuzione spaziale. In Fig. 16.8.3 è riportato un esempio di mappatura relativo all'auditorium del Carmine a Parma; nel seguito i passi seguiti nel corso della misurazione.

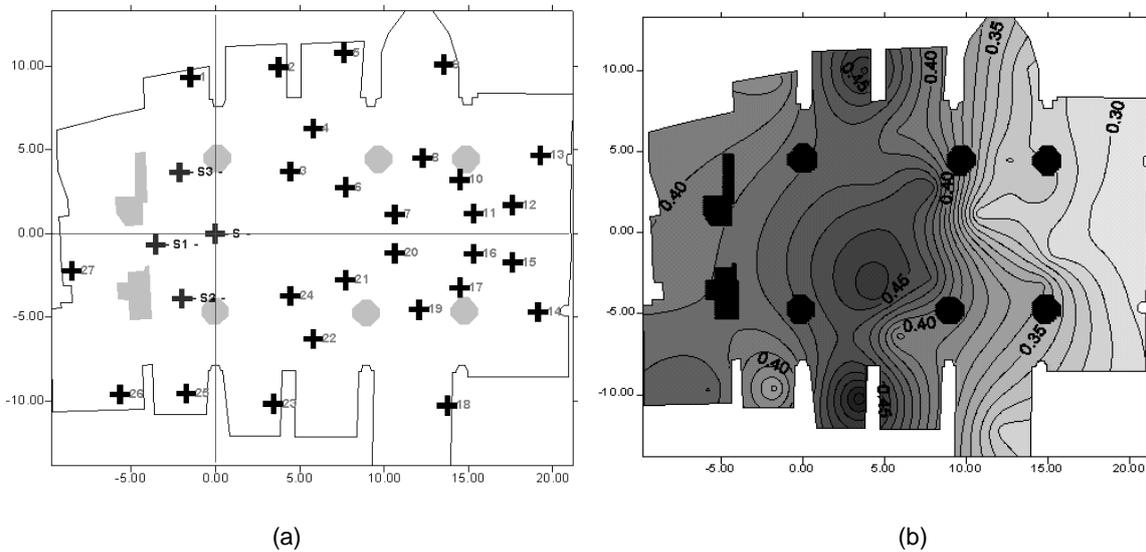


Figura 16.8.3 Auditorium del Carmine di Parma: (a) posizione della sorgente S e dei punti di misurazione; (b) mappa della distribuzione dello STI.

La sorgente sonora irradia un rumore rosa equalizzato, con livello di pressione sonora di 60 dB(A) alla distanza di 1 m, e uno spettro standardizzato rappresentativo dell'emissione sonora della voce umana maschile o femminile, come mostrato in Fig. 16.8.4. A sorgente sonora correttamente equalizzata (tipicamente per l'emissione sonora maschile), viene calibrato il microfono di misura, registrando un tratto del segnale generato da un calibratore acustico. In ciascun punto di misura nella sala vanno effettuate tre registrazioni:

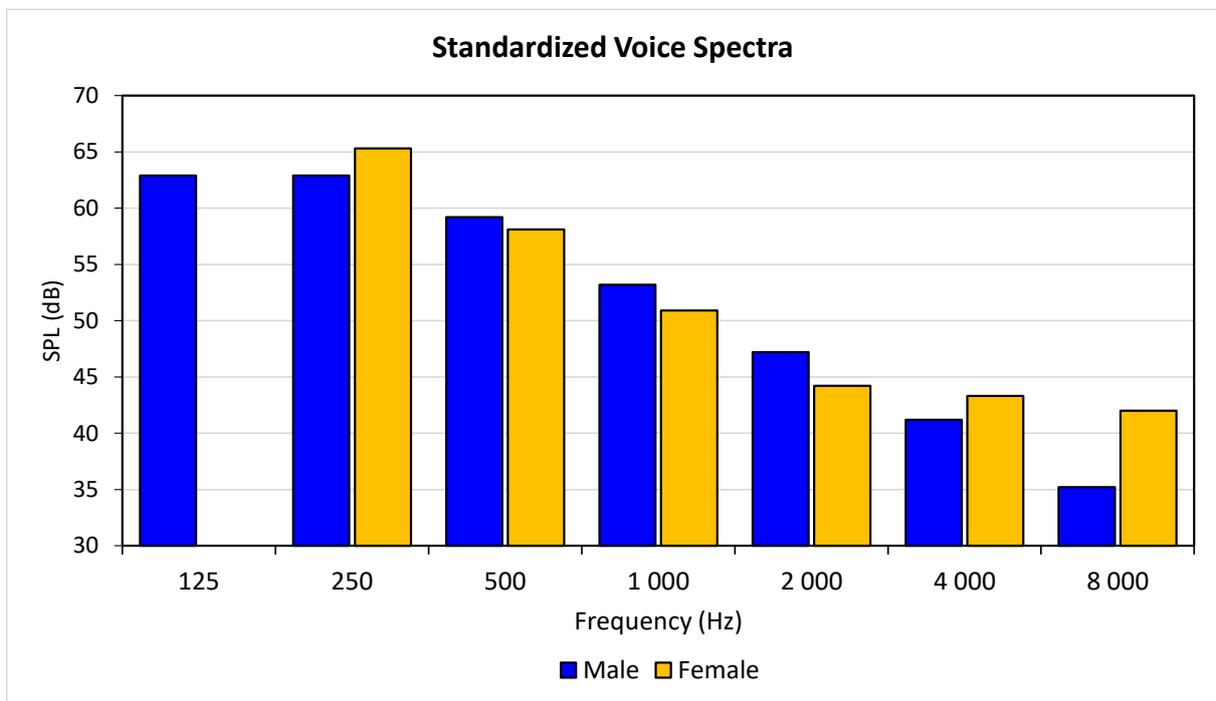


Figura 16.8.4 Spettro standardizzato della voce umana maschile e femminile.

- registrazione del rumore di fondo, con bocca artificiale spenta;
 - registrazione del rumore rosa equalizzato emesso dalla bocca artificiale;
- sembra che il livello del segnale non corrisponda a quello dell'immagine 16.8.4. Qui si riporta uno sforzo vocale normale che è piuttosto basso e difficile da rilevare in ambiente senza l'influenza del rumore di fondo. Mentre nella figura 16.8.5 il livello del parlato (nella posizione dell'ascoltatore) è 68.9 dB(A), e quindi presuppone uno sforzo vocale molto più alto.
- risposta all'impulso, ottenuta da un segnale *exponential sine sweep* emesso dalla bocca artificiale e operando l'opportuna deconvoluzione.

Il software di calcolo dello *STI* elabora le tre registrazioni, effettuando anzitutto un'analisi per bande d'ottava delle prime due, e determinando, per ciascuna banda, il valore del rapporto segnale/rumore. Viene poi analizzata la risposta all'impulso in base alla (16.8.2), infine calcolato il valore della *MTF*, inclusiva del rapporto segnale/rumore con la (16.8.3). In Fig. 16.8.4 sono riportate le tre registrazioni, ciascuna corredata della relativa analisi. In conclusione, risultano gli 98 valori di *MTF* riportati in forma di tabella nella Fig. 16.8.5 e in forma grafica in Fig. 16.8.6.

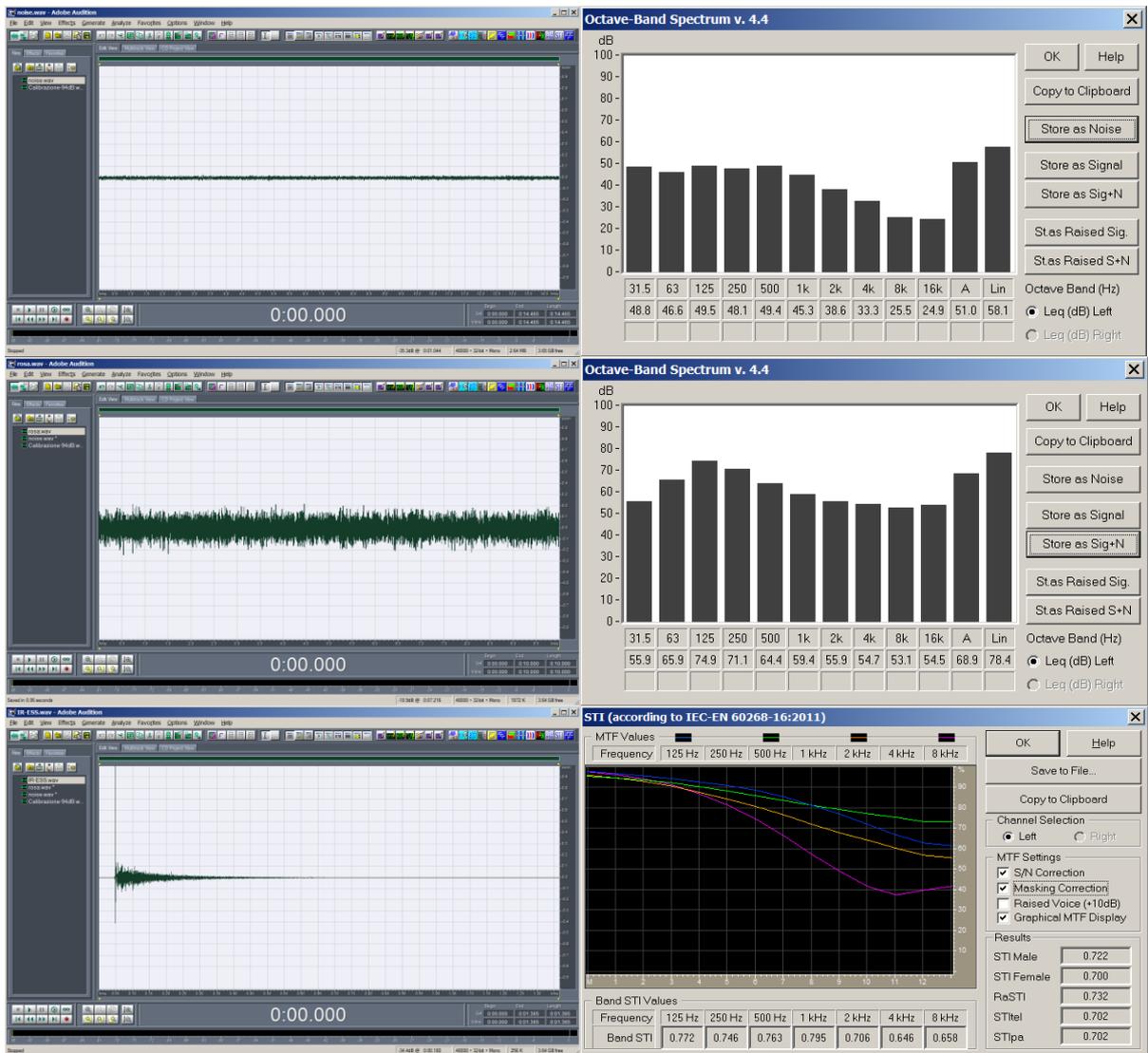


Figura 16.8.5 Le tre registrazioni effettuate: dall'alto, rumore di fondo, segnale, risposta all'impulso.

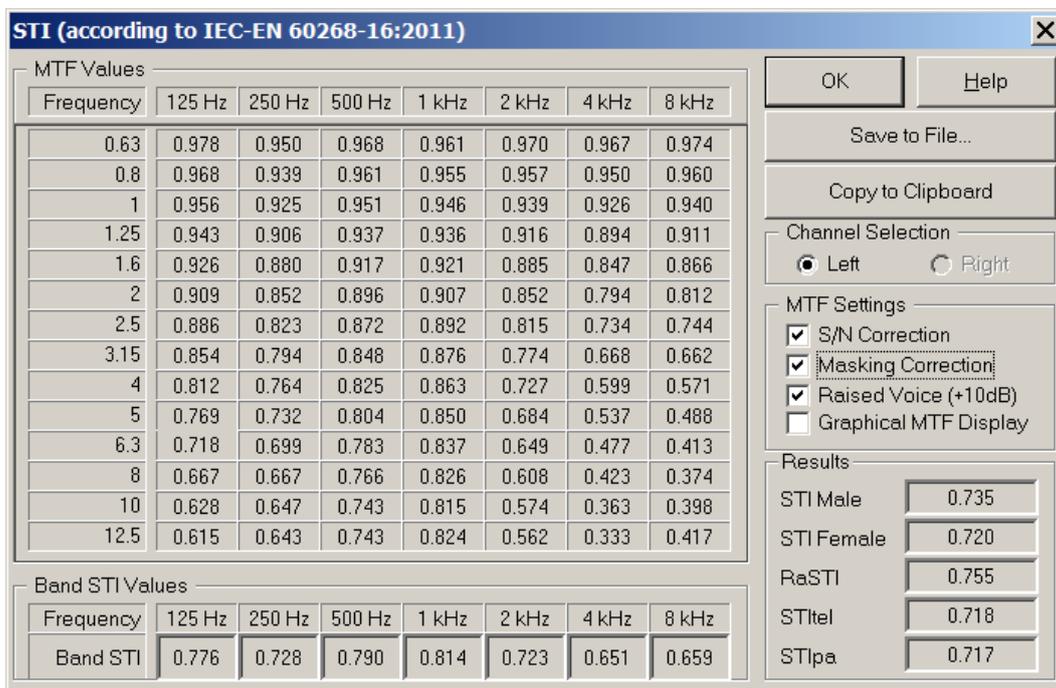


Figura 16.8.6 Valori di *MTF* in formato tabellare e calcolo dello *STI* e derivati.

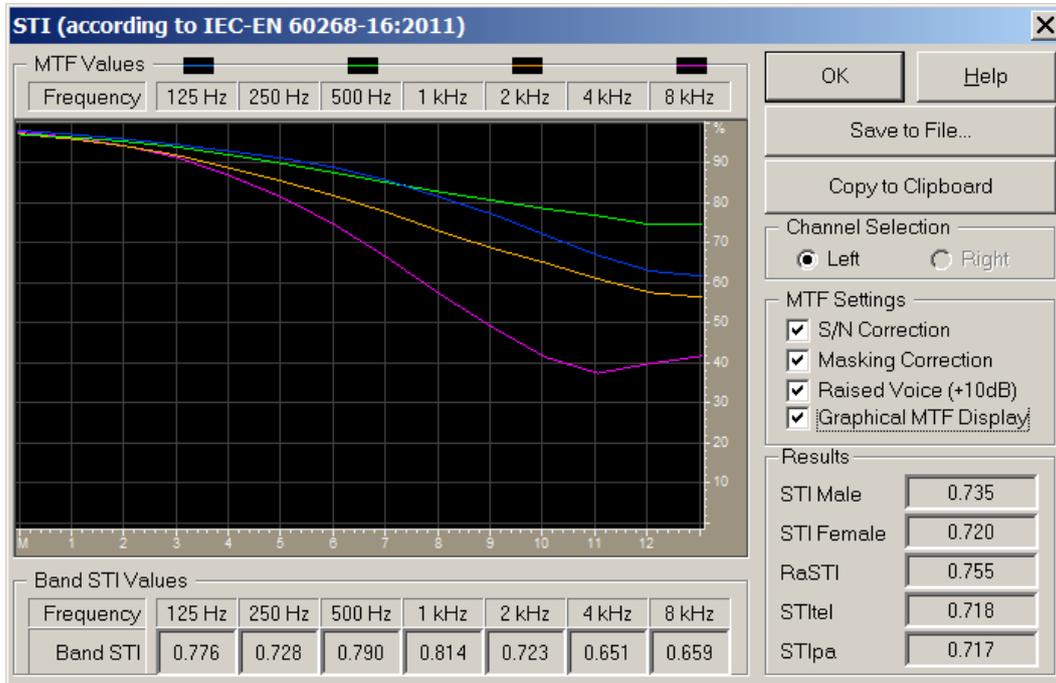


Figura 16.8.7 Valori di *MTF* in formato grafico e calcolo dello *STI* e derivati.

Se le curve della *MTF* sono sostanzialmente piatte, significa che il valore di *STI* è penalizzato dal rapporto segnale/rumore (che, come già osservato, non è influenzato dalla frequenza di modulazione). Se invece le curve partono dal valore 1 e scendono progressivamente al crescere della frequenza di modulazione *F* (come in effetti risulta in Fig. 16.8.6), il valore di *STI* è penalizzato principalmente da fenomeni di riverberazione e/o di eco.

È anche possibile, senza dover nuovamente eseguire la misura, valutare di quanto migliorerebbe lo *STI* nel caso in cui il parlatore alzasse il proprio volume vocale (*raised voice*), ipotizzando un incremento del rapporto segnale/rumore di 10 dB a tutte le frequenze. Viene inoltre effettuata simultaneamente anche l'analisi dello *STI female* (il software corregge automaticamente lo spettro del segnale) e delle versioni rapide RASTI, STITEL e STIPA.

Poiché le tre registrazioni sono disponibili separatamente, diventa anche possibile effettuare eventuali simulazioni mirate, per esempio riducendo il rumore di fondo (a seguito di un ipotetico intervento di bonifica acustica), aumentando e riequalizzando il segnale (prevedendo l'effetto di un sistema di amplificazione elettroacustica), o abbassando la pendenza della risposta all'impulso (con un trattamento fonoassorbente capace di ridurre il tempo di riverberazione).

L'analisi dello *STI* basata sulla misura separata di rumore di fondo, segnale e risposta all'impulso consente pertanto non solo di analizzare lo stato di fatto e comprendere l'origine degli eventuali problemi di intelligibilità, ma anche di progettare e verificare l'efficacia di possibili interventi correttivi. Sono questi vantaggi ad aver determinato il fatto che la tecnica di misura convenzionale, con generazione di rumore rosa filtrato e modulato, sia ormai sostanzialmente caduta in disuso, a favore del metodo basato sulla risposta all'impulso qui descritto.

BIBLIOGRAFIA

- OPPENHEIM A.V., SCHAFER R.W., 1975, *Digital signal processing*, Prentice Hall Inc., Englewood Cliffe, New Jersey.
- TORGER A., FARINA A., 2001, *Real-time partitioned convolution for Ambiophonics surround sound*, IEEE Workshop Applic. Signal Proc. Audio Acoust., Mountain House New Paltz, New York.
- SMITH III J.O., 2007, *Introduction to Digital Filters with Audio Applications*, <http://ccrma.stanford.edu/~jos/filters/>.
- GARDNER W.G., 1995, *Efficient convolution without input-output delay*, «J. Audio Eng. Soc.», 43, pp. 127-36.
- BARRY K.D., 1988, *Digital Equalization Using Fourier Transform Techniques*, 85th AES Convention, preprint 2694.
- DOUGLAS D.R., VANDERKOOY J., 1989, *Transfer-Function Measurement with Maximum-Length Sequences*, «J. Audio Eng. Soc.», 37, pp. 419.
- FARINA A., 2000, *Simultaneous measurement of impulse response and distortion with a swept-sine technique*, 110th AES Convention, Paris.

- SCHRÖDER M.R., 1965, *A new method of measuring reverberation time*, «J. Acoust. Soc. Am.», 37, pp. 409–12. 1981,
- SCHRÖDER M.R., 1981, *Modulation Transfer Functions: Definition and Measurement*, «Acustica», Volume 49, Number 3, 1 November 1981, pp. 179-182.
- COOLEY J.W., TUKEY J.W., 1965, *An algorithm for the machine calculation of complex Fourier series*, «Math. Computation», 19, pp. 297-301.
- BORISH J., ANGELL J.B., 1983, *An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise*, «J. Audio Eng. Soc.», 31, pp. 478-88.
- MÜLLER S., MASSARANI P., 2001, *Transfer-function measurements with sweeps*, «J. Audio Eng. Soc.», 49), pp. 443-71.
- SVENSSON W.P., NIELSEN J.L., 1999, *Errors in MLS Measurements Caused by Time Variance in Acoustic Systems*, «J. Audio Eng. Soc. », 47, pp. 907-27.
- TORRAS-ROSELL A., JACOBSEN F., 2011, *A new interpretation of distortion artifacts in sweep measurements*, «J. Audio Eng. Soc.», 59, pp. 283-89.
- KEMP J., PRIMACK H., 2011, *Impulse Response Measurement of Nonlinear Systems: Properties of Existing Techniques and Wide Noise Sequences*, «J. Audio Eng. Soc.», 59, p. 953.
- GUSKI M., VÖRLANDER M., 2014, *Comparison of Noise Compensation Methods for Room Acoustic Impulse Response Evaluations*, «Acta Acustica united with Acustica», 100, pp. 320-27.
- HOUTGAST T., STEENEKEN H.J.M., 1985, *The modulation transfer function in room acoustics*, «B&K Tech. Rev. », 3, pp. 3-12.
- HOUTGAST T., STEENEKEN H.J.M., PLOMP R., 1980, *Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function: I. General Room Acoustics*, «Acustica», 46, pp. 60-72.